

هم‌ترازی دنباله

Sequence Alignment



DNA, Nucleotides, and Gene

□ Nucleotide

- The basic building block of DNA.
- Consists of a nitrogenous base (A, T, G, C), a sugar (deoxyribose), and a phosphate group.
- Multiple nucleotides link together to form DNA strands.

□ DNA

- A long molecule made up of many nucleotides arranged in a double-helix structure.
- Stores genetic information in the sequence of its nucleotides.
- Found in the nucleus of cells and carries instructions for all biological processes.

□ Gene

- A specific section of DNA that codes for a particular protein.
- Determines hereditary traits and controls cellular functions.
- Can be turned on or off to regulate biological activity.

DNA, Nucleotides, and Gene

- **Nucleotides** are like letters.
- **DNA** is the book they create.
- **Genes** are the specific sentences in that book that hold meaningful instructions.

Nucleotide

Each nucleotide has a **nitrogenous base** that pairs **only** with its complementary base:

- ▣ **Adenine (A) always pairs with Thymine (T)**
- ▣ **Guanine (G) always pairs with Cytosine (C)**

This strict pairing maintains the **double-helix structure** and ensures genetic information is copied correctly during DNA replication.

A diagram of a DNA double helix. The two strands are blue, and the base pairs are represented by vertical lines in red, green, and black. A section of the helix is unwound to show the base sequence. In the top right corner, there is a blue icon of a chromosome (a thick X-shape) with the word "Chromosome" written below it.

Sequence Alignment Algorithms

- They are essential tools in **bioinformatics**, used to compare and analyze biological sequences like **DNA**.
- These algorithms help identify similarities, detect evolutionary relationships, and find conserved regions in sequences.

Types of Sequence Alignment

□ Global Alignment

- Aligns entire sequences from start to end.
- Used when sequences are **similar in length and content**.
- Example: **Needleman-Wunsch algorithm**.

□ Local Alignment

- Focuses on the **most similar regions** within sequences.
- Useful for **finding conserved motifs or functional domains**.
- Example: **Smith-Waterman algorithm**.

LCS vs Needleman-Wunsch algorithm

- ❑ The **Longest Common Subsequence (LCS) algorithm** and **Needleman-Wunsch algorithm** both deal with sequence comparison, but they differ in their goals and applications.
- ❑ In LCS, Gaps are ignored. But Gaps have penalties in Needleman-Wunsch algorithm

LCS

The LCS problem can be broken down into **three cases**:

1. **Characters match:** If the last characters of both sequences match, they are part of the LCS:

$$LCS(X[0..m], Y[0..n]) = LCS(X[0..m-1], Y[0..n-1]) + 1$$

2. **Characters don't match:** If the last characters are different, we take the **maximum** of two possibilities:

$$LCS(X[0..m], Y[0..n]) = \max(LCS(X[0..m-1], Y[0..n]), LCS(X[0..m], Y[0..n-1]))$$

3. **Base case:** If either sequence is empty, LCS is 0:

$$LCS("", Y) = 0, \quad LCS(X, "") = 0$$

LCS Implementation

```
import numpy as np
def lcs_matrix(X, Y):
    m, n = len(X), len(Y)
    # Create DP table (matrix)
    dp = np.zeros((m + 1, n + 1), dtype=int)
    # Fill the matrix using the LCS recurrence relation
    for i in range(1, m + 1):
        for j in range(1, n + 1):
            if X[i - 1] == Y[j - 1]: # Matching characters
                dp[i][j] = dp[i - 1][j - 1] + 1
            else:
                # Choosing the best from previous
                dp[i][j] = max(dp[i - 1][j], dp[i][j - 1])
```

Example 1

			L	O	N	G	E	S	T
		0	1	2	3	4	5	6	7
	0	0	0	0	0	0	0	0	0
S	1	0	0	0	0	0	0	1	1
T	2	0	0	0	0	0	0	1	2
O	3	0	0	1	1	1	1	1	2
N	4	0	0	1	2	2	2	2	2
E	5	0	0	1	2	2	3	3	3
				O	N		E		

Example 2

		a	b	c	d	a	f
	0	0	0	0	0	0	0
a	0	1	1	1	1	1	1
c	0	1	1	2	2	2	2
b	0	1	2	2	2	2	2
c	0	1	2	3	3	3	3
f	0	1	2	3	3	3	4

Red arrows indicate a path through the grid, starting from the top-left cell (0,0) and ending at the bottom-right cell (7,5). The path follows the sequence of cells: (0,0) → (1,1) → (2,2) → (3,3) → (4,4) → (5,4) → (6,4) → (7,5).