# Lecture 15

If we have a protein, how do we find its gene?

It's the BIG question, and it will take several lectures to give a satisfactory answer.

Remembrance of cDNA. We have already discussed how to make cDNAs with reverse transcriptase. Today we're going to extend the discussion to discuss how RNA is extracted and purified from cells, and how temperate bacteriophage vectors are used to clone cDNAs.

Total RNA Isolation The total RNA in a cell is a mixture of rRNA, tRNA, snRNA, and mRNA. About 95-98% of the mass is rRNA, however so extraction of total RNA may give a gel that looks like this:
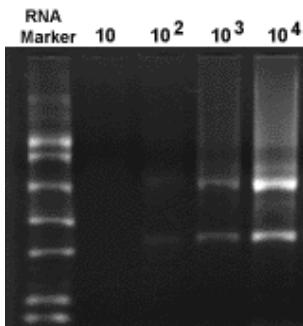

http://www.zymor.com/r1005_minirna_print.html

From: "ZYMO RESEARCH. Figure 2. Total cellular RNA isolated from 10 to 10^4 cells using the Mini RNA Isolation System."

Note the the total RNA appears to be a smear with two prominent bands. These bands are the large and small rRNA molecules, and the smear is probably precursors and mRNA of various sizes from the cell.

Well, here's another example for you:

The GStract^{TM} Total RNA Isolation Kit
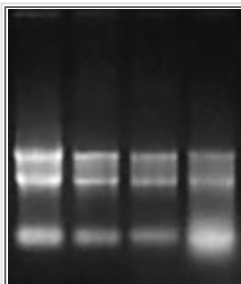
Total RNA isolated from Mouse tissues using EXT-0003.
lane 1: Mouse brain total RNA.
lane 2: Mouse heart total RNA.
lane 3: Mouse lung total RNA.
lane 4: Mouse spleen total RNA.

http://www.maximbio.com/totalrnaisolation.htm

Once again, you see two large bands which are the large and small rRNA. This time you also see a smaller band, which is probably the tRNA fraction. The point of these gels is to indicate what total RNA looks like when it is extracted and is in "good shape" (i.e. not degraded).

If you are interested in rRNA, this is terrific, but what if you are working on expression of protein coding genes? In a eukaryote you have a tool you can use, and that is to take advantage of the poly(A) tail that nearly every mRNA carries. We discussed this in a previous lecture, in

the context of making a first strand of a cDNA.

First Strand cDNA synthesis

Extension of an oligo(dT) primer using reverse transcriptase, an RNA-dependent DNA polymerase of retroviruses

You may make use of the poly(A) tail for purification as well, using oligo(dT)-cellulose as an affinity purification substrate. You prepare a column (or a batch job) using oligo(dt) cellulose and the mRNA selectively binds to the oligo(dT) at high ionic strength. You may release the mRNA by lowering the ionic strength.

Here are some specifications from the Worthington Inc. Catalog, just as an example of an oligo(dT) cellulose product that one might use in the laboratory:

> # Worthington Catalog
>
> Oligo (dT)-Cellulose
>
> Source: Thymidylic acid and cellulose
>
> Oligothymidylic acid is bound to microcrystalline cellulose through the free hydroxyl group. Oligo (dT)-cellulose is particularly useful in the affinity purification of polyA-rich mRNA. It can be used either in batchwise separation or in column and spin-column chromatography. When fully hydrated, one gram will swell to 3-4 ml bed and will bind >40 A260 units of polyA in 10 mM Tris, pH 7.5 and 0.5 M NaCl.
>
> [NOTE THAT ONE A260 UNIT IS APPROXIMATELY 40 ug RNA]
>
> Uses: Isolation and purification of mRNA from mammalian cellular RNA.
>
> Reference
>
> Gilham, P.T., J. Am. Chem. Soc., 86, 4982 (1971).
>
> http://www.worthington-biochem.com/priceList/O/OligodTC.html

Expression of a specific RNA in tissues

As you are no doubt aware, some mRNAs are "housekeeping" and are expressed at similar levels in nearly every tissue. Often however, you may be interested in the RNAs that are not expressed at even levels betwen tissues, because those differences in expression may be functionally significant.

It is really very simple: RNA steady-state levels in a cell are defined by rates of

- Transcription

- Degradation

Here is an example of a commercially-available northern blot from Origene Inc., with mRNAs from 12 human tissues shown. This blot has been probed with the housekeeping gene actin, which was the basis for normalization of the samples.)The RNAs were diluted, that is, so that the actin mRNA was at the same concentration.)

Some RNAs are found at high steady state levels because they are transcribed at a high rate, others at high levels because they are very stable in the cell.

Catalog #: HB-2010

http://www.origene.com/noblot.htm
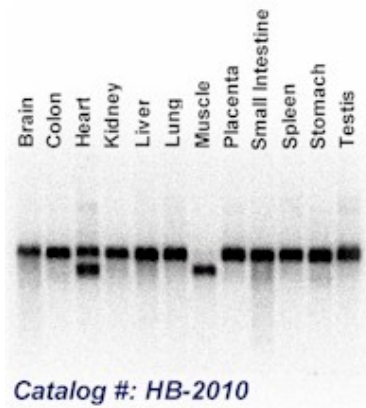
This raises an important point about RNA - how do you prepare northern blots so that you are representing the amount of a specific mRNA?

- normalize to total cell number in the extracted material?

- normalize to total RNA (essentially rRNA)?

- normalize to poly(A)+ RNA?

- normalize to a housekeeping RNA such as actin?

The results will vary depending on which of these methods is chosen. Often times, people will load a specific amount of poly(A)+ RNA on a northern blot from each of several tissue types, then once the first northern is visualized they will re-probe the northern with a housekeeping gene to establish that similar amounts were loaded, and that the RNA is intact.

Keep it clean

RNase contamination can be a major problem when working with RNA. RNase is a very simple enzyme that easily self-folds after it has been denatured. Sure, you can put it in the autoclave and heat it to 121 C for 20 minutes, and it will denature, but as the autoclave is cooling it will refold into functional RNase again. Did you ever see the movie Terminator-2? That's how you should think about RNase contamination.

Here are some common methods of preventing or dealing with RNase problems.

- Wearing gloves and changing them often

- Use of new plasticware

- DEPC-treated glassware and water

- alkali-treatment of glassware (0.2 M NaOH overnight)

- phenol extraction of samples

- Good water supply (not moldy!)

- Baking glassware at 250 C overnight - turns RNase to ash.

- RNasin, or similar inhibitors

Here's a word or two about RNasin, from Promega Inc. A key point is that it requires DTT for activity, so you must have DTT present if you are using RNasin to protect your RNA sample.

Description

Recombinant RNasin® Ribonuclease Inhibitor, developed in the laboratories of Promega, has broad spectrum RNase inhibitory properties including the inhibition of eukaryotic RNases of the neutral type (1). The 50kDa protein exerts its inhibitory effect by binding noncovalently to RNases in a 1:1 ratio with an association constant greater than 1014 (2). Promega's RNasin® Ribonuclease Inhibitor is tested for DNA exo- and endonuclease and RNase activity.

Feature(s)

- Inhibits common eukaryotic RNases including RNase A, RNase B, RNase C and human placental RNase.
- Does not inhibit RNase H, S1 Nuclease, SP6, T7 or T3 RNA Polymerase, AMV or M-MLV Reverse Transcriptase, Taq DNA Polymerase, RNase T1 or RNaseONE; Ribonuclease.
- Active over a broad pH range; requires DTT for activity.

Application(s)

- Useful in any applications where eukaryotic RNase contamination is a potential problem.
- Protection of mRNA in cDNA synthesis reactions.
- In vitro transcription/translation systems.
- Increased yields and activity of polysomes.
- Improvement of in vitro virus replication.
- Improvement of RNA translation in homologous systems.
- Preparation of RNase-free antibody.
- Enhancement of RNA yields from Riboprobe® System RNA synthesis reactions.

http://catalog.promega.com/catalog/catinfo.asp?idx=1031

Storage of RNA - must be done properly for samples to be stable

- Purified RNA can be stored in an aqueous solution at -70 for many years. If it is RNase-free when frozen, it will remain intact.

- Ethanol precipitated can be stored at -20C, either dessicated or in an ethanolic mixture

The BIG question Now we are ready to tackle a big question. Suppose you have purified a protein from an organism and want to obtain the cloned gene. How would you go about finding it?

Well, the answer depends a bit on the organism you are studying. Is the sequence of the organism already determined? If so, you may wish to obtain a bit of protein sequence by Edman degradation.

Purified protein → Proteolytic fragments → N-terminal sequence

PRETTYFLY
MCHAMMER
CANTTCHTHIS
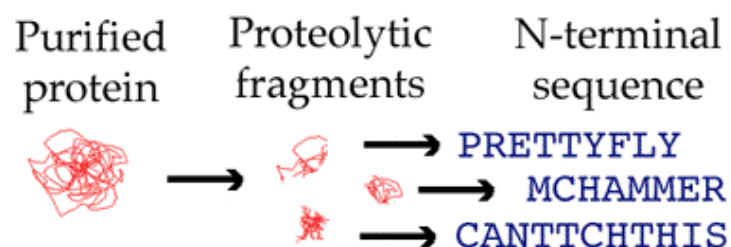
Once you have the protein sequence for a few peptide fragments, you can easily go to the sequence database for the sequenced organism and figure out what gene (or genes) might encode it. Then, you will have a pretty good idea of the boundaries of the coding sequence, and you can set about cloning the gene from genomic DNA (if a prokaryote) or cDNA (if the gene has introns).

Edman degradation is not such a bad approach, even if you don't have the full sequence for the organism. With a bit of scattered protein sequence data for a gene, you can make degenerate oligonucleotides to try to generate a segment of the gene by PCR. For example, the protein sequence PRETTYFLY could be encoded as follows, with the possible codon differences indicated vertically.

```
 P   R   E   T   T   Y   F   L   Y

CCA CGA GAA ACA ACA TAC TTC CTA TAC
 G   G   G   G   G   T   T   G   T
 C   C       C   C           C
 T   T       T   T           T
     A                       T
```

Why are the oligos degenerate? Because the genetic code is degenerate. Given a protein sequence, there are a number of ways it might be encoded in nucleic acid, and you want to give each way a chance of working. Obviously, if your protein is rich in amino acids like arginine, leucine, or serine, each of which have six codons specifying incorporation of the amino acid during translation, it will be a more difficult task. The example above includes 11 points of degeneracy, and an overall degeneracy of:

$$(4)(2)(4)(2)(4)(4)(2)(2)(2)(4)(2) = 65,536$$

When you sequence the N-termini of a series of proteolytic fragments, you don't know the order in which the fragments actually appear in the protein. It could be:

PRETTYFLY --- CANTTCHTHIS --- MCHAMMER

or

CANTTCHTHIS --- PRETTYFLY --- MCHAMMER

or

PRETTYFLY --- MCHAMMER --- CANTTCHTHIS

To clone a segment of the gene, you have to make pairs of degenerate PCR oligos and try a few different orders, like this (where the oligos are meant to indicate PCR of the segment of the gene):

PRETTYFLY --- CANTTCHTHIS --- MCHAMMER

```
   5'---->                              <----5'    Will it work?   Who knows?
```

or

CANTTCHTHIS --- PRETTYFLY --- MCHAMMER

```
   5'---->                              <----5'    Will it work?   Who knows?
```

or

PRETTYFLY --- MCHAMMER --- CANTTCHTHIS

```
   5'---->                              <----5'    Will it work?   Who knows?
```

With a segment of the gene cloned by PCR, you have a usable probe and can isolate the remainder more easily. How do you know if you've got the right piece of DNA? Well, suppose the second trial PCR happened to produce a product:

CANTTCHTHIS --- PRETTYFLY --- MCHAMMER

```
   5'---->--------------------------

            --------------------------<----5'
```

You sequence it, and find that it encodes PRETTYFLY in the interior coding sequence. Your PCR oligos were specifying two of the proteolytic fragment sequences, and you found the third fragment encoded between them in the gene. Joy! It worked! You're going to graduate after all!

A different approach Sometimes this just isn't the right way to find a gene, but there is another common way. You can develop an antibody directed against the protein, and use a cDNA expression system. Remember that if the mRNA for the gene is at a low steady-state level in the cell, the cDNA will also be at a low level in the collection of products. Cloning the cDNA is one way to identify the product unambiguously, and people often use bacteriophage vectors for the purpose. We'll learn more about that in the next lecture, but here's where we start the story:

From the finest sewers of Paris, direct to your benchtop!

The usual picture we carry in our minds of viral infection, is one of death and destruction. We think of viruses that enter a cell, take over and subvert the macromolecular machinery, and produce hundreds of progeny viruses that explode out of the dying cell.

Virulent infections = certain death!
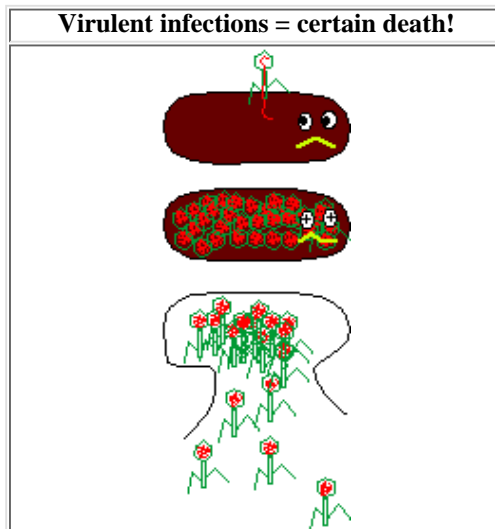
In fact, there are a variety of viral life cycle strategies, and not all of them involve certain death for the cell. We discussed one such example in a previous lecture, when we looked at the filamentous bacteriophages such as M13, fd and f1.

Persistent infection by filamentous bacteriophage
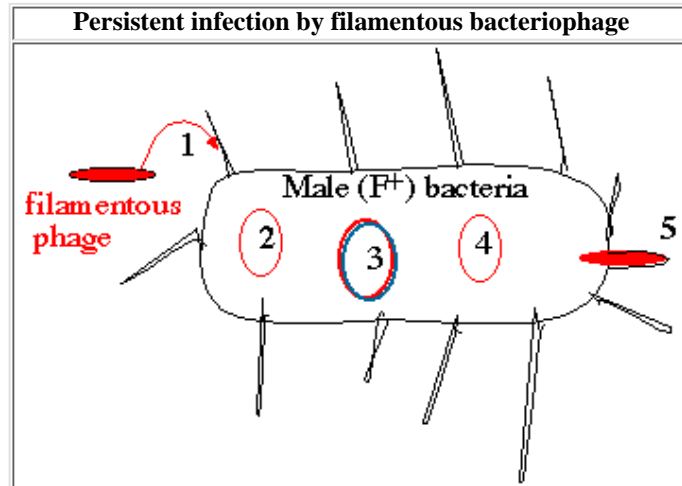
As you recall, those viruses infect male (pili +) bacteria, injecting a single stranded DNA genome that is converted to a double stranded plasmid. The infected cell is not killed, but rather produces and secretes viral particles indefinitely. The infected cell pays "a price" in that its growth rate is slower than that of its uninfected peers.

Temperance in the face of adversity The **temperate** bacteriophage are different still, in that they may enter a cell and produce no progeny virus whatever! They lie silently, allowing the DNA replication machinery of the cell to copy their genomes during the course of the normal cell cycle, and having little discernable effect on the health of the host. At some point, in response to an environmental trigger, the virus leaves its cryptic state (see #7-8 below) and enters a lytic cycle that leads to host cell death and virus release (see #3-6 below).

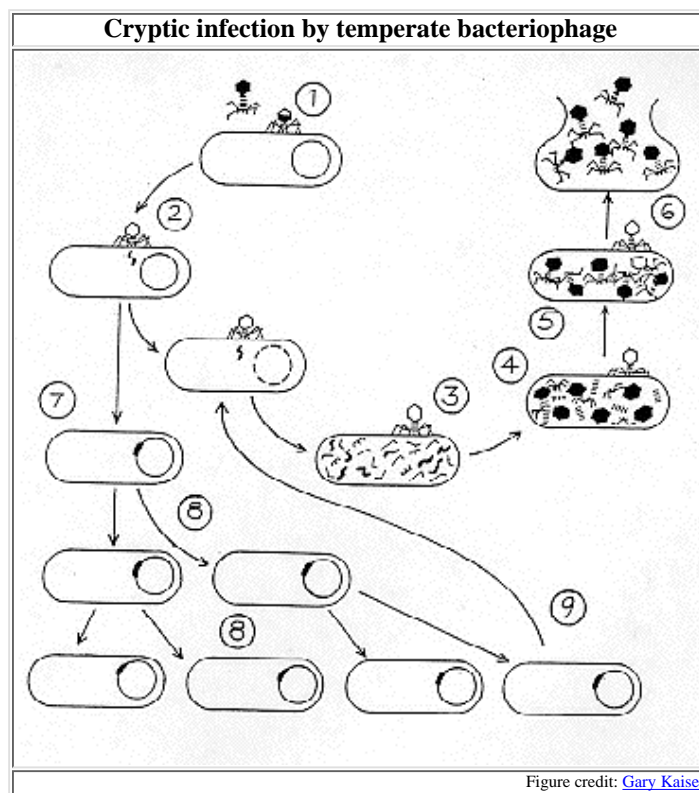**Cryptic infection by temperate bacteriophage**

Figure credit: Gary Kaiser

The most thoroughly studied temperate bacteriophage is lambda, which was pulled out of a Paris sewer 50 years ago by Lwoff, Jacob and Monod. They found that certain strains of E. coli, when

exposed to ultraviolet light, generated viral plaques on a plate - that is, small areas of bacterial lysis. The word "**lysogeny**" was coined to describe this type of cryptic infection. Plaques generated by lambda are typically "**turbid**" rather than "**clear**" because after an initial round of lytic growth on rapidly dividing E. coli, there follows an overgrowth of bacteria carrying lambda lysogenically. These lysogenic bacteria are "**immune**" to lytic lambda infection, because they already harbor the virus!

The nature of immunity was discovered through the isolation of "clear mutants" of lambda that were unable to enter a lysogenic state. These mutated phage were completely virulent and only able to cause a lytic infection. Thus the name "clear" to describe the characteristics of their plaques on a plate. Three loci in the lambda genome could harbor mutations leading to a "clear" phenotype, and these were named cI, cII, and cIII. Infection of a lysogenic bacterium with clear mutants of lambda did not lead to lysis, however! E. coli lysogenic for lambda was immune even to clear mutants.

Virus	Host	Result
wild type lambda	lambda-free E. coli	turbid plaques
	lysogenic E. coli	no plaques
cI mutant of lambda	lambda-free E. coli	clear plaques
	lysogenic E. coli	no plaques

In the next lecture, we'll learn how bacteriophage lambda can be used to answer the BIG question.

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