1
Protein Purification

Richard R. Burgess

1.1
Introduction

The magnitude of the challenge of protein purification becomes clearer when one considers the mixture of macromolecules present in a cell extract. In addition to the protein of interest, several thousand other proteins with different properties are present in the extract, along with nucleic acids (DNA and RNA), polysaccharides, lipids, and small molecules. The proteins present in the bacterium *Escherichia coli* may be dramatically visualized after resolution by two-dimensional gel electrophoresis as shown in Figure 1.1. A given protein may be present at more than 10% or at less than 0.001% of the total protein in the cell. Enzymes are found in different states and locations: soluble, insoluble, membrane-bound, DNA-bound, in organelles, cytoplasmic, periplasmic, and nuclear. The challenge, therefore, is to separate the protein of interest from all of the other components in the cell, especially the unwanted contaminating proteins, with reasonable efficiency, speed, yield, and purity, while retaining the biological activity and chemical integrity of the polypeptide.

Sections 1.1–1.6 provide background on classical protein fractionation and purification; that is, the isolation of a protein from its natural source. Sections 1.7 and 1.8 give a brief introduction to overproduction and purification of recombinant proteins cloned and overexpressed in a bacterial host expression system.

1.2
Types of Molecular Interactions and Variables that Affect Them

With regard to protein structure and stability and the interaction between an individual protein and other proteins, DNA, or materials used in protein purification, one must understand the molecular forces involved and how the strength of these forces varies as one varies conditions such as temperature, pH, and ionic strength of a solution. The atomic interactions that seem to be

Proteomics of the Nervous System. Edited by H.G. Nothwang and S.E. Pfeiffer

ISBN: 978-3-527-31716-5

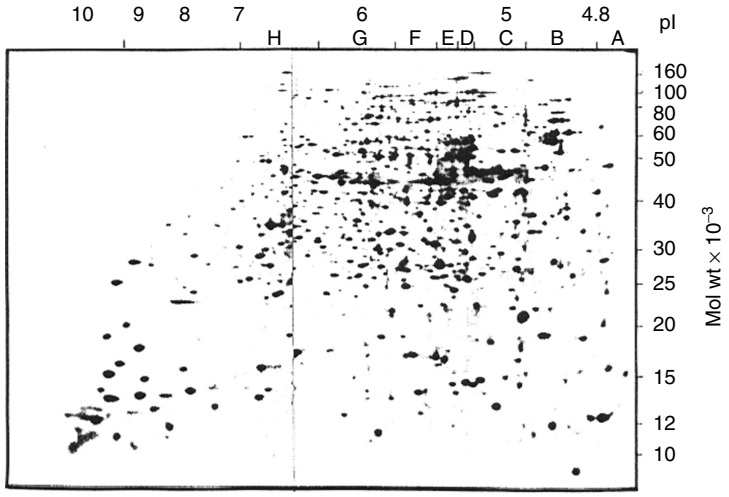

Figure 1.1 *E. coli* proteins resolved on a two-dimensional gel. The approximate isoelectric point and molecular weight scales are indicated. *E. coli* K12 strain W3110 was labeled with $^{35}SO_4$ during growth in glucose minimal medium at 37 °C. A composite autoradiogram was made from non-equilibrium (left side) and pH 5–7 (right side) isoelectric focusing gels. (Adapted, with permission, from Neidhardt and Phillips (1985).)

the most important with regard to protein interactions are hydrogen bonds, hydrophobic interactions, and ionic interactions. These are described briefly below (Also see Creighton, 1993).

1.2.1 Hydrogen Bonds

Hydrogen bonds (Figure 1.2) occur when a proton is shared between a proton donor (NH and OH) and a proton acceptor (OC and N). Optimal hydrogen bonds have a linear geometry and a distance between the donor and acceptor atoms between 2.6 and 3.1 Å. Hydrogen bonds are stronger at low temperature and are weakened as the temperature is raised.

1.2.2 Hydrophobic Interactions

Non-polar residues (isoleucine, leucine, valine, phenylalanine, and tryptophan) cannot make favorable hydrogen bonds with water. In order to avoid water, they tend to come together in a so-called hydrophobic interaction (see Figure 1.2), usually resulting in their being buried in the interior of a protein. Hydrophobic interactions are strengthened at high salt levels and high temperature.

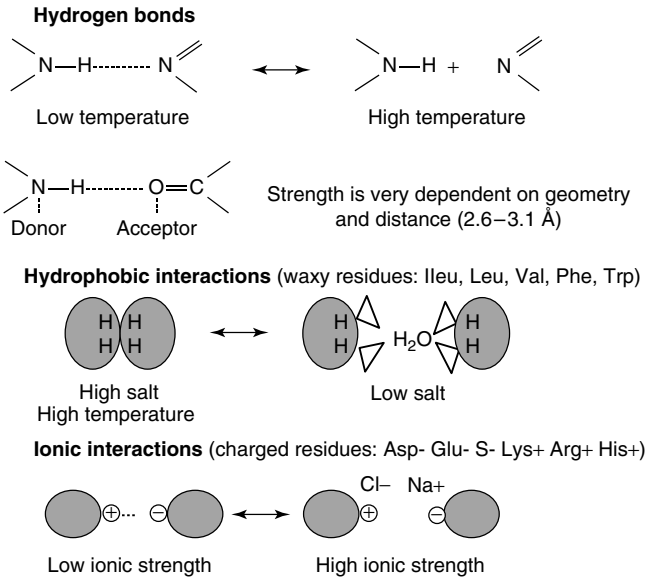

Figure 1.2 Hydrogen bonds, hydrophobic interactions, ionic interactions. (Please find a color version of this figure in the color plates.)

1.2.3 Ionic Interactions

Ionic interactions (see Figure 1.2) occur between charged molecules, with like charges repelling and opposite charges attracting. The force of the electrostatic interaction is given by an approximation of Coulomb's law, $E = Z_A Z_B e^2 / D r_{AB}$, where r_{AB} is the distance between two charges, A and B, Z_A, and Z_B are their respective number of unit charges, e is one unit of electronic charge, and D is the dielectric constant of the solvent. The strength of ionic interactions is therefore inversely proportional to the distance between the charges and the dielectric constant of the solvent, which varies from 2 in non-polar solvents like hexane to 80 in highly polar solvents such as water. Ionic interactions are weakened as the ionic strength of the solvent increases and the charge is shielded by counterions. Ionic interactions are affected by the pH of the solution, since pH determines the number of charged residues.

1.2.4 Variables that Affect Molecular Forces

Conditions can be varied to affect the relative strength of the above molecular forces. Temperature, ionic strength, ion type, dielectric constant, and pH, for example, can all be varied. In a few cases, researchers have also varied pressure.

1.3
Protein Properties that can be Used as Handles for Purification

A single protein can be purified from a mixture of thousands of proteins because proteins vary tremendously in a number of their physical and chemical properties. These properties are the result of proteins having different numbers and sequences of amino acids. The amino acid residues attached to the polypeptide backbone may be positively or negatively charged, neutral and polar, or neutral and hydrophobic. In addition, the polypeptide is folded in a very definite secondary structure (α-helices, β-sheets, and various turns) and tertiary structure to create a unique size, shape, and distribution of residues on the surface of the protein. By exploiting the differences in properties between the protein of interest and other proteins in the mixture, a rational series of fractionation steps can be designed. These properties include: size, shape, charge, isoelectric point, charge distribution, hydrophobicity, solubility, density, ligand binding, metal binding, reversible association, posttranslational modifications and specific sequences or structures

1.3.1
Size

Proteins may vary in size from peptides of a few amino acids (with molecular weights of a few hundred) to very large proteins containing over 10 000 amino acids (with molecular weights of over 1 000 000). Most proteins have molecular weights in the range 10 000–150 000 (see Figure 1.1). Proteins that are part of multi-subunit complexes may reach much larger sizes. Proteins are often fractionated on the basis of size (really on the basis of effective radius or Stokes' radius) by passing down a gel-filtration column (or size-exclusion column, SEC). The column is filled with porous beads with characteristic pore sizes. The largest proteins cannot penetrate into the bead and are excluded and elute first in what is called the void or excluded volume. Very small proteins and salts easily pass in and out of the beads and see the entire volume of the column (the column volume). Other intermediate-sized proteins elute between the void and the column volume based on how much time they spend outside and inside the beads.

1.3.2
Shape

Protein shapes range from approximately spherical (globular) to quite asymmetric. The shape of a protein influences its movement through a solution during centrifugation, through small pores in membranes, into beads during gel filtration, or through gels during electrophoresis. For example, consider two monomeric proteins of the same mass where one is spherical and the

other is cigar shaped. During centrifugation through a glycerol gradient, the spherical protein will have a smaller Stokes' radius and, thus, will encounter less friction as it sediments through the solution. It will sediment faster, and thus, appear to be larger than the cigar-shaped protein. On the other hand, during size-exclusion chromatography, the same spherical protein with its smaller Stokes' radius will more readily diffuse into the pores of a gel-filtration bead and will elute later, thus appearing smaller than the cigar-shaped protein.

1.3.3 Charge

The net charge of a protein is determined by the sum of the positively and negatively charged amino acid residues. If a protein has a preponderance of aspartic and glutamic acid residues, it has a net negative charge at pH 7 and is termed an acidic protein. If it has a preponderance of lysine and arginine residues, it is considered to be a basic protein. The equilibrium between charged and uncharged groups and hence the charge of a protein is determined by the pH of the solution. The charge of the ionizable groups found on unmodified proteins as a function of pH is shown in Table 1.1. In general, a positively charged resin (an anion-exchange column) is used to bind a negatively charged protein and a negatively charged resin (a cation-exchange column) to bind a positively charged protein. The protein is bound to the column at low salt (e.g. 0.1 M NaCl) and eluted with an increasing salt gradient. At some stage, the ionic attraction of the protein to the column resin will become weak enough to cause the protein to dissociate from the column and elute.

Table 1.1 The charge of the ionizable groups found on unmodified proteins as a function of pH.

Ionizable group	pK_a[a]	pH 2					pH 7															pH 12
C-terminal (COOH)	4.0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Aspartate (COOH)	4.5	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Glutamate (COOH)	4.6	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Histidine (imidazole)	6.2	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	0
N-terminal (amino)	7.3	+	+	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0	0	0	0	0
Cysteine (SH)	9.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−	−	−	−	−
Tyrosine (phenol)	10.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−	−	−	−
Lysine (amino)	10.4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	0	0	0
Arginine (guanido)	12.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0

[a] pK_a is the pH at which the ionizable group is half ionized. The precise pK_a value for a given ionizable group can be influenced by the immediate local environment.

1.3.4
Isoelectric Point

The isoelectric point is the pH at which the charge on a protein is zero and is determined by the number and titration curves of the positively and negatively charged amino acid residues on the protein. Protein pI values generally range from 4 to 10 (Figure 1.1). An example of a theoretical titration curve and pI determination of *E. coli* RNA polymerase transcription factor, sigma32 (σ^{32}), is shown in Figure 1.3.

1.3.5
Charge Distribution

The charged amino acid residues may be distributed uniformly on the surface of the protein or they may be clustered such that one region is highly positive while another region is highly negative. Such non-random charge distribution can be used to discriminate among proteins. An example is the *E. coli* σ^{32} protein (Figure 1.3). At pH 7.9, σ^{32} has a negative charge of −46 and a positive charge of +40, giving a net charge of −6. It is able to bind reasonably tightly

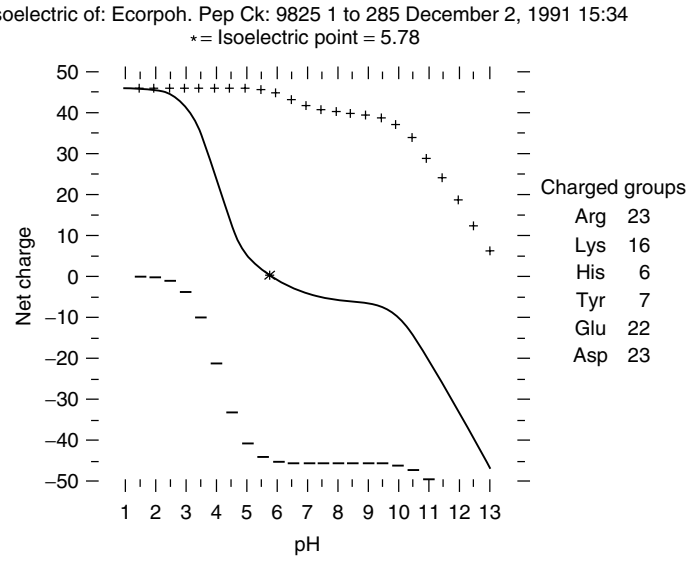

Figure 1.3 Titration curve and isoelectric point (pI) of *E. coli* σ^{32}. This graph shows theoretical plots of the number of positively charged and negatively charged groups and the net charge as a function of pH for the *E. coli* RNA polymerase transcription factor σ^{32}, based on its amino acid sequence. The pI is indicated by the asterisk and is 5.78. The charged groups are Arg (23), Lys (16), His (6), Tyr (7), Glu (22), and Asp (23). This plot was generated using the Genetics Computer Group Sequence Analysis Software package.

to both anion- and cation-exchange columns, apparently because its charged residues are not evenly distributed on the surface. This property can be used to purify this protein because most proteins will not bind to both types of ion-exchange columns under a single solvent condition.

1.3.6 Hydrophobicity

Most hydrophobic amino acid residues are buried on the inside of a protein, but some are found on the surface. The number and spatial distribution of hydrophobic amino acid residues present on the surface of the protein determine the ability of the protein to bind to hydrophobic column materials (as in hydrophobic interaction chromatography or HIC) and, therefore, can be exploited in fractionation. In general, a protein mixture is loaded onto an HIC column at high salt (e.g. 1 M ammonium sulfate), where hydrophobic interactions are strongest, and then eluted with a decreasing salt gradient to successively elute more and more tightly bound proteins.

1.3.7 Solubility

Proteins vary dramatically in their solubility in different solvents, all the way from being essentially insoluble ($<10\ \mu g\ mL^{-1}$) to being very soluble ($>300\ mg\ mL^{-1}$). Key variables that affect the solubility of a protein include pH, ionic strength, the nature of the ions, temperature, and the polarity of the solvent. Proteins are generally less soluble at their isoelectric point where there is less charge repulsion. Proteins are commonly fractionated by adding higher and higher concentrations of the mild salt ammonium sulfate. In general, the solubility of a given protein will decrease about 10-fold as the ammonium sulfate increases about 6% in saturation. (It takes about 760 g of ammonium sulfate added to a liter of water to give a 100% saturated solution, which is about 4.1 M, at 20 °C.) Since ammonium sulfate is mild and stabilizing to proteins, relatively inexpensive and pure, and highly soluble, it is the most common material used to fractionate proteins on the basis of solubility.

1.3.8 Density

The density of most proteins is between 1.3 and $1.4\ g\ cm^{-3}$ and this is not generally a useful property for fractionating proteins. However, proteins containing large amounts of phosphate (e.g. phosvitin, density $= 1.8\ g\ cm^{-3}$) or lipid moieties (e.g. β-lipoprotein, density $= 1.03\ g\ cm^{-3}$) are substantially different in density compared with the average protein and may be separated from the bulk of proteins using density methods.

1.3.9
Ligand Binding

Many enzymes bind substrates, effector molecules, cofactors, or DNA sequences quite tightly. This binding affinity can be used to bind an enzyme to a column to which the appropriate ligand or DNA sequence has been immobilized. For example, the transcription factor AP-1 is purified by binding to a specific DNA affinity column.

1.3.10
Metal Binding

Many enzymes bind certain chelated metal ions (e.g. Cu^{2+}, Zn^{2+}, Ca^{2+}, Co^{2+}, and Ni^{2+}) quite tightly, usually through interactions with cysteine or histidine residues. This binding can be used to bind an enzyme to a column to which the appropriate chelated metal ion has been immobilized. See Section 1.3.15 on the use of a metal-chelate column for purification of protein tagged by the addition of 6–10 terminal histidines.

1.3.11
Reversible Association

Under certain solution conditions, some enzymes aggregate to form dimers, tetramers, and so on. For example, the ability of *E. coli* RNA polymerase to be a dimer under one condition (0.05 M NaCl) and a monomer under another condition (0.3 M NaCl) can be used if two fractionations based on size are carried out sequentially under those two different conditions (Burgess, 1969).

1.3.12
Posttranslational Modifications

After protein synthesis, many proteins are modified by the addition of carbohydrates, acyl groups, phosphate groups, or a variety of other moieties. In many cases, these modifications provide handles that can be used in fractionation. For example, proteins containing carbohydrates on their surface can often be bound to columns containing plant lectins, which are molecules capable of binding tightly to certain carbohydrate moieties on glycoproteins. Phosphoproteins will in some cases bind to a chelated Fe^{2+} column.

1.3.13
Specific Sequence or Structure

The precise geometric presentation of amino acid residues on the surface of a protein can be used as the basis of a separation procedure. For example,

an antibody that recognizes only a particular site (epitope) on a protein can usually be obtained. An immunoaffinity column can be prepared by attaching a monospecific antibody (which binds only to the protein of interest) to a resin. Immunoaffinity chromatography (IAC) can result in highly selective separation and provides a very effective purification step (Burgess and Thompson, 2002). A protein of interest can also be immobilized and used to specifically bind another protein out of a complex protein extract. This process is called protein affinity chromatography.

1.3.14
Unusual Properties

In addition to the types of properties mentioned above, certain proteins have unusual properties that can be exploited during their purification—an example is unusual thermostability. Most proteins unfold and coagulate or precipitate when heated to 95 °C. A protein that remains soluble and active after such heat treatment can be separated easily from the bulk of the other cellular proteins. Another such property is unusual resistance to proteases. These two properties often go hand in hand. An interesting example of a purification involving these properties is that of *E. coli* alkaline phosphatase. The cellular extract is heated and the insoluble coagulated proteins are removed by centrifugation. The supernatant that contains the phosphatase is then treated with a protease, which digests the remaining contaminating proteins, leaving an essentially pure preparation of alkaline phosphatase.

1.3.15
Genetically Engineered Purification Handles

With the advent of genetic engineering, it has become relatively easy to clone the cDNA encoding a given protein. It is then possible to construct an overproducing strain of *E. coli* that can be induced to produce large amounts of a desired gene product. Recently, it has become common to alter the cDNA in such a way as to add a few extra amino acids on the N-terminus or the C-terminus of the protein being expressed. This added ''tag'' can be used as an effective purification handle. One of the most popular tags is addition of 6–10 histidines onto the N-terminus of a protein (Hochuli *et al.*, 1988; Ford *et al.*, 1991; Porath, 1992). The protein is then purified by its ability to bind tightly to a column containing chelated Ni^{2+} or Co^{2+} in which it can be washed and then eluted with free imidazole or by lowering the pH to 5.9, where histidine becomes fully protonated and no longer binds to chelated metal.

1.3.16
What Can Be Learnt from the Amino Acid Sequence of a Protein that is Useful in Purification?

These days it is common to purify a protein in which the gene has been sequenced. Thus, one can easily deduce the amino acid sequence of the

corresponding protein. Can this knowledge help in designing a purification scheme? The answer is that it can be somewhat, but not very, helpful. It is easy to determine the precise molecular weight of the polypeptide chain, but not to predict whether it forms dimers or tetramer or is part of a multi-subunit complex. Its charge versus pH and its isoelectric point can be determined as shown in Figure 1.3, which gives some idea as to which type of ion-exchange column to use, but again this will only be useful if it is not associated with other proteins. It is possible to calculate its extinction coefficient on the basis of its tryptophan and tyrosine content, which is very useful when it is pure (Gill and von Hippel, 1989). It is possible to determine if it has membrane-spanning regions or it has potential modification sites. One may be able to deduce that it is a member of a larger family of proteins by sequence alignment or by the presence of conserved sequence motifs that suggest cofactor affinity. However, its shape is not known, since the three-dimensional structure from sequence cannot yet be reliably predicted. Its multi-subunit features cannot be predicted, nor can its ammonium sulfate precipitation properties or its surface features such as hydrophobic patches, charge distribution, or antigenic sites. Therefore, it must be concluded that protein purification is still an empirical science.

1.4 Types of Separation Methods

There are a large number of separation processes that can be utilized to fractionate proteins on the basis of the properties listed above. These are summarized in Table 1.2. The sequential use of several of these separation processes will allow the progressive purification of almost any protein. If the processes are chosen carefully, and if proper attention is paid to separation conditions, and to maintaining the stability of the protein, the purification will result in reasonable efficiency, speed, yield, and purity, while retaining the biological activity and chemical integrity of the polypeptide (See Burgess, 1987; Coligan *et al.*, 1997; Deutscher, 1990).

An example of a hypothetical protein fractionation scheme is shown in Figure 1.4. This scheme relies on three of the most common fractionation methods, ammonium sulfate precipitation, ion-exchange chromatography, and gel-filtration or size-exclusion chromatography. The purification summary in Table 1.3 is a typical way of summarizing the yield and specific activity of each of the major steps in a purification scheme.

1.5 Protein Inactivation and How to Prevent It

A protein purification scheme will generally not be considered successful if the result is a protein that is pure, but inactive. Therefore, one of the key considerations in working with a protein is to prevent it from becoming inactivated. Table 1.4 summarizes some of the main reasons why a protein might become

Table 1.2 Separation processes that can be utilized to fractionate proteins.

Separation process	Basis of separation
Precipitation	
Ammonium sulfate	Solubility
Acetone	Solubility
Polyethyleneimine	Charge, size
Isoelectric	Solubility, pI
Phase partitioning (e.g. with polyethylene glycol)	Solubility
Chromatography	
Ion exchange (IEX)	Charge, charge distribution
Hydrophobic interaction (HIC)	Hydrophobicity
Reverse-phase HPLC	Hydrophobicity, size
Affinity	Ligand-binding site
DNA affinity	DNA-binding site
Lectin affinity	Carbohydrate content and type
Immobilized metal affinity (IMAC)	Metal binding
Immunoaffinity (IAC)	Specific antigenic site
Chromatofocusing	pI
Gel filtration/size exclusion (SEC)	Size, shape
Electrophoresis	
Gel electrophoresis (PAGE)	Charge, size, shape
Isoelectric focusing (IEF)	pI
Centrifugation	Size, shape, density
Ultrafiltration	Size, shape

inactivated and what can be done to prevent this inactivation. In general, working quickly and at low temperature (in a cold room or ice bucket) will help to avoid proteolytic degradation. Avoiding foaming or undue exposure to oxygen and adding a reducing agent should prevent oxidation. A buffer is used to maintain pH and a chelating agent like EDTA to protect against heavy-metal ions. Addition of 5% glycerol seems to stabilize most proteins and reduce adsorption to the walls of the container. A low salt concentration (e.g. 100 mM) helps increase solubility and prevent ionic adsorption to surfaces. Table 1.5 gives a good all-purpose buffer that in most cases will keep a protein active and happy.

1.6 Protein Purification Strategy

How a purification scheme is designed will, in large part, determine how successful it will be in achieving the goal of a protein purification: getting a high yield of highly pure and active protein in the minimal number of steps. Achieving a high final yield requires a high recovery at each step. Four steps at 80% step yield will be $(0.8)^4 = 0.41 = 41\%$ final yield, while four steps at 60% step yield will give a 13% final yield. Final purity will be guided by the intended

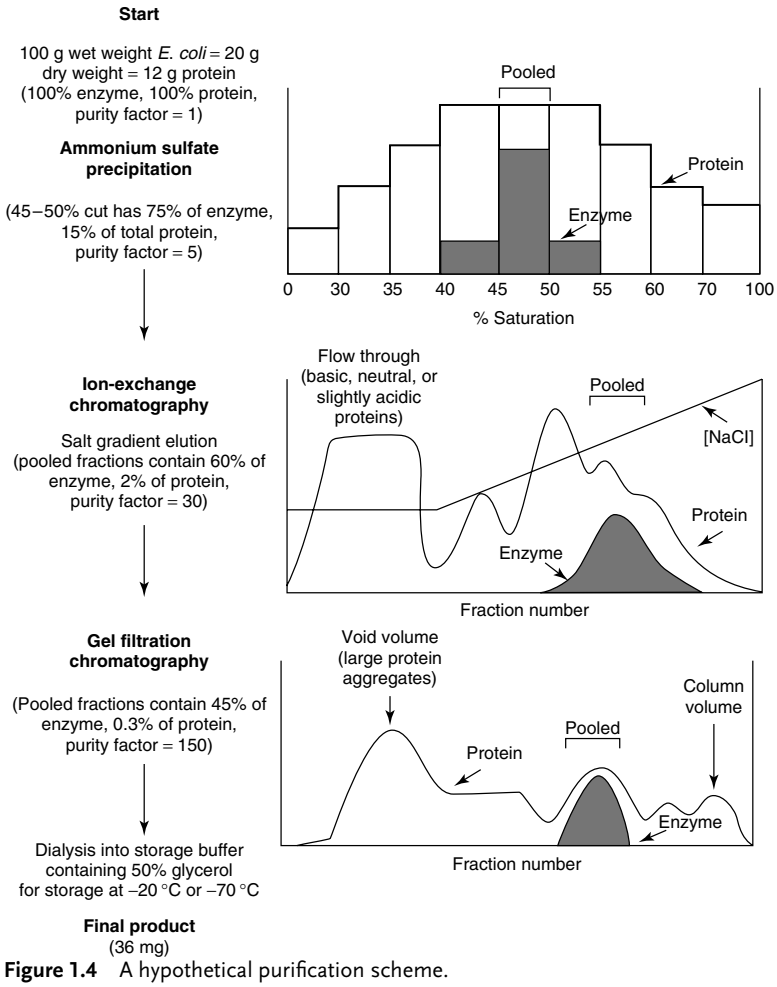

Figure 1.4 A hypothetical purification scheme.

use of the protein, but it should be free of major contaminants and any traces of enzymes that interfere with the intended use. High activity will depend on maintaining the stability of the protein as discussed in Section 1.5. By choosing high-resolution fractionation steps, a given fold purification can be achieved in fewer steps. For example, if the target protein is 0.01% of the total protein in the extract, it will require a 10^4-fold purification. This can be achieved in four steps, each giving a 10-fold purification, or three steps with a 22-fold purification, or two steps capable of 100-fold purification. The fewer the steps, the faster the preparation, the lower the protein losses, and the lower the cost of the purification procedure. Some of the key considerations in designing a purification procedure are (i) to have a convenient assay to follow purification; (ii) choose a starting material rich in protein; (iii) take precautions to minimize damage, inactivation or loss; (iv) use the minimal number of steps; (v) remove

Table 1.3 Summary of hypothetical purification in Figure 1.4.

Fraction	Total protein (mg)	Total activity (%)	Specific activity	Step yield (%)	Overall yield (%)
Extract	12 000	100	= 1	75	= 100
Ammonium sulfate precipitation (45–50%)	1800	75	5	80	75
IEC (pooled peak)	240	60	30	75	60
Gel filtration (peak)	36	45	150-fold purification		45 final yield
Pure standard			150		

IEC, ion exchange chromatography.

Table 1.4 Protein inactivation and ways of preventing it.

Reasons for inactivation	How to prevent it
Oxidation, foaming	Add DTT or TCEP, store under argon
Protease degradation	Add protease inhibitors, cooler, purer
Adsorption to container	Use polypropylene tubes, BSA carrier, glycerol, non-ionic detergent, protein more concentrated
Aggregation and precipitation	Store less concentrated, add salt, pH away from pI
Heavy metals	Add EDTA, cleaner tube, reagents
Temperature inactivation	Store cooler, add ligand or glycerol to stabilize
Bacterial growth	Use Tris, EDTA, azide, avoid PO_4, OAc^-
Enzymatic reaction (phosphatase)	Cooler, purer, add specific inhibitor
Dissociation of subunits/cofactors	Store more concentrated
pH changed	Avoid CO_2 in room, Tris changes pH with temperature
Inactive/misfolded conformation	Incubate at 37 °C to anneal the structure

the bulk of material quickly; (vi) avoid unnecessary duplication, dialysis, and delay; (vii) generally use fractionation steps in the order: precipitation, ion exchange, affinity, sizing; and (viii) use high-resolution steps where possible.

1.7 Overproducing Recombinant Proteins

The advent of genetic engineering has given us the ability to routinely clone genes and overproduce their gene products. This has changed the way we

Table 1.5 A good all-purpose buffer for keeping proteins happy.

TGED + 0.1 M NaCl	
Buffer	50 mM Tris-HCl, pH 7.9 at 20 °C
Stabilizer	5% glycerol
Chelator	0.1 mM EDTA
Reducing agent	0.1 mM DTT (dithiothreitol)
Salt	0.1 M NaCl

Storage buffer—similar to above but has 50% glycerol. Will not freeze at −20 °C. Best stored at −70 °C.

think about protein purification. For most protein purifications, it is no longer necessary to start with large amounts of naturally occurring material. Instead, the gene of interest can be cloned, inserted into a suitable expression vector and the vector transformed into a suitable expression host (e.g. *E. coli*). The cells can then be grown, transcription of the gene of interest induced, the cells harvested, the cells broken open and the overproduced recombinant protein purified. By using an expression vector where the target gene has a strong, inducible promoter, it is possible to express the target to levels as high as 20–40% of the total cellular protein.

The most commonly used bacterial expression system was developed by Bill Studier and colleagues at Brookhaven National Laboratory (Studier *et al.*, 1990). The host strain BL21(DE3) is a derivative of *E. coli* B that is deficient in several proteases to help prevent proteolysis of the recombinant protein and that has an inducible copy of the T7 phage RNA polymerase integrated as a phage lambda lysogen in the bacterial chromosome. The T7 RNA polymerase is kept repressed by the lactose operon repressor due to the placement of a lac operator near the promoter. The gene of interest is inserted into a multicopy expression vector under the control of a T7 RNA polymerase promoter, also with a lac operator to keep it off. The vector also contains an extra copy of the lac repressor gene to enhance repression. The repressor and the presence of an additional plasmid (pLysS, that encodes a T7 lysozyme to inhibit low levels of T7 RNA polymerase activity) help keep uninduced transcription of the target gene low. When the cells have been grown to the desired cell density, the lac operon inducer, IPTG, is added to about 1 mM, causing the repressor to dissociate from the lac operators and allowing expression of the T7 RNA polymerase. The T7 RNA polymerase in turn actively transcribes the many copies of the plasmid-encoded recombinant gene and the resulting mRNA is efficiently translated into the protein of interest. The result can be the production of very large amounts of the recombinant protein, to as much as 20–40% of the total cell protein within about 4 h after induction. This means that it is possible to purify as much as 30–50 mg of recombinant protein from 1 g of wet weight bacterial cell paste.

Several refinements have been introduced in the last few years to improve the chances that overproduction will be successful, especially if it is a mammalian,

Table 1.6 Problems of poor recombinant protein expression and their solutions.

Problem	Solution
1. Target gene contains rare *E. coli* codons	Supplement host *E. coli* with rare tRNAs (Novy *et al.*, 2001)
2. Target mRNA is degraded	Use *E. coli* strain deficient in RNase E (Lopez *et al.*, 1999)
3. Target protein is toxic to host cells	Use tighter repression, lower copy plasmid
4. Target protein is a membrane protein	Use a strain that has extra internal membranes (Miroux and Walker, 1996)
5. Target protein needs to form disulfide bonds	Use strain that is deficient in several key reductases, Gor, TrxB (Bessette *et al.*, 1999)
6. Product normally forms stable heterodimers	Simultaneous coexpression of two different proteins in one *E. coli* strain (Held *et al.*, 2003)

plant, or archaeal protein that is being overproduced. One problem is that human proteins often use codons that are rarely used in *E. coli*. These codons correspond to *E. coli* tRNAs that are very low in abundance in the cell. When overexpression of a gene containing many of these rare codons is attempted, the result is that very little, if any, of the protein is produced. This problem was originally solved by changing the DNA sequence of the recombinant gene so that it did not contain rare codons, but it contained the preferred *E. coli* codons. A much easier and more elegant solution has now been developed and is marketed by several biotechnology research products companies. This involves creation of an improved host bacterial strain that has had 3–5 of its rare tRNAs augmented. Many poorly expressed proteins can now be expressed at very high levels. Table 1.6 lists this and several other problems that have been encountered in protein overexpression in *E. coli* along with how these problems have been solved or alleviated. There are a wide variety of elegantly engineered expression vectors and bacterial expression hosts available from many different biotechnology research products companies worldwide. In addition to many *E. coli*-based expression hosts, there are also expression hosts such as: *Bacillus subtilis*, *Pichia pastoris*, *Aspergillus* spp., baculovirus/insect cells, mammalian cells, plants, and animals.

1.8 Refolding Proteins Solubilized from Inclusion Bodies

One of the most common problems in overexpressing a recombinant protein in *E. coli* is the fact that while large amounts of the protein are produced, most of it is not soluble, but is found as an insoluble inclusion body.

Apparently, the newly synthesized protein, when partially folded into its native structure, exposes some hydrophobic regions and is quite sticky and prone to interaction with other partially folded proteins, leading to aggregation and inclusion body formation. There are two main approaches to dealing with inclusion bodies: (i) try to increase the proportion of the overproduced protein that is soluble (Schein, 1989); and (ii) purify the inclusion body, solubilize it by dissolving it in a protein denaturant, and then refold it into its native structure (Marschak *et al.*, 1996).

1.8.1 Increasing Production of Soluble Protein

To purify the soluble material, it is necessary to increase as much as possible the proportion of the overproduced protein that is soluble. The most common approach is to induce the overproduction in cells growing at 20–25 °C. Apparently, the slower growth rate and lower temperatures results in more refolded protein and less aggregation and inclusion body formation. People have tried coexpressing cloned chaperone proteins to facilitate proper folding, but this is not common as it is only effective in some cases. An elegant recent approach has been to grow the cells at 37 °C, shift the temperature briefly to 42 °C to induce the heat shock response, and then shift the cells to 20 °C for induction. Often if two proteins that normally form stable heterodimers are individually overexpressed they are insoluble, but if coexpressed in the same cell they form soluble, native heterodimers. Finally, many proteins remain soluble when overexpressed as genetic fusions with known proteins that readily fold to form stable native structures (protein fusion partners like NusA, GST, TrxA, and maltose binding protein, MBP).

1.8.2 Refolding Inclusion Bodies

If a protein solubilized from inclusion bodies is to be refolded, it is common to first wash the inclusion bodies with a non-ionic detergent like Triton X-100 to solubilize membranes and break any unbroken cells. The washed inclusion body is then almost pure, and is ready to be solubilized. The real challenge is not the purification, but the refolding. The key to refolding without reaggregation and precipitation is to refold under low protein concentration. Under these conditions, the concentration of sticky, partially refolded material is lower, decreasing the opportunity for interaction and aggregation. However, for larger preparations, the large volumes become a major problem. Usually, refolding conditions can be found that give efficient refolding yields at reasonable protein concentrations.

1.8.3
A General Procedure for Refolding Proteins from Inclusion Bodies

A general method that the author has found to be quite effective for many proteins is given below.

1. Grow cells and induce overexpression of target protein.
2. Harvest cells, weigh cell pellet, store frozen at −80 °C.
3. Break cells by sonication (pLysS cells are easy to break because of the presence of some T7 lysozyme). Otherwise adding lysozyme helps.
4. Centrifuge cell lysate, wash the inclusion body pellet with 1% Triton X-100 to solubilize membranes and membrane proteins, then wash with buffer to remove Triton X-100.
5. Solubilize inclusion bodies with a denaturant such as 6 M guanidium hydrochloride or 0.3% Sarkosyl to about 1 mg protein mL^{-1}. Difficult-to-refold proteins may need to be diluted to 0.1 mg mL^{-1} in denaturant; 8 M urea can also be used, but there is a risk of carbamylation of the protein.
6. Centrifuge out any undissolved material and slowly drip dilute the solubilized protein into 15–60 volumes of suitable refolding buffer. Additives or various buffer variables are often used to improve folding efficiency of a particular protein. These include 0.5 M arginine, 25% glycerol, varying the pH, temperature, presence of divalent ions, and presence of redox buffers.
7. Pass dilute refolded protein over a suitable high-resolution ion-exchange column, wash the column, and then elute with an increasing salt gradient. This step accomplishes several important things: (i) it concentrates the dilute protein; (ii) it removes the denaturant; (iii) it removes impurities that do not bind to the column or bind weaker or stronger than the target protein; and finally (iv) it often separates refolded monomer from soluble multimers that tend to bind tighter and elute later in the gradient.
8. The resulting protein is usually fully active, homogeneous, and capable of forming crystals suitable for three-dimensional structure determination.

References

Bessette, P.H., Aslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the E. coli cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 13703–13708.

Burgess, R.R. (1969) A new method for the large-scale purification of E. coli DNA-dependent RNA polymerase. *Journal of Biological Chemistry*, **244**, 6160–6167.

Burgess, R.R. (1987) Protein Purification, in *Protein Engineering* (eds D. Oxender

and C.F. Fox), A.R. Liss, New York, pp. 71–82.
Burgess, R.R. and Thompson, N.E. (2002) Advances in gentle immunoaffinity chromatography. *Current Opinions in Biotechnology*, **13**, 304–308.
Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W. and Wingfield, P. (1997) *Current Protocols in Protein Science*, John Wiley & Sons, Inc., New York.
Creighton, T.E. (1993) *Proteins: Structures and Molecular Properties*, 2nd edn., W.H. Freeman, San Francisco, CA.
Deutscher, M.P. (1990) *Guide to Protein Purification, Methods in Enzymology*, vol. 182, Academic Press, New York.
Ford, C.F., Suominen, I. and Glatz, C.E. (1991) Fusion tails for the recovery and purification of recombinant proteins. *Protein Expression and Purification*, **2**, 95–107.
Gill, S. and von Hippel, P. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Biochemistry*, **182**, 319–326.
Held, D., Yaeger, K. and Novy, R. (2003) Co-expression vectors. *Innovations*, **18**, 4–6.
Hochuli, E., Bannworth, W., Dobeli, R., Gentz, R. and Studber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Biotechnology*, **6**, 1321–1325.
Lopez, P.J., Marchand, I., Joyce, S.A. and Dreyfus, M. (1999) RNase E, the C-terminal half of which organizes the *E . coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Molecular Microbiology*, **33**, 188–199.
Marschak, D., Kadonaga, J., Burgess, R., Knuth, M., Brennan, W. and Lin, S.-H. (1996) *Strategies for Protein Purification and Characterization: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor.
Miroux, B. and Walker, J. (1996) Over-production of proteins in *E. coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, **260**, 289–298.
Neidhardt, F.C. and Phillips, T.A. (1985) The protein catalog of E. coli, in *Two-Dimensional Gel Electrophoresis of Proteins* (eds J.E. Celis and R. Bravo), Academic Press, New York, pp. 417–444.
Novy, R., Drott, D., Yaeger, K. and Mierendorf, R. (2001) Overcoming codon bias of E. coli for enhanced protein expression. *Innovations*, **12**, 1–3.
Porath, J. (1992) Immobilized metal ion affinity chromatography. *Protein Expression and Purification*, **3**, 206–281.
Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *Biotechnology*, **7**, 1141–1149.
Scopes, R. (1994) *Protein Purification: Principles and Practice*, 3rd edn., Springer-Verlag, New York.
Simpson, R.J. (ed.) (2004) *Purifying Proteins for Proteomics: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor.
Studier, W., Rosenberg, A., Dunn, J. and Dubendorff, J. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology*, **185**, 60–89.