

Protein Purification

[PRINT](#)

The PROTEIN PURIFICATION AND CHARACTERIZATION Team headed by George Phillips, Jr., PhD, is responsible for purifying all recombinant protein produced by the project. This team also confirms that the expressed protein corresponds to the expected product.



Goals for CESG Protein Purification Team

- High-throughput purification of labeled and unlabeled proteins suitable for subsequent structural determination studies.
- Develop and utilize screening methods to assess cleavage and solubility characteristics.
- Assess, verify, and record the results of all screening methodologies.

- [Bioinformatics](#)
- [Cell-Free](#)
- [Cloning](#)
- [Crystallography](#)
- [Expression Testing](#)
- [NMR](#)
- [Protein Production](#)
- [Protein Purification](#)
- [Quality Assurance](#)
- [Sesame](#)

Overview of Protein Purification at CESG

An efficient and semi-automated pipeline system has been established for high-throughput purification of *E. coli*-expressed proteins. The protocols have been optimized to work with unlabeled proteins, proteins labeled with Se-Met for X-ray crystallography, and proteins labeled with ^{15}N or ^{13}C ; ^{15}N for NMR spectroscopy.

Key Steps of Protein Purification

1. Initial purification of (His)6-MBP-tagged fusion proteins from cell lysates.
2. TEV protease cleavage.
3. Removal of the liberated (His)6-MBP tag from the target protein.
4. Target protein evaluation and concentration.

Two-Step Purification of Recombinant Protein. Recombinant protein is purified using a two-step chromatography procedure, with an optional polishing step where ionic or size exclusion columns are used to improve the purity of the target protein. Once proteins have been processed through the concentration stages, all samples are sent to the ESI-MS and MALDI-MS analysis for quality assurance before X-ray crystal screening or NMR HSQC data acquisition.

Key Features of the System. The key features of this system include an HPLC that uses a binary gradient pump to purify six proteins sequentially by applying gradient elution from six independent Ni-IDA columns. One-step desalting and subtractive Ni-IDA chromatography removes His-tagged proteins from the target protein. Three such systems are presently in use.

The construct used always yields an N-terminal serine following TEV protease cleavage. Sesame (LIMS) is used for data capture and analysis. Details and statistics for each purification step, including % TEV cleavage, yields, and purity, are recorded.

Optimization of Purification Protocols. In order to optimize the pipeline, we developed protocols for automated protein production. We found that the presence of chaotropic agents such as ethylene glycol and imidazole in the initial sonication buffer are critical to obtain high purity of fusion proteins. Optimum concentration of ethylene glycol and imidazole are 20% (w/v) and 35 mM, respectively. The optimized protocol allows purification of native, SeMet-, ^{15}N -, and ^{13}C ; ^{15}N -labeled proteins up to 150 mg of protein from 2L culture volume. The purity of target proteins is typically greater than 90%, suitable for structural determination by X-ray crystallography and NMR.

Experimental bioinformatic data are used to improve efficiency of current *E. coli* protein production

experimental bioinformatic data are used to improve efficiency of current *E. coli* protein production pipeline. The most common reason for failures of target purification is poor TEV proteolysis, particularly when the percentage of cleavage is less than 70%. The results emphasize the requirement for screening methods that can assess the TEV proteolysis at small-scale cell culture stage. We are in the process of testing expression vectors that contain different proteases cleavage sites and different linker sequences and are developing high-throughput methods for screening for cleavage and solubility at the small-scale expression stage.

We also utilize a highly ordered data storage system, Lamp module in [Sesame \(LIMS\)](#), to monitor the quality control of purification processes and to store the biochemical properties of purified protein such as mobility and purity on SDS-polyacrylamide gel, UV-visible spectra, MALDI-MS, and ESI-MS.

Expression Vectors and Large-Scale Cell Growth. We have designed a (His)_n-MBP fusion tag system ($n = 6$ or 8) to overcome the low solubility of recombinant eukaryotic proteins and to provide a generic Ni-IMAC purification strategy. The pVP13 and pVP16 expression vectors used for these studies were derived from pQE80 (Qiagen, Valencia, CA) to express an N-terminal fusion protein consisting of (His)_n-MBP and a linker region containing the TEV protease site contiguous with the second residue of the target protein. Either *E. coli* Rosetta or B834 strains were used to produce unlabeled-, ¹⁵N-, and ¹⁵N/¹³C-, or SeMet-labeled proteins, respectively. Cells were inoculated in a 2-liter polyethylene terephthalate bottle which contained 500 ml of Terrific Broth or auto-induction medium and incubated in a shaker at 250 rpm, 25°C for 22-24 h. Cells were harvested by centrifugation at 5000 x g for 20 min.

Protein Purification Protocols. The overall philosophy of the protein purification pipeline is to automate the protocols as much as possible while preserving protein quality sufficient for structural studies. Protein purification processes are as follows:

- Step 1: Cell lysis and preparation of the soluble fraction
- Step 2: 1st IMAC capture of His-tagged fusion proteins
- Step 3: Desalting of fusion proteins into TEV proteolysis buffer
- Step 4: TEV proteolysis of fusion tags
- Step 5: 2nd IMAC removal of tag and isolation of target proteins
- Step 6: Desalting of targets
- Step 7: Concentration of targets
- Step 8: Drop-freezing of targets

Steps 2, 3, 5, and 6 are performed on the ÄKTA Purifier controlled by Unicorn 4.12 software. Detail description of purification processes was reported.

Confirm Identify and Integrity of Proteins. All purified proteins are subjected to MALDI-TOF and ESI mass spectrometry to confirm identity and integrity, determine oligomeric state, and investigate possible ligands. Incorporation levels of Se-Met and ¹⁵N and ¹³C isotopes are also determined by ESI-MS.

Methodology and Technology Development

+ High-Throughput Processing Using Data Input / Recording Tablet

One of the most important parts of any research effort is recording experimental parameters and results. For the PSI, the time spent recording results is also essential to populate the NIH KnowledgeBase. However, as the number of samples and the complexity of the associated processing grow in large-scale efforts such as the PSI, the time required to collect and store the experimental information can become a significant bottleneck. And, this is especially the case in Protein Purification where a myriad of samples are processed.



In order to address this bottleneck of experimental information capture, CESG has begun to explore the use of a lightweight tablet PC with built-in wireless connection and the ability to recognize handwriting as the user interface to record data and observations during the purification process. This system has been also linked to a barcode inventory control system for identification and retrieval of the frozen cell pastes available for purification. The tablet and scanner is shown at left. The tablet is a Fujitsu Stylistic ST5112 Tablet PC and the scanner is a Datalogic Lynx BT reader.

These provide the vehicle to interface with an Excel-based spreadsheet and the Sesame (LIMS) system. Thus, the data that the report generates can be easily imported into Excel. The tablet has also been coupled with a Bluetooth enabled scanner which allows for inventory control of cell pastes available for purification.

Advantages of the Tablet Input System

- Allows data recording and progress assessment to happen anywhere in the lab, and in real-time.
- As data is entered into this spreadsheet, all necessary calculations are automatically processed, and as an electronic form, it could be sent to anyone around the world. Coupled with on-line, 24 h access to Sheherazade, which is straightforward, supporting gels and other experimental data files can also be viewed in real-time.
- The lightweight and portability of the tablet makes its use practical in the laboratory setting.
- Time saving arises from the availability of automatic calculations and error reduction is provided by the clear text format, which now replaces what a scrawled number or word written on the fly during a procedure.
- Like the paper form it replaced, the tablet has the ability to capture graphic images. For example, a critical point can be circled or an arrow can be drawn to bring attention.

- A WORKER AT THE CHROMATOGRAPHY CHAMBER, RECORDING GEL FILTRATION COLUMN DATA AND RUN PARAMETERS (LEFT),
- AT THE GEL BOX, ASSESSING PURITY AND SIZE OF PURIFIED PROTEIN TARGETS (MIDDLE),
- AND SCANNING CELL PASTES AND ADDING THEM TO INVENTORY (RIGHT).



Use of the System for Inventory Control: All cell pastes grown in Large-Scale get a barcode that links to Sheherazade. The record contains the expression, solubility, and cleavage results for the protein candidate for purification, plus cell mass. An inventory report was developed in Sheherazade that returns all the testing results, vector information, and whether or not a particular sample is an outside request. This report is fed into an Access based database, which has been coded to sort the cell pastes, allowing for easy selection. Outside requests receive the highest priority and, then, cell pastes are sorted based on suitability scores.

Protein Purification and Characterization Team

- George Phillips, Jr., PhD (PI)
- Frank Vojtik (Team Leader)
- Karl Nichols

