

## Mini-Projet : Bio-Algorithmique

### Alignement multiple de séquences

#### Indications

Travail à effectuer **en binôme (ou monôme)**

A envoyer au plus tard le **samedi 18 mai 2024** à l'adresse [recupspace@gmail.com](mailto:recupspace@gmail.com)

Le travail (compte rendu) doit être tapé dans un éditeur et enregistré au **format pdf**

Nom de fichier = « MiniProjetBioAlgo-nomsEtudiants.pdf »

L'objet du mail : « MiniProjet-BioALGO »

L'objectif de ce mini-projet est de se pencher sur la question d'alignement de séquences (biologiques), notamment l'alignement multiple (plus de deux séquences), analyser la performance et la complexité des méthodes utilisées.

Deux types d'alignement existent : un alignement global et un alignement local. De plus, on parle d'alignement par paire (entre deux séquences) ou d'alignement multiple (plus de deux séquences)

#### Alignement par paire

- Expliquer le principe de l'algorithme de **Needleman et Wunsch** (vu en cours)
- Montrer des exemples d'alignement de **deux séquences**
- Motiver l'utilisation de l'alignement par paires de séquences (cas d'applications)
- Implémenter l'algorithme (utiliser la matrice de similarité BLOSUM62 et Indel = -2). Afficher le score de l'alignement (sur le chemin optimal) et le résultat de l'alignement.
- Effectuer plusieurs tests sur des séquences de longueur 20, 30, 50, 100, ... Montrer quelques résultats d'alignement ainsi que le score de l'alignement.
- Evaluer la complexité expérimentale (temps d'exécution) et reporter les résultats dans un tableau.

#### Alignement multiple

- Définir la notion d'alignement multiple de séquences
- Montrer des exemples d'alignement de **trois / quatre / cinq séquences**
- Motiver l'utilisation de l'alignement multiple de séquences (cas d'applications)

#### Méthodes d'alignement multiple

Les méthodes de l'alignement multiple suivent différentes approches dont : *l'approche exacte* (programmation dynamique, ...), *l'approche progressive* (notion de profil) et *l'approche itérative* (DiAlign, SAGA, ...).

- Expliquer le principe de chaque approche et dire quelles sont ses limites
- **Implémenter une méthode d'alignement multiple** selon l'approche progressive (**utiliser la notion de profil**). Afficher les alignements obtenus progressivement.
- Effectuer plusieurs tests, en gardant la même longueur de séquences (exemple L= 50) et en augmentant le nombre de séquences 3, 4, 5, ... 10. Montrer les résultats des alignements.

- Reportez les résultats des expérimentations (temps d'exécution) dans un tableau et commentez.
- Effectuez d'autres tests en augmentant les tailles des séquences ( $L = 100, 150, 200, \dots$ ) ; reportez les résultats dans un tableau
- Conclure par rapport aux limites de la méthode (optimalité des résultats, nombre maximum de séquences, exactitude, performance, ...)

#### **Autre méthode d'alignement multiple**

- **Implémenter une autre méthode d'alignement multiple** selon une autre approche
- Effectuer plusieurs tests, en gardant la même longueur de séquences (exemple  $L = 50$ ) et en augmentant le nombre de séquences 3, 4, 5, ... 10.
- Conclure par rapport aux limites de la méthode (optimalité des résultats, nombre maximum de séquences, exactitude, performance, ...)
- Comparer les deux méthodes implémentées

#### **Remettre un rapport avec la structure suivante :**

- Une page de garde (en tête, titre, noms, ...)
- Une introduction dans laquelle on présente la bio-informatique et les objectifs du projet
- Notions de base : séquences biologiques, ADN, ARN, ... l'alignement de séquences, exemples, les matrices de similarité, exemples,
- Alignement par paire de séquences : suivre les points donnés dans le sujet (algorithme, tests, analyse ...)
- Alignement multiple de séquences : suivre les points donnés dans le sujet (différentes approches, algorithme progressif, tests, analyse, autre méthode, ...)
- Une conclusion.
- **Une annexe comportant le code source (bien commenté) des algorithmes**