Introduction to Recombinant Protein Purification

5

Nitu Singh and Kakoli Bose

Abstract

Purified form of recombinant proteins is a prerequisite for undertaking in vitro biochemical and structural analyses of these macromolecules. Unfortunately, due to other proteins from the expression host, such as *E. coli,* the task of obtaining the desired protein from the heterologous system at highest purity and in sufficient quantity is arduous. With the increasing demand for recombinantly purified proteins both in basic and industrial research, over the last five decades, a plethora of research endeavors have been directed toward developing efficient protein purification techniques that would precisely amalgamate time and yield cost-effectively. However, it is extremely important to put careful forethought prior to developing a purification flow-scheme for a target protein to obtain the best possible output. This chapter outlines the general considerations to be undertaken while designing and streamlining these protocols with the help of recent advances in protein purification methodologies. It also provides an overview of the various chromatographic techniques that will be further elaborated in the succeeding chapters of this book.

Keywords

Purification · Protein stability · Salting · Inclusion bodies · Dialysis · Chromatography

N. Singh
Regional Centre for Biotechnology, NCR Biotech Science Cluster, Faridabad, Haryana, India

K. Bose (✉)
Integrated Biophysics and Structural Biology Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Mumbai, India

Homi Bhabha National Institute, BARC Training School Complex, Mumbai, India
e-mail: kbose@actrec.gov.in

K. Bose (ed.), *Textbook on Cloning, Expression and Purification of Recombinant Proteins*, https://doi.org/10.1007/978-981-16-4987-5_5

5.1 Introduction

Protein purification gained major attention, when James B. Sumner in 1926 started purification and crystallization of urease from yeast [1]. Since then, biochemists, in particular, have put tremendous effort in developing convenient purification methods for obtaining recombinant proteins from the bacterial or eukaryotic expression systems using the differences in the physicochemical properties of the proteins [2, 3]. Over the past five decades, the methods for separation and purification of recombinant proteins have contributed immensely to the advancements in the fields of biosciences and bioengineering. This has been mainly possible due to the increased use of tags for separation of recombinant proteins, development of new chromatographic techniques, and use of computerized sophisticated instruments [4, 5]. The challenge in protein purification is self-evident, given the complex mixture of biomolecules present inside an expression host, commonly used for recombinant protein extraction. With the advent of new generations of chromatographic media and automated systems, gone are those days when an investigator used to spend several months in establishing a purification protocol. However, not all problems are resolved by using sophisticated column packing and laboratory equipment. Difficulties still persist in finding optimal conditions for sample pretreatment, choosing an appropriate buffer condition, or dealing with insolubility of the proteins. Therefore, prior to stepping into purification of a particular protein of interest, it is important to consider a few things such as the purpose of purifying the recombinant protein, its purity, and its required concentration as well as storage condition [6]. All these factors are critical in designing and executing a procedure for purifying a target protein to a sufficient degree in a cost-effective and timely manner. In this chapter, we discuss purification steps for isolating recombinant proteins from hosts, such as *E. coli*, emphasizing on the conditions for cell lysis, protein solubility and stability. It also includes an overview of protein separation techniques using different chromatographic platforms, with each individual technique discussed in more detail in the subsequent chapters of this book.

5.2 Databases and Tools to Determine Physicochemical Properties of Protein

Now that more than thousands of whole genomes have been sequenced, determining the physicochemical properties of the proteins from their amino acid sequences can help in designing an effective purification strategy for that protein. Some of the parameters that we can learn from the amino acid sequence of the protein using a sequence analysis software is described below.

5.2.1 Physicochemical Parameters Important in Initial Designing of the Purification Procedure

5.2.1.1 Molecular Weight of the Protein

The amino acid composition of the protein can be easily used to calculate the molecular weight of the polypeptide, such as for a cloned protein. This information is very helpful in the initial estimation of the level of protein expression on SDS-PAGE (in case the expressed protein is large enough to give a distinct band). However, it is important to note that the calculated molecular weight and the apparent weight based on the mobility on SDS-PAGE can sometimes deviate due to several other factors [7]. A more quantitative method of molecular mass estimation, such as mass spectrometry and analytical ultracentrifugation, can be used in such cases [8, 9]. In addition, the knowledge of the molecular weight is also important for the initial selection of the size-exclusion chromatography medium for separation of the proteins under native conditions [10]. In cases where the amino acid sequence is unavailable, one can combine the size exclusion chromatography (size-based separation technique) and the activity assays to estimate the molecular weight [10].

5.2.1.2 Isoelectric Point, pI

Theoretically, the isoelectric point of the protein can be estimated from the primary sequence or can be determined experimentally using isoelectric focusing [11]. The pI value is helpful toward selecting a suitable matrix for purification using ion-exchange chromatography (charge-based separation technique) [12]. Additionally, since a protein tends to have lowest solubility at its pI [13], one can also consider an isoelectric reversible precipitation step during purification considering that the target is not in a stable complex with other proteins [14]. This helps in isolating the recombinant protein form the cellular lysate of the host, such as *E. coli*, prior to its loading on a chromatographic column.

5.2.1.3 Molar Extinction Coefficient/Absorptivity Coefficient

Using the amino acid sequence of a protein, one can calculate its molar extinction coefficient. This is one of the important parameters for the estimation of the protein concentration. At a wavelength of 280 nm, tyrosine, tryptophan, and cystine (cysteine does not absorb much at wavelengths >260 nm, while cystine does) exhibit strong UV-light absorption. Cystine is formed when a pair of cysteine molecules join together by a disulfide bond. Using the Beer-Lambert equation, the extinction coefficient (ε) of the native protein is computed [14, 15]:

$$\varepsilon\,(\text{Prot}) = (\text{Number of Tyr}) * \varepsilon(\text{Tyr}) + (\text{Number of Trp}) * \varepsilon(\text{Trp}) + (\text{Number of Cystine}) * \varepsilon(\text{Cystine})$$

where $\varepsilon(\text{Tyr}) = 1490$, $\varepsilon(\text{Trp}) = 5500$, $\varepsilon(\text{Cystine}) = 125$; (for proteins in water measured at 280 nm)

The absorption of the UV-light is proportional to the aromatic amino acid content and total concentration of the protein. Using the estimated extinction coefficient (with fixed amino acid composition of a given protein), we can easily calculate the protein's concentration in solution from its absorbance. The online tool ProtParam (http://www.expasy.org/tools/protparam) gives two values based on the above equation [16]. The first value is based on the assumption that all cysteines in the sequence occur as half cystines (i.e., all pairs of cysteine residues form cystines), while the second value assumes that all pairs of cysteines are in reduced state. It is reported that this computation is considerably reliable for Trp containing proteins; however there may be around 10% error in proteins without Trp [17]. Nevertheless, this method is not valid if the protein sample contains nucleic acid contaminants or other components that show significant absorption at 280 nm such as iron–sulfur centers, or bound cofactors. Moreover, it is not accurate in case of complex protein solutions, such as cell lysates, which contain a mixture of proteins with unknown extinction coefficients.

5.2.1.4 Cysteine Content

Knowing the number of cysteines in the protein sequence will help decide whether the purification buffer should contain a reducing agent. The reducing agents such as beta mercaptoethanol or dithiothreitol in the buffer can prevent any unwanted intra- or inter-disulfide bond formation and aid in protein solubility as well as stability.

5.2.1.5 Stability

The knowledge of the protein stability with respect to pH, salt, temperature, proteases, or aggregation can be very helpful during purification and storage of the proteins [18]. However, most of these parameters are only experimentally derived. Using bioinformatics tools, such as ProtParam [16], one can estimate the in vivo half-life and instability index of the protein from its amino acid sequence. Half-life is the time taken for half of the amount of protein in a cell to disappear after its synthesis. The calculation of half-life in ProtParam relies on the N-end rule [19]. This rule that originated from the studies performed in model systems such as mammalian cells, yeast, and *E. coli* demonstrate importance of N-terminal residues in determining the stability of the protein [20].

The instability index (II) is estimated by calculating the frequency of occurrences of dipeptides in the protein of interest, which is then compared with a set of test proteins that are known to be stable or unstable [21]. The instability value less than 40 is predicted as stable, while a value above 40 indicates that the protein might be unstable. This information is useful during protein purification as a protein predicted to be unstable needs to be handled with special care such as use of low temperature during purification and addition of salts and/or protease inhibitor cocktails in the extraction buffers. Furthermore, a shorter purification and storage time could also be considered.

5.2.1.6 Hydrophobicity

It is possible to predict if the protein is hydrophilic or hydrophobic by analyzing the amino acids in the sequence. One such method was developed by Kyte and Doolittle that plots the hydropathy value over the length of the protein sequence [22]. Based on the hydrophobic or hydrophilic properties of the 20 amino acids, the hydropathy scale is calculated at each point in the sequence. This can help in determining the hydrophobic core of the protein and potential membrane spanning region, which will certainly be important in designing the purification scheme. The ProtParam tool uses this method to predict the GRAVY (grand average of hydrophobicity) value for a protein, which is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence [22].

5.2.2 Bioinformatics Resources

One of the most frequently used online tools for computing the physicochemical properties of proteins using the amino acid sequence is the ProtParam feature of ExPASy (Expert Protein Analysis Software) (http://www.expasy.org/tools/protparam) as mentioned above. ProtParam calculates parameters such as molecular weight, amino acid composition, molar extinction coefficient, atomic composition, instability index, aliphatic index, and GRAVY [16].

To use this tool, go to ExPASy ProtParam tools and either enter the Swiss-Prot/TrEMBL accession number (for example, Q9VFJ3- human mitochondrial serine protease HtrA2) or a sequence identifier (ID) (for example, HTRA2_DROME), or you can paste the raw amino acid sequence (in one-letter code) in the box and click "compute parameters." The result obtained can be saved or printed for further analysis.

Another commercial package is "Protean" from the Lasergene protein module of DNASTAR (http://www.dnastar.com). Similar to ProtParam, it also performs protein sequence analysis, wherein it computes parameters such as protein stability. In addition, it also allows for residue-specific change in the protein 3D-structure and helps in predicting whether these changes promote stabilizing or destabilizing effects. Using the Lasergene "Protein's advanced protein design software" one can make accurate protein stability predictions in minutes for the designed mutants.

5.3 Lysis and Protein Extraction

5.3.1 Source Material for the Protein

Although purification of proteins from their natural sources is relatively common, with the recent developments in the field of gene manipulation and recombinant expression, one usually targets overexpression systems such as bacteria, yeast, insect, or mammalian cells. Each individual host system has its own advantages and disadvantages. For instance, *E. coli* gives the highest yield, typically an average

yield of 2.5 g/l of culture, while it lacks the posttranslational modifications, if any [23]. On the other hand, if the protein requires glycosylation, the yeast system provides more extensive glycosylation than insect and mammalian cells [24]. The choice depends on the properties of the protein as well as on its downstream application.

Proteins overexpressed in genetically modified organisms or cultured eukaryotic cells are majorly localized in the cytoplasm or can be targeted for secretion into the medium. In case of bacterial expression, such as in *E. coli*, they can also be targeted to the periplasmic space or selectively released in the growth medium by altering the growth conditions [25]. This can help in eliminating the need for cell lysis, thus providing a substantial level of purity at the initial steps. Therefore, choice of the host and overexpression strategy is one of the rate-limiting steps of protein purification.

5.3.1.1 Extraction Methods

In most cases, the extraction procedures depend on the source of the protein, which could be bacterial, yeast, or mammalian cells and either intracellular or extracellular. Extraction from an intracellular source often faces compromised recovery and purity. The main objective of the extraction should be to achieve the desired protein in a non-degraded or non-denatured form with minimal or no contaminants.

The extraction protocol is usually optimized by strategic variations in the parameters such as extraction medium, time, temperature, equipment for lysis, and energy input (agitation speed, pressure, etc.). It is to be noted that the choice of the method should be such that it is as gentle as possible because too vigorous or harsh conditions might denature the desired protein or release the endo-proteolytic enzymes and cause general acidification. Also, one should keep in mind that the extraction should be performed quickly at low temperatures (at 4 °C or on ice) in a suitable buffer to maintain the ionic strength and pH to stabilize the protein [3]. One of the major problems that is confronted during extraction is proteolysis or contamination with nucleic acids. However, to some extent this can be tackled by performing the extraction at low temperatures in the presence of protease inhibitors and with inclusion of nucleases in the extraction medium. Therefore, for an optimized design, it is essential to standardize preliminary experiments in a small scale for maximization of protein content and activity, which can later be scaled-up effectively.

The composition of the extraction medium should be such that protein remains stable and is effectively released from the cells with maximum recovery and purity. Following are the factors that should be taken into consideration while preparing the extraction medium/lysis buffer: buffer salt, pH, reducing agent, chaotropic agents, detergents, metal ions, proteolytic inhibitor, and DNase [3].

5.3.1.2 Extraction Medium/Lysis Buffer

Buffer Salt and pH

There are several factors that should be considered while selecting a buffer. Firstly, pH optimum of the protein is important in determining the best buffer condition. In

Table 5.1 Buffers and pKa values

Buffer	pKa (25 °C)	pH range for use
Phosphoric acid	2.12	1.1–3.1
Formic acid	3.75	2.8–4.8
Acetic acid	4.75	3.8–5.8
Sulfonic acid	6.91	5.9–7.9
Dihydrogen phosphate ion	7.21	6.2–8.2
Ammonium ion	9.25	8.2–10.2
Hydrogen phosphate ion	12.66	11.3–13.3

Table 5.2 Detergents used for protein solubilization [28]

Detergent	Ionic character	Critical micelle concentration (% w/v)
Triton-X 100	Non-ionic	0.02
Nonidet P-40	Non-ionic	0.012
Octylglucoside	Non-ionic	0.73
Tween 80	Non-ionic	0.002
Sodium deoxycholate	Anionic	0.21
CHAPS 3-[(3-Cholamidopropyl dimethyl amino propanesulfonic acid)	Zwitterionic	1.4

practice, the pH value is chosen such that the protein activity is maximum. Since the buffering capacity is maximal within one pH unit from the pKa value, the selected buffer preferably should be close to this value (Table 5.1, buffers with the pKa values). Secondly, good buffers should be relatively free of side effects. For example, TRIS and primary amine buffers can form Schiff base adducts with aldehydes and ketones and interfere with Bradford dye-binding assay [26]. Some inorganic buffers can remove metal ions by chelation, thus inhibiting metal-dependent protein activity. Lastly, in protein purification, the cost and compatibility of the buffer with different purification techniques are important parameters to be considered. Once an ideal buffer is chosen, one needs to select a suitable ionic strength of the buffer. Most of the proteins show maximum solubility and activity at low to moderate ionic concentrations, 0.05–0.1 M. It is to remember that proteins also act as buffers, and therefore it is important to carefully monitor the pH after addition of large amounts of proteins to a weakly buffered solution.

5.3.1.3 Detergents and Chaotropic Agents

In some cases, the desired protein could be associated with the membrane or might be aggregated due to its hydrophobic nature. In such cases, use of detergent or chaotropic agents helps in solubilization of the protein, thereby allowing its separation. The detergents are amphipathic molecules that during solubilization help the membrane proteins to partition into apolar lipid bilayers [27]. They also aid in masking the hydrophobic surface of the proteins, thus preventing protein aggregation. Some of the commonly used detergents are listed in Table 5.2. Many of these at a concentration below the critical micelle concentration (CMC), i.e., the

concentration of a surfactant above which it starts to form micelle, do not denature or interfere with the protein's biological activity, with few exceptions such as SDS (sodium dodecyl sulfate) [29]. Mostly, detergents are added in the first step of purification, i.e., in the lysis buffer. This is because the presence of detergents in the later steps often complicates the purification process, especially in column chromatography [6]. In a few cases, there might be a requirement of the detergent throughout the purification process, leading to separation of purified protein-detergent complexes. Beside detergent, one can also use chaotropic agents, such as urea, guanidine hydrochloride, or polyethylene glycol (moderate hydrophobic organic compound), to solubilize the aggregates [30]. Chaotropes are agents that denature proteins by disordering the surrounding water molecules through disruption of the hydrogen bonding network among the water molecules. These chaotropic agents are commonly used in purification of the recombinant proteins from the inclusion bodies as discussed in the following sections.

5.3.1.4 Reducing Agents

Proteins with free cysteine residues (i.e., exposed thiol groups) can easily get oxidized to disulfides, sulfinic acid (—SOOH), or sulfonic acid (—SO_2OH) during the process of purification [31]. This is generally due to lower redox potential inside the cell as compared to the surrounding medium used for protein purification. The oxidation of the thiol groups could be problematic as it would affect the solubility and stability of the proteins. Reducing agents such as 1,4-dithioerythritol (DTE), dithiothreitol (DTT), beta mercaptoethanol (BME), or Tris (2-carboxyethyl) phosphine (TCEP) at concentrations 1–20 mM can be safely added to protect the thiol group, without reducing the internal disulfide bridges [32]. However, one should be careful while choosing the type of reducing agent. For instance, a protein sample should not be stored in BME for too long as it is highly volatile and susceptible to air oxidation. The oxidized form of BME can react with the reduced cysteines to form disulfide [31]. However, since BME is comparatively cheaper, it can be added during the purification process, but for storage, a more stable agent like TCEP should be used. Alternatively, the oxidation of the reducing agent can also be hindered by addition of chelating agents such as EDTA (ethylene diamine tetra acetic acid) in the concentration range of 10–25 mM [33]. However, one should not use it in combination with the divalent metal ion-dependent separation techniques such as IMAC (immobilized metal ion affinity chromatography) and AIEX (anion exchange chromatography) or with proteins that are dependent on divalent metal ions for their activity.

5.3.1.5 Stabilizing Additives

Several additives can be added to the extraction buffer to stabilize the target protein. The choice of the additive should be made considering its possible effect on the downstream chromatographic separation. Most often the additive is required only in the initial steps of lysate preparation and can be safely excluded in the subsequent steps of the purification. Commonly used additives are sucrose, glucose, glycerol, and other polyols. Glycerol in the concentration range of 5–50% is frequently used to

prevent aggregation and stabilize the protein by increasing protein compactness and decreasing interactions of the hydrophobic surfaces [34]. Polyols such as mannitol, sucrose, propylene, and polyethylene glycol, typically at a concentration of 10%, can be added to stabilize the proteins [35].

5.3.1.6 Nucleases

Although not mandatory, addition of nucleases to the lysis buffer is beneficial in several ways. The nucleic acid released during the cell disruption process might cause aggregation of cell debris and hinder the chromatographic purification either by binding to the chromatographic medium or to the target protein. It might also lead to increase in the viscosity of the sample solution. Most effective engineered nuclease, Benzonase, that breaks down both RNA and DNA can be added in the lysis buffer at a concentration of 1–20 μg/ml before cell disruption [36].

5.3.1.7 Protease Inhibitors

A key threat to protein stability during purification is from the proteases of the source organism. The simplest way to protect against proteolytic degradation is by working quickly at low temperature, suitably on ice. In addition, a mixture of protease inhibitors should be added during purification, especially in the lysis buffer (Table 5.3 provides the list of protease inhibitors) [37]. In some cases, the proteases can be inactivated by adjusting the pH of the solution to a value at which the proteases are inactive, while maintaining the activity of the desired protein.

5.3.2 Clarification of the Extract

The crude extract is often turbid and contains insoluble residues and cellular debris, which might block a chromatography column. Therefore, in a purification protocol, which includes a chromatographic separation, it is important to perform pre-clearing of the crude extract. This could be achieved by centrifugation and/or filtration before applying the sample to the column.

Table 5.3 Protease inhibitors used in protein purification

Protease inhibitor	Molecular weight (Da)	Target protease	Inhibitor type
AEBSF	239.5	Serine proteases	Irreversible
Aprotinin	6611.5	Serine proteases	Reversible
Bestatin	308.38	Amino peptidases	Reversible
EDTA	372.4	Metalloproteases	Irreversible
E-64	357.4	Cysteine proteases	Irreversible
Leupeptin	475.6	Serine and cysteine proteases	Reversible
Pepstatin A	685.9	Aspartic acid proteases	Reversible

Abbreviation: *AEBSF* 4-(2-aminoethyl) benzene sulfonyl fluoride, *E-64* l-*trans*-epoxy succinyl-l-leucylamido-(4-guanido) butane

5.3.2.1 Centrifugation

In a laboratory-scale setup, centrifugation is a common method for clarification of the crude extract. For larger volume of cell lysates, one should use the following setting: 40000 × *g* for 30 min at 4 °C. In case of small sample volumes, a benchtop centrifuge at the highest available g-force, such as 14,000 × *g* for 15–30 min, is sufficient. Important to note here is that the centrifuge should be pre-cooled along with the rotor before placing the samples in it [38].

5.3.2.2 Filtration

Alternative to centrifugation, clarification of the crude extract can also be achieved by filtration, which is relatively less time-consuming than the centrifugation method. The sample to be clarified is passed through a 0.45 μm pore size filter fitted to the syringe or filtration assembly. The 0.45 μm has been recognized as standard pore size membranes that can efficiently remove large bacteria or particulate matter. The common membrane filters with least nonspecific binding include cellulose acetate, polyethersulfone (PES), or polyvinylidene fluoride (PVDF). However, often the filters become saturated after being repeatedly used as they tend to have a certain capacity. Therefore, depending on the sample volume, the filter and syringe capacity should be carefully decided prior to setting up the experiment.

5.4 Checking Solubility and Designing Purification Strategies

5.4.1 Protein Solubility and Precipitation

Protein solubility is a key parameter in any protein purification. The solubility differs markedly from one protein to another, and is highly dependent on the solvent and physicochemical properties of the proteins. The parameters that largely influence protein solubility include solvent's pH, ionic strength, temperature, and the type of exposed amino acid side chains on the proteins' surface. Proteins with lower ratio of the solvent-exposed charged and hydrophobic amino acids tend to be less soluble [39]. Since it is difficult to accurately predict the solubility properties of a protein, one should carefully design pilot studies by varying different conditions and checking protein solubility on SDS-PAGE.

As discussed in Sect. 5.3.2, composition of the extraction medium is critical in protein solubility. For instance, addition of salt such as NaCl, pH (close to protein pI value), and/or inclusion of stabilizing additives should be considered to prevent aggregation and precipitation of the proteins. In cases, where the sample volumes are too large, one can also exploit the protein's tendency of differential precipitation in the presence of neutral salts (ammonium sulfate), polymers (polyethylene glycol), or organic solvents (e.g., ethanol or acetone). This process of purification is termed "salting out," which relies on the principle that at high concentrations of salt, the salt molecules compete with proteins for binding with water, thus leading to protein precipitation [40]. This happens because protein molecules preferentially interact with each other due to energetically favorable protein-protein interactions instead of

protein-solvent interaction. Salting-out offers an alternate means to concentrate the proteins at a very early stage of purification, and with two–three-fold degrees of purity.

5.4.2 Salting-out

Salting-out, which is generally known as salt-induced precipitation or salt fractionation, is based on the interaction of protein with the salt (solute). Salts tend to dissociate in the aqueous solution (solvent), which forms the basis of the salting-out process [41]. In a condition where the salt concentration is increased, the water molecules start getting attracted by the salt ions, which in turn decreases the number of water molecules that can interact with the charged portion of the protein. As a result, protein molecules tend to associate with each other due to stronger protein-protein interaction than the solute-solvent interaction causing protein aggregation and subsequent precipitation. This process is known as salting-out (Fig. 5.1). Importantly, the salt concentration needed to precipitate a protein differs from protein to protein. One of the most common salts used for protein precipitation is ammonium sulfate.

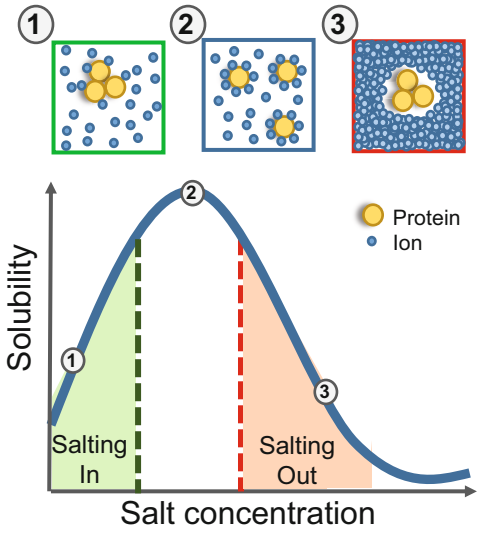

Fig. 5.1 Dependence of protein solubility on salt concentration. A schematic two-dimensional solubility curve of a protein as a function of salt concentration. The solubility curve divides the space in two areas—salting in (*green*) and salting out (*pink*). During the salting in process (phase-1), the salt molecules increase the solubility of the proteins by reducing the ionic interactions between the protein molecules. As the concentration of the salt increases, the excess ions start competing with proteins for the solvent. Gradually, post phase-2, the protein-protein interactions become energetically more favorable than protein-solvent interaction, and the proteins tend to precipitate and come out of the solution (phase-3). This effect is referred to as "salting-out" of the protein

5.4.3 Ammonium Sulfate Precipitation

Ammonium sulfate is the most common salt used for protein precipitation since it is relatively inexpensive, highly soluble in water, and very stabilizing to protein structure. The amount of salt needed for protein precipitation differs from one to another, and also varies with the temperature. Often it is desirable to perform precipitation at low temperatures to avoid protein denaturation. The concentration of ammonium sulfate required to precipitate a particular protein depends on the molecular weight of the protein, solvent pH, temperature, as well as number and position of polar groups [40]. Preferably screening for the percentage saturation of ammonium sulfate to precipitate the target protein or any contaminant should be first established in a small-scale setup. Following is a general protocol for the ammonium sulfate precipitation of a crude lysate [40].

1. Keep the clarified extract on ice and add pre-chilled 50 mM HEPES or Tris–HCl buffer, pH 8.0 to maintain the pH of the sample solution, since addition of ammonium sulfate acidifies the solution.
2. Slowly add fresh, desiccated powder of ammonium sulfate with gentle stirring. It is important to make sure that the salt is completely dissolved before adding more solid, and care should be taken to avoid foaming of the solution. Adding the salt very slowly ensures that local concentration around the addition site does not exceed the required salt concentration.
3. Carefully calculate the amount of ammonium sulfate to be added to attain the chosen percent saturation. Online calculators such as from EnCor Biotechnology Inc. (http://www.encorbio.com/protocols/AM-SO4.htm) can be used to calculate the amount of ammonium sulfate to be added to a specific volume of a solution to reach a particular percentage saturation at a specific temperature.
4. Allow the sample to precipitate for 30 min at 4 °C with continuous stirring.
5. Collect the precipitate by centrifugation at 10,000 to 50,000 × *g* for 30 min–1 h at 4 °C. Note that a solution that is highly saturated in ammonium sulfate is often dense and quite difficult to pellet.
6. Separate the pellet from supernatant and proceed to the next required percent saturation, if supernatant contains the target protein.
7. Resuspend recovered pellets in a volume of buffer that is equal to the volume of the extract. To get rid of excess ammonium ions, dialysis (as described below) can be carried out using the desired buffer before proceeding to the next stage of protein purification or prior to protein storage.

The process of ammonium sulfate precipitation offers an easy and relatively inexpensive way to concentrate and purify proteins. However, it is important to consider the associated drawbacks while selecting this method of separation. Most importantly, it is essential to get rid of the salt from the protein sample, so the downstream processing steps in the form of either dialysis or chromatography will be required.

5.4.4 Salting-in

Protein solubility is affected by the ionic strength of the solution. If a protein is placed in aqueous solution like water, the only ionic component in the solution will be protein molecules. Although water is polar, it only slightly ionizes and therefore proteins tend to aggregate based on the protein-protein ionic interactions [42, 43]. This interaction between the proteins is more favored than protein-water interactions resulting in an irreversible precipitation. At low ionic concentrations of salts such as NaCl, the presence of other ionic species now can compete with the ionic protein-protein interactions. These ions in the solution tend to shield the protein molecules from the charge of the other protein molecules. The decrease in the electrostatic interaction between the protein molecules eventually increases the solubility of the protein, referred to as "salting-in" (Fig. 5.1). However, at a point when the ionic strength starts getting too high, it imparts a negative effect on the protein solubility, resulting in "salting-out," as discussed in the previous section.

Salting-in of the proteins occurs generally near its isoelectric point (pI) [13]. In addition to the electrostatic effect, the limited charge on the surface of the protein affects the water associated with the protein [44]. All in all, the pairing of salt ions with the charged groups on protein molecules increases protein solubility resulting in "salting-in" of the proteins.

5.4.5 Dealing with Proteins in the Inclusion Bodies

Often high expression of the recombinant proteins in *E. coli* results in the formation of insoluble and aggregated proteins referred to as "inclusion bodies." In addition to being insoluble, the inclusion bodies have a non-native structure and therefore require solubilization and refolding of the target protein to its native structure [30]. This generally requires a lot of optimization and is also time-consuming. In addition, one is expected to face significant loss in the protein amount during the process of refolding. The formation of the inclusion bodies in *E. coli* can be avoided by modifying several conditions during protein expression. Most commonly, use of lower temperature such as 18 °C and < 0.5 mM IPTG (isopropyl thiogalactoside) can be used to tackle the problem of inclusion body formation [2]. In the presence of higher concentrations of IPTG and/or at high temperatures (37 °C), the protein expression takes place at a high translational rate, which often exhausts the protein quality control system of *E. coli* resulting in partially folded or misfolded protein aggregates in the form of inclusion bodies [45]. However, in certain conditions, wherein expression of large amounts of proteins would otherwise be toxic to the host cell, expression in the form of inclusion bodies allows its large-scale preparation. In addition, the sequestration of proteins in the inclusion bodies prevents it from proteolysis by the cytosolic proteases [30].

For the isolation of inclusion bodies, the cells expressing the target protein are harvested, lysed by mechanical lysis, and then centrifuged at high speed ($20{,}000 \times g$), 15 min at 4 °C. The inclusion bodies are obtained as a pellet. The

pellet is then washed with detergents such as 1% Triton-X, followed by denaturation of the protein using chemical denaturants, viz.8 M Urea or 6 M Guanidine-Hydrochloride [30]. The denaturation step can be performed multiple times to achieve maximum recovery. Often, more than 90% purity is obtained by following these steps. In case, the target protein contains impurities, it can be subjected to affinity purification or salting-out precipitation to increase its purity. The denatured protein is then refolded by gradual removal of the denaturant either using dilution or dialysis method. In the dilution method, a drop-by-drop solution of the denatured protein is added in a buffer solution (100× the volume of the denatured protein) [46]. In the case of dialysis, the denatured protein is placed in the dialysis tubing of a particular molecular weight cutoff (depending on the size of the protein), and is suspended in the refolding buffer solution. Some of the critical parameters during refolding include temperature, pH, presence of reducing agents (such as DTT, TCEP), and additives (often in combination) [47]. However, most importantly, screening of multiple conditions and subsequent optimization is required for successful refolding. Often, the success rate for refolding of proteins is not very encouraging. Therefore, a large number of refolding conditions should be tested such that a biologically active form of the protein is obtained in large amounts at the highest purity level. Apart from dilution or dialysis methods, the denaturing agents in the unfolded protein solution can also be removed using different chromatographic techniques. Here, the protein is either slowly allowed to migrate through a column (e.g., gel filtration on a Superdex 75 column) or is bound to the matrix (e.g., affinity chromatography using Ni-NTA agarose) and then eluted in a buffer with decreasing concentration of the denaturant.

5.5 Overview of Chromatography

Column chromatography, most commonly liquid chromatography (LC), is used in purification of the recombinant proteins. In this method of chromatography, the stationary phase is packed in a column and the liquid mobile phase is allowed to pass through the column using a pump or under gravity flow. The sample mixture is introduced at one end, followed by elution with the mobile phase at the other end of the column. The separation of the components in the mixture depends on partitioning of the molecules between the mobile and the stationary phases, which is based on the differences in their molecular weights. For recombinant protein purification, affinity chromatography, ion-exchange chromatography, and size exclusion chromatography are commonly applied to achieve satisfactory purity and homogeneity of the target protein [48].

5.5.1 Affinity Chromatography

There are numerous ways in which the affinity-based method is employed for purification of the recombinant proteins. The most common strategy for

Table 5.4 The list of fusion tags commonly selected for purification of recombinant proteins [14]

Tag	Size (no. of amino acids or kDa)	Ligand	Separation method
Polyhistidine	5–15 a.a.	Ni^{+2} or Cu^{+2}	AC
Glutathione S-transferase	26 kDa	Glutathione	AC
Maltose binding domain	40 kDa	Amylose	AC
FLAG	8 a.a.	mAb based	AC
Strep tag I	9 a.a.	Streptavidin	AC
Strep tag II	8 a.a.	Streptactin	AC
T7-tag	11–16 a.a	mAb based	AC
c-myc	10 a.a.	mAb based	AC
S-tag peptide	15 a.a.	S-protein	AC
Polyaspartic acid	5–16 a.a.	Ion-exchange or precipitation	IEX
VSV tag	11 a.a.	mAb based	AC
Calmodulin binding peptide	26 a.a.	Calmodulin	AC

AC Affinity chromatography, *IEX* Ion-exchange chromatography, *a.a.* Amino acid, *kDa* kilodalton

affinity-based purification involves use of fusion tags (i.e., amino acid sequences attached to the recombinant proteins) that have affinity for ligands immobilized on the column. Some of the commonly used tags are outlined in Table 5.4.

Particularly, the histidine (His) tag (a sequence of six or more histidine residues) is added to either N- or C-terminus of the recombinant protein, and is frequently used for purification of the target protein from a mixture of proteins in the cellular lysate. The His-tagged protein shows affinity toward divalent metal ions such as Ni^{+2} or Zn^{+2}, and therefore the chromatographic technique is known as immobilized metal affinity chromatography (IMAC). IMAC uses matrices such as iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA) to chelate transition metals through three or four coordination sites, respectively [49]. Since the metal ion is only weakly bound via three-coordination sites in IDA compared to four sites in case of NTA, often metal leaching from the IDA matrix occurs during purification resulting in lower yield and impure proteins. Therefore, advancements in NTA-based chemistry have facilitated development of matrices, which securely coordinate metal ions with four-coordination sites while leaving two of the transition metal coordination sites exposed to interact with the His-tag [49]. Due to the presence of the electron donor groups on histidine, the His-tag promotes strong interaction with the immobilized transition metal and gets retained on the IMAC column. After subsequent washes of the matrix to remove nonspecifically bound proteins, the His-tagged recombinant protein gets eluted either by adjusting the pH of the elution buffer or by adding histidine analog, imidazole (concentration of 100–500 mM) [50]. At lower pH (4.5–5.3), the imidazole nitrogen atom of the histidine residue (pKa 6.0) gets protonated that disturbs the coordination bond between histidine and

the transition metal [49], while imidazole at concentrations >100 mM acts as a competitor resulting in the elution of the bound his-tagged protein. The major advantage of the IMAC system is that it can tolerate a wide range of buffer conditions, including the presence of additives like detergent and chemical denaturants. In addition, these resins can be regenerated and reused several times, thus enabling development of economical purification strategies in both academia and industrial settings. A schematic overview of Ni-NTA affinity chromatography is provided in Fig. 5.2.

One of the major disadvantages of the polyhistidine tag is the nonspecific binding of the proteins to the IMAC column. Nevertheless, the His-tag offers several advantages due to its small size, which rarely affects protein function. Also in most cases, one can achieve purity up to 90–95% in a single step of purification. The IMAC resin remains unaffected by protease or nuclease activities in the cell lysate making it suitable for purification with crude lysates. One of the advantages of the His-tag is that it can be combined with other affinity tags (listed in Table 5.4) to the same protein to provide great flexibility during the purification process. Overall, IMAC offers a rapid and an inexpensive purification method compared to other affinity-based purification methods.

Similar to polyhistidine tag, alternative affinity tags such as MBP or GST are also used frequently; however, due to large size, these tags need to be removed via a proteolytic cleavage at specific sites introduced between the tag and target protein. Due to the additional processing step involved in the large affinity tags, the use of these tags increases the downstream processing cost for the purification of the recombinant proteins [14, 51].

5.5.2 Ion Exchange Chromatography (IEX)

IEX is based on the electrostatic interactions between the charged groups of the protein with the matrix of the column. Separation of the proteins in IEX depends on surface charge of the proteins, pH, and salt concentration of the mobile phase [52]. IEX is most frequently applied for purification of recombinant proteins that are untagged. Since proteins are amphoteric molecules (act as both acids and bases), any protein can bind the ion-exchange matrix once the pH of the solution is optimized for its binding. The chosen matrix has an ion-load opposite to that of the protein to be separated. A positively charged ion-exchange matrix (i.e., anion-exchanger) adsorbs negatively charged proteins, while a negatively charged matrix (cation-exchanger) adsorbs the positively charged molecules. As a rule, the selected mobile phase should have low to medium salt concentration (i.e., low to medium conductivity). Moreover the chosen pH should lie between the isoelectric point (pI) *aka* acid dissociation constants (pKa) of the charged molecule, and that of the ion on the matrix [53]. For example, in cation-exchange chromatography, where the pKa of matrix is 1.2, the mobile phase of pH 6.0 should be ideally used for a protein with a pI of 8.2. On the other hand, in anion-exchange chromatography where the

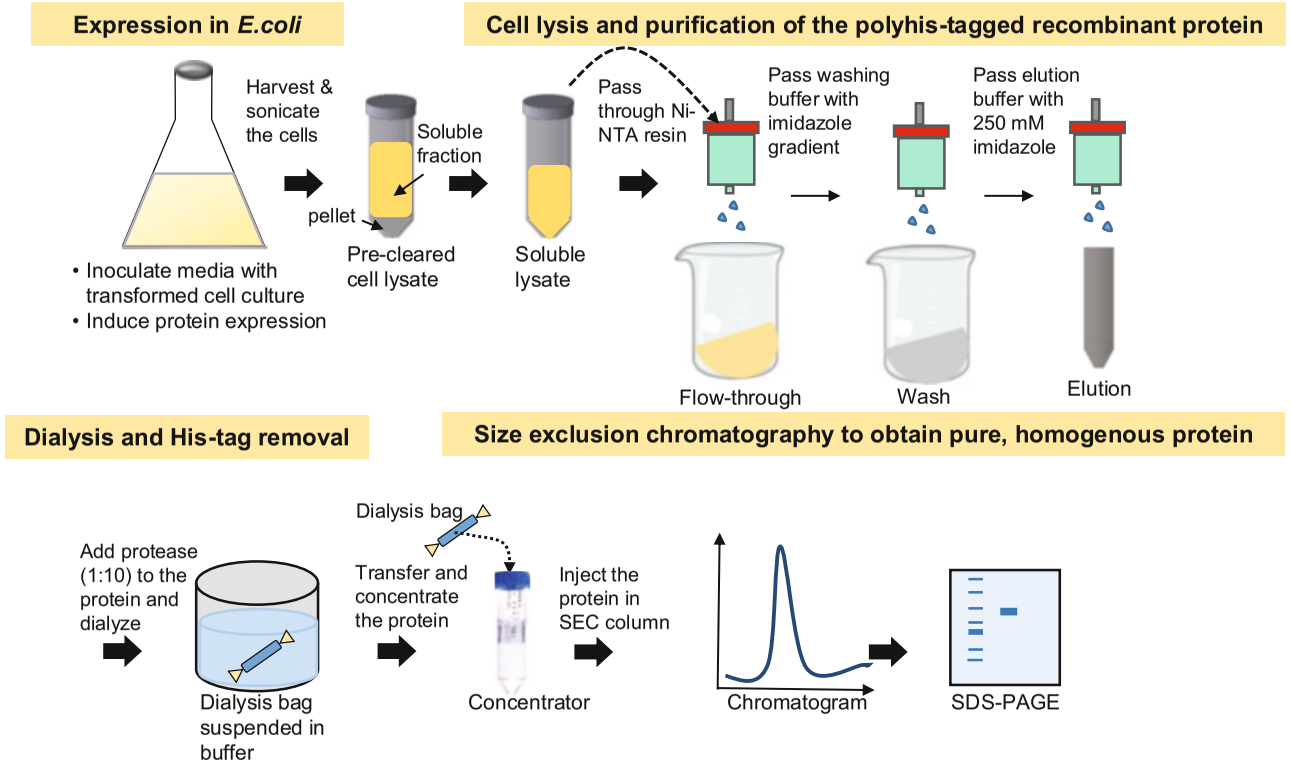

Fig. 5.2 A schematic representation for polyhis-tagged protein purification using Ni-NTA affinity chromatography. The 6X-His-tagged recombinant protein is expressed in the host, such as *E. coli*. Cells over-expressing the protein are harvested, resuspended in the lysis buffer followed by cell lysis using sonication. To remove the cellular debris, the cell lysate is centrifuged at high speed to separate the soluble fraction and pellet. The pre-cleared lysate is then passed through a Ni-NTA affinity column, and the flow-through is collected. The column is given several washes with washing buffer containing low concentrations of imidazole. The desired protein is then eluted using ≥250 mM imidazole buffer. The purified protein is treated with site-specific proteases, such as thrombin or TEV

Fig. 5.2 (continued) protease to remove the His-tag, and then dialyzed overnight in the desired buffer (optional step). In order to obtain highly pure and homogenous form of the recombinant protein, the dialyzed protein is concentrated using a concentrator (such as Centricon), and then subjected to size exclusion chromatography (SEC). The eluted fractions are analyzed on SDS-PAGE to check for purity and confirmation of the molecular weight

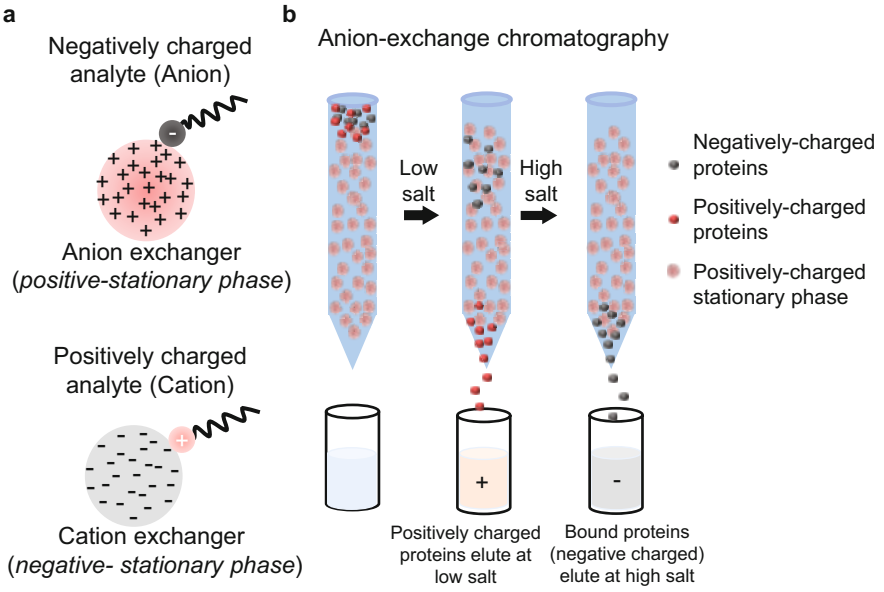

Fig. 5.3 Schematic representation of ion-exchange chromatography. (**a**) Anion and cation exchanger bind to negatively and positively charged molecules, respectively. (**b**) Flow-scheme of anion-exchange chromatography. The positively charged stationary phase of the anion-exchanger allow binding of negatively charged proteins at low ionic concentrations of the mobile phase. In the presence of low salt, the positively charged molecule in the protein mixture elutes first. As the concentration of the salt increases, the bound negatively charged proteins are exchanged by the salt ions and gets eluted from the column

matrix pKa is 10.3, a mobile phase of pH 8.0 should be used for protein molecules with a pI of 6.8.

Binding of the protein onto the column depends on the surface charge of the protein and charge on the matrix (Fig. 5.3a). This interaction, which is reversible, is then disrupted using a linear gradient of salt or varying pH to elute the bound proteins [54]. For instance, the negatively charged proteins can be displaced using negatively charged salt ions. This is because the negatively charged salt ions competitively displace the negatively charged proteins from the functional groups of the matrix. The elution profile of the protein therefore follows a low-to-high salt concentration gradient (Fig. 5.3b). Alternatively, varying pH can also be used to separate the proteins. In cation-exchange chromatography, increasing the buffer pH makes the protein less protonated (i.e., less positively charged), thus disrupting its ionic interaction with the negatively charged matrix and subsequent elution. While in the anion-exchange chromatography, decrease in the pH of the mobile phase results in protonation of the protein (i.e., more positively charged) which promotes protein elution [48].

Under mild conditions, IEX provides high binding capacity and resolution for separation of recombinant proteins. In addition, its ability to scale-up (particularly

for recombinant proteins), moderate cost, and broad applicability have led it to become one of the most widely used and versatile liquid chromatography techniques.

5.5.3 Size Exclusion Chromatography

Size exclusion chromatography (SEC), also known as gel filtration chromatography (GFC), separates the macromolecules based on differences in their hydrodynamic volume and molecular weight [55]. The stationary phase consists of an inert spherical bead or gel with pores of a specific size distribution. Proteins larger than the pore size of the gel cannot permeate into the gel particles, thereby eluting first by rapidly passing through the space in-between the beads. On the other hand, proteins smaller than the gel pore size get diffused into the pores and elute at proportionally longer retention times (i.e., the time taken from injection to detection of the protein) [56]. Hence, SEC is also commonly referred to as gel permeation chromatography. Unlike affinity chromatography or IEX, in SEC, the protein molecules do not directly interact with the mobile phase, so the buffer composition does not affect the column resolution (i.e., the degree of separation of the peaks). The retention time of larger proteins are shorter than the smaller proteins, which enables separation of these proteins [54, 57]. Apart from the gel pore size, the column resolution is also influenced by bed height, flow rate, volume of the sample, and the molecular weight of the protein [58–60]. Generally, the highest possible resolution can be obtained with slow to moderate flow rate, long and narrow columns, small pore-size gel and sample volumes (1–5% of the total column volume). A general layout of the size exclusion chromatography is given in Fig. 5.4.

SEC is widely used during the final polishing steps in recombinant protein purification due to its excellent desalting properties. This is usually done when the volume of the sample has been reduced and the major goal remains is to remove aggregates or change the protein buffer (also known as desalting). For desalting purpose, since the difference in the molecular mass of the protein and salt is typically very large, the peaks are very less likely to overlap. Therefore, even sample volumes as much as 30% of the bed volume can be safely applied to the column. In addition to the purification purpose, SEC can also be used for analytical use. With proper molecular weight standards, one can calculate the molecular weight of the unknown protein [61]. In this method, a few proteins of known molecular weights are initially used for calibrating the column. The calibration standard also contains a very large molecular weight protein (such as blue dextran (1000 kDa) to determine the void volume (V_o) of the column, which refers to the excluded volume, i.e., the space between the particles. One can then determine the elution volume (V_e) of the individual standards to determine the molecular weight of the unknown protein. A standard plot can be generated by dividing "V_e" of the standards by the "V_0" (V_e/V_0), and plotting this value against the log of the molecular weights (log M) of the standard. The molecular weight of the unknown protein can then be extrapolated from the standard plot. However, care should be taken while interpreting the data, as

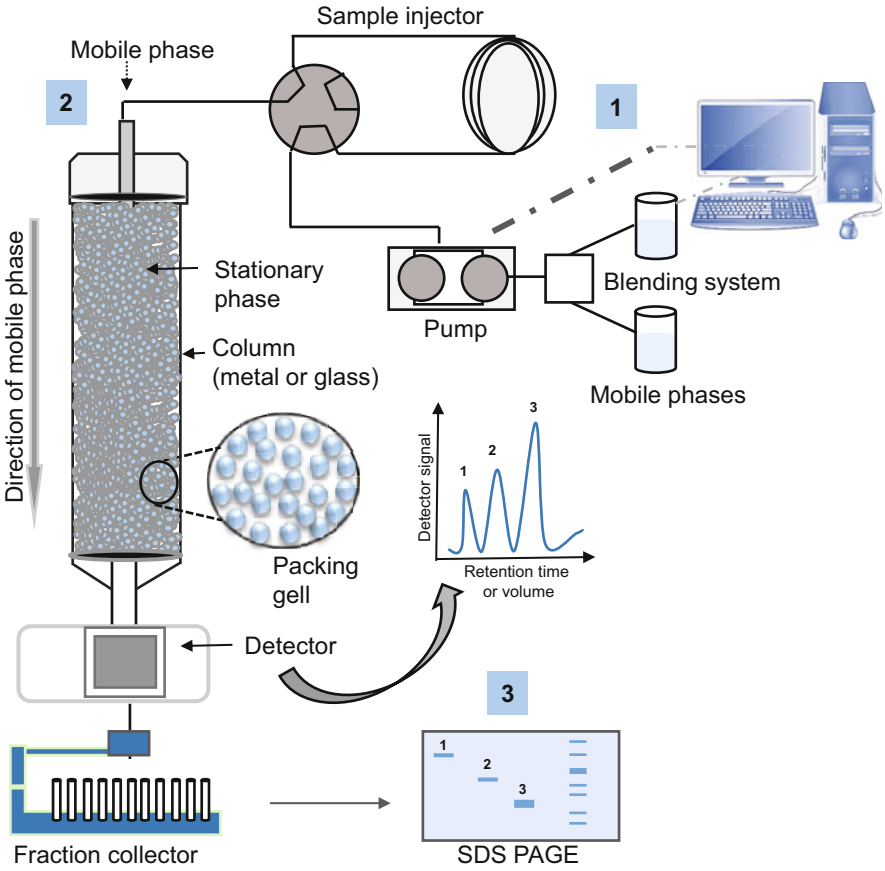

Fig. 5.4 Schematic overview of size exclusion chromatography (SEC). SEC or gel filtration chromatography mostly involves an automated system constituting a pump, sample injector, column, and fraction collector coordinated through a software. With the help of a pump, the mobile phase is passed through the column at a particular flow rate. The protein mixture is injected through a sample injector, and separation is achieved based on the pore size of the packing gel and proteins molecular weight. Protein with higher molecular weight elutes first followed by the smaller proteins. The elution of the protein is monitored in real time in the form of chromatogram, generated by the detector. The eluted fractions are collected either manually or with the help of fraction collector. The collected fractions are further analyzed on SDS-PAGE for purity and confirmation of the molecular weight of the separated proteins

SEC is quite accurate for globular proteins, while it is less accurate for flattened or extended protein molecules.

5.5.4 Concluding Remarks

The bottleneck in the production of recombinant proteins is the cost of purification of the protein. Therefore, over the past several decades, tremendous effort has been directed toward development of new purification methods as well as improvement in the existing strategies. Here, we present general steps for the purification of recombinant proteins that are expressed in a bacterial system such as *E. coli*, which is one of the most preferred microbial cell factories. Although it is suitable for stably expressing folded, globular proteins, difficulties are often encountered in expression and purification of membrane or membrane-associated proteins. In this chapter, we describe the possible routes to meet those challenges of expression and purification of recombinant proteins in *E. coli*. Although many of these outlined approaches might fail at several stages, one has to find ways to standardize their protocols and do extensive troubleshooting to overcome those obstacles since expression and purification of recombinant proteins are often protein specific.

Acknowledgments NS gratefully acknowledges financial support from the DBT-RA Program in Biotechnology and Life Sciences for completion of the manuscript. The authors acknowledge Ms. Chanda Baisane, Bose Lab, ACTREC for formatting the manuscript.

Problems

Multiple Choice Questions

1. Protein purification techniques are based on the following properties except:
   (a) Solubility of the protein
   (b) Charge on the protein
   (c) Viscosity of the protein
   (d) Specific binding affinity of the protein
2. Salting-out refers to
   (a) Precipitation of proteins using ammonium sulfate
   (b) Precipitation of proteins using copper sulfate
   (c) Precipitation of proteins using sodium chloride
   (d) Both (a) and (c)
3. You find that your protein sample shows lots of additional bands of lower molecular weight apart from the desired protein. What can you do about this?
   (a) Add an additional purification step
   (b) Use a protease inhibitor during lysis and purification
   (c) Perform each step as quickly as possible, in a cold-room
   (d) All of the above

Subjective Questions

1. To estimate the molecular mass of an unknown protein, you decide to run a size exclusion chromatography. Next, you run a series of proteins with the known molecular mass and the unknown protein on a Sephadex G-200 column. Below are the elution volumes (V_e) for each protein. The measured void volume (V_0) of the column is 36 mL. Using these values, calculate the molecular mass of the unknown protein.

Protein	Molecular weight (kDa)	V_e (mL)
Lysozyme	14	100
Ovalbumin	45	79
Serum albumin	66	70
Aldolase	150	51
Urease	489	26
Unknown	?	88

2. To separate a mixture of proteins with different pI values, you try anion exchange chromatography using DEAE-cellulose column. For this, you first equilibrate your column with phosphate buffer, pH 6.5. Then you pass the following mixture of proteins, AS (pI – 4.6), BS (pI – 5.0), and CS (pI – 7.0). The proteins are eluted first with weak ionic strength buffer (50 mM NaCl, pH 6.5) and then in the same buffer but with increasing NaCl concentration. Explain what order will the proteins elute?

References

1. Sumner JB. The isolation and crystallization of the enzyme urease preliminary paper. J Biol Chem. 1926;69:435–41.
2. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.
3. Structural Genomics, C., C. China Structural Genomics, C. Northeast Structural Genomics, Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schutz A, Heinemann U, Yokoyama S, Bussow K, Gunsalus KC. Protein production and purification. Nat Methods. 2008;5:135–46.
4. Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif. 2006;48:1–13.

5. Tripathi NK, Shrivastava A. Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Front Bioeng Biotechnol. 2019;7:420.
6. Wingfield PT. Overview of the purification of recombinant proteins. Curr Protoc Protein Sci. 2015;80:6.1.1–6.1.35.
7. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci U S A. 2009;106:1760–5.
8. Cole JL, Lary JW, Moody TP, Laue TM. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. Methods Cell Biol. 2008;84:143–79.
9. Park ES, Wallace WE, Guttman CM, Flynn KM, Richardson MC, Holmes GA. A general method for quantitative measurement of molecular mass distribution by mass spectrometry. J Am Soc Mass Spectrom. 2009;20:1638–44.
10. Hong P, Koza S, Bouvier ES. Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J Liq Chromatogr Relat Technol. 2012;35:2923–50.
11. Audain E, Ramos Y, Hermjakob H, Flower DR, Perez-Riverol Y. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. Bioinformatics. 2016;32:821–7.
12. Ahamed T, Chilamkurthi S, Nfor BK, Verhaert PD, van Dedem GW, van der Wielen LA, Eppink MH, van de Sandt EJ, Ottens M. Selection of pH-related parameters in ion-exchange chromatography using pH-gradient operations. J Chromatogr A. 2008;1194:22–9.
13. Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci. 2001;10:1206–15.
14. Saraswat M, Musante L, Ravida A, Shortt B, Byrne B, Holthofer H. Preparative purification of recombinant proteins: current status and future trends. Biomed Res Int. 2013;2013:312709.
15. Lakowicz JR. Principles of fluorescence spectroscopy. Switzerland: Springer Science & Business Media; 2013.
16. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. New York: Springer; 2005. p. 571–607.
17. Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989;182:319–26.
18. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng Des Sel. 1990a;4:155–61.
19. Varshavsky A. The N-end rule pathway of protein degradation. Genes Cells. 1997;2:13–28.
20. Mogk A, Schmidt R, Bukau B. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 2007;17:165–72.
21. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 1990b;4:155–61.
22. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157:105–32.
23. Waegeman H, Soetaert W. Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. J Ind Microbiol Biotechnol. 2011;38:1891–910.
24. Vieira Gomes AM, Souza Carmo T, Silva Carvalho L, Mendonca Bahia F, Parachin NS. Comparison of yeasts as hosts for recombinant protein production. Microorganisms. 2018;6
25. Wurm DJ, Slouka C, Bosilj T, Herwig C, Spadiut O. How to trigger periplasmic release in recombinant Escherichia coli: a comparative analysis. Eng Life Sci. 2017;17:215–22.
26. Brady PN, Macnaughtan MA. Evaluation of colorimetric assays for analyzing reductively methylated proteins: biases and mechanistic insights. Anal Biochem. 2015;491:43–51.
27. Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta. 2004;1666:105–17.

28. Ersson B, Rydén L, Janson JCJPPP. High resolution methods, and applications. In: Introduction to protein purification. Hoboken, NJ: John Wiley & Sons, Inc.; 2011. p. 1–22.
29. Otzen DE. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. Biophys J. 2002;83:2219–30.
30. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Factories. 2015;14:41.
31. Poole LB. The basics of thiols and cysteines in redox biology and chemistry. Free Radic Biol Med. 2015;80:148–57.
32. Lebendiker M, Maes M, Friedler A. A screening methodology for purifying proteins with aggregation problems. Methods Mol Biol. 2015;1258:261–81.
33. Trivedi MV, Laurence JS, Siahaan TJ. The role of thiols and disulfides on protein stability. Curr Protein Pept Sci. 2009;10:614–25.
34. Middelberg AP. Preparative protein refolding. Trends Biotechnol. 2002;20:437–43.
35. Singh S, Singh J. Effect of polyols on the conformational stability and biological activity of a model protein lysozyme. AAPS PharmSciTech. 2003;4:E42.
36. Lin MZ, Tsien RY. TimeSTAMP tagging of newly synthesized proteins. Curr Protoc Protein Sci. Chapter 26. 2010;Unit 26:25.
37. Wingfield PT. Preparation of soluble proteins from Escherichia coli. Curr Protoc Protein Sci. 2014;78:6.2.1–6.2.22.
38. Cvjetkovic A, Lotvall J, Lasser C. The influence of rotor type and centrifugation time on the yield and purity of extracellular vesicles. J Extracell Vesicles. 2014;3
39. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Hierarchical structure of proteins. In: Molecular cell biology. 4th ed. Bethesda, MD: WH Freeman; 2000.
40. Wingfield P. Protein precipitation using ammonium sulfate. Curr Protoc Protein Sci. 2001; Appendix 3:Appendix 3F.
41. Hyde AM, Zultanski SL, Waldman JH, Zhong Y-L, Shevlin M, Peng F. General principles and strategies for salting-out informed by the hofmeister series. Org Process Res Dev. 2017;21:1355–70.
42. Damodaran S. Amino acids, peptides and proteins. Boca Raton, FL: CRC Press; 2008.
43. Mattos C. Protein–water interactions in a dynamic world. Trends Biochem Sci. 2002;27:203–8.
44. Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. Biophys J. 2012;102:1907–15.
45. Carrio MM, Villaverde A. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J Bacteriol. 2005;187:3599–601.
46. Yamaguchi H, Miyazaki M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomol Ther. 2014;4:235–51.
47. Willis MS, Hogan JK, Prabhakar P, Liu X, Tsai K, Wei Y, Fox T. Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions. Protein Sci. 2005;14:1818–26.
48. Coskun O. Separation techniques: chromatography. North Clin Istanb. 2016;3:156–60.
49. Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 2000;326:245–54.
50. Porath J. Immobilized metal ion affinity chromatography. Protein Expr Purif. 1992;3:263–81.
51. Mahmoodi S, Pourhassan-Moghaddam M, Wood DW, Majdi H, Zarghami NJCB. Current affinity approaches for purification of recombinant proteins. Cogent Biol. 2019;5(1):1665406.
52. Jungbauer A, Hahn R. Ion-exchange chromatography. In: Methods in enzymology, vol. 463. Netherland: Elsevier; 2009. p. 349–71.
53. Hedhammar M, Karlström AE, Hober SJS. Chromatographic methods for protein purification. Seattle, WA: Semantic Scholar; 2006. p. 1–31.
54. Mori S, Barth HG. Size exclusion chromatography. Singapore: Springer Science & Business Media; 2013.

55. Sun T, Chance RR, Graessley WW, Lohse DJ. A study of the separation principle in size exclusion chromatography. Macromolecules. 2004;37:4304–12.
56. Fekete S, Beck A, Veuthey J-L, Guillarme D. Theory and practice of size exclusion chromatography for the analysis of protein aggregates. J Pharm Biomed Anal. 2014;101:161–73.
57. Burgess RR. A brief practical review of size exclusion chromatography: rules of thumb, limitations, and troubleshooting. Protein Expr Purif. 2018;150:81–5.
58. Afeyan N, Gordon N, Mazsaroff I, Varady L, Fulton S, Yang Y, Regnier FJ. Flow-through particles for the high-performance liquid chromatographic separation of biomolecules: perfusion chromatography. J Chromatogr A. 1990;519:1–29.
59. Granath KA, Kvist BE. Molecular weight distribution analysis by gel chromatography on sephadex. J Chromatogr A. 1967;28:69–81.
60. Striegel A, Yau WW, Kirkland JJ, Bly DD. Modern size-exclusion liquid chromatography: practice of gel permeation and gel filtration chromatography. Hoboken, NJ: John Wiley & Sons; 2009.
61. Erickson HP. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol Proced Online. 2009;11:32.