PLNT2530
2025

Unit 6c

Finding a Gene in a Library

Molecular Biotechnology (Ch 4)

Analysis of Genes and Genomes (Ch 6)

Two scenarios

A. You know the gene you are trying to isolate

B. You know the trait but don't know what the gene responsible is.

A. When you know the gene you are seeking

Screening a library

- Direct screening screen the library for the DNA sequence using a nucleotide probe

- Indirect screening of an expression library for the presence of the gene product

How do we get the probes?

a) heterologous probe: When a gene has been isolated from a closely related species, that gene sequence will likely share sufficient sequence homology to the gene from your plant of interest to be able to form a stable hybrid.

eg. Most cereal genes share considerable homology

wheat, barley, oat, rice, maize, rye

eg. Finding additional members of multigene family using one gene

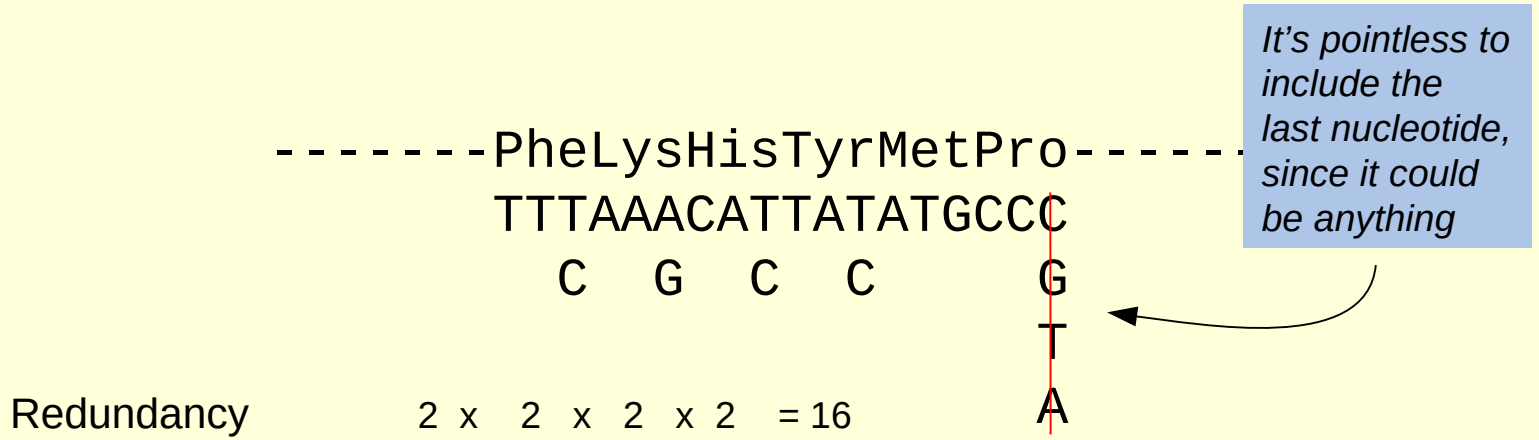
Highly conserved genes are suitable for even distantly related sources

– eg. histones, rRNA, ubiquitin, actin

How do we get the probes?

b) Use a synthetic oligonucleotide probe based on the amino acid sequence of the protein gene product (if this information is known). The oligonucleotide probe sequences are based on the codon sequences for an amino acid sequence in the protein

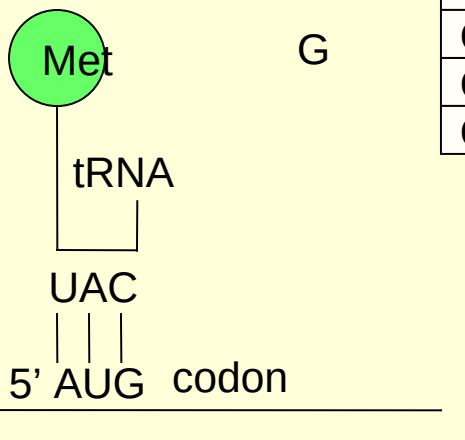
The synthetic oligonucleotide probes are normally redundant 17-26 nts long



There are 16 potential sequence combinations – only one correct for each gene!

Genetic code

First base	Second base: U	Second base: C	Second base: A	Second base: G
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys
	UUA Leu	UCA Ser	UAA STOP	UGA STOP
	UUG Leu	UCG Ser	UAG STOP	UGG Trp
C	CUU Leu	CCU Pro	CAU His	CGU Arg
	CUC Leu	CCC Pro	CAC His	CGC Arg
	CUA Leu	CCA Pro	CAA Gln	CGA Arg
	CUG Leu	CCG Pro	CAG Gln	CGG Arg
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser
	AUC Ile	ACC Thr	AAC Asn	AGC Ser
	AUA Ile	ACA Thr	AAA Lys	AGA Arg
	AUG Met	ACG Thr	AAG Lys	AGG Arg
G	GUU Val	GCU Ala	GAU Asp	GGU Gly
	GUC Val	GCC Ala	GAC Asp	GGC Gly
	GUA Val	GCA Ala	GAA Glu	GGA Gly
	GUG Val	GCG Ala	GAG Glu	GGG Gly



Oligonucleotide probes are normally redundant mixtures

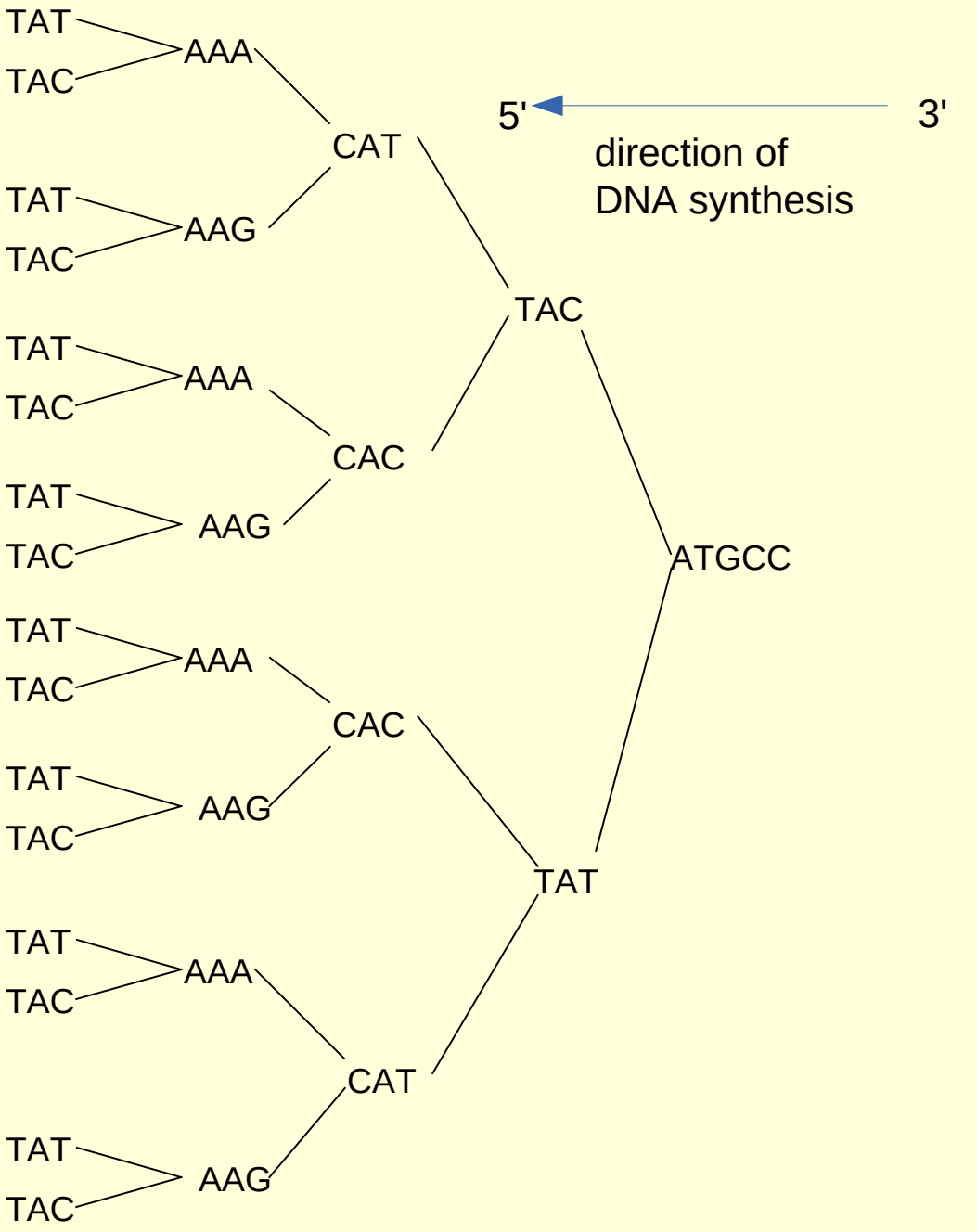
-------Phe Lys His Tyr Met Pro-------
TTT AAA CAT TAT ATG CC
C G C C

TTtAAgCAtTAcATGCC
TTcAAgCAtTAcATGCC
TTtAAaCAtTAcATGCC
TTcAAaCAtTAcATGCC
TTtAAgCAcTAcATGCC
TTcAAgCAcTAcATGCC
TTtAAaCAcTAcATGCC
TTcAAaCAcTAcATGCC
TTtAAaCAtTAtATGCC
TTcAAaCAtTAtATGCC
TTtAAgCAtTAtATGCC
TTcAAgCAtTAtATGCC
TTtAAaCAcTAtATGCC
TTcAAaCAcTAtATGCC
TTtAAgCAcTAtATGCC
TTcAAgCAcTAtATGCC

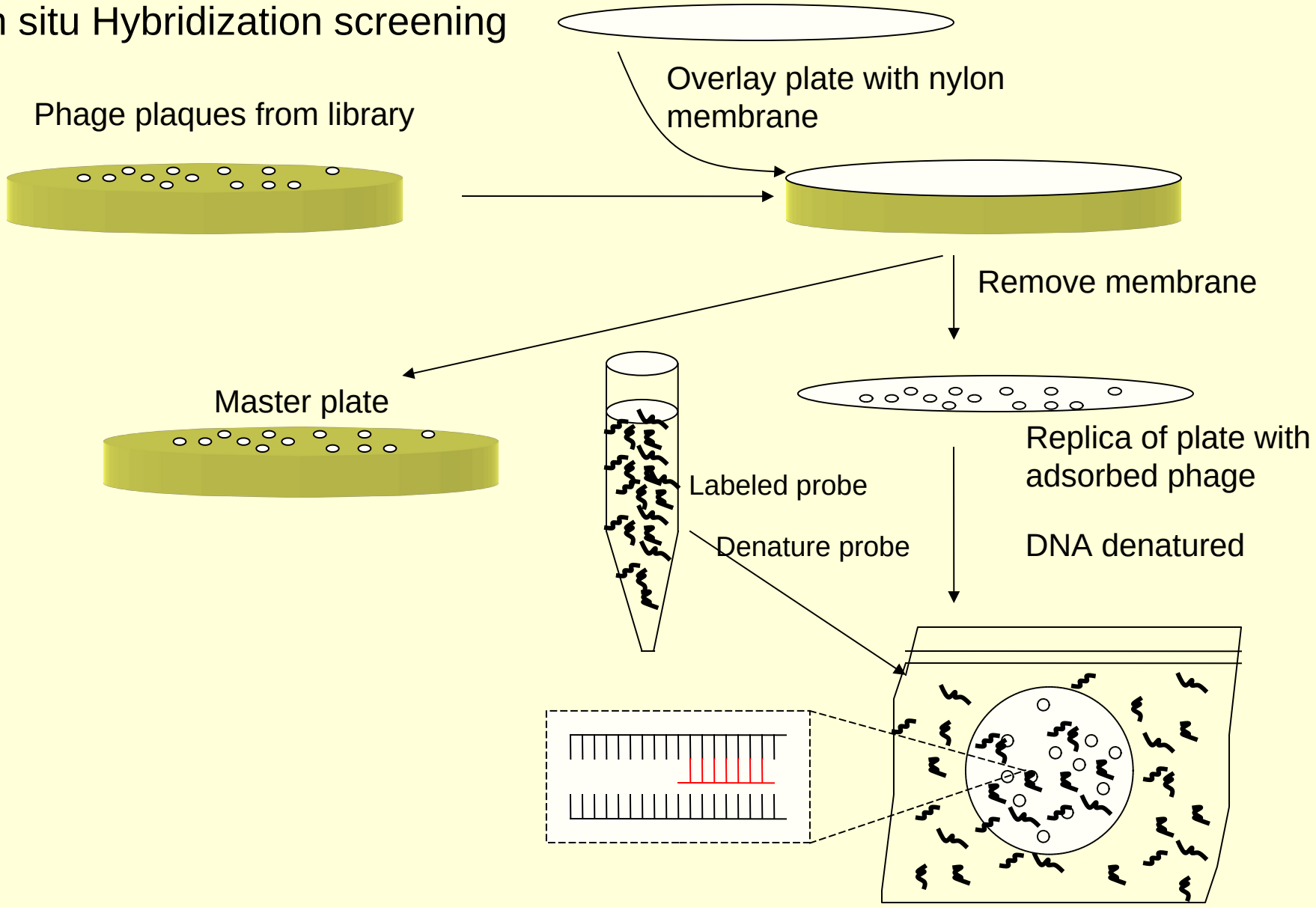
Oligonucleotides are normally labeled using polynucleotide kinase (PNK), which adds a phosphate group to the 5' end.

Chemiluminescent or flourescent nucleotides can be used for end labeling.

Synthesis pattern of oligonucleotide from the 3’ end with the 16 fold redundancy shown

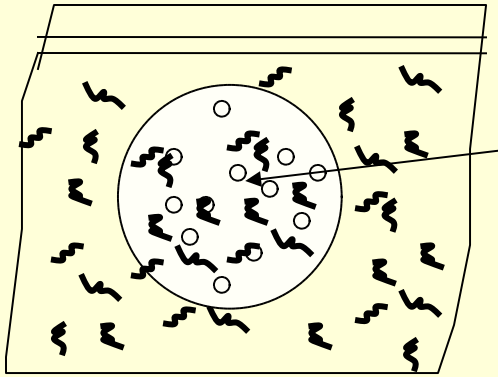


In situ Hybridization screening



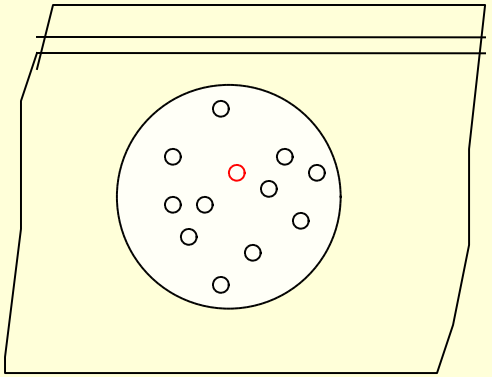
Hybridize probe to filter membrane under salt and temperature to find near perfect match

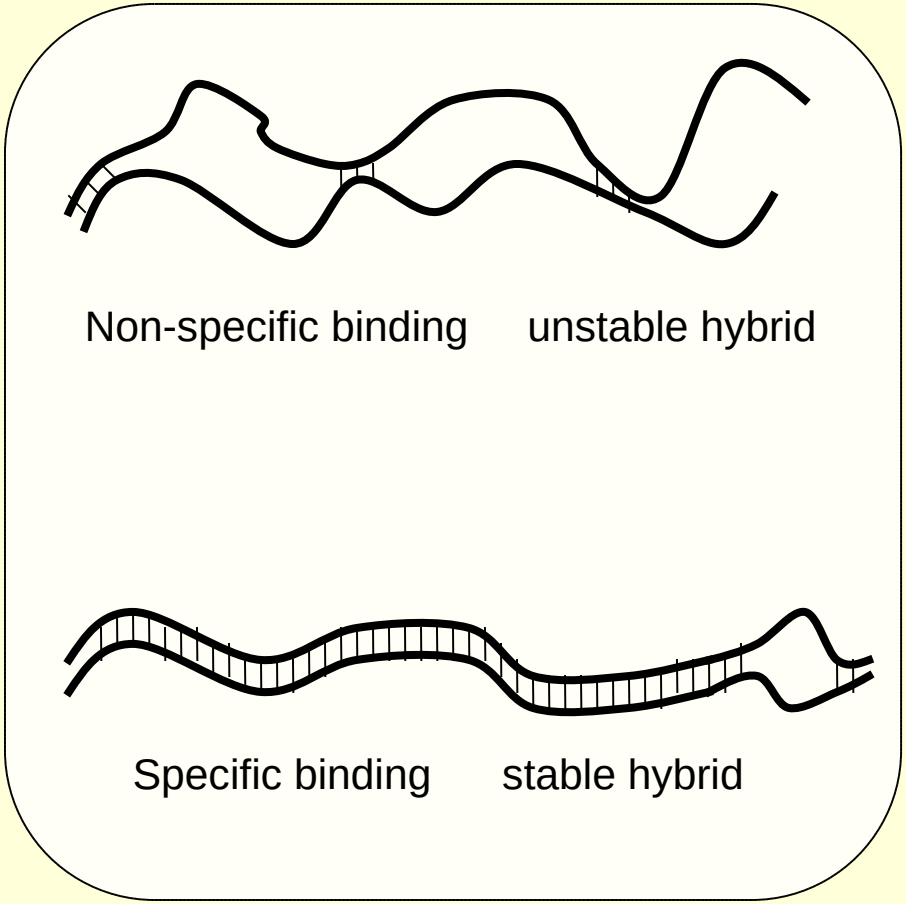
Hybridization screening



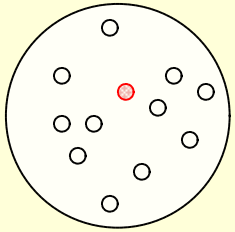
Probe hybridizes specifically to phage insert

Remove excess probe and wash membrane to remove non-specific binding

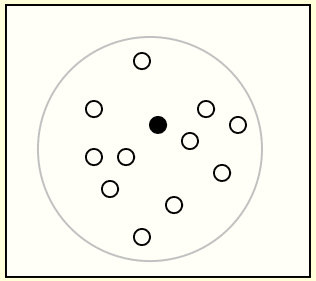




Hybridization screening

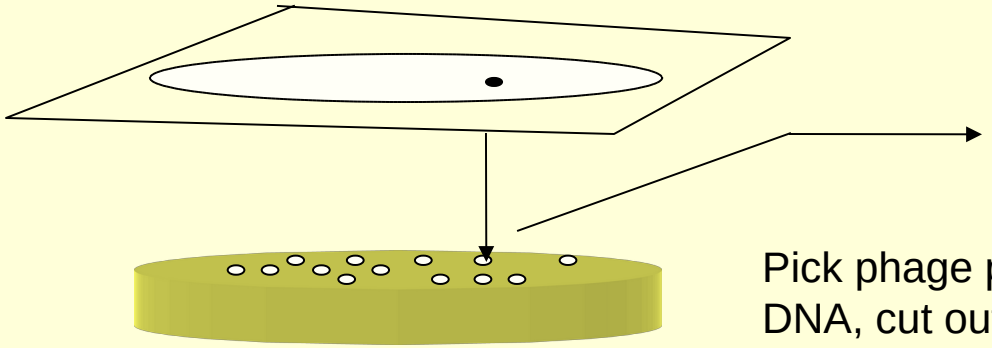


Expose to X-ray film



Exposed film blackened by region with bound chemiluminescent probe

Align film to master plate to find correct plaque



Master plate

Pick phage plaque and isolate phage DNA, cut out insert and transfer insert to a plasmid for further work

Generalizations -- hybridization screening

- If the vector used for the cDNA library was a plasmid rather than a lambda vector, the same process would be used except you would be dealing with colonies on the plate rather than plaques (colonies would have to be lysed on the membrane to release the DNA).

- Oligonucleotide probes are labeled with polynucleotide kinase at the 5’ end only (1 label per probe molecule)

- Heterologous probe (much larger) would be labeled by random hexanucleotide labeling using Klenow enzyme (many labeled sites along probe) Stronger signal, more sensitive. Probe can be labeled with fluorescent signals as an alternative to radioactivity.

Screening of cDNA libraries

Screening is about detecting a specific gene by its sequence or its expression

- Direct screening (in situ hybridization) involves use of an oligonucleotide probe to detect a specific sequence

- Indirect screening involves demonstrating the presence of the expressed protein product of the gene rather than the gene sequence itself.

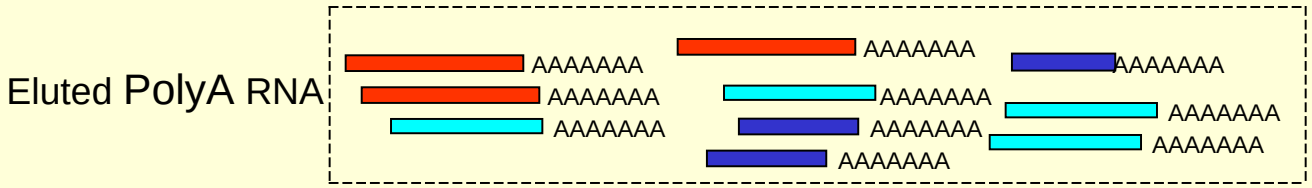
  - **Requires:** Expression of the genes in the cDNA library in their bacterial hosts (to protein products)

  - **Requires:** An efficient and specific means of detecting the expressed gene product.

    - If target gene encodes an enzyme and there was a way to demonstrate the presence of the enzyme in a bacterial colony—
      eg. something like the Xgal substrate for LacZ

    - Immunoscreening involves using antibodies that specifically recognize the target protein and bind to it.

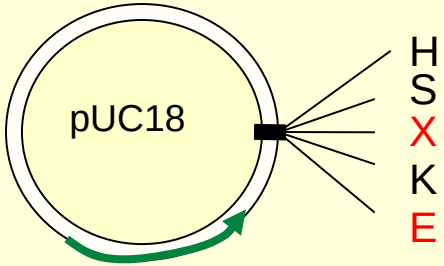
Directional cloning



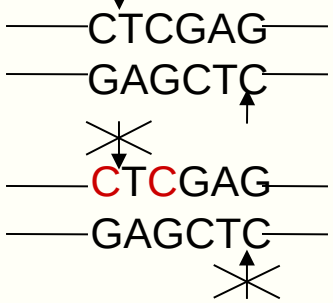
Using Reverse Transcriptase (RNA template-dependent DNA polymerase) and oligo-dT (15-18mer) primer, mRNA copied into complementary DNA (cDNA)

In non-directional cloning

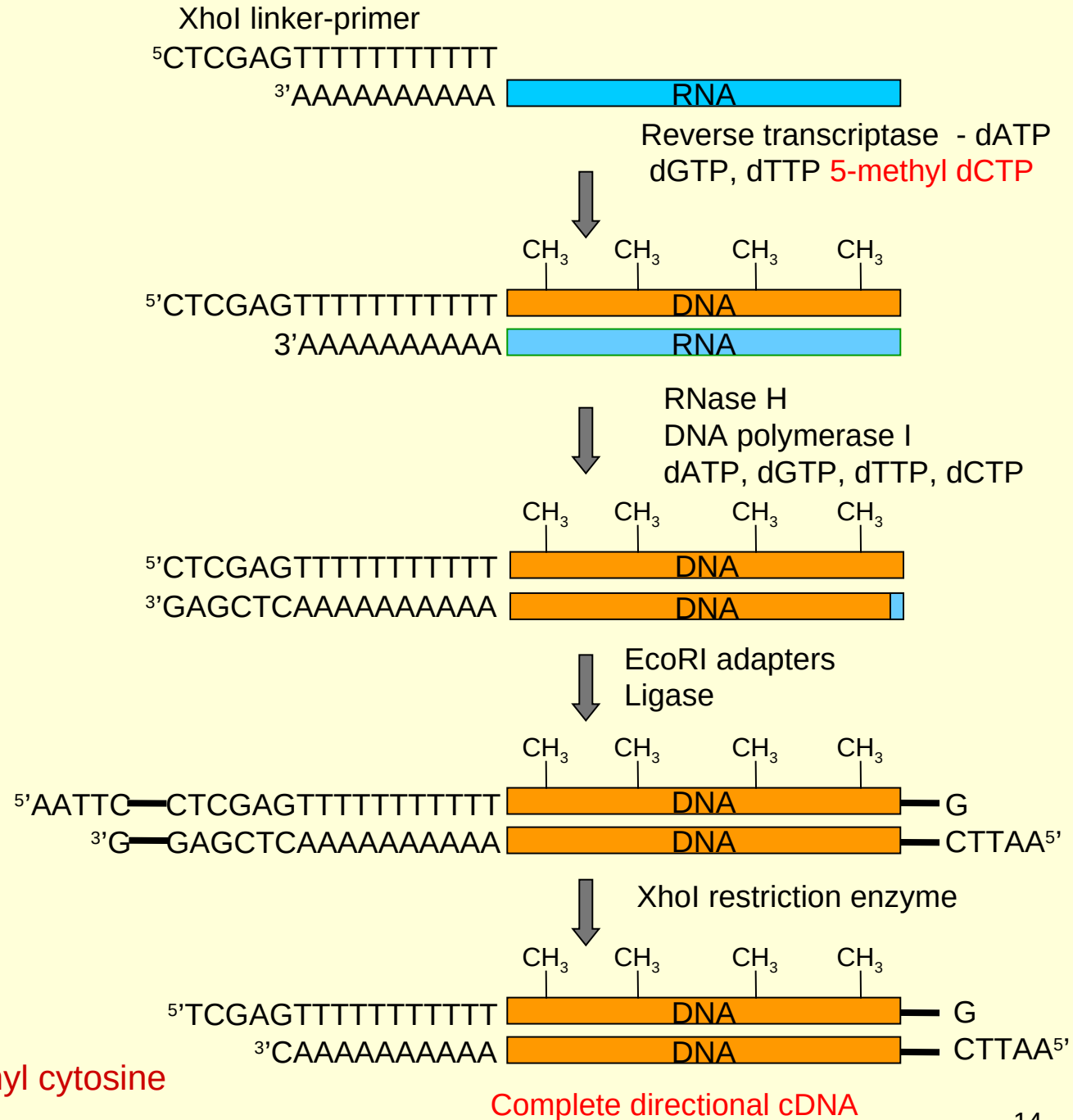
Stratagene Directional Cloning strategy



LacZ promoter



C -methyl cytosine

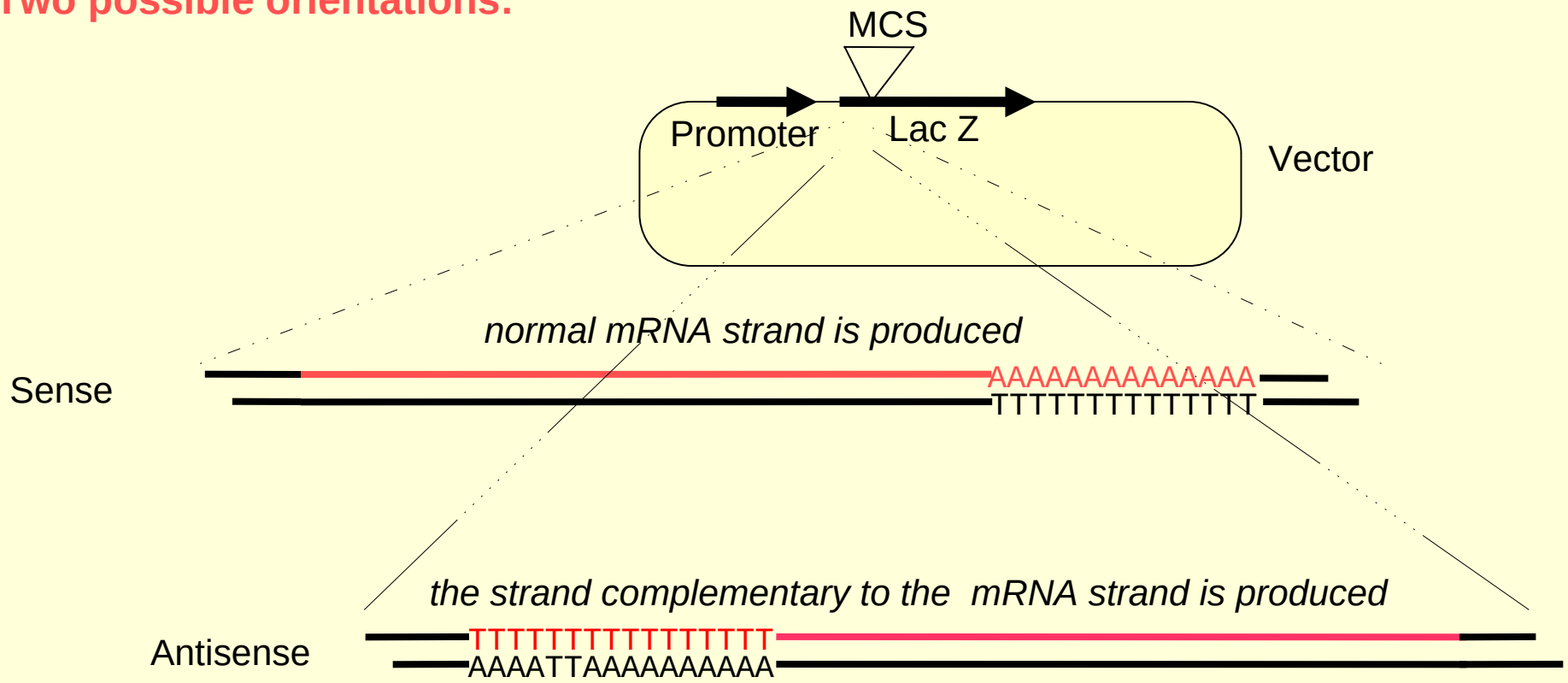


Complete directional cDNA

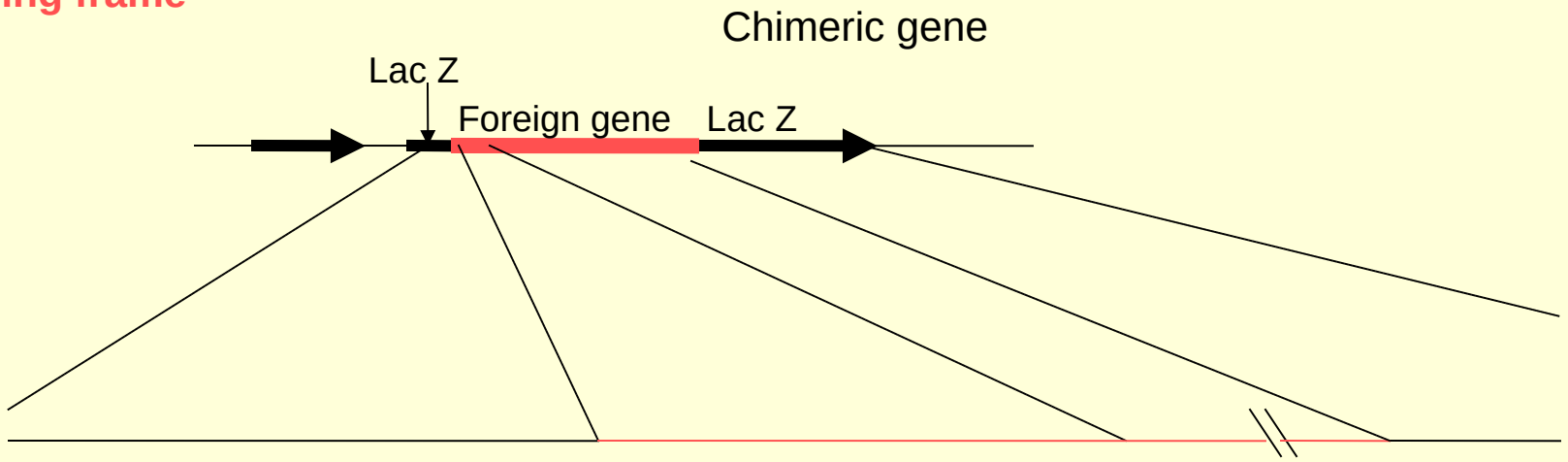
Expression of the genes in the cDNA library

- Expression requires use of the selectable marker gene promoter
  1. Correct (sense) orientation of gene sequence in the vector.
  2. Correct reading frame with the Lac Z gene

Two possible orientations:



Reading frame

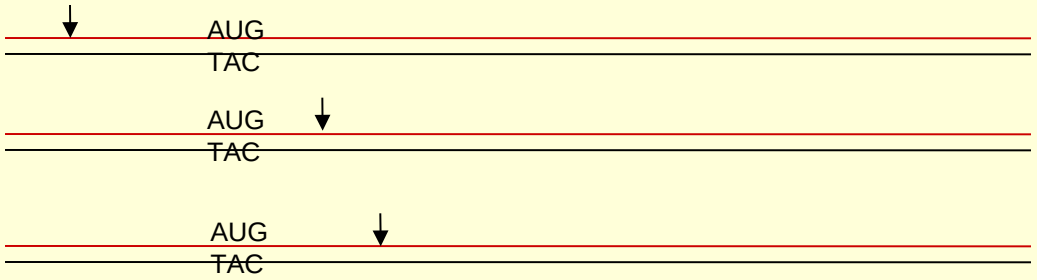


ATG ACC ATG ATT xxx xxx xxx xxx xxx xxx ATG GGA AGT AAG TAT ---

These codons must be in same reading frame

Met Thr Met Ile yyy yyy yyy yyy uuu uuu Met Gly Ser Tyr

What is the likelihood of this being the case?



3 mRNA from same gene in a heteroduplex

Cleavage by RNase H is random so there is an equal chance that subsequent insertion into the Lac Z gene will be in any of its 3 reading frames. Cleavage could either be in the 5'UTR, or, more commonly, in the coding region itslef.

True for all mRNAs - hence at best only 1/3 of the clones in a library will be able to correctly express the encoded protein if insert is in the correct orientation.

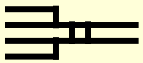
If the cDNA gene has the same adapters on each end and can be inserted in either orientation then the potential for correct expression drops to 1/6

Thus it is important with a cDNA expression library to have the best selection of tissue and time for mRNA isolation for your gene of interest so that there are many copies present.

Can double chances by directional cloning

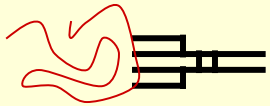
Antibodies -as specific probes

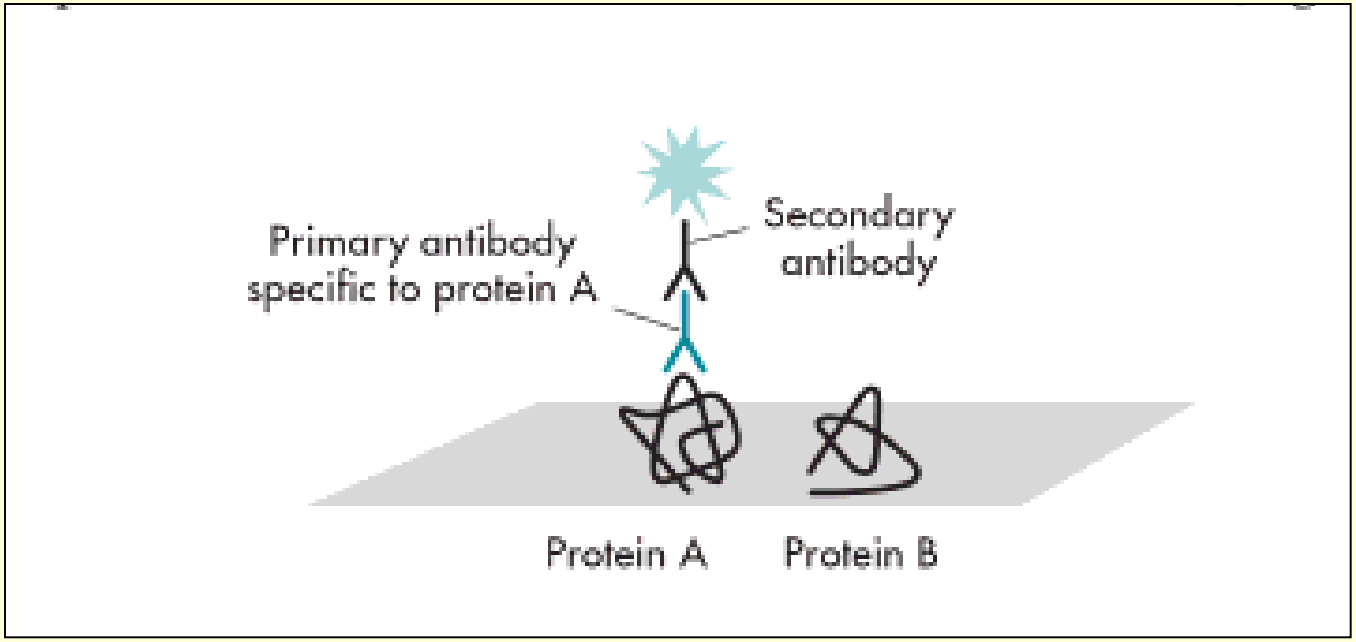
- Are part of the defensive immune system of animals and birds
  - Injection of foreign proteins into any animal of bird results in the generation by the animal of a mixture of antibodies which will recognize and bind to the foreign protein and trigger its removal from the system



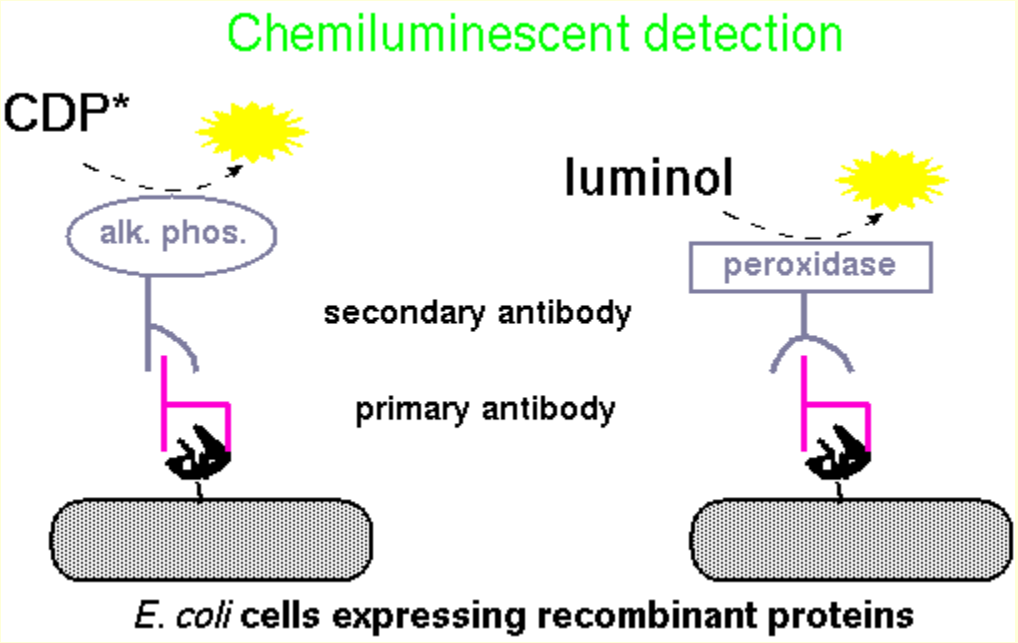
- To prepare specific antibodies for screening, you must isolate the pure protein for the target gene, inject it into a rabbit, rat, or mouse (normally) where it serves as the antigen. The animal produces antibodies in their blood which can be recovered by collecting the blood serum – antiserum.

- When antiserum is exposed to the antigen in a test-tube under proper conditions it will bind to it tightly.





Gene cloning: principles and applications: Julia Lodge, Peter A. Lund and Steve Minchin



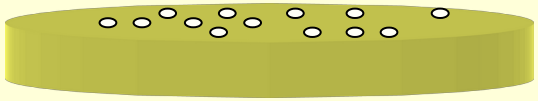
Chemiluminescent detection is typically done using substrates that emit photons of light when broken down by enzyme-conjugated antibodies. The antibody binds to a specific protein on the E. coli cell surface, carrying with it the enzyme.

For alkaline phosphatase, the substrate is CDP*

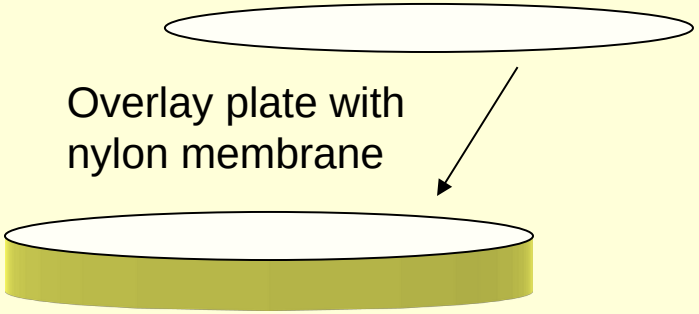
For horseradish peroxidase the substrate luminol peroxide from Clontech Inc.

Indirect Screening using Antibodies (Immunoscreening)

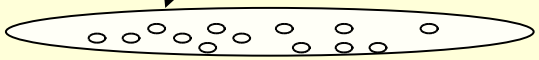
Colonies/plaques from cDNA library



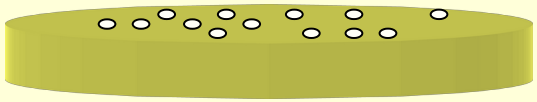
Overlay plate with nylon membrane



Remove membrane



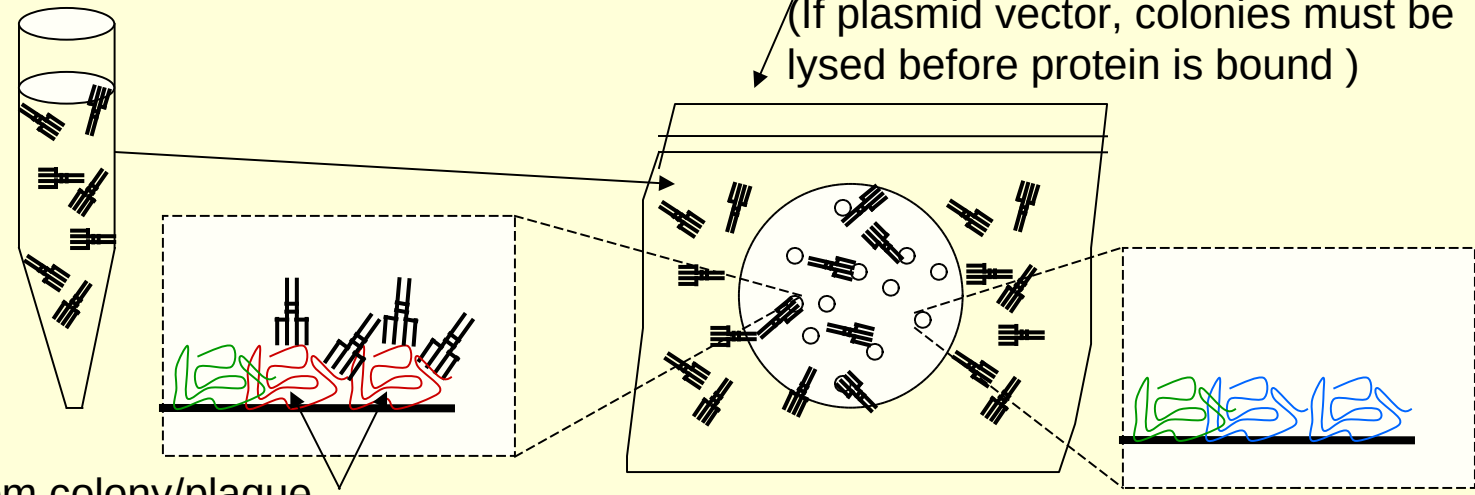
Master plate



Replica of plate with adsorbed protein and DNA from each plaque

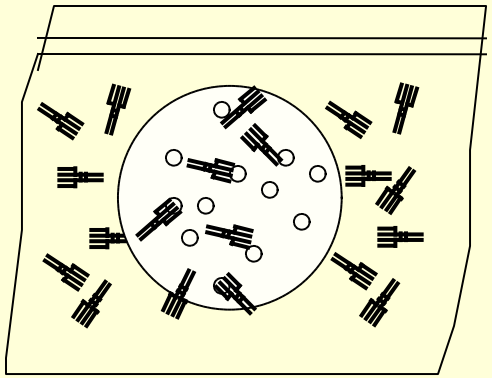
(If plasmid vector, colonies must be lysed before protein is bound)

Primary antiserum to target protein

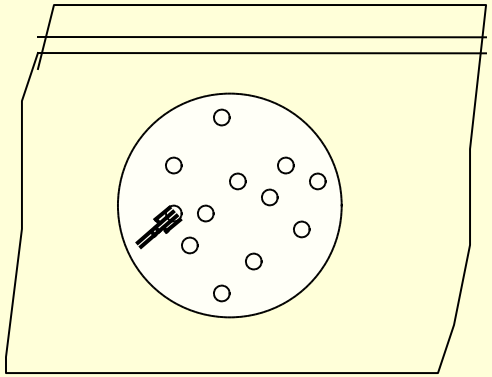


Proteins bound from colony/plaque

Antibodies bind to specific protein if it is present

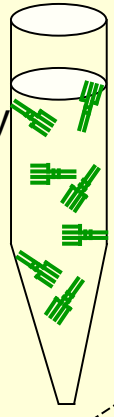


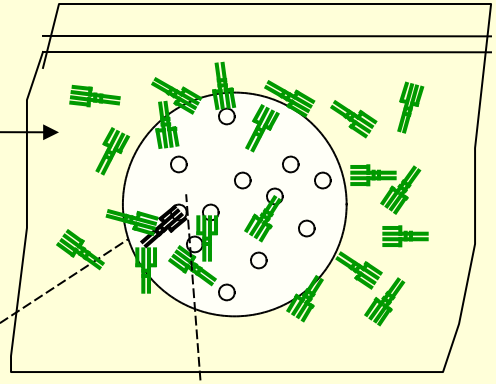
After allowing binding wash unbound primary antibody away

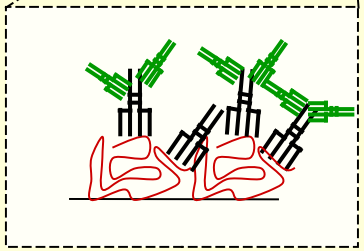


Add labeled secondary antibody that binds to the primary antibody

Secondary antibody

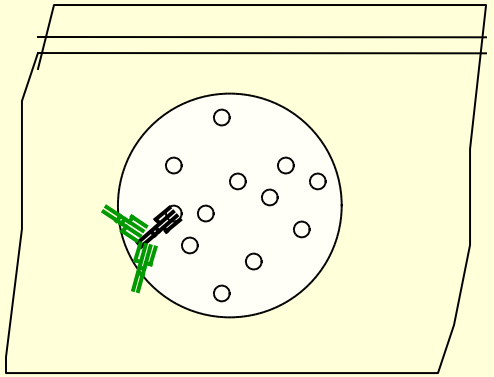






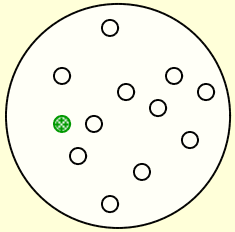
Wash away any unbound antibody

Add substrate

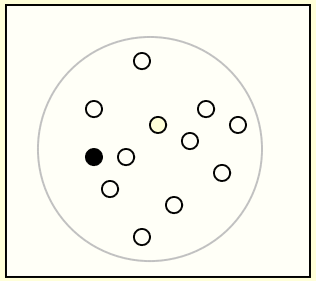


Expose to film to find where specific antibody was bound

Immunoscreening (cont'd)

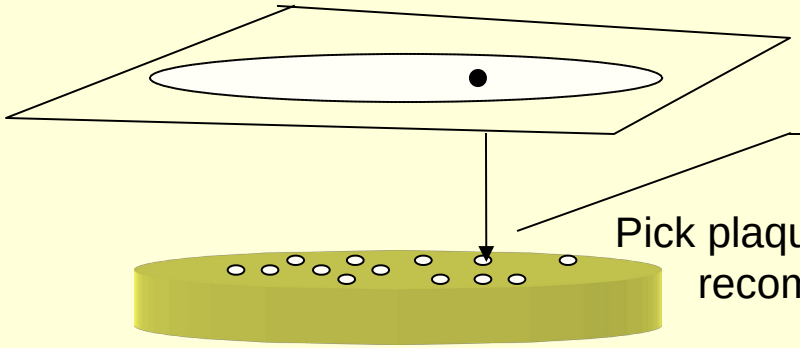


Expose to X-ray film



Exposed film blackened by region with bound labeled probe

Align film to master plate to find correct plaque



Pick plaque/colony and isolate recombinant vector and insert DNA

Master plate

Summary of gene isolation from a cDNA library

- Make cDNA library

- Screen library
  - Nucleotide probe (direct screening for gene)
  - Antibody probe (indirect screening for protein gene product)

- Select positive clone
  - If phage library –amplify selected clone and isolate insert and transfer to a plasmid vector for future work
  - If plasmid library – amplify cells from positive colony and isolate plasmid and retain cell line with selected clone

- Characterize the insert by sequencing it

- To verify you have the correct gene,
  - compare the sequence to other sequences in the available databases to see if there is similarity (sequence or motifs). This can be done at http://blast.ncbi.nlm.nih.gov

    See BLASTX search on next 2 slides

  - If you originally had protein sequence information you could verify if the nucleotide sequence would encode the protein by translating all reading frames

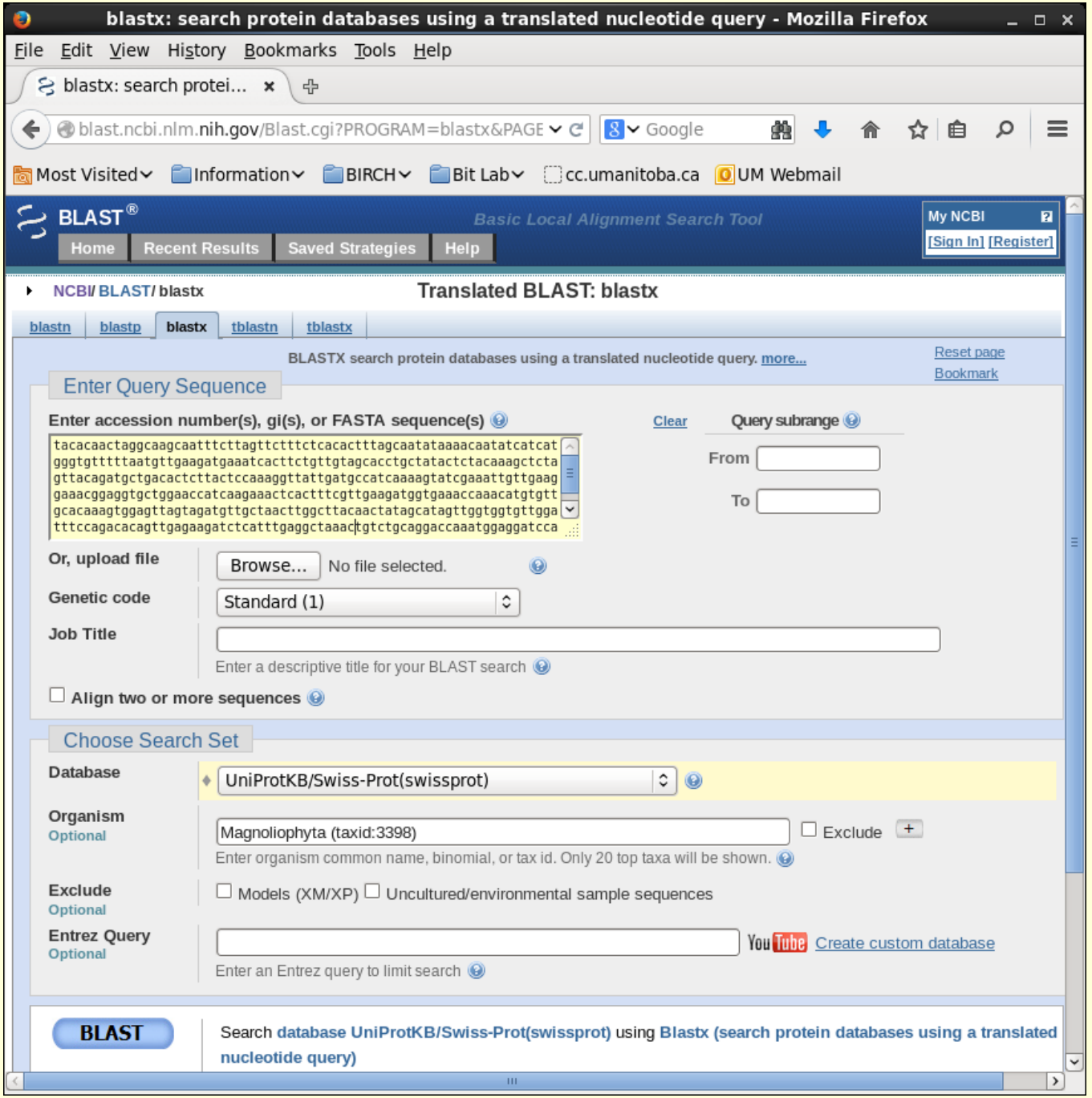
- Advantage of cDNA library is its simplicity but if you want the promoter and total unprocessed gene you would need to go to a genomic library.

  However, if you have the correct cDNA clone this is now easy!!

- This is how a gene is isolated when you know what the gene is --- i.e its sequence or identity based on the gene product

NCBI BLAST search

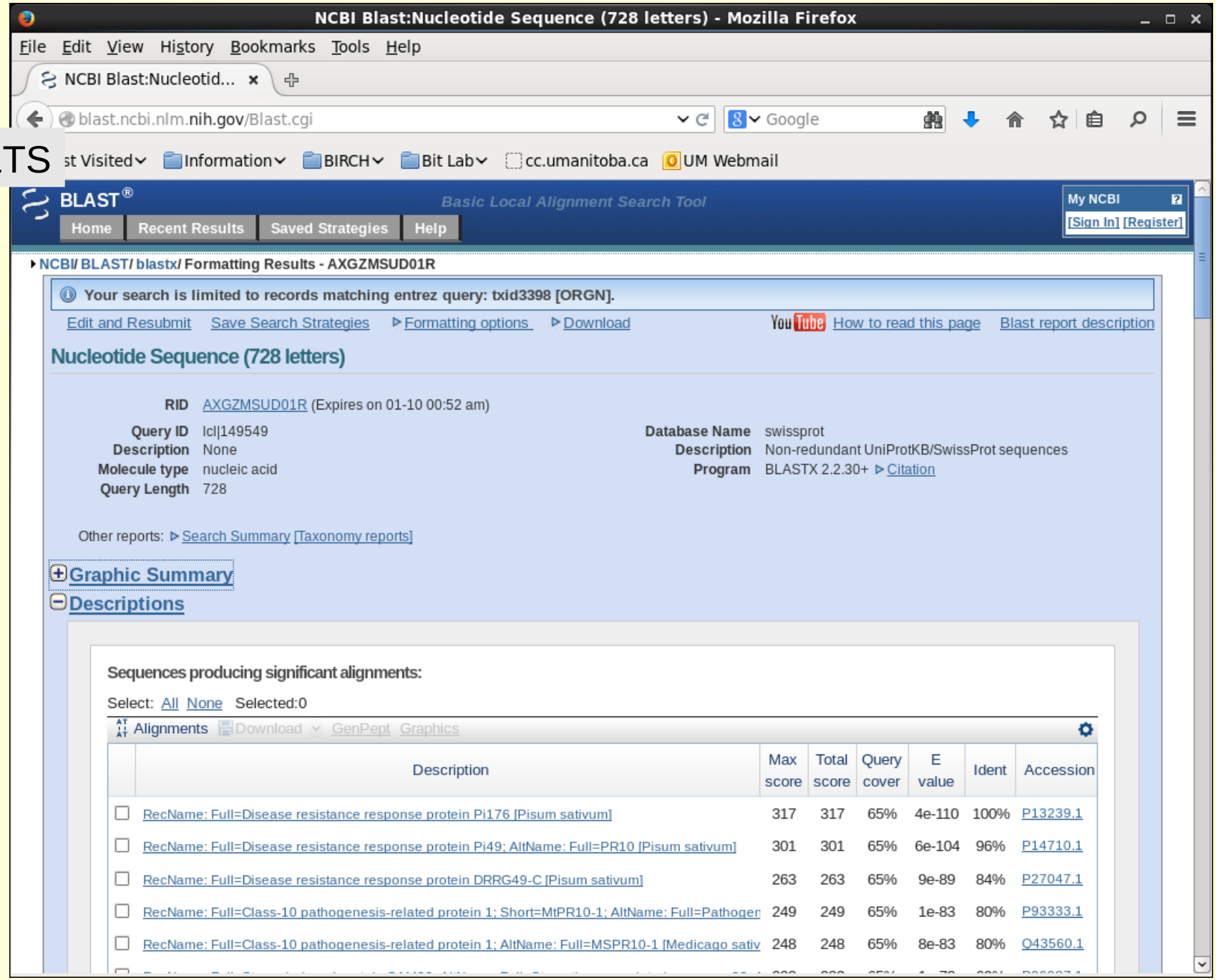
Set parameters, and paste your query sequence into the box



Note that we are searching the UniProt database, which is probably the most carefully annotated database for proteins.

Since the query is a plant sequence, we limit the search to the Magnoliophyta (flowering plants).

RESULTS


NCBI Blast:Nucleotide Sequence (728 letters) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NCBI Blast:Nucleotid...
blast.ncbi.nlm.nih.gov/Blast.cgi
Google
st Visited Information BIRCH Bit Lab cc.umanitoba.ca UM Webmail
BLAST®
Basic Local Alignment Search Tool
Home Recent Results Saved Strategies Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST/ blastx/ Formatting Results - AXGZMSUD01R
Your search is limited to records matching entrez query: txid3398 [ORGN].
Edit and Resubmit Save Search Strategies Formatting options Download YouTube How to read this page Blast report description
Nucleotide Sequence (728 letters)
RID AXGZMSUD01R (Expires on 01-10 00:52 am)
Query ID lcl|149549
Description None
Molecule type nucleic acid
Query Length 728
Database Name swissprot
Description Non-redundant UniProtKB/SwissProt sequences
Program BLASTX 2.2.30+ Citation
Other reports: Search Summary [Taxonomy reports]
Graphic Summary
Descriptions
Sequences producing significant alignments:
Select: All None Selected:0
Alignments Download GenPept Graphics
Description
Max score
Total score
Query cover
E value
Ident
Accession
RecName: Full=Disease resistance response protein Pi176 [Pisum sativum] 317 317 65% 4e-110 100% P13239.1
RecName: Full=Disease resistance response protein Pi49; AltName: Full=PR10 [Pisum sativum] 301 301 65% 6e-104 96% P14710.1
RecName: Full=Disease resistance response protein DRRG49-C [Pisum sativum] 263 263 65% 9e-89 84% P27047.1
RecName: Full=Class-10 pathogenesis-related protein 1; Short=MtPR10-1; AltName: Full=Pathogen 249 249 65% 1e-83 80% P93333.1
RecName: Full=Class-10 pathogenesis-related protein 1; AltName: Full=MSPR10-1 [Medicago sativ 248 248 65% 8e-83 80% Q43560.1