

Protein Purification - I

Today I want to cover aims and strategies in protein purification, and the making of a crude extract. I hope that before the end of the next lecture I will actually be talking about purification methods.

Before starting to describe protein purification in the strict sense, I should review questions of aims and strategies. By aims I mean, what are you purifying the protein for? How much of it do you need? How pure do you want it? Must it be active, or will SDS-denatured protein do? I am generally assuming that you are trying to purify a protein in its native, biologically active conformation - whether or not you have an *in vitro* assay for this activity - and that you are trying to purify it on a research laboratory scale in order to study its properties. As is shown in the figure on the first page of Scopes' review which I will hand out, this is only a part of the spectrum of amounts and uses for which proteins are purified. If one simply wants to determine some sequence of the protein purified from the natural source, in order to design oligonucleotide probes to look for the cDNA or to elicit antibodies which will recognize the protein produced from the cDNA, you can isolate less than 1 μg in a denatured state by gel electrophoresis, transfer to a polyvinylidene difluoride membrane, and sequence; mass spectrometric sequencing uses amounts down to 1 ng. However, you have to be sure that that small sample represents the protein you are interested in. On the other extreme, the detergent industry uses tons of proteolytic and lipolytic enzymes, though naturally it doesn't need them to be highly purified. At the extremes, the techniques you can use become limited, because of the difficulty of detection at the low end and the cost of purification materials and equipment at the high end. But in the range I am most concerned with, from perhaps 0.1 mg final product at the low end to 100 mg at the high end, these constraints are not too important. Usually we want to purify enough to study the protein physically and chemically, as well as determining at least some of its sequence, and want at least 90% purity. Clinical and therapeutic purposes may require 99.999% purity to satisfy the FDA that it won't cause bad reactions in patients. If you need very high purity, you need to start with more crude extract, because you will continue to lose protein when you are removing that last 0.1% contaminant; but only therapeutic proteins need be that pure.

By "strategy" I mean the relationships between the choices of individual steps of the procedure, to maximize efficiency, economy and effectiveness. This will be apparent as we go through methods of purification; for instance, precipitation with $(\text{NH}_4)_2\text{SO}_4$ yields a concentrated but high-salt fraction which would have to be desalted before ion exchange chromatography but is appropriate for a gel filtration column, which will incidentally desalt it but spread it out, make it more dilute. Therefore, the order of steps precipitation/gel filtration/ion exchange is good strategy, the order precipitation/desalting/ion exchange/concentration/gel filtration is not, because it requires additional steps to get the sample ready for the next step. "Strategy" also involves using procedures which are low resolution but can easily and economically be carried out on large amounts of crude material early in the procedure, using expensive, low capacity but high resolution procedures late in the purification. The exception to this is the use of affinity chromatography, defined as binding the protein to some special group on a chromatographic support by some special characteristic of the protein, then eluting it off, by a specific means if possible. This may be appropriate at the beginning of a purification in two extreme circumstances: you have a large amount of very dilute protein solution, such as hundreds of liters of urine from which you want to purify urokinase; or, you have a small amount of extract containing a fairly high amount of protein with a hexahistidine 'tag' at the end of the sequence. In either case an expensive, specially designed and prepared affinity column may be the best procedure to use. The use of immobilized metal ion columns with his-tagged proteins is

now so common that some may wonder why I am teaching anything else; but sometimes it doesn't work.

The first strategic choice, of course, is whether to purify from the native source or from a foreign host in which the protein is produced from an introduced foreign gene - a cloned protein from a recombinant source, for short. (For those who are just taking General Biochemistry and don't know how this is done, just accept that it can be done. Read about it in Rosenberg, pp. 335-350.) This has major advantages over purifying from the native source: you can use a standard cell growth procedure, and most of all you can generally produce much more per gram cells. The main disadvantage is that you or someone must first spend time, probably at least a couple of years, cloning the gene from the natural source - which may include purifying the protein from the natural source, unless the DNA sequence of the same protein from a different source is already known - and maximizing its expression in a foreign host. If the native protein has other groups stuck on it - what we call post-translationally modified - you may not get the same modifications stuck on in the same places. You certainly won't if you produce it in a bacterial host; you have to use insect or mammalian cell culture, which is much slower and more prone to contamination. See Rosenberg, pp. 350-361. I will talk about special problems and procedures of working with recombinant proteins later in this lecture, or in the next one.

Purification of a protein is at each step a compromise between three considerations: retaining activity - this is usually paramount, and includes not having the activity *changed* in any way as well as not *losing* biological activity; maximizing the degree of purification at the step; and maximizing yield, how much of what you had before the step you recover after it. It is the last two which are frequently in conflict, and you often have to decide what compromise to accept. It is essential to determine activity as soon as possible after carrying out a purification step, especially when you are exploring methods, and to calculate the recovery of total units, especially when you change the volume in the procedure; units/ml x ml = total units.

The purification summary includes the following headings:

Step	volume (ml)	units/ ml	total units	mg pro- tein/ml	total mg	specific activity	yield	purif. factor
------	----------------	--------------	----------------	--------------------	-------------	----------------------	-------	------------------

The actual experimental data are volume, units/ml and mg protein/ml. The most important values are **total units**, units/ml x volume, and **specific activity**, units/mg protein, calculated from units/ml divided by mg/ml. Specific activity should increase through the purification, reaching a maximum value when the protein is pure. Yield is total units at a given step divided by total units in the crude extract; it decreases through the purification, and in a long purification the final yield may be only a few %. This is undesirable, but may be unavoidable if purity requires purifying 100,000-fold. Your objective in purification is to maintain total units and yield while steadily increasing specific activity. Purification factor, not always shown in the table, is specific activity at a given step divided by specific activity in the crude extract; the final purification factor may range from two-fold for a well-expressed recombinant protein to 100,000 fold for a minor protein such as the N- and C-proteases in collagen maturation. Total mg protein is mg/ml x volume, and often isn't shown.

You may also calculate recovery and purification *for a single step*, i.e. total units at that step/total units at the preceding step and specific activity at that step/specific activity at the preceding step. I feel that a minimal criterion for a purification step should be that the recovery - normally a number less than 1 - multiplied by the purification - a number greater than 1 - should yield a number greater than one, and preferably much larger. However, this may not be possible late in a purification when you are trying to remove the last 5% contamination, and the purification can't exceed 1.05. [I once worked out a step, for a bacterial cholinesterase, which involved shaking the extract with *sec*-butanol for half an hour at room temperature, which gave 10-fold purification with 144% yield; but that was lucky.]

You often want, particularly for precipitation procedures, to try out the procedure first on very small aliquots of your protein solution, checking recovery and purification on a number of samples which have been treated in different ways. The microcentrifuge is very useful for this, you can use samples of 50 or 100 μL and assay the supernatant.

The per cent recovery is important to know what is going on, especially to recognize removal of an inhibitor or dispersal of a complex. Once I heard Dr. Abou-Sabé talk about glucose stimulation of an adenylyl cyclase. In crude extract the enzyme was stimulated about 80% by glucose, but after $(\text{NH}_4)_2\text{SO}_4$ precipitation glucose didn't stimulate. I asked whether total activity after $(\text{NH}_4)_2\text{SO}_4$ precipitation corresponded only to total **unstimulated** activity in the crude extract, or to total **stimulated** activity. In the first case one would suspect that stimulation involved another factor (not necessarily a protein) which was not precipitated by $(\text{NH}_4)_2\text{SO}_4$ with the adenylyl cyclase; in the second case one would suspect that the crude extract contained an inhibitor of adenylyl cyclase, counteracted by glucose, which was separated from the enzyme by $(\text{NH}_4)_2\text{SO}_4$ precipitation.

Disappearance of activity in a procedure which should not cause inactivation, such as $(\text{NH}_4)_2\text{SO}_4$ precipitation or gel filtration, may be due to separation of two or more factors, all of which are required for activity. The simplest case is loss of an essential metal ion, or a cofactor which stabilizes the protein. [Right now in my lab we are trying to purify a nitroaromatic reductase, which doesn't seem to like being precipitated with $(\text{NH}_4)_2\text{SO}_4$. We tested what stabilized it at 37° , and the cofactor FMN helped, so now we are adding that, especially when redissolving after $(\text{NH}_4)_2\text{SO}_4$ precipitation.] Another way of recognizing this is in the assay for the activity; if this is non-linear with amount of extract at low levels, as if a certain amount of crude extract must be added before *any* activity is seen - I sometimes call this a "concentration lag" - it suggests that interaction of two or more factors is required, or even that the protein must be activated by some other enzyme before it can be active, as with the enzyme I did my thesis on. Another possibility, when the total amount of protein added to the assay is very low, is that some of it is sticking to the walls of the test tube or cuvette and isn't active in the assay. This can be tested and cured by adding a little bovine serum albumin, say 0.1%, to the assay mix to saturate such sites.

An early question, even before preparing an extract, is, where is the protein? For proteins of higher organisms it is intracellular, but if made in a microorganism it may be periplasmic or extracellular. Ideally a cloned protein is secreted from the cell, because there will be very few other proteins secreted and your purification will be a lot simpler, and because in the best cases you can get a lot of it made, it isn't limited by filling up the cell. But this means including a signal sequence in the gene, which directs the cell to extrude the protein through the membrane. If it is periplasmic it will be selectively released from the cell by a little osmotic shock, but the amount present will be much more limited. If it is intracellular you have a full protein purification problem. You have to ask, is it cytoplasmic, and released just by sufficient homogenization, or in an organelle, or membrane-bound? It may be worth purifying the organelle or membranes and then releasing the protein from them; or you may be interested in a specific isoform found in an organelle or membrane, and purify this, generally by differential centrifugation, before extracting the protein from it. Release of proteins from membranes is a problem with which I don't have much experience; generally you get a kit of detergents and try them to see which one releases the protein in good yield with as little other protein as possible. Rosenberg has a whole chapter on membrane-bound proteins and their release, pp. 135-153. You may have to keep detergent present throughout the purification to keep the protein soluble, or you may be able to remove the detergent once you have separated the protein from membrane material, or at least greatly decrease the concentration from that needed to solubilize. See Scopes and Rosenberg on this. I have a sheet from Calbiochem about an adsorbent for removing

detergents. When detergent is present, membrane proteins are likely to be in micelles, and behave as such rather than as true soluble proteins - they will be excluded from gel filtration materials, for instance.

Another question is, what source? I have already mentioned the choice between a native source and a recombinant source. You may not have a choice of a native source: you are specifically investigating a protein in a particular organism, perhaps a particular organ; or an enzyme known only from a particular species of bacterium. But if you do have a choice, you normally choose an organism and tissue which has lots of the protein. Sometimes ease of extraction is more important, or ready availability of the tissue. If it is a microorganism, other than baker's yeast, you will probably have to grow it, which is easy for *E. coli*, more difficult for a slow-growing anaerobe or something growing on an unusual substrate.

Extraction is usually a compromise between breaking up the tissue and breaking open the cells, complete extraction, separation of non-protein material such as cell walls, carbohydrates and fats, and keeping the protein active. The first two are the method, while complete extraction and keeping the protein active depend mainly on the extraction medium. You generally use the gentlest method which solubilizes the protein well. If you want to isolate organelles as an intermediate step, you must have an osmotic support such as 0.25 M sucrose or sorbitol present. Organelles are isolated by differential centrifugation, not only pelleting at different speeds but equilibrium density gradient centrifugation.

For animal sources the prime problem is breaking down the tissue, for plants breaking the cell walls. Sometimes one makes an **acetone powder**, by grinding the tissue in acetone and collecting the precipitated protein. This is mainly useful if the tissue contains a lot of fat, or a lot of water as in fruit. But not all proteins survive in active form. Plant tissues tend to contain phenolic compounds which oxidize and polymerize on extraction, yielding brown polyphenols which precipitate proteins irreversibly. One uses some sort of antioxidant to prevent this, and often also insoluble polyvinylpyrrolidone which precipitates polyphenols.

Bacteria have no tissue, but breaking the cell wall is a problem. There are physical means, such as sonic oscillation, sudden release of pressure in a French press, shaking with very small glass beads, grinding with alumina (a Stone Age procedure). And there are biochemical means - lysis with enzymes such as lysozyme, or with EDTA or non-ionic detergent. These methods are not so good for large-scale work, and yield very viscous solutions due to unsheared DNA.

Fungi and yeast are very tough. Blending with glass beads and pushing through the French press are two methods. Dr. Macmillan once developed a yeast cell wall degrading enzyme from the actinomycete *Oerskovia*, but it was never produced commercially. Snail gut enzymes are used on a small scale. Sometimes bacteria and yeast autolyse - digest their own cell walls - if exposed to organic solvents such as toluene or ethyl acetate. Sometimes repeated freezing and thawing of cells breaks them.

For extraction one usually uses 2 to 2.5 ml buffer per gram wet weight of material, less with plants because there is less cytoplasm. For most materials 0.25 M salt maximizes extraction of intracellular proteins; for bacteria even 0.5 M may be good. You should initially extract a small sample with a large amount of buffer to determine what the maximum amount extractable is, then decrease this amount to the lowest still extracting nearly all the protein. Sometimes a lower ionic strength may extract your protein selectively (rabbit GPDH with 0.03 M KOH). Also, inclusion of 10% glycerol in the medium often improves protein stability - and if you want to store it at -20° 30% glycerol will keep it from freezing as well as stabilizing. Restriction enzymes are usually in 30 to 50% glycerol.

This **crude homogenate** contains many insoluble materials. These are normally removed by centrifugation - filtration is slow and ineffective. Unless one is trying to isolate

mitochondria or other large organelles, one usually centrifuges 10 to 30 min at 18,000 rpm, 40,000 x g, top speed in the Servall refrigerated centrifuge, or in a continuous flow centrifuge for really large amounts. This leaves ribosomes and membrane fragments still in solution. These may be centrifuged down at 100,000 x g for an hour, but the capacity of high speed centrifuge heads is smaller. To remove ribosomes, particularly from bacterial extracts, you can precipitate with polycations such as protamine sulfate, streptomycin sulfate, spermine, or polyethyleneimine, or with manganese salts. Or if your protein is not precipitated or inactivated at pH 5.0, you can adjust to that pH and precipitate ribosomes.

You now have a **crude extract**. At this point you are probably worried about stability of your protein, particularly if you want to go home at night and leave it in the refrigerator. First of all the extract should be well buffered, particularly if it is a plant extract - plant tissues tend to contain acidic vesicles - or may metabolize further. You might at this point test **pH stability** - both at what pH, usually acidic, is it precipitated immediately (maybe isoelectric precipitation, maybe denaturation), and at what pH is it most stable to long term storage: set up samples at various pH values, centrifuge down any immediately precipitated protein and assay the supernatant for activity remaining, then store overnight in the refrigerator and assay again. Two other worries are stability to oxygen - intracellular proteins readily have their cysteine sulfhydryl groups oxidized and lose activity - and protease activity. Dithiothreitol or dithioerythritol, 1 - 5 mM, generally protect against SH oxidation, but sometimes, especially for extracts of anaerobic bacteria, stronger compounds are necessary, or even working under a nitrogen atmosphere. Proteases are dealt with by adding protease inhibitors to the extraction solution - phenylmethanesulfonyl fluoride for serine proteases, pepstatin for aspartic proteases, antipain for cysteine proteases, EDTA or *o*-phenanthroline for metal-dependent proteases. But you need to know whether your protein *requires* a metal ion for activity, and is inhibited by EDTA or other chelators. There are also many proteins which are more or less specific protease inhibitors, but you would hardly use them in purification. See Rosenberg pp. 410-12, or the Boehringer Mannheim guide he copied, for more details.

You usually keep the protein cold whenever possible - though there are cold-labile proteins. Proteins from thermophilic organisms may not need to be kept cold, except to keep other bacteria from growing on them. Na azide will inhibit bacteria, though also heme enzymes like peroxidase.