# Sequence Alignment using Dynamic Programming

## *Introduction:*

In bioinformatics, understanding the similarity of gene sequences is critical. The Sequence Alignment Problem tackles this issue by calculating the closeness of two sequences. This report describes a dynamic programming algorithm for determining the optimal alignment of two gene sequences given a score matrix.

X=TCCCAGTTATGTCAGGGGACACGAGCATGCAGAGAC
Y= AATTGCCGCCGTCGTTTTCAGCAGTTATGTCAGATC

| δ | A | G | T | C | - |
|---|---|---|---|---|---|
| A | 1 | -0.8 | -0.2 | -2.3 | -0.6 |
| G | -0.8 | 1 | -1.1 | -0.7 | -1.5 |
| T | -0.2 | -1.1 | 1 | -0.5 | -0.9 |
| C | -2.3 | -0.7 | -0.5 | 1 | -1 |
| - | -0.6 | -1.5 | -0.9 | -1 | n/a |

## *Algorithm Overview:*

The algorithm uses dynamic programming to generate a matrix, dpMatrix, of size (n+1) x (m+1), where n and m are the lengths of sequences x and y. Each cell indicates the highest possible score for aligning the matching x and y prefixes. Iteratively traversing the matrix, the algorithm considers three operations: match, delete, and insert, accumulating scores based on the provided scoring matrix.

The time complexity is O(mn), achieved through nested loops iterating over the entire dynamic programming matrix.

## *Implementation:*

The Python implementation defines the sequence_alignment function, which takes as inputs the sequences x and y, as well as a score matrix. The function returns the alignment with the highest score, as well as the aligned sequences and the alignment score. The alignment is determined by backtracking through dpMatrix, constructing the final aligned sequences with gap characters ("-").

## *Performance:*

A scoring matrix and a test case with the sequences x and y verify the implementation. The program outputs aligned sequences and an alignment score of 7.1, aligning sequences correctly. The aligned sequence (result) are:

```
Aligned X: ---TCCCAGTTATGTCAGGGGACACG-AG-CATG-CAGAGAC
Aligned Y: AATTGCC-G-C-CGTC-GTTTTCA-GCAGTTATGTCAGAT-C
```

## *Conclusion:*

For sequence alignment, the dynamic programming technique offers an effective solution (O(mn)). In bioinformatics, this method is frequently employed to compare gene sequences, which helps researchers comprehend evolutionary relationships.