ACS SyntheticBiology
Research Article
Cite This: ACS Synth. Biol. 2018, 7, 1618−1628
pubs.acs.org/synthbio

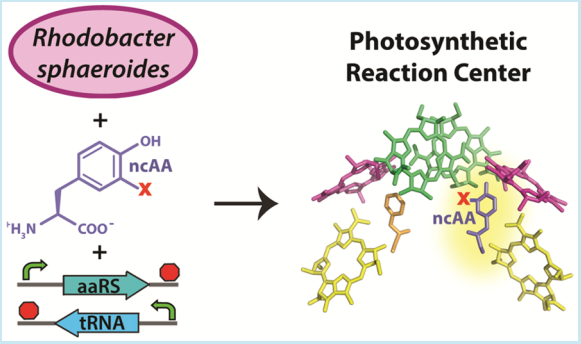
Genetic Code Expansion in *Rhodobacter sphaeroides* to Incorporate Noncanonical Amino Acids into Photosynthetic Reaction Centers

Jared Bryce Weaver and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Supporting Information

ABSTRACT: Photosynthetic reaction centers (RCs) are the membrane proteins responsible for the initial charge separation steps central to photosynthesis. As a complex and spectroscopically complicated membrane protein, the RC (and other associated photosynthetic proteins) would benefit greatly from the insight offered by site-specifically encoded noncanonical amino acids in the form of probes and an increased chemical range in key amino acid analogues. Toward that goal, we developed a method to transfer amber codon suppression machinery developed for *E. coli* into the model bacterium needed to produce RCs, *Rhodobacter sphaeroides*. Plasmids were developed and optimized to incorporate 3-chlorotyrosine, 3-bromotyrosine, and 3-iodotyrosine into RCs. Multiple challenges involving yield and orthogonality were overcome to implement amber suppression in *R. sphaeroides*, providing insights into the hurdles that can be involved in host transfer of amber suppression systems from *E. coli*. In the process of verifying noncanonical amino acid incorporation, characterization of this membrane protein *via* mass spectrometry (which has been difficult previously) was substantially improved. Importantly, the ability to incorporate noncanonical amino acids in *R. sphaeroides* expands research capabilities in the photosynthetic field.



KEYWORDS: *amber suppression, noncanonical amino acids, Rhodobacter sphaeroides, photosynthetic reaction centers, orthogonality*

Photosynthetic reaction centers (RCs) are the membrane proteins responsible for the initial charge separation steps central to photosynthesis.[1,2] They capture solar energy in a highly efficient combination of energy and electron transfer steps. RCs possess two branches of chromophores arranged in nearly identical paths, L and M (or A and B). A pair of excitonically coupled bacteriochlorophylls, known as the special pair (P), absorbs a photon directly or by energy transfer, and the excited state transfers an electron to neighboring chromophores as illustrated in Figure 1. Following photoexcitation, however, electrons are only transferred down the L branch of chromophores. Despite intense effort, the origin(s) of this unidirectional electron transfer is not fully understood, and even the role of intermediary electron acceptors is unclear.[1,3,4]

In this manuscript, we introduce site-specific noncanonical amino acid (ncAA) incorporation as a method to analyze protein design features in this complicated membrane protein complex from *R. sphaeroides*, consisting of three protein subunits (denoted L, M and H) and 9 prosthetic groups (6 chromophores, 2 quinones, a nonheme Fe, and a carotenoid). Conventional site-directed mutagenesis has been extensively applied to perturb photosynthetic charge separation, but to date, tools for introducing ncAAs into the RC have not been described. Their inclusion would allow greater chemical control than the often-large chemical changes which ensue in canonical amino acid mutations. For instance, the tyrosine at M210 (Figure 1) has long been indicated as a key factor for L-branch electron transfer because it is a clear deviation in symmetry between the L and M branch protein environments, and it is the only 1 out of the 28 tyrosines in the RC whose hydroxyl

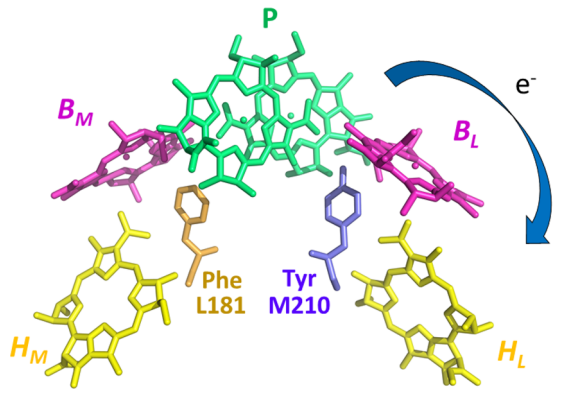


Figure 1. Chromophore arrangement within the RC as determined by the wild-type *R. sphaeroides* crystal structure (Protein Data Bank: 2J8C; quinones and accessory carotenoid not shown for simplicity). Chromophore phytyl side chains are removed for clarity. Blue arrow indicates direction of electron flow along L branch. Tyr at M210, the target of the current work, is labeled in blue and its symmetry-related counterpart, Phe at L181, is shown labeled in gold.

Received: March 7, 2018
Published: May 15, 2018

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

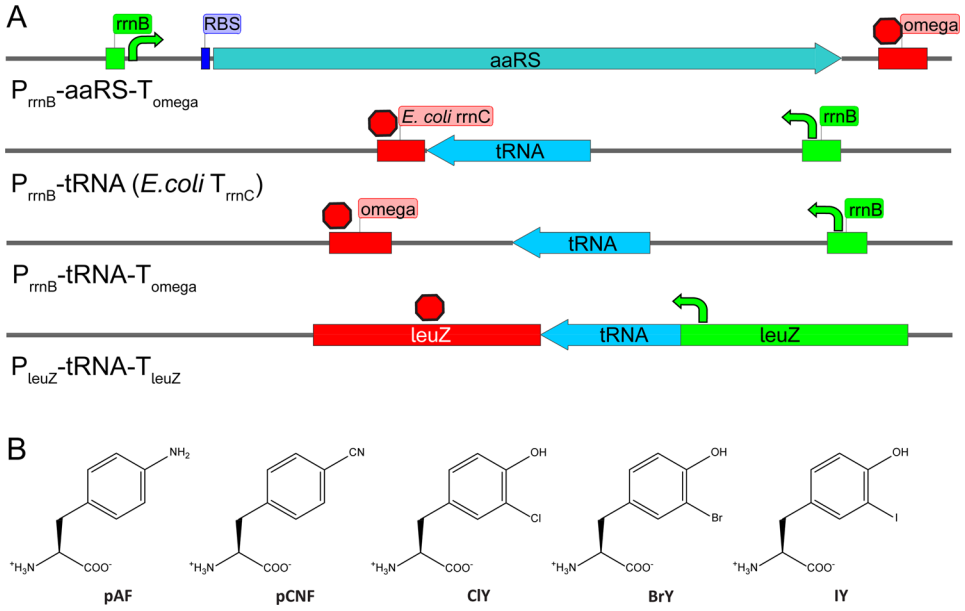


Figure 2. Constructs and ncAAs used in this work. (A) Promoter, RBS, and/or terminator layout of modified tRNA and aaRS gene cassettes (see Figure S1 for an example of a plasmid map containing the aaRS and tRNA gene cassettes). Promoters, RBS, and terminators are either derived from or known to be functional in *R. sphaeroides* unless otherwise stated. The specific aaRS or tRNA contained in the general gene cassettes depicted above varied depending on plasmid construct (see Table 2; *i.e.*, [**8**] contained pAFRS and $tRNA^{Tyr}$ as the aaRS and tRNA in cassettes P_{rrnB}-aaRS-T_{omega} and P_{rrnB}-$tRNA^{Tyr}$ (*E. coli* T_{rrnC})). (B) Different ncAAs discussed in this study, *p*-aminophenylalanine (pAF), *p*-cyanophenylalanine (pCNF), 3-chlorotyrosine (ClY), 3-bromotyrosine (BrY), and 3-iodotyrosine (IY).

group is not hydrogen-bonded.[6] Past simulation results have indicated that the orientation and dipole of the tyrosine hydroxyl play an important role in stabilizing initial charge transfer intermediates.[7] Recent results from our lab demonstrated the impact of introducing a putative hydrogen bond partner near Y(M210), a change which could reorient the tyrosine −OH dipole away from maximally stabilizing interactions for charge transfer indicated by these simulation results.[6] This change dramatically slowed electron transfer from 3−4 to 50 ps. However, by using ncAAs such as halogenated tyrosines, for example, the effect of hydroxyl dipole strength could be tuned more precisely. While the tyrosine at M210 is an important initial target, ncAA incorporation in the form of chemical probes and other amino acid analogues in *R. sphaeroides* could revitalize study at many sites in this important protein, prompt investigation in other associated photosynthetic complexes, and have applications in biotechnology as we gain exquisite control over these photoactive membrane proteins. We will report detailed biophysical and structural characterization of the RC variants produced in this study and others in subsequent manuscripts.

While amber codon suppression has been developed for well over a decade, the range of bacterial hosts for which this technique is available is very limited.[5,8−12] This is due to multiple factors—as in any heterologous gene expression, the aminoacyl tRNA synthetase (aaRS) and tRNA must have transcriptional and translational regulatory control compatible with their host organism.[10] This involves choosing appropriate promoters, ribosome-binding sites (RBSs), and terminators for different aaRS/tRNA pairs, along with other genetic features. The relative strength or ability of these regulatory elements to up/downregulate aaRS and tRNA expression in a new host is also of importance both to avoid negative effects on growth and to obtain sufficient yields of the ncAA-incorporated proteins.[11−14] While these issues can generally be managed by screening different aaRS and tRNA regulatory control features and plasmid vectors, perhaps the greatest challenge is the potential lack of orthogonality of the new aaRS or tRNA in the given host organism. While other types of orthogonality are involved in a successful amber suppressor aaRS/tRNA pair, we will focus on the orthogonality, which is perhaps more changeable and detrimental if lost upon amber suppression host transfer (see SI Section A for discussion). Orthogonality in this case refers to the property by which the introduced aaRS binds only to its cognate amber suppressor tRNA in the presence of the other endogenous host tRNAs. Similarly, an orthogonal tRNA must be recognized only by the aaRS evolved for ncAA incorporation and none of the other endogenous aaRSs. While any given aaRS/tRNA amber suppressor pair is evolved in *E. coli* under positive and negative pressure to be orthogonal to endogenous *E. coli* aaRSs and tRNAs, orthogonality deviations in evolved pairs can still remain following selection or can occur upon transfer to organisms other than *E. coli*.[15−17] They can result in either misincorporation at the amber-encoded-ncAA site of interest or greater stop codon read through at amber stop codon-terminating genomic genes, as nonorthogonal amber suppressor tRNAs are aminoacylated with not just ncAAs but also canonical amino acids.[18−20] Likewise, incorrect amino acid incorporation and consequent increased translation error can occur as nonorthogonal aaRSs aminoacylate endogenous tRNAs with ncAAs or not-encoded canonical amino acids.[15−17] Both issues are detrimental to yield and obtaining desired protein constructs and can result in decreased growth rate and toxicity in less optimal cases. Notwithstanding the challenges in transferring amber suppressor aaRS/tRNA pairs from the organism in which they were originally developed, referred to here as amber suppression host transfer, many proteins cannot be expressed in model hosts like *E. coli*. The RC is an example due to the number of proteins and unique prosthetic groups required for

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

Table 1. Plasmids Used in This Study and Their General Purpose[a]

construct	description	reference
Group 1. Vectors and Original Amber Suppression Plasmids		
[1] pIND4-RC	Midcopy pIND vector, IPTG inducible RC expression	ref. 35
[2] pDule-pAF	Amber suppressor MjTyrRS and $tRNA^{Tyr}$ for pAF	ref. 25
[3] pDule-pCNF	Amber suppressor MjTyrRS and $tRNA^{Tyr}$ for pCNF	ref. 36
[4] pDule-HaloY	Amber suppressor MbTyrRS and $tRNA^{Pyl}$ for ClY, BrY, and IY	Ryan Mehl
Group 2. GFP and *E. coli*-Regulated aaRS/tRNA for pAF		
[5] pBad-sfGFP	GFP under arabinose induction	ref. 26
[6] pBad-sfGFP-150TAG	GFP under arabinose induction with TAG stop at 150	ref. 26
[7] pIND4-RC-pAF0	pIND4-RC with pAFRS/tRNA under original *E. coli* regulatory control	This Study
Group 3. Initial aaRS/tRNA for pAF, Toxic in *R. sphaeroides*		
[8] pIND4-RC-pAF1	pIND4-RC with P_{rrnB}-pAFRS-T_{omega} and P_{rrnB}-$tRNA^{Tyr}$ (*E. coli* T_{rrnC})	This Study
[9] pIND4-RC-pAF2	pIND4-RC with P_{rrnB}-pAFRS-T_{omega} and P_{rrnB}-$tRNA^{Tyr}$-T_{omega}	This Study
[10] pIND4-RC-pAF3	pIND4-RC with P_{rrnB}-pAFRS-T_{omega} and P_{leuZ}-$tRNA^{Tyr}$-T_{leuZ}	This Study
Group 4. GFP inserts in *R. sphaeroides* to Check Regulatory Control Toxicity		
[11] pIND4-RC-GFP1	pIND4-RC with GFP in aaRS cassette,[b] P_{rrnB}-GFP-T_{omega}	This Study
[12] pIND4-RC-GFP2	pIND4-RC with GFP in tRNA cassette,[b] P_{rrnB}-GFP-T_{omega}	This Study
[13] pIND4-RC-GFP3	pIND4-RC with GFP in aaRS and tRNA cassette,[b] P_{rrnB}-GFP-T_{omega}	This Study
Group 5. Single aaRS or tRNA Inserts for Toxicity Identification		
[14] pIND4-RC-pAFtRNA1	pIND4-RC with P_{rrnB}-$tRNA^{Tyr}$ (*E. coli* T_{rrnC})	This Study
[15] pIND4-RC-pAFtRNA2	pIND4-RC with P_{rrnB}-$tRNA^{Tyr}$-T_{omega}	This Study
[16] pIND4-RC-pAFtRNA3	pIND4-RC with P_{leuZ}-$tRNA^{Tyr}$-T_{leuZ}	This Study
[17] pIND4-RC-pAFRS	pIND4-RC with P_{rrnB}-pAFRS-T_{omega}	This Study
Group 6. Toxic pCNF aaRS		
[18] pIND4-RC-pCNFRS	pIND4-RC with P_{rrnB}-pCNFRS-T_{omega}	This Study
Group 7. Functional Nontoxic aaRS/tRNA for HaloY		
[19] pIND4-RC-HaloY1	pIND4-RC with P_{rrnB}-HaloYRS-T_{omega} and P_{rrnB}-$tRNA^{Pyl}$ (*E. coli* T_{rrnC})	This Study
[20] pIND4-RC-HaloY2	pIND4-RC with P_{rrnB}-HaloYRS-T_{omega} and P_{rrnB}-$tRNA^{Pyl}$-T_{omega}	This Study
[21] pIND4-RC-HaloY3	pIND4-RC with P_{rrnB}-HaloYRS-T_{omega} and P_{leuZ}-$tRNA^{Pyl}$-T_{leuZ}	This Study

[a]All constructs containing a pAF or HaloY aaRS had an *R. sphaeroides* RBS before the pAF and HaloY aaRSs except in the case of pIND4-RC-pAF0 [7]. The respective aaRS/tRNA pairs in each construct in this study were derived from pDule-pAF, pDule-pCNF, or pDule-HaloY (denoted pDule1-HaloTyr by Ryan Mehl but referred to as pDule-HaloY throughout this study for brevity in plasmid nomenclature and description). While not explicitly listed, for each pAF or HaloY aaRS/tRNA construct used to screen ncAA incorporation in RCs, a plasmid was also constructed that contained the Y(M210)TAG mutation. See Figure S1 and SI Section B for general plasmid map, sequence, and details on plasmid construction in this study. [b]While identical in promoter and terminator usage, the aaRS and tRNA cassettes differed by orientation of the gene (aaRS and tRNA in opposing orientations) and minor sequence changes for surrounding noncoding linker sequences in each gene cassette. Both were tested separately and together to determine if metabolic stress from two strongly promoted genes due to new regulatory control was detrimental to cell viability and growth.

its assembly and function. The following describes how we transferred a functional, orthogonal aaRS/tRNA pair to *R. sphaeroides*, the challenges we encountered, and how we addressed them, further informing on obstacles generally encountered in aaRS/tRNA host transfer.

RESULTS AND DISCUSSION

Plasmid Construction. To encode site-specific ncAA incorporation in *R. sphaeroides* we initially used the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase and its associated amber suppressor $tRNA_{CUA}$ (MjTyrRS/$tRNA^{Tyr}$).[21] Incorporation of ncAAs with the MjTyrRS/$tRNA^{Tyr}$ pair has been implemented in multiple species, and it is widely used in *E. coli*, which we thought might be closely enough related to the *Rhodobacter* genus (both Gram-negative bacteria). Importantly it has been evolved to incorporate multiple ncAAs, among them several tyrosine and phenylalanine analogues (3-chlorotyrosine, *p*-cyanophenylalanine, *p*-azidophenylalanine, *etc.*).[22−24] The first MjTyrRS/$tRNA^{Tyr}$ pair we implemented was one in which the aaRS had been evolved for *p*-aminophenylalanine (pAF, Figure 2B) incorporation encoded on the plasmid pDule-pAF [2] (Table 1, Group 1).[25,26]

The pIND4-RC plasmid [1], a pMG160 derivative, was chosen as a vector for pAF aaRS (pAFRS)/tRNA constructs (Table 1, Group 1 and Figure S1) due to its higher copy number in both *E. coli* and *R. sphaeroides* (18−23 in pIND4-RC vectors as opposed to 4−7 in commonly used pRK vectors in *R. sphaeroides* cells) making it more amenable to plasmid construction.[27] The pIND4-RC plasmid also contained the H, M, and L genes for RC production under control of the IPTG-inducible *lac* operon which facilitated coordination of RC induction with ncAA addition for amber suppression.[27−29] The pIND4-RC plasmid and the *R. sphaeroides* RCx strain (an electroporatable cell strain with genes RC and light harvesting complexes chromosomally removed) described below were kindly provided by Drs. Thomas Beatty and Daniel Jun.

Initially the *E. coli* promoters, RBS, and terminators associated with the pAFRS/tRNA pair were left unchanged in its insertion into the *R. sphaeroides* vector pIND4-RC. This generated the plasmid pIND4-RC-pAF0 [7], a plasmid which also contained the genes for RC production (Table 1, Group 2). The pIND4-RC vector already contained the genes for the RC H, L, and M subunits, so to test ncAA incorporation, the codon for tyrosine at residue M210 was mutated to the amber stop codon, making a pIND4-RC-pAF0 variant with Y(M210)-

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

Table 2. Amber Suppression Constructs and Their Toxicity

Construct	Transferrable to R. sphaeroides[a]	Transferrable to E. coli
E. coli-regulated MjTyr-derived aaRS/tRNA for pAF incorporation		
pIND4-RC-pAF0[b]	✓	✓
R. sphaeroides-regulated aaRS/tRNA for pAF		
pIND4-RC-pAF1	X	✓
pIND4-RC-pAF2	X	✓
pIND4-RC-pAF3	X	✓
GFP Constructs Testing aaRS and/or tRNA Regulatory Control Toxicity		
pIND4-RC-GFP1	✓	✓
pIND4-RC-GFP2	✓	✓
pIND4-RC-GFP3	✓	✓
Single aaRS or tRNA Inserts for Toxicity Identification in pAF Constructs		
pIND4-RC-pAFtRNA1	✓	✓
pIND4-RC-pAFtRNA2	✓	✓
pIND4-RC-pAFtRNA3	✓	✓
pIND4-RC-pAFRS	X	✓
Toxic MjTyrRS-derived pCNFRS Construct		
pIND4-RC-pCNFRS	X	✓
Non-toxic MbPyl-derived aaRS/tRNA for HaloY Incorporation		
pIND4-RC-HaloY1	✓	✓
pIND4-RC-HaloY2	✓	✓
pIND4-RC-HaloY3	✓	✓

[a]Transferrable, in this context, refers to the ability of a plasmid to produce colonies on agar plates under antibiotic selection following electroporation and outgrowth of transformed cells and is used as a measure of plasmid toxicity, as transformed cells do not produce growth on agar plates in contrast to positive controls attempted simultaneously. Plasmids then that were not transferrable were designated toxic. All constructs that did not yield colonies following transformation had successful positive controls (pIND4-RC transformation) with at least 1.0×10^2 cfu/μg. [b]While pIND4-RC-pAF0 [7] was not toxic in *R. sphaeroides* cells, amber suppression was not functional (Figure 4).

TAG. When the pIND4-RC-pAF0 Y(M210)TAG and wild-type (WT) variants were then compared in side-by-side screens to which pAF was added or withheld, no RC production was observed in Y(M210)TAG containing cells (see experimental details in Amber Suppression Screening below). This indicated aaRS/tRNA expression was absent or impaired due to improper regulatory control.

To ensure translation of the aaRS/tRNA pair, one aaRS gene cassette was generated where the aaRS was placed under control of the strong constitutive P_{rrnB} promoter, RBS, and T_{omega} terminator (P_{rrnB}-pAFRS-T_{omega}). All three had been shown to function in *R. sphaeroides* and were chosen due to their characterized nature and their regulatory strength in promoting high gene expression.[30−33] Due to the incompletely characterized nature of post-transcriptional tRNA processing in *R. sphaeroides*, transcriptional control of the tRNA was varied. Three tRNA gene cassettes were utilized (Figure 2A): one with the *R. sphaeroides* P_{rrnB} promoter and T_{rrnC} terminator originally associated with the amber suppressor $tRNA^{Tyr}$ in *E. coli* plasmid pDule-pAF (P_{rrnB}-$tRNA^{Tyr}$ (*E. coli* T_{rrnC})), one with the *R. sphaeroides* P_{rrnB} promoter and *R. sphaeroides* T_{omega} terminator (P_{rrnB}-$tRNA^{Tyr}$-T_{omega}), and the last with the promoter and terminator found on the strongly transcribed native *R. sphaeroides* tRNA (P_{leuZ}-$tRNA^{Tyr}$-T_{leuZ}).[34] The combination of the pIND4-RC vector, the aaRS gene cassette (Figure 2A), and the three tRNA gene cassettes (Figure 2A), resulted in the construction of three plasmids: pIND4-RC-pAF1 [**8**], pIND4-RC-pAF2 [**9**], and pIND4-RC-pAF3 [**10**] (Table 1, Group 3).

Amber Suppression Toxicity. In our initial attempts to carry out amber suppression in *R. sphaeroides*, we transformed strain RCx with pIND4-RC-pAF1 [**8**], pIND4-RC-pAF2 [**9**], and pIND4-RC-pAF3 [**10**] by electroporation and plated on RLB/agar plates containing the appropriate antibiotic (kanamycin (Kan) for pIND4-based constructs, see Methods). Here RLB is an LB-based media, modified for *R. sphaeroides* growth.[35] Despite repeated plasmid transformation attempts (more than 3 attempts), no cell growth was observed following transformation of Group 3 constructs for pAF incorporation under *R. sphaeroides* transcriptional control. In contrast, positive controls performed in each round of electroporation consisting of pIND4-RC [**1**] transformations consistently produced 1×10^2 to 1×10^3 cfu/μg (in general *R. sphaeroides* competency is low), and negative controls, where no plasmid had been added, consistently produce no *R. sphaeroides* colonies.[37] While it is likely that multiple factors (cell shock, low competency) contributed to the inability to transfer plasmids into viable *R. sphaeroides* cells, given the consistent and unambiguous success of positive controls, it seemed apparent that some attribute of the current constructs was toxic to transformed cells. This toxicity continued when transformed cells were supplemented with 1 mM pAF in RLB/pAF media during the outgrowth step and during antibiotic selection on RLB/Kan/pAF agar plates. To check that the aaRS and tRNA in amber suppression constructs were functional and assembled properly, the amber suppression plasmid with aaRS/tRNA genes under *E. coli* transcriptional control, pIND4-RC-pAF0 [**7**], was tested in *E. coli* for ncAA incorporation. The plasmid pIND4-RC-pAF0 [**7**] was cotransformed with pBad-sfGFP-150TAG [**6**], a reporter plasmid encoding GFP with an inserted stop codon at residue 150, into Thermo DH10-B cells.[26] Site 150 is a surface-exposed residue amenable to mutagenesis; in the absence of ncAA incorporation at position 150, only truncated non-fluorescent peptide is produced. In separate, parallel experiments, a WT pBad-sfGFP [**5**] was cotransformed with pIND4-RC-pAF0 [**7**] as a positive control (similar to WT and Y(M210)TAG screening with RCs). By providing or withholding pAF (1 mM) from cotransformed cells, amber

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

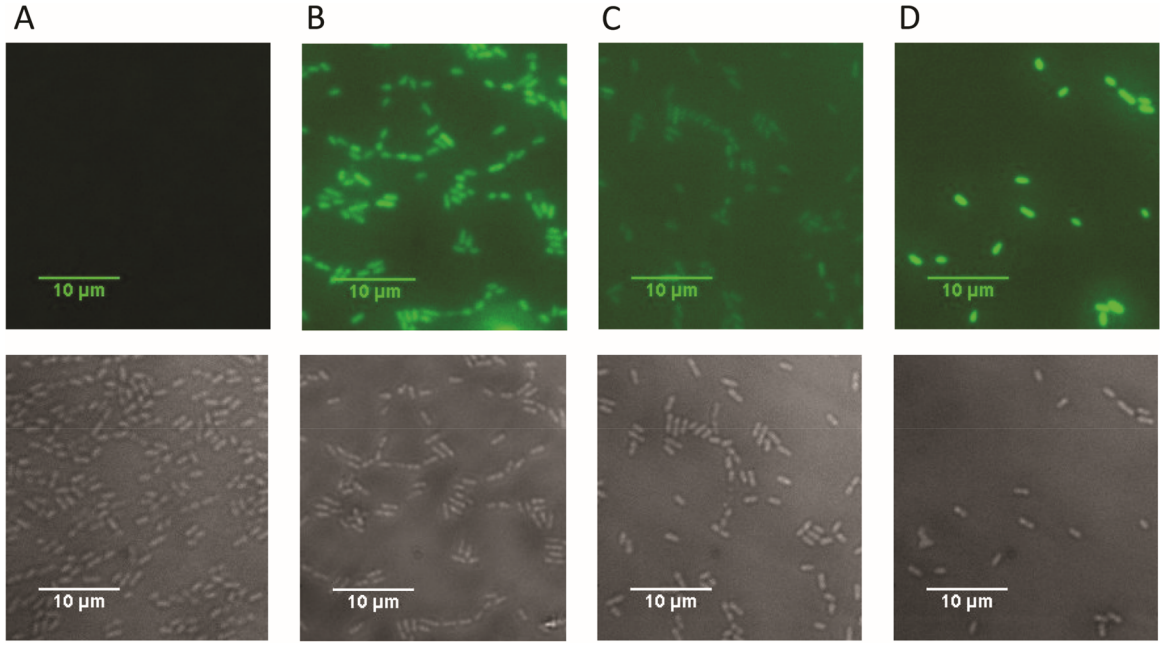


Figure 3. GFP expression in *R. sphaeroides* where GFP is placed in the aaRS and/or tRNA cassette to test promoter/RBS/terminator function. Upper panels are green channel fluorescent images of cell fluorescence, while the lower panels show the respective brightfield images. Images shown are typical and representative of each sample. (A) Negative control, RCx cells with no GFP transformed; (B) GFP in pAFRS cassette, pIND4-RC-GFP1 [**11**]; (C) GFP in tRNA cassette, pIND4-RC-GFP2 [**12**]; (D) GFP in aaRS and tRNA cassettes, pIND4-RC-GFP3 [**13**].

suppression function was verified by GFP fluorescence present only when pAF had been added or in positive controls (Figure S2).

To test the general toxicity of non-aaRS/tRNA sequence features in aaRS/tRNA gene cassettes, namely regulatory elements of the tRNA and aaRS (promoters/RBS for the aaRS/terminators), aaRS and tRNA orientation, and contextual aaRS- or tRNA-specific linker sequences, control plasmids were constructed derived from the toxic plasmid, pIND4-RC-pAF2 [**9**]. This was done to ensure all aspects of the plasmid, even linking sequences and the orientation of the aaRS and tRNA genes (which should both have no function) were nontoxic. Here GFP was inserted in the place of the pAFRS and tRNA genes in aaRS and tRNA gene cassettes respectively. Then one or both cassettes were inserted into pIND4-RC as performed in Group 3 constructs. This was done with the P_{rrnB}-aaRS-T_{omega} cassette to create pIND4-RC-GFP1 [**11**], the P_{rrnB}-tRNA-T_{omega} cassette to create pIND4-RC-GFP2 [**12**], and both the P_{rrnB}-aaRS-T_{omega} and the P_{rrnB}-tRNA-T_{omega} cassettes to create pIND4-RC-GFP3 [**13**] (Table 1, Group 4). While transformation efficiency was lowered (2.0×10^1 cfu/μg to 3.8×10^2 cfu/μg) it was not enough to impede successful plasmid transfer (Table 2). Growth in liquid culture was also typical with final turbidity following induction (OD_{700} = 2.9 ± 0.1) near equivalent to other pIND4-RC construct-containing RCx cultures (OD_{700} = 2.8 ± 0.2; OD_{700} measured due to 600 nm absorbance of photosynthetic chromophores produced by cells).[35] Positive controls likewise produced 1.4×10^3 cfu/μg and negative controls produced no colonies (see Methods). Additionally, all transformed cells displayed constitutive GFP expression (Figure 3) while cells lacking GFP plasmids only exhibited weak autofluorescence.

Given that all other features of inserted aaRS/tRNA gene cassettes (contextual gene linker sequences and regulatory elements) and the pIND4-RC vector backbone were shown to be nontoxic and functional at expressing GFP, and amber suppression was shown to be functional in *E. coli*, we reasoned that the toxicity of pIND4-RC-pAF1 [**8**], pIND4-RC-pAF2 [**9**], and pIND4-RC-pAF3 [**10**] was due to a lack of orthogonality in the aaRS and/or tRNA. To better identify the cause of cell growth inhibition, control plasmids were made where only the aaRS or one of each of the three tRNA gene cassettes had been introduced: pIND4-RC-pAFtRNA1 [**14**], pIND4-RC-pAFtRNA2 [**15**], pIND4-RC-pAFtRNA3 [**16**], and pIND4-RC-pAFRS [**17**] (Table 1, Group 5). While all three transformations of tRNA insert constructs had typical transformation efficiencies (1.1×10^3 to 2.4×10^3 cfu/μg), the construct containing only the aaRS cassette produced no colonies upon plating (Table 2). Again, our positive control had typical transformation efficiency (1.4×10^3 cfu/μg) and our negative control produced no growth. The inability to transfer pIND4-RC-pAFRS [**17**] into *R. sphaeroides* cells led us to conclude that the pAFRS is toxic and therefore not orthogonal in *R. sphaeroides*. Furthermore, it suggests that the toxicity of Group 3 plasmid constructs was due, at least in part, to this lack of orthogonality in the pAFRS (for discussion on the potential mechanism of toxicity see SI Section C).

Due to the apparent toxicity of the aaRS evolved for pAF incorporation, other strategies for reducing toxicity and/or choosing other possible nontoxic aaRS/tRNA pairs were pursued. Initially, the same *R. sphaeroides*-regulated aaRS/tRNA cassettes from three Group 3 plasmids were inserted into the lower copy pRK vector, given past reports that lower-copy plasmids ameliorated toxicity, but pRK construct transformations also failed to yield cell growth (SI Section D).[18] Attempts to place the pAFRS under an inducible promoter, while successful in yielding colonies in the absence of pAFRS induction by IPTG, failed to produce RCs. This was perhaps due to the interplay of identical induction for RC and pAFRS genes (both *lac*-inducible promoters, SI Section E). A different

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

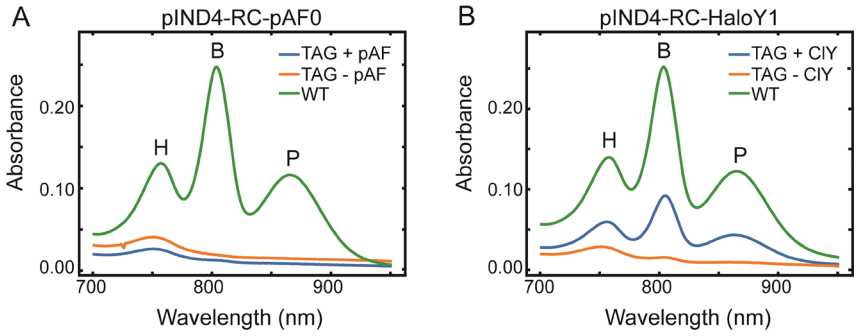


Figure 4. Near-IR unprocessed absorption spectra of a single replicate of cell lysates demonstrating amber suppression screens in *R. sphaeroides* cells transformed with amber suppression plasmids expressing RCs. Band assignments for the H, B, and P chromophores in the RC are indicated above each peak. (A) TAG − pAF and TAG + pAF consist of cell lysate from pIND4-RCM210TAG-pAF0 transformed cells to which pAF was added (+pAF) or withheld (−pAF). WT consists of cell lysate from wild-type RC plasmid pIND4-RC-pAF0. Absorbance in TAG + pAF or TAG − pAF culture is identical and lacks the three NIR peaks characteristic of RC expression, indicating amber suppression is not functional with *E. coli* promoters and terminators. (B) TAG + ClY and TAG − ClY consist of cell lysates from cells transformed with pIND4-RCM210TAG-HaloY1 to which ClY is added (+ClY) or withheld (−ClY). In contrast to (A), TAG + ClY cell lysate displays RC expression, albeit at a somewhat reduced level compared to the WT positive control (cells were lysed at comparable densities), indicating ncAA incorporation. The background peak in the TAG + pAF and TAG − pAF spectra in (A) and in the TAG − ClY in (B) at approximately 750 nm is free bacteriochlorophyll and does not correspond to RC expression.

MjTyrRS/tRNA^{Tyr} pair, which had been developed for *p*-cyanophenylalanine (pCNF) incorporation, was then investigated.[36] To test for toxicity, the pCNF aaRS (pCNFRS) from the plasmid pDule-pCNF [**3**] was inserted into the pIND4-RC [**1**] backbone alone (since the tRNA for pCNF incorporation was the same as that used in pAF incorporation) making pIND4-RC-pCNFRS [**18**] (Table 1, Group 6). Like the MjTyrRS evolved for pAF, however, the pCNFRS also displayed toxicity upon transformation (Table 2).

The toxicity induced by the *M. jannaschii* TyrRSs evolved for pAF and pCNF incorporation was a surprising result. More often in current literature for amber suppressor aaRS/tRNA pairs where a lack of orthogonality has been observed, the tRNA has been focused on as the source of impaired cellular fitness.[18,19,38] Also, the absence of orthogonality in the host transfer of an archaeal-derived aaRS/tRNA pair already evolved in *E. coli*, to another Gram-negative bacterial host, was unexpected and demonstrates the potential for species-specific toxicity in amber suppressor aaRSs.

Though it is possible that with optimization (proper combination of inducible promoters for the pAFRS and RC genes, lower-strength promoters, or different media composition) an MjTyr-derived aaRS/tRNA pair could function in *R. sphaeroides*, other aaRS/tRNA pairs of different origins were pursued. We hypothesized that an aaRS with a different mechanism for substrate (tRNA and/or amino acid) recognition might have a higher chance of success in achieving incorporation of tyrosine derivatives into *R. sphaeroides*. To this end, an aaRS/tRNA pair was examined which had been isolated from another organism and evolved from a different aaRS, the *Methanosarcina barkeri* pyrrolysyl tRNA synthetase (MbPylRS).[39] The particular aaRS/tRNA pair selected was one that has recently been evolved for incorporation of 3-chlorotyrosine (ClY) and other halogenated tyrosines (HaloY) and was generously provided by Dr. Ryan Mehl at Oregon State University (Table 1, plasmid [**4**] see footnote *a*). As with pAF, the MbPylRS/tRNA^{Pyl} pair for HaloY incorporation was inserted into the tRNA and aaRS gene cassettes in the pIND4-RC vector yielding three plasmids: pIND4-RC-HaloY1 [**19**], pIND4-RC-HaloY2 [**20**], and pIND4-RC-HaloY3 [**21**] (Table 1, Group 7). Encouragingly, all constructs displayed transformation efficiency similar to that of pIND4-RC positive controls (1×10^2 to 1×10^3 cfu/μg).

Amber Suppression Screening and Verification. Amber suppression was then tested on RCs for site-specific 3-chlorotyrosine, 3-bromotyrosine (BrY), and 3-iodotyrosine (IY) incorporation as was done with pIND4-RC-pAF0. As with pIND4-RC-pAF0, Y(M210)TAG variants were created for each of the nontoxic HaloY incorporating plasmids: pIND4-RC-HaloY1 [**19**], pIND4-RC-HaloY2 [**20**], and pIND4-RC-HaloY3 [**21**]. These plasmids and their wild-type at M210 counterparts were then compared side-by-side in screens to which halogenated tyrosine was added or withheld to Y(M210)TAG RC variants. Expression of RCs in Y(M210)-TAG variants was compared to RC expression in the corresponding aaRS/tRNA plasmid containing wild-type RC. RC production was monitored by the characteristic near-IR (NIR) RC absorption (Figure 4, Figure S3). For each cell culture to which ncAA had been supplemented to the growth media (1 mM), RC production was observed, while in corresponding negative controls to which no ncAA had been added, no RC production occurred. This contrasted with the pIND4-RC-pAF0 plasmid, where the pAFRS and tRNA^{Tyr} were still under *E. coli* regulatory control. It was likely only transferrable to viable *R. sphaeroides* cells due to the lack of expression of amber suppressor genes due to improper regulatory control, as it did not produce any RCs in identical amber suppression screening upon addition of pAF. For all HaloY ncAAs screened here, none appeared to change final cell turbidity relative to samples where ncAA was withheld, indicating no toxicity was present for any ncAA.

To verify ncAA incorporation, liquid chromatography−mass spectrometry (LC−MS) was performed on isolated RCs to yield an intact subunit mass characterization. As a hydrophobic, three-subunit, ∼100 kDa membrane protein complex, the RC has been difficult to analyze *via* mass spectrometry and has not been completely characterized in the literature, especially at the level of single-Dalton mass accuracy for intact subunits. In this study, LC separation of the three RC subunits and subsequent MS determined subunit mass with 10 ppm mass accuracy (approximately ±1 Da for a 30 kDa subunit, Figures S4−S7). LC−MS spectra supported halogenated tyrosine incorporation

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

with expected 34, 78, and 126 Da mass increases on the M-subunit peak for the ClY, BrY, and IY RC variants, respectively (Figure 5). The peak for the M-subunit showed little to no

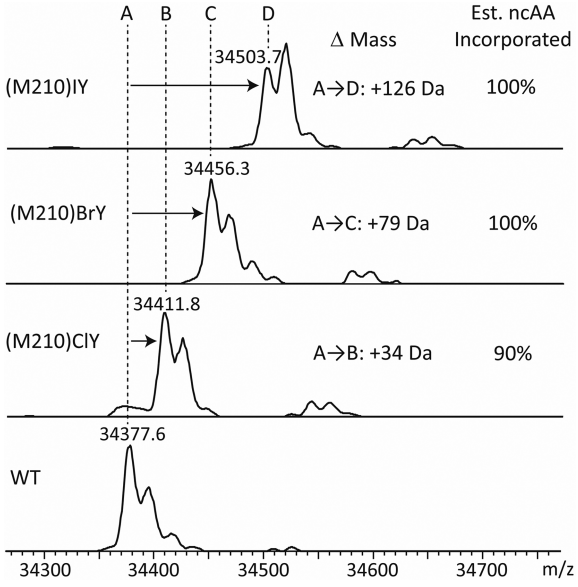


Figure 5. Deconvoluted ESI-MS spectra of the RC M-subunit. Change in mass from WT to ncAA incorporated variants is indicated for each, with the M-subunit major low m/z peak used for mass change determination indicated with a dashed line. Change in mass from A (the WT peak) to each M210 variant (B = ClY, C = BrY, D = IY) is indicated on each. High estimated ncAA incorporation is seen for each ncAA, with only ∼10% WT tyrosine incorporation seen in the (M210)ClY variant mass spectra and no detectable WT RC is seen in other variant mass spectra. Estimations in ncAA incorporation were determined by the ratio of intensity at ncAA m/z peaks to any residual WT m/z peak in each RC variant spectra.[5] Mass change for each variant at the M-subunit matches the change in mass expected for each respective ncAA incorporation, while both the L and H subunits showed no change in mass relative to WT (Table S2).

detectable amount of WT M-subunit from a tyrosine mis-aminoacylation in the variant MS spectra (Figure 5), and the L- and H-subunits show no significant change in mass (Table S2). This supports the subunit-specific nature of ncAA incorporation, the lack of recognition of ncAAs by endogenous *R. sphaeroides* aaRSs, and the high level of specificity of MbPylRS variants for ClY, BrY, or IY over canonical amino acids. Low levels of tyrosine incorporation which were present in ClY variants may have been partly because rich media that was elevated in tyrosine concentrations was employed as opposed to a minimal or more defined media.[40]

Site-specificity of HaloY incorporation at position M210 was further validated by liquid chromatography tandem-mass spectrometry (LC−MS/MS) data. As with the intact subunit MS, sequence coverage for LC−MS/MS experiments performed on enzymatic protein digests has proven quite incomplete in the past. Previous work from our lab in which RCs were solubilized with ProteaseMAX and doubly digested with trypsin and α-chymotrypsin, increased coverage (60% H, 54% M, and 41% L).[6] By using a different set of proteases (AspN and Trypsin/LysC), an elution gradient optimized for more hydrophobic peptide fragments, and a more firmly packed and smaller 0.2 μm pore nano LC column, sequence coverage was greatly enhanced (100% H, 100% M, and 88% L). In these experiments, the detergent used for RC isolation, dimethyldodecylamine N-oxide (LDAO, Sigma), did not overly suppress ionization efficiency. This allowed us to skip protein precipitation, which normally must be performed to remove detergent, and this likely increased signal to allow for greater protein coverage. LC−MS/MS results verified that ClY incorporation was site-specific and essentially complete at M210 (Figure S8). Nonspecific ClY incorporation into the RC (detected *via* CID fragmentation) occurred at only 4 of the potential 27 off-target sites and was less than 10% (within experimental error) in each case (SI Section F, Tables S4 and S5).

The M210 tyrosine chemical variants produced only minor changes in RC UV−vis−NIR absorption spectrum (Figure 6),

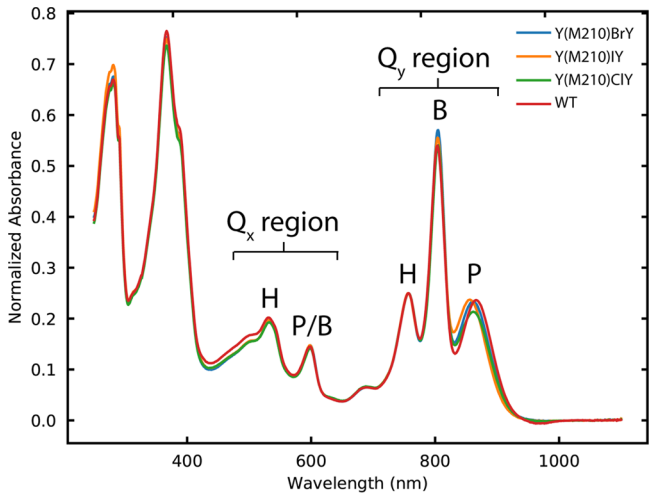


Figure 6. Absorbance spectra of purified RCs containing ClY, BrY, or IY at M210. Absorbance spectra were normalized to the absorbance maxima of the Q_Y H-band (758 nm).

relative to wild-type RC. This is in line with past results in the literature where perturbative mutations (ex. tyrosine mutated to tryptophan) caused only small (<10 nm) shifts in the NIR (or Q_y) region.[41,42] Further structural and electron transfer kinetic characterization of these RC variants is ongoing.

Amber Suppression Efficiency Comparison. With a functional and nontoxic aaRS in hand (MbPylRS evolved for HaloY), we returned to the question of which regulatory elements could optimize the expression of this synthetase/tRNA in *R. sphaeroides*, and as a result, the efficiency of amber suppression. To do this, we screened the amber suppression plasmids mentioned earlier, pIND4-RC-HaloY1 [**19**], pIND4-RC-HaloY2 [**20**], and pIND4-RC-HaloY3 [**21**]. The ability of the three different constructs to express RCs with ClY at M210 was assessed by comparing RC screens performed as listed below. For each plasmid two samples contained the Y(M210)-TAG mutation and the third was wild-type at M210. Again, by testing protein expression in the presence and absence of 1 mM ncAA, amber suppression efficiency was assayed. Experiments were performed in triplicate, and, using NIR RC absorbance on detergent-solubilized RCs in cell lysates, crude yields were measured for each amber suppression plasmid where tRNA regulatory elements had been varied. A significant difference in amber suppression efficiency was observed. The most efficient construct, pIND4-RC-HaloY1 [**19**], had a yield (2.6 mg/L or 30 ± 2% WT yield), over an order of magnitude higher than the least efficient (pIND4-RC-HaloY2 [**20**] produced 0.16 mg/

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

L or 1.8 ± 0.5% WT yield, Figure S9 and Figure S3). Surprisingly, the construct in which the tRNA had been placed under the same regulatory control as that of a native highly expressed tRNA, produced 5 times less RC than the most efficient construct (0.5 mg/L or 11 ± 2% WT yield). For the other variants, pIND4-RC-HaloY1 (with TAG at M210) produced RCs containing BrY at M210 with similar yields (2 mg/L) as ClY variants, while IY containing variants were produced with less than half that yield (0.7−0.9 mg/L). Even for the lowest yielding RC variant, IY at M210, yields were tractable for transient absorption, X-ray crystallography, and other biophysical characterization of RC variants under typical RC production conditions (3−6 L cultures).

In conclusion, we report here the first example of site-specific ncAA incorporation in *R. sphaeroides via* amber suppression. The difficulty initially faced in transferring the MjTyrRS/$tRNA^{Tyr}$ pair was unexpected, given the pair was derived from archaea and then evolved to be orthogonal in another Gram-negative bacterium (*E. coli*). A potential explanation is that organisms in the same domain (bacteria) may have distinct tRNA pools so an aaRS that is orthogonal in one organism may not be in another (SI Section C). Through use of the more othogonal *M. barkeri* PylRS/$tRNA_{CUA}$ pair, these problems with toxicity were overcome. The MbPylRS/$tRNA^{Pyl}$ pair has been found to be orthogonal across a wide range of organisms, a point attributed to the exotic sequence features of the Mb $tRNA^{Pyl}$, such as its unusually short variable arm, and as found in recent structural studies, the N- and C-terminal domains of its cognate aaRS which possess more interfacial surface area than any of the known 20 canonical aaRSs.[43,44] It is possible the combination of these features allows pyrrolysl-tRNA synthetases to recognize features which are truly distinct and not likely to be found on tRNAs from many organisms, and in our case *R. sphaeroides* tRNAs, whereas the orthogonality of MjTyrRS to *E. coli* tRNAs may be more incidental.[43] Regulation of tRNA transcription seen in the three combinations of promoter/terminators used here had a major influence on amber suppression efficiency, with over an order of magnitude difference in yields. These improvements highlight the importance of optimizing tRNA regulation.

In the process of verification of ncAA incorporation, LC−MS and LC−MS/MS characterization of this membrane complex was dramatically enhanced. These methods provide valuable tools for additional ncAA incorporation and general protein engineering verification of this important photoactive protein. Introduction of ClY, BrY, and IY into M210 in the RC has shown minimal perturbation in the absorption spectrum, and studies of electron transfer kinetics are underway to test the role of tyrosine M210.

METHODS

Plasmid Construction and Cell Strains. All plasmids were constructed through use of the NEBuilder HiFi DNA Assembly Master Mix. All aaRS and tRNA gene cassettes were ordered as G-blocks from IDT, along with GFP inserts flanked by aaRS or tRNA gene cassettes. The pIND4-RC vector was linearized with the Kpn2I restriction enzyme and G-blocks for aaRS, tRNA, and GFP inserts were ordered with necessary homologous sequences for vector/insert assembly. All aaRSs and tRNAs were designed based on sequence information from pDule-pAF, pDule-pCNF, and pDule-HaloY (see SI Section B). *E. coli*-derived promoters and terminators used in this paper were also derived from promoters and terminators originally located on the pDule-pAF and pDule-HaloY aaRS/tRNA genes developed for *E. coli*. Plasmids were transformed into NEB 5-alpha Competent *E. coli* cells. *E. coli* strains were cultured at 37 C and 225 rpm in LB broth. Mutagenesis to insert the amber stop codon at M210 was performed with the QuikChange Lightning Site-Directed Mutagenesis Kit on RC-containing pIND4-RC vectors prior to Hifi assemblies. All plasmids generated in this study and mutations made were verified by DNA sequencing (see SI for further detail on vector and insert sequences and plasmid construction).

Plasmids were transformed into the *R. sphaeroides* rshI-, LHI-, LHII-, and RC- (H, L, and M) deletion strain, RCx; here LHI and LHII are light harvesting complexes I and II. Due to the chromosomal deletion of the rshI gene from the RCx strain produced by the Beatty lab (and in a similar deletion strain produced by Philip Laible and Deborah Hansen) plasmids could be transferred *via* electroporation, greatly facilitating plasmid transfer.[35,45] Electroporation cell preparation was similar to previously published literature: 2 × 80 mL RCx cells were grown in semiaerobic growth in 2 × 125 mL Erlenmeyer flask until at early to mid log phase growth (OD600 = 0.4) at 33 C.[35,37,45] Cells were then pelleted at 4600*g* and resuspended at 4 °C three times with 2 × 40 mL, 2 × 20 mL, and 1 × 20 mL of ice cold sterile Millipore water. Cells were then resuspended in 10 mL of 10% glycerol, pelleted again at 4600*g* and finally resuspended in 1 mL 10% glycerol. Aliquots of 40 μL were electroporated with 100 ng of plasmid at 2500 V, 25 μF, and 200 ohms, and then 900 μL of RLB media (see below) was added. Cells were then allowed to recover by culturing at 35 C and 150 rpm for 4 h and then plated on RLB agar plates supplemented with 25 μg/mL Kan. In each transformation testing toxicity of plasmid constructs (Table 2), pIND4-RC [**1**] plasmids were transformed as positive controls (every plasmid constructed in this study had a pIND4-RC vector backbone). Each transformation testing toxicity also had a negative control where plasmid was withheld and electroporated cells were allowed to recover and plated as in other construct and positive control transformations. In each transformation of plasmids containing a pAFRS (Table 1, Group 3−5 Plasmids), pAF was not added to RLB/Kan agar plates. Transformations repeated on Group 3 plasmids where pAF was supplemented to RLB/Kan agar plates also had identical results, demonstrating pAF addition did not alleviate toxicity.

For liquid culture, *R. sphaeroides* cells were grown at 30 C and 200 rpm in RLB, LB broth which has been supplemented with 810 μM $MgCl_2$ and 510 μM $CaCl_2$.[35] Kanamycin selection concentrations were 25 μg/mL for *R. sphaeroides* and 30 μg/μL for *E. coli*.

Amber Suppression Screening and Efficiency. Amber suppression function was verified by the presence of RC expression. *R. sphaeroides* cell cultures were made in groups of 3 × 50 mL in 250 mL Erlenmeyer flasks to allow for aerobic growth. In two of the 50 mL cultures the plasmid of interest (possessing aaRS and tRNA gene cassettes) contained the amber codon at M210, while the third 50 mL culture was wild-type at M210. Normally aerobic growth conditions for production of RCs are avoided because the RC is generally under control of the *puf* or *puc* promoter and the presence of O_2 down-regulates genes under *puf* or *puc* control.[35] RC induction is then produced by naturally decreasing oxygen concentrations as cell turbidity increases during cell culture. The RCx deletion strain and the pIND4-RC plasmid developed

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

by Jun *et al.* and others, remove O_2 suppression and allowed the RC genes to be placed under control of the IPTG-inducible *lac* operon.[35] This feature not only allowed aerobic culture, which sped up cell growth rate, but was also helpful in allowing RC production to be coordinated with addition of the ncAA (ClY, BrY, IY) for initiating amber suppression.

RC expression was induced by addition of 50 μL of 1 M IPTG once cell culture reached early log phase (OD600 = 0.2) to achieve a final concentration of 1 mM IPTG in cell media. At the same time 0.5 mL of 100 mM of the ncAA of interest was added, taking the final ncAA final concentration to 1 mM. While ClY and BrY were solubilized to 100 mM in water (with the help of gentle heating below 55 C for less than an hour and sonication), IY required the addition of 160 μL of 8 M NaOH for 10 mL of 100 mM IY to fully solubilize similar to Mehl *et al.* 2007.[46] Induction was allowed to proceed for 36 h at 30 C and 200 rpm.

Cells were pelleted and resuspended in 5 mL Lysis Buffer (10 mM Tris, 150 mM NaCl) then frozen in liquid nitrogen. To proceed with the preparation, cells were thawed, minimal DNase was added (several crystals) and for each gram of cells (wet mass), 1 μL of 1 M $MgCl_2$ was added. Thawed cells were then lysed with a Fisher Scientific CL-334 probe sonicator by two rounds of 8 × 10 s pulses with 10 s rest on ice between pulses and 10−15 min on ice between each round.

The insoluble phase (cell debris) was collected by centrifuging the lysates for 20 min at 12 100*g* and 4 °C and the remaining liquid lysate decanted. LDAO was added from an approximately 30% stock until the liquid lysate was to a final LDAO concentration of 1% to solubilize RC chromatophores.[35] Membrane debris was removed from this suspension by spinning the cell lysate at 26 100*g* at 4 °C and collecting the liquid phase. Solubilized chromatophores were then decanted and UV−vis−NIR spectra acquired on a PerkinElmer Lambda-25 UV−vis spectrometer at a scanning rate of 480 nm/min.

Yields of RCs in amber suppression screens testing tRNA transcriptional control (testing plasmids pIND4-RC-HaloY1 [**19**], pIND4-RC-HaloY2 [**20**], and pIND4-RC-HaloY3 [**21**] with TAG at M210) were calculated by first taking the absorbance spectra for TAG + ClY cell lysate and then subtracting off TAG − ClY lysate to remove low levels of cell lysate background. The resulting spectra were then baselined by uniformly subtracting off average absorbance of the 1000 to 1100 nm region. Absorbances of the B-band maxima at ∼804 nm were then used to measure RC yield, where the molar absorptivity of RCs at 804 nm is 288 000 M^{-1} cm^{-1} and the molecular weight is approximately 94.6 kDa based off mass-spectra and protein sequence.[47] Yields for RCs with BrY or IY incorporated (using pIND4-RC-HaloY1 [**19**] with TAG at M210) were calculated similarly.

Amber suppression of pIND4-RC-pAF0 performed in *E. coli* was performed according to literature protocol.[46] Specifically, GFP- and Amber suppression plasmids (pBad-sfGFP [**5**] and pIND4-RC-pAF0 [**7**] or pBad-sfGFP-150TAG [**6**] and pIND4-RC-pAF0 [**7**]) were cotransformed into DH10B cells and grown in autoinduction media. Subculturing and autoinduction followed autoinduction protocol following Hammill *et al.* and GFP was allowed to grow and induce for 24 h following subculture step.[46]

Mass Spectrometry. LC−MS to verify incorporation of ClY, BrY and IY into the RC M-subunit was performed on an Agilent Pursuit 5 diphenyl 150 × 2.1 mm column in an Agilent 1260 HPLC and Bruker MicroTOF-QII.[48] Temperature on the column was 80 C, the flow rate was 0.3 mL/min and the injection volume was 5 μL. Samples were injected at a concentration of 3.5 μM RC in 10 mM Tris, 0.03% LDAO, pH = 8.0 buffer. To elute peptides a gradient containing 0.1% formic acid was run from water (with 0.05% trifluoroacetic acid) to acetonitrile (both also contained formic acid, Table S1). Data was collected in full scan MS mode with a mass scan range of 400−4000 Da. Samples were ionized with ESI and the collision RF setting was 800 Vpp.

LC−MS/MS was used to verify site-specificity of ClY incorporation at M210 in doubly enzymatically digested RC samples. Sample containing 20 μg of RCs in 4.3 μL of 0.03% LDAO, 10 mM Tris, pH = 8.0 was taken to 50 μL with 50 mM ammonium bicarbonate. The sample was then reduced by taking the solution to 10 mM dithiothreitol for 30 min at 55C and reduced cysteines capped by addition of acrylamide to 30 mM overall concentration and incubation for 30 min at room temperature. After reduction and alkylation proteins were digested with AspN for 2 h at 37 C followed by Trypsin/LysC overnight at 37 C in the presence of 0.02% ProteaseMax (Promega). Formic acid was then added to a final concentration of 2.5% to terminate digestion. Samples were then analyzed by a Thermo Scientific Orbitrap Fusion with a nanoLC column. Peptides were eluted in an 80 min gradient and were ionized by ESI (Table S3). Fragmentation for MS2 spectra was obtained by three different methods, electron-transfer dissociation (ETD), higher-energy collision-induced dissociation (HCD), and collision-induced dissociation (CID) to maximize sequence coverage.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00100.

Orthogonality discussions, Plasmid map and sequence information, lower plasmid copy and IPTG-induction attempts to decrease pAFRS toxicity, amber suppression in *E. coli*, mass spectrometry spectra and experimental details, tRNA transcriptional regulation effect on yield and NIR absorption screens (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sboxer@stanford.edu.

ORCID

Jared Bryce Weaver: 0000-0002-3823-422X

Steven G. Boxer: 0000-0001-9167-4286

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. Thomas Beatty and Daniel Jun at the University of British Columbia for their generous gifts of the plasmid pIND4-RC and the RCx strain, along with many helpful discussions. We likewise thank Dr. Ryan Mehl, Rachel Henson, and Dr. Joseph Porter at Oregon State University for the maps and sequence information on the pDule-pAF, pDule-pCNF, and pDule-HaloY, many helpful discussions and their generous provision of pDule-HaloY prior to publication. We thank Dr. Stephen Fried for his advice on trouble-shooting toxicity in amber suppression plasmid constructs and critical reading of

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

the manuscript, and Drs. Philip Laible and Deborah Hanson of the Argonne National Laboratory for their help in working with *R. sphaeroides* and assessing early challenges in this study. We also thank Theresa McLaughlin, Dr. Ryan Leib, Dr. Christopher Adams, and the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry for all their support performing the LC−MS and LC−MS/MS in this study. This work is supported by a grant from the NSF Biophysics Program (MCB-1408785).

■ ABBREVIATIONS

RC, reaction center; P, special pair; ncAA, noncanonical amino acid; aaRS, aminoacyl tRNA synthetase; MjTyrRS, *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase; pAF, *p*-aminophenylalanine; RBS, ribosome binding site; Kan, kanamycin; WT, wild-type; IPTG, isopropyl *β*-D-1-thiogalactopyranoside; pCNF, *p*-cyanophenylalanine; MbPylRS, *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase; pAFRS, *p*-aminophenylalanyl-tRNA synthetase; pCNFRS, *p*-cyanophenylalanyl-tRNA synthetase; ClY, 3-chlorotyrosine; BrY, 3-bromotyrosine; IY, 3-iodotyrosine; HaloYRS, halotyrosyl-tRNA synthetase; LC−MS, liquid chromatography−mass spectrometry; LC−MS/MS, liquid chromatography−tandem mass spectrometry; LDAO, dimethyldodecylamine N-oxide; ETD, electron-transfer dissociation; HCD, higher-energy collision-induced dissociation; CID, collision-induced dissociation; LHI, light harvesting complex I; LHII, light harvesting complex II

■ REFERENCES

(1) Parson, W., and Warshel, A. (2009) Mechanism of Charge Separation in Purple Bacterial Reaction Centers, In *The Purple Phototrophic Bacteria* (Hunter, N., Daldal, F., Thurnauer, M. C., and Beatty, J. T., Eds.), pp 356−357, Springer, Dordrecht.

(2) Blankenship, R. (2014) *Molecular Mechanism of Photosynthesis*, 2nd ed., Wiley-Blackwell, New York, NY.

(3) Breton, J., Martin, J.-L., Lambry, J.-C., Robles, S. J., and Youvan, D. C. (1990) Ground State and Femtosecond Transient Absorption Spectroscopy of a Mutant of Rhodobacter Capsulatus Which Lacks the Initial Electron Acceptor Bacteriopheophyrin, In *Structure and Function of Bacterial Photosynthetic Reaction Centers* (Michel-Beyerle, M. E., Ed.), pp 293−302, Springer-Verlag, New York.

(4) Carter, B., Boxer, S. G., Holten, D., and Kirmaier, C. (2009) Trapping the P+B(L)- initial intermediate state of charge separation in photosynthetic reaction centers from Rhodobacter capsulatus. *Biochemistry 48*, 2571−2573.

(5) He, J., Van Treeck, B., Nguyen, H. B., and Melancon, C. E., 3rd (2016) Development of an Unnatural Amino Acid Incorporation System in the Actinobacterial Natural Product Producer Streptomyces venezuelae ATCC 15439. *ACS Synth. Biol. 5*, 125−132.

(6) Saggu, M., Carter, B., Zhou, X., Faries, K., Cegelski, L., Holten, D., Boxer, S. G., and Kirmaier, C. (2014) Putative hydrogen bond to tyrosine M208 in photosynthetic reaction centers from Rhodobacter capsulatus significantly slows primary charge separation. *J. Phys. Chem. B 118*, 6721−6732.

(7) Alden, R. G., Parson, W. W., Chu, Z. T., and Warshel, A. (1996) Orientation of the OH Dipole of Tyrosine (M)210 and Its Effect on Electrostatic Energies in Photosynthetic Bacterial Reaction Centers. *J. Phys. Chem. 100*, 16761−16770.

(8) Wang, F., Robbins, S., Guo, J., Shen, W., and Schultz, P. G. (2010) Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. *PLoS One 5*, e9354.

(9) Chemla, Y., Friedman, M., Heltberg, M., Bakhrat, A., Nagar, E., Schwarz, R., Jensen, M. H., and Alfonta, L. (2017) Expanding the Genetic Code of a Photoautotrophic Organism. *Biochemistry 56*, 2161−2165.

(10) Lin, S., Zhang, Z., Xu, H., Li, L., Chen, S., Li, J., Hao, Z., and Chen, P. R. (2011) Site-specific incorporation of photo-cross-linker and bioorthogonal amino acids into enteric bacterial pathogens. *J. Am. Chem. Soc. 133*, 20581−20587.

(11) Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., Reed, S. A., Yang, P. Y., Wang, R. E., Javahishvili, T., Schultz, P. G., and Young, T. S. (2016) Recombinant thiopeptides containing non-canonical amino acids. *Proc. Natl. Acad. Sci. U. S. A. 113*, 3615−3620.

(12) Gan, Q., Lehman, B. P., Bobik, T. A., and Fan, C. (2016) Expanding the genetic code of Salmonella with non-canonical amino acids. *Sci. Rep. 6*, 39920.

(13) Schmied, W. H., Elsasser, S. J., Uttamapinant, C., and Chin, J. W. (2014) 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.

(14) Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010) Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting. *J. Mol. Biol. 404*, 70−87.

(15) Javahishvili, T., Manibusan, A., Srinagesh, S., Lee, D., Ensari, S., Shimazu, M., and Schultz, P. G. (2014) Role of tRNA orthogonality in an expanded genetic code. *ACS Chem. Biol. 9*, 874−879.

(16) Chin, J. W., Cropp, T. A., Chu, S., Meggers, E., and Schultz, P. G. (2003) Progress Toward an Expanded Eukaryotic Genetic Code. *Chem. Biol. 10*, 511−519.

(17) Lesjak, S., Boko, D., and Weygand-Durasevi, I. (2010) Seryl-tRNA synthetases from methanogenic archaea: suppression of bacterial amber mutation and heterologous toxicity. *Food Technol. Biotechnol. 48*, 512−518.

(18) Tian, H., Deng, D., Huang, J., Yao, D., Xu, X., and Gao, X. (2013) Screening system for orthogonal suppressor tRNAs based on the species-specific toxicity of suppressor tRNAs. *Biochimie 95*, 881−888.

(19) Reynolds, N. M., Vargas-Rodriguez, O., Soll, D., and Crnkovic, A. (2017) The central role of tRNA in genetic code expansion. *Biochim. Biophys. Acta, Gen. Subj. 1861*, 3001−3008.

(20) Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods 4*, 239−244.

(21) Wang, L., Magliery, T. J., Liu, D. R., and Schultz, P. G. (2000) 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.

(22) Liu, X., Jiang, L., Li, J., Wang, L., Yu, Y., Zhou, Q., Lv, X., Gong, W., Lu, Y., and Wang, J. (2014) Significant expansion of fluorescent protein sensing ability through the genetic incorporation of superior photo-induced electron-transfer quenchers. *J. Am. Chem. Soc. 136*, 13094−13097.

(23) Schultz, K. C., Superkova, L., Ryu, Y., Xie, W., Perera, R., and Schultz, P. G. (2006) A Genetically Encoded Infrared Probe. *J. Am. Chem. Soc. 128*, 13984−13985.

(24) Young, D. D., Jockush, S., Turro, N. J., and Schultz, P. G. (2011) Synthetase polyspecificity as a tool to modulate protein function. *Bioorg. Med. Chem. Lett. 21*, 7502−7504.

(25) Mehl, R. A., Anderson, J. C., Santoro, S. W., Wang, L., Martin, A. B., King, D. S., Horn, D. M., and Schultz, P. G. (2003) Generation of a Bacterium with a 21 Amino Acid Genetic Code. *J. Am. Chem. Soc. 125*, 935−939.

(26) Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., and Mehl, R. A. (2010) Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry 49*, 1667−1677.

(27) Inui, M., Nakata, K., Roh, J. H., Vertes, A. A., and Yukawa, H. (2003) Isolation and Molecular Characterization of pMG160, a Mobilizable Cryptic Plasmid from Rhodobacter blasticus. *Appl. Environ. Microbiol. 69*, 725−733.

(28) Ind, A. C., Porter, S. L., Brown, M. T., Byles, E. D., de Beyer, J. A., Godfrey, S. A., and Armitage, J. P. (2009) Inducible-expression

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

plasmid for Rhodobacter sphaeroides and Paracoccus denitrificans. *Appl. Environ. Microbiol. 75*, 6613−6615.
(29) Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta, Proteins Proteomics 1844*, 1059−1070.
(30) Dryden, S. C., and Kaplan, S. (1993) Identification of cis-Acting Regulatory Regions Upstream of the rRNA Operons of Rhodobacter sphaeroides. *J. Bacteriol. 175*, 6392−6402.
(31) Dryden, S. C., and Kaplan, S. (1990) Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. *Nucleic Acids Res. 18*, 7267−7277.
(32) Lu, W., Ye, L., Xu, H., Xie, W., Gu, J., and Yu, H. (2014) Enhanced production of coenzyme Q10 by self-regulating the engineered MEP pathway in Rhodobacter sphaeroides. *Biotechnol. Bioeng. 111*, 761−769.
(33) Yeliseev, A. A., Krueger, K. E., and Kaplan, S. (1997) A mammalian mitochondrial drug receptor functions as a bacterial "oxygen" sensor. *Proc. Natl. Acad. Sci. U. S. A. 94*, 5101−5106.
(34) Cheng, D., Wang, R., Prather, K. J., Chow, K. L., and Hsing, I. M. (2015) Tackling codon usage bias for heterologous expression in Rhodobacter sphaeroides by supplementation of rare tRNAs. *Enzyme Microb. Technol. 72*, 25−34.
(35) Jun, D., Saer, R. G., Madden, J. D., and Beatty, J. T. (2014) Use of new strains of *Rhodobacter sphaeroides* and a modified simple culture medium to increase yield and facilitate purification of the reaction centre. *Photosynth. Res. 120*, 197−205.
(36) Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry 48*, 5953−5962.
(37) Donohue, T. J., and Kaplan, S. (1991) Genetic Techniques in Rhodospirillaceae. *Methods Enzymol. 204*, 459.
(38) Maranhao, A. C., and Ellington, A. D. (2017) Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. *ACS Synth. Biol. 6*, 108−119.
(39) Blight, S. K., Larue, R. C., Anirban, M., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004) Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*. *Nature 431*, 333−335.
(40) Kunjapur, A. M., Stork, D. A., Kuru, E., Vargas-Rodriguez, O., Landon, M., Soll, D., and Church, G. M. (2018) Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proc. Natl. Acad. Sci. U. S. A. 115*, 619−624.
(41) Nagarajan, V., Parson, W. W., Davis, D., and Schenck, C. C. (1993) Kinetics and Free Energy Gaps of Electron-Transfer Reactions in *Rhodobacter sphaeroides* Reaction Centers. *Biochemistry 32*, 12324−12336.
(42) Parson, W. W. (2015) *Modern Optical Spectroscopy*, 2nd ed., Springer.
(43) Suzuki, T., Miller, C., Guo, L. T., Ho, J. M. L., Bryson, D. I., Wang, Y. S., Liu, D. R., and Soll, D. (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol. 13*, 1261−1266.
(44) Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C., and Söll, D. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U. S. A. 104*, 3141−3146.
(45) Laible, P. D., and Hanson, D. K. (2012) *Transformable Rhodobacter Strains, Method for Producing Transformable Rhodobacter Strains*, UChicago Argonne, LLC.
(46) Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., and Mehl, R. A. (2007) Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat. Protoc. 2*, 2601−2607.
(47) Gray, K. A., Farchaus, J. W., Breton, J., and Oesterhelt, D. (1990) Initial characterization of site-directed mutants of tyrosine M210 in the reaction centre of Rhodobacter sphaeroides. *EMBO J. 9*, 2051−2070.
(48) Hambly, D. M., and Gadgil, H. S. (2011) Identification and Quantification of Degradants and Impurities in Antibodies, in Characterization of Impurities and Degradants Using Mass Spectrometry, In *Characterization of Impurities and Degradants Using Mass Spectrometry* (Pramanik, B. N., Lee, M. S., and Chen, G., Eds.), John Wiley & Sons, Inc., Hoboken, NJ.

DOI: 10.1021/acssynbio.8b00100
ACS Synth. Biol. 2018, 7, 1618−1628

Genetic code expansion in *Rhodobacter sphaeroides* to incorporate non-canonical amino acids into photosynthetic reaction centers

Supporting Information

Authors: Jared B. Weaver, Steven G. Boxer
Affiliations: Department of Chemistry, Stanford University, Stanford, CA

Table of Contents

A. Fidelity of suppressor aaRSs and Orthogonality of ncAAs upon Host Transfer

In the main text, orthogonality discussed mainly consisted of the suppressor tRNA orthogonality with respect to the endogenous aaRS pool and the suppressor aaRS orthogonality with respect to endogenous tRNA pool. There are two other more commonly discussed aspects to functioning amber codon suppression system that could be thought of as a form orthogonality. One point is the ability of the suppressor aaRS to maintain orthogonality to the host amino acid pool (and other small amino acid-like metabolites) – in the main text here we chose to refer to this as the fidelity of the suppressor aaRS as it is more typically called.[2-4] The other aspect is the non-canonical amino acid incorporated must be orthogonal to endogenous aaRSs to avoid translation errors across the proteome.[5]

For the host transfer of an amber suppressor aaRS/tRNA pair, the canonical amino acids or structurally similar metabolites available in a new host which an aaRS can bind remain largely unchanged. In contrast, the tRNA structure and the tRNA-binding site on aaRSs may be much more variable since there does not seem to be a common feature maintaining canonical tRNA structural sequence evolutionarily in any given host. We realize this may not be true in all cases, but it was true here and we have not yet found any cases involving significant toxicity (leading to inability to transfer aaRS/tRNA pair) of non-orthogonality due to canonical amino acids binding to suppressor aaRSs. Though changes in the fidelity of certain evolved aaRSs have been observed upon host transfer, they did not report a significant increase in toxicity or any difficulty in plasmid transfer.[6] The ncAA could also not be orthogonal in a new host, but this seems to occur less often in any given amber suppression host transfer in the literature, possibly due to similar mechanisms of recognition of amino acids across different species since the amino acid pool in any species is largely unchanged.

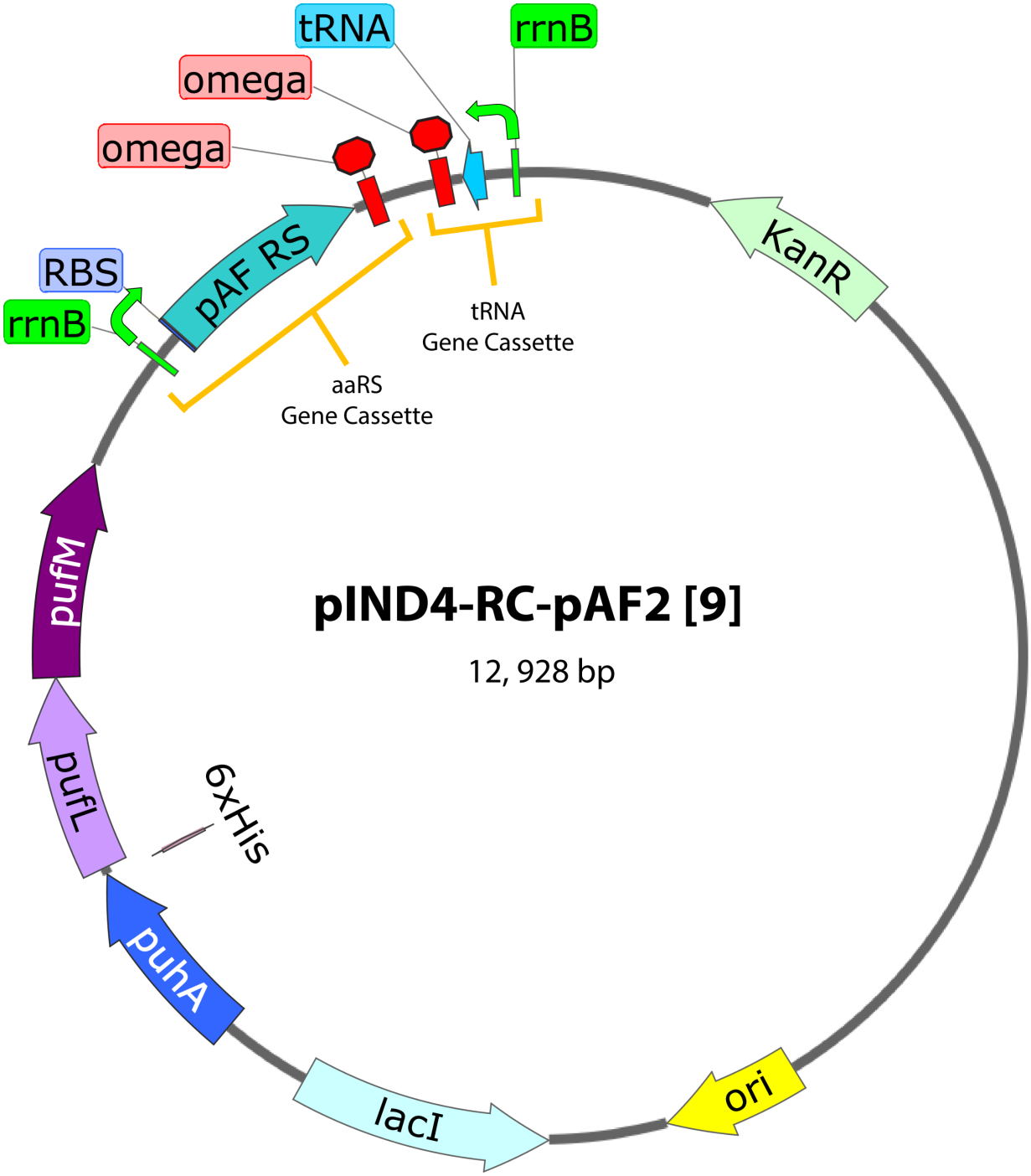


Figure S1. Plasmid map of pIND4-RC-pAF2 [**9**] to illustrate general design of amber suppression/RC plasmids. Green arrows and red stops, respectively, indicate promoter and terminators for the amber suppression genes, in this case P_{rrnB} for the promoter for both the tRNA and aaRS and T_{omega} also for the both the tRNA and aaRS. The boundaries for aaRS and tRNA gene cassettes discussed in this paper (Figure 2A) are outlined above in yellow. The RC, also encoded on the pIND4-RC backbone, is denoted as *puhA*, *pufL*, and *pufM* for the H, L, and M subunits respectively, with a 6xHis-tag on the H-subunit.

B. Sequence detail and plasmid construction for each group of plasmids listed in Table 1 in the main text.

Group 1: Vector Sequences:

pIND4-RC [**1**]: For sequence information, contact Dr. Thomas Beatty at j.beatty@ubc.ca.[7]

pDule-pAF [**2**], pDule-pCNF [**3**], pDule-HaloY [**4**]:
Sequence detail on pDule-pAF [**2**] and pDule-pCNF [**3**] are deposited on Addgene (Addgene plasmids #85502 and #85494 respectively) and in the literature.[8-10] For pDule-HaloY [**4**] sequence information, contact Dr. Ryan Mehl (see Oregon State University Unnatural Protein Facility, http://upf.science.oregonstate.edu/content/contact-us).

Group 2: *E. coli*-regulated aaRS/tRNA for pAF

Gene cassette inserts: **Ec-pAF**, **Ec-tRNA**

Vector: pIND4-RC [**1**]

Plasmid(s) Constructed: pIND4-RC-pAF0 [**7**]

Ec-pAF: Promoter (*E. coli* P_{lpp}), E. coli RBS, pAF aaRS, terminator (*E. coli* $T_{rrnB\text{-}T1}$)
CTCAGAATAAGAAATGAGGCCGCTCATGGCGTTCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACAACC
CTGGCGCCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAAAAAATATTCTCAACATAA
AAAACTTTGTGTAATACTTGTAACGCTGAGTTTACGCTTTGAGGAATCCCCCATGGACGAATTTGAAATG
ATAAAGAGAAACACATCTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAAGATGAAAA
ATCTGCTACTATAGGTTTTGAACCAAGTGGTAAAATACATTTAGGGCATTATCTCCAAATAAAAAAGATG
ATTGATTTACAAAATGCTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCTATTTAAACCAGAA
AGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAA
AGGCAAAATATGTTTATGGAAGTACGTTCCAGCTTGATAAGGATTATACACTGAATGTCTATAGATTGGC
TTTAAAAACTACCTTAAAAAGAGCAAGAAGGAGTATGGAACTTATAGCAAGAGAGGATGAAAATCCAA
AGGTTGCTGAAGTTATTTATCCAATAATGCAGGCTAATCCGTTGCATTATGCTGGCGTTGATGTTGCAGT
TGGAGGGATGGAGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTA
TTCACAACCCTGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAG
CTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCCAGCTGGAGTTGTTG
AAGGAAATCCAATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCAGAAA
AATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGC
ATCCAATGGATTTAAAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGAT
TATAACTGCAGTTTCAAACGGGTACCATATGGGAATTCGAAGCTTGGGCCCGAACAAAAACTCATCTCA
GAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCATTGAGTTTAAACGGTCTCCAGCTTGG
CTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGAT
AAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGA
AACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAAT

AAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTG
AGTAGGACAAATCCGCCGGGAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAA
AGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCA
GTGAAGTGCTTCATGTGGC

Ec-tRNA: Promoter (*E. coli* P_{lpp}), tRNA, terminator (*E. coli* T_{rrnC})
GCTGGTTTACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTT
TGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCATTTATTCTGCCTCCCAGAGCATGATA
AAAACGGTTAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGC
AGATCCGGAACATAATGGTGCAGGGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG
GGACTGTTGGGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGC
GACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTACCGGCAGCGGTGCGGACTGTTGTAACTCAGA
ATAAGAAATGAGGCCGCTCATGGCGTTCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACAACCCTGGCG
CCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAAAAAATATTCTCAACATAAAAAACT
TTGTGTAATACTTGTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCC
GCATGGCGCTGGTTCAAATCCGGCCCGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTT
AAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC
CTTGCAGCACATCCCCCTTTCGCCAGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTA
GTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATG

Ec-pAF Fwd Hifi Primer: CCCGCCTGATGAATGCTCATCTCAGAATAAGAAATGAGGCCGCTCATG
Ec-pAF Rvs Hifi Primer: GCCGCAAGGAATGGTGCATGGCCACATGAAGCACTTCACTGACAC
Ec-tRNA Fwd Hifi Primer: AGTGAAGTGCTTCATGTGGCCATGCACCATTCCTTGCGG
Ec-tRNA Rvs Hifi Primer: CCGTCTTTCATTGCCATACGAAATTGCTGGTTTACCGGTTTATTGACTACCG

Plasmid construction: The plasmid pIND4-RC-pAF0 (**7**) was constructed in this group by PCR-amplifying the Ec-pAF and Ec-tRNA gene cassettes above from the plasmid pDule-pAF (**2**, provided by Dr. Ryan Mehl) plasmid with the four *E. coli* Hifi Primers mentioned above. Then through linearizing the pIND4-RC vector with restriction enzyme Kpn2I (Thermo-Fisher), the vector, aaRS, and tRNA were ligated together with the NEBuilder® HiFi DNA Assembly Cloning Kit. In constructed plasmids, the orientation of the tRNA gene cassette is reversed with respect to the orientation of the pAF gene cassette orientation (Figure S1).

Group 3: Initial aaRS/tRNA for pAF, Toxic in *R. sphaeroides*

Gene cassette inserts: **P_{rrnB}-pAFRS-T_{omega}, P_{rrnB}-tRNA (*E.coli* T_{rrnC}), P_{rrnB}-tRNA-T_{omega}, P_{leuZ}-tRNA-T_{leuZ}**

Vector: pIND4-RC [**1**]

Plasmid(s) Constructed: pIND4-RC-pAF1 [**8**], pIND4-RC-pAF2 [**9**], and pIND4-RC-pAF3 [**10**]

P_{rrnB}-pAFRS-T_{omega}: Promoter (P_{rrnB}), RBS, pAF aaRS, terminator (T_{omega})

GCTAGCGAGCTCGCCGCTCATGGCGTTCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACAACCCTGGCG
CCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAATTGTTACGGAGCCCAAAAAATCC
GCTTGCGCCCGGGGCCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCGGACGG
AGACGAGGGAGCGGATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGATCGCTG
GTTGTCTGCATCAACGGAGGTCCCCCATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA
TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAAGATGAAAAATCTGCTACTATAGGTTTTGAACCAA
GTGGTAAAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGA
TATAATTATATTGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAA
ATAGGAGATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAAGTACG
TTCCAGCTTGATAAGGATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAA
GAAGGAGTATGGAACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATTTATCCAATAA
TGCAGGCTAATCCGTTGCATTATGCTGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA
CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTCACAACCCTGTCTTAACGGGTTTGG
ATGGAGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTA
GGGCTAAGATAAAGAAAGCATACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGATAGCT
AAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATA
GCTATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTAAAAAATGCTGTAG
CTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATAAGATCCGGTGGATGACCTTTTGA
ATGACCTTTAATAGATTATATTACTAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAAATTTT
TTCACAAAACGGTTTACAAGCATAAAGCTTGCTCAATCAATCACCCATCAGCCCTGAGGTCGGGC

P_{rrnB}-tRNA (*E.coli* T_{rrnC}): Promoter (P_{rrnB}), tRNA, terminator (*E. coli* T_{rrnC})
CGAAATAGTACTTCCGGACTCTAGAGTGAAAATTGTTACGGAGCCCAAAAAATCCGCTTGCGCCCGGGG
CCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCGGACGGAGACGAGGGAGCG
GATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGATCGCTGGTTGTCTGTGTAACG
CTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCGCTGGTTCAAAT
CCGGCCCGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGCTTGGCACTGGCCGTC
GTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTT
CGCCAGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAG
CACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGCCCGACCTCAGGGCTGATG

P_{rrnB}-tRNA-T_{omega}: Promoter (P_{rrnB}), tRNA, terminator (T_{omega})
CGAAATAGTACTTCCGGACTCTAGAGTGAAAATTGTTACGGAGCCCAAAAAATCCGCTTGCGCCCGGGG
CCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCGGACGGAGACGAGGGAGCG
GATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGATCGCTGGTTGTCTGTGTAACG
CTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCGCTGGTTCAAAT
CCGGCCCGCCGGACCACTGCAGATGATCCGGTGGATGACCTTTTGAATGACCTTTAATAGATTATATTAC
TAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAAATTTTTTCACAAAACGGTTTACAAGCAT
AAAGCTTGCTCAATCAATCACCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTAC
CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGACGCTCTCCCTTATGCGACTCCTGCATTAGGA
AGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGCCC
GACCTCAGGGCTGATG

P_{leuZ}-tRNA-T_{leuZ}: Promoter ($P_{Native\ -\ LeuZ}$), tRNA, terminator ($T_{Native-\ LeuZ}$)

CGAAATAGTACTTCCGGACTCTAGAGTGACCGCAGCCTGTCAATCGCCGCGCCGCGGCGTCAGGCCGCG
GATCGGGCTCTTCCGGCGCCGTCCGGTCCCCCGCGACGGATCGGCCGGCGGGCGGGGCGCCCCACTTT
TCCCTTGCCGGTTCCGGGTGCGCGTCCTATCCGGCACTCGGGCTGAATTCCCGGCGGTAGTTCAGCAGG
GCAGAACGGCGGACTCTAAATCCGCATGGCGCTGGTTCAAATCCGGCCCGCCGGACCACTGCAGATGC
GCCCGAGCCGCCGTAGCGGCCCCGGTGCGGATCCGGAAGAGTGGTGGAGGGCAGGGGCGTCCCCGAG
TCCGGAAAAGAATGCGGTCGGGCGCTGCTGCCCGCTCGCGATCCCGCGATGAAGCGAGAGGCTGCGCC
CCTGCGTCGGCGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCA
GTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGCCCGACCTCAGG
GCTGATG

pAFRS Fwd Hifi Primer: CCCGCCTGATGAATGCTCATGCTAGCGAGCTCGCCG
pAFRS Rvs Hifi Primer: GCCCGACCTCAGGGCTGATG
tRNA Fwd Hifi Primer: CATCAGCCCTGAGGTCGGGC
tRNA Rvs Hifi Primer: CCGTCTTTCATTGCCATACGAAATTCGAAATAGTACTTCCGGACTCTAGAGTGA

Plasmids constructed: The plasmids pIND4-RC-pAF1 (**8**), pIND4-RC-pAF2 (**9**), and pIND4-RC-pAF3 (**10**) in this group were constructed by PCR-amplifying the aaRS and tRNA cassettes above from ordered IDT G-blocks. The four primers, pAFRS Fwd and Rvs Hifi Primer and the tRNA Fwd and Rvs Hifi Primer above, were used to generate homologous ends for each Hifi assembly. Again, through linearizing the pIND4-RC vector with restriction enzyme Kpn2I (Thermo-Fisher), the vector, the aaRS cassette, and the appropriate tRNA cassette were ligated together with the NEBuilder® HiFi DNA Assembly Cloning Kit. In constructed plasmids, the orientation of the tRNA gene cassettes is reversed with respect to the orientation of the pAF gene cassette orientation (Figure S1).

Group 4: GFP inserts in *R. sphaeroides* to Check Regulatory Control Toxicity

Gene cassette inserts: **aaRSGFP**, **tRNAGFP**

Vector: pIND4-RC [**1**]

Plasmid(s) Constructed: pIND4-RC-GFP1 [**11**] (single aaRSGFP inserted), pIND4-RC-GFP2 [**12**] (single tRNAGFP inserted), and pIND4-RC-GFP3 [**13**] (combo of both aaRSGFP and tRNAGFP inserted)

aaRSGFP: aaRS promoter (P_{rrnB}), RBS, GFP, terminator (T_{omega})
ATCCAAGCTAGCGAGCTCGCCGCTCATGGCGTTCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACAACC
CTGGCGCCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAATTGTTACGGAGCCCAAA
AAATCCGCTTGCGCCCGGGGCCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCG
GACGGAGACGAGGGAGCGGATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGAT
CGCTGGTTGTCTGCATCAACGGAGGTCCCCCATGAGCCAGGGCGAAGAACTGTTTACCGGCGTGGTGCC
GATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGGCGAAGGCGAAGGC

GATGCGACCATTGGCAAACTGACCCTGAAATTTATTTCCACCACCGGCAAACTGCCGGTGCCGTGGCCG
ACCCTGGTGACCACCCTGAGCTATGGCGTGCAGGCCTTTAGCCGCTATCCGGATCACATGAAACGCCAT
GATTTTTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGC
AAATATAAAACCCGCGCGGTGGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGG
CACCGATTTTAAAGAAGATGGCAACATTCTGGGCCATAAACTGGAATATAACTTTAACAGCCATAACGTG
TATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTGCGCCATAACGTGGAAGAT
GGCAGCGTGCAGCTGGCGGATCATTATCAGCAGAACACCCCGATTGGCGATGGCCCGGTGCTGCTGCC
GGATAACCATTATCTGAGCACCCAGACCGTGCTGAGCAAAGATCCGAACGAAAAACGCGATCACATGGT
GCTGCATGAATATGTGAACGCGGCGGGCATTACCCATGGCATGGATGAACTGTATGGTTCCAGCCATCA
CCACCATCATCACTAAGATCCGGTGGATGACCTTTTGAATGACCTTTAATAGATTATATTACTAATTAATT
GGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAAATTTTTTCACAAAACGGTTTACAAGCATAAAGCTTGC
TCAATCAATCACCCATCAGCCCTGAGGTCGGGC

tRNAGFP: aaRS promoter (P_{rrnB}), RBS, GFP, terminator (T_{omega})
CGAAATAGTACTTCCGGACTCTAGAGTGAAAATTGTTACGGAGCCCAAAAAATCCGCTTGCGCCCGGGG
CCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCGGACGGAGACGAGGGAGCG
GATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGATCGCTGGTTGTCTGTGTAACG
ATGAGCCAGGGCGAAGAACTGTTTACCGGCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAA
CGGCCATAAATTTAGCGTGCGCGGCGAAGGCGAAGGCGATGCGACCATTGGCAAACTGACCCTGAAAT
TTATTTCCACCACCGGCAAACTGCCGGTGCCGTGGCCGACCCTGGTGACCACCCTGAGCTATGGCGTGC
AGGCCTTTAGCCGCTATCCGGATCACATGAAACGCCATGATTTTTTTAAAAGCGCGATGCCGGAAGGCT
ATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGCAAATATAAAACCCGCGCGGTGGTGAAATTTG
AAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGG
GCCATAAACTGGAATATAACTTTAACAGCCATAACGTGTATATTACCGCGGATAAACAGAAAAACGGCA
TTAAAGCGAACTTTACCGTGCGCCATAACGTGGAAGATGGCAGCGTGCAGCTGGCGGATCATTATCAGC
AGAACACCCCGATTGGCGATGGCCCGGTGCTGCTGCCGGATAACCATTATCTGAGCACCCAGACCGTGC
TGAGCAAAGATCCGAACGAAAAACGCGATCACATGGTGCTGCATGAATATGTGAACGCGGCGGGCATT
ACCCATGGCATGGATGAACTGTATGGTTCCAGCCATCACCACCATCATCACTAAGATCCGGTGGATGACC
TTTTGAATGACCTTTAATAGATTATATTACTAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAA
AATTTTTTCACAAAACGGTTTACAAGCATAAAGCTTGCTCAATCAATCACCACTGGCCGTCGTTTTACAAC
GTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGACG
CTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGC
CGCAAGGAATGGTGCATGCAAGGAGCCCGACCTCAGGGCTGATG

Single aaRSGFP Fwd Hifi Primer: see Group 5: Single aaRS Fwd Hifi Primer
Single aaRSGFP Rvs Hifi Primer: see Group 5: Single aaRS Rvs Hifi Primer

Single tRNAGFP Fwd Hifi Primer: see Group 5: Single tRNA Fwd Hifi Primer
Single tRNAGFP Rvs Hifi Primer: see Group 5: Single tRNA Rvs Hifi Primer

Combo aaRSGFP Fwd Hifi Primer: see Group 3: pAFRS Fwd Hifi Primer
Combo aaRSGFP Rvs Hifi Primer: see Group 3: pAFRS Rvs Hifi Primer
Combo tRNAGFP Fwd Hifi Primer: see Group 3: tRNA Fwd Hifi Primer
Combo tRNAGFP Rvs Hifi Primer: see Group 3: tRNA Rvs Hifi Primer

Plasmids constructed: The plasmids pIND4-RC-GFP1 [**11**], pIND4-RC-GFP2 [**12**], and pIND4-RC-GFP3 [**13**] in this group were constructed by PCR-amplifying the aaRSGFP and tRNAGFP gene cassettes above from ordered IDT G-blocks with the appropriate primers. Single aaRSGFP Fwd and Rvs Hifi Primers were used to generate homologous ends on aaRSGFP for the vector to construct the pIND4-RC-GFP1 plasmid. Single tRNAGFP Fwd and Rvs Hifi Primers were used to generate homologous ends on the tRNAGFP cassette for the vector to construct the pIND4-RC-GFP2 plasmid. Combo aaRSGFP Fwd and Rvs Hifi Primers were used on the aaRSGFP gene cassette to make homologous ends for the pIND4-RC to 5' aaRSGFP junction and 3' aaRSGFP to 5' tRNAGFP junction. Then, Combo tRNAGFP Fwd and Rvs Hifi Primers were used on the tRNAGFP gene cassette to generate homologous ends for the 3' aaRSGFP to 5' tRNAGFP junction and 3' tRNAGFP to pIND4-RC junction to and both Combo PCR-amplicons were used to construct the pIND4-RC-GFP3 plasmid. Again, through linearizing the pIND4-RC vector with restriction enzyme Kpn2I (Thermo-Fisher), the vector, the aaRSGFP and/or tRNAGFP cassette(s) were ligated together with the NEBuilder® HiFi DNA Assembly Cloning Kit. In constructed plasmids, the orientation of the tRNAGFP gene cassettes is reversed with respect to the orientation of the aaRSGFP gene cassette orientation (Figure S1).

Group 5: Single aaRS or tRNA Inserts for Toxicity Identification

Gene cassette inserts: **P_{rrnB}-pAFRS-T_{omega}**, **P_{rrnB}-tRNA-(*E. coli* T_{rrnC})**, **P_{rrnB}-tRNA-T_{omega}**, **P_{leuZ}-tRNA-T_{leuZ}**

Vector: pIND4-RC

Plasmid(s) Constructed: pIND4-RC-pAFRS [**14**], pIND4-RC-pAFtRNA1 [**15**], and pIND4-RC-pAFtRNA2 [**16**], pIND4-RC-pAFtRNA3 [**17**]

P_{rrnB}-pAFRS-T_{omega}: see Group 3: **P_{rrnB}-pAFRS-T_{omega}** sequence above
P_{rrnB}-tRNA-(*E. coli* T_{rrnC}): see Group 3: **P_{rrnB}-tRNA-(*E. coli* T_{rrnC})**sequence above
P_{rrnB}-tRNA-T_{omega}: see Group 3: **P_{rrnB}-tRNA-T_{omega}** sequence above
P_{leuZ}-tRNA-T_{leuZ}: see Group 3: **P_{leuZ}-tRNA-T_{leuZ}** sequence above

Single pAFRS Fwd Hifi Primer: see Group 3: pAF Fwd Hifi Primer
Single pAFRS Rvs Hifi Primer: CCGTCTTTCATTGCCATACGAAATTGCCCGACCTCAGGGCTGATG

Single tRNA Fwd Hifi Primer: CCCGCCTGATGAATGCTCATCATCAGCCCTGAGGTCGGGC
Single tRNA Rvs Hifi Primer: see Group 3 tRNA Rvs Hifi Primer

Plasmids constructed: The plasmids pIND4-RC-pAFtRNA1 [**14**], pIND4-RC-pAFtRNA2 [**15**], pIND4-RC-pAFtRNA3 [**16**], and pIND4-RC-pAFRS [**17**] in this group were constructed by PCR-amplifying the pAFRS or tRNA gene cassettes above with the appropriate primers. As in single-GFP insertions, Single pAF Fwd and Rvs Hifi Primers were used to generate homologous ends on the pAFRS cassette for the pIND4-RC vector to construct the pIND4-RC-pAF [**17**] plasmid. Similarly,

single tRNA Fwd and Rvs Hifi Primers were used to generate homologous ends on the different tRNA cassettes for the pIND4-RC [**1**] vector to make the pIND4-RC-pAFtRNA1 [**14**], pIND4-RC-pAFtRNA2 [**15**], and pIND4-RC-pAFtRNA3 [**16**] plasmids. Again, through linearizing the pIND4-RC vector with restriction enzyme Kpn2I (Thermo-Fisher), the vector, the pAF casette or the appropriate tRNA cassette were ligated together with the NEBuilder® HiFi DNA Assembly Cloning Kit. In constructed plasmids, the orientation of the pAF gene cassette is reversed with respect to the orientation of the tRNA gene cassette orientation (Figure S1).

Group 6: Toxic pCNF aaRS

Gene cassette inserts: **P_{rrnB}-pCNFRS-T_{omega}**

Vector: pIND4-RC [**1**]

Plasmid(s) Constructed: pIND4-RC-pCNFRS [**18**]

P_{rrnB}-pCNFRS-T_{omega}: Promoter (P_{rrnB}), RBS, pCNF aaRS, terminator (T_{omega})
ATCCAAGCTAGCGAGCTCGCCGCTCATGGCGTTCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACAACC
CTGGCGCCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAATTGTTACGGAGCCCAAA
AAATCCGCTTGCGCCCGGGGCCGTCTGCTCCTAGAAACCGCTTCACCGAGACGAAGACCGGCAGCGCCG
GACGGAGACGAGGGAGCGGATGACAGAAACGTCGGCCGCGACAATTGAAGATGAGGCGGACGGGAT
CGCTGGTTGTCTGCATCAACGGAGGTCCCCCATGGACGAATTTGAAATGATAAAGAGAAACACATCTGA
AATTATCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAAGATGAAAAATCTGCTCTGATAGGTTTTGA
ACCAAGTGGTAAAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGA
TTTGATATAATTATAGTGTTGGCTGATTTACATGCCTATTTAAACCAGAAAGGAGAGTTGGATGAGATTA
GAAAAATAGGAGATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAA
GTGAATGGATGCTTGATAAGGATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAG
AGCAAGAAGGAGTATGGAACTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTATC
CAATAATGCAGGTTAATGGTCCTCATTATCTTGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAA
AAATACACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTCACAACCCTGTCTTAACGGG
TTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGA
GATTAGGGCTAAGATAAAGAAAGCATACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGA
TAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTTGACAGT
TAATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTAAAAAATGCT
GTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATAAGATCCGGTGGATGACCTT
TTGAATGACCTTTAATAGATTATATTACTAATTAATTGGGGACCCTAGAGGTCCCCTTTTTTATTTTAAAA
ATTTTTTCACAAAACGGTTTACAAGCATAAAGCTTGCTCAATCAATCACCCATCAGCCCTGAGGTCGGGC

pCNFRS Fwd Hifi Primer: See Group 5: Single pAFRS Fwd Hifi Primer
pCNFRS Rvs Hifi Primer: See Group 5: Single pAFRS Rvs Hifi Primer

Plasmids constructed: The plasmid, pIND4-RC-pCNF[**18**] in this group, was constructed by PCR-amplifying the pCNF gene cassettes above with the primers listed above. As in the single-GFP and the single pAFRS insertions, Group 4: Single pAFRS Fwd and Rvs Hifi Primers were used to

generate homologous ends on the pCNFRS cassette for the pIND4-RC [**1**] vector to be able to construct the pIND4-RC-pCNF plasmid. Through linearizing the pIND4-RC [**1**] vector with restriction enzyme Kpn2I (Thermo-Fisher), the pIND4-RC [**1**] vector and the pCNF gene cassette were ligated together with the NEBuilder® HiFi DNA Assembly Cloning Kit. The pCNFRS gene cassette is in the same orientation as the pAFRS gene cassette-containing plasmids in Figure S1.

All primers were ordered from the Protein and Nucleic Acid Facility (PAN) at Stanford University. PCR was performed on aaRS and tRNA All PCR was performed using the NEB Phusion® High-Fidelity DNA Polymerase according to manufacturer-recommended thermocycling temperatures and times at each annealing, elongation, and melting step given our primers and the length of the desired tRNA and aaRS amplicons. HF buffer and 3% DMSO concentration were used in all PCR reactions. The sequence information for aaRS and tRNA cassettes for ClY, BrY, and IY incorporation (what would be Group 6) are not included above because work in which the Mehl lab evolved the amber suppressor MbPyrRS/tRNA pair for halogenated tyrosine incorporation has not yet been published.[11] Sequence information for the HaloY aaRS cassette and the three tRNA cassettes (P_{rrnB}-tRNA-(*E. coli* T_{rrnC}), P_{rrnB}-tRNA-T_{omega}, P_{leuZ}-tRNA-T_{leuZ}) is identical to the pAF and pCNF aaRS and tRNA cassettes, except for the identity and sequence of the tRNA and aaRS. Similarly, construction was also identical to plasmids created in Group 3 since only the aaRS and tRNA sequences and not the 5' or 3' bordering sequence elements were changed for aaRS and tRNA gene cassettes. In every plasmid construction, Fwd and Rvs Primers for tRNA or tRNA-derived cassettes (tRNAGFP) are only Fwd and Rvs with respect to the orientation of the aaRS and vector backbone. Therefore, Fwd Hifi primers for the tRNA actually bind down-stream of the 3’ end of the tRNA because of its opposite orientation to the aaRS (Figure S1) and vice-versa for the Rvs primers. As mentioned in the main text, every plasmid in which amber codon suppression was screened,

had a duplicate made which contained the TAG codon in place of the WT tyrosine TAC codon at 210 on the M-subunit in the reaction center. All constructed plasmids were sequenced to verify plasmids were made as designed through ELIM Biopharmaceuticals, Inc. (Hayward, CA). Full sequence information for all plasmids discussed here (see Table 1 in manuscript) will be deposited in Addgene following publication by the Mehl lab of work describing the HaloYRS and tRNA .

C. Proposed Mechanism of pAFRS and Group 3 Plasmid Construct Toxicity

Given toxicity was present whether or not pAF is supplemented to media in Group 3 plasmid transformations, we propose the simplest interpretation is that the pAFRS is toxic in *R. sphaeroides* because it can charge an endogenous non-tyrosyl-tRNA with tyrosine resulting in proteotoxic stress. In other words, the lack of orthogonality in the pAFRS is due to its inability to distinguish its own amber suppressor tRNA from the *R. sphaeroides* native tRNA pool. This is consistent with a non-orthogonal pAFRS, since toxicity is likely caused by endogenous tRNA aminoacylation by the pAFRS, which still retains some affinity to canonical amino acid substrate. In this case, substrate was likely tyrosine given the pAFRS was evolved from an MjTyrRS and other groups' past results note similar evolved MjTyrRSs still display tyrosine affinity.[6] This affinity can be also be seen in negative controls in previous work which display low amounts of protein expression from amber-codon interrupted constructs when ncAA is withheld (which is unsurprising given the high structural similarity between pAF and tyrosine).[4, 6, 9] This is not to say tyrosine aminoacylation occurs in pAF-supplemented inductions where an operating aaRS/tRNA pair for pAF incorporation is present, as competition between pAF and tyrosine now occurs. Site-specific incorporation of pAF in past results suggest that under pAF supplementation, pAF out-competes tyrosine for binding to the aaRS and is the primary amino acid incorporated at amber codons.[8] The toxicity which remained upon pAF supplementation to Group 3 plasmid transformations also supports the conclusion a non-orthogonal aaRS is present, since a non-tyrosyl-tRNA charged with pAF would be similarly detrimental to cell growth as one charged with tyrosine.

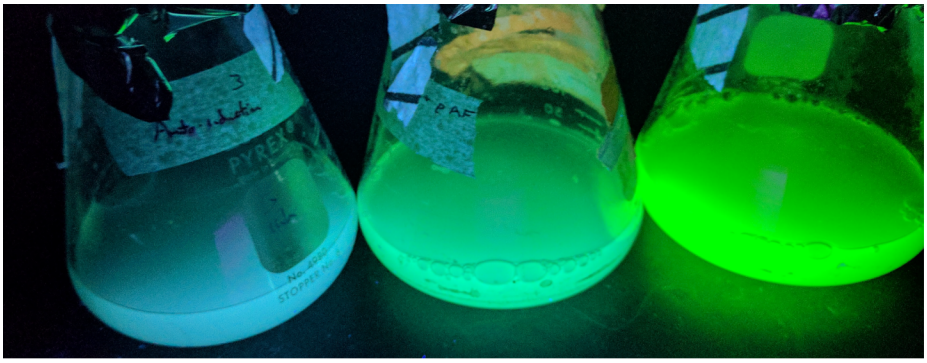
While it is possible that in the presence of the *M. jannaschii*-tRNA^{Tyr}, the mechanism of toxicity induced by the aaRS could change to one like that of amber codon read-through in genomic genes resulting in toxic undesired elongation of essential proteins and stress on translational resources, it seems less likely. In this study, we describe a successful amber suppressor aaRS/tRNA pair (HaloYRS/tRNA^{Pyl}) which even upon addition of its cognate ncAAs, had largely unchanged cell growth as measured by final OD700 following inductions. The mechanism proposed above appears to be the most probable given the toxicity of plasmid constructs constructed here (Table 2) and past reports of aaRS non-orthogonality in infidelitous aminoacylation of non-tyrosyl tRNAs.[12]

D. Toxicity of aaRS/tRNA in Lower-copy Vector

Due to past reports of the decreased toxicity of lower copy-number plasmids in a non-orthogonal aaRS/tRNA pair, the Mj pAF aaRS/tRNA pair was inserted into the lower copy (4-7 copy) pRK-415 vector.[13, 14] Results, however, were similar to those found with the pIND4-RC vector (see Table 1, Table 2). When the tRNA and aaRS promoters and terminators remained unaltered, plasmids were transformable but did not produce RCs. Then, when the aaRS/tRNA were placed under the control of *R. sphaeroides* promoters, RBS, and terminators (as in pIND4-RC-pAF1 [**8**], pIND4-RC-pAF2 [**9**], and pIND4-RC-pAF3 [**10**]), plasmid constructs did not produce cell growth on plasmid transformations while positive controls displayed typical transformation efficiencies with 1 x 10^2 to 1 x 10^3 cfu/ug seen in positive controls consisting of pRK-415 vector transformations.

E. IPTG-induction to Decrease Amber Suppression Toxicity

The pAF aaRS was placed under the control of the IPTG-inducible *lac* promoter (generating pIND4-RC-pAF4 [**not in Table 1**]). As might be expected the plasmid was still toxic on transformations with IPTG and transformable on IPTG absent plates. However, when amber suppression genes and RCs were later induced with IPTG, no RCs were produced in either (M210)TAG samples supplemented with 1 mM pAF or in positive controls where RCs were wild-type (tyrosine) at M210. This continued even when IPTG concentration was increased to 2 mM and then 4 mM. The dual control of the aaRS and RC expression were likely tied to these unsuccessful attempts in RC expression.



- pAF + pAF WT

Figure S2. Amber suppression in *E. coli*. amber suppression capacity of pIND4-RC-pAF0 [**7**] when co-transformed with a plasmid bearing an amber codon interrupted GFP gene (pBad-sfGFP-150TAG [**6**]) in *E. coli*. GFP fluorescence is noticeably present in the WT culture (WT) and in the GFP-150TAG culture to which pAF is added (+ pAF) while only auto-fluorescence is seen in GFP-150TAG culture where pAF was retained (- pAF). WT cell culture contained cells co-transformed with a plasmid bearing WT GFP (pBad-sfGFP [**5**]) and pIND4-RC-pAF0 [**7**] as a positive control.

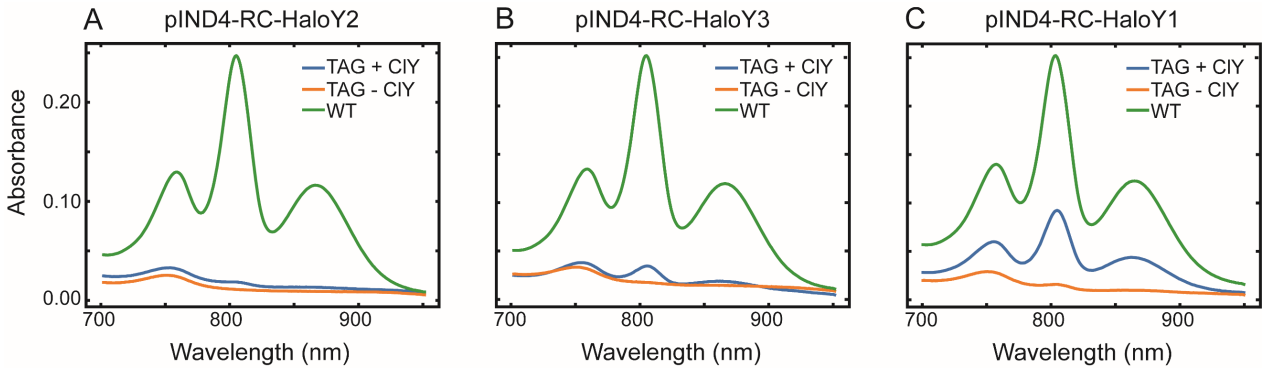


Figure S3. Near-IR unprocessed absorption screen of a single replicate of cell lysates to test amber suppression function in *R. sphaeroides*. Cell lysate containing solubilized RC chromatophores show three maxima past 700 nm characteristic of RC expression. TAG + ClY spectra refer to lysate from cells containing the construct listed above with an amber stop codon (TAG) at M210 which have been supplemented with the relevant non-canonical amino acid, ClY. In contrast, TAG - ClY have TAG at M210 but have not had ClY supplemented to media. WT does not have the TAG at M210 mutation and shows wild-type RC expression in the presence of amber suppression genes.

Table S1: LC Elution Gradient for LC/MS

Time (min)	Solvent A	Solvent B
0	95	5
1.5	95	5
2	65	35
34	20	80
35	5	95
37	5	95
38	95	5
45	95	5

Table S1. Gradient used in LC elution of intact subunits for LC/MS experiments. Solvent A was 0.05% trifluoroacetic acid, 0.09% formic acid in water, Solvent B was 0.1% formic acid in acetonitrile.

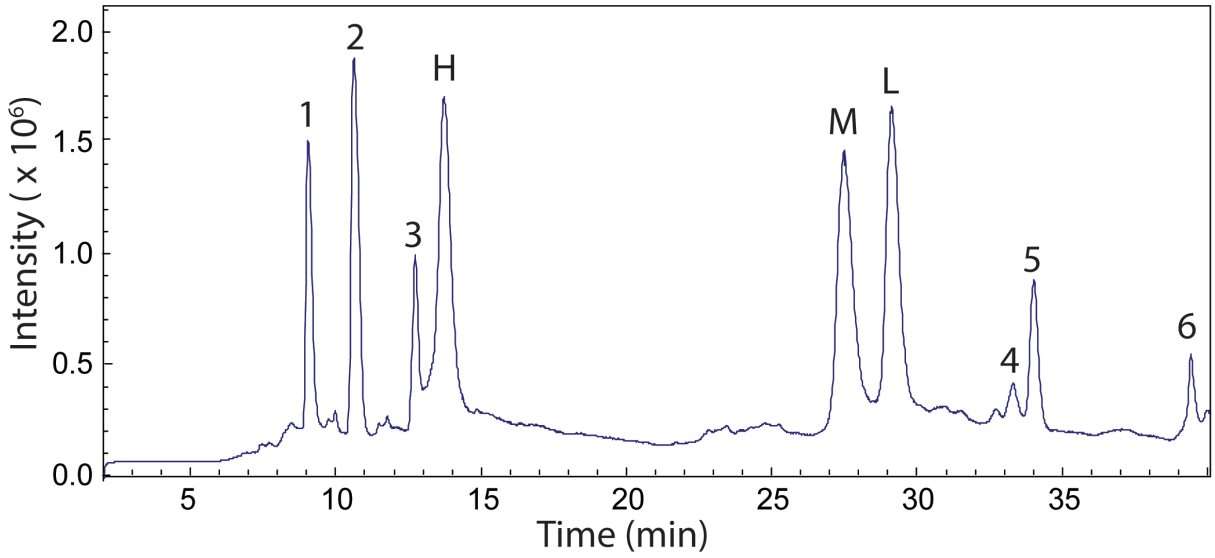


Figure S4. LC Chromatogram of RC with IY at M210 displayed as a representative chromatogram of intact RC subunits and associated small molecule cofactors chromatograms (all chromatograms looked similar) and intensity on y-axis correlated with total ion count from mass spectrometry. When under 4 C sample storage for limited time (<1 week) prior to running samples, all RC variant and WT proteins produce a near identical chromatogram. Peaks corresponding to H-, M-, and L-subunits are indicated their corresponding eluate peak while small molecules peaks are denoted numerically (see Table S1 for elution gradient). While the methods used for mass determination were intended for large subunit mass determination and were not suitable for accurate masses for small molecule peaks, general estimates between 200 and 1000 Da were found for small molecule peaks. This is as expected and would be consistent with RC cofactor masses, such as bacteriochlorophyll and residual detergent used to solubilize RCs.

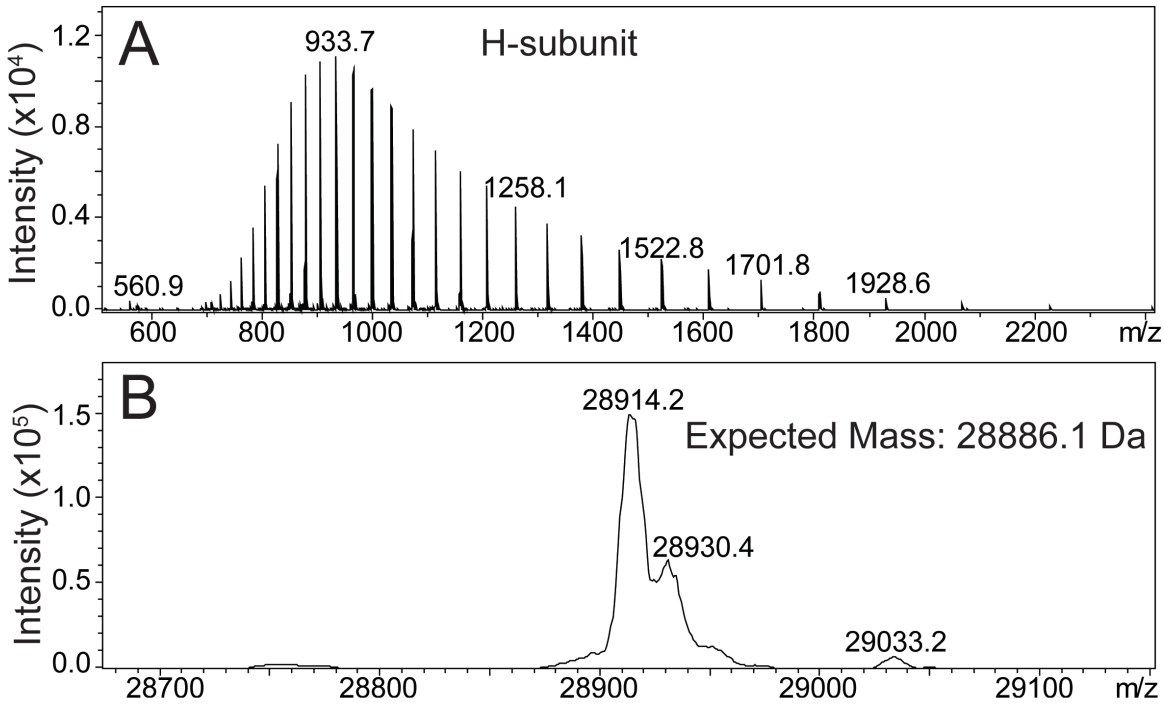


Figure S5. (A) Raw MS spectra of the H-subunit (peak H at 13.5 min in Figure S4) from LC/MS experiment, representative of general H-subunit mass spectra as other RC variants were not significantly different. (B) Deconvoluted MS spectra of the H-subunit. The low M/Z peak (peak at 28914.2 Da) was used for subunit mass comparison across different RC variants. The theoretical mass was calculated with the initial methionine included given LC/MS-MS results showing Methionine presence in H-subunit peptide sequencing and past crystallographic structures (Protein Data Bank, 2J8C). The L- and M-subunits (Figures S6 and S7) did not include Methionine mass for the same reasons. While there was a change in mass in the experimentally observed 28914.2 Da peak, this shift remained constant across all RC variants. This +27.9 Da shift from the expected mass is potentially due to the formyl group addition (+28.0) or to some other small chemical modification caused by sample preparation.

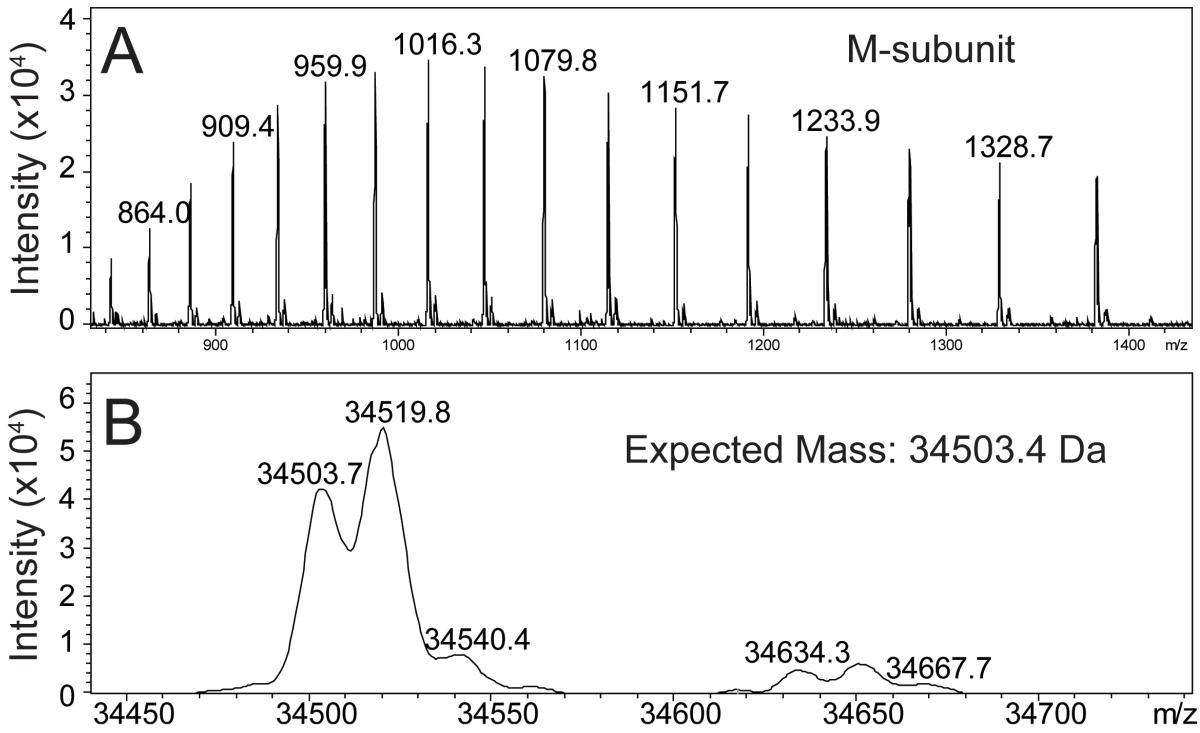


Figure S6. (A) Raw MS spectra of the M-subunit (peak M at 27.5 min in Figure S4) from LC/MS experiment, though masses varied depending on the RC tyrosine variant, the spectra above are representative of M-subunit mass spectra as other RC variants were not significantly different in peak shape. (B) Deconvoluted MS spectra of the M-subunit. The low M/Z peak (peak at 34503.7 Da) was used for subunit mass comparison across different RC variants. Since these are mass spectra of (M210)IY RCs, the expected mass was calculated with the mass change for modification of tyrosine to 3-iodootyrosine (+125.9 Da) taken into account.

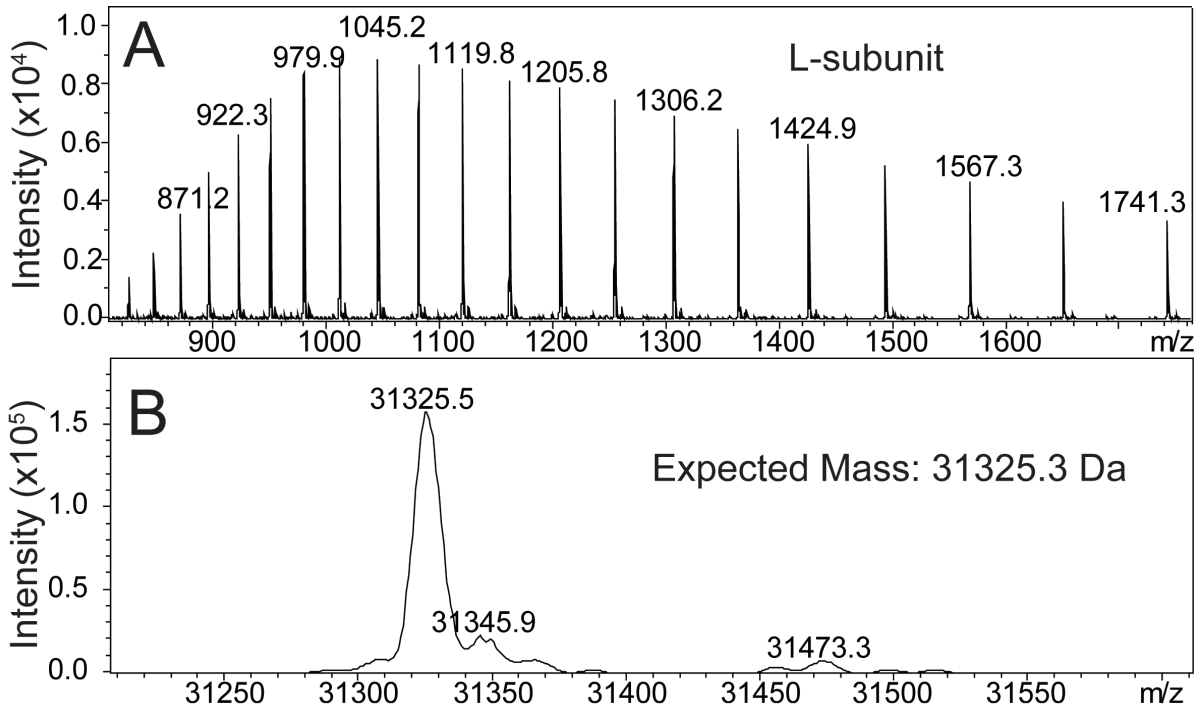


Figure S7. (A) Raw MS spectra of the L-subunit (peak L in Figure S4) from LC/MS experiment, representative of general L-subunit mass spectra as other RC variants were not significantly different. (B) Deconvoluted MS spectra of the L-subunit. Deconvoluted MS spectra of the M-subunit. The low M/Z peak (peak at 31325.5 Da) was used for subunit mass comparison across different RC variants.

Table S2: LC/MS Subunit Delta Mass Comparison between RC variants and WT RCs

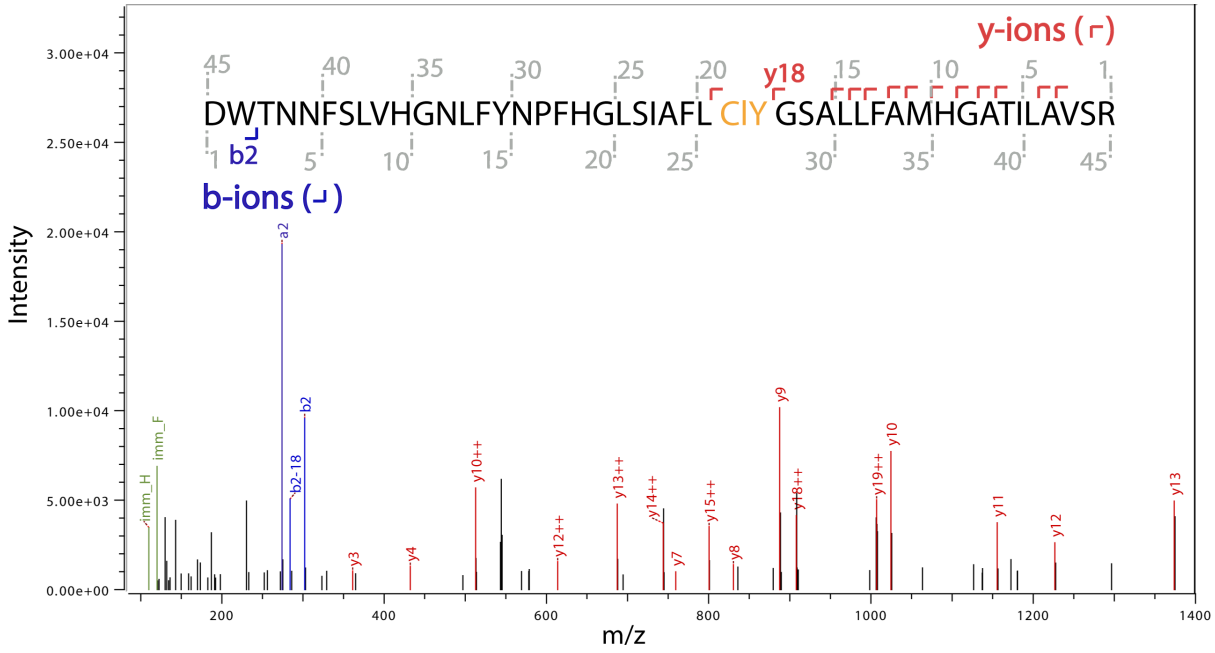
RC Variant	H (Da)	H Delta Mass	M (Da)	M Delta Mass (Da)	L (Da)	L Delta Mass
WT	28914	NA	34378	NA	31325	NA
Y(M210)ClY	28914	0	34412	34	31325	0
Y(M210)BrY	28914	0	34456	79	31325	0
Y(M210)IY	28914	0	34504	126	31326	1

Table S2. Mass change from WT mass for each subunit across the different M210 tyrosine variant from whole-subunit LC/MS. Mass change was calculated as the difference in mass between the major low mass peak of the variant subunit minus the mass of the WT peak at the same elution time. Additional peaks on each subunit are likely due to mass change via oxidation due to the formic acid in the elution gradient, mass change due chemical adducts, or mass change due to retaining the initial methionine

Table S3: LC Elution Gradient for LC/MS-MS

Time (min)	% Solvent A	% Solvent B
0	98	2
0.1	94	6
20	80	20
25	65	35
35	55	45
71	30	70
75	5	95
77	5	95
78	98	2
80	98	2

Table S3. Step gradient used in LC elution of digested RCs for LC/MS-MS experiments. Solvent A was 0.2% Formic Acid + 99.8% Water and Solvent B was 0.2% Formic Acid + 99.8% Acetonitrile. Flow was 0.45 μL/minute throughout the 80-minute gradient.



B ion M/Z Calculated vs. Observed for Peptides Verifying ClY incorporation

Sequence	Ion	Calc. Mass	Obs. Mass	Delta Mass
$^{+}$HGATILAVSR	Y10+	1024.59	1024.595	0.0054
$^{2+}$ClYGSALFAMHGATILAVSR	Y19++	1007.017	1007.011	-0.0059
$^{2+}$GSALFAMHGATILAVSR	Y18++	908.0009	907.9984	-0.0025
$^{+}$GATILAVSR	Y9+	887.5309	887.5345	0.0036

Figure S8. MS-MS spectra of RCs with ClY incorporated at M210. B ions are indicated in blue and y ions are indicated in red. Numbering of b ions is given below the peptide sequence while y ion numbering is written above the peptide sequence since b-ions fragment N-terminally and y-ions C-terminally. Fragmentation sites for b- and y-ions observed are indicated above the peptide sequence with either red (y-ions) or blue (b-ions) corner symbols. Doubly charged ions are indicated with "++". ClY in the peptide sequence is colored in yellow. The table below displays a subset of MS-MS peaks in which the peptide with ClY incorporation (Y19++) is also indicated with ClY in yellow.

F. LC/MS-MS Search Parameters

LC/MS-MS data was processed in Byonic™ version 2.12.0 by Protein Metrics Inc. In searches of peptide MS-MS spectra, the following modifications were allowed: tyrosine chlorination (+34.968853 Da), asparagine deamidation (+0.984016 Da), glutamine deamidation (0.984016 Da), glutamic acid amidation (-0.984016 Da), aspartic acid amidation (-0.984016 Da), N-terminal acetylation (+42.010565 Da), propionamide addition at cysteine (+71.037114 Da), and methionine oxidation (+15.994915 Da). Only one N-terminal acetylation was allowed and for all other modifications only two occurrences were allowed each. A complex cleavage was allowed, with C-terminal cleavage at arginine or lysine and an N terminal cleavage at glutamic acid, allowing for ragged N-terminal end of the peptides. A maximum of two missed cleavages was allowed to account for digest inefficiency. Mass accuracy search tolerances for the precursor scan(s) were 12 ppm, with fragmentation mass tolerances of 12 ppm (for HCD collected in the orbitrap) and 0.4 Da (for ETD and CID collected in the ion trap). A protein false discovery value (FDR) of 1% was used in the search. A maximum precursor mas of 10,000.0 Da was allowed. Due to experimental conditions (DTT reduction followed by acrylamide-cysteine capping), disulfide and trisulfide modifications were not allowed.

Table S4 ClY Site-specificity via LC/MS-MS with ETD/HCD Fragmentation

Residue	(H29)Y	(H30)Y	(L144)Y	(M101)Y	(M210)Y
ClY Hits	2	1	1	1	15
Total Hits	199	199	59	26	15
% Incorp.	1%	1%	2%	4%	100%
% Error	7%	7%	13%	20%	26%

Table S4. The relative frequency of ClY incorporation at the site indicated above as determined by MS-MS sequencing, where the MS2 used for peptide fragmentation was electron-transfer dissociation/higher-energy collisional dissociation (ETD/HCD). Here the total number of spectra where a site on the RC was identified to contain a ClY modification by MS-MS sequencing is indicated on the "ClY Hits" row and number of spectra corresponding to sequence coverage at that site is indicated below as "Total Hits". MS-MS sequencing is indicated in "Total Hits". Only data from MS-MS sequencing with a protein-aware posterior error probability (PEP2D) of less than 0.001 (in both total hits at that residue and ClY incorporated hits) were used for ClY incorporation rate calculations. Incorporation of ClY at M210 in the table is boxed in yellow.

Table S5: ClY Site-specificity via LC/MS-MS with CID Fragmentation

Residue	(H18)Y	(M3)Y	(M198)Y	(M210)Y
ClY Hits	2	1	1	14
Total Hits	98	71	15	15
% Incorp.	2%	1%	7%	93%
% Error	10%	12%	26%	26%

Table S5. The relative frequency of ClY incorporation at the site indicated above as determined by MS-MS sequencing, where the MS2 used for peptide fragmentation was accomplished via collision-induced fragmentation (CID). ClY Hits, Total Hits, % incorporations and % error were calculated and reported identically to Table S4. Similarly, incorporation of ClY at M210 in the table is boxed in yellow.

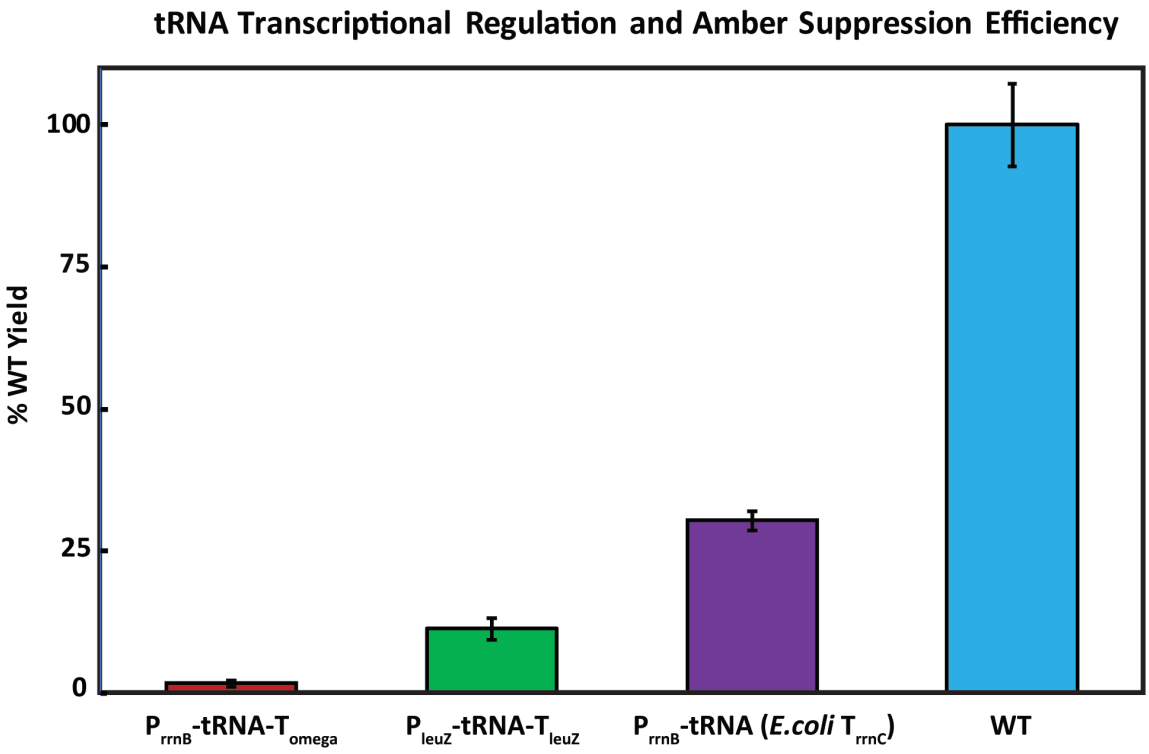


Figure S9. Crude RC yields of RC variants containing ClY at M210 as a function of tRNA transcriptional regulation. Yields were based on absorbance at 804 nm and a molar absorptivity of 288,000 $M^{-1}cm^{-1}$.[1] Amber suppression efficiency or the ability of the evolved ClY RS/tRNA pair to incorporate ClY at M210 is shown to vary dramatically depending on the composition of the cassette surrounding the tRNA. WT RC yields pertain to WT RCs produced in plasmids containing RCs without the amber codon at M210. In each ClY tRNA-variant, spectra were baselined by subtracting negative control UV-Vis-NIR spectra from the same induction (in which no ClY was added) from samples containing ClY. Absorbance between 1000-1100 nm where the RC does not absorb (due to instrumental error and small amounts of scattering) was then averaged and subtracted off to better baseline the spectra and more accurately and reproducibly determine RC yield.

References

[1] Gray, K. A., Farchaus, J. W., Breton, J., and Oesterhelt, D. (1990) Initial characterization of site-directed mutants of tyrosine M210 in the reaction centre of Rhodobacter sphaeroides, *The EMBO Journal 9*, 2051-2070.

[2] Monk, J. W., Leonard, S. P., Brown, C. W., Hammerling, M. J., Mortensen, C., Gutierrez, A. E., Shin, N. Y., Watkins, E., Mishler, D. M., and Barrick, J. E. (2017) Rapid and Inexpensive Evaluation of Nonstandard Amino Acid Incorporation in Escherichia coli, *ACS Synth Biol 6*, 45-54.

[3] Antonczak, A. K., Simova, Z., Yonemoto, I. T., Bochtler, M., Piasecka, A., Czapinska, H., Brancale, A., and Tippmann, E. M. (2011) Importance of single molecular determinants in the fidelity of expanded genetic codes, *Proc Natl Acad Sci U S A 108*, 1320-1325.

[4] Kunjapur, A. M., Stork, D. A., Kuru, E., Vargas-Rodriguez, O., Landon, M., Soll, D., and Church, G. M. (2018) Engineering posttranslational proofreading to discriminate nonstandard amino acids, *Proc Natl Acad Sci U S A 115*, 619-624.

[5] Minajigi, A., Deng, B., and Francklyn, C. S. (2011) Fidelity escape by the unnatural amino acid beta-hydroxynorvaline: an efficient substrate for Escherichia coli threonyl-tRNA synthetase with toxic effects on growth, *Biochemistry 50*, 1101-1109.

[6] He, J., Van Treeck, B., Nguyen, H. B., and Melancon, C. E., 3rd. (2016) Development of an Unnatural Amino Acid Incorporation System in the Actinobacterial Natural Product Producer Streptomyces venezuelae ATCC 15439, *ACS Synth. Biol. 5*, 125-132.

[7] Jun, D., Saer, R. G., Madden, J. D., and Beatty, J. T. (2014) Use of new strains of *Rhodobacter sphaeroides* and a modified simple culture medium to increase yield and facilitate purification of the reaction centre, *Photosynth. Res. 120*, 197-205.

[8] Mehl, R. A., Anderson, J. C., Santoro, S. W., Wang, L., Martin, A. B., King, D. S., Horn, D. M., and Schultz, P. G. (2003) Generation of a Bacterium with a 21 Amino Acid Genetic Code, *J. Am. Chem. Soc. 125*, 935-939.

[9] Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., and Mehl, R. A. (2010) Generating permissive site-specific unnatural aminoacyl-tRNA synthetases, *Biochemistry 49*, 1667-1677.

[10] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry 48*, 5953-5962.

[11] Oregon State University Unnatural Protein Facility. http://upf.science.oregonstate.edu/content/contact-us (February 5, 2018).

[12] Javahishvili, T., Manibusan, A., Srinagesh, S., Lee, D., Ensari, S., Shimazu, M., and Schultz, P. G. (2014) Role of tRNA orthogonality in an expanded genetic code, *ACS Chem Biol 9*, 874-879.

[13] Tian, H., Deng, D., Huang, J., Yao, D., Xu, X., and Gao, X. (2013) Screening system for orthogonal suppressor tRNAs based on the species-specific toxicity of suppressor tRNAs, *Biochimie 95*, 881-888.

[14] Inui, M., Nakata, K., Roh, J. H., Vertes, A. A., and Yukawa, H. (2003) Isolation and Molecular Characterization of pMG160, a Mobilizable Cryptic Plasmid from Rhodobacter blasticus, *Appl. Environ. Microbiol. 69*, 725-733.