

Video Article

Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an *Escherichia coli* Cell-free Transcription-translation System

Emanuel G. Worst¹, Matthias P. Exner², Alessandro De Simone², Marc Schenkelberger¹, Vincent Noireaux³, Nediljko Budisa², Albrecht Ott¹¹Department of Experimental Physics, Saarland University²Institute of Chemistry, Technische Universität Berlin³School of Physics and Astronomy, University of MinnesotaCorrespondence to: Albrecht Ott at albrecht.ott@physik.uni-saarland.deURL: <http://www.jove.com/video/54273>DOI: [doi:10.3791/54273](https://doi.org/10.3791/54273)Keywords: Molecular Biology, Issue 114, Cell-Free Expression, Noncanonical Amino Acid Analog, Residue-Specific Incorporation, L-Arginine Analog, L-Canavanine, Bioengineering, Synthetic Biology, *Escherichia coli* Cell Extract, Unnatural Amino Acid

Date Published: 8/1/2016

Citation: Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., Ott, A. Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an *Escherichia coli* Cell-free Transcription-translation System. *J. Vis. Exp.* (114), e54273, [doi:10.3791/54273](https://doi.org/10.3791/54273) (2016).

Abstract

The canonical set of amino acids leads to an exceptionally wide range of protein functionality. Nevertheless, the set of residues still imposes limitations on potential protein applications. The incorporation of noncanonical amino acids can enlarge this scope. There are two complementary approaches for the incorporation of noncanonical amino acids. For site-specific incorporation, in addition to the endogenous canonical translational machineries, an orthogonal aminoacyl-tRNA-synthetase-tRNA pair must be provided that does not interact with the canonical ones. Consequently, a codon that is not assigned to a canonical amino acid, usually a stop codon, is also required. This genetic code expansion enables the incorporation of a noncanonical amino acid at a single, given site within the protein. The here presented work describes residue-specific incorporation where the genetic code is reassigned within the endogenous translational system. The translation machinery accepts the noncanonical amino acid as a surrogate to incorporate it at canonically prescribed locations, *i.e.*, all occurrences of a canonical amino acid in the protein are replaced by the noncanonical one. The incorporation of noncanonical amino acids can change the protein structure, causing considerably modified physical and chemical properties. Noncanonical amino acid analogs often act as cell growth inhibitors for expression hosts since they modify endogenous proteins, limiting *in vivo* protein production. *In vivo* incorporation of toxic noncanonical amino acids into proteins remains particularly challenging. Here, a cell-free approach for a complete replacement of L-arginine by the noncanonical amino acid L-canavanine is presented. It circumvents the inherent difficulties of *in vivo* expression. Additionally, a protocol to prepare target proteins for mass spectral analysis is included. It is shown that L-lysine can be replaced by L-hydroxy-lysine, albeit with lower efficiency. In principle, any noncanonical amino acid analog can be incorporated using the presented method as long as the endogenous *in vitro* translation system recognizes it.

Video Link

The video component of this article can be found at <http://www.jove.com/video/54273/>

Introduction

The genetic code is universal to the biosphere. It codes for a set of 20 canonical amino acids, which is sometimes extended by selenocysteine¹ or pyrrolysine². It is the ribosome that translates the genetic code with the help of tRNAs into chains of amino acids that fold into proteins. The functional groups of the canonical amino acids, in combination with posttranslational modifications, contribute to an exceptionally wide range of protein function^{3,4}. In principle, functional limitations due to the limited set of canonical amino acids can be overcome by incorporating further, noncanonical amino acids (ncAAs) that enable new chemistries and new functionalities^{3,4}.

There are two complementary approaches for the incorporation of ncAAs: the site- or the residue-specific incorporation. The former method entails considerable technical difficulties, since the canonical set of aminoacyl-tRNA-synthetases (aaRS) and tRNAs must be expanded by an orthogonal aaRS-tRNA pair that must not interact with the endogenous translation machinery. Based on careful engineering, this approach incorporates the ncAAs as single point mutations at the desired protein sites. Site-specific incorporation of ncAAs is genetically encoded by a codon that is not assigned to a canonical amino acid (cAA), usually a stop codon⁵⁻⁹. This method entails changes in function at a given site rather than across the entire protein¹⁰⁻¹³.

In contrast, residue-specific incorporation relies on erroneous recognition of the noncanonical amino acid by the canonical translation machinery. The incorporation occurs due to the lack of substrate specificity of the aaRS. The residue-specific incorporation of ncAAs, built on the work of Cohen and coworkers¹⁴, has led to important applications^{3,10}, among them bio-orthogonal labeling¹⁵⁻¹⁷ of proteins or structure elucidation of proteins in X-ray crystallography¹⁸.

As natural aaRS generally prefer their cognate amino acid over an isostructural ncAA, efficient *in vivo* residue-specific incorporation usually requires an auxotrophic expression host not capable of synthesizing the canonical analog of the ncAA. The host cells are cultivated in growth medium that delivers only a low concentration of the analogous cAA. Its exhaustion in combination with the consecutive supplementation with the ncAA forces the expression host to incorporate the ncAA into the model protein at multiple, canonically prescribed sites. In contrast to the site-specific approach, this generally has a deep impact on the entire protein structure, leading to considerably modified physical and chemical properties of proteins^{19,20}. However, most of the nCAAs are growth inhibitors for the expression host³, as they are incorporated into many other proteins besides those of interest during recombinant gene expression. This clearly limits the *in vivo* approach. The *in vivo* incorporation of amino acids that are toxic or have strong influence on the protein structure remains particularly challenging. However, these molecules are among the most promising to engineer proteins with extraordinary functions.

One example is the toxic, noncanonical, naturally occurring L-canavanine (Can), an analog of L-arginine (Arg). It affects and blocks Arg associated regulatory and catalytic reaction pathways, and its presence in the living cell can lead to immediate death^{3,21-23}. Its incorporation into proteins at arginine positions can reduce protein stability²¹⁻²³. Due to the resulting toxicity, expression of canavanine containing proteins in *Escherichia coli* (*E. coli*) and other common expression hosts remains a challenge. For these reasons, complete *in vivo* incorporation of Can at all Arg positions has appropriately been confirmed only once²⁴, using an elaborated single-protein production system. However, Can has been proposed as an anti-cancer agent²⁵⁻²⁷, and as a stimulator for autoimmune diseases in humans²⁸. Additionally, it is subject of various studies on its anti-metabolic, antibacterial, antifungal and antiviral properties²⁵. These properties raise a demand for efficient and easy-to-perform methods to express Can containing proteins for pharmaceutic, medical and functional studies.

Although many problems that are connected to *in vivo* production can be circumvented using cell-free expression systems, *in vitro* residue-specific approaches have only been poorly explored. The cell-free residue-specific incorporation of an L-tryptophan analog²⁹ and multiple nCAAs³⁰ have been reported. These methods are based on the highly efficient T7 RNA polymerase. The T7 RNA polymerase entails bacteriophage-like transcription, thereby reducing genetic functionality in comparison to endogenous transcription.

The complete residue-specific incorporation of Can into a model protein at all Arg positions was recently reported³¹, using a cell-free expression system³². A slight modification of the same system enabled site-specific incorporation of different pyrrolysine analogs into a model protein via stop codon suppression³³. The employed cell-free system³¹⁻³³ is based on an all *E. coli* transcription-translation system. Nevertheless, it enables protein expression as efficiently as in current bacteriophage systems (0.5 - 1 mg/ml of recombinant protein)³², while retaining much of the original transcription-translation modularity.

In this work, a detailed protocol is provided on how the residue-specific incorporation of nCAAs can be realized, using this all *E. coli* cell-free system³². Additionally, further steps to prepare the expressed proteins for appropriate evaluation via HPLC-ESI mass spectroscopy are proposed. To expand the properties of this cell-free system, this work does not only refer to the published incorporation of Can³¹ but also presents new data related to the noncanonical L-lysine analog L-hydroxy-lysine.

The following protocol for the residue-specific incorporation of nCAAs is an adaptation of a protocol recently published in JoVE³⁴. The latter protocol describes how to perform highly efficient cell-free expression with standard amino acids. Furthermore, it presents the preparation of the crude cell free extract, the amino acid solution, the energy stock solution and the energy buffer used in this approach. The following protocol focuses on modified steps in comparison to the previous protocol in order to enable the residue-specific incorporation of nCAAs. Calibrated pipets, low-binding pipette tips and micro-centrifuge tubes are recommended for the preparation. In the following, IUPAC abbreviations for the amino acids are used.

Protocol

Caution! Please consult all relevant material safety data sheets (MSDS) before use. Several of the used chemicals are acutely toxic. Personal protective equipment is required (eyeshield, dust mask, gloves, lab coat, full length pants, closed-toe shoes) as well as working in a fume hood.

1. Amino Acid Solution Preparation

1. Stock solution preparation of the ncAA (168 mM)

NOTE: The stock solution preparation of the ncAA is described for the Arg analog Can as an example. Accordingly adapt the values for other nCAAs.

1. Place a 1.5 ml reaction tube onto a microbalance. Weigh out 46.1 mg of Can inside the reaction tube for the preparation of 1 ml of a 168 mM solution. Use a sterile microspatula. For a racemic mixture of the ncAA, double the concentration of the stock solution.
2. Add 977 μ l of sterile ddH₂O. Thoroughly vortex until Can is in complete dissolution.

NOTE: For a total solution volume of 1 ml, the physical volume of the dissolved amino acid has to be compensated. For any of the amino acids, estimate half of the solid mass in mg as the corresponding volume increase in μ l (100 mg solid will take a volume of 50 μ l in the solution)³⁵. Most amino acids can be dissolved at this concentration. If not, reduce the concentration until complete dissolution.

3. Directly use the ncAA stock solution for the preparation of the amino acid solutions in section 1.2 or flash freeze it in liquid nitrogen and store it at -20 °C. **CAUTION!** For safety, wear an eyeshield and Cryo-gloves to be protected from liquid nitrogen splashes.

2. Preparation of the amino acid solutions

NOTE: For the preparation of the amino acid solutions, use the amino acid sampler providing the L-isomers of the 20 cAAs in separate stock solutions (1.5 ml, buffered with HEPES/KOH, < 0.1% NaN₃, pH 7.5), each at a concentration of 168 mM, except for L-Leucine (140 mM). For a homemade preparation of these stock solutions (buffered with KOH), follow this protocol³⁵.

1. Thaw the stock solutions of the 20 cAAs (amino acid sampler or prepared according to³⁵) and of the ncAA (prepared in section 1.1) at RT.

2. After thawing, frequently vortex the stock solutions to redissolve any precipitated amino acids. As some amino acids are harder to dissolve, incubate them in a heating block at 37 °C until complete dissolution. Cys may not fully dissolve. Place all amino acids on ice, except for Asn, Phe and Cys — keep these at RT to avoid precipitation.
3. Use the following values to use one-seventh of the complete kit.
NOTE: Scale down appropriately to work with smaller volumes and to save parts of the kit for further experiments. Scale up for incorporation of ncAAs into model proteins on a large scale. To prevent frequent thawing, which is likely to reduce the stability of the amino acids, aliquot the individual amino acid stock solutions into volumes of 200 µl. This 200 µl aliquot volume and the aliquot volumes used in step 1.2.4.1 account for losses due to pipetting.
4. First, prepare an amino acid master mix solution that will be split in step 1.2.4.3 to finalize the preparation of 3 differently composed amino acid solutions (sections 1.2.5 - 1.2.7). In these solutions, concentrate all amino acids at 6 mM, except for Leu (5 mM).
 1. Transfer 1.4 ml of sterile ddH₂O into a 15 ml centrifuge tube. Put it on ice. Add 175 µl of each amino acid stock solution. Add one after another, **except for the stock solution of the cAA** (e.g., Arg) to be replaced by the ncAA (e.g., Can). Thoroughly vortex after each addition and put the solution back on ice.
NOTE: Leu is at 5 mM in the 3 differently composed amino acid solutions, compared to 6 mM for the other amino acids. The reduced concentration does not reduce the expression efficiency. Scaling up to 6 mM is suitable as well.
 2. Transfer the amino acid stock solutions in the following order to avoid precipitation: Ala, Arg, Asn, Asp, Gln, Glu, Gly, His, Ile, Lys, Met, Phe, Pro, Ser, Thr, Val, Trp, Tyr, Leu, and Cys. Remember **to not add** the stock solution of the cAA (e.g., Arg) that is analogous to the ncAA (e.g., Can). Finally, thoroughly vortex. Incubate at 37 °C to make the solution as clear as possible.
 3. Split this amino acid master mix solution into three equal volumes of 1.35 ml. Transfer each of the split volumes into 1.5 ml reaction tubes. Keep them on ice.
5. Prepare an amino acid solution that consists of all 20 cAAs at a concentration of 6 mM each, except for Leu that is 5 mM. To the first volume of 1.35 ml, as a result of the split in step 1.2.4.3, add 50 µl of the 168 mM stock solution of the cAA (e.g., Arg) that is analogous to the ncAA (e.g., Can). Thoroughly vortex.
 1. Put back on ice. Aliquot this 1.4 ml solution in volumes of 16 µl into reaction tubes. Note that this solution volume approximately leads to 85 aliquots. Label these aliquots "+ cAA" (e.g., + Arg).
 2. Flash freeze the aliquots in liquid nitrogen and store at -80 °C. **CAUTION!** For safety, wear an eyeshield and Cryo-gloves to be protected from liquid nitrogen splashes.
6. Prepare an amino acid solution that is composed of 19 cAAs except for the cAA (e.g., Arg) that is analogous to the ncAA (e.g., Can). Add each amino acid to a concentration of 6 mM, except for Leu (5 mM).
 1. Add 50 µl of sterile ddH₂O to the second volume of 1.35 ml, as a result of the split in step 1.2.4.3. Thoroughly vortex and put back on ice. Aliquot this 1.4 ml solution in volumes of 16 µl into reaction tubes. Note that this solution volume approximately leads to 85 aliquots. Label these aliquots "- cAA" (e.g., - Arg).
 2. Flash freeze the aliquots in liquid nitrogen and store at -80 °C. **CAUTION!** For safety, wear an eyeshield and Cryo-gloves to be protected from nitrogen splashes.
7. Prepare an amino acid mixture that contains 19 cAAs and the ncAA (e.g., Can) that substitutes the canonical one (e.g., Arg). Add each amino acid to a concentration of 6 mM, except for Leu (5 mM). To the last 1.35 ml volume, as a result of the split in step 1.2.4.3, add 50 µl of the 168 mM stock solution of the ncAA (e.g., Can). Label it "+ ncAA" (e.g., + Can). Thoroughly vortex and put back on ice.
 1. Aliquot this 1.4 ml solution in volumes of 16 µl into reaction tubes. Note that this solution volume approximately leads to 85 aliquots. Label these aliquots "+ ncAA" (e.g., + Can).
 2. Flash freeze the aliquots in liquid nitrogen and store at -80 °C. **CAUTION!** For safety, wear an eyeshield and Cryo-gloves to be protected from nitrogen splashes.
NOTE: The 16 µl aliquot volumes used in the steps 1.2.5.1, 1.2.6.1 and 1.2.7.1 are slightly higher than required to account for losses due to pipetting.

2. Energy Buffer Preparation

NOTE: Each batch of crude extract is unique and requires optimized concentrations of Mg- and K-glutamate³⁴. The crude extract aliquot volume depends on the protein concentration³⁴. Use the provided calculation template (Supplemental Material 1) for different values. Find further instructions in Supplemental Material 1 figure legend, explaining how to employ this template.

1. Prepare and store at -80 °C the 14x energy solution and crude extract aliquots according to the unmodified protocol³⁴. Calibrate the crude extract depending on the Mg- and K-glutamate concentrations for optimized expression efficiency³⁴.
NOTE: The final composition of 14x energy solution is: 700 mM HEPES (pH 8), 21 mM ATP, 21 mM GTP, 12.6 mM CTP, 12.6 mM UTP, 2.8 mg/ml tRNA, 3.64 mM CoA, 4.62 mM NAD, 10.5 mM cAMP, 0.95 mM folinic acid, 14 mM spermidine, 420 mM 3-PGA.
2. Thaw on ice the 100 mM Mg-glutamate stock solution, 3 M K-glutamate stock solution, 14x energy solution and 40% PEG-8000 to prepare the master mix. Keep them on ice.
3. Mix 9.18 µl of 100 mM Mg-glutamate stock solution, 3.06 µl of 3 M K-glutamate stock solution, 21.86 µl of 14x energy solution, 15.3 µl of 40% PEG-8000 and 1.6 µl sterile ddH₂O in a reaction tube. Thoroughly vortex this master mix after each addition, and keep it on ice.
4. Aliquot the master mix (51 µl) into volumes of 16 µl (3 aliquots) into reaction tubes. Frequently vortex the master mix during aliquoting. Flash freeze the aliquots in liquid nitrogen.
NOTE: The 16 µl aliquot volume as well as the master mix volume are slightly higher than required to account for losses due to pipetting.
5. Use a strainer to collect the energy buffer tubes. Store the tubes at -80 °C. **CAUTION!** Wear an eyeshield and Cryo-gloves to be protected from nitrogen splashes.

3. Preparation and Execution of Cell-free Reactions for Residue-specific Incorporation of ncAAs

- First, prepare the DNA vector solution in ddH₂O.
 - For highly efficient protein expression, use the expression vector pBEST-OR2-OR1-Pr-UTR1-deGFP-T500³². Clone the gene that codes for the model protein into this vector^{36,37}.
NOTE: Alternatively, use other promoters that are recognized by σ^{70} as well, but note that expression efficiency may be reduced.
 - Transform^{37,38} the vector in *E. coli* strain KL 740³² (Yale CGCS#:4382), purify amplified vector DNA^{37,39} and quantify the concentration of the DNA solution⁴⁰⁻⁴². Store the DNA solution at -20 °C or directly use it for the cell-free reaction preparation (steps 3.4.1, 3.4.2 and 3.4.3).
- Calibrate the cell-free expression efficiency depending on the used vector construct concentration according to the unmodified protocol³⁴. Use the optimal concentration that leads to highest protein yields for the cell-free reaction preparation (steps 3.4.1, 3.4.2 and 3.4.3).
NOTE: The preparation of cell-free reactions is exemplified using a 90 nM vector DNA stock solution that leads to a final vector concentration of 10 nM in the cell-free reaction and follows the above optimal values of Mg- and K-glutamate as well as the extract aliquot volume. Use the calculation template for different values.
- Thaw on ice 3 crude extract aliquots each of 30 μ l volume (prepared according to unmodified protocol³⁴), 1 amino acid solution aliquot labeled "+ cAA" (e.g., + Arg), 1 amino acid solution aliquot labeled "- cAA" (e.g., - Arg) and 1 amino acid solution aliquot labeled "+ ncAA" (e.g., + Can) (prepared in sections 1.2.5 - 1.2.7), 3 energy buffers aliquots (prepared in section 2) and the vector DNA solution (prepared in section 3.1).
NOTE: The crude extract is slightly viscous and it may contain air bubbles. Remove air bubbles by centrifuging at 10,000 x g for 30 sec at 4 °C. Put crude extract aliquots back on ice.
- Prepare 3 differently composed cell-free reactions (each 90 μ l final volume) by mixing crude extract (33.33%), energy buffer (16.67%), 1 of the 3 differently composed amino acid solution aliquots (16.67%) and vector DNA solution. Optionally, add further biomolecules (DNA, proteins, tRNA, etc.), but appropriately reduce the volume of ddH₂O.
 - Prepare the reference cell-free reaction (90 μ l) expressing the unmodified model protein.
 - Add the 15 μ l of energy buffer, 15 μ l of the amino acid solution aliquot labeled "+ cAA" (e.g., + Arg), 10 μ l of 90 nM vector DNA solution and 20 μ l of sterile ddH₂O to the 30 μ l of crude extract. Mix by pipetting up and down and gently vortex after the addition of each ingredient.
 - Aliquot the 90 μ l of cell-free reaction in 15 equal volumes of 6 μ l. Transfer each of the 15 volumes into a separate reaction tube. Close the tubes and put them back on ice. Label all reaction tubes as "CFR(+cAA)" (e.g., CFR(+Arg)).
 - Prepare the negative control cell-free reaction (90 μ l), where neither the cAA (e.g., Arg) nor the noncanonical analog (e.g., Can) are added.
 - Add the 15 μ l of energy buffer, 15 μ l of the amino acid solution aliquot labeled "- cAA" (e.g., - Arg), 10 μ l of 90 nM vector DNA solution and 20 μ l of sterile ddH₂O to the 30 μ l of crude extract. Mix by pipetting up and down and gently vortex after the addition of each ingredient.
 - Aliquot the 90 μ l of cell-free reaction in 15 equal volumes of 6 μ l. Transfer each of the 15 volumes into a separate reaction tube. Close the tubes and put them back on ice. Label all reaction tubes as "CFR(-cAA)" (e.g., CFR(-Arg)).
 - Prepare the cell-free reaction (90 μ l) that is supposed to residue-specifically incorporate the ncAA (e.g., Can) into the expressed model protein.
 - Add the 15 μ l of energy buffer, 15 μ l of the amino acid solution aliquot labeled "+ ncAA" (e.g., + Can), 10 μ l of 90 nM DNA solution and 20 μ l of sterile ddH₂O to the 30 μ l of crude extract. Mix by pipetting up and down and gently vortex after the addition of each ingredient.
NOTE: The expression efficiency may be reduced in comparison to the expression with standard amino acids. If necessary, simply scale up the cell-free reaction volume.
 - Aliquot the 90 μ l of cell-free reaction in 15 equal volumes of 6 μ l. Transfer each of the 15 volumes into a separate reaction tube. Close the tubes and put them back on ice. For increased reaction volumes, accordingly aliquot in further volumes of 6 μ l. Label all reaction tubes as "CFR(+ncAA)" (e.g., CFR(+Can)).
- Incubate all tubes at 29 °C O/N.
NOTE: Only small reaction volumes enable adequate oxygen diffusion into the reaction that is crucial for highly efficient protein expression. Reaction volumes greater than 10 μ l require active oxygenation through agitation³⁴. For large volumes, split the reaction into volumes smaller than 15 μ l.
- After cell-free expression, pool all identically composed split cell-free reactions of 6 μ l. First, pool all 15 split cell-free reactions of 6 μ l labeled "CFR(+cAA)" (e.g., CFR(+Arg)). Then, pool all 15 split cell-free reactions of 6 μ l labeled "CFR(-cAA)" (e.g., CFR(-Arg)). Finally, pool all 15 split cell-free reactions of 6 μ l labeled "CFR(+ncAA)" (e.g., CFR(+Can)). For increased reaction volumes, accordingly pool further volumes of 6 μ l.
- Check the expression level of the model protein in all three pooled, differently composed cell-free reactions by performing denaturing SDS-PAGE^{43,44} (section 4.1).
NOTE: Use this method for a preliminary evaluation of the incorporation experiment.
- Prepare the cell-free expressed model proteins for an appropriate analysis via HPLC-ESI mass spectroscopy (section 4.4). First, purify⁴⁵ them (section 4.2). Finally, exchange the buffer (section 4.3) to avoid high background noise during mass spectroscopy.
NOTE: The sections 3.4.1, 3.4.2 and 3.4.3 lead to typical cell-free reaction conditions³⁴: 8.9 - 9.9 mg/ml protein (from crude extract), 4.5 - 10.5 mM Mg-glutamate, 40 - 160 mM K-glutamate, 1 mM of each amino acid except leucine, 0.83 mM leucine, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/ml tRNA, 0.26 mM CoA, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30

mM 3-PGA, 2% PEG-8000 and 10 nM pBEST-OR2-OR1-Pr-UTR1-gene_of_model_protein-T500. If desired, a different procedure of cell-free reaction preparation can be conducted that leads to the reaction conditions above.

4. Preliminary Evaluation via SDS-PAGE^{43,44} and Preparation of the Cell-free Expressed Model Proteins for HPLC-ESI Mass Spectrometry

1. SDS-PAGE of the cell-free reactions

NOTE: Perform denaturing SDS-PAGE for a quick and preliminary analysis of the expressed proteins without any further purification or extraction, and execute the following steps.

- Thaw the protein standard on ice. Ensure that it consists of proteins with molecular weights similar to the expressed model protein for its localization on the gel (step 4.1.13).
NOTE: Here, the used standard provides proteins over a wide range of molecular weights (myosin: 212 kDa, maltose-binding-protein- β -galactosidase: 158 kDa, β -galactosidase: 116 kDa, phosphorylase b: 97 kDa, serum albumin: 66 kDa, glutamic dehydrogenase: 56 kDa, maltose-binding-protein: 43 kDa, thioredoxin reductase: 35 kDa, triosephosphate isomerase: 27 kDa, trypsin inhibitor: 20 kDa, lysozyme: 14 kDa, aprotinin: 7 kDa, insulin A: 3 kDa, B chain: 2 kDa).
- Since cell-free reactions are slightly viscous due to the high protein concentration, dilute them with sterile ddH₂O 5 - 10 times before SDS-PAGE to ensure appropriate gel migration and to avoid saturation after staining. In order not to waste too much sample, dilute 1 μ l of cell-free reaction in 4 μ l of sterile ddH₂O. Prepare the dilution in reaction tubes, thoroughly vortex and shortly spin the liquid down with a mini centrifuge for 2 - 3 sec at 2,000 x g.
- Add 5 μ l of 2x loading buffer to 5 μ l of diluted cell-free reaction. Thoroughly vortex and spin down with a mini centrifuge for 2 - 3 sec at 2,000 x g.
NOTE: Here, after addition, the biomolecules are dissolved in 62.5 mM Tris/Cl⁻, 10% glycerol, 2% SDS, and 0.0025% bromophenol blue at pH 6.8, a typical composition. Using other loading dyes that lead to slightly different dilution conditions may be suitable as well.
- Thoroughly vortex the protein standard and transfer 15 μ l of it into a reaction tube.
NOTE: Depending on gel size and for other protein standards, the recommended volume can differ from the above.
- Place the reaction tubes into a heating block. Check if the lids of the tubes are properly closed. Heat at 95 - 100 °C for 3 - 5 min. Do this to denature the proteins and to enable SDS-wrapping around the protein backbone.
NOTE: Some protein standards must NOT be heated, see supplier instructions.
- Meanwhile, transfer the running buffer into the gel electrophoresis chamber. The content of 1x running buffer is 25 mM Tris, 192 mM glycine, 0.1% SDS at pH 8.3 (with HCl).
- Fix the precast 4 - 20% gradient Tris-Glycine-SDS gel (10 cm x 10 cm x 1 mm) to the electrophoresis chamber. Remove the comb and rinse the wells with a syringe loaded with running buffer.
NOTE: The gel concentration strongly depends on the size and nature of the expressed model protein. The above concentration separates the *E. coli* crude extract proteins as well as low molecular weight model proteins. Self-cast gels are suitable as well.
- Remove the tubes from the heating block. Spin the liquid down with a mini centrifuge for 2 - 3 sec at 2,000 x g. Shortly vortex and repeat spinning down for 2 - 3 sec at 2,000 x g.
- Transfer the 15 μ l of protein standard (24 - 48 μ g protein) and 10 μ l (11 - 22 μ g protein) of each sample into the wells and start electrophoresis. Here, SDS-PAGE is performed at 125 V and 20 - 40 mA for approximately 90 min, a typical protocol.
- Carefully extract the gel from the cassette. Transfer it into the fixing solution (50% methanol, 10% acetic acid, 40% ddH₂O) for 30 min.
CAUTION! Methanol is toxic by inhalation and in contact with skin. Wear protective gloves and work under a fume hood.
- Transfer the gel into staining solution (0.025% Coomassie brilliant blue G-250, 10% acetic acid, 90% ddH₂O). Stain for 60 min.
- Transfer the gel into destaining solution (10% acetic acid, 90% ddH₂O) for 60 - 120 min.
NOTE: Fixing, staining and destaining strongly depend on the size and nature of the expressed protein. This protocol applies to a wide range of proteins of rather small molecular weights⁴⁶.
- Displace the gel onto a matte transparency on a white background that generates an appropriate contrast for the stained protein bands.

2. Purification of the His-tagged⁴⁵ model proteins from the cell-free reactions

NOTE: For protein purification, several methods exist that deliver similar results. This protocol purifies cell-free expressed model proteins that have a C-terminal polyhistidine-tag (His-tag). It uses a purification kit suitable for proteins that are expressed in the small reaction volumes of cell-free protein expression. It is identical for the purification of native or modified model proteins. Thus, it is generally described on the basis of one single cell-free reaction.

- For an appropriate HPLC-ESI mass spectroscopic analysis, extract and purify the model proteins from the cell-free reaction and execute the following steps.
 - Mix equal volumes of 90 - 150 μ l of His-binding buffer and 90 - 150 μ l of cell-free reaction. Pipette up and down, followed by gentle vortexing.
NOTE: Using His-binding buffer is recommended. However, cell-free reaction medium can be a starting material as long as the model protein is soluble, pH is between 7.5 - 8, imidazole/His concentration is < 10 mM, concentration of strong reducing agents is < 15 mM and no metal-chelating agents are present. If your mixture volume exceeds 300 μ l, split it into equal volumes and aliquot these volumes into different reaction tubes for the following purification steps.
 - Prepare the column system. Thoroughly vortex the stock solution of the His-affinity gel until the gel resin is completely dissolved. Transfer 250 μ l of gel resin into the column. Use a 1 ml pipette tip for the viscous gel resin. Place the column into a collection tube.
 - Centrifuge column/collection tube for 5 - 10 sec at 13,000 - 15,000 x g. Ensure that the gel resin is completely drained. If not, extend the centrifugation time by additional 5 - 10 sec, but pay attention that the affinity gel is not over-dried. Accordingly extend the centrifugation time in steps 4.2.1.5, 4.2.1.7 and 4.2.1.9.
NOTE: The gel resin is completely drained if it becomes stiff and no supernatant remains on top.
 - Transfer 150 - 300 μ l of the cell-free reaction/His-binding buffer to the column. If the mixture volume exceeds the recommended volume, split over several spin columns. Resuspend the gel resin by frequent tapping and gentle vortexing during an incubation time of at least 2 min. For volumes greater than 200 μ l, incubate for additional 1 - 2 min.

NOTE: Sufficient incubation time is crucial for binding of the model proteins to the gel resin.

5. Centrifuge column/collection tube for 5 - 10 sec at 13,000 - 15,000 x g. Discard the flow-through and place the column back into the collection tube.
6. Add 250 μ l of His-wash buffer. Resuspend the gel resin by tapping and gentle vortexing.
7. Centrifuge column/collection tube for 5 - 10 sec at 13,000 - 15,000 x g. Discard the flow-through and place the column back into the collection tube. Repeat step 4.2.1.6 and centrifuge again for 5 - 10 sec at 13,000 - 15,000 x g.
8. Place the column into a standard 1.5 ml reaction tube. Add 150 μ l of elution buffer and resuspend the gel resin by tapping and gentle vortexing. Reduce the volume to 100 μ l for higher purified model protein concentrations after elution in the next step 4.2.1.9.

NOTE: The total abundance of purified model protein may be reduced due to possible incomplete elution of all gel resin-bound proteins, if the elution buffer volume is reduced.

9. Centrifuge column/reaction tube for 5 - 10 sec at 13,000 - 15,000 x g to dilute the purified protein in the elution buffer. If split before, pool all solutions into one stock solution.
10. Before executing the buffer exchange, determine the protein concentration using standard methods, e.g., the Bradford assay^{47,48}. NOTE: For an appropriate HPLC-ESI mass spectroscopy (section 4.4), use optimal protein concentrations that are usually about 0.5 mg/ml with required volumes of 20 - 50 μ l. If the concentration is below 0.2 mg/ml, concentrate the protein solution after the buffer exchange using a spinning vacuum concentrator. In this case, first read section 4.3.8 to appropriately adapt the exchanging buffer.

3. Buffer exchange for HPLC-ESI mass spectroscopy

NOTE: Exchange the His-tag elution buffer to avoid high background noise in mass spectroscopic analysis due to high concentration of imidazole (> 150 mM) and other salts such as NaH_2PO_4 (> 300 mM) or NaCl (> 50 mM)⁴⁹. The following buffer exchange protocol uses a prehydrated gel filtration spin column system. It is identically executed for native or modified model proteins. Thus, it is generally described on the basis of one single cell-free reaction. Note that there are other easy-to-perform buffer exchange methods, e.g., mini-dialysis cartridges.

1. Prepare 100 ml of protein storage buffer (50 mM Tris-Cl pH 8, 100 mM NaCl, and 10% glycerol). Weigh out into an autoclavable 100 ml bottle 605.7 mg Tris-Base, 584.4 mg NaCl and add 10 ml of glycerol. Fill up to approximately 80 ml and adjust the pH value to 8 by adding NaOH. Steadily check the pH value with a pH meter. Finally, fill up to the 100 ml mark.

NOTE: Use this buffer to stabilize a wide range of model proteins and not to disturb mass spectroscopic analysis, as long as HPLC is performed prior to the mass spectroscopic measurement. However, depending on the nature of the expressed model protein, use other buffers that might be better suitable. If your target protein is not stable at pH 8, adjust the pH accordingly. **Do not use higher glycerol concentrations.**

2. Warm up the gel filtration spin columns for at least 15 min to RT.
3. Resuspend the prehydrated gel by gentle tapping or vortexing and remove air bubbles.
4. First remove the bottom cap, and then take the top cap away. Place the column into a wash tube (at least 2 ml). Transfer the column/wash tube into the centrifuge. Each column possesses an orientation mark. Centrifuge at 1,000 x g for 2 min to remove the gel storage buffer. Discard the flow-through.

NOTE: Pay attention to proceed in this order. Ensure that the column has the same orientation in all further centrifugation steps.

5. Add up to 400 μ l of the protein storage buffer. Centrifuge for 2 min at 1,000 x g. Repeat this to completely load the protein storage buffer into gel. Carefully add up to 100 μ l of the solution of the purified model protein. Pipette directly onto the center of the gel bed. NOTE: Solution volumes of purified proteins, particularly of modified proteins, might be higher. Split such solutions over several buffer exchange columns. Proteins must have molecular weights higher than 5 kDa for this kind of buffer exchange.

6. Place the column into a collection tube and centrifuge for 2 min at 1,000 x g. NOTE: This step leads to the dilution of the purified model protein in the protein storage buffer. If split before, pool all solutions into one stock solution.

7. Flash freeze the protein solution in liquid nitrogen and store at -80 °C or directly analyze it via HPLC-ESI mass spectroscopy⁵⁰ (section 4.4). **CAUTION!** Wear an eyeshield and Cryo-gloves to be protected from nitrogen splashes.

8. Perform the following steps of the protocol to concentrate the protein solution using a spinning vacuum concentrator that carefully evaporates the solvent⁵¹.

NOTE: These steps are required if the protein concentration is too low for HPLC-ESI mass spectroscopy (< 0.2 mg/ml). The solution volume and model protein concentration is roughly the same before and after buffer exchange. The concentration process is described by means of the following example: After His-tag purification, the model protein is dissolved in 100 μ l His elution buffer (step 4.2.1.9) and has a concentration of 0.07 mg/ml (step 4.2.1.10).

1. To ensure that the protein concentration and the solution volume are after evaporation at least 0.2 mg/ml and 20 μ l, respectively, dilute, before buffer exchange, the protein storage buffer prepared in step 4.3.1 at least threefold but at most fivefold. Equally evaporate in the following at least two third (66.67 μ l) or at most four fifth (80 μ l).

NOTE: Dilution of the protein storage buffer before the buffer exchange is crucial to take into account that concentration values of this buffer (step 4.3.1) are still realized in spite of evaporation. After dilution of the protein storage buffer, first perform the buffer exchange (steps 4.3.2 - 4.3.6) before executing the following steps 4.3.8.2 - 4.3.8.4.

2. After buffer exchange, put the complete 100 μ l of model protein solution in an open tube into the spinning vacuum concentrator. NOTE: The model protein is dissolved in the diluted protein storage buffer. Incline it towards the rotation center to avoid liquid-loss during rotation. Ensure that an open blank of the same volume of H_2O is symmetrically placed to balance the spinning system.

3. Close the lid of the concentrator and start rotation. To assure protein stability, concentrate at RT or at low temperatures up to 30 °C.

NOTE: The used concentrator automatically accelerates to 170 x g and establishes the vacuum. In the following it accelerates to its maximum speed of 240 x g.

4. Frequently interrupt the concentration process to survey the liquid volume. Stop the concentration, when the remaining solution volume is between 20 μ l and 33 μ l.

NOTE: Accordingly adapt the steps 4.3.8.1 - 4.3.8.4 for other solution volumes (step 4.2.1.9) and other protein concentrations (step 4.2.1.10). Flash freeze the concentrated protein solution in liquid nitrogen and store at -80°C , or directly analyze it via HPLC-ESI mass spectroscopy⁵⁰. CAUTION! Wear an eyeshield and Cryo-gloves to be protected from nitrogen splashes.

4. HPLC-ESI mass spectrometric analysis⁵⁰ of model proteins

1. Perform HPLC separation of 5 - 15 μl protein solution (prepared in section 4.3) on a C5 reverse-phase column (3 μm , 100 x 2.1 mm) with a gradient of 20% - 90% acetonitrile and 0.1% of formic acid over 25 min and subsequent ESI mass spectrometry with detection on a time-of-flight mass analyser in the range of 300 - 3,000 m/z.

NOTE: Depending on the nature of the expressed model protein, use other solvents or columns that might be better suitable.

2. Deconvolute measured mass spectra using appropriate software⁵² to calculate zero-charge masses.

Representative Results

This protocol guides through the cell-free residue-specific incorporation of ncAAs into model proteins. It proposes SDS-PAGE for a preliminary evaluation of the incorporation experiment and further steps to prepare the model proteins for an appropriate HPLC-ESI mass spectroscopic analysis.

Here, representative results of the cell-free residue-specific incorporation of the Arg analog Can, as well as the Lys analog L-hydroxy-lysine (Hyl) are presented. The different amino acid solutions, the energy buffer, the vector DNA coding for the model protein and cell-free reactions are prepared as described above. The reference cell-free reaction is provided with the amino acid solution consisting of the 20 cAAs. For each experiment, a negative control cell-free reaction is supplied with the amino acid solution that lacks the canonical analog of the ncAA in question. For each approach, one cell-free reaction expresses the model protein in the presence of the amino acid solution, where the cAA is substituted by the noncanonical analog. His-tag purification, buffer exchange and HPLC-ESI mass spectrometric analysis are executed according to the above described protocol.

The model protein is the C-terminal His-tagged deGFP³², a truncated version of EGFP⁵³. Its one letter amino acid sequence can be found in (Supplemental Material 2). This model protein contains 6 Arg and 18 Lys positions, respectively. The expression vector is pBEST-OR2-OR1-Pr-UTR1-deGFP-T500.

In the case of complete incorporation of the ncAA, one can assume that the negative control reaction does not express deGFP, since one of the 20 cAAs is missing. Contrary, deGFP must be detectable in the other two reactions: the native one in the reference cell-free reaction and the modified protein in the cell-free reaction that is provided with the ncAA.

Figure 1A shows the preliminary SDS-PAGE evaluation of the Can incorporation experiment. The reference cell-free reaction has the highest deGFP expression level. In the cell-free reaction that is provided with Can, deGFP is expressed at slightly lower concentration. No deGFP expression can be detected in the negative control. This SDS-PAGE result is a good indication for a successful incorporation of Can into the target protein deGFP.

To prove the hypothesized complete incorporation of Can into deGFP, both purified model proteins, visualized in **Figure 1B**, are analyzed through HPLC-ESI mass spectroscopy. **Figure 1C** shows the deconvoluted mass spectra of the purified deGFP molecules. The deconvoluted mass of deGFP that is expressed in the reference cell-free reaction is 26,192.8 Da. For deGFP expressed in the Can containing cell-free reaction a mass of 26202.5 Da appears. The expected masses for the native deGFP6Arg and the modified deGFP6Can with Arg being fully replaced by Can are 26,193 Da and 26,204 Da, respectively. The mass difference of 1.5 Da for deGFP6Can is within the error of spectrum deconvolution. Thus, the full incorporation of Can into deGFP at all 6 Arg positions is confirmed.

The two peaks of reduced intensity correspond to the native deGFP6Arg and modified deGFP6Can that did not attain their mature fluorophore. The fluorophore is autocatalytically generated by elimination of a H_2O molecule, followed by oxidation. This leads to a mass increased by 20 Da if this process does not proceed.

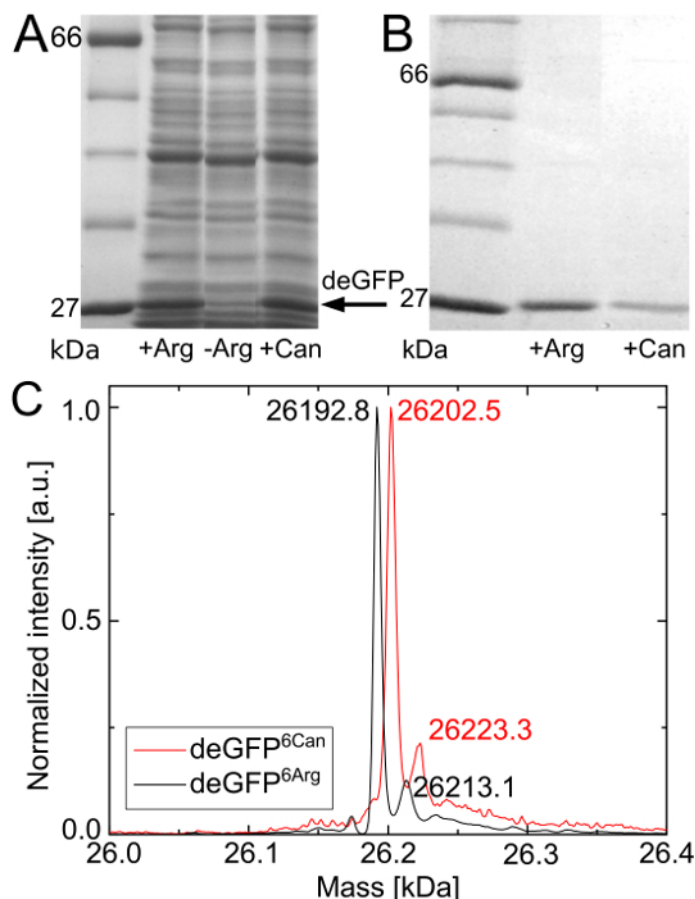


Figure 1. SDS-PAGE evaluation of the Can incorporation experiment and HPLC-ESI mass spectroscopy of the cell-free expressed and purified deGFP molecules. (A) Preliminary evaluation of the Can incorporation experiment using SDS-PAGE. From left to right: Protein standard, reference cell-free reaction, negative control and cell-free reaction providing Can instead of Arg. (B) SDS-PAGE after His-tag purification and buffer exchange of the expressed deGFP molecules. From left to right: Protein standard, purified deGFP from the reference reaction, purified deGFP from the Can containing reaction. (C) Confirmation of full incorporation of Can by HPLC-ESI mass spectroscopy. The expected masses of the native deGFP and the modified deGFP with Arg fully replaced by Can are 26,193 Da and 26,204 Da, respectively. Each spectrum is normalized to its highest intensity (counts). Peak positions are indicated in Da. For visualization purposes, the gel lanes are extracted from the gel pictures, joined together, are converted into gray scale format, size is optimized and contrast as well as brightness are enhanced. Original gel lanes are presented in Supplemental Material 3. [Please click here to view a larger version of this figure.](#)

Figure 2A shows the preliminary SDS-PAGE evaluation of the Hyl incorporation experiment. In the cell-free reaction that is provided with Hyl, deGFP is expressed. Contrary to the first experiment, a weak deGFP band can be observed in the negative control reaction. This might be due to Lys residues in the cell-free reactions. This enables a faint deGFP expression in the negative control reaction, where neither Lys nor Hyl are added.

For HPLC-ESI mass spectroscopy, the deGFP molecules of the cell-free reaction provided with Hyl are purified and the buffer is exchanged (**Figure 2B**).

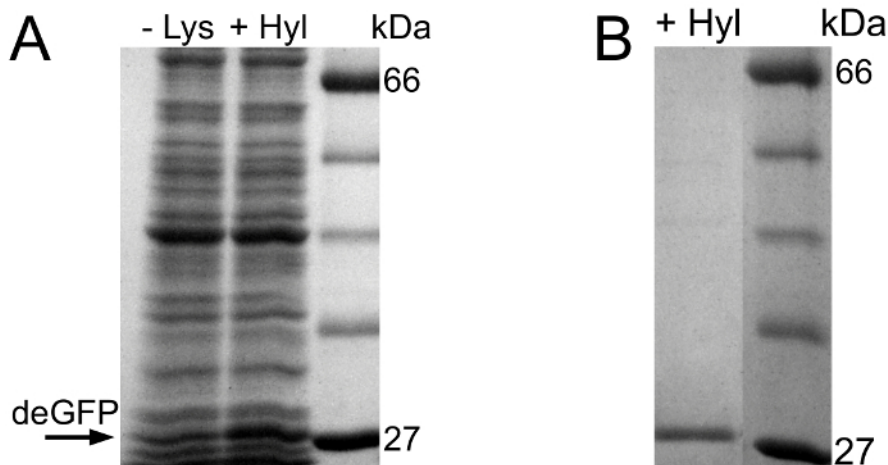


Figure 2. SDS-PAGE evaluation of the Hyl incorporation experiment. (A) From left to right: Negative control cell-free reaction, cell-free reaction containing Hyl and protein standard. (B) SDS-PAGE after His-tag purification and buffer exchange of the expressed deGFP molecules. From left to right: Purified deGFP from the Hyl containing reaction and protein standard. For visualization purposes, the gel lanes are extracted from the gel pictures, joined together, are converted into gray scale format, size is optimized and contrast as well as brightness are enhanced. Original gel lanes are presented in Supplemental Material 3. [Please click here to view a larger version of this figure.](#)

Figure 3 shows the deconvoluted mass spectrum of purified deGFP molecules. Figure 3A confirms the hypothesis of already present Lys residues in the cell-free reactions. The predominant peak of the spectrum corresponds to the native deGFP (expected mass: 26,193 Da). Again, deGFP molecules of a 20 Da higher mass that did not produce their fluorophore can be detected. The Lys residues are preferentially loaded onto the tRNA^{Lys} by lysyl-tRNA-synthetase leading to a high expression level of the native deGFP18Lys.

The mass difference between Hyl and Lys is 16 Da. Due to the presence of Hyl that is in competition to the Lys residues all possible deGFP species are generated (deGFP18Hyl, deGFP17Hyl+1Lys, ..., deGFP16Hyl+2Lys) (Figure 3B). Admittedly, the peak of deGFP1Hyl+17Lys overlaps with the peak of the native deGFP that did not produce its fluorophore (Figure 3A) and the mass of some peaks differs more than 2 Da from the expected mass. These mass differences can be attributed to high noise due to low amounts of the deGFP species. However, Hyl is generally incorporated by the cell-free system. Further improvements have to be done to abolish Lys residues in the cell-free reactions.

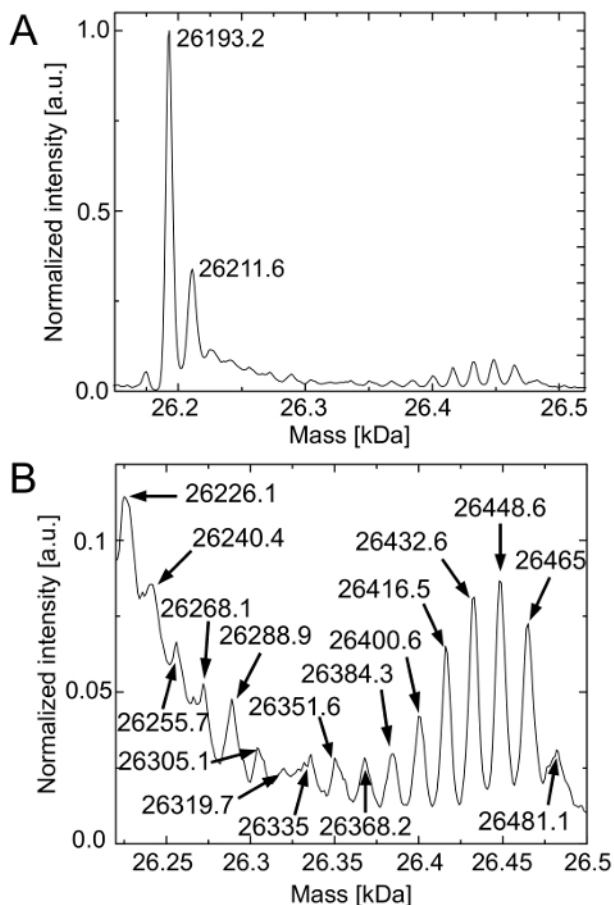


Figure 3. HPLC-ESI mass spectroscopy of the purified deGFP molecules of the cell-free reaction containing Hyl. (A) The native deGFP18Lys is predominantly detected (expected masses: with fluorophore 26,193 Da, without mature fluorophore 26,213 Da). (B) Magnification reveals the existence of all possible deGFP species (deGFP18Hyl, deGFP17Hly+1Lys, deGFP16Hyl+2Lys,..., deGFP1Hyl+17Lys). Their expected masses are 26,193 Da + N x 16 Da (N = 1, ..., 18). The spectrum is normalized to its highest intensity (counts). Peak positions are indicated in Da. [Please click here to view a larger version of this figure.](#)

1) Volume of crude extract aliquots, cell-free reactions and energy buffer aliquots

Enter the volume of crude extract aliquots into the orange field.
 Note: The aliquot volume depends on the total protein concentration of the produced crude extract batch.^{**} The maximal aliquot volume is 30 µl for protein concentrations between 27 - 30 mg/µl^{**}, it is scaled down for higher protein concentrations^{**} (e.g. 32 mg/ml protein concentration → 28.1 µl aliquots).
 Crude extract volume is 1/3 of the total cell-free reaction volume.
 Consequently, the final volume of the cell-free reaction is: µl
 Crude extract volume and energy buffer volume are 1/2 of the total cell-free reaction volume.
 Thus, the volume of energy buffer added to the cell-free reaction is: µl

2) Energy buffer composition and preparation

2.1) Enter the optimal Mg²⁺ and K⁺ glutamate concentrations for cell-free expression into the orange fields:

| Component | Stock solution concentration | Concentration in cell-free reaction |
|-----------------|------------------------------|-------------------------------------|
| Crude Extract | 100 % | 33.33 % |
| Mg glutamate | 100 mM | 2 mM |
| K glutamate | 3000 mM | 40 mM |
| Energy Solution | 14 x | 1 x |
| PEG-8000 | 40 % | 2 % |

The 2nd column of this table indicates the stock solution concentration of the energy buffer components.
 The 3rd column of this table presents the concentration of the energy buffer components in the final cell-free reaction.

Based on the above stock solution concentrations and optimal Mg²⁺ & K⁺ glutamate concentrations, the composition of a 15 µl energy buffer added to the cell-free reaction is:

| Component | Composition of 15 µl energy buffer | Component | Composition of 30 µl energy buffer aliquots |
|------------------------|------------------------------------|------------------------|---|
| Mg glutamate (100 mM) | 1.00 µl | Mg glutamate (100 mM) | 1.00 µl |
| K glutamate (3000 mM) | 1.33 µl | K glutamate (3000 mM) | 1.33 µl |
| Energy Solution (14 x) | 1.00 µl | Energy Solution (14 x) | 0.80 µl |
| PEG-8000 (40 %) | 1.00 µl | PEG-8000 (40 %) | 0.80 µl |
| Water to add* | 1.67 µl | Water to add* | 3.33 µl |

* If negative, increase stock solution concentrations of Mg²⁺ or K⁺ glutamate

2.2) Enter the number of desired energy buffer aliquots each of a volume of 16 µl that you want to produce into the orange field: = 16 µl

Prepare the energy buffer master mix by mixing the following volumes of the energy buffer components:

| Component | Composition of master mix for 100 energy buffer aliquots |
|------------------------|--|
| Mg glutamate (100 mM) | 200.00 µl |
| K glutamate (3000 mM) | 133.00 µl |
| Energy Solution (14 x) | 728.73 µl |
| PEG-8000 (40 %) | 510.00 µl |
| Water to add* | 121.27 µl |
| | 1700 µl |

* If negative, increase stock solution concentrations of Mg²⁺ or K⁺ glutamate
 Note: The 16 µl aliquot volume as well as the master mix volume are slightly higher than required to account for losses due to pipetting.

3) Preparation of the cell-free reaction

3.1) Enter the optimal vector concentration of pBEST-ORF-DNA gene_of_model_protein-T500 that leads to the highest expression yield of the model protein into the orange field: nM

3.2) Enter the stock solution concentration of the vector pBEST-ORF-DNA gene_of_model_protein-T500 into the orange field: nM

Based on the DNA concentrations, prepare the cell-free reaction by mixing the following volumes of the different components:

| Component | Composition of cell-free reaction | Concentration in cell-free reaction |
|--|-----------------------------------|-------------------------------------|
| Crude Extract | 27.00 µl | 33.33 % |
| Energy buffer | 16.00 µl | 20.00 % |
| Amino acid solution* | 14.00 µl | 17.78 % |
| pBEST-ORF-DNA gene_of_model_protein-T500 | 1.00 µl | 1.28 % |
| Water to add** | 15.00 µl | 19.00 % |
| | 80 µl | 100.00 % |

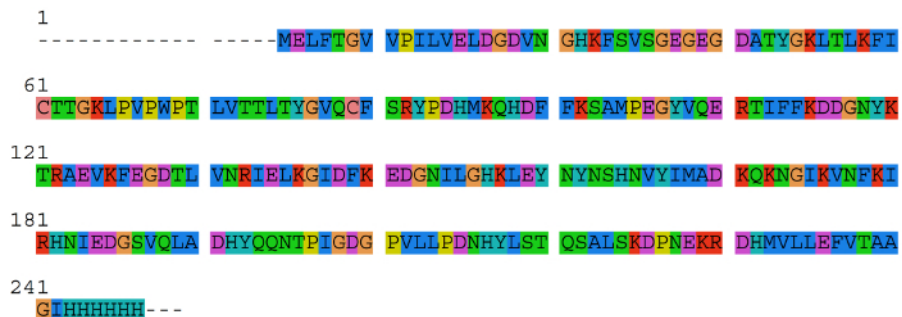
* leads to a concentration of 1 mM for each amino acid except Leu (0.83 mM Leu) in a cell-free reaction

** If negative, increase concentration of DNA in stock solution

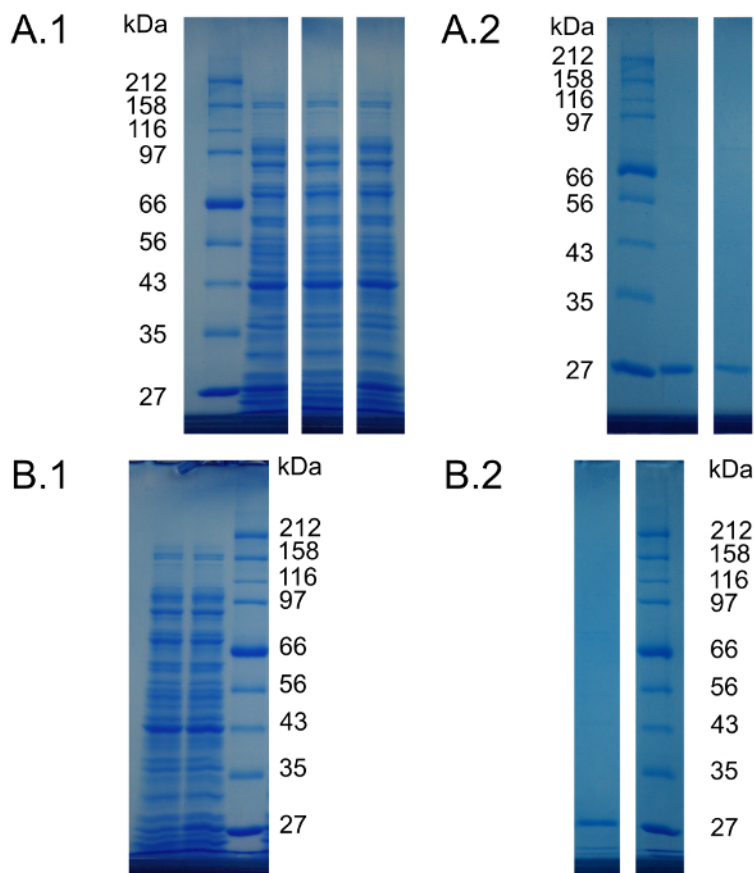
Supplemental Material 1. Calculation template. In section 2, the preparation of the energy buffer master mix is exemplified using crude extract aliquots of 30 µl and optimal Mg- and K-glutamate concentrations of respectively 3 mM and 30 mM. This example leads to a master mix volume that yields 3 energy buffer aliquots. In section 3, the preparation of the cell-free reaction is exemplified using a 90 nM DNA stock solution leading to an optimal vector DNA concentration of 10 nM in the cell-free reaction. [Please click here to download this file.](#)

For an appropriate traceability, the use of the calculation template is exemplified by insertion of these typical values that differ from the above example: Crude extract volume: 28 µl, optimal Mg-glutamate: 2 mM, optimal K-glutamate: 40 mM, number of desired buffer aliquots: 100, optimal vector concentration in cell-free reaction: 8 nM, DNA vector stock solution: 150 nM.

First enter 28 µl as crude extract volume into the orange field of the first template section. Then, enter into the second template section 2 mM and 40 mM as optimal Mg- and K-glutamate concentrations into the orange fields. Taking into account the optimal Mg- and K-concentrations, the composition of a 15 µl energy buffer, as well as a corresponding, scaled up 16 µl aliquot is calculated. Below, accordingly enter 100 as desired number of energy buffer aliquots (16 µl). The template adapts the volumes of the different buffer components for the 1,700 µl master mix as follows: 204 µl of 100 mM Mg-glutamate stock solution, 136 µl of 3 M K-glutamate stock solution, 728.73 µl of 14x energy solution, 510 µl of 40% PEG-8000 and 121.27 µl sterile ddH₂O. Finally, in the third template section, enter 8 nM and 150 nM as optimal vector concentration in the cell-free reaction and respectively, vector DNA stock solution concentration. The template adapts the volumes of the different components that must be added to the 28 µl of crude extract to finalize the preparation of a 90 µl cell-free reaction as follows: 15 µl of energy buffer, 15 µl of one of the 3 differently composed amino acid solutions, 4.80 µl of vector DNA solution (150 nM), and 27.20 µl of sterile ddH₂O.



Supplemental Material 2. One letter amino acid sequence of the model protein deGFP. This model protein contains 6 Arg and 18 Lys positions. [Please click here to view a larger version of this figure.](#)



Supplemental Material 3. Full length and unmodified gel lanes that correspond to the gel pictures presented Figure 1 and 2. The gel lanes presented in each individual subfigure are extracted from the same SDS polyacrylamide gel. In **Figure 1** and **2** these lanes were joined together for presentation purposes. Molecular weights of the protein standard bands are indicated beside the figures. (A.1) Uncropped gel lanes of **Figure 1A**. From left to right: Protein standard, reference cell-free reaction, negative control and cell-free reaction providing Can instead of Arg. (A.2) Uncropped gel lanes of **Figure 1B**. From left to right: Protein standard, purified deGFP from the reference reaction, purified deGFP from the Can containing reaction. (B.1) Uncropped gel lanes of **Figure 2A**. From left to right: Negative control cell-free reaction, cell-free reaction containing Hyl and protein standard. (B.2) Uncropped gel lanes of **Figure 2B**. From left to right: Purified deGFP from the Hyl containing reaction and protein standard. [Please click here to view a larger version of this figure.](#)

Discussion

An easy-to-use cell-free expression system as a viable strategy to residue-specifically incorporate ncAAs into proteins, is presented. To this end, the crude extract is supplemented with vector DNA coding for the protein of interest, the energy buffer and the corresponding amino acids. Note that the crude extract aliquot volume depends on the crude extract protein concentration³⁴. The cell-free expression efficiency is optimized depending on the vector DNA construction concentration. The volumes of the energy buffer components vary as function of optimized Mg- and K-glutamate concentrations in order to enable high yields of the cell-free expressed model protein.

A preliminary evaluation of the incorporation experiment can be obtained by performing SDS-PAGE of the unpurified cell-free reaction medium. For a more detailed analysis, HPLC-ESI mass spectroscopy is proposed as a means to check for complete, residue-specific incorporation of the ncAA. As a preparation for the latter, spin column systems are used to enable His-tag purification and buffer exchange with the small volumes that we use in this protocol.

Including HPLC-ESI mass spectroscopy, the entire protocol can be performed within 2 days. It does not include any particularly critical steps. However, concentration optimizations of Mg- and K-glutamate as well as of vector DNA are crucial in order to express high yields of the model protein. The use of the highly efficient expression vector pBEST-OR2-Or1-Pr-UTR1-gene_of_model_protein-T500 is strongly recommended. Elution of His-tagged proteins is usually due to high concentration of imidazole (> 150 mM) and other salts such as NaH₂PO₄ (> 300 mM) or NaCl (> 50 mM) that generate high background noise in mass spectroscopic analysis⁴⁹. Exchange of such elution buffers with a suitable protein storage buffer stabilizes the model protein and drastically reduces background noise during mass spectroscopic analysis.

As a result, Can replaces Arg at all six positions within the model protein. In the expression system, no Arg residues can be detected. This simplifies the residue-specific incorporation of Arg analogs compared to other expression systems that require further depletion strategies^{29,30}. The presented cell-free approach circumvents the inherent limitations of *in vivo* approaches that are due to Can toxicity, or the strong

dependency on mRNA sequence in single protein production strategies^{24,31}. Contrary to the employed *in vitro* system, *in vivo* cleavage of Can to homoserine and hydroxyguanidine does occur³¹.

However, the cell-free system retains a sufficient amount of Lys to compete with analogs such as Hyl. The HPLC-ESI mass spectroscopic analysis shows that the model protein contains both, the canonical as well as the noncanonical analog in different proportions. The residue-specific incorporation of Lys is possible in general, but for complete substitution, further depletion strategies, or specially engineered aaRS and tRNA optimized for the recognition of ncAAs need to be developed.

We achieved excellent yields of cell-free expressed, modified model proteins by adding the ncAA at the same concentration as the canonical ones. The incorporation efficiency depends on the nature of the ncAA to be incorporated. Even higher yields might still be realizable by optimizing the concentration of the ncAA.

The presented results demonstrate the applicability of the employed system for the residue-specific incorporation of ncAAs as long as they are accepted by the canonical endogenous translational system. For the residue-specific incorporation of specific ncAAs, one further needs to check if the residues of the analogous cAA disturb the expression system.

Cell-free transcription-translation systems can be engineered from different organisms to respond to different demands⁵⁴. The all *E. coli* transcription-translation machineries of the here presented cell-free system enable the usage of bacteriophage and *E. coli* promoters, and they can act in parallel or consecutively in cascades⁵⁵. The general applicability and usability make the method a potent tool for further research in amino acid toxicity and therapeutic application.

Disclosures

The authors have nothing to disclose.

Acknowledgements

E.G. Worst and A. Ott acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) within the collaborative research center SFB 1027 as well as Saarland University. E.G. Worst, A. Ott and V. Noireaux further acknowledge financial aid by the Human Frontiers Science Program Organization (HFSP). The authors thank Tobias Baumann and Stefan Oehm (Institute of Chemistry, Technische Universität Berlin) for critical reading.

References

- Böck, A., *et al.* Selenocysteine: the 21st amino acid. *Mol. Microbiol.* **5** (3), 515 - 520 (1991).
- Srinivasan, G., James, C.M., Krzycki, J.A. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science*. **296** (5572), 1459 - 1462, (2002).
- Budisa, N. Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire. *Angew. Chem. Int. Ed. Engl.* **43**, 6426 - 6463 (2004).
- Wang, L., Xie, J., Schultz, P.G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225 - 249 (2006).
- Chin, J.W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **83**, 379 - 408 (2014).
- Neumann, H. Rewiring translation - Genetic code expansion and its applications. *FEBS Lett.* **586** (15), 2057 - 2064 (2012).
- Liu, C.C., Schultz, P.G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413 - 444 (2010).
- Goerke, A.R., Swartz, J.R. High-Level Cell-Free Synthesis Yields of Proteins Containing Site-Specific Non-Natural Amino Acids. *Biotechnol. Bioeng.* **102** (2), 400 - 416 (2009).
- Albayrak, C., Swartz, J.R. Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.* **41** (11), 5949 - 5963 (2013).
- Johnson, J.A., Lu, Y.Y., Van Deventer, J.A., Tirrell, D.A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14** (6), 774 - 780 (2010).
- Xiu, X., Puskar, N.L., Shanata, J.A.P., Lester, H.A., Dougherty D.A. Nicotine binding to brain receptors requires a strong cation- π interaction. *Nature*. **458** (7237), 534 - 537 (2009).
- Grünewald, J., *et al.* Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids. *Proc. Natl. Acad. Sci. USA*. **106** (11), 4337 - 4342 (2009).
- Nikić, I., Lemke, E.A. Genetic code expansion enabled site-specific dual-color protein labeling: superresolution microscopy and beyond. *Curr. Opin. Chem. Bio.* **28**, 164 - 173 (2015).
- Munier, R., Cohen, G.N. Incorporation d'analogues structuraux d'acides aminés dans les protéines bactériennes. *Biochim. Biophys. Acta*. **21** (3), 592 - 593 (1956).
- Lepthien, S., Merkel, L., Budisa, N. *In Vivo* Double and Triple Labeling of Proteins Using Synthetic Amino Acids. *Angew. Chem. Int. Ed.* **49** (32), 5446 - 5450 (2010).
- Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., Schuman, E.M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. USA*. **103** (25), 9482 - 9487 (2006).
- Dieterich, D.C. *et al.* *In situ* visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13** (7), 897 - 905 (2011).
- Hendrickson, W.A., Horton, J.R., LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9** (5), 1665 - 1672, (1990).
- Hoesl, M.G., *et al.* Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem*. **3** (1), 213-221 (2011).
- Bae, J.H., *et al.* Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins. *J. Mol. Biol.* **328** (5), 1071 - 1081 (2003).

21. Schachtele, C.F., Rogers, P. Canavanine death in *Escherichia coli*. *J. Mol. Biol.* **14** (2), 474 - 489 (1965).
22. Rosenthal, G.A. The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine. *Q. Rev. Biol.* **52** (2), 155 - 178 (1977).
23. Rosenthal, G.A., Dahlman, D.L. Incorporation of L-Canavanine into Proteins and the Expression of Its Antimetabolic Effects. *J. Agric. Food Chem.* **39** (5), 987 - 990 (1991).
24. Ishida, Y., Park, J.H., Mao, L., Yamaguchi, Y., Inouye, M. Replacement of All Arginine Residues with Canavanine in MazF-bs mRNA Interferase Changes Its Specificity. *J. Biol. Chem.* **288** (11), 7564 - 7571 (2013).
25. Thomas, D.A., Rosenthal, G.A., Gold, D.V., Dickey, K. Growth Inhibition of a Rat Colon Tumor by L-Canavanine. *Cancer Res.* **46** (6), 2898 - 2903, (1986).
26. Bence, A.K., Worthen, D.R., Adams, V.R., Crooks, P.A. The antiproliferative and immunotoxic effects of L-canavanine and L-canaline. *Anticancer Drugs.* **13** (3), 313 - 320 (2002).
27. Bence, A.K., Crooks, P.A. The Mechanism of L-Canavanine Cytotoxicity: Arginyl tRNA Synthetase as a Novel Target for Anticancer Drug Discovery. *J. Enzyme Inhib. Med. Chem.* **18** (5), 383 - 394 (2003).
28. Akaogi, J., et al. Role of non-protein amino acid L-canavanine in autoimmunity. *Autoimmun. Rev.* **5** (6), 429 - 435 (2006).
29. Singh-Blom, A., Hughes, R.A., Ellington A.D. An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J. Biotechnol.* **178**, 12 - 22 (2014).
30. Oh, S.-J., Lee, K.-H., Kim, H.-C., Catherine, C., Yun, H., Kim, D.-M. Translational Incorporation of Multiple Unnatural Amino Acids in a Cell-free Protein Synthesis System. *Bioprocess. Eng.* **19** (3), 426 - 432 (2014).
31. Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., Ott, A. Cell-free expression with the toxic amino acid canavanine. *Bioorg. Med. Chem. Lett.* **25** (17), 3658 - 3660 (2015).
32. Shin, J., Noireaux, V. Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* **4** (8) (2010).
33. Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V., Alfonta, L. Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* **112** (8), 1663 - 1672 (2015).
34. Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., Noireaux, V. Protocols for Implementing an *Escherichia coli* Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp.* (79), e50762 (2013).
35. Caschera, F., Noireaux, V. Preparation of amino acid mixtures for cell-free expression systems. *Biotechniques.* **58** (1), 40 - 43 (2015).
36. JoVE Science Education Database. Molecular Cloning. *Basic Methods in Cellular and Molecular Biology*. JoVE, Cambridge, MA (2015).
37. Green, M.R., Sambrook, J. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2012).
38. Froger, A., Hall, J.E. Transformation of Plasmid DNA into *E. coli* Using the Heat Shock Method. *J. Vis. Exp.* (6), e253 (2007).
39. Zhang, S., Cahalan, M.D. Purifying Plasmid DNA from Bacterial Colonies Using the Qiagen Miniprep Kit. *J. Vis. Exp.* (6), e247 (2007).
40. Moreno, L.A., Cox, K.L. Quantification of dsDNA using the Hitachi F-7000 Fluorescence Spectrophotometer and PicoGreen Dye. *J. Vis. Exp.* **45**, e2465 (2010).
41. Desjardins, P., Conklin, D. NanoDrop Microvolume Quantitation of Nucleic Acids. *J. Vis. Exp.* **45**, e2565 (2010).
42. Sukumaran, S., Concentration Determination of Nucleic Acids and Proteins Using the Micro-volume Bio-spec Nano Spectrophotometer. *J. Vis. Exp.* **48**, e2699, (2011).
43. Laemmli, U.K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature.* **227** (5259), 680 - 685 (1970).
44. JoVE Science Education Database. *Basic Methods in Cellular and Molecular Biology*. Separating Protein with SDS-PAGE. JoVE, Cambridge, MA (2015).
45. Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., Stüber, D. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nature Biotechnology.* **6** (11), 1321 - 1325 (1988).
46. Schägger, H., Aquila, H., Von Jagow, G. Coomassie blue-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for direct visualization of polypeptides during electrophoresis. *Anal. Biochem.* **173** (1), 201 - 205 (1988).
47. Bradford, M.M. A Rapid Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **72**, 248 - 254 (1976).
48. Ernst, O., Zor, T. Linearization of the Bradford Protein Assay. *J. Vis. Exp.* **38**, e1918 (2010).
49. Hurst, R., Kobs, G., Johnson, T. *Mass Spectrometric Analysis of MagneHis Purified Proteins*. Promega Corporation Web site. <http://www.promega.de/resources/pubhub/enotes/mass-spectrometric-analysis-of-magnehis-purified-proteins/> Updated 2003. Accessed October 12, (2015).
50. Banerjee, S., Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**, 1 - 40 (2012).
51. Fujiwara, K., Nomura, S.M. Condensation of an additive-free cell extract to mimic the conditions of live cells. *PLoS One.* **8** (1), e54155 (2013).
52. Zhang, Z., Marshall, A.G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass. Spectrom.* **9** (3), 225 - 233 (1998).
53. Li, X., Zhang, G., Ngo, N., Zhao, X., Kain, S.R., Huang, C.C. Deletions of the *Aequorea victoria* Green Fluorescent Protein Define the Minimal Domain Required for Fluorescence. *J. Biol. Chem.* **272** (45), 28545 - 28549 (1997).
54. Gagoski, D., Polinkovsky, M.E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y., Alexandrov, K. Performance benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.* **113** (2), 292 - 300 (2016).
55. Shin, J., Noireaux, V. An *E. coli* Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **1** (1), 29 - 41 (2012).