

Sequence Alignment Basics

Sequence alignment is the process of arranging **DNA, RNA, or protein sequences** to identify **regions of similarity**. These similarities may indicate:

- **Evolutionary relationships** between organisms.
- **Functional or structural similarities** between genes/proteins.
- **Predicted functions** of unknown sequences.

Imagine comparing two essays. You align sentences **word by word** to see which parts are similar or different. Sequence alignment works in the same way for biological sequences.

Types of Sequence Alignment

i. Global Alignment

- Aligns the **entire length** of two sequences.
- Best for sequences that are **similar in length**.
- **Algorithm:** Needleman–Wunsch.

Compare two full DNA barcodes to see overall similarity.

Algorithm: Needleman–Wunsch.

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

0	-2
-2	0

-4	4	1
----	---	---

-2	-4
1	0

-1	-6	-3
----	----	----

Result:

A	T	G	C	T
A	-	G	C	T

ii. Local Alignment

- Finds the **best matching region** between two sequences.
- Useful when one sequence is **much longer** than the other.
- **Algorithm:** Smith–Waterman.

Find a matching paragraph in two different books.

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1	0	0	0	0
G	0	0	0	1	0	0
C	0	0	0	0	2	0
T	0	0	1	0	0	3

0	0
0	0

-2	-2	1
----	----	---

0	0	1
---	---	---

0	0
1	0

-1	-2	-1
----	----	----

0	0	0
---	---	---

[negative values becomes zero]

Result: [Choose the values that are greater than 1 in the matrix, should move up to 0]

A	T	G	C	T
	A	G	C	T

3. Scoring Matrices

To quantify similarity and penalize mismatches or gaps.

Rules (simple scoring example)

- **Match:** +1
- **Mismatch:** -1
- **Gap (insertion/deletion):** -2

Example:

Align sequences

A T G C

with

A T T C

Common Matrices

- **PAM (Point Accepted Mutation)** : Measures evolutionary distance.
- **BLOSUM (BLOCKS Substitution Matrix)** : Measures amino acid substitutions; widely used for proteins.

Use **PAM** for closely related sequences, **BLOSUM** for more divergent protein sequences.

Practical Session: Pairwise Alignment (Basics)

Perform pairwise sequence alignment manually and with tools.

1. Take two short DNA sequences:

Seq1: ATGCT

Seq2: ATTT

2. Try to align them manually (introduce gap).
3. Calculate score using simple rules:
 - o Match = +1, Mismatch = -1, Gap = -2.
4. Use [EMBOSS Needle](#) (global)
and [EMBOSS Water](#) (local).

Tasks:

- Compare results of Needle vs Water.
- Which alignment gives a better score for partial matches?

Quick Review Questions

1. What's the main difference between global and local alignment?
2. Why do we need scoring matrices?
3. Which matrix (PAM or BLOSUM) is better for protein sequence alignment?