

Corrigé type de l'Examen de Bioinformatique**Exercice 1 : (12 pts)***I- Cochez la ou les bonnes réponses :***QCM1 : Concernant les bases de données biologiques → (4 pts = 2*2)**

- A. Les banques généralistes de séquences nucléotidiques sont GenBank, TrEMBL et DDBJ
- B. La banque de données PDB « Protein Data Bank » est une banque de structure 3 D des protéines ✓**
- C. Les banques spécialisées contiennent des données hétérogènes établie auteur d'une thématique
- D. Les banques généralistes contiennent une grande diversité des organismes représentés ✓**
- E. La base de données KEEG pathway et PFAM sont spécialisées dans l'étude des voies métaboliques

QCM 2 : L'entrée écrite dans les lignes ci-dessous est obtenue par le format → (2 pts)

- SQ Sequence 581 BP; 132 A; 164 C; 145 G; 140 T; 0 other;
ccatttagag gaagtaaaag tcgtaacaag gtctccgttg gtgaaccaggaggatc 60
ttacgagagt gtcaccactc ccaaccact gttacctac ccgtccaccg tgcttcggca 120
ggcagtcctg tgggacaggg cctcgcccc ctccgggggg tgccctgccg

A. EMBL ✓ **B. FASTA** **C. GenBank** **D. DDBJ** **E. TrEMBL**

QCM 3 : La famille des matrices PAMs → (4 pts = 2*2)

- A. Se sont des matrices asymétriques (20 × 20)
- B. Pour les construire on utilise 504 groupes de protéines avec un taux de mutations très élevées
- C. Il s'agit d'un travail qui fait publier par Margaret Dayhoff (1978) ✓**
- D. Pour les construire on utilise des arbres phylogénétiques ✓**
- E. Lorsqu'il y a une sorte de divergence entre deux séquences, on utilise une matrice PAM avec un nombre réduit de mutations

QCM 4 : La famille des matrice BLOSUMs → (2 pts)

- A. Pour les construire on utilise exclusivement la famille des protéines cytochrome C
- B. Se sont des matrices de substitution nucléiques
- C. Le chiffre venant juste après BLOSUM (ex BLOSUM 62) indique le nombre des mutations acceptés dans chaque 100 acides aminés
- D. Lorsqu'il y a une sorte de conservation entre deux séquences, on utilise généralement une matrice BLOSUM avec un taux d'identité faible
- E. Aucune des propositions ci-dessus n'est exacte ✓**

Exercice 2 : (8 pts)

1. En utilisant la matrice de substitution ci-dessous et un coût de gap à -5, quel alignement a le meilleur score ?

D	6				
E	2	5			
F	-3	-3	6		
G	-1	-2	-3	6	
W	-4	-3	1	-2	11
	D	E	F	G	W

Alignement (1)	Alignement (2)
DFDW-FE DF-WDGF	DEDW-FE FE-WDWE
Alignement (1) = $6 + 6 - 5 + 11 - 5 - 3 - 3 = 23 - 16$ = 7 → 1 pts	Alignement (2) = $-3 + 5 - 5 + 11 - 5 + 1 + 5$ = 9 → 1 pts
L'alignement 2 a le score plus élevé. 0.5 pts	

2. Soit la matrice de programmation dynamique suivante :

		M	V	S	S	D
	0	-2	-4	-6	-8	-10
M	-2	2	0	-2	-4	-6
V	-4	0	4	2	0	-2
S	-6	-2	2	6	4	2
D	-8	-4	0	4	5	6

a. Quel est le type de l'alignement optimal recherché : **global (0,5 pts)**; Justifiez : **parce que l'initialisation se fait par des valeurs multiples de Δ (pour la 1^{ère} ligne et la 1^{ère} colonne). (1 pts)**

b. Quel est le score de cet alignement ? **6 (0,5 pts)**; Comment l'avez-vous obtenu? **a été obtenu dans la case à droite en bas (dernier élément dans la table T(i, j)). (0,5 pts)**

c. Déterminez à partir de la matrice le ou les meilleur(s) chemin ainsi que l'alignement optimal correspondant ?

✓ Chemin (1) = 0, 2, 4, 6, 4, 6 → 0 ; 5 pts;

✓ Chemin (2) = 0, 2, 4, 2, 4, 6 → 0,5 pts

Alignement (1) =

M	V	S	S	D
M	V	S	-	D

2 pts (2* 1)

Alignement (2) =

M	V	S	S	D
M	V	-	S	D