

# Bioinformatique :

## Principes de base de l'alignement de séquences

Une des utilisations de la bioinformatique consiste à comparer des séquences (d'acides aminés pour les protéines ou d'acides nucléiques).

Les buts sont nombreux, on peut donner comme rôles :

- Identifier au sein d'une banque une séquence obtenue en laboratoire de biologie.
- Localiser une séquence d'acide nucléique au sein du génome d'un organisme
- Identifier un rôle à une molécule séquencée par comparaison avec des molécules de fonctions similaires déjà répertoriées.
- Réaliser une étude phylogénétique
- Prédire la structure secondaire (tertiaire) d'une protéine

### Principe de l'alignement

**Aligner des séquences** = Rechercher le maximum d'appariements entre les résidus des séquences comparées. L'alignement est d'autant plus parfait qu'il n'y a pas de mésappariements et de brèches.

**Mesure du degré de similitude:** La plupart des méthodes d'alignement de séquences biologiques, et en particulier les méthodes d'alignement de séquence de protéines cherchent à optimiser un score d'alignement. Ce score est relié au taux de similarité entre les deux séquences comparées.

$$\text{Score} = \Sigma \text{ scores élémentaires} - \Sigma \text{ pénalités}$$

Exemple de détermination de score avec une matrice unitaire (appariement vaut +1, mésappariement vaut 0 et une brèche vaut -1)

Alignement sans brèches	Alignement avec brèches
Séquence 1    ATGACTGGGCCACT     . . . .     .   .       Séquence 2    A T A C T G G G A C A A C T	Séquence 1    ATGACTGGGCC-ACT                 .           Séquence 2    AT-ACTGGGACA A C T
8 appariements (« match ») et 6 mésappariements (« mismatch ») score= 8-0 =8	12 appariements et 1 mésappariements et 2 brèches score= 12-2 =10

Exemple de matrices de score (Matrices nucléiques)

	<b>A</b>	<b>T</b>	<b>C</b>	<b>G</b>
<b>A</b>	1	0	0	0
<b>C</b>	0	1	0	0
<b>G</b>	0	0	1	0
<b>T</b>	0	0	0	1

Matrice unitaire

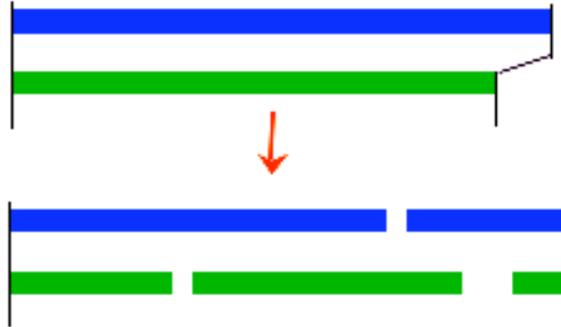
	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
<b>A</b>	3	0	1	0
<b>C</b>	0	3	0	1
<b>G</b>	1	0	3	0
<b>T</b>	0	1	0	3

Matrice transition-transversion

## Trois processus d'alignements : global, multiple et local

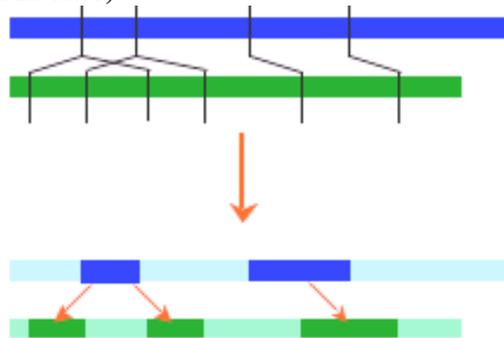
**Alignement global :** Alignement de deux séquences sur la totalité de leur longueur en tenant compte de tous les résidus. Si les longueurs sont différentes, des insertions / délétions sont introduites pour aligner les deux extrémités des deux séquences.

Cet alignement permet de mesurer le degré de similitude entre deux séquences

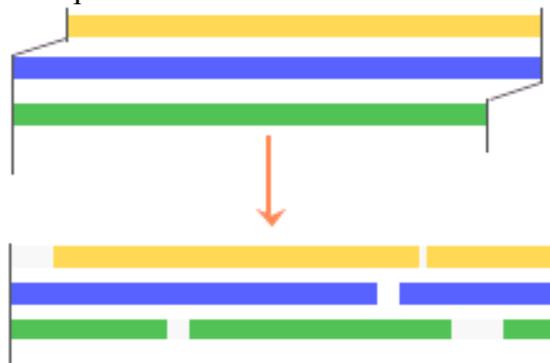


**Alignement local:** Alignement de deux séquences sur des régions isolées et permettant de trouver des segments qui ont un haut degré de similitude.

Utilisé pour la recherche dans les bases de données (comparaison d'une séquence avec les séquences contenues dans la base)



**Alignement Multiple:** Alignement portant sur plusieurs séquences à la fois et dans leur intégralité. Il permet de mettre en évidence les relations entre séquences que l'on ne peut pas visualiser en comparant les séquences deux à deux.



À chaque type d'alignement est associé un programme informatique permettant d'optimiser le traitement

Alignement global :	Alignement local:	Alignement Multiple:
Needle Stretcher	BLAST (Basic Local Alignment Tool) FASTA	T-Coffee