Lecture 12

Protein Expression

When you're ready to start "engineering proteins", there are terrific ways of getting them expressed.

Expression of proteins in vitro Can ribosomes be trained to do their thing in a test tube? Yes indeed! This method pre-dates the advent of recombinant DNA techniques by many years.

For many applications, one may use a rabbit reticulocyte or wheat germ lysate to synthesize proteins in vitro. One simply adds an RNA template of choice, and lets the ribosomes attach and run their course. In the old days, one would have to treat rabbits with something like phenylhydrazine to induce formation of reticulocytes (immature red blood cells), then spend an afternoon bleeding, lysing, and freezing the cells. Now it's easier just to get a purchase order number and have the cells "overnighted".

Here are two examples of translation-competent lysates you can buy:

Rabbit Reticulocyte Lysate Translation Systems are utilized in the identification of mRNA species, the characterization of their products and the investigation of transcriptional and translational control.

Rabbit Reticulocyte Lysate is prepared from New Zealand white rabbits using a standard protocol which ensures reliable and consistent reticulocyte production in each lot. The reticulocytes are purified to remove contaminating cells which could otherwise alter the translational properties of the final extract. After the reticulocytes are lysed, the extract is treated with micrococcal nuclease to destroy endogenous mRNA and thus reduce background translation to a minimum.

The lysate contains the cellular components necessary for protein synthesis (tRNA, ribosomes, amino acids, and initiation, elongation and termination factors).

References:
Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247.

(from the Rabbit reticulocyte lysate technical manual): http://www.promega.com/tbs/tm232/tm232.html

Cell-free extracts of wheat germ support the in vitro translation of a wide variety of viral, prokaryotic and eukaryotic mRNAs.

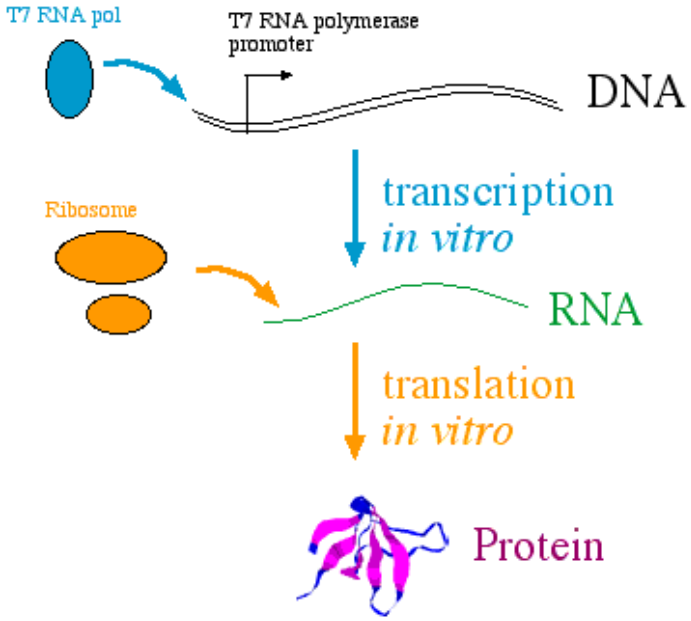
Promega's Wheat Germ Extract is prepared by grinding wheat germ in an extraction buffer followed by centrifugation to remove cell debris. The supernatant is then separated by chromatography from endogenous amino acids and plant pigments which inhibit translation. The extract is also treated with micrococcal nuclease to destroy endogenous mRNA and thus reduce background translation to a minimum.

Wheat Germ Extract contains the cellular components necessary for protein synthesis (tRNA, ribosomes, initiation, elongation and termination factors).

(from the Wheat germ extract technical manual): http://www.promega.com/tbs/tm230/tm230.html

How might you make use of these lysates and extracts?

Suppose you have performed some biochemical studies on a purified enzyme, and believe you have cloned the gene encoding that protein. How do you demonstrate this? One way would be to take an RNA transcript from the cloned sequence and translate it in vitro. The scheme would look like this:



The protein synthesized in the rabbit reticulocyte lysate or wheat germ extract may have enzymatic activity, or may be recognized by an antiserum you have developed against the native protein. In either case, you will have generated data that shows a correspondence between your "protein of interest" and your "candidate gene sequence."

We'll talk more about transcription a bit later, so for now we'll concentrate particularly on the translation extract.

Here are instructions for one particular in vitro protein synthesis kit (from Fermentas Corporation) which are downloaded so that you can see how the components are used:

Experimental Procedure: *In Vitro* Translation:

C.2.2 Set-up of *Modulis*™ translation with purified mRNA

Modulis™ translation mix	12 µL
^{35}S-Met (10.2mCi/mL;1175 Ci/mmol)	2 µL
Purified RNA	x µL
10X Adaptor buffer	2 µL
DEPC-treated water	4- x µL
	20 µL

- mix all of the above components in a microcentrifuge tube on ice
- incubate at 30ºC for 30 min
- analyze translated products by denaturing gel electrophoresis and/or filter TCA assay
- the translated products may be stored at -20ºC for future applications

Notes:

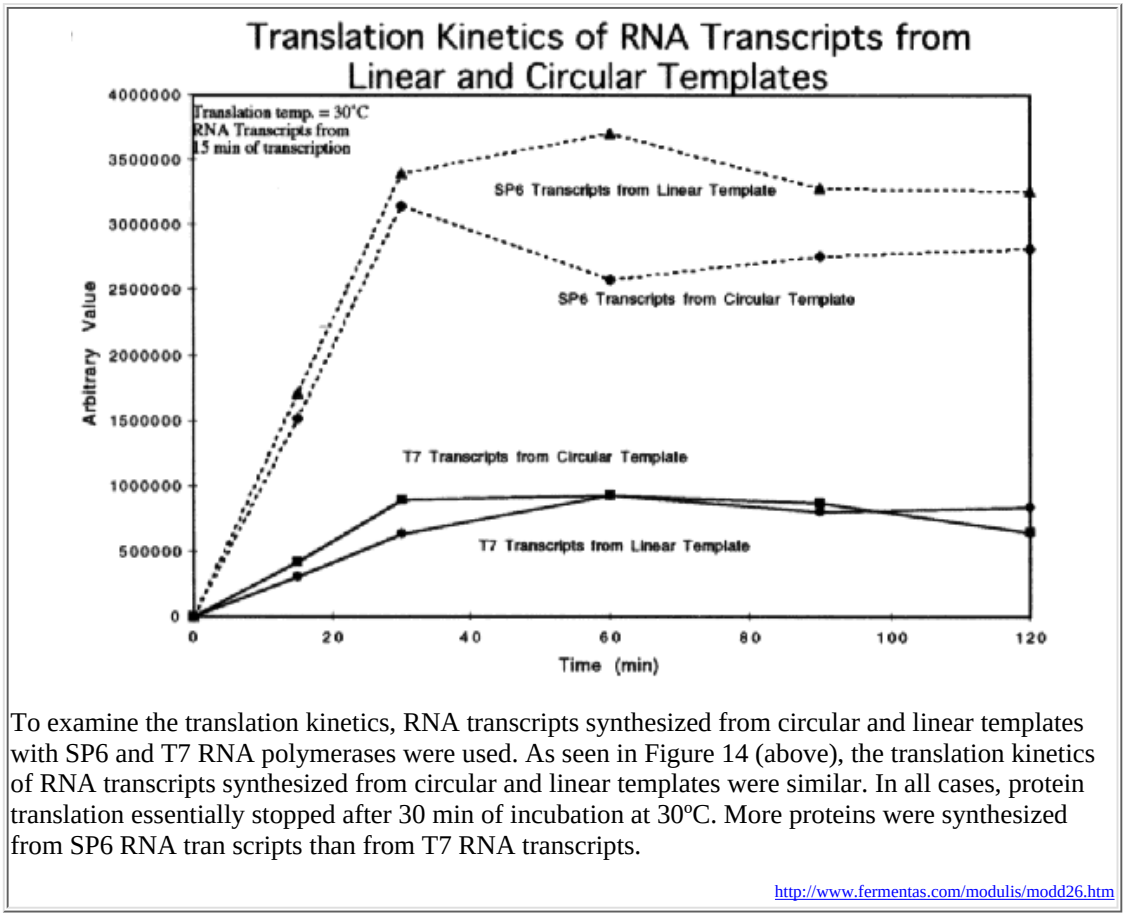
1. **Important:** The translation mix should be thawed quickly by holding the tube between fingers. Once thawed, it should be kept on ice while setting up the reaction mixture. **It is very important not to thaw the translation mix in water baths.**

   After use, the translation mix should be quick-frozen with liquid nitrogen and stored at -70ºC. If the above procedures are followed, the translation mix can be frozen and thawed at least six times without affecting the translation activity.

2. The amount of purified RNA added to a standard *Modulis*™ translation reaction should be between 2 - 6 µg for SP6 transcripts or 1 - 3 µg for T7 transcripts.

http://www.fermentas.com/modulis/modc22.htm

How rapidly does the in vitro protein synthesis reaction progress? Here's an example of data downloaded from Fermentas:



To examine the translation kinetics, RNA transcripts synthesized from circular and linear templates with SP6 and T7 RNA polymerases were used. As seen in Figure 14 (above), the translation kinetics of RNA transcripts synthesized from circular and linear templates were similar. In all cases, protein translation essentially stopped after 30 min of incubation at 30ºC. More proteins were synthesized from SP6 RNA tran scripts than from T7 RNA transcripts.

http://www.fermentas.com/modulis/modd26.htm

Post-transcriptional modifications - are they important?

The synthesis of protein from a DNA sequence in vitro takes two steps, of course. The first is transcription of an RNA copy, perhaps using a phage RNA promoter, and the second is the translation of a protein. You might have several questions about whether post-transcriptional modifications are important for RNA sequences from eukaryotes:

- **Polyadenylation.** In vitro protein synthesis does not depend on having a polyadenylated RNA, but if having a poly(A) tail is essential for some other purpose, one may use a vector that has a stretch of about 100 A residues incorporated into the polylinker region. That way, the poly(A) tail is "built in" by the synthetic method.

- **5' end capping.** Eukaryotic ribosomes read RNAs more efficiently if they have a 5' methyl guanosine cap. If you need to have an RNA that is capped, you have the choice of either incorporating a cap by initiation of transcription using a capped base analogue, or adding a cap in a separate in vitro reaction post-transcriptionally.

Post-translational modifications - are they important?

Here's where you discover the real disadvantages in the synthesis of proteins in vitro. Proteins need to be folded properly to be functional, and often times they are cleaved and modified as a normal process of expression. In vitro reactions can fall short on both counts.

- **Correct protein folding.** Do you feel lucky today? That's about all you have to go on as far as protein synthesis in vitro is concerned. If your aim is only to generate a synthetic protein that can be sized on an SDS polyacrylamide gel (a process that denatures the protein) or perhaps used in Western analyses, then that's a simple matter. If your aim is to generate a protein that has the full functional activity of a native product then that's quite another thing!

- **Cleavage, phosphorylation, myristylation, glycosylation, etc...** It is possible to use in vitro lysates that are slightly fancier (than a rabbit reticulocyte lysate) and contain mammalian cell microsomes (endoplasmic reticulum and other membranous compartments). Proteins may then be synthesized, translocated and processed appropriately, but this is still a bit of an "iffy" proposition.

If you are working on a project where you need to have specific modifications in your protein of interest, then perhaps the in vitro approach will not be right for you.

Combined transcription/translation reactions - doing it all in one tube.

Why run two reactions when one will do? Combined transcription/translation systems are available, in which both phage RNA polymerases (such as T7 or SP6) and eukaryotic ribosomes are present. One example of a kit is the TNT® system from Promega Corporation.

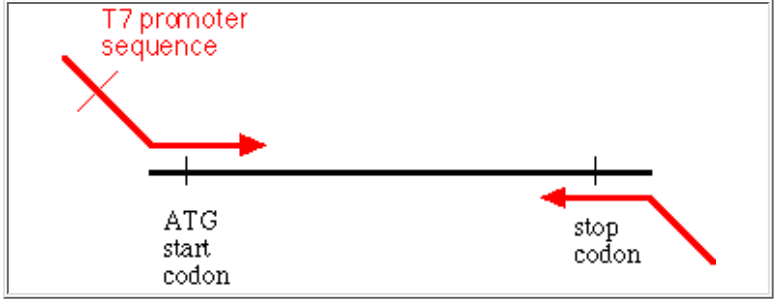
Here is a description from the Promega Corp. web site, for how the TNT® system is used:

> To use these systems, 0.2-2.0µg of circular plasmid DNA (or linear DNA for the T3 and T7 systems) are added directly to TNT® Lysate and incubated in a 50µl reaction volume for 1-2 hours at 30°C. The synthesized proteins then are analyzed by SDS gel electrophoresis (SDS-PAGE) and autoradiography.
>
> http://www.promega.com/tbs/tb126/tb126.html

An example Here's an example of how an experiment might progress, using several of these methods. Suppose we are studying the response of the mammalian immune system to an infectious agent, and we recognize that a particular protein in the organism is immunogenic (i.e. recognized by the host immune system, which makes antibodies directed against it). If we've cloned and sequenced the gene for this immunogenic protein, how can we map the epitopes (regions on the protein that are recognized)?

We could mutagenize the gene in a plasmid, make a small library of the different products, then use the polymerase chain reaction to isolate the coding sequence from each plasmid candidate. While amplifying the sequence, we may tack on a polymerase promoter by just adding it to the 5' end of the synthesized oligonucleotide.

Addition of the T7 promoter sequence to a coding region, by PCR



We will then have a collection of different PCR fragments, each having a T7 RNA polymerase promoter connected to a mutagenized coding region for the gene of interest. We can now simply put that DNA template into a transcription/translation extract (for example, one made by Ambion Corporation), and collect the specific protein an hour or so later. These proteins can be tested with immune sera (from humans) on a Western blot to see which modifications have changed the antigenicity of the protein. Consequently, we can rapidly map the epitopes involved in antibody recognition.

Of course, the caveat mentioned previously still holds! Any post-translational modifications of the protein (such as phosphorylations or glycosylations) will not take place, so the epitopes one can detect are defined by the choice of protocol.

Expression of proteins in vivo What are the advantages of expressing a protein of interest in vivo?

- You can easily make milligram quantities of your protein
- There is a reasonable chance that the protein will be correctly folded
- You may increase the liklihood that it is correctly processed post-translationally

Of course we have already been exposed, to a small extent, to the idea of expressing proteins in vivo.

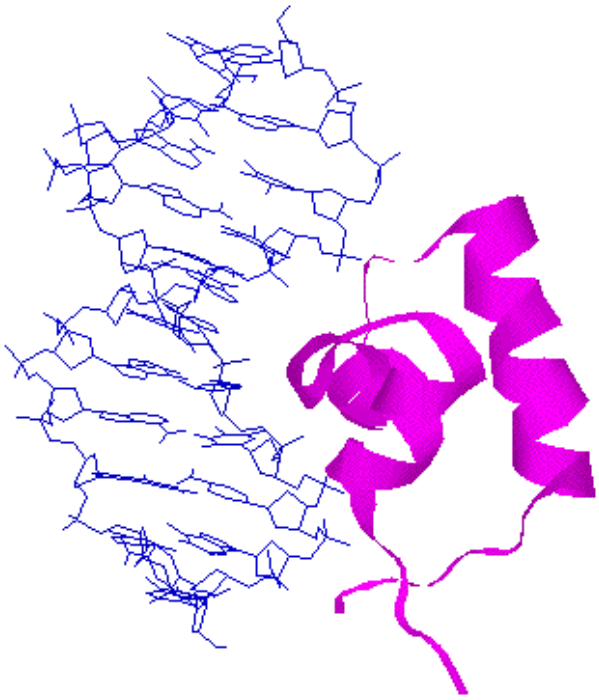
- Recall that in the laboratory, we are using a pGEX-based system to express mutagenized peptides as a fusion protein.
- Recall that the blue/white screening method and allied methods are fundamentally cases of protein expression, albeit of a reporter gene.

The choice of host

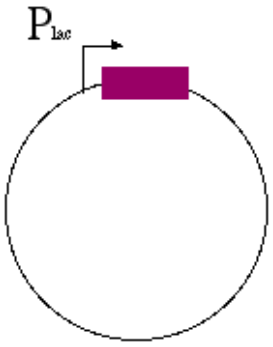
Why regulate at all?

It may seem odd that one would want to regulate the expression of a protein, if the entire purpose is to "make buckets of it." As we've discussed however, many proteins turn out to be toxic when they are overexpressed in a cell. We need to have some ability to turn off expression, simply so that the cells stay alive long enough.

We are already very familiar with the regulation of the lac operon, and its derepression with IPTG. Here's the crystal structure lac repressor complexed with DNA (PDB code 1lcd):



When we use the lac operon for regulation of gene expression in a vector, one way is represented by the pGEX vector, where a "lac" or "tac" promoter is used as an inducible promoter:



You can refresh your memory of the regulatory system of the lac operon on John W. Kimball's page. The lac operon regulatory region is pretty small. In fact, here's the guts of it, with the transcribed sequence (the part that becomes RNA) shown in red..

```
TAGGCACCCCAGGCTTTACACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA...
ATCCGTGGGGTCCGAAATGTGAAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTT...
```

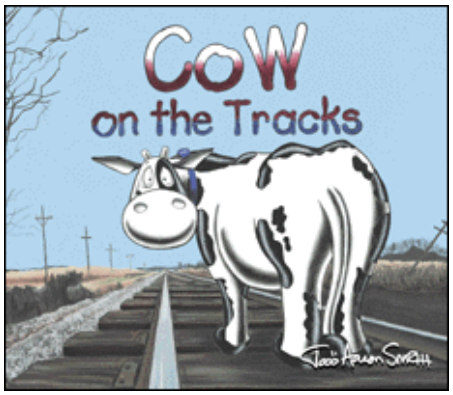
Here's the part that represents (roughly) the RNA polymerase binding site (in green), based on RNA protection studies.

```
TAGGCACCCCAGGCTTTACACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA...
ATCCGTGGGGTCCGAAATGTGAAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTT...
```

Finally, here's the part that represents (roughly) the lac repressor binding site (in blue)

```
TAGGCACCCCAGGCTTTACACTTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA...
ATCCGTGGGGTCCGAAATGTGAAAATACGAAGGCCGAGCATACAACACACCTTAACACTCGCCTATTGTTAAAGTGTGTCCTTT...
```

So you see, the lac repressor binds and really gets in the way of the RNA polymerase. Like a cow on the tracks!



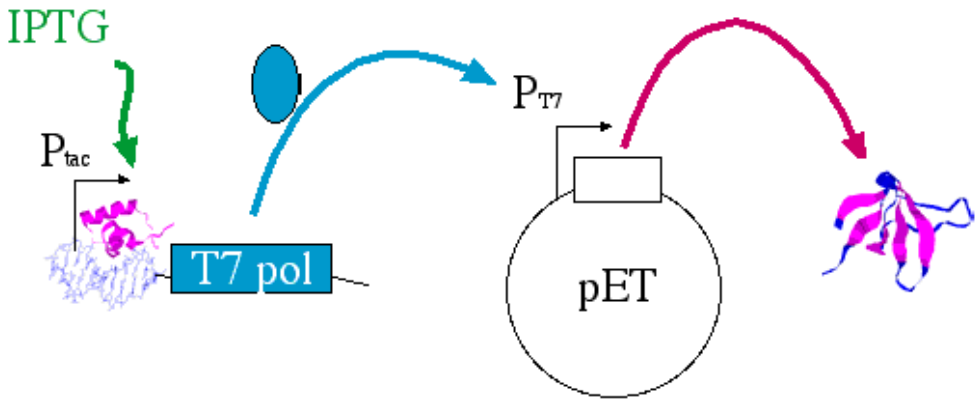
(Cow on the Tracks, by Todd Aaron Smith)

How do we get this cow off of the tracks? Using IPTG as the carrot.

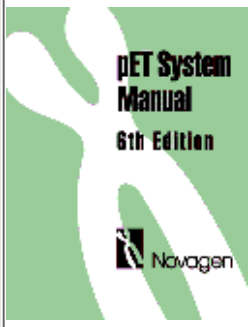
Besides pGEX, the lac (tac) operon is used to control expression from pET.

- In the case of pGEX vectors, the regulation is direct, in that the gene of interest is cloned directly in front of a lac promoter and operator region.
- In the case of pET vectors, the regulation is indirect. The gene of interest is under the control of a T7 RNA polymerase promoter, and IPTG induces expression of the T7 RNA polymerase.

Follow the bouncing arrows to make your recombinant protein:



From Novagen's pET System Manual



The pET System* is the most powerful system yet developed for the cloning and expression of recombinant proteins in E. coli. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 transcription and (optionally) translation signals; expression is induced by providing a source of T7 RNA polymerase in the host cell.

T7 RNA polymerase is so selective and active that almost all of the cell's resources are converted to target gene expression; the desired product can comprise more than 50% of the total cell protein after a few hours of induction.

Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the uninduced state.

Target genes are initially cloned using hosts that do not contain the T7 RNA polymerase gene, thus eliminating plasmid instability due to the production of proteins potentially toxic to the host cell. Once established in a non-expression host, plasmids are then transferred into expression hosts containing a chromosomal copy of the T7 RNA polymerase gene under lacUV5 control, and expression is induced by the addition of IPTG. Two types of T7 promoter and several hosts that differ in their stringency of suppressing basal expression levels are available, providing great flexibility and optimizing the expression of a wide variety of target genes.

The pET vectors were originally constructed by Studier and colleagues (Studier and Moffatt 1986; Rosenberg et al. 1987, Studier et al. 1990). The newer pET derivatives developed at Novagen have been designed with enhanced features to permit easier subcloning, detection and purification of target proteins. Two general categories of vectors are available:

1. Transcription vectors, which are designed for expression of target genes that already carry their own prokaryotic ribosome binding site and ATG start codon.
2. Translation vectors, which contain the highly efficient ribosome binding site from the phage T7 major capsid protein.

In general, the translation vectors are used for the expression of target genes derived from eukaryotic sources, whereas the transcription vectors are used for prokaryotic genes (which usually carry compatible ribosome binding sites).

http://www.novagen.com

Purification of GST fusions (pGEX)

Here is a good place to have already planned ahead! You really have two choices when expressing a protein in a host. You may either express it as a fusion protein, meaning that your sequence of interest is fused to a "tag" or handle of some sort, or you may express your protein in its native state, meaning that the ribosome is initiating translation at the AUG start codon of your gene. Some cloning vectors offer both scenarios as possibilities, and you make your choice at the time you clone your gene.

What are the advantages of having a gene expressed, starting with its own AUG start codon?

- If you are interested in the function of the gene, you may be relieved not to have "uninvited" residues attached. They might confound your interpretation of the function of your gene product.

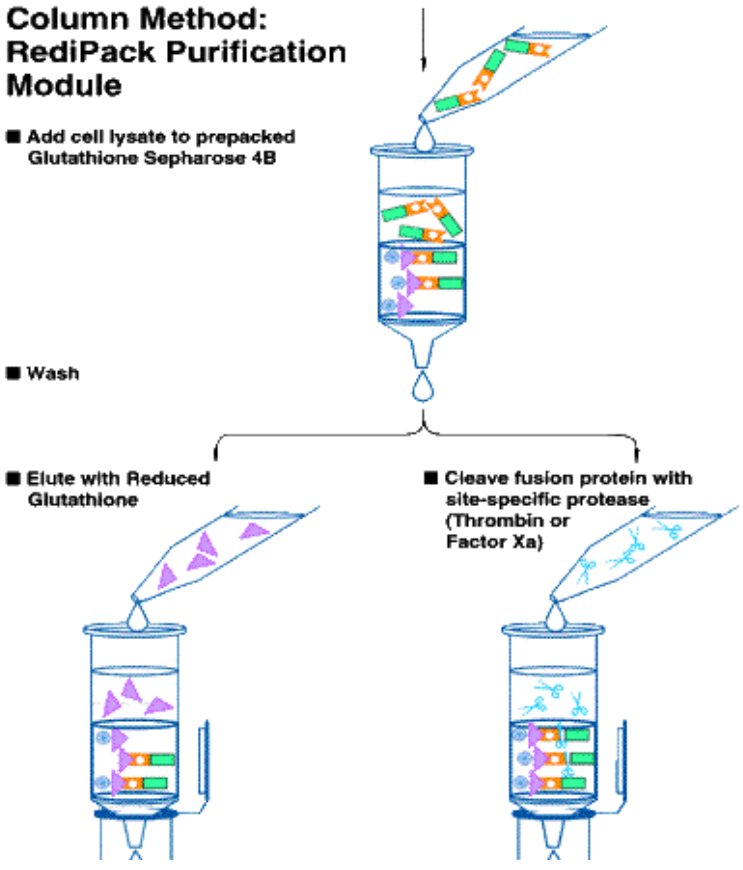
What are the advantages of fusion proteins?

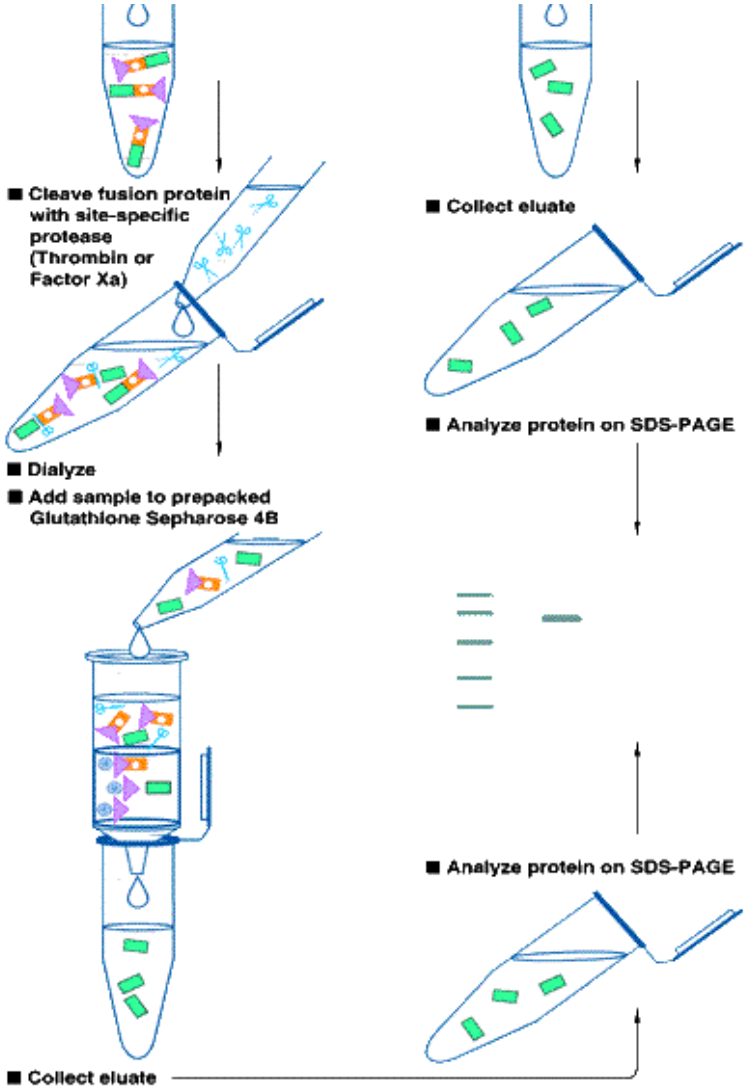
- The chimera may be more stable or more soluble in the host cell, than would be the native protein.
- The fusion partner may be a convenient "tag" for detection or purification.
- The fusion partner segment may be cut off using a sequence specific protease such as thrombin, factor Xa, or enterokinase.
- If you are dealing with only small segments of a coding region in a gene, a fusion protein prepared with those segments can make the products large enough to see on a gel. They may also act as a hapten to ensure that the segments you care about are immunogenic.

Here is an example of a purification method, based on the GST (glutathione S transferase) fusion system of pGEX.

Suppose this schematic diagram represents the fusion protein, with the N-terminal GST part shown in orange and the protein of interest shown in green.

Then we can follow the purification of the chimeric protein on glutathione sepharose, the so-called "RediPack" method from Pharmacia.





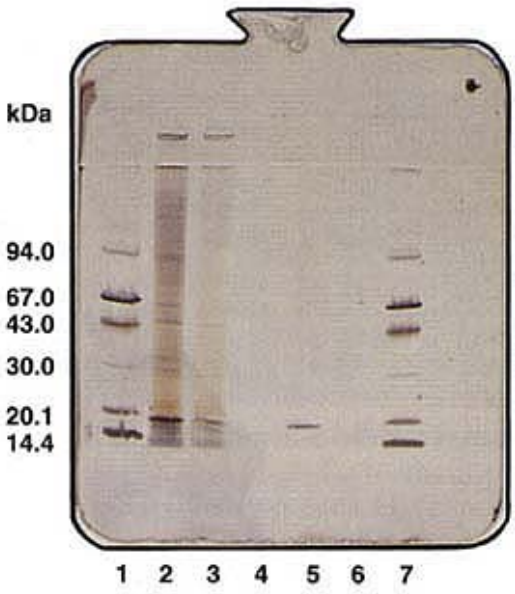
Expressed fusion proteins are easily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Cleavage of the desired protein from the fusion product is achieved using a site-specific protease whose recognition sequence is located immediately upstream from the multiple cloning site (MCS). The GST System has been used successfully in many applications such as molecular immunology, the production of vaccines and studies involving protein-protein and DNA-protein interactions.

Glutathione Sepharose 4B consists of the ligand glutathione, coupled via a 10-carbon spacer arm to the oxirane group of epoxy-activated Sepharose 4B.

Source: http://www.biotech.pharmacia.se/gene_xpr/gene_xpr.htm

A different type of tag One very popular method for purifying proteins is the "His tag" system. Your gene of interest is cloned as a fusion protein with 6 to 10 consecutive histidine residues as a "tag" on the amino or carboxy terminus. When you are ready to purify your protein product, you take advantage of the fact that consecutive histidines can join forces to bind divalent cations such as nickel. You can apply your lysate to a nickel chelation resin, where the his-tagged protein will stick, and wash away the unbound proteins that you don't want. You then elute your tagged protein with imidazole as a competitor. This system is popular because it works under a variety of conditions. You may bind and release a protein from the column under gentle conditions that maintain structure and function, or you alternatively under harsh denaturing conditions (as when you are trying to "persuade" inclusion bodies to go into solution). A small 2.5 ml metal chelation resin column has a capacity of 20 mg recombinant protein.

Samples: 8 ml cell extract containing $(His)_{10}$-tagged protein
The clone was a kind gift from Dr. C. Fuller and S. Brasher, Department of Biochemistry, University of Cambridge, UK
Column: HiTrap Chelating 1 ml, Ni^{2+}-loaded according to the instructions
Start Buffers: 1 × Phosphate Buffer, 100 mM imidazole, 8 M Urea
Elution Buffers: 1 × Phosphate Buffer, 500 mM imidazole, 8 M Urea
Flow rate: ≈ 4 ml/min, 624 cm/h
Instrumentation: Syringe



Lane 1: Low Molecular Weight Calibration Kit (LMW)
Lane 2: Starting material, cell extract, dil. 1:20
Lane 3: Flow-through, dil. 1:10
Lane 4: Wash
Lane 5: Elution Ia (two first ml)
Lane 6: Elution Ib (two last ml)
Lane 7: LMW

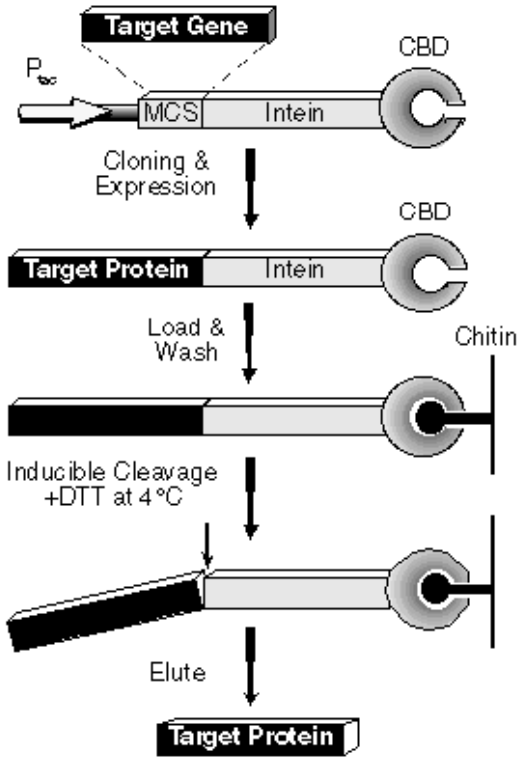
Purification of $(His)_{10}$-tagged protein from inclusion bodies in 8 M urea using HisTrap and syringe operation. SDS electrophoresis on PhastSystem using PhastGel 10-15 and silver staining.

http://www.novagen.com

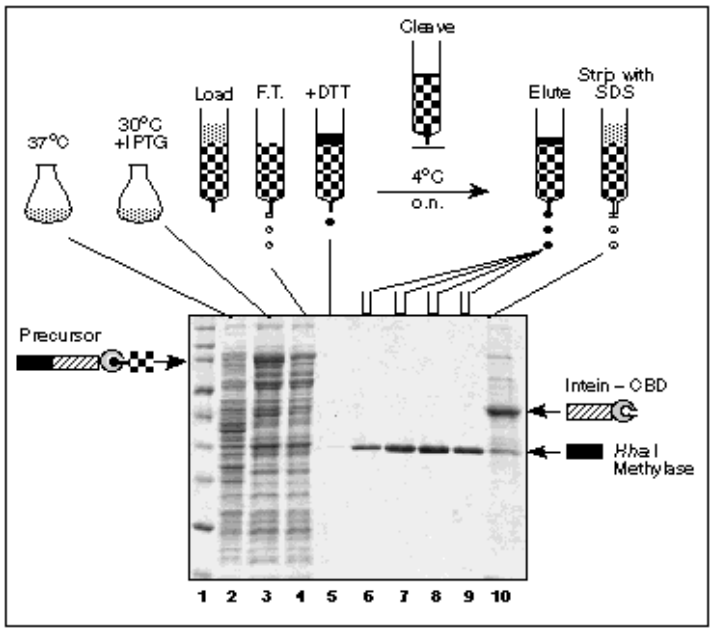
Cleavage of fusion proteins You have several choices, when working with fusion proteins, for how to separate the fusion partner from the peptide sequence of interest. With some vectors, there are sites engineered just upstream of the point of fusion, that allow digestion with specific proteases such as thrombin, factor Xa, or enterokinase. The pGEX vectors have these capabilities, as did the pLIC (ligation independent cloning vector).

A new idea: Proteins that cleave themselves!

The IMPACT system from New England Biolabs



Here's an example: Purification of a Hha I methylase, using the Impact I system.



References

1. Chong, S. et al. (1997) Sindle-column purification of free recombinant proteins using a self-cleavable

affinity tag derived from a protein splicing element. Gene 192, 271-281.

2. Chong, S. et al. (1996) Protein splicing involving the Saccharomyces cerevisiae VMA intein: the steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J. Biol. Chem. 271, 22159-22168.

http://www.neb.com/neb/products/pfp/impactT7_frame.html

Stan Metzenberg
Department of Biology
California State University Northridge
Northridge CA 91330-8303
stan.metzenberg@csun.edu