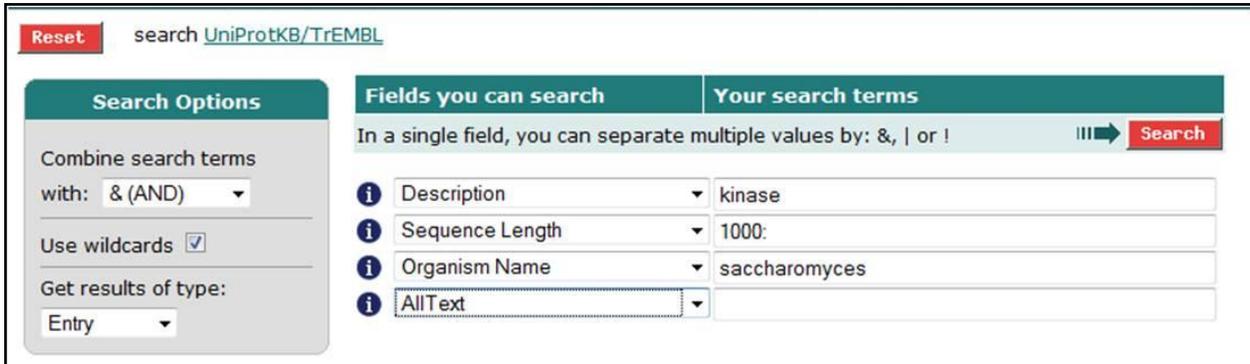


Examen : Bioanalyse (3L5BC5M) – Janvier 2009

Exercice 1 : banque de données et recherches d'information

1. Une requête a été réalisée. Les résultats sont présentés ci-dessous. Avec quel outil cette requête a-t-elle été effectuée ? Quel sera le résultat ?



The screenshot shows a search interface for UniProtKB/TrEMBL. It includes a 'Reset' button, a search bar with the text 'search UniProtKB/TrEMBL', and a 'Search Options' panel. The 'Search Options' panel has three sections: 'Combine search terms with: & (AND)', 'Use wildcards' (checked), and 'Get results of type: Entry'. To the right, there are two columns: 'Fields you can search' and 'Your search terms'. The 'Fields you can search' column lists 'Description', 'Sequence Length', 'Organism Name', and 'AllText'. The 'Your search terms' column contains 'kinase', '1000:', and 'saccharomyces'. A 'Search' button is located at the bottom right of the search terms area.

Exercice 2 : petit QCM de bioanalyse

Parmi les affirmations suivantes, indiquez sur votre copie lesquelles sont **VRAIES** :

- A. Blast est un programme d'alignement global plus rapide que ses prédécesseurs
- B. Blast calcule une E-value : une valeur comprise entre 0 et 1
- C. Blast calcule une E-value reflétant le nombre attendu d'alignements significatifs de score supérieur ou égal à celui obtenu entre la séquence requête et la séquence de la banque.
- D. La programmation dynamique est une méthode utilisée uniquement pour réaliser un alignement global entre 2 séquences
- E. Le numéro d'accèsion d'une séquence est définitif
- F. On choisira la matrice PAM350 plutôt que BLOSUM62 pour chercher des séquences éloignées
- G. Le dot-plot permet de visualiser les régions répétées présentes dans une séquence
- H. Le score de l'alignement ne dépend pas de sa longueur
- I. Le pourcentage de similarité est toujours supérieur ou égal au pourcentage d'identité
- J. Un score élémentaire peut-être négatif
- K. PUBMED est une méthode pour prédire les sites actifs d'une protéine
- L. G-A-[ILV]-X-D est une signature PROSITE
- M. Un motif protéique ne peut pas être présent dans un domaine fonctionnel
- N. La phylogénie a pour but de retracer l'histoire évolutive des espèces
- O. La phylogénie moléculaire permet d'obtenir des arbres, avec des nœuds, des branches et des fleurs
- P. En phylogénie, on utilise généralement 3 types de méthodes : la parcimonie, le maximum de vraisemblance et des méthodes de distance.

Exercice 3 : comparaison de deux séquences

Les 60 premiers acides aminés de deux protéines ont été alignés en utilisant un alignement global mais en faisant varier des paramètres.

Problème (Lire attentivement la totalité du sujet avant de répondre aux questions)

- 1) Vous avez séquencé un fragment d'ADN. L'analyse informatique de cette séquence révèle la présence d'un gène et vous permet d'obtenir sa séquence protéique. Expliquer la démarche que vous adopteriez pour réaliser la prédiction fonctionnelle (fonction putative, famille protéique, motifs etc.). Justifier chacune des étapes.
- 2) A l'issue de votre démarche en 1) vous avez sélectionné 11 séquences chez l'homme, la souris et la grenouille (RANES). L'alignement multiple est donné en Annexe 1. Vous avez réalisé l'analyse en domaines fonctionnels des séquences humaines sur le site du NCBI en effectuant votre recherche sur la banque CD (Conserved Domain) et obtenu les résultats présentés en Annexe 2.
 - Pouvez-vous proposer une fonction pour vos séquences ?
 - Combien de sous-familles obtenez-vous ? Justifiez votre réponse.
 - En utilisant l'architecture en domaines et l'alignement multiple, identifiez à quelle sous-famille appartiennent les sept autres séquences.
- 3) Vous avez ensuite recherché les motifs PROSITE sur la séquence SEQ7_HUMAN et trouvé qu'elle possédait 4 motifs liant le Ca^{2+} appelés EF-Hand et localisés comme suit :

SEQ7_HUMAN  (160 aa)

28 - 40: DADGGGDISvkEL

64 - 76: DEDGSGTIDfeEF

104 - 116: DRNADGYIDpeEL

140 - 152: DKNMDGRIDfdEF

Sachant qu'un motif EF-Hand fonctionnel débute toujours par l'acide aminé D, et connaissant l'architecture en domaines, identifiez sur l'alignement multiple ces motifs dans l'ensemble des protéines (les encadrer). Etablir la signature PROSITE du quatrième motif EF-Hand identifié.

- 4) Il existe en fait deux types de motifs EF-Hand, ceux qui lient spécifiquement le calcium (Ca^{2+}) et ceux qui lient le calcium ou le magnésium ($\text{Ca}^{2+}/\text{Mg}^{2+}$). Sachant que la séquence SEQ7_HUMAN correspondant à la troponine C du muscle squelettique possède deux sites Ca^{2+} et deux sites $\text{Ca}^{2+}/\text{Mg}^{2+}$ et que la séquence SEQ4_HUMAN correspondant à la parvalbumine alpha possède deux sites $\text{Ca}^{2+}/\text{Mg}^{2+}$, déterminer le type des motifs EF-Hand des deux autres protéines humaines (SEQ1_HUMAN et SEQ10_HUMAN).

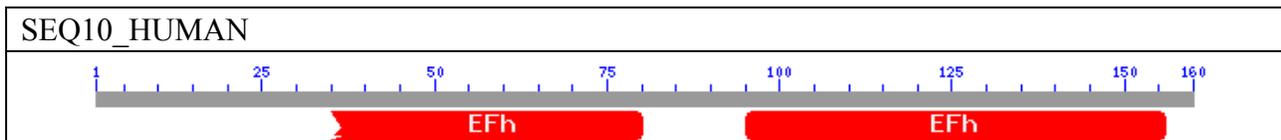
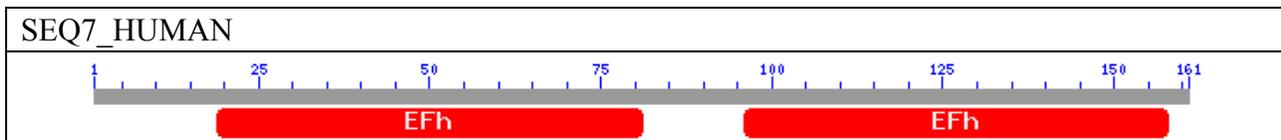
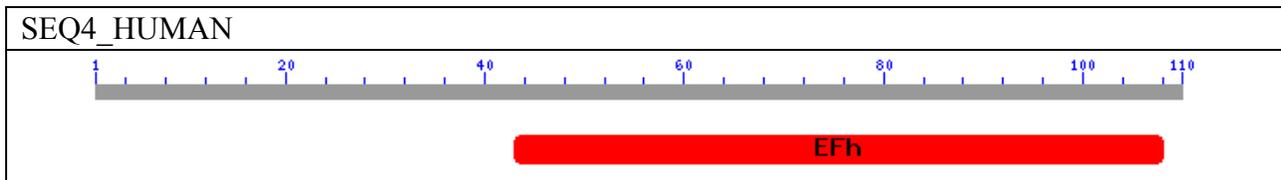
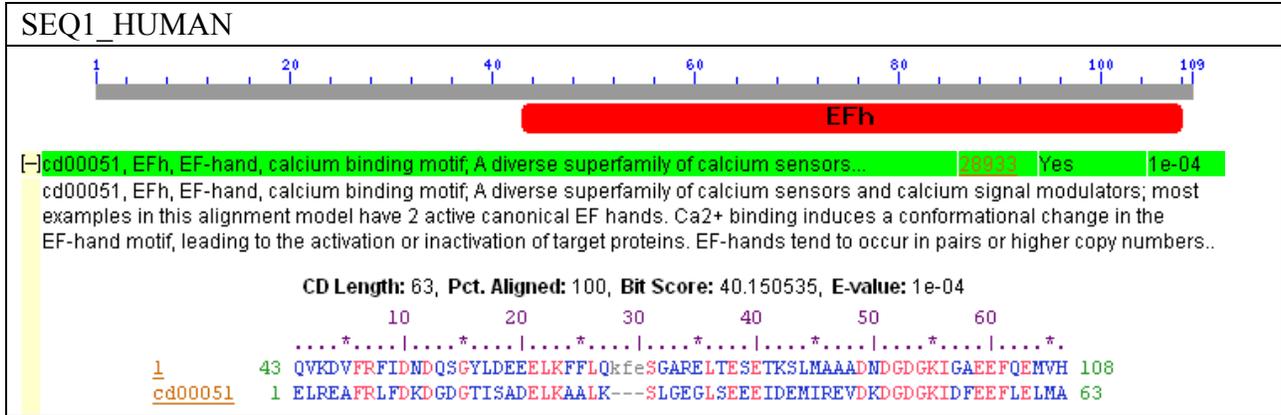
Annexe 1 : Alignement multiple

```
SEQ1_HUMAN -----MSITDVLSADDIA
SEQ2_MOUSE -----MSITDILSADDIA
SEQ3_MOUSE -----MSMTDVLSAEDIK
SEQ4_HUMAN -----MSMTDLLNAEDIK
SEQ5_RANES -----PMTDLLAAGDIS
SEQ6_RANES -----SITDIVSEKID
SEQ7_HUMAN --MTDQQAEARSYLSEEMIAEFKAAFDMF-DADGGGDISVKELGTVMRMLGQTPTKEELD
SEQ8_MOUSE --MTDQQAEARSYLSEEMIAEFKAAFDMF-DADGGGDISVKELGTVMRMLGQTPTKEELD
SEQ9_RANES AQPTDQQMDARSFLSEEMIAEFKAAFDMF-DTDGGGDISVKELGTVMRMLGQTPTKEELD
SEQ10_HUMAN --MDDIYKAAVEQLTEEQKNEFKAAFDFVFLGAEDGCISTKELGKVMRMLGQNPTPEELQ
SEQ11_MOUSE --MDDIYKAAVEQLTEEQKNEFKAAFDFVFLGAEDGCISTKELGKVMRMLGQNPTPEELQ
```

```
SEQ1_HUMAN AALQECQD--PDTFEPQKFFQTS-----LSKMSANQVKDVFRFIDNDQSGYLDEEELK
SEQ2_MOUSE AALQECQD--PDTFEPQKFFQTS-----LSKMSASQLKDFQFIDNDQSGYLDEDELK
SEQ4_MOUSE KAIGAFSA--ADSFHKKFFQMV-----LKKKNPDEVKKVFHILDKDKSGFIEDELG
SEQ4_HUMAN KAVGAFSA--TDSFDHKKFFQMV-----LKKKSADDVKKVFHMLDKDKSGFIEDELG
SEQ5_RANES KAVSAFAA--PESFNHKKFFELCG-----LKSKEIMQKVFHVLDDQDQSGFIEQDELG
SEQ6_RANES AALESVKA--AGSFNYKIFFQKVG-----LAGKSAADAKKVFIEILDRDKSGFIEDELG
SEQ7_HUMAN AIIIEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRI FDRNADGYIDPEELA
SEQ8_MOUSE AIIIEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRI FDRNADGYIDAEELA
SEQ9_RANES AIIIEVDEDGSGTIDFEEFLVMMVRQMKEDAQKSEEELAECFRI FDKNADGYIDSEELG
SEQ10_HUMAN EMIDEVDEDGSGTVDFDFEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELK
SEQ11_MOUSE EMIDEVDEDGSGTVDFDFEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELK
```

```
SEQ1_HUMAN FFLQKFESGARELSETKSLMAAADNDGDGKIGAEFQEMVHS--
SEQ2_MOUSE YFLQRFQSDARELSETKSLMDAADNDGDGKIGADEFQEMVHS--
SEQ3_MOUSE SILKGFSSDARDLSAKETKTLAAGDKDGDGKIGVEEFSTLVAES-
SEQ4_HUMAN FILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES-
SEQ5_RANES LILKGFTPTEGRSLSDKETTALLAAGDKDGDGKIGVDEFVTLVSES-
SEQ6_RANES LFLQNFRASARVLSDAETS AFLKAGSDGDGKIGVEEFQALVKA--
SEQ7_HUMAN EIFR---ASGEHVTDEEIESLMKDGDKNNNDGRIDDFDFLKMMEGVQ
SEQ8_MOUSE EIFR---ASGEHVTEEEIESLMKDGDKNNNDGRIDDFDFLKMMEGVQ
SEQ9_RANES EILR---SSGESITDEEIEELMKDGDKNNDGKIDFDFLKMMEGVQ
SEQ10_HUMAN IMLQ---ATGETITEDDIEELMKDGDKNNDGRIDYDFDFLEFMKGVE
SEQ11_MOUSE MMLQ---ATGETITEDDIEELMKDGDKNNDGRIDYDFDFLEFMKGVE
```

Annexe 2 : recherche de domaines fonctionnels



Par soucis de clarté, la définition du domaine fonctionnel détecté EFh a été conservée uniquement dans le cas de l'analyse de la séquence SEQ1_HUMAN. Pour les trois autres séquences, seul le résultat graphique est reporté.