How to isolate proteins

Manju Kapoor

Background

Numerous authoritative books, excellent reviews and articles have been written on this subject. While general methods for isolation and purification of proteins are applicable to all organisms, it is invariably necessary to develop unique strategies for isolation of the target protein of interest. Unlike research with DNA, no manual of standard protocols or "recipes" is available, outlining a stepwise approach applicable to all proteins. Furthermore, there are no organism-specific procedures that can allow one to plan a course of action with a predictable outcome. Protein purification has been described as "more of an art than a science". The design of an appropriate procedure for isolation of a given protein should be tailored in accordance with the objective(s) of the research project, which may require relatively pure product in modest amounts for analytical purposes (e.g. enzyme kinetics) or a highly purified, homogeneous preparation for physicochemical or structural studies. Isolation and purification of a single protein from cells containing a mixture of thousands of unrelated proteins is achievable because of the remarkable variation in the physical and chemical attributes of proteins. Characteristics unique to each protein—amino acid composition, sequence, subunit structure, size, shape, net charge, isoelectric point, solubility, heat-stability, hydrophobicity, ligand/metal binding properties and post-translational modifications—can be exploited in formulation of a strategy for purification. Based on these properties a combination of various methods, listed below, can be used for separation of cellular proteins (Refs 1, 2).

Procedure

Separation Method	Property
1. Precipitation	
Ammonium sulfate	Solubility
Polyethylene glycol	Solubility
2. Chromatography	
Ion-exchange (anion or cation)	Net charge

Hydrophobic interaction	Surface hydrophobicity
Metal affinity	Metal-binding sites
Ligand affinity	Ligand-binding sites (e.g. NAD, NADP)
Gel filtration	Subunit/oligomer size, shape
3. Centrifugation	Size, shape

Generic outline for protein purification

In general, protein purification entails essentially five types of steps: 1) efficient extraction from biological material; 2) separation from non-protein components (nucleic acids and lipids); 3) precipitation steps, initially to recover the bulk protein from a crude extract, followed by preliminary resolution into manageable fractions; 4) use of ion-exchange chromatography/size fractionation or hydrophobic chromatography columns to further separate the target protein-containing fraction from the bulk protein; 5) a more refined set of steps including an "affinity" matrix to enable recovery of the target protein in a highly purified state along with a high yield. A variety of agarose-based matrices with immobilized reactive dyes, covalently bound nucleotides, metals and numerous other ligands are commercially available (supplied by Sigma, Amicon, etc.).

In order to evaluate the progress of purification, a convenient assay procedure—based on enzymatic activity or some other easily monitored property specific to the protein—should be available. A spectrophotometric or colorimetric method for enzymatic activity measurement is most convenient and a progressive increase in specific activity (for enzymes, activity in units /mg protein) is an excellent indicator of the efficacy of the purification step. For proteins lacking a readily measurable biological activity, it may be feasible to use an immunochemical procedure such as western blotting or ELISA (Enzyme-Linked-Immuno-sorbent Assay), provided suitable antibodies are available. In this case, electrophoretic resolution of the protein population in samples at each stage of purification will be required.

Purification of native proteins

While purification of the native proteins is a challenging exercise, several reliable approaches have stood the test of time. Compared to soluble proteins, membrane-bound

proteins are more difficult to purify. Solubilization of membrane proteins can be achieved by the use of detergents but removal of the detergent is necessary for subsequent analytical manipulations. A detailed treatment of the properties of various detergents and their applications is available in reference 1. In the following a representative procedure for purification of soluble *Neurospora* proteins is outlined.

1. Preparation of crude extracts: Efficient extraction of the total protein from the starting material is vital for success of any purification procedure. Complete disruption of cells and release of contents from cellular debris is the most important step in the process. For purification of *Neurospora* proteins in the native state, the first step involves the extraction of bulk protein fraction from mycelial cells. All steps in the procedure are carried out at 4°C to minimize protein degradation. Mycelial cultures are grown for 18 to 20 h in a medium conducive to optimal production of the target protein, harvested, lyophilized and stored at –70°C. Ten to 20 g of lyophilized mycelial powder is suspended in 10 volumes of an extraction buffer (50 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 1 mM β-mercaptoethanol or dithiothreitol) and the mixture is stirred for 45 min in the cold room. The presence of EDTA serves to inhibit protease action and β-mercaptoethanol (or DTT) is necessary for maintenance of a reducing environment. This slurry is homogenized using a glass homogenizer and the homogenate is centrifuged at 12 000 x g for 20 min (to remove cellular debris) in a refrigerated Centrifuge. The pellet is discarded and the supernatant is used in subsequent steps. At this stage it may prove helpful to add a mixture of protease inhibitors (Complete cocktail: Roche or Sigma) if the target protein is suspected to be unstable. [**Note**: Nucleic acids can be removed from the extract by addition of protamine sulfate to a final concentration of 0.2%, while stirring. The precipitated nucleic acids are removed by centrifugation. For most purposes, nucleic acid removal is not necessary; the precipitate may also bind the protein of interest].

2. Precipitation of proteins: Several methods are available for precipitation of proteins utilizing changes in pH and temperature, or addition of salts and organic solvents. Ammonium sulfate is the most commonly used precipitant for salting out of proteins. At saturation (3.9 M at 0°C and 4.04 M at 20°C) it precipitates most proteins and protects proteins in solution from denaturation and bacterial growth. To the supernatant from step

1, sufficient solid $(NH_4)_2SO_4$ (Ultrapure reagent or Enzyme grade) is added to achieve 40% saturation [See Ref.1 for Table showing relationship between $(NH_4)_2SO_4$ concentration and % saturation]. To avoid surface denaturation, the solution should not be stirred vigorously and $(NH_4)_2SO_4$ should be added gradually, in small amounts, allowing each successive batch to dissolve completely before addition of the next. The precipitated protein is removed by centrifugation at 12 000 x g for 10 min and to the supernatant more $(NH_4)_2SO_4$ is added to yield 80% saturation. The fraction of precipitated proteins between 40 and 80% saturation is recovered by centrifugation, resuspended gently in 5 to 10 ml of a suitable buffer (e.g. 20 mM Tris-HCl, pH 7.5, 20 mM NaCl, 10 mM $MgCl_2$) and dialyzed in the cold room against several, 4-L changes of the same buffer over a 16-h period to remove residual $(NH_4)_2SO_4$. The dialyzed suspension is then centrifuged at 12 000 x g for 10 min to remove insoluble particulate matter and the supernatant is tested for the presence of the target protein (pX).

3. Ion-exchange chromatography: The dialyzed fraction is applied to a 16 mm x 30 cm column packed with an anion-exchanger, DEAE-cellulose (Sigma Fast Flow Fibrous DEAE Cellulose) or DEAE-Sepharose, previously equilibrated against the above-mentioned dialysis buffer. The column is connected to a Pharmacia P-1 pump and a Frac-100 fraction collector and is washed with ~60-100 ml of buffer to remove unbound proteins. The protein fraction bound to the matrix (including the target protein) is eluted with 150 ml of a linear 0 to 1.75 M NaCl or KCl gradient, prepared in the same buffer, generated by a Pharmacia GM-1 gradient mixer. [**Note**: See instructions for column packing in Ref. 2].

Alternatively, the fraction can be chromatographed by passage through a Mono Q anion-exchange column (HR 5/5) attached to a Pharmacia Fast Protein Liquid Chromatography (FPLC) system. The sample is clarified by centrifugation, loaded onto the column and eluted with a discontinuous gradient consisting of steps of 0 to 0.3 M, 0.3 to 44 M and 0.44 to 1.2 M NaCl, as an example. The fractions enriched in pX are pooled and centrifuged at 12 000 x g for 10 min to remove insoluble material. The supernatant is dialyzed for 4 to 6 h against 20 mM Tris-HCl, pH 7.5 to remove NaCl, brought to 80% $(NH_4)_2SO_4$ and the precipitated fraction is resuspended in 20 mM Tris-HCl, pH 7.5. If

necessary, residual salt can be removed by passage through a small gel filtration column. For applications requiring cation-exchange columns carboxymethyl (CM)-cellulose or CM-Sepharose can be employed.

4. Separation by hydrophobic interaction: The protein sample from either of the preceding steps—in 30% saturated $(NH_4)_2SO_4$—can be applied to a column containing a hydrophobic matrix, such as Phenyl- or Octyl Sepharose, pre-equilibrated with 20 mM Tris-HCl (pH 7.5) containing 30% saturated $(NH_4)_2SO_4$. The column is washed successively with buffer containing 25%, 20%, 15%, 10% and 5% saturated $(NH_4)_2SO_4$. Finally the protein is eluted with the original buffer. The fractions containing pX are combined and concentrated using Centricon filter concentrators (10 or 30 kDa cutoff).

5. Affinity chromatography: An example of affinity chromatography for separation of NAD(P)-binding proteins is the use of agarose or sepharose-bound reactive dyes. The protein sample is loaded onto a 10 mm x 20 cm column packed, for instance, with Cibacron Blue-agarose 3GA resin (immobilized on cross-linked 4% beaded agarose, type 3000 CL; Sigma), pre-equilibrated with a buffer. Proteins bound to the matrix are eluted with ~80 ml of a linear 0 to 1.75 M NaCl gradient, fractions with pX are pooled and centrifuged to pellet insoluble proteins. The supernatant is concentrated and electrophoresed in SDS-polyacrylamide gels to determine the state of purity of the protein. Other types of affinity matrices, such as ADP-, ATP-agarose or Con A-Sepharose (containing immobilized adenosine nucleotides or Concanavalin A) can be employed to capture ATP/ADP-binding proteins or glycoproteins, respectively.
[**Note**: Selection of the appropriate combination of purification steps will depend on the properties of the target protein. If it is a known protein, such as an enzyme that has been isolated and characterized from another organism, a good starting point will be to try out some of the steps in the published procedure.]

   The state of purity of the sample is judged by SDS-PAGE; the number of stained polypeptide bands will decrease with progressive removal of contaminating proteins. Ultimately, a single, stained band should result, indicating a nearly homogeneous sample. An efficient purification protocol should also result in high enough yield to justify the

effort involved. Ideally, no more than five or six purification steps should be required for purification of a given protein. The entire process should be completed in four to five days as a small proportion of the target protein is inevitably lost in the course of each step, due to denaturation, protease action and storage conditions. On starting with ~20 g of lyophilized mycelium, the yield of inducible enzymes and environmentally-regulated proteins in the vicinity of 5-10 mg is normal. Higher yields can be attained if losses during individual steps can be kept to a minimum. The final homogeneous preparation should be stored in a concentrated solution ($\sim$ 10 mg/ml) at -20° C or -70°C, along with 30% glycerol for maximum stability. All of the above steps can be scaled-up.

Purification of over-expressed proteins

Over-expression of complete proteins, or fragments, is desirable if large quantities of the product are required for structural studies/x-ray crystallography.   The availability of cDNA clones encoding the proteins of interest simplifies the purification procedures. By constructing suitable bacterial strains large quantities of the required product can be generated. *E. coli* cells are employed most commonly for heterologous expression of fungal and other eukaryotic proteins and a variety of cloning vectors—harboring selectable markers, replication origin sequences, promoters, translation initiation codon and a suitable "tag" sequence, initiation and termination sequences—are available commercially. Several host systems/expression vectors with different tags to facilitate purification of the over-expressed proteins are available. For instance, the Pin-point Xa-Protein purification system (Promega) is designed to permit purification of proteins that are biotinylated *in vivo*.

An example of a popular expression vector is the pUC-derived pRSET series (Invitrogen; described in Ref. 3) harbouring the T7 promoter, followed by ATG, the metal-binding domain comprised of 6 histidine residues, the enterokinase cleavage recognition sequence, the multiple cloning site and the β-lactamase gene. The recombinant plasmids are transformed into a suitable host strain, such as *E.coli* BL21(DE3)pLysS cells for expression. The transformants are selected based on ampicillin resistance and grown under conditions conducive for maximum expression.

The resulting, over-expressed protein is captured readily by an appropriate "affinity" matrix.

Representative purification procedure: Isolation of the protein, resulting from fusion of the coding sequence with the metal-binding tag sequence, involves the following: (1) growth and induction of transformed bacterial cultures; (2) lysis of cells in a suitable buffer containing a detergent and lysozyme; (3) DNase and RNase treatment for removal of the nucleic acid fraction; (4) passage of the extract through an affinity resin (Ni^+-NTA agarose for His-tagged proteins) and (5) elution of bound protein. Many of the so-called "one-step" purification systems, offered by commercial suppliers, frequently require multiple steps for recovery of a homogenous product.

Expression and purification of the fusion protein: The initial inoculum is prepared in SOB-amp medium (20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 0.186 g KCl, 0.01M $MgCl_2$ per litre containing 50 µg/ml ampicillin) by addition of 3-5 µl of the stock culture and incubation at 35^oC for 4-6 h while shaking. Two-ml aliquots of the culture are used to inoculate fresh 50-ml SOB-amp medium which is incubated at 30^oC while shaking for 12 h. For expression, 20 to 30 ml/litre of this culture is used to inoculate SOB-amp medium, yielding a starting OD_{600nm} of 0.07 to 0.11. Incubation is resumed in the shaker at 30^oC for 3.5 to 4.5 h, until OD_{600} of 0.35 to 0.45 is attained. Addition of $1-2$ mM IPTG during the last hour of growth can, often though not invariably, enhance the yield of the fusion protein. The culture is harvested by centrifugation and the pellet frozen at -20^oC, thawed on ice, resuspended in cold Extraction Buffer (50 mM phosphate buffer, 300 mM NaCl, pH 8.0, containing 20 mM imidazole and 10 mM β-mercaptoethanol) and the cells are homogenized using a glass homogenizer. A detergent such as Triton X-100 is added to a final concentration of 0.01% followed by freshly prepared lysozyme (100 µg/ml) and the mixture shaken at room temperature for 20 min.

Next, DNase (5 µg/ml) and RNase (6.25 µg/ml) are added and shaking continued for an additional 30 min and the crude cell extract is centrifuged at 12 000 g for 20 min. The resulting supernatant is clarified by re-centrifugation, mixed with 1 to 1.5 ml of the affinity resin (Ni^+-NTA agarose; Qiagen), pre-equilibrated with 20 mM imidazole in phosphate buffer (50 mM NaH_2PO4, 300 mM NaCl, pH 8.0). The mixture is shaken for

90 min at 4°C to allow binding of the His-tagged protein. The resin is poured into a column and washed in two steps: first with the above phosphate buffer containing 20 mM imidazole, followed by two washes with 40 mM imidazole to remove the non-specifically bound protein fraction [also see Ref. 3].

Elution is conducted in a batchwise manner with 2 ml of 150 mM, 3 ml of 200 mM and 3 ml of 250 mM imidazole. Alternatively, a linear gradient of 100 mM to 400 mM imidazole can be employed if larger volumes of the affinity resin and the supernatant are to be used. The eluted fractions are monitored for the presence of the fusion protein by SDS-PAGE. The fractions containing high levels of the fusion protein are pooled and concentrated using Centricon filters (with a cutoff of 30K or 10K depending on the size of the protein). The filters are washed with 1 ml of appropriate buffer (e.g. 20 mM Tris-HCl, pH 7.5) to remove residual imidazole. The purity of the protein should be checked by SDS-PAGE and its identity verified by western blotting, if specific antibodies are available or by measurements of specific activity if the expressed protein is an enzyme. When contaminating proteins are present, additional purification steps including size fractionation or ion-exchange chromatography will be necessary.

Problems are often encountered during the initial cell lysis step resulting in incomplete liberation of the expressed protein from bacterial cell debris. Therefore, it is essential to explore different methods of cell breakage, such as sonication, freeze-thaw cycles, etc. to optimize the yield. A number of commercial cell-lysis solutions—claimed to be suitable for bacterial, yeast or mammalian cells— are marketed by various suppliers but home-made mixtures are just as satisfactory. A sizeable fraction of the protein can also be lost by formation of insoluble aggregates in the so-called inclusion bodies. Some of the aggregated protein may be recoverable by using one or more cycles of denaturation-renaturation. However, the latter approach can lead to inactivation of the protein and reduced yield. Low yields may also result from proteolytic degradation of the target protein during one or more of the purification steps. Inclusion of protease inhibitors in the lysis buffer and/or subsequent steps is helpful.

While many eukaryotic proteins have been successfully expressed in bacterial hosts, recovery of biologically active proteins is often difficult due to the lack of post-translational modifications that can occur only in a eukaryotic intracellular milieu. As an

alternative, *Pichia pastoris* cells can be used for expression of fungal proteins. The main advantage is the ability of *P. pastoris* cells to perform appropriate post-translational modifications.   However, many proteins are not readily expressed in this system.

References
1. Methods in Enzymology (1990) volume 182. A Guide to Protein Purification. Edited by M. P. Deutscher.   Academic Press.

2. Ausubel, F.M., Brett, R., Kingston, R. E., Moore, D. D., Seidman, J.G., Smith, J.A., and Struhl, L. 1991. Current Protocols in Molecular Biology, Vol. 1. Wiley. New York.

3. Xpress^{TM} System Protein Purification: A Manual of Methods for Purification of polyhistidine-containing Recombinant Proteins. Invitrogen.