UNIVERSITY OF GHARDAÏA
DEPARTMENT OF BIOLOGY

GENETIC ENGINEERING

By: ALLA EDDINE MAHAMEDI, PH.D.

// Course for students in Biochemistry, 3rd year //

– GHARDAÏA 2024 –

PREFACE

Genetic engineering stands at the forefront of modern science, offering transformative potential across medicine, agriculture, environmental management, and beyond. This course is designed to introduce students to the principles, techniques, and the main applications through the modification of genetic material. As we explore the fundamental concepts of DNA, gene expression and DNA modification, we also delve into real-world applications—from the development of gene therapies to the creation of genetically modified organisms that promise enhanced crop yields and disease resistance.

Throughout this course, critical analysis and hands-on learning will encourage students to understand not only the scientific mechanisms but also the broader implications of genetic manipulation.

The aim is to cultivate a deep understanding of theoretical insights and practical examples, this course invites you to explore how genetic engineering is reshaping our world and to consider how you might help guide its future development.

Welcome to a journey into the remarkable world of genetic engineering—where curiosity meets innovation and new technologies that drive sustainable scientific progress.

CONTENT

CHAPTER 1

INTRODUCTION TO GENETIC ENGINEERING

INTRODUCTION

The discovery in the **1950s** of the role of **DNA** and the **genes** it contains gave rise to an applied discipline: **genetic engineering**. Based on a set of **techniques**, it aims to **study** and **exploit** genes, mainly in the **medical**, **pharmaceutical** and **agri-food** fields. The exploitation of genes assumes that they can be **isolated**, their structure **determined** and **introduced** into **cells** or whole **organisms** in order to **modify** their biological properties.

Although there are many unique and intricate procedures involved, the basic concepts of genetic modification are reasonable. **Genetic material** provides a **valuable resource** in the form of information recorded by the sequence of bases in DNA. The technology is based on the assumption that genetic information, encoded by **DNA** and organized into **genes**, is a resource that can be **manipulated** in a variety of ways to achieve **specific goals** in both pure and practical research and medicine. Genetic modification has several applications, including the following:

- **Fundamental study** on gene structure and function.
- Production of **valuable proteins** using innovative technologies.
- Production of **transgenic** plants and animals.
- **Medical** diagnosis and therapy.
- **Genome analysis** using DNA sequencing.

1.1. Molecular Cloning

Utilizing a variety of molecular tools originating from microbes, **molecular cloning** is a collection of techniques used to create recombinant DNA and incorporate it into a host organism. The first molecular cloning procedure was completed in 1973 when Herbert Boyer and Stanley Cohen successfully cloned and inserted genes from the African clawed frog (*Xenopus laevis*) into a bacterial plasmid, which was subsequently inserted into the bacterial host *Escherichia coli*. The following figure 1 shows the main steps in a gene cloning experiment.

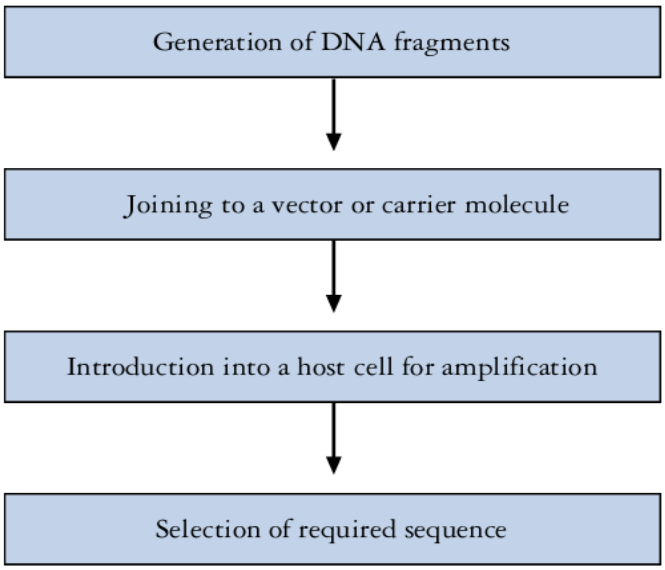


Fig. 1. The main steps in a gene cloning experiment.

1.1.1. Basic concept in genetic engineering

Genetic engineering involves the use of **recombinant DNA technology**, the process by which a DNA sequence is manipulated *in vitro*, thus creating **recombinant DNA molecules** that have new combinations of genetic material. The recombinant DNA is then introduced into a **host** organism. If the DNA that is introduced comes from a different species, the host organism is now considered to be transgenic. The general process implicates:

1. Extraction of the **gene** encoding the **desired protein**
2. Plasmid **extraction** (vector) and opening by **restriction enzyme**
3. Insertion of the gene into the plasmid in the presence of **ligase** (a recombined plasmid is obtained)
4. Culture of **bacteria** in the presence of plasmid (recombinant DNA) to be integrated by bacteria.
5. Identification of transformed cells: **expression** of the gene in question.
6. Extraction of protein by osmotic shock techniques.

Example of production of human insulin (Fig. 2). The insulin gene from humans was inserted into a plasmid. This recombinant DNA plasmid was then inserted into bacteria. As a result, these transgenic microbes are able to produce and secrete human insulin. Many prokaryotes are able to acquire foreign DNA and incorporate functional genes into their own genome through "mating" with other cells (conjugation), viral infection (transduction), and taking up DNA from the environment (transformation). Recall that these mechanisms are examples of horizontal gene transfer—the transfer of genetic material between cells of the same generation.

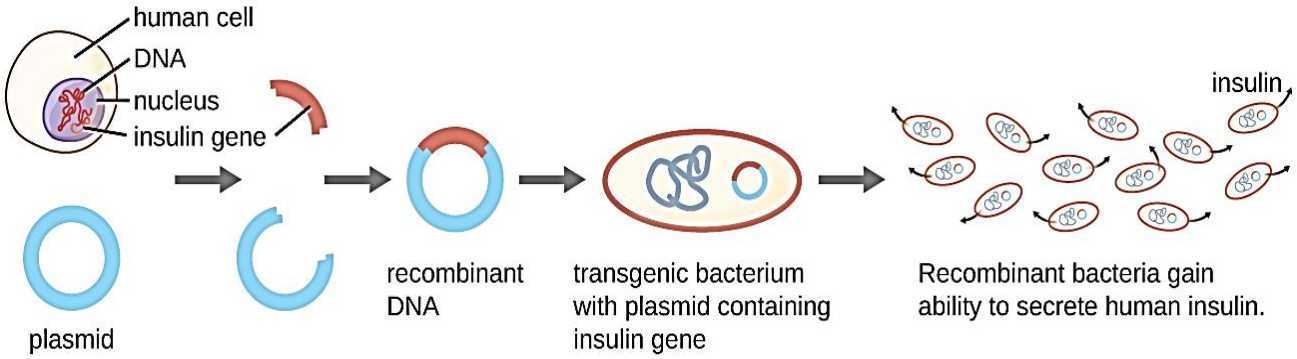


Fig. 2. Recombinant DNA technology is the artificial recombination of DNA from two organisms.

1.1.2. Concept map

Here is a general map that shows the main steps, tools, methods and applications of genetic engineering (Fig. 3).

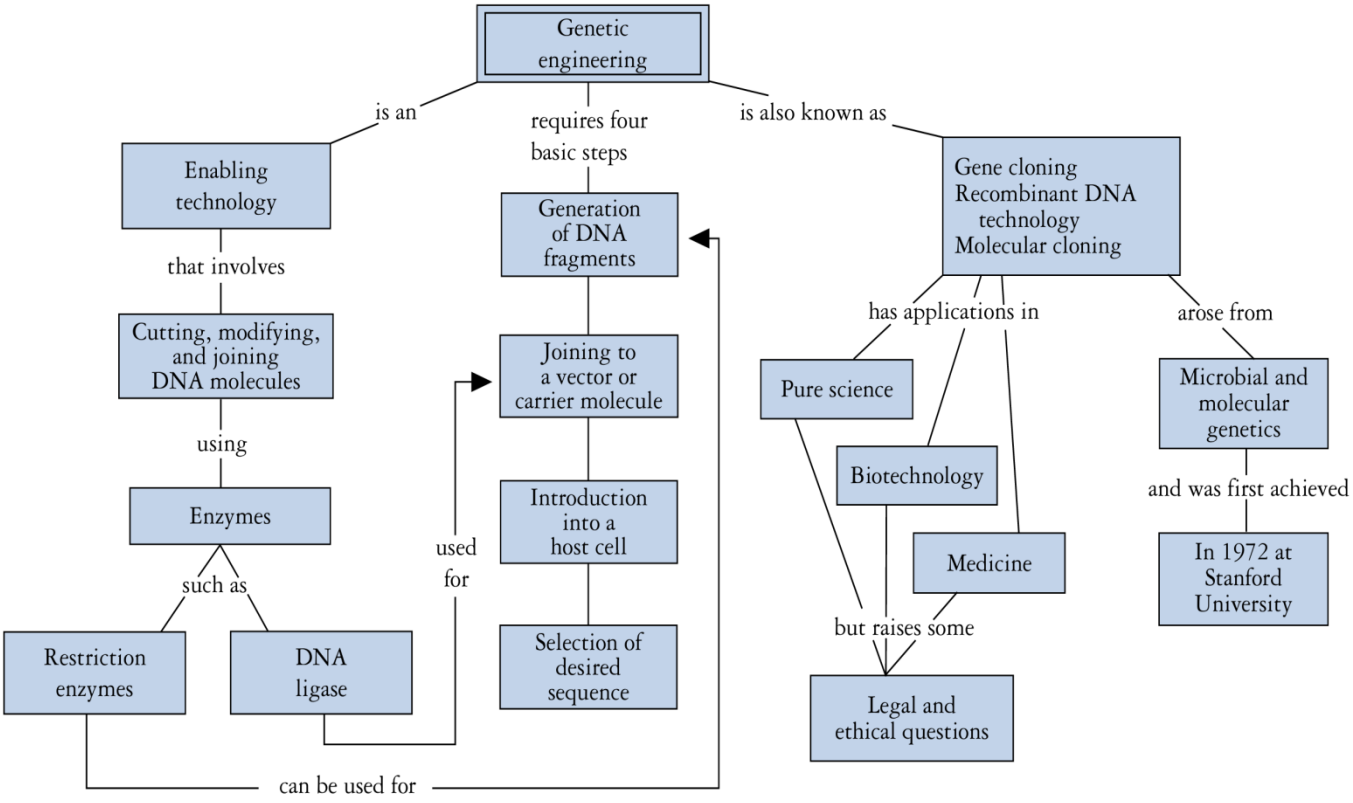


Fig. 3. Concept map of genetic engineering.

CHAPTER 2

ENZYMATIC TOOLS OF GENETIC ENGINEERING

In recombinant DNA technology, DNA molecules are manipulated using **naturally** occurring enzymes derived mainly from **bacteria** and **viruses**. The creation of recombinant DNA molecules is possible due to the use of naturally occurring restriction **endonucleases** (restriction enzymes), bacterial enzymes produced as a protection mechanism to **cut** and **destroy** foreign cytoplasmic DNA that is most commonly a result of bacteriophage infection. Stewart Linn and Werner Arber discovered restriction enzymes in their 1960s studies of how *E. coli* limits bacteriophage replication on infection. Today, we use restriction enzymes extensively for cutting DNA fragments that can then be spliced into another DNA molecule to form recombinant molecules.

The basic methodology for manipulating genes and cloning recombinant DNA molecules requires various **restricting, joining** and **modifying enzymes**.

2.1. Restriction enzymes – cutting DNA

The restriction enzymes, which **cut** DNA at **defined sites**, represent one of the most important groups of enzymes for the manipulation of DNA. These enzymes are found in bacterial cells, where they function as part of a protective mechanism called the restriction-modification system.

Restriction enzymes are of three types (**I, II**, or **III**). Most of the enzymes commonly used today are **type II enzymes**, which have the **simplest** mode of action. These enzymes are **nucleases** and as they cut at an **internal position** in a DNA strand (**endonucleases**).

— **Type I** and **III systems** : recognize specific DNA sequences, the sites of actual cleavage are **so far**, and can be **hundreds** of bases **away**. Both require **ATP** for their proper function.

— **Type II systems** : make cleavages at very specific sites that are **within** or **close** to the recognition sequence. The vast majority of known restriction enzymes are of type II, and it is these that find the **most** use as **laboratory** tools. The first to be discovered and utilized was **EcoRI**, which is staggered and its recognition sequence is **5'-GAATT**C-3'.

2.1.1. Naming of restriction enzymes

Restriction enzymes are named based on the bacteria in which they are isolated in the following way. For example **EcoRI :**

E ***Escherichia*** (genus)
Co ***coli*** (species)
R **RY13** (strain)
I **First** identified Order ID'd in bacterium

Several restriction enzymes are known. The following table shows the main restriction enzymes widely used in genetic engineering techniques.

Table 2.1. Restriction enzymes widely used in genetic engineering

Enzyme	Recognition sequence	Cutting sites	Ends
*Bam*HI	5′-GGATCC-3′	G↓GATCC CCTAG↑G	5′
*Eco*RI	5′-GAATTC-3′	G↓AATTC CTTAA↑G	5′
*Hae*III	5′-GGCC-3′	GG↓CC CC↑GG	Blunt
*Hpa*I	5′-GTTAAC-3′	GTT↓AAC CAA↑TTG	Blunt
*Pst*I	5′-CTGCAG-3′	CTGCA↓G G↑ACGTC	3′
*Sau*3A	5′-GATC-3′	↓GATC CTAG↑	5′
*Sma*I	5′-CCCGGG-3′	CCC↓GGG GGG↑CCC	Blunt
*Sst*I	5′-GAGCTC-3′	GAGCT↓C C↑TCGAG	3′
*Xma*I	5′-CCCGGG-3′	C↓CCGGG GGGCC↑C	5′

2.1.2. Restriction enzymes functioning

Each restriction enzyme cuts DNA at **a characteristic recognition site**, a specific, usually **palindromic**, DNA sequence typically between four to six base pairs in length. A palindrome is a sequence of letters that reads the same forward as backward. (The word "*level*" is an example of a palindrome.) Palindromic DNA sequences contain the same base sequences in the 5' to 3' direction on one strand as in the 5' to 3' direction on the complementary strand.

A restriction enzyme **recognizes** the DNA palindrome and **cuts each backbone** at identical positions in the palindrome. Some restriction enzymes cut to produce molecules that have **complementary overhangs** (sticky ends) while others cut without generating such overhangs, instead producing **blunt ends** (fig. 4).

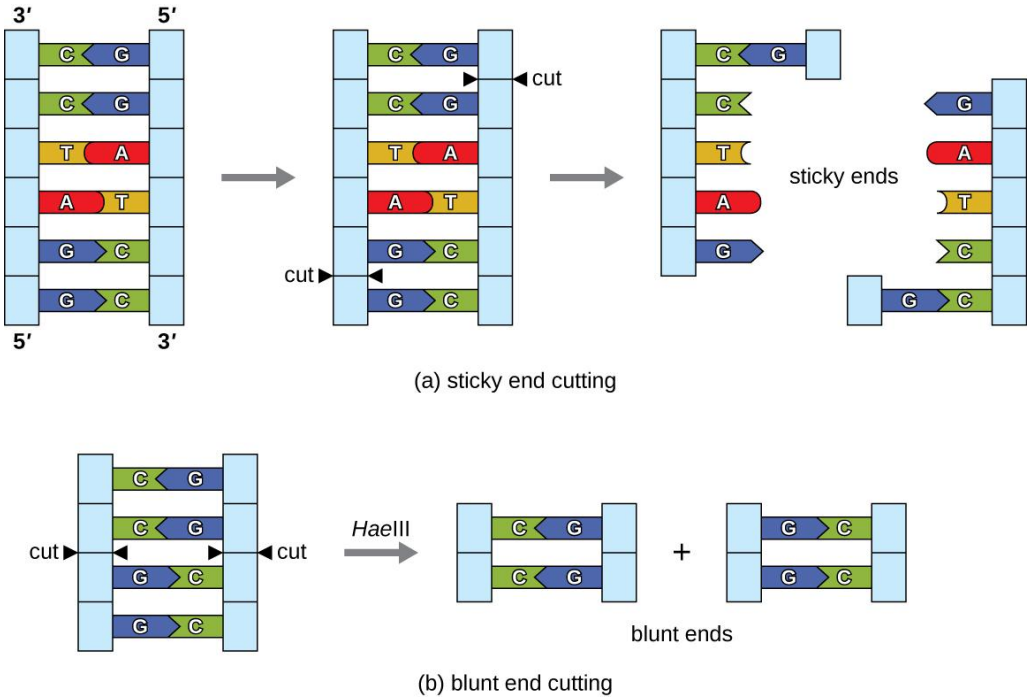


Fig. 4. Sticky end (a) and blunt end (b) cuts.

2.1.3. Isoschizomers and Neoschizomers

There are many different phages infecting a variety of different host strains, all of which have developed restriction/ modification systems as a form of self-defense. Therefore, it is not surprising that enzymes from different bacteria may recognize the same recognition motif if they are infected with phage carrying the very same motif. Here we can have two possibilities:

- If two enzymes isolated from different bacteria (hence different enzyme names) recognize the **same sequence** and cut at exactly the **same position**, these enzymes are called **isoschizomers**.
- If the two enzymes recognize the **same DNA sequence** motif but cleave at **different positions**, then they are called **neoschizomers**.

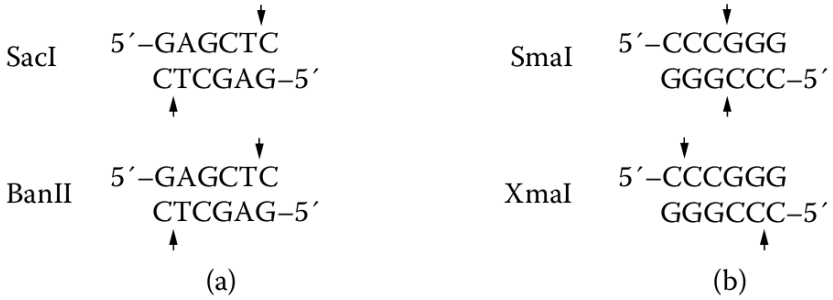


Fig. 5. Isoschizomers (SacI and BanII ; a) and neoschizomers (SmaI and XmaI ; b).

Examples of isoschizomers and neoschizomers enzymes: The following table shows some of the known isoschizomers and neoschizomers enzymes used in genetic engineering.

Table 2. Examples of isoschizomers and neoschizomers

Enzyme	*Isoschizomer(s)*	*Recognition Sequence*
AclNI	SpeI	A/CTAGT
Bsp19I	NcoI	C/CATGG
Bsp106I	ClaI, Bsu15I	AT/CGAT
Eco32I	EcoRV	GAT/ATC
Sac II	SstII, KspI	CCGC/GG
XhoI	PaeR7I	C/TCGAG

2.1.4. Restriction mapping

Most sequences of DNA will have recognition sites for **various restriction** enzymes, and it is often beneficial to know the relative locations of some of these sites. The technique used to obtain this information is known as restriction mapping. This involves cutting a DNA fragment with a **selection** of restriction enzymes, singly and in various combinations. The fragments produced are run on an agarose gel and their sizes determined (Fig. 6). From the data obtained, the relative locations of the restriction sites can be worked out.

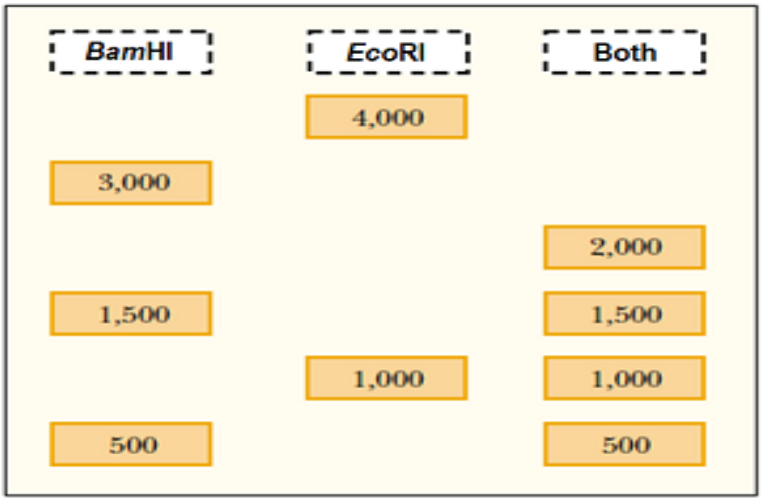


Fig. 6. Example of cutting a linear DNA fragment of 5 kb using BamHI et EcoRI restriction enzymes.

The analysis of the digestion results allows us to determine the different **possibilities** of the **localization** of the restriction sites for each enzyme (Fig. 7a,b).

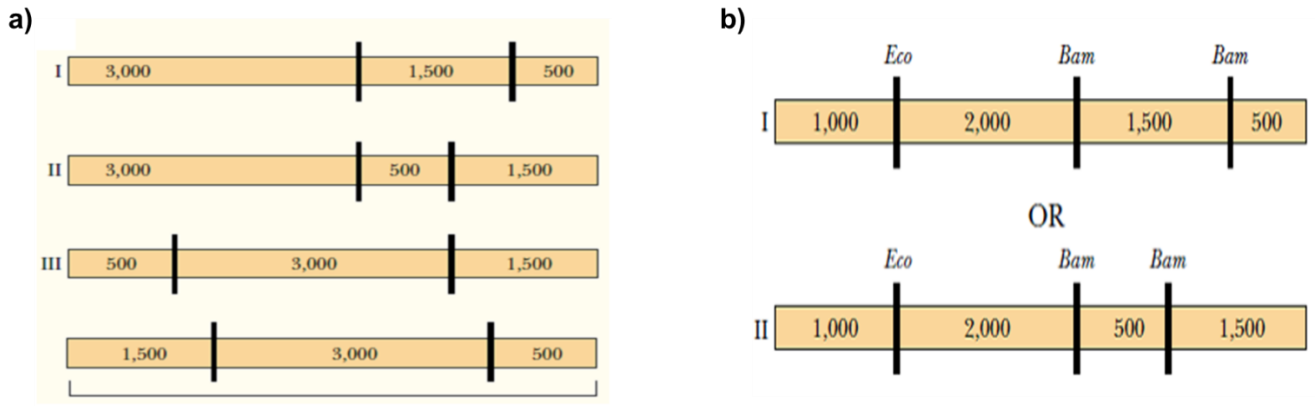


Fig. 7. Possible restriction sites locations by using BamHI (a) and both enzymes (EcoRI and BamHI).

2.1.5. DNA methylation

DNA methylation is the **modification** of DNA bases by adding a **methyl group** by a **DNA methyltransferase**. It is important to consider methylation when planning a restriction digest, as methylated DNA may block or impair the binding of restriction enzymes to DNA. The 3 most common types of methylation that impact restriction digestions are **dam** and **dcm** methylation, which are found in prokaryotes such as *E. coli*, and **CpG** methylation, which is found in eukaryotes such as plants and mammals.

If present within the recognition site, a methyl group can **prevent** the **restriction** enzyme from binding and cleaving. However, dam and dcm methylation will only impact restriction digestion of a plasmid if the restriction and methylation sites overlap.

a. Dam methylase

Dam methylase methylates adenine in the sequence 5'-GA*TC-3'. The recognition sites of several restriction enzymes such as PvuI, BamHI, BclI, BglII, XhoII, MboI and Sau3AI contain this sequence as well as a certain proportion of the sites recognized by ClaI (1 out of 4 sites), TaqI (1 out of 16 sites), MboII (1 site out of 16) and HphI (1 site out of 16). Some enzymes do not cut if adenine is methylated. The MboI enzyme does not cut the GATC sequence if it is methylated while the Sau3AI enzyme recognizes the same sequence as MboI and is not affected by methylation dam.

b. Dcm methylase

This methylase adds a methyl group to the internal cytosine of the 5'-CC*AGG-3' and 5'-CC*TGG-3' sequences. One enzyme affected by dcm methylation is EcoRII. We can work around the problem by using BstNI which recognizes the same sequence as EcoRII, but does not cut it in the same place. If BstNI is not suitable, DNA can be prepared from *E. coli* strains that are dcm-.

Example of *E. coli* restriction modification system: The recognition motifs for the **EcoRI** restriction endonuclease in the host genome are modified by the EcoRI methylase, which covalently adds a **methyl group** to the adenine nucleotide.

This modification does **not affect** the structure of the host DNA, but simply **disables** the endonuclease from recognizing the motif, thus the host genome is protected from cleavage (Fig. 8).

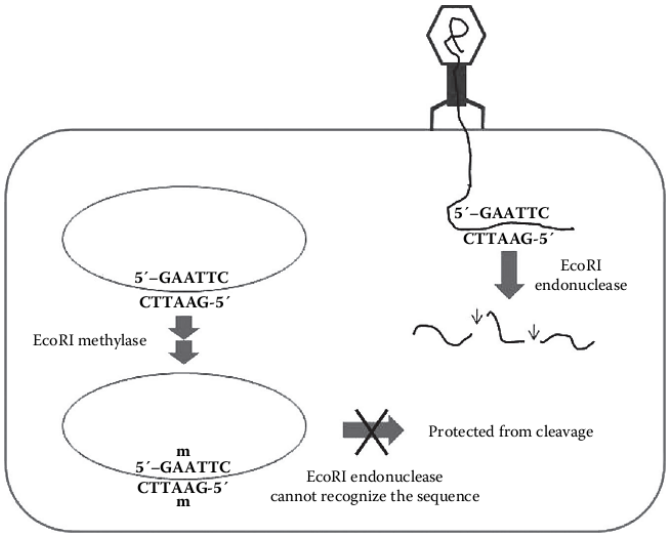


Fig. 8. Restriction modification system for *E. coli.*

2.2. DNA modifying enzymes

Besides restriction enzymes and DNA ligase (cutting and joining functions), other enzymes are used in genetic engineering may be loosely termed as **DNA modifying enzymes**, with the term used here to include **degradation**, **synthesis**, and **alteration** of DNA. Here are some of the most commonly used enzymes.

2.2.1. Nucleases

Nuclease enzymes degrade nucleic acids by breaking the **phosphodiester bond** that holds the nucleotides together. These enzymes could cut **within** a DNA strand (**endonucleases**), but others from nucleases could degrade DNA from the **termini** of the molecule (**exonucleases**).

a. DNase

DNase acts on the bonds adjacent to the pyrimidic nucleosides (C, T) of single or double strand DNA. The activity of this enzyme is dependent on the presence of Mg++ or Mn++

- Mg++: Random hydrolysis

– Mn++: Sequence-dependent hydrolysis

This enzyme is used for the analysis of **protein-binding sites –DNA** by the technique of **foot-printing** (footprint). Because the DNA-binding protein protects it from hydrolysis by this enzyme. This enzyme is extracted from the beef pancreas.

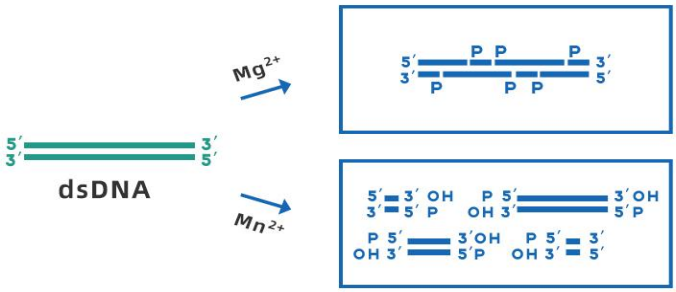


Fig. 9. Functioning of Dnases

b. S1 Nuclease

It is a metalloprotein (PM: 32000 Da) produced by the fungus *Aspergillus oryzae*, it is a **single strand DNA** (or **RNA**) **specific** nuclease, although at high concentrations it acts on hybrids. It is used in several applications:

– To make **blunt ends** of double-stranded DNA fragments.
– To **hydrolyze single strand** DNA fragments at the gaps (even if only one link is missing).
– To **open hairpins** formed during cDNA synthesis.
– To **isolate** DNA/DNA, DNA/RNA, or RNA/RNA **hybrids** (a tool of choice for measuring hybridization rate).

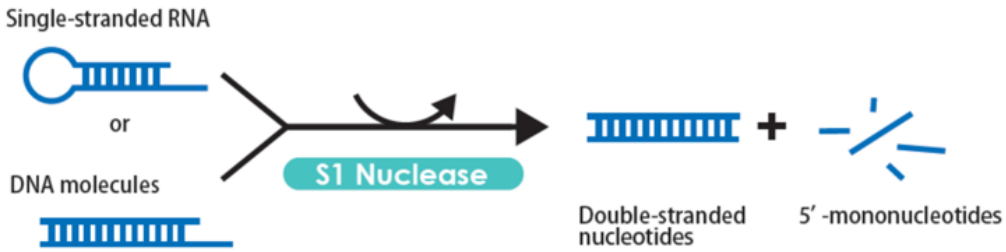


Fig. 10. Functioning of S1 Nuclease

c. Lambda exonuclease

It is also encountered in T4 or T7, as well as in animals (DNase IV) hydrolyzes preferably the 5'-phosphate ends of double-stranded DNA continuing to the 3' side (5'→3' exonuclease). It produces 5'-phosphate nucleosides. This enzyme is produced by phage-infected bacterial cells.

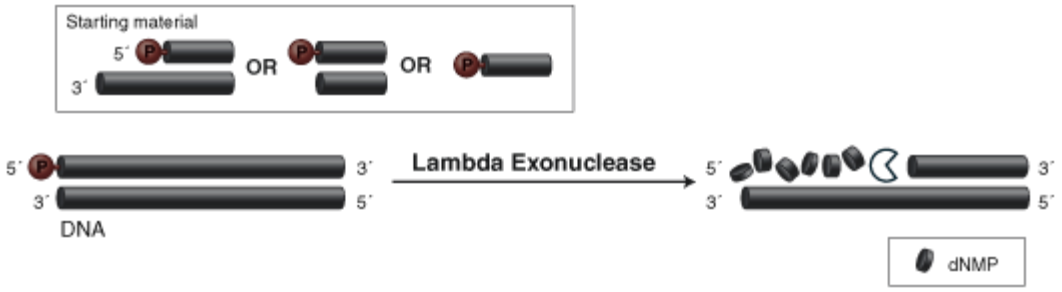


Fig. 11. Functioning of Lambda exonucléase.

d. Exonuclease III

This enzyme produced by *E. coli*, and also by *Haemophilus influenza*, in contrast to lambda exonuclease, the exonuclease III hydrolyzes preferably the **3' ends** of double-stranded DNA up to the 5' side (3'→5' exonuclease). It produces **5'-phosphate** nucleotides. This enzyme can also hydrolyze a phosphorylated radical linked to the 3'-OH function of the last nucleotide (phospho-mono-esterase action).

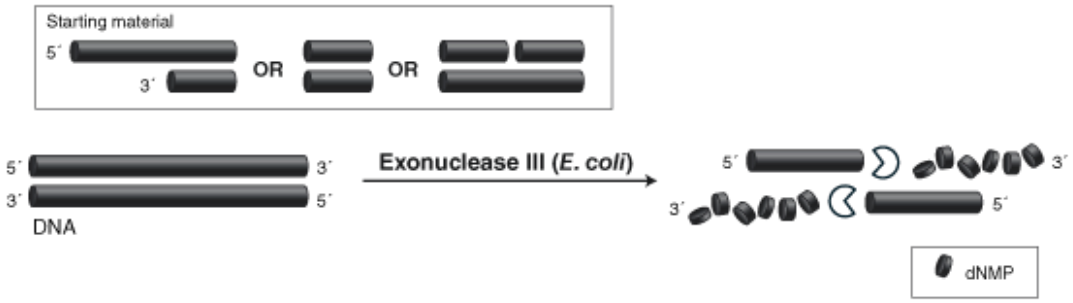


Fig. 12. Functioning of Exonuclease III.

e. Exonuclease Bal 31

This multifunctional enzyme is produced by *Alteromonas espegiana*. It has both:

- **Exonuclease** activity simultaneously hydrolyzing both ends 3' and 5' of double stranded DNA.
- Highly specific **endodeoxyribonuclease** activity of single stranded DNA.

The Bal31 acts sequentially, as **exonuclease** in the directions 3' → 5' then by **endonucleotide** elimination of single outgoing strands.

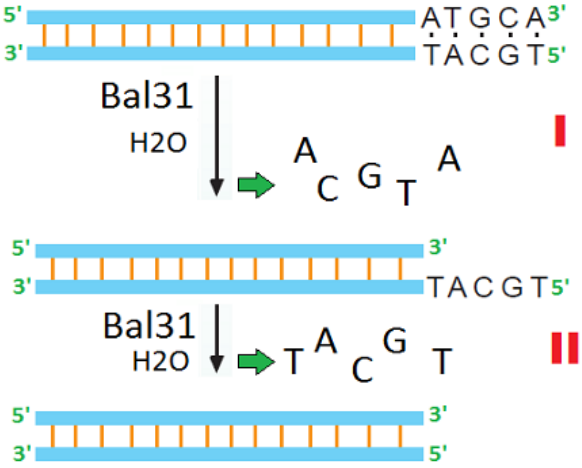


Fig. 13. Functioning of Bal 31 Exonuclease.

In addition to DNA-specific nucleases, there are **ribonucleases (RNases)**, which act on RNA. These may be required for many of the stages in the preparation and analysis of recombinants and are usually used to get rid of unwanted RNA in the preparation.

f. RNase

It acts as an **endonuclease**, preferentially **after** pyrimidine nucleotides (U, C). Hydrolyzing the bond between phosphate and 5' carbon of the next nucleotide. It hydrolyzes **single-stranded RNA** (inactive on double-stranded RNA) to a mixture of oligonucleotides all ending in a pyrimidine nucleotide esterified by a 3' phosphate.

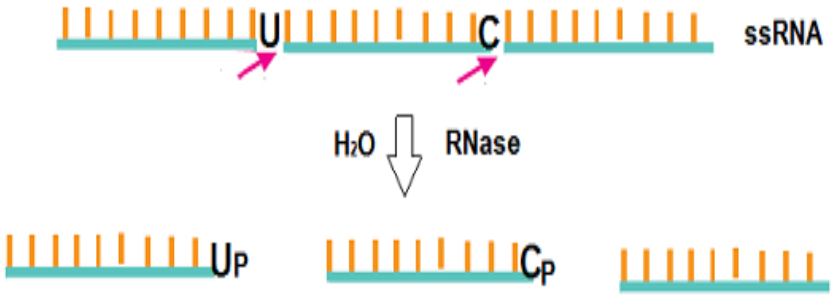


Fig. 14. Functioning of RNase.

N.B. *RNase inhibitors, Ex. SDS (Sodium Dodecyl Sulphate), proteins extracted from the placenta (these inhibitors are often used to protect RNA during extraction or enzymatic reactions).*

g. Ribonuclease T1

This nuclease is produced by *Aspergillus oryzae*, has a PM of 11000 Da, and specifically hydrolyzes **ssRNAs** by breaking **3'-5' phosphoester** bonds downstream of the **GMP**, such that the product is a Guanosine 3'-phosphate or oligonucleotide ending in Guanosine 3'-phosphate. This enzyme is mainly used in RNA sequencing.

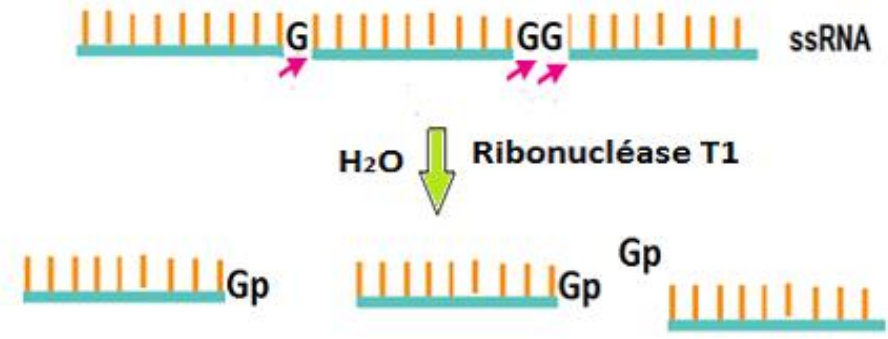


Fig. 15. Functioning of Ribonuclease T1.

h. Ribonuclease T2

As the T1-ribonuclease, this enzyme is produced by *Aspergillus oryzae*, has a PM of 36 000 Da, and specifically hydrolyzes **sb RNAs** by breaking **3'-5' phosphoester** bonds downstream of adenosine (**AMP**). It is also used in RNA sequencing.

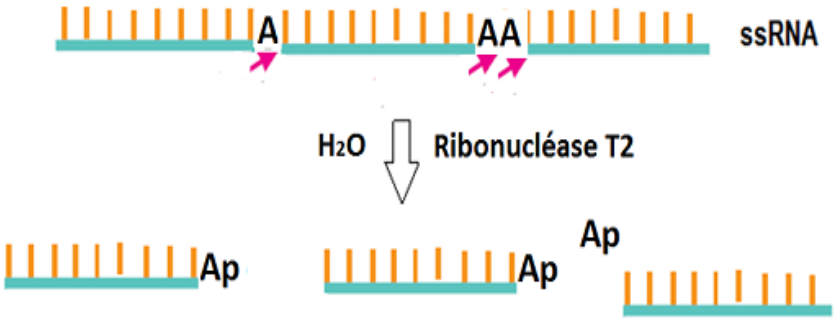


Fig. 16. Functioning of Ribonuclease T2.

2.2.2. Polymerases

Polymerase enzymes **synthesise** copies of nucleic acid molecules and are used in many genetic engineering procedures. When describing a polymerase enzyme, the terms '**DNA-dependent**' or '**RNA-dependent**' may be used to indicate the type of nucleic acid template that the enzyme uses.

- **DNA-dependent** DNA polymerase copies **DNA** into **DNA**;
- **RNA-dependent** DNA polymerase copies **RNA** into **DNA**;
- **DNA-dependent** RNA polymerase transcribes **DNA** into **RNA**.

a. DNA polymerase I

It has, in addition to its polymerase function, **5' → 3'** and **3' → 5' exonuclease** activities. The enzyme catalyses a strand-replacement reaction, where the 5'→3' exonuclease function degrades the non-template strand as the polymerase synthesises the new copy. A **major** use of this enzyme is in the nick translation procedure for radiolabelling DNA.

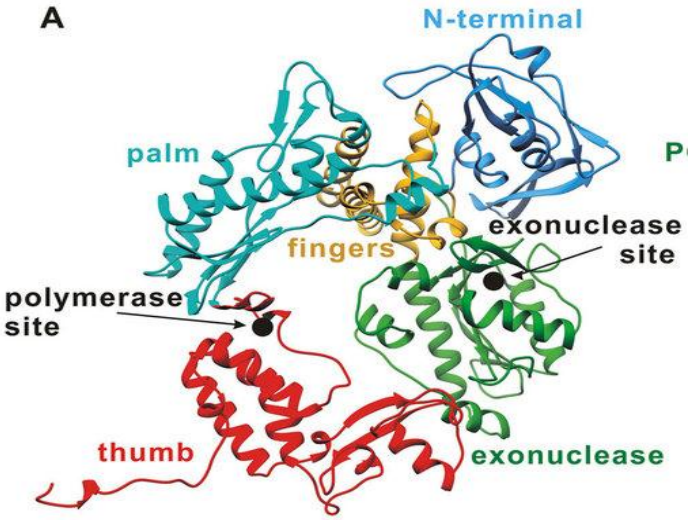


Fig. 17. The three-dimensional structure of DNA polymerase I.

b. Klenow fragment

It is the result of removing the 5' → 3' exonuclease function of DNA polymerase I. This retains the **polymerase** and **3' → 5' exonuclease** activity.

The Klenow fragment is used where a single-stranded DNA molecule needs to be copied; because the 5' → 3' exonuclease function is missing, the enzyme cannot degrade the **non-template strand** of **dsDNA** during synthesis of the new DNA.

Major uses for the Klenow fragment include **radiolabelling** by primed synthesis and **DNA sequencing** by the dideoxy method.

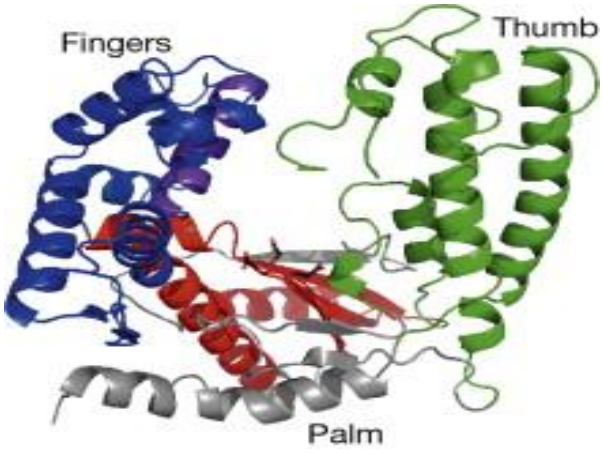


Fig. 18. The three-dimensional structure of Klenow fragment

c. Reverse transcriptase (RTase)

It is an **RNA-dependent DNA polymerase**, and therefore produces a DNA strand from an RNA template. It has no associated exonuclease activity. The enzyme is used mainly for copying **mRNA** molecules in the preparation of **cDNA** (complementary or copy DNA) for cloning, although it will also act on DNA templates.

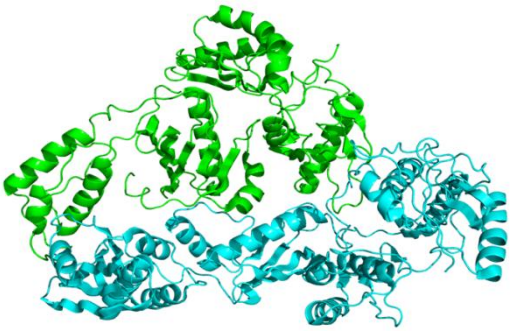


Fig. 19. The three-dimensional structure of Reverse transcriptase (RTase)

2.2.3. Enzymes that modify the ends of DNA molecules

The enzymes **alkaline phosphatase**, **polynucleotide kinase**, and **terminal transferase** act on the termini of DNA molecules and provide important functions that are used in a variety of ways.

a. phosphatase and **b. kinase enzymes**, as their names suggest, are involved in the **removal** or **addition** of phosphate groups (Fig. 20). Bacterial alkaline phosphatase (there is also a similar enzyme, calf intestinal alkaline phosphatase) removes phosphate groups from the 5’ ends of DNA, leaving a 5’-OH group.

The enzyme is used to prevent **unwanted ligation** of DNA molecules, which can be a problem in certain cloning procedures. It is also used prior to the addition of **radioactive phosphate** to the 5’ ends of DNAs by polynucleotide kinase.

c. Terminal transferase (terminal deoxynucleotidyl transferase).

This enzyme **repeatedly** adds **nucleotides** to any available **3' terminus**. Although it works best on protruding 3' ends, conditions can be adjusted so that blunt-ended or 3'-recessed molecules may be utilised. The enzyme is mainly used to add **homopolymer tails** to DNA molecules prior to the construction of recombinants (Fig. 20).

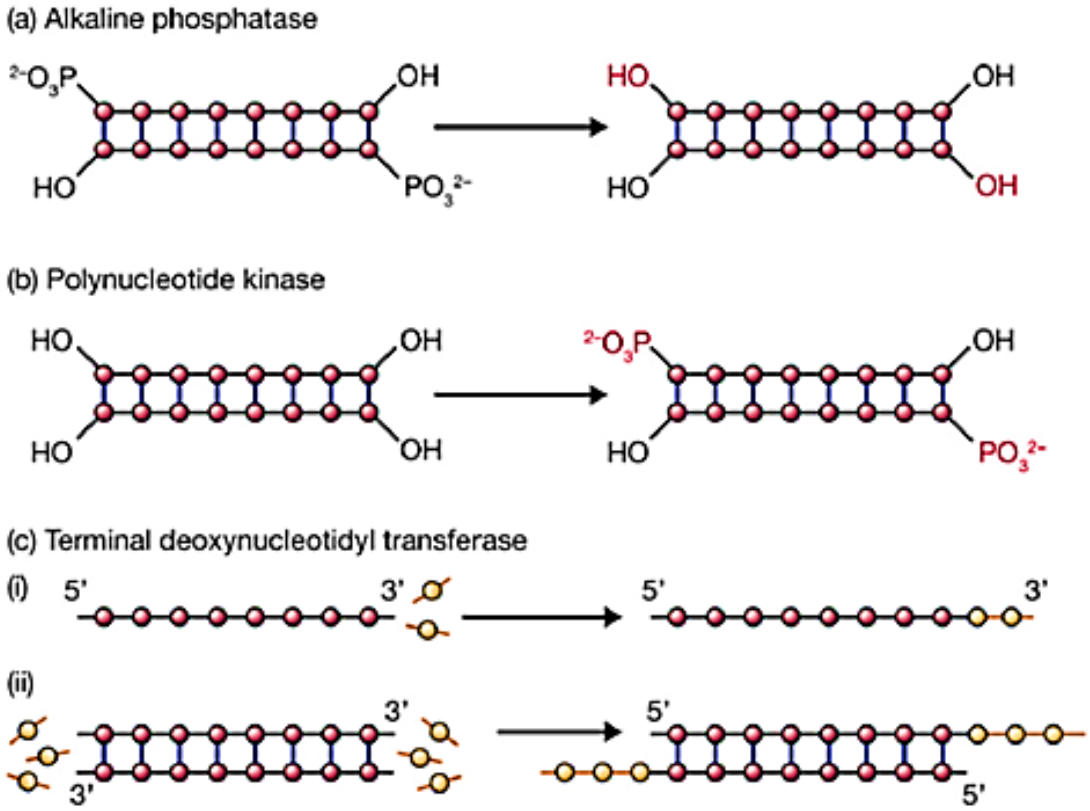


Fig. 20. Fonctionning of alkaline phosphatase (a), polynucleotide kinase (b), and terminal transferase (c).

2.3. DNA ligase – joining DNA molecules

2.3.1. DNA ligase

It is an important cellular enzyme, as its function is to **repair broken phosphodiester** bonds that may occur at random or as a consequence of DNA replication or recombination.

In genetic engineering it is used to **seal** discontinuities in the **sugar—phosphate** chains that arise when recombinant DNA is made by **joining** DNA molecules from different sources (molecular glue which is a key enzyme in genetic engineering.

a. T4 DNA ligase

It is the enzyme used most often in experiments, which is purified from *E. coli* cells infected with bacteriophage T4. The enzyme is most efficient when sealing **gaps** in fragments that are held together by **cohesive ends**, it will also join **blunt-ended** DNA molecules together under **appropriate conditions**.

The enzyme works best at 37 °C, but is often used at much lower temperatures (4-15 °C) to **prevent thermal denaturation** of the short base-paired regions that hold the cohesive ends of DNA molecules together.

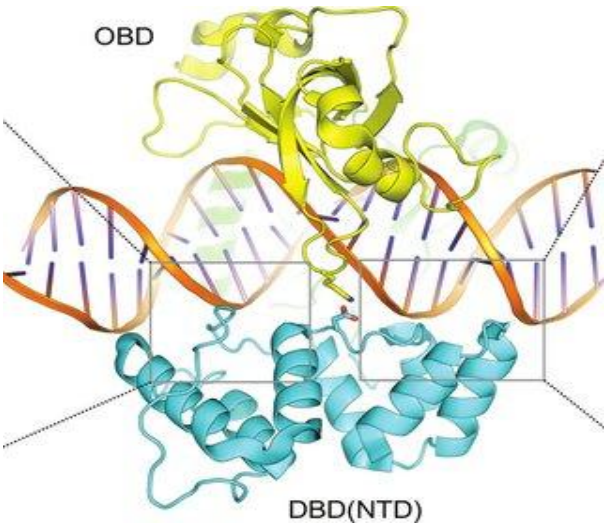


Fig. 21. The three-dimensional structure of T4 DNA ligase.

The T4 DNA ligases are the most used in genetic engineering. The amount of DNA ligase required in each reaction and the activity of this enzyme depend on a number of factors:

- Nature of ligature DNA fragments (blunt or cohesive ends).
- Length of cohesive ends.
- Stability of hydrogen bonds.
- The incubation temperature.
- Concentration of fragments.

Example: The ligation speed of restriction fragments of pBR322, of the same length and all with cohesive ends, varies according to the sequence of these ends.

Summary on enzymes used in genetic engineering

The ability to **cut**, **modify**, and **join** DNA molecules gives the genetic engineer the freedom to **create recombinant** DNA molecules.

The technology involved is a **test-tube technology**, with no requirement for a living system. However, once a recombinant DNA fragment has been generated ***in vitro***, it usually has to be amplified so that **enough material** is available for subsequent manipulation and analysis.

Amplification usually requires a biological system, unless the polymerase chain reaction (PCR) is used. We must, therefore, examine the types of **living systems** that can be used for the **propagation** of recombinant DNA molecules.

CHAPTER 3

HOST-VECTOR SYSTEMS AND GENE CLONING

3.1. Vector

Definition: A cloning vector is a **small**, **self-replicating** genetic element used to produce **multiple copies** of the gene of interest. These cloning vectors are specifically designed to allow the **integration** of the **exogenous DNA** portion into a specific site without affecting its own replication.

There are several types of vectors that can introduce recombinant DNA molecules into bacteria, yeasts, plant cells or mammalian and human cells.

The ideal vector : First, the cloning vector must have a known replication **origin** (*Ori*) by the host cell, which allows its replication. It must also contain a **selector marker** (antibiotic resistance gene), which allows the selection of transformed clones. A cloning vector must have a single **restriction site** for one or **more** restriction enzymes (MCS; Fig. 22).

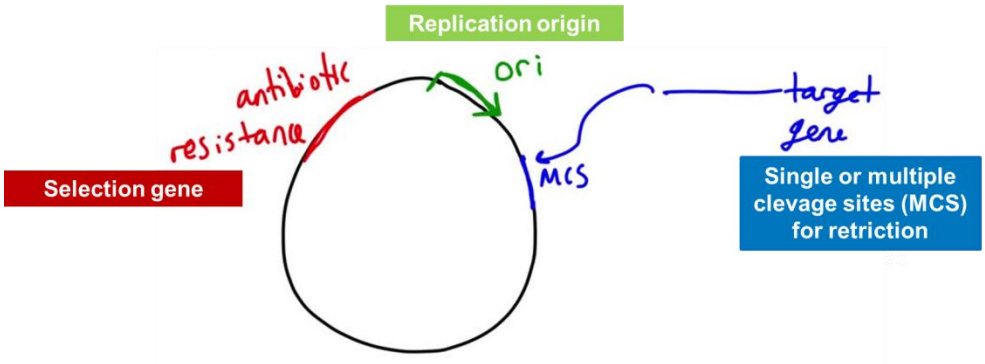


Fig. 22. Composition of a cloning vector.

In genetic engineering several vectors from different sources are used:

3.1.1. Plasmids

Plasmids are **small extra-chromosomal** genetic elements with **autonomous** replication, typically **double-stranded**, **circular** DNA molecules, ranging in size from 1 kb to two or three hundred kb (< 5% of bacterial chromosome size). They contain genes that often code for proteins that give an advantage(s) to the host cell. For example:

- **Antibiotic** resistance
- Resistance to **heavy metals**
- Degradation of **aromatic** compounds
- Production of **toxins**
- Production of **antibiotics**
- Induction of **tumours** in plants

Benefits of using plasmids

Plasmids are good cloning vectors in bacteria because they:

- Multiply in **large** numbers
- **Easily** purified.
- Presence of **selection markers** (e.g. antibiotic resistance genes) that allow the identification of recombinant bacteria (transformed) they carry.

In genetic engineering, several vectors are used for molecular cloning:

a. Plasmide pBR322

The pBR322 belongs to a series of **first generation** cloning vectors, partially built by genetic engineering. Which was built on the basis of a chimera gathering the interesting elements of natural plasmids and increasing the number of unique sites for cutting by restriction enzymes (Fig. 23).

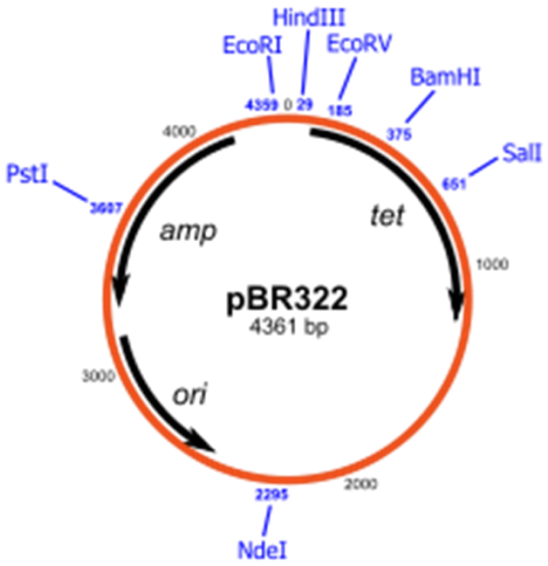


Fig. 23. The plasmide pBR322.

▪ Characteristics of the plasmide pBR322

1. pBR322 is a small plasmid consisting of **4361** base pairs, whose nucleotide sequence is completely known.
2. It is maintained **stably** in its host at a level close to 20 to 30 copies per cell.
3. Its production can be **increased** to several thousand copies (1000 to 3000 copies per cell) when protein synthesis is inhibited by the addition of chloramphenicol in the culture.
4. It is easily **purified** in supercoiled form by usual extraction and isolation techniques.

5. It is possible to insert a fragment of DNA of **good size** without exceeding the size of 10 kpb under penalty of plasmid instability.
6. It has two **antibiotic resistance** genes: one for **ampicillin** (ApR), the other for **tetracycline** (TcR), the expression of one of these two genes facilitates the selection of recombinant clones.
7. It also has **twenty** unique sites for restriction enzymes.
8. It is easily **transferable** by transformation or electroporation.

▪ Twin antibiotic resistance screening

If a target DNA fragment is ligated into coding region of one of the resistance genes, the gene will become **insertionally inactivated**, and can be determined by the antibiotic resistance exhibited by the transformants.

- **Amp^R** gene used to discriminate between Transformed and non Transformed **bacteria**.
- **Tet^R** gene used to discriminate between Recombinant DNA in a bacterium and that with non Recombinant DNA.

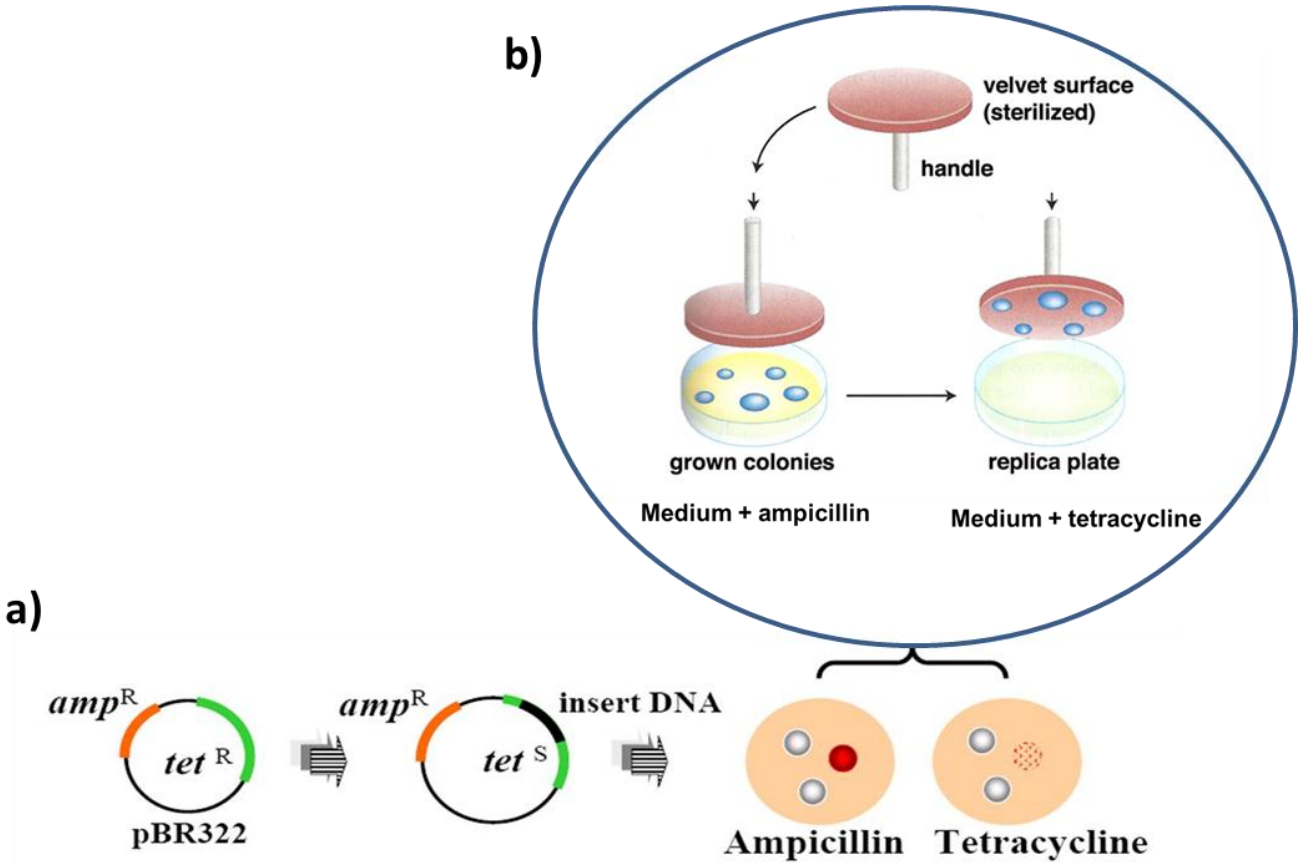


Fig. 24. Twin antibiotic resistance screening (a) using Replica plating method (b).

b. Second Generation Vectors

New generation of increasingly powerful plasmids have been developed since pBR322 and these derivatives. This is the case of the **pUC** family called (second generation cloning vectors).

The second generation cloning vectors are **small plasmids** of about **2700pb**. The plasmid **pUC19** contains the **ampicillin** resistance gene of pBR322, but in addition it has a part of the **lacZ** gene into which a multiple cloning site (**Polylinker**) was introduced, containing a whole series of single cut sites.

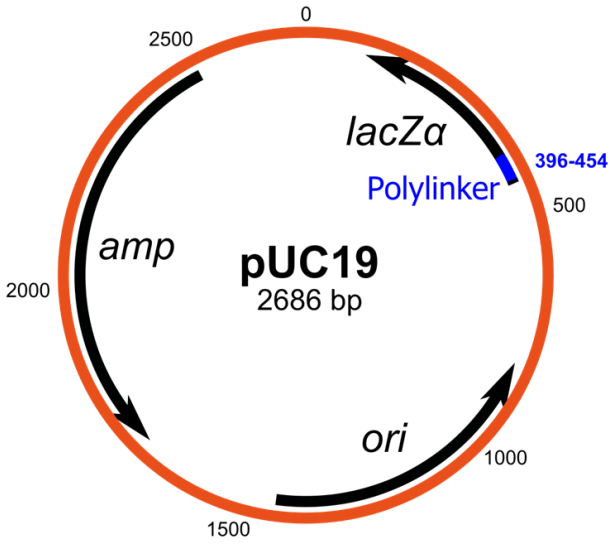


Fig. 25. The plasmide pUC19

- ### Bleu white colony screening

The fact that a polylinker is inserted into the lacZ gene that is involved in lactose catabolism makes it easy to reveal the integration of an insert by «**insertional inactivation**). The use of a coloured inductor such as X-gal (5-bromo-4chloro-3-indonyl-β-D-galactoside), hydrolysis of the latter into 5,5'-dibromo-4, 4'-dichloro-indigo (**blue**), indicates the production of **β-galactosidase**.

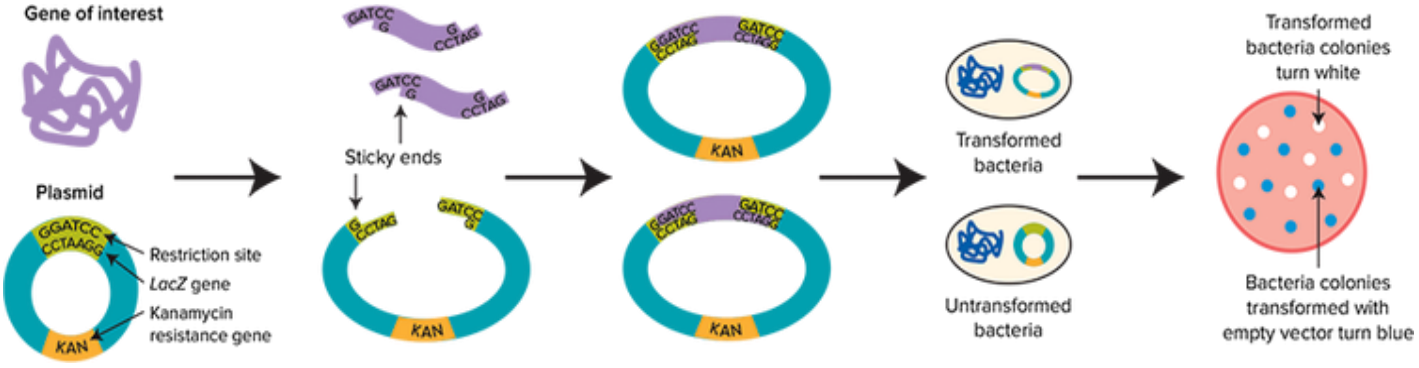


Fig. 26. Bleu white colony screening.

- **Bleu white colony screening reaction**

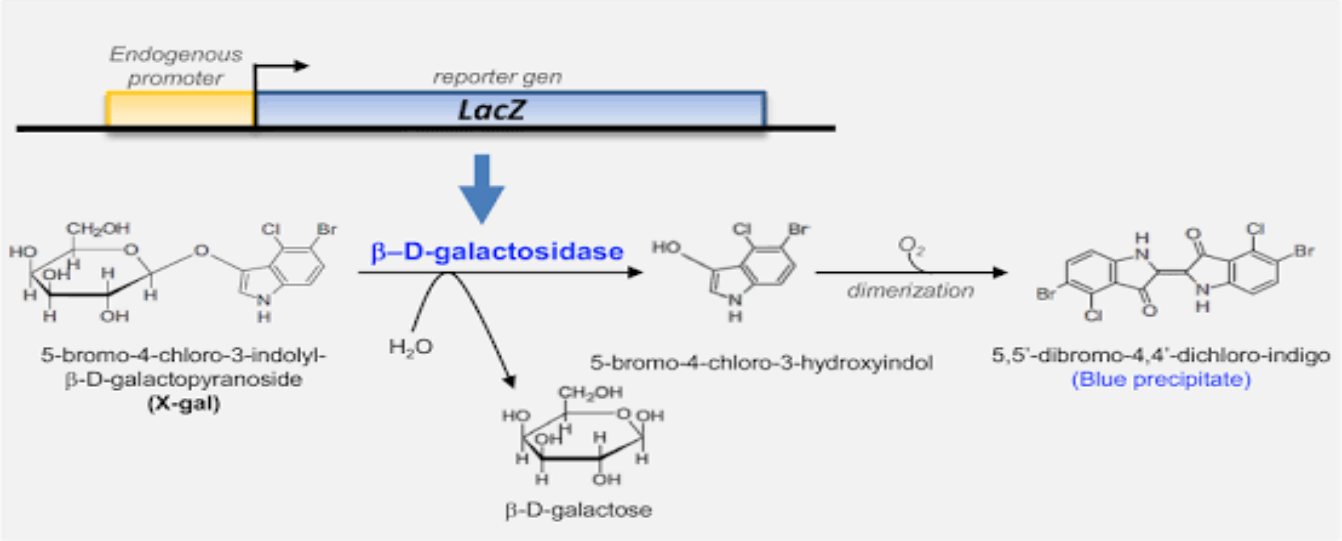


Fig. 27. Bleu white colony screening reaction.

3.1.2. Lambda Bacteriophage

The λ Bacteriophage was discovered by Esther M. Lederberg in 1950. It is an *E. coli* virus, the DNA of this phage is a 48 kb double-stranded linear DNA molecule. At each end 5' is a single-stranded region of 12 nucleotides; one complementary to the other and their association gives a **circular structure** to the DNA in the host cell. The association of these natural cohesive ends forms the site **cos** [cos: Important elements for the **replication** and **encapsidation** of bacteriophage λ] (Fig. 28).

- **Characteristics of λ bacteriophage**

The bacteriophage λ gave rise to the first phagic vectors because:

- Its **Biology** is well **known**
- It is possible to insert up to **22kb**, after elimination of the essential part of the phage life cycle.
- *In vitro* **packaging** of recombinant phagic DNA in phage heads.
- It has a very **fast** host **infection** (transfection) capability.
- The **number** of **copies** per cell is considerable.
- The **yield** of this transfection is much higher than what is obtained when the bacterium is transformed by plasmids.

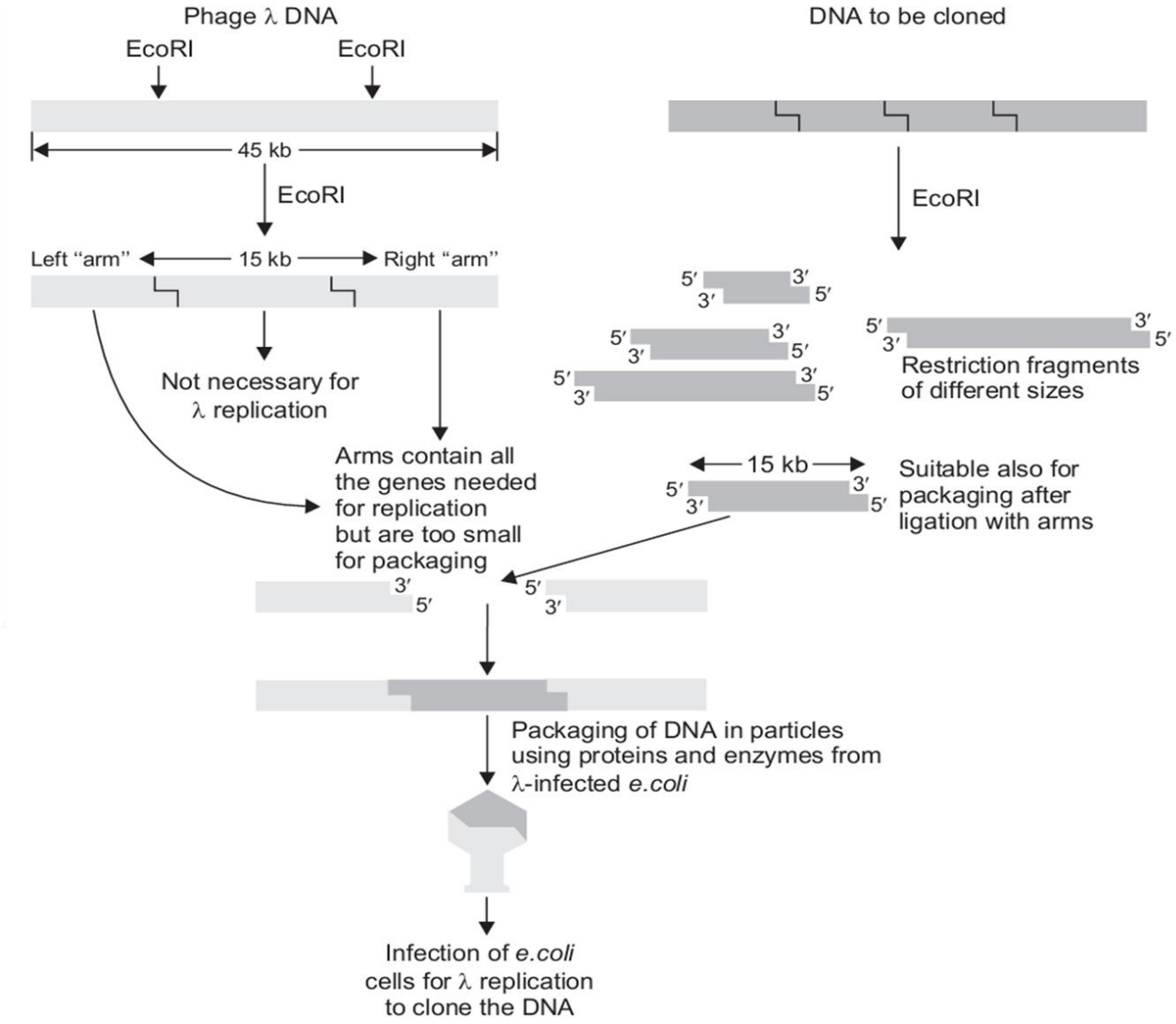


Fig. 28. λ Phage based cloning vectors for cloning inserts of large size.

3.1.3. M13 Bacteriophage

This is a **male-specific** (infects *E. coli* having F. pili), lysogenic filamentous phage with a circular single-stranded DNA genome about 6,407 bp (6.4 kb) in length.

On infection, this molecule is transferred to *E. coli* and converted into the **double-stranded** replicative form (RF). The replication continues and when there are more than 100 copies of DNA in the cell, the DNA replications become asymmetric and produce copies of the original single-stranded molecule, which are packaged into infective particles and extruded from the cell. The host is never lysed but continues to grow throughout the infection, although at a significantly reduced rate.

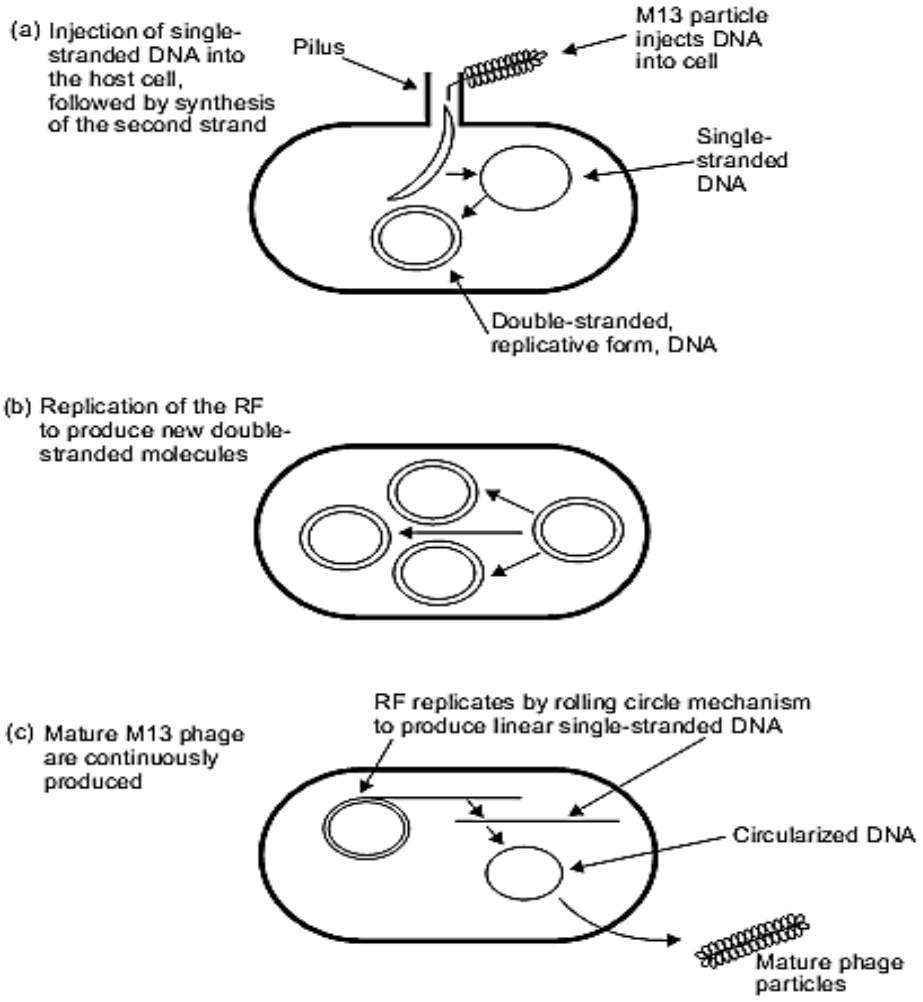


Fig. 29. Infection and life cycle of M13 phage.

▪ Applications of M13 Bacteriophage

M13 has derivatives containing part of the **lacZ gene** (peptide α), with a **polylinker** for 13 unique restriction sites (54pb). This allows the insertion of DNA fragments into these sites, and the selection of white colonies on boxes containing the X-gal analogue.

This vector and these derivatives can be used:

- To **sequence** DNA fragments even from sequences unknown by Sanger technique.
- For **cloning** DNA fragments up to **six times** larger than viral DNA.
- For transfection of competent *E. coli* cells by both forms (SS, DS).
- For generate **directed mutagenesis** (fig. 30).

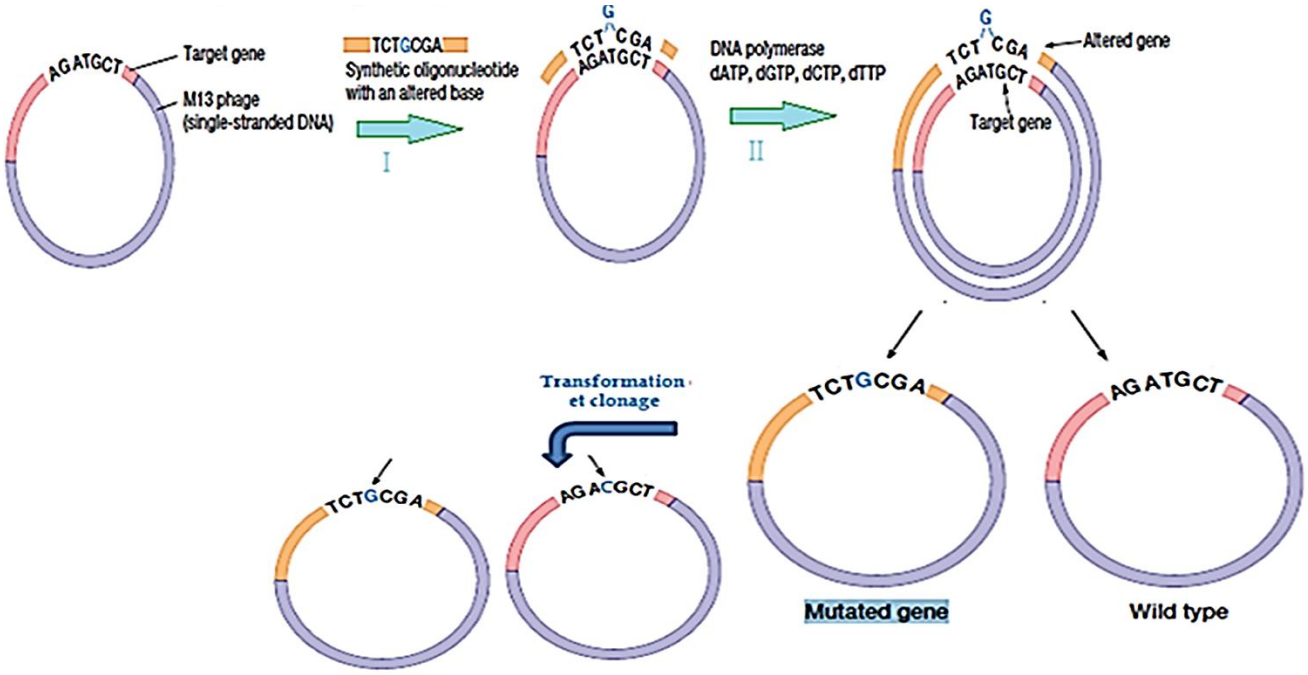


Fig. 30. Site-directed Mutagenesis using M13 phage.

3.1.4. Cosmids

Cosmids are artificial vectors (≈ 5kb) consisting of a classical plasmid to which have been added the **λ phage cos sequences**. These vectors gather both the interesting properties of plasmids as:

- The **origin** of replication
- **Antibiotic resistance** gene And bacteriophage
- *In vitro* **encapsidation** of large DNA fragment.

▪ Characteristics of cosmids

The transformed cells are selected on a medium containing antibiotic (ampicillin). As a conclusion, cosmids allow:

- **Encapsidation** of a modified (recombinant) plasmid into a virion.
- To obtain integration **yields** much higher than those given by bacterial transformation by a plasmid.
- Cloning of a larger DNA fragment (**32 to 47 kb**) than that carried by a plasmid (10 kb) or by the λ bacteriophage (22 kb).
- Requires **fewer** clones to create a genomic bank.
- Cosmids are more **stable** than plasmids.

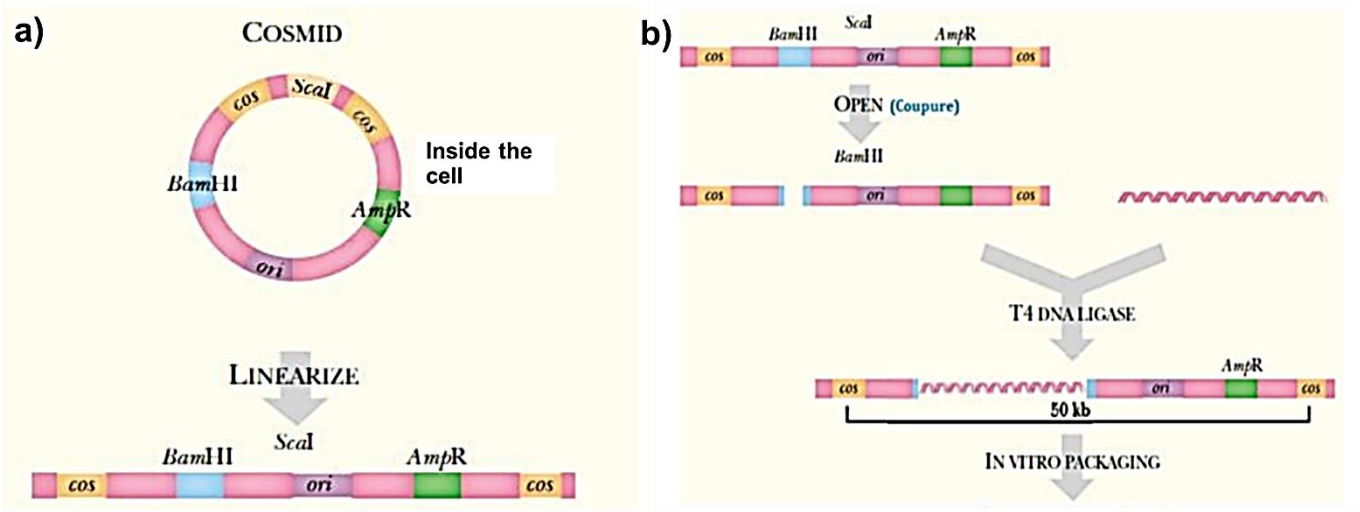


Fig. 31. Circular and linear forms of cosmid (a) and *In vitro* packaging (b)

3.1.5. Bacterial Artificial Chromosomes (BACs)

Bacterial artificial chromosomes (BACs) are the cloning vectors based on the extra-chromosomal plasmids of ***E. coli*** called **F factor** (fertility factor) . These vectors enable the construction of artificial chromosomes, which can be cloned in *E. coli*.

This vector is useful for cloning DNA fragments up to **350 kb**, but can be handled like regular bacterial plasmid vectors, and is very useful for **sequencing large stretches** of chromosomal DNA.

Like any other vector, BACs contain **ori** sequences derived from *E. coli* plasmid F factor, multiple cloning sites (**MCS**) having unique restriction sites, and suitable **selectable markers**.

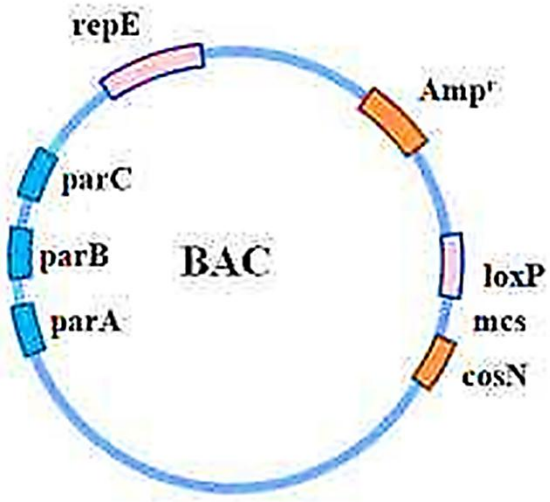


Fig. 32. Bacterial Artificial Chromosome.

3.1.6. Yeast Artificial Chromosomes (YACs)

YACs are the vectors that enable the formation of artificial chromosomes with the **foreign DNA** fragments and cloning into **yeast**. These are used for the cloning of very **large** DNA fragments in the range of 500 to 1,000 kb. YACs (Yeast Artificial Chromosomes) must have:

- A replication **origin**
- **Telomeres** for DNA replication at the ends of the chromosome
- A **centromere** (segregation during mitosis).
- Multiple cloning site or **polylinker**
- Selection **marker**.

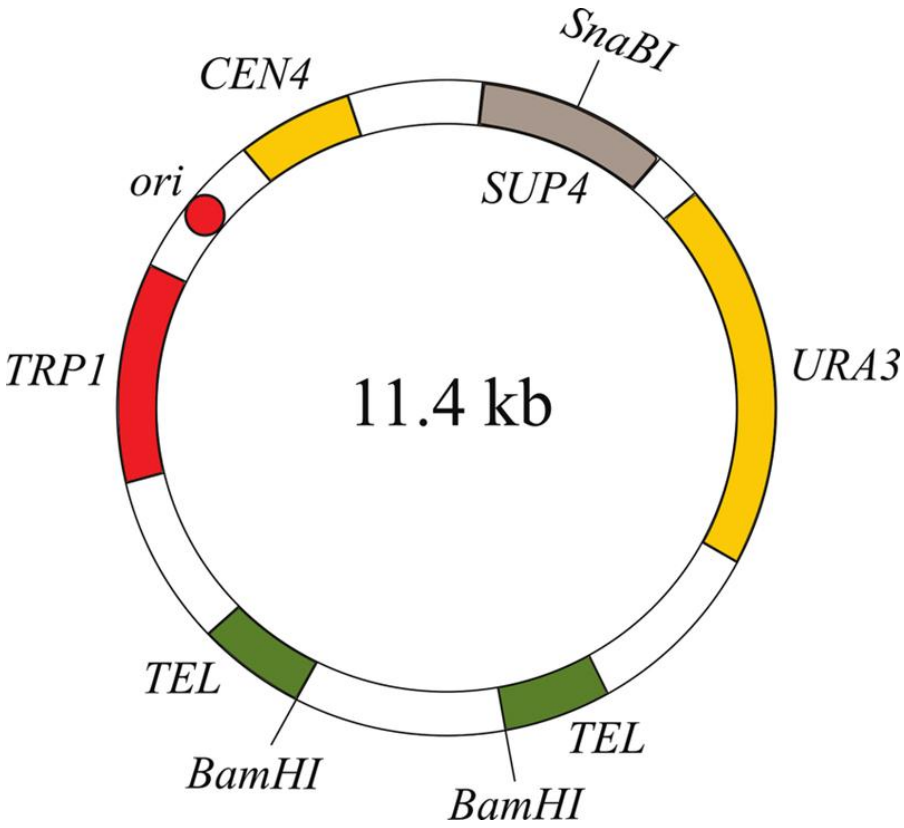


Fig. 33. Yeast Artificial Chromosome.

3.2. Host

Genetic engineering is used for commercial purposes such as the production of new vaccines, large quantities of recoverable proteins, or the introduction of specific genes into an animal or plant organism. In each case, the choice of the host is essential since it will direct us towards a suitable vector.

3.2.1. Type of host cell

The type of host cell used for a particular application will depend mainly on the **purpose** of the cloning procedure.

- If the aim is to isolate a gene for **structural analysis**, the requirements may call for a **simple system** that is easy to use.
- If the aim is to **express** the genetic information in a higher eukaryote such as a plant, a more **specific system** will be required.

These two aims are not necessarily mutually exclusive; often a simple **primary host** is used to isolate a sequence that is then introduced into a **more complex** system for expression. The following table shows types of host used for genetic engineering.

Table 3. Different types of hosts used in genetic engineering.

Major group	Prokaryotic/eukaryotic	Type	Examples
Bacteria	Prokaryotic	Gram –	*Escherichia coli*
		Gram +	*Bacillus subtilis*
			Streptomyces spp.
Fungi	Eukaryotic	Microbial	*Saccharomyces cerevisiae*
		Filamentous	*Aspergillus nidulans*
Plants	Eukaryotic	Protoplasts	Various types
		Intact cells	Various types
		Whole organism	Various types
Animals	Eukaryotic	Insect cells	*Drosophila melanogaster*
		Mammalian cells	Various types
		Oocytes	Various types
		Whole organism	Various types

3.2.2. The ideal host

To obtain large amounts of cloned DNA the ideal host must:

- **Multiply rapidly** in a low-cost growing environment.
- Be **non-pathogenic**.

- Be able to **incorporate** DNA.
- Be **stable** in culture
- Should have **enzymes suitable** for vector replication.

The hosts meeting these criteria are eukaryotic or prokaryotic microorganisms whose genomes are **well known**, fully **sequenced**, and genetically **manipulable**.

3.2.3. The main hosts used for genetic engineering

The following table present the main hosts used for genetic engineering.

Table 4. Main hosts used in genetic engineering.

	Advantages	**disadvantages**
E. coli **(Procaryotic; Gram negative bacteria)**	— Genetically very well known — Many strains available — Prokaryote the most famous.	— Potentially pathogenic — Periplasm could trap proteins
Bacillus subtilus **(Procaryotic; Gram positif bacteria; spore +)**	— Easily transferable — Non-pathogenic — Naturally secreted proteins — Formation of endospores facilitating cultures	— Genetically unstable — Less known genetics compared to *E. coli*
Saccaromyces cerivisae **(yeast; Eucaryotic)**	— Well-known genetically — Non-pathogenic — Ensures the maturation of mRNAs and proteins — Easy to grow	— Unstable plasmids — No replication for most procaryotic plasmids.

- **Characteristics**

— ***E. coli*** (most used organism in molecular cloning) is counted among the **normal flora** of the intestine of humans and animals, it is also a **potential pathogen** (especially wild strains). And also the synthesis of **endotoxins** (LPS) that can contaminate finished products, this is a potential problem especially for intravenous pharmaceuticals. Also the problem, that *E. coli* retains extracellular proteins in its periplasm.

- With ***Bacillus subtilus*** the major disadvantage remains the **difficulty** of **maintaining plasmid replication** in subcultures, which often leads to the loss of cloned DNA.

- The advantage of using ***Saccharomyces cerivisae*** is that it has complex RNAs and post translational systems necessary for the synthesis of gene products from higher organisms. Post translation processes can cause cloning problems.

- **Mammalian cell** culture has **high cost** and large scale production difficulties. In addition, the level of **expression** of cloned genes is often **low** (also for insects, plants, etc.).

3.2.4. Introducing recombinant molecules into Eukaryotic hosts

The use of bacterial hosts for genetic engineering laid the foundation for recombinant DNA technology; however, researchers have also had great interest in **genetically engineering eukaryotic cell**s, particularly those of plants and animals. The introduction of recombinant DNA molecules into eukaryotic hosts is called **transfection**. Genetically engineered plants, called transgenic plants, are of significant interest for agricultural and pharmaceutical purposes.

N.B. In bacteria, the DNA to be cloned can be transferred by three methods: **transformation**, **transduction** and **conjugation.**

a. Electroporation

This technique involves the exposure of the host to **electric discharges** in order to open the **pores** (temporarily) in the membrane through which the cloned DNA, added to the medium, can penetrate without cell lysis (Fig. 34).

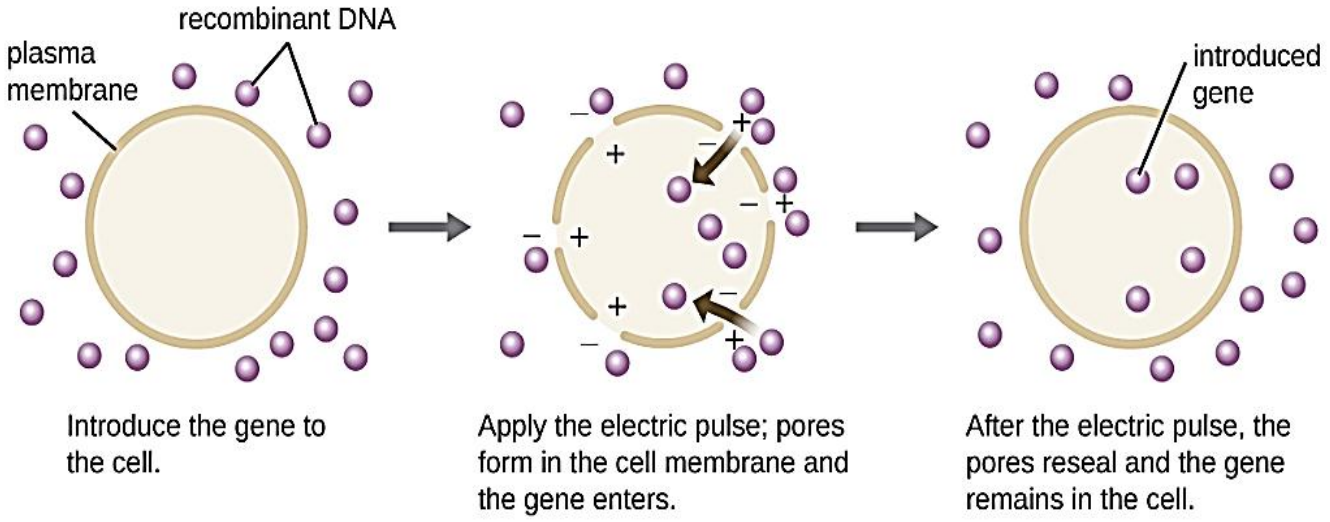


Fig. 34. Electroporation technique.

b. Gene Guns (*biolistics*)

The transfection of target cells is done by **metal beads** (usually tungsten) covered with **nucleic acids**, piercing plasma walls and membranes without causing cell lysis. This technique has been used on yeasts, algae, plant cells, and even mitochondria and chloroplasts. In addition, unlike electroporation, this technique can be used to introduce DNA into intact tissues such as plant seeds (Fig. 35).

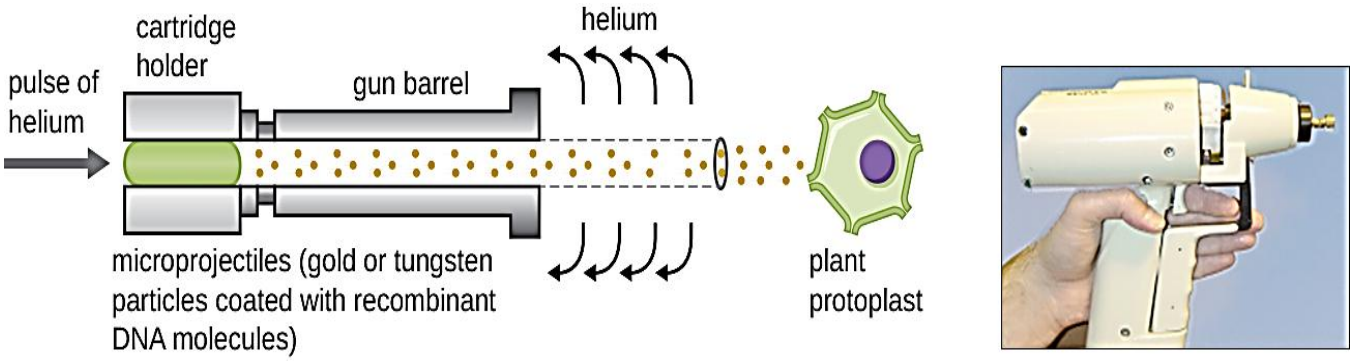


Figure 35. Bombardment with DNA-Coated Microparticles « Gene Gun ».

c. Microinjection

An alternative method of transfection is called microinjection. Because eukaryotic cells are typically **larger** than those of prokaryotes, DNA fragments can sometimes be **directly injected** into the cytoplasm using a glass micropipette (Fig. 36). The microinjection technique is usually used to introduce DNA into **animal cells** (eggs, oocytes, and embryos) or **plant protoplasts**.

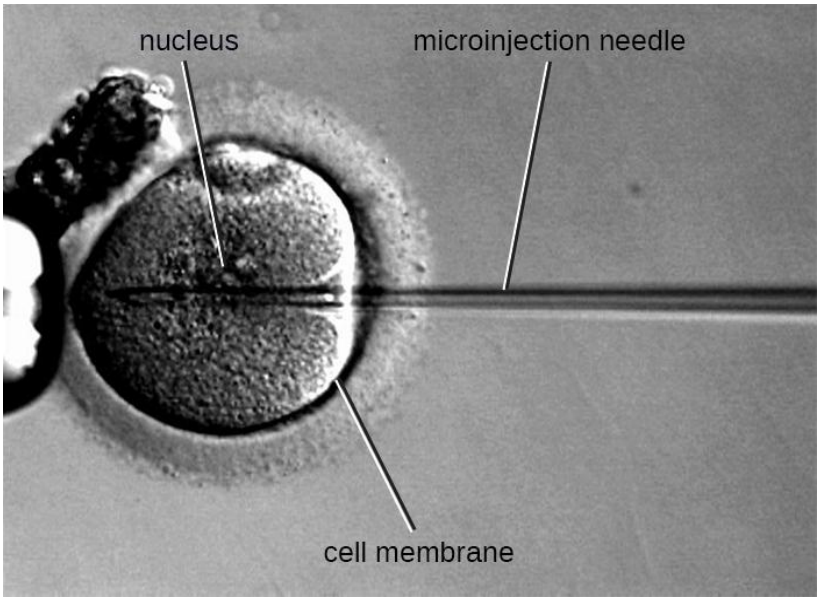


Fig. 36. The microinjection technique

d. Shuttle Vectors

Another method of transfecting plants involves **shuttle vectors**, plasmids that can move between bacterial and eukaryotic cells. The tumor-inducing (T_i) plasmids originating from the bacterium ***Agrobacterium tumefaciens*** are commonly used as shuttle vectors for incorporating genes into plants.

In nature, the T_i plasmids of *A. tumefaciens* cause plants to **develop tumors** when they are transferred from bacterial cells to plant cells. Researchers have been able to manipulate these naturally occurring plasmids to **remove** their **tumor-causing** genes and **insert desirable** DNA fragments (Fig. 37).

The resulting recombinant T_i plasmids can be transferred into the plant genome through the natural transfer of T_i plasmids from the bacterium to the plant host. Once inside the plant host cell, the gene of interest recombines into the plant cell's genome (Fig. 37).

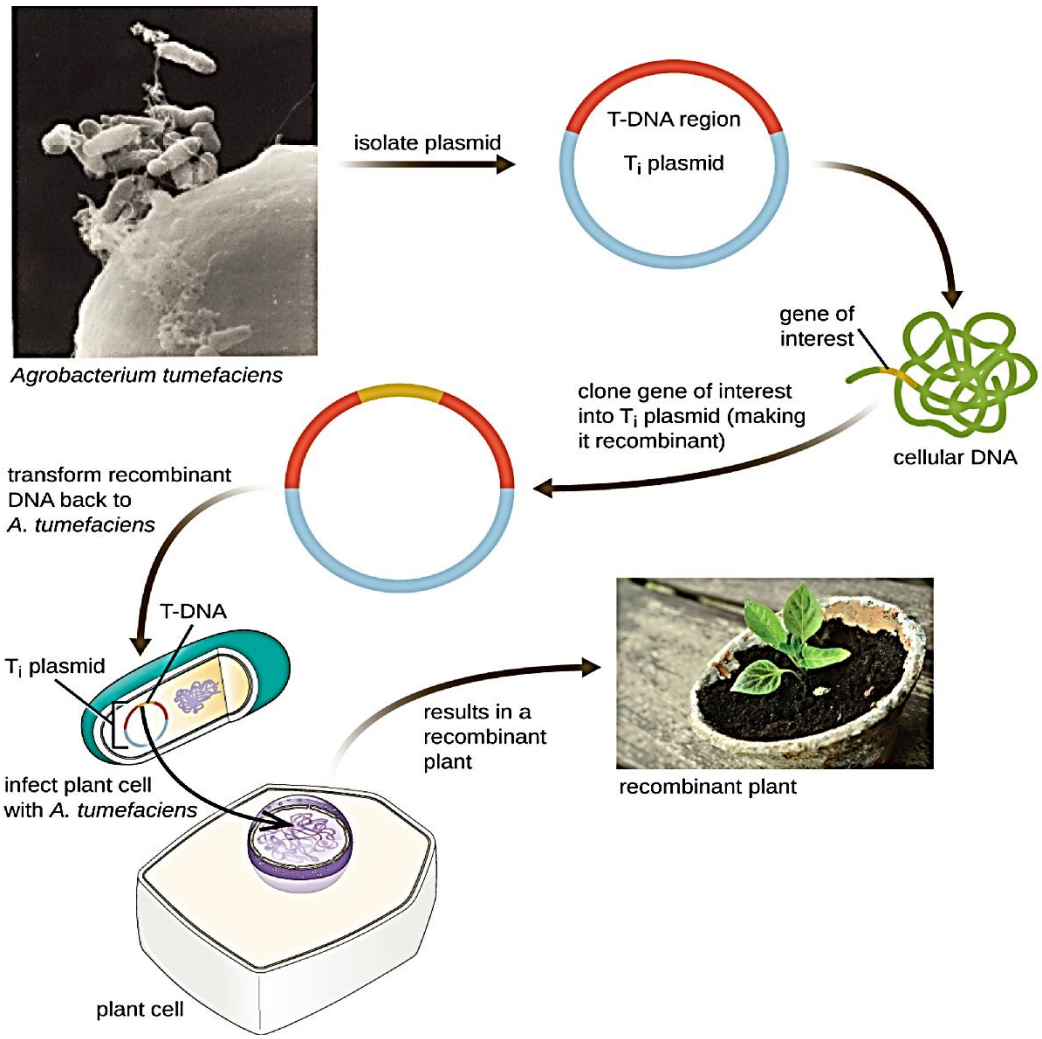


Fig. 37. Transformation of plant cell using a shuttle vector (*Agrobacterium tumefaciens*).

e. Viral Vectors

Viral vectors can also be used to **transfect eukaryotic cells**. In fact, this method is often used in **gene therapy** to introduce healthy genes into human patients suffering from diseases that result from genetic mutations.

Viral genes can be deleted and replaced with the **gene to be delivered** to the patient; the virus then infects the host cell and delivers the foreign DNA into the genome of the targeted cell.

Adenoviruses are often used for this purpose because they can be grown to high titer and can infect both nondividing and dividing host cells. However, use of viral vectors for gene therapy can pose some risks for patients (Fig. 38).

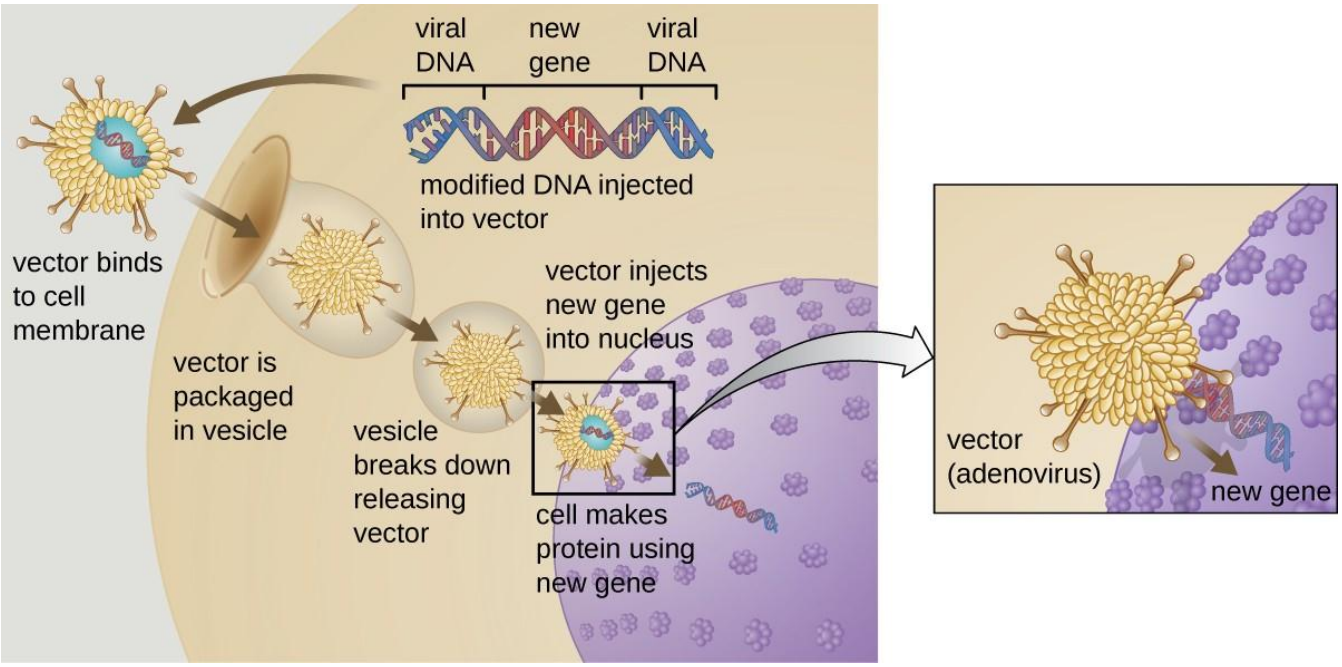


Fig. 38. Process of introducing foreign DNA into human cells by using adenovirus.

CHAPTER 4

MOLECULAR HYBRIDIZATION, PROBES AND DNA LABELLING

Molecular hybridization is a fundamental property of nucleic acids based on the rules of complementarity. For the case of two complementary strands of the same DNA molecule (homologous strands) we speak of renaturation (100% hybridization). Molecular hybridization is used mainly in the detection of homology between DNA molecules from different sources. Complementarity will depend on specificity and sensitivity.

4.1. Denaturation

During denaturation process, we remove the secondary or primary structure of the DNA or RNA. Heat to separate the two strands (at melting temperature or Tm: melting).

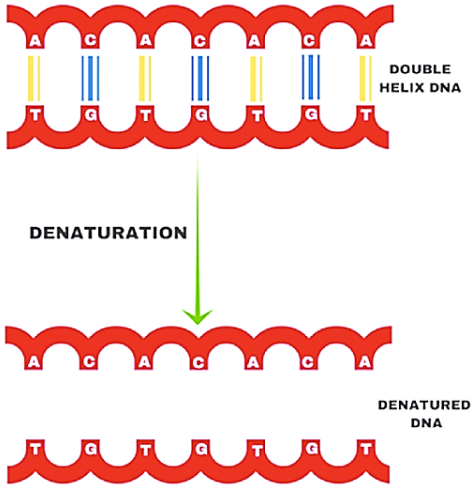


Fig. 39. Thermal denaturation of a bicatenary DNA into a single-stranded DNA form.

4.1.1. DNA melting temperature

- This is a temperature, also called **half denaturation temperature** (Tm) which corresponds to the opening or unwinding of **50%** of the heated DNA chain.
- It is the **hyperchromic** effect that corresponds to the **rupture** of hydrogen bonds (weak bonds) and the separation of the 2 chains leads to an increase in **U.V absorption of 40% at 260 nm**. The DNA Tm value is a linear function of the percentage of (G + C) DNA:
- The result of the change from bicatenary to single-stranded DNA is an increase in the extinction coefficient of DNA: this is the **hyperchrome effect**. The nitrogenous bases that absorb at this wavelength are very orderly in the bicaténaire form and each base is no longer an isolated absorbent center. However, the nitrogenous bases of the single strand DNA hide less and have no more ordered structure (minimal quenching). The molecular extinction coefficient is higher.

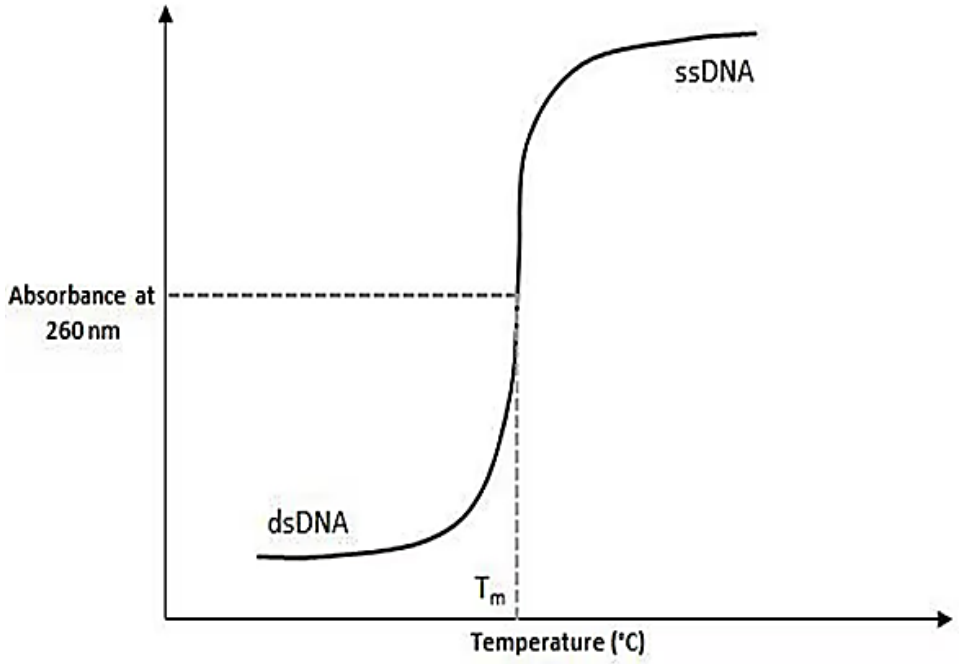


Fig. 40. Oligonucleotide melting temperature of double-stranded DNA (dsDNA) and Single-stranded DNA (ssDNA)

Factors that can influence the Tm value

- **The number of H-H** links depends on the fragment length (up to 150 links). This number is important for small fragments (50 to 60 bases): a long DNA contains more bonds to dissociate than a shorter DNA.
- **The composition in GC / AT** bases: the number of hydrogen bonds is not the same for each base couple. It is therefore obvious that the number of bases will be a significant factor in the calculation of Mt.
- **The presence of mis-matching**: In general, the lowering of the Tm is 1 °C for a value of 1% mismatch.
- **The composition of the medium**: an increase in monovalent salts elevates Tm. An increase of urea, formamide allows decreasing the Tm.
- **Extreme pH decreases** Tm.

Note: an empirical formula can be used that groups all these parameters together for fragments smaller than 100 bps:

Tm = 16,6 log [M] + 0,41 (% G+C) + 18,5 – (% mismatch) – (675 / length in bases) – 0,65 (% of formamide)

With [M] = Na+ ion concentration

▪ Properties changed when denaturing DNA

- Its **density** increases; in isopycnic centrifugation, single strand DNA is at the level of RNA.
- **Hydroxyapatite** (calcium phosphate) **columns** retain double strand DNA and allow the single strand to pass through.
- **Nylon** or **nitrocellulose** membranes retain the single strand and allow the double strand to pass through.
- Hybridization depends on **temperature**, **concentration** and **time**.

4.2. Hybridization

If the separation of the strands of DNA is followed by a gradual and slow cooling, there will be gradual reassembly of the two complementary strands of DNA (This is the property of a monostrand DNA molecule to spontaneously associate itself in a specific and reversible way with another monostrand molecule if it is complementary to it): this is the phenomenon of **hybridization**. This **reassociation** can be carried out between **DNA** and **RNA** sequences which allow obtaining more stable **DNA/RNA** hybrids. The separation of the hybridization products (single strand DNA, double strand DNA and DNA/RNA hybrid) is done by centrifugation in a Cesium chloride concentration gradient.

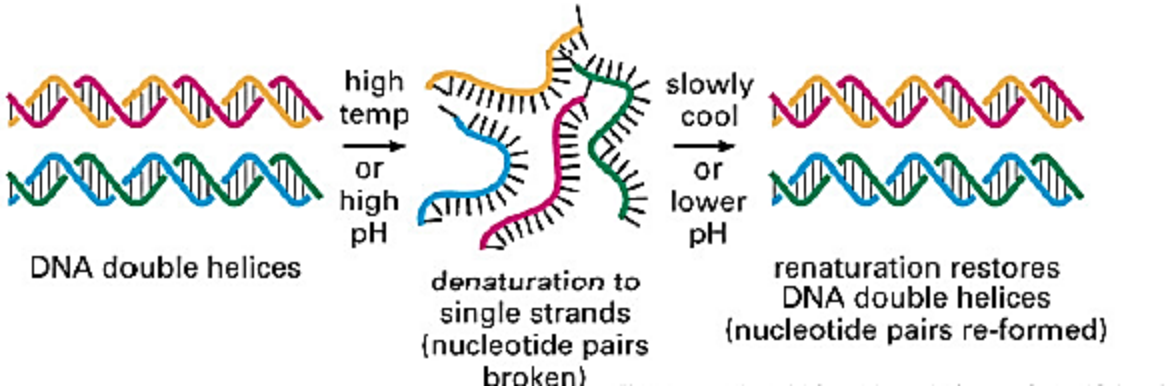


Fig. 41. Denaturation and Renaturation.

▪ Factors that can influence the phenomenon of hybridization

- **DNA concentration and time**: concept of CoT and RoT. The hybridization of nucleic sequences is random. However, if the concentration of DNA is large, the number of hybridized copies will be large. As a result, the rate of hybridization increases as the DNA concentration increases. It is the same for the time factor: the probability of association of the complementary strands is important when the time is long.

- **Temperature**: it favors the meeting of the two complementary sequences, thus the speed of hybridization.

- **The size of the nucleic fragme**nt: If the two complementary sequences are perfectly identical, the reassociation speed of these two strands increases proportionally with the length of the fragments considered.

- **The nature of nucleic acid**s: The rate of reassociation depends on both the nature of the hybrid and its concentration. If the hybridization involves a DNA/DNA mixture and if the DNA is in excess of the RNA, the RNA/DNA hybridization rate is 5 times lower than the DNA/DNA reassociation. On the other hand, if the RNA is in large excess, the rate of reassociation will be identical to that of DNA/ DNA.

- **Ionic strength**: NaCl concentration plays an important role in the reassociations of the complementary segments s. For a NaCl concentration close to 1 M, the rate of reassociation can be multiplied by 10.

▪ The different types of hybridization

a. Liquid phase hybridization

The complementary segments are placed in a solution containing a buffer and **formamide**. The thermal agitation ensures the connection between the complementary fragments. This temperature is generally **15 °C** lower than the **Tm** of the DNA concerned.

Note: For the same DNA concentration, the smaller (or less complex) the genome, the faster the renaturation.

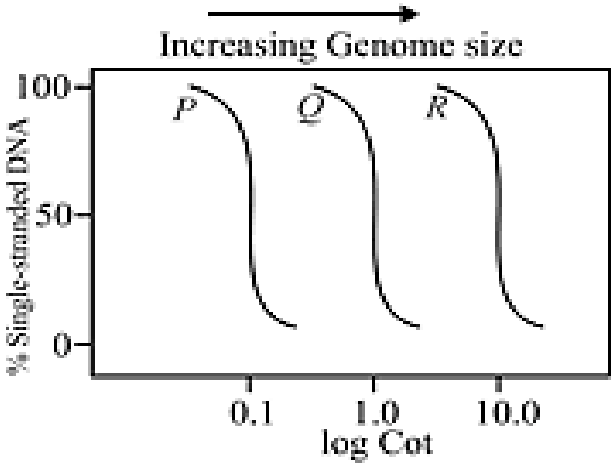


Fig. 42. Realtion between the length of the DNA and the necessary time for renaturation.

b. Hybridization on solid support

The target complementary sequence is fixed (immobilized) on a **solid** support. This method facilitates the **separation** of hybridized and non-hybridized fractions. However, the hybridization rate is significantly lower than that of the liquid phase (up to 10 times).

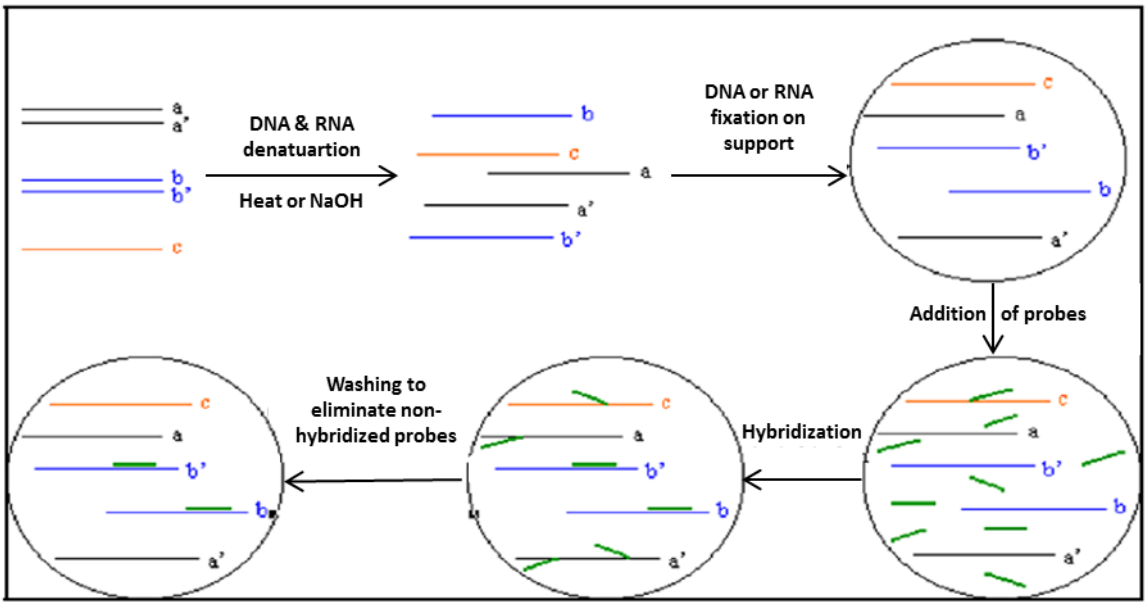


Fig. 43. Denaturation and probes hybridization on solid support.

Hybridization on a support is carried out with several types of supports that make it possible to immobilize the nucleic acid strands:

- **Nitrocellulose:** is the **first** immobilization medium used for nucleic acid fixation. This support, requires a high **ionic concentration**, which makes it possible to create **irreversible** bonds under vacuum and at 80 °C.

- **Synthetic membranes** (nylon based): Similarly, they require strong **ionic strength**. These membranes allow more stable bonds than those obtained with nitrocellulose because of their treatment by **short UV rays** (254 nm). This type of bond, will allow more their dissociations and hybridizations.

c. In situ hybridization (HIS)

It is a technique that allows, through the use of probes, to **highlight** and **locate**, in **cells** or **tissues**, sequences of known nucleic acids. It is very close, in principle, to the Southern and Northern Blot and relies, like them, on the hybridization of a nucleic acid probe (DNA or RNA) marked with a complementary sequence of nucleic acids that is sought to identify and locate. The only difference is that the Southern and Northern Blots are made on tissue crushes, while the HIS is made on a **histological** section of tissue.

In the laboratory, the main applications of HIS are the **localization** of genes on metaphase chromosomes and the search for **bacteria** that have integrated a plasmid or a recombinant phage (selection).

Several probes can be used to perform HIS: DNA (double strand or more rarely monostrand) or RNA-messenger or synthetic oligonucleotides (20 to 50 nucleotides). This figure explains the principle of hybridization by FISH (In situ fluorescent hybridization technique).

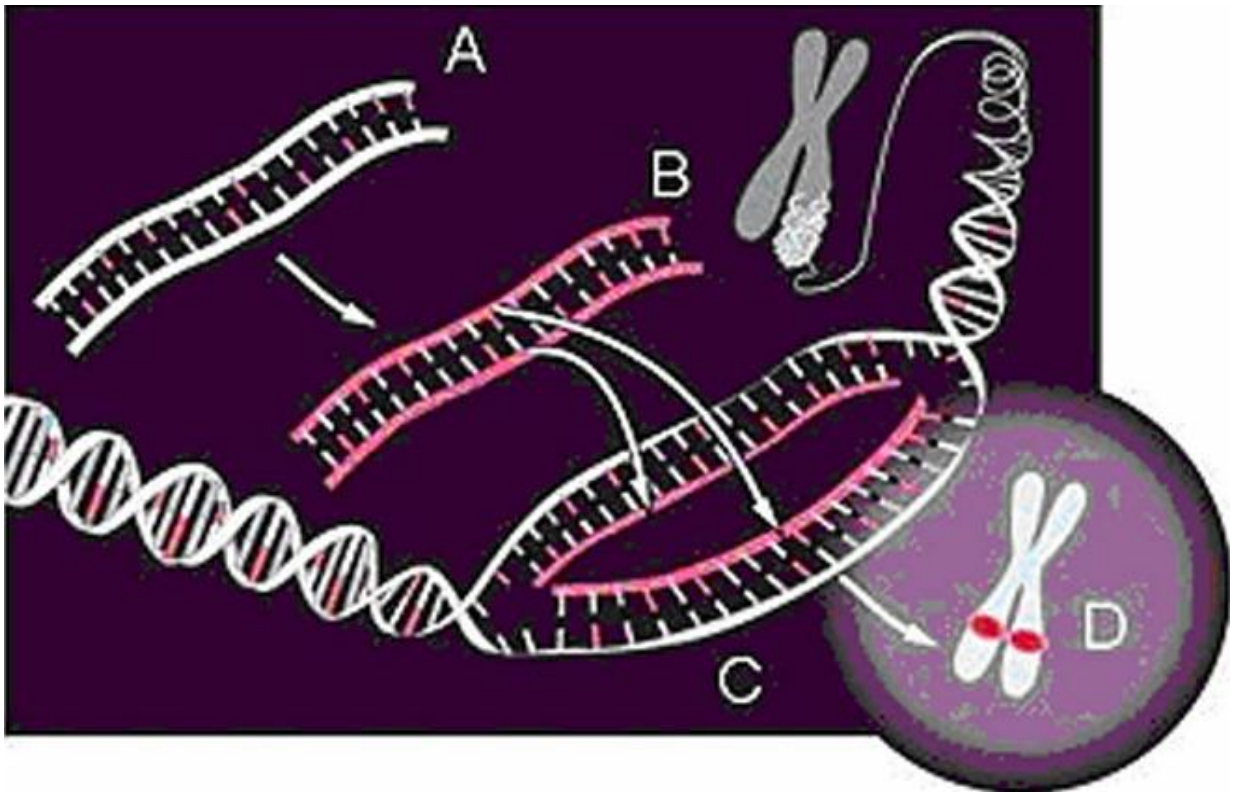


Fig. 44. *In situ* fluorescent hybridization technique. In **A:** probe. **B**: probe colored with fluorochrome. **C**: hybridization with nuclear DNA. **D**: appearance of the metaphasic chromosome where the probe is attached.

4.3. Probes and DNA labelling

A major problem encountered in many cloning procedures is that of keeping track of the small amounts of nucleic acid involved. This problem is magnified at each stage of the process, because losses mean that the amount of material usually diminishes after each step. One way of tracing the material is to label the nucleic acid with a marker of some sort, so that the material can be identified at each stage of the procedure. So what can be used as the label?

Probe: A small DNA or RNA sequence marked (by a fluorescent compound, radioisotope or enzyme) that is used to hybridize and detect homologous complementary sequences. Hybridization can occur both *in vivo* and *in vitro* and involves fragments of 15 nucleotides. Probes must be specific and sensitive. They associate only with a complementary sequence and are easily identifiable and quantifiable.

There are actually several strategies for marking probes:

3.4.1. End labelling

In the end labelling technique, the enzyme polynucleotide kinaseis used to transfer the terminal phosphate group of ATP onto 5-hydroxyl termini of nucleic acid molecules. If the ATP donor is radioactively labelled, this produces a labelled nucleic acid of relatively low specific activity, as only the termini of each molecule become radioactive (Fig. 45).

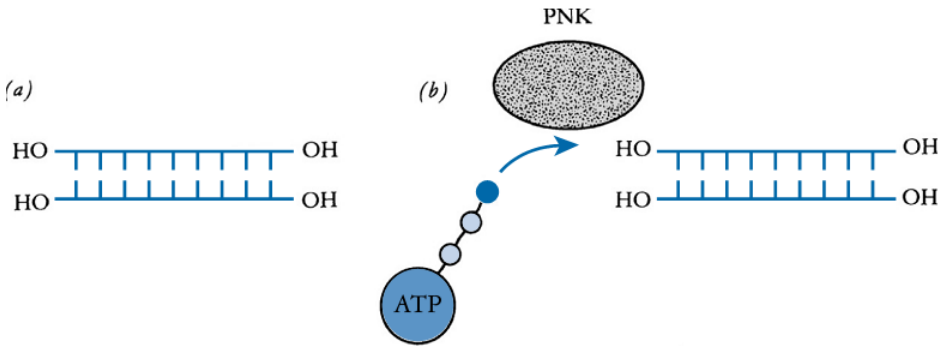


Fig. 45. End labelling DNA using polynucleotide kinase (PNK). (a) DNA dephosphorylatin. (b) The terminal phosphate of [γ-32P]ATP (solid circle) is then transferred to the 5' terminus by PNK.

3.4.2. The Nick translation

DNase I randomly cut through the strands of bicatenary DNA. The ends 3'OH thus released become a binding site for DNA polymerase I that will:

1. Destroy DNA through activity exonucleasic (direction 5'→ 3')

2. Resynthesize (polymerase activity) a new chain with nucleotides of which at least one is radioactive present in the enzymatic incubation medium.

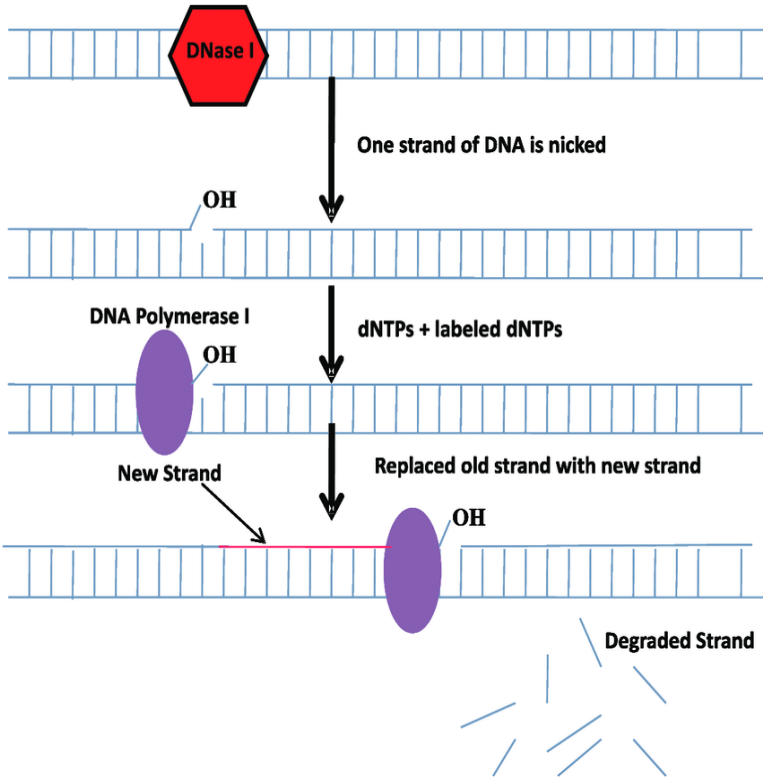


Fig. 46. Nick translation labelling method.

3.4.3. Multirandom priming

The dual-stranded DNA is denatured by heating at 100 °C to separate the two strands, and then cooled abruptly to prevent the two strands from reassociating.

Each of the two strands is randomly hybridized at 25 °C with small primers (6 nucleotides). The sequences of these hexa-nucleotides are different. It is a combination of 4 nucleotides among 6 different positions; so there will certainly be some hexa-nucleotides that will hybridize with both of DNA.

The enzyme DNA polymerase I (Klenow fragment), ensures the synthesis of the two complementary strands in the direction 5' → 3' using the nucleotides triphosphates (ATP, CTP, TTP, and GTP) of which one is marked. The neosynthesized DNA strands are then marked:

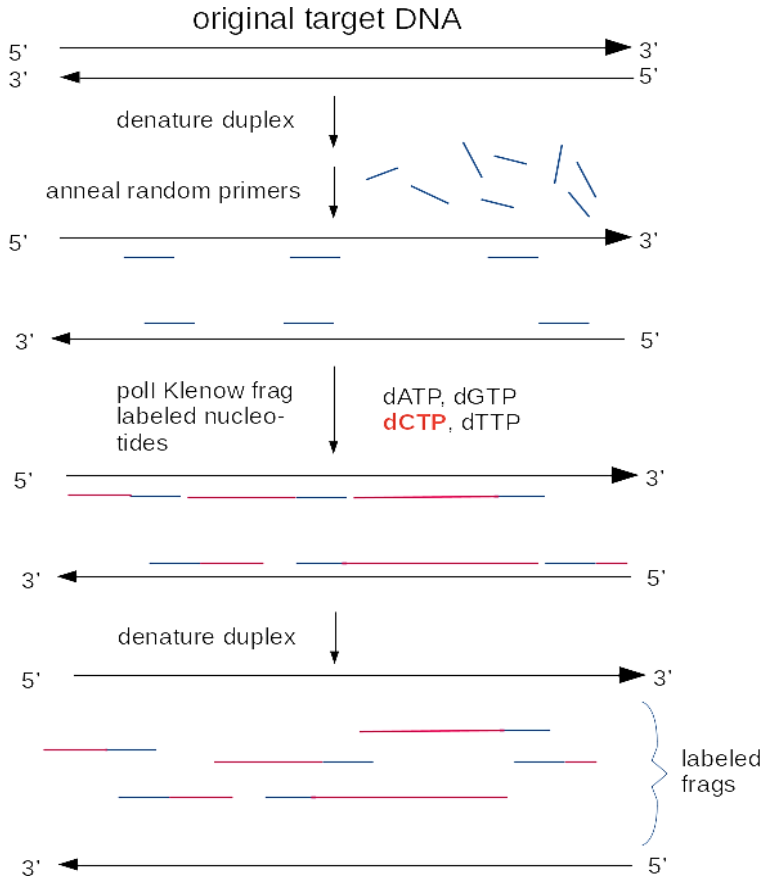


Fig. 47. Multirandom priming labelling technique.

- **How DNA probes are build?**

It is possible to build a DNA probe long before knowing the gene sequence itself! This is done by working with the protein product of the gene. Remember that in our description of protein manufacturing, we said that a gene was transcribed into messenger RNA (mRNA), according to the

simple rules of complementarity of bases. Messenger RNA is transported out of the nucleus and used as a matrix for the formation of an amino acid chain, which turns into a protein.

We can isolate the protein produced by the gene we are interested in, and find out what are the first 30 amino acids of the protein. Based on this information, we can determine the first 90 nucleotides of this protein messenger RNA matrix (remember that each nucleotide triplet codes for an amino acid, hence the 90:30 ratio). And since the messenger RNA matrix is complementary to the sought gene, we know that our DNA probe should have a complementary sequence of the first 90 nucleotides of the sought gene. To build a DNA probe, we use a «gene machine» capable of synthesizing in just a few hours a short single-stranded DNA molecule containing the desired nucleotide sequence.

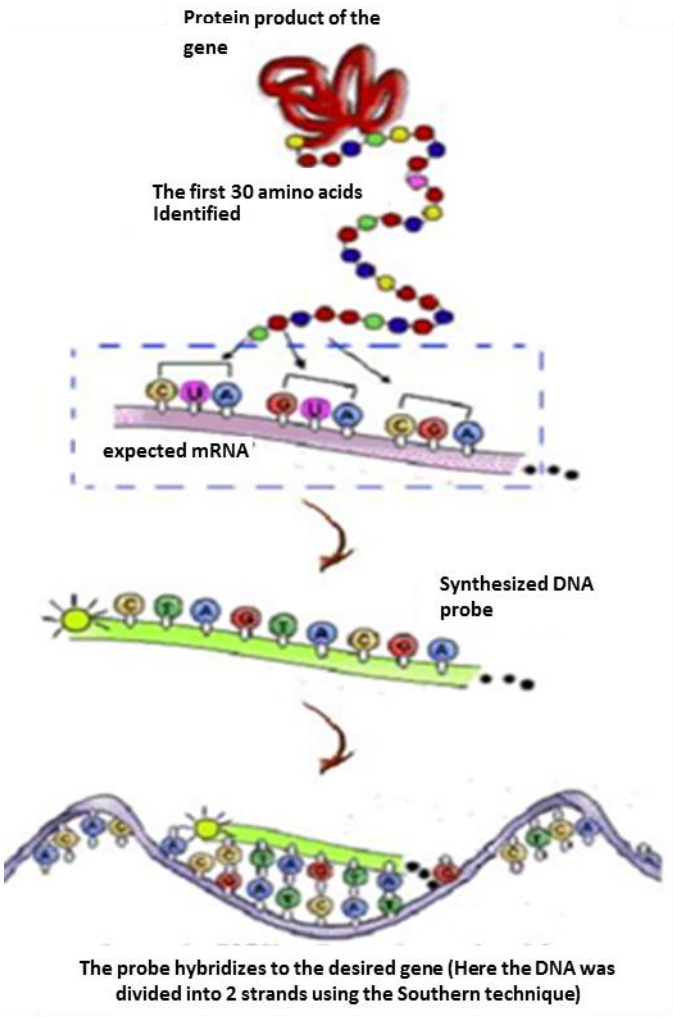


Fig. 48. Process of making probes based on peptide composition.

- **Example: use of DNA probes:** search for a gene of interest associated with the suspected pathogen.

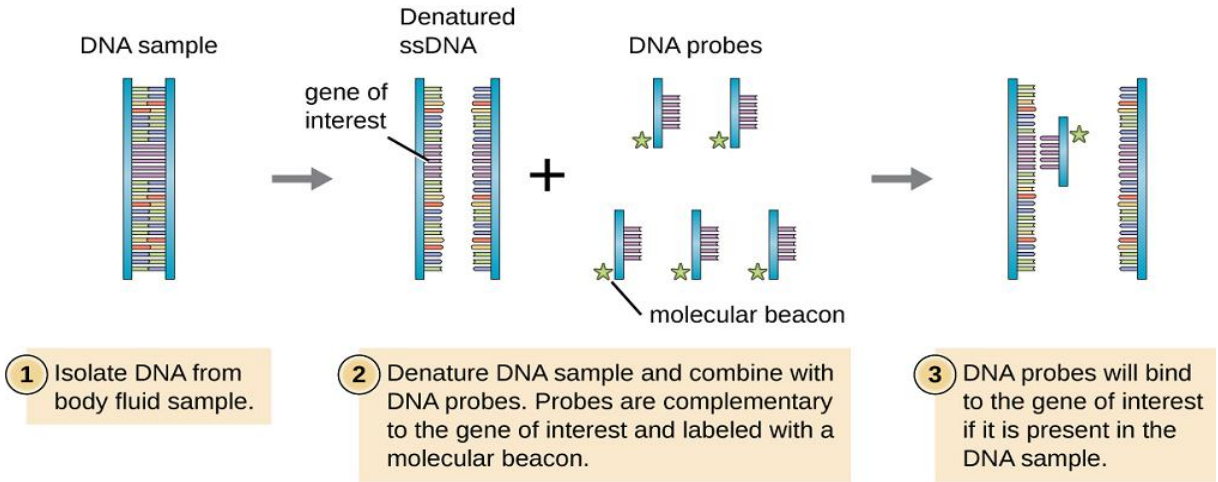


Fig. 49. DNA probes can be used to confirm the presence of a suspected pathogen in patient samples.

CHAPTER 5

GENE AMPLIFICATION, EXPRESSION AND MODIFICATION TECHNIQUES

5.1. Gel electrophoresis

This technique is **vital** to the genetic engineer, as it represents the **main** way by which nucleic acid fragments may be **visualised directly**. The method relies on the fact that nucleic acids are polyanionic at neutral pH; that is, they carry multiple **negative charges** because of the **phosphate** groups on the phosphodiester back-bone of the nucleic acid strands. This means that the molecules will migrate towards the positive electrode when placed in an electric field. As the negative charges are distributed evenly along the DNA molecule, the charge/mass ratio is **constant**; thus, mobility depends on **fragment length**. The technique is carried out using a gel matrix, which separates the nucleic acid molecules according to size (Fig. 50).

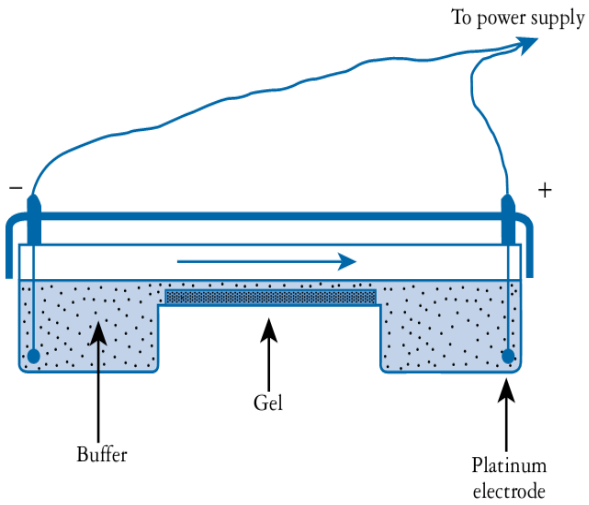


Fig. 50. A typical system used for agarose gel electrophoresis. The gel is just covered with buffer.

The type of matrix used for electrophoresis has important consequences for the degree of separation achieved, which is dependent Separation of biomolecules by on the porosity of the matrix. Two gel types are commonly used: **agarose** and **polyacrylamide**. Agarose is extracted from seaweed and can be purchased as a dry powder that is melted in buffer at an appropriate concentration, normally in the range 0.3--2.0% (w/v). On cooling, the agarose sets to form the gel. Agarose gels are usually run in the apparatus shown in Fig. 51, using the **submerged agarose gel electrophoresis** (**SAGE**) technique (Fig. 50). **Polyacrylamide-based gel electrophoresis** (**PAGE**) is sometimes used to separate small nucleic acid molecules; in applications such as DNA sequencing, as the pore size is smaller than that achieved with agarose. The useful separation ranges of agarose and polyacrylamide gels are shown in Table 5.

Table 5. Seperation characteristics of agarose and polyacrylamide gels.

Gel type	Separation range (base pairs)
0.3% agarose	50000 to 1000
0.7% agarose	20000 to 300
1.4% agarose	6000 to 300
4% polyacrylamide	1000 to 100
10% polyacrylamide	500 to 25
20% polyacrylamide	50 to 1

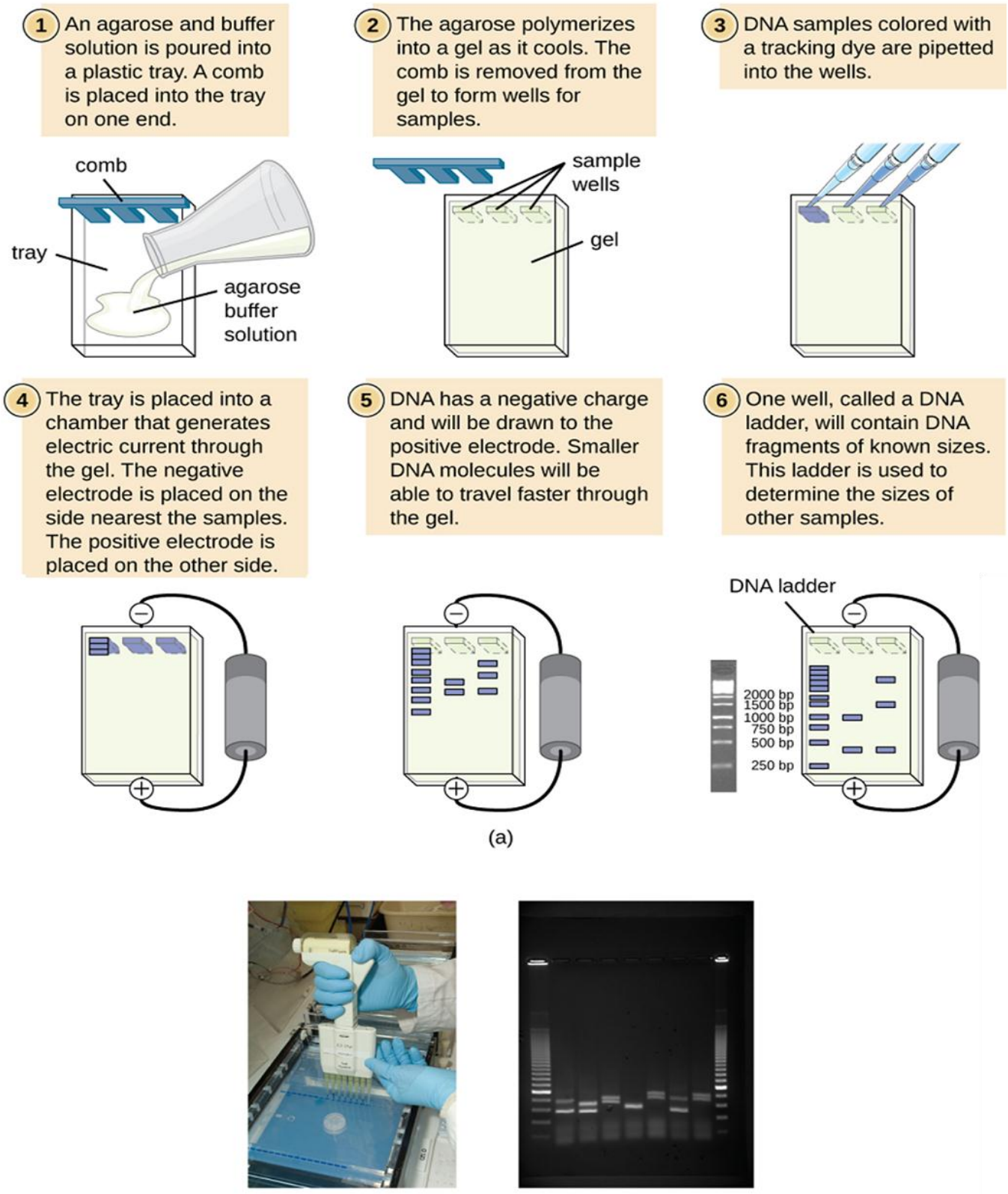


Fig. 51. The process of agarose gel electrophoresis.

In addition to its use in the analysis of nucleic acids, **PAGE** is used extensively for the analysis of **proteins**. The methodology is different from that used for nucleic acids, but the basic principles are similar. One common technique is SDS-PAGE, in which the detergent SDS (sodium dodecyl sulphate) is used to denature multisubunit proteins and cover the protein molecules with negative charges. In this way the inherent charge of the protein is masked, and the charge/mass ratio becomes constant. Thus, proteins can be separated according to their size in a similar way to DNA molecules. PAGE can be further modified to separate proteins based on **two characteristics**, such as their charges at various **pHs** as well as their **size**, through the use of **two-dimensional PAGE**. In any of these cases, following electrophoresis, proteins are visualized through staining, commonly with either **Coomassie blue** or a silver stain.

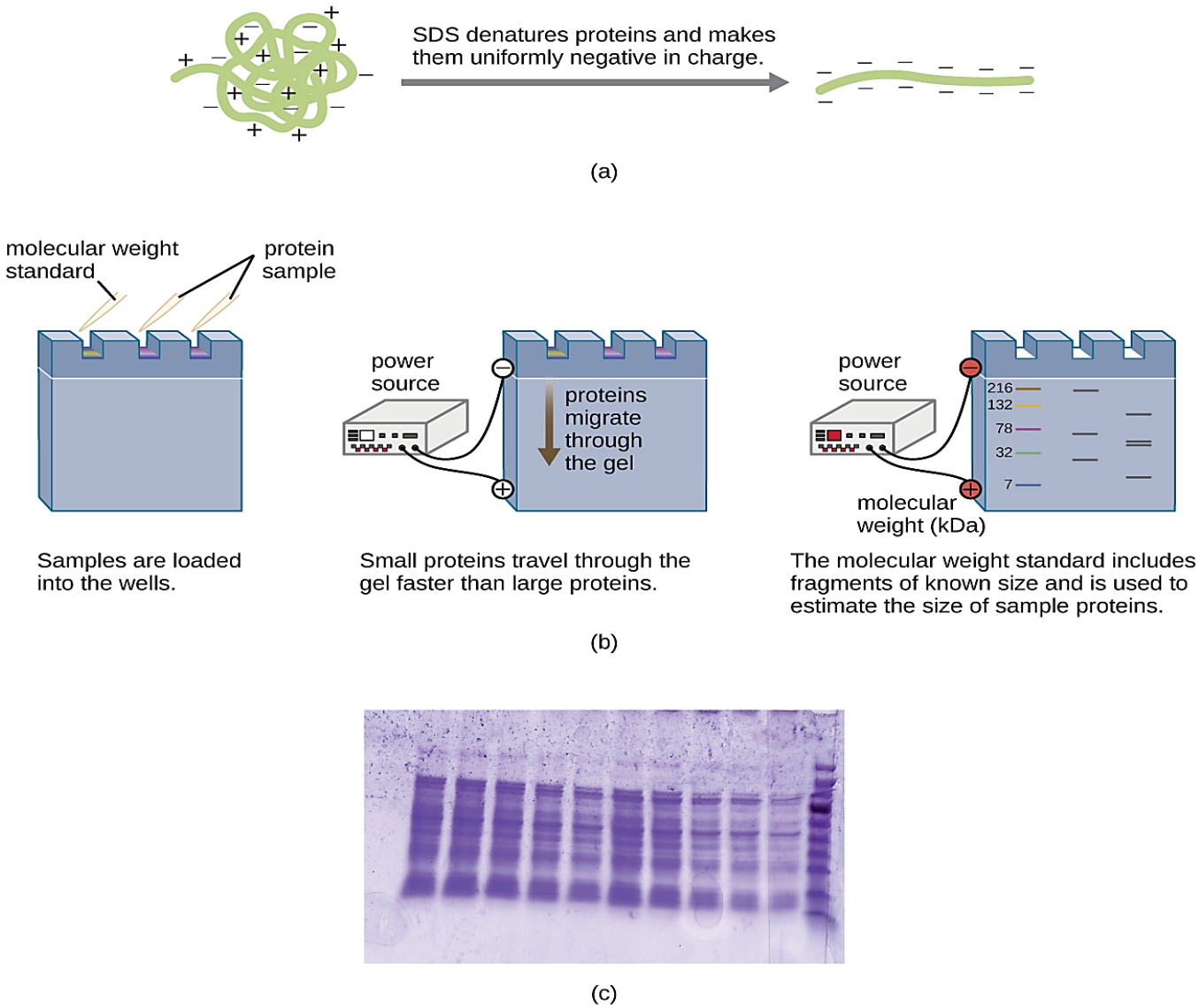


Fig. 52. SDS-PAGE protocol. (a) SDS detergent denatures proteins and making them uniformly negatively charged. (b) The process of SDS-PAGE. (c) A photograph of an SDS-PAGE gel shows Coomassie stained bands.

NOTE. *PAGE is typically performed using a vertical gel apparatus.*

5.2. Restriction Fragment Length Polymorphism (RFLP) Analysis

Restriction enzyme recognition sites are short (only a few nucleotides long), sequence-specific palindromes, and may be found throughout the genome. Thus, differences in DNA sequences in the genomes of individuals will lead to differences in distribution of restriction-enzyme recognition sites that can be visualized as distinct banding patterns on a gel after agarose gel electrophoresis.

Restriction fragment length polymorphism (RFLP) analysis compares DNA banding patterns of different DNA samples after restriction digestion (Fig. 52).

RFLP analysis has many practical applications in both medicine and forensic science. For example, epidemiologists use RFLP analysis to track and identify the source of specific microorganisms implicated in outbreaks of food poisoning or certain infectious diseases. RFLP analysis can also be used on human DNA to determine inheritance patterns of chromosomes with variant genes, including those associated with heritable diseases or to establish paternity.

Forensic scientists use RFLP analysis as a form of DNA fingerprinting, which is useful for analyzing DNA obtained from crime scenes, suspects, and victims by comparing the banding patterns of samples collected from the crime scene against those collected from suspects or victims.

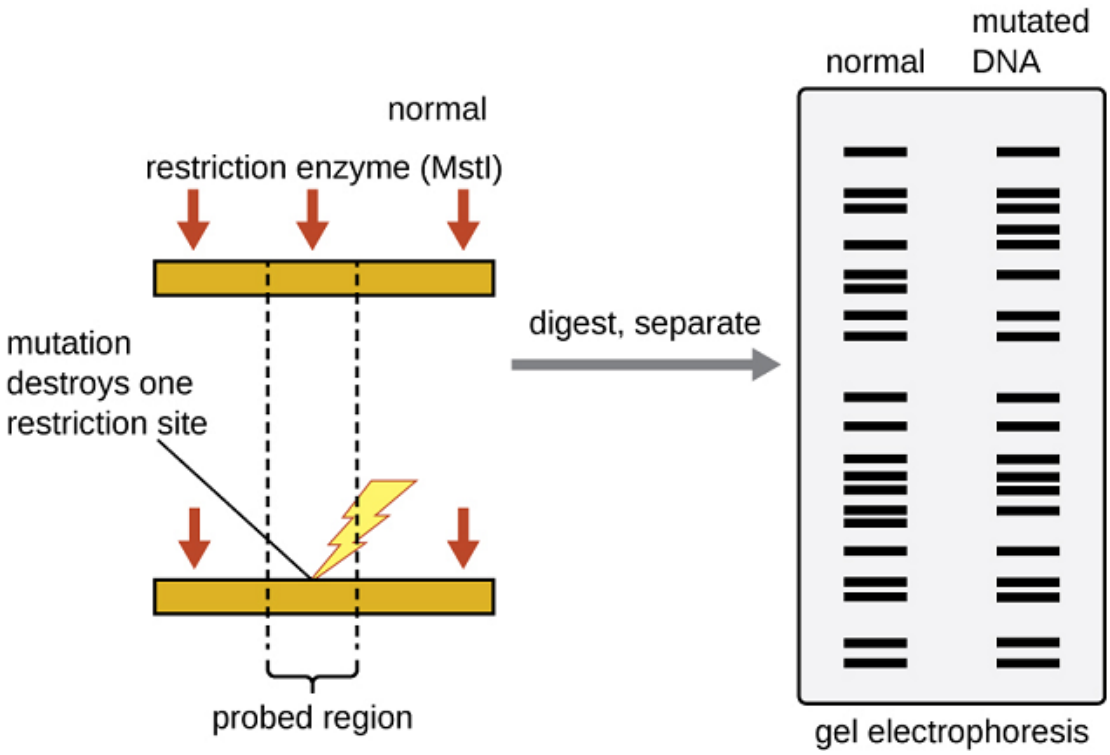


Fig. 53. RFLP analysis can be used to differentiate DNA sequences of normal and mutated chromosomes.

5.3. Microarray Analysis

This technique also capitalizes on the **hybridization** between **complementary** nucleic acid sequences. Microarray analysis is useful for the **comparison** of **gene-expression** patterns between different cell types—for example, cells infected with a virus vs. uninfected cells, or cancerous cells vs. healthy cells (Fig. 54). Typically, **DNA** or **cDNA** from an experimental sample is deposited on **a glass slide** alongside known DNA sequences. DNA fragments (encompassing an organism's entire genomic library) or cDNA fragments (corresponding to an organism's full complement of expressed genes) can be individually spotted on a glass slide. Once deposited on the slide, genomic **DNA** or **mRNA** can be isolated from the two samples for comparison. If **mRNA** is isolated, it is **reverse-transcribed** to cDNA using reverse transcriptase. Then the two samples of genomic DNA or cDNA are **labeled** with different fluorescent dyes (typically **red** and **green**). The labeled genomic DNA samples are then combined in equal amounts, added to the **microarray chip**, and allowed to hybridize to complementary spots on the microarray.

Hybridization of sample genomic DNA molecules can be monitored by measuring the **intensity** of **fluorescence** at particular spots on the microarray. **Differences** in the amount of hybridization between the samples can be **readily observed**. If only one sample's nucleic acids hybridize to a particular spot on the microarray, then that spot will appear either green or red. However, if both samples' nucleic acids hybridize, then the spot will appear yellow due to the combination of the red and green dyes (Fig. 54).

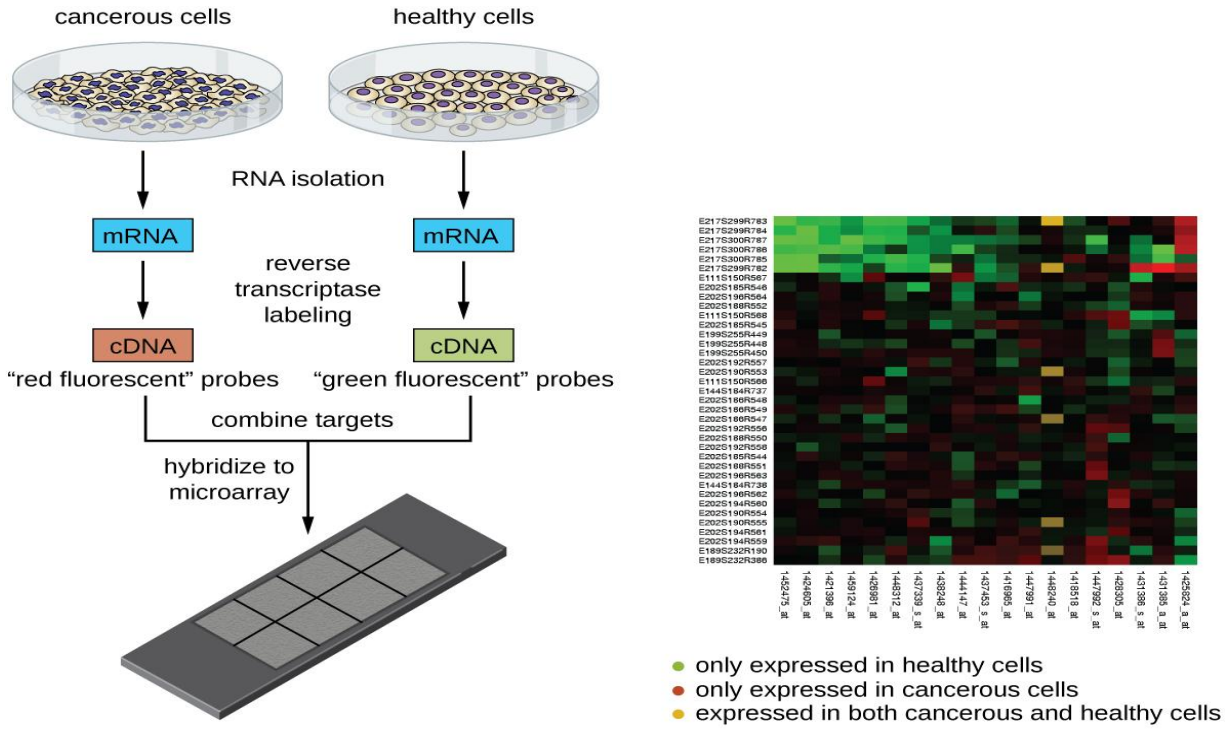


Fig. 54. Illustration of the main steps of microarray analysis.

5.4. Southern Blots

It is the method of analysis of the DNA imagined by Southern in 1975 to visualize the genes or any sequence of a genomic DNA, by a hybridization with a probe, marked and specific, with fragments of restriction of DNA, previously separated by electrophoresis, denatured and transferred to a membrane. One or more restriction enzymes may be used to achieve this cleavage.

The restriction fragments in the gel are denatured with an alkaline solution and then transferred to a nitrocellulose sheet or nylon membrane. This transfer is done by blotting (Fig. 55).

▪ Experimental protocol

1. The DNA to be analyzed is subject to the action of restriction enzymes.

2. Sample collection for agarose gel electrophoresis (90 mV/1h).

3. DNA staining by ethidium bromide (BET): a reddish color is revealed under UV (254 nm).

4. Bicatenial DNA is alkaline (e.g. NaOH) treated to separate the two strands.

5. A nitrocellulose or nylon membrane is applied to the agarose gel and then absorbent paper is placed on the agarose gel. This allows the single strands to pass from the gel to the membrane by simple capillarity: it is blotting or blotting.

6. The membrane, which has become an electrophoretic migration copy of agarose gel, is subjected either to UV to bind single strand DNA to the membrane (covalent bonds) in the case of nylon membrane, or to heat in the case of nitrocellulose membrane.

7. The membrane is placed in a bath of a hybridization solution that contains radioactive single-stranded DNA probes. These will hybridize (if compatible) specifically with the monostrand fragments of DNA contained in the nitrocellulose membrane. After hybridization, the excess probe is removed from the membrane by different washes.

8. A photographic film is placed on this membrane. The radiations emitted by the marked probes allow the printing on the photographic film: it is the autoradiography. Thanks to this film we will be able to visualize the location of the hybridization in the form of genetic fingerprints.

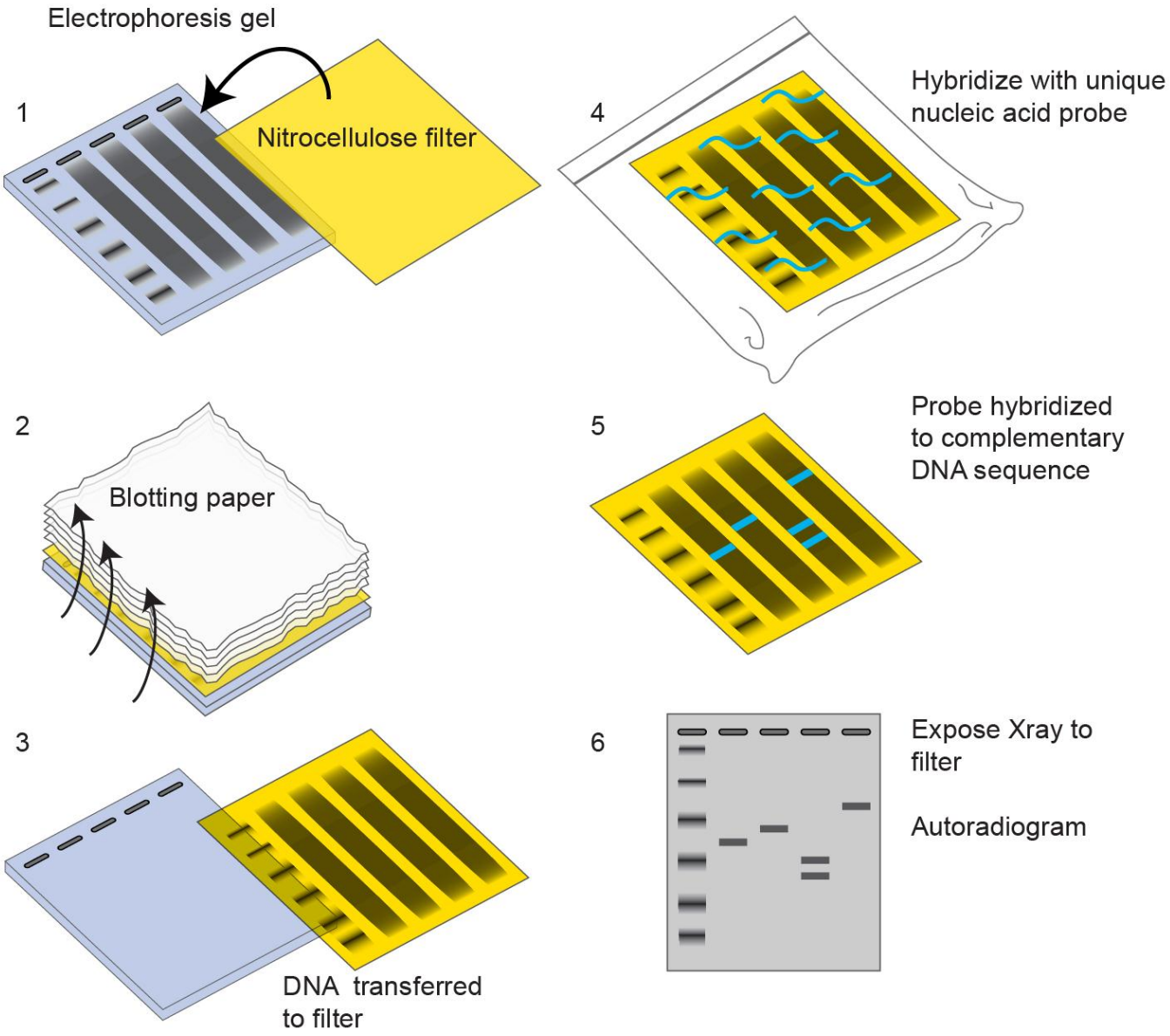


Fig. 55. Southern blot technique.

5.5. Northern blot

Its principle is identical to the Southern blot but in this case we work with RNA, with small fragments and there is no digestion. This technique is applied for the study of gene expression and also RNA size (Fig. 56).

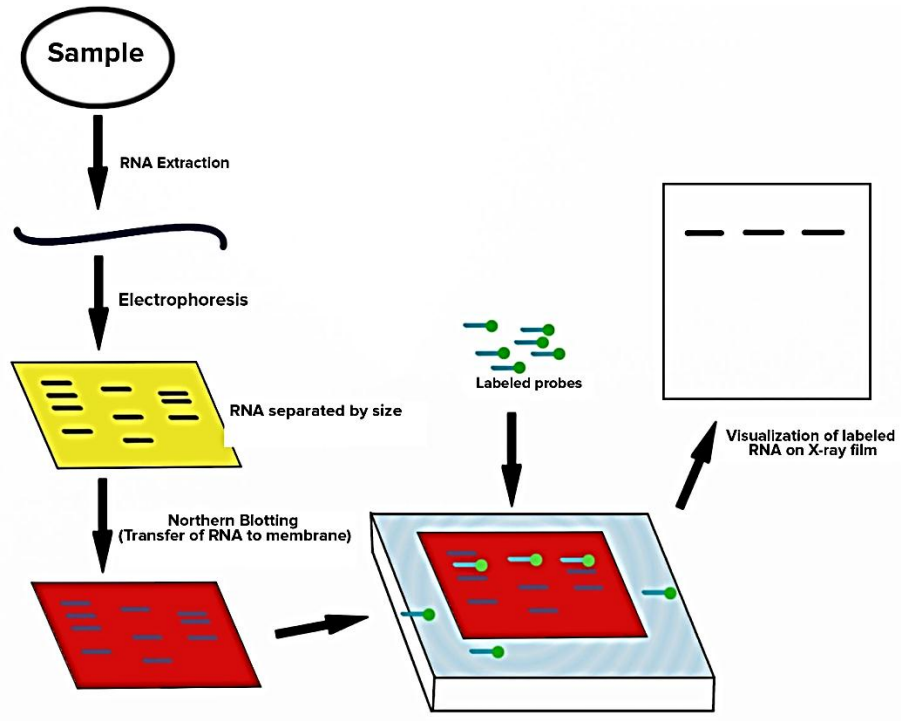


Fig. 56. Northern blotting technique.

5.6. Polymerase Chain Reaction (PCR)

Most methods of DNA analysis, such as restriction enzyme digestion and agarose gel electrophoresis, or DNA sequencing **require large amounts of a specific DNA fragment**. In the past, large amounts of DNA were produced by growing the host cells of a genomic library. However, libraries take time and effort to prepare and DNA samples of interest often come in minute quantities.

The polymerase chain reaction (PCR) permits **rapid amplification** in the number of copies of specific DNA sequences for further analysis (Fig. 57). One of the most **powerful techniques** in molecular biology, PCR was developed in 1983 by Kary Mullis while at Cetus Corporation. PCR has specific applications in **research**, **forensic**, and **clinical** laboratories, including:

- determining the sequence of nucleotides in a specific region of DNA
- amplifying a target region of DNA for cloning into a plasmid vector
- identifying the source of a DNA sample left at a crime scene
- analyzing samples to determine paternity
- comparing samples of ancient DNA with modern organisms
- determining the presence of difficult to culture, or unculturable, microorganisms in humans or environmental samples

PCR is an ***in vitro*** laboratory technique that takes advantage of the **natural process** of DNA replication. The heat-stable DNA polymerase enzymes used in PCR are derived from hyperthermophilic prokaryotes. ***Taq DNA polymerase***, commonly used in PCR, is derived from the *Thermus aquaticus* bacterium isolated from a hot spring in Yellowstone National Park. DNA replication requires the use of primers for the initiation of replication to have free 3'-hydroxyl groups available for the addition of nucleotides by DNA polymerase. However, while primers composed of RNA are normally used in cells, DNA primers are used for PCR. DNA primers are preferable due to their stability, and DNA primers with known sequences targeting a specific DNA region can be chemically synthesized commercially. These DNA primers are functionally similar to the DNA probes used for the various hybridization techniques described earlier, binding to specific targets due to complementarity between the target DNA sequence and the primer (Fig. 57).

The PCR reaction consists of three steps that constitute a cycle in which the amount of DNA to be amplified is doubled. These cycles are renewed between 20 and 50 times depending on the amount of target DNA and the desired goal.

• **denaturing** the DNA to be amplified at 94 °C,

• **hybridization** (annealing) with primer at (56 – 64 °C)

• **extension** of primer at 70–72 °C by *Taq polymerase*.

Amplification is performed by repeating cycles which ensures exponential duplication of each strand (Fig. 55).

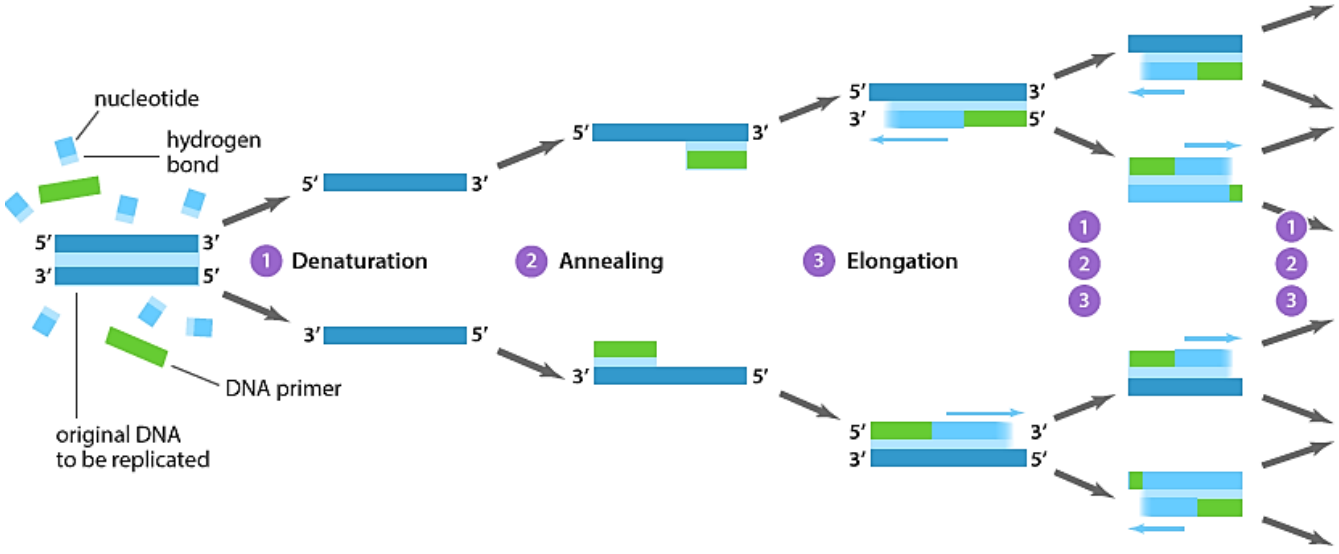


Fig. 57. The polymerase chain reaction (PCR) is used to produce many copies of a specific sequence of DNA.

▪ PCR Variations

Several later modifications to PCR further increase the utility of this technique.

a) Reverse transcriptase PCR (RT-PCR) is used for obtaining DNA copies of a specific mRNA molecule. RT-PCR begins with the use of the reverse transcriptase enzyme to convert **mRNA** molecules into **cDNA**. That cDNA is then used as a template for traditional PCR amplification. RT-PCR can detect whether a specific gene has been expressed in a sample (Fig. 58).

b) Real-time PCR, also known as **quantitative PCR (qPCR)**. Standard PCR and RT-PCR protocols are not quantitative because any one of the reagents may become limiting before all of the cycles within the protocol are complete and samples are only analyzed at the end. Because it is not possible to determine when in the PCR or RT-PCR protocol a given reagent has become limiting, it is not possible to know how many cycles were completed prior to this point, and thus it is not possible to determine how many original template molecules were present in the sample at the start of PCR. In qPCR, however, the use of **fluorescence** allows one to monitor the increase in a double-stranded template during a PCR reaction as it occurs. These kinetics data can then be used to **quantify** the amount of the original target sequence. The use of qPCR in recent years has further expanded the capabilities of PCR, allowing researchers to determine the number of DNA copies, and sometimes organisms, present in a sample (Fig. 58).

PS. *Both **RT-PCR** and **qPCR** techniques could be combined in one single protocol named **RT-qPCR**.*

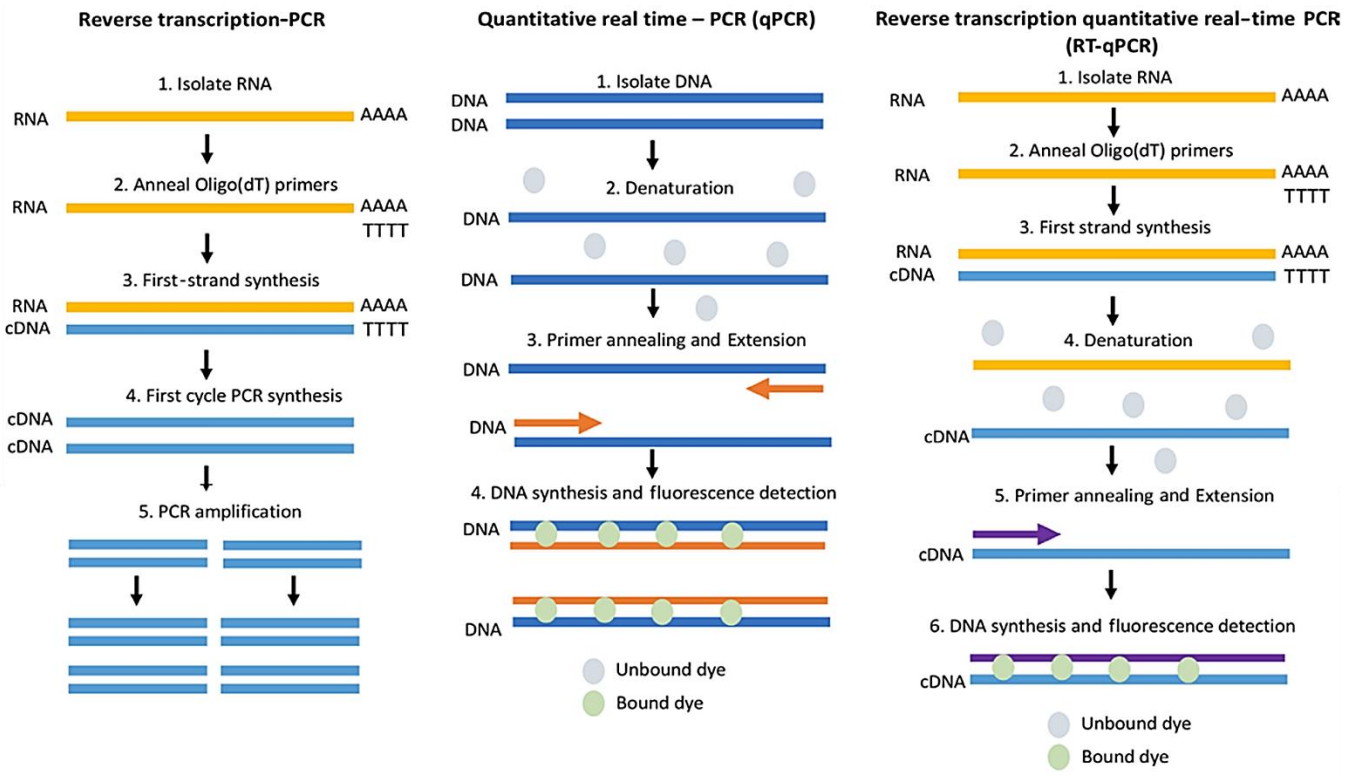


Fig. 58. Schematic comparing of RT-PCR, qPCR and RT-qPCR workflows.

CHAPTER 6

DNA SEQUENCING, GENOMIC DNA AND cDNA LIBRARY

6.1. DNA sequencing

To understand the complexities of gene **structure**, its **expression**, its **regulation**, protein **interactions**, and molecular mechanisms of genetic **diseases**—the **detailed** and **exact sequences** of the bases in DNA is very essential. During the late 1970s two different sequencing techniques were developed. They are:

1. The chemical cleavage method developed by Allan Maxam and Walter Gilbert (*the Maxam-Gilbert Method*). This method was very important in the beginning but is no longer practiced today.

2. The enzyme-mediated chain termination method developed by Frederick Sanger along with Andrew Coulson is popularly known as the *Sanger sequencing method* or the enzymatic method of DNA sequencing. Both methods involve the **labeling** of the **terminal** nucleotide followed by **separation** and **detection** of the generated oligonucleotides.

Principles of DNA sequencing

By definition, the determination of a DNA sequence requires that the bases are identified in a sequential technique that enables the processive identification of each base in turn. There are three main requirements for this to be achieved:

- **DNA** fragments need to be **prepared** in a form suitable for sequencing.
- The **technique** used must achieve the aim of presenting **each base** in turn in a form suitable for identification.
- The **detection** method must permit **rapid** and **accurate** identification of the bases.

6.1.1. Maxam–Gilbert (chemical) sequencing

The double-stranded DNA to be sequenced is first **marked** with ^{32}P at the level of the **5' phosphate**. It must then be **digested** by a **restriction endonuclease** into two different size fragments, which are separated by **electrophoresis**. This results in a fragment with only one **5 end marked** (Fig. 59).

The principle of the method consists in **modifying a base** type specifically by a **chemical reagent** and **cleaving** the DNA chain at the level of this modified nucleotide. The modifying agents are used in such a **concentration** that we will have only **one cut** per molecule, and these at all possible positions (Fig. 59). The reagents used are:

- **DMS (dimethyl sulfate) - purine methylation then 0.1 M NaOH → cleaving after G then 0.1 M HCl → cleaving after A ->**
- **Hydrazine - cleaving after C and T → hydrazine + NaCl → cleaving after C.**

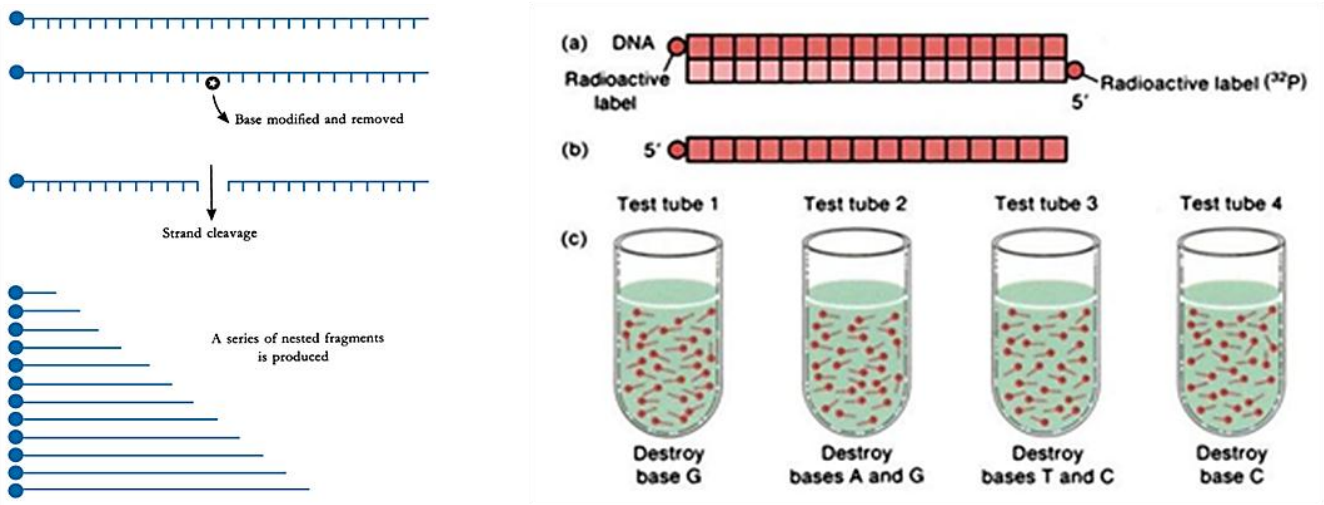


Fig. 59. DNA sequencing principale using the chemical (Maxam–Gilbert) method.

After electrophoresis on polyacrylamide gel of the **4 reaction** products and autoradiography, the relative positions of the different bases are deduced from the comparison of migration distances of the fragments marked in 5'. The following figure shows an example of a fragment sequencing using the Maxam–Gilbert method (Fig. 60).

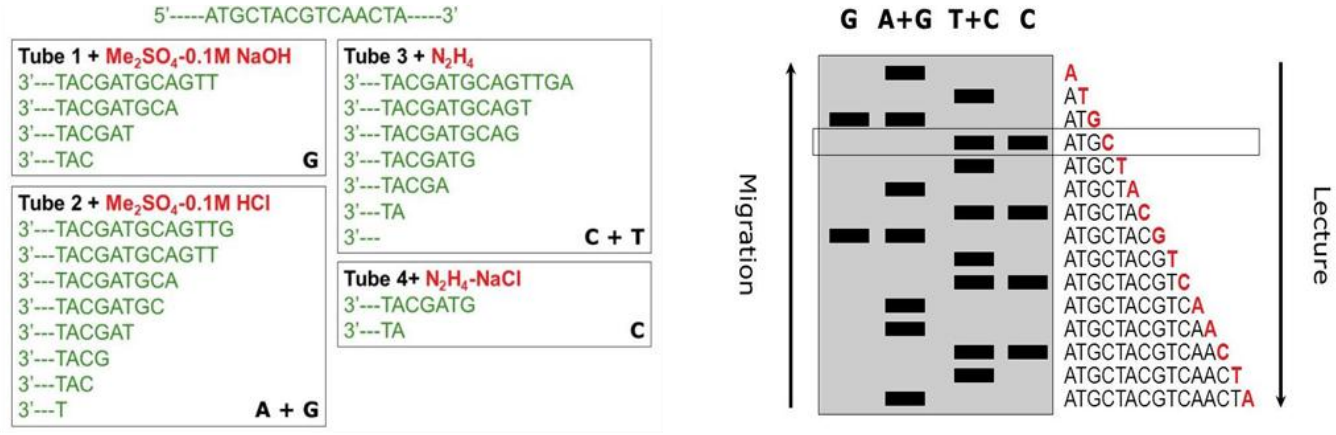


Fig. 60. Example of a fragment sequencing using the Maxam–Gilbert method.

6.1.2. Sanger–Coulson (dideoxy or enzymatic) sequencing

This is the most **widely** used method. This enzymatic sequencing method developed by Sanger by incorporation didoxyribonucleotides chain terminators has been universally adopted. This method takes advantage of the absence of a 3' hydroxyl of a **ddXTP** (Fig. 61) that does not allow the formation of a phosphodiester bond. The consequence is a **stop** of elongation when a didoxyribonucleotide is incorporated into a strand of synthetic DNA. This phenomenon is the basis of Sanger's method.

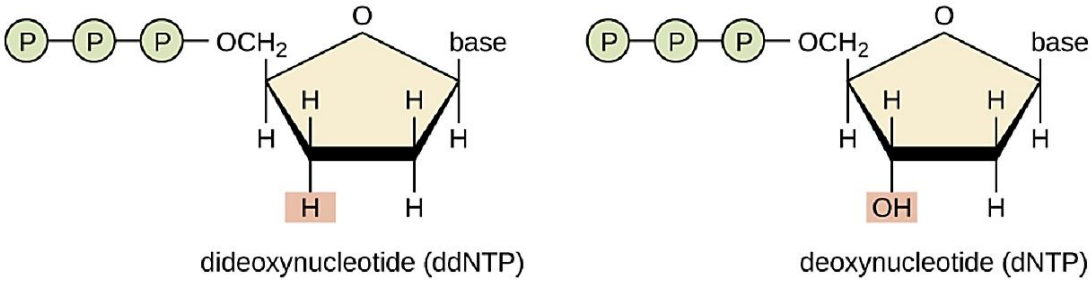


Fig. 61. Dideoxynucleotide structure comparing to a deoxynucleotide (3' hydroxyl group missing)

Principle: controlled **interruption** of enzymatic replication called didoxynucleotide method.

- A **single** strand DNA and small complementary primer are used.
- **DNA polymerase** synthesizes the complementary strand in the presence of ^{32}P-labelled dNTP
- The chain elongation is stopped by an analogue of each dNTP (4 different experiments) a **didesoxy** that stops the chain elongation.

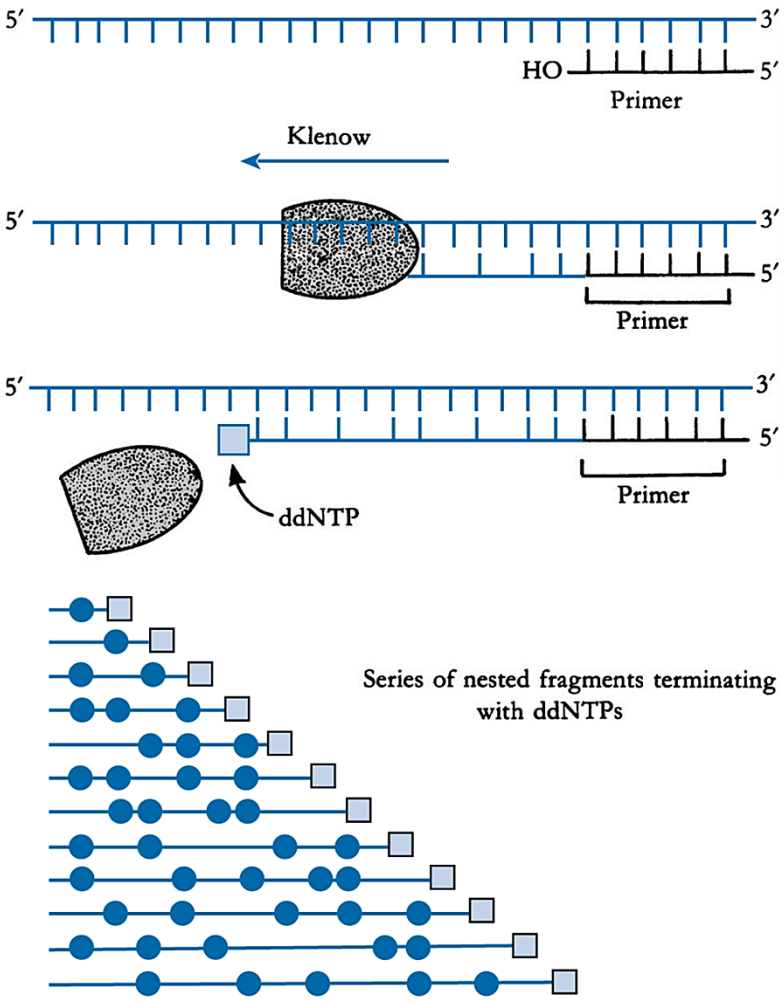


Fig. 62. DNA sequencing using the dideoxy chain termination (Sanger–Coulson).

DNA polymerase adds nucleotides until the addition of a didoxyribonucleotide which stops the synthesis, the enzyme can not add nucleotide to a didoxyribonucleotide (there is no more hydroxyl function). This results in a series of more or less long DNA molecules in the tube.

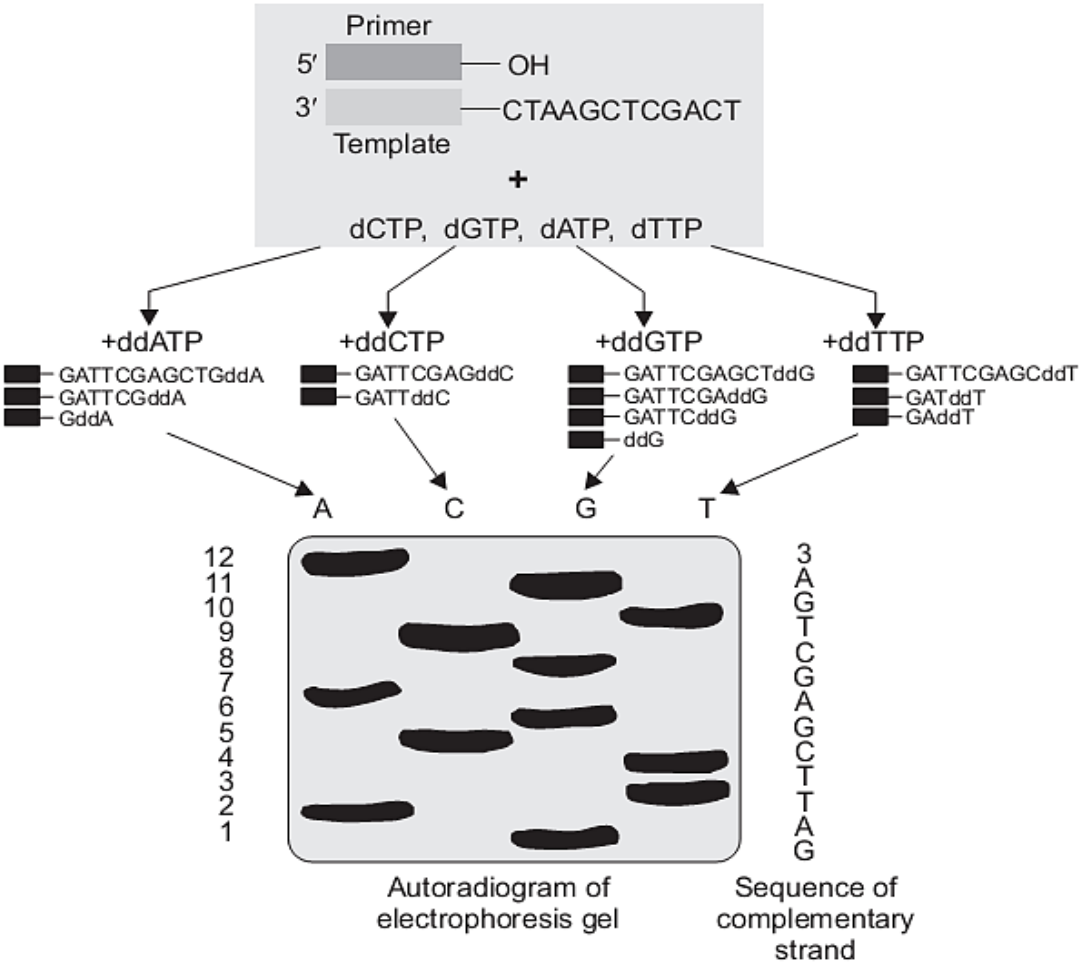


Fig. 63. Example of sequencing process using the Sanger's method.

DNA Sequencing reading

In Sanger's method, four reactions were set up for each DNA molecule being sequenced, each reaction containing only one of the four possible ddNTPs. Each ddNTP was **labeled** with a **radioactive phosphorus molecule**. The products of the four reactions were then run in separate lanes side by side on long, narrow PAGE gels, and the bands of varying lengths were detected by **autoradiography**. Today, this process has been simplified with the use of ddNTPs, each labeled with a **different colored fluorescent dye** or **fluorochrome** (Fig. 64), in one sequencing reaction containing all four possible ddNTPs for each DNA molecule being sequenced). These fluorochromes are detected by **fluorescence spectroscopy**. Determining the fluorescence color of each band as it passes by the **detector** produces the nucleotide sequence of the template strand (Fig. 64).

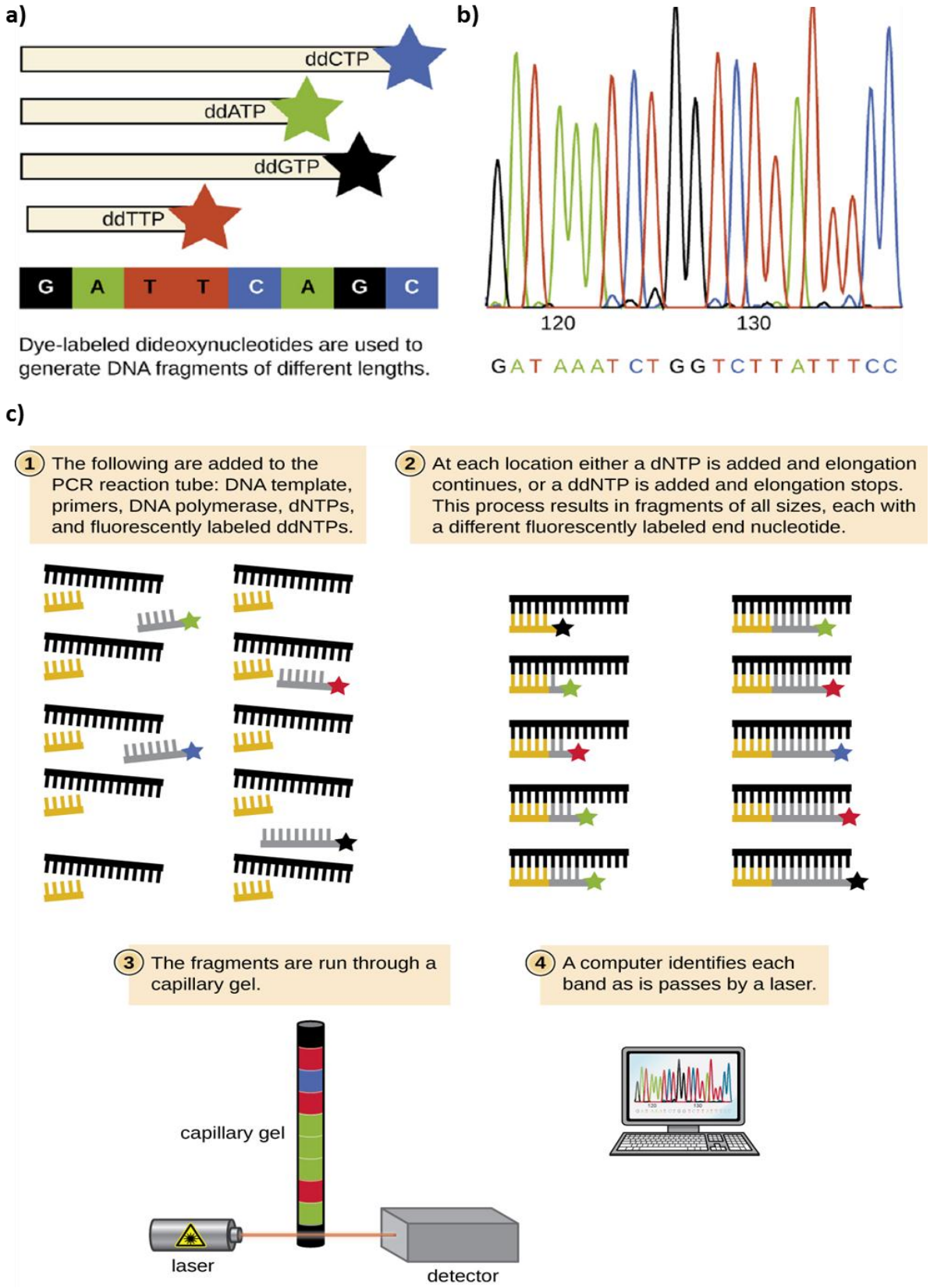


Fig. 64. Illustration of (a) Sanger's dideoxy chain termination method using ddNTPs tagged with (b) fluorochromes and (c) fluorescence detection.

6.1.3. Pyrosequencing

Pyrosequencing is a technique of sequencing by direct synthesis of oligonucleotides. Its realization requires the use of a cascade of 4 enzymes:

- **Taq polymerase** for incorporation of oligonucleotides (**dNTP**);
- **ATP sulfurylase** to catalyze inorganic pyrophosphate (**ppi**) into energy, source of **ATP**.
- **Luciferase** to generate **light** or **fluorescent signal** from ATP.
- **Pyrase** to **degrade ATP** and excess unincorporated nucleotides.

Unlike automatic sequencing by the Sanger technique, Pyrosequencing uses 2'-deoxyadenosine-5'-O-(1-thiotriphosphate) (**dATPαS**) instead of **dATP**. This nucleotide is recognized by both **polymerase** and **luciferase**, unlike ATP nucleotide which is recognized only by polymerase. The four oligonucleotides are introduced one by one.

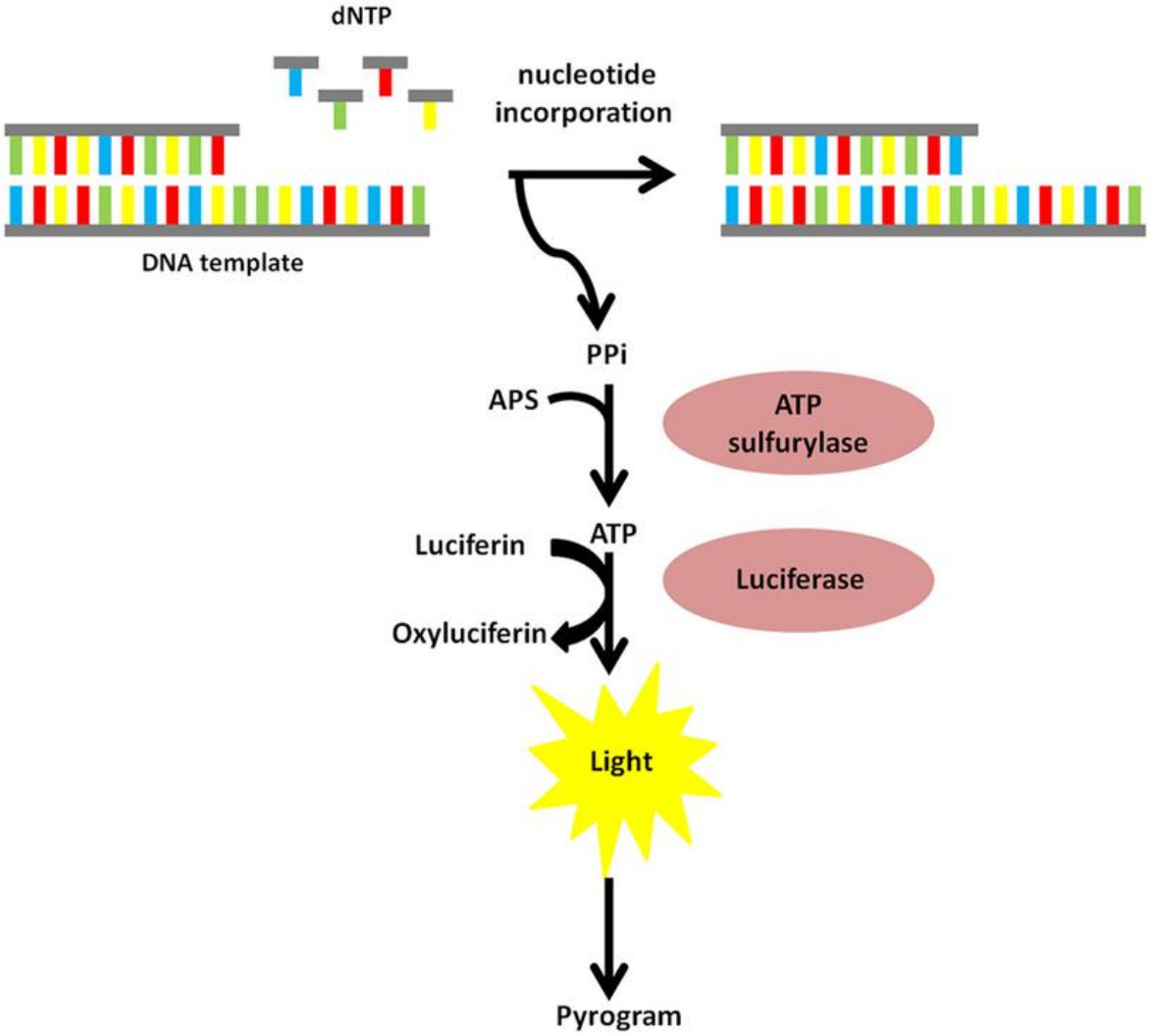
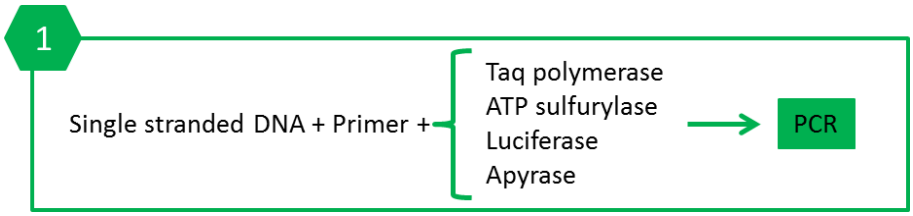


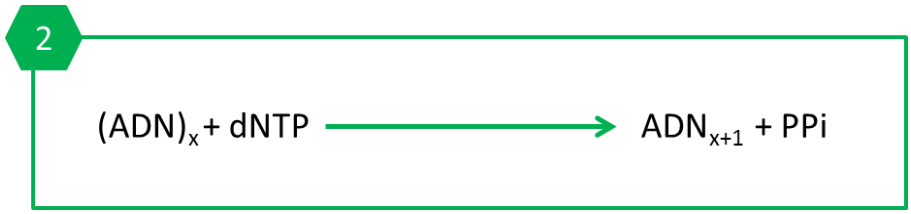
Fig. 65. The principle of pyrosequencing.

Pyrosequencing is performed in 5 main steps:

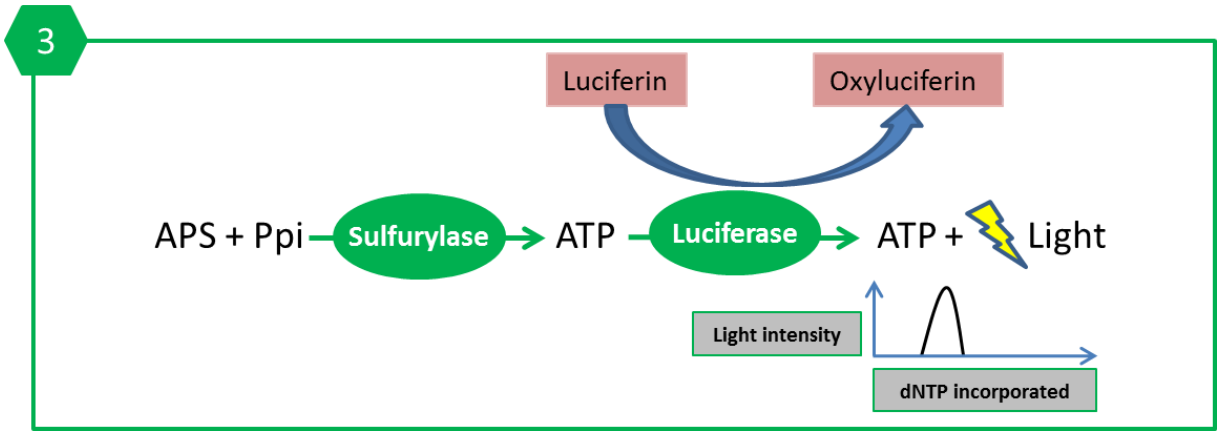
Step 1: the DNA analyzed is amplified in the presence of a sequencing primer. The whole is then incubated in the presence of the 4 enzymes: Taq polymerase, ATP sulfurylase, luciferase and apyrase, the substrate adenosine 5' phosphosulfate APS is finally luciferin.



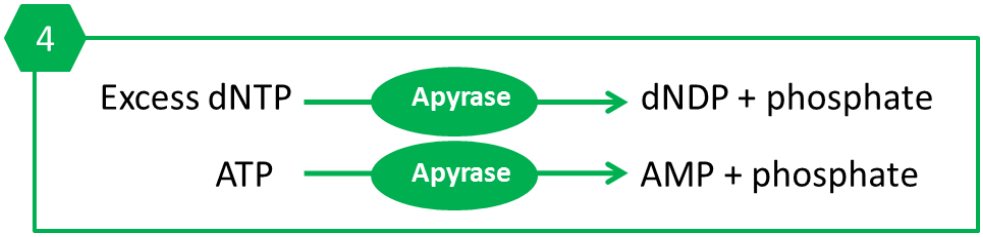
Step 2: the first of the 04 dNTPs is added in the reaction. DNA polymerase allows the incorporation of this dNTP in the strand being synthesized, if the latter is complementary to the base located in the sequence to be analyzed. Each incorporation of dNTP is followed by a release of a pyrophosphate molecule (ppi). The amount of (ppi) released is proportional to the amount of dNTP incorporated.



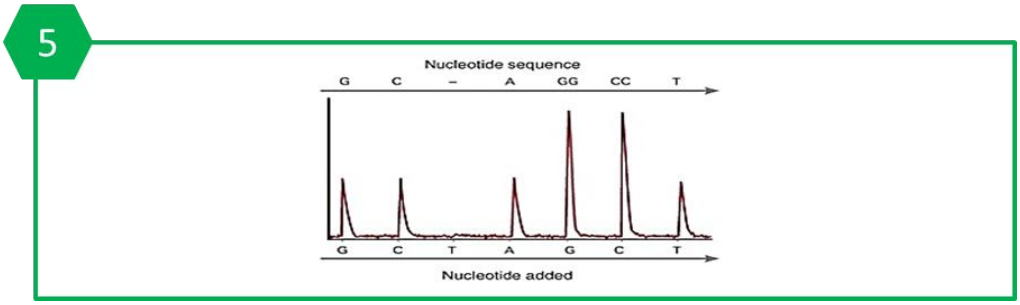
Step 3: in the presence of adenosine 5' phosphosulfate (APS) ATP sulfurylase transforms (ppi) into ATP equimolar. This ATP will be used as a substrate by luciferase during the transformation of luciferin into oxyluciferin and allows the emission of a light signal whose intensity is proportional to the amount of ATP. This light signal will be detected by a CCD camera connected to the sequencer and visualized as a peak or pyrogram. The height of this peak will be proportional to the numbers of (ppi) released which is itself proportional to the numbers of embedded nucleotides.



Step 4: The excess of unincorporated dNTP and ATP are continuously degraded by the last enzyme in the cascade: pyrase.



Step 5: The dNTPs are introduced one after the other, according to the established order. The dNTP thus incorporated form a sequence that will be determined from the specific light signal of each nucleotide and then captured and recorded on the pyrogram.



6.1.4. Next Generation Sequencing

Next generation sequencing, or alternatively **high- throughput sequencing**, actually dates back to the 1990s—to the massively parallel signature sequencing, developed by Lynx Therapeutics (founded by Sydney Brenner and Sam Eletr). The idea behind high-throughput technologies in general is to **screen** a **large amount of data** for a **minimum price** and in as **short a time** as possible. As such, it is not surprising that next- generation sequencers generate data that is equivalent to **several hundred Sanger**-type sequencers, operated by one person in one day.

Recently, **Roche** Diagnostics released 454 **pyrosequencing** (also known as sequencing by synthesis) technologies, where DNA is captured by immobilization on beads and amplified inside water-oil emulsion. The machine uses many wells with **picoliter** volume to amplify from this template DNA, and uses **luciferase** for monitoring (the addition of new nucleotides results in a light signal); however, the cost per base that is sequenced is too high.

Illumina, on the other hand, has developed another high-throughput method that uses reversible dye terminators. As these nucleotides are added to the DNA one at a time, the fluorescence of the nucleotides is recorded. Last, ABI's SOLiD (Sequencing by Oligonucleotide Ligation and Detection) sequencing relies on ligation, where oligonucleotides of a fixed length are captured on magnetic beads (a library of one fragment per bead), with an **adapter sequence**.

When a universal primer anneals, a library of probes is added to the reaction, allowing the ligation of the probe to the primer, followed by ligation at each cycle.

a. Workflow of Next-Generation Sequencing (NGS)

Next-Generation Sequencing (NGS) involves several key steps. First, **DNA** or **RNA** is **extracted** from the sample of interest, such as a patient's blood or a piece of environmental DNA, containing vital nucleotide information. Next, the extracted genetic material is **fragmented** into **smaller** pieces. These fragments are then **prepared** into a **sequencing library** by attaching unique molecular identifiers (barcodes) and adapters. The prepared library is loaded onto a **high-throughput sequencer**, where each nucleotide fragment is **individually sequenced**. Finally, the raw sequencing data is processed and analyzed to reconstruct the original genetic sequence (Fig. 66).

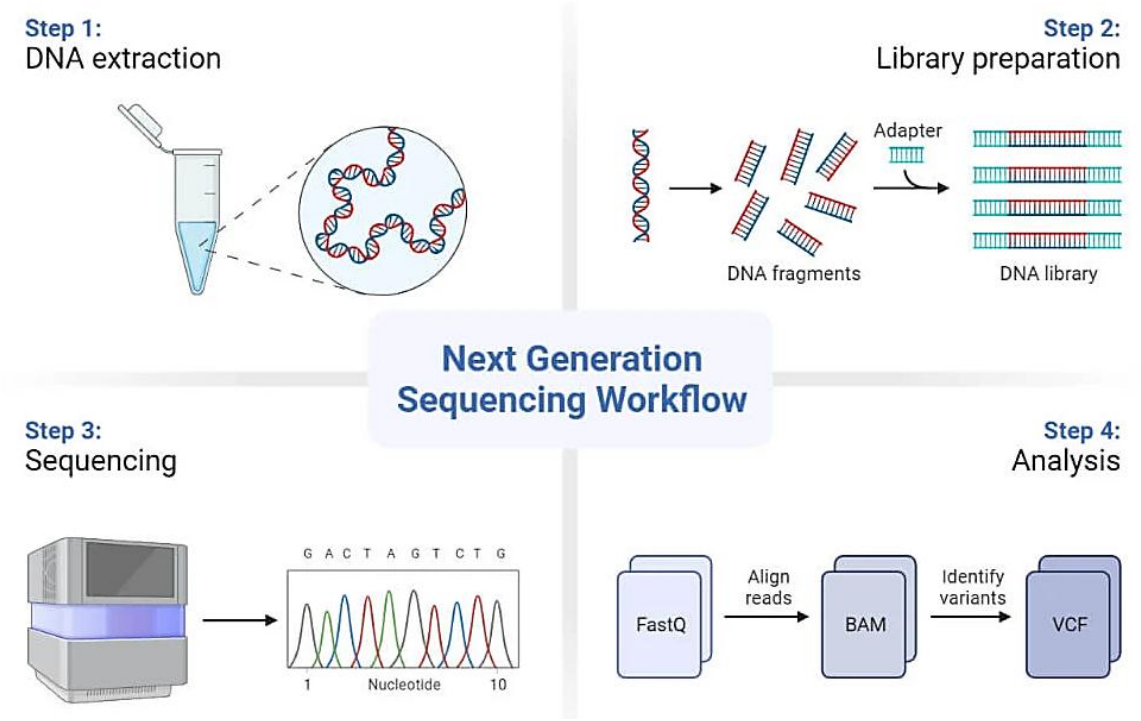


Fig. 66. Basic principle and workflow of Next-Generation Sequencing.

The key components of NGS include:

1) DNA extraction: This initial step involves **isolating DNA** from the biological sample. Depending on the sample type, various methods can be used, such as chemical extraction or mechanical disruption.

2) Library preparation: The genetic material is **fragmented** into **smaller pieces** after DNA extraction. These fragments are then **tagged** with **adapters** and **barcodes**, allowing the sequencer

to distinguish individual DNA molecules. This step is crucial for high-throughput sequencing, enabling parallel processing of numerous DNA fragments.

3) Sequencing: The prepared library is loaded onto a sequencer, which **reads** the sequence of **each fragment**. Modern NGS platforms can process **millions** to **billions** of fragments simultaneously, resulting in an enormous amount of sequencing data.

Illumina Platforms

While all NGS platforms perform sequencing of millions of small fragments of DNA or cDNA, there are several different sequencing technologies. Some platforms can produce more reads or different read lengths than others. The most prevalent and successful sequencing technology was pioneered by Illumina. Illumina's instrument portfolio includes benchtop (iSeq, MiniSeq, MiSeq, NextSeq) and production-scale (NextSeq, HiSeq, and NovaSeq) sequencing platforms.

Illumina sequencers use a glass flowcell coated with millions of oligonucleotides that are complimentary to the sequencing adaptors.

Each library fragment hybridizes with the primers and is further amplified to generate millions to billions of clonal clusters. Then, fluorescently labeled nucleotides are used to synthesize a complementary strand for each fragment. After the addition of each tagged nucleotide, the flow cell is imaged and the emission from each cluster is recorded. The fluorescent emission wavelength and intensity are used to identify the sequence of the templates.

4) Data analysis: The raw sequencing data is processed through **bioinformatics** pipelines. This involves aligning the short sequencing reads to a reference genome or assembling them de novo to reconstruct the original genetic sequence. Various bioinformatics tools are used to identify genetic variations, gene expression levels, or other relevant information.

b. Mains applications of NGS

Next-generation sequencing technology has fundamentally changed the kinds of questions scientists can ask and answer. Innovative sample preparation and data analysis options enable a broad range of applications. For example, NGS allows labs to:

- **Rapidly** sequence **whole genomes**
- **Deeply** sequence **target** regions
- Utilize RNA sequencing (RNA-Seq) to discover **novel RNA** variants and splice sites, or **quantify** mRNAs for gene expression analysis
- Analyze **epigenetic** factors such as genome-wide DNA methylation and DNA-protein interactions
- Sequence **cancer** samples to study rare somatic variants, tumor subclones, and more
- Study the human **microbiome**
- Identify **novel** pathogens

6.2. Genomic DNA and cDNA library

6.2.1. Genomic DNA

Genomic-DNA sequences are sequences of the DNA **directly isolated** from the chromosomes or the total genome. They include both **coding** parts and **non-coding** parts. Coding parts of the genome are the genes of proteins and other types of RNAs.

6.2.2. Creating a Genomic Library

Molecular cloning may also be used to generate a **genomic library**. The library is a **complete** (or nearly complete) **copy** of an organism's genome contained as recombinant DNA plasmids engineered into unique clones of bacteria. Having such a library allows a researcher to create large **quantities** of each fragment by growing the bacterial host for that fragment. These fragments can be used to **determine** the sequence of the DNA and the **function** of any genes present.

One method for generating a genomic library is to **ligate** individual restriction enzyme-digested **genomic fragments** into **plasmid** vectors cut with the same restriction enzyme (Fig. 65). After transformation into a bacterial host, **each transformed bacterial** cell takes up a single **recombinant plasmid** and grows into a colony of cells. All of the cells in this colony are identical clones and carry the same recombinant plasmid. The resulting library is a collection of colonies, each of which contains a fragment of the original organism's genome, that are each separate and distinct and can each be used for further study. This makes it possible for researchers to screen these different clones to discover the one containing a gene of interest from the original organism's genome (Fig. 65).

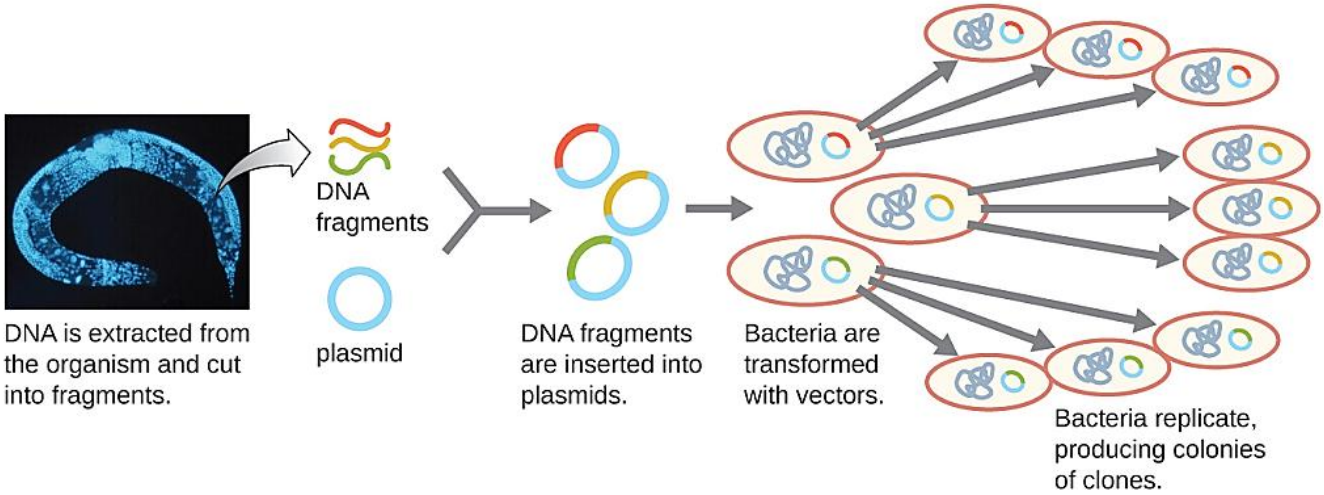


Fig. 67. Generation of a genomic library facilitates the discovery of the genomic DNA fragment that contains a gene of interest.

To construct a genomic library using **larger fragments** of genomic DNA, an *E. coli* **bacteriophage**, such as **lambda**, can be used as a host (Fig. 66). Genomic DNA can be sheared or enzymatically **digested** and **ligated** into a pre-digested bacteriophage lambda DNA vector. Then, these recombinant phage DNA molecules can be packaged into **phage particles** and used to infect *E. coli* host cells on a plate. During infection within each cell, each recombinant phage will make **many copies** of itself and lyse the *E. coli* lawn, forming a plaque. Thus, each plaque from a phage library represents a unique recombinant phage containing a distinct genomic DNA fragment. Plaques can then be screened further to look for genes of interest. One **advantage** to producing a library using phages instead of plasmids is that a phage particle holds a **much larger insert of foreign DNA** compared with a plasmid vector, thus requiring a much smaller number of cultures to fully represent the entire genome of the original organism (Fig. 66).

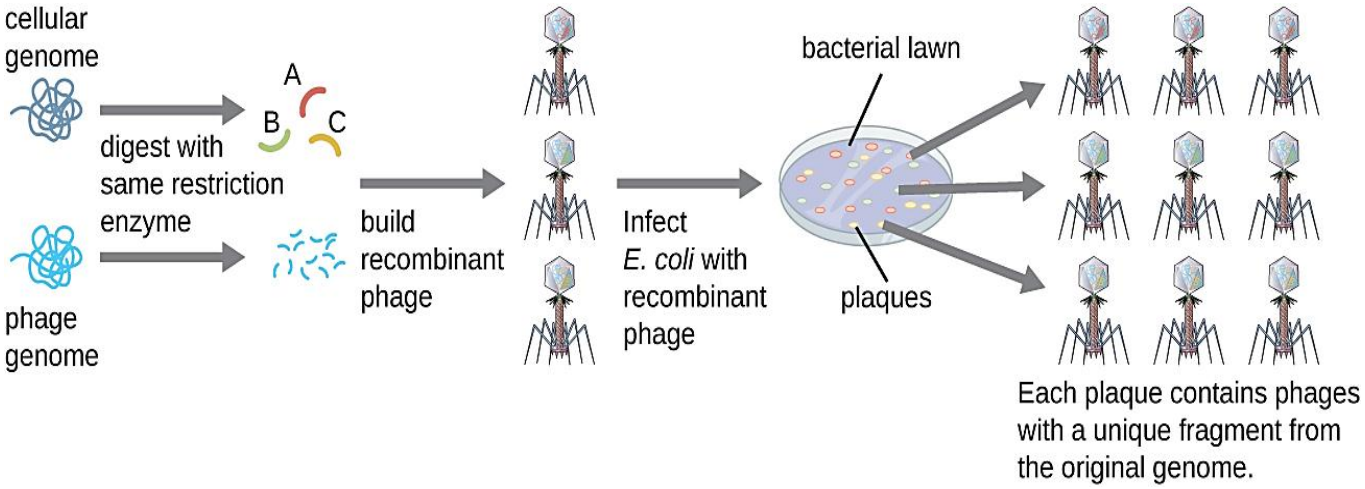


Fig. 68. Recombinant phage DNA molecules are made by ligating digested phage particles with fragmented genomic DNA molecules.

6.2.3. Creating cDNA library

To focus on the **expressed genes** in an **organism** or even a **tissue**, researchers construct libraries using the organism's messenger RNA (**mRNA**) rather than its genomic DNA. Whereas all cells in a single organism will have the same genomic DNA, **different tissues** express **different genes**, producing **different complements of mRNA** (For example, all human cells' genomic DNA contains the gene for insulin, but only cells in the pancreas express mRNA directing the production of insulin).

Because **mRNA** cannot be cloned directly, in the laboratory mRNA must be used as a template by the **retroviral enzyme reverse transcriptase** to make complementary DNA (**cDNA**). A cell's full complement of mRNA can be reverse-transcribed into cDNA molecules, which can be used as a template for DNA polymerase to make double-stranded DNA copies; these fragments can subsequently be ligated into either plasmid vectors or bacteriophage to produce a cDNA library.

The benefit of a cDNA library is that it contains DNA from **only the expressed genes** in the cell. This means that the introns, control sequences such as promoters, and DNA not destined to be translated into proteins are not represented in the library. The focus on translated sequences means that the library cannot be used to study the sequence and structure of the genome in its entirety. The construction of a cDNA genomic library is shown in Figure 67.

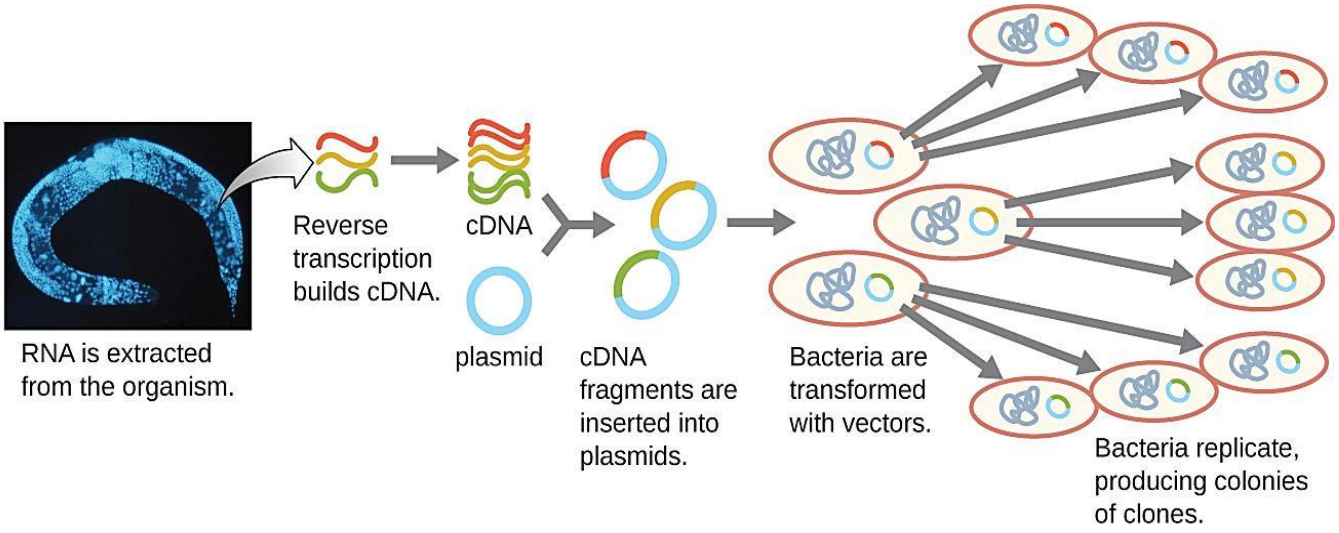


Fig. 69. Complementary DNA (cDNA) is made from mRNA by the retroviral enzyme reverse transcriptase, converted into double-stranded copies, and inserted into either plasmid vectors or bacteriophage, producing a cDNA library.

▪ cDNA synthesis

The cDNA synthesis is catalyzed by **reverse transcriptases (RT)**. These enzymes are DNA-dependent RNA polymerases able of using an **RNA** strand as a matrix to catalyse the synthesis of a **complementary DNA strand**. Like all DNA polymerases, reverse transcriptases cannot initiate the synthesis of a DNA strand alone. They need a **primer** with a **free 3'-OH end**. When the RNA to be amplified are polyadenylated into 3' (eukaryotic mRNA for example), the chosen primer can simply be a **polyT** sequence (12 to 18) consisting of a succession of deoxythymidines (as in the diagram below in red). In this case, all mRNAs are a priori copied to cDNA. **Ribonuclease H** or **RNAse H** is an enzyme that hydrolyzes the RNA strand in a hybrid DNA: RNA double strand. It releases 3'-OH and 5'-phosphate ends. RNAse H does not hydrolyze single- or double-stranded RNA (Fig. 68.a).

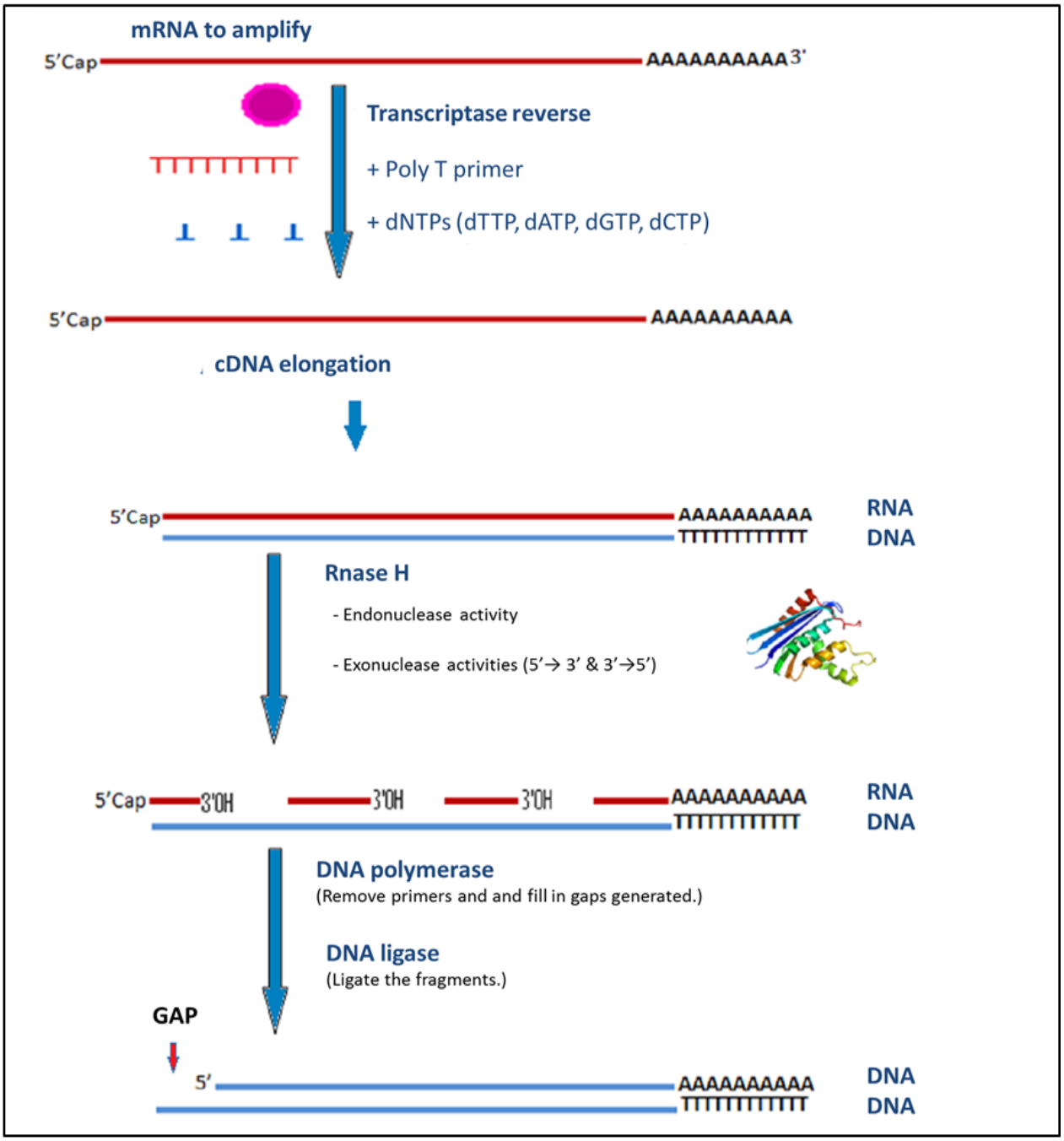
▪ Inserting cDNA in an appropriate vector

There are two methods of inserting cDNA into vectors:

1. Addition of a homopolymeric tail.
2. Addition of a joint (linker).

The addition of a homopolymeric tail is done by an unusual DNA polymerase found in eukaryotic cell types called «pre-lymphocytes». This enzyme is called terminal transferase, it is able to add deoxynucleotide in part 3'-OH of a DNA strand (does not require a primer and matrix). In the presence of divalent cations (Mg++, Co++), the enzyme catalyses the addition of dNTPs (Purines: Mg++; Pyrimidines: Co+) at the 3' end of the DNA. Depending on the reaction conditions three to a few thousand nucleotides are added (Fig. 68.b) .

a)



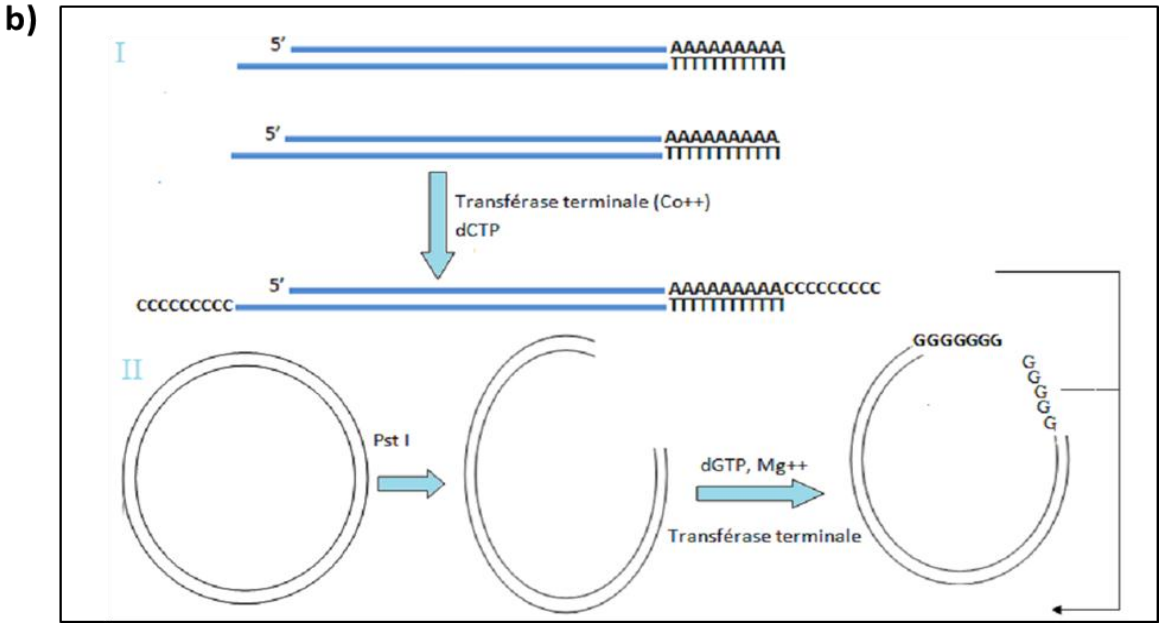


Fig. 70. (a) cDNA synthesis and (b) insertion into the vector.

CHAPTER 7

BIOTECHNOLOGICAL APPLICATIONS OF RECOMBINANT DNA

In recent decades, recombinant DNA technology has been used for a wide range of applications in different sectors (medical, pharmaceutical, food, etc.). In this section, we present some well-known applications of genetic engineering to biotechnology.

8.1. The enzyme luciferase

The enzyme luciferase catalyzes the reaction of **luciferin** with ATP. The reaction is accompanied by an emission of **light** and this is what explains the luminescence of the **firefly** in flight (the firefly produces this enzyme hence its name). This enzyme is currently used to monitor the regulation and genotype of tobacco.

A transgenic tobacco plant that expresses the luciferase gene lights up in the dark if it is brushed with a luciferin solution.

The luciferase gene can therefore be used as an **indicator** to study, during development, the regulation of the expression of one or the other gene of the plant or integrated DNA.

8.2. Transgenic maize

Every year farmers around the world lose 1/10 of their maize crops to insect pests. Transgenic maize contains a gene in its genome that codes for a **protein** that insects **hate** and that in principle should not cause harm to humans. This variety increases the yield by eliminating the insect factor.

8.3. Flavrsavr tomatoes

These transgenic tomatoes contain a gene that **inhibits** the **ripening** of the fruit, which gives them a better resistance to transport and a long shelf life.

Now with the advent of genetic engineering, not only can the quality of animal and plant species be improved, but these transgenic organisms are used as factories to produce «useful» proteins encoded by genes of foreign organisms.

8.4. Plant protection from viral infection

Plant **viruses** are a serious problem for agriculture because their infections cause a decrease in plant growth, crop yield and quality. Tobacco mosaic virus (TMV) infects tobacco plants. A transgenic plant that expresses the coat protein (CP) of TMV is **resistant** to infection by this virus. The transgenic plant is already produced in the United States.

8.5. Insect Pest Control

Every year billions of dollars in crops are lost to insects. The weapons most used to control insects are «insecticides and pesticides» which currently pose serious pollution problems to the environment.

Bacillus thuringensis (*Bt*), when sporulated, produces a crystallized protein that is toxic to the larvae of a number of crop insect pests and does not appear to be harmful to vertebrates. Transgenic tobacco, tomato and cotton plants that express the toxin gene are resistant to insect larvae.

The use of transgenic expression of serine protease inhibitors of insect digestive systems in tomatoes, potatoes, peas and other cereals also protects plants from insect pests.

8.6. Plant production of monoclonal antibodies

Plants can also be used as **bioreactors** for **antibody** production. This aspect represents a commercial interest for the production of proteins including monoclonal antibodies. The successful experiment used a heavy chain (H) and a light chain (L) in two separate experiments to transfer the antibody genes via T-DNA into two **tobacco** plants. Transgenic plants containing the heavy and light chain genes were crossed to produce progeny containing both genes. The leaves of these plants produce the inserted monoclonal antibody (about 1.5% of the translated proteins).

8.7. Bio-insecticide and malaria control

In 1976, an entomologist noticed a large amount of larvae of dead insects in the Negev desert, still very infested by mosquitoes and flies responsible for two large plagues. After a thorough study in the laboratory, he found that bacteria of the genus *Bacillus thuringiensis* present in the lakes of this desert systematically killed the larvae of mosquitoes and flies. From this discovery, a new strategy to control the plasmodium and trypanosome anopheles and glossin vectors was initiated using the following strategy:

- Production of bio-insecticides by *Bacillus thuringiensis* capable of killing fly and mosquito larvae
- Use of air transport means (helicopters) for the spraying of these bio-insecticides at rivers and water bodies infested by these insect vectors.

8.8. Transgenic mice

Any gene can currently be inserted into the mouse to obtain gene expression for this transgene. In this transgenesis, the mouse that integrates for example the gene of the human growth hormone (GH) acquires a huge double or triple mass compared to its twin sisters.

This type of approach can be applied to all mammals to increase their yield in weight and therefore in economic interest.

Transgenic organisms can be produced by injecting vectors into fertilized eggs and reimplanted into females. During embryogenesis, exogenous DNA makes its way to the germ cells and now behaves like an endogenous gene. It therefore passes to progeny as a normal nuclear gene.

The embryo with the transgene can also be transferred to the uterus of a healthy pseudo-pregnant mouse. The mouse that will be put down will carry the characters of the transgene but not the other sisters.

8.9. Cloning whole organisms

A mammal clone is a genetic copy of an entire organism. *Dolly* (first sheep cloned in England), is a clone of his mother, or his genes. Like sheep, will men also switch to the "*photocopier*".

The cloning technique is relatively simple in principle. In the case of the Dolly sheep, a cell was taken from the udder of a Finn Set white-faced sheep. An egg was also taken from another sheep and the nucleus containing the genetic baggage was removed. Why an egg? So that it could become an embryo. With the help of an electric shock, the udder cell, which contains all its genes, was fused in vitro and the egg emptied of any hereditary material. This is the reason why Dolly will have for any genetic material only that which contained the cell of the udder. The egg thus "*electrified*" divides and the life process begins. After dividing a sufficient number of times, the embryo (blastula stage) is placed in the uterus of a carrier sheep. Dolly was born of this technique, identical in all respects to the sheep that provided the cell of the udder.

The process of cloning an entire organism is simple:

a. An egg cell is taken and emptied of its nucleus.

b. The nucleus of a cell that is to be cloned is removed and put into the nucleus-private egg

c. Implantation in the uterus of a carrier

d. The organism born bears the characteristics of the nucleus transferred from the cloned organism. However, certain egg characters remain present because of the genetic material contained in the mitochondria.

8.10. Genetic engineering and the fight against HIV.

Today, gene therapy, the instrument of genetic engineering, which consists in intervening on the genes of a cell to **modify** some of these functions or abilities, is very advanced in the therapy of cancers and much other genetic pathology. As far as the treatment of HIV infection is concerned,

there are already many lines of research, some protocols for trials in humans have been approved and some are even being evaluated.

8.10.1. How to modify a cell's genes?

Two therapeutic approaches are now used to combat HIV:

Direct modification of the patient's cell genes:

We inject the patient with a virus carrying the gene we want to modify. This virus will integrate into the DNA of the cells and insert the transgene. The cells thus acquire the new properties conferred by this gene, for example, by producing a toxin against the AIDS virus.

Use of cells that have already been modified:

The patient is injected with human cells, for example, immune cells such as CD4 lymphocytes, which have been genetically modified by the method described above in the laboratory. These chimera cells act as a drug in fighting the disease. They are now able to resist the AIDS virus and can provide immunity against infections.

8.10.2. Strategies to combat HIV

- Intracellular immunization,
- Selective destruction of infected cells,
- Secretion of inhibitory proteins,
- Genetic pharmacomodulation,
- Genetic immunotherapy.

8.2. Risks, Benefits, and Perceptions of Genetic Engineering

Many types of genetic engineering have yielded clear benefits with **few apparent risks**. Few would question, for example, the value of our now abundant supply of human insulin produced by genetically engineered bacteria. However, many emerging applications of genetic engineering are much more **controversial**, often because their potential benefits are pitted against significant risks, real or perceived. This is certainly the case for **gene therapy**, a clinical application of genetic engineering that may one day provide a cure for many diseases but is still largely an **experimental approach** to treatment.

8.2.1. Mechanisms and Risks of Gene Therapy

Human diseases that result from **genetic mutations** are often **difficult** to treat with drugs or other traditional forms of therapy because the signs and symptoms of disease result from abnormalities in a patient's genome. For example, a patient may have a genetic mutation that prevents the expression of a specific protein required for the normal function of a particular cell type. This is the case in patients with Severe Combined Immunodeficiency (SCID), a genetic disease that impairs the function of certain white blood cells essential to the immune system. **Gene therapy** attempts to **correct genetic abnormalities** by introducing a nonmutated, functional gene into the patient's genome. The nonmutated gene encodes a functional protein that the patient would otherwise be unable to produce. Viral vectors such as **adenovirus** are sometimes used to introduce the functional gene; part of the viral genome is removed and replaced with the desired gene. More advanced forms of gene therapy attempt to **correct the mutation** at the original site in the genome, such as is the case with treatment of SCID. So far, gene therapies have proven relatively ineffective, with the possible exceptions of treatments for cystic fibrosisand adenosine deaminase deficiency, a type of SCID. Other trials have shown the **clear hazards of attempting genetic manipulation in complex multicellular organisms** like humans. In some patients, the use of an **adenovirus** vector can trigger an unanticipated **inflammatory** response from the immune system, which may lead to **organ failure**. Moreover, because viruses can often target multiple cell types, the virus vector may **infect cells not targeted** for the therapy, damaging these other cells and possibly leading to illnesses such as cancer. Another potential risk is that the modified virus could revert to being **infectious** and cause **disease** in the patient. Lastly, there is a risk that the inserted gene could unintentionally **inactivate** another **important gene** in the patient's genome, disrupting normal cell cycling and possibly leading to tumor formation and **cancer**. Because gene therapy involves so many risks, candidates for gene therapy need to be fully informed of these risks before providing informed consent to undergo the therapy.

8.2.2. Ethical Concerns

Beyond the health risks of gene therapy, the ability to genetically modify humans poses a number of ethical issues related to the limits of such "therapy." While current research is focused on gene therapy for genetic diseases, scientists might one day **apply these methods to manipulate other genetic traits not perceived as desirable**. This raises questions such as:

The ability to alter reproductive cells using gene therapy could also **generate new ethical dilemmas**. To date, the various types of gene therapies have been targeted to somatic cells, the non-reproductive cells within the body. Because somatic cell traits are not inherited, any genetic changes accomplished by somatic-cell gene therapy would not be passed on to offspring. However, should scientists successfully introduce new genes to germ cells (eggs or sperm), **the resulting traits could be passed on to offspring**. This approach, called **germ-line gene therapy**, could potentially be used to combat heritable diseases, but it could also lead to **unintended consequences** for future generations. Moreover, there is the question of informed consent, because those impacted by germ-line gene therapy are **unborn** and therefore **unable to choose** whether they receive the therapy. For these reasons, the governments do not currently fund research projects investigating germ-line gene therapies in humans.

Summary

- While gene therapy shows great promise for the treatment of genetic diseases, there are also significant risks involved.
- There is considerable federal and local regulation of the development of gene therapies by pharmaceutical companies for use in humans.
- Before gene therapy use can increase dramatically, there are many ethical issues that need to be addressed by the medical and research communities, politicians, and society at large.

BIBLIOGRAPHY

Bibliography

Adams, G. (2020). A beginner's guide to RT-PCR, qPCR and RT-qPCR. *The Biochemist, 42*(3), 48-53.

Aouf, A. (2016). Cours de Biologie Moléculaire et Génie Génétique. Polycopié destiné aux étudiants de 3ème année Licence de Microbiologie. Université Ferhat Abbas-Sétif1-Algeria. 133 p.

Brown T. A. (2010). Gene Cloning and DNA Analysis: An Introduction. 6thEd, Weley Blackwell. London-UK. 782p.

Gubta, P. K. (2005). Microbiology, Cell Physiology and Biotechnology. 2ndEd. National Offset Printers, Meerut, India. 311 p.

Illumina (2024). Introduction to NGS. URL: https://sapac.illumina.com/science/technology/next-generation-sequencing.html. Accessed March 1st, 2024.

Kurnaz, I. A. (2015). Techniques in genetic engineering. CRC press. 277 p.

Nair, A. J. (2008). Introduction to biotechnology and genetic engineering. Laxmi Publications, Ltd.. 798 p.

Natesan, M & Ulrich. G. R. (2010). Protein Microarrays and Biomarkers of Infectious Disease. *International Journal of Molecular Sciences, 11*(12): 5165–5183.

Nicholl, S. T. D. (2008). An introduction to Genetic Engineering. 3rd Edition. Cambridge University Press. UK. 292 p.

OpenStax CNX. (2024). Microbiology. Open Education Resource (OER: LibreTexts Project (https://LibreTexts.org).

Prescott, M. L., Harley, J. P. & Klein, D. A. (2007). Microbiologie. 2ndEd. Edition De Boek Université. Paris. 1137 p.

Shikara, M. (2009). Genetic Engineering. Guide to genetic engineering: Fourth Year. Biotechnology Division, Department of Applied Sciences, University of Technology, Baghdad, Iraq. 131 p.

Sibbald, B. (2001). Death but one unintended consequence of gene-therapy trial. *Canadian Medical Association. Journal*, *164*(11), 1612.

Singh, B. D. (2007). Molecular Biology and genetic engineering. Kalyani Publishers. India. 311p.

Spigaglia, P., Barbanti, F., Dionisi, A. M., & Mastrantonio, P. (2010). *Clostridium difficile* isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype 018. *Journal of clinical microbiology, 48*(8), 2892-2896.

Watson, J. T., Baker, S., Bell, A., Gann, M., Levine, R. L. (2009). Biologie Moléculaire du gène. 6ème Ed. Pearson Education- Paris-France. 688 p.

GLOSSARY

Glossary

Agarose gel electrophoresis: is a technique that allows for the separation of DNA molecules based on size.

Alkaline phosphatase: An enzyme that removes phosphate groups from substrate molecules preferably in alkaline conditions.

Antibiotic screening: Screening of transformation using anti-biotic selection marker.

Bacterial artificial chromosome (**BAC**): an F plasmid-based vector for cloning large inserts, commonly used for sequencing projects.

Bacteriophage (phage): A virus that infects bacteria.

Biolistic: A method for DNA delivery into cells that are other -wise difficult to transform. It involves gold or tungsten microparticles coated with DNA ("bullets"), which are then accelerated to a high velocity so as to pen -etrate even cell walls. Also called gene gun or particle bombardment.

Biotechology : is the science of utilizing living systems to benefit humankind. In recent years, the ability to directly alter an organism's genome through **genetic engineering** has been made possible due to advances in **recombinant DNA technology**, which allows researchers to create **recombinant DNA** molecules with new combinations of genetic material.

Blue-white screening allows selection of bacterial transformants that contain recombinant plasmids using the phenotype of a reporter gene that is disabled by insertion of the DNA fragment.

Blunt end: The base-paired end of a double-stranded DNA molecule.

cDNA library: A compilation of cDNA fragments corresponding to most of the mRNAs expressed in that cell or tis-sue, cloned into an appropriate library vector and host organism.

cDNA: Also called complementary DNA , cDNA is a DNA molecule that is reverse transcribed using mRNA as a template (hence is "complementary" to mRNA). In eukaryotes, because mRNA is processed and introns are removed, cDNA corresponds to the "intronless" message.

Cell culture: A technique to grow, propagate, and manipulate live cells under defined physical and chemical conditions (i.e., pH, temperature, medium, and supplements).

Cloning: The process of creating genetically identical copies (animal cloning); the technique used to create copies of (and amplify) fragments of target DNA using a vector and a host organism, such as bacteria (biotechnology).

Cosmid: A type of artificially engineered cloning vector (commonly for library construction), which includes elements from phade genome (the "cos" sites) and some features of plasmids (the "-mid"), thereby allowing cloning of larger size DNA fragments.

DNA fingerprinting: A method that distinguishes each individual's DNA with relatively high significance, by analyzing highly variable repeats on the human genome that is unique to each individual (just like a fingerprint, hence the name).

DNA sequencing: The method of determining the sequence of nucleotides in a DNA strand.

DNA: Deoxyribonucleic acid, a double- stranded polymer of a nucleotide unit, which consists of a deoxyribose sugar, a phosphate group, and a nitrogen-containing base; functions as the hereditary molecule and abbreviated as DNA.

Endonuclease: Enzyme that cuts from within a nucleic acid sequence, by cleaving phosphodiester bonds between two nucleotides.

Expression vector: A vector DNA that is specifically used for the cloning and expression of a coding sequence of a given gene in a particular host cell or organism.

Gene therapy: A therapeutic approach to diseases using genetic engineering tools. Ideally it involves replacement of a mutated or dysfunctional gene with a healthy, normal version, although other approaches are available.

Genetic engineering: The use of molecular biology tools and techniques to change an organism's genome by introducing a novel DNA.

Genome: The entire genetic material of an organism.

Genomic DNA library: A compilation of genomic fragments of a target organism, cloned into an appropriate library vector and host organism.

Genomic libraries can be made by cloning genomic fragments from one organism into plasmid vectors or into bacteriophage.

Genomics: The study of the complete genome of an organism.

In vitro: Outside the natural environment or organism.

Insertional vector: A vector where a specific site in the mid-dle is digested and the target sequence can be inserted.

Klenow fragment: The large fragment of *E. coli* DNA polymerase I, which retains the 5′-to-3′ polymerase and 3′-to- 5′-exonuclease activities.

Ligase (DNA ligase): A modifying enzyme that can ligate, or form a phosphodiester bond between, a 5′-phosphate group and a nearby free 3′-hydroxyl group in a DNA molecule.

Marker: Any genetic or other way to "mark" a particular cell, tissue, or organism; (in cloning) a selectable marker such as an antibiotic resistance gene identifies cells that have taken up the recombinant DNA; in transgenic animals or plants, fluorescent markers or antibiotic resistance markers, et cetera, help identify the organisms that have stably incorporated the transgene.

Methylase: DNA methylase, which adds methyl groups to DNA molecules at specific sites (such as adenine within a 5′-GATC-3′ in the case of Dam methylase).

Methylation: Chemical linkage of a methyl group to a sub-strate (such as DNA or protein).

Microarray technology: is a nucleic acid hybridization technique that allows for the examination of many thousands of genes at once to find differences in genes or gene expression patterns between two samples of genomic DNA or cDNA,

Molecular cloning: a technique that involves methods used to construct recombinant DNA and facilitate their replication in host organisms. These methods include the use of restriction enzymes (to cut both foreign DNA and plasmid vectors), ligation (to paste fragments of DNA together), and the introduction of recombinant DNA into a host organism (often bacteria).

Multiple cloning site (*polylinker*): A short engineered sequence within a cloning or expression vector that contains a number of unique restriction enzyme recognition sequences not present in the base vector.

Mutagenesis: Any process which causes a change in the DNA sequence.

Palindrome: A specific DNA sequence motif that "reads" the same 5′-to- 3′ on both strands of the sequence.

Plasmid: Small, circular, extrachromosomal, self-replicating DNA unit.

Polyacrylamide gel electrophoresis (PAGE): is a technique that allows for the separation of proteins by size, especially if native protein charges are masked through pretreatment with SDS.

Polymerase chain reaction: is a technique that allows for the rapid amplification of a specific DNA sequence. Variations of PCR can be used to detect mRNA expression (reverse transcriptase PCR) or to quantify a particular sequence in the original sample (real-time PCR). Although the development of Sanger DNA sequencing was revolutionary, advances in next generation sequencing allow for the rapid and inexpensive sequencing of the genomes of many organisms, accelerating the volume of new sequence data.

Polymerase: An enzyme that generates nucleic acid polymers from nucleotide building blocks.

Primer: A short stretch of single- stranded DNA that acts as a starting point or a primer of DNA replication.

Probe: A relatively long (in the range of 100–1000 bp) stretch of DNA or RNA that is radioactively or fluorescently labeled, and used for detection of a target.

Real-time PCR: A version of PCR reaction that is modified so as to use fluorescent dye molecules to optically monitor the reaction in real time.

Recombinant DNA: A new combination of DNA, generated by genetic engineering from two different DNA molecules.

Restriction enzyme (restriction endonuclease)**:** Endonucleases that recognize spe-cific DNA sequence motifs and cleave the sugar-phosphate backbone on DNA.

Restriction fragment length polymorphism (**RFLP**): A method devised to analyze genetic variations among samples, which are associated with variations in restriction enzyme recognition motifs within a DNA region.

Restriction map: A map of all known restriction sites on a given DNA region, with the nucleotide positions with respect to each other.

Reverse transcriptase: RNA- directed DNA polymerase of retroviruses, used for synthesis of cDNA from mRNA template.

RT-PCR: Reverse transcription polymerase chain reaction, used to PCR-amplify expressed sequences after converting them from mRNA to cDNA.

Screening: (In molecular cloning) a method to identify true recombinants among a group of transformed cells or organisms.

SDS-PAGE (sodium dodecyl sulphate poplyacrylamide gel electrophoresis**):** A method for separating denatured proteins according to size.

Southern blot: is a technique that allows researchers to find a particular DNA sequence within a sample whereas northern blot analysis allows researchers to detect a particular mRNA sequence expressed in a sample.

Sticky end: The non-base-paired end of a double-stranded DNA molecule, where a short stretch is "sticking" out.

Strain: A genetic variant of a microorganism, cell, plant, or animal used in a laboratory, with distinct and unique features, yet not categorized as an entirely different breed or variety.

Ti plasmid: The tumor-inducing (Ti) plasmid of the bacterium *Agrobacterium tumafeciens* that infects plants.

Transfection of eukaryotic hosts can be achieved through various methods using electroporation, gene guns, microinjection, shuttle vectors, and viral vectors.

Transgenic: An organism that has been genetically modified by a foreign DNA (the transgene) to achieve a novel property.

Vector: A vehicle DNA that is used to transfer foreign DNA into a host cell or organism.

Western blot: A method used to detect target protein species within a mixture through antigen-antibody recognition and signal detection.

Yeast artificial chromosome (YAC): Genetically engineered artificial chromosomes that carry key elements of yeast chromosomes.