L'algorithme de Needleman et Wunsch dans le cas des protéines :

Première méthode :

Dans le cas où l'utilisateur ne souhaite pas utiliser une des matrices de scores décrites précédemment (PAM, BLOSUM …etc.), il est primordial de définir au préalable 3 scores (**identité**, **substitution** et **gap**) afin d'appliquer convenablement l'algorithme de **Needleman et Wunsch.**

$$\mathbf{S(i,j) = Max} \begin{cases} \mathbf{S(i-1,j-1) + se(i,j)} \\ \mathbf{S(i-1,j) + gap} \\ \mathbf{S(i,j-1) + gap} \end{cases}$$

Exemple d'alignement : On considère les deux séquences suivantes :

S1 = MPRCLCQR

S2 = PYRCKCR

Afin d'aligner ces deux séquences, il est important de définir le score d'identité, de substitution et de gap.

Dans cet exemple, on admet que le score **d'identité = 3**, le score de **substitution = -1** et le score de **gap = -2.**

1ere étape : Construction de la matrice initiale

Dans un premier temps, les deux séquences S1 et S2 sont insérées dans une matrice dite **initiale** de sorte que S1 soit à l'horizontal (axe des abscisses) et S2 à la verticale du tableau (axe des ordonnées). Puis, les cases de cette matrice doivent être remplies par des scores élémentaires appropriés (**3** si les deux acides aminés des deux séquences sont identiques et **-1** pour une substitution).

	M	**P**	**R**	**C**	**L**	**C**	**Q**	**R**
P	-1	3	-1	-1	-1	-1	-1	-1
Y	-1	-1	-1	-1	-1	-1	-1	-1
R	-1	-1	3	-1	-1	-1	-1	3
C	-1	-1	-1	3	-1	3	-1	-1
K	-1	-1	-1	-1	-1	-1	-1	-1
C	-1	-1	-1	3	-1	3	-1	-1
R	-1	-1	3	-1	-1	-1	-1	3

2ème étape : construction de la matrice transformée

Dans un deuxième temps, il faut créer une deuxième matrice à i+2 colonnes et j+2 lignes, dans laquelle la 1ère ligne et la 1ère colonne seront initialisées non pas à zéro mais en utilisant le score des gaps (-2) comme suit :

		j	1	2	3	4	5	6	7	8
			M	**P**	**R**	**C**	**L**	**C**	**Q**	**R**
i		0	-2	-4	-6	-8	-10	-12	-14	-16
1	**P**	-2								
2	**Y**	-4								
3	**R**	-6								
4	**C**	-8								
5	**K**	-10								
6	**C**	-12								
7	**R**	-14								

C'est à ce niveau qu'on applique l'algorithme de **Needleman-Wunsch** afin d'aligner les deux séquences S1 et S2 :

$$\mathbf{S(i,j) = Max} \begin{cases} \mathbf{S(i-1,j-1) + se(i,j)} \\ \mathbf{S(i-1,j) + gap} \\ \mathbf{S(i,j-1) + gap} \end{cases}$$

Si on commence par la case (1,1), l'algorithme est appliqué comme suit :

$$S(1,1) = Max \begin{cases} S(0,0) + se(1,1) = 0 + -1 = -1 \\ S(0,1) + -2 = -4 \\ S(1,0) + -2 = -4 \end{cases}$$

Le maximum entre -1, -4 et -4 est bien -1. Donc le score à mettre dans la case i= 1 et j=1 c'est **-1.**

		M	**P**	**R**	**C**	**L**	**C**	**Q**	**R**
	0	-2	-4	-6	-8	-10	-12	-14	-16
P	-2	-1							
Y	-4								
R	-6								
C	-8								
K	-10								
C	-12								
R	-14								

L'application de l'algorithme de **Needleman-Wunsch** permet de remplir la matrice transformée comme suit :

		M	**P**	**R**	**C**	**L**	**C**	**Q**	**R**
	0	-2	-4	-6	-8	-10	-12	-14	-16
P	-2	-1	1	-1	-3	-5	-7	-9	-11
Y	-4	-3	-1	0	-2	-4	-6	-8	-10
R	-6	-5	-3	2	0	-2	-4	-6	-5
C	-8	-7	-5	0	5	3	1	-1	-3
K	-10	-9	-7	-2	3	4	2	0	-2
C	-12	-11	-9	-4	1	2	7	5	3
R	-14	-13	-11	-6	-1	0	5	6	8

3ème étape : traceback (traçage en arrière)

le parcours de la matrice transformée commence par le plus haut score, vers le plus haut score parmi les trois cases (i-1, j-1) (i-1,j) et (i, j-1) et ainsi de suite jusqu'à la case (1,1). Dans cet exemple, on commence par la case (7,8) ayant le plus haut score = 8. Dans ce cas, les scores des 3 cases (6,7) (6,8) et (7,7) sont respectivement 5, 3 et 6. Donc, le parcours de la matrice sera vers la case (7,7) où le score est le plus grand soit 6. Le parcours final de la matrice transformée est le suivant :

	S1	M	P	R	C	L	C	Q	R
S2	0	-2	-4	-6	-8	-10	-12	-14	-16
P	-2	**-1**	**1**	-1	-3	-5	-7	-9	-11
Y	-4	-3	-1	**0**	-2	-4	-6	-8	-10
R	-6	-5	-3	**2**	0	-2	-4	-6	-5
C	-8	-7	-5	0	**5**	3	1	-1	-3
K	-10	-9	-7	-2	3	**4**	2	0	-2
C	-12	-11	-9	-4	1	2	**7**	5	3
R	-14	-13	-11	-6	-1	0	5	**6**	**8**

4ème étape : génération de l'alignement et calcul des scores

En suivant le parcours de la matrice transformée tracé précédemment, les acides aminés en diagonal représentent soit appariement (identité **|**) ou une substitution (**:**).

Il est très important de signaler que les acides aminés en horizontal dans le parcours de la matrice transformée représentent soit une insertion dans la séquence S1 ou une délétion dans la séquence S2. Par exemple, le premier et le deuxième acides aminés de la séquence S1 (M et P respectivement), sont en horizontal dans le parcours de la matrice transformée. Cela signifie que soit la séquence S1 a subit une insertion de l'acide aminé M (car M a un plus faible score par rapport à P) ou alors la séquence S2 a subit une délétion de cette acide aminé au cours de l'évolution.

-De même, les acides aminés en vertical représentent soit une insertion dans la séquence S2 ou une délétion dans la séquence S1. Par exemple, le deuxième et le troisième acides aminés de la séquence S2 (Y et R respectivement) sont en vertical dans le parcours de la matrice transformée. Cela signifie que soit il y'a eu une insertion de l'acide aminé Y (plus faible score par rapport à l'acide aminé R) dans la séquence S2, ou alors une délétion de cette acide aminé dans la séquence S1 et ce par nécessité adaptative.

S1	M	P	**▬**	R	C	L	C	Q	R
	*	**|**	*	**|**	**|**	:	**|**	*	**|**
S2	**▬**	P	Y	R	C	K	C	**▬**	R

Le trou ( ) en position 3 entre l'acide aminé P et R de la séquence S1 représente un gap ou indel. Il est représenté en * lors de l'alignement.

Il est également important de signaler qu'une substitution a eu lie au niveau de la sixième position où l'acide aminé **L** de la séquence S1 a été remplacé par un **K** dans la séquence S2 et ce par nécessité évolutive et d'adaptation à l'environnement.

Calcul des scores :

Le pourcentage d'identité (%id) = (nombre d'identités / taille de la séquence après alignement)

Dans cet exemple %id= (5/9)* 100 = 55.55%

Le pourcentage des gaps : = (nombre gaps / taille de la séquence après alignement)*100

= (3/9)*100= 33.33%

Le pourcentage des substitutions : (nombre de substitutions / taille de la séquence après alignement)*100

= (1/9)*100= 11.11%.

Score d'alignement : (nombre d'identités * score d'identité) + (nombre de substitutions*score des substitutions) + (nombre de gaps * score des gaps)

(5*3)+(1*-1)+(3*-2)=8