



Published in final edited form as:

*Methods Enzymol.* 2021 ; 659: 37–70. doi:10.1016/bs.mie.2021.07.007.

## Use of tandem affinity–buffer exchange chromatography online with native mass spectrometry for optimizing overexpression and purification of recombinant proteins

Stella M. Lai<sup>1,2,3</sup>, Pankajavalli Thirugnanasambantham<sup>1,2,†</sup>, Vaishnavi Sidharthan<sup>1,2,†</sup>, Andrew Norris<sup>1,2,3,‡</sup>, Jamison D. Law<sup>1,2,4,‡</sup>, Venkat Gopalan<sup>1,2,\*</sup>, Vicki H. Wysocki<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA

<sup>2</sup>Center for RNA Biology, The Ohio State University, Columbus, OH 43210, USA

<sup>3</sup>Resource for Native Mass Spectrometry-Guided Structural Biology, The Ohio State University, Columbus, OH 43210, USA

<sup>4</sup>Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA

### Abstract

Purification of recombinant proteins typically entails overexpression in heterologous systems and subsequent chromatography-based isolation. While denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis is routinely used to screen a variety of overexpression conditions (e.g., host, medium, inducer concentration, post-induction temperature and/or incubation time) and to assess the purity of the final product, its limitations, including aberrant protein migration due to compositional eccentricities or incomplete denaturation, often preclude firm conclusions regarding the extent of overexpression and/or purification. Therefore, we recently reported an automated liquid chromatography–mass spectrometry-based strategy that couples immobilized metal affinity chromatography (IMAC) with size exclusion-based online buffer exchange (OBE) and native mass spectrometry (nMS) to directly analyze cell lysates for the presence of target proteins. IMAC-OBE-nMS can be used to assess whether target proteins (1) are overexpressed in soluble form, (2) bind and elute from an IMAC resin, (3) oligomerize, and (4) have the expected mass. Here, we use four examples to demonstrate the potential of IMAC-OBE-nMS for expedient optimization of overexpression and purification conditions for recombinant poly-His-tagged protein production.

### Keywords

liquid chromatography–mass spectrometry; native mass spectrometry; online buffer exchange; protein overexpression; immobilized metal affinity chromatography

\*To whom correspondence should be addressed: [VG] tel. +1 614 292 1332, gopalan.5@osu.edu; [VHW] tel.: +1 614 292 8687; wysocki.11@osu.edu.

†These authors contributed equally to this work as second authors

‡These authors contributed equally to this work as third authors

*Conflict of interest statement.* The desalting cartridge used in this study is a prototype of a commercial product that is being developed by Thermo Scientific with input and testing by the OSU team.

## 1 Introduction

*In vitro* structure–function studies and biotherapeutic initiatives rely heavily on the ability to overexpress and purify large quantities of recombinant proteins of interest to near-homogeneity. Typically, well-characterized and cost-effective heterologous overexpression systems such as *Escherichia coli* are used for protein production. Column-based chromatographic separation strategies are then used to exploit intrinsic properties of target proteins (e.g., isoelectric point, hydrophobicity (Jungbauer & Hahn, 2009; McCue, 2009)) and/or genetically engineered affinity tags (e.g., poly-His-tags (Structural Genomics Consortium et al., 2008)) to separate proteins of interest from host cell proteins. While this workflow has enabled biochemical and biophysical characterization of numerous proteins from organisms in all three domains of life, some proteins remain challenging to overexpress in non-native hosts for a variety of reasons. Factors range from host cell translation issues (e.g., rare codon usage, translation rates that cause misfolding) to target protein attributes (e.g., poor solubility, propensity to aggregate, cellular toxicity (Francis & Page, 2010; Lebendiker & Danieli, 2014; Rosano & Ceccarelli, 2014)).

To study previously uncharacterized proteins, it is common to perform small-scale pilot overexpression experiments in which conditions are individually varied to assess their effect on the overall yield of soluble protein. Along with traditional overexpression optimization strategies like titrating inducer concentrations or changing post-induction temperatures and incubation times, genetically engineered overexpression strains (e.g., *E. coli* SixPack cells (Lipinszki et al., 2018)), supplemented growth media (e.g., Terrific Broth (Tartof & Hobbs, 1987), Dynamite medium (Taylor, Denson, & Esposito, 2017)), and co-expressed molecular chaperones (e.g., GroESL, DnaK, GrpE, ClpB (Marco, 2007)) offer possibilities for improving the yield of otherwise poorly overexpressed or misfolded recombinant proteins. Outcomes of these pilot tests, some of which are depicted in Figure 1, are typically evaluated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). SDS-PAGE is an inexpensive and routinely used method that can provide a rapid readout on whether a protein has been successfully overexpressed and is approximately the correct molecular weight, thereby facilitating the identification of optimal overexpression conditions. However, SDS-PAGE is not without its limitations. For example, misidentification can occur if proteins migrate in a manner inconsistent with their predicted molecular weight due to either amino acid composition (e.g., a high proportion of charged residues) or overexpression as a truncated protein (Guan et al., 2015; R. Verheijen, M. Salden, & W.J. van Venrooij, 1996; Shi et al., 2012). Additionally, the presence of a band corresponding to a poorly overexpressed target protein may be obscured by high background from the presence of abundant host cell proteins, some of which may co-migrate with the protein of interest. These challenges are compounded by the abovementioned difficulties associated with recombinant protein overexpression, and multiple rounds of troubleshooting and optimization are often required to maximize the yield of pure recombinant protein. Thus, robust and sensitive methods that are complementary to SDS-PAGE are needed to help rapidly identify suitable protein overexpression and purification conditions.

Advances in mass spectrometry (MS)-based methods have enabled the characterization of macromolecules under “native” conditions that preserve non-covalent interactions. For example, offline direct-infusion nanoelectrospray ionization (nESI) has been successfully employed to directly analyze proteins and their interactions in crude cell lysates (Ali & Imperiali, 2005; Ben-Nissan et al., 2018; Catcott, Yan, Qu, Wysocki, & Zhou, 2017; Vimer, Ben-Nissan, & Sharon, 2020a, 2020b). We recently developed an automated online method that utilizes liquid chromatography (LC)-driven separation strategies to couple immobilized metal affinity chromatography (IMAC) with size exclusion-based online buffer exchange (OBE) and native MS (nMS) for screening overexpressed target proteins in cell lysates with minimal sample preparation (see Figure 2 for instrument setup) (Busch, VanAernum, Lai, Gopalan, & Wysocki, 2021; VanAernum et al., 2020). IMAC-OBE-nMS allows for the specific enrichment of poly-His-tagged proteins in cell lysates, simultaneous removal of endogenous host cell proteins, and buffer exchange into non-denaturing, MS-compatible solutions (e.g., ammonium acetate). Solutions containing volatile electrolytes are typically used to transfer intact species from the solution phase to the gas phase for nMS analysis and have been shown, either alone or with charge-reducing reagents, to preserve non-covalent inter- and intramolecular interactions for a variety of macromolecular species and complexes (Heck & van den Heuvel, 2004; Hernández & Robinson, 2007), although there are exceptions (e.g., (Agasid, Sørensen, Umer, Yan, & Robinson, 2021; Xia, DeGrandchamp, & Williams, 2019)).

In this chapter, we describe how to use IMAC-OBE-nMS to streamline assessment of cell lysates for successful protein overexpression and IMAC-based purification to determine whether target proteins (i) are overexpressed in a soluble form, (ii) bind to and elute from an IMAC column, (iii) are present as monomers or higher-order oligomers, and (iv) have the expected mass. The high mass accuracy of nMS provides a decisive advantage over other methods, including SDS-PAGE, when verifying the identity of an overexpressed protein of interest. To highlight the advantages of such an approach, we conducted pilot overexpression trials with four proteins under an assortment of conditions (see Table 1 for more details). In each case, IMAC-OBE-nMS either successfully identified the protein of interest or provided insights about the overexpressed protein that could not be gleaned from SDS-PAGE analysis alone (see Discussion). Thus, both in this chapter and our previous report on the development of this method (Busch et al., 2021), we show that IMAC-OBE-nMS can successfully capture and characterize a variety of poly-His-tagged proteins in cell lysates. While we focus on *E. coli* here as our overexpression host of choice (given its widespread use), IMAC-OBE-nMS may be readily adapted for screening protein overexpression conditions with any suitable overexpression host.

## 2 Materials

Prepare all solutions using analytical-grade reagents and ultrapure water (ddH<sub>2</sub>O) that has been reverse osmosis-filtered or double distilled/deionized to a resistivity of 18.2 MΩ cm at 25°C. Store all prepared reagents and solutions at room temperature (RT; ~22°C) unless otherwise stated below.

## 2.1 Recombinant protein overexpression in *E. coli*

1. Overexpression plasmids encoding poly-His-tagged proteins of interest (*see* Notes 1 and 2)
2. *E. coli* competent cells suitable for the protein of interest
3. Lysogeny broth (LB): combine 10 g tryptone, 5 g yeast extract, and either 0.5, 5, or 10 g NaCl (*see* Note 3) in a final volume of 1 L ddH<sub>2</sub>O, and sterilize by autoclaving for 20 min at 121°C.
4. LB-agar plate: combine 1.5 g agar per 100 mL LB medium, sterilize by autoclaving for 20 min at 121°C, and cool before adding appropriate antibiotics and pouring into Petri dishes (*see* Notes 4 and 5).
5. Equipment and materials for competent cell transformation
6. Temperature-controlled shaker
7. Overexpression media, such as:
  - i. LB medium: *see* (3) above
  - ii. Terrific Broth medium (Cold Spring Harbor Laboratory, 2006; Tartof & Hobbs, 1987): combine 12 g tryptone, 24 g yeast extract, and 4 mL 100% (v/v) glycerol in a final volume of 900 mL ddH<sub>2</sub>O, sterilize by autoclaving for 20 min at 121°C, and allow media to cool to 60°C before adding 100 mL sterile phosphate buffer (0.17 M KH<sub>2</sub>PO<sub>4</sub>, 0.72 M K<sub>2</sub>HPO<sub>4</sub>) (*see* Note 6).
  - iii. Dynamite medium (Taylor et al., 2017): combine 12 g tryptone, 24 g yeast extract, 6.3 mL 100% (v/v) glycerol, 3.8 g KH<sub>2</sub>PO<sub>4</sub>, 12.5 g K<sub>2</sub>HPO<sub>4</sub>, 5 g glucose, and 0.195 g anhydrous MgSO<sub>4</sub> in a final volume of 1 L ddH<sub>2</sub>O, and sterilize by autoclaving for 20 min at 121°C (*see* Note 7).
8. *Optional.* 40% (w/v) glucose: dissolve in ddH<sub>2</sub>O, and filter sterilize using a 0.22 µm membrane filter (*see* Note 8).
9. 1 M isopropyl-β-D-thio-galactoside (IPTG): dissolve in autoclaved ddH<sub>2</sub>O, and store at -20°C until use.
10. Spectrophotometer or cell density meter (e.g., Biowave cell density meter CO8000)

## 2.2 Preparation and SDS-PAGE analysis of cell lysates

1. 1X phosphate-buffered saline (1X PBS): dissolve commercially available tablets in ddH<sub>2</sub>O per manufacturer's instructions, and filter sterilize using a 0.22 µm membrane filter (*see* Notes 9 and 10).
2. Ethylenediaminetetraacetic acid-free protease inhibitor (*see* Note 11)
3. Probe-based sonicator (*see* Notes 12 and 13)

4. Refrigerated microcentrifuge
5. 96-well plate or LC vials (*see* Note 14)
6. Equipment and materials for SDS-PAGE

### 2.3 Tandem immobilized metal affinity-buffer exchange chromatography online with nMS (IMAC-OBE-nMS)

LC-MS solutions, including mobile phases and wash solutions, should be prepared using ddH<sub>2</sub>O (as defined in Section 2.1) and LC-MS-grade reagents (where possible) and then purified through a 0.22 μm membrane filter and degassed prior to use (*see* Notes 15 and 16). The general IMAC-OBE-nMS instrument setup is depicted in Figure 2A while a detailed view of the dual switching valve system is shown in Figure 2B.

1. ddH<sub>2</sub>O, or mobile phase A
2. 1 M ammonium acetate, or mobile phase B: use the highest-grade ammonium acetate (99.99% trace metals basis; e.g., Sigma-Aldrich, cat. no. 431311) where possible
3. 10% (v/v) methanol for the seal, pump, and syringe washes
4. 5 M imidazole
5. Water bath sonicator or vacuum pump for degassing LC-MS solutions (*see* Note 16)
6. Analytical flow UHPLC system equipped with an autosampler, dual pump, and two six-port switching valves (*see* Notes 17 and 18 as well as Figure 2B)
7. ProPac IMAC-10 column (1 × 50 mm, 1.7 μm; Thermo Scientific, cat. no. 063617) or similar LC-compatible IMAC column: the column must be charged with Ni<sup>2+</sup> (*see* Note 19) as per manufacturer's instructions prior to use (*see* Notes 20 and 21).
8. Size-exclusion-based desalting column (OBE column; *see* Note 22)
9. Resistor tubing (75 μm internal diameter, 650 mm length; *see* Note 23)
10. Mass spectrometer optimized for ion transmission and detection over a wide *m/z* range and under native conditions
11. ESI source with a heated electrospray probe and a high-flow metal needle (100 μm inner diameter)
12. Any poly-His-tagged protein for initial calibration of the switching valve divert time and routine quality control

### 2.4 Software-based analysis of IMAC-OBE-nMS data

1. Xcalibur, version 4.1.31.9 (Thermo Scientific)—used to analyze chromatograms and mass spectra collected using the Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific)

2. MS deconvolution software (*see* Note 24)

### 3 Methods

#### 3.1 Recombinant protein overexpression in *E. coli*

Overexpression conditions (e.g., host, medium, inducer concentration, post-induction temperature and/or incubation time) may be varied as desired to maximize soluble target protein overexpression.

1. Transform 1–10 ng overexpression plasmid into ~50  $\mu\text{L}$  *E. coli* competent cells, and allow cells to recover in 150–500  $\mu\text{L}$  LB medium at 37°C with shaking for 1 h before streaking or spreading an aliquot of the transformation sample on a pre-warmed LB-agar plate. Incubate the plate at 37°C overnight (~16 h).
2. Prepare overnight seed cultures using appropriate antibiotics (*see* Note 5), and (optional) supplement with glucose to trigger catabolite repression and inhibition of basal expression of the target protein during overnight growth (*see* Note 8). Use aseptic technique to inoculate the culture with a single, robust colony, and incubate at 37°C with shaking overnight (~16 h).
3. Prepare pilot overexpression cultures of the desired volume (*see* Notes 25 and 26) and/or media using appropriate antibiotics, and inoculate with 1:100 dilutions of the overnight seed cultures. Grow at 37°C with shaking to an  $\text{OD}_{600}$  ~0.6–0.8 (i.e., log phase of *E. coli* cell growth).
4. Divide each culture evenly to generate a pair of “uninduced” and “induced” samples for each test overexpression condition.
5. Induce protein overexpression by adding a sufficient volume of IPTG to each “induced” sample to achieve the desired final concentration, and grow each pair of “uninduced” and “induced” overexpression cultures at specific temperatures with shaking for defined time periods.
6. Harvest final “uninduced” and “induced” samples by centrifugation (*see* Note 26), discard supernatants, and immediately store cell pellets at either –20°C or –80°C (for short- and long-term storage, respectively) until use (*see* Note 27).

#### 3.2 Preparation and analysis of cell lysates

1. Immediately prior to cell resuspension, add protease inhibitor to the lysis buffer for a final concentration of 1X (*see* Note 11).
2. Add one-tenth culture volume of protease inhibitor-supplemented lysis buffer to each cell pellet (e.g., 500  $\mu\text{L}$  to a 5-mL cell pellet; *see* Notes 28 and 29), and pipette up and down to completely resuspend cells.
3. Incubate the resuspended samples on ice for 10 min to allow the cells to swell, thereby facilitating more efficient cell lysis using physical disruption methods.

4. Use a probe-based sonicator (*see* Notes 12, 13, and 29) to individually lyse each resuspended overexpression sample. Keep the sample on ice at all times to avoid overheating.
5. Centrifuge crude cell lysates ( $20,000 \times g$ , 30 min;  $4^{\circ}\text{C}$ ) to remove cellular debris and insoluble components, and transfer the soluble fractions (*i.e.*, supernatants) to a 96-well plate or LC vials for immediate IMAC-OBE-nMS analysis.
6. Perform SDS-PAGE analysis of post-induction samples at sequential stages of processing: whole-cell sample following overexpression and boiled in SDS-PAGE sample loading dye, whole-cell crude lysate post-sonication, and soluble fraction post-centrifugation. Adjust the polyacrylamide gel percentage based on the expected molecular weight of target proteins.

### 3.3 Tandem immobilized metal affinity-buffer exchange chromatography online with nMS (IMAC-OBE-nMS)

For all IMAC-OBE-nMS experiments, the mobile phase should be maintained at a constant flow rate of  $100 \mu\text{L}/\text{min}$  (*see* Note 30). While the ammonium acetate concentration in the mobile phase may require optimization, 200 and 500 mM appear well-suited for a wide range of proteins.

#### 3.3.1 Instrument setup

1. Connect the UHPLC system, IMAC column, switching valves, OBE column, resistor tubing, ESI source with a HESI probe, and mass spectrometer as shown in Figure 2. Purge both pumps after connecting mobile phase A ( $\text{ddH}_2\text{O}$ ) and B (1 M ammonium acetate) to remove any air bubbles from the system. Equilibrate the system with the desired mobile phase concentration (*e.g.*, 200 or 500 mM ammonium acetate) until the spray current on the mass spectrometer is near constant (usually  $\sim 15$  min at  $100 \mu\text{L}/\text{min}$ , depending on the volume of the system). In the meantime, cool the autosampler to  $4^{\circ}\text{C}$  to minimize protein degradation during IMAC-OBE-nMS analysis, and verify that all fittings are tightly connected and leak-free.
2. Use the parameters outlined in Tables 2 and 3 to set up two LC-MS methods in the instrument method manager: “Load” and “Elute”. A switching valve divert time will be added to the “Elute” method after it has been empirically determined as described below in Section 3.3.2 (*see* Note 30). Adjust other settings as needed for detection of target proteins, and include a syringe wash before each sample injection (*e.g.*, wash mode: after draw; wash time, 10 s; wash speed,  $20 \mu\text{L}/\text{s}$ ).

**3.3.2 Calibration of the switching valve divert time**—The switching valve divert time refers to the time at which the valves begin to divert flow to waste following protein elution from the IMAC column (*see* Note 30).

1. Load an aliquot of 5 M imidazole into a single well of a 96-well plate or an LC vial. In another well or vial, load an aliquot of poly-His-tagged protein (~2–4 µg) (see Note 31).
2. Load the 96-well plate or LC vials into the pre-cooled autosampler, and use the “Load” and “Elute” methods from step 2 of Section 3.3.1 to inject, elute (using the 5 M imidazole), buffer exchange, and analyze the sample.
3. The dual switching valve system has two positions: “Load” and “Elute” (see Figure 2B).
  - a. When both switching valves are in the “Load” position, the poly-His-tagged protein is injected from the autosampler onto the IMAC column. Pump L directs the IMAC flow-through to waste and washes the column with ammonium acetate mobile phase, effectively removing any unbound proteins and non-volatile species from the system. Meanwhile, Pump R washes the downstream OBE column with mobile phase to prepare for subsequent OBE.
  - b. Following sample injection, mobile phase is washed over both columns for 6 min.
  - c. When both switching valves are in the “Elute” position and the mass spectrometer is set to begin acquisition upon injection, 3 µL 5 M imidazole is injected from the autosampler. Pump L pushes the IMAC column eluate to the OBE column, which separates the sample from imidazole and non-volatile salts prior to detection by nMS.
  - d. Following a 2-min MS acquisition time, mobile phase is washed over both columns for an additional 8 min.
4. Upon initiation of the “Elute” method, monitor the chromatogram closely. The poly-His-tagged protein will elute from the OBE column first, followed by imidazole. When the imidazole peak begins to elute from the column, change both switching valves back to the “Load” position, stop mass spectral acquisition, and turn off the electrospray voltage to avoid introducing additional imidazole into the mass spectrometer. Note the elution time of the protein and imidazole peaks (see Note 32). The selected switching valve divert time should precede the elution time of the latter imidazole peak.
5. Edit the “Elute” method from step 2 to include the selected divert time; the upper switching valve should now be programmed to change from the “Elute” (1\_2) to “Load” (6\_1) position before imidazole begins to elute from the OBE column to the mass spectrometer.

### 3.3.3 IMAC-OBE-nMS analysis of overexpression samples

1. Before beginning each sequence, perform a test run with a poly-His-tagged protein (or protein standard; see Note 31) to ensure that the system is calibrated properly and running smoothly.



2. To screen for poly-His-tagged proteins of interest, use the updated “Load” and “Elute” methods (see Sections 3.3.1–3.3.2 and Table 3) to analyze “induced” overexpression samples (see Note 33). To analyze unbound proteins (e.g., host cell proteins), use only the “Elute” method.
3. Monitor the system backpressure over the course of successive runs. Increasing backpressure may indicate that protein aggregation is occurring and/or that proteins remain bound to the column. Back pressure that exceeds manufacturer’s recommendations is detrimental to the integrity and performance of the column. Strip and recharge the column with Ni<sup>2+</sup> as per manufacturer’s instructions.

### 3.4 Software-based analysis of IMAC-OBE-nMS data

Provided below is a brief overview of how to use UniDec and MetaUniDec software (Kostelic & Marty, 2020; Marty et al., 2015); see Note 34 for the UniDec deconvolution parameters used in this study to generate waterfall plots and bar charts for Figures 3–6.

#### 3.4.1 Data analysis using UniDec

1. Load a mass spectrum into UniDec by selecting one of the options in the ‘File’ dropdown menu or by manually pasting a spectrum list into the program using Control + g.
2. Select one of the available options under ‘Presets’ in the ‘File’ dropdown menu. Alternatively, the options under the ‘Additional Data Processing Parameters’ menu on the right side can be customized according to the user’s preferences.
3. Under ‘Data Processing’, set the m/z range, and select ‘Process Data’.
4. Under ‘UniDec Parameters’, set wide ranges for ‘Charge Range’ and ‘Mass Range’ to account for all species present in the sample. Then select ‘Run UniDec’ to begin deconvolving the spectrum.
5. After deconvolution is complete, the fitted data (shown in red) should overlay and align well with the raw mass spectrum in the ‘MS Data v. Fit’ tab. To improve data fitting (if needed), adjust the ‘Peak FWHM (Th)’ under ‘Additional Deconvolution Parameters’. A deconvolved zero-charge mass spectrum can be found under the ‘Mass Distribution’ tab.
6. Under ‘Peak Selection and Plotting’, choose an appropriate ‘Peak Detection Range (Da)’ (i.e., the mass range between peaks in Da) and ‘Peak Detection Threshold’ (i.e., the minimum relative intensity for peak identification). Click ‘Peak Detection’.
7. To label identified peaks in the deconvolved zero-charge mass spectrum, select ‘Plot Peaks’.
8. Additional data analysis options may be found under the ‘Tools’, ‘Analysis’, ‘Advanced’, and ‘Experimental’ dropdown menus.
9. Save processed spectra as PDF, EPS, or PNG files using the ‘Save Figure Presets’ option from the ‘File’ dropdown menu.

**3.4.2 Data analysis using MetaUniDec**—MetaUniDec can be used to simultaneously batch-process multiple spectra.

1. Open the HDF5 Import Wizard in the UniDec launcher window.
2. Navigate to the folder containing the files of interest, select each spectrum to be imported, and click ‘Add’. Repeat until all of the spectra to be deconvolved have been added to the HDF5 file list. Designate variables and an averaging time window for each spectrum if desired before selecting ‘Load to HDF5 File’ to save the HDF5 file.
3. Open MetaUniDec and the HDF5 file generated in step 2.
4. Batch process the spectra in a similar manner as described above in **3.4.1** for a single spectrum.
5. When deconvolution is complete, a waterfall plot can be generated (select ‘Waterfall Plot’ in the ‘Experimental’ dropdown menu) to facilitate comparison of the  $m/z$  or deconvolved mass data across multiple conditions in a single plot.
6. Adjust the Data Type, Elevation, and Azimuth, and then middle-click on the waterfall plot to save the image.
7. Save processed spectra as PDF, EPS, or PNG files using the ‘Save Figure Presets’ option from the ‘File’ dropdown menu.

## 4 Notes

1. Overexpression plasmids encoding poly-His-tagged proteins of interest may be generated in-house using standard molecular biology techniques (such as those described in (Sambrook, Fritsch, & Maniatis, 1989)) or purchased commercially. In addition to the requisite poly-His-tag for IMAC-based purification, overexpression plasmids may be engineered to include other fusion tags and sequence modifications to enhance protein production and/or to facilitate subsequent purification. Solubility tags, such as maltose-binding protein (Fox & Waugh, 2003; Kapust & Waugh, 1999; Tropea, Cherry, Nallamsetty, Bignon, & Waugh, 2007) and small ubiquitin-like modifier (SUMO) protein (Butt, Edavettal, Hall, & Mattern, 2005; Guerrero, Ciragan, & Iwai, 2015; Lima & Mossessova, March 22, 2011; Malakhov et al., 2004; Marblestone et al., 2006; Panavas, Sanders, & Butt, 2009), are often used to enhance soluble protein expression; protease cleavage sites inserted between the target protein coding region and poly-His- and solubility tags facilitate removal following protein isolation. Tobacco etch virus (TEV) protease (Tropea et al., 2007) is commonly used for general tag removal while SUMO tags are specifically removed *via* Ulp1 cleavage (Butt et al., 2005; Guerrero et al., 2015; Lima & Mossessova, March 22, 2011; Malakhov et al., 2004; Marblestone et al., 2006; Mossessova & Lima, 2000; Panavas et al., 2009). Additionally, the introduction of preferred sequence motifs at codon positions 3 through 5, or a so-called translational “ramp”, has been shown to increase protein expression

((Moreira et al., 2019; Verma et al., 2019); see also (Zahurancik, Szkoda, Lai, & Gopalan, 2020)).

2. It is critical that the target protein coding region is sequenced in its entirety to ascertain the expected amino acid sequence of the final, translated protein for molecular weight calculations. Depending on the length of the coding region, two sequencing primers that bind to either the RNA polymerase promoter or terminator sequence, may be sufficient; for longer constructs, an internal primer may be necessary to ensure full sequence coverage.
3. There are three formulations of LB, which differ in their NaCl concentration to allow for the selection of osmotic conditions for specific cell strains and culture conditions: LB Miller (Bertani, 1951; Luria & Burrous, 1957) contains 10 g/L medium while LB Lennox (Lennox, 1955) and Luria contain only 5 g/L and 0.5 g/L, respectively. LB Miller is the original and most commonly used formulation while LB Lennox and Luria are generally used for salt-sensitive applications.
4. Given the thermal instability of antibiotics, they should be added to the LB-agar solution when the flask is warm to the touch but cool enough to hold for several seconds (~50°C; (Cold Spring Harbor Laboratory, 2011)). If the solution cools too much, the agar will begin to solidify, making it difficult to pour plates.
5. The choice of antibiotics is dependent on the antibiotic resistance genes present in the expression plasmid(s) and/or host cell strain that will be plated on the LB-agar. Antibiotic stocks should be prepared in the appropriate solvent (e.g., ddH<sub>2</sub>O, methanol) and stored at -20°C to prevent degradation and/or contamination.
6. It has been reported that cells grown in Terrific Broth have higher plasmid yields and reach higher final optical densities (OD<sub>600</sub>) than those grown in LB (Danquah & Forde, 2007; Tartof & Hobbs, 1987).
7. Taylor *et al.* reported that cells grown in their optimized Dynamite medium reached an average final optical density (OD<sub>600</sub>) value/mL culture >30, which was a ~12-fold increase over that of cells grown in LB medium (Taylor et al., 2017).
8. Glucose is often included in growth media to decrease basal target protein expression resulting from the stimulation of RNA polymerase by cyclic AMP receptor protein and cyclic AMP. We have had success adding 1–2% (w/v) glucose to overnight seed cultures while omitting it from overexpression cultures (Novy & Morris, 2001).
9. To make 1X PBS in-house: combine 8 g NaCl, 0.2 g KCl, 1.44 g Na<sub>2</sub>HPO<sub>4</sub>, and 0.24 g KH<sub>2</sub>PO<sub>4</sub> in 800 mL ddH<sub>2</sub>O, adjust the pH to 7.4 with HCl, bring to a final volume of 1 L with ddH<sub>2</sub>O, and sterilize by autoclaving for 20 min at 121°C or passing through a 0.22 µm membrane filter.
10. While 1X PBS was used for all experiments described here and in (Busch et al., 2021), the IMAC-OBE-nMS setup includes (1) a switching valve after the IMAC

column that directs non-volatile salts from the cell lysate to waste and (2) a subsequent online buffer exchange step that removes any remaining non-volatile components (e.g., imidazole). Therefore, most cell lysis buffers, regardless of their compatibility with MS analysis, may be used, though care should be taken to ensure that additives do not (1) bind and elute from the IMAC resin, as they will appear as “contaminating” species in the final mass spectrum; (2) strip the metal from the IMAC resin (e.g., EDTA); and/or (3) significantly alter the viscosity of samples, thereby increasing column and system backpressure. Viscous samples may also be difficult for the small-bore LC autosampler syringe needle to draw and inject. If detergents or glycerol are used as additives, it is best not to exceed twice the critical micelle concentration or 10–20% (v/v), respectively.

11. Protease inhibitor will minimize proteolysis of lysed overexpression samples during subsequent IMAC-OBE-nMS analyses. As most commercially available protease inhibitors are comprised of small molecules and/or non-His-tagged polypeptides (< ~10 kDa), their presence in the lysis buffer should not adversely affect nMS-based analysis.
12. Because probe-based sonicators can only lyse one overexpression sample at a time, commercially available water bath sonicators that can simultaneously lyse up to 16 samples/run may be used to increase throughput.
13. While other methods may be used to lyse cells for IMAC-OBE-nMS analyses, it is important to carefully consider the points discussed in Note 10 before selecting an alternative approach.
14. Depending on the LC system used for IMAC-OBE-nMS analyses, the requisite type of 96-well plate may vary. Some systems, such as the Vanquish Duo UHPLC system, come equipped with a bar code reader that is used to detect empty shelves and to verify rack/well-plate identity location, so they require manufacturer-specific plates (e.g., WebSeal Well Plates, barcoded for Vanquish UHPLC Systems; Thermo Scientific, cat. no. 60180-P206B). LC vials (e.g., National Target DP Vials; Thermo Scientific, cat. no. C4000–11) may be used in place of a 96-well plate if desired.
15. Minimizing trace impurities and particle levels in LC-MS solutions protects the LC column and system components, especially the pumps, and maintains steady flow with minimal backpressure.
16. Although many UHPLC systems come equipped with a built-in mobile phase degasser, any solution used for LC-MS-based experiments should still be manually degassed before use to prevent the introduction of air bubbles into the system, which could cause backpressure spikes that damage the LC column and/or system. Unsealed solutions may be degassed by sonicating for 15 min prior to use.

17. Minimize the length of all solution-bearing lines within the UHPLC system, especially between the column and detector. Using wide-bore and/or extended tubing will decrease ionization efficiency and peak resolution.
18. While a dual six-port switching valve system is not absolutely required for IMAC-OBE-nMS analyses, it ensures that non-volatile salts are sent to waste well in advance of the mass spectrometer. See (Busch et al., 2021) for IMAC-OBE-nMS implementation using only a single switching valve; in that study, imidazole and remaining non-volatile salts were left to drip out of the ESI needle and into a waste tube connected to the ion source-housing drain.
19. Commercially available LC-compatible IMAC columns may be charged with a variety of transition metal ions. While  $\text{Ni}^{2+}$  is most commonly used for IMAC-based purification of poly-His-tagged proteins, it can be substituted with  $\text{Co}^{2+}$ ,  $\text{Cu}^{2+}$ , and/or  $\text{Zn}^{2+}$  (where compatible with the chemistry of the column's solid support) as needed based on the metal ion-binding preference of the target protein.
20. If an LC system is used to charge the IMAC column with metal, lines should first be flushed with ddH<sub>2</sub>O. After the column has been charged, rinse the lines with ddH<sub>2</sub>O and then 50 mM EDTA to chelate any remaining metal ions that could otherwise corrode metal components within the system.
21. Including a guard column or frit filter upstream of the IMAC column will extend the life of the column.
22. In this study, we used a desalting cartridge prototype (2.1 mm × 50 mm, 80 Å pore size; packed with 3 μm silica particles) provided by Thermo Fisher Scientific (Sunnyvale, CA). However, in (Busch et al., 2021), PEEK tubing (0.03 in. internal diameter, 12 cm length) was self-packed with P6 gel medium (Bio-Rad) to generate our own in-house size exclusion-based desalting columns (see (VanAernum et al., 2020) for the column packing protocol).
23. Tubing is placed between the HESI probe and ion source's stainless-steel grounding union as "resistor tubing" (see Figure 2A) to maintain an electrospray current below the maximum limit set by the instrument software; otherwise, a loss of electrospray and decreased sensitivity could result. Changing the length of resistor tubing will also alter the spray voltage needed for efficient sample ionization. As indicated in Note 18, increasing the length and/or decreasing the inner diameter of the resistor tubing generally decreases the requisite spray voltage.
24. Various MS deconvolution software suites are available from commercial and academic sources, including UniDec ((Marty et al., 2015); freely available for download at <https://github.com/michaelmarty/UniDec/releases>), Intact Mass (Protein Metrics Inc., San Carlos, CA), and BioPharma Finder (Thermo Scientific).

25. For all *E. coli* growth cultures, the total culture volume should not exceed 25% of the total culture vessel volume to ensure optimal aeration and cell growth (Sambrook et al., 1989).
26. Overexpression culture volumes may be adjusted and optimized as needed. In our experience, IMAC-OBE-nMS analyses worked well when cell pellets from 5-mL cultures were resuspended and lysed in 400–500  $\mu$ L lysis buffer (*see* Note 29 for more details), though we did not test smaller volumes. For SDS-PAGE, we recommend saving  $\sim$ 1 mL cell culture for harvesting and resuspension in SDS loading dye. Therefore, for parallel IMAC-OBE-nMS and SDS-PAGE analyses, we recommend a minimum culture volume of 6 mL per overexpression condition.
27. While cell pellets can generally be stored at  $-20^{\circ}\text{C}$  or  $-80^{\circ}\text{C}$  for extended periods of time without issue, it is critical to proceed quickly once cells have been thawed and resuspended in order to minimize proteolysis in the cell lysate.
28. While resuspension in one-tenth culture volume of lysis buffer is generally recommended to maximize solubility of the target protein, the solubility of each protein of interest will vary, so appropriate resuspension volumes may need to be empirically determined. Resuspension in a smaller volume of lysis buffer can lead to protein aggregation and/or precipitation in some cases.
29. For probe-based sonication (e.g., with a Vibra-Cell Ultrasonic Processor, cat. no. GEX130), we recommend a minimum cell pellet resuspension volume of 500  $\mu$ L to avoid generating excessive heat that could cause target protein loss. While cycle settings and sonication time can be optimized for specific proteins of interest, we saw successful lysis (with minimal protein loss) for a variety of samples when sonicating for 1 min at 40% amplitude with pulse cycles of 2 s ON and 5 s OFF. Lower resuspension volumes and longer sonication times led to soluble target protein loss.
30. Any changes to the mobile phase flow rate and/or the UHPLC system's plumbing, especially to tubing length or inner diameter, will alter sample arrival time at the mass spectrometer, so the divert time for the switching valves in the "Elute" method (*see* Table 3) must be calibrated anew each time these parameters are changed.
31. If a poly-His-tagged protein is unavailable, a protein standard such as bovine albumin serum (BSA) may be used to calibrate the switching valve divert time and/or perform system quality control tests. To do so, skip the "Load" method and run only the "Elute" method; the other steps remain the same.
32. The amount of upstream LC tubing determines the timing delay between sample elution from the OBE column and arrival time at the mass spectrometer.
33. If overall peak intensities are low and/or the target protein is not readily apparent, increasing the sample injection volume may improve the signal. Depending on the extent of target protein overexpression and the abundance of contaminating species, a 1–10  $\mu$ L sample injection volume is generally sufficient.

34. All spectra presented in Figures 3–6 were either deconvolved manually or using UniDec and MetaUniDec, version 4.4.0 (Kostelic & Marty, 2020; Marty et al., 2015), and the following deconvolution parameters: time start, 0.9; time end, 1.4; preset, high-resolution native;  $m/z$  range, 1000–11000; charge range, 1–25; mass range, 5000–225000; sample mass every (Da), 10; automatic peak width tool, on. To facilitate cross-spectra comparisons of peak and species abundances for the purpose of evaluating overexpression conditions, Data, Peak, and Extract Normalization options were turned off so that peak intensities were adjusted relative to the highest peak across all spectra.
35. Presence of the SID device did not significantly influence the selection of standard tune settings for IMAC-OBE-nMS.
36. With the appropriate adjustments (as illustrated for OBE-nMS in (VanAernum et al., 2020)), other LC-MS systems may be used for IMAC-OBE-nMS analyses.
37. To generate the *E. coli* SixPack strain, Lipinski et al. inserted extra copies of the genes encoding the six least abundant *E. coli* tRNAs (*proL*, *argY*, *ileX*, *argX*, *glyT*, and *leuW*) into one of the ribosomal RNA operons in *E. coli* BL21 (DE3). They reported that protein overexpression in SixPack cells was equal to or up to 20-fold better than in control strains BL21 (DE3) and Rosetta2(DE3)pLysS, as determined by the amount of target protein per cell mass or culture volume (Lipinski et al., 2018).
38. *E. coli* BL21 (DE3) is used for routine T7-based overexpression while enhanced strains like Rosetta allow for more tunable protein expression (e.g., by improving production of proteins containing codons rarely used by *E. coli*).
39. Scan range may be adjusted to encompass the expected masses ( $m/z$  values) of target proteins.
40. Depending on the mass spectrometer and source used, the sheath gas pressure, which is critical for ionization and desolvation during HESI, may be lowered.
41. If necessary, the capillary temperature may be increased up to 320°C to improve sample desolvation.
42. Activation energy, either in the form of in-source trapping (IST) or higher-energy collisional dissociation collision energy (HCD CE, which occurs in the HCD cell rather than in the source region of the instrument, as IST does) can be used to help desolvate and de-adduct species in the sample. (Descriptors like IST and HCD and the locations of collisional activation vary between vendors; the terms used here refer to Thermo Scientific Orbitrap systems.)

## 5 Discussion

For this study, we coupled a Vanquish Duo Ultra-High-Performance LC (UHPLC) system (Thermo Scientific) equipped with a dual switching valve setup to a Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) modified with a customized device for performing surface-induced dissociation (SID; see Note 35) (Harvey,

Author Manuscript

VanAernum, Kostelic, Marty, & Wysocki, 2020; VanAernum et al., 2019). The UHMR was fitted with an Ion Max ion source (Thermo Scientific) that had a heated electrospray ionization (HESI) probe (HESI-II, Thermo Scientific) with a high-flow metal needle (Thermo Scientific, cat. no. OPTON-53010). Figure 2A shows the overall IMAC-OBE-nMS instrument setup (see Note 36) while Figure 2B provides a detailed schematic of the dual switching valve system used to divert unbound proteins and non-volatile salts to waste (see Note 18).

Author Manuscript

Following pilot overexpression trials with four proteins under a variety of conditions (see Table 1) and subsequent SDS-PAGE-based analyses, we used IMAC-OBE-nMS to screen overexpression samples and obtain accurate mass measurements for four target proteins: *Salmonella* AsnA and AsnB (see Figures 3 and 4), *Homo sapiens* GalT (see Figure 5), and *Methanocaldococcus jannaschii* HARP (see Figure 6). We describe below payoffs from IMAC-OBE-nMS analyses in each case.

### ***Salmonella* AsnA and AsnB**

Author Manuscript

AsnA and AsnB are asparagine synthetases that catalyze the conversion of Asp to Asn. When *Salmonella* AsnA and AsnB overexpression samples were analyzed using SDS-PAGE [10% (w/v) polyacrylamide], the apparent AsnA and AsnB bands migrated slower than expected given their molecular weights (see Figures 3A and 4A) providing clear examples of anomalous migration in SDS-PAGE [which is rather pronounced for AsnB, even with 8% (w/v) polyacrylamide]. To verify the identity and mass of the overexpressed proteins observed using SDS-PAGE, the same overexpression samples were lysed and subjected to IMAC-OBE-nMS analyses, which confirmed the presence of AsnA and AsnB monomers with the expected masses (albeit without the N-terminal methionine, a common post-translational modification (Ben-Bassat et al., 1987; Giglione, Boularot, & Meinel, 2004; Wingfield, 2017)) (see Figures 3B–C and 4B–C).

Author Manuscript

IMAC-OBE-nMS analyses proved additionally instructive as we were able to identify AsnA dimers, consistent with its *E. coli* homolog (Nakatsu, Kato, & Oda, 1998). However, when we analyzed *Salmonella* AsnB overexpression samples, no AsnB dimers were observed, contrary to previous reports for its *E. coli* homolog (Larsen et al., 1999). The basis for this difference in observed oligomeric states for AsnA and AsnB is unclear. Finally, IMAC-OBE-nMS results for *Salmonella* AsnB overexpression samples grown at 18°C or 37°C indicate that AsnB was more abundant in the 18°C sample than in the 37°C one (see Figure 4D), an observation that is also borne out by SDS-PAGE (see Figure 4A).

### ***H. sapiens* GalT**

Author Manuscript

In the Leloir pathway, galactose-1-phosphate uridylyl transferase (GALT) converts galactose-1-phosphate and UDP-glucose to glucose-1-phosphate and UDP-galactose, respectively. SDS-PAGE-based analyses of whole-cell *H. sapiens* GalT (Duarte variant N314D; (Lai, Langley, Dembure, Hjelm, & Elsas II, 1999)) overexpression samples indicated that GalT expresses well in both Terrific Broth and Dynamite media (see OE lanes, Figure 5A), though only a small amount was expressed as soluble protein (see SF lanes, Figure 5A). IMAC-OBE-nMS-based analyses of the same samples post-lysis confirmed



these findings and, importantly, the presence of GalT as both a monomer and a dimer, regardless of the overexpression medium used (*see* Figure 5B and 5C). IMAC-OBE-nMS results also revealed that GalT was more abundant in the Dynamite medium sample than in the Terrific Broth one (*see* Figure 5D), suggesting that Dynamite medium may be more optimal for soluble GalT production.

IMAC-OBE-nMS analyses also provided unexpected mass information: while the observed masses for overexpressed GalT are generally consistent with the observed mass for purified GalT (S.M. Lai, P. Thirugnanasambantham, and V. Gopalan, unpublished data), which is present primarily as a dimer as seen in the crystal structure (McCorvie et al., 2016), they are larger than the expected mass by ~336 Da/monomer, suggesting the presence of an unknown modification or ligand. This mass discrepancy may offer a possible explanation for why the recombinant protein expressed in and purified from *E. coli* (P. Thirugnanasambantham and V. Gopalan, unpublished data) has a lower  $k_{cat}$  than the same protein expressed and purified from yeast (Crews, Wilkinson, Wells, Perkins, & Fridovich-Keil, 2000). Additional MS experiments have been planned to identify the source of this 336 Da discrepancy and to better understand how this alteration affects catalytic activity.

### ***M. jannaschii* HARP**

Homologs of Aquifex RNase P (HARP), a tRNA-processing endonuclease, represent an RNA-free, protein-only form present in bacteria and archaea (Daniels, Lai, Chen, & Gopalan, 2019; Nickel et al., 2017). For pilot overexpression trials of *M. jannaschii* SUMO-HARP, the inducer (IPTG) concentration was titrated (e.g., 0.1 mM, 0.5 mM, and 1 mM), and the final induced samples analyzed using SDS-PAGE, which indicated that there were similar amounts of full-length SUMO-HARP across all three conditions (*see* Figure 6A). There also appeared to be a band corresponding to the non-SUMO-tagged HARP in all three gels (*see* Figure 6A), which may indicate processing of SUMO-HARP either during overexpression or post-lysis. While IMAC-OBE-nMS analysis did not identify the presence of any full-length SUMO-HARP, there was, as in the SDS-PAGE gels, some non-SUMO-tagged HARP monomers present (*see* Figure 6B–C). Interestingly, despite similar intensities across all three SDS-PAGE gels, IMAC-OBE-nMS results suggested that non-SUMO-tagged HARP was most abundant in the 0.1 mM IPTG sample (*see* Figure 6D).

Given the presence of SUMO-HARP in the soluble fraction (*see* SF lanes, Figure 6A) and successful purification (V. Sidharthan and V. Gopalan, unpublished data), its absence in the IMAC-OBE-nMS results appears to indicate a potential “false negative”. However, since non-SUMO-tagged HARP is more stable (i.e., less likely to precipitate) than SUMO-HARP at lower NaCl concentrations (V. Sidharthan and V. Gopalan, unpublished data), it is possible that SUMO-HARP may have aggregated on the IMAC or OBE column upon buffer exchange into the ammonium acetate mobile phase, thereby precluding its detection by nMS. Thus, SUMO-HARP highlights one potential limitation of IMAC-OBE-nMS: it is primarily suited for analyzing soluble fractions that are stable in volatile, MS-compatible solutions. However, the incorporation of two switching valves as well as OBE allows us to add moderate amounts of non-volatile solutes to samples that contain proteins that may only

be soluble and stable within a narrow range of conditions, thereby permitting analysis by IMAC-OBE-nMS.

## 6 Summary

Here, we described the protocol for an automated approach to screening cell lysates for successful overexpression and IMAC-based purification of poly-His-tagged proteins of interest using our recently reported IMAC-OBE-nMS method (Busch et al., 2021) and demonstrated its utility by comparing soluble protein yields for four different proteins across a variety of overexpression conditions. The benefits of our approach are two-fold: (1) IMAC-based purification enriches for poly-His-tagged proteins of interest, even with a high endogenous host cell protein background; and (2) nMS can, with high mass accuracy, unambiguously verify the identity and mass of overexpressed target proteins, which is particularly useful in cases of anomalous protein migration in SDS-PAGE or low levels of soluble protein expression. Additionally, nMS can pinpoint mass discrepancies that may arise from amino acid misincorporation (Tsai et al., 1988), proteolysis during overexpression (Rozkov & Enfors, 2004), and/or post-translation modifications, which are too small to observe using SDS-PAGE, and identify higher-order oligomeric states (if any), which generally do not survive SDS-based denaturation. Thus, IMAC-OBE-nMS can streamline assessment of cell lysates for successful protein overexpression and IMAC-based purification and provides additional characterization of overexpressed proteins of interest.

While our work thus far has focused on protein overexpression in *E. coli* and subsequent IMAC-based purification, the workflow described here may be adapted for samples from any overexpression host and other affinity-based purification strategies (e.g., Protein A, Protein G, glutathione, and streptavidin), thereby broadening the range of samples that can be analyzed. Therefore, IMAC-OBE-nMS has the potential to expedite structural biology and biotherapeutics initiatives by streamlining the process of optimizing overexpression and purification conditions for large-scale protein production in heterologous systems.

## Acknowledgments

We thank Drs. Florian Busch and Zachary VanAernum for initial development of the OBE-nMS and IMAC-OBE-nMS methods and for their input on subsequent experiments; Dr. Zsuzsanna Gyorfy (Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary) for providing the *E. coli* SixPack strain; Drs. Rosa Viner and Shane Bechler (Thermo Fisher Scientific, Sunnyvale, CA) for providing the ProPac IMAC-10 column and desalting cartridge prototype; and Dr. Michael Marty (University of Arizona) for his assistance with MetaUniDec-based data analyses. We also gratefully acknowledge research support from the National Institutes of Health (P41 GM128577 to V.H.W.; GM120582 to V.G. and V.H.W.; AI140541 to V.G.; and T32 GM86252 to J.D.L.).

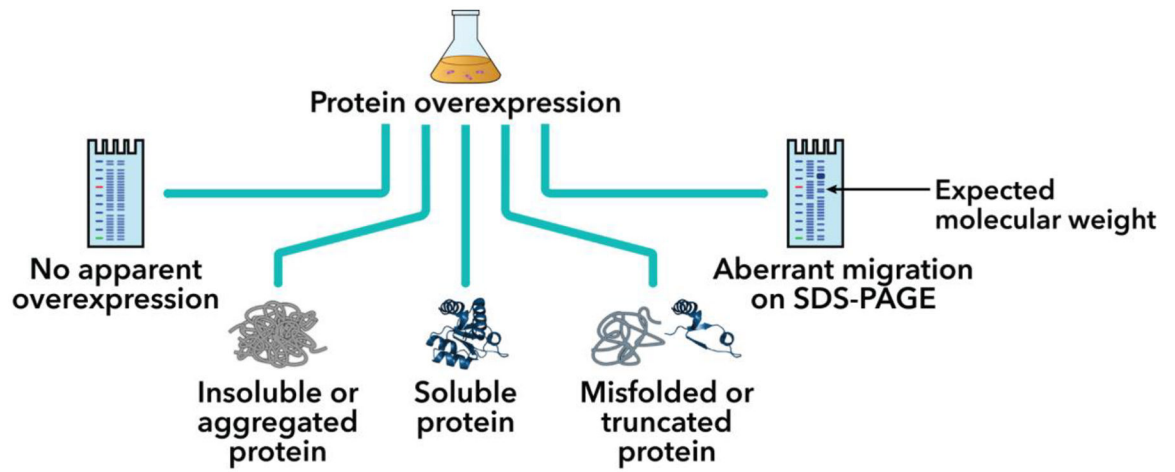
## References

- Agasid MT, Sørensen L, Umer LH, Yan J, & Robinson CV (2021). The effects of sodium ions on ligand binding and conformational states of G protein-coupled receptors—insights from mass spectrometry. *J. Am. Chem. Soc.*, 143(11), 4085–4089. [PubMed: 33711230]
- Ali MH, & Imperiali B (2005). Protein oligomerization: how and why. *Bioorg. Med. Chem.*, 13(17), 5013–5020. [PubMed: 15993087]
- Ben-Bassat A, Bauer K, Chang S-Y, Myambo K, Boosman A, & Chang S (1987). Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.*, 169(2), 751–757. [PubMed: 3027045]

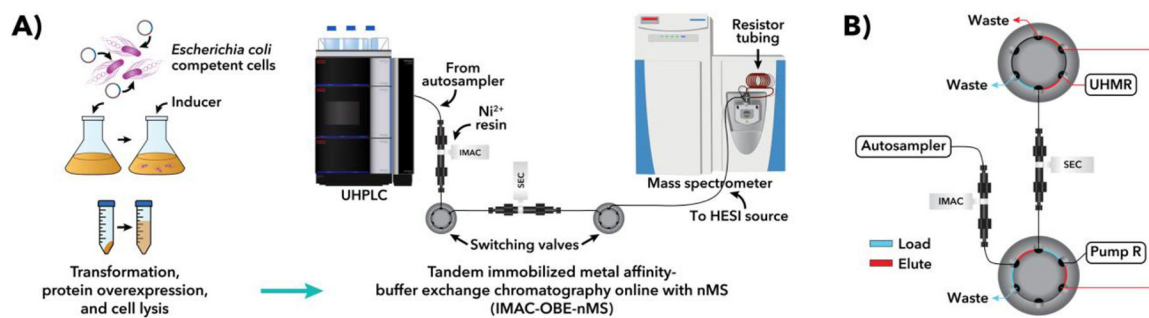
- Ben-Nissan G, Vimer S, Warszawski S, Katz A, Yona M, Unger T, ... Sharon M (2018). Rapid characterization of secreted recombinant proteins by native mass spectrometry. *Commun. Biol*, 1, 213. [PubMed: 30534605]
- Bertani G (1951). Studies on lysogenesis: I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol*, 62(3), 293–300. [PubMed: 14888646]
- Busch F, VanAernum ZL, Lai SM, Gopalan V, & Wysocki VH (2021). Analysis of tagged proteins using tandem affinity-buffer exchange chromatography online with native mass spectrometry. *Biochemistry*. doi:10.1021/acs.biochem.1c00138
- Butt TR, Edavettal SC, Hall JP, & Mattern MR (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expression Purif*, 43(1), 1–9.
- Catcott KC, Yan J, Qu W, Wysocki VH, & Zhou ZS (2017). Identifying unknown enzyme–substrate pairs from the cellular milieu with native mass spectrometry. *ChemBioChem*, 18(7), 613–617. [PubMed: 28140508]
- Cold Spring Harbor Laboratory. (2006). Terrific Broth. *Cold Spring Harb. Protoc* doi:doi:10.1101/pdb.rec8620
- Cold Spring Harbor Laboratory. (2011). Agar plates with LB medium and ampicillin (50 µg/mL). *Cold Spring Harb. Protoc* doi:doi:10.1101/pdb.rec12396
- Crews C, Wilkinson KD, Wells L, Perkins C, & Fridovich-Keil JL (2000). Functional consequence of substitutions at residue 171 in human galactose-1-phosphate uridylyltransferase. *J. Biol. Chem*, 275(30), 22847–22853. [PubMed: 10811638]
- Daniels CJ, Lai LB, Chen T-H, & Gopalan V (2019). Both kinds of RNase P in all domains of life: surprises galore. *RNA*, 25, 286–291. [PubMed: 30578286]
- Danquah MK, & Forde GM (2007). Growth medium selection and its economic impact on plasmid DNA production. *J. Biosci. Bioeng*, 104(6), 490–497. [PubMed: 18215636]
- Fox JD, & Waugh DS (2003). Maltose-binding protein as a solubility enhancer. *Methods Mol. Biol*, 205, 99–117. [PubMed: 12491882]
- Francis DM, & Page R (2010). Strategies to optimize protein expression in *E. coli*. *Curr. Protoc. Protein Sci*, 61(1), 5.24.21–25.24.29.
- Giglione C, Boularot A, & Meinnel T (2004). Protein N-terminal methionine excision. *Cell. Mol. Life Sci*, 61, 1455–1474. [PubMed: 15197470]
- Guan Y, Zhu Q, Huang D, Zhao S, Jan Lo L, & Peng J (2015). An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Scientific Reports*, 5, 13370. [PubMed: 26311515]
- Guerrero F, Ciragan A, & Iwai H (2015). Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expression Purif.*, 116, 42–49.
- Harvey SR, VanAernum ZL, Kostelic MM, Marty MT, & Wysocki VH (2020). Probing the structure of nanodiscs using surface-induced dissociation mass spectrometry. *Chem. Commun. (Camb.)*, 56(100), 15651–15654. [PubMed: 33355562]
- Heck AJR, & van den Heuvel RHH (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev*, 23(5), 368–389. [PubMed: 15264235]
- Hernández H, & Robinson CV (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc*, 2(3), 715–726. [PubMed: 17406634]
- Jungbauer A, & Hahn R (2009). Ion-exchange chromatography. *Methods Enzymol*, 463, 349–371. [PubMed: 19892182]
- Kapust RB, & Waugh DS (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 8, 1668–1674. [PubMed: 10452611]
- Kostelic MM, & Marty MT (2020). Deconvolving native and intact protein mass spectra with UniDec. *ChemRxiv*. doi:10.26434/chemrxiv.13417118.v1
- Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680–685. [PubMed: 5432063]

- Lai K, Langley SD, Dembure PP, Hjelm LN, & Elsas LJ II (1999). The Duarte allele impairs biostability of galactose-1-phosphate uridylyltransferase in human lymphoblasts. *Hum. Mutat*, 11(1), 28–38.
- Larsen TM, Boehlein SK, Schuster SM, Nigel GJ Richards, Thoden JB, Holden HM, & Rayment I (1999). Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product. *Biochemistry*, 38(49), 16146–16157. [PubMed: 10587437]
- Lebendiker M, & Danieli T (2014). Production of prone-to-aggregate proteins. *FEBS Lett*, 588(2), 236–246. [PubMed: 24211444]
- Lennox ES (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*, 1(2), 190–206. [PubMed: 13267987]
- Lima CD, & Mossesso E (March 22, 2011). United States of America Patent No. US7910364B2.
- Lipinszki Z, Vernyik V, Farago N, Sari T, Puskas LG, Blattner FR, ... Gyorfy Z (2018). Enhancing the translational capacity of *E. coli* by resolving the codon bias. *ACS Synth. Biol*, 7(11), 2656–2664. [PubMed: 30351909]
- Luria SE, & Burrous JW (1957). Hybridization between *Escherichia coli* and *Shigella*. *J. Bacteriol*, 74(4), 461–476. [PubMed: 13475269]
- Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, & Butt TR (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics*, 5(1–2), 75–86. [PubMed: 15263846]
- Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, & Butt TR (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci*, 15(1), 182–189. [PubMed: 16322573]
- Marco A. d. (2007). Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. *Nat. Protoc*, 2(10), 2632–2639. [PubMed: 17948006]
- Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, & Robinson CV (2015). Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem*, 87(8), 4370–4376. [PubMed: 25799115]
- McCorvie TJ, Kopec J, Pey AL, Fitzpatrick F, Patel D, Chalk R, ... Yue WW (2016). Molecular basis of classic galactosemia from the structure of human galactose 1-phosphate uridylyltransferase. *Hum. Mol. Genet*, 25(11), 2234–2244. [PubMed: 27005423]
- McCue JT (2009). Theory and use of hydrophobic interaction chromatography in protein purification applications. *Methods Enzymol*, 463, 405–414. [PubMed: 19892185]
- Moreira MH, Barros GC, Requião RD, Rossetto S, Domitrovic T, & Palhano FL (2019). From reporters to endogenous genes: the impact of the first five codons on translation efficiency in *Escherichia coli*. *RNA Biology*, 16(12), 1806–1816. [PubMed: 31470761]
- Mossesso E, & Lima CD (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell*, 5(5), 865–876. [PubMed: 10882122]
- Nakatsu T, Kato H, & Oda J. i. (1998). Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. *Nat. Struct. Biol*, 5(1), 15–19. [PubMed: 9437423]
- Nickel AI, Wäber NB, Gößringer M, Lechner M, Linne U, Toth U, ... Hartmann RK (2017). Minimal and RNA-free RNase P in *Aquifex aeolicus*. *Proc. Natl. Acad. Sci. U.S.A.*, 114(42), 11121–11126. [PubMed: 29073018]
- Novy R, & Morris B (2001). Use of glucose to control basal expression in the pET System. *inNovations*, 13, 8–10.
- Panavas T, Sanders C, & Butt TR (2009). SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods Mol. Biol*, 497, 303–317. [PubMed: 19107426]
- Verheijen R, Salden M, & van Venrooij WJ. (1996). Protein blotting (Venrooij W. J. v. & Maini RN Eds.). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Rosano GL, & Ceccarelli EA (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol*, 5, 172. [PubMed: 24860555]

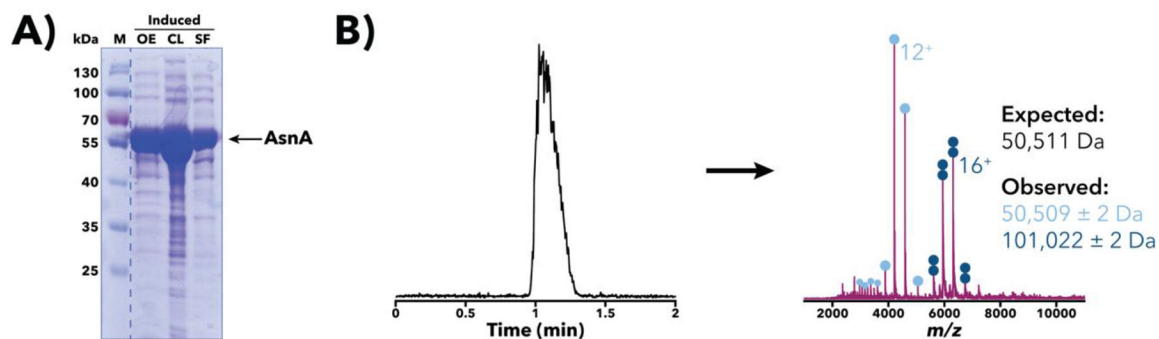
- Rozkov A, & Enfors S-O (2004). Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol*, 89, 163–195. [PubMed: 15217159]
- Sambrook J, Fritsch EF, & Maniatis T (1989). *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, ... Shaw BF (2012). Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein Sci*, 21(8), 1197–1209. [PubMed: 22692797]
- Structural Genomics Consortium. Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, ... SPINE2-Complexes. (2008). Protein production and purification. *Nat. Methods*, 5, 135–146. [PubMed: 18235434]
- Tartof KD, & Hobbs CA (1987). Improved media for growing plasmid and cosmid clones. *Focus*, 9, 12.
- Taylor T, Denson J-P, & Esposito D (2017). Optimizing expression and solubility of proteins in *E. coli* using modified media and induction parameters. *Methods Mol. Biol*, 1586, 65–82. [PubMed: 28470599]
- Tropea JE, Cherry S, Nallamsetty S, Bignon C, & Waugh DS (2007). A generic method for the production of recombinant proteins in *Escherichia coli* using a dual hexahistidine-maltose-binding protein affinity tag. *Methods Mol. Biol*, 363, 1–19. [PubMed: 17272834]
- Tsai LB, Lu HS, Kenney WC, Curless CC, Klein ML, Lai P-H, ... Mann MB (1988). Control of misincorporation of *de novo* synthesized norleucine into recombinant interleukin-2 in *E. coli*. *Biochem. Biophys. Res. Commun*, 156(2), 733–739. [PubMed: 3056402]
- VanAernum ZL, Busch F, Jones BJ, Jia M, Chen Z, Boyken SE, ... Wysocki VH (2020). Rapid online buffer exchange for screening of proteins, protein complexes, and cell lysates by native mass spectrometry. *Nat. Protoc*, 15(3), 1132–1157. [PubMed: 32005983]
- VanAernum ZL, Gilbert JD, Belov ME, Makarov AA, Horning SR, & Wysocki VH (2019). Surface-induced dissociation of noncovalent protein complexes in an extended mass range Orbitrap mass spectrometer. *Anal. Chem*, 91(5), 3611–3618. [PubMed: 30688442]
- Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, ... Djuranovic S (2019). A short translational ramp determines the efficiency of protein synthesis. *Nat. Commun*, 10, 5774. [PubMed: 31852903]
- Vimer S, Ben-Nissan G, & Sharon M (2020a). Direct characterization of overproduced proteins by native mass spectrometry. *Nat. Protoc*, 15(2), 236–265. [PubMed: 31942081]
- Vimer S, Ben-Nissan G, & Sharon M (2020b). Mass spectrometry analysis of intact proteins from crude samples. *Anal. Chem*, 92(19), 12741–12749. [PubMed: 32897050]
- Wingfield P (2017). N-terminal methionine processing. *Curr. Protoc. Protein Sci*, 88, 6.14.11–16.14.13. [PubMed: 28369664]
- Xia Z, DeGrandchamp JB, & Williams ER (2019). Native mass spectrometry beyond ammonium acetate: effects of nonvolatile salts on protein stability and structure. *Analyst*, 144(8), 2565–2573. [PubMed: 30882808]
- Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, ... Li H (2010). Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell*, 39(6), 939–949. [PubMed: 20864039]
- Zahurancik WJ, Szkoda BE, Lai LB, & Gopalan V (2020). Ramping recombinant protein expression in bacteria. *Biochemistry*, 59(23), 2122–2124. [PubMed: 32493006]



**Figure 1.** Potential outcomes of recombinant protein overexpression. [Protein structure: PDB 3NVI (Xue et al., 2010)]



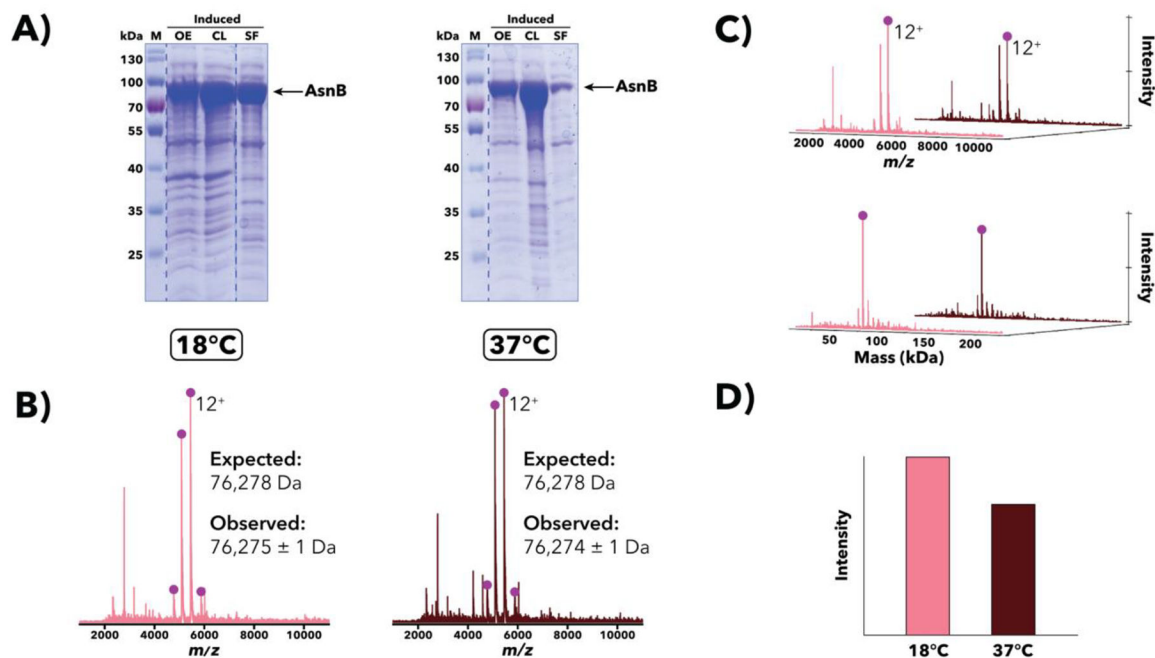
**Figure 2.** Recombinant protein overexpression and characterization using IMAC-OBE-nMS. (A) General workflow and instrument setup. (B) Schematic of dual six-port switching valve system, which has two positions: “Load” (blue) and “Elute” (red). See Section 3.3.2 for more details. Adapted from (Busch et al., 2021) with permission from ACS Publications.



**Figure 3.**

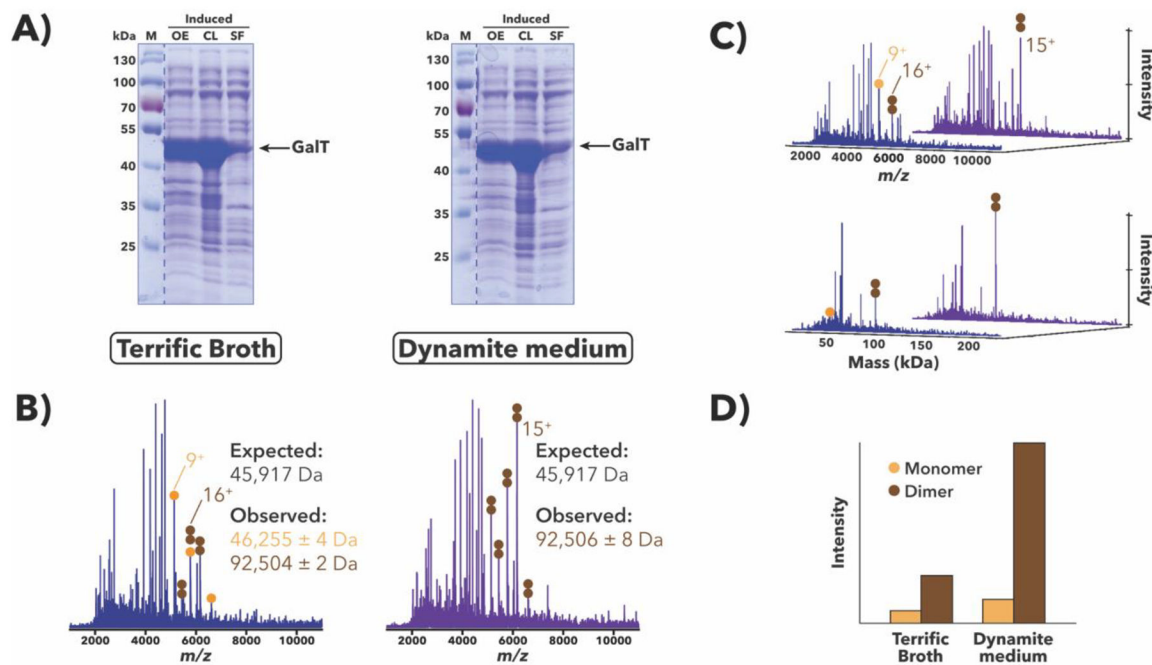
Characterization of a *Salmonella* AsnA overexpression sample. (A) SDS-PAGE [10% (w/v) polyacrylamide] analysis of a post-induction sample at sequential stages of processing. Despite an expected mass of ~50.5 kDa, the apparent AsnA band runs slower than the 55 kDa marker, an example of aberrant protein migration in SDS-PAGE. Gel image has been spliced at the dotted line to only show relevant samples. M, size markers; OE, whole-cell sample following overexpression; CL, whole-cell crude lysate; SF, soluble fraction post-centrifugation. (B) IMAC-OBE-nMS analysis of the soluble fraction using a 500 mM ammonium acetate mobile phase. A representative total ion chromatogram (left) and mass spectrum (right; IST 100 V, HCD CE 5 eV) are shown. The expected mass noted above accounts for loss of the N-terminal methionine, a common post-translational modification (Ben-Bassat et al., 1987; Giglione et al., 2004). Charge state distributions for monomeric and dimeric AsnA species are indicated with light and dark blue circles, respectively, and the main charge state for each species is labeled. The bimodal charge state distribution for the monomer, with its low intensity, higher charge state peaks marked with smaller light blue circles, suggests that there is partially unfolded monomer present in the sample. The y-axis (not shown) represents relative intensity.



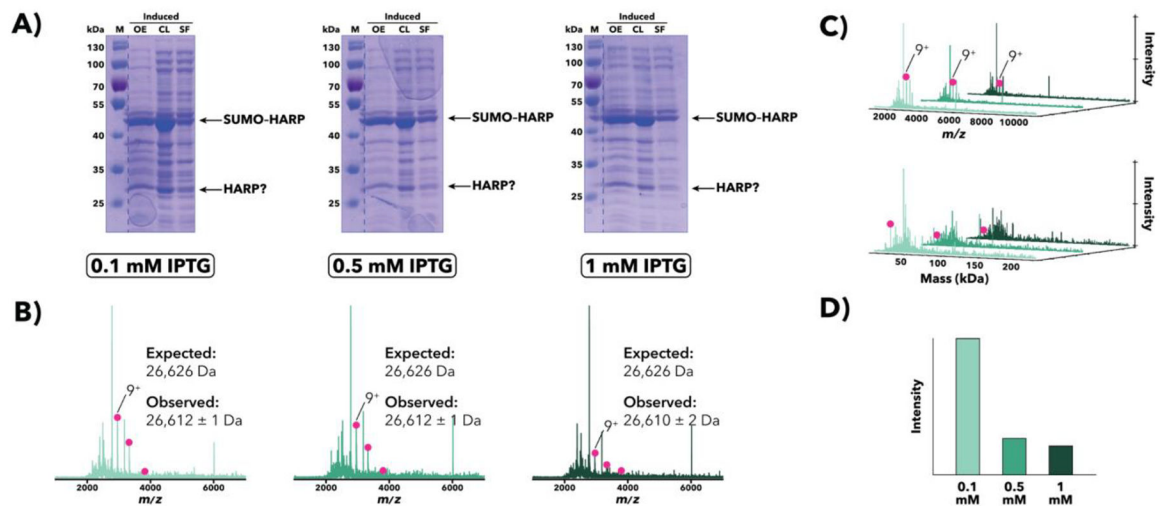


**Figure 4.**

Comparison of *Salmonella* AsnB overexpression samples induced and grown at either 18°C or 37°C. (A) SDS-PAGE [10% (w/v) polyacrylamide] analysis of post-induction samples at sequential stages of processing. Despite an expected mass of ~76.3 kDa, the apparent AsnB band runs much slower than the 70 kDa marker, an example of aberrant protein migration in SDS-PAGE. (B) IMAC-OBE-nMS analysis of soluble fractions using a 500 mM ammonium acetate mobile phase. Mass spectra (IST 100 V, HCD CE 5 eV) for 18°C (pink) and 37°C (dark red) are shown. The expected mass noted above accounts for loss of the N-terminal methionine, a common post-translational modification (Ben-Bassat et al., 1987; Giglione et al., 2004). The charge state distribution for AsnB is indicated, and the main charge state is labeled. The y-axis (not shown) represents relative intensity. (C) Waterfall plots showing  $m/z$  (top) and deconvolved mass (bottom) data. The main charge state (top) and deconvolved mass (bottom) for AsnB is labeled. (D) Relative abundance of AsnB across both samples, as quantified using MetaUniDec software.

**Figure 5.**

Comparison of *H. sapiens* GalT overexpression samples grown in either Terrific Broth (Tartof & Hobbs, 1987) or Dynamite media (Taylor et al., 2017). (A) SDS-PAGE [10% (w/v) polyacrylamide] analysis of post-induction samples at sequential stages of processing. While GalT appears to overexpress well in both media (OE lanes), only a small amount is apparently expressed in a soluble form (SF lanes). (B) IMAC-OBE-nMS analysis of soluble fractions using a 200 mM ammonium acetate mobile phase. Mass spectra (IST 100 V, HCD CE 5 eV) for the Terrific Broth (blue) and Dynamite medium (purple) samples are shown. While there are other proteins present in both samples, charge state distributions for monomeric and dimeric GalT species are indicated with orange and brown circles, respectively, and the main charge state for each species is labeled. The y-axis (not shown) represents relative intensity. The observed masses are larger than the expected mass (by ~336 Da/monomer) (see Discussion). (C) Waterfall plots showing  $m/z$  (top) and deconvolved mass (bottom) data, which were generated using MetaUniDec software (Kostelic & Marty, 2020; Marty et al., 2015). The main charge state (top) and deconvolved mass (bottom) for each species is labeled. (D) Relative abundance of GalT species across both samples, as quantified using MetaUniDec software.



**Figure 6.**

Comparison of *M. jannaschii* SUMO-HARP overexpression samples induced with either 0.1, 0.5, or 1 mM IPTG. (A) SDS-PAGE [10% (w/v) polyacrylamide] analysis of post-induction samples at sequential stages of processing. Apparent bands corresponding to full-length SUMO-HARP as well as non-SUMO-tagged HARP are indicated. (B) IMAC-OBE-nMS analysis of soluble fractions using a 500 mM ammonium acetate mobile phase. Mass spectra (IST 100 V, HCD CE 5 eV) for 0.1 mM (light green), 0.5 mM (middle green), and 1 mM (dark green) IPTG are shown. While no apparent full-length SUMO-HARP is present in the spectrum, there is some non-SUMO-tagged HARP (pink circles). The charge state distribution for non-SUMO-tagged HARP is indicated, and the main charge state is labeled. The y-axis (not shown) represents relative intensity. (C) Waterfall plots showing  $m/z$  (top) and deconvolved mass (bottom) data. The main charge state (top) and deconvolved mass (bottom) for non-SUMO-tagged HARP is labeled. (D) Relative abundance of non-SUMO-tagged HARP across all samples, as quantified using MetaUniDec software.

**Table 1.**

Overexpression conditions for recombinant poly-His-tagged proteins analyzed using IMAC-OBE-nMS.

Protein	Overexpression conditions				
	<i>E. coli</i> cell strain	Medium	[IPTG]	Post-induction temperature	Post-induction time
SUMO- <i>Salmonella</i> AsnA	SixPack ( <i>see Note 37</i> ; (Lipinski et al., 2018))	Lysogeny broth (LB) Miller ( <i>see Note 3</i> )	0.5 mM	37°C	4 h
<i>H. sapiens</i> GalT (Duarte variant N314D) (Lai et al., 1999)		Terrific Broth (Tartof & Hobbs, 1987)	0.1 mM	18°C	16 h
		Dynamite medium (Taylor et al., 2017)			
SUMO- <i>M. jannaschii</i> HARP	BL21 (DE3) ( <i>see Note 38</i> )	LB Miller	0.1 mM	37°C	3 h
			0.5 mM		
			1 mM		
5SUMO- <i>Salmonella</i> AsnB			0.5 mM	18°C	16 h
		37°C	3 h		

**Table 2.**

Tune settings used for the Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer. Adapted from (Busch et al., 2021) with permission from ACS Publications.

Setting	Value
Scan range ( $m/z$ ) (see Note 39)	1,000–16,000
Resolution (at 400 $m/z$ )	12,500
Microscans	5
Max inject time (ms)	200
Sheath gas (psi) (see Note 40)	50
Aux gas (psi)	0
Sweep gas (psi)	0
Spray voltage (kV)	3.75
Capillary temperature ( $^{\circ}$ C) (see Note 41)	275
S-lens RF level (V)	200
In-source dissociation (V)	0
In-source trapping (V) (see Note 42)	Variable
HCD CE (V) (see Note 42)	Variable
Source DC offset (V)	21
Injection flatapole DC (V)	5
Inter flatapole lens(V)	4
Bent flatapole DC (V)	2
Trapping gas pressure	4

**Table 3.**

IMAC-OBE-nMS method. Total run time for each sample is 16 min. The “Elute” method is divided into two phases: during phase 1, the upper switching valve is in the 1\_2 position to direct protein eluate from the IMAC column to the OBE column; later, during phase 2, the upper switching valve changes to the 6\_1 position to divert imidazole (and any other remaining non-volatile salts) to waste. The lower switching valve remains in the 1\_2 position for the duration of the “Elute” method. After 2 min, mass spectral acquisition ends, the source gas and ESI are turned off, and both columns are washed with the ammonium acetate mobile phase for the remaining 8 min of the “Elute” method. Adapted from (Busch et al., 2021) with permission from ACS Publications.

Method	Time	Autosampler	LC pump	Lower switching valve	Upper switching valve	MS
Load	6 min	1–10 $\mu$ L injection of cell lysate	100 $\mu$ L/min	Position 6_1 (to waste)	Position 6_1 (to waste)	OFF (no source gas, no ESI voltage)
Elute	10 min	3 $\mu$ L injection of 5 M imidazole, pH 7.5	100 $\mu$ L/min	(1–2) Position 1_2 (to OBE column)	(1) Position 1_2 (to MS); (2) Position 1–6 (to waste)	ON (source gas, ESI voltage)