


4-2017

# Genetic Code Expansion in Biochemical Investigations and Biomedical Applications

Nanxi Wang

University of Nebraska-Lincoln, nwang3@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/chemistrydiss>

 Part of the [Biochemical and Biomolecular Engineering Commons](#), [Biology Commons](#), and the [Biomedical Engineering and Bioengineering Commons](#)

Wang, Nanxi, "Genetic Code Expansion in Biochemical Investigations and Biomedical Applications" (2017). *Student Research Projects, Dissertations, and Theses - Chemistry Department*. 79.  
<http://digitalcommons.unl.edu/chemistrydiss/79>

This Article is brought to you for free and open access by the Chemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Student Research Projects, Dissertations, and Theses - Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

GENETIC CODE EXPANSION IN BIOCHEMICAL INVESTIGATIONS AND  
BIOMEDICAL APPLICATIONS

by

Nanxi Wang

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Jiantao Guo  
Lincoln, Nebraska

April, 2017

# GENETIC CODE EXPANSION IN BIOCHEMICAL INVESTIGATIONS AND BIOMEDICAL APPLICATIONS

Nanxi Wang, Ph.D.

University of Nebraska, 2017

Advisor: Jiantao Guo

Genetic code expansion provides a powerful tool for site-specific incorporation of unnatural amino acids (unAAs) with novel biochemical and physiological properties into proteins in live cells and organisms. To achieve this, a nonsense codon suppression system, which consists of an orthogonal aminoacyl-tRNA synthetase (aaRS) and tRNA pair that specifically decodes a nonsense codon (e.g., amber codon and quadruplet codon) with an unAA but do not “cross talk” with their endogenous counterparts, was established. This Ph.D. thesis presents our efforts on evolution and application of nonsense codon suppression systems for biochemical and biomedical investigations.

In Chapter 1, a brief overview of genetic code expansion technique and recent advances in this area of research was given. To improve unAA incorporation efficiency, we focused on systematic evolution of two most commonly used orthogonal aaRS/tRNA pairs: PylRS/tRNA<sup>Pyl</sup> and MjTyrRS/tRNA<sup>Tyr</sup>. We enhanced quadruplet codon decoding efficiency of PylRS/tRNA<sup>Pyl</sup> pairs by completely randomizing the anticodon-stem loop of tRNA<sup>Pyl</sup> (Chapter 3). In addition, we improved amber suppression efficiency of MjTyrRS/tRNA<sup>Tyr</sup> derivatives by engineering the anticodon binding pocket of MjTyrRS (Chapter 4). All these efforts lead to a further improvement in current nonsense codon suppression systems and may expand their applications in unAA mutagenesis. Next, we reported the application of an amber suppression system as an unnatural

genetic switch to manipulate the expression of essential HIV-1 proteins, which resulted in either single-cycle or multicycle live-attenuated HIV-1 viruses (Chapter 2). These genetically modified viruses can be potentially used as preventive vaccines to protect against HIV-1 infection. Our methodology can also be applied to the generation of vaccines against other pathogens.

## Table of Contents

List of Figures .....	iv
List of Tables .....	vi
Lists of Abbreviations.....	vii
<b>CHAPTER 1</b> Genetic Code Expansion: An Overview and Recent Advances in Biochemical and Biomedical Research .....	1
1.1 Introduction: an overview of genetic code expansion.....	1
1.2 Orthogonal aaRS/tRNA pairs.....	2
1.2.1 MjTyrRS/tRNA <sup>Tyr</sup> .....	4
1.2.2 PyIRS/tRNA <sup>Pyl</sup> .....	4
1.2.3 EcTyrRS/BstRNA <sup>Tyr</sup> .....	5
1.3 Nonsense Codons.....	6
1.3.1 Amber stop codon .....	6
1.3.2 Quadruplet codon.....	7
1.3.3 Rare codons.....	8
1.4 Unnatural Amino Acids .....	9
1.5 Approaches to improve the efficiency of unAA incorporation.....	12
1.5.1 Evolution of orthogonal aaRS/tRNA pairs .....	12
1.5.2 Release factor 1 knockout and an amber-free host to enhance amber suppression .....	14
1.5.3 Refining the interactions between the orthogonal tRNA and cellular components .....	15
1.5.4 Orthogonal ribosomes for enhancing nonsense codon suppression efficiency.....	16
1.6 Application of genetic code expansion in vaccine development .....	17
1.6.1 Design therapeutic vaccines using genetic code expansion.....	17
1.6.2 Design preventive vaccines using genetic code expansion.....	18
1.7 Summary and remarks .....	19
1.8 Figures and tables .....	21
<b>CHAPTER 2</b> Development of Live-attenuated HIV-1 Viruses via Genetic Code Engineering .....	25
2.1 Introduction: A novel approach towards a HIV-1 live-attenuated vaccine.....	25
2.2 Approaches .....	26
2.2.1 Single-cycle amber suppression-dependent HIV-1 variants.....	26
2.2.2 Multi-cycle amber suppression-dependent HIV-1 variants .....	34
2.2.3 Attempts to improve the replication capacity of amber suppression-dependent HIV-1 variants .....	39
2.3 Summary and remarks .....	42

2.4 Figures and tables .....	45
<b>CHAPTER 3 Systematic Evolution and Study of Quadruplet Codon Decoding tRNAs .....</b>	<b>70</b>
3.1 Introduction: Quadruplet codon decoding tRNAs for genetic code expansion .....	70
3.2 Approaches .....	71
3.2.1 tRNA (with an expanded anticodon loop) library construction and selection for UAGN (N = A, G, U, C) codon decoding. ....	71
3.2.2 Characterization and validation of evolved tRNAs. ....	73
3.2.3 Incorporation of other unAAs using the evolved tRNAs.....	74
3.2.4 Analysis of the evolved tRNA mutants.....	75
3.2.5 Cross-decoding among UAGN codons.....	75
3.2.6 tRNA (with regular 7-base anticodon loop) library construction and selection for UAGN (N = A, G, U, C) decoding. ....	76
3.2.7 Cross-decoding against UAG codons. ....	76
3.2.8 Quantification of UAGN decoding efficiency with fluorescence reporters.....	77
3.3 Summary and remarks .....	78
3.4 Figures and tables .....	80
<b>CHAPTER 4 Development of A General Approach to Evolve Aminoacyl-tRNA synthetase for Efficient Nonsense Codon Suppression.....</b>	<b>93</b>
4.1 Introduction: Evolution of aminoacyl-tRNA synthetase for more efficient unAA incorporation ...	93
4.2 Approach.....	95
4.2.1 pAcFRS library construction and selection for improving the incorporation efficiency of pAcF .....	95
4.2.2 Attempts to transfer the beneficial mutations to other aminoacyl-tRNA synthetases .....	97
4.2.2 BpaRS library construction and selection for improving the incorporation efficiency of Bpa..	97
4.2.3 sTyrRS library construction and selection for improving the incorporation efficiency of sTyr	98
4.2.4 Expression of GFPuv mutants with pAcFRS and MjtRNA <sub>CUA</sub> <sup>Tyr</sup> variants.....	99
4.2.5 Quantification of aminoacyl-tRNA synthetase expression level .....	99
4.2.6 Growth curves of cells containing the evolved aminoacyl-tRNA synthetase .....	100
4.3 Summary and remarks .....	100
4.4 Figures and tables .....	100
<b>CHAPTER 5 Experimental: Materials and methods .....</b>	<b>111</b>
5.1 General materials and methods .....	111
5.1.1 General materials .....	111
5.1.2 General methods .....	112
5.2 Experimental for Chapter 2.....	117
5.2.1 Plasmid construction.....	117

5.2.2 Transfection and generation of HIV-1 .....	122
5.2.3 Virus ultracentrifugation .....	123
5.2.4 Quantification of viral load .....	123
5.2.5 Infection assay .....	124
5.2.6 Fluorescence spectroscopy and cell imaging .....	125
5.2.7 Analysis of AzFRS transcription using quantitative PCR .....	125
5.2.8 Viral transcription and provirus detection using RNAscope and DNAscope in situ hybridization .....	126
5.2.9 Amplification and sequencing for aaRS-tRNA <sub>CUA</sub> insertion .....	126
5.2.10 Protein expression and purification.....	127
5.2.11 LC/MS/MS.....	127
5.3 Experimental for Chapter 3.....	128
5.3.1 Plasmid construction.....	128
5.3.2 Positive selection.....	131
5.3.3 Hit verification .....	131
5.3.4 Fluorescence analysis of bacterial culture.....	131
5.3.5 Protein expression and purification.....	132
5.4 Experimental for Chapter 4.....	133
5.4.1 Plasmid construction .....	133
5.4.2 Positive selection.....	134
5.4.3 Negative selection .....	134
5.4.4 Hit verification .....	135
5.4.5 Fluorescence analysis of bacterial culture.....	135
5.4.6 Protein expression and purification.....	135
<b>References.....</b>	<b>137</b>

## List of Figures

<b>Figure 1</b>	Reprogramming the genetic code.	21
<b>Figure 2</b>	An orthogonal aaRS/tRNA pair for genetic code expansion.	22
<b>Figure 3</b>	Examples of unnatural amino acids.	23
<b>Figure 4</b>	A general <i>in vivo</i> selection method for identifying evolved aaRS.	24
<b>Figure 5</b>	An unnatural genetic switch for regulation of HIV-1 replication.	45
<b>Figure 6</b>	Amber suppression of EGFP in 293T cells.	47
<b>Figure 7</b>	The HIV-1 genome organization.	48
<b>Figure 8</b>	p24 assay after transfection with pSUMA variants and pAzFRS/tRNA <sup>Tyr</sup> pair.	50
<b>Figure 9</b>	p24 assay after transfection with pSUMA variants and plodoFRS/tRNA <sup>Tyr</sup> or pAcFRS/tRNA <sup>Tyr</sup> pair.	51
<b>Figure 10</b>	Control the expression of EGFP with photocaged unAAs.	52
<b>Figure 11</b>	Control the replication of HIV-1 by amber suppression.	53
<b>Figure 12</b>	Infection of TZM-bl with pSUMA-wt.	54
<b>Figure 13</b>	Expression and infection with pNL-GI variants.	55
<b>Figure 14</b>	Infection of TZM-bl cells with pSUMA variants in the presence of TyrRS/tRNA <sup>Tyr</sup> pair.	56
<b>Figure 15</b>	Control the replication of HIV-1 with an unnatural genetic switch.	57
<b>Figure 16</b>	Infection of TZM-bl cells with HIV-1-iGFP and HIV-1-NL-GI strains.	58
<b>Figure 17</b>	Relative replication efficiency of four pNL4-3 variants.	59
<b>Figure 18</b>	Construction and examination of HIV-AzFRS-tRNA variants.	60
<b>Figure 19</b>	Infectivity assay of the pNL43-Trp36-AzFRS-tRNA-2 and the pNL43-Tyr59-AzFRS-tRNA-2 mutants in the presence and the absence of pAzF.	61
<b>Figure 20</b>	Viral RNA (vRNA) and DNA (vDNA) detection using RNAscope and DNAscope <i>in situ</i> hybridization (ISH)	62
<b>Figure 21</b>	Infection assay of the amber-codon-free pNL4-3 variants	65



<b>Figure 22</b>	Relative expression levels of pAzFRS.	66
<b>Figure 23</b>	Expression and infection of pNL4-3 variants with two genomic copies of tRNA <sup>Tyr</sup> .	67
<b>Figure 24</b>	Rescue of the pNL43-Tyr59-AzFRS-tRNA-2 infectivity by viral protein supplement.	68
<b>Figure 25</b>	Suppression of an amber codon in EGFP in the presence of different concentration of pAzF.	69
<b>Figure 26</b>	Library construction and selection for UAGN-decoding tRNA <sub>NCUA</sub> <sup>PyI</sup> variants.	80
<b>Figure 27</b>	GFPuv expression in cells containing tRNA <sub>NCUA</sub> <sup>PyI</sup> variants.	81
<b>Figure 28</b>	Purification of GFPuv-Asn149BocK mutants.	82
<b>Figure 29</b>	Mass spectrometry analysis of GFPuv containing BocK at position 149	83
<b>Figure 30</b>	UAGN codons decoding with AbK.	84
<b>Figure 31</b>	UAGN codons decoding with ONBK.	85
<b>Figure 32</b>	Cross-activity of tRNA <sub>NCUA</sub> <sup>PyI</sup> variants among UAGN codons.	87
<b>Figure 33</b>	UAGN and UAG codon decoding efficiency of tRNA <sub>NCUA</sub> <sup>PyI</sup> variants.	88
<b>Figure 34</b>	Quantification of UAGN decoding efficiency with pStrepII-GFPuv-UAGN reporters.	89
<b>Figure 35</b>	Quantification of UAGN decoding efficiency with pStrepII-GST-X-GFPuv reporters.	90
<b>Figure 36</b>	Two classical models for quadruplet codon decoding.	91
<b>Figure 37</b>	A new model for quadruplet decoding (+1 frameshift).	92
<b>Figure 38</b>	Recognition of anticodon region of Mj tRNA <sub>GUA</sub> <sup>Tyr</sup> by MjTyrRS.	101
<b>Figure 39</b>	Positive and negative selections to identify MjTyrRS mutants with enhanced unAA incorporation efficiency.	102
<b>Figure 40</b>	Characterization and validation of pAcFRS mutants.	103
<b>Figure 41</b>	GFPuv fluorescence assay of cells containing MjTyrRS mutants.	104
<b>Figure 42</b>	Characterization and validation of BpaRS mutants.	105
<b>Figure 43</b>	Characterization and validation of sTyrRS mutants.	106

<b>Figure 44</b>	GFPuv fluorescence assay of cells containing MjTyrRS and MjtRNA <sub>CUA</sub> <sup>Tyr</sup> variants.	107
<b>Figure 45</b>	Purification of GFPuv mutants in the presence of evolved pAcFRS and MjtRNA <sub>CUA</sub> <sup>Tyr</sup> variants.	108
<b>Figure 46</b>	Western blot analysis of expression level of pAcFRS variants.	109
<b>Figure 47</b>	Growth curves of cells containing MjTyrRS variants.	110

### List of Tables

<b>Table 1</b>	pSUMA variants being constructed and examined in this work.	49
<b>Table 2</b>	Evaluation of the multi-cycle infectivity of the pNL43-Trp36-AzFRS-tRNA-2 and the pNL43-Tyr59-AzFRS-tRNA-2 mutants in the presence and the absence of pAzF.	64
<b>Table 3</b>	Sequences of tRNA <sub>NCUA</sub> <sup>Py1</sup> variants that have improved UAGN decoding efficiency.	86

## Lists of Abbreviations

aaRS	Aminoacyl-tRNA synthetase
AbK	3'-azibutyl-N-carbamoyl-lysine
BocK	N <sup>ε</sup> -( <i>tert</i> -butoxycarbonyl)-L-lysine
Bpa	<i>p</i> -benzoyl-L-phenylalanine
Cm <sup>R</sup>	Chloramphenicol acetyltransferase (CAT)
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide solution mix
<i>E. coli</i>	<i>Escherichia coli</i>
GFP	Green fluorescent protein
HIV	Human immunodeficiency virus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LAV	Live attenuated vaccine
LB	Lysogeny broth
<i>Mb</i>	<i>Methanosarcina barkeri</i>
<i>Mj</i>	<i>Methanocaldococcus jannaschii</i>
<i>Mm</i>	<i>Methanosarcina mazei</i>
ONBK	<i>o</i> -nitrobenzyl-oxycarbonyl-N <sup>ε</sup> -L-lysine
ONBY	<i>o</i> -nitrobenzyl-tyrosine
pAcF	<i>p</i> -acetyl-L-phenylalanine
pAzF	<i>p</i> -azido-L-phenylalanine
PCR	Polymerase chain reaction
pIodoF	<i>p</i> -iodo-L-phenylalanine
PTM	Post-translational modification
rRNA	Ribosomal ribonucleic acid
sTyr	Sulfotyrosine
tRNA	Transfer ribonucleic acid
unAA	Unnatural amino acid

## CHAPTER 1

**Genetic Code Expansion: An Overview and Recent Advances in Biochemical and Biomedical Research****1.1 Introduction: an overview of genetic code expansion**

Genetic code expansion allows for site-specific incorporation of unnatural amino acids (unAAs), beyond the 20 canonical ones, into proteins in live cells and organisms. The major challenge for genetic code expansion is to introduce unAAs exclusively into proteins at pre-determined sites. Since the first report demonstrating that an archaeal aminoacyl-tRNA synthetase (aaRS) and tRNA pair could be used to encode unAA in response to an amber nonsense codon in *E. coli* (**Figure 1**),<sup>1,2</sup> this system has been widely adopted. Different sets of cross-species aaRS/tRNA pairs have been developed to adapt to a range of different expression hosts. Through directed evolutions and functional screenings,<sup>3,4</sup> aaRS/tRNA pairs with altered amino acid specificity have been selected to direct the incorporation of novel unAAs into proteins in both prokaryotic and eukaryotic cells, and recently animals.<sup>5,6</sup>

The ability to reprogram genetic code greatly expanded the scope of unAA mutagenesis for biochemical and biomedical research. Unlike traditional protein engineering strategies, genetic code expansion is not limited to the natural building blocks, but offers much broader structural diversity and functional versatility. Distinctive unAAs with unique biochemical and physiological properties have been installed in the target proteins using this system, such as spectroscopic probes,<sup>7,8</sup> post-translational modifications,<sup>9,10</sup> photo-crosslinkers,<sup>11</sup> photocaged or chemically-caged groups,<sup>12-14</sup> metal chelators,<sup>15</sup> and chemically reactive handles.<sup>16,17</sup> Site-specific incorporation of these unAAs enables the observation and control of target proteins under physiological conditions in live cells or even in live animals. Further enrichment in protein

complexity can be achieved by selective chemical modifications at unAA-mutation sites for the study of more complicated cellular events. Genetically encoded unAAs have taken active roles in probing protein structure and functions, tracking protein localization, monitoring protein-protein interactions, and controlling enzymatic activities.<sup>6,18-20</sup> More importantly, genetic code expansion has opened new frontiers for biomedical applications. For example, genetically encoded posttranslational modifications (PTMs) and mimics may enhance the immunogenicity of self-proteins, which makes them promising candidates for therapeutic vaccines.<sup>21-23</sup> The insertion of amber stop codons in viral genome provides a simple and effective strategy to develop preventive vaccines.<sup>24-26</sup>

To sum up, genetic code expansion has been proven particularly useful in expanding chemistry of living systems and turned into a key fundamental chemical biology tool. Herein, this chapter provides an overview of the genetic code expansion, and recent advances in the development and application of this technique.

## 1.2 Orthogonal aaRS/tRNA pairs

The key component of genetic code expansion is the orthogonal aaRS/tRNA pair (**Figure 2**). To apply the orthogonal aaRS/tRNA pair in live cells, several prerequisites needed to be fulfilled. First, no “cross-talk” should happen between this aaRS/tRNA pair and their counterparts in the host cells. To avoid misincorporation of canonical amino acids, it demands that the nonsense suppressor tRNA is only recognized by its cognate aaRS but not endogenous aaRSs. Second, the substrate specificity of aaRS should be altered towards a specific unAA while excluding all other canonical amino acids. Third, this orthogonal aaRS is required to charge the unAA onto its cognate tRNA but not onto any endogenous tRNA. In addition, the aminoacylated tRNA needs to be transported efficiently to a host ribosome or an engineered orthogonal ribosome, and positioned

correctly in response to either a nonsense triplet or a quadruplet codon to direct the incorporation of unAA into the nascent peptide. Last, the nonsense suppressor tRNA must outcompete the translation termination signal mediated by release factors in the host cells.

To date, a tyrosyl-tRNA synthetase (TyrRS)/tRNA<sup>Tyr</sup> pair derived from *Methanocaldococcus jannaschii* and a pyrrolysyl-tRNA synthetase (PylRS)/pyrrolysyl-tRNA (tRNA<sup>Pyl</sup>) pair derived from *Methanosarcina mezei* or *Methanosarcina bakeri* have emerged as the two major systems for genetic code expansion.<sup>6</sup> MjTyrRS/tRNA<sup>Tyr</sup> variants are commonly used in *E. coli*. While PylRS/tRNA<sup>Pyl</sup> variants are compatible with a much broader ranges of hosts, including both prokaryotes and eukaryotes.<sup>13,27-31</sup> One noteworthy exception demonstrated that MjTyrRS/tRNA<sup>Tyr</sup> mutants could be functional in mammalian cells through the evolution of the orthogonality-determining elements.<sup>32</sup>

To further expand the diversity of unAAs that could be incorporated into proteins in *E. coli*, orthogonal aaRS/tRNA pairs from other organisms have been developed, including a lysyl-tRNA synthetase/tRNA<sup>Lys</sup> pair<sup>33</sup> and a glutamyl-tRNA synthetase/tRNA<sup>Glu</sup> pair<sup>34</sup> derived from *Pyrococcus horikoshii* and a tryptophanyl-tRNA synthetase/tRNA<sup>Trp</sup> pair<sup>35</sup> from *Saccharomyces cerevisiae*. In yeast and mammalian cells, an *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/tRNA<sup>Tyr</sup> pair,<sup>36,37</sup> an *E. coli* leucyl-tRNA synthetase (EcLeuRS)/tRNA<sup>Leu</sup> pair,<sup>38</sup> and an *E. coli* tryptophanyl-tRNA synthetase (EcTrpRS)/tRNA<sup>Trp</sup> pair<sup>39</sup> have been utilized for unAA incorporation. In addition, the heterologous *E. coli* TyrRS (EcTyrRS)/*Bacillus stearothermophiles* tRNA (BstRNA<sup>Tyr</sup>) pair has been used as an efficient amber suppressor in mammalian cells<sup>40</sup> and transgenic animals,<sup>12</sup> which provides a valuable tool for genetic code expansion.

### 1.2.1 MjTyrRS/tRNA<sup>Tyr</sup>

The first orthogonal MjTyrRS/tRNA<sup>Tyr</sup> pair that functioned properly in *E. coli* was reported in 2000.<sup>1</sup> Although its orthogonality had yet to be optimized, this pair was used for the first *in vivo* amber suppression to incorporate *p*-methoxyphenylalanine.<sup>2</sup> The initial evolution efforts were focused on enhancing the orthogonality and suppression efficiency of MjtRNA<sup>Tyr</sup>. Later, the research emphasis was shifted to engineer MjTyrRS to accommodate unAAs with novel functions and structures.

The rapid expanded pool of MjTyrRS/tRNA<sup>Tyr</sup> variants allows for the site-specific incorporation of a large number of phenylalanine analogs with *para*-, *meta*-, or *ortho*-substitutions. For example, unAAs containing unsaturated carbon-carbon bonds, such as allyl,<sup>4</sup> propargyl,<sup>41</sup> alkynyl<sup>42</sup>, and allenyl (unpublished data) groups have been incorporated into proteins. The incorporation of unAAs bearing chemically active handles, such as aryl halides,<sup>43</sup> azide,<sup>44</sup> ketone,<sup>45</sup> and boronic acid<sup>46</sup> allowed for further modifications of unAA-containing proteins. Sulfated- or acetylated amino acids, such as sulfotyrosine (sTyr),<sup>47</sup> and *p*-acetyl-L-phenylalanine (pAcF), have been used to generate homogeneously modified proteins, which may enhance our understanding of PTMs.<sup>45</sup> Alternative aromatic rings, such as naphthyl,<sup>43</sup> hydroxyl- and methyl-coumarins<sup>48</sup> have been used as protein probes to study protein structure and functions. Photo-caged tyrosine analogues, such as *o*-nitrobenzyl-tyrosine (ONBY)<sup>49</sup> has been incorporated into proteins for light activation of enzymatic activity.

### 1.2.2 PylRS/tRNA<sup>Pyl</sup>

In nature, pyrrolysyl-tRNA synthetase (PylRS) is responsible for charging the 22<sup>nd</sup> proteinogenic amino acid pyrrolysine (Pyl) onto pyrrolysyl-tRNA (tRNA<sup>Pyl</sup>), a naturally occurring amber suppressor.<sup>50</sup>

The wild type PylRS possesses a deep and bulky hydrophobic pocket.<sup>51</sup> The PylRS active site only forms two hydrogen bonds with Pyl in addition to the hydrophobic interactions. Due to the lack of specific interactions between Pyl and the binding pocket, PylRS displays a high tolerance toward variations in the side chains of its amino acid substrates, which enabled the evolution of PylRS with an expanded substrate flexibility for genetic code expansion.<sup>52</sup> The wild-type and evolved PylRS/tRNA<sup>Pyl</sup> pairs provide a versatile system for the genetic incorporation of a large variety of unAAs into bacteria, yeasts, mammalian cells, and animals with unique structural scaffolds and biochemical properties.<sup>28-30,53,54</sup> In *E. coli*, more than 100 unAAs and several  $\alpha$ -hydroxy acids have been genetically encoded by employing the PylRS/tRNA<sup>Pyl</sup> system.<sup>52</sup>

According to the crystal structure of *Desulfitobacterium hafniense* PylRS (PDB ID: 2ZNI), the anticodon loop of tRNA<sup>Pyl</sup> does not directly interact with PylRS and is not used as an identity element by PylRS.<sup>55</sup> Thus, the PylRS/tRNA<sup>Pyl</sup> system provides an additional advantage of a high tolerance towards changes in the tRNA<sup>Pyl</sup> anticodon loop, allowing for the decoding of other non-sense codons, such as UAA (ochre), UGA (opal), AGGA, AGUA, UAGA and CUAG.<sup>30,54,56,57</sup> In addition, the PylRS/tRNA<sup>Pyl</sup> pair can be directly evolved in *E. coli* and transferred to eukaryotic cells for unnatural amino acid mutagenesis since it is both orthogonal in prokaryotic and eukaryotic cells.<sup>54</sup>

### 1.2.3 EcTyrRS/BstRNA<sup>Tyr</sup>

In mammalian cells, both the EcTyrRS/EctRNA<sup>Tyr</sup> and the EcTyrRS/BstRNA<sup>Tyr</sup> pairs have been used for amber suppression. However, BstRNA<sup>Tyr</sup>, harboring internal A- and B-box promoters, expressed more efficiently in mammalian cells than EctRNA<sup>Tyr</sup>.<sup>58</sup> Therefore, EcTyrRS/BstRNA<sup>Tyr</sup> has been developed as a major system for the incorporation of unAAs into proteins in mammalian cells. A panel of functional groups, such as aryl halides, ketone, benzophenone, azide, and



propargyl have been genetically encoded by EcTyrRS/BstRNA<sup>Tyr</sup> pairs to facilitate the investigation of protein structure and function in mammalian cells.<sup>40,59,60</sup>

### 1.3 Nonsense Codons

Another important component of genetic code expansion is a unique codon to be assigned to encode an unAA. The standard genetic code, is well-conserved in nearly all organisms and uses 61 codons to encode the 20 canonical amino acids. The remaining 3 codons (UAA, ochre; UGA, opal; and UAG, amber) are used as termination signals for the protein translation. Extensive effort has been made to tackle the challenges facing genetic code reprogramming by creating additional “nonsense” or “blank” codons, such as quadruplet codons and rare codons.

#### 1.3.1 Amber stop codon

In *E. coli*, in-frame amber UAG codon occurs about 300 times in the genome, representing the least used stop codon (7%) compared to UAA and UGA codons.<sup>61</sup> In mouse and human, the frequency of UAG codon usage elevates slightly but still remains the lowest among the three stop codons. Thus, the reassignment of UAG codon into a sense codon should pose minimum impact on the native proteome and is expected to be well-tolerated by the host cells.

We and others have successfully incorporated unnatural amino acids in response to amber nonsense codons. To achieve this, one or more UAG codons were introduced in the coding gene of a protein of interest and a UAA or UGA codon was used as a “real” stop codon. To recognize the in-frame UAG codon, the anticodon of an orthogonal tRNA was mutated to CUA. Further evolution of the amber suppressor tRNA and cognate aaRS dramatically increased their amber suppression efficiency.<sup>3,4</sup> Generally, the UAG codon should not be decoded by any cellular components in the absence of an aminoacylated (charged) amber suppressor tRNA. Instead, truncated proteins should be formed due to the presence of termination machinery in the host cells.

Only in the presence of a charged amber suppressor tRNA can a full-length protein be formed by read-through of the amber codon as a sense codon.

To date, the UAG codon has been the most frequently used nonsense codon for unAA incorporation, while unAA mutagenesis at UAA or UGA codon has been minimally used.<sup>56,62</sup> The UAG codon has also been used together with other nonsense codons<sup>30,31,62,63</sup> to incorporate multiple unAAs into the same protein.

### 1.3.2 Quadruplet codon

The ability to incorporate multiple unAAs at defined positions in a single protein will provide a valuable tool to investigate protein structure, function, and dynamics. The number of unAAs that can be incorporated simultaneously is determined by the number of unique “blank” codons presented in a single gene. Since only two stop codons can currently be used, more “blank” codons are required to direct the incorporation of multiple unAAs. One solution is to redefine the meaning of naturally existing sense codons to create additional triplet nonsense codons, which will be discussed in the next section. An alternative solution is to employ unnatural quadruplet codons, which in theory provides 256 “blank” codons for unAA incorporation.

The naturally occurring quadruplet codon decoding tRNA mutants normally contain an expanded anticodon loop.<sup>64</sup> Transplanting such features into triplet decoding tRNAs rendered the generation of novel tRNAs capable of decoding quadruplet codons with unAAs. The Schultz group evolved a *P. horikoshii* lysyl tRNA synthetase/tRNA<sup>Lys</sup> pair for a less efficient insertion of L-homoglutamine in response to an AGGA codon.<sup>33</sup> An evolved MjTyrRS/tRNA<sup>Tyr</sup> pair was then used to incorporate *p*-acetyl-L-phenylalanine (pAcF) by suppressing UAGN codons in a genomically recoded bacteria strain.<sup>65</sup> In addition to the above efforts, we showed for the first time that an evolved tRNA<sup>Pyl</sup> variant could incorporate N<sup>ε</sup>-(*tert*-butyloxycarbonyl)-L-lysine (BocK) into

proteins in response to an AGGA codon in mammalian cells.<sup>54</sup> We also conducted a systematic evolution and investigation of UAGN-decoding tRNA<sup>Py1</sup> variants in a genomically recoded bacteria strain.<sup>66</sup>

Multi-site unAA incorporation has been achieved with the use of both amber codon and quadruplet codons. The Chin group developed an AGGA suppressor tRNA derived from *M. jannaschii*, which enables the incorporation of multiple *p*-azido-L-phenylalanine (pAzF) at specific sites in GST-MBP fusion protein.<sup>67</sup> This newly evolved MjTyrRS/tRNA<sup>Tyr</sup> pair was shown to be mutually orthogonal to the MbPylRS/tRNA<sup>Py1</sup> pair in *E. coli* and enabled the expression of full length GST-CaM fusion protein containing both pAzF and N6-[(2-propynyloxy)carbonyl]-L-lysine (CAK) in response to AGGA and UAG codons, respectively.<sup>67</sup> The incorporation of the above two unAAs enabled protein cyclization via a Cu(I)-catalysed Huisgen's [2+3] cycloaddition reaction. In a separate work, the same group evolved AGGA, AGUA, UAGA and CUAG suppressor tRNA<sup>Py1</sup> variants.<sup>30</sup> An azide and a norbornene were introduced into GST-CaM fusion protein by suppressing UAG and AGUA codons, which allowed for the protein to be selectively labeled with a FRET pair containing a strained alkyne and a tetrazine, respectively.<sup>30</sup>

### 1.3.3 Rare codons

An “amber-free” system utilizing sense codons (preferably rare codons) highlights the potential of an alternative approach for unAA incorporation. In nature, multiple redundant codons may encode a single amino acid. Theoretically, one codon is sufficient for coding each of the canonical 20 amino acids. Therefore, those rarely used sense codons offer a great potential to be redefined to new meanings without competing with release factors.<sup>68</sup>

However, significant challenges needed to be overcome in order to achieve efficient sense codon reassignment. A major concern is the impact on the host proteome. Pioneering work by the

Church group showed that synonymous replacement of rare codons in the essential genes significantly impaired the fitness of the host cells for unknown reasons.<sup>69</sup> In the same way, rare codon reassignment with unAAs may be unfavorable. Another concern is the competition with endogenous tRNAs at the codon of interest. This can be circumvented by engineering endogenous tRNAs to separate their assigned decoding pathways.<sup>70</sup> Despite all of these challenges, recent progress has stimulated more attentions in this area of research. A rare sense codon, AGG, has been successfully redefined to encode different unAAs both *in vitro* and *in vivo*.<sup>63,70-72</sup> In another example, the rare codon AUA has been made available for reassignment via deletion of the native tRNA and replacement with a tRNA from another species,<sup>73</sup> but incorporation of unAAs using this system is still in development.

#### 1.4 Unnatural Amino Acids

Unnatural amino acids play a crucial role in reprogramming the genetic code. The genetic incorporation and application of novel unAAs in living cells will not only expand the repertoire of building blocks for protein synthesis, but also deepen our understanding of molecular basis of life.

Similar to a canonical amino acid, an unAA contains an amine group, a carboxyl group, and a side chain that is specific to each amino acid. Theoretically, the side chain can be substituted with any desirable functional group, which provides potentially unlimited structural diversity and functional versatility. However, in a practical sense, only unAAs that possess proper physical, chemical, and biological properties can be genetically encoded. Appropriate solubility, stability, sizes, structures, and physicochemical properties are normally expected. In an attempt to mimic the natural substrates of MjTyrRS/tRNA<sup>Tyr</sup> and PylRS/tRNA<sup>Pyl</sup> pairs, many unAAs maintain the core structure of phenylalanine or lysine. In general, these unAAs are larger than canonical amino acids due to complex side chains.

To date, more than 200 unAAs have been selectively introduced into proteins and this number is likely to grow rapidly.<sup>6</sup> Site-specific incorporation of unAAs has enabled a broad range of applications in biochemical and biomedical research (**Figure 3**). The introduction of chemically reactive handles allows for site-selective labeling and modification of proteins, via bio-orthogonal reactions. A large number of bio-orthogonal reactions that can proceed selectively and efficiently under physiological conditions have been developed, such as Cu(I)-catalyzed or copper-free click reactions, the Staudinger ligation, and the inverse electron demand Diels-Alder reaction.<sup>5</sup> Genetic encoding of these reactive handles into proteins allows for corresponding reactions to occur in living cells. Besides selective labeling, these unAAs are also particularly useful in generating bio-conjugates and bio-materials with remarkable homogeneity and novel functions.<sup>74,75</sup>

Genetic incorporation of unAAs containing spectroscopic probes provides a powerful tool for the analysis of protein structure and function. Spectroscopic probes may also provide useful information on protein dynamics, localization, protein-protein interactions, and other cellular processes. For example, isotope-labeled unAAs, such as 2-amino-3-(4-(trifluoromethoxy)phenyl)propanoic acid (OCF<sub>3</sub>Phe), <sup>13</sup>C/<sup>15</sup>N-labeled *p*-methoxyphenylalanine (OMePhe), and <sup>15</sup>N-labeled *o*-nitrobenzyl-tyrosine (ONBY) have been used to probe protein structure, dynamics and ligand binding using NMR.<sup>76</sup> unAAs with specific infrared signatures, such as *p*-cyano-L-phenylalanine (pCNPhe)<sup>77</sup> and *p*-azido-L-phenylalanine (pAzF)<sup>78</sup> have been incorporated into proteins to probe protein dynamics and interactions. Site-directed spin-labeling has also been achieved with genetic incorporation of pAcF followed by chemical modification.<sup>79</sup> Direct incorporation of fluorescent unAAs provides a convenient tool to probe protein expression, localization and conformations in living cells.<sup>7,80</sup>

It has been known that PTMs play important roles in biological processes and cellular events. Some well-known PTMs include phosphorylation, sulfation, nitration, glycosylation, methylation, acetylation, and ubiquitination. However, it remains challenging to the biological role of PTMs due to the difficulty in the preparation and detection of PTM-proteins. The site-specific incorporation of unAAs containing proper PTMs or their mimics offers unprecedented control on the stoichiometry and homogeneity of the PTMs, allowing for the generation of homogeneously modified proteins. Thus, these unAAs may facilitate the elucidation of the biological roles and functions of the PTMs.

The site-specific incorporation of photo-caged unAAs provides precise temporal and spatial control on the structure and function of user-defined proteins. When introduced into proteins, the photo-caging group may temporarily block the molecular function of target proteins by either masking activity of essential amino acids or providing steric hindrance at or near the active site. After UV irradiation, the photo-caging group can be removed and the protein function can be restored. For example, photo-caged lysine, tyrosine, cysteine, selenocysteine, and serine have been genetically incorporated into proteins in various living cells, which enables a specific and rapid activation of protein function.<sup>49,81-85</sup>

Photo-crosslinkers represent another group of photo-activatable reactive unAAs. The introduction of photo-crosslinkers enables the formation of covalent bonds between the unAA-bearing proteins and its interacting partners under UV irradiation. Using this method, interactions that are weak, transient, or even unknown can be captured. For example, *p*-benzoyl-L-phenylalanine (Bpa) has been proven particularly useful in mapping interactions between a target protein and its ligand or other macromolecules.<sup>11,86</sup>

## 1.5 Approaches to improve the efficiency of unAA incorporation

To achieve optimal efficiency for unAA incorporation, extensive efforts have been devoted to evolving individual components of the translation machinery. Since genetic code expansion relies on an orthogonal aaRS/tRNA pair to direct the incorporation of unAA, systematic evolution approaches have been conducted to improve the inherent orthogonality of aaRS/tRNA pairs<sup>4,87,88</sup> and the catalytic activity of aaRS variants.<sup>27</sup> The competition between amber suppressor tRNA and release factor 1 (RF1) was shown to be responsible for low unAA incorporation in response to amber nonsense codon. This has been successfully addressed with RF1 knock-out strains.<sup>31,89</sup> As genetic code expansion utilizes the host translational machinery, proper interactions between the aaRS/tRNA pair and cellular components is highly appreciated. Previous work highlighted the potential of improving unAA incorporation through maintaining optimal interactions between tRNA charged with unAA and EF-Tu.<sup>90,91</sup> Also, the introduction of an orthogonal translation system which is only responsible for nonsense codon suppression provides advantageous options to enhance the incorporation efficiency.<sup>67,92,93</sup> Adjusting other experimental details, such as the sequence context flanking the nonsense codons,<sup>94</sup> the organization of expression vectors,<sup>95</sup> and the gene delivery method<sup>96</sup>, may also contribute to an improved unAA incorporation system.

### 1.5.1 Evolution of orthogonal aaRS/tRNA pairs

A general *in vivo* selection method (**Figure 4**) to identify MjTyrRS/tRNA<sup>Tyr</sup> mutants with better incorporation efficiency has been established based on amber suppression of a toxic gene (negative selection) and an antibiotic-resistant reporter gene (positive selection).<sup>3</sup> The negative selection was based on the suppression of amber mutations at permissive sites (Gln2 and Asp44) of barnase in the absence of unAA. The synthetases that could use any endogenous amino acids would produce functional barnase that would result in cell death. The positive selection was based

on the suppression of an amber mutation at a permissive site (Asp112) of chloramphenicol acetyl transferase-encoding gene (CAT or Cm<sup>R</sup>) in the presence of unAA. Cells can only survive if the synthetase is able to incorporate unAA or natural amino acids. In order to obtain MjtRNA<sup>Tyr</sup> mutants with improved orthogonality, a MjtRNA<sup>Tyr</sup> library was generated and subjected to both positive and negative selections to identify functional and orthogonal MjtRNA mutants.<sup>3</sup> The best candidate MjtRNA<sup>Tyr</sup> displayed a remarkable enhancement in the orthogonality and has been widely used.

In order to alter the specificity of MjTyrRS towards different substrates, key residues in the amino acid binding pocket of MjTyrRS were randomized.<sup>2,4</sup> The subsequent MjTyrRS library was subjected to both positive and negative selections to identify variants with the desired change in amino acid substrate specificity.<sup>4</sup> By varying the stringency, order, and cycles of selections, the activity and fidelity of selected MjTyrRS can be adjusted to achieve an optimal incorporation efficiency. This selection scheme was proved generally applicable for the evolution of other aaRS/tRNA pairs, such as PylRS/tRNA<sup>Pyl</sup> and EcTyrRS/tRNA<sup>Tyr</sup>.

Positive-negative selections can be performed with different signal outputs by simply replacing the selection motifs. For example, a dual reporter combining T7-GFP and CAT gene was developed.<sup>4</sup> Amber stop codons were placed in both the T7 RNA polymerase gene and the CAT gene. Amber suppression efficiency can then be determined by changes in the fluorescence intensity or by the concentration of chloramphenicol in the presence or absence of unAA.

In another example, Taq polymerase and *E. coli* phenylalanyl-synthetase (EcPheRS) were used as a positive- and negative-selection reporter, respectively.<sup>88</sup> The positive selection was performed in emulsified droplets. Functional tRNAs, which suppressed the amber stop codon, led to the generation of full-length Taq polymerases. The resulting tRNA mutants were amplified by



PCR with the Taq polymerase. In the negative selection, amber suppression with non-orthogonal tRNAs resulted in the generation of EcPheRS, which can utilize *p*-Cl-phenylalanine to cause cell death. By varying the concentration of *p*-Cl-phenylalanine and duration of negative selection, the selection stringency can be tuned.

Since the conventional selection strategy requires multiple steps of plasmid extraction and re-transformation, new strategies with fewer steps and faster speed are highly desirable. By integrating the target aaRS and the negative selection reporter *tolC* in the *E. coli* chromosome, and the positive selection reporter GFP on a separate plasmid, the whole selection process for identifying evolved aaRS candidates has been shortened to 1-2 days.<sup>97</sup>

### **1.5.2 Release factor 1 knockout and an amber-free host to enhance amber suppression**

One major drawback of the amber suppression system is the low incorporation efficiency, due to the ambiguous functions of UAG codon as both a termination signal and an unAA-encoding codon. Although the amber suppressor tRNAs are generally efficient, they cannot completely outcompete the translation termination machinery. The resulting truncated protein products may interfere with the function of the target protein and can be deleterious to the host cells. In addition, extremely low efficiency of unAA insertions at multi-sites is usually observed due to competition with the termination signals. Furthermore, the suppression of genomic UAG codons of essential genes by exogenous amber suppressor tRNAs may have unpredictable consequences and cause destructive effects to the host cells.

In prokaryotes, RF1 (release factor 1, recognizing UAA and UAG) and RF2 (release factor 2, recognizing UAA and UGA) signal the termination of translation by direct binding to the ribosome A site, which mediates the dissociation of the nascent peptide from the P site. To address the low amber suppression efficiency, initial efforts were focused on minimizing the RF1 mediated

translation termination. By decreasing the functional interaction between an orthogonal ribosome and RF1, the efficiency of amber suppression has been enhanced by more than 3 folds with a single amber codon and 20 folds with two amber codons.<sup>98</sup> An RF1 knock-out in *E. coli* resulted in a dramatically enhanced amber suppression efficiency.<sup>89,99,100</sup> Recent advances in genome editing allows for complete reassignment of UAG to a sense codon for protein translation. To achieve this, the Church group replaced all the instances of UAG with UAA in *E. coli* MG1655 followed by the removal of the RF1 encoding gene to generate C321.ΔA. This genomic recoded strain resulted in a tremendous enhancement in amber suppression efficiency.<sup>65,97</sup> One noteworthy example using this strain resulted in the incorporation of up to 30 unAAs into a single protein with excellent yields (~50 mg/L) and fidelity (>95%).<sup>97</sup> In addition, the C321.ΔA strain also works as an ideal expression host for quadruplet UAGN codon decoding. The Schultz group reported that the C321.ΔA strain enhanced the suppression efficiency of UAGN codon with a MjTyrRS/tRNA<sup>Tyr</sup> pair.<sup>65</sup> Using the same strain, we performed a systematic evolution of UAGN suppressors and successfully identified several tRNA<sup>Pyl</sup> mutants with robust quadruplet codon decoding efficiency for different unAAs.<sup>66</sup>

### 1.5.3 Refining the interactions between the orthogonal tRNA and cellular components

In prokaryotes, EF-Tu (elongation factor thermo unstable) recognizes and transfers the charged tRNA to the ribosome A site. Since the sequence of archaeal tRNA differs significantly from that of *E. coli* tRNA, a compromised compatibility of archaeal tRNA with the *E. coli* translational machinery may occur. Non-optimal interactions between EF-Tu and tRNA charged with unAA may result in low efficiency of protein synthesis.

Two approaches have been taken to obtain optimal interactions between archaeal tRNA and EF-Tu in host cells. One approach utilizes evolved tRNAs to resolve such unfavorable interaction.

From tRNA libraries containing mutations at the binding interface with EF-Tu, Guo *et al* identified multiple MjtRNA mutants with significant enhanced amber suppression efficiency with a panel of distinct unAAs.<sup>90</sup> Targeting the same sets of nucleotides, Maranhan *et al* successfully evolved tRNAs that specifically incorporate *meta*-halo-tyrosines.<sup>88</sup> An alternative approach is to engineer a specific EF-Tu to facilitate the transfer of tRNA charged with unAAs. For example, Söll and co-workers successfully incorporated *o*-phosphoserine (Sep) in response to amber stop codon using a Sep-specific aaRS/tRNA pair and an engineered EF-Tu (Sep-Tu).<sup>91</sup>

#### 1.5.4 Orthogonal ribosomes for enhancing nonsense codon suppression efficiency

The natural ribosomes are less efficient at amber suppression or quadruplet codon decoding due to their natural function in synthesizing a triplet codon-based proteome. Conversely, the orthogonal ribosomes works in parallel, but independent of natural ribosomes; and are only responsible for translating the message from an orthogonal mRNA containing an orthogonal ribosome-binding site (RBS).<sup>92</sup> In principle, more efficient site-direct incorporation of unAAs at amber stop codons and quadruplet codons can be achieved with evolved orthogonal ribosomes without affecting the synthesis of the native proteome.

To generate an orthogonal ribosome, the Chin group created a 16S rRNA library containing 7 randomized nucleotides in the 530 loop which is proximal to both tRNA and RF1 binding site.<sup>101</sup> Mutations that favored interactions with an amber suppressor tRNA over the translation termination were identified through a function screen. The resultant ribo-X showed significantly enhanced tRNA-dependent amber suppression with orthogonal mRNAs containing one or more amber stop codons. Using ribo-X as a starting point, 11 saturation mutagenesis libraries covering 127 nucleotides in the tRNA binding site of ribosome were constructed in order to discover orthogonal ribosomes that acquire the ability to efficiently decode quadruplet codons while

maintaining amber suppression ability.<sup>67</sup> Surprisingly, one resultant ribosome mutant, ribo-Q1, contained only two mutations (A1196G and A1197G), but the mutations drastically improved the decoding efficiency for a variety of quadruplet codons with tRNAs derived from *M. jannaschii*<sup>67</sup> or *Methanosarcina* species<sup>30</sup>. Using this O-ribosome, an insertion of multiple unAAs was achieved with two mutually orthogonal aaRS/tRNA pairs.

## 1.6 Application of genetic code expansion in vaccine development

Therapeutic and preventive vaccines, are important bio-products that help develop immunity against particular diseases. The scientific community has spent decades trying to develop therapeutic vaccines to cure various cancers. However, little progress has been made until now in the development of therapeutic vaccine partially due to the complexity of human immune system. Conversely, preventive vaccines have achieved great success in preventing a variety of diseases, such as polio, measles, and tetanus. However, key hurdles still exist that hinder developing efficient and safe vaccines for other diseases, such as AIDS.<sup>102</sup>

Since previous efforts have not provided solutions for some irremediable diseases, novel vaccine strategies are urgently needed. Genetic code expansion has shown great potential in vaccine development. First, it offers a simple and effective method to increase immunogenicity of therapeutic vaccines. Second, it provides a useful tool for generating virus like-particles or conditionally-replicated live viruses that can be used as preventive vaccines.

### 1.6.1 Design therapeutic vaccines using genetic code expansion

The development of therapeutic vaccines against cancer or chronic degenerative diseases is a well-known challenge due to our natural tolerance against self-proteins. An autologous protein modified with an immunogenic unAA offers a simple and efficient strategy to circumvent self-

tolerance. In one example, a murine tumor necrosis factor- $\alpha$  (mTNF- $\alpha$ ) with a Tyr86pNO<sub>2</sub>F (*p*-nitro-L-phenylalanine) mutation induced a high titer of antibody responses against both wild-type and pNO<sub>2</sub>F containing mTNF- $\alpha$ . Critically, it protected against lipopolysaccharide (LPS)-induced death in mice.<sup>21</sup> High titers of antibody responses were also observed with mTNF- $\alpha$  containing pNO<sub>2</sub>F mutations at different locations.<sup>22</sup> Similar results were obtained with mTNF- $\alpha$  and EGF proteins containing different pNO<sub>2</sub>F, sTyr or 3-nitrotyrosine (3NO<sub>2</sub>Tyr) mutations, suggesting an important role for naturally occurring PTMs in antibody production.<sup>23</sup>

### 1.6.2 Design preventive vaccines using genetic code expansion

The applications of genetic code expansion in viral genome engineering have gained considerable interests. Site-specific incorporation of unAAs have been achieved with several human viral systems, including lentiviral vectors,<sup>103</sup> Hepatitis-D,<sup>104</sup> adeno-associated virus 2,<sup>105</sup> HIV-1,<sup>25,26,106</sup> and Influenza-A.<sup>24</sup> The further investigation of genetically modified viruses will provide valuable information regarding the host-pathogen interactions and related immune responses. Furthermore, the introduction of control elements, including nonsense codons and orthogonal aaRS/tRNA pairs, can be used to switch on or off the replication of live viruses. The resulting conditionally replicated viruses can be used as preventive vaccines to elicit robust immune responses without concerns of acquiring disease.

We previously reported a general approach to generate live-attenuated HIV-1 virus, which can be potentially used as a preventive HIV vaccine.<sup>25</sup> In this approach, one or multiple amber stop codons were installed in the essential genes in the HIV-1 genome and an evolved EcTyrRS/BstRNA<sup>Tyr</sup> pair was supplied exogenously. In the absence of an unAA, the amber stop codon cannot be decoded and the viral replication is switched off. In the presence of the unAA,

full-length viral protein can be expressed and the replication of HIV-1 can be switched on. To successfully apply this strategy *in vivo*, an important prerequisite is to elicit a robust immune response through sustained viral replication. We sought to solve this by integrating the control elements in the HIV-1 genome<sup>26</sup> to enable the co-expression of the EcTyrRS/BstRNA<sup>Tyr</sup> pair and the HIV-1 viral proteins. The resultant virus strain obtains the ability of multicycle infection in T cells, which represents the first step towards a safe and effective HIV-1 live-attenuated vaccine.

Using a similar approach, a live but replication-incompetent influenza A vaccine was generated and evaluated in animal models.<sup>24</sup> In this approach, multiple amber stop codons were inserted in the viral genome to avoid reversion to wild-type during viral propagation. In conventional 293T cells or transgenic 293T cells stably expressing an evolved PylRS/tRNA<sup>Pyl</sup> pair, the influenza virus was packaged in the presence of N<sup>ε</sup>-2-azidoethoxycarbonyl-L-lysine. The virus was only observed to reproduce in the transgenic cells, suggesting a replication-incompetent feature. The *in vivo* tests demonstrated that the virus vaccine was safe in mice, ferrets, and guinea pigs. The virus vaccine elicited robust humoral, mucosal, and cell-mediated immunity in the tested animals, which was comparable to or stronger than the commercial inactivated or live-attenuated influenza vaccines. Potent protection was observed in animals challenged with wild-type or antigenically distant viral strains. In addition, the virus vaccine was shown to inhibit the propagation of existing virus.

## 1.7 Summary and remarks

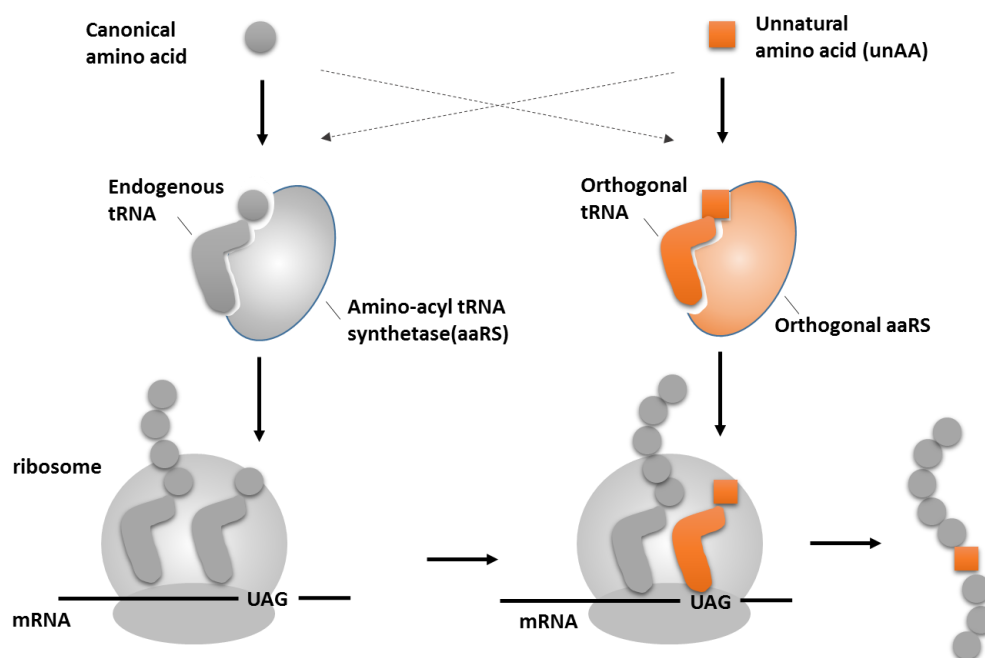
Genetic code expansion provides a versatile approach to generate unAA-modified proteins. Over the last decade, the repertoire of unAAs has rapidly expanded. A large variety of functional groups with distinctive structures have been selectively introduced into numerous target proteins in live cells. The site-specific incorporation of these unAAs relies on the suppression of nonsense

codons with orthogonal aaRS/tRNA pairs. High fidelity and good suppression efficiency have been achieved with aaRS/tRNA pairs derived from different species, allowing for unAA mutagenesis in all common protein expression hosts, including bacteria, yeasts, insect cells, and mammalian cells. Different approaches have been taken to improve the efficiency of unAA incorporation.

Although the evolved aaRS/tRNA pairs have shown good orthogonality and suppression efficiency, a deeper understanding of the interaction between aaRS and tRNA is still highly desirable. To achieve this, we investigated two of the most commonly used orthogonal pairs: PylRS/tRNA<sup>Pyl</sup> and MjTyrRS/tRNA<sup>Tyr</sup>. We improved quadruplet codon decoding efficiency through systematic evolution of PylRS/tRNA<sup>Pyl</sup> pairs (Chapter 3). We enhanced amber suppression efficiency of evolved MjTyrRS/tRNA<sup>Tyr</sup> pairs by engineering the anticodon binding pocket of MjTyrRS (Chapter 4). All of these efforts lead to a further improvement in orthogonal aaRS/tRNA pairs, which will likely expand the applications of unAA mutagenesis in biochemical and biomedical research.

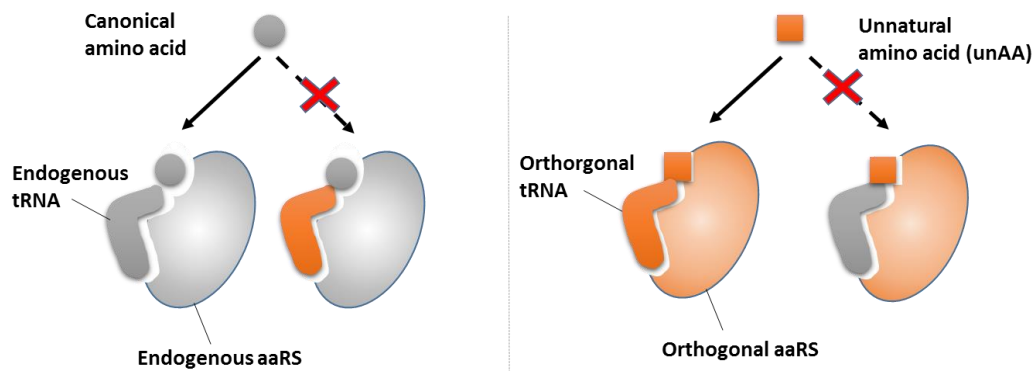
We and others have employed unAA mutagenesis for vaccine development. As an example, developing a safe and effective HIV-1 vaccine may provide an ultimate solution for the worldwide pandemic. Since HIV-1 hijack host translational machinery for protein synthesis, a reprogrammed protein translation system may be used to control viral replication. We reported the application of genetic code expansion in HIV-1 virus, resulting in either single-cycle or multicycle live-attenuated HIV-1 viruses (Chapter 2). These genetically modified viruses may be used as preventive vaccines to protect against HIV-1 infection. Our strategy may be applied to the generation of vaccines against other pathogens.

## 1.8 Figures and tables

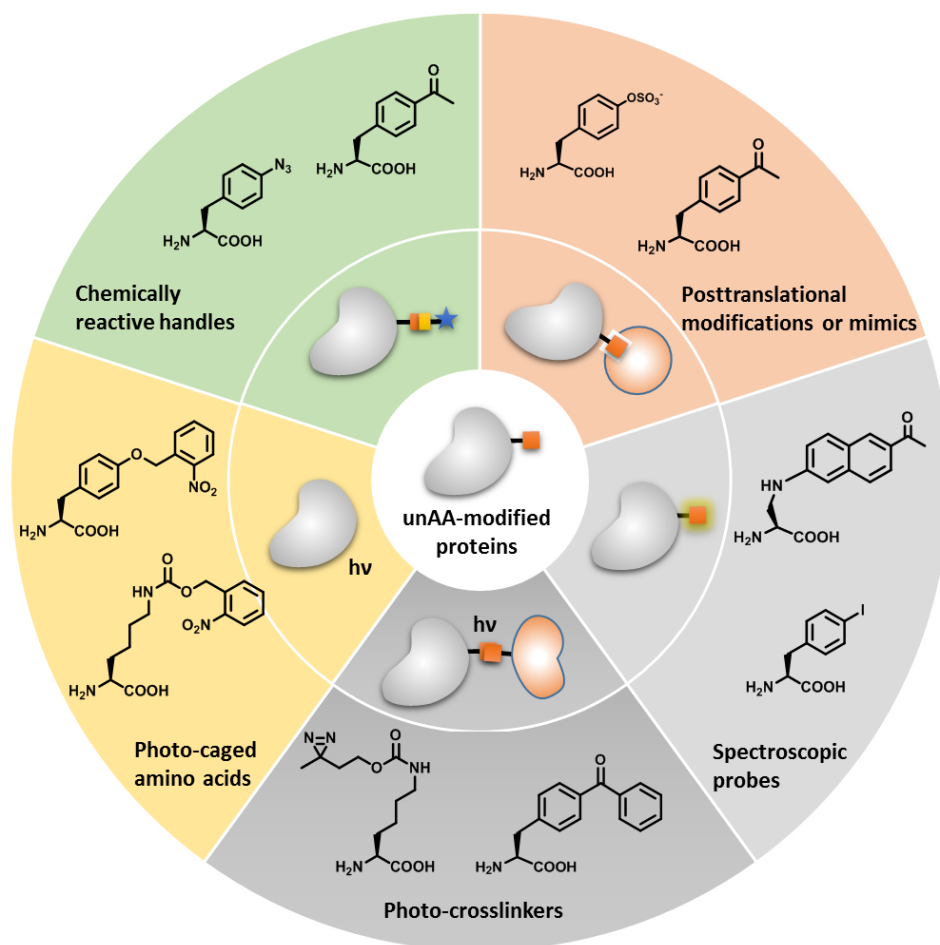


**Figure 1.** Reprogramming the genetic code. The unnatural amino acid (unAA) is specifically recognized by an orthogonal aminoacyl-tRNA synthetase (aaRS) and charged onto its cognate amber suppressor tRNA. The unAA is subsequently added into the growing polypeptide chain during ribosomal protein synthesis in response to the UAG amber stop codon.

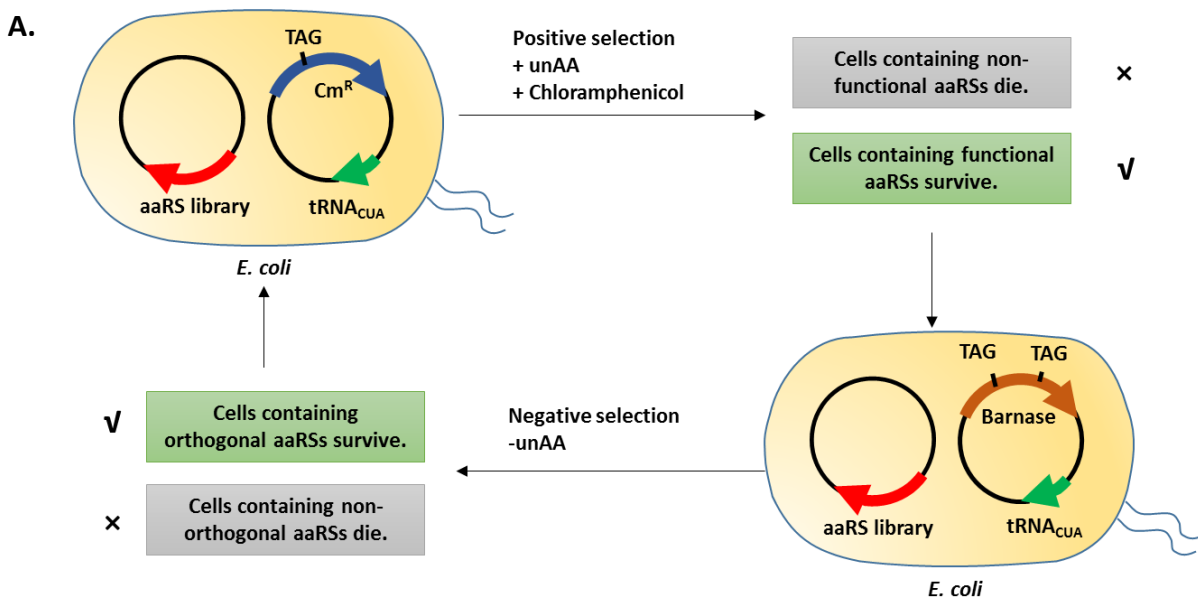




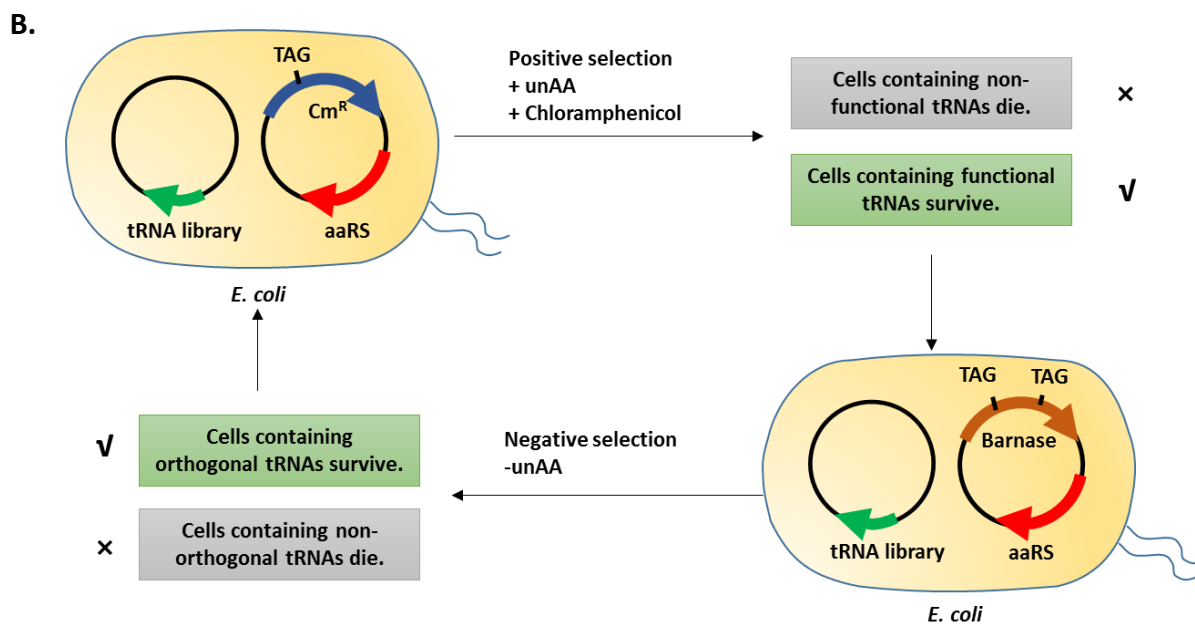
**Figure 2.** An orthogonal aaRS/tRNA pair for genetic code expansion. The orthogonal aaRS/tRNA pair should not “crosstalk” with the endogenous aaRS/tRNA pairs. The unAA should only be recognized by its orthogonal aaRS and charged onto the cognate tRNA, but not the endogenous tRNA. The orthogonal tRNA should not be recognized by endogenous tRNA to avoid misincorporation of canonical amino acids.



**Figure 3.** Examples of unnatural amino acids. Based on functions, unAA can be classified into five major categories, including chemically reactive handles, posttranslational modifications or mimics, spectroscopic probes, photo-crosslinkers, and photo-caged amino acids.



### *In vivo* selections to identify evolved aaRS variants



### *In vivo* selections to identify evolved tRNA variants

**Figure 4.** A general *in vivo* selection method for identifying evolved aaRS and tRNA<sub>CUA</sub> variants.

(A) *In vivo* screening of an aaRS library to identify aaRS variants with altered amino acid specificity. (B) *In vivo* screening of a tRNA library to identify tRNA<sub>CUA</sub> variants with improved orthogonality.

## CHAPTER 2

**Development of Live-attenuated HIV-1 Viruses via Genetic Code Engineering**

Reproduced with permission from “Angewandte Chemie International Edition. 2014, 53, 4867.”, “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”, and “Journal of chemical technology & biotechnology. 2017, 92, 455.”

**2.1 Introduction: A novel approach towards a HIV-1 live-attenuated vaccine**

Since the first discovery of human immunodeficiency virus type 1 (HIV-1) in the last century, over 70 million people worldwide have been infected with HIV-1 and about half of them have died of this virus.<sup>107,108</sup> In order to control the global HIV-1 pandemic, the development of a safe and effective vaccine is of critical importance. Unfortunately, after more than three decades of effort, an HIV-1 vaccine remains elusive. The development of an HIV-1 vaccine faces unprecedented scientific obstacles, such as the early establishment of latent viral reservoir,<sup>109</sup> the enormous viral genetic diversities and mutation rates,<sup>110</sup> the ability of the virus to escape from immune surveillance,<sup>111</sup> and incapability of eliciting broadly neutralizing antibodies.<sup>112</sup> To tackle these challenges, a deeper understanding of HIV-1 pathogenesis and immunology and novel approaches beyond the scope of traditional vaccinology are urgently needed.

Among all the vaccine modalities developed and evaluated so far, a HIV-1 live attenuated vaccine (LAV) that was constructed by the deletion of *nef* gene (referred to as  $\Delta$ *nef*-LAV hereafter) has conferred the best protective efficacy (95%) in a simian immunodeficiency virus (SIV)/rhesus macaque model.<sup>113-117</sup> However, the SIV  $\Delta$ *nef*-LAV retained pathogenic potentials and could lead to AIDS and even death in vaccinated neonatal and adult macaques,<sup>118-120</sup> which was likely caused by uncontrolled viral replication. In addition, the  $\Delta$ *nef*-LAV strains showed potentials to regain

virulence through long-term virus evolution.<sup>121-123</sup> Therefore, HIV-1 LAV strategy has not been actively pursued in clinical trials.

To solve the safety problems of conventional HIV-1 LAVs, we developed a novel approach to regulate the replication of HIV-1 with an unnatural genetic switch<sup>25,26,106</sup> based on an amber suppression system. This genetic switch consists of (1) a unique amber suppressor tRNA, (2) an aaRS that specifically charges an unAA onto the amber suppressor tRNA, and (3) an unAA (**Figure 5**). The regulation of HIV-1 replication is achieved through the control of the biosynthesis of essential viral proteins. To this end, one or more amber mutations were inserted into essential genes of HIV-1. In the presence of the aaRS/tRNA pair and the unAA, full-length HIV-1 proteins can be synthesized and the replication of HIV-1 is thus turned on (on-state). In the absence of an unAA, only truncated viral proteins can be produced and the replication of HIV-1 is turned off (off-state). We have successfully demonstrated that the replication of HIV-1 can be turned on and off using this genetic switch *in vitro* by exogenously expressing the aaRS/tRNA pair. We also engineered all-in-one live-attenuated HIV-1 variants that contain a genomic copy of the amber suppression machinery. These strains gained the capability of multicycle replication and infection in mammalian cells. The use of an unAA-mediated genetic switch in vaccine design represents a completely novel approach towards the development of a safe and effective HIV-1 LAV. Potentially, this approach can be applied in the research and discovery of other live-attenuated or live but replication-incompetent vaccines.

## 2.2 Approaches

### 2.2.1 Single-cycle amber suppression-dependent HIV-1 variants

We sought to control the replication of HIV-1 by our proposed unnatural genetic switch. In this work, one major challenge was to identify a suitable genetic switch with good suppression

efficiency and excellent fidelity on HIV-1 virus. We mainly focused on the EcTyrRS/BstRNA<sup>Tyr</sup> pair and its derivatives, including *p*-azido-L-phenylalanyl-tRNA synthetase (pAzFRS)/tRNA<sup>Tyr</sup>, *p*-iodo-L-phenylalanyl-tRNA synthetase (pIodoFRS)/tRNA<sup>Tyr</sup>, and *p*-acetyl-L-phenylalanyl-tRNA synthetase (pAcFRS)/tRNA<sup>Tyr</sup> pairs.<sup>40</sup> We also evaluated PylRS/tRNA<sup>Pyl</sup> pairs, and specifically on *o*-nitrobenzyl-tyrosyl-tRNA synthetase (ONBY-RS)/tRNA<sup>Pyl</sup> and *o*-nitrobenzyl-oxycarbonyl-N<sup>ε</sup>-L-lysyl tRNA synthetase (ONBK-RS)/tRNA<sup>Pyl</sup> pairs.<sup>49,124,125</sup> First, we examined whether the pAzFRS/tRNA<sup>Tyr</sup> pair can be applied as a genetic switch in mammalian cells via a fluorescence based functional assay. Next, we investigated the incorporation of natural or unnatural amino acids into viral proteins with the aaRS/tRNA pairs mentioned above. At the meantime, we evaluated the effects of amber suppression on HIV-1 viability using p24-based infection assay by using TZM-bl as host cell. Last, we successfully demonstrated that HIV-1 replication can be precisely controlled by proposed unnatural genetic switch strategy when pAzFRS/tRNA<sup>Tyr</sup> pair was employed.

### 2.2.1.1 Detection of suppression efficiency and fidelity of the AzFRS/tRNA<sup>Tyr</sup> pair

We first examined the suppression efficiency and fidelity of the pAzFRS/tRNA<sup>Tyr</sup> pair in mammalian cells. We co-transfected 293T cells with a plasmid pAzFRS that encodes the pAzFRS/tRNA<sup>Tyr</sup> pair and a plasmid pEGFP that encodes an enhanced green fluorescent protein (EGFP) with an in-frame amber mutation at position Tyr40. After transfection, 293T cells were incubated for an additional 12 h in DMEM media with or without 1 mM pAzF. As shown in **Figure 6**, fluorescence signals were only detected in the presence of pAzF. While in the absence of pAzF, no detectable fluorescence signals were observed. In addition, tandem-mass spectrometry data (**Figure 6C**) confirmed that 4-aminophenylalanine (the reduction product of pAzF) was

incorporated at position 40 of EGFP exclusively. Taken together, the pAzFRS/tRNA<sup>Tyr</sup> pair showed an excellent suppression efficiency and fidelity in 293T cells.

### 2.2.1.2 Screening of permissive sites in HIV-1 genome for amber mutations

The HIV-1 genome is highly compact and harbors multiple overlapping regions.<sup>126</sup> As shown in **Figure 7**, the HIV-1 genome contains three major genes (*gag*, *pol*, and *env*) and six accessory genes (*tat*, *rev*, *vif*, *vpu*, *vpr*, and *nef*) flanked by 5'- and 3'-long terminal repeats (LTRs). The *gag* gene encodes viral structural proteins, including matrix protein p17 (MA), capsid protein p24 (CA), spacer peptide 1 (SP1), nucleocapsid protein (NC), spacer peptide 2 (SP2), and P6 protein. The *pol* gene encodes essential viral enzymes, including protease (PR), reverse transcriptase (RT), RNase H and integrase (IN). The *env* gene encodes glycoprotein gp120 or surface protein (SU) and gp41 or transmembrane protein (TM). The six accessory genes encode regulatory proteins (Tat and Rev) and auxiliary proteins (Vif, Vpu, Vpr and Nef). In addition, the primary HIV-1 transcript undergoes a sophisticated splicing process and can be processed to yield more than 40 different mRNA species.<sup>127</sup> Thus, it is important to identify a suitable site for amber mutation or other genetic modifications.

To screen for a proper mutation site, multiple HIV-1 variants were constructed (**Table 1**). A wild-type transmitter/founder HIV-1 virus pSUMA (catalog #11748, from Dr. John Kappes and Dr. Christina Ochsenbauer through NIH AIDS Reagent Program) was used as the parental template for subsequent cloning work. Amber mutations were inserted at *gag*, *pol*, and *tat* genes respectively by site-directed mutagenesis. The *env* gene was not included as a target gene due to a relatively higher mutation rate.

### 2.2.1.3 Genetic incorporation of tyrosine analogs into HIV-1 proteins

When applying the unnatural genetic switch to control HIV-1 replication, the choice of unAA is of great importance. Ideally, unAAs should not interfere with the structure and function of viral proteins. Initially, we chose three close analogs of tyrosine, pAzF, pIodoF, and pAcF (1, 2, and 3 in **Figure 5C**). In addition to having similar sizes to tyrosine, these unAAs contain chemically reactive handles that allow for selective modifications<sup>128-131</sup> of viral proteins, which provide opportunities for remodeling the properties of vaccine strains in the future.

We first evaluated the incorporation of pAzF into viral proteins. The pSUMA variants MA-A119, MA-Y132, RT-F216, RT-F271, RT-L365, or Tat-T20 were co-transfected with pAzFRS into 293T cells. After 48 h of cultivation either in the presence or absence of pAzF, the resulting viruses were harvested and subjected to a p24 assay to quantify the viral titer. The presence and intensity of a blue color indicate the existence and the level of p24 being produced, respectively. As shown in **Figure 8**, in the presence of 1 mM pAzF, p24 expression was detected with all pSUMA variants although with different levels. In the absence of pAzF, p24 synthesis of MA-Y132 and MA-A119 was undetectable. The harvested viruses were subsequently used to infect TZM-bl cells, an engineered cell line that express CD4 and CCR5 on its surface and being highly sensitive to infection with diverse isolates of HIV-1. TZM-bl cell also harbors a genomic copy of  $\beta$ -galactosidase gene under the control of HIV-1 promoter. Thus, HIV-1 infection can be detected with X-gal assay using this cell line. As shown in **Figure 11G** and **11H**, extremely weak infection was observed with MA-A119 and RT-L365 variants. While no live virus was detected with other HIV-1 variants (**Figure 11E** and **11F**).



Next, the incorporation of pIodoF and pAcF was examined by co-transfection 293T cells with a pSUMA variant (MA-A119, MA-Y132, CA-L268, RT-F216, RT-F271, RT-L365, or Tat-T20) and a plasmid pIodoFRS (encodes the pIodoFRS/tRNA<sup>Tyr</sup> pair) or a plasmid pAcFRS (encodes the pAcFRS/tRNA<sup>Tyr</sup> pair). As shown in **Figure 9**, p24 synthesis was almost undetectable with MA-Y132, MA-A119, or Tat-T20 variants in the presence of pIodoF, suggesting that the incorporation of pIodoF in MA or Tat might disrupt the expression of p24. In contrast, strong p24 signals were obtained with these mutants in the presence of pAcF. However, the incorporation of pIodoF or pAcF did not lead to any infectious virus.

#### 2.2.1.4 Genetic incorporation of photocaged unAAs into HIV-1 proteins

Genetically encoded photocaged unAAs can be used to regulate protein function and biological process with spatial and temporal precision in living cells.<sup>6</sup> We envisaged that substitution of a natural amino acid with a photocaged unAA at protease cleavage sites in Gag may transiently block the cleavage process. After UV irradiation, the caged group can be removed and the cleavage site can be restored for proteolytic cleavage, which will result in generation of live viruses. We chose ONBY and ONBK (5 and 6 in **Figure 5C**) for the test.

We first examined the incorporation of ONBY at MA/CA cleavage site using MA-Y132 variant. After UV irradiation at 365nm, ONBY should be converted into tyrosine at position 132 of matrix protein and result in a wild-type HIV-1 proteome. However, no infection activity was detected with MA-Y132ONBY before or after UV treatment. Next, we evaluated the incorporation of ONBK at CA/P2 cleavage site by using a HIV-1 variant CA-K359. After UV irradiation at 365nm, ONBK should be de-caged and produce a lysine residue at position 359 of capsid protein. However, no infectious viruses were formed with or without UV treatment.

As a control, we examined the incorporation of ONBY and ONBK into EGFP in 293T cells. Green fluorescence was only detected in the presence of 0.4 mM ONBY (**Figure 10A**), suggesting efficient incorporation of ONBY. Genetically encoded ONBK in EGFP also led to a strong green fluorescence signal (**Figure 10C**), suggesting sufficient amber suppression efficiency. While in the absence of ONBY (**Figure 10B**) and ONBK (**Figure 10D**), no significant fluorescence signal was detected. These results suggest that the incorporation of ONBY or ONBK is not the limiting factor in our experimental design. Combining the above results, we reasoned that the failure in virus production was not due to low suppression efficiency but rather the disruption of protein functions by the encoded unAA. Since we have not fully explored available photocaged unAAs or screened sufficient amber mutation sites, we remain optimistic that optical-regulation of HIV-1 replication can be achieved in the future.

### 2.2.1.5 Genetic incorporation of tyrosine into HIV-1 proteins

We sought to investigate whether the incorporation of unAAs caused the deficiency in virus production. As a comparison to unAA mutagenesis, we evaluated the effects of natural amino acid substitution on HIV-1 viability. We co-transfected MA-Y132 variant with a plasmid that encodes the TyrRS/tRNA<sup>Tyr</sup> pair into 293T cells. The incorporation of tyrosine at position 132 should result in wild-type Gag. Based on infection assays (**Figure 11A**), MA-Y132 showed near wild-type viability, suggesting that synonymous mutagenesis was tolerable at position 132 of Gag. In the control experiment, no infection was observed without the TyrRS/tRNA<sup>Tyr</sup> pair (**Figure 11B**). These results suggested that the incorporation of unAA at Y132 indeed interfered with the expression and/or function of Gag. Next, we tested the incorporation of tyrosine into MA-A119, CA-L268, RT-F216, RT-F271, RT-L365, and Tat-T20 variants respectively in the presence of the TyrRS/tRNA<sup>Tyr</sup> pair. Among all the HIV-1 variants being tested, only RT-F216 (**Figure 11C**) and

RT-L365 (**Figure 11D**) showed some infectivity, suggesting that natural amino acids substitution may also affect viral viability.

#### 2.2.1.6 Effects of suppression at genomic amber codons

As the formation of virions was hardly observed in above experiments, we decided to assess potential impact of suppression at genomic amber codons. Since wild-type pSUMA (referred to as pSUMA-wt) has five amber stop codons, undesirable suppression at any of these positions may affect protein syntheses and following life cycle of the virus, which are crucial for HIV-1 infection. We co-transfected pSUMA-wt with plasmid pAzFRS into 293T cells. The viruses were harvested and subjected to p24 assay and infection assay to detect viral titer and viability. No obvious differences were observed in the presence or absence of the pAzFRS/tRNA<sup>Tyr</sup> pair (**Figure 12**). Therefore, amber suppression did not have an apparent negative effect on HIV-1 life cycles.

Next, we constructed an HIV-1 variant based on pNL-GI, which harbors a genomic copy of hGFP and can directly indicate HIV-1 replication by fluorescence signals. An amber mutation was made at position 40 of hGFP to replace a tyrosine codon and the entire HIV-1 genome remained unchanged. The resulting HIV-1 variant pNL-GI-N40 was co-transfected with plasmid pAzFRS into 293T cells. We observed strong fluorescence with pNL-GI-N40 (**Figure 13B**) in the presence of AzF, which is comparable to the wild-type pNL-GI (**Figure 13A**), indicating a good suppression efficiency with the pAzFRS/tRNA<sup>Tyr</sup> pair. In addition, a near wild-type infection efficiency was observed with pNL-GI-N40TAG in the presence of pAzF (**Figure 13D**). Based on the above results, we concluded that amber suppression did not pose a significant pressure on viral proteome or viability.

### 2.2.1.7 To regulate the replication of HIV-1 using an unnatural genetic switch

Since amber suppression with MA-Y132 and pNL-GI-N40 resulted in near wild-type infectivity, we decided to look for other suitable sites for unAA incorporation. We constructed a new HIV-1 variant, PR-Y59, which contains an amber mutation at position 59 of protease. Since Tyr59 is away from the catalytic center (PDB ID: 1EBZ),<sup>132</sup> an unAA at this position might not cause a significant change in the structure and function of the protease. To test the efficiency of amber suppression, we first examined the incorporation of tyrosine at this position. As shown in **Figure 14A**, in the presence of the TyrRS/tRNA<sup>Tyr</sup> pair, the assembly of live viruses was observed with PR-Y59. To reduce the chance of HIV-1 to revert back to the virulent form, we installed two amber mutations at position 36 and 127 of matrix protein based on crystal structure (PDB ID:2H3F).<sup>133</sup> It was shown clearly that the presence of the TyrRS/tRNA<sup>Tyr</sup> pair and MA-W36Q127 variant led to infectious viruses (**Figure 14C**). In addition, we replaced Phe44 of matrix protein with tyrosine using the TyrRS/tRNA<sup>Tyr</sup> pair and obtained similar level of infection (**Figure 14B**).

Next, we examined the amber suppression of PR-Y59 and MA-W36Q127 with the pAzFRS/tRNA<sup>Tyr</sup> pair. Encouragingly, only in the presence of 1 mM pAzF, PR-Y59 and MA-W36Q127 showed infection activity (**Figure 15**). As a control, no live viruses were generated with PR-Y59 or MA-W36Q127 in the absence of pAzF. Since the infected cells do not contain the genetic switch (pAzFRS/tRNA<sup>Tyr</sup> pair and pAzF), the infection lasts for only one-cycle and no progeny viruses will be produced, which partially addresses the safety concerns on HIV-1 LAVs. Then we tested the stringency of the unnatural genetic switch by comparing the tissue culture infectious dose 50 (TCID<sub>50</sub>) values in the presence (on-state) or the absence (off-state) of pAzF (**Figure 15**). The TCID<sub>50</sub> of PR-Y59 was 0.00 at off-state in comparison to  $1.31 \times 10^3$  at on-state, suggesting a high fidelity of the pAzFRS/tRNA<sup>Tyr</sup> pair. A slightly higher TCID<sub>50</sub> ( $5.12 \times 10^3$ ) was

observed with MA-W36Q127 at on-state, while no infection activity (0.00) was observed at off-state, which also confirms the tight control on the viral replication by our unnatural genetic switch.

In summary, we have demonstrated that a stringent control of HIV-1 replication can be achieved by using an unnatural genetic switch *in vitro*. The ability to incorporate different functional groups with *E. coli* and archaeal aaRS/tRNA pairs allows for further modifications of viral proteins, which can be potentially used to improve the safety and efficacy of our vaccine strains.

### 2.2.2 Multi-cycle amber suppression-dependent HIV-1 variants

In previous sections, we have demonstrated a precise control of the replication of HIV-1 by an unnatural genetic switch. Since the host cells do not have the unique amber suppressor aaRS/tRNA pair, these variants cannot complete a second round of replication or infection. Therefore, these single-cycle HIV-1 LAV strains should be safe but may not be effective to induce desirable immune responses in the vaccination host due to a short window of infection. To overcome this potential drawback, HIV-1 variants that are able to undergo multiple cycles of replication and remain under a stringent control are needed.

To achieve this goal, we constructed all-in-one live-attenuated HIV-1 variants by sequentially inserting all control elements of the unnatural genetic switch into HIV-1 genome (pNL4-3 strain). First, we screened genomic sites that allow for the insertion of a large piece of foreign DNA. Next, we inserted a genomic copy of aaRS between *env* and *nef* of pNL4-3 genome and examined the infectivity of the resulting pNL43-aaRS mutants. Then, we inserted a DNA cassette containing tRNA<sup>Tyr</sup> under the control of human U6 promoter into pNL43-AzFRS at different locus and orientations to afford four pNL43-AzFRS-tRNA variants. After evaluation of both viral infectivity

and proper function of the inserted pAzFRS/tRNA<sup>Tyr</sup> pair, we chose the best candidate pNL43-AzFRS-tRNA-2 for subsequent experiments. pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 were constructed by inserting amber mutation into HIV-1 matrix protein and protease respectively. The infectivity and multicycle replication of the above two strains were examined. We demonstrated that the unnatural genetic switch could be used to turn on and off the replication of multicycle HIV-1 variants.

### 2.2.2.1 Screening of genomic insertion sites for aaRS/tRNA pairs

HIV-1 genome is highly compact and contains many overlapping regions. The insertion of a large piece of foreign DNA may interfere with HIV-1 genome, proteome and even normal life cycle. Therefore, it is critical to identify a proper insertion site for the aaRS/tRNA pair. Previously, GFP-encoding gene (~720bp) has been successfully inserted into HIV-1 genome at two different positions. In HIV-1-iGFP, *gfp* gene is inserted between MA and CA.<sup>134</sup> In HIV-1-NL-GI or pNL-GI, *gfp* gene is inserted at the 5'-end of *nef* gene.<sup>135</sup> We compared the effects of genomic insertion sites by evaluating the infectivity of these two HIV-1 variants using TZM-bl cell based infection assay. As shown in **Figure 16**, both HIV-1-iGFP and HIV-1-NL-GI showed strong infections, and HIV-1-NL-GI exhibited relatively higher infectivity. In addition, the study of  $\Delta$ *nef*-HIV vaccine showed that the deletion in the *nef* gene and surrounding regions is tolerable. Thus, we decided to insert encoding genes of the aaRS/tRNA pair into *nef* region.

### 2.2.2.2 Construction and evaluation of HIV-aaRS variants

To investigate the genomic insertion of aaRS/tRNA encoding gene at the *nef* region, we chose pNL4-3 (catalog #114, from Dr. Malcolm Martin through NIH AIDS Reagent Program) which encodes a full-length HIV-1 genome as the parental template for subsequent cloning. We focused

on *E. coli* based pAzFRS/tRNA<sup>Tyr</sup> pair, which has successfully worked as a genetic switch for regulating single-cycle HIV-1 variants. pNL43-AzFRS was constructed by inserting pAzFRS-encoding gene (1284 bp, 52% GC content) at the intergenic region between *env* and *nef* gene. As a control, EGFP-encoding gene (720 bp, 62% GC content) was inserted at the same position to afford pNL43-EGFP. Each of the above two constructs was transfected into 293 T cells and the viruses were used to infect TZM-bl cells. As shown in **Figure 17**, both pNL43-AzFRS and pNL43-EGFP resulted in the assembly of infectious viruses, albeit with lower titers comparing to the wild-type pNL4-3 (pNL43-wt). To further validate the genetic plasticity at *nef* region, we constructed pNL43-BocKRS by inserting BocKRS-encoding gene (1260 bp, 43% GC content) at the same position. Similar to ONBY-RS and ONBK-RS, BocKRS was derived from PylRS. The BocKRS/tRNA<sup>Pyl</sup> pair specifically recognizes N<sup>ε</sup>-(*tert*-butoxycarbonyl)-L-lysine (BocK) and can be potentially used as another class of unnatural genetic switch for the generation of HIV-1 LAVs. As shown in **Figure 17**, pNL43-BocKRS also led to a moderate infectivity, which was relatively lower than that of pNL43-wt. Again, it confirms that the intergenic region between *env* and *nef* gene is a proper locus for the insertion of foreign DNA with varying length and GC contents.

### 2.2.2.3 Construction and evaluation of HIV-AzFRS-tRNA variants

Next, we inserted another essential control element of the genetic switch, the amber suppressor tRNA, into HIV-1 genome. A human RNA polymerase III U6 promoter was used to drive the transcription of prokaryotic tRNA<sup>Tyr</sup>. Four new constructs were generated by inserting the U6-tRNA<sup>Tyr</sup> cassette (351bp, 30% GC content) at different loci with different orientations. First, the U6-tRNA<sup>Tyr</sup> cassette was inserted between *AzFRS* and *nef*, in the same (pNL43-AzFRS-tRNA-1) or the opposite (pNL43-AzFRS-tRNA-2) direction of *AzFRS* transcription. Then, the tRNA transcription cassette was inserted between *env* and *AzFRS*, in the same (pNL43-AzFRS-tRNA-3)

or the opposite (pNL43-AzFRS-tRNA-4) direction of *AzFRS* transcription. These constructs were designed to cover all possible combinations of *AzFRS* and U6-tRNA<sup>Tyr</sup> cassettes at the *env/nef* locus. By examining these constructs (**Figure 18A**), we also evaluated the potential polar effects resulting from transcription and translation of neighboring genes. We performed TZM-bl-based infection assay to assess the infectivity of these constructs. As shown in **Figure 18B**, the insertion of U6-tRNA<sup>Tyr</sup> between *AzFRS/nef* led to strong infectivity in both pNL43-AzFRS-tRNA-1 and pNL43-AzFRS-tRNA-2. While the insertion at *env/AzFRS* only led to pNL43-AzFRS-tRNA-4 that displayed similar level of infectivity.

Next, we examined whether the inserted *AzFRS* and U6-tRNA<sup>Tyr</sup> cassettes can function properly as a genetic switch in mammalian cells. To achieve this, we devised a fluorescence-based functional assay by co-transfecting each of these four constructs with a plasmid pEGFP into 293T cells. Since pEGFP has an in-frame amber mutation, only when the p*AzFRS*-tRNA<sup>Tyr</sup> pair was transcribed and processed correctly, they could decode the amber mutation and lead to fluorescence signal in the presence of p*AzF*. As shown in **Figure 18C**, the inserted p*AzFRS*/tRNA<sup>Tyr</sup> pair was functional in all four constructs.

#### 2.2.2.4 Construction and evaluation of multicycle HIV-UAG-AzFRS-tRNA variants.

Since pNL43-AzFRS-tRNA-2 gave the best infectivity and amber suppression efficiency, we decided to use it for subsequent cloning. Based on pNL43-AzFRS-tRNA-2, we generated two new constructs by installing amber stop codon at either position 36 of matrix protein (pNL-Trp36-AzFRS-tRNA-2) or at position 59 of protease (pNL-Tyr59-AzFRS-tRNA-2).

First, we examined the infection activity of these two constructs by TZM-bl cell-based infection assay and TCID<sub>50</sub> measurement. As shown in **Figure 19A**, both of the pNL-Trp36-



AzFRS-tRNA-2 and pNL-Tyr59-AzFRS-tRNA-2 variants led to infectious viruses in the presence of pAzF, while no live viruses were generated in the absence of pAzF. According to the TCID<sub>50</sub> values (**Figure 19B**), the infectivity of both constructs was significantly attenuated compared to that of the wild-type pNL4-3. However, the number of viral particles formed by these two constructs were close to that of wild-type viruses (**Figure 19B**). The discrepancy between total number of viral particles and the number of virions (infectious viral particles) produced by pNL-Trp36-AzFRS-tRNA-2 and pNL-Tyr59-AzFRS-tRNA-2 indicated that the amber suppression did not significantly affect the function of HIV-1 genome, however, it might interfere with the generation of infectious viruses at the translational level.

Next, we examined the multicycle replication of pNL-Trp36-AzFRS-tRNA-2 and pNL-Tyr59-AzFRS-tRNA-2 variants. These two constructs were transfected into 293T cells in the presence of 1 mM pAzF. 48 h post transfection, viruses were harvested and used to infect SupT1 cells that express high level of CD4 on the surface and can support multicycle replication of HIV-1. One day post infection, SupT1 cells were washed twice to remove any free virus particles in the culture media. Then the infected SupT1 cells were cultivated in the presence or absence of 1 mM pAzF. As shown in **Table 2**, multicycle replication of pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 were achieved. According to TCID<sub>50</sub> measurements, mild but continuous replication was observed with both constructs in SupT1 cells after two weeks' cultivation in the presence of pAzF. pNL43-Tyr59-AzFRS-tRNA-2 displayed better infectivity than pNL43-Trp36-AzFRS-tRNA-2. In the control experiment, no live viruses were generated after two weeks' cultivation in the absence of pAzF, suggesting a good stringency of the unnatural genetic switch. Comparing with the parental wild-type pNL4-3 viruses, infectivity of both constructs was

significantly lower. In addition, pNL4-3 showed slightly lower titer in the presence of pAzF than in the absence of pAzF.

We examined HIV-1 viral RNA (vRNA) and viral DNA (vDNA) in SupT1 cells that were infected with pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 using RNAscope *in situ* hybridization and DNAscope ISH. Viral transcription was observed both in the presence and absence of pAzF. In the presence of pAzF, significant higher level of viral transcription was observed in both constructs. In addition, the transcription level of pNL43-Trp36-AzFRS-tRNA-2 was slightly lower than that of pNL43-Tyr59-AzFRS-tRNA-2, which is consistent with the infectivity measurements. As a negative control, no transcription was detected in the uninfected SupT1 cells. As a positive control, robust transcription signals were detected in ACH2 cells with PMA stimulation. Based on DNAscope ISH, integration of pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 DNA was observed in infected SupT1 cells in the presence of absence of pAzF.

## **2.2.3 Attempts to improve the replication capacity of amber suppression-dependent HIV-1 variants**

### **2.2.3.1 Construction and examination of an amber-codon-free HIV-1 variant**

The wild-type pNL4-3 uses five amber stop codons to terminate protein translation. When applying the unnatural genetic switch in HIV-1, the amber suppressor tRNA may outcompete release factors and lead to read-through at these amber stop codons, which may cause negative effects on HIV-1 protein synthesis, virus assembly, and infectivity. Therefore, we decided to construct an amber-codon-free HIV-1 variant. To achieve this, we first replaced the amber stop codon (UAG) of HIV-1 *vif*, *vpu*, *vpr*, and *rev* genes with ochre stop codon (UAA) respectively to

afford pNL43-vif-UAA, pNL43-vpu-UAA, pNL43-vpr-UAA, and pNL43-rev-UAA. As shown in **Figure 21**, the infectivity of these four mutants was similar to that of wild-type pNL4-3 virus. Next, we generated an amber-codon-free pNL4-3 mutant, pNL43-all-UAA, by replacing all five UAG codons with UAA codons. According to the infection assay, pNL43-all-UAA retains a similar level of infection activity as the parental pNL4-3 virus, suggesting that amber stop codons were nonessential in HIV-1 virus. By preventing undesired amber suppression, the amber-free construct provides a useful tool for HIV-1 LAV research.

### 2.2.3.2 Attempts to increase transcription level of the aaRS/tRNA pair

As shown in **Figure 6** and **Figure 18**, the intensity of fluorescence signal was significantly weakened after transferring the encoding genes of pAzFRS/tRNA<sup>Tyr</sup> pair onto HIV-1 genome, which suggested a reduction in its amber suppression activity. Since the inherent enzymatic activity of pAzFRS remains unchanged, we reasoned that the reduction might result from position effect of the pAzFRS-U6-tRNA<sup>Tyr</sup> cassette.

To confirm our hypothesis, we measured the transcription level of pAzFRS. As shown in **Figure 22A**, the levels of pAzFRS mRNA transcribed from pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 were significantly lower than that from a separate plasmid pAzFRS. To increase the transcription level of pAzFRS in HIV-1 genome, we introduced a strong CMV promoter in the front of *AzFRS* gene to generate pNL43-Trp36-CMV-AzFRS-tRNA-2 and pNL43-Tyr59-CMV-AzFRS-tRNA-2 mutant. However, no improvement in infectivity was observed with these mutants **Figure 22B**. While when the pAzFRS/tRNA<sup>Tyr</sup> pair was supplied on a separate plasmid, the infectivity was significantly improved (**Figure 22B**), indicating that the transcription

level of the pAzFRS/tRNA<sup>Tyr</sup> pair may be one of the limiting factors for the low replication capacity of amber-suppression dependent HIV-1 variants.

Several research groups have demonstrated that the use of multiple tRNA transcription cassettes could increase the amber suppression efficiency in mammalian cells.<sup>136,137</sup> Using pNL43-AzFRS-tRNA-2 as a parental plasmid, two constructs pNL43-AzFRS-2tRNA-F and pNL43-AzFRS-2tRNA-R were generated by inserting an additional U6-tRNA<sup>Tyr</sup> cassette between *env* and *AzFRS* gene at the same or opposition direction of *AzFRS* gene. We evaluated whether transcription from the novel tRNA<sup>Tyr</sup>-pAzFRS-tRNA<sup>Tyr</sup> cassette was beneficial for amber suppression. Based on fluorescence assay, the inserted cassette was functional (**Figure 23**). But unfortunately, no live virus was generated, suggesting that the insertion of the second tRNA transcription cassette may have interfered with the expression or function of viral proteins.

### 2.2.3.3 Supplement of viral proteins

We reasoned that the insertion of amber stop codon and foreign DNA in essential genes of HIV-1 may interfere the expression of downstream viral proteins. To test our hypothesis, we replenished each viral protein from a separate plasmid to assist viral production. We co-transfected pNL43-Tyr59-AzFRS-tRNA-2 variant with a plasmid<sup>138-144</sup> that over-expresses one of the nine HIV-1 proteins, namely Gag, Env, Pol, Nef, Tat, Vif, Vpr, Rev, or Vphu into 293T cells. Since over-expression of protease is toxic to the cells, only integrate and reverse transcriptase were included in the Pol supplement experiment. 48 h after transfection, the viruses were harvested and used to infect TZM-bl cells. As shown in **Figure 24**, there was no significant difference between each testing group. One plausible explanation is that the amber suppression in the protease did not affect the expression or function of other viral proteins but posed negative effects on protease itself.

#### 2.2.3.4 Effect of unAA concentration on amber suppression

Generally, 1 mM unAA is employed for amber suppression both *in vitro* and *in vivo*, *albeit* higher concentration (up to 10 mM) of unAA was also well tolerated.<sup>12,136,137</sup> We attempted to reduce the working concentration of unAA for amber suppression. Firstly, a low concentration of unAA may reduce potential long-term toxicity of our vaccine strains. Secondly, it will be more economically feasible for future investigations on animal models. To achieve this, we determined amber suppression with different concentrations of pAzF in 293T cells. As shown in **Figure 25**, lower concentration of pAzF, such as 0.1 mM, led to efficient amber suppression. There was no significant difference in the expression levels of GFP between 0.1 and 1.0 mM pAzF test groups. The above results are consistent with a previous report where an even lower concentration (0.01 mM) of unAA was sufficient for protein expression.<sup>7</sup> Therefore, the concentration of unAA can be adjusted within a large dynamic range based on different experiment requirements, and should not be a limiting factor in our vaccine strategy.

### 2.3 Summary and remarks

In this chapter, we demonstrated the use of amber suppression machinery as an unnatural genetic switch to turn on and off HIV-1 replication through the addition or the clearance of an unAA. Based on this strategy, two generations of amber suppression-dependent HIV-1 variants have been constructed and examined. The first-generation vaccine variants contain one or more in-frame amber mutations within essential genes of HIV-1. Consequently, the biosynthesis of corresponding viral proteins becomes dependent on the presence of the exogenous amber-suppressing aaRS/tRNA pair and an unAA. Since the host cells do not have the amber suppression machinery, the viruses can only complete one cycle of infection or replication and may not be able

to mount sufficient protective immune responses. To solve this problem, the aaRS-tRNA encoding gene was inserted onto HIV-1 genome to generate the second-generation all-in-one HIV-1 variants. Multicycle replication of HIV-1 has been achieved through co-localization of the amber suppression machinery and HIV-1 provirus in the same infected cells. These multicycle attenuated HIV-1 variants can be employed to mimic the prime-boost vaccine regimen. After obtaining desired immune responses with the prime, the replication of the vaccine strain can be switched off by simply withdrawing the unAA. Once a booster vaccination is needed, the replication of HIV-1 can be turned on transiently by providing the unAA and turned off later. This approach can be potentially extended for the development of LAVs for other pathogens.

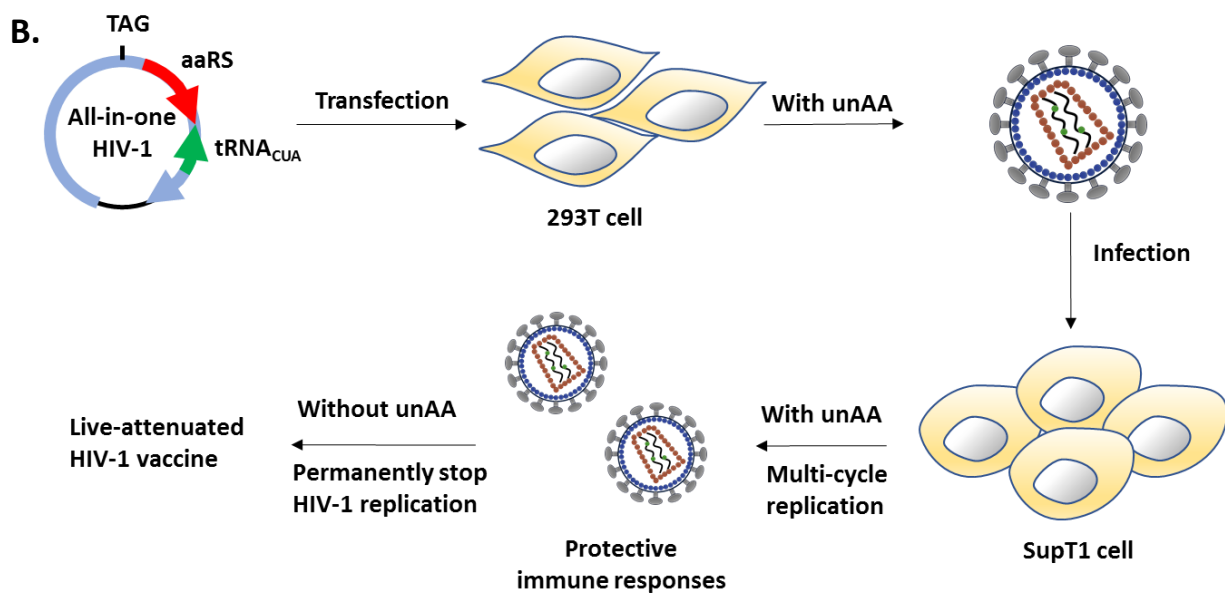
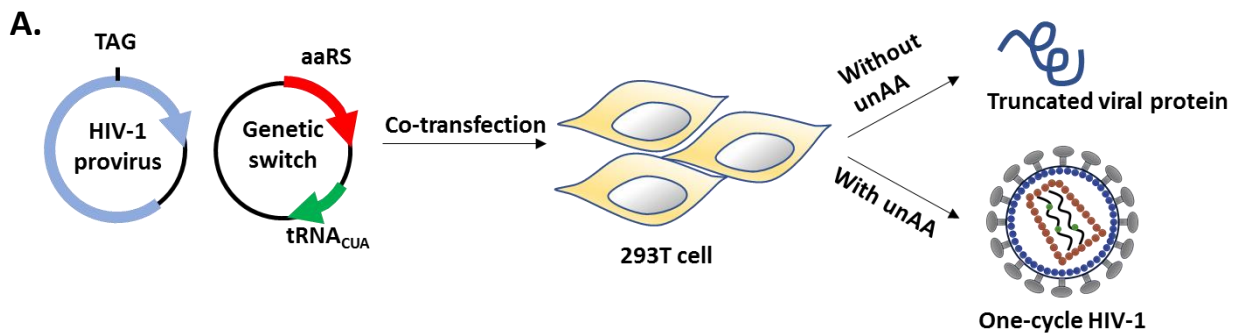
Our strategy provides several advantages over conventional HIV-1 LAV strategies. The safety of  $\Delta nef$ -LAVs has been questioned due to their pathogenic potentials. Multiple deletions or mutations in accessory genes may improve the safety of these vaccine strains,<sup>145</sup> but they usually lead to a significant reduction in protective efficacy due to decreased replication capacity. We prevented uncontrolled virus replication by introducing an unnatural genetic switch into HIV-1. The replication of our vaccine strains can be permanently switched off by simply discontinuing the administration of the unAA. In addition, our vaccine strains maintain the intact HIV-1 genome and all of the viral immunogens. The only changes are the substitutions of natural amino acids with their unnatural amino acid analogs at permissive positions as well as the insertion of a genomic copy of aaRS-tRNA encoding gene. Theoretically, these changes should not significantly reduce the vaccine efficacy. The immunogenicity of our vaccine strains can be further improved by the incorporation of immunogenic unAAs.<sup>21-23</sup>

One concern for all HIV-1 LAVs is the possibility to regain virulence during virus replication due to error-prone transcriptions. In our strategy, this can be prevented by the installation of

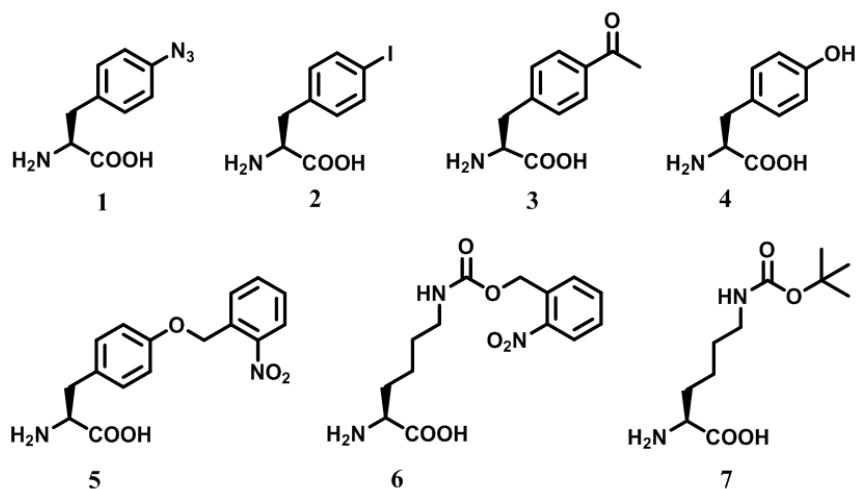
multiple amber stop codons. Based on our calculations, the possibility of reverting all amber codons to sense codons is approximately  $1.0 \times 10^{-2}$ ,  $1.0 \times 10^{-4}$ , and  $1.0 \times 10^{-6}$  respectively (in 100 years) when one, two, and three amber codons were used. In fact, this chance can be even lower. Based on a recent amber-suppression dependent influenza A vaccine, the escape frequencies in mammalian cells decreased from  $1.0 \times 10^{-8}$  to  $1.2 \times 10^{-10}$  and to less than  $1.0 \times 10^{-11}$  when one, two, and three ambers codons are presented.<sup>24</sup> Another concern of our vaccine strategy is the genetic stability of the inserted aaRS-tRNA encoding gene, which may lose the control activity due to mutations or deletions. Although the loss of pAzFRS-tRNA<sup>Tyr</sup> cassette might occur in a very long term, during our experiments (2 weeks' cultivation) we did not observe such event according to RT-PCR and DNA sequencing results. In fact, the loss of pAzFRS-tRNA<sup>Tyr</sup> function will result in a replication-incompetent strain, which can be beneficial in terms of safety.

In summary, we have demonstrated that the replication and viability of HIV-1 can be precisely controlled by an unnatural genetic switch based on genetic code engineering. The ability to control the timing and duration of HIV-1 replication represents an important step towards the construction of a safe and effective live-attenuated HIV-1 vaccine. In addition, unAA mutagenesis of HIV-1 viral proteins may enhance their biochemical or immunological properties, and facilitate the study of HIV-1. This strategy can be potentially extended to the development of other live-attenuated vaccines.

## 2.4 Figures and tables

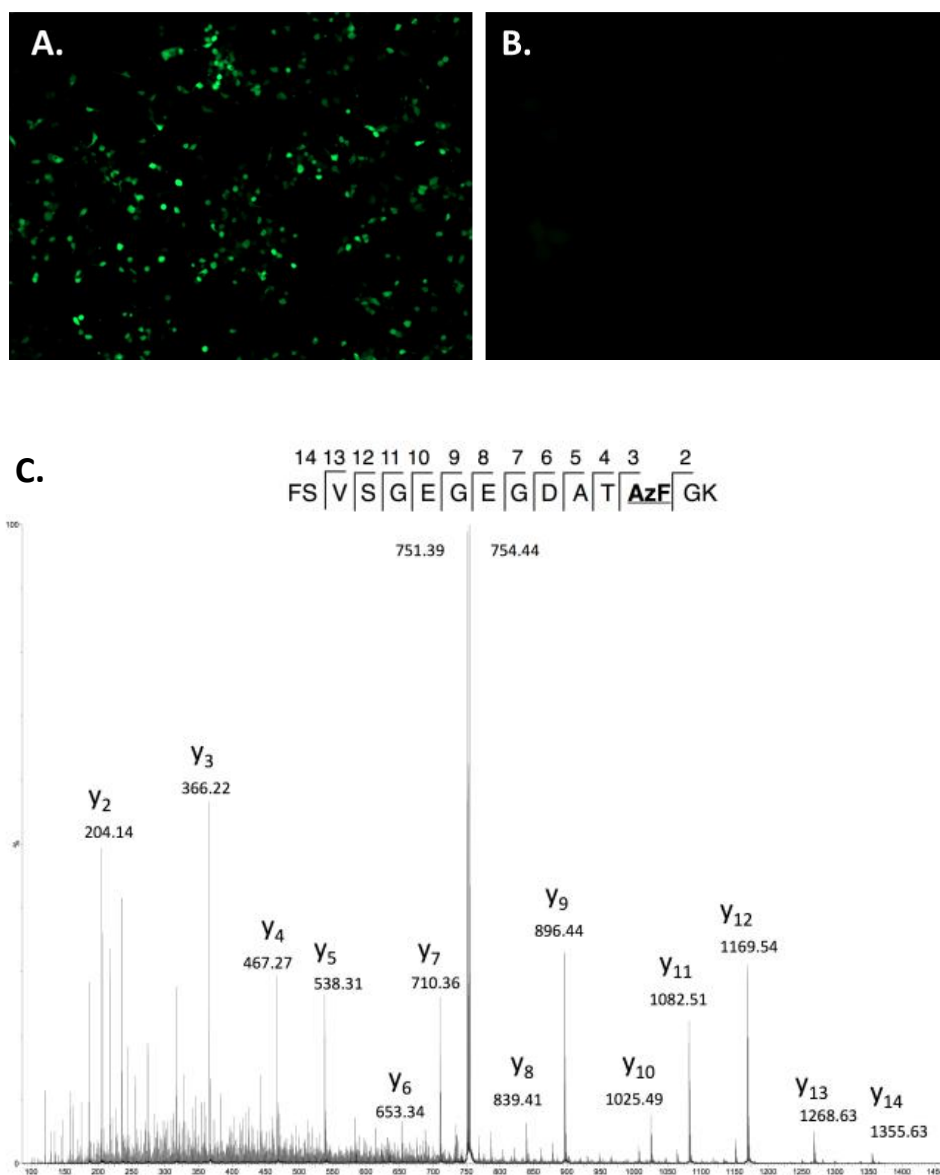


C.



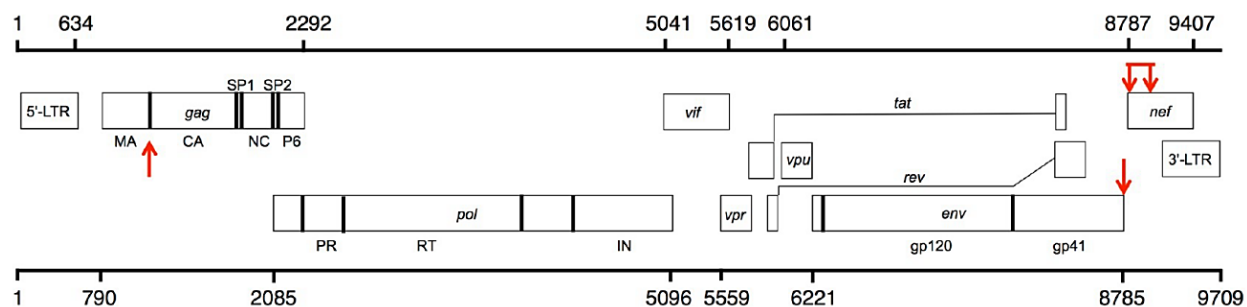


**Figure 5.** An unnatural genetic switch for regulation of HIV-1 replication. (A) Amber mutations were installed in the essential genes of HIV-1. The replication of resulted HIV-1 variants can be turned on in the presence of unAA and exogenously expressed unnatural genetic switch (aaRS/tRNA<sub>CUA</sub> pair). In the absence of unAA, the viral replication is turned off. Since host cells do not have unnatural genetic switch, the virus can only replicate for single-cycle. (B) A genomic copy of unnatural genetic switch was inserted onto HIV-1 genome to afford all-in-one HIV-1 variants, the replication of which is under stringent control of the presence or absence of unAA. Since the aaRS/tRNA<sub>CUA</sub> pair is co-expressed with essential viral proteins, these HIV-1 variants can undergo multi-cycle replication. (C) Natural and unnatural amino acids used in this chapter, namely: 1, *p*-azido-L-phenylalanine (pAzF); 2, *p*-iodo-L-phenylalanine (pIodoF); 3, *p*-acetyl-L-phenylalanine (pAcF); 4, tyrosine; 5, *o*-nitrobenzoyl-tyrosine (ONBY); 6, *o*-nitrobenzyl-oxycarbonyl-N<sup>ε</sup>-L-lysine (ONBK); and 7, N<sup>ε</sup>-(*tert*-butoxycarbonyl)-L-lysine (BocK).



**Figure 6.** Amber suppression of EGFP in 293T cells. (A) EGFP expression in the presence of pAzFRS/tRNA<sup>Tyr</sup> pair and 1 mM pAzF; (B) EGFP expression in the presence of pAzFRS/tRNA<sup>Tyr</sup> pair, but without pAzF; (C) Tandem-mass spectra of EGFP-Y40pAzF. The reduction product (4-aminophenylalanine) of pAzF was detected exclusively at position 40 of EGFP.

Reproduced with permission from “Angewandte Chemie International Edition. 2014, 53, 4867.”

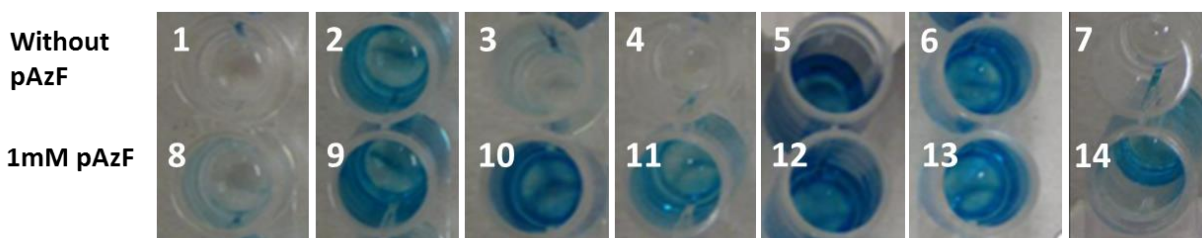


**Figure 7.** The HIV-1 genome organization. Abbreviations: MA, p17 matrix protein; CA, p24 capsid protein; NC, nucleocapsid protein; SP1, Spacer peptide 1; SP2, Spacer peptide 2; LTR, long terminal repeat; PR, protease; RT, reverse transcriptase; IN, integrase. The red arrows indicated insertion sites for foreign genes.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”

SUMA mutants	Gene	Protein	Mutations
MA-Y132	<i>gag</i>	MA	Y132TAG
MA-A119	<i>gag</i>	MA	A119TAG
MA-W36Q127	<i>gag</i>	MA	W36TAG, Q127TAG
MA-F44	<i>gag</i>	MA	F44TAG
CA-K359	<i>gag</i>	CA	K359TAG
CA-L268	<i>gag</i>	CA	L268TAG
PR-Y59	<i>pol</i>	PR	Y59TAG, G226S
RT-F216	<i>pol</i>	RT	F216TAG
RT-F271	<i>pol</i>	RT	F271TAG
RT-L365	<i>pol</i>	RT	L365TAG
Tat-T20	<i>tat</i>	Tat	T20TAG

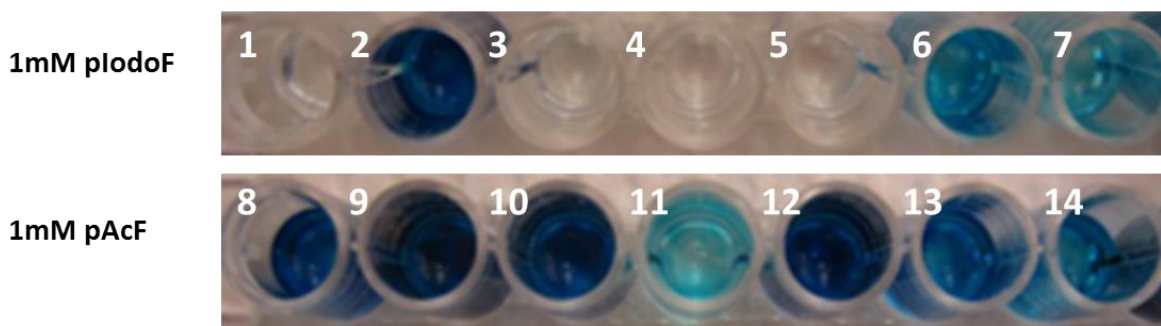
**Table 1.** pSUMA variants being constructed and examined in this work.



OD values of above p24 assay

Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7
0.062	2.734	0.499	0.061	3.981	3.613	0.067
Well 8	Well 9	Well 10	Well 11	Well 12	Well 13	Well 14
0.514	3.607	3.801	2.188	3.669	3.773	1.425

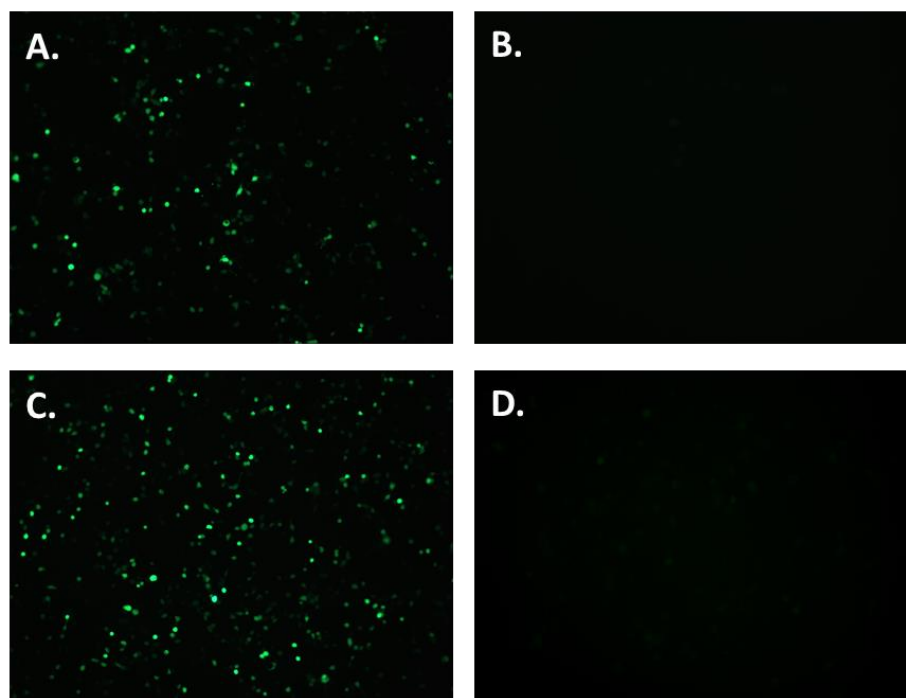
**Figure 8.** p24 assay after transfection with pSUMA variants and pAzFRS/tRNA<sup>Tyr</sup> pair. (Well 1-6) Co-transfection of pAzFRS/tRNA<sup>Tyr</sup> pair with the following pSUMA variants in the absence of pAzF: (well 1) MA-Y132; (well 2) RT-F216; (well 3) Tat-T20; (well 4) MA-A119; (well 5) RT-F271; (well 6) RT-L365. (Well 8-13) Co-transfection of pAzFRS/tRNA<sup>Tyr</sup> pair with the following pSUMA variants in the presence of 1 mM pAzF: (well 8) MA-Y132; (well 9) RT-F216; (well 10) Tat-T20; (well 11) MA-A119; (well 12) RT-F271; (well 13) RT-L365. (Well 7) negative control, no p24; (well 14) positive control, 125 pg/mL p24.



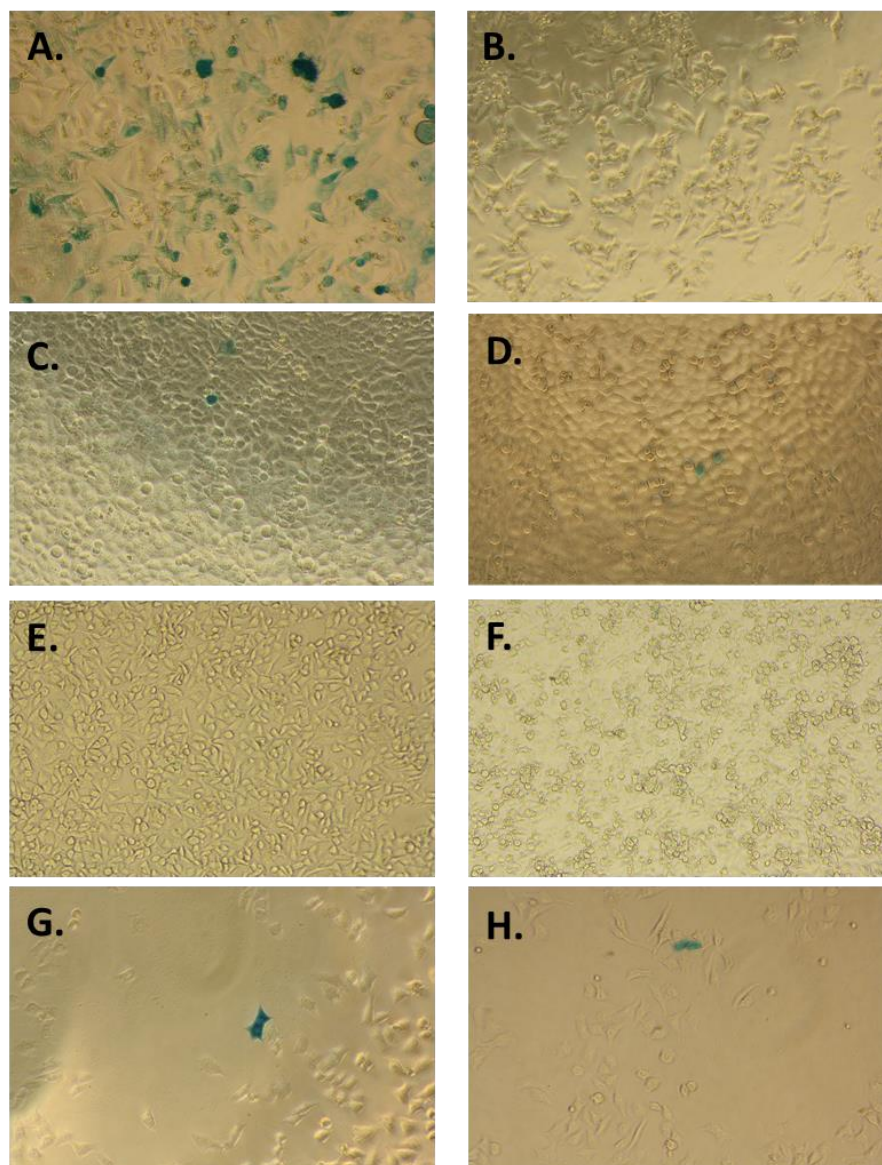
OD values of above p24 assay

Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7
0.064	OVRFLW	0.053	0.046	0.052	2.463	1.973
Well 8	Well 9	Well 10	Well 11	Well 12	Well 13	Well 14
OVRFLW	OVRFLW	OVRFLW	1.425	OVRFLW	OVRFLW	3.262

**Figure 9.** p24 assay after transfection with pSUMA variants and pIodoFRS/tRNA<sup>Tyr</sup> or pAcFRS/tRNA<sup>Tyr</sup> pair. (Well 1-7) Co-transfection of pIodoFRS/tRNA<sup>Tyr</sup> pair with the following pSUMA variants in the presence of 1 mM plodoF: (well 1) MA-Y132; (well 2) RT-F216; (well 3) Tat-T20; (well 4) MA-A119; (well 5) CA-L268; (well 6) RT-F271; (well 7) RT-L365. (Well 8-14) Co-transfection of pAcFRS/tRNA<sup>Tyr</sup> pair with following pSUMA variants in the presence of 1 mM pAcF: (well 8) MA-Y132; (well 9) RT-F216; (well 10) Tat-T20; (well 11) MA-A119; (well 12) CA-L268; (well 13) RT-F271; (well 14) RT-L365.



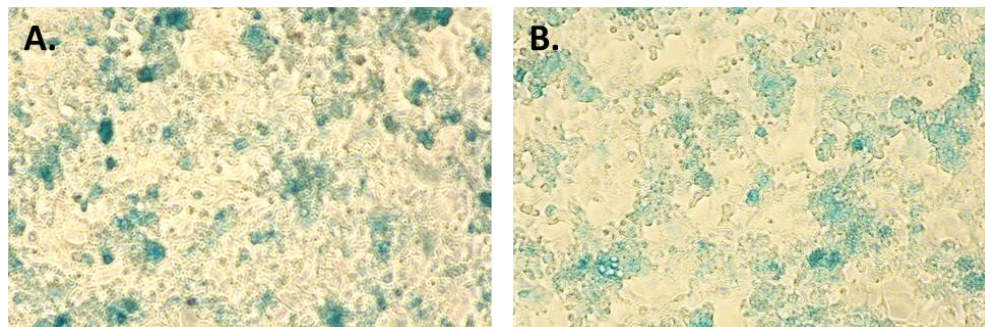
**Figure 10.** Control the expression of EGFP with photocaged unAAs. (A) EGFP expression in the presence of the ONBY-RS/tRNA<sup>Pyl</sup> pair and 0.4 mM ONBY; (B) EGFP expression in the presence of the ONBY-RS/tRNA<sup>Pyl</sup> pair; (C) EGFP expression in the presence of the ONBK-RS/tRNA<sup>Pyl</sup> pair and 1 mM ONBK; (D) EGFP expression in the presence of the ONBK-RS/ tRNA<sup>Pyl</sup> pair.



**Figure 11.** Control the replication of HIV-1 by amber suppression.

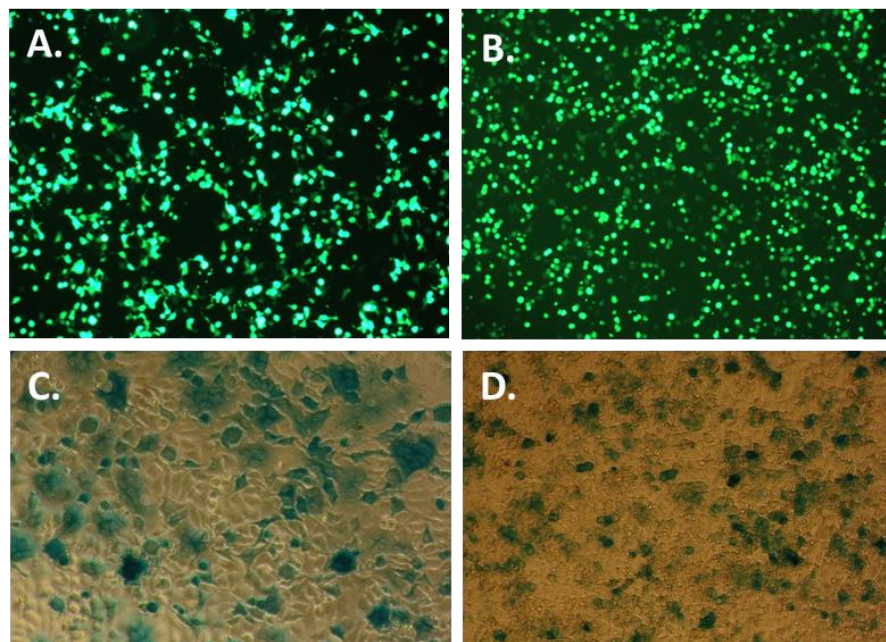
(A-H) Infection assays with TZM-bl cells. (A) MA-Y132 + TyrRS/tRNA<sup>Tyr</sup> pair; (B) MA-Y132; (C) RT-F216 + TyrRS/tRNA<sup>Tyr</sup> pair; (D) RT-L365 + TyrRS/tRNA<sup>Tyr</sup> pair; (E) MA-Y132 + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF; (F) MA-Y132 + pAzFRS/tRNA<sup>Tyr</sup> pair; (G) MA-A119 + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF; (H) RT-L365 + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF.





**Figure 12.** Infection of TZM-bl with pSUMA-wt.

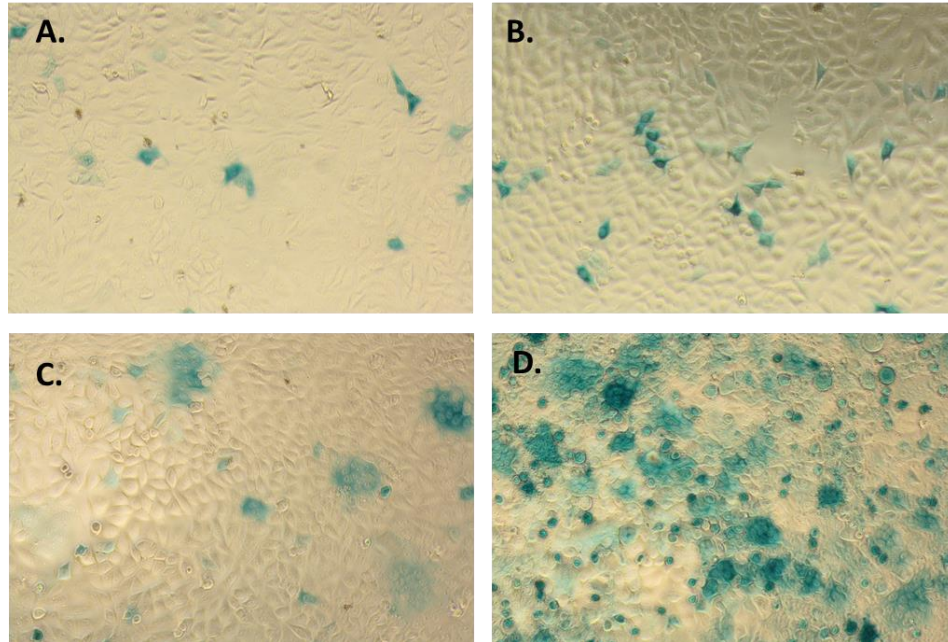
(A) pSUMA-wt; (B) pSUMA-wt + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF.



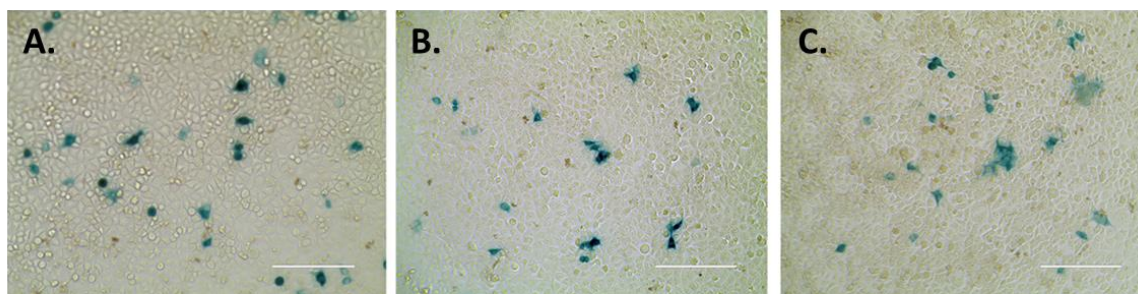
**Figure 13.** Expression and infection with pNL-GI variants.

(A) Fluorescence image of wild-type pNL-GI; (B) Fluorescence image of pNL-GI-N40TAG + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF; (C) Infection of TZM-bl cells with viruses collected from (A); (D) Infection of TZM-bl cells with viruses collected from (B).

Reproduced with permission from “Angewandte Chemie International Edition. 2014, 53, 4867.”



**Figure 14.** Infection of TZM-bl cells with pSUMA variants in the presence of TyrRS/tRNA<sup>Tyr</sup> pair. (A-D) Infection assay with TZM-bl cells. (A) PR-Y59 + TyrRS/tRNA<sup>Tyr</sup> pair; (B) MA-F44 + TyrRS/tRNA<sup>Tyr</sup> pair; (C) MA-W36Q127 + TyrRS/tRNA<sup>Tyr</sup> pair; (D) pSUMA-wt.



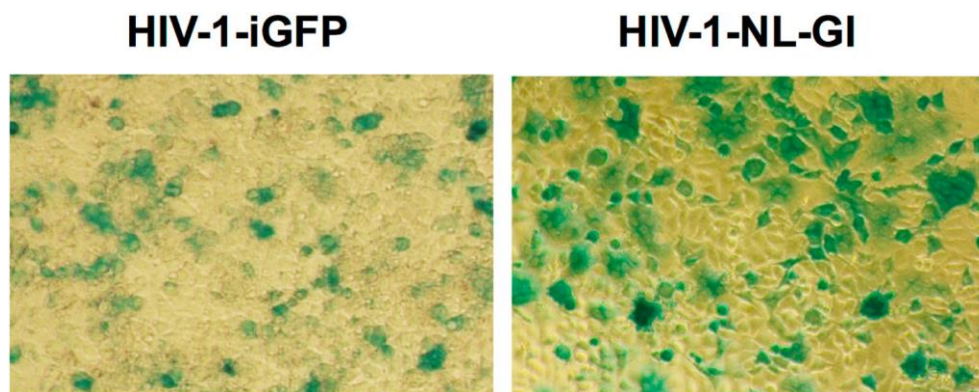
**Tissue Culture Infectious Dose 50 (TCID<sub>50</sub>/ml)**

pSUMA-wt	PR-Y59	MA-W36Q127
$5.10 \times 10^3$	$1.31 \times 10^3$	$5.12 \times 10^3$

**Figure 15.** Control the replication of HIV-1 with an unnatural genetic switch. (A-C) Infection assay with TzM-bl cells. (A) pSUMA-wt; (B) PR-Y59 + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF; (C) MA-W36Q127 + pAzFRS/tRNA<sup>Tyr</sup> pair, with 1 mM pAzF. Scale bars, 200 nm. The doses used for infection are shown as TCID<sub>50</sub> values. The wild-type pSUMA was diluted 4-fold and the pSUMA mutants were concentrated 15-fold in order to adjust the infectious dose of each variant to similar levels.

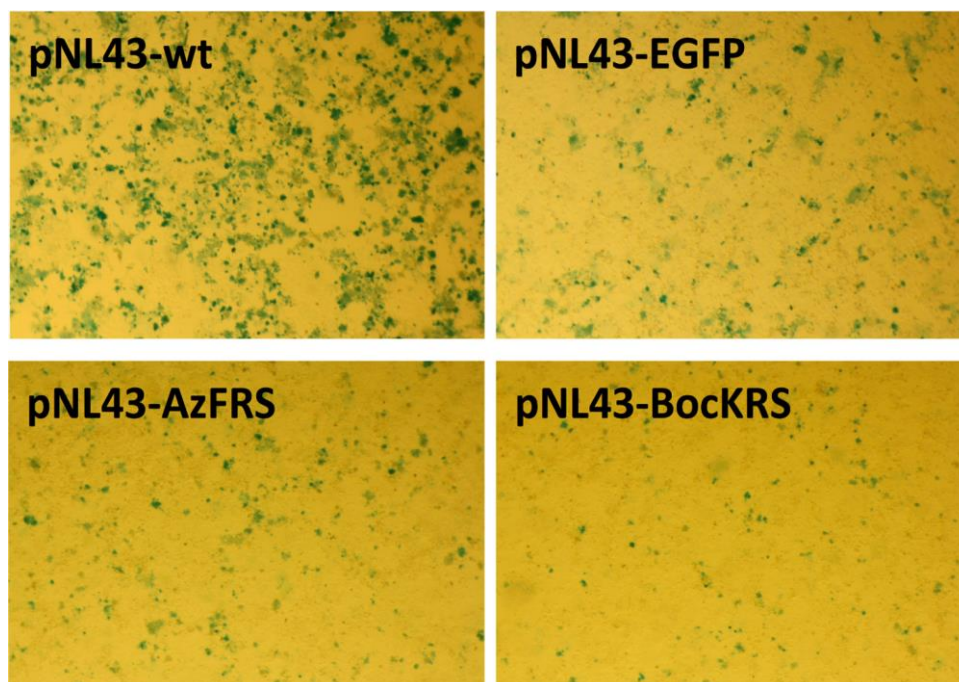
Reproduced with permission from “Angewandte Chemie International Edition. 2014, 53, 4867.”





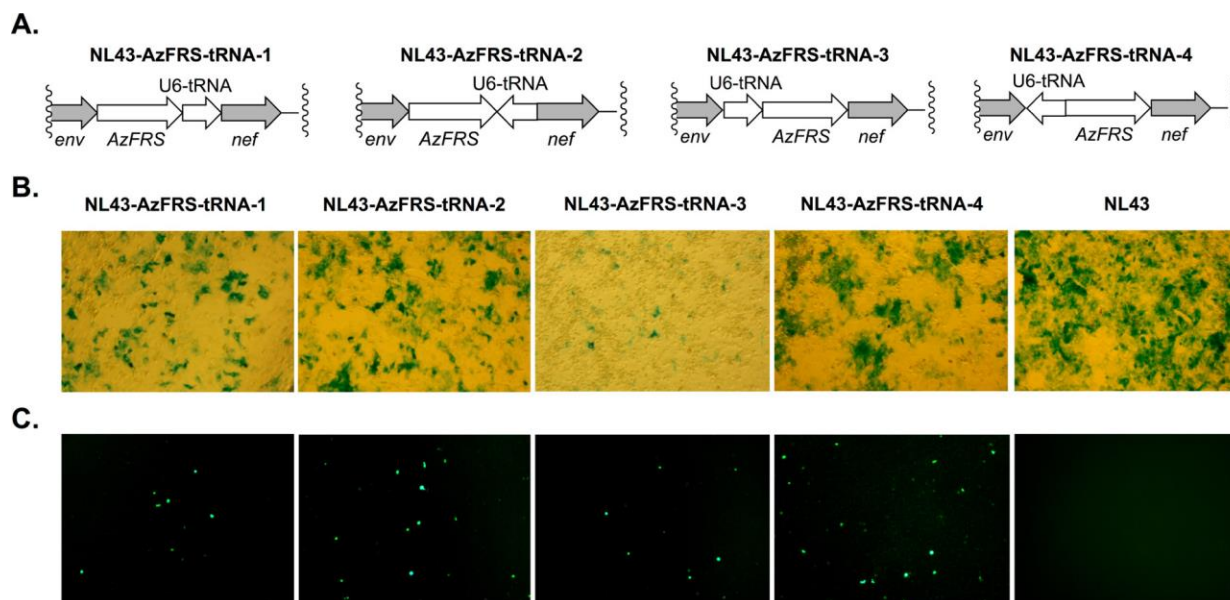
**Figure 16.** Infection of TZM-bl cells with HIV-1-iGFP and HIV-1-NL-GI strains.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”



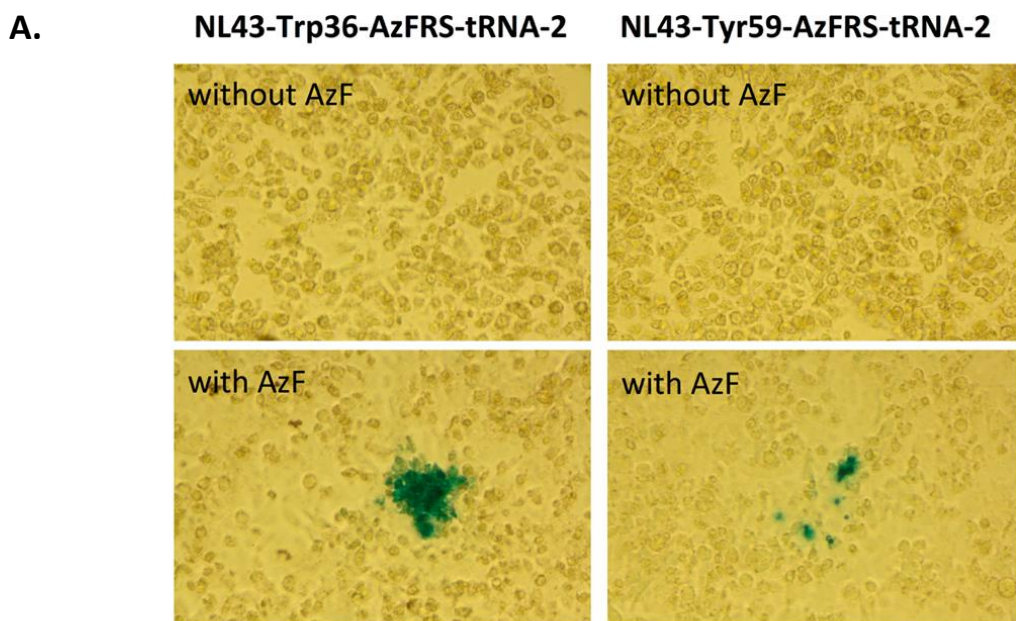
**Figure 17.** Relative replication efficiency of four pNL4-3 variants.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”



**Figure 18.** Construction and examination of HIV-AzFRS-tRNA variants. (A) The four HIV-AzFRS-tRNA variants with different positions and orientations of tRNA<sup>Tyr</sup>; (B) Infectivity assay of the four HIV-AzFRS-tRNA variants and the parental pNL43-wt virus; (C) Functional assay of the inserted AzFRS-tRNA<sup>Tyr</sup> pair of the four HIV-AzFRS-tRNA variants and the parental pNL43-wt virus in the presence of 1 mM pAzF.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”



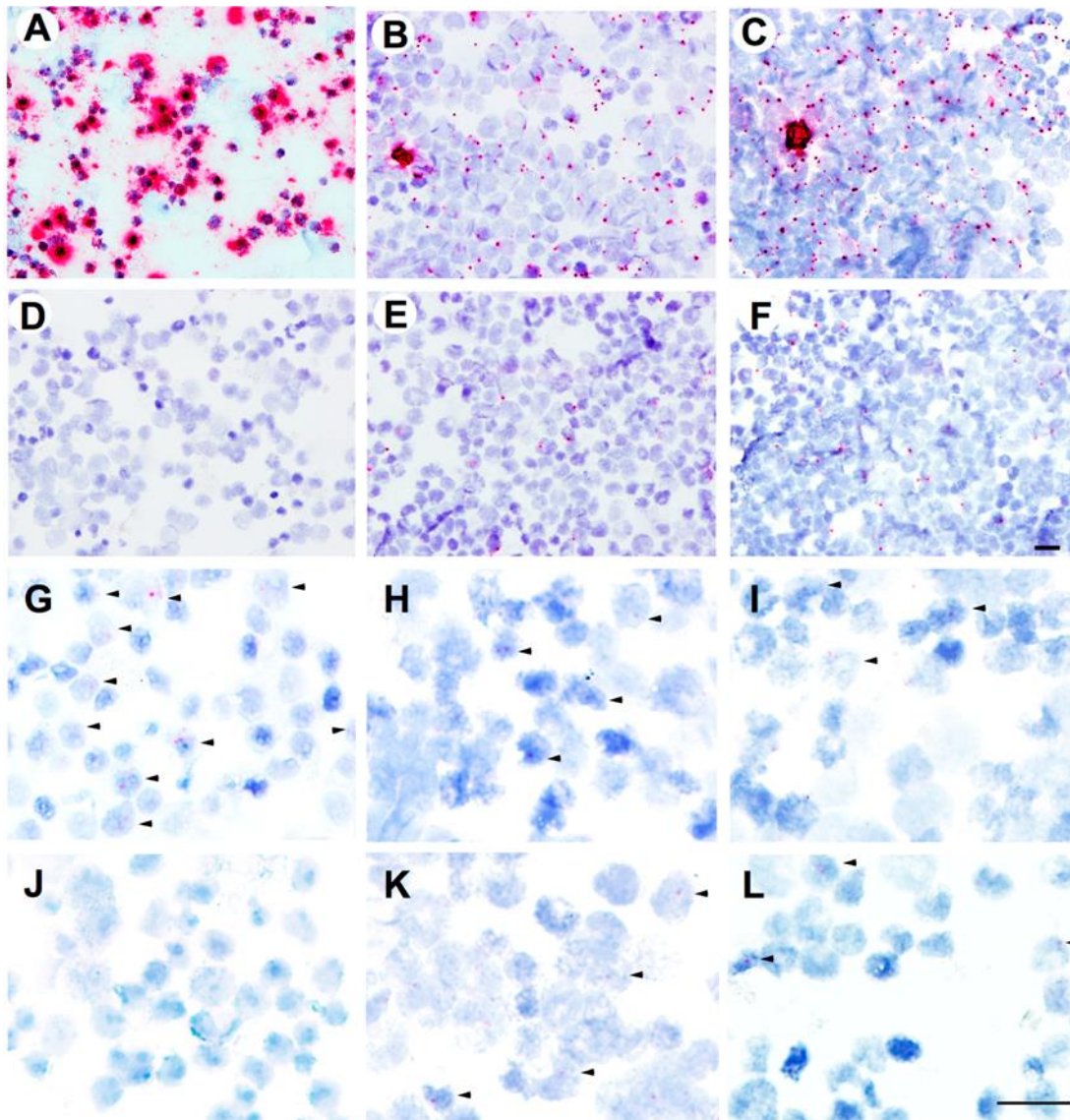
**B.**

HIV mutants	with AzF		without AzF	
	virion number/ml	TCID <sub>50</sub> /ml	virion number/ml	TCID <sub>50</sub> /ml
NL43-Trp36-AzFRS-tRNA-2	$3.45 \times 10^8$	$2.56 \times 10^3$	$2.55 \times 10^8$	0
NL43-Tyr59-AzFRS-tRNA-2	$6.69 \times 10^9$	$2.03 \times 10^3$	$3.03 \times 10^9$	0
NL43 (transfection control)			$6.68 \times 10^9$	$5.20 \times 10^7$

**Figure 19.** Infectivity assay of the pNL43-Trp36-AzFRS-tRNA-2 and the pNL43-Tyr59-AzFRS-tRNA-2 mutants in the presence and the absence of pAzF. (A) Infection of TZM-bl cells with pNL43-Trp36-AzFRS-tRNA-2 and the pNL43-Tyr59-AzFRS-tRNA-2 mutants; (B) TCID<sub>50</sub> values and virion numbers.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”





**Figure 20.** Viral RNA (vRNA) and DNA (vDNA) detection using RNAscope and DNAscope *in situ* hybridization (ISH). (A-F) *In situ* detection of HIV-1 vRNA. HIV-1 vRNA was detected using RNAscope ISH with V-HIV Clade B antisense probes and RNAscope 2.0 HD red reagent kit. (A) Positive control, ACH2 cells stimulated with PMA; (B) SupT1 cells infected with pNL43-Trp36-AzFRS-tRNA-2 and in the presence of pAzF; (C) SupT1 cells infected with pNL43-Tyr59-AzFRS-tRNA-2 and in the presence of pAzF; (D) Negative control, SupT1 cells without HIV-1 infection; (E) SupT1 cells infected with pNL43-Trp36-AzFRS-tRNA-2 and in the absence of pAzF;

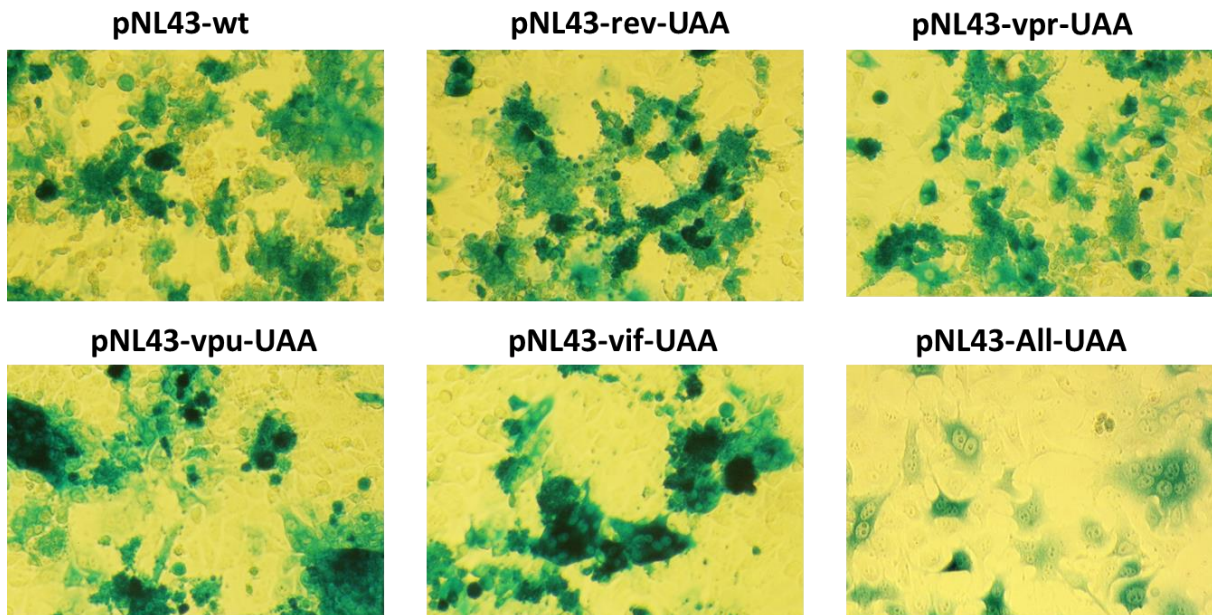
(F) SupT1 cells infected with pNL43-Tyr59-AzFRS-tRNA-2 and in the absence of pAzF. (G-L) *In situ* detection of HIV-1 vDNA. HIV-1 vDNA was detected using DNAscope ISH with V-HIV Clade B sense probes and RNAscope 2.0 HD red reagent kit. Arrows indicate the cells harboring HIV-1 vDNA (red dots). (G) Positive control, ACH2 cells; (H) SupT1 cells infected with pNL43-Trp36-AzFRS-tRNA-2 and in the presence of pAzF; (I) SupT1 cells infected with pNL43-Tyr59-AzFRS-tRNA-2 and in the presence of pAzF; (J) Negative control, SupT1 cells without HIV-1 infection; (K) SupT1 cells infected with pNL43-Trp36-AzFRS-tRNA-2 and in the absence of pAzF; (L) SupT1 cells infected with pNL43-Tyr59-AzFRS-tRNA-2 and in the absence of pAzF. Scale bars, 20  $\mu$ m.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”

HIV mutants	Week 0	Week 2 with AzF		Week 2 without AzF	
	TCID <sub>50</sub> /ml	virion number/ml	TCID <sub>50</sub> /ml	virion number/ml	TCID <sub>50</sub> /ml
NL43-Trp36-AzFRS-tRNA-2	0	$7.57 \times 10^7$	$1.10 \times 10^2$	0	0
NL43-Tyr59-AzFRS-tRNA-2	0	$1.10 \times 10^8$	$5.08 \times 10^2$	$1.02 \times 10^3$	0
NL43	0	$2.67 \times 10^8$	$2.62 \times 10^6$	$5.31 \times 10^9$	$5.24 \times 10^6$

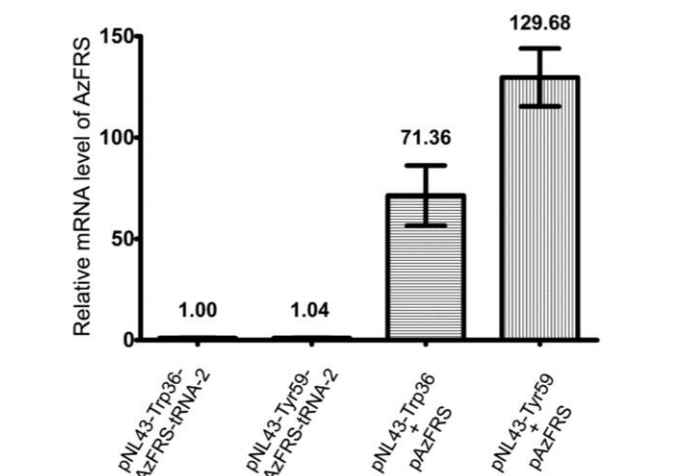
**Table 2.** Evaluation of the multi-cycle infectivity of the pNL43-Trp36-AzFRS-tRNA-2 and the pNL43-Tyr59-AzFRS-tRNA-2 mutants in the presence and the absence of pAzF.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”

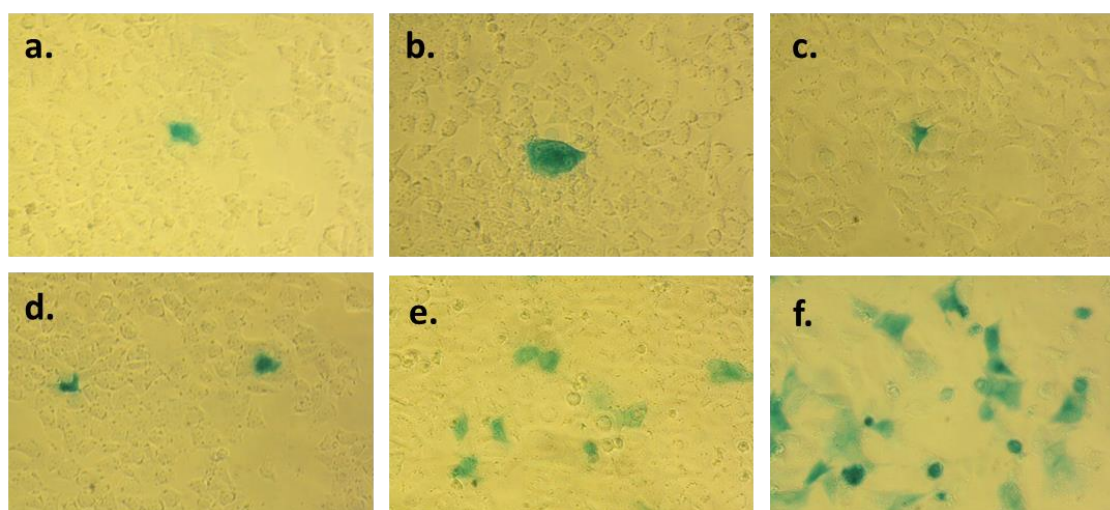


**Figure 21.** Infection assay of the amber-codon-free pNL4-3 variants.

A.



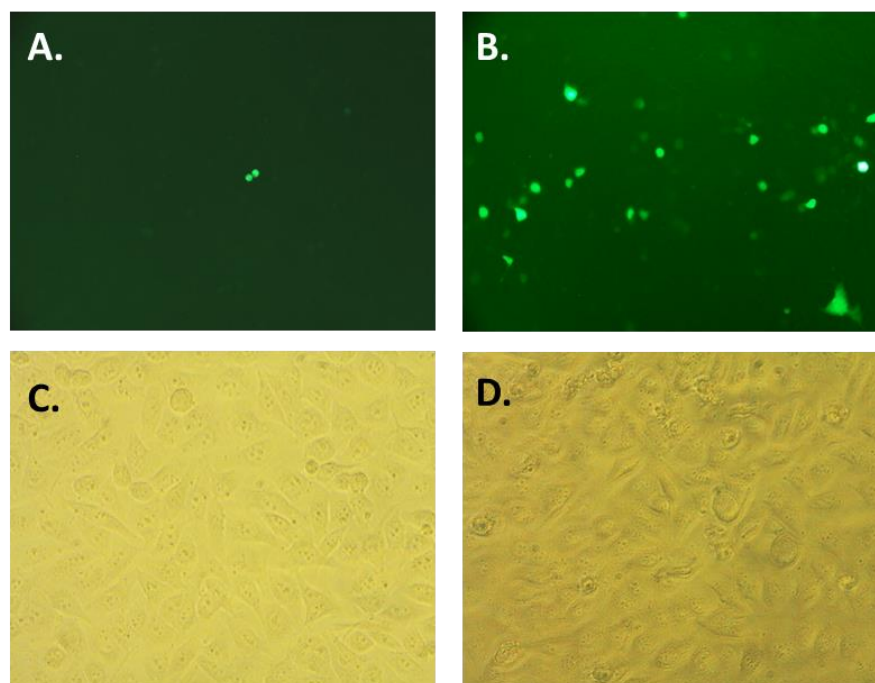
B.



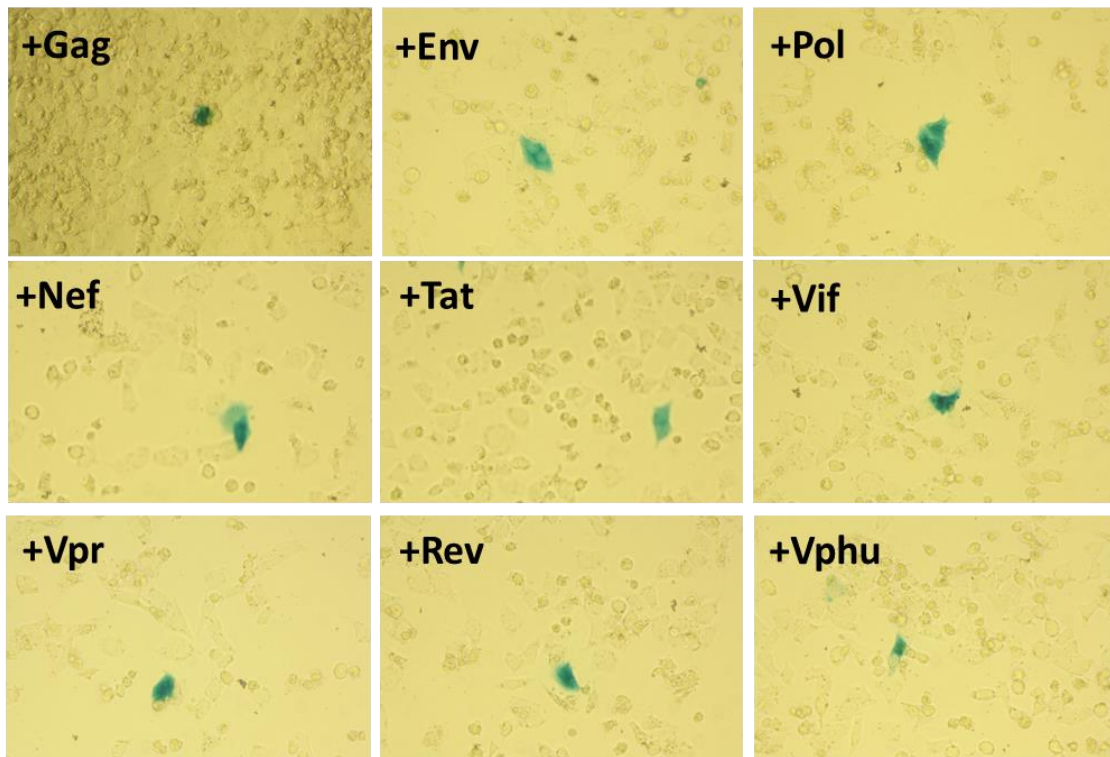
**Figure 22.** Relative expression levels of pAzFRS. (A) The pAzFRS mRNA level. (B) Infection of TZM-bl cells with pNL4-3 variants containing CMV promoter. (a) pNL43-Trp36-CMV-AzFRS-tRNA-2, with 1 mM pAzF; (b) pNL43-Trp36-AzFRS-tRNA-2, with 1 mM pAzF; (c) pNL43-Tyr59-CMV-AzFRS-tRNA-2, with 1 mM pAzF; (d) pNL43-Tyr59-AzFRS-tRNA-2, with 1 mM pAzF; (e) pNL43-Trp36 + pAzFRS plasmid, with 1 mM pAzF; (f) pNL43-Tyr59 + pAzFRS plasmid, with 1 mM pAzF. The pNL43-all-UAA variant was used as the parental plasmid for all above pNL43 constructs. The pAzFRS plasmid encodes pAzFRS/tRNA<sup>Tyr</sup> pair. Reproduced with



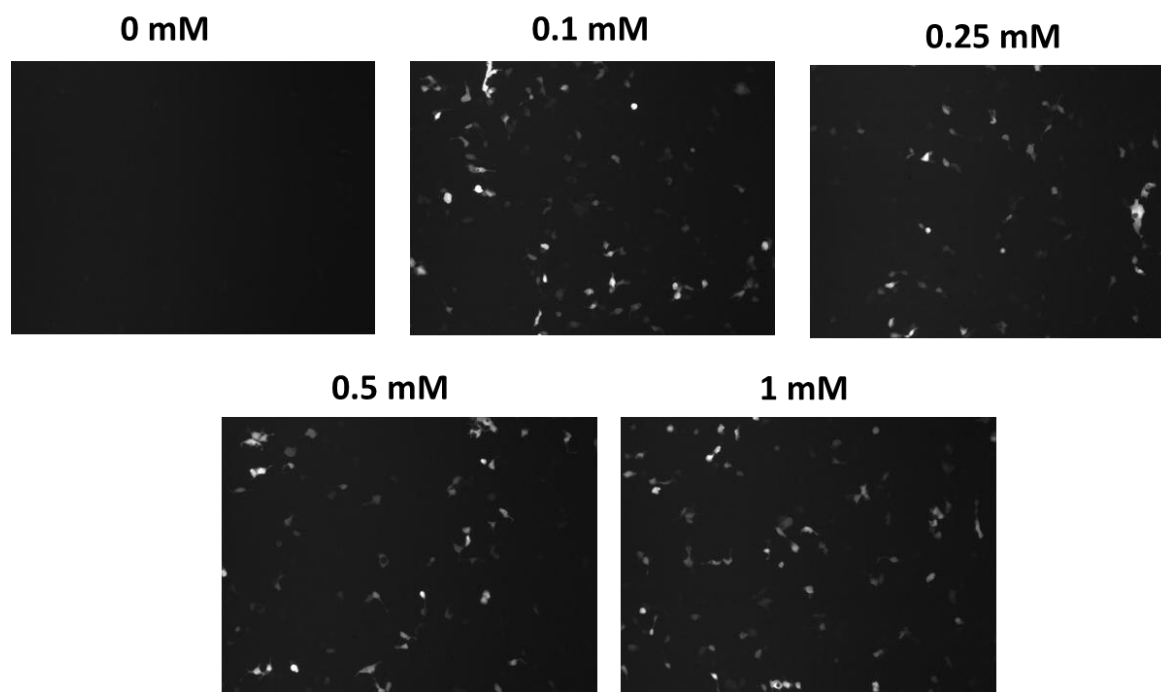
permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”



**Figure 23.** Expression and infection of pNL43 variants with two genomic copies of tRNA<sup>Tyr</sup>. (A) Fluorescence image of pNL43-AzFRS-2tRNA-F + pEGFP, with 1 mM pAzF; (B) Fluorescence image of pNL43-AzFRS-2tRNA-R + pEGFP, with 1 mM pAzF; (C) TZM-bl infection with viruses collected from (A); (D) TZM-bl infection with viruses collected from (B).



**Figure 24.** Rescue of the pNL43-Tyr59-AzFRS-tRNA-2 infectivity by viral protein supplement.



**Figure 25.** Suppression of an amber codon in EGFP in the presence of different concentration of pAzF. 293T cells were co-transfected with pAzFRS and pEGFP plasmids.

Reproduced with permission from “ACS Synthetic Biology, in press. Unpublished work copyright 2017 American Chemical Society.”



## CHAPTER 3

**Systematic Evolution and Study of Quadruplet Codon Decoding tRNAs**

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”

**3.1 Introduction: Quadruplet codon decoding tRNAs for genetic code expansion**

Genetic code is the language that translates genetic information into proteins, which is essentially universal in all organisms. The canonical genetic code consists of 64 triplet codons that are used to encode the twenty natural amino acids and terminate polypeptide synthesis. In recent years, amber stop codon reassignment has enabled site-specific incorporation of unAAs into proteins in a variety of organisms.<sup>5,6,52</sup> To date, more than 200 structurally and functionally distinctive unAAs have been genetically encoded. The ability to simultaneously incorporate different unAAs into target proteins at defined sites will vastly expand our tools for investigating protein structure and function, and for installing novel functions into biomolecules. However, the number of available nonsense triplet codons has severely hindered our ability to introduce different unAAs at multiple sites. Thus, more “blank” codons that have not been used to encode proteinogenic amino acids are needed.

In theory, a quadruplet codon table would provide a total of  $4^4=256$  “blank” codons, which could be used to further expand the genetic code. Quadruplet codons (known as +1 frameshift) have been shown to be decoded by naturally occurring tRNA mutants and to incorporate natural amino acids.<sup>146-154</sup> These tRNA mutants usually contain an extra nucleotide in the anticodon loop (8-bases instead of 7-bases). Following this rationale, unAA mutagenesis has been achieved using quadruplet codons and chemically aminoacylated tRNA bearing an expanded anticodon loop in a cell-free translation system.<sup>155-158</sup> Simultaneous incorporation of two distinctive unAAs has been realized in *E. coli* by decoding a quadruplet codon and an amber stop codon with two mutually

orthogonal aaRS/tRNA pairs.<sup>33,67</sup> Nevertheless, genetically encoding multiple unAAs is still at an early stage of development and requires further improvement with respect to decoding efficiency and the availability of novel quadruplet anticodon-codon pairs.

Recently, we<sup>54</sup> and others<sup>30</sup> reported a general approach to improve the quadruplet codon decoding efficiency with AGGA, UAGA, CUAG, and AGUA codons through tRNA engineering. Following the same strategy, we systematically evolved tRNAs for UAGN (N = A, G, U, or C) codon decoding in a genomically recoded *E. coli* strain C321.ΔA (with all 321 UAG codons changed to UAA and an RF1 knock-out).<sup>66</sup> We first constructed a tRNA library, in which each tRNA mutant harbored an expanded anticodon stem loop (ASL) with saturated mutations at positions 29, 30, 31, 32, 33, 33.5 and positions 38, 39, 40, 41. The library was then subjected to antibiotic-based positive selection to isolate functional library members. The selected library members were sequenced and validated through fluorescence-based assays. We also characterized the efficiency and fidelity of quadruplet codon decoding of the evolved tRNAs. Our efforts on engineering UAGN decoding tRNAs may help both uncover the mechanism of quadruplet codon decoding and broaden the application of genetic code expansion with quadruplet codons.

## 3.2 Approaches

### 3.2.1 tRNA (with an expanded anticodon loop) library construction and selection for UAGN (N = A, G, U, C) codon decoding.

We chose tRNA<sub>CUA</sub><sup>Pyl</sup>, a naturally occurring amber suppressor tRNA from *Methanosarcina maize* as a model for investigating quadruplet codon decoding. Based on the crystal structure, PylRS does not directly interact with the anticodon loop of tRNA<sub>CUA</sub><sup>Pyl</sup>.<sup>55</sup> This may allow tRNA<sub>CUA</sub><sup>Pyl</sup> to be evolved for quadruplet codon decoding by introducing mutations at the anticodon stem loop. In order to identify tRNA<sub>NCUA</sub><sup>Pyl</sup> variants that are able to decode UAGN (N=

A, G, U or C) codons, we created a tRNA<sub>NCUA</sub><sup>Pyl</sup> (NCUA-11) library in which six nucleotides in the anticodon stem (positions 29, 30, 31, 39, 40, and 41), and five nucleotides in the anticodon loop (positions 32, 33, 33.5, 37, and 38) were completely randomized (**Figure 26A**). The actual library size was  $5.2 \times 10^6$  which covered >99% of the theoretical diversity. Multiple research groups have shown that four-base complementarity is nonessential but highly favored in quadruplet codon decoding.<sup>159-161</sup> The insertion of a randomized nucleotide at position 33.5 (between U33 and C34; the first base of anticodon) in tRNA<sub>CUA</sub><sup>Pyl</sup> may allow us to assess whether complete quadruplet codon-anticodon base-pairing is required for UAGN decoding.

To identify functional tRNA<sub>NCUA</sub><sup>Pyl</sup> variants that can suppress UAGN codons, we constructed positive selection reporters (**Figure 26B**) that contain an in-frame quadruplet codon of interest in the chloramphenicol acetyltransferase gene (pREP-BocKRS-UAGN). The selection was conducted by growing cells that contained the tRNA<sub>NCUA</sub><sup>Pyl</sup> library, BocKRS (a PylRS derivative that charges N<sup>ε</sup>-(*tert*-butyloxy-carbonyl)-L-lysine (BocK) onto tRNA<sub>CUA</sub><sup>Pyl</sup>), and chloramphenicol acetyltransferase reporter gene in the presence of varying concentration of chloramphenicol (from 34 to 75 μg/mL) and 5 mM BocK. The surviving cells were replicated on LB agar plates containing a range of different concentrations of chloramphenicol (50, 75, and 100 μg mL<sup>-1</sup>) in the presence or absence of BocK. A clone that grew at 100 μg mL<sup>-1</sup> chloramphenicol in the presence of BocK but did not grow at 50 μg mL<sup>-1</sup> chloramphenicol in the absence of BocK were considered a hit. Multiple hits were selected for UAGA, UAGU, and UAGG codons. We were not able to obtain a hit for UAGC decoding after several attempts. From each selection, three hits that conferred the highest level of chloramphenicol resistance were further characterized. The selected tRNA<sub>NCUA</sub><sup>Pyl</sup> hits were named UAGN-X, where N represents A, G, or U; and X represents the hit number.

### 3.2.2 Characterization and validation of evolved tRNAs.

We examined the efficiency and fidelity of quadruplet codon decoding tRNAs using a fluorescence-based assay. The four targeting quadruplet codons, UAGA, UAGG, UAGU, and UAGC, were inserted at a permissive site in GFPuv encoding gene to yield GFPuv-Asn149UAGA, GFPuv-Asn149UAGG, GFPuv-Asn149UAGU, and GFPuv-Asn149UAGC reporters. The encoding gene for each tRNA<sub>NCUA</sub><sup>Pyl</sup> mutant was introduced into a corresponding GFPuv-Asn149UAGN reporter to generate a pGFPuv-NCUA-X plasmid (N=A, G, U, or C; X is hit number). If a tRNA can decode a quadruplet codon of interest, a full-length GFPuv is produced. Therefore, the intensity of GFPuv fluorescence can be directly correlated to the quadruplet codon decoding efficiency of a tRNA mutant. As a control, the encoding gene for tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt (N=A, G, U, or C), in which the triplet CUA anticodon was replaced with a quadruplet anticodon of interest, was also introduced into the reporters to generate pGFPuv-NCUA-wt plasmids.

pGFPuv-NCUA-X and pGFPuv-NCUA-wt plasmids were each transformed into C321.ΔA cells that contained pBK-BocKRS plasmid. The cells were first grown to saturation at 30 °C and then subcultured in fresh LB medium containing 0.1 mM IPTG with or without 5 mM BocK. As shown in **Figure 27**, a significant enhancement in quadruplet codon decoding efficiency was observed with all mutants when BocK was supplemented. The best three hits, UAGA-1, UAGU-1 and UAGG-2 displayed approximately 55, 9, and 40 fold-improvement in quadruplet codon decoding in comparison to the corresponding tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt. The suppression efficiency of UAGA and UAGG codons was much higher than that of the UAGU codon. In the absence of BocK, the fluorescence of these mutants was negligible, indicating that tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants were selectively charged with BocK by BocKRS but not with any natural amino acids by BocKRS or endogenous aaRSs. Although we did not obtain a hit for the UAGC codon, we tested the efficiency

of UAGC codon decoding by tRNA<sub>GCUA</sub><sup>Pyl</sup>-wt and a few tRNA hits. No detectable decoding was observed.

GFPuv mutants from the above experiments were purified with Ni Sepharose Fast 6 resins (**Figure 28**). Consistent with the observation in the fluorescence assays, suppression of the UAGA codon by tRNA hit UAGA-1 resulted in the highest protein yield (~54 mg/L). Suppression of the UAGG codon by tRNA hit UAGG-2 also led to a large protein yield (~21 mg/L). The lowest yield (~9 mg/L) was obtained with tRNA hit UAGU-1 for UAGU decoding. The GFPuv-Asn149BocK samples were then subjected to in-gel digestion for mass spectrometry analysis. At position 149 of GFPuv, only BocK or Lys (after cleavage of Boc group from BocK) was detected (**Figure 29**), suggesting high fidelity of the evolved tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants for quadruplet codon decoding.

### 3.2.3 Incorporation of other unAAs using the evolved tRNAs.

Since the ASL of tRNA<sub>CUA</sub><sup>Pyl</sup> does not directly interact with PylRS,<sup>162,163</sup> the evolved tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants can be potentially recognized by other PylRS derivatives with different amino acid specificity. To demonstrate that tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants can be used to incorporate a range of unAAs in response to UAGN (N= A, G, or U) codons, we randomly chose two unAAs, 3'-azibutyl-N-carbamoyl-lysine (AbK)<sup>164</sup> and *o*-nitrobenzyl-oxycarbonyl-N<sup>ε</sup>-L-lysine (ONBK)<sup>125</sup> for the study. Cells containing UAGA-1, UAGU-1 and UAGG-2 tRNA mutants were grown in the presence or absence of 1 mM AbK (**Figure 30**) and 1 mM ONBK (**Figure 31**), respectively. All of the selected hits displayed significant enhanced quadruplet codon decoding efficiency with both unAAs compared to tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt. The relative decoding efficiency at UAGA, UAGG, and UAGU codons by BocK, AbK, and ONBK remained consistent, though in general, the incorporation of BocK and AbK was more efficient than ONBK. Notably, the UAGU-1 mutant led to a significant increase in UAGU codon decoding efficiency when pairing with AbKRS and

without further changes in the nucleotide sequence of tRNA, which suggested that the nature of aaRS also affected quadruplet codon decoding efficiency.

### 3.2.4 Analysis of the evolved tRNA mutants.

Due to potential post-transcriptional modifications and stable secondary structures, it is hard to directly sequence tRNA transcripts through standard methods. Therefore, we isolated and sequenced the plasmid DNA that encode evolved tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants. The sequencing results for these tRNA<sub>NCUA</sub><sup>Pyl</sup> hits are shown in **Table 3**. Surprisingly, a 5'-UCUA-3' anticodon was conserved in all the hits. In addition, tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants sharing the same sequences were identified from different selections, such as UAGA-2 and UAGU-3, UAGG-1 and UAGU-1, UAGG-3 and UAGU-2. In fact, all UAGU variants were also found in selections against UAGA and UAGG. U33 and A37 seemed to be essential for UAGN decoding since no mutations were found at these two positions. All the evolved tRNA variants contained A31G and U39C mutations. The replacement of the A31-U39 pair with the G31-C39 pair likely strengthened the anticodon stem region that is directly connected with the anticodon loop. Most of these mutants did not have Watson-Crick base-pairing at position 29-41 or 30-40. The C40U mutation led to a weaker base-pairing interaction at position 30-40. These mutations in the ASL may result in a favorable conformational change, which allows tRNA variants to be correctly positioned in the ribosome A site and translocated during peptide elongation.

### 3.2.5 Cross-decoding among UAGN codons.

Since the three best hits, UAGA-1, UAGU-1, and UAGG-2 all contained 5'-UCUA-3' anticodon, we cross-tested the decoding efficiency of the quadruplet codons that were not selected. New fluorescence reporter plasmids pGFPuv-Asn149UAGN-BocKRS were constructed to simplify the experimental design. The evolved tRNA<sub>NCUA</sub><sup>Pyl</sup> variants were transformed into cells

containing the corresponding fluorescence reporters. All the tRNA variants showed some level of cross-activity against other codons. The UAGG-2 mutant displayed the highest cross activity (**Figure 32**) with the UAGA codon. It showed approximately 87% suppression efficiency relative to the UAGA-1 mutant. This mutant showed approximately 122% activity relative to the UAGU-1 mutant with the UAGU codon. The UAGA-1 mutant was relatively more selective for the UAGA codon. Interestingly, UAGU-1 displayed a much higher suppression efficiency towards the UAGA and UAGG codons than to its cognate UAGU codon.

### **3.2.6 tRNA (with regular 7-base anticodon loop) library construction and selection for UAGN (N = A, G, U, C) decoding.**

Since UAGN decoding does not require a perfect complementarity at the fourth base of the anticodon-codon, we sought to investigate whether tRNAs with regular 7-base anticodon loops were capable of decoding UAGN codons. We constructed a tRNA<sub>CUA</sub><sup>Pyl</sup> library (CUA-10) in which six nucleotides in the anticodon stem (positions 29-31, 39-41) and four nucleotides in the anticodon loop (positions 32,33,37,38) were completely randomized (**Figure 26A**). The actual library size was  $8.7 \times 10^6$ , which covered >99% theoretical diversity of the library. The resulting library was subsequently subjected to positive selections against UAGA, UAGG, UAGU, and UAGC codons, respectively, using pREP-BocKRS-UAGN reporters. We did not identify any hits from the selections. These results suggested that an expanded anticodon loop was essential for UAGN decoding.

### **3.2.7 Cross-decoding against UAG codons.**

Since the last three bases of 5'-NCUA-3' anticodon of tRNA<sub>NCUA</sub><sup>Pyl</sup> variants can form Watson-Crick base-pairing with the amber UAG codon, they might be able to decode the amber codon. To examine this possibility, a fluorescence reporter pGFPuv-Asn149UAG-BocKRS was

constructed. The amber suppression efficiency of tRNA variants, UAGA-1, UAGU-1, and UAGG-2 were examined by the expression of full-length GFPuv in the presence of 5 mM BocK. As shown in **Figure 33**, UAGA-1 and UAGG-2 mutants displayed a good selectivity towards UAGA and UAGG codons. The UAGU-1 mutant showed a similar level of suppression efficiency towards both UAGU and UAG codons. This may result from low decoding efficiency of UAGU-1 towards the UAGU codon. All three mutants showed a significant lower amber suppression efficiency than tRNA<sub>CUA</sub><sup>Pyl</sup> (decreased by approximately 40-60 folds).

### 3.2.8 Quantification of UAGN decoding efficiency with fluorescence reporters

To quantitatively analyze the decoding efficiency of the hits against UAGN (N= A, G, U, or C) codons versus UAG codon, we generated two series of fluorescence reporters, pStrepII-GFPuv-UAGN and pStrepII-GST-X-GFPuv. The pStrepII-GFPuv-UGAN reporters (**Figure 34A**) contained one copy of GFPuv gene with a quadruplet codon of interest at position 149 and two additional mutations GCA<sub>154</sub>GCT, and GAC<sub>155</sub>GAT. As shown in **Figure 34A**, only if UAGN is decoded as a quadruplet codon, can a full-length GFPuv be expressed. On the other hand, if UAGN is translated as an ordinary three-base, then a premature termination at position 155 will result in a truncated form of GFPuv, which is not fluorescent. Since a StrepII-tag was fused to the N-terminus of GFPuv, the expression levels of both the full-length and truncated protein can be simultaneously quantified by western blot (**Figure 34B**). We observed the formation of both full length and truncated GFPuv in *E. coli* GeneHogs, but not in C321. ΔA (Data not shown).

The pStrepII-GST-X-GFPuv reporters contained a short linker G\*ADI (\* represents tyrosine encoded by UAG, or BocK encoded by UAGA) between GST and GFPuv protein (**Figure 35A**). Therefore, when UAGA is correctly decoded, a GST-GFPuv fusion protein (51 kDa) is produced. If the first three bases of UAGA are decoded as a triplet codon, then a GST with three additional



amino acids at the C-terminus (25 kDa) is generated. We first titrated the GST-GFPuv fusion protein (produced by pStrepII-GST-UAC-GFPuv) and GST protein (produced by pStrepII-UAG-GFPuv) using western blot (**Figure 35B**). Surprisingly, the signals detected from GST-GFPuv were much weaker than from the same amount of GST. One plausible explanation is a less efficient transfer of GST-GFPuv than GST from the SDS-PAGE gel to the nitrocellulose due to its larger size. Next, we compared the decoding efficiency of tRNAs UAGA-1 and UAGA-wt towards the UAGA codon in cell lysates. The expression of GST-GFPuv was almost non-detectable in both samples (**Figure 35C**). Strikingly, although the positive control (pStrepII-GST-UAG-GFPuv with tRNA<sub>CUA</sub><sup>Pyl</sup>) showed a strong fluorescence, the expression of GST-GFPuv was not detected by western-blot. Since this method led to a discrepancy in signal outputs for two target proteins, new reporters based on different tags or reporting genes are needed.

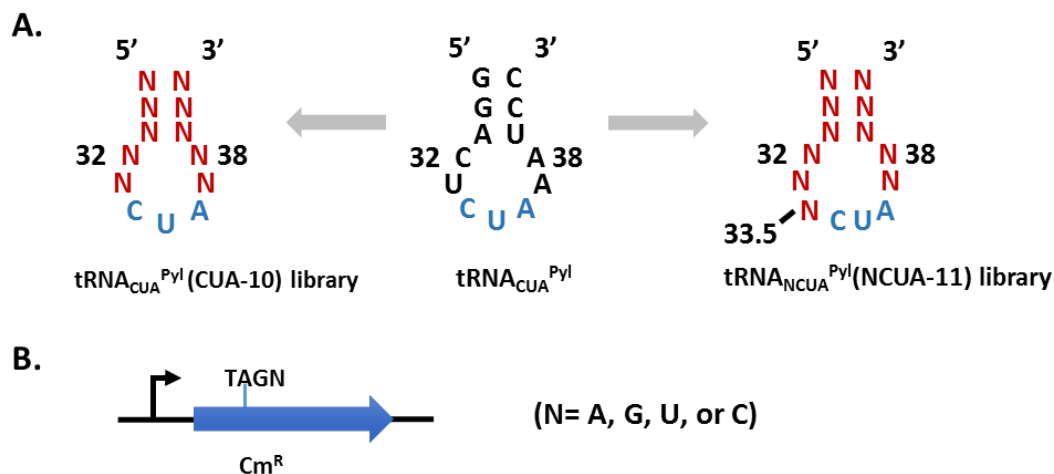
### 3.3 Summary and remarks

In this chapter, we presented the first example of systematic evolution of tRNAs for UAGN codon (N= A, G, U, or C) decoding in a genetically recoded *E. coli* strain C321.ΔA. By fully randomizing selected nucleotides in the ASL of the parental tRNA, we successfully engineered several tRNA variants that displayed a significantly higher efficiency in UAGN decoding in comparison to their corresponding wild-type controls. The best hit (UAGA-1) led to a 55-fold improvement in UAGA decoding efficiency in comparison to tRNA<sub>UCUA</sub><sup>Pyl</sup>-wt. This hit also showed a significant decoding activity towards the UAGA codon in *E. coli* GeneHogs, indicating that this tRNA can overcome the competition from RF1 at the ribosome A site. In addition, some of the hits contained convergent sequences and displayed cross-activity against the quadruplet codons that they were not selected against.

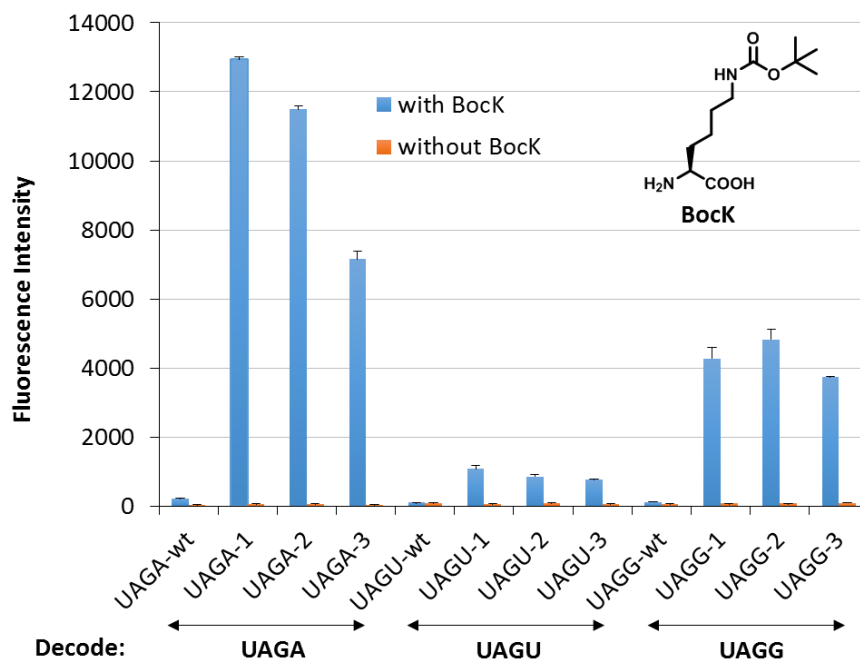
Our data suggests that an expanded anticodon loop (8-base) is required for efficient UAGN decoding. However, a complete four-base pair is not essential for UAGN decoding. This result cannot be easily explained using classical models, the yardstick model or the slippery model (**Figure 36**)<sup>165,166</sup> for quadruplet codon decoding (+1 frameshift). The yardstick model states that the anticodon forms a complete four-base interaction with the quadruplet codon at the ribosome A site, and a subsequent quadruplet translocation from the A site to the P site causes the +1 frameshift. However, a few evolved tRNA<sub>NCUA</sub><sup>Pyl</sup> variants apparently cannot form a complete quadruplet codon-anticodon interaction at the A site. Alternatively, the slippery model proposes a triplet anticodon-codon interaction at the A site and a triplet translocation from the A site to the P site followed by slippage of the mRNA by one base and re-pairing between the anticodon and the mRNA. In our UAGN study, the re-pairing event is unlikely to happen between UAGN codons (N=A, G, or U) and 5'-UCUA-3' anticodon in tRNA<sub>NCUA</sub><sup>Pyl</sup> variants due to the lack of Watson-Crick base-pairing after the re-pairing event. Therefore, we proposed a modified slippery model for UAGN decoding (**Figure 37**). It is likely that the extra nucleotide U33.5 in tRNA<sub>NCUA</sub><sup>Pyl</sup> interacts with the fourth base of UAGN sequence in the P site which induces steric hindrance to the incoming tRNA in the A site and facilitates the +1 frameshift.

In summary, we evolved a set of tRNAs that are capable of incorporating different unAAs in response to UAGN codons. We also confirmed the essential role of an expanded anticodon loop for efficient UAGN decoding. In addition, we demonstrated that four-base complementarity was not required for UAGN decoding. This work will facilitate the mechanistic study of quadruplet codon decoding tRNAs bearing expanded anticodon loops.

## 3.4 Figures and tables

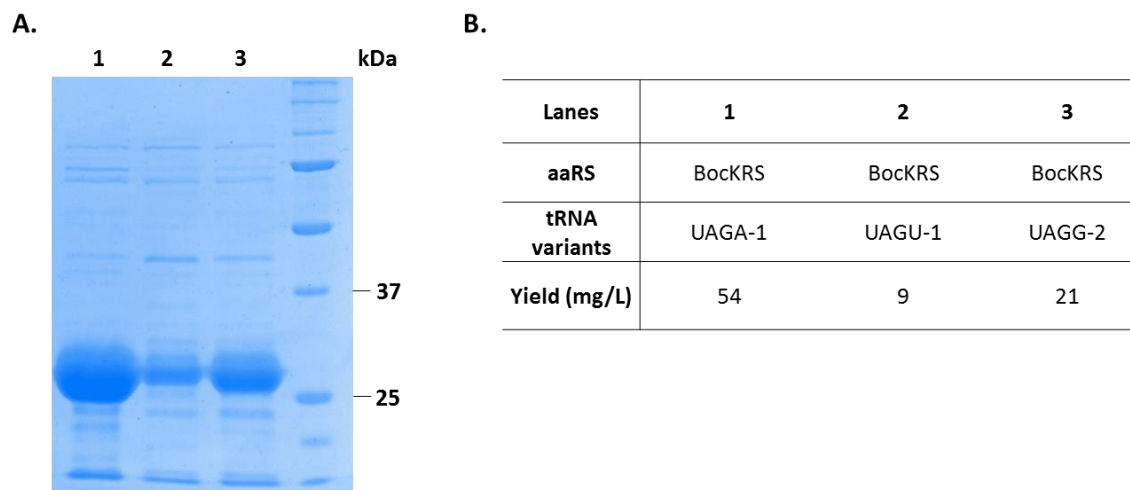


**Figure 26.** Library construction and selection for UAGN-decoding tRNA<sub>NCUA</sub><sup>Pyl</sup> variants. (A) Construction of tRNA<sub>CUA</sub><sup>Pyl</sup> (CUA-10) library and tRNA<sub>NCUA</sub><sup>Pyl</sup> (NCUA-11) library (N = A, G, U, or C). Mutation sites are highlighted in red and the three-base anticodon of wild-type tRNA<sup>Pyl</sup> is in blue. (B) The chloramphenicol acetyltransferase reporter with Gln98UAGN (N = A, G, U, or C) mutations. Only cells containing tRNA<sub>NCUA</sub><sup>Pyl</sup> variants that can efficiently decode UAGN codons can survive in the presence of high concentration of chloramphenicol.



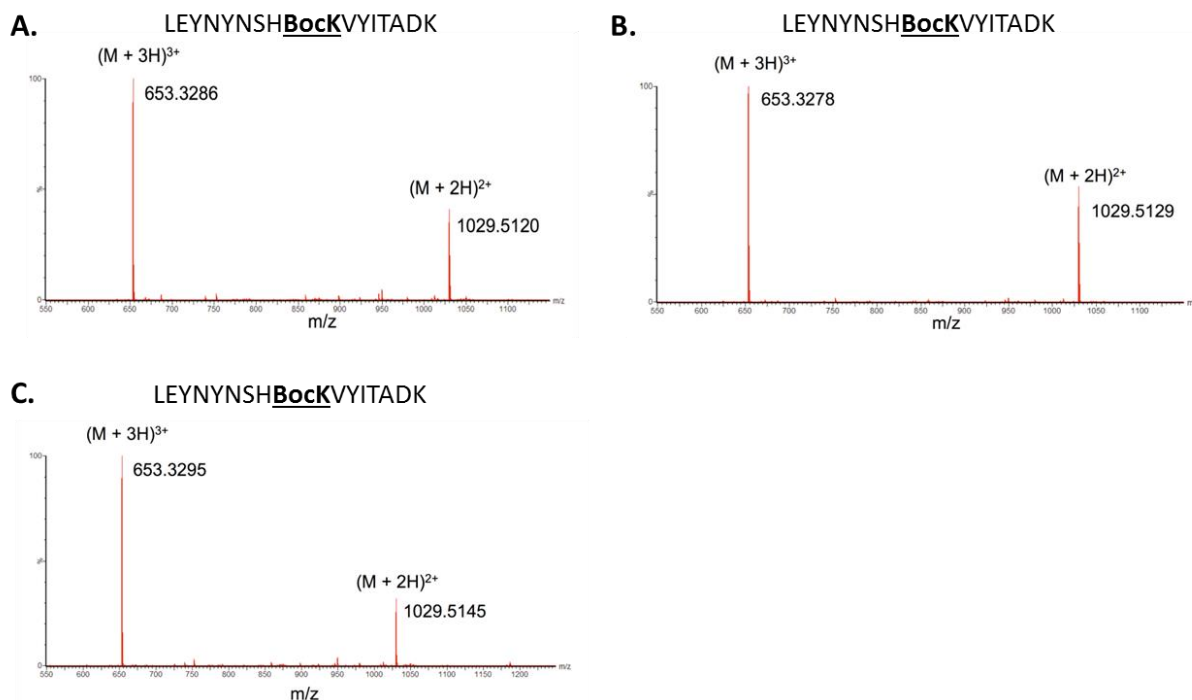
**Figure 27.** GFPuv expression in cells containing  $tRNA_{NCUA}^{Pyl}$  variants. GFPuv expression was conducted in cells containing  $tRNA_{NCUA}^{Pyl}$  mutants or  $tRNA_{NCUA}^{Pyl-wt}$ , BocKRS, and corresponding GFPuv-Asn149UAGN reporter in the presence or absence of 5 mM BocK. Fluorescence intensity was normalized to cell density. The data shown was an average of triplicate measurements.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



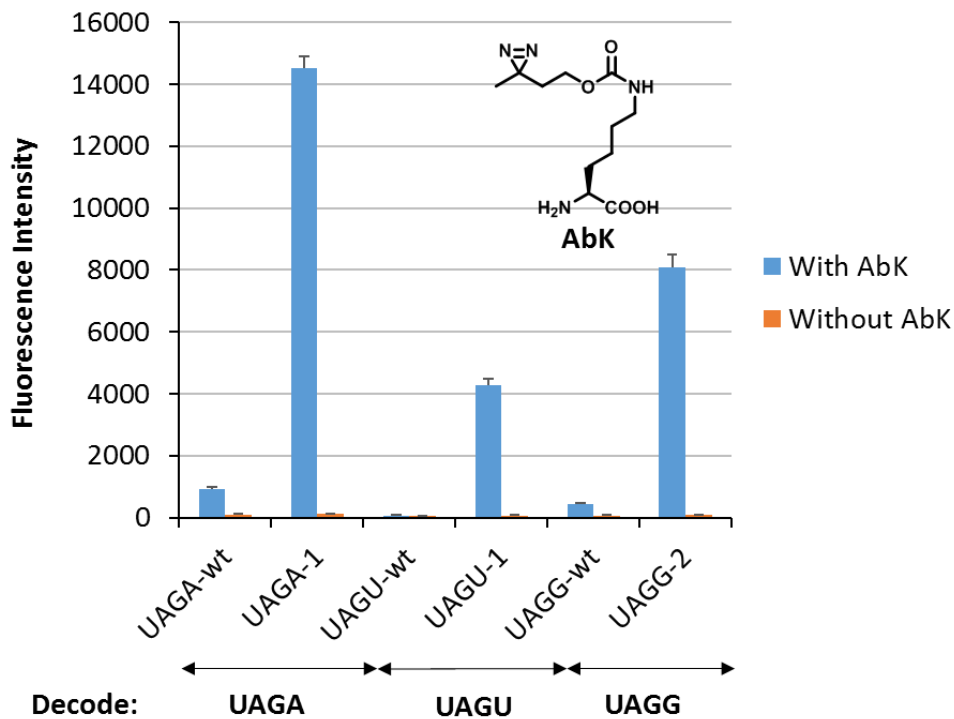
**Figure 28.** Purification of GFPuv-Asn149BocK mutants. (A) Lane 1, GFPuv mutant expressed in the presence of GFPuv-Asn149UAGA and tRNA UAGA-1, with 5 mM BocK; Lane 2, GFPuv mutant expressed in the presence of GFPuv-Asn149UAGU and tRNA UAGU-1, with 5 mM BocK; Lane 3, GFPuv mutant expressed in the presence of GFPuv-Asn149UAGG and tRNA UAGG-2, with 5 mM BocK. (B) Protein yields in (A).

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



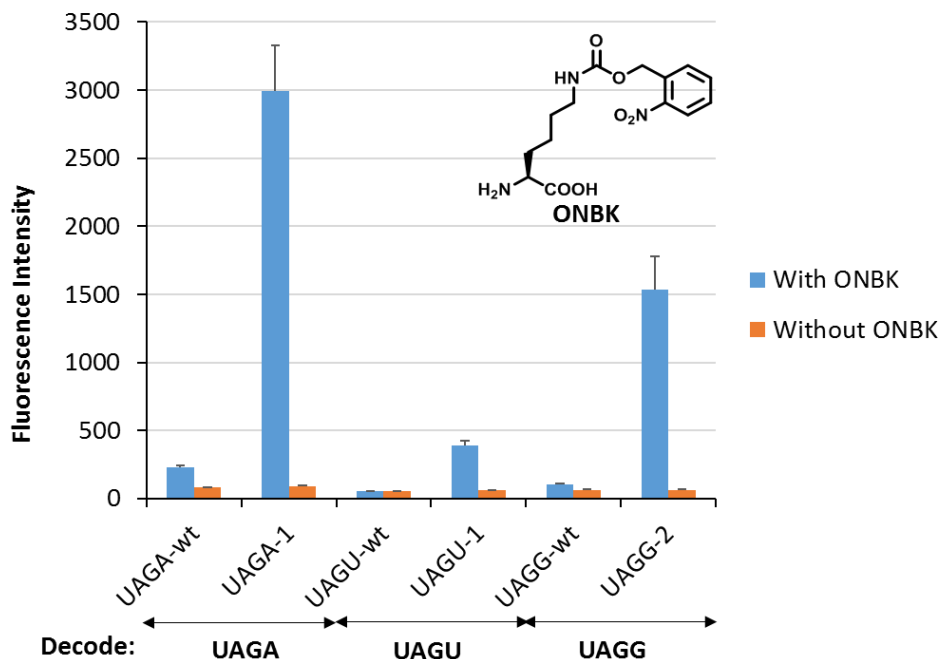
**Figure 29.** Mass spectrometry analysis of GFPuv containing BocK at position 149. (A) Suppress UAGU codon; (B) Suppress UAGG codon; (C) Suppress UAGA codon. The amino acid sequence of the peptide fragment, LEYNYNSHBocKVYITADK, from mutant GFPuv containing Boc-Lys is shown on top. Two mass speaks were observed. The  $(M + 2H)^{2+}$  peak corresponds to LEYNYNSHBocKVYITADK that contains an intact BocK residue at position 149. The  $(M + 3H)^{3+}$  peak corresponds to LEYNYNSHKVYITADK that contains a lysine residue at position 149 due to the loss of Boc group under the mass spectrometry conditions. (Note: The BocKRS cannot charge the tRNA with lysine. The observed peptide that contains a lysine at position 149 must be derived from the cleavage of the Boc group).

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



**Figure 30.** UAGN codons decoding with AbK. GFPuv expression was conducted in cells containing  $tRNA_{NCUA}^{Pyl}$  mutants or  $tRNA_{NCUA}^{Pyl-wt}$ , AbKRS, and corresponding GFPuv-Asn149UAGN reporter in the presence or absence of 1 mM AbK. Fluorescence intensity was normalized to cell density. The data shown was an average of triplicate measurements.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



**Figure 31.** UAGN codons decoding with ONBK. GFPuv expression was conducted in cells containing tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants or tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt, ONBK-RS, and corresponding GFPuv-Asn149UAGN reporter in the presence or absence of 1 mM ONBK. Fluorescence intensity was normalized to cell density. The data shown was an average of triplicate measurements.

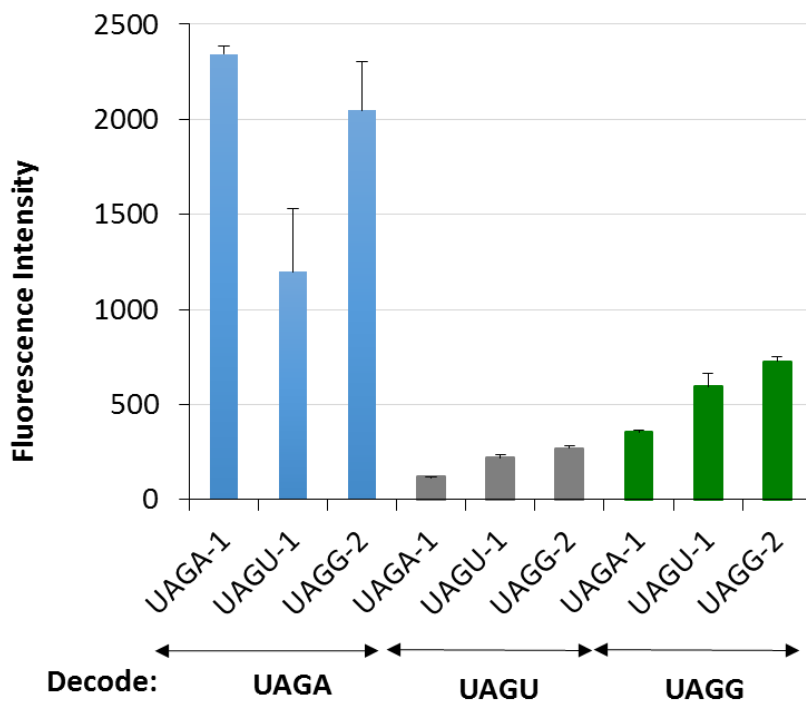
Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



Codon	tRNA variants	Positions				
		29-31	32,33	33.5	37, 38	39-41
UAGA	UAGA- wt	GGA	CU	U	AA	UCC
	UAGA-1	GGG	CU	U	AU	CCU
	UAGA-2	UGG	AU	U	AC	CUU
	UAGA-3	CGG	AU	U	AC	CUU
UAGU	UAGU- wt	GGA	CU	A	AA	UCC
	UAGU-1	UGG	CU	U	AU	CUU
	UAGU-2	AGG	CU	U	AU	CUU
	UAGU-3	UGG	AU	U	AC	CUU
UAGG	UAGG- wt	GGA	CU	C	AA	UCC
	UAGG-1	UGG	CU	U	AU	CUU
	UAGG-2	GGG	CU	U	AU	CUU
	UAGG-3	AGG	CU	U	AU	CUU

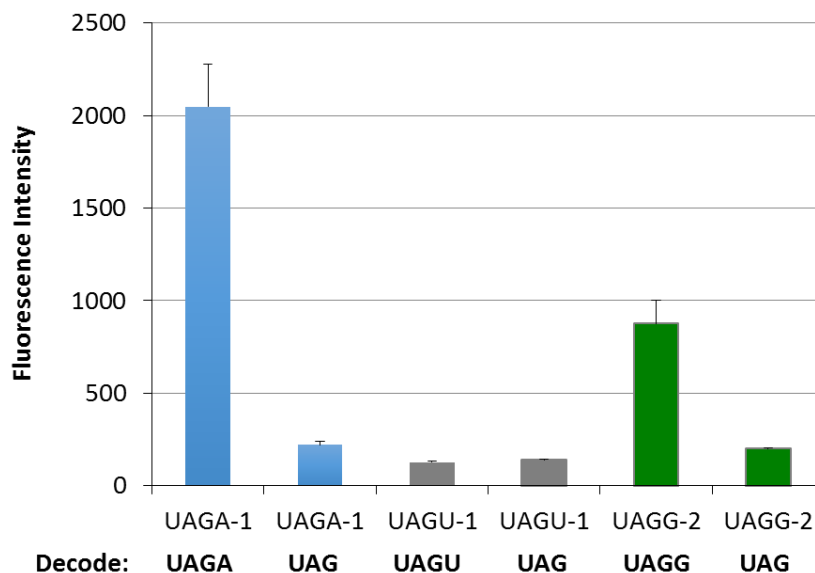
**Table 3.** Sequences of tRNA<sub>NCUA</sub><sup>Pyl</sup> variants that have improved UAGN decoding efficiency.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



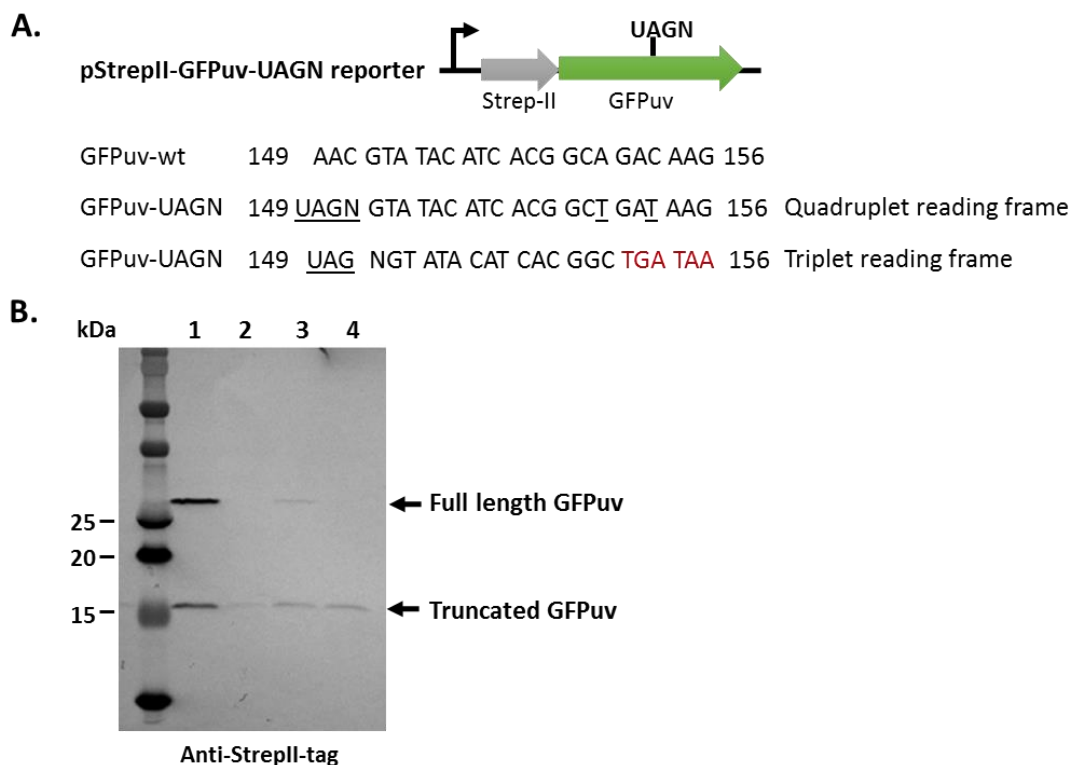
**Figure 32.** Cross-activity of tRNA<sub>NCUA</sub><sup>Pyl</sup> variants among UAGN codons. GFPuv expression was conducted in cells containing tRNA<sub>NCUA</sub><sup>Pyl</sup> mutants, BockRS, and corresponding GFPuv-Asn149UAGN-BockRS reporters in the presence of 5 mM Bock. Fluorescence intensity was normalized to cell density. The data shown was an average of triplicate measurements.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”



**Figure 33.** UAGN and UAG codon decoding efficiency of  $tRNA_{NCUA}^{Pyl}$  variants. GFPuv expression was conducted in cells contained  $tRNA_{NCUA}^{Pyl}$  mutants, BocKRS, and pGFPuv-Asn149UAG or pGFPuv-Asn149UAGN reporters in the presence of 5 mM BocK. Fluorescence intensity was normalized to cell density. The data shown was an average of triplicate measurements.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”

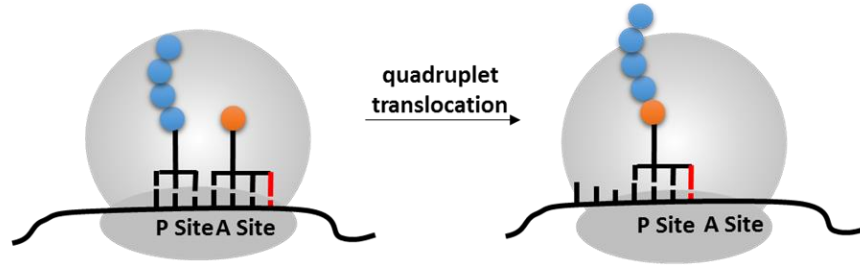


**Figure 34.** Quantification of UAGN decoding efficiency with pStrepII-GFPuv-UAGN reporters.

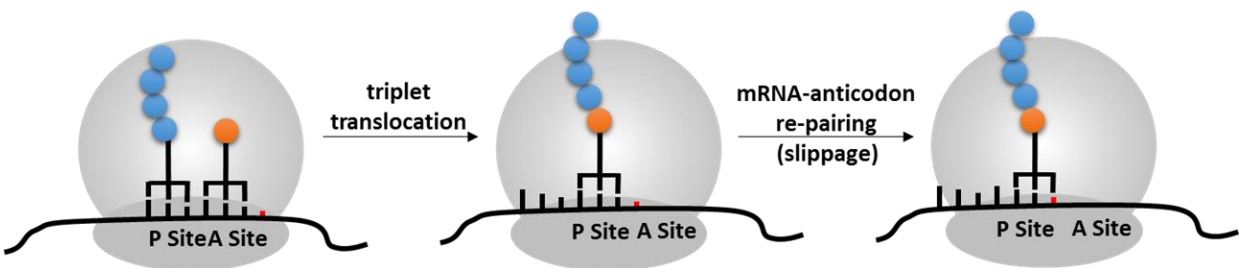
(A) Partial DNA sequence of GFPuv-wt and GFPuv-UAGN. Quadruplet codon decoding at position 149 leads to a full length GFPuv (27 kDa). Triplet codon decoding at position 149 leads to a truncated GFPuv (17 kDa). (B) Western blot of cell lysates containing pStrepII-GFPuv-UAGN (N=A, G, U, or C) reporters and tRNA variant UAGA-1. Lane 1, pStrepII-GFPuv-UAGA + tRNA UAGA-1, with 5 mM BocK; Lane 2, pStrepII-GFPuv-UAGU + tRNA UAGA-1, with 5 mM BocK; Lane 3, pStrepII-GFPuv-UAGG + tRNA UAGA-1, with 5 mM BocK; Lane 4, pStrepII-GFPuv-UAGC + tRNA UAGA-1, with 5 mM BocK. GFPuv expression was conducted in *E. coli* strain GeneHogs.



**A. Yardstick model**

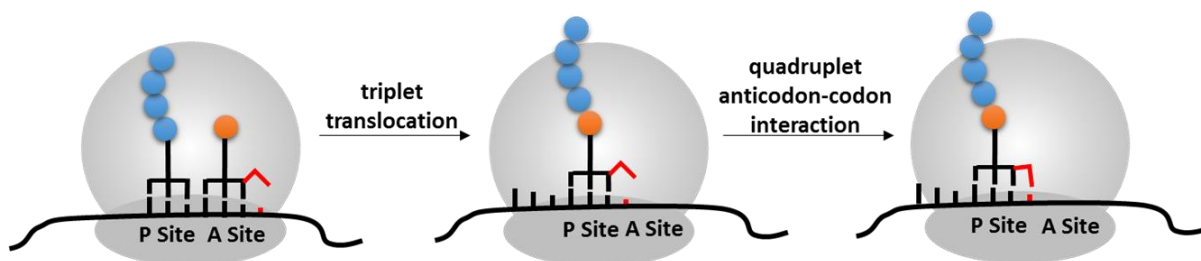


**B. Slippery model**



**Figure 36.** Two classical models for quadruplet codon decoding. (A) Yardstick model; (B) Slippery model.

### Modified slippery model



**Figure 37.** A new model for quadruplet decoding (+1 frameshift). This model features a triplet translocation followed by a quadruplet anticodon-codon interaction in the P site of ribosome.

Reproduced with permission from “Scientific Report. 2016, 6, a21898.”

## CHAPTER 4

## Development of A General Approach to Evolve Aminoacyl-tRNA synthetase for Efficient Nonsense Codon Suppression

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”

### 4.1 Introduction: Evolution of aminoacyl-tRNA synthetase for more efficient unAA incorporation

*Methanococcus jannaschii*-derived tyrosyl-tRNA synthetase (MjTyrRS) and tyrosyl-tRNA (MjtRNA<sup>Tyr</sup>) pair has been extensively used for the genetic incorporation of aromatic unAAs in *E. coli*.<sup>6</sup> To create a highly efficient and specific MjTyrRS/MjtRNA<sup>Tyr</sup> pair for unAA incorporation, several requirements need to be fulfilled: (1) an accurate and efficient aminoacylation of MjtRNA<sup>Tyr</sup> with a unique unAA by its cognate MjTyrRS; (2) no cross activity between the unAA-specific MjTyrRS/MjtRNA<sup>Tyr</sup> pair and endogenous aaRS/tRNA pairs; (3) favorable interaction between the MjTyrRS/MjtRNA<sup>Tyr</sup> pair and the host translational machinery; and (4) nonsense codon (e.g., amber nonsense codon or quadruplet codons) decoding capacity of the charged MjtRNA<sup>Tyr</sup>.

In previous studies, the evolution of MjTyrRS has focused on altering its specificity towards unAAs with different chemical structures and functions. The classical evolution approach takes two steps.<sup>2,4,11,44,45,167</sup> The first step is to create a MjTyrRS library in which key residues within the amino acid recognition site are fully randomized. The second step involves consecutive rounds of *in vivo* selection to identify library members with desirable properties. Generally, a toxic gene is exploited in the negative selection to remove unspecific MjTyrRS variants that recognize natural amino acids. In the positive selection, functional MjTyrRS hits that can decode nonsense codon



in an antibiotic resistance gene (or other selectable markers) are enriched. The order and number of positive and negative selections can be adjusted based on different experiment needs to obtain the best MjTyrRS variants. Following the above approach, a variety of MjTyrRS variants that possess good fidelity and high efficiency for unAA incorporation have been successfully identified.

In the meantime, MjtRNA<sup>Tyr</sup> has been engineered to allow efficient amber suppression. The first base of the anticodon of wild-type MjtRNA<sup>Tyr</sup> (MjtRNA<sub>GUA</sub><sup>Tyr</sup>), G34, was mutated to C34 to generate an amber suppressor MjtRNA<sup>Tyr</sup> variant (MjtRNA<sub>CUA</sub><sup>Tyr</sup>).<sup>1</sup> The orthogonality of MjtRNA<sub>CUA</sub><sup>Tyr</sup> was then further improved through directed evolution.<sup>3</sup> However, based on an *in vitro* study, the aminoacylation efficiency of MjtRNA<sub>CUA</sub><sup>Tyr</sup> by MjTyrRS was significantly lower than that of MjtRNA<sub>GUA</sub><sup>Tyr</sup>,<sup>168</sup> which is undesired in unAA incorporation. According to an X-ray crystallography study (**Figure 38**), G34 plays a critical role in anticodon recognitions. The base of G34 is sandwiched between Phe261 and His283, and N1 and N2 of G34 form hydrogen bonds with Asp286 in MjTyrRS.<sup>32,87,169</sup> Therefore, G34C mutation in MjtRNA<sub>CUA</sub><sup>Tyr</sup> may lead to lower aminoacylation activity due to suboptimal anticodon recognition by MjTyrRS. A more comprehensive understanding of the interaction between the anticodon of MjtRNA<sub>CUA</sub><sup>Tyr</sup> and the anticodon binding pocket of MjTyrRS may facilitate an improvement in unAA incorporation efficiency. In a prior report, it was demonstrated that MjTyrRS with an Asp286Arg mutation in the anticodon binding pocket had a higher aminoacylation rate and amber suppression efficiency.<sup>169</sup> However, no systematic evolution at the anticodon recognition site of MjTyrRS has been reported.

In this work, we utilized a directed evolution approach to identify MjTyrRS variants that displayed higher unAA incorporation efficiency. To create MjTyrRS libraries, four residues (Phe261, His283, Met285, and Asp286) that are proximal to C34 were fully randomized. Through

multiple rounds of selections, we obtained a collection of MjTyrRS variants with significantly improved unAA incorporation efficiency. The mutations in the evolved MjTyrRS variants were found to be specific for each unAA and could not be functionally transferred. In addition, the unAA incorporation efficiency has been further improved by combining our evolved MjTyrRS variants with a previously reported MjtRNA<sub>CUA</sub><sup>Tyr</sup> mutant. This evolution approach will not only enhance the performance and utility of current unAA incorporation system, but also facilitate the discovery of highly selective and efficient aaRSs for novel unAAs. Our study can be theoretically extended to improving the incorporation efficiency of unAA against other nonsense codons, such as quadruplet codons.

## 4.2 Approach

### 4.2.1 pAcFRS library construction and selection for improving the incorporation efficiency of pAcF

Site-specific incorporation of *p*-acetyl-L-phenylalanine (pAcF) that contains a keto group provides a useful tool for protein modifications and bio-conjugation.<sup>45,79,130</sup> To improve the incorporation efficiency of pAcF, we constructed a pAcFRS library, in which residues Phe261, His283, Met285 and Asp286 in the anticodon recognition pocket of wild-type pAcFRS (referred to as pAcFRS-wt hereafter) were randomized into all possible combinations of amino acids.<sup>170</sup> Complete randomization was achieved by overlapping PCR with synthetic oligonucleotide primers containing NNK (N= A, G, U, or C; K= U or G) codons. We confirmed a coverage greater than the 99% theoretical diversity of the library.

The resulting library was subjected to a positive selection using a dual-reporter (pREP) that contains amber mutations in the encoding genes for chloramphenicol acetyltransferase (Cm-Asp112UAG) and T7 RNA polymerase (RNAP-Gln107UAG) (**Figure 39**). The positive selection

was conducted in the presence of pAcFRS library, MjtRNA<sub>CUA</sub><sup>Tyr</sup>, and the dual-reporter, with 1 mM pAcF and chloramphenicol. Only cells containing pAcFRS mutants that can charge either pAcF or natural amino acids onto MjtRNA<sub>CUA</sub><sup>Tyr</sup> survived. Plasmids that encode pAcFRS mutants were isolated from the surviving cells and subjected to a negative selection in the absence of pAcF. A toxic barnase gene containing two amber mutations (Gln2UAG and Asp44UAG) was used. The pAcFRS mutants that could recognize natural amino acids led to the expression of full-length barnase and subsequent cell death. Conversely, cells containing orthogonal pAcFRS mutants survived. The clones that survived from two rounds of positive selection and one round of negative selection were screened for chloramphenicol resistance in the presence or absence of 1 mM pAcF. The clones that survived at 150 µg/mL chloramphenicol in the presence of 1 mM pAcF and that did not grow at 75 µg/mL chloramphenicol in the absence of pAcF was further characterized. Mutants pAcFRS-8G and pAcFRS-12B conferred the fastest growth rate in the presence of chloramphenicol and the brightest GFP fluorescence (driven by T7 RNAP). Another mutant pAcFRS-2B showed a relatively lower amber suppression efficiency.

To determine whether incorporation efficiency of pAcF can be improved using the obtained mutants, each of the plasmids encoding pAcFRS-8G, pAcFRS-12B, pAcFRS-wt and pAcFRS-D286R was transformed into *E. coli* GeneHogs containing a GFP reporter with an amber mutation at a permissive site (GFPuv-Asn149UAG). As shown in **Figure 40**, both pAcFRS-8G and pAcFRS-12B mutants displayed significant improvements in the incorporation efficiency of pAcF in comparison to the pAcFRS-wt and pAcFRS-D286R controls.<sup>169</sup> The mutations for these evolved pAcFRS derivatives are listed in **Figure 40**. The results show that the stacking interactions at position 34 with Phe261 and His283 was not essential for the recognition of MjtRNA<sub>CUA</sub><sup>Tyr</sup> by MjTyrRS variants. Interestingly, residues Met285 and Asp286 were mutated into smaller amino

acids, which contradicted our expectation that larger amino acids are needed to fill the anticodon-binding pocket caused by the G34C mutation. pAcFRS-8G was chosen for further characterization since it displayed the highest amber suppression efficiency and better reproducibility.

#### 4.2.2 Attempts to transfer the beneficial mutations to other aminoacyl-tRNA synthetases

MjTyrRS derivatives that recognize different unAAs generally share a conserved anticodon-binding pocket. We examined whether the beneficial mutations observed in pAcFRS-8G could be directly transferred to other MjTyrRS variants to improve the incorporation efficiency of unAAs. We examined BpaRS<sup>11</sup> (a MjTyrRS derivative that specifically recognizes *p*-benzoyl-L-phenylalanine, or Bpa), sTyrRS<sup>47</sup> (a MjTyrRS derivative that specifically recognizes sulfotyrosine, or sTyr), pIodoFRS<sup>171</sup> (a MjTyrRS derivative that specifically recognizes *p*-iodo-L-phenylalanine, or pIodoF), and pAzFRS<sup>44</sup> (a MjTyrRS derivative that specifically recognizes *p*-azido-L-phenylalanine, or pAzF). BpaRS-8G, sTyrRS-8G, pIodoFRS-8G, and pAzFRS-8G mutants were constructed by transplanting the Phe261Gly, His283Leu, Met285Val, and Asp286Gln mutations from pAcFRS-8G into the corresponding MjTyrRS variants. There was no obvious enhancement in the incorporation efficiency of Bpa, pIodoF, or pAzF after installation of the beneficial mutations according to GFPuv-based fluorescence assay (**Figure 41**). The incorporation efficiency of sTyr was dramatically decreased. These results suggest that the recognition of an anticodon and the recognition of unAA are mutually dependent.

#### 4.2.2 BpaRS library construction and selection for improving the incorporation efficiency of Bpa

##### Bpa

Next, we investigated whether the evolution of the BpaRS at anticodon binding pocket can improve the incorporation efficiency of Bpa. To achieve this, a BpaRS library was constructed by fully randomizing residues Phe261, His283, Met285 and Asp286. The BpaRS library was screened

following the same selection strategy for pAcFRS. Three hits (BpaRS-8E, BpaRS-11D, BpaRS-11H) displayed better incorporation efficiency than BpaRS-wt. BpaRS-8E showed the best incorporation efficiency (**Figure 42**). BpaRS-8E and BpaRS-11D, which had a convergent amino acid sequence, displayed relatively lower amber suppression efficiency than BpaRS-8E. Sequencing results showed that the stacking interaction between C34 and residues Phe261 and His283 was not required for anticodon recognition. Met285 of the selected BpaRS variants was replaced with smaller amino acids. In BpaRS-11D and BpaRS-11H, the negatively charged Asp286 remained unchanged or was mutated to a positively charged Arg. This was consistent with a previous finding that the Asp286Arg mutation can improve aminoacylation activity and amber suppression efficiency.<sup>95,169</sup>

#### **4.2.3 sTyrRS library construction and selection for improving the incorporation efficiency of sTyr**

Inspired by our success in evolving pAcFRS and BpaRS, we constructed a sTyrRS library in which the same set of residues (Phe261, His283, Met285, and Asp286) were fully randomized. The resulting library was subjected to multiple rounds of positive and negative selections to identify desirable sTyrRS variants. Two hits (sTyrRS-2A, sTyrRS-5A) displayed a slight enhancement in amber suppression efficiency compared to sTyrRS-wt (**Figure 43**). Again, the mutations at position 261 and 283 suggested that the stacking interaction required for the wild-type MjtRNA<sub>GUA</sub><sup>Tyr</sup> was no longer required for the amber suppressor MjtRNA<sub>CUA</sub><sup>Tyr</sup>. Since position 285 was mutated to smaller amino acids in all selections, we concluded that this position might not directly engage in anticodon recognition. Amino acids with shorter side chains were also observed at position 286.

#### 4.2.4 Expression of GFPuv mutants with pAcFRS and MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants

MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1 is a previously reported MjtRNA<sub>CUA</sub><sup>Tyr</sup> variant with beneficial T-stem mutations.<sup>90</sup> These mutations repaired a suboptimal interaction between archaeal tRNA and prokaryotic EF-Tu, and led to a significantly improved amber suppression efficiency.<sup>90</sup> We investigated whether beneficial mutations of MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1 and our evolved MjTyrRS variants can be combined to facilitate amber suppression. Since the T-stem of MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1 does not directly interact with MjTyrRS, we envisaged that it could still be efficiently recognized by pAcFRS variants. We compared the expression levels of GFPuv-Asn149UAG in *E. coli* GeneHogs in the presence of pAcFRS variants, MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants, and 1 mM pAcF. As shown in **Figure 44**, the combination of MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1 with pAcFRS variants significantly enhanced the GFPuv expression level relative to MjtRNA<sub>CUA</sub><sup>Tyr</sup>.

We performed a large-scale purification of GFPuv-Asn149UAG in the presence of pAcFRS and MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants (**Figure 45**). The pAcFRS-8G/MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1 pair resulted in the highest protein yield (~76.2 mg/L), which was approximately 2-fold higher than the pAcFRS-8G/MjtRNA<sub>CUA</sub><sup>Tyr</sup> pair (~33.9 mg/L) and 4-fold higher than the pAcFRS-wt/MjtRNA<sub>CUA</sub><sup>Tyr</sup> pair (~17.9 mg/L). The previously reported mutant pAcFRS-D286R showed only minor improvement in incorporation efficiency (~22.7mg/L).

#### 4.2.5 Quantification of aminoacyl-tRNA synthetase expression level

We examined the expression level of pAcFRS-wt and pAcFRS-8G by western blot. In order to quantify the two MjTyrRS variants in crude cell lysates, a 6× His-tag was fused to the C-terminus of each protein during sub-cloning. As shown in **Figure 46**, we did not observe a difference between the two samples, which ruled out the possibility that the improvement in pAcF incorporation efficiency was due to a higher expression level of pAcFRS-8G.

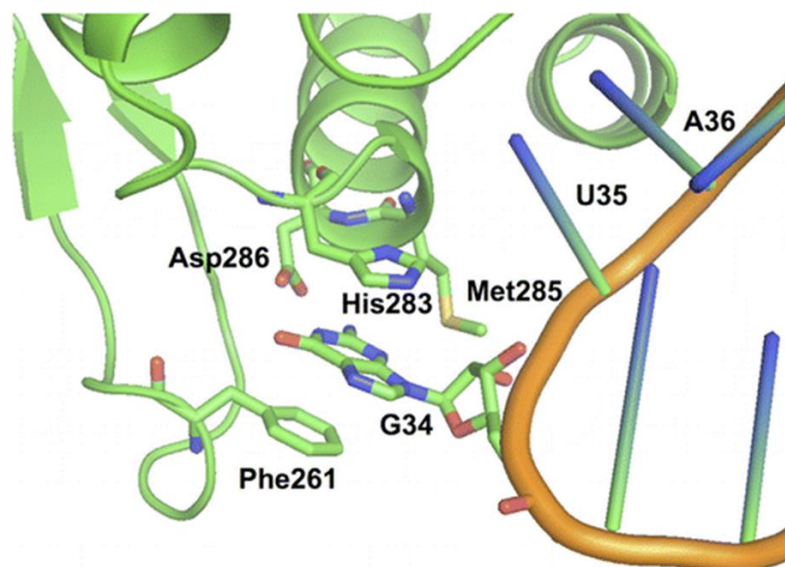
#### 4.2.6 Growth curves of cells containing the evolved aminoacyl-tRNA synthetase

We evaluated the growth rate of *E. coli* cells harboring MjTyrRS variants. As shown in **Figure 47**, pAcFRS-12B and BpaRS-8E grew at a similar rate to controls. Conversely, pAcFRS-8G and sTyrRS-5A displayed a relatively slower growth rate compared to controls. These results suggest the improvement in unAA incorporation was not due to faster growth rates of the evolved MjTyrRS variants.

#### 4.3 Summary and remarks

In this chapter, we introduced a directed evolution approach to evolve aminoacyl-tRNA synthetase. By mutating key residues at the anticodon-binding interface, we identified MjTyrRS variants with notable enhancements in the unAA incorporation efficiency. The beneficial effects were likely due to an optimized recognition of the anticodon region in MjtRNA<sub>CUA</sub><sup>Tyr</sup> by the evolved MjTyrRS variants. It was not caused by higher expression levels or lower toxicity of these synthetases. Significantly higher protein expression levels were achieved by the combination of evolved MjTyrRS variants and an optimized MjtRNA<sub>CUA</sub><sup>Tyr</sup>. We also found that the optimal efficiencies of MjTyrRS-MjtRNA<sub>CUA</sub><sup>Tyr</sup> pairs are unAA-specific. This strategy can potentially be applied to optimize the interaction between aaRSs and other nonsense codon suppressor tRNAs, such as quadruplet codon decoding tRNAs.

#### 4.4 Figures and tables

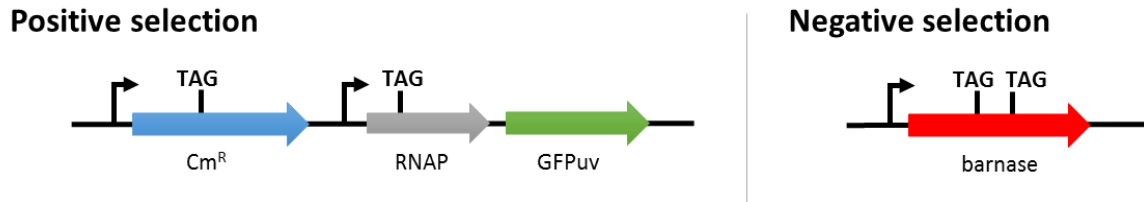


**Figure 38.** Recognition of anticodon region of MjtRNA<sub>GUA</sub><sup>Tyr</sup> by MjTyrRS (PDB ID: 1J1U).

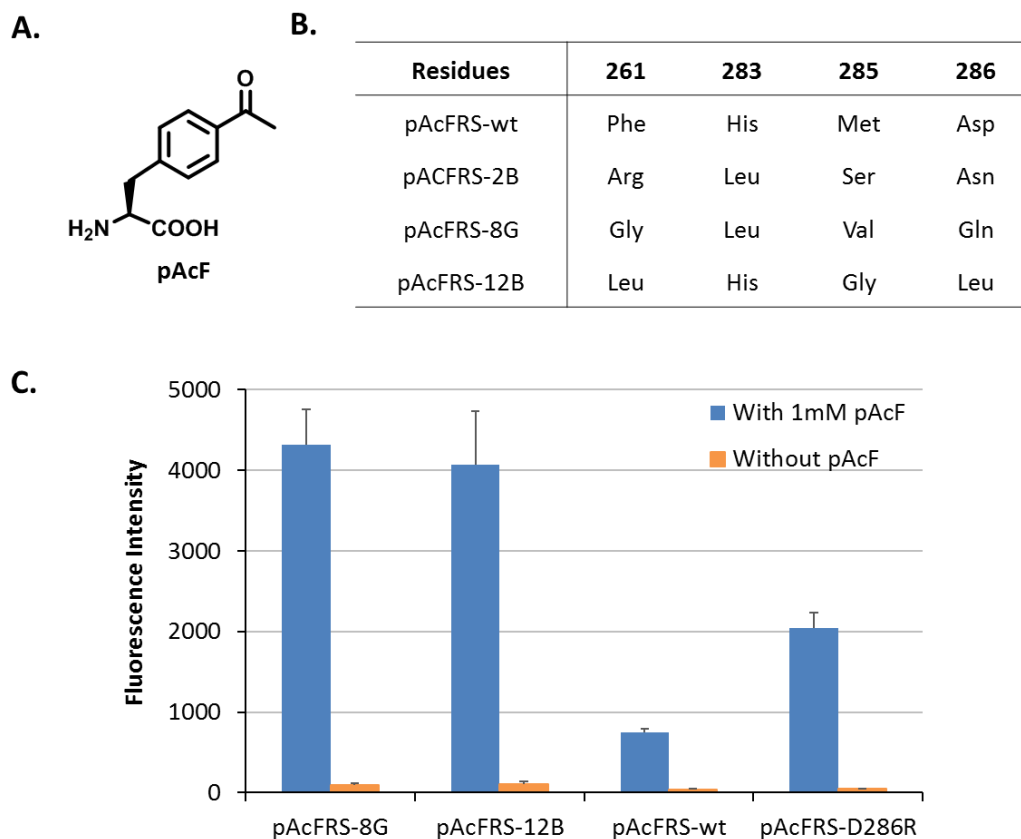
Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015

American Chemical Society.”



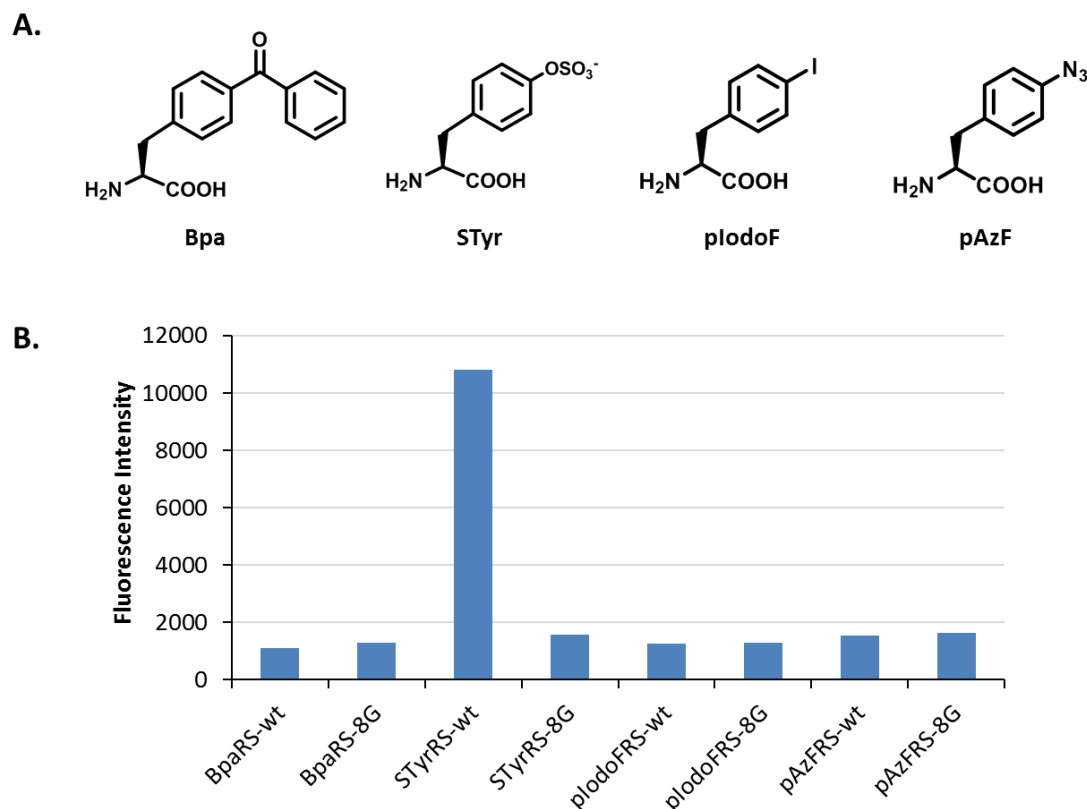


**Figure 39.** Positive and negative selections to identify MjTyrRS mutants with enhanced unAA incorporation efficiency.

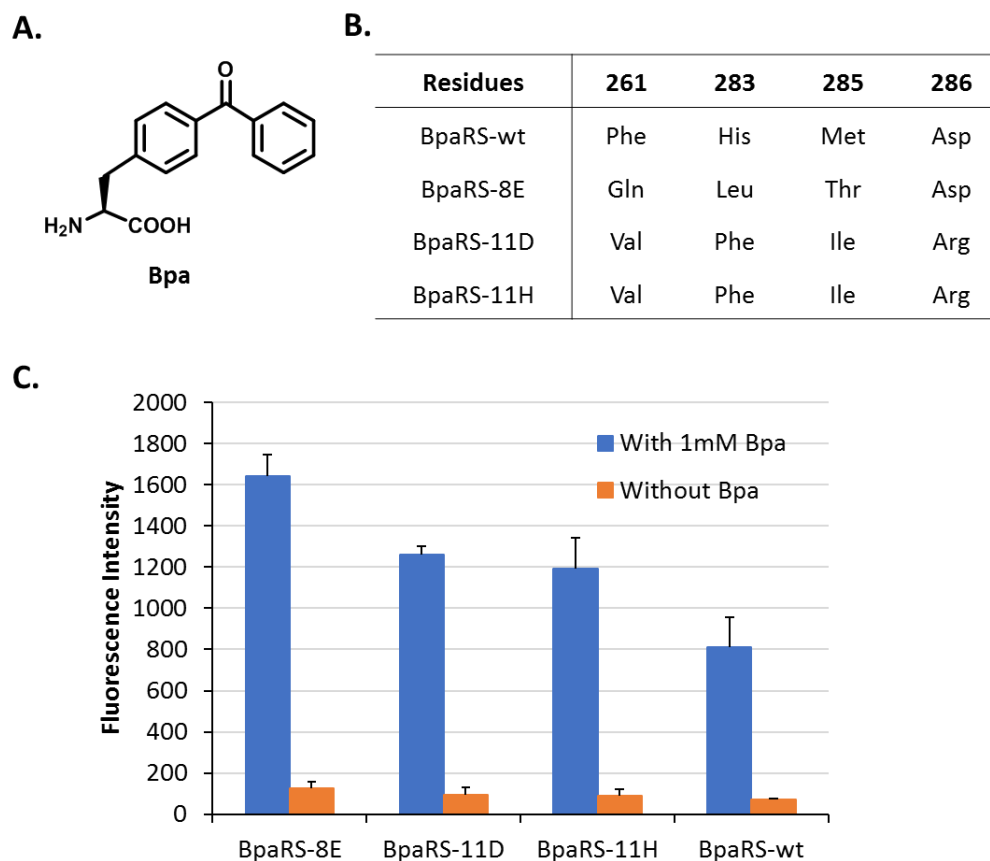


**Figure 40.** Characterization and validation of pAcFRS mutants. (A) Structure of *p*-acetyl-L-phenylalanine (pAcF). (B) Sequences of pAcFRS variants. (C) GFPuv fluorescence assay of cells containing evolved pAcFRS derivatives and MjtRNA<sub>CUA</sub><sup>Tyr</sup> in the presence or absence of 1 mM pAcF. Fluorescence intensity was normalized to cell density. The data shown was an average of duplicate measurements.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”

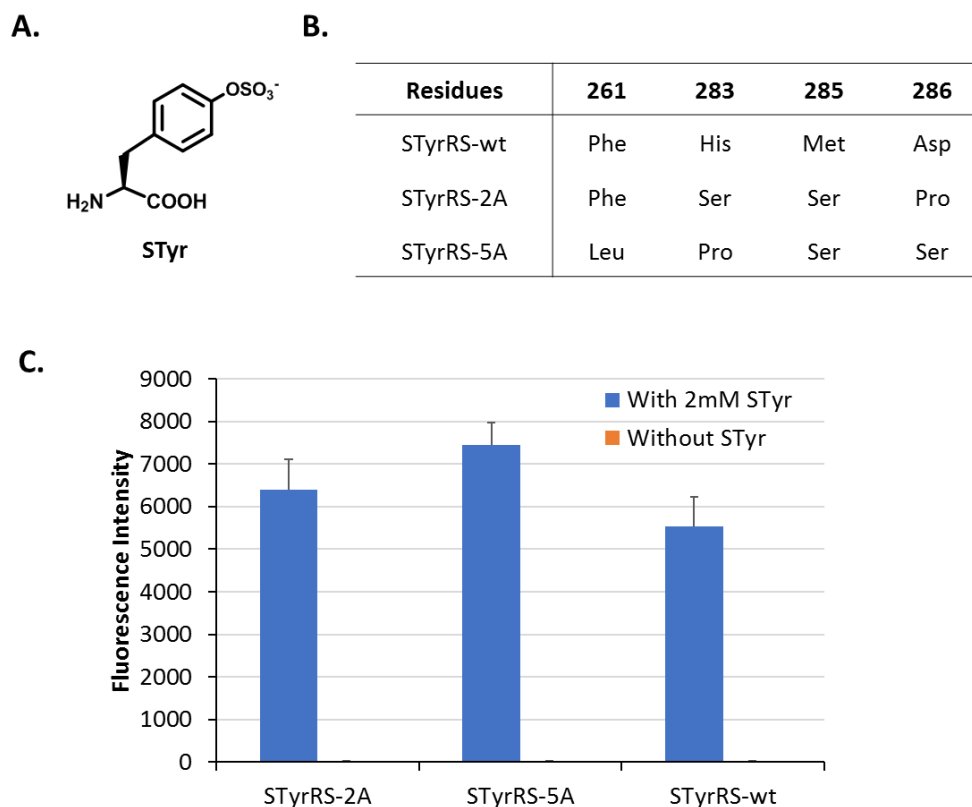


**Figure 41.** GFPuv fluorescence assay of cells containing MjTyrRS mutants. These MjTyrRS mutants share the same anticodon binding pocket with pAcFRS-8G. (A) Structure of Bpa, sTyr, plodoF, and pAzF. (B) Fluorescence readings of cells containing MjTyrRS derivatives. The GFPuv expression was conducted in the presence of MjtRNA<sub>CUA</sub><sup>Tyr</sup>, MjTyrRS derivatives and corresponding unAAs, respectively. The culture media was supplemented with 1 mM Bpa, 10 mM sTyr, 1 mM plodoF, or 1 mM pAzF respectively. Fluorescence intensity was normalized to cell density. The experiment was repeated multiple times and one representative set of data was presented in this figure.



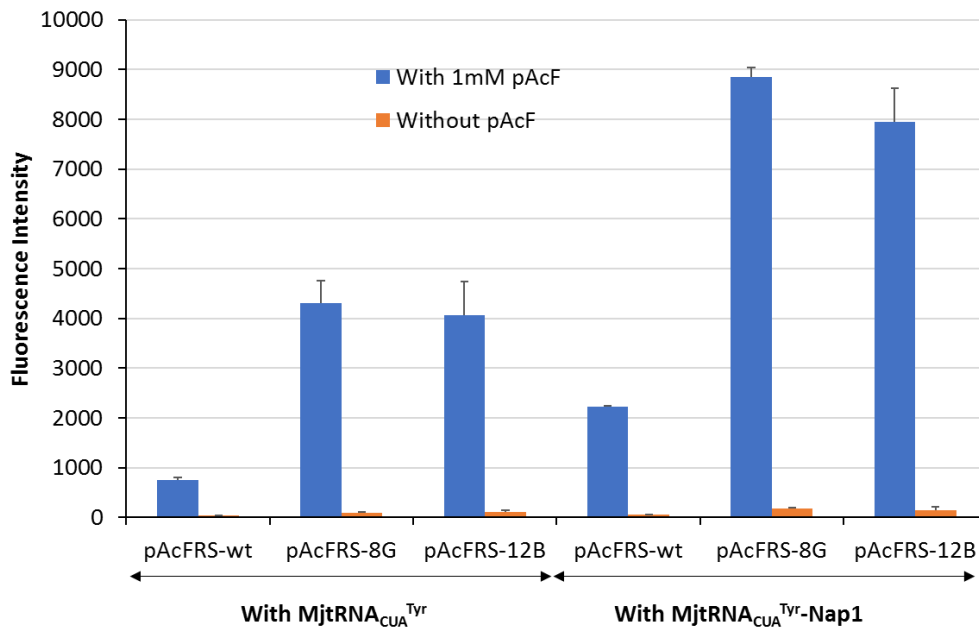
**Figure 42.** Characterization and validation of BpaRS mutants. (A) Structure of *p*-benzoyl-L-phenylalanine (Bpa). (B) Sequences of BpaRS variants. (C) GFPuv fluorescence assay of cells containing evolved BpaRS derivatives and MjtRNA<sub>CUA</sub><sup>Tyr</sup> in the presence or absence of 1 mM Bpa. Fluorescence intensity was normalized to cell density. The data shown was an average of duplicate measurements.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”



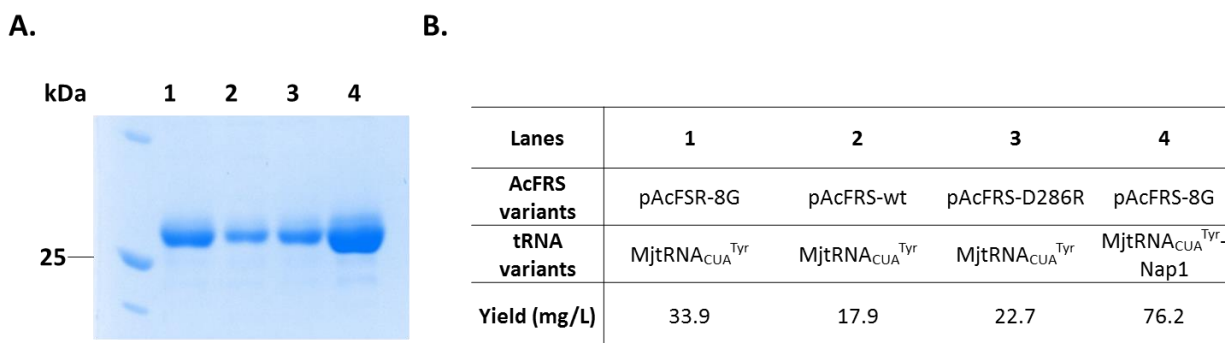
**Figure 43.** Characterization and validation of sTyrRS mutants. (A) Structure of sulfotyrosine (sTyr). (B) Sequences of sTyrRS variants. (C) GFPuv fluorescence assay of cells containing evolved sTyrRS derivatives and MjtRNA<sub>CUA</sub><sup>Tyr</sup> in the presence or absence of 2 mM sTyr. Fluorescence intensity was normalized to cell density. The data shown was an average of duplicate measurements.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”



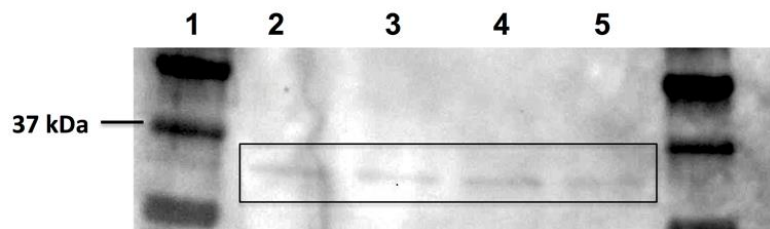
**Figure 44.** GFPuv fluorescence assay of cells containing MjTyrRS and MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”



**Figure 45.** Purification of GFPuv mutants in the presence of evolved pAcFRS and MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants. (A) SDS-PAGE of GFPuv mutants. GFPuv expression was conducted in the presence pAcFRS, MjtRNA<sub>CUA</sub><sup>Tyr</sup> variants, and 1 mM pAcF. Lane 1, pAcFRS-8G + MjtRNA<sub>CUA</sub><sup>Tyr</sup>; Lane 2, pAcFRS-wt + MjtRNA<sub>CUA</sub><sup>Tyr</sup>; Lane 3, pAcFRS-D286R + MjtRNA<sub>CUA</sub><sup>Tyr</sup>; Lane 4, pAcFRS-8G + MjtRNA<sub>CUA</sub><sup>Tyr</sup>-Nap1. (B) Protein yields of each sample in (A). 1 mM of pAcF was supplemented in the culture media.

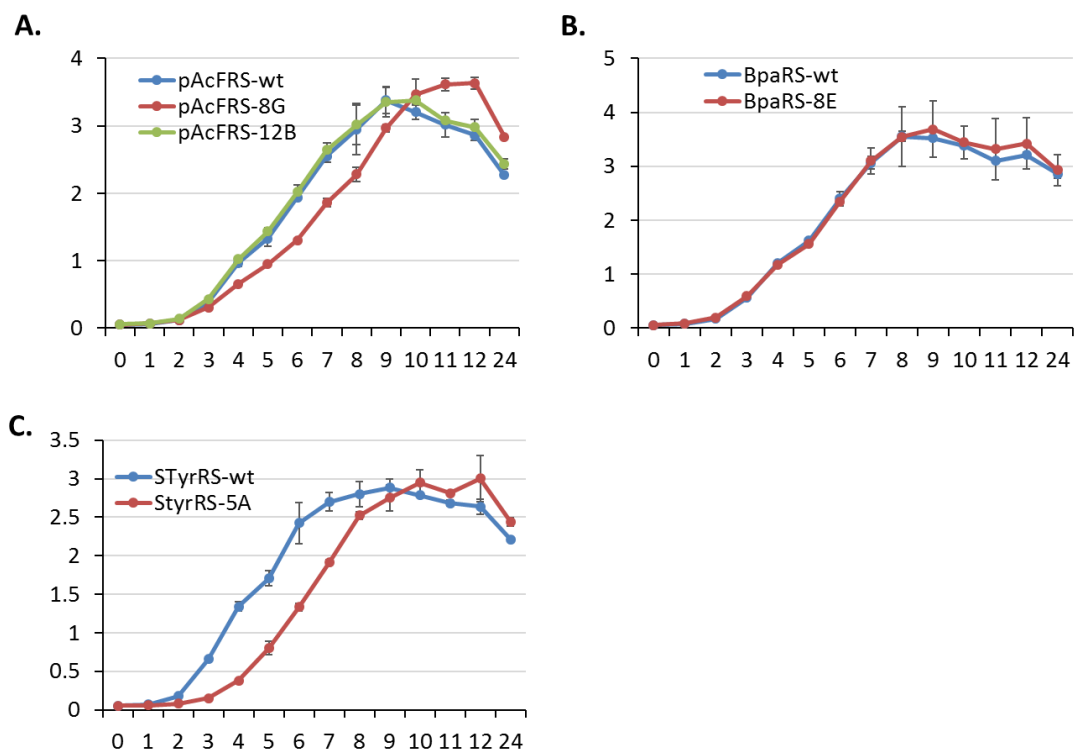
Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”



**Figure 46.** Western blot analysis of expression level of pAcFRS variants. Lane 1, protein maker; lane 2, pAcFRS-8G; lane 3, pAcFRS-8G; lane 4, pAcFRS-wt; lane 5, pAcFRS-wt. Protein expression was conducted in two biological repeats for pAcFRS-8G and pAcFRS-wt. Each sample was normalized to the same number of cells by OD<sub>600</sub> measurement.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”





**Figure 47.** Growth curves of cells containing MjTyrRS variants. (A) *E. coli* GeneHogs containing pAcFRS variants. (B) *E. coli* GeneHogs containing BpaRS variants. (C) *E. coli* GeneHogs containing sTyrRS variants. The data shown was an average of triplicate measurements.

Reproduced with permission from “ACS Synthetic Biology. 2015, 4, 207. Copyright 2015 American Chemical Society.”

## CHAPTER 5

**Experimental: Materials and methods****5.1 General materials and methods****5.1.1 General materials**

**DNA manipulation.** Restriction enzymes, Antarctic phosphatase (AP) were purchased from New England Biolabs (NEB). T4 DNA ligase was purchased from Thermo Fisher Scientific. KOD hot start DNA polymerase was purchased from EMD Millipore. All DNA oligonucleotides were purchased from Eurofins MWG operon or Sigma. Standard molecular biology techniques were used throughout.<sup>172</sup> Site-directed mutagenesis, insertion and deletion of genes were carried out by overlapping PCR.

***E. coli* strains.** *E. coli* GeneHogs (Thermo Fisher Scientific) was used for routine cloning and DNA propagation for non-viral genes. *E. coli* Max Efficiency Stbl-2 (Life Technologies) was used for manipulation of HIV-1 genes. *E. coli* C321.ΔA.exp (Addgene) and C321.ΔA (Addgene) were used for quadruplet decoding tRNA selection and evaluation. *E. coli* GeneHogs was used for MjTyrRS selection and evaluation.

**Medium and antibiotics.** All solutions were prepared in deionized water which was further treated by Barnstead Nanopure ultrapure water purification system (Thermo Fischer Scientific Inc). All purchased chemicals were used without further purification. LB broth (1 L) was prepared by adding 25 g LB broth media (1% tryptone, 0.5% yeast extract, and 1% NaCl) into 1 L of ddH<sub>2</sub>O and autoclaved. SOC medium (1 L) was prepared by adding 32.6 g SOC broth media (2% tryptone, 0.5% yeast extract, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 20 mM MgSO<sub>4</sub>, 20 mM glucose) in 1 L of ddH<sub>2</sub>O and autoclaved. LB agar (1 L) was prepared by adding 25 g LB broth media and 15 g agar into 1 L of ddH<sub>2</sub>O and autoclaved. Antibiotics were added where appropriate to following final

concentration: ampicillin, 100  $\mu\text{g}/\text{mL}$ ; kanamycin, 50  $\mu\text{g}/\text{mL}$ ; tetracycline, 12.5  $\mu\text{g}/\text{mL}$ ; chloramphenicol 34  $\mu\text{g}/\text{mL}$  or other concentrations for positive selections.

**Unnatural amino acids (unAAs).** The unnatural amino acids (unAAs), *p*-azido-L-phenylalanine (pAzF), *p*-iodo-L-phenylalanine (pIodoF), *p*-benzyol-L-phenylalanine (Bpa), and  $\text{N}^{\epsilon}$ -(*tert*-butoxycarbonyl)-L-lysine (BocK) were purchased from Bachem. *p*-acetyl-L-phenylalanine (pAcF),<sup>45</sup> sulfotyrosine (sTyr),<sup>173</sup> *o*-nitrobenzyl-oxycarbonyl- $\text{N}^{\epsilon}$ -L-lysine (ONBK),<sup>125</sup> *o*-nitrobenzyl-tyrosine (ONBY),<sup>49</sup> and 3'-azibutyl-N-carbamoyl-lysine (AbK)<sup>164</sup> were synthesized following previously reported methods.

**Protein purification and western blot reagents.** Ni Sepharose 6 Fast Flow resin (GE Healthcare) was used for purification of His-tagged fusion proteins. Glutathione Sepharose 4B resin (GE Healthcare) was used for purification of GST-tagged fusion proteins. BugBuster protein extraction reagent (EMD Millipore) was used for small-scale cell lysis and protein purification. 6x-His Tag monoclonal antibody (mouse) from Thermo Fisher Scientific was used for probing His-tagged fusion proteins in western blot. StrepTactin-HRP conjugate (Bio-Rad) was used to detect and quantify proteins fused with Strep-tag II in western blot. Goat-anti-mouse IgG (H+L)-HRP conjugate (Bio-Rad) was used as a secondary antibody for detecting mouse IgG primary antibody.

### 5.1.2 General methods

**PCR.** All PCR reactions were performed following standard protocols being described before.<sup>172</sup> A typical PCR reaction solution (25  $\mu\text{L}$ ) contained 2.5  $\mu\text{L}$  of 10X buffer for KOD Hot Start DNA polymerase, 1.5  $\mu\text{L}$  of  $\text{MgSO}_4$  (25 mM), 2.5  $\mu\text{L}$  of dNTPs (2 mM each), 0.75  $\mu\text{L}$  of primer mixture (10  $\mu\text{M}$  each), 0.5  $\mu\text{L}$  of DNA template (10-20 ng/ $\mu\text{L}$ ), 0.5  $\mu\text{L}$  of KOD Hot Start DNA polymerase, and 16.75  $\mu\text{L}$  of ddH<sub>2</sub>O. For library construction, a larger scale (100-200  $\mu\text{L}$ )

of PCR reaction was performed to obtain higher yields of amplification products. A typical 3-step PCR reaction was conducted under following conditions: a hot start at 95 °C for 2 min, and then 35 cycles of 95 °C for 20 s (denature), 56 °C for 15 s (annealing), and 70 °C for certain time (extension), and then a final extension at 70 °C for 3 min.

The reaction solution for overlapping PCR was prepared in a similar way as described above. But the primer mixture was omitted from the initial solution and DNA fragments with overlapping overhangs were used as the template. The PCR reaction was conducted under following conditions: 95 °C for 2 min, and then 10 cycles of 95 °C for 20 s, 50 °C for 15s, and 70 °C for certain time, and then paused at 4 °C. After the addition of the primer mixture, the PCR reaction was resumed, and conducted under following conditions: 30 cycles of 95 °C for 20 s, 56 °C for 15 s, and 70 °C for certain time, and then a final extension at 70 °C for 3 min.

**Restriction enzyme digestion.** Generally, the vector and insert were digested with the same restriction enzymes to create blunt ends or compatible sticky ends for ligation. A typical digestion reaction (40 µL) contained 10 µL of DNA (~100 ng/µL), 4 µL of 10X digestion buffer, 1 µL of each restriction enzyme, and proper amount of ddH<sub>2</sub>O. The reaction was incubated at suggested temperature (normally 37 °C) for 3 h.

**Agarose gel electrophoresis.** Agarose gels were commonly prepared at concentrations of 0.8% to 2% (w/v) in 1X TAE buffer based on the size of DNA fragments needed to be isolated. Ethidium bromide (0.5 µg/mL) was added into the agarose solution to allow for visualization of DNA fragments under UV light. To analyze or purify DNA fragments (e.g., plasmid after restriction enzyme digestion), the samples were mixed with 6X loading dye and then loaded into a freshly poured agarose gel alongside a DNA ladder. Then the gel was run in 1X TAE buffer at 100 V for 45-60 min. The size of DNA fragments was determined by comparison to the DNA ladder. The

gel bands containing desired DNA fragments were carefully cut out using razor blades and directly used for gel extraction.

**DNA gel extraction.** Zymoclean DNA Gel Extraction kit was used to extract DNA from excised agarose gel slice. In a 1.5 mL Eppendorf tube, one volume of gel slice was mixed with three volumes of agarose dissolving buffer (ADB). The resulting suspension was incubated at 52 °C until the gel was fully dissolved. The dissolved DNA-gel solution was loaded onto a Zymo-spin column, extracted, and washed following manufacturer's instructions. The purified DNA was eluted in 10-20  $\mu$ L ddH<sub>2</sub>O and stored at 4 °C.

**Ligation.** To ligate the digested insert into a vector (recipient plasmid), a standard ligation reaction was performed. In general, a vector + insert ligation reaction (5  $\mu$ L) contained 50 ng of vector, insert (the molar ratio of insert: vector was 3:1), 0.5  $\mu$ L of 10X T4 ligation buffer, 0.25  $\mu$ L of T4 ligase, and proper amount of ddH<sub>2</sub>O. A vector only control was used to determine the "background" level of self-ligation of the vector. For a regular cloning, the ligation reaction was incubated at room temperature for 1 h before transformation. For the construction of an aaRS or tRNA library, the ligation reaction was performed at 14 °C for 16 h in a larger scale (1  $\mu$ g of vector + appropriate amount of insert).

**Sequence and ligation independent cloning (SLIC cloning).** Unlike traditional ligation dependent cloning, SLIC is not constrained by DNA sequence contents. This approach relies on T4 DNA polymerase's exonuclease activity to create homologous overhangs in both vector and insert, which allows for subsequent annealing and assembly. A SLIC cloning reaction (20  $\mu$ L) containing 150 ng of linearized plasmid DNA, insert (with a 2:1 molar ratio of insert: vector), 2  $\mu$ L Buffer 2.1 (NEB), 1  $\mu$ L T4 DNA polymerase (1:6 dilution in Buffer 2.1), and ddH<sub>2</sub>O, was prepared on ice. Then the reaction mixture was incubated at 22 °C for 30 min (3' to 5' exonuclease

digestion), 75 °C for 15 min (heat inactivation of T4 DNA polymerase), and 37 °C for 20 min (annealing). Then 2 µL of the reaction mixture was transformed into GeneHogs chemical competent cells to obtain single colonies.

**Preparation of chemical competent cells.** An *E. coli* strain of interest was streaked on LB agar containing appropriate antibiotics (or no antibiotics) to obtain single colonies. Then a single colony was cultured in 5 mL LB at 37 °C (GeneHogs and C321.ΔA.exp) or 30 °C (C321.ΔA and Stbl2) overnight with shaking. 1 mL of overnight culture was sub-cultured in 100 mL of fresh LB medium at 30 °C with shaking until reaching OD<sub>600</sub> of 0.4 to 0.6. The cells were harvested and centrifuged at 5,000g and 4 °C for 10 min. The cells were resuspended in 1 mL of pre-chilled CCMB solution on ice and centrifuged again. The supernatant was carefully removed and the cells were resuspended in 4 mL of pre-chilled CCMB solution on ice. The cell suspension was aliquoted into 50 µL portions in 1.5 mL Eppendorf tubes and stored at -80 °C.

**Transformation of chemical competent cells.** An aliquot of chemical competent cells (50 µL) was thawed on ice, followed by addition of 10-100 ng of DNA sample, such as, ligation products, SLIC cloning products, and plasmids. The cells were incubated on ice for 15-30 min, followed a heat shock at 42 °C for 25-30 s. The cells were immediately moved back to ice bath and incubated for another 2 min. Then 450 µL of LB or SOC medium was added into each sample and the cells were allowed to recover at 37 °C (GeneHogs and C321.ΔA.exp) for 1 h or at 30 °C (C321.ΔA and Stbl2) for 1.5 h. 100 µL of recovered cells were plated on LB agar containing appropriate antibiotics and incubated overnight to obtain single colonies.

**Preparation of electrocompetent cells.** The *E. coli* strain of interest was cultured in 1 L of LB medium following the same protocol when preparing chemical competent cells. When OD<sub>600</sub> of 0.6-1.0 was reached, the cells were harvested and centrifuged at 5,000g and 4 °C for 10 min.

The cells were resuspended in 100-200 mL of pre-chilled autoclaved ddH<sub>2</sub>O on ice, and centrifuged at 5,200g and 4 °C for 10 min. After carefully removing the supernatant, the cells were resuspended in 100-200 mL of pre-chilled autoclaved 10% glycerol solution on ice, and centrifuged at 5,200g and 4 °C for 10 min. The washing step was repeated twice with glycerol solution. Each volume of cell pellet was subsequently resuspended in half-volume of 10% glycerol solution on ice, aliquoted into 60 µL portions into 1.5 mL Eppendorf tubes and used freshly for electroporation.

**Electroporation of electrocompetent cells.** In this study, electroporation was used for the construction and selection of a tRNA or aaRS library. Briefly, 2 µL of concentrated ligation products or plasmid DNA (100-250 ng/µL) was added into an aliquot of freshly prepared electrocompetent cells on ice and mixed by gently stirring. The cell suspension was transferred into a pre-chilled electroporation cuvette on ice. The cells were electroporated at 2500 V (Eppendorf Electroporator 2510) and immediately resuspended in 1 mL of pre-warmed SOC medium. The cuvette was rinsed with an additional 1 mL of SOC medium. The combined cell suspension was transferred into a pre-warmed 50 mL Erlenmeyer flask and recovered at 37 °C for 1 h with shaking. Then the cells were transferred into pre-warmed LB broth with appropriate antibiotics and cultivated overnight at 37 °C with shaking. After 18 h of cultivation, 2 tubes of glycerol stocks were made by mixing 0.75 mL of cell culture with 0.5 mL of 50% glycerol solution in a 2 mL cryo-tube and stored at -80 °C. Certain number of cells ( $>4.6 \times$  library size) were plated on LB agar containing appropriate antibiotics for the selection.

**Transfection.** Transfection was commonly used to introduce DNA or RNA into mammalian cells. Briefly, the cells were seeded in a 6-well plate to be 50-70% confluent at transfection. Before transfection, the cell culture medium was replaced with 1 mL of fresh DMEM medium without

FBS. In one 1.5 mL Eppendorf tube, 1 µg of each plasmid DNA was diluted in 500 µL DMEM. In another tube, 10 µL of lipofectamine 2000 transfection reagent was diluted in 500 µL DMEM. The diluted DNA was added into diluted lipofectamine solution. The DNA-lipid complex was incubated at room temperature for 20 min before being added to cells. 48 h after transfection, the cells were imaged and virus-containing cell culture was collected for future analysis.

## 5.2 Experimental for Chapter 2

### 5.2.1 Plasmid construction

**pcDNA3.1-tRNA<sub>CUA</sub>.** The amber suppressor tRNA<sub>CUA</sub> was derived from *Bacillus stearothermophilus* tRNA<sup>Tyr</sup>. The G35 of the tRNA<sup>Tyr</sup> was mutated to C to generate a CUA anticodon. The tRNA<sub>CUA</sub> gene lacking 3'-CCA but with a 3'-TTTTTCT sequence was PCR amplified. A human U6 promoter was PCR amplified from pCMV-pylRSNBK-1<sup>174</sup> and added in front of the tRNA<sub>CUA</sub> gene by overlapping PCR. The U6-tRNA<sub>CUA</sub> gene cassette was PCR amplified and inserted into pcDNA3.1/hygro+ (Life Technologies) in front of the CMV promoter to afford pcDNA3.1-tRNA<sub>CUA</sub>.

**pAzFRS.** The AzFRS-encoding gene was PCR amplified from plasmid pSWAN-pAzpaRS<sup>175</sup> and inserted into pcDNA3.1-tRNA<sub>CUA</sub> behind the non-regulated CMV promoter to afford plasmid pAzFRS.

**pAcFRS.** The AcFRS-encoding gene was PCR amplified from plasmid pSWAN-pApaRS<sup>175</sup> and inserted into pcDNA3.1-tRNA<sub>CUA</sub> behind the non-regulated CMV promoter to afford plasmid pAcFRS.



**pIodoFRS.** The IodoFRS-encoding gene was PCR amplified from plasmid pSWAN-pIpaRS<sup>175</sup> and inserted into pcDNA3.1-tRNA<sub>CUA</sub> behind the non-regulated CMV promoter to afford plasmid pAzFRS.

**pEGFP.** The EGFP-encoding gene containing an amber nonsense codon at a permissive site (Tyr40TAG) was inserted into pcDNA3.1-tRNA<sub>CUA</sub> behind the non-regulated CMV promoter to afford plasmid pEGFP.

**pSUMA variants.** HIV-1 clone, pSUMA.c/2821(referred to as pSUMA-wt), was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Kappes and Dr. Chistina Ochsenbauer. Overlapping PCR was used to introduce amber mutations onto the HIV-1 genome encoded by pSUMA.

**pNL4-3 variants.** HIV-1 clone, pNL4-3(referred to as pNL43-wt), was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH from Dr. Malcolm Martin. Plasmids HIV-1-iGFP and HIV-1-NL-GI were generous gifts from Dr. Benjamin K. Chen.<sup>176,177</sup> To facilitate the insertion of foreign genes into HIV-1 genome, two unique restriction enzyme cutting sites, *ClaI* and *XmaI*, were introduced into pNL43-wt to yield pNL43-*ClaI-XmaI*. The two cutting sites were installed between env and nef genes by overlapping PCR using primers SUMA.Env.Nef SeqF, NL43-*ClaI*-R, NL43-*XmaI*-F, and NL43-Nef-R3. The AzFRS-encoding gene was PCR amplified using primers AzF-*ClaI* and AzF-*XmaI* and inserted into pNL43-*ClaI-XmaI* between *ClaI* and *XmaI* sites to afford pNL43-AzFRS. EGFP and BocKRS-encoding genes were inserted at the same locus to give pNL43-EGFP and pNL43-BocKRS respectively. The U6-tRNA<sub>CUA</sub> cassette was PCR amplified and inserted into *ClaI* or *XmaI* site of pNL43-AzFRS to yield four different constructs, pNL43-AzFRS-tRNA-X (X=1, 2, 3, and 4). Amber mutations at position Trp36 of p17 and position Tyr59 of HIV-1 protease were introduced into pNL43-AzFRS-tRNA-2 using overlapping

PCR. All of the five genomic UAG codons were replaced with UAA codons in pNL43-wt to afford pNL43-All-UAA. The modifications on pNL43-Trp36-AzFRS-tRNA-2 and pNL43-Tyr59-AzFRS-tRNA-2 were transplanted into pNL43-All-UAA to generate amber free versions of HIV-1 variants. CMV promoter were inserted before AzFRS-encoding gene to give pNL43-Trp36-CMV-AzFRS-tRNA-2 and pNL43-Tyr59-CMV-AzFRS-tRNA-2.

The following primers were used for the construction of HIV-1 variants in this study:

### **SUMA-MA-Y132**

Mut1-F1	CTAGAACGATTCGCAGTTAACCCCTG
Mut1-R1	GTTCTGCACTATAGGCTAATTTTGGCTGACCTGGCTG
Mut1-F2	TAGCCTATAGTGCAGAACCTCCAG
Mut1-R2	GTGCCTATAGCTTTGTGTCCACAG

### **SUMA-RT-F216**

Mut2-F1	CAAAGTAAGACAATATGATCAGGTAAC
Mut2-R1	ATGGCCTATACTGGAGTATTGTATGGATTTTC
Mut2-F2	CTCCAGTATAGGCCATAAAGAAAAAAGACAGTAC
Mut2-R2	GCTTCCCATGTTTCTCTTTGTATG

### **SUMA-Tat-T20**

Mut3-F1-2	GCAGTGCAAAGGTAGCCACAGAAAG
Mut3-R1	GCAAGTGGTACAAGCCTACTTAGGCTGACTTCCTGGATGCTTC
Mut3-F2	GGCTTGTACCACTTGCTATTGTAAAAAG
Mut3-R2	GTAACACAGAGTGGAGTTAATCTTAC

### **SUMA-MA-A119**

Mut1-F1	CTAGAACGATTCGCAGTTAACCCCTG
Mut4-R1	TTTCCTGCGTCAGCCTATGCTTGCTGTGCTTTTTTCTTAC
Mut4-F2	CTGACGCAGGAAACAACAGCCAG
Mut1-R2	GTGCCTATAGCTTTGTGTCCACAG

### **SUMA-CA-L268**

Mut1-F1	CTAGAACGATTCGCAGTTAACCCCTG
---------	----------------------------

Mut5-R1 AATCCCTAGATTATCCATCTTTTATAAAATTTCTCCTAC  
 Mut5-F2 GGATAATCTAGGGATTAATAAAAATAGTAAGGATGTATAG  
 Mut1-R2 GTGCCTATAGCTTTGTGTCCACAG

**SUMA-RT-F271**

Mut2-F1 CAAAGTAAGACAATATGATCAGGTAAC  
 Mut6-R1 GGAAGTACTAATATGCATCACCCACATCTAG  
 Mut6-F2 ATATTAGTCAGTTCCCTTAGATAAAGAATTCAG  
 Mut2-R2 GCTTCCCATGTTTCTCTTTGTATG

**SUMA-RT-L365**

Mut2-F1 CAAAGTAAGACAATATGATCAGGTAAC  
 Mut7-R1 CACCTCTACAGATGTTCTCTCAGTTCCTC  
 Mut7-F2 CATCTGTAGAGGTGGGGATTTACCACAC  
 Mut2-R2 GCTTCCCATGTTTCTCTTTGTATG

**SUMA-MA-W36Q127**

Mut8-F1-BamHI GTAGAGGATCCACTAGTAAC  
 Mut8-R1 TAGCTCCCTGCTTGCCTATACTATATGTTTTAATTG  
 Mut8-F2 GGCAAGCAGGGAGCTAGAAC  
 Mut1-R2 GTGCCTATAGCTTTGTGTCCACAG

**SUMA-MA-F44**

Mut8-F1-BamHI GTAGAGGATCCACTAGTAAC  
 Mut9-R1 CAGGGTAACTG CTATCGTTCTAGCTCCCTGCTTG  
 Mut9-F2 GCAGTTAACCCTGGCCTGTTAG  
 Mut1-R2 GTGCCTATAGCTTTGTGTCCACAG

**SUMA-RT-Y59**

Mut1-F1 CTAGAACGATTCGCAGTTAACCCTG  
 Mut10-R-BstEII ATTTCTATGGTTACCTGATCCTATTGTCTTACTTTGATAAAAC

**SUMA-CA-K359**

SUMA-HpaI-F1 CAGGGAGCTAGAACGATTCGCAGTTAACCCT  
 Gag-K359-R1 CATTGCTTCCGCCAAAACCTCTTGCCTAATGGCCGGGTCTCCAC  
 Gag-K359-F2 GCAAGAGTTTTGGCGGAAGCAATG  
 SUMA-BstEII-R2 TGTGTCCACAGATTTCTATGGTTACCTGATC

**NL-GI-N40TAG**

N40TAG-F1 GCTATAAGACGCGTCCACCATG  
 N40TAG-R1 CCTAGGTGGCATCGCCCTC  
 N40TAG-F2 CGATGCCACCTAGGGAAAGCTGACCCTGAAGTTC  
 N40TAG-R2 CGTCTAGATTACTTGTACAGCTCATC

**pNL43-ClaI-XmaI**

SUMA.Env.Nef SeqF CCTCAAATATTGGTGGAAATC  
 NL43-Cla1-R CCCGGGGGTGGCTATCGATTTATAGCAAAAATC  
 NL43-Xma1-F ATCGATAGCCACCCCGGGATGGGTGGCAAGTGGTC  
 NL43-Nef-R3 CTTCTAGCCAGGCACAAGC

**pNL43-AzFRS**

AzF-Cla1 ATAAATCGATAGCCACCATGGCAAGCAGTAACTTG  
 AzF-Xma1 TTTCCCGGGTTAAACGGGCCCTTTCCAG

**pNL43-EGFP**

EGFP-Cla1 GAAAGGGCTTTACTATAAATCGATAGCCACCATGGTGAGCAAGGGCGAC  
 EGFP-Xma1 TGGCCCGGGTTACTTGTACAGCTCGTC

**pNL43-BocKRS**

PylRS-Cla1 ATAAATCGATAGCCACCATGGATAAAAAACCATTAG  
 PylRS-Xma1 TGGCCCGGGTTACAGATTGGTTGAAATC

**pNL43-AzFRS-tRNA-1/pNL43-AzFRS-tRNA-2**

tRNA-Xma1-F GTAACCCGGGACAGAAAAATGGAG  
 tRNA-Xma1-R ATATCCCGGGTCGGGCAGGAAGAG

**pNL43-AzFRS-tRNA-3/pNL43-AzFRS-tRNA-4**

tRNA-Cla1-F GGAAATCGATACAGAAAAATGGAG  
 tRNA-Cla1-R TTATATCGATTTCGGGCAGGAAGAG

**pNL43-Trp36-AzFRS-tRNA-2**

Mut8-F1 GAAAAGTGCCACCTGACGTCTAAG  
 Mut8-R1 TAGCTCCCTGCTTGCCTATACTATATGTTTTAGTTTATATTG  
 Mut8-F2 GGCAAGCAGGGAGCTAGAAC  
 Mut8-R2 CATCTGGCCTGGTGCAATAG

**pNL43-Tyr59-AzFRS-tRNA-2**

Mut10-F1	GAAGTGACATAGCAGGAACTAC
Mut10-R1	TATGAGTATCTGATCCTACTGTCTTACTTTGATAAAACCTC
Mut10-F2	GATCAGATACTCATAGAAATCTGCGGAC
Mut10-R2	TCCTGAAGTCTTTATCTAAGGGAAGTCT
<b>pNL43-all-UAA</b>	
NL43-AgeI-Up	GGAACCAAAGCACTAACAG
NL43-Vif-TAA-R	GTGTCCATTCATTGTATGG
NL43-Vif-TAA-F	CATACAATGAATGGACACTAAAGCTTTTAGAGGAACTT
NL43-Vpr-TAA-R	GGATCTACTGGCTCCATTTC
NL43-Vpr-TAA-F	ATGGAGCCAGTAGATCCTAAACTAGAGCCCTGGAAGCATCCAG
NL43-Vpu-TAA-R	CAGATCATCAATATCCCAAG
NL43-Vpu-TAA-F	GGGATATTGATGATCTGTAATGCTACAGAAAAATTGTG
NL43-NheI-Down	CAATTAAAAGTGTGCGTTACAA
NL43-NheI-Up	GAGAGCATTGTGTTACAATAG
NL43-Tat-TAA-R	TTCCTTCGGGCCTGTCTG
NL43-Tat-TAA-F	CGACAGGCCCGAAGGAATAAAAGAAGAAGGTGGAGAG
NL43-Rev-TAA-HpaI	TGAGCAAGTTAACAGCATTATTCTTTAGTTCCTGACTC
<b>pNL43-Trp36-CMV-AzFRS-tRNA-2/pNL43-Tyr59-CMV-AzFRS-tRNA-2</b>	
CMV-F-ClaI	ATAAATCGATGTTGGCATTGATTA
AzFRS-R-XbaI-Bst	CATTTTTCTGTTCTAGATTAACGGGCCC
AzFRS-F-XbaI-Bst	GCCCGTTTAATCTAGAACAGAAAAATG
NL43-Nef-R3	CTTCTAGCCAGGCACAAGC

### 5.2.2 Transfection and generation of HIV-1

293T cells were grown in 2 mL DMEM medium containing 10% FBS and 2 mM L-glutamine at 37 °C in 6-well flat bottom plates in a humidified atmosphere of 5% CO<sub>2</sub>. When 60-70% confluency was reached, cells were transfected with 1µg of appropriate plasmid(s) using Lipofectamine 2000 (Life Technologies) following manufacturer's instructions. 6 h post-transfection, the culture medium containing transfection reagent was carefully removed and

replaced with 2 mL fresh medium with or without unAA. A final concentration of 1 mM pAzF, 1 mM pIodoF, 1 mM pAcF, 0.4 mM ONBY, or 1 mM ONBK was added into the medium for corresponding studies. After 48 h of cultivation, the supernatant was collected separately from each well. The FBS concentration in the harvested culture medium was adjusted to 20%. The culture medium was filtered through a 0.45- $\mu$ m filter. Viral load of the filtered culture medium was determined by p24-assay or qRT-PCR. Then the culture medium was directly used for infection assay or stored at -150 °C for later analysis.

### 5.2.3 Virus ultracentrifugation

293T cells in T175 flasks were transfected with appropriate plasmids (60  $\mu$ g) using Lipofectamine 2000 (120  $\mu$ L). After 48 h of cultivation, cell culture medium was collected from each flask separately and filtered through a 0.45- $\mu$ m filter. The filtrate (35 mL) from each flask was centrifuged at 25,000 rpm in Optima L-100X ultracentrifuge (Beckman Coulter) at 4 °C for 90 min. The virus pellet was resuspended in 1 mL fresh medium and aliquoted into 200  $\mu$ L portions in sterile screw-cap vials and stored at -150 °C.

### 5.2.4 Quantification of viral load

**p24 assay.** The presence and level of p24 (HIV-1 capsid protein) in collected culture medium was measured with Retroteck HIV-1 p24 Antigen ELISA 2.0 kit (ZeptoMetrix Corporation) following manufacturer's instructions. Briefly, 200  $\mu$ L viral-containing culture medium, the positive control (125 pg/mL p24) and the negative control were added separately into micro-plate wells coated with anti-p24 monoclonal antibody and incubated at 37 °C for 1.5 h. Then the wells were washed and incubated with antibody-HP conjugate at 37 °C for 1 h. After multiple washes, HRP-substrate was added into the wells for color development. The reactions were stopped by the

addition of 1 M H<sub>2</sub>SO<sub>4</sub> and absorbance at 450 nm were examined. The amount of p24 was obtained by interpolation from a point-to-point plot or from a linear regression analysis of the standard curve.

**qRT-PCR.** Viral RNAs were extracted from collected culture medium using QIAamp Viral RNA Mini kit (Qiagen). The viral load was determined by qRT-PCR using C1000 Thermo Cycler and CFX96 Real-time system (Bio-Rad) with Taqman Fast Virus 1-Step Master Mix (Life Technologies), using primers V1: 5'-GCCTCAATAAAGCTTGCCTTGA-3', V2: 5'-GGGCGCCACTGCTAGAGA-3', and probe V3: 5'-FAM/CCAGAGTCACACAACAGACGGGCACA/BHQ\_1/-3'.

### 5.2.5 Infection assay

**Qualitative infection assay.** The infectivity of HIV-1 variants was evaluated using TZM-bl cells and X-gal staining (Genlantis). Briefly, a mixture of virus-containing culture medium, 10% DMEM growth media and DEAE-dextran (final concentration of 40 mg/mL), were added into each well of 6-well flat bottom plates that contained TZM-bl cells. After 48 h of cultivation at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>, TZM-bl cells were washed, fixed, and stained using X-gal solution for 2 h at 37 °C. Infectivity was evaluated by counting X-gal-positive cells under a light microscope.

**Tissue Culture Infectious Dose 50 (TCID<sub>50</sub>).** Infectious titers of HIV-1 variants were quantified by standard tissue culture infectious dose 50 (TCID<sub>50</sub>) method using X-gal staining assay (as mentioned above). Briefly, 4-fold serial dilutions of viruses were performed in quadruplicate in 96-well flat bottom plates and applied for X-gal staining assay. The TCID<sub>50</sub> values were calculated by Spearman-Kärber formula according to the negative end-point and dilution factors. The TZM-bl cell-only sample and the wild-type HIV-1 virus were included as negative and positive controls respectively in all TCID<sub>50</sub> measurements.

**Multicycle infection assay.** Concentrated viruses from ultracentrifugation (200  $\mu$ L) were added to T25 flasks containing  $1 \times 10^6$  Sup-T1 cells. After 24 h of cultivation, cells were collected by centrifugation and the supernatants containing unadsorbed viruses were discarded. The cell pellets were resuspended in 5 mL fresh medium and was centrifuged again to remove residual free viruses. The cell pellets were resuspended in 5 mL culture with or without 1 mM pAzF and cultivated at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>. The status of cells was monitored everyday by a light microscope. After 2 weeks of cultivation, viruses were collected by ultracentrifugation and TCID<sub>50</sub> values were measured.

### 5.2.6 Fluorescence spectroscopy and cell imaging

The fluorescent images and bright-field images were obtained by a Nikon Eclipse TE3000 microscope and an EVOS FL Auto Imaging System with DIC, respectively. The samples were excited at 488 nm to acquire EGFP fluorescence images at 530/25 nm.

### 5.2.7 Analysis of AzFRS transcription using quantitative PCR

48 h post-transfection, the total RNA was extracted from transfected 293T cells using RNeasy Plus Mini Kit (Qiagen). The quantity and purity of RNA samples were determined by NanoDrop 2000 (Thermo Scientific). The cDNA was synthesized using an oligo-dT primer (IDT) and Superscript III reverse transcriptase (Life Technologies) on C1000 Thermo Cycler (Bio-Rad). The qPCR was conducted on the CFX96 Real-Time detection system (Bio-Rad) using a hot start (95 °C for 5 min), and 40 amplification cycles (95 °C for 15s, 60 °C for 30s). For amplification and detection of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) control, primers Q1: 5'-ACATCATCCCTGCCTCTACT-3' and Q2: 5'-TCAGGTCCACCACTGACA-3', and probe Q3: 5'-/56-FAM/CAAGGTCAT/ZEN/CCCTGAGCTGAACGG/3IABkFQ/-3' were used. For amplification and detection of *AzFRS* gene, primers Q4: 5'-GCATCAGAATCAGGTGTTTGG-3'



and Q5: 5'-TGGTAGAATTTGTACGGGCTG-3', and probe Q6: 5'-/56-FAM/TGCTTTAGT/ZEN/GATCAGCGGAACGGT/3IABkFQ/-3' were used.

### **5.2.8 Viral transcription and provirus detection using RNAscope and DNAscope *in situ* hybridization**

Viral RNA and DNA were detected using single-copy sensitivity DNAscope and RNAscope *in situ* hybridization (ISH) techniques.<sup>178</sup> Briefly, after 72 h of cultivation with or without pAzF, Sup-T1 cells were collected and centrifuged, spotted on glass-slide, and fixed with 4% PFA for 1 h. HIV-1 viral RNA was detected with V-HIV Clade B antisense probes and RNAscope 2.0 HD red reagent kit (Advanced Cell Diagnostics, CA). HIV-1 viral DNA was detected with V-HIV Clade B sense probes and RNAscope 2.0 HD red reagent (Advanced Cell Diagnostics, CA). All experiments were conducted according to manufacturer's instructions.

### **5.2.9 Amplification and sequencing for aaRS-tRNA<sub>CUA</sub> insertion**

Cells infected with wild-type pNL4-3 virus, the pNL43-Trp36-AzFRS-tRNA-2 variant, or the pNL43-Tyr59-AzFRS-tRNA-2 variant were collected from Sup-T1 cells after 2 weeks of cultivation in the presence of pAzF. Viral RNA was extracted from cell culture using QIAamp Viral RNA Mini kit (Qiagen). RNA quality was verified using NanoDrop 2000 (Thermo Scientific). The cDNA of extracted RNA was synthesized by reverse transcription using Superscript III reagents (Life Technologies). The mixture of viral RNA and a gene-specific antisense primer (5'-AAGATCTACAGCTGCCTT-3') was heated at 65 °C for 5 min and then incubated on ice for 2 min. A master mix consisting of following reagents was then added: 5× First-Strand buffer, dithiothreitol, RNaseOUT recombinant RNase Inhibitor (40 units/μL), SuperScriptIII RT (200 units/μL), and RNase free water. The mixture was incubated at 50 °C for 60 min, 70 °C for 15 min, followed by the addition of 1 μL RNase H (5U/μL) and incubation at

37 °C for additional 20 min. The cDNA was stored at -20 °C for future analysis. PCR amplification of AzFRS-tRNA<sub>CUA</sub> insertion was performed using High Fidelity Platinum Taq polymerase (Invitrogen) in a 20 µL of reaction with forward primer 5'-TTATAGAAGTATTACAAG-CAGCTTATAG-3' and reverse primer 5'- AAGATCTACAGCTGCCTT-3' under the following conditions: 94 °C for 5 min, 36 cycles of 94 °C for 30 s, 53 °C for 30 s, and 68 °C for 2 min, and a final extension at 68 °C for 5 min. The amplicons of full-length insertion were sequenced at Sequetech (Mountain View, CA) using the Sanger method.

### 5.2.10 Protein expression and purification

293T cells were grown in T75 flasks with DMEM medium containing 10% FBS, and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>. When 60-70% confluency was reached, 293T cells were transfected with 12 µg of pEGFP-40TAG plasmid and 12 µg of pAzFRS plasmid with 36 µL Lipofectamine 2000 according to manufacturer's instructions. 6 h post-transfection, the culture medium was replaced with 12 mL fresh medium containing 1 mM pAzF. Cells were cultivated for additional 36 h before being washed with DPBS, lysed with RIPA buffer (Thermo Scientific), and partially purified using Ni-NTA resin (GE Healthcare).

### 5.2.11 LC/MS/MS

The corresponding protein band of EGFP-40pAzF was cut from SDS-PAGE and in-gel digested with trypsin (in 50 mM ammonium bicarbonate, pH 8.0) overnight at 37 °C. The resulting peptide fragment were extracted with 0.1% formic acid/75% acetonitrile, and then subjected to LC/MS/MS analysis using a Waters Q-TOF Ultima. Database searches were performed on an in-house Mascot server (Matrix Science Ltd., London, UK). For pAzF substitution site mapping on EGFP, pAzF substitution for tyrosine was included as a variable modification.

## 5.3 Experimental for Chapter 3

### 5.3.1 Plasmid construction

**pBK-tRNA<sub>NCUA</sub><sup>Pyl</sup>-N11 library.** Procedures for library construction were adapted from a previously reported method.<sup>179</sup> The tRNA mutants with complete randomization at 11 selected positions in ASL were obtained by overlapping PCR using pBK-mmPylT as the template.<sup>125</sup> The resulting PCR product containing *lpp* promoter-RNA mutants-rrnC terminator was digested with *Nco*I and *Xho*I and inserted into the same sites of pBK vector to generate pBK-tRNA<sub>NCUA</sub><sup>Pyl</sup>-N11 library. The quality of this library was confirmed by DNA sequencing.

**pBK-tRNA<sub>CUA</sub><sup>Pyl</sup>-N10 library.** The tRNA<sub>CUA</sub><sup>Pyl</sup>-N10 library was constructed following the same protocol described above. Briefly, the tRNA mutants with complete randomization at 10 selected positions in ASL was inserted into *Nco*I and *Xho*I sites of pBK-vector to afford pBK-tRNA<sub>CUA</sub><sup>Pyl</sup>-N10 library. The quality of this library was confirmed by DNA sequencing.

**pBK-BocKRS.** Plasmid pBK-BocKRS was constructed by inserting BockRS-encoding gene behind the constitutive *glnS* promoter (*P<sub>glnS</sub>*) on pBK vector.<sup>180</sup>

**pREP-BocKRS-UAGN (positive selection reporter).** This reporter was constructed by modifying plasmid pRep-Cm12b.<sup>90</sup> Briefly, the UAG codon at position 98 in the chloramphenicol acetyltransferase-encoding gene on pRep-Cm12b was changed to UAGN (N= A, U, G, or C) by site-directed mutagenesis. The resulting plasmid, pRepCM12b-UAGN, was digested with *Xba*I and ligated to a DNA fragment encoding *P<sub>glnS</sub>*-BocKRS cassette, which was amplified from pBK-BocKRS, to yield plasmid pREP-BocKRS-UAGN. Individual plasmids with different quadruplet codons of interest, pREP-BocKRS-UAGA, pREP-BocKRS-UAGU, pREP-BocKRS-UAGG, and pREP-BocKRS-UAGC were isolated from single *E. coli* colonies and verified by DNA sequencing.

**pGFPuv-UAGN (GFPuv reporter).** Four different pGFPuv-UAGN(N=A, U, G, or C) reporters were constructed by inserting a DNA fragment encoding GFPuv-Asn149UAGN into *NdeI* and *SacI* sites of plasmid pLei-GFPuv-Asn149UAG<sup>90</sup> to replace the original GFPuv reporter.

**pGFPuv-NCUA-wt (N= A, G, U, or C).** A single insertion 33.5N (N= A, G, U, or C) was introduced into tRNA<sub>CUA</sub><sup>Pyl</sup>-wt by overlapping PCR to yield tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt. The encoding genes for tRNA<sub>NCUA</sub><sup>Pyl</sup>-wt were inserted between *SpeI* and *PstI* sites of pGFPuv-UAGN to yield pGFPuv-NCUA-wt.

**pGFPuv-NCUA-X (N= A, G, U, or C; X= hit number).** The encoding genes for tRNA hits were inserted between *SpeI* and *PstI* sites of corresponding pGFPuv-UAGN reporters by SLIC-cloning to give pGFPuv-NCUA-X variants.

**pGFPuv-UAGN-BocKRS (N= A, G, U, or C).** The pGFPuv-UAGN-BocKRS reporters were generated by deleting the original tRNA fragment and inserting a DNA fragment encoding *P<sub>glnS</sub>*-BocKRS cassette, which was amplified from pBK-BocKRS, into *SpeI* and *PstI* sites of pGFPuv-NCUA-wt by SLIC-cloning.

**pGFPuv-UAG-BocKRS.** The pGFPuv-UAG-BocKRS reporter was constructed by deleting the original tRNA fragment and inserting a DNA fragment encoding *P<sub>glnS</sub>*-BocKRS cassette, which was amplified from pBK-BocKRS, into *SpeI* and *PstI* sites of pLei-GFPuv-Asn149UAG<sup>90</sup> by SLIC-cloning.

The following primers are used in the quadruplet codon decoding tRNA evolution and evaluation:

**pBK-tRNA<sub>NCUA</sub><sup>Pyl</sup>-N11 library**

tRNA-T-Lib-F1- CATGCCATGGGTTCCACAGGGTAGCCAGCAGC  
NcoI  
tRNA-T-Lib-R1 ATTCGATCTACATGATCAGGT  
tRNA-T-Lib-TAGN- TCATGTAGATCGAATNNNNNCTANNNNNGTTCAGCCGGGTTAGATTC  
F2  
tRNA-T-Lib-R2- CCGCTCGAGCAGAACATATCCATCGCGTCCGC  
XhoI

**pBK-tRNAN<sub>CUA</sub><sup>Pyl</sup>-N10 library**

tRNA-T-Lib-F1- CATGCCATGGGTTCCACAGGGTAGCCAGCAGC  
NcoI  
tRNA-T-Lib-R1 ATTCGATCTACATGATCAGGT  
tRNA-T-Lib-TAG- TCATGTAGATCGAATNNNNNCTANNNNNGTTCAGCCGGGTTAGATTC  
F2  
tRNA-T-Lib-R2- CCGCTCGAGCAGAACATATCCATCGCGTCCGC  
XhoI

**pREP-BocKRS-UAGN**

Cm-F1 CGCTCTAGACAATTGGTGCAC  
Cm-R1 CTCATGGAAAACGGTGTAACAAG  
Cm-F2 ACCGTTTTCCATGAGTAGNACTGAAACGTTTTTCATCGCTCTG  
Cm-R2 CCACTCATCGCAGTACTGTTG

**pLei-GFPuv-UAGN**

pLei-GFP-F-NdeI GAGAAATTACATATGAGTAAAG  
pLei-GFP-R1 GTGTGAGTTATAGTTGTAICTC  
pLei-GFP- GTACA ACTATAACTCACACTAGAGTATAACATCACGGCAGAC  
N149TAGA-F2  
pLei-GFP- GTACA ACTATAACTCACACTAGTGTATAACATCACGGCAGAC  
N149TAGT-F2  
pLei-GFP- GTACA ACTATAACTCACACTAGCGTATAACATCACGGCAGAC  
N149TAGC-F2  
pLei-GFP- GTACA ACTATAACTCACACTAGGGTATAACATCACGGCAGAC  
N149TAGG-F2

pLei-GFP-R-SacI      GATGGAGCTCTTTGTAGAGTTC  
**pLei-GFPuv-NCUA-wt/pLei-GFPuv-NCUA-X**  
 SLIC-pLei-PylT-PstI    TTCGCTAAGGATCTGCAGTGGCGGAAACC  
 SLIC-pLei-PylT-      GTTGCCCGTCTCACTAGTGAAAAGAAAAACAACCCTGGCG  
 SpeI

### 5.3.2 Positive selection

The library plasmids were transformed into *E. coli* C321.ΔA.exp or C321.ΔA electrocompetent cells containing plasmid pREP-BocKRS-UAGN (N=A, G, U, or C). Transformants were cultivated in LB media containing kanamycin and tetracycline. After 12 h of cultivation, cells were harvested. Based on calculation, a certain number of cells ( $>4.6 \times$  the size of the library) were plated on LB agar containing kanamycin (50  $\mu\text{g}/\text{mL}$ ), tetracycline (12.5  $\mu\text{g}/\text{mL}$ ), BocK (5 mM), and chloramphenicol (concentration ranging from 34  $\mu\text{g}/\text{mL}$  to 75  $\mu\text{g}/\text{mL}$ ). The selection plates were incubated at 37 °C (C321.ΔA.exp) or 30 °C(C321.ΔA) for 48 h.

### 5.3.3 Hit verification

Selected numbers of single colonies were further screened by replication onto LB agar plates with varying concentrations (50, 75, or 100  $\mu\text{g}/\text{mL}$ ) of chloramphenicol in the presence or absence of 5 mM BocK. Only the one that grew in the presence of BocK but did not grow in the absence of BocK were further characterized.

### 5.3.4 Fluorescence analysis of bacterial culture

*E. coli* C321.ΔA containing plasmid pBK-BocKRS and a pGFPuv-NCUA variant (wt or tRNA hits) was cultured in 1 mL LB media containing ampicillin, kanamycin, and chloramphenicol at 30 °C. After 18 h of cultivation, 50  $\mu\text{L}$  culture was sub-cultured in 1 mL fresh LB medium containing ampicillin, kanamycin, chloramphenicol, IPTG (0.1 mM), and 5 mM BocK (or 5 mM

Lys, or none). After addition 16 h of cultivation at 30 °C, 1 mL of culture was collected, washed, and resuspended in 1 mL of potassium phosphate buffer (50 mM, pH 7.4). The processed cells were directly used for fluorescence and cell density measurements using a Synergy H1 Hybrid plate reader (BioTek Instruments). The fluorescence of GFPuv was monitored at  $\lambda_{\text{Ex}}=390$  nm and  $\lambda_{\text{Em}}=510$  nm. The cell density was measured by absorbance at 600 nm. The fluorescence intensity was normalized to cell density for comparison of different samples. The incorporation of AbK and ONBK was conducted using the same protocol described above with previously reported PylRS variants. The cross test among different UAGN codons were conducted using a combination of pBK-tRNA-hit and pGFPuv-UAG(N)-BocKRS. Reported data was the average of three measurements with standard deviations.

### 5.3.5 Protein expression and purification

*E. coli* C321.ΔA containing plasmid pBK-BocKRS and a pGFPuv-NCUA-X variant was cultured in 5 mL LB media containing kanamycin, chloramphenicol at 30 °C. After 18 h of cultivation, 0.5 mL culture was sub-cultured in 50 mL medium. The protein expression was induced at OD<sub>600</sub> of 0.6 by addition of IPTG (0.5 mM) and BocK (5 mM). Cells were collected by centrifugation at 5,000 g and 4 °C for 15 min. Harvested cells were resuspended in lysis buffer containing potassium phosphate (20 mM, pH 7.4), NaCl (150 mM) and imidazole (10 mM). Then cells were lysed by sonication. Cell debris was removed by centrifugation (21,000 g, 4 °C, 30 min). The cell-free lysate was applied to Ni Sepharose 6 Fast Flow resin (GE Healthcare). Protein purification was conducted by following manufacturer's instructions. Protein concentrations were determined by Bradford assay (Bio-Rad). Purified protein was isolated by SDS-PAGE and digested with trypsin prior to MS analysis.

## 5.4 Experimental for Chapter 4

### 5.4.1 Plasmid construction

**pBK-pAcFRS-library.** Residues Phe261, His283, Met285 and Asp286 in pAcFRS were mutated to NNK (N=A, G, U, or C; K=U, G) by overlapping PCR using pAcFRS-wt<sup>45</sup> as template. The PCR products were digested with *NdeI* and *PstI* and then ligated into *NdeI* and *PstI* sites of pBK<sup>181</sup> vector to give a pAcFRS library.

**pBK-BpaRS-library.** Residues Phe261, His283, Met285 and Asp286 in BpaRS were mutated to NNK (N=A, G, U, or C; K=U, G) by overlapping PCR using BpaRS-wt<sup>11</sup> as template. The PCR products were digested with *NdeI* and *PstI* and then ligated into *NdeI* and *PstI* sites of pBK<sup>181</sup> vector to give a BpaRS library.

**pBK-sTyrRS-library.** Residues Phe261, His283, Met285 and Asp286 in sTyrRS were mutated to NNK (N=A, G, U, or C; K=U, G) by overlapping PCR using sTyrRS-wt<sup>47</sup> as template. The PCR products were digested with *NdeI* and *PstI* and then ligated into *NdeI* and *PstI* sites of pBK<sup>181</sup> vector to give a sTyrRS library.

The following primers were used in the evolution of MjTyrRS variants:

#### pBK-MjTyrRS-library

pBK-TyrRS-NdeI-F1	TTGAGGAATCCCATATGGACGA
TyrRS-Lib1-R1	CAATTCCTTATTTTTAAATAA
TyrRS-Lib1-F2	TTATTTAAAAATAAGGAATTGNKCCANNKNNKTTAAAAAATGC
TyrRS-Lib1-R2	GCGAACGCCTTATCCGGCCTG
TyrRS-F261-R	TTTTTCTGGCCTTTTTATGGT
TyrRS-F261-F	ACCATAAAAAGGCCAGAAAAANNKGGTGGAGATTTGACAGTTAA

#### pBK-pAcFRS-D286R

pBK-TyrRS-NdeI-F1	TTGAGGAATCCCATATGGACGA
-------------------	------------------------



TyrRS-Lib1-R1	CAATTCCTTATTTTTAAATAA
pBK-wtAcF-D286R-F2	AAAAATAAGGAATTGCATCCAATGCGCTTAAAAAATGCTG
TyrRS-Lib1-R2	GCGAACGCCTTATCCGGCCTG
<b>pBK-MjTyrRS-8G variants</b>	
pBK-TyrRS-NdeI-F1	TTGAGGAATCCCATATGGACGA
8G-R1	CTCCATCCCTCCAACAGC
8G-F2	GCTGTTGGAGGGATGGAG
TyrRS-Lib1-R2	GCGAACGCCTTATCCGGCCTG

#### 5.4.2 Positive selection

The library DNAs were transformed into *E. coli* GeneHogs electrocompetent cells containing plasmid pREP<sup>181</sup> that harbors MjtRNA<sub>CUA</sub><sup>Tyr</sup> and a chloramphenicol acetyltransferase-encoding gene with an amber mutation at position Asp112. Transformants were cultivated in LB media containing kanamycin and tetracycline. After 12 h of cultivation, cells were harvested. Based on calculation, a certain number of cells ( $>4.6 \times$  the size of the library) were plated on LB agar containing kanamycin, tetracycline, unAA (e.g., 1 mM pAcF), and chloramphenicol (concentration ranging from 50  $\mu\text{g}/\text{mL}$  to 250  $\mu\text{g}/\text{mL}$ ). The selection plates were incubated at 37 °C for 24 h. Survived cells were pooled and library plasmids were isolated by DNA gel electrophoresis.

#### 5.4.3 Negative selection

*E. coli* GeneHogs was cotransformed with plasmid pNEG<sup>181</sup> (containing MjtRNA<sub>CUA</sub><sup>Tyr</sup> and a barnase-encoding gene with two amber mutations at position Gln2 and Asp44) and library plasmids isolated from the positive selection. Transformants were plated on LB agar containing ampicillin, kanamycin, and 0.2% L-arabinose. The selection plates were incubated at 37 °C for 12 h. Survived cells were then pooled the library plasmids were isolated for another round of positive selection.

#### 5.4.4 Hit verification

Selected number of single colonies from the last round positive selection were screened by replication onto LB agar plates with varied concentration of chloramphenicol (34, 50, 75, 100, 150, or 250  $\mu\text{g}/\text{mL}$ ) in the presence or absence of the appropriate unAA. Only the clones that grew in the presence of unAA but did not grow in the absence of unAA were selected for further evaluation.

#### 5.4.5 Fluorescence analysis of bacterial culture

*E. coli* GeneHogs harboring plasmids pBK-MjTyrRS variant and pLei-GFPuv-Asn149UAG<sup>90</sup> was cultured in 5 mL LB media containing kanamycin and chloramphenicol at 37 °C. The expression of GFPuv was induced at the OD<sub>600</sub> of 0.6 by additions of IPTG (0.1 mM) in with or without unAA. Following cultivation at 37 °C for additional 16 h, 1 mL of cell culture were collected, washed, and resuspended in 1 mL of potassium phosphate buffer (50 mM, pH 7.4). The processed cells were directly used for fluorescence and cell density measurement using Synergy H1 Hybrid plate reader (Bio-Teck Instruments). The fluorescence of GFPuv was monitored at  $\lambda_{\text{Ex}} = 390$  nm and  $\lambda_{\text{Em}} = 510$  nm. The cell density was estimated by absorbance at 600 nm. The fluorescence intensity was normalized to cell density for comparison of different samples. Reported data was the average of two or more measurements with standard deviations.

#### 5.4.6 Protein expression and purification

Similar cell cultivation procedure for fluorescence analysis was applied to preparing 50 mL of *E. coli* culture for protein purification. Cells were collected by centrifugation at 5,000g and 4 °C for 15 min. Harvested cells were resuspended in lysis buffer containing potassium phosphate (20 mM, pH 7.4), NaCl (300 mM), and imidazole (10 mM). Cells were subsequently lysed by sonication. Cell debris were removed by centrifugation at 21,000g and 4 °C for 30 min. The cell-free lysate was applied to Ni Sepharose 6 Fast Flow resin (GE Healthcare). Protein was purified

following manufacturer's instructions. Protein concentrations were determined by Bradford assay (Bio-Rad). Purified proteins were analyzed by SDS-PAGE.

## References

- 1 Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A New Functional Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J Am Chem Soc* **122**, 5010-5011, doi:10.1021/ja000595y (2000).
- 2 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of *Escherichia coli*. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).
- 3 Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology* **8**, 883-890, doi:[http://dx.doi.org/10.1016/S1074-5521\(01\)00063-1](http://dx.doi.org/10.1016/S1074-5521(01)00063-1) (2001).
- 4 Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat Biotech* **20**, 1044-1048, doi:[http://www.nature.com/nbt/journal/v20/n10/supinfo/nbt742\\_S1.html](http://www.nature.com/nbt/journal/v20/n10/supinfo/nbt742_S1.html) (2002).
- 5 Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chemical Science* **6**, 50-69, doi:10.1039/C4SC01534G (2015).
- 6 Xiao, H. & Schultz, P. G. At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harb Perspect Biol* **8**, doi:10.1101/cshperspect.a023945 (2016).
- 7 Chatterjee, A., Guo, J., Lee, H. S. & Schultz, P. G. A Genetically Encoded Fluorescent Probe in Mammalian Cells. *J Am Chem Soc* **135**, 12540-12543, doi:10.1021/ja4059553 (2013).
- 8 Fleissner, M. R. *et al.* Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proceedings of the National Academy of Sciences* **106**, 21637-21642, doi:10.1073/pnas.0912009106 (2009).
- 9 Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat Chem Biol* **11**, 496-503, doi:10.1038/nchembio.1823  
<http://www.nature.com/nchembio/journal/v11/n7/abs/nchembio.1823.html#supplementary-information>  
(2015).

- 10 Liu, C. C., Cellitti, S. E., Geierstanger, B. H. & Schultz, P. G. Efficient expression of tyrosine-sulfated proteins in *E. coli* using an expanded genetic code. *Nature protocols* **4**, 1784-1789, doi:10.1038/nprot.2009.188 (2009).
- 11 Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proceedings of the National Academy of Sciences* **99**, 11020-11024, doi:10.1073/pnas.172226299 (2002).
- 12 Chen, Y. *et al.* Heritable expansion of the genetic code in mouse and zebrafish. *Cell Res*, doi:10.1038/cr.2016.145 (2016).
- 13 Wang, J. *et al.* Palladium-Triggered Chemical Rescue of Intracellular Proteins via Genetically Encoded Allene-Caged Tyrosine. *J Am Chem Soc* **138**, 15118-15121, doi:10.1021/jacs.6b08933 (2016).
- 14 Li, J. & Chen, P. R. Development and application of bond cleavage reactions in bioorthogonal chemistry. *Nat Chem Biol* **12**, 129-137, doi:10.1038/nchembio.2024 (2016).
- 15 Liu, X. *et al.* Significant Expansion of the Fluorescent Protein Chromophore through the Genetic Incorporation of a Metal-Chelating Unnatural Amino Acid. *Angewandte Chemie International Edition* **52**, 4805-4809, doi:10.1002/anie.201301307 (2013).
- 16 Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **114**, 4764-4806, doi:10.1021/cr400355w (2014).
- 17 Shang, X. *et al.* Fluorogenic protein labeling using a genetically encoded unstrained alkene. *Chemical Science* **8**, 1141-1145, doi:10.1039/C6SC03635J (2017).
- 18 Neumann-Staubitz, P. & Neumann, H. The use of unnatural amino acids to study and engineer protein function. *Current Opinion in Structural Biology* **38**, 119-128, doi:<http://dx.doi.org/10.1016/j.sbi.2016.06.006> (2016).
- 19 Ge, Y., Fan, X. & Chen, P. R. A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. *Chemical Science* **7**, 7055-7060, doi:10.1039/C6SC02615J (2016).

- 20 Ko, W., Kim, S., Lee, S., Jo, K. & Lee, H. S. Genetically encoded FRET sensors using a fluorescent unnatural amino acid as a FRET donor. *RSC Advances* **6**, 78661-78668, doi:10.1039/C6RA17375F (2016).
- 21 Grünewald, J. *et al.* Immunochemical termination of self-tolerance. *Proceedings of the National Academy of Sciences* **105**, 11276-11280, doi:10.1073/pnas.0804157105 (2008).
- 22 Grünewald, J. *et al.* Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc Natl Acad Sci U S A* **106**, 4337-4342, doi:10.1073/pnas.0900507106 (2009).
- 23 Gauba, V. *et al.* Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids. *Proc Natl Acad Sci U S A* **108**, 12821-12826, doi:10.1073/pnas.1110042108 (2011).
- 24 Si, L. *et al.* Generation of influenza A viruses as live but replication-incompetent virus vaccines. *Science* **354**, 1170-1173, doi:10.1126/science.aah5869 (2016).
- 25 Wang, N. *et al.* Construction of a Live-Attenuated HIV-1 Vaccine through Genetic Code Expansion. *Angewandte Chemie International Edition* **53**, 4867-4871, doi:10.1002/anie.201402092 (2014).
- 26 Yuan, Z. *et al.* Controlling Multicycle Replication of Live-Attenuated HIV-1 Using an Unnatural Genetic Switch. *ACS Synthetic Biology*, doi:10.1021/acssynbio.6b00373 (2017).
- 27 Rauch, B. J., Porter, J. J., Mehl, R. A. & Perona, J. J. Improved Incorporation of Noncanonical Amino Acids by an Engineered tRNATyr Suppressor. *Biochemistry* **55**, 618-628, doi:10.1021/acs.biochem.5b01185 (2016).
- 28 Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. *Nat Chem Biol* **12**, 776-778, doi:10.1038/nchembio.2160 (2016).
- 29 Elsasser, S. J., Ernst, R. J., Walker, O. S. & Chin, J. W. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature methods* **13**, 158-164, doi:10.1038/nmeth.3701 (2016).
- 30 Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem* **6**, 393-403, doi:10.1038/nchem.1919 (2014).

- 31 Schmied, W. H., Elsasser, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J Am Chem Soc* **136**, 15577-15583, doi:10.1021/ja5069728 (2014).
- 32 Thibodeaux, G. N. *et al.* Transforming a pair of orthogonal tRNA-aminoacyl-tRNA synthetase from Archaea to function in mammalian cells. *PLoS One* **5**, e11263, doi:10.1371/journal.pone.0011263 (2010).
- 33 Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc Natl Acad Sci U S A* **101**, 7566-7571, doi:10.1073/pnas.0401517101 (2004).
- 34 Santoro, S. W., Anderson, J. C., Lakshman, V. & Schultz, P. G. An archaeobacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*. *Nucleic Acids Res* **31**, 6700-6709 (2003).
- 35 Chatterjee, A., Xiao, H., Yang, P. Y., Soundararajan, G. & Schultz, P. G. A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli*. *Angew Chem Int Ed Engl* **52**, 5106-5109, doi:10.1002/anie.201301094 (2013).
- 36 Chin, J. W. *et al.* An Expanded Eukaryotic Genetic Code. *Science* **301**, 964-967, doi:10.1126/science.1084772 (2003).
- 37 Shen, B. *et al.* Genetically Encoding Unnatural Amino Acids in Neural Stem Cells and Optically Reporting Voltage-Sensitive Domain Changes in Differentiated Neurons. *STEM CELLS* **29**, 1231-1240, doi:10.1002/stem.679 (2011).
- 38 Kang, J.-Y. *et al.* In vivo Expression of a Light-activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* **80**, 10.1016/j.neuron.2013.1008.1016, doi:10.1016/j.neuron.2013.08.016 (2013).
- 39 Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biol* **advance online publication**, doi:10.1038/nchembio.2312

<http://www.nature.com/nchembio/journal/vaop/ncurrent/abs/nchembio.2312.html#supplementary-information> (2017).

- 40 Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Meth* **4**, 239-244, doi:[http://www.nature.com/nmeth/journal/v4/n3/supinfo/nmeth1016\\_S1.html](http://www.nature.com/nmeth/journal/v4/n3/supinfo/nmeth1016_S1.html) (2007).
- 41 Deiters, A. & Schultz, P. G. In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorganic & Medicinal Chemistry Letters* **15**, 1521-1524, doi:<http://dx.doi.org/10.1016/j.bmcl.2004.12.065> (2005).
- 42 Miyake-Stoner, S. J. *et al.* Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* **48**, 5953-5962, doi:10.1021/bi900426d (2009).
- 43 Wang, L., Xie, J., Deniz, A. A. & Schultz, P. G. Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *The Journal of Organic Chemistry* **68**, 174-176, doi:10.1021/jo026570u (2003).
- 44 Chin, J. W. *et al.* Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia coli. *J Am Chem Soc* **124**, 9026-9027, doi:10.1021/ja027007w (2002).
- 45 Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences* **100**, 56-61 (2003).
- 46 Brustad, E. *et al.* A Genetically Encoded Boronate Amino Acid. *Angewandte Chemie (International ed. in English)* **47**, 8220-8223, doi:10.1002/anie.200803240 (2008).
- 47 Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat Biotechnol* **24**, 1436-1440, doi:10.1038/nbt1254 (2006).
- 48 Ugwumba, I. N. *et al.* Improving a Natural Enzyme Activity through Incorporation of Unnatural Amino Acids. *J Am Chem Soc* **133**, 326-333, doi:10.1021/ja106416g (2011).



- 49 Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A Genetically Encoded Photocaged Tyrosine. *Angewandte Chemie International Edition* **45**, 2728-2731, doi:10.1002/anie.200600264 (2006).
- 50 Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science* **296**, 1459-1462, doi:10.1126/science.1069588 (2002).
- 51 Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences* **104**, 11268-11273 (2007).
- 52 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* **1844**, 1059-1070, doi:10.1016/j.bbapap.2014.03.002 (2014).
- 53 Mukai, T. *et al.* Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and Biophysical Research Communications* **371**, 818-822, doi:<http://dx.doi.org/10.1016/j.bbrc.2008.04.164> (2008).
- 54 Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem Biol* **8**, 1640-1645, doi:10.1021/cb4001662 (2013).
- 55 Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase-tRNA<sup>Pyl</sup> structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167, doi:[http://www.nature.com/nature/journal/v457/n7233/supinfo/nature07611\\_S1.html](http://www.nature.com/nature/journal/v457/n7233/supinfo/nature07611_S1.html) (2009).
- 56 Wan, W. *et al.* A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angewandte Chemie International Edition* **49**, 3211-3214, doi:10.1002/anie.201000465 (2010).
- 57 Odoi, K. A., Huang, Y., Rezenom, Y. H. & Liu, W. R. Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the Escherichia coli BL21(DE3) cell strain. *PLoS One* **8**, e57035, doi:10.1371/journal.pone.0057035 (2013).

- 58 Sakamoto, K. *et al.* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Research* **30**, 4692-4699, doi:10.1093/nar/gkf589 (2002).
- 59 Ye, S., Huber, T., Vogel, R. & Sakmar, T. P. FTIR analysis of GPCR activation using azido probes. *Nature chemical biology* **5**, 397-399, doi:10.1038/nchembio.167 (2009).
- 60 Ye, S. *et al.* Site-specific Incorporation of Keto Amino Acids into Functional G Protein-coupled Receptors Using Unnatural Amino Acid Mutagenesis. *Journal of Biological Chemistry* **283**, 1525-1533, doi:10.1074/jbc.M707355200 (2008).
- 61 Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **285**, 11039-11044, doi:10.1074/jbc.R109.091306 (2010).
- 62 O'Donoghue, P. *et al.* Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett* **586**, 3931-3937, doi:10.1016/j.febslet.2012.09.033 (2012).
- 63 Cui, Z. *et al.* Combining Sense and Nonsense Codon Reassignment for Site-Selective Protein Modification with Unnatural Amino Acids. *ACS Synth Biol*, doi:10.1021/acssynbio.6b00245 (2016).
- 64 Wang, K., Schmied, W. H. & Chin, J. W. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew Chem Int Ed Engl* **51**, 2288-2297, doi:10.1002/anie.201105016 (2012).
- 65 Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M. & Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *Chembiochem* **15**, 1782-1786, doi:10.1002/cbic.201402104 (2014).
- 66 Wang, N., Shang, X., Cerny, R., Niu, W. & Guo, J. Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria. *Sci Rep* **6**, 21898, doi:10.1038/srep21898 (2016).
- 67 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* **464**, 441-444, doi:10.1038/nature08817 (2010).

- 68 Lajoie, M. J., Söll, D. & Church, G. M. Overcoming Challenges in Engineering the Genetic Code. *Journal of Molecular Biology* **428**, 1004-1021, doi:<http://dx.doi.org/10.1016/j.jmb.2015.09.003> (2016).
- 69 Lajoie, M. J. *et al.* Probing the limits of genetic recoding in essential genes. *Science* **342**, 361-363, doi:10.1126/science.1241460 (2013).
- 70 Mukai, T. *et al.* Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res* **43**, 8111-8122, doi:10.1093/nar/gkv787 (2015).
- 71 Iwane, Y. *et al.* Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat Chem* **8**, 317-325, doi:10.1038/nchem.2446 (2016).
- 72 Zeng, Y., Wang, W. & Liu, W. R. Toward Reassigning the Rare AGG Codon in *Escherichia coli*. *Chembiochem : a European journal of chemical biology* **15**, 1750-1754, doi:10.1002/cbic.201400075 (2014).
- 73 Bohlke, N. & Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *Fems Microbiology Letters* **351**, 133-144, doi:10.1111/1574-6968.12371 (2014).
- 74 Teramoto, H. & Kojima, K. Production of *Bombyx mori* Silk Fibroin Incorporated with Unnatural Amino Acids. *Biomacromolecules* **15**, 2682-2690, doi:10.1021/bm5005349 (2014).
- 75 Axup, J. Y. *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16101-16106, doi:10.1073/pnas.1211023109 (2012).
- 76 Cellitti, S. E. *et al.* In vivo Incorporation of Unnatural Amino Acids to Probe Structure, Dynamics, and Ligand Binding in a Large Protein by Nuclear Magnetic Resonance Spectroscopy. *J Am Chem Soc* **130**, 9268-9281, doi:10.1021/ja801602q (2008).
- 77 Schultz, K. C. *et al.* A Genetically Encoded Infrared Probe. *J Am Chem Soc* **128**, 13984-13985, doi:10.1021/ja0636690 (2006).

- 78 Thielges, M. C. *et al.* Two-Dimensional IR Spectroscopy of Protein Dynamics Using Two Vibrational Labels: A Site-Specific Genetically Encoded Unnatural Amino Acid and an Active Site Ligand. *The Journal of Physical Chemistry B* **115**, 11294-11304, doi:10.1021/jp206986v (2011).
- 79 Fleissner, M. R. *et al.* Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc Natl Acad Sci U S A* **106**, 21637-21642, doi:10.1073/pnas.0912009106 (2009).
- 80 Kalstrup, T. & Blunck, R. Dynamics of internal pore opening in KV channels probed by a fluorescent unnatural amino acid. *Proceedings of the National Academy of Sciences* **110**, 8272-8277 (2013).
- 81 Neumann, H. Rewiring translation – Genetic code expansion and its applications. *FEBS Letters* **586**, 2057-2064, doi:10.1016/j.febslet.2012.02.002 (2012).
- 82 Rakauskaitė, R. *et al.* Biosynthetic selenoproteins with genetically-encoded photocaged selenocysteines. *Chemical Communications* **51**, 8245-8248, doi:10.1039/C4CC07910H (2015).
- 83 Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A Genetically Encoded Photocaged Amino Acid. *J Am Chem Soc* **126**, 14306-14307, doi:10.1021/ja040175z (2004).
- 84 Nguyen, D. P. *et al.* Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J Am Chem Soc* **136**, 2240-2243, doi:10.1021/ja412191m (2014).
- 85 Gautier, A., Deiters, A. & Chin, J. W. Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *J Am Chem Soc* **133**, 2124-2127, doi:10.1021/ja1109979 (2011).
- 86 Chin, J. W. & Schultz, P. G. In vivo photocrosslinking with unnatural amino Acid mutagenesis. *Chembiochem* **3**, 1135-1137, doi:10.1002/1439-7633(20021104)3:11<1135::AID-CBIC1135>3.0.CO;2-M (2002).
- 87 Javahishvili, T. *et al.* Role of tRNA orthogonality in an expanded genetic code. *ACS Chem Biol* **9**, 874-879, doi:10.1021/cb4005172 (2014).

- 88 Maranhao, A. C. & Ellington, A. D. Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. *ACS Synth Biol* **6**, 108-119, doi:10.1021/acssynbio.6b00145 (2017).
- 89 Johnson, D. B. F. *et al.* Release Factor One Is Nonessential in Escherichia coli. *ACS Chemical Biology* **7**, 1337-1344, doi:10.1021/cb300229q (2012).
- 90 Guo, J., Melançon, C. E., Lee, H. S., Groff, D. & Schultz, P. G. Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Unnatural Amino Acid-Containing Proteins. *Angewandte Chemie (International ed. in English)* **48**, 9148-9151, doi:10.1002/anie.200904035 (2009).
- 91 Park, H. S. *et al.* Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **333**, 1151-1154, doi:10.1126/science.1207203 (2011).
- 92 Rackham, O. & Chin, J. W. A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* **1**, 159-166, doi:10.1038/nchembio719 (2005).
- 93 Chen, I. A. & Schindlinger, M. Quadruplet codons: one small step for a ribosome, one giant leap for proteins: an expanded genetic code could address fundamental questions about algorithmic information, biological function, and the origins of life. *Bioessays* **32**, 650-654, doi:10.1002/bies.201000051 (2010).
- 94 Pott, M., Schmidt, M. J. & Summerer, D. Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids. *ACS Chemical Biology* **9**, 2815-2822, doi:10.1021/cb5006273 (2014).
- 95 Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **395**, 361-374, doi:10.1016/j.jmb.2009.10.030 (2010).
- 96 Chatterjee, A., Xiao, H., Bollong, M., Ai, H.-W. & Schultz, P. G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proceedings of the National Academy of Sciences* **110**, 11803-11808, doi:10.1073/pnas.1309584110 (2013).
- 97 Amiram, M. *et al.* Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat Biotech* **33**, 1272-1279, doi:10.1038/nbt.3372

- <http://www.nature.com/nbt/journal/v33/n12/abs/nbt.3372.html#supplementary-information> (2015).
- 98 Barrett, O. P. & Chin, J. W. Evolved orthogonal ribosome purification for in vitro characterization. *Nucleic Acids Res* **38**, 2682-2691, doi:10.1093/nar/gkq120 (2010).
- 99 Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* **7**, 779-786, doi:<http://www.nature.com/nchembio/journal/v7/n11/abs/nchembio.657.html#supplementary-information> (2011).
- 100 Heinemann, I. U. *et al.* Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Letters* **586**, 3716-3722, doi:10.1016/j.febslet.2012.08.031 (2012).
- 101 Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotech* **25**, 770-777, doi:[http://www.nature.com/nbt/journal/v25/n7/supinfo/nbt1314\\_S1.html](http://www.nature.com/nbt/journal/v25/n7/supinfo/nbt1314_S1.html) (2007).
- 102 Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature* **455**, 613-619, doi:10.1038/nature07352 (2008).
- 103 Zheng, Y. *et al.* Broadening the versatility of lentiviral vectors as a tool in nucleic acid research via genetic code expansion. *Nucleic Acids Res* **43**, e73, doi:10.1093/nar/gkv202 (2015).
- 104 Lin, S. *et al.* Site-Specific Engineering of Chemical Functionalities on the Surface of Live Hepatitis D Virus. *Angewandte Chemie International Edition* **52**, 13970-13974, doi:10.1002/anie.201305787 (2013).
- 105 Zhang, C. *et al.* Development of next generation adeno-associated viral vectors capable of selective tropism and efficient gene delivery. *Biomaterials* **80**, 134-145, doi:<http://dx.doi.org/10.1016/j.biomaterials.2015.11.066> (2016).
- 106 Wang, N., Yuan, Z., Niu, W., Li, Q. & Guo, J. Synthetic biology approach for the development of conditionally replicating HIV-1 vaccine. *Journal of Chemical Technology & Biotechnology* **92**, 455-462, doi:10.1002/jctb.5174 (2017).

- 107 Brown, J., Excler, J. L. & Kim, J. H. New prospects for a preventive HIV-1 vaccine. *J Virus Erad* **1**, 78-88 (2015).
- 108 Persaud, D. *et al.* Absence of Detectable HIV-1 Viremia after Treatment Cessation in an Infant. *New England Journal of Medicine* **369**, 1828-1835, doi:doi:10.1056/NEJMoa1302976 (2013).
- 109 Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine* **9**, 727-728, doi:[http://www.nature.com/nm/journal/v9/n6/supinfo/nm880\\_S1.html](http://www.nature.com/nm/journal/v9/n6/supinfo/nm880_S1.html) (2003).
- 110 Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2**, doi:10.1101/cshperspect.a006882 (2012).
- 111 Kijak, G. H. *et al.* Lost in translation: implications of HIV-1 codon usage for immune escape and drug resistance. *AIDS Rev* **6**, 54-60 (2004).
- 112 Jardine, J. G. *et al.* HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* **351**, 1458 (2016).
- 113 Wyand, M. S. *et al.* Protection by live, attenuated simian immunodeficiency virus against heterologous challenge. *J Virol* **73**, 8356-8363 (1999).
- 114 Daniel, M. D., Kirchoff, F., Czajak, S. C., Sehgal, P. K. & Desrosiers, R. C. Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. *Science* **258**, 1938-1941 (1992).
- 115 Ruprecht, R. M. Live attenuated AIDS viruses as vaccines: promise or peril? *Immunol Rev* **170**, 135-149 (1999).
- 116 Desrosiers, R. C. HIV with Multiple Gene Deletions as a Live Attenuated Vaccine for AIDS. *AIDS Research and Human Retroviruses* **8**, 411-421, doi:10.1089/aid.1992.8.411 (1992).
- 117 Koff, W. C. *et al.* HIV vaccine design: insights from live attenuated SIV vaccines. *Nat Immunol* **7**, 19-23 (2006).
- 118 Baba, T. W. *et al.* Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques. *Nature medicine* **5**, 194-203 (1999).

- 119 Hofmann-Lehmann, R. *et al.* Live attenuated, nef-deleted SIV is pathogenic in most adult macaques after prolonged observation. *Aids* **17**, 157-166, doi:10.1097/01.aids.0000042942.55529.01 (2003).
- 120 Baba, T. W. *et al.* Pathogenicity of live, attenuated SIV after mucosal infection of neonatal macaques. *Science* **267**, 1820-1825 (1995).
- 121 Whatmore, A. M. *et al.* Repair and evolution of nef in vivo modulates simian immunodeficiency virus virulence. *J Virol* **69**, 5117-5123 (1995).
- 122 Stahl-Hennig, C. *et al.* Attenuated SIV imparts immunity to challenge with pathogenic spleen-derived SIV but cannot prevent repair of the nef deletion. *Immunol Lett* **51**, 129-135 (1996).
- 123 Berkhout, B., Verhoef, K., van Wamel, J. L. & Back, N. K. Genetic instability of live, attenuated human immunodeficiency virus type 1 vaccine strains. *J Virol* **73**, 1138-1145 (1999).
- 124 Luo, J. *et al.* Genetically encoded optical activation of DNA recombination in human cells  
†Electronic supplementary information (ESI) available: Experimental protocols. See DOI: 10.1039/c6cc03934k Click here for additional data file. *Chemical Communications (Cambridge, England)* **52**, 8529-8532, doi:10.1039/c6cc03934k (2016).
- 125 Chen, P. R. *et al.* A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells. *Angewandte Chemie (International ed. in English)* **48**, 4052-4055, doi:10.1002/anie.200900683 (2009).
- 126 Mayrose, I. *et al.* Synonymous site conservation in the HIV-1 genome. *BMC Evolutionary Biology* **13**, 164, doi:10.1186/1471-2148-13-164 (2013).
- 127 Martin Stoltzfus, C. in *Advances in virus research* Vol. Volume 74 1-40 (Academic Press, 2009).
- 128 Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-Mediated Cell-Surface Labeling. *J Am Chem Soc* **134**, 800-803, doi:10.1021/ja209352s (2012).
- 129 Tsao, M.-L., Tian, F. & Schultz, P. G. Selective Staudinger Modification of Proteins Containing p-Azidophenylalanine. *ChemBioChem* **6**, 2147-2149, doi:10.1002/cbic.200500314 (2005).



- 130 Axup, J. Y. *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proceedings of the National Academy of Sciences* **109**, 16101-16106, doi:10.1073/pnas.1211023109 (2012).
- 131 Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M. & Schultz, P. G. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorganic & Medicinal Chemistry Letters* **14**, 5743-5745, doi:<http://dx.doi.org/10.1016/j.bmcl.2004.09.059> (2004).
- 132 Andersson, H. O. *et al.* Optimization of P1–P3 groups in symmetric and asymmetric HIV-1 protease inhibitors. *European Journal of Biochemistry* **270**, 1746-1758, doi:10.1046/j.1432-1033.2003.03533.x (2003).
- 133 Saad, J. S. *et al.* Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11364-11369, doi:10.1073/pnas.0602818103 (2006).
- 134 Hübner, W. *et al.* Sequence of Human Immunodeficiency Virus Type 1 (HIV-1) Gag Localization and Oligomerization Monitored with Live Confocal Imaging of a Replication-Competent, Fluorescently Tagged HIV-1. *Journal of Virology* **81**, 12596-12607, doi:10.1128/jvi.01088-07 (2007).
- 135 Hübner, W. *et al.* Quantitative 3D Video Microscopy of HIV Transfer Across T Cell Virological Synapses. *Science* **323**, 1743-1747, doi:10.1126/science.1167525 (2009).
- 136 Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. *Nat Chem Biol* **12**, 776-778, doi:10.1038/nchembio.2160  
<http://www.nature.com/nchembio/journal/v12/n10/abs/nchembio.2160.html#supplementary-information>  
(2016).
- 137 Kang, J. Y. *et al.* In vivo expression of a light-activatable potassium channel using unnatural amino acids. *Neuron* **80**, 358-370, doi:10.1016/j.neuron.2013.08.016 (2013).

- 138 Lewis, N., Williams, J., Rekosh, D. & Hammarskjöld, M. L. Identification of a cis-acting element in human immunodeficiency virus type 2 (HIV-2) that is responsive to the HIV-1 rev and human T-cell leukemia virus types I and II rex proteins. *Journal of Virology* **64**, 1690-1697 (1990).
- 139 Frankel, A. D. & Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **55**, 1189-1193, doi:[http://dx.doi.org/10.1016/0092-8674\(88\)90263-2](http://dx.doi.org/10.1016/0092-8674(88)90263-2) (1988).
- 140 Freed, E. O., Myers, D. J. & Risser, R. Mutational analysis of the cleavage sequence of the human immunodeficiency virus type 1 envelope glycoprotein precursor gp160. *Journal of Virology* **63**, 4670-4675 (1989).
- 141 Gao, F. *et al.* Codon Usage Optimization of HIV Type 1 Subtype C gag, pol, env, and nef Genes: In Vitro Expression and Immune Responses in DNA-Vaccinated Mice. *AIDS Research and Human Retroviruses* **19**, 817-823, doi:10.1089/088922203769232610 (2003).
- 142 Nguyen, K.-L. *et al.* Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression. *Virology* **319**, 163-175, doi:<http://dx.doi.org/10.1016/j.virol.2003.11.021> (2004).
- 143 Schaeffer, E., Geleziunas, R. & Greene, W. C. Human Immunodeficiency Virus Type 1 Nef Functions at the Level of Virus Entry by Enhancing Cytoplasmic Delivery of Virions. *Journal of Virology* **75**, 2993-3000, doi:10.1128/jvi.75.6.2993-3000.2001 (2001).
- 144 Raney, A., Kuo, L. S., Baugh, L. L., Foster, J. L. & Garcia, J. V. Reconstitution and Molecular Analysis of an Active Human Immunodeficiency Virus Type 1 Nef/p21-Activated Kinase 2 Complex. *Journal of Virology* **79**, 12732-12741, doi:10.1128/JVI.79.20.12732-12741.2005 (2005).
- 145 Desrosiers, R. C. HIV with multiple gene deletions as a live attenuated vaccine for AIDS. *AIDS Res Hum Retroviruses* **8**, 1457, doi:10.1089/aid.1992.8.1457 (1992).
- 146 Riddle, D. L. & Carbon, J. Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. *Nat New Biol* **242**, 230-234 (1973).
- 147 Yourno, J. & Kohno, T. Externally suppressible proline quadruplet ccc U. *Science* **175**, 650-652 (1972).

- 148 Bossi, L. & Smith, D. M. Suppressor sufJ: a novel type of tRNA mutant that induces translational frameshifting. *Proc Natl Acad Sci U S A* **81**, 6105-6109 (1984).
- 149 Bossi, L., Kohno, T. & Roth, J. R. Genetic characterization of the sufJ frameshift suppressor in *Salmonella typhimurium*. *Genetics* **103**, 31-42 (1983).
- 150 Bossi, L. & Roth, J. R. Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor sufJ. *Cell* **25**, 489-496 (1981).
- 151 Moore, B., Nelson, C. C., Persson, B. C., Gesteland, R. F. & Atkins, J. F. Decoding of tandem quadruplets by adjacent tRNAs with eight-base anticodon loops. *Nucleic Acids Res* **28**, 3615-3624 (2000).
- 152 Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. *J Mol Biol* **307**, 755-769, doi:10.1006/jmbi.2001.4518 (2001).
- 153 Anderson, J. C., Magliery, T. J. & Schultz, P. G. Exploring the limits of codon and anticodon size. *Chem Biol* **9**, 237-244 (2002).
- 154 Walker, S. E. & Fredrick, K. Recognition and positioning of mRNA in the ribosome by tRNAs with expanded anticodons. *J Mol Biol* **360**, 599-609, doi:10.1016/j.jmb.2006.05.006 (2006).
- 155 Taki, M., Hohsaka, T., Murakami, H., Taira, K. & Sisido, M. Position-specific incorporation of a fluorophore-quencher pair into a single streptavidin through orthogonal four-base codon/anticodon pairs. *J Am Chem Soc* **124**, 14586-14590 (2002).
- 156 Hohsaka, T. & Sisido, M. Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology* **6**, 809-815, doi:[http://dx.doi.org/10.1016/S1367-5931\(02\)00376-9](http://dx.doi.org/10.1016/S1367-5931(02)00376-9) (2002).
- 157 Murakami, H., Hohsaka, T., Ashizuka, Y. & Sisido, M. Site-Directed Incorporation of p-Nitrophenylalanine into Streptavidin and Site-to-Site Photoinduced Electron Transfer from a Pyrenyl Group to a Nitrophenyl Group on the Protein Framework. *J Am Chem Soc* **120**, 7520-7529, doi:10.1021/ja971890u (1998).

- 158 Hohsaka, T., Ashizuka, Y., Murakami, H. & Sisido, M. Incorporation of Nonnatural Amino Acids into Streptavidin through In Vitro Frame-Shift Suppression. *J Am Chem Soc* **118**, 9778-9779, doi:10.1021/ja9614225 (1996).
- 159 Moore, B., Persson, B. C., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Quadruplet codons: implications for code expansion and the specification of translation step size1. *Journal of Molecular Biology* **298**, 195-209, doi:<http://dx.doi.org/10.1006/jmbi.2000.3658> (2000).
- 160 Phelps, S. S. *et al.* Translocation of a tRNA with an Extended Anticodon Through the Ribosome. *Journal of Molecular Biology* **360**, 610-622, doi:<http://dx.doi.org/10.1016/j.jmb.2006.05.016> (2006).
- 161 Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of “shifty” four-base codons with a library approach in *Escherichia coli*1. *Journal of Molecular Biology* **307**, 755-769, doi:<http://dx.doi.org/10.1006/jmbi.2001.4518> (2001).
- 162 Herring, S., Ambrogelly, A., Polycarpo, C. R. & Söll, D. Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic Acids Research* **35**, 1270-1278, doi:10.1093/nar/gkl1151 (2007).
- 163 Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase:tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167, doi:10.1038/nature07611 (2009).
- 164 Ai, H.-w., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing Protein–Protein Interactions with a Genetically Encoded Photo-crosslinking Amino Acid. *ChemBioChem* **12**, 1854-1857, doi:10.1002/cbic.201100194 (2011).
- 165 Jager, G., Nilsson, K. & Bjork, G. R. The phenotype of many independently isolated +1 frameshift suppressor mutants supports a pivotal role of the P-site in reading frame maintenance. *PLoS One* **8**, e60246, doi:10.1371/journal.pone.0060246 (2013).
- 166 Farabaugh, P. J. & Björk, G. R. How translational accuracy influences reading frame maintenance. *The EMBO Journal* **18**, 1427-1434, doi:10.1093/emboj/18.6.1427 (1999).

- 167 Wang, L., Brock, A. & Schultz, P. G. Adding 1-3-(2-Naphthyl)alanine to the Genetic Code of *E. coli*. *J Am Chem Soc* **124**, 1836-1837, doi:10.1021/ja012307j (2002).
- 168 Fechter, P., Rudinger-Thirion, J., Tukalo, M. & Giegé, R. Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA<sup>Tyr</sup> are conserved but expressed differently. *European Journal of Biochemistry* **268**, 761-767, doi:10.1046/j.1432-1327.2001.01931.x (2001).
- 169 Kobayashi, T. *et al.* Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Biol* **10**, 425-432, doi:10.1038/nsb934 (2003).
- 170 Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A* **100**, 56-61, doi:10.1073/pnas.0234824100 (2003).
- 171 Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotech* **22**, 1297-1301 (2004).
- 172 Sambrook, J., Russell, D. W. & Sambrook, J. *The condensed protocols from Molecular cloning : a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2006).
- 173 Liu, C. C., Choe, H., Farzan, M., Smider, V. V. & Schultz, P. G. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* **48**, 8891-8898, doi:10.1021/bi9011429 (2009).
- 174 Chen, P. R. *et al.* A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem., Int. Ed.* **48**, 4052-4055, doi:10.1002/anie.200900683 (2009).
- 175 Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4**, 239-244, doi:10.1038/nmeth1016 (2007).
- 176 Chen, P., Hübner, W., Spinelli, M. A. & Chen, B. K. Predominant Mode of Human Immunodeficiency Virus Transfer between T Cells Is Mediated by Sustained Env-Dependent Neutralization-Resistant Virological Synapses. *Journal of Virology* **81**, 12582-12595, doi:10.1128/JVI.00381-07 (2007).

- 177 Cohen, G. B. *et al.* The Selective Downregulation of Class I Major Histocompatibility Complex Proteins by HIV-1 Protects HIV-Infected Cells from NK Cells. *Immunity* **10**, 661-671, doi:[http://dx.doi.org/10.1016/S1074-7613\(00\)80065-5](http://dx.doi.org/10.1016/S1074-7613(00)80065-5) (1999).
- 178 Wang, F. *et al.* RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues. *The Journal of Molecular Diagnostics : JMD* **14**, 22-29, doi:10.1016/j.jmoldx.2011.08.002 (2012).
- 179 Noren, K. A. & Noren, C. J. Construction of high-complexity combinatorial phage display peptide libraries. *Methods* **23**, 169-178, doi:10.1006/meth.2000.1118 (2001).
- 180 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).
- 181 EXPANDING THE GENETIC CODE. *Annual Review of Biophysics and Biomolecular Structure* **35**, 225-249, doi:10.1146/annurev.biophys.35.101105.121507 (2006).