

# *Expression Cloning Vectors*

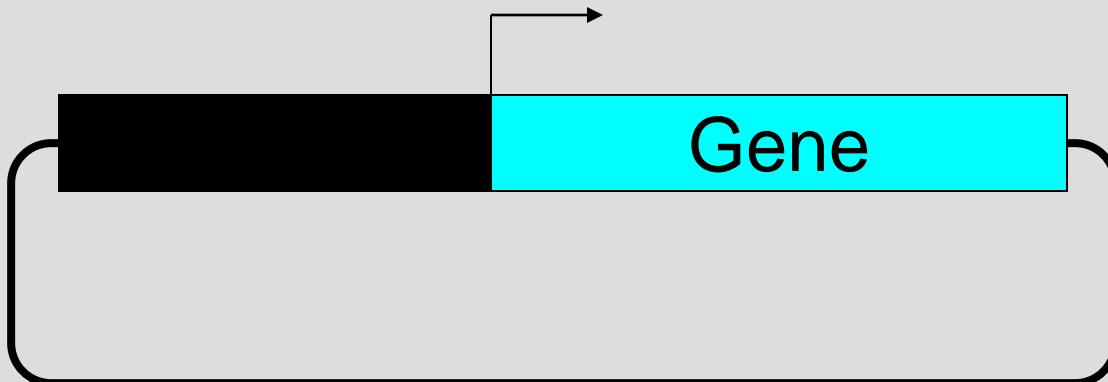
# **Factors required for high-level expression of genes**

- 1) transcriptional promoter & terminator**
- 2) Shine Dalgarno Sequence (ribosome binding site)**
- 3) Efficiency of translation**
- 4) Stability of the protein**
- 5) Final cellular location (secreted?)**
- 6) Number of copies of cloned gene**
- 7) Host organism**

# *Gene Expression from a Strong, Regulatable Promoter*

---

Expression Plasmids have promoters upstream of cloning sites for expression of genetic info encoded by DNA fragment



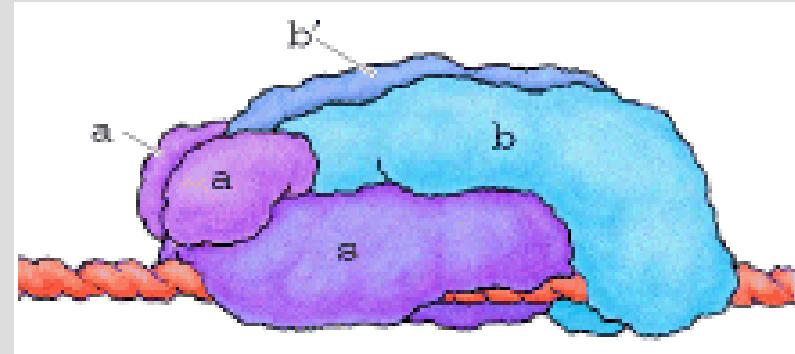
# *Gene Expression from a Strong, Regulatable Promoter*

---

- **Strong**
  - high affinity for RNA polymerase
  - tight binding - frequently transcribed
  - weak binding - RNA Pol falls off, no txn
- **Regulatable**
  - researcher can control when gene is expressed
  - use inducers / co-repressors

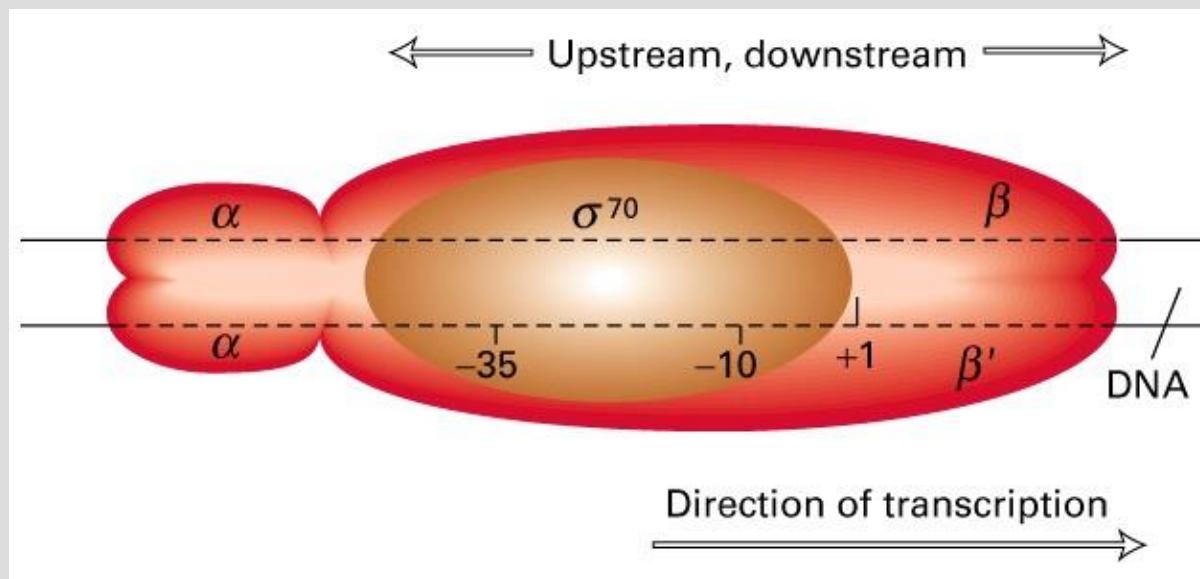
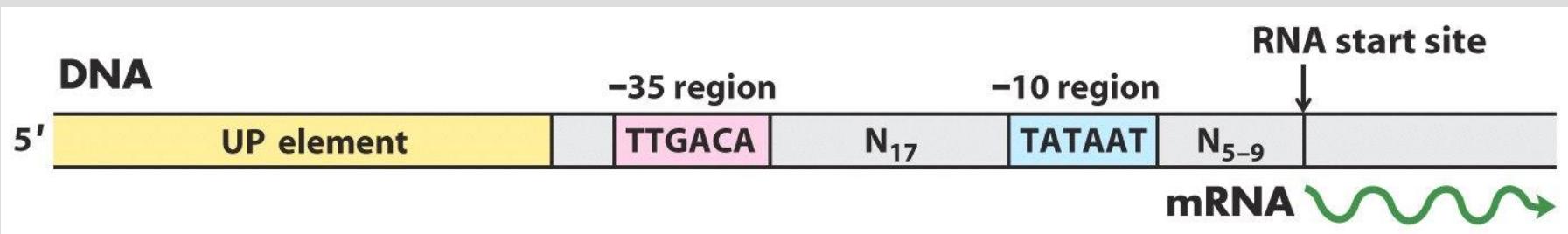
# Initiation

- RNA polymerase  $\alpha\alpha\beta\beta'\sigma$
- Transcription factors
- Sigma factor ( $\sigma$ )- determines promoter specificity



Start site of txn is +1

# *E. coli* Promoter Sites



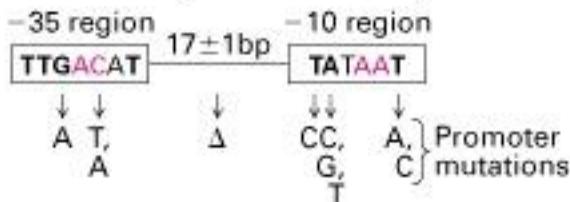
# *E. coli* Promoter Sites

## (a) Strong *E. coli* promoters

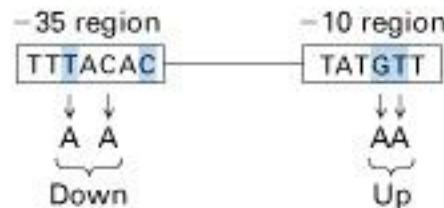
The figure shows a sequence alignment of various *E. coli* promoters. The promoters are listed on the left, and their sequences are shown on the right. The -35 promoter region is highlighted in yellow, the -10 region in yellow, and the start site (+1) in red. The start site is indicated by a red arrow. The sequences are aligned to show conservation of these key regions across different promoters.

Promoter	Sequence
tyr tRNA	TCTCAACGTAACACTTTACAGCGGGCG**CGTCATTGATATGATGC*GCCCGCTTCCCATAAAGGG
rrn D1	GATCAAAAAAAATACTTGTGCAAAAAAA**TTGGGATCCCTATAATGCGCCTCCGTTGAGACGACAACG
rrn X1	ATGCATTTTCCGCTTGTCTTCTGA**GCCGACTCCCTATAATGCGCCTCCATCGACACGGCGGAT
rrn (DXE) <sub>2</sub>	CCTGAAATTCAAGGGTTGACTCTGAAA**GAGGAAAGCGTAATATAAC*GCCACCTCGCGACAGTGAGC
rrn E1	CTGCAATTTCCTATTGCGGCTCGCG**GAGAACTCCCTATAATGCGCCTCCATCGACACGGCGGAT
rrn A1	TTTTAAATTTCCTCTTGTCAAGGCCGG**AATAACTCCCTATAATGCGCCACCACTGACACGGAAACAA
rrn A2	GCAAAATAAAATGCTTGTACTCTGTAG**CGGGAAAGGCGTATTATGC*ACACCCCCGCGCTGAGAA
$\lambda$ P <sub>R</sub>	TAACACCGTGCCTGTTGACTATTAA•CCTCTGGCGGTGATAATGG•TTGCATGTACTAAGGAGGT
$\lambda$ P <sub>L</sub>	TATCTCTGGCGGTGTTGACATAATAA•CACTGGCGGTGATAACTGGA•GCACATCAGCAGGACGCAC
T7 A3	GTGAAACAAAAACGGTTGACAACATGA•AGTAAACACGGTACGATGT•ACCACATGAAACGACAGTGA
T7 A1	TATCAAAAAGAGTATTGACTTAAAGT•CTAACCTATAGGATACTTA•CAGCCATCGAGAGGGACACG
T7 A2	ACGAAAAAACAGGTATTGACAACATGAAGTAACATGCAGTAAGATAAC•AAATCGCTAGGTAACACTAG
fd VIII	GATACAAATCTCCGTTGACTTTGTT**TCGCCTTGGTATAATCG*CTGGGGGTCAAAGATGAGTG

## (b) Consensus sequences of $\sigma^{70}$ promoters

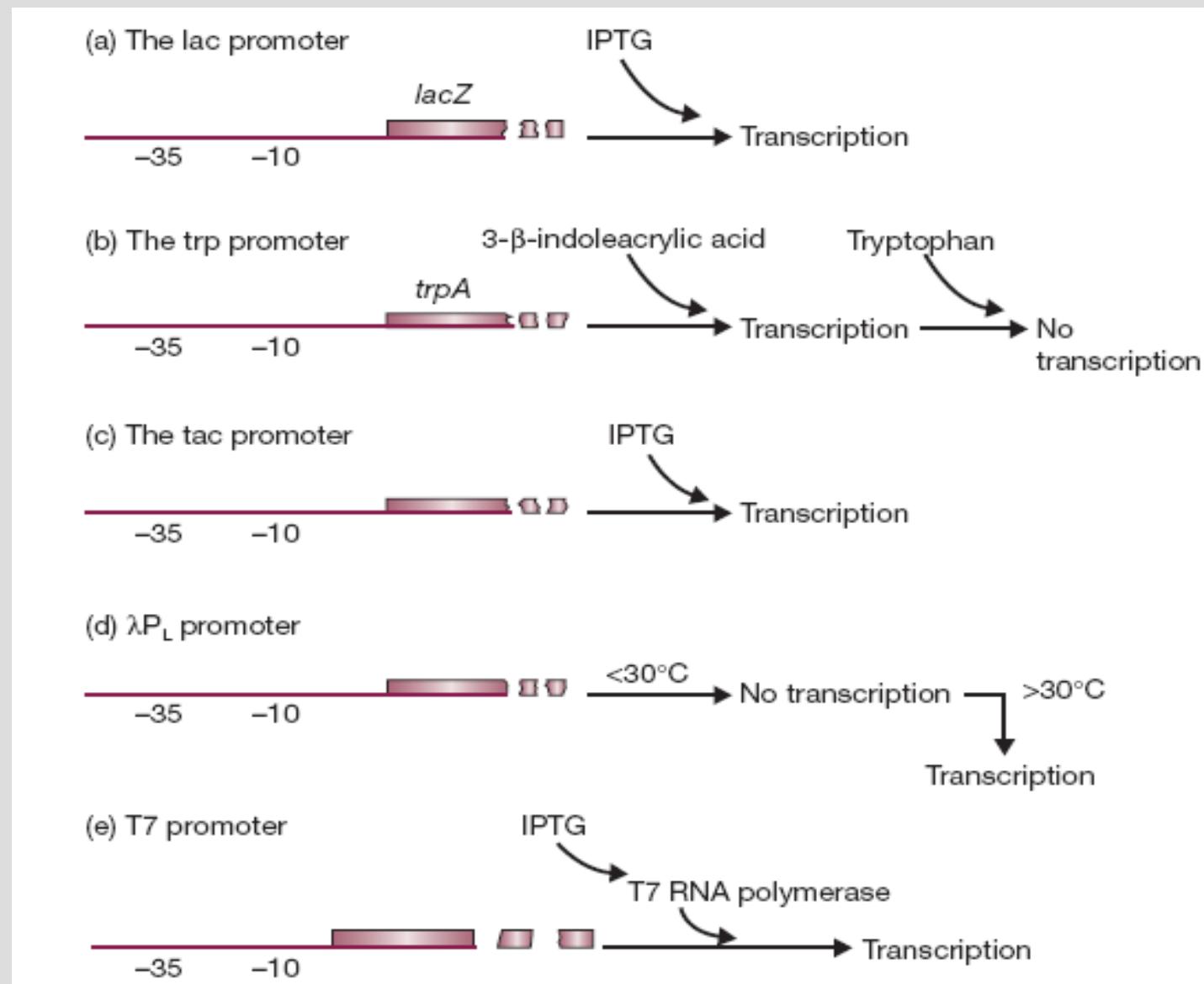


## (c) *Lac* promoter sequence



Deviation from consensus -10 , -35 sequence leads to weaker gene expression

# Promoters of Importance in Biotech



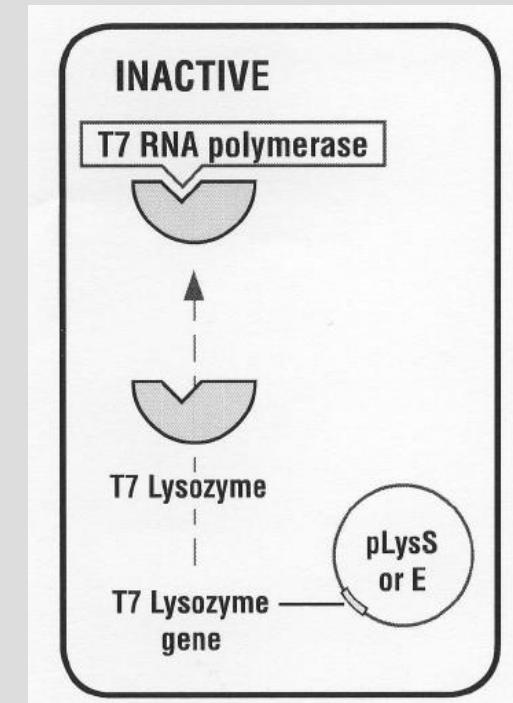
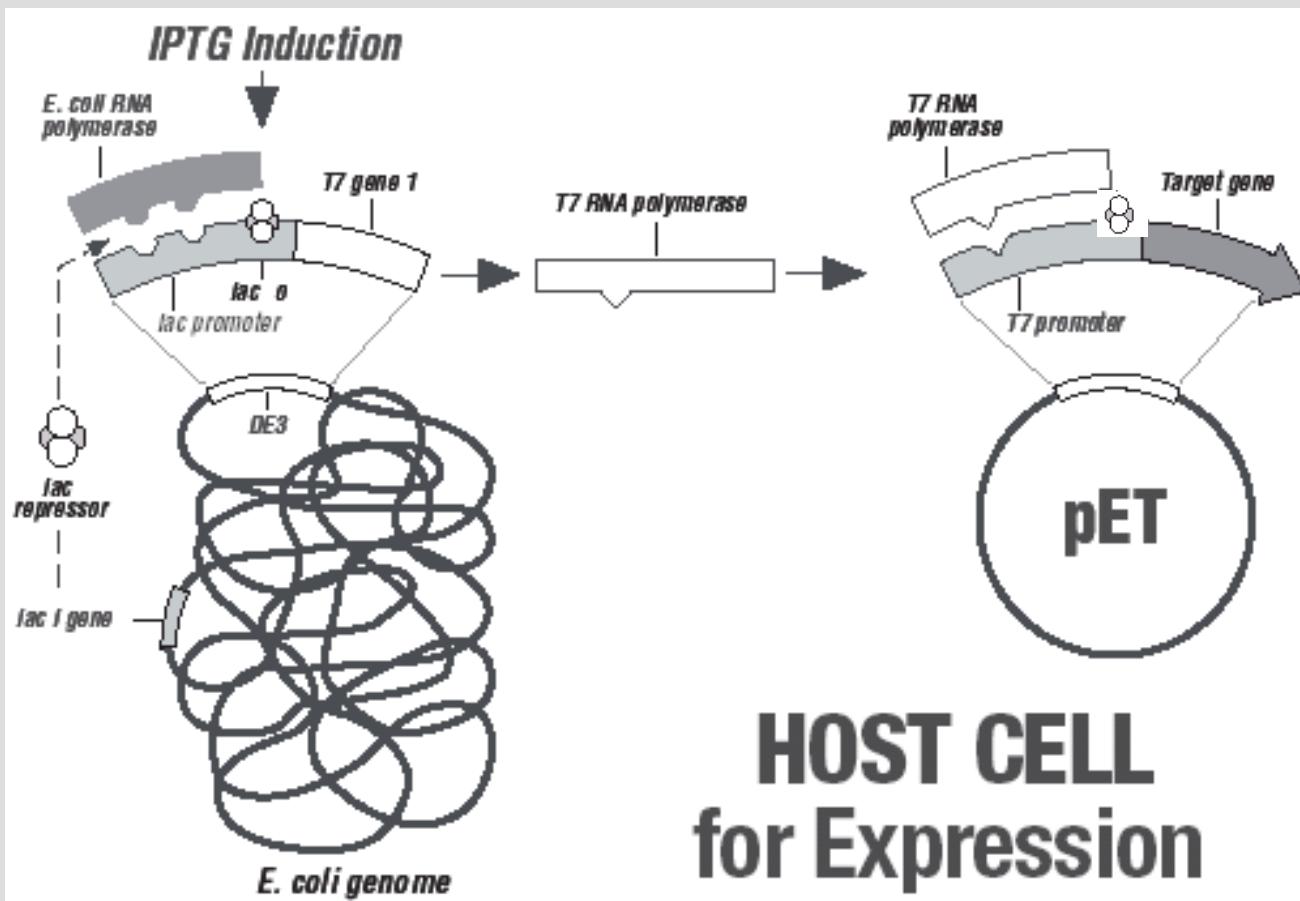
# *tac (trc) promoter*

---

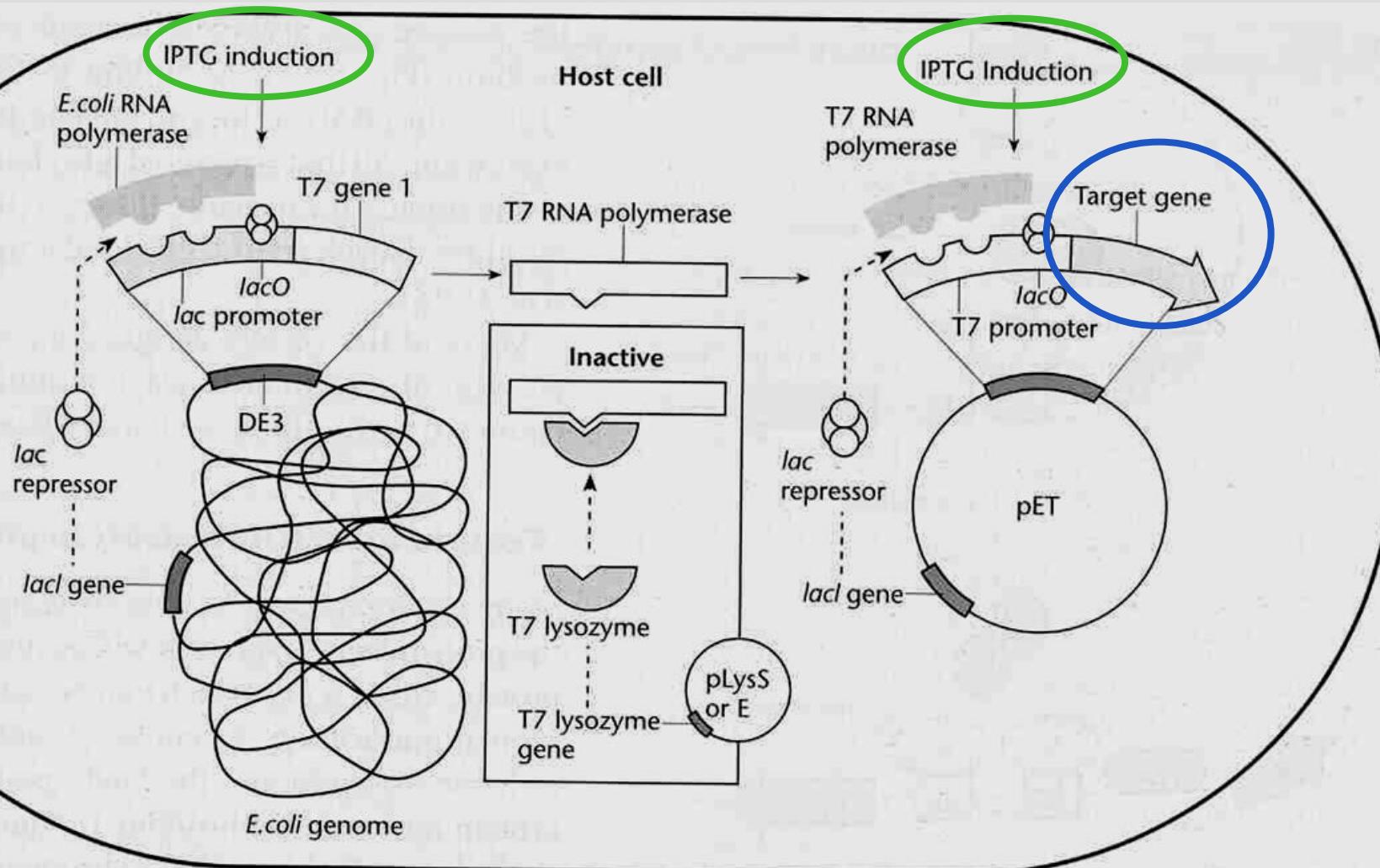
- Hybrid of lac and trp promoters
  - 35 region from trp
  - 10 region from lac
    - separated by 16bp = tac promoter
    - separated by 17bp = trc promoter
- 3x stronger than trp
- 5-10x stronger than lac

# *T7 promoter*

- From T7 phage
- Use T7 polymerase – need it



# pET vectors: protein expression



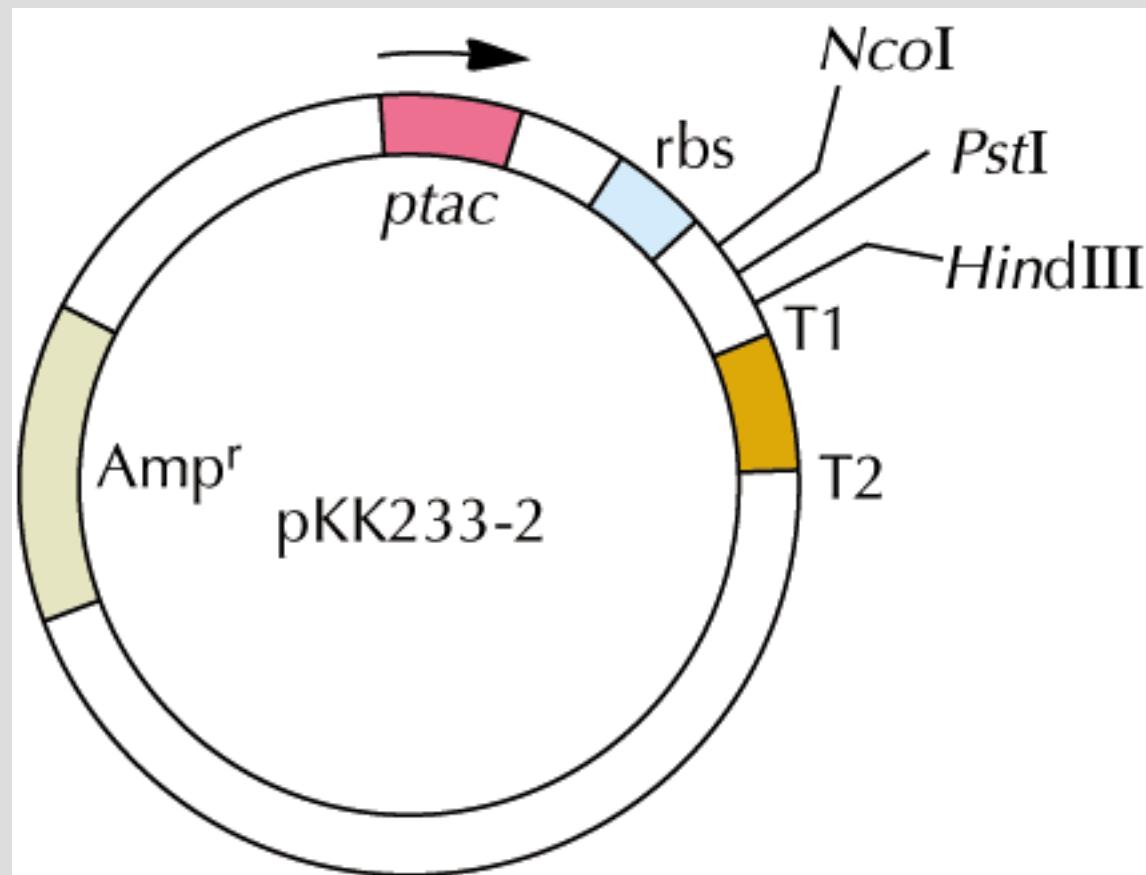
# Shine Dalgarno Sequence

An efficient expression vector requires not only a strong, regulatable promoter, but also an *E. coli ribosome binding sequence and a terminator.*

- In Prokaryotes
  - translation signal is RBS
  - ribosome binding site
  - Shine-Dalgarno sequence (AGGAGGU)
  - 6-8 nt in length
  - located short distance (~10nt) upstream of AUG translation start codon

# *Expression Vector pKK233-2*

selectable marker  
tac promoter  
**RBS**  
unique REase sites  
Txn terminators  
  
not shown:  
ori of replication

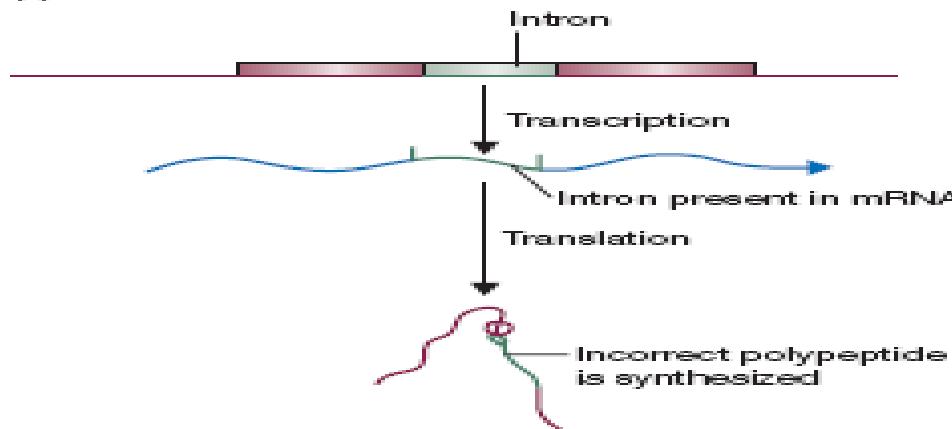


# **General Problems with the production of foreign recombinant protein expressed in *E.coli***

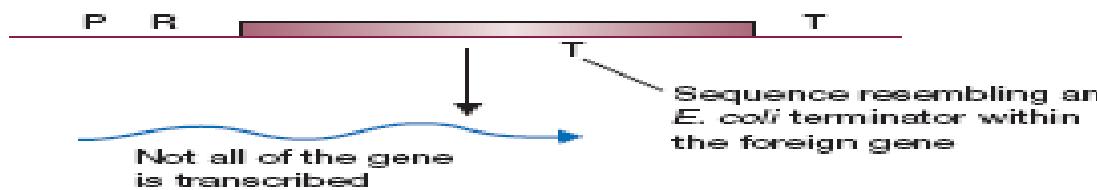
1. Problems caused by the sequence of the foreign gene
2. Problems caused by the *E.coli*

# General Problems with foreign protein expressed in E.coli

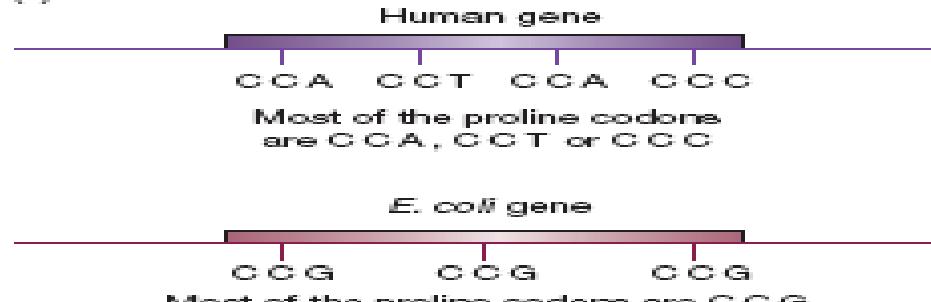
(a) *E. coli* cannot excise introns



(b) Premature termination of transcription



(c) Codon bias



Result: *E. coli* has difficulty translating the proline codons in a human gene

# ***Problems caused by E.coli***

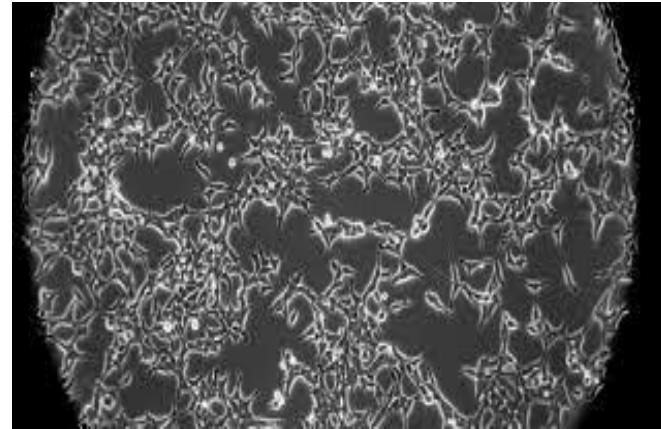
---

- ***E.coli*** NOT always the best organism
  - *E. coli might not process the recombinant protein correctly (post translational modifications)*
    - *Correct disulfide bond formation*
    - *Proteolytic cleavage of inactive precursor*
    - *Glycosylation - addition of sugar residues*
    - *Alteration of amino acids in protein*
      - » *phosphorylation*
      - » *acetylation*
      - » *sulfation*
      - » *fatty acid addition*
  - *E. coli might not fold the recombinant protein correctly (S-S bonds, incorrect folding, inclusion bodies, inactive)-Use chaperone overexpressing strains*
  - *E. coli might degrade the foreign recombinant protein- Use protease deficient strains*

# Eukaryotic Expression Systems

- *Saccharomyces cerevisiae*
- *Pichia pastoris*
- Mammalian systems
  - Transient expression:
    - Vero cells are derived from the kidney of an African green monkey, and are one of the more commonly used mammalian continuous cell lines in microbiology, and molecular and cell biology research
    - BHK (Baby hamster kidney) cells line- like BHK-21 is one of the most commonly used cell lines for the expression of biopharmaceuticals, and it is also among the top three cell types that have been most frequently used for transient expression. BHK cells are originally isolated by polyoma transformation of hamster cells and have been extensively used as substrates for virus propagation for vaccine and more generally for viral mediated expression
    - Human Embryonic Kidney Cell line- HEK293 cells
  - Long-term expression: **Chinese hamster ovary** and mouse myeloma cells.

# HEK293 Cells



HEK293 cells are Human Embryonic Kidney cells, originally isolated and grown by Dutch biologist Alex Van der Eb in the early 1970s. They were transfected with sheared adenovirus 5 (Ad5) DNA by Frank Graham, a postdoc in Van der Eb's lab. It was his 293rd experiment, which is why they got the tag HEK293 A 4-kb Ad5 DNA fragment encoding the E1A/E1B proteins was later shown to have integrated into chromosome 19, resulting in transformation.

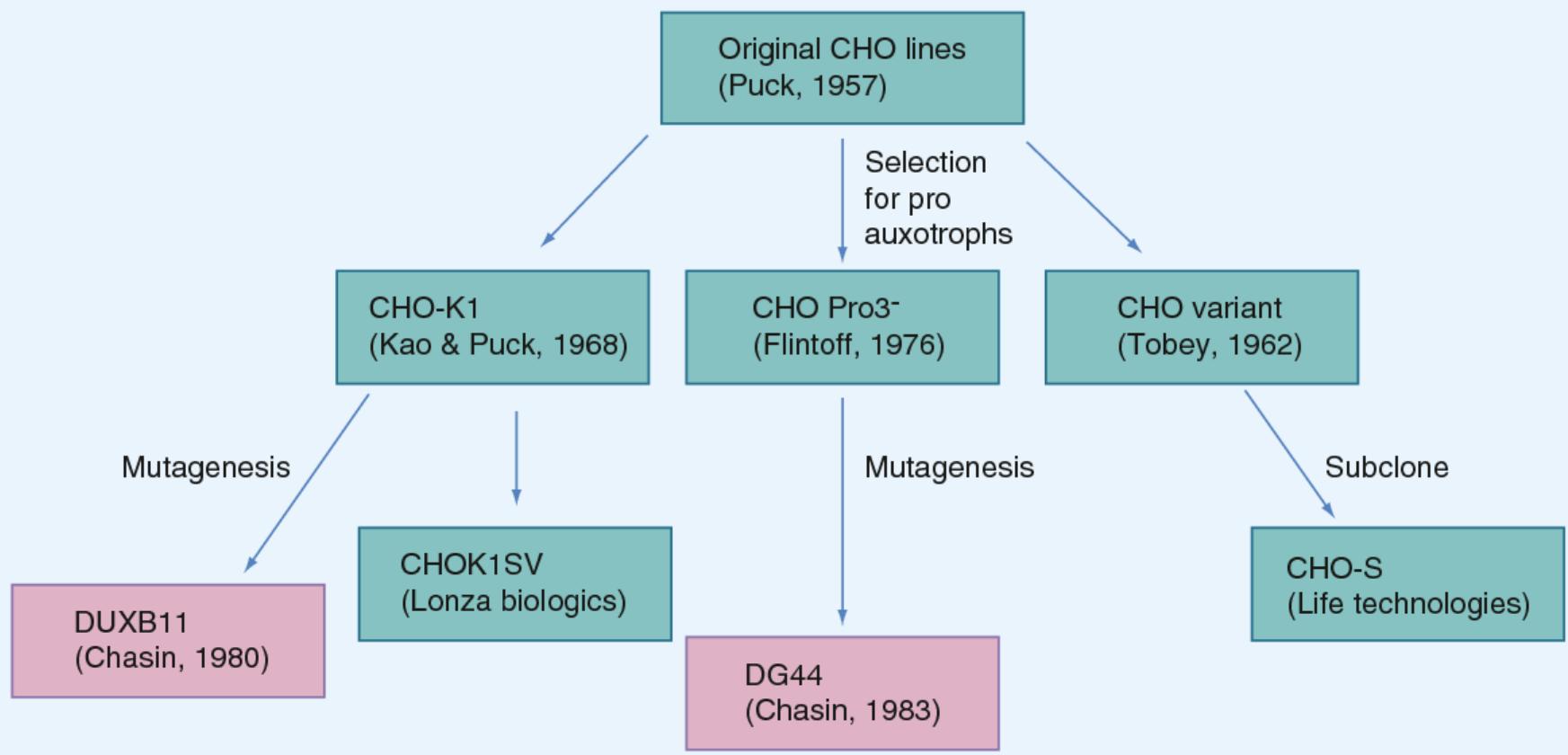
Incorporating the adenoviral genes into the HEK cell genome resulted in the cells becoming very efficient at producing high amounts of recombinant proteins from plasmid vectors carrying the CMV promoter region.

There are many advantages of using HEK293 cells. They are a hardy, semi-adherent, low-maintenance cell line and divide rapidly, doubling about every 36 hours. They can be utilized for both transient and stable expression, can be cultured in suspension or as a monolayer, are easy to transfect (and can be transfected via a variety of methods) and are able to produce large amounts of recombinant proteins. HEK293 cells are used in cancer research, vaccine development, protein production, signal transduction and protein interaction studies, drug testing etc

# HEK293T

- The ‘T’ in the name of this daughter cell line comes from the incorporation of the SV40 large T antigen into the HEK293 genome – this means they are able to produce large amounts of protein from plasmids vectors carrying the SV40 origin of replication.

# CHO Cell lines

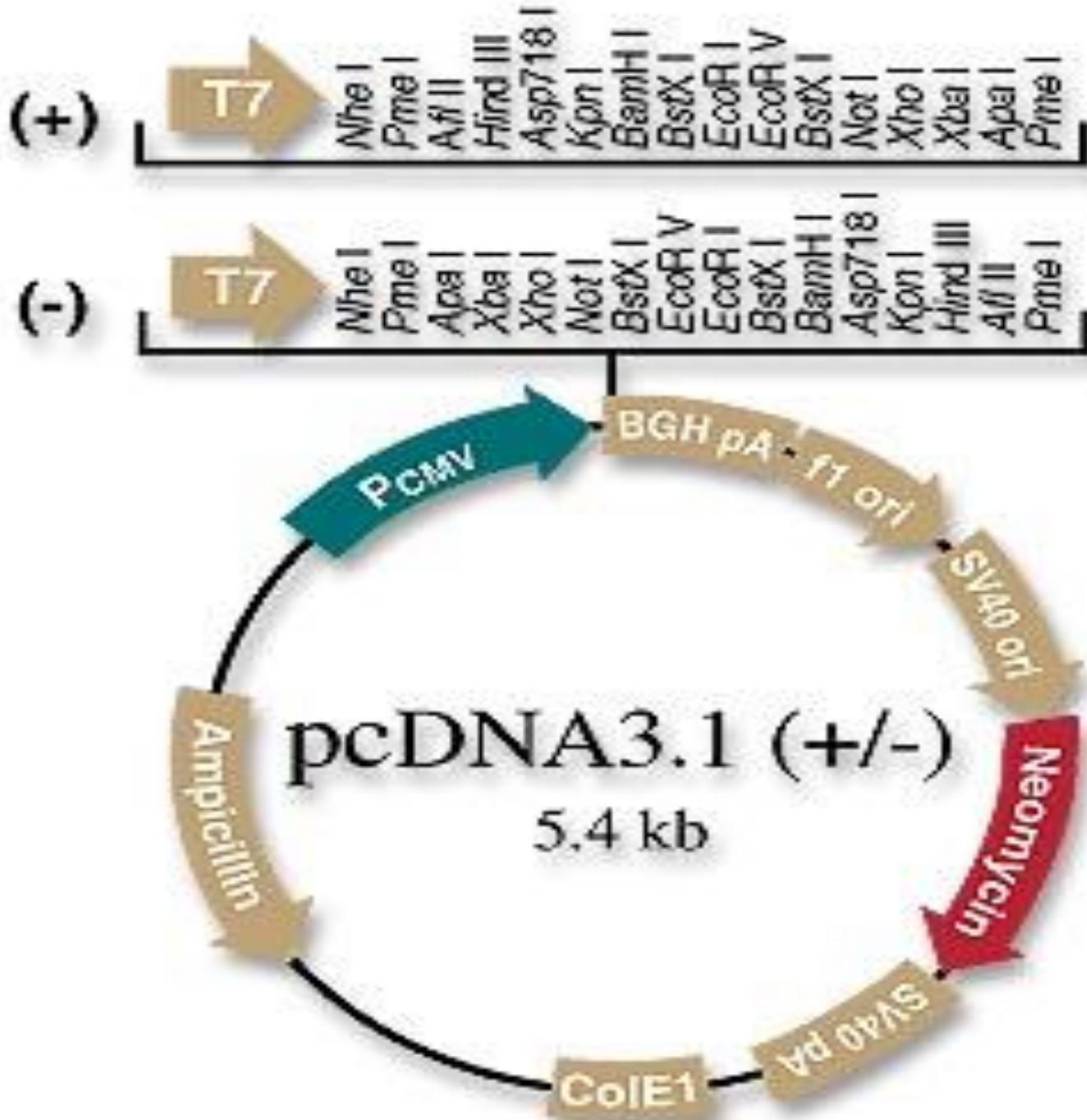


**Figure 3. History and lineage of Chinese hamster ovary cell lines.** Cell lines derived by mutagenesis to be DHFR

DUXB11 and DH44- A cell line deficient in DHFR activity, which requires the addition of glycine, hypoxanthine, and thymidine (GHT) in the medium for survival, allows for the implementation of a selection system based on the insertion of a cloned dhfr gene in combination with the gene of interest.

# Eukaryotic Expression Vectors

- Same sorts of genetic features
  - eukaryotic promoter—generally derived from animal viruses or from highly expressed mammalian genes (SV40, cytomegalovirus (CMV), herpes simplex virus (HSV), Elongation factor 1 (EF1))
  - ori of replication (eukaryote..usually viral, SV40)
  - selectable marker (for eukaryotic cell)
  - mRNA polyadenylation signal
  - ori of replication (Plasmid)
  - selectable marker (bacteria)



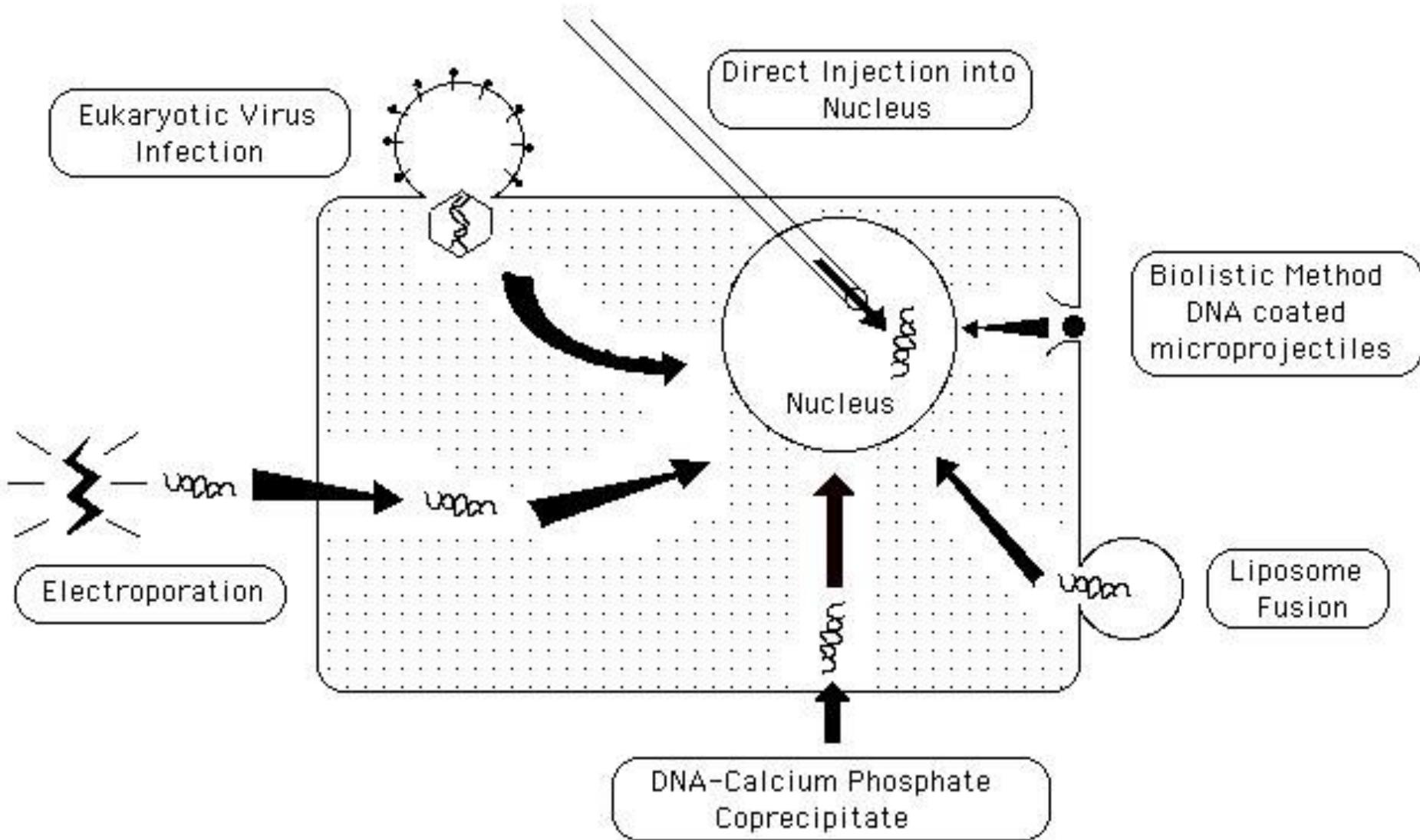
# Selectable Markers for mammalian Expression Vectors

TABLE 7.7 Selective marker gene systems for mammalian cells

Selective agent	Action of selective agent	Marker gene	Action of marker gene protein
Xyl-A	Damages DNA	Adenine deaminase ( <i>ada</i> )	Deaminates Xyl-A
Blasticidin S	Inhibits protein synthesis	Blasticidin S deaminases ( <i>Bsr</i> , <i>BSD</i> )	Deaminates blasticidin S
Bleomycin	Breaks DNA strands	Bleomycin-binding protein ( <i>Ble</i> )	Binds to bleomycin
G-418 (Geneticin)	Inhibits protein synthesis	Neomycin phosphotransferase ( <i>neo</i> )	Phosphorylates G-418
Histidinol	Produces cytotoxic effects	Histidinol dehydrogenase ( <i>hisD</i> )	Oxidizes histidinol to histidine
Hygromycin B	Inhibits protein synthesis	Hygromycin B phosphotransferase ( <i>Hph</i> )	Phosphorylates hygromycin B
MSX	Inhibits glutamine synthesis	Glutamine synthetase ( <i>GS</i> )	Cells that produce excess glutamine synthetase survive.
MTX	Inhibits DNA synthesis	Dihydrofolate reductase ( <i>dhfr</i> )	Cells that produce excess dihydrofolate reductase survive.
PALA	Inhibits purine synthesis	Cytosine deaminase ( <i>codA</i> )	Lowers cytosine levels in the medium by converting cytosine to uracil
Puromycin	Inhibits protein synthesis	Puromycin N-acetyltransferase ( <i>Pac</i> )	Acetylates puromycin

MSX, methionine sulfoximine; MTX, methotrexate; PALA, N-(phosphoacetyl)-L-aspartate; Xyl-A, 9-β-D-xylofuranosyl adenine.

# *Transfection Methods*



# Viral Vectors

## Retroviruses

- Murine leukemia virus (MuLV)
- Human immunodeficiency virus (HIV)
- Human T-cell lymphotropic virus (HTLV)

## DNA Viruses

- Adenovirus
- Adeno-associated virus (AAV)
- Herpes simplex virus (HSV)

**Retroviruses** — a class of viruses that can create double-stranded DNA copies of their RNA genomes; these copies can be integrated into the chromosomes of host cells. HIV is a retrovirus.

**Adenoviruses** — a class of viruses with double-stranded DNA genomes that cause respiratory, intestinal, and eye infections in humans. The virus that causes the common cold is an adenovirus.

**Adeno-associated viruses** — a class of small, single-stranded DNA viruses that can insert their genetic material at a specific site on chromosome 19.

**Herpes simplex viruses** — a class of double-stranded DNA viruses that infect a particular cell type, neurons. Herpes simplex virus type 1 is a common human pathogen that causes cold sores.

# Viral Attributes

Viral Vector	DNA Insert Size	Maximum Titer	Cell Type	Expression	Pitfalls
Retroviral	8 kb	$1 \times 10^9$	Dividing cells	Stable	Random insertion site
Lentivirus	9 kb	$1 \times 10^9$	Dividing cells Nondividing cells	Stable	Random insertion site
Adenovirus	8 kb	$1 \times 10^{13}$	Dividing cells Nondividing cells	Transient	Highly immunogenic
Adeno-associated virus (AAV)	5 kb	$1 \times 10^{11}$	Dividing cells Nondividing cells	Stable, site-specific location	Requires helper virus to grow; difficult to remove helper virus
Herpes simplex virus	30–40 kb	$1 \times 10^9$	Dividing cells Nondividing cells	Transient	No gene expression during latent infection
Vaccinia virus	25 kb	$3 \times 10^9$	Dividing cells	Transient	Potential cytopathic effects

## Adenoviruses

Adenoviruses are DNA viruses with broad cell tropism that can transiently transduce nearly any mammalian cell type. The adenovirus enters target cells by binding to the Coxsackie/Adenovirus receptor (CAR) (Bergelson et al., 1997). After binding to the CAR, the adenovirus is internalized via integrin-mediated endocytosis followed by active transport to the nucleus, where its DNA is expressed episomally (Hirata and Russell, 2000). Although adenoviral vectors work well for transient delivery in many cell types, for some difficult cell lines such as non-dividing cells and for stable expression, lentiviral vectors are preferred. The packaging capacity of adenoviruses is 7–8 kb.

## Retroviruses

Retroviruses are positive-strand RNA viruses that stably integrate their genomes into host cell chromosomes. When pseudotyped with an envelope that has broad tropism, such as vesicular stomatitis virus glycoprotein (VSV-G), these viruses can enter virtually any mammalian cell type. However, most retroviruses depend upon the breakdown of nuclear membrane during cell division to infect cells and are thus limited by the requirement of replicating cells for transduction. Other disadvantages of retroviruses include the possibility of insertional mutagenesis and the potential for the activation of latent disease. Like adenoviruses, retroviruses can carry foreign genes of around 8 kb.

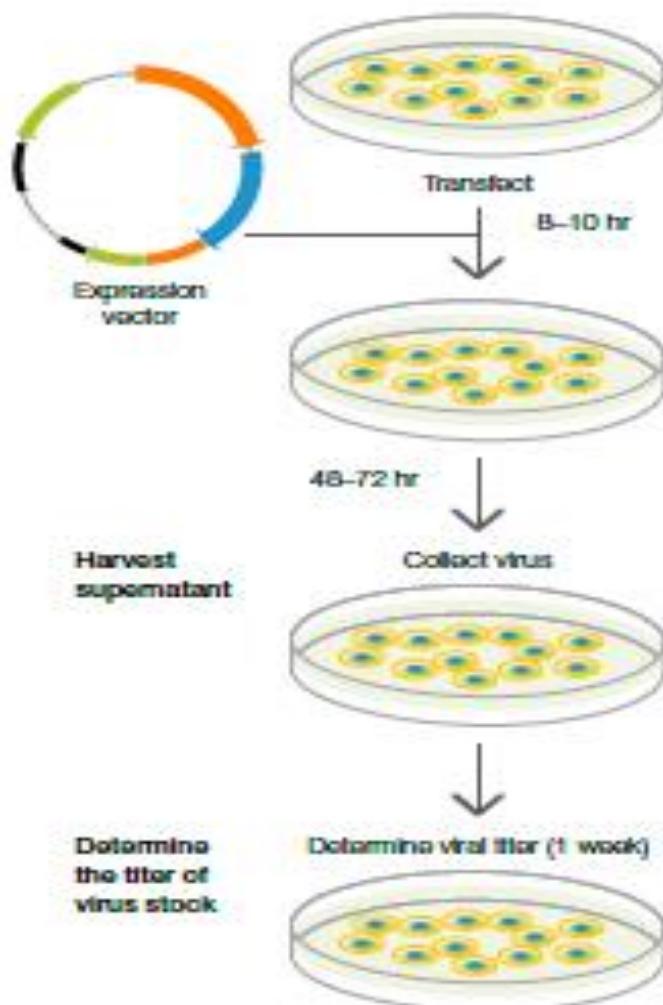
## Lentiviruses

Lentiviruses are a subgroup of the retrovirus family; as such, they can integrate into the host cell genome to allow stable, long-term expression (Anson, 2004). In contrast to other retroviruses, lentiviruses are more versatile tools as they use an active nuclear import pathway to transduce non-dividing, terminally differentiated cell populations such as neuronal and hematopoietic cells.

## Adeno-associated viruses

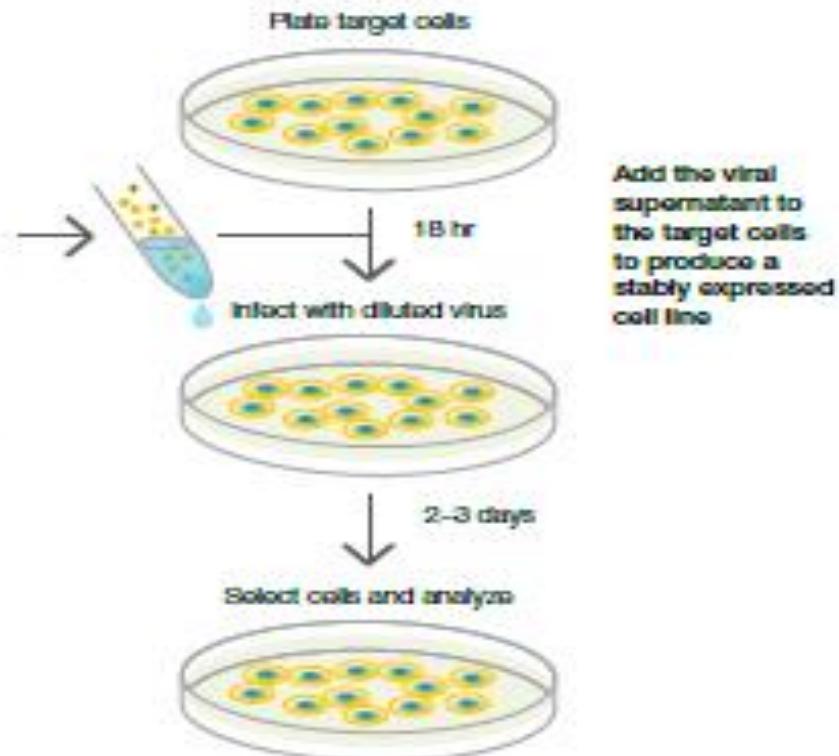
Adeno-associated viruses are capable of transducing a broad range of dividing and non-dividing cell types, but they require coinfection with a helper virus like adenovirus or herpes virus to produce recombinant virions in packaging cells. This causes difficulties in obtaining high quality viral stocks that are free of helper viruses. Furthermore, adeno-associated viruses have only limited packaging capacity of up to 4.9 kb. On the other hand, adeno-associated viruses show low immunogenicity in most cell types, and they have the ability to integrate into a specific region of the human chromosome, thereby avoiding insertional mutagenesis.

## Virus Preparation



Transfect a 293 producer cell line with the gene of interest inserted into the expression vector

## Target Cell Infection



# Pros and Cons

## Advantages of Virus-Based Methods

- Very high gene delivery efficiency, 95–100%
- Simplicity of infection

## Disadvantages of Virus-Based Methods

- Labor intensive
- Best for introducing a single cloned gene that is to be highly expressed
- P2 containment required for most viruses
  - Institutional regulation and review boards required
  - Viral transfer of regulatory genes or oncogenes is inherently dangerous and should be carefully monitored
  - Host range specificity may not be adequate
- Many viruses are lytic
- Need for packaging cell lines

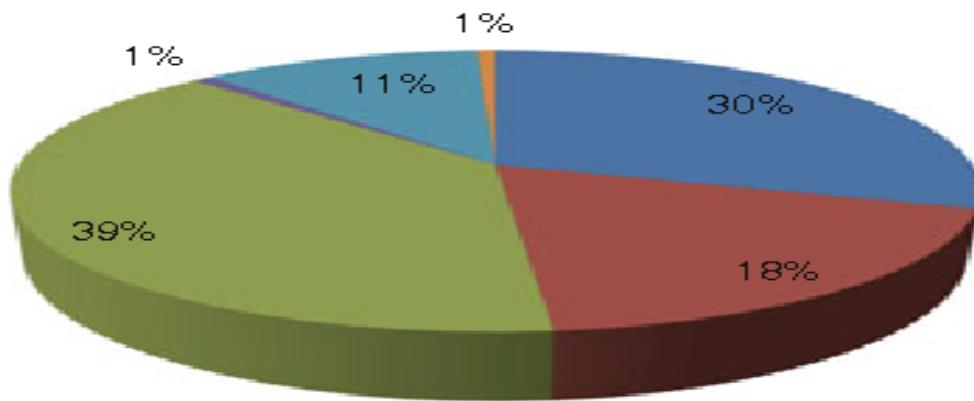
# ***Protein Drugs Produced by Eukaryotic Cell Culture***

---

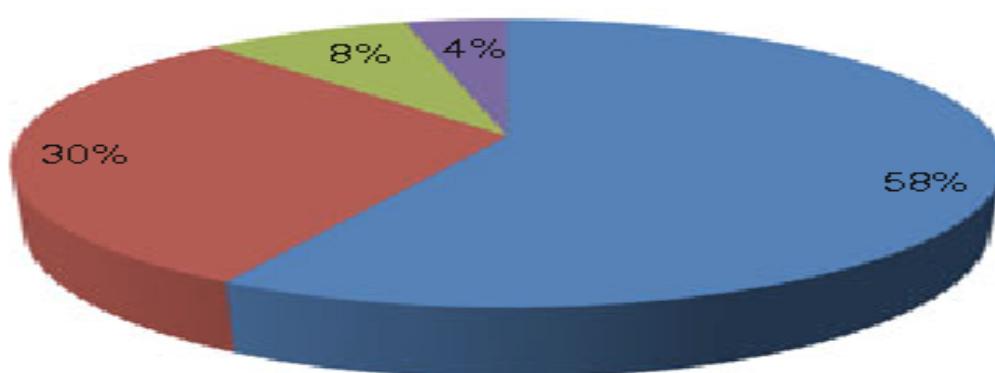
Protein	Condition
Factor IX & VIIIc	hemophiliacs
CD4 receptor	AIDS
erythropoetin	cancer
$\beta$ & $\gamma$ interferons	cancer
Interleukin-2	cancer
tissue plasminogen activator	heart attack/stroke
Hepatitis B surface antigen	vaccine
monoclonal antibodies	various

**A**

### Biopharmaceutical market

**B**

### Industrial enzyme market



# Fusion Tags



# Fusion Proteins

---

- Protein fusion done at level of DNA (genes)
  - coding region of cellular protein fused in frame to coding region of target protein gene
  - when transcribed and translated generates a ***fusion protein***

# ***Advantages of Gene fusions***

1. Efficient translation of the mRNA produced from the cloned gene depends not only on the presence of a ribosome binding site, but is also affected by the nucleotide sequence at the start of the coding region.
2. The presence of the bacterial peptide at the start of the fusion protein may stabilize, solubilize the molecule and prevent it from being degraded by the host cell.
3. The bacterial segment may constitute a signal peptide, responsible for directing the *E. coli* protein to its correct position in the cell. (secretion signal)
4. The bacterial segment may also aid purification by enabling the fusion protein to be recovered by affinity chromatography.
5. Protein localization
6. Protein protein interactions (co-localization)

Details on Fusion Tags

<https://www.frontiersin.org/articles/10.3389/fmicb.2014.00063/full>

# Fusion Tags

Fusion tags can improve protein expression, stability, resistance to proteolytic degradation and solubility. A wide range of fusion tags are available from small peptides to relatively large proteins, each with its own unique characteristics. Many solubility tags are engineered for use in bacterial expression systems to overcome poor protein solubility.

Fusion Tag	Function	Size (kDa)	Description
Polyhistidine (e.g. 6xHis, 10xHis)	Affinity	1-2	The most commonly used affinity tag, binds to metal ions
Strep-tag II	Affinity	1	High affinity for engineered streptavidin
Thioredoxin (Trx)	Solubility	12	Aids in refolding proteins that require a reducing environment
Small Ubiquitin-like Modifier (SUMO)	Solubility	12	Contains a native cleavage sequence enabling tag removal with SUMO protease
Glutathione S-transferase (GST)	Solubility, affinity	26	High affinity for glutathione, often needs to be removed due to large size
Maltose Binding Protein (MBP)	Solubility, affinity	41	Binds to maltose, often needs to be removed due to large size

## Combinatorial fusion tags



A combination of fusion tags can be used to maximise their functionality. A popular method is to utilise a solubility tag (e.g. GST or MBP) and an affinity tag (e.g. polyhistidine). This combination promotes soluble protein production and provides multiple options for affinity purification.

# ***Secretion of Protein of Interest***

---

- ***Why secrete Protein of Interest?***
  - stability of the protein may be increased
  - protein may be easier to purify
- ***Stability of the protein may be increased***
  - remove from cytoplasm and its proteases
  - recombinant human proinsulin (rh-proinsulin) stability is increased 10X if secreted

The secretion signals offered include mal, gIII, ompA, pelB, phoA, ompC, ompT, dsbA, torT, sufl, torA, STII, EOX, lamb, MgIB, SfmC, TolB and MmAp.

# Secretion of Protein of Interest

---

- ***What is required for secretion?***
  - **Signal peptide sequence** required for passage through cell membrane
  - added at NH<sub>2</sub> end of protein (5' end of gene)

NH<sub>2</sub> - signal protein of interest - COOH

must be in-frame in gene!

# *Increase stability*

---

- Insolubility may be due to aggregation caused by incorrect folding
  - Solution:
    - ***make a fusion protein with thioredoxin***
    - 11.7 kdal protein
    - can keep the fusion protein soluble even when it makes up 40% of total protein

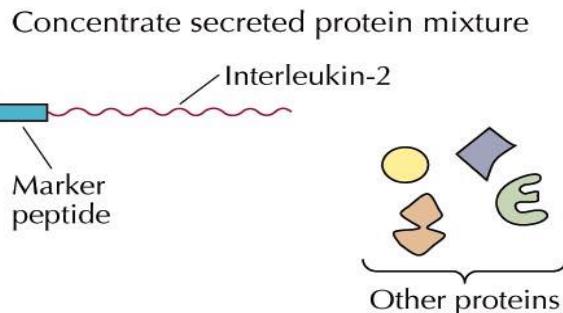
# ***Construction of Fusion Proteins might help purification***

---

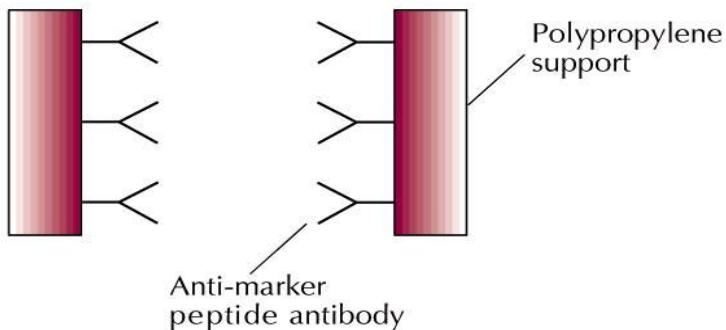
- Fusion partner (“tag”) binds to small molecule or Antibody (Ab)
- Small molecule or Ab can be linked to inert matrix
- Fusion protein will bind, other proteins won’t elute purified protein from column

# Affinity Chromatography Purification

1

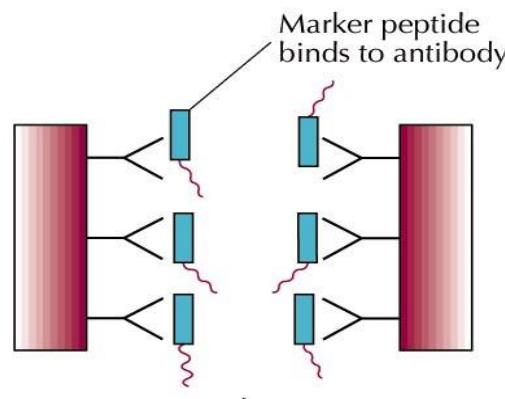


2

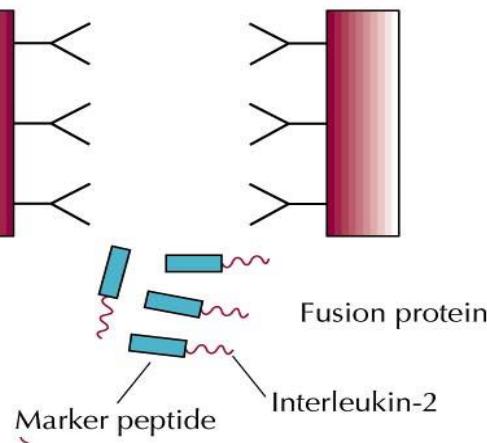


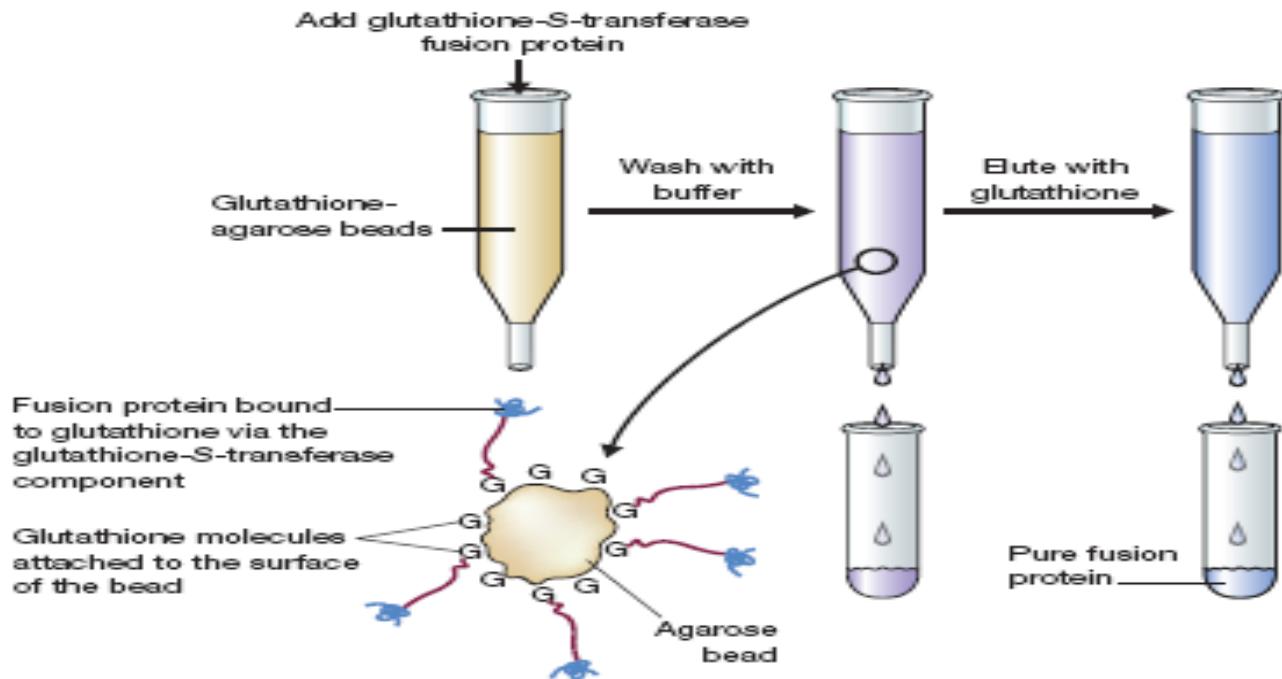
3

Add secreted protein mixture to the column



4 Elute fusion protein



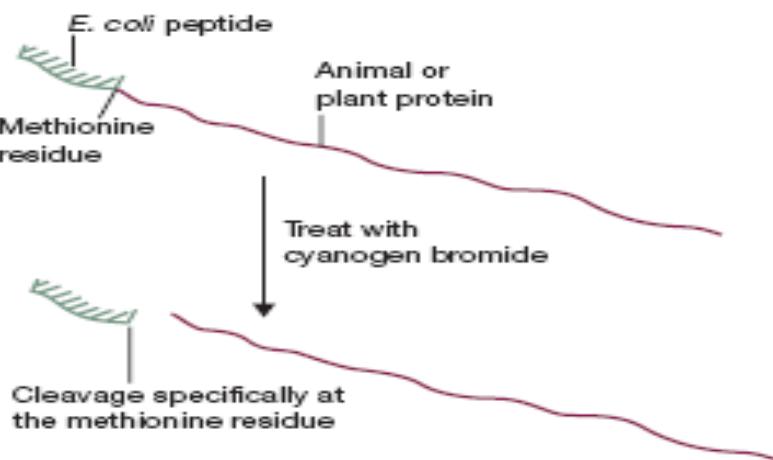


**Figure 13.12**

The use of affinity chromatography to purify a glutathione-S-transferase fusion protein.

**Figure 13.13**

One method for the recovery of the foreign polypeptide from a fusion protein. The methionine residue at the fusion junction must be the only one present in the entire polypeptide; if others are present cyanogen bromide will cleave the fusion protein into more than two fragments.



# Helper tags for protein production and purification

- **6/7 histidine tag:** interacts very specifically with Ni<sup>2+</sup> ions, which can be immobilized on columns or beads
- **GST tag-** binds to glutathione
- **Epitopes** (e.g. c-myc) for specific antibodies can be included as tags--purify on antibody column
- Tags can be engineered to be removable

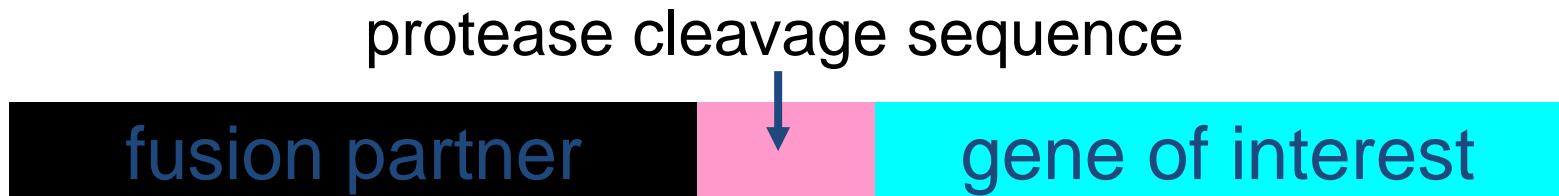
# Fusion Proteins

---

- Produce fusion protein
  - prevent degradation
  - increase ease of purification
- ***May need to remove fusion “tag”***
  - may effect biological functioning
  - may make it unsuitable for clinical use
    - FDA might not grant approval

# Fusion Proteins

- Need Proteases equivalent to REases
  - Several have been developed
  - cleave a short defined aa sequence
  - insert protease cleavage sequence between fusion partner gene (tag) and G of I



# Translation control elements

---

## Transcribed Region of Gene



- 1 AUG (Kozak sequence = CCRCC**AUGG**)
- 2 Signal sequence for secretion
- 3 Affinity tag for purification
- 4 Proteolytic cleavage site

# Tag removal



In many cases it is desirable to remove fusion tags during purification to restore native protein structure. Removal of the tag is achieved by including a cleavage site between the fusion tag and the gene sequence. Commonly used cleavage proteases include:

Protease	Type	Recognition Site
Human Rhinovirus 3C (HRV3C)	Cysteine	LEVLFQ/GP
Tobacco Etch Virus Protease (TEVp)	Cysteine	ENLYFQ/G
Ubiquitin-like Specific Protease (SUMOp)	Cysteine	Recognises tertiary structure
Thrombin	Serine	LVPA/GS

A number of challenges can be encountered during tag removal, including:

- Incomplete cleavage
- Difficulty in separating the protease and tag from the native protein
- Loss of protein from the cleavage process
- Loss of solubility following cleavage

# Fusion protein cleavage

Protease              Source

Factor Xa              blood clotting factors

Thrombin

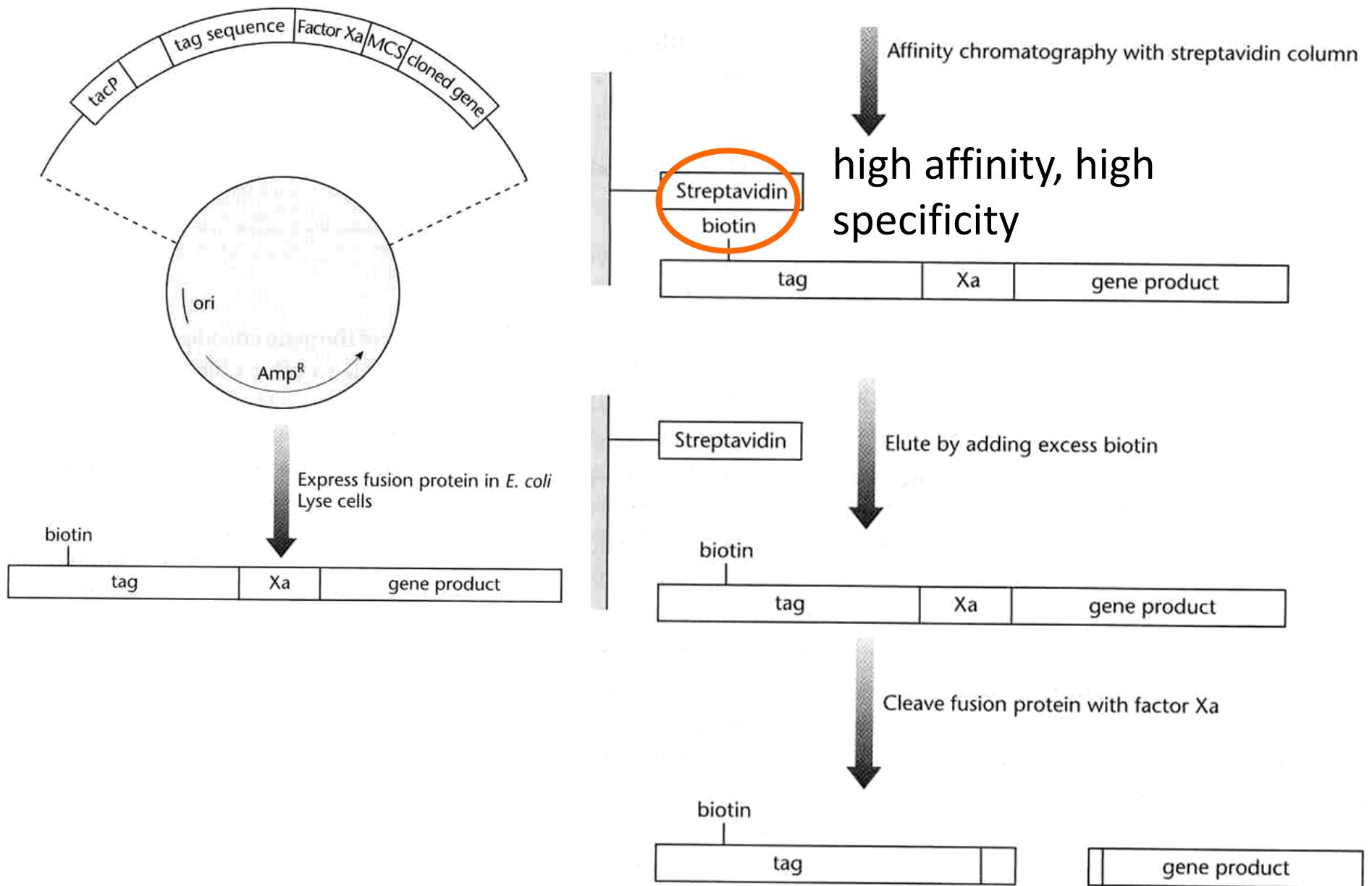
Enterokinase

Site of  
cleavage of  
factor X<sub>a</sub>

X<sub>a</sub> linker sequence

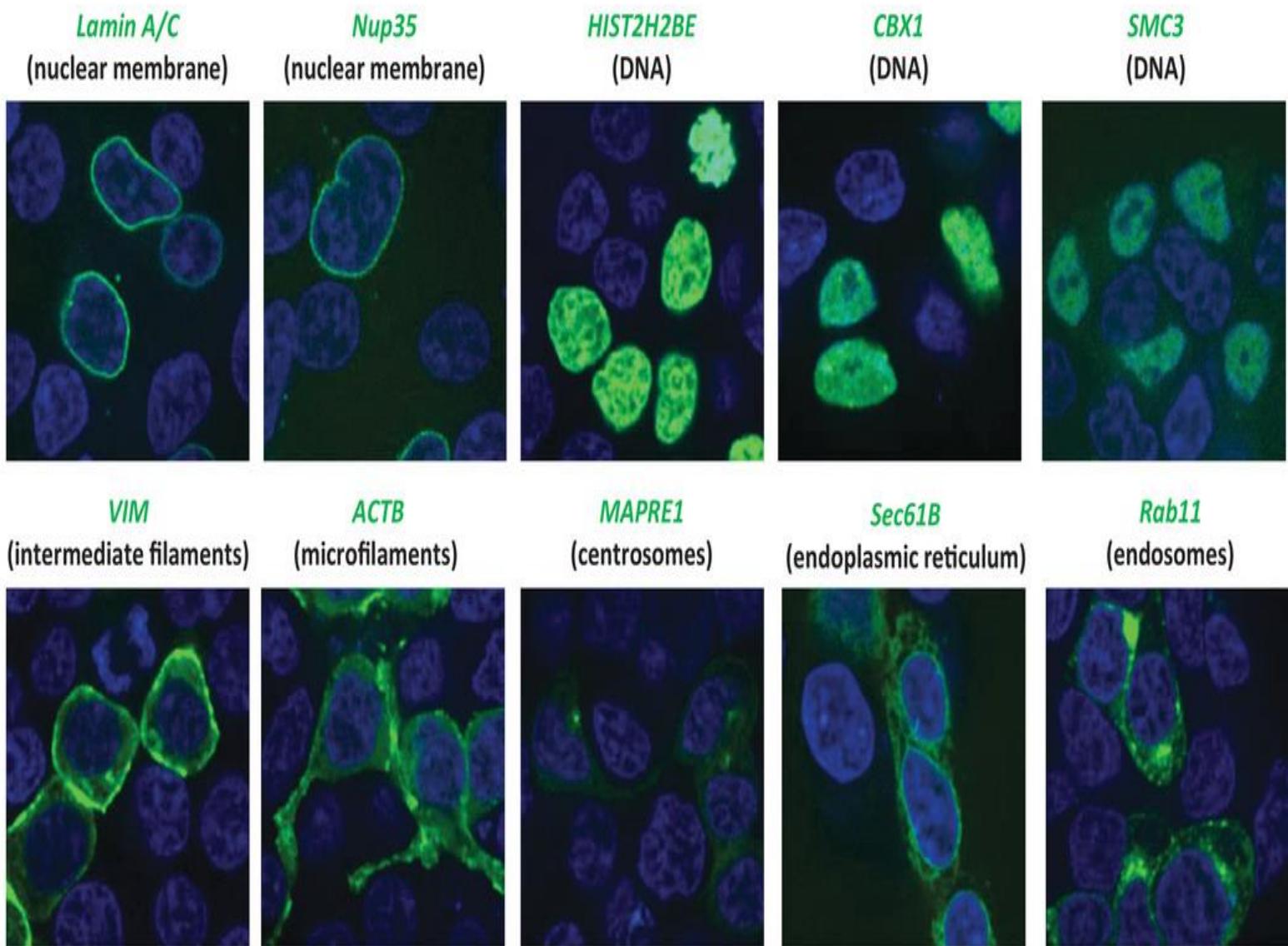


... Thr-Ala-Glu-Gly-Gly-Ser-Ile-Glu-Gly-Arg-Val-His-Leu ...

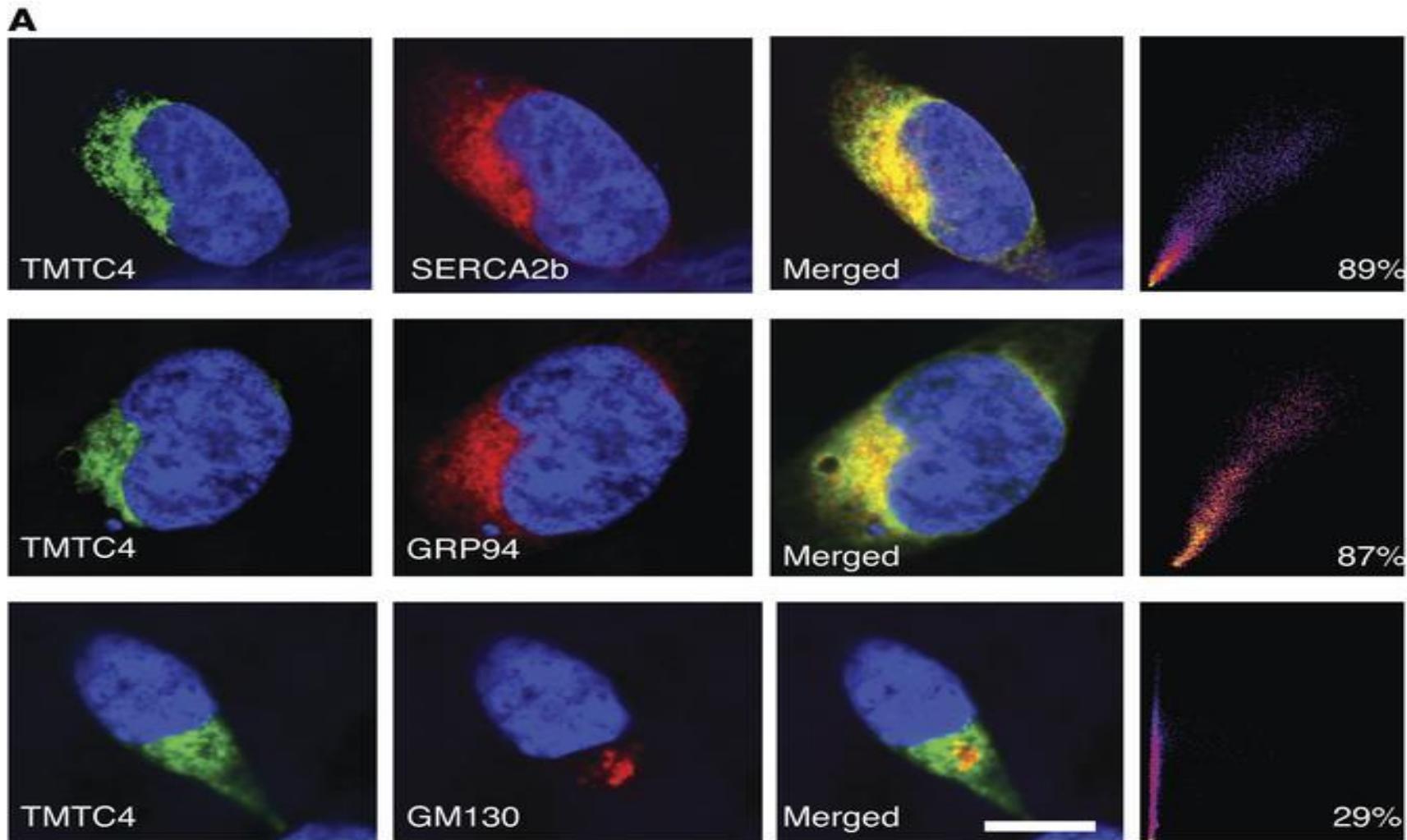


Using tags in protein purification

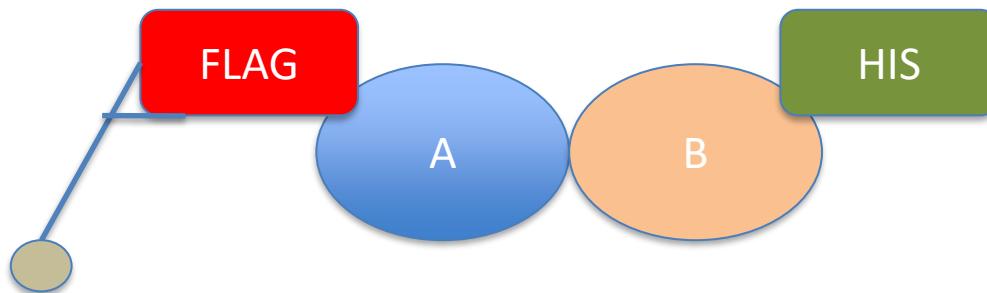
## C GFP localization patterns



# Protein co-localization



# Protein protein interaction



Pull down with FLAG Ab and perform western with His Ab

	Input	IP-FLAG
HIS	[ ]	[ ]

Conclude- Protein A and B Interact

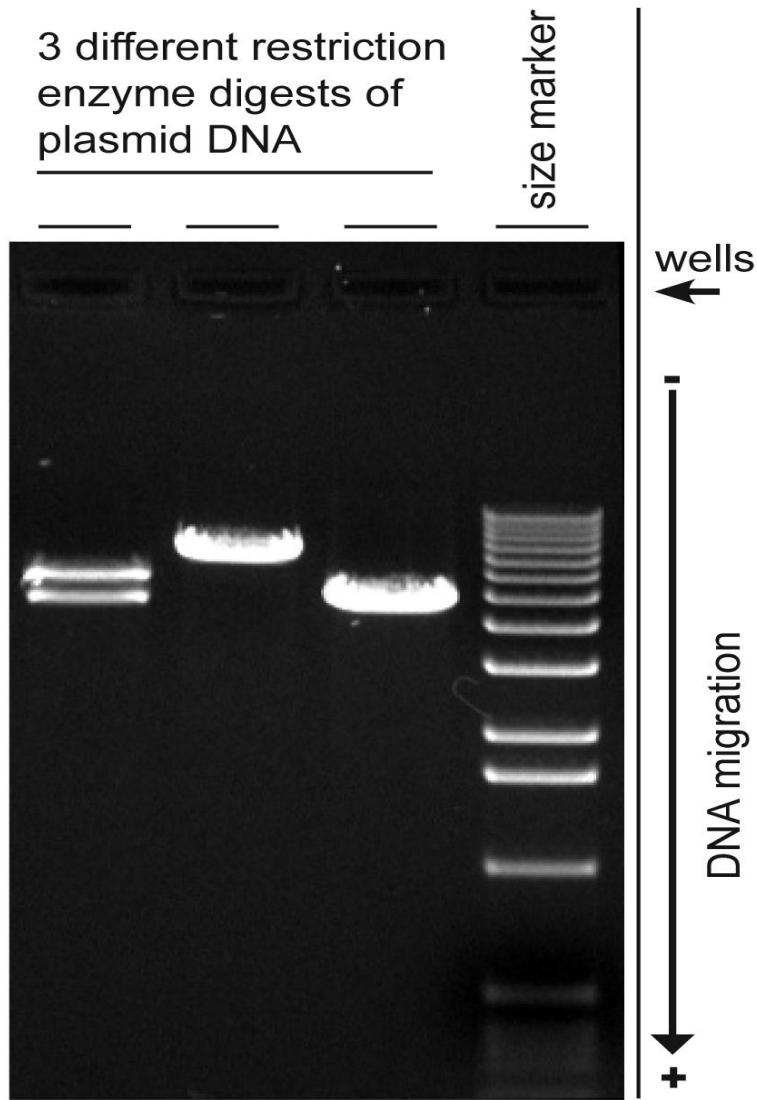
	Input	IP-FLAG
HIS	[ ]	

Conclude- Protein A and B do not Interact

Note- You can Pull down with HIS Ab and perform western with FLAG Ab

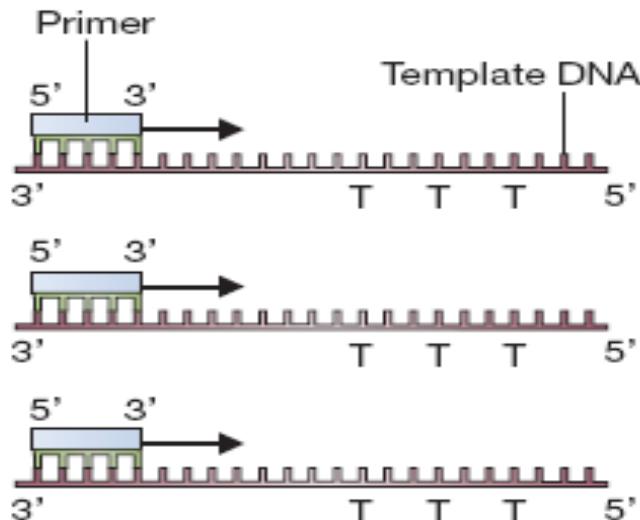
# Gene Sequencing

AAGGGACCCCTCTCCTCATTGAGAG....

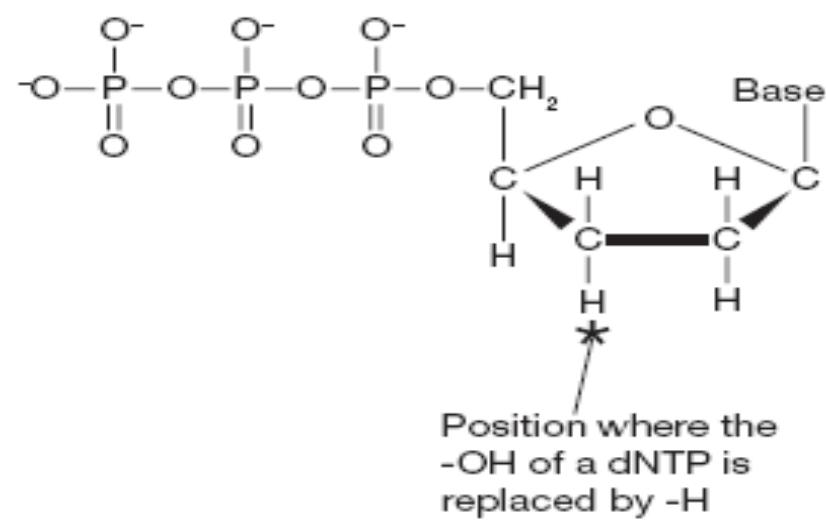


# Sanger/Chain Termination Sequencing

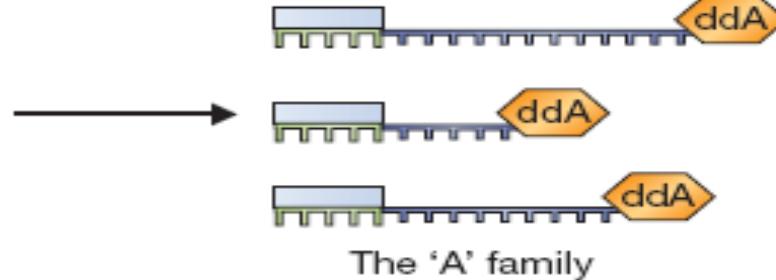
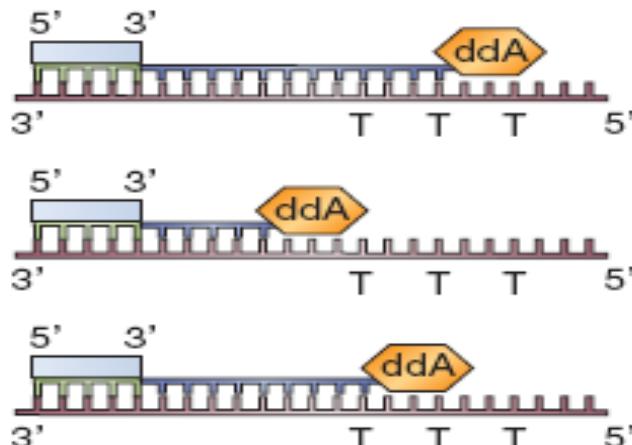
(a) Initiation of strand synthesis



(b) A dideoxynucleotide

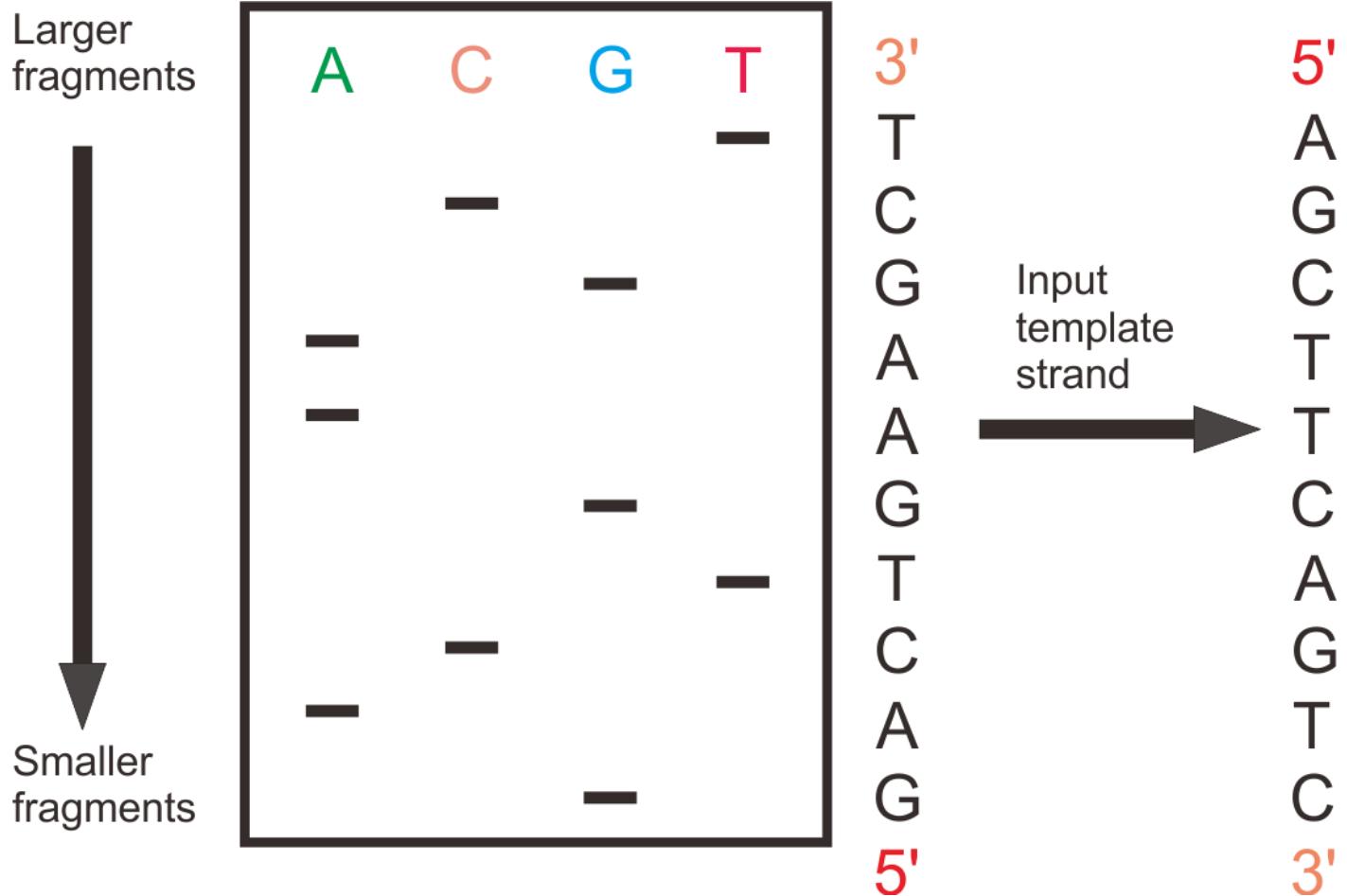


(c) Strand synthesis terminates when a ddNTP is added

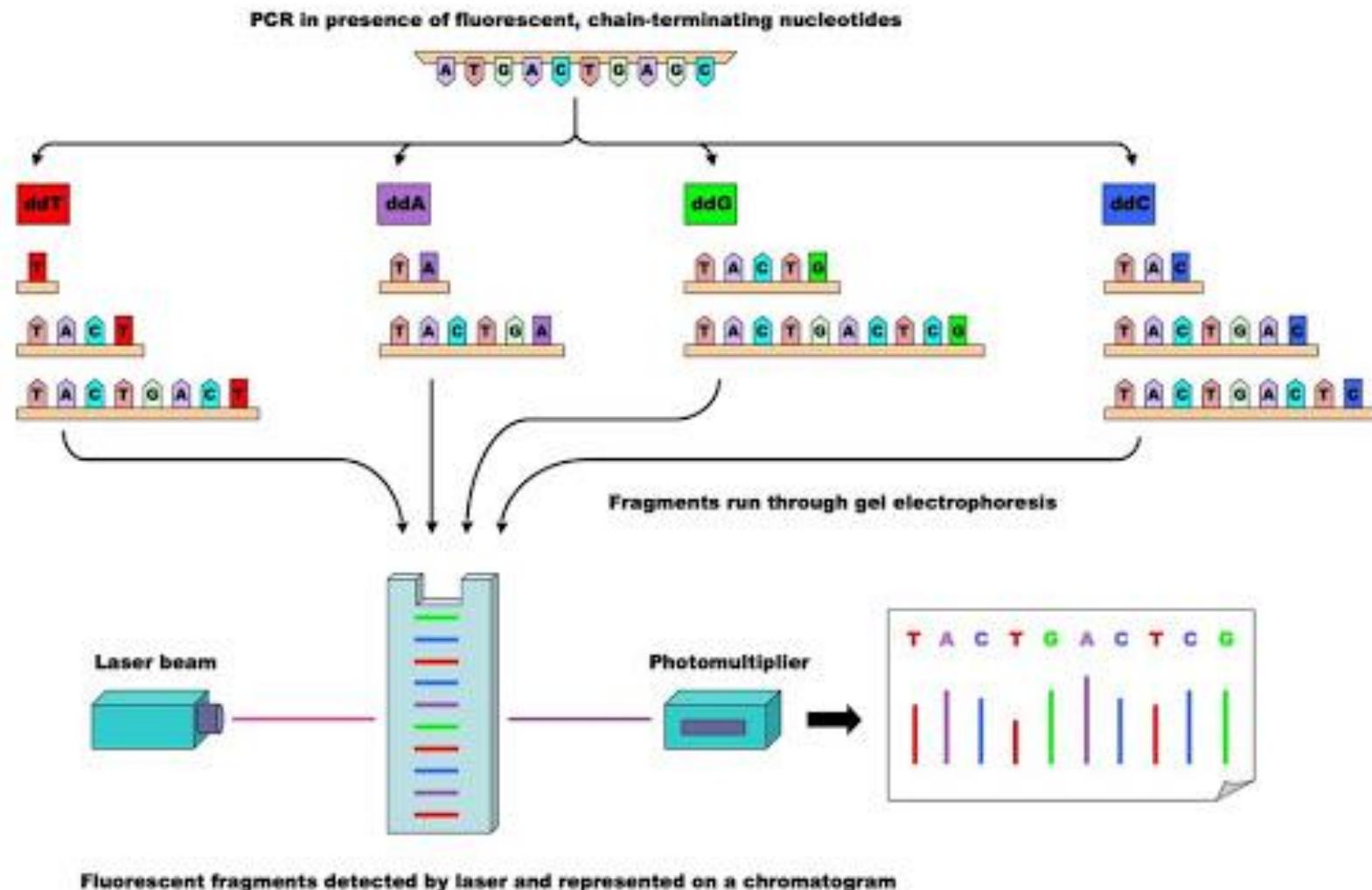


The 'A' family

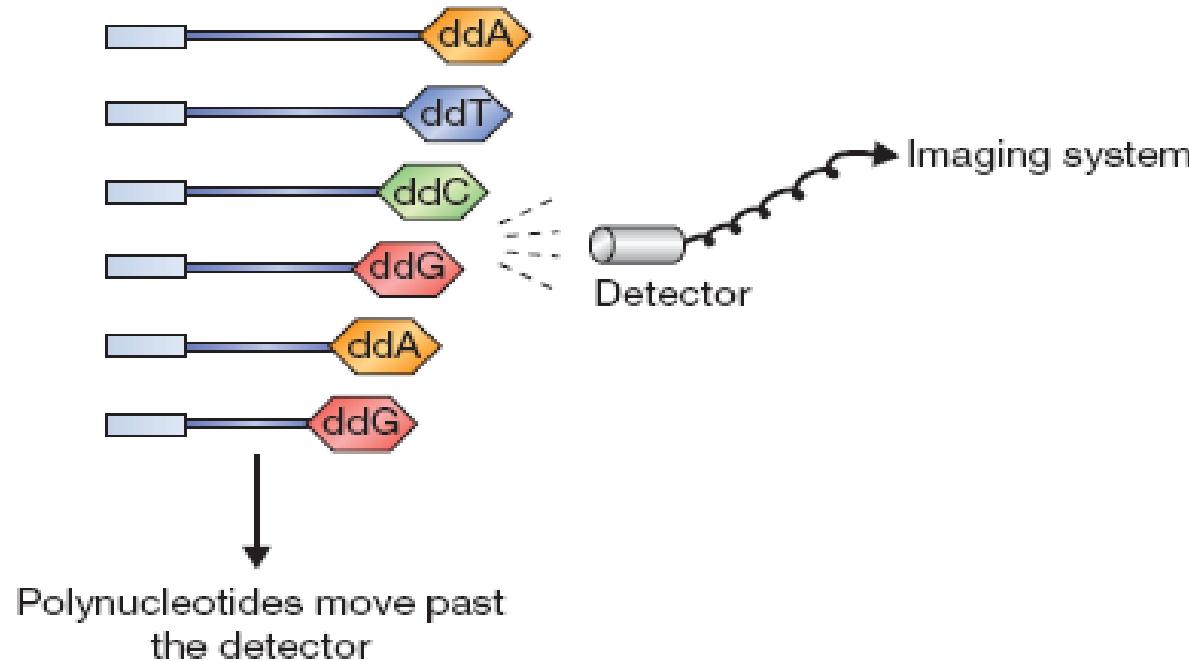
# Sequencing Gel



# Automated Sequencing

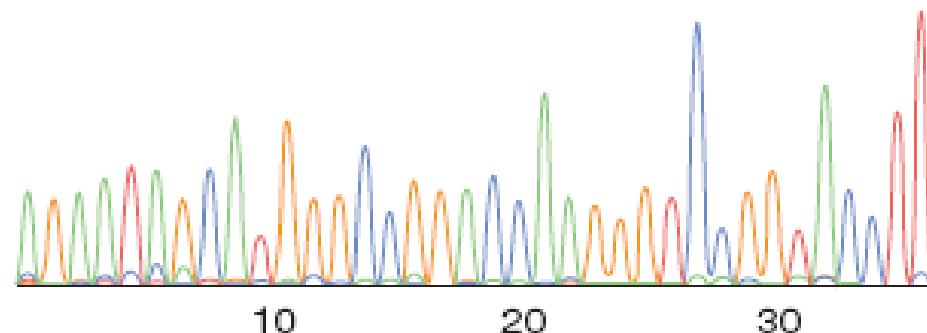


(a) Detection of chain-terminated polynucleotides

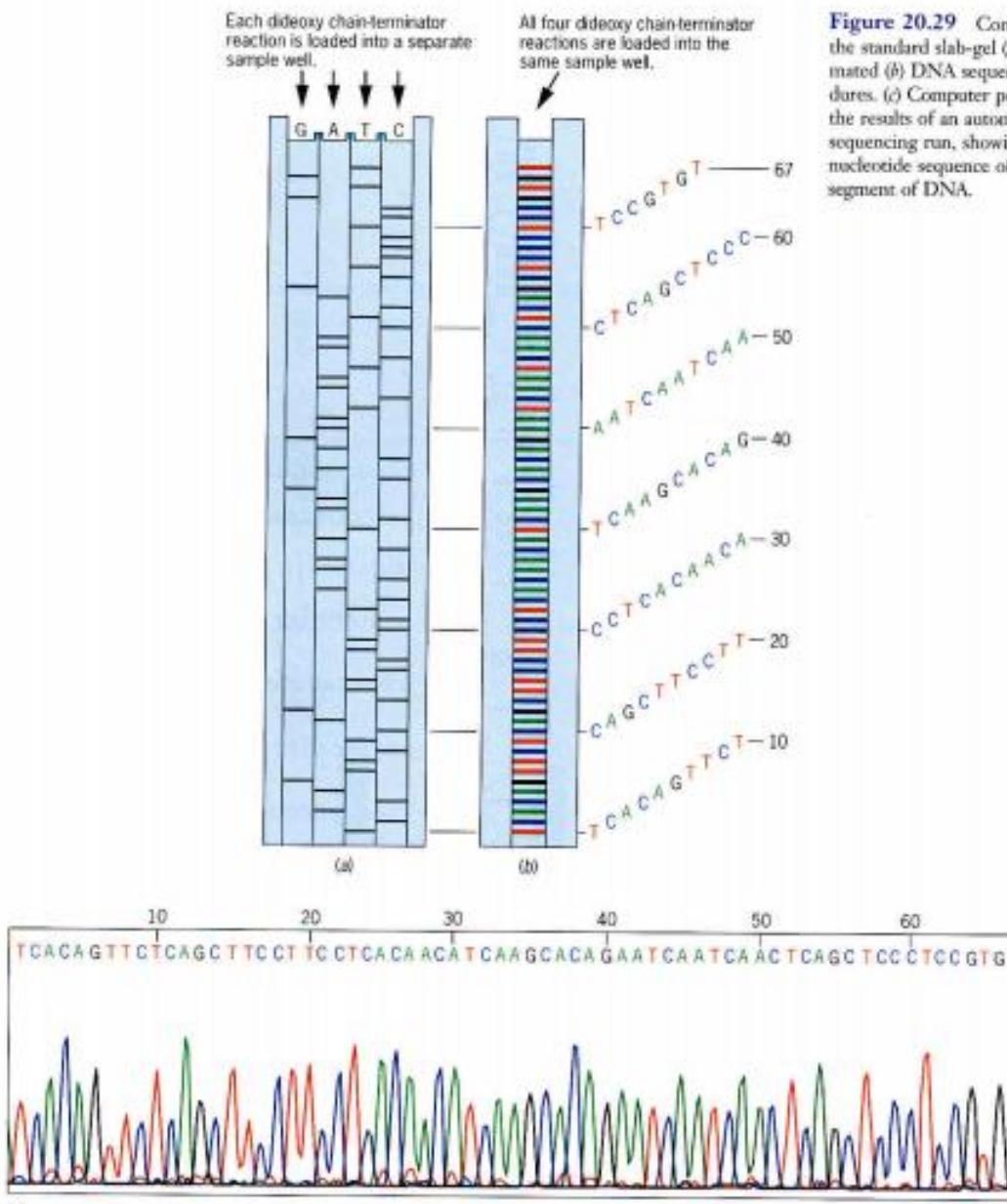


(b) The print out from an automated sequencer

CACCGCATCGAAATTAACTTCCAAAGTTAAGCTTGG



# Automated DNA sequencing



# DNA polymerase for sequencing

- Klenow polymerase has low **processivity, meaning that it can only synthesize a relatively short DNA strand before dissociating from the template due to natural causes.** This limits the length of sequence that can be obtained from a single experiment to about 250 bp.
- To avoid this problem, most sequencing today makes use of a more specialized enzyme, such as **Sequenase**, a modified version of the DNA polymerase encoded by bacteriophage T7.
- Sequenase has high processivity and no exonuclease activity and so is ideal for chain termination sequencing, enabling sequences of up to 750 bp to be obtained in a single experiment.

# Journey of Sequencing Enzyme

Read this article- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997047/>

High Processivity, no exonuclease, no discrimination between dNTP and ddNTP

**E.coli DNA polymerase I**

Low processivity



**T7 DNA polymerase- gp5 + E.coli thioredoxin**

Thioredoxin increase the processivity of the polymerase more than 100-fold

**T7 DNA polymerase**

(exonuclease activity of T7 DNA polymerase could be specifically inactivated in an oxidation reaction by oxygen, a reducing agent and ferrous ion)



**T7 DNA polymerase**

(in place of Manganese does not differentiate between dNTP and ddNTP)

**T7 DNA polymerase**

(chemical and mutagenesis screening a mutant lacking 28 amino acids in the N-terminal exonuclease domain had no detectable exonuclease activity, while its polymerase activity is significantly higher than that of the wild-type protein- This was **Sequenase 2**)

**Thermostability- Cycle Sequencing**



**Taq DNA polymerase into “Thermo Sequenase”**

Tabor and Richardson swapped the five most conserved regions in the crevice responsible for binding DNA and NTPs between T7 DNA polymerase and *E. coli* DNA polymerase I

Further mutagenesis in this helix O region revealed that the tyrosine-526 in T7 DNA polymerase or the homologous position phenylalanine-762 in *E. coli* DNA polymerase I was the single determinant for discrimination against ddNTP.

When the corresponding residue, F667 in Taq DNA polymerase was replaced with tyrosine, the modified Taq DNA polymerase F667Y actually preferred ddNTP 2-fold over dNTP, comparing to the 6000-fold discrimination against ddNTP by the wide-type enzyme

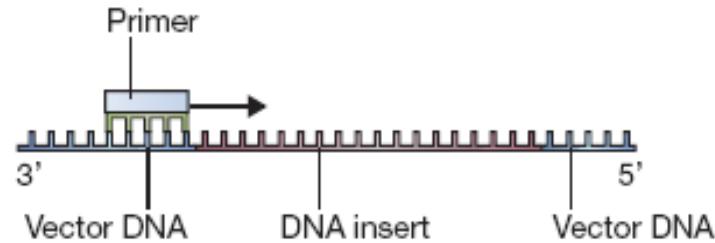
A thermostable enzyme like Taq DNA polymerase is superior for cycle sequencing, in which multiple rounds of DNA synthesis are carried out from the same template, with the newly synthesized DNA strand released after each cycle by heat denaturation.

Thermostability is desired but the strong discrimination against ddNTP (at least 100-fold, often 10,000-fold) by most thermostable DNA polymerase was a significant obstacle. Further, Manganese had several disadvantages compared with magnesium such as narrow working concentration, precipitation, and less activity of DNA polymerase than that supported by magnesium ion

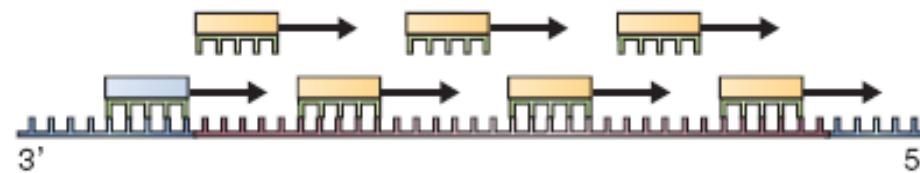
## Different types of primer for chain termination sequencing

For sequencing longer fragments  
Universal Primer and Internal Primers

(a) A universal primer



(b) Internal primers

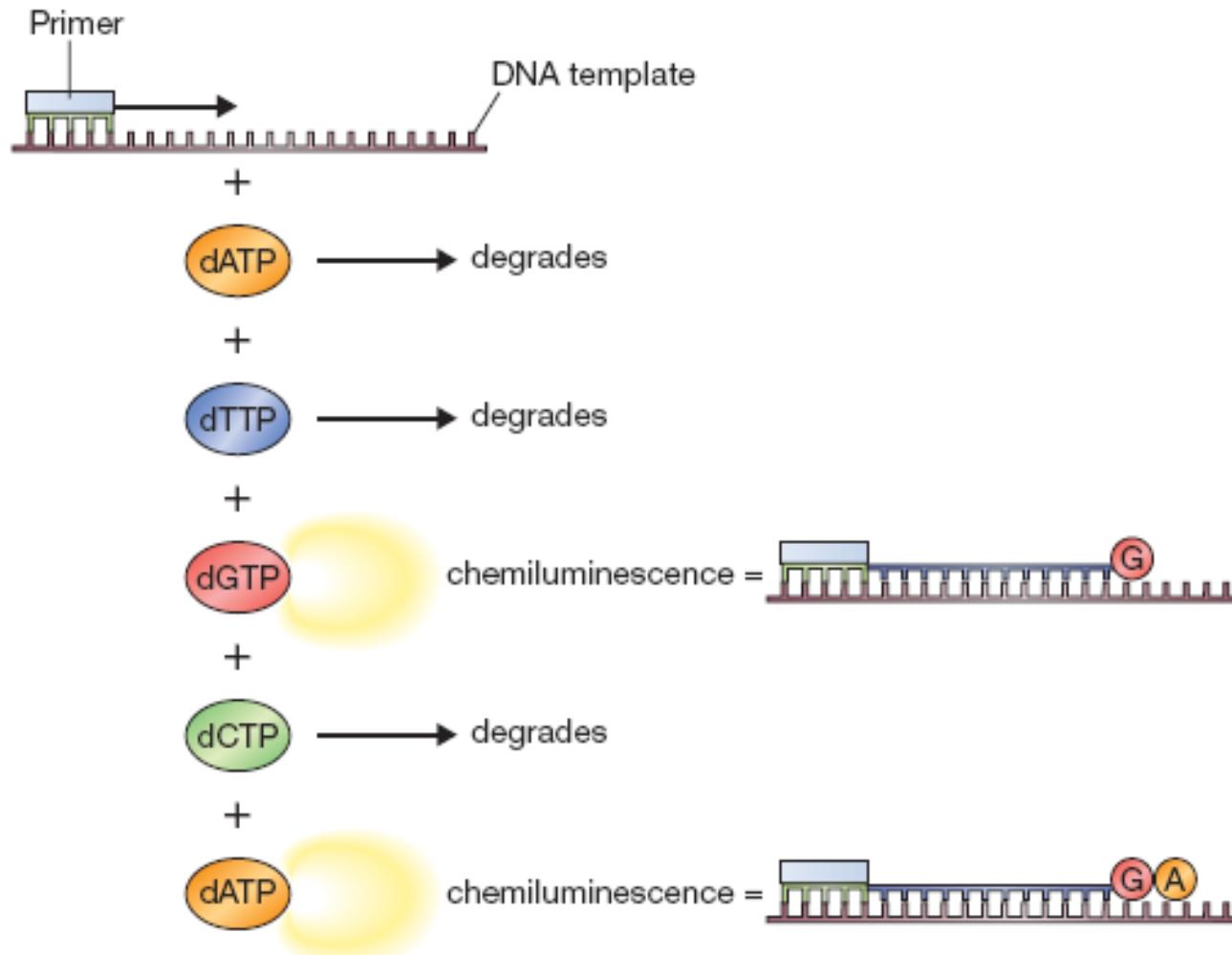


Universal primer



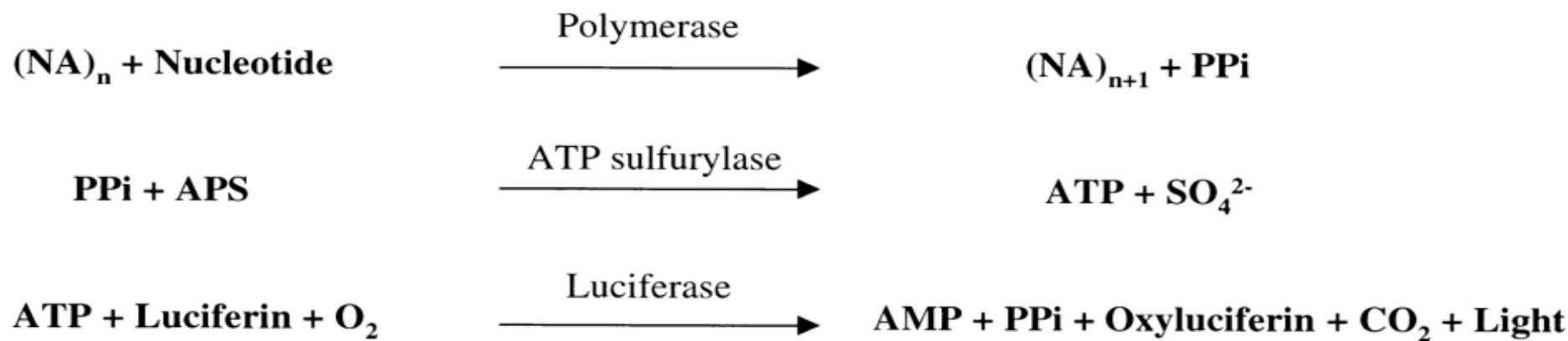
Internal primer

# Pyrosequencing



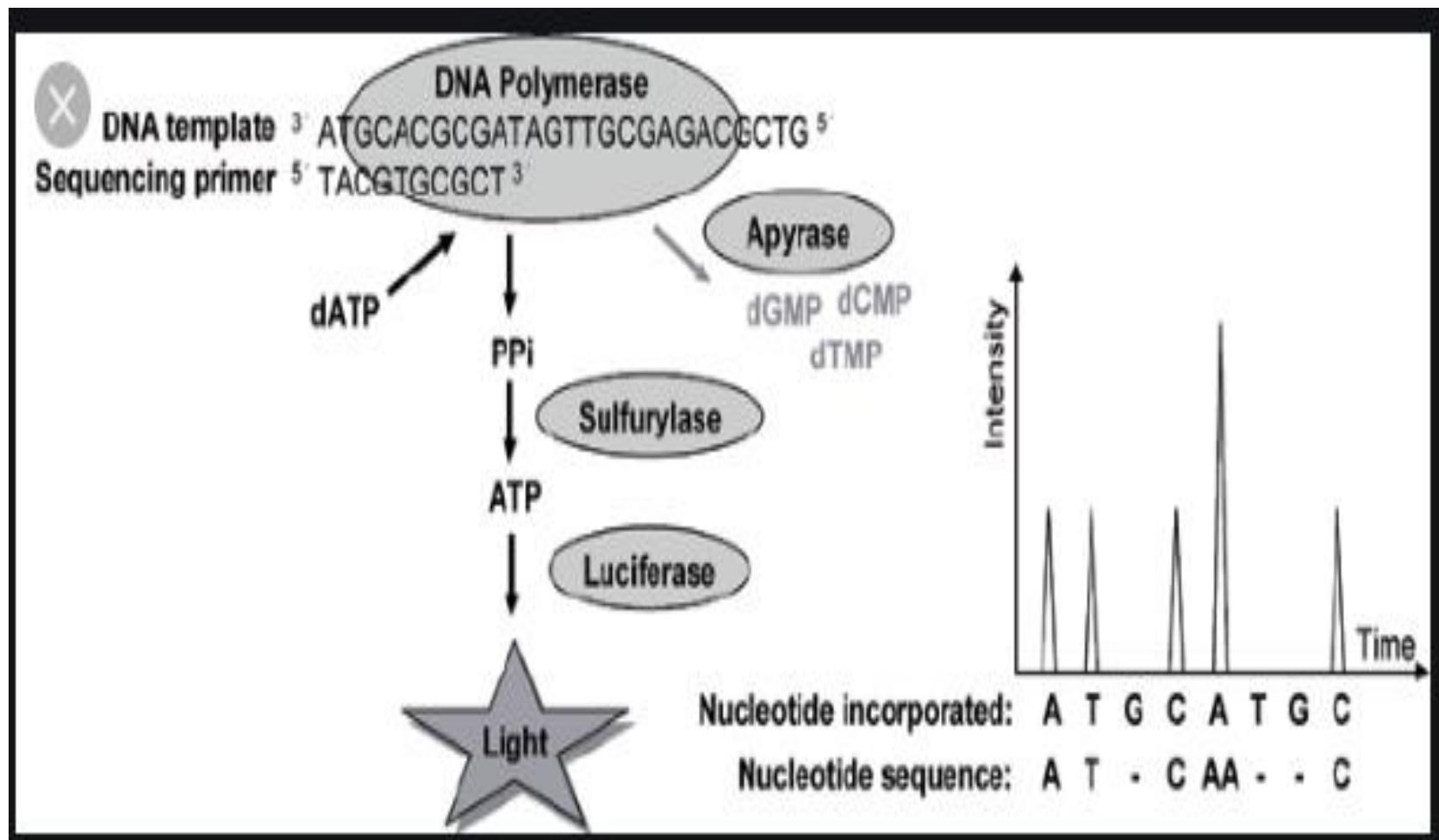
# Pyrosequencing

---



ATP sulfurylase then converts pyrophosphate to adenosine triphosphate (ATP) in the presence of adenosine 5' phosphosulfate. ATP then takes part in the luciferase-mediated conversion of luciferin to oxyluciferin. This process emits light proportionately to the amount of ATP taking part in the conversion, which is picked up by a detector.

<https://www.youtube.com/watch?v=WdTX1yykLks>



# Massive Parallel Pyrosequencing

(a) Break genomic DNA into fragments



(b) Ligate adaptors



(c) Separate strands and attach to beads



(d) PCRs in an oil emulsion

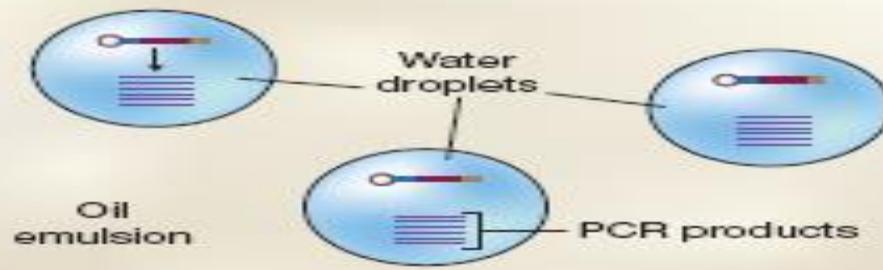


FIGURE 8

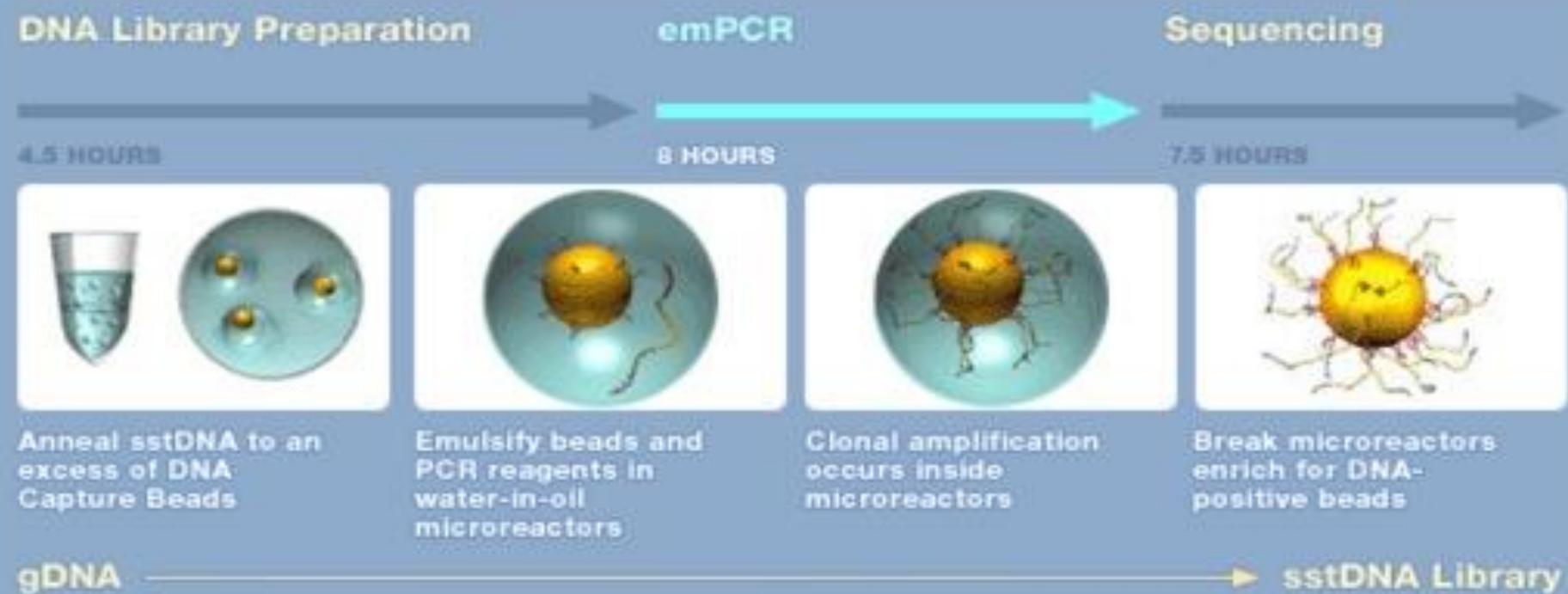


FIGURE 9

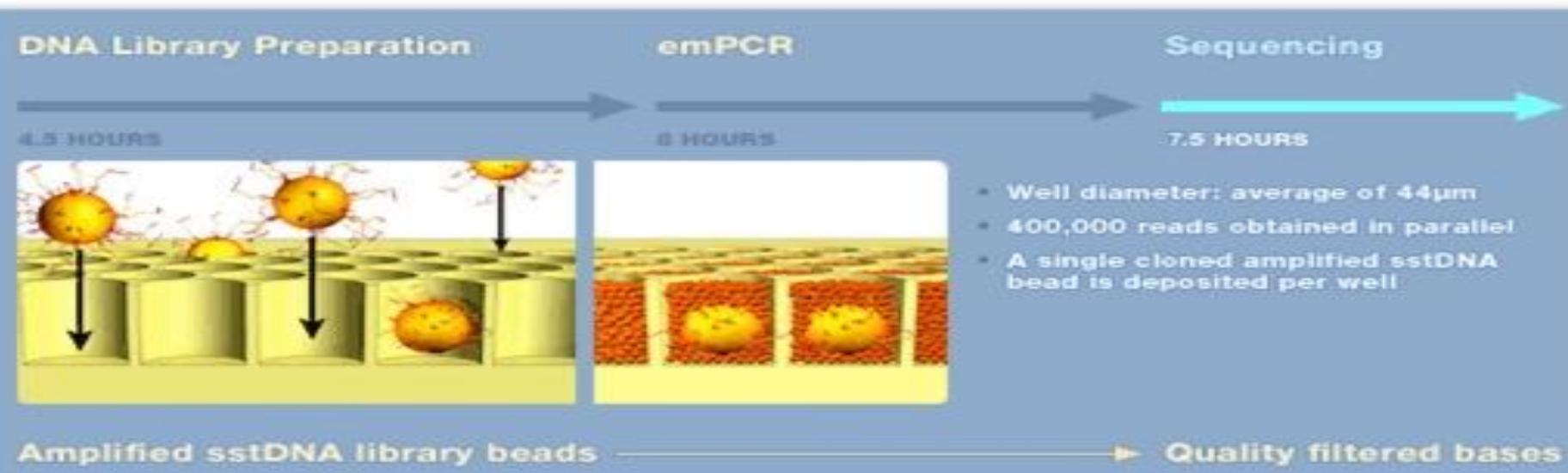
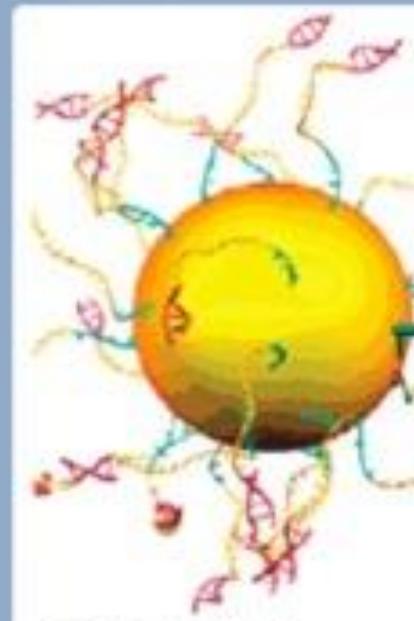


FIGURE 10

### DNA Library Preparation

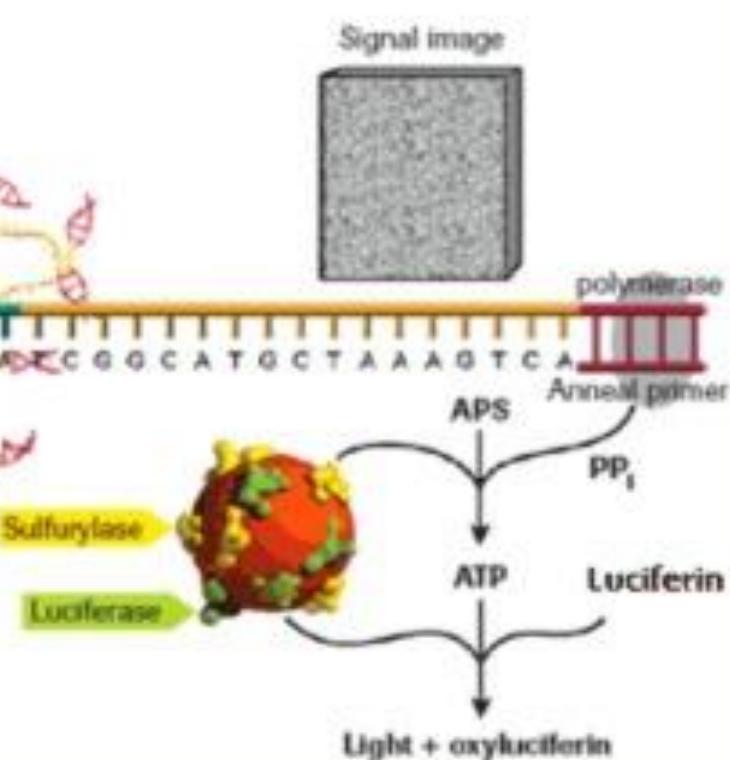
4.5 HOURS



DNA Capture Bead  
containing millions of  
copies of a single  
clonal fragment

### emPCR

8 HOURS



### Sequencing

7.5 HOURS

- 4 bases (TACG) cycled 100 times
- Chemiluminescent signal generation
- Signal processing to determine base sequence and quality score

Amplified sstDNA library beads →

Quality filtered bases

# Genome Assembly

# Genome Sequences

The first DNA molecule to be completely sequenced was the 5386 nucleotide genome of bacteriophage  $\phi$ X174, which was completed in 1975. This was quickly followed by sequences for SV40 virus (5243 bp) in 1977 and pBR322 (4363 bp) in 1978. Gradually sequencing was applied to larger molecules. Professor Sanger's group published the sequence of the human mitochondrial genome (16.6 kb) in 1981 and of bacteriophage  $\lambda$  (49 kb) in 1982. Nowadays sequences of 100–200 kb are routine and most research laboratories have the necessary expertise to generate this amount of information.

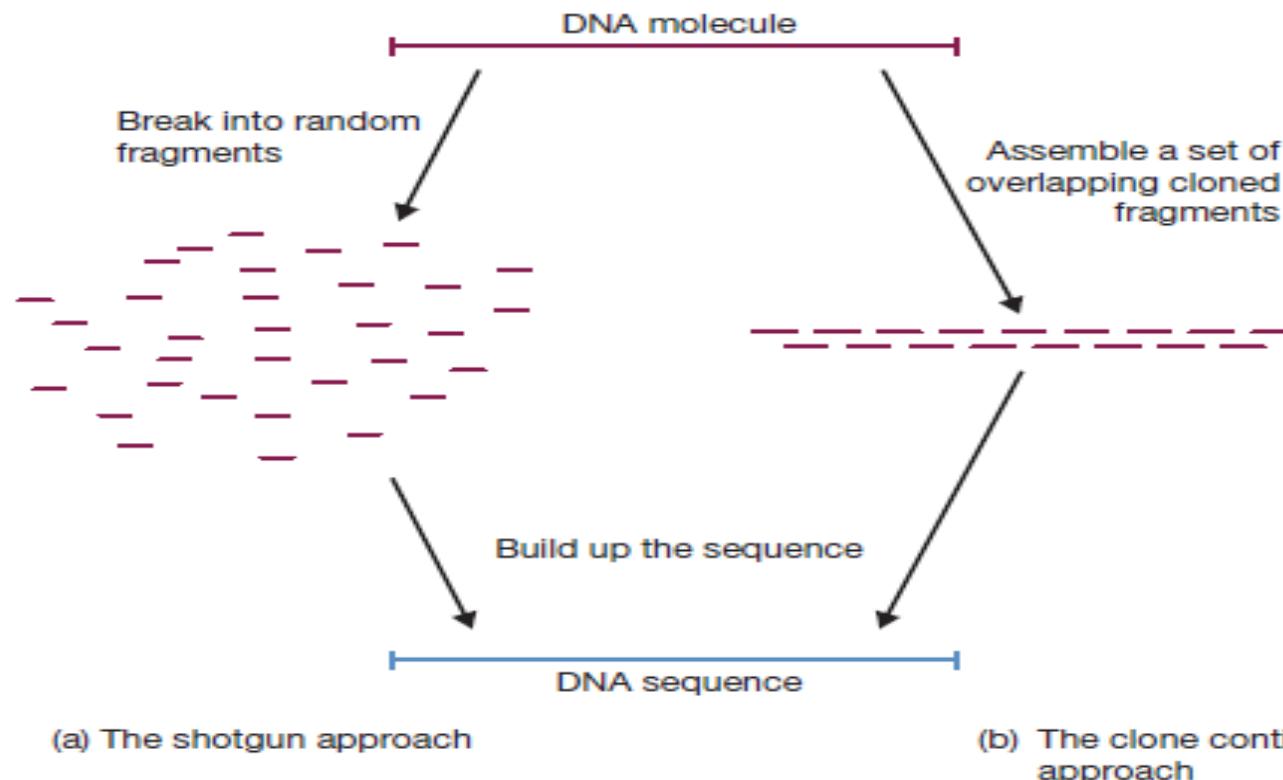
A single chain termination sequencing experiment produces about 750 bp of sequence, and a single pyrosequence yields up to 150 bp. But the total size of a fairly typical bacterial genome is 4,000,000 bp and the human genome is 3,200,000,000 bp (Table 10.1). Clearly a large number of sequencing experiments must be carried out in

Sizes of representative genomes.

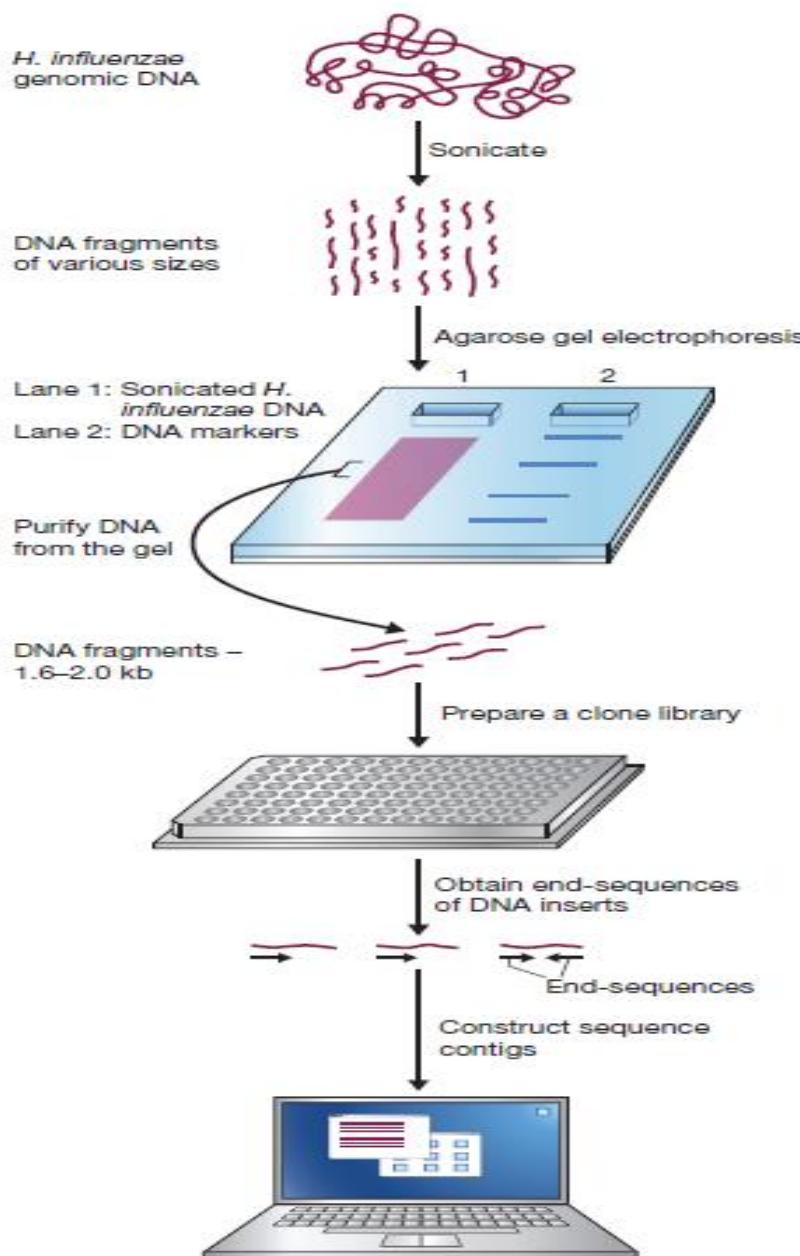
SPECIES	TYPE OF ORGANISM	GENOME SIZE (Mb)
<i>Mycoplasma genitalium</i>	Bacterium	0.58
<i>Haemophilus influenzae</i>	Bacterium	1.83
<i>Escherichia coli</i>	Bacterium	4.64
<i>Saccharomyces cerevisiae</i>	Yeast	12.10
<i>Caenorhabditis elegans</i>	Nematode worm	97.00
<i>Drosophila melanogaster</i>	Insect	180.00
<i>Arabidopsis thaliana</i>	Plant	125.00
<i>Homo sapiens</i>	Mammal	3200.00
<i>Triticum aestivum</i>	Plant (wheat)	16,000.00

# Two methods of genome assembly

- The **shotgun approach**, in which the genome is randomly broken into short fragments. The resulting sequences are examined for overlaps and these are used to build up the contiguous genome sequence.
- The **clone contig approach**, which involves a pre-sequencing phase during which a series of overlapping clones is identified. This contiguous series is called a **contig**. Each piece of cloned DNA is then sequenced, and this sequence placed at its appropriate position on the contig map in order to gradually build up the overlapping genome sequence.



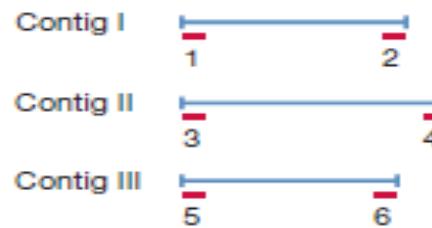
# The shotgun approach was 1st used successfully in *Haemophilus influenzae*



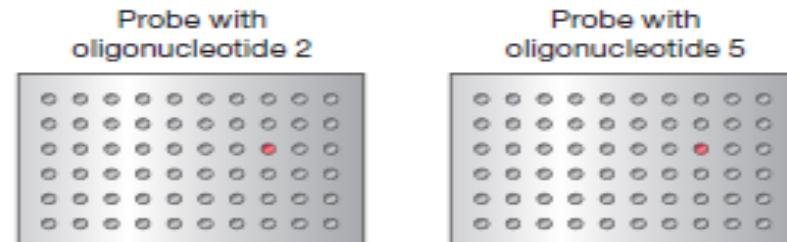
## Gap filling

Using oligonucleotide hybridization to close gaps in the *H. influenzae* genome sequence. Oligonucleotides 2 and 5 both hybridize to the same λ clone, indicating that contigs I and III are adjacent. The gap between them can be closed by sequencing the appropriate part of the λ clone.

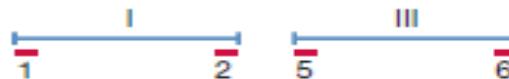
### (a) Prepare oligonucleotide probes



### (b) Probe a genomic library

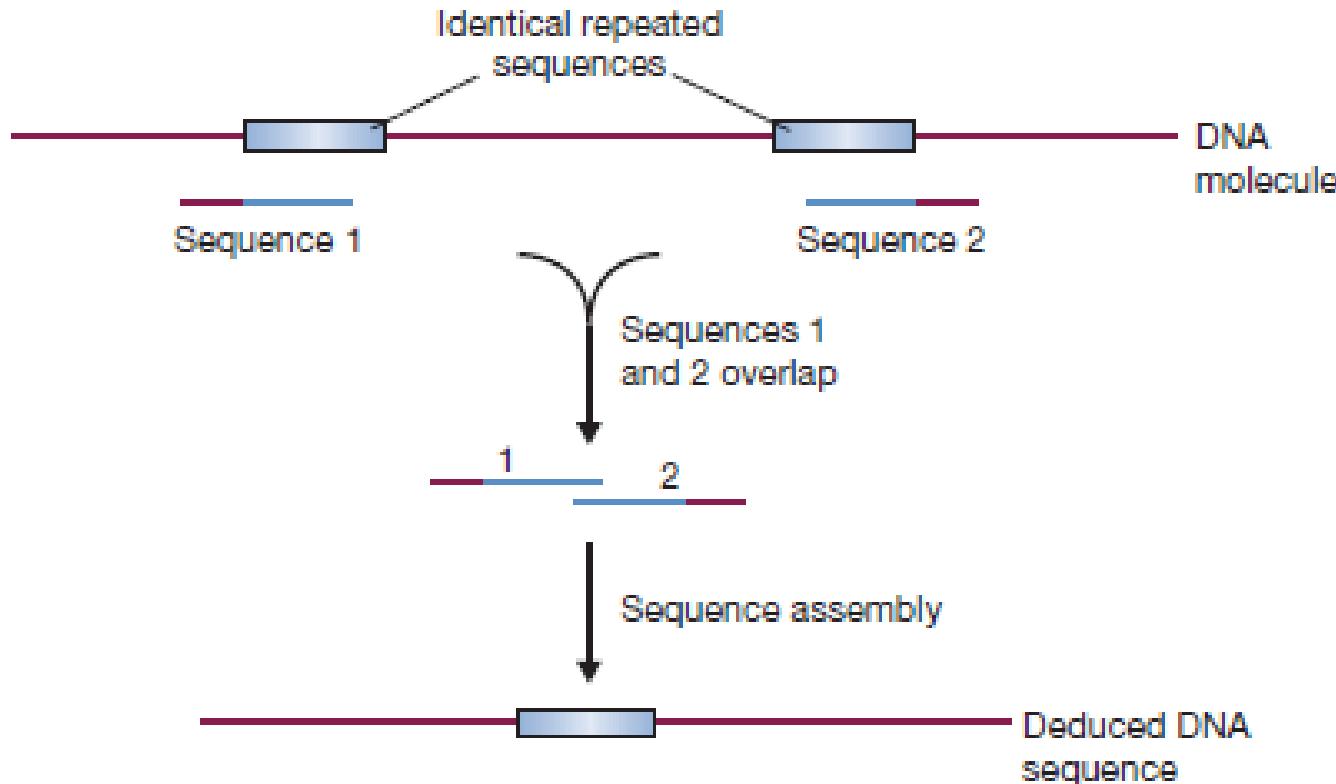


Conclusion:



Contigs I and III are adjacent in the genome

# Problem with shotgun sequencing



The probability of making mistakes increases with larger genome sizes, so the shotgun approach has been used mainly with the smaller bacterial genomes.

# Clone contig approach

The clone contig approach does not suffer from the limitations of shotgun sequencing and so can provide an accurate sequence of a large genome that contains repetitive DNA. Its drawback is that it involves much more work and so takes longer and costs more money. The additional time and effort is needed to construct the overlapping series of cloned DNA fragments. Once this has been done, each cloned fragment is sequenced by the shotgun method and the genome sequence built up step by step (see Figure 10.9).

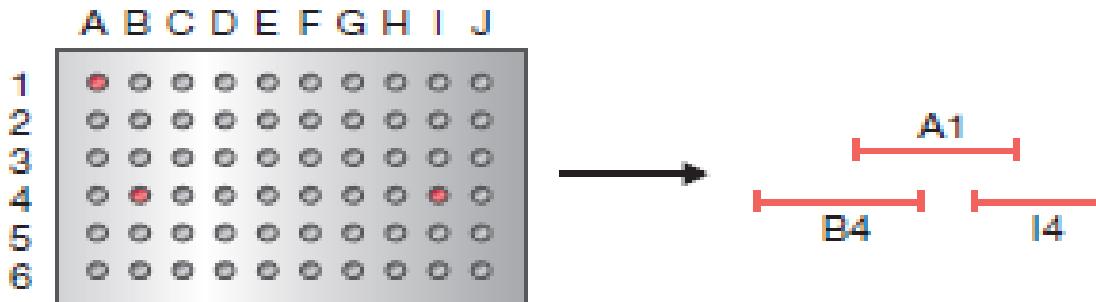
The first eukaryotic chromosome to be sequenced—chromosome III of *Saccharomyces cerevisiae* – was initially cloned in a cosmid vector (p. 101) with the resulting contig comprising 29 cloned fragments. Chromosome III is relatively short, however, and the average size of the cloned fragments was just 10.8 kb. Sequencing of the much longer human genome required 300,000 bacterial artificial chromosome (BAC) clones (p. 103). Assembling all of these into chromosome-specific contigs was a massive task.

Three methods:

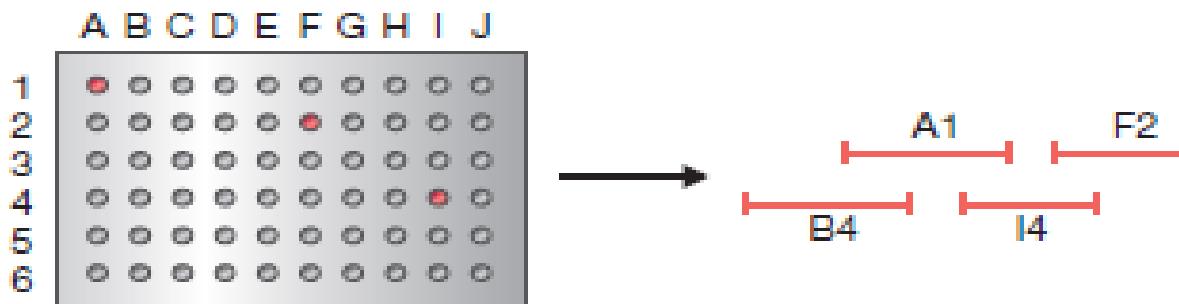
1. Chromosome Walking
2. IRE-PCR
3. STS-assembly

# Chromosome Walking

(a) Probe the library with clone A1

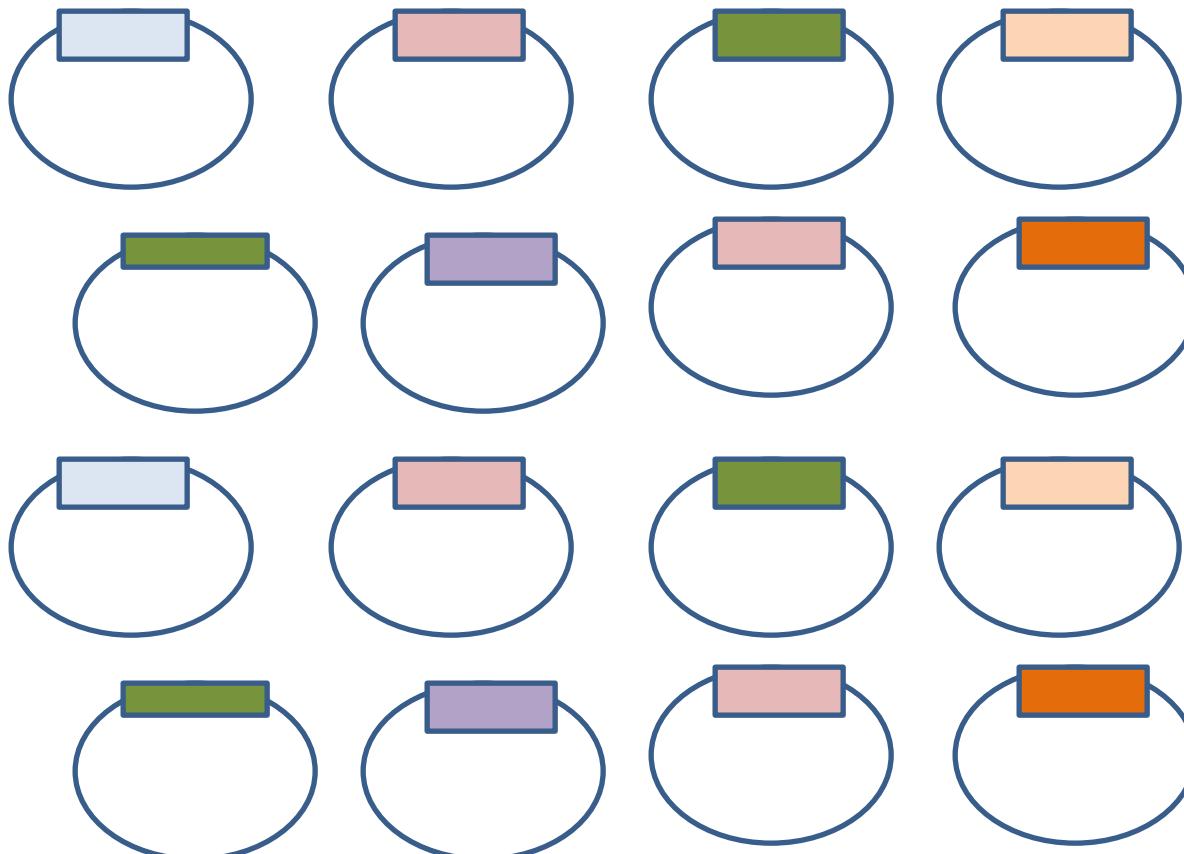


(b) Probe the library with clone I4



Gradually the clone contig is built up in a step-by-step fashion. But this is a laborious process and is only attempted when the contig is for a short chromosome and so involves relatively few clones, or when the aim is to close one or more small gaps between contigs that have been built up by more rapid methods.

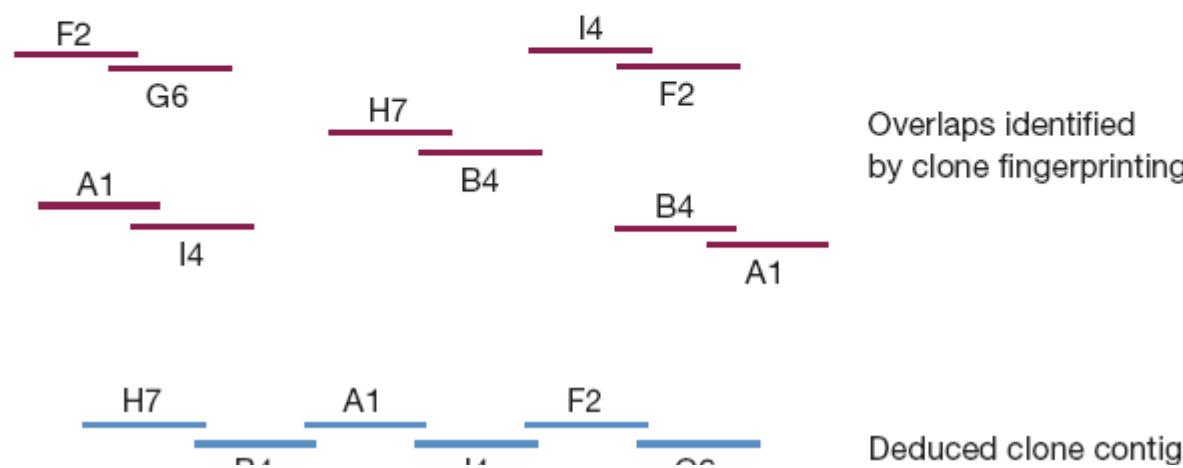
# Genomic Library- Need for rapid approach



# Rapid methods for clone contig assembly

- Clone fingerprinting

Clone fingerprinting is based on the identification of sequence features that are shared by a pair of clones. The simplest approach is to digest each clone with one or more restriction endonucleases and to look for pairs of clones that share restriction fragments of the same size, excluding those fragments that derive from the vector rather than

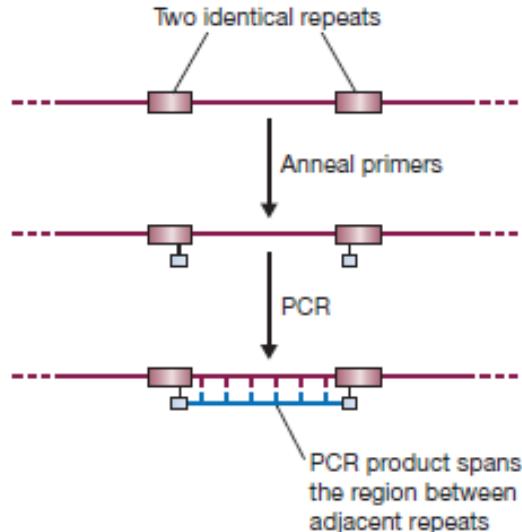


the inserted DNA. This technique might appear to be easy to carry out, but in practice it takes a great deal of time to scan the resulting agarose gels for shared fragments. There is also a relatively high possibility that two clones that do not overlap will, by chance, share restriction fragments whose sizes are indistinguishable by agarose gel electrophoresis.

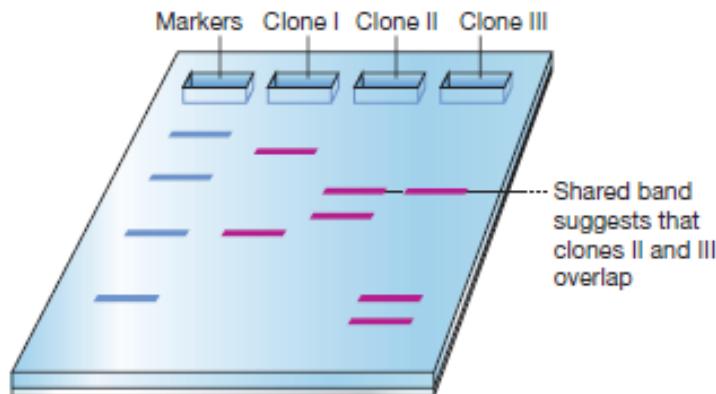
# *Rapid methods for clone contig assembly*

## Interspersed repeat element PCR (IRE-PCR).

(a) The basis to IRE-PCR

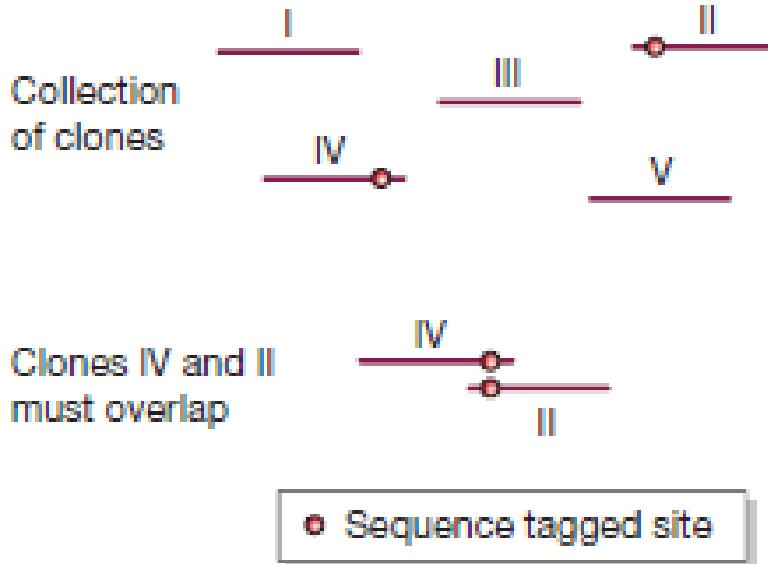


(b) Interpreting the results



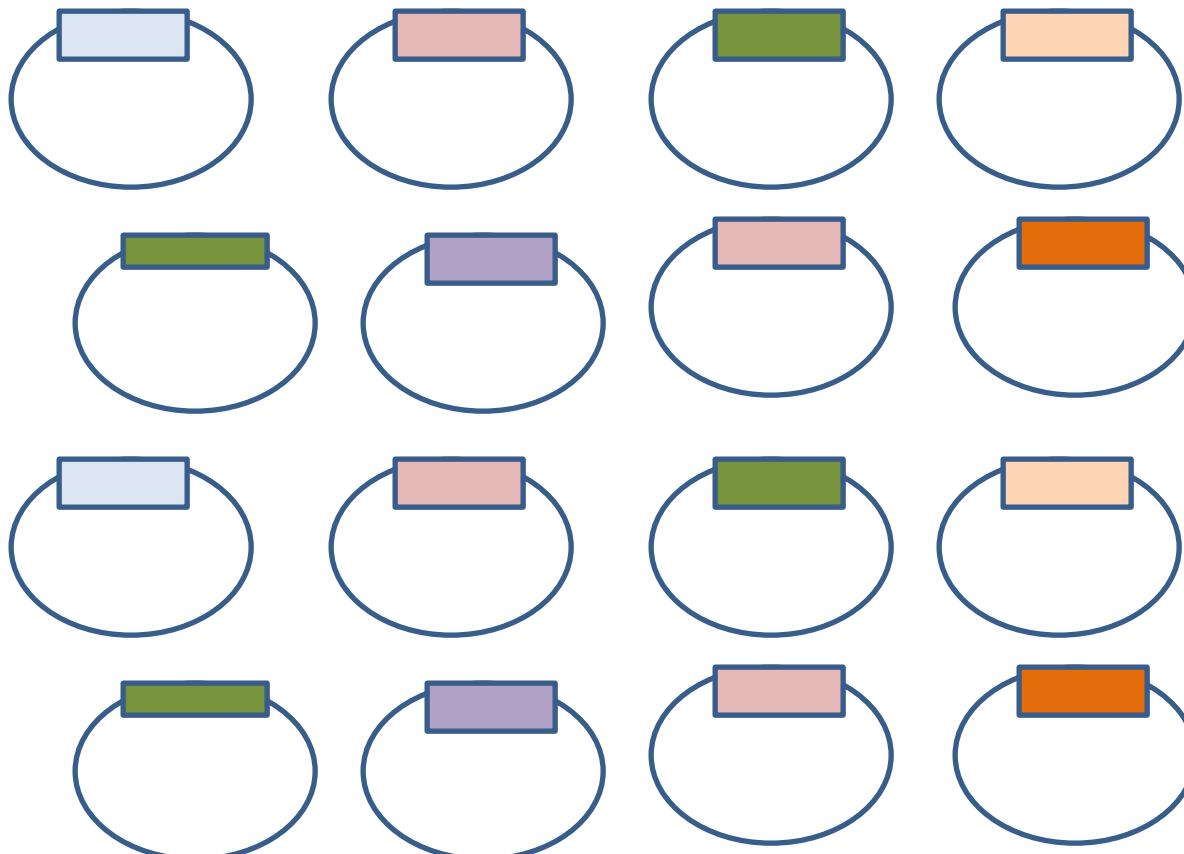
# *Rapid methods for clone contig assembly*

## Sequence Tagged Sites (STS) based Mapping



The STS need not necessarily be a gene but can be any short piece of DNA sequence, the only requirement being that it occurs just once in the genome.

# Genomic Library- Need for rapid approach



# Genetic Maps

- **Mendelian Inheritance (Pedigree Analyses)**
- More recently, techniques have been devised for genetic mapping of DNA sequences that are not genes but which still display variability in the human population
  - **Single nucleotide polymorphisms (SNPs)**
  - **Restriction fragment length polymorphisms (RFLPs)**
  - **Short tandem repeats (STRs), also called microsatellites,**
- All of these DNA markers are variable and so exist in two or more allelic forms. Their inheritance patterns can be determined by analysis of DNA prepared from the parents and offspring from a genetic cross, and the data used to place the DNA markers on a genetic map, in exactly the same way as genes are mapped.

# Genetic Map

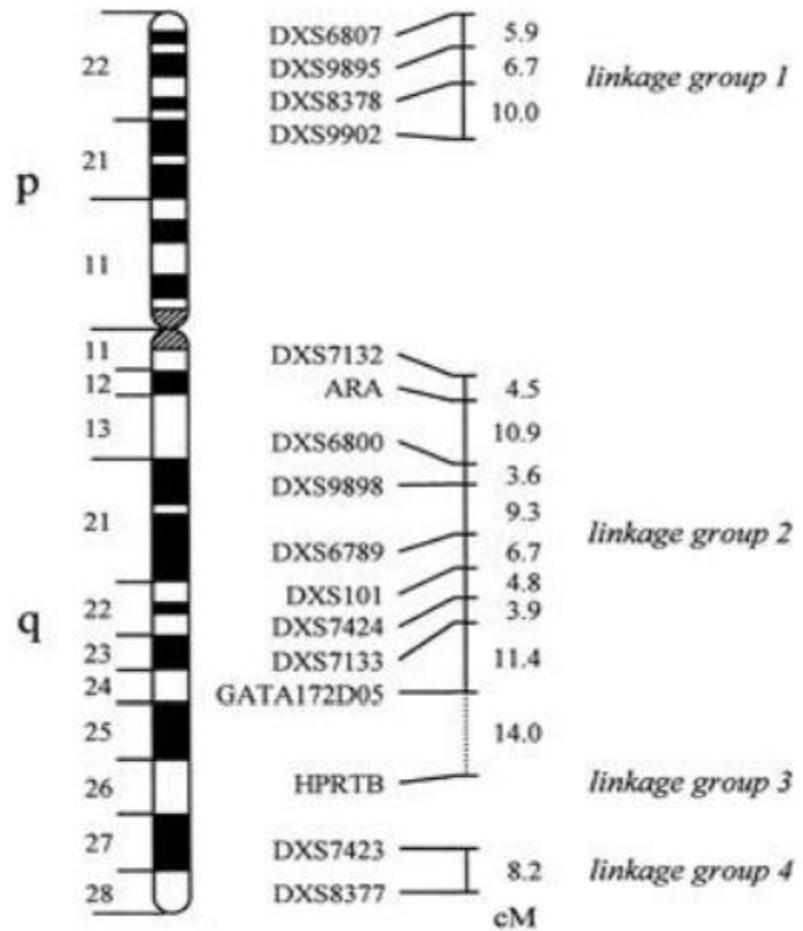
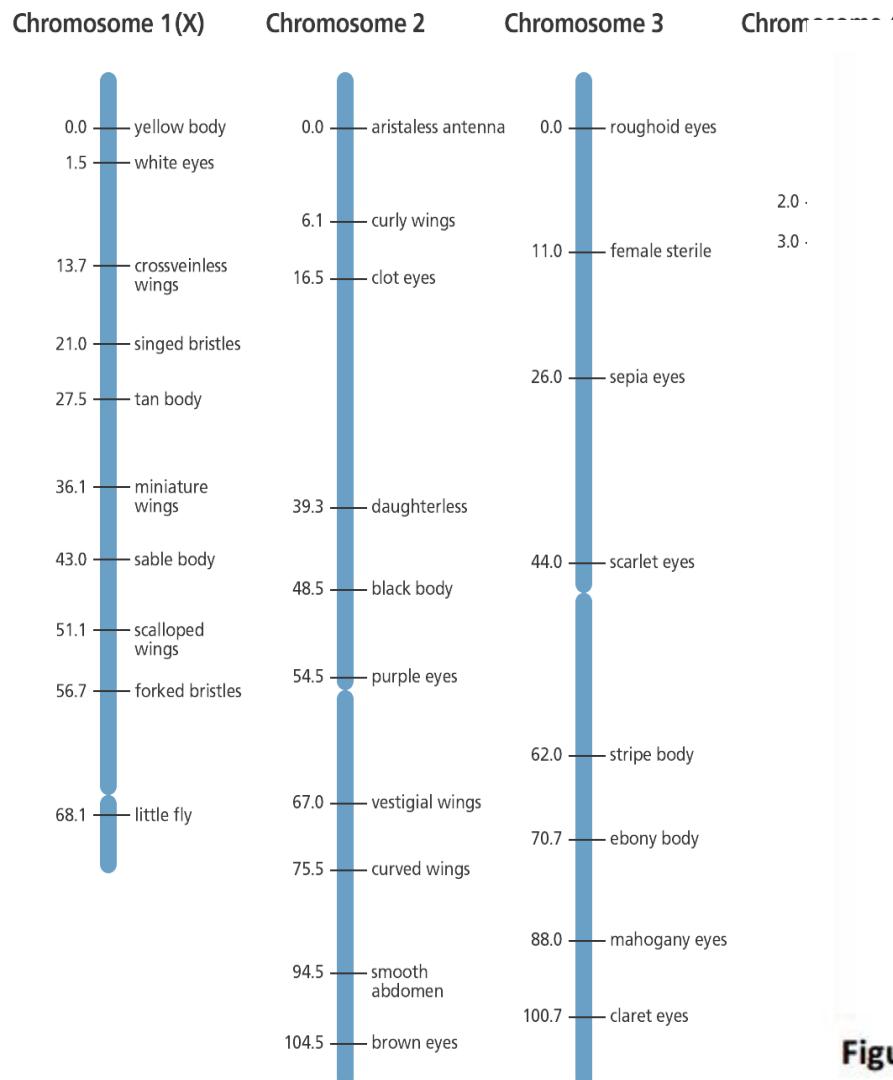
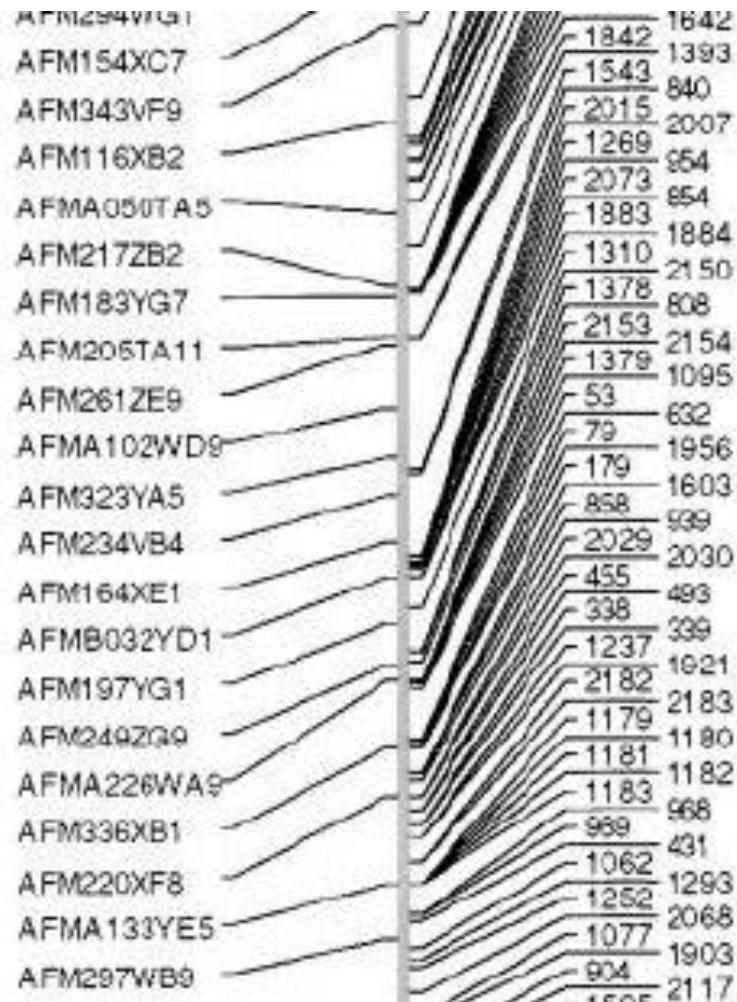


Figure 11: Localisation of ChrX STRs used in forensic practise.

# SNP map

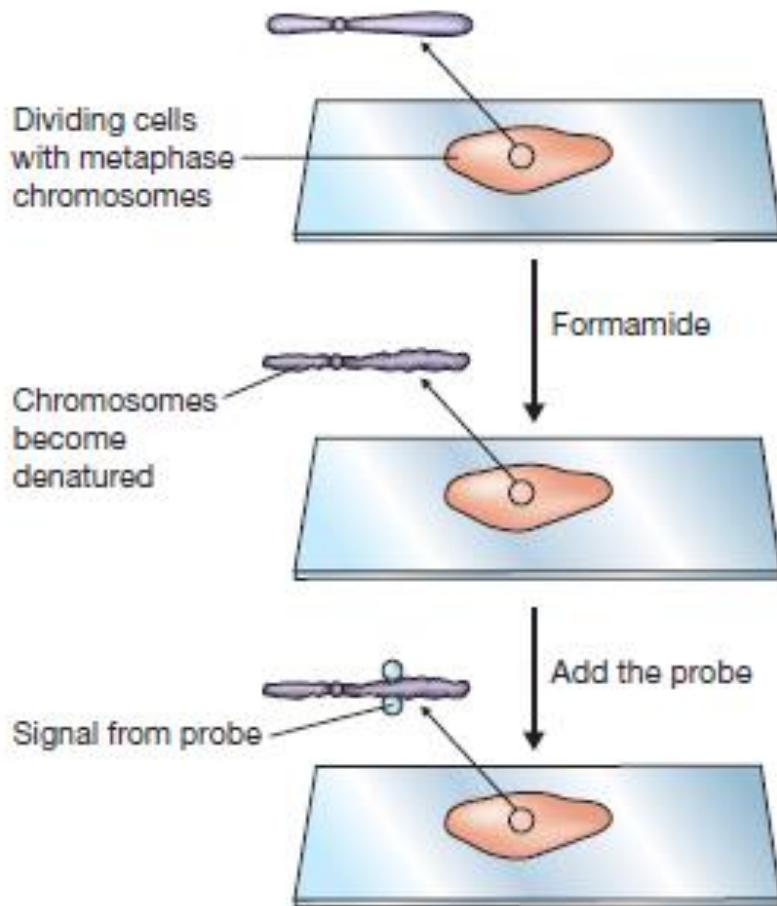


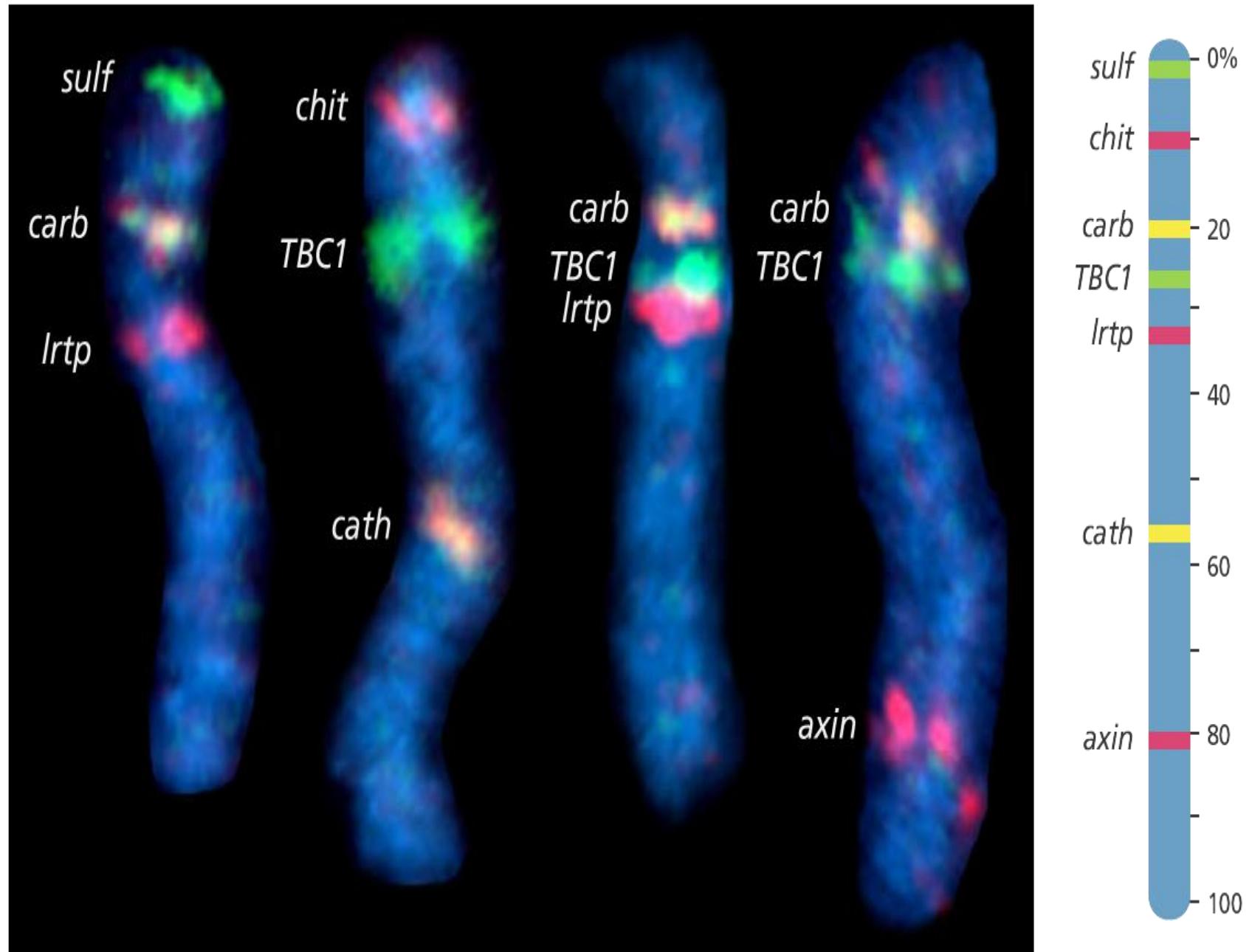
# *Physical maps*

A physical map is generated by methods that directly locate the positions of specific sequences on a chromosomal DNA molecule. As in genetic mapping, the loci that are studied can be genes or DNA markers. A **method used for physical mapping is:**

- **Fluorescence *in situ* hybridization (FISH)**

# FISH





# Library Preparation and Screening

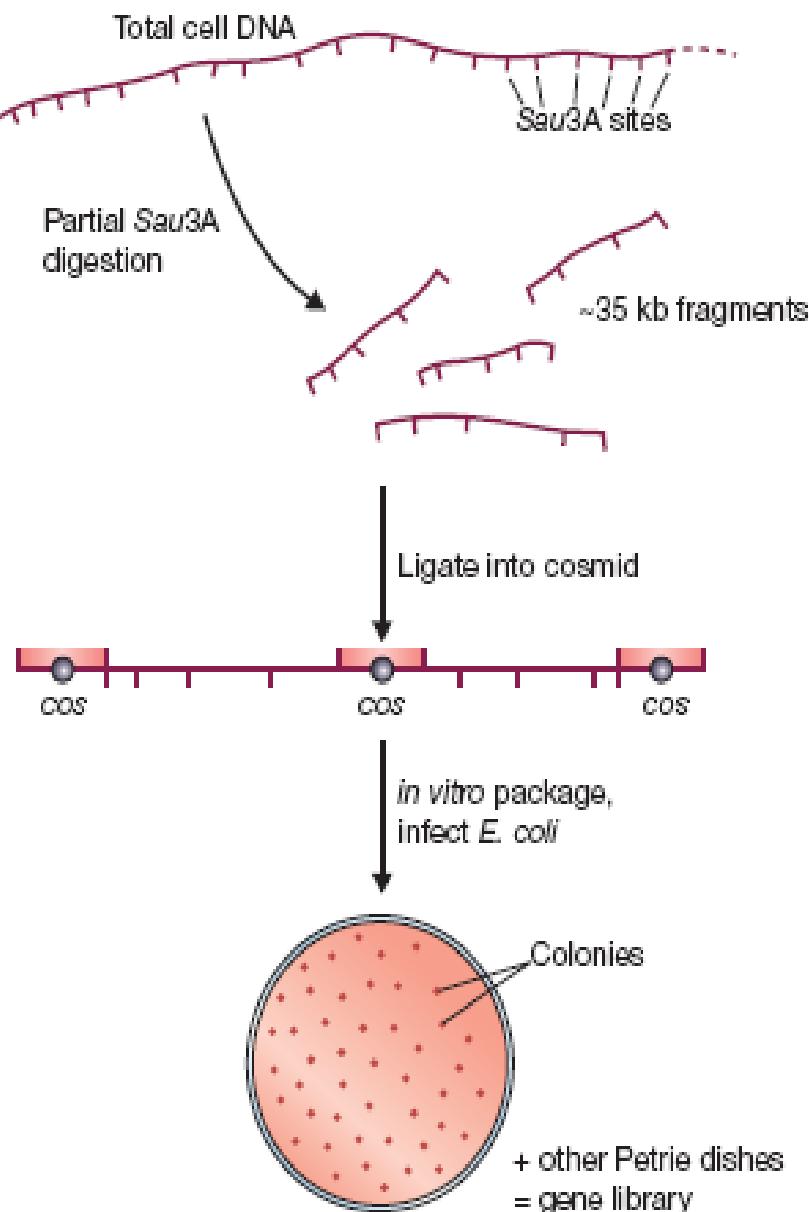
Chapter 8 (TA Brown)

# Outline

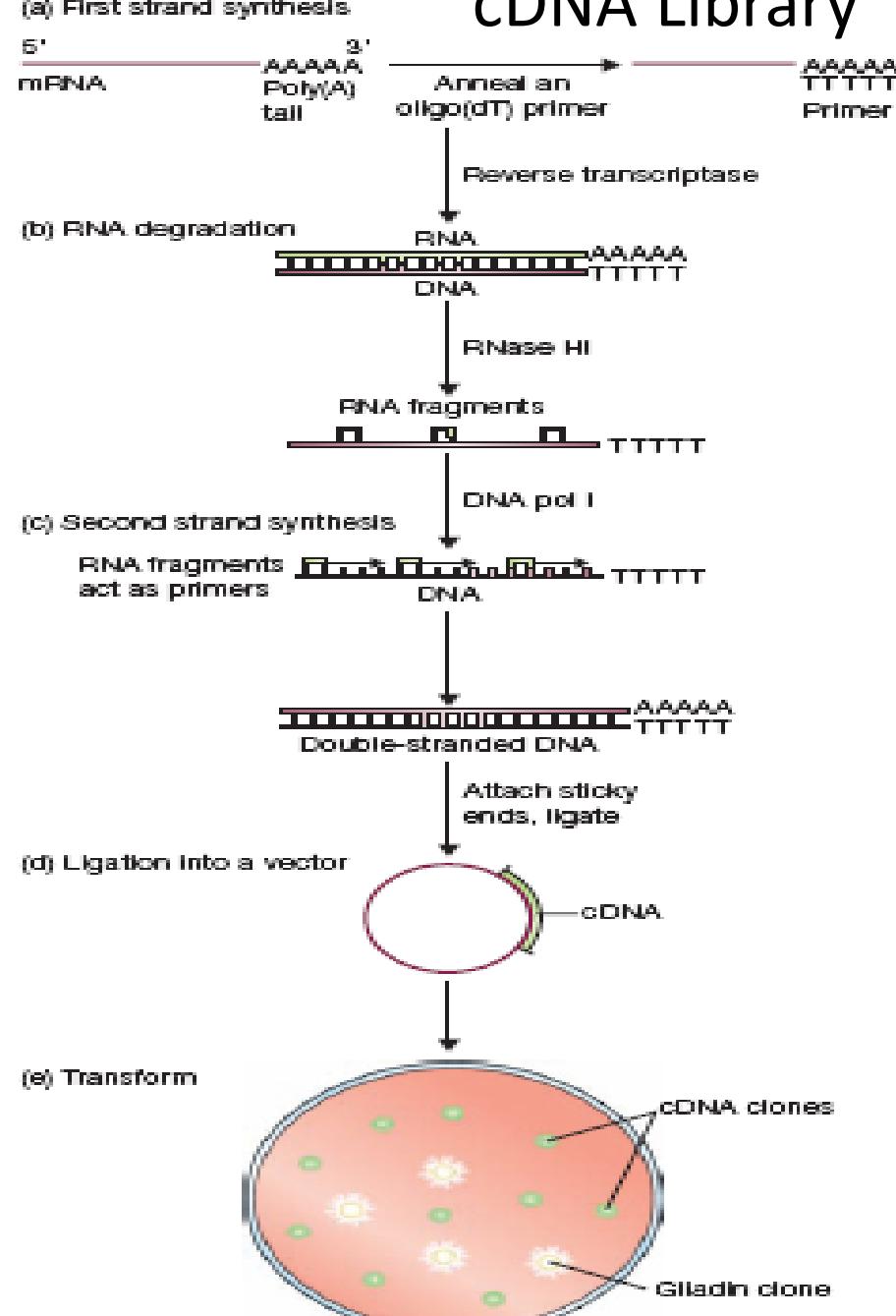
- Genomic and cDNA library preparation
- Screening for the desired clone
  - Direct Clone Selection
  - Nucleic Acid Hybridization
    - Probe Preparation
      - Labelling Methods
        - » Radioactive
        - » Non-radioactive
      - Oligonucleotide synthesis
    - Probe Sequence
      - Exact Nucleotide Match
      - Heterologous Probing
    - Southern Hybridization
  - Immunoscreening

Genomic Library	cDNA Library
<ul style="list-style-type: none"> <li>● It includes all possible fragments of DNA from a given cell or organism.</li> <li>● It is larger.</li> <li>● Represents the entire genome of an organism having both coding and non coding regions.</li> <li>● Expression of genes taken from genomic library is difficult in prokaryotic system like bacteria due to absence of splicing mechanism.</li> <li>● Vectors used in genomic library include plasmid, cosmid, lambda phage, BAC and YAC in order to accommodate large fragments.</li> </ul>	<ul style="list-style-type: none"> <li>● Carries only expressed gene sequences.</li> <li>● It is smaller.</li> <li>● Represents only the expressed part of the genome and contain only coding sequences called ESTs.</li> <li>● cDNA has only coding sequences therefore can be directly expressed in prokaryotic system.</li> <li>● Vectors used in cDNA library include plasmid, phagemid, lambda phage etc. to accommodate small fragments as cDNA has no introns.</li> </ul>

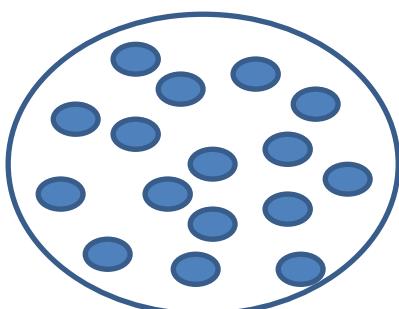
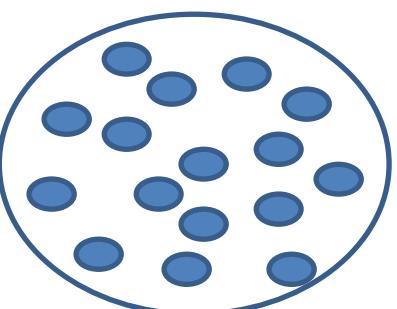
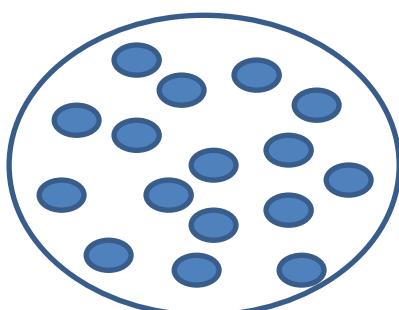
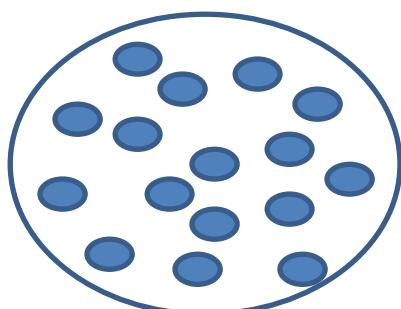
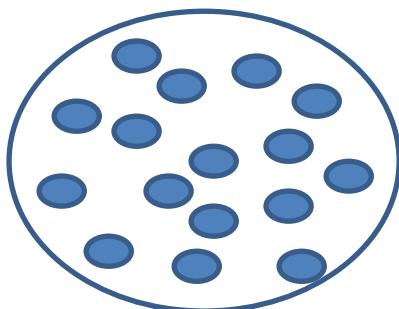
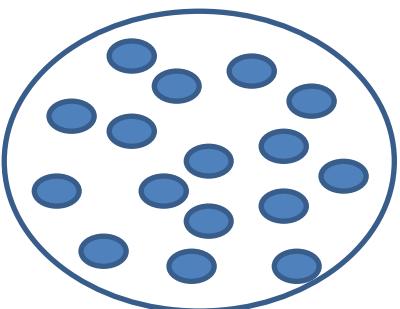
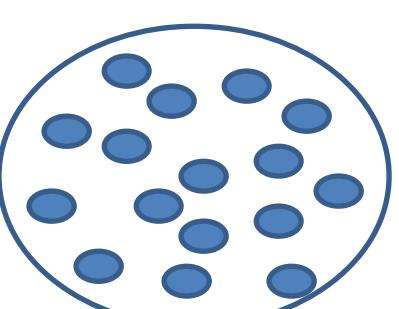
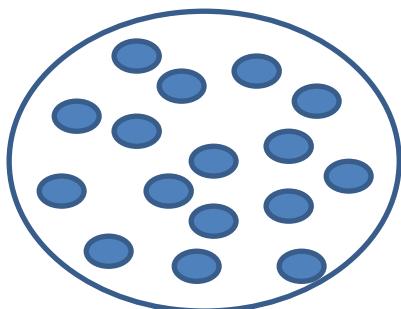
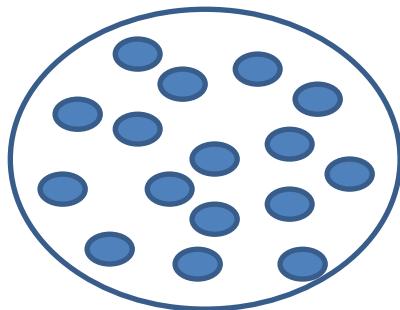
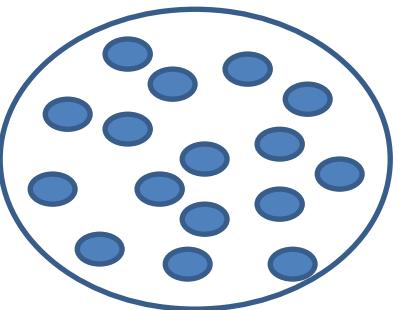
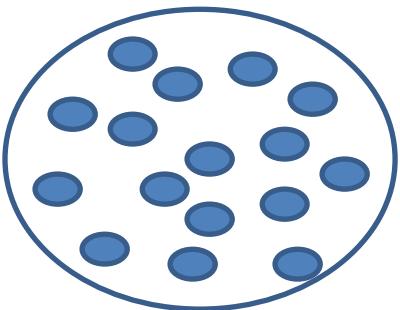
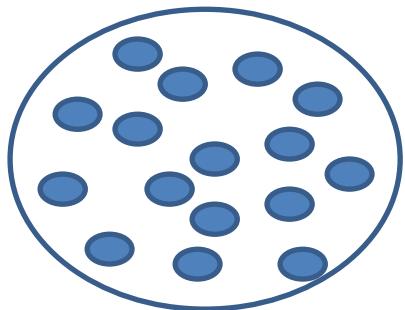
# Genomic Library



# cDNA Library

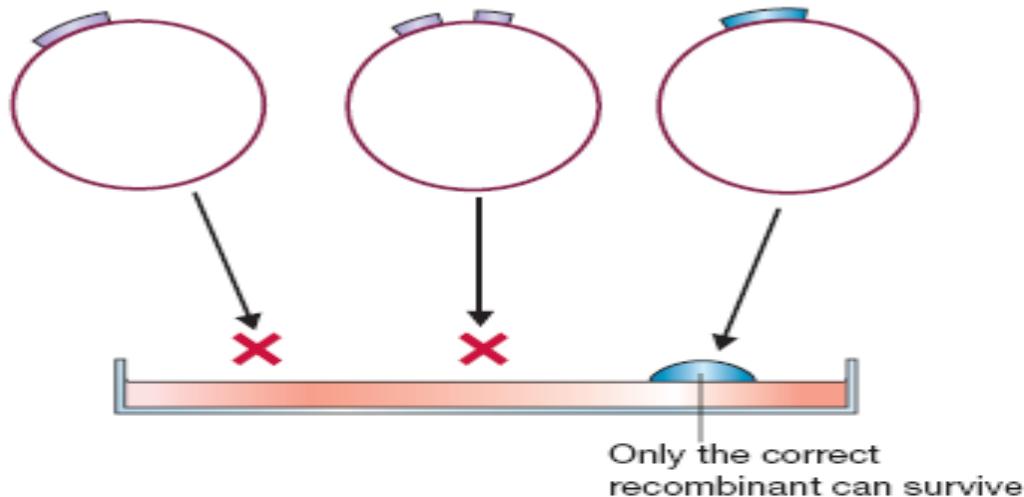


# Clone Identification

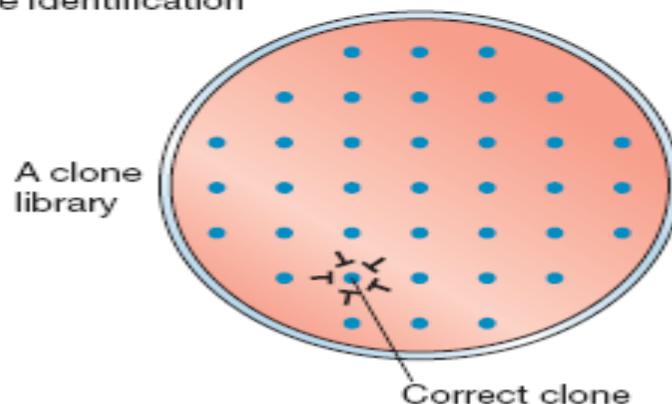


# Clone Identification

(a) Direct selection



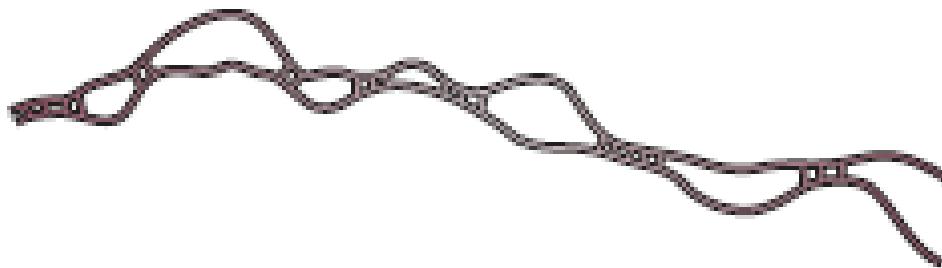
(b) Clone identification



# Methods of Clone Identification

## Nucleic Acid Hybridization

(a) An unstable hybrid



(b) A stable hybrid



(c) A DNA-RNA hybrid

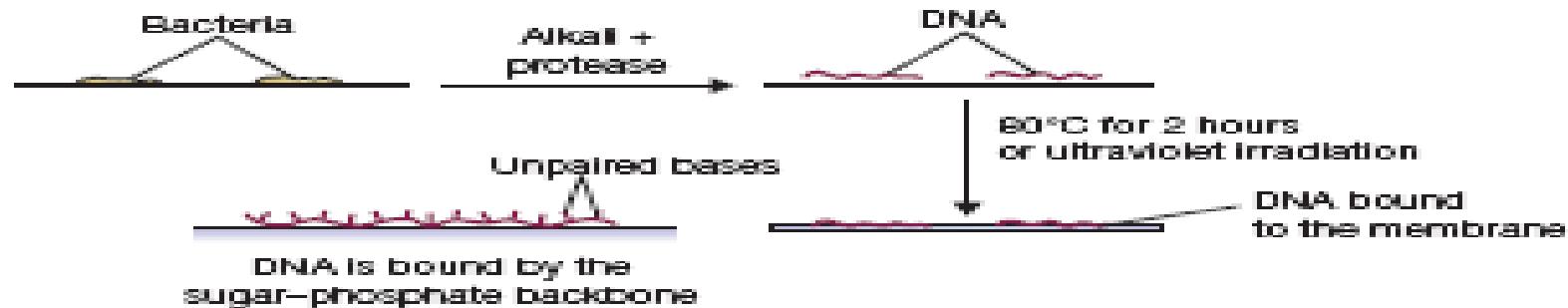


# Colony Hybridization

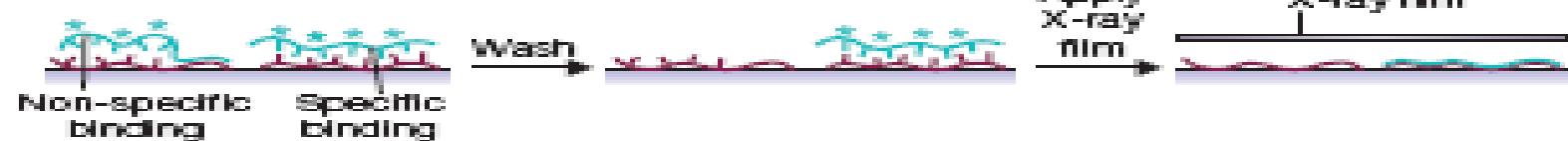
(a) Transfer colonies to nitrocellulose or nylon



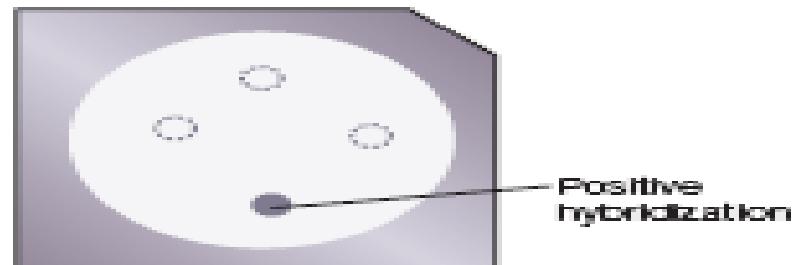
(b) Degradate cells, purify DNA



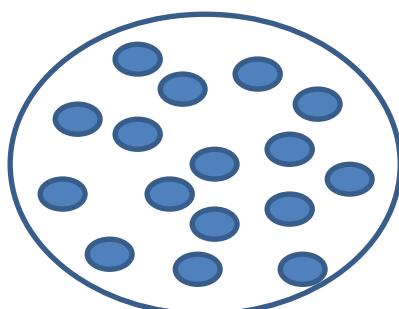
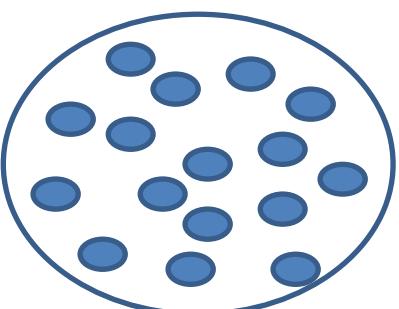
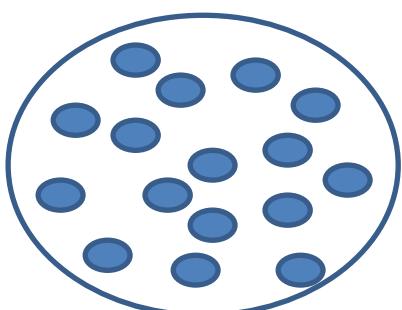
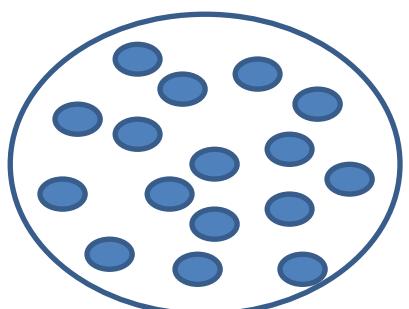
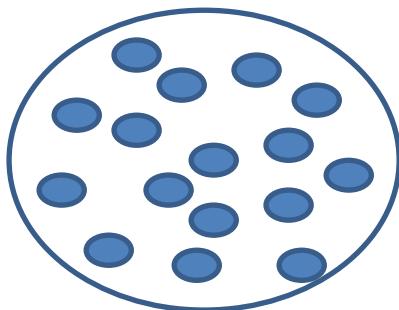
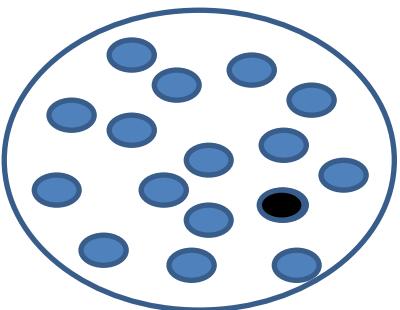
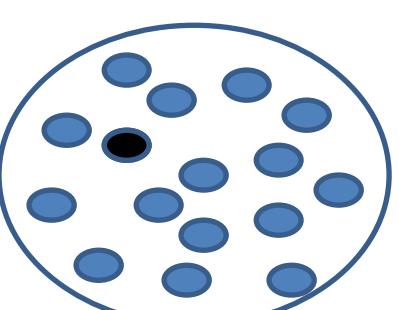
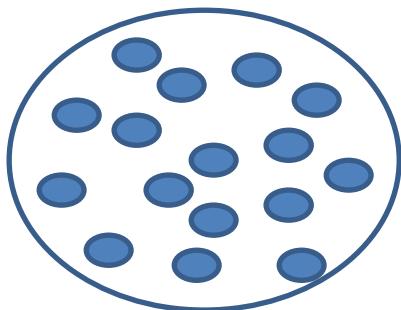
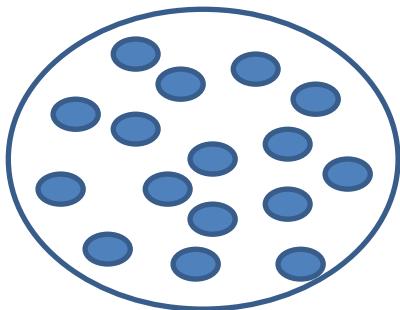
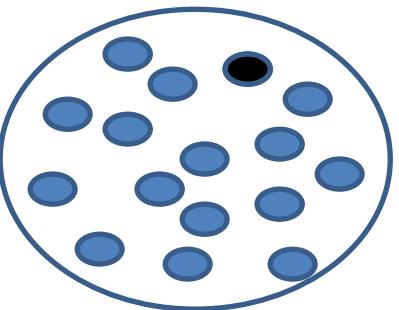
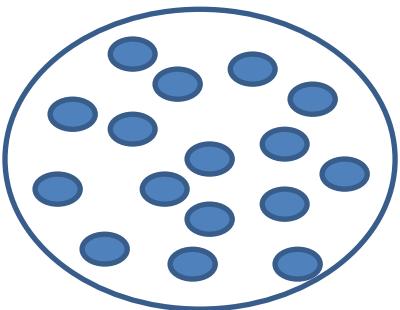
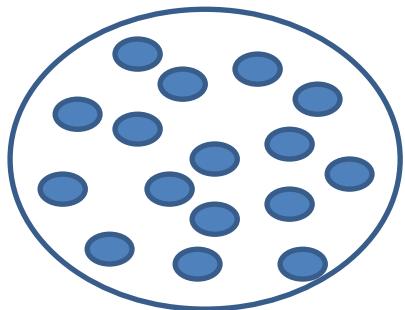
(c) Probe with labelled DNA



(d) The resulting autoradiograph



# Clone Identification

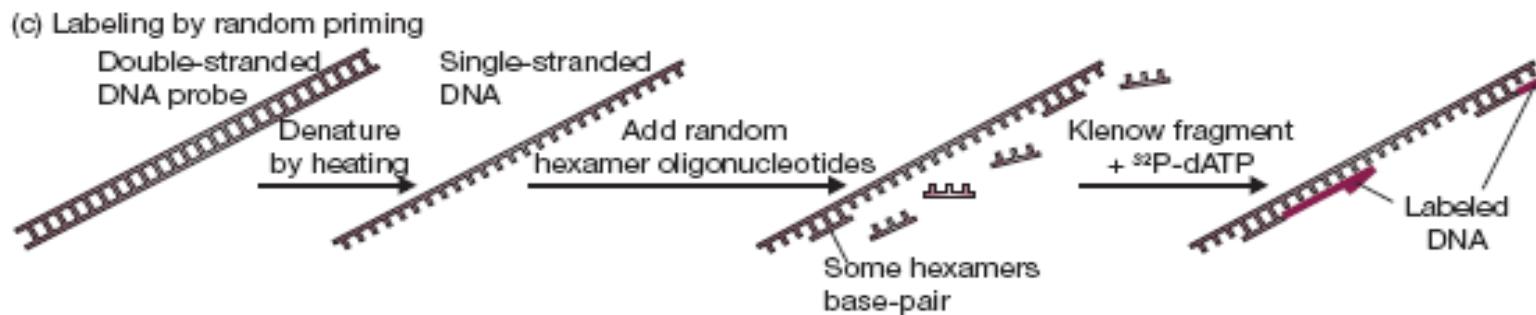
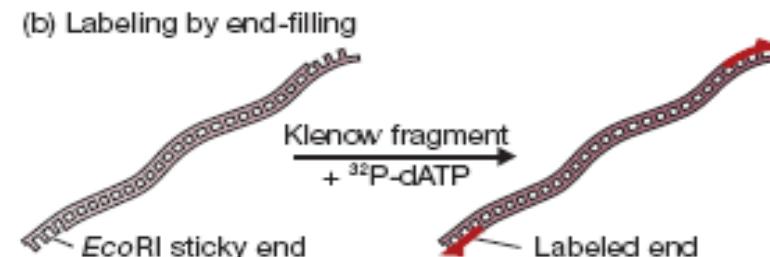
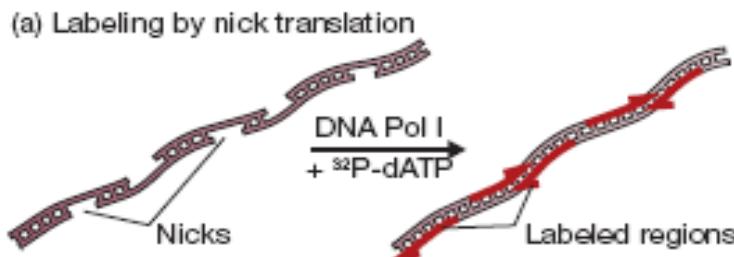
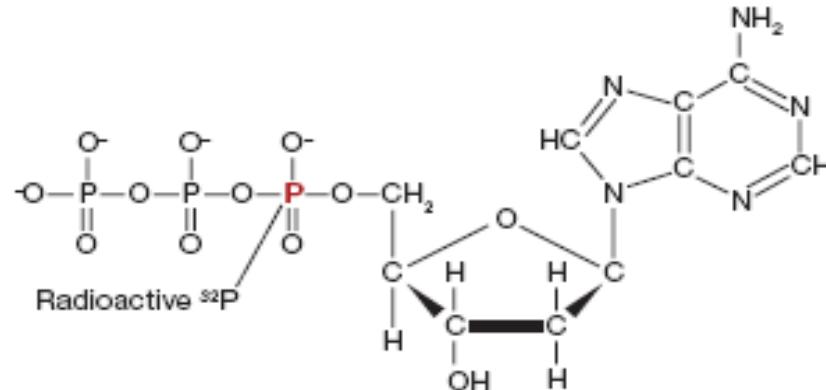


# Methods for DNA Labelling

## Radioactive Labeling

Figure 8.10

The structure of  $\alpha$ - $^{32}\text{P}$ -deoxyadenosine triphosphate ( $[\alpha\text{-}^{32}\text{P}]$ dATP).



# Handling Radioactive Material



# Methods for DNA Labeling

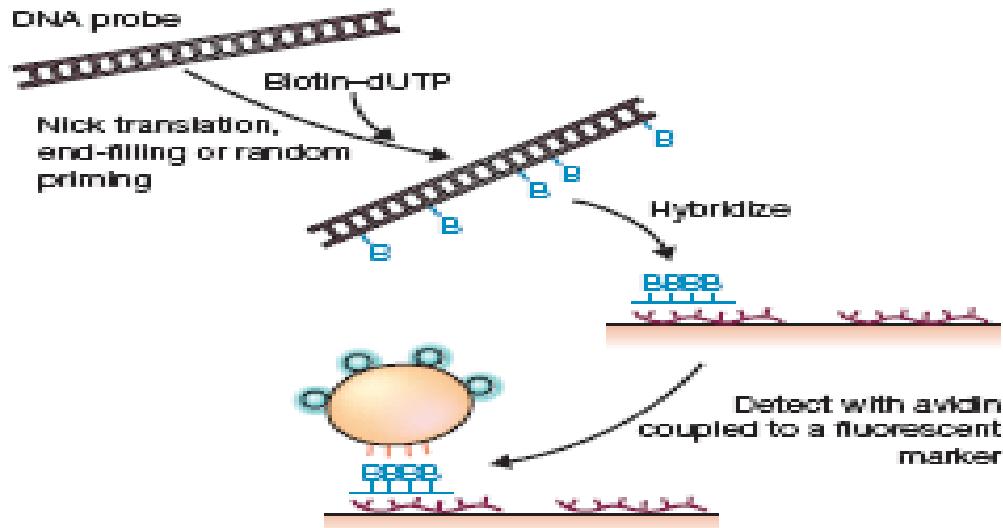
**Figure 8.12**

Two methods for the non-radioactive labelling of DNA probes.

## Non-Radioactive Labeling

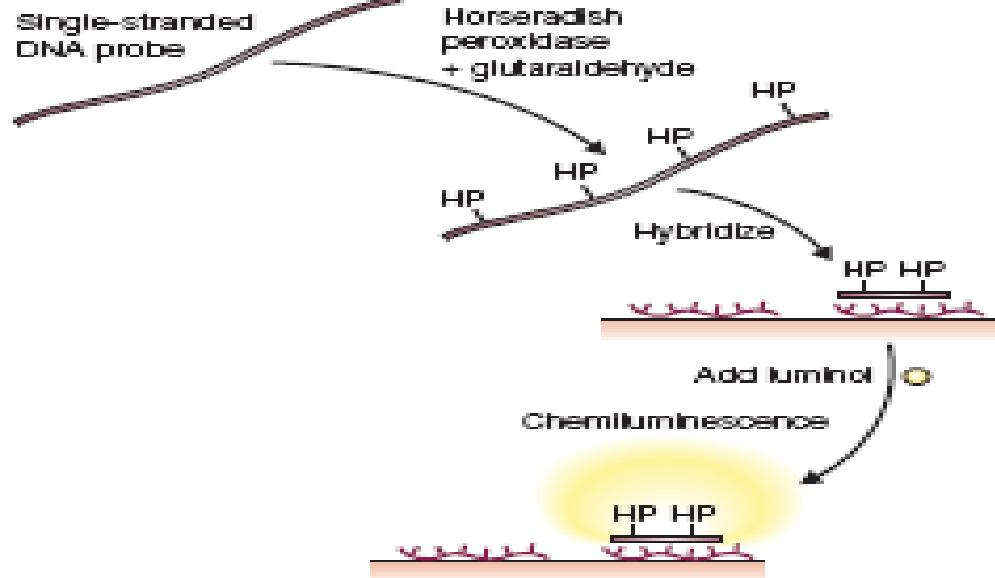
(a) Labelling with a biotinylated nucleotide

DNA probe



(b) Labelling with horseradish peroxidase

Single-stranded  
DNA probe



# Probe Sequence

- Exact Complementary Sequence
- Extract it from Protein Sequence
- Partial Complementary Sequence  
(Heterologous Probing)

# Probe- Cytochrome C

GLY-SER-ALA- LYS- LYS-GLY-ALA-THR- LEU- PHE- LYS- THR-ARG-CYS-GLU-  
LEU-CYS- HIS- THR-VAL-GLU-LYS- GLY-GLY-PRO- HIS- LYS- VAL- GLY-PRO-  
ASN-LEU- HIS- GLY- ILE- PHE-GLY-ARG- HIS- SER- GLY-GLN- ALA- GLN-GLY-  
TYR-SER-TYR-THR-ASP-ALA-ASN- ILE- LYS- LYS- ASN-VAL- LEU- **TRP-ASP-**  
**GLU-ASN-ASN-MET**-SER-GLU-TYR-LEU-THR-ASN-PRO-LYS- LYS- TYR- ILE-  
PRO-GLY-THR- LYS- MET-ALA-PHE-GLY-GLY- LEU- LYS- LYS- GLU- LYS- ASP-  
ARG-ASN-ASP-LEU- ILE- THR-TYR-LEU- LYS- LYS- ALA-CYS-GLU

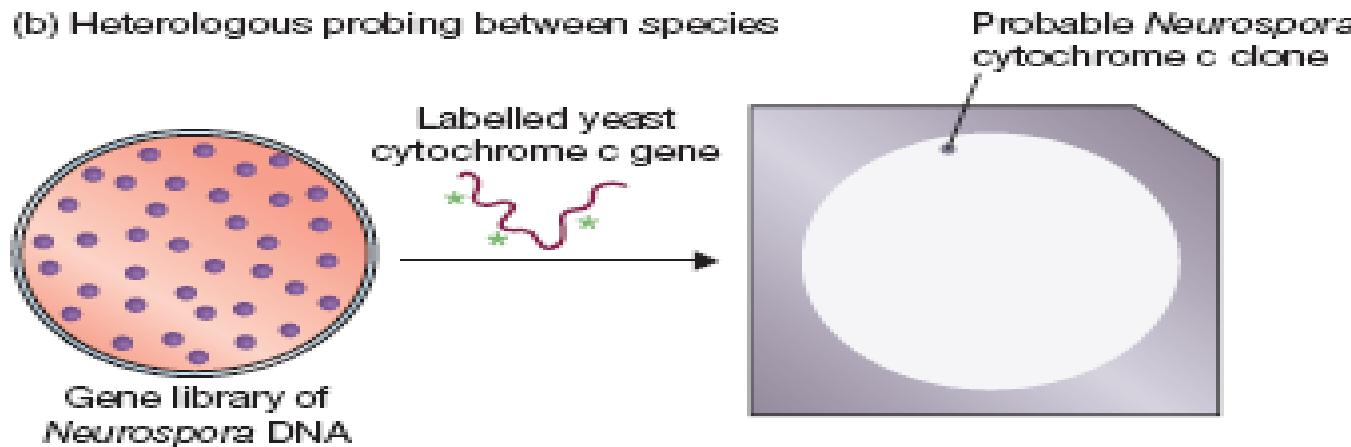
The cytochrome c protein from yeast was sequenced in 1963, with the result shown in Figure 8.14. This sequence contains a segment, starting at amino acid 59, that runs Trp-Asp-Glu- Asn-Asn-Met. The genetic code states that this hexapeptide is coded by TGG-GAT-GAA-AA<sub>C</sub><sup>T</sup>-AA<sub>C</sub><sup>T</sup>-ATG. Although this represents a total of 16 different possible sequences, 14 of the 18 nucleotides can be predicted with certainty.

# Heterologous Probing

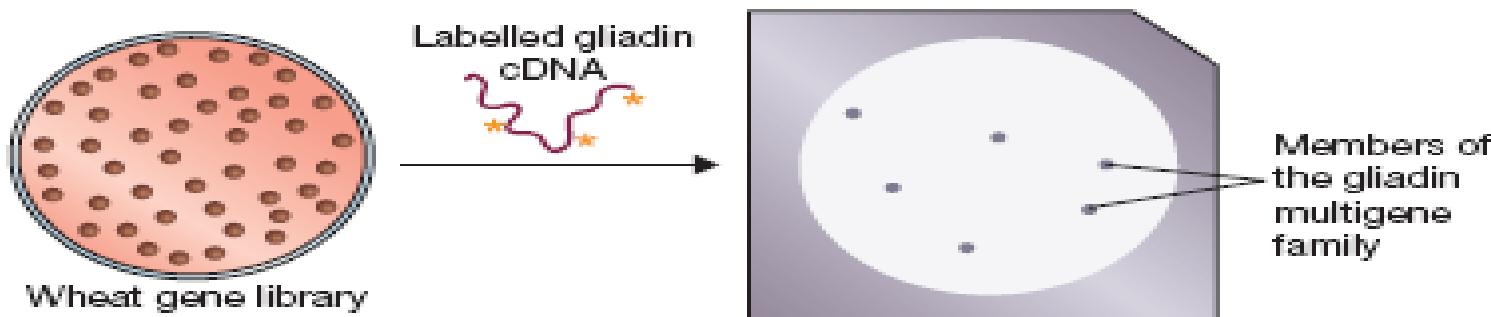
(a) A hybrid between two related DNA strands



(b) Heterologous probing between species



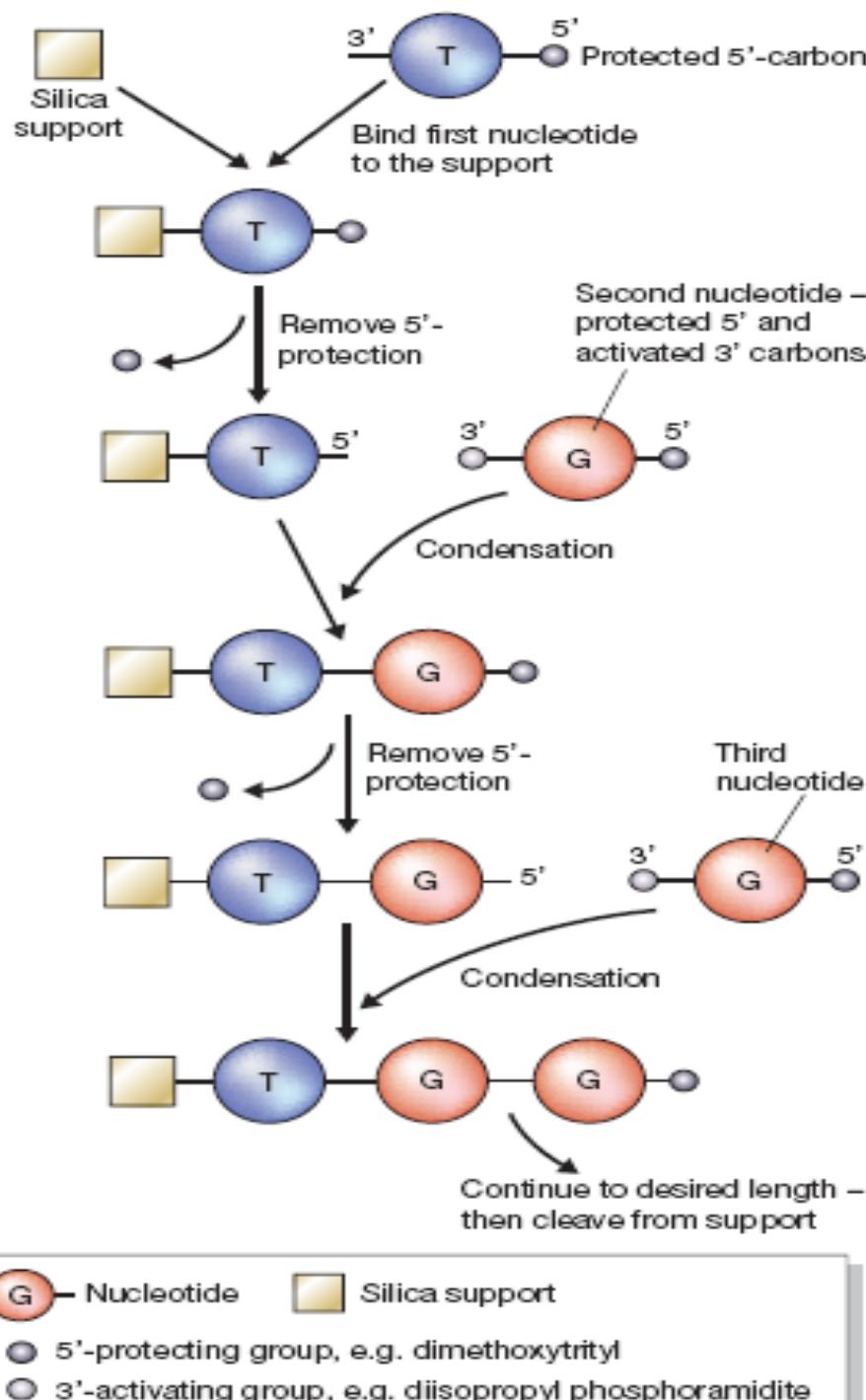
(c) Heterologous probing within a species



**Figure 8.15**

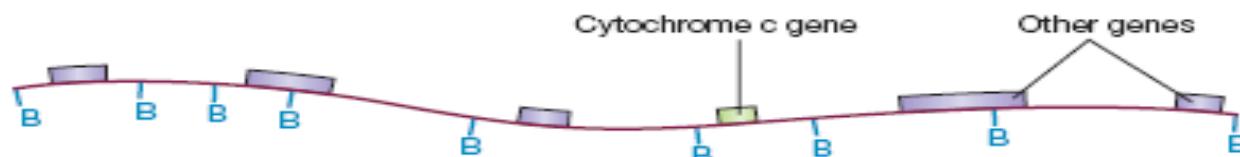
A simplified scheme for oligonucleotide synthesis. Each nucleotide is modified by attachment of an activating group to the 3' carbon and a protecting group to the 5' carbon. The activating group enables the normally inefficient process of nucleotide joining to proceed much more rapidly. The protecting group ensures that individual nucleotides cannot attach to one another, and instead react only with the terminal 5' group of the growing oligonucleotide, this 5' group being deprotected by chemical treatment at the appropriate point in each synthesis cycle.

## Oligonucleotide Synthesis



# Southern Hybridization

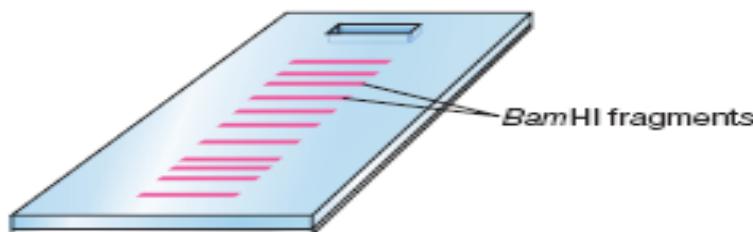
*Southern hybridization enables a specific restriction fragment containing a gene to be identified*



**Figure 8.18**

A long cloned DNA fragment may contain several genes in addition to the one in which we are interested. B = *Bam*HI restriction site.

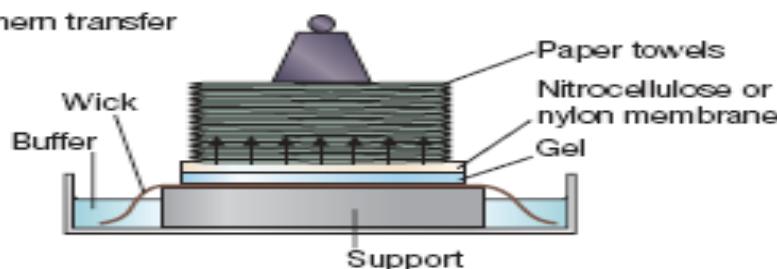
(a) Electrophoresis *Bam*HI-restricted DNA



**Figure 8.19**

Southern hybridization.

(b) Southern transfer



(c) Result of hybridization probing



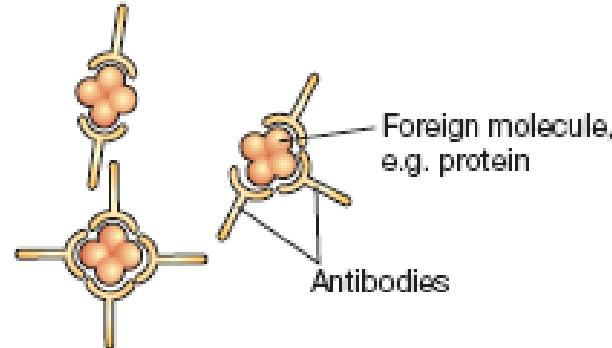
# Immunoscreening

## Detection of Translated Products

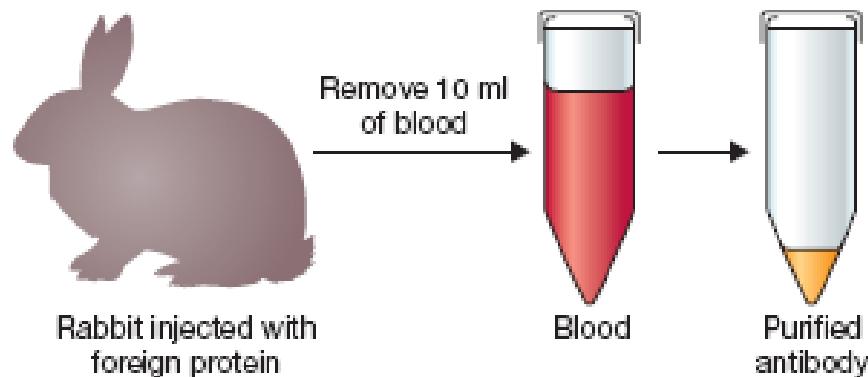
Figure 8.20

Antibodies. (a) Antibodies in the bloodstream bind to foreign molecules and help degrade them.  
(b) Purified antibodies can be obtained from a small volume of blood taken from a rabbit injected with the foreign protein.

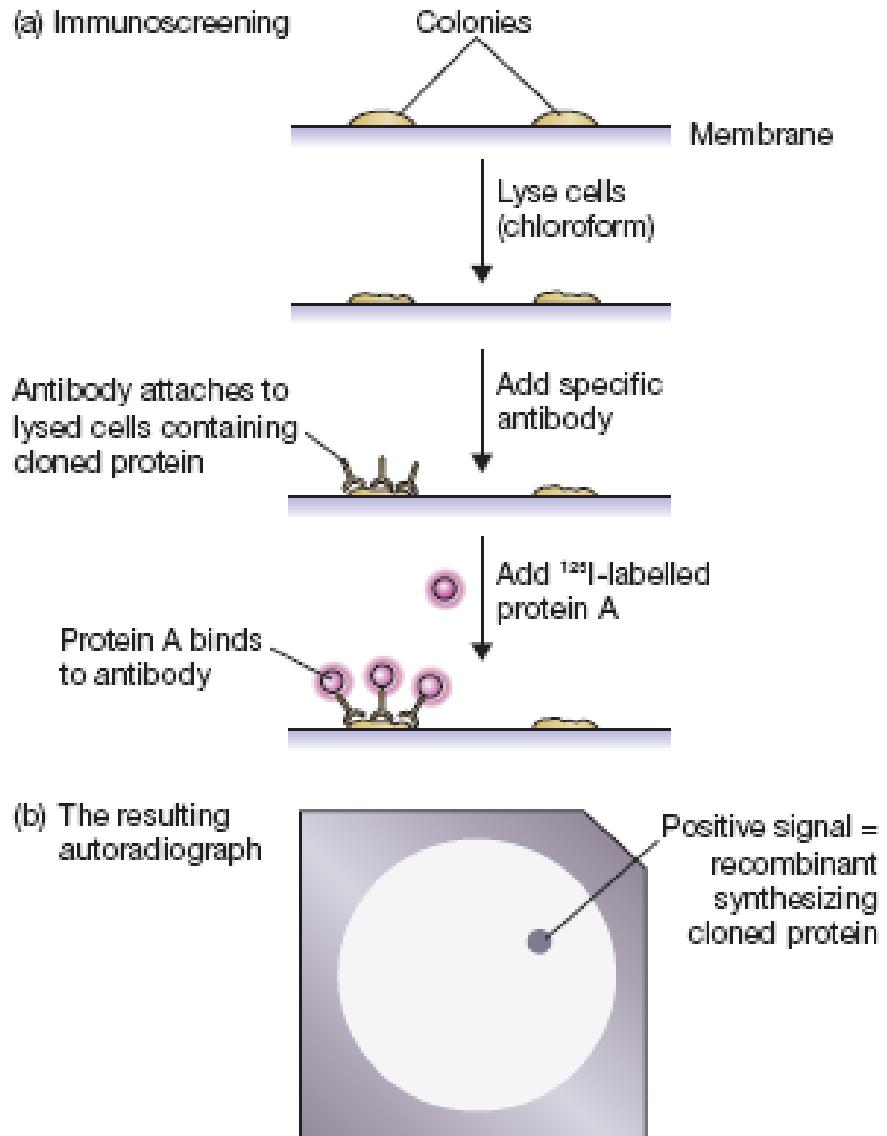
(a) Antibodies bind to foreign molecules



(b) Antibody purification



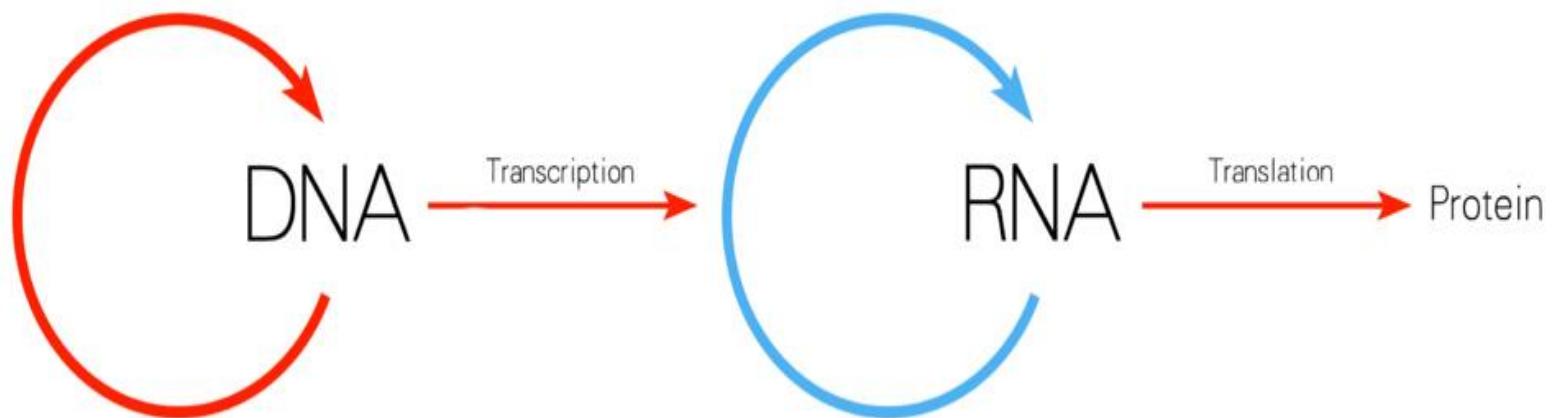
# Immunoscreening



**Figure 8.21**

Using a purified antibody to detect protein in recombinant colonies. Instead of labelled protein A, the antibody itself can be labelled, or alternatively a second labelled antibody which binds specifically to the primary antibody can be used.

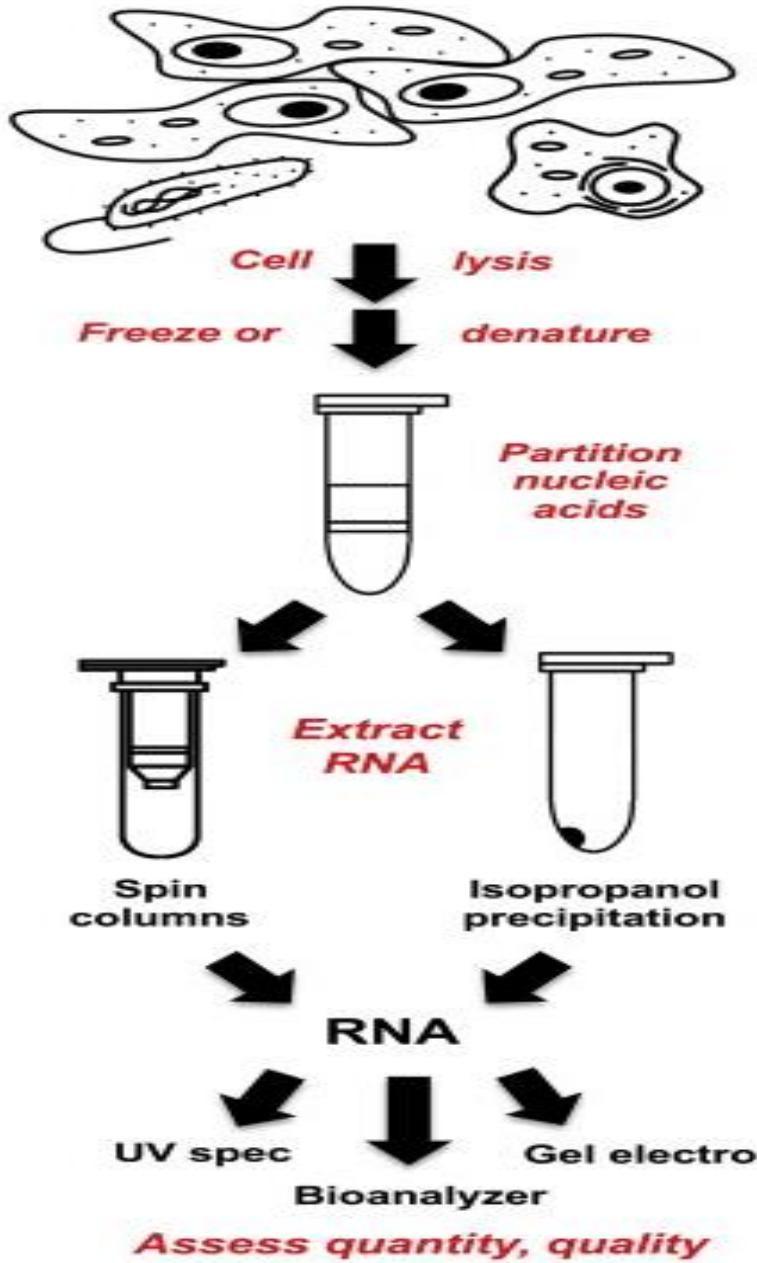
# Gene Expression Studies



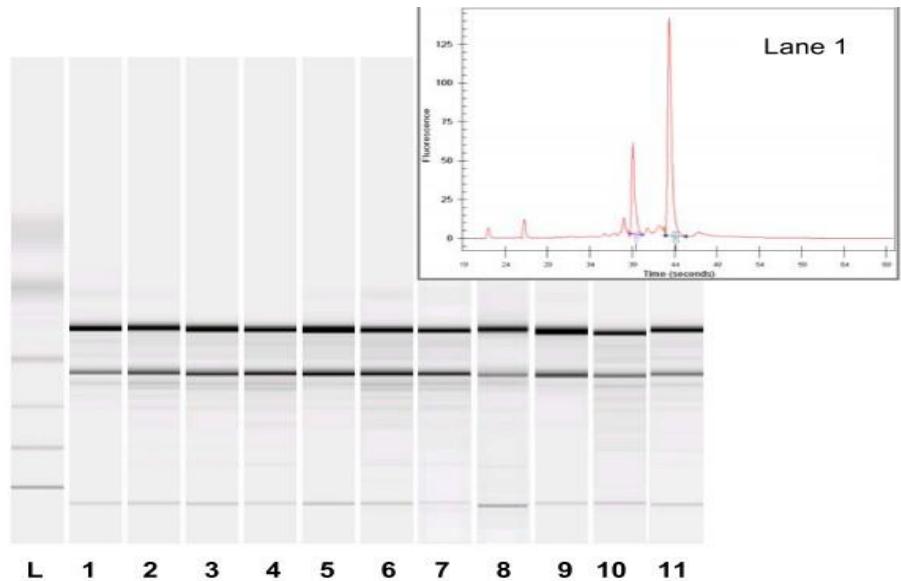
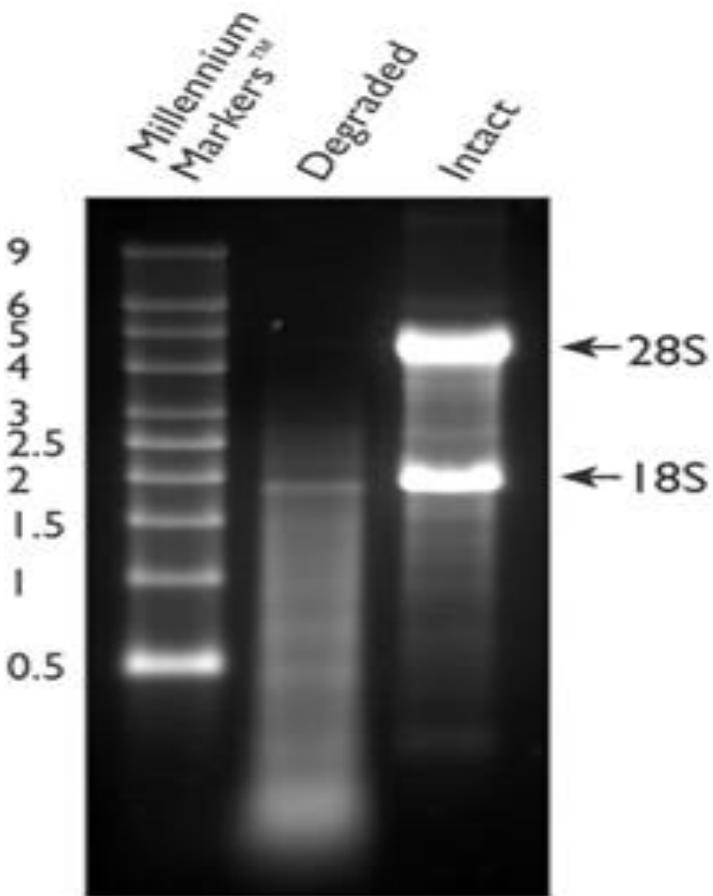
# Lecture Outline

- Detect Transcript, measure transcript levels
- Detect alternatively spliced transcripts
- Transcriptome studies
- Transcript Mapping
  - S1 Nuclease transcript mapping
  - Primer Extension
  - RACE (Rapid Amplification of Transcript Ends)
- Promoter Studies
  - Promoter Luciferase Assays
  - Transcription Factor Binding Sites
  - Electrophoretic Mobility Shift Assay
  - DNasel footprinting Assay

# Method of RNA isolation



# RNA Gel

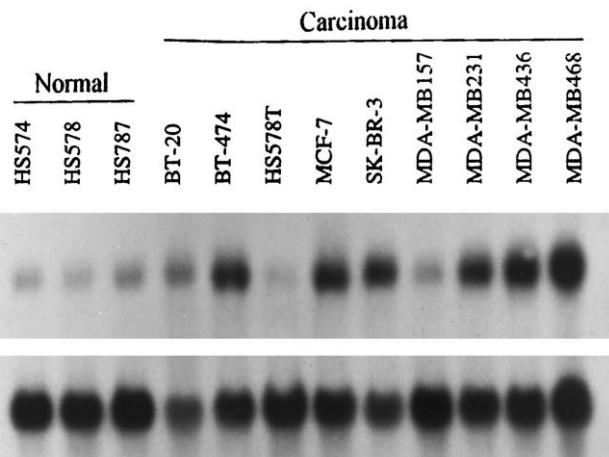
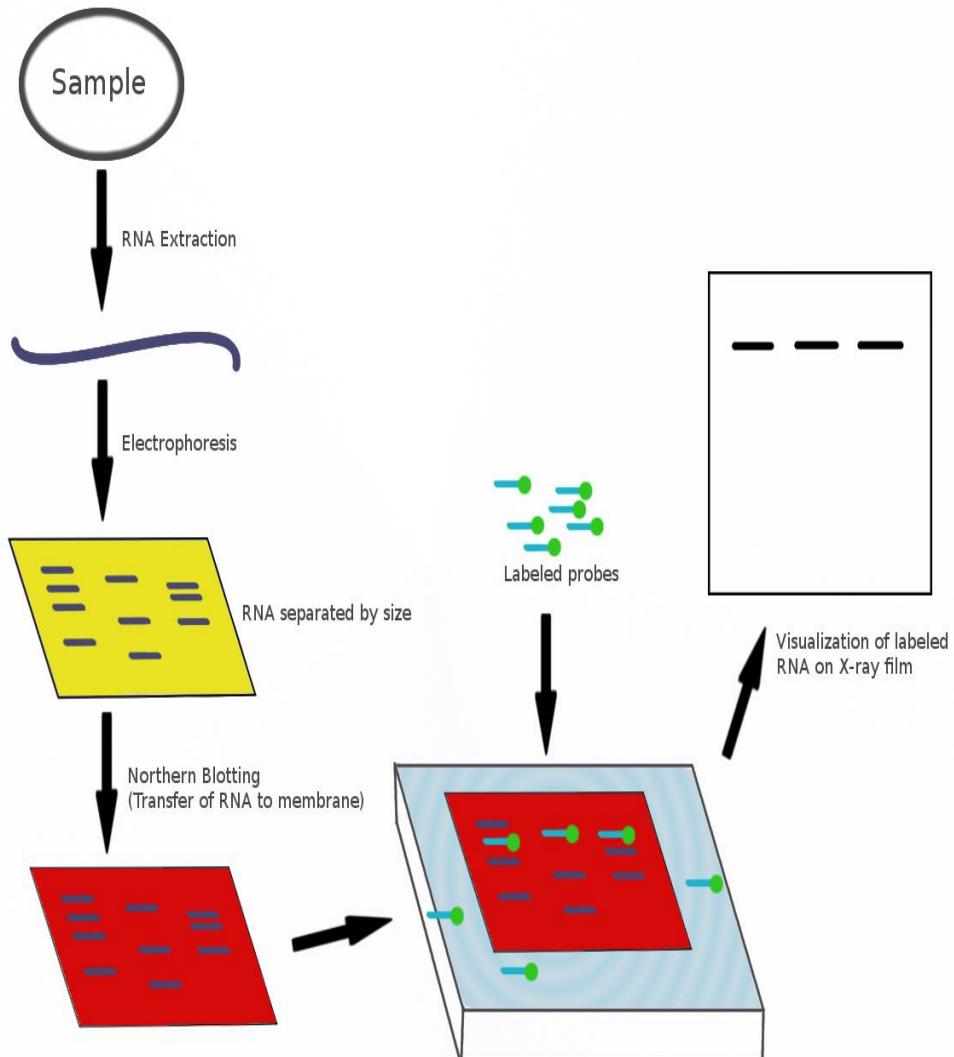


The **Bioanalyzer** is a chip-based capillary electrophoresis machine to analyse RNA, DNA, and protein. It is produced by Agilent and widely used, among other things, in RNA quality control measurements before downstream experiments like microarrays.

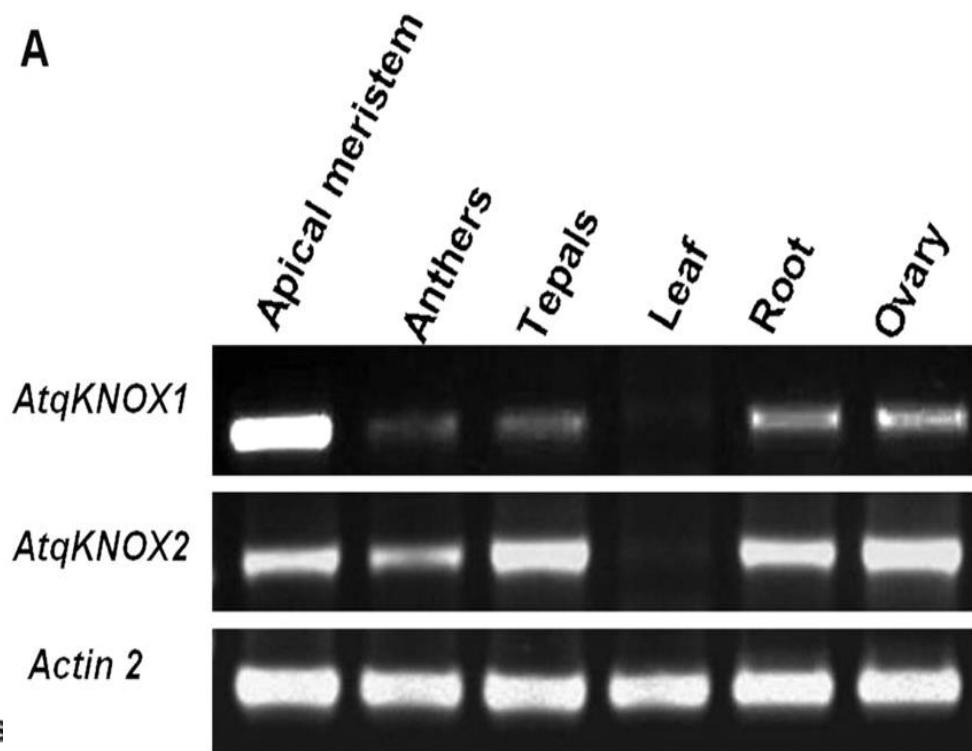
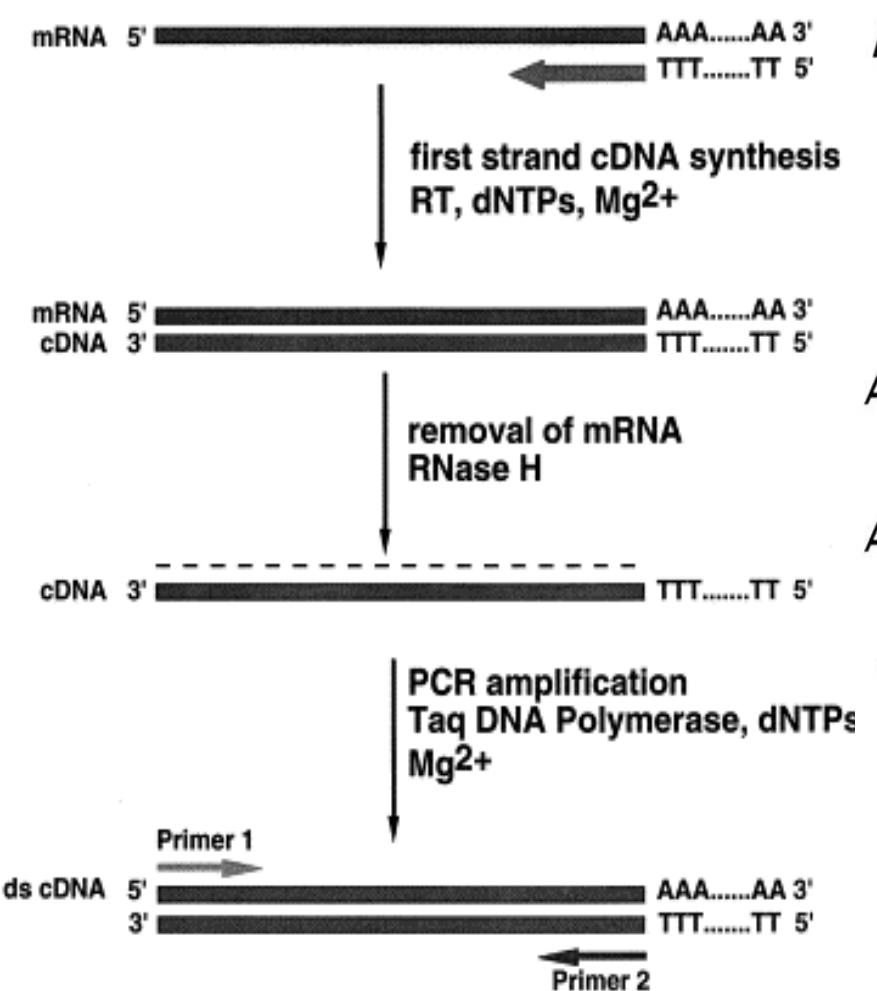
Note: in prokaryotes u will have 23S and 16S bands)

# Northern Blotting-

## A technique to study transcript levels

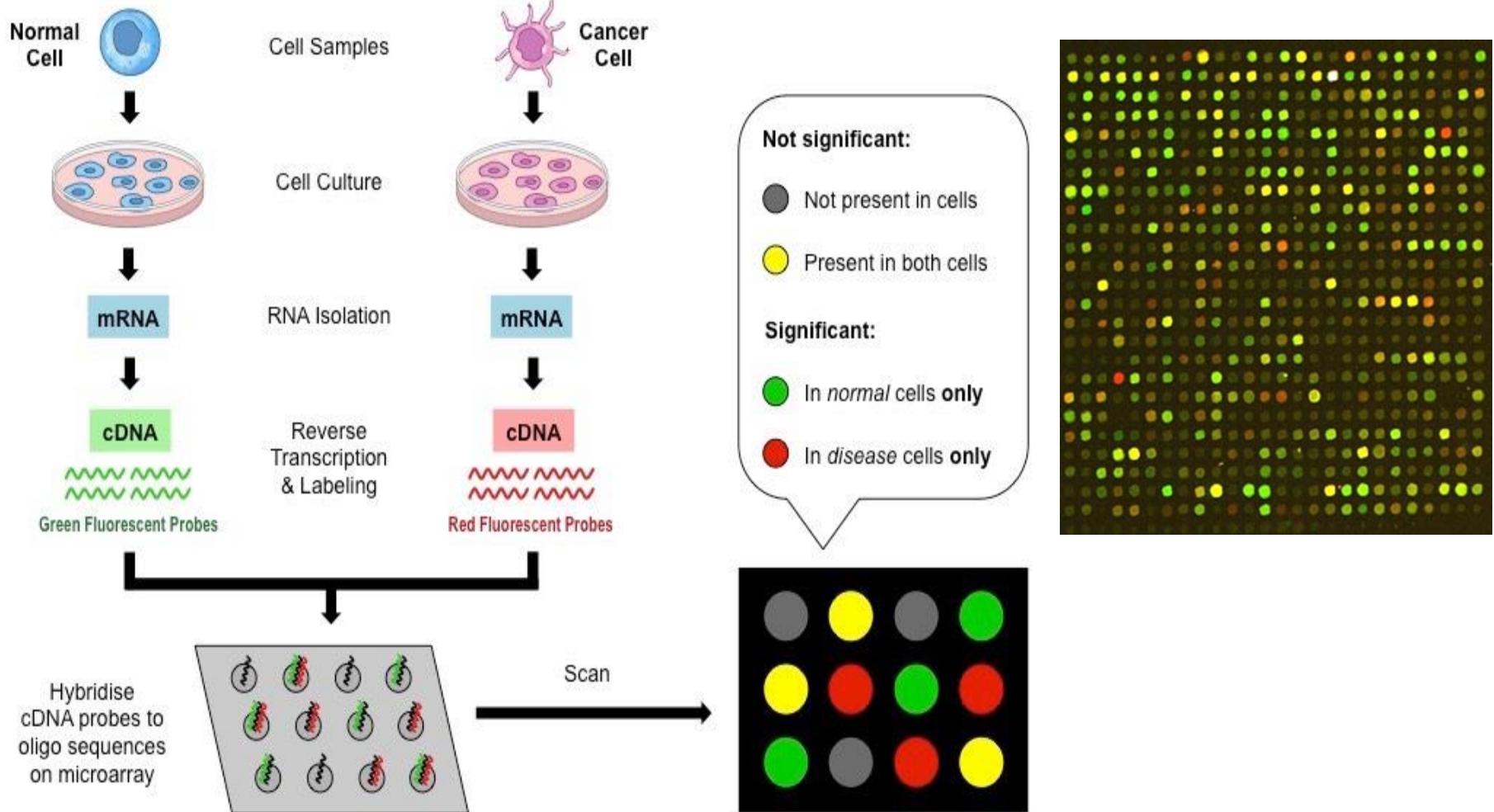


# Reverse-Transcriptase PCR (RT-PCR)



# Microarray

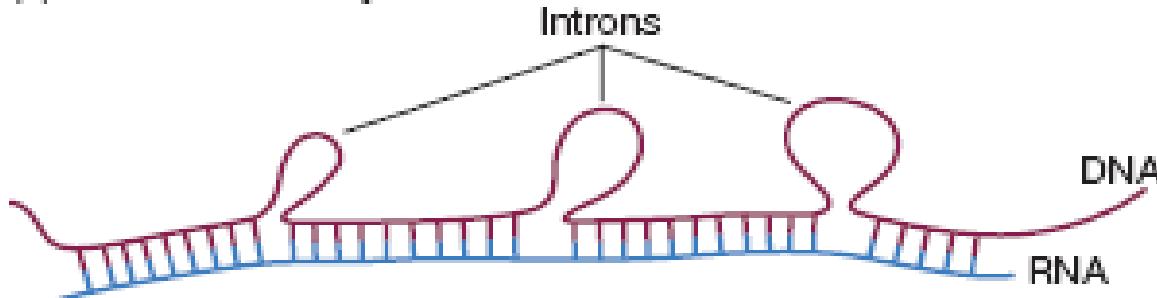
A **microarray** is a multiplex lab-on-a-chip. It is a two-dimensional array on a solid substrate—usually a glass slide or silicon thin-film cell—that assays (tests) large amounts of biological material using high-throughput screening miniaturized, multiplexed and parallel processing and detection methods.



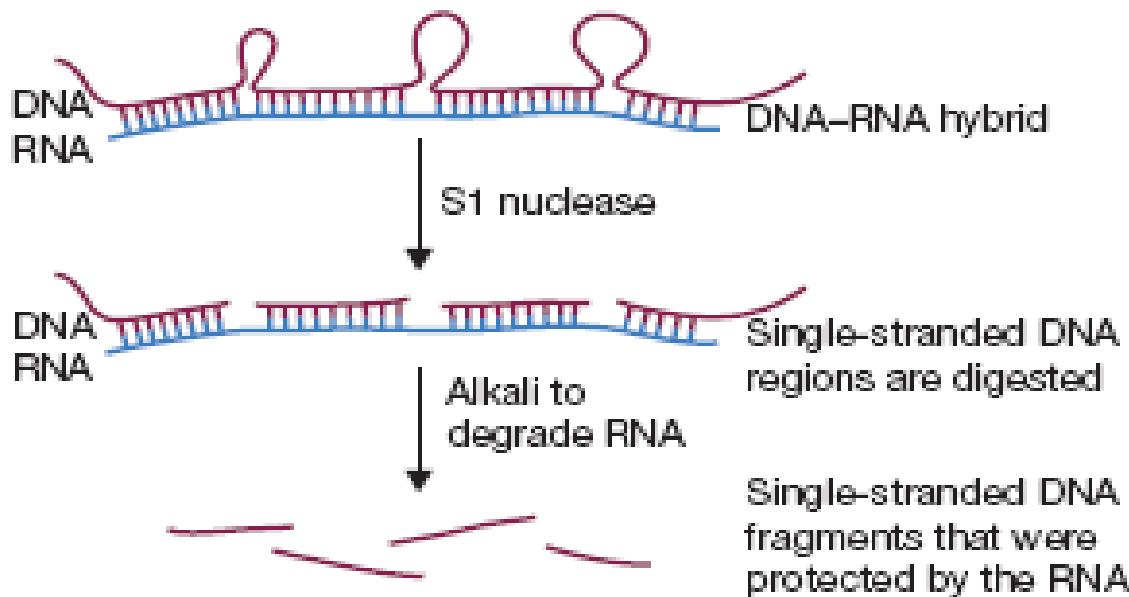
# Transcript Characterization

# Mapping Intron Exon Boundary

(a) A DNA-mRNA hybrid

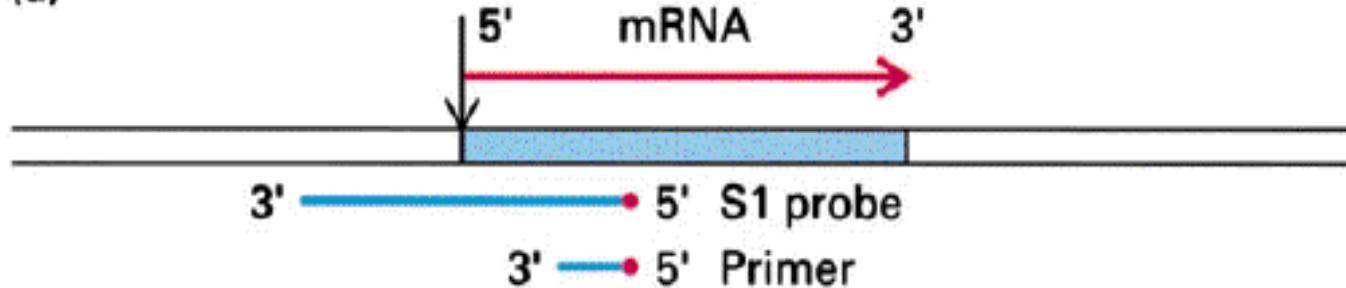


(b) Treatment of the hybrid with S1 nuclease

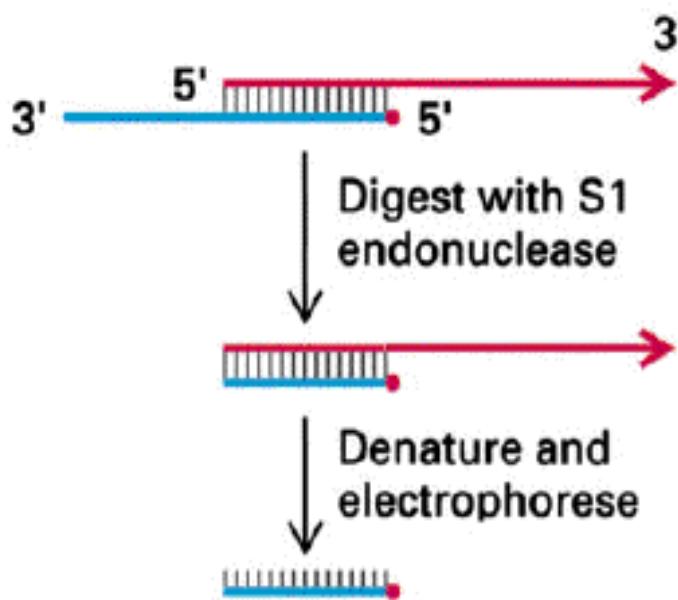


# Location of Transcription start site by S1 Nuclease Transcript Mapping

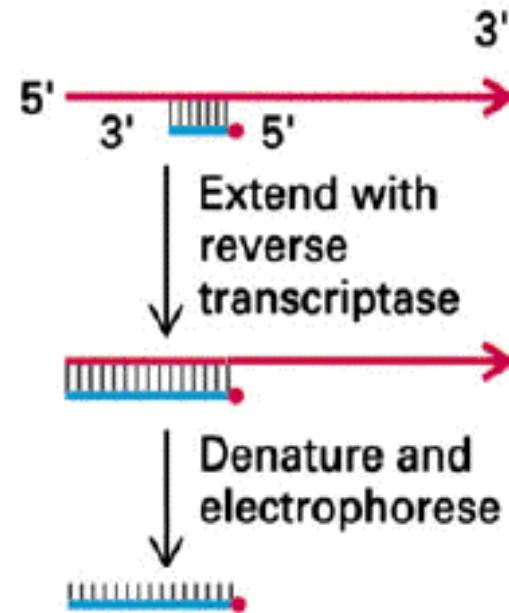
(a)



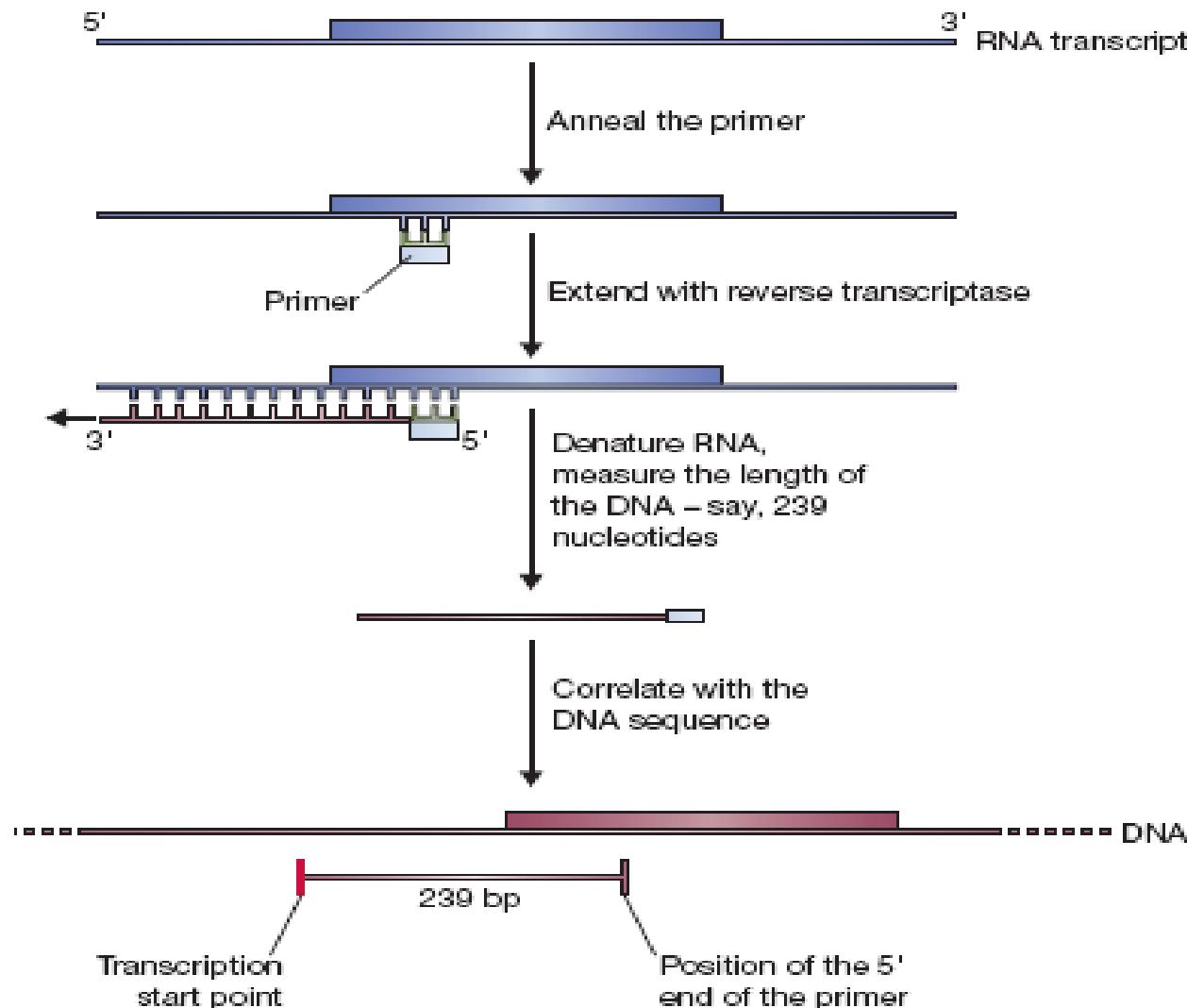
(b) S1 Nuclease mapping



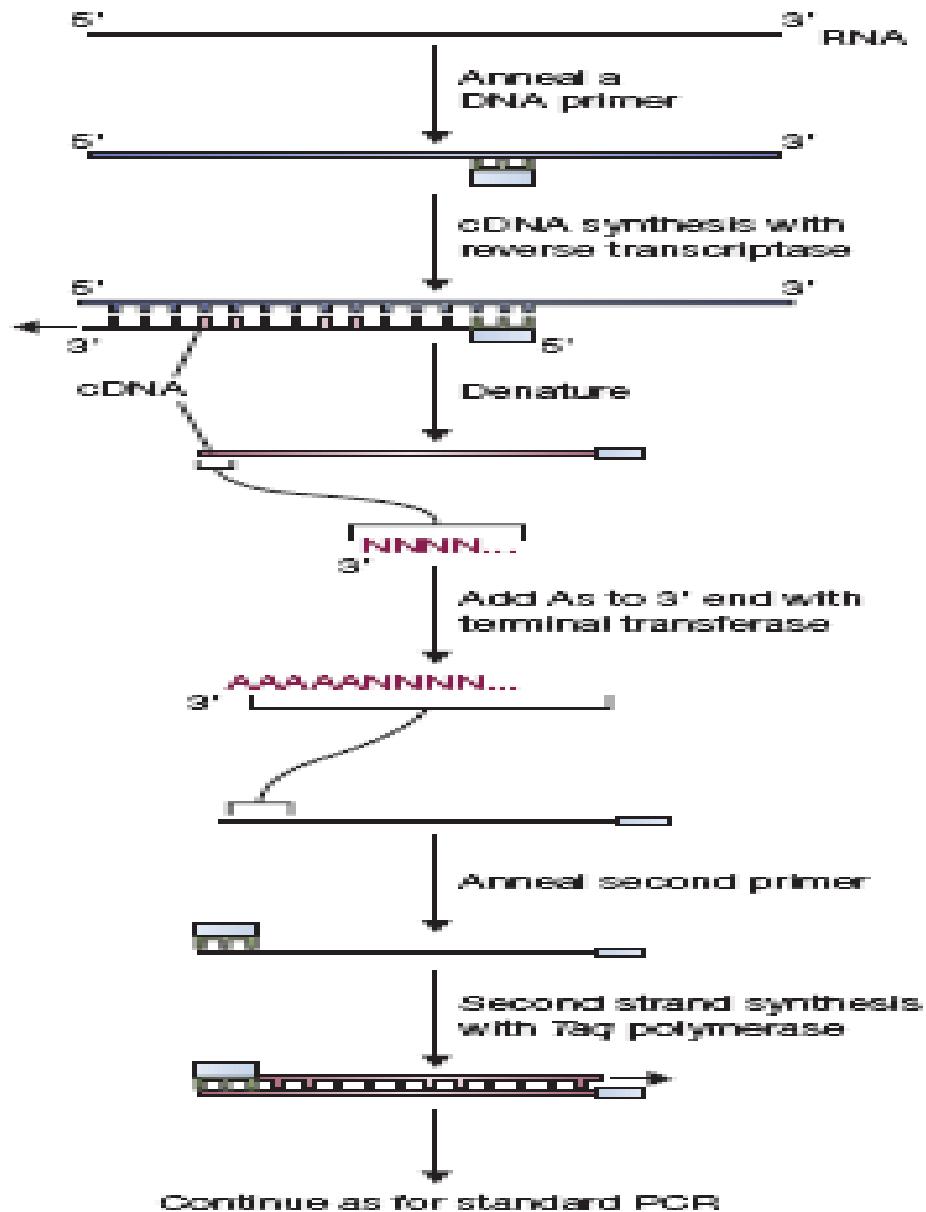
(c) Primer Extension



# Location of transcription start site by primer extension assay



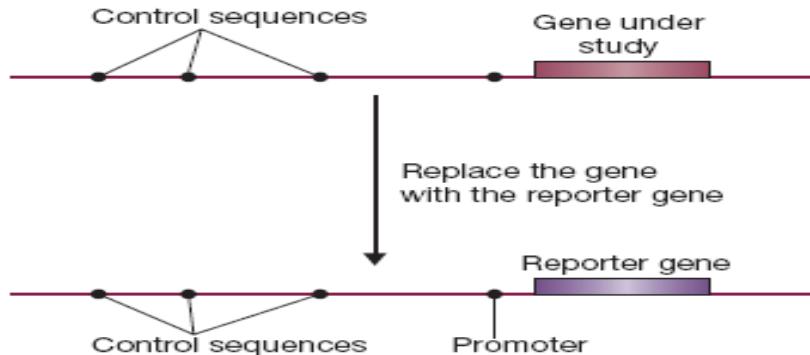
# Rapid amplification of cDNA ends



**Figure 11.7**  
One version of RACE.

# Regulation of Gene Expression

# Promoter Reporter Assays



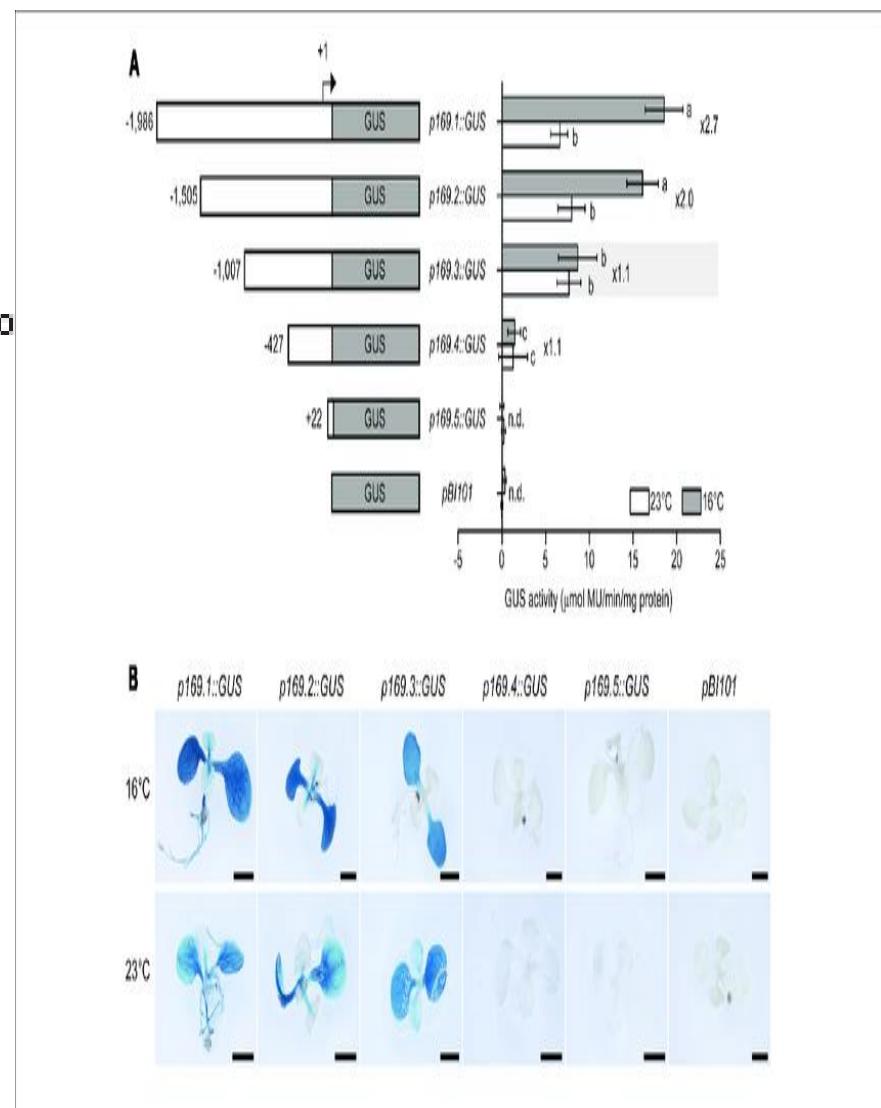
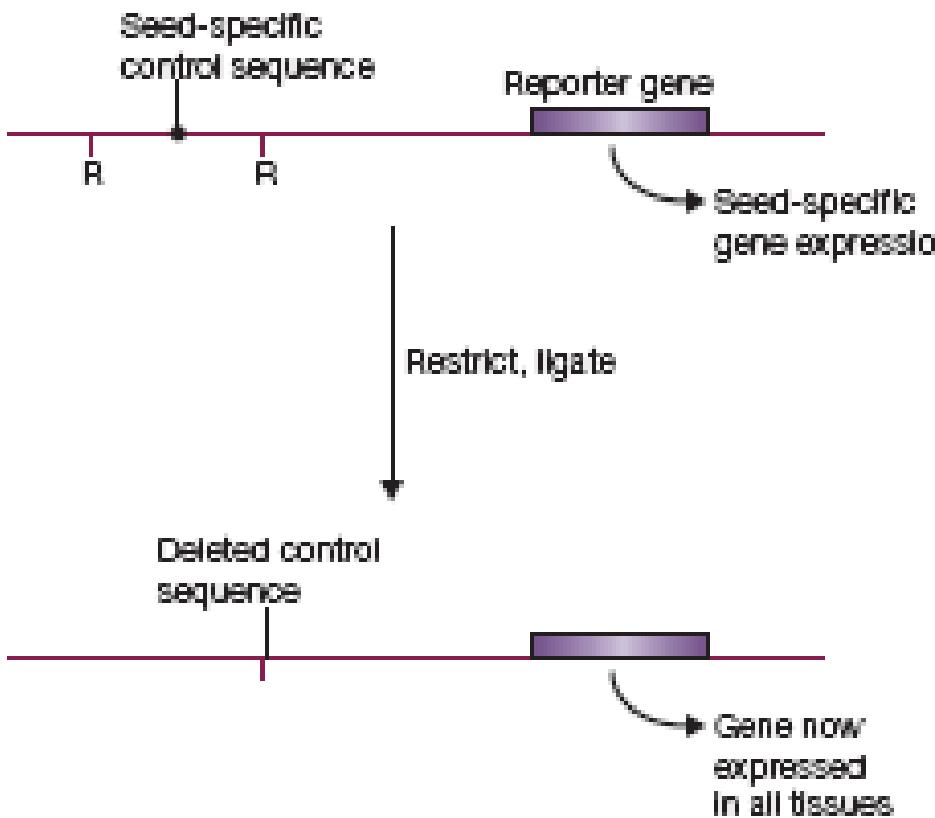
## Reporter Gene Assay

- 1) Join the regulatory sequence of interest to a reporter gene or some marker carried in an expression vector.
- 2) The recombinant construct is introduced into the cell and the activity of the regulatory sequence (i.e. promoter activity) can be determined.
- 3) Reporter genes or markers provide a convenient means to identify and analyze the regulatory elements of genes.
- 4) Reporter system measures the transcriptional activity (the interaction of the cis-elements on the promoter with the trans-acting factors).

A few examples of reporter genes used in studies of gene regulation in higher organisms.

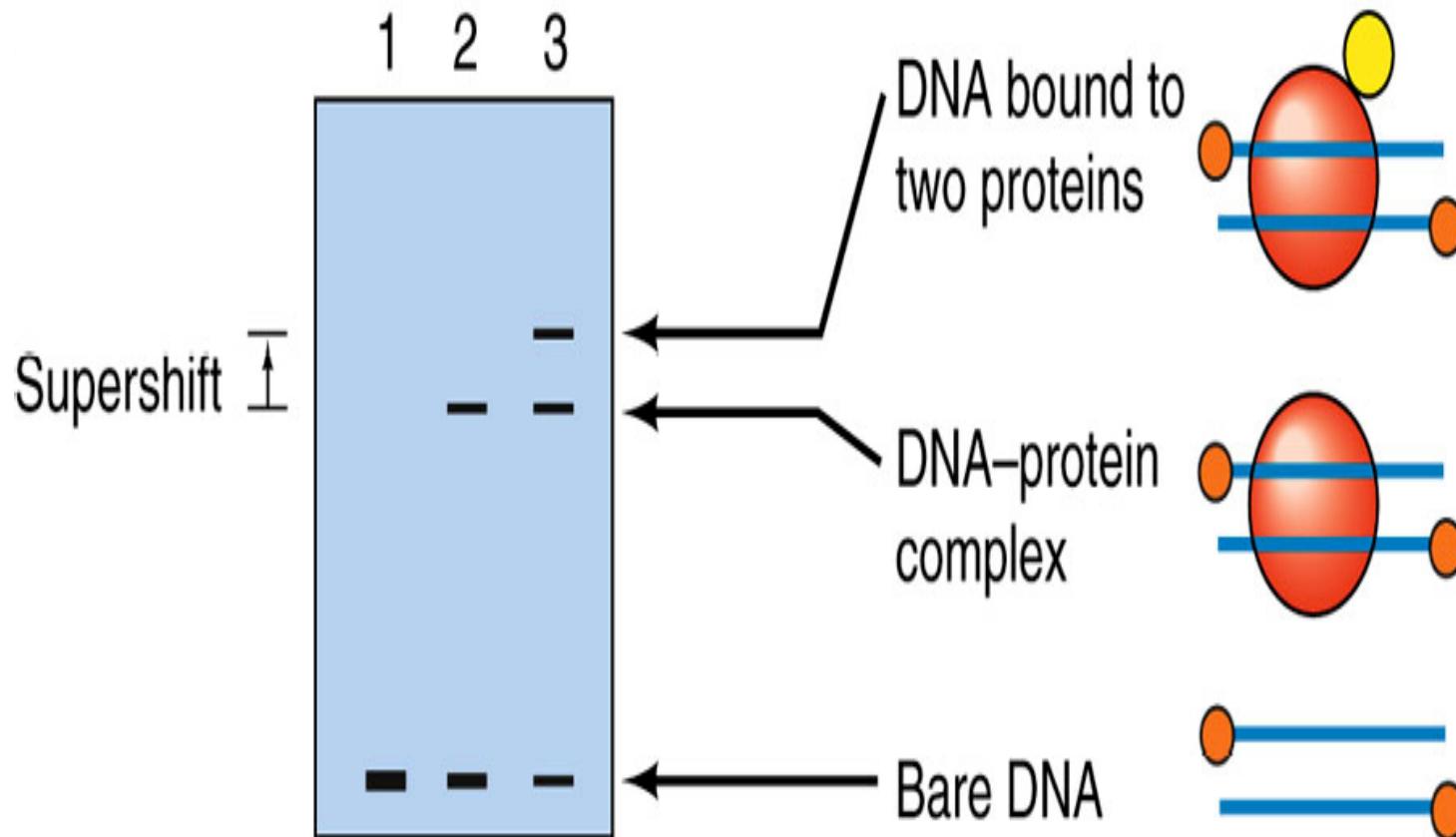
GENE*	GENE PRODUCT	ASSAY
<i>lacZ</i>	$\beta$ -Galactosidase	Histochemical test
<i>neo</i>	Neomycin phosphotransferase	Kanamycin resistance
<i>cat</i>	Chloramphenicol acetyltransferase	Chloramphenicol resistance
<i>dhfr</i>	Dihydrofolate reductase	Methotrexate resistance
<i>aphIV</i>	Hygromycin phosphotransferase	Hygromycin resistance
<i>lux</i>	Luciferase	Bioluminescence
GFP	Green fluorescent protein	Fluorescence
<i>uidA</i>	$\beta$ -Glucuronidase	Histochemical test

# Deletion Analysis

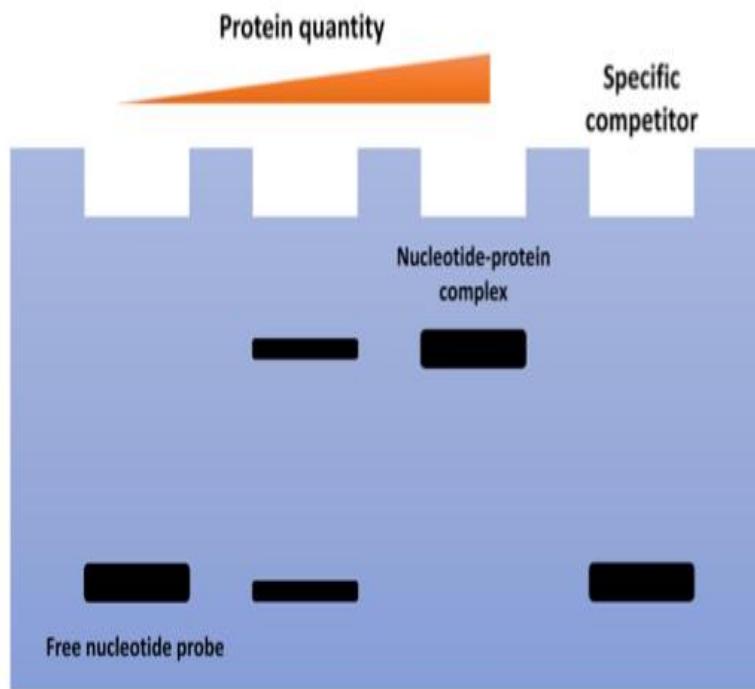


# Electrophoretic mobility shift assay (EMSA)

Detect interaction between a protein and DNA by the retardation of the electrophoretic mobility of DNA that occurs upon binding to the protein



## Gel mobility shift assay



Lane

EBNA Extract

Unlabeled EBNA DNA

Unlabeled Oct-1 DNA

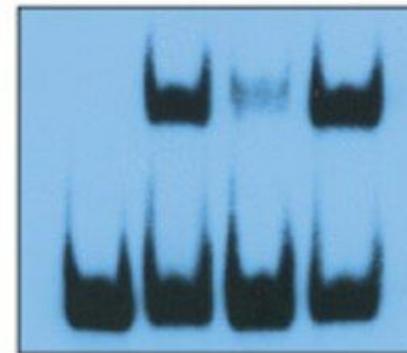
1

2

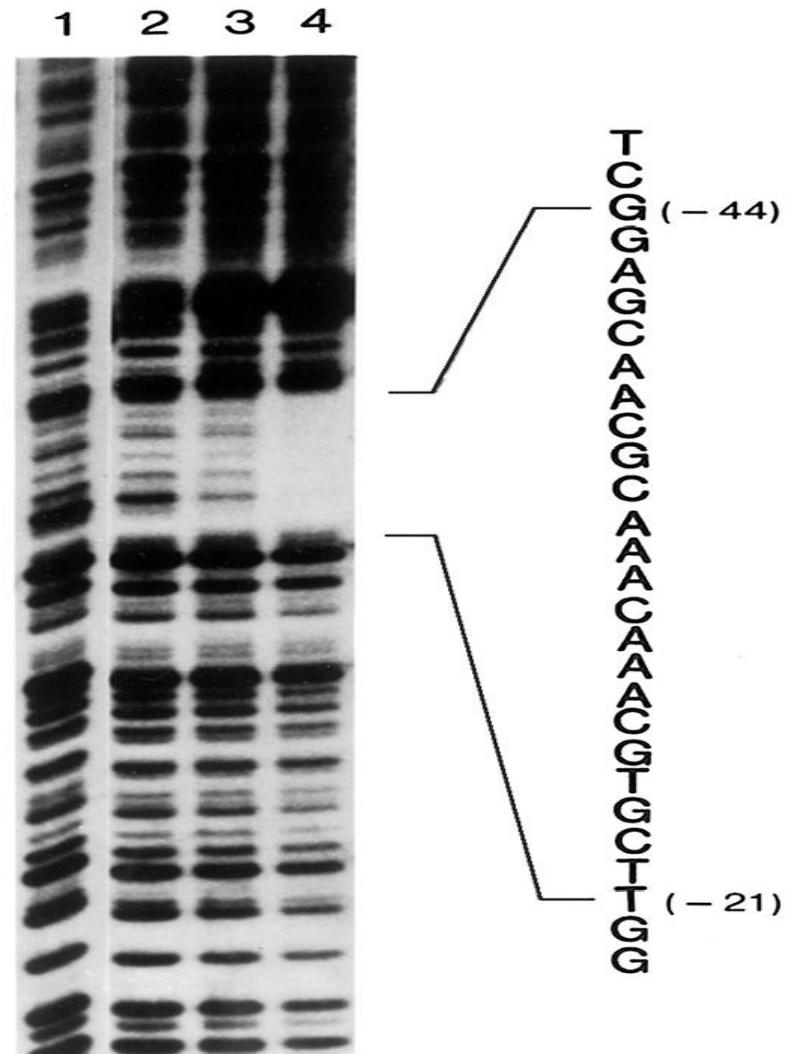
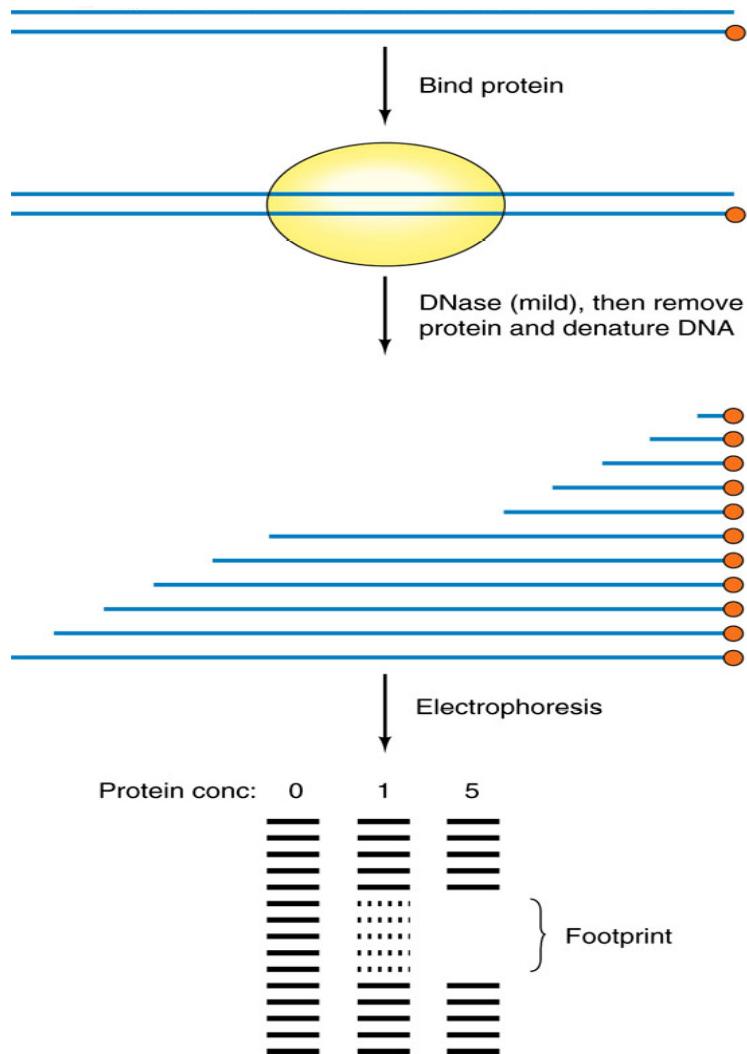
3

4

	1	2	3	4
EBNA Extract	-	+	+	+
Unlabeled EBNA DNA	-	-	+	-
Unlabeled Oct-1 DNA	-	-	-	+



## Footprinting assay on DNA-protein interaction



- **Gene of Interest- Cyclin D**
- Questions: Is CyclinD highly expressed in Cancer?
- Question : Does it have alternatively spliced forms?
- Question : What effect Cyclin D overexpression has on the cellular transcriptome?
- Question : Which transcription factors are involved in the regulation of Cyclin D levels?
- Question : Where do these TFs bind?

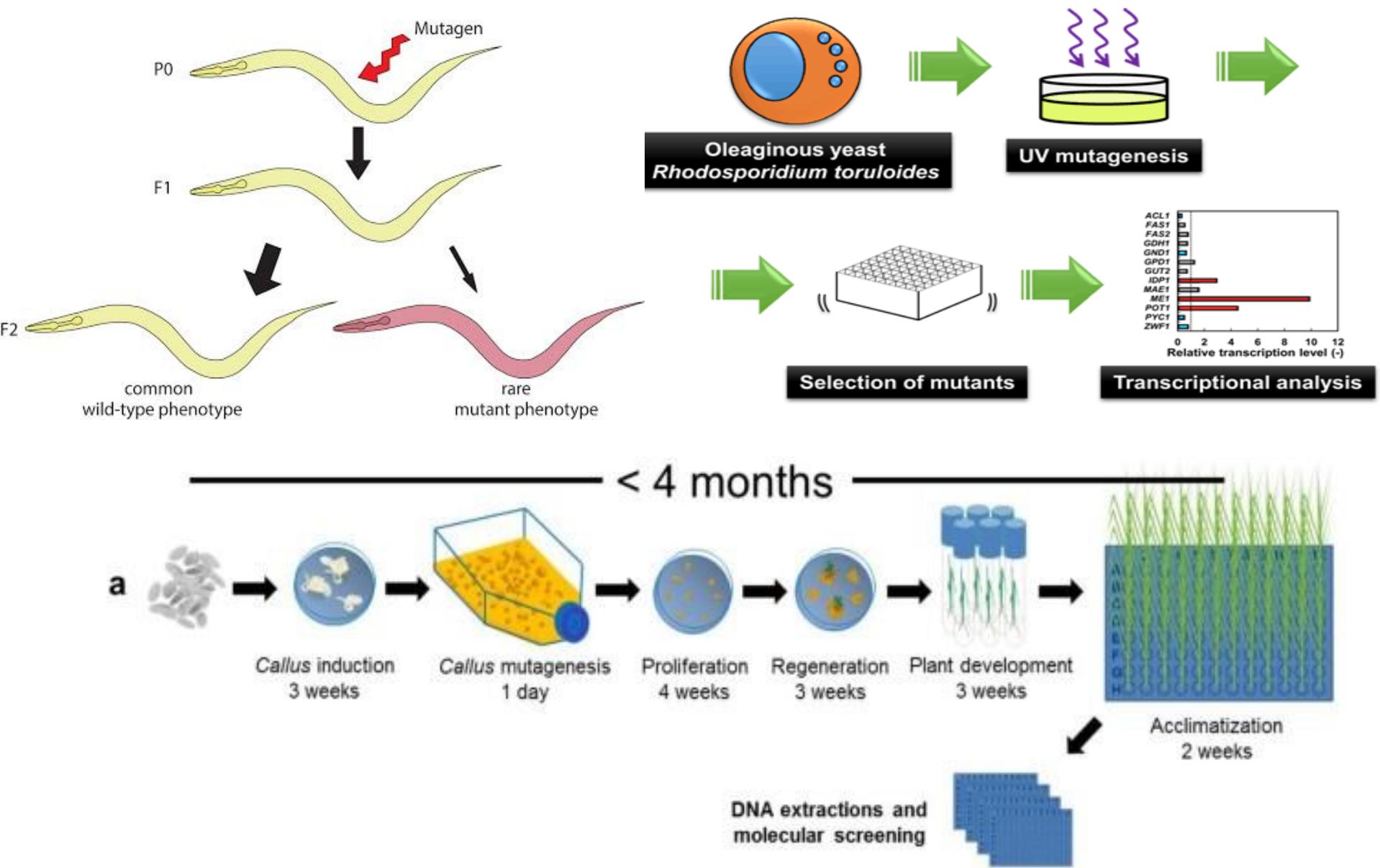
# Mutagenesis Techniques

# Mutagenesis

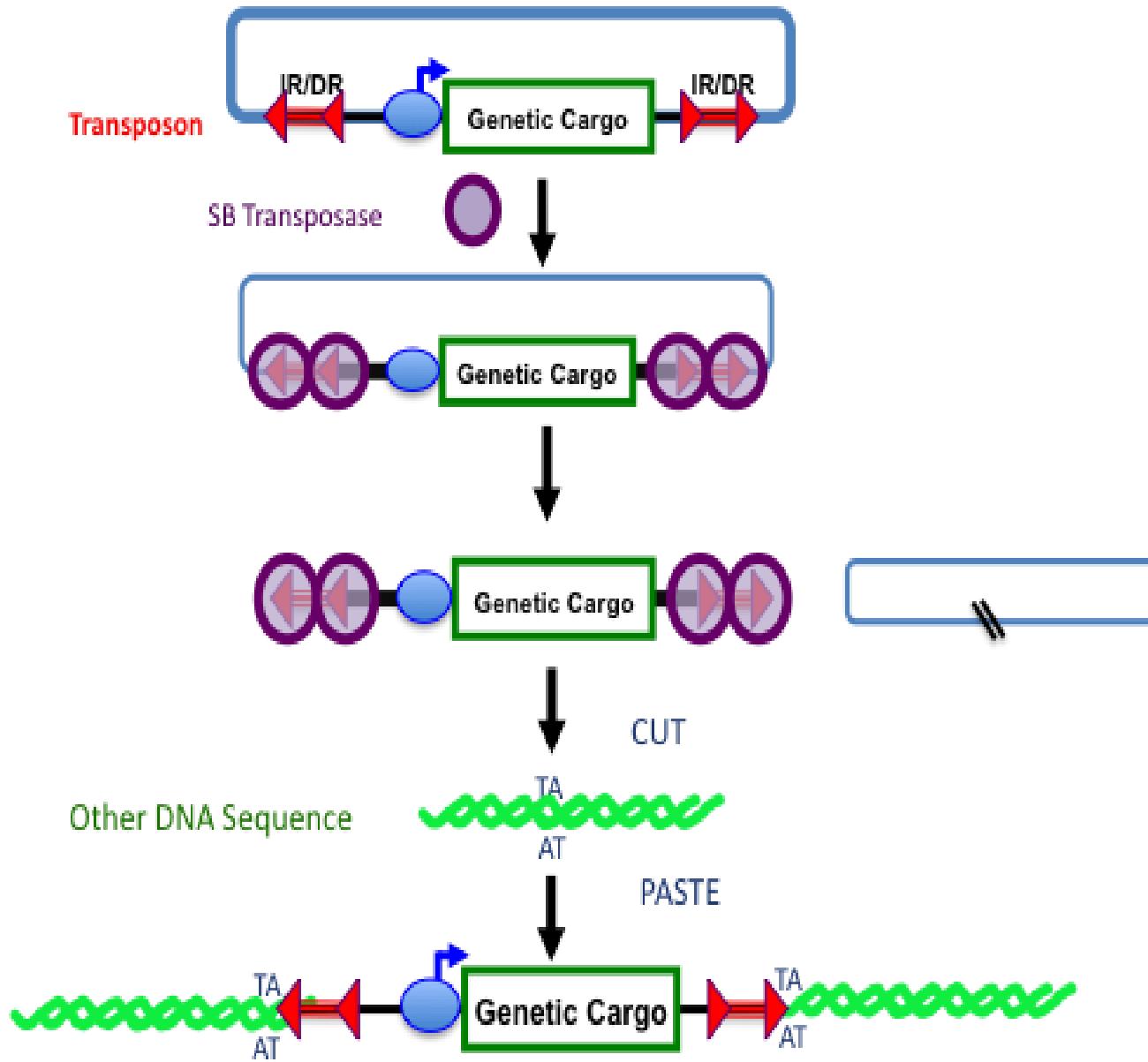
- Random mutagenesis
  - UV rays, X-ray irradiated, chemical (alkylating agents like Ethylmethanesulfonate (EMS))
  - Transposon Mutagenesis
  - Error Prone PCR
    - Mn<sup>2+</sup>/MgCl<sub>2</sub> mediated
    - Error Prone Taq polymerase
    - DNA shuffling
- Site Directed Mutagenesis
  - Restriction enzyme dependent
  - Oligonucleotide Directed
  - Artificial Gene Synthesis
  - PCR mediated
    - Substitution
    - Deletion
    - Insertion
- Examples

# Random Mutagenesis

# Random Mutagenesis



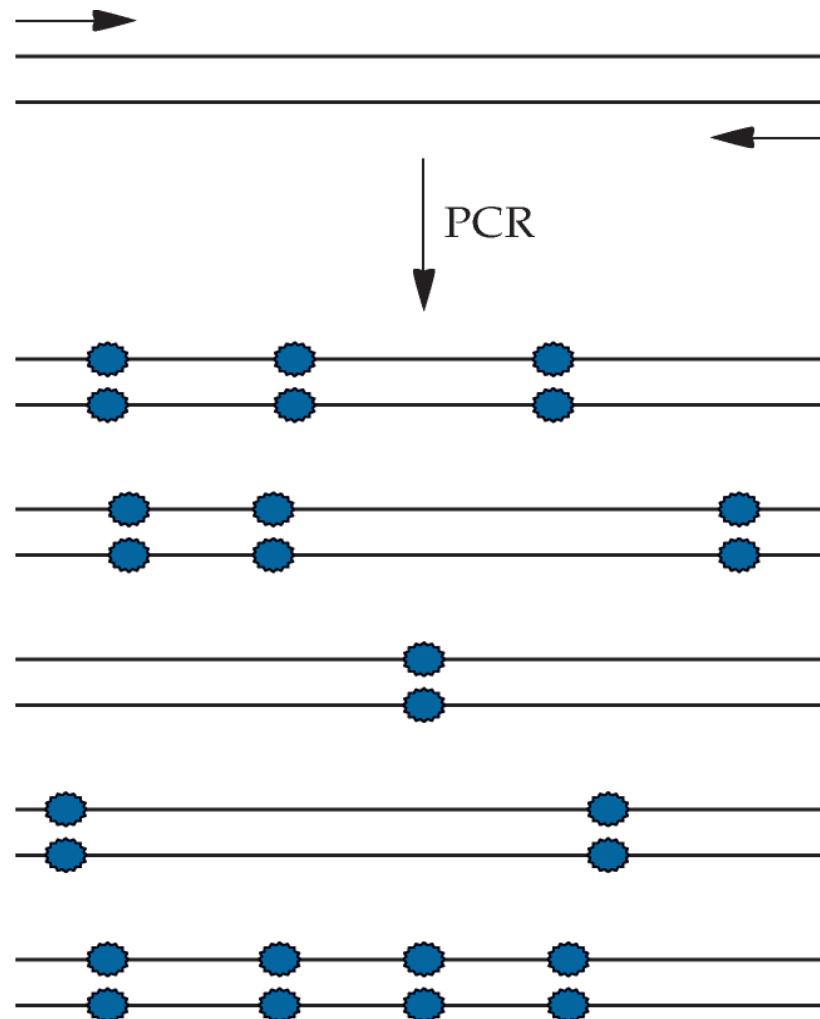
# Transposon Mutagenesis



## Error-Prone PCR

- It is a powerful method for random mutagenesis.
- *Taq* DNA polymerase lacks proof reading activity .
- Adding  $Mn^{2+}$ , increasing the concentration of  $Mg^{2+}$ , and adding unequal amount of 4 dNTPs will increase the error rate.
- Use high error prone *Taq* polymerases or mutazymes
- Following the PCR, the randomly mutagenized DNA is cloned into expression vectors and screened for altered or improved protein activity.
- The desired clones are then isolated and sequenced so the changes can be elaborated.

- Error – prone PCR of a target gene yields a variety of mutated forms of the gene.



# DNA Shuffling

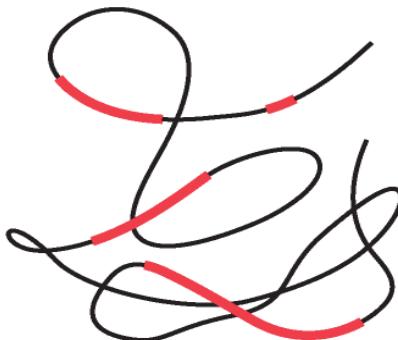
- Some proteins are encoded by a group of related genes called a gene family.
- It is possible using common restriction site to recombine domains from different members of the family to look for proteins with unusual characteristics.
- Also, some of the hybrid proteins may combine important attributes of two or more of the original proteins, e.g., high activity and thermostability.



Wild type

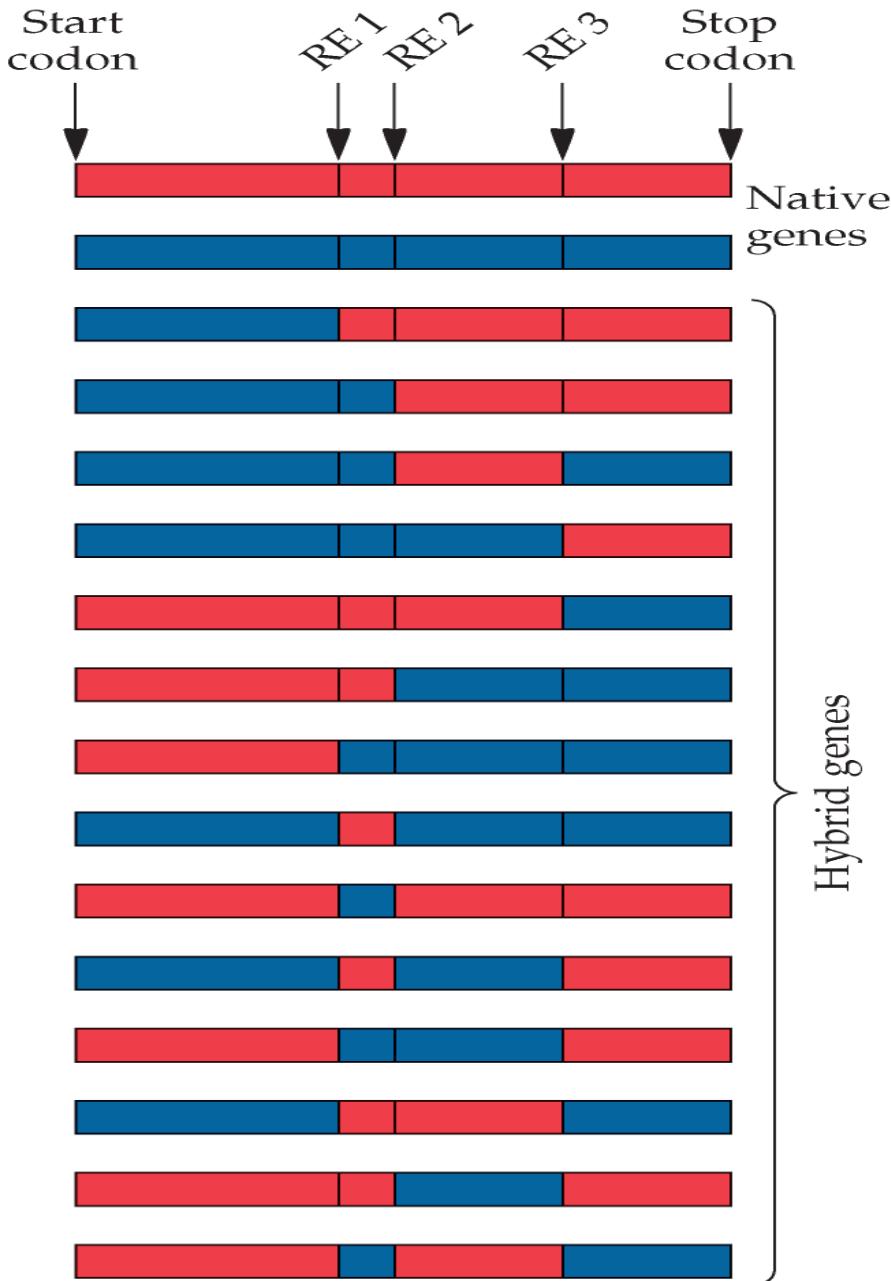


Random  
mutagenesis  
or error-prone PCR



DNA shuffling

- Random mutagenesis or error-prone PCR causes single-amino-acid substitutions.
- DNA shuffling, in which genes are formed with large regions from different sources.

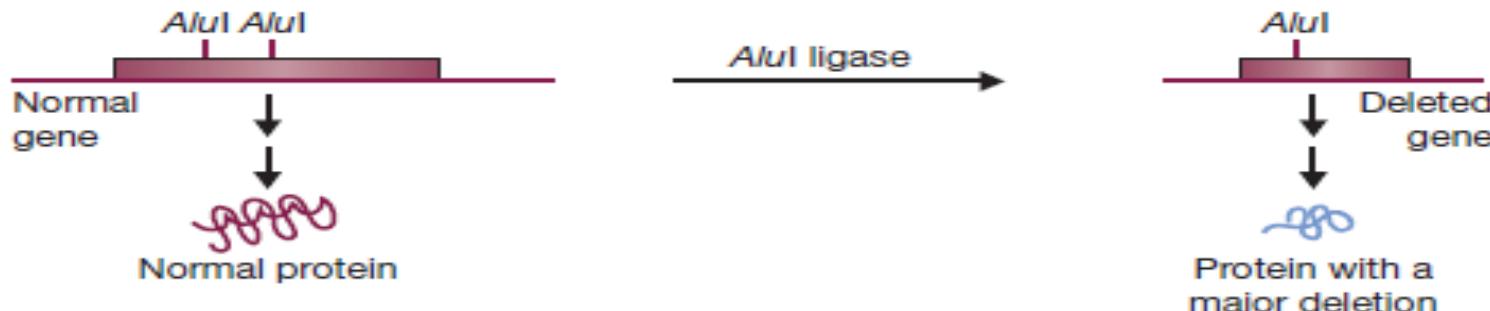


- Digestion of two or more of the DNAs that encode the native forms of similar proteins with one or more restriction enzymes that cut the DNA in the same place.
- Followed by ligation of the mixture of DNA fragments, a large number of hybrids are generated.

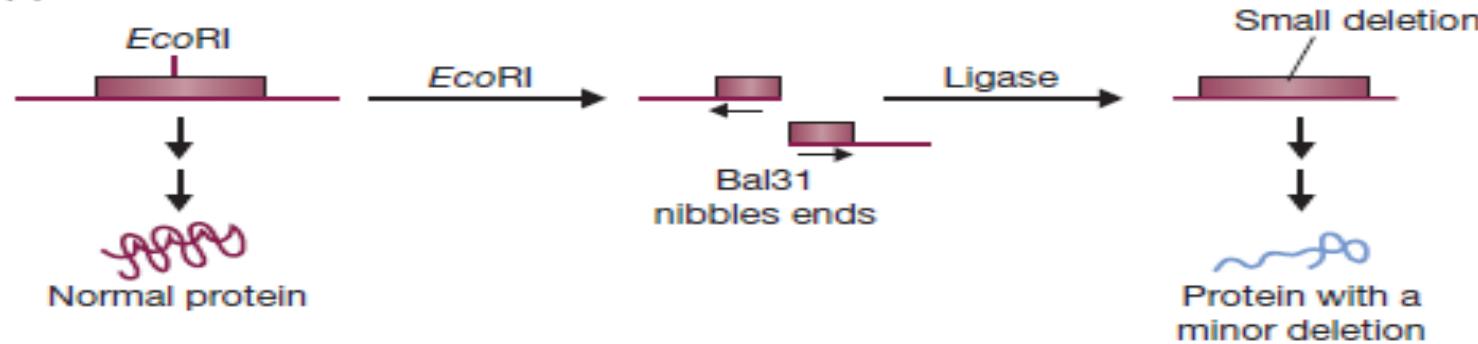
# Site-directed mutagenesis

# Directed *in vitro* mutagenesis

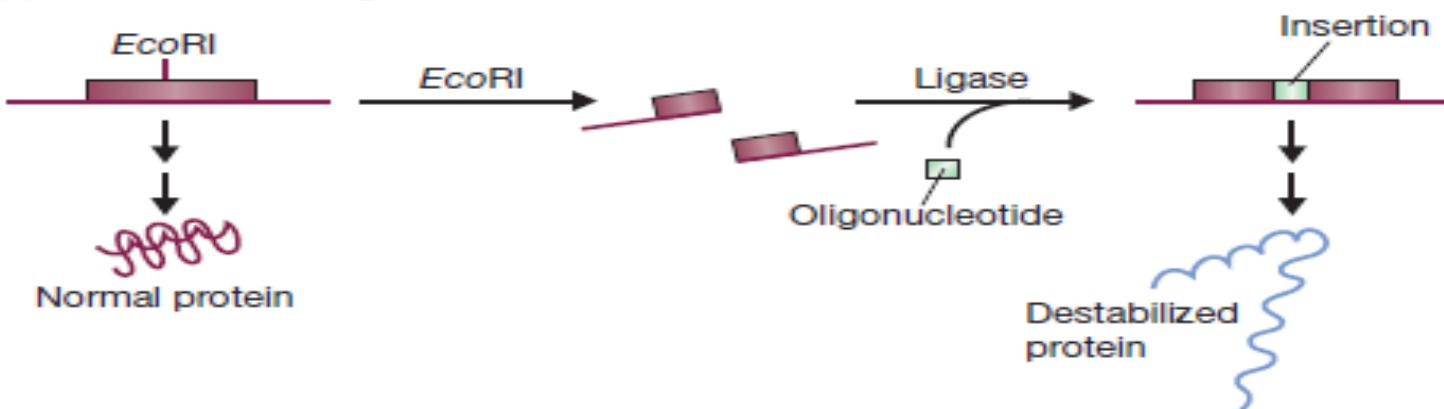
## (a) Restriction fragment deletion



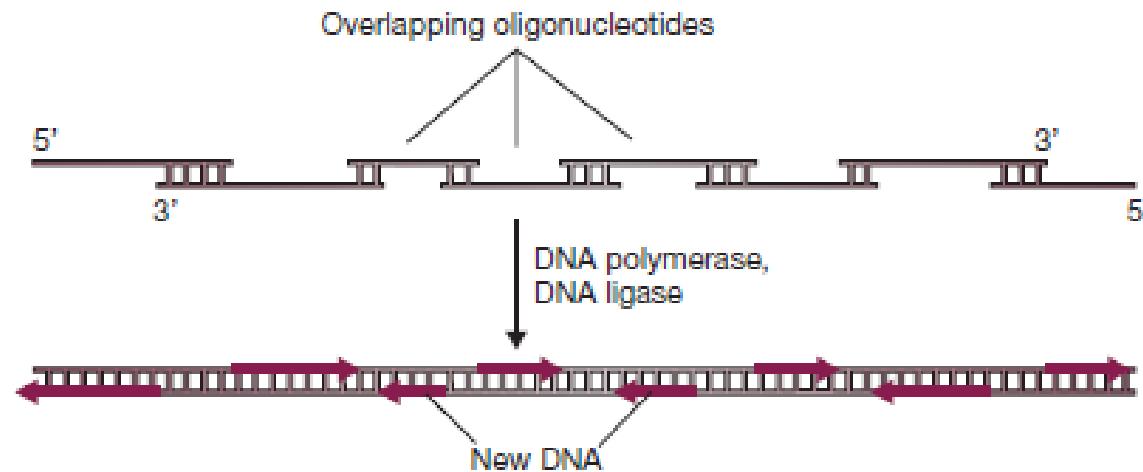
## (b) Nucleotide removal at restriction site



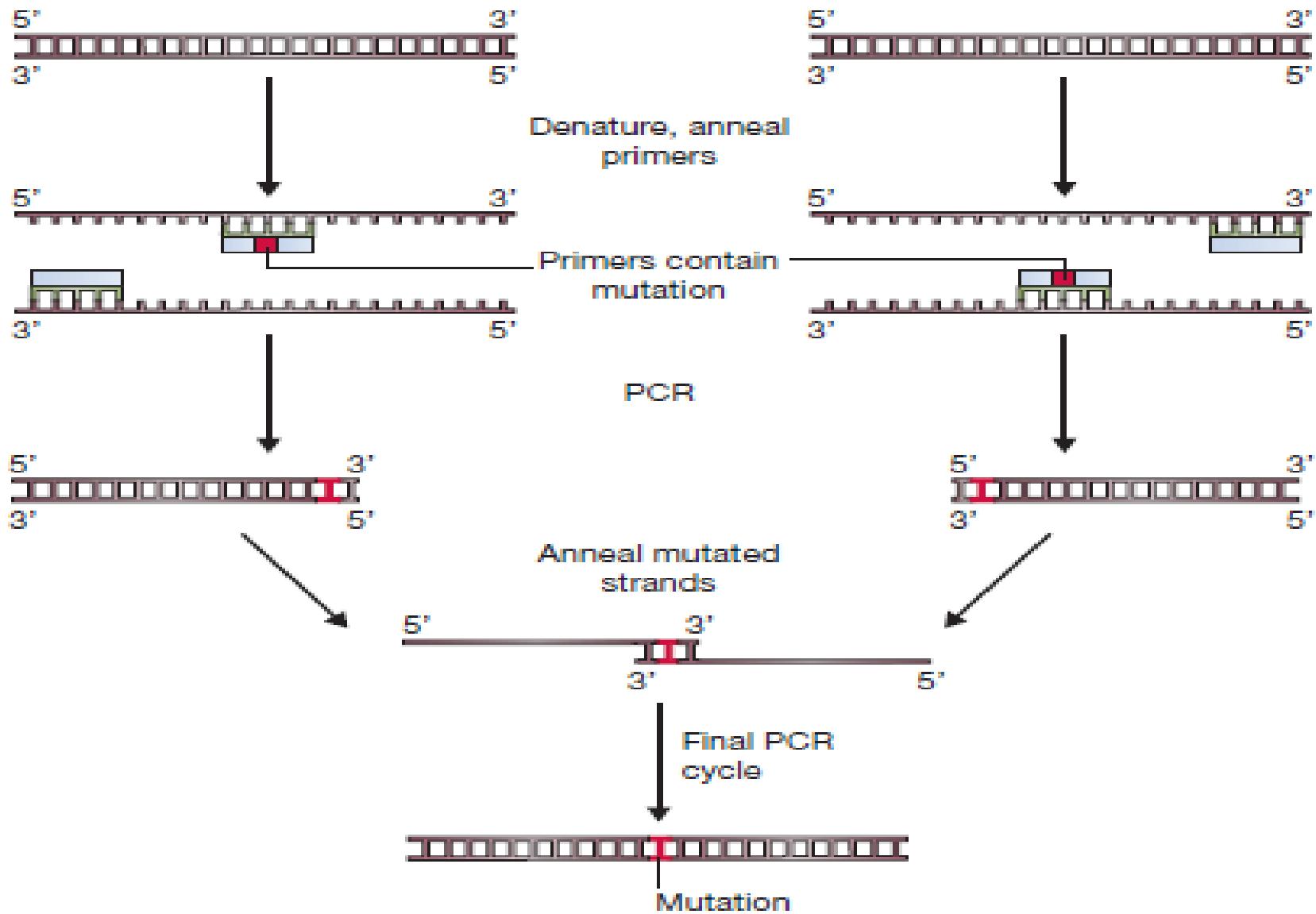
## (c) Insertion of an oligonucleotide



# Artificial Gene Synthesis to introduce multiple mutations

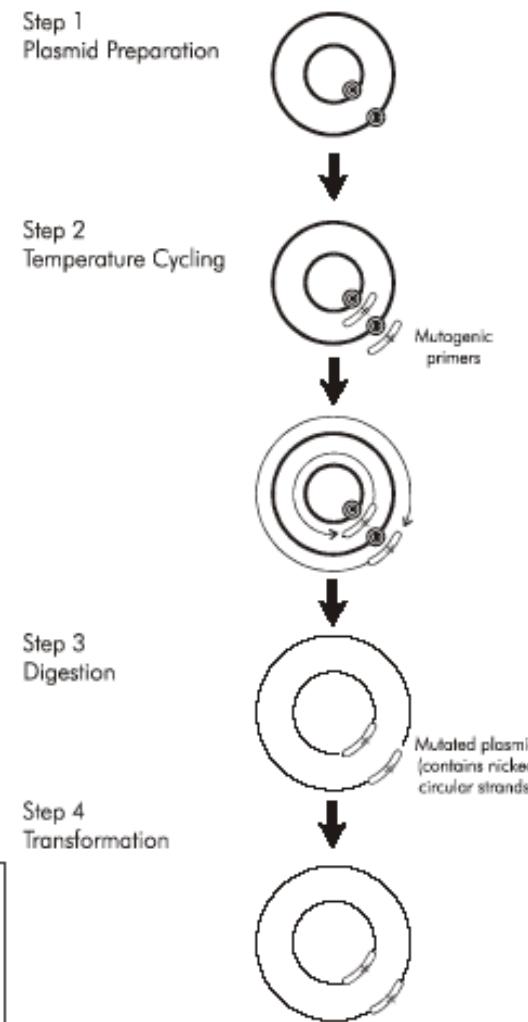


# PCR for in-vitro mutagenesis



# PCR mediated Site-directed mutagenesis

## Workflow (Stratagene)



Gene in plasmid with target site mutation

Denature the plasmid and anneal the oligonucleotide primers containing the desired mutation

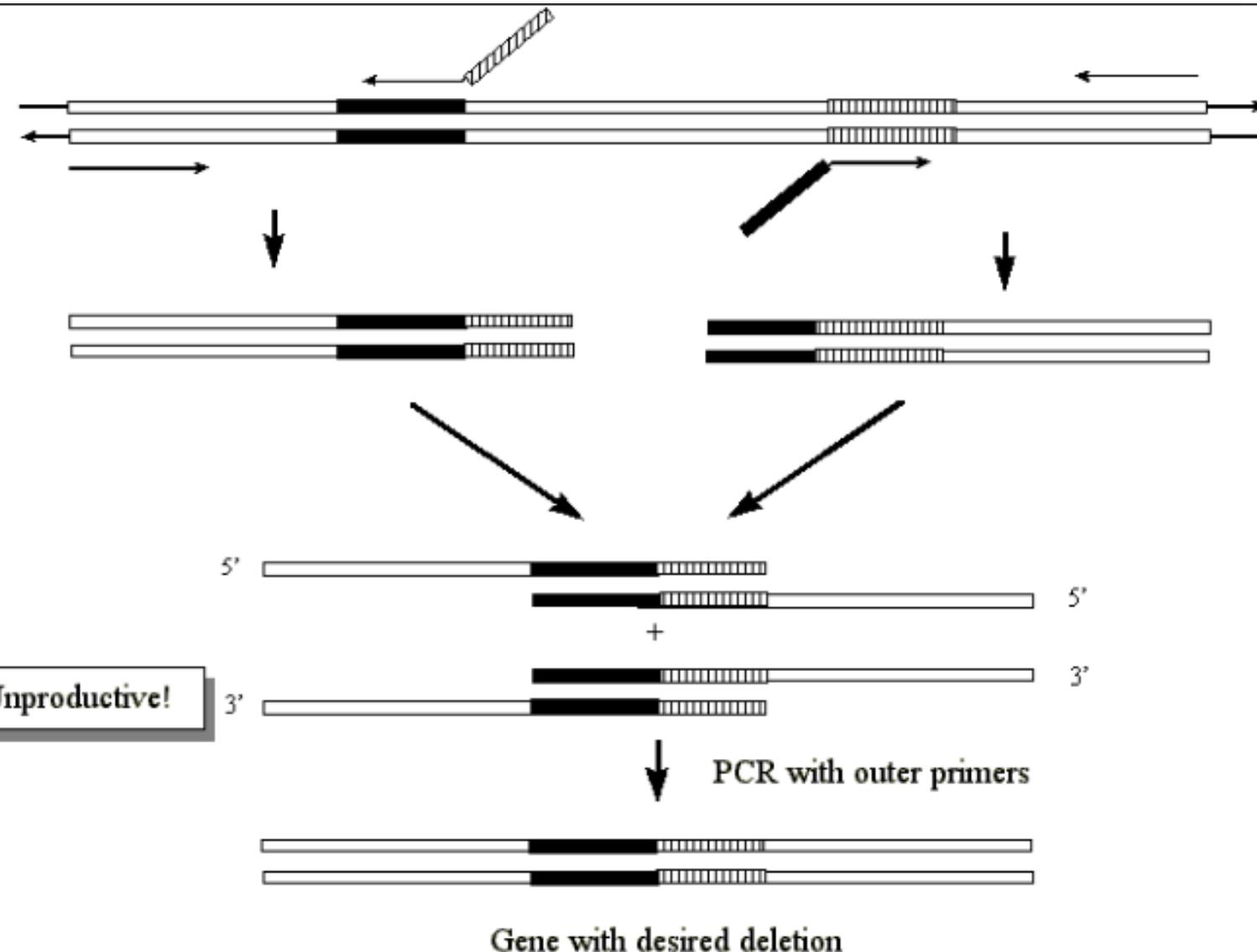
Using the non-strand-displacing action of *PfuTurbo* polymerase, extend and incorporate the mutagenic primers resulting in nicked circular strands

Digest the methylated, nonmutated parental DNA template with Dpn I

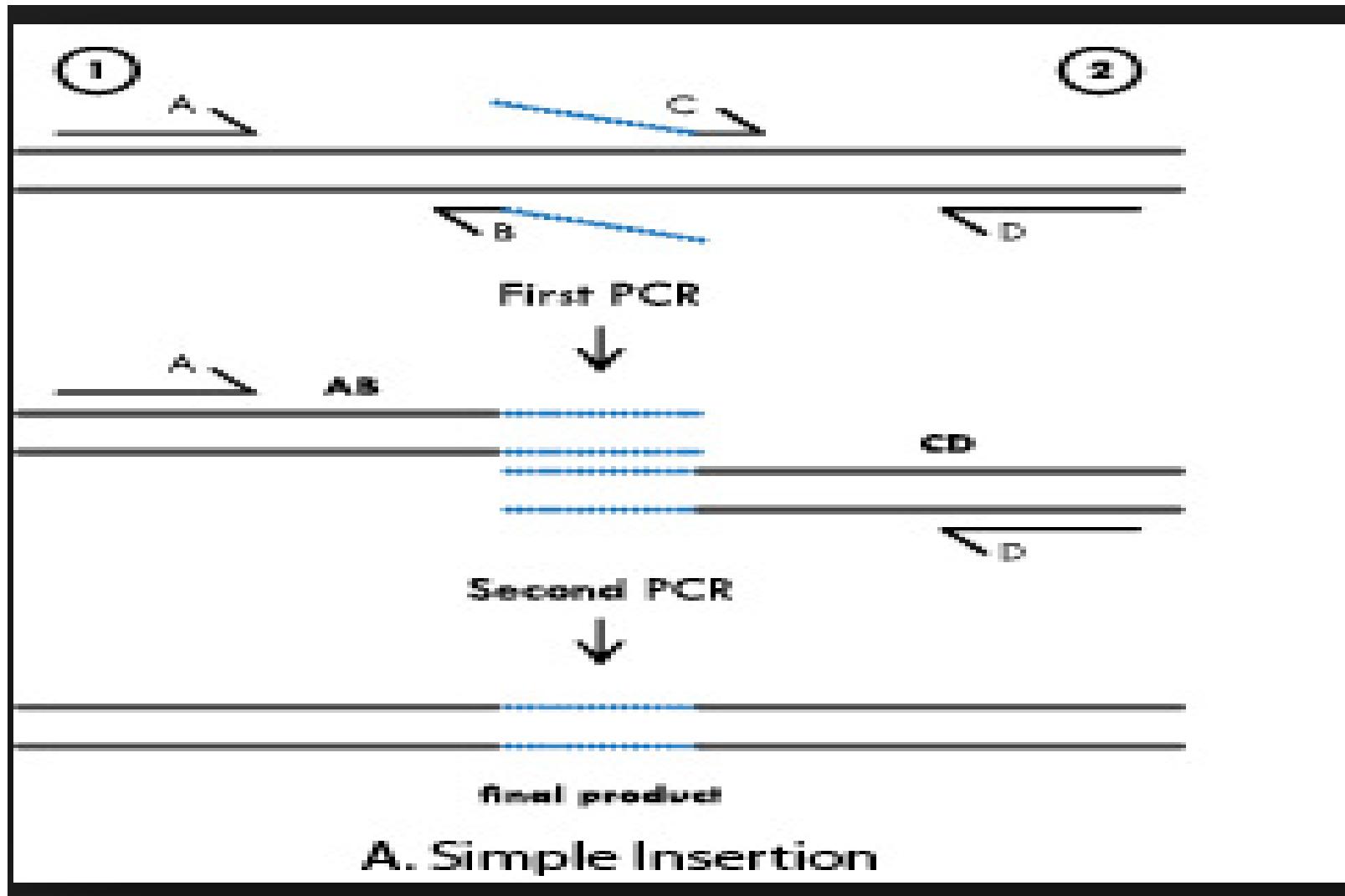
Transform the circular, nicked dsDNA into supercompetent cells

After transformation the supercompetent cells repair the nicks in the mutated plasmid

# Generation of deletion mutants



# Generation of insertion mutants



# Protein Engineering

- Of the many thousand enzymes characterized, only few are used greatly in industrial processes.
- The major barrier is finding enzymes with the desired activities and well suited for a highly specialized industrial application.
- These naturally occurring enzymes are easily denatured by high temperature and organic solvents used in many processes.

# Protein Engineering

**TABLE 8.1** Some industrial enzymes and their commercial uses

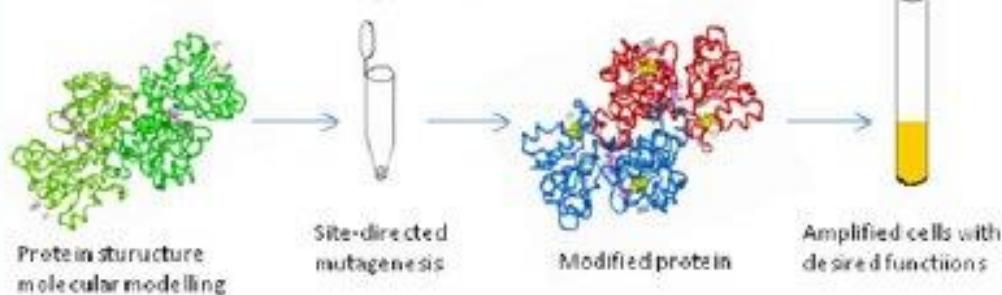
Enzyme	Industrial use(s)
$\alpha$ -Amylase	Beer making, alcohol production
Aminoacylase	Preparation of L-amino acids
Bromelain	Meat tenderizer, juice clarification
Catalase	Antioxidant in prepared foods
Cellulase	Alcohol and glucose production
Ficin	Meat tenderizer, juice clarification
Glucoamylase	Beer making, alcohol production
Glucose isomerase	Manufacture of high-fructose syrups
Glucose oxidase	Antioxidant in prepared foods
Invertase	Sucrose inversion
Lactase	Whey utilization, lactose hydrolysis
Lipase	Cheese making, preparation of flavorings
Papain	Meat tenderizer, juice clarification
Pectinase	Clarifying fruit juices, alcohol production
Protease	Detergent, alcohol production
Rennet	Cheese making

# In vitro mutagenesis

- Km &/or Vmax
- Temperature or pH stability
- Reactivity in non-aqueous solvents
- Cofactor requirements
- Substrate binding or specificity
- Stability to proteases
- Allosteric regulation

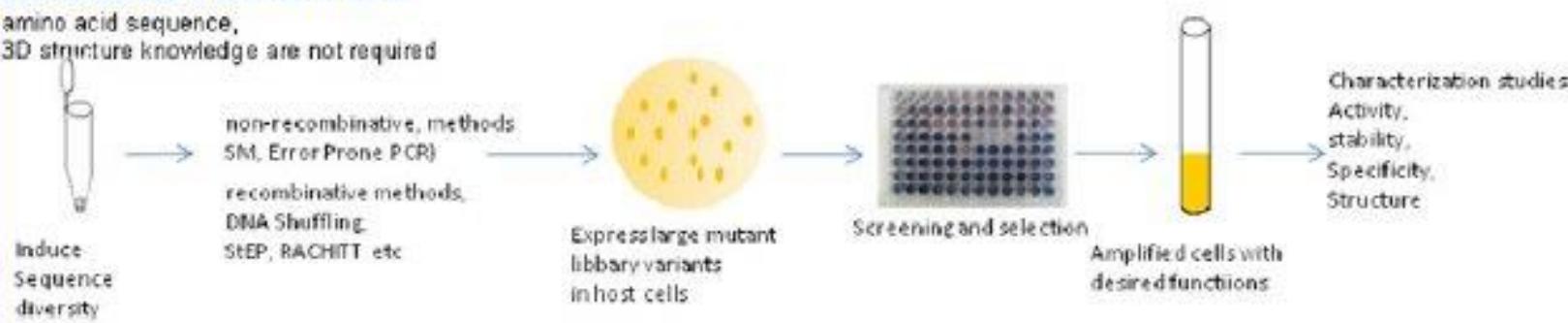
## RATIONAL DESIGN

amino acid sequence,  
3D structure and function knowledge are required



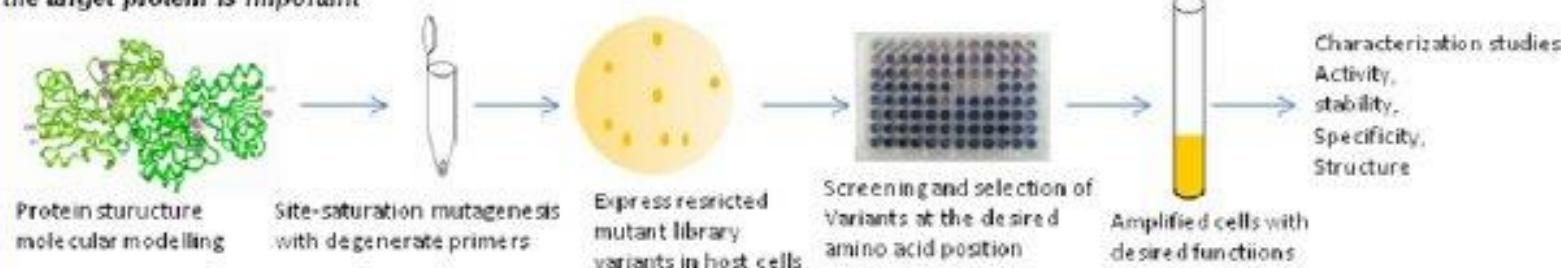
## DIRECTED EVOLUTION

amino acid sequence,  
3D structure knowledge are not required



## SEMI-RATIONAL DESIGN

Structure-function relationship of related site of the target protein is important



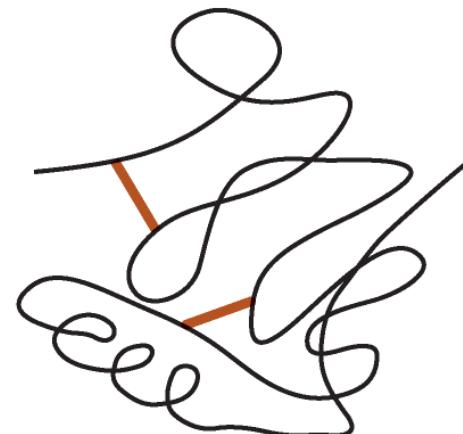
Some Examples for your reference  
Please go through these

# Adding disulfide bonds

- There is generally a direct correlation between the number of disulfide bonds and the thermo-stability of a protein.
- To increase stability of a variety of enzymes disulfide bonds have been added to their structure.
- The problem is whether extra disulfide bonds perturb the normal function.



Native protein



Engineered protein

# Adding disulfide bonds

- T4 lysozyme-6 variants were created using oligonucleotide-directed mutagenesis of amino acid whose R-groups were in close proximity to incorporate more cysteine residues.
- The variants were tested and were generally found to be more thermostable.
- Some variants however lacked activity.
- It is a trial-and-error process.
- However, it is clear that increasing disulfide bonds to enhance protein stability is feasible.

# Adding disulfide bonds

**TABLE 8.2** Properties of T4 lysozyme and six engineered variants

Enzyme	Amino acid at position:							No. of -S-S-	% Activity	$T_m$ (°C)
	3	9	21	54	97	142	164			
wt	Ile	Ile	Thr	Cys	Cys	Thr	Leu	0	100	41.9
pwt	Ile	Ile	Thr	Thr	Ala	Thr	Leu	0	100	41.9
A	<u>Cys</u>	Ile	Thr	Thr	<u>Cys</u>	Thr	Leu	1	96	46.7
B	Ile	<u>Cys</u>	Thr	Thr	Ala	Thr	<u>Cys</u>	1	106	48.3
C	Ile	Ile	<u>Cys</u>	Thr	Aka	<u>Cys</u>	Leu	1	0	52.9
D	<u>Cys</u>	<u>Cys</u>	Thr	Thr	<u>Cys</u>	Thr	<u>Cys</u>	2	95	57.6
E	Ile	<u>Cys</u>	<u>Cys</u>	Thr	Ala	<u>Cys</u>	<u>Cys</u>	2	0	58.9
F	<u>Cys</u>	<u>Cys</u>	<u>Cys</u>	Thr	<u>Cys</u>	<u>Cys</u>	<u>Cys</u>	3	0	65.5

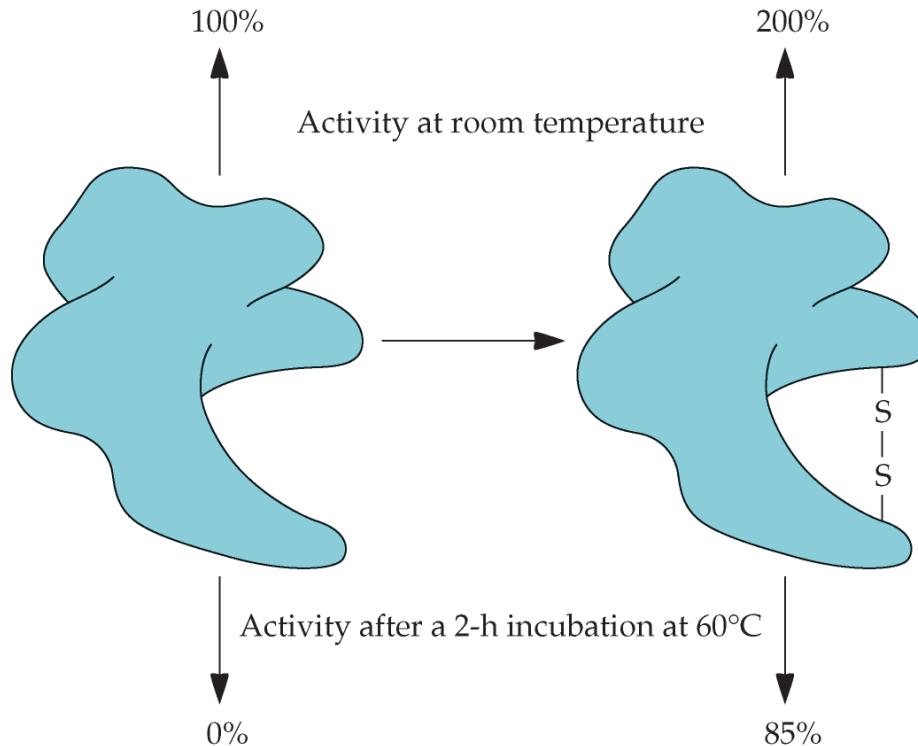
Adapted from Matsumura et al., *Nature* 342:291–293, 1989.

wt, wild-type T4 lysozyme; pwt, pseudo-wild-type enzyme; A through F, six engineered cysteine variants; -S-S-, disulfide bonds;  $T_m$ , “melting” temperature (a measure of thermostability).

# Adding disulfide bonds

- Similar experiments were used to produce a more thermostable xylanase.
- An enzyme used in the degradation of hemi-cellulose, the compound generally darkens paper.
- The use of the variant xylanase that function efficiently at high industrial temperature allows the use of less bleach in paper production.
- Less pollution

- Addition of a disulfide bond stabilizes the protein.
- Its activity is doubled at room temperature, and protected against heat inactivation.



# Changing Asparagine to Other Amino Acids

- Asparagine and glutamine residues can undergo deamination at high temperature becoming aspartic and glutamic acid, respectively.
- The change could lead to improper folding and loss of enzyme activity.
- A yeast enzyme served as the model and showed increased thermostability when Asn was changed to threonine and isoleucine.
- Similarly, a modified longer-lasting insulin was produced by changing Asp to Gly (approved for therapeutic use).

# Changing Asparagine to Other Amino Acids

**TABLE 8.3** Stability at 100°C of the yeast enzyme triosephosphate isomerase and its engineered derivatives

Enzyme	Amino acid at position:		Half-life (min)
	14	78	
Wild type	Asn	Asn	13
Variant A	Asn	Thr	17
Variant B	Asn	Ile	16
Variant C	Thr	Ile	25
Variant D	Asp	Asn	11

Adapted from Ahern et al., *Proc. Natl. Acad. Sci. USA* **84**:657–679, 1987.

Enzyme stability is expressed as the half-life, or rate of enzyme inactivation, at 100°C. A longer half-life indicates a more stable enzyme.

# Increasing Enzymatic Activity

- Experiments were performed using an Tyrosyl-tRNA synthetase for which the active site was well characterized.
- A threonine at position 51 of the protein was targeted for modification.
- If it was replaced with an alanine, the  $K_m$  of the enzyme decreased 2X without altering the rate.
- If it was replaced with proline, the enzyme bound ATP 100X more tightly than normal and increased its efficiency.

# Increasing Enzymatic Activity

**TABLE 8.4** Aminoacylation activity of native (Thr-51) and modified (Ala-51 and pro-51) tyrosyl-tRNA synthetases

Enzyme	$k_{\text{cat}} (\text{s}^{-1})$	$K_m (\text{mM})$	$k_{\text{cat}}/K_m (\text{s}^{-1} \text{ M}^{-1})$
Thr-51	4.7	2.5	1,860
Ala-51	4.0	1.2	3,200
Pro-51	1.8	0.019	95,800

Adapted from Wilkenson et al., *Nature* 307:187–188, 1984.

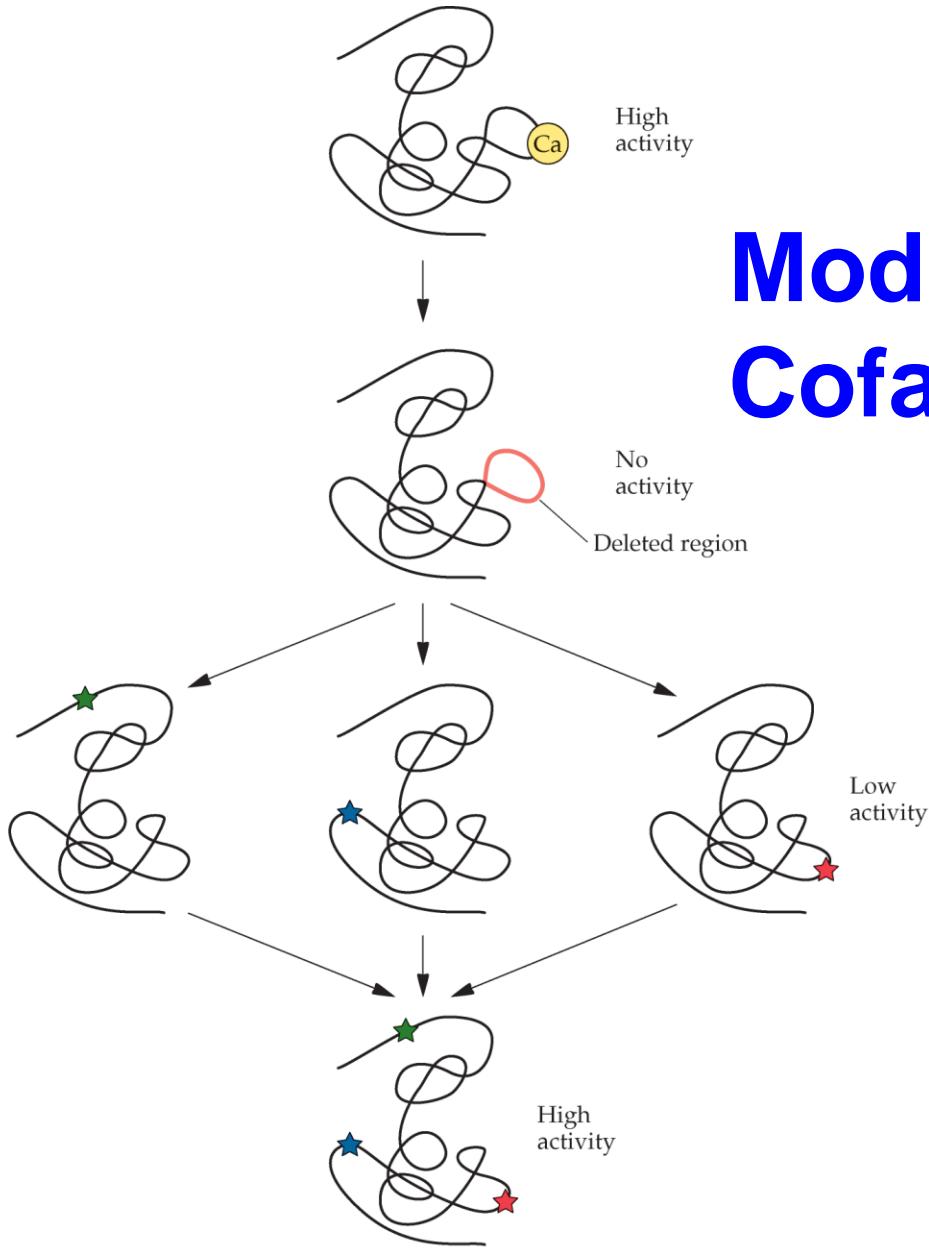
The units for  $K_m$ , the binding constant of the enzyme for ATP, are millimolar units (mM); the units for  $k_{\text{cat}}$ , the catalytic rate constant, are reciprocal seconds ( $\text{s}^{-1}$ ); and the units for  $k_{\text{cat}}/K_m$ , the catalytic efficiency, are  $\text{s}^{-1} \text{ M}^{-1}$ .

# Modifying Metal Cofactor Requirements

- Subtilisin was studied. This enzyme from gram-positive bacteria is widely used as biodegradable cleaning agents in laundry detergent.
- It requires  $\text{Ca}^{+2}$  as a cofactor to stabilize the enzyme.
- The enzyme is used in many industrial settings where there are a large number of metal-chelating agents that can bind to effectively remove calcium.
- These enzymes are rapidly inactivated under these conditions.

# Modifying Metal Cofactor Requirements

- To avoid this, the portion of the enzyme required for  $\text{Ca}^{+2}$  binding was determined and eliminated from the protein (= inactive protein).
- The researchers began changing amino acids to restore enzyme activity.
- The result was a modified subtilisin that retains its activity but no longer requires  $\text{Ca}^{+2}$  ions for improved industrial use.



# Modifying Metal Cofactor Requirements

# Modifying Metal Cofactor Requirements

**TABLE 8.5** Effects of random mutations of selected amino acid residues on the stability of subtilisin BPN lacking a calcium-binding domain

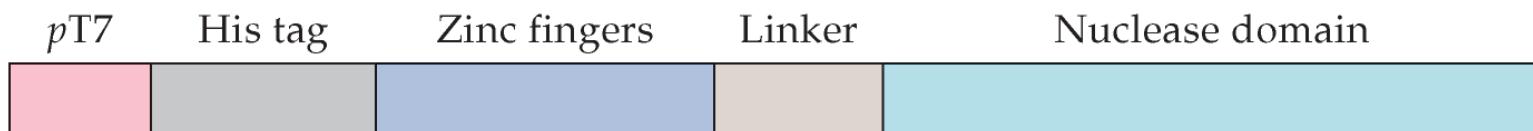
Region of protein	Amino acid residue	Stabilizing mutation	Fold increase in half-life
N terminus	2	Gln→Lys	2.0
	3	Ser→Cys	17.0
	4	None found	None
	5	Pro→Ser	1.2
Omega loop	41	Asp→Ala	1.2
	44	Lys→Asn	1.2
α-Helix	73	Ala→Leu	2.6
	74	None found	None
β-Pleat structure	206	Gln→Cys	17.0
	214	None found	None

Adapted from Strausberg et al., *Bio/Technology* 13: 669–673, 1995.

The mutations at positions 3 and 206 to Cys occur in the same clone and provide such a high level of stability because of the formation of the disulfide bridge between these residues.

# Modifying Protein Specificity

- *FokI* restriction enzyme from *Flavobacterium okeanokoites* was used as a model. It is a relatively non-specific nuclease.
- To add specificity to the enzyme, the portion of the gene encoding the catalytic domain was fused to another segment of DNA encoding histidine, three zinc finger domains, a linker peptide to confer flexibility.
- The modified enzyme was shown to recognize the expected site and to a lesser degree another two sites.



- Antibodies have also been used to demonstrate this.
- By modifying the hypervariable regions of the protein, it should be possible to generate antibodies that are directed against a wide range of antigenic determinants.
- Done by introducing mutations in the hypervariable complementarity-determining regions (CDR). Together, the six of these CDRs determine the specificity of an antibody.



- Three CDR on the heavy chain were modified using random mutagenesis with mixed degenerate primers.
- The first PCR modified the CDR 1.
- The second PCR modified CDR 2 and 3.
- The third PCR combined the three altered CDR into a single fragment.
- In one instance, a Fab fragment of a monoclonal antibody that was specific for the compound 11-deoxycortisol was altered to be specific for cortisol.
- This approach can be used to create Fab fragments directed toward any antigenic determinant depending on the method used in screening the library.

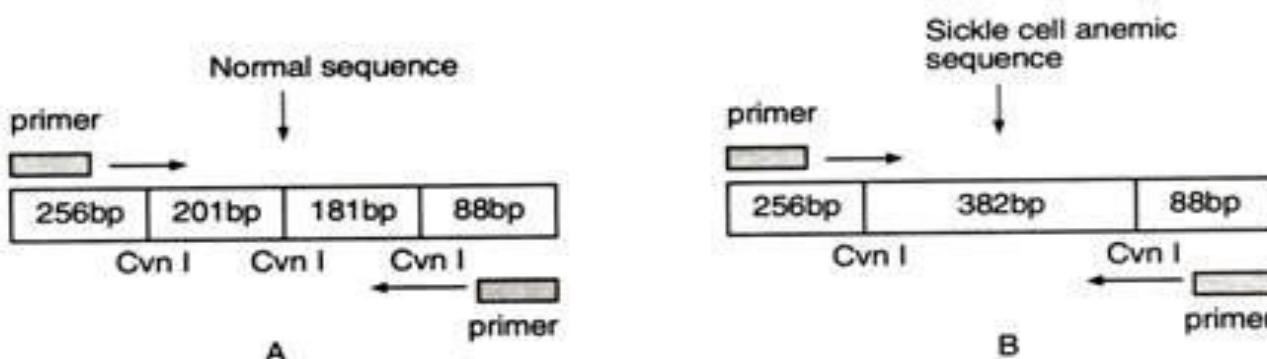
# **Applications of Recombinant DNA Technology in Medicine**

# Applications

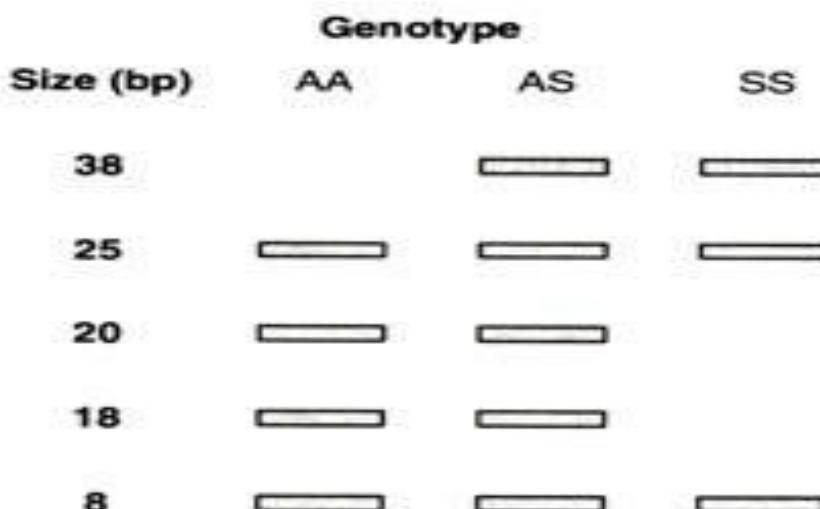
1. Diagnosis of Genetic Diseases – (PCR, RT-PCR, Probing etc.)
2. DNA Typing (DNA Fingerprinting)
3. Gene Therapy (Gene Overexpression/inhibition)
  1. Production of Therapeutic Proteins (eg Insulin), growth hormones (eg. Somatostatin) or clotting factor etc.
  2. Recombinant Vaccines (eg Hepatitis B Vaccine)
  3. Production of recombinant proteins by Pharming
    - a) Pharming using live animals
    - b) Pharming using live plants
4. Somatic or Germ Cell Therapy

# Disease Diagnosis

Sickle cell anemia- PCR based DNA diagnostic assay.



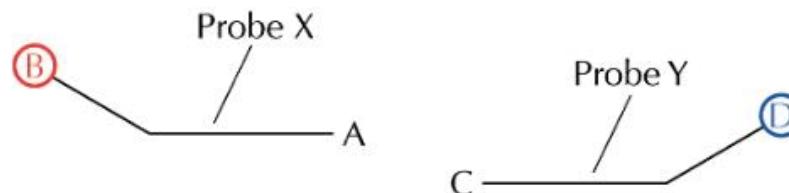
**Fig. 19.1.** Detection of Sick cell anemia gene at the DNA level. (A) Normal genotype (B) Mutation at the Cvn I site resulting in sickle cell anemia and loss of restriction site.



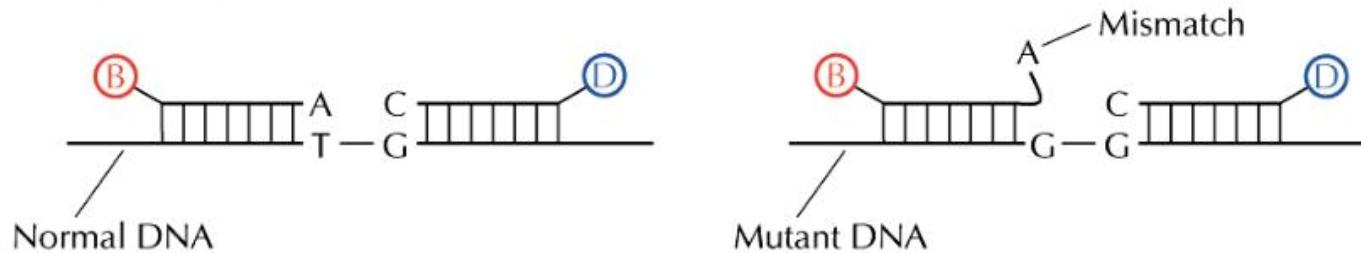
**Fig. 19.2.** Electrophoretic pattern of sickle cell diseased and healthy restriction digested PCR amplified  $\beta$ -globin DNA. AA, homozygous normal; AS, heterozygous; SS, homozygous sickle cell.

# Polymerase Chain Reaction-Oligonucleotide Ligation Assay (PCR-OLA) –For SNP detection

**A** Synthesize a pair of oligonucleotide probes



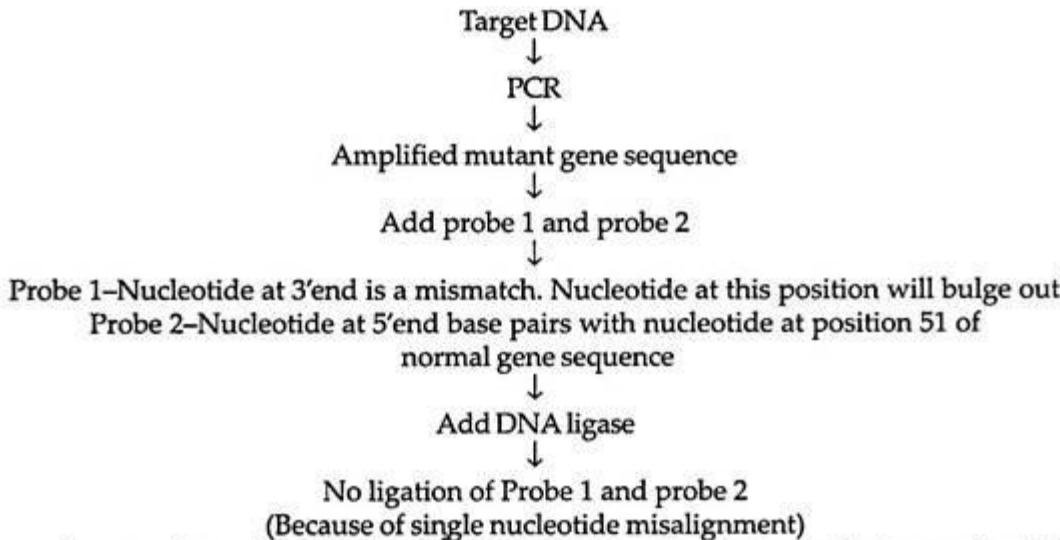
**B** Hybridize probes to PCR-amplified DNA



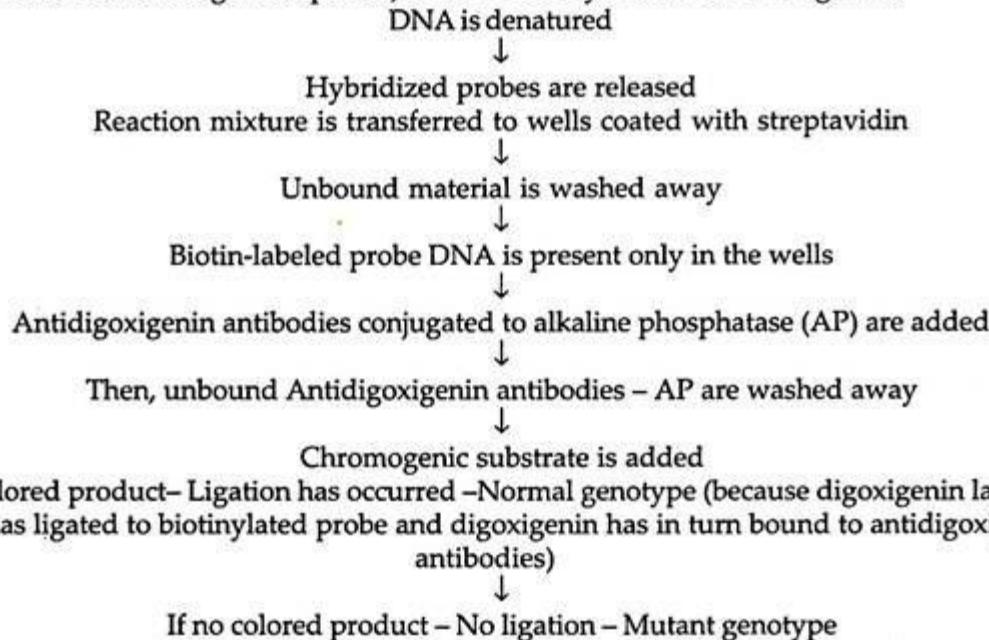
**C** Add ligase to hybridized DNA



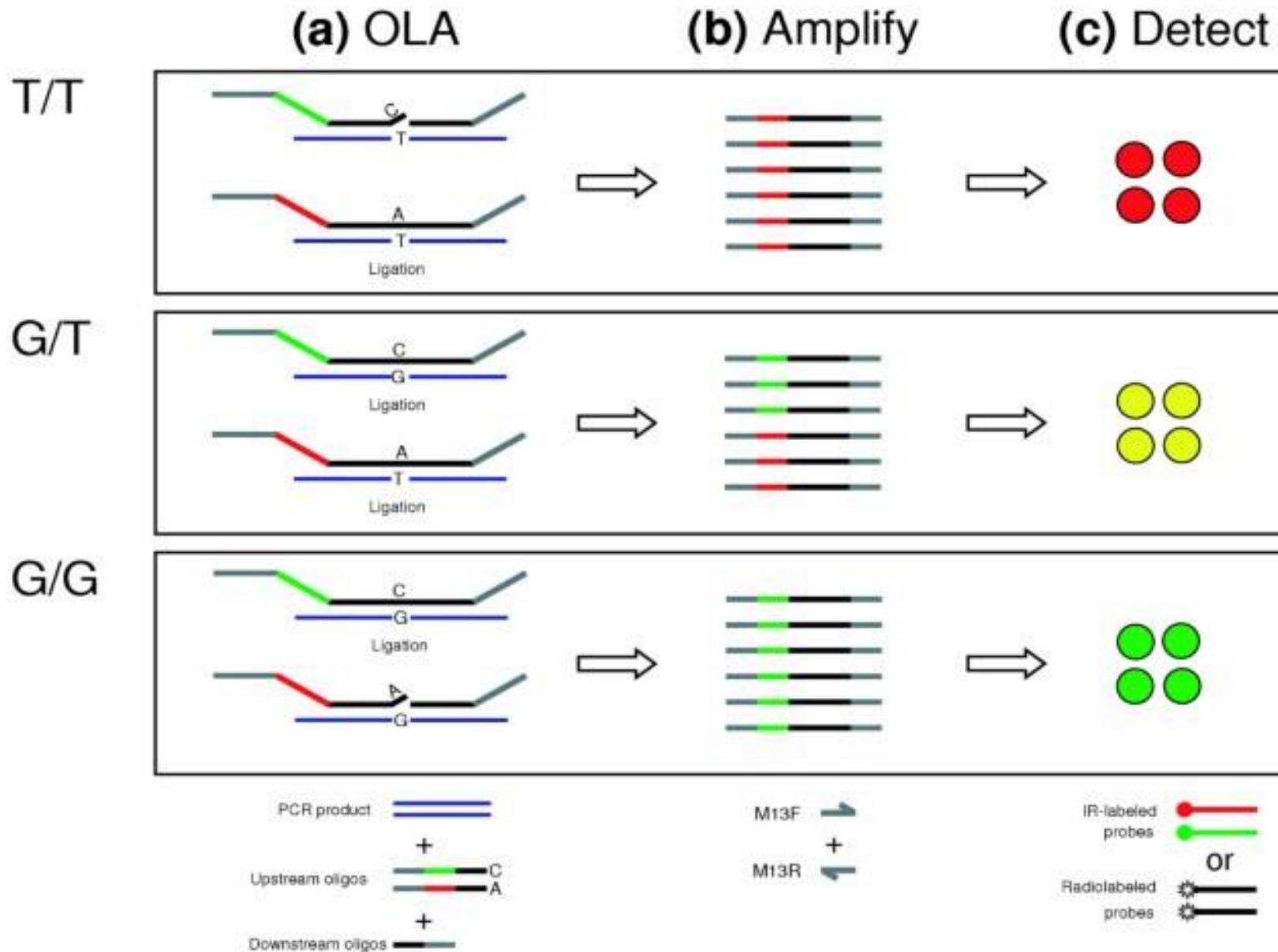
# Polymerase Chain Reaction-Oligonucleotide Ligation Assay (PCR-OLA) –For SNP detection



Now, we have to determine whether ligation has taken place between the two probes in both cases (normal and mutant gene sequence) or not. After hybridization and ligation,

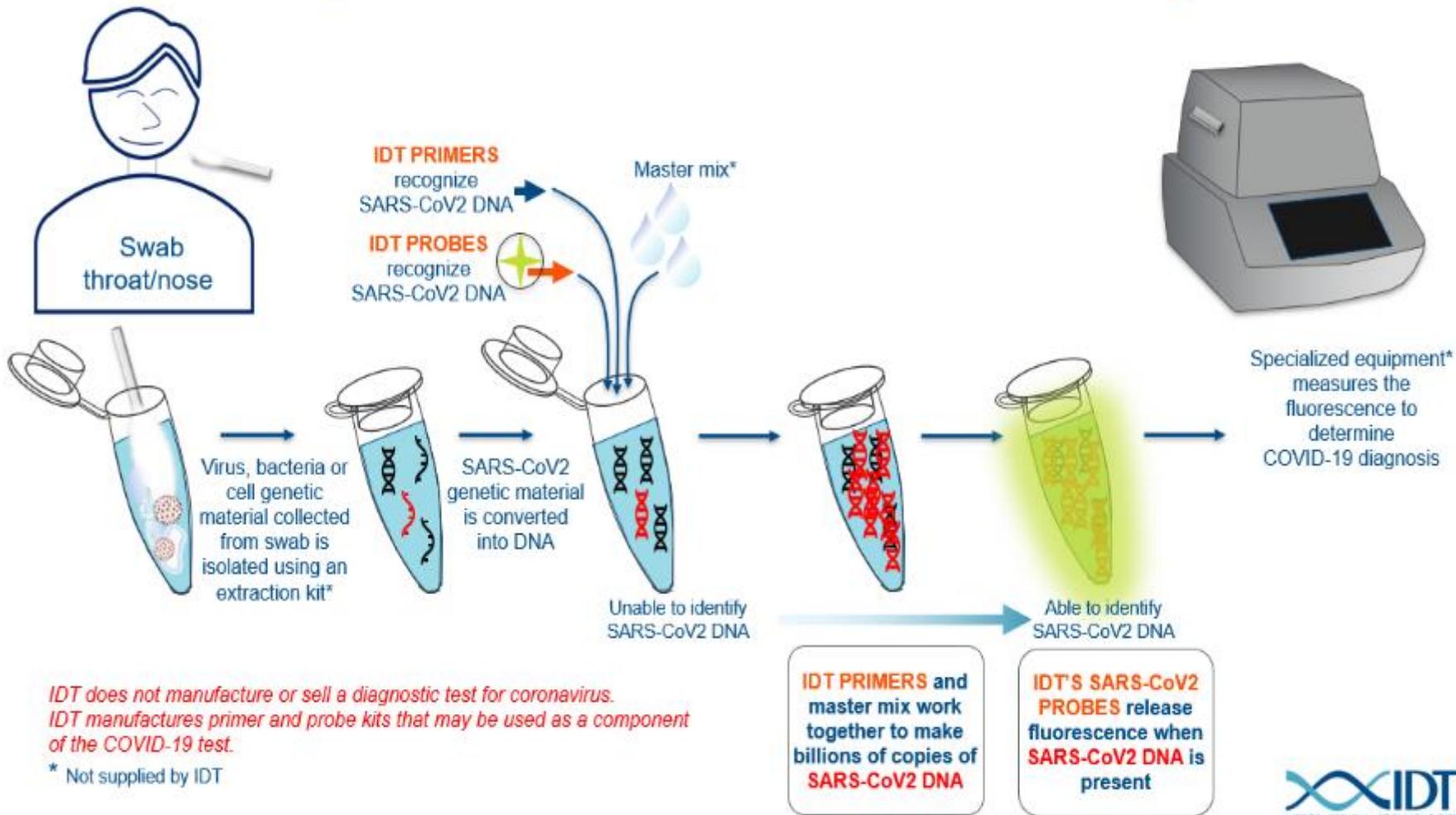


# Genotyping with Fluorescent Labelled Primers



# RT-PCR- Covid testing

## Using IDT Primers and Probes for COVID-19 Testing



# DNA fingerprinting

## STR Analysis

- Short tandem repeats consists of short sequences of nucleotides that are repeated many times tandemly.

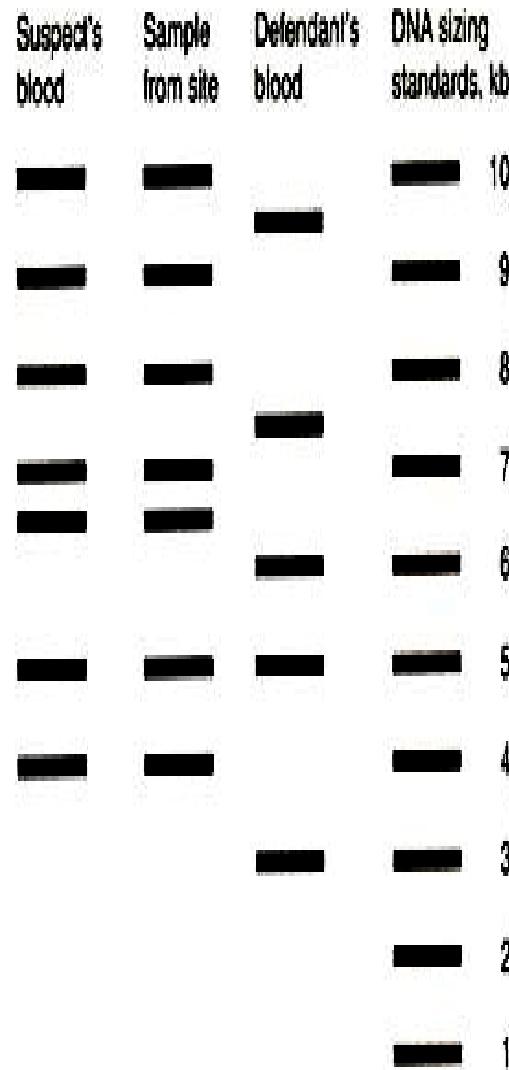
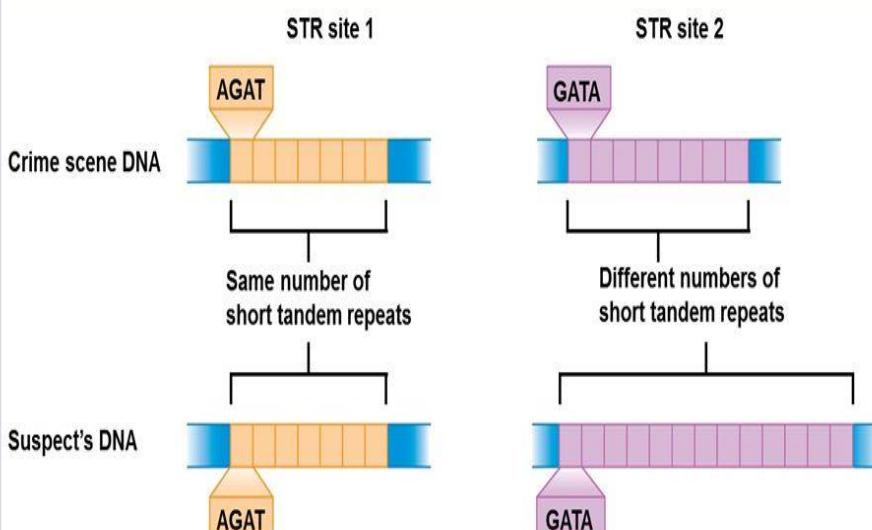
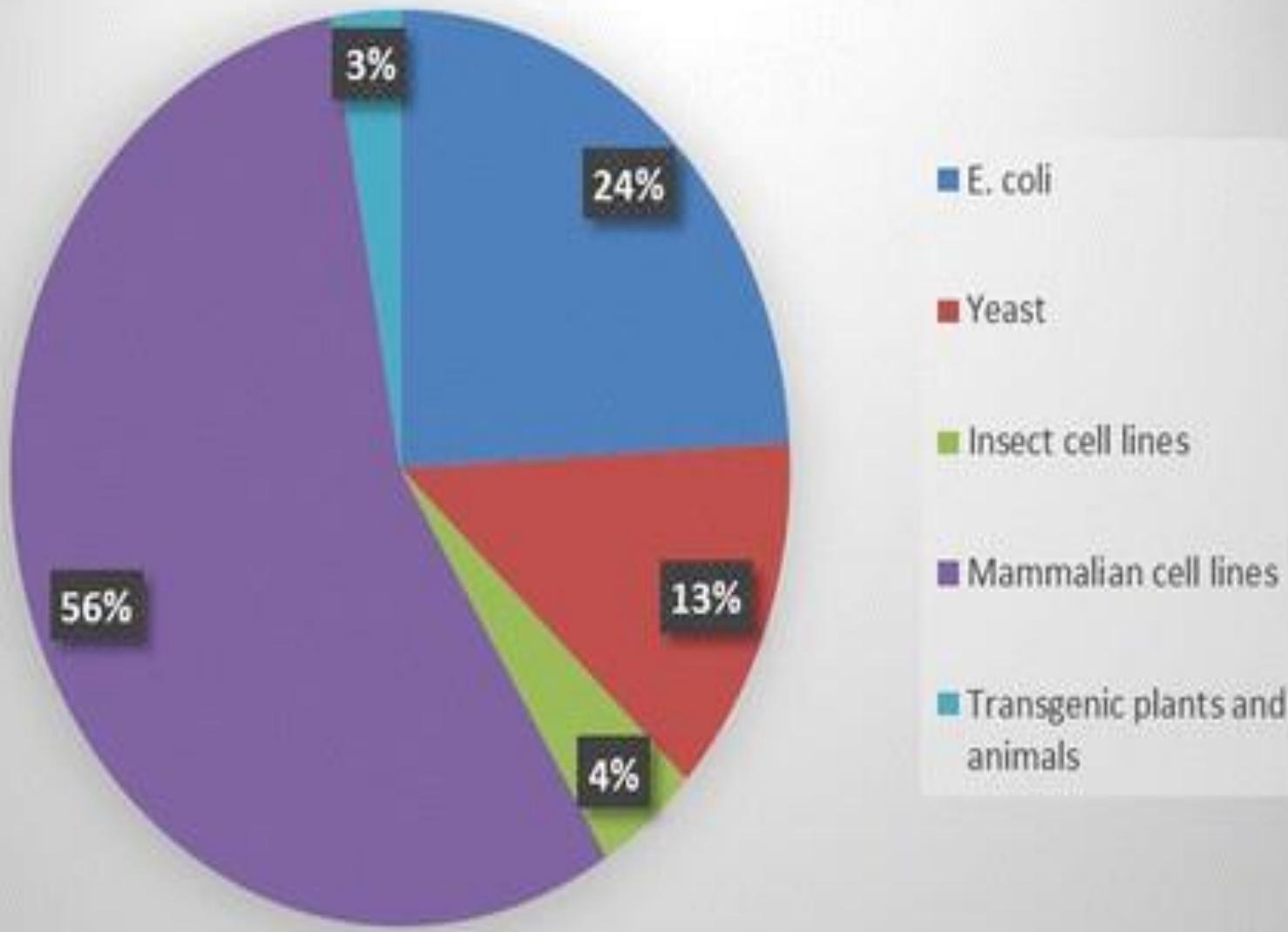


Fig. 19.4. Southern blot of a forensic DNA sample. Here the banding pattern of the suspected person's blood matches with the blood sample collected from the site of crime, thus confirming the victim.

# Production of recombinant pharmaceuticals

- A number of human disorders can be traced to the absence or malfunction of a protein normally synthesized in the body.
- Administering the human protein, then obtaining sufficient quantities will be a major problem unless donated blood can be used as the source.
- Animal proteins are therefore used whenever these are effective, but there are not many disorders that can be treated with animal proteins, and there is always the possibility of side effects such as an allergenic response.
- Several success stories--

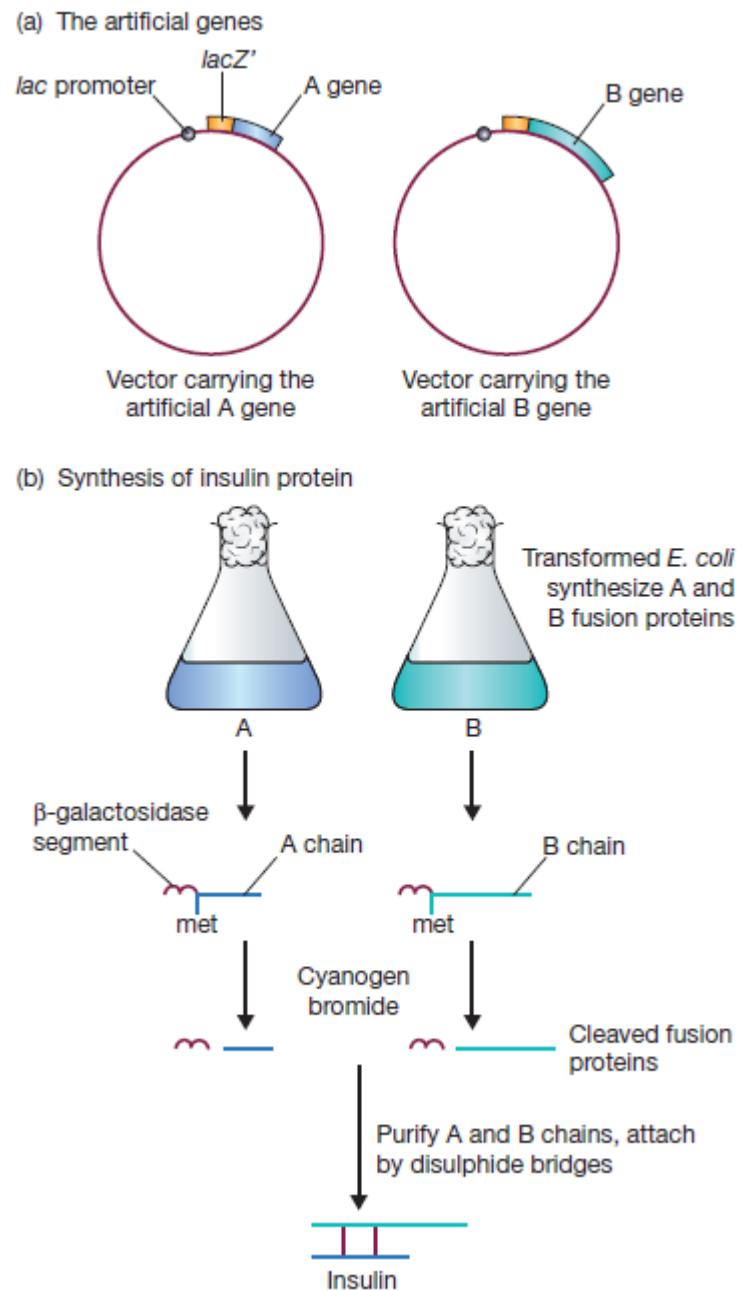
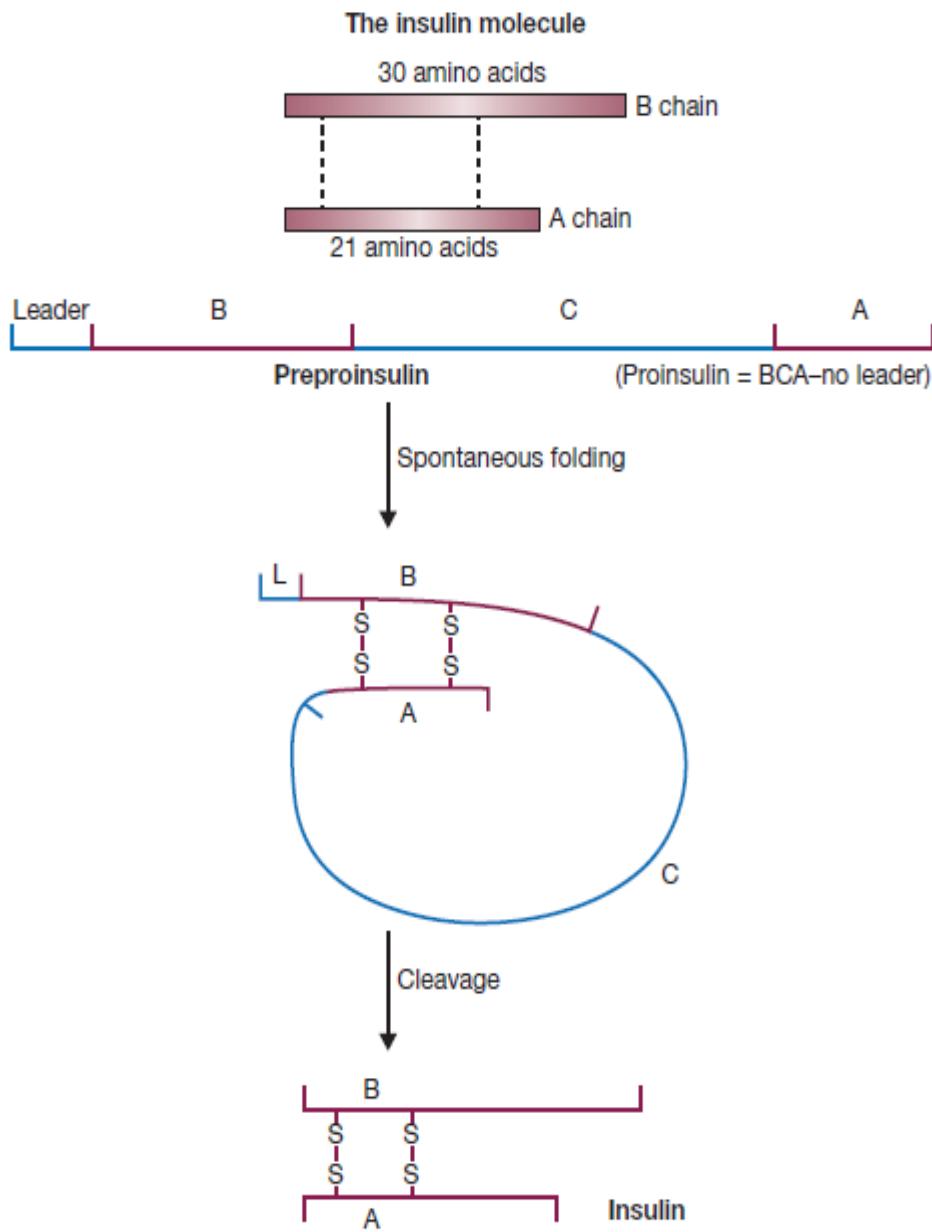


# Success Stories

Some of the human proteins that have been synthesized from genes cloned in bacteria and/or eukaryotic cells or by pharming.

PROTEIN	USED IN THE TREATMENT OF
$\alpha_1$ -Antitrypsin	Emphysema
Deoxyribonuclease	Cystic fibrosis
Epidermal growth factor	Ulcers
Erythropoietin	Anemia
Factor VIII	Hemophilia
Factor IX	Christmas disease
Fibroblast growth factor	Ulcers
Follicle stimulating hormone	Infertility treatment
Granulocyte colony stimulating factor	Cancers
Insulin	Diabetes
Insulin-like growth factor 1	Growth disorders
Interferon- $\alpha$	Leukemia and other cancers
Interferon- $\beta$	Cancers, AIDS
Interferon- $\gamma$	Cancers, rheumatoid arthritis
Interleukins	Cancers, immune disorders
Lung surfactant protein	Respiratory distress
Relaxin	Used to aid childbirth
Serum albumin	Used as a plasma supplement
Somatostatin	Growth disorders
Somatotrophin	Growth disorders
Superoxide dismutase	Free radical damage in kidney transplants
Tissue plasminogen activator	Heart attack
Tumor necrosis factor	Cancers

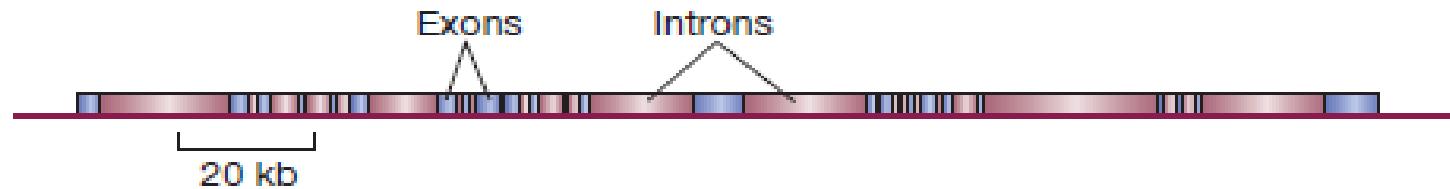
# Recombinant insulin



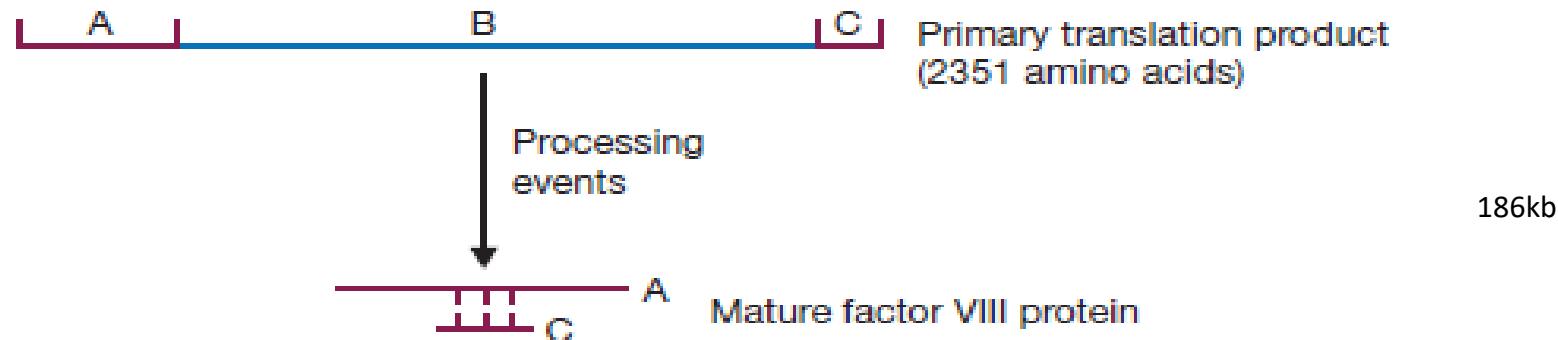
# *Recombinant factor VIII- Blood clotting factor*

Haemophilia treatment- Injection of purified factor VIII protein, obtained from human blood provided by donors. Purification of factor VIII is a complex procedure, treatment is expensive, Hepatitis and acquired immune deficiency syndrome (AIDS) can be and have been passed on to hemophiliacs via factor VIII injections..

## (a) The factor VIII gene



## (b) Post-translational processing of factor VIII



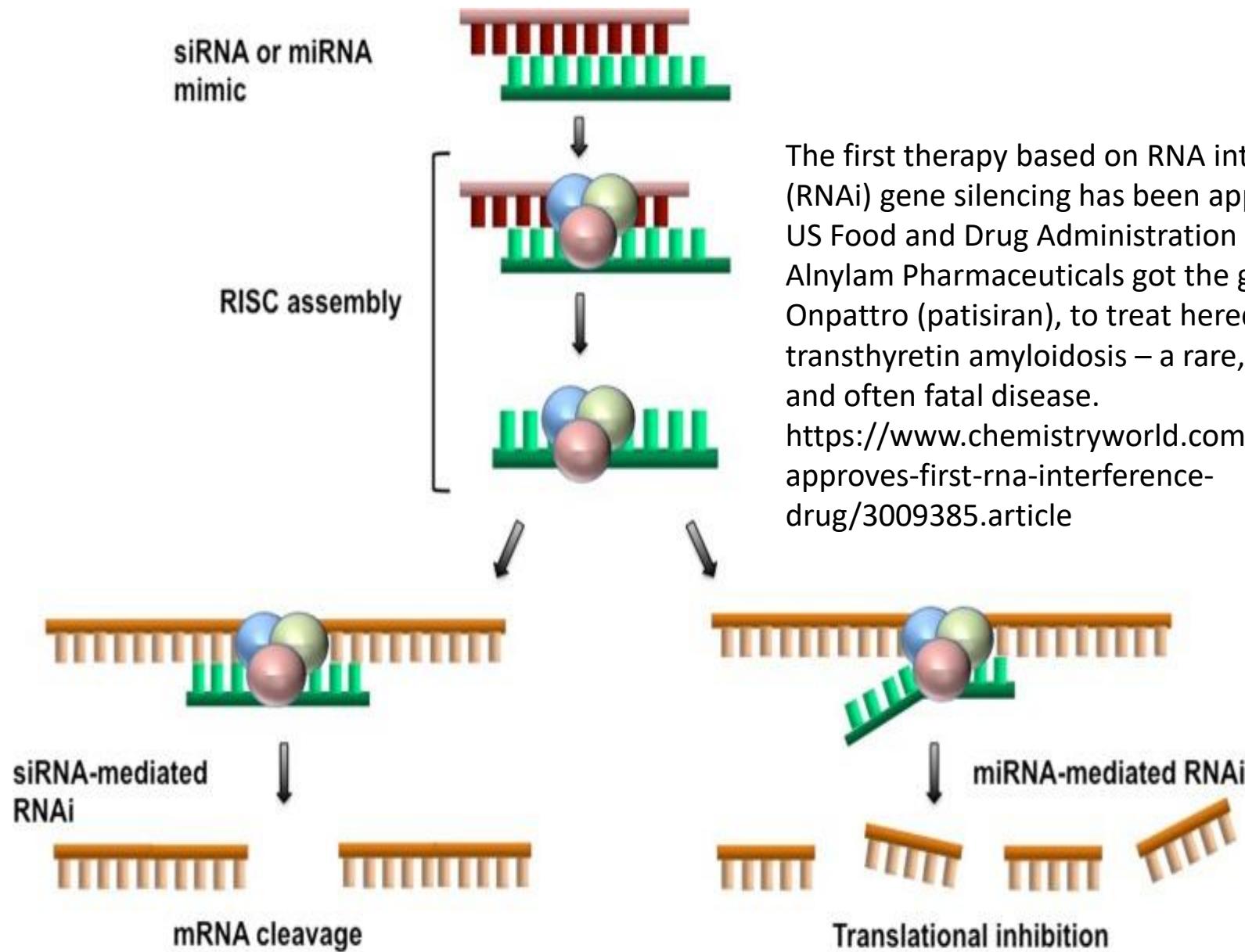
17 disulfide bonds and number of glycosylated sites

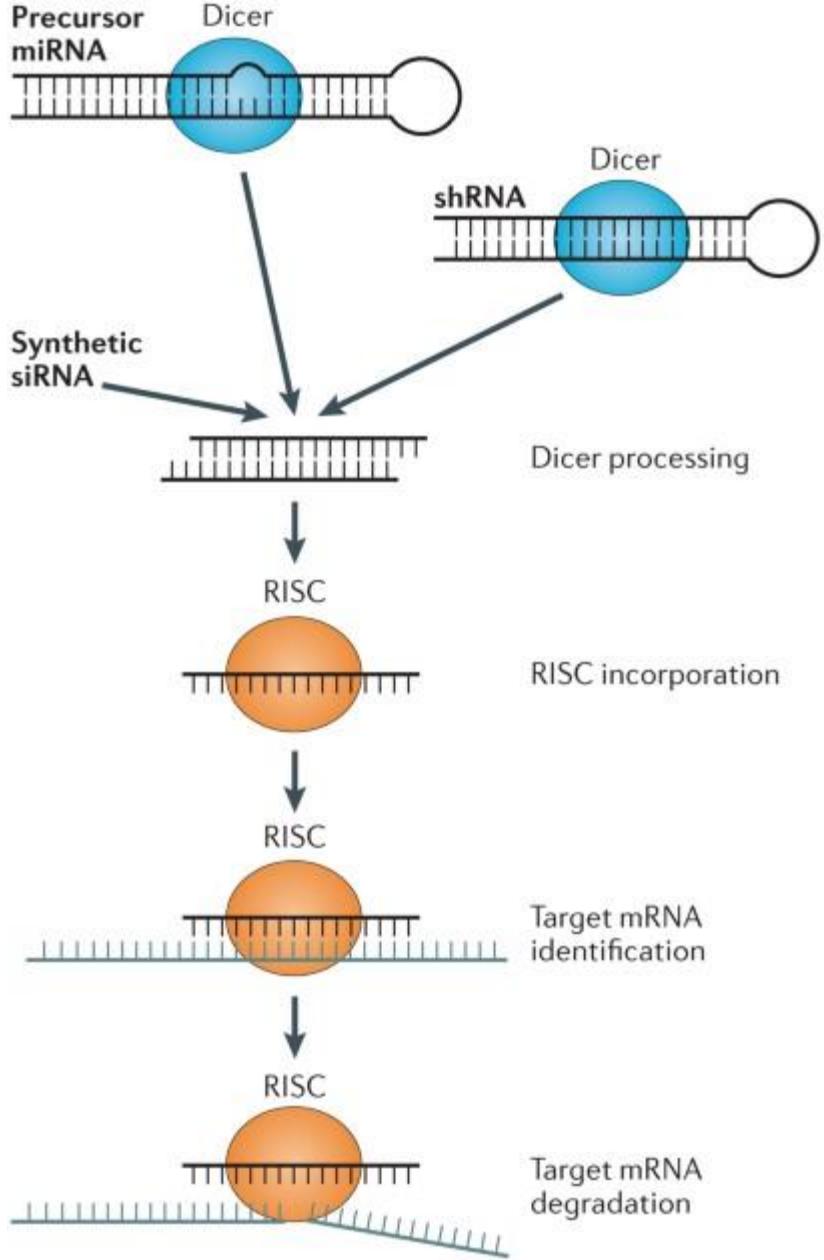
# Recombinant Factor VIII

- Initial attempts to obtain recombinant factor VIII therefore involved mammalian cells. In the first experiments to be carried out the entire cDNA was cloned in hamster cells, but yields of protein were disappointingly low.
- Each cDNA fragment was ligated into an expression vector, downstream of the Ag promoter (a hybrid between the chicken b-actin and rabbit b-globin sequences) and upstream of a polyadenylation signal from SV40 virus. The plasmid was introduced into a hamster cell line and recombinant protein obtained. The yields were over ten times greater
- 



# Gene Inhibition Approach





## ***Pharming—recombinant protein from live animals and plants***

- The use of silkworms for recombinant protein production is an example of the process often referred to as **pharming**, where a **transgenic** organism acts as the host for protein synthesis. Pharming is a recent and controversial innovation in gene cloning.
- This is a lengthy procedure and transgenic animals are therefore expensive to produce, but the technology is cost-effective because once a transgenic animal has been made it can reproduce and pass its cloned gene to its offspring according to standard Mendelian principles.

- Although proteins have been produced in the **blood** of transgenic animals, and in the **eggs** of transgenic **chickens**, the most successful approach has been to use **farm animals** such as **sheep or pigs**, with the **cloned gene attached to the promoter for the animal's b-lactoglobulin gene**. This promoter is active in the mammary tissue which means that the recombinant protein is secreted in the milk.
- For example, the average cow produces some 8000 liters of milk per year, yielding 40–80 kg of protein. Because the protein is secreted, purification is relatively easy.

# Human factor VIII- Pharming

- The complete human cDNA has been attached to the **promoter for the whey acidic protein gene** of pig, leading to synthesis of human factor VIII in **pig mammary tissue and subsequent secretion of the protein in the milk**. The factor VIII produced in this way appears to be exactly the same as the native protein and is fully functional in blood clotting assays.
-

## Tracy (1990-1997): Transgenic Ewe

- Genetically modified so that her milk produced a human protein called alpha antitrypsin, a potential treatment for the disease cystic fibrosis.

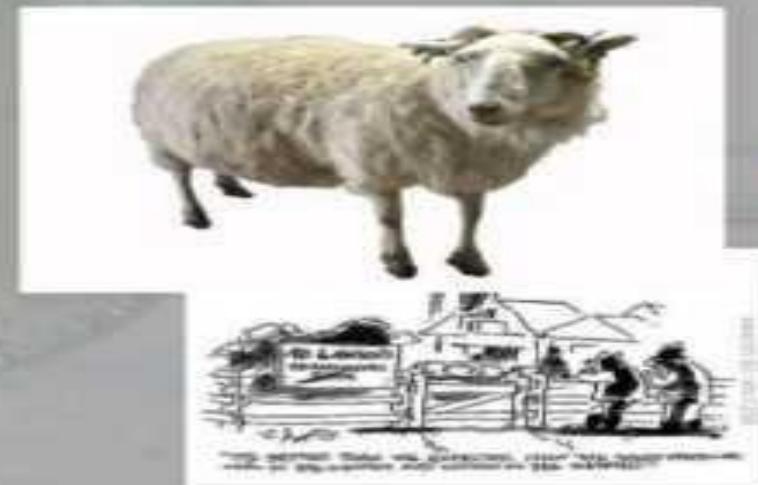


Plate 1: Tracy, the first mammal engineered successfully for Gene Pharming

Source: [www.slideshare.net](http://www.slideshare.net)

Alpha 1-antitrypsin comprised 50% of the total protein in Tracy's milk, a remarkably high level maintained after lactation. Similar levels were detected in the milk produced by her granddaughters.[3] A deficiency in this protein in humans can produce lung diseases, and its artificial creation was thought to be a potential success in the diseases' treatment.[2] Clinical trials for the engineered protein in 1998 revealed that it developed breathing problems in patients, and research for the milk as a remedy for the diseases has not continued since then.[4]

# ATryn



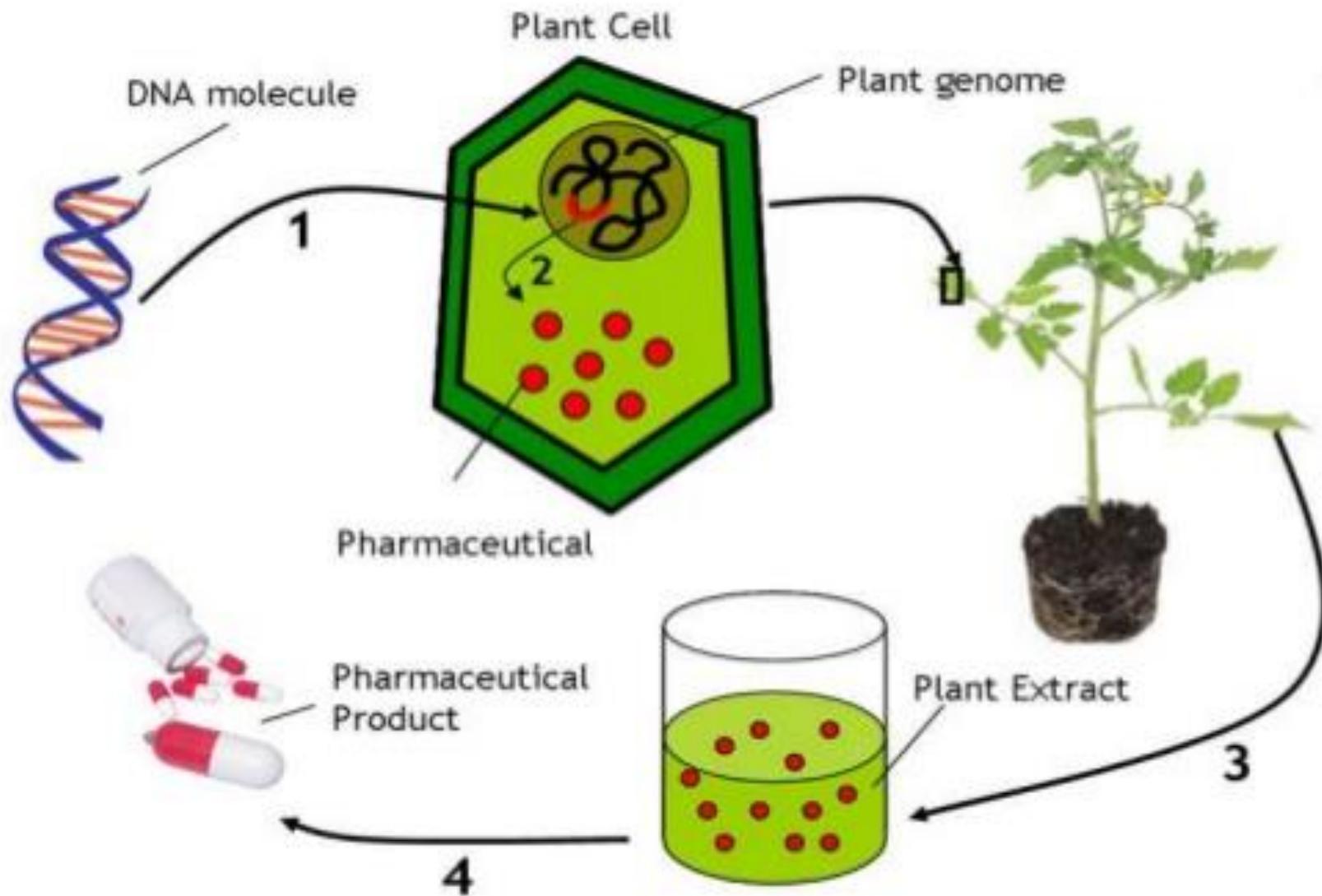
[https://www.youtube.com/watch?v=aH\\_SvupfkgU](https://www.youtube.com/watch?v=aH_SvupfkgU)

In February 2009 the US FDA granted marketing approval for the first drug to be produced in genetically modified livestock.<sup>[18]</sup> The drug is called ATryn, which is antithrombin protein purified from the milk of genetically modified goats. Marketing permission was granted by the European Medicines Agency in August 2006.<sup>[19]</sup>

# *Recombinant proteins from plants*

- Plants provide the final possibility for production of recombinant protein. **Plants and animals have similar protein processing activities**, although there are slight differences in the glycosylation pathways.
- **Plant cell culture is a well established technology** that is already used in the commercial synthesis of natural plant products.
- Alternatively, **intact plants can be grown** to a high density in fields. The latter approach to recombinant protein production has been used with a variety of crops, such as maize, tobacco, rice, and sugarcane.

# Plant Pharming



# Recombinant Vaccines

- Problems with attenuated viral vaccines:
  - The inactivation process must be 100% efficient, as the presence in a vaccine of just one live virus particle could result in infection.
  - The large amounts of virus particles needed for vaccine production are usually obtained from tissue cultures. Unfortunately some viruses, notably hepatitis B virus, do not grow in tissue culture.

# Recombinant vaccines

- The greatest success with this approach has been with hepatitis B virus.
- A person who recovers from hepatitis B is immune to future infection because their blood contains antibodies to the **hepatitis B surface antigen (HBsAg)**, which is one of the virus coat proteins.
- This protein has been synthesized in both ***Saccharomyces cerevisiae*, using a vector based on the 2 um plasmid, and in Chinese hamster ovary (CHO) cells.**
- In both cases, the protein was obtained in reasonably high quantities, and when injected into test animals provided protection against hepatitis B.

# *Recombinant vaccines in transgenic plants*

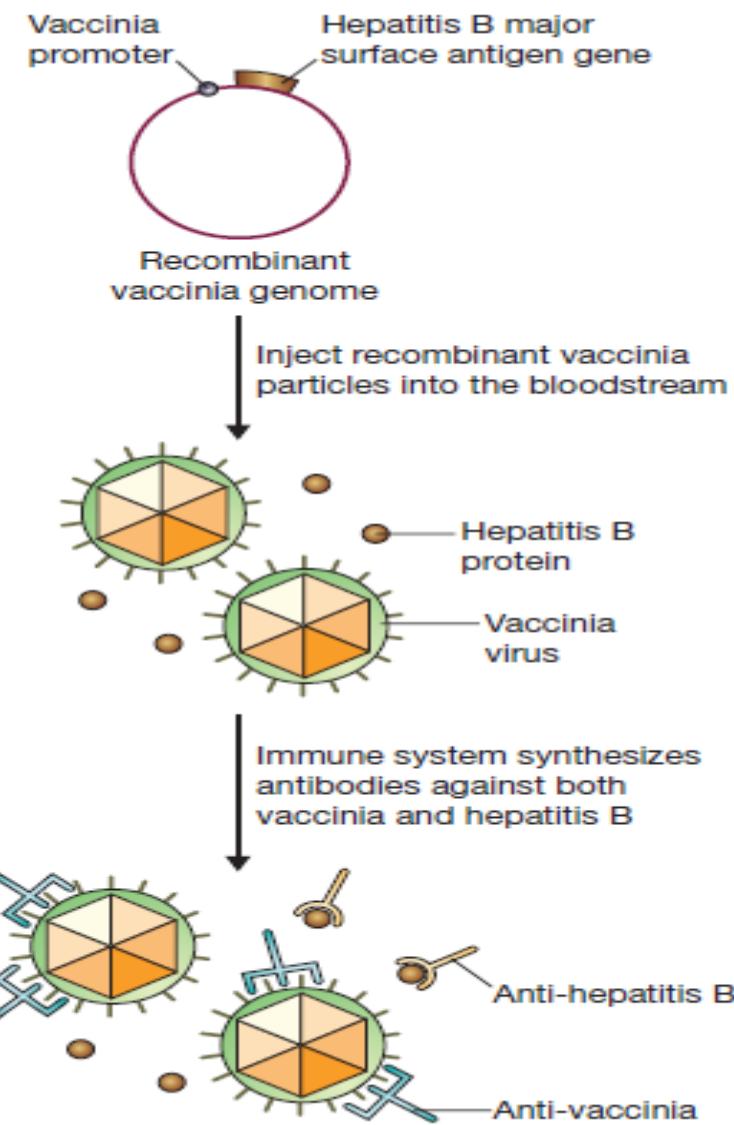
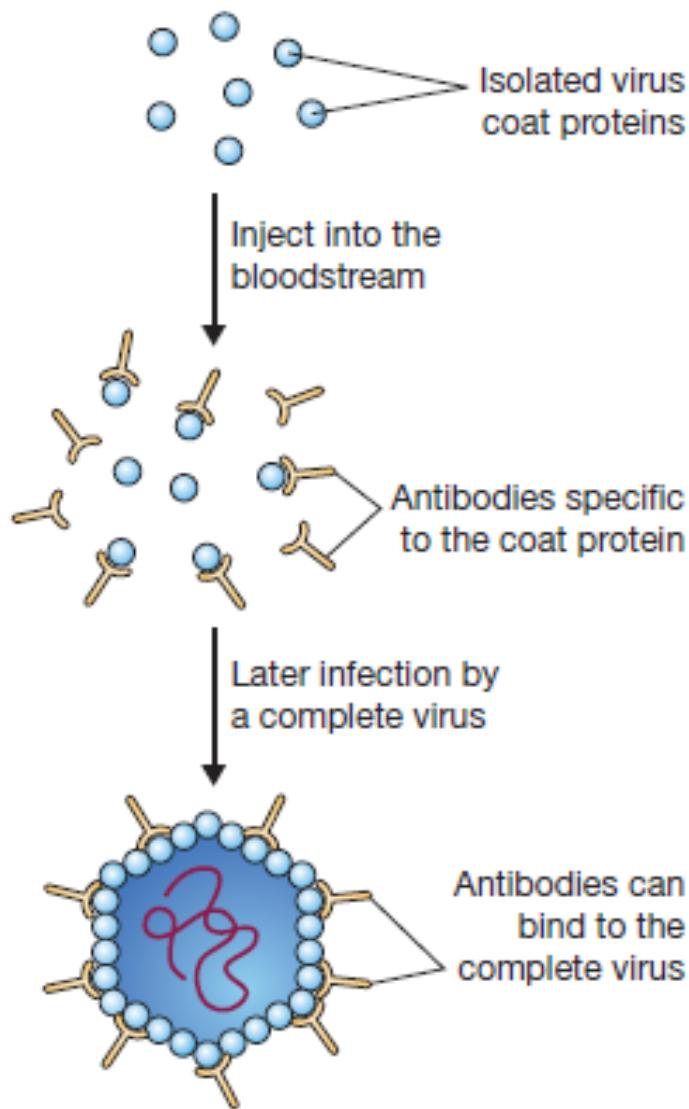
- If the recombinant vaccine is effective after oral administration, then **immunity could be acquired simply by eating part or all of the transgenic plant.** A simpler and cheaper means of carrying out a mass vaccination program.
- The feasibility of this approach has been demonstrated by trials with vaccines such as HbsAg and the coat proteins of measles virus and respiratory syncytial virus.

In 2006 [Dow AgroSciences](#) received USDA approval to market a vaccine for poultry against [Newcastle disease](#), produced in plant cell culture – the first plant-produced vaccine approved in the U.S

# *Recombinant vaccines in transgenic plants*

- The main problem currently faced by the companies developing this technology is
  - that the **amount** of recombinant protein synthesized by the plant is **often insufficient** to stimulate complete immunity against the target disease. To be completely effective the yield of the vaccine needs to make up 8–10% of the soluble protein content of the part of the plant which is eaten, but in practice yields are much less than this, usually not more than 0.5%. **Variability in the yields** between different plants in a single crop is also a concern.
  - A partial solution is provided by **placing the cloned gene in the chloroplast genome rather than the plant nucleus**, as this generally results in much higher yields of recombinant protein. However, **proteins made in the chloroplast are not glycosylated** and so those vaccines that require post-translational modification will be inactive if produced in this way.
  - *Vibrio cholerae* B subunit, which can be used to confer immunity against diseases such as cholera. This protein has been synthesized in transgenic tobacco, tomato, and rice plants and shown to elicit an anti-cholera immune response when leaves, fruits, or seeds are fed to mice.

# Recombinant Vaccines- Use of vaccinia virus



Some of the foreign genes that have been expressed in recombinant vaccinia viruses.

GENE
<i>Plasmodium falciparum</i> (malaria parasite) surface antigen
Influenza virus coat proteins
Rabies virus G protein
Hepatitis B surface antigen
Herpes simplex glycoproteins
Human immunodeficiency virus (HIV) envelope proteins
Vesicular somatitis coat proteins
Sindbis virus proteins

# Gene Therapy

- This is the name originally given to methods that aim to cure an inherited disease by providing the patient with a correct copy of the defective gene.
- There are two basic approaches to gene therapy:
  - **germline therapy and somatic cell therapy.**
  - In germline therapy, a fertilized egg is provided with a copy of the correct version of the relevant gene and re-implanted into the mother. If successful, the gene is present and expressed in all cells of the resulting individual.

Because gene therapy involves making changes to the body's set of basic instructions, it raises many unique ethical concerns. The ethical questions surrounding gene therapy include:

- How can “good” and “bad” uses of gene therapy be distinguished?
- Who decides which traits are normal and which constitute a disability or disorder?
- Will the high costs of gene therapy make it available only to the wealthy?
- Could the widespread use of gene therapy make society less accepting of people who are different?
- Should people be allowed to use gene therapy to enhance basic human traits such as beauty, intelligence, or athletic ability?

# Gene Therapy- Ethical Issues

The idea of germline gene therapy is controversial. While it could spare future generations in a family from having a particular genetic disorder, it might affect the development of a fetus in unexpected ways or have long-term side effects that are not yet known. Because people who would be affected by germline gene therapy are not yet born, they can't choose whether to have the treatment. Because of these ethical concerns, the U.S. Government does not allow federal funds to be used for research on germline gene therapy in people.

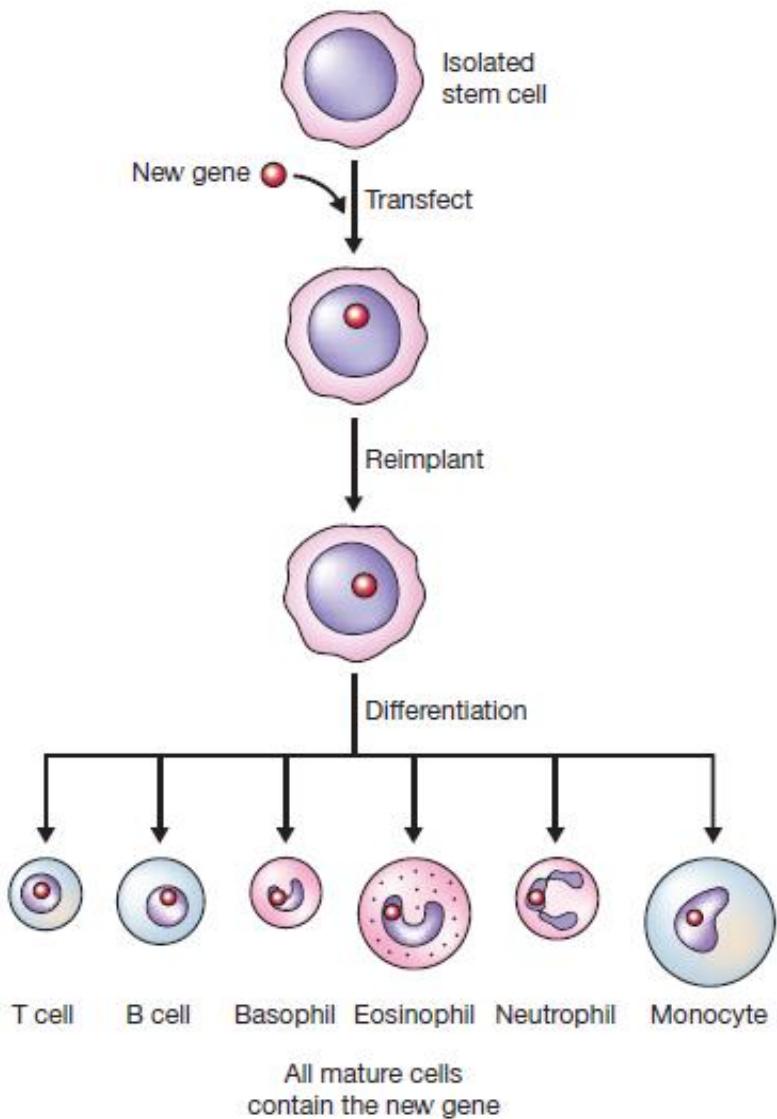
# Gene Therapy

Thus, the current gene therapy research has focused on treating individuals by targeting the therapy to body cells such as bone marrow or blood cells. This type of gene therapy cannot be passed to a person's children.

Somatic cell therapy involves manipulation of cells, which either can be removed from the organism, transfected, and then placed back in the body, or transfected *in situ* without removal.

The vector can be injected or given **intravenously (by IV) directly into a specific tissue** in the body, where it is taken up by individual cells. **Alternately, a sample of the patient's cells can be removed and exposed to the vector in a laboratory setting. The cells containing the vector are then returned to the patient.** If the treatment is successful, the new gene delivered by the vector will make a functioning protein.

# Gene Therapy- Somatic Cell Therapy



- The technique has **most promise for inherited blood diseases (e.g., hemophilia and thalassaemia)**, with genes being introduced into stem cells from the bone marrow, which give rise to all the specialized cell types in the blood.
- Somatic cell therapy **also has potential** in the treatment of **lung diseases such as cystic fibrosis**, as DNA cloned in adenovirus vectors or contained in liposomes is taken up by the epithelial cells in the lungs after introduction into the respiratory tract via an inhaler. However, this has not yet been developed into an effective means of treating cystic fibrosis.

# Applications of RDT in Agriculture

# Agricultural Biotechnology

- **Genetically modified organisms (GMO)** with inserted genes.
- Genetically Modified plants can be **resistant to disease, stress, frost, pest, herbicide**
- GM can be a **factory for pharmaceuticals**: tobacco plant that produces Hemoglobin
- Plants that yield a **healthier and higher than normal crop** to improve our food supply
- **Improved shelf life,**
- production of useful goods such as **biofuel or drugs**
- ability to **absorb toxins** and for use in **bioremediation** of pollution

➤ **Insect Resistance- Bt Crops**

➤ **Herbicide Resistant**

➤ **Other Examples- Flavr Savr Tomato. Golden Rice**

➤ **Terminator Technology**

# Insect Resistant Plants

# The need for Insect Resistant Plants

To reduce losses, crops are regularly sprayed with insecticides. Most conventional insecticides (e.g., pyrethroids and organophosphates) are

1. **relatively non-specific poisons** that kill a broad spectrum of insects, not just the ones eating the crop.
2. Because of their **high toxicity**, several of these insecticides also have potentially harmful side effects for other members of the local biosphere, including in some cases humans.
3. **movement of the chemicals in the ecosystem** cannot be controlled.
4. Furthermore, **insects that live within the plant**, or on the undersurfaces of leaves, can sometimes avoid the toxic effects altogether.

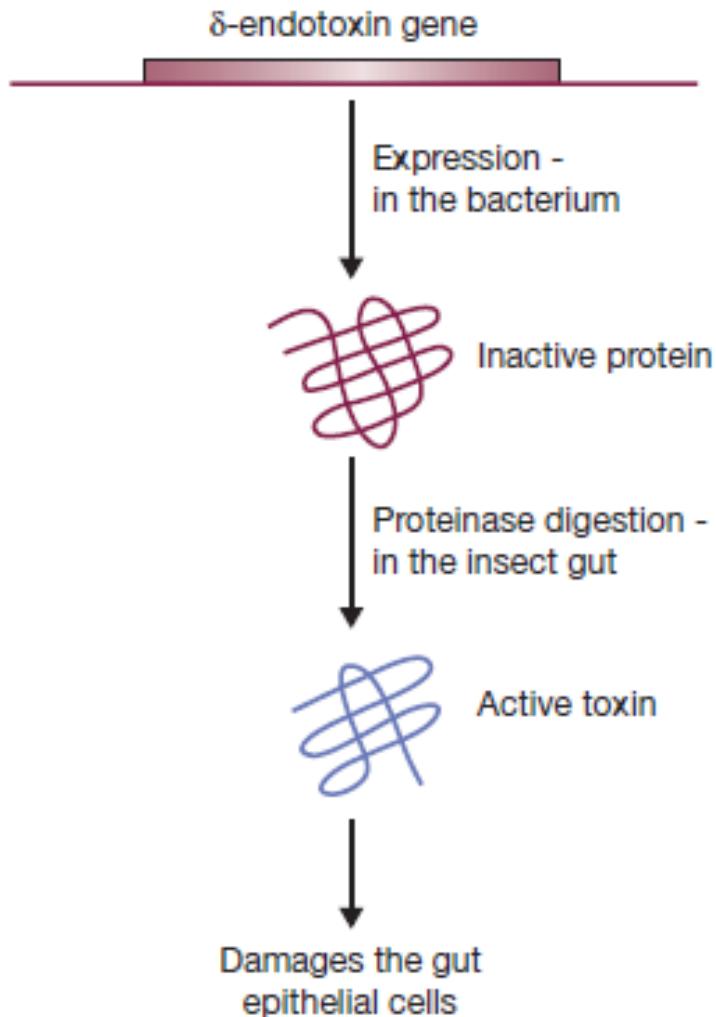
## Ideal insecticide-

1. **Highly selective**, so that the insecticide is harmless to other insects and is not poisonous to animals and to humans.
2. The insecticide should **be biodegradable**, so that any residues that remain after the crop is harvested, or which are carried out of the field by rainwater, do not persist long enough to damage the environment.
3. And it should be possible to apply the insecticide in such a way that **all parts of the crop**, not just the upper surfaces of the plants, are protected against insect attack.

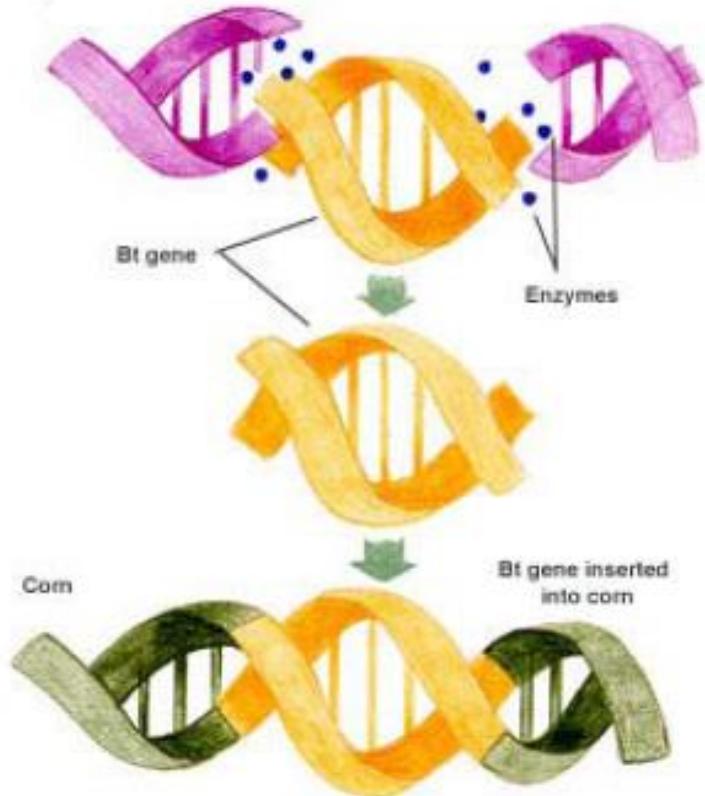
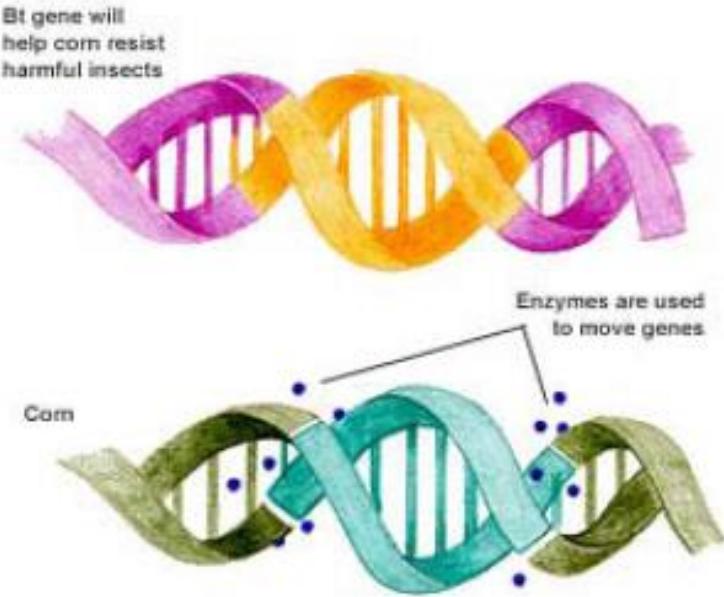
The ideal insecticide has not yet been discovered. The closest we have are the delta-endotoxins produced by the soil bacterium ***Bacillus thuringiensis***.

Several types of bacteria have evolved defense mechanisms against insect predation, an example being *B. thuringiensis* which, during sporulation, forms intracellular crystalline bodies that contain an insecticidal protein called the  **$\delta$ -endotoxin**.

The activated protein is highly poisonous to insects, some 80,000 times more toxic than organophosphate insecticides, and is relatively selective, different strains of the bacterium synthesizing proteins effective against the larvae of different groups of insects.



Transfer of Bt genes from *Bacillus thuringiensis* which code for proteins that are converted into toxins in the gut tracts of insects.



When spliced into the genome of corn, it is hoped that these genes will produce the proteins which will convey insect resistance to this particular strain of corn.

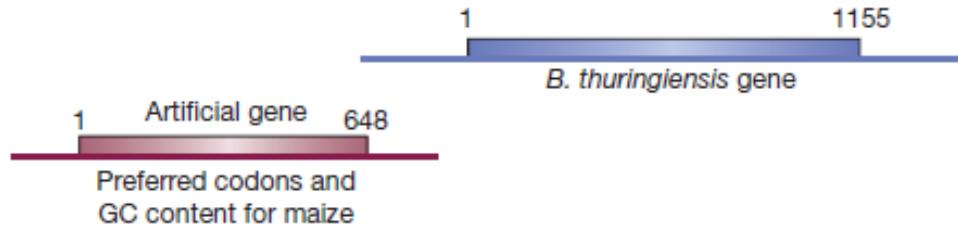
# *Cloning a endotoxin gene in maize*

- Maize is an example of a crop plant that is not served well by conventional insecticides.
- A major pest is the European corn borer (*Ostrinia nubilalis*), which tunnels into the plant from eggs laid on the undersurfaces of leaves, thereby evading the effects of insecticides applied by spraying.

# Cloning a $\delta$ -endotoxin gene in maize

Toxic activity is from 29-607 aa.(Cry1ab gene)

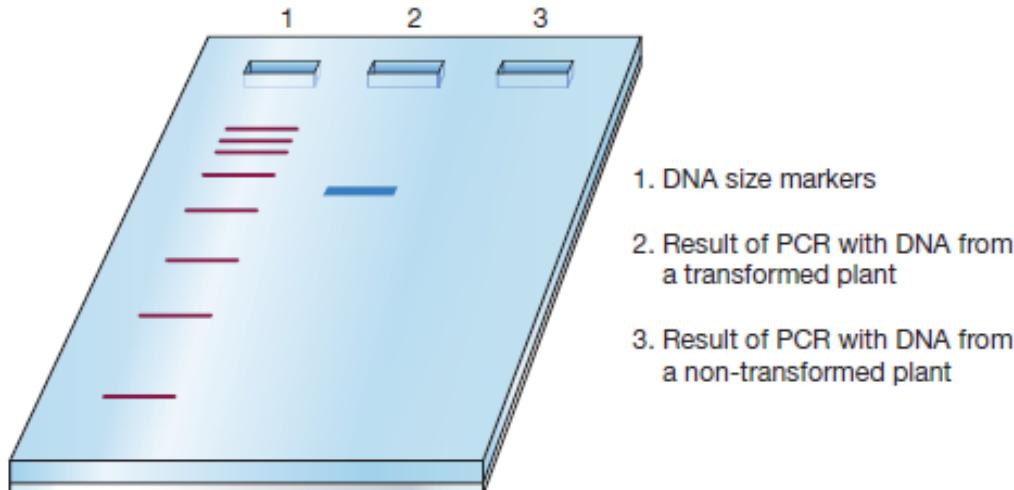
(a) Synthesis of an artificial  $\delta$ -endotoxin gene



(b) Attachment of a promoter and polyadenylation signal



(c) PCR analysis of mature plants



Cassette Vector introduced into maize embryos by bombardment with DNA-coated microprojectiles. The embryos were grown into mature plants, and transformants identified by PCR analysis of DNA extracts, using primers specific for a segment of the artificial gene.

$\delta$ -endotoxin being produced varied from plant to plant, from about 250–1750 ng of toxin per mg of total protein, mainly due to positional effects

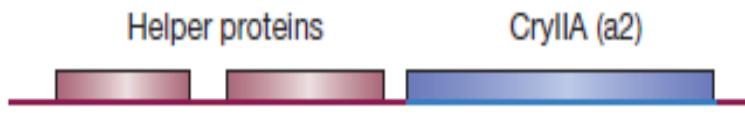
In particular, the average length of the larval tunnels was reduced from 40.7 cm for the controls to just 6.3 cm for the engineered plants.

# *Cloning δ-endotoxin genes in chloroplasts*

- A transgene located in **the chloroplast genome cannot escape via pollen** for the simple reason that pollen does not contain chloroplast.
- One advantage of using chloroplasts as the sites of recombinant protein synthesis is that the gene expression machinery of chloroplasts, being related to that of bacteria is able to **express all the genes in an operon**.
- In contrast, each gene that is placed in a plant (or animal) nuclear genome must be cloned individually, with its own promoter and other expression signals.
- The amounts of CryIIA(a2) protein produced in the tissues of these GM plants was quite remarkable, the toxin making up over 45% of the total soluble protein in tobacco, more than previously achieved in any plant cloning experiment.

**Figure 15.4**

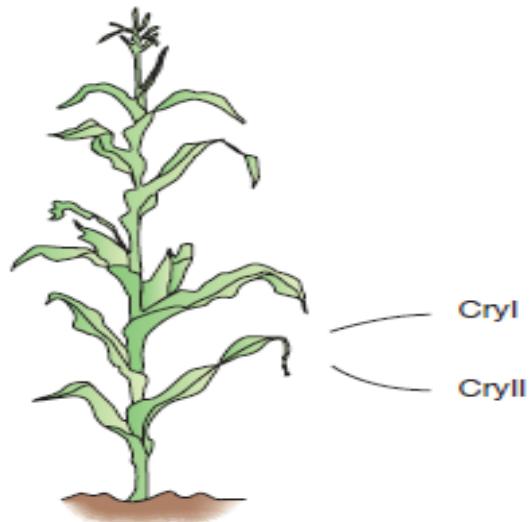
The CryIIA(a2) operon.



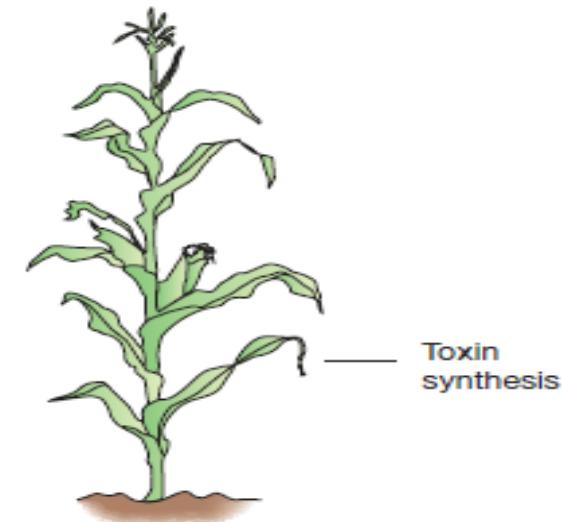
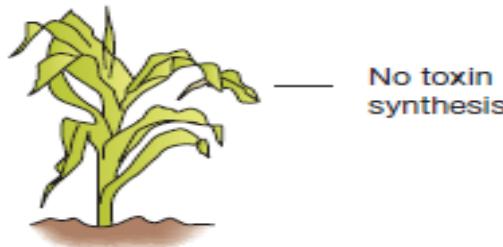
Five days after being placed on the GM plants, all cotton bollworm and beet armyworm larvae were dead, with appreciable damage being visible only on the leaves of the plants exposed to armyworms, which have a relatively high natural resistance to δ-endotoxins. Attempts to repeat this experiment with maize, cotton, and other more useful crops have been hampered by the difficulties in achieving chloroplast transformation with plants other than tobacco

# Countering insect resistance to $\delta$ -endotoxin crops

(a) Production of two toxins



(b) Temporal control over toxin synthesis



(c) Mixed planting

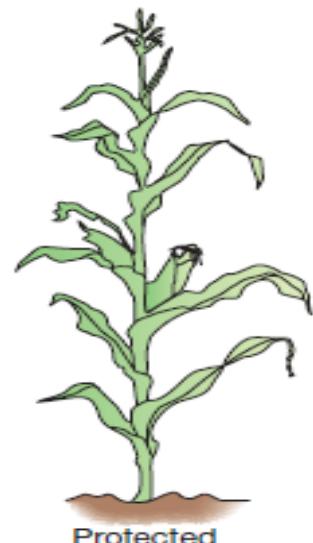


Figure 15.5

Three strategies for countering the development of insect resistance to  $\delta$ -endotoxin crops.

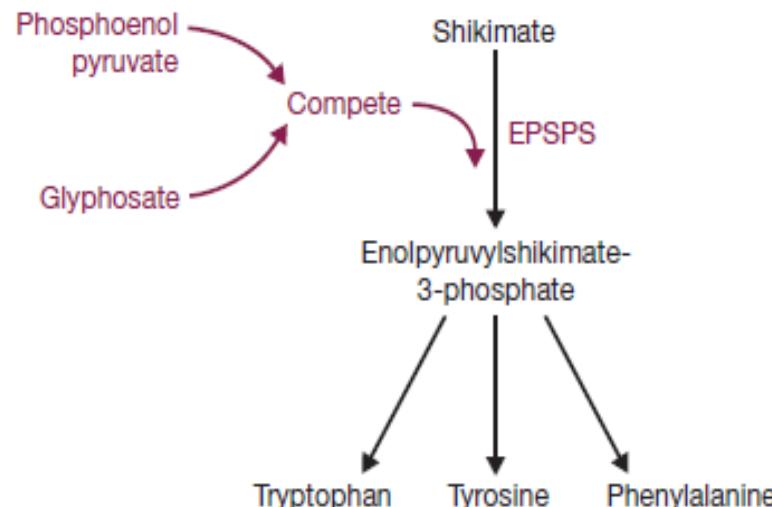
# Herbicide Resistant Plants

# Herbicide resistant crops (Roundup Ready Crops)

The first crops to be engineered for glyphosate resistance were produced by Monsanto Co. and called “Roundup Ready”, reflecting the trade name of the herbicide

**Figure 15.6**

Glyphosate competes with phosphoenol pyruvate in the EPSPS catalyzed synthesis of enolpyruvylshikimate-3-phosphate, and hence inhibits synthesis of tryptophan, tyrosine, and phenylalanine.



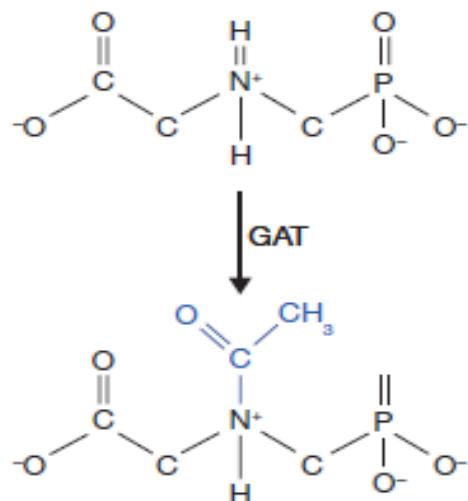
## Search for EPSPS enzyme showing resistance to glyphosate

After testing the genes from various bacteria, as well as mutant forms of *Petunia* that displayed glyphosate resistance, the EPSPS gene from *Agrobacterium* strain CP4 was chosen, because of its combination of high catalytic activity and high resistance to the herbicide.

EPSPS is located in the plant chloroplasts, so the *Agrobacterium* EPSPS gene was cloned in a Ti vector as a fusion protein with a leader sequence that would direct the enzyme across the chloroplast membrane and into the organelle. Biolistics was used to introduce the recombinant vector into soybean callus culture. After regeneration, the GM plants were found to have a threefold increase in herbicide resistance.

# A new generation of glyphosate resistant crops

(a) Detoxification of glyphosate by GAT



(b) Directed evolution to produce a highly active GAT enzyme

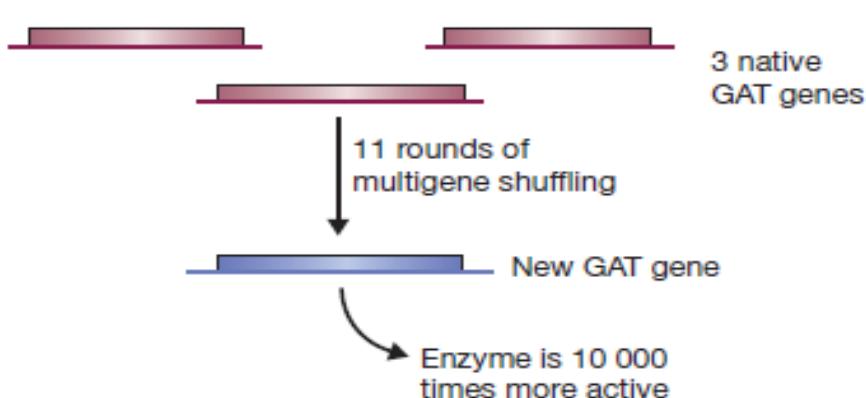


Figure 15.7

Use of glyphosate *N*-acetyltransferase to generate plants that detoxify glyphosate. (a) GAT detoxifies glyphosate by adding an acetyl group (shown in blue). (b) Creation of a highly active GAT enzyme by multigene shuffling.

The most active detoxifier known is a strain of *B. licheniformis*, but even this bacterium detoxifies glyphosate at rates that are too low to be of value if transferred to a GM crop.

After 11 rounds, a gene specifying a GAT with 10,000 times the activity of the enzymes present in the original *B. licheniformis* strain was obtained. This gene was introduced into maize, and the resulting GM plants were found to tolerate levels of glyphosate six times higher than the amount normally used by farmers to control weeds, without any reduction in the productivity of the plant.

Examples of gene addition projects with plants.

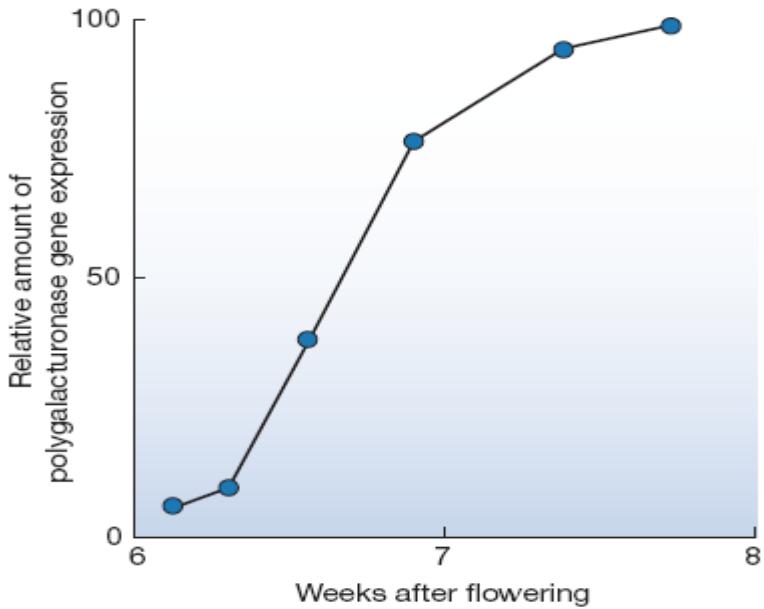
GENE FOR	SOURCE ORGANISM	CHARACTERISTIC CONFERRED ON MODIFIED PLANTS
δ-Endotoxin	<i>B. thuringiensis</i>	Insect resistance
Proteinase inhibitors	Various legumes	Insect resistance
Chitinase	Rice	Fungal resistance
Glucanase	Alfalfa	Fungal resistance
Ribosome-inactivating protein	Barley	Fungal resistance
Ornithine carbamyltransferase	<i>Pseudomonas syringae</i>	Bacterial resistance
RNA polymerase, helicase	Potato leafroll luteovirus	Virus resistance
Satellite RNAs	Various viruses	Virus resistance
Virus coat proteins	Various viruses	Virus resistance
2'-5'-Oligoadenylate synthetase	Rat	Virus resistance
Acetolactate synthase	<i>Nicotiana tabacum</i>	Herbicide resistance
Enolpyruvylshikimate-3-phosphate synthase	<i>Agrobacterium</i> spp.	Herbicide resistance
Glyphosate oxidoreductase	<i>Ochrobactrum anthropi</i>	Herbicide resistance
Glyphosate N-acetyltransferase	<i>B. licheniformis</i>	Herbicide resistance
Nitrilase	<i>Klebsiella ozaenae</i>	Herbicide resistance
Phosphinothricin acetyltransferase	<i>Streptomyces</i> spp.	Herbicide resistance
Phosphatidylinositol-specific phospholipase C	Maize	Drought tolerance
Barnase ribonuclease inhibitor	<i>Bacillus amyloliquefaciens</i>	Male sterility
DNA adenine methylase	<i>E. coli</i>	Male sterility
Methionine-rich protein	Brazil nuts	Improved sulphur content
1-Aminocyclopropane-1-carboxylic acid deaminase	Various	Modified fruit ripening
S-Adenosylmethionine hydrolase	Bacteriophage T3	Modified fruit ripening
Monellin	<i>Thaumatococcus danielli</i>	Sweetness
Thaumatin	<i>T. danielli</i>	Sweetness
Acyl carrier protein thioesterase	<i>Umbellularia californica</i>	Modified fat/oil content
Delta-12 desaturase	<i>Glycine max</i>	Modified fat/oil content
Dihydroflavanol reductase	Various flowering plants	Modified flower color
Flavonoid hydroxylase	Various flowering plants	Modified flower color

# *Gene Inhibition Approach*

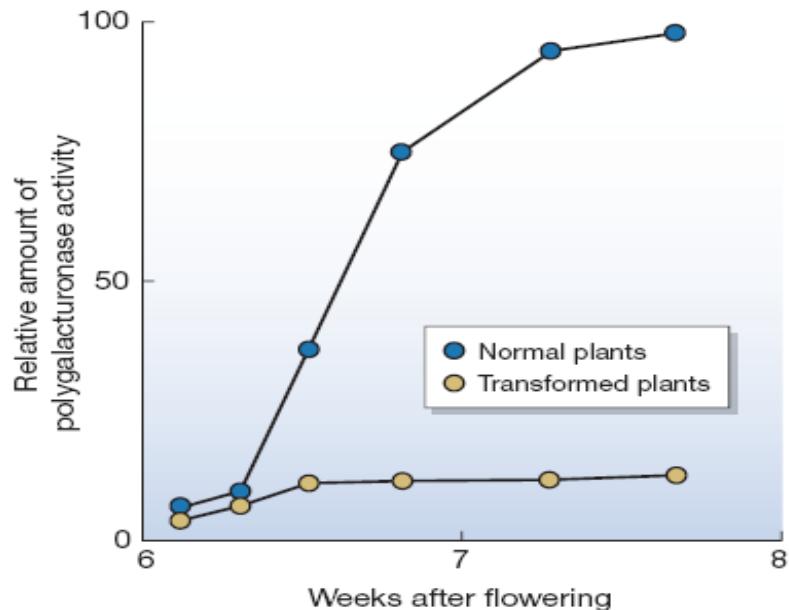
---

The Flavr Savr tomato is a genetically altered tomato developed by Calgene. It contains an antisense RNA which inhibits the expression of a gene that normally causes fruit to soften, therefore, the fruit stays firm longer. This allows producers a greater period of time for transportation and the opportunity for mechanical harvesting with little bruising.





The increase in polygalacturonase gene expression seen during the later stages of tomato fruit ripening.



### Polygalacturonase gene

Partial inactivation of the polygalacturonase gene increases the time between flavor development and spoilage of the fruit.

# Ethylene Inhibition

The penultimate step in the ethylene synthesis pathway is conversion of S-adenosyl-methionine to 1-aminocyclopropane-1-carboxylic acid (ACC), which is the immediate precursor for ethylene.

After regeneration, the engineered plants were grown to the fruiting stage and found to make only 2% of the amount of ethylene produced by non-engineered plants. This reduction was more than sufficient to prevent the fruit from completing the ripening process. These tomatoes have been marketed as the “Endless Summer” variety.

[DNA Plant Technology](#) (DNAP), [Agritope](#) and [Monsanto](#) developed tomatoes that delayed ripening by preventing the production of [ethylene](#),<sup>[8]</sup> a [hormone](#) that triggers ripening of fruit.<sup>[9]</sup>

All three tomatoes inhibited ethylene production by reducing the amount of [1-aminocyclopropane-1-carboxylic acid](#) (ACC), the [precursor](#) to ethylene. DNAP's tomato, called Endless Summer, [inserted a truncated version of the ACC synthase gene](#) into the tomato that interfered with the [endogenous ACC synthase](#).<sup>[8]</sup> Monsanto's tomato was engineered with the [ACC deaminase gene from the soil bacterium \*Pseudomonas chlororaphis\*](#) that lowered ethylene levels by breaking down ACC.<sup>[10]</sup> Agritope introduced [an S-adenosylmethionine hydrolase \(SAMase\) encoding gene derived from the \*E. coli\* bacteriophage T3](#), which reduced the levels of [S-adenosylmethionine](#), a precursor to ACC.<sup>[11]</sup>

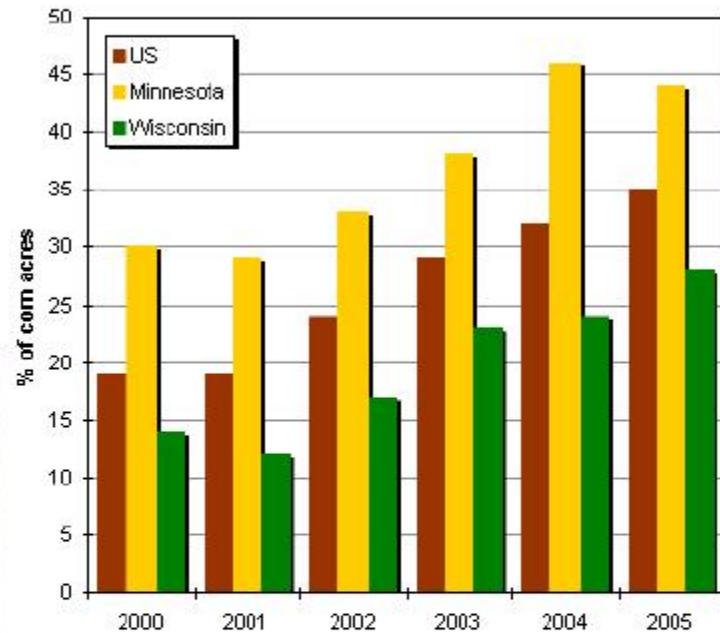
Scientists in India have delayed the ripening of tomatoes [by silencing two ripening specific genes encoding N-glycoprotein modifying enzymes, α-mannosidase and β-D-N-acetylhexosaminidase](#). The fruits produced were not visibly damaged after being stored at room temperature for 45 days, whereas unmodified tomatoes had gone rotten.<sup>[13]</sup> In India, where 30% of fruit is wasted before it reaches the market due to a lack of refrigeration and poor road infrastructure, the researchers hope genetic engineering of the tomato may decrease wastage.<sup>[14]</sup>

## Table 15.3

Examples of gene subtraction projects with plants.

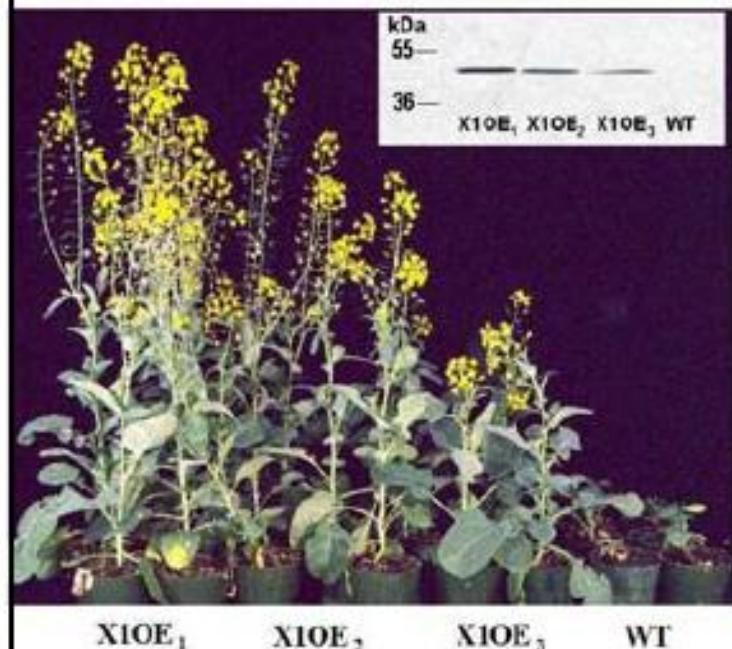
TARGET GENE	MODIFIED CHARACTERISTIC
Polygalacturonase	Delay of fruit spoilage in tomato
1-Aminocyclopropane-1-carboxylic acid synthase	Modified fruit ripening in tomato
Polyphenol oxidase	Prevention of discoloration in fruits and vegetables
Starch synthase	Reduction of starch content in vegetables
Delta-12 desaturase	High oleic acid content in soybean
Chalcone synthase	Modification of flower color in various decorative plants
1D-myo-inositol 3-phosphate synthase	Reduction of indigestible phosphorus content of rice grains

Using biotechnology methods, plants have been made to be more resistant to viral and bacterial diseases, insects infestations, etc.



**Bt corn adoption trends in US,  
Minnesota and Wisconsin 2000-2005**

Other crop plants have been engineered to resist drought, withstand excessive soil salinity, and soil acidity, produce more seeds or fruits, be resistant to herbicides used to control weeds, etc...



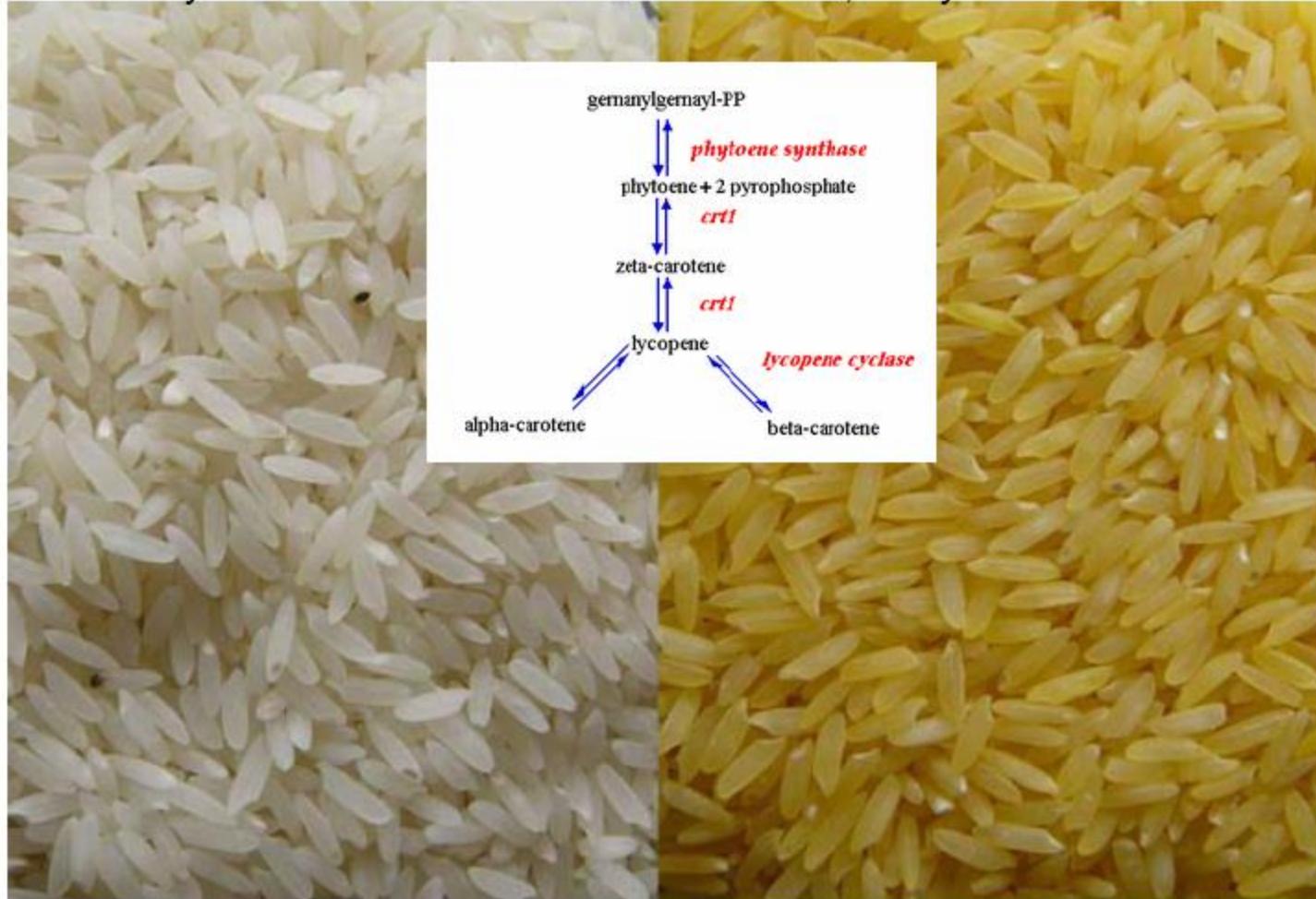
Salt-tolerant, transgenic Brassica plants



*On the left are CMV infected nontransgenic tomato plants, and on the right are CMV resistant transgenic tomato plants. Note the differences in growth and fruiting.*

---

"Golden" rice has been developed with increased vit. A production through the increased production of carotenoids.<sup>1</sup> Other strains of rice have been modified to increase iron accumulation by three times over normal rice. Iron deficiency affects more than 500 million children, many of whom die from it.



# THUMBS-UP TO TRANSGENIC

Some of the 28 no-objection certificates, the companies that have obtained them, and the special traits they have promised for their genetically modified crops (list not exhaustive):



Rice

## COMPANY

Bayer Bio Science,  
Haryana

## PROMISED TRAIT

Insect resistance and  
herbicide tolerance

## CROP



Wheat

## COMPANY

Maharashtra  
Hybrid Seeds,  
Mumbai

## PROMISED TRAIT

Herbicide  
tolerance

## CROP



Cotton

## COMPANY

Monsanto  
India

## PROMISED TRAIT

Insect resistance and  
herbicide tolerance

## CROP



Maize

## COMPANY

Dow Agro  
Sciences India

## PROMISED TRAIT

Insect resistance and  
herbicide tolerance

## CROP



Brinjal

## COMPANY

Ankur  
Seeds

## PROMISED TRAIT

Insect  
resistance

## CROP



Maize

## COMPANY

Pioneer Overseas  
Corporation

## PROMISED TRAIT

Insect resistance and  
herbicide tolerance

## CROP



Maize

## COMPANY

Syngenta Bio  
Sciences, Pune

## PROMISED TRAIT

Insect  
tolerance

## **Common GMO crops**

Currently, only a small number of genetically modified crops are available for purchase and consumption in the United States. The USDA has approved soybeans, corn, canola, sugar beets, papaya, squash, alfalfa, cotton, apples, and potatoes. GMO apples (arctic apples) are non-browning apples and eliminate the need for anti-browning treatments, reduce food waste, and bring out flavor.

The production of Bt cotton has skyrocketed in India, with 10 million hectares planted for the first time in 2011, resulting in a 50% insecticide application reduction.

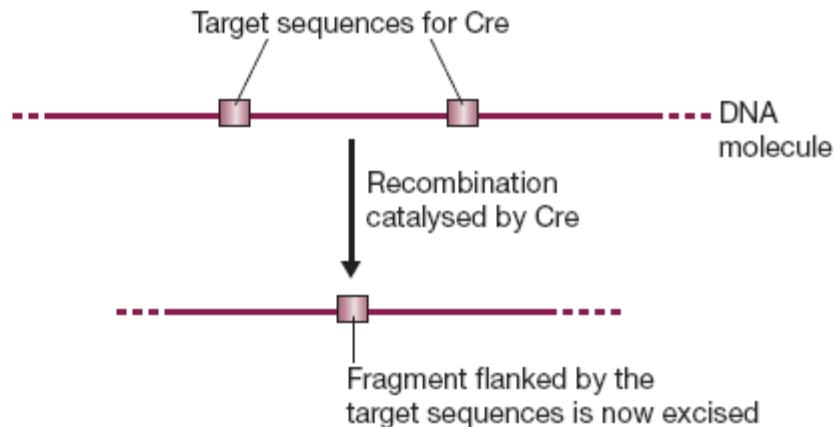
## *Safety concerns with selectable markers*

- Could the  $kan^R$  gene contained in a genetically modified foodstuff be passed to bacteria in the human gut, making these resistant to kanamycin and related antibiotics?
- Could the  $kan^R$  gene be passed to other organisms in the environment, and would this result in damage to the ecosystem?

### Removal of Kan resistance gene by Cre recombinase enzyme

**Figure 15.11**

DNA excision by the Cre recombinase enzyme.

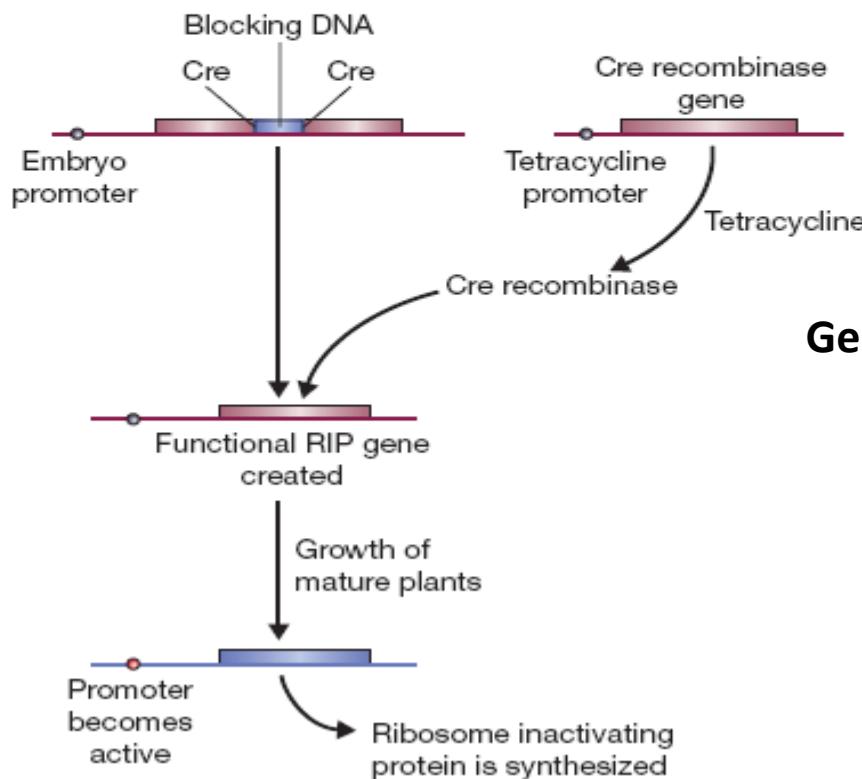


The Cre recombination system also underlies one of the most controversial aspects of plant genetic engineering, the so-called **terminator technology**. This is one of the processes by which the companies who market GM crops attempt to protect their financial investment by ensuring that farmers must buy new seed every year, rather than simply collecting seed from the crop and sowing this second generation seed the following year.

(a) The RIP gene



(b) The terminator system



**Figure 15.12**

The terminator technology. (a) The RIP gene codes for a protein that blocks protein synthesis. (b) The system that is used to allow first generation seeds to be produced.

## Genetic use restriction technology (GURT) or Terminator Technology

## **Safety testing and government regulations**

Agricultural biotechnology regulation in the US falls under three main government agencies: The Department of Agriculture (USDA), the Environmental Protection Agency (EPA), and the Food and Drug Administration (FDA). The USDA must approve the release of any new GMOs, EPA controls the regulation of insecticide, and the FDA evaluates the safety of a particular crop sent to market. On average, it takes nearly 13 years and \$130 million of research and development for a genetically modified organism to come to market. The regulation process takes up to 8 years in the United States.[9] The safety of GMOs has become a topic of debate worldwide, but scientific articles are being conducted to test the safety of consuming GMOs in addition to the FDA's work. In one such article, it was concluded that Bt rice did not adversely affect digestion and did not induce horizontal gene transfer.

Genetic Engineering Appraisal Committee (GEAC)- India's apex regulator of transgenic products

After delays due to activists attempting to prevent the work from being carried out, the UK research team reported their findings in 2003. The study involved 273 field trials throughout England, Wales, and Scotland, and included glyphosate resistant sugar beet as well as maize and spring rape engineered for resistance to a second herbicide, glufosinate-ammonium. The results, as summarized in the official report (see Burke (2003) in *Further Reading*), were as follows:

The team found that there were differences in the abundance of wildlife between GM crop fields and conventional crop fields. Growing conventional beet and spring rape was better for many groups of wildlife than growing GM beet and spring rape. There were more insects, such as butterflies and bees, in and around the conventional crops because there were more weeds to provide food and cover. There were also more weed seeds in conventional beet and spring rape crops than in their GM counterparts. Such seeds are important in the diets of some animals, particularly some birds. In contrast, growing GM maize was better for many groups of wildlife than conventional maize. There were more weeds in and around the GM crops, more butterflies and bees around at certain times of the year, and more weed seeds. The researchers stress that the differences they found do not arise just because the crops have been genetically modified. They arise because these GM crops give farmers new options for weed control. That is, they use different herbicides and apply them differently. The results of this study suggest that growing such GM crops could have implications for wider farmland biodiversity. However, other issues will affect the medium- and long-term impacts, such as the areas and distribution of land involved, how the land is cultivated and how crop rotations are managed. These make it hard for researchers to predict the medium- and large-scale effects of GM cropping with any certainty. In addition, other management decisions taken by farmers growing conventional crops will continue to impact on wildlife.