

Cours IFT3290
Démo 2 : le 27 janv., 03

Le code génétique

La translation d'une séquence de nucléotides est en une séquence d'acides aminés correspondant à la protéine, se fait selon le code génétique.

L'ARN messenger qui sert d'intermédiaire entre l'ADN et la protéine, pour être lu par groupe de trois nucléotides dans un ordre séquentiel. Chacun des triplets est appelé codon spécifiant un acide aminé. 64 acides aminés sont possibles vu que les triplets sont codés à partir d'un alphabet de longueur 4. Uniquement 20 sont retrouvés dans les protéines. Un codon est utile pour indiquer le début du processus de translation AUG et trois autres pour indiquer la fin du processus UAA, UAG, UGA.

Voici la table du code génétique

		Second base of codon					
		U	C	A	G		
First base of codon	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } SER UCA } UCG }	UAU } Tyr UAC } UAA UAG	UGU } Cys UGC } UGA UGG } Trp	U	C
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U	C
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thy ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U	C
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U	C
						A	G
						Third base of codon	

The genetic code, written by convention in the form in which the Codons appear in mRNA. The three terminator codons, UAA, UAG, and UGA, are boxed in red; the AUG initiator codon is shown in green.

Les codes des acides aminés :

A : Ala (Alanine)
B : Asp (acide Aspartique) ou Asn (Asparagine) (D ou N)
C : Cys (Cystine)
D : (Asp) Aspartate
E : (Glu) Glutamate
F : (Phe) Phénylalanine)
G : (Gly) Glycine
H : (His) Histidine
I: (Ile) Isoleucine
K: (Lys) Lysine
L: Leu (Leucine)
M: (Met) Méthionine
N : (Asn) Asparagine
P : (Pro) Proline
Q: (Gln) Glutamine
R: (Arg) Arginine
S : (Ser) Serine
T : (Thr) Thréonine
V: (Val) Valine
W : (Trp) Thryptophane
Y: (Tyr) Tyrosine
Z: (Glx) E ou Q
X: acide amine indéterminé

Alignement Optimal avec trous :

Un exemple d'alignement global entre deux séquences de protéines en utilisant le modèle avec une pénalité de trou constante : $d=8$, Pour calculer cet alignement, il a été utilisé une matrice de pondérations des acides aminés nommée Blosom50 dont un extrait utile pour le cas suivant.

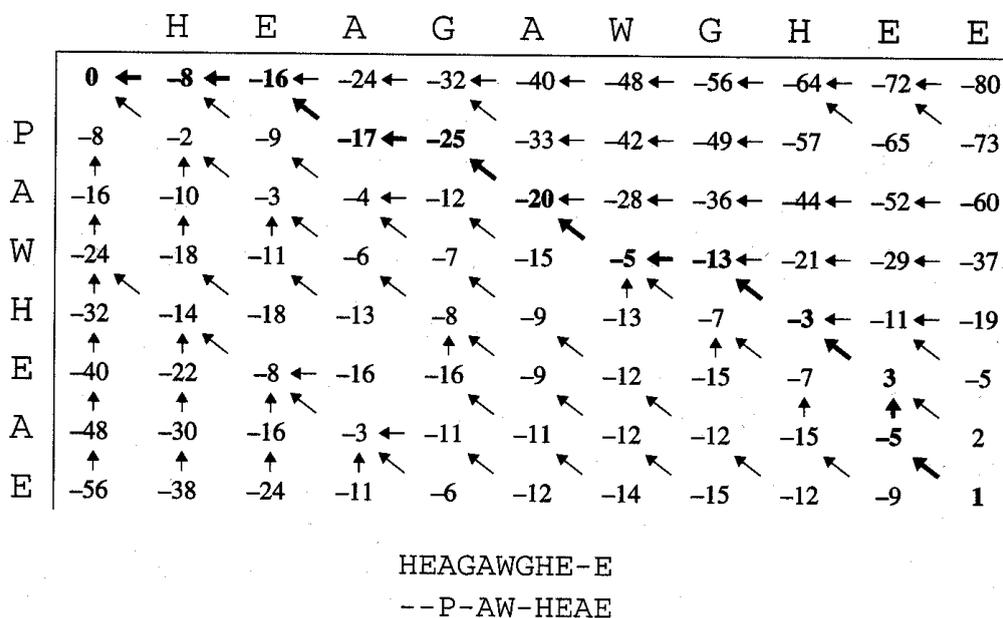


Figure 2.5 Above, the global dynamic programming matrix for our example sequences, with arrows indicating traceback pointers; values on the optimal alignment path are shown in bold. Below, a corresponding optimal alignment, which has total score 1.

	H	E	A	G	A	W	G	H	E	E
P	-2	-1	-1	-2	-1	-4	-2	-2	-1	-1
A	-2	-1	5	0	5	-3	0	-2	-1	-1
W	-3	-3	-3	-3	-3	15	-3	-3	-3	-3
H	10	0	-2	-2	-2	-3	-2	10	0	0
E	0	6	-1	-3	-1	-3	-3	0	6	6
A	-2	-1	5	0	5	-3	0	-2	-1	-1
E	0	6	-1	-3	-1	-3	-3	0	6	6

Figure 2.3 The two example sequences we will use for illustrating dynamic programming alignment algorithms, arranged to show a matrix of corresponding BLOSUM50 values per aligned residue pair. Positive scores are in bold.

Exemple d'un alignement local avec le modèle de pénalité constante

	H	E	A	G	A	W	G	H	E	E
	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0
A	0	0	0	5	0	5	0	0	0	0
W	0	0	0	0	2	0	20	12	4	0
H	0	10	2	0	0	0	12	18	22	14
E	0	2	16	8	0	0	4	10	18	28
A	0	0	8	21	13	5	0	4	10	20
E	0	0	6	13	18	12	4	0	4	16

AWGHE
 AW-HE

Figure 2.6 Above, the local dynamic programming matrix for the example sequences. Below, the optimal local alignment, with score 28.

Modèle d'alignement avec une pénalité de trou en fonction de sa longueur

Ce modèle tient compte de la présence du trou ainsi que de sa longueur, La pénalité sera donc la somme de $Wg + qWs$ avec Ws et Wg des constantes = 0 et $q = 1$

On assume que $s(x, -) = s(-, x) = 0$, vu que les espaces introduits vont être pénalisés comme faisant partie du trou. Le but sera de maximiser les scores en déduisant les poids induits par les trous ainsi que les espaces.

L'algorithme de programmation dynamique

L'algorithme procède par l'alignement de $S[1] \dots [i]$ avec $T[1] \dots [j]$. En plus de l'alignement global des deux séquences, on peut trouver deux autres, dans lesquels, on retrouve une suite d'espaces dans l'une ou l'autre séquence. On définit donc les variables suivantes :

1. $V(i,j)$ la valeur d'un alignement optimal de $S[1] \dots S[i]$ et $T[1] \dots T[j]$
2. $G(i,j)$ la valeur d'un alignement optimal de $S[1] \dots S[i]$ et $T[1] \dots T[j]$ dont la dernière paire fait correspondre $S[i]$ avec $T[j]$.
3. $F(i,j)$ la valeur de l'alignement optimal de $S[1] \dots S[i]$ et $T[1] \dots T[j]$ dont la dernière paire fait correspondre $S[i]$ avec un blanc.
4. $E(i,j)$ est la valeur de l'alignement optimal de $S[1] \dots S[i]$ et $T[1] \dots T[j]$ dont la dernière paire fait correspondre un blanc à $T[j]$

I G A x_i
L G V y_j

A I G A x_i
G V y_j - -

G A x_i - -
S J G V y_i

Conditions de Bases :

$$M(0,0) = I_x(0,0) = I_y(0,0) = 0$$

$$M(i,0) = I_x(i,0) = -d \cdot i \cdot e, \text{ pour } i > 0$$

$$M(0,j) = I_y(0,j) = -d \cdot j \cdot e, \text{ pour } j > 0$$

Récurrance : Pour $i > 0$ et $j > 0$

$$M(i, j) = \max \begin{cases} M(i-1, j-1) + s(x_i, y_j), \\ I_x(i-1, j-1) + s(x_i, y_j), \\ I_y(i-1, j-1) + s(x_i, y_j); \end{cases}$$

$$I_x(i, j) = \max \begin{cases} M(i-1, j) - d, \\ I_x(i-1, j) - e; \end{cases}$$

$$I_y(i, j) = \max \begin{cases} M(i, j-1) - d, \\ I_y(i, j-1) - e. \end{cases}$$

Dans le calcul de I_x (I_y), on tient compte du cas où l'alignement se termine avec plus d'un espace, donc un trou. Dans ce cas trois tables sont nécessaires.

Dans le cas où la pénalité varie en fonction de la longueur du trou, les changements suivants sont à apporter :

$$M(0,0) = I_x(0,0) = I_y(0,0) = 0$$

$$M(i,0) = I_x(i,0) = -w(i), \text{ pour } i > 0$$

$$M(0,j) = I_y(0,j) = -w(j), \text{ pour } j > 0$$

$I_x(i,j) = \text{Max}(M[i-k,j] - w(k), I_x(i-k,j))$ pour $1 \leq k \leq i$ donc deux boucles
 $O(mn^2)$

Idem pour $I_y(i,j)$ mais pour $1 \leq k \leq j$, $O(nm^2)$

La complexité est $O(mn^2 + nm^2)$