

Special problems of purifying recombinant proteins

I have already made some mention the production of recombinant proteins in *E. coli* and other hosts; here I want to address the special problems involved. If your protein is present only in very low concentrations in its natural source, or the source is not readily available, such as humans, an alternative is to clone the gene and produce the protein in a bacterium, yeast, or mammalian or insect cells in tissue culture. This is discussed in Rosenberg, pp. 335-350. Pages 350-361 discuss use of eukaryotic cells, though much is on getting the foreign DNA into the host cell.

If a bacterial host is used, there are two big problems, whether necessary post-translational modifications occur, and whether the proteins accumulate as insoluble "inclusion bodies", which typically also include the 4 subunits of RNA polymerase, the outer membrane proteins OmpA, C and F, and ribosomal RNA. It is not entirely clear why this happens, but accumulation of a large amount of a single protein at high concentration is probably a major factor. The protein in inclusion bodies typically is misfolded, and will have either incorrect disulfide bonds or intermolecular bonds rather than intramolecular. The proteins are released from these inclusion bodies only by powerful denaturing agents, 6 M urea or 8 M guanidinium HCl; they must then be renatured by dilution and dialysis away of the denaturing agent, typically in steps, followed by concentration of the dilute protein. During this process formation of incorrect disulfide bonds frequently occurs. This problem can be minimized in principle by adding protein disulfide isomerase, which corrects such problems in the cell, or more practically by passing through a column of thiol-Sephadex, which was shown almost 30 years ago to be very effective in reducing protein disulfides, and appears to act like a disulfide isomerase; or by carrying out the renaturation in the presence of an empirically established mixture of oxidized and reduced forms of a thiol compound - this was done well by one of our undergraduates working at Schering. Another problem is that if there are any proteases present, they just love chewing on unfolded proteins, and must be inhibited by inhibitors such as PMSF. See Rosenberg pp. 339-340 for a sample procedure.

It is better to avoid forming the inclusion bodies in the first place. This can be done by: producing the protein in eukaryotic cells - though even this doesn't guarantee soluble protein, it may be stuck in the Golgi apparatus, and in any case this is slower, more prone to contamination, and yields lower concentrations of the protein. Secreting the protein from cells, after genetically fusing a secretory signal peptide to the amino terminal of the protein, is better. *Bacillus subtilis* and streptomycetes are much better secretors than *E. coli*, if you clone the gene in an appropriate vector. Sometimes simply growing the cells at 30° rather than 37° helps a lot, for unknown reasons, not connected with incipient heat denaturation, as even heat-stable proteins such as RNase A can form inclusion bodies at 37°. It may be possible to improve solubility by specific mutations in the protein which do not affect activity but change the solubility; negatively charged and hydrophilic proteins are secreted better than positively charged, hydrophobic proteins. It is hoped that co-cloning of 'foldases' will prevent IB formation in the cell, but they don't prevent it in lysates.

When a gene has been cloned, it can be fused with another gene, so that a fusion protein is produced with properties of two proteins, or a bit of sequence can be added. The first reason for this is that the fusion protein is induced with the inducer of the fused bacterial protein - for instance isopropyl thiogalactoside for β -galactosidase. But it is now normal simply to clone the gene in a vector with the desired promoter, whether or not β -galactosidase is produced. The second is that the fusion protein can be recognized by either the activity or the antigenicity of the fused bacterial protein. And fusion can be taken advantage of for purification of the cloned protein, especially if the other protein can be separated by a specific proteolytic cleavage. For instance, the cloned gene can be coupled with that for a maltose-binding protein, and the fusion protein can then be isolated by adsorption on an amylose column. The maltose-binding protein used for this has a C-terminal linker sequence which is cleaved specifically by

the blood-clotting protease Factor Xa, to allow the cloned protein to be freed from the fusion. Similarly, the cloned gene can be coupled with glutathione S-transferase, the fusion protein isolated by adsorption on an immobilized glutathione column, eluted with free glutathione, and cleaved by thrombin, which like Factor Xa is fairly specific for cleavage at exposed arginine residues, probably in specific sequences. However, there is always the possibility that the cleaving protease will also cleave the desired protein somewhere.

Another type of modification of a cloned protein is the attachment of a hexahistidine peptide. This interacts tightly with immobilized Cu^{++} , Ni^{++} or Co^{++} ions, chelated by iminodiacetate or other chelator attached to a solid support), permitting separation from almost all other proteins, followed by elution with EDTA to compete for the Cu^{++} or imidazole to compete for Ni^{++} . Such a method can be used to purify the protein while still unfolded. While I haven't worked with any of these methods myself, I think the hexahistidine method is probably best, because it usually doesn't interfere with the function of the protein and so the tag doesn't have to be removed and you don't have to take a chance with proteases. It's my impression that this is winning out as a method for purifying cloned proteins. See Rosenberg, pp. 345-6. Qiagen sells expression vectors and Ni-NTA columns for the purpose. I gave you copies of their material, and of a competitor's.

However, one may not attach any tag, and purify the protein by conventional means, thus being sure that the tag does not affect the protein's activity.

Yet another method mentioned in Rosenberg is the FLAG[□] epitope, the sequence Asp-Tyr-Asp-Asp-Asp-Asp-Lys, which is bound by an antibody in presence of Ca^{++} . It is then eluted with EDTA, and the FLAG cleaved off with the protease enterokinase, which recognizes this sequence, which is essentially the amino terminal sequence of trypsin, naturally activated by enterokinase.

A final remark is that protein stability may be modified by site-specific mutagenesis, but "such mutations are rather easy to produce but time-consuming to characterize; even selective mutagenesis may deplete the graduate student supply long before all the possibilities are exhausted."