

# Examen de bioinformatique

Février 2009

Durée : 2h (ou 2h30) - Documents interdits

## Première partie (4 points)

1) La séquence ci-dessous est-elle au format fasta (justifiez votre réponse) ? **1 pt**  
*Is the sequence below in fasta format (justify your answer) ?*

```
>tr | A0A098 | A0A098_CHLRE  
MASMAAELRPSDGGSSLHMLDLSLLMMGLSSGGGVGGGGSSQSQILDSAGAAELAALLLPQ  
HSNDPLHLMSTGDAALGLAGPMAAAEHHQHHPHHQHHSVPATAGFPSQTPPPPLFSNATA  
GAAPATRVRAAGSCGSGGVAGGTTSHSSEDGVFHSADPHHHHQHLQQPQPQQQ
```

2) Quel problème rencontrera-t-on avec cette séquence lors d'une recherche de similarité ? **1.5 pt**  
*Which problem will be encountered with this sequence in a similarity search?*

3) Définissez la banque GO. Define the GO database. **1.5 pt**

## Deuxième partie (7.5 points)

Un alignement multiple de séquences eucaryotes est présenté page 2. *A multiple alignment of eucaryotic sequences is shown page 2.*

1) Quels groupes pouvez-vous distinguer dans cet alignement ? **2 pt**  
Donnez 2 résidus discriminants pour chaque groupe.  
*Which group can you distinguish in this alignment? Give 2 discriminative residues for each group.*

2) Donnez une erreur de séquence probable dans cet alignement. **1.5 pt**  
*Give a probable sequence error in this alignment.*

3) Quelle est la relation d'homologie entre zn143\_human et znf76\_human ? zn143\_human et q8ci27\_mouse ? zn143\_human et znf76\_mouse ? **1.5 pt**  
*Give the homology relation between zn143\_human and znf76\_human ? zn143\_human and q8ci27\_mouse ? zn143\_human and znf76\_mouse ?*

4) Un arbre a été construit à partir de cet alignement selon la méthode du neighbor-joining. **2.5pt**  
Dans cet arbre (page 3), 3 identifiants de séquences ont été remplacés par x, y et z.  
*A tree has been constructed from this alignment using the neighbor-joining method. In this tree (page 3), 3 sequence identifiers have been replaced par x, y, and z.*

a) Est-ce que cet arbre est en accord avec votre analyse de l'alignement ? Justifiez votre réponse.

*Is this tree in agreement with your alignment analysis? Justify your answer.*

b) Selon vous, à quelles séquences correspondent respectivement X, Y et Z ?

*According to you, which sequences correspond respectively to X, Y and Z?*

```

                *      20      *      40      *      60      *      80      *      100      *
ZN143_HUMAN : MEGVSLQAVTLADGSTAYIQHNSK----DAKLIDGQVIQLEDGSAAYVQHVP-IPKSTGDSLRLLEDGQAVQLEDGTTAFIHHTSKDSYDQSALQAVQLEDGTTAYIEHAVQ
A6QQW0_BOVIN : MEGVSLQAVTLADGSTAYIQHNS-----KDGSAAYVQHVP-IPKTTGDSLRLLEDGQAVQLED-----SYDQSALQAVQLEDGTTAYIEHAVQ
Q8CI27_MOUSE : MEGVSLQAVTLADGSTAYIQHNSK----DGRLIDGQVIQLEDGSAAYVQHVP-IPKS-----NSYDQSSLQAVQLEDGTTAYIEHAVQ
Q6VQB0_FUGRU : MDTIVSLQAVTLADGSTAYIQHDSKASFSGQINDGQVIQLEDGSAAYVQHVSMPKAGGDSLQLEDGQTVQLEDGTTAYIHTP-KETYDQSGLQEVQLEDGSTAYIQHTVH
A0AUQ7_DANRE : MDTIVSLQAVTLVVGSTAYIQHSPKVSLTENKINEGQVIQLEDGSAAYVQHLPMSKTGGEGRLLEDGQAVQLEDGTTAYTHAP-KETYDQGLQAVQLEDGTTAYIQH---
Q4S173_TETNG : MDTIVSLQAVTLADGSTAYIQHDSKASFPDQINDGQVIQLEDGSAAYVQHVSMPKAGGESLQLEDGQTVQLEDGTTAYIHTAP-KETYDQSGLQEVQLEDGSTAYIQHTVH
Q6GPP5_XENLA : MESMSLQAVTLADGSTAYIQHNTK----DGKLMGQVIQLEDGSAAYVQHIP----KGDLSLEDGQAVQLEDGTTAYIHHSKESYDQSSVQAVQLEDGTTAYIEHAVQ
ZNF76_MOUSE : MESLGLQTVRLSDGTTAYVQQAVK----GEKLEEGQVIQLEDGTTAYIQVTVI---QKESFSFEDGQPVQLEDGSMAYIHHTPKEGCDPSALEAVQLEDGSTAYIHPVVP
ZNF76_HUMAN : MESLGLHTVRLSDGTTAYVQQAVK----GEKLEEGQVIQLEDGTTAYIQVTVI---QKEALSFEEDGQPVQLEDGSMAYIHRTPREGYDPSLEAVQLEDGSTAYIEHPVA

```

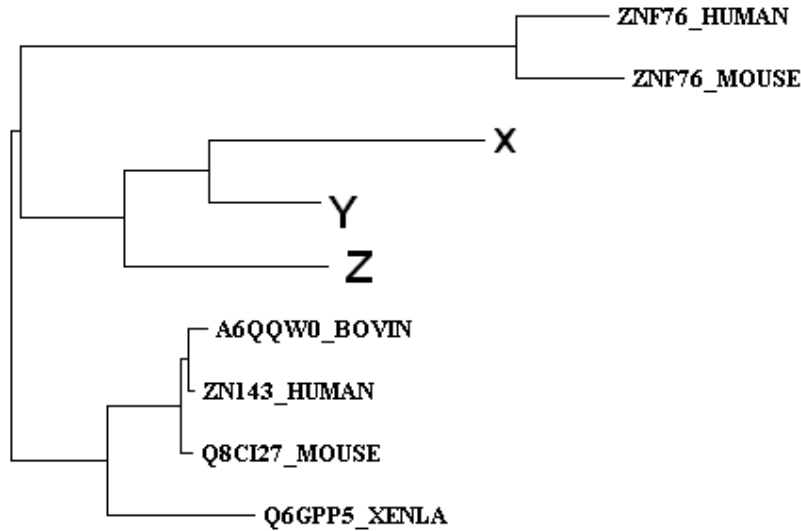
```

                120      *      140      *      160      *      180      *      200
ZN143_HUMAN : VPQSDTILAIQADGTVAGLHT-GDATIDPDTIISALEQYAAKVSIDGSEVAGTGMIGENEQEKMQIVLQGHATRVTAKSQQSCEKAFRCBYDGGC-
A6QQW0_BOVIN : VPQSDTILAIQADGTVAGLHT-GDAAIDPDTIISALEQYAAKVSIDGSEGVGTSGIIGENEQEKMQIVLQGHATRVTAKSQQSCEKAFRCYDGGC-
Q8CI27_MOUSE : VPQSDTILAIQADGTVAGLHT-GDATIDPDTIISALEQYAAKVSIDGSDGVTSTGMIGENEQEKMQIVLQGHATRVTPKSQQSCEKAFRCYDGGC-
Q6VQB0_FUGRU : MPQSNITILAIQADGTIADLQA-DATGLNPETIISVLEQYATKVESIENQLG--SYSRAEADNGVHVRIVLQDQDNFQS-RSTNVGEKSFRCFYEYEGCG-
A0AUQ7_DANRE : MPQSNITILAIQADGTVADLQT-EGT-IDAETIISVLEQYSTKMEATECGTG--LIGRGDSG-VVHMQIVLQGDQCRSP-RIQHVCEKAFRCHEHGGC-
Q4S173_TETNG : MPQSNITILAIQADGTIADLQA-DAAGLNPETIISVLEQYATKVPVLSGLRLRLWAGGHYRKPVGLLQPAGGGEREPH-ADCFTRSRQQAVAEHQCGR-
Q6GPP5_XENLA : VPQSDTILAIQADGTVAGLHT-GEASIDPDTIISALEQYAAKVSIEGEGAGSNALITSESEKMQIVLS-HGSEVPVKVPQTNEKAFRCYDGGC-
ZNF76_MOUSE : VPSDSAIIAVQTEAGLEDLAAEDEEGFGTDTVVALEQYASKVLHDS-----PASHNGKGQOVCDRAFRCYKGGC-
ZNF76_HUMAN : VPESEITILAVQTEVGLLEDLAAEEDDEGFSADAVVALEQYASKVLHDS-----QIPRNGKGQOVCDRAFRCYKGGC-

```

9 species, 129 sites (global gap removal)  
 Neighbor Joining Method  
 Observed divergence  
 100 bootstrap replicates

0.109



### Troisième partie (8.5 points)

Une entrée de la banque EMBL vous est présentée partiellement page 4.  
*An entry from the EMBL database is partially represented page 4.*

1) Selon vous, à quelle classe et à quelle division appartient cet enregistrement ? **1 pt**  
*According to you, which class and which division does this entry belong to?*

2) Que signifie la ligne ci-dessous ? *What does the line below mean?* **1 pt**

FT gene <1..>3374

3) Quelle est la nature de l'exon 1 ? *What is the nature of exon 1?* **1 pt**

4) On désire récupérer les régions codantes des gènes humains annotés HLA-A avec SRS. **2.5 pt**  
 Précisez la banque choisie et indiquez comment procéder en utilisant le formulaire de requête standard (cf ci-dessous).

*We want to get the coding regions of the human genes annotated as HLA-A using SRS. Precise the database and the query using the standard query form (see below).*

<i>Combine search with :</i>		<b>Champs (fields)</b>	<b>Mots-clés (Keywords)</b>	<i>Retrieve entries of type :</i>	
AND	<input type="checkbox"/>			ENTRY	<input type="checkbox"/>
OR	<input type="checkbox"/>			REFERENCES	<input type="checkbox"/>
BUTNOT	<input type="checkbox"/>			FEATURES	<input type="checkbox"/>
				COUNTERS	<input type="checkbox"/>

```

...
DE Homo sapiens MHC class I antigen (HLA-A) gene, HLA-A*01 variant allele,
DE alternatively spliced.
...
XX
FH Key Location/Qualifiers
FT source 1..3374
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT gene <1..>3374
FT /gene="HLA-A"
FT /allele="HLA-A*01 variant"
FT mRNA join(<1..373,504..773,1015..1266,1870..2145,2248..2364,
FT 2807..2839,2982..3029,3199..>3374)
FT exon <1..373
FT /number=1
FT 5' UTR <1..300
FT /allele="HLA-A*01 variant"
FT CDS join(301..373,504..773,1015..1266,1870..2145,2248..2364,
FT 2807..2839,2982..3029,3199..3203)
FT /gene="HLA-A"
FT /product="MHC class I antigen"
FT /protein_id="AAW30165.1"
FT /translation="MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F F T S V S R P G R G E P R
FT F I A V G Y V D D T Q F V R F D S D A A S Q K M E P R A P W I E Q E G P E Y W D Q E T R N M K A H S Q T D R A N L G T
FT L R G Y N Q S E D G S H T I Q I M Y G C D V G P D G R F L R G Y R Q D A Y D G K D Y I A L N E D L R S W T A A D M A
FT A Q I T K R K W E A V H A A E Q R R V Y L E G R C V D G L R R Y L E N D P P K T H M T H H P I S D H E A T L R C W A L
FT G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G
FT L P K P L T L R W E L S S Q P T I P I V G I I A G L V L L G A V I T G A V V A A V M W R R K S S D R K G G S Y T Q A A
FT S S D S A Q G S D V S L T A C K V "
FT exon 504..773
FT /number=2
FT exon 1015..1266
FT /number=3
FT variation 1268
FT /note="alternatively spliced compared to HLA-A*010101;
FT results in altered exon and protein length; no membrane
FT expression detected"
FT /replace="g"
FT /gene="HLA-A"
FT exon 1870..2145
FT /number=4
FT exon 2248..2364
FT /number=5
FT exon 2807..2839
FT /number=6
FT exon 2982..3029
FT /number=7
FT exon 3199..>3374
FT /number=8
FT 3' UTR 3204..>3374
...

```

5) Une recherche blastp a été effectuée à partir de la séquence protéique HLA-A (357 aa). Cette protéine est similaire à des immunoglobulines comme le montrent les alignements avec la protéine MUCM\_RABIT.

3 pt

*A blastp search has been performed with the HLA-A sequence (357 aa). This protein is similar to immunoglobulins as shown by the alignments with the protein MUCM\_RABIT.*

a) Représentez schématiquement les 2 protéines en indiquant les régions conservées.  
*Represent schematically the 2 proteins with their conserved regions.*

b) Représentez le résultat d'une comparaison des deux protéines par la méthode de la matrice de points.

*Represent the result of a dotplot comparison between the two proteins.*

```
>sp|P04221|MUCM_RABIT  Ig mu chain C region membrane-bound form
      Length = 479

Score = 45.1 bits (105), Expect = 3e-06
Identities = 29/94 (30%), Positives = 49/94 (52%), Gaps = 11/94 (11%)

Query: 214 EATLRCWALGFYPAEITLTWQRDGED-----QTQDTELVETRPAGDGTQKWAADVVPSPG 268
      ++ L C A GF P +I+++W RDG+          T+ E ET+ AG TF  + + +
Sbjct: 132 KSRLICQATGFSPKQISVSWLRDQKVESGVLTKPVE-AETKGAGPATFSSISMLTITES 190

Query: 269 E---EQRYTCHVQHEGL--PKPLTLRWELSSQPT 297
      +   + YTC V H G+   K +++ E S+ P+
Sbjct: 191 DWLSQSLYTCRVDHRGIFFDKNVSMSSSECSTTPS 224

Score = 40.4 bits (93), Expect = 7e-05
Identities = 24/81 (29%), Positives = 37/81 (45%), Gaps = 6/81 (7%)

Query: 215 ATLRCWALGFYPAEITLTWQRDGEDQTQD---TELVETRPAGDGTQKWAADVVPSP---G 268
      AT+ C   GF PA++ + WQ+ G+   + D   T       P   G +   + + V
Sbjct: 352 ATVTCLVKGFSPADVFVQWQQRGQPLSSDKYVTSAPAPEPQAPGLYFTHSTLTVTEEDWN 411

Query: 269 EEQRYTCHVQHEGLPKPLTLR 289
      + +TC V HE LP  +T R
Sbjct: 412 SGETFTCVVGHEALPHMVTER 432

Score = 31.2 bits (69), Expect = 4e-03
Identities = 23/85 (27%), Positives = 37/85 (43%), Gaps = 10/85 (11%)

Query: 219 CWALGFYPAEITLTWQRDGEDQTQDTELVETRPAAA---GDGTFQKWAADVVPSP-----GEE 270
      C A F P+ +T +W      +   + V T P   GD +   + V+VPS   G E
Sbjct: 28  CLARDFLPSSVTFWSFKNNSEIS-SRTVRTFPVVKRGD-KYMATSQVLVPSKDVLPQTE 85

Query: 271 QRYTCHVQHEGLPKPLTLRWELSSQ 295
      +   C VQH   + L + + + S+
Sbjct: 86  EYLVCKVQHSNSNRDLRVSFPVDSE 110
```

# Bioinformatique

Janvier 2010

Durée : 2h (2h30) - Documents interdits

## Première partie (7.5 points)

On veut rechercher par SRS les régions promotrices annotées sur le génome de la souris (*Mus musculus*).

*We want to retrieve promotor regions annotated on the mouse (Mus musculus) genome using SRS.*

1) Quelle(s) banque(s) doit-on choisir ? *Which database(s) do we have to choose?* 1 pt

- Uniref50
- Uniprot
- EMBL
- TrEMBL
- Unigene

2) Donnez la requête SRS en vous aidant du formulaire ci-dessous. 2 pts

*Give the SRS query using the query form below.*

<i>Combine search with :</i>	<b>Champs (fields)</b>	<b>Mots-clés (Keywords)</b>	<i>Retrieve entries of type :</i>
AND			ENTRY
OR			REFERENCES
BUTNOT			FEATURES
			COUNTERS

3) Un utilisateur de SRS a sélectionné une liste de séquences protéiques. A présent, il veut sélectionner parmi ces séquences seulement les entrées des séquences protéiques possédant une structure 3D connue. Comment peut-il procéder en utilisant SRS ? 1.5 pts

*A SRS user has selected a list of protein sequences. Now, he wants to retrieve among these sequences only the protein sequence entries with a known 3D structure. How can he proceed using SRS?*

4) Quelles sont les principales différences entre Swissprot et SpTrEmbl ? 1.5 pts

*What are the main differences between Swissprot and SpTrEmbl?*

5) Quels avantages présente la banque Interpro par rapport aux banque Pfam ou Prosite par exemple ? 1.5 pts

*What are the advantages of the Interpro database compared to databases such as Pfam or Prosite for instance?*



## Fasta

>>EM\_PAT:CS723756; CS723756 Sequence 14 from Patent WO20 (4700 nt)  
initn: 378 initl: 238 opt: 549 Z-score: 279.9 bits: 65.5 E(): 7.6e-09  
58.0% identity (69.2% similar) in 357 nt overlap (631-982:1319-1667)

```

          610      620      630      640      650      660
Sequen  GAGGCCAUGGAGGUUGCUAAUCAGACUAGGCCAGAUGGUACAUGCAAUGAGAACUAUUGGG
          . . . . . : . : . . . . . : . . . :
EM_PAT  GCUGACAGACUAACAGACUGUUCCUUCCAUGGGUCUUUUCUGCAGUCACCGUCGUCGAC
          1290      1300      1310      1320      1330      1340

          670      680      690      700      710
Sequen  ACUCAU--CCUAGCUCCAGUGCUGGUCU-GAAAGAUGACCUUCUUG-AAAAUUUGCAGGC
          :: . . : . : . . . . : : : : : : : : : : : : : : : : : : : : : : : :
EM_PAT  ACGUGUGAUCAGAUAUCGCGGCCGCUCUAGAGAUAUCGCCACCAUGCAGUACAUCAAGGC
          1350      1360      1370      1380      1390      1400

          720      730      740      750      760      770
Sequen  CUACCAGAAGCGAAU-GGGAGUGCAGAUGCAGCGAUUCAAGUGAUCCUCUCGUCAUUGCA
          : : : . : : : . : : : : : : : : : : . . : : : . : : : : : : : :
EM_PAT  CAACAGCAAGUUAUCGGCAUCACAGAGCUGUCUCUGCUGACAGAAGUGGAGAC-CCCUA
          1410      1420      1430      1440      1450      1460

          780      790      800      810      820      830
Sequen  GCAAUAUCAUUGGGAUCUUGCACCUGAUAUUGUGGAUUACUGAUCGUCUUUUUUUCAAA
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
EM_PAT  CCAGAAAUGAGUGGGA--GUGCA---GAUGUAG-CGAUAGC-GACCGGCUGUUCUUCAAG
          1470      1480      1490      1500      1510      1520

          840      850      860      870      880      890
Sequen  UGUUUUUACUGCGCUUUAAAUCGGUUUGAAAAGAGGGCCUUCUACGGAAGGAGUGCCU
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
EM_PAT  UGCAUCUACCGGAGACUGAAUGGACUGAAGAGAGGACCUGCCACAGCCGGAGUGCCU
          1530      1540      1550      1560      1570      1580

          900      910      920      930      940      950
Sequen  GAGUCCAUGAGGGAAGAAUAUCAACAGGAACAGCAGAGUGCUGUGGAUGUUGACGAUGGU
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
EM_PAT  GAAUCUAUGCGGGAGGAGUAUAGACAGGAACAGCAGAGCGCCGUGGAUGUGGAUGAUGGC
          1590      1600      1610      1620      1630      1640

          960      970      980
Sequen  CAUUUUGUCAACAUAGAGCUAGAGUAA
          : : : : : : : : : : : : : : : : : : : : : : : :
EM_PAT  CACUUCGUGAAUAUCGAGCUGGAGUGAACACGUGGGAUCCAGAUCUGCUGUGCCUUCUAG
          1650      1660      1670      1680      1690      1700
```



## Alignement Water

GQ229385.1	586	A-----GTGAACAGGCA--GCGGAGGCCATGGAGGTTGCTAATCAG--AC	626
CS723756.1	1347	ACACGTGTGATCAGATATCGCGGCCGCTCTAGAGAT-----ATC-GCCAC	1390
GQ229385.1	627	TAGGCAGATGGTACATGCAATGAGAAC-----TATTGGGACTCATCC	668
CS723756.1	1391	CATGCA----GTACAT-CAAGGCCAACAGCAAGTTTATCGG----CATC-	1430
GQ229385.1	669	TAGCTCCAGTGTCTGGT-----CTGAAAGATGACCTTCTTGAAAATTTGCA	713
CS723756.1	1431	-----ACAGAGCT-GTCTCTGCTGACAGAAG-----TGGA	1459
GQ229385.1	714	G--GCCTACCAGAAGCGAATGGGAGTGCAGATGCAGCGATTCAAGTGATC	761
CS723756.1	1460	GACCCCTACCAGAAATGAGTGGGAGTGCAGATGTAGCGAT-----	1499
GQ229385.1	762	CTCTCGTCATTGCAGCAAATATCATTTGGGATCTTGCACCTGATATTGTGG	811
CS723756.1	1500	-----AGC-----	1502
GQ229385.1	812	ATTACTGATCGTCTTTTTTTTCAATGTATTTATCGTCGCTTTAAATACGG	861
CS723756.1	1503	-----GACCGGCTGTTCTTCAAGTGCATCTACCGGAGACTGAAGTATGG	1546
GQ229385.1	862	TTTGAAGAGAGGGCCTTCTACGGAAGGAGTGCCTGAGTCCATGAGGGAAG	911
CS723756.1	1547	ACTGAAGAGAGGACCTGCCACAGCCGAGTGCCTGAATCTATGCGGGAGG	1596
GQ229385.1	912	AATATCAACAGGAACAGCAGAGTGTCTGTGGATGTTGACGATGGTCATTTT	961
CS723756.1	1597	AGTATAGACAGGAACAGCAGAGCGCCGTGGATGTGGATGATGGCCACTTC	1646
GQ229385.1	962	GTCAACATAGAGCTAGAGTAA	982
CS723756.1	1647	GTGAATATCGAGCTGGAGTGA	1667

## Troisième partie (6.5 points)

1) Est-il préférable d'utiliser un motif ou un profil pour rechercher des homologues divergents ? Justifiez votre réponse. 1.5 pts

*Is it better to use a pattern or a profile to search for divergent homologs? Justify your answer.*

2) Une recherche psi-blast a été effectuée à partir d'une séquence protéique de levure (750 aa). Commentez le résultat de la 1<sup>ère</sup> itération illustré schématiquement (page 5). 1 pts

*A psi-blast search has been performed for a yeast protein sequence (750 aa). Comment on the first iteration result schematically illustrated (page 5).*

3) Les alignements de la séquence de levure et de la protéine SURE\_METJA obtenus lors de la première et la seconde itération sont présentés (page 6).

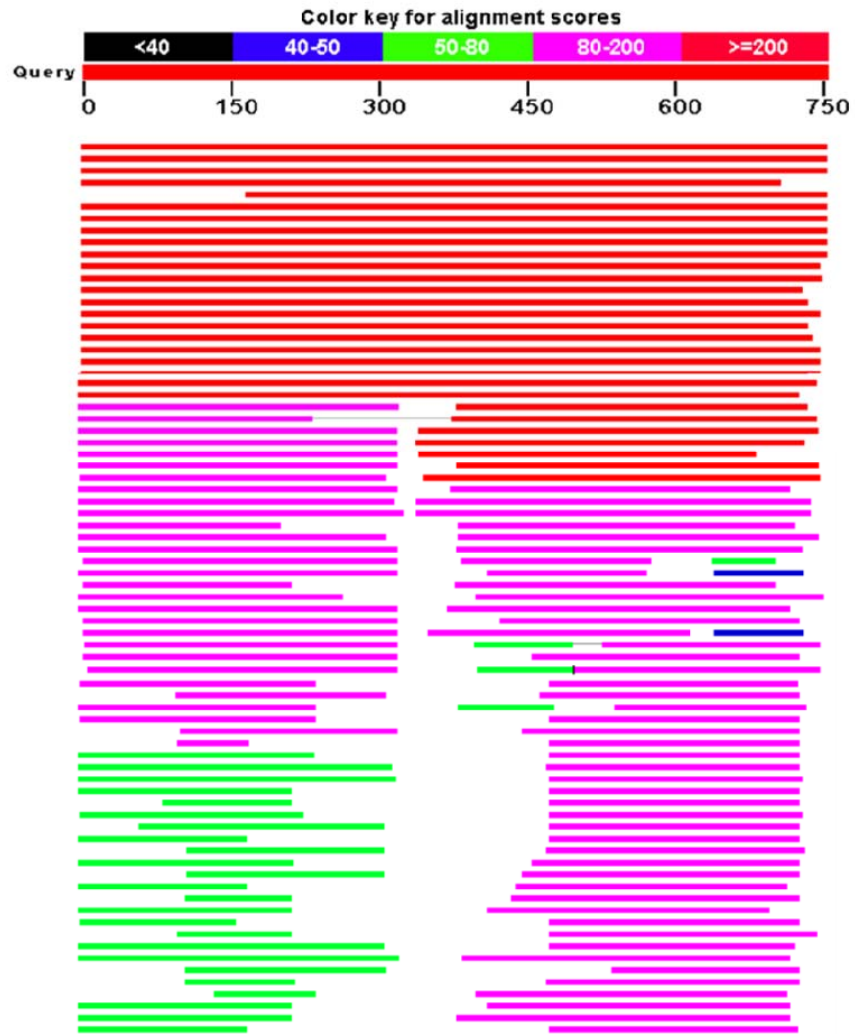
*The alignments between the yeast sequence and the SURE\_METJA protein obtained at the first and second iteration are shown (page 6).*

a) Quelle est la signification exacte de « Expect = 3e-06 » ? What is the exact meaning of « Expect = 3e-06 » ? 1.5 pts

b) Pourquoi le score de l'alignement a-t-il change entre les deux itérations ? 1 pt  
*Why has the alignment score changed between the two iterations?*

c) Que pouvez-vous dire sur la similarité entre les 2 protéines?  
*What can you say about the similarity between the two proteins?*

1.5 pts



## FIRST iteration

>sp|Q57979.2|SURE\_METJA RecName: Full=5'-nucleotidase surE; AltName:  
Full=Nucleoside 5'-monophosphate phosphohydrolase  
Length=266

Score = 58.9 bits (141), Expect = 3e-06, Method: Compositional matrix adjust.  
Identities = 55/219 (25%), Positives = 97/219 (44%), Gaps = 45/219 (20%)

```
Query 1 MRVLITNDGPLSDQFSPYIRPFIQHIKRNYPEWKITVCPVHVQKSWVGKAHLAAGKNLTA 60
M +LI NDDG +SP + +K + + IT+ P Q+S +G+A
Sbjct 1 MEILIVNDDG----IYSPSLIALYNALKEKFSANITIVAPTNNQSGIGRAI----- 48

Query 61 QFIYSKVDAEDNTFWGPFIQPQIRSENSKLPYVLNAEIPKDTIEWILIDGTPASCANIGL 120
+ + P +++ + KD + + + GTP C +G+
Sbjct 49 -----SLFEPLRMTKVK-----LAKDIVGY-AVSGTPTDCVILGI 82

Query 121 HLLSNEPFDLVLGSPNVGRNTSAAAYITSSGTVGGAMESVITGNTKAIASWAYFN--GL 177
+ + + DLV+SG N+G N I +SGT+G A E+ G K+IA S +
Sbjct 83 YQILKKVPDLVISGINIGENLQTE-IMTSGTLGAAFEAAHHG-AKSIASSLQITSDHLKLF 140

Query 178 KNVS-PLLMEKASKRSLDVIKHLVKNWDPKTDLYSINIP 215
K + P+ E +K + + + + ++D D+ +INIP
Sbjct 141 KELDIPINFEIPAKITAKIAEKYL-DYDMPCDVLNINIP 178
```

## SECOND iteration

>sp|Q57979.2|SURE\_METJA RecName: Full=5'-nucleotidase surE; AltName:  
Full=Nucleoside  
5'-monophosphate phosphohydrolase  
Length=266

Score = 210 bits (534), Expect = 5e-52, Method: Composition-based stats.  
Identities = 67/318 (21%), Positives = 119/318 (37%), Gaps = 71/318 (22%)

```
Query 1 MRVLITNDGPLSDQFSPYIRPFIQHIKRNYPEWKITVCPVHVQKSWVGKAHLAAGKNLTA 60
M +LI NDDG +SP + +K + + IT+ P Q+S +G+A + L
Sbjct 1 MEILIVNDDGI---YSPSLIALYNALKEKFSANITIVAPTNNQSGIGRAISLFEPLRM 56

Query 61 QFIYSKVDAEDNTFWGPFIQPQIRSENSKLPYVLNAEIPKDTIEWILIDGTPASCANIGL 120
+ D I + GTP C +G+
Sbjct 57 TKVKLAKD-----IVGYAVSGTPTDCVILGI 82

Query 121 HLLSNEPFDLVLGSPNVGRNTSAAAYITSSGTVGGAMESVITGN---TKAIASWAYFNGL 177
+ + + DLV+SG N+G N I +SGT+G A E+ G ++ I+ +
Sbjct 83 YQILKKVPDLVISGINIGENLQTE-IMTSGTLGAAFEAAHHGAKSIASSLQITSDHLKLF 141

Query 178 KNVSPLLMEKASKRSLDVIKHLVKNWDPKTDLYSINIPLVESLSDDTKVYAPIWENRWI 237
+ P+ E +K + + + + P D+ +INIP E+ + +T + + +
Sbjct 142 ELDIPINFEIPAKITAKIAEKYLDYDMP-CDVLNINIP--ENATLETPIEITRLARKMYT 198

Query 238 PIFNGPHINLENSFAEIEDGNESSSISFNWAPKFGAHKDSIHYMDEYKDRVLTDAEVI- 296
+E+ + S+ W D +E +D TD V+
Sbjct 199 -----THVEERIDPRGRSYYW-----IDGYPIFEEED----TDVYVLR 233

Query 297 ESEMISVTPMKATFKGVN 314
+ IS+TP+ N
Sbjct 234 KKRHISITPLTLDTTIKN 251
```