Introduction to Biotechnology

Biotechnology is the use of microbes, animal/plant cells and their products to synthesize, break down or transform materials. Primarily it includes the use of recombinant DNA technology and genetic engineering techniques to improve upon the quality of processes.

Introduction to Biotechnology

Traditional biotechnology refers to the conventional techniques that have been used for many centuries to produce beer, wine, cheese etc.

Introduction to Biotechnology

Modern Biotechnology embraces all methods of genetic modification by recombinant DNA & cell fusion techniques together with the modern developments of traditional biotechnological processes.

Branches of Biotechnology

White Biotechnology

Development of processes and microorganisms for Industrial processes.

Key concept

Scale of operation

Low volume high price

Branches of Biotechnology

White Biotechnology

Example

- Enzyme production

Branches of Biotechnology

Red Biotechnology

It is concerned with the discovery and development of innovative drugs and treatments.

Branches of Biotechnology

Green Biotechnology

It is concerned with the modification of the genetic composition of plants to enhance existing traits or add new ones.

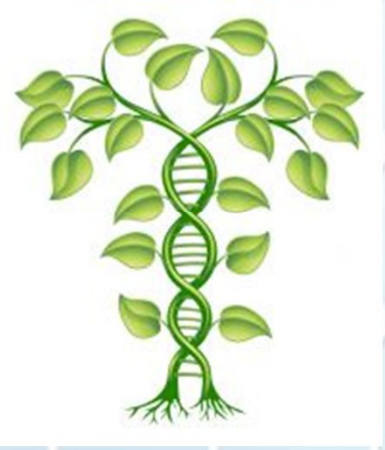
Branches of Biotechnology

Green Biotechnology

Example

Bt corn from Syngenta

Bt cotton from Monsanto



Modern Biotechnology (Principles & Applications)

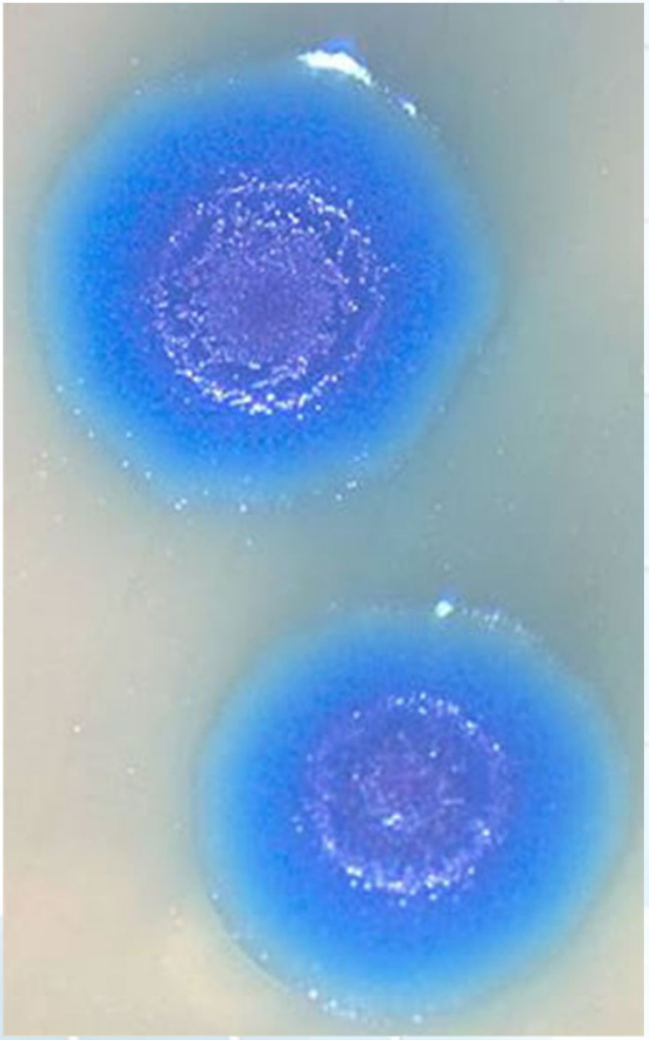
Course Credits 3

Course Instructor

Dr. Muhammad Safwan Akram

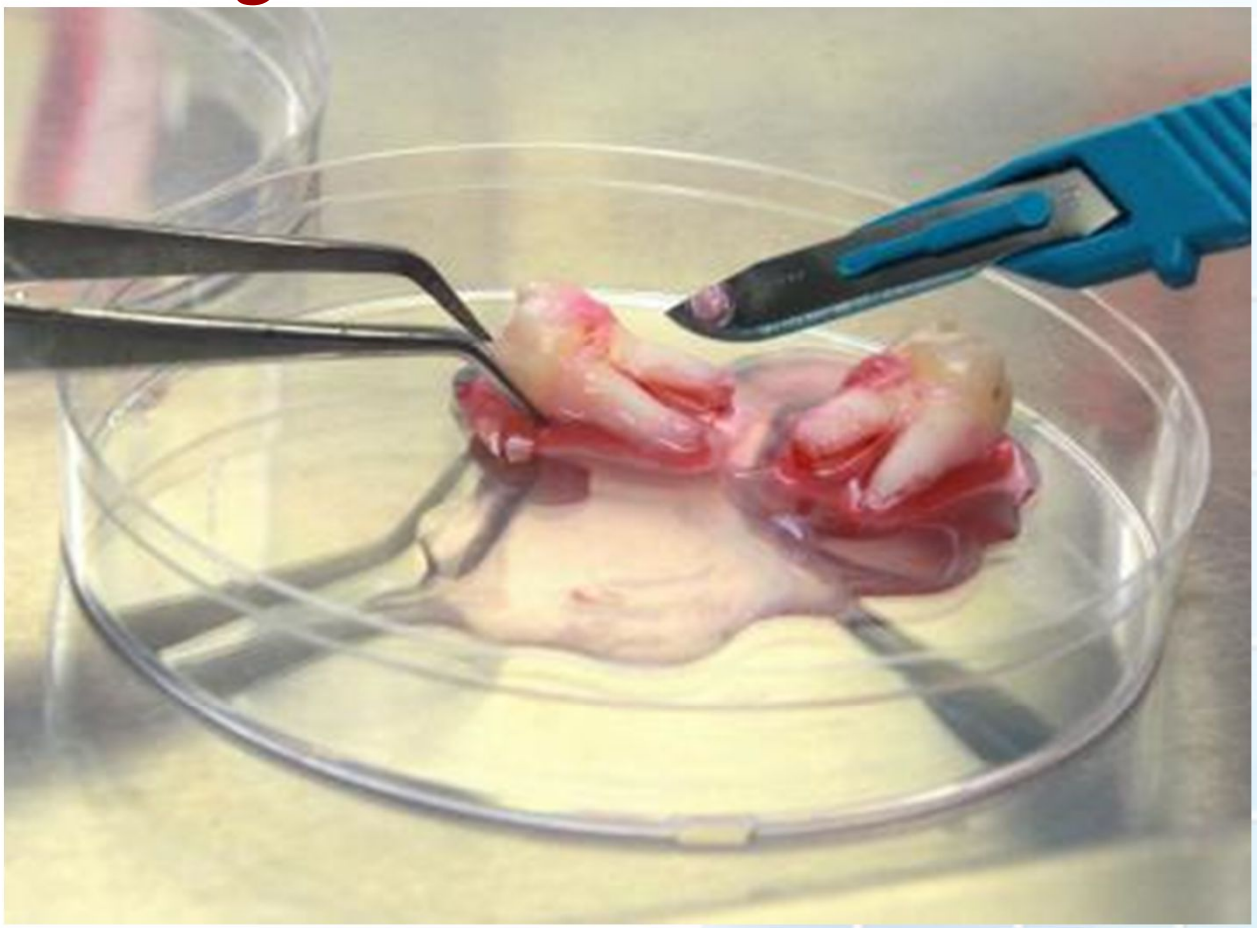
Ph.D. Biotechnology
(Univ. of Cambridge, UK)
M.Phil Bioscience Enterprise
(*Univ. of Cambridge, UK*)
M.Sc. Biotechnology
(*Univ. of the Punjab, Lahore, Pakistan*)
B.Sc. (Hons.) Biochemistry
(Univ. of the Punjab, Lahore, Pakistan)

Biotechnology in the 21st Century

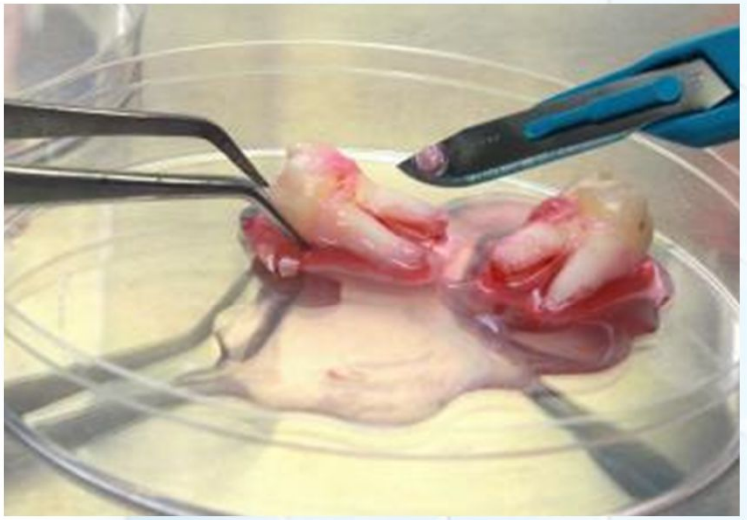


Biotechnology in the 21st Century

Growing Teeth

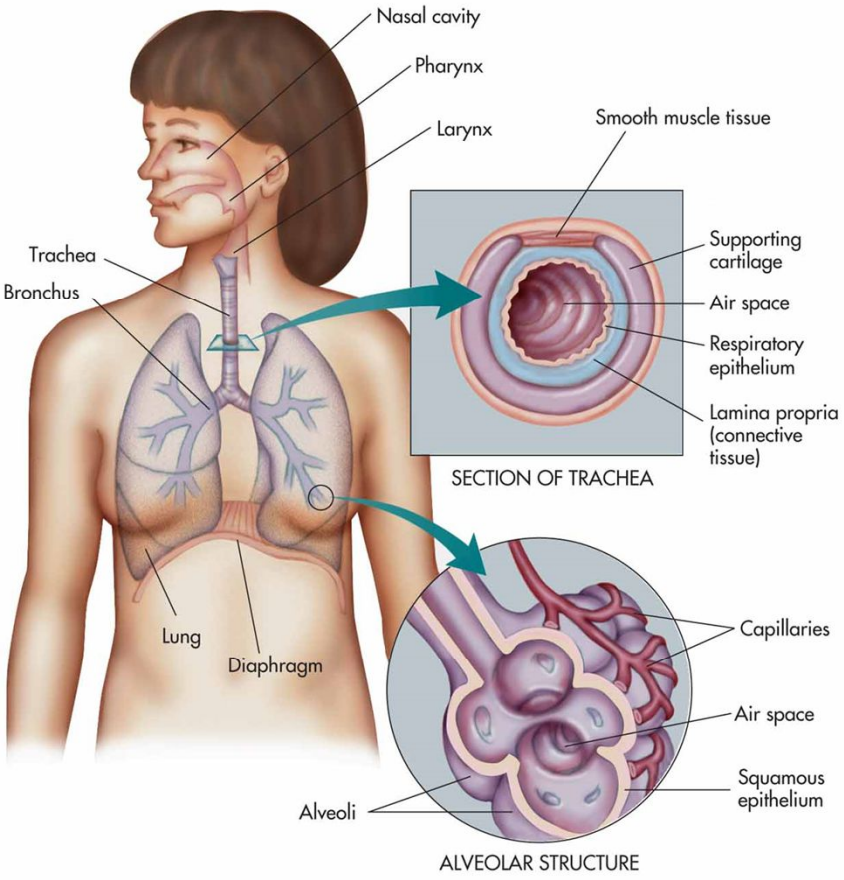


Biotechnology in the 21st Century

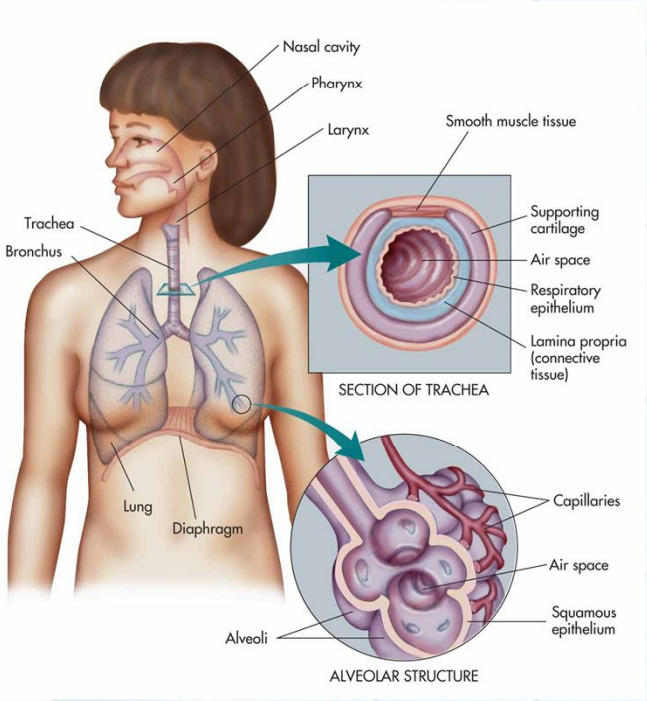


Biotechnology in the 21st Century

Claudia's Trachea

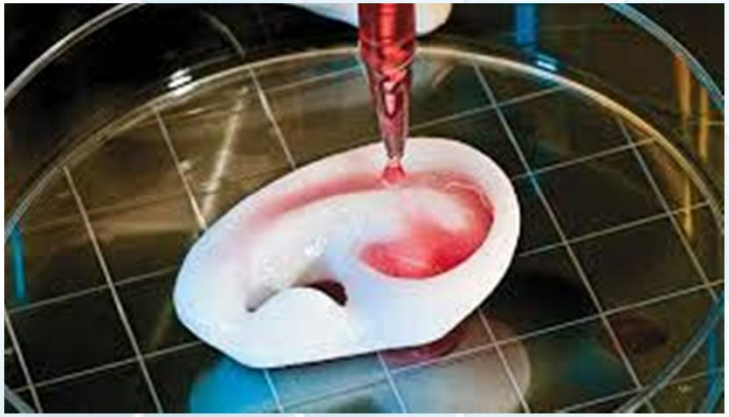


Biotechnology in the 21st Century



Biotechnology in the 21st Century

Growing Organs



Biotechnology in the 21st Century

Growing Blood Vessels

Biotechnology in the 21st Century

Chicken with an extra leg

HARVARD MEDICAL SCHOOL

Biotechnology in the 21st Century

Growing Meat

Roots of Biotechnology

6000 BC

Sumerians & Babylonians used **anaerobic respiration** of yeast to make beer.

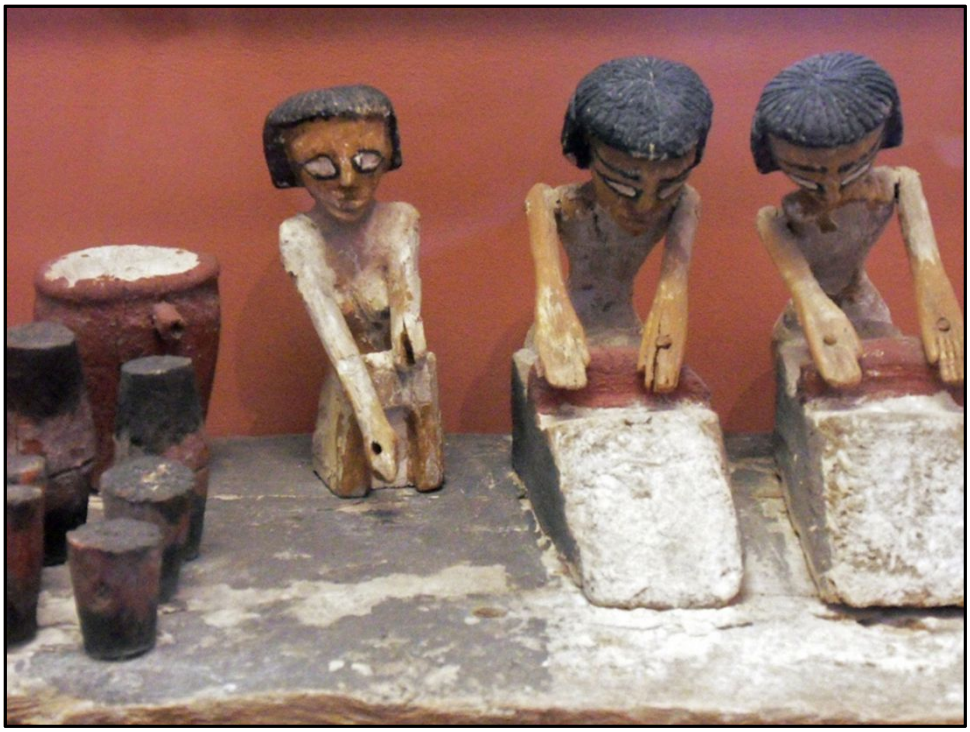
Sumerians are people from Southern Iraq

Roots of Biotechnology

4000 BC

Egyptians used **aerobic respiration** of yeast to leaven or rise bread and used mold to flavour cheese.

Fermented food is protected for longer and is easier to digest



Excavation in Egypt has revealed figurines showing bread making process

Roots of Biotechnology

1350 BC

Pregnancy test

Barley grows – male child

Wheat grows – female child

A papyrus described a test in which a woman who might be pregnant could urinate on wheat and barley seeds over the course of several days: “If the barley grows, it means a male child. If the wheat grows, it means a female child. If both do not grow, she will not bear at all.” Testing of this theory in 1963 found that 70 percent of the time, the urine of pregnant women did promote growth, while the urine of non-pregnant women and men did not. Scholars have identified this as perhaps the first test to detect a unique substance in the urine of pregnant women, and have speculated that elevated levels of estrogens in pregnant women’s urine may have been the key to its success

Roots of Biotechnology

1350 BC

Pregnancy test

Barley grows – male child

Wheat grows – female child

A papyrus described a test in which a woman who might be pregnant could urinate on wheat and barley seeds over the course of several days: “If the barley grows, it means a male child. If the wheat grows, it means a female child. If both do not grow, she will not bear at all.” Testing of this theory in 1963 found that 70 percent of the time, the urine of pregnant women did promote growth, while the urine of non-pregnant women and men did not. Scholars have identified this as perhaps the first test to detect a unique substance in the urine of pregnant women, and have speculated that elevated levels of estrogens in pregnant women’s urine may have been the key to its success

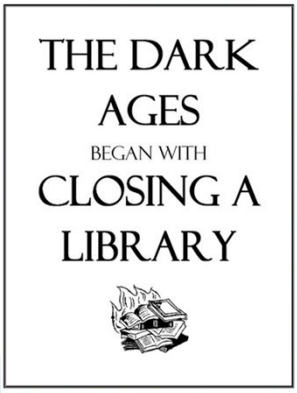
Roots of Biotechnology

400 BC

Use of salts to preserve food specially meat.

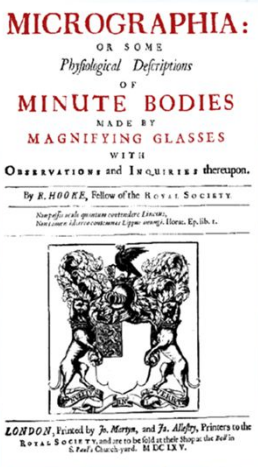
Roots of Biotechnology

European Dark Period &
Islamic Golden Era



Roots of Biotechnology

1665 AD



MICROGRAPHIA:

OR SOME

Physiological Descriptions

OF

MINUTE BODIES

MADE BY

MAGNIFYING GLASSES

WITH

OBSERVATIONS and INQUIRIES thereupon.

By R. HOOKE, Fellow of the ROYAL SOCIETY

LONDON, Printed by Jo. Martyn, and Ja. Allestry, Printers to the ROYAL SOCIETY, and are to be sold at their Shop at the Bell in S. Paul's Church-yard. M DC LX V.

First scientific best seller, written by Robert Hooke brought microscopy into scientific limelight

Roots of Biotechnology

1676 AD

Antonie Van Leeuwenhoek

Observation of microscopic single celled organisms

Very fine lens maker, considered father of microbiology. He made around 200 microscopes and 500 lenses with various magnifications. His lenses were ultimately replicated by C. L Strong in 1957 almost after 300 years.

Roots of Biotechnology

1828 AD

Synthesis of Urea

Fredrick Wohler

Pioneer in inorganic chemistry for synthesizing urea. This discovery became refutation of vitalism, the hypothesis that living things are alive because of some special "vital force"

Scientist were pondering the nature of living things? What is the difference between living things and inanimate things? Theodore Schwann defined cell as a membrane bound structure containing nucleus. The body was not infused by a mysterious force, and life originated from the actions of cells. 1834 to 1839 …..he was active as a scientists.

Roots of Biotechnology

1834 - 1839 AD

Cell Theory

Theodore Schwann

"All living things are composed of cells and cell products"

Yeast Pee!

Intellectual blind spot and 5 years of brilliance: Scientist were pondering the nature of living things? What is the difference between living things and inanimate things? Theodore Schwann defined cell as a membrane bound structure containing nucleus. The body was not infused by a mysterious force, and life originated from the actions of cells. 1834 to 1839 …..he was active as a scientist. In that final year, he suffered a vicious personal attack by Fredrick Wohler. They mocked his idea that alcoholic fermentation was the result of yeast acting on sugars, insultingly drawing cartoons of yeast excreting wine through their wine bottled shaped imaginary asses. Prevailing theory at that time was that sugar transformed by reacting with air and nitrogenous substances in fruit juices. Most highly respected authorities denied Schwann any funds and his career was over.

Roots of Biotechnology

1857 AD

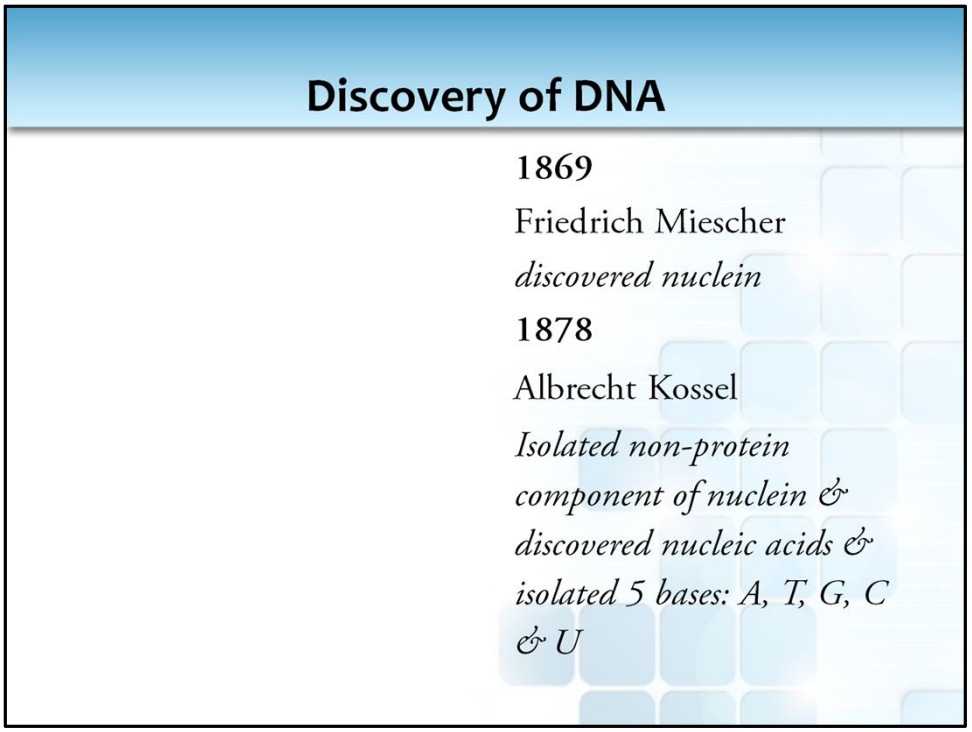
Rudolf Virchow

Omnis cellula e cellula

"Every cell arise from another cell"

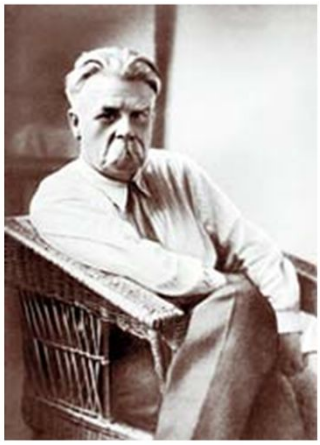
End

We will discuss vaccines and the role they played in a separate lecture.



Sumerians are people from Southern Iraq

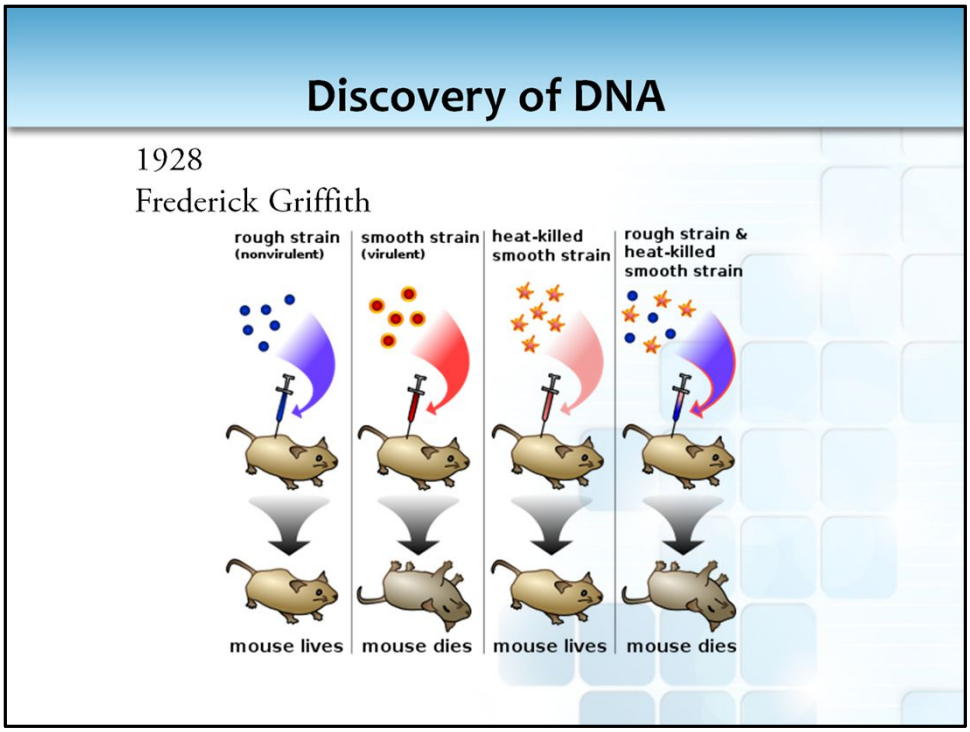
Discovery of DNA



1927

Nikolai Konstantinovich Koltsov

Inherited traits would be inherited via a "giant hereditary molecule" which would be made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template"



Discovered transformation, proof that DNA contains genetic information,

different strains of streptococcus pneumoniae, Type 3S and Type 2R

Discovery of DNA

1944

Avery, MacLeod & McCarty Experiment

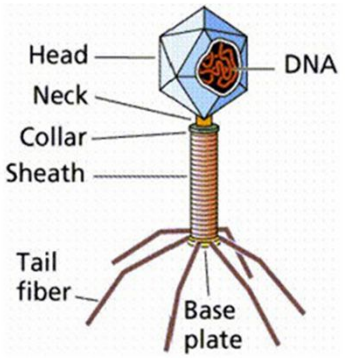
Discovered transformation, proof that DNA contains genetic information,

different strains of streptococcus pneumoniae, Type 3S and Type 2R

Discovery of DNA

1952

Hershey-Chase Experiment

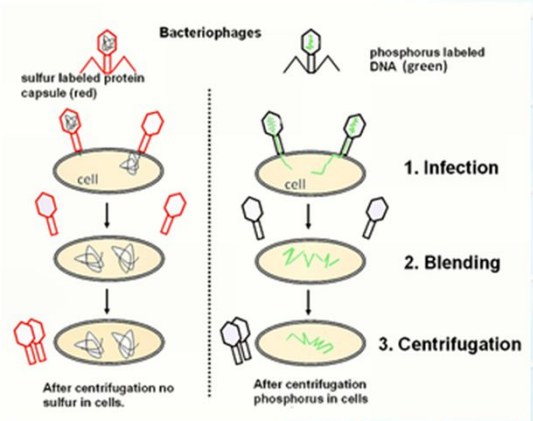


Radioactive phosphorus 32 and Sulfur 35, Sulfur is only present in protein

Discovery of DNA

1952

Hershey-Chase Experiment



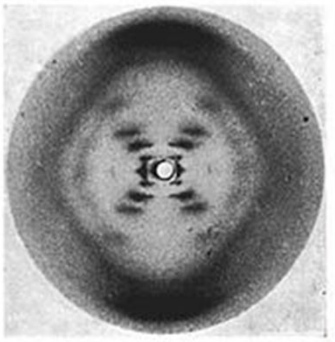
Radioactive phosphorus 32 and Sulfur 35, Sulfur is only present in protein

Discovery of DNA

Photo 51

Rosalind Franklin

Maurice Wilkins



Rosalind Franklin (died at the age of 37) supervisor of Maurice Wilkins. Wilkins said its alpha

helical while Franklin denied.

Discovery of DNA

Watson & Crick 1953

Rosalind Franklin (died at the age of 37) supervisor of Maurice Wilkins. Wilkins said its alpha

helical while Franklin denied.

Discovery of DNA

No. 4356 April 25, 1953 NATURE 737

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons:

(1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press), in his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxy-ribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions.

Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain, does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain, is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in time following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J.D.W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

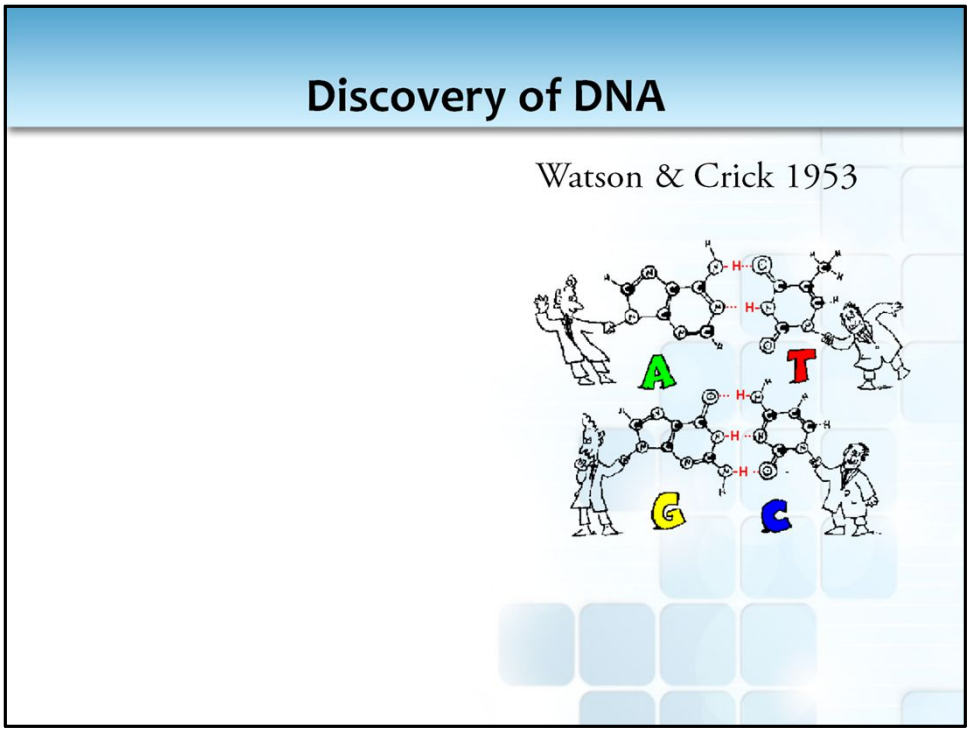
J.D. WATSON
F.H.C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.

[1] Pauling, L., and Corey, R. B., *nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G.R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., *Symp. Soc. Exp. Biol. 1, Nucleic Acid*, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F. and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Rosalind Franklin (died at the age of 37) supervisor of Maurice Wilkins. Wilkins said its alpha

helical while Franklin denied.

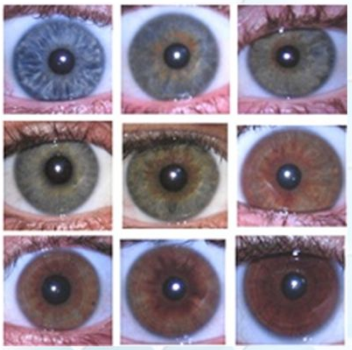


Rosalind Franklin (died at the age of 37) supervisor of Maurice Wilkins. Wilkins said its alpha

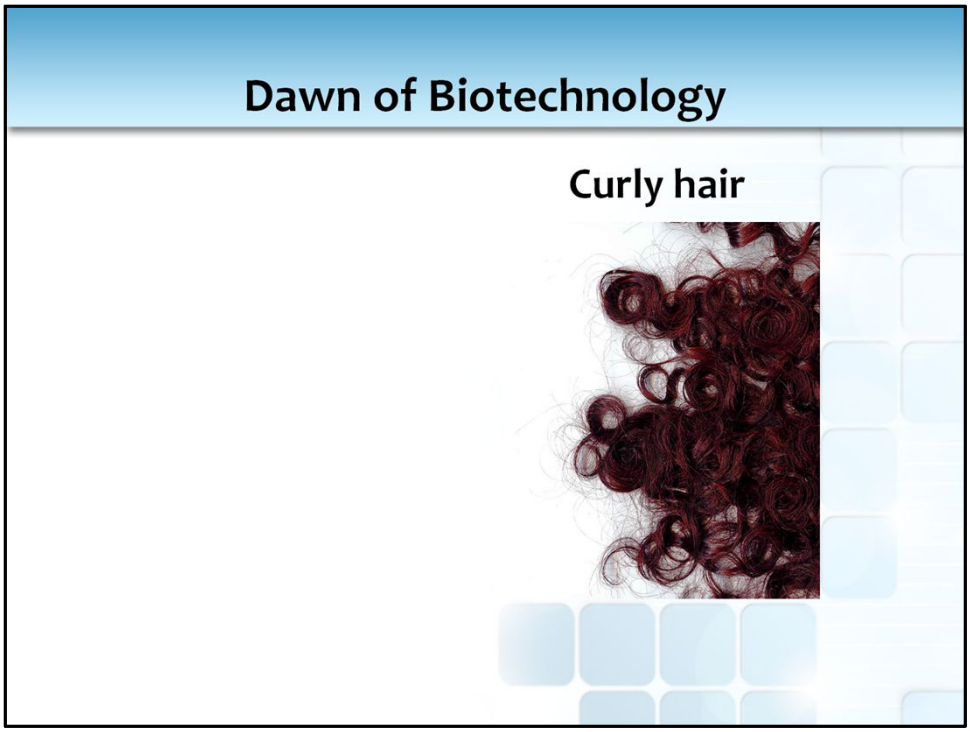
helical while Franklin denied.

Dawn of Biotechnology

Spectrum of Eye Colour



Once the structure of DNA was discovered, hunt for genes become red hot.

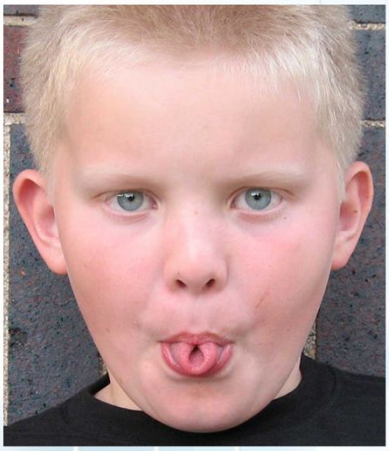


Round hair follicles make straight hair, flattened or c-shaped hair follicles make curly hair, and oval hair follicles make wavy hair. Hair texture is a continuous trait, meaning that hair can be straight or curly or anywhere in between.

Multiple genes control hair texture, and different variations in these genes are found in different populations. For instance, curly hair is common in African populations, rare in Asian populations, and in-between in Europeans. Straight hair in Asians is mostly caused by variations in two genes—different genes from the ones that influence hair texture in Europeans.

Dawn of Biotechnology

Tongue Rolling



Controlled by one gene but some environmental factors may play a role as people can

learn to

Dawn of Biotechnology

Can we control the traits?
Can we isolate the genes?
Can we edit the genes?
Can we introduce them in new species?

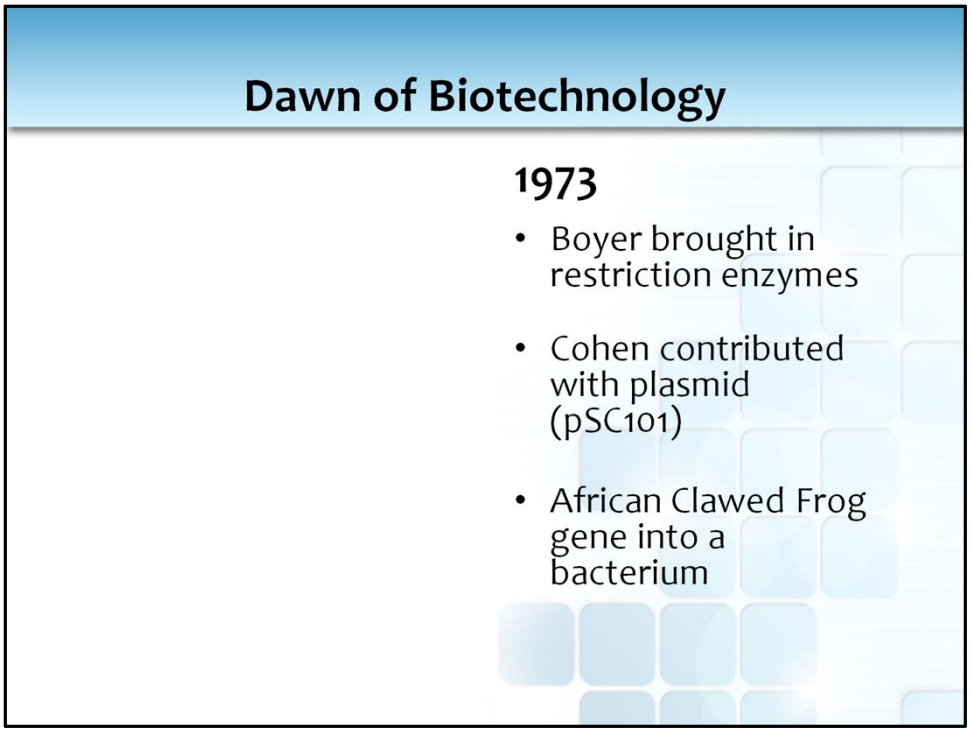
Freckles

Freckles are controlled primarily by the single MC1R gene

Dawn of Biotechnology

Restriction Endonucleases

Introduce new genes we required something to cut the DNA at a specific location
Herbert Boyer, Stanley Norman Cohen and Paul Berg

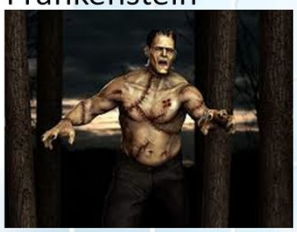


pSC101 had tetracycline resistance

Dawn of Biotechnology

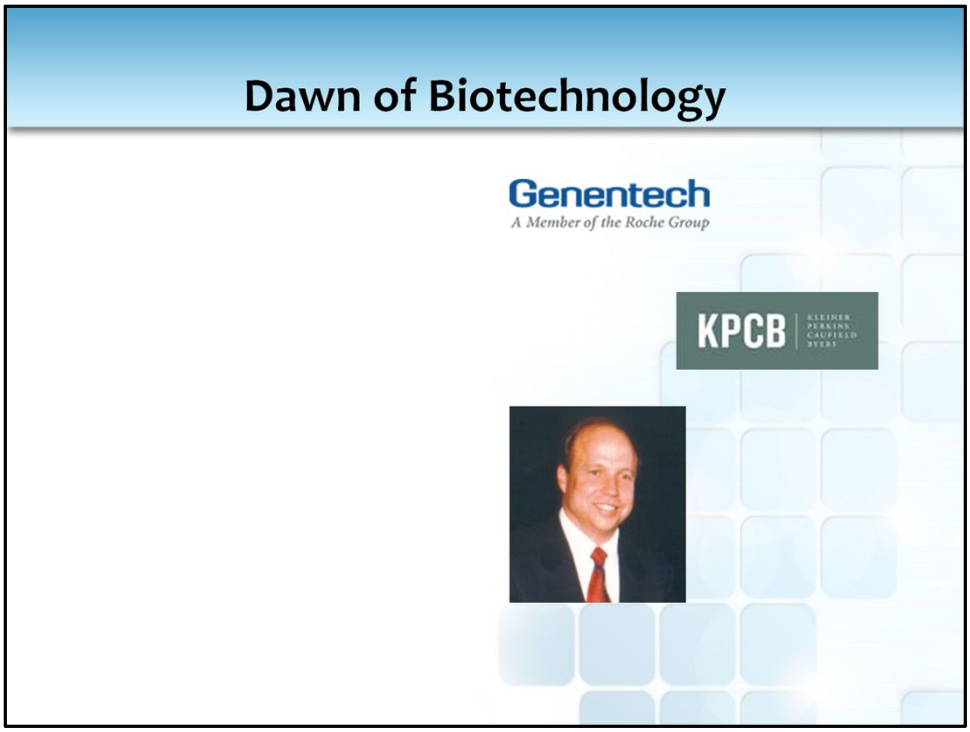
Negative Perception

"Genetic Engineering can give rise to something which can crawl out of the laboratory such as Frankenstein"



Local politician in Massachusetts

Dawn of Biotechnology

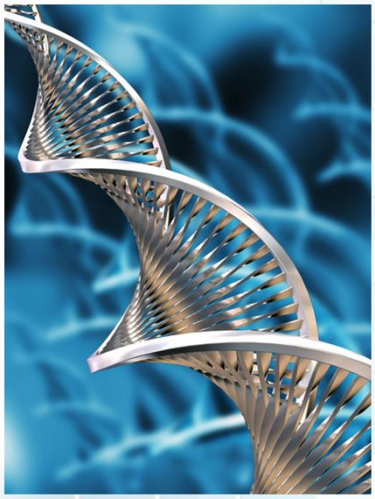


Boyer went onto make Genentech in 1976. Robert Swanson was the first VC to invest in Biotechnology.

Dawn of Biotechnology

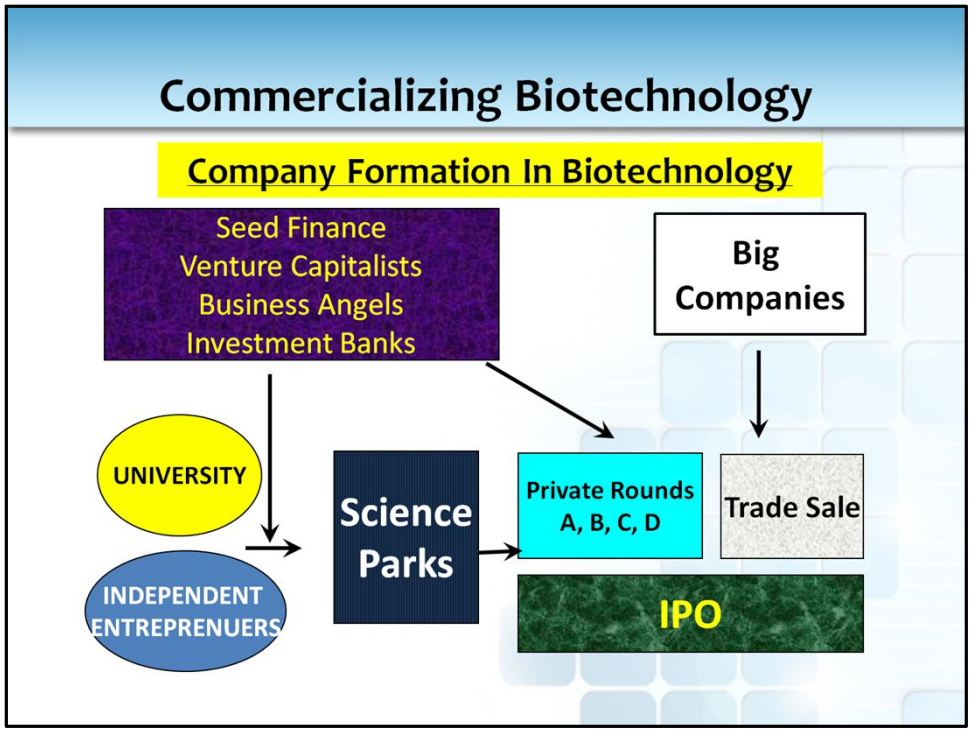
Somatostatin was the first protein to be cloned in E. coli to produce a biopharmaceutical at large scale and Humulin was the first recombinant Insulin produced by Genentech but marketed by Eli Lilly

Commercializing Biotechnology



In 1982, Amgen tendered an IPO raising $42.3 million. This came on the heels of a remarkably successful IPO by rival Genentech, which created a short-lived biotech frenzy on Wall Street.

Commercializing Biotechnology



Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
Genentech	10/80	$35M
Cetus	03/81	$107M
Genetic Systems	04/81	$6M
Ribi Immunochem	05/81	$1.8M
Genome Therapeutics	O5/82	$12.9M
Centocor	12/82	$21M
Scios	01/83	$12M

In 1982, Amgen tendered an IPO raising $42.3 million. This came on the heels of a remarkably successful IPO by rival Genentech, which created a short-lived biotech frenzy on Wall Street.

Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
Biotechnology General	09/83	$8.9M
Immunex	03/83	$16.5M
Amgen	06/83	$42.3M
Biogen	06/83	$57.5M
Chiron	08/83	$17M
Immunomedics	11/83	$2.5M
Repligen	04/86	$17.5M

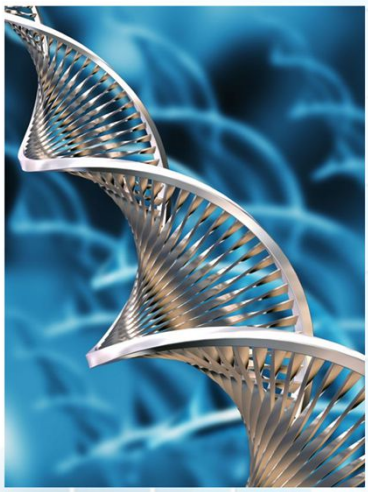
Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
OSI	04/86	$13.8M
Cytogen	06/86	$35.6M
Xoma	06/86	$32M
Genzyme	06/86	$28M
ImClone	06/86	$32M
Genetics Institute	05/86	$79M

Reference: from Alchemy to IPO (The business of Biotechnology) by Cynthia Robbins-Roth (2000)

Commercializing Biotechnology



Commercializing Biotechnology

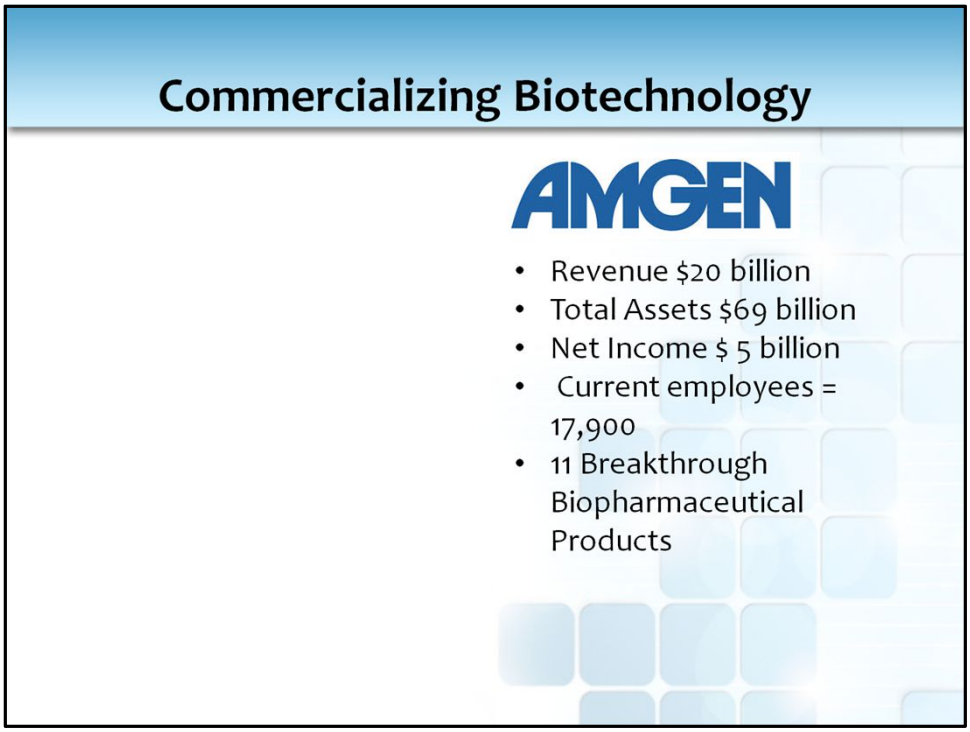
- Roche acquired Genentech for $46.8 billion
- Current employees = 12,895
- 19 Breakthrough Biopharmaceutical Products

Commercializing Biotechnology

Genentech
IN BUSINESS FOR LIFE

Roche

DR. HERBERT BOYER



Applied Molecular Genetics

Commercializing Biotechnology

AMGEN

DR. WINSTON SALSER

Applied Molecular Genetics

Commercializing Biotechnology

Biogen

- Revenue $9.7 billion
- Total Assets $15 billion
- Net Income $ 3 billion
- Current employees = 7,550
- 10 Breakthrough Biopharmaceutical Products

Applied Molecular Genetics

Commercializing Biotechnology

Biogen

THE UNIVERSITY OF EDINBURGH

Sir Kenneth Murray

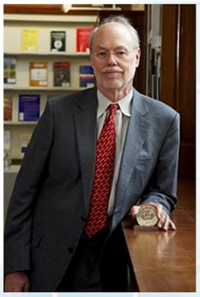
UNIVERSITAS TURICENSIS MDCCC XXXIII

Charles Weissmann

Commercializing Biotechnology

Biogen

Phillip Sharp



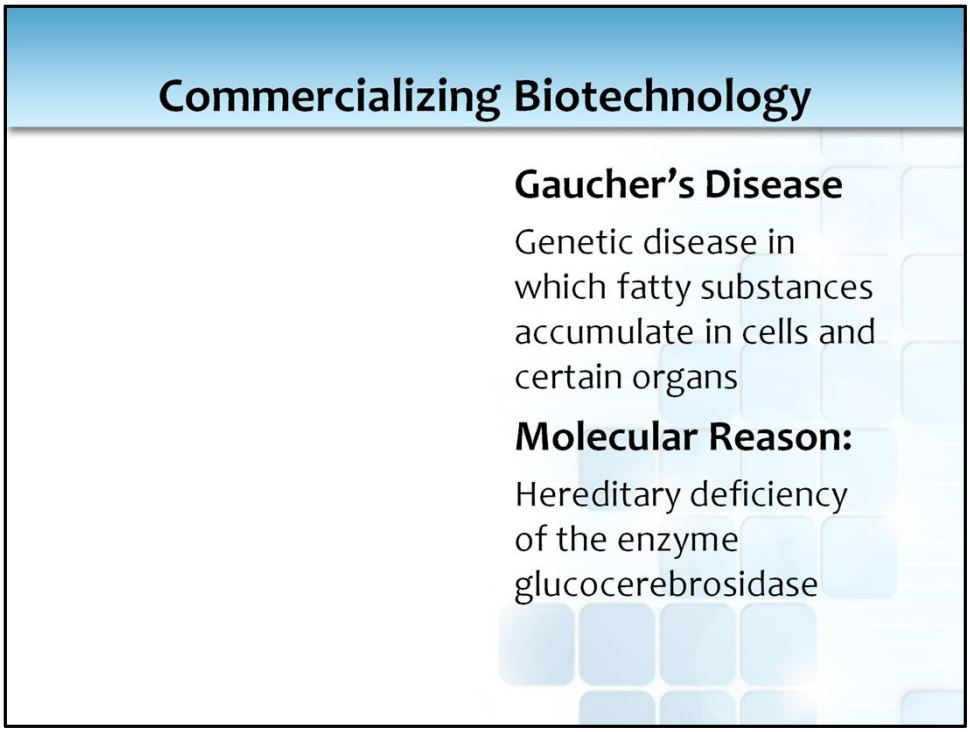
MIT

Commercializing Biotechnology

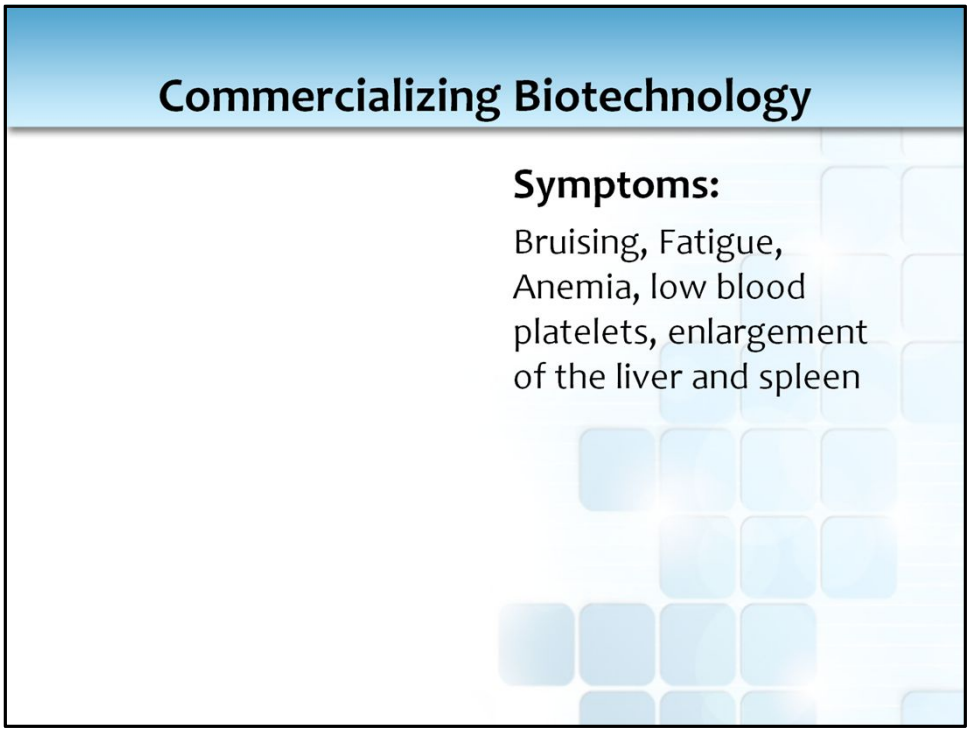
genzyme

- Revenue $4.6 billion
- Net Income $ 421 million
- Current employees = 12,000
- 16 innovative Biopharmaceutical Products

Applied Molecular Genetics



Accumulation of a sphingolipid named ceramide that doesn't get metabolised due to lack of glucosidase enzyme



Accumulation of a sphingolipid named ceramide that doesn't get metabolised due to lack of glucosidase enzyme

Commercializing Biotechnology

Prior to this, they had to collect 22,000 placentas to treat Gaucher's patients

Commercializing Biotechnology

genzyme

Tufts UNIVERSITY

DR. HENRY BLAIR

Applied Molecular Genetics

Commercializing Biotechnology

genzyme

DR. HENRY BLAIR

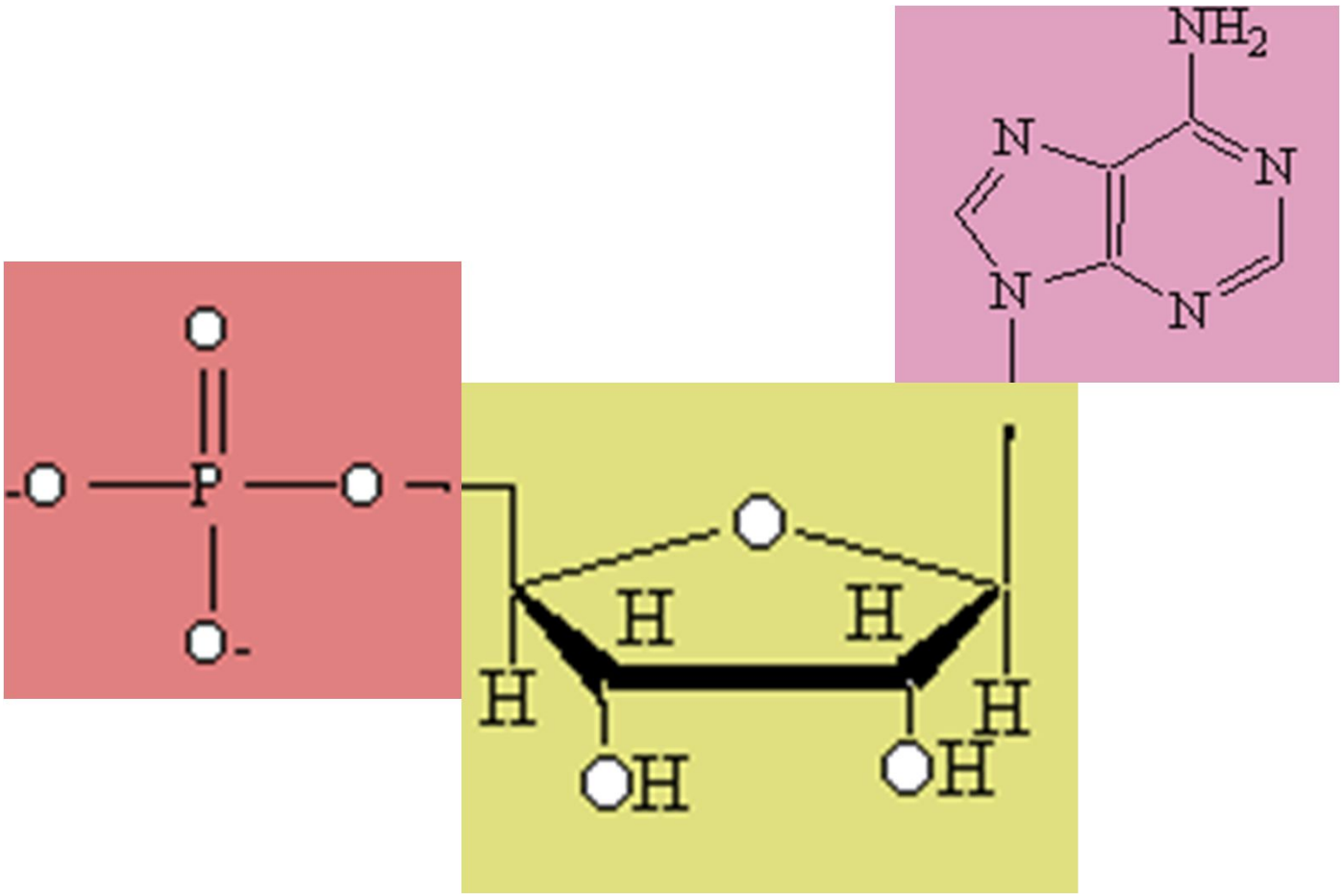
Tufts UNIVERSITY

Applied Molecular Genetics

Nucleoside & Nucleotide

- A molecule containing all these three components is called a nucleotide.
- While a molecule without the phosphate group is called a nucleoside.

Nucleoside & Nucleotide



Nucleoside & Nucleotide

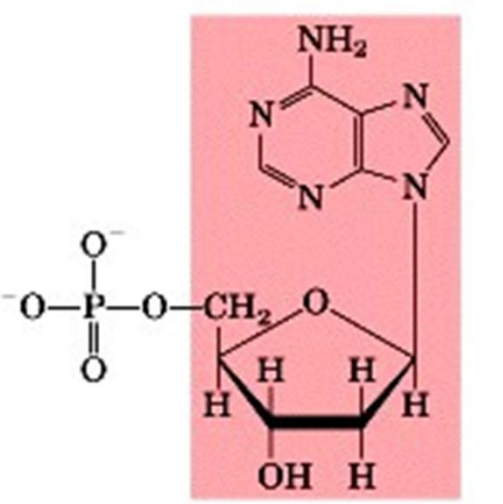
Nucleotide =
Nucleoside +
Phophoric acid
&
Nucleoside =
Nucleotide –
Phoshoric acid

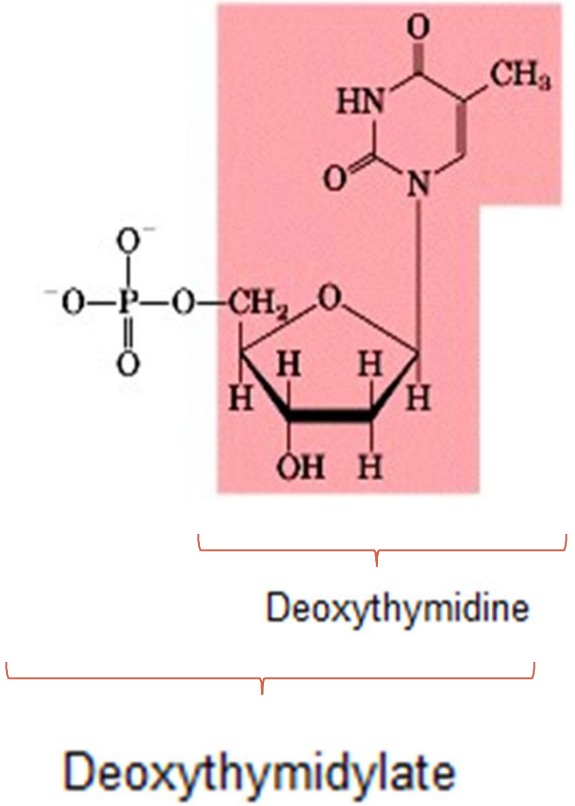
end

Types of Deoxyribonucleotides

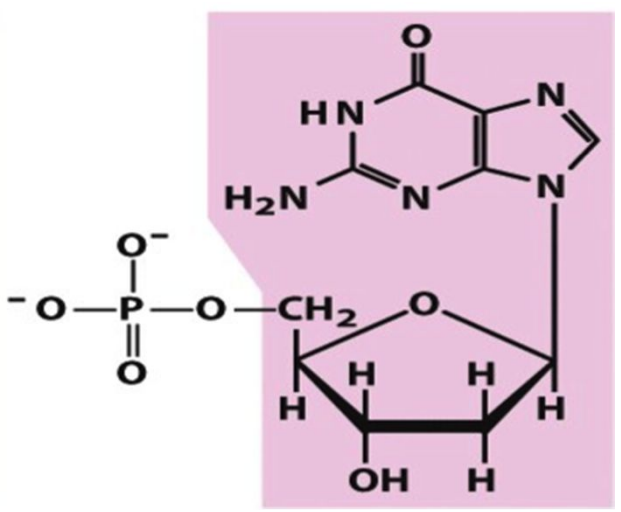
- There are four types of Deoxy-ribonucleotides.

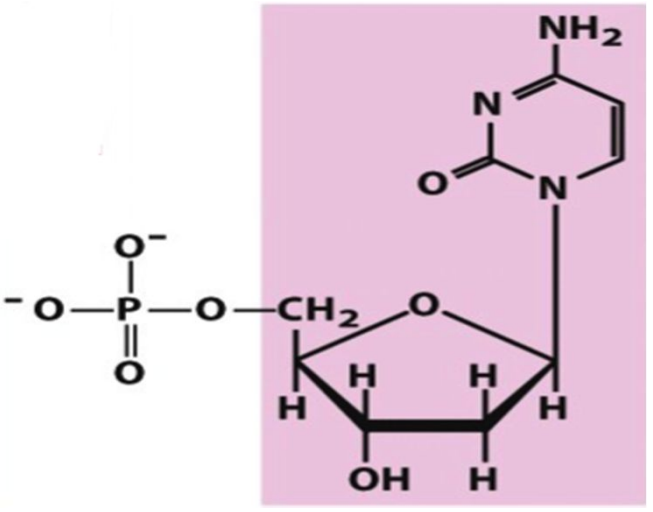
Types of Deoxyribonucleotides





Types of Deoxyribonucleotides





Types of Deoxyribonucleotides

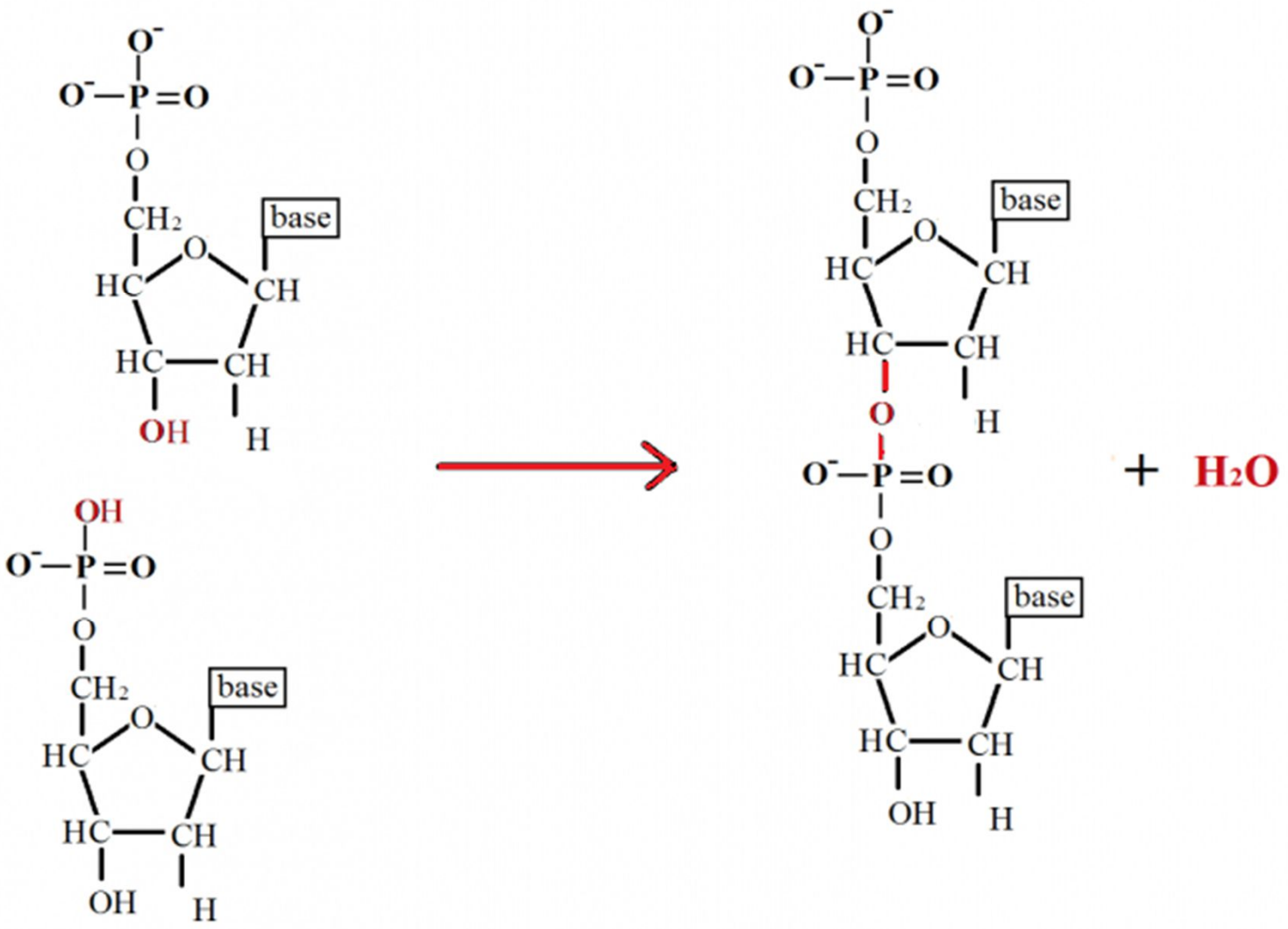
- These four deoxy-ribonucleotides make the structural units of DNA.

end

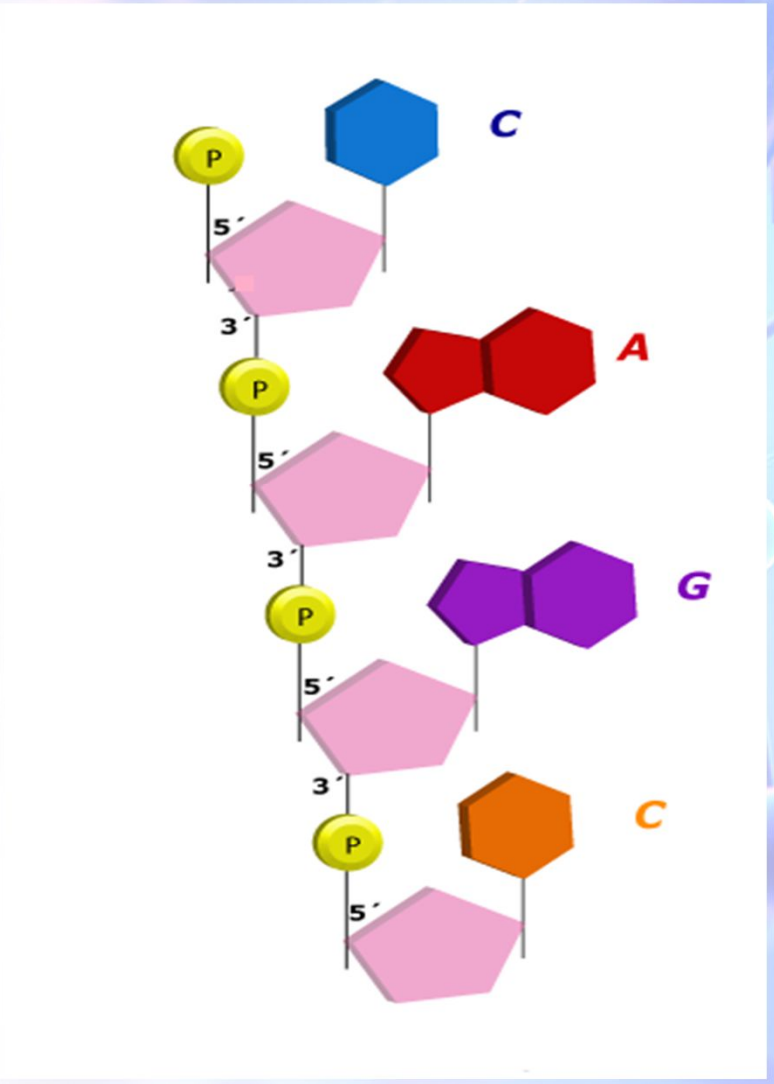
How do Deoxyribonucleotides Join?

- The successive nucleotides of DNA are joined together through phospho-diester linkages.

How do Deoxyribonucleotides Join?



How do Deoxyribonucleotides Join?



Structure of DNA

Work of Chargaff (Late 1940s)

- The discovery of the structure of DNA is one of the greatest events in the history of science.

Structure of DNA

Work of Chargaff (Late 1940s)

- Erwin Chargaff and his colleagues provided a most important clue to the structure of DNA.
- The work of Chargaff led him to following conclusions, also called “Chargaff Rules”:-

Structure of DNA

Work of Chargaff (Late 1940s)

1. Base composition of DNA varies from one species to another.

Structure of DNA

Work of Chargaff (Late 1940s)

2. The DNA isolated from different tissues of the same species have the same base composition.

Structure of DNA

Work of Chargaff (Late 1940s)

3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.

Structure of DNA

Work of Chargaff (Late 1940s)

4. In DNA, the number of adenosine residues is equal to the number of thymidine (A=T) and the number of guanosine residues is equal to the number of cytidine (G=C).

Structure of DNA

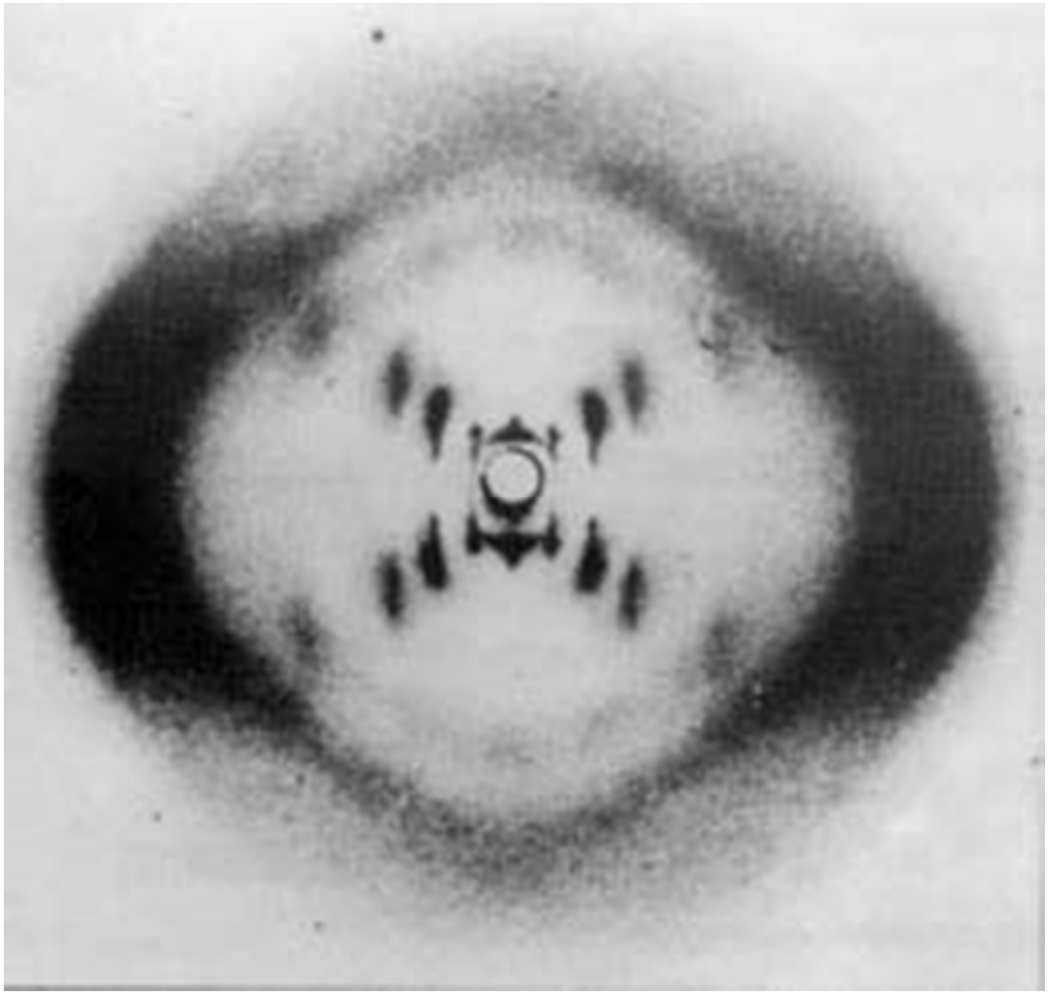
Work of Chargaff (Late 1940s)

- It means that the sum of the purine residues equals the sum of the pyrimidine residues (AG=TC).

Work of Franklin & Wilkins (1950s)

- Rosalind Franklin and Maurice Wilkins performed the x-ray diffraction analysis of DNA fibers.
- They showed that DNA produces a characteristic x-ray diffraction pattern.

Work of Franklin & Wilkins (1950s)



Work of Franklin & Wilkins (1950s)

- From this pattern, they made two important findings:-

1. DNA molecules are helical.
2. The helices have two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å.

Structure of DNA

Work of Watson & Crick

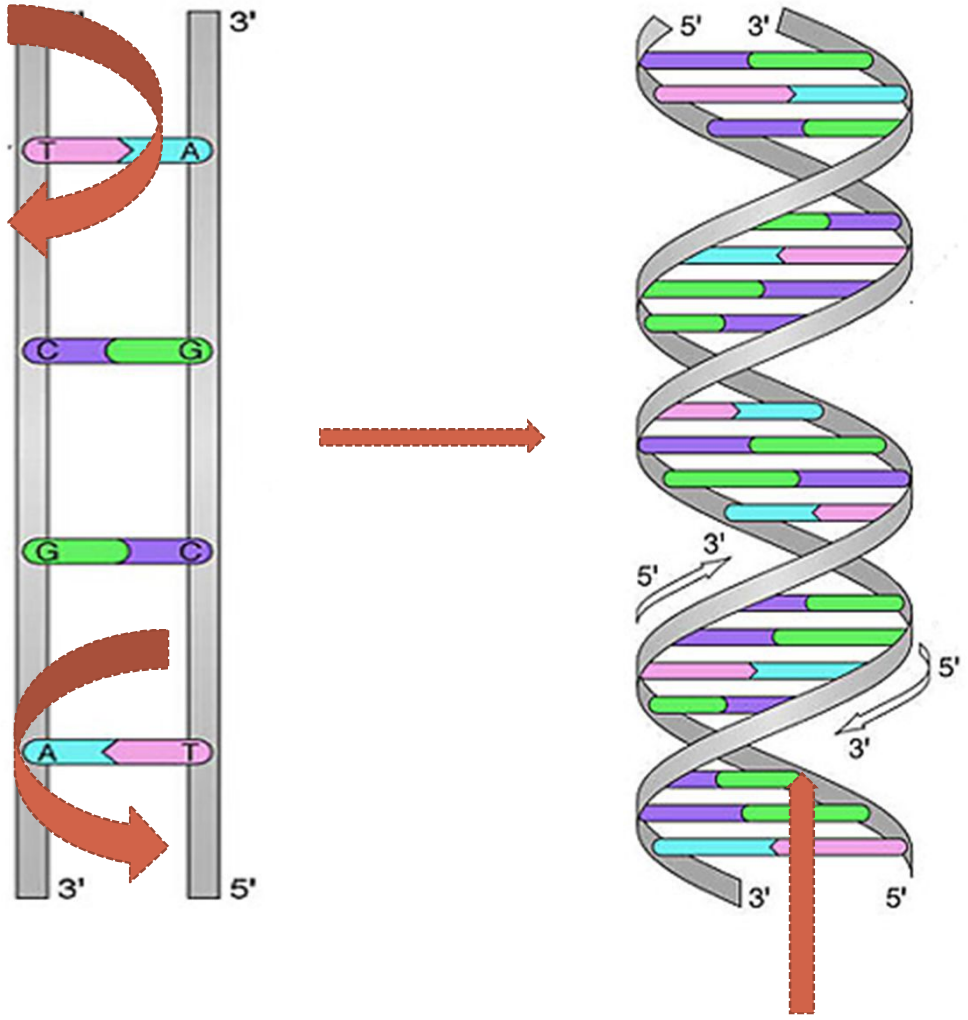
- James Watson and Francis Crick postulated a three dimensional model of DNA structure in 1953. The major features of this model are as follow:-

Structure of DNA

Work of Watson & Crick

1. DNA consists of two helical polynucleotide strands which are wound around the same axis to form a right handed double helix.

Structure of DNA



Structure of DNA

Work of Watson & Crick

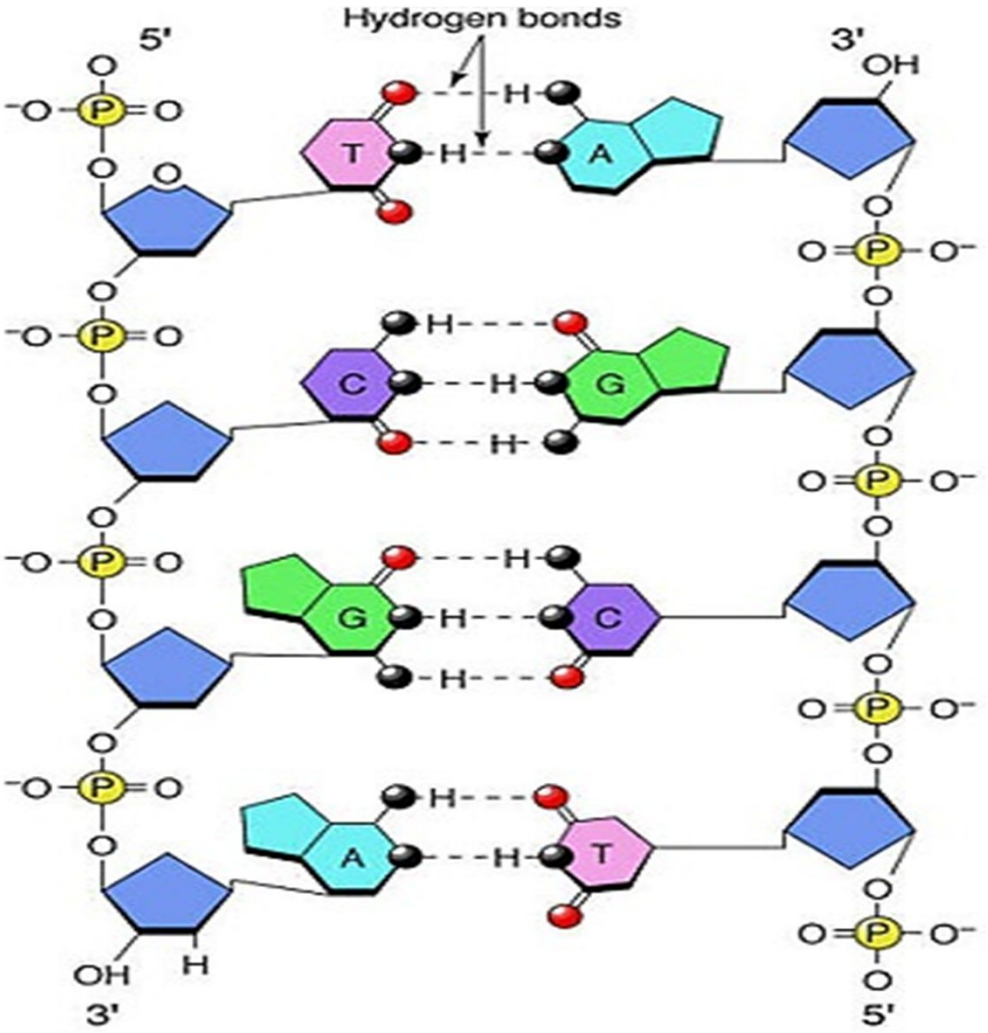
2. The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water.

Structure of DNA

Work of Watson & Crick

3. The nitrogenous bases of both strands are stacked inside the double helix lying perpendicular to the long axis of the helix.

Structure of DNA

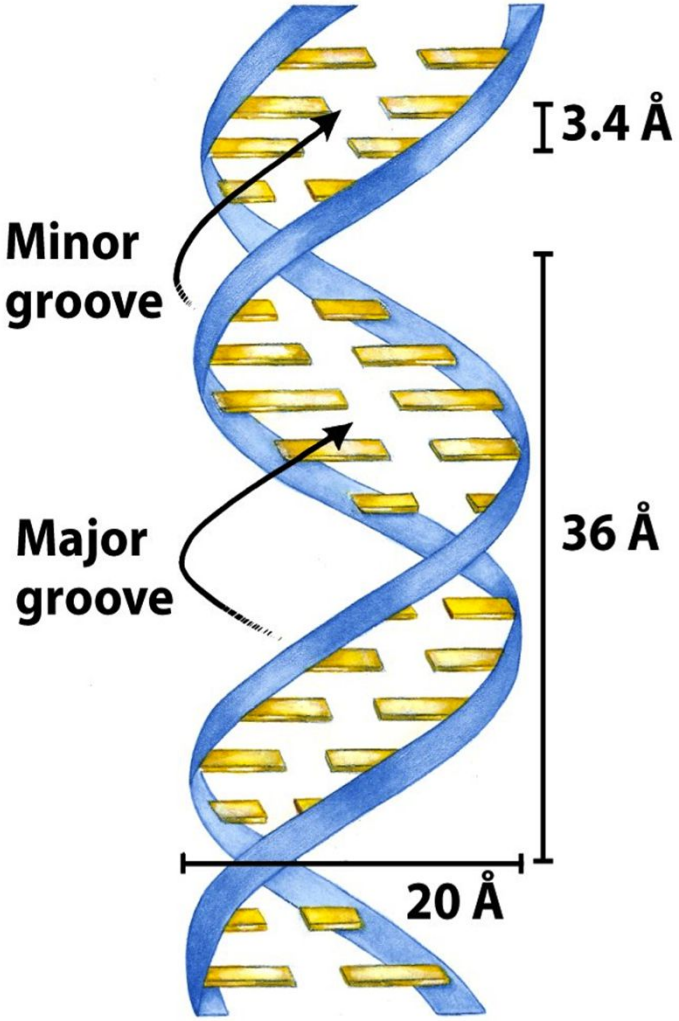


Structure of DNA

Work of Watson & Crick

4. The pairing and coiling of the two strands create a major groove and minor groove on the surface of the helix.

Structure of DNA

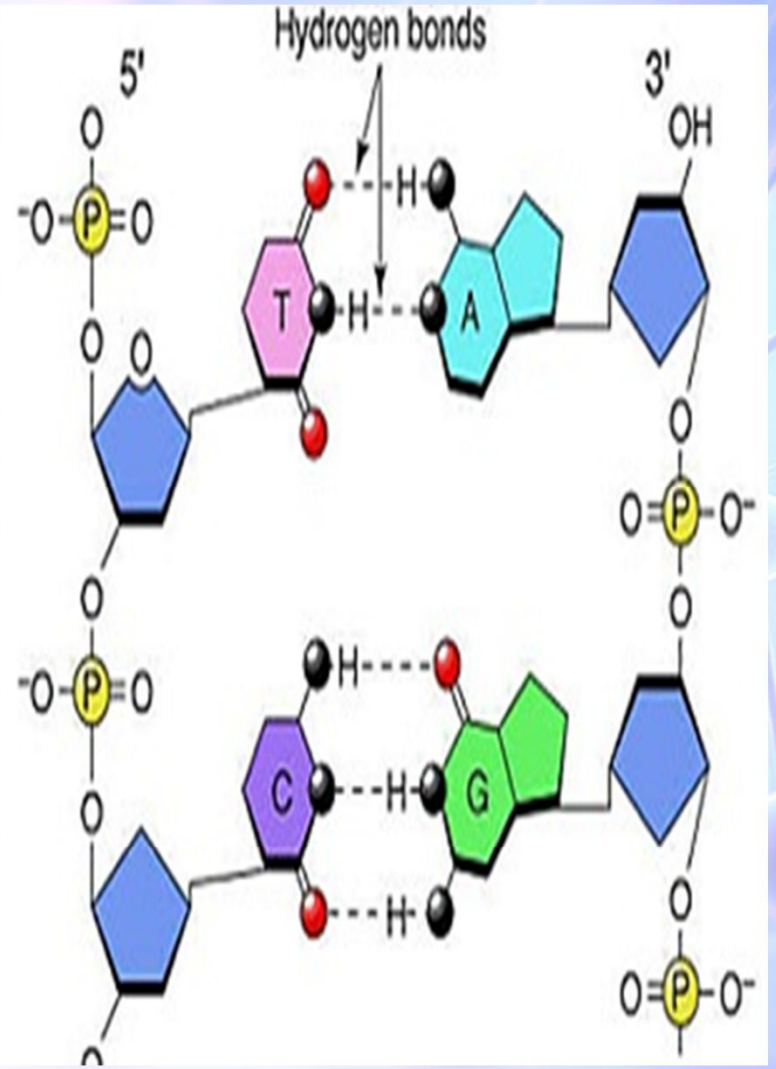


Structure of DNA

Work of Watson & Crick

5. Each nucleotide base of one strand is paired in the same plane with a base of the other strand.

Structure of DNA



Structure of DNA

Work of Watson & Crick

6. They also found that G pairs with C and A pairs with T due to the reason that they fit best within the structure.

Structure of DNA

Work of Watson & Crick

7. The two strands in DNA are present in antiparallel orientation i.e; their 5,3-phosphodiester bonds run in the opposite directions.

Structure of DNA

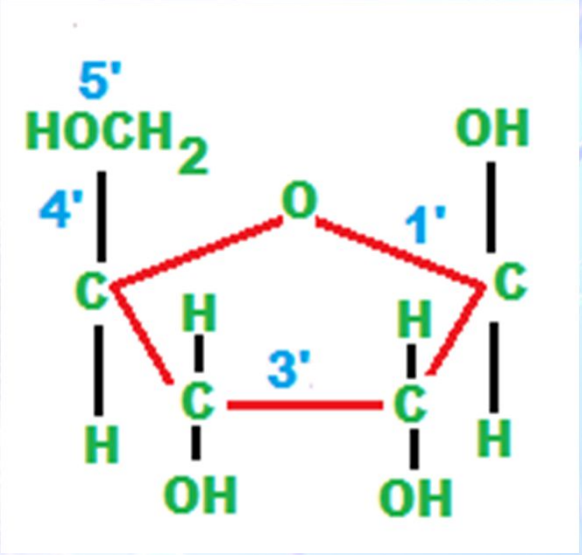
Work of Watson & Crick

8. The two antiparallel strands of double-helical DNA are complementary to each other.

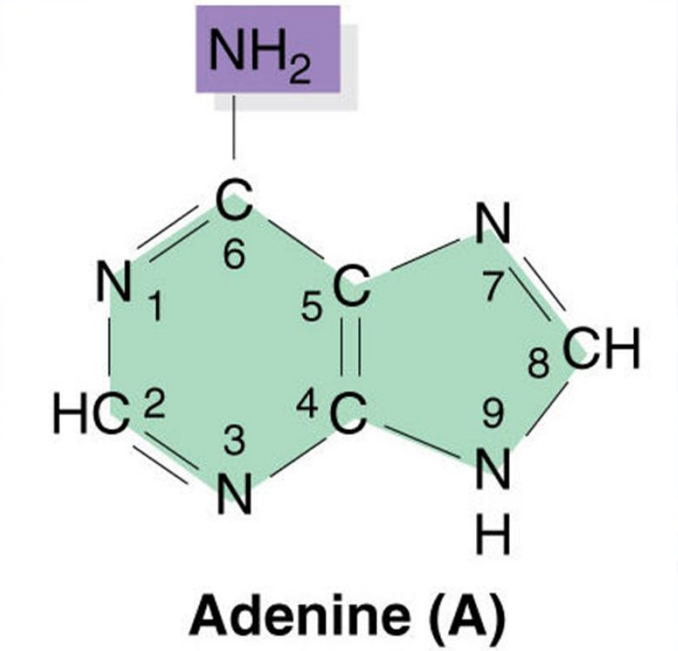
Chemical composition of RNA

- RNA (Ribonucleic acid) is a polymer of ribonucleotides.
- Each ribonucleotide is composed of three components:
  - A ribose sugar
  - A Nitrogenous Base
  - A Phosphoric acid

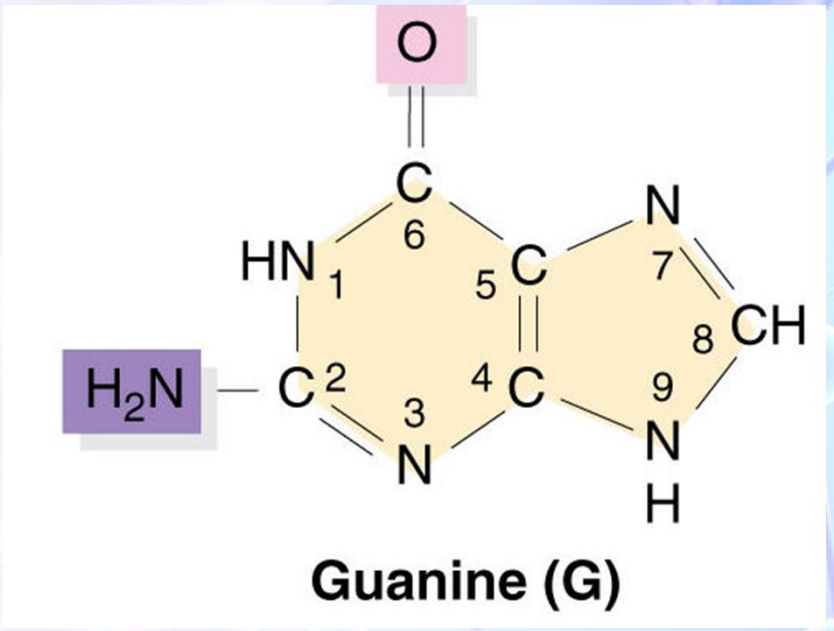
Ribose (a pentose sugar)



Nitrogenous Bases



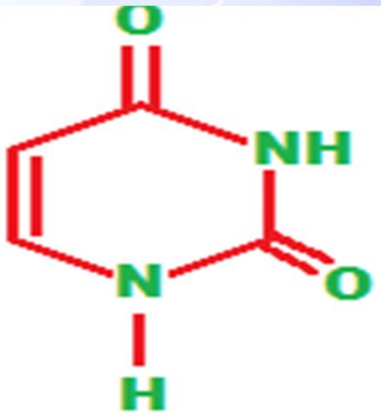
Adenine (A)



Guanine (G)

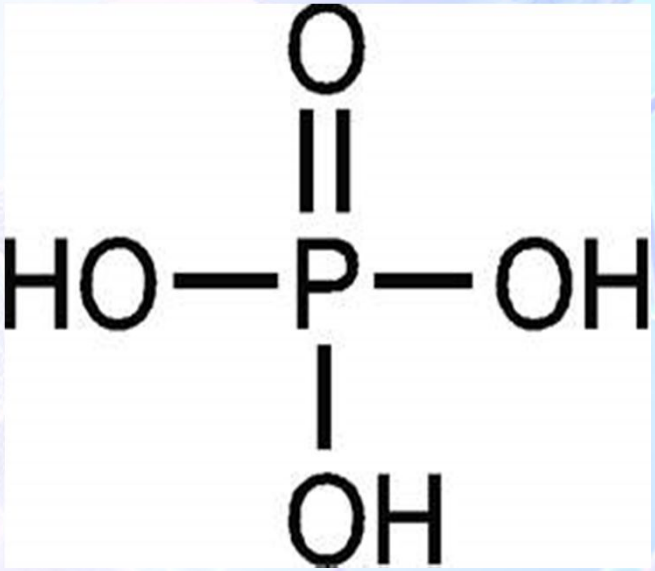
Nitrogenous Bases

Cytosine (C)

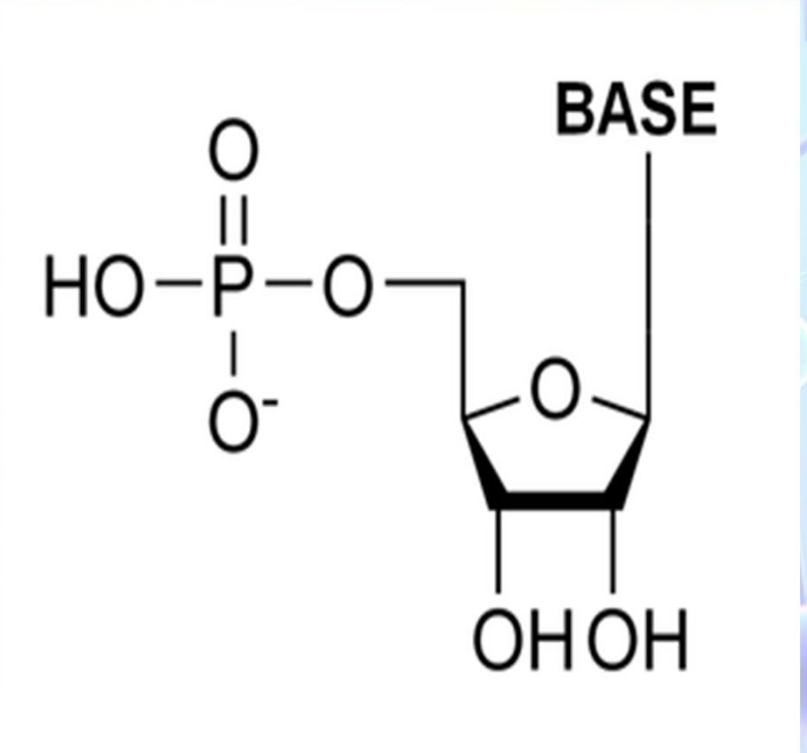


Uracil (U)

Phosphoric acid



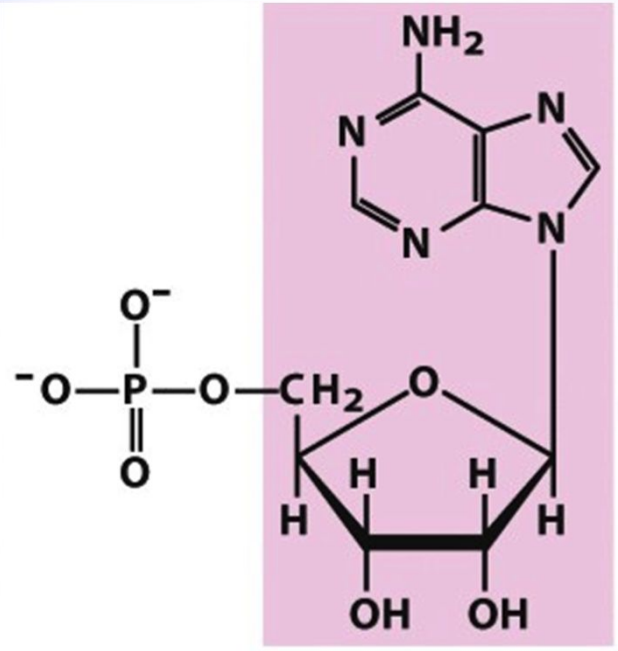
A Ribonucleotide



Types of Ribonucleotides

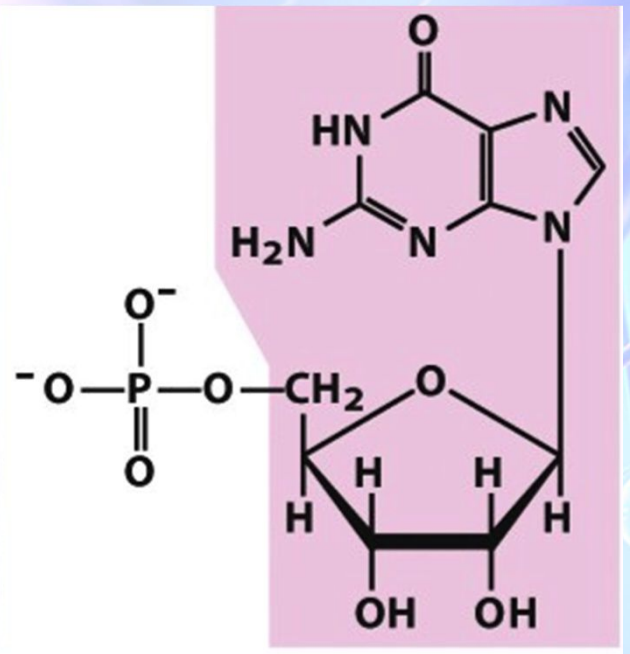
- There are mainly four types of ribonucleotides depending upon the types of nitrogenous bases present in RNA.

Types of Ribonucleotides



Adenylate (adenosine 5′-monophosphate)

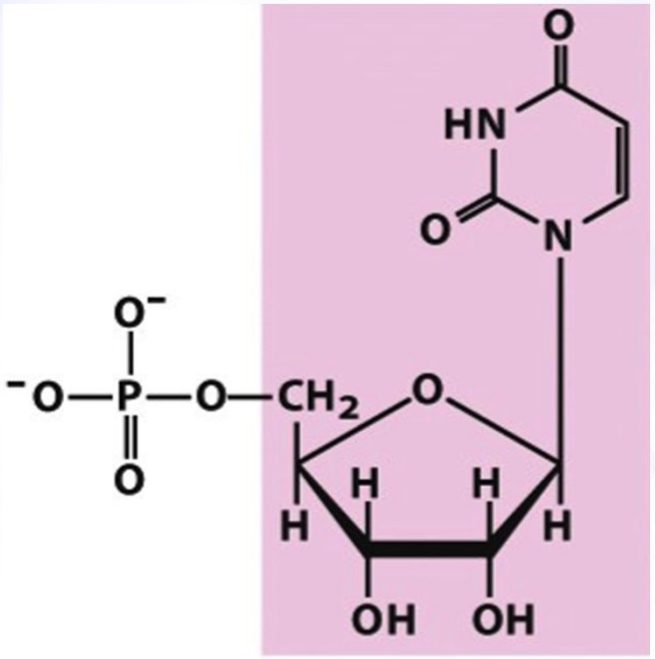
Adenosine



Guanylate (guanosine 5′-monophosphate)

Guanosine

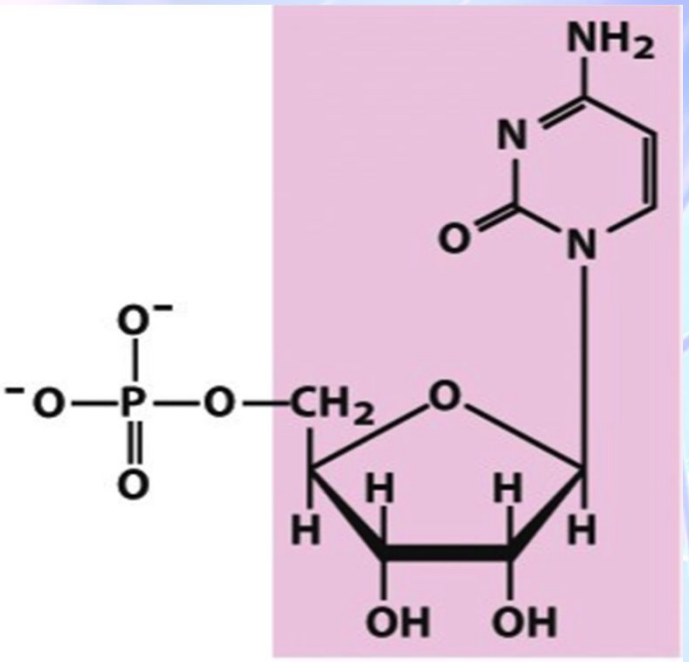
Types of Ribonucleotides



Uridylate (uridine 5′-monophosphate)

U, UMP

Uridine



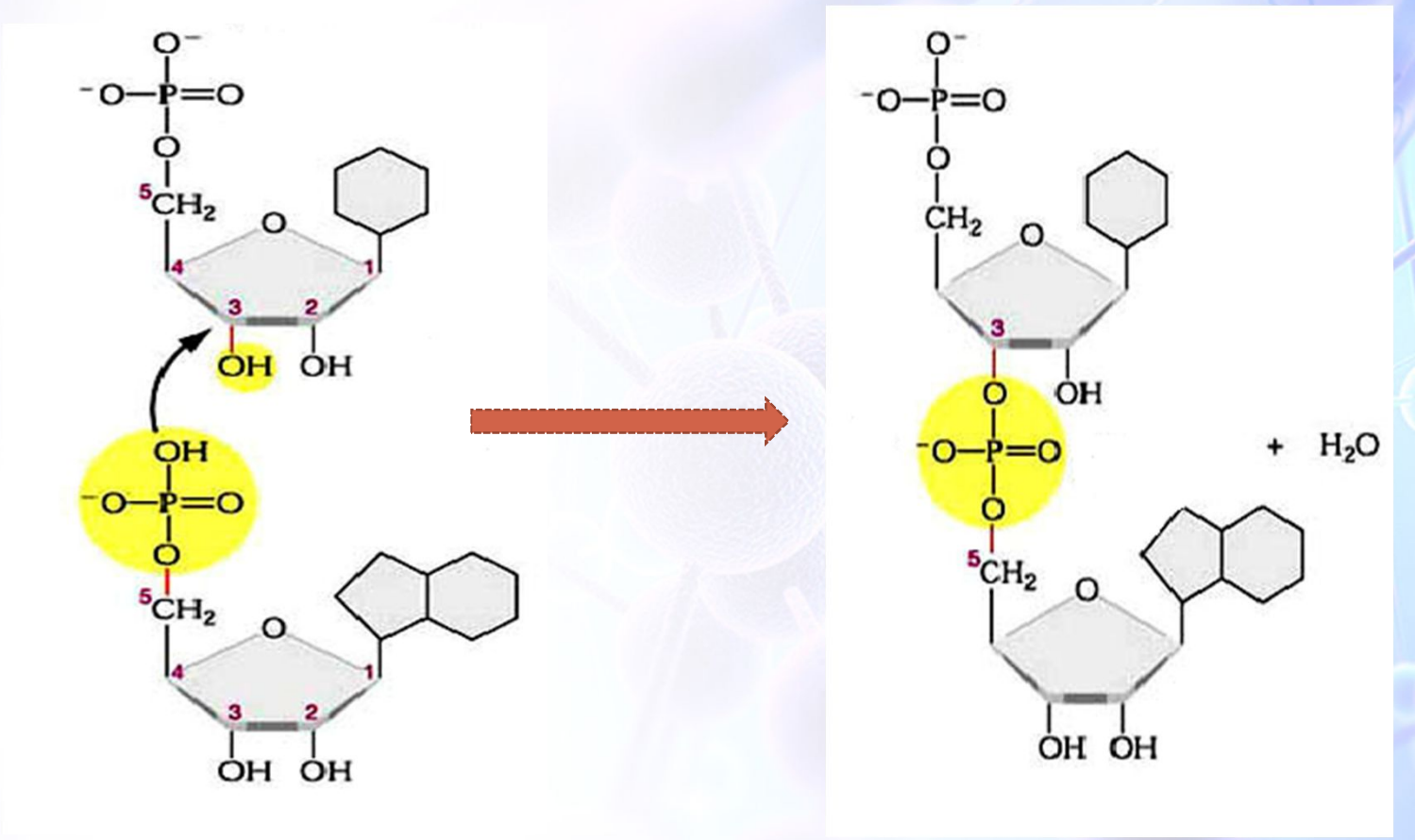
Cytidylate (cytidine 5′-monophosphate)

C, CMP

Cytidine

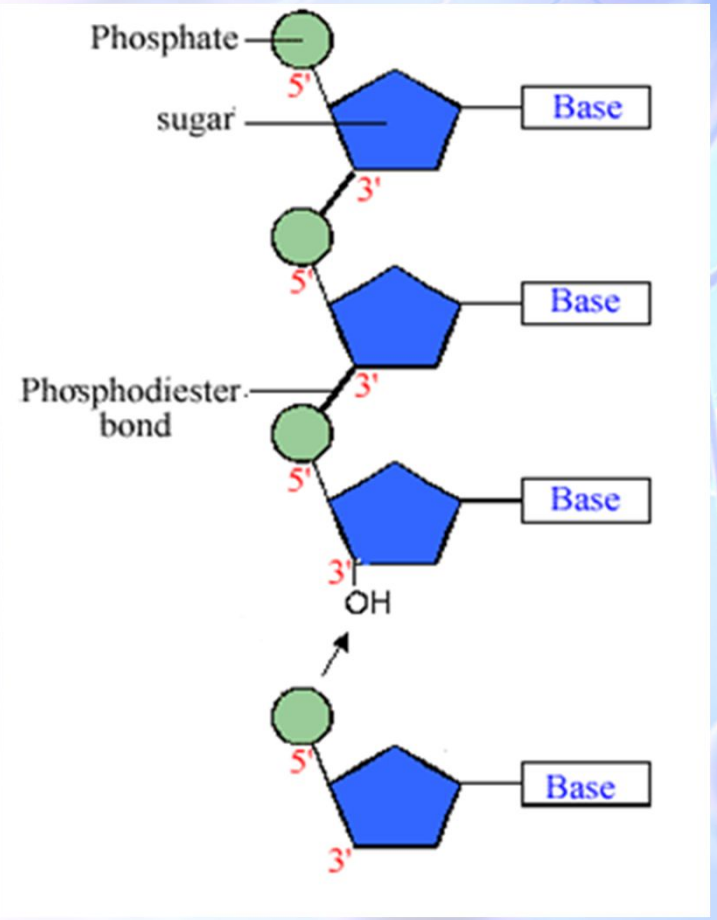
Types of Ribonucleotides

How do Ribonucleotides Join?



Types of Ribonucleotides

A Poly-Ribonucleotide



Types of RNAs

- There are mainly three types of Ribonucleic acids (RNAs) present in the cells of living organisms.
  - Messenger RNA (mRNA)
  - Transfer RNA (tRNA)
  - Ribosomal RNA (rRNA)

Types of RNAs

Messenger RNA (mRNA)

- It is the type of RNA that carries genetic information from DNA to the protein biosynthetic machinery of the ribosome.
- It provides the templates that specify amino acid sequences in polypeptide chains.
- The process of forming mRNA on a DNA template is known as transcription.

Types of RNAs

Messenger RNA (mRNA)

- It may be monocistronic or polycistronic.
- The length of mRNA molecules is variable and it depends on the length of gene.

Types of RNAs

Transfer RNA (tRNA)

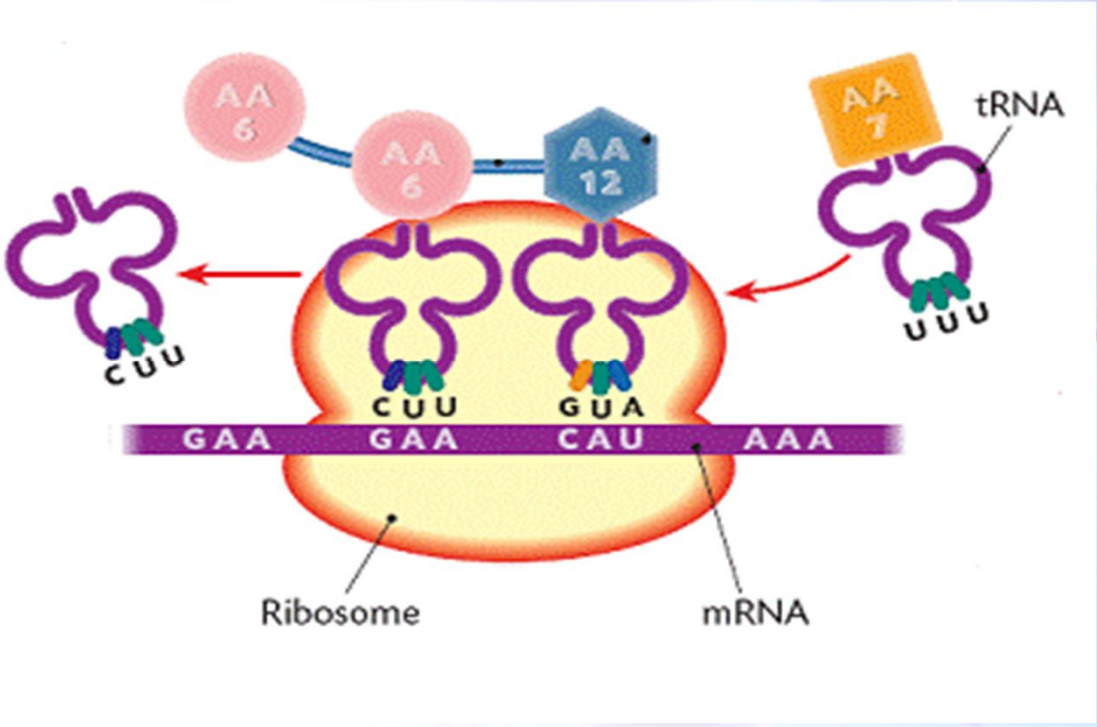
- Transfer RNAs serve as adapter molecules in the process of protein synthesis.
- They are covalently linked to an amino acid at one end.

Types of RNAs

Transfer RNA (tRNA)

- They pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence.

Types of RNAs



Types of RNAs

Ribosomal RNA (rRNA)

- Ribosomal RNAs are components of ribosomes.
- rRNA is a predominant material in the ribosomes constituting about 60% of its weight.
- It has a number of functions to perform in the ribosomes.

Structures of RNAs

- mRNA is always single stranded when it is formed from DNA.
- But this single strand assumes a double helical conformation soon after its formation.
- This confirmation is achieved mainly due to base stacking interactions.

Structures of RNAs

Messenger RNA (mRNA)

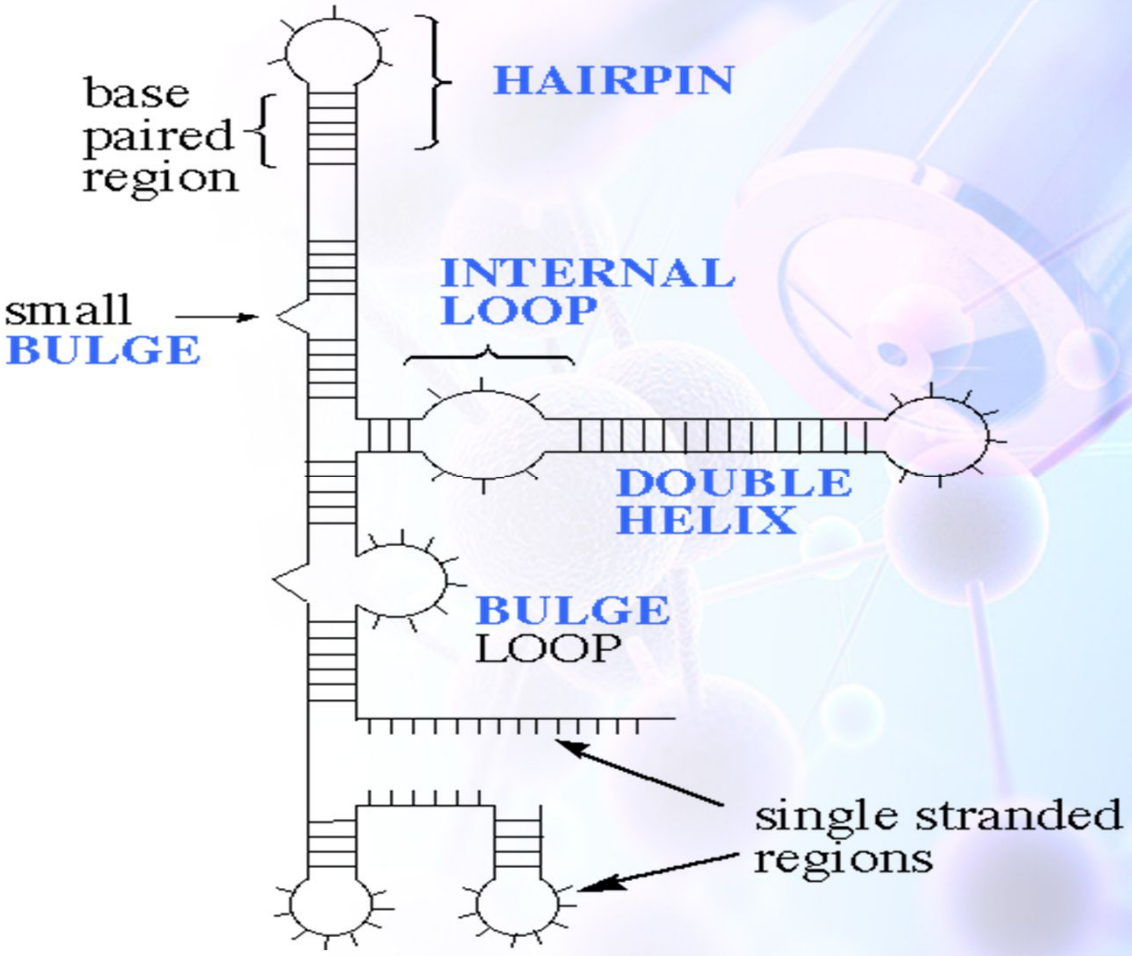
Structures of RNAs

- Self-complementary sequences may occur in the RNA molecules which produce more complex structures.
- So RNA can base-pair with complementary regions of either RNA or DNA.
- RNA has no any regular secondary structure that serves as a reference point. The three-dimensional structures of many RNAs are complex and unique.

Structures of RNAs

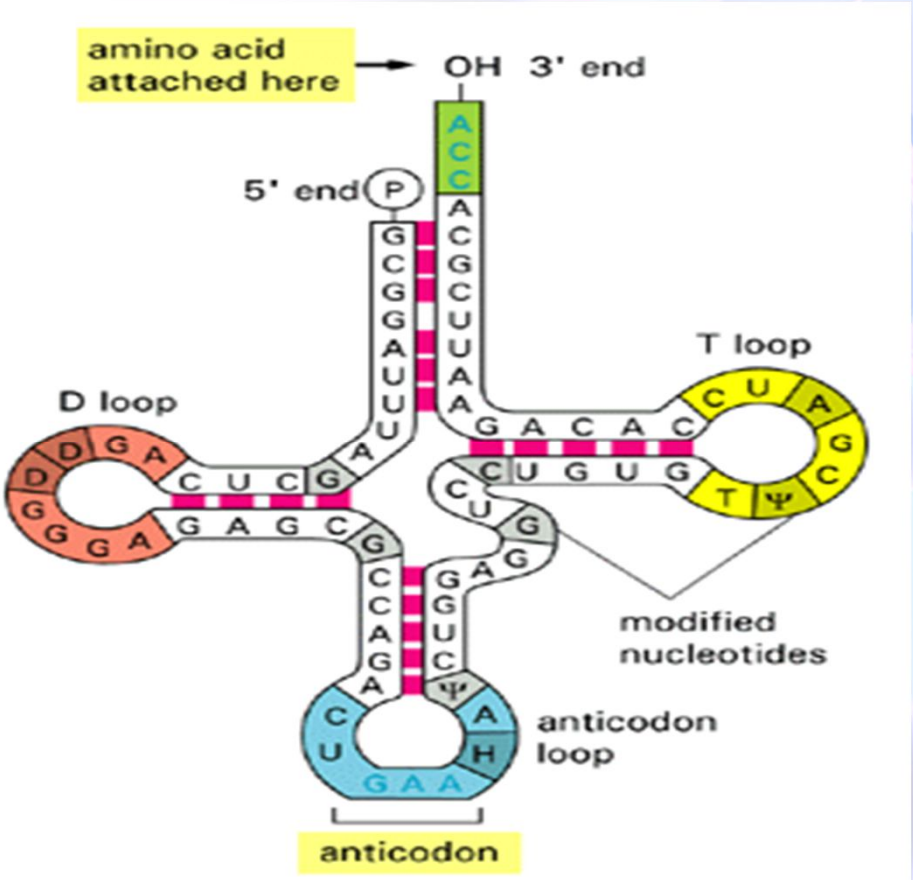
- Breaks in the helix caused by mismatched or unmatched bases in one or both strands are common and result in bulges or internal loops.
- Hairpin loops form between nearby self-complementary sequences.

Structures of RNAs



Structures of RNAs

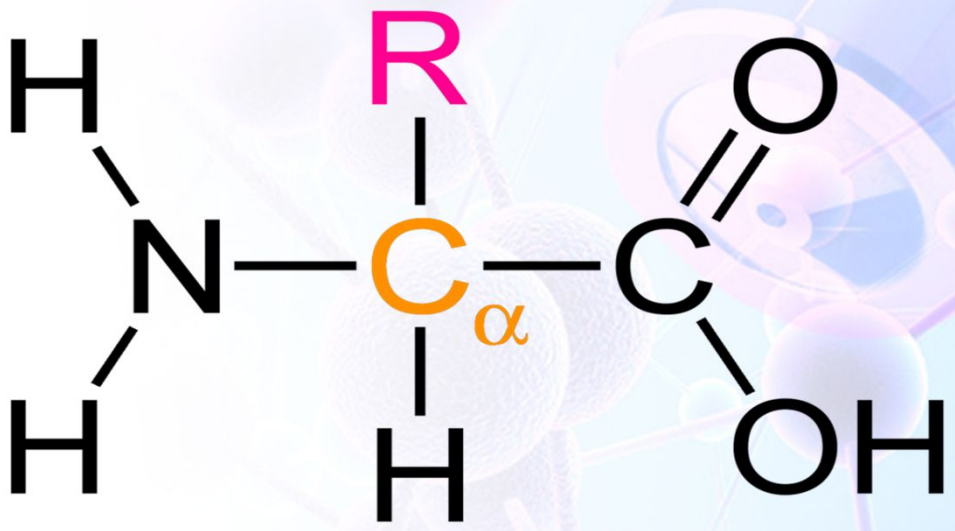
Transfer RNA (tRNA)



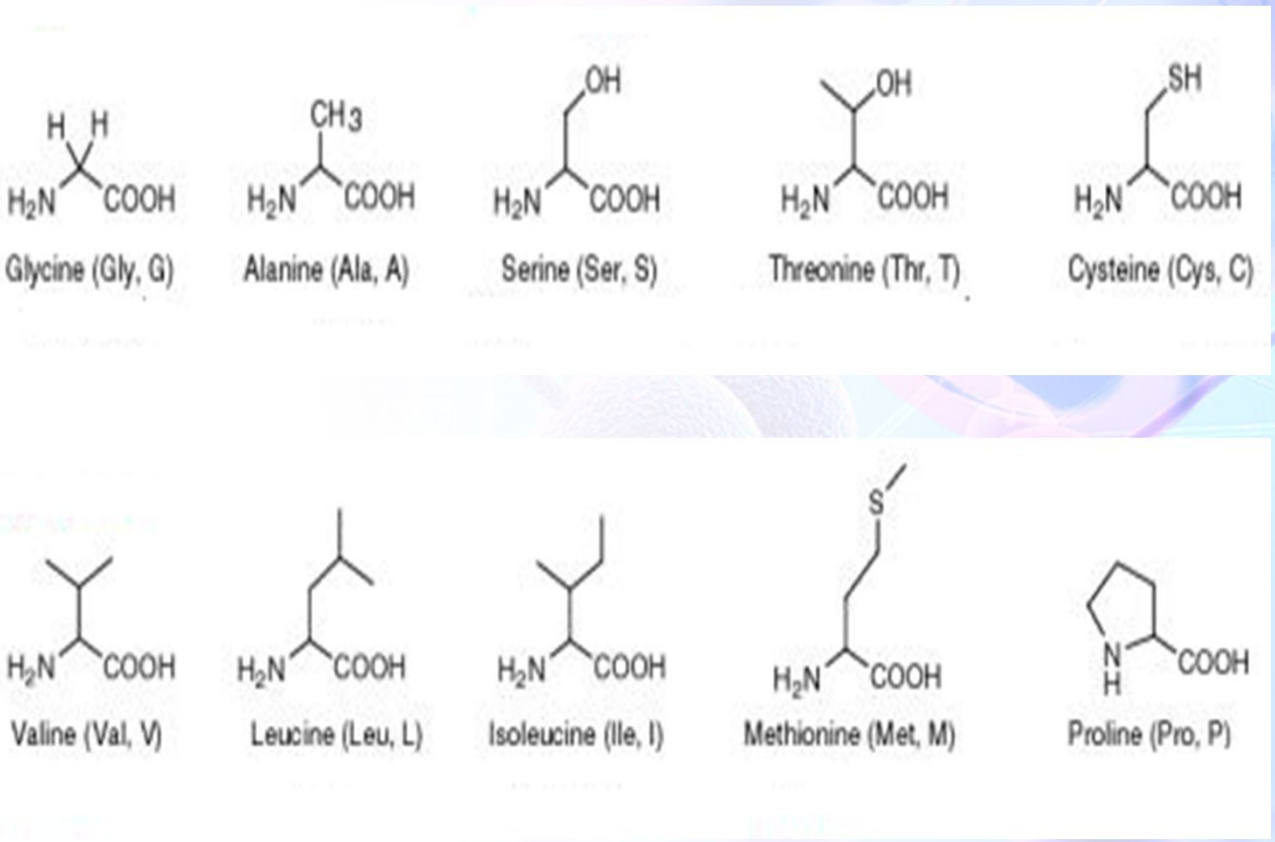
Chemical composition of proteins

- Proteins are polymers of amino acids.
- They range in size from small to very large.
- All the proteins are made up of Twenty different types of amino acids. So these amino acids are called standard amino acids.

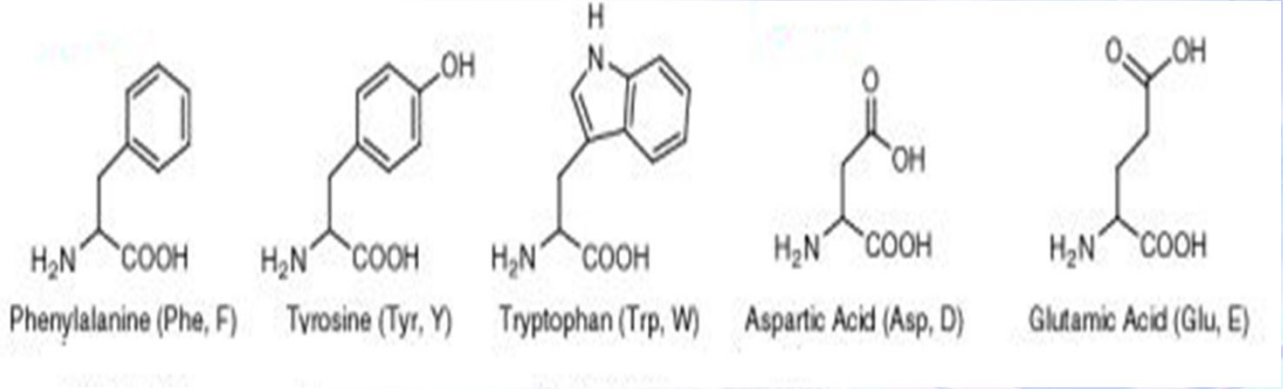
Chemical composition of proteins

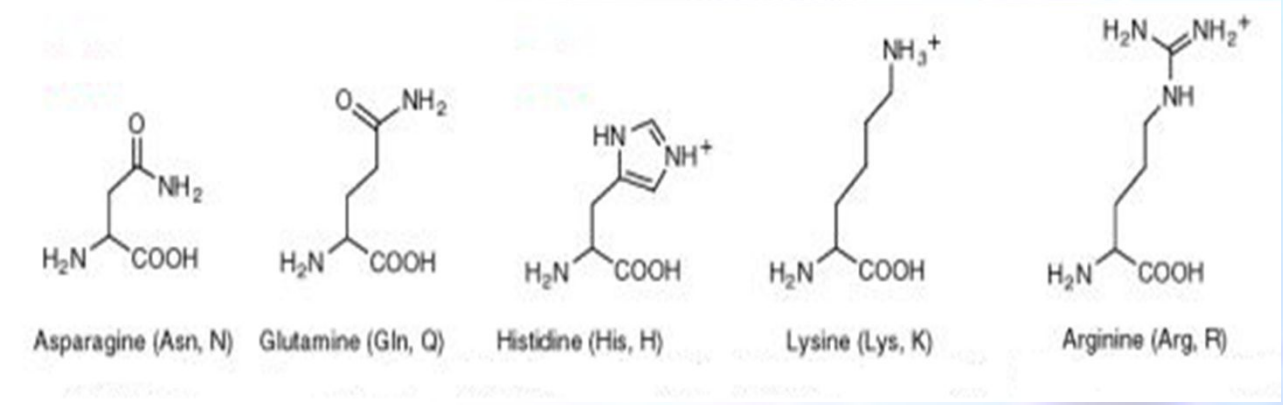


Chemical composition of proteins



Chemical composition of proteins

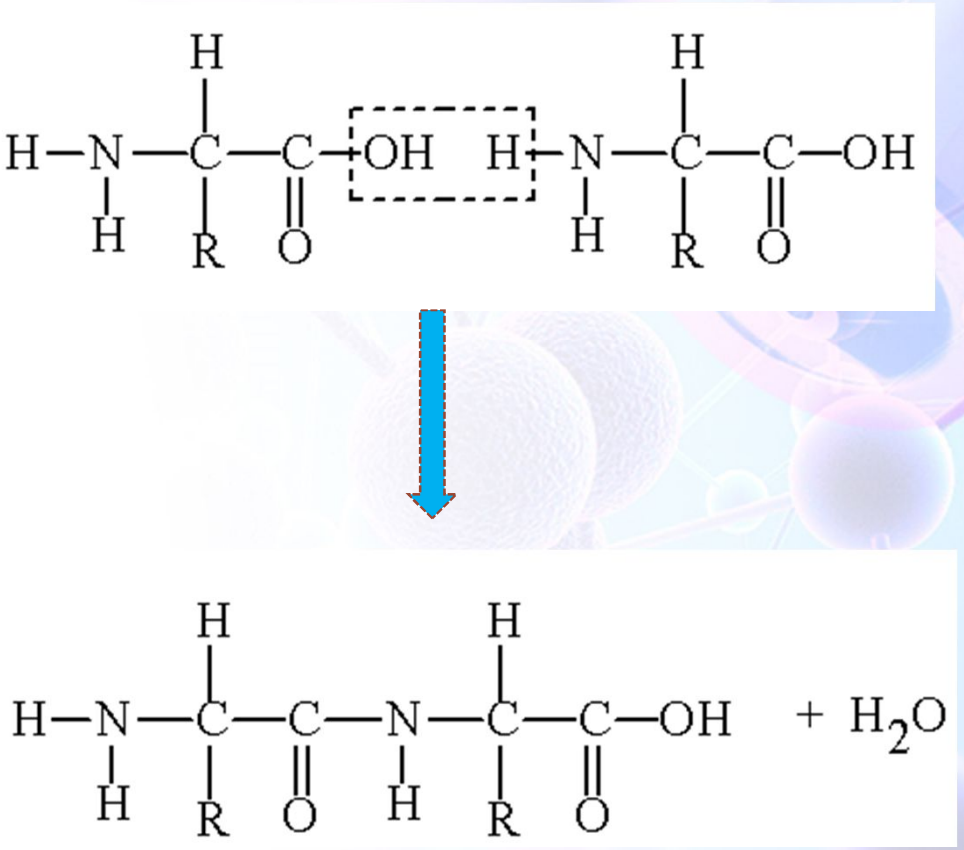




Chemical composition of proteins

- In a protein molecule, each amino acid residue is joined to its neighbour by a specific type of covalent bond which is called Peptide Bond.

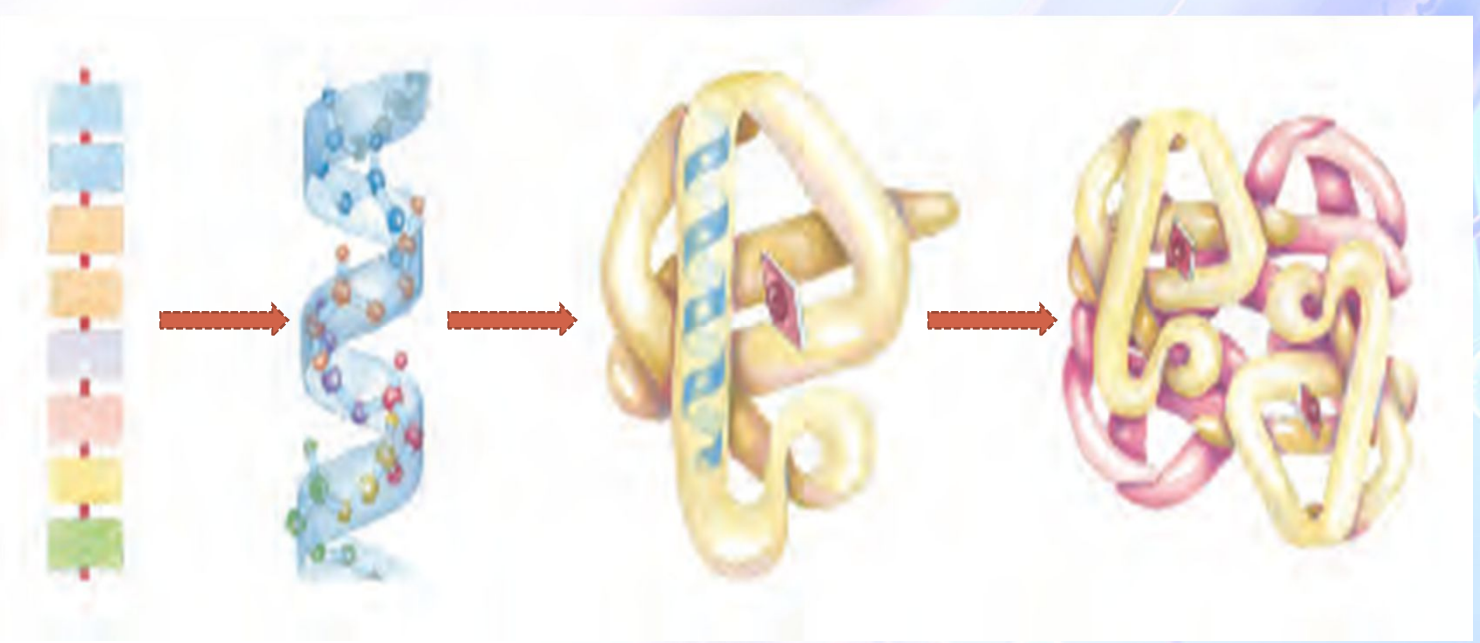
Chemical composition of proteins



Chemical composition of proteins

- Amino acids can successively join to form dipeptides, tripeptides, tetrapeptides, oligo peptides and polypeptides.

Chemical composition of proteins



Primary structure of proteins

- Primary structure or covalent structure of protein refers to the amino acid sequence of its polypeptide chain.
- Each type of protein has a unique amino acid sequence.

Primary structure of proteins

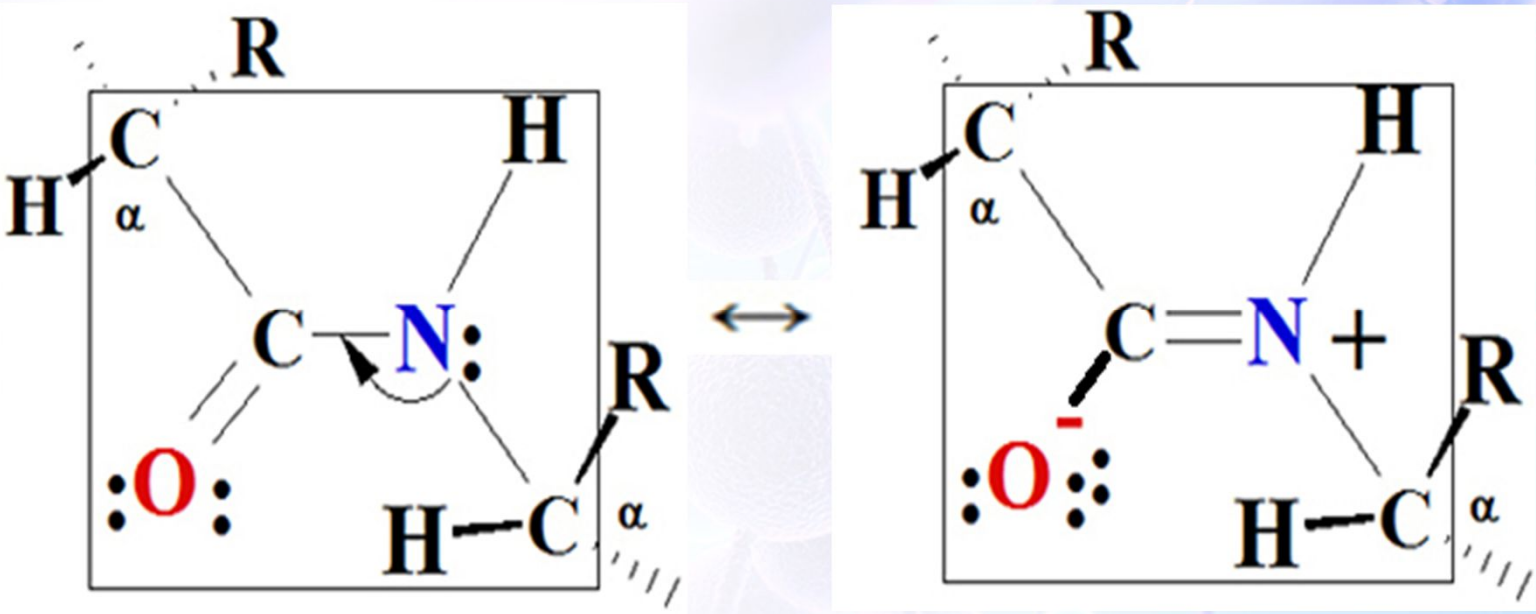
Peptide Bond Is Rigid and Planar

- Linus Pauling and Robert Corey carefully analyzed the peptide bond.
- Their findings laid the foundation for our present understanding of protein structure.

Primary structure of proteins

- They demonstrated that the peptide C - N bond is somewhat shorter than the C - N bond in a simple amine.

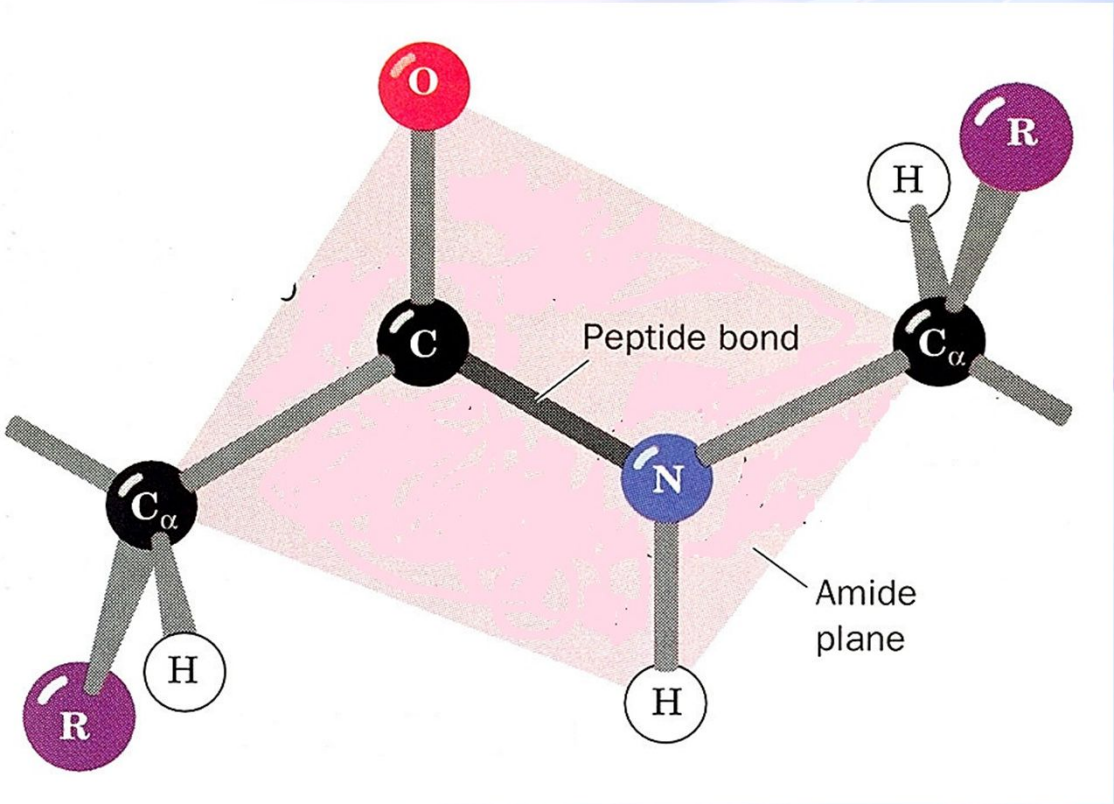
Primary structure of proteins



Primary structure of proteins

- The six atoms of the peptide group are co-planar i.e., lie in a single plane, with the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen trans to each other.

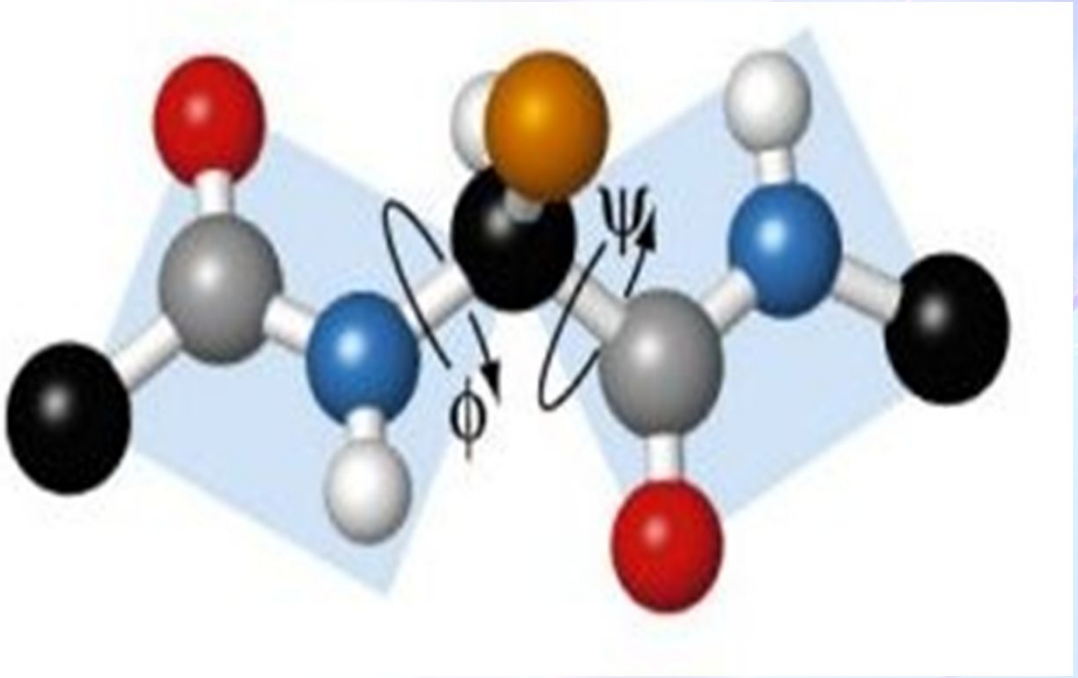
Primary structure of proteins



Primary structure of proteins

- Pauling and Corey concluded that the peptide C - N bonds are unable to rotate freely because of their partial double-bond character.
- Rotation is permitted about the N - αC and the αC - C bonds.

Primary structure of proteins



Primary structure of proteins

- The bond angles resulting from rotations at C are labelled ϕ (phi) for the N - αC bond and ψ(psi) for the αC - C bond.
- In principle, ϕ and ψ can have any value between +180 & -180.

Secondary structure of proteins

- Secondary structure of proteins refers to the local conformation of some part of a polypeptide.

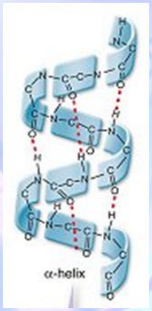
Secondary structure of proteins

Secondary structure of proteins

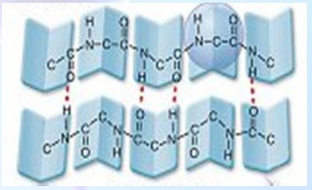
- A few types of secondary structures are particularly stable and occur widely in proteins.

Secondary structure of proteins

- The most prominent are:-
  - α-helix



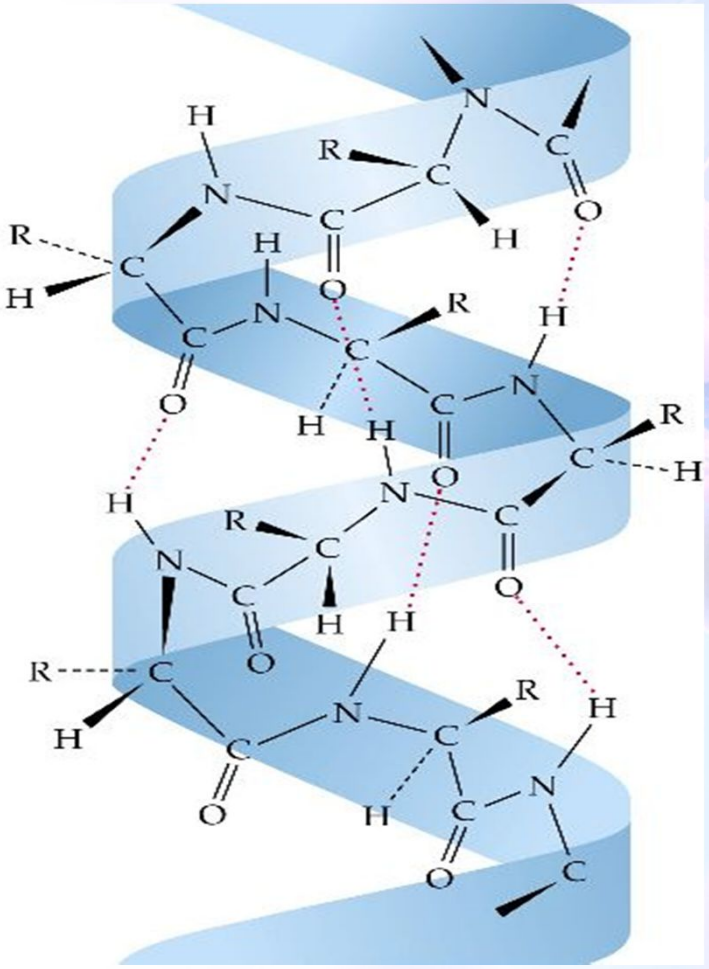
  - β- conformations.



α- Helix

- The simplest arrangement which a polypeptide chain could assume with its rigid peptide bonds is a helical structure, which Pauling and Corey called the **α-helix**.

α- Helix



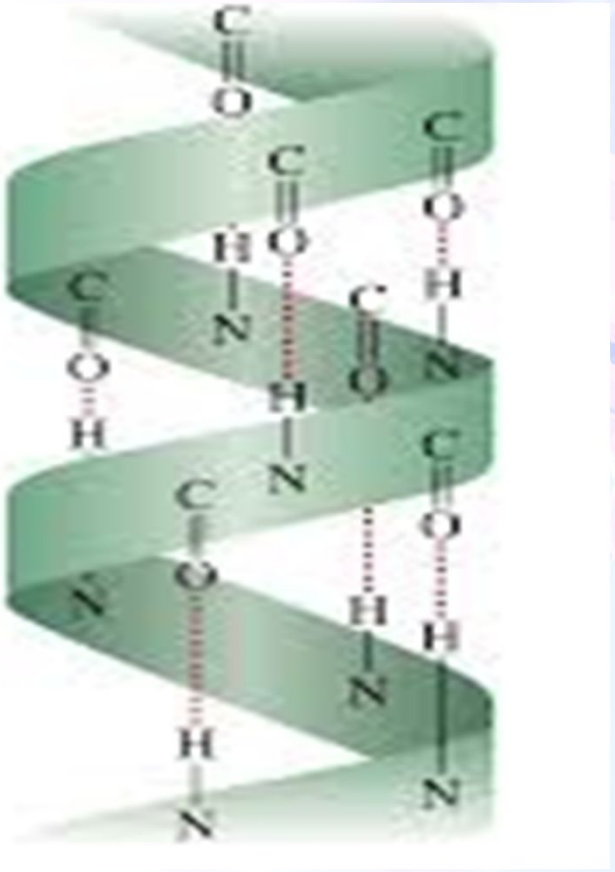
α- Helix

- The helical twist of the *α*-helix found in all proteins is right-handed.
- The repeating unit is a single turn of the helix, which extends about 5.4 Å (includes 3.6 amino acid residues) along the long axis.

α- Helix

- The amino acid residues in an helix have conformations with psi = –45 to –50 and phi = – 60.
- An helix makes optimal use of internal hydrogen bonds.

α- Helix



α- Helix

- About one-fourth of all amino acid residues in polypeptides are found in α-helices while in some proteins it is the predominant structure.

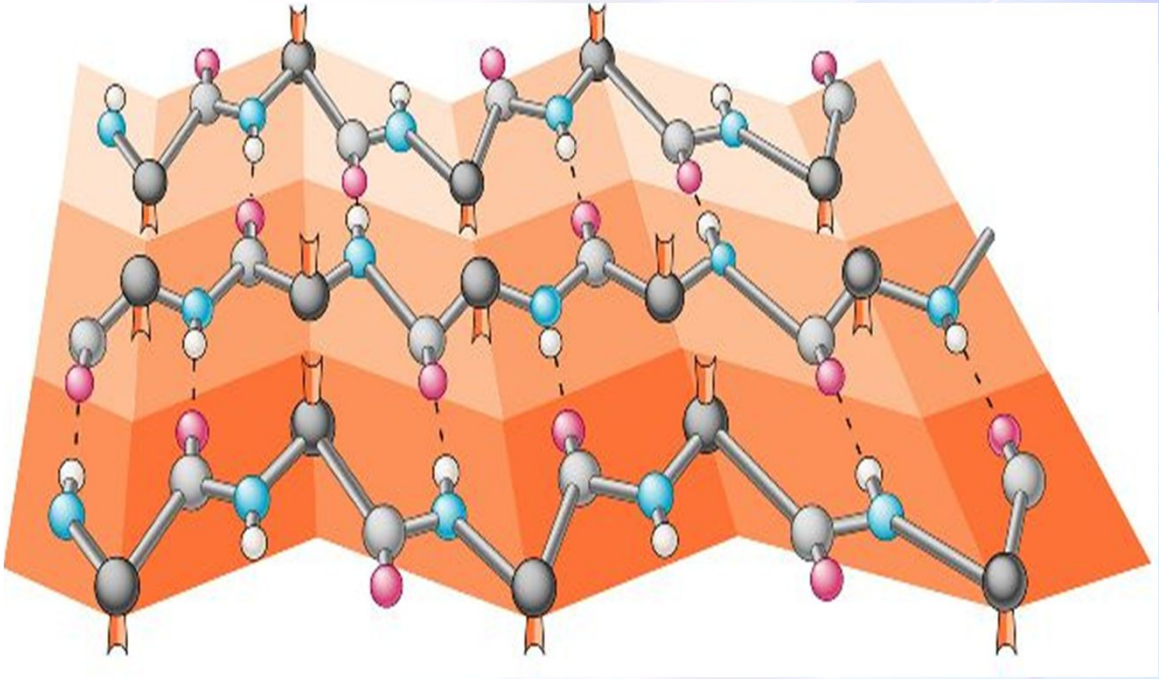
β- Pleated Sheets

- Pauling and Corey predicted a second type of secondary structure which they called **β-sheets.**
- This is a more extended conformation of polypeptide chains.

β- Pleated Sheets

- The backbone of the polypeptide chain is extended into a zigzag structure.
- The zigzag polypeptide chains are arranged side by side to form a structure resembling a series of pleats.

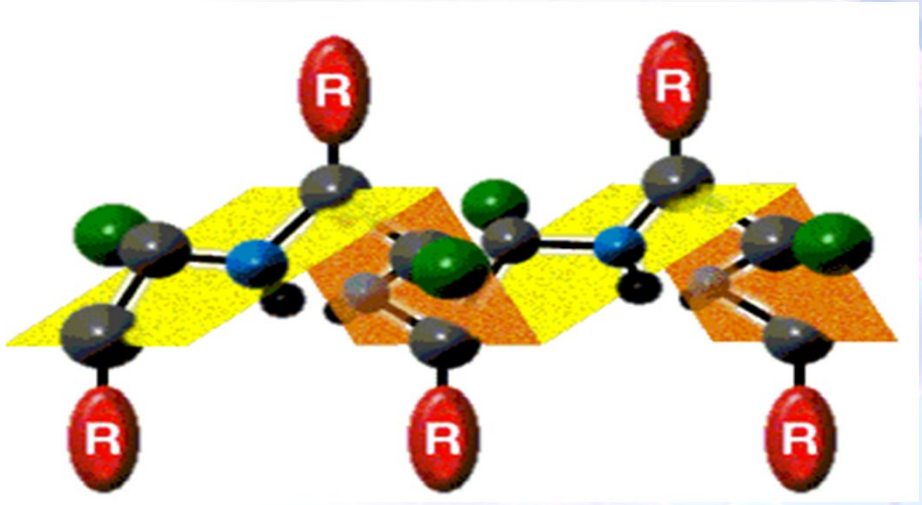
β- Pleated Sheets



β- Pleated Sheets

- The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions.

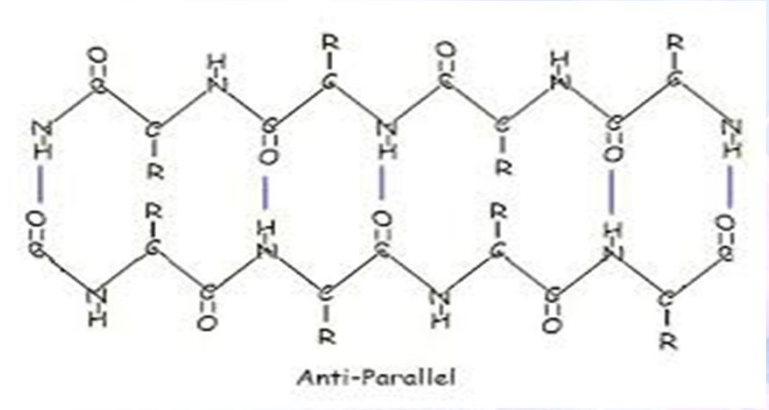
β- Pleated Sheets



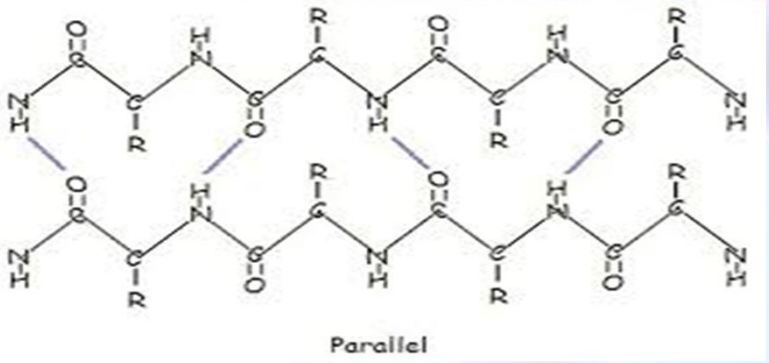
β- Pleated Sheets

- Hydrogen bonds are formed between adjacent segments of polypeptide chain.
- The adjacent polypeptide chains in a sheet can be either parallel or antiparallel.

β- Pleated Sheets



Anti-Parallel

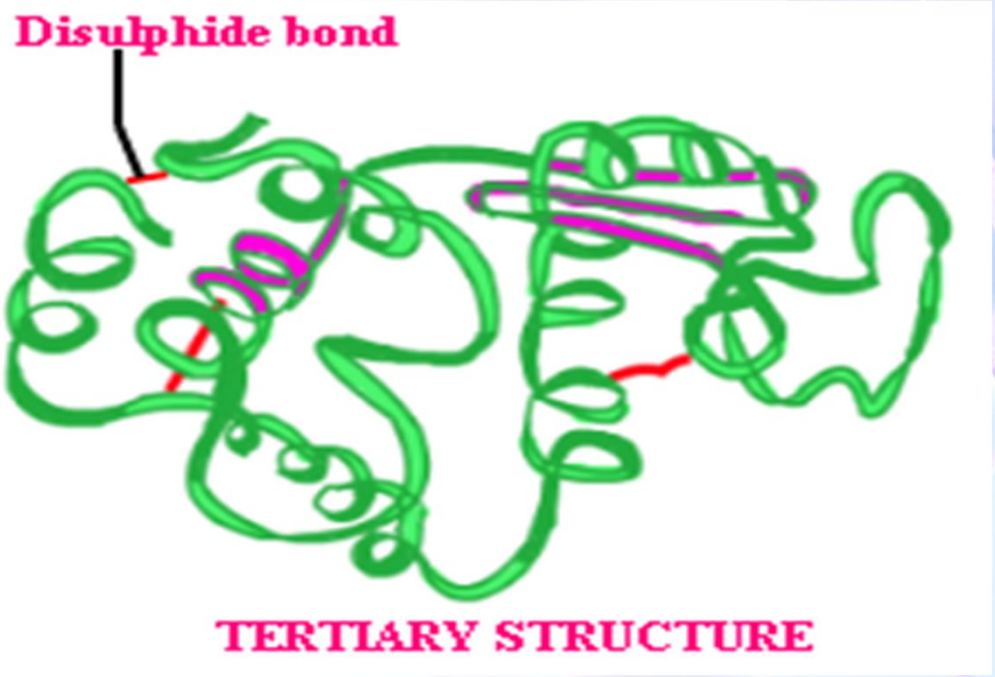


Parallel

Tertiary Structure of Proteins

- The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure.**

Tertiary Structure of Proteins



Tertiary Structure of Proteins

- It includes longer-range aspects of amino acid sequence.
- Amino acids that are far apart in the polypeptide chain may interact within the completely folded structure of a protein.

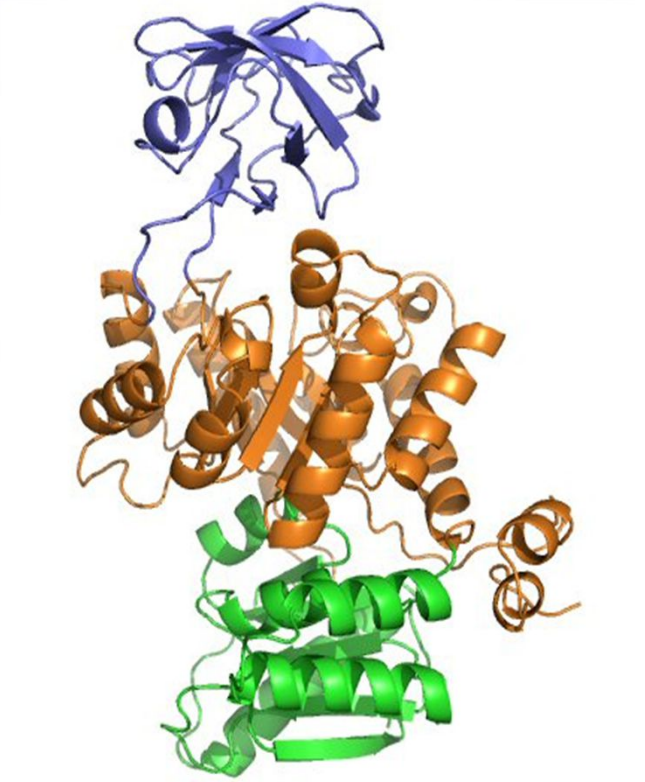
Tertiary Structure of Proteins

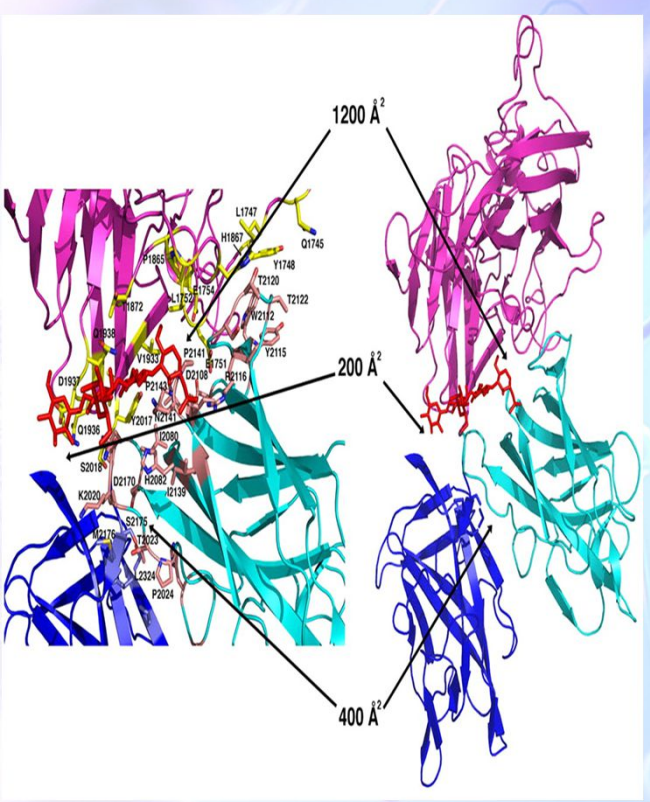
- Interacting segments of polypeptide chains are held in their characteristic tertiary positions by different kinds of weak interactions (and sometimes by covalent bonds) between the segments.

Tertiary Structure of Proteins

- Large polypeptide chains usually fold into two or more globular clusters known as **domains,** which often give these proteins a bi- or multilobal appearance.

Tertiary Structure of Proteins





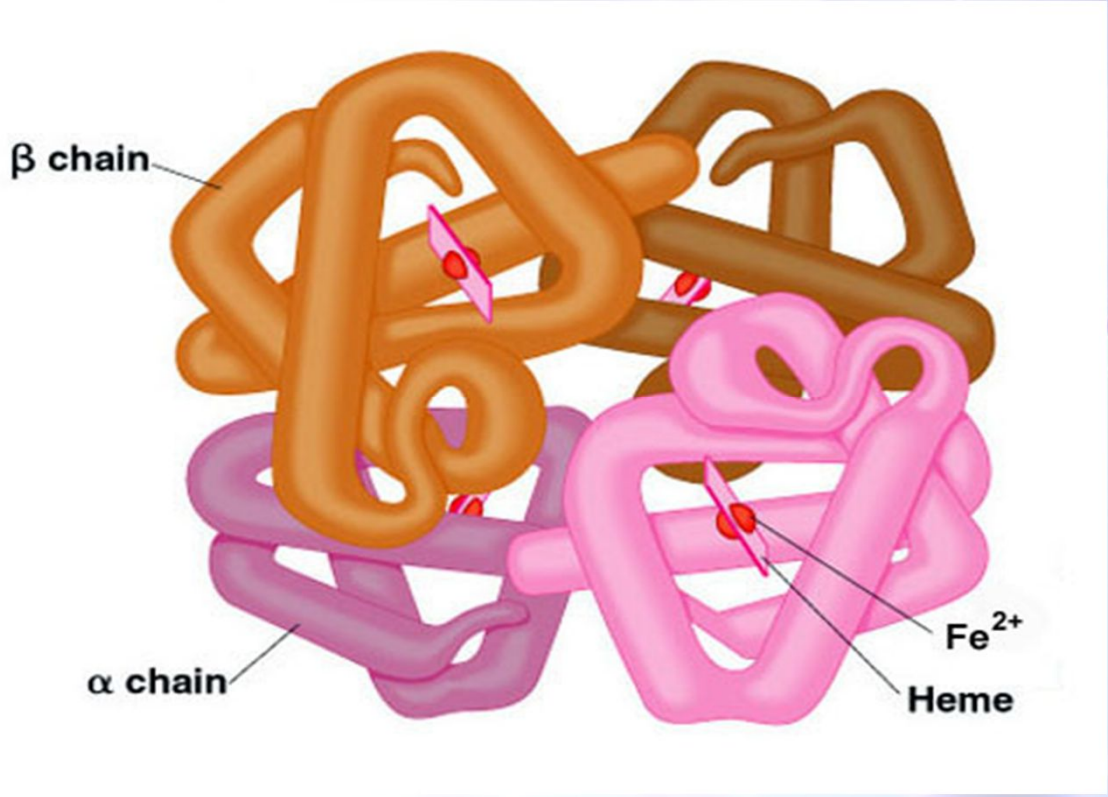
Quaternary Structure of Proteins

- Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different.
- The spatial arrangement of these subunits is known as a protein's quaternary structure.

Quaternary Structure of Proteins

- A multi-subunit protein is also referred to as a multimer.
- A multimer with just a few subunits is called as oligomer and a single subunit or a group of subunits, is called a protomer.

Quaternary Structure of Proteins



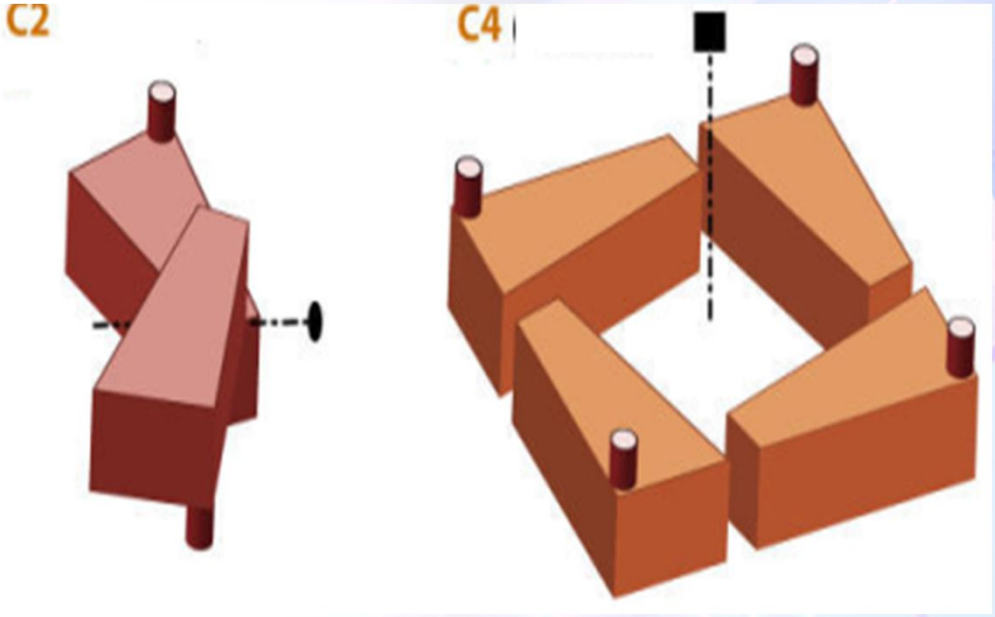
Quaternary Structure of Proteins

- Identical subunits of multimeric proteins are generally arranged in a symmetric patterns.
- Oligomers can have either rotational symmetry or helical symmetry.

Quaternary Structure of Proteins

- There are several forms of rotational symmetry. The simplest is cyclic symmetry, involving rotation about a single axis.

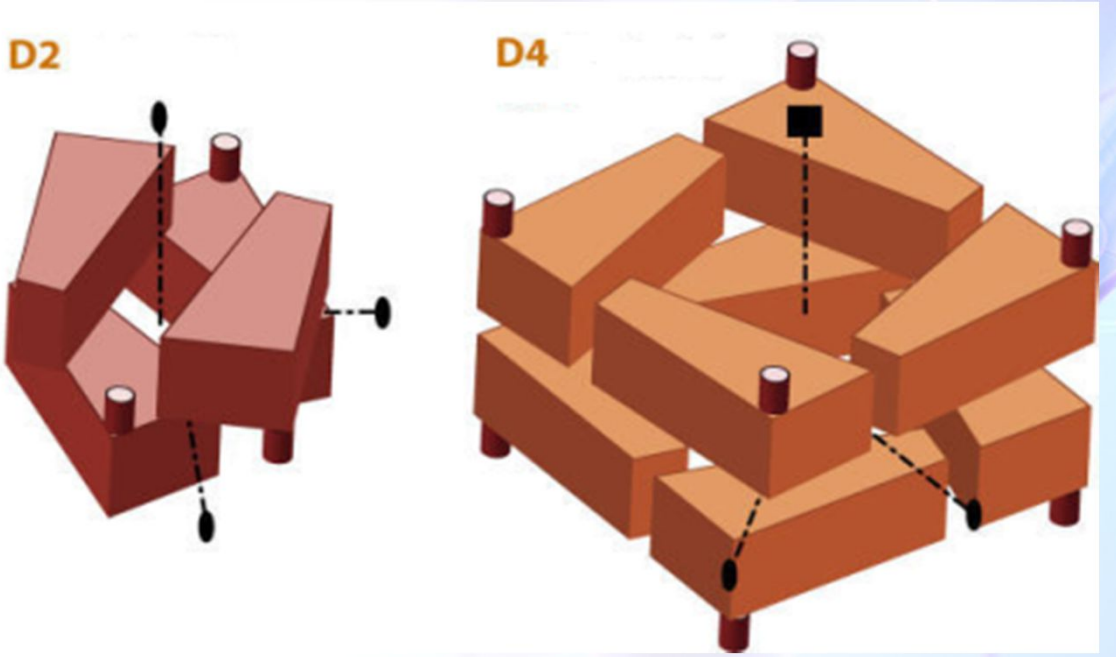
Quaternary Structure of Proteins



Quaternary Structure of Proteins

- A somewhat more complicated rotational symmetry is dihedral symmetry, in which a twofold rotational axis is present.

Quaternary Structure of Proteins



Quaternary Structure of Proteins

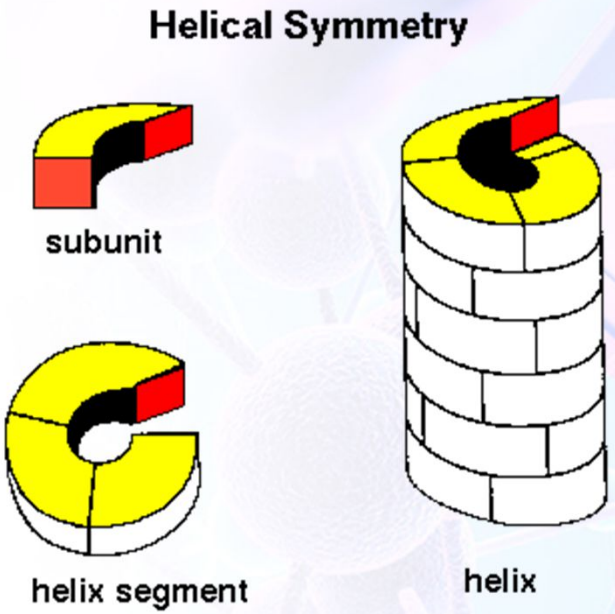
- More complex rotational symmetries include icosahedral symmetry.
- An icosahedron is a regular 12-cornered polyhedron having 20 triangular faces.

Quaternary Structure of Proteins

Quaternary Structure of Proteins

- The other major type of symmetry found in oligomers is helical symmetry.

Quaternary Structure of Proteins

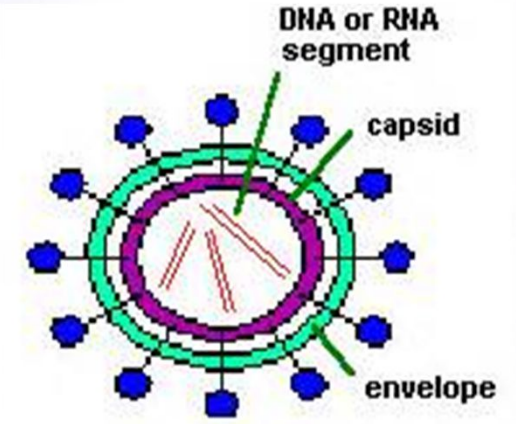


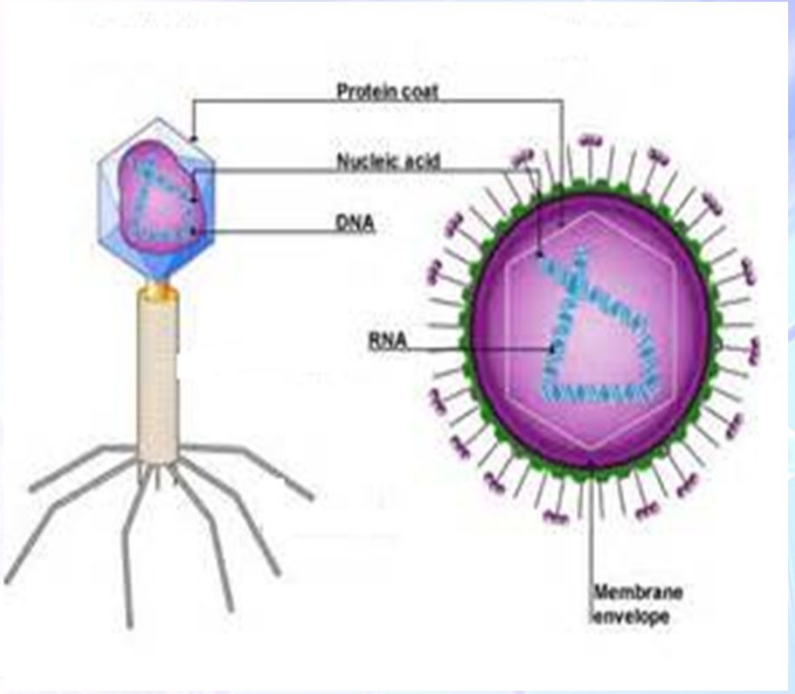
(adapted from Voet and Voet, 1990)

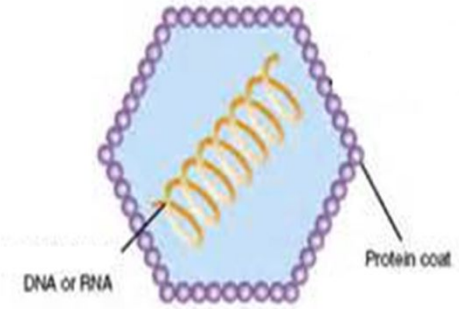
Genetic Materials in Viruses

- Viruses are exceptionally simple and extremely small microorganisms.
- They have a very simple structural organization consisting of a molecule of nucleic acid and a protein coat.

Genetic Materials in Viruses







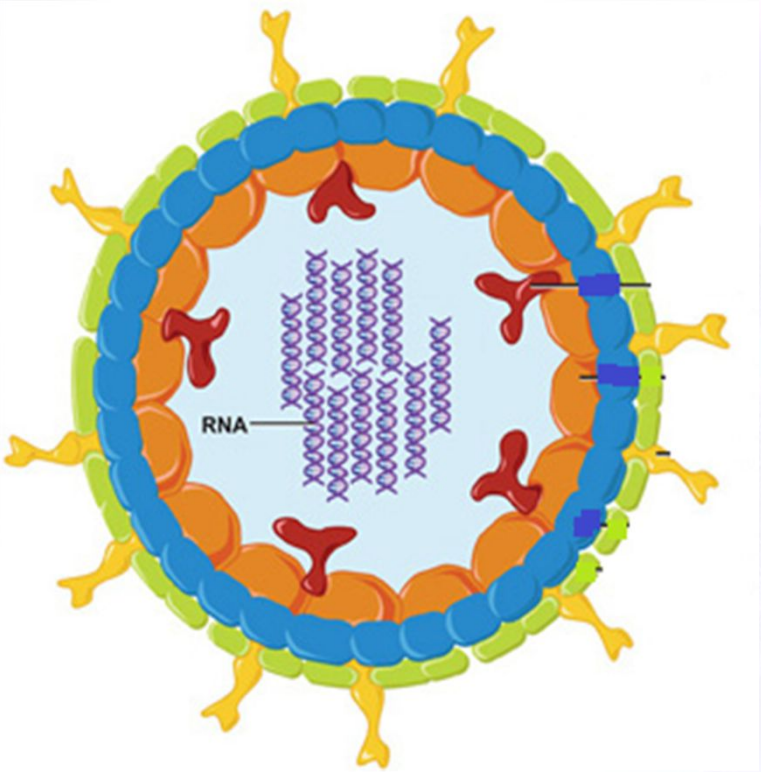
Genetic Materials in Viruses

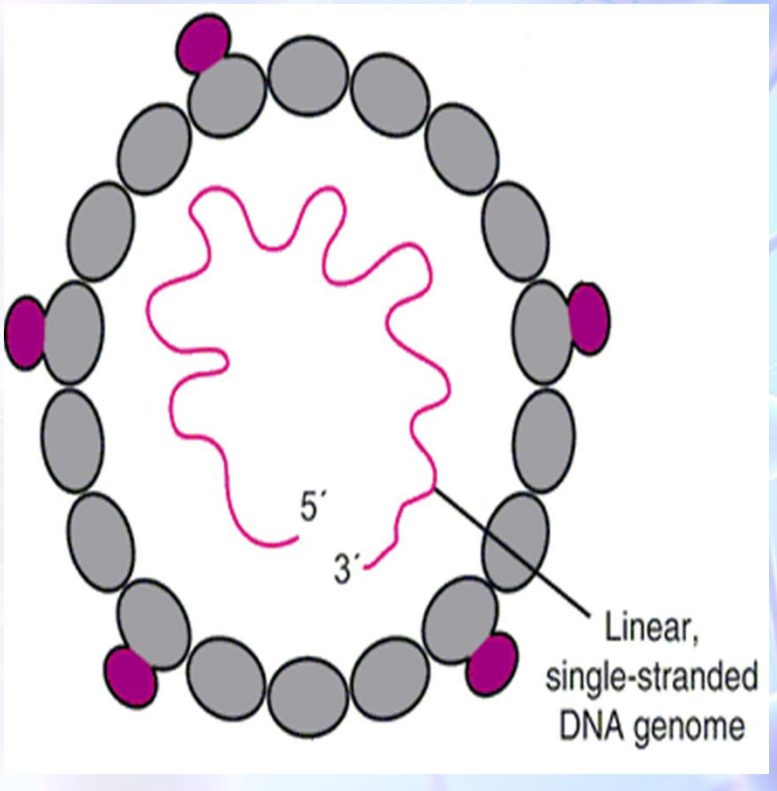
- Viruses can have either DNA or RNA as a genetic material but never both.
- The nucleic acid of a virus can be single-stranded or double-stranded.

Genetic Materials in Viruses

- There are viruses with the familiar double-stranded DNA, with single-stranded DNA, with double-stranded RNA and with single-stranded RNA.

Genetic Materials in Viruses

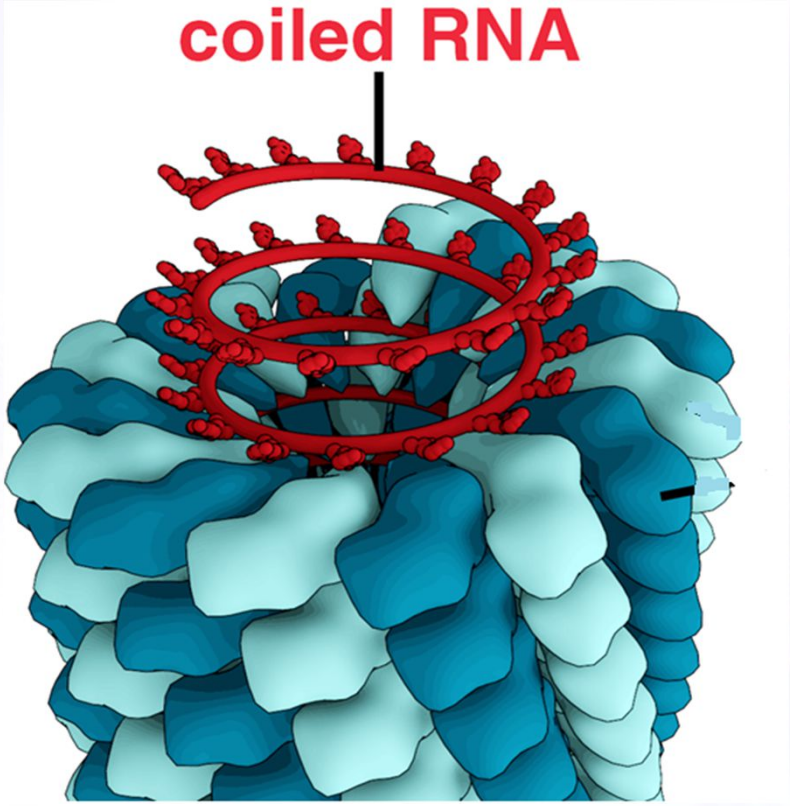




Genetic Materials in Viruses

- The viral nucleic acid can be linear or circular.
- Some viruses may contain coiled RNA.
- In some viruses , the nucleic acid is in several separate segments.

Genetic Materials in Viruses



Genetic Materials in Viruses

- The percentage of nucleic acid in relation to protein is about 1% for the influenza virus and about 50% for some bacteriophages.

Genetic Materials in Viruses

- The total amount of nucleic acid varies from a few thousand nucleotides to as many as 250,000 nucleotides.
- *E. coli's* chromosome consists of approx. 4 million nucleotide pairs.

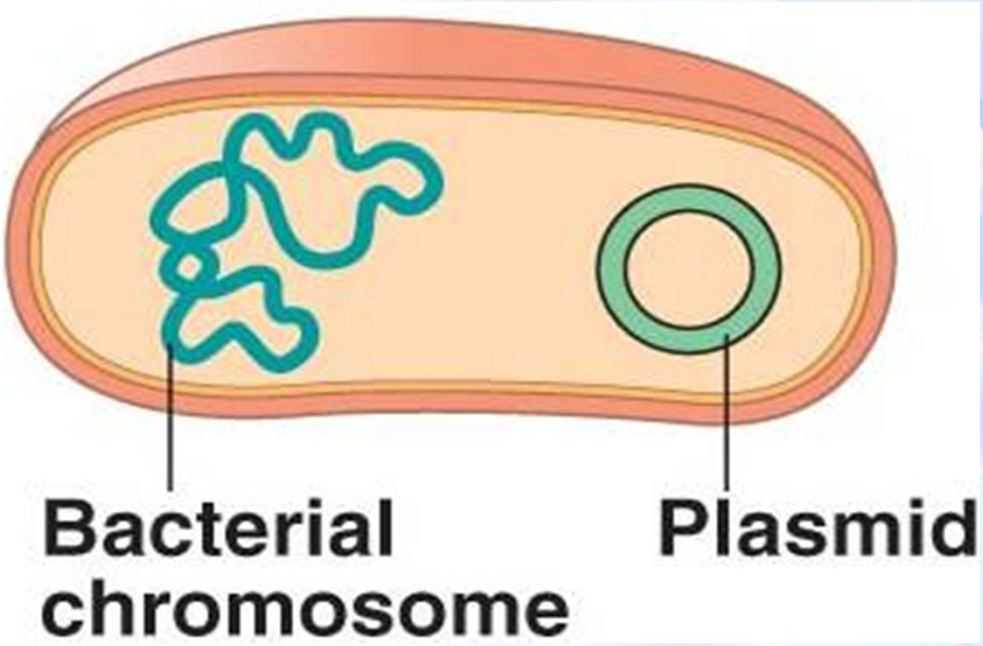
Organization of Genetic Material in Bacteria

- Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins.

Organization of Genetic Material in Bacteria

- The bacterial chromosome is a very long (up to 1mm).
- It is looped and folded and attached at one or several points to the plasma membrane.

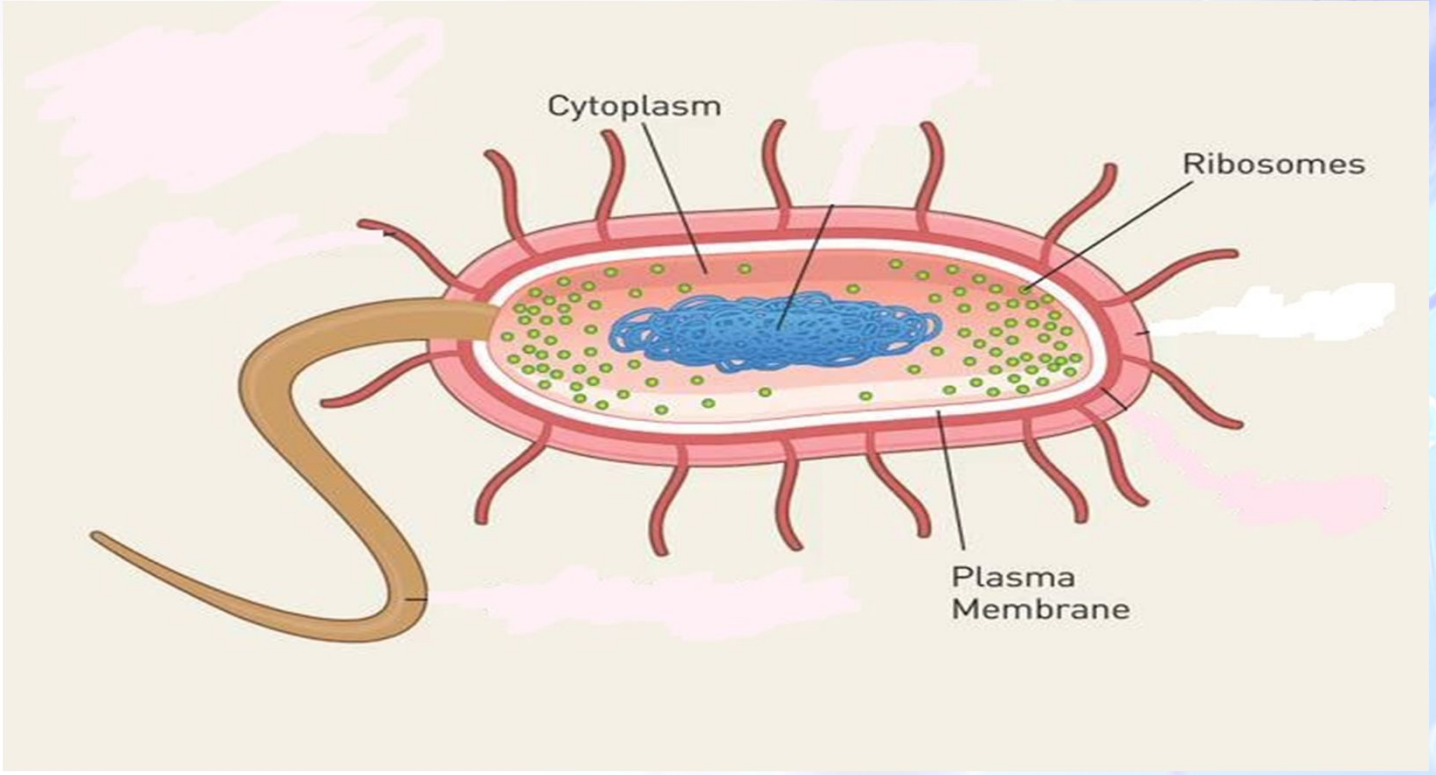
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the **nucleoid.**

Organization of Genetic Material in Bacteria



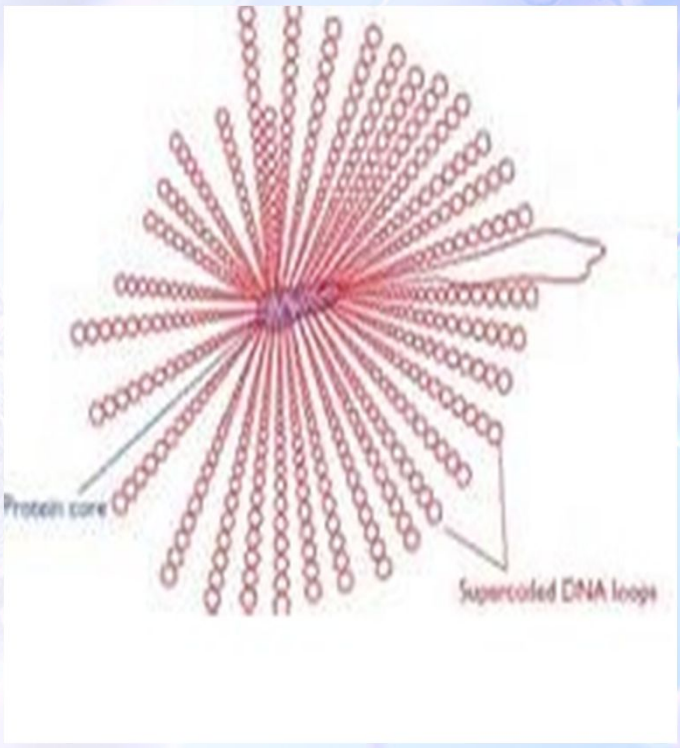
Organization of Genetic Material in Bacteria

- Bacterial chromatin can be released from the cell by gentle lysis of the cell.
- Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin.

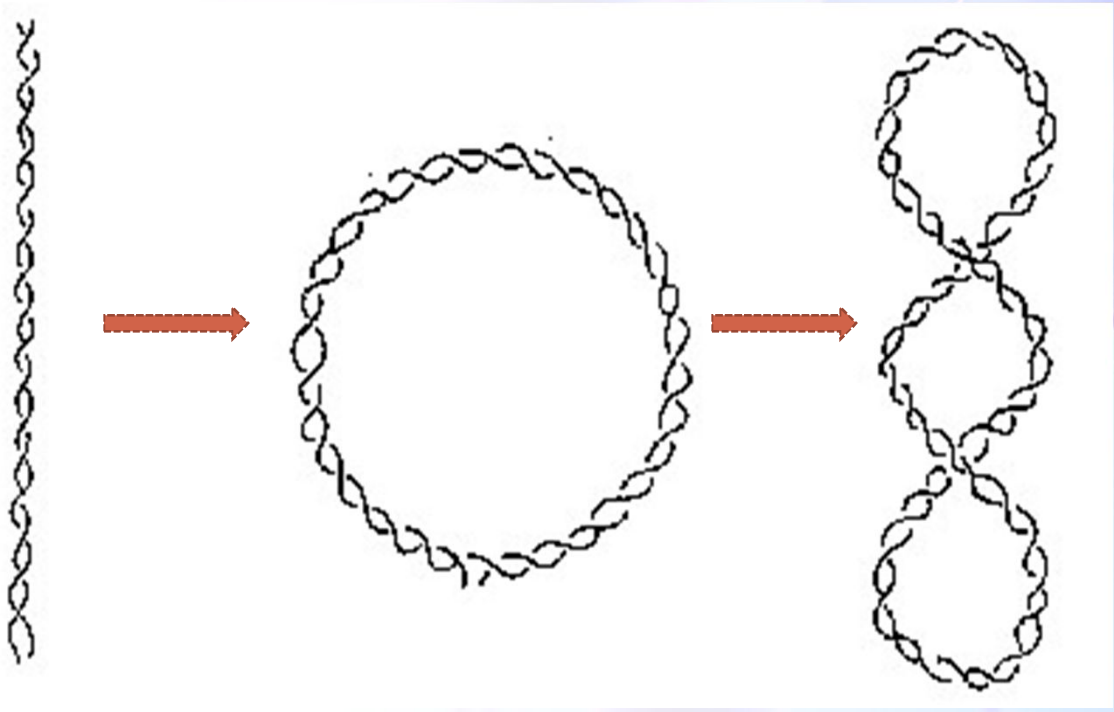
Organization of Genetic Material in Bacteria

- Some of the loops are super-coiled while some are relaxed.
- Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.

Organization of Genetic Material in Bacteria



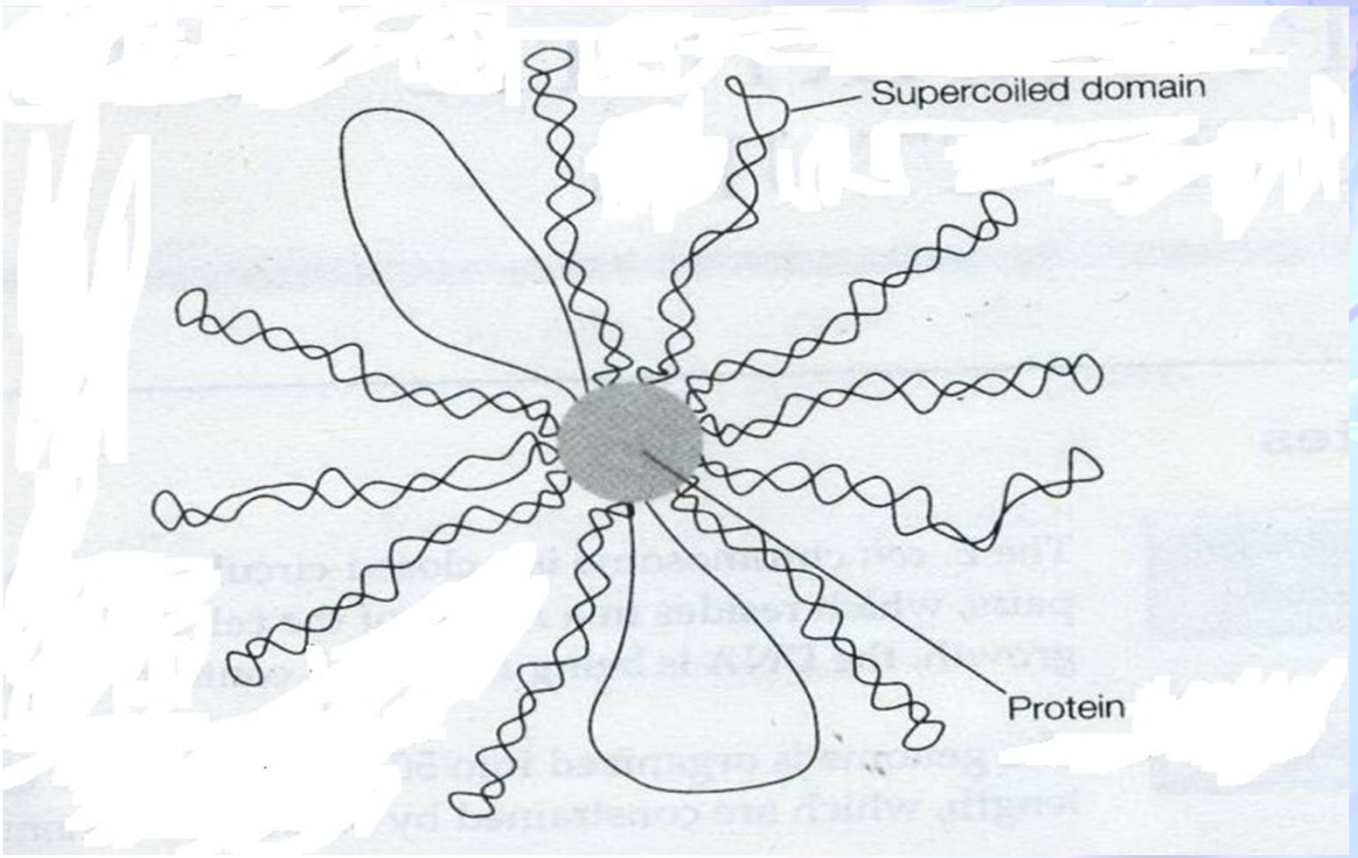
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- If a super-coiled DNA molecule receives a nick, the strain of under-winding is immediately removed, and all the super-coiling is lost.
- Studies confirm that continued nuclease treatment increases number of relaxed loops.

Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others.

Organization of Genetic Material in Bacteria

- So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.
- The super-coiled loops are dynamic structures which change during cell growth & division.

Organization of Genetic Material in Bacteria

- An *E. coli* chromosome is estimated to have about 400 super-coiled loops.
- Each loop has an average length of about 10-20 kbp.

Organization of Genetic Material in Bacteria

- The DNA compaction in a bacterial cell is contributed by super-coiling of loops, macromolecular crowding and DNA-binding proteins.

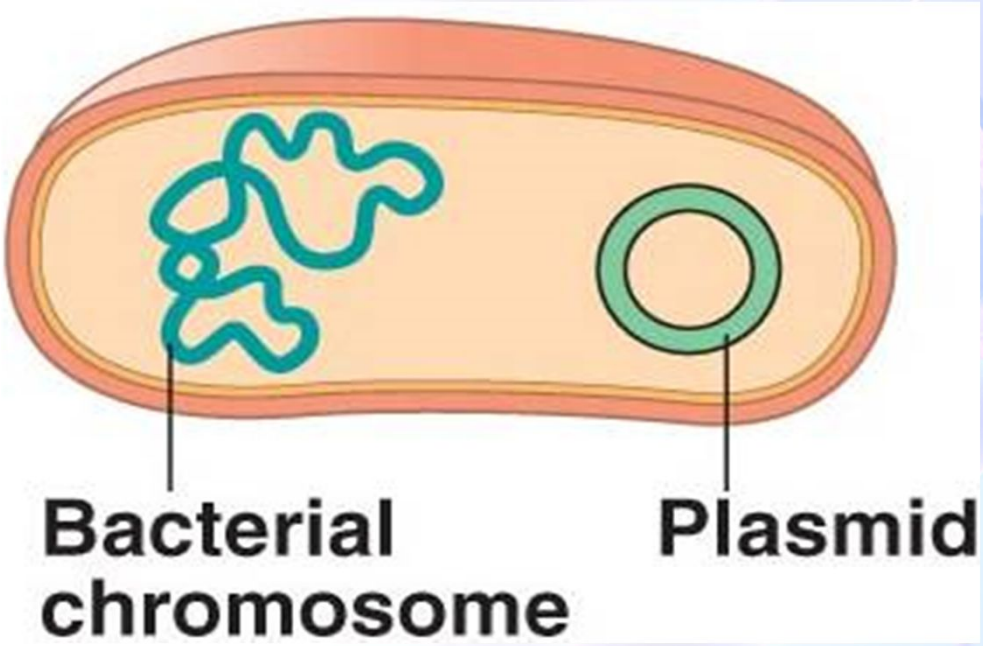
Organization of Genetic Material in Bacteria

- Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins.

Organization of Genetic Material in Bacteria

- The bacterial chromosome is a very long (up to 1mm).
- It is looped and folded and attached at one or several points to the plasma membrane.

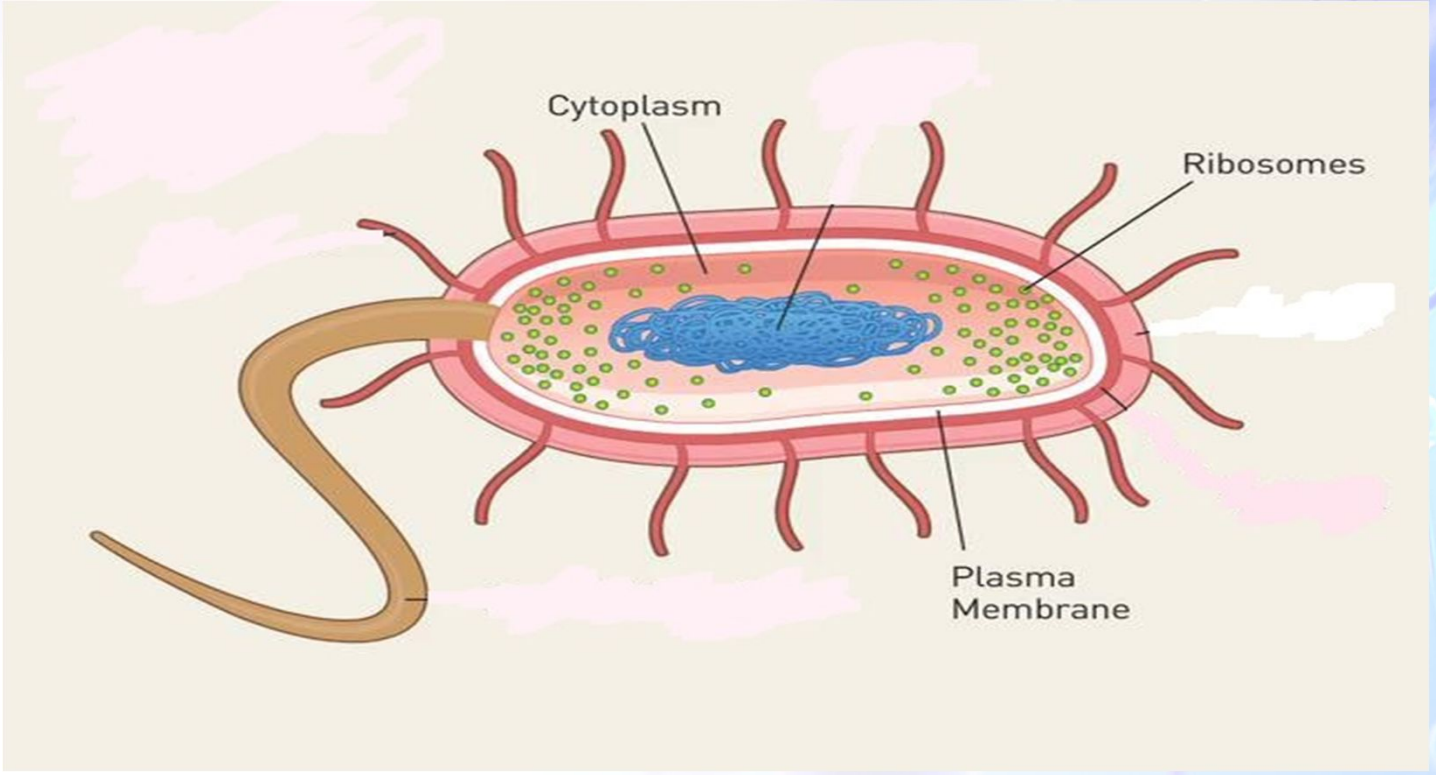
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the **nucleoid.**

Organization of Genetic Material in Bacteria



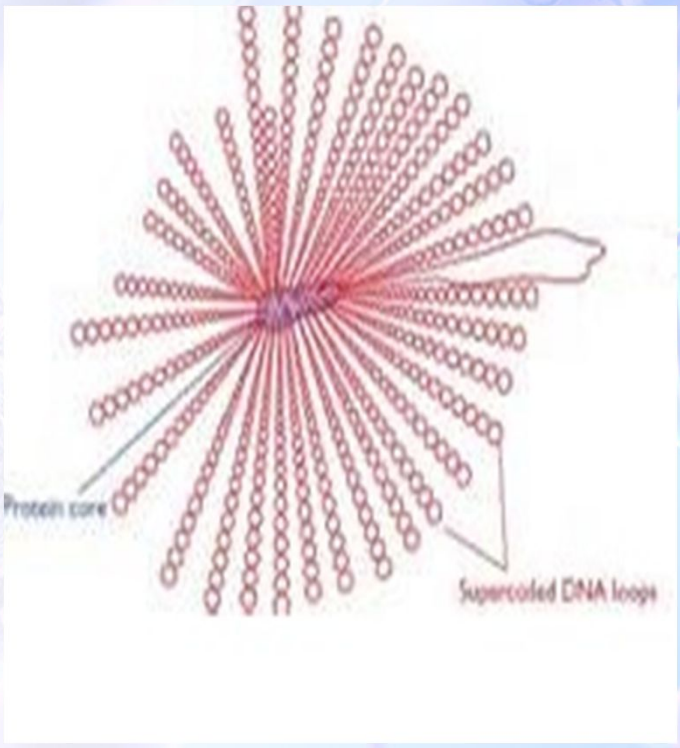
Organization of Genetic Material in Bacteria

- Bacterial chromatin can be released from the cell by gentle lysis of the cell.
- Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin.

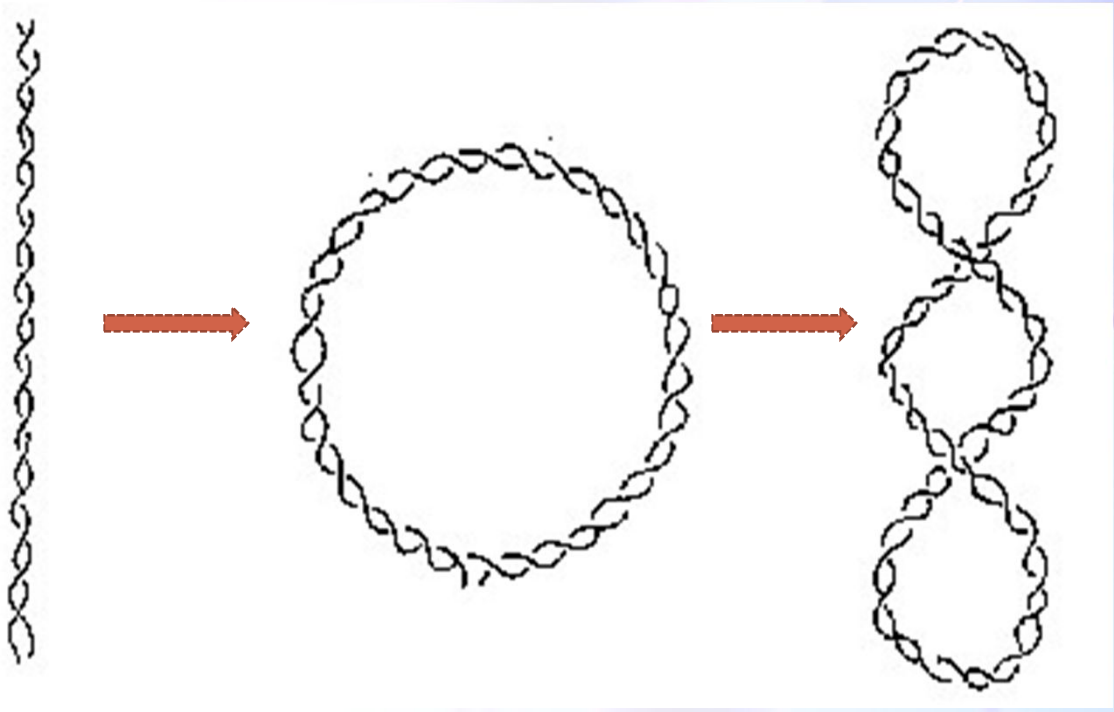
Organization of Genetic Material in Bacteria

- Some of the loops are super-coiled while some are relaxed.
- Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.

Organization of Genetic Material in Bacteria



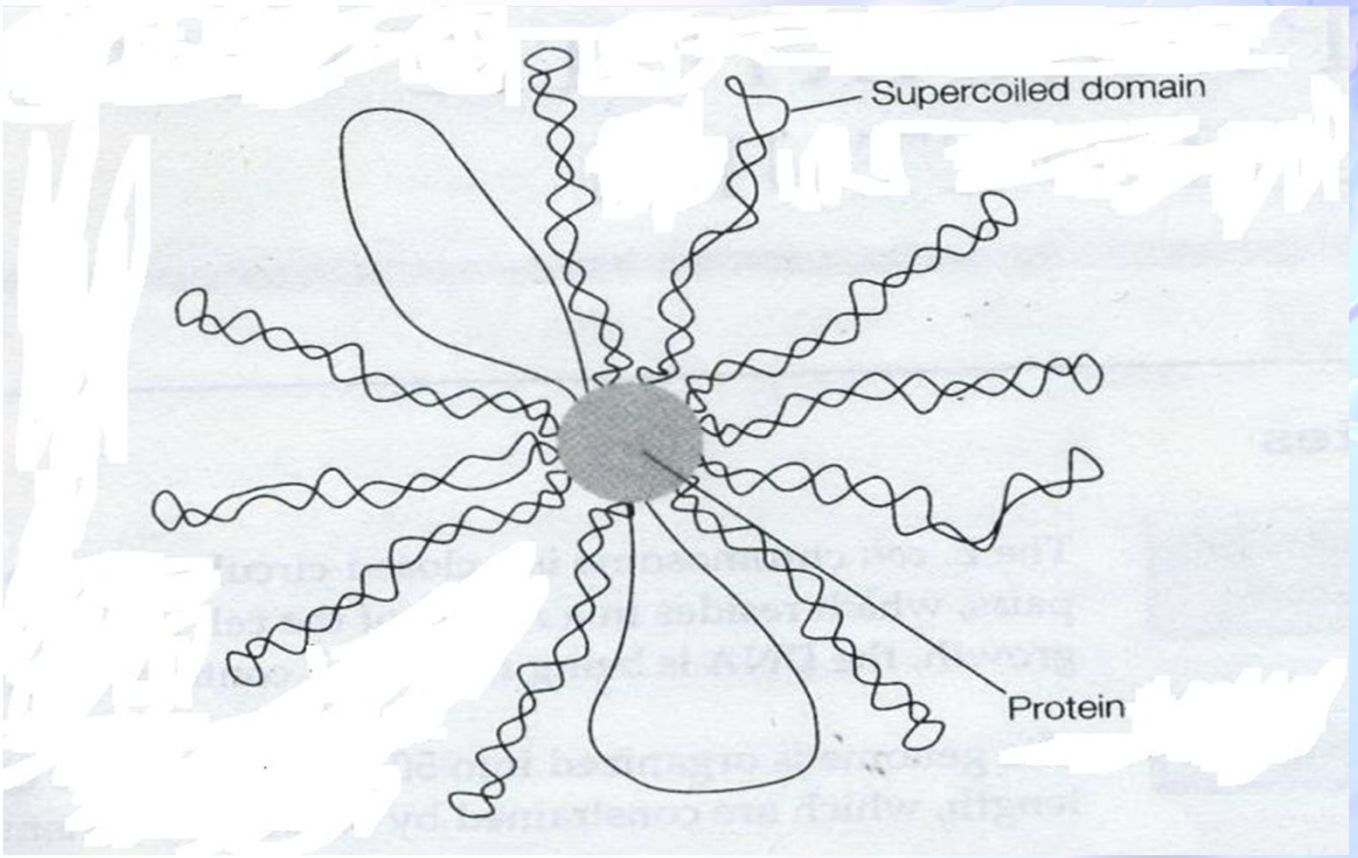
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- If a super-coiled DNA molecule receives a nick, the strain of under-winding is immediately removed, and all the super-coiling is lost.
- Studies confirm that continued nuclease treatment increases number of relaxed loops.

Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others.

Organization of Genetic Material in Bacteria

- So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.
- The super-coiled loops are dynamic structures which change during cell growth & division.

Organization of Genetic Material in Bacteria

- An *E. coli* chromosome is estimated to have about 400 super-coiled loops.
- Each loop has an average length of about 10-20 kbp.

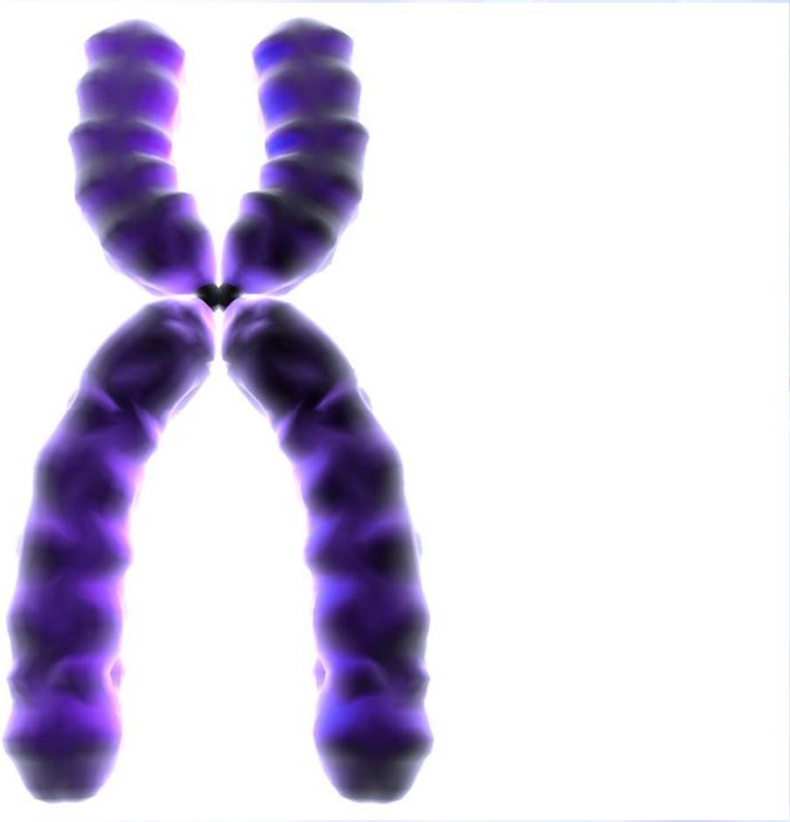
Organization of Genetic Material in Bacteria

- The DNA compaction in a bacterial cell is contributed by super-coiling of loops, macromolecular crowding and DNA-binding proteins.

Organization of Genetic Material in Eukaryotes

- The genetic material (DNA) of eukaryotic organisms is organized in the form of chromosomes.
- The chromosomes of eukaryotic cells are larger and more complex than those of prokaryotes.

Organization of Genetic Material in Eukaryotes



Organization of Genetic Material in Eukaryotes

- Each un-replicated chromosome consists of a single molecule of DNA.
- If stretched out, some human chromosomes would be several centimetres long.

Organization of Genetic Material in Eukaryotes

- To package such a tremendous length of DNA into this small volume, each DNA molecule is coiled again and again and tightly packed around histone proteins.

Organization of Genetic Material in Eukaryotes

- As eukaryotic chromosomes are not circular, so instead of super-coiling, the mechanism of packaging involves winding the DNA around special proteins, the **histones.**

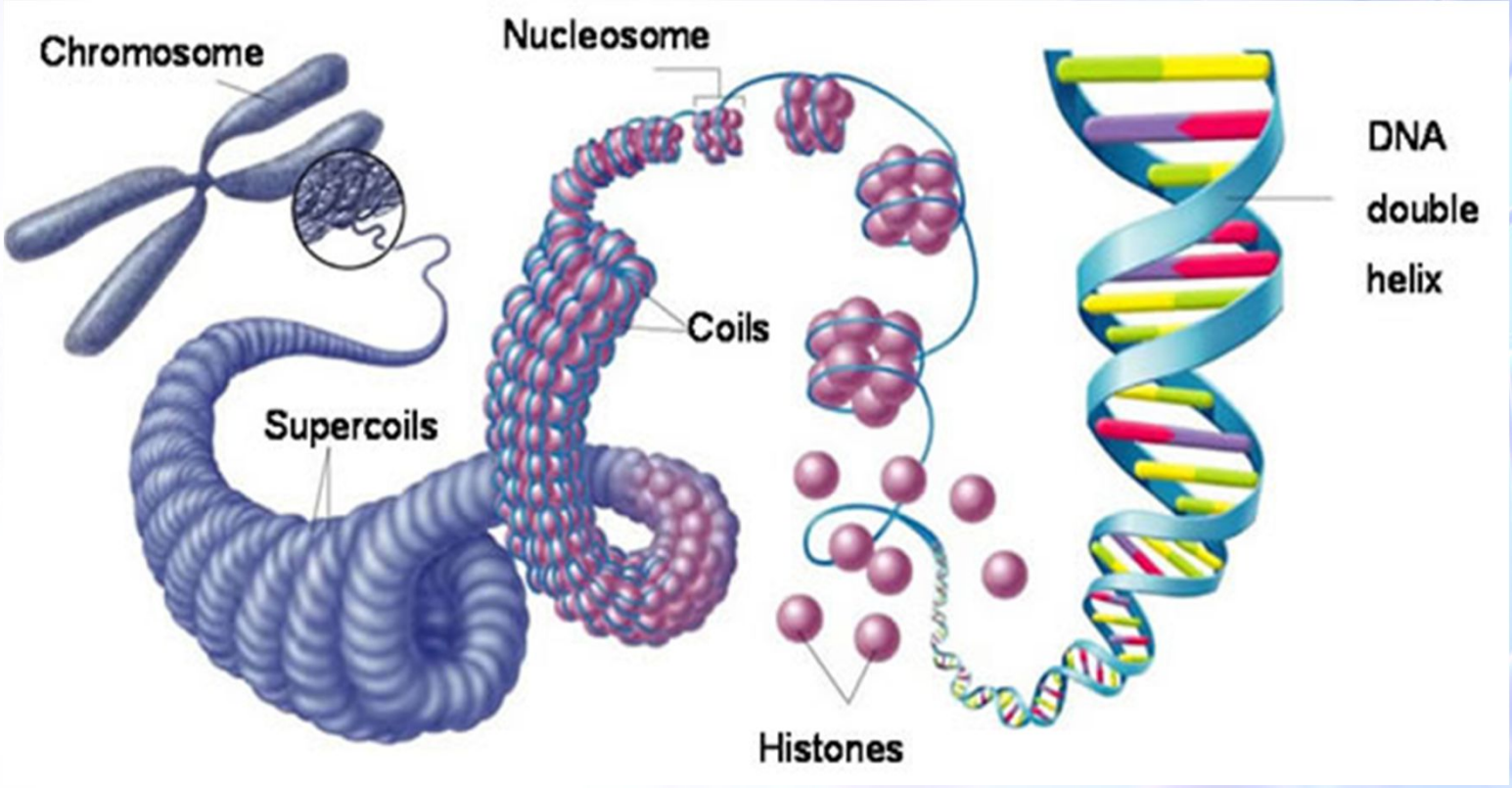
Organization of Genetic Material in Eukaryotes

- DNA with bound histones in the eukaryotes is called as **chromatin.**
- Chromatin consists of roughly spherical subunits, the **nucleosomes**, each containing approx. 200 bp of DNA and nine histones.

Organization of Genetic Material in Eukaryotes

- A condensed mitotic chromosome is about 50,000 times shorter than fully extended DNA.
- Highly condensed chromatin is known as **heterochromatin**.
- The more extended form is known as **euchromatin.**

Organization of Genetic Material in Eukaryotes



Genetics and Genomics

Genetics and Genomics - Difference

Genetics and Genomics-Difference

Genetics

- Genetics is the study of heredity, or how the characteristics of living organisms are transmitted from one generation to the next generation through DNA.

Genetics and Genomics-Difference

Genetics

- Genetics involves the study of specific and limited numbers of genes that have a known function.
- Genetics deals that how genes guide the body's development, cause disease or affect response to drugs.

Genetics and Genomics-Difference

Genomics

- Genomics, in contrast, is the study of the entirety of an organism's genes – called the genome.

Genetics and Genomics-Difference

Genomics

- Using high-performance computing and math techniques known as bioinformatics, genomics analyzes enormous amounts of DNA sequence data to find variations.

Genetics and Genomics-Difference

Genomics

- Genomics particularly deals with genetic variants that affect health, disease or drug response.
- In humans that means searching through about 3 billion units of DNA across 23,000 genes.

Genetics and Genomics-Difference

Genomics

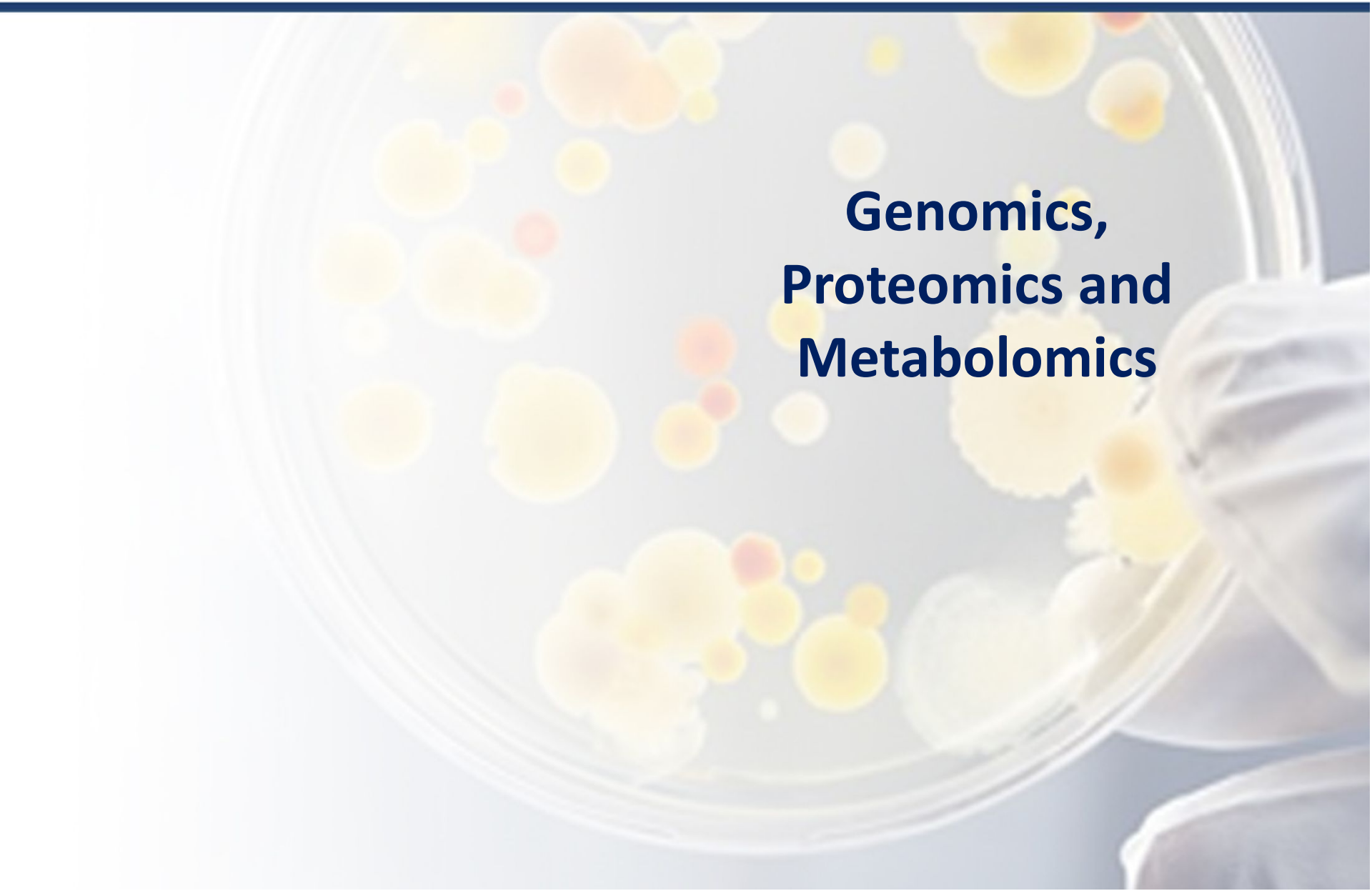
- Genomics is a much newer field than genetics and became possible only in the last couple of decades due to technical advances in DNA sequencing and computational biology.

Genetics and Genomics-Difference

Conclusion

- Genetics: How the characteristics of living organisms are transmitted from one generation to the next generation.
- Genomics: study of the entirety of an organism's genes – called the genome.

Genetics and Genomics



Genomics, Proteomics and Metabolomics

Genomics, Proteomics, Metabolomics

Genome and Genomics

- The complete set of DNA found in each cell is known as the genome and study is called as genomics.

Genomics, Proteomics, Metabolomics

Proteome and Proteomics

- The complete set of proteins found in each cell is known as the proteome.

Genomics, Proteomics, Metabolomics

Proteome and Proteomics

- Proteins concentration (and activity) may be different than gene expression due to post-translational modification

Genomics, Proteomics, Metabolomics

Metabolomics

- The complete set of metabolites found in each cell is known as the metabolome.
- Use of high-throughput mass spectrometry to analyze the metabolic components of cell.

Genomics, Proteomics, Metabolomics

Metabolomics

- Useful for determining the effects of the environment or gene transformation on the metabolism of the plants/animals.

Genomics, Proteomics, Metabolomics

Conclusion

- Genomics, proteomics and metabolomics will give an integrated, wholistic view of the cell.
- Can be used to monitor or modify organisms in a comprehensive way.

Genomics, Proteomics, Metabolomics

Conclusion

- Bioinformatics - the key to understand the plethora of information and modeling the cell.

Genetics and Genomics

Why Sequence Genomes

Why Sequence Genomes

Why Sequence Genomes

- To identify gene numbers, their locations on genomes, and to study their functions.
- Genes regulation

Why Sequence Genomes

Why Sequence Genomes

- DNA sequence
- Genome organization
- Chromosomal structure and organization

Why Sequence Genomes

Why Sequence Genomes

- Noncoding DNA types, amount, distribution and functions.
- Coordination of gene expression, protein synthesis, and post-translational events.

Why Sequence Genomes

Why Sequence Genomes

- Interaction of proteins in complex molecular machines
- Predicted vs experimentally determined gene function
- Evolutionary conservation

Why Sequence Genomes

Why Sequence Genomes

- Proteins structure and function.
- Proteomes (total protein content and function) in organisms.
- Correlation of SNPs with health and disease

Why Sequence Genomes

Why Sequence Genomes

- Disease-susceptibility prediction based on gene sequence variation
- Genes involved in complex traits and multigene diseases

Why Sequence Genomes

Novel Diagnostics

- Complex systems biology, developmental genetics.
- To provide platform for microchips and DNA microarrays.
- Gene expression - RNA

Why Sequence Genomes

Novel Diagnostics

- Complex systems biology, developmental genetics and genomics

Why Sequence Genomes

Novel Therapeutics

- Drug target discovery
- Rational drug design
- Molecular docking
- Gene therapy
- Stem cell therapy

Why Sequence Genomes

Understanding Metabolism

- To understand the metabolism of cells and tissues within different organisms.

Why Sequence Genomes

Understanding mechanism of diseases

- Inherited diseases
- Infectious diseases
- Pathogenic bacteria
- Viruses

Why Sequence Genomes

Conclusions

- Better understanding of the genomes would be possible by sequencing of the genomes.

Genetics and Genomics

Major Techniques used for Genomes Characterization

Genomes Characterization -Techniques

Major Techniques used for Genomes Characterization

- Cloning
- Hybridization
- PCR amplification
- Sequencing
- Computational tool

Genomes Characterization - Techniques

Genomes Characterization

Techniques - Cloning

- Genomes digested with restriction enzymes and inserted in vectors to produce genomic libraries.
- BACs
- YACs

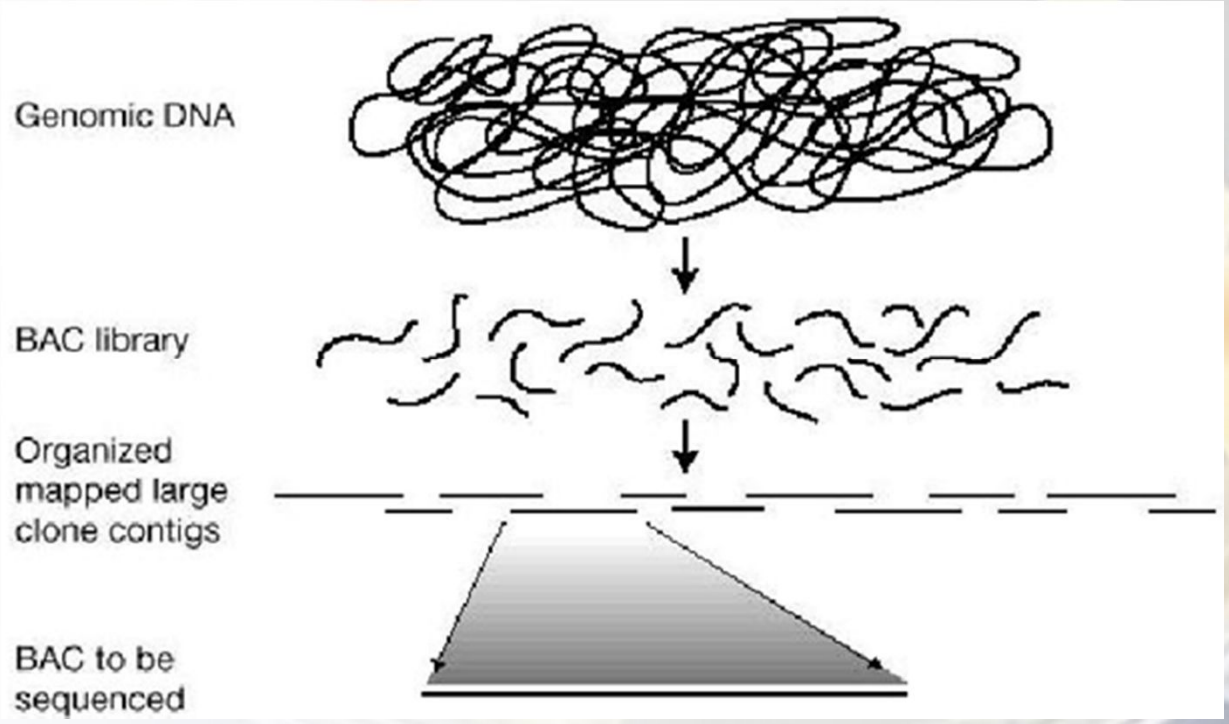
Genomes Characterization - Techniques

Genomes Characterization Techniques - Hybridization

- To arrange large contigs of genomes to produce genetic maps and physical maps of genomes.

Genomes Characterization - Techniques

Genomes Characterization Techniques – Hybridization: To arrange large contigs of genomes to produce genetic and physical maps



Genomes Characterization - Techniques

Genomes Characterization Techniques – PCR

- Technique to amplify the DNA. Different variants of the technique used

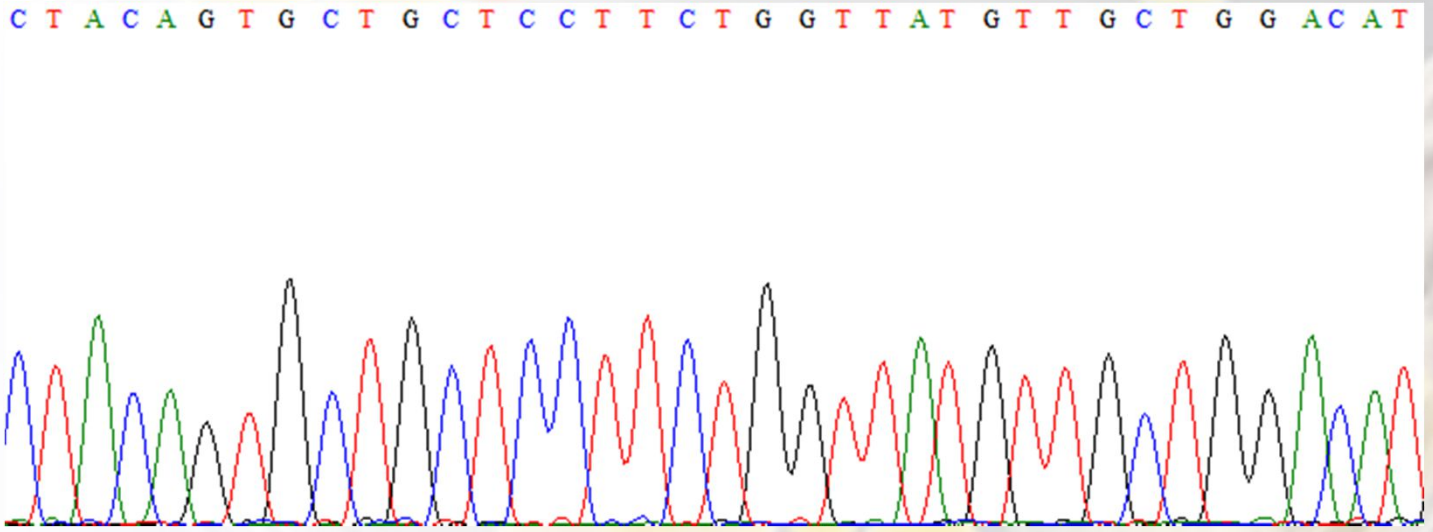
Genomes Characterization - Techniques

Genomes Characterization Techniques – DNA Sequencing

- One of the important technique used to characterize the genomes
- To study structure and function of genomes.

Genomes Characterization - Techniques

Genomes Characterization Techniques – DNA Sequencing: DNAs are amplified and sequenced



Genomes Characterization - Techniques

Genomes Characterization Techniques – Computational Tools

- Used to align the sequenced DNA to produce physical maps of the genomes.

Genomes Characterization - Techniques

Genomes Characterization Techniques – Conclusion

- Different techniques used for genomes characterizations.

Genetics and Genomics

Steps of Genomes Analysis

Genomes Analysis - Steps

Steps of Genomes Analysis

- Genome sequence assembled
- Identify repetitive sequences – mask out
- Gene prediction – train a model for each genome

Genomes Analysis - Steps

Steps of Genomes Analysis

- Look for EST and cDNA sequences
- Genome annotation
- Microarray analysis
- Metabolic pathways and regulation
- Protein 2D gel electrophoresis

Genomes Analysis - Steps

Steps of Genomes Analysis

- Functional genomics
- Gene location/gene map
- Self-comparison of proteome
- Comparative genomics

Genomes Analysis - Steps

Steps of Genomes Analysis

- Identify clusters of functionally related genes
- Evolutionary modeling

Genetics and Genomics

Benefits of Genomes Research

Genomes Research - Benefits

Genomes Research – Molecular Medicine

- Improve diagnosis of disease
- Detect genetic predispositions to disease (cancer, diabetes etc)
- Create drugs based on molecular information

Genomes Research - Benefits

Genomes Research – Molecular Medicine

- Use gene therapy and control systems as drugs

Genomes Research - Benefits

Genomes Research – Risk Assessment

- Evaluate the health risks faced by individuals who may be exposed to radiations and to cancer causing chemicals and toxins.

Genomes Research - Benefits

Genomes Research – Bioarchaeology, Anthropology, Evolution and Human Migration

- Study evolution through genetic variants in lineages.
- Study of migration of different populations

Genomes Research - Benefits

Genomes Research – Bioarchaeology, Anthropology, Evolution and Human Migration

- Study mutations on the Y chromosome to trace lineage and migration of males
- Evolution of mutations with ages of populations

Genomes Research - Benefits

Genomes Research – DNA Forensics

- Identify potential suspects whose DNA may match evidence left at crime scenes.
- Exonerate persons wrongly accused of crimes.
- Identify catastrophe victims.

Genomes Research - Benefits

Genomes Research – DNA Forensics

- Establish paternity and other family relationships.
- Identify endangered and protected species as an aid to wildlife officials.

Genomes Research - Benefits

Genomes Research

- Detect bacteria and other organisms that may pollute air, water, soil and food.
- Match organ donors with recipients in transplant programs
- Determine pedigree for seed or livestocks

Genomes Research - Benefits

Genomes Research – Disease-resistant crops and disease-resistant animals

- Grow disease/insect resistant and drought-resistant crops.
- Breed healthier, more productive, disease-resistant farm animals.

Genomes Research - Benefits

Genomes Research – Agriculture, Livestock Breeding, and Bioprocessing

- Develop biopesticides.
- Incorporate edible vaccines incorporated into food products.

Genomes Research - Benefits

Genomes Research – Microbial Genomics

- Rapidly detect and treat pathogens (disease-causing microbes).
- Develop new energy sources (biofuels)
- monitor environment to detect pollutants.

Genomes Research - Benefits

Genomes Research – Microbial Genomics

- Protect populations from biological and chemical warfare
- Clean up toxic waste safely and efficiently

Genetics and Genomics

Genes and Size of Genomes

Genes and Size of Genomes

Size of Genomes

- Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb).
- Genomes of eukaryotes are usually larger

Genes and Size of Genomes

Size of Genomes

- Most plants and animals have genomes greater than 100 Mb.
- Humans have genome size of 3,000 Mb

Genes and Size of Genomes

Size of Genomes

- Within each domain there is no systematic relationship between genome size and phenotype

Genes and Size of Genomes

Size of Genomes

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
Haemophilus influenzae	1.8	1,700	940
Escherichia coli	4.6	4,400	950
Archaea			
Archaeoglobus fulgidus	2.2	2,500	1,130
Methanosarcina barkeri	4.8	3,600	750

Genes and Size of Genomes

Size of Genomes

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Eukaryotes			
Saccharomyces cerevisiae (yeast, a fungus)	12	6,300	525
Caenorhabditis elegans (nematode)	100	20,100	200
Arabidopsis thaliana (mustard family plant)	120	27,000	225
Drosophila melanogaster (fruit fly)	165	13,700	83
Oryza sativa (rice)	430	42,000	98
Zea mays (corn)	2,300	32,000	14
Mus musculus (house mouse)	2,600	22,000	11
Ailuropoda melanoleuca (giant panda)	2,400	21,000	9
Homo sapiens (human)	3,000	<21,000	7

Genes and Size of Genomes

Conclusion

- Although most eukaryotes have large size of genomes.
- Within each domain there is no systematic relationship between genome size and phenotype

Genetics and Genomics

Viral Genomes

Viral Genomes

Genomes of Viruses

- Viral genomes can be
- ssRNA
- dsRNA
- ssDNA
- dsDNA
- Linear
- Ciruclar

Viral Genomes

Viruses Genomes

- A viral genome is the genetic material of the virus.
- Also termed the viral chromosome.
- Viral genomes vary in size -few thousand to more than a hundred thousand nucleotides.

Viral Genomes

Viruses with RNA Genomes

- Almost all plants viruses and some bacterial and animal viruses
- Genomes are rather small (a few thousands nucleotides)

Viral Genomes

Viruses with DNA Genomes

- Often a circular genome
- lambda = 48,502 bp

Viral Genomes

Replicative form of Viral Genomes

- All ssRNA viruses produce dsRNA molecules
- Many linear DNA molecules become circular

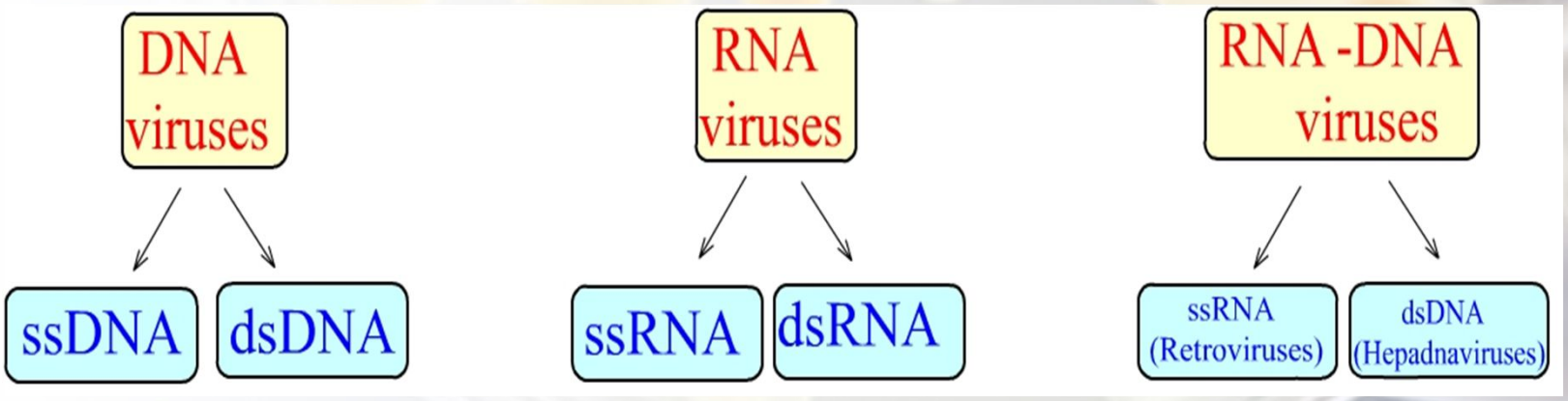
Viral Genomes

Viruses and Kingdoms

- Many plants viruses contain ssRNA genomes.
- Many fungal viruses contain dsRNA genomes.
- Many bacterial viruses contain dsDNA genomes.

Viral Genomes

Genomes in Virions: The genomes of viruses can be composed of either DNA or RNA, and some use both as their genomic material at different stages in their life cycle. However, only one type of nucleic acid is found in the virion of any particular type of virus.



Viral Genomes

Viruses and Number of Genes

Virus	Host	Type of Nucleic Acid	Number of Genes
Parvovirus	Mammals	ssDNA	5
Phage fd	*E. coli*	ssDNA	10
Lambda	*E. coli*	dsDNA	36
T4	*E. coli*	dsDNA	>190
Qβ	*E. coli*	ssRNA	4
TMV	Many plants	ssRNA	6
Influenza virus	Mammals	ssRNA	12

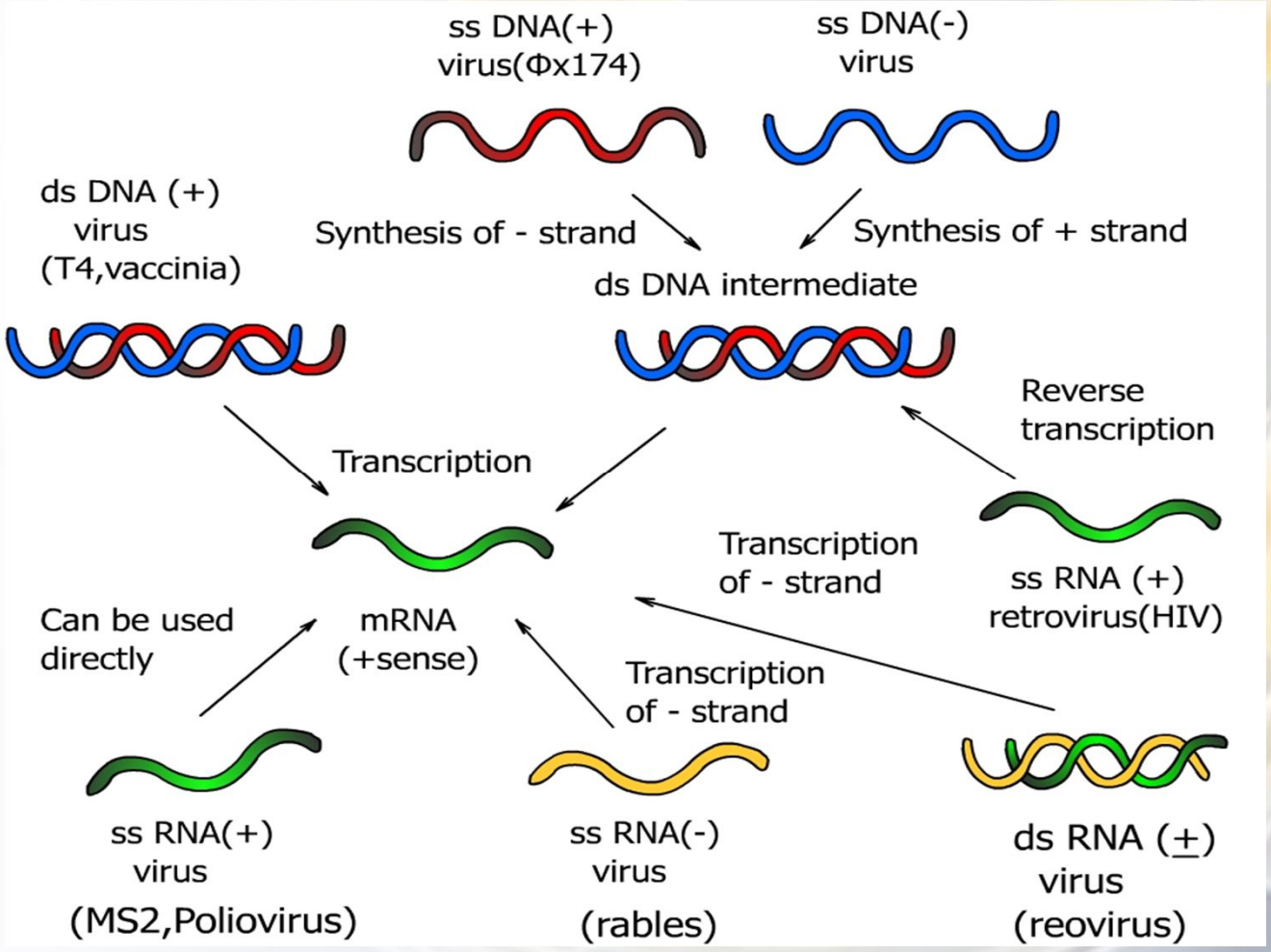
Viral Genomes

Viruses and Number of Genes

Virus	Genome structure	Genome size (kb)	Number of genes
Adenovirus	Double-stranded linear DNA	36.0	30
Hepatitis B	Partly double-stranded circular DNA	3.2	4
Influenza virus	Single-stranded segmented linear RNA	22.0	12
Parvovirus	Single-stranded linear DNA	1.6	5
Poliovirus	Single-stranded linear RNA	7.6	8
Reovirus	Double-stranded segmented linear RNA	22.5	22
Retroviruses	Single-stranded linear RNA	6.0–9.0	3
SV40	Double-stranded circular DNA	5.0	5
Tobacco mosaic virus	Single-stranded linear RNA	6.4	6
Vaccinia virus	Double-stranded circular DNA	240	240

Viral Genomes

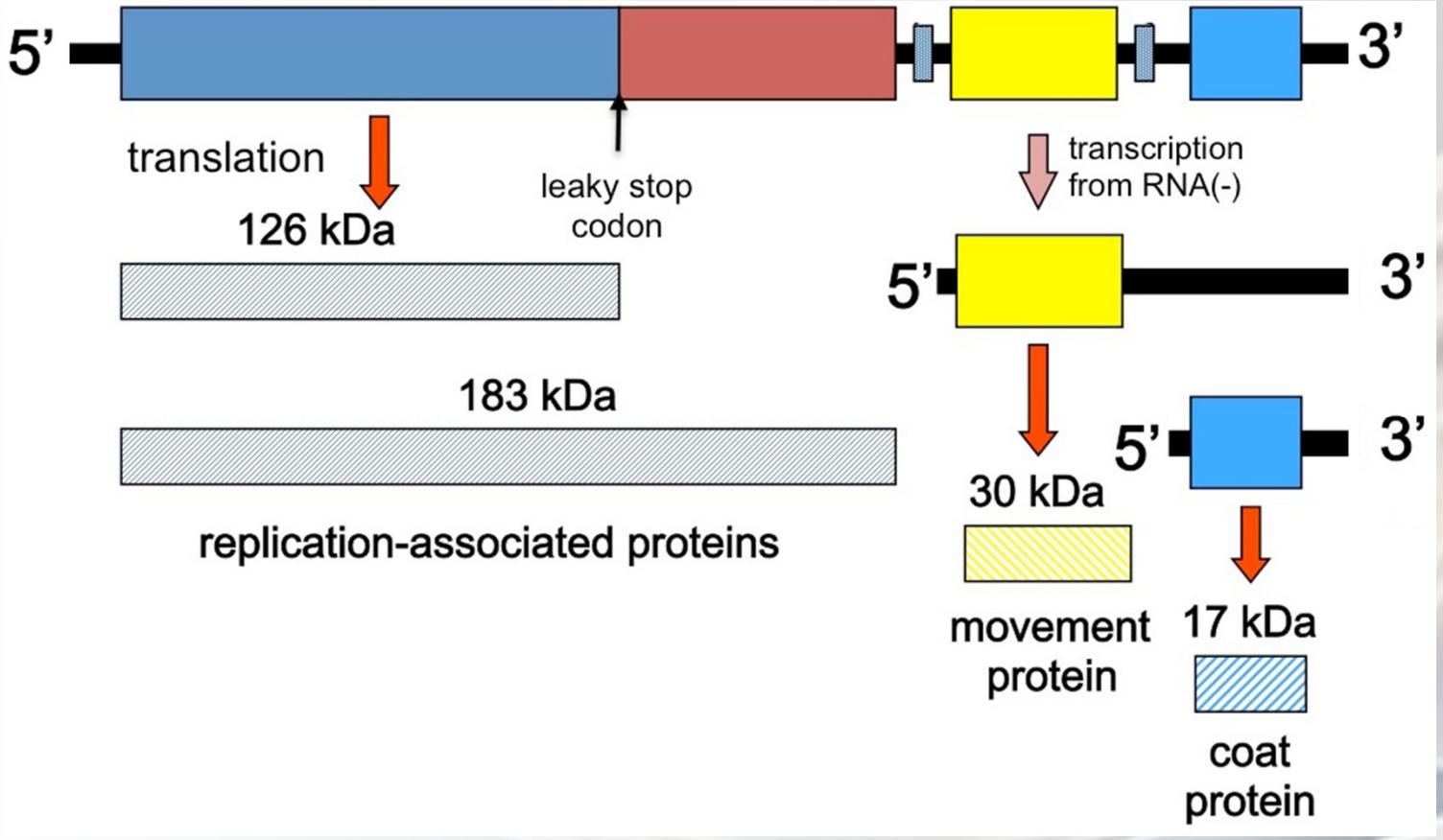
Genomes in Virions



Viral Genomes

Genome of Tobacco Mosaic Virus

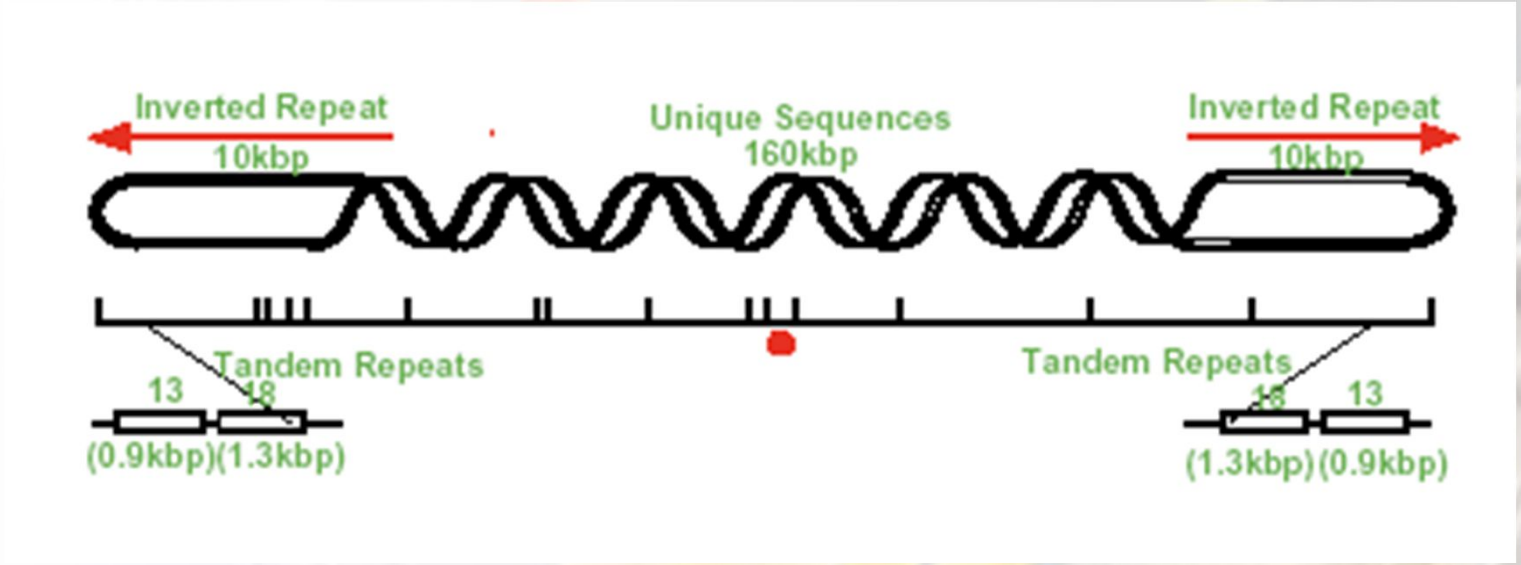
- Single, 6400 nucleotides RNA, 3 Essential Genes



Viral Genomes

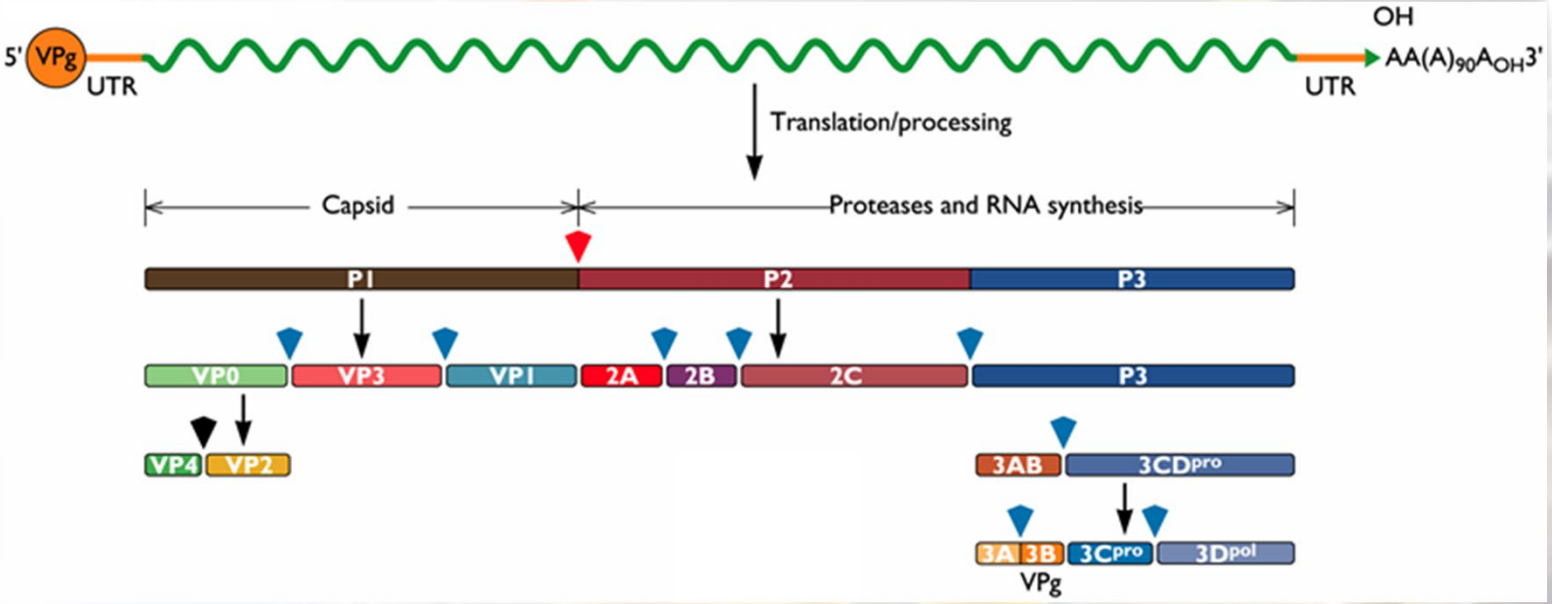
Genome of Poxvirus –A typical large dsDNA Virus

- 180 kb DNA, >100 Essential Genes



Viral Genomes

Genome of Polio Virus: Single-stranded positive-sense RNA genome that is about 7500 nucleotides long



Viral Genomes

Genome of Pox Virus

- Linear dsDNA 130-375 kbp; covalently closed termini.
- Large hairpin structure at each terminus - up to 10 kb total at each end is repeat sequence.
- Encode 150-300 proteins.
- Coding regions are closely spaced, no introns.
- Coding regions are on both strands of genome, and are not tightly clustered with respect to time of expression or function.

Viral Genomes

Viral Genomes

- Viral genomes can be
- ssRNA
- dsRNA
- ssDNA
- dsDNA
- Linear
- Ciruclar

Genetics and Genomics

Bacterial Genomes

Bacterial Genomes

Genomes of Bacteria

- Small organisms carry high coding density (85-90%)
- 1 gene per 1000 bases in prokaryotes
- Large variation in genome size between bacteria

Bacterial Genomes

Genomes of Bacteria – Large Variation

- *Tremblaya princeps* 140kb, 121 coding sequences
- *Sorangium cellulosum*
- 14000kb
- 11599 coding sequences

Bacterial Genomes

Genomes of Bacteria

Comparison of regulatory genes in bacterial genomes

Microorganism	# Genes in the Genome	# Regulatory Proteins	% of Total
Pseudomonas aeruginosa	5570	468	8.4
Escherichia coli	4289	250	5.8
Bacillus subtilis	4100	217	5.3
Mycobacterium tuberculosis	3918	117	3.0
Helicobacter pylori	1566	18	1.1

Bacterial Genomes

Distribution of genes among selected bacterial genomes and their sizes

Organism	Genome Size (Mbp)	No. of ORFs (% coding)		Unknown Function		Unique ORFs	
Aeropyrum pernix K1	1.67	1,885	(89%)				
A. aeolicus VF5	1.50	1,749	(93%)	663	(44%)	407	(27%)
A. fulgidus	2.18	2,437	(92%)	1,315	(54%)	641	(26%)
B. subtilis	4.20	4,779	(87%)	1,722	(42%)	1,053	(26%)
B. burgdorferi	1.44	1,738	(88%)	1,132	(65%)	682	(39%)
Chlamydia pneumoniae AR39	1.23	1,134	(90%)	543	(48%)	262	(23%)
Chlamydia trachomatis MoP_n	1.07	936	(91%)	353	(38%)	77	(8%)
C. trachomatis serovar D	1.04	928	(92%)	290	(32%)	255	(29%)
Deinococcus radiodurans	3.28	3,187	(91%)	1,715	(54%)	1,001	(31%)
E. coli K-12-MG1655	4.60	5,295	(88%)	1,632	(38%)	1,114	(26%)
H. influenzae	1.83	1,738	(88%)	595	(35%)	237	(14%)
H. pylori 26695	1.66	1,589	(91%)	744	(45%)	539	(33%)
Methanobacterium thermotautotrophicum	1.75	2,008	(90%)	1,010	(54%)	496	(27%)

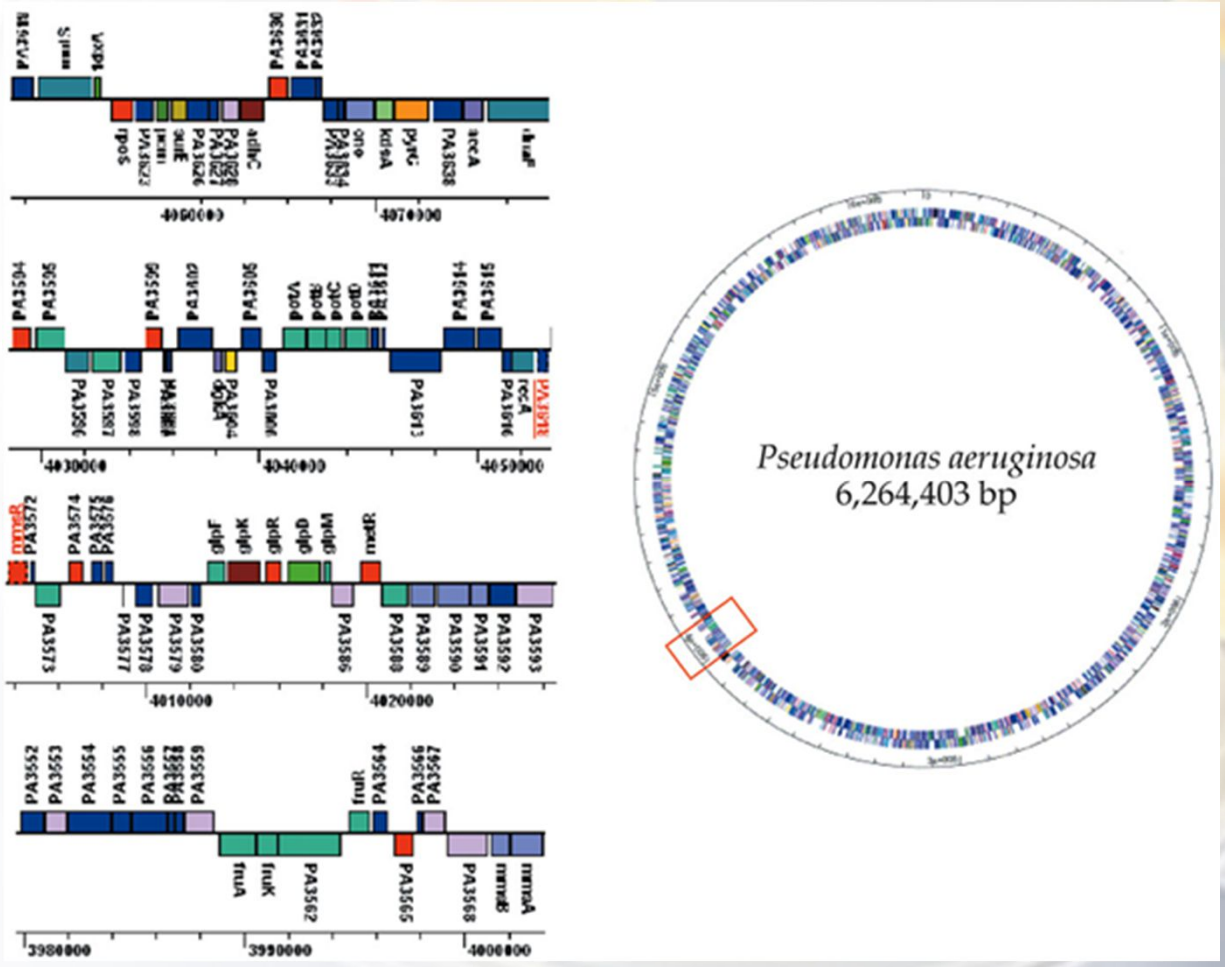
Bacterial Genomes

Distribution of genes among selected bacterial genomes and their sizes

Organism	Genome Size (Mbp)	No. of ORFs (% coding)		Unknown Function		Unique ORFs	
Methanococcus jannaschii	1.66	1,783	(87%)	1,076	(62%)	525	(30%)
M. tuberculosis CSU#93	4.41	4,275	(92%)	1,521	(39%)	606	(15%)
M. genitalium	0.58	483	(91%)	173	(37%)	7	(2%)
M. pneumoniae	0.81	680	(89%)	248	(37%)	67	(10%)
N. meningitidis MC58	2.24	2,155	(83%)	856	(40%)	517	(24%)
Pyrococcus horikoshii OT3	1.74	1,994	(91%)	589	(42%)	453	(22%)
Rickettsia prowazekii Madrid E	1.11	878	(75%)	311	(37%)	209	(25%)
Synechocystis sp.	3.57	4,003	(87%)	2,384	(75%)	1,426	(45%)
T. maritma MSB8	1.86	1,879	(95%)	863	(46%)	373	(26%)
T. pallidum	1.14	1,039	(93%)	461	(44%)	280	(27%)
Vibrio cholerae El Tor N1696	4.03	3,890	(88%)	1,806	(46%)	934	(24%)
	50.60	52,462	(89%)	22.358	(43%)	12,161	(23%)

Bacterial Genomes

Genes in a portion of a bacterial genome



Bacterial Genomes

Bacterial Genomes - Conclusion

- Small organisms carry high coding density.
- Large variation in genome size between bacteria.

Genetics and Genomics

Yeast Genome

Yeast Genome

Yeast Genome

- The nuclear genome consists of 16 chromosomes.
- In addition, there is a mitochondrial genome and a plasmid, 2 micron circle.

Yeast Genome

Yeast Genome

- The haploid yeast genome consists of ~ 12.1 Mb
- Yeast genome was completely sequenced by 1996

Yeast Genome

Yeast Genome - Characteristics

- Small and compact
- Small intergenic sequences
- Few transposable elements
- Few introns
- Limited RNA interference

Yeast Genome

Yeast Genome

- The yeast genome is predicted to contain about 6,200 genes
- 274 tRNA
- 287 introns
- Small percentage of yeast genes have introns
- The intergenic space between genes is only between 200bp - 1,000bp

Yeast Genome

Yeast Genome: Genome of Yeast Cell

Characteristic	Chromosomes	Plasmid	Mitochondria
Relative amount (%)	85	5	10
Number of copies	2 x 16	60-100	~50 (8-130)
Size (kbp)	~ 12,100	6.318	70-76

Yeast Genome

Yeast Genome

- The largest known regulatory sequences are spread over about 2,800bp
- MUC1/FLO11

Yeast Genome

Yeast Genome

- Yeast genes have names consisting of three letters and up to three numbers
- GPD1, HSP12, PDC6
- Usually they are meaningful

Yeast Genome

Yeast Genome – Genes Nomenclature

- Wild type genes are written with capital letters in italics: *TPS1*, *RHO1*, *CDC28*
- Recessive mutant genes are written with small letters in italics: *tps1, rho1, cdc28*

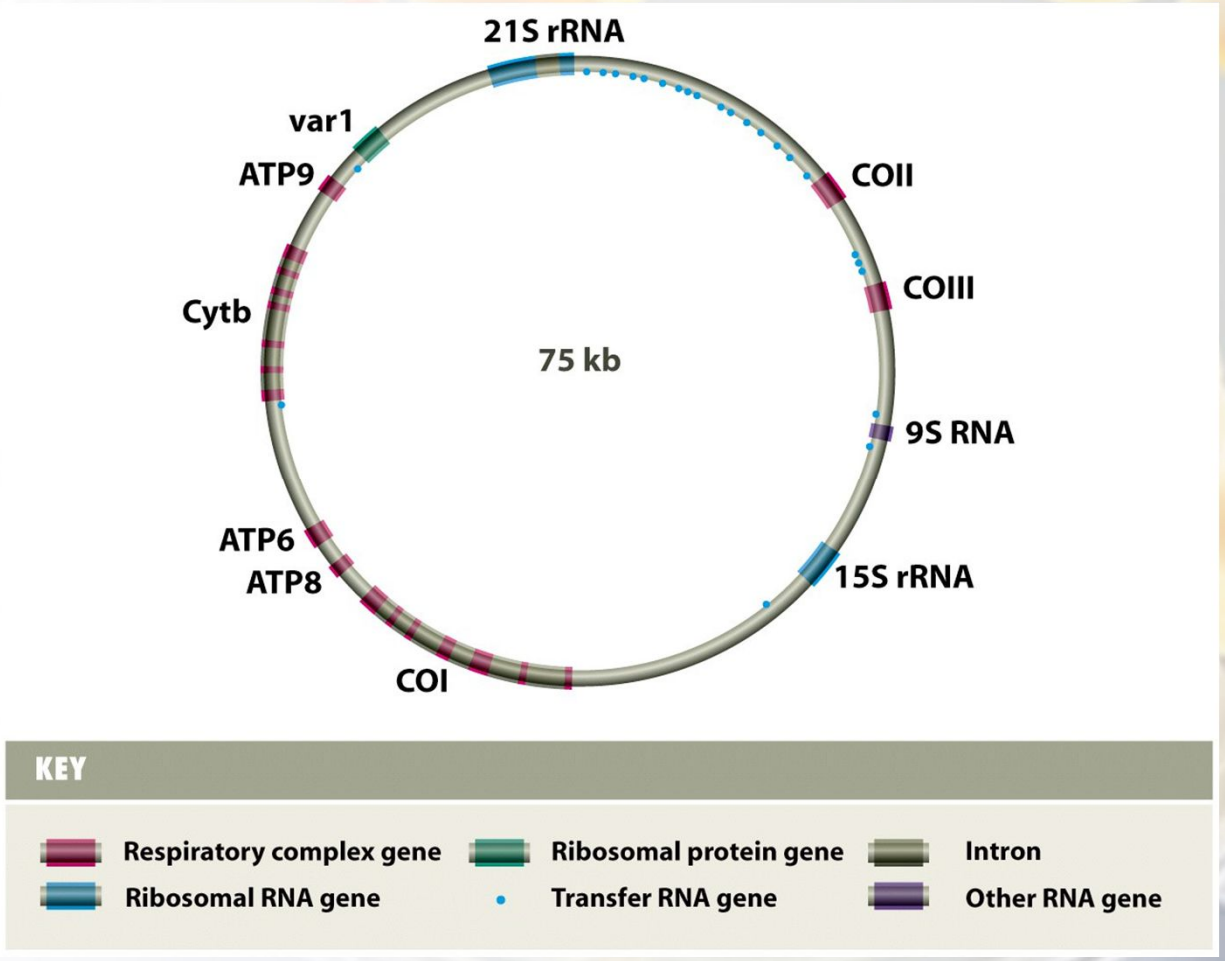
Yeast Genome

Yeast Genome – Genes Nomenclature

- Three letters provides information about a function, mutant phenotype, or process related to that gene.
- CDC - Cell Division Cycle ; ADE-ADEnine biosynthesis

Yeast Genome

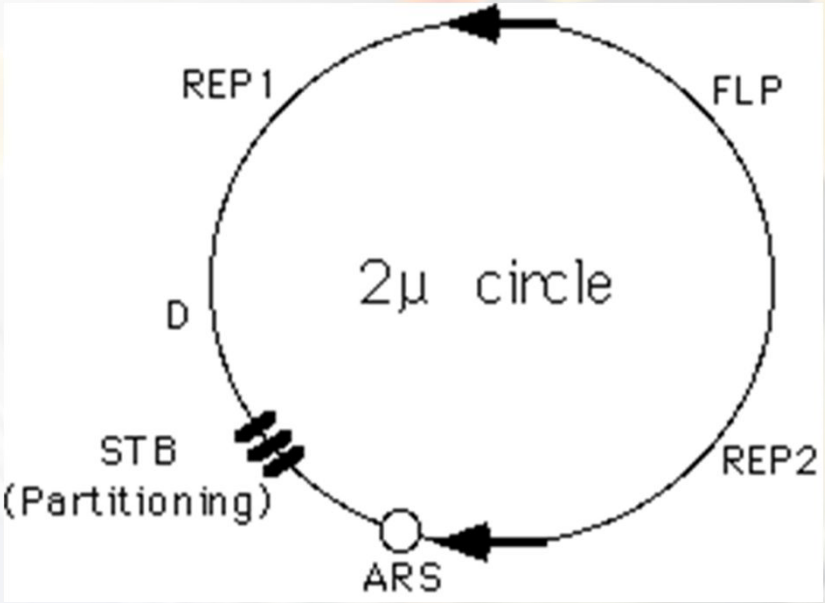
Yeast – Mitochondrial DNA



Yeast Genome

Yeast – Plasmid DNA

- The 2u circle is a 6.3 kb
- 50 to 100 copies per haploid genome of the yeast cells
- ARS, the FLP gene, the three genes which encode proteins required for regulation of FLP expression (REP2, REP1, and D)
- Set of small direct repeats (called "STB") required for partitioning into daughter cells during mitosis and meiosis.



Yeast Genome

Yeast Genome

- Yeast nuclear genome has 16 chromosomes.
- A mitochondrial genome.
- A plasmid.

Mitochondrial Genome

Mitochondrial Genome

Mitochondrial Genome

- Multiple identical circular chromosomes
- ~15-16 Kb in animals
- ~ 200 kb to 2,500 kb in plants

Mitochondrial Genome

Mitochondrial Genome

- Over 95% of mitochondrial proteins are encoded in the nuclear genome.
- Often A+T rich genomes.

Mitochondrial Genome

Human Mitochondrial Genome

- Circular, double stranded, 16.6 kb
- The two strands are notably different in base composition, leading to one strand being heavy (H strand) and the other light (L strand).

Mitochondrial Genome

Human Mitochondrial Genome

- Both strands encode genes, although more are on the H strand.
- A short region (1121 bp), the D loop is a DNA triple helix: two overlapping copies of the H strand.

Mitochondrial Genome

Human Mitochondrial Genome

- The D loop is also the site where most of replication and transcription is controlled.
- Genes are tightly packed, with almost no non-coding DNA outside of D loop.

Mitochondrial Genome

Human Mitochondrial Genome

- Human mitochondrial genes contain no introns, although introns are found in the mitochondria of other groups (plants)

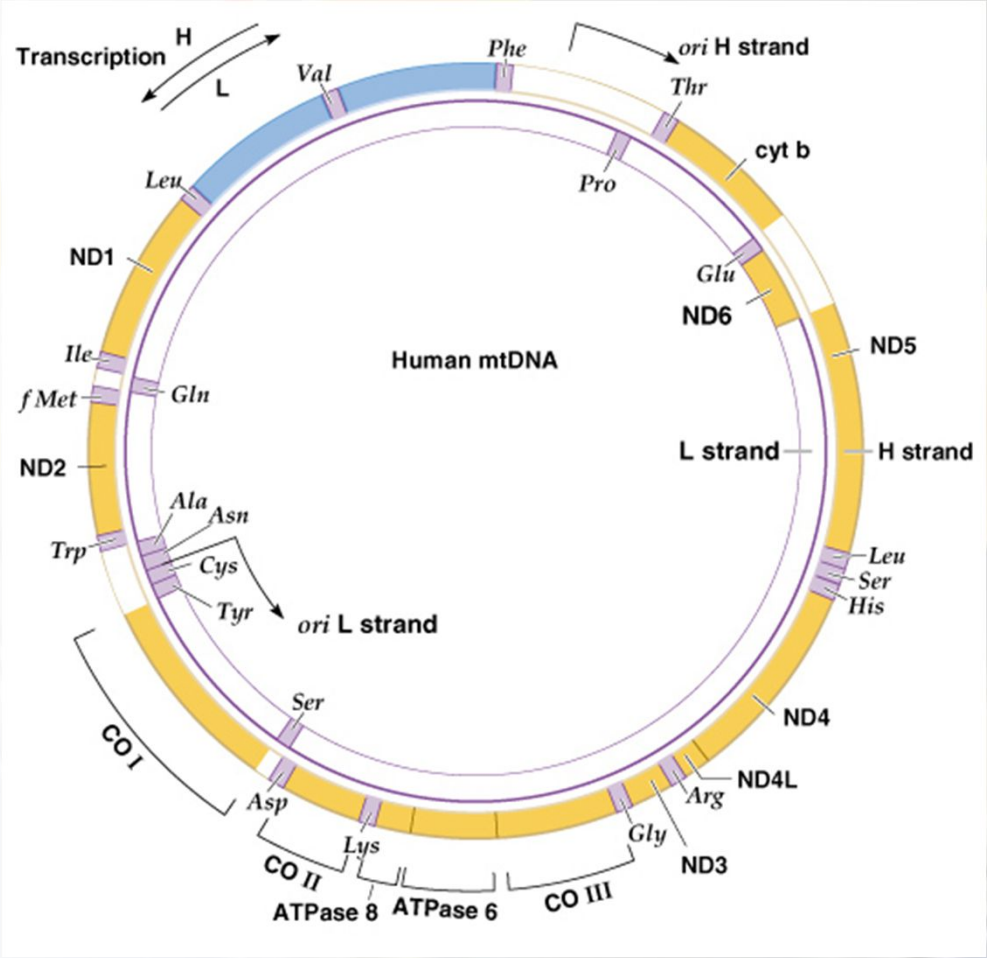
Mitochondrial Genome

Human Mitochondrial Genome

- 37 Genes
- 22 tRNAs
- 2 rRNAs
- 13 polypeptides
- tRNA: only 60 of the 64 codons code for amino acids.

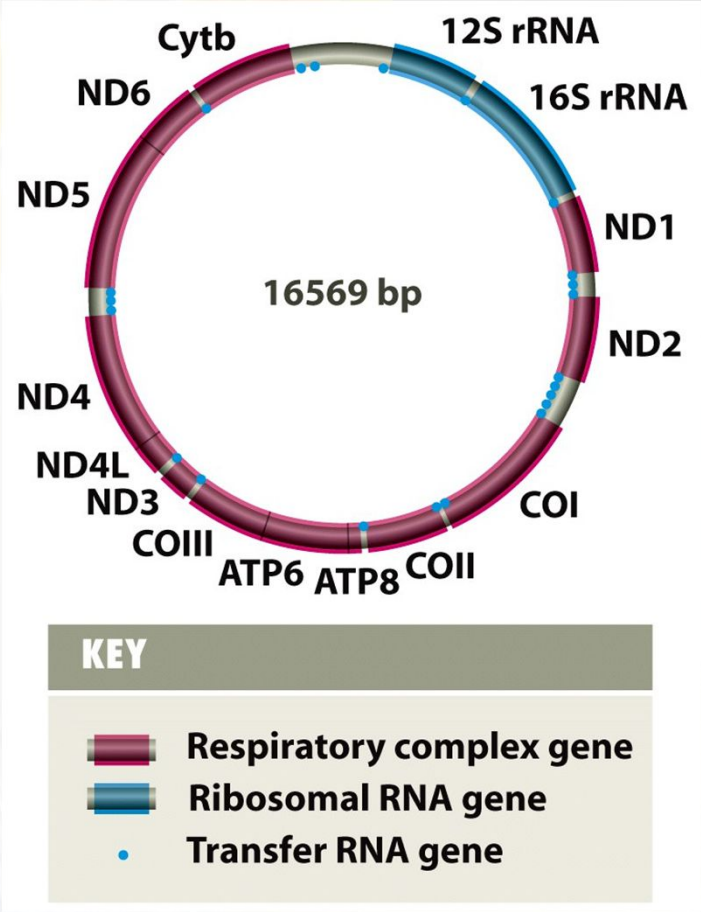
Mitochondrial Genome

Mitochondrial Genome - Humans



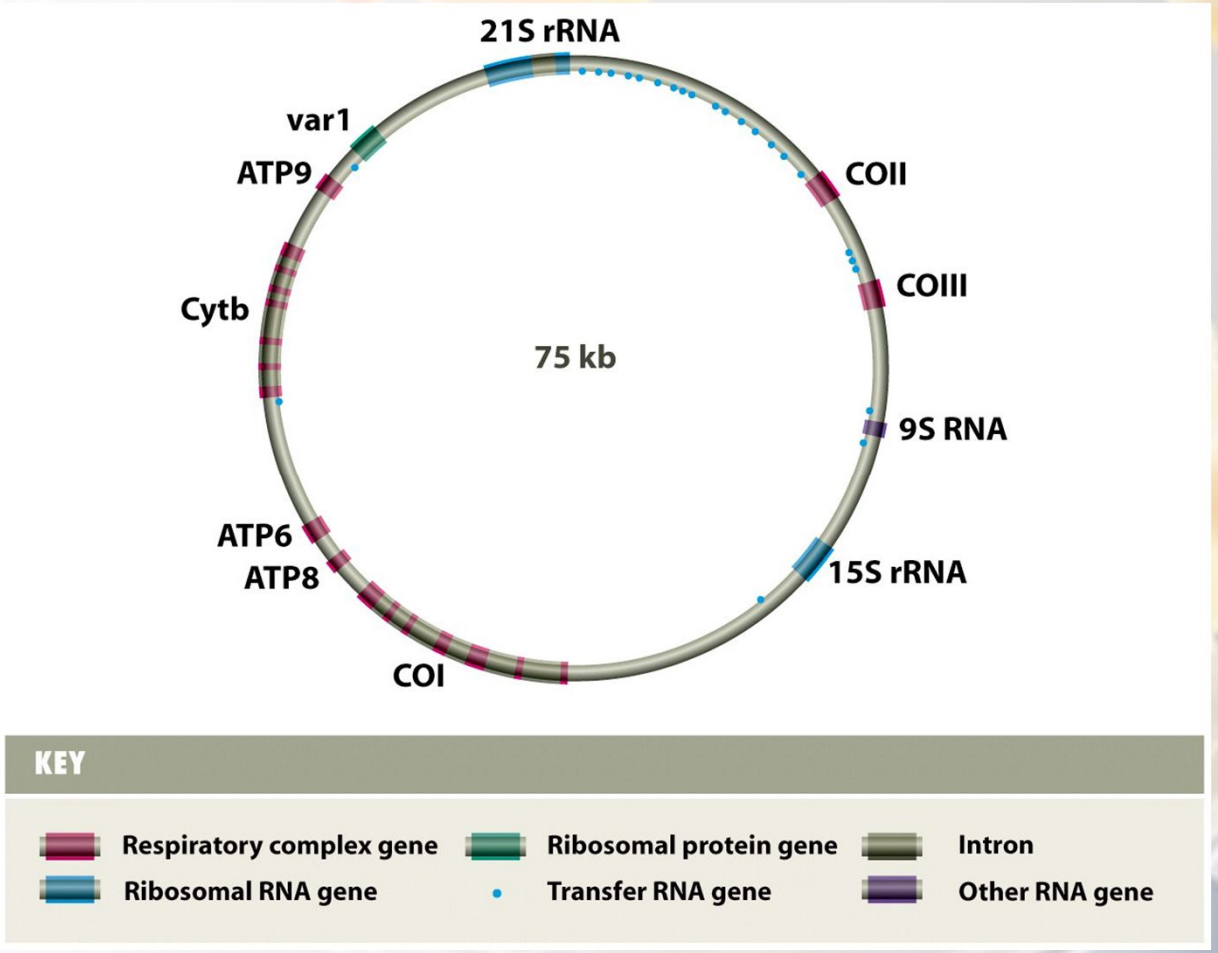
Mitochondrial Genome

Mitochondrial Genome - Humans



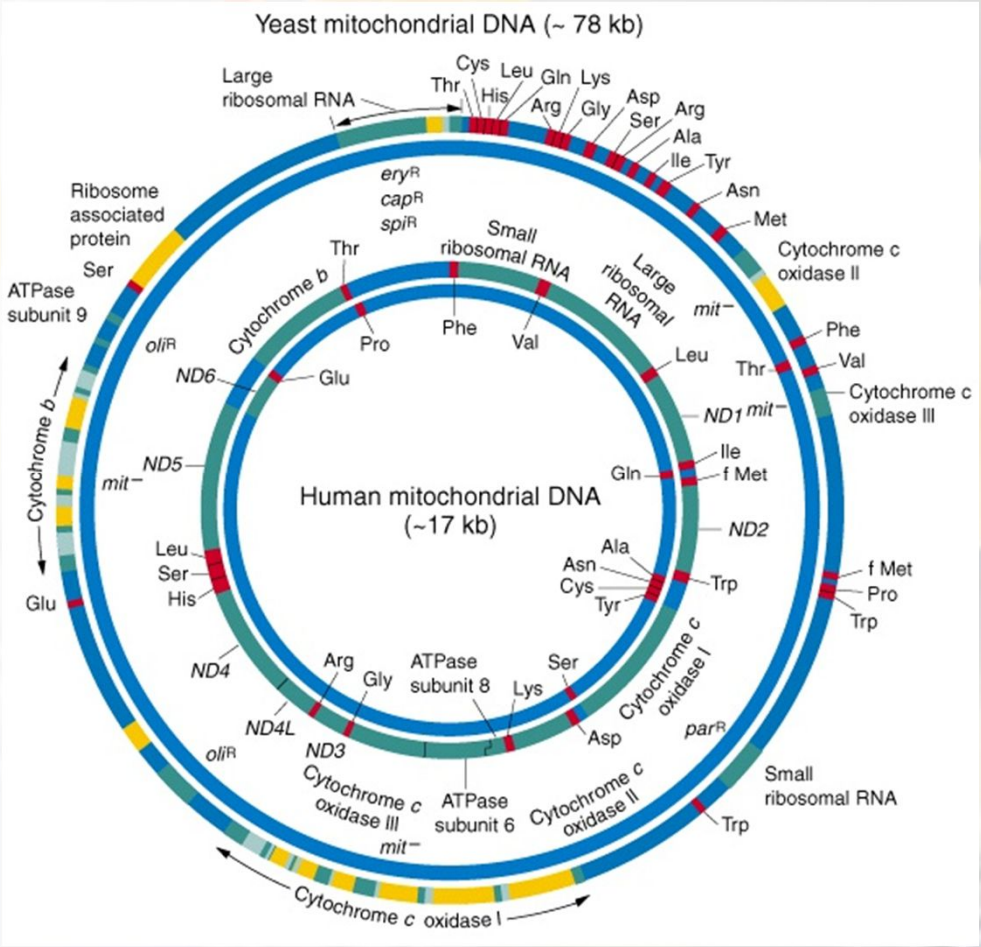
Mitochondrial Genome

Mitochondrial Genome – Yeast



Mitochondrial Genome

Mitochondrial Genome – Human and Yeast



Mitochondrial Genome

Mitochondrial Genomes

Species	Type of organism	Genome size (kb)
Mitochondrial genomes		
Plasmodium falciparum	Protozoan (malaria parasite)	6
Chlamydomonas reinhardtii	Green alga	16
Mus musculus	Vertebrate (mouse)	16
Homo sapiens	Vertebrate (human)	17
Metridium senile	Invertebrate (sea anemone)	17
Drosophila melanogaster	Invertebrate (fruit fly)	19
Chondrus crispus	Red alga	26
Aspergillus nidulans	Ascomycete fungus	33
Reclinomonas americana	Protozoa	69
Saccharomyces cerevisiae	Yeast	75
Suillus grisellus	Basidiomycete fungus	121
Brassica oleracea	Flowering plant (cabbage)	160
Arabidopsis thaliana	Flowering plant (vetch)	367
Zea mays	Flowering plant (maize)	570
Cucumis melo	Flowering plant (melon)	2500

Mitochondrial Genome

Mitochondrial Genomes

Feature	*Plasmodium falciparum*	*Chlamydomonas reinhardtii*	*Homo sapiens*	*Saccharomyces cerevisiae*	*Arabidopsis thaliana*	*Reclinomonas americana*
Total number of genes	5	12	37	35	52	92
Types of genes						
Protein-coding genes	3	7	13	8	27	62
Respiratory complex	3	7	13	7	17	24
Ribosomal proteins	0	0	0	1	7	27
Transport proteins	0	0	0	0	3	6
RNA polymerase	0	0	0	0	0	4
Translation factor	0	0	0	0	0	1
Functional RNA genes	2	5	24	27	25	30
Ribosomal RNA genes	2	2	2	2	3	3
Transfer RNA genes	0	3	22	24	22	26
Other RNA genes	0	0	0	1	0	1
Number of introns	0	1	0	8	23	1
Genome size (kb)	6	16	17	75	367	69

Mitochondrial Genome

Universal Code and Mitochondrial Code

Mitochondrial code

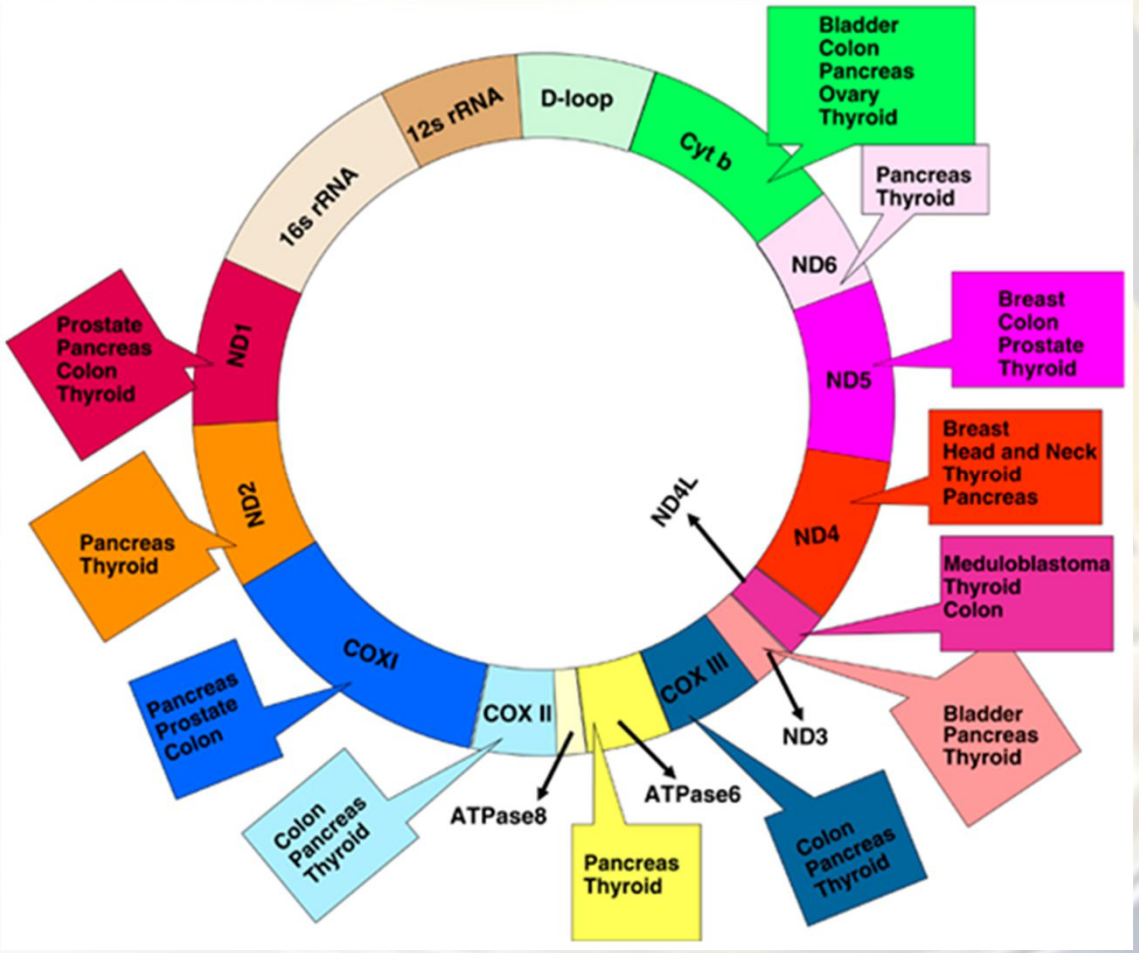
First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	*Stop*	(*Stop*) Trp	A
	Leu	Ser	*Stop*	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile (Met)	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	(Ile) Met	Thr	Lys	(Arg) *Stop*	A
	Ile	Thr	Lys	(Arg) *Stop*	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Universal code

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

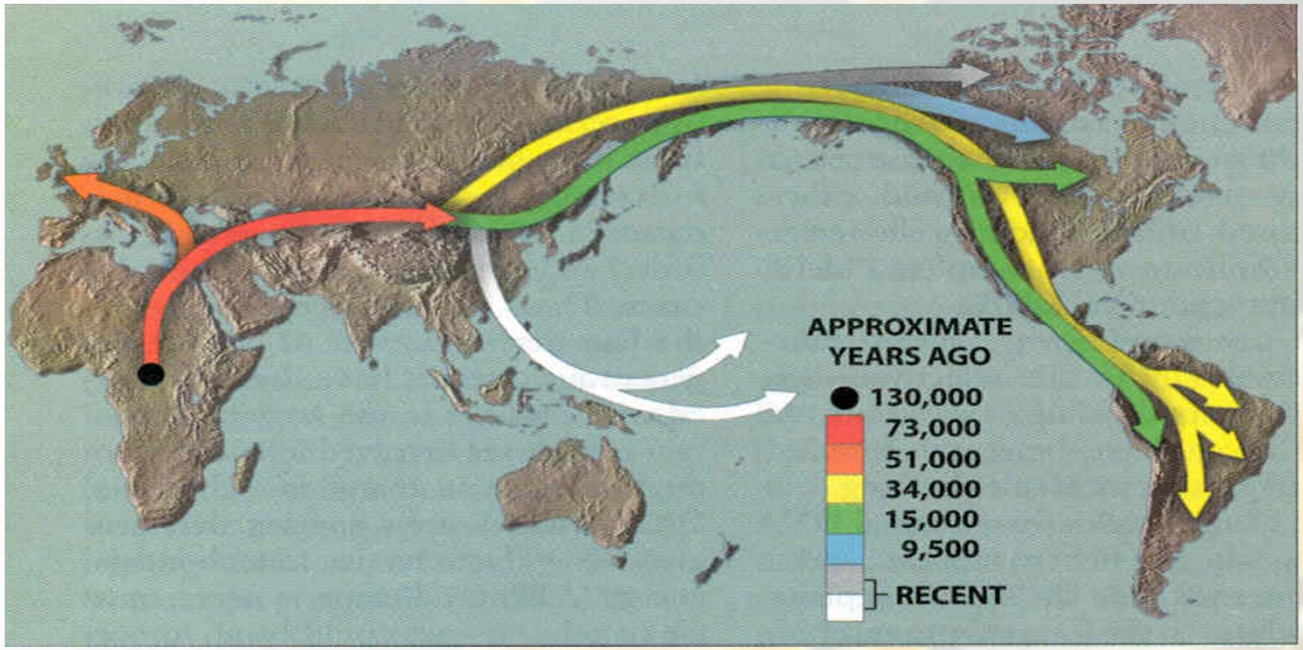
Mitochondrial Genome

Mitochondrial Genome – Tumors due to mutations



Mitochondrial Genome

Mitochondrial DNA polymorphisms track human migrations



All humans descend from a small group of Africans
This group originated in central Africa ~200,000 years ago
The founding group was small (10^2-10^4 people)

Mitochondrial Genome

Conclusion

- Multiple identical circular chromosomes
- 15-16 Kb in animals
- 200 kb to 2,500 kb in plants

Chloroplast Genome

Chloroplast Genome

Chloroplast Genome

- Multiple circular molecules
- Size ranges from 70 kb to 2000 kb
- Land plants typically 120 – 170 kb
- ~ 70 kb – Epifagus
- ~2,000 kb – Acetabularia

Chloroplast Genome

Chloroplast Genome

- Similar to mtDNA
- Many chloroplast proteins are encoded in the nucleus (separate signal sequence)

Chloroplast Genome

Chloroplast Genome

- Double stranded DNA molecule
- Chloroplasts genomes are relatively larger
- Multiple copies of genome

Chloroplast Genome

Chloroplast Genome

- Large enough to code 50-100 proteins as well as rRNAs & tRNAs
- cpDNA regions includes Large Single Copy & Small Single-Copy (SSC) regions, and Inverted Repeats (IRA & IRB).

Chloroplast Genome

Chloroplast Genomes: Size in different Taxa

Taxa	Genome size (In kb)	Inverse duplication (In kb)
Angiospermae		
Nicotiana tabacum	156	25
Spinacia oleracea	150	24
Pelargonium hortorum	217	76
Pisum sativum	120	Not present
Epifragus virginiana	70	22
Oryza sativa	134	21
Gymnospermae		
Pinus	120	Not present
Ginkgo biloba	158	17
Pteridophyta		
Osmunda cinnamomea	144	10
Bryophyta		
Marchantia polymorpha	121	10
Chlorophyta		
Codium fragile	85	Not present
Chlamydomonas reinhardtii	195	22
Chlamydomonas moewusii	292	41
Rhodophyta		
Cyanophora paradoxa	127	10
Chromophyta		
Dictyota dichotoma	123	5

Chloroplast Genome

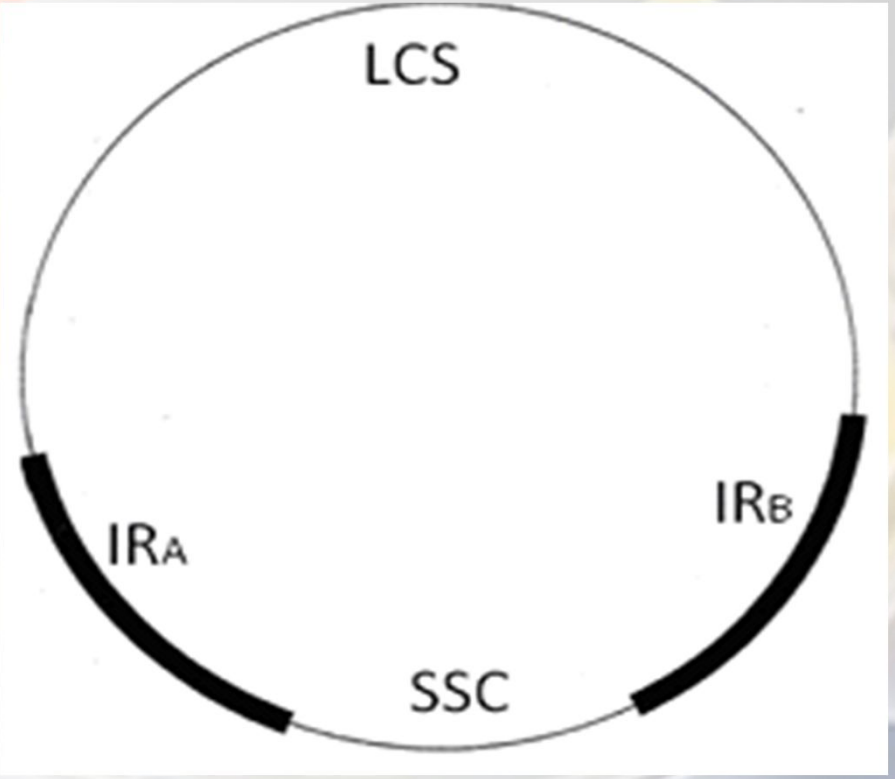
Chloroplast Genome Size

Species	Type of organism	Genome size (kb)
Chloroplast genomes		
Pisum sativum	Flowering plant (pea)	120
Marchantia polymorpha	Liverwort	121
Oryza sativa	Flowering plant (rice)	136
Nicotiana tabacum	Flowering plant (tobacco)	156
Chlamydomonas reinhardtii	Green alga	195

Chloroplast Genome

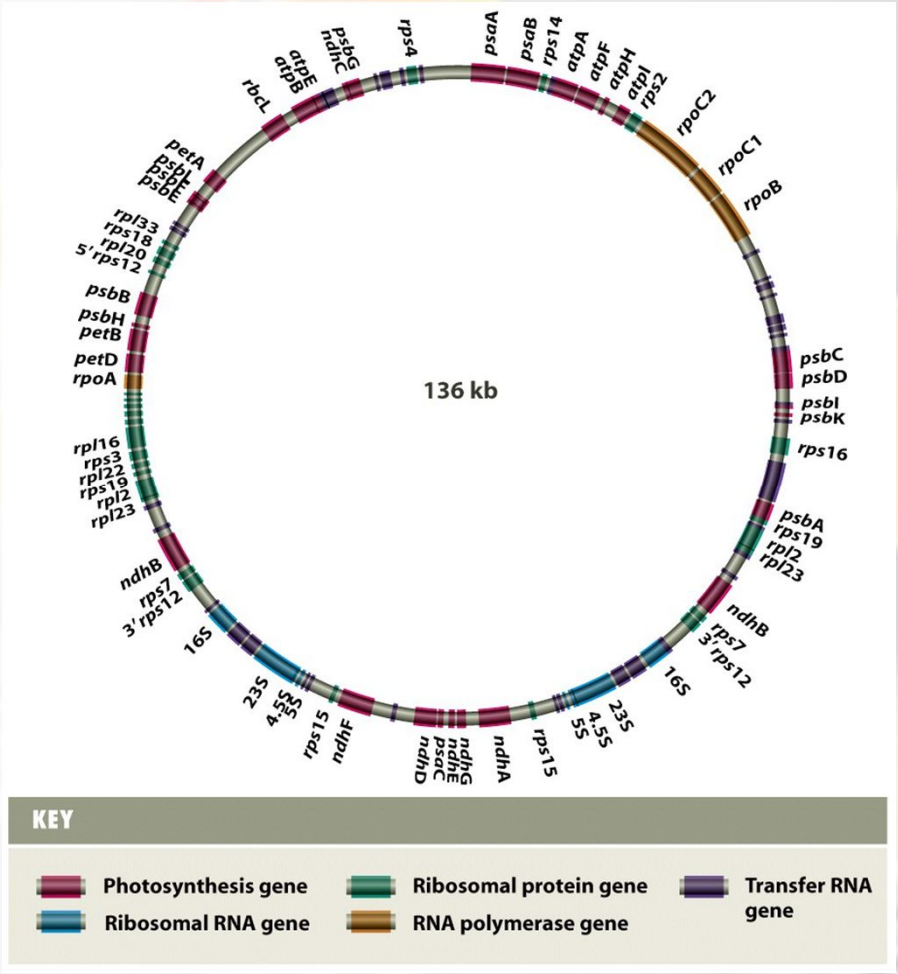
Chloroplast Genome:

- Large Single Copy Region
- Small Single Copy Region
- Inverted Repeats



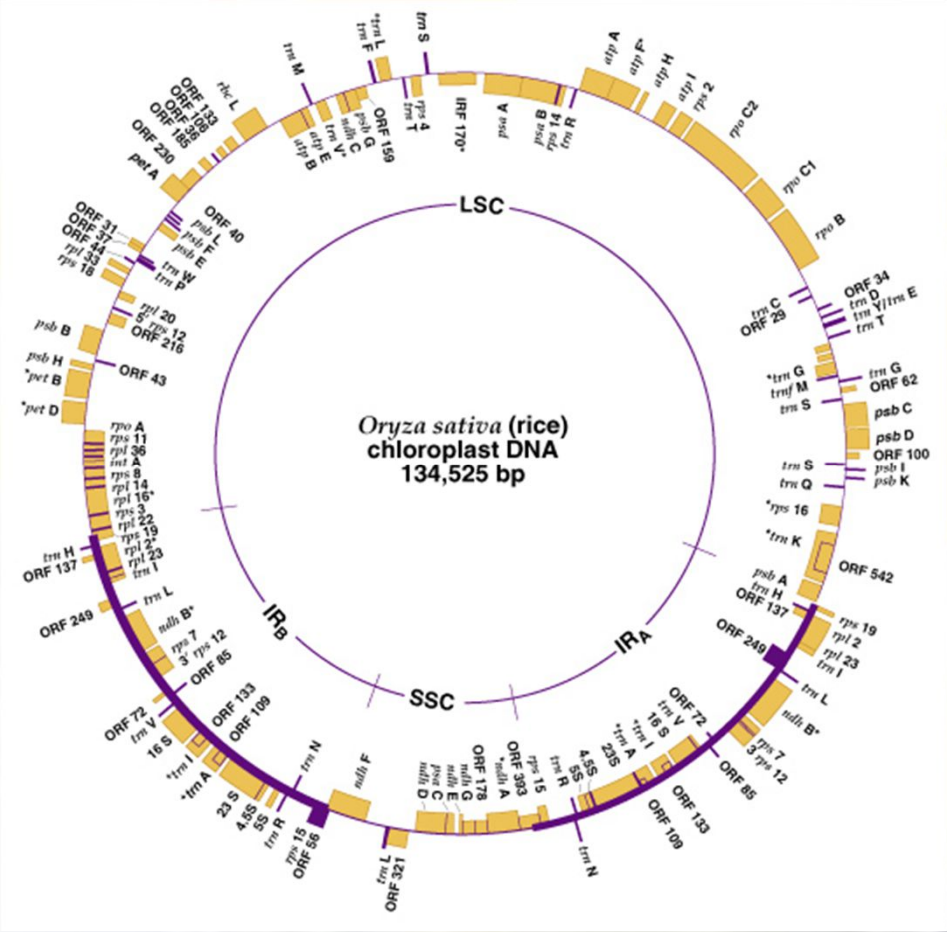
Chloroplast Genome

Chloroplast Genome of Rice



Chloroplast Genome

Chloroplast Genome of Rice



Chloroplast Genome

Functions of Chloroplast Genes

- Most cp genes fall into two functional groups:
- Genes involved in replication, transcription, translation
- Genes involved in photosynthesis

Chloroplast Genome

Genes Nomenclature

- Based on bacterial naming system, which uses lower case letters, and a descriptive prefix, based on the probable function. If the gene product is part of a multi-subunit complex, a letter of the alphabet is used to denote different subunits.
- psa for genes of photosystem I (psaA, psaB, etc.)
- psb for genes of photosystem II (psbA, psbB, etc.)

Chloroplast Genome

Properties of Chloroplast Genome

- Non-mendelian inheritance
- Self replication
- Somatic segregation in plants
- Inherited independently of nuclear genes

Chloroplast Genome

Properties of Chloroplast Genome

- Conservative rate of nucleotide substitution enables to resolve plant phylogenetic relationships at deep levels of evolution.

Chloroplast Genome

Chloroplast Genome - Conclusion

- Multiple circular molecules
- Size ranges from 70 kb to 2000 kb

Genetics and Genomics

Eukaryotic Genomes

Eukaryotic Genomes

Eukaryotic Genomes

- Located on several chromosomes
- Relatively low gene density
- Carry organelles genome in addition to nuclear genome

Eukaryotic Genomes

Eukaryotic Genomes

- Contains repetitive sequences
- SINEs (short interspersed elements)
- LINEs (long interspersed elements)

Eukaryotic Genomes

LINES and SINES

- LINES
- 1-5 kb
- 10-10,000 copies

- SINES
- 200-300 bp
- 100,000 copies

Eukaryotic Genomes

Highly repetitive

- Minisatellites
- Microsatellites
- Telomeres

Eukaryotic Genomes

Minisatellites

- Repeats of 14-500 bp
- 1-5 kb long
- Scattered throughout the genome

Eukaryotic Genomes

Microsatellites

- Repeats up to 13 bp

Eukaryotic Genomes

Telemeres

- Short repeats (6 bp)
- 250-1,000 at ends of chromosomes

Eukaryotic Genomes

Elements of Eukaryotic Genomes

Chromosomes: linear, centromeres, telomeres, origins of replication, replicons

Protein-coding genes and **spliceosomal introns**

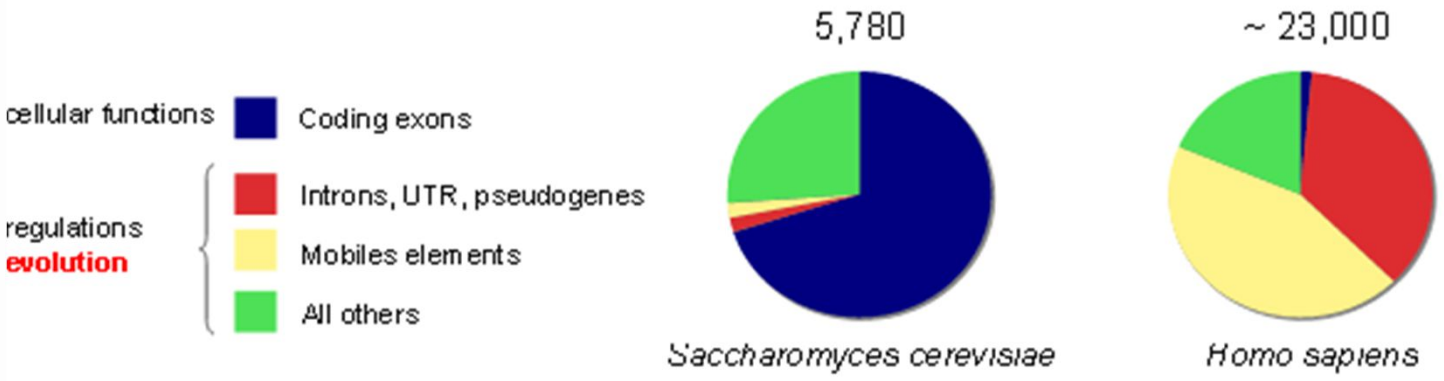
Genes for non coding RNAs: rRNAs, tRNAs, snoRNAs, snRNAs, microRNAs

Mobile genetic elements: and their remnants

Pseudogenes: and processed pseudogenes

Satellite DNAs: micro-, minisatellites, repeated sequences

Fragments of organellar DNAs: NUMTs and NUPTs



Eukaryotic Genomes

Comparison - Prokaryotic & Eukaryotic Genomes

Prokaryotic	Eukaryotic
• Usually circular • Smaller • Found in the nucleoid region • Less elaborately structured and folded	• Complex with a large amount of protein to form chromatin • Highly extended and tangled during interphase • Found in the nucleus

Eukaryotic Genomes

Conclusion

- Located on several chromosomes
- Relatively low gene density
- Chromosomes vary in length
- Carry organelles genome

Genetics and Genomics

Genomes
Comparisons

Genomes Comparisons

Genomes vary in size

- Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb)
- Most plants and animals have genomes greater than 100 Mb; humans have 3,000 Mb

Genomes Comparisons

Genomes vary in genes numbers

- Free-living bacteria and archaea have 1,500 to 7,500 genes
- Fungi have about 5,000 genes and multicellular eukaryotes upto 40,000 genes

Genomes Comparisons

Genomes vary in genes numbers

- Number of genes is not correlated to genome size
- Nematode *C. elegans* has 100 Mb and 20,000 genes, while *Drosophila* has 165 Mb and 13,700 genes

Genomes Comparisons

Genomes vary in genes numbers

- Vertebrate genomes can produce more than one polypeptide per gene because of alternative splicing of RNA transcripts

Genomes Comparisons

Humans and Mammals have low gene density

- Humans and other mammals have the lowest gene density, or number of genes, in a given length of DNA. Multicellular eukaryotes have many introns within genes and noncoding DNA between genes

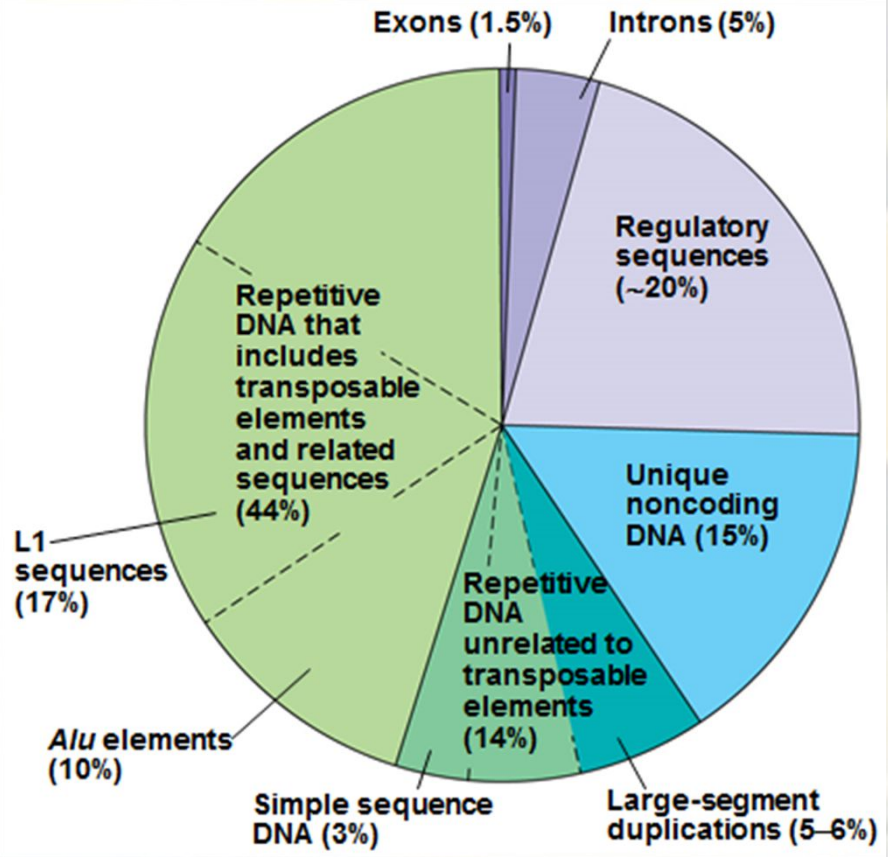
Genomes Comparisons

Multicellular eukaryotes have much noncoding DNA and multigene families

- Most of eukaryotic genomes neither encodes proteins nor functional RNAs
- Evidence indicates that noncoding DNA plays important roles in the cell

Genomes Comparisons

Human Genome: Distribution of coding and non-coding DNA



Genomes Comparisons

Comparing Genomes

- Significant similarity between genomes of ”distant” species (Man – Yeast 23%)
- Similarity increases for taxonomically close species.

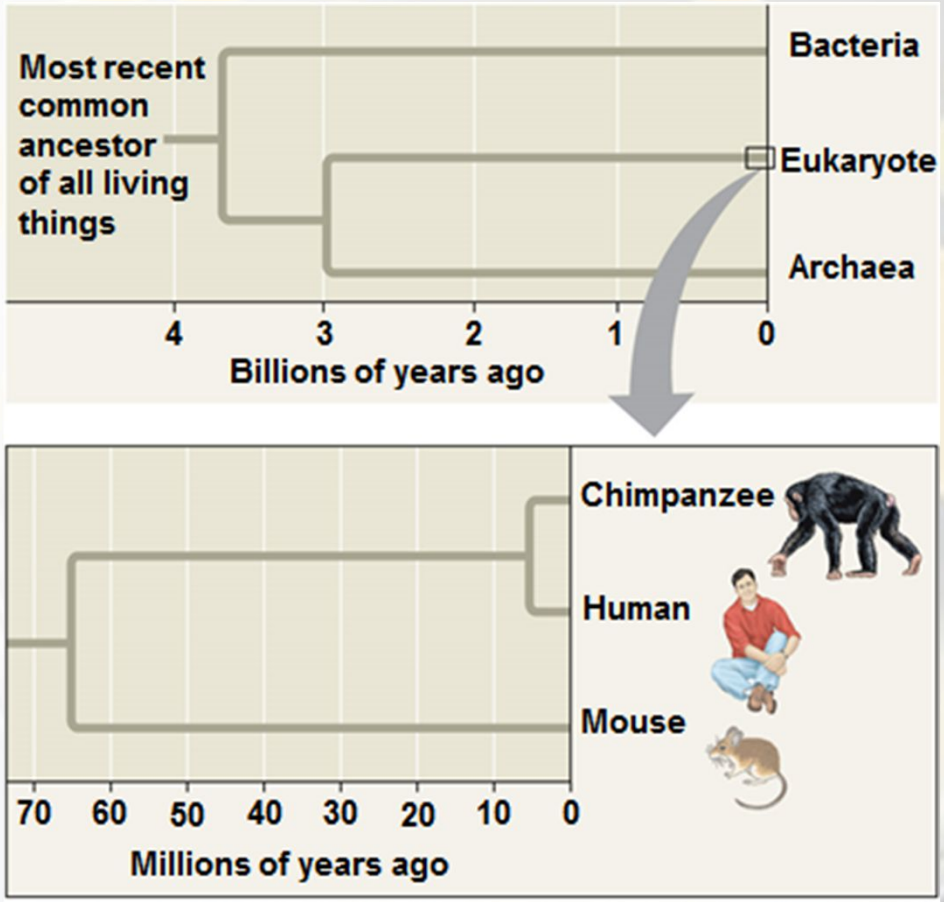
Genomes Comparisons

Comparing Genomes

- Closely related species help us understand recent evolutionary events
- Distantly related species help us understand ancient evolutionary events

Genomes Comparisons

Comparing Genomes: Bacteria, archaea, and eukaryotes diverged from each other between 2 and 4 billion years ago



Genomes Comparisons

Comparing Genomes

	Bacteria	Archaea	Eukaryotes
Genome size	Most are 1–6 Mb		Most are 10–4,000 Mb, but a few are much larger
Number of genes	1,500–7,500		5,000–40,000
Gene density	Higher than in eukaryotes		Lower than in prokaryotes (Within eukaryotes, lower density is correlated with larger genomes.)
Introns	None in protein-coding genes	Present in some genes	Unicellular eukaryotes: present, but prevalent only in some species Multicellular eukaryotes: present in most genes
Other noncoding DNA	Very little		Can be large amounts; generally more repetitive noncoding DNA in multicellular eukaryotes

Genomes Comparisons

Comparing Genomes

- Human and chimpanzee genomes differ by 1.2%, at single base-pairs, and by 2.7% because of insertions and deletions

Genetics and Genomics

Comparing distantly/closely related species

Distantly/Closely Related Species

Comparing distantly related species

- Highly conserved genes have changed very little over time
- These help to clarify relationships among species that diverged from each other long ago

Distantly/Closely Related Species

Comparing distantly related species

- Bacteria, archaea, and eukaryotes diverged from each other 2 and 4 billion years ago
- Highly conserved genes can be studied in one model organism.

Distantly/Closely Related Species

Comparing closely related species

- Genetic differences between closely related species can be correlated with phenotypic differences

Distantly/Closely Related Species

Comparing closely related species

- Genetic comparison of several mammals with non-mammals helps to identify what make mammals

Distantly/Closely Related Species

Comparing closely related species

- Human and chimpanzee genomes differ by 1.2%, at single base-pairs, and by 2.7% because of insertions and deletions

Distantly/Closely Related Species

Comparing closely related species

- Several genes are evolving faster in humans than chimpanzees

Distantly/Closely Related Species

Comparing closely related species

- Genes involved in defense against malaria and tuberculosis and in regulation of brain size, genes code for transcription factors

Distantly/Closely Related Species

Comparing closely related species

- Humans and chimpanzees differ in the expression of the *FOXP2* gene, whose product turns on genes involved in vocalization

Distantly/Closely Related Species

Comparing closely related species

- Differences in the *FOXP2* gene may explain why humans but not chimpanzees communicate by speech

Distantly/Closely Related Species

Conclusion

- Highly conserved genes have changed very little over the time
- These help to clarify relationships among species that diverged from each other long ago

Anatomy and Organization of Genomes

Genomes Anatomy/Organization

Genome Anatomy

- Anatomy of different genomes differ from each other.
- Eukaryotes and prokaryotes genomes differ very significantly.

Genomes Anatomy/Organization

Genome Anatomy

- Size of genomes - 1000 fold difference between eukaryotes and prokaryotes.
- ~ 30 fold between genomes of different eukaryotes.

Genomes Anatomy/Organization

Genome Anatomy

- In humans ~ 23,000
- Bacterial genomes ~ 1,500 – 2,000 genes.

Genomes Anatomy/Organization

Genome Anatomy - Eukaryotes

- Eukaryotic genomes are full of simple repeats, numerous types of transposable elements and other sequences.

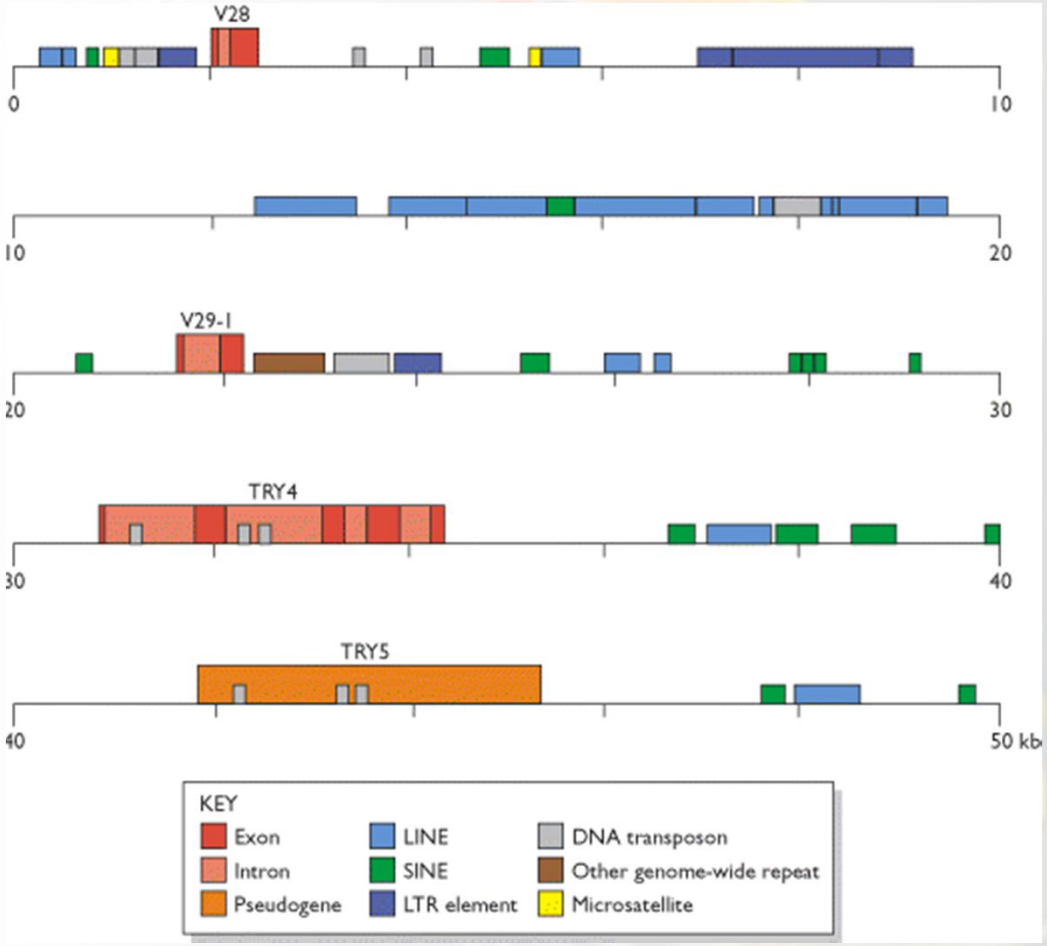
Genomes Anatomy/Organization

Genome Anatomy - Prokaryotes

- Prokaryotes have a few repeats and transposable elements and their genomes consist of mainly the genes.

Genomes Anatomy/Organization

Genome Anatomy: Segment of human chromosome 7



1 Gene: TRY4

2 Gene Segments: V28 & V29-1

1 Pseudogene: TRY5

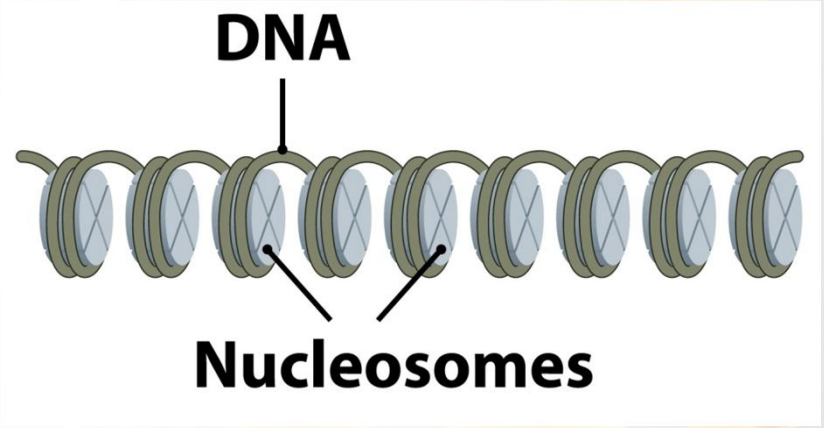
52 genome wide repeat sequences:

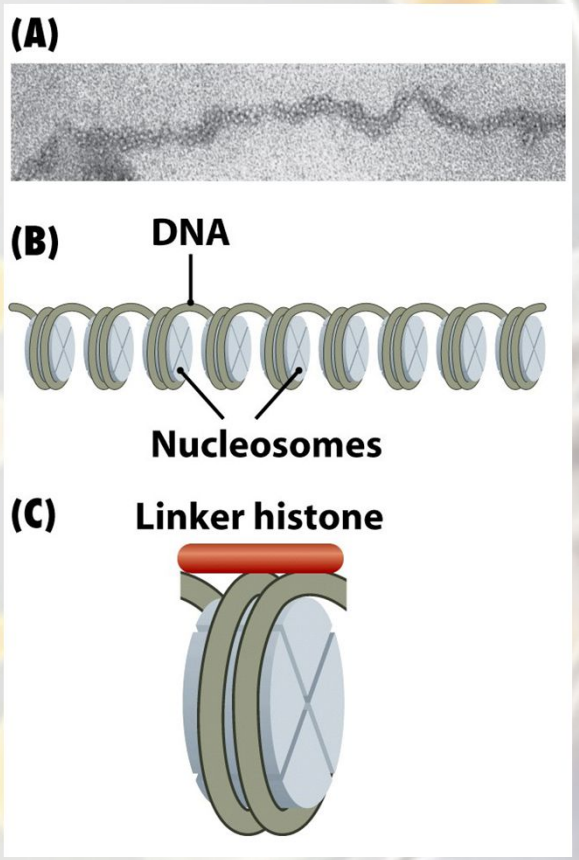
LINE, SINE, LTR, & DNA transposon.

Two Microsatellites

Genomes Anatomy/Organization

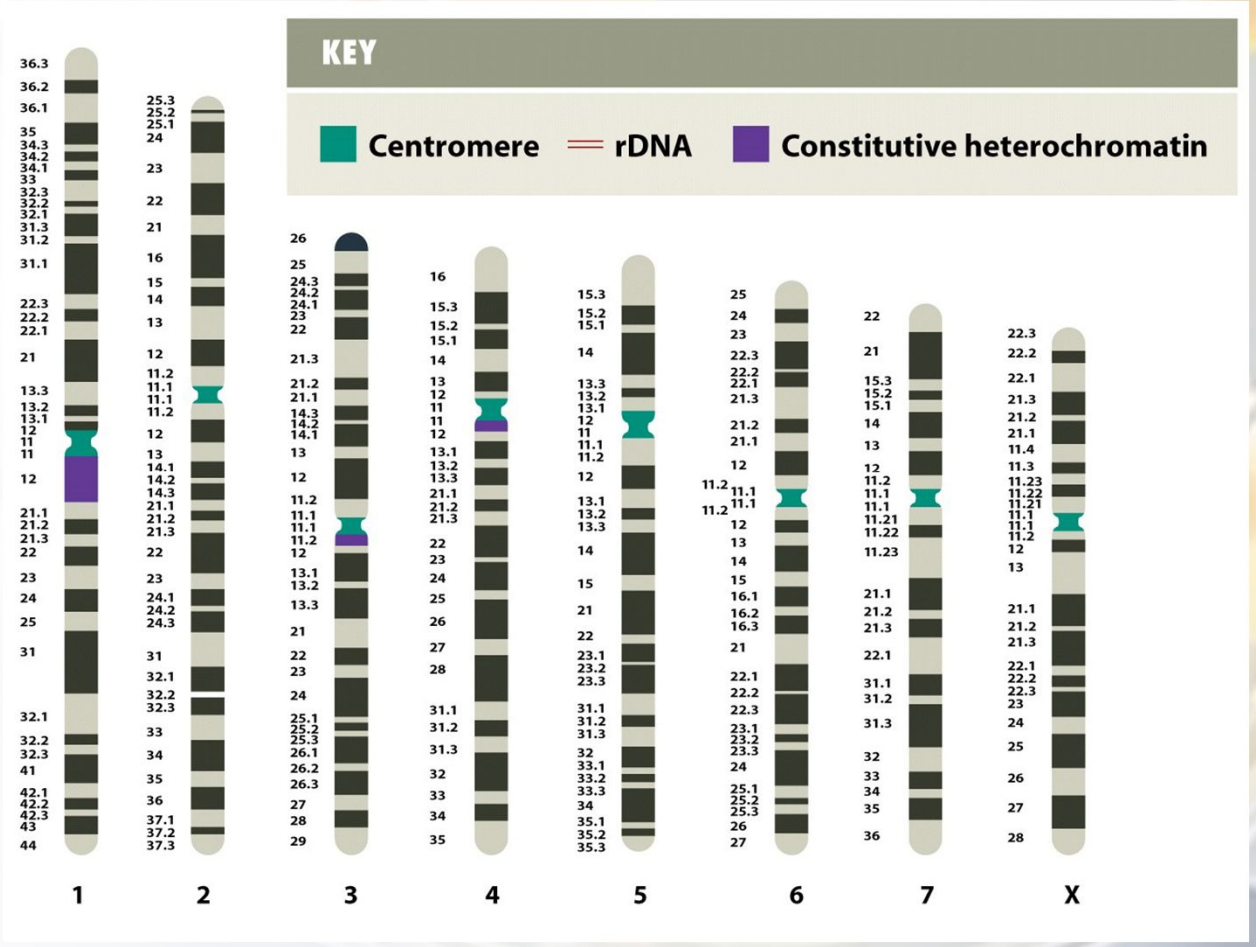
Genome Organization





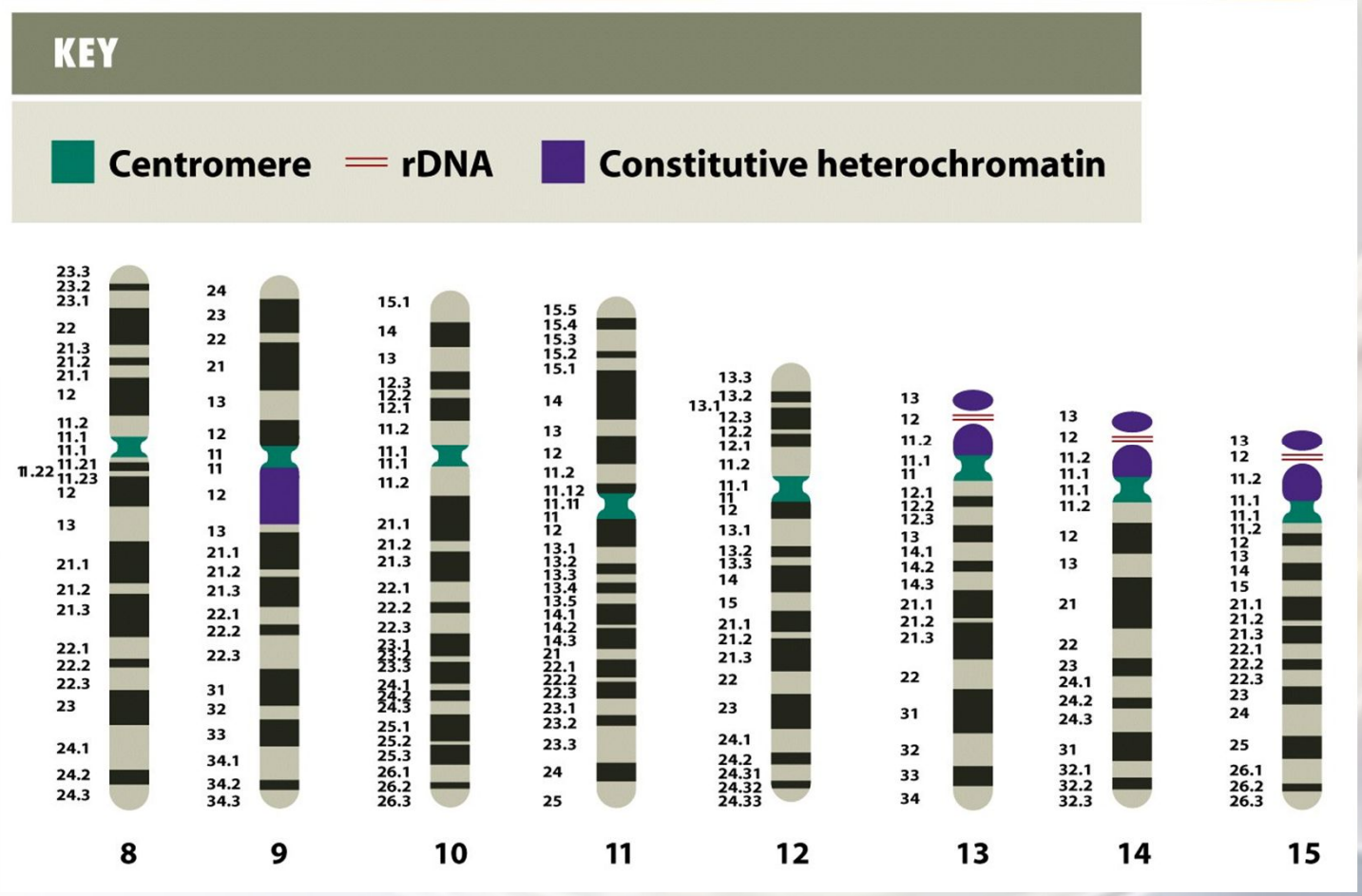
Genomes Anatomy/Organization

Genome Organization



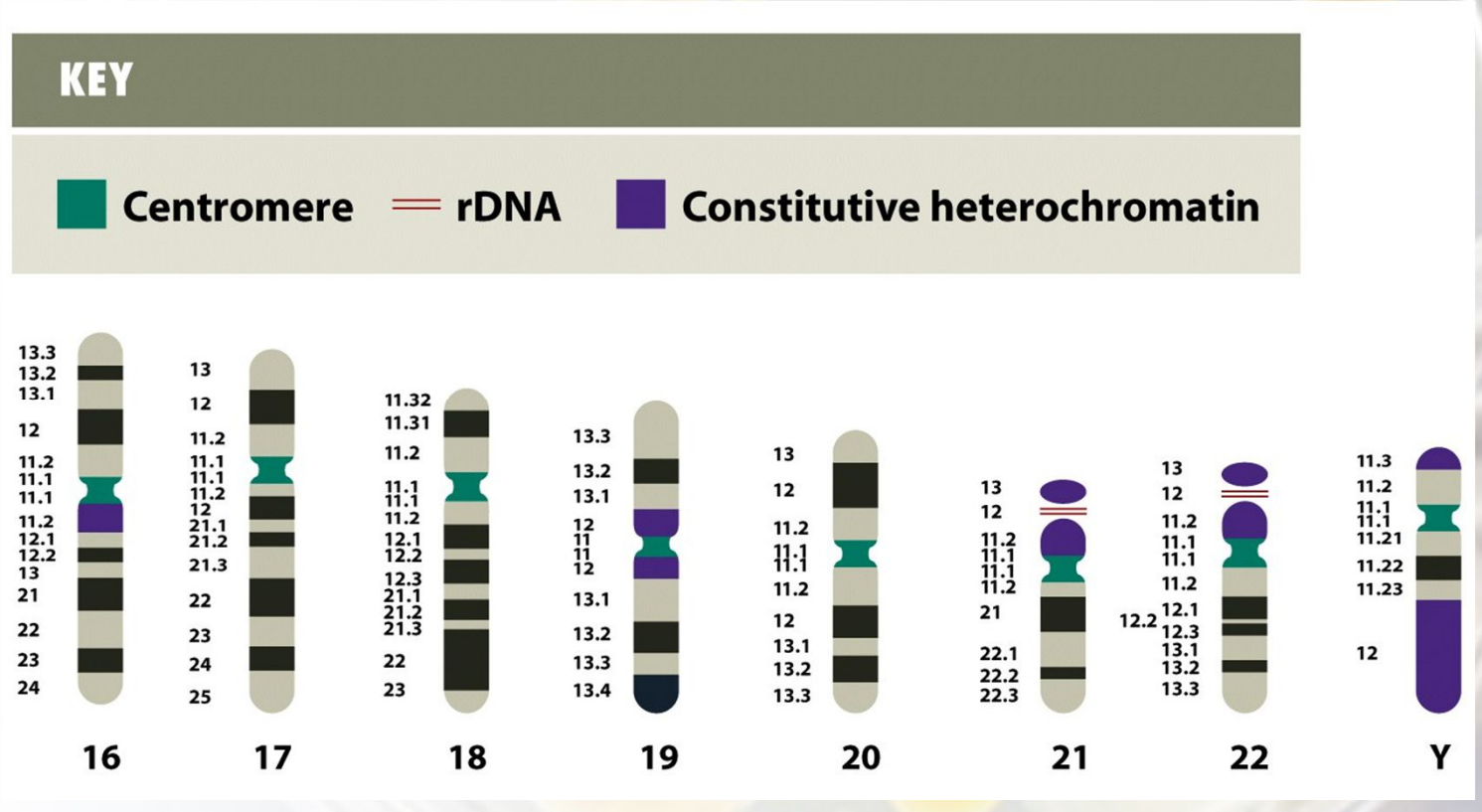
Genomes Anatomy/Organization

Genome Organization



Genomes Anatomy/Organization

Genome Organization



Genomes Anatomy/Organization

Genome Organization in Prokaryotes

Species	DNA molecules	Genome organization	
		Size (Mb)	Number of genes
Escherichia coli K12	One circular molecule	4.639	4405
Vibrio cholerae El Tor N16961	Two circular molecules		
	Main chromosome	2.961	2770
	Megaplasmid	1.073	1115
Deinococcus radiodurans R1	Four circular molecules		
	Chromosome 1	2.649	2633
	Chromosome 2	0.412	369
	Megaplasmid	0.177	145
	Plasmid	0.046	40

Genomes Anatomy/Organization

Genome Organization in Prokaryotes

Species	Size of genome (Mb)	Approximate number of genes
Bacteria		
Mycoplasma genitalium	0.58	500
Streptococcus pneumoniae	2.16	2300
Vibrio cholerae El Tor N16961	4.03	4000
Mycobacterium tuberculosis H37Rv	4.41	4000
Escherichia coli K12	4.64	4400
Yersinia pestis CO92	4.65	4100
Pseudomonas aeruginosa PA01	6.26	5700
Archaea		
Methanococcus jannaschii	1.66	1750
Archaeoglobus fulgidus	2.18	2500

Genomes Anatomy/Organization

Genome Organization: Comparisons

Species	Genome size (Mb)
Fungi	
Saccharomyces cerevisiae	12.1
Aspergillus nidulans	25.4
Protozoa	
Tetrahymena pyriformis	190
Invertebrates	
Caenorhabditis elegans	97
Drosophila melanogaster	180
Bombyx mori (silkworm)	490
Strongylocentrotus purpuratus (sea urchin)	845
Locusta migratoria (locust)	5000

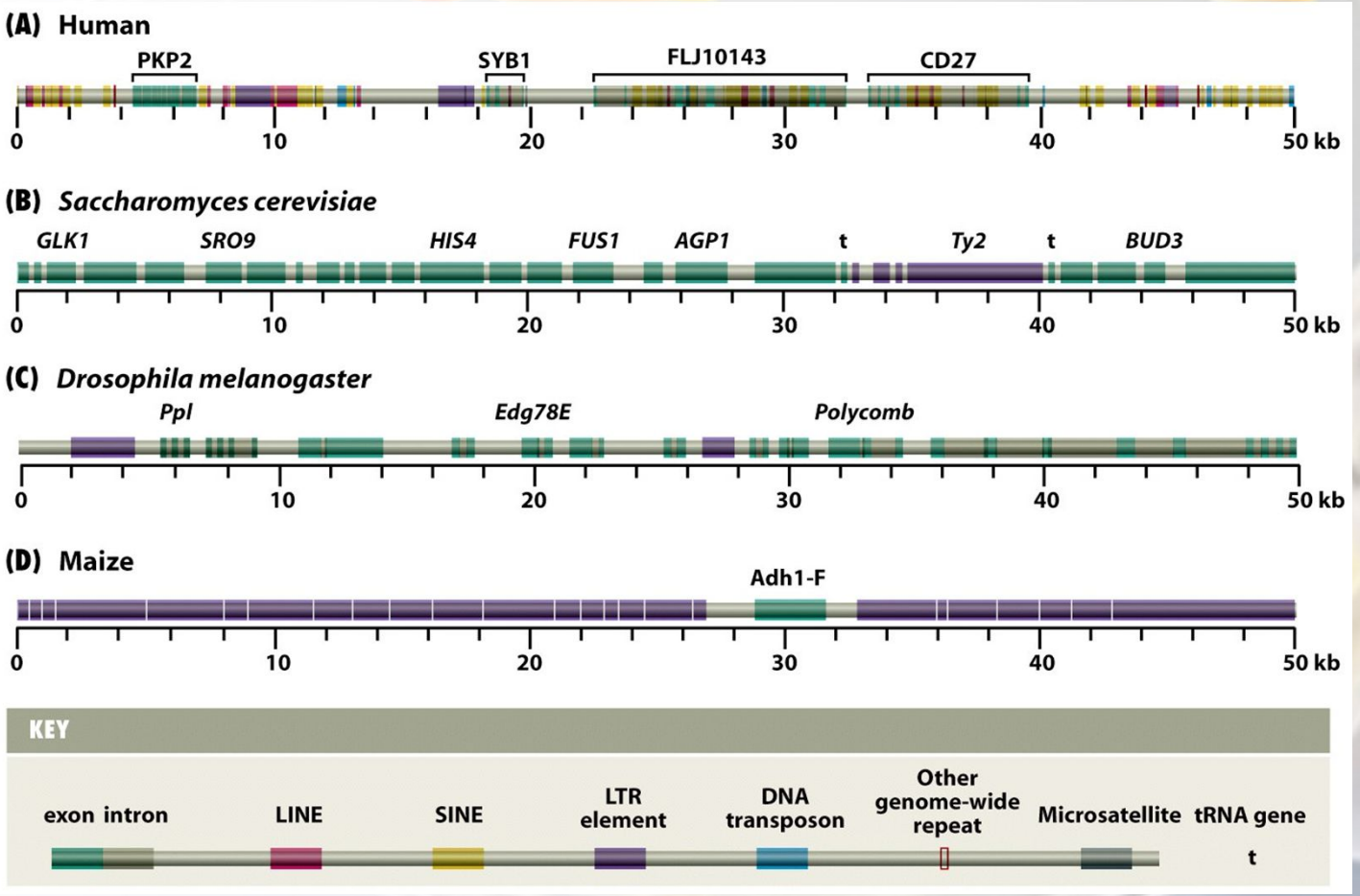
Genomes Anatomy/Organization

Genome Organization: Comparisons

Species	Genome size (Mb)
Vertebrates	
Takifugu rubripes (pufferfish)	400
Homo sapiens	3200
Mus musculus (mouse)	3300
Plants	
Arabidopsis thaliana (vetch)	125
Oryza sativa (rice)	466
Zea mays (maize)	2500
Pisum sativum (pea)	4800
Triticum aestivum (wheat)	16,000
Fritillaria assyriaca (fritillary)	120,000

Genomes Anatomy/Organization

Genome Organization: Human, Yeast, Fruit Fly , Maize



Genomes Anatomy/Organization

Compactness of Genomes

Feature	Yeast	Fruit fly	Human
Gene density (average number per Mb)	496	76	11
Introns per gene (average)	0.04	3	9
Amount of the genome that is taken up by genome-wide repeats	3.4%	12%	44%

Genomes Anatomy/Organization

Conclusion

- Anatomy of different genomes differ from each other.
- Eukaryotes and prokaryotes genomes differ very significantly.

Genetics and Genomics

Gene Anatomy

Gene Anatomy

What is Gene

- A piece of DNA (or RNA) that contains the primary sequence to produce a functional biological gene product (RNA or protein).

Gene Anatomy

What is Gene

- Entire nucleic acid sequence necessary for the synthesis of a functional polypeptide (protein chain) or functional RNA

Gene Anatomy

Genetic information is stored in DNA

- Segments of DNA that encode proteins or other functional products are called genes.
- Gene sequences are transcribed into messenger RNA (mRNA).

Gene Anatomy

mRNA is translated into Proteins

- mRNA intermediates are translated into proteins that perform most of the life functions.

Gene Anatomy

Gene Anatomy

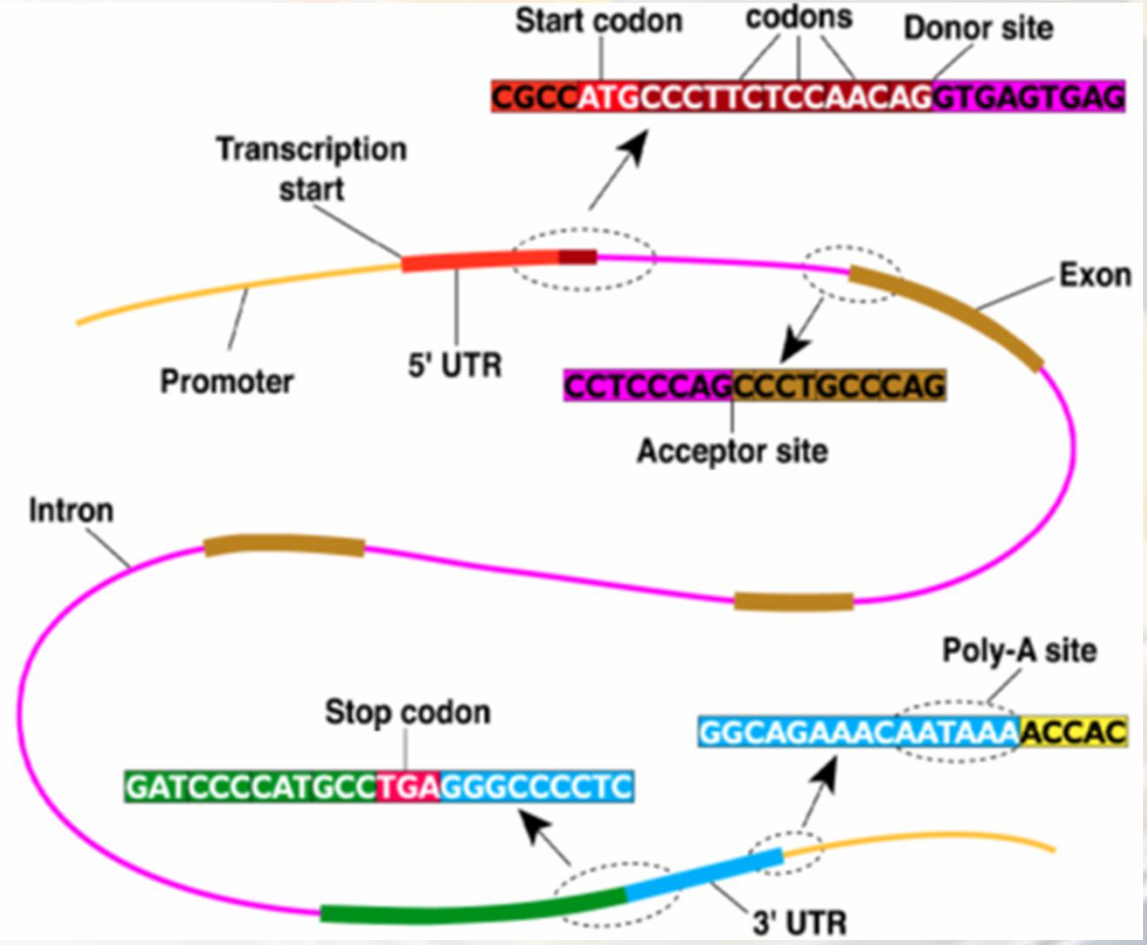
- Three components
- Open Reading Frame: From start codon (ATG) to stop (TGA, TAA, TAG)
- Upstream region with binding site. (e.g. TATA box, GC box, CAAT box).
- Poly-a tail

Gene Anatomy

Gene Anatomy – Typical Prokaryotic Gene

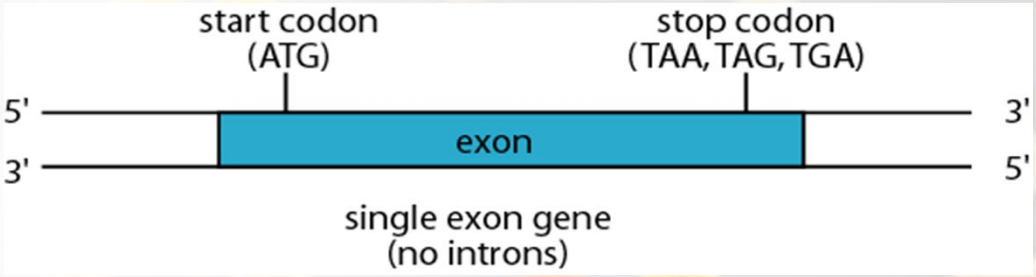
Gene Anatomy

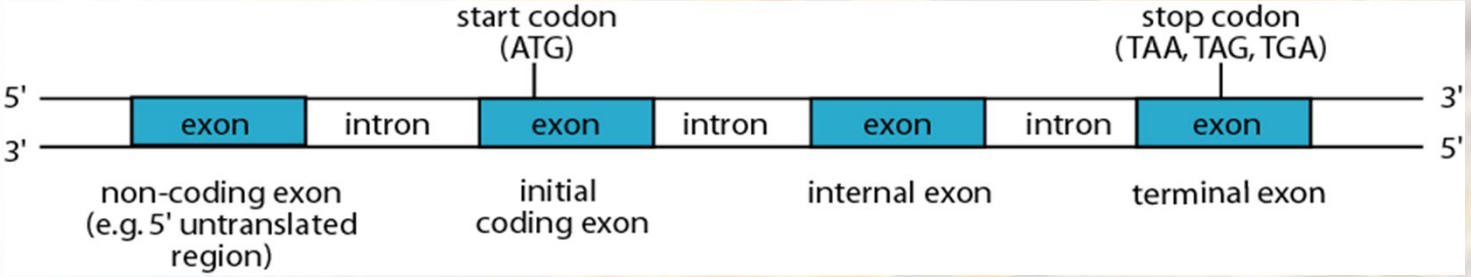
Gene Anatomy – Typical Eukaryotic Gene



Gene Anatomy

Single Exon Gene and Multiple Exons Gene





Gene Anatomy

What is Gene

- A piece of DNA (or RNA) that contains the primary sequence to produce a functional biological gene product (RNA or protein).

Genetics and Genomics

Prokaryotic Gene and Eukaryotic Gene

Prokaryotic Gene/Eukaryotic Gene

Bacterial Gene

- Most do not have introns
- Many are organized in operons: contiguous genes, transcribed as a single polycistronic mRNA, that encode proteins with related functions

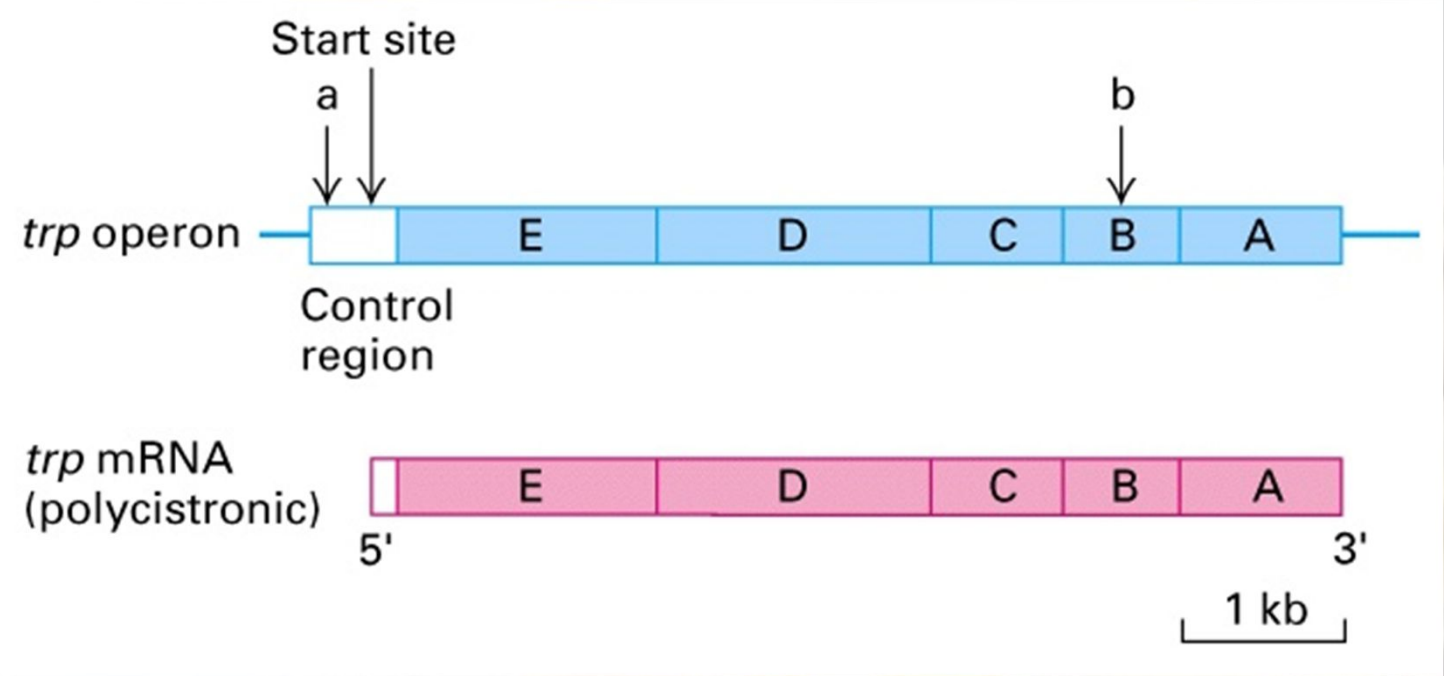
Prokaryotic Gene/Eukaryotic Gene

Bacterial Gene

- Polycistronic mRNA encodes several proteins

Prokaryotic Gene/Eukaryotic Gene

Bacterial Gene: Polycistronic mRNA encodes several proteins



Prokaryotic Gene/Eukaryotic Gene

Eukaryotic Gene: Exons and Introns

- Introns: intervening sequences within a gene that are not translated into a protein sequence.
- Exons: sequences within a gene that encode protein sequence.

Prokaryotic Gene/Eukaryotic Gene

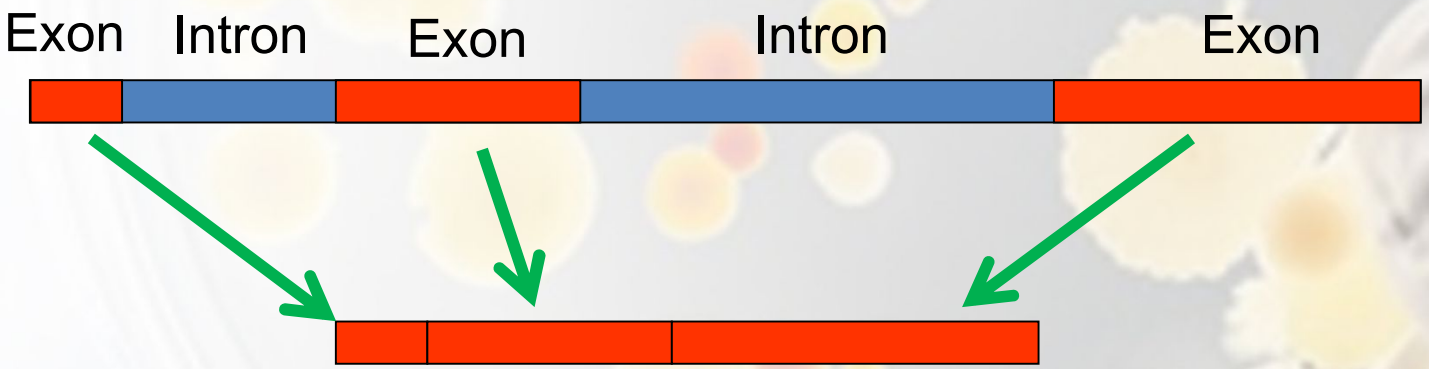
Eukaryotic Gene

- Splicing: Removal of introns from the mRNA molecule

Prokaryotic Gene/Eukaryotic Gene

Eukaryotic Gene

- Splicing: Removal of introns from the mRNA molecule



Prokaryotic Gene/Eukaryotic Gene

Eukaryotic Gene

- Organize expression of genes' (function calls)
- Promoter region (binding site), usually near coding region
- Binding can block (inhibit) expression

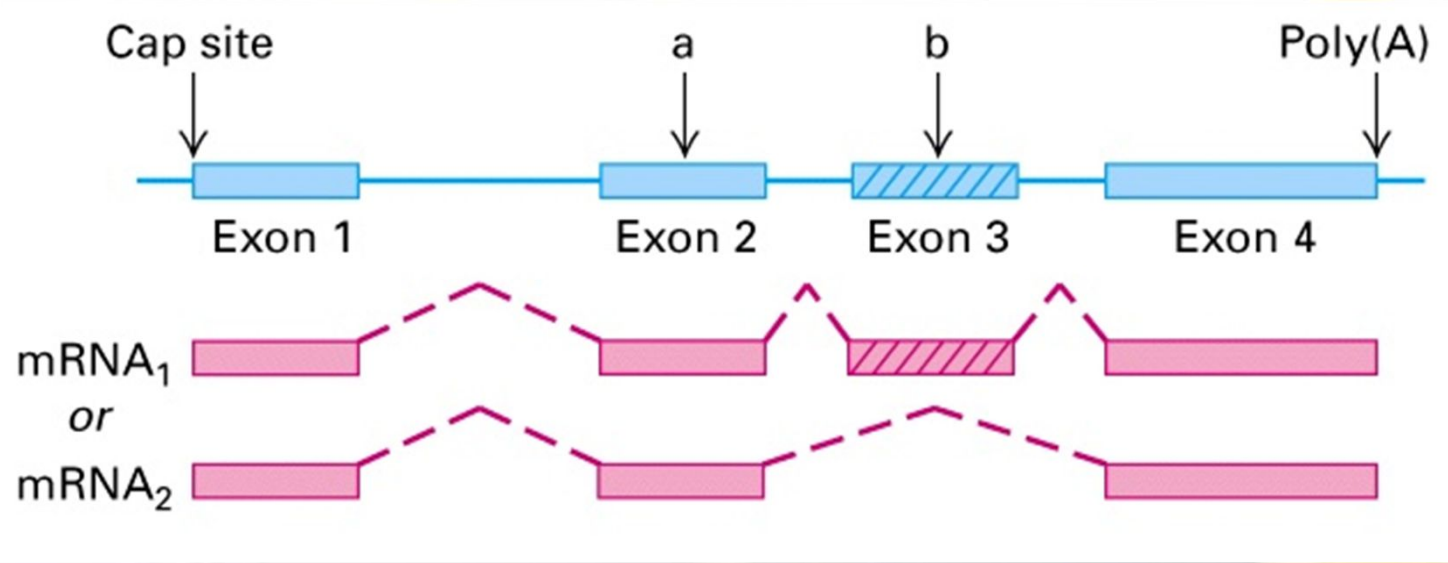
Prokaryotic Gene/Eukaryotic Gene

Alternative Splicing in Eukaryotic Genes

- Most have introns
- Produce monocistronic mRNA
- Large in size

Prokaryotic Gene/Eukaryotic Gene

Eukaryotic Gene: Alternative Splicing



Prokaryotic Gene/Eukaryotic Gene

Eukaryotic Gene

Computational challenges

Identify binding sites

Correlate sequence to expression

Prokaryotic Gene/Eukaryotic Gene

Conclusion

- Most of prokaryotic genes are without introns and in the are polycistronic.
- Eukaryotic genes have introns and alternative splicing.

Types of Eukaryotic DNA

Types of Eukaryotic DNA

Types of Eukaryotic DNA

- Protein coding genes
- Tandemly repeated genes
- Repeated DNA
- Unclassified spacer DNA

Types of Eukaryotic DNA

Types of Eukaryotic DNA

- Protein coding genes can also be in the form of:
- Duplicated and diverged genes
- Functional gene families and non-functional pseudo-genes

Types of Eukaryotic DNA

Types of Eukaryotic DNA

- Tendemly repeated genes encoding rRNA, 5sRNA, tRNA and histones.

Types of Eukaryotic DNA

Repetitive DNA

- Simple sequence DNA
- Moderately repeated DNA (mobile DNA elements)
- Tranposones
- Retrotransposons

Types of Eukaryotic DNA

Repetitive DNA

- Long interspersed elements
- Short interspersed elements
- Unclassified spacer DNA

Types of Eukaryotic DNA

- Protein-coding genes
  - Solitary genes
  - Duplicated and diverged genes (functional gene families and nonfunctional pseudogenes)
- Tandemly repeated genes encoding rRNA, 5S rRNA, tRNA, and histones
- Repetitious DNA
  - Simple-sequence DNA
  - Moderately repeated DNA (mobile DNA elements)
    - Transposons
    - Viral retrotransposons
    - Long interspersed elements (LINES; nonviral retrotransposons)
    - Short interspersed elements (SINES; nonviral retrotransposons)
- Unclassified spacer DNA

Types of Eukaryotic DNA

Major Classes of Eukaryotic DNA in Human Genome

Class	Length	Copy Number in Human Genome	Fraction of Human Genome, %
Protein-coding genes			
Solitary genes	Variable	1	≈15* (0.8)†
Duplicated or diverged genes in gene families	Variable	2–≈1000	≈15* (0.8)†
Tandemly repeated genes encoding rRNAs, tRNAs, snRNAs, and histones	Variable	20–300	0.3
Repetitious DNA			
Simple-sequence DNA	1–500 bp	Variable	3
Interspersed repeats			
DNA transposons	2–3 kb	300,000	3
LTR retrotransposons	6–11 kb	440,000	8
Non-LTR retrotransposons			
LINEs	6–8 kb	860,000	21
SINEs	100–300 bp	1,600,000	13
Processed pseudogenes	Variable	1–≈100	≈0.4
Unclassified spacer DNA	Variable	n.a.‡	≈25

Types of Eukaryotic DNA

Types of Eukaryotic DNA

- Protein coding genes
- Tandemly repeated genes
- Repeated DNA
- Unclassified spacer DNA

Duplicated Genes and Pseudo-Genes

Duplicated Genes/Pseudo-Genes

Duplicated Genes

- Encode closely related(homologous) proteins
- Clustered together in genome
- Formed by duplication of an ancestral gene followed by mutation

Duplicated Genes/Pseudo-Genes

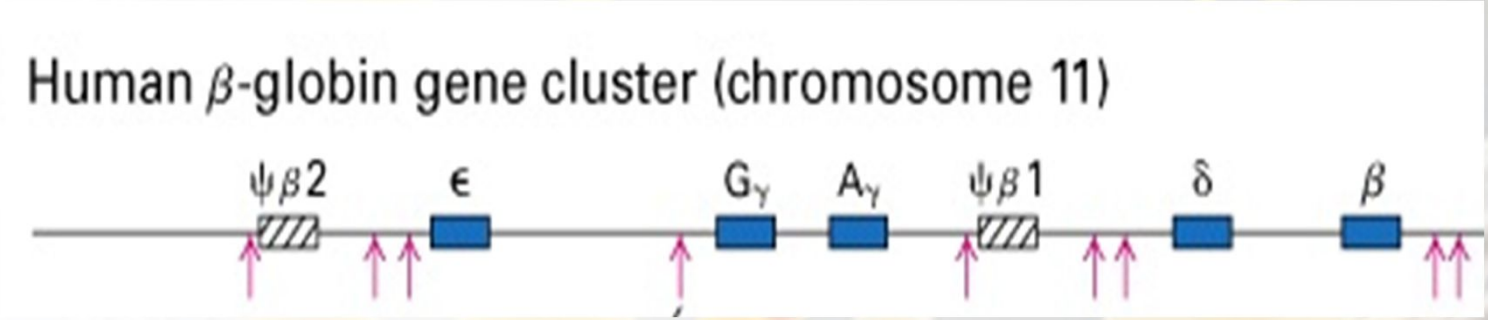
Pseudo-Genes

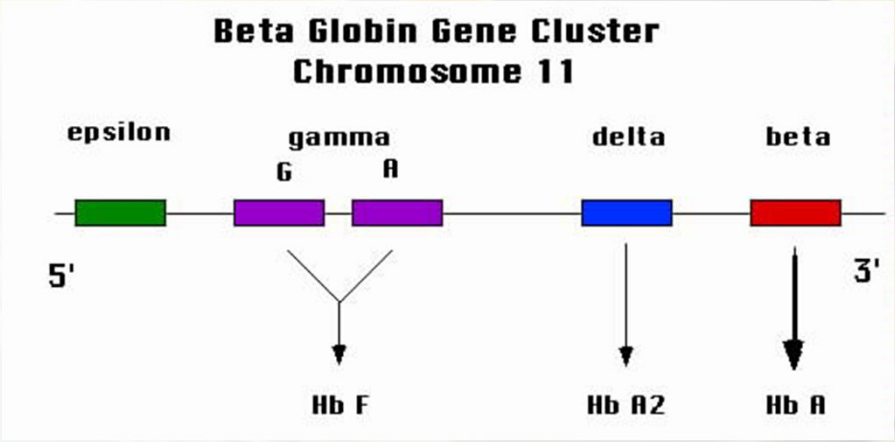
- Nonfunctional copies of genes
- Formed by duplication of ancestral gene, or reverse transcription and integration.

Duplicated Genes/Pseudo-Genes

Duplicated Genes

- Five functional genes and two pseudo-genes





Duplicated Genes/Pseudo-Genes

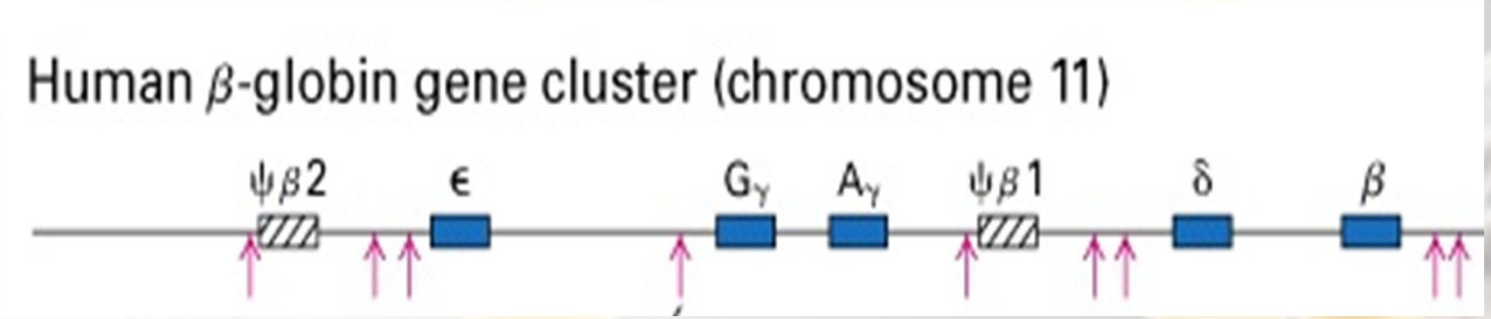
Pseudo-Genes

- Not expressed due to mutations that produce
- A stop codon
- Prevent mRNA processing
- Due to lack of regulatory sequences

Duplicated Genes/Pseudo-Genes

Pseudo-Genes

- Five functional genes and two pseudo-genes



Duplicated Genes/Pseudo-Genes

Genes duplication: Globin family

- Ancestral globin gene was duplicated ~ 500 million years ago.
- Mutations in both genes to differentiate them - α and β present in all higher vertebrates

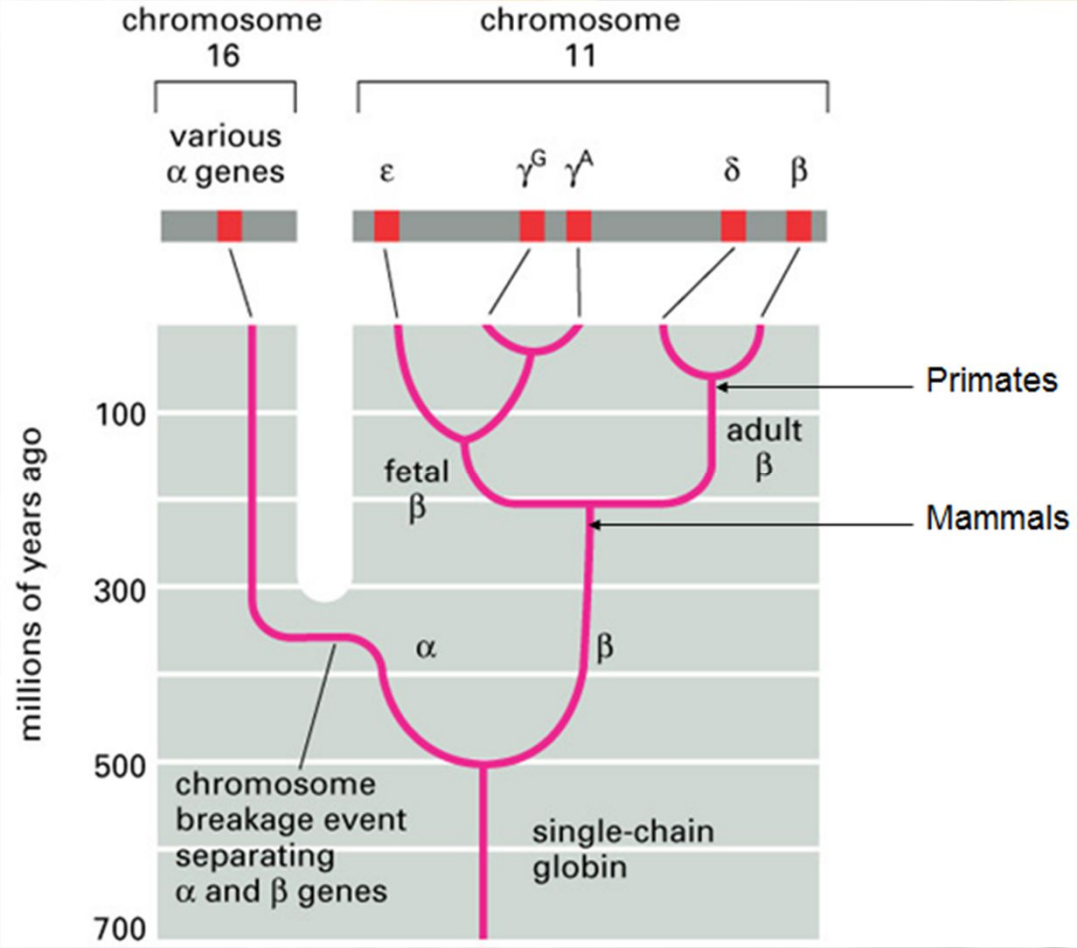
Duplicated Genes/Pseudo-Genes

Genes duplication: Globin family

- Further gene duplications produced alternative forms in mammals and in primates

Duplicated Genes/Pseudo-Genes

Genes Duplication: Globin Family



Duplicated Genes/Pseudo-Genes

DNA Duplications:

- Two phenomenon are common in DNA duplications:
- Slipped strand mispairing
- Unequal crossover during recombination

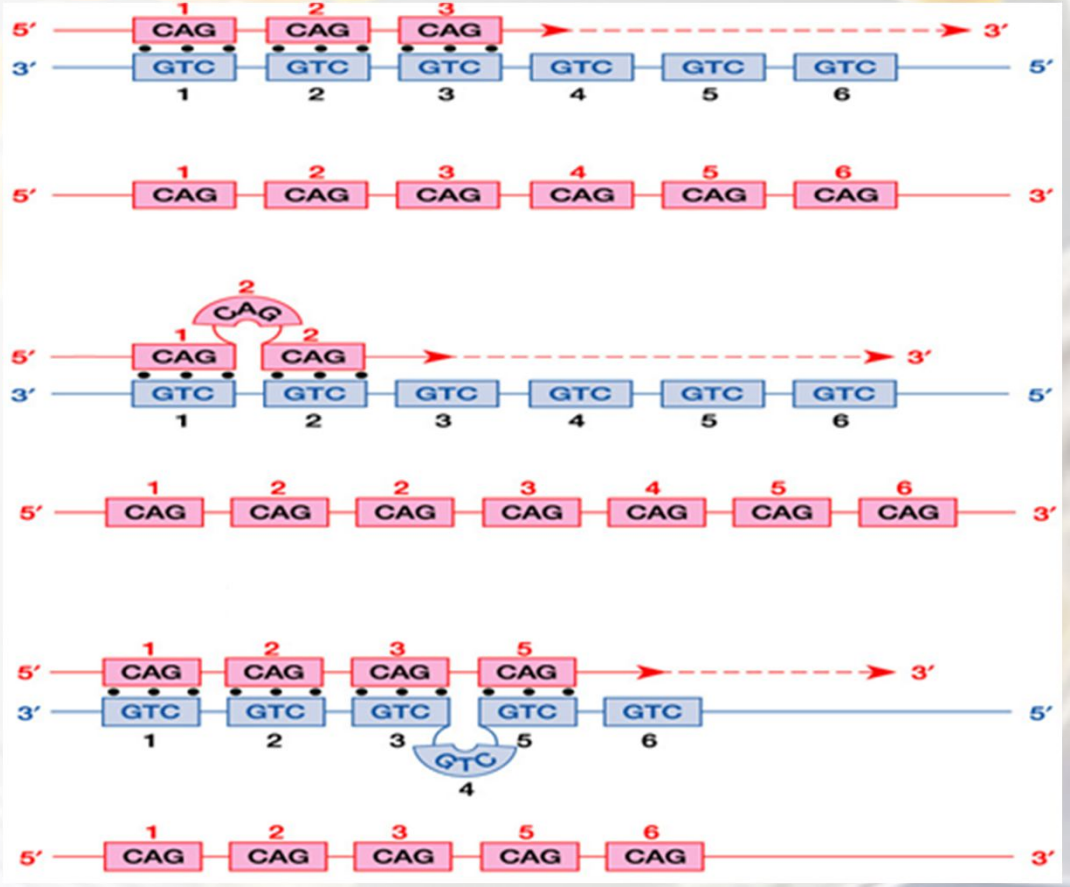
Duplicated Genes/Pseudo-Genes

DNA Duplications: Slipped strand mispairing

Normal replication

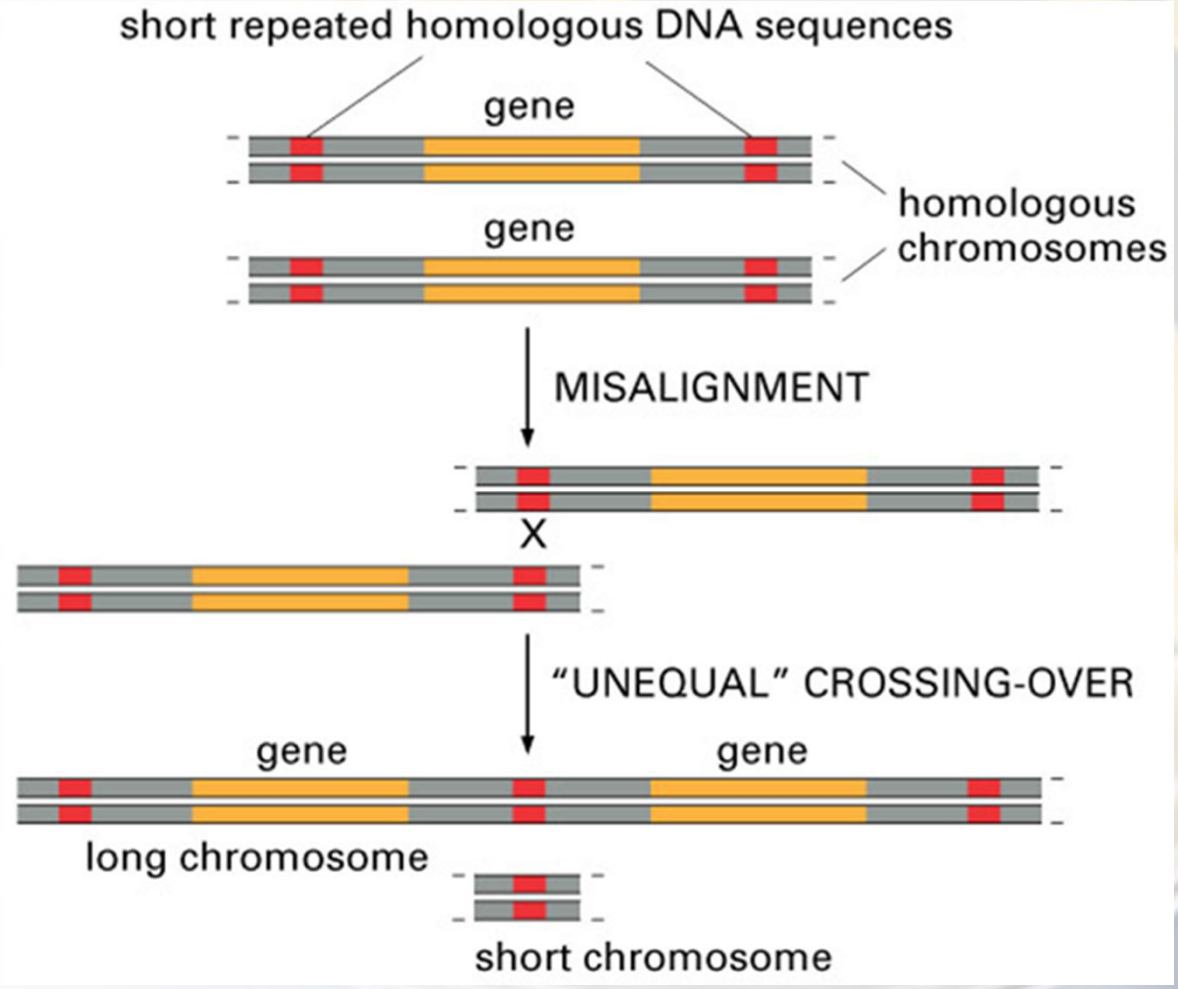
Backward slippage causes insertions

Forward slippage slippage causes deletions



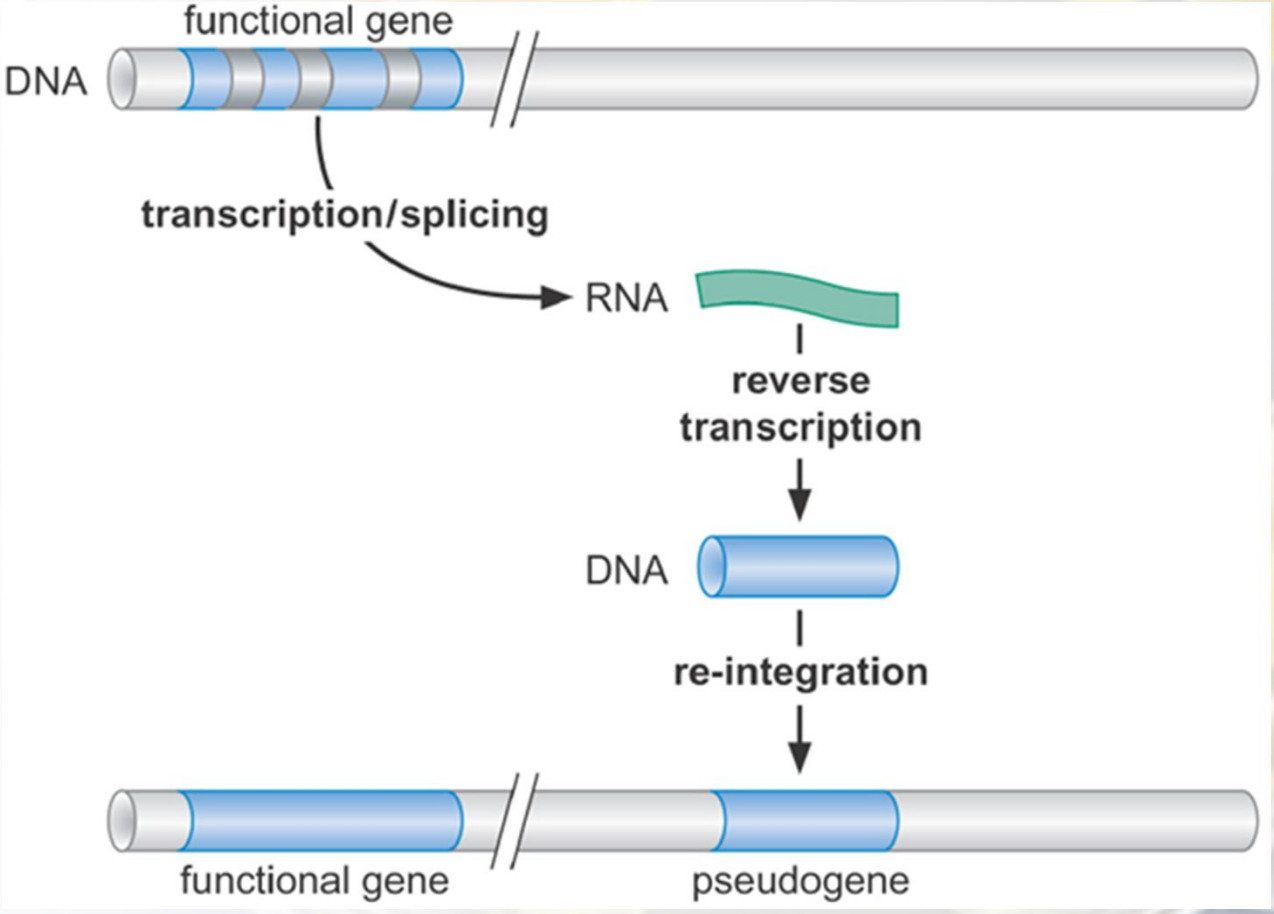
Duplicated Genes/Pseudo-Genes

DNA Duplications: Unequal Cross Over



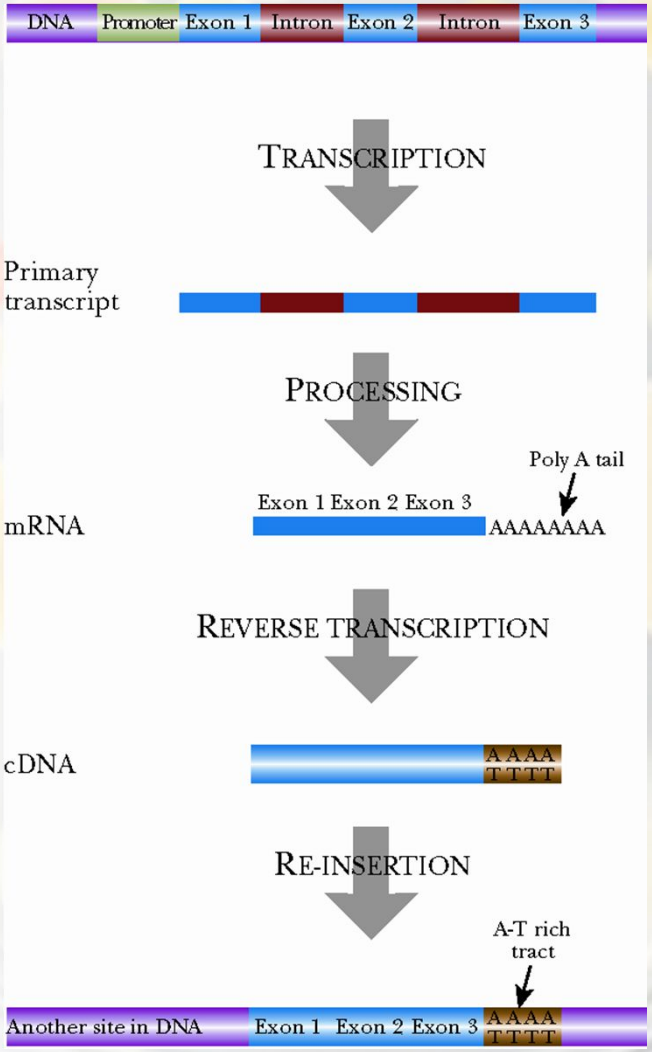
Duplicated Genes/Pseudo-Genes

Pseudo-Genes



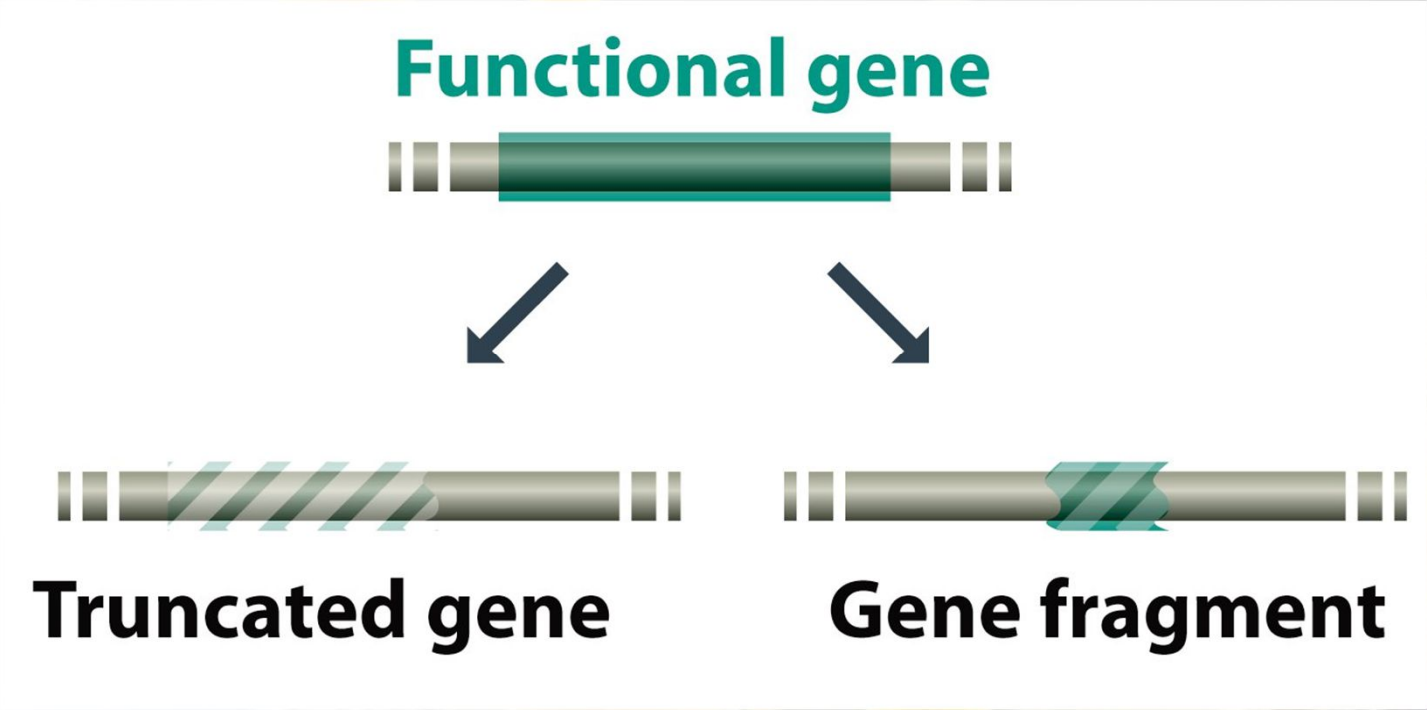
Duplicated Genes/Pseudo-Genes

Pseudo-Genes



Duplicated Genes/Pseudo-Genes

Functional Genes: Process of evolution produced truncated genes and gene fragments



Duplicated Genes/Pseudo-Genes

Duplicated Genes and Pseudo Genes

- Clustered together in genome.
- Formed by duplication of an ancestral gene
- Nonfunctional copies of genes

Genetics and Genomics

Repetitive DNA

Repetitive DNA

Chromosomes - highly dynamic

- Whole genome duplication
- The genomes of two distinct species can merge
- An individual can acquire an extra copy of a chromosome

Repetitive DNA

Chromosomes - highly dynamic

- Chromosomes can fuse; e.g. human chromosome 2 derived from a fusion of two ancestral primate chromosomes

Repetitive DNA

Chromosomes - highly dynamic

- Chromosomal regions can be inverted or deleted
- Segmental and other duplications can occur

Repetitive DNA

Five main classes of Repetitive DNA

- Interspersed repeats (RNA/DNA transposon-derived)
- Approximately 45% of human genome (e.g. LINES, SINES, Alu)

Repetitive DNA

Interspersed Repeats

- Retrotransposons constitute over 40% of the human genome and consist of several millions of family members.
- They play important roles in shaping the structure/evolution of the genome.

Repetitive DNA

Processed Pseudogenes

- These genes have a stop codon, frameshift mutation, or loss of promoter activity and do not encode a functional protein.

Repetitive DNA

Processed Pseudogenes

- They commonly arise from retrotransposition, or following gene duplication and subsequent gene loss.

Repetitive DNA

Simple Sequence Repeats

- Simple sequence repeats
- Microsatellites (1-12 bp)
- Minisatellites (12-500 bp)

Repetitive DNA

Segmental Duplications

- Segmental duplications blocks of about 1 kb to 300 kb that are copied within or between chromosomes
- ~5% of human genome.

Repetitive DNA

Blocks of Tandem Repeats

- Blocks of tandem repeats includes telomeric and centromeric repeats and can span millions bp
- Often species specific.

Repetitive DNA

Repeats in Mouse genome

Type	No. of Repeats	Size	Percent of genome
Highly repetitive	> 1 Million	< 10 bp	10 %
Moderately repetitive	> 1000	~ 150 - ~300 bp	20 %

Repetitive DNA

Conclusion

- Main classes of Repetitive DNA
- Simple sequence of DNA
- DNA transposons
- LTR retrotransposons
- Non LTR retrotranposons like LINE, SINE

Genetics and Genomics

Mobile DNA

Mobile DNA

Mobile DNA

- Move within genomes
- Most of moderately repeated DNA sequences found throughout higher eukaryotic genomes
- Most mobile DNA - transposable elements

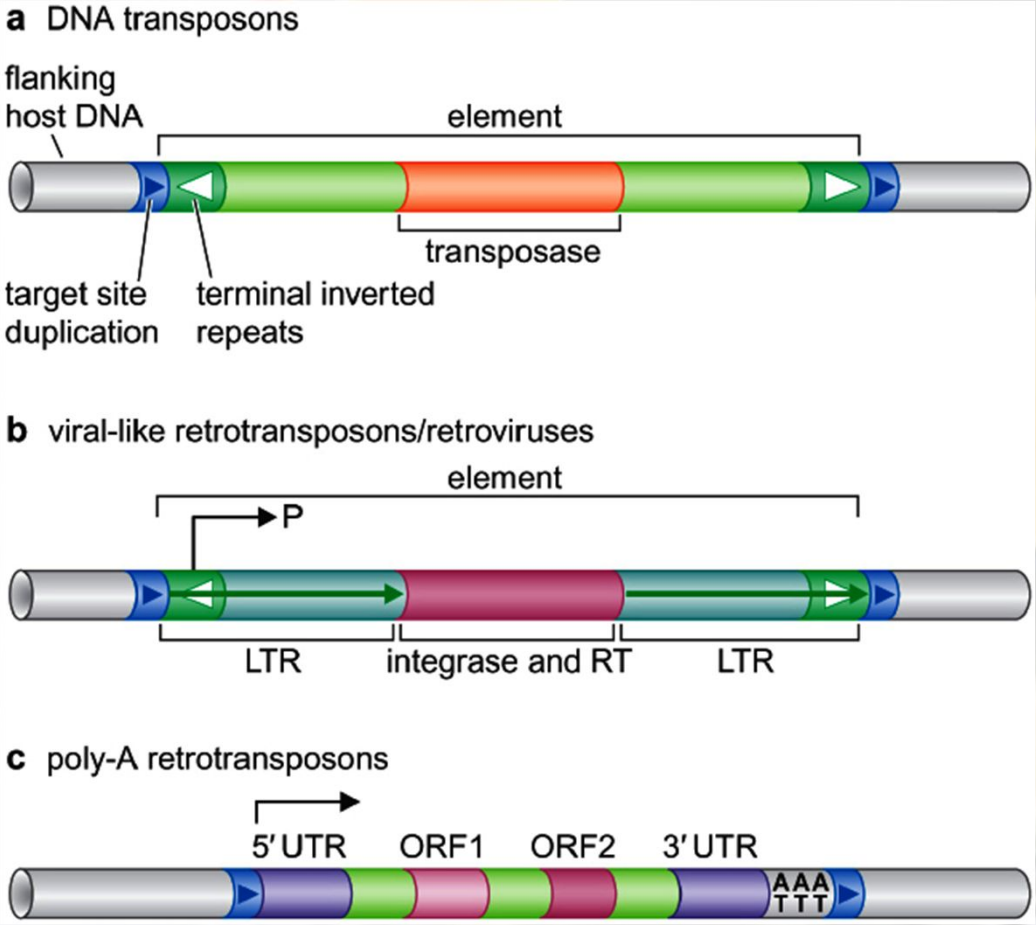
Mobile DNA

Mobile DNA

- Includes DNA-based transposons and retro-transposons.
- Some encode enzymes that catalyze movement.

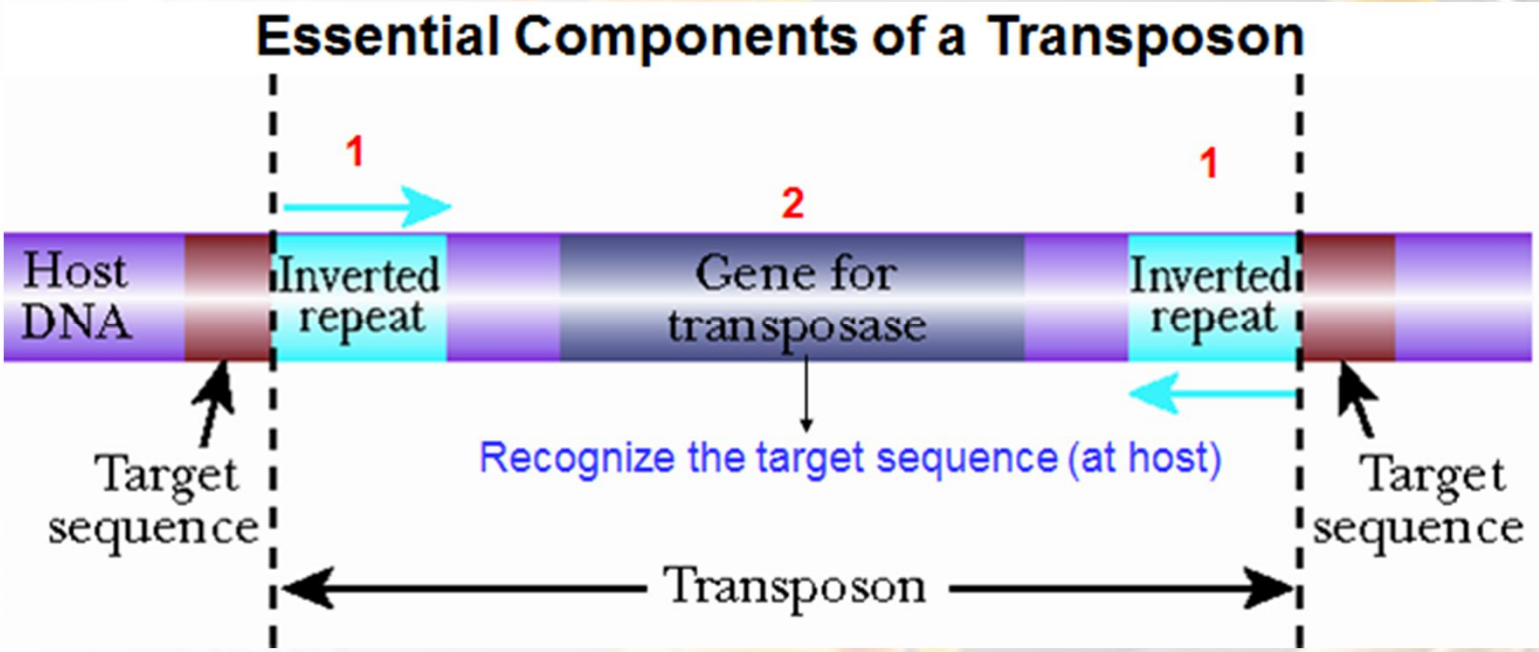
Mobile DNA

Mobile DNA – Types of Transposons



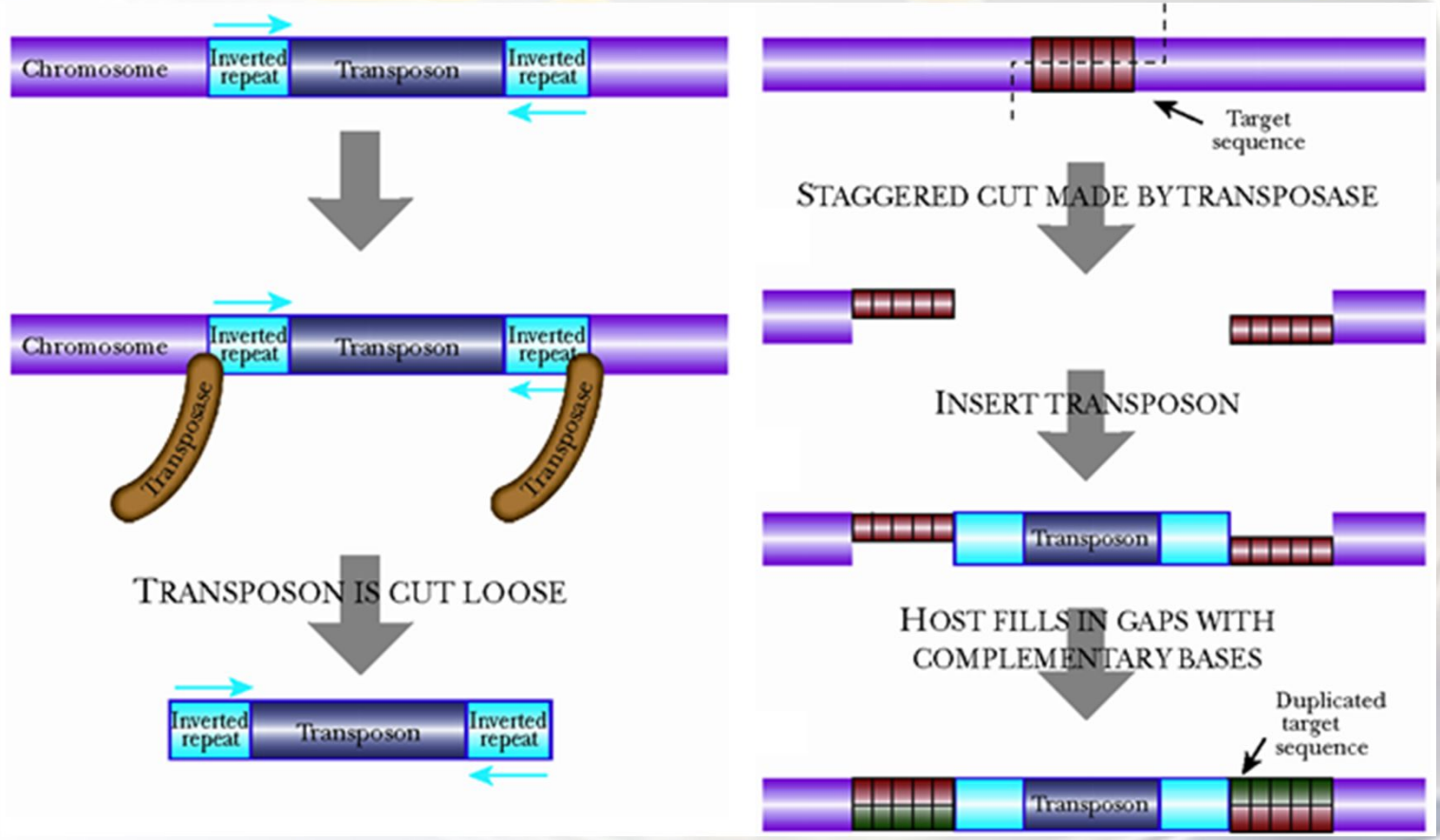
Mobile DNA

Mobile DNA



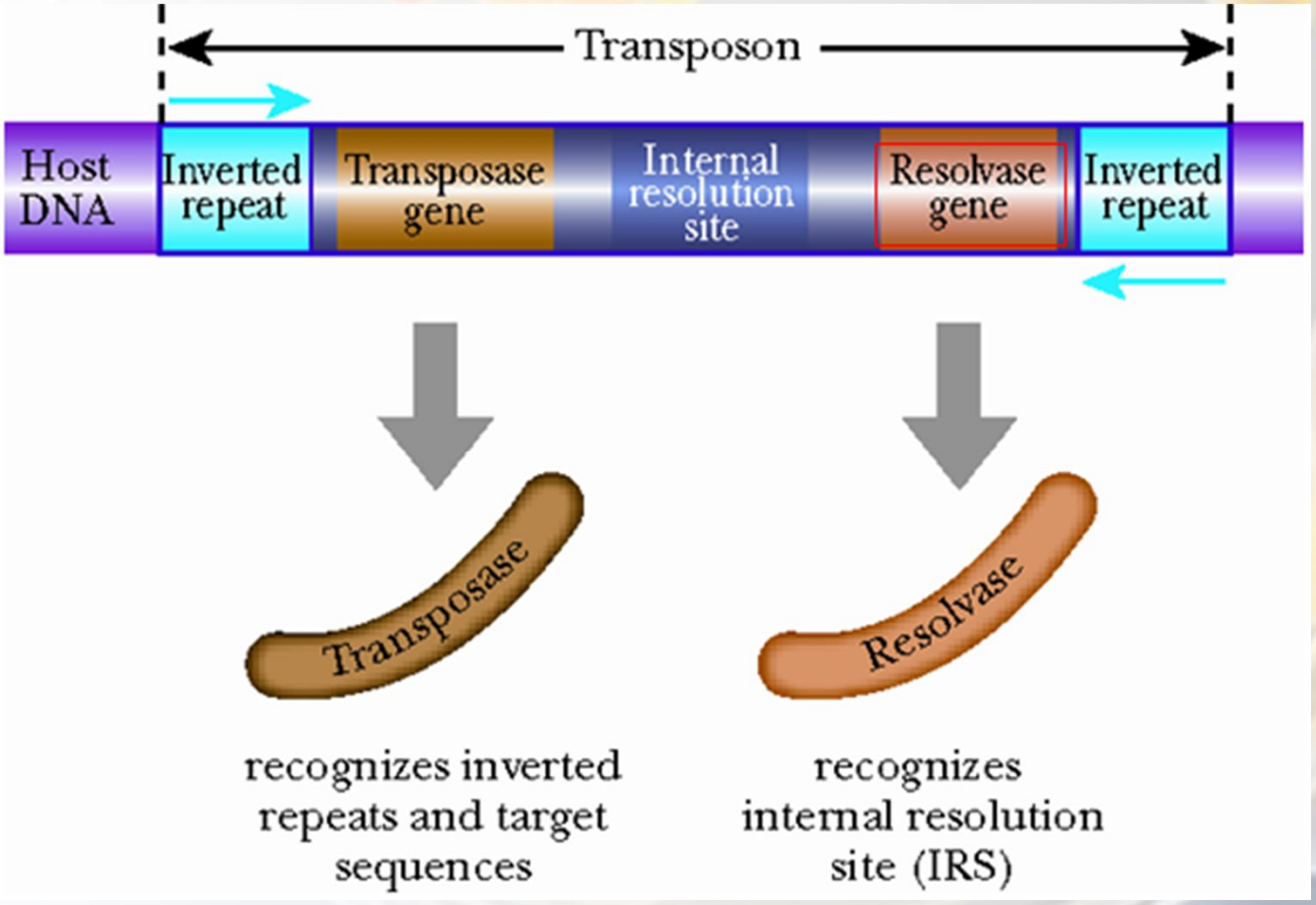
Mobile DNA

Transposon is cut off by transposase and inserted on another site



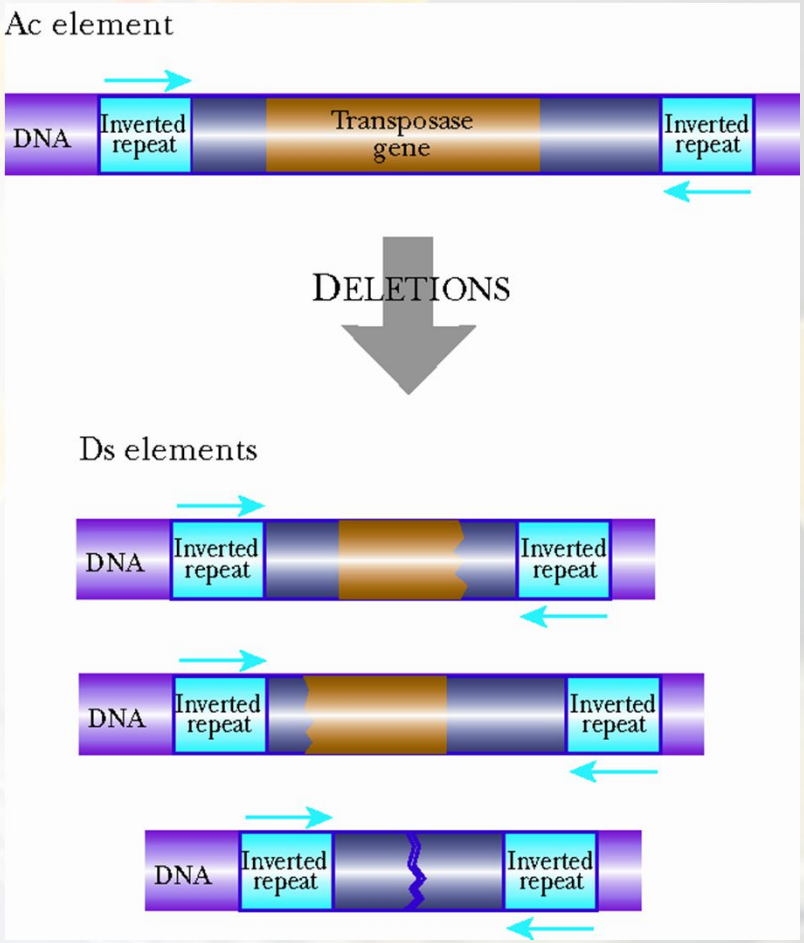
Mobile DNA

Components of a Complex Transposon



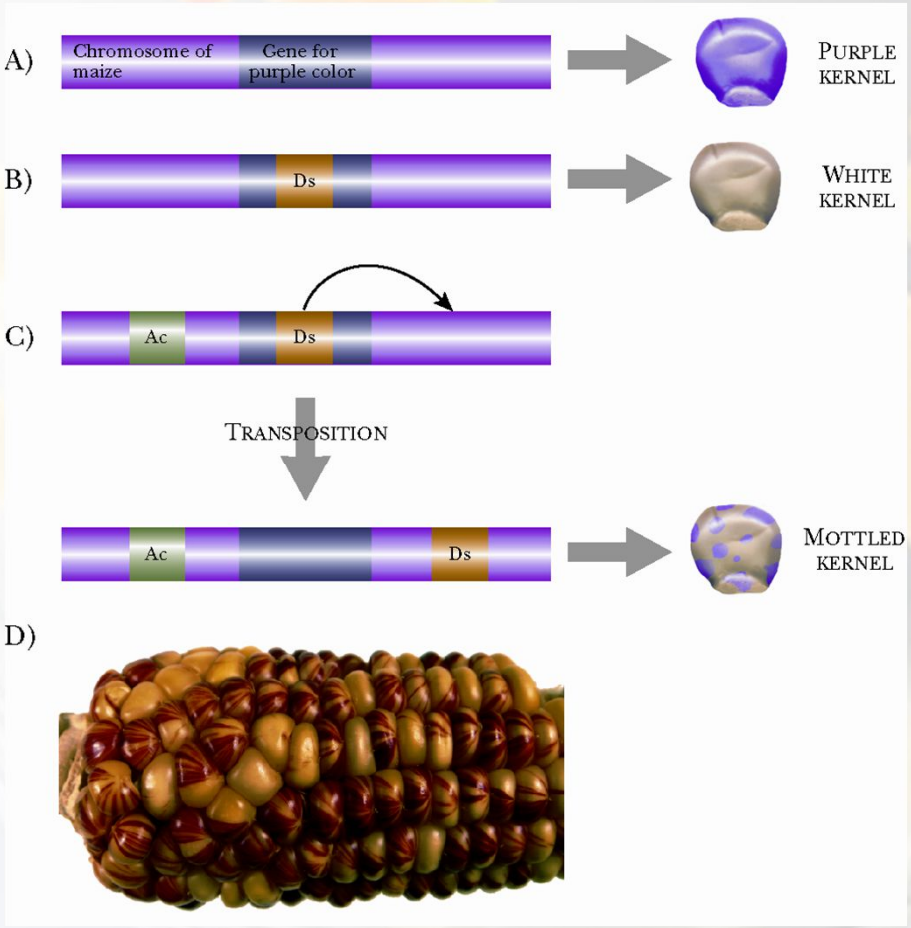
Mobile DNA

Ac and Ds Elements of Maize – Higher Life Form



Mobile DNA

Movement of Ds Elements Gives Mottled Corn



Mobile DNA

Mobile DNA - Transposition

- Transposition is movement of mobile DNA
- Involves copying of mobile DNA element and insertion into new site in genome
- Molecular parasite: “selfish DNA”

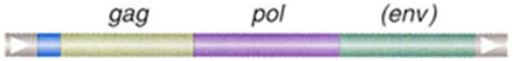
Mobile DNA

Mobile DNA - Retrotransposition

- Moving in the form of RNA by element coding for reverse transcriptase
- LINEs
- SINEs
- retrovirus-like elements (long terminal repeat)

Mobile DNA

Mobile DNA – Interspersed Repeats in Human Genome

Element	Transposition	Structure	Length	Copy number	Fraction of genome
LINEs	Autonomous		1–5 kb	20,000–40,000	21%
SINEs	Nonautonomous		100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Nonautonomous		1.5–3 kb		
DNA transposons	Autonomous		2–3 kb	300,000	3%
	Nonautonomous		80–3000 bp		

Mobile DNA

Mobile DNA – Fuel for Evolution

- Probably have significant effect on evolution by facilitating gene duplication, which provides the fuel for evolution, and exon shuffling

Movement of Transposons and Retro-transposons

Transposons/Retrotransposons-Movement

Movement of Transposons

- Eukaryotic transposable elements - two types
- **Transposons**, which move by means of a DNA intermediate
- **Retrotransposons**, which move by means of an RNA intermediate

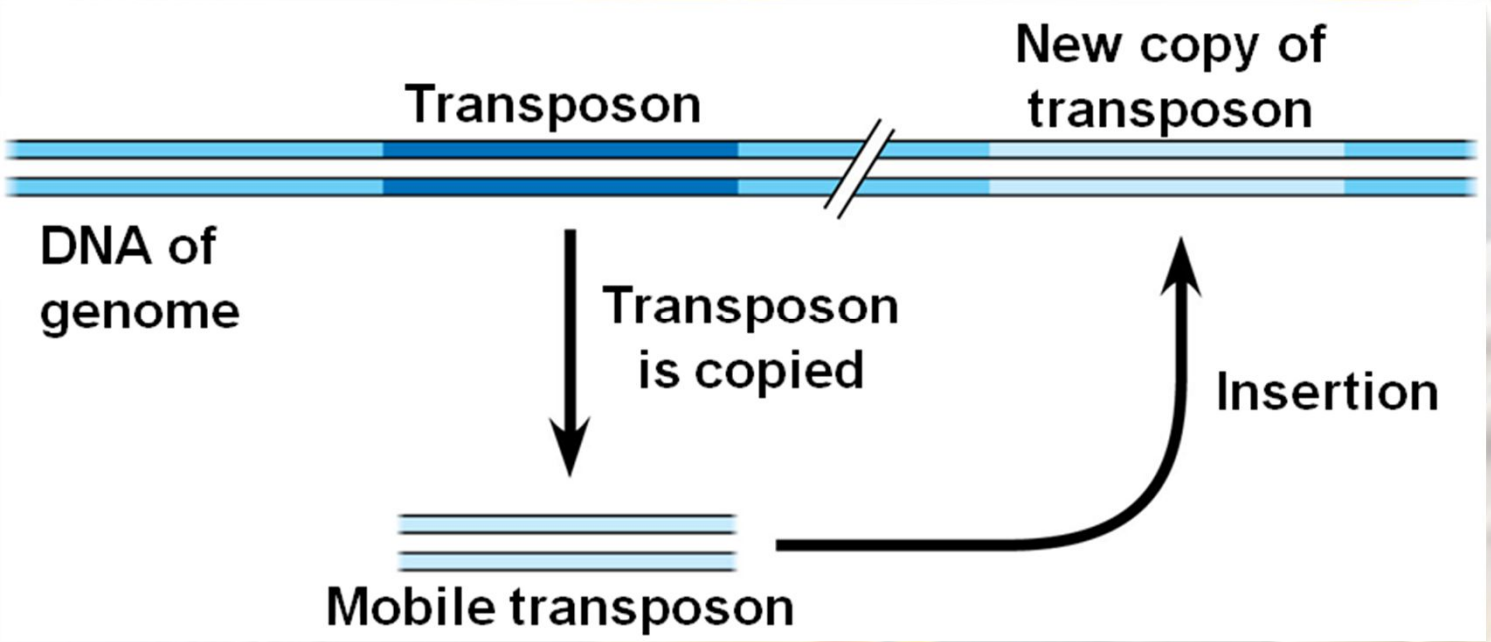
Transposons/Retrotransposons-Movement

Transposons

- Transposons - major content of eukaryotic genomes
- ~50% of Human/Mouse
- ~75% of the maize genome
- ~85% of the barley genome
- ~98% of the iris genome

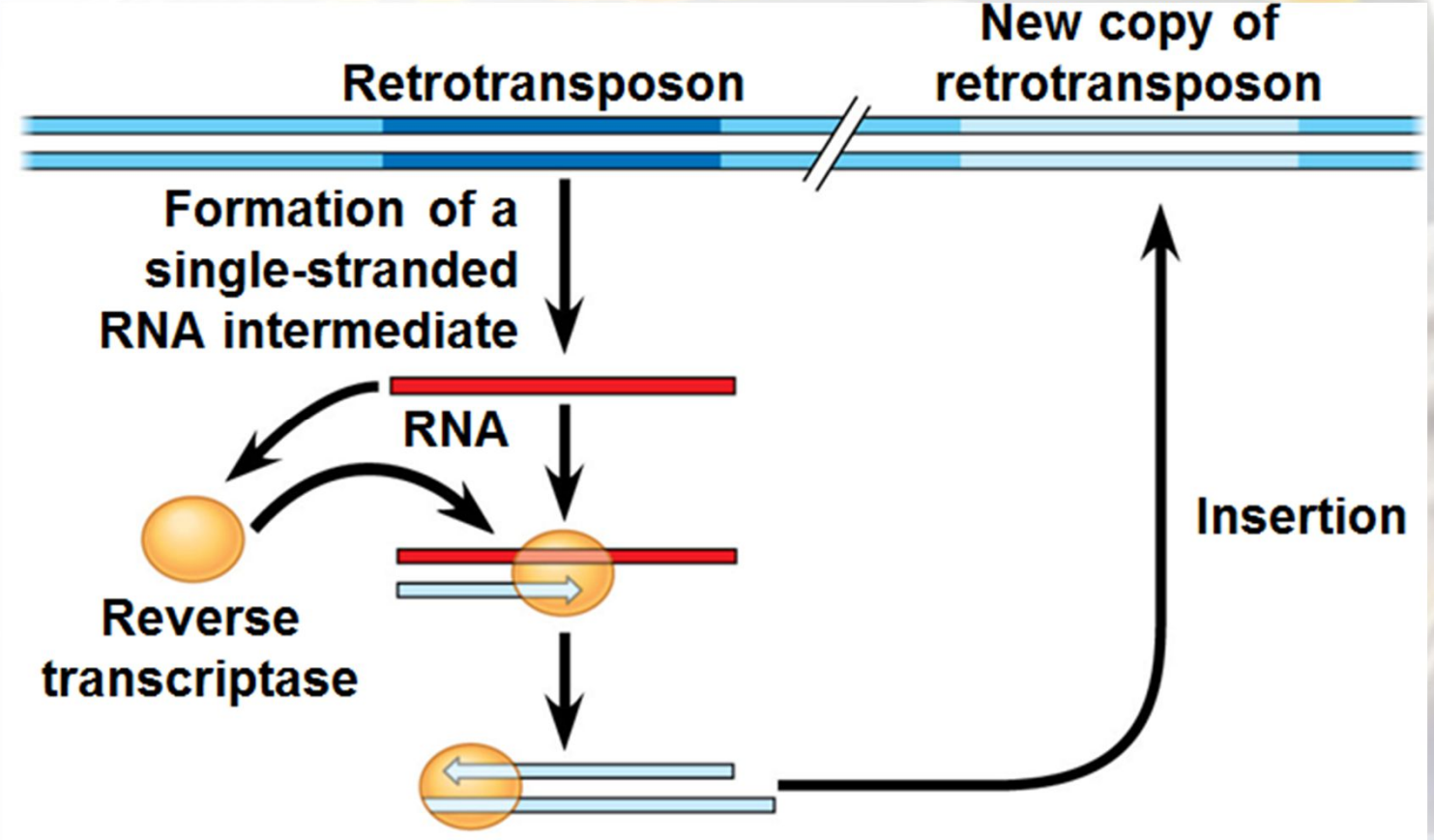
Transposons/Retrotransposons-Movement

Movement of Transposons



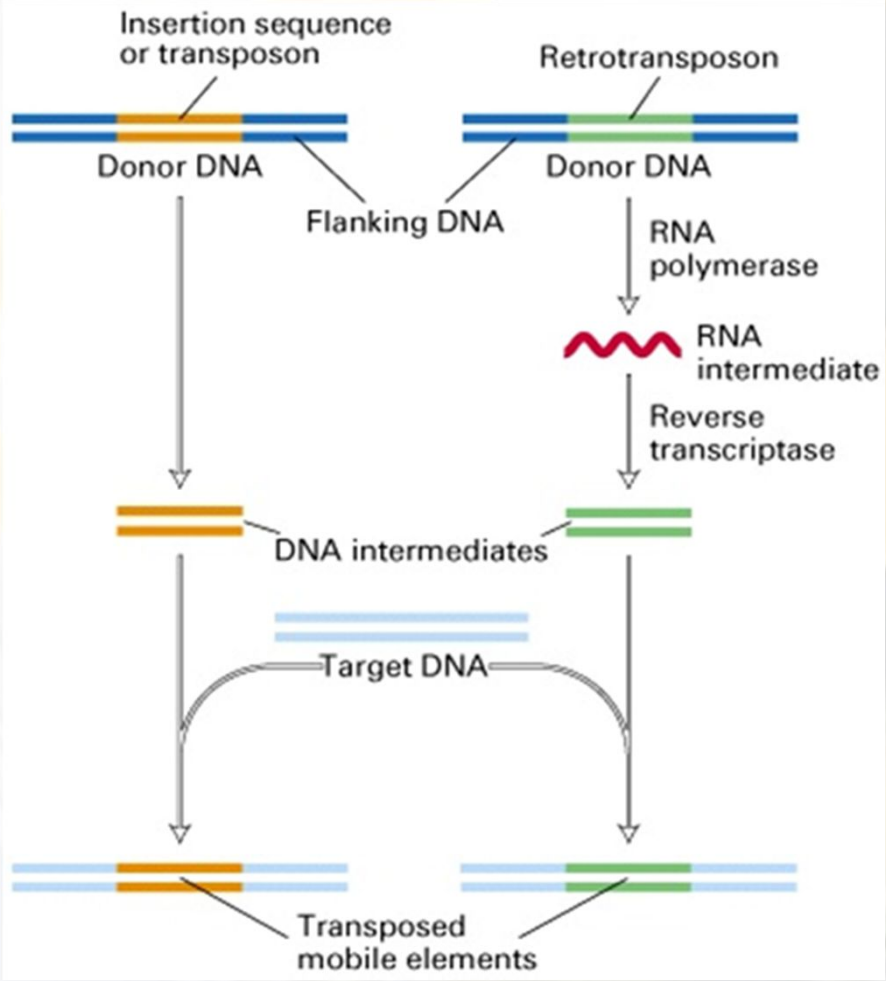
Transposons/Retrotransposons-Movement

Movement of Retrotransposons



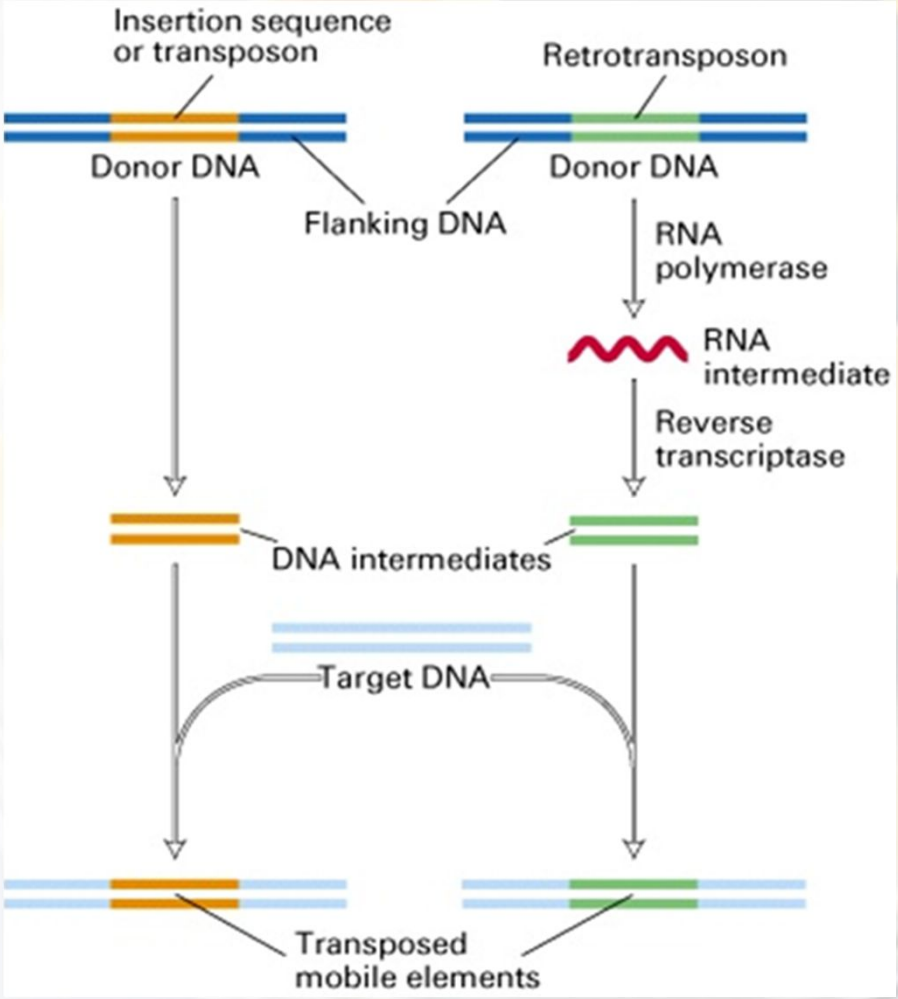
Transposons/Retrotransposons-Movement

Transposons moves using DNA intermediate



Transposons/Retrotransposons-Movement

Retrotransposon moves using RNA intermediate



Transposons/Retrotransposons-Movement

Conclusion

- Eukaryotic transposable elements - two types
- **Transposons**, which move by means of a DNA intermediate
- **Retrotransposons**, which move by means of an RNA intermediate

Transposons in Prokaryotes and Eukaryotes

Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes

- Two types of transposons in prokaryotes i.e.
- Insertion Sequence (IS)
- Transposons (Tn) and Bateriophage Mu

Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: IS Elements

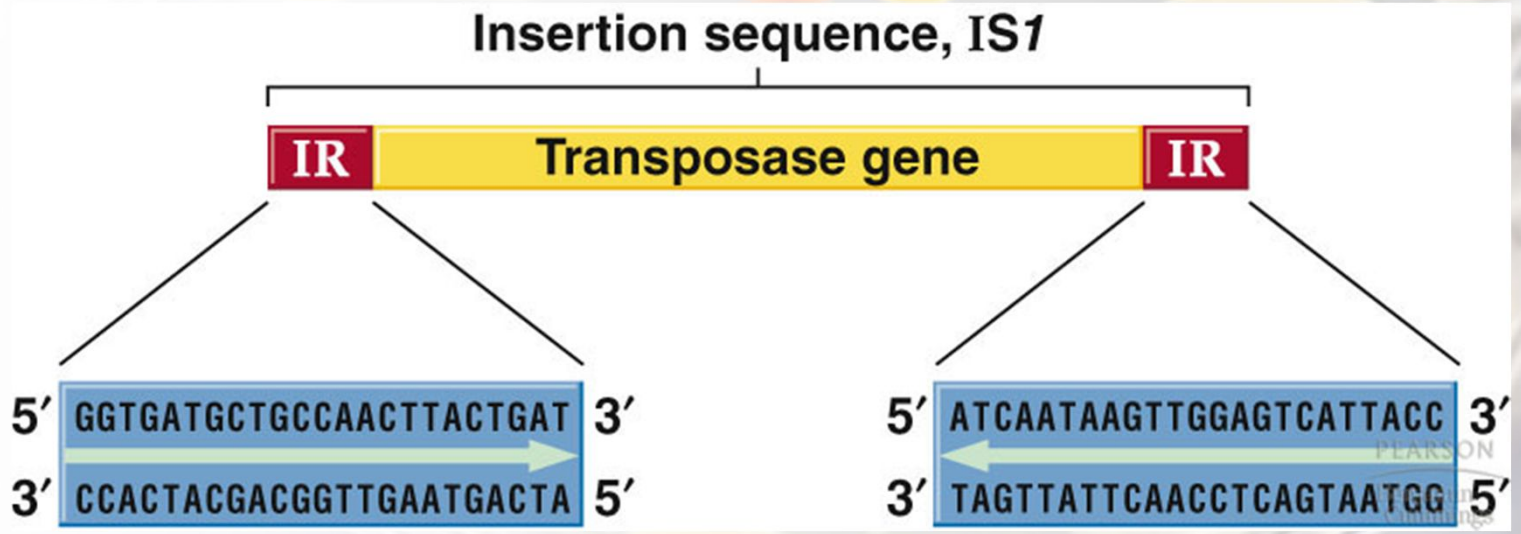
- Prokaryotic IS elements range in size from 768 bp to over 5 kb.
- IS1 is 768 bp long, and present in 4–19 copies on the E. coli chromosome.
- IS2 has 0–12 copies on the chromosome, and 1 copy on the F plasmid
- IS10 is found in R plasmids

Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: IS elements

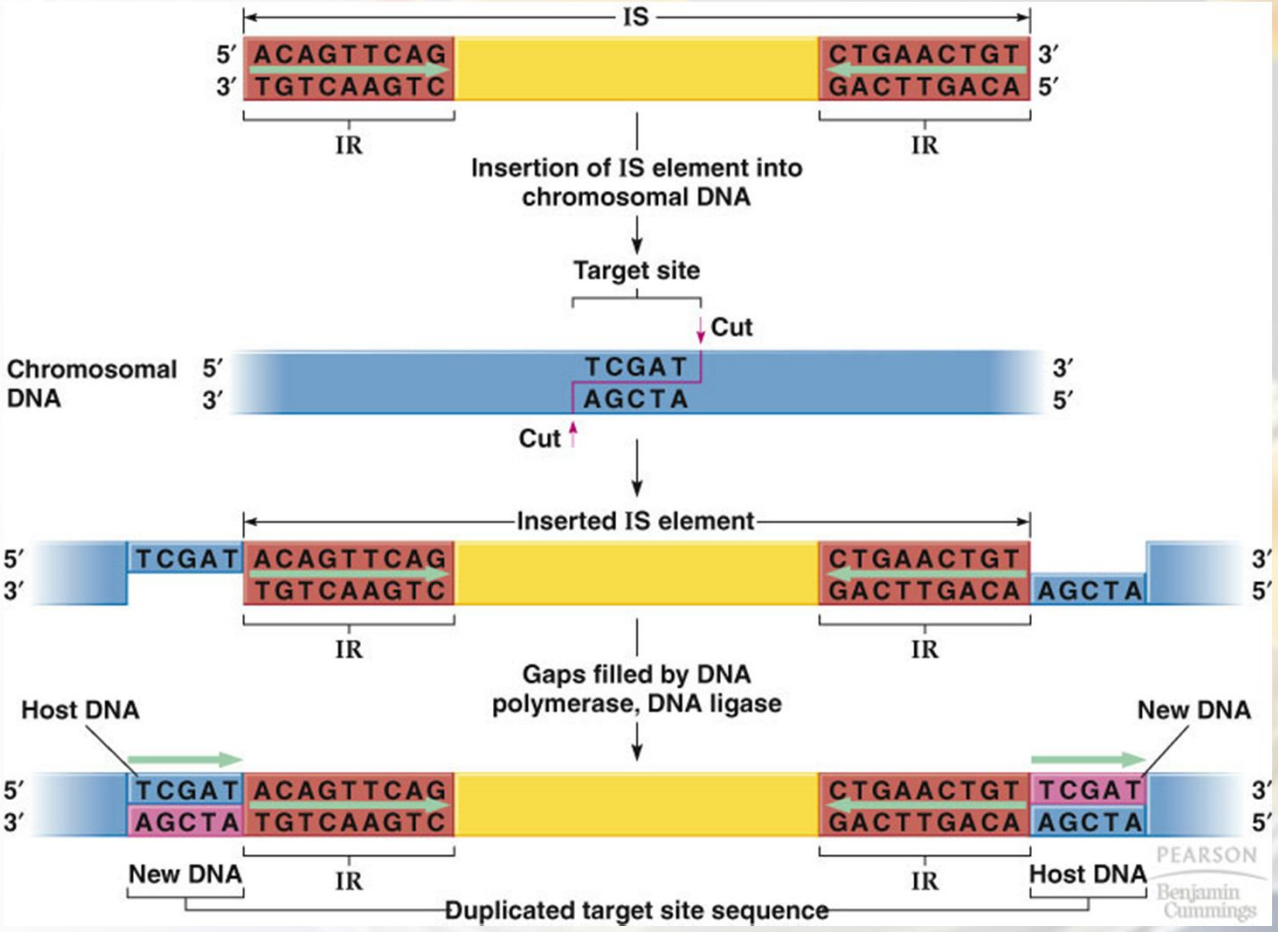
The ends of all sequences (IS elements) show inverted terminal repeats (IRs) of 9–41 bp

(e.g., IS1 has 23 bp of nearly identical sequence).



Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: IS Elements



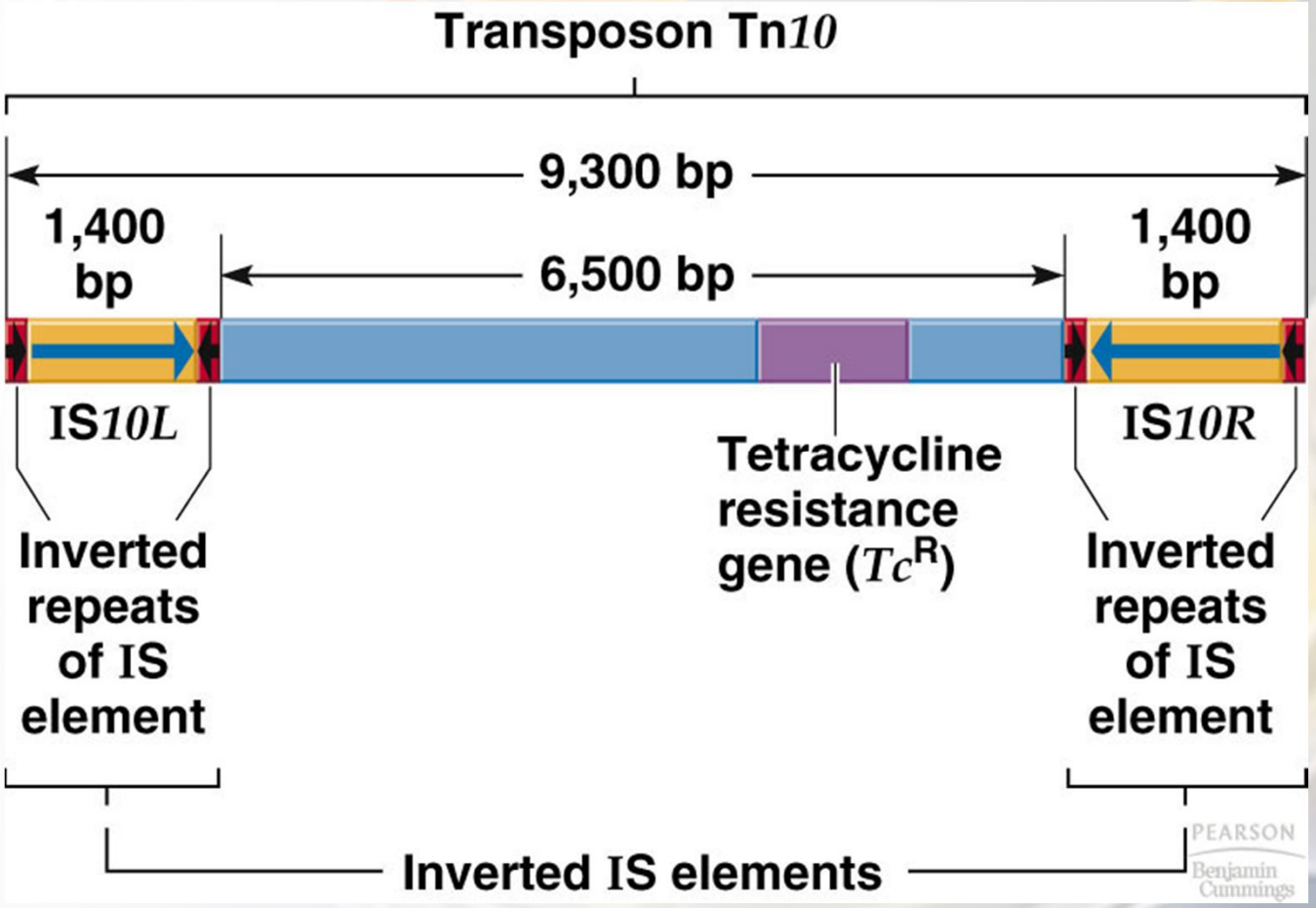
Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes

- Transposons are similar to IS elements, but carry additional genes, and have a more complex structure.
- Two types of prokaryotic transposons:
- Composite transposons carry genes (e.g., antibiotic resistance) flanked on both sides by IS elements (IS modules). Example Tn10
- Non-composite transposons also carry genes (e.g., drug resistance) but do not terminate with IS elements.

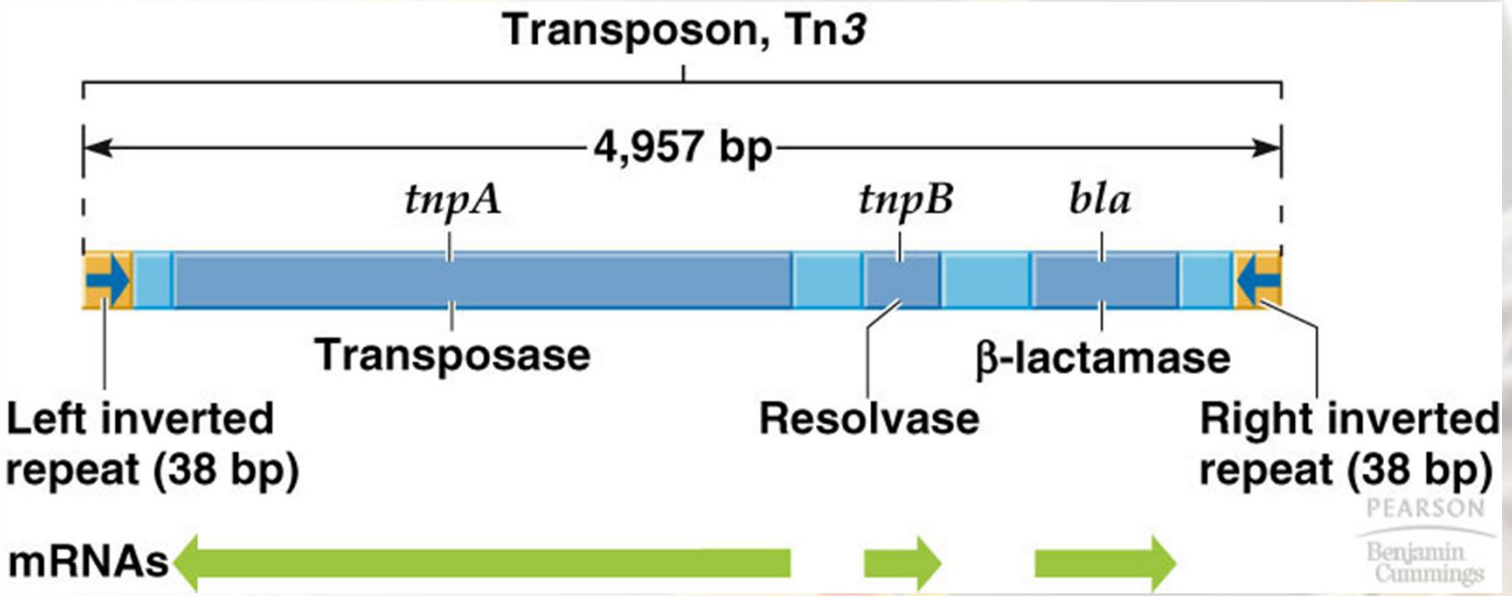
Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: Composite Tn10



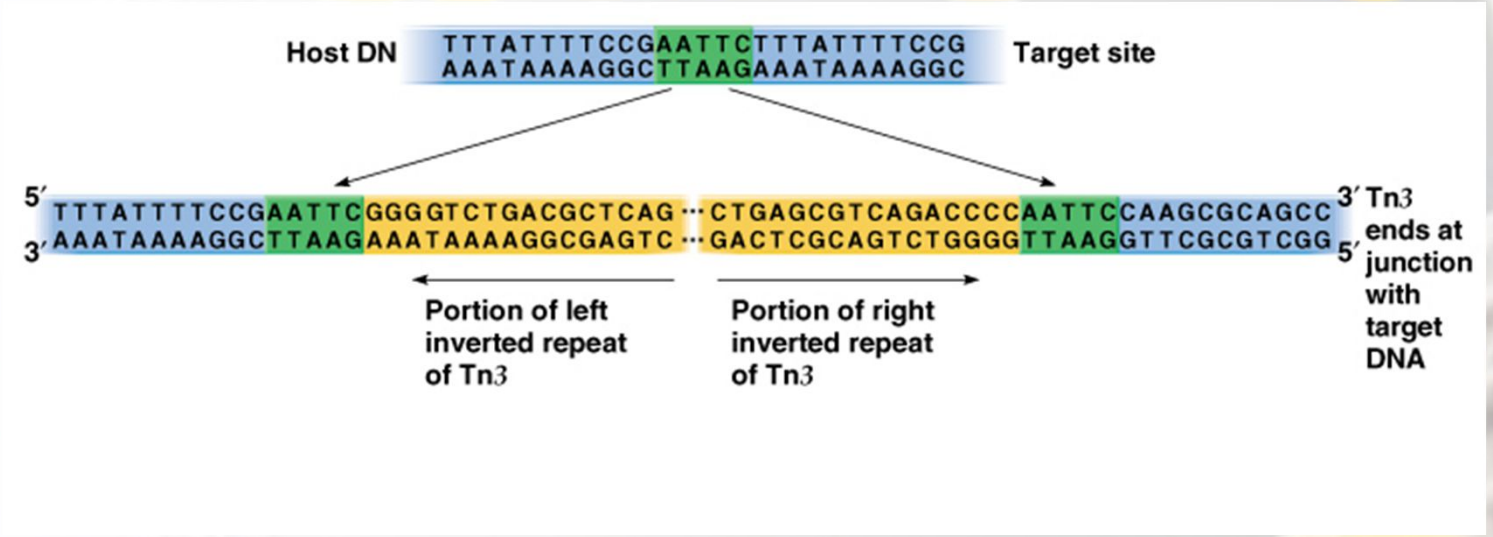
Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: Non-composite Tn3



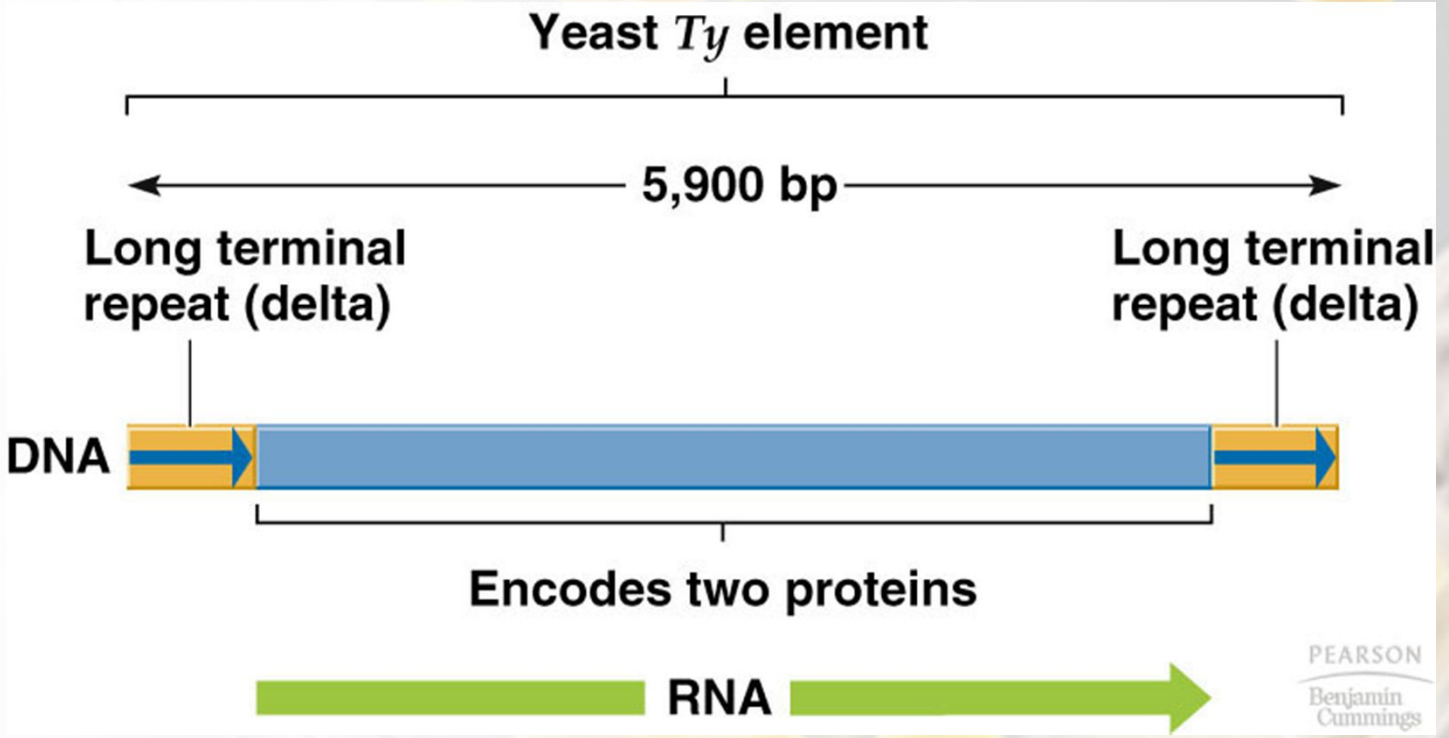
Transposons in Prokaryotes/Eukaryotes

Transposons in Prokaryotes: Non-composite Tn3



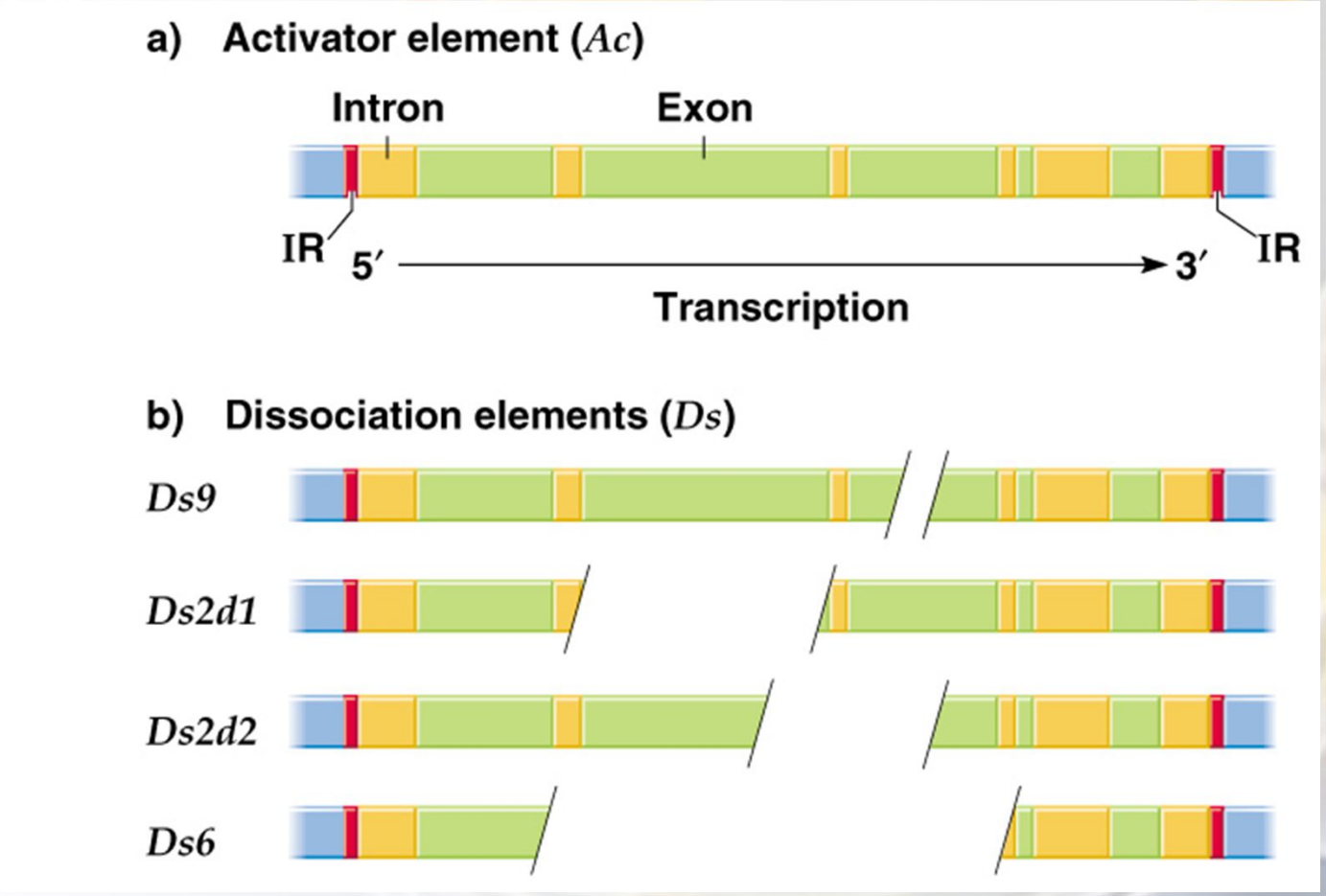
Transposons in Prokaryotes/Eukaryotes

Transposons in Eukaryotes: Ty Elements in Yeast



Transposons in Prokaryotes/Eukaryotes

Transposons in Eukaryotes: Ac and Ds Elements



Transposons in Prokaryotes/Eukaryotes

Transposons in Eukaryotes: Ac and Ds Elements in Maize

- Ac is 4,563 bp, with 1 1-bp imperfect terminal IRs and 1 transcription unit producing a 3.5 kb mRNA encoding an 807 amino acid transposase.
- Ac activates Ds to transpose or break the chromosome where it is inserted.
- Ds elements vary in length and sequence, but all have the same terminal IRs as Ac, and many are deleted or rearranged versions of Ac

Transposons in Prokaryotes/Eukaryotes

Transposons

- Transposons are found in both Prokaryotes and Eukaryotes.

Long Terminal Repeats

Long Terminal Repeats

Long Terminal Repeats

- Eukaryotic retrotransposons fall into two major groups:
- LTR retrotransposons
- Non-LTR retrotransposons

Long Terminal Repeats

Long Terminal Repeats

- LTRs stand for Long direct Terminal Repeats.
- LTRs consist of 250-600 bp direct repeat sequences located at the ends of the retrotransposon coding region

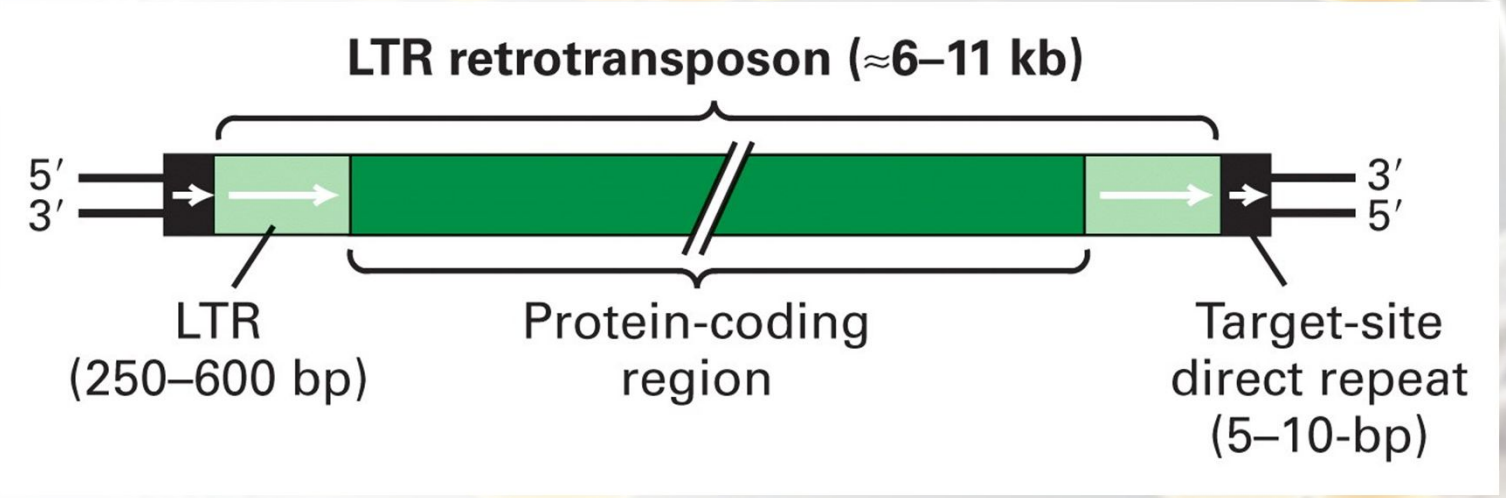
Long Terminal Repeats

Long Terminal Repeats

- Flank viral retrotransposons and retroviruses
- Contain regulatory sequences like transcription start site and poly (A) site

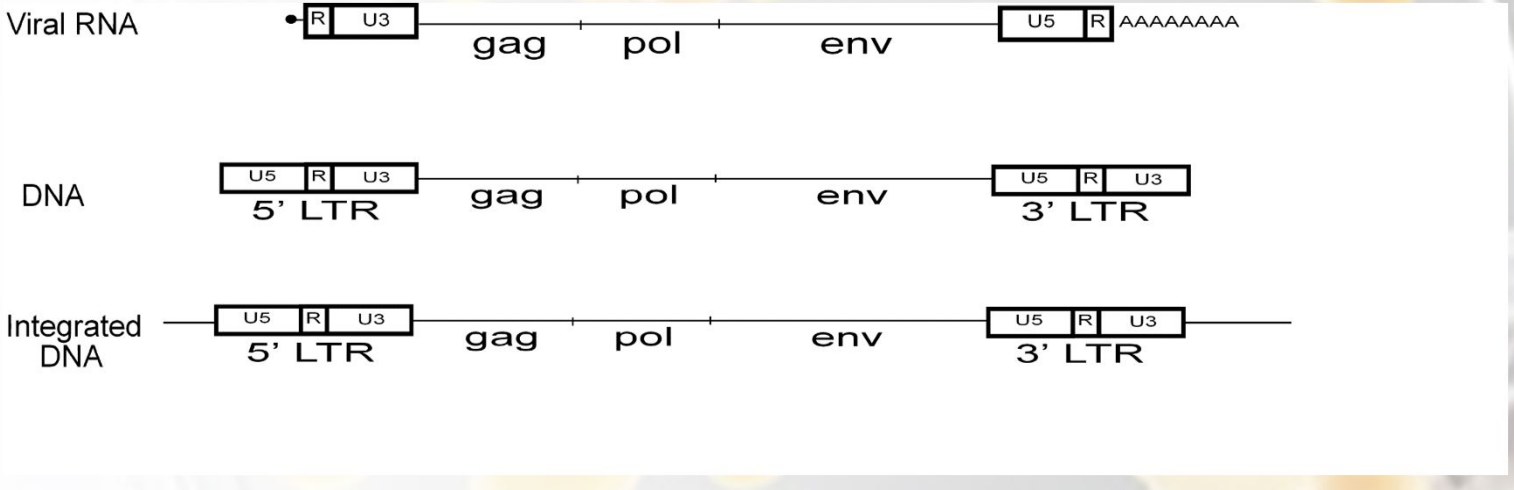
Long Terminal Repeats

Long Terminal Repeats



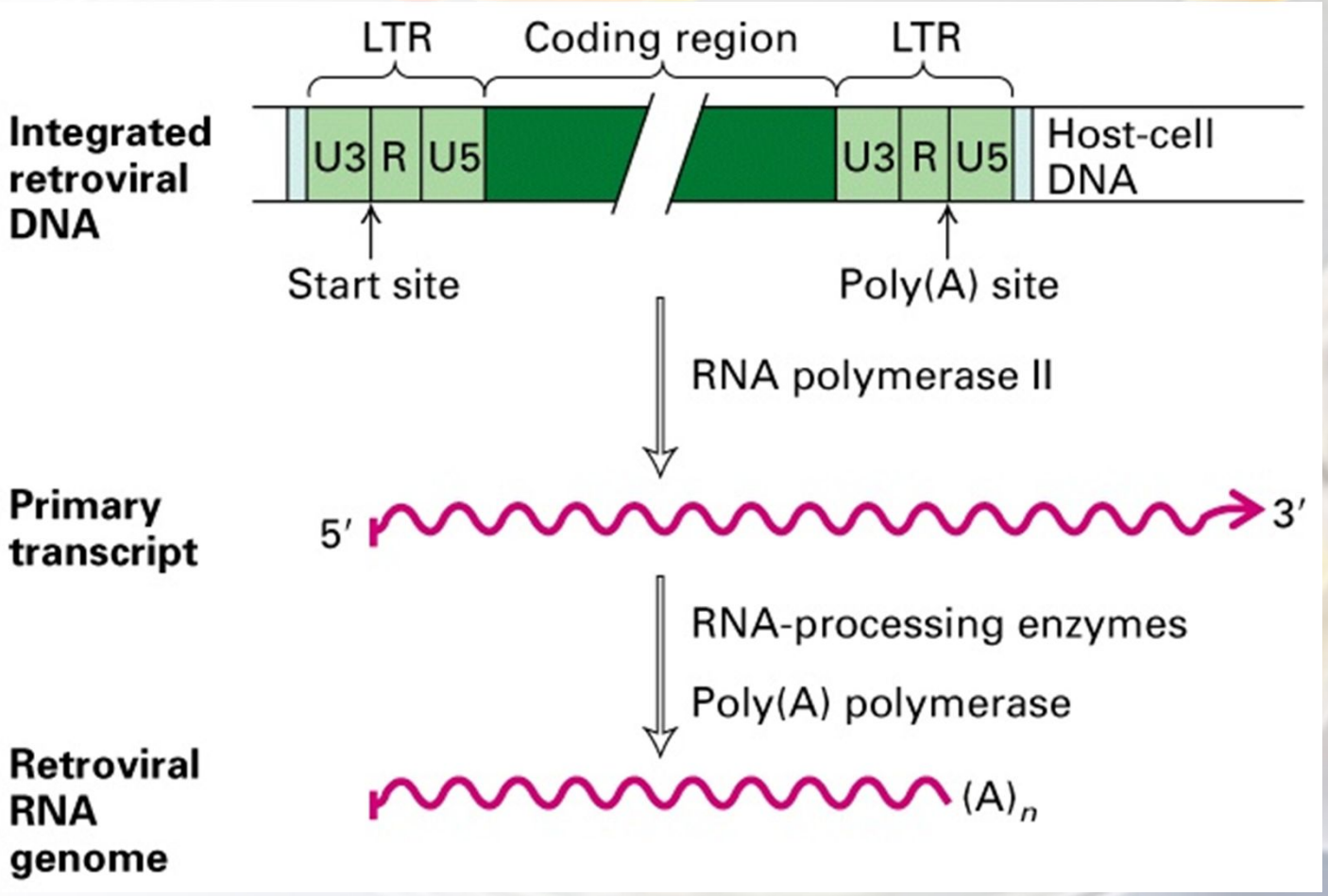
Long Terminal Repeats

Long Terminal Repeats



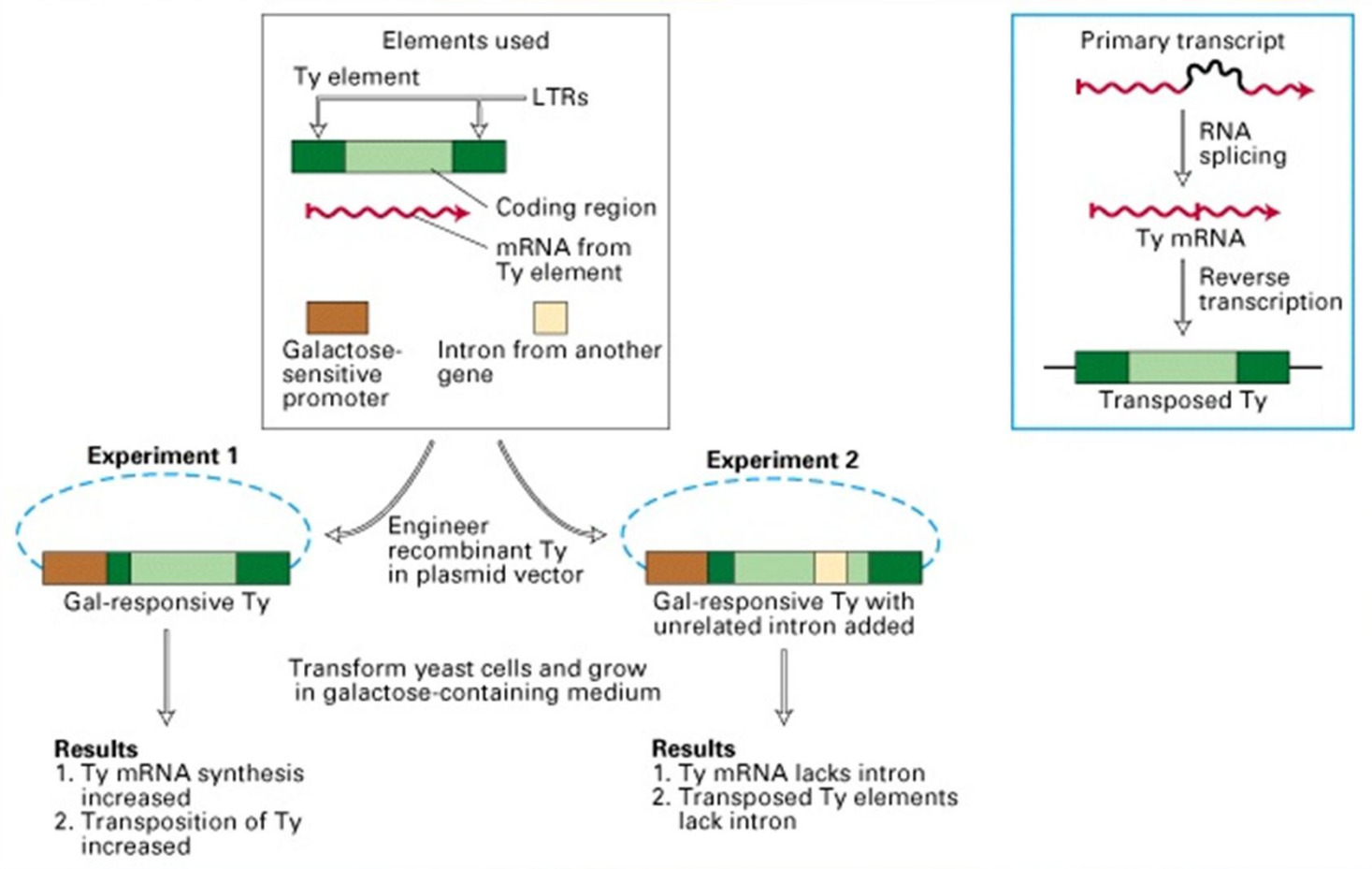
Long Terminal Repeats

Long Terminal Repeats



Long Terminal Repeats

Long Terminal Repeats



Long Terminal Repeats

Conclusion

- Eukaryotic retrotransposons fall into two major groups:
- LTR retrotransposons
- Non-LTR retrotransposons

Genetics and Genomics

Long Interspersed Nuclear Elements

and

Short Interspersed Nuclear Elements

LINEs and SINEs

LINEs and SINEs

- Non-LTR retrotransposons consist of two sub-types;
- Long Interspersed Nuclear Elements (LINEs)
- Short Interspersed Nuclear Elements (SINEs)

LINEs and SINEs

LINEs and SINEs

- High copy numbers, in the plants species
- Widespread in eukaryotic genomes
- LINEs possess two ORFs, which encode all the functions needed for retrotransposition

LINEs and SINEs

LINEs

- Functions of ORFs include
- Reverse transcriptase
- Endonuclease activities, in addition to a nucleic acid-binding property needed to form a ribonucleoprotein

LINEs and SINEs

SINEs

- SINEs, on the other hand, use the LINE machinery and function as non-autonomous retro-elements

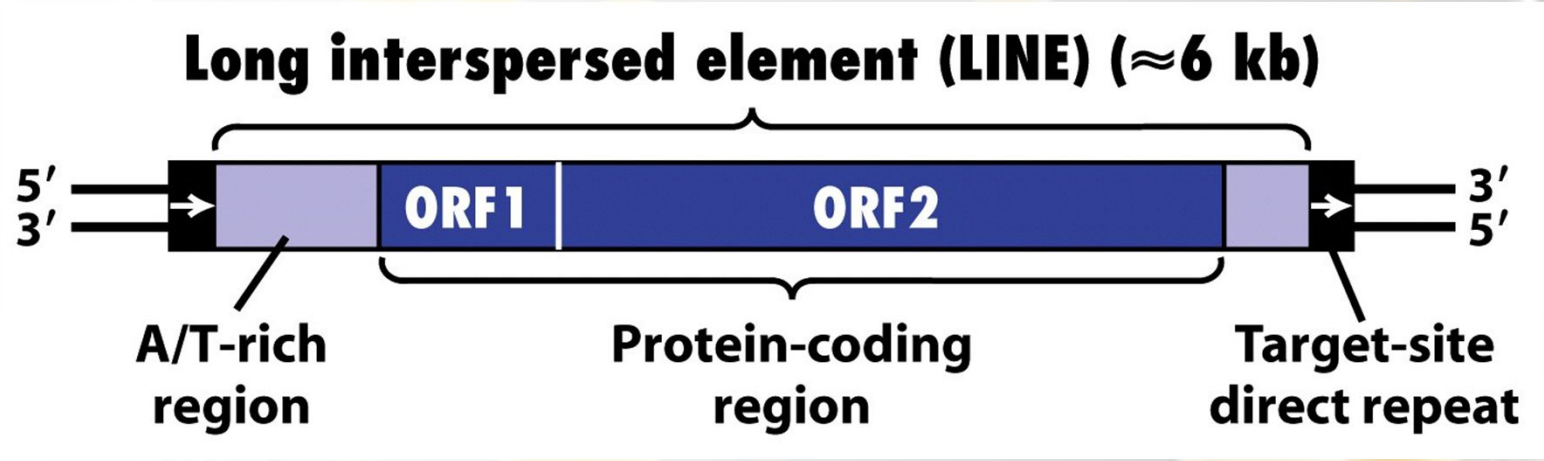
LINEs and SINEs

LINEs

- Several subgroups, such as L1, L2 and L3.
- Human coding L1 begin with an untranslated region (UTR) that includes an RNA polymerase II promoter, two non-overlapping open reading frames (ORF1 and ORF2), and ends with another UTR.
- ORF1 encodes an RNA binding protein and ORF2 encodes a protein having an endonuclease

LINEs and SINEs

LINEs



LINEs and SINEs

LINEs

- The 5' UTR contains the promoter sequence, while the 3' UTR contains a polyadenylation signal (AATAAA) and a poly-A tail
- Human genome contains about 500,000 LINEs

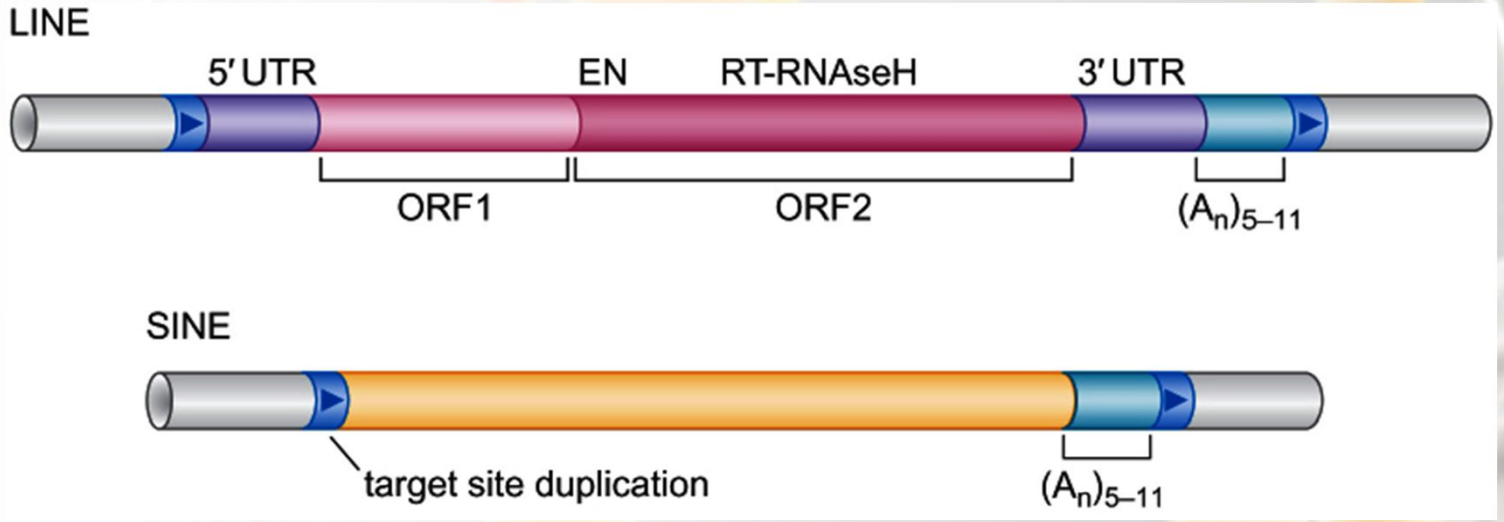
LINEs and SINEs

Short Interspersed Elements

- Short Interspersed Elements
- ~300 base pairs
- *Alu*
- ~ 11-13% of human genome

LINEs and SINEs

LINEs and SINEs



The sequences of LINE and SINE look like simple genes.

LINEs and SINEs

Short Interspersed Elements

- Short DNA sequences ~ 300 bases that represent reverse-transcribed RNA molecules
- SINEs do not encode a functional reverse transcriptase protein and rely on other mobile elements for transposition
- In some cases they may have their own endonuclease that will allow them to cleave their way into the genome

LINEs and SINEs

Short Interspersed Elements

- The most common SINEs in primates are called Alu sequences.
- Alu elements are approximately 350 base pairs long, do not contain any coding sequences, and can be recognized by the restriction enzyme AluI
- SINEs make up ~ 11-13% of human genome

LINEs and SINEs

LINEs and SINEs - Conclusion

- LINES
- 6-7 kb long
- SINES
- ~ 300bp long
- Example: Alu ~ 350 bp long

Genetics and Genomics

Genetic Variations

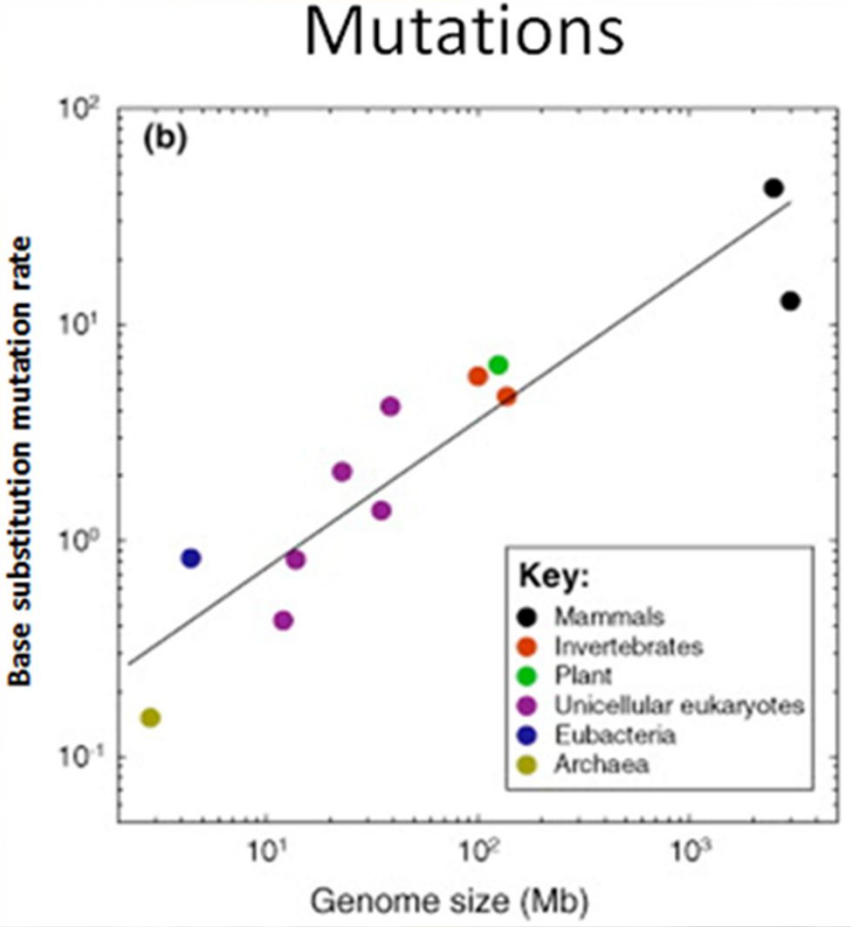
Genetic Variations

Genetic Variations

- Variation between individuals of a population (within species) can be due to differences in the nucleotide sequence
- How genetic variations emerge ?

Genetic Variations

Genetic Variations can be due to the Mutations



Genetic Variations

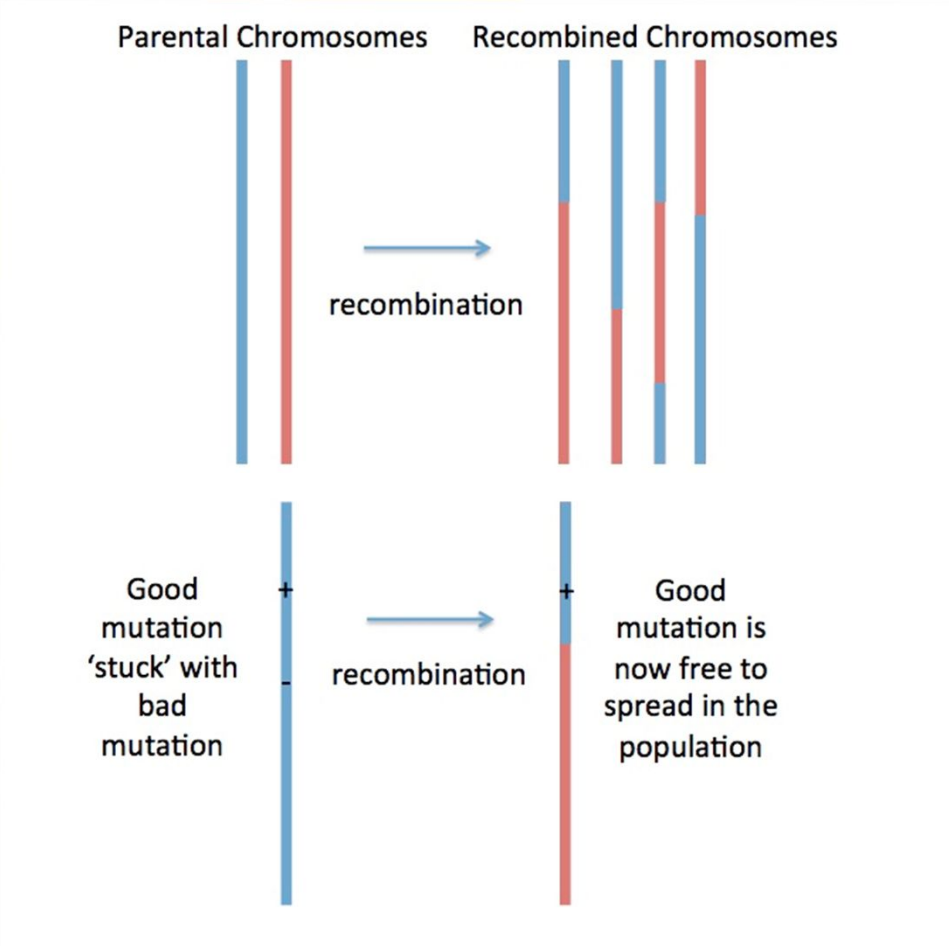
Genetic Variations: Recombinations

Recombination

Shuffling gene variants (alleles) in a population

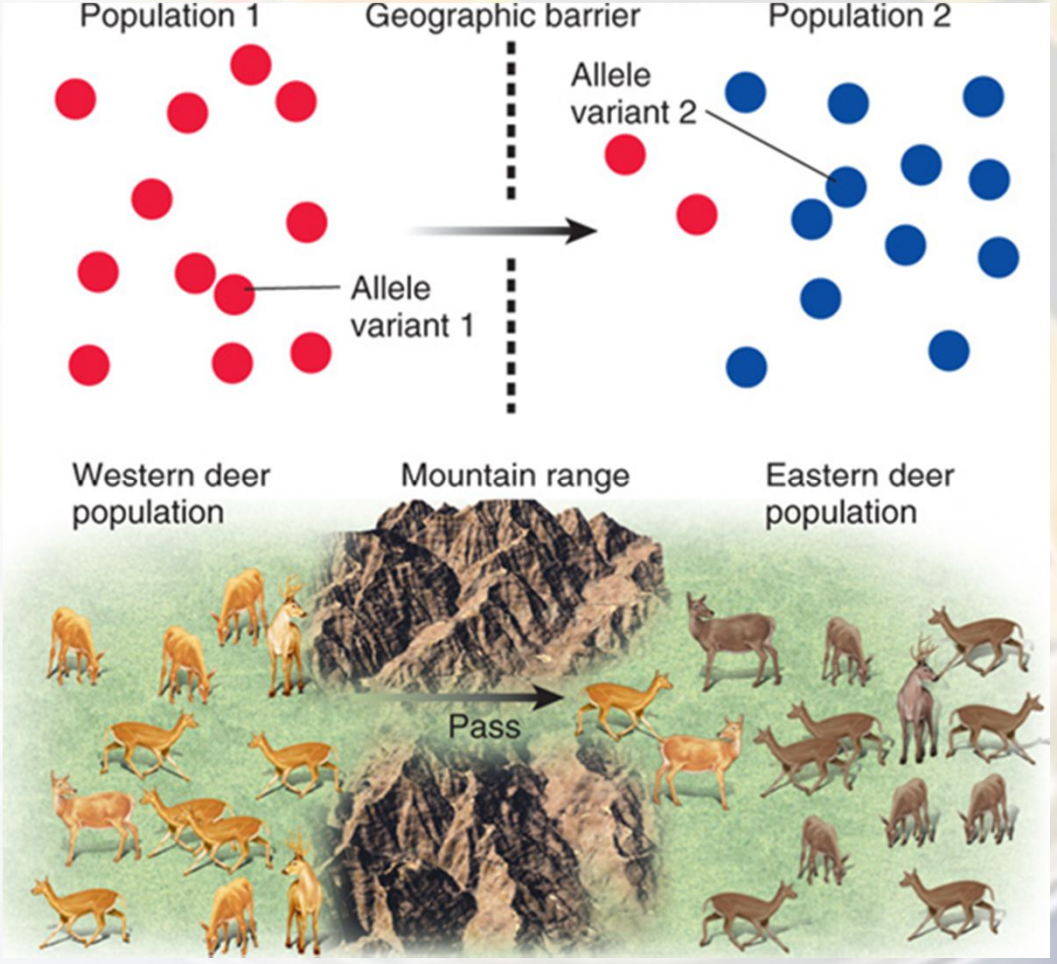
Genetic Variations

Genetic Variations: Recombinations



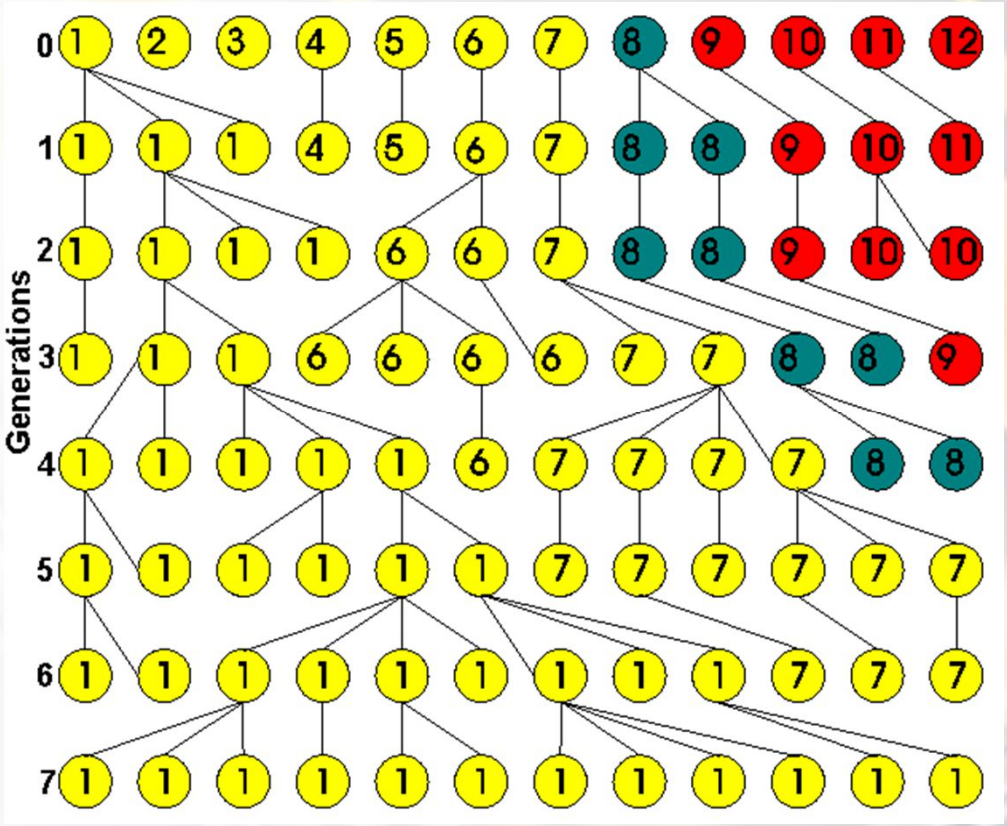
Genetic Variations

Genetic Variations: Genes Flow



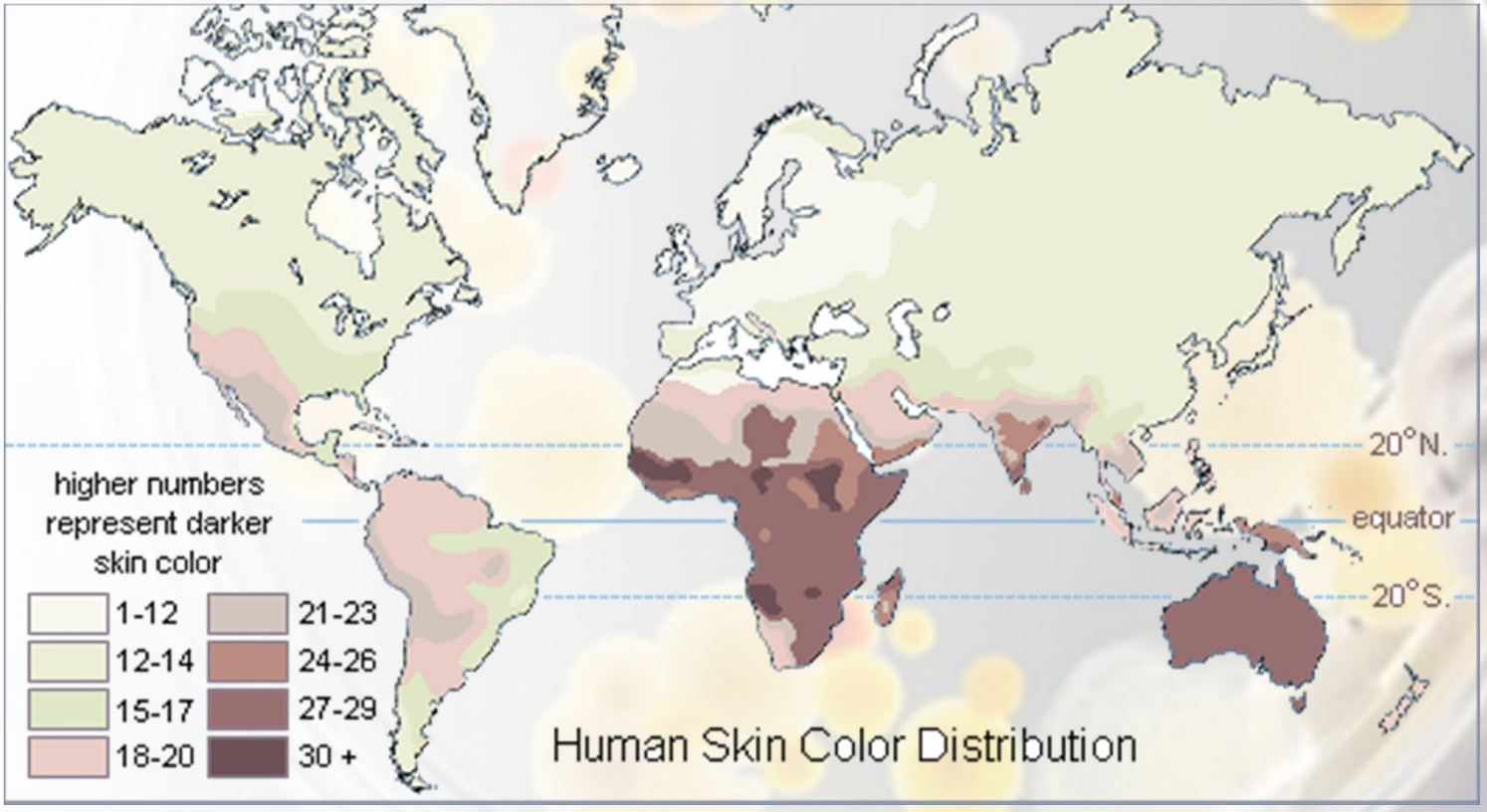
Genetic Variations

Genetic Variations: Genetic Drift: **Change in the frequency of a gene variant (allele) in a population due to random sampling of organisms**



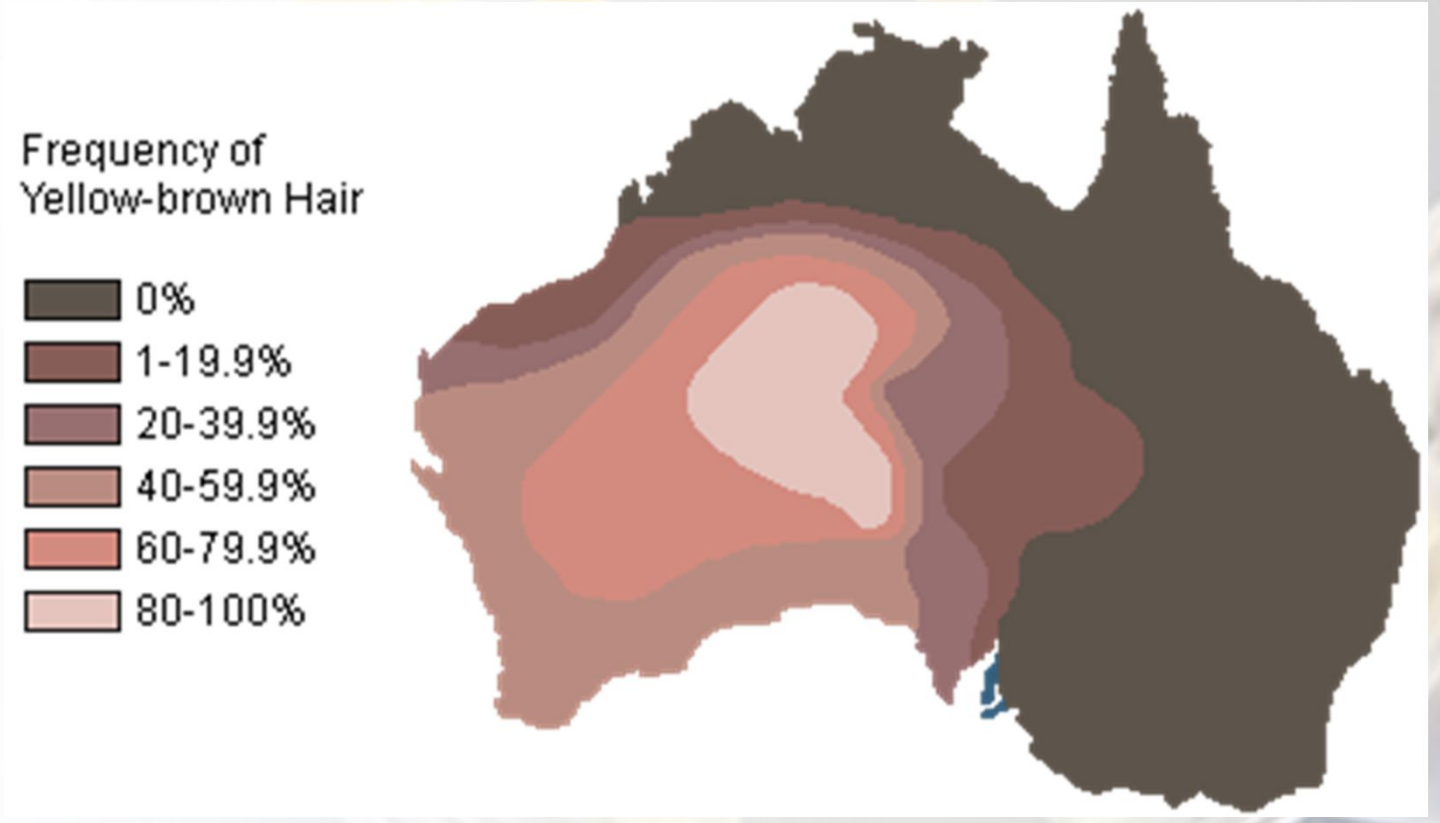
Genetic Variations

Genetic Variations: Humans Skin Color



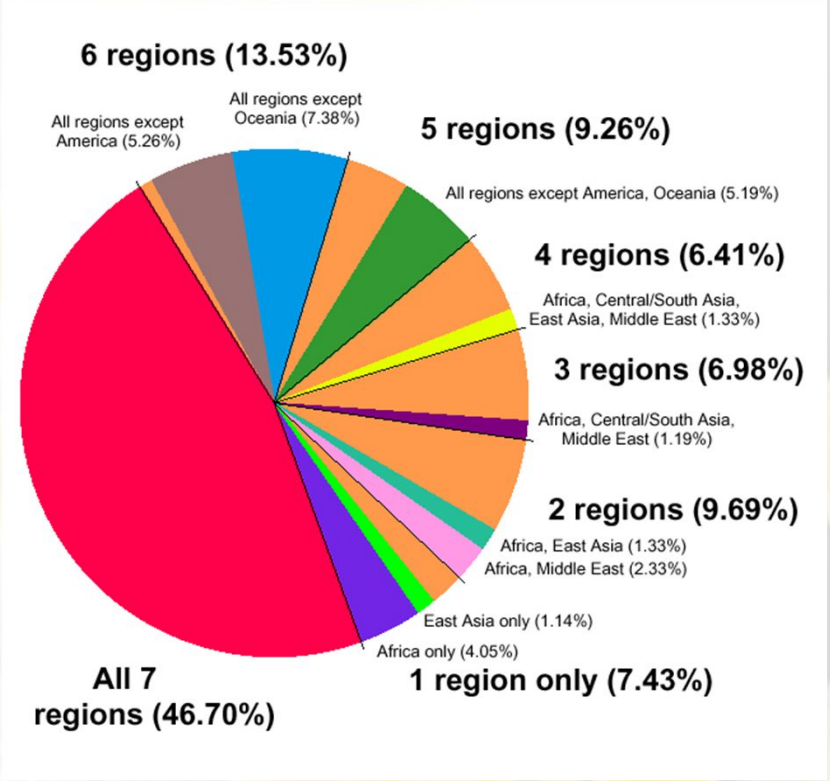
Genetic Variations

Genetic Variations: Yellow Brown Hair Australian natives



Genetic Variations

Genetic Variations: Microsatellites



377 autosomal microsatellites loci in 1056 individuals from 52 populations of seven regions

Genetic Variations

Genetic Variations - Humans

- Two genomes are roughly 99.9% identical to each other.
- If human genome is 3.0 billion bp then, there are 3.0 million differences between any two genomes.

Basic techniques

The solutions: basic techniques

- The fragments of DNA can be inserted to a suitable replicons-cloning vehicles
- Plasmids and bacteriophages are most suitable vectors

Basic techniques

Isolation of genomic DNA from *E. coli*

- Harvesting of cell pellet by centrifugation
- Cell lyses by SDS and proteinase K
- DNA extraction with phenol and chloroform
- Precipitation with isoprepenol

Basic techniques

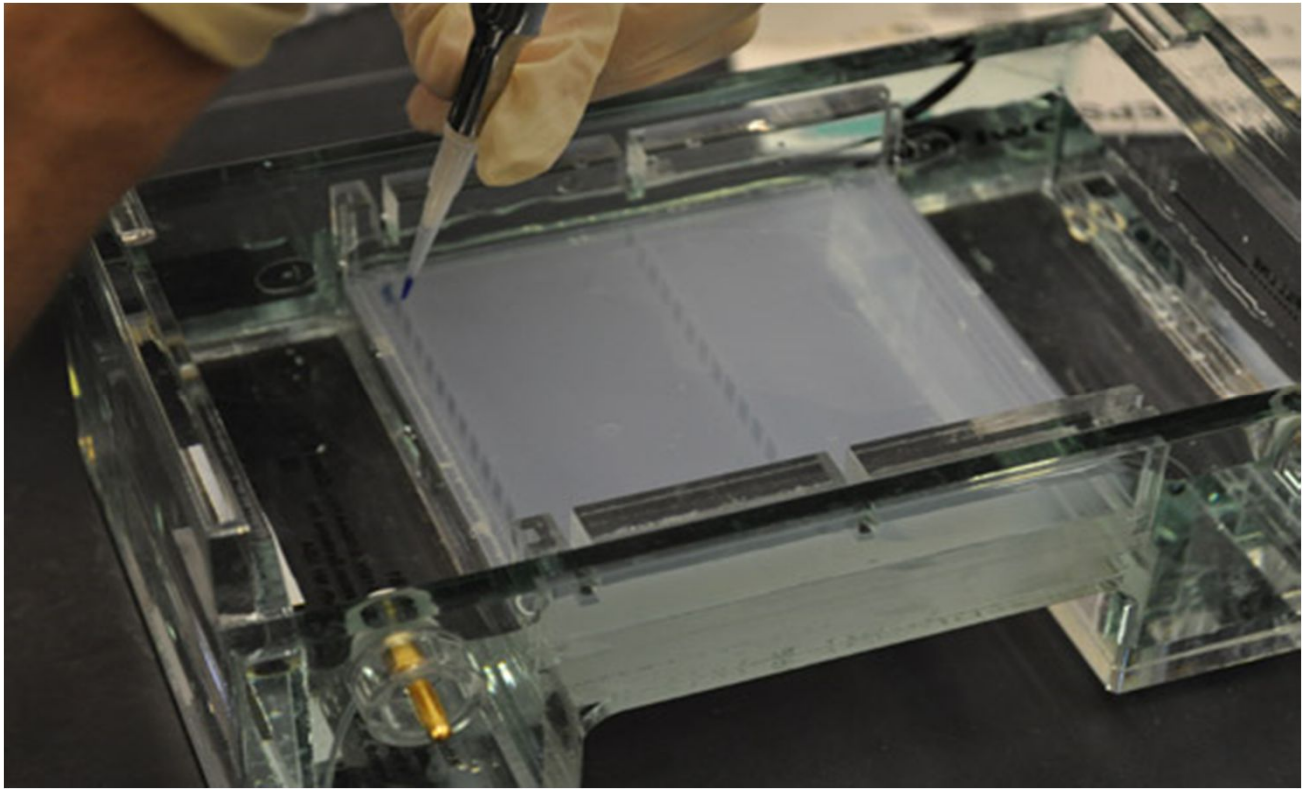
Agarose gel electrophoresis

- Agarose is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb.

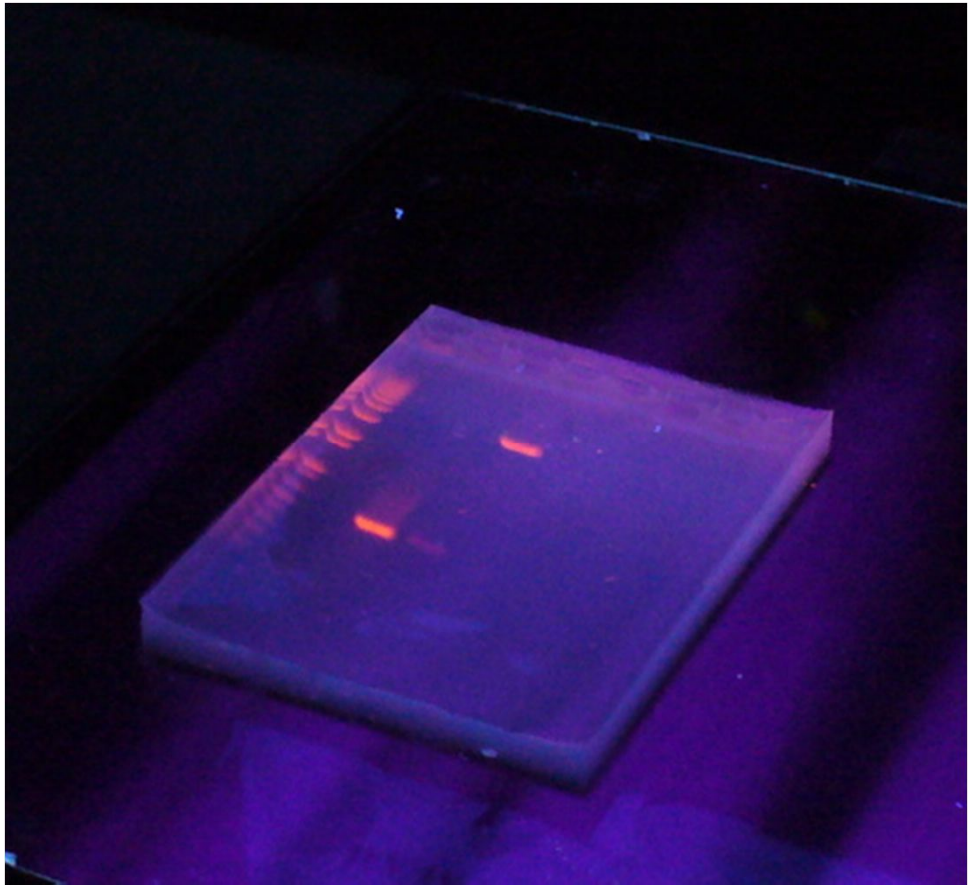
Basic techniques

Polyacrlamide gel electrophoresis

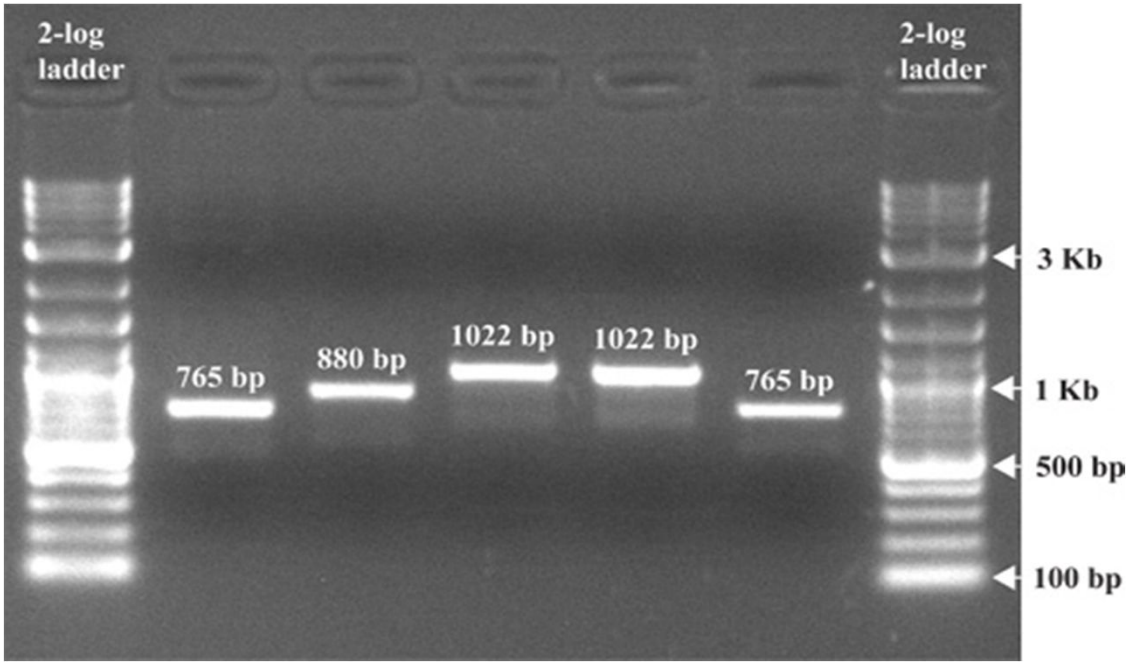
- **Polyacrlamide is preferred for smaller DNA fragments**



Gel electrophoresis apparatus



Visualization of gel under UV light



Visualization of gel on gel doc system

Basic techniques

Nucleic acid blotting

- Blotting describes the immobilization of sample nucleic acids or proteins on to a solid support, generally nylon or nitrocellulose membranes
- Southern blotting
- Northern blotting
- Western blotting

Basic techniques

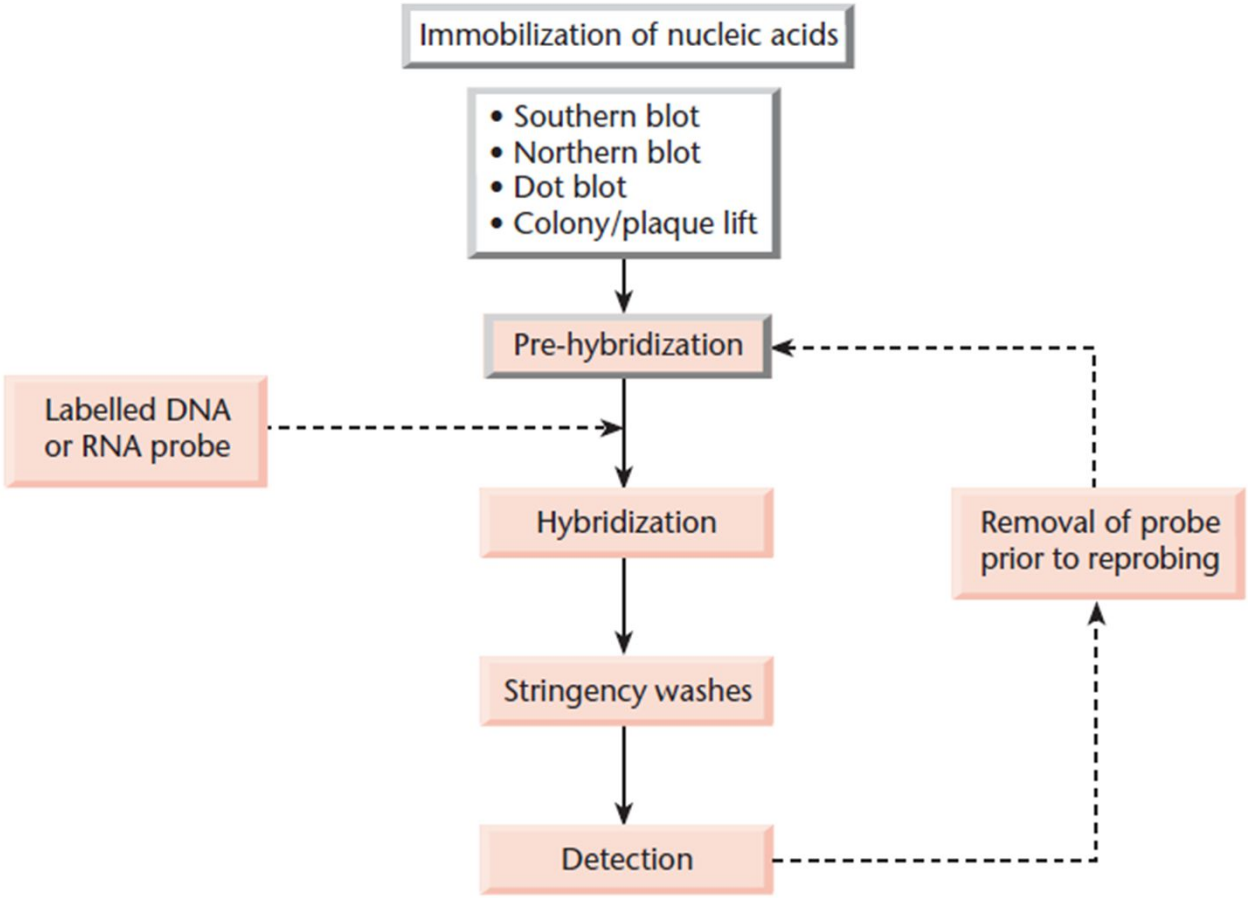
Nucleic acid Hybridization

- The hybridization of nucleic acids on membranes is a widely used technique in gene manipulation

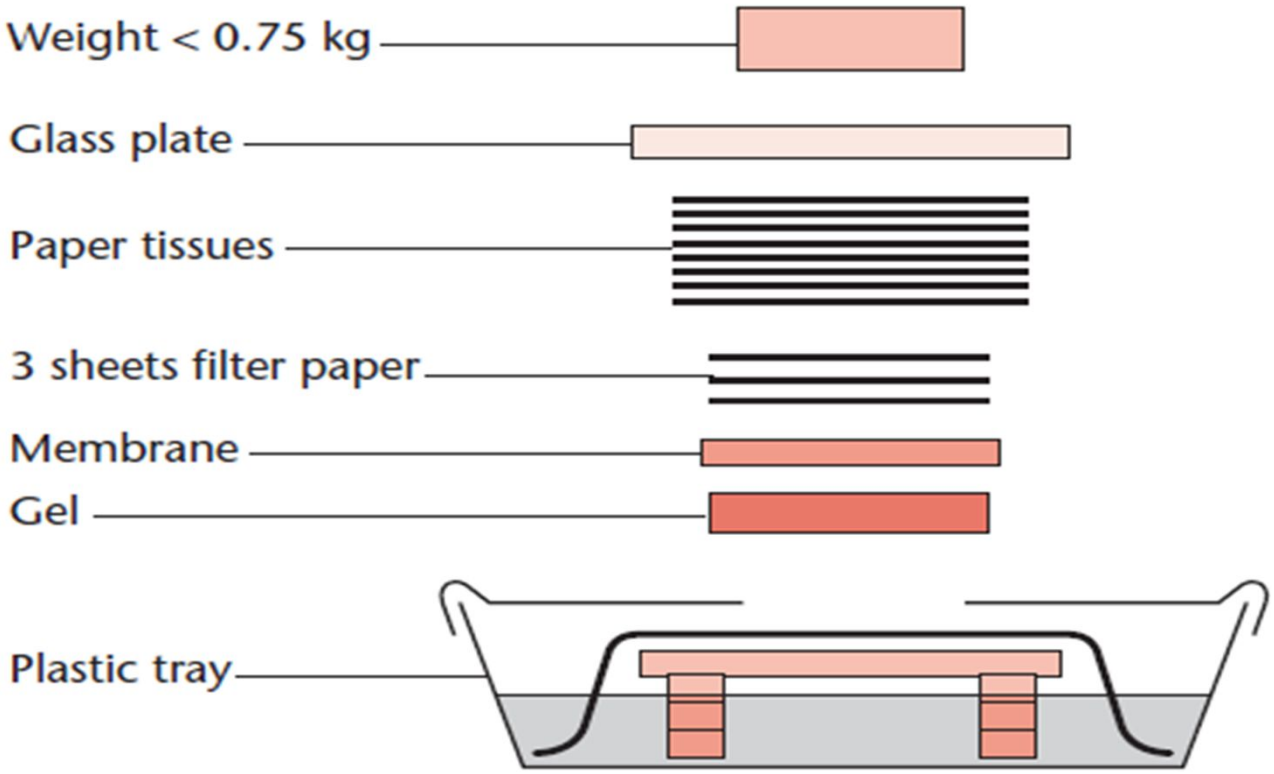
Basic techniques

Stringency control

- Stringency can be regarded as the specificity with which a particular target sequence is detected by hybridization probe
- Probe is a fragment of DNA or RNA which is used to detect the sequences of target DNA



Nucleic acid blotting and hybridization

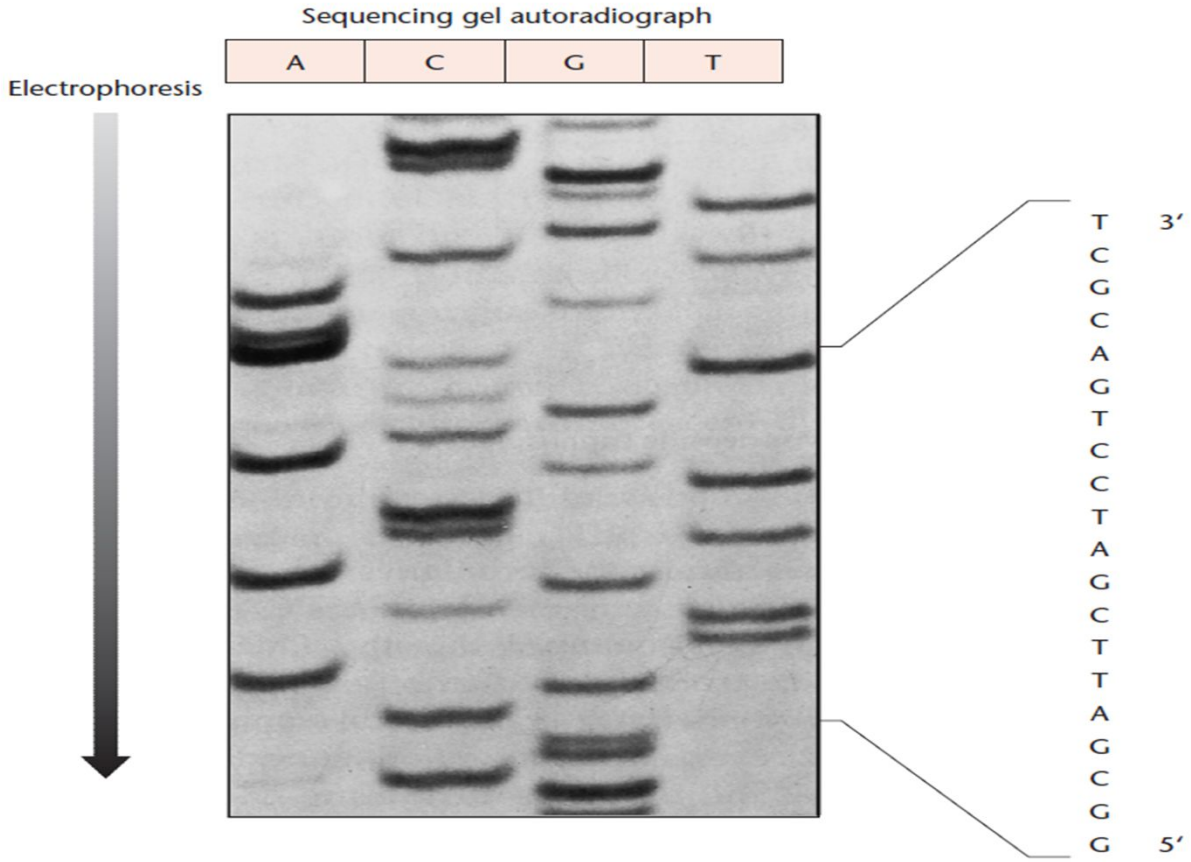


A typical capillary blotting appratus

Basic techniques

Autoradiography

- A technique using X-ray film to visualize fragments of molecules that have been radioactively labeled
- For example, it can be used to analyze the length and number of DNA fragments separated by gel electrophoresis



Autoradiograph of sequencing gel

Basic techniques

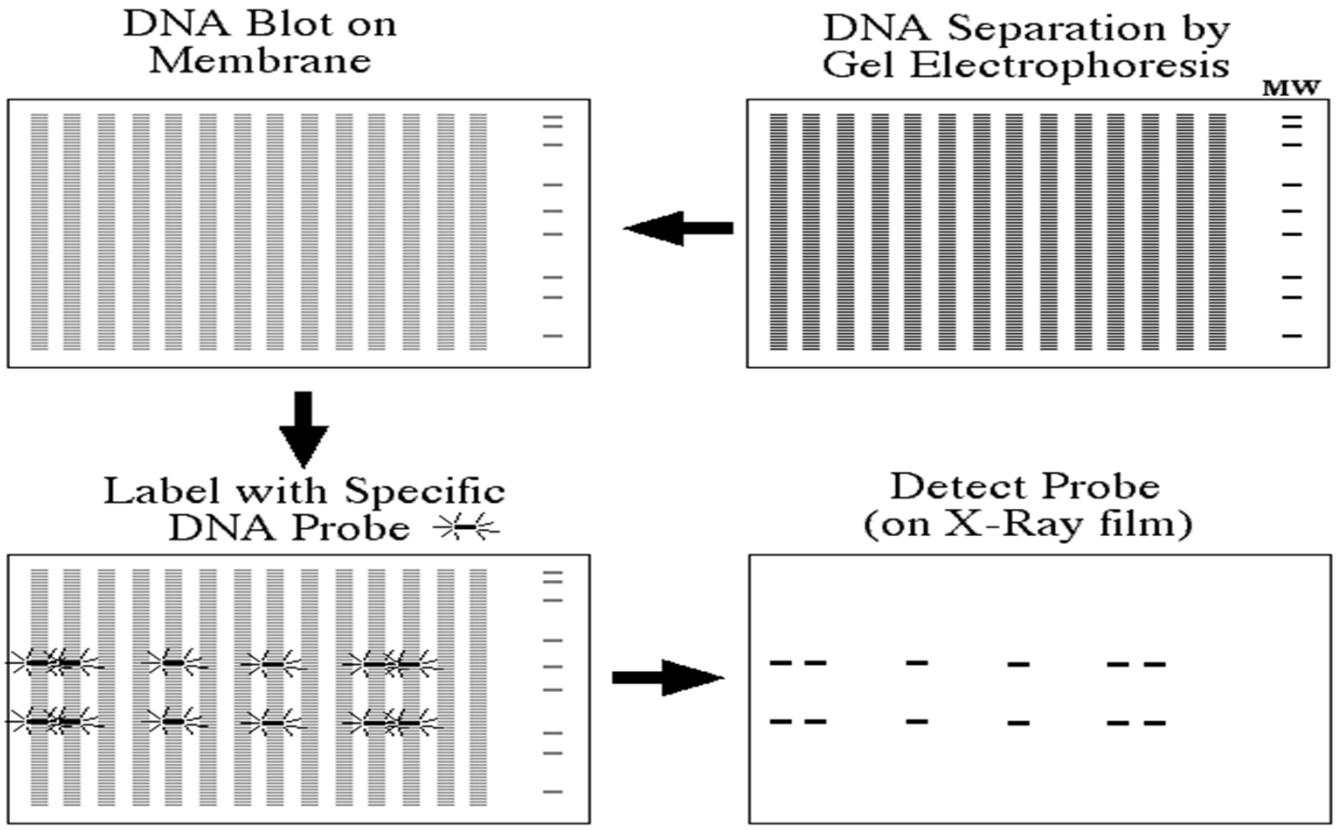
Southern blotting

- A Southern blot is a laboratory method used to detect specific DNA fragment from a mixture of DNA molecules
- The technique was named after its inventor, Edward Southern (1975)
- For efficient blotting, gel pretreatment is important

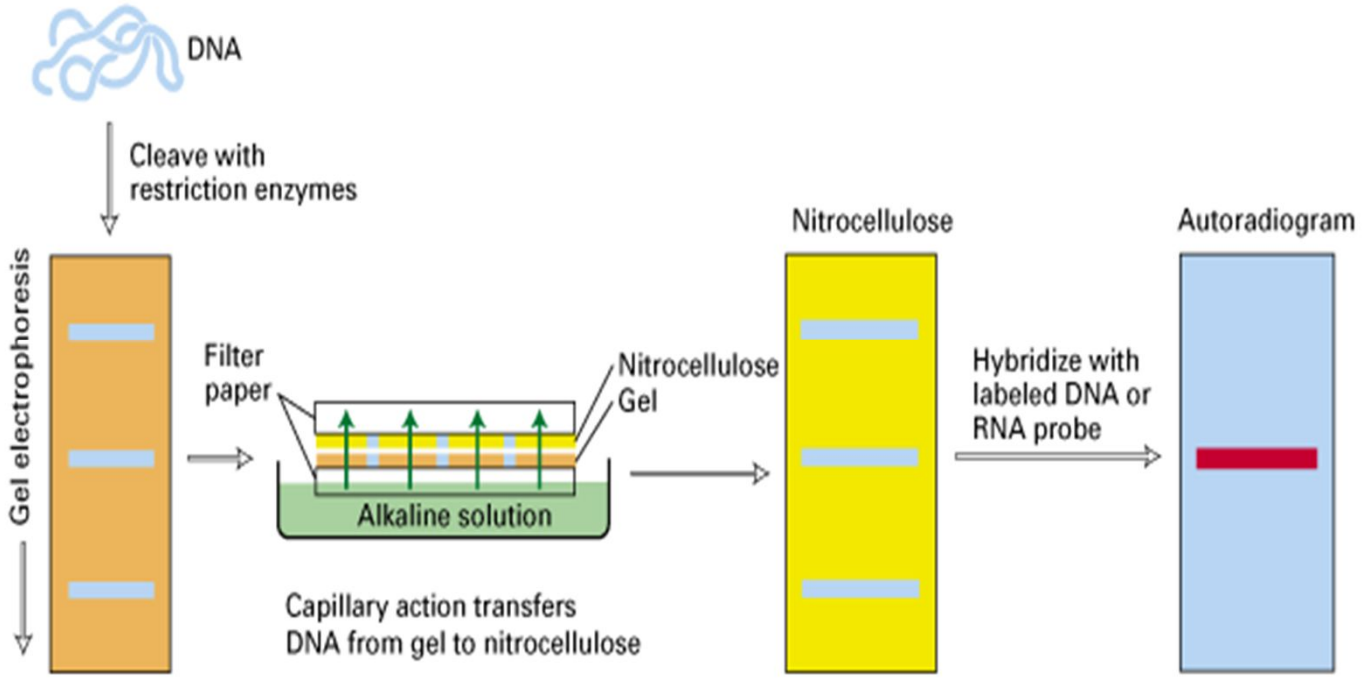
Basic techniques

Southern blotting

- After transfer, the nucleic acid needs fixation
- After fixation, the membrane is placed in a solution of labelled probe
- After hybridization reaction, membrane is washed and hybridization are detected by autoradiography



Overview of Southern blotting

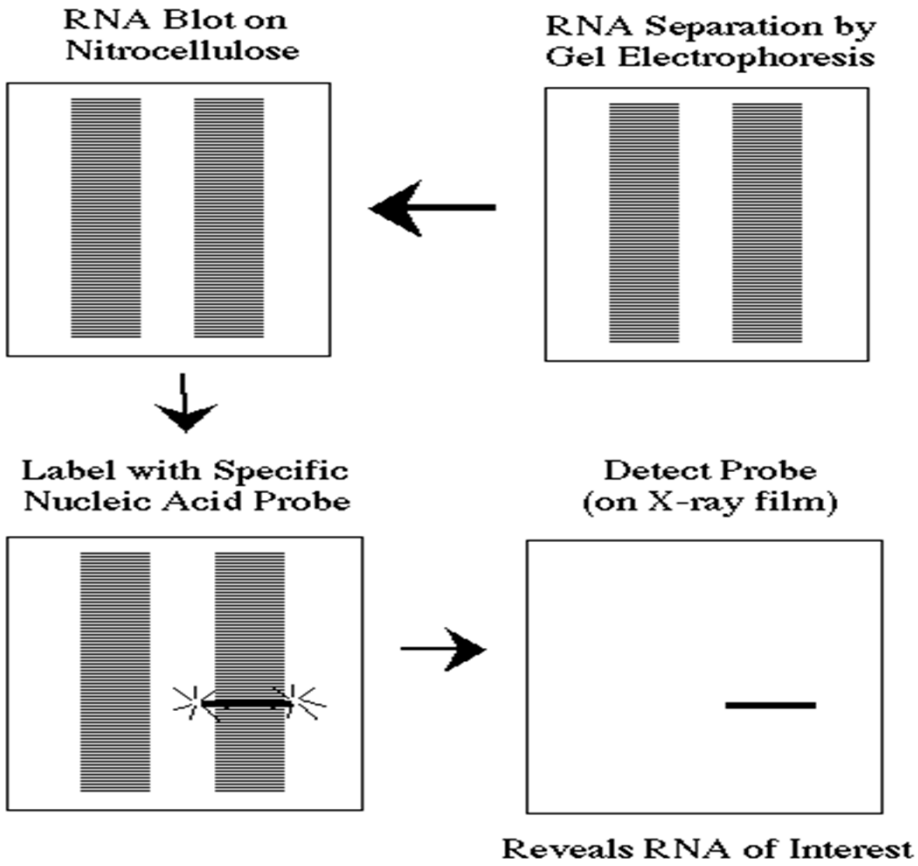


Southern blot hybridization

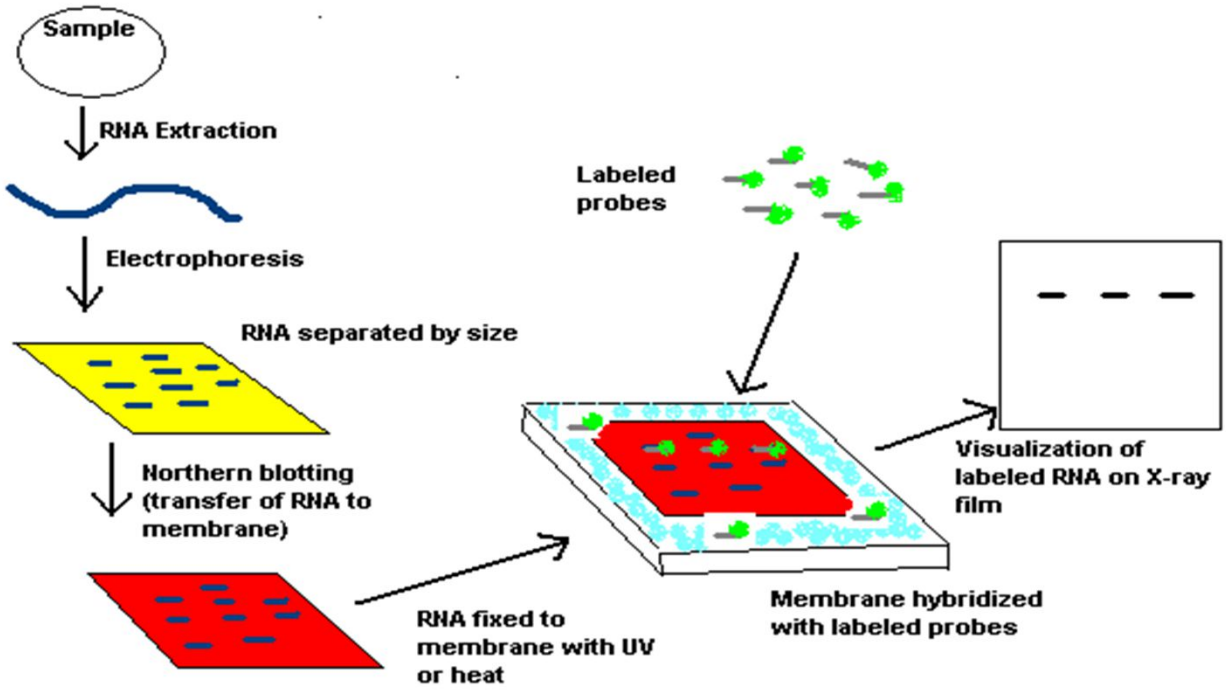
Basic techniques

Northern blotting

- A northern blot is a method used to detect specific RNA molecules among a mixture of RNA
- Northern blotting can be used to analyze a sample of RNA from a particular tissue or cell type in order to measure the RNA expression of particular genes



Overview of Northern blotting

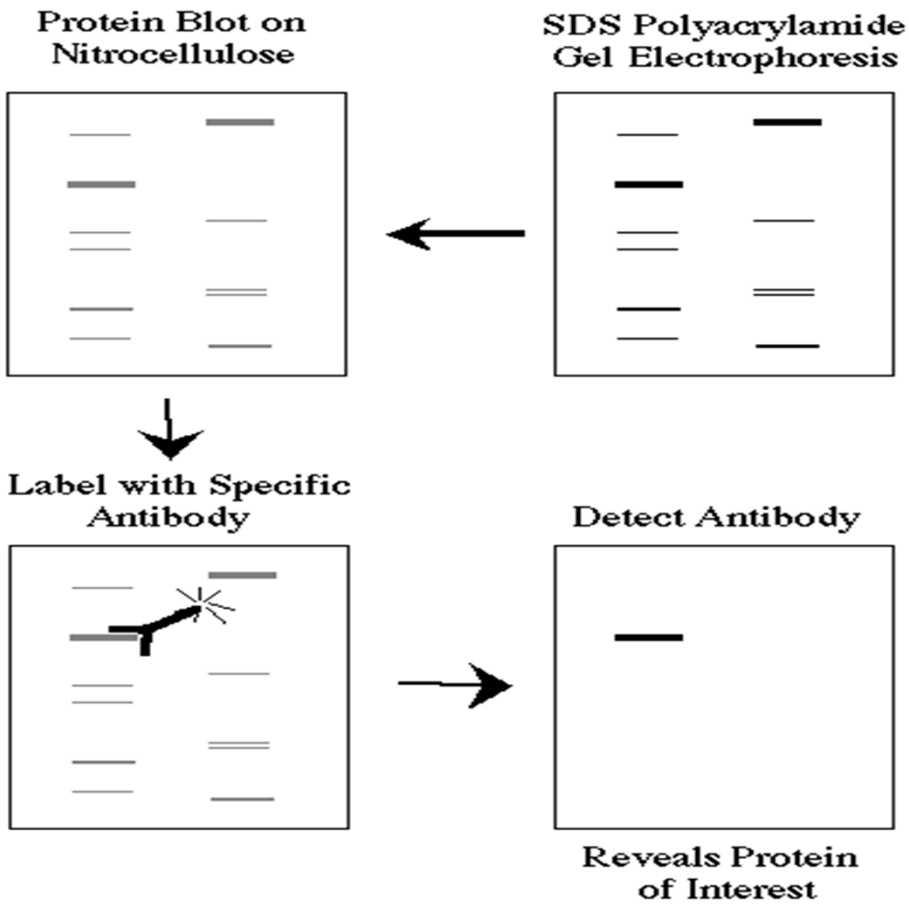


RNA detection by Northern blotting

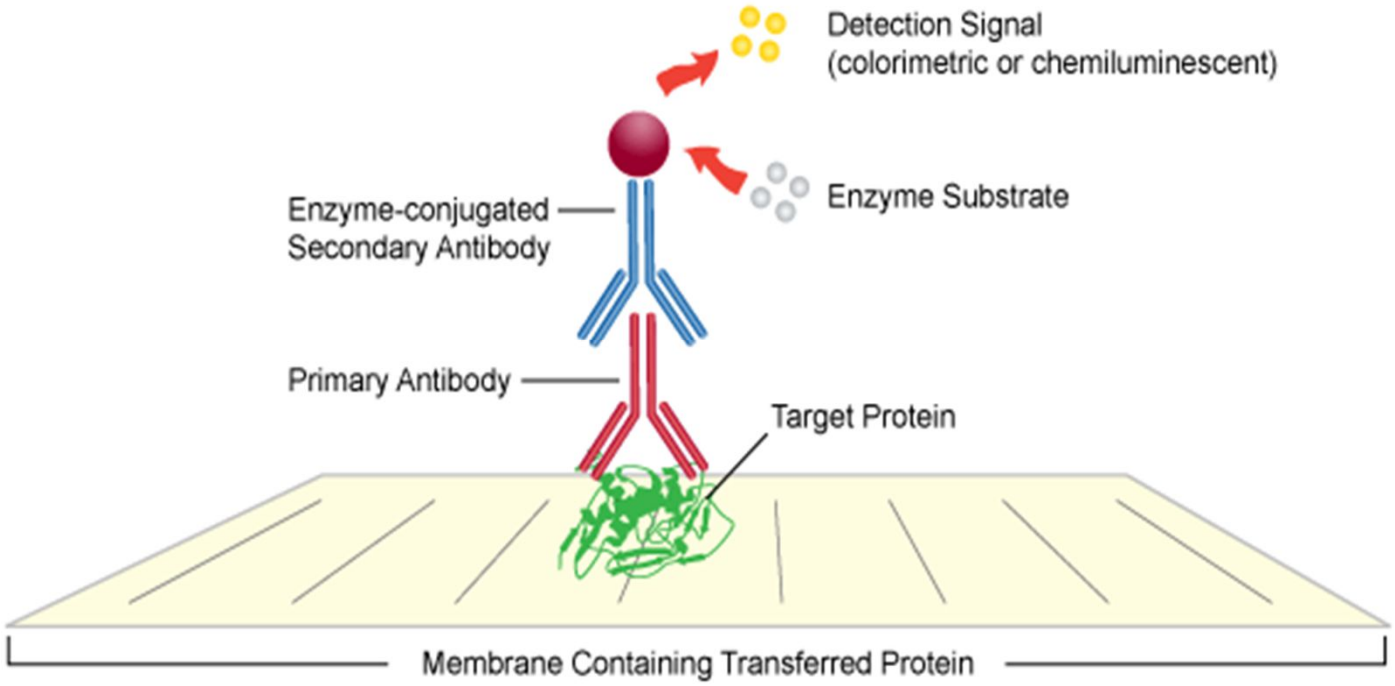
Basic techniques

Western blotting

- It involves the transfer of electrophoresed protein bands from a gel on to a membrane
- The bounded proteins are then available for analysis
- Most commonly, antibodies are used to detect specific antigens



Western blot method



Detection in Western blot

Basic techniques

Transformation of *E. coli*

- **Transformation is the process of getting the recombinant vector from a reaction mixture or vector solution into *E. coli* cells**
- **To enable the cells to take up circular vector DNA they have to be made competent**

Basic techniques

Electroporation

- A rapid and simple technique for introducing cloned genes into a wide variety of microbial, plant and animal cell, including *E. coli* is electroporation
- In electroporation, electric field is applied to cells in order to increase permeability of the cell membrane to take DNA

Basic techniques

Transformation with other organisms

- ***E. coli*** **often remains the host of choice for cloning**
- ***Bacillus subtilis*** **is a particularly attractive alternative cloning host**
- **Animal cells, protoplasts of yeast and plant are susceptible to transformation by liposomes**

Basic techniques

Polymerase chain reaction (PCR)

- The impact of PCR upon molecular biology has been profound
- The reaction is easily performed and leads to the amplification of specific DNA by an enormous factor

Cycle number	Number of double-stranded target molecules
1	0
2	0
3	2
4	4
5	8
6	16
7	32
8	64
9	128
10	256
11	512
12	1024
13	2048
14	4096
15	8192
16	16,384
17	32,768
18	65,536
19	131,072
20	262,144
21	524,288
22	1,048,576
23	2,097,152
24	4,194,304
25	8,388,608
26	16,777,216
27	33,554,432
28	67,108,864
29	134,217,728
30	268,435,456

Theoretical PCR amplification of a target fragment

Cutting and Joining DNA molecules

Cutting DNA molecules

- Before 1970 there was no method of cleaving DNA at discrete points
- Mechanical shearing was used for DNA fragmentation
- During 1960s phage biologist elucidated the phenomenon of restriction and modification

Cutting DNA molecules

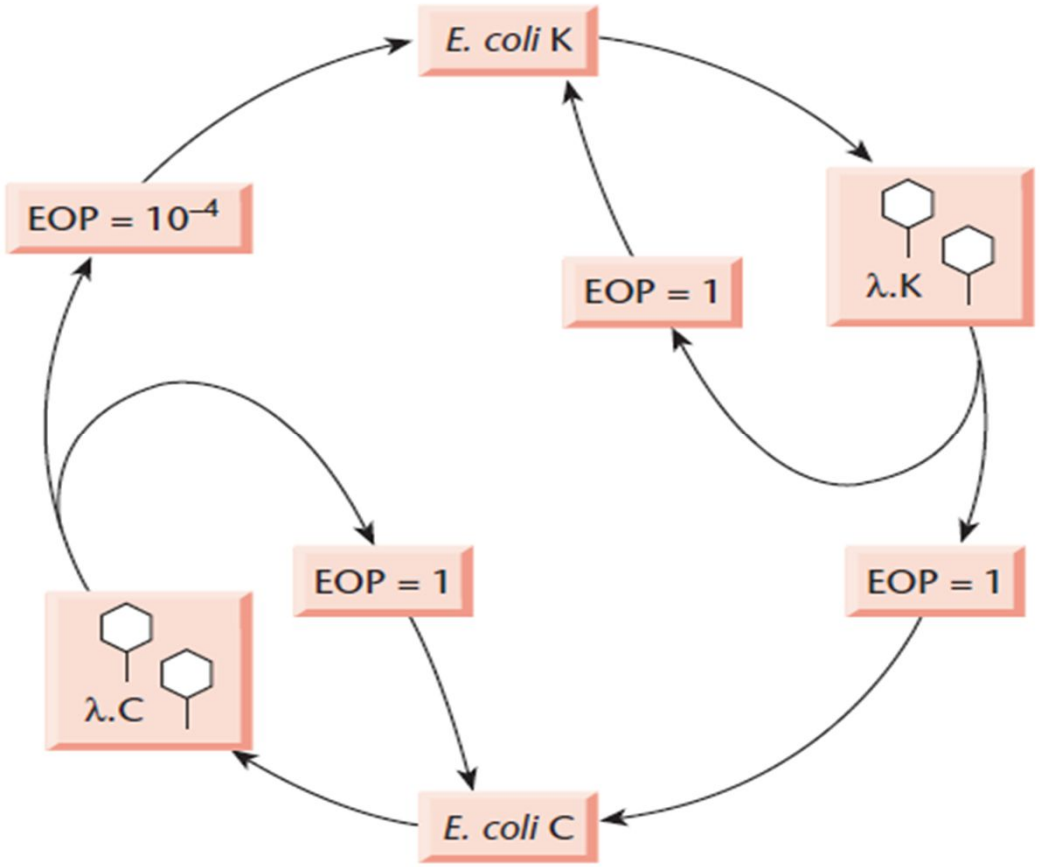
Host-controlled restriction/ modification

- **Restriction system allow bacteria to monitor the origin of incoming DNA**
- **When the incoming DNA is a bacteriophage genome, the effect is to reduce the efficiency of plating i.e. to reduce the number of plaques in plating test**

Cutting DNA molecules

Host-controlled restriction/ modification

- The phenomenon of restriction and modification were well illustrated and studied by the behavior of phage λ on two *E. coli* strains
- *E. coli* C and *E. coli* K



Host-controlled restriction and modification of phage λ

Cutting DNA molecules

Host-controlled restriction/ modification

- Restrictive host must, of course protects its own DNA from restriction endonucleases by modification
- Modification involves methylation of certain bases that constitute recognition sites

Cutting DNA molecules

Types of restriction and modification (R-M) system

- At least four R-M systems are known
- Type I
- Type II
- Type III
- Type IIs

Cutting DNA molecules

Type I

- Type I systems were the first to be characterized from *E. coli* K12
- The active enzyme consists of two restriction subunit, two modification subunit and one recognition subunit
- Type I systems are of little value for gene manipulation

Cutting DNA molecules

Type II

- Most of the useful R-M system is Type II
- Type II enzymes recognize a defined sequence and cut within it

Cutting DNA molecules

Type III

- Type III enzymes have symmetrical recognition sequences but otherwise resemble type I systems and are of little value

Cutting DNA molecules

Type IIs

- Type IIs systems have similar cofactors and structure to type II but restriction occurs at a distance from recognition site that limits their usefulness

Cutting DNA molecules

Nomenclature

- A suitable system was proposed by Smith and Nathans (1973)
- The species name of host organisms is identified by the first letter of genus and first two letters of specific epithet
- *E. coli* = Eco
- *H. influenzae* = Hin

Cutting DNA molecules

Nomenclature

- **Strain identification is written as *Eco*K**
- **In case, host strain has several restriction and modification systems, these are identified by roman numerals, for example, in case of *H. influenzae***
- ***HindI, HindII, HindIII* etc**

Cutting DNA molecules

Nomenclature

- **All restriction enzymes have general name endonuclease R and modification-methylase M followed by the system name, for example, in case of *H. influenzae***
- **R. *HindIII* or M. *HindIII***

Examples of restriction endonuclease nomenclature

Enzyme	Enzyme source	Recognition sequence
SmaI	***Serratia marcescens***, 1st enzyme	**CCCGGG**
HaeIII	***Haemophilus aegyptius***, 3rd enzyme	**GGCC**
HindII	***H. influenzae***, strain d, 2nd enzyme	**GTPyPuAC**
HindIII	***H. influenzae***, strain d, 3rd enzyme	**AAGCTT**
HamHI	***Bacillus amyloliquefaciens***, strain H, 1st enzyme	**GGATCC**

Cutting DNA molecules

Target sites

- **Type II endonucleases recognize and cleave DNA within particular sequences of four to eight nucleotides which have a twofold axis of rotational symmetry i.e. referred as *palindromes***

5′-GAATTC-3′

5′-CTTAAG-3′

5′-G A A T T C-3′

3′-C T T A A G-5′

5′-G/ A A* T T C-3′

3′-C T T A* A/ G-5′

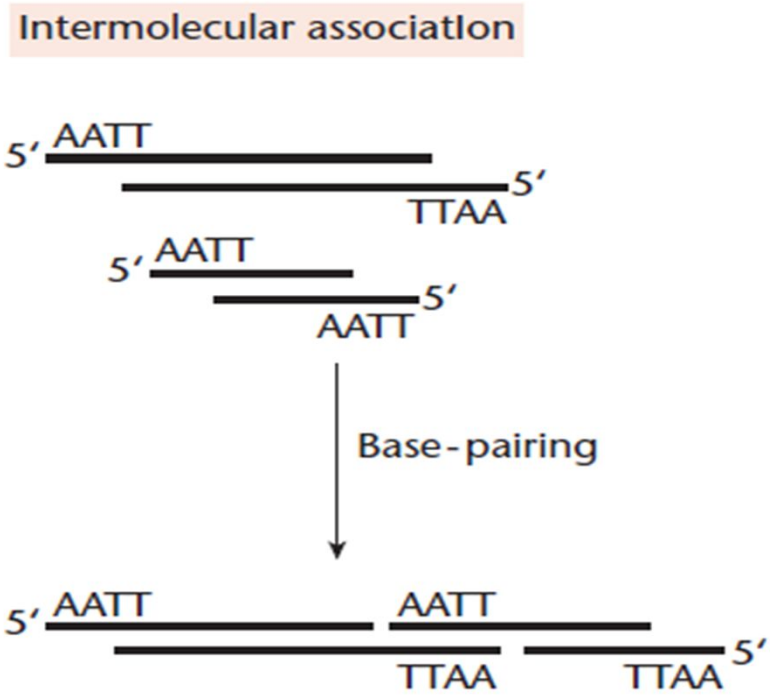
5′-G

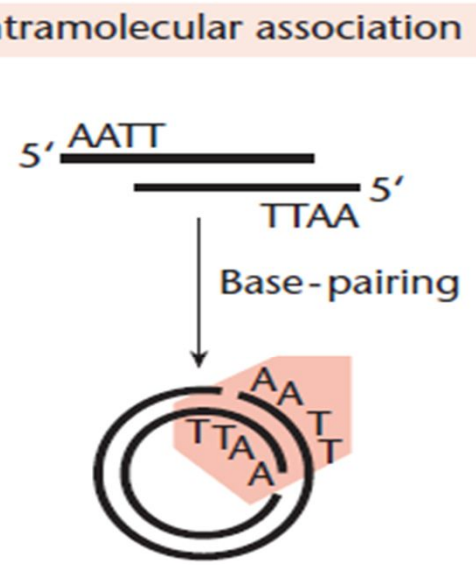
3′-C T T A A-5′

5′- A A T T C-3′

G-5′

Single stranded breaks by *Eco*R1





Cohesive fragments of DNA produced by digestion with EcoR1

Cutting DNA molecules

Number and size of restriction fragments

- **The number and size of fragments generated by restriction enzyme depend on the frequency of occurrence of the target site in the DNA to be cut**

Cutting DNA molecules

Number and size of restriction fragments

- Four base recognition site occurs every 4^4 (256) bp
- Six base recognition site occurs every 4^6 (4096) bp
- Eight base recognition site occurs 4^8 (65,536) bp

Average fragment size (bp) produced by different enzymes

Enzyme	Target	Arabidopsis	E. Coli	Human
*Apa*I	GGGCCC	25000	15000	2000
*Bam*HI	GGATCC	6000	5000	5000
*Spe*I	ACTAGT	8000	60000	10000

Cutting DNA molecules

Summary of restriction endonucleases

- **Type II restriction enzymes are heavily responsible for the current explosion in the field of gene manipulation in that they are essential in forming recombinant DNA molecules**

Joining DNA molecules

DNA modifying enzymes

- Nucleases
- DNA Polymerase
- Reverse transcriptases
- DNA ligases

Joining DNA molecules

Nucleases

- Nucleases or DNases are the enzymes that degrade DNA
- Two broad classes of nucleases
  - i). Exonucleases
  - ii). Endonucleases

Joining DNA molecules

DNA Polymerase

- An enzyme that catalyzes template-dependent synthesis of DNA

Joining DNA molecules

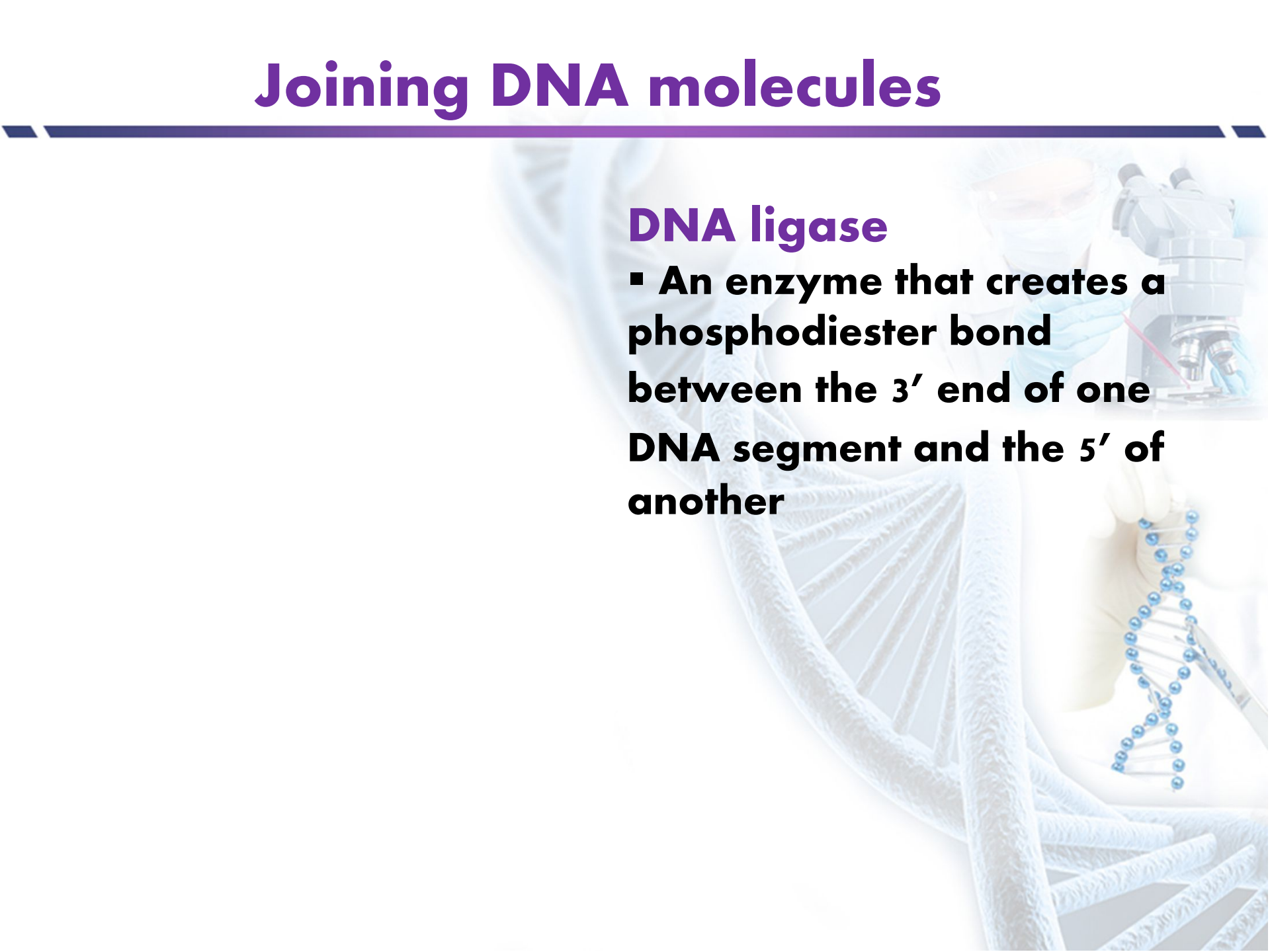
Reverse transcriptases

- **An RNA-directed DNA polymerase in retroviruses; capable of making DNA complementary to an RNA**

Joining DNA molecules

DNA ligase

- An enzyme that creates a phosphodiester bond between the 3′ end of one DNA segment and the 5′ of another



Joining DNA molecules

Methods of joining DNA fragments

- **Mainly three methods are used for joining DNA in vitro**

i). Joining covalently annealed cohesive ends by DNA ligase

ii). Joining blunt-ended fragments by DNA ligase from phage T4 infected *E. coli*

Joining DNA molecules

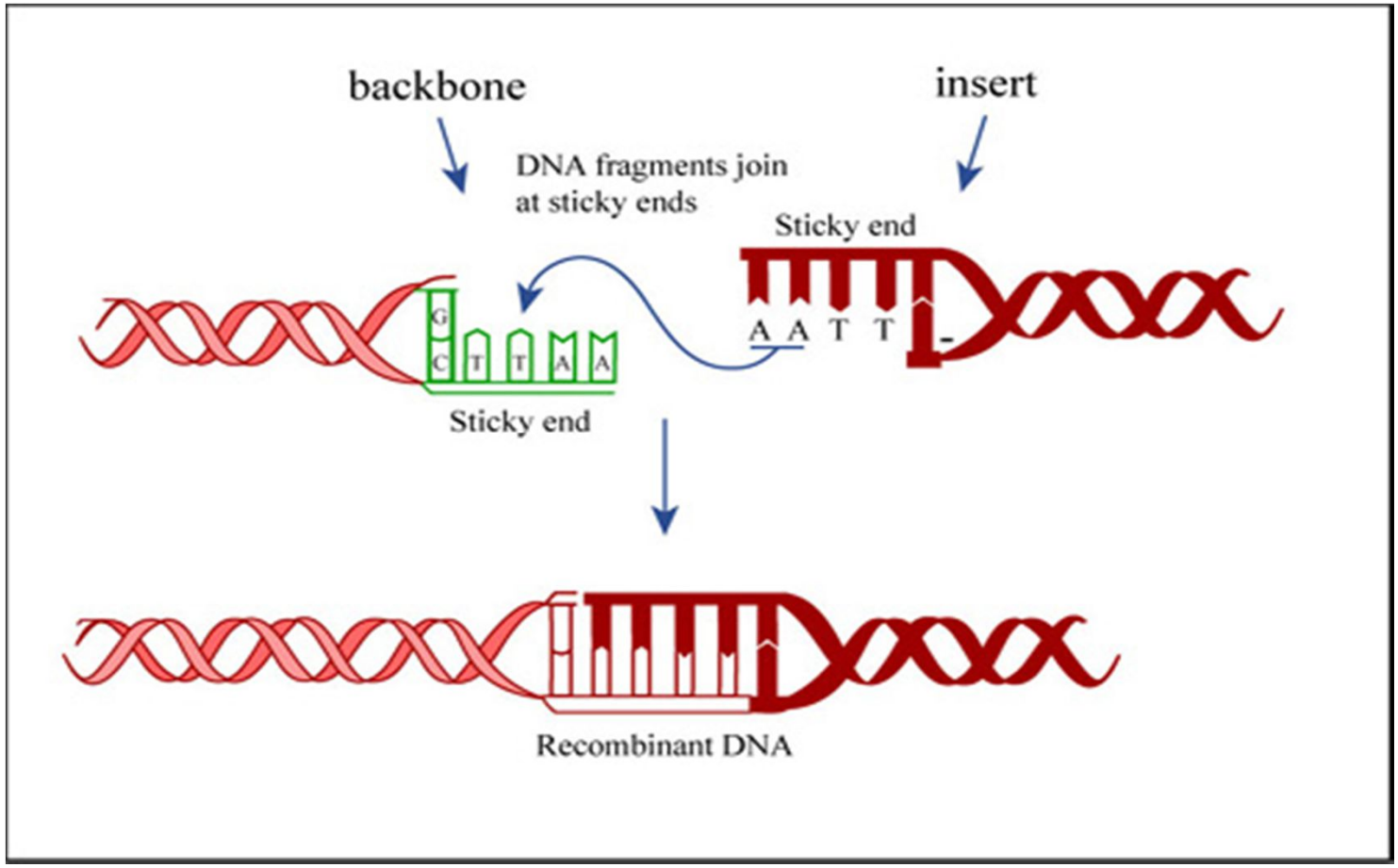
Methods of joining DNA fragments

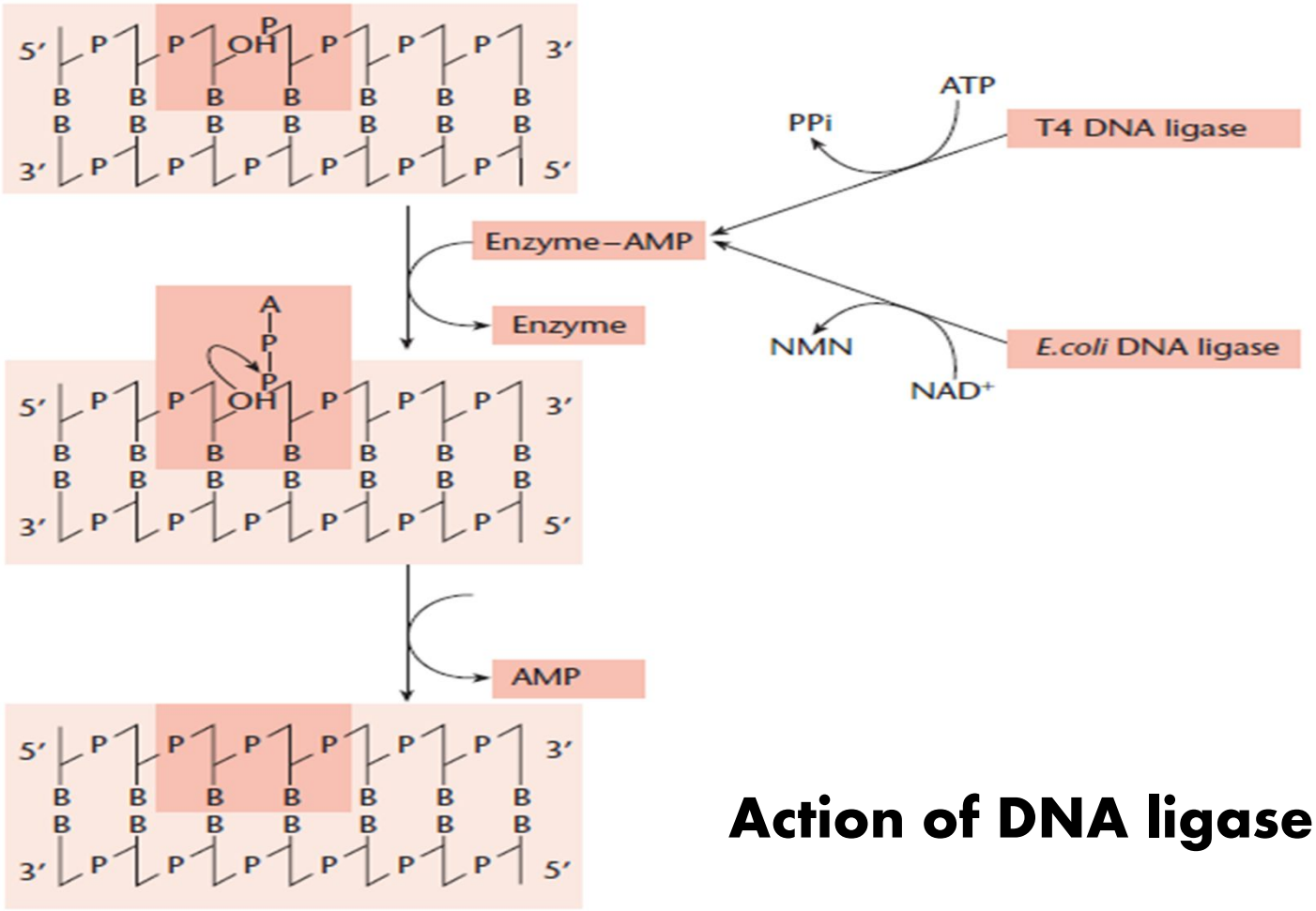
iii). The third utilizes the enzyme terminal deoxynucleotidyltransferase to synthesize homopolymeric 3′ single stranded tails at the ends of fragments

Joining DNA molecules

DNA ligase

- ***E. coli*** **and phage T4 encode enzyme, DNA ligase, which seals single stranded nicks between adjacent nucleotides in a duplex DNA chain**


backbone
insert
DNA fragments join
at sticky ends
Sticky end
G
C T T A A
Sticky end
A A T T
Recombinant DNA



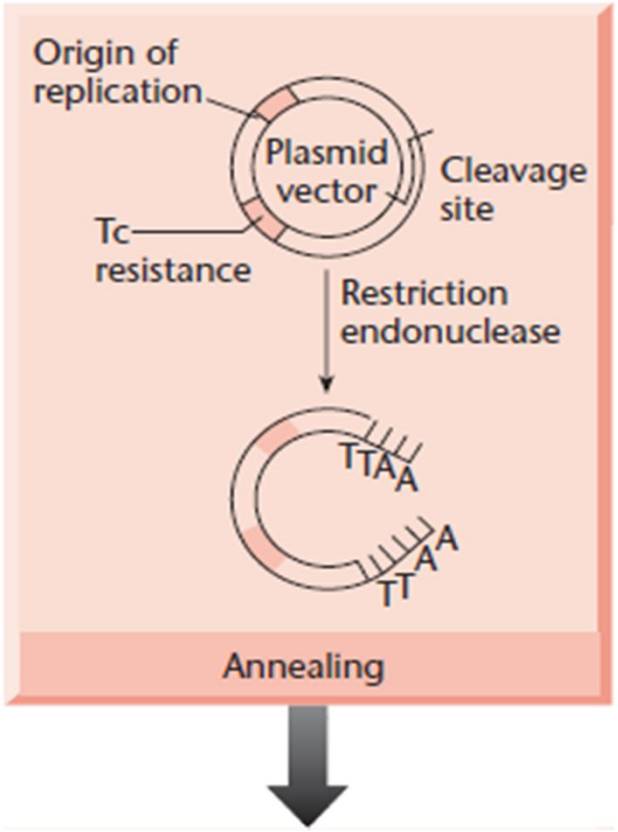
Action of DNA ligase

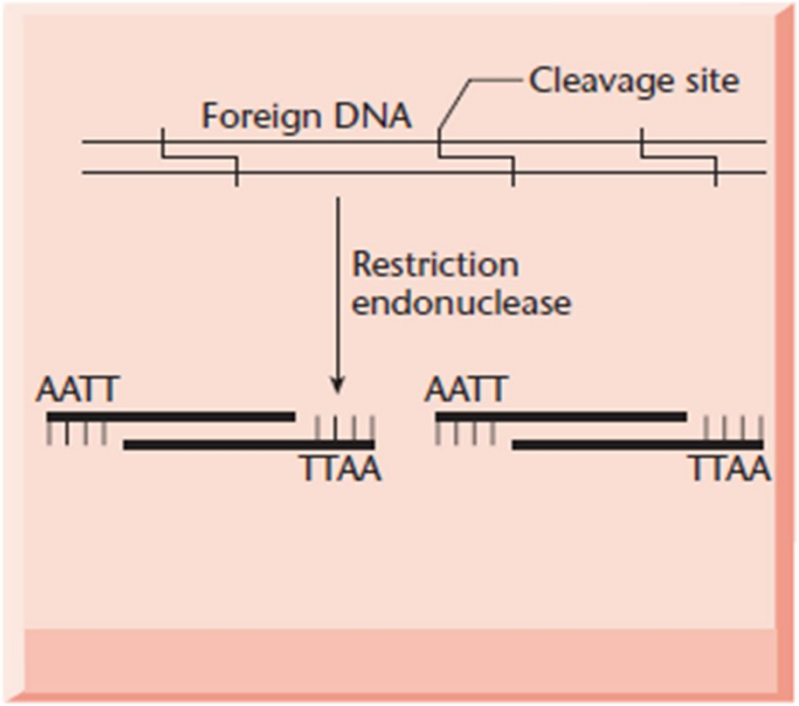
Joining DNA molecules

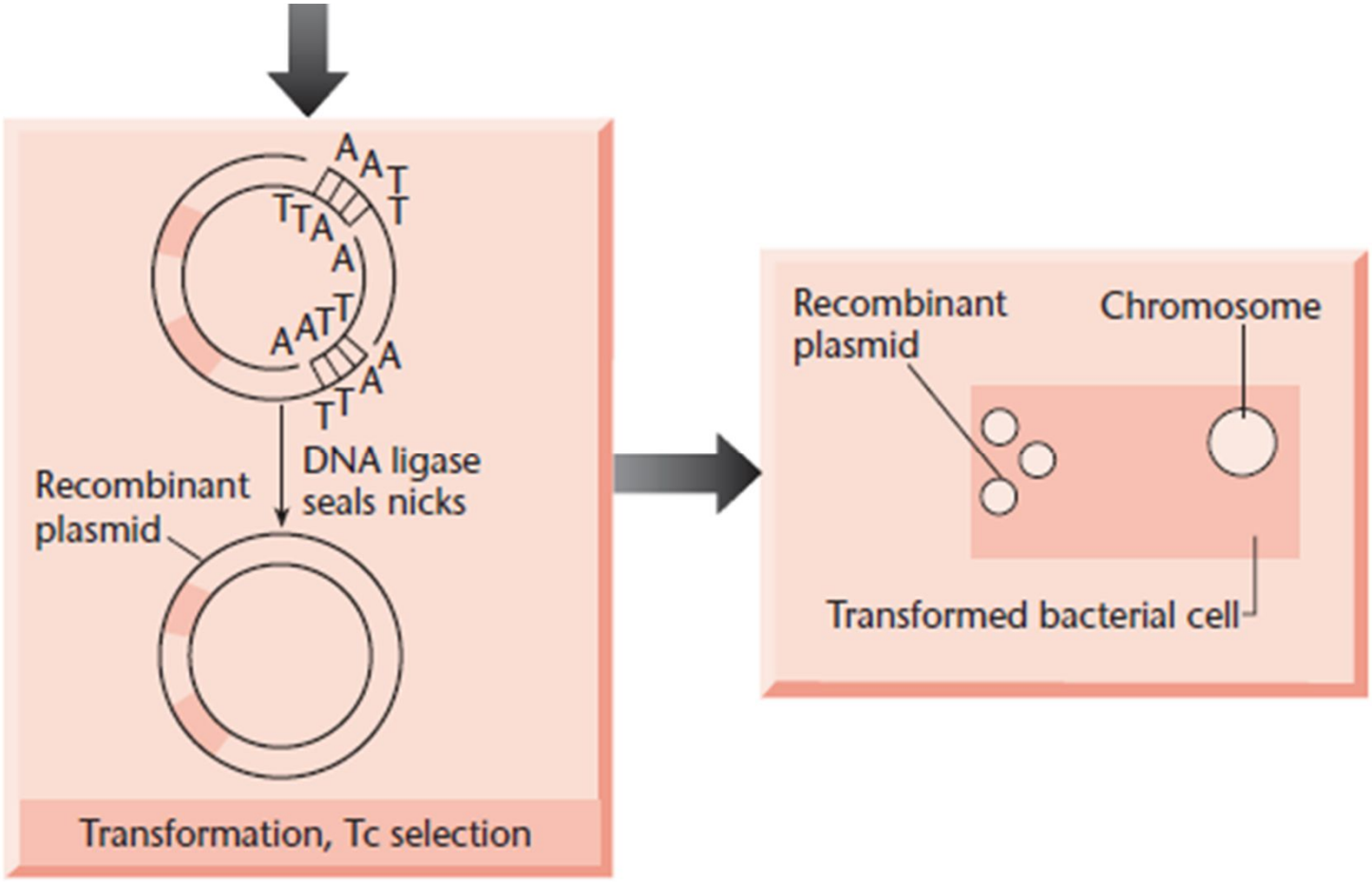
DNA ligase to create covalent recombinant DNA

- **DNA ligase can repair the nicks produced after the association of cohesive ends of DNA strands**

Use of DNA ligase to create a covalent DNA recombinant






A A T
T T A
A
A A T T
A A
T T
DNA ligase
seals nicks
Recombinant
plasmid
Transformation, Tc selection
Recombinant
plasmid
Chromosome
Transformed bacterial cell

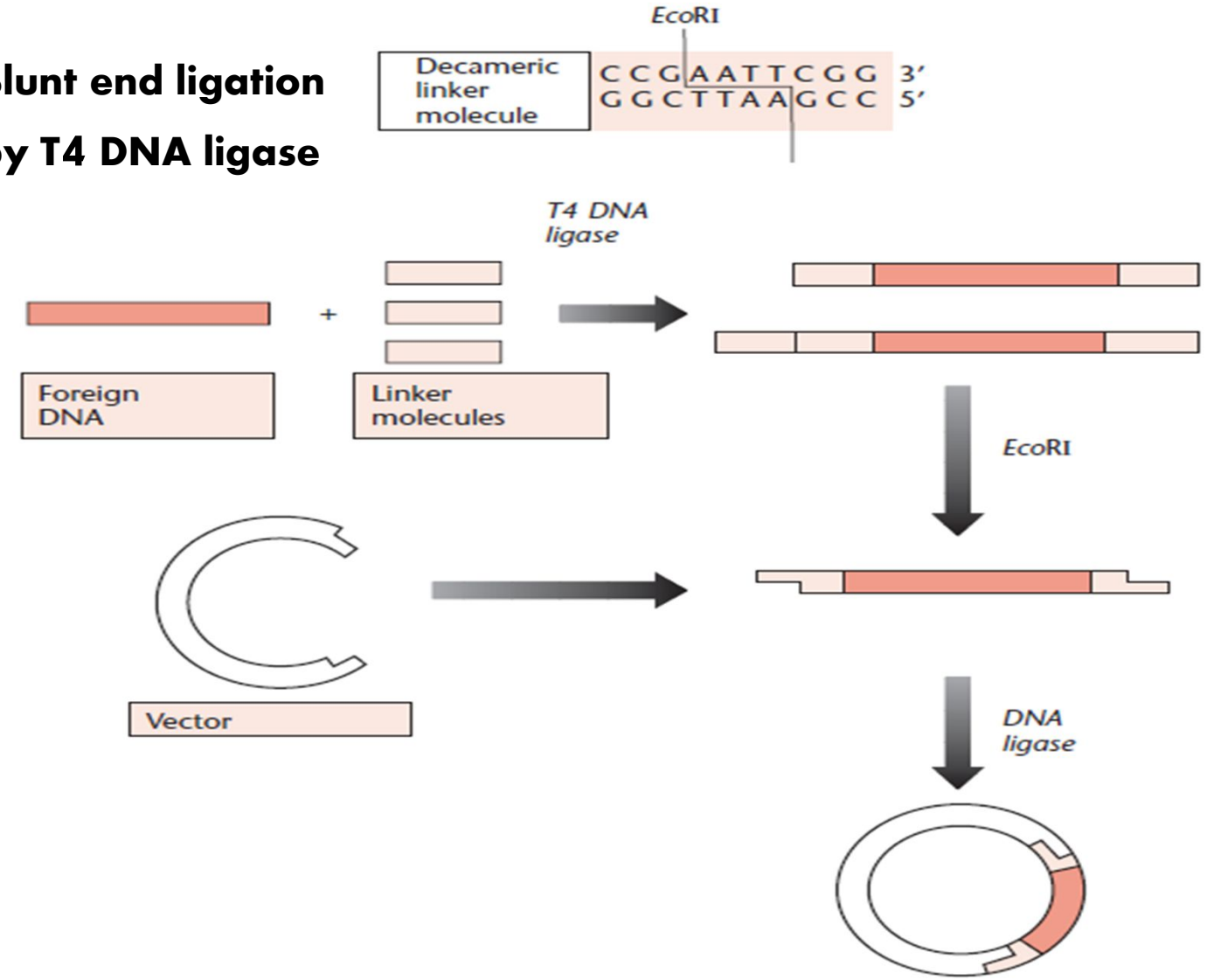
Joining DNA molecules

Blunt end ligation via linker molecules

- Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process
- T4 DNA ligase has been used to joint blunt-ended DNA molecules

Blunt end ligation

by T4 DNA ligase



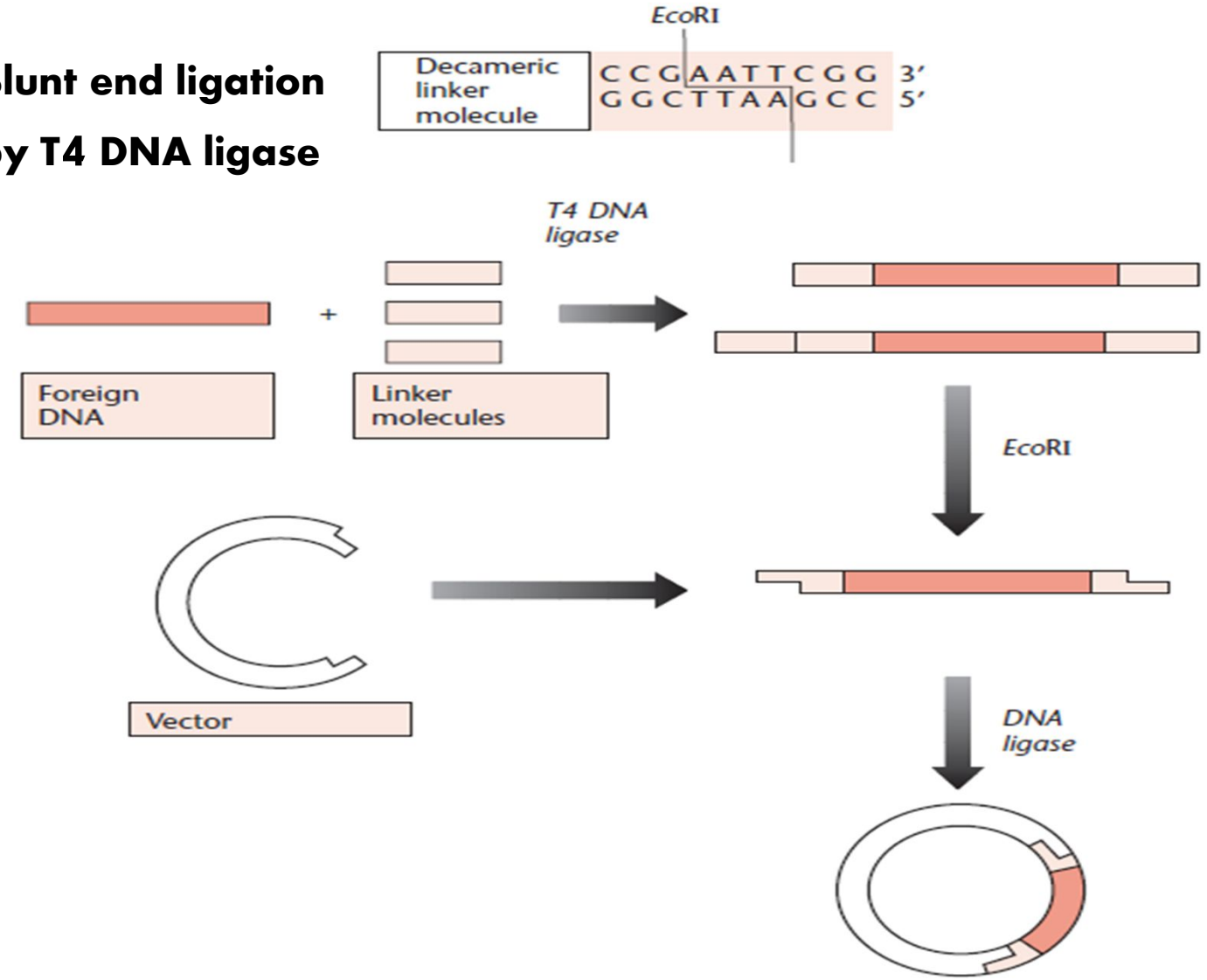
Joining DNA molecules

Blunt end ligation via linker molecules

- Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process
- T4 DNA ligase has been used to joint blunt-ended DNA molecules

Blunt end ligation

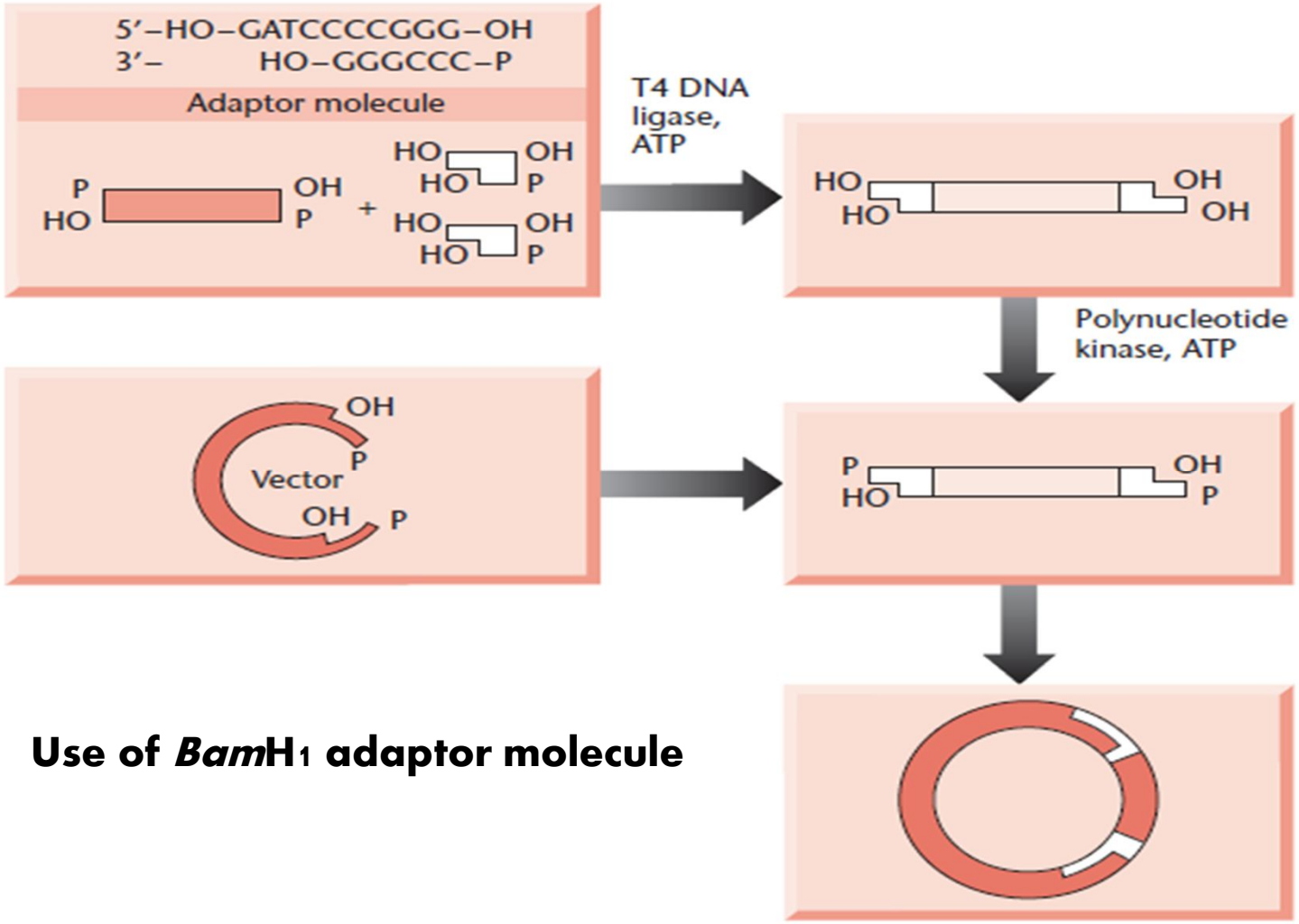
by T4 DNA ligase



Joining DNA molecules

Adaptors

- **Chemically synthesized adaptor molecules which have a preformed cohesive end can be used to insert blunt ended foreign DNA into vector**



Use of *Bam*H1 adaptor molecule

Joining DNA molecules

Adaptors

- **The main difference between linker and adaptor is that former having blunt ends while later have one cohesive end**

Joining DNA molecules

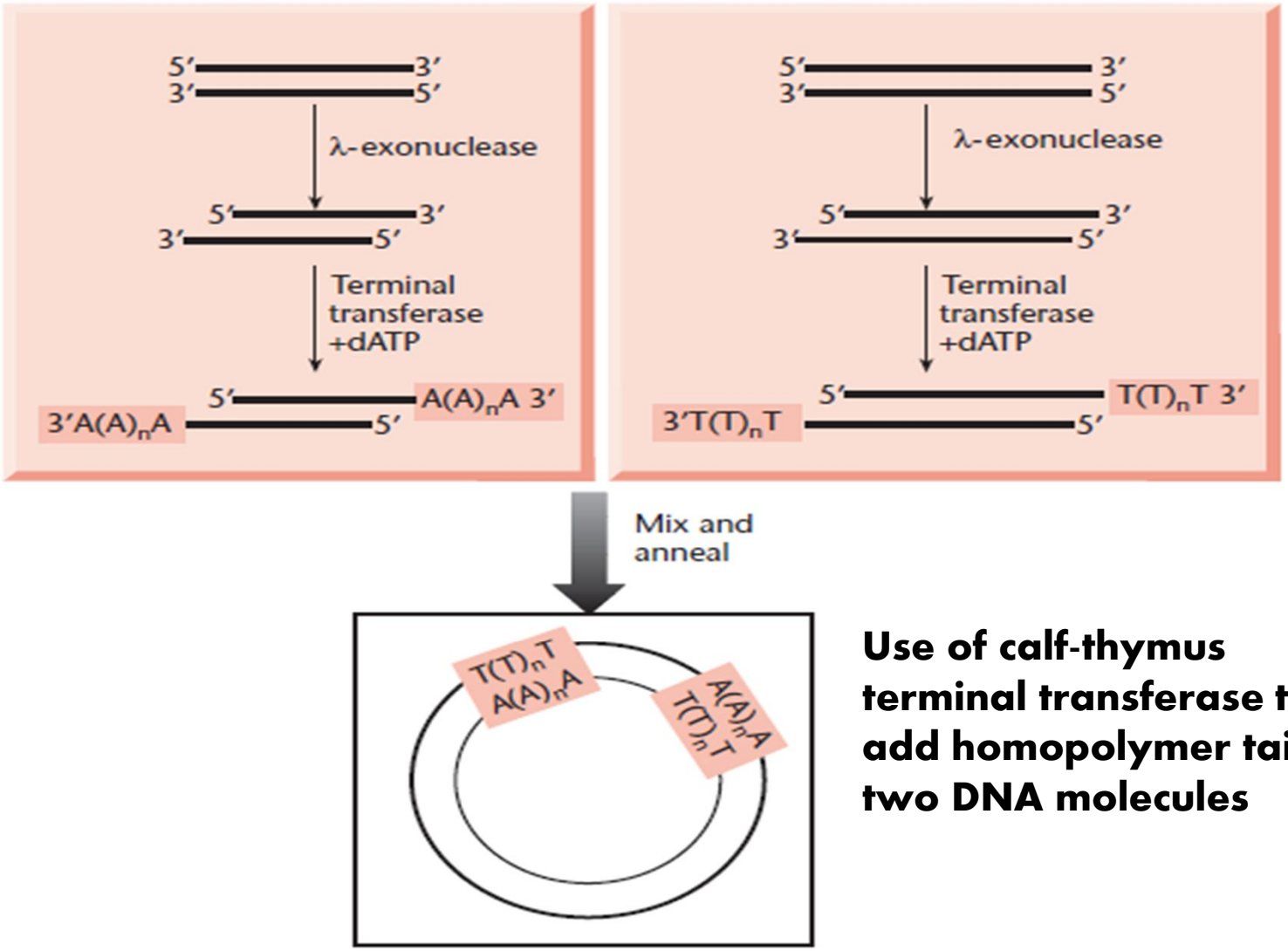
Homopolymer tailing

- A general methods for joining DNA molecules makes use of the annealing of complementary homopolymer sequences by adding oligo(dA) sequences to 3′ ends of one population of DNA molecules and oligo(dT) blocks to the 3′ ends of another population

Joining DNA molecules

Homopolymer tailing

- An enzyme purified from calf thymus, terminal deoxynucleotidyltransferase, provides the means by which the homopolymeric extensions at 3′ OH termini can be synthesized, if presented with a single type of nucleotide



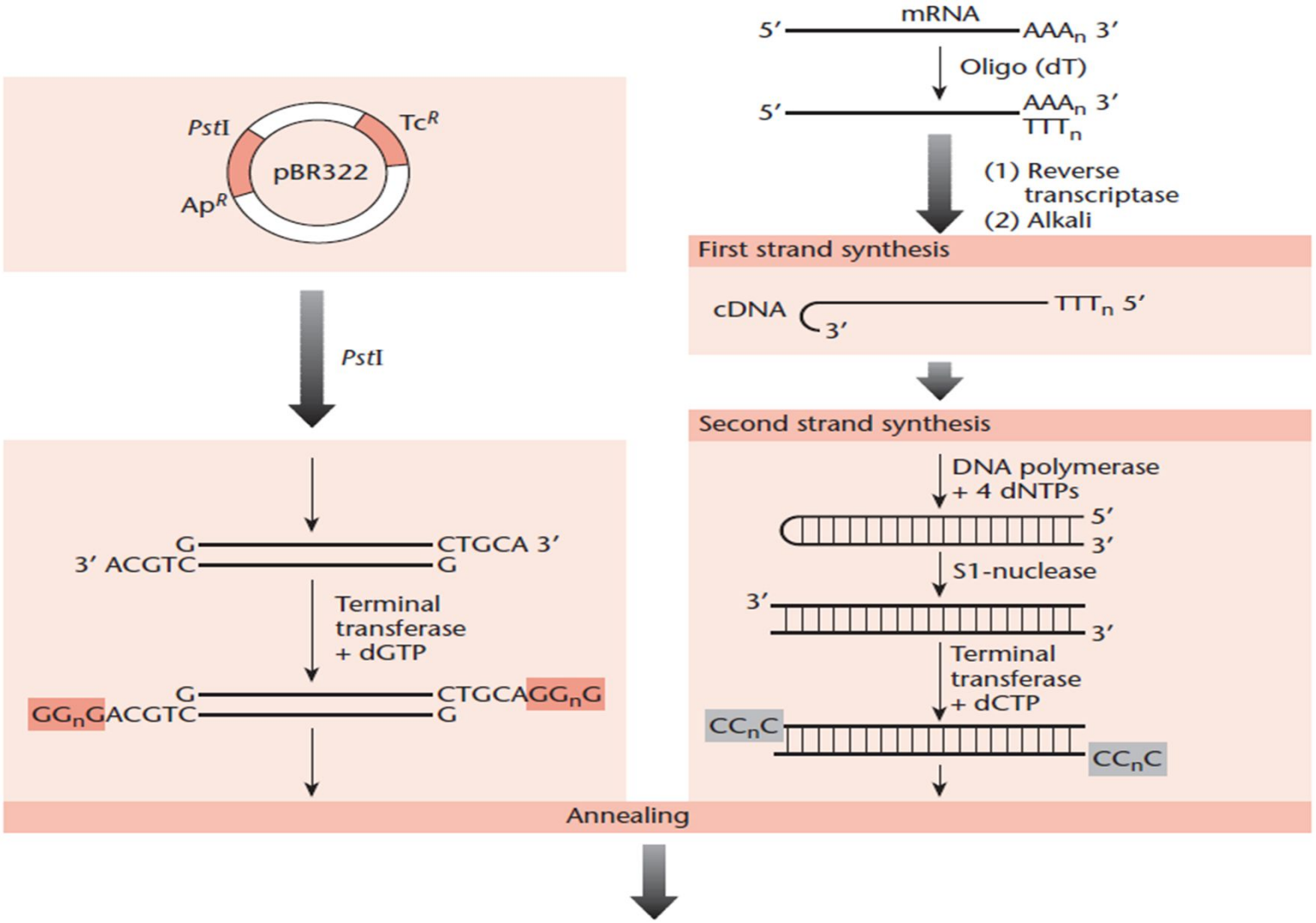
Use of calf-thymus terminal transferase to add homopolymer tails to two DNA molecules

Joining DNA molecules

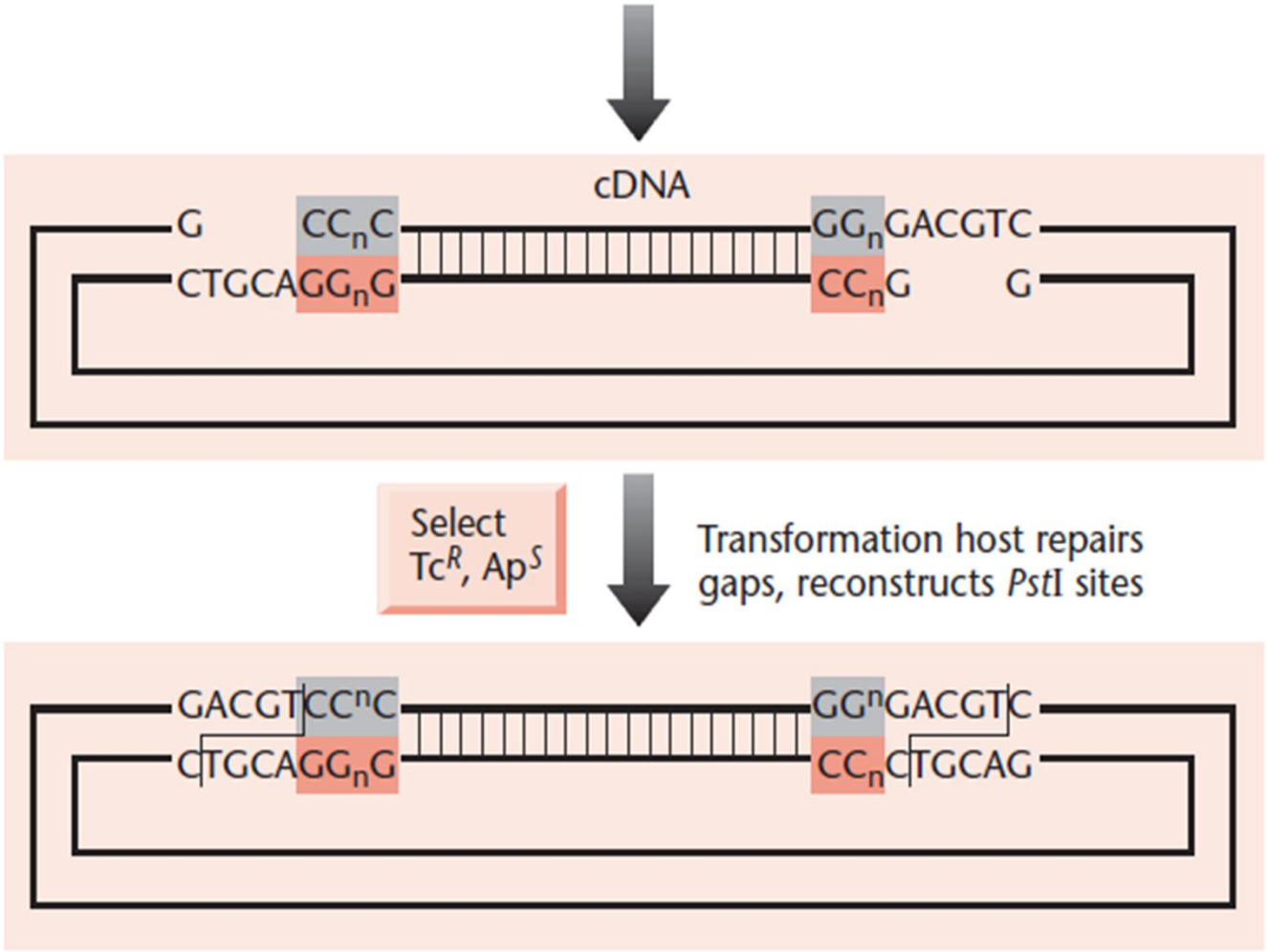
Cloning of cDNA by Homopolymer tailing

- **Eukaryotic mRNA can be cloned in vector after converting it to cDNA by using an enzyme reverse transcriptase**

Insertion of cDNA into vector by homopolymer tailing



Insertion of cDNA into vector by homopolymer tailing



Insertion of cDNA into vector by homopolymer tailing

Joining DNA molecules

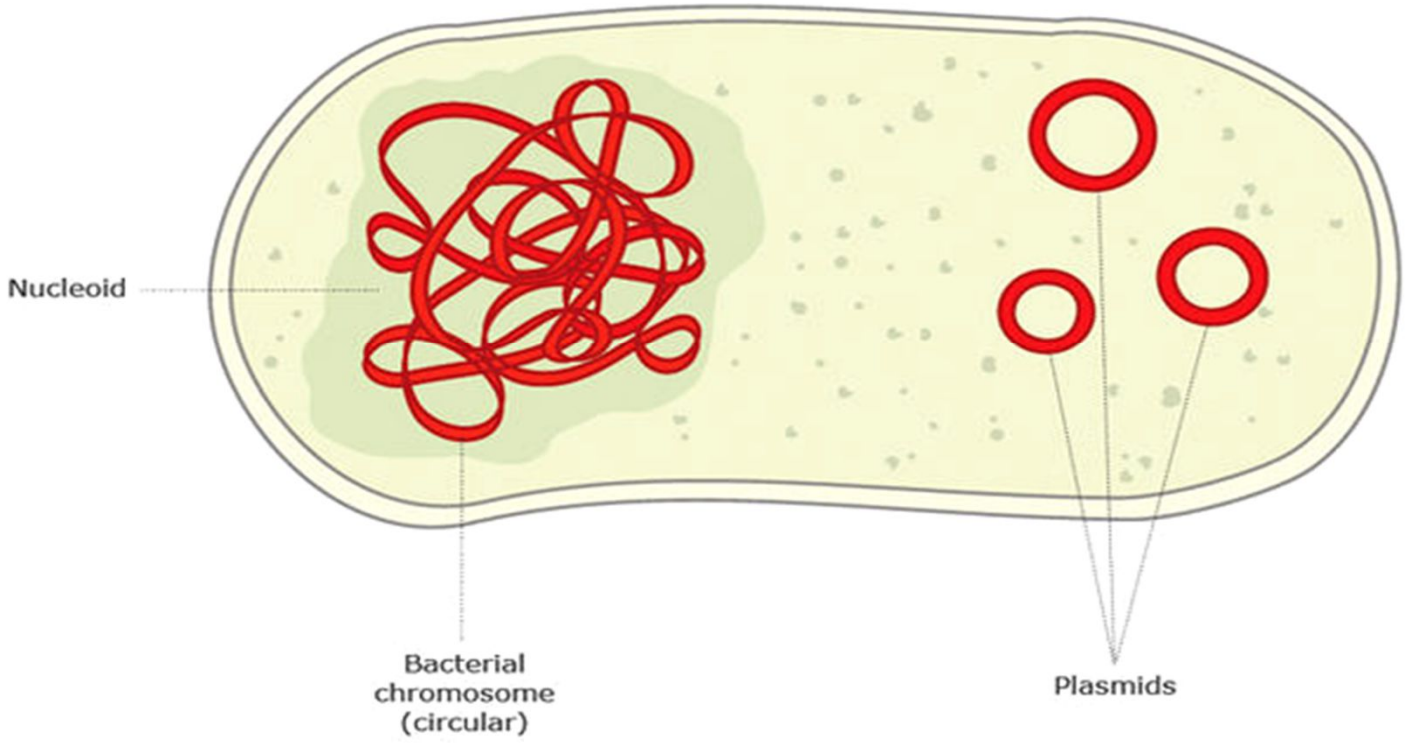
- **cutting and joining DNA molecules is basic step in gene manipulation**

Basic Biology of plasmid and phage vectors

Plasmids

- Plasmids are mostly circular DNA molecules that replicate separately from the host chromosome
- Naturally occurring bacterial plasmids range in size from 5,000 to 400,000 bp

Plasmid Biology



Plasmid Biology

Plasmids

- **Plasmid help bacteria to survive stress**

Plasmid Biology

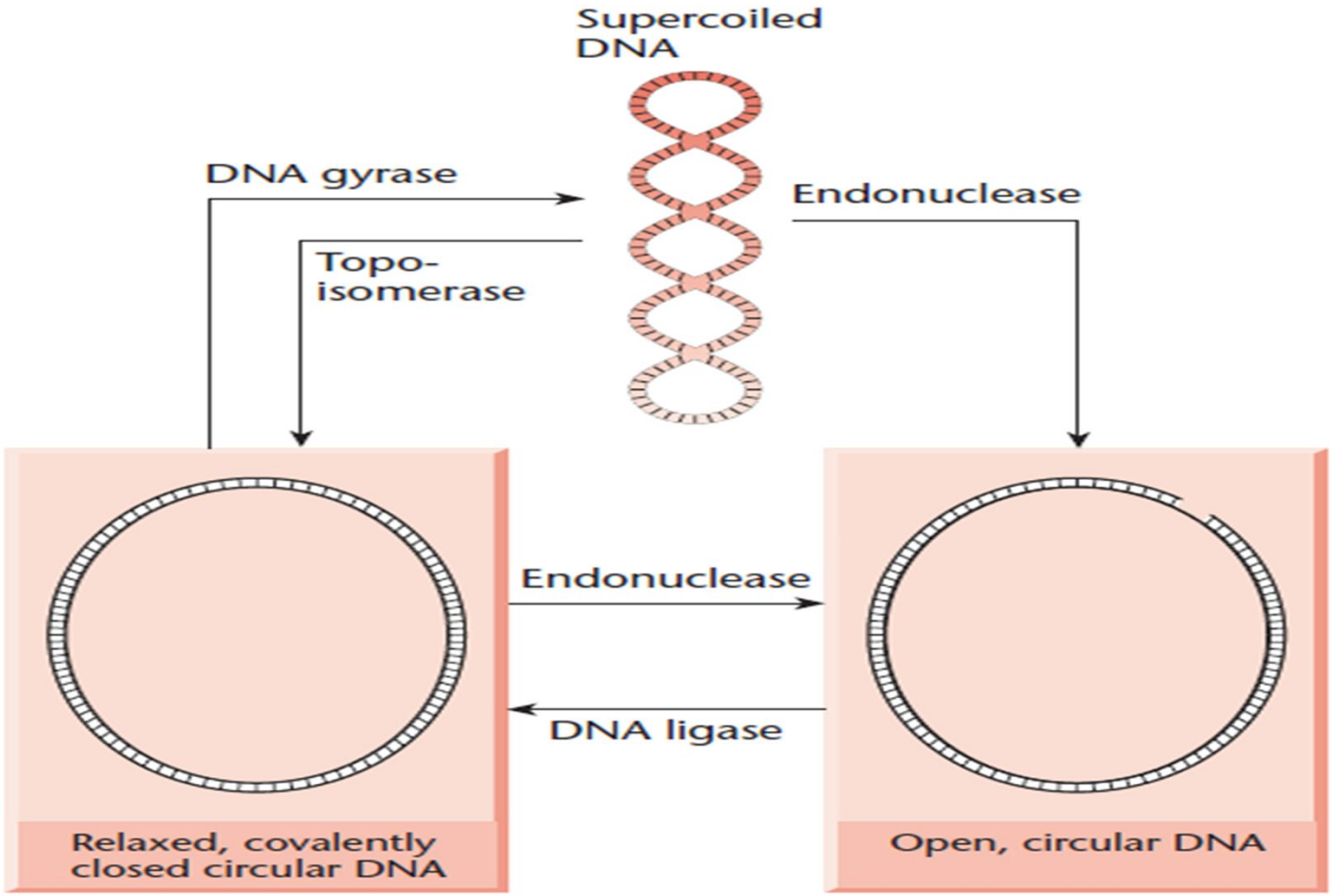
Interconversion of plasmid DNA

- Most plasmids exist as double-stranded circular DNA molecules
- If both strands of DNA are intact circles the molecules are describes as covalently closed circles or CCC DNA.

Plasmid Biology

Interconversion of plasmid DNA

- **Open circle or OC DNA have one strand intact**



Interconversion of supercoiled, relaxed CCC DNA and OC DNA

Plasmid Biology

Effect of ethidium bromide on supercoiling of DNA

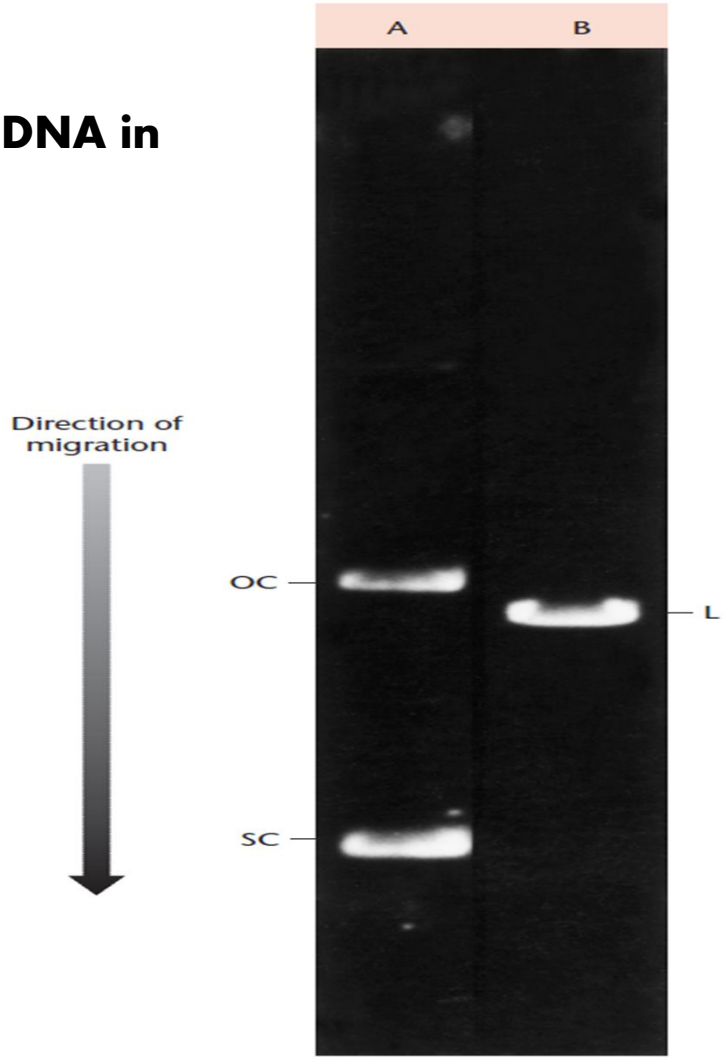
- DNA bands can be visualized by soaking the gel in a solution of ethidium bromide

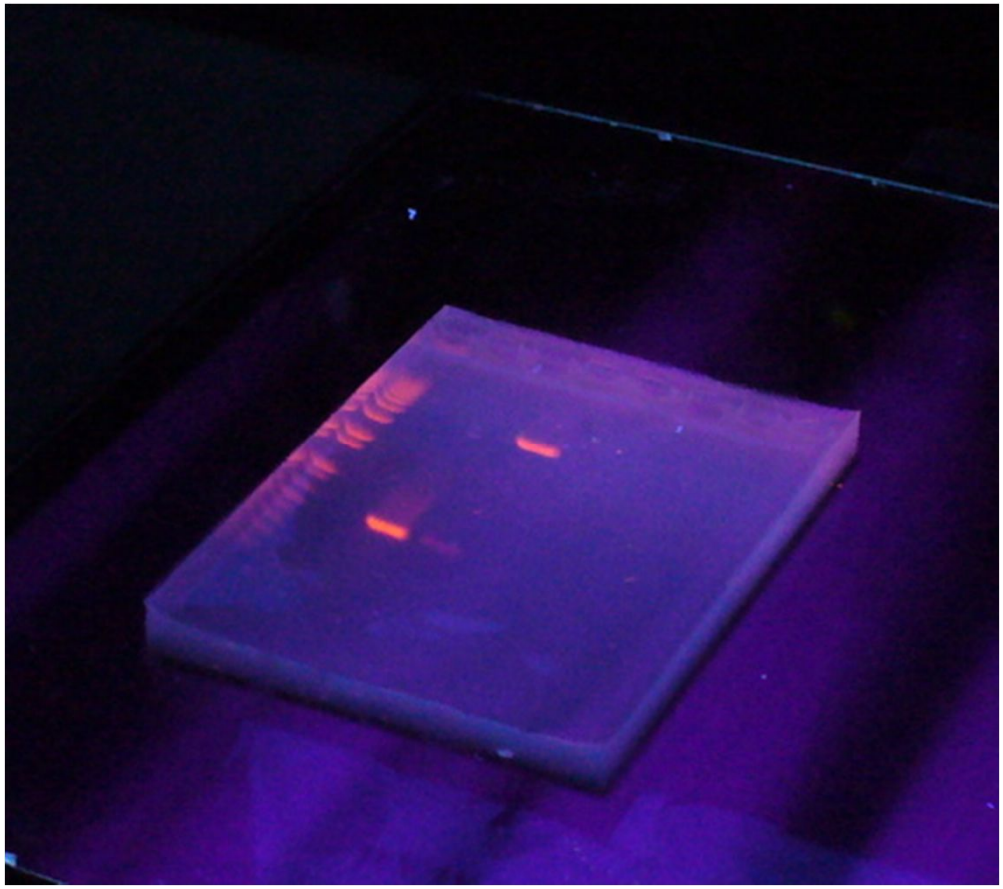
Plasmid Biology

Effect of ethidium bromide on supercoiling of DNA

- As the amount of intercalating agent i.e. ethidium bromide increases, the double helix untwists until the open form of circular DNA is produced

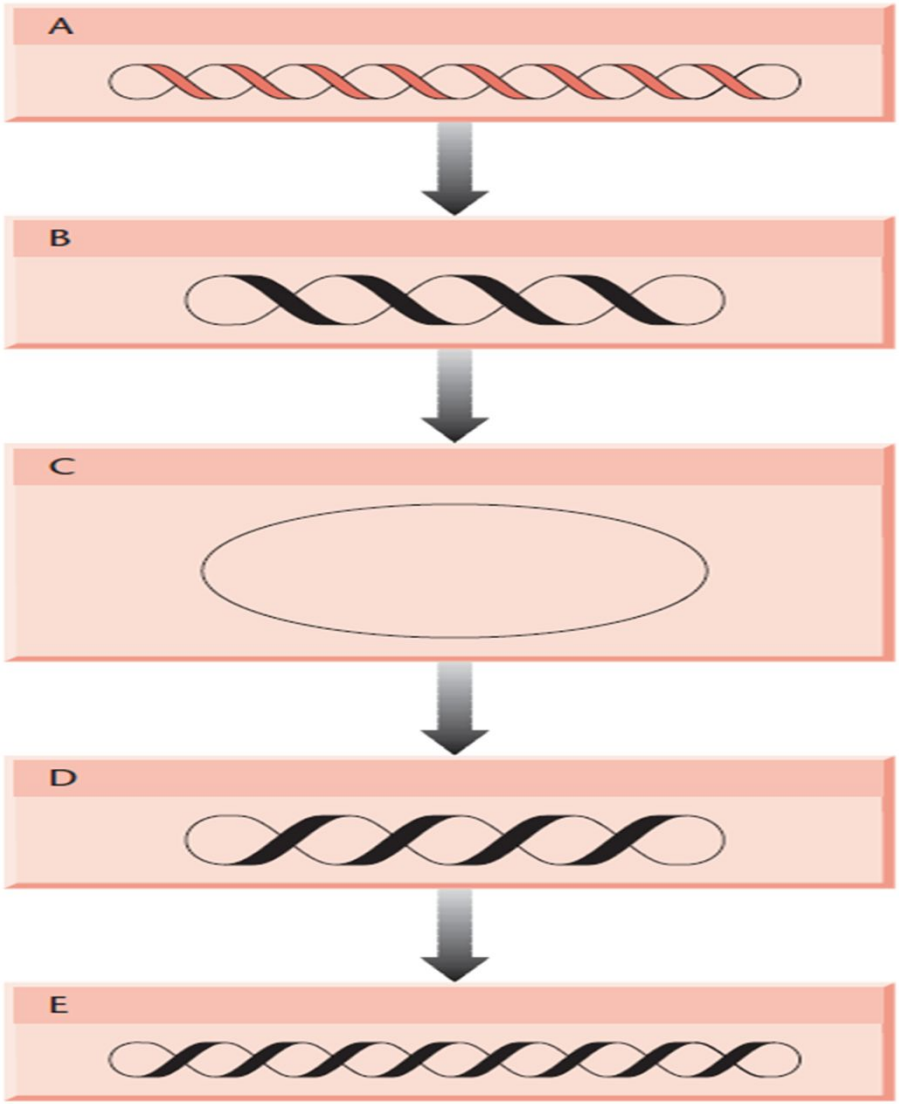
Electrophoresis of DNA in agarose gel





Visualization of gel under UV light after staining with ethidium bromide

Effect of ethidium bromide on supercoiling of DNA



Plasmid Biology

Effect of ethidium bromide on supercoiling of DNA

- **Ethidium bromide may be a mutagen, a carcinogen, or a teratogen**

Plasmid Biology

Phenotypic traits exhibited by plasmids

- Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 to greater than 200×10^6 daltons

Plasmid Biology

Some phenotypic traits exhibited by plasmid-carried genes

Antibiotic resistance
Antibiotic production
Degradation of aromatic compounds
Haemolysin production
Sugar fermentation
Enterotoxin production
Heavy-metal resistance
Bacteriocin production
Induction of plant tumours
Hydrogen sulphide production
Host-controlled restriction and modification

Plasmid Biology

Phenotypic traits exhibited by plasmids

- Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 to greater than 200×10^6 daltons

Plasmid Biology

Some phenotypic traits exhibited by plasmid-carried genes

Antibiotic resistance
Antibiotic production
Degradation of aromatic compounds
Haemolysin production
Sugar fermentation
Enterotoxin production
Heavy-metal resistance
Bacteriocin production
Induction of plant tumours
Hydrogen sulphide production
Host-controlled restriction and modification

Plasmid Biology

Properties of Conjugative and non-conjugative plasmids

- **Plasmids can be categorized into conjugative or non-conjugative depending whether or not they carry a set of transfer genes, called *tra* genes, which promoted bacterial conjugation**

Plasmid Biology

Properties of some conjugative and non-conjugative plasmids of Gram-negative organisms

Plasmid	Size (Mda)	Conjugative	No. of plasmid copies	Phenotype
ColE1	4.2	No	10-15	ColE1 production
RSF1030	5.6	No	20-40	Ampicillin resistance
cloDF13	6	No	10	Cloacin production
R6k	25	Yes	13-38	Ampicillin and streptomycin resistance
F	62	Yes	1-2	-
RI	RI	Yes	3-6	Multiple drug resistance
Ent P 307	65	Yes	1-3	Enterotoxin production

Plasmid Biology

Host range of plasmids

- Plasmids encode only a few of the proteins required for their own replication
- Host range of plasmid is determined by *ori* region
- Plasmids whose *ori* derived from Col E1 have restricted host range

Plasmid Biology

Host range of plasmids

- **RP4 and RSF1010 and many plasmids from *Staphylococcus aureus* have broad host range**

Plasmid Biology

Partitioning and segregative stability of plasmids

- The loss of plasmids due to defective partioning is called segregative instability
- Naturally occurring plasmids are stably maintained because they contain *par* regions

Plasmid Biology

Incompatibility of plasmids

- **The inability of two different plasmids to coexist in the same cell in the absence of selection pressure**

Maximizing the expression of cloned gene

Structural instability

- Structural instability of plasmids may arise by deletions or rearrangements of DNA

Plasmid Biology

Desirable properties of plasmid cloning vehicles

- **An ideal cloning vehicle would have the following three properties**

i). Low molecular weight

ii). Ability to confer readily selectable phenotypic traits on host cells

Plasmid Biology

Desirable properties of plasmid cloning vehicles

iii). **Single sites for number of restriction endonucleases**

Plasmid Biology

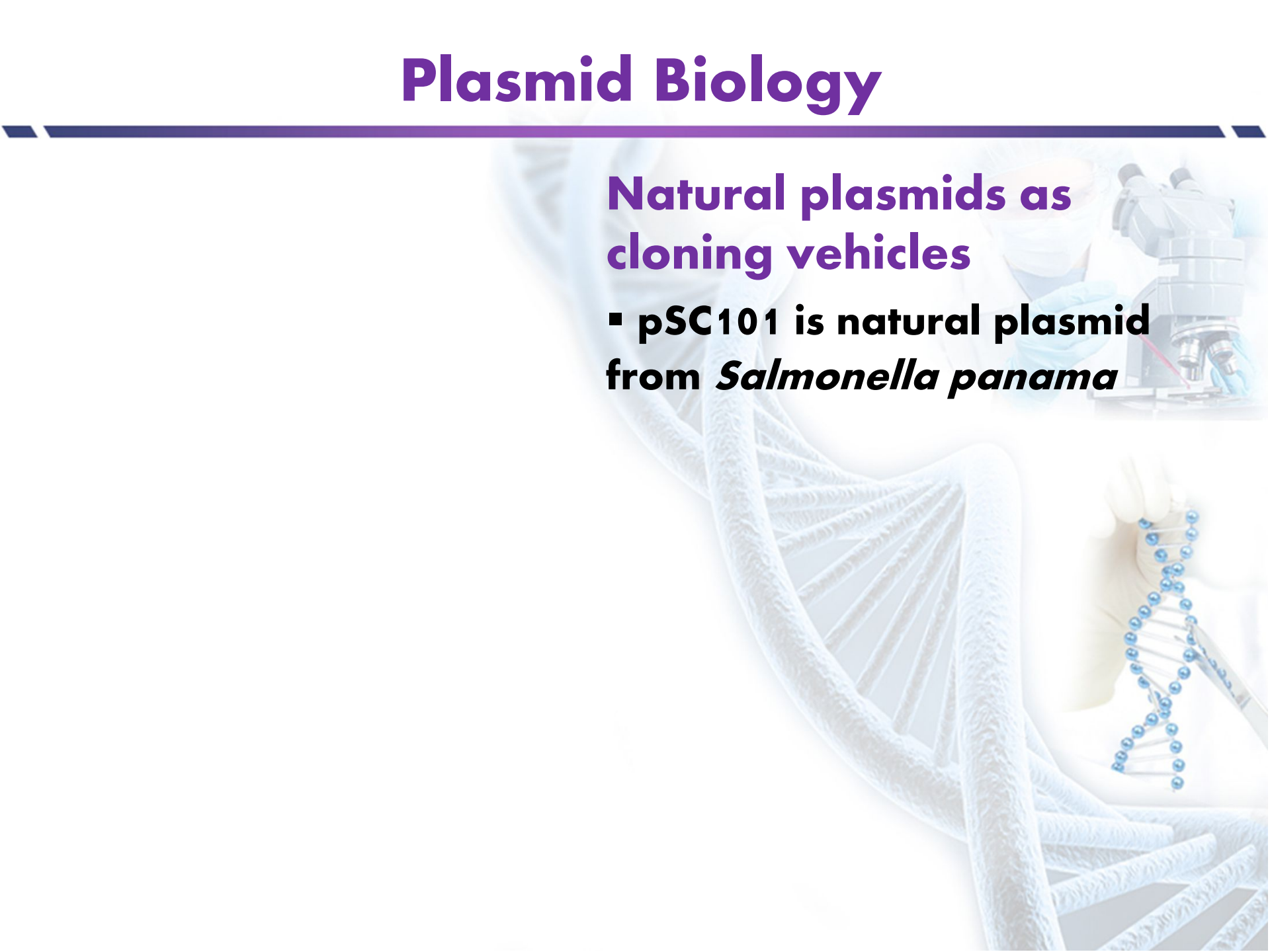
Natural plasmids as cloning vehicles

- Plasmids which were not constructed *in vitro* for the sole purpose of cloning are called natural plasmids
- Col E1 is a natural occurring plasmid
- RSF2124 is a derivative of Col E1

Plasmid Biology

Natural plasmids as cloning vehicles

- **pSC101 is natural plasmid from *Salmonella panama***



Properties of some 'natural' plasmids used for cloning DNA

Plasmid	Size (Mda)	Sites of endonucleases	Selectable markers	Insertional inactivation
pSC101	5.8	*Xho*I, *Eco*RI, *Pvu*II, *Hinc*II, *Hpa*I	Tetracycline resistance	---
		*Hind*III, *Bam*HI, *Sal*I	__	Tetracycline resistance
Col E1	4.2	*Eco*RI	Immunity to colicin E1	Colicin E1 production
RSF2124	7.4	*Eco*RI, *Bam*HI	Ampicillin resistance	Colicin E1 production

Plasmid Biology

Natural plasmids as cloning vehicles

- **Cloning in pSC101, ColiE1 and RSF2124**

Plasmid Biology

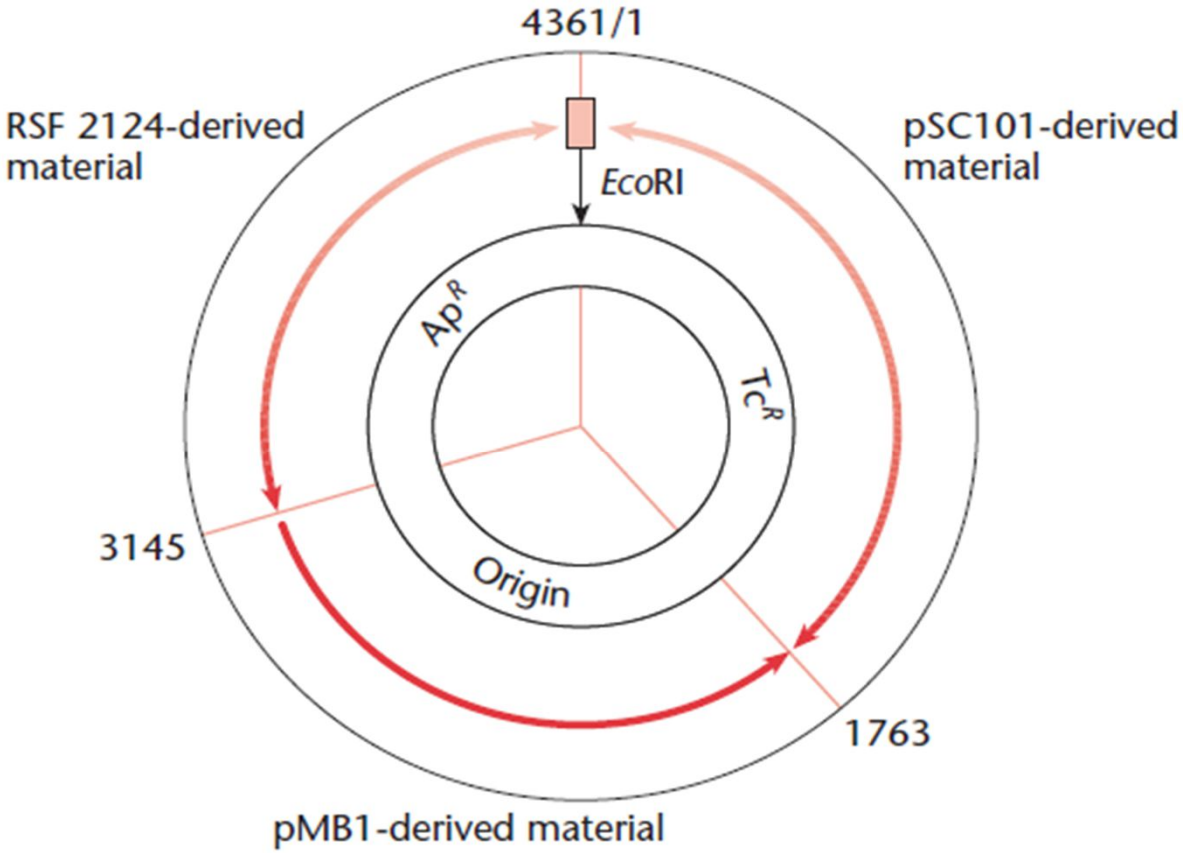
Use of pSC101 for cloning

- Expression of *Staphylococcus* plasmid genes in *E. coli*
- pSC101 contain Tetracycline resistance *P1248 from Staphylococcus*

Plasmid Biology

pBR322, a purpose-built cloning vehicle

- pBR322 is an example of *in vitro* constructed cloning vehicle
- pBR322 contains the Ap^R and Tc^R genes of RSF2124 and pSC101, combined with replication element of pMB1

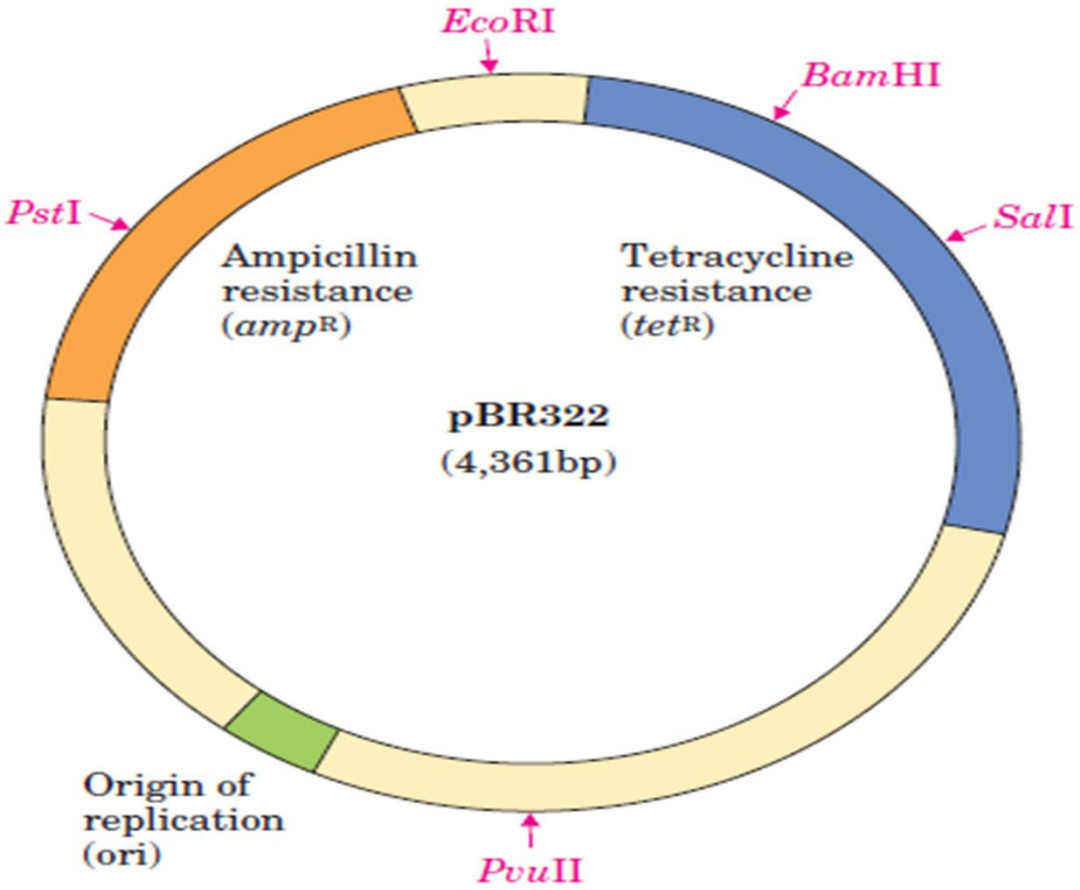


The origin of plasmid pBR322

Plasmid Biology

pBR322, a purpose-built cloning vehicle

- It has 40 target sites for different restriction enzymes



The constructed E. coli plasmid pBR322

Plasmid Biology

pBR322, a purpose-built cloning vehicle

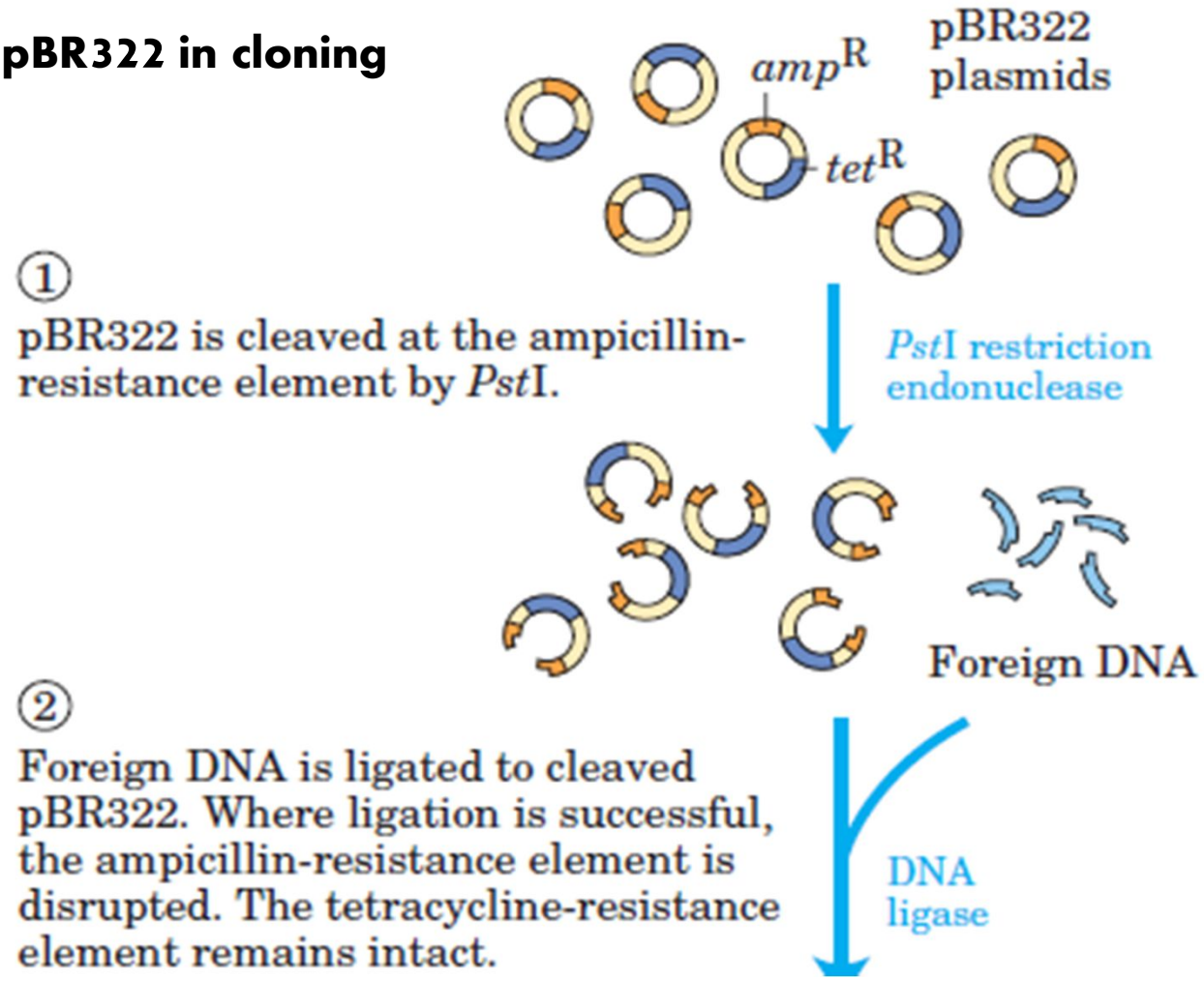
- It is difficult to clone DNA segments longer than about 15,000 bp when plasmids are used as the vector

Plasmid Biology

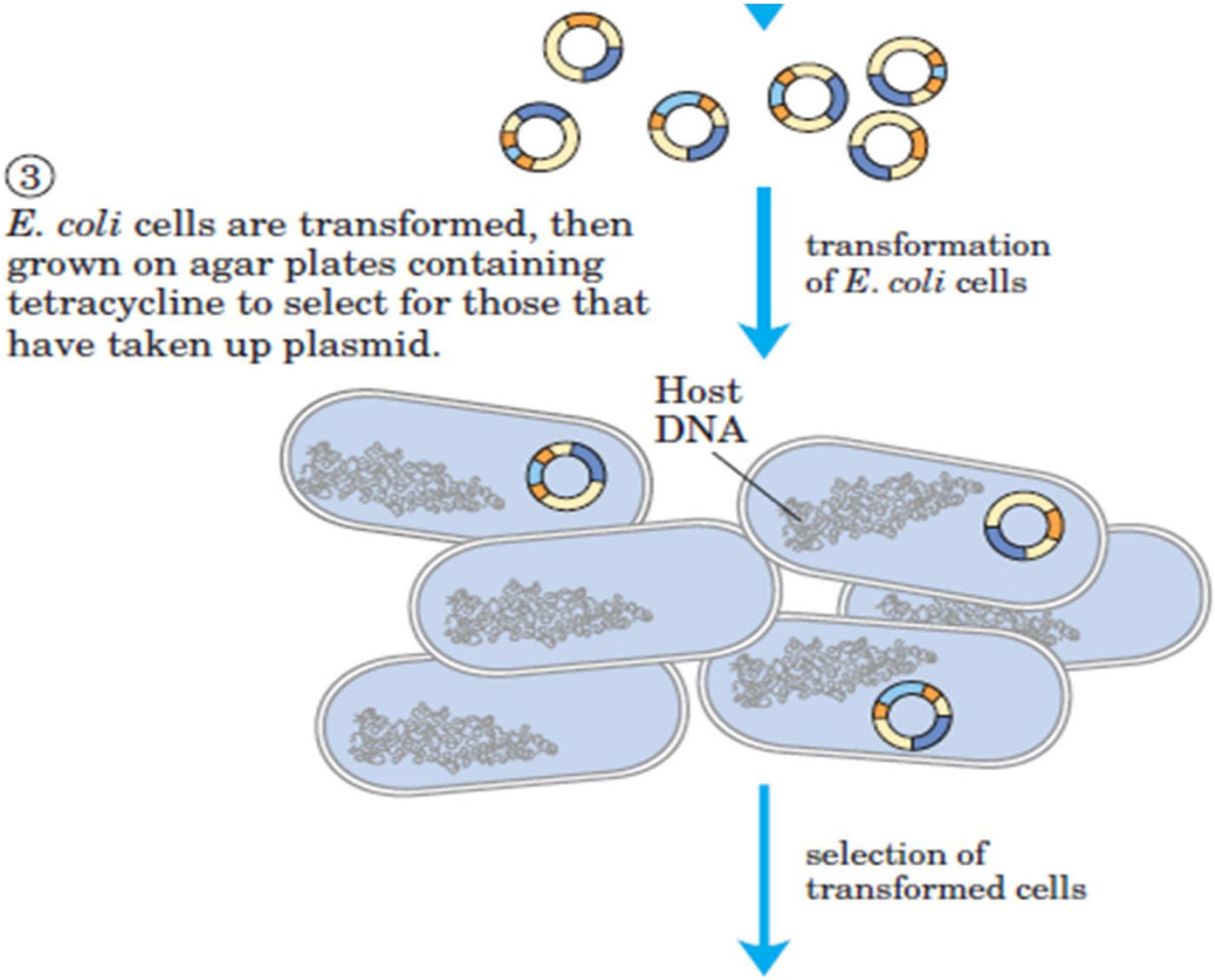
Cloning with pBR322

- **Plasmid pBR322 has a number of characteristics that make it suitable as cloning vehicle**

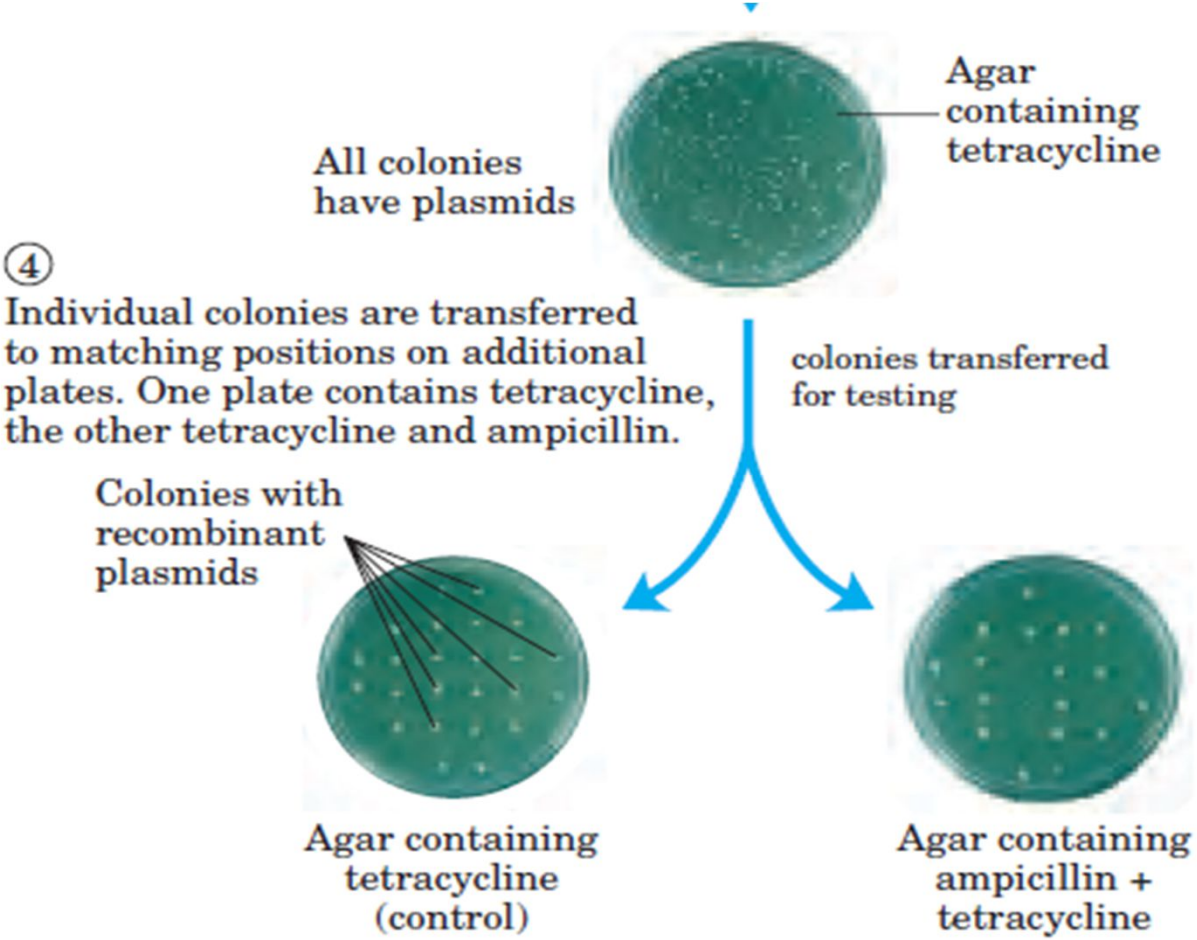
Use of pBR322 in cloning



Use of pBR322 in cloning



Use of pBR322 in cloning



Plasmid Biology

Improved vectors derived from pBR322

- Over the years numerous different derivatives of pBR322 have been constructed
- pBR325 encodes chloramphenicol resistance in addition to ampicillin and tetracycline

Plasmid Biology

Improved vectors derived from pBR322

- pUC19 carrying sites for many different restriction endonucleases
- pAT153 another derivative of pBR322

Plasmid Biology

Runaway plasmid vectors

- Loss of copy number control –runaway replication

Plasmid Biology

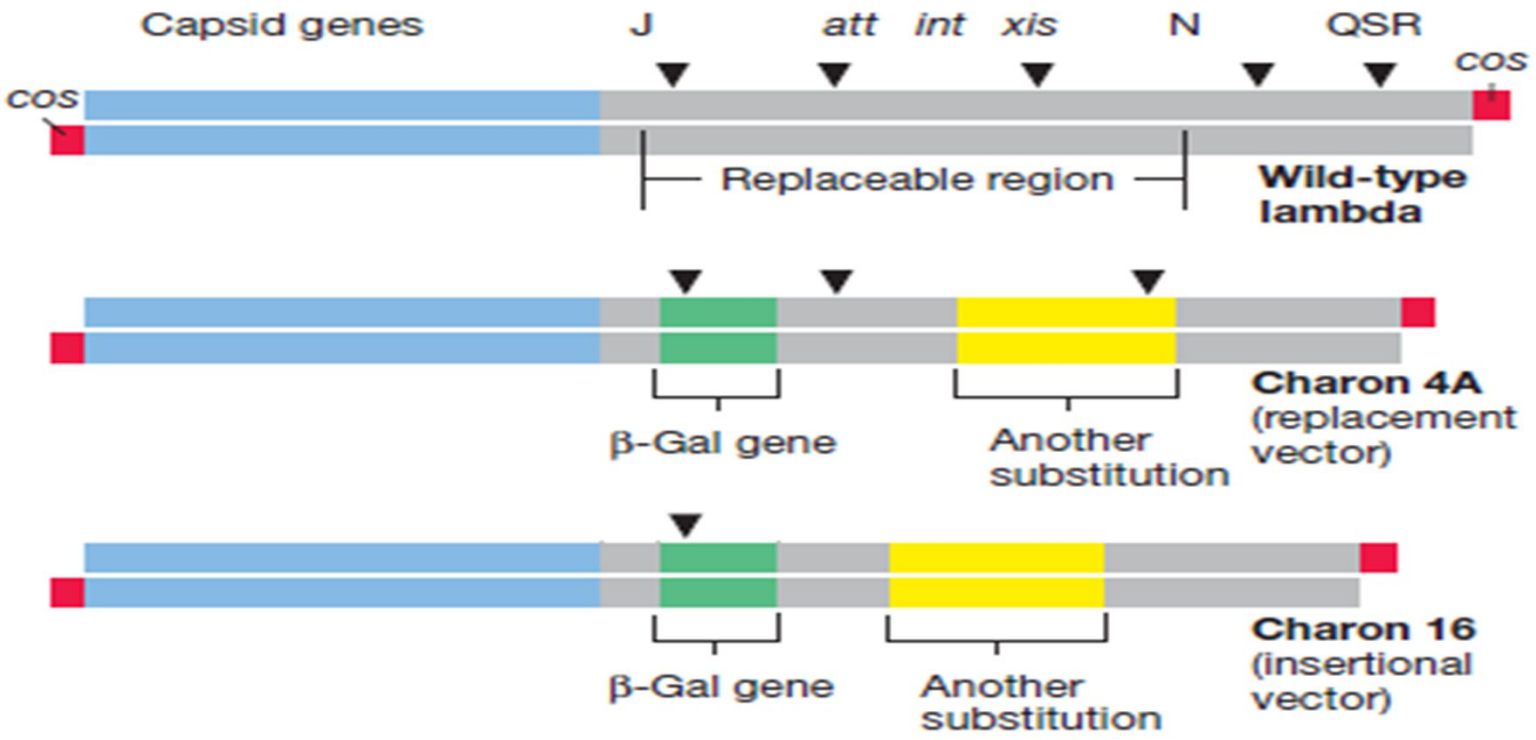
Runaway plasmid vectors

- Loss of copy number control –runaway replication

Phage vectors

Modified λ phages

- Wild type is not suitable as a cloning vector because it has too many restriction enzymes sites
- Insertional vectors
- Replacement vectors



Modified λ cloning vectors

Phage vectors

Modified λ phages

- Charon 4A and 16 are both derivatives of lambda with various substitutions and deletions in the nonessential region
- Each has the *lacZ* gene, encoding enzyme β-galactosidase, which permits detection of phage containing inserted clones

Phage vectors

Steps in cloning with λ

- Cloning with lambda replacement vectors involves the following steps

(i). Isolation and cutting vector DNA with appropriate restriction enzymes

(ii). Connecting the two lambda fragments to foreign DNA by using DNA ligase

Phage vectors

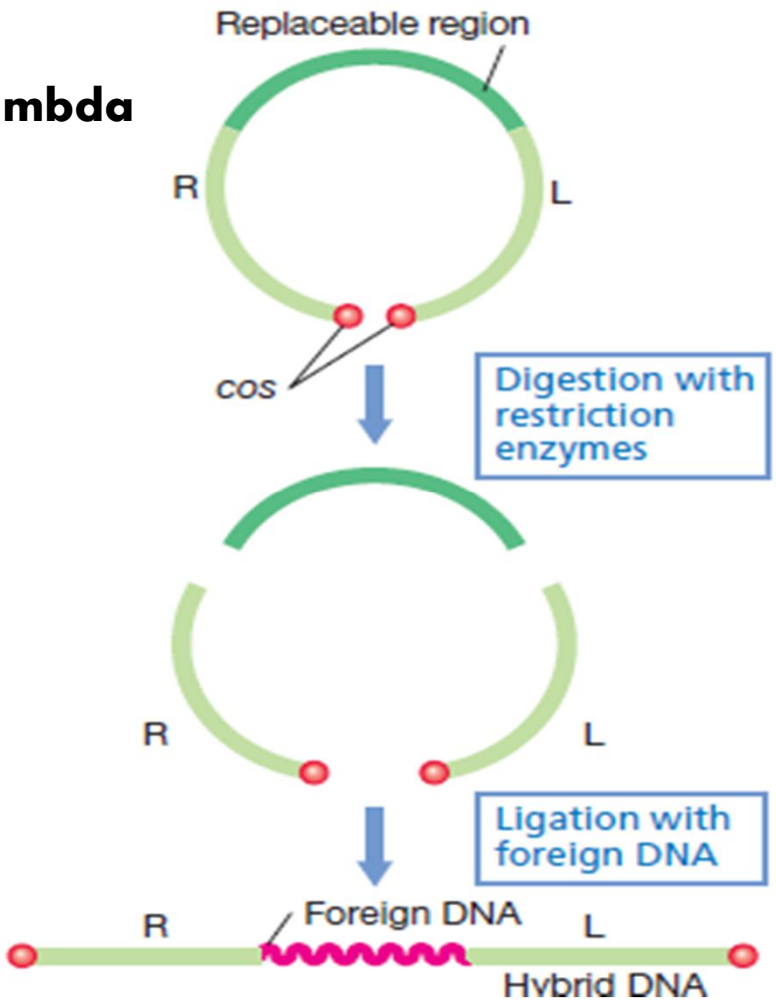
Steps in cloning with λ

(iii). ***In vitro*** **packaging of recombinant DNA**

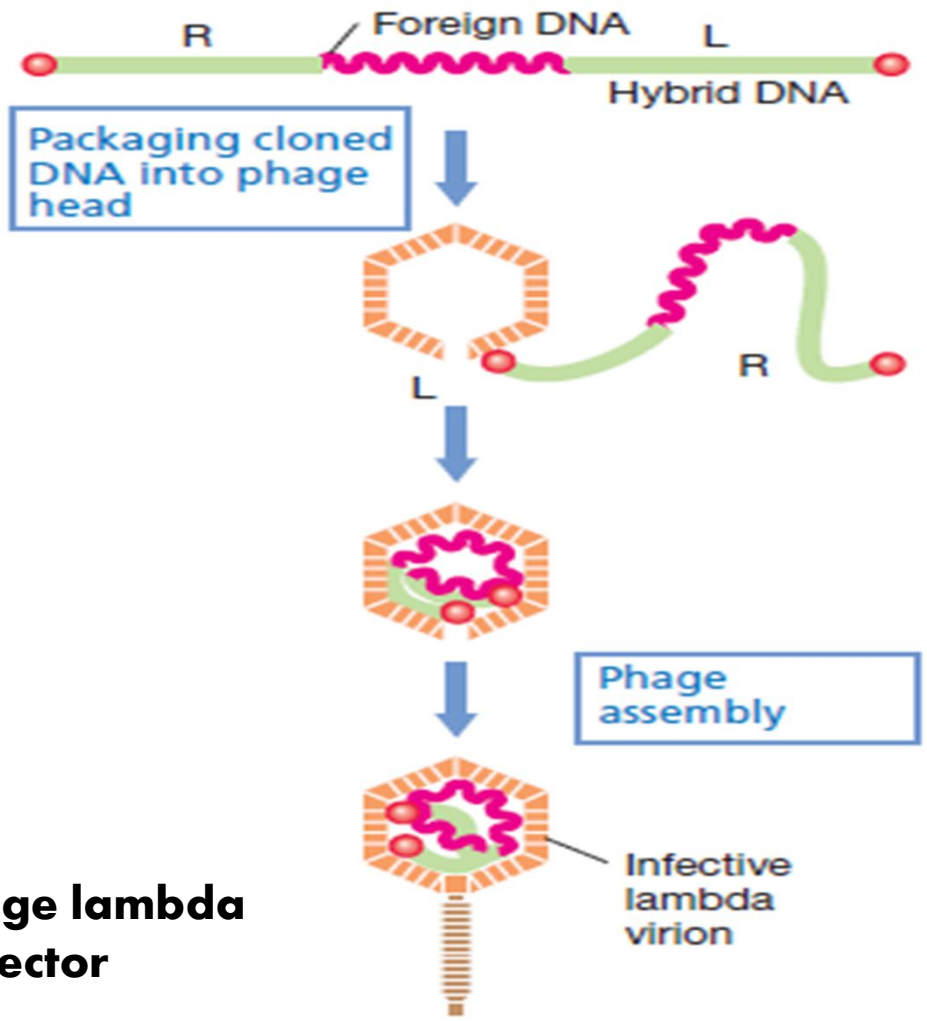
(iv). **Infection of *E. coli* cells and isolation of phage clones by picking plaques on a host strain**

(v). **Checking recombinant phage for the presence of foreign DNA**

Bacteriophage lambda as cloning vector



Bacteriophage lambda as cloning vector



Phage vectors

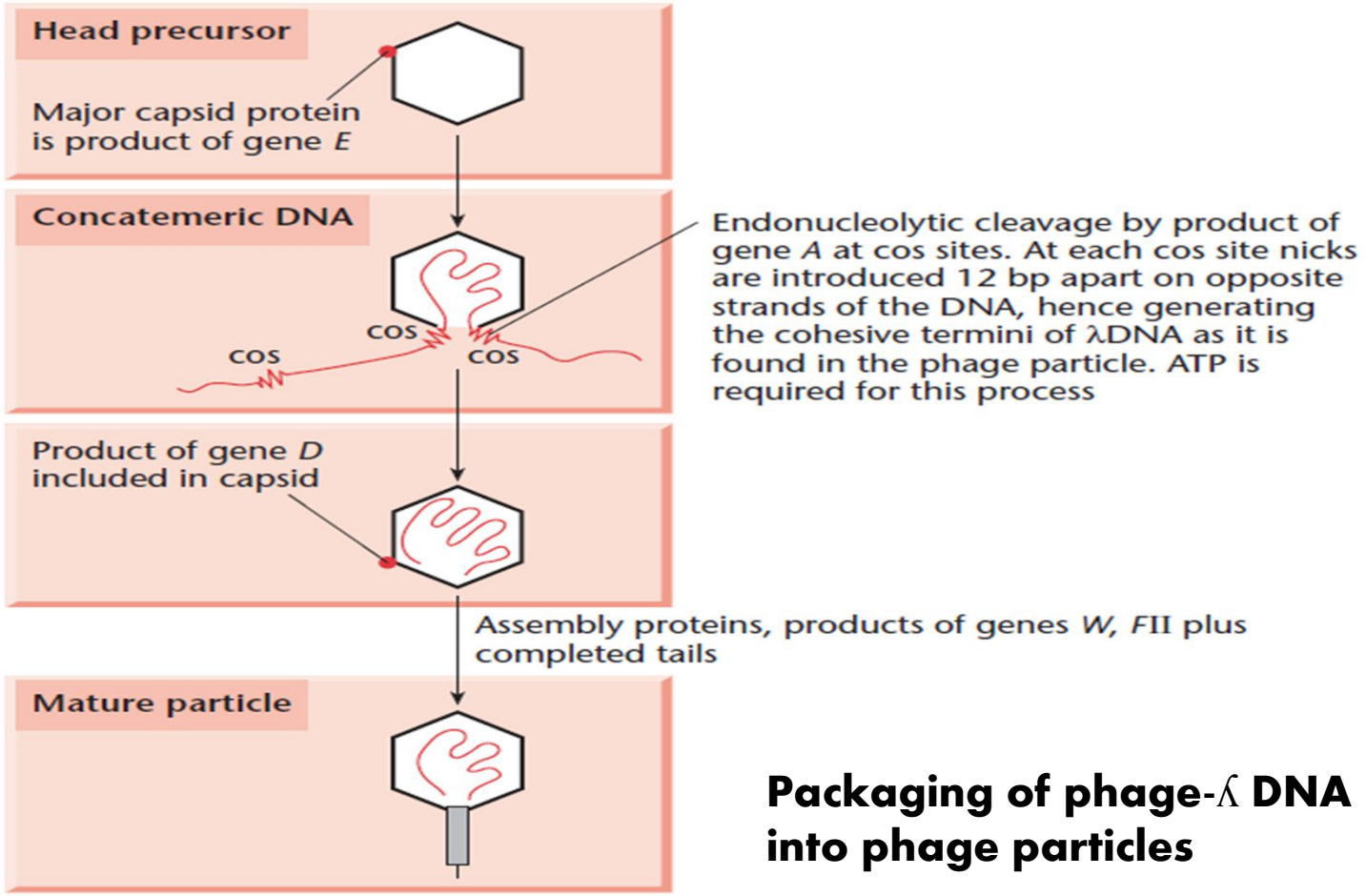
Steps in cloning with λ

- **Both Charon vectors are also engineered to contain reporter genes such as for β-galactosidase for detection of recombinant phages**

Phage vectors

Packaging phage-λ DNA *in vitro*

- Recombinant DNA in a phage coat allows it to be introduced into the host bacteria by the normal processes of phage infection
- Packaging *in vitro* yields about 10^6 plaques/ µg of vector DNA



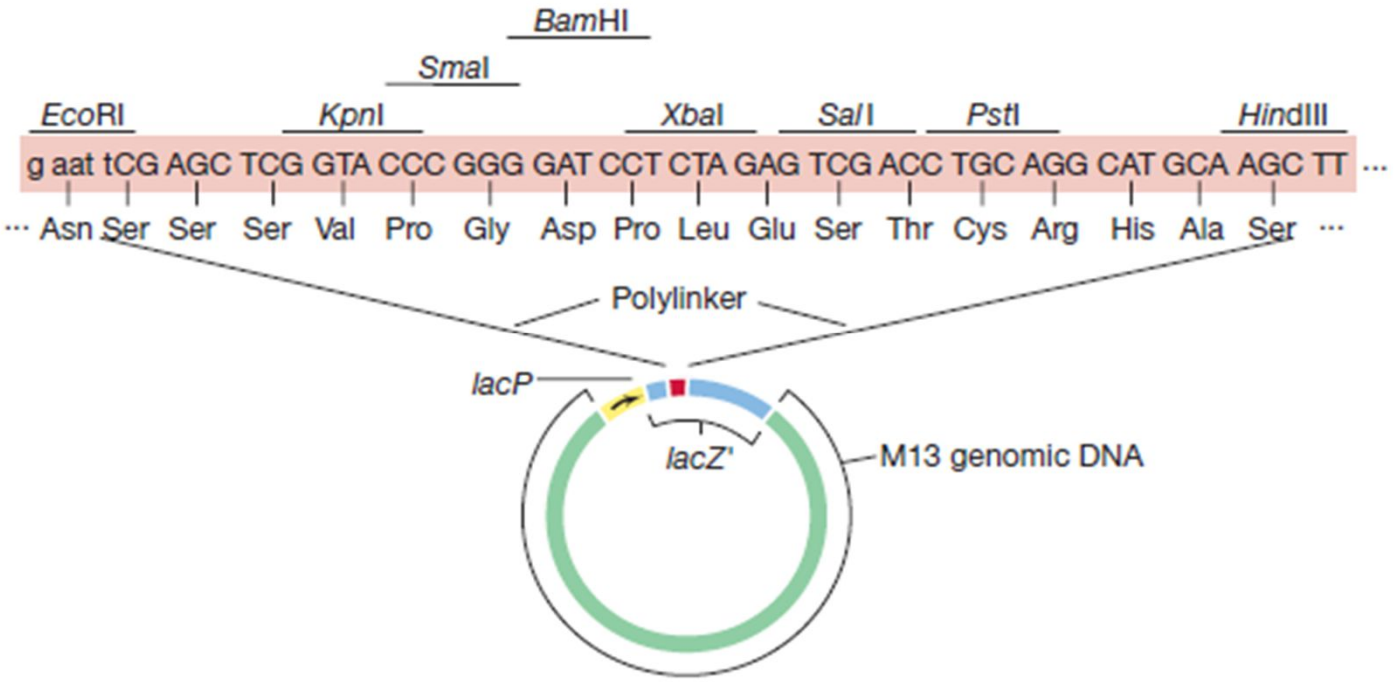
Packaging of phage-ʎ DNA into phage particles

Phage vectors

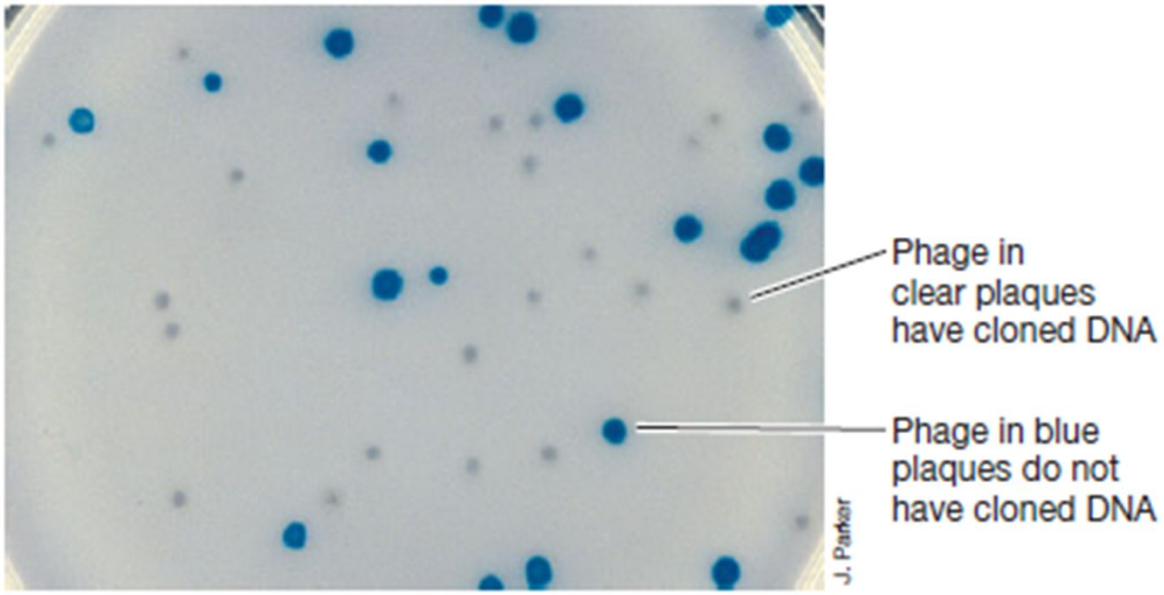
Vectors for DNA sequencing:

bacteriophage M13

- M13 is a filamentous bacteriophage with single stranded DNA that replicates without killing its host
- Variable lengths of DNA upto about 5 kbp, can be cloned



Cloning using bacteriophage M13mp18



Cloning using bacteriophage M13mp18

Phage vectors

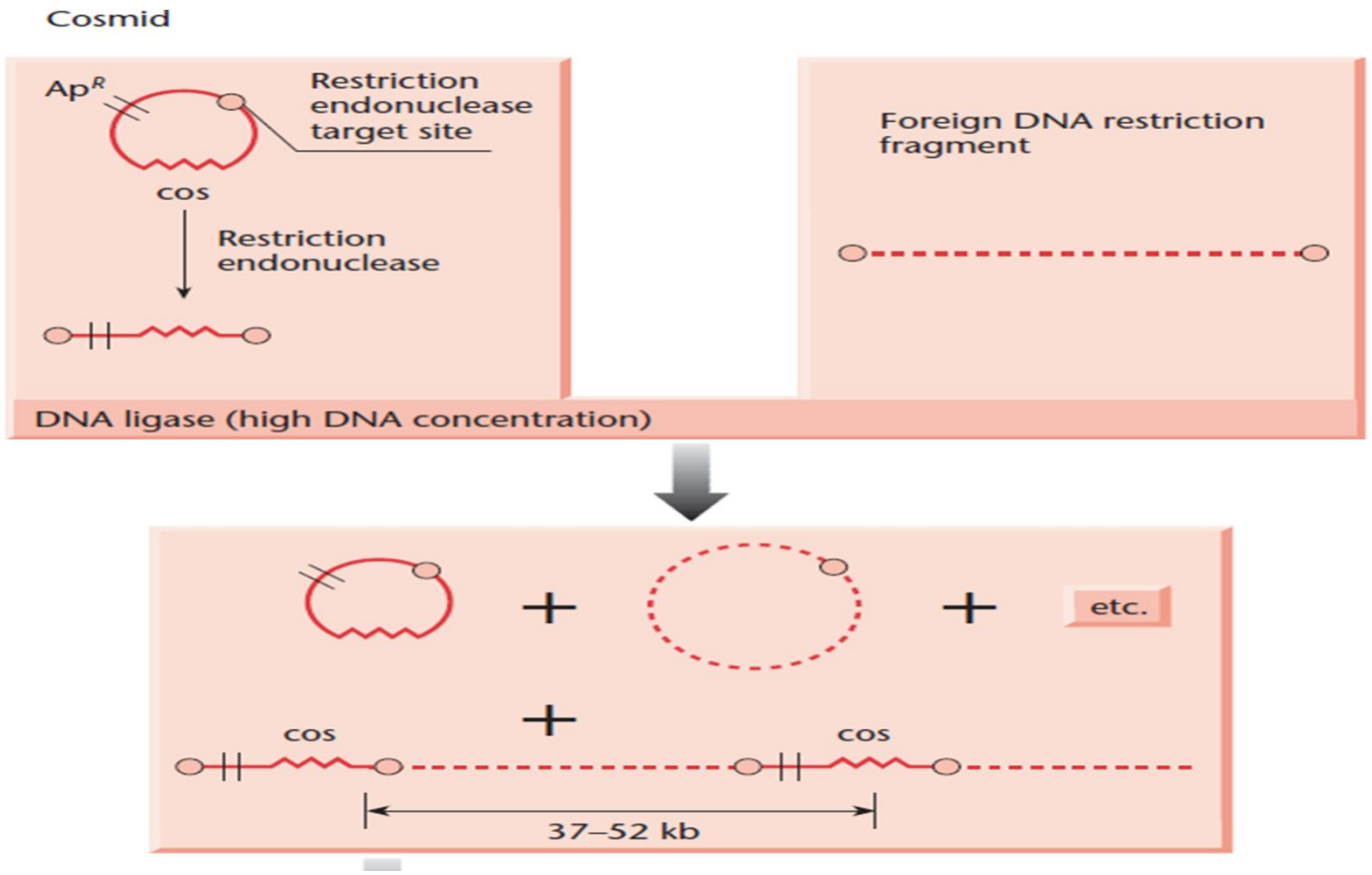
Vectors for DNA sequencing: bacteriophage M13

- **Use of M13 in molecular cloning**

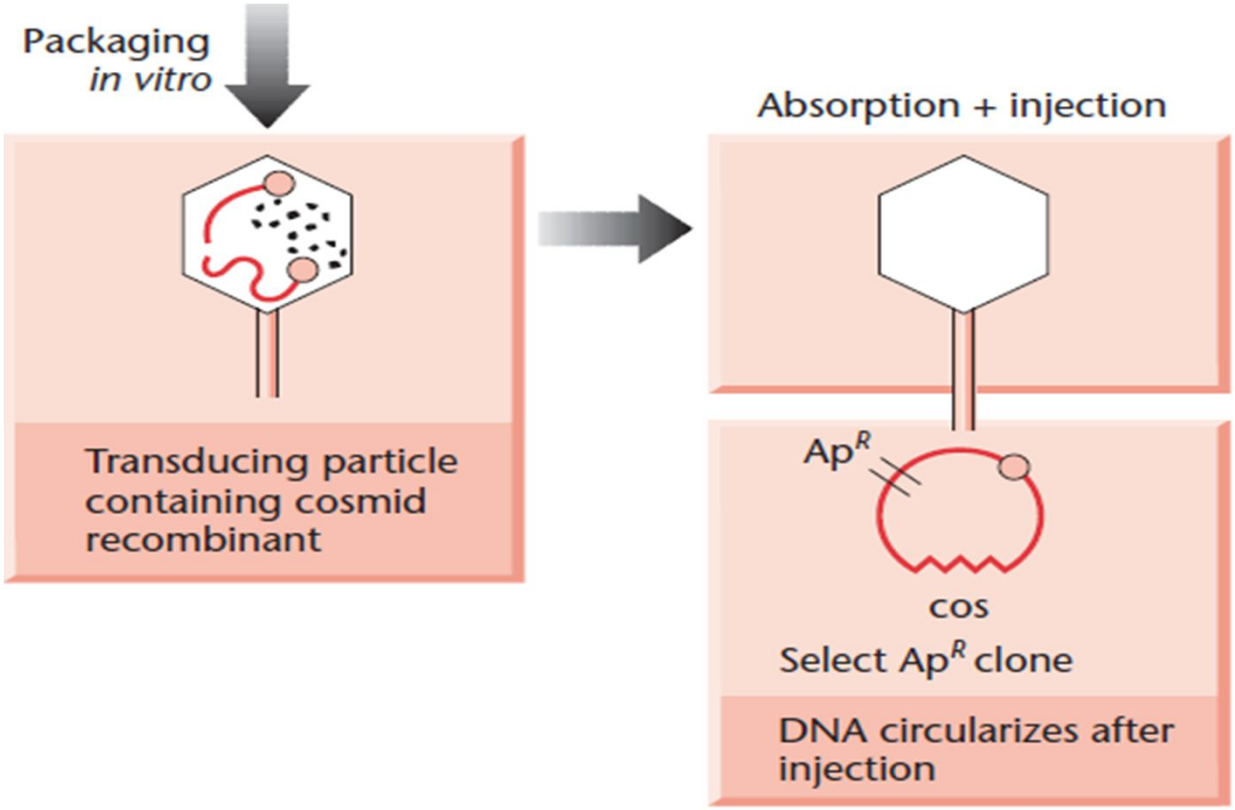
Cosmids, phasmids and other vectors

Cosmid vectors

- Plasmids have been constructed which contain a fragment of λ DNA including '*cos*' site
- These plasmids have been termed cosmids and can be used as gene-cloning vectors in conjunction with the *in vitro* packaging system



Simple scheme for cloning in a cosmid vector



Simple scheme for cloning in a cosmid vector

Cosmids, phasmids and other vectors

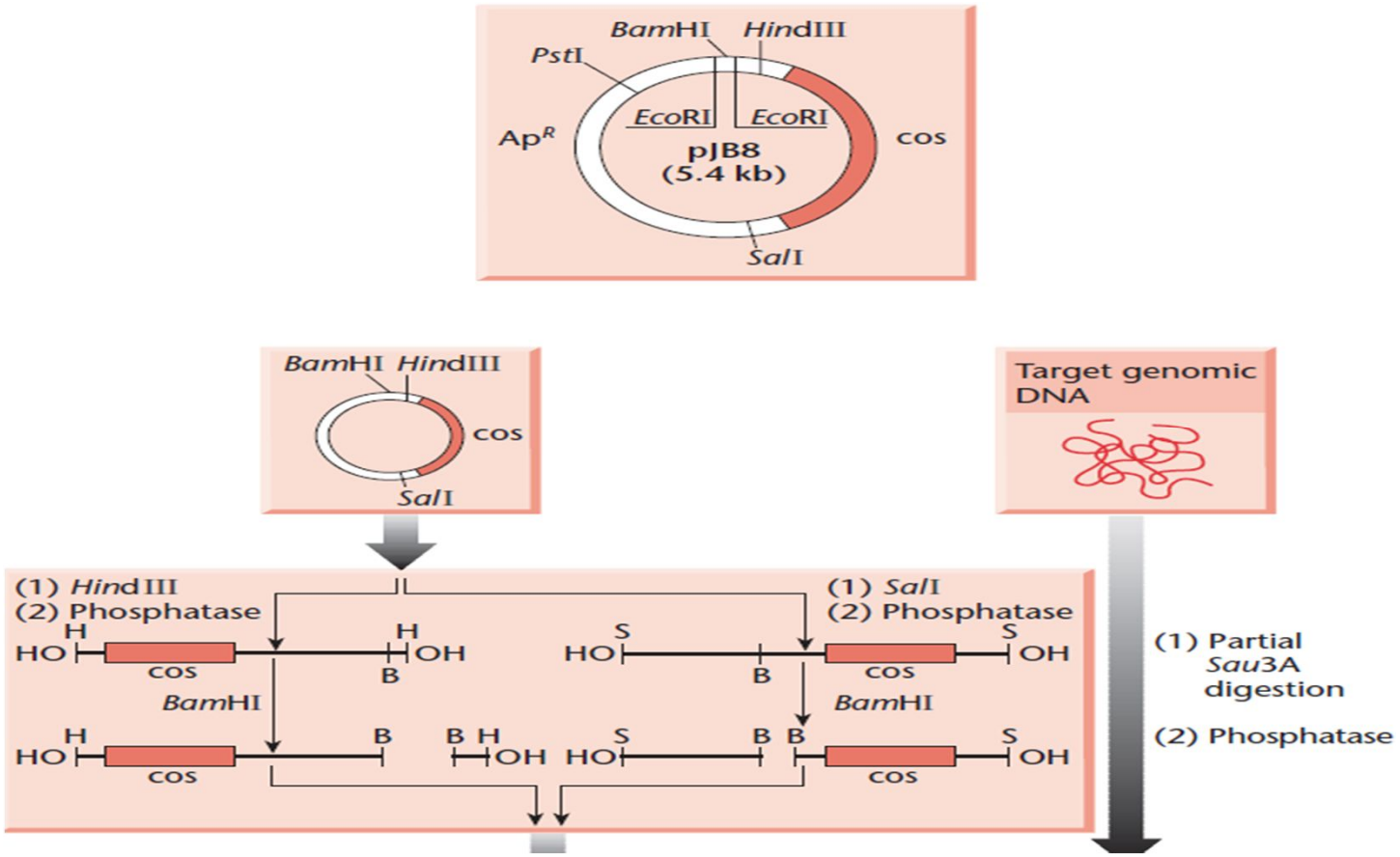
Cosmid vectors

- Cosmids provide an efficient means of cloning large pieces of foreign DNA
- Cosmids are particularly attractive vectors for constructing libraries of eukaryotic genomic fragments

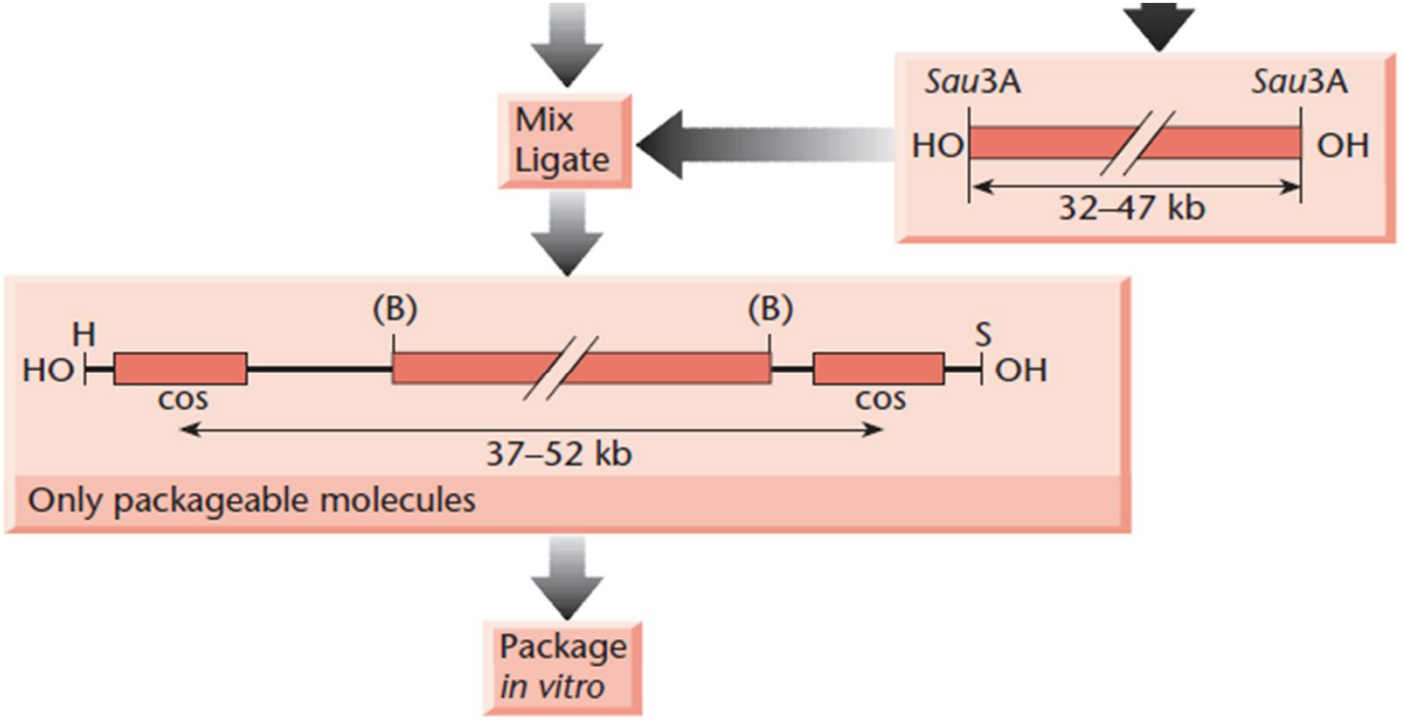
Cosmids, phasmids and other vectors

Modified schemes for cloning in Cosmid vectors

- **Difficulties in cloning with cosmids have been overcome in a cosmid cloning procedure devised by Ish-Horowicz and Burke (1981)**



Plasmid cloning scheme of Ish-Horowicz and Burke (1981)

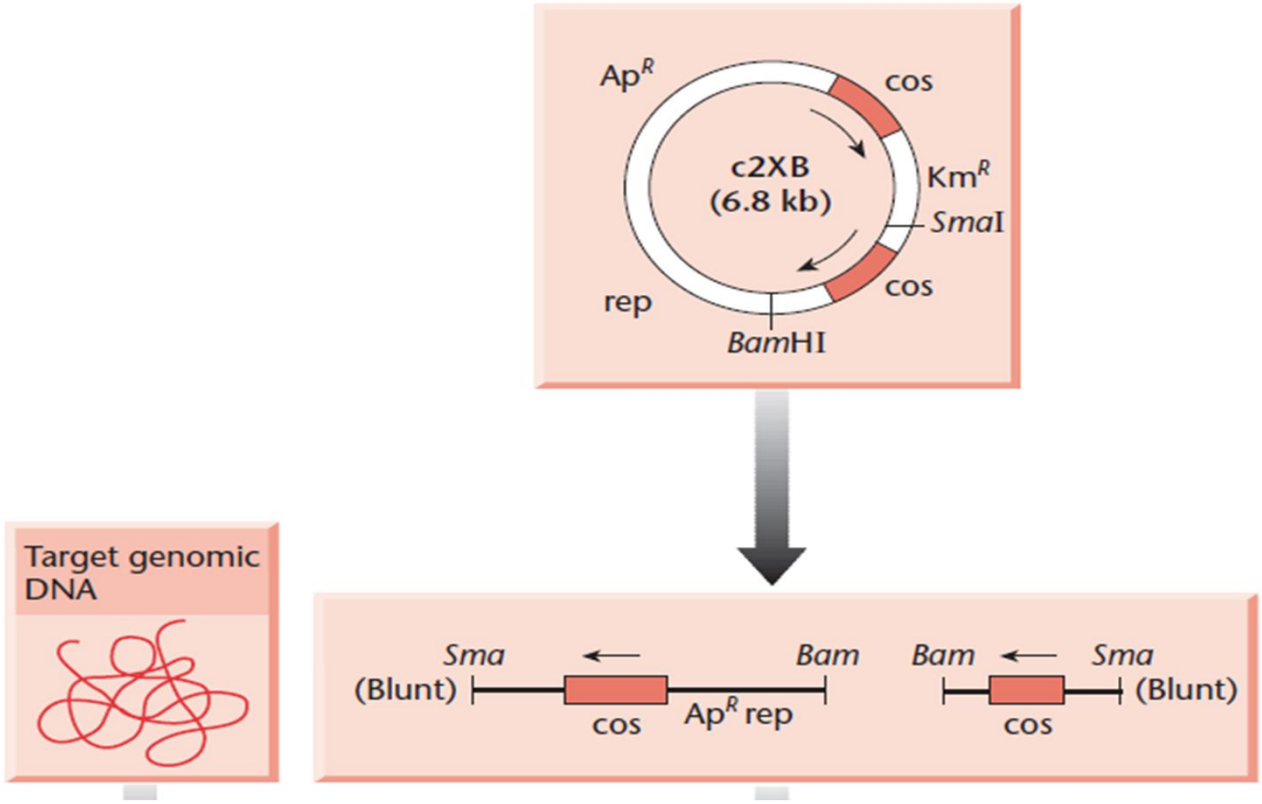


Plasmid cloning scheme of Ish-Horowicz and Burke (1981)

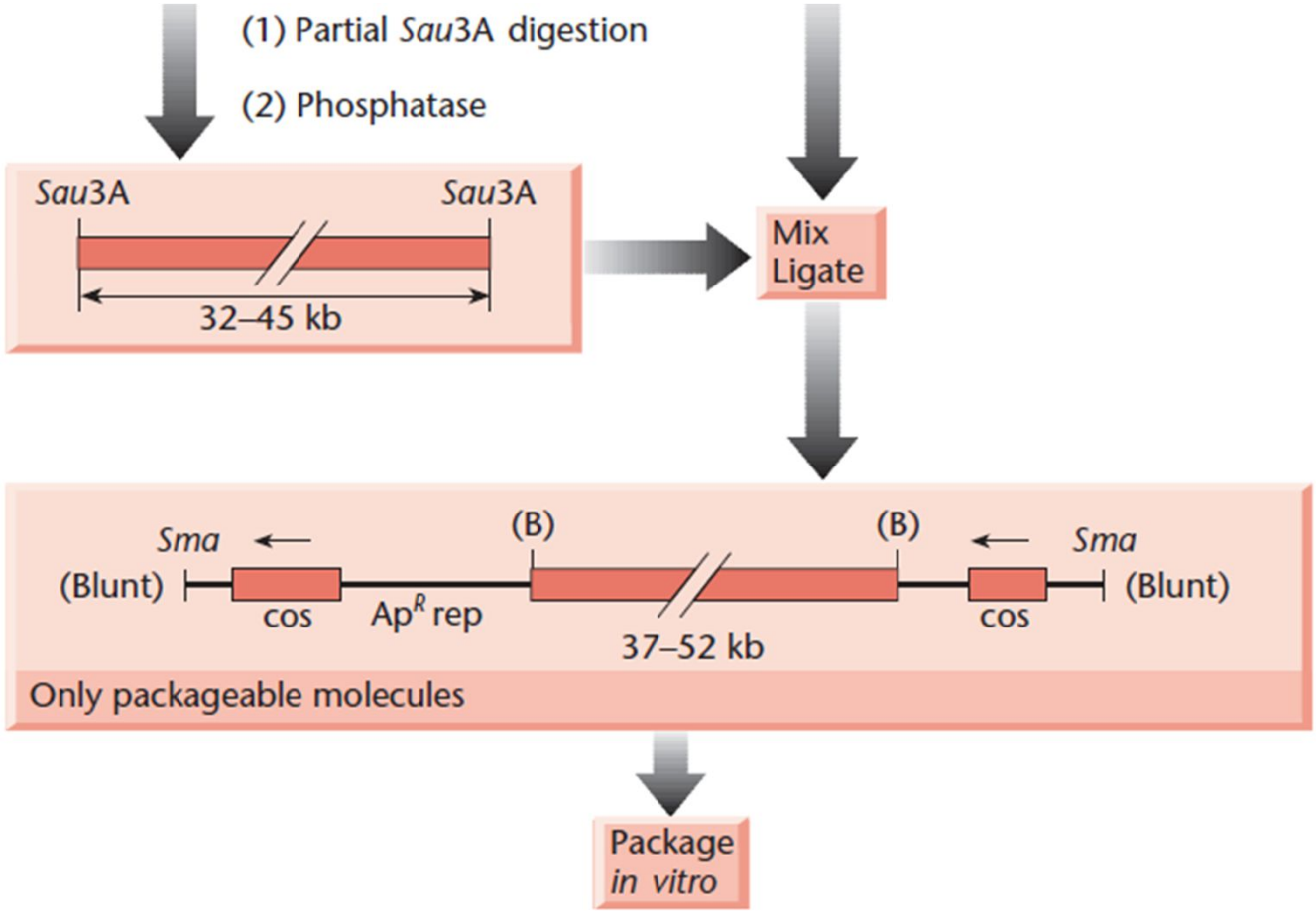
Cosmids, phasmids and other vectors

Modified schemes for cloning in Cosmid vectors

- **An alternate solution to these problems has been devised by Bates and Swift (1983) Who constructed cosmid c_2XB**



Cosmid cloning scheme of Bates and Swift (1983)



Cosmid cloning scheme of Bates and Swift (1983)

Cosmids, phasmids and other vectors

Phasmid vectors

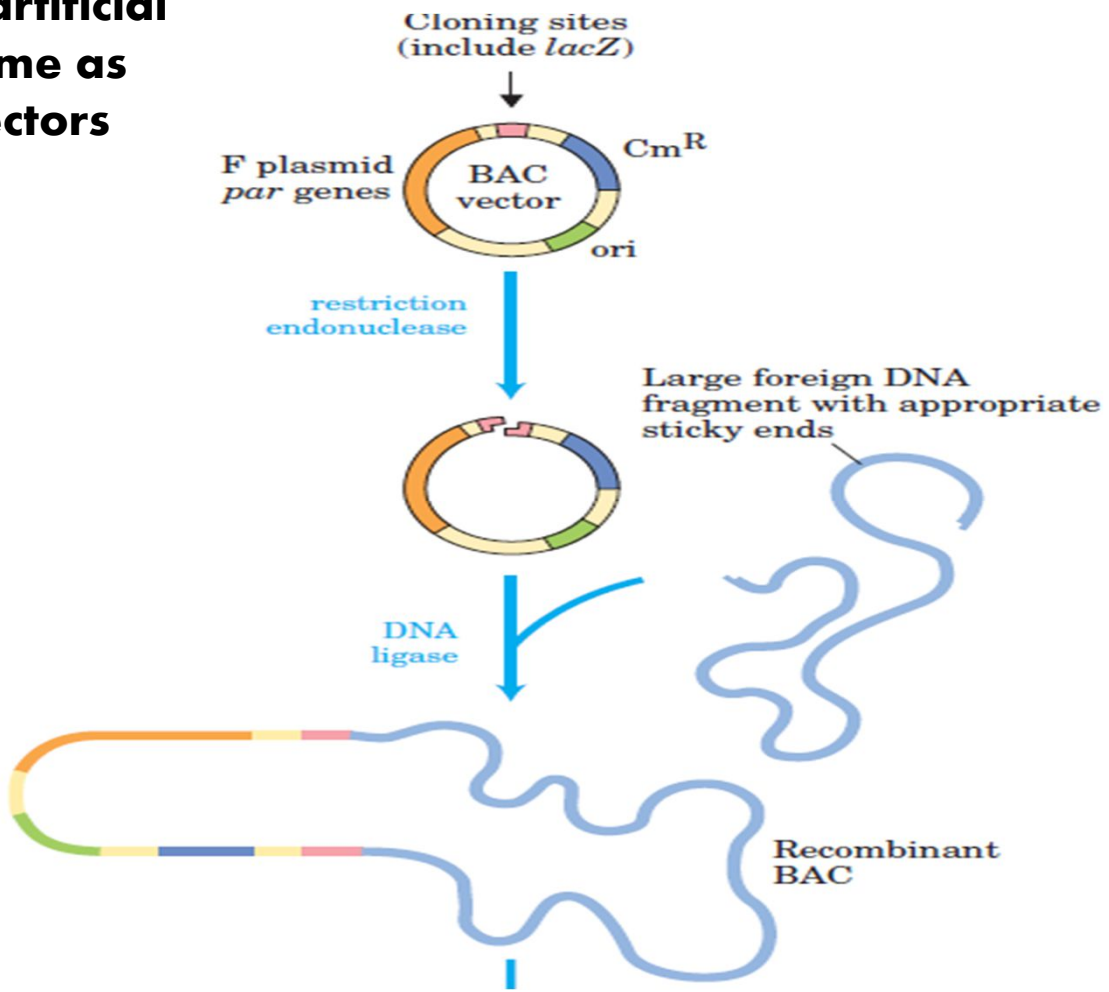
- **Phasmid vectors are combination of plasmid and λ phage sequences**
- **It consists of a plasmid vector carrying a λ attachment (λatt) site**

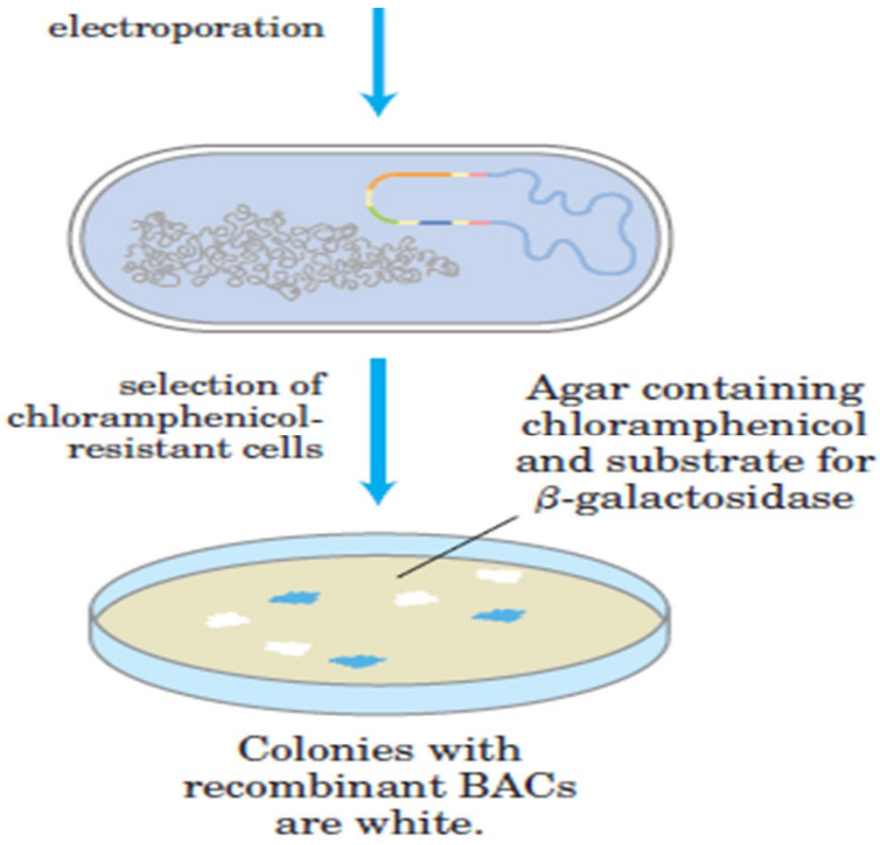
Cosmids, phasmids and other vectors

Bacterial artificial chromosomes (BACs)

- Bacterial artificial chromosomes are simply plasmids designed for the cloning of very long segments of DNA (100,000 to 300,000 bp)

Bacterial artificial chromosome as cloning vectors



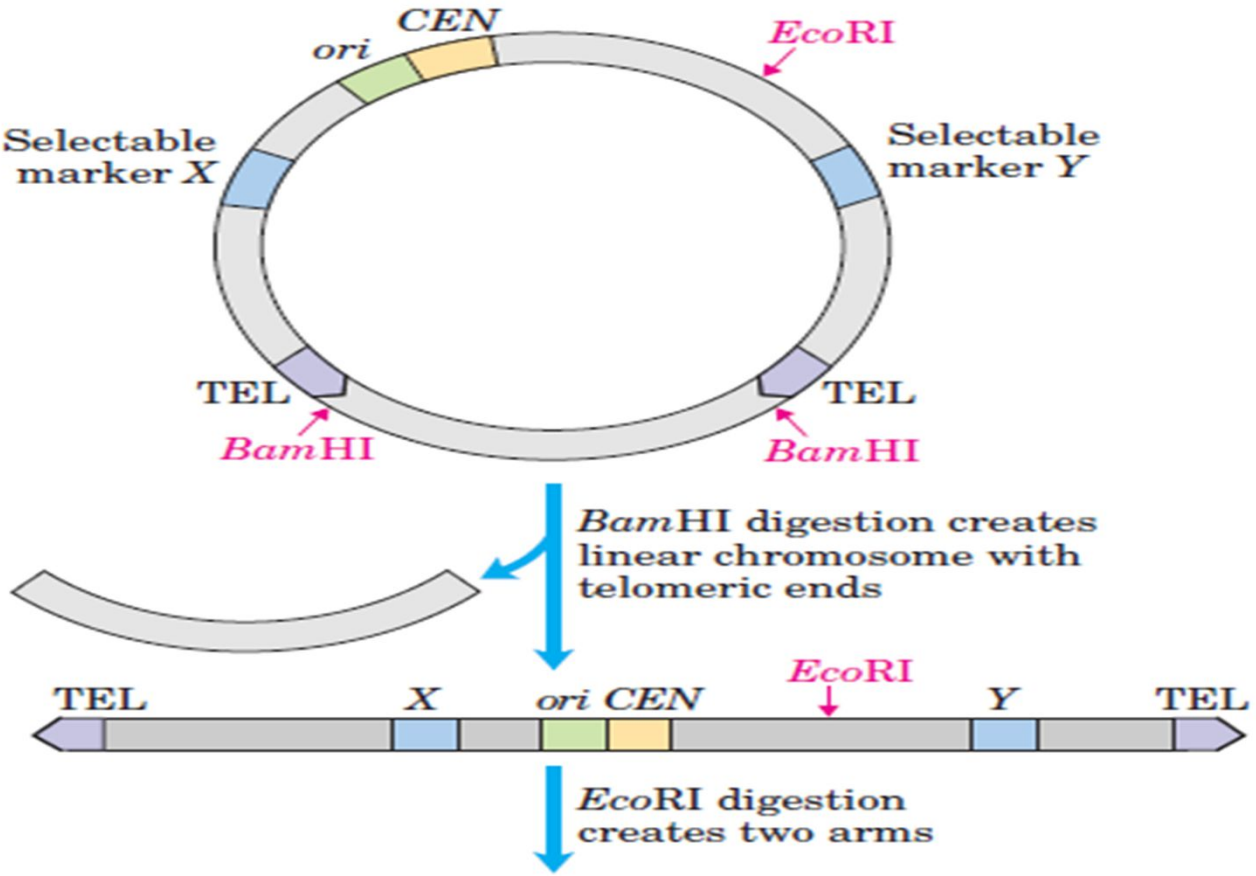


Bacterial artificial chromosome as cloning vectors

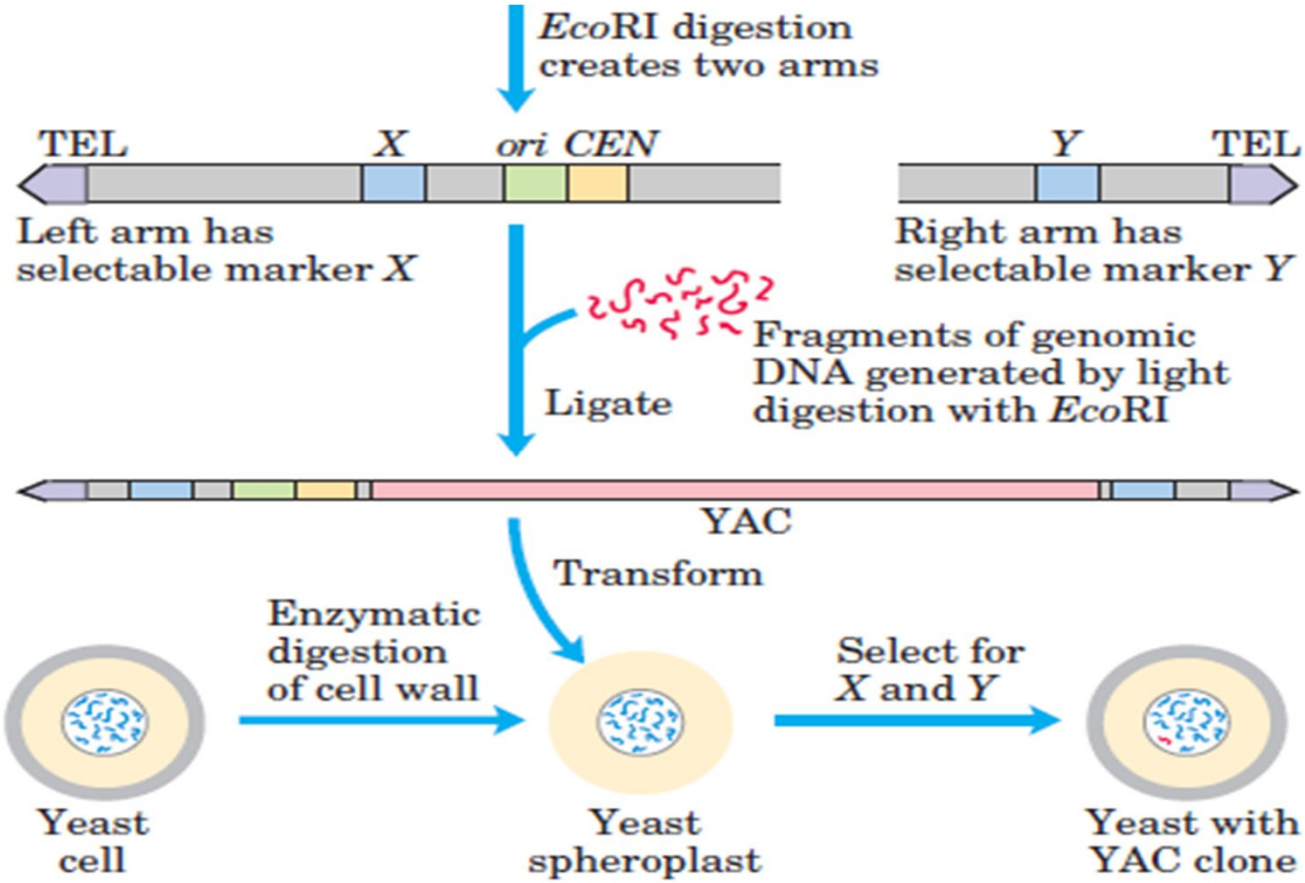
Cosmids, phasmids and other vectors

Yeast artificial chromosomes (YACs)

- These vectors replicate in yeast like normal chromosome but they have sites where very large DNA fragments can be inserted



Construction of yeast artificial chromosome



Construction of yeast artificial chromosome

Cosmids, phasmids and other vectors

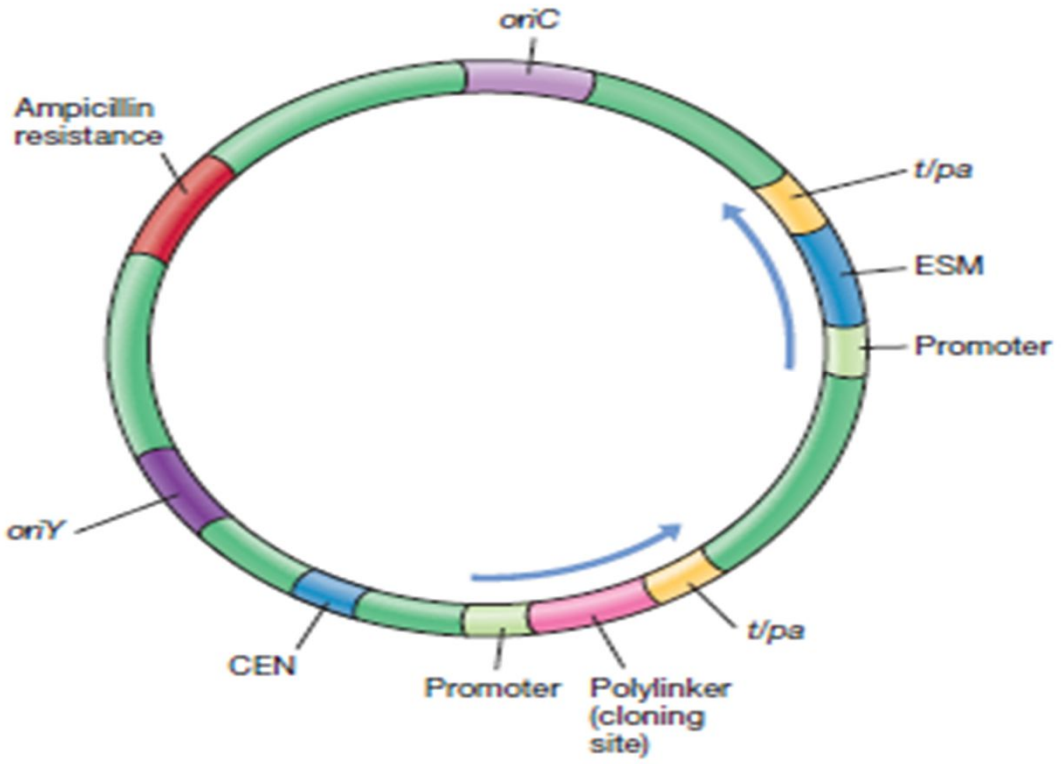
Yeast artificial chromosomes (YACs)

- YACs with inserts more than 150,000 bp are nearly as stable as normal cellular chromosomes

Cosmids, phasmids and other vectors

Shuttle and Expression vectors

- **Vectors that can replicate and are stably maintained in two or more unrelated host organisms are called shuttle vectors**



Genetic map of a shuttle vector used in yeast

Cosmids, phasmids and other vectors

Shuttle and Expression vectors

- To control the expression of cloned genes expression vectors are used

Cosmids, phasmids and other vectors

Comparison of different cloning vectors-summary

- Different cloning vectors can carry variable sizes of target DNA into host organisms

Cosmids, phasmids and other vectors

Table. Maximum DNA insert possible with different cloning vectors

Vector	Host	Insert size
pBR322	*E. coli*	10-15 kb
ʎ phage	*E. coli*	2-25 kb
ʎ Cosmids	*E. coli*	35-45 kb
P1 phage	*E. coli*	70-100 kb
PACs	*E. coli*	100-300 kb
BACs	*E. coli*	≤ 300 kb
YACs	*S. cerevisiae*	200-2000 kb

Cloning strategies

OVERVIEW OF CLONING STRATEGIES

- **Any cell based cloning procedure has four essential parts**

i). A method of generating DNA fragment for cloning

ii). A reaction that inserts that fragment into the chosen cloning vectors

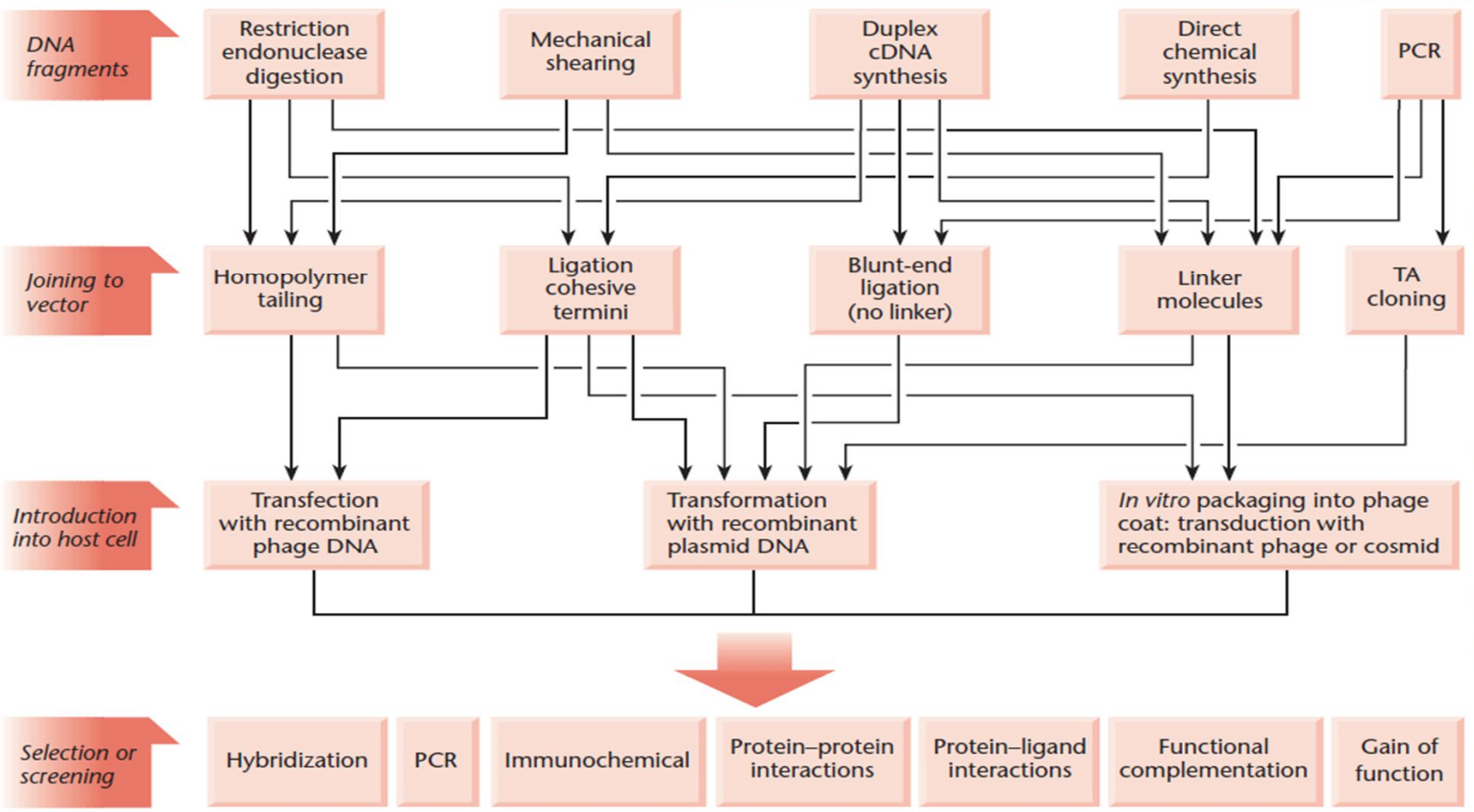
Cloning strategies

Introduction

iii). A means for introducing that recombinant vector into a host cell wherein it is replicated

iv). A method of selecting recipient cells that have acquired the recombinant

Cloning strategies



Generalized overview of cloning strategies

Cloning strategies

Introduction

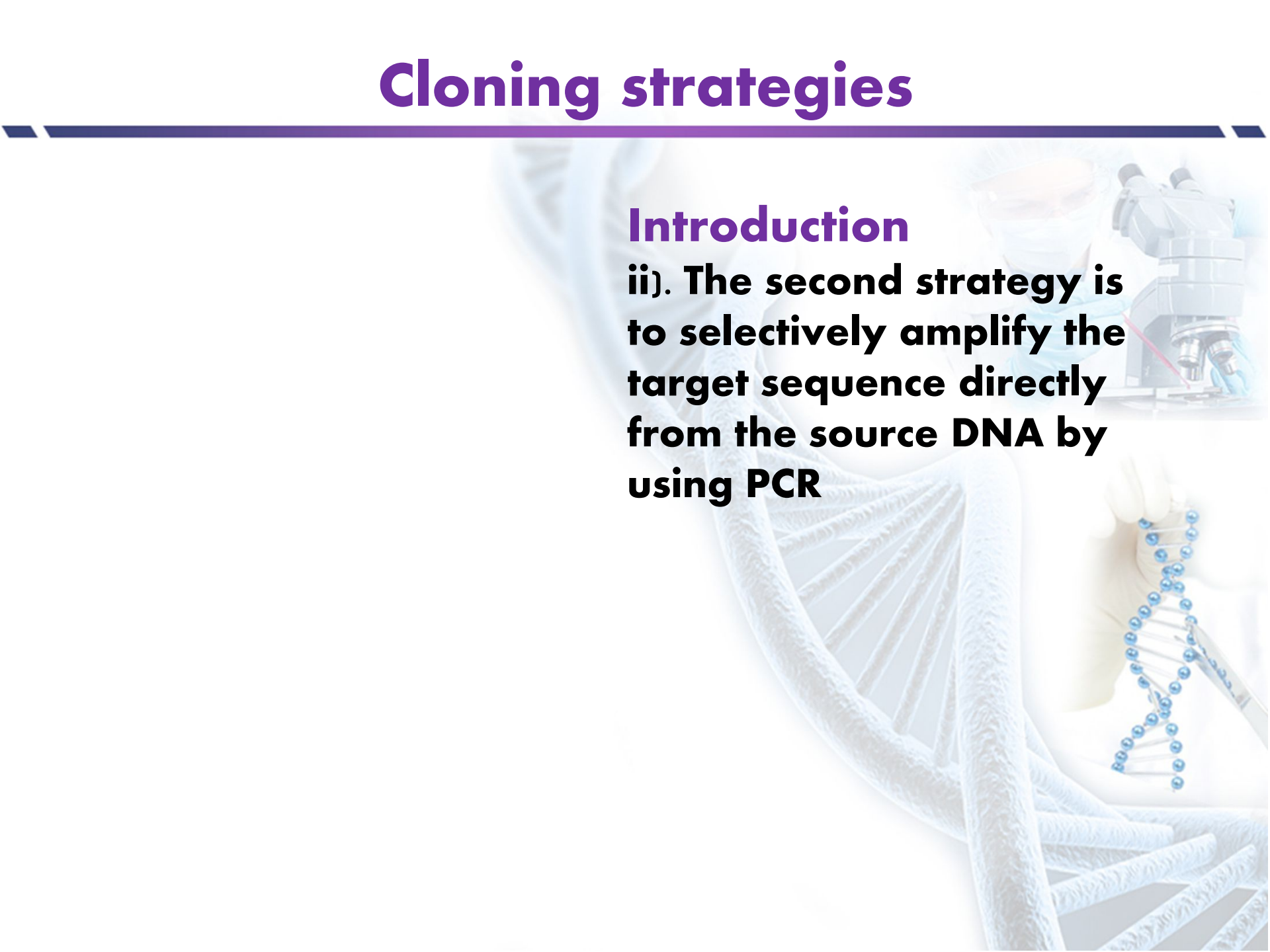
- There are two major strategies for isolating sequences from complex sources

i). Cell based cloning strategy, is to divide the source DNA into manageable fragments and clone everything

Cloning strategies

Introduction

ii). The second strategy is to selectively amplify the target sequence directly from the source DNA by using PCR



Cloning genomic DNA

Genomic DNA libraries

- **Producing representative genomic libraries in λ cloning vector**

Cloning genomic DNA

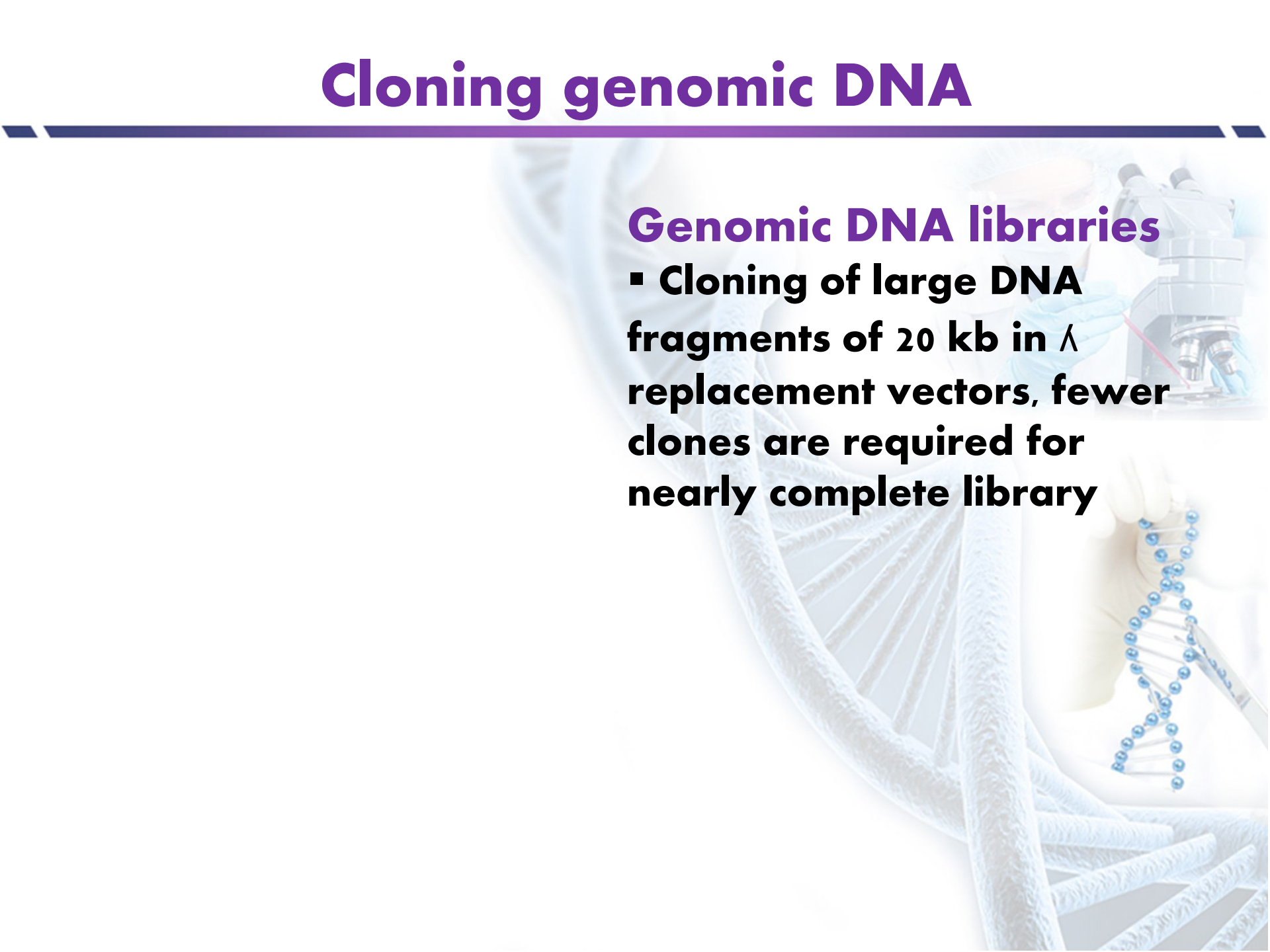
Table 1. Genome size of selected organisms

Organism	Genome size (kb) (haploid where appropriate)
Escherichia coli	4.0×10^3
Yeast (*Saccharomyces cerevisiae*)	1.35×10^4
Arabidopsis thaliana (higher plant)	1.25×10^5
Tobacco	1.6×10^6
Wheat	5.9×10^6
Zea mays	1.5×10^7
Drosophila melanogaster	1.8×10^5
Mouse	2.3×10^6
Human	2.8×10^6
Xenopus laevis	3.0×10^6

Cloning genomic DNA

Genomic DNA libraries

- Cloning of large DNA fragments of 20 kb in λ replacement vectors, fewer clones are required for nearly complete library



Cloning genomic DNA

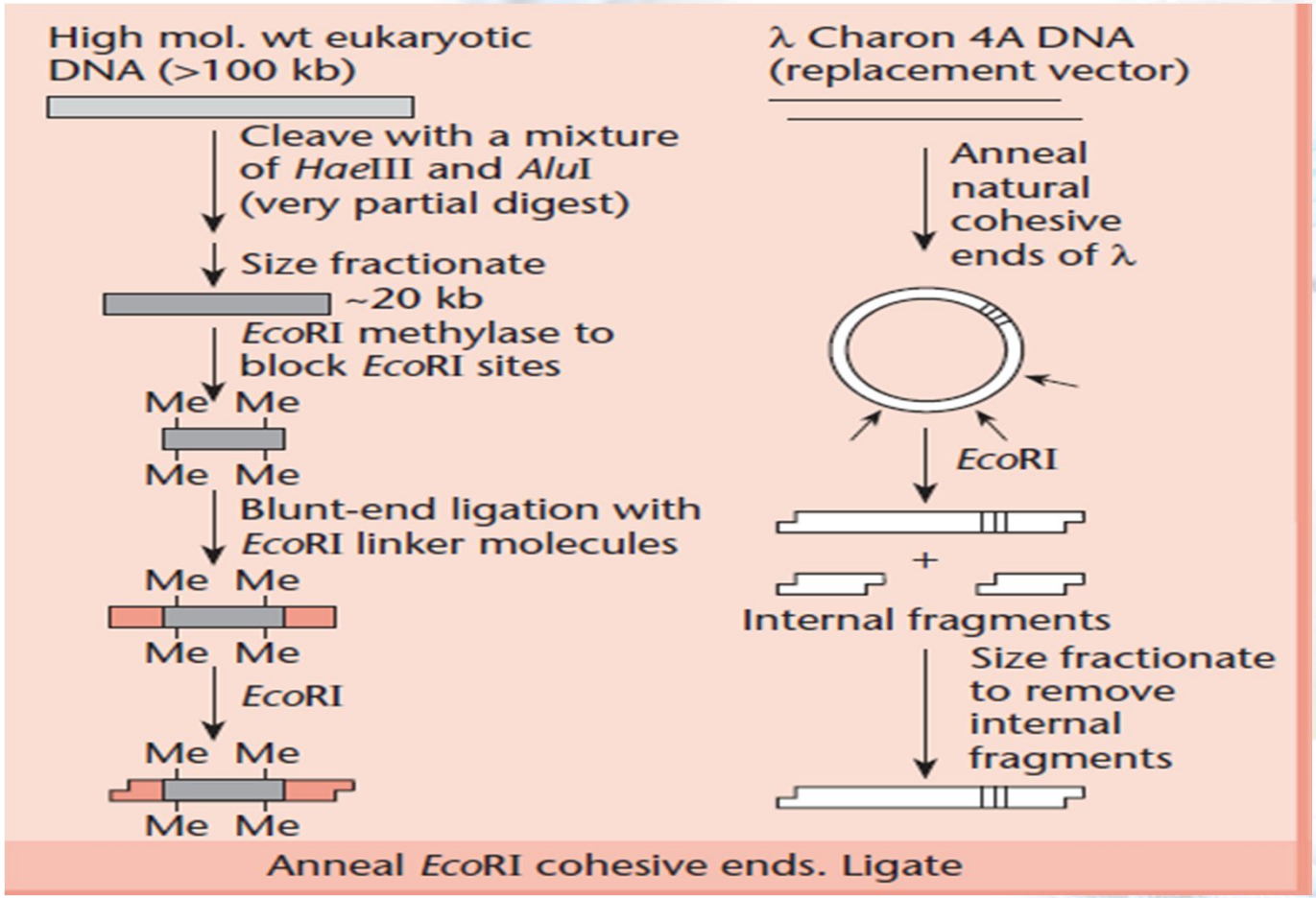
Genomic DNA libraries

- How can appropriately sized random fragments be produced?

i). Average fragment size can be controlled by mechanical shearing

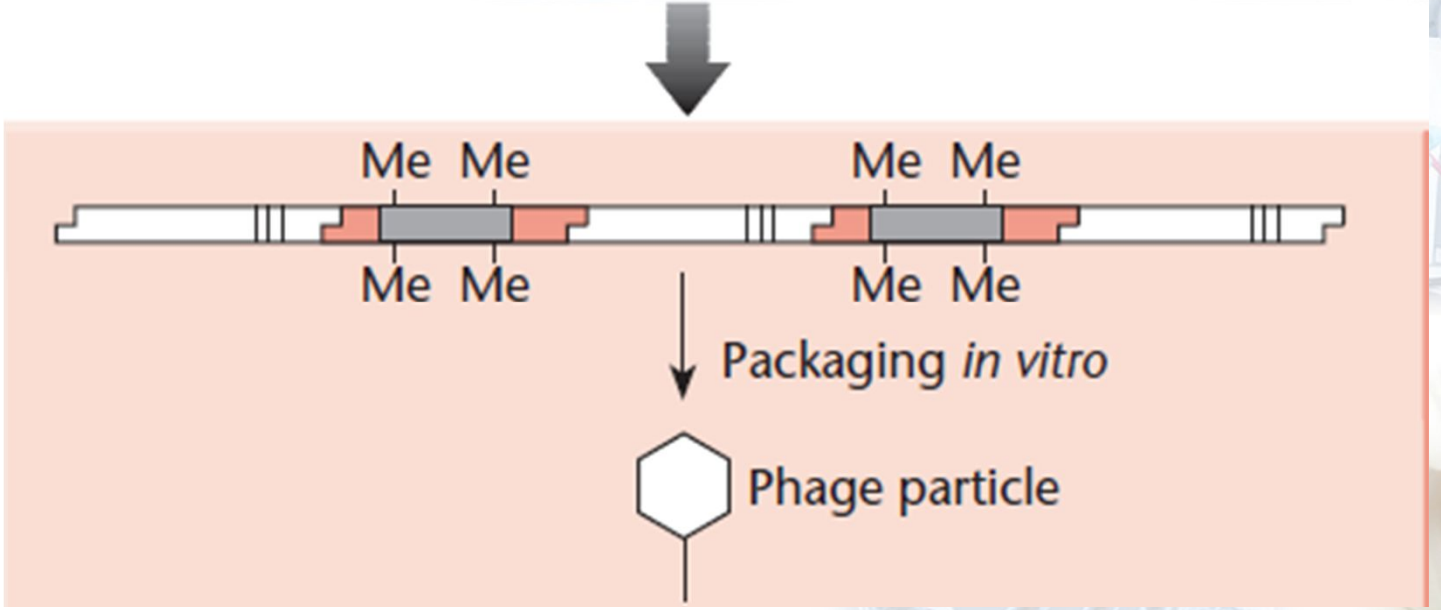
ii). The more commonly used procedure is restriction endonucleases

Cloning genomic DNA



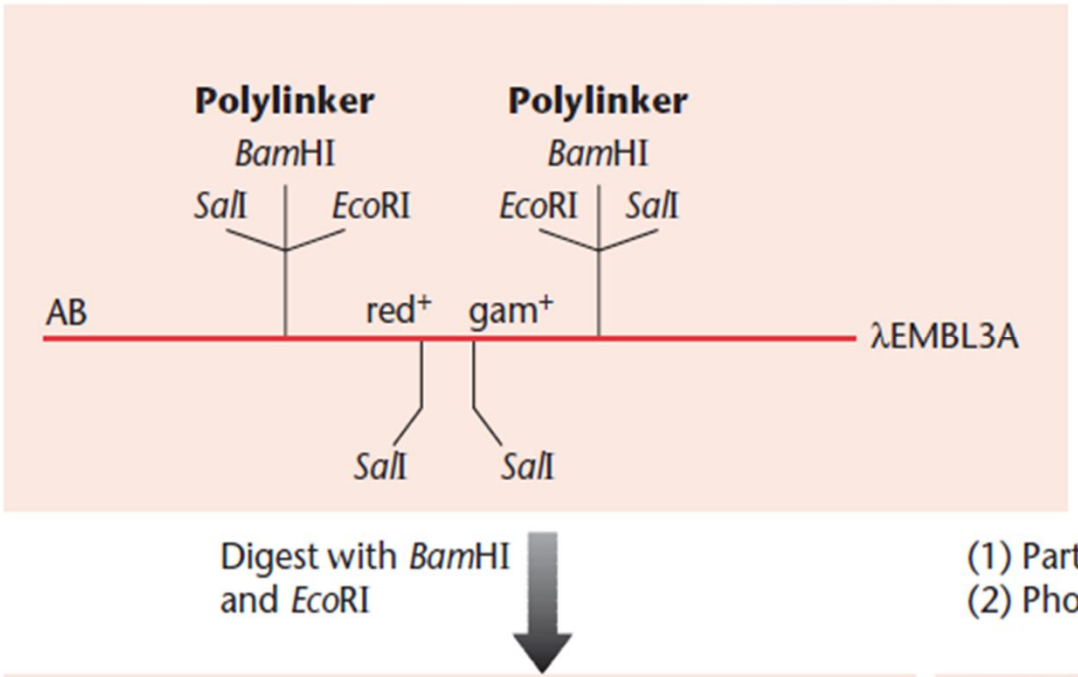
Maniatis' strategy for producing a representative gene library

Cloning genomic DNA



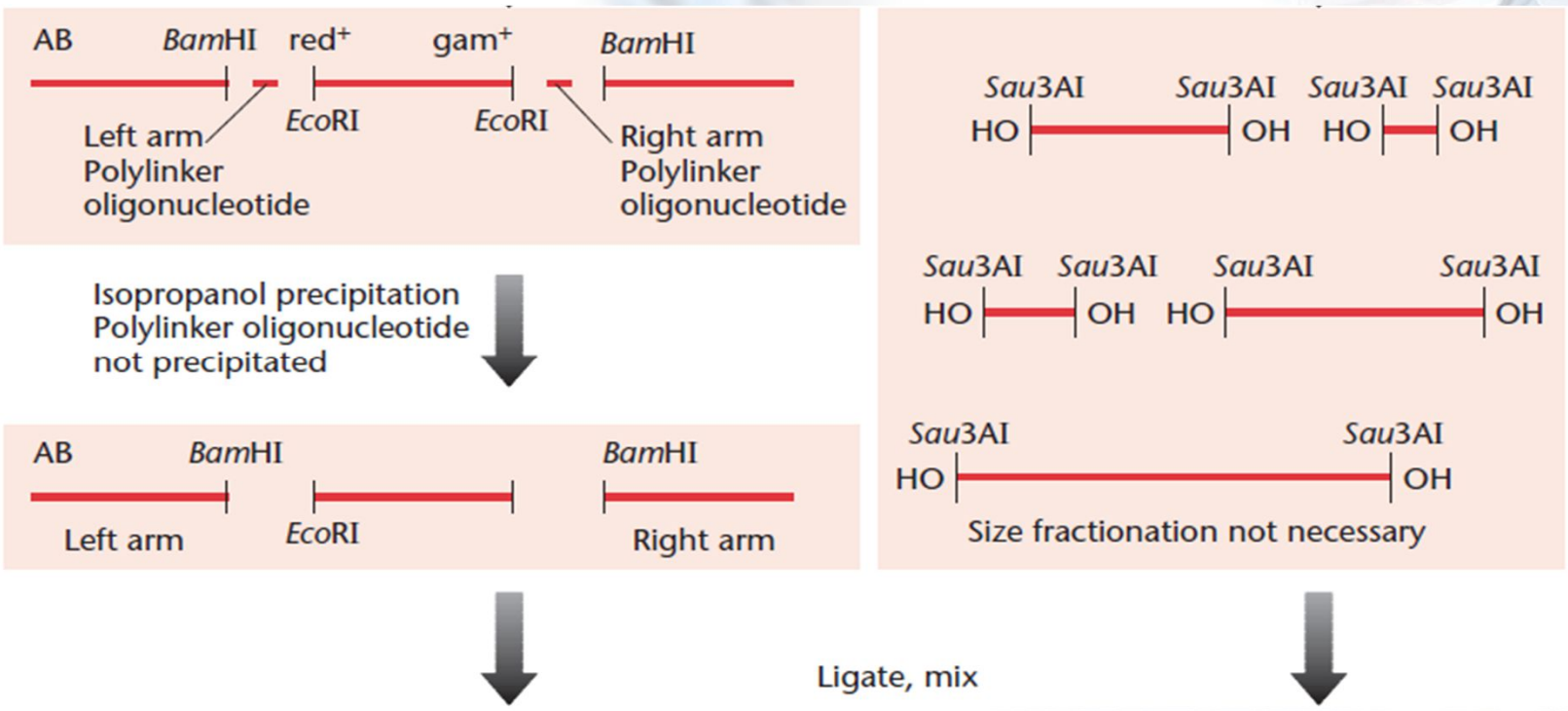
Maniatis' strategy for producing a representative gene library

Cloning genomic DNA



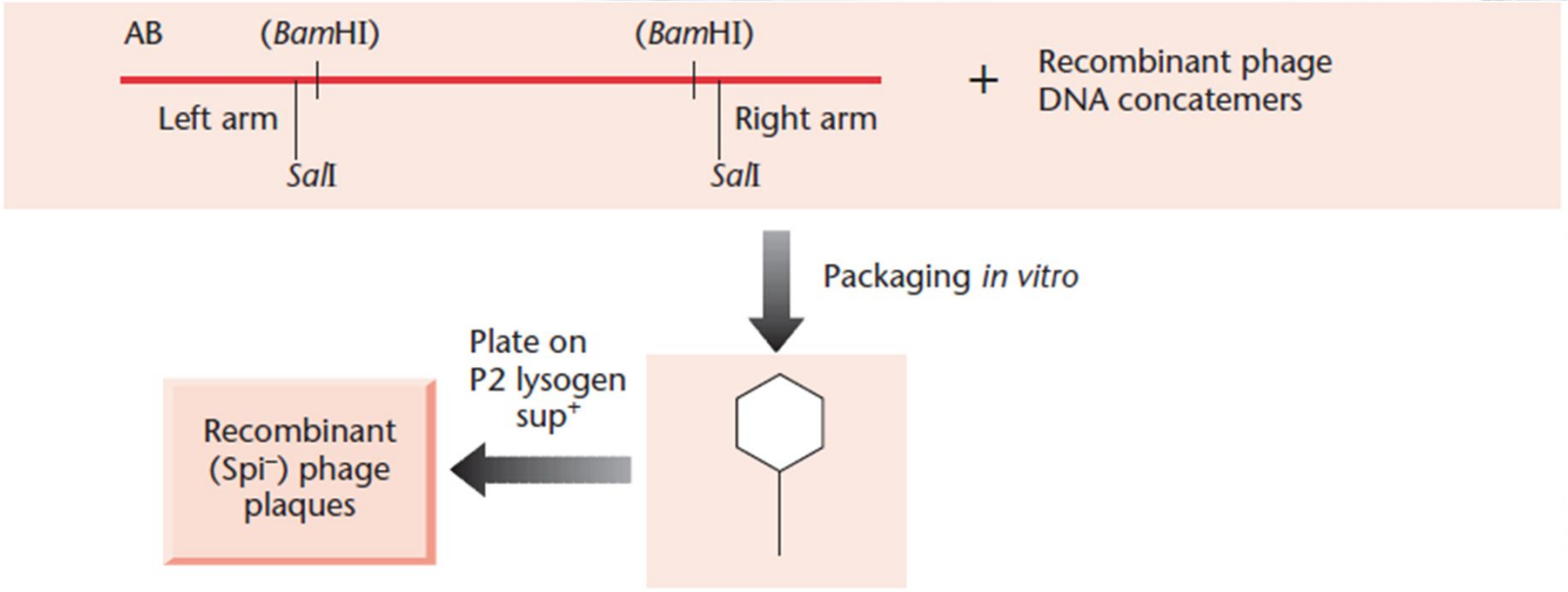
Creation of genomic DNA library using the phage-ʎ vector EMBL3A

Cloning genomic DNA



Creation of genomic DNA library using the phage-λ vector EMBL3A

Cloning genomic DNA



Creation of genomic DNA library using the phage-ʎ vector EMBL3A

Cloning genomic DNA

Genomic libraries in high-capacity vectors

- In place of phage-λ derivatives, a number of high capacity cloning vectors such as cosmids, bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) are available for construction of genomic libraries

Cloning genomic DNA

PCR as an alternative to genomic DNA cloning

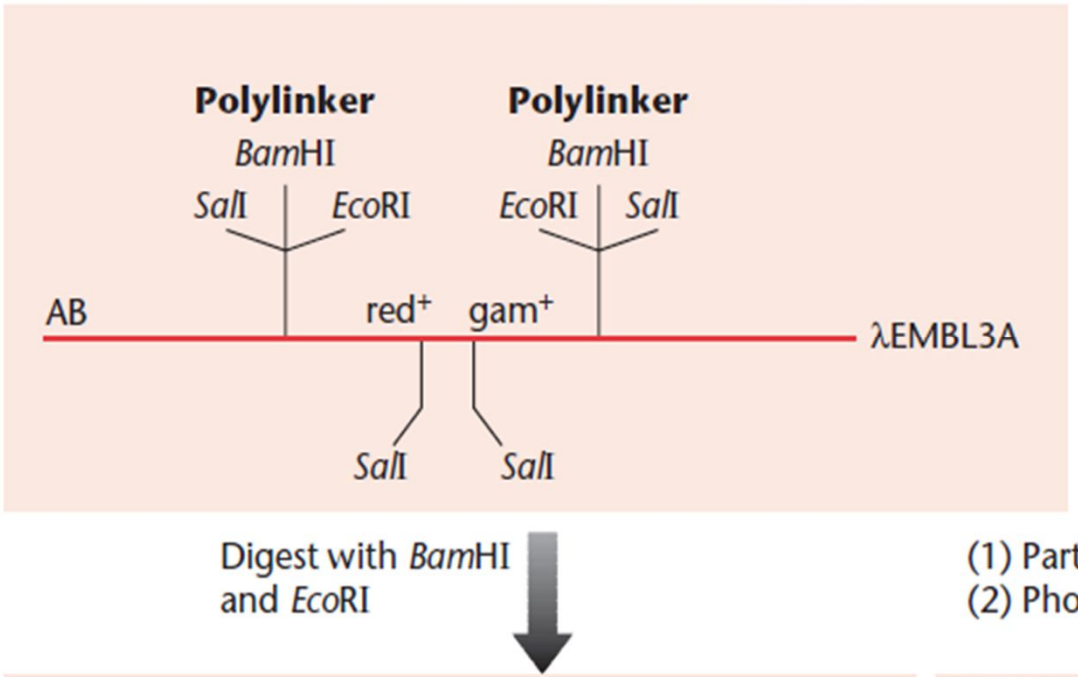
- **PCR with specific primers could be used to isolate genes directly from genomic DNA, obviating the need for the production of genomic libraries**

Cloning genomic DNA

λEMBL vectors for genomic library construction

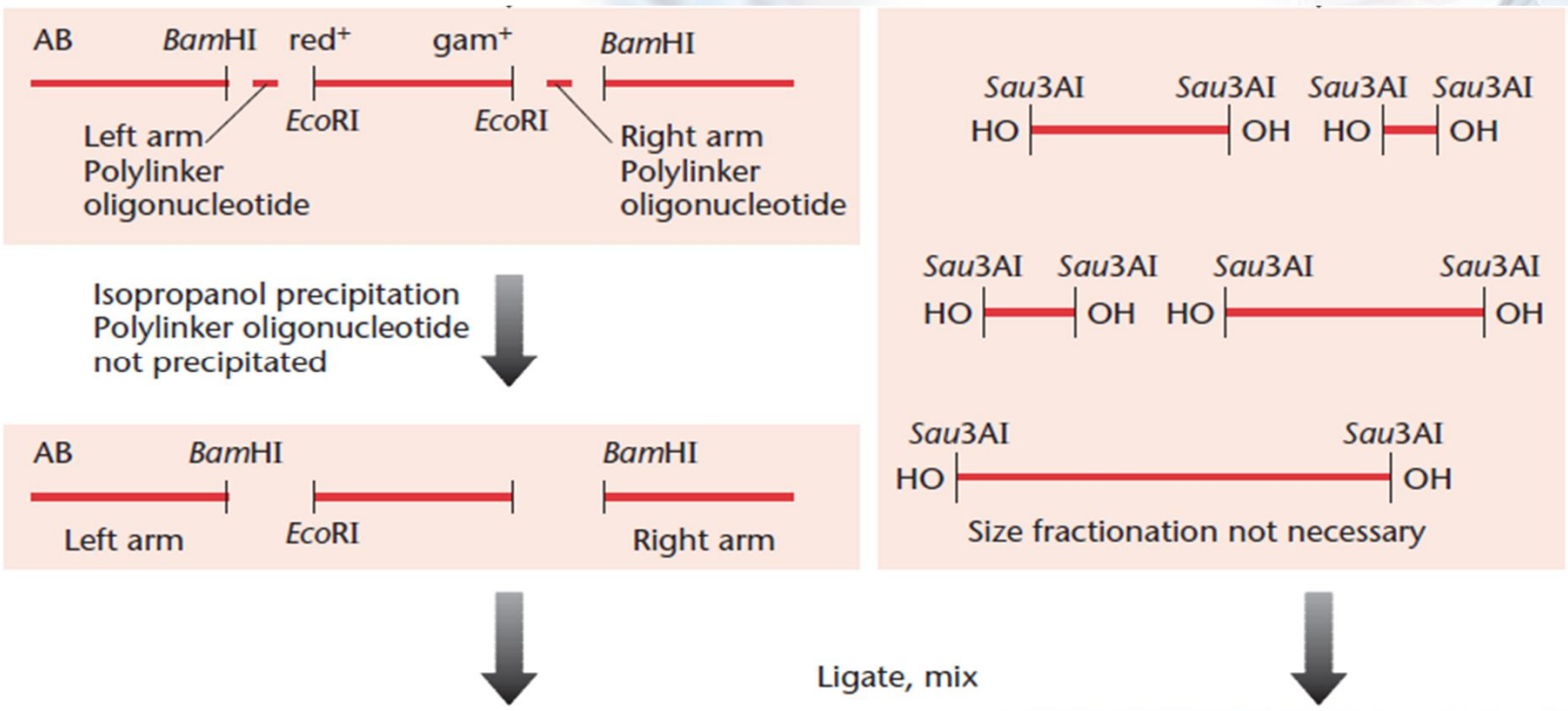
- High molecular weight genomic DNA is digested with *Sau*3AI and subsequently ligated in λEMBL vector digested with *Bam*H1

Cloning genomic DNA



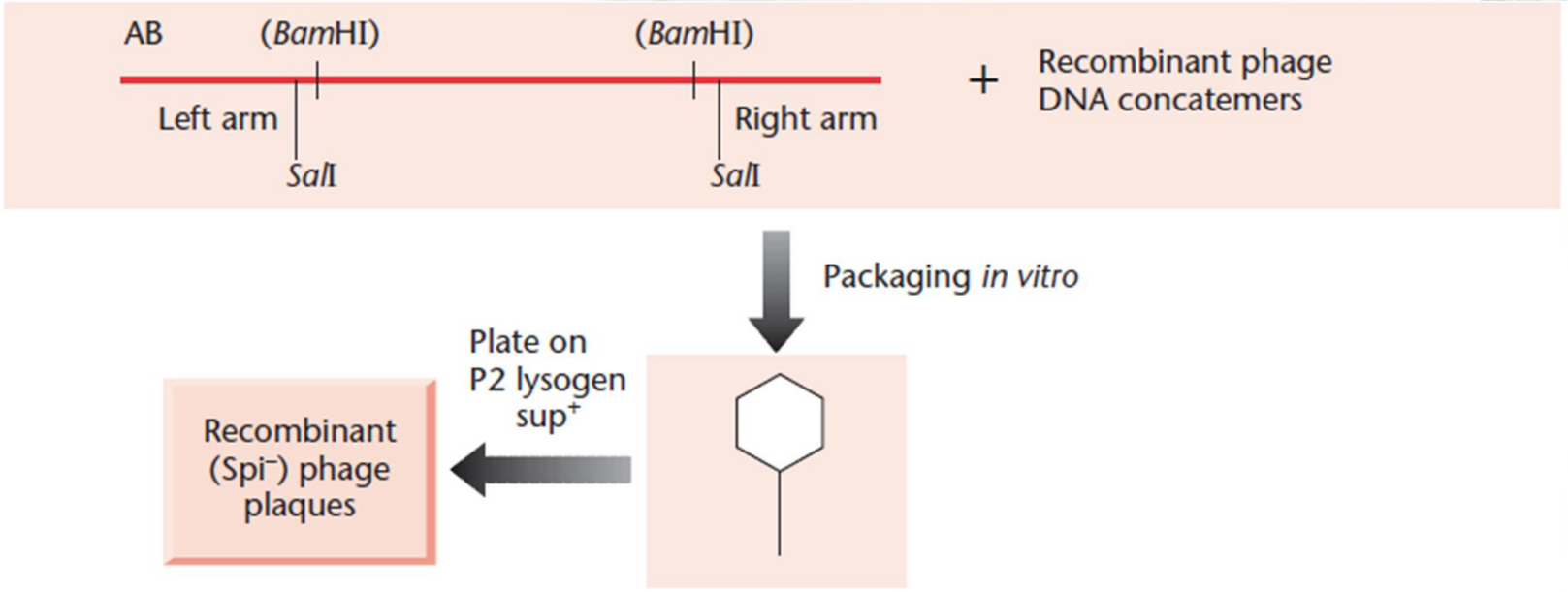
Creation of genomic DNA library using the phage-ʎ vector EMBL3A

Cloning genomic DNA



Creation of genomic DNA library using the phage-λ vector EMBL3A

Cloning genomic DNA



Creation of genomic DNA library using the phage-ʎ vector EMBL3A

Cloning genomic DNA

Genomic libraries in high-capacity vectors

- In place of phage-λ derivatives, a number of high capacity cloning vectors such as cosmids, bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) are available for construction of genomic libraries

Cloning genomic DNA

PCR as an alternative to genomic DNA cloning

- **PCR with specific primers could be used to isolate genes directly from genomic DNA, obviating the need for the production of genomic libraries**

cDNA cloning

Properties of cDNA

- cDNA is prepared by reverse-transcribing cellular RNA
- cDNA lack introns and other non-coding sequences present in the corresponding genomic DNA
- Introns are rare in bacteria but occur in genes of higher eukaryotes

cDNA cloning

cDNA libraries

- cDNA library is a combination of cloned cDNA fragments inserted into a collection of host cells
- cDNA library is representative of the RNA population from which it was derived

cDNA cloning

Table. Abundance classes of typical mRNA populations

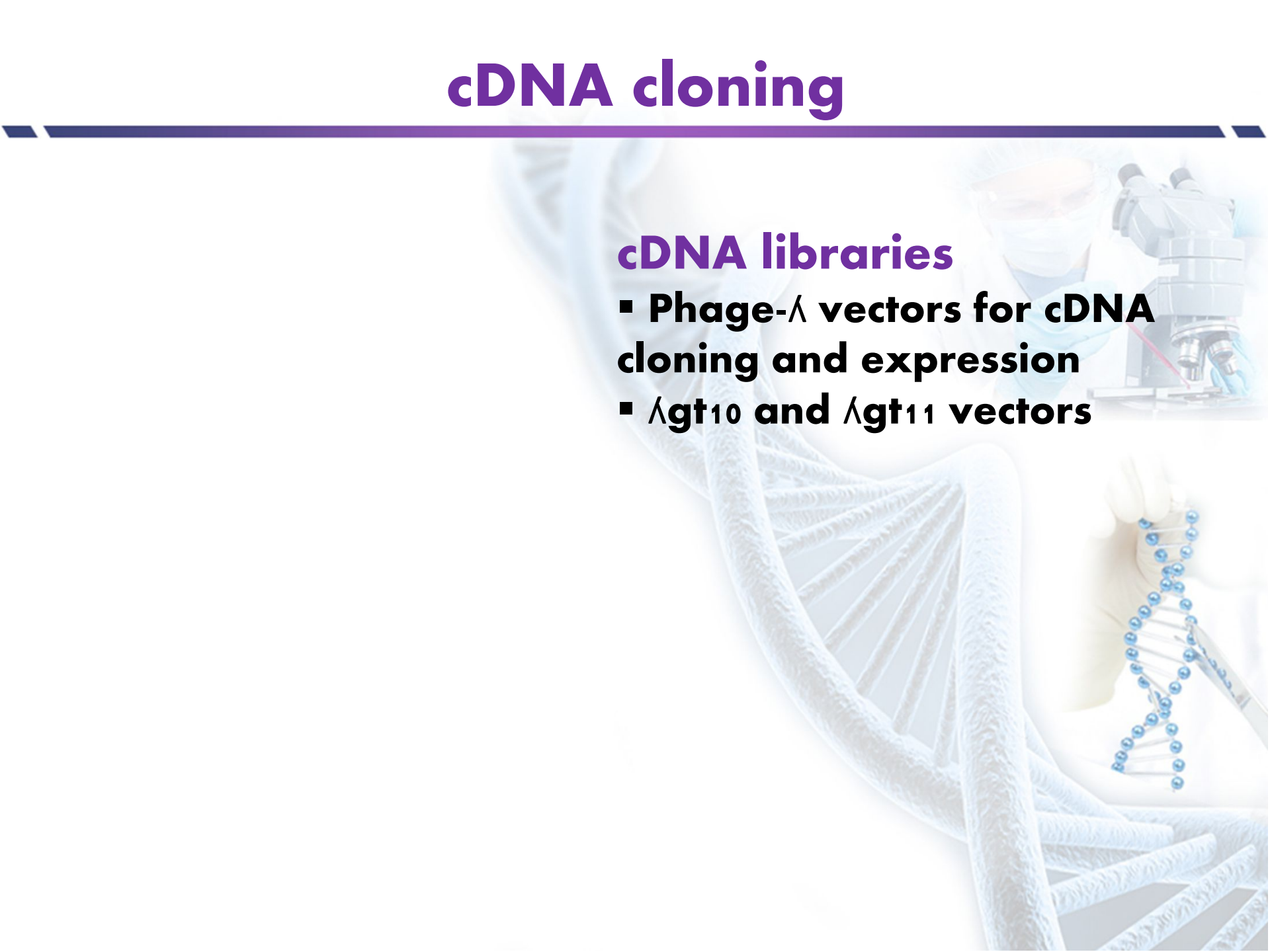
Source	Number of different mRNAs	Abundance (molecules/cell)
Mouse liver cytoplasmic poly(A)$^+$	9	12 000
	700	300
	11 500	15
Chick oviduct polysomal poly(A)$^+$	1	100 000
	7	4 000
	12 500	5

References: mouse (Young *et al.* 1976); chick oviduct (Axel *et al.* 1976).

cDNA cloning

cDNA libraries

- **Phage-λ vectors for cDNA cloning and expression**
- **λgt10 and λgt11 vectors**

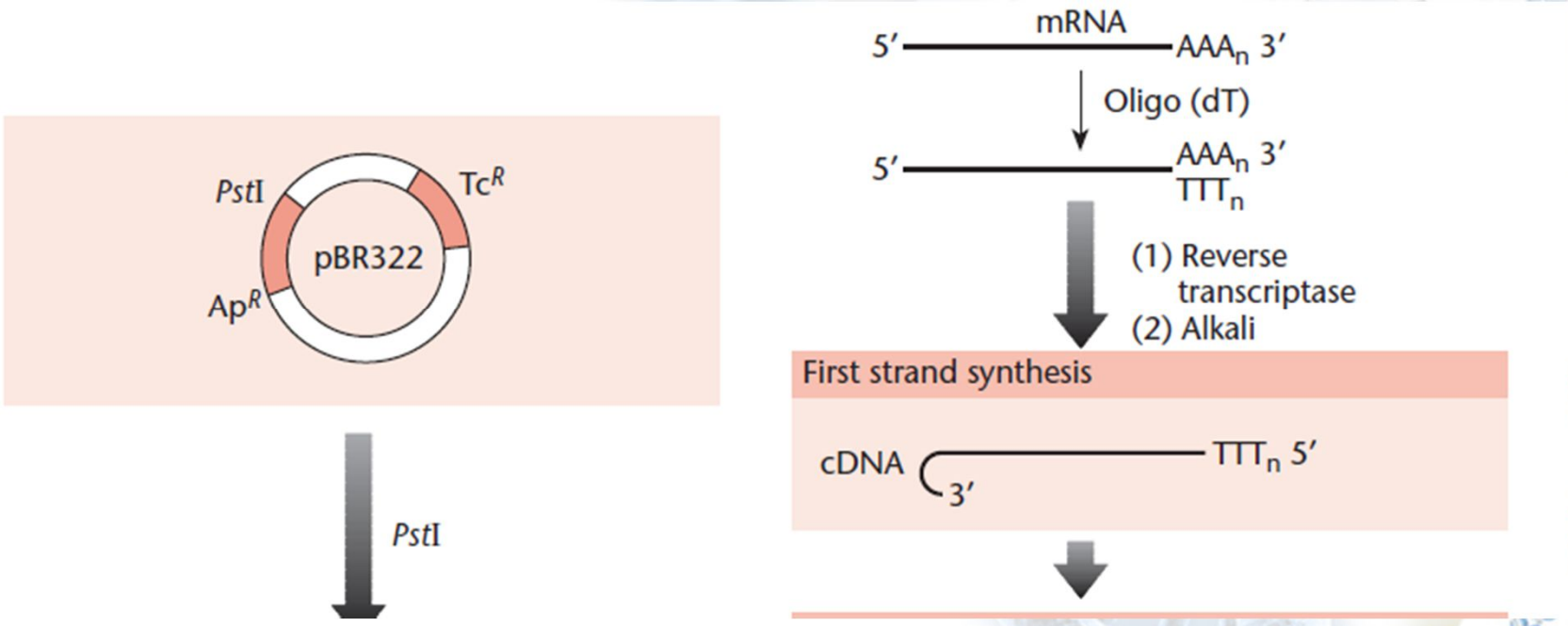


cDNA cloning

Preparation of cDNA for library construction

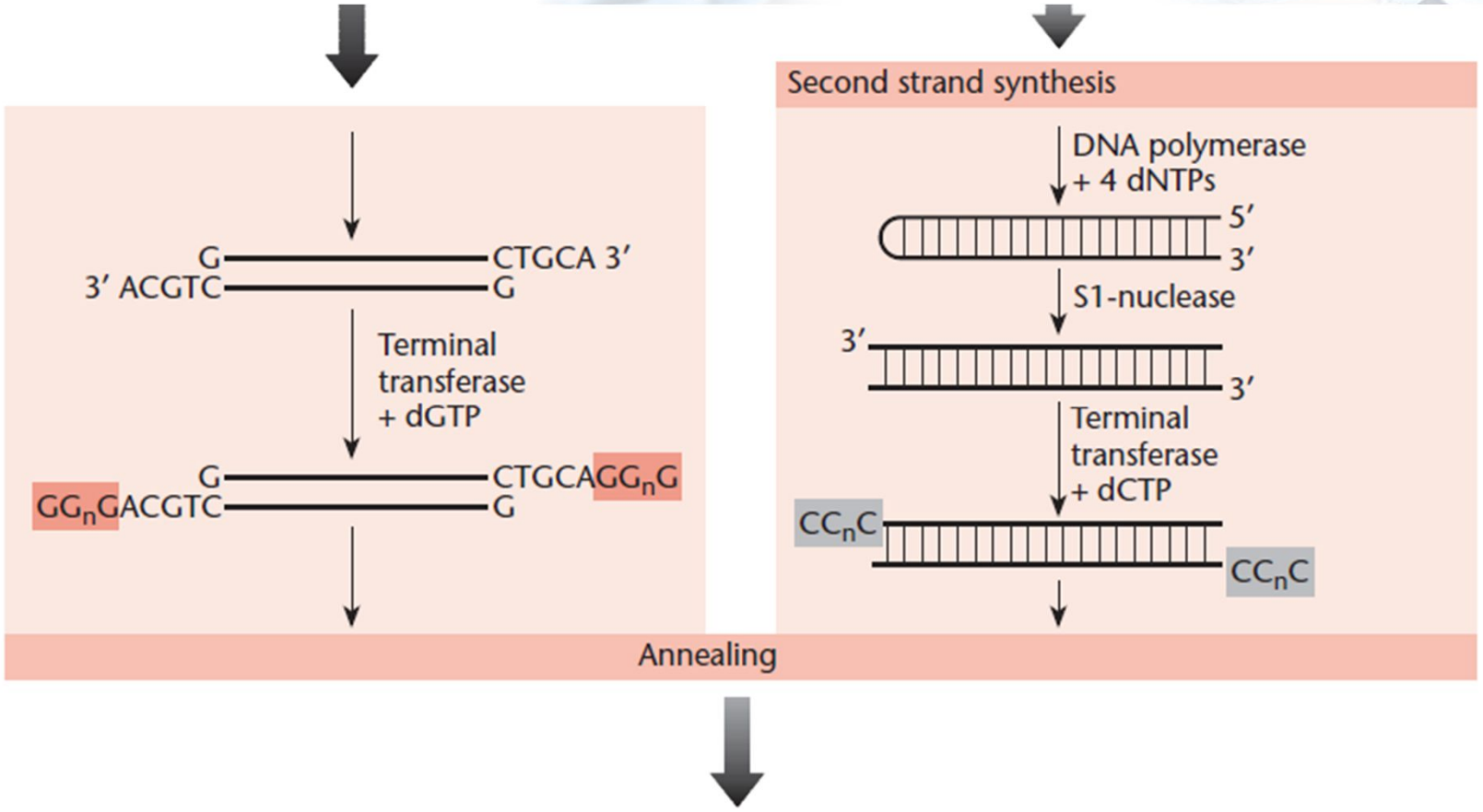
- The cDNA synthesis reaction
- Development of cDNA cloning strategies

cDNA cloning



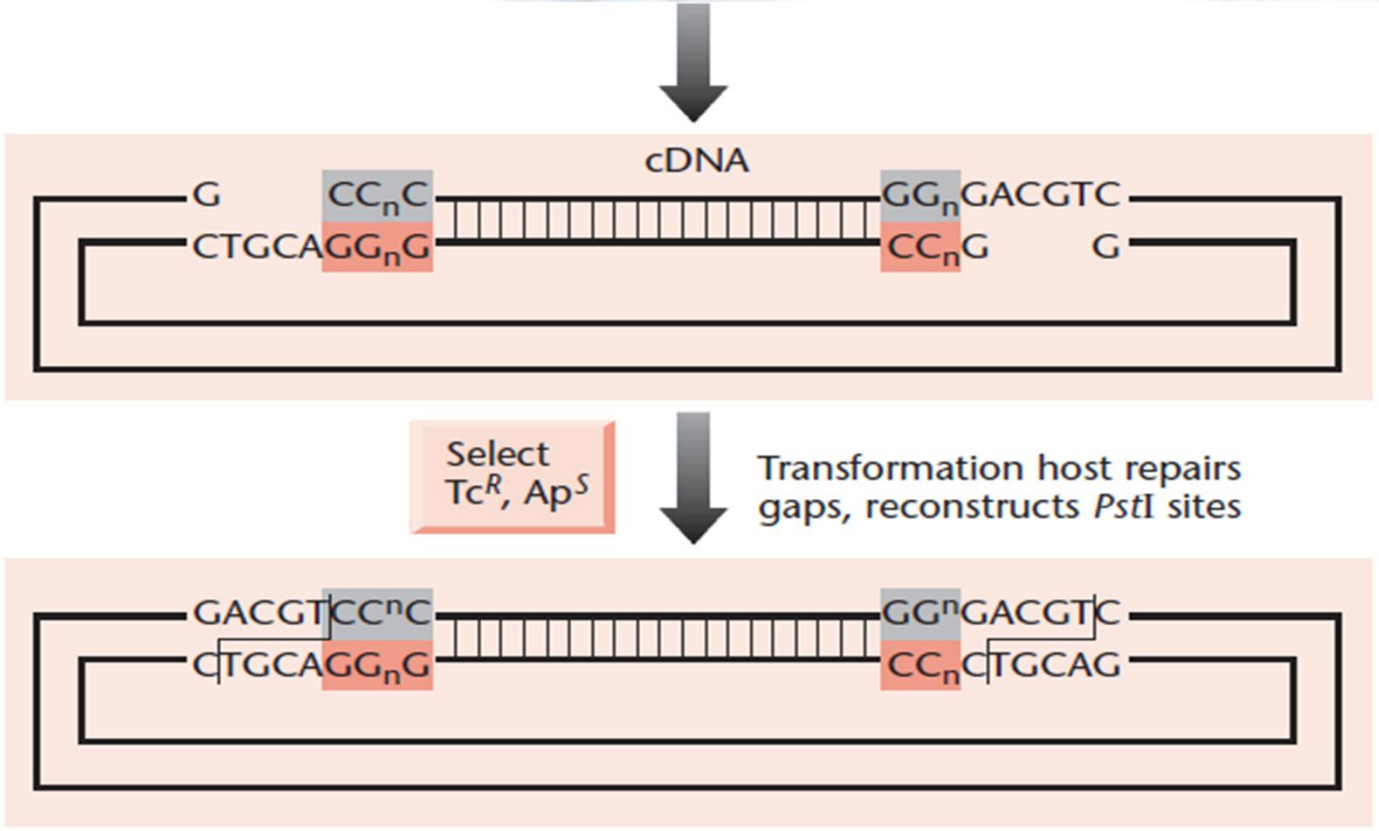
An early cDNA cloning strategy

cDNA cloning



An early cDNA cloning strategy

cDNA cloning



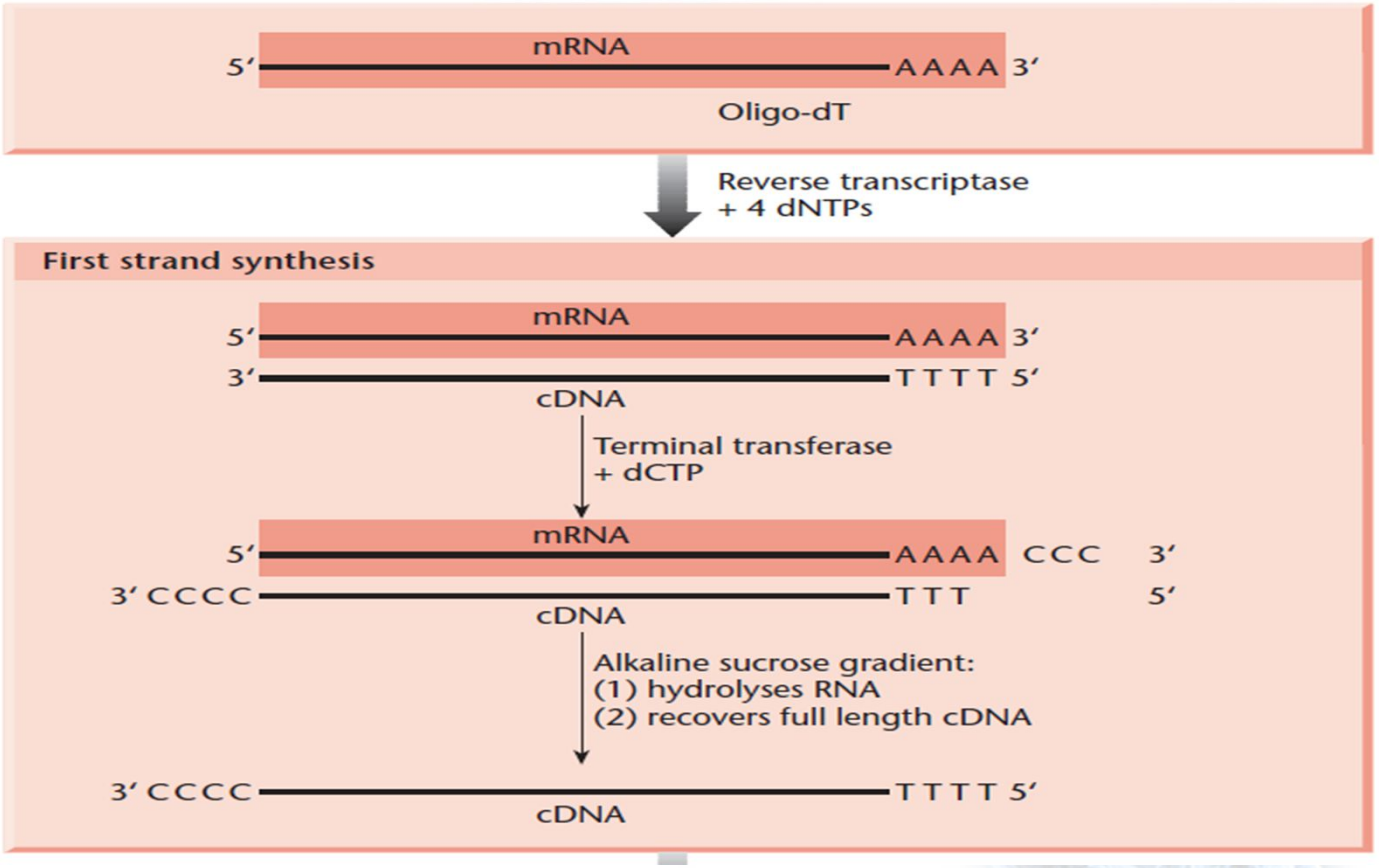
An early cDNA cloning strategy

cDNA cloning

Improved methods for cDNA cloning

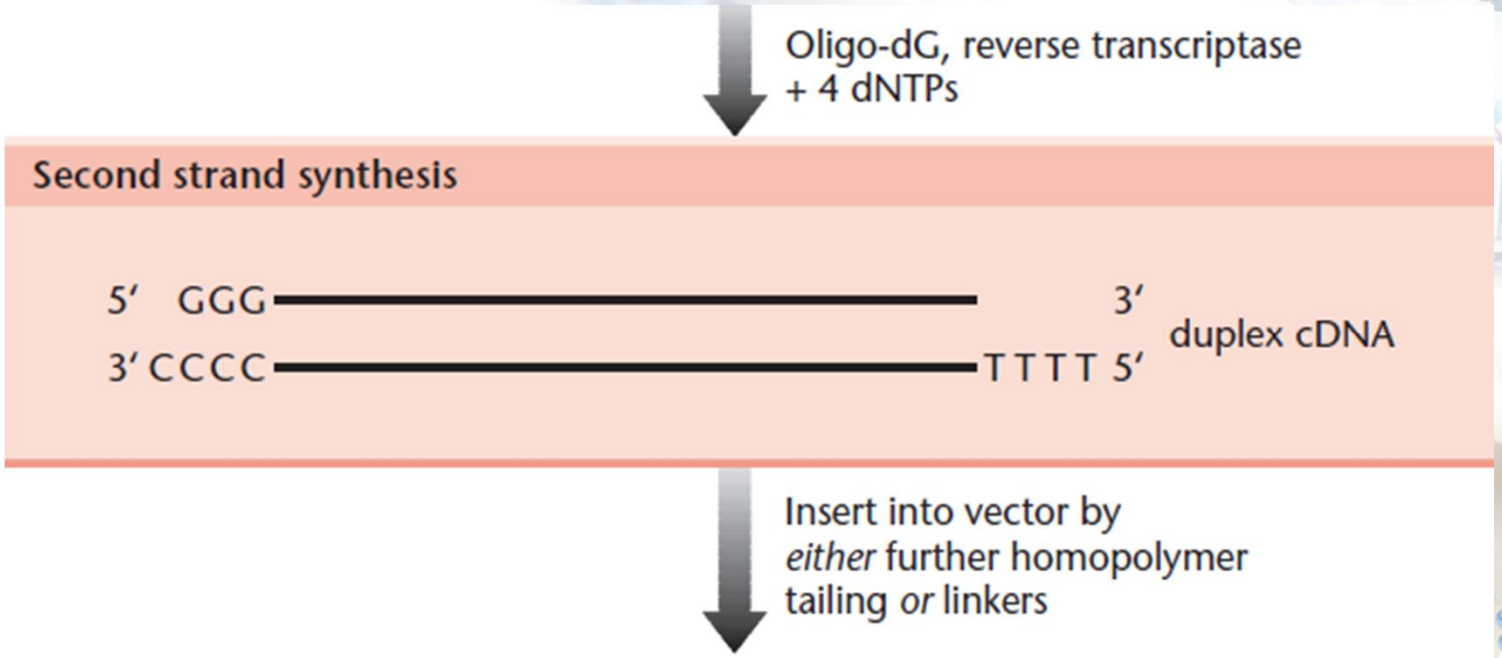
- A serious disadvantage of the hairpin method is that cleavage with S_1 nuclease results in the loss of sequences at the 5′ end of the clone
- This strategy has therefore been superseded with other methods

cDNA cloning



Improved method for cDNA cloning (Land et al. 1981)

cDNA cloning



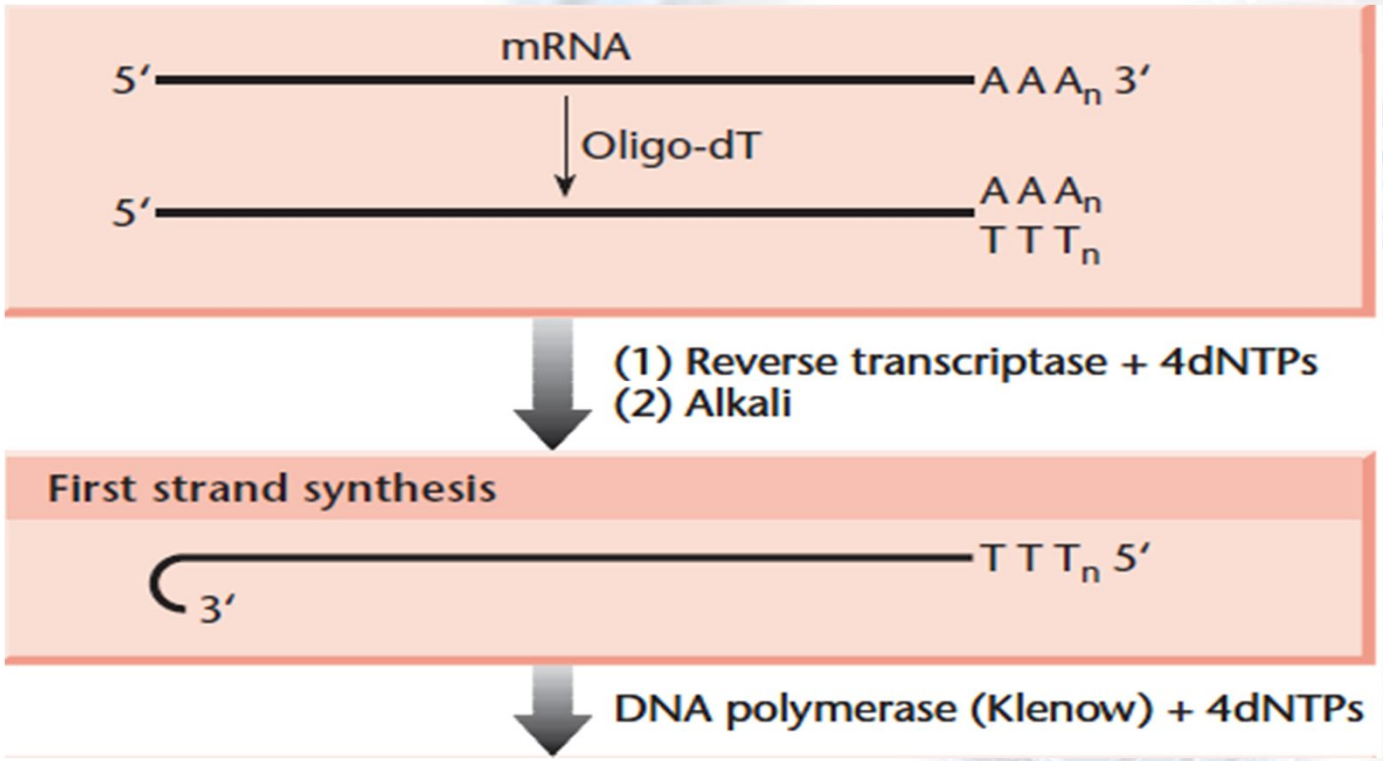
Improved method for cDNA cloning (Land et al. 1981)

cDNA cloning

Improved methods for cDNA cloning

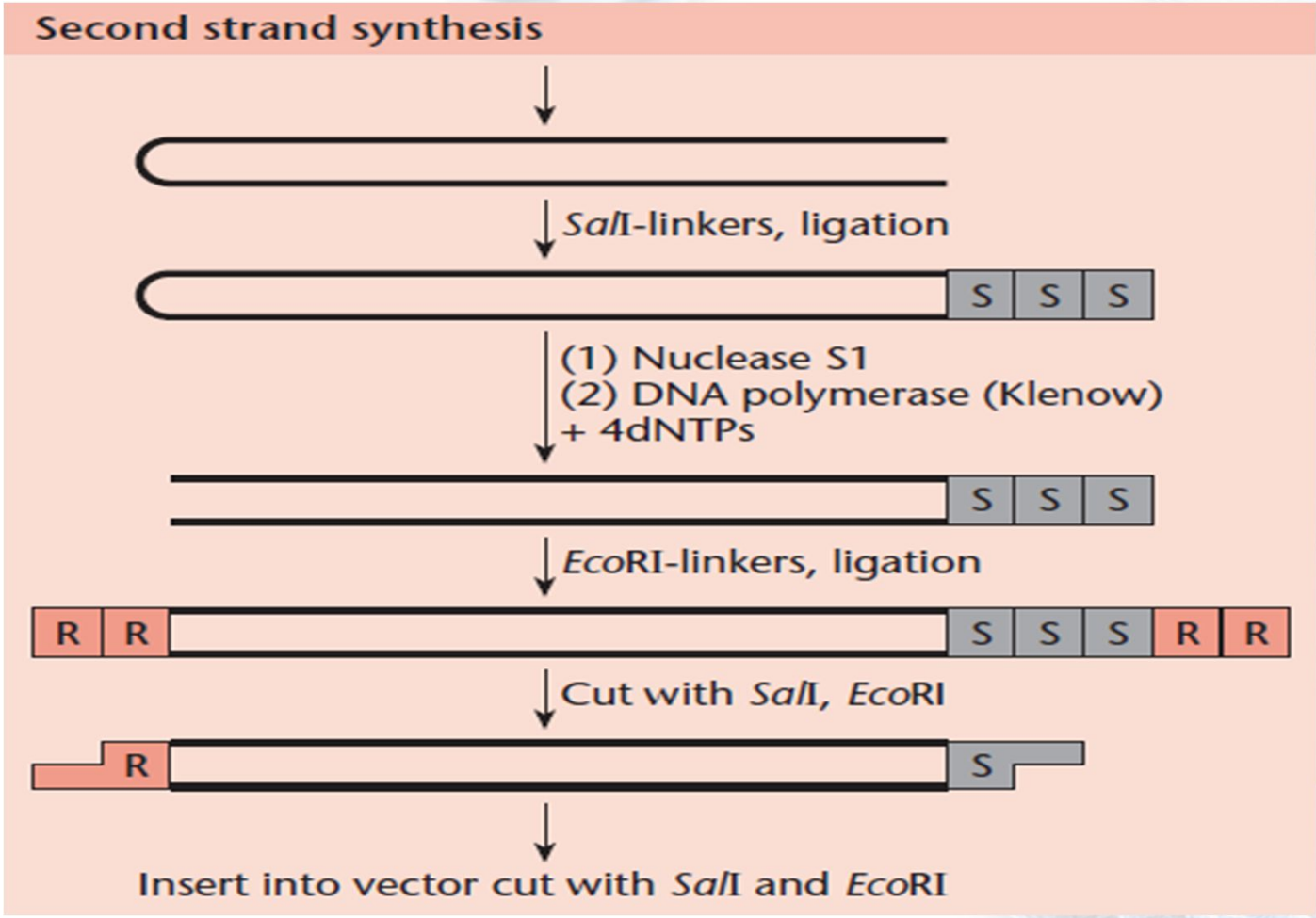
- For cDNA expression libraries, it is advantageous if the cDNA can be inserted into the vector in the correct orientation
- This can be achieved by self-priming method by adding linker molecule to double stranded cDNA

cDNA cloning



Method for direction cDNA cloning

cDNA cloning



Method for direction cDNA cloning

cDNA cloning

PCR as an alternative for cDNA cloning

- Reverse transcription followed by the PCR (RT-PCR) leads to the amplification of RNA sequences in cDNA form
- RT-PCR is a sensitive means for detecting, quantifying and cloning specific cDNA molecules

Screening strategies

- **Major screening strategies involve**

1). Genetic methods

2). Sequence-dependent screening

3). Screening expression libraries

Genetic methods

Selection for the presence of vector

- All useful vector molecules carry a selectable genetic marker or property
- Plasmid and cosmid vectors carry drug resistance or nutritional marker
- In phage vectors, plaque formation is itself the selected property

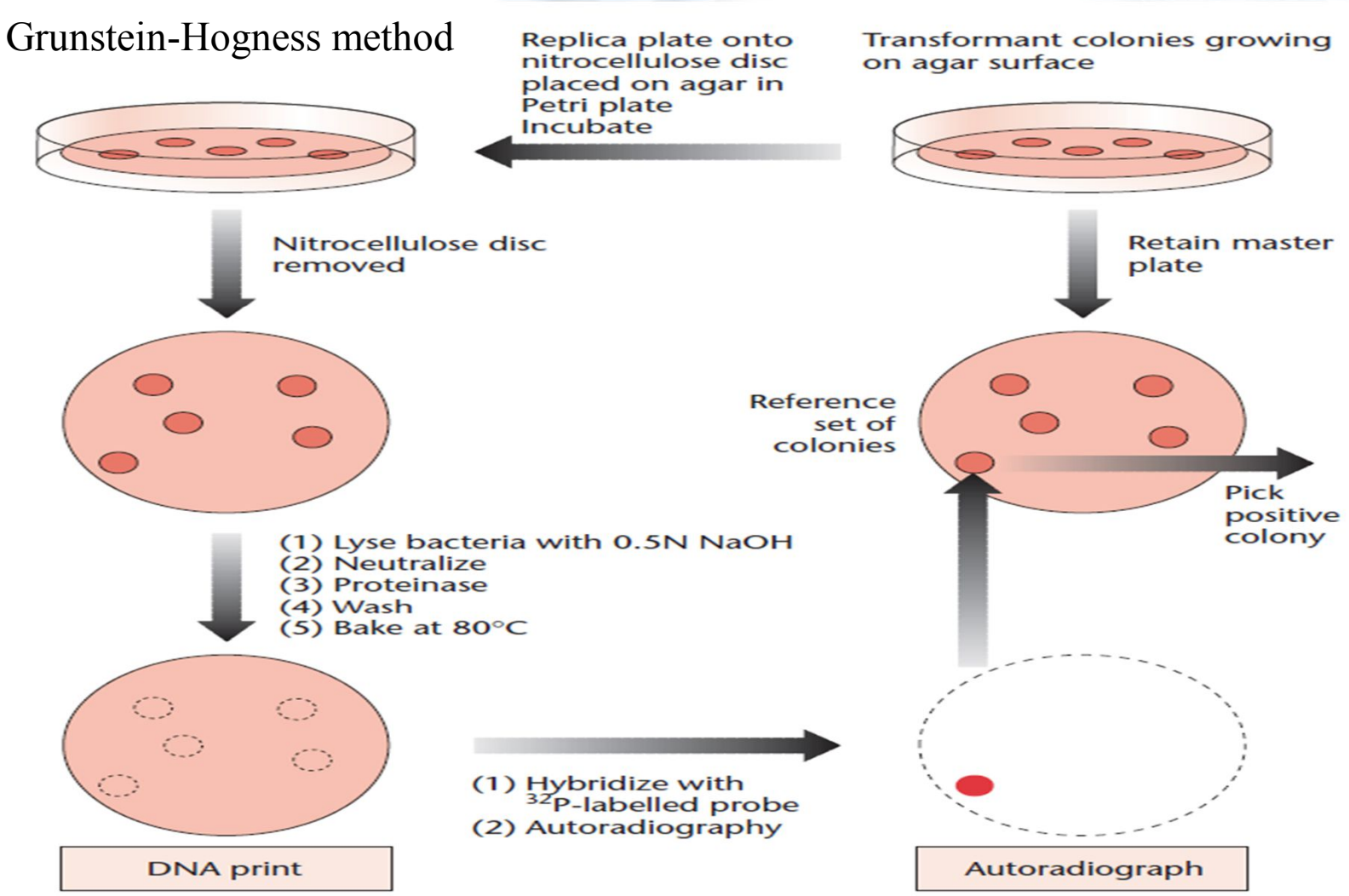
Sequence-dependent screening

Screening by hybridization

- Nucleic acid hybridization is the most commonly used method of library screening
- Grunstein and Hogness (1975) developed a screening procedure to detect DNA sequences in transformed colonies by *in situ* hybridization

Sequence-dependent screening

Grunstein-Hogness method



Sequence-dependent screening

Screening by hybridization

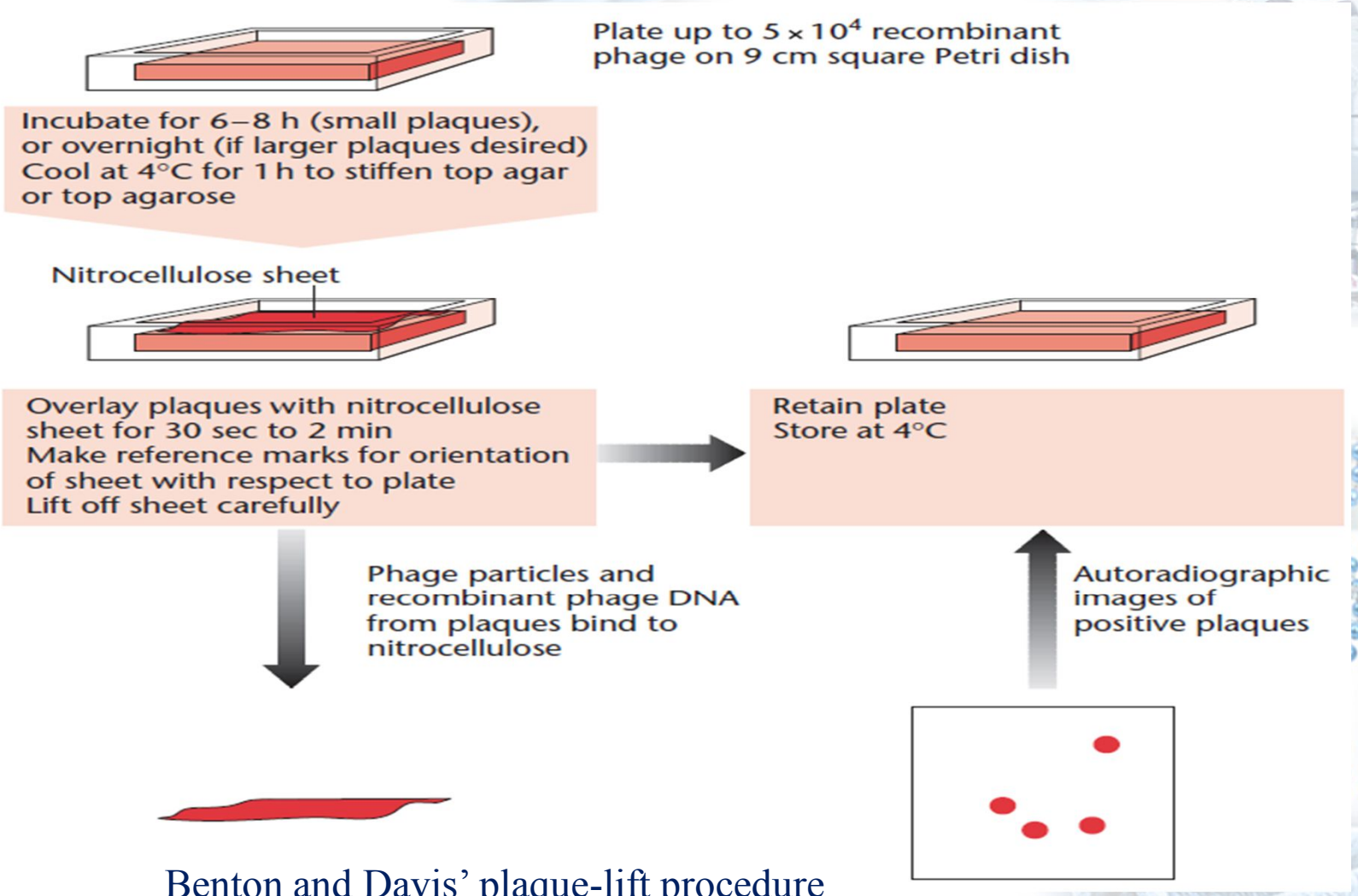
- The results of the hybridization can be monitored by autoradiography

Sequence-dependent screening

Benton and Davis' plaque lift procedure

- **Benton and Davis (1977) devised a method called plaque lift, in which nitrocellulose filter is applied to the upper surface of agar plates, making direct contact between plaques and filter**

Sequence-dependent screening

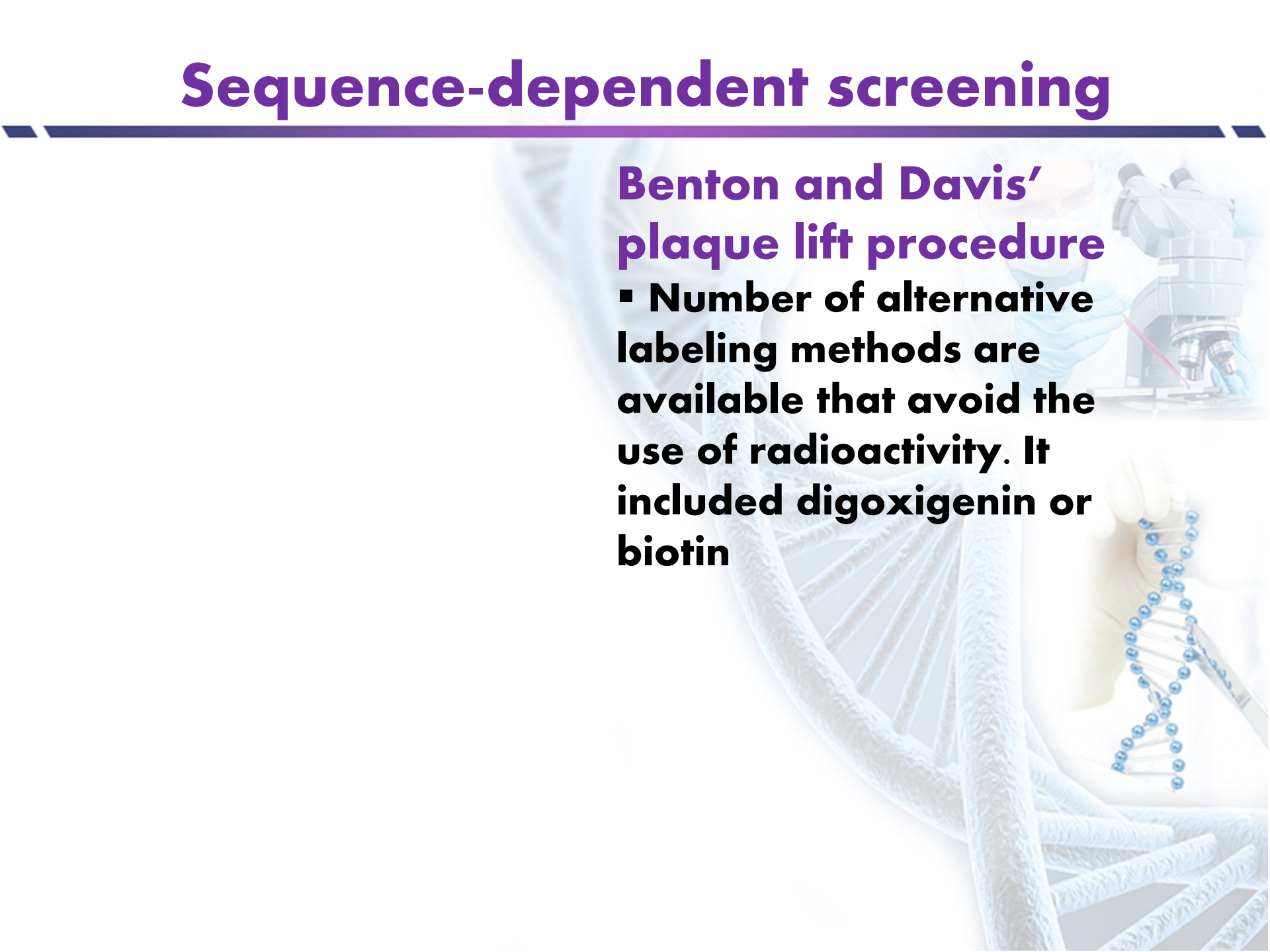


Benton and Davis' plaque-lift procedure

Sequence-dependent screening

Benton and Davis' plaque lift procedure

- **Number of alternative labeling methods are available that avoid the use of radioactivity. It included digoxigenin or biotin**



Sequence-dependent screening

Probe design

- A great advantage of hybridization for library screening is that it is extremely versatile
- Conditions can be used in which hybridization is very stringent, so that only sequences identical to the probe are identified

Sequence-dependent screening

Chromosome walking

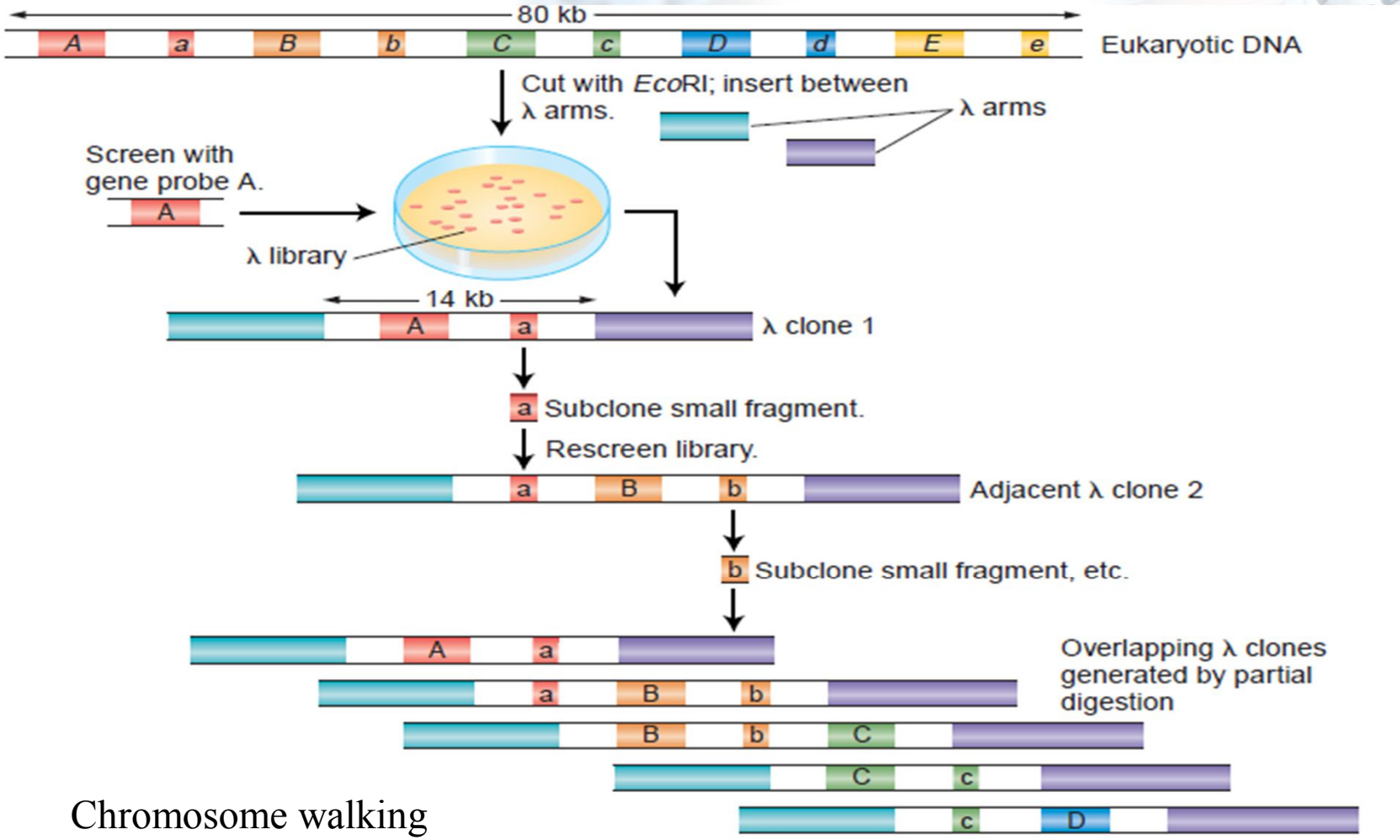
- Walking along the chromosome is a term used to describe an approach which allows the isolation of gene sequences whose function is quite unknown but whose genetic location is known
- A cloned genomic fragment must be found as a starting point for the walk

Sequence-dependent screening

Chromosome walking

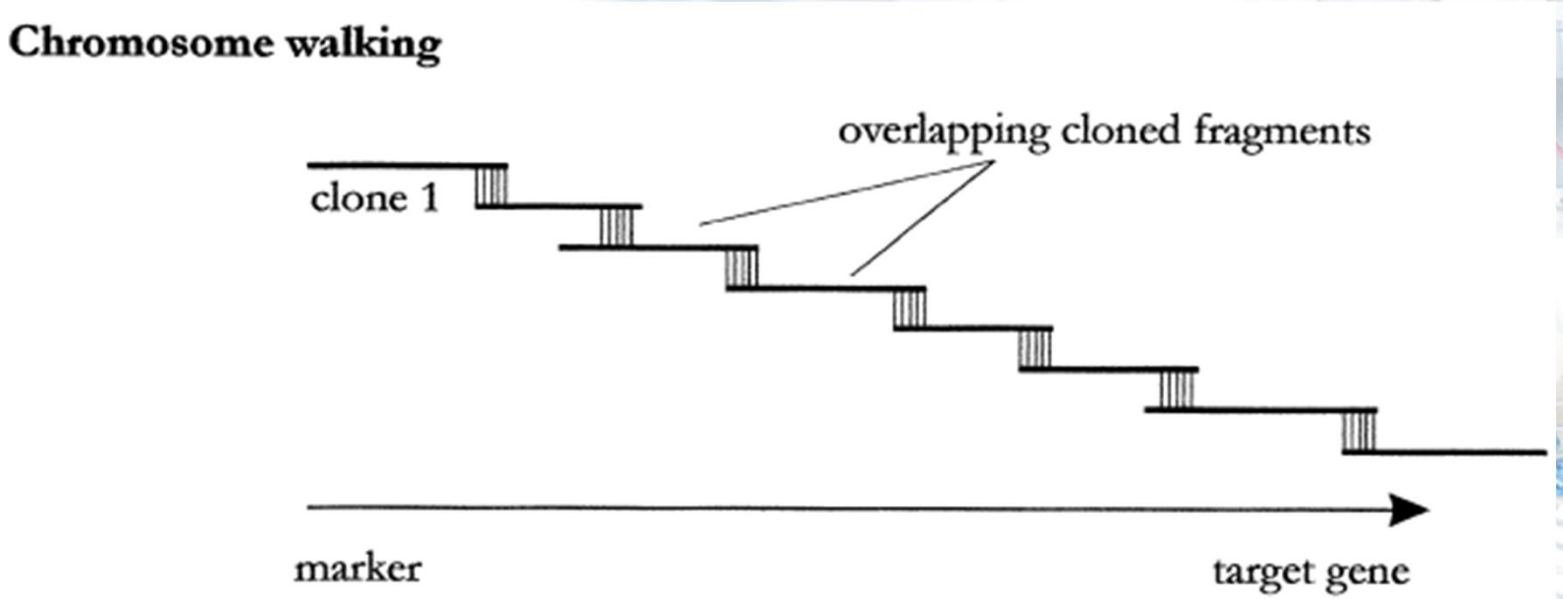
- A cloned genomic fragment must be found as a starting point for the walk
- In human genome, the starting point may be a restriction fragment length polymorphism (RFLP) sequence that is closely linked to disease locus

Sequence-dependent screening



Chromosome walking

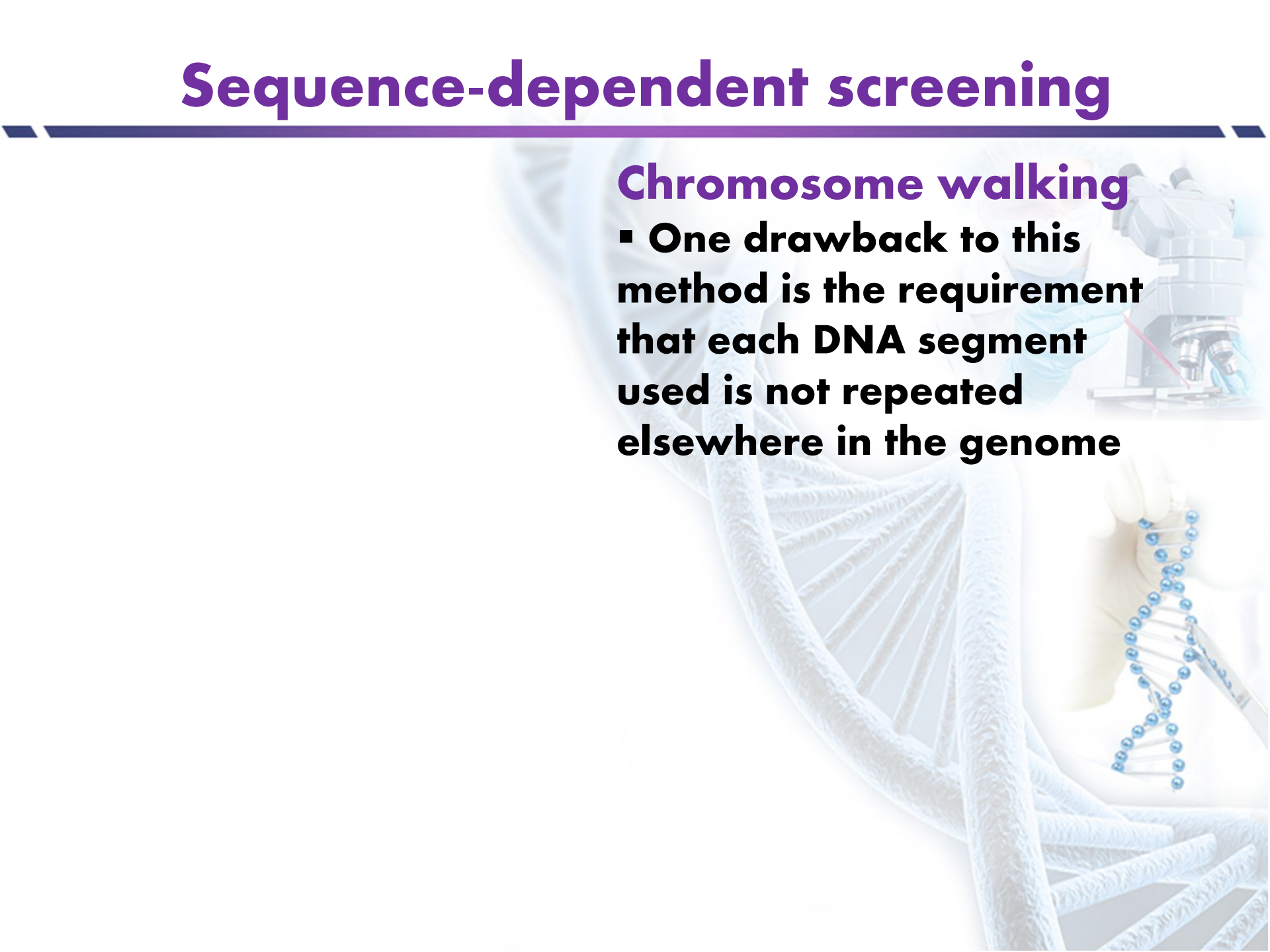
Sequence-dependent screening



Sequence-dependent screening

Chromosome walking

- One drawback to this method is the requirement that each DNA segment used is not repeated elsewhere in the genome



Sequence-dependent screening

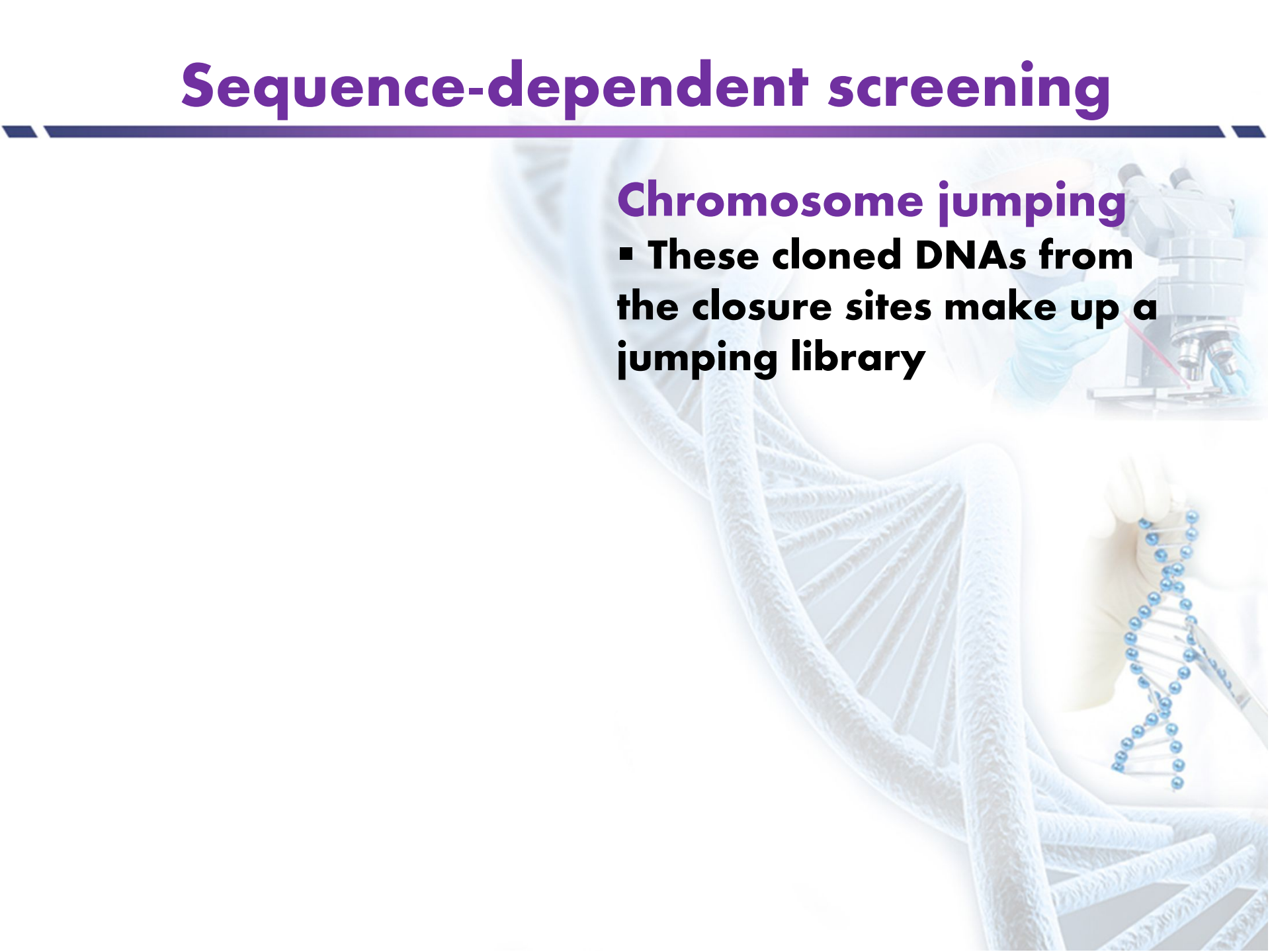
Chromosome jumping

- In chromosome jumping, the DNA of interest is identified, cut into fragments with restriction enzymes and circularized
- It brings together DNA sequence that were originally located a considerable distance apart in the genome

Sequence-dependent screening

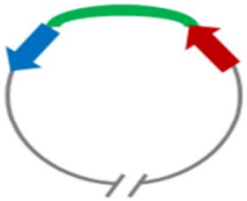
Chromosome jumping

- These cloned DNAs from the closure sites make up a jumping library



Screening strategies

CHROMOSOME



Chromosome jumping involving a series of biochemical manipulations

JUMPING CLONE

Sequence-dependent screening

Chromosome jumping

- One of the application of chromosome walking and jumping is in the cloning of human cystic fibrosis gene

Sequence-dependent screening

Screening by PCR

- The PCR is widely used to isolate specific DNA sequences from uncloned genomic DNA or cDNA but it also a useful technique for library screening
- To isolate specific clone, PCR is carried out with gene-specific primers that flank a unique sequence in the target

Sequence-dependent screening

Screening by PCR

- There are several applications where the use of degenerate primes is favorable
- A degenerate primer is a mixture of primers, all of similar sequences but with variations at one or more positions

Screening expression libraries

Expression cloning

- If DNA library is established using expression vectors, each individual clone can be expressed to yield a polypeptide
- Expression libraries are useful because they allow a range of alternative techniques to be employed

Screening expression libraries

Immunochemical screening

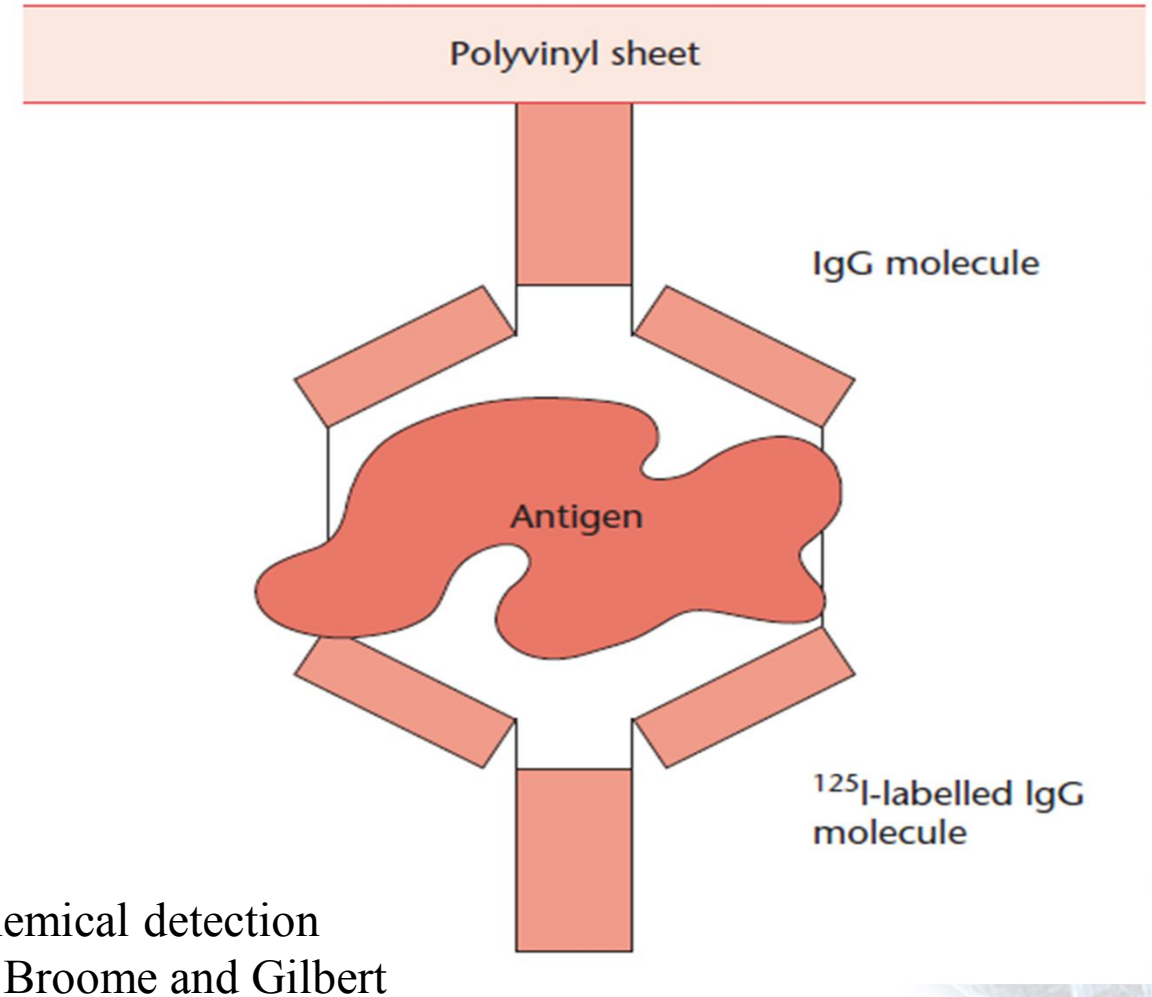
- **Immunochemical screening involves the use of antibodies that specifically recognize antigenic determinants on the polypeptides synthesized by a target clone**
- **The molecular target for recognition is generally an epitope**

Screening expression libraries

Immunochemical screening

- The method of Broome and Gilbert (1978) exploited the fact that antibodies adsorb very strongly to certain types of plastic such as polyvinyl and that IgG antibodies can be readily labelled with ^{125}I by iodination *in vitro*

Screening expression libraries



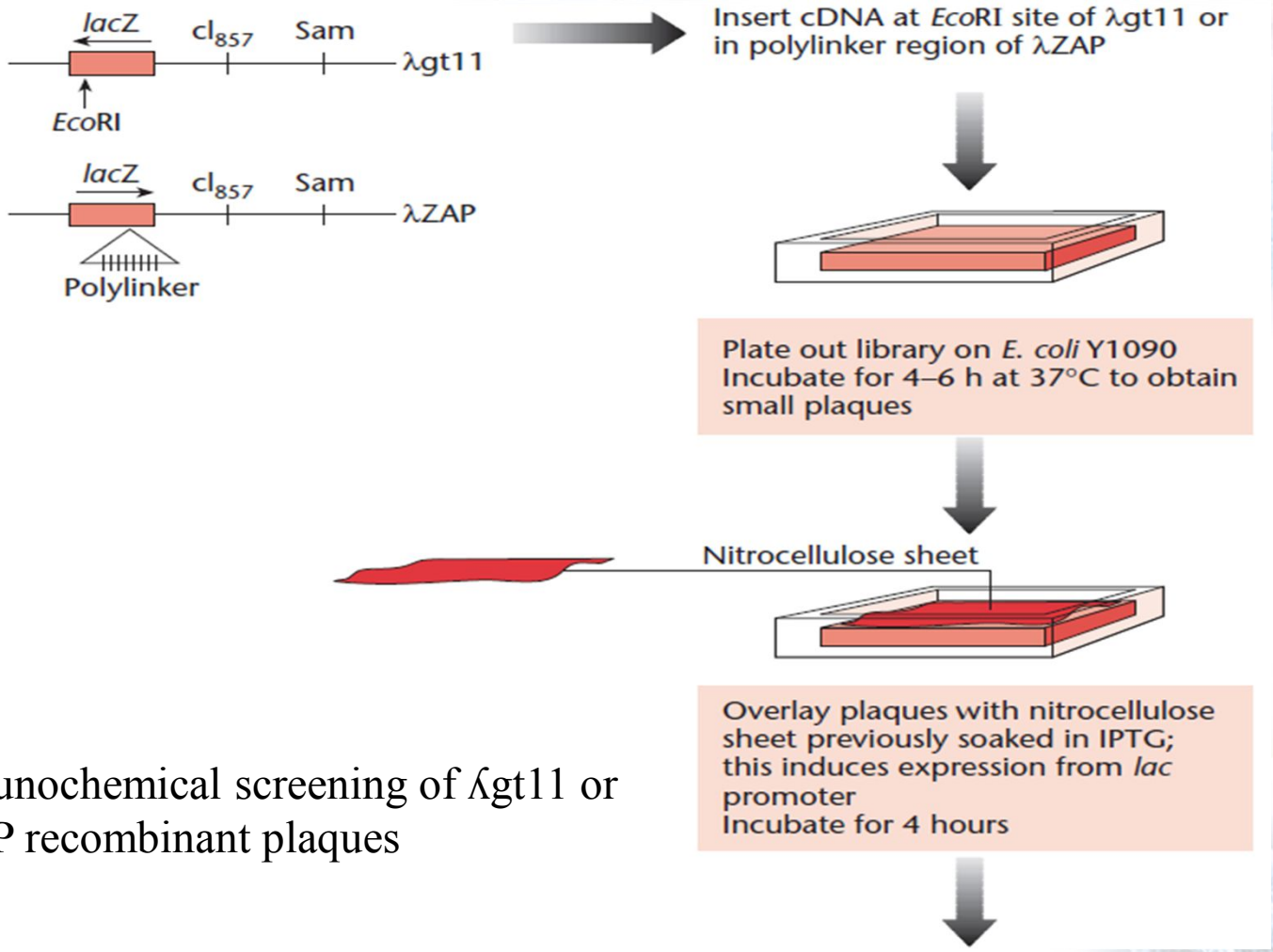
Immunochemical detection method of Broome and Gilbert

Screening expression libraries

Immunochemical screening of λgt11

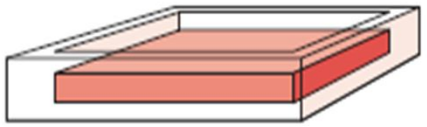
- It is much more convenient to use bacteriophage-λ insertion vectors, because these have a higher capacity and the efficiency of *in vitro* packaging allows large numbers of recombinants to be prepared and screened

Screening expression libraries



Immunochemical screening of ʎgt11 or ʎZAP recombinant plaques

Screening expression libraries



Retain plate
Pick positive plaque from retained plaque

Carefully remove nitrocellulose sheet, which will have adsorbed fusion proteins expressed in recombinant phage plaques, i.e. plaque-lift

Screen nitrocellulose plaque-lift with specific antibody to detect fusion protein
Identify positive plaque

Immunochemical screening of λgt11 or λZAP recombinant plaques

Screening expression libraries

Immunochemical screening of λgt11

- The original detection method using iodinated antibodies has been superseded by more convenient methods using non-isotopic labels, which are also more sensitive and have a lower background of non-specific signal

Screening expression libraries

South-western and north-western blotting

- It involves the screening and isolation of clones expressing sequence specific DNA or RNA-binding proteins
- Screening is carried out without using an antibody, by incubating the membranes with radiolabelled double stranded DNA probe

Screening expression libraries

Screening by functional complementation

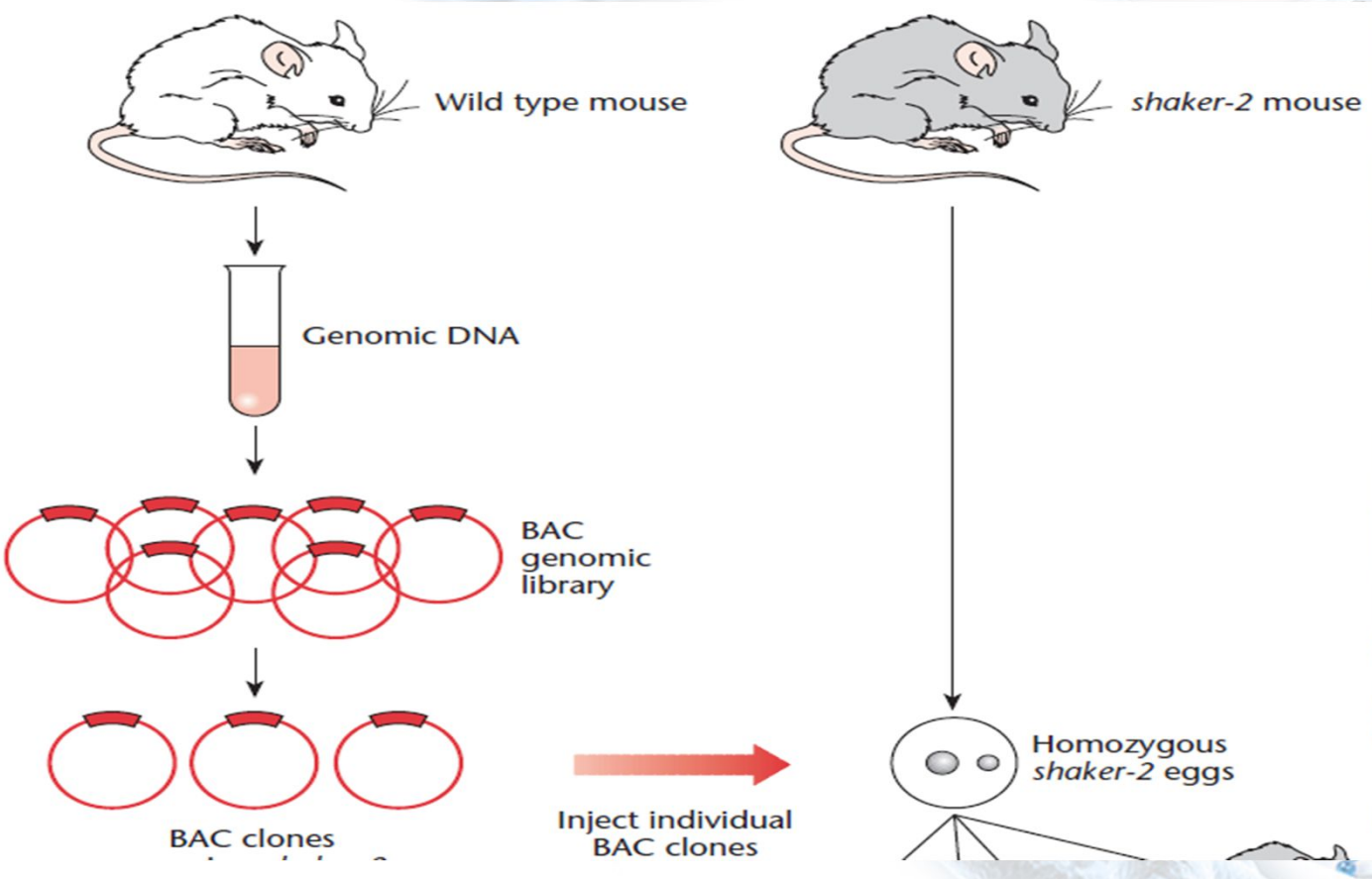
- **Functional complementation is the process by which a particular DNA sequence compensates for a missing function in a mutant cell and thus restores the wild type phenotype**

Screening expression libraries

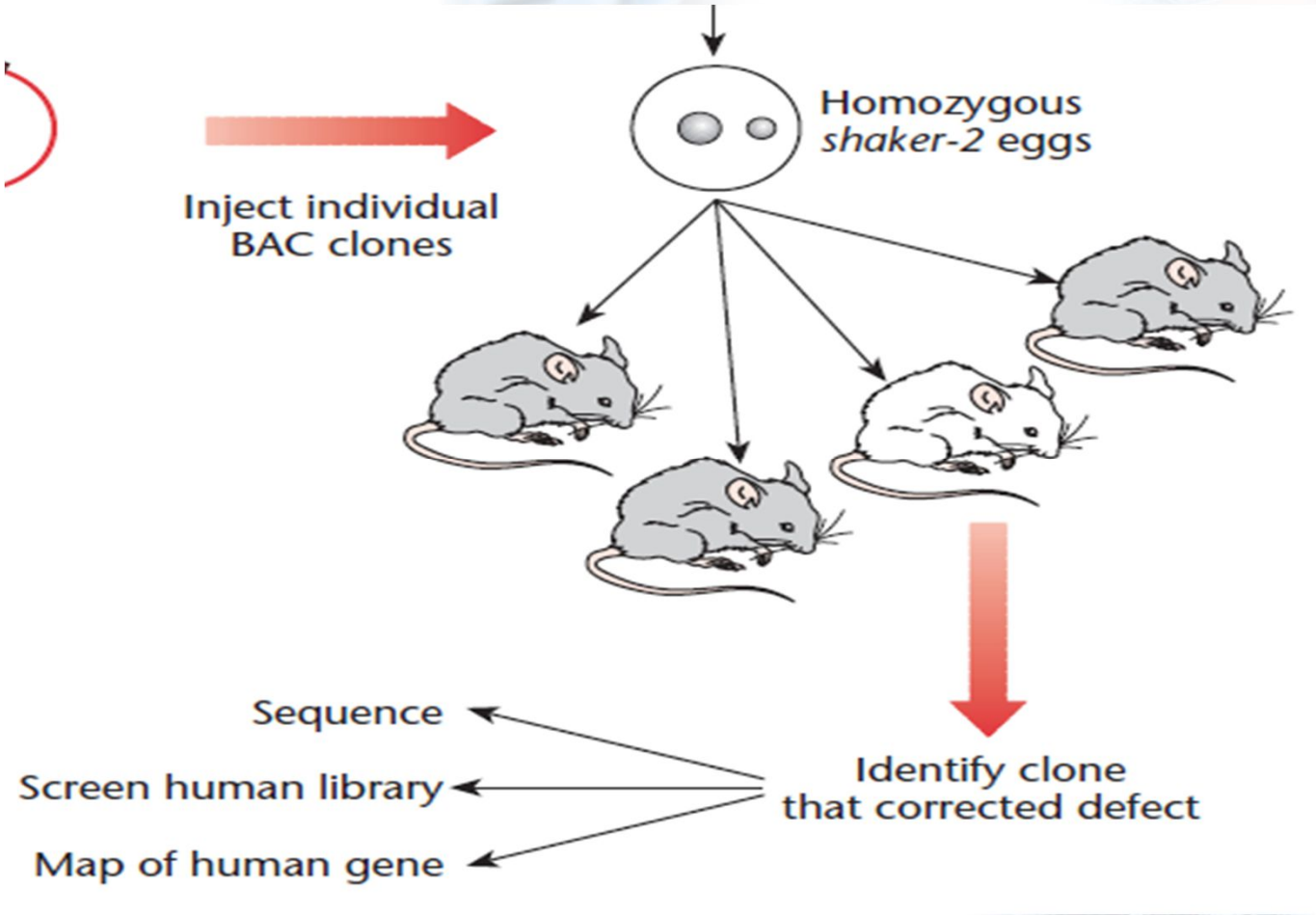
Screening by functional complementation

- Functional complementation is also possible in transgenic animals and plants
- It has been used for complementation in transgenic mice to isolate the *Shaker-2* gene

Screening expression libraries



Screening expression libraries



Screening expression libraries

Screening by functional complementation

- **Functional complementation is the process by which a particular DNA sequence compensates for a missing function in a mutant cell and thus restores the wild type phenotype**

Expression in *E. coli* of cloned DNA

Requirement for expression in *E. coli*

- **Synthesis of functional protein depends upon transcription of appropriate gene, efficient translation of mRNA and in many cases, post-translational processing and compartmentalization of nascent polypeptide**

Expression in *E. coli* of cloned DNA

Secretion of proteins

- **Gram-negative bacteria such as *E. coli* have a complex wall-membrane structure comprising an inner, cytoplasmic membrane separated from an outer membrane by a cell wall and periplasmic space**

Expression in *E. coli* of cloned DNA

Protein trafficking

- The bacterial inner membrane, periplasmic space and outer membrane all contain proteins not found in the cytoplasm
- All of these proteins are first synthesized in the cytoplasm but somehow find their way out

Expression in *E. coli* of cloned DNA

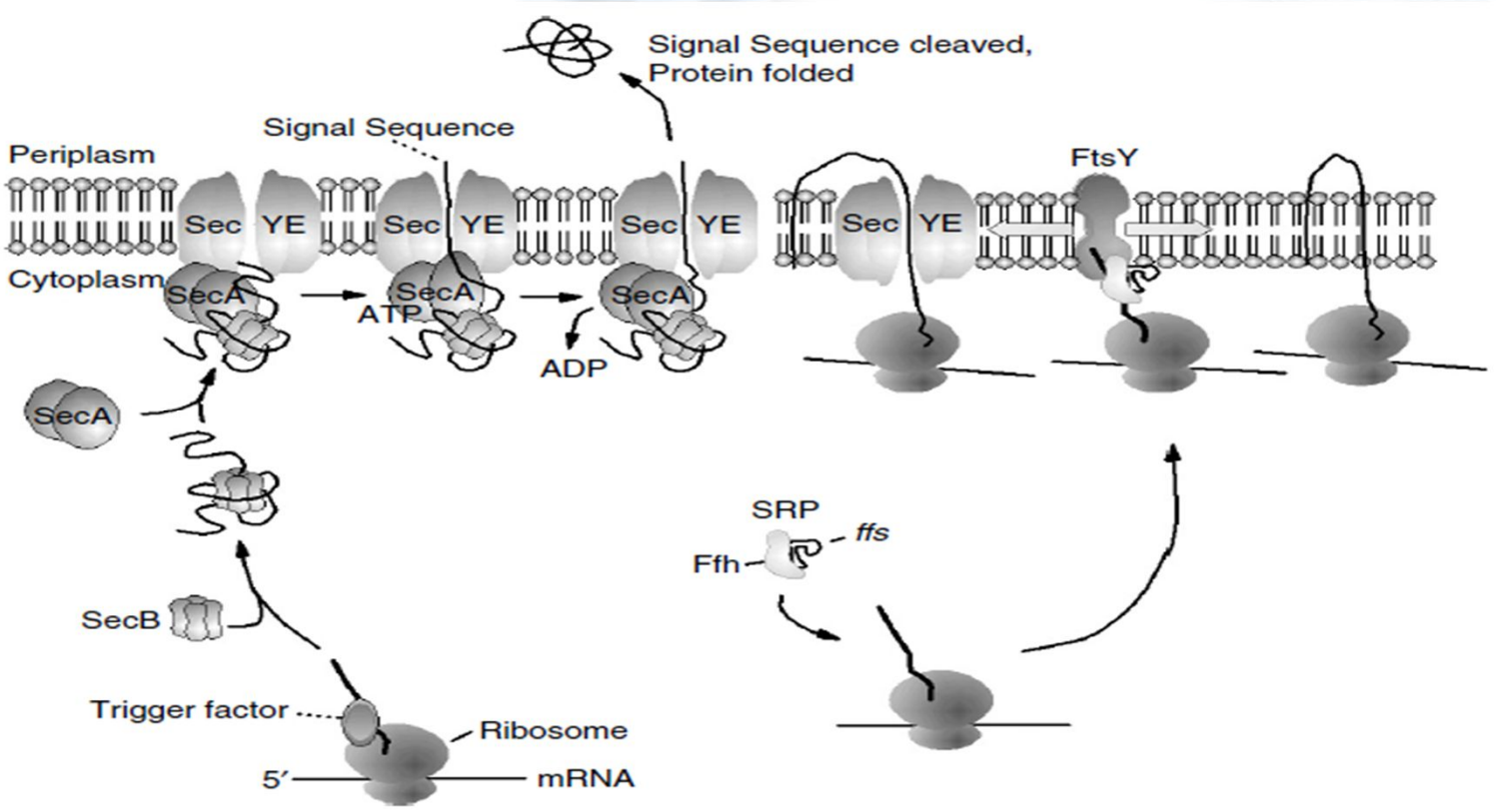


Figure. General secretion pathway

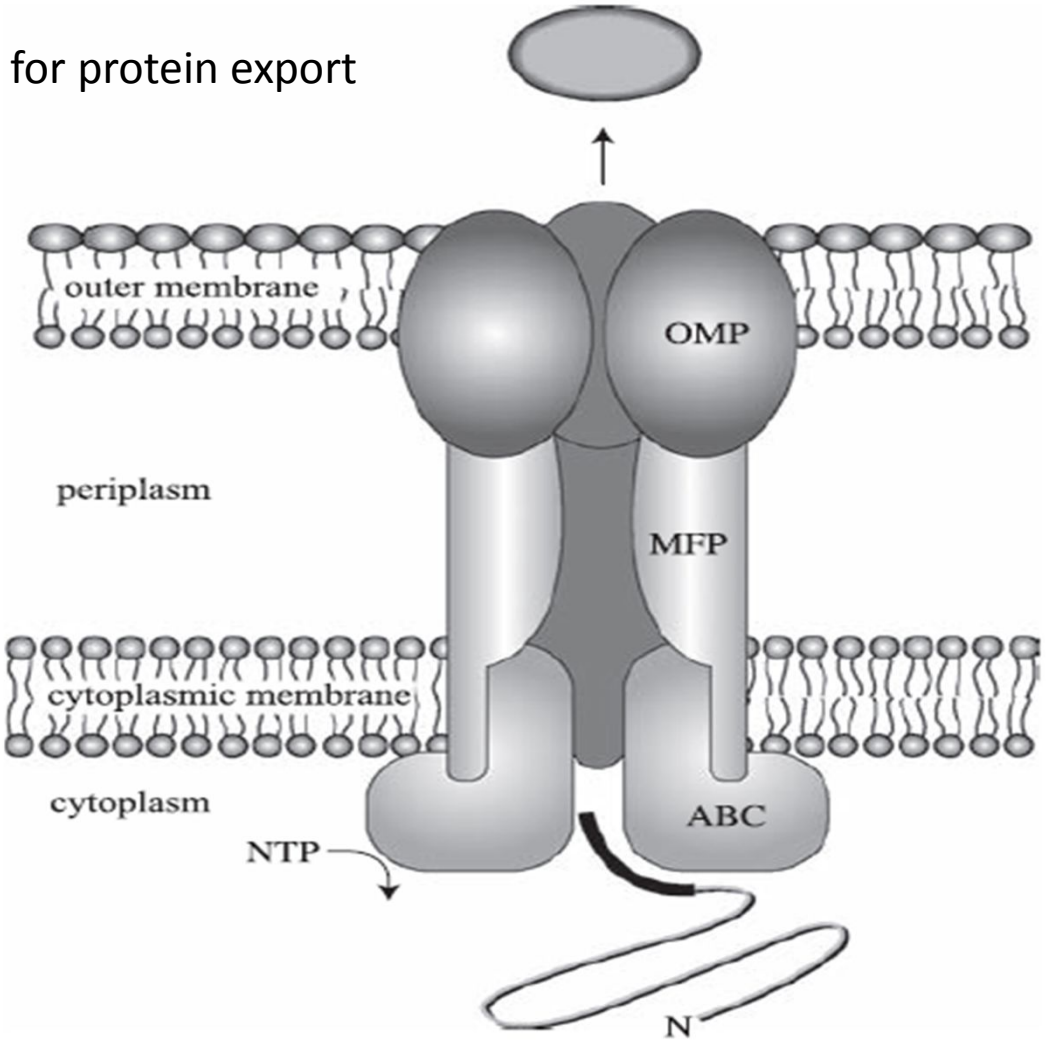
Expression in *E. coli* of cloned DNA

Protein trafficking

- There are actually several general mechanisms of protein export in bacteria

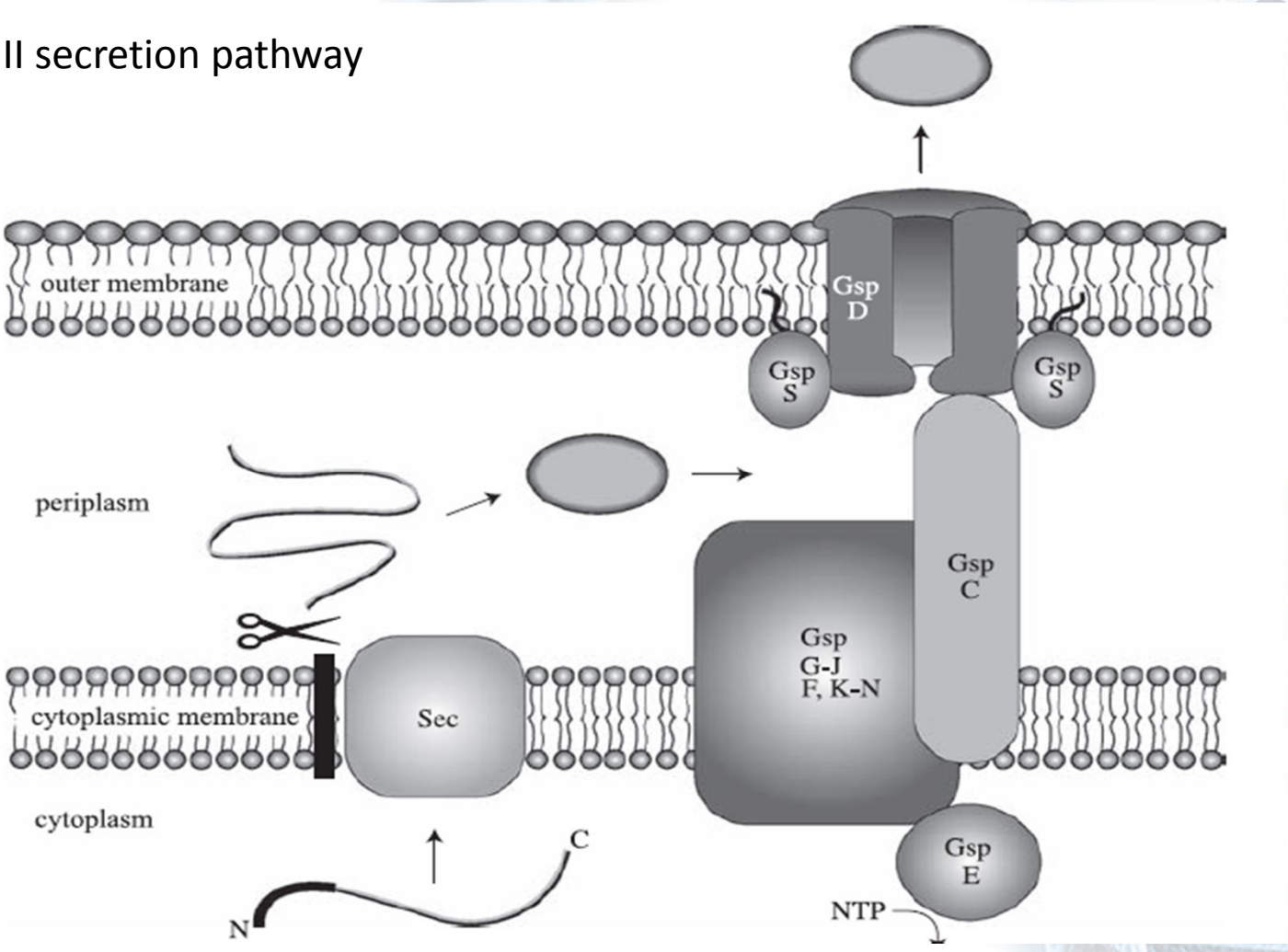
Expression in *E. coli* of cloned DNA

ABC pathway for protein export



Expression in *E. coli* of cloned DNA

Type II secretion pathway



Expression in *E. coli* of cloned DNA

Protein trafficking

- Each system transport a particular type of proteins

Expression in *E. coli* of cloned DNA

Stability of foreign proteins in *E. coli*

- **Various strategies have been developed to cope with the instability of foreign proteins in *E. coli***
- **In the case of somatostatin, degradation was prevented by producing a fused protein consisting of somatostatin and β-galactosidase**

Maximizing the expression of cloned gene

Constructing the optimal promoter

- Large number of promoters for *E. coli* have been analysed
- Many promoters has led to the formulation of a consensus sequence which consists of the -35 region (5′-TTGACA-) and -10 region or Pribnow box (5′-TATAAT-)

Maximizing the expression of cloned gene

	-35 Region		-10 Region	
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17		
CONSENSUS	• • •	TTGACA • • • • • • • • • • • • • • • • •	TATAAT	• •
lac	GGC	TTTACA CTTTATGCTTCCGGCTCG	TATATTGT	
trp	CTG	TTGACA ATTAATCAT CGAACTAG	TTAACTAG	
λP_L	GTG	TTGACA TAAATACCA CTGGCGG	TGATACTGA	
rec A	CAC	TTGATA CTGTATGAA GCATACAG	TATAATTG	
tacI	CTG	TTGACA ATTAATCAT CGGCTCG	TATAATGT	
tacII	CTG	TTGACA ATTAATCAT CGAACTAG	TTTAATGT	

Figure. The base sequence of the -10 and -35 regions for natural and hybrid promoters

Maximizing the expression of cloned gene

Constructing the optimal promoter

- Expression from a strong promoter can represent 20-40% of cloned gene product of toral cell protein

Maximizing the expression of cloned gene

Optimizing translation initiation

- Complementarity of Shine-Dalgarno (S-D) sequence with 16s rRNA can affect the rate of translation
- Composition of triplet immediately preceding the AUG start codon also affects the efficiency of translation

Maximizing the expression of cloned gene

Stability of mRNA and codon choice

- The rate of synthesis of a particular protein will depend on the steady-state of mRNA in the cell
- Degradation of mRNA usually proceeds by a combination of endonuclease and 3′-exonuclease attack

Maximizing the expression of cloned gene

The effect of plasmid copy number

- The number of ribosomes in a cell far exceeds any one class of mRNA
- One way of increasing the expression of a cloned gene is to increase the number of the corresponding transcript

Maximizing the expression of cloned gene

Plasmid stability

- Having maximized the expression of a particular gene it is important to consider what effects this will have on the bacterium harbouring the recombinant plasmid
- The loss of plasmids due to defective partitioning is called segregative instability

Maximizing the expression of cloned gene

Plasmid stability

- Structural instability of plasmids may arise by deletions or rearrangements of DNA

Maximizing the expression of cloned gene

Host cell physiology can affect the level of expression

- Factors which will be important include the choice of nutrients and environmental parameters

Expression in *E. coli* of cloned DNA

Stability of foreign proteins in *E. coli*

- **Various strategies have been developed to cope with the instability of foreign proteins in *E. coli***
- **In the case of somatostatin, degradation was prevented by producing a fused protein consisting of somatostatin and β-galactosidase**

Maximizing the expression of cloned gene

Constructing the optimal promoter

- Large number of promoters for *E. coli* have been analysed
- Many promoters has led to the formulation of a consensus sequence which consists of the -35 region (5′-TTGACA-) and -10 region or Pribnow box (5′-TATAAT-)

Maximizing the expression of cloned gene

	-35 Region		-10 Region	
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17		
CONSENSUS	• • •	TTGACA • • • • • • • • • • • • • • • • •	TATAAT	• •
lac	GGC	TTTACA CTTTATGCTTCCGGCTCG	TATATTGT	
trp	CTG	TTGACA ATTAATCAT CGAACTAG	TTAACTAG	
λP_L	GTG	TTGACA TAAATACCA CTGGCGG	TGATACTGA	
rec A	CAC	TTGATA CTGTATGAA GCATACAG	TATAATTG	
tacI	CTG	TTGACA ATTAATCAT CGGCTCG	TATAATGT	
tacII	CTG	TTGACA ATTAATCAT CGAACTAG	TTTAATGT	

Figure. The base sequence of the -10 and -35 regions for natural and hybrid promoters

Maximizing the expression of cloned gene

Constructing the optimal promoter

- Expression from a strong promoter can represent 20-40% of cloned gene product of toral cell protein

Maximizing the expression of cloned gene

Optimizing translation initiation

- Complementarity of Shine-Dalgarno (S-D) sequence with 16s rRNA can affect the rate of translation
- Composition of triplet immediately preceding the AUG start codon also affects the efficiency of translation

Maximizing the expression of cloned gene

Stability of mRNA and codon choice

- The rate of synthesis of a particular protein will depend on the steady-state of mRNA in the cell
- Degradation of mRNA usually proceeds by a combination of endonuclease and 3′-exonuclease attack

Maximizing the expression of cloned gene

Stability of mRNA and codon choice

- The rate of synthesis of a particular protein will depend on the steady-state of mRNA in the cell
- Degradation of mRNA usually proceeds by a combination of endonuclease and 3′-exonuclease attack

Maximizing the expression of cloned gene

The effect of plasmid copy number

- The number of ribosomes in a cell far exceeds any one class of mRNA
- One way of increasing the expression of a cloned gene is to increase the number of the corresponding transcript

Maximizing the expression of cloned gene

Plasmid stability

- Having maximized the expression of a particular gene it is important to consider what effects this will have on the bacterium harbouring the recombinant plasmid
- The loss of plasmids due to defective partitioning is called segregative instability

Maximizing the expression of cloned gene

Structural instability

- Structural instability of plasmids may arise by deletions or rearrangements of DNA

Maximizing the expression of cloned gene

Host cell physiology can affect the level of expression

- Factors which will be important include the choice of nutrients and environmental parameters

Analysing DNA sequences

DNA sequencing: Benefits and Applications

- **DNA sequence information is a prerequisite for planning any substantial manipulation of the DNA**
- **Information is useful to various fields i.e. molecular and evolutionary biology, metagenomics, medicine, forensics etc.**

Analysing DNA sequences

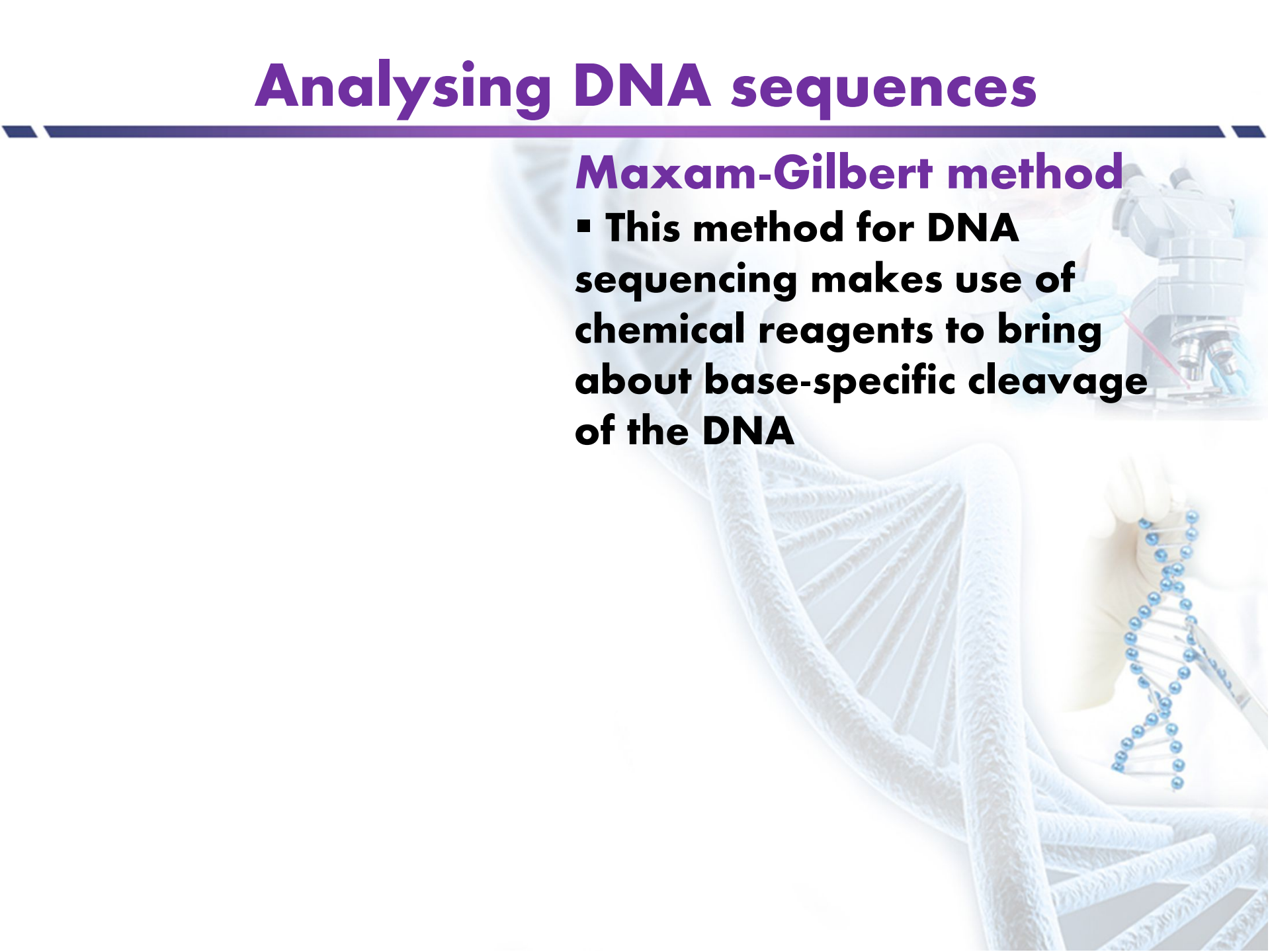
DNA sequencing: Benefits and Applications

- Techniques for DNA sequencing became available in the late 1970s

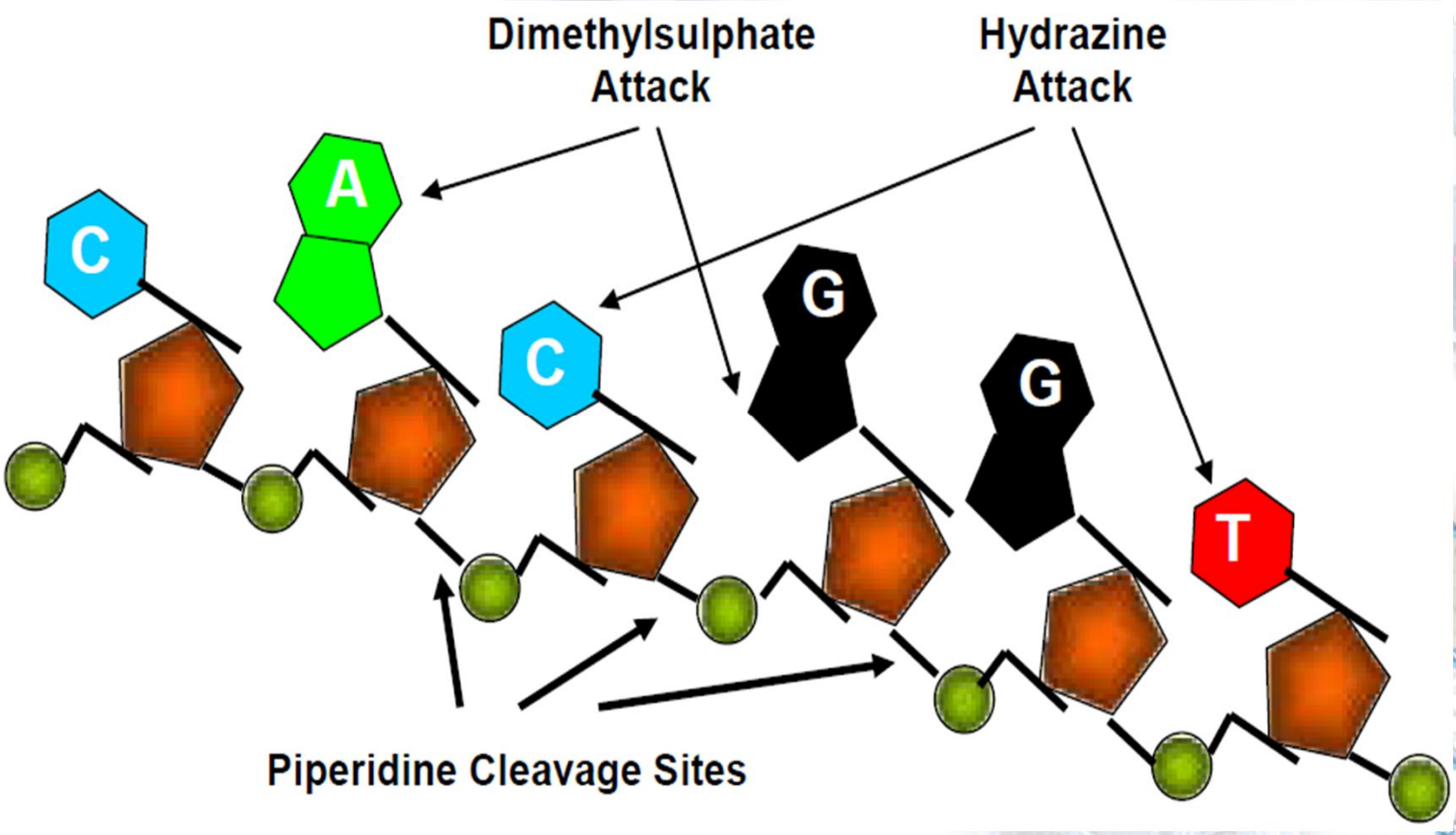
Analysing DNA sequences

Maxam-Gilbert method

- This method for DNA sequencing makes use of chemical reagents to bring about base-specific cleavage of the DNA



Analysing DNA sequences



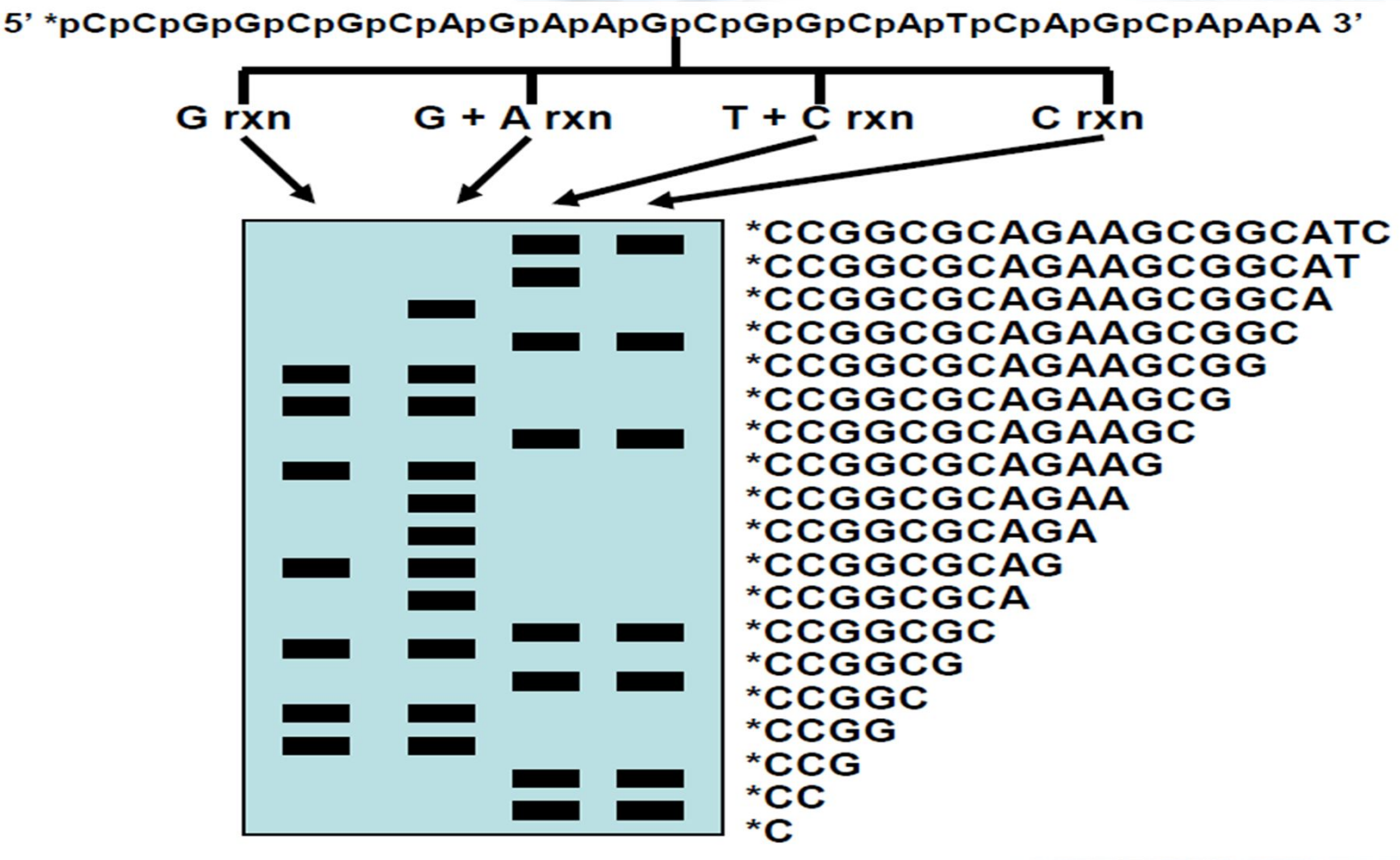
Chemical targets in Maxam and Gilbert DNA sequencing strategy

Analysing DNA sequences

Table. Reagents for Maxam and Gilbert DNA sequencing

Base specificity	Base reaction	Altered base removal	Strand cleavage
G	Dimethylsulphate	Pepridine	Pepridine
G+A	Acid	Acid-catalysed depurination	Pepridine
T+C	Hydrazine	Pepridine	Pepridine
C	Hydrazing+NaCl	Pepridine	Pepridine
A > C	NaOH	Pepridine	Pepridine

Analysing DNA sequences



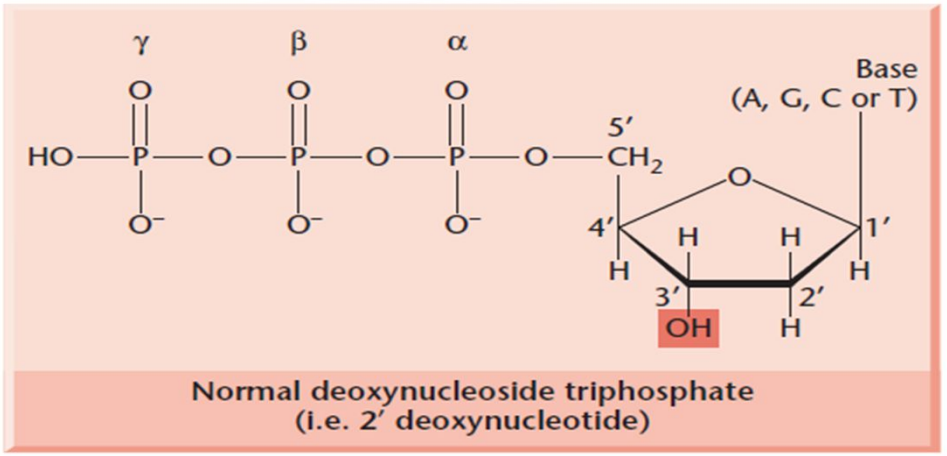
The Maxam and Gilbert mannual sequencing scheme

Analysing DNA sequences

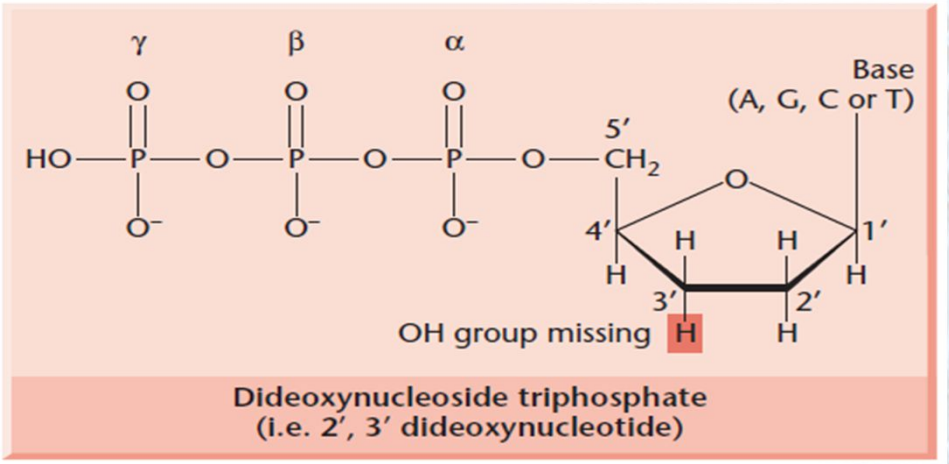
Chain termination or dideoxy procedure

- This method was developed by Sanger et al. (1977)
- It capitalizes on the ability of DNA polymerase to incorporate dideoxynucleoside triphosphates as substrate

Analysing DNA sequences



Normal deoxynucleoside triphosphate
(i.e. 2′ deoxynucleotide)



Dideoxynucleoside triphosphate
(i.e. 2′, 3′ dideoxynucleotide)

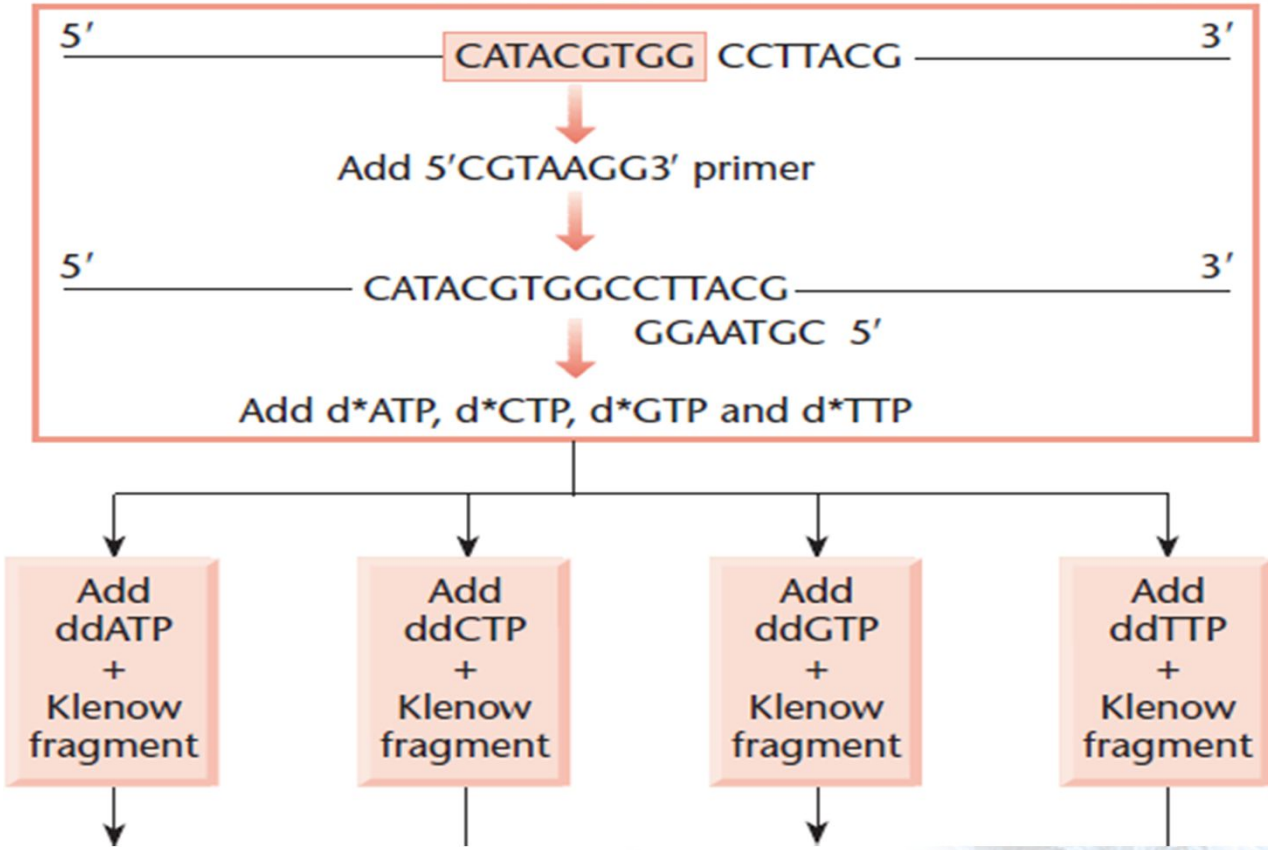
Dideoxynucleoside triphosphate act as chain terminator

Analysing DNA sequences

Chain termination or dideoxy procedure

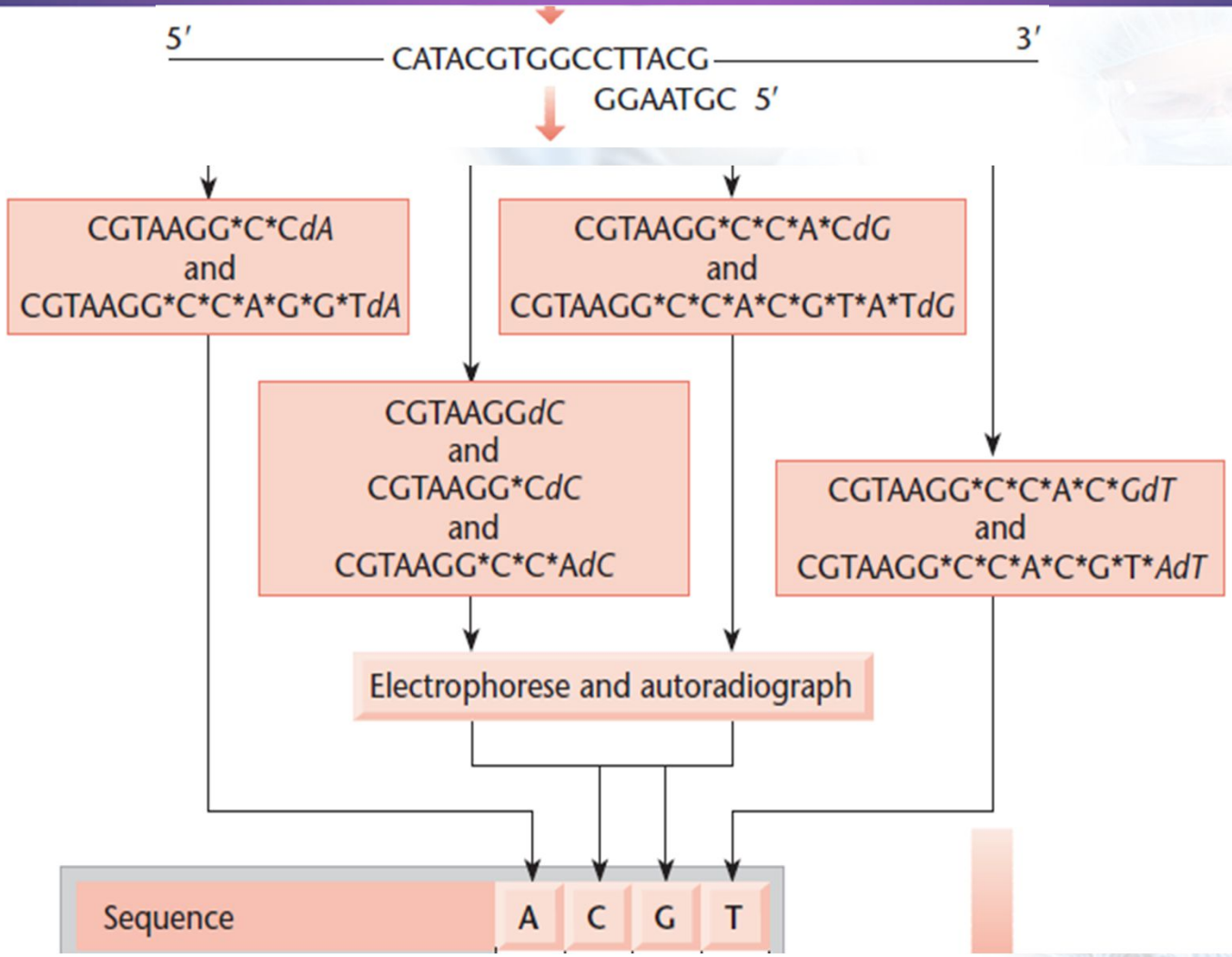
- DNA synthesis is carried out in the presence of the four deoxynucleoside triphospahtes, one or more of which is labelled with ^{32}P and in four separate incubation mixes

Analysing DNA sequences

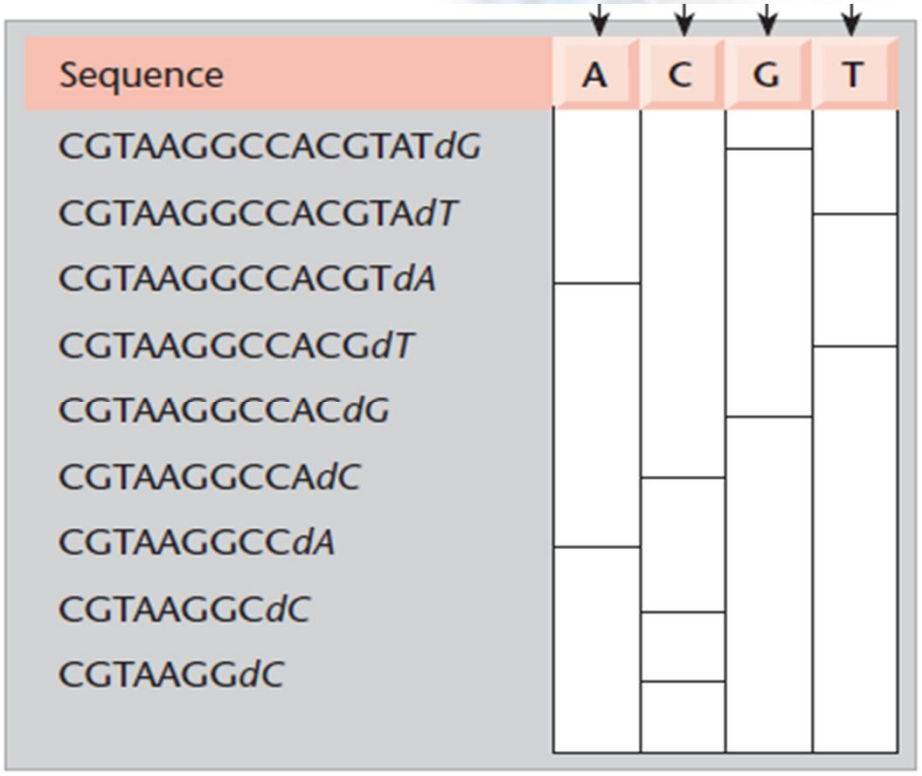


DNA sequencing by chain terminator method

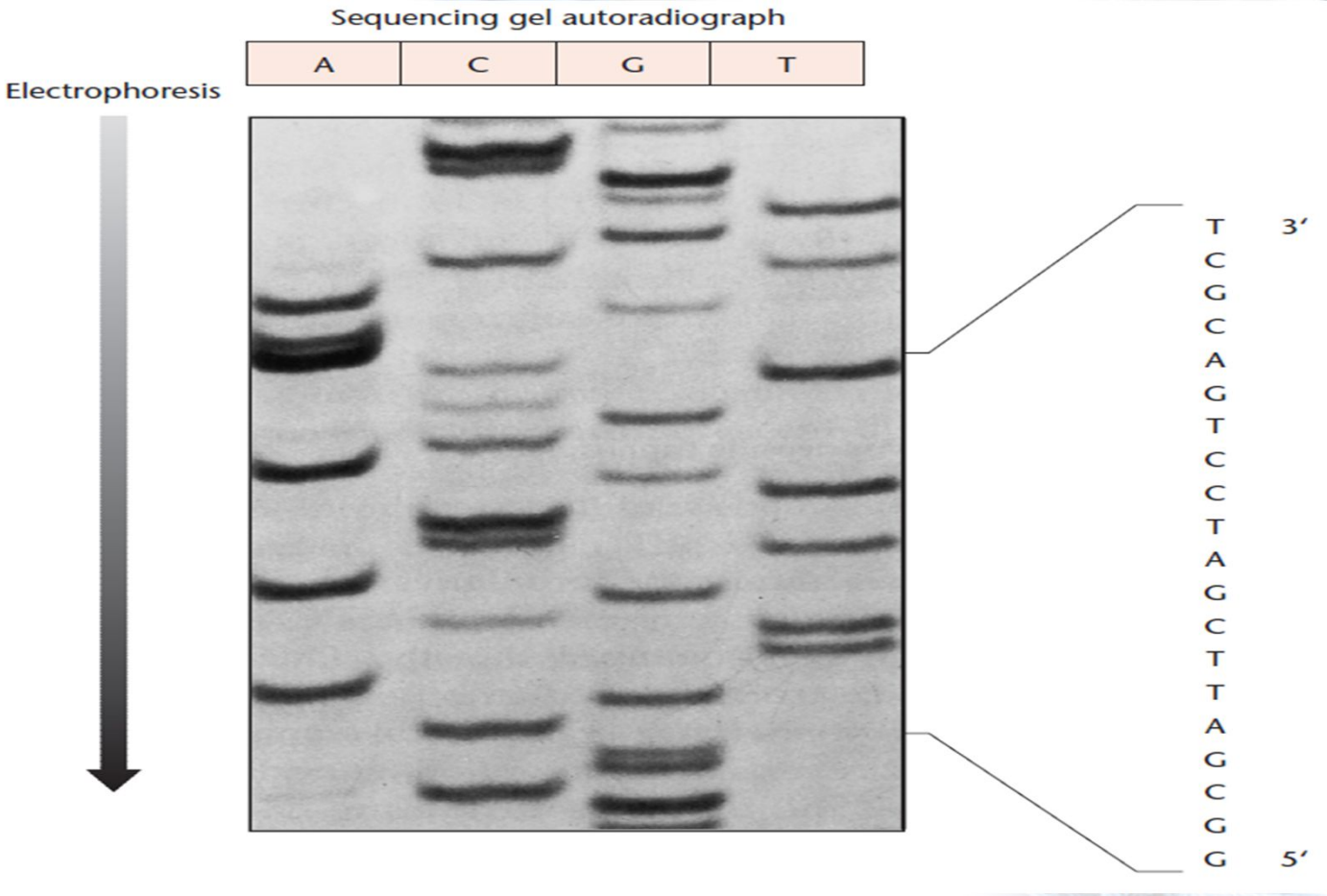
Analysing DNA sequences



Analysing DNA sequences



Analysing DNA sequences



Analysing DNA sequences

Modifications of chain terminator sequencing

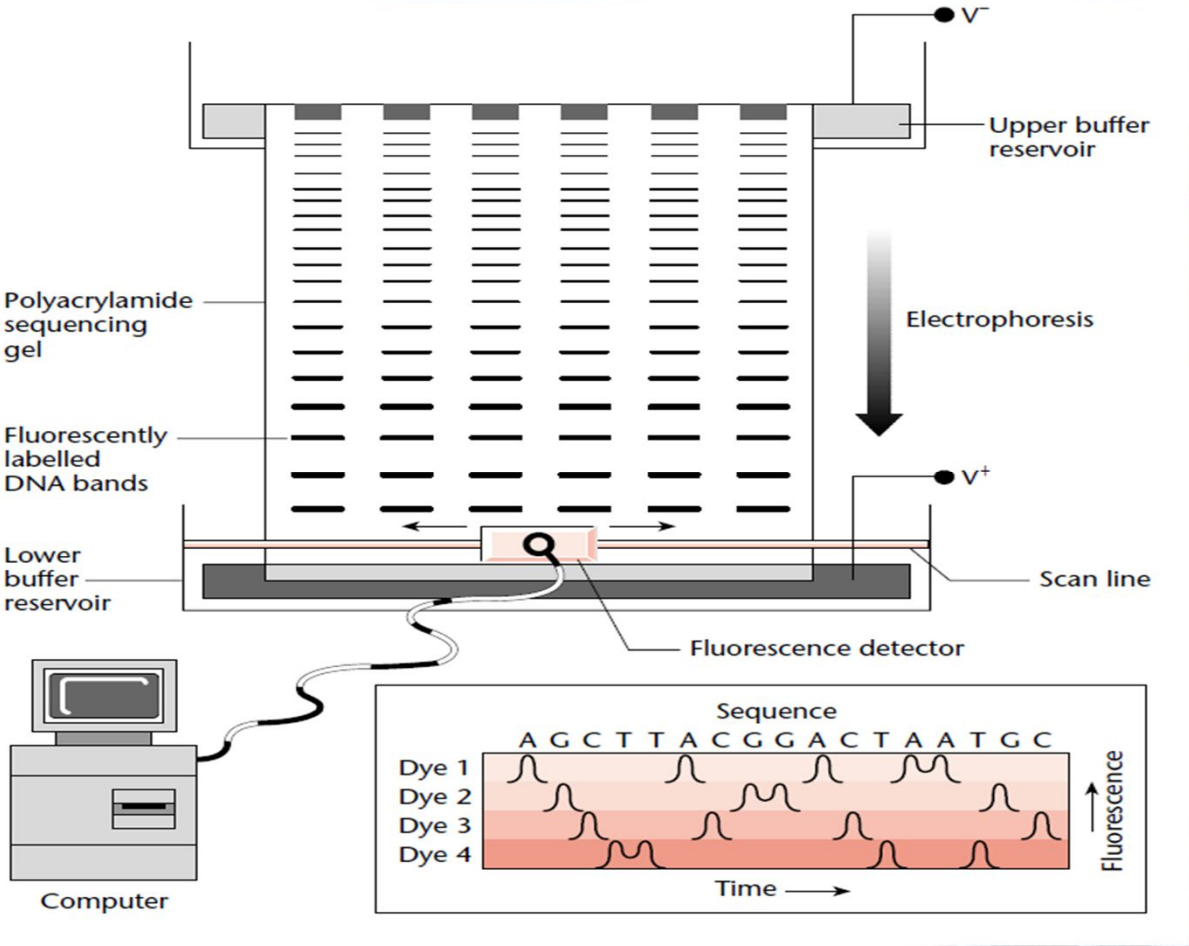
- Improvements to original Sanger's method have been made by replacing the Klenow fragment of *E. coli* DNA polymerase I
- The combination of chain terminator and M_{13} vectors to produce single stranded DNA is very powerful

Analysing DNA sequences

Automated DNA sequencing

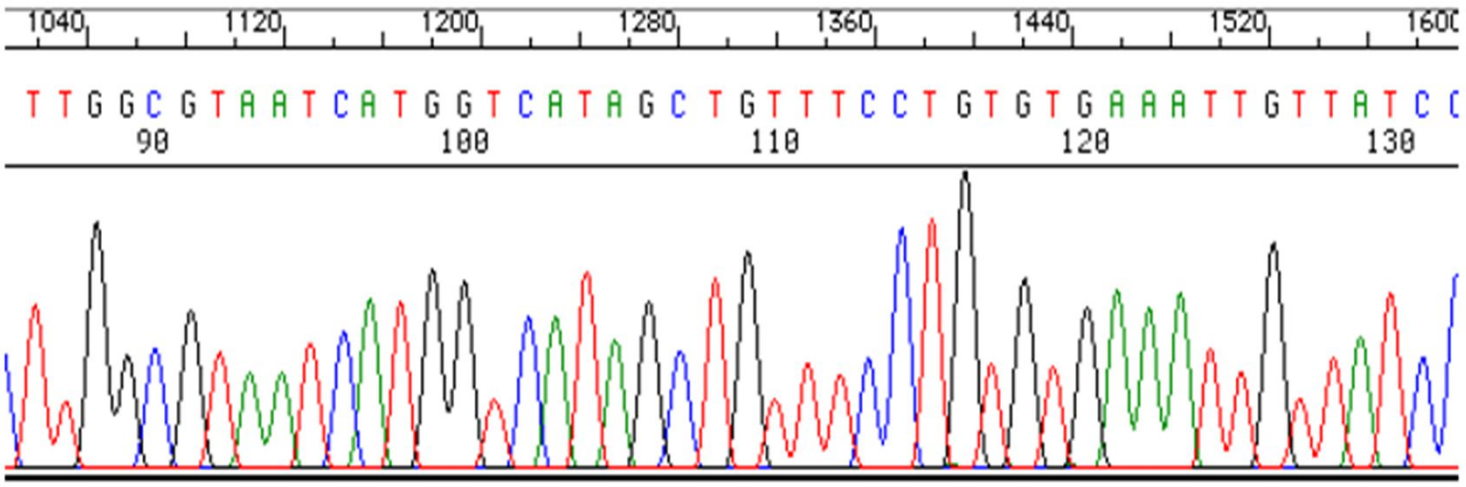
- To automate the process, it is desirable to acquire sequence data in real time by detecting the DNA bands within the gel during the electrophoretic separation

Analysing DNA sequences



Block diagram of an automated DNA sequencer

Analysing DNA sequences



Sequence electropherogram

Analysing DNA sequences

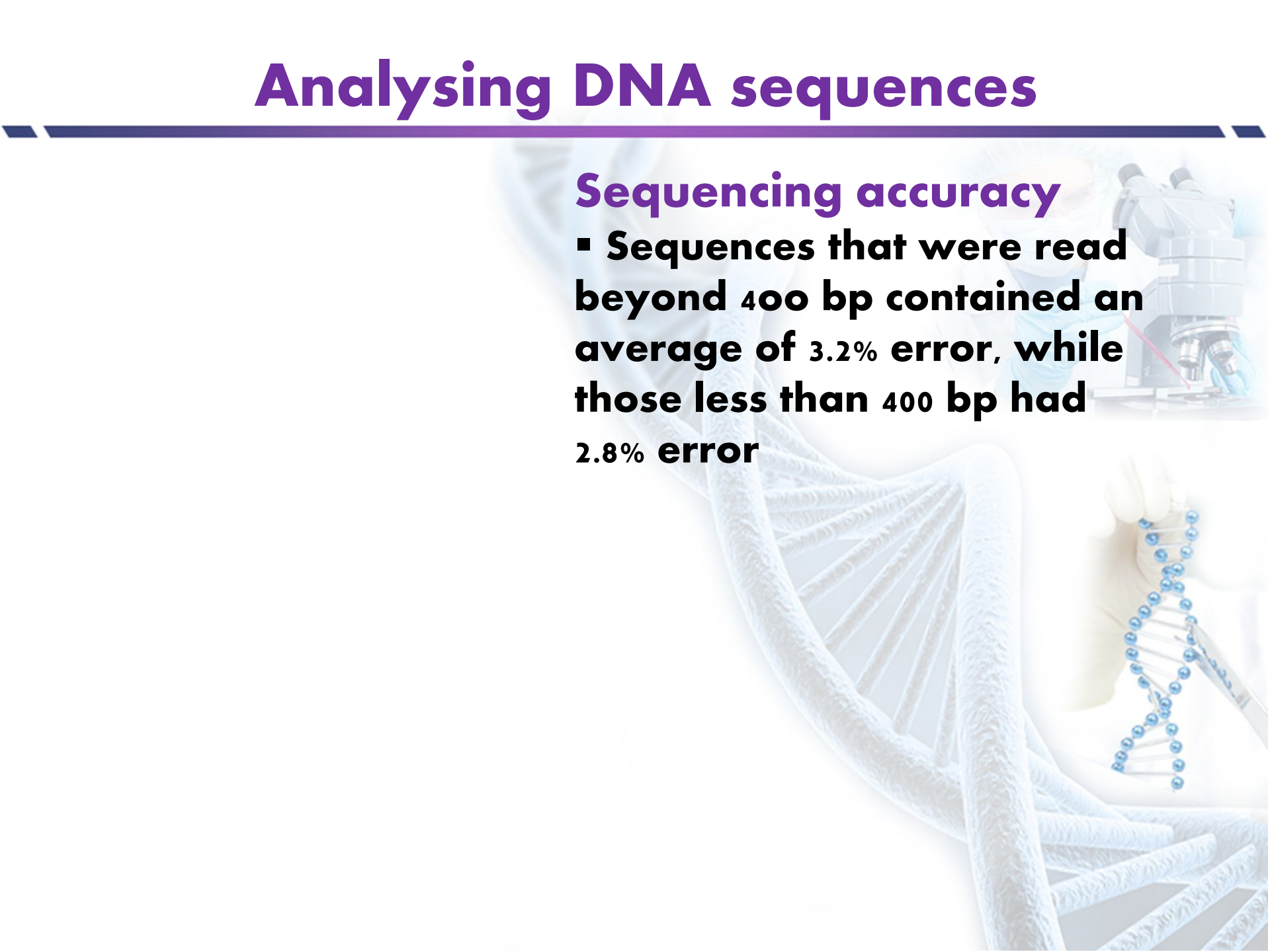
Automated DNA sequencing

- **Automated DNA sequencing offers a number of advantages**

Analysing DNA sequences

Sequencing accuracy

- Sequences that were read beyond 400 bp contained an average of 3.2% error, while those less than 400 bp had 2.8% error



Analysing DNA sequences

DNA sequence databases

- **Since the development of current DNA sequencing technology large amount of sequence data has accumulated that is maintained in 3 data bases**

i). National center for Biotechnology Information

ii). DNA Databank of Japan

iii). European Bioinformatic institute-UK

Changing genes: site-directed mutagenesis

Mutagenesis

- Mutagenesis is a process to change the genetic information of an organism
- It can occur naturally, or as a result of exposure to mutagens or induced experimentally in laboratory

Changing genes: site-directed mutagenesis

Mutagenesis

- **Three different methods of site-directed mutagenesis have been devised**

i). Cassette mutagenesis

ii). Primer extension

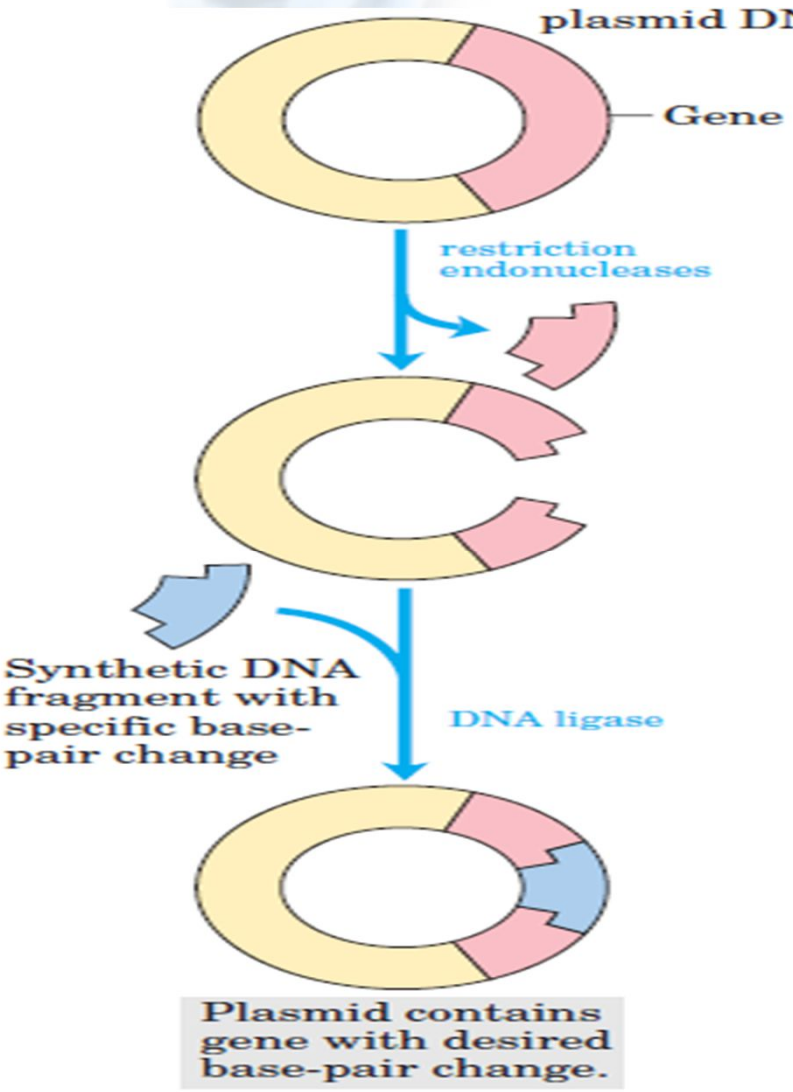
iii). Procedures based on PCR

Changing genes: site-directed mutagenesis

Cassette mutagenesis

- A synthetic DNA fragment containing the desired mutant sequence is used to replace the corresponding sequence in the wild-type gene

Changing genes: site-directed mutagenesis

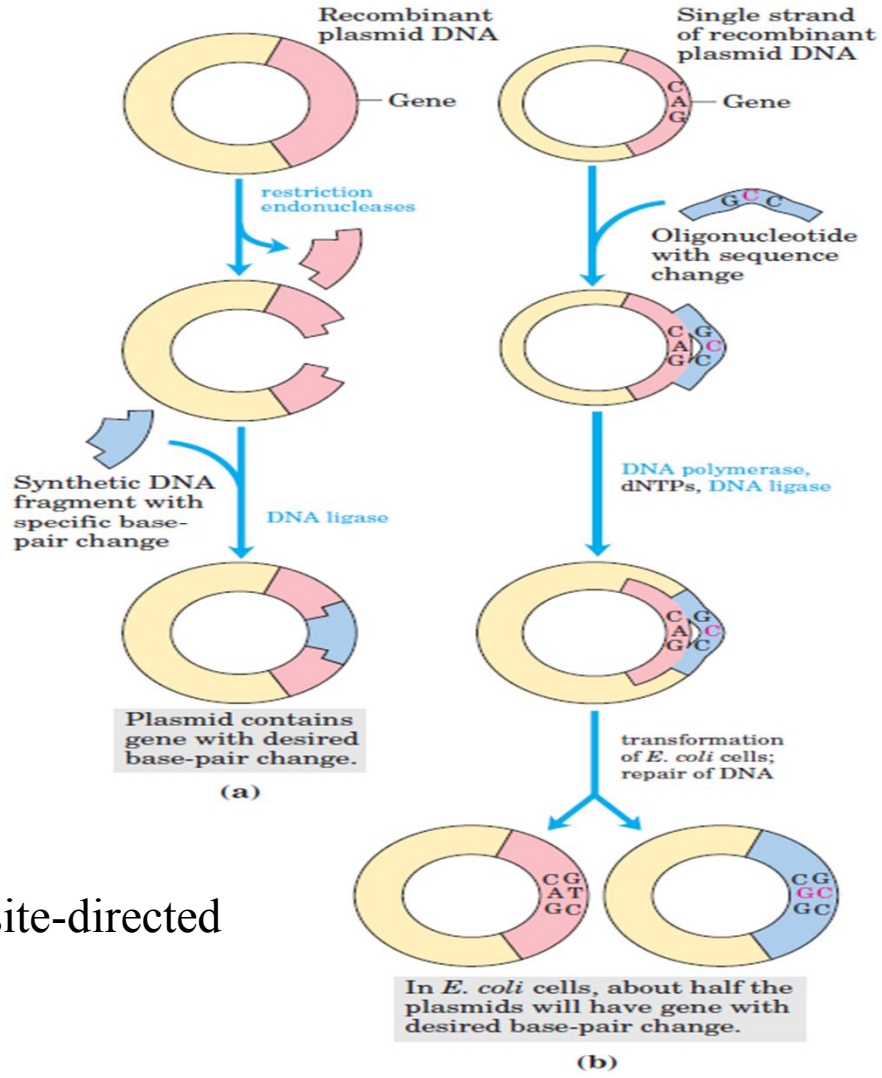


Changing genes: site-directed mutagenesis

Primer extension: the single primer method

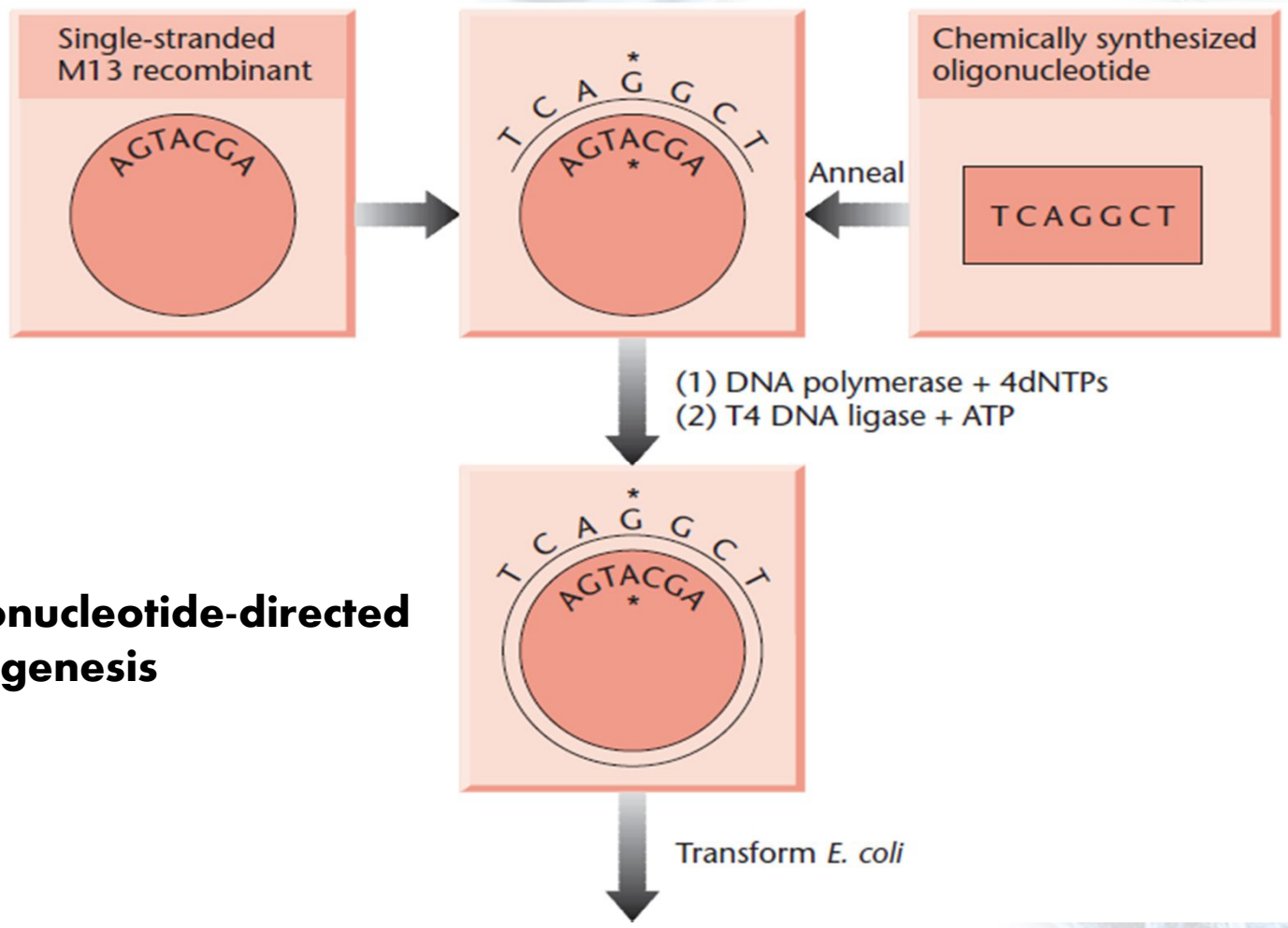
- It involves priming in vitro DNA synthesis with a chemically synthesized oligonucleotide (7-20 bp) that carries a base mismatch with the complementary sequence

Changing genes: site-directed mutagenesis



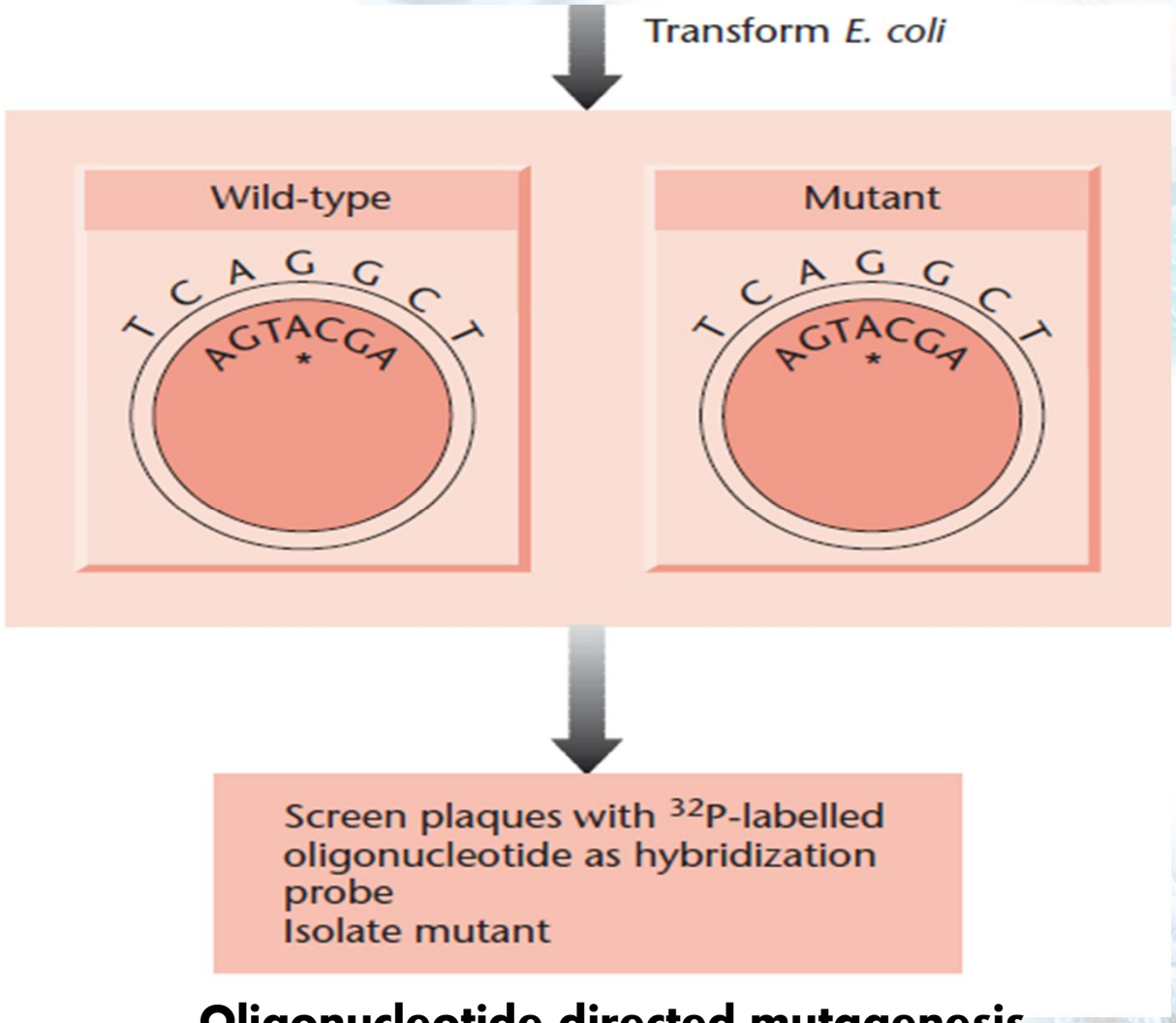
Two approaches to site-directed mutagenesis

Changing genes: site-directed mutagenesis



Oligonucleotide-directed mutagenesis

Changing genes: site-directed mutagenesis



Oligonucleotide-directed mutagenesis

Changing genes: site-directed mutagenesis

Primer extension: the single primer method

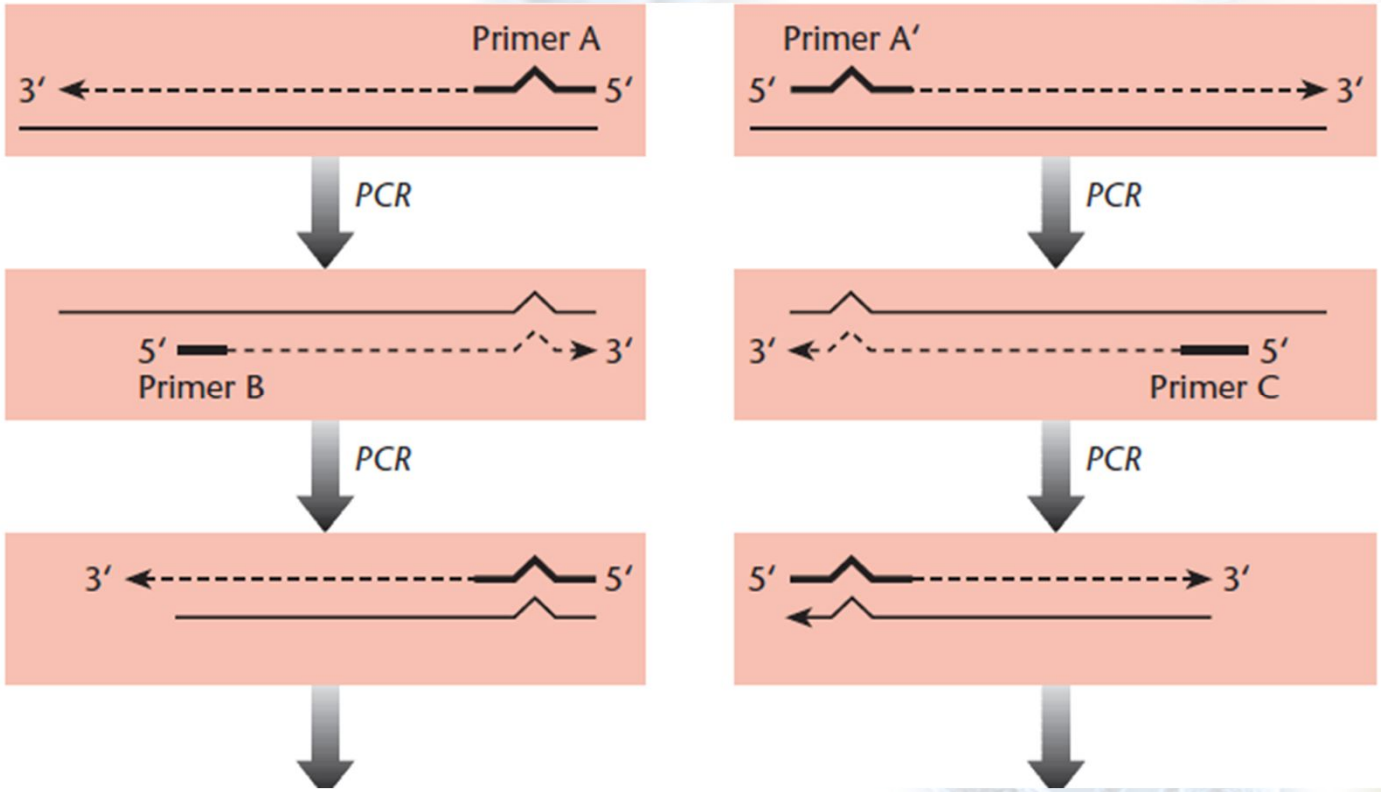
- **After *E. coli* transformation, heteroduplex that are formed are either that of original wild type DNA or that containing the mutated base**

Changing genes: site-directed mutagenesis

PCR methods for site-directed mutagenesis

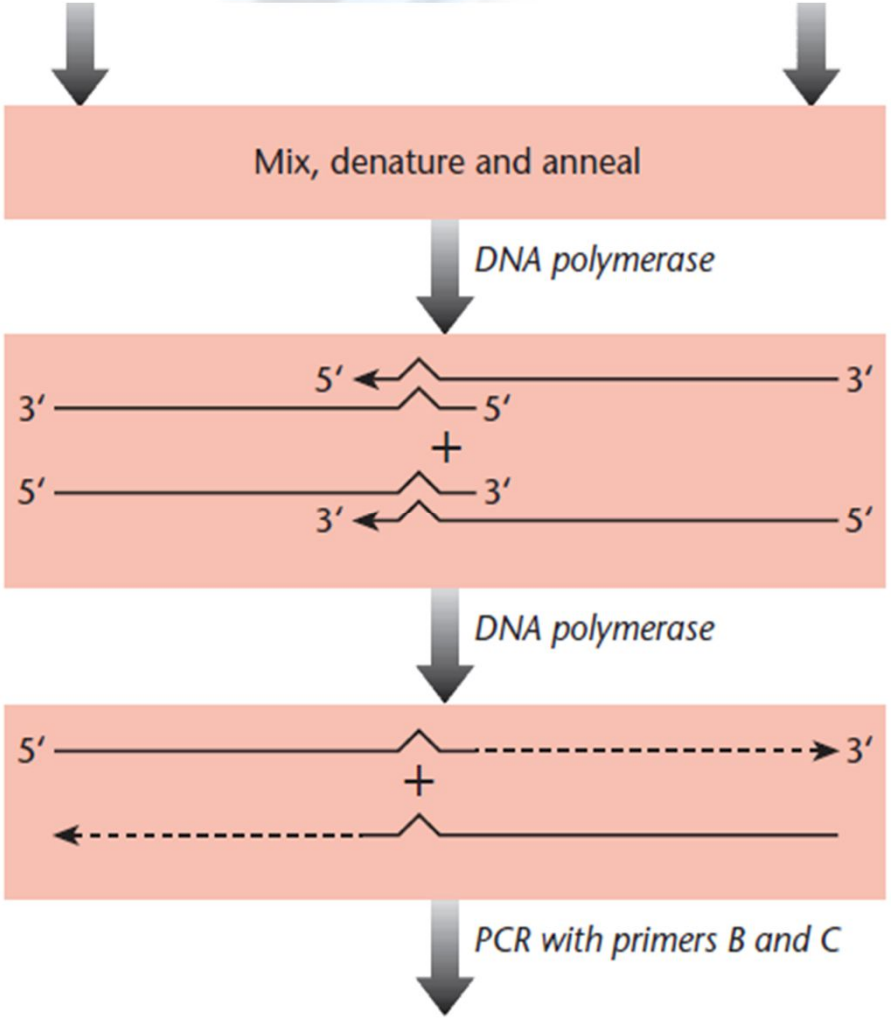
- Single bases mismatched between the amplification primer and the template become incorporated into the template sequence as a result of amplification

Changing genes: site-directed mutagenesis



Site directed mutagenesis by means of the PCR

Changing genes: site-directed mutagenesis



Site directed mutagenesis by means of the PCR

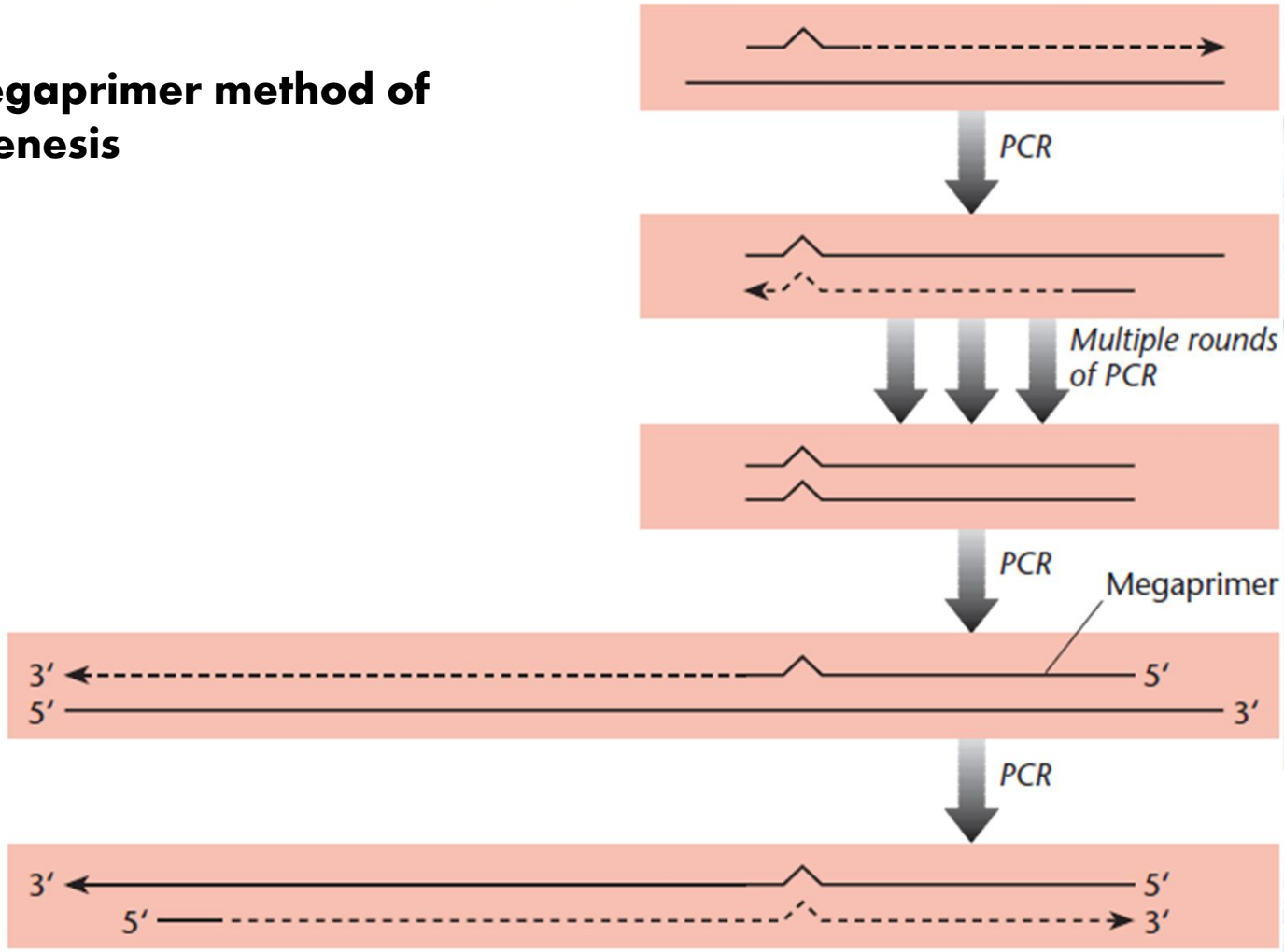
Changing genes: site-directed mutagenesis

PCR methods for site-directed mutagenesis

- In megaprimer method, the product of the first PCR is used for the second PCR

Changing genes: site-directed mutagenesis

The megaprimer method of mutagenesis



Changing genes: site-directed mutagenesis

PCR methods for site-directed mutagenesis

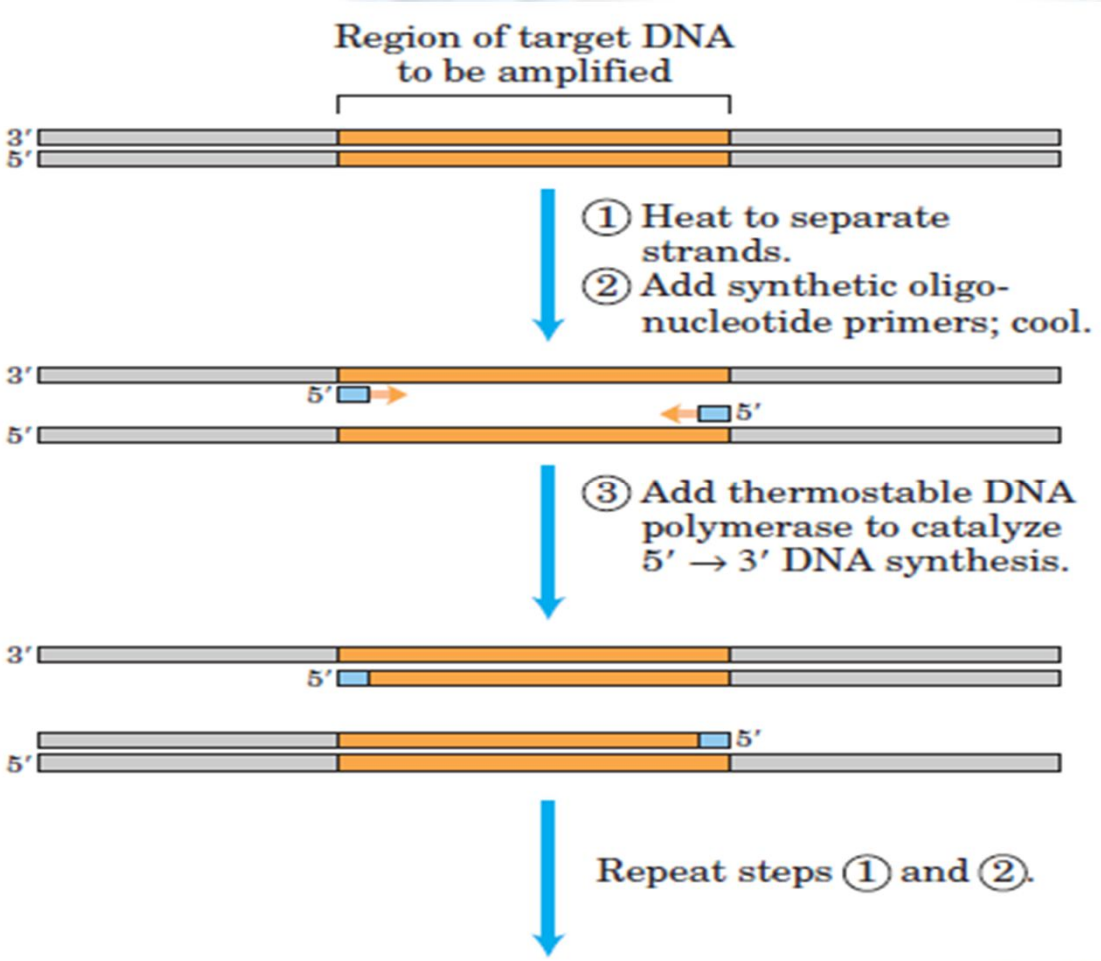
- The advantage of a PCR-based mutagenic protocol is that the desired mutation is obtained with 100% efficiency

Polymerase chain reaction (PCR)

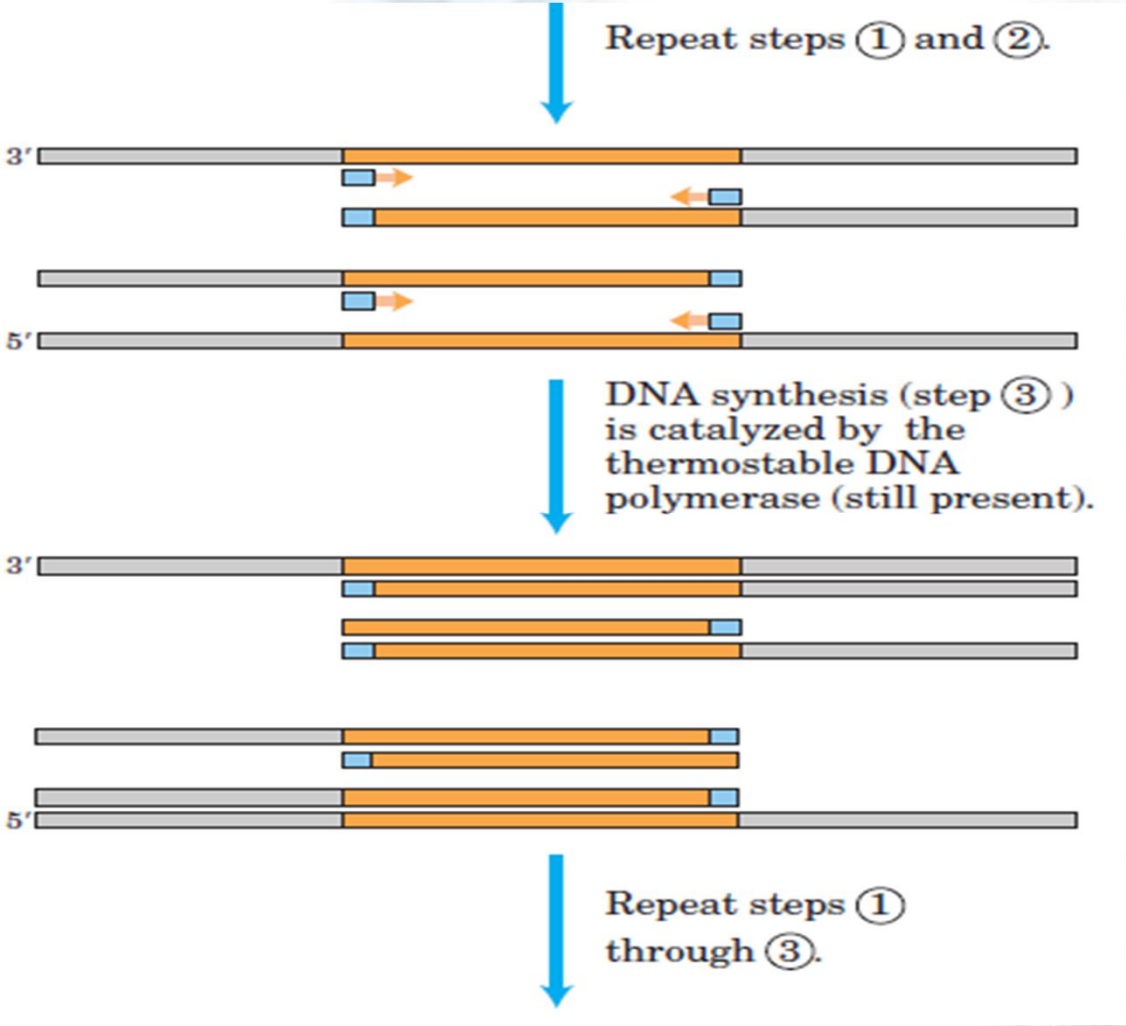
Basic PCR reaction

- PCR is a technology in molecular biology used to amplify a single or a few copies of a target DNA across several orders of magnitude, generating millions of copies of a particular DNA sequence

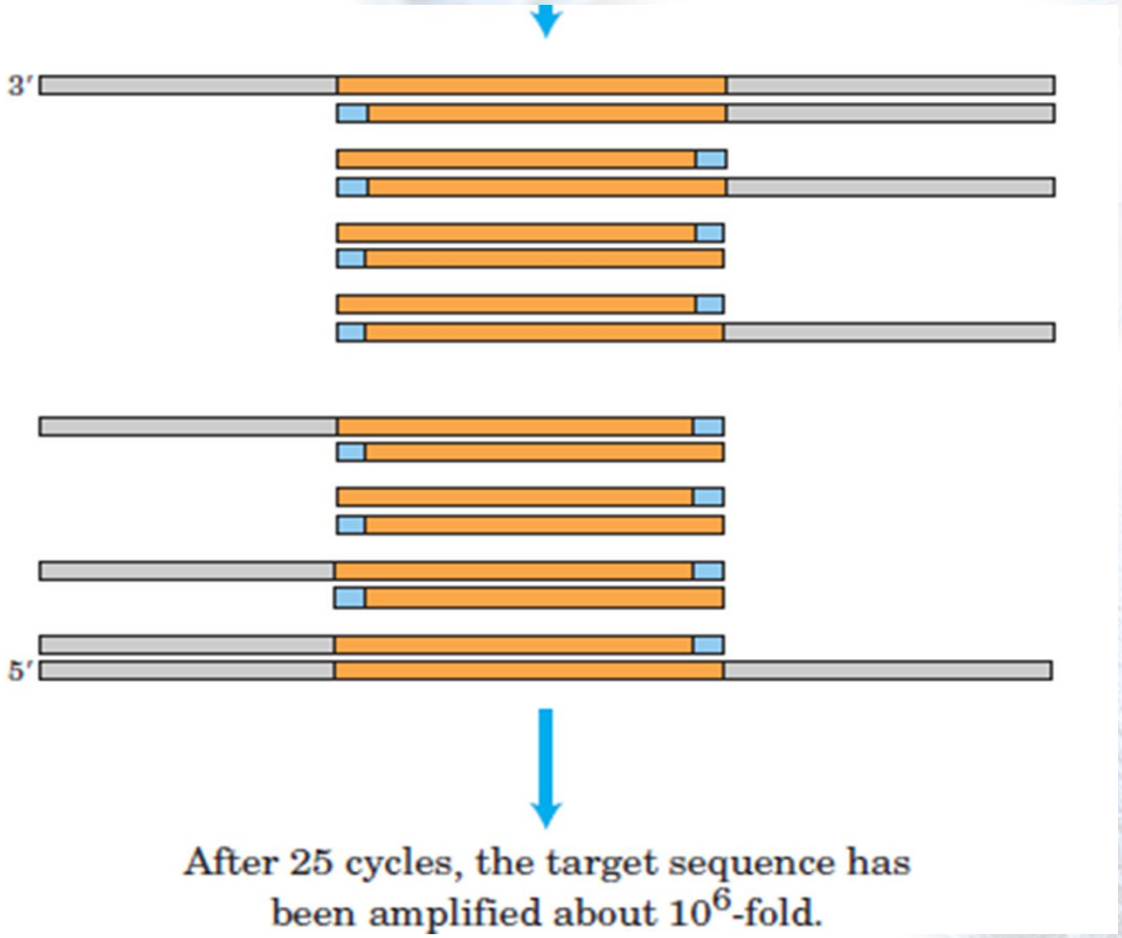
Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)

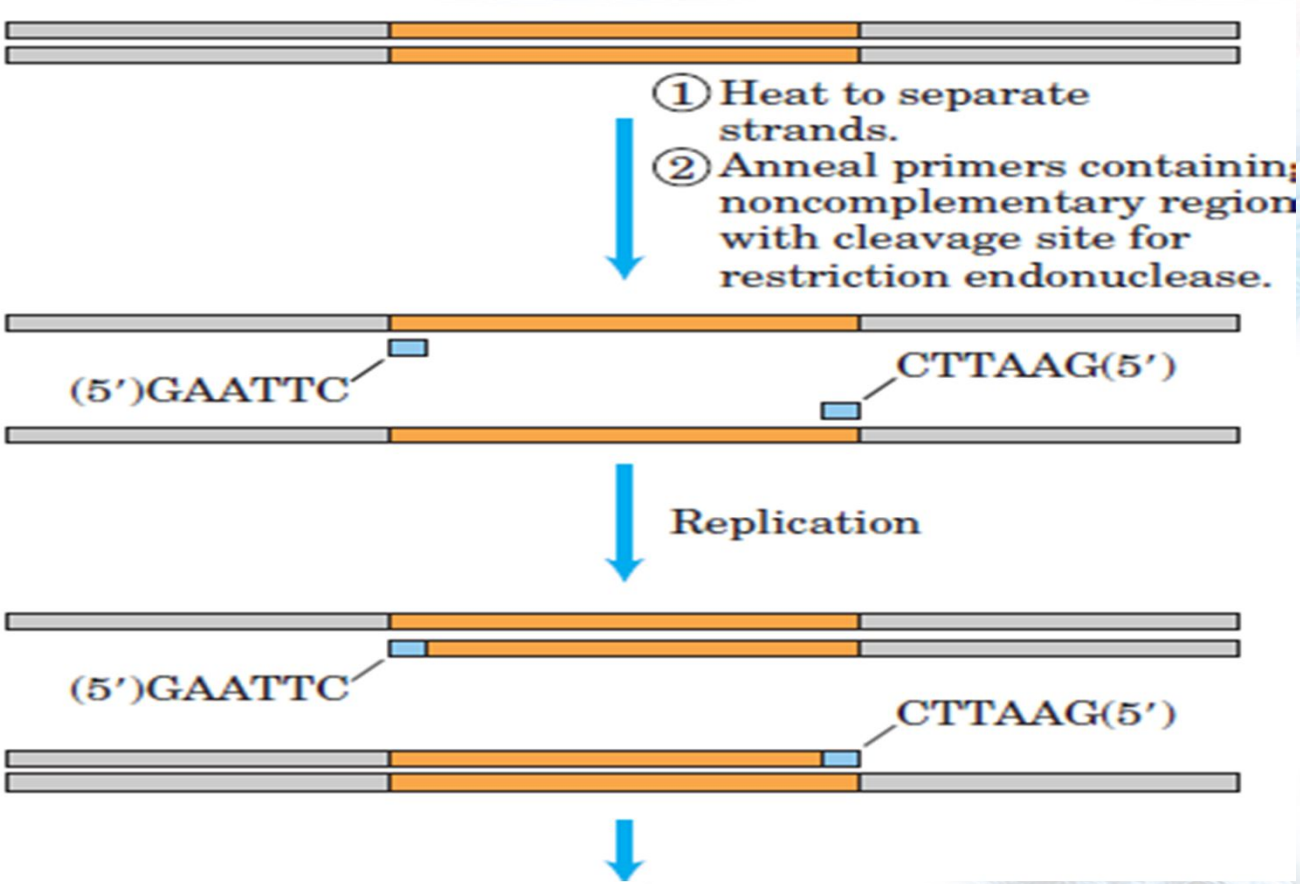
Cycle number	Number of double-stranded target molecules
1	0
2	0
3	2
4	4
5	8
6	16
7	32
8	64
9	128
10	256
11	512
12	1024
13	2048
14	4096
15	8192
16	16,384
17	32,768
18	65,536
19	131,072
20	262,144
21	524,288
22	1,048,576
23	2,097,152
24	4,194,304
25	8,388,608
26	16,777,216
27	33,554,432
28	67,108,864
29	134,217,728
30	268,435,456

Polymerase chain reaction (PCR)

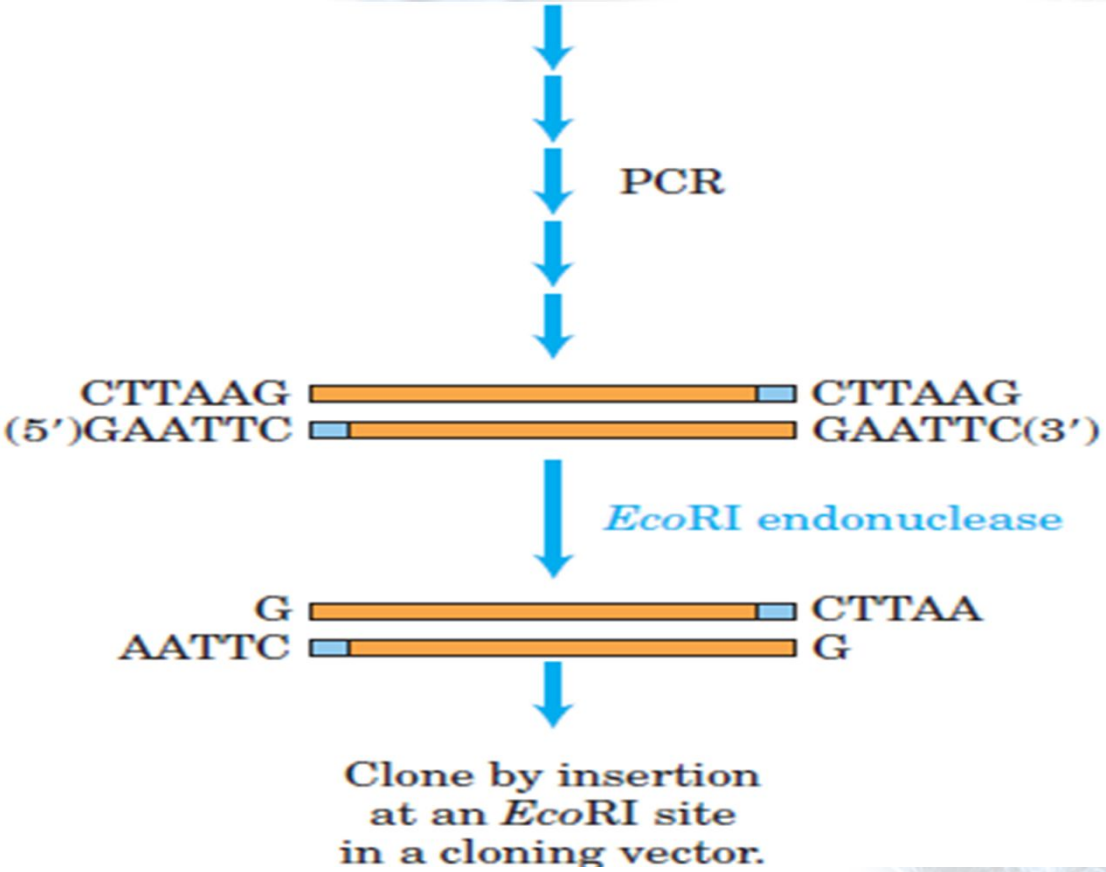
Basic PCR reaction

- PCR can also be used to generate DNA fragments for gene cloning by using primers with specific restriction sites

Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)

Basic PCR reaction

- PCR is a technology in molecular biology used to amplify a single or a few copies of a target DNA across several orders of magnitude, generating millions of copies of a particular DNA sequence

Polymerase chain reaction (PCR)

DNA polymerases

- **DNA polymerases are enzymes that synthesize DNA molecules by adding nucleotide, the building blocks of DNA**

Polymerase chain reaction (PCR)

Table. Comparison of DNA polymerases of E. coli

	DNA polymerase		
	I	II	III
Structural gene*	*polA*	*polB*	*polC* (*dnaE*)
Subunits (number of different types)	1	7	≥10
M_r	103,000	88,000†	791,500
3′→5′ Exonuclease (proofreading)	Yes	Yes	Yes
5′→3′ Exonuclease	Yes	No	No
Polymerization rate (nucleotides/s)	16–20	40	250–1,000
Processivity (nucleotides added before polymerase dissociates)	3–200	1,500	≥500,000

Polymerase chain reaction (PCR)

Taq DNA polymerase

- **Taq polymerase is a thermostable DNA polymerase isolated from thermophilic bacterium *Thermus aquaticus* and is used in PCR to amply DNA fragments**

Polymerase chain reaction (PCR)

Table. Sources of thermostable DNA polymerases

DNA polymerase	Source
Tma	*Thermotoga maritima*
Deep Vent™	*Pyrococcus* sp.
Tli	*Thermococcus litoralis*
Pfu	*Pyrococcus furiosus*
Pwo	*Pyrococcus woesi*

Polymerase chain reaction (PCR)

Primers

- Primer is a short strand of oligonucleotide that serves as a starting point for DNA synthesis

Polymerase chain reaction (PCR)

Degenerate primers

- A degenerate primer is actually a mixture of primers, all of similar sequence but with variations at one or more positions

Polymerase chain reaction (PCR)

Table. Amino acids by codon specificity

One Codon	Two Codons	Three Codons	Four Codons	Six Codons
Met (M)	Cys (C)	Ile (I)	Ala (A)	Leu (L)
Trp (W)	Asp (D)		Gly (G)	Arg (R)
	Glu (E)		Pro (P)	Ser (S)
	Phe (F)		Thr (T)	
	His (H)		Val (V)	
	Lys (K)			
	Asn (N)			
	Gln (Q)			
	Tyr (Y)			

Polymerase chain reaction (PCR)

Table. Standard nucleotide coding system

Symbol	Base represented
A	A
C	C
G	G
T	T
R	A, G
Y	C, T
M	A, C
K	G, T
H	A, C, T
B	C, G, T
V	A, C, G
N	A, C, G, T

Polymerase chain reaction (PCR)

Degenerate primers

- A degenerate primer is actually a mixture of primers, all of similar sequence but with variations at one or more positions

Polymerase chain reaction (PCR)

Types of PCR

- **Recently, variants have been developed from basic PCR method to improve performance and specificity and to amplify other molecules such as RNA**

Polymerase chain reaction (PCR)

RT-PCR

- **Reverse transcriptase PCR (RT-PCR) was designed to amplify RNA sequences (mainly mRNA) through synthesis of cDNA by reverse transcriptase**

Polymerase chain reaction (PCR)

RT-PCR

- Various strategies can be adopted for first-strand cDNA synthesis

Polymerase chain reaction (PCR)

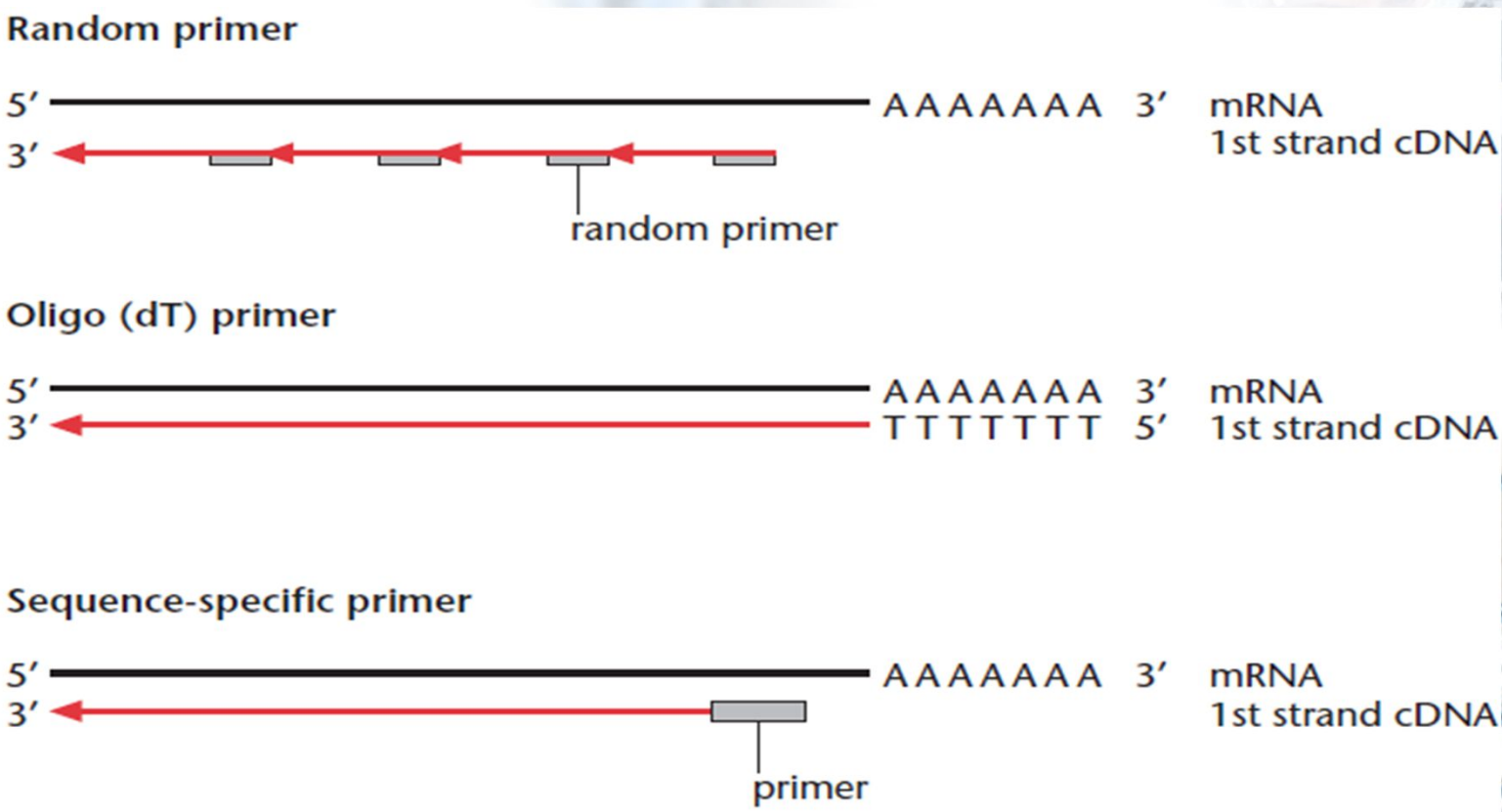


Fig. Strategies for synthesis of first-strand cDNA

Polymerase chain reaction (PCR)

RT-PCR

- Various strategies can be adopted for first-strand cDNA synthesis

Polymerase chain reaction (PCR)

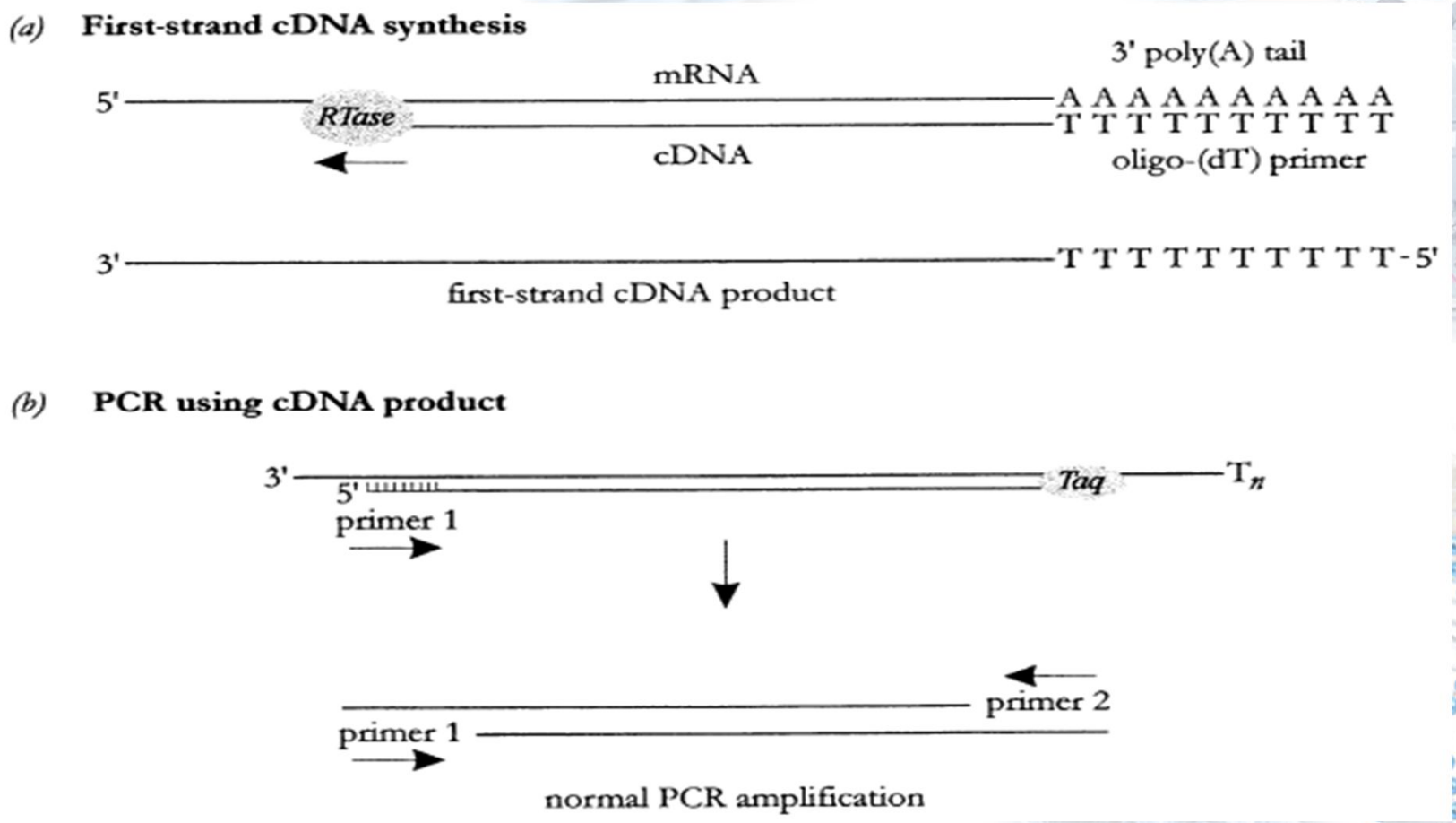


Fig. RT-PCR

Polymerase chain reaction (PCR)

Competitor RT-PCR

- One use of RT-PCR is in determining the amount of mRNA in a sample (competitor RT-PCR)

Polymerase chain reaction (PCR)

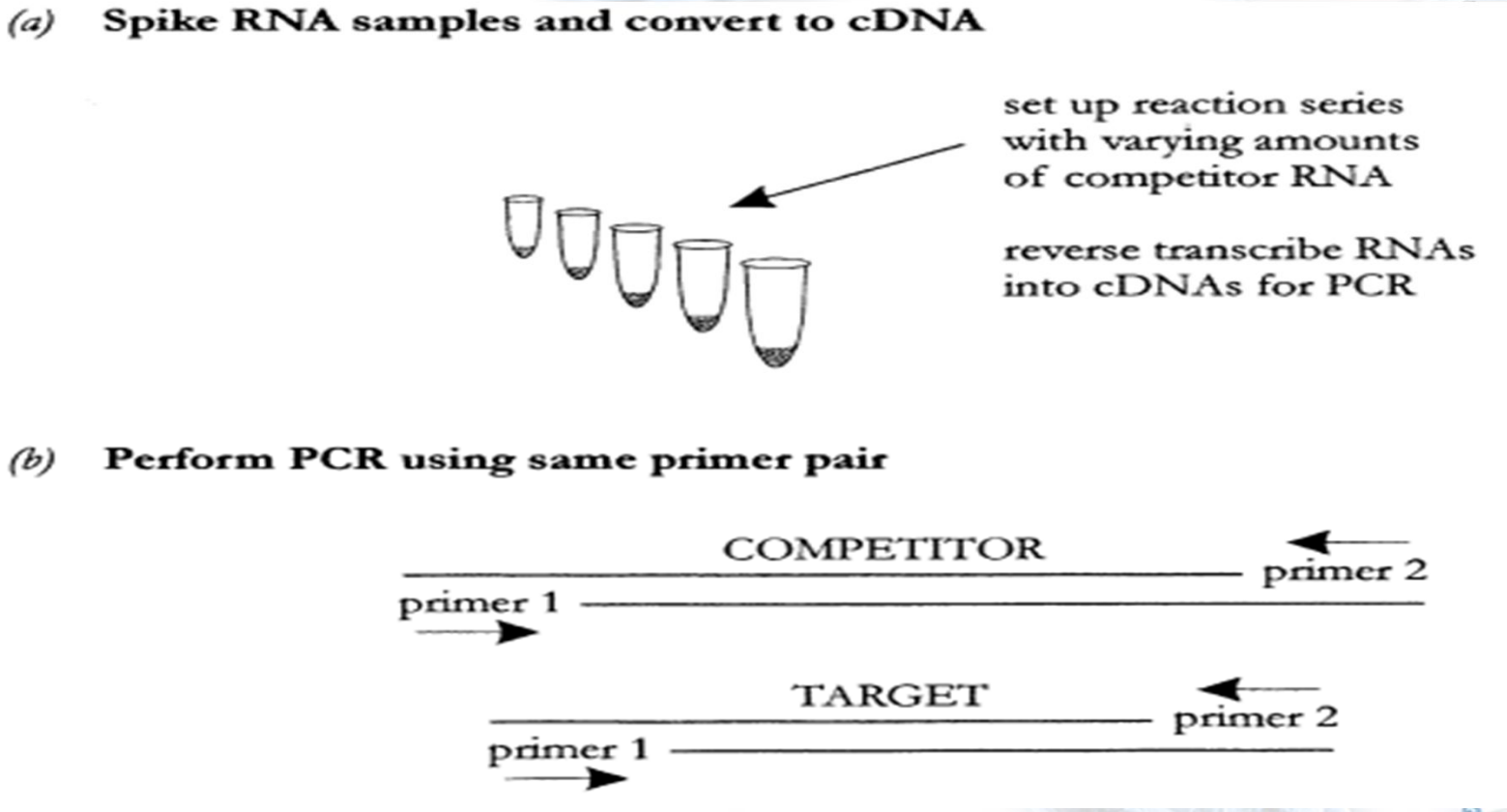


Fig. Competitor RT-PCR

Polymerase chain reaction (PCR)

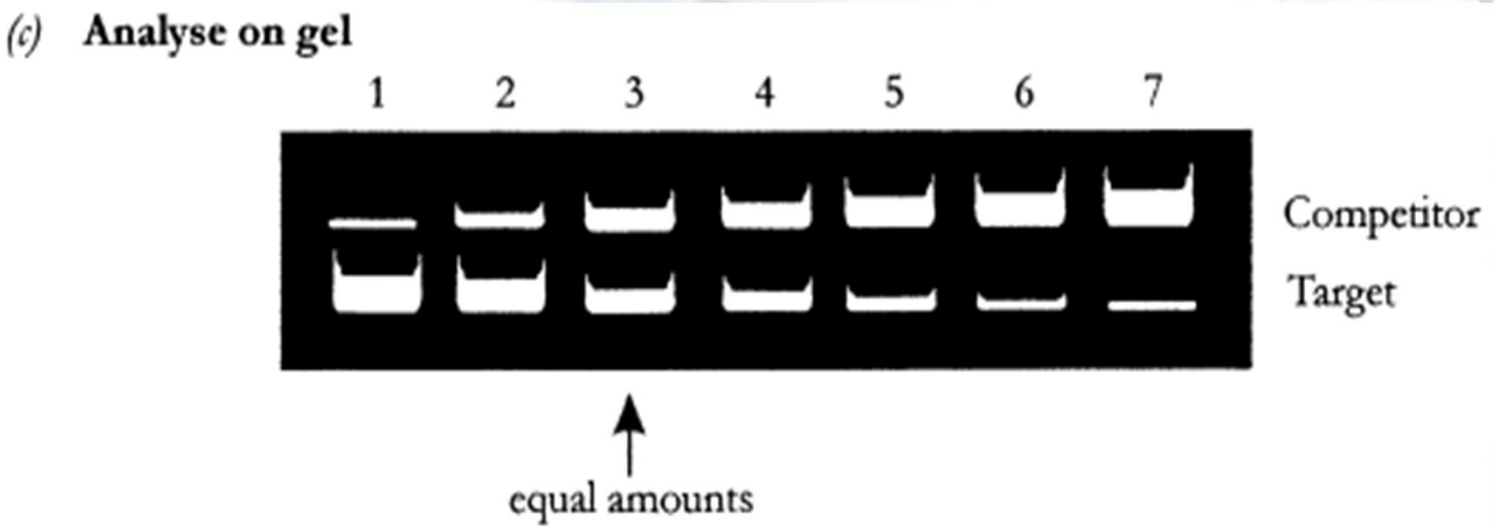


Fig. Competitor RT-PCR

Polymerase chain reaction (PCR)

RT-PCR

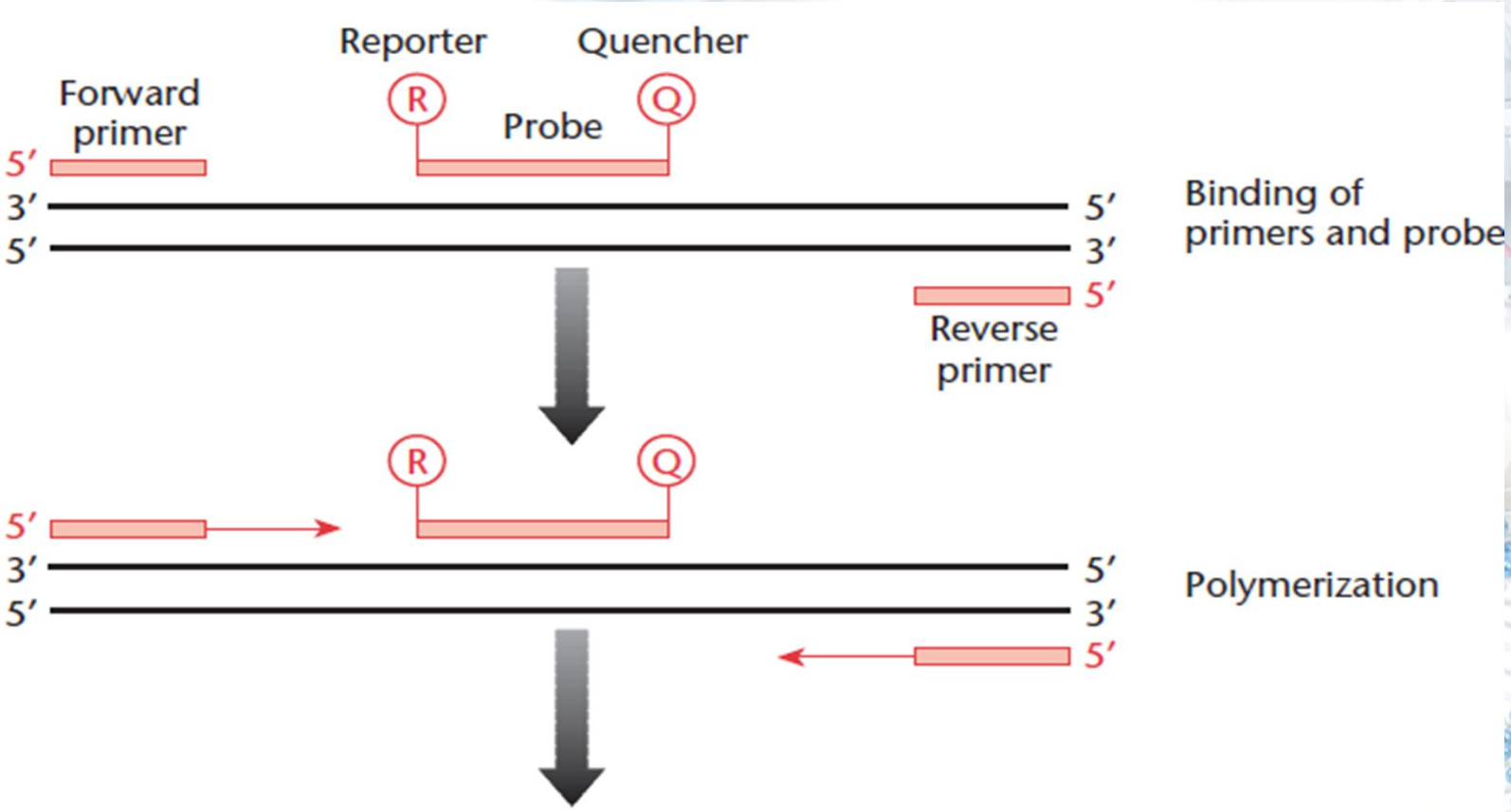
- **Reverse transcriptase PCR (RT-PCR) was designed to amplify RNA sequences (mainly mRNA) through synthesis of cDNA by reverse transcriptase**

Polymerase chain reaction (PCR)

Real-time quantitative PCR (qPCR)

- qPCR is used to measure the quantity of a target sequence
- It is used to measure starting amounts of DNA, cDNA, or RNA

Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)

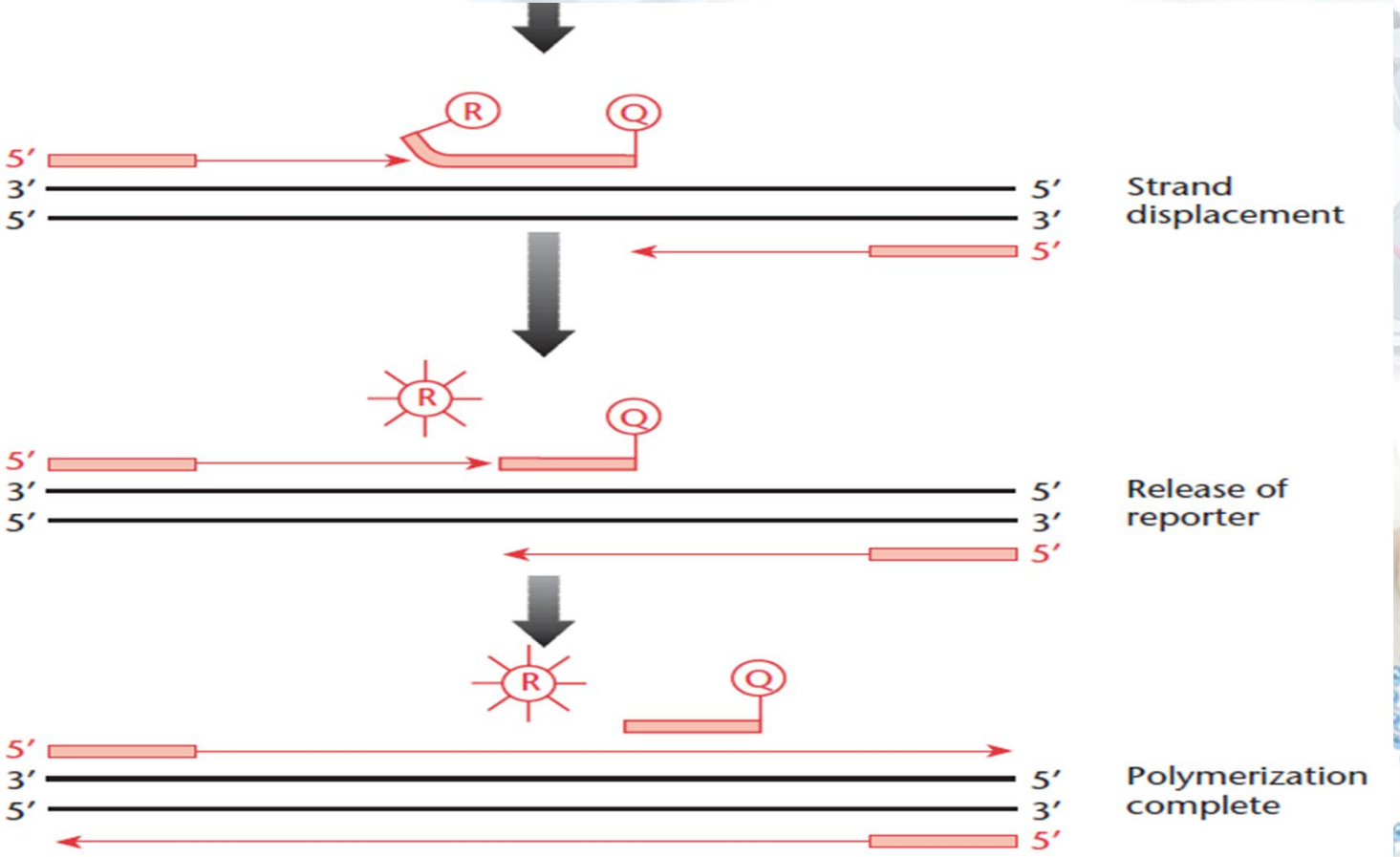


Fig. Real-time quantitative PCR

Polymerase chain reaction (PCR)

Real-time quantitative PCR (qPCR)

- qPCR is used to measure the quantity of a target sequence
- It is used to measure starting amounts of DNA, cDNA, or RNA

Polymerase chain reaction (PCR)

Nested PCR

- In nested PCR, two sets of primers are used in two successive PCRs
- First set of primers allows a first amplification
- Product of first PCR is subjected to a second PCR using the second set of primers

Polymerase chain reaction (PCR)

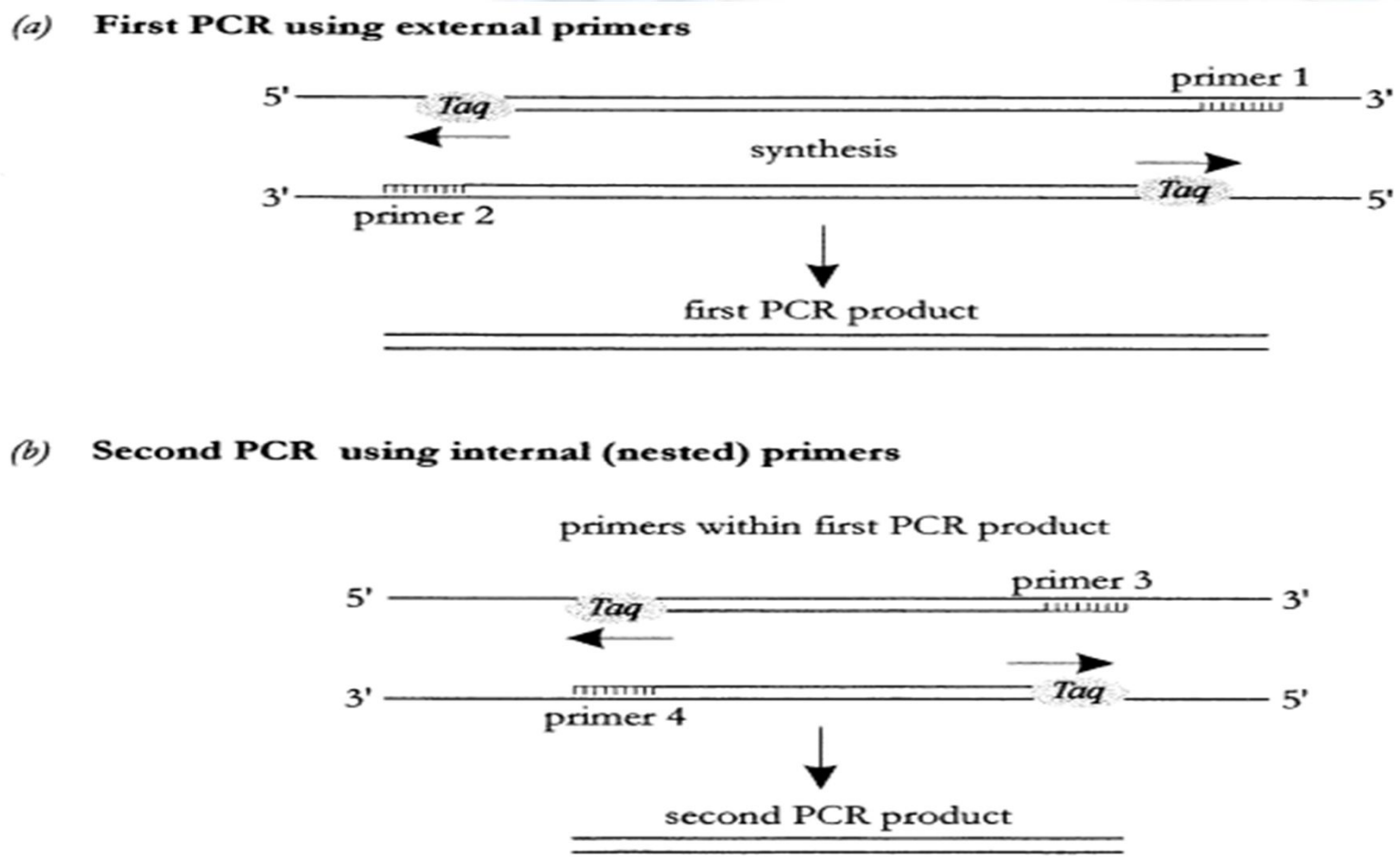


Fig. Nested PCR

Polymerase chain reaction (PCR)

Nested PCR

- In nested PCR, two sets of primers are used in two successive PCRs
- First set of primers allows a first amplification
- Product of first PCR is subjected to a second PCR using the second set of primers

Polymerase chain reaction (PCR)

Inverse PCR

- **Inverse PCR is used to identify the flanking sequences around genomic inserts**

Polymerase chain reaction (PCR)

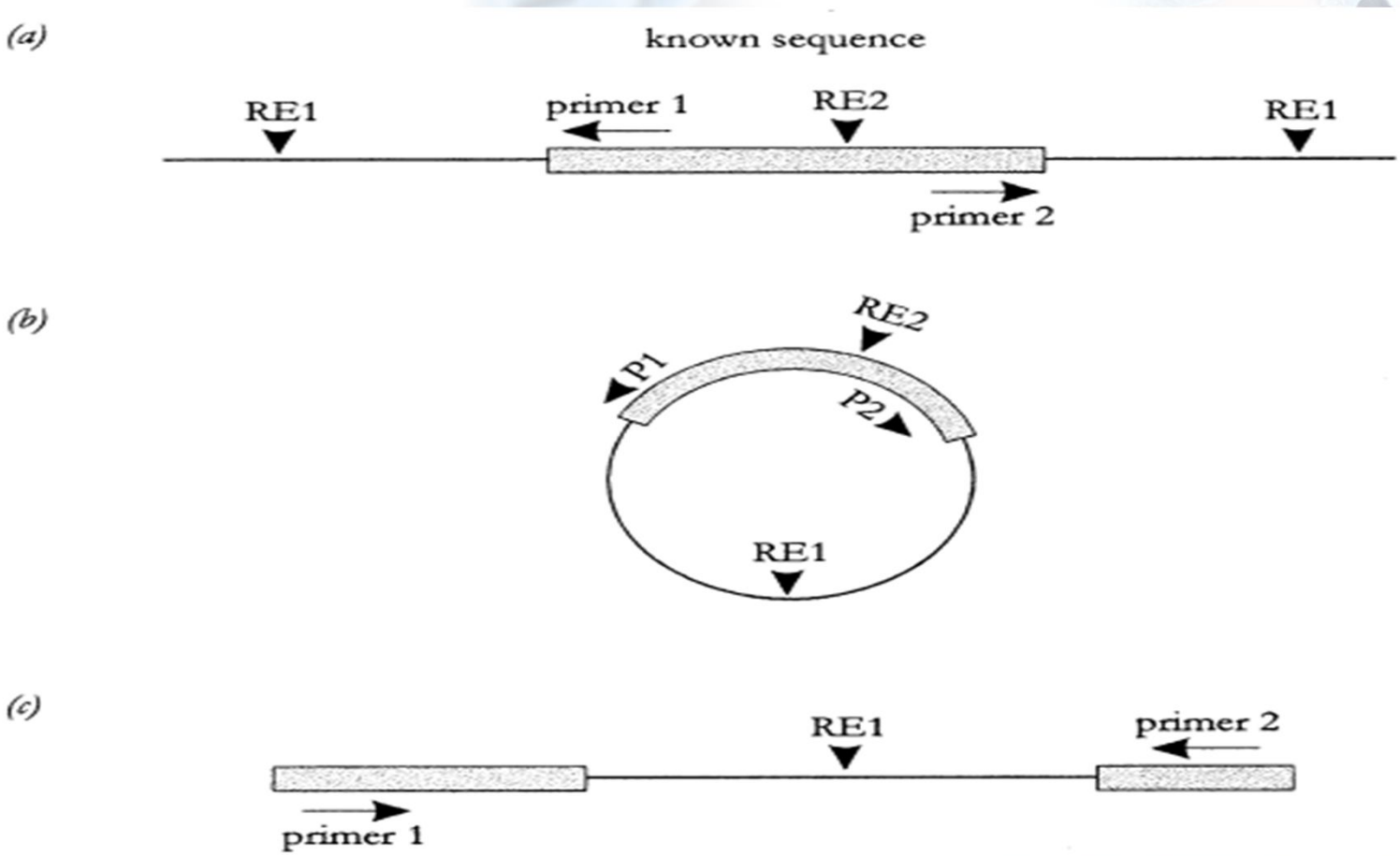


Fig. Inverse PCR

Polymerase chain reaction (PCR)

Multiplex PCR

- **Multiplex PCR is used to amplify several different DNA sequences simultaneously**

Polymerase chain reaction (PCR)

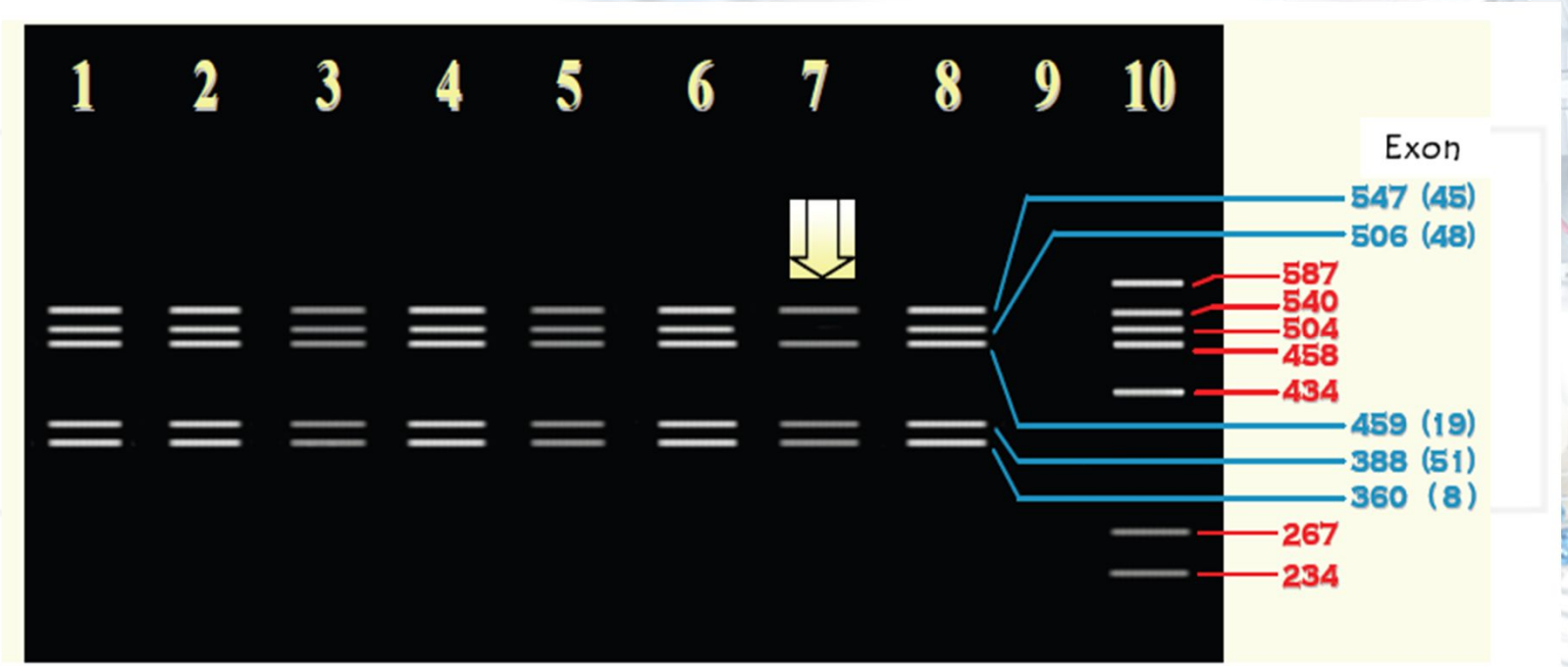


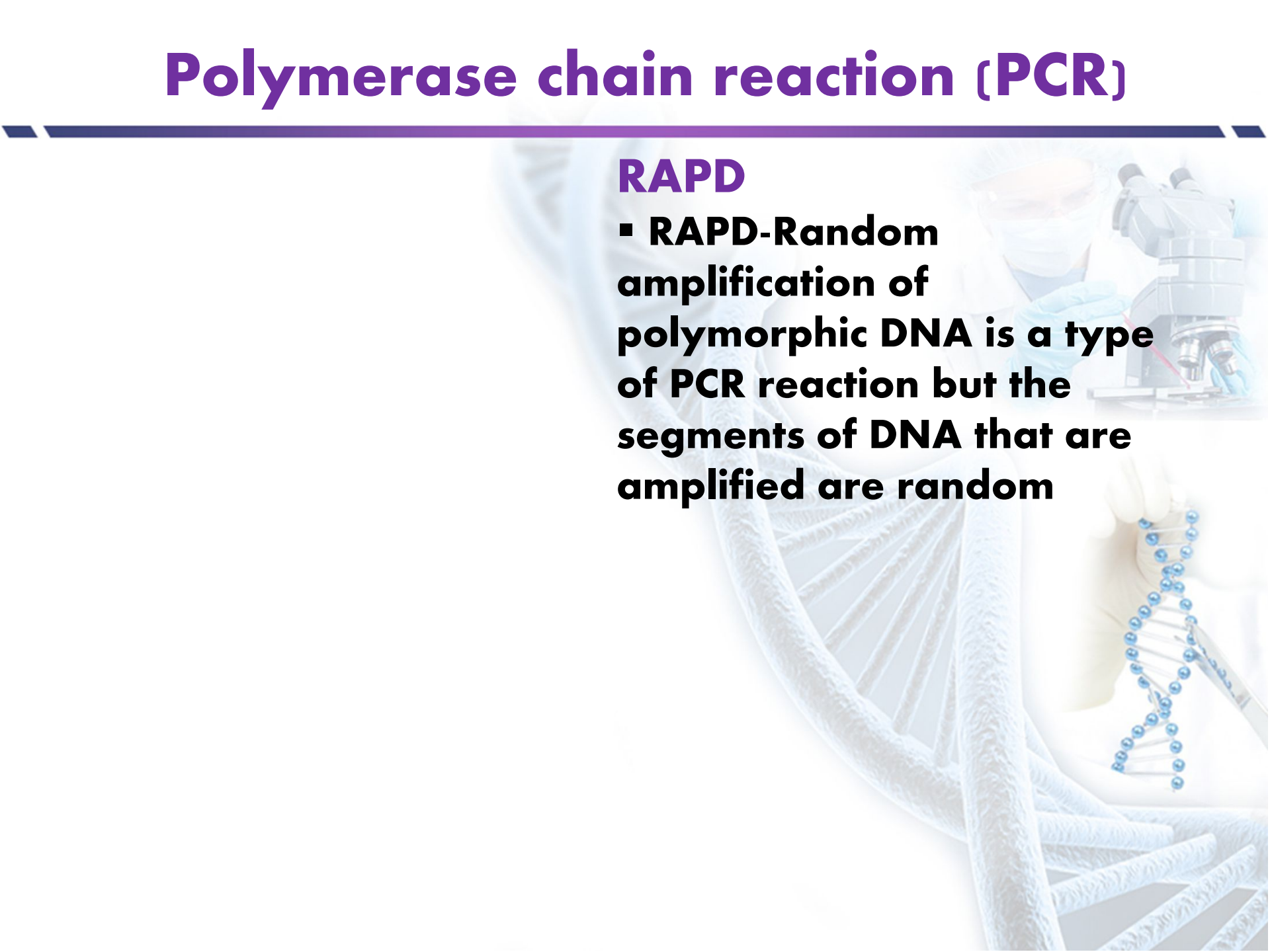
Fig. Results of multiplex PCR in a patient with Duchenne Muscular Dystrophy Inverse PCR

Hernandex-Rodriguez et al., 2000;
Hernandex-Rodriguez & Restrepo, 2002
(www.intechopen.com)

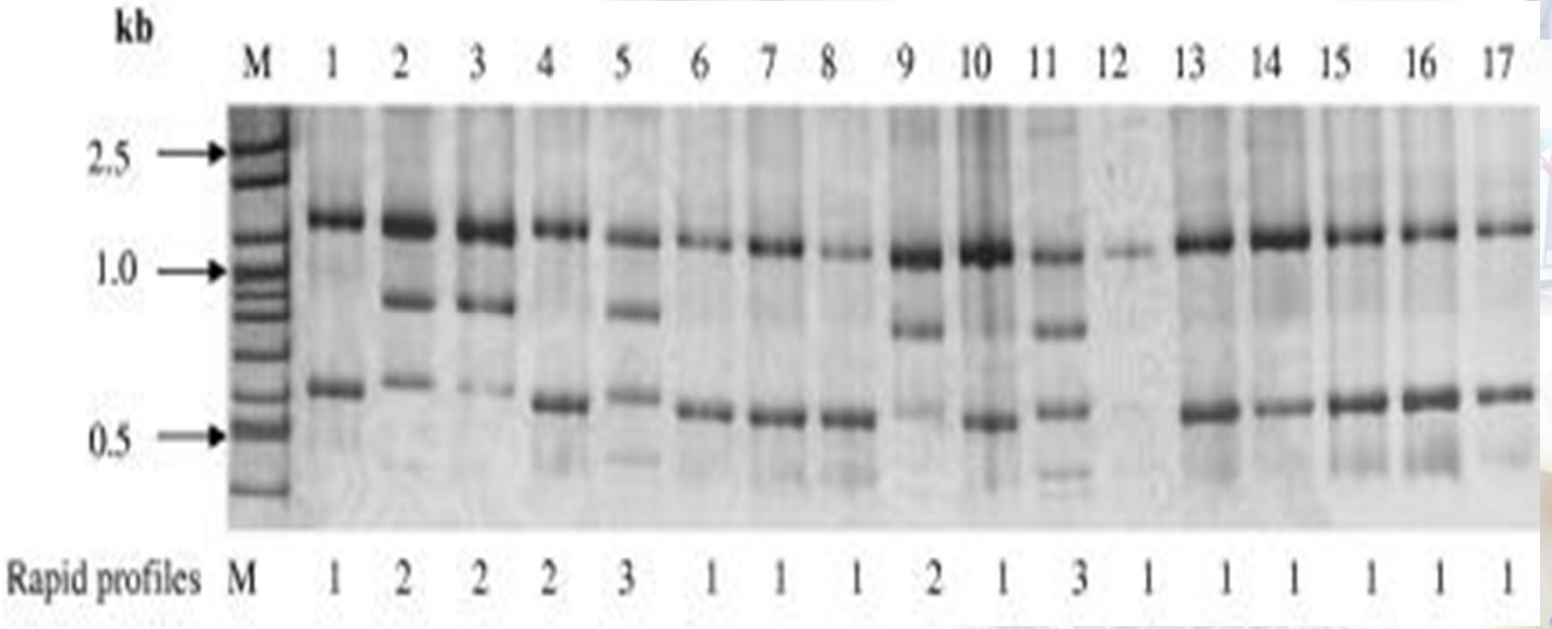
Polymerase chain reaction (PCR)

RAPD

- **RAPD-Random amplification of polymorphic DNA is a type of PCR reaction but the segments of DNA that are amplified are random**



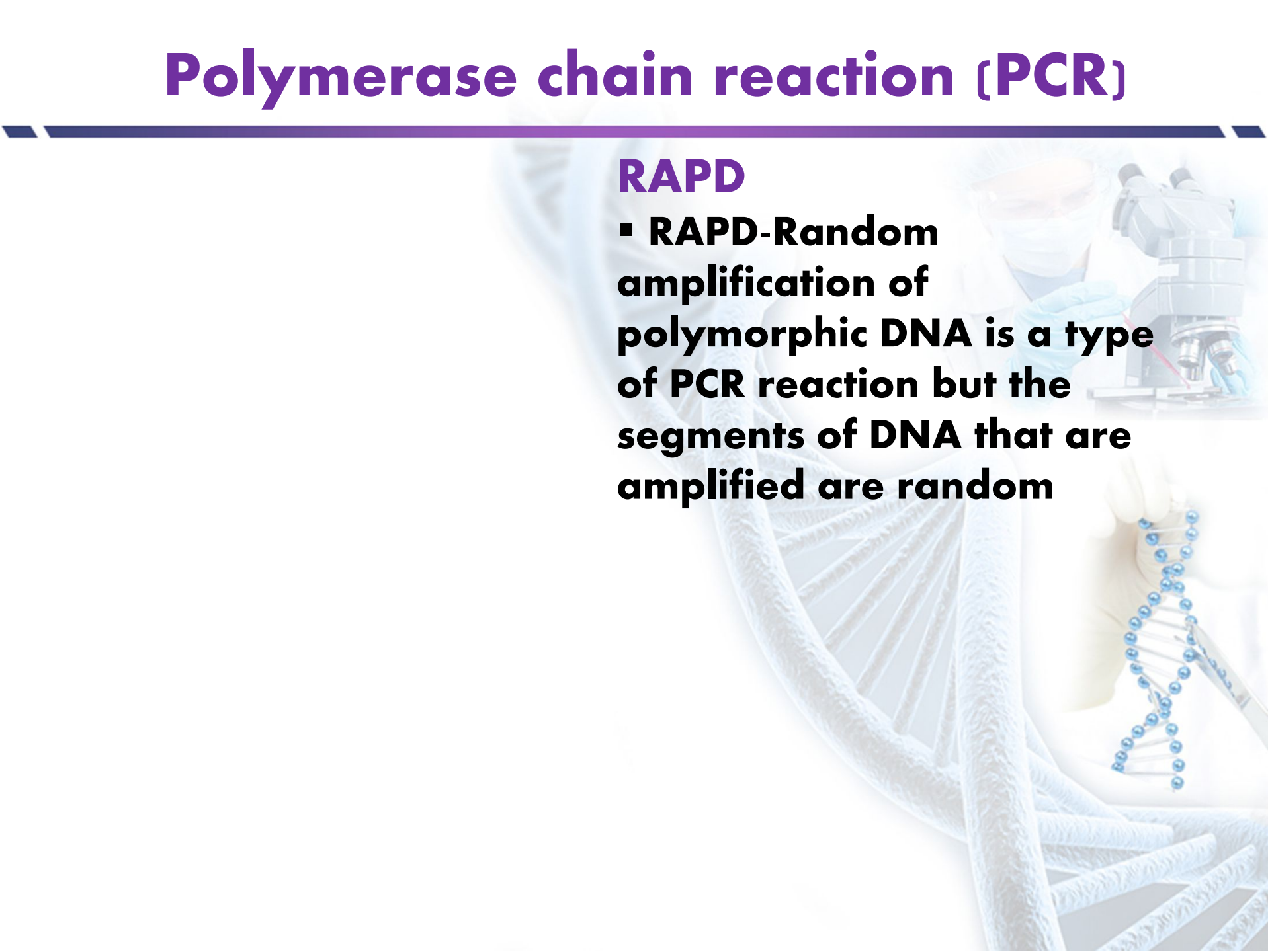
Polymerase chain reaction (PCR)



Polymerase chain reaction (PCR)

RAPD

- **RAPD-Random amplification of polymorphic DNA is a type of PCR reaction but the segments of DNA that are amplified are random**

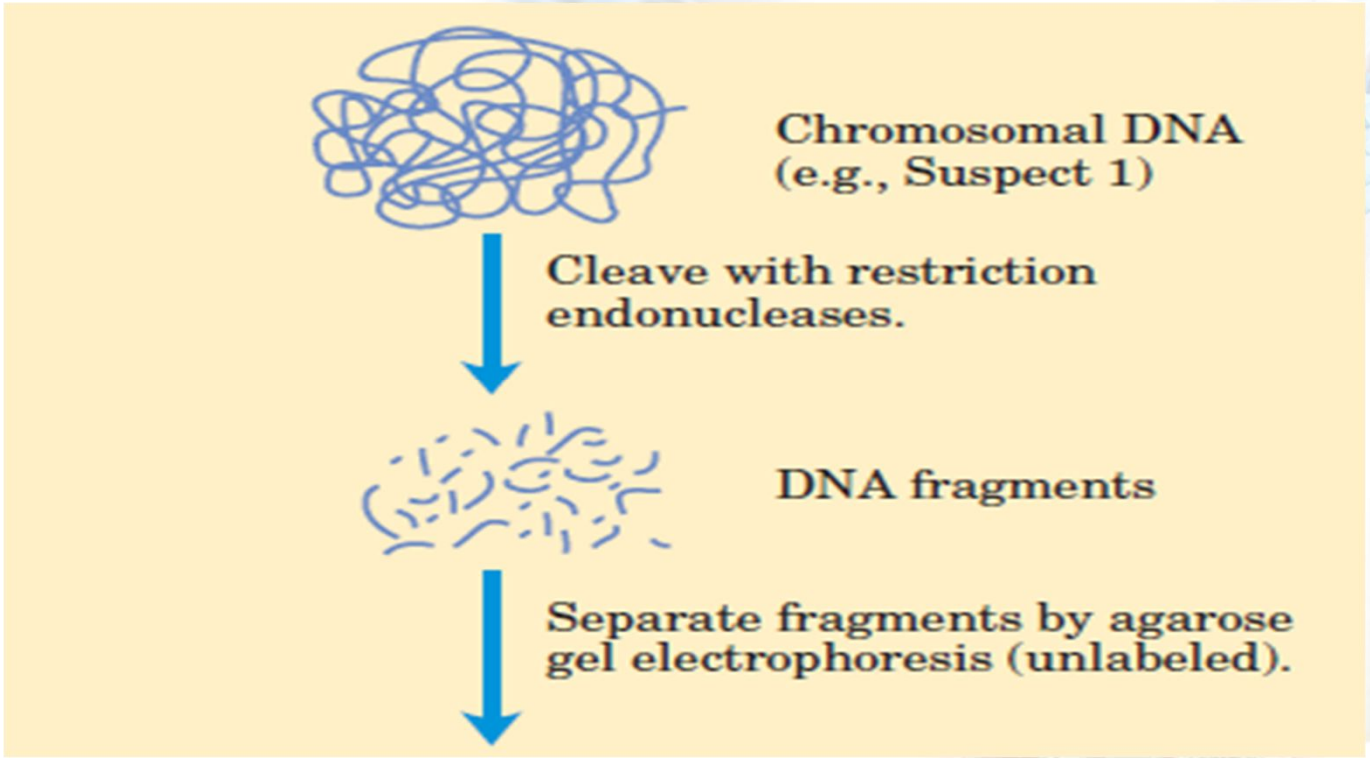


Polymerase chain reaction (PCR)

RFLP

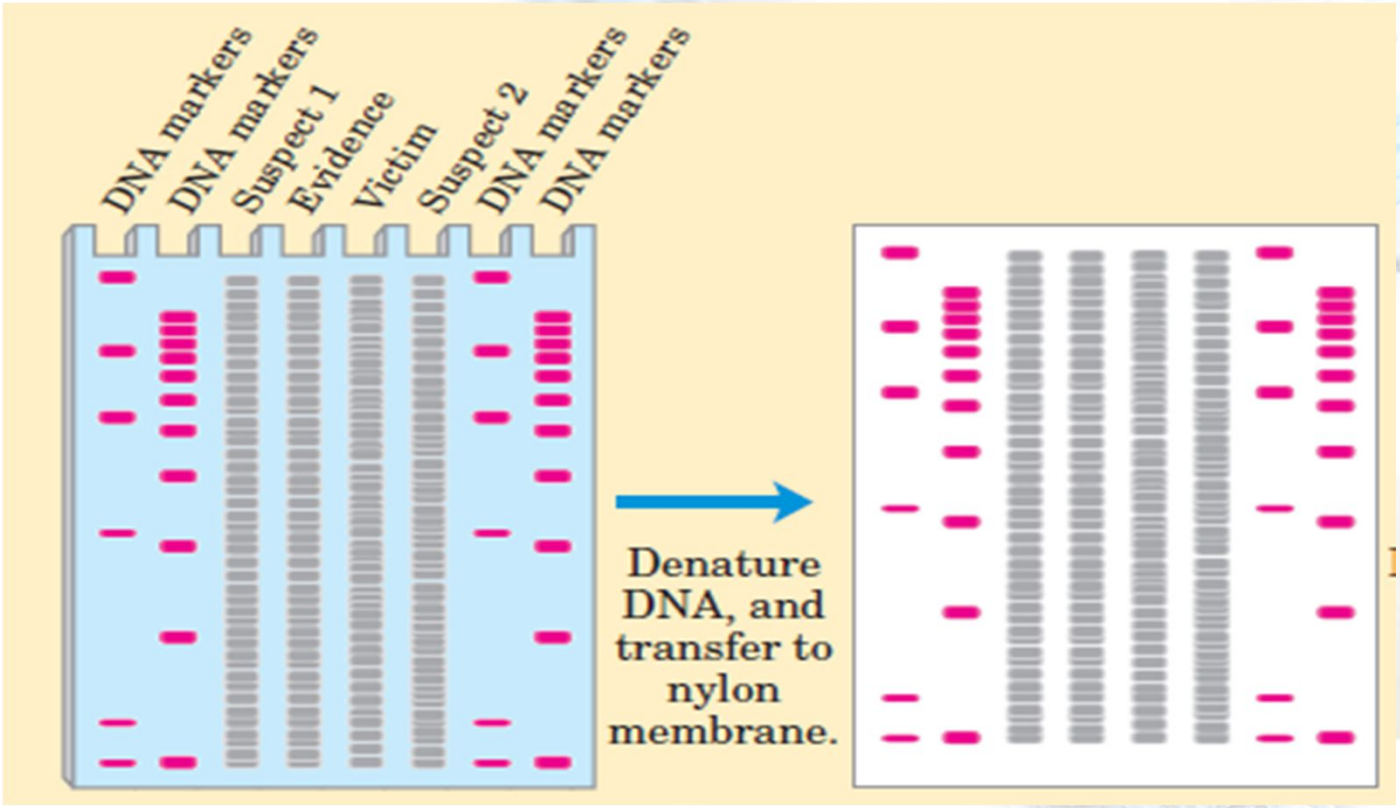
- RFLP-Restriction fragment length polymorphism exploits variations in homologous DNA sequences
- In RFLP, DNA sample is digested by restriction enzymes and separated by gel electrophoresis

Polymerase chain reaction (PCR)



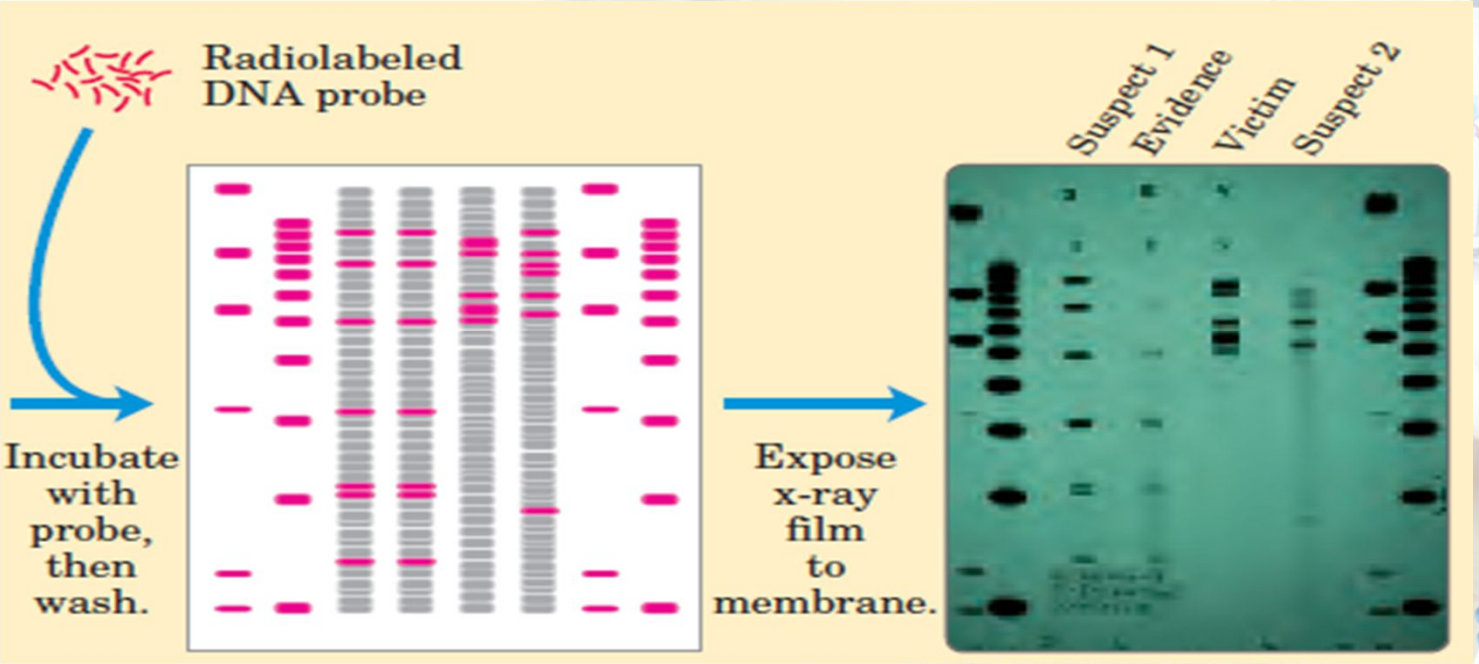
DNA fingerprinting

Polymerase chain reaction (PCR)



DNA fingerprinting

Polymerase chain reaction (PCR)



DNA fingerprinting

Polymerase chain reaction (PCR)

RFLP

- RFLP-Restriction fragment length polymorphism that exploits variations in homologous DNA sequences
- In RFLP, DNA sample is digested by restriction enzymes and separated by gel electrophoresis

Polymerase chain reaction (PCR)

AFLP-PCR

- **AFLPs-Amplified fragment length polymorphisms are differences in restriction fragments length caused by SNPs or INDELS**
- **AFLP is a PCR based tool used in genetic engineering.**

Polymerase chain reaction (PCR)

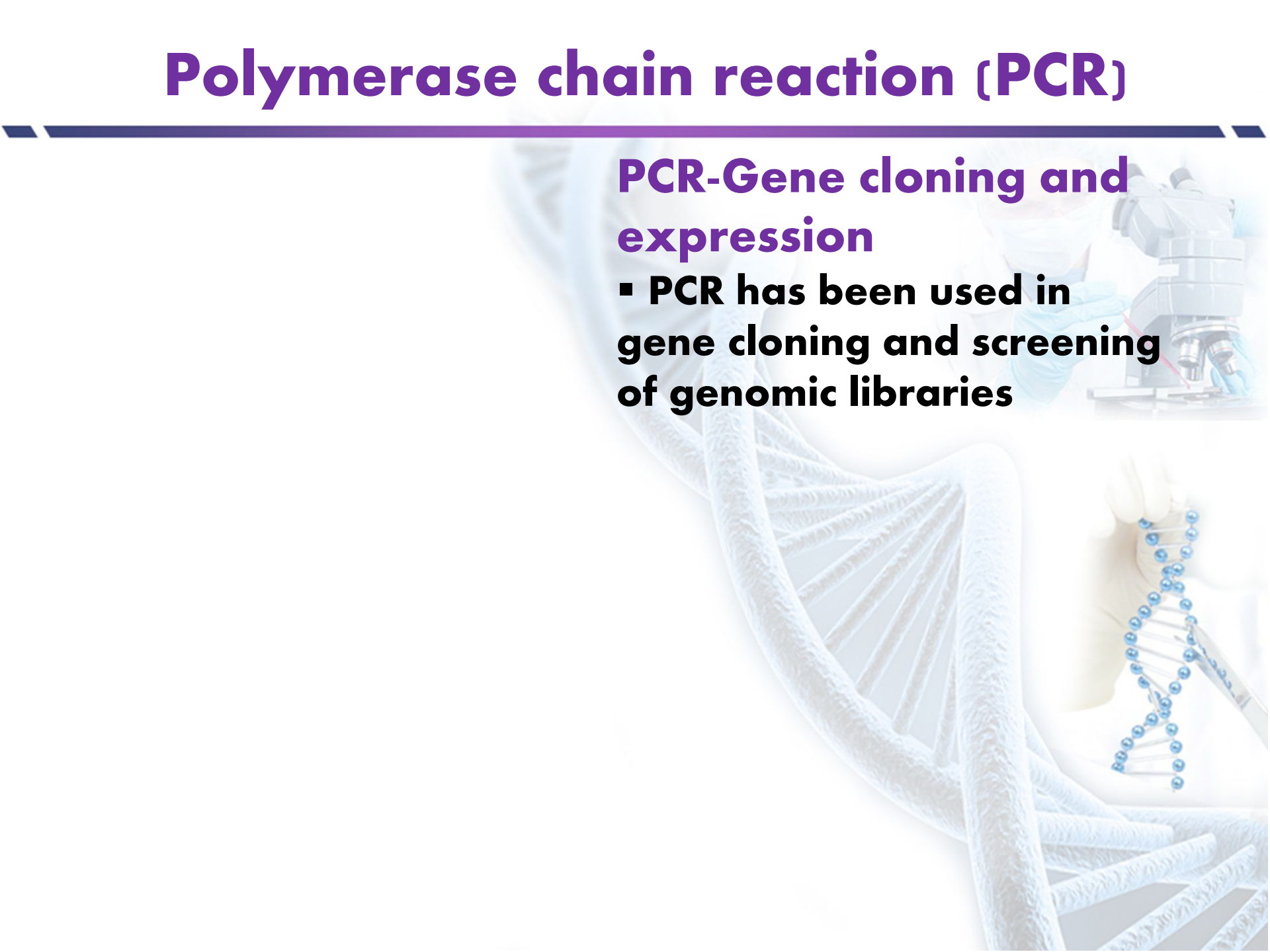
Applications of PCR

- **PCR has widespread applications in various fields of life sciences including genetic engineering, medical, forensic, agriculture, environment etc.**

Polymerase chain reaction (PCR)

PCR-Gene cloning and expression

- **PCR has been used in gene cloning and screening of genomic libraries**



Polymerase chain reaction (PCR)

PCR-Medicine

- PCR has major impact on medicine especially in the field of clinical microbiology or diagnosis
- Molecular tools have also allowed to perform prenatal genetic diagnosis

Polymerase chain reaction (PCR)

PCR-Forensic sciences

- Forensic science is the application of scientific procedures to solve criminal and legal matters
- Molecular methods are used to established the filiations of a person or to obtain evidence from minimal samples of saliva, semen or other tissues

Polymerase chain reaction (PCR)

PCR-DNA profiling

- **DNA profiling or DNA fingerprinting is a forensic technique used to identify individuals by characteristics of their DNA**

Polymerase chain reaction (PCR)

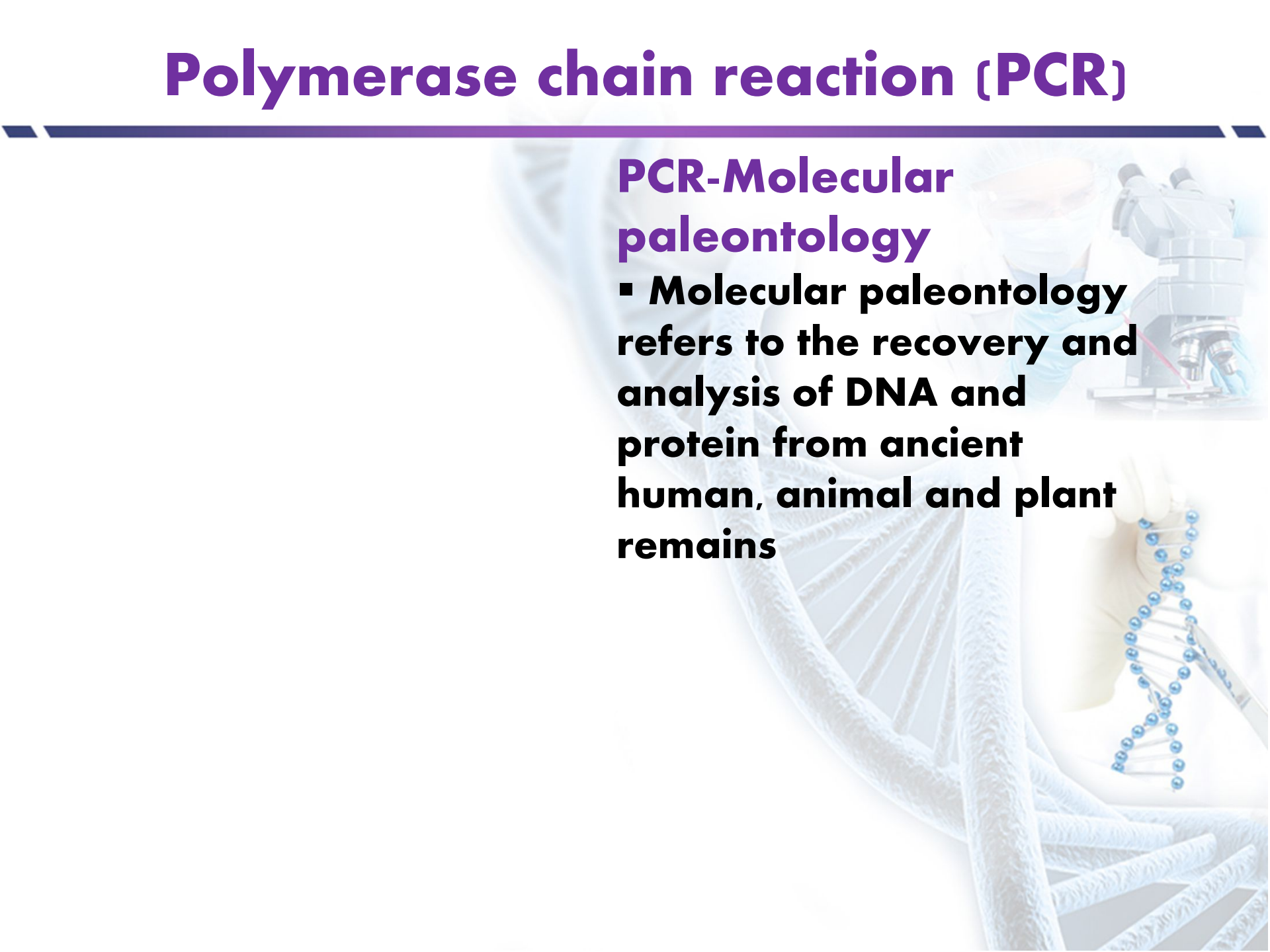
PCR-Agricultural sciences and environment

- PCR has also facilitated research in detection of pathogens in plants, animals and environment

Polymerase chain reaction (PCR)

PCR-Molecular paleontology

- **Molecular paleontology refers to the recovery and analysis of DNA and protein from ancient human, animal and plant remains**



Mapping and sequencing genomes

Genome mapping

- Assigning of a specific gene to particular region of a chromosome and determining the location of and relative distances between genes on the chromosome
- Types of maps:
    1. Linkage maps
    2. Physical maps

Mapping and sequencing genomes

Markers for genome mapping

- Any identifiable feature of the genome can serve as a marker in mapping
- The landmarks on genome map might include short DNA sequences, regulatory sites and genes themselves

Mapping and sequencing genomes

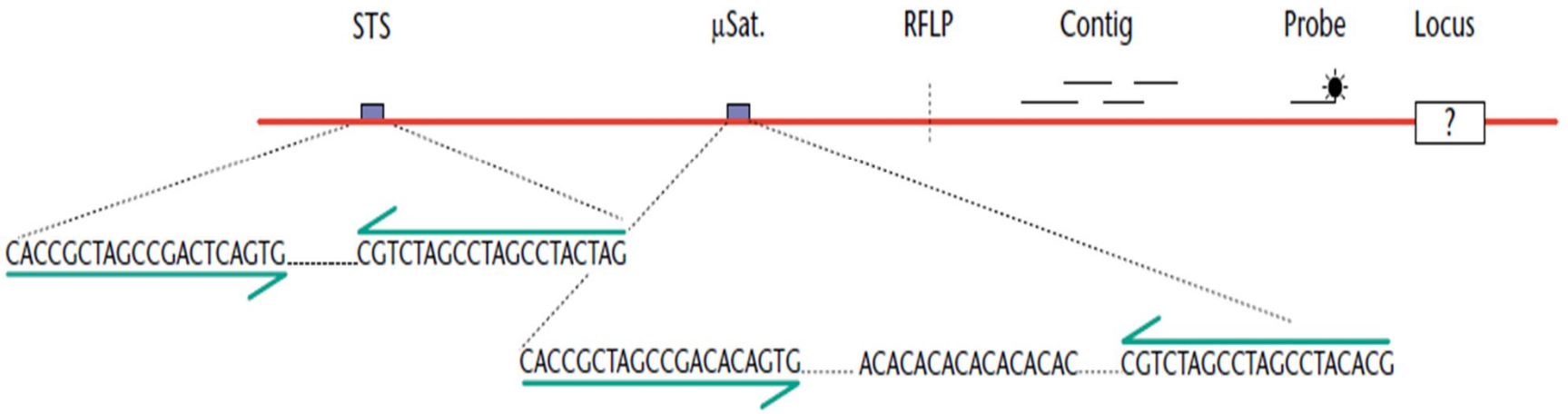


Figure. Markers used in genome mapping

(Paul H Dear, 2001, Encyclopedia of life sciences)

Methods for genome mapping

Genetic linkage mapping

- Genetic linkage maps illustrate the order of genes or genetic markers on a chromosome and the relative distances between those genes

Methods for genome mapping

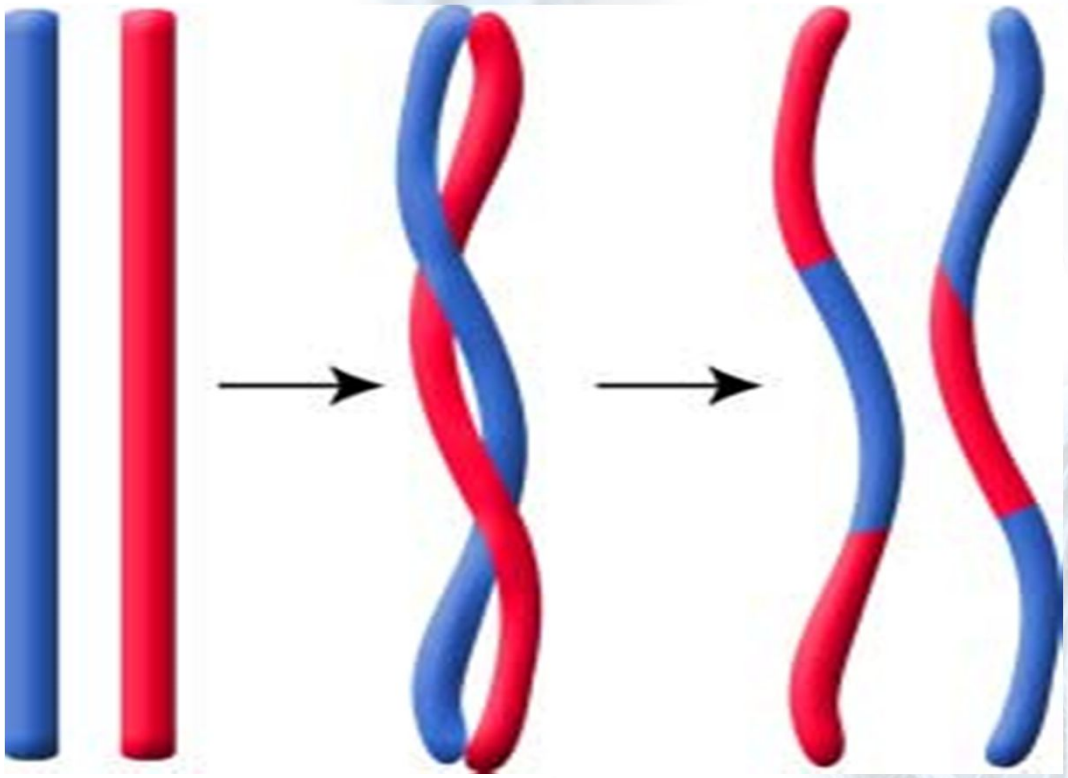


Figure. Crossing over in chromosomes
Source: Genomic News Network (GNN)

Methods for genome mapping

Genetic linkage mapping

- Genetic linkage maps illustrate the order of genes or genetic markers on a chromosome and the relative distances between those genes

Methods for genome mapping

Physical mapping

- **Physical maps give the DNA base pair distances from one landmark to another**

Methods for genome mapping

Physical versus linkage maps

- Chromosomal maps of bacteria has been obtained by combining the mapping techniques of interrupted mating, recombination mapping, transformation and transduction

Methods for genome mapping

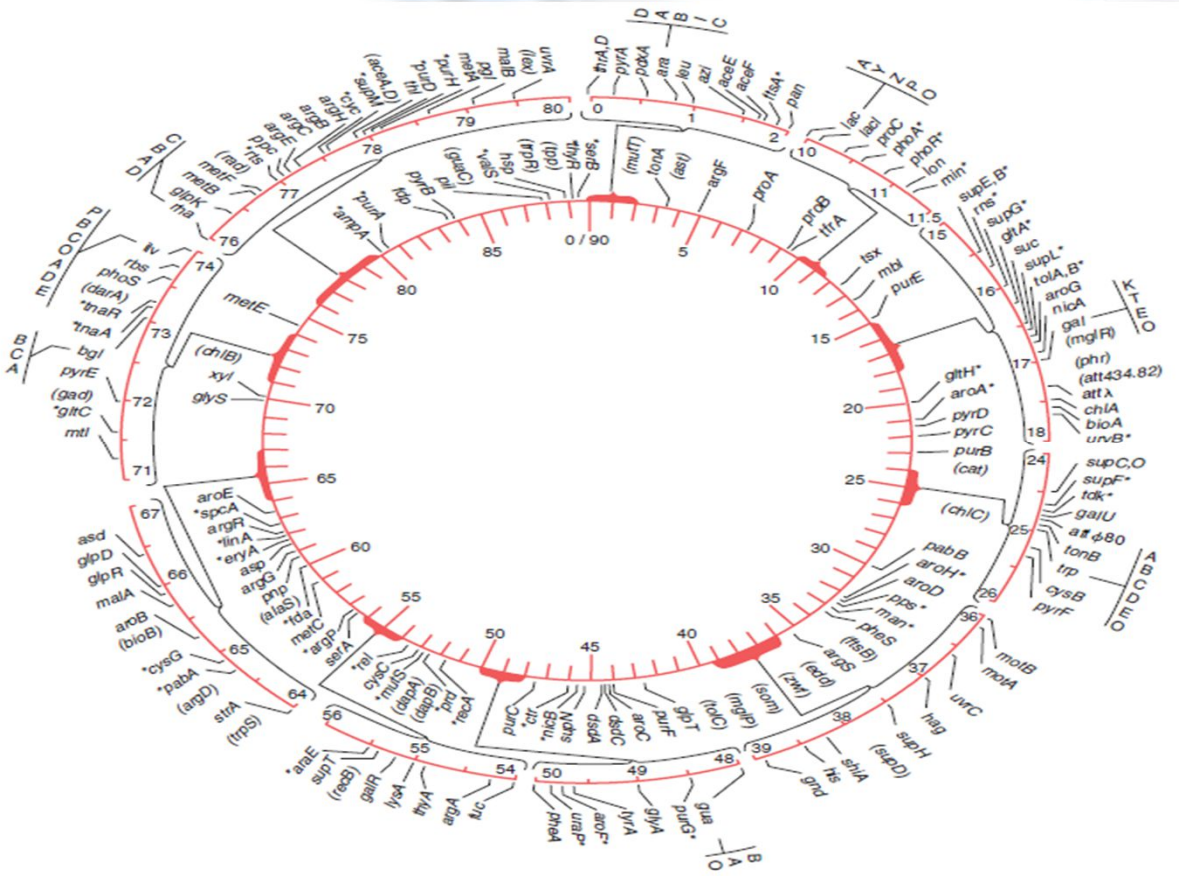


Figure. The 1963 map *of E. coli*

Methods for genome mapping

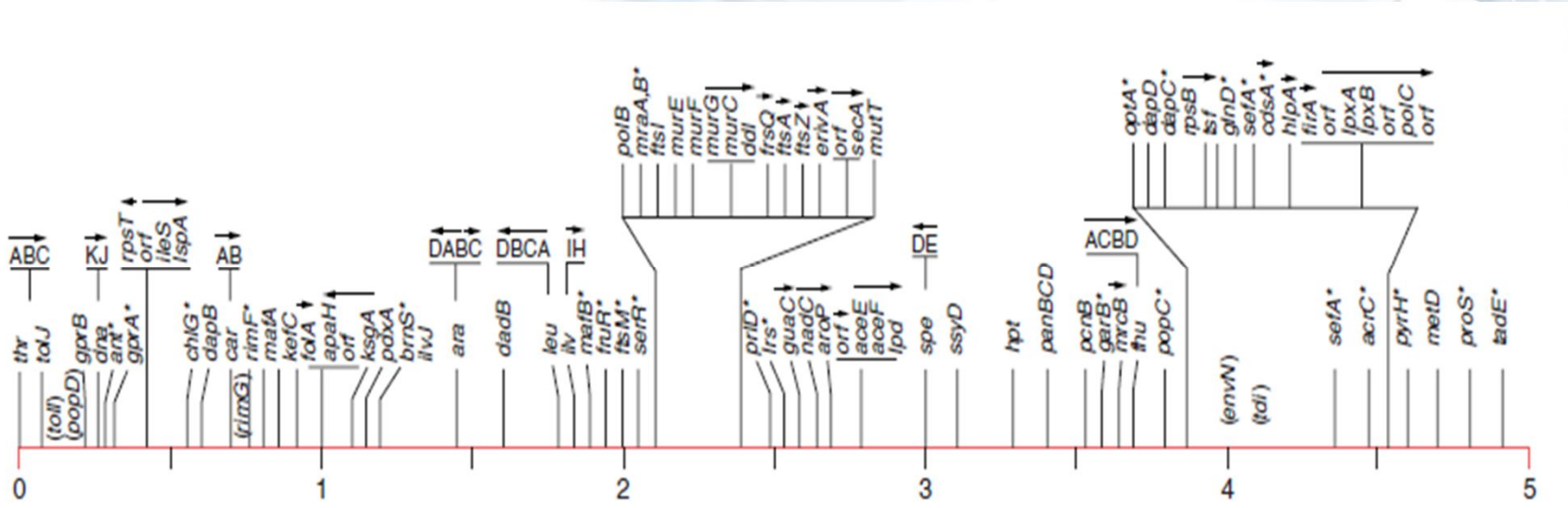


Figure. Linear scale drawing of a 5-minute section of the 100-minute 1990 *E. coli* linkage map

Methods for genome mapping

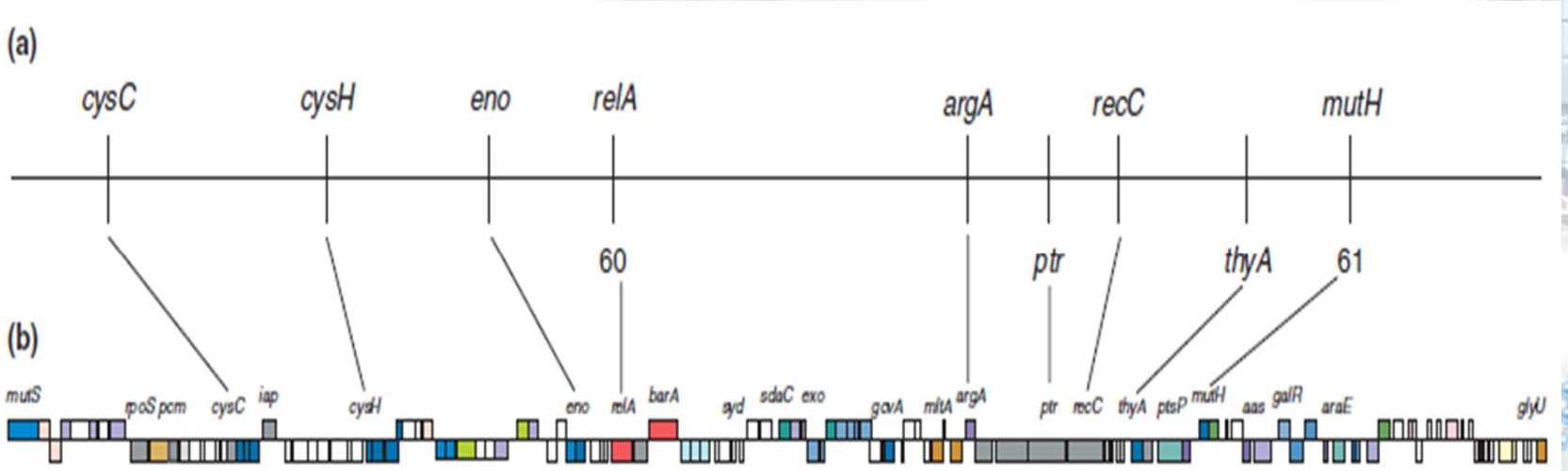


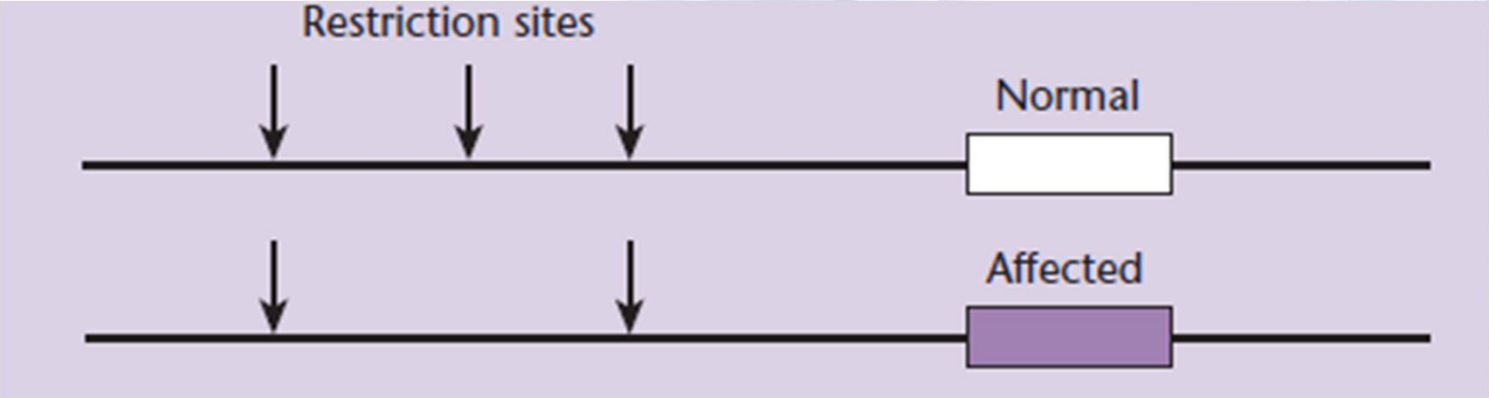
Figure. Correlation of the genetic and physical maps

Methods for genome mapping

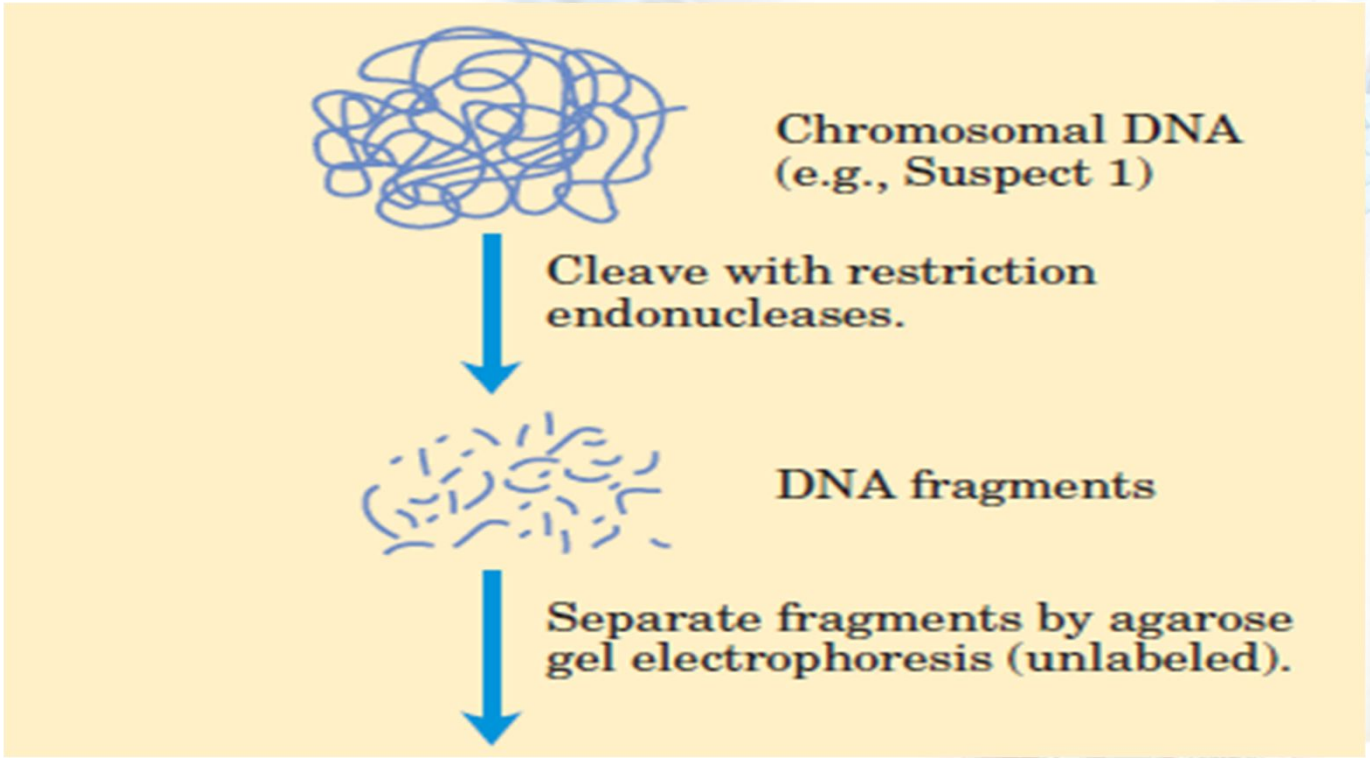
The use of RFLPs in physical maps

- **RFLP-Restriction fragment length polymorphism exploits variations in homologous DNA sequences**
- **Botstein et al. (1980) were the first to recognize that DNA probes that target RFLPs can be used for mapping**

Methods for genome mapping

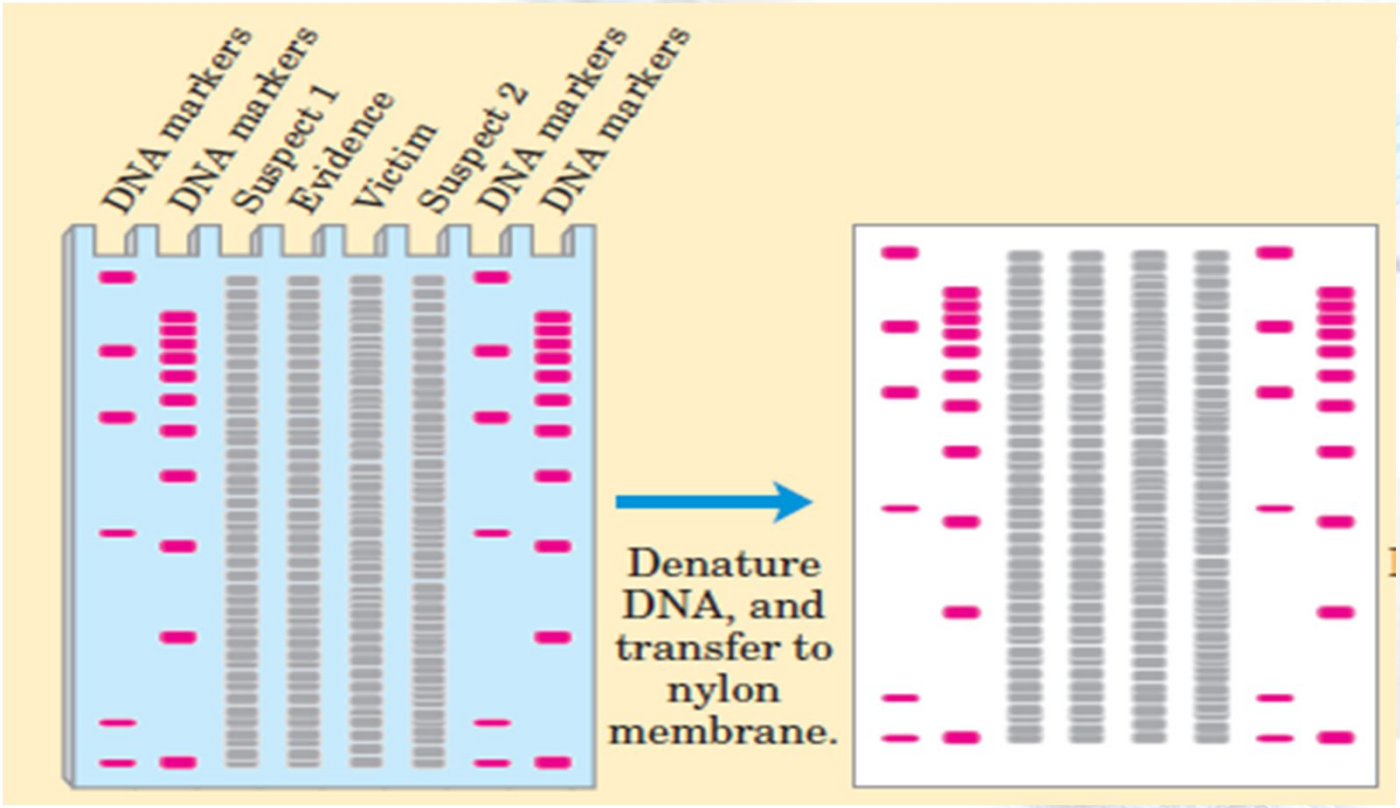


Polymerase chain reaction (PCR)



DNA fingerprinting

Polymerase chain reaction (PCR)



DNA fingerprinting

Methods for genome mapping

STS in physical maps

- **Sequence-tagged sites (STS) are more convenient markers than RFLPs because they do not use southern blotting**

Methods for genome mapping

Figure. Examples of a sequence-tagged site (STS)

Methods for genome mapping

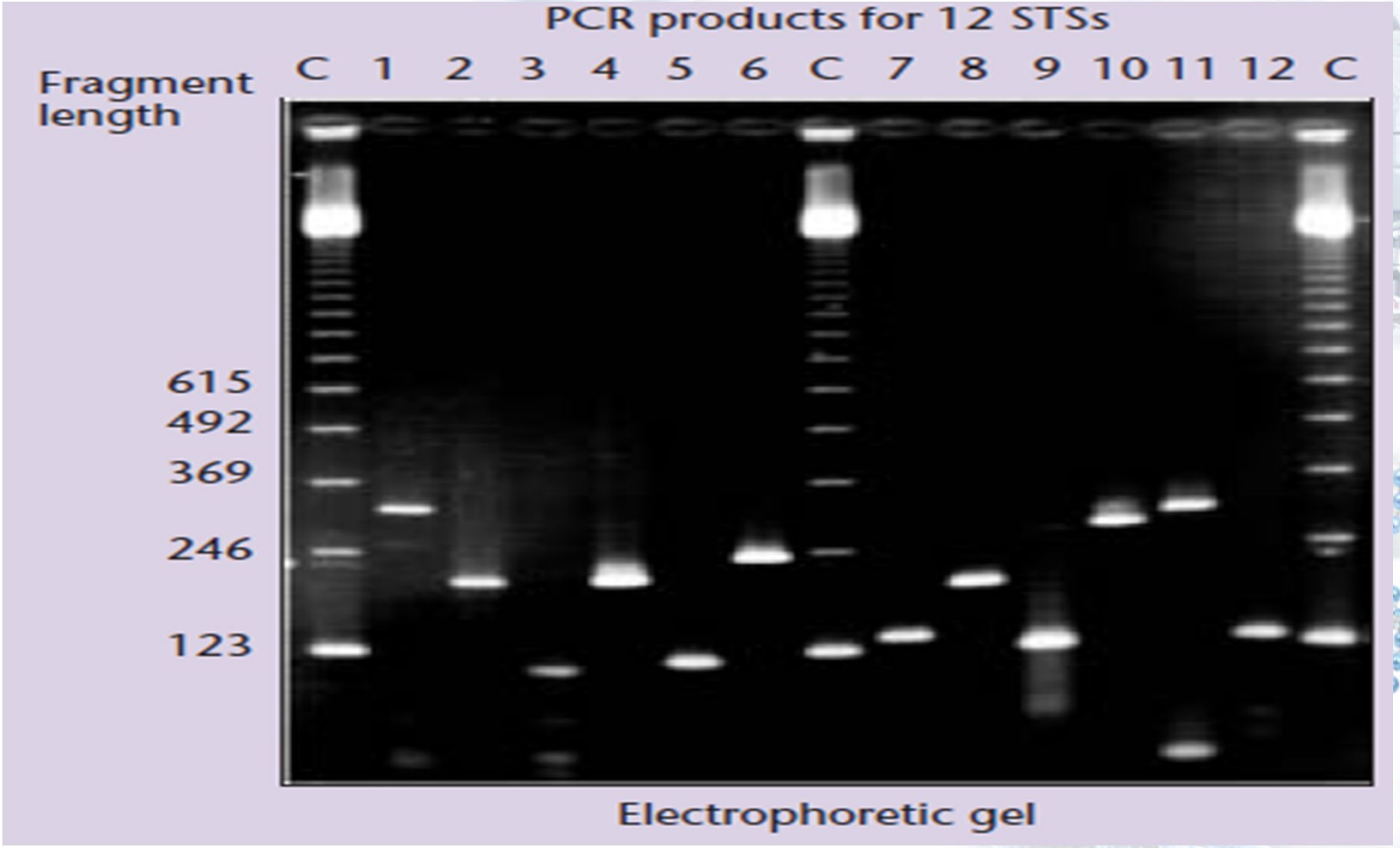


Figure. Confirmation that an STS is a unique sequence on the genome

Methods for genome mapping

STS in physical maps

- **Sequence-tagged sites (STS) are more convenient markers than RFLPs because they do not use southern blotting**

Methods for genome mapping

SNPs as physical markers

- Single nucleotide polymorphisms (SNPs) are single base-pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population
- SNPs probably are the most important sequence markers for physical mapping of genomes

Methods for genome mapping

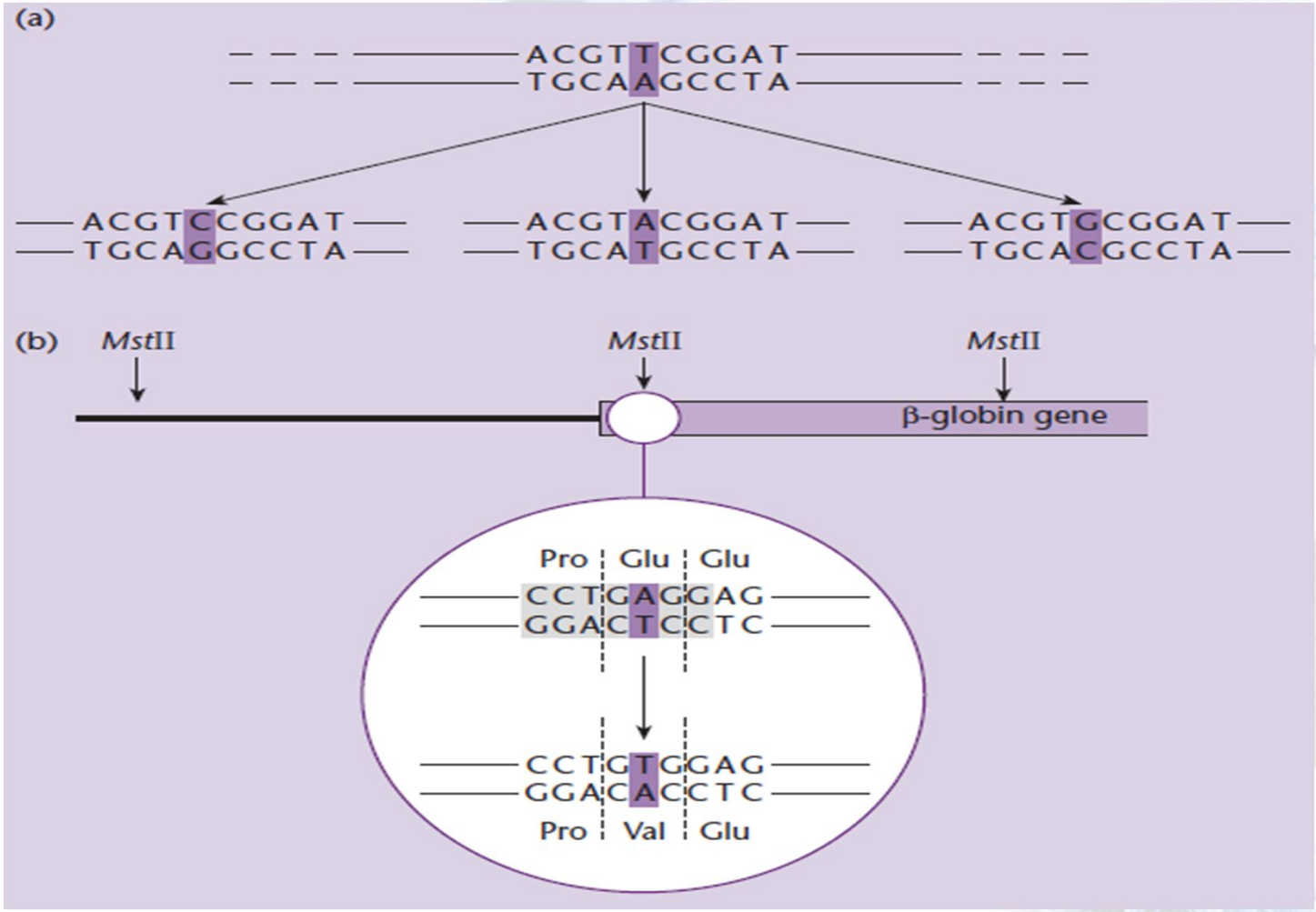


Figure. Examples of single nucleotide polymorphisms

Methods for genome mapping

SNPs as physical markers

- Different methods to detect SNPs have been developed that based on enzymatic, electrophoretic, solid phase or chromatographic analysis

Methods for genome mapping

Polymorphic DNA detection in the absence of sequence information

- **Polymorphic DNA can be detected by amplification in the absence of the target DNA sequence information used to generate STSs**

Methods for genome mapping

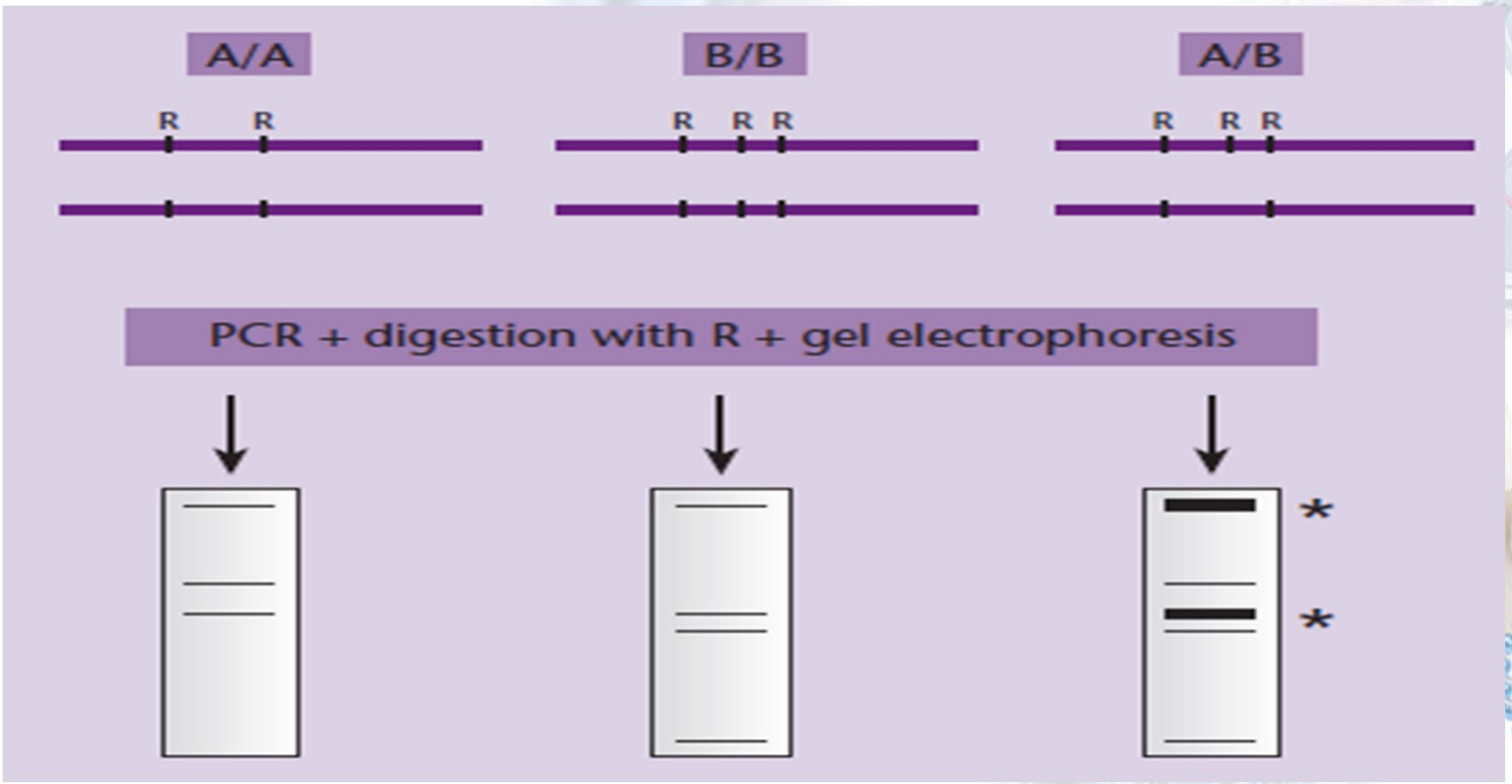


Figure. Generation and visualization of CAPS markers

Methods for genome mapping

AFLPs detection in the absence of sequence information

- **Amplified fragment length polymorphisms (AFLP) is a diagnostic fingerprinting technique that detects genomic restriction fragments and in that respect resembles the RFLP technique**

Methods for genome mapping

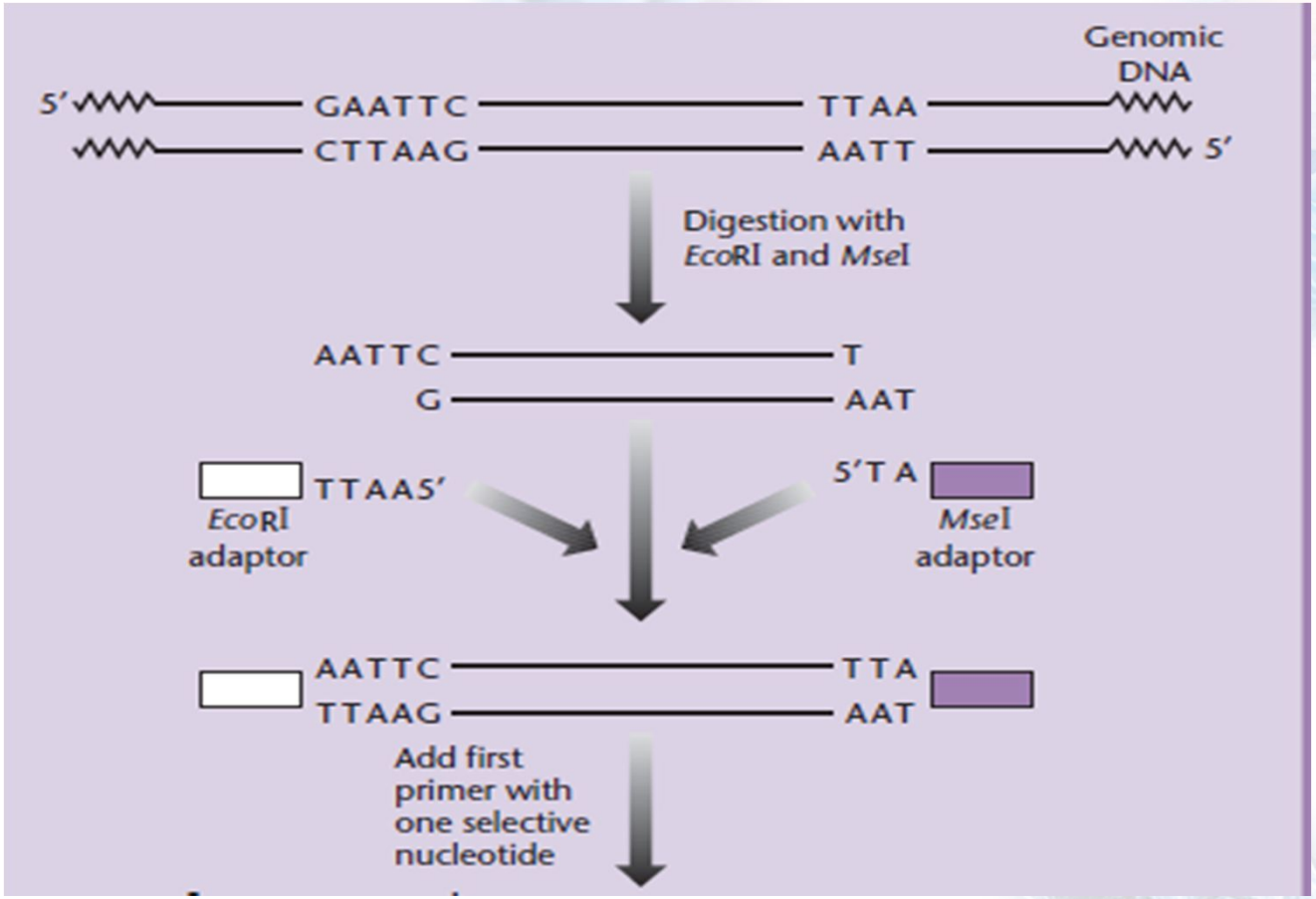


Figure. Principle of the AFLP method

Methods for genome mapping

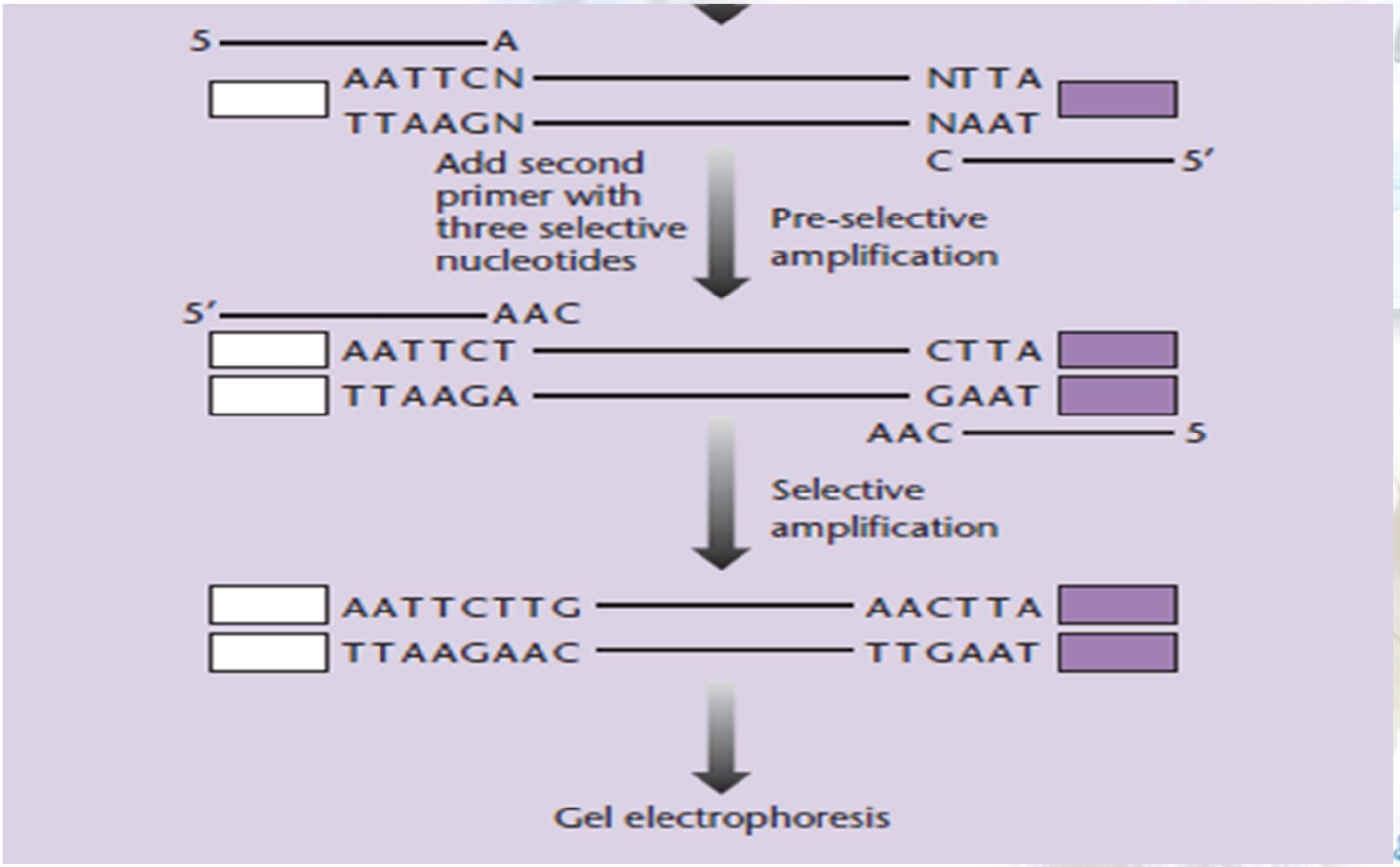


Figure. Principle of the AFLP method

Methods for genome mapping

AFLPs detection in the absence of sequence information

- **Amplified fragment length polymorphisms (AFLP) is a diagnostic fingerprinting technique that detects genomic restriction fragments and in that respect resembles the RFLP technique**

Methods for genome mapping

Fluorescence in situ hybridization (FISH)

- **FISH is a cytogenetic method that used fluorescent probes that bind to only those parts of the chromosomes with a high degree of sequence complementarity**

Methods for genome mapping

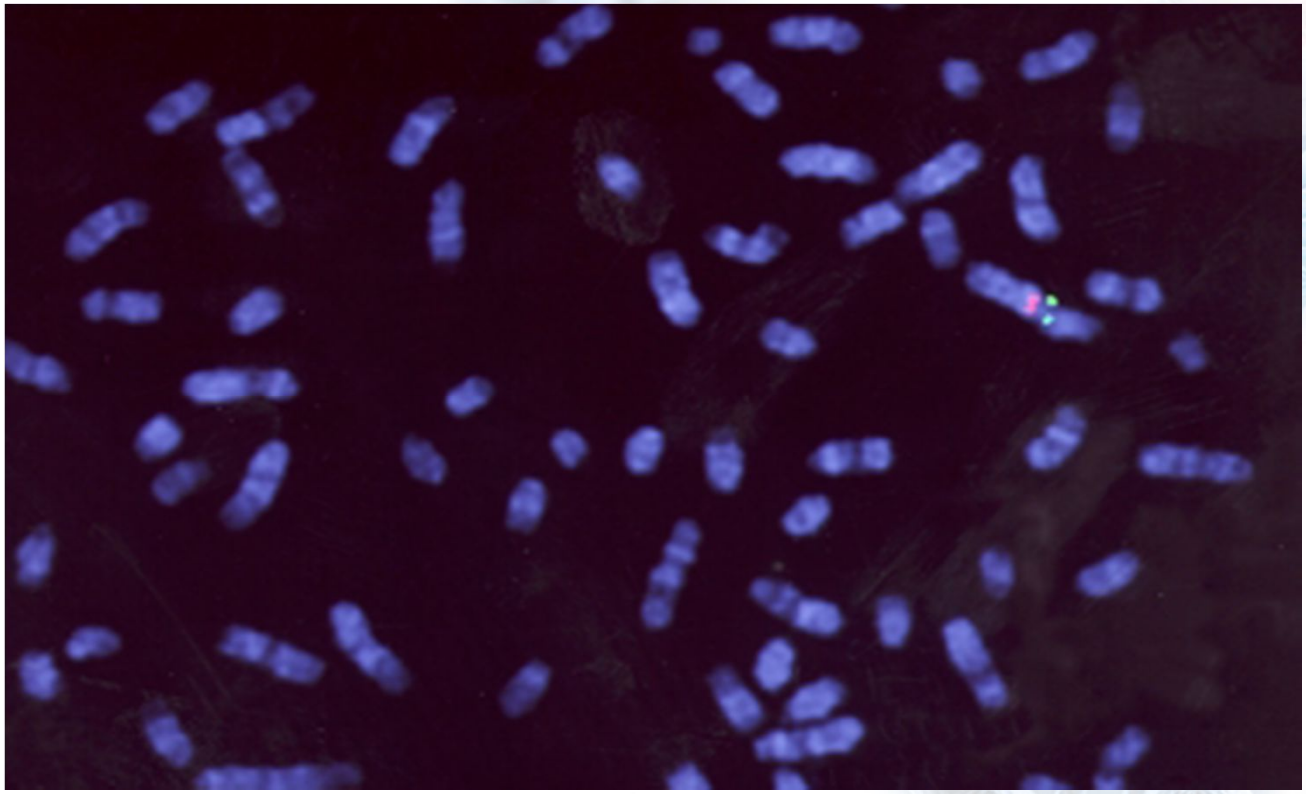


Figure. Dual-color fluorescence in situ hybridization

Methods for genome mapping

Radiation hybrid (RH) mapping

- RH mapping used X-ray breakage of chromosomes to determine the distances between DNA markers as well as their order on the chromosome

Methods for genome mapping

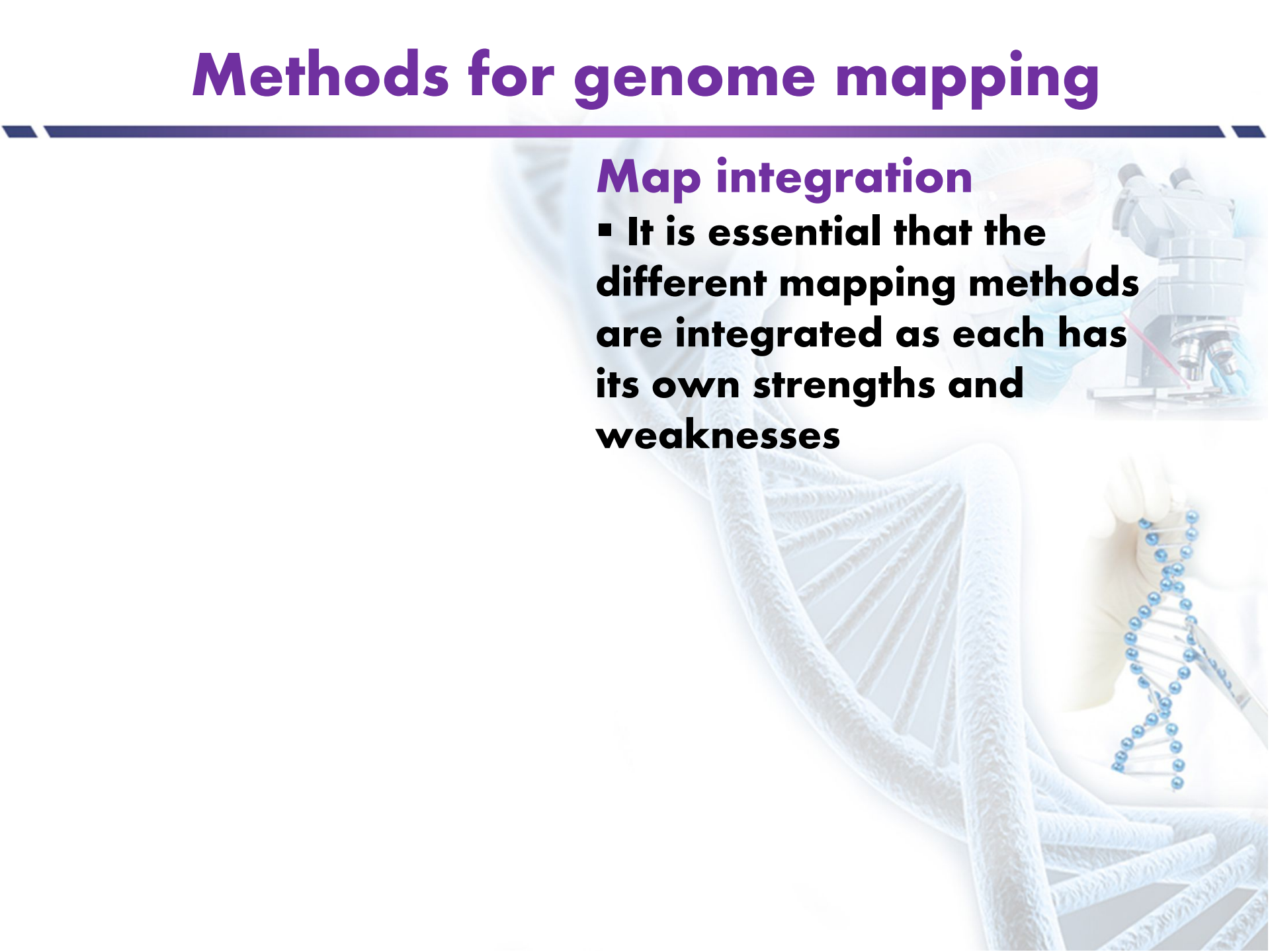
Happy mapping

- **RH mapping have a few problems that arise from the *in vivo* step. Happy mapping overcomes these limitations, being an entirely *in vitro* technique**

Methods for genome mapping

Map integration

- It is essential that the different mapping methods are integrated as each has its own strengths and weaknesses



Mapping and sequencing genomes

Sequencing genome

- The process of determining the exact order of nucleotides within a DNA or RNA molecule is called sequencing
- In 1970s, DNA sequencing was a very tedious process which involved determining only a few hundred nucleotides at a time

Mapping and sequencing genomes

High-throughput sequencing

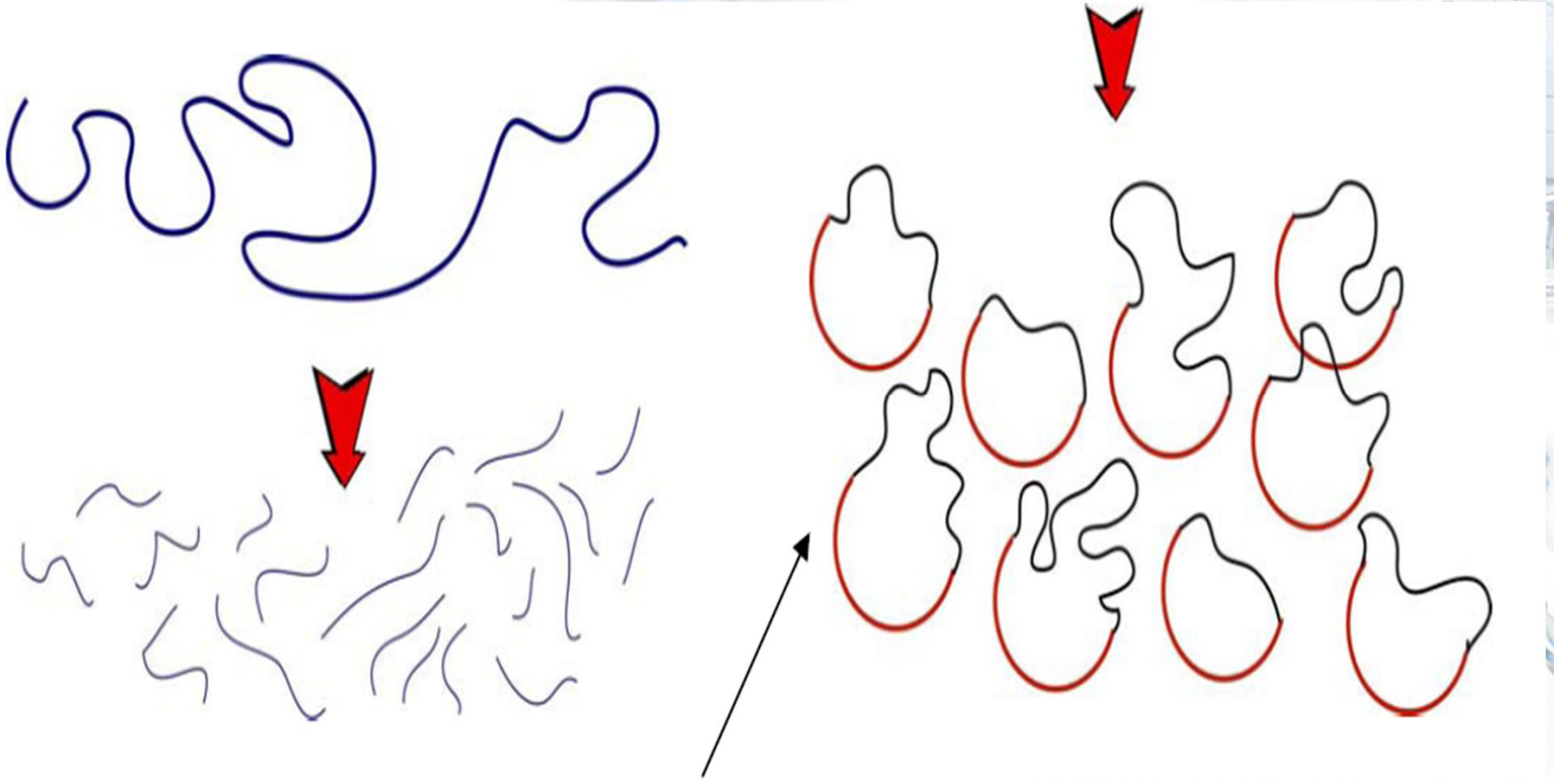
- Recent advances in DNA sequencing have make it possible to sequence data very rapidly and at a substantially lower cost
- It mainly describe a number of different modern sequencing technologies to sequence genome

Mapping and sequencing genomes

Shotgun sequencing

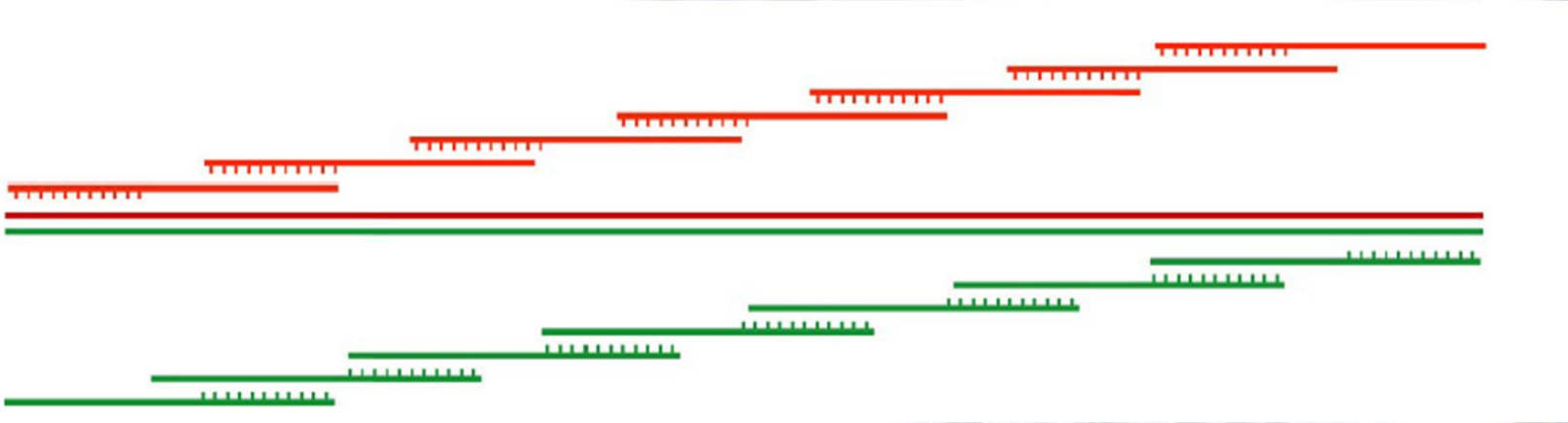
- Shotgun sequencing is used for sequencing long DNA strands
- In this strategy, DNA shredded into smaller fragments that can be sequenced individually and then reassembled into original order

Mapping and sequencing genomes



Shotgun library

Mapping and sequencing genomes

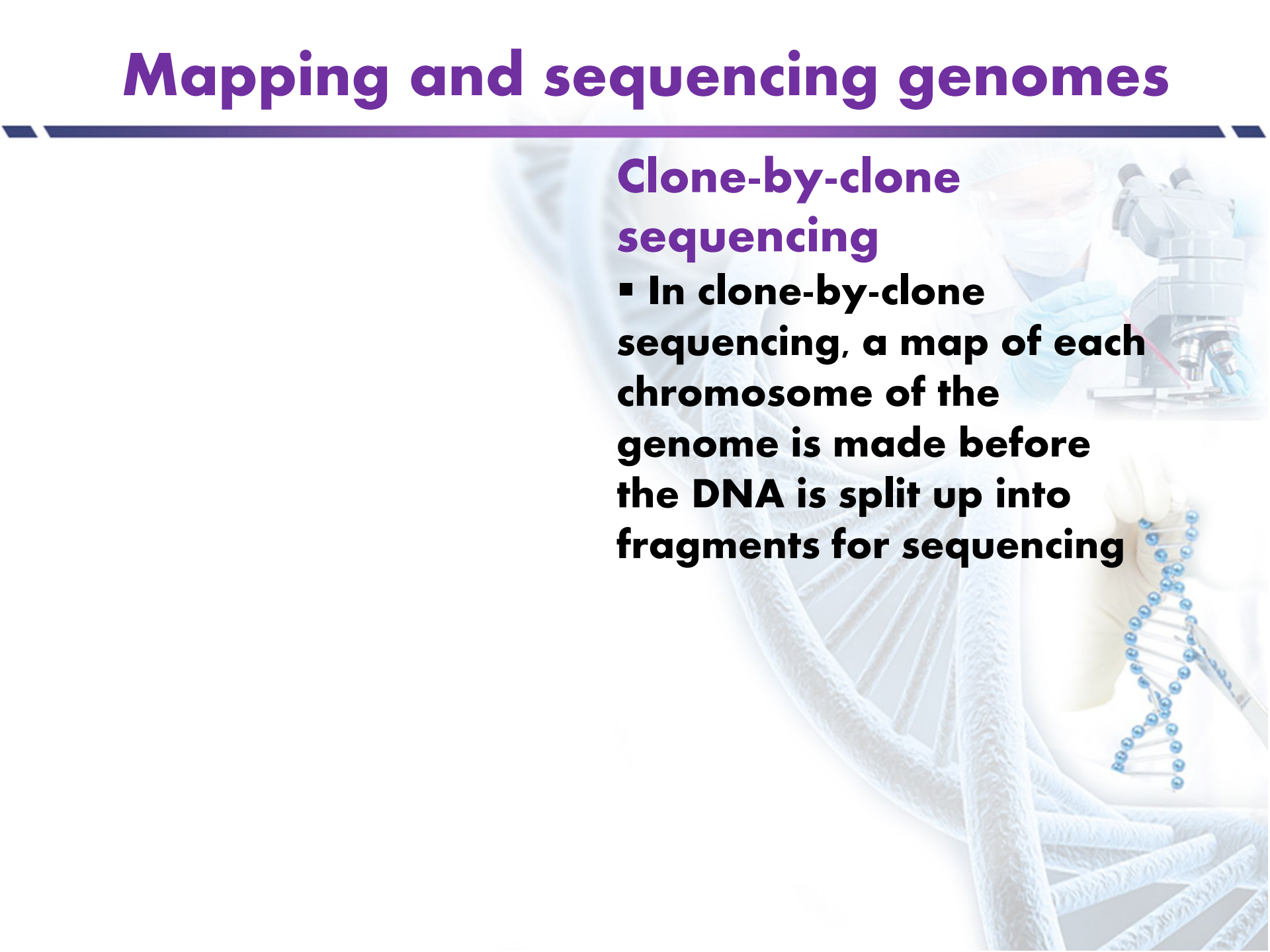


Shotgun library

Mapping and sequencing genomes

Clone-by-clone sequencing

- In clone-by-clone sequencing, a map of each chromosome of the genome is made before the DNA is split up into fragments for sequencing



Comparative genomics

Orthologs and paralogs

- **Orthologs are homologous genes in different organisms that encode proteins with the same function and which have evolved by direct vertical descent**

Comparative genomics

Comparative genomics of bacteria

- Genome of bacteria may be of variable in size. For example, it may vary from 0.49 Mb (*Nanoarchaeum equitans*) to 9.1 Mb (*Bradyrhizobium japonicum* or *Streptomyces*)

Comparative genomics

Comparative genomics of organelles

- Mitochondrial genomes exhibit an amazing structural diversity
- Gene transfer has occurred between mtDNA and nuclear DNA

Comparative genomics

Comparative genomics of eukaryotes

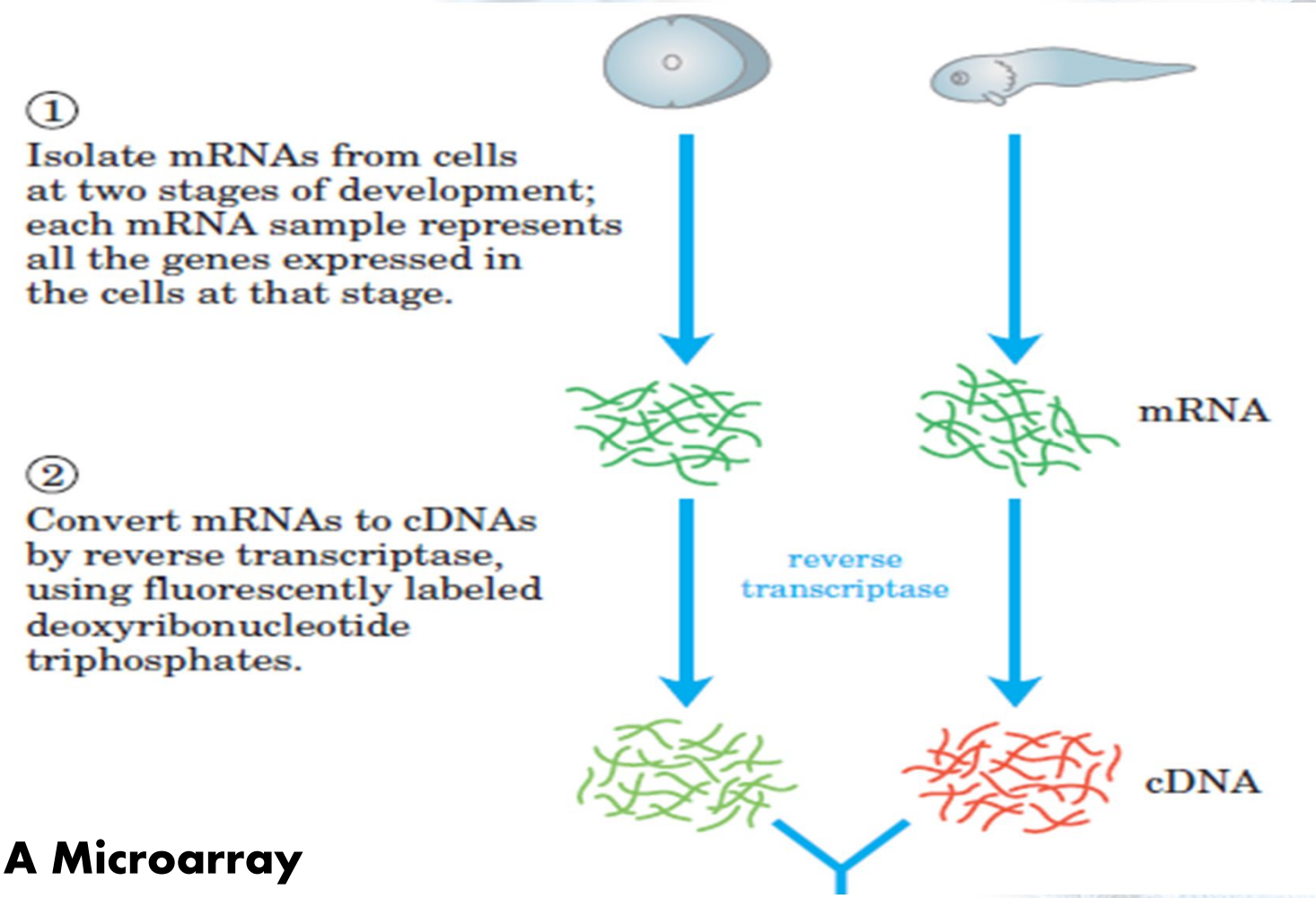
- The minimal eukaryotic genome is smaller than many bacterial genomes
- Comparative genomics can be used to identify genes and regulatory elements
- It gives insight into the evolution of key proteins

Functional genomics and proteomics

DNA Microarrays

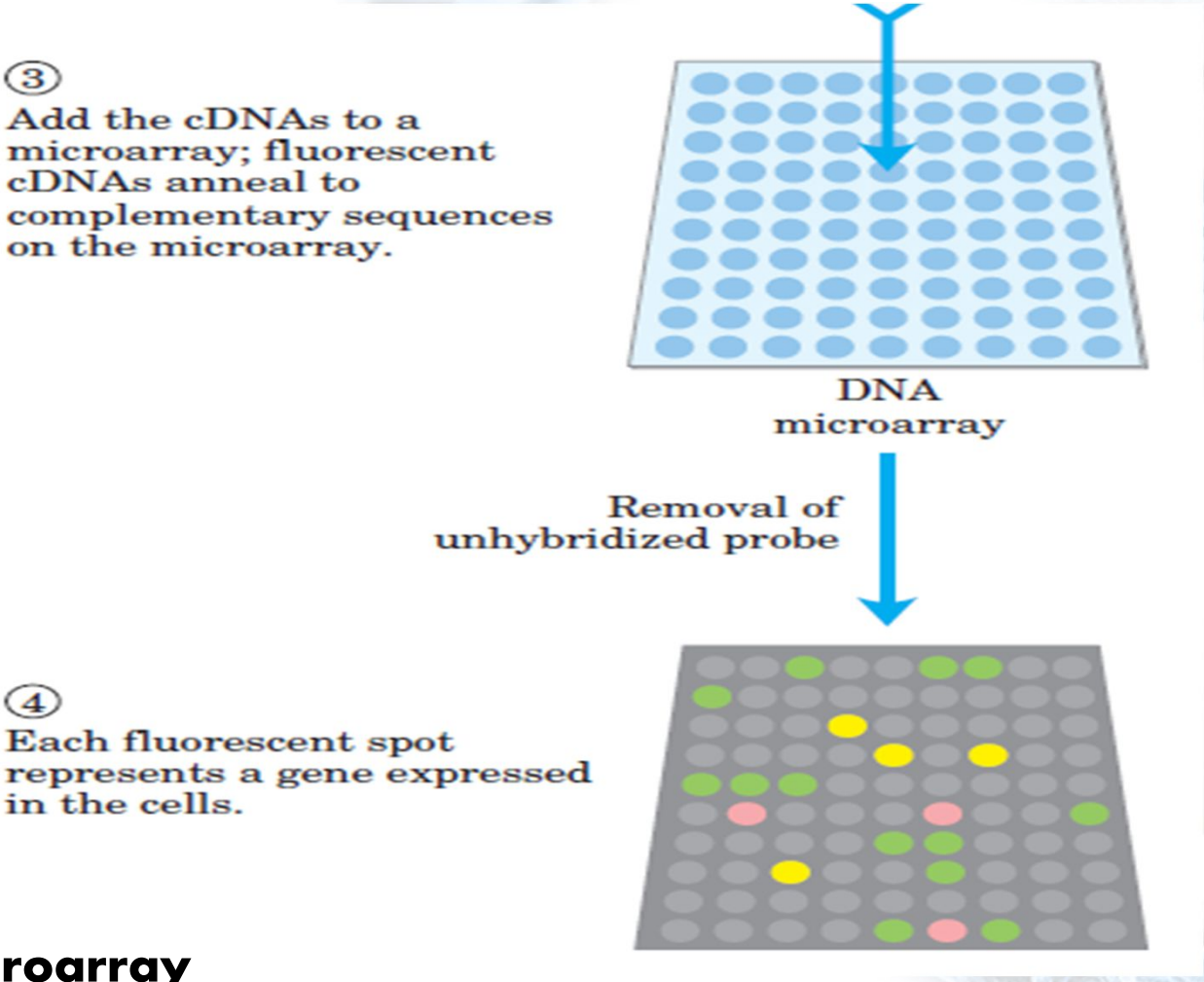
- Microarrays are used to study the expression of many genes at once. It involves the immobilization of thousands of gene sequences on glass slide that are then detected with complementary base pairing between the sample and the gene sequences on the chip

Functional genomics and proteomics



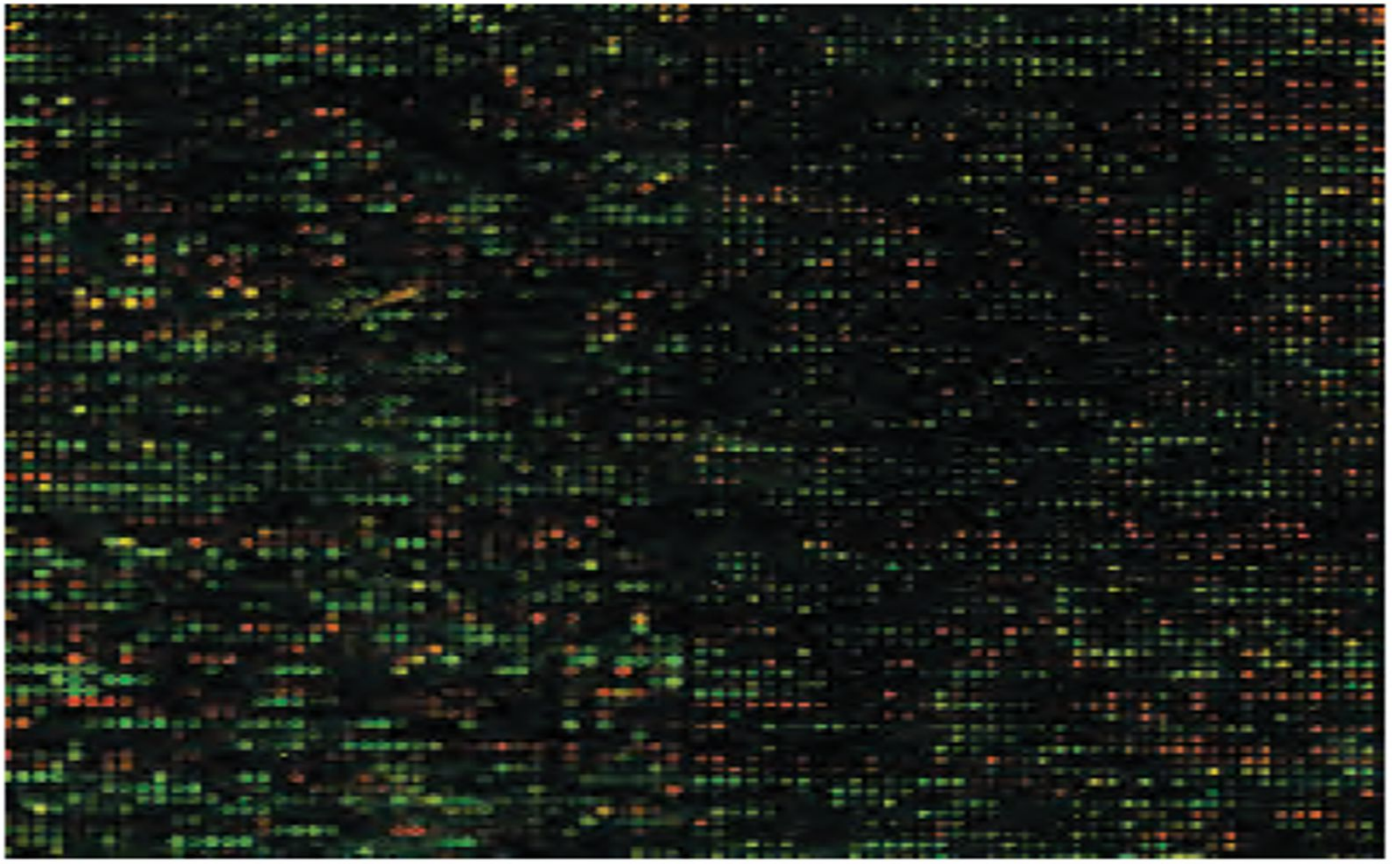
DNA Microarray

Functional genomics and proteomics



DNA Microarray

Functional genomics and proteomics



Enlarged image of a DNA Microarray

Functional genomics and proteomics

DNA Microarrays

- **Two major types of DNA array are used in expression analysis**

i). Spotted DNA arrays

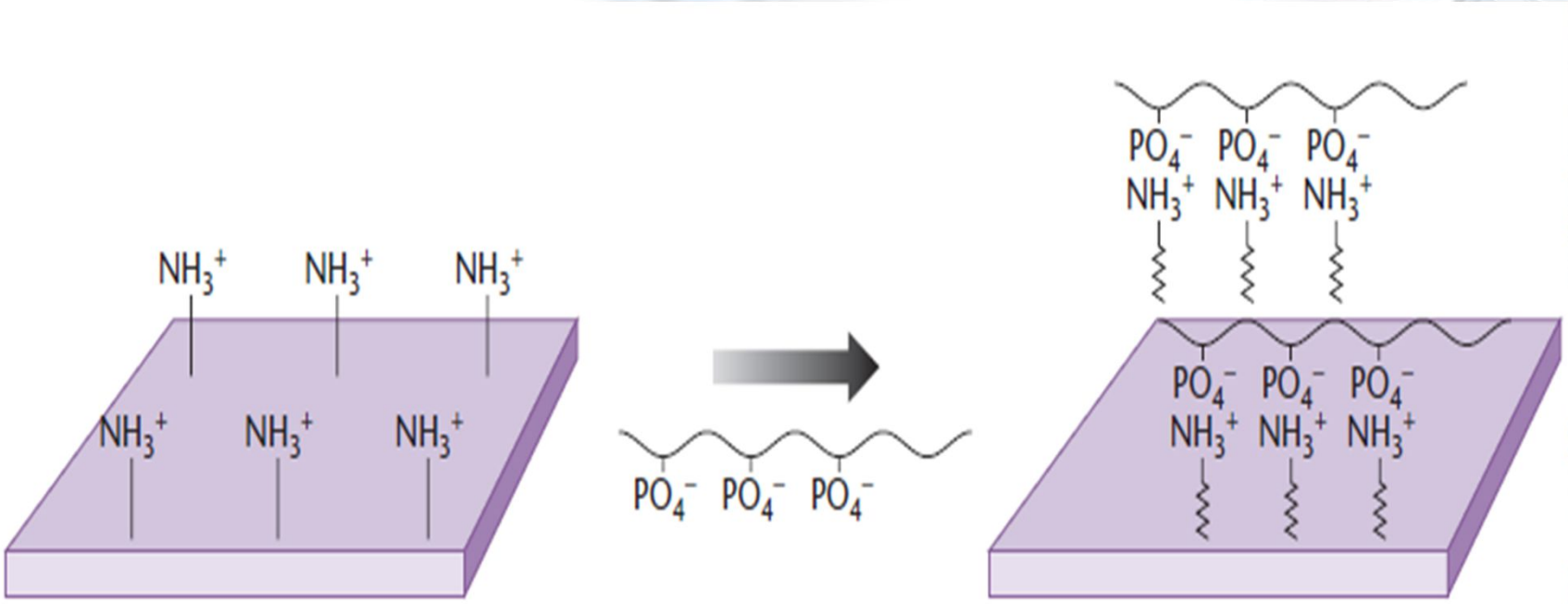
ii). Printed oligonucleotide chips

Functional genomics and proteomics

Spotted DNA arrays

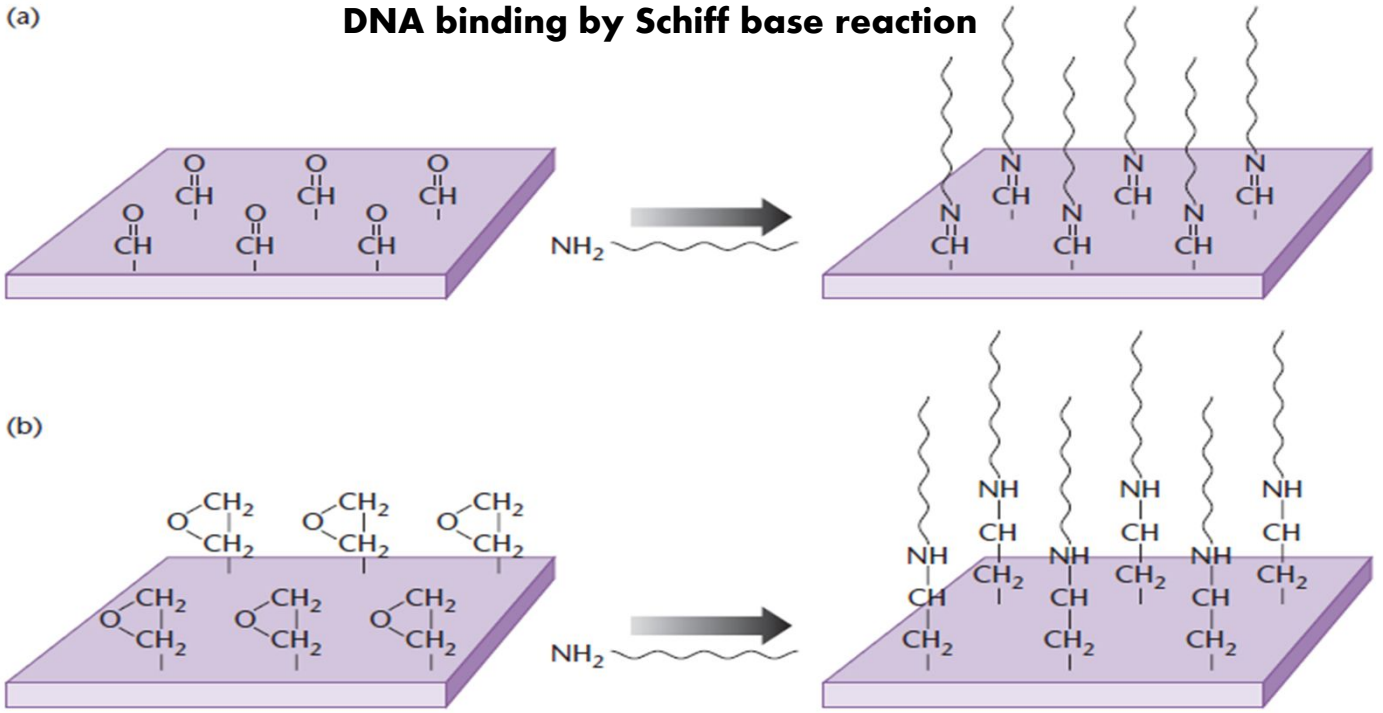
- **Spotted DNA arrays are produced by printing DNA samples on treated microscope slides. It is made by transferring or spotting DNA clones or PCR product individually onto a solid support where they are immobilized**

Functional genomics and proteomics



DNA binding by electrostatic interactions

Functional genomics and proteomics



b)- Binding of DNA molecules to epoxy-derivatized surfaces

Functional genomics and proteomics

Spotted DNA arrays

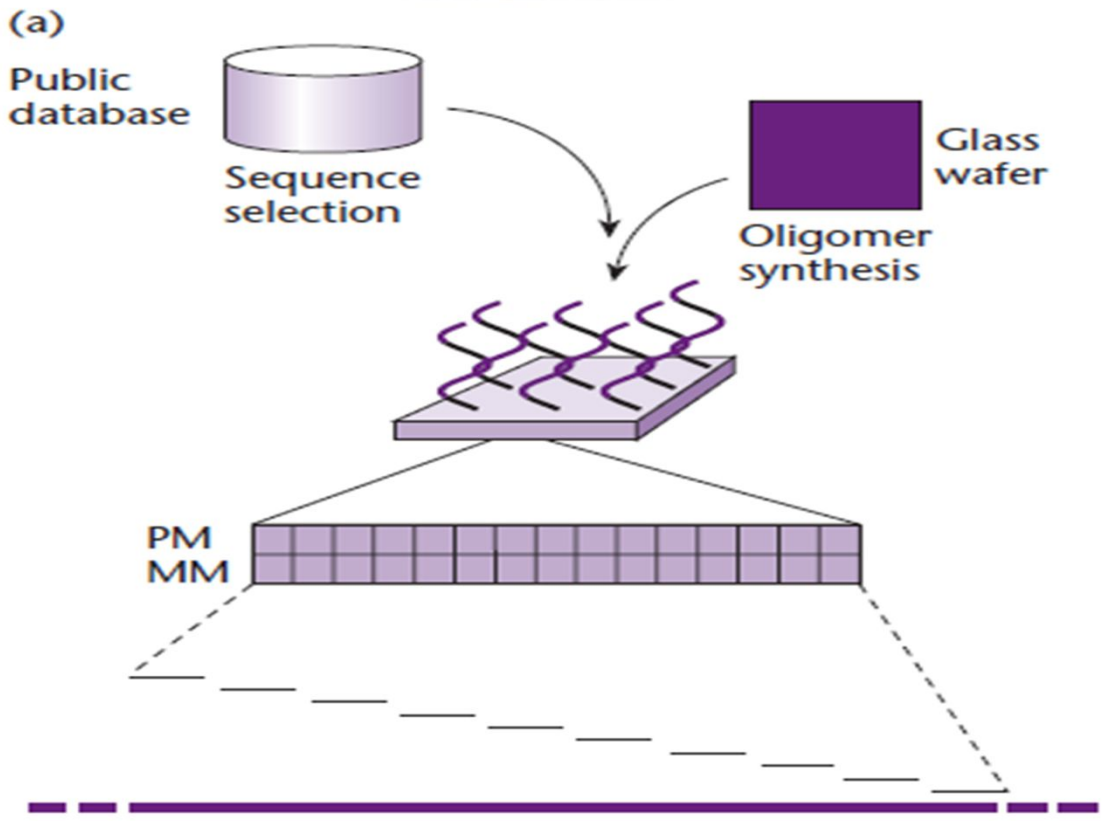
- **Spotted DNA arrays are produced by printing DNA samples on treated microscope slides. It is made by transferring or spotting DNA clones or PCR product individually onto a solid support where they are immobilized**

Functional genomics and proteomics

Oligonucleotide chips

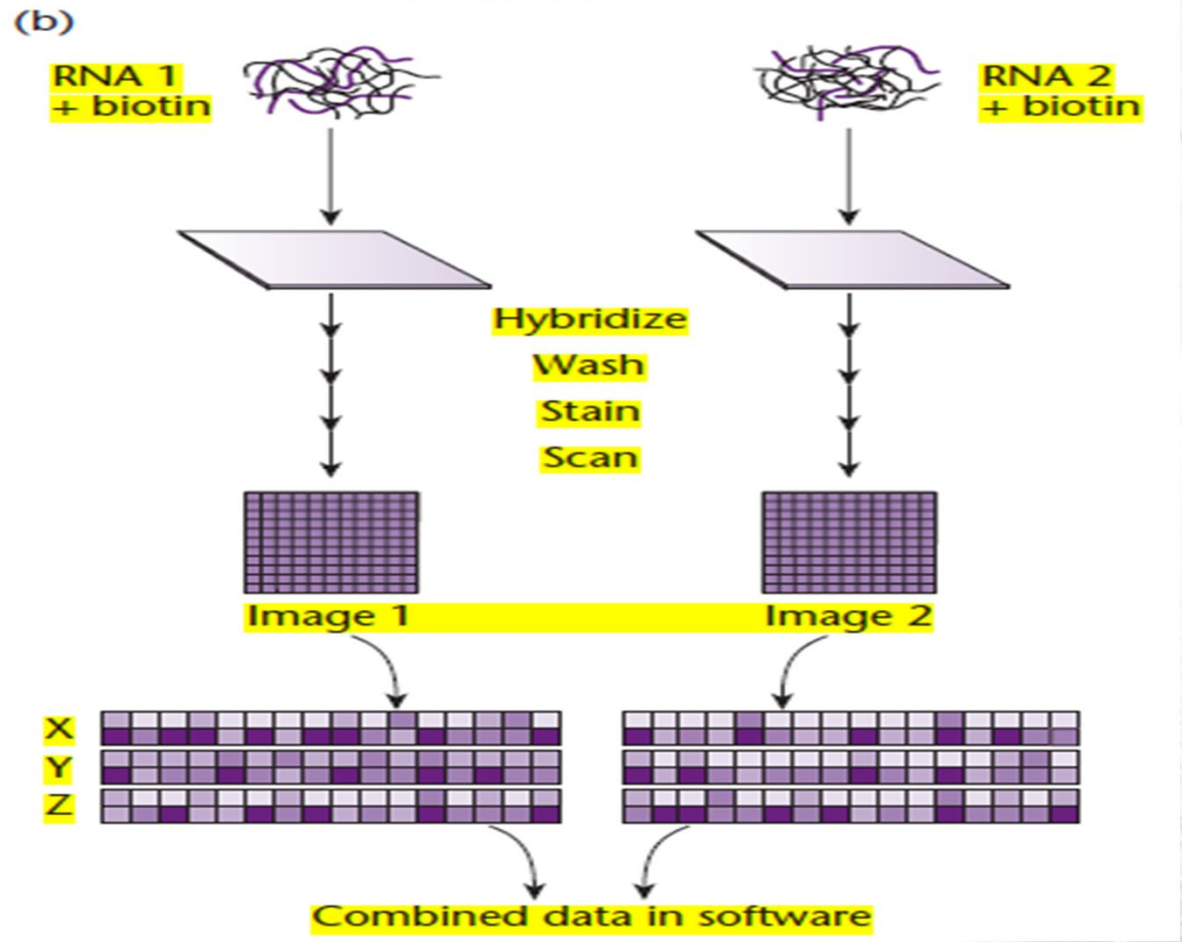
- **Oligonucleotide chips are manufactured by in situ oligonucleotide synthesis**

Functional genomics and proteomics



Principle of chip manufacture

Functional genomics and proteomics



Principle of chip hybridization

Applications of microarrays

Microbial gene expression analysis

- The predominate application of DNA microarray has been to measure gene expression levels

Applications of microarrays

Profiling in human disease

- Arrays have been used to investigate transcriptional profiles associated with human disease and to identify novel disease markers and potential new drug targets

Applications of microarrays

Genotyping

- Microarrays have been widely used as single nucleotide polymorphism (SNP) genotyping platforms

Functional genomics and proteomics

Phage display

- **Phage display is the technology that allows expression of exogenous polypeptides on the surface of phage particles**

Functional genomics and proteomics

Screening phage display libraries

- The most common screening method is based on enriching the phage clones with binding affinity for the target by a process called biopanning

Functional genomics and proteomics

Applications of phage display

- Applications of phage display includes determination of protein to protein interactions, to determine enzyme specificity and to generate target specific antibodies

Functional genomics and proteomics

Knock outs

- A gene knock out is a genetic technique in which one of an organism's gene is made inoperative that is simply called as knocked out

Functional genomics and proteomics

Knock ins

- In molecular cloning, a knock-in refers to a gene manipulation method that involves the insertion of a protein coding cDNA sequence at a particular locus in an organism's chromosome

Functional genomics and proteomics

siRNA technology

- **Small interfering RNA (siRNA) is the most commonly used RNA interference (RNAi) tool for inducing short-term silencing of protein coding genes. siRNA is a synthetic RNA duplex designed to specifically target a particular mRNA for degradation**

Functional genomics and proteomics

Applications of siRNA

- siRNA) is the method of choice to target specific genes for silencing and has provided immense potential as therapeutic tools