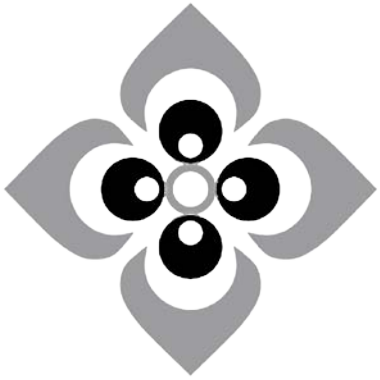

Consortium for Educational Communication

Module
on
Protein Purification

By
NAZISH NAZIR
Research Scholar
Centre of Research for Development
University of Kashmir
+91-9622618814
email. naazish.zargar@gmail.com

TEXT

- **Protein Purification and its brief history**

Proteins are the molecules that do interesting things in cells, they are involved in the production of all other molecules and cell structure, and they are involved in a cell's ability to respond to changes in its environment. It's not clear how many proteins there are in an organism, but a "best guess" for humans is about 50,000, and for bacteria may be 10,000 or so. There are a variety of ways to investigate protein function. Some of these are made possible by advances in molecular biology and these are useful, because they look at the protein in its cellular context. But often it is necessary to purify the protein away from the other thousands of cellular proteins, so that its structure can be examined, or its interactions with other molecules studied more specifically. Thus, a great deal of protein chemistry has to do with the principles and problems of protein purification, and even when not related to purification, it can be instructive and applicable to all sorts of protein work. The tremendous variety in protein structure and chemistry makes the task of purification sensible but difficult. The overall challenge is to find methods that will separate proteins into two (or more) different fractions based on some property (charge, size, density etc). The protein of interest is thus put into a smaller group of proteins that are similar in one way (let's say, proteins that are large), and then take this group and separate them based on whether they bind to (+) charged surfaces. One always looks for the protein of interest, hoping to find one fraction that has lots of the protein of interest, but few other proteins. One important thing to understand is that the process of designing purification strategies is mostly empirical. A method is applied, the investigators find the protein of interest, characterize the different fractions and decide if this was a good method to use (or not) and then they go again. These steps also tell the investigators something about the protein. It is true that our ability to learn the primary structure of an interesting protein before we even begin to try to purify it (by having identified its DNA sequence and de-coding it), but that only takes us part way. The folded, native structure of a protein dictates its interactions with other molecules, and hence, the relevant purification strategies. It is also true that genetic engineering has allowed us to manipulate proteins by attaching tags that allow purification schemes that have little to do with the proteins native structure. Still, these methods are seldom used as

the sole purification step, and they are still based on the same sorts of principles that drive a more classical protein purification approach and optimization of such steps is still largely empirical.

The term “protein” comes from the Greek, “standing in front.” The term was suggested by Jons Jacob Berzelius in 1838. At the time there were ongoing discussions about whether proteins were macromolecules or colloidal aggregates, a debate that lasted until 1930. In 1901 Hermann Emil Fischer synthesized a dipeptide, demonstrating that amino acids can be linked together, and the following year Franz Hofmeister suggested that proteins are amino acids linked by peptide bonds. Many years later (1926) James B. Sumner crystallized the enzyme urease and showed by chemical analysis that the substance was a protein. This was the first proof that enzymes are proteins, Linus Pauling and co-workers proposed in 1951 that the primary structural motifs of proteins are α helices and β sheets. In 1957 Christian B and colleagues showed that ribonuclease (RNase) could be refolded into its native structure and regain its activity after complete unfolding and reduction (disulfide bonds broken). Later it was suggested that the native structure of a protein is thermodynamically most stable structure with the given amino acid sequence. The same year, Max Perutz and John Kendrew determined the structure of myoglobin.

Importance of Protein Purification:

The study of proteins and their function is central to understanding both cells and organisms. We have a few of the reasons why proteins are important in biology:

- They serve as catalysts that maintain metabolic processes in the cell.
- They serve as structural elements both within and outside the cell.
- They are signals secreted by one cell or deposited in the extracellular matrix that are recognized by other cells.
- They are receptors that convey information about the extracellular environment to the cell.
- They serve as intracellular signaling components that mediate the effects of

receptors.

- They are key components of the machinery that determines which genes are expressed and whether mRNAs are translated into proteins.

They are involved in manipulation of DNA and RNA through processes such as DNA replication, DNA recombination, RNA splicing or editing. Beyond this, there are a number of important reasons for purifying proteins, and a brief list of these may serve to emphasize why individuals spend so much time and effort in this pursuit. By purifying a protein it can be clearly established that a particular biological activity (enzymatic activity, signaling capacity, etc.) actually resides in a unique protein. Purified proteins serve as extremely valuable biochemical reagents. It is remarkably valuable to be able to obtain things like purified growth factors or hormones, proteases, DNA polymerases, reverse transcriptases, ligases, phosphatases, or antibodies that recognize a particular epitope of interest. Once a protein is purified, it is possible to study its enzymology, understand its affinity for particular substrates, or dissect its ability to catalyze enzymatic reactions. Such approaches have allowed us to understand how biological molecules can act as catalysts in metabolic processes or as transducers that will convert chemical energy into ionic gradients or mechanic forces. The availability of purified proteins allows bio-organic chemists to modify specific residues to help understand how these residues confer particular structures or allow the protein to operate as catalyst. Purified proteins can be sequenced either by Edmund degradation, which obtains sequence information from the N-terminal sequence of a protein or a peptide derived from a protein. This approach is rapidly being replaced by mass spectroscopy. Sequence information can be used to design probes to isolate cDNAs and this information can be used to deduce the primary sequence of the entire protein. MALDI-TOF (Matrix Assisted Laser Desorption Ionization Time of Flight) is a technology that can provide much information about a protein, even when only small quantities are available. Partially purified proteins can often be identified on the basis of fragments generated by ionizing the proteins or peptides derived from them and performing mass spectroscopy. Traditional structure-function studies have become remarkably more powerful because of the ability to isolate and purify not only conventional enzymes but also mutant

forms of those proteins. With the advent of site-directed mutagenesis and expression systems, it is possible to produce and purify a protein with essentially any mutation of interest. This allows the testing of a broad spectrum of hypotheses as well as the design of “improved” protein molecules. Using genetic engineering, it is possible to create novel proteins or combination of protein elements that can serve a specific function. For example, it is possible to combine a region of one protein that is capable of activating transcription with a region of another protein that recognizes a particular site on the DNA producing a novel DNA-binding protein. It is even possible to modify proteins so that particularly desirable characteristics like heat stability or protease resistance can be enhanced. If a protein with a particular enzymatic activity or signaling capacity can be purified it serves as an important indication of how biological systems work. As a cell contains a particular metabolic pathway is dramatically improved by isolating enzymes that catalyze components of that pathway. Isolation of a particular polypeptide growth factor can be a powerful indication of how an organism regulates particular cellular or tissue processes like angiogenesis. For example, our understanding of angiogenesis is dramatically enhanced by the isolation of the protein VEGF that stimulates capillary formation. Likewise understanding of how a growth cone can navigate within the body to make appropriate connections is dramatically improved when molecules capable of collapsing the growth cone (collapsin) are isolated and purified. Once a protein has been isolated, it is possible to produce reagents (antibodies) that are capable of determining the location of that protein *in vivo* which can give important support to interesting hypotheses and disprove incorrect speculations. By understanding the structure-function relationships of proteins, it is often possible to design specific reagents that can be used to test the function of a protein *in vivo.* For example, understanding the ways that G-proteins can be mutated so that the protein can either be active (always on) or be incapable of ever being activated (constitutively off), allows biologists to explore the role of these individual proteins *in vivo*. You might already know that G proteins bind GTP and GDP and act as molecular switches and that GTP bound is ‘on’ while GDP bound is ‘off’. Many of these ideas and their structural basis were worked out with a protein called ras, but subsequently many dozens of similar proteins (e.g., rac and rho) were isolated and they all work by a similar mechanism. Thus, understanding one makes it much easier to understand the others. Likewise,

understanding how DNA-binding proteins recognize DNA sequences allows us: to understand the significance of particular DNA mutations, to design proteins that can recognize modified DNA sequences or recognize a known DNA sequence and to design mutant proteins that interfere with the activity of proteins normally present in the cell. For example, if a protein has a bHLH motif (basic Helix Loop Helix) it is possible to make a good prediction of what sequences interact with DNA and what mutations might prevent binding of the protein to DNA or which sequences might compete with an endogenous proteins for DNA binding. Finally, the increasing power of X-ray diffraction and 2-D NMR to determine protein structure requires the availability of larger quantities of purified proteins. Understanding 3-D protein structure has begun to allow us, to understand how protein folding is controlled and it has provided a remarkable insight into the way the proteins act *in vivo.* The ability to visualize an ion channel, even at low resolution, must forever modify the way a scientist understands ion flow through the membranes. Likewise, understanding how proteins recognize specific DNA sequences changes the way our community understand transcriptional activation or repression. The impact of this type of 3-D image of a protein cannot help but have a dramatic impact on the way we visualize biological processes. Determining new protein structure is a major undertaking. To use either method, large amounts of pure protein must be available. For crystallography, protein crystals must be made and defraction information must be analyzed. Interpreting this data requires one to solve the phase problem, which usually means getting additional data about the defraction pattern of a protein that contains a heavy metal. This site provides a more detailed discussion of difraction and crystallography 2D-NMR avoids the need to make a crystal, but this approach is limited to smaller proteins (less than about 20,000-50,000 Daltons) and requires careful analysis. We can find a description of the theories and practical issues in NMT at the sites to understanding how particular amino acid residues are involved in protein function, especially when it is combined with knowledge of the 3-D structure of the protein, helps understand how particular sequences in proteins are involved in biological functions. The accumulation of this type of knowledge derived from the study of proteins, combined with tremendous amount of information in DNA databases have allowed a construction of what might be best called “protein descriptors.” These descriptors can consist either of a particular conserved sequence of an amino acid in the primary sequence or a particular pattern of distributed amino

acids in the protein. Frequently, these descriptors can be recognized with the primary sequence of a protein, but understanding how a particular descriptor might function in an actual protein can dramatically increase the value such a descriptor. For example, understanding the enzymology of proteolysis and the importance of a catalytic triad of Asp-His- Ser makes it worthwhile to look for the pattern in novel amino acid sequences. Being able to predict that a particular sequence of amino acid is likely to form an alpha-helix or a random coil can be important to understanding the properties and frequently the function of a protein. Visualizing the biological distribution of leucine residues in a 'leucine zipper' makes this pattern of leucine repeats of great interest. Recognizing that many of the proteins having the 'zinc finger motif' are transcription factors will often give a clue to the function of a protein. Although purified proteins are often studied in purified systems (*in vitro*), the advent of microinjection systems allow purified proteins to be introduced into single cells where their biological activity can be determined. In marked contrast to *in vitro* systems, microinjectiom allows a purified protein to be introduced into a cell where it can interact with a vast number of proteins and structures. It is possible to study complex changes (e.g., cell behavior, cell shape, cell division) which can't be easily reconstituted *in vitro*. It is a quick way of introducing a variety of proteins into different cell types in the absence of using an expression vector. This type of study, which is between the classic *in vivo* and *in vitro* studies, might be called in '*vivtro*.' It is an important approach to studying protein function. Likewise, it is often possible to study the effect of a particular protein on development or a physiological function by introducing it into an organism. Purification of a protein can also help purify a nucleic acid of interest. For example, if the hnRNA, snRNA proteins are purified, they can be co-purified with associated RNA, and bound RNA can be amplified. Briefly, the steps in the procedure are, make purified protein-RNA complex, purify the bounded RNA, reverse transcribe, add polynucleotide promoter tail, PCR amplify. The RNAs which specifically bind to the purified protein can be affinity selected, amplified and sequenced.

- **Different techniques/methods of protein purification:**

Sometimes, the amount of desired protein is so small, and the amount of macromolecular contaminant is so high, that one needs to employ nearly

every “trick of the trade” to achieve high purity. Imagine wanting to isolate milligrams of a precious protein from thousands of liters of crude jellyfish extract. Sometimes, purifying a protein to homogeneity, from such large volumes of highly viscous starting material, may involve separating one milligram of the protein-of-interest (POI) from 100 mg of initial total protein. This is called a 100-fold purification. In other cases the required purification factor may be on the order of 1000-fold or 10,000-fold. The issues facing a scientist working on a difficult protein purification project are many. Some of them are:

1. Choosing or developing a sensitive, reproducible, and selective assay for the protein of interest (POI).

2. Establishing conditions under which the POI is stable and biologically active.

3. Finding conditions under which the POI can be stored safely between steps.

4. Choosing the best biological starting material (natural source or recombinant).

5. Developing or choosing appropriate methods for gross extraction.

6. Decreasing viscosity of crude extracts and removing particulates from those extracts.

7. Reducing volume.

8. Finding the substrate(s), inhibitors, activators, allosteric effectors, etc., if the protein of

interest is an enzyme.

Early steps in designing protein purification strategies some very useful information can be acquired, unambiguously, if a small sample of pure protein can be obtained. It is so easy to make major errors if you try to over analyze a crude sample. Acquiring a pure sample of the protein-of-interest may be difficult (if the specific purification methods have not been optimized). But, obtaining a small amount of pure protein can be very useful for future optimization of purification. Unless the protein of interest is pure, data on its characteristics can be very misleading.

a) Solubility in water, salt solutions, organic solvents, etc.

b) Presence of isoforms or isoenzymes.

c) Molecular weight.

d) Degree of oligomerization (monomer, dimer, tetramer, aggregation, etc).

e) Isoelectric point.

f) Partial amino acid sequence (needed if the mRNA is to be found).

g) Post translational alterations (phosphorylation, glycosylation, blocked N-terminus,etc).

h) Amino acid analysis.

i) Relative hydrophobicity (as determined by Hydrophobic Interaction Chromatography or ammonium sulfate precipitation).

j) Antibodies to the protein of interest. Essential cofactors, prosthetic groups, stabilizing agents, etc.

Various methods of protein purification are:

The methods used in protein purification, can be analytical and preparative methods. The distinction is not exact, but the deciding factor is the amount of protein, that can practically be purified with that method. Analytical methods aim to detect and identify a protein in a mixture, whereas preparative methods aim to produce large quantities of the protein for other purposes, such as structural biology or industrial use. In general, the preparative methods can be used in analytical applications, but not the other way around.

Extraction:

Depending on the source, the protein has to be brought into solution by breaking the tissue or cells containing it. There are several methods to achieve this; Repeated freezing and thawing, sonication, homogenization by high pressure or permeabilization by organic solvents. The method of choice depends on how fragile the protein is and how sturdy the cells are. After the extraction process soluble protein will be in the solvent, and can be separated from cell membranes, DNA, etc. by centrifugation. The extraction process also extracts proteases, which will start digesting the proteins in

the solution. If the protein is sensitive to proteolysis, it is usually desirable to proceed quickly, and keep the extract cooled, to slow down proteolysis.

Precipitation and Differential Solubilisation:

In bulk protein purification, a common first step to isolate proteins is precipitation with ammonium sulphate $(NH_4)_2SO_4$. This is performed by adding increasing amounts of ammonium sulphate and collecting the different fractions of precipitate protein. One advantage of this method is that it can be performed inexpensively with very large volumes. The first proteins to be purified are water-soluble proteins. Purification of integral membrane proteins requires disruption of the cell membrane in order to isolate any one particular protein from others that are in the same membrane compartment. Sometimes a particular membrane traction can be isolated first, such as isolating mitochondria from cells before purifying a protein located in a mitochondrial membrane.

A detergent such as sodium dodecyl sulphate (SDS) can be used to dissolve cell membranes and keep membrane proteins in solution during purification; however, because SDS causes denaturation, milder detergents such as Triton X-100 or CHAPS can be used to retain the protein's native conformation during purification.

Ultracentrifugation:

Centrifugation is a process that uses centrifugal force to separate mixtures of particles of varying masses or densities suspended in a liquid. When a vessel (typically a tube or bottle) containing a mixture of proteins or other particulate matter, such as bacterial cells, is rotated at high speeds, the angular momentum yields an outward force to each particle that is proportional to its mass. The tendency of a given particle to move through the liquid because of this force is offset by the resistance the liquid exerts on the particle. The net effect of "spinning" the sample in a centrifuge is that massive, small, and dense particles move outward faster than less massive particles or particles with more "drag" in the liquid. When suspensions of particles are "spun" in a centrifuge, a "pellet" may form at the bottom of the vessel that is enriched for the most massive particles with low drag in the liquid. The remaining, non-compacted particles still remaining mostly in the liquid are called the "supernatant" and can be removed from the vessel to separate the supernatant from the pellet. The rate of centrifugation is specified by the angular acceleration applied to the sample, typically measured

in comparison to the g. If samples are centrifuged long enough, the particles in the vessel will reach equilibrium wherein the particles accumulate specifically at a point in the vessel where their buoyant density is balanced with centrifugal force. Such an "equilibrium" centrifugation can allow extensive purification of a given particle.

➢ Chromatographic Methods

In the course of developing a start-to finish protocol for any given protein, unexpected information is uncovered along the way. Long after developing a working protocol, one may discover, for example, that the POI is glycosylated. Following this discovery, one might want to experiment with affinity chromatography using an immobilized lectin or may wish to try a boronate column that binds vicinal hydroxyl groups on sugar residues. I want you to understand that no purification protocol is ever final. There are always alternate ways that could improve or streamline an earlier protocol. This is one of many places that the artistry of protein purification comes into play.

▪ Ion Exchange Chromatography (IEX)

In this method, different types of proteins are separated based on their net charge. As such, columns can either be prepared to facilitate both anion exchange and cation exchange.

Once viscosity has been largely eliminated and once the crude protein sample is particle free, it may be time to use ion exchange chromatography the most frequently employed chromatographic method for proteins. Early, small-scale testing with a relatively salt-free sample is advised. There are simple, syringe-operated ion exchange columns both anion exchange columns and cation exchange columns. These columns can be used to determine (within one-half of a pH unit) the isoelectric point of the protein. This is accomplished by equilibrating the two columns with low ionic strength buffers of varying pH values. The most common cation exchange functional group is carboxymethyl, abbreviated CM. CM is essentially immobilized acetic acid and, like acetic acid, CM takes on a negative charge at pH values of 4 and above. CM is designated a weak cation exchanger as it has little binding capacity below

pH 4. Sulfonated or phosphorylated exchangers are called strong cation exchangers because they can be used at pH 2. For the POI to bind to CM, the protein must be positively charged (below its isoelectric point). CM is not satisfactory for GFP purification as GFP is unstable below its pI of 5.3. When GFP takes on a positive charge (below pH 5.3) the protein slowly denatures, losing its fluorescence. So, it is not possible to use CM with GFP in any slow process like column chromatography. But, if GFP exposure time is kept at a minimum, the pI of GFP can be estimated by its binding to CM at pH's below 5.3. Diethylaminoethyl (DEAE) is the most commonly used anion exchanger. The DEAE functional group is a tertiary amine, protonated (and positively charged) at pH values below 10. DEAE is designated a weak anion exchanger as it cannot be used effectively above pH 10. But, a bead-bound quarternary amine extends the range of anion exchange to pH 12. So any medium designated Q (or QAE, for quarternary amino ethyl) is called a strong anion exchanger. All four of these types of these media (weak and strong cation exchangers and weak and stronganion exchangers) are available in small, syringe operated columns. If one of these DEAE columns is equilibrated at a variety of pH values (10, 9, 8, 7, 6, 5, and 4), GFP will bind from pH 10 to pH 5.5, but not at pH 5, indicating that the pI of GFP is below 5.5.Once the pI has been determined and the anion exchanger has been chosen, a preparative column can now be poured. Most ion exchangers can bind 30 to 50 mg of protein per 1 ml of swollen gel. One can estimate the total amount of protein in the sample by absorbance at 280nm, ascribing one absorbance unit to one mg/ml of protein. But, high levels of DNA and moderate turbidity will artificially elevate this absorbance number (sometimes greatly). It is good practice to test, experimentally, the capacity of an ion exchange material in a small trial. Using 1 ml of swollen gel, add crude extract in successive 100 micro liter volumes until the gel becomes saturated with protein. The saturation limit can be determined by taking POI activity measurements after each incremental addition of extract.For enzymatic measurement, remove just a few microliters of the supernatant after the gel settles (so the aqueous volume remains about the same). When the activity appears in the supernatant, you will have determined the saturation point in terms of mg of extract per ml of gel. Now fill a chromatography column with at least 5-times as much gel as your preliminary testing indicates you will need for total binding. Short, stout columns are usually better than long thin ones. Resolution comes not from column dimensions, but from the rate

at which the eluting strength of the salt (usually sodium chloride) is raised in the elution phase. Take note of the fact that an ion exchanger is an excellent buffer, so pH equilibration of the gel requires many column volumes of dilute buffer solution. Alternatively, a very high concentration of buffer may be used to titrate the column, first. But, after titration, at least one column volume of the dilute (low ionic strength) buffer must be passed through the column. It is also necessary to use a buffering salt that has the same charge as the ion exchange gel. When using positively charged DEAE columns, positively charged Tris (hydroxyl methyl amino methane) buffer in the chloride form (generally abbreviated as Tris) is commonly used. For negatively charged CM, negatively charged sodium phosphate buffers are recommended. The protein of interest should be equilibrated in the same dilute buffer. For best resolution, a shallow, continuous gradient (50 column volumes or greater) from 0.0 M NaCl to 0.5 M NaCl is recommended. To achieve near base line resolution of 5 GFP iso forms (differing from each other by one or two amino acids), h 100 ml DEAE column with 80 column volumes (8 liters) of sodium chloride solution shoud be eluted from 0.05 to 0.25 M. In this case (and in all other cases) the salt solutions need to be prepared in the same buffer used to equilibrate the column.

- **Hydrophobic Interaction Chromatography (HIC)**

HIC media are available in several strengths. The hydrophobic ligands are usually attached to the porous hydrophilic gels via a 3-carbon spacer based on epichlorohydrin chemistry. From strongest binding to weakest binding ligands, the order is Phenyl >Octyl> Butyl > Methyl. Strongly hydrophobic ligands are appropriate for weakly hydrophobic proteins and weakly hydrophobic ligands for strongly hydrophobic proteins. Early testing, calculation of gel volume, and choice of column dimensions are carried out in a similar fashion as the protocols used for ion exchange. Hydrophobic binding is favored by very high salt concentration (up to 3 molar ammonium sulfate, in some cases). Elution is accomplished by lowering the salt concentration in increments (step gradient) or by applying a continuous linear gradient of decreasing salt concentration. Be aware that gradients of ammonium sulfate produce gradients of refractive index, easily confused by a spectrophotometer as a higher UV absorbance value or a lower

UV absorbance value. If precise 280 nm absorbance measurements are desired following gradient elution of proteins from an HIC column, it is necessary to have a continuously changing blank that closely matches the salt concentrations of the samples. An advantage to having HIC follow IEX is that one need not remove the NaCl in the fractions eluted from the IEX column. NaCl neither favors nor inhibits hydrophobic interaction nor does it interfere with spectroscopic measurements as much as ammonium sulfate. If the two steps are reversed, ammonium sulfate must be removed entirely before going on to IEX.

- **Affinity Chromatography**

This technique is considered to be the most selective chromatography technique, it is known to give the purest results and is therefore used in completing the protein purification process. Some prefer to use affinity chromatography very early in a protein purification process as a "one-step purification method". Often contaminants remain in affinity-purified proteins. Commonly, those contaminants are large protein aggregates that result from the almost inevitable leaching of " bound" ligand. That released ligand then forms a high molecular weight complex with the protein of interest. When we purify 'anti-GFP' antibodies on an immobilized GFP affinity column we almost always detect, by SEC-HPLC, a high molecular weight aggregate that is distinctly fluorescent, suggesting that an antigen (GFP)-antibody complex has formed. Because most affinity columns are quite expensive and could be plugged by crude starting samples, one should prefer to use affinity chromatography late in a protocol. The principle is easy. Take for example, that a ligand, recognized by an enzyme, is covalently bound to the matrix (usually agarose). That ligand may be a pseudo-substrate, a cofactor, an inhibitor, or an antibody. Binding is easy, but elution may be difficult. It is preferable to use, as the eluting solvent, a solution containing a competing ligand (the pseudo-substrate, cofactor, inhibitor, or antibody). But, sometimes the competing ligand is very expensive, unavailable, or irreversibly bound to the enzyme. In such cases, other eluting solvents must be used. Dilute solutions of ethylene glycol in buffer are sometimes used. So are buffers of low pH, a variety of salts, metal chelators, etc. Many other forms of affinity chromatography exist. We have to purify anti-GFP antibodies on a column to which GFP is covalently immobilized.

We normally elute with a concentrated pH 3.0 solution of sodium citrate. The pH 3 buffer temporarily denatures both the antibody and the GFP. Both column-bound GFP and the eluted antibody are rapidly renatured with a strong pH 8 buffer. Based upon analytical techniques (including size exclusion (SEC), HPLC, SDS gel electrophoresis, UV absorption spectroscopy and western blotting) purity of GFP-specific antibody can approach 99% However, if purity greater than 99% is desired, affinity chromatography requires a follow up step. Most commonly we use preparative gel filtration to remove protein aggregates that may form when a small quantity of bound ligand leaches from the column. For recombinant proteins, the favorite affinity column is an immobilized (chelated) metal ion column (abbreviated IMAC for immobilized metal ion affinity chromatography). In IMAC columns, nickel ions or cobalt ions are bound to the column in a chelate ion complex. The column-bound chelator is usually nitrilo tri acetic acid. The metal ion, chelated to the IMAC column, can be co-chelated, non-specifically, by the R-groups of histidine, cysteine, and tryptophan. Binding may occur if one or more of these amino acids are exposed on the surface of the protein-of-interest (or any protein contaminant in the mixture). Almost universally, recombinant proteins that are subjected to generic affinity chromatography are processed by IMAC. But to achieve specificity (and tight binding), the recombinant proteins are genetically modified by the addition of a string of 6 histidine residues, sometimes on the C-terminus, sometimes on the N-terminus, and sometimes within exposed loop regions. The string of 6 histidines (the HIS-tag) is a strong co-chelator and the tag is sufficiently exposed that the His-tag almost always out-competes any naturally occurring co-chelators found in high abundance on the surface of a protein contaminant. The method is carried out at pH of 8 or higher and it must be performed in the absence of other metal chelators such as EDTA, citrate, oxalate, ammonium ion, etc. Concentrated solutions of imidazole are usually used for elution. In my experience, all affinity chromatography columns, each time they are used, leach a bit of their covalently bound ligand, often as high molecular weight complexes with the POI. That ligand winds up in the fractions that have eluted from the column. So, in every case in which IMAC is used, it is wise to follow this step with a gel filtration run.

- **Gel Filtration Chromatography**

This technique is generally used to separate larger proteins from smaller ones by using a minimal volume of eluate. It exhibites good sensitivity and does not lead to

sample loss mainly because the solutes do not interact with the stationary phase.

Low pressure gel filtration is the easiest chromatographic method in principle, but it is the hardest method to administer properly. Because gel filtration seems so straight forward, liberties are sometimes taken in utilizing the method. For best results, attention to detail is essential. Gel filtration (or size exclusion as the method is called in HPLC) separates macromolecules by size. Size exclusion chromatography (SEC) is generally used as an analytical HPLC method while gel filtration is used primarily in preparative protein separations. Size exclusion HPLC utilizes small, rigid, uniform, spherical beads of 5 micrometer or 10 micrometer diameter. Small, porous, silica beads used in SEC provide higher resolution than low pressure gel filtration. Low pressure gels are comprised of small (20 to 300 micrometers) porous beads which, unlike Fast Flow adsorption beads, have blind cul-de-sacs that provide differential flow paths through the column. The largest molecules are unable to enter any pores, so they must travel around the beads. This means that large molecules exit first while smaller molecules spend some time inside the beads, so they exit later. The volume in which the very large molecules exit (DNA, proteoglycans, ribosomes, lipid micelles, and protein complexes) is called the void volume. The void volume, usually 25% of the total column volume, is often measured by the elution position of Blue Dextran , a covalently-dyed sugar polymer having a molecular weight of 2 million Daltons. So, if the column volume (π r2 h) is 200 cubic centimeters (200 ml), the center of the Blue Dextran-calibrated void volume peak will appear close to the 50 ml mark. The next 25% of the column volume is the resolving zone, accessible to moderate size proteins. The final 50% of the column volume (100 ml) is the zone in which peptides, very small proteins, oligonucleotides, other small molecules and salt ions will elute. The total liquid volume in the column (salt volume) is accepted as being either the total volume of the column, as calculated from π r2 h, or it is the volume measured by adding a measurable salt to the applied sample. The salt can be sodium chloride, detectable by conductivity, or sodium nitrite, detected by its fairly strong absorbance at 280 nm. Gel filtration is, intrinsically, a low resolution separation method for proteins, yet it is frequently used in protein purification. Gel filtration is gentle to the sample and it is the best preparative method for fractionating native proteins by size and shape. Passage though the partially accessible pores in the beads will generate broad elution bands, each band lying within just 25 percent of the total column volume, thus the

intrinsically low resolution of the method. Generally, the highest resolving columns, containing very small beads of soft gel materials, like Sephadex G-100 Super fine or BioGel P-100 minus 400 mesh, operate under low gravitational force fields (50 cm pressure head, or smaller). Beads used for relatively large proteins must have low degrees of cross-linking, making the gels soft and highly compressible. For the most compressible beads, for example, pressure heads may need to be as small as 15 cm. In general, gel filtration columns are able to give baseline resolution for no more than 4 proteins, each differing in molecular weight by a factor of 2. So, under the best of conditions, a mixture of globular proteins of MW 200,000, 100,000, 50,000, and 25,000 Daltons can be baseline resolved.

- **Physical Set-ups in Column Chromatography**

Those new to column chromatography often ask, “What size column should we use and what are the most appropriate dimensions of length and width”? Clearly there is no one correct answer. But there are some appropriate generalities that can help with column selection in adsorption chromatography. Adsorption chromatography includes ion exchange, hydrophobic interaction, affinity chromatography, and all other forms of chromatography in which the analyte binds to the stationary phase (all methods other than gel filtration and SEC). Protein resolution in adsorption chromatography depends upon the rate of change of the eluting solvent, not upon the length or width of the column. Better resolution results from gradual change in the strength of eluting solvent. The limit of “slow rate of change” is no change at all. In chromatography, we call “no change at all” isocratic elution. Isocratic elution, at the right solvent concentration, generates, the highest possible resolution, but peak spreading will be greater in isocratic elution than in gradient elution. In general, the amount of adsorbent in a column should have the capacity to bind three-to-five-times the amount of protein being loaded. The length and width of the column are not critical. It is not unreasonable to use a short, stout column for adsorption chromatography, a column with length 2- to 3-times the column diameter, for example. Such columns allow very high flow rates, so a large volume of eluent can be used in a fairly short period of time. But, one should not greatly extend column width at the expense of length (eg. dimensions of a cake pan are problematic). A wide diameter column, where the eluent exits through a port

at the center of the cylinder, provides early elution of protein that happen to migrate down the center of the column. Protein (of the same type) that migrates near the circumference of the column will exit significantly later. This differential elution (side vs center of the cylinder) produces smearing of a band that might otherwise be sharper (if the column had more “normal” dimensions). An exceedingly long and thin column is not desirable either. Flow rates will be very slow, especially if the gel is soft. If, in an attempt to speed up flow, pressure is increased, the gel may compress and flow will slow down. In extreme cases, flow may stop altogether. Even if the adsorbent is rigid and non-compressible (as in size exclusion HPLC) a column with a small cross-sectional area may over-pressure if particles collect on the surface. It is common to use long, narrow columns for gel filtration, but here as well, columns need not have such extreme dimensions. The problem of particles collecting on the surface of the column may still occur. But even if the sample is particle free, flow rate may be much slower than desired. A 50 cm column, with a diameter of 2.5 cm, can give excellent resolution in gel filtration as well as in adsorption chromatography.

- **HPLC (high performance liquid chromatography)**

High performance liquid chromatography (HPLC) is basically a highly improved form of column liquid chromatography. Instead of a solvent being allowed to drip through a column under gravity, it is forced through under high pressures of up to 400 atmospheres. That makes it much faster. All chromatographic separations, including HPLC operate under the same basic principle; separation of a sample into its constituent parts because of the difference in the relative affinities of different molecules for the mobile phase and the stationary phase used in the separation.

There are following variants of HPLC, depending upon the phase system (stationary) in the process :

1. **Normal Phase HPLC:**
   This method separates analytes on the basis of polarity. NP-HPLC uses polar stationary phase and non-polar mobile phase. Therefore, the stationary phase is usually silica and typical mobile phases are hexane, methylene chloride, chloroform, diethyl ether, and mixtures of these. Polar samples are thus

retained on the polar surface of the column packing longer than less polar materials.

2. **Reverse Phase HPLC:**
   The stationary phase is nonpolar (hydrophobic) in nature, while the mobile phase is a polar liquid, such as mixtures of water and methanol or acetonitrile. It works on the principle of hydrophobic interactions hence the more nonpolar the material is, the longer it will be retained.

3. **Size-exclusion HPLC:** The column is filled with material having precisely controlled pore sizes, and the particles are separated according to its their molecular size. Larger molecules are rapidly washed through the column; smaller molecules penetrate inside the porous of the packing particles and elute later.

4. **Ion-Exchange HPLC:**
   The stationary phase has an ionically charged surface of opposite charge to the sample ions. This technique is used almost exclusively with ionic or ionizable samples. The stronger the charge on the sample, the stronger it will be attracted to the ionic surface and thus, the longer it will take to elute. The mobile phase is an aqueous buffer, where both pH and ionic strength are used to control elution time.

HPLC instrumentation includes a pump, injector, column, detector and integrator or acquisition and display system. The heart of the system is the column where separation occurs.

1. **Solvent Resorvoir** : Mobile phase contents are contained in a glass resorvoir. The mobile phase, or solvent, in HPLC is usually a mixture of polar and non-polar liquid components whose respective concentrations are varied depending on the composition of the sample.

2. **Pump** : A pump aspirates the mobile phase from the solvent resorvoir and forces it through the system's column and detecter. Depending on a number of factors including column dimensions, particle size of the stationary phase, the flow rate and composition of the mobile phase, operating pressures of up to 42000 kPa

(about 6000 psi) can be generated.

3. **Sample Injector** : The injector can be a single injection or an automated injection system. An injector for an HPLC system should provide injection of the liquid sample within the range of 0.1-100 mL of volume with high reproducibility and under high pressure (up to 4000 psi).

4. **Columns** : Columns are usually made of polished stainless steel, are between 50 and 300 mm long and have an internal diameter of between 2 and 5 mm. They are commonly filled with a stationary phase with a particle size of 3–10 µm. Columns with internal diameters of less than 2 mm are often referred to as microbore columns. Ideally the temperature of the mobile phase and the column should be kept constant during an analysis.

5. **Detector** : The HPLC detector, located at the end of the column detect the analytes as they elute from the chromatographic column. Commonly used detectors are UV-spectroscopy, fluorescence, mass-spectrometric and electrochemical detectors.

6. **Data Collection Devices** : Signals from the detector may be collected on chart recorders or electronic integrators that vary in complexity and in their ability to process, store and reprocess chromatographic data. The computer integrates the response of the detector to each component and places it into a chromatograph that is easy to read and interpret.

Applications of HPLC: The information that can be obtained by HPLC includes resolution, identification and quantification of a compound. It also aids in chemical separation and purification. HPLC separation of protein samples prior to mass spectrometric detection presents the most widely used separation method beside capillary electrophoresis. Due to the complexity of proteomic samples, one dimensional separation of peptides or proteins is not sufficient , additional separation dimensions must be introduced in order to reduce the samples complexity. Today nano HPLC is the most widely separation method used for protein sepration due to increase in sensitivity compared to conventional HPLC and the possibility to analyse low amounts of sample with none or very low dilution.

- **Batch Methods for Protein Purification**

Occasionally one finds a batch method that works as well in purifying a particular protein as a variety of chromatographic methods. Batch methods are particularly useful in early stages when a sample is highly viscous or full of fine colloidal material. Such batch methods include ammonium sulfate precipitation, precipitation from other salt solutions, from aqueous solutions at low pH, or from organic solvents (usually acetone, ethanol, ethylene glycol, or polyethylene glycol). In some cases, recrystallization from salt solutions is possible. Even if crystals do not form, differential precipitation can be an effective purification method. A particularly effective batch method is isoelectric precipitation in which the pH of a dilute aqueous buffer is adjusted to the isoelectric point (pI) of the POI. The protein-of-interest, or contaminants in the POI mixture, can be adsorbed to Celite, alumina gels, calcium phosphate gels (hydroxyapatite), and other media. If antibodies are available, the protein of interest can be selectively bound to those antibodies. If aggregates form upon such treatments, the aggregate can be collected by centrifugation and then dissociated into free antibody and free POI by a variety of methods including application of low pH buffers.

- **A-Free IgG**

While exploring repeated rounds of ammonium sulfate precipitation, the purification of rabbit-derived antibodies, goat anti-rabbit IgG, and chicken IgY. Because this process does not utilize Protein-A, and this method is called “A-Free.” For rabbit-derived antibodies, the “A-free IgG” procedure works at least as well as chromatography on columns of Protein-A. Goat-derived antibodies, that are not as amenable to purification on Protein-A columns, and chicken-derived IgY, that cannot be purified on Protein-A at all, respond equally well to the “A-Free” method. Although very commonly used in purifying therapeutic monoclonal antibodies, Protein-A is quite expensive. Despite its being covalently bound to the affinity column matrix, Protein-A is able to leach from the column matrix during the elution phase. Traces of Protein-A in therapeutic monoclonals could present a health hazard, as Protein-A may bind to other essential antibodies in the patient. We have not found formal regulations limiting the use of

Protein-A in purifying therapeutic monoclonals, but manufacturers might prefer a safer, more cost-effective method. We have a satisfactory replacement for Protein-A in the method we call “A-Free IgG.” This method has submitted, through Rutgers University, as a provisional patent application. As often occurs in experimental science, the “A-Free IgG” method of antibody purification arose from an accident. In the course of purifying IgG from rabbit serum by a traditional single round of ammonium sulfate fractionation, a mistake that was picked up by size exclusion HPLC. The SEC profile showed more contaminants than seen previously. So, to remove those additional contaminants, we repeat the entire process. The second round of ammonium sulfate precipitation produced a cleaner IgG sample than previously seen with just one round of precipitation. But the redissolved pellet was still slightly pink (not all the transferrin was removed), and the HPLC profile still showed a tiny shoulder of albumin. So the same precipitation process is repeated third time. This time, the HPLC profile showed 99% pure IgG virtually no high molecular weight contaminant and no indication of any albumin. The SDS gel profile showed strongly stained heavy chain and more weakly stained light chain (normal for IgG) and a very weakly staining contaminant or these side-by-side experiments show that the “A-Free IgG” method actually out-performs Protein-A affinity chromatography. The time involvement is similar and the price is much lower. On occasion we perform a 4th ammonium sulfate precipitation, obtaining a sample marginally cleaner than that resulting from three rounds of precipitation. The method works equally well with goat anti-rabbit IgG, an antibody less amenable to Protein-A purification. We have a large supply of chicken egg yolk containing anti-GFP antibodies (IgY) for which Protein-A is totally ineffective. The A-Free method is suitable with IgY so long as the large amount of lipid has been removed by a freeze-thaw method.

- **Criteria for Protein Purity**

Demonstrating purity of a given protein is not an easy task. But, without achieving protein homogeneity, serious errors and experimental artifacts may arise. Even a 1% contaminant may contribute to erroneous observations. A minor contaminant (protein or otherwise) could significantly raise or lower an enzyme’s apparent activity level. If an impure protein of-interest is used to generate antibodies, a very immunogenic contaminant could induce more antibody than the POI. Some biochemists and some journals will accept, as the sole criterion of purity, a photograph or a densitometry

trace of a Coomassie-stained SDS polyacrylamide gel that shows one stained band. But, a case in which a “single band” on an SDS gel, accepted by a prominent journal as proof of purity, turned out to be a 97% contaminant of the protein-of-interest. The actual POI represented only 1% of the total “pure protein”. Errors of this magnitude can be avoided by using a variety of different criteria for evaluating protein purity.

1) Constant specific activity across a broad portion of the peak in the final preparative chromatography column.

2) Single, symmetric band by size exclusion HPLC.

3) Single band, in the correct MW region, on an SDS gel (or, for hetero-oligomers, the appropriate number of bands in the correct positions.

4) Unambiguous, single amino acid detected in N-terminal amino acid analysis. Inability to detect an N-terminal amino acid may also be taken as evidence of purity—not a very strong criterion as many other proteins have blocked N-terminal amino acids.

5) Single band on a native polyacrylamide gradient gel (or appropriate number of bands of correct MW for heterodimers, heterotetramers, etc).

6) Single, sharp band by isoelectric focusing in an acrylamide gel or in a capillary isoelectric focusing system (or the appropriate number of bands for hetero-oligomeric proteins).

7) Unambiguous N-terminal peptide sequence by Edman degradation.

8) Single band by Western blot, if antibodies are available.

9) Single MW form by Maldi TOF (matrix assisted laser desorption time-of-flight mass spectrometry).