Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System

Sviatlana Smolskaya1,2, Zhiwen Jonathan Zhang3*, Lital Alfonta1,2*

1 Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 2 The Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 3 Department of Bioengineering, School of Engineering, Santa Clara University, Santa Clara, California, United States of America

Abstract

Using a commercial protein expression system, we sought the crucial elements and conditions for the expression of proteins with genetically encoded unnatural amino acids. By identifying the most important translational components, we were able to increase suppression efficiency to 55% and to increase mutant protein yields to levels higher than achieved with wild type expression (120%), reaching over 500 μg/mL of translated protein (comprising 25 μg in 50 μL of reaction mixture). To our knowledge, these results are the highest obtained for both in vivo and in vitro systems. We also demonstrated that efficiency of nonsense suppression depends greatly on the nucleotide following the stop codon. Insights gained in this thorough analysis could prove useful for augmenting in vivo expression levels as well.

Introduction

The methodology based on unnatural amino acids (UAAs) incorporation into desired loci of the protein of interest is widely used for understanding protein structure-function relationships, investigating protein-based biological processes, and generating proteins and organisms with new properties [1]. Over the past two decades, the most established methods to site-specifically incorporate UAA in vivo were based on genetic code expansion. This is accomplished by supplying organisms with a non-endogenous aminoacyl-tRNA synthetase/tRNA pair, referred to as an orthogonal pair, that directs site-specific incorporation of UAA in response to a unique codon [2]. The orthogonal aminoacyl-tRNA synthetase (aaRS) aminoacylates a cognate orthogonal tRNA (but no other cellular tRNAs) with UAA. The orthogonal tRNA is a substrate for the orthogonal aaRS but is not aminoacylated by any endogenous aaRS [3]. The orthogonal aaRS/tRNA pair should, however, be compatible with the translational machinery of the host cell. The first orthogonal aaRS/tRNA pair used in Escherichia coli originated from the archaeon, Methanoclostridium jannaschii, and was generated from the tyrosyl-tRNA synthetase and its cognate tRNA pair (MjTyrRS/ tRNA<sup>Tyr</sup>) [4]. A unique codon is required to specify the UAA, and two main strategies are generally used: nonsense suppression, i.e. UAA incorporation in response to the least used stop codon recognized by a specific suppressor tRNA [5,6]; and frame-shift suppression based on the application of four- or five-base extended codons and cognate suppressors [7–9].

Nowadays, genetic code expansion in E. coli using the amber suppression strategy and evolved variants of orthogonal M. jannaschii TyrRS/MjtRNA<sup>Tyr</sup><sub>Clua</sub> is considered to be the most established and robust methodology to site-specifically incorporate UAAs. However, this methodology has not been reported to achieve high protein yields. Suppression efficiency depends greatly on the degree of orthogonal tRNA compatibility with the translational apparatus of host cell. The reasons for such an incompatibility are low affinity of the elongation factor Tu (EF-Tu) to UAA-charged MjtRNA<sub>Clua</sub> derived from archael tRNA<sup>Tyr</sup> [10] or its inability to recognize and deliver orthogonal tRNA to the ribosomal A-site [11]. On the other hand, suppression efficiency can be interrupted by the release factors, RF1 and RF2, which are responsible for the release of the growing polypeptide chain from the ribosome. RF functionality depends on the particular stop codon encoded in the mRNA sequence, as well as the fourth base following the stop codon [12,13]. Recently, tremendous efforts were geared towards the production of greater yields of recombinant proteins containing UAAs, including design of specific vectors encoding multiple copies of MjtRNA<sub>Clua</sub> [14], identification of newly optimized suppressor tRNA<sub>Clua</sub><sup>opt</sup> with modified T-stem [15–17], selection of the orthogonal ribosome [18], and selection and application of RF1-depleted E. coli variants [19,20]. However, given the in vivo nature of the methodology, overall yield of recombinant protein with site-specifically incorporated UAA has never been reported to exceed 50% comparing to wild type.

Citation: Smolskaya S, Zhang ZJ, Alfonta L (2013) Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System. PLoS ONE 8(7): e68363. doi:10.1371/journal.pone.0068363

Editor: Jörg D. Hoheisel, Deutsches Krebsforschungszentrum, Germany

Received March 8, 2013; Accepted June 3, 2013; Published July 2, 2013

Copyright: © 2013 Smolskaya et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The research leading to these results received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 260647 and from the Edmond J. Safra Center for the Design and Engineering of Functional Biopolymers. LA holds the Elaine S. and Alvin W. Wene Career Development Chair in Biotechnology Engineering. SS acknowledges the Harbour Foundation for a Fay and Bert Harbour Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zzhang@scu.edu (ZJJ); alfonta@bgu.ac.il (LA)
Additionally, the evolution of cognate aaRS/tRNA pairs and the incorporation of UAs into proteins in living organisms is not possible for certain amino acids. These include toxic UAs, amino acids with extreme redox potentials or hydrophilicity that are not able to cross the cell membrane or UAAs which cannot be generated in large amounts, due to their difficult multi-step chemical synthesis. Moreover, site-directed UAAs incorporation into eukaryotic proteins expressed in mammalian cells is problematic, since no efficient system is available yet.

Herein we suggest complementing caveats that exist in the in vivo translational approach by using an efficient cell-free system for the different applications of this robust technology. The cell-free protein translation system can be considered as a good alternative to current in vivo incorporation of UAAs [21,22] due to several advantages over current in vivo processes, such as possibility to direct all of the cellular resources towards the production of a single protein [23]; to control the level of an orthogonal aaRS/tRNA pair and of the UAA employed in protein expression due to the absence of a cell wall. Another advantage of the in vitro approach is the possible use of aforementioned UAAs which can not be applied in vivo.

Here, we describe a general strategy based on the use of commercially available cell-free expression systems, combined with orthogonal M. jannaschii synthetases and cognate MrnA\textsubscript{UA}s with orthogonal \textit{MrnA}\textsubscript{UA}s Opt\textsuperscript{4,24}, to obtain high yields of UAA-labeled proteins. This approach allowed us to incorporate tyrosine, as well as p-acetyl-L-phenylalanine (pAcPhe), p-benzoyl-L-phenylalanine (pBpa) and p-iodo-L-phenylalanine (pPhe; Fig. 1) into GFP mutants.

### Materials and Methods

**GFP Mutants**

The X-ray crystallographic structure of template GFP (PDB accession number: 1EMA), encoded by the GFP control vector (RTS, 5 PRIME, Hamburg, Germany), was analyzed to choose sites for UAA incorporation. To examine the effect of the nucleotide following the stop codon on protein yields, we selected four amino acid residues on two external β-sheet of GFP and its adjacent loop. The selection of these residues for substitution by an alternative for in vivo translation by using an efficient cell-free system combined with optimal \textit{MrnA}\textsubscript{UA}s and cognate aaRSs. The final yield of modified proteins varied widely and under optimal conditions reached roughly 50–120% of the wild type expression levels, depending on the type of suppressor tRNA, aaRS used and UAA used.

![ Figure 1. Structures of tyrosine and the three unnatural amino acids used in this study.](doi:10.1371/journal.pone.0068363.g001)
**Preparation of MjtRNA_CUA, trRNA_CUA Opt***

cDNA copies of all the RNA molecules under regulation of the T7 promoter were obtained by annealing the following two synthetic oligonucleotides: MjtRNA_CUA Forward 5'-TTACATGTGAGTCGTATTA-3'; and trRNA_CUA Opt 5'-TGGTCCGGCGGGCCGGAGTTAAGCCGCCATGCGGACC-3' and MjtRNA_CUA Reverse 5'-TGGTCCGGCGGGCCGGAGTTAAGCCGCCATGCGGACC-3' and trRNA_CUA Opt 5'-TGGTCCGGCGGGCCGGAGTTAAGCCGCCATGCGGACC-3' and trRNA_CUA Opt 5'-TGGTCCGGCGGGCCGGAGTTAAGCCGCCATGCGGACC-3'. Obtained proteins were used for in vitro transcription by TranscriptionAid High Yield Transcription Kit (Fermentas, Vilnius, Lithuania). Purified RNA molecules were heated to 90 °C for 2 minutes, placed on ice for 2 minutes, and folded by addition of RNA Structuring Buffer (10 mM Tris-HCl, pH 7, 0.1 M KCl, 10 mM MgCl) and kept at 37 °C for 20 minutes.

**In-Vitro Translation with Unnatural Amino Acids**

Following UAA incorporation using the cell-free expression system, 5 μL of reaction solution were precipitated with acetone to avoid protein aggregation. The protein pellet was resuspended in an equal volume of LDS loading buffer (Invitrogen). The resulting solution was incubated at 70 °C for 10 minutes and resolved by SDS-PAGE. Proteins of interest were visualized by Coomassie staining using SimplyBlue SafeStain (Invitrogen). For Western blot analysis, proteins were transferred to a nitrocellulose membrane using Immuno-Blot PVDF membrane apparatus (Bio-Rad, Hercules, CA). To visualize proteins in the Western blot, we used primary mouse monoclonal IgG directed against His-tag (Santa Cruz Biotechnology, Santa Cruz, CA) and horseradish peroxidase-conjugated goat polyclonal secondary antibodies to mouse IgG1 heavy chain (Abcam, Cambridge, UK). Chemiluminescence was detected with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, Rockford, IL). Absolute quantities of expressed proteins were estimated by purification of WT and mutated GFP by Ni-NTA agarose (Qiagen), concentration by Amicon Ultra-centrifuge device 10 kDa MWCO and resolved by SDS-PAGE. Proteins of interest were visualized by Coomassie staining using SimplyBlue SafeStain (Invitrogen), the desired proteins bands were cut from the gel and eluted using Model 422 Electro-Eluter (Bio-Rad). The concentration of proteins was measured by Implen Nanophotometer (Labfish, Germany). The relative quantity of expressed proteins was analyzed by densitometry using GeneTools software (SynGene, Cambridge, UK). All of the proteins expressed using the RTS 100 E. coli HY Kit were purified according to the manufacturer's manual using Ni-NTA Spin Columns (Qiagen, Hilden, Germany).

**Preparation of M. jannaschii Tyrosyl-tRNA Synthetase and its Evolved Derivatives**

E. coli BL21 (DE3) cells transformed with one of the above plasmids were grown to an OD600 of 0.5–0.7 in 1 L Luria-Bertani (LB) medium. Isopropl-B-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM and the cells were grown for additional 4–5 h at 37 °C. Cells were harvested at 8,000 g for 10 min at 4 °C. The cell pellet was resuspended in 4 mL of lysis buffer (300 mM NaCl, 10 mM imidazole, 50 mM NaH2PO4, pH 8.0) per gram of cell paste. A cell lysate was prepared using BugBuster Protein Extraction Reagent (Novagen, Darmstadt, Germany) with addition of benzamidine nuclease (Novagen) and Protease Inhibitor Cocktail Set III (Merck, Darmstadt, Germany). The lysate was centrifuged at 16,000 g at 4 °C for 30 minutes. The His-tagged synthetase was then purified using Ni-NTA agarose (Qiagen). The Ni-NTA agarose beads were washed twice with wash buffer (300 mM NaCl, 20 mM imidazole, 50 mM NaH2PO4, pH 8.0) and the protein was eluted with elution buffer (300 mM NaCl, 250 mM imidazole, 50 mM NaH2PO4, pH 0.0). The eluate was dialyzed against sterile PBS buffer pH 7.4 three times and concentrated using an Amicon Ultra-centrifuge device 10 kDa MWCO (Millipore, Beverly, MA).

**Coupled in vitro Transcription/translation Reaction**

Cell-free protein expression was performed using the RTS 100 E. coli HY Kit (5 PRIME, Hamburg, Germany) at 30 °C for 6 h in 10 (for Western blot) or in 50 μL (for protein purification and MS analysis) of reaction mixture. Expression of GFP with an incorporated UAA was achieved by mixing the RTS 100 E. coli HY Kit reaction mixture containing 0.5 μL of modified control vector GFP with purified M. jannaschii aaRS derivatives (100–450 μg/mL, final concentration) and orthogonal suppressor either MjtRNA_CUA or trRNA_CUA Opt (490–600 μg/mL) in the absence or presence of the corresponding UAA (1 mM).

**Proteins Quantitative Analysis and Purification**

Proteins were estimated by the manufacturer's manual using Ni-NTA Spin Columns (Qiagen, Hilden, Germany). All of the proteins expressed using the RTS 100 E. coli HY Kit were purified according to the manufacturer's manual using Ni-NTA Spin Columns (Qiagen, Hilden, Germany).
were extracted from the gel slices by washing once with 25 mM NH₄HCO₃, ACN and 1% formic acid (FA). The samples were then dried in a vacuum centrifuge. The extracted peptides were purified and concentrated using ZipTip pipette tips (Millipore), following the manufacturer’s instruction.

EASI-MS/MS analysis was performed using reverse phase nano-LC (Agilent Technology) connected directly to the LTQ XL Orbitrap ETD mass spectrometer (Thermo Electron, Wien, Austria) at the Analytical Research Services & Instrumentation Unit, BGU. The peptides were eluted with an increasing ACN gradient (Solvent A, 0.1% FA, 5% ACN; Solvent B, 0.1% FA, 80% ACN) over a period of 70 min. MS/MS spectra were acquired in a data-dependent fashion. Instrument control was performed using the Xcalibur software package (Thermo Electron).

Theoretical monoisotopic masses for the peptides generated by trypsin digestion of WT GFP were predicted with PeptideMass (http://web.expasy.org/peptide_mass/), while fragmentation of the FSVSGEGEGDATYGK peptide and theoretical molecular masses of the peptide species were calculated with the MS-Product software at the ProteinProspector web service (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct). Theoretical molecular masses for peptides containing UAA were adjusted manually.

Results

Site-specific Incorporation of Tyrosine into GFP in Response to a UAG-stop Codon in a Cell-free Expression System

To incorporate tyrosine in response to TAG stop codon we expressed plasmid GFP Y39TAG obtained by site-directed mutagenesis in the RTS 100 E. coli HY Kit mixture supplied with external components, purified Mj TyrRS and a suppressor Mj tRNACUA we employed GFP, encoded by a control vector of the kit, as a reporter protein. Western blot with anti His-antibodies enabled to visualize full-length and not truncated GFP at a size of 28 kDa, as well as Mj TyrRS of 36 kDa (Fig. 2A). The addition of purified Mj TyrRS and synthetic Mj tRNACUA to the reaction mixture permitted site-specific incorporation of tyrosine in response to the stop codon at GFP Y39TAG-mutated proteins, while no bands corresponding in size to GFP were detected in the reaction mixture supplied only with Mj TyrRS, indicating orthogonality of M. jannaschii synthetase to endogenous tRNA molecules. Since estimated band intensity corresponding to GFP Y39TAG did not exceed 10% of the WT expression level, we further adjusted Mj TyrRS and Mj tRNACUA concentrations in the reaction mixture. The synthetase concentration required for maximal suppression efficiency was found to vary widely and depended on the concentration of Mj tRNACUA (Fig. 2B). At a Mj TyrRS concentration equal to 400–450 μg/mL and in the presence of 60 μg/mL Mj tRNACUA, the expression level of Y39TAG GFP reached the maximum possible under these conditions and was roughly 20% of WT expression level, while in the presence of 450 μg/mL Mj tRNACUA, the maximal expression level reached 35% of WT expression level and required only 100 μg/mL of Mj TyrRS. To determine whether the suppressor tRNA structure is limiting, we sought another suppressor. We then adjusted the concentration of both Mj tRNACUA and T-stem-modified tRNACUA for optimal tyrosine incorporation. Using both the original and alternate suppressor, the expression of full-length GFP was demonstrated to depend greatly on the nonsense suppressor concentration (Fig. 2C). A maximum yield of Y39TAG GFP constituting 55% and 115% of the WT expression level was achieved at Mj tRNACUA concentration of 600 μg/mL and of a tRNACUAOpt concentration of 480 μg/mL, respectively. This experiment revealed that the nature of suppressor tRNA significantly affects the efficiency of TAG nonsense suppression in recombinant proteins, while application of tRNACUAOpt allows one to reach at least similar protein yields as that of WT protein.

The Effect of the Nucleotide following the Stop Codon on the Expression of UAA-containing Proteins

As it is mentioned above, the TAG stop codon specifying the desired position for UAA incorporation into a recombinant protein could be recognized by either RF1 or by the cognate suppressor tRNA. It has been shown that the identity of the nucleotide following nonsense codon impinge on the selection rate of RF1 [12], i.e. a low rate of stop signal recognition by RF1 means that mRNA interaction with near-cognate aminoaeryl-tRNA or frame-shifting occurs faster than does RF1 binding to the ribosomal A-site [12]. Based on the literature screened [17,19,22,25,26], we suggested that the forth nucleotide in tetranucleotide stop signal could affect the rate of nonsense suppression [12,13]. To verify this notion, we selected several amino acids in two different GFP β-sheets and their adjacent loop that were substituted to TAG, the location was selected in a way that afforded that the immediate nucleotide downstream from the stop codon differed from one location to another. Thus, we constructed plasmids encoding GFP Y39TAG, K41TAG, L42TAG and K45TAG mutants, where the amber codon was followed by a guanine (G), a cytosine (C), an adenine (A) and a thymine (T), respectively; and GFP H148TAG, N149TAG, V150TAG and Y151TAG were stop codon was followed by A, G, T and C, respectively. All these constructs were expressed using the cell-free expression kit supplemented by Mj TyrRS (150 μg/mL in the absence or presence of either Mj tRNACUA or tRNACUA by (480 μg/mL). Western blot analysis (Fig. 3A, 3B) revealed that the expression level of full-length GFP depended on the specific nucleotide following TAG stop codon in the cell-free protein translation system based on an E. coli lysate. The fourth nucleotide hierarchy for efficient suppression was demonstrated to be A>G>C>T for both variants of nonsense suppressor tRNAs. A literature screen revealed that E. coli-based cell-free protein synthesis had been successfully employed for UAA incorporation into proteins when the TAG stop codon was followed by G [22,26], A [19,25], and C [22]. To the best of our knowledge, no expression of protein containing UAA has been reported when T followed the stop codon. To demonstrate that indeed the “following base” and not position change in the protein affects the rate of suppression efficiency, we have substituted K41Y, L42Y and K45Y to tyrosine and tested GFP expression level for these mutants relative to the WT. It should be noted that, the fourth nucleotide after the tyrosine codon remained as in the native sequence: G after the codon in the native protein GFP29Y, C after GFP K41Y, A after GFP L42Y and T after K45Y. Western blot analysis of these control mutants revealed that expression levels did not differ from those of WT and from one another (Fig. 3C), verifying that indeed the effects that we have observed imply context dependence.

Genetic Incorporation of UAA in Response to the Amber Stop Codon

To test the generality of the developed platform, we examined its ability to incorporate diverse UAA at position 39 of GFP in response to the TAG stop codon, applying both types of
Figure 2. Western Blot of WT GFP and GFP Y39TAG mutant expression in a cell-free translation system. Synthesis of WT GFP and the GFP Y39TAG mutant was performed using the RTS E. coli HY Kit, to which the corresponding plasmid (500 μg/mL), purified MjTyrRS and cognate suppressor MjtRNA_{CUA} (tRNA) or T-stem modified tRNA_{CUA}Opt (denoted as *) were added. (A) Expression of WT GFP and the GFP Y39TAG mutant in the presence of MjTyrRS (300 μg/mL) and synthetic MjtRNA_{CUA} (60 μg/mL). The band at 28 kDa corresponds to full-length GFP. (B) Western blot analysis demonstrates enhanced GFP Y39TAG protein expression as a function of increased MjTyrRS concentrations in a cell-free reaction medium supplied with MjtRNA_{CUA} (60 μg/mL – top panel and 450 μg/mL – bottom panel). (C) Dependence of GFP Y39TAG yield on the type and concentration of nonsense suppressor, as visualized by Western blot. doi:10.1371/journal.pone.0068363.g002

Figure 3. Cell-free expression of WT GFP and tyrosine-incorporating mutant GFP, as visualized by Western blot. (A and B) Co-translational incorporation of tyrosine at different positions in response to the amber stop codon was achieved by adding purified MjTyrRS (200 μg/mL) and two types of suppressor tRNA (480 μg/mL) to the reaction mixture (tRNA denotes synthetic MjtRNA_{CUA}, *– tRNA_{CUA}Opt). (C) Western blot visualization of the expression level of GFP WT and tyrosine-substituted proteins. doi:10.1371/journal.pone.0068363.g003
suppressor tRNAs and three variants of MjTyrRS derivatives. The three evolved variants of M. jannaschii aaRS, i.e. AcRS [27], BpaRS [28] and IPheRS [29], were tested for the ability to suppress the amber stop codon in GFP Y39TAG mutants together with either MjtRNACUA or tRNACUA Opt in the absence or presence of their cognate UAA in a cell-free translation system. The expression of full-length GFP Y39TAG was shown (Fig. 4A and 5A) to depend on the presence of pBpa and pIPhe. GFP expression was not detected in the absence of pBpa and pIPhe. Although AcRS has been widely used for site-specific protein labeling in vivo [17,30,31], its application in cell-free reaction medium led to background suppression in the absence of pAcPhe (Fig. 6A). The reason for background suppression in vivo is from mis-acylation of the suppressor tRNA molecules by the evolved synthetase with an endogenous amino acid, such as tyrosine or phenylalanine, in the rich media [17]. The overall level of background suppression was estimated to be less than 2 and 4.5% of GFP WT expression level for MjtRNACUA and tRNACUA Opt, respectively; however, since the main disadvantage of using previously reported eukaryotic-based cell-free systems for UAA incorporation was a high degree of mis-acylation with endogenous amino acids [21], site-specifically modified GFP Y39TAG were further characterized by mass spectrometry (MS).

The relative yields of GFP Y39TAG protein with site-specifically incorporated pIPhe and pAcPhe were measured to be 45–52% for the reactions utilizing MjtRNACUA and an approximately 85% for tRNACUA Opt (Fig. 5A and 6A). The highest efficiency of UAA incorporation was obtained by applying BpaRS to the cell-free reaction mixture; the expression level of GFP Y39-mutated proteins comprised an approximately 80 and 120% of GFP WT level for MjtRNACUA and tRNACUA Opt, respectively (Fig. 4A).

Mass Spectrometry Analysis to Determine Specificity and Fidelity of Incorporation

To examine the specificity and fidelity of M. jannaschii aaRS derivatives, WT GFP and GFP Y39TAG were expressed using a cell-free translation system, purified and separated by SDS-PAGE. The bands corresponding in size to GFP, as visualized by Coomassie staining, were excised, trypsin-digested and analyzed with a LTQ XL Orbitrap ED mass spectrometer. The peptide of interest was predicted by the PeptideCutter software to be FSVSGEGGDATY*GK, where Y* denotes either tyrosine in

![Figure 4. Site-specific Bpa incorporation into GFP in a cell-free expression system.](image)

(A) Western blot visualization of WT GFP and Bpa-incorporating GFP Y39TAG. The synthesis of GFP was performed using an in vitro translation kit. Co-translational incorporation of Bpa was achieved by addition of M. jannaschii BpaRS (100 μg/mL), different types of suppressor tRNA (480 μg/mL) and Bpa (1 mM) to the reaction medium. tRNA denotes synthetic MjtRNACUA Opt – tRNACUA Opt. (B) Annotated MS/MS spectrum of the FSVSGEGGDATY*GK peptide from GFP Y39Bpa. The mass shift of 88 Da between Bpa in GFP Y39Bpa and tyrosine in WT GFP is clearly observed. For clarity, only the most abundant "y" ions are assigned. doi:10.1371/journal.pone.0068363.g004
WT GFP or UAA in GFP Y39TAG. The monoisotopic mass calculated by PeptideMass software for this peptide generated by trypsin digest of the WT protein was 1503.6597 Da; the mass of the same peptide from GFP Y39TAG was adjusted manually (depending on the incorporated UAA mass). MS analysis of WT GFP identified the desired peptide, showing a doubly charged ion \([\text{M}^+2]\) peak of \(m/z\) 752.829, in a good agreement with the calculated mass. Peaks observed in the average MS/MS spectrum (Fig. 7) were also in a good agreement with predicted masses (calculated with the ProteinProspector MS-Product software) of “y” and “b” ions generated by FSVSGEGEGDATY*GK peptide fragmentation (Table S1). MS analysis of GFP Y39TAG revealed peaks with doubly charged ion masses of 765.84 (calculated mass, 765.85), 807.79 (calculated mass, 807.78) and 796.85 (calculated mass, 796.88) for FSVSGEGEGDATY*GK peptides containing pAcPhe, pBpa and pIPhe, respectively. Although Western blot revealed some degree of background suppression of amber stop codon of GFP Y39TAG obtained by using AcRS in the absence of pAcPhe, no peptide signals corresponding to GFP sequence containing tyrosine or phenylalanine at position 39 were detected, indicating that similarly to the observed in vivo process [17,27] misacylation of suppressor tRNA by near-cognate endogenous amino acids was inhibited by addition of cognate UAA. Addition of higher concentrations of the UAA to the reaction, eliminate the competition fully, as no molecules with \(m/z\) equivalent to trypsin digested WT GFP were found for proteins obtained by addition of UAA as the cell-free reaction. Moreover, no peptides containing canonical amino acids from near-cognate suppression of stop codon were detected, confirming good selectivity and fidelity of the analyzed \(Mj\) aaRSs. MS/MS analysis of GFP Y39TAG peptides (Fig. 4B, 5B and 6B) demonstrated a characteristic mass shift of 88.11, 109.9 and 26.04 Da relative to WT GFP, values that exactly match the differences between tyrosine and pBpa, pIPhe and pAcPhe, respectively. Mass shifts were also detected for peaks corresponding to “y” ions (from \(y_3\) to \(y_{14}\)), as well as for \(b_{13}\) ion, indicating the site of UAA incorporation to be position 39 of GFP. In general, we did not detect any differences in the GFP Y39TAG proteins obtained by utilization of either \(Mj\) tRNACUA or tRNACUA \(^{Opt}\), apart from the different protein yields.

**Discussion**

Since it has been shown that protein synthesis is a ribosome-mediated process that does not require cell integrity, cell-free protein expression systems have been used for a variety of purposes, including site-specific UAA incorporation into recom-
binant proteins [21,26]. In our efforts to develop an *E. coli*-based cell-free protein expression system to produce high yields of UAA-containing recombinant proteins, we employed the most established *M. janaschii* orthogonal synthetases and nonsense suppressor molecules. In contrast to previously described cell-free expression systems adapted for protein synthesis with encoded UAAs, we employed purified *Mj* TyrRS and its derivatives along with suppressor tRNA as additional components of defined reaction mixture allowing for control of the protein synthesis environment. In this manner, we were able to produce GFP Y39TAG with tyrosine with an absolute yield of 270 \( \mu \text{g/mL} \) and a suppression efficiency of 55% by adjusting *Mj* TyrRS and *Mj* tRNACUA concentrations. Although suppressor tRNA concentration is the major factor limiting production of proteins containing UAAs, further augmentation of the proportion of *Mj* tRNACUA in the reaction medium (exceeding a final concentration of 600 \( \mu \text{g/mL} \)) led to the inhibition of recombinant protein biosynthesis, presumably because of translational apparatus overloading with non-endogenous elements. Replacement of the orthogonal *Mj* tRNACUA suppressor by T-stem-modified tRNACUA optimized for efficient recognition and binding by *E. coli* EF-Tu factor resulted in further enhancement of recombinant protein yields, which were estimated to be 120% of WT GFP expression. It was previously shown that the degree of improvement in suppression efficiency of evolved tRNACUA varied, depending on the specific aaRS and cognate UAA used [17,24]. We have also observed the ability of modified tRNACUA to suppress amber stop codon *in vitro* with different efficiencies, depending on the UAA being incorporated. The application of optimized suppressors in a cell-free reaction medium for most aaRSs resulted in an enhanced protein yield, ranging from 85 to 110% of WT levels. Still, we cannot exclude the possibility that some of the evolved *Mj* TyrRSs could have lower affinity to such a suppressor. For both tRNA molecules, the high fidelity of *M. jannaschii* aaRS derivatives in the cell-free expression system was confirmed by mass spectrometry. The efficiency of UAAs incorporation in response to a stop codon is known to depend on the position of the mutation site and the nature of the recombinant protein, in particular the encoded amino acids and corresponding codon surrounding the amber codon. Although this phenomenon is usually taken into consideration when designing experiments, the precise reason and

Figure 6. Site-specific pAcPhe incorporation into GFP in a cell-free expression system. (A) Western blot visualization of WT GFP and pAcPhe-incorporating GFP Y39TAG. The synthesis of GFP was performed using an *in vitro* translation kit. Co-translational incorporation of pAcPhe was achieved upon addition of the corresponding plasmid (500 \( \mu \text{g/mL} \)) *M. jannaschii* pAcPheRS (100 \( \mu \text{g/mL} \)), different types of suppressor tRNA (480 \( \mu \text{g/mL} \)), and pAcPhe (1 mM) to the reaction medium. tRNA denotes synthetic *Mj* tRNACUA Opt – tRNACUA Opt. (B) Annotated average MS/MS spectrum of peptides with \( m/z \) 765.84 corresponding to pAcPhe-incorporated FSVSEGEDGATY*GK peptide of GFP Y39TAG. The Y* ion corresponding to pAcPhe (calculated \( m/z \), 393.24; observed \( m/z \), 393.23) can be easily detected. The distinguishing mass shift of 26 Da can be observed for most abundant “y” ions. For clarity, only the most abundant “y” ions are assigned.
doi:10.1371/journal.pone.0068363.g006
mechanism leading to this observation was not reported. It is well known that the identity of nucleotide following stop codon influence the efficiency of translation termination. Depending on the nucleotide downstream the certain stop codon, the decoding site of 16S ribosomal RNA (rRNA) favors either translational termination by binding to the RF-factors, “read-through” of stop codon by augmentation of near-cognate aminoacyl-tRNA binding or a “frame-shift” [12,32]. We hypothesized that the effectiveness of suppressor tRNA binding to its cognate nonsense codon would also depend on the following nucleotide. The expression of GFP Y39TAG, K41TAG, L42TAG and K45TAG mutants (where the fourth nucleotide was G, C, A, and T, respectively), and of GFP H148TAG, N149TAG, V150TAG and Y151TAG (where the fourth nucleotide was – A, G, T and C, respectively) demonstrated that the identity of the base following the amber codon determined the efficiency of UAA-charged suppressor MjtRNACUA or tRNACUA Opt interaction with UAG stop codon and, as a result, overall protein yields. According to our study, the strength of UAG stop codon selection and interaction was predicted by the fourth base hierarchy to be A>G>C>T for both MjtRNACUA and tRNACUA Opt. The purines as the forth nucleotide improve the termination in the presence of RF1, however, we suggest that in the cell-free reaction medium prepared by us concentration of suppressor tRNA exceeds the concentration of RF1 many times rendering RF1 unavailable to the ribosomes; under these conditions, a purine nucleotide downstream the stop codon would augment UAG interaction with its cognate suppressor tRNA over the possibilities of mis-acylation or frame-shift. Analysis of the literature failed to find an example of the incorporation of any UAA in response to the amber stop codon followed by a T. Instead, the base following the stop codon was either G or A for the majority of UAA-containing proteins produced in vitro and in vivo that we have screened. These findings indirectly confirm our observation concerning the effect of the nucleotide downstream of the stop codon on the efficiency of UAA incorporation. Nonetheless, the actual mechanism by which the fourth base modulates the efficiency of stop codon suppression has yet to be revealed. In addition we would like to stress that a statistical analysis was not performed, hence we base our conclusions only on experimental evidence in our study combined with a literature screen.

It is also important to note that all of the reported protein yields are not absolute quantities and are reported as relative values and as percent of the WT expression levels, which were set to 100% in this study.
Conclusions

The cell-free translation system, as modified here to reportedly encode proteins containing UAAs resulted in increased amounts of recombiant proteins with very good fidelity. Concentrations of added UAAs with cognate RSs that have shown lower fidelity, can be tuned in this system in a controlled manner, thus eliminating possible competition of incorporation of natural amino acids. Competition of suppressor tRNA with RF1 can be reduced significantly by using controlled and higher concentrations of suppressor tRNA, thus affording higher suppression efficiencies. The ability to control the concentrations of the different orthogonal components in this system afford reduced competition from natural components in the translational machinery.

The major advantage of the methodology reported here is its generality. Due to the availability of commercial cell-free translation systems with variety of modifications, it is possible to produce both prokaryotic and eukaryotic UAA-encoded proteins. The nature of the in vitro approach enables one to incorporate UAAs into nascent polypeptides that are not available for living organisms, provided that the right aaRS is available. It is also our belief that through this approach, more than one UAA could be incorporated into a protein with only a small loss in protein yield.

References

1. Wang Q, Parrish AR, Wang L. (2009) Expanding the genetic code for biological studies. Chem Biol 16: 323–336. 10.1016/j.chembiol.2009.03.001.
2. Xie J, Schultz PG (2006) A chemical toolkit for proteins - an expanded genetic code. Nat Rev Mol Cell Biol 7: 775–782. 10.1038/nrm2005.
3. Liu DR, Schultz PG (1999) Progress toward the evolution of an organism with an expanded genetic code. Proc Natl Acad Sci U S A 96: 4780–4785.
4. Wang L, Magliery TJ, Liu DR, Schultz PG (2000) A new functional suppressor tRNA/Aminoseryl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J Am Chem Soc 122: 5010–5011. 10.1021/ja000305y. Available: http://dx.doi.org/10.1021/ja000305y via the Internet.
5. Bain JD, Diala ES, Glabe CG, Dix TA, Chamberlin AR (1989) Biosynthetic site-specific incorporation of a non-natural amino acid into a polysaccharide. J Am Chem Soc 111: 8013–8014. 10.1021/ja00202a052. Available: http://dx.doi.org/10.1021/ja00202a052.
6. Noren C, Anthony-Cahill S, Griffith M, Schultz P (1989) A general method for specific incorporation of nonnatural amino acids into proteins. Science 244: 182–188. 10.1126/science.2449509.
7. Hohsaka T, Ashizuka Y, Murakami H, Siido M (2001) Five-base codons for incorporation of nonnatural amino acids into proteins. Nucleic Acids Res 29: 3646–3651.
8. Hohsaka T, Ashizuka Y, Murakami H, Siido M (1996) Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. J Am Chem Soc 118: 9778–9779. 10.1021/ja9614225. Available: http://dx.doi.org/10.1021/ja9614225.
9. Siido M, Hohsaka T (2001) Introduction of specialty functions by the position-specific incorporation of nonnatural amino acids into proteins through four-base codon/anticodon pairs. Appl Microbiol Biotechnol 57: 274–281.
10. LaRiviere JF, Wollons AD, Ulsenbeck OC (2001) Uniform binding of an expanded genetic code. Nat Biotechnol 19: 1094–1098. 10.1038/nbt1091.
11. Park HS, Hohn MJ, Umehara T, Guo LT, Osborne EM, et al. (2011) Expanding the genetic code of Escherichia coli with phosphosine. Science 333: 1151–1154. 10.1126/science.1207263.
12. Poole ES, Brown CM, Tate WP (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. J Mol Biol 245: 131–140.
13. Poole ES, Major LL, Manning SA, Tate WP (1998) Translational termination in Escherichia coli: Three bases following the stop codon crosstalk to release factor 2 and affect the decoding efficiency of UGA-containing signals. Nucleic Acids Res 26: 595–601.
14. Rys Y, Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. Nat Methods 3: 263–265. 10.1038/nmeth0606.
15. Schrader JM, Chapman SJ, Ulsenbeck OC (2009) Understanding the sequence specificity of tRNA binding to elongation factor Tu using δRNA mutagenesis. J Mol Biol 386: 1255–1264.
16. Eargle J, Black AA, Sethi A, Trabuco LG, Ludhey-Schulten Z (2006) Dynamics of recognition between tRNA and elongation factor Tu. J Mol Biol 377: 1302–1405. 10.1016/j.jmb.2006.01.073.

Supporting Information

Table S1 Calculation of “g” and “b” ions of the FSVSGEGEGDATY*GK fragment (“Y” denotes either tyrosine in WT GFP or UAA in the GFP Y39TAG mutants). Masses for the WT GFP-derived FSVSGEGEGDATY*GK peptide fragmentation were predicted by the MS-Product software of the ProteinProspector web service, while masses for GFP Y39TAG mutants were adjusted manually. (DOC)

Acknowledgments

Mr. Ziv Roth and Dr. Isam Khalaila are thanked for their help in analyzing the MS data. Prof. Peter G. Schultz is gratefully acknowledged for generously supplying us with the pBSup plasmid.

Author Contributions

Conceived and designed the experiments: LA ZJZ SS. Performed the experiments: SS. Analyzed the data: SS. Contributed reagents/materials/analysis tools: LA. Wrote the paper: SS LA.

In-Vitro Translation with Unnatural Amino Acids

July 2013 | Volume 8 | Issue 7 | e68363