Screening tagged proteins using tandem affinity-buffer exchange chromatography online with native mass spectrometry

Florian Busch[1-3,*], Zachary L. VanAernum[1,2,*], Stella M. Lai[1,2,*], Venkat Gopalan[1], and Vicki H. Wysocki[1-3,#]

[*]Contributed equally
[#]Corresponding author: Vicki H. Wysocki, wysocki.11@osu.edu

[1]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210 USA
[2]Resource for Native Mass Spectrometry-Guided Structural Biology, The Ohio State University, Columbus, OH 43210 USA
[3]Campus Chemical Instrument Center, Mass Spectrometry and Proteomics, The Ohio State University, Columbus, OH 43210 USA

Abstract

Protein overexpression and purification are critical for *in vitro* structure-function characterization studies. However, some proteins are difficult to express robustly in heterologous systems due to host-related (e.g., codon usage, translation rate) and/or protein-specific (e.g., toxicity, aggregation) challenges. Therefore, it is often necessary to screen multiple overexpression and purification conditions to maximize the yield of functional protein, particularly for resource-heavy downstream applications (e.g., biocatalysts, tertiary structure determination, biotherapeutics). Here, we describe an automatable liquid chromatography–mass spectrometry-based method for rapid, direct analysis of target proteins in cell lysates. This online approach is facilitated by coupling immobilized metal affinity chromatography (IMAC), which leverages engineered poly-histidine tags in proteins of interest, with size exclusion-based buffer exchange (OBE) and native mass spectrometry (nMS). The use of IMAC-OBE-nMS to optimize conditions for large-scale protein production should expedite structural biology and biotherapeutic initiatives.

Introduction

Protein purification is a critical prerequisite for biochemical and biophysical characterization of proteins and protein-ligand complexes *in vitro*. Therefore, well-characterized, cost-effective overexpression systems like *Escherichia coli* and *Saccharomyces cerevisiae* are commonly used to generate large amounts of recombinant protein. The method involves overexpression of the protein of interest in an appropriate host and subsequent chromatography-based purification to obtain the pure protein of interest (**Figure 1A and 1B**). Typically, the gene encoding the target protein is cloned into an expression vector that is, in turn, transformed into host cells. The target gene is usually placed under the control of an inducible promoter (1-3), which allows for regulated overexpression of the protein of interest. A chemical inducer is typically added

during the log phase of cell growth to induce target gene expression, and the culture is grown at a specific temperature for a defined time period post-induction to promote maximal protein synthesis. Recombinant protein production can usually be scaled up relatively easily and cost effectively by simply increasing the cell culture volume. Following confirmation of successful overexpression, the target protein is isolated from host cell proteins using column-based chromatographic purification methods that exploit intrinsic properties of the protein of interest [e.g., isoelectric point (pI), hydrophobicity] (4,5). Alternatively, genetically engineered affinity tags can be fused to either terminus of the protein to enable affinity-based purification (3).
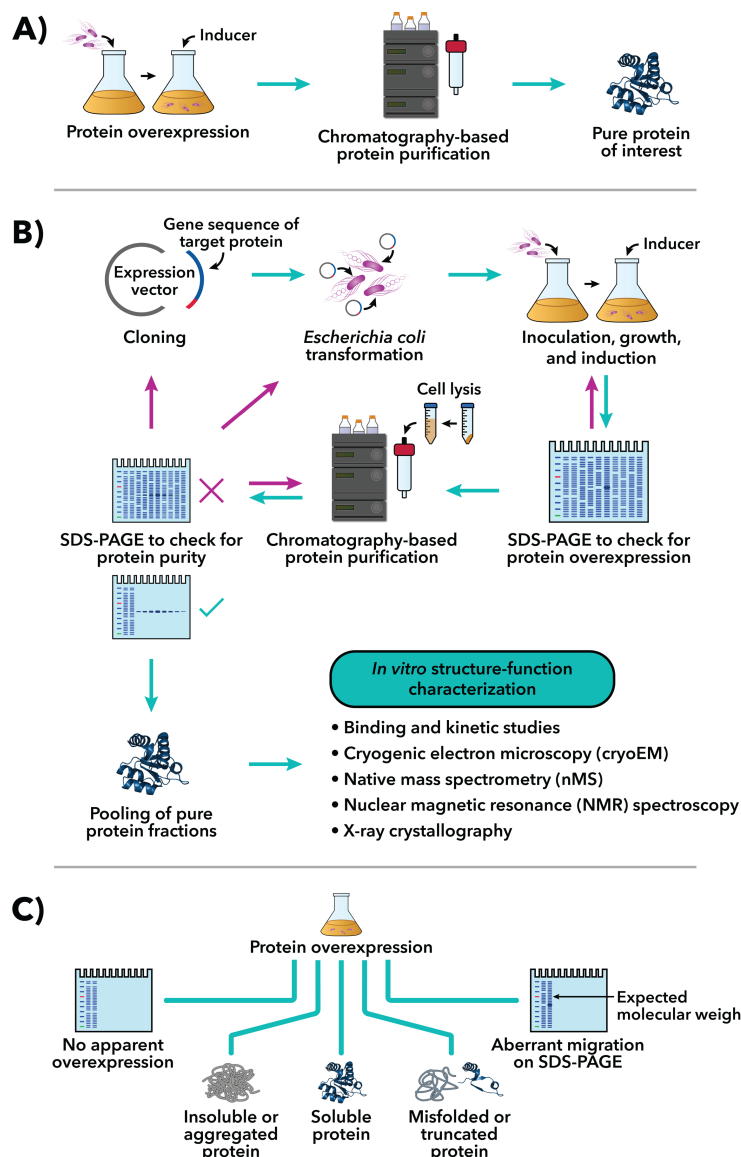

Figure 1. (A) Simplified method for preparing recombinant proteins. (B) Workflow for overexpression and purification of recombinant proteins of interest. Teal-colored arrows indicate the expected experimental workflow while magenta-colored arrows denote steps that may require troubleshooting and optimization. (C) Potential outcomes and pitfalls of recombinant protein overexpression. [PDB: 3NVI (6)]

Unfortunately, obtaining large amounts of soluble, correctly folded, and active recombinant protein is frequently not straightforward. Some of the major challenges that undermine expression of certain recombinant proteins arise from issues ranging from host-cell translation (e.g., codon usage differences, translation rate) to target protein attributes (e.g., poor solubility, propensity to aggregate, toxicity (7-9)). Additionally, specific incompatibilities between the protein of interest and the selected heterologous host system may culminate in misfolding or premature translation termination, generating truncated proteins. Therefore, different expression constructs are often evaluated in small-scale pilot tests to assess total overexpression and solubility. To this end, cells are harvested, lysed, and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (10). The absence or presence of the target protein, as inferred from a band of the expected molecular weight, is then used to make a rapid and qualitative assessment of protein overexpression.

SDS-PAGE, which is inexpensive and routinely used, has been a cornerstone of modern biochemistry. However, its drawbacks often cause researchers to either prematurely discard overexpression constructs as inviable or proceed to downstream processing steps with false optimism. First, determining whether the target protein has even been successfully overexpressed in its full-length form is complicated by the frequent aberrant migration of proteins on SDS-PAGE, which is caused by compositional eccentricities (e.g., above-average presence of charged residues) of the protein of interest (11-13). Unfortunately, this simple limitation is further compounded by proteolysis of misfolded foreign proteins in different heterologous hosts. Second, the target protein may be obscured by co-migrating host proteins and thus scored as absent despite being present at low levels. Consequently, opportunities to isolate even modest amounts of the target protein may be lost. Lastly, because SDS-PAGE denatures proteins, the presence of a target protein band does not preclude the possibility that the protein is part of a soluble aggregate in the crude lysate. Therefore, confirmation of overexpression and solubility of the target protein by SDS-PAGE does not reliably forecast subsequent successful purification. These shortcomings of SDS-PAGE, coupled with the challenges of heterologous gene expression, often necessitate laborious and time-consuming rounds of screening, troubleshooting, and optimization to maximize the yield of pure recombinant protein (**Figure 1B**). To overcome these limitations, we sought to establish an automatable liquid chromatography (LC)-based method that would allow for screening of protein overexpression and purification strategies.

Specifically, we have coupled immobilized metal affinity chromatography (IMAC) (14), which leverages engineered poly-histidine tags (poly-His-tags) in proteins of interest, with our recently developed method for performing size exclusion-based buffer exchange online with native mass spectrometry (OBE-nMS) (15). IMAC-OBE-nMS allows for rapid, direct analysis of target proteins in cell lysates without prior purification and overcomes potential ambiguities with SDS-PAGE by using non-denaturing methods to measure the accurate mass and determine the oligomeric state of proteins of interest. This method has the potential to expedite structure- and biotherapeutic-focused initiatives by facilitating the rapid screening of multiple samples to assess successful overexpression and IMAC-based purification. Here, we selected five proteins to illustrate the utility of the IMAC-OBE-nMS method for (i) selectively capturing and enriching

for overexpressed His$_6$-tagged target proteins, (ii) determining the oligomeric state of proteins of interest, and (iii) assessing the ability of solubility tags to prevent protein aggregation and precipitation.

Materials and Methods

Protein overexpression and lysis

Purified His$_6$-tagged *Salmonella* YidA and overexpression plasmids encoding His$_6$-tagged *Methanococcus maripaludis* (*Mma*) L7Ae, *Salmonella* FrlB, *Salmonella* FraR, and *Salmonella* FraR fused to maltose-binding protein (MBP-FraR) were used in this study. The motivation for this selection of plasmids was two-fold: (i) the in-house availability of purified protein and clones (Gopalan, unpublished data) and (ii) the desire to exploit our method to advance ongoing projects. Overexpression conditions and purification protocols for these proteins have not been previously reported and will be described elsewhere. Briefly, *E. coli* BL21 (DE3) or Rosetta (DE3) strains were transformed with plasmids encoding the His$_6$-tagged proteins mentioned above. Overnight seed cultures were used to inoculate 5 mL lysogeny broth (LB) media containing the appropriate antibiotics, which were subsequently grown at 37°C with shaking to an OD$_{600}$ ~ 0.6. Protein overexpression was induced with appropriate amounts of isopropyl-β-D-thio-galactoside (IPTG). Post-induction, these cultures were allowed to grow further under previously optimized conditions (Gopalan, unpublished data). Cells were then harvested by centrifugation (20,000 x *g*; 2 min; ~22°C) and lysed by sonication in 400 μL 1X phosphate-buffered saline (PBS) supplemented with Halt protease inhibitor (Thermo Scientific). Following centrifugation (20,000 x *g*; 30 min; 4°C), supernatants were transferred to 96-well plates for MS analysis.
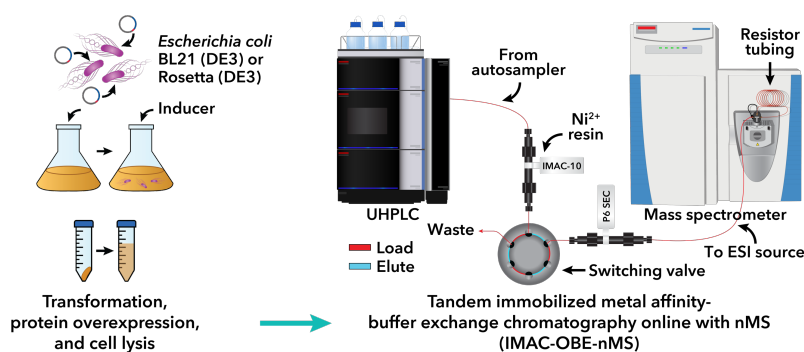

Figure 2. Analysis of recombinant protein overexpression and IMAC-based purification conditions using IMAC-OBE-nMS. The switching valve (shown in gray) has two positions: "Load" (red) and "Elute" (blue). When the autosampler injects clarified cell lysate onto the IMAC-10 column, the switching valve is set to "Load", which sends the flow-through to waste. However, when the autosampler injects imidazole onto the column to elute the His-tagged protein, the switching valve is set to the "Elute" position, and the eluate is directed to the P6 SEC column for buffer exchange prior to MS analysis.

LC-based IMAC-OBE-nMS

For all experiments, a mobile phase of 200 mM ammonium acetate, pH 7.5 (pH adjusted using ammonium hydroxide), was maintained at a flow rate of 100 µL/min. A diagram of the instrument setup described below is shown in **Figure 2**. A Vanquish Duo Ultra-High-Performance LC (UHPLC) system (Thermo Scientific) equipped with a dual pump and autosampler was used to load crude cell lysates onto a Ni^{2+}-charged IMAC column (ProPac IMAC-10, 1 x 50 mm, 1.7 µm Thermo Scientific, 063617). A six-port switching valve downstream of the IMAC column was used to direct flow: unbound species were directed to waste while bound proteins were eluted using 3 µL 5 M imidazole, pH 7.5, and subsequently buffer exchanged into 200 mM ammonium acetate and separated from imidazole using a size exclusion chromatographic (SEC) column made of P6 gel medium (Bio-Rad) that was self-packed in PEEK tubing (0.03 in. internal diameter, 12 cm length) (15). Desalted samples were then analyzed using a Q Exactive Ultra-High Mass Range (UHMR) Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific) that was modified with a customized device for performing surface-induced dissociation (16,17) and equipped with an Ion Max ion source (Thermo Scientific) and a heated electrospray ionization (HESI-II) probe (Thermo Scientific). The instrument was tuned to optimize ion transmission over a wide *m/z* range; see **Table 1** for specific ESI and MS tune settings. Because nMS analysis was carried out in a mobile phase of moderately high ionic strength (200 mM), PEEK tubing (0.005 in. internal diameter, 10 ft. length) was placed between the HESI-II probe and the Ion Max source's stainless-steel grounding union as "resistor tubing" to ensure that the electrospray current did not exceed the maximum limit set by the instrument software. Exceeding this threshold would cause a loss of electrospray and decrease sensitivity (15).

Setting	Value
Scan range (*m/z*)	1,000-14,000
Resolution (at 400 *m/z*)	12,500
Microscans	5
Max inject (ms)	200
Sheath gas (psi)	15
Aux gas (psi)	0
Sweep gas (psi)	0
Spray voltage (kV)	4.0
Capillary temperature (°C)	320
S-lens RF level (V)	200
In-source dissociation (V)	10
In-source trapping (V)	Variable
Source DC offset (V)	21
Injection flatapole DC (V)	5
Inter flatapole lens (V)	4
Bent flatapole DC (V)	2
Trapping gas pressure	3

Table 1. Tune settings for the Q Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer.

Data analysis

All data analyses were conducted using the Intact Mass software (version w2.15-294-gba5daea4b; Protein Metrics Inc., San Carlos, CA) (18) and the following deconvolution parameters: minimum difference between mass peaks, 10; charge vector spacing, 1; baseline radius (m/z), 15; smoothing sigma (m/z), 0.02; spacing (m/z): 0.04; mass smoothing sigma, 3; mass spacing, 0.5; iteration max, 10.

Results

Instrument setup for tandem affinity-buffer exchange chromatography online with nMS

For implementation of IMAC-OBE-nMS, an analytical flow LC system equipped with an autosampler is coupled to a mass spectrometer via a multi-port switching valve (**Figure 2**). Located between the LC autosampler and the mass spectrometer is a Ni^{2+}-charged IMAC column; a six-port switching valve, which can be used to direct flow as described below; and an SEC column for removal of non-volatile components and buffer exchange into ammonium acetate for nMS analysis.

The switching valve has two positions: "Load" and "Elute" (**Figure 2**). When the autosampler injects crude cell lysate onto the IMAC column, the switching valve is set to the "Load" position to direct the flow-through toward waste, allowing for the removal of unbound proteins and other non-volatile species from the sample. In contrast, when the autosampler injects imidazole onto the column to elute the tagged protein, the switching valve is set to the "Elute" position. Subsequently, the eluate is directed to the size exclusion column, which separates the sample from imidazole and non-volatile salts prior to detection by MS. To analyze unbound proteins (e.g., host cell proteins), the switching valve can remain in the "Elute" position during the load step. However, if cellular debris is not carefully removed after cell lysis or if protein aggregation occurs, the performance of the size exclusion column may be hampered.

A mobile phase of 200 mM ammonium acetate, pH 7.5, was selected for two primary reasons: (i) it is a volatile electrolyte, which is required to transfer samples from the solution phase to the gas phase without extensive adduction for MS analysis; and (ii) ammonium acetate is generally suitable and commonly used for preserving non-covalent inter- and intramolecular interactions for a wide range of macromolecular samples and complexes (19,20). In all cases, the pH of the mobile phase should be neutral to slightly alkaline (pH 7-8) to ensure that the poly-histidine tag on the protein of interest binds the Ni^{2+}-charged resin.

General workflow for IMAC-OBE-nMS experiments

Following overexpression of poly-His-tagged proteins of interest, cells are resuspended in an appropriate buffer and then lysed. While we used 1X PBS for all experiments described here, inclusion of a switching valve that can direct non-volatile salts to waste and an online buffer exchange step should permit the use of any buffer, regardless of its compatibility with MS

Step	Time	Autosampler	LC pump	Switching valve	MS
Load	4 min	1 μL injection of crude cell lysate	100 μL/min	Position 1-6 (to waste)	Tune-file2 (no source gas, no ESI voltage)
Wash	4 min	5 μL injection of mobile phase	100 μL/min	Position 1-2 (to MS)	Tune-file2 (no source gas, no ESI voltage)
Elute	4 min	3 μL injection of 5 M imidazole, pH 7.5	100 μL/min	Position 1-2 (to MS)	Tune-file1 (source gas, ESI voltage) during protein elution; switch to Tune-file2 before imidazole elutes after ~1 min (no source gas, no ESI voltage)

Table 2. Method used for OBE-nMS and IMAC-OBE-nMS.

analysis. With the switching valve set to the "Load" position, the autosampler is used to inject the sample onto the Ni^{2+}-charged IMAC column. Poly-His-tagged proteins remain bound to the column while unbound species are directed to waste. A subsequent injection of mobile phase minimizes carryover from the previous injection. When the valve is then switched to the "Elute" position, imidazole is injected to elute bound species from the IMAC column. The eluate is directed toward the SEC column, where proteins are separated from imidazole and concomitantly buffer exchanged into ammonium acetate prior to nMS analysis.

Mass spectral acquisition is timed to permit ionization of proteins, but not imidazole, upon elution from the SEC column – due to the large size discrepancy, imidazole elutes later than proteins during size exclusion-based separation. To this end, the MS method is comprised of two separate tune files (**Table 2**): the first file turns on the source gas and applies electrospray voltage as proteins elute from the SEC column while the second one turns off both before imidazole begins to elute at ~1.5 min. This strategy allows for the efficient detection of proteins while imidazole and any remaining non-volatile salts drip out of the ESI needle and into a waste tube connected to the ion source housing drain.

To evaluate the contribution of the added IMAC purification step, we also performed OBE-nMS-only experiments on the same samples subjected to IMAC-OBE-nMS analysis. For OBE-nMS, the IMAC column is removed from the instrument setup described above, and the autosampler is used to direct samples onto the SEC column via the switching valve. As with IMAC-OBE-nMS, two separate tune files (**Table 2**) are used to ensure that non-volatile salts are not electrosprayed and the instrument is kept clean for subsequent experiments.
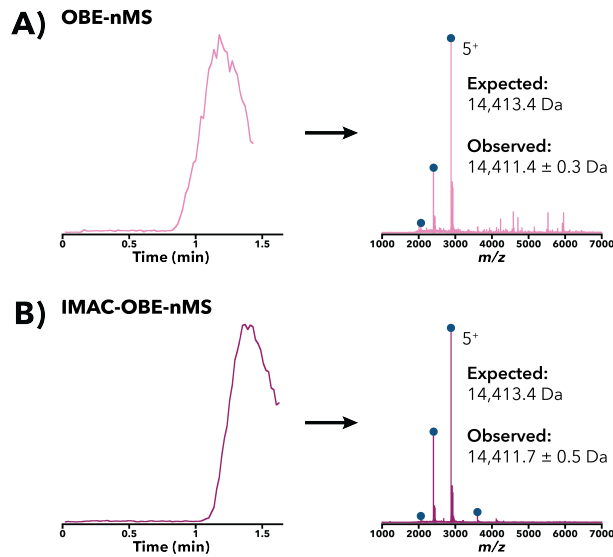
Figure 3. Total ion chromatograms and mass spectra of *Mma* L7Ae analyzed using (A) OBE-nMS or (B) IMAC-OBE-nMS. The expected mass indicated above accounts for loss of the N-terminal methionine, which is a common post-translational modification (21-23). Charge state distributions for L7Ae are indicated with blue circles, and the main charge state is labeled. In all plots, the y-axis (not depicted) represents relative intensity.

Validating IMAC-OBE-nMS as a tool for screening protein overexpression and purification conditions

We used several proteins to highlight the ability of IMAC-OBE-nMS to selectively capture and enrich for overexpressed poly-His-tagged proteins of interest in crude lysate. First, His$_6$-tagged *Methanococcus maripaludis* (*Mma*) L7Ae, a soluble protein that ionizes well under a variety of conditions, was used as proof of concept (**Figure 3**). When a cell lysate containing overexpressed L7Ae was subjected to OBE-nMS (**Figure 3A**), L7Ae was the dominant species, though the spectrum revealed the presence of other, less abundant *E. coli* host cell proteins. In contrast, combining the IMAC purification step and OBE-nMS removed almost all of the contaminating species, demonstrating that IMAC-OBE-nMS effectively enriches for His$_6$-tagged proteins (**Figure 3B**). However, due to the relatively high ionization efficiency and overexpression level of *Mma* L7Ae, the benefit of our new method is not particularly striking. Therefore, to better establish the merits of IMAC-OBE-nMS, we intentionally spiked a pure His$_6$-tagged protein into crude lysate and compared the results of OBE-nMS and IMAC-OBE-nMS (**Figure 4**).

For this next set of experiments, we used His$_6$-tagged *Salmonella* YidA as the target protein because its ionization efficiency is lower than that of *Mma* L7Ae, so it may be more representative of an average protein of interest. A fixed amount of purified YidA (0.033 mg/mL), either alone or in combination with varying amounts of *E. coli* cell lysate (0.16 or 0.31 mg/mL), was subjected to OBE-nMS and IMAC-OBE-nMS. Samples containing higher crude lysate concentrations were intended to mimic instances in which host cell protein levels exceed
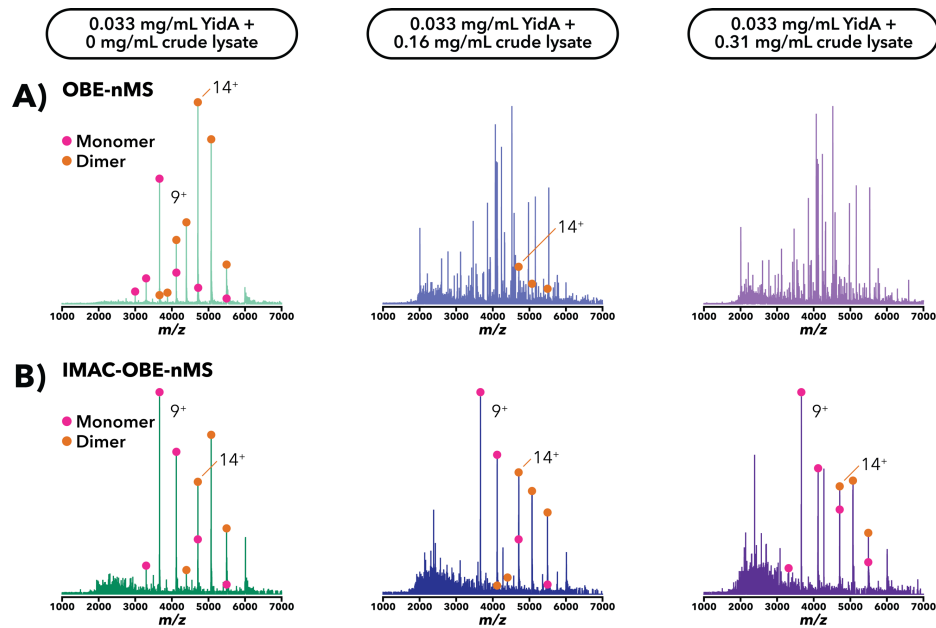
Figure 4. Mass spectra of purified *Salmonella* YidA, either alone or diluted as indicated with *E. coli* crude lysate, and analyzed using (A) OBE-nMS or (B) IMAC-OBE-nMS. For the monomeric and dimeric YidA species shown here, the observed masses (averaged across all six conditions) are 32,975 ± 1.2 Da and 65,951 ± 4.2 Da. Charge state distributions for monomeric and dimeric species are indicated with pink and orange circles, respectively, and the main charge state for each species is labeled. In all plots, the y-axis (not shown) represents relative intensity.

that of the target protein, and SDS-PAGE cannot be used to unambiguously assess the extent of protein expression. Consistent with the expectation that both methods should be gentle enough to preserve and screen for the oligomeric state of overexpressed proteins of interest (15), a mixture of monomeric and dimeric YidA species was observed when YidA alone was analyzed using either method. However, addition of crude lysate to the sample clearly distinguished the two methods: whereas the YidA signal for OBE-nMS was lost with increasing amounts of cell lysate (**Figure 4A**), YidA was still detectable at a 1:10 YidA:crude lysate dilution with IMAC-OBE-nMS (**Figure 4B**). The loss of YidA signal with OBE-nMS is characteristic of increased charge competition and ion suppression from the higher abundance of host cell proteins (24). Thus, the results of this experiment suggest that the addition of the IMAC step increases the dynamic range of detection by selectively enriching for tagged protein while depleting endogenous proteins.

Nearly 40% of all enzymes with known structures function as dimers (25,26). Structure-function studies of homodimers, in particular, are complicated by the fact that introducing any mutations that target the active site could potentially disrupt oligomerization. Therefore, prior to initiating large-scale purification and functional characterization efforts, it is useful to carry out preliminary small-scale studies to determine whether introduced mutations impair homodimer formation and lead to a loss of activity. Because SDS denatures proteins and disrupts protein–protein interactions, SDS-PAGE is not suitable for assessing the oligomeric state of overexpressed proteins. To demonstrate the utility
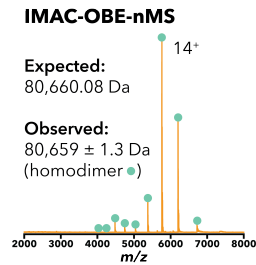
IMAC-OBE-nMS

Expected:
80,660.08 Da

Observed:
80,659 ± 1.3 Da
(homodimer ●)

14+

2000 3000 4000 5000 6000 7000 8000
m/z

Figure 5. Mass spectrum of *Salmonella* FrlB, as analyzed using IMAC-OBE-nMS. The charge state distribution for the FrlB homodimer is indicated with green circles, and the main charge state is labeled. The y-axis (not shown) represents relative intensity.

of IMAC-OBE-nMS for such pilot studies, we selected His$_6$-tagged *Salmonella* FrlB deglycase as a test case because previous nMS experiments had shown that two monomers of wild-type FrlB assemble to form a functional homodimeric enzyme (Kovvali, Di Capua, Gopalan, and Wysocki, unpublished data; see also (27,28)). Our IMAC-OBE-nMS results showed that a FrlB E224Q/H240N mutant mirrors the wild-type in preferentially forming a homodimer, as there were no detectable peaks corresponding to the FrlB monomer (**Figure 5**).

One notable challenge of using heterologous systems for recombinant protein production is that a significant number of proteins are not expressed in a soluble form, accumulating instead in inclusion bodies and precluding the use of non-denaturing purification (7-9). While there are many potential reasons for protein insolubility, the primary cause is misfolding, which may result from suboptimal translation rates or a lack of post-translational modifications, chaperones, or binding partners required for proper folding. A subset of these insoluble proteins can be rescued by fusing a solubility tag such as maltose-binding protein (MBP) to one terminus of the polypeptide chain, which can significantly increase the yield of overexpressed, soluble protein (29,30). To test whether IMAC-OBE-nMS can be used to screen the ability of solubility tags to prevent protein aggregation and precipitation, overexpression samples containing either His$_6$-tagged wild-type *Salmonella* FraR or a variant that is fused to MBP (MBP-FraR) were analyzed (**Figure 6**). Because wild-type FraR is largely expressed as an insoluble protein in *E. coli* (Lai and Gopalan, unpublished data), it was not detected by either OBE-nMS or IMAC-OBE-nMS (data not shown). Even as a fusion construct, though, the apparent low abundance and ionization efficiency of MBP-FraR prevented detection via OBE-nMS, and no peaks corresponding to the target protein were observed (**Figure 6A**). However, IMAC-OBE-nMS successfully enriched for MBP-FraR while simultaneously depleting *E. coli* host cell proteins (**Figure 6B**), thereby facilitating detection of the protein of interest and revealing that addition of the MBP tag helped promote the production of soluble target protein.
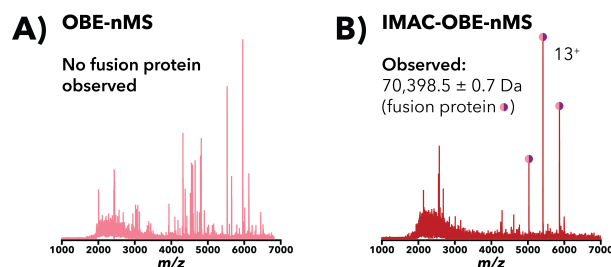
Figure 6. Mass spectra of MBP-*Salmonella* FraR analyzed using (A) OBE-nMS or (B) IMAC-OBE-nMS. The charge state distribution for MBP-FraR is indicated with pink and purple circles, and the main charge state is labeled. In both plots, the y-axis represents relative intensity.

Discussion

We have developed and validated a laboratory-scale approach for rapidly screening cell lysates for successful overexpression and IMAC-based purification of poly-His-tagged proteins of interest without extensive sample preparation. By adding an upstream IMAC-based enrichment step to our recently described OBE-nMS approach (15), we have developed an automated method to determine whether target proteins (i) have been overexpressed in a soluble form, (ii) are able to bind and elute from an IMAC column, (iii) exist as monomers or higher-order oligomers, and (iv) importantly, have the expected mass, with much greater accuracy than is attainable by other methods, including SDS-PAGE. Given the propensity for amino acid misincorporation and proteolysis during recombinant protein production (31,32), having a method that allows for the rapid identification of such issues through accurate mass measurements affords a decisive advantage over SDS-PAGE for even inexpensive small-scale trials.

Our approach is particularly effective for analyzing protein overexpression because it can be used for a variety of different samples, including those with low overexpression and poor protein solubility. In the future, we plan to show that the method may be adapted for use with an even wider range of samples by, for example, substituting the IMAC resin with other media (e.g., Protein A, Protein G, glutathione, streptavidin) that exploit different affinity tags. In addition, while *E. coli* overexpression cultures were used for this study, IMAC-OBE-nMS should, in principle, be compatible with cell lysates or secretomes from any bacterial, archaeal, or eukaryotic host. Thus, IMAC-OBE-nMS can expedite the overall process of optimizing overexpression and purification conditions for large-scale protein production, making it both a powerful stand-alone technique and a potentially complementary addition to other MS-based approaches (33-37) for studying protein structure and function.

Author Information

Florian Busch
Present address: Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany

Zachary L. VanAernum
Present address: Analytical Research and Development Mass Spectrometry, Merck & Co., Inc., 2000 Galloping Hill Road, Kenilworth, NJ 07033 USA

Corresponding author: Vicki H. Wysocki, wysocki.11@osu.edu

Funding and Acknowledgments

References

1. Tabor S and Richardson CC. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1074-1078.
2. Tabor S. (1990) Expression using the T7 RNA polymerase/promoter system. *Curr. Protoc. Mol. Biol.*, **11**, 16.12.11-16.12.11.
3. 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, Joint Center for Structural Genomics, Midwest Center for Structural Genomics, New York Structural GenomiX Research Center for Structural Genomics, Northeast Structural Genomics Consortium, Oxford Protein Production Facility, Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine, RIKEN Structural Genomics/Proteomics Initiative, and SPINE2-Complexes. (2008) Protein production and purification. *Nat. Methods*, **5**, 135-146.
4. Jungbauer A and Hahn R. (2009) Ion-exchange chromatography. *Methods Enzymol.*, **463**, 349-371.
5. McCue JT. (2009) Theory and use of hydrophobic interaction chromatography in protein purification applications. *Methods Enzymol.*, **463**, 405-414.
6. Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Maxwell ES, and Li H. (2010) Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Mol. Cell*, **39**, 939-949.

7.   Francis DM and Page R. (2010) Strategies to optimize protein expression in *E. coli*. *Curr. Protoc. Protein Sci.*, **61**, 5.24.21-25.24.29.

8.   Lebendiker M and Danieli T. (2014) Production of prone-to-aggregate proteins. *FEBS Lett.*, **588**, 236-246.

9.   Rosano GL and Ceccarelli EA. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.*, **5**, 172.

10.   Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

11.   Rath A, Glibowicka M, Nadeau VG, Chen G, and Deber CM. (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1760-1765.

12.   Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, Slunt-Brown H, Borchelt DR, and Shaw BF. (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein Sci.*, **21**, 1197-1209.

13.   Guan Y, Zhu Q, Huang D, Zhao S, Jan Lo L, and 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.

14.   Sulkowski E. (1985) Purification of proteins by IMAC. *Trends Biotechnol.*, **3**, 1-7.

15.   VanAernum ZL, Busch F, Jones BJ, Jia M, Chen Z, Boyken SE, Sahasrabuddhe A, Baker D, and Wysocki VH. (2020) Rapid online buffer exchange for screening of proteins, protein complexes, and cell lysates by native mass spectrometry. *Nat. Protoc.*, **15**, 1132-1157.

16.   VanAernum ZL, Gilbert JD, Belov ME, Makarov AA, Horning SR, and Wysocki VH. (2019) Surface-induced dissociation of noncovalent protein complexes in an extended mass range Orbitrap mass spectrometer. *Anal. Chem.*, **91**, 3611-3618.

17.   Harvey SR, VanAernum ZL, Kostelic MM, Marty MT, and Wysocki VH. (2020) Probing the structure of nanodiscs using surface-induced dissociation mass spectrometry. *Chem. Commun. (Camb.)*, **56**, 15651-15654.

18.   Bern M, Caval T, Kil YJ, Tang W, Becker C, Carlson E, Kletter D, Sen KI, Galy N, Hagemans D, Franc V, and Heck AJR. (2018) Parsimonious charge deconvolution for native mass spectrometry. *J. Proteome Res.*, **17**, 1216-1226.

19.   Heck AJR and van den Heuvel RHH. (2004) Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.*, **23**, 368-389.

20.   Hernández H and Robinson CV. (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.*, **2**, 715-726.

21.   Sherman F, Stewart JW, and Tsunasawa S. (1985) Methionine or not methionine at the beginning of a protein. *Bioessays*, **3**, 27-31.

22.   Ben-Bassat A, Bauer K, Chang S-Y, Myambo K, Boosman A, and Chang S. (1987) Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.*, **169**, 751-757.

23.   Giglione C, Boularot A, and Meinnel T. (2004) Protein N-terminal methionine excision. *Cell. Mol. Life Sci.*, **61**, 1455-1474.

24.   Annesley TM. (2003) Ion suppression in mass spectrometry. *Clin. Chem.*, **49**, 1041-1044.

25.   Marianayagam NJ, Sunde M, and Matthews JM. (2004) The power of two: protein dimerization in biology. *Trends Biochem. Sci.*, **29**, 618-625.

26. Matthews JM and Sunde M. (2012) Dimers, oligomers, everywhere. *Adv. Exp. Med. Biol.*, **747**, 1-18.

27. Sengupta A, Wu J, Seffernick JT, Sabag-Daigle A, Thomsen N, Chen T-H, Capua AD, Bell CE, Ahmer BMM, Lindert S, Wysocki VH, and Gopalan V. (2019) Integrated use of biochemical, native mass spectrometry, computational, and genome-editing methods to elucidate the mechanism of a *Salmonella* deglycase. *J. Mol. Biol.*, **431**, 4497-4513.

28. Graf von Armansperg B, Koller F, Gericke N, Hellwig M, Jagtap PKA, Heermann R, Hennig J, Henle T, and Lassak J. (2020) Transcriptional regulation of the Nε-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues. *Mol. Microbiol.*

29. Fox JD and Waugh DS. (2003) Maltose-binding protein as a solubility enhancer. *Methods Mol. Biol.*, **205**, 99-117.

30. Tropea JE, Cherry S, Nallamsetty S, Bignon C, and 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.

31. Tsai LB, Lu HS, Kenney WC, Curless CC, Klein ML, Lai P-H, Fenton DM, Altrock BW, and Mann MB. (1988) Control of misincorporation of *de novo* synthesized norleucine into recombinant interleukin-2 in *E. coli*. *Biochem. Biophys. Res. Commun.*, **156**, 733-739.

32. Rozkov A and Enfors S-O. (2004) Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.*, **89**, 163-195.

33. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, and Pico AR. (2014) Affinity purification–mass spectrometry and network analysis to understand protein-protein interactions. *Nat. Protoc.*, **9**, 2539-2554.

34. Zhang Y, Sun H, Zhang J, Brasier AR, and Zhao Y. (2017) Quantitative assessment of the effects of trypsin digestion methods on affinity purification–mass spectrometry-based protein–protein interaction analysis. *J. Proteome Res.*, **16**, 3068-3082.

35. Ehkirch A, D'Atri V, Rouviere F, Hernandez-Alba O, Goyon A, Colas O, Sarrut M, Beck A, Guillarme D, Heinisch S, and Cianferani S. (2018) An online four-dimensional HICxSECx-IMxMS methodology for proof-of-concept characterization of antibody drug conjugates. *Anal. Chem.*, **90**, 1578-1586.

36. Vimer S, Ben-Nissan G, and Sharon M. (2020) Direct characterization of overproduced proteins by native mass spectrometry. *Nat. Protoc.*, **15**, 236-265.

37. Takano K, Arai S, Sakamoto S, Ushijima H, Ikegami T, Saikusa K, Konuma T, Hamachi I, and Akashi S. (2020) Screening of protein-ligand interactions under crude conditions by native mass spectrometry. *Anal. Bioanal. Chem.*, **412**, 4037-4043.