YOUR MENTOR GURU

# CSIR NET Life Sciences & Biotechnology — Unit 11

## Evolution and Behavior

Study & revision companion | Topic explanations, core facts, and active-recall questions

## How to use this guide

Read each topic, explain the key mechanism in your own words, then answer the self-check prompts without looking at the notes. Use your detailed textbook or class notes for full derivations and additional examples.

## Topic overview


1. 1. Biological Databases and File Formats
2. 2. Pairwise Alignment and Scoring
3. 3. BLAST and Sequence Similarity
4. 4. Phylogenetic Inference
5. 5. Genomics and Transcriptomics
6. 6. Protein Structure and Molecular Modeling
7. 7. Statistics, Programming and Machine Learning

## 1. Biological Databases and File Formats

GenBank and ENA archive nucleotide sequences; UniProt provides protein sequence and functional annotations.

FASTA stores sequences with a header; FASTQ adds per-base quality scores for sequencing reads.

PDB archives experimentally determined macromolecular structures and related metadata.

Accession identifiers provide stable database references, while annotation versions can change.

Sequence, expression and structure databases store different types of biological evidence.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: a sequence match does not prove identical function.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 2. Pairwise Alignment and Scoring

Global alignment spans sequences end-to-end; local alignment finds high-scoring matching regions.

Dynamic programming underlies Needleman–Wunsch and Smith–Waterman algorithms.

Substitution matrices such as PAM and BLOSUM encode empirical amino-acid substitution scores.

Gap-open and gap-extension penalties reflect different evolutionary assumptions.

Local and global alignments answer different questions; BLAST uses heuristic local similarity searches.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: consider E-value, identity and query coverage together.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 3. BLAST and Sequence Similarity

BLAST searches for local similarities using heuristic seed-and-extend approaches.

E-value estimates the expected number of chance hits with at least that score under the search model.

Lower E-values indicate stronger statistical evidence, but database size affects the value.

Identity, coverage and biological context must be considered together.

Orthologs arise through speciation and paralogs through duplication; phylogenetic trees infer relationships.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: homology is not expressed as a percentage.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 4. Phylogenetic Inference

Multiple sequence alignment identifies putatively corresponding sites before tree inference.

Neighbor joining is distance-based; maximum likelihood compares models and trees using sequence probabilities.

Bootstrap resampling assesses stability under resampled alignments.

Long-branch attraction and model misspecification can mislead inferred relationships.

Coding regions, regulatory features and variants can be predicted using sequence, structure and comparative evidence.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: predicted pathogenicity requires validation.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 5. Genomics and Transcriptomics

Short-read and long-read sequencing have different error, length and assembly tradeoffs.

RNA-seq quantifies transcript abundance and differential expression under statistical models.

Read mapping aligns sequencing reads to references; de novo assembly reconstructs sequences without a complete reference.

Normalization and multiple-testing correction are essential in differential-expression analysis.

Homology modeling, threading and AI-based prediction estimate protein structures with variable confidence.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: a predicted structure is not a validated mechanism.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 6. Protein Structure and Molecular Modeling

Primary sequence influences fold; secondary structures include helices and sheets.

Structure prediction outputs models whose confidence varies by region and context.

Docking explores candidate binding poses but scoring errors can limit affinity prediction.

Molecular dynamics integrates approximate equations of motion using force fields and finite time steps.

Force fields approximate molecular interactions; simulations explore conformational change over time.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: simulations depend on assumptions and sampling.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## 7. Statistics, Programming and Machine Learning

Training, validation and test sets serve distinct roles; data leakage inflates apparent performance.

Precision = TP/(TP+FP); recall = TP/(TP+FN); specificity = TN/(TN+FP).

Cross-validation estimates generalization under appropriate sampling assumptions.

Class imbalance can make accuracy misleading; use precision–recall metrics when relevant.

Docking, QSAR, pharmacophores and ADMET predictions prioritize compounds for testing.

When solving a question, distinguish the underlying mechanism from its observable outcome. Compare the normal situation with the altered condition, and check whether the evidence supports a direct effect or only an association. The central distinction to retain is: a favorable docking score does not prove efficacy.

Self-check: Explain the mechanism or relationship above; identify one common misconception; predict the effect of changing a key variable.

## Active-recall practice

1. Explain the essential principles of 1. Biological Databases and File Formats and give one experimental or real-world application.

2. Explain the essential principles of 2. Pairwise Alignment and Scoring and give one experimental or real-world application.

3. Explain the essential principles of 3. BLAST and Sequence Similarity and give one experimental or real-world application.

4. Explain the essential principles of 4. Phylogenetic Inference and give one experimental or real-world application.

5. Explain the essential principles of 5. Genomics and Transcriptomics and give one experimental or real-world application.

6. Explain the essential principles of 6. Protein Structure and Molecular Modeling and give one experimental or real-world application.

7. Explain the essential principles of 7. Statistics, Programming and Machine Learning and give one experimental or real-world application.

## Final revision checklist

- 1. Biological Databases and File Formats
- 2. Pairwise Alignment and Scoring
- 3. BLAST and Sequence Similarity
- 4. Phylogenetic Inference
- 5. Genomics and Transcriptomics
- 6. Protein Structure and Molecular Modeling
- 7. Statistics, Programming and Machine Learning

## Further study

Use the corresponding unit notes, topic-wise detailed lessons, MCQs, practice tests and full-length mock tests on yourmentorguru.com. This compact PDF is a revision companion, not a replacement for full syllabus textbooks.