ACADEMIC YEAR 2025-2026, SEMESTER – V
STUDY MATERIAL FOR B.Sc. MICROBIOLOGY
BASIC BIOTECHNOLOGY AND BIOINFORMATICS

STUDY MATERIAL FOR B.Sc. MICROBIOLOGY

BASIC BIOTECHNOLOGY AND BIOINFORMATICS

SEMESTER – V

ACADEMIC YEAR 2025-26

PREPARED BY

MICROBIOLOGY DEPARTMENT

INDEX

UNIT - I

1. History and Development

- Biotechnology is the use of living organisms or biological systems to develop products or technologies for human use.

- Ancient use: Fermentation for bread, alcohol (6000–4000 BC).

- Modern era: Began with recombinant DNA technology (1970s).

Key Milestones:

- 1953 – DNA structure discovered (Watson & Crick)

- 1973 – First recombinant DNA molecule

- 1983 – PCR (Polymerase Chain Reaction) developed

- 1997 – First cloned mammal (Dolly the sheep)

2. Scope and Importance

- Offers solutions to agriculture, medicine, environment, and industry.

- Plays a major role in:

    o Disease diagnosis and treatment

    o Development of genetically modified crops

    o Industrial enzyme production

    o Waste management and pollution control

3. Applications of Biotechnology

Field	Applications
Medical	Production of vaccines (e.g., Hepatitis B), insulin, gene therapy, diagnostics
Agriculture	GM crops (Bt cotton, Golden Rice), pest-resistant, high-yield crops
Industrial	Fermentation, enzyme production, biofuels, bioplastics
Environmental	Bioremediation, waste treatment, biosensors
Food	Fermentation (yogurt, cheese), food preservation, flavour enhancers

----

Applications of Biotechnology – Brief Notes

1. Agriculture

- **GM crops:** Pest-resistant (e.g., Bt cotton), drought-tolerant, high yield.

- **Tissue culture:** Rapid multiplication of disease-free plants.

- **Bio fertilizers & Bio pesticides:** Environment-friendly alternatives.

- **Marker-assisted breeding:** Faster and more accurate plant breeding.

2. Medicine

- **Recombinant drugs:** Insulin, growth hormones, interferons

- **Gene therapy:** Correcting defective genes (e.g., SCID).

- **Vaccine production:** Hepatitis B, HPV (using recombinant technology).

- **Molecular diagnostics:** PCR, ELISA for disease detection.

3. Environment

- **Bioremediation:** Use of microbes to clean oil spills and pollutants.

- **Bioleaching:** Extraction of metals using microbes.

- **Biosensors:** Detect environmental toxins or pollutants.

- **Waste management:** Biodegradable plastics and composting.

4. Veterinary Sciences

- **Animal vaccines:** Rabies, foot and mouth disease.

- **Artificial insemination & embryo transfer:** Improve breeds.

- **Disease diagnosis:** PCR for animal diseases.

- **Transgenic animals:** For better productivity and disease models.

5. Food Industry

- **Fermentation:** Production of yogurt, cheese, vinegar.

- **Enzyme use:** Amylase, protease for improving food processing.

- **Genetically modified yeast:** Improved baking and alcohol fermentation.

- **Preservation:** Biotech-based additives and packaging.

----

6. Chemical Industry

- **Biocatalysts/Enzymes:** Used in detergents, textiles, paper.

- **Biopolymers:** Biodegradable plastics like PHB, PLA.

- **Green chemistry:** Eco-friendly chemical synthesis using microbes or enzymes.

7. Pharmaceutical Industry

- **Drug production:** Antibiotics, anti-cancer drugs.

- **Monoclonal antibodies:** Used in cancer and autoimmune diseases.

- **Personalized medicine:** Based on genetic profiling.

- **Bio similars:** Affordable versions of biological drugs.

8. Forensic Science

- **DNA fingerprinting:** Identifying individuals in crime scenes.

- **Paternity testing:** Using STR markers or Y-DNA analysis.

- **Biological evidence analysis:** Hair, blood, saliva.

- **Gene profiling:** Criminal investigations and missing person cases.

Bioremediation and Waste Treatment Biotechnology – Brief Notes

1. What is Bioremediation?

- Bioremediation is the use of microorganisms or plants to remove, degrade, or neutralize pollutants from soil, water, or air.

- It is an eco-friendly and cost-effective waste management method.

2. Types of Bioremediation

Type	Description	Example
In situ	Treatment at the contamination site	Oil spill cleanup using bacteria
Ex situ	Contaminated material is removed and treated elsewhere	Composting solid waste

3. Mechanisms Used

- **Microbial degradation** of organic pollutants (e.g., hydrocarbons, pesticides).

- **Phytoremediation:** Use of plants to absorb or detoxify pollutants.

- **Bio augmentation:** Adding specific microbes to enhance cleanup.

- **Bio adsorption:** Microbes adsorb heavy metals onto their cell walls.

4. Waste Treatment Using Biotechnology

- **Sewage treatment:** Use of microbes in activated sludge processes.

- **Composting:** Aerobic microbial breakdown of organic solid waste.

- **Anaerobic digestion:** Microbes break down waste to produce biogas (methane).

- **Bio filters and bioreactors:** Treat air and water pollution using microbial films.

5. Applications

- Cleaning oil spills and industrial waste.

- Treating heavy metal contamination.

- Wastewater treatment in sewage plants.

- Recycling agricultural waste into bio fertilizers.

1. Global Trends in Biotechnology

- Gene editing (e.g., CRISPR-Cas9) for precision medicine and agriculture.

- Personalized medicine and genomic diagnostics.

- Synthetic biology for developing artificial organisms.

- Biopharmaceuticals: Increasing demand for vaccines, bio similars.

- Green biotechnology: Focus on sustainable agriculture and environment.

- Bioinformatics: Integration of big data and AI in life sciences.

2. Biotechnology in India

India is a rapidly growing hub for biotech, especially in healthcare, agriculture, and industrial enzymes.

Key Features:

- 12th largest biotech economy globally.

- Focus on bio pharma, agri-biotech, bio-services, and bio-industrial products.

- Supported by government initiatives like:

  o Biotechnology Industry Research Assistance Council (BIRAC)

  o Department of Biotechnology (DBT)

----

3. Major Biotechnology Institutes in India

Institute	Location	Focus Area
National Institute of Immunology (NII)	New Delhi	Immunology, vaccine development
Indian Institute of Science (IISc)	Bangalore	Multidisciplinary biotech research
National Centre for Cell Science (NCCS)	Pune	Cell biology, genomics
Institute of Genomics and Integrative Biology (IGIB)	Delhi	Genomics, bioinformatics
Centre for Cellular and Molecular Biology (CCMB)	Hyderabad	Molecular biology, gene editing

4. Major Biotechnology Companies in India

Company	Headquarters	Specialization
Biocon	Bangalore	Biopharmaceuticals, insulin, bio similars
Serum Institute of India	Pune	Vaccines (e.g., Covishield)
Panacea Biotec	New Delhi	Vaccine and drug development
Dr. Reddy's Laboratories	Hyderabad	Bio similars, generic drugs
Shantha Biotechnics	Hyderabad	Hepatitis B vaccine, insulin

UNIT - II

Plant Biotechnology – Brief Exam Notes

1. Plant Tissue Culture

- Technique to grow plant cells/tissues in nutrient media under sterile conditions.

- Used for propagation, genetic modification, and conservation.

Sterilisation

- Essential to prevent microbial contamination.

- Methods:

  o Instruments: Autoclaving

  o Media: Filtration or autoclaving

  o Explants: Surface sterilization using ethanol or sodium hypochlorite

Media Requirements

- MS (Murashige and Skoog) medium – most commonly used.

- Contains: Macronutrients, micronutrients, vitamins, plant growth regulators (auxins, cytokinins), sugar.

2. Micro propagation

- Rapid cloning of plants using tissue culture techniques.

Stages:

1. Initiation

2. Multiplication (shoot formation)

3. Rooting

4. Acclimatization

3. Callus and Protoplast Culture

- Callus: Unorganized mass of cells from explants induced by auxins and cytokinins

- Protoplast: Plant cell without a cell wall, obtained using cellulase and pectinase. Used for somatic hybridization.

4. Agrobacterium and Crown Gall Tumors

- Agrobacterium tumefaciens causes Crown Gall disease in plants.

- Transfers part of its plasmid (T-DNA) into plant genome.

Ti Plasmid

- Tumour-inducing plasmid.

- Modified for use as a vector in plant genetic engineering.

Ri Plasmid

- Found in Agrobacterium rhizogenes induces hairy root disease.

- Used for root culture studies.

Animal Biotechnology – Brief Exam Notes

1. Principles of Animal Cell Culture

- In vitro growth of animal cells in nutrient-rich media.

- Requires sterile conditions and controlled environment (CO_2, temperature, humidity).

Media

- Examples: DMEM, RPMI-1640, MEM

- Contains: Amino acids, vitamins, glucose, serum (FBS), antibiotics.

Equipment

- CO_2 incubator, biosafety cabinet, centrifuge, inverted microscope, laminar airflow hood.

2. Primary and Secondary Cultures

- Primary Culture: Directly from tissues, limited lifespan.

- Secondary Culture (Sub-culture): Transfer of cells from primary culture to new media; allows expansion.

3. Cell Lines

- Cultures that can be propagated indefinitely (immortalized).

Types:

- Finite cell lines: Limited number of divisions.

- Continuous cell lines: Immortal; derived from tumors or transformed cells (e.g., HeLa, CHO).

Establishment & Maintenance

- Regular sub-culturing

- Monitoring for contamination

- Proper storage (cryopreservation)

4. Transgenic Animals

- Animals with foreign genes inserted into their genome.

- Methods: Microinjection, viral vectors, CRISPR.

- Applications: Disease models, drug production, organ donation (xenotransplantation)

5. Industrially Important Products

Product	Use
Recombinant proteins	Insulin, growth hormones
Vaccines	Hepatitis B, COVID-19 (subunit)
Immuno toxins	Targeted cancer therapy
Monoclonal antibodies	Autoimmune diseases, diagnostics
Hybridoma technology	Produces monoclonal antibodies using fusion of B-cells and myeloma cells

----

UNIT - III

Genetic Engineering

Definition

Genetic engineering is the direct manipulation of an organism's DNA using biotechnology techniques to alter its characteristics or produce desired proteins.

History of Genetic Engineering

- **1972–1973:** Paul Berg, Stanley Cohen & Herbert Boyer developed recombinant DNA (rDNA) technology.

- **1977:** First artificial plasmid vector (pBR322) developed.

- **1982:** First genetically engineered product (human insulin) approved.

- **1994:** First GM food (Flavr Savr tomato) released.

- **2000s:** Advances in CRISPR-Cas9 genome editing.

Scope of Genetic Engineering

- **Medicine:** Gene therapy, vaccine production, monoclonal antibodies.

- **Agriculture:** GMO crops (Bt cotton, Golden rice).

- **Industry:** Production of enzymes, biofuels.

- **Research:** Gene function studies, model organisms.

Concepts in Genetic Engineering

- **Recombinant DNA (rDNA):** Combining DNA from two different sources.

- **Cloning:** Making identical copies of DNA or organisms.

- **Vectors:** DNA molecules (e.g., plasmids) used to deliver genes into cells.

- **Host cells:** Bacteria or eukaryotic cells used to express foreign genes.

- **Restriction enzymes:** Cut DNA at specific sites.

- **DNA ligase:** Joins DNA fragments.

Principles of rDNA Technology

- **Isolation** of genetic material (DNA).

- **Digestion** with restriction enzymes.

- **Ligation** of DNA into vectors using ligase.

- **Transformation** into host organism.

- **Selection and screening** for recombinant organisms.

Applications of rDNA Technology

- **Medicine:** Insulin, growth hormones, vaccines.

- **Agriculture:** Disease-resistant and pest-resistant crops.

- **Environment:** Bioremediation using engineered microbes.

- **Research:** Study of gene expression, genetic diseases.

Concepts, Principles, and Applications of rDNA Technology

1. Concepts of rDNA Technology

Recombinant DNA (rDNA) technology involves combining DNA from two different organisms to form a new, functional DNA molecule that is then introduced into a host organism.

Key Concepts:

1. Gene Cloning – Making identical copies of a gene or DNA sequence.

2. Vectors – DNA carriers (e.g., plasmids, phages) used to deliver foreign genes into host cells.

3. Host Organism – Microbial or eukaryotic cells (e.g., E. coli, yeast) used to express the recombinant DNA.

4. Restriction Enzymes (RE) – Enzymes that cut DNA at specific sequences to allow insertion into vectors.

5. DNA Ligase – Enzyme that joins DNA fragments by forming phosphodiester bonds.

6. Selectable Markers – Genes (like antibiotic resistance) used to identify transformed cells.

7. Transformation – The process of introducing recombinant DNA into host cells.

2. Principles of rDNA Technology

Recombinant DNA technology is based on a step-by-step process involving the following:

Step-by-Step Principles:

1. **Isolation of Genetic Material (DNA):**

   - Extract DNA from donor and vector organisms.

2. **Cutting DNA at Specific Sites:**

   - Use restriction endonucleases to cut donor and vector DNA at specific sequences.

----

3. **Insertion of Gene into Vector:**

- Use DNA ligase to insert foreign DNA into vector DNA (plasmid/phage).

4. **Introduction into Host:**

- Introduce recombinant vector into a host cell using transformation, electroporation, or microinjection.

5. **Selection of Recombinant Cells:**

- Use antibiotic resistance or blue-white screening to identify transformed cells.

6. **Expression of Gene of Interest:**

- In suitable hosts, the inserted gene is transcribed and translated into a functional protein.

7. **Harvesting the Product:**

- Desired proteins or gene products are extracted, purified, and used for applications.

3. Applications of rDNA Technology

Recombinant DNA technology has revolutionized medicine, agriculture, and industry.

In Medicine:

- Insulin production (first rDNA product approved by FDA).
- Human Growth Hormone (HGH), erythropoietin, interferons.
- Gene therapy for correcting genetic disorders.
- Production of vaccines, e.g., Hepatitis B vaccine.

In Agriculture:

- Genetically Modified (GM) Crops like Bt cotton and Golden Rice.
- Insect-resistant, herbicide-tolerant, disease-resistant plants.
- Improved yield, shelf life, and nutrition.

In Research:

- Gene function studies using reporter genes.
- Development of genetic markers, DNA probes.
- Production of transgenic animals and model organisms.

In Industry:

- Production of enzymes (amylase, protease).

- Biodegradable plastics using engineered microbes.

- Bioremediation using bacteria to clean pollutants.

Summary Table for Exams

Step	Description
1. Isolation	Extract DNA from donor and vector organisms
2. Cutting	Use restriction enzymes to cut DNA
3. Ligation	Insert gene into vector using ligase
4. Transformation	Introduce rDNA into host cells
5. Selection	Use markers to identify recombinants
6. Expression	Production of protein/gene product
7. Purification	Extract the desired product

Isolation & Purification of DNA from Cells

Introduction

Isolation of DNA is a fundamental step in genetic engineering. It involves separating DNA from cellular components such as proteins, lipids, RNA, and other organelles to obtain pure DNA for downstream applications like cloning, PCR, sequencing, etc.

Steps in DNA Isolation

1. Cell Lysis (Breaking Open the Cells)

- Goal: To disrupt the cell wall/membrane and nuclear membrane to release DNA.
- Methods:
  - **Mechanical:** Grinding, homogenization.
  - **Chemical:** Use of detergents like SDS (Sodium Dodecyl Sulfate) to dissolve lipids.
  - **Enzymatic:**
    - Lysozyme (for bacterial cell walls),
    - Proteinase K (to digest proteins).

----

2. Removal of Proteins and Other Contaminants

- Phenol-Chloroform Extraction:

    o Phenol denatures proteins.

    o Chloroform helps in phase separation.

    o Centrifugation separates DNA (aqueous layer) from proteins (organic layer).

- Protease treatment can also be used to digest proteins before extraction.

3. DNA Precipitation

- DNA is precipitated by adding cold ethanol or isopropanol.

- DNA becomes visible as a white thread-like precipitate.

- Carried out in the presence of salts (e.g., sodium acetate) to neutralize charges.

4. Washing the DNA

- The DNA pellet is washed with 70% ethanol to remove residual salts and impurities.

- After washing, the pellet is dried.

5. Resuspension of DNA

- Pure DNA is resuspended in TE buffer (Tris-EDTA) or nuclease-free water.

- Stored at -20°C or -80°C for long-term use.

Reagents Used in DNA Isolation

Reagent	Function
SDS (Detergent)	Breaks membranes
EDTA	Chelates Mg^{2+} (inhibits nucleases)
Proteinase K	Digests proteins
Phenol-Chloroform	Removes proteins
Ethanol/Isopropanol	Precipitates DNA
RNase A	Removes RNA contaminants

Types of DNA Isolation Protocols

- Genomic DNA Isolation: For long fragments from nucleus/chromosome.

- Plasmid DNA Isolation: For small circular DNA from bacteria using alkaline lysis.

- Total DNA Isolation: Includes both genomic and organelle DNA (mitochondrial/chloroplast).

Precautions

- Use sterile and nuclease-free reagents.

- Avoid shearing of DNA by gentle mixing.

- Keep DNA cold to avoid degradation.

Purity Check of Isolated DNA

- Spectrophotometry:

- Ratio of absorbance at 260/280 nm:

- ~1.8 → pure DNA

- <1.8 → protein contamination

- Agarose Gel Electrophoresis:

- To check size, integrity, and yield of DNA.

Applications of Purified DNA

- PCR (Polymerase Chain Reaction)

- DNA sequencing

- Cloning and recombinant DNA work

- DNA fingerprinting

- Gene expression studies

Summary (One-Liner Points for Exam)

- DNA is isolated via lysis, extraction, precipitation, and purification.

- SDS, EDTA, and proteases play key roles.

- Purity checked by 260/280 nm ratio and gel electrophoresis.

- Used in PCR, cloning, and molecular diagnostics.

DNA Ligases and DNA Modifying Enzymes

1. DNA Ligases

Definition:

DNA ligases are enzymes that join breaks in the DNA backbone by forming phosphodiester bonds between adjacent nucleotides. They are essential for DNA replication, repair, and recombinant DNA technology.

Types of DNA Ligases:

Type	Source	Cofactor	Application
T4 DNA Ligase	Bacteriophage T4	ATP	Joins sticky and blunt ends in cloning
E. coli DNA Ligase	E. coli	NAD^+	Mostly used in DNA repair studies

Functions:

- Seals nicks in the DNA during replication and repair.

- Joins DNA fragments during cloning.

- Facilitates ligation of vector and insert DNA in rDNA technology.

Applications in Genetic Engineering:

- Cloning recombinant plasmids.

- Ligation reactions after restriction digestion.

- Blunt-end ligation (with T4 ligase).

- In vitro mutagenesis and gene assembly.

2. DNA Modifying Enzymes

These enzymes alter DNA structure or chemical composition and are essential for molecular cloning, DNA labeling, and sequencing.

Major Types and Functions:

Enzyme	Function	Use in rDNA Technology
Kinases (e.g., T4 Polynucleotide Kinase)	Adds phosphate groups to 5' ends of DNA	Labelling DNA ends, preparing for ligation

----

Enzyme	Function	Use in rDNA Technology
Phosphatases (e.g., Alkaline Phosphatase)	Removes phosphate from 5' ends	Prevents vector self-ligation
Nucleases (e.g., Exonucleases, DNase I)	Degrade DNA from ends or internally	DNA cleanup, mapping, mutagenesis
DNA Polymerases (e.g., Taq, Klenow fragment)	Synthesizes DNA from a template	PCR, fill-in blunt ends
Reverse Transcriptase	Converts RNA to cDNA	cDNA library construction
Methylases	Add methyl groups to DNA bases	Protect DNA from restriction enzymes

Commonly Used Modifying Enzymes:

1. **T4 Polynucleotide Kinase (PNK):**

   o Adds phosphate to 5'-OH ends.

   o Used for labelling and preparing DNA ends for ligation.

2. **Calf Intestinal Alkaline Phosphatase (CIAP):**

   o Removes 5' phosphate to prevent self-ligation of plasmid.

3. **Klenow Fragment (of DNA Pol I):**

   o Used to fill in 5' overhangs or create blunt ends.

4. **Exonuclease III:**

   o Digests DNA from the 3' end; used in mapping and site-directed mutagenesis.

5. **Terminal Transferase:**

   o Adds nucleotides to the 3' ends of DNA; useful in adding poly-A tails or labelling.

Summary (One-liner Points for Exams):

- DNA ligase joins DNA fragments by forming phosphodiester bonds.

- T4 DNA ligase is most widely used in molecular biology.

- DNA modifying enzymes alter DNA ends or structure to make cloning possible.

- Examples: kinases, phosphatases, nucleases, polymerases, methylases.

Eukaryotic and Prokaryotic Hosts for Cloning

Introduction

In genetic engineering, host organisms are used to accept and replicate recombinant DNA introduced via vectors. Selection of an appropriate host depends on factors like growth rate, ease of transformation, gene expression system, and post-translational modifications.

Hosts are broadly categorized into:

- Prokaryotic hosts (e.g., *E. coli*)

- Eukaryotic hosts (e.g., yeast, mammalian cells)

1. Prokaryotic Hosts for Cloning

Most Common Host: Escherichia coli (E. coli)

Features of E. coli as Host:

- Well-characterized genome.

- Rapid growth and easy to culture.

- High transformation efficiency.

- Multiple well-developed cloning vectors (e.g., pBR322, pUC19).

- Easy recovery of plasmids and proteins.

- Suitable for producing simple recombinant proteins.

Common E. coli Strains Used:

Strain	Special Features
DH5α	High plasmid yield, lacks endonucleases
JM109	Ideal for blue-white screening
BL21 (DE3)	Used for protein expression (T7 system)
TOP10	High transformation efficiency

Limitations of E. coli:

- Cannot perform post-translational modifications (glycosylation, folding).

- Endotoxins (LPS) in outer membrane problematic for pharmaceuticals.

- Inefficient expression of eukaryotic genes with introns.

2. Eukaryotic Hosts for Cloning

Eukaryotic hosts are chosen for functional expression of eukaryotic proteins, especially when post-translational modifications are required.

a) Yeast (*Saccharomyces cerevisiae*)

Advantages:

- Eukaryotic system – can perform glycosylation and folding.

- Grows quickly and is easy to genetically manipulate.

- Can maintain plasmids (e.g., YEp, YIp, YCp vectors).

- Safe and inexpensive.

Applications:

- Expression of recombinant proteins (e.g., insulin, vaccines).

- Used in making YACs (Yeast Artificial Chromosomes) for large DNA inserts.

b) Insect Cell Lines (e.g., Sf9 cells from *Spodoptera frugiperda*)

- Used with baculovirus expression system.

- High-level protein expression.

- Useful for producing functional eukaryotic proteins.

c) Mammalian Cells (e.g., CHO, HEK293)

Advantages:

- Capable of accurate post-translational modifications.

- Required for expression of complex human proteins (e.g., monoclonal antibodies).

- Used in therapeutic protein production.

Limitations:

- Slow growth.

- Expensive media.

- Requires advanced sterile techniques.

d) Plant and Algal Cells

- Used for transgenic plant production (e.g., Agrobacterium-mediated transformation).

- Useful in biopharming—production of pharmaceuticals in plants.

Comparison Table: Prokaryotic vs. Eukaryotic Hosts

Feature	Prokaryotic Hosts (E. coli)	Eukaryotic Hosts (Yeast, CHO)
Growth rate	Fast	Moderate to slow
Cost	Low	High
Gene expression	Easy for bacterial genes	Suitable for eukaryotic genes
Post-translational mods	Not possible	Possible
Scalability	Easy	Complex
Introns	Cannot process introns	Can process via splicing machinery
Protein folding	May form inclusion bodies	Proper folding and processing

Criteria for Choosing a Host System

1. Nature of the gene (prokaryotic or eukaryotic).

2. Requirement for post-translational modifications.

3. Scale of production.

4. Ease of culture and transformation.

5. Availability of suitable vectors and promoters.

Conclusion

Selection of the right host is critical for the success of gene cloning. E. coli remains the most commonly used host due to its simplicity, while yeast and mammalian cells are preferred for complex proteins requiring proper folding and modifications.

Restriction Enzymes – Types and Sources

Introduction

Restriction enzymes, also called restriction endonucleases, are enzymes that recognize and cut DNA at specific sequences, called restriction sites. They are essential tools in recombinant DNA (rDNA) technology.

Discovered in 1970 by Werner Arber, Daniel Nathans, and Hamilton Smith, these enzymes naturally occur in bacteria and archaea, acting as a defense mechanism against invading viruses (bacteriophages).

Restriction enzyme Definition:

Restriction enzymes are endonucleases that recognize specific palindromic sequences in double-stranded DNA and cleave both strands at or near the recognition site.

Concept of Palindromic Sequences:

A palindrome is a sequence of DNA that reads the same on both strands in the 5' to 3' direction.

 Example:

5' – GAATTC – 3'

3' – CTTAAG – 5'

(Cleaved by EcoRI)

Classification / Types of Restriction Enzymes

Restriction enzymes are grouped into three main types based on their structure, cofactor requirements, and site of cleavage:

Type I Restriction Enzymes

- Recognize specific DNA sequences but cut at random sites far from the recognition site.

- Multi subunit complex: Has both restriction and methylation activities.

- Require ATP, Mg^{2+}, and S-adenosyl methionine (SAM).

- Rarely used in genetic engineering due to unpredictable cutting.

Example: EcoKI from E. coli K12

Type II Restriction Enzymes

- Recognize and cut within or near specific palindromic recognition sites.

- Require only Mg^{2+} (no ATP).

- Most commonly used in molecular biology and rDNA technology.

- Highly specific, producing either blunt ends or sticky (cohesive) ends.

Example:

- EcoRI → Sticky ends

- HindIII, BamHI, AluI, SmaI

ACADEMIC YEAR 2025-2026, SEMESTER – V
STUDY MATERIAL FOR B.Sc. MICROBIOLOGY
BASIC BIOTECHNOLOGY AND BIOINFORMATICS

Type III Restriction Enzymes

- Recognize specific sequences and cut at short distance (~25 bp) away from the site.

- Require ATP and SAM.

- Less commonly used in labs.

Example: EcoP15I

Sources of Restriction Enzymes

- Naturally found in prokaryotes (bacteria and archaea).

- Named based on:

    o Genus and species

    o Strain

    o Order of discovery

Naming Convention Example:

EcoRI

- E → Genus (Escherichia)

- co → Species (coli)

- R → Strain (RY13)

- I → First enzyme discovered

Common Enzymes and Sources:

Enzyme	Source Organism	Recognition Site	Cut Type
EcoRI	E. coli	GAATTC	Sticky ends
HindIII	Haemophilus influenzae	AAGCTT	Sticky ends
BamHI	Bacillus amyloliquefaciens	GGATCC	Sticky ends
AluI	Arthrobacter luteus	AGCT	Blunt ends
SmaI	Serratia marcescens	CCCGGG	Blunt ends
TaqI	Thermus aquaticus	TCGA	Sticky ends

Types of DNA Cuts

Cut Type	Description	Example
Sticky Ends	Overhanging single-stranded ends	EcoRI, BamHI
Blunt Ends	Straight cuts with no overhang	SmaI, AluI

Applications of Restriction Enzymes in rDNA Technology

1. Gene Cloning – Cutting both vector and insert DNA with same enzyme ensures compatibility.

2. DNA Mapping – Determining gene location using restriction maps.

3. Genome Editing – Site-specific cleavage enables insertion/deletion.

4. Diagnostic Tools – RFLP (Restriction Fragment Length Polymorphism) in forensic and disease studies.

5. Creating Recombinant Vectors – Seamless ligation of gene of interest into plasmids.

Cloning Vectors – Plasmid-Based Vectors

Definition of Cloning Vector

A cloning vector is a DNA molecule used to carry foreign genetic material into a host cell for the purpose of replication and expression. Vectors must be capable of independent replication and should have features that allow for the insertion, selection, and identification of recombinant molecules.

Plasmid-Based Vectors

Plasmids are small, circular, double-stranded DNA molecules found naturally in bacteria. They replicate independently of the chromosomal DNA and are widely used as vectors in genetic engineering.

Essential Features of Plasmid Cloning Vectors

Feature	Function
Origin of replication (ori)	Initiates replication inside host
Selectable marker genes	e.g., antibiotic resistance to identify transformed cells
Multiple Cloning Site (MCS)	Region containing many restriction sites for gene insertion

----

Feature	Function
Small size	Enhances transformation efficiency
High copy number	Increases plasmid yield

Natural Plasmid-Based Vectors

1. pSC101

- Origin: Derived from Salmonella plasmid.

- Size: ~9.3 kb

- Features:

    o Low copy number (~5 copies/cell).

    o Carries tetracycline resistance (Tet^R) gene.

    o Used in early cloning experiments.

- Limitations:

    o Low yield of cloned DNA.

    o No multiple cloning site.

2. pSF2124

- Origin: Naturally occurring in Salmonella.

- Features:

    o Carries multiple antibiotic resistance genes.

    o Moderate copy number.

- Use: Less commonly used in modern cloning due to limited versatility.

3. pMB1

- Origin: Derived from E. coli.

- Significance:

    o Its origin of replication (ori) is used in constructing many modern artificial vectors like pBR322 and pUC.

    o High replication rate; used to create high-copy number plasmids.

Artificial Plasmid-Based Vectors

Artificial vectors are engineered plasmids created to overcome limitations of natural plasmids and to enhance cloning efficiency.

1. pBR322

Feature	Description
Size	4.36 kb
Origin	Constructed by Bolivar and Rodriguez (BR = Bolivar & Rodriguez)
Selectable Markers	Ampicillin resistance (amp^R) and Tetracycline resistance (tet^R)
Cloning Sites	Several unique restriction sites located within amp^R and tet^R genes
Origin of Replication	Derived from pMB1 ori
Copy Number	Moderate (~15–20 copies/cell), can be increased by mutation

Advantages:

- Dual antibiotic resistance enables insertional inactivation screening.
- Small size, stable replication in E. coli.

Limitations:

- Lacks MCS (Multiple Cloning Site).
- Moderate copy number compared to newer plasmids.

2. pUC Series (e.g., pUC18, pUC19)

Feature	Description
Size	~2.7 kb (smaller than pBR322)
Origin	Derived from pBR322 with improved features
Selectable Marker	Ampicillin resistance (amp^R)
MCS	Present within the lacZα gene
Screening	Blue-white screening via X-gal/IPTG
Copy Number	High (~500–700 copies/cell) due to modified pMB1 ori

----

Advantages:

- High yield of plasmid DNA.

- Efficient selection using lacZ complementation.

- Small and easy to manipulate.

- Widely used in modern cloning labs.

Limitation:

- Only one antibiotic resistance marker (amp^R).

Comparison Table: pBR322 vs pUC

Feature	pBR322	pUC18/19
Size	4.36 kb	2.7 kb
Copy number	Moderate	High
Antibiotic resistance	Amp^R, Tet^R	Amp^R only
MCS	No	Yes (in lacZ gene)
Screening method	Antibiotic inactivation	Blue-white screening
Origin	pMB1	pMB1 (modified)

Phage-Based Vectors: Construction and Lambda (λ) Phage Vectors

Introduction

Phage-based vectors are derived from bacteriophages (viruses that infect bacteria). These vectors are especially useful in gene cloning, DNA library construction, and transduction because of their ability to efficiently infect and replicate within bacterial cells.

Among all phages, the lambda (λ) phage is the most widely used in genetic engineering due to its well-characterized genome and ability to accommodate foreign DNA.

1. What is Lambda (λ) Phage?

- A temperate bacteriophage that infects Escherichia coli (E. coli).

- Genome: ~48.5 kb linear double-stranded DNA.

- Has cohesive (cos) ends – 12-base single-stranded overhangs that allow circularization inside the host.

- Capable of entering lysogenic or lytic cycle.

2. Key Features of Lambda Phage Genome

Region	Function
cos sites	Cohesive ends, essential for packaging DNA into phage heads
att site	Site for integration into host chromosome (lysogeny)
Genes for head and tail proteins	Involved in phage assembly
Lytic and lysogenic control genes	Determine lifecycle (Cro, cI, etc.)

3. Construction of λ Phage-Based Vectors

To use lambda phage as a vector, non-essential genes are removed and replaced with foreign DNA (insert). There are two major strategies for constructing λ vectors:

a) Insertion Vectors

- A single restriction site is introduced into the non-essential region (e.g., between genes not required for lytic growth).

- DNA inserts of up to 7–8 kb can be cloned.

- Example: λgt10 vector.

Advantages:

- Simple cloning system.

- Suitable for smaller inserts.

b) Replacement Vectors

- Large non-essential regions (e.g., between the stuffer fragment) are removed and replaced with foreign DNA.

- Can carry larger inserts: up to 20–25 kb.

- Stuffer DNA is used to maintain size required for packaging.

- Example: λgt11, λEMBL3, λZAP.

Advantages:

- Efficient for genomic or cDNA library construction.

- Better suited for larger DNA fragments.

Packaging Size Constraint:

- For λ DNA to be packaged into phage heads, its size must be between 38–52 kb.

- Vectors outside this range won't be packaged or infectious.

4. Derivatives of λ Phage Vectors

Vector	Type	Key Features	Application
λgt10	Insertion	Carries lacZ gene for blue-white screening	Cloning smaller DNA fragments
λgt11	Expression	Contains lacZ for fusion protein expression	Production of protein from inserted gene
λZAP	Replacement	Allows excision of cloned DNA into plasmid (phagemid)	Library screening and sub cloning
λEMBL3 & EMBL4	Replacement	Improved stability and capacity (~23 kb inserts)	Genomic library construction

Steps in Using λ Phage Vectors for Cloning:

1. Restriction digestion of vector and insert DNA.

2. Ligation of foreign DNA into λ vector.

3. In vitro packaging into phage particles (using packaging extract).

4. Infection of E. coli host cells.

5. Screening for recombinant phage plaques.

Advantages of Lambda Phage Vectors

- High efficiency of DNA delivery via infection.

- Can accommodate larger DNA inserts than plasmids.

- Useful for genomic and cDNA library construction.

- Can be screened using plaque hybridization techniques.

Limitations

- Not suitable for cloning very small DNA fragments (<5 kb).

- Requires specific E. coli host strains for propagation.

- More complex than plasmid cloning.

Hybrid Vectors: Phagemid, Phasmid, Cosmid, BAC, and YAC

Introduction

Hybrid vectors are engineered DNA molecules combining features of two or more natural vectors to improve cloning efficiency, insert size capacity, and versatility in genetic engineering. They are vital for cloning large DNA fragments and for advanced recombinant DNA technologies.

1. Phagemid Vectors

Definition:

Phagemids are hybrid vectors containing elements of both plasmids and filamentous phage (M13).

Features:

- Contain plasmid origin of replication (ori) for replication in E. coli.

- Contain phage origin of replication (f1 ori) for packaging into single-stranded DNA by helper phage.

- Carry antibiotic resistance genes for selection.

Uses:

- Can propagate as plasmids in bacteria.

- Upon infection with helper phage, phagemids produce single-stranded DNA, useful in site-directed mutagenesis and DNA sequencing.

- Insert size: typically up to 15 kb.

Example:

- pBlue script

2. Phasmid Vectors

Definition:

Phasmids are hybrids of plasmids and bacteriophage λ.

Features:

- Carry plasmid ori for replication in bacteria.

- Carry λ phage sequences allowing packaging into phage particles.

- Can exist as plasmids or as phage particles, combining advantages of both.

Uses:

- Useful for cloning DNA fragments that require phage packaging efficiency.

- Less common than cosmids or phagemids.

3. Cosmid Vectors

Definition:

Cosmids are plasmids engineered to carry λ phage cos sequences, which are required for packaging DNA into phage heads.

Features:

- Contain plasmid ori and selectable markers (e.g., antibiotic resistance).

- Include cos sites from λ phage, allowing packaging into λ phage heads.

- Can carry large inserts: up to 40–45 kb.

Uses:

- Efficient cloning of large DNA fragments.

- Used extensively in genomic library construction.

- Packaged in vitro into phage particles for infection of E. coli.

4. Bacterial Artificial Chromosome (BAC)

Definition:

BACs are plasmid-based vectors derived from the F-factor plasmid of E. coli, designed for cloning very large DNA fragments.

Features:

- Size: Can carry inserts from 100 kb to 300 kb.

- Low copy number (1–2 copies per cell) for stability of large DNA.

- Contain oriS (F plasmid origin) and partitioning system to ensure stable inheritance.

- Carry selectable markers (e.g., chloramphenicol resistance).

Uses:

- Construction of genomic libraries for large genomes.

- Used in genome sequencing projects (e.g., Human Genome Project).

- Stable maintenance of large DNA fragments with minimal rearrangements.

5. Yeast Artificial Chromosome (YAC)

Definition:

YACs are vectors that behave like artificial chromosomes in yeast (Saccharomyces cerevisiae), capable of carrying extremely large DNA inserts.

Features:

- Can carry very large inserts: 100 kb to over 1 Mb.

- Contain essential yeast chromosomal elements:

- ARS (Autonomously Replicating Sequence) for replication.

- CEN (Centromere) for proper segregation during cell division.

- TEL (Telomeres) for chromosome stability.

- Contain selectable markers for yeast (e.g., URA3) and bacterial selection markers.

Uses:

- Cloning of very large DNA fragments.

- Useful in mapping and studying large genomic regions.

- Important for physical mapping and functional analysis of genomes.

Limitations:

- Instability and rearrangement of large inserts sometimes occur.

- More complex handling than bacterial systems.

Summary Table of Hybrid Vectors

Vector Type	Insert Size Capacity	Host	Key Features	Applications
Phagemid	Up to ~15 kb	E. coli + Helper phage	Plasmid + M13 phage ori, single-strand DNA production	DNA sequencing, mutagenesis

----

Vector Type	Insert Size Capacity	Host	Key Features	Applications
Phasmid	Moderate	E. coli + λ phage	Plasmid + λ phage sequences	Cloning requiring phage packaging
Cosmid	35–45 kb	E. coli via phage packaging	Plasmid + λ cos sites	Genomic libraries for large DNA
BAC	100–300 kb	E. coli	F-factor ori, stable low copy	Large genomic libraries, sequencing
YAC	100 kb – >1 Mb	Yeast	ARS, CEN, TEL elements	Very large DNA fragments, genome mapping

----

UNIT- IV

Bioinformatics: Biological Databases

Introduction to Biological Databases

A biological database is a repository that stores biological information such as nucleotide sequences, protein sequences, protein structures, gene expression data, and metabolic pathways. These databases are essential for bioinformatics research, providing organized, searchable, and accessible data to the scientific community.

1. Types of Biological Databases

Biological databases can be broadly classified based on the type of data they store:

Database Type	Description	Examples
Nucleic Acid Sequence Databases	Store DNA and RNA sequences	GenBank, EMBL, DDBJ
Protein Sequence Databases	Store amino acid sequences of proteins	UniProt, Swiss-Prot, TrEMBL
Protein Structure Databases	Store 3D structures of proteins and nucleic acids	PDB (Protein Data Bank)
Gene Expression Databases	Contain data on gene expression profiles under different conditions	GEO (Gene Expression Omnibus), ArrayExpress
Metabolic Pathway Databases	Provide information on biochemical pathways and interactions	KEGG, Reactome

2. Nucleic Acid Sequence Databases

Purpose:

To collect and provide access to nucleotide sequences from diverse organisms.

Key Databases:

- **GenBank (NCBI):**
  - Largest and most comprehensive public nucleotide sequence database.
  - Includes raw sequence data submitted by researchers worldwide.

- **EMBL (European Molecular Biology Laboratory):**

    o European nucleotide sequence database, collaborates with GenBank and DDBJ.

- **DDBJ (DNA Data Bank of Japan):**

    o Japanese nucleotide sequence database, part of the International Nucleotide Sequence Database Collaboration (INSDC) with GenBank and EMBL.

Features:

- Store sequences with annotations such as gene locations, function, and literature references.

- Support sequence retrieval, BLAST searches, and data submission.

3. Protein Sequence Databases

Purpose:

To store and provide access to amino acid sequences of proteins with functional annotations.

Key Databases:

- **UniProt:**

    o Universal Protein Resource; central hub for protein information.

    o Divided into Swiss-Prot (manually curated, high-quality) and TrEMBL (automatically annotated).

- **Swiss-Prot:**

    o High-quality protein sequences with extensive manual annotations, functional information, and cross-references.

- **TrEMBL:**

    o Automated, preliminary protein sequence database supplementing Swiss-Prot.

Features:

- Provide sequence similarity search tools (BLASTP).

- Include functional domains, post-translational modifications, and protein family classifications.

4. Protein Structure Databases

Purpose:

Store experimentally determined 3D structures of proteins and nucleic acids.

Key Database:

- **Protein Data Bank (PDB):**

    o Repository for 3D structural data obtained by X-ray crystallography, NMR spectroscopy, or cryo-electron microscopy.

    o Provides atomic coordinates, structure factors, and biological annotations.

Features:

- Enables visualization of molecular structures.

- Essential for structure-based drug design and functional analysis.

5. Gene Expression Databases

Purpose:

Store gene expression data obtained through microarray, RNA-seq, and other high-throughput techniques.

Key Databases:

- **Gene Expression Omnibus (GEO):**

    o NCBI's public repository for gene expression and other functional genomics data.

- **Array Express:**

    o European database of gene expression data maintained by EMBL-EBI.

Features:

- Store data from different tissues, developmental stages, or experimental treatments.

- Allow meta-analysis and expression profiling.

6. Metabolic Pathway Databases

Purpose:

Provide detailed information on biochemical pathways, metabolic reactions, and molecular interactions.

Key Databases:

- **KEGG (Kyoto Encyclopedia of Genes and Genomes):**

    o Integrates genomic, chemical, and systemic functional information.

    o Includes pathways for metabolism, genetic information processing, and environmental information processing.

- **Reactome:**

  o Curated database of biological pathways with detailed reactions and interactions.

  o Useful for systems biology and pathway analysis.

Summary Table

Database Category	Examples	Data Type	Key Features
Nucleic Acid	GenBank, EMBL, DDBJ	DNA/RNA sequences	Large, annotated nucleotide sequences
Protein Sequence	UniProt, Swiss-Prot, TrEMBL	Protein sequences	Curated sequences, functional annotation
Protein Structure	PDB	3D molecular structures	Atomic-level protein/nucleic acid models
Gene Expression	GEO, ArrayExpress	Expression profiles	High-throughput expression data
Metabolic Pathways	KEGG, Reactome	Pathways & interactions	Detailed biochemical pathways

Important Points for Exam

- Biological databases are vital for managing large-scale biological data.

- Nucleotide and protein sequence databases support sequence alignment and homology studies.

- Protein structure databases enable understanding of molecular function and drug design.

- Gene expression databases facilitate understanding gene regulation.

- Metabolic pathway databases help analyze complex biological systems and interactions.

Bioinformatics: Data Storage, File Formats, Sequence Alignment & Phylo genetics

1. Mode of Data Storage in Bioinformatics

Biological data is stored in digital formats that allow easy access, sharing, and analysis. The data is organized as files in specific standardized formats for sequences, structures, and annotations.

----

2. Common File Formats

a) FASTA Format

- Most widely used format for storing nucleotide and protein sequences.

- Simple text-based format.

- Begins with a header line starting with ">" followed by sequence identifier and description.

- The next lines contain the sequence (A, T, G, C for DNA; 20 amino acids for proteins).

Example:

shell

CopyEdit

>gi|12345|human hemoglobin

ATGCGTACGTTAGCT...

b) Gen Bank Format

- Rich text format used by Gen Bank database.

- Contains detailed annotations: source organism, gene features, coding sequences (CDS), mutations, etc.

- Multi-line and complex but very informative.

c) UniProt Format

- Protein sequence entries with annotations.

- Contains detailed information on protein function, structure, modifications, variants.

- Can be downloaded in plain text or XML formats.

3. Major Biological Databases

a) NCBI (National Centre for Biotechnology Information)

- Central hub for various databases.

- Houses Gen Bank (nucleotide sequences), RefSeq, PubMed, BLAST, etc.

- Offers tools for sequence search, alignment, and analysis.

b) DDBJ (DNA Data Bank of Japan)

- Collaborates with NCBI and EMBL for nucleotide sequences.

- Shares and synchronizes data worldwide.

----

c) PDB (Protein Data Bank)

- Stores 3D structural data of proteins and nucleic acids.

- Provides atomic coordinates, ligands, and experimental details.

4. Sequence Alignment

Purpose:

To identify regions of similarity that may indicate functional, structural, or evolutionary relationships between sequences.

a) Types of Sequence Alignment

i) Local Sequence Alignment

- Finds the best matching region(s) between two sequences.

- Useful when sequences are suspected to share conserved domains or motifs.

- Algorithm example: Smith-Waterman.

ii) Global Sequence Alignment

- Aligns two sequences end-to-end over their entire length.

- Useful for sequences of similar length and high similarity.

- Algorithm example: Needleman-Wunsch.

b) Pairwise vs Multiple Sequence Alignment

Type	Description	Example Tool
Pairwise Alignment	Comparison between two sequences	BLAST, FASTA
Multiple Sequence Alignment (MSA)	Simultaneous alignment of more than two sequences to identify conserved regions	ClustalW, MUSCLE, MAFFT

5. Phylogenetic Trees

Purpose:

To represent evolutionary relationships among organisms or genes based on sequence data.

----

a) Types of Phylogenetic Trees

Type	Description
Rooted Tree	Has a common ancestor (root) representing the evolutionary origin
Unrooted Tree	Shows relationships but not the direction of evolution

b) Approaches for Phylogenetic Tree Construction

Method	Description	Examples
Distance-based Methods	Use pairwise distance matrix to infer trees	Neighbor-Joining, UPGMA
Character-based Methods	Use sequence characters (nucleotides or amino acids) directly	Maximum Parsimony, Maximum Likelihood
Bayesian Methods	Use probabilistic models and Bayesian inference	MrBayes

c) Summary of Methods

Method	Strengths	Weaknesses
Neighbour-Joining	Fast, suitable for large datasets	Less accurate if rate variation is high
UPGMA	Simple, assumes molecular clock	Less accurate if evolutionary rates vary
Maximum Parsimony	Finds simplest tree with minimum changes	Computationally intensive, may give multiple trees
Maximum Likelihood	Statistically robust, models evolution	Computationally expensive
Bayesian Inference	Provides posterior probabilities	Requires good prior knowledge

----

UNIT – V

Protein Structure and Prediction

1. Introduction

Proteins are linear polymers of amino acids that fold into specific three-dimensional structures essential for their biological functions. Understanding protein structure is crucial for insights into protein function, interaction, and for drug design.

2. Hierarchy of Protein Structure

Protein structure is organized at four distinct levels:

a) Primary Structure

- The linear sequence of amino acids in a polypeptide chain.

- Determined by the gene sequence encoding the protein.

- Connected by peptide bonds.

- Determines all higher levels of structure.

- Changes or mutations here can drastically affect protein function.

b) Secondary Structure

- Regular, localized folding patterns stabilized mainly by hydrogen bonds between backbone atoms.

- Common types:

    - α-helix: Right-handed coil stabilized by hydrogen bonds between the carbonyl oxygen of one residue and the amide hydrogen of the residue four positions ahead.

    - β-sheet: Polypeptide chains aligned side-by-side, connected by hydrogen bonds; can be parallel or antiparallel.

    - Turns and loops: Connect helices and sheets; often involved in functional sites.

c) Tertiary Structure

- The overall 3D folding of a single polypeptide chain.

- Stabilized by interactions between side chains (R-groups):

    - Hydrophobic interactions

    - Hydrogen bonds

    - Ionic bonds (salt bridges)

- o Disulfide bridges (covalent bonds between cysteine residues)

- Determines the protein's functional conformation.

- Can include distinct structural domains.

d) Quaternary Structure

- The assembly of multiple polypeptide chains (subunits) into a functional protein complex.

- Subunits may be identical or different.

- Stabilized by similar interactions as tertiary structure.

- Examples: Hemoglobin (tetramer), DNA polymerase (multi-subunit).

3. Protein Structure Prediction

Protein structure prediction aims to determine the 3D structure of a protein from its amino acid sequence.

a) Methods of Structure Prediction

Method	Description	Advantages	Limitations
Homology (Comparative) Modelling	Uses known structures of homologous proteins as templates	High accuracy if good template available	Requires related structure
Threading (Fold Recognition)	Fits the sequence onto known structural folds even if no clear homology	Useful for distant relationships	Less accurate than homology modelling
Ab initio (De novo) Prediction	Predicts structure from physical principles without templates	Can predict novel folds	Computationally intensive, less accurate

b) Tools for Structure Prediction

- SWISS-MODEL: Automated homology modelling.

- Phyre2: Combines homology and threading.

- Rosetta: Ab initio structure prediction.

- Alpha Fold: AI-based highly accurate prediction of protein structures.

----

4. Importance of Protein Structure

- Structure determines function.

- Insight into enzyme active sites, binding pockets.

- Helps in drug design and understanding disease mechanisms.

- Mutations affecting structure can lead to loss of function or gain of harmful function.

Summary

Level	Description	Stabilizing Forces	Functional Relevance
Primary	Amino acid sequence	Peptide bonds	Determines all higher structures
Secondary	α-helix, β-sheet	Hydrogen bonds (backbone)	Local folding patterns
Tertiary	3D folding of single chain	Side-chain interactions	Functional 3D conformation
Quaternary	Assembly of multiple chains	Non-covalent and disulfide bonds	Multi meric protein function

Protein Structure, Prediction, and In Silico Drug Design

1. Primary, Secondary, and Tertiary Structures

a) Primary Structure

- Linear sequence of amino acids linked by peptide bonds.

- Encoded directly by the gene.

- Determines protein folding and function.

b) Secondary Structure

- Local spatial arrangement of polypeptide backbone.

- Stabilized by hydrogen bonds between backbone amide and carbonyl groups.

- Main types:

  o α-helix: Right-handed coil, hydrogen bonding every 4th residue.

  o β-sheet: Parallel or antiparallel strands connected by hydrogen bonds.

  o Turns and loops: Connect secondary structures; often on protein surface.

----

c) Tertiary Structure

- Overall 3D structure of a single polypeptide chain.

- Stabilized by interactions among side chains (R-groups):

    o Hydrophobic interactions drive folding by excluding water.

    o Hydrogen bonds, ionic bonds (salt bridges), and disulfide bridges provide additional stability.

- Determines protein's biological function.

2. Motifs, Folds, and Domains

a) Motifs

- Simple combinations of secondary structures forming recognizable patterns.

- Examples:

    o Helix-turn-helix: DNA-binding motif.

    o β-α-β motif: Common in enzymes.

- Motifs are smaller than domains and often part of functional sites.

b) Folds

- The overall arrangement of secondary structures in 3D space.

- Represents a recurring architecture shared by different proteins.

- Examples:

    o Rossmann fold: Nucleotide-binding.

    o TIM barrel: Common enzyme fold.

c) Domains

- Independently folding, stable units within a protein.

- Can function autonomously and often have specific functions.

- Proteins can be single-domain or multi-domain.

- Domains can evolve, shuffle, and recombine.

3. Protein Structure Prediction

a) Importance

- Experimental structure determination is time-consuming and costly.

- Prediction helps understand function, interactions, and design drugs.

b) Methods

Method	Description	Pros	Cons
Homology Modelling	Uses known structures of homologous proteins as templates to predict unknown structure	High accuracy if template is closely related	Requires suitable template
Threading (Fold Recognition)	Aligns sequence to known folds without close homology	Useful for distant relationships	Lower accuracy
Ab Initio Modelling	Predicts structure from sequence based on physical principles	Can predict novel folds	Computationally intensive, less accurate

c) Homology Modelling Steps

1. Identify homologous template protein with known structure.

2. Align target sequence with template.

3. Build backbone model based on template.

4. Add side chains and loops.

5. Refine model energetically.

6. Validate model quality.

4. In Silico Drug Discovery and Design

a) Overview

- Computational methods used to identify and optimize drug candidates.

- Saves time and cost compared to traditional experimental methods.

b) Key Steps

1. Target Identification: Protein involved in disease.

2. Structure Determination: Experimental or predicted protein structure.

3. Virtual Screening: Docking large libraries of compounds to protein binding sites.

4. Molecular Docking: Predict binding affinity and mode of ligands.

----

5. Lead Optimization: Improve potency, selectivity, and pharmacokinetic properties.

6. ADMET Prediction: Assess absorption, distribution, metabolism, excretion, and toxicity.

c) Tools and Techniques

- Molecular Docking Software: AutoDock , Glide.

- Pharma cophore Modelling: Identify essential features for activity.

- QSAR (Quantitative Structure-Activity Relationship): Correlate chemical structure with biological activity.

- Molecular Dynamics: Simulate ligand-protein interactions over time.

Summary Table

Concept	Description	Importance
Primary Structure	Amino acid sequence	Basis of folding & function
Secondary Structure	α-helix, β-sheet, loops	Local folding patterns
Tertiary Structure	3D folding of polypeptide	Functional conformation
Motifs	Small conserved patterns	Functional sites
Folds	Overall 3D arrangement	Structural classification
Domains	Independent folding units	Functional & evolutionary units
Homology Modeling	Structure prediction using templates	Widely used, accurate with templates
In Silico Drug Design	Computational drug discovery	Cost-effective, speeds up drug development