**Multiple Sequence Alignment (MSA)**

**Why Multiple Sequence Alignment?**

* Pairwise = only 2 sequences.
* Biology often needs **dozens/hundreds** of sequences aligned at once.
* Applications:
  + Identify **conserved motifs** (functionally important regions).
  + Study **evolutionary relationships**.
  + Assist in **protein structure prediction**.

**Analogy**:

Pairwise = comparing 2 essays.

MSA = aligning paragraphs from **many essays** to find the shared theme.

**Principles of MSA**

1. **Progressive Alignment**
   * Start with the **most similar pair**.
   * Add sequences step by step to build the alignment.
   * Example tool: **Clustal Omega**.
2. **Iterative Refinement**
   * Start with a guess alignment.
   * Re-align sequences multiple times for improvement.
   * Example tool: **MUSCLE**.

**Interpreting MSA**

* **Conserved regions** (same letters across sequences) → often functional/structural importance.
* **Gaps** → insertions or deletions in evolution.
* **Consensus sequence** → the most common residues across sequences.

👉 **Example (Protein snippet):**

Seq1: ATGCTAGC

Seq2: ATG--AGC

Seq3: ATGCTTGC

Consensus: ATGCTAGC

**Lab Session: Hands-on MSA**

**Objective**: Perform MSA using online tools.

1. Collect 4–5 hemoglobin protein sequences from **NCBI Protein**.
2. Use **Clustal Omega (EMBL-EBI)**:
   * Paste sequences.
   * Run alignment.
3. Observe output:
   * Conserved amino acids marked with “\*”.
   * Similar residues with “: or .”.
4. Compare with **MUSCLE** results.

Task:

* Identify the most conserved stretch in hemoglobin sequences.
* Discuss why that region might be functionally important.

**📝 Quick Review Questions**

1. Why do we need multiple sequence alignment instead of just pairwise?
2. What is the difference between progressive and iterative alignment?
3. How can conserved regions in MSA guide us in protein function studies?