HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2022 April 14.

Published in final edited form as:

Methods Enzymol. 2021 ; 656: 545–571. doi:10.1016/bs.mie.2021.05.008.

Incorporation of proline analogs into recombinant proteins expressed in *Escherichia coli*

Stephanie L. Breunig, **David A. Tirrell***

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA.

Abstract

Proline residues are unique in the extent to which they constrain the conformational space available to the protein backbone. Because the conformational preferences of proline cannot be recapitulated by any of the other proteinogenic amino acids, standard mutagenesis approaches that seek to introduce new chemical functionality at proline positions unavoidably perturb backbone flexibility. Here, we detail the incorporation of proline analogs into recombinant proteins in *Escherichia coli* via a residue-specific mutagenesis strategy. This approach results in global replacement of proline residues with high yields of the recombinant protein of interest, minimal genetic manipulation, and maintenance of backbone conformational constraints.

Keywords

Proline analog; non-canonical amino acid; residue-specific mutagenesis; protein conformation; recombinant protein; *Escherichia coli*

INTRODUCTION

As the sole α-imino acid among the canonical building blocks of proteins, proline is unique. Its pyrrolidine ring imposes conformational constraints on the polypeptide backbone that are essential to protein structure and function (Figure 1a,b). The backbone dihedral angle ϕ of proline residues is restricted to 63±15° (MacArthur & Thornton, 1991), limiting the conformational trajectories that are available to the polypeptide chain. The reduced conformational space sampled by the peptide backbones of proline residues is illustrated by the Ramachandran plots for proline and the acyclic amino acids (Figure 1a-b). Compared to linear amino acid side chains, the pyrrolidine side chain is similarly limited in its conformational flexibility, and generally adopts one of only two rapidly-interconverting ring puckers: C$^{\gamma}$-*endo* or C$^{\gamma}$-*exo* (Figure 1c-d, Shoulders & Raines, 2009). Because the proline amide linkage cannot serve as a hydrogen-bond donor, the presence of proline residues within α-helical or β-strand structural motifs is disfavored (Levitt, 1978). Proline's *cis* and *trans* conformers are nearly isoenergetic, and the barrier to *cis-trans* isomerization is reduced in comparison to the other amino acids. Consequently, *cis* isomers are more common for proline than for any other canonical amino acid (MacArthur & Thornton, 1991), and

*Corresponding author: tirrell@caltech.edu.

cis-trans isomerization at proline residues can play important roles in protein folding (Lu, Finn, Lee, & Nicholson, 2007) and function (Lummis, Beene, Lee, Lester, Broadhurst, & Dougherty, 2005).

Non-canonical amino acids (ncAAs) have found extensive use in chemical biology and related fields (Chin, 2017; Johnson, Lu, Van Deventer, & Tirrell, 2010; Mukai, Lajoie, Englert, & Söll, 2017). They serve many roles, such as providing chemical handles for protein modification (Link, Vink, & Tirrell, 2004), serving as probes in time-resolved and cell-selective proteomic analyses (Dieterich, Link, Graumann, Tirrell & Schuman, 2006), interrogating the effects of post-translational modifications (Luo, et al., 2017), identifying protein-protein interaction partners (Chin, Martin, King, Wang, & Schultz, 2002), tracking protein location *in vivo* (Wang, Xie, & Schultz, 2006), and probing the importance of non-covalent interactions in protein behavior (Xiu, Puskar, Shanta, Lester & Dougherty, 2009).

The utility of standard mutagenesis at proline sites is limited by its impact on chain conformation. ncAAs provide a means of addressing this limitation, as replacement of proline by non-canonical analogs allows introduction of new chemical functionality while maintaining conformational constraints. For instance, the hydroxyl groups of the non-canonical proline (ncPro) variants 4*R*-hydroxyproline (4*R*-OH) and 4*S*-hydroxyproline (4*S*-OH, Table 1) permit hydrogen-bonding interactions; one such hydrogen bond has been suggested to alter the behavior of an engineered insulin variant (Lieblich, et al., 2017). Expanding or contracting the five-membered pyrrolidine ring can add (piperidine-2-carboxylic acid, Pip, Table 1) or remove (azetidine-2-carboxylic acid, Aze, Table 1) hydrophobic packing interactions (Fang, Lieblich, & Tirrell, 2019). The alkene functionality in 3,4-dehydroproline (Dhp, Table 1) has been used as a functional handle to modify protein-based materials (Deming, Fournier, Mason & Tirrell, 2006). Fluorinated proline variants can be used as conformational reporters in ^{19}F NMR experiments (Verhoork Killoran, & Coxon, 2018). For instance, *cis/trans* isomerization of the sole 4,4-difluoroproline (4,4-diF, Table 1) residue in β2 microglobulin could be monitored by ^{19}F NMR (Torbeev & Hilvert, 2013).

Many ncPro analogs have well-documented conformational biases. The presence of a 4*R*-electron-withdrawing substituent (as in 4*R*-OH and 4*R*-fluoroproline, 4*R*-F; Table 1) stabilizes the C$^{\gamma}$-*exo* ring pucker through a gauche effect (Figure 1c). The *exo* ring pucker, in turn, pre-organizes the amide in the *trans* conformation. Conversely, ncPro residues with 4*S*- electron-withdrawing groups (such as 4*S*-OH and 4*S*-fluoroproline, 4*S*-F; Table 1) favor the *endo* ring pucker (Figure 1d), and have a higher propensity for the *cis* amide isomer compared to canonical proline. The presence of a 4-fluoro substituent (and presumably any electron-withdrawing group) lowers the energy barrier to cis/trans isomerization, as the substituent decreases the bond order of the preceding amide (Renner, Alefelder, Bae, Budisa, Huber, & Moroder, 2001). These established conformational preferences provide a means of interrogating the effects of conformational properties (such as pyrrolidine ring pucker or amide isomerization) on protein structure and function (Ganguly & Basu, 2020; Shoulders & Raines, 2009). For instance, ncPro residues have been used to identify a key proline *cis-trans* isomerization event in ion channel opening, (Lummis, Beene, Lee, Lester, Broadhurst, & Dougherty, 2005), demonstrate the stereoelectronic basis of collagen stability (Shoulders

& Raines, 2009), and modulate the properties of protein–based materials (Kim, McMillan, Snyder, & Conticello, 2005).

Two general strategies have been developed for *in vivo* introduction of ncAAs into proteins: site-specific and residue-specific replacement. Both approaches rely on the availability of aminoacyl-tRNA synthetases (aaRSs) able to charge their cognate tRNAs with the ncAAs of interest (Figure 2a). Site-specific approaches (including nonsense suppression and related techniques) cause minimal perturbation of protein sequence, and important advances in such methods (for example, Sakamonto et al., 2002; Lajoie et al, 2013; Dunkelmann, Willis, Beattie, & Chin, 2020) have been accomplished since the first report of incorporation in *Escherichia coli* (Furter, 1998). In site-specific ncAA mutagenesis, a re-assigned codon (most often the amber stop codon) is matched with the ncAA of interest by an engineered, orthogonal tRNA/aaRS pair. The resulting ncAA-charged tRNA competes with release factors for amber codon recognition within the ribosomal complex, and successful translational read-through positions the ncAA at the amber site (Chin, 2017). Challenges in implementing site-specific methods include the formation of truncation products that reduce recombinant protein yields, especially when ncAAs are incorporated at multiple positions; and the requirement for development of orthogonal tRNA/aaRS pairs, which can be difficult.

In contrast to site-specific methods, residue-specific ncAA mutagenesis (Johnson, Lu, Van Deventer, & Tirrell, 2010) results in the global replacement of a canonical amino acid with a non-canonical counterpart. This technique typically uses defined expression media and amino acid auxotrophs (i.e. strains deficient in biosynthesis of the amino acid of interest) as expression hosts. A generalized workflow for residue-specific ncAA mutagenesis is depicted in Figure 2b. In some cases, the endogenous aaRS is promiscuous enough to allow the desired substitution (Dieterich, Link, Graumann, Tirrell & Schuman, 2006); in other cases, overexpression of the endogenous aaRS (Kiick, van Hest, & Tirrell, 2000) or expression of a mutant aaRS is required (Tanrikulu, Schmitt, Mechulam, Goddard III, & Tirrell, 2009). Residue-specific replacement often requires less genetic manipulation than site-specific techniques, and can produce high yields of the recombinant protein of interest: typical yields are 50-60% of those obtained from expression in media that contain the canonical amino acid.

Residue-specific ncAA mutagenesis has been used in a variety of contexts. Global incorporation of the methionine analog azidohomoalanine (Aha) enables ‘bio-orthogonal non-canonical amino acid tagging’ (BONCAT), a proteomic method capable of enriching for newly synthesized proteins (Dieterich, Link, Graumann, Tirrell & Schuman, 2006). Various methionine analogs have also been incorporated by residue-specific methods to study purified proteins, including elastin-based biomaterials (Teeuwen, et al., 2009) and prion proteins (Wolschner, Giese, Kretzchmar, Huber, Moroder, & Budisa, 2009). Leucine analogs have been used to alter the properties of coiled-coil proteins (Montclare, Son, Clark, Kumar, & Tirrell, 2008) and introduced into GFP (Yoo, Link, & Tirrell, 2007).

Global incorporation of proline analogs has been used to modulate the properties of purified, recombinant proteins. The first example of biosynthetic ncPro incorporation in recombinant

proteins expressed in *E. coli* was the replacement of five proline residues in human annexin V with 1,3-thiazolidine-4-carboxylic acid (Thz; Budisa, Minks, Medrano, Lutz, Humer, & Moroder, 1998). Since then, additional proline variants have been reported to show translational activity in *E. coli* (Kim, George, Evans, & Conticello, 2004; Renner, Alefelder, Bae, Budisa, Huber, & Moroder, 2001). The most commonly used analogs are 4*R*-F and 4*S*-F, which have been found to alter the folding and fluorescence properties of GFP (Steiner, Hess, Bae, Wiltschi, Moroder, & Budisa, 2008), the activity of a DNA polymerase (Holzberger & Marx, 2010) and the melting temperatures of elastin-mimetic peptides (Kim, McMillan, Snyder, & Conticello, 2005; Kim, Hardcastle, & Conticello, 2006). Similar approaches have been used to incorporate proline analogs into recombinantly-expressed barstar (Renner, Alefelder, Bae, Budisa, Huber, & Moroder, 2001), ubiquitin (Crespo & Rubini, 2011), thioredoxin (Rubini, Schärer, Capitani, & Glockshuber, 2013), and 4-oxalocrotonate tautomerase (Lukesch, Pavkov-Keller, Gruber, Zangger, & Wiltschi, 2019). We have found that replacement of proline B28 of human insulin by a variety of proline analogs can be used to engineer the therapeutically-relevant biophysical properties of the protein (Lieblich, et al., 2017; Fang, Lieblich, & Tirrell, 2019). Because no aaRS/tRNA pair capable of site-specific incorporation of proline analogs has been described to date, residue-specific ncAA mutagenesis provides the only option currently available for introducing modified proline residues into recombinant proteins.

In this chapter, we discuss the residue-specific incorporation of proline analogs into proteins expressed in *E. coli*. The *E. coli* prolyl-tRNA synthetase (ProRS) and downstream translational machinery have been shown to accommodate a variety of structural analogs of proline. For these analogs, simple over-expression of the *E. coli* ProRS (or point mutants thereof), combined with increased expression of proline transporters under hyperosmotic conditions (Grothe, Krogsrud, McClellan, Milner, & Wood, 1989), enables high levels of proline replacement (Kim, George, Evans, & Conticello, 2004). Proline analogs that have been incorporated into recombinant proteins are detailed in Table 1, along with reported expression conditions. The protocols discussed here are illustrated by the incorporation of proline analogs into recombinant proinsulin expressed from a pQE-80L vector backbone. However, similar approaches should enable the incorporation of other amino acid analogs and expression of other recombinant proteins.

MATERIALS AND EQUIPMENT

General equipment

- Agarose gel electrophoresis equipment (e.g. Mini-Run Gel Electrophoresis System, Bulldog Bio)
- Agarose gel imager
- Standing incubator, 37°C
- Shaking incubator, 37°C
- Electroporator
- Spectrophotometer

- Centrifuges (able to cool to 4°C)
- SDS gel electrophoresis chamber (e.g. XCell *SureLock* Mini-Cell, Thermo Fisher Scientific)
- SDS gel electrophoresis power supply (e.g. PowerEase Power Supply, Thermo Fisher Scientific)
- Sonicator or French press (or alternative method for bacterial cell lysis)
- MALDI-TOF mass spectrometer

General materials

- PCR tubes (0.2 mL)
- Microcentrifuge tubes (1.7 mL)
- LB medium (10 g/L casein, 5 g/L yeast extract, 10 g/L NaCl)
- LB agar plates (10 g/L casein, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar)
- Ampicillin (100 mg/mL stock solution)
- 0.5x TBE buffer (40 mM Tris hydrochloride, pH 8.3; 45 mM boric acid; 1 mM Ethylenediaminetetraacetic acid, EDTA)
- Cast agarose gels for DNA electrophoresis, containing dye for DNA visualization (e.g. 1-1.5% w/v agarose in 0.5x TBE + 1x GelGreen Nucleic Acid Stain, Millipore)
- Gel loading dye, 6x (NEB; often packaged with restriction enzyme orders)
- DNA ladder (e.g. TriDye 1kb Plus DNA Ladder, NEB)
- DNA gel recovery kit (e.g. Zymoclean Gel DNA Recovery Kit, Zymo Research)
- Plasmid miniprep kit (e.g. ZR Plasmid Miniprep, Zymo Research)
- Serological pipettes (10 mL, 25 mL)
- Sterile culture tubes
- Sterile Erlenmeyer flasks (125 mL, 2.5 L)
- Sterile filters, 0.2 μm
- Syringes
- Disposable plastic cuvettes
- Isopropyl β-d-1-thiogalactopyranoside (IPTG, 1 M stock solution)
- Sterile 500 mL centrifuge bottles (e.g. autoclaved Nalgene PPCO Centrifuge Bottles with Sealing Closure, Thermo Fisher Scientific)

Note: To facilitate wash steps during the medium shift, it is helpful to mark 100 mL on each bottle prior to protein expression.

- SDS-PAGE gels (e.g. NuPAGE 4-12%, Bis-Tris, Mini Protein Gel)
- SDS loading buffer (e.g. 4x NuPAGE LDS Sample Buffer, Thermo Fisher Scientific; add 0.5% v/v 2-mercaptoethanol immediately before protein denaturation for reducing loading buffer)
- Protein ladder (e.g. SeeBlue Plus2 Pre-stained Protein Standard, Thermo Fisher Scientific)
- 1x SDS running buffer (e.g. NuPAGE MES SDS Running Buffer, Thermo Fisher Scientific)
- Coomassie staining solution (e.g. InstantBlue Ultrafast Protein Stain, MilliporeSigma)

Strain preparation and cloning

- Competent cloning *E. coli* strain (e.g. NEB 10-beta)
- pQE-80L vector (Qiagen)
- Primers for *proS* Gibson Assembly (10 μM stock solution)
  - proS insert, fwd primer (proS-ins-fwd)
  - proS insert, rev primer (proS-ins-rev)
  - pQE80 vector, fwd primer (proS-vec-fwd)
  - pQE80 vector, rev primer (proS-vec-rev)
- Gene corresponding to the protein of interest (POI), flanked by restriction enzyme cut sites (see Note 1)
- Restriction enzymes (e.g. EcoRI-HF & BamHI-HF, NEB)
- DNA ligase (e.g. T4 DNA Ligase, NEB)
- High-fidelity DNA polymerase (e.g. Q5 High-Fidelity 2x Master Mix, NEB)
- Gibson Assembly enzymes (e.g. repliQa HiFi Assembly Mix, Quantabio)
- Competent proline auxotrophic *E. coli* strain for protein expression (e.g. CAG18515, see Note 2).

Note: Competent bacterial samples can be prepared by a variety of methods, e.g. see Inoue 1990.

Protein expression (2 x 1.25 L scale)

Note: It is recommended to conduct batch expression experiments with 2.5 L 1x 20aa M9 medium, split into two flasks (2 x 1.25 L). One floor centrifuge (e.g. with a JA-10 rotor) can accommodate this culture volume, which facilitates the medium shift. The corresponding volumes of 1.25x 19aa M9 medium are 2 x 1 L. However, this protocol can easily be altered to accommodate other culture volumes.

- 1x 20aa M9 medium (2.5 L)

See Table 2 for preparation of stock solutions, sterilization, and suggested storage conditions; and Table 3 for medium recipe.

1x M9 medium working concentrations: 8.5 mM NaCl; 18.7 mM NH_4Cl; 22 mM KH_2PO_4; 47.8 mM Na_2HPO_4; 20 mM D-glucose; 50 mg/L each canonical amino acid; 0.1 mM $CaCl_2$; 1 mM $MgSO_4$; 3 mg/L $FeSO_4$; 35 mg/L thiamine·HCl; 10 mg/L biotin; 1 μg/L each Cu^{2+}, Mn^{2+}, Zn^{2+}, MoO_4^{2-}; 100 mg/L ampicillin.

- 1.25x 19aa M9 minimal medium (−proline; 2 L)

  See Table 3 for medium recipe. Composition of this medium is identical to 1x 20aa M9, but omits 5x proline.

- 0.9% w/v NaCl (9 g/L NaCl, sterilize by autoclaving. Cool to 4 °C before expression experiment.)

- 5x ncPro solution (2.5 mM ncPro, 1.5 M NaCl; 0.5 L, sterile filter. Working concentration: 0.5 mM ncPro, 0.3 M NaCl)

Note: We have found that 0.5 mM ncPro and 0.3 mM NaCl often provide high levels of analog incorporation and suggest these concentrations in general. However, these conditions can be optimized; see Note 3.

Alternatives: Other chemically defined medium may be used in place of M9, as long as it is possible to remove proline in the 19aa formulation.

Protein purification and analysis

- Ni-NTA agarose (e.g. HisPur Ni-NTA resin, Thermo Fisher Scientific), protein purification column, and relevant purification buffers (or other method for protein purification, dependent upon POI; see Note 4)
- Dithiothreitol (DTT, 100 mM stock solution)
- Iodoacetamide (300 mM stock solution)
- 100 mM ammonium bicarbonate (NaH_4HCO_3) pH 8.0
- Glu-C (0.5 μg/μL stock solution), or other relevant peptidase (see Note 5)
- 5% Trifluoroacetic acid (TFA) in H_2O
- Acetonitrile (ACN)
- 0.1% TFA in H_2O
- 0.1% TFA in a 1:1 ACN:H_2O solution
- ZipTip with C^{18} resin, MilliporeSigma (or other desalting columns/tips)

STEP-BY-STEP METHOD DETAILS

Strain preparation and cloning

Timing: ~5 days—Note: We describe one possible cloning approach below; however, other cloning approaches may be more appropriate depending upon the situation; see Note 1.

1. Insert the gene corresponding to the POI into the pQE-80L vector backbone using standard restriction enzyme ligation approaches. In this case, a gBlock gene fragment encoding the RBS and an N-terminal hexahistidine-tagged proinsulin, and flanked by EcoRI and BamHI cut sites, was inserted to yield pQE80-PI.
2. Transform the ligation product into an *E. coli* cloning strain. Purify the resulting plasmid product, and verify correct assembly by sequencing.
3. Insert the gene encoding the *E. coli* ProRS downstream of the POI using standard Gibson Assembly methods. In this case, the plasmid pQE80-PI-proS (Figure 3) was prepared with repliQa HiFi Assembly Mix, following the manufacturer's protocols. The following polymerase chain reaction (PCR)-amplified DNA fragments were used in the assembly:
    a. The *E. coli proS* gene and its endogenous promoter were amplified by colony PCR using the primers proS-ins-fwd and proS-ins-rev (Q5 PCR conditions: 66°C annealing temperature; 1 min extension time).
    b. pQE80-PI was amplified using the primers proS-vec-fwd and proS-vec-rev (Q5 PCR conditions: 64°C annealing temperature; 2.5 min extension time).
4. Transform the Gibson Assembly product into an *E. coli* cloning strain. Purify the resulting plasmid product, and verify correct assembly by sequencing.
5. Transform the proline auxotrophic *E. coli* strain of choice (see Note 2) with the sequence-verified construct. In this case, strain CAG18515 was transformed with pQE80-PI-proS.

Note: Efficient incorporation of some proline analogs requires ProRS mutants (e.g. ProRS-C443G has been used to incorporate ncPro **Pip**, Table 1). These mutations can be prepared via standard site-directed mutagenesis techniques (e.g. QuikChange, Agilent; or Q5 Site-Directed Mutagenesis, NEB).

Protein expression

Timing: 2 days—Note: Protein expression conditions may need to be optimized, depending upon the POI (see Note 3).

Day 1:

4. Prepare and sterilize 5x ncPro/NaCl solution, 0.9% NaCl solution, 1x 20aa M9 medium, and 1.25x M9 medium.

Note: Biotin, thiamine, and $FeSO_4$ should be dissolved the day of medium preparation to minimize degradation/oxidation. Prepared complete medium should be used within one day.

5. From a single colony of CAG18515/pQE80-PI-proS, inoculate 30 mL LB medium in a 125 mL Erlenmeyer flask. Grow overnight at 37°C, with shaking. Ensure that stationary phase is reached before dilution into M9 medium the next morning.

Day 2:

Growth

6. Remove and set aside 10 mL prepared 1x 20aa M9 medium, to be used as a blank for measuring the optical density at 600 nm (OD_{600}).
7. Inoculate each 1.25 L 1x 20aa M9 flask with 12.5 mL overnight culture.
8. Incubate each culture at 37°C, with shaking.
9. Monitor bacterial growth over time by measuring OD_{600}. Allow the culture to grow just until the end of exponential phase (generally, OD_{600} ~0.8).
10. During cell growth, prepare to perform a medium shift:
    a. Cool centrifuges to 4°C.
    b. Ensure 0.9% NaCl solution is at 4°C; keep on ice during the medium shift.
11. When the bacterial cultures reach OD_{600} ~0.8, immediately remove the flasks from the incubator and place on ice for 15 min to cool cultures and stop growth. Swirl cultures occasionally as they cool.

Critical: To ensure good protein yield and ncPro incorporation, the medium shift must be performed within a fairly small OD_{600} window (0.75-0.85). Beginning a medium shift too early can reduce cell density and lead to poor overall yield, and too late can result in poor protein expression and ncPro incorporation.

Medium Shift: Critical: All steps during medium shift should be performed efficiently (within about one hour), and cells should be kept at 4°C throughout the process.

12. Transfer the cooled cultures to six sterile 500 mL centrifuge bottles.
13. Centrifuge 4 kg, 8 min, 4°C.
14. Decant the medium, leaving a cell pellet.
15. Resuspend each cell pellet in 100 mL cold, sterile 0.9% NaCl by gently pipetting up and down with a 25 mL serological pipette.
16. Repeat steps 13-15, for a total of two washes, each with 100 mL cold 0.9% NaCl.
17. Resuspend each cell pellet in ~25 mL 1.25x 19aa M9. Transfer the resuspended cell mass to the larger 1.25x 19aa M9 medium flasks, evenly dividing the cell pellets between the two 1 L volumes.

Proline depletion & ncPro uptake

18. Incubate 19aa cultures at 37°C, with shaking, for 30 min, to deplete residual proline.
19. Add 250 mL sterile 5x ncPro/NaCl solution to each proline-depleted culture.

20. Incubate at 37°C, with shaking, for 30 min, to allow for ncPro uptake.

Protein expression

21. Induce protein expression. Here, proinsulin expression was induced by adding 1.25 mL 1 M IPTG (working concentration: 1 mM IPTG).

22. Incubate at 37°C, with shaking, for 2.5 hours. Expression times may vary depending upon the POI, see Note 3.

23. Transfer cultures to 500 mL centrifuge tubes.

24. Harvest cells by centrifugation at 5 kg for 10 min, RT.

25. Transfer cell pellet to tared conical tubes.

26. Determine and record the mass of the resulting cell pellet. This value is often used in cell lysis and protein purification protocols.

27. Store cell pellets at −80°C, at least overnight.

Pause Point: Cell pellets can be stored indefinitely at −80°C before proceeding to cell lysis, protein purification, and analysis.

Analysis of ncPro incorporation

Timing: 1-2 days

Protein purification: **Note**: Purification protocols will vary depending upon the POI (see Note 4).

28. Thaw bacterial pellets from full expression cultures.

29. Lyse cells. A variety of methods can be used, including chemical lysis (e.g. B-PER Complete, Thermo Fisher Scientific), French press, and sonication. In this case, B-PER Complete was used.

30. Purify the protein of interest using an appropriate purification protocol. In this case, hexahistidine-tagged proinsulin was isolated from the inclusion body fraction by immobilized metal affinity chromatography (IMAC) purification performed under denaturing conditions (8 M urea). A pH 3.0 elution buffer was used to elute the POI. An example gel of the inclusion body fraction after proinsulin expression, and fractions collected during IMAC purification, is shown in Figure 4a.

31. Verify proper isolation of the desired protein using techniques such as SDS-PAGE and MALDI-TOF.

Protein digestion: **Note**: Steps 32 and 33 are only necessary if cysteine residues are present in the POI.

32. To a 30 μL sample of purified protein, add 1.5 μL 20x DTT stock (100 mM; working concentration: 5 mM). Incubate sample for 20 min at 55°C to reduce cysteine residues.

33. Add 1.58 μL 20x iodoacetamide stock (300 mM; working concentration: 15 mM) and incubate for 15 min at RT to alkylate reduced cysteine residues.

Note: Be sure to cover the sample, and avoid exposing the alkylation reaction mixture to light.

34. Digest alkylated protein with Glu-C.

Note: Other peptidases besides Glu-C (such as trypsin) can be used (see Note 5). Digestion conditions may vary with the chosen peptidase.

a. Dilute 10 μL of the alkylated sample with 90 μL of 100 mM NH_4HCO_3, pH 8.0.

b. Add 0.6 μL Glu-C stock (0.5 μg/μL; working concentration: 3 ng/μL)

c. Incubate sample for 2.5 h at 37°C to digest protein.

d. Quench the digestion reaction by adding 10 μL 5% TFA, and immediately proceed to MALDI-TOF sample preparation.

Determination of ncPro incorporation by MALDI-TOF

35. Desalt the quenched digestion reactions by ZipTips, following the manufacturer's protocol. Elute peptides in 1:1 ACN:H_2O, 0.1% TFA.

36. Dilute 1 μL of the eluted fraction with 3 μL saturated α-cyanohydroxycinnamic acid matrix.

37. Spot 0.75 μL matrix/sample mixture onto a MALDI target plate, and analyze by MALDI-TOF mass spectrometry.

Note: To achieve good peptide ionization, alternative MALDI-TOF matrices, or mass spectrometry techniques (e.g. electrospray ionization) may be needed.

38. Determine the incorporation efficiency of the ncPro analog into the recombinant protein by the following formula:

$$\%\ Incorporation = \frac{A}{P + A} \times 100\,\%$$

Where P and A are determined by integrating the peaks corresponding to the proline- and analog-containing peptides, respectively. See Figure 4b for an example MALDI-TOF spectrum, and Figure 4c for a representative analysis.

EXPECTED OUTCOMES

This method should yield a purified, recombinant protein with high (≥85%) incorporation efficiencies of the desired proline analog. Incorporation efficiency will depend upon the proline analog used (Table 1). In general, we have found that expression levels of the recombinant POI correlate with proline analog incorporation.

NOTES AND CONSIDERATIONS

Note 1: Design a plasmid and cloning scheme that enables expression of both the POI and the *E. coli* ProRS. We typically use inducible expression of the POI (e.g. via the IPTG-inducible T5 promoter), and ProRS overexpression with its endogenous promoter. Both proteins can be expressed from the same plasmid backbone; see Figure 3 for the scheme of the final construct described here. We have also found success with different construct designs (alternative vector backbones, promoters, two plasmid approaches, etc.) that facilitate expression of both the POI and the ProRS. Cloning approaches other than those described here may be more convenient, depending upon the situation.

Note that ProRS over-expression may not be necessary for the incorporation of some ncPro analogs with particularly high incorporation efficiencies; for example, 3*R*-F, 3*S*-F, 4*R*-F, 4*S*-F, and Dhp are reported to have been incorporated without ProRS over-expression (Kim, George, Evans, & Conticello, 2004; Kim, Hardcastle, & Conticello, 2006). However, as there are conflicting reports in different expression systems (for example, Lukesch, Pavkov-Keller, Gruber, Zangger, & Wiltschi, 2019), we recommend ProRS over-expression in general.

Note 2: Obtain an appropriate strain of *E. coli* for expression of the POI. A bacterial strain unable to synthesize proline (i.e. a proline auxotroph) is necessary for efficient proline analog incorporation into the recombinant POI. Strains deficient in proline biosynthesis via disruption of *proA, proB* or *proC* are available from the Coli Genetic Stock Center (CGSC, https://cgsc.biology.yale.edu). We generally use the strain CAG18515; however, we have also had success with other strains (such as those from the Keio collection).

In some cases, authors have noted oxidative degradation of the proline analogs Thz and Dhp via L-proline dehydrogenase (*putA*) and Δ1-pyrroline-5-carboxylate reductase (*proC*). Strain UMM5, which contains mutations in both of these genes, resulted in enhanced incorporation of these ncPro residues (Kim, George, Evans, & Conticello, 2004).

Note 3: Optimize expression conditions for the POI in M9 medium using the proline auxotroph of choice. These efforts typically involve screening parameters such as expression time, temperature, and inducer concentrations. Some reports in the literature have suggested that osmolyte identity (e.g. NaCl vs sucrose) and concentration can influence ncPro incorporation efficiencies and protein yield (for example, Buechter et al, 2003; Kim, George, Evans, & Conticello, 2004), and so we also suggest screening these parameters in the relevant expression system. Relative protein expression levels can be evaluated by methods such as SDS-PAGE, and extent of ncPro replacement by MALDI-TOF.

Note 4: Determine a method of purification for the POI. One approach often used to isolate a hexahistidine-tagged protein (common for proteins encoded on a pQE-80L vector) is IMAC using Ni-NTA agarose resin (e.g. HisPur Ni-NTA Resin, ThermoFisher Scientific). In this case, *E. coli* cells are lysed with B-PER Complete after expression. Hexahistidine-tagged proinsulin was isolated from the inclusion body fraction by IMAC purification performed under denaturing conditions (8 M urea), using a low pH (3.0) elution buffer. However, purification methods will likely vary depending upon the POI and affinity tag

used. For example, denaturants such as urea are typically not used if the POI is soluble; instead, the protein is purified under native conditions. Resources that describe protein purification protocols are available; for instance, we have found the QIAexpressionist (which can be found on qiagen.com) to be helpful in planning purification of hexahistidine-tagged proteins.

Note 5: Digestion of the POI and MALDI-TOF analysis of the resulting peptide fragments are often necessary to permit accurate determination of ncPro incorporation efficiencies. The protocol here describes digestion with the peptidase GluC-C; however, the choice of peptidase will depend on the sequence of the POI. We recommend using a peptidase that results in at least one peptide fragment that contains only one proline residue. This makes the analysis of proline analog incorporation more straightforward.

ADVANTAGES

Incorporation of proline analogs into recombinant proteins is especially useful in producing high molecular weight proteins that would be difficult to prepare by solid phase peptide synthesis, and in producing large quantities of a POI. Expression yields are usually high: yields of ncPro-containing proteins are typically ~50-60% of those obtained by expression in rich medium.

LIMITATIONS

Residue-specific proline replacement results in global replacement of proline residues by the proline analog of interest; specific replacement of one proline residue (in a POI with multiple proline residues) is not possible with this method. The method is limited to proline analogs that are accepted by the *E. coli* translational machinery (and in particular, to those that can be activated by the endogenous or mutant forms of the ProRS). Bulky or charged analogs may yield low or undetectable levels of incorporation.

The proline analog will be incorporated into all newly synthesized proteins, not just the recombinant POI. This phenomenon impacts cell growth and limits the utility of the technique in studies of ncPro-modified proteins in live *E. coli*.

Quantitative replacement of proline is often not achieved. Low levels of proline-containing proteins are generally observed, likely due to residual intracellular proline stores and leaky expression during the initial growth phase in proline-containing medium. These effects are diminished for proteins that express well, and for proline analogs that are efficiently activated by the *E. coli* ProRS.

OPTIMIZATION AND TROUBLESHOOTING

Weak/unobservable peptide peak by MALDI-TOF

- Ensure that the peptidase used is active by running control digestion reactions.
- Prepare fresh TFA/H_2O/ACN solutions for peptide desalting by ZipTip.
- Try alternative MALDI-TOF matrices.

- Use alternative mass spectrometry approaches (e.g. electrospray ionization).

Low ncPro incorporation efficiency

- Ensure that wash steps during the medium shift were performed well. After each centrifugation step, the supernatant should be cleared of obvious cell density and fully decanted. Cell pellets should be fully resuspended and dispersed in the 0.9% NaCl wash solution during wash steps, or 19aa M9 medium before the depletion step.
- Optimize protein expression conditions by screening strains, medium recipes, inducer concentrations, osmolyte concentration, expression temperatures, and expression times. Expression of the recombinant POI is often correlated with proline analog incorporation efficiency.
- Certify proline auxotrophy in the expression strain by comparing growth in medium with and without proline. The bacterial strain should not grow in the absence of proline.
- Using analytical chemistry techniques such as HPLC and NMR analysis, check that the proline analog used is not contaminated by canonical proline.
- Test proline analog incorporation in alternative strains. Oxidative degradation of Thz and Dhp in *E. coli* has been reported to occur; these effects been alleviated by using strains lacking the genes *putA* and *proC* (such as strain UMM5, Kim, George, Evans, & Conticello, 2004).

SAFETY CONSIDERATIONS AND STANDARDS

General PPE (such as safety glasses and gloves) should be worn at all times.

Use a chemical hood when handling TFA and 2-mercaptoethanol.

ALTERNATIVE METHODS/PROCEDURES

A similar method that does not involve a medium shift has been described by other laboratories (for example, Budisa, Minks, Medrano, Lutz, Humer, & Moroder, 1998). In this approach (designated “selective pressure incorporation;” SPI), cells are grown in proline-limiting medium until proline has been depleted (as determined by growth arrest). At this point, the ncPro is added to the culture, and protein expression is induced. This approach is less resource- and labor-intensive than the medium shift protocol described here. Furthermore, it may be more adaptable to large-scale industrial applications. However, we generally observe slightly higher ncPro incorporation rates when a medium shift is included in the expression protocol. For instance, we have achieved 84% incorporation of 4*R*-OH using the SPI method, compared to 89% incorporation using the medium shift expression protocol described here.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health (1R01GM134013-01). S.L.B. was supported by a National Science Foundation Graduate Research Fellowship under grant number 1745301. We thank our colleagues Alex Chapman, Katharine Fang, and Seth Lieblich for their contributions to our work on proline analogs.

REFERENCES

Budisa N, Minks C, Medrano FJ, Lutz J, Huber R, & Moroder L (1998). Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: Structure and stability of the per-thiaproline mutant of annexin V. Proc. Natl. Acad. Sci. USA, 95, 455–459. [PubMed: 9435213]

Buechter DD, Paolella DN, Leslie BS, Brown MS, Mehos KA, & Gruskin EA (2003). Co-translational incorporation of *trans*-4-hydroxyproline into recombinant proteins in bacteria. J. Biol. Chem, 278, 645–650. [PubMed: 12399455]

Chin JW, Martin AB, King DS, Wang L, & Schultz PG (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 99, 11020–11024. [PubMed: 12154230]

Chin JW (2017). Expanding and reprogramming the genetic code. Nature, 550, 53–60. [PubMed: 28980641]

Crespo MD & Rubini M (2011). Rational Design of Protein Stability: Effect of (2S,4R)-4-Fluoroproline on the Stability and Folding Pathway of Ubiquitin. PLoS ONE, 6, e19425. [PubMed: 21625626]

Deming TJ, Fournier MJ, Mason TL & Tirrell DA (1997). Biosynthetic incorporation and chemical modification of alkene functionality in genetically engineered polymers. J. Macromol. Sci. A, 34, 2143–2150.

Dieterich DC, Link AJ, Graumann J, Tirrell DA & Schuman EM (2006). Selective identification of newly synthesized proteins in mammalian cells using Bioorthogonal Noncanonical Amino Acid Tagging. Proc. Natl. Acad. Sci. USA, 103, 9482–9487. [PubMed: 16769897]

Dunkelmann DL, Willis JCW, Beattie AT, & Chin JW (2020) Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. Nat. Chem, 12, 535–544. [PubMed: 32472101]

Fang KY, Lieblich SA, & Tirrell DA (2019). Replacement of ProB28 by pipecolic acid protects insulin against fibrillation and slows hexamer dissociation. J. Polym. Sci. A. Polym. Chem, 57, 264–267.

Furter R (1998). Expansion of the genetic code: Site-drected *p*-fluoro-phenylalanine incorporation in *Escherichia coli*. Protein Sci., 7, 419–426. [PubMed: 9521119]

Ganguly HK & Basu G (2020). Conformational landscape of substituted prolines. Biophys. Rev, 12, 25–39. [PubMed: 31953795]

Grothe S, Krogsrud RL, McClellan DJ, Milner JL, & Wood JM (1986). Proline transport and osmotic stress response in *Escherichia coli* K-12. J. Bacteriol, 166, 253–259. [PubMed: 3514577]

Holzberger B & Marz A (2010). Replacing 32 proline residues by noncanonical amino acid results in a highly active DNA polymerase. J. Am. Chem. Soc, 132, 15708–15713. [PubMed: 20961065]

Inoue H, Nojima H & Okayama H (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene, 96, 23–28. [PubMed: 2265755]

Johnson JA, Lu YY, Van Deventer JA, & Tirrell DA (2010). Residue-specific incorporation of non-canonical amino acids into proteins: Recent developments and applications. Curr. Opin. Chem. Biol, 14, 774–780. [PubMed: 21071259]

Kiick KL, van Hest JCM, & Tirrell DA (2000). Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. Angew. Chemie Int. Ed, 39, 2148–2152.

Kim W, George A, Evans M, & Conticello VP (2004). Co-translational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system. ChemBioChem, 5, 928–936. [PubMed: 15239049]

Kim W, McMillan A, Snyder JP, & Conticello VP (2005). A stereoelectronic effect on turn formation due to proline substitution in elastin-mimetic polypeptides. J. Am. Chem. Soc, 127, 18121–18132. [PubMed: 16366565]

Kim W, Hardcastle KI, & Conticello VP (2006). Fluoroproline flip-flop: Regiochemical reversal of a stereoelectronic effect on peptide and protein structures. Angew. Chem. Int. Ed 45, 8141–8145.

Lajoie MJ, Rovner AJ, Goodman DB, Aerni H-R, Haimovich AD, Mercer JA, et al. (2013). Genomically recoded organisms expand biological functions. Science, 342, 357–360. [PubMed: 24136966]

Levitt M (1978). Conformational preferences of amino acids in globular proteins. Biochemistry, 17, 4277–4285. [PubMed: 708713]

Lieblich S (2017). Non-canonical amino acid mutagenesis of position B28 in insulin with proline analogs (Doctoral thesis, California Institute of Technology, Pasadena, CA, USA). Retrieved from https://thesis.library.caltech.edu/10166

Lieblich SA, Fang KY, Cahn JKB, Rawson J, LeBon J, Ku HT, & Tirrell DA (2017). 4*S*-Hydroxylation of insulin at ProB28 accelerates hexamer dissociation and delays fibrillation. J. Am. Chem. Soc, 139, 8384–8387. [PubMed: 28598606]

Link AJ, Vink MKS, & Tirrell DA (2004). Presentation and detection of azide functionality in bacterial cell surface proteins. J. Am. Chem. Soc 126, 10598–10602. [PubMed: 15327317]

Luo X, Fu G, Wang RE, Zhu X, Zambaldo C, Liu R et al. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. (2017). Nat. Chem. Biol 13, 845–849. [PubMed: 28604693]

Lu KP, Finn G, Lee TH, & Nicholson LK (2007). Prolyl *cis-trans* isomerization as a molecular timer. Nat. Chem. Biol, 3, 619–629. [PubMed: 17876319]

Lukesch MS, Pavkov-Keller T, Gruber K, Zangger K, & Wiltschi B (2019). Substituting the catalytic proline of 4-ozalcrotonate tautomerase with non-canonical analogues reveals a finely tuned catalytic system. Sci. Rep, 9, 2697. [PubMed: 30804446]

Lummis SCR, Beene DL, Lee LW, Lester HA, Broadhurst RW & Dougherty DA (2005). *Cis-trans* isomerization at proline opens the pore of a neurotransmitter-gated ion channel. Nature, 438, 248–252. [PubMed: 16281040]

MacArthur MW & Thornton JM (1991). Influence of proline residues on protein conformation. J. Mol. Biol, 218, 397–412. [PubMed: 2010917]

Montclare JK, Son S, Clark GA, Kumar K, & Tirrell DA (2008). Biosynthesis and stability of coiled-coil peptides containing (2*S*,4*R*)-5,5,5-trifluoroleucine and (2*S*,4*S*)-5,5,5-trifluoroleucine. ChemBioChem, 10, 84–86.

Mukai T, Lajoie MJ; Englert M, & Söll D (2017). Rewriting the genetic code. Annu. Rev. Microbiol, 71, 557–577. [PubMed: 28697669]

Renner C, Alefelder S, Bae JH, Budisa N, Huber R, & Moroder L (2001). Fluoroprolines as tools for protein design and engineering. Angew. Chem. Int. Ed, 40, 923–925.

Rubini M, Schärer MA, Capitani G, & Glockshuber R (2013). (4*R*)- and (4*S*)-fluoroproline in the conserved *cis*-prolyl peptide bond of the thioredoxin fold: Tertiary structure context dictates ring puckering. ChemBioChem, 14, 1053–1057. [PubMed: 23712956]

Sakamonto K, Hayashi A, Sakamonto A, Kiga D, Nakayama H, Soma A, et al. (2002). Site-specific incorporation of unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res., 30, 4692–4699. [PubMed: 12409460]

Shoulders MD & Raines RT (2009). Collagen structure and stability. Annu. Rev. Biochem, 78, 929–958. [PubMed: 19344236]

Steiner T, Hess P, Bae JH, Wiltschi B, Moroder L, & Budisa N (2008). Synthetic biology of proteins: Tuning GFPs folding and stability with fluoroproline. PLoS ONE, 3, e1680. [PubMed: 18301757]

Tanrikulu IC; Schmitt E; Mechulam Y; Goddard III WA; & Tirrell DA (2009) “Discovery of *Escherichia coli* methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine *in vivo*. Proc. Natl. Acad. Sci. USA, 106, 15285–15290. [PubMed: 19706454]

Teeuwen RLM; van Berkel SS, van Dulmen THH, Schoffelen S, Meeuwissen SA, Zuilhof H, et al. (2009) “Clickable” elastins: Elastin-link polypeptides functionalized with azide or alkyne groups. Chem. Commun, 27, 4022–4024.

Torbeev VY, & Hilvert D (2013). Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate β2 microglobulin amyloid assembly. Proc. Natl. Acad. Sci. USA, 110, 20051–20056. [PubMed: 24262149]

Wang J, Xie J, & Schultz PG (2006). A genetically encoded fluorescent amino acid. J. Am. Chem. Soc 128, 8738–8739. [PubMed: 16819861]

Wolschner C, Giese A, Kretzchmar HA, Huber R, Moroder L, Budisa N (2009). Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. Proc. Natl. Acad. Sci. USA, 106, 7756–7761. [PubMed: 19416900]

Xiu X, Puskar NL, Shanata JAP, Lester HA, & Dougherty DA (2009). Nicotine binding to brain receptors requires a strong cation-π interaction. Nature, 458, 534–537. [PubMed: 19252481]

Yoo TH, Link AJ, & Tirrell DA (2007). Evolution of a fluorinated green fluorescent protein. Proc. Natl. Acad. Sci. USA, 104, 13887–13890. [PubMed: 17717085]

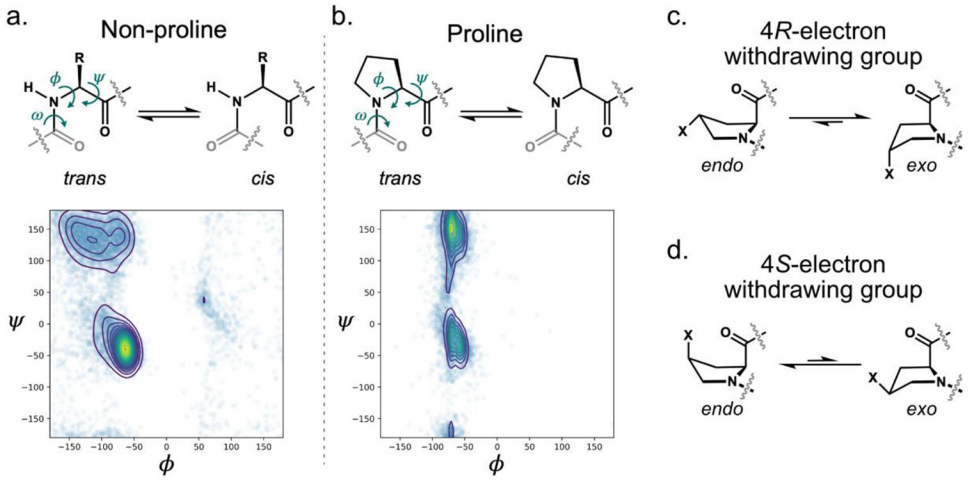


Figure 1.

Proline conformational preferences. **a-b**. Backbone conformations of and Ramachandran plots for non-proline (**a**) and proline (**b**) residues found in structural data deposited in the Protein Data Bank (PDB). The torsional angles ψ, ϕ, and ω, along with *cis* and *trans* amide isomers are indicated in the amino acid structures. **c-d**. *Endo* and *exo* ring pucker preferences for proline analogs containing electron withdrawing substituents at the 4 position. ncPro residues with 4*R*-electron withdrawing groups prefer the *exo* pucker (**c**); those with 4*S*-electron withdrawing groups prefer *endo* (**d**).

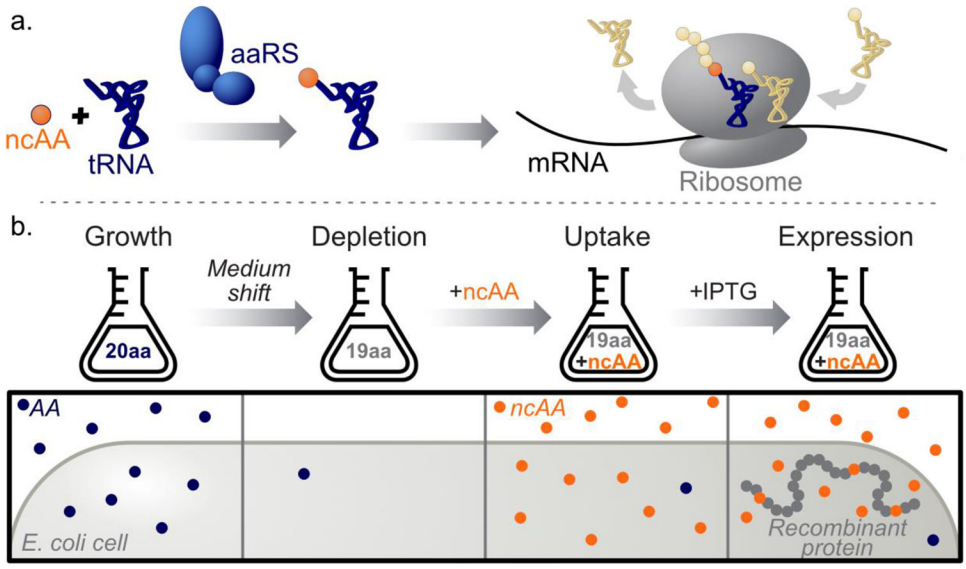


Figure 2.

Non-canonical amino acid (ncAA) mutagenesis. **a**. For both site-specific and residue-specific approaches, incorporation of ncAA residues into recombinant proteins relies on the ability of the translational machinery of the host to accommodate the ncAA of interest. In many cases, the limiting step is aminoacylation by the relevant aminoacyl-tRNA synthetase (aaRS, left). Upon aminoacylation, the charged tRNA is delivered to the ribosome, which adds the ncAA to the growing polypeptide chain (right). **b**. Residue-specific ncAA mutagenesis workflow. After growth in medium that contains the canonical amino acid (AA) to be replaced, a shift to AA-depleted medium is performed. After addition of the ncAA of interest, expression of the recombinant POI is induced. The ncAA replaces the canonical AA residues in all newly synthesized proteins, including the POI.

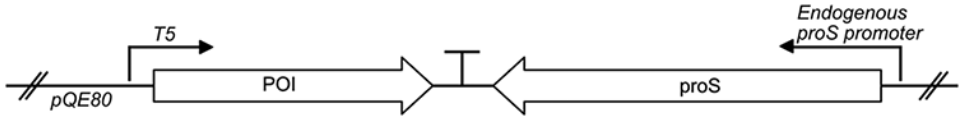


Figure 3.

The plasmid pQE80-PI-proS described here. The POI (in this case, proinsulin) was inserted into the multiple cloning site of the vector pQE-80L using restriction enzyme cloning. Expression is controlled by the T5 promoter. The gene encoding the *E. coli* ProRS was amplified by colony PCR and inserted downstream of the POI by Gibson Assembly.

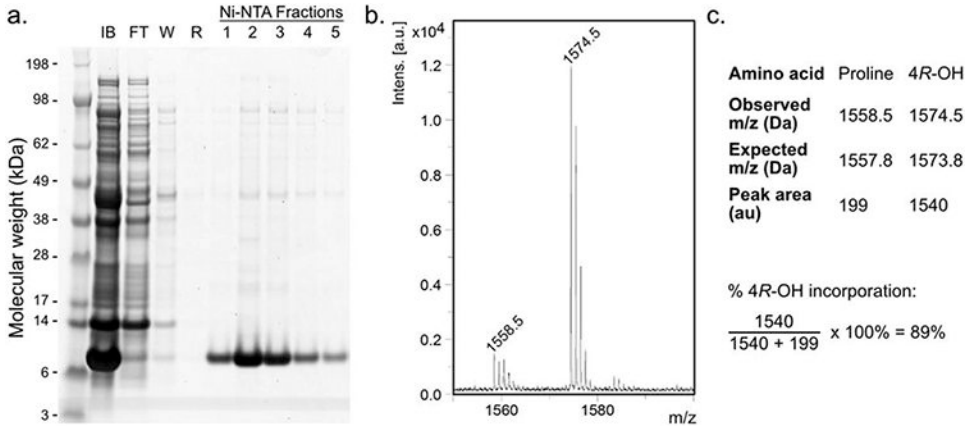
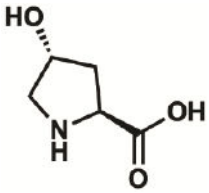
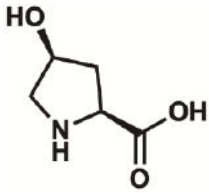
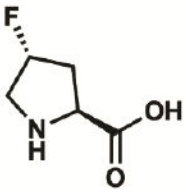
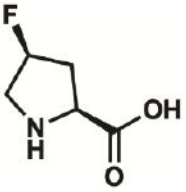
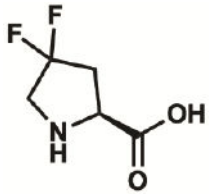
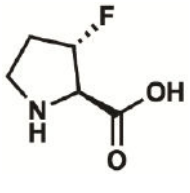


Figure 4.

Representative analysis of proline analog incorporation after POI expression. **a**. SDS-PAGE analysis of proinsulin expression and purification. After proline analog incorporation into proinsulin (10.7 kDa), the inclusion body (IB) fraction was solubilized in 8 M urea. After incubation with Ni-NTA resin, the flow-through (FT), wash (W; 20 mM imidazole), and rinse (R; no imidazole) fractions were collected. Proinsulin was eluted under low pH (3.0) conditions, and fractions were collected, then pooled. **b, c**. MALDI-TOF spectrum (**b**) and analysis (**c**) of proline analog incorporation after GluC digestion. Relevant proinsulin fragments: ^{33}RGFFYT**P**KTRRE (expected: 1557.8 Da) and ^{33}RGFFYT–**4*R*-OH**–KTRRE (1573.8 Da).

Table 1:

Proline analogs amenable to residue-specific incorporation into recombinant proteins

Proline analog	Proline structure	CAS	ProRS variant	Reported [NaCl] (M)	Reported level of proline replacement	Reference
4*R*-OH		51-35-4	wt	0.3	88%	Lieblich, et al., 2017
4*S*-OH		618-27-9	wt	0.3	91%	Lieblich, et al., 2017
4*R*-F		2507-61-1	wt	0.3	96%	Lieblich, 2017
4*S*-F		2438-57-5	wt	0.3	97%	Lieblich, 2017
4,4-diF		52683-81-5	wt	0.3	86%	Lieblich, 2017
3*S*-F	 *	261350-70-3	–	–	“virtually complete”	Kim, Hardcastle, & Conticello, 2006

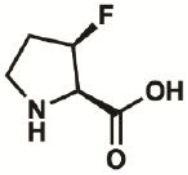
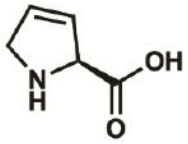
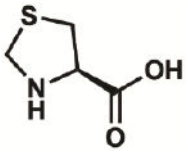
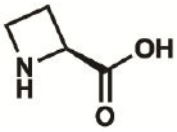
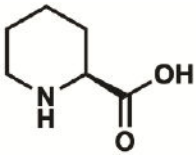
Proline analog	Proline structure	CAS	ProRS variant	Reported [NaCl] (M)	Reported level of proline replacement	Reference
3*R*-F		261350-69-0	–	–	"virtually complete"	Kim, Hardcastle, & Conticello, 2006
Dhp		4043-88-3	wt	0.3	≤100%	Fang, Lieblich, & Tirrell, 2019
Thz		34592-47-7	wt	0.3	90%	Fang, Lieblich, & Tirrell, 2019
Aze		2133-34-8	wt	0.3	≥100%	Fang, Lieblich, & Tirrell, 2019
Pip		3105-95-1	C443G	0.3	89%	Fang, Lieblich, & Tirrell, 2019

Table 2:

Stock solutions for M9 Medium

Component	Stock concentration	Mass	Stock Volume	Sterilization	Suggested storage conditions
19aa (5x) *	250 mg/L each aa	125 mg each aa	500 mL	Sterile filter	4 °C, 1 month
Proline (5x)	250 mg/L	125 mg	500 mL	Sterile filter	4 °C, 1 month
M9 salts (5x)	8.5 mM NaCl 18.7 mM NH_4Cl 22 mM KH_2PO_4 47.8 mM Na_2HPO_4	2.5 g NaCl 5.0 g NH_4Cl 15.0 g KH_2PO_4 64.0 g Na_2HPO_4·7H2O	1 L	Autoclave	RT
Glucose (10x)	20%	36.03 g	1 L	Sterile filter	RT
Biotin (100x) ‡	1 mg/mL	12 mg	12 mL	Sterile filter	Use day-of
Thiamine (1,000x)	35 g/L	420 mg	12 mL	Sterile filter	Use day-of
$FeSO_4$ (1,000x)	3 g/L	66 mg $FeSO_4$·7H2O	12 mL	Sterile filter	Use day-of
$MgSO_4$ (1,000x)	1 M	2.95 g $MgSO_4$·7H2O	12 mL	Sterile filter	RT
$CaCl_2$ (1,000x)	0.1 M	147 mg $CaCl_2$·2H2O	12 mL	Sterile filter	RT
Trace metals (10,000x) §	10 mg/L each metal	3.9 mg $CuSO_4$·$5H_2O$ 3.6 mg $MnCl_2$·$4H_2O$ 4.4 mg $ZnSo_4$·$7H_2O$ 1.8 mg $(NH_4)_6Mo7O_{24}$·$4H_2O$	100 mL	Sterile filter	RT
Water				Autoclave	RT

* Stir ~1hr on stir plate at room temperature (RT) for all amino acids to dissolve.

‡ Add ~10 μL of 10 M NaOH to fully dissolve biotin before sterile filtering.

§ It is recommended to prepare the trace metals solution as a 100,000x stock in 10 mL, then dilute 1 mL 1:10 with H_2O to obtain a 10,000x stock solution.

Table 3:

M9 Medium

Component	Stock volume
19aa (5x)	250 mL
Proline (5x) *	250 mL
M9 salts (5x)	250 mL
Glucose (10x)	125 mL
Biotin (100x)	1.25 mL
Thiamine (1,000x)	1.25 mL
$FeSO_4$ (1,000x)	1.25 mL
$MgSO_4$ (1,000x)	1.25 mL
$CaCl_2$ (1,000x)	1.25 mL
Trace metals (10,000x)	125 μL
Water	365 mL
Total	**1.25 L**

* Omit 5x proline solution for 1.25x 19aa M9 preparation

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
Competent *E. coli* for cloning (e.g. NEB 10-beta)	NEB	C3019H
E. coli proline auxotroph (e.g. CAG18515; see Note 1)	CGSC	7331
Chemicals, Peptides, and Recombinant Proteins		
Restriction enzymes (e.g. EcoRI-HF & BamHI-HF)	NEB	R3101; R3136
T4 DNA ligase	NEB	M0202
Q5 High-Fidelity 2x Master Mix	NEB	M0492
repliQa HiFi Assembly Mix	Quantabio	95190
Glu-C protease (or other appropriate protease; see Note 5)	Promega	V1651
M9 Medium Components		
20 proteinogenic amino acids	Sigma-Aldrich	various
Sodium chloride (NaCl, CAS: 7647-14-5)	Sigma-Aldrich	S7653
Ammonium chloride (NH_4Cl, CAS: 12125-02-9)	Sigma-Aldrich	A9434
Potassium phosphate, monobasic (KH_2PO_4, CAS: 7778-77-0)	Sigma-Aldrich	P5655
Sodium phosphate dibasic heptahydrate ($Na_2HPO_4{\cdot}7H_2O$, CAS: 7782-85-6)	Sigma-Aldrich	S9390
D-Glucose (CAS: 50-99-7)	Sigma-Aldrich	G7021
Biotin (CAS: 58-85-5)	Sigma-Aldrich	B4639
Thiamine hydrochloride (CAS: 67-03-8)	Sigma-Aldrich	T1270
Iron(II) sulfate heptahydrate ($FeSO_4{\cdot}7H_2O$, CAS:7782-63-0)	Sigma-Aldrich	F8633
Magnesium sulfate hepahydrate ($MgSO_4{\cdot}7H_2O$, CAS: 10034-99-8)	Sigma-Aldrich	M2773
Calcium chloride dihydrate ($CaCl_2{\cdot}2H_2O$, CAS: 10035-04-8)	Sigma-Aldrich	C7902
Copper(II) sulfate pentahydrate ($CuSO_4{\cdot}5H_2O$, CAS: 7758-99-8)	Sigma-Aldrich	C8027
Manganese(II) chloride tetrahydrate ($MnCl_2{\cdot}4H_2O$, CAS: 13446-34-9)	Sigma-Aldrich	63535
Zinc sulfate heptahydrate ($ZnSO_4{\cdot}7H_2O$, CAS: 7446-20-0)	Sigma-Aldrich	Z0251
Ammonium molybdate tetrahydrate [$(NH_4)6Mo_7O_{24}{\cdot}4H_2O$, CAS: 12054-85-2]	Sigma-Aldrich	M1019
Commercial Kits		
Plasmid Miniprep Kit	Zymo Research	D4015
Gel DNA Recovery Kit	Zymo Research	D4001
Oligonucleotides		
proS-ins-fwd: 5’- GTGAGAATCCAAGCTAGCTCAGCCTTTAATCT GTTTCACCAG-3’	IDT	
proS-ins-rev: 5’- CGTATAATATTTGCCCATGGATTCACGCCCTT CT CTTTTGAC -3’	IDT	
proS-vec-fwd: 5’- GAGAAGGGCGTGAATCCATGGGCAAATATTA TACGCAAG -3’	IDT	
proS-vec-rev: 5’- GTGAAACAGATTAAAGGCTGAGCTAGCTTGG ATTCTCACC-3’	IDT	
Recombinant DNA		
pQE-80L	Qiagen	
Gene encoding POI	IDT	gBlock gene fragment