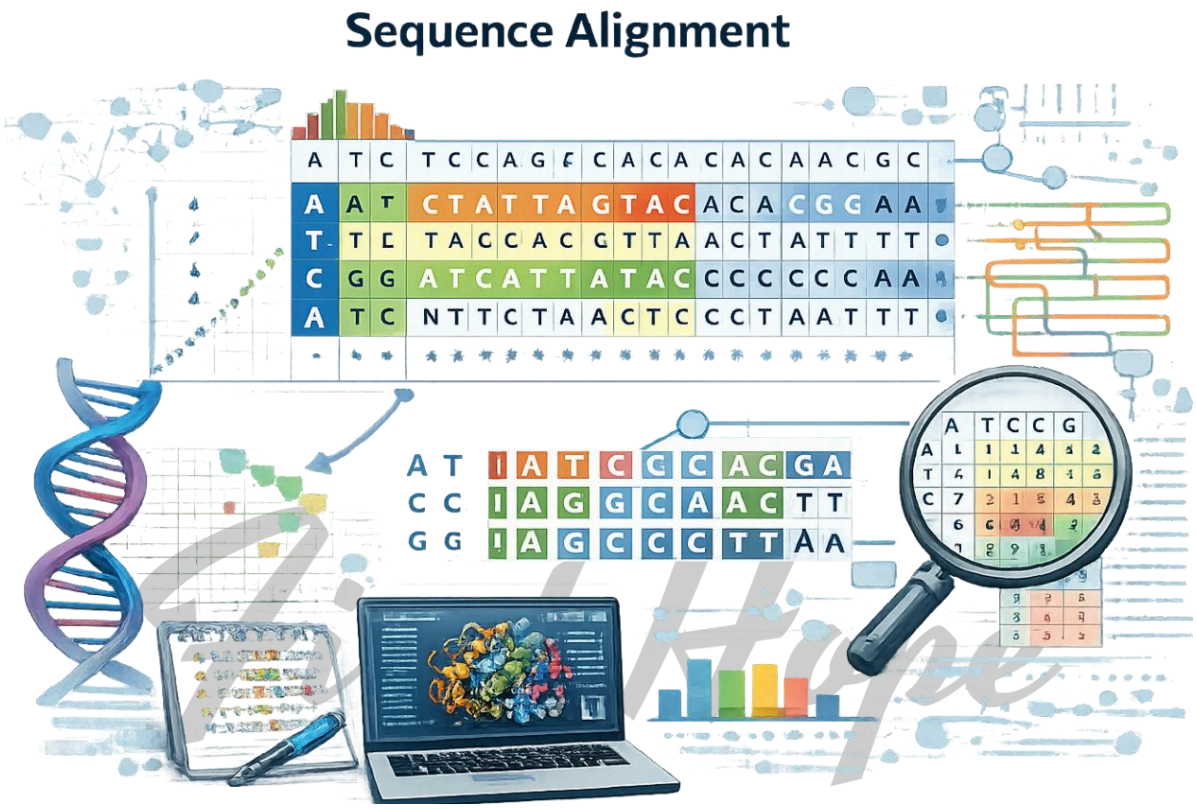
Sequence Analysis

Sequence Alignment

Definition of Sequence Alignment

- Sequence alignment is the process of **arranging two or more biological sequences** (DNA, RNA, or protein) to identify **regions of similarity**.



- **These similarities may indicate:**
  - Functional relationships
  - Structural similarities
  - Evolutionary relationships

Purpose of Sequence Alignment

- Identify conserved regions in sequences
- Study evolutionary relationships
- Predict function of unknown genes or proteins
- Detect mutations and variations

Types of Sequence Alignment

1. Global Alignment

- Aligns entire length of sequences
- Best for sequences of similar length and high similarity

Copyright by Engacy Learning Private Limited. All rights reserved. Reproduction or distribution without written consent is strictly prohibited.

2. Local Alignment

- Aligns only similar regions within sequences
- Useful when sequences are different in length

Basic Concepts in Alignment

- **Match**: identical residues
- **Mismatch**: different residues
- **Gap**: insertion or deletion in sequence

Scoring in Alignment

1. Match Score

- Given when two residues are identical.

2. Mismatch Penalty

- Given when residues differ.

3. Gap Penalty

- Applied when gaps are introduced.

Importance of Sequence Alignment

- Identifying homologous genes
- Predicting protein function
- Studying evolutionary relationships
- Detecting mutations

First Hope

Pairwise Alignment Techniques

Definition

- Pairwise alignment compares **two sequences at a time** to find the best possible alignment.

Types of Pairwise Alignment

1. Global Pairwise Alignment

- Aligns sequences from start to end
- Best for closely related sequences
- **Algorithm Used**: Needleman–Wunsch algorithm

2. Local Pairwise Alignment

- Finds the best matching region within sequences

Copyright by Engacy Learning Private Limited. All rights reserved. Reproduction or distribution without written consent is strictly prohibited.

- Useful for distantly related sequences
- **Algorithm Used**: Smith–Waterman algorithm

Common Algorithms

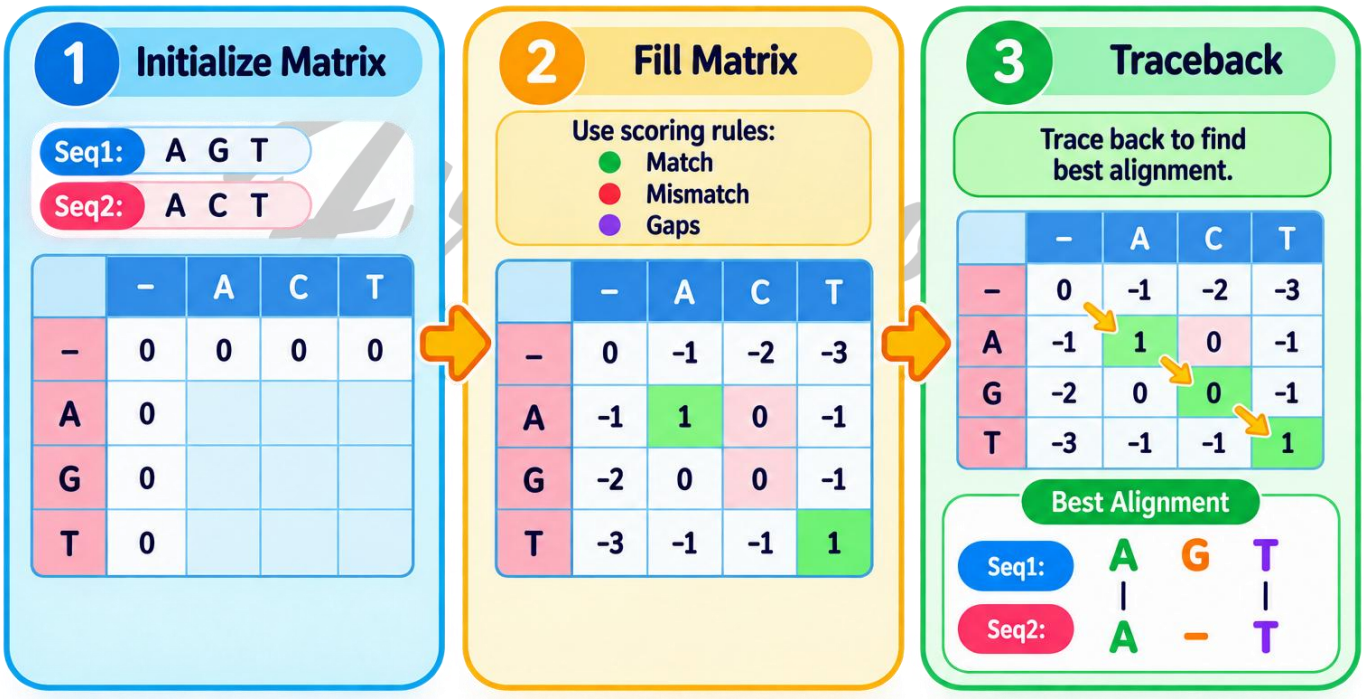
Needleman–Wunsch Algorithm

- Used for global alignment
- Based on dynamic programming
- Gives the best alignment across the full sequence

Smith–Waterman Algorithm

- Used for local alignment
- Finds highly similar regions within sequences
- Also uses dynamic programming

Steps in Pairwise Alignment



1. Initialize a scoring matrix.
2. Fill the matrix using scoring rules (match, mismatch, gaps).
3. Perform traceback to find the best alignment.

Scoring Matrices

- **Used to assign scores for matches and mismatches**
  - PAM (Point Accepted Mutation)
  - BLOSUM (Blocks Substitution Matrix)

Applications

Copyright by Engacy Learning Private Limited. All rights reserved. Reproduction or distribution without written consent is strictly prohibited.

- Identifying homologous sequences
- Studying mutations
- Functional prediction of genes

Advantages

- Simple and precise for two sequences
- Provides detailed comparison

Limitations

- Not suitable for large number of sequences
- Time-consuming for long sequences

Multiple Sequence Analysis

Definition

- **Multiple sequence analysis refers to:**
  - The process of examining more than two biological sequences together
  - Identifying patterns, similarities, and differences across a group of sequences
- **It is broader than alignment:** Includes alignment, comparison, interpretation, and biological inference

First Hope

Basic Concept

- Instead of comparing only two sequences, many sequences are analyzed together
- Regions that are similar across all sequences are called **conserved regions**
- These regions often indicate important structural or functional roles

Purpose of Multiple Sequence Analysis

- To identify conserved regions across many sequences
- To study evolutionary relationships among organisms
- To detect functional domains in proteins
- To analyze gene families

Key Concepts

A. Conserved Regions

- Regions that remain unchanged across sequences
- Indicate important functional or structural roles

Copyright by Engacy Learning Private Limited. All rights reserved. Reproduction or distribution without written consent is strictly prohibited.

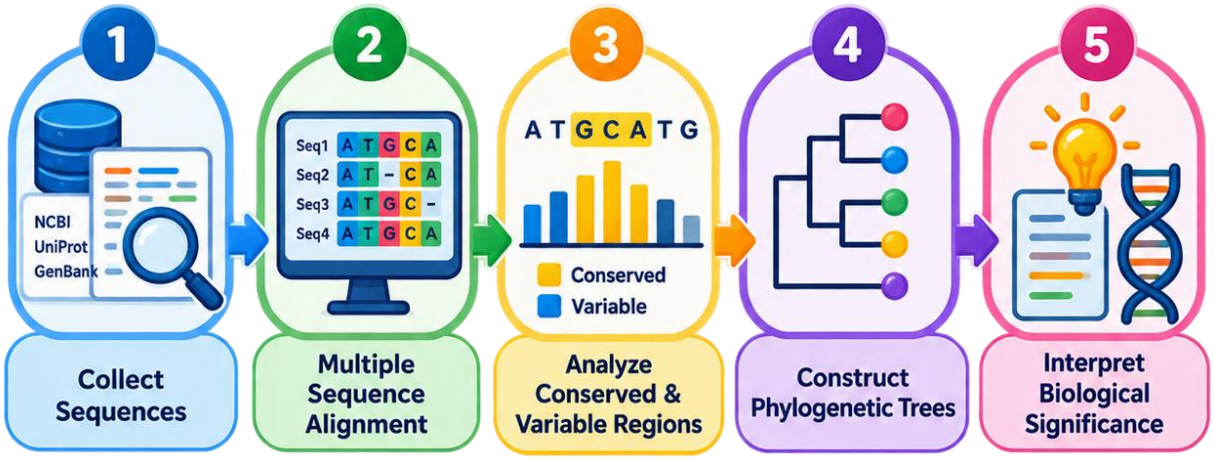
B. Consensus Sequence

- A sequence that represents the most common residues at each position

C. Sequence Homology

- Similarity due to common ancestry

Steps in Multiple Sequence Analysis



- Collect related sequences from databases
- Perform multiple sequence alignment (MSA)
- Analyze conserved and variable regions
- Construct phylogenetic trees
- Interpret biological significance

First Hope

Key Features

- Detects sequence conservation
- Identifies important motifs and domains
- Helps in building phylogenetic relationships

Methods Used

A. Progressive Methods

- Align sequences step by step
- Start with the most similar sequences

B. Iterative Methods

- Repeatedly improve the alignment for better accuracy

Applications

- Phylogenetic analysis
- Identification of functional domains

Copyright by Engacy Learning Private Limited. All rights reserved. Reproduction or distribution without written consent is strictly prohibited.

First Hope

FREE PREVIEW

Click here to visit our website

PHARMACY PREMIUM NOTES

Complete subject and semesters are available on FirstHope app only.

Scan or click here

Click here to get the app >

Available on Google Play

JOIN THE FIRSTHOPE COMMUNITY

Join our Telegram channel

Click here to join Telegram >

Scan or click here

@FirstHopeNotes

Instagram

@firsthope.official

Click here to follow >

YouTube channel

FirstHope-Pharmacy / Live streams

Click here to watch >

WhatsApp channel

Follow FirstHope on WhatsApp

Click here to join >

FirstHope website

firsthope.co.in

Click here to visit >

B.Pharm / D.Pharm / M.Pharm / Pharm.D

Preview notes. Complete subjects and semesters in the FirstHope app.