*Chapter 2*

# HOW TO DO A CELL-SELECTIVE PROTEOMICS EXPERIMENT USING BIOORTHOGONAL NONCANONICAL AMINO ACID TAGGING (BONCAT)

## 2.1 Abstract

This protocol describes how to cell-selectively tag and identify newly-synthesized proteins using azide-containing noncanonical amino acids. Tagged proteins can be analyzed with conventional biochemical methods or liquid chromatography-tandem mass spectrometry (LC-MS/MS). The protocol involves an initial cloning step, but the tagging, detection, and enrichment steps can proceed over 3-5 days. Notably, this protocol does not require depletion of any amino acids in the media and can even be used in animals.

This work was a collaboration with Judy Shon, Graham D. Hamblin, Weslee S. Glenn, Brett M. Babin, Kai P. Yuet, Alborz Mahdavi, Annie Moradian, Michael J. Sweredoski, and Sonja Hess.

## 2.2 Introduction

Experiments that require proteomic analysis of subpopulations of cells within a complex multicellular environment often can be like looking for needles in a haystack. Whether an investigator is looking at proteins secreted by a pathogen into host cells (1), or defining the proteome of particular cell types such as neurons or glia within an entire organism (2, 3), cell-selective metabolic labeling methods are useful tools for distinguishing the needles from the hay.

Our lab has developed cell-selective bioorthogonal noncanonical amino acid tagging (BONCAT) as a metabolic labeling method that can target specific subpopulations of cells. Achieving this specificity requires the expression of a mutant aminoacyl-tRNA synthetase (aaRS) under a promoter that is only active in the cell type of interest (Fig 2.1). The mutant aaRS has been engineered to incorporate a noncanonical amino acid (ncAA) in a residue-selective manner. For example, we have created both a mutant methionyl-tRNA synthetase (MetRS) (1, 4, 5) and phenylalanyl-tRNA synthetase (PheRS) (3), which can incorporate azido analogs of methionine and phenylalanine residues, respectively (Fig 2.2). The azide moiety allows the newly synthesized proteins of only distinct cell types to be identified and enriched from the rest of the proteins using strain-promoted azide-alkyne cycloaddition (SPAAC) (Fig 2.1B) (4, 6). Cells within the population that do not express the mutant aaRS cannot incorporate the ncAA into their proteins. Because the mutant aaRSs charge tRNAs with the ncAA at high enough rates to compete with charging of canonical substrates, this method does not require depletion

of amino acids or use of minimal medium, and can be accomplished in whole organisms. Our goal is to make this method straightforward so that even non-specialized laboratories can perform a cell-selective proteomic experiment. Mutant aaRSs are available through Addgene, and all reagents are commercially available. Analyzing subpopulations of cells does not have to be nearly as daunting as our original needles and haystack analogy suggests.

**Applications of the method**

While complete depletion of a canonical amino acid may perturb biological systems, BONCAT is non-toxic and has no obvious effect on protein synthesis or degradation in media that has canonical amino acids present (7, 8). Previous BONCAT experiments that were not cell-selective relied on the cell's endogenous protein synthesis machinery to incorporate ncAAs such as azidohomoalanine (Aha) and homopropargylglycine (Hpg) (Fig 2.2). To achieve high incorporation with these ncAAs, cells were often first depleted of their natural methionine (Met) reserves. Cell-selective BONCAT does not require such depletion, as the mutant aaRSs generally prefer the ncAAs over the canonical residue. Depending on the expression level and activity of the endogenous aaRS in the organism of interest, optimal labeling is usually achieved at either equimolar concentrations of free natural and noncanonical residues or reasonable 10- to 50-fold excesses of the noncanonical residue. For example, the mutant *E. coli* NLL-MetRS prefers Anl 1.2 times more than Met (4), and the mutant PheRS prefers 4-azidophenylalanine (Azf) 42.7 times more than phenylalanine (Phe) (9). In comparison,

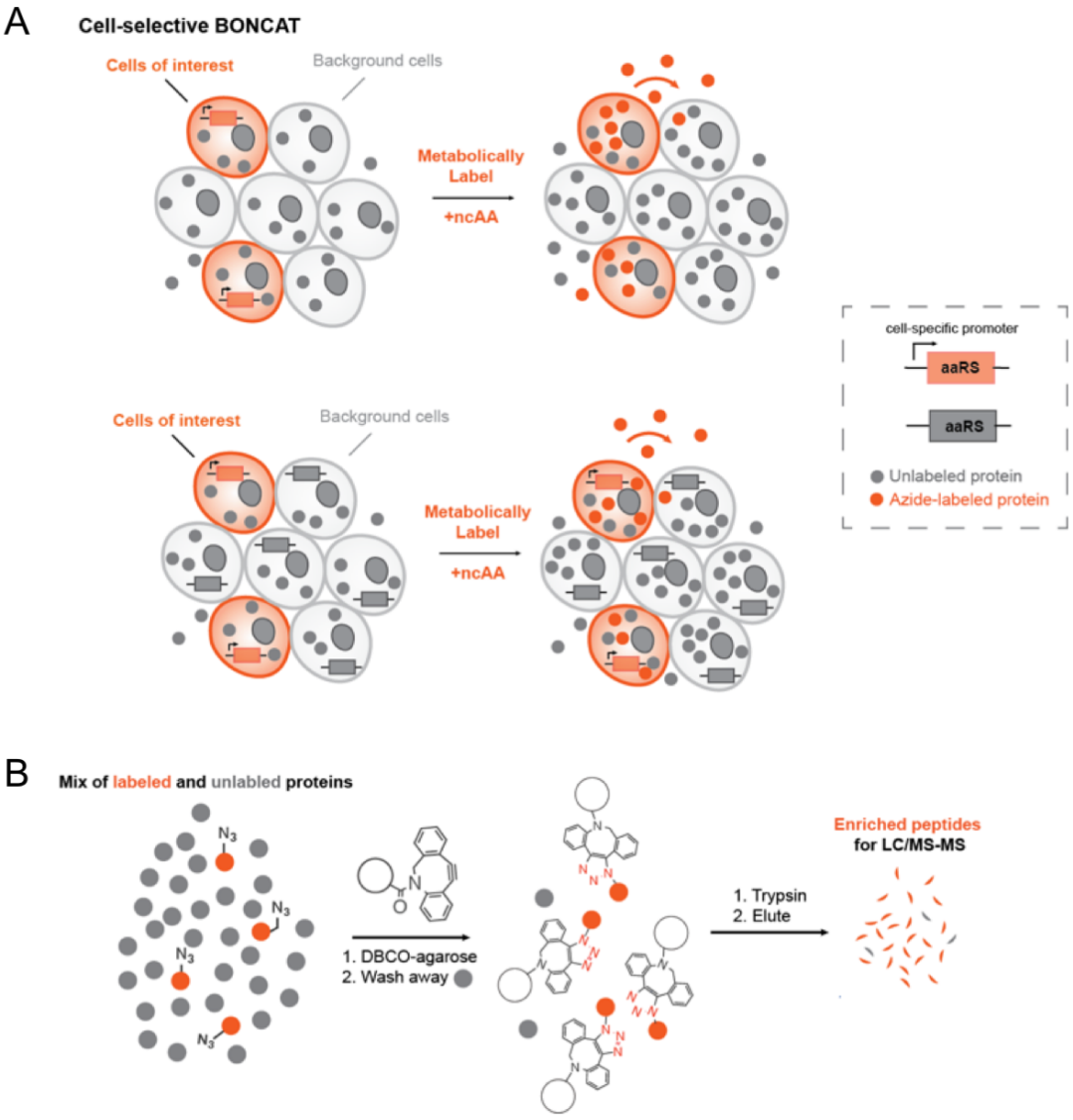


**Figure 2.1:** A) Scheme depicting two different versions of cell-specific BONCAT. In the first scheme, cells have the mutant aaRS in the presence of cells that do not. In the second scheme, all cells have the mutant aaRS but a cell-specific promoter drives expression only in the cells of interest. B) Scheme depicting strain-promoted azide-alkyne cycloaddition (SPAAC) based enrichment strategy for detection of tagged proteins. Azide-tagged proteins are conjugated to DBCO-agarose beads and untagged proteins can be washed away. On-bead digestion of enriched proteins reveals many peptides that can be analyzed using liquid chromatography tandem mass spectrometry (LC-MS/MS).

the $k_{cat}/ k_M$ (a measure of substrate preference) is 400-fold higher for the natural substrate of wild-type *E. coli* MetRS than for the ncAA analog incorporates Aha (6), which explains why previous BONCAT experiments needed to deplete cells of Met and add prohibitively high Aha concentrations in order to achieve a high degree of labeling. The mouse L274GMetRS was engineered to incorporate Anl about 4 times less than its natural substrate Met when the two compounds are in equal concentrations (2, 10). Cell-selective BONCAT can work under normal cell growth conditions and has even been used to label proteins inside living animals (2).

While we use strain-promoted azide-alkyne cycloaddition (SPAAC) to enrich azide-labeled proteins in this protocol, we envision similar design strategies may also be useful for other click chemistries, such as tetrazine ligation or copper-catalyzed alkyne-azide cycloadditions (CuAAC). Furthermore, this method is compatible with other metabolic labeling methods, including stable isotope labeling with amino acids in culture (SILAC) for quantitative proteomics (3).

**Comparisons with other methods**

The design strategies outlined here apply to other useful metabolic labeling methods in the literature, such as stochastic orthogonal recoding of translation (SORT) (11, 12). Physical sorting methods like fluorescence activated cell sorting (FACS) are required when cells come from clinical samples or are genetically intractable; however, these methods can produce artifacts during the sample preparation process.

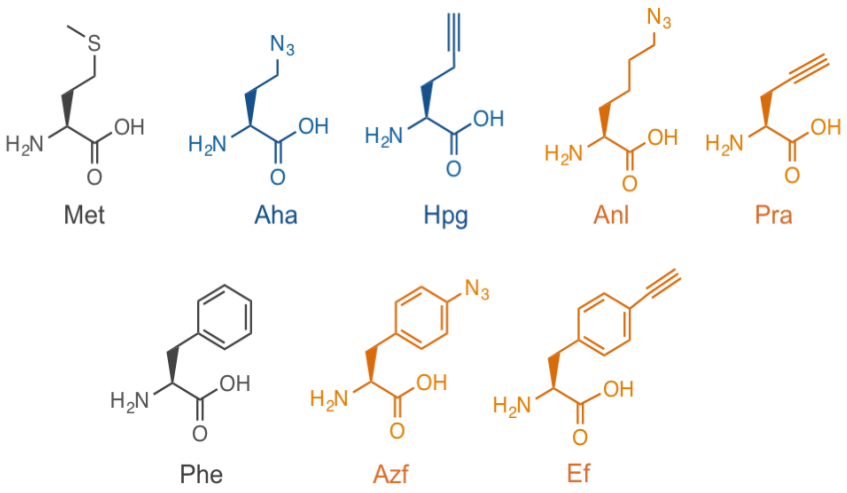


**Figure 2.2:** Noncanonical amino acids (ncAAs) and the corresponding residue they replace. Orange residues require the expression of a mutant synthetase in order to be incorporated into proteins, and thus, can be used for cell-selective BONCAT. Blue residues can be incorporated by cell's endogenous protein synthesis machinery and cannot be used for cell-selective BONCAT. **Met**: methionine. **Aha**: azidohomoalanine. **Hpg**: homopropargylglycine. **Anl**: azidonorleucine. **Pra**: propargylglycine. **Phe**: phenylalanine. **Azf**: 4-azido-phenylalanine. **Ef**: 4-ethynyl-phenylalanine

### Limitations of the method

Cell-selective BONCAT can only be used in genetically-tractable systems due to its reliance on the expression of a mutant aaRS. Promoters that are only active in the cellular subset of interest must also be established or characterized prior to use. In order to initiate labeling, one needs a way to deliver the ncAA to the cells of interest in sufficient quantities. Finally, the investigator must have the organism's genome available to theoretically digest the proteome and compare it to identified fragments.

**Design of a cell-specific BONCAT experiment**

Figure 2.3 illustrates the workflow of the entire procedure, and the details are described below. The investigator must first choose a unique promoter from the cell type of interest and clone the mutant synthetase to be under the control of that promoter. Next, expression of the synthetase and labeling efficiency should be tested *in vitro*; an appropriate negative control is required. Labeling should be compared with cells lacking the mutant synthetase by using click chemistry in combination with either in-gel fluorescence or fluorescence microscopy. After confirmation of cell-type specific expression, the experiment should be run in (at least) biological triplicate with enough material for the proteomics enrichment.

The choice of synthetase may be limited by the cell system of interest. We recommend testing both the MetRS and PheRS to see which system exhibits strong labeling without compromising cellular growth. Elliot *et al.* suggest that some bias in the set of proteins identified can depend on the residue chosen for noncanonical replacement and bioorthogonal reaction (11); labeling with the mutant PheRS or mutant MetRS may also exhibit a similar bias.

It is important to note that total replacement of the chosen residue is not necessary or probably even desirable for detection and enrichment of proteins from the rest of the cellular milieu. In fact, one ncAA per protein is ideal as it is enough to attach the protein to the enrichment resin, and would minimize perturbation of function and dynamics.

This approach also leaves the rest of the protein to be digested by trypsin for detection and identification in the mass spectrometer. In our experience, searching for the site of ncAA incorporation is difficult and unnecessary for identification and quantification of cell-specific proteins.

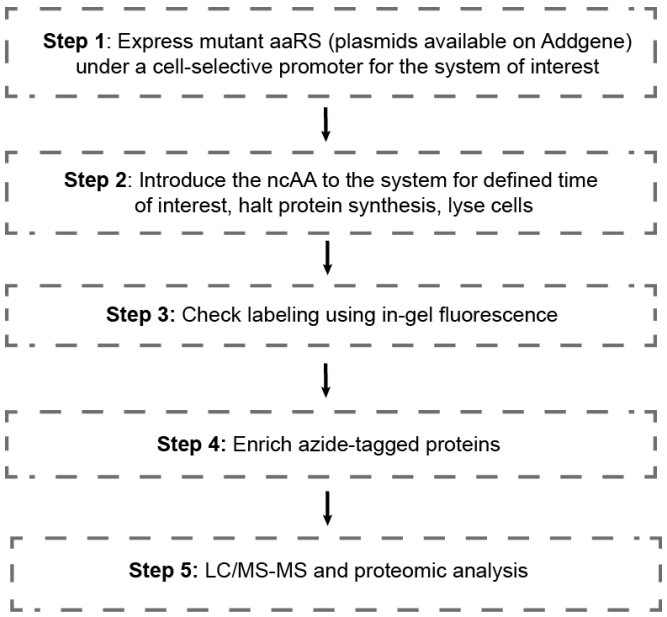


**Figure 2.3:** Workflow of cell-selective BONCAT protocol outlined in the text.

**Materials**

Reagents:

- DNA for cloning the mutant aaRS into cells of interest (MetRS and PheRS variants for prokaryotic and eukaryotic expression available at https://www.addgene.org/David_Tirrell/)
- Corresponding azido-noncanonical
    - L-azidolysine hydrochloride (Iris Biotech cat. no. HAA1625)
    - 4-azido-L-phenylalanine hydrochloride (Iris Biotech cat. no. HAA1850)
- 4-12% NuPage Bis-Tris polyacrylamide gels (Thermo Fisher cat. no NP0322BOX) or similar
- Complete EDTA-free Protease Inhibitor Tablets (Roche, cat. no. 1873580)
- Dithiothreitol (DTT)
- Benzonase (Sigma, cat. no. E1014)
- Iodoacetamide (Sigma, cat. no. I1149): 0.5 M in molecular biology grade water, prepare fresh before use
- DBCO-TAMRA (Click Chemistry Tools, cat. no. A131)
- 4x Laemmli Sample Buffer (Bio Rad cat. no. 161-0747)
- InstantBlue (Expedeon cat. no. ISB1L)
- Protein ladder (such as SeeBlue)
- Destain solution (10% acetic acid, 40% methanol, 50% dd$H_2O$)
- SDS wash buffer: 0.8% SDS, 0.15 M NaCl, in 100 mM Tris pH 8.0
- Urea wash buffer: 8 M urea, 0.15 M NaCl, in 100 mM Tris pH 8.0
- DBCO-Agarose (Click Chemistry Tools, 1034-2)
- Poly-Prep Chromatography Columns (Bio-Rad #7311550)
- Sequencing-grade Trypsin (Promega, cat. no. V5111)
- Ammonium bicarbonate
- Formic Acid
- DMSO
- Acetonitrile
- HiPPR Detergent Removal Spin Column Kit (Thermo Fisher cat. no. 88305)
- Bicinchoninic acid (BCA) protein quantification (Pierce, cat. no 23227)

Equipment:

- Protein gel electrophoresis system
- Gel imaging system (such as GE Healthcare's Typhoon gel imager)

- Rotator
- Rocker
- Temperature-controlled shaker for microcentrifuge tubes (such as Thermomixer from Eppendorf)
- Speedvac
- Orbitrap MS

Software tools:

- MaxQuant (freely available (13))

**Procedure**

**Preparation of solutions**

- DBCO-TAMRA stock solution (5 mM): Dissolve DBCO-TAMRA in DMSO to obtain a final concentration of 5 mM (1000x stock). Prepare 10 µL aliquots of the stock solution in individual tubes and store them at −20 °C for up to 2 years. DBCO-TAMRA is light sensitive and should be kept in the dark.
- Trypsin: Resuspend 500 µg in 50 µL 1 N HCl in HPLC grade water. Prepare 5 µL aliquots of this 10 µg/µL solution, which can be stored at -20 °C for up to 1 year.

**Step 1: Plasmid construction and expression of the mutant aaRS**

1. Obtain the desired cell line or organism required for the experiment.
2. Insert the aaRS gene in a vector of your choice under a cell-specific promoter. As a negative control, do not insert the aaRS.
   Depending on the species of interest and the species of the aaRS, codon-optimization may be necessary. If the aaRS does not express well in your species of interest, check the codon usage of each.
   In this protocol, we have cloned a codon-optimized version of the *E. coli* NLL-MetRS (Addgene #51401) into *Staphylococcus aureus* under a tetracycline-

inducible promoter (Addgene #26252). Our control is this vector without the inserted MetRS.

Critical Step: The choice of promoter is very important, as it will determine the level of specificity of the experiment. It is recommended to check expression and cell specificity of the chosen promoter throughout the labeling period using a fluorescent protein or Western blot. Even a small degree of non-specific expression of the aaRS will result in off-target synthesis of tagged proteins and hinder results.

**Step 2: Cell-specific labeling**

5. Culture the cells as appropriate. There is no need to deplete the cells of amino acids.
6. Add the ncAA for the time period of interest.

   The amount of ncAA to add depends on the labeling time, the concentration of free amino acids in the growth medium, the rate of protein synthesis, and the natural expression level of endogenous MetRS. For most systems, we generally add 30x the amount of free Met when using Anl and equimolar amounts of Azf and Phe in solution. Using higher concentrations will increase labeling but also increase the probability of deleterious effects such as inhibition of growth. Investigators may wish to determine the optimal ncAA concentration and labeling time for each cell type on a small scale first.

   For example, the free methionine concentration in serum is ~30 μM; therefore, concentrations of 1 mM Anl procure sufficient cell-specific labeling for downstream analysis (13).
7. To stop labeling, add both a protein synthesis inhibitor, such as chloramphenicol (10 μg/mL) for bacterial cells or cycloheximide for mammalian cells (100 μg/mL), and protease inhibitors (1x) for 5 min.

   Without this step, protein labeling will continue during lysis and separation, and stress-response proteins will be found highly expressed.

8. Lyse cells in the presence of an alkylation agent.

   For *E. coli*, we often spin down the cells at 3500 rcf (relative centrifugal force, g) for 5 min, then lyse in 10% of the original volume in 2% SDS in 100 mM Tris (pH 8.0) with 100 mM chloroacetamide. For mammalian cells, the addition of 2% SDS in 100 mM Tris (pH 8.0) and 100 mM chloroacetamide lyses the cells. Heat the lysates at 65 °C to ensure alkylation of thiols.

   Critical: The addition of chloroacetamide will alkylate free thiols, which would nonspecifically react with cyclooctyne reagents (14). We have also had success with iodoacetamide and use the two interchangeably. Both of these solutions need to be made fresh on the day of the experiment and kept from light.

   If the lysates are highly viscous due to the presence of DNA, sonication or benzonase treatment should be performed.

   We recommend saving 10% of the volume of these labeled lysates in a separate small aliquot to test labeling efficiency using DBCO-TAMRA prior to enrichment. Additionally, they can be used as an "unenriched" control to approximate the degree of enrichment by using filter-aided sample preparation (FASP) (15) prior to LC-MS/MS analysis.

**Pause point: Labeled cell lysates can be stored at -80 °C for several months without any harmful effect on click chemistry enrichments.

**Step 3: Check protein concentration and degree of labeling of small aliquot**

9. Thaw lysates (if frozen). Spin 5 min at 12-14k rcf to clarify and keep the supernatant.
10. Perform BCA assay or other method of determining protein concentration as per manufacturer's instructions.
11. Take 20 µg of protein from each sample and bring to 5 µM DBCO-TAMRA for 15 min.
12. Add 4x Laemmli sample buffer with 50 mM DTT and boil for 2 min.

13. Allow samples to cool to <50 °C, then load ~10 μg into protein gel wells, along with a protein ladder.
14. Run protein gel at constant voltage (170 V) for 1 hour.
15. Carefully remove protein gel from cast and submerge in Destain solution (enough to cover the gel). Leave the gel in this solution on a rocker, covered from light, for at least 4 hours at room temperature, to both remove leftover unbound dye and fix the proteins within the gel.

**Pause point: The gel can be left for up to several days in this Destain solution as long as it is kept from light prior to imaging.

16. After disposing of the Destain solution, allow the gel to rehydrate in deionized water ($dH_2O$) for 15 min, then visualize the gel using a gel imaging system with appropriate laser and bandpass filter settings. For TAMRA ($\lambda_{ex}$ = 555 nm and $\lambda_{em}$ = 580 nm), we excite with a green laser at 532 nm and detect signal with a 580 band-pass 30 nm filter. Only the cells that expressed the mutant aaRS should show labeling in the protein lanes, exemplified in Fig 2.4.
17. Stain for total protein by adding enough InstantBlue to cover the gel for 15 min. Protein bands should become visible by eye, and can be imaged on the gel imaging system. Protein loading should be about equal between lanes.

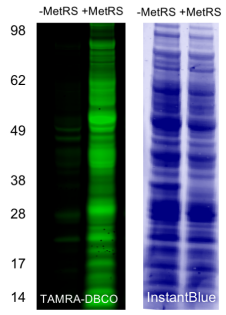


**Figure 2.4:** TAMRA gel of a cell-selective BONCAT experiment. Cells expressing the NLL-MetRS incorporate azides into their proteins, which can be conjugated to DBCO-TAMRA

**Step 4: Enrichment**

18. Thaw lysates (if frozen). Spin 5 min at 12-14k rcf to clarify and keep the supernatant.
19. Add equal volume 8 M urea/0.15 M NaCl/protease inhibitor in PBS, made fresh so the protein is resuspended in buffer with 0.5-1% SDS, and 4 M urea.
20. Wash ~25 µL per sample DBCO-agarose in 1 mL of 1% SDS three times. Resuspend in original volume of 1% SDS.
    Critical: Always centrifuge the resin at 1500 rcf or less. Spinning at faster speeds can result in destruction of the resin. We generally use ~25 µL of resin / 5mg enrichment. Using high amounts of resin results in much higher proteomic background.
21. Add washed resin to samples and rotate end over end for 16-24 hours.
22. Wash resin, now with covalently bound proteins, with 1 mL wash buffer to remove unbound proteins
    We often keep the supernatants at this step, in case the proteins did not bind.
23. Reduce bound proteins with 0.5 mL 5 mM DTT in SDS wash buffer for 30 min. Remove supernatant.
24. Alkylate bound proteins with 100 mM chloroacetamide or iodoacetamide in the dark for 45 min at 50 °C.
25. Transfer resin to Poly-prep chromatography column.

26. Wash resin with the following solutions:
    a. 8 x 5 mL SDS wash buffer
    b. 8 x 5 mL 8 M urea in 100 mM Tris pH 8.0
    c. 8 x 5 mL 20% acetonitrile
27. Transfer beads to Eppendorf tubes using 10% ACN in 50 mM Ammonium Bicarbonate (AmBi).
28. Spin 5 min at 1500 rcf, remove supernatant down to 100 µL.
29. Add 100 ng trypsin to each sample and incubate overnight at 37 °C.
30. Collect supernatant, then wash resin with 150 µL 20% acetonitrile twice and combine washes with supernatant. Be careful to avoid carrying over resin during transfer steps.
31. Speedvac to dryness.

**Pause point: Peptides can be stored at -20 °C for several months without any harmful effects.

32. Follow StageTip protocol to desalt peptides (17).
33. Resuspend in 8 µL 0.2% formic acid for injection onto the LC-MS/MS

**Step 5: LC-MS/MS and proteomic analysis (~8-48 hours, depending on the number of samples and replicates)**

34. Inject ~100 ng of the enriched lysate onto a liquid chromatography system coupled to an Orbitrap mass spectrometer, equipped with a nano-electrospray ion source.
35. Separate the peptide using a chromatographic separation for 1-3 hours using an elution gradient from 2 to 30% acetonitrile at a flow rate of 220 nL/min, and operate the mass spectrometry in data-dependent mode (18).
    We typically collect full scan mass spectra with 400-1600 m/z, and collect the top 20 most intense ions from the survey scan for fragmentation in the linear ion

trap by collision-induced dissociation (CID). We use precursor ion charge state screening to reject singly charged and unassigned charge states.

36. Take the raw files from the LC-MS/MS run and process using MaxQuant as previously described (13, 19).

    Set carbamidomethylation of cysteine as a fixed modification, and protein N-terminal acetylation, N-terminal formylation, and methionine oxidation as variable modifications. We also include variable modifications of methionine corresponding to Anl (+23.0450) and reduced Anl (-2.9455), but often do not find many sites of Anl labeling because they are left on the resin during the digestion step.

**Table 2.1: Troubleshooting Table**

| Step | Problem | Possible Reason | Solution |
|---|---|---|---|
| 8 | Viscous samples after lysis | Insufficient genomic DNA lysis | Use more Benzonase to facilitate lysis of genomic DNA, shear DNA with a syringe and a needle, or sonicate samples |
| 17 | Low levels of tagged proteins | Labeling time too short or not enough ncAA added | Add the ncAA for a longer time or increase concentration |
| 35 | Few proteins found, LC trace has most signal towards end of run | Leftover detergent in samples | Use HiPPR detergent removal columns or increase the number of washes of the resin |
| 35 | PEG present in LC-MS | PEG contamination | Use as few transfers of the material as possible into new Eppendorf tubes, as the sample will pick up trace plasticizers over time, do not use autoclaved plastics as this increases the amount of plastics |

**Timing**

Step 1: Cloning of aaRS into cells of interest (variable)

Varies depending on system chosen

Step 2: Cell labeling and lysis (1-24 hours)

Varies depending on time window of interest to label newly synthesized proteins

Step 3: Testing labeling using in-gel fluorescence (6 hours)

Step 4: Enrichment and preparation of proteins for MS (48 hours)

Step 5: LC-MS/MS and data analysis (2 days to 2 weeks**)**

**Anticipated Results**

Using this method, we often find 1000-5000 proteins from the cell-type of interest.

**References**


1. Mahdavi A, *et al*. (2014) Identification of secreted bacterial proteins by noncanonical amino acid tagging. *Proc Natl Acad Sci* 111(1):433-438.
2. Erdmann I, *et al*. (2015) Cell-selective labelling of proteomes in *Drosophila melanogaster*. *Nat Comm* 6.
3. Yuet KP, *et al*. (2015) Cell-specific proteomic analysis in *Caenorhabditis elegans*. *Proc Natl Acad Sci* 112(9):2705-2710.
4. Tanrikulu IC, Schmitt E, Mechulam Y, Goddard 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* 106(36):15285-15290.
5. Ngo JT, *et al*. (2009) Cell-selective metabolic labeling of proteins. *Nat Chem Biol* 5(10):715-717.
6. Kiick KL, Saxon E, Tirrell DA, & Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci* 99(1):19-24.
7. Bagert JD, *et al*. (2014) Quantitative, Time-Resolved Proteomic Analysis by Combining Bioorthogonal Noncanonical Amino Acid Tagging and Pulsed Stable Isotope Labeling by Amino Acids in Cell Culture. *Mol Cell Prot* 13(5):1352-1358.
8. Wang JG, *et al*. (2017) Nonradioactive quantification of autophagic protein degradation with (L)-azidohomoalanine labeling. *Nat Protoc* 12(2):279-288.
9. Carrico IS (2004) Protein engineering through *in vivo* incorporation of phenylalanine analogs. Thesis Dissertation. Chapter 4.

10. Mahdavi A, *et al*. (2016) Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis. *J Am Chem Soc* 138(13):4278-4281.
11. Elliott TS, Bianco A, Townsley FM, Fried SD, & Chin JW (2016) Tagging and Enriching Proteins Enables Cell-Specific Proteomics. *Cell Chem Biol* 23(7):805-815.
12. Elliott TS, *et al*. (2014) Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat Biotech* 32(5):465-U186.
13. Stein WH & Moore S (1954) The Free Amino Acids of Human Blood Plasma. *J Biol Chem* 211(2):915-926.
14. van Geel R, Pruijn GJM, van Delft FL, & Boelens WC (2012) Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide-Alkyne Cycloaddition. *Bioconj Chem* 23(3):392-398.
15. Wisniewski JR, Zougman A, Nagaraj N, & Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Meth* 6(5):359-U360.