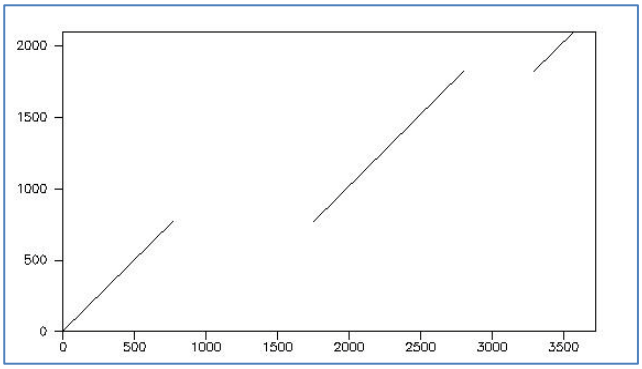
Examen intermédiaire 2013/2014
M1MABS: Harmonisation des Connaissances, partie Bioanalyse
Exam time : 2H (Write your answer on the exam paper)

Exercice 1 : Indicate on your copy, the CORRECT sentences

1. NCBI is a resource center regarding genetic diseases
2. Prosite is a protein database
3. A position weight matrix (PWM) is used to generate a profile
4. BLAST is a multiple alignment software
5. K-K-[DEN]-X-[MFP]-T-[LIV]-G-[HNT] is a profile
6. When using a distance score, indels penalties are less important than substitution penalties.
7. A FASTA sequence starts with the symbol <
8. ClustalW is a software to perform multiple alignment
9. A dot plot is a point matrix
10. Orthologous sequences are homologous

Exercice 2 :



1. Which software has been used to obtain this graph?
2. What kind of sequences are represented on the 2 axes of the graph?
3. How many local alignments will be generated?

Exercice 3 :

Get GL985084 sequence

1. From which is this sequence organism coming?
2. Has this organism been sequenced?
3. Represent as a small schema the architecture of the protein. Precise the protein length in amino acids, as well as the positions and roles of the functional domains.
4. What is the function of the protein?

Exercice 4 :

1. What was the objective of the below analysis?
2. Indicate the meaning of the term « Identities » and of the symbol «+» presented in the alignment
3. Is the result significative? Comment your answer.

```
>emb|CAA90081.1|  small GTP-binding protein [Pisum sativum]
Length=215

 Score =  285 bits (728),  Expect = 2e-94,
 Identities = 138/202 (68%), Positives = 161/202 (80%), Gaps = 7/202 (3%)

Query  8    DFLIKLLLIGDSGVGKSCCLLRFSEDSFTPSFITTIGIDFKIRTIELDGKRVKLQIWDTA  67
            D+LIKLLLIGDSGVGKSC LLRFS+ SFT SFITTIGIDFKIRTIELDGKR+KLQIWDTA
Sbjct  13   DYLIKLLLIGDSGVGKSCLLLRFSDGSFTTSFITTIGIDFKIRTIELDGKRIKLQIWDTA  72

Query  68   GQERFRTITTAYYRGAMGILLVYDVTDERSFNNIRTWFANVEQHATEGVNKILIGNKCDW  127
            GQERFRTITTAYYRGAMGILLVYDVTDE SFNNIR W  N+EQHA++ VNKIL+GNK D
Sbjct  73   GQERFRTITTAYYRGAMGILLVYDVTDEASFNNIRNWIRNIEQHASDNVNKILVGNKADM  132

Query  128  EE-KRVVSTERGQQLADELGIPFLEVSAKSNINIDKAFYSLAADIKKRLIDNQKNEQPAA  186
            +E KR V T +GQ LADE GI F E SAK+N+N+++ F+S+A DIK+RL D     +P
Sbjct  133  DESKRAVPTSKGQALADEYGIKFFETSAKTNMNVEEVFFSIARDIKQRLADTDSKSEPQT  192

Query  187  SGVNVGESSGSGGK------CC  202
              +N  + + +GG+      CC
Sbjct  193  IKINQQDPAANGGQAATKSACC  214
```

Exercice 5:

Is the below sequence a coding sequence?

Explain the experimental procedure you follow to answer this question.

>SeqInconnue
ATTTCCGAATATGCTGACTTTGTTCTGTGTCGTTGTTGGTGAGGGAAGAC
CTTTCTCCATCGACATTGATGTTGGAAAGACCGTCGATCATCTCAAGAAG
AAGATCAAGGAAGAGAACAAGAACACAATTTCTTGTGATGCGAAGGATCT
CCAGCTTTATCTGGCTTTGAAGGGTGGTTTACAGTTAAAGGATGGTGCGT
GGCTGTCTGACGAAGACCCTGATTTGGAAGGCCTTTCTCAACCCGCTGAA
GGAAACACAGTGTTACCAAAGTATGTCAATGAAGAAAGAAAGATGAGAGA
AACCAAGAAGCTTTCCAACTACTTTTCTGGTGGTGAAGATTACCCTGAAT
ATTGCGACGAAAAAATTCATGTGGTGGTGATTGTTCCAGAAGTTCCTTTG
TTGAAGGTGACCGCTCTAGAACCCTCAGTGCCAGTGCATCCCAGTGTTGA
CAGGAAGAGGCGATTTGATGAATTGAATCAAATCCTATCACAAGCTGAAA
TTGACGCATCAAATGATTCAAACAAGAAGCCAAAGAAATCTTCGAATTTT
TCTTCAATCAAATGGGAATTGGTCGCACCCTTGTTTAGCCGCGTTATGTC
GGCATATGAACAAGAAGAAAAAGCCATTCCGCGTGAAATTCTGCAAGAAC
TCCAGGATTACTCTGCCCGTGCGTTCACATGTTTCGAGCTGTCCAGTTGT
TCGGAGGCCACTCTCAACATCTTTATTGCCCCAGTGCTGGTCCAAGTATG
TGCATTATTTAACGGTGACATCAAAATCTTTGGAAAAGAAACTCTGAAAG
GGAAATATGTGAAGGCAAATGGTCGTTTTGAATTTGTATTGAGGAGAGGA
CTGAAGAGCATTTTCATTGTTGAAGCGAAGAAAGAGGATTTCGATCAAGG
TGCTGCGCAAGAATTGGTTGGGGCGGAAGTTGCGGCTGAGTTGGGAAGTT
TGAATGTTGTTTATGGGATCGTGACAAACTTCAAGGAATGGGTGTTCTTC
AAGAGCTCGAATACCAAAATTGAGAAAGATGCATCTTTCATGTATCATCC
ACCCAAACCATATTCAATGGAAACAATGTTGGCGAAAGCGACTGCCAAAA
TTTACGCCATACTTTTTGAATAACAATTTTATCAATTGTTGGCTCAGTGG
GTAAACAGCTAATTCATTCATCGGTCGTGTTTTTTTGCTTGAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA