

TP de Bioinformatique L3

Etude d'une séquence associée à une pathologie

Avant propos

A la fin de la séance de travaux pratiques, un compte rendu de 3 pages vous sera demandé. Veillez à prendre des notes et rédiger au fur et à mesure de la progression du TP car vous rendrez le compte rendu une semaine après la séance. Répondez aux questions posées et donnez l'interprétation de chaque analyse réalisée. Dans votre conclusion, faites une synthèse des résultats et proposez des analyses bioinformatiques complémentaires. La notation se fera sur la base de ce travail écrit mais aussi sur votre participation et sur l'intérêt que vous aurez porté à ce TP.

1) Introduction

Une équipe de zoologistes avec laquelle vous collaborez travaille sur une famille de tigres de Sibérie élevés en captivité. Les trois individus de la nouvelle portée souffrent d'une desquamation sévère et ont des problèmes de vision affectant leur santé. Les vétérinaires ont remarqué que ces symptômes sont proches des symptômes d'une carence en vitamine A. Après analyse de leur alimentation, ils constatent pourtant que l'apport nutritionnel en vitamine A est théoriquement largement suffisant. Malgré un apport supplémentaire mesuré (du fait des risques de survitaminose) en vitamine A, les troubles pathologiques persistent.

Cette espèce étant protégée, les zoologistes ont fait appel à votre laboratoire pour tenter d'identifier le ou les problèmes touchant cette portée. Le métabolisme de la vitamine A étant lié au foie, les investigations se sont portées sur cet organe. Des biopsies ont été réalisées et les échantillons analysés dans votre laboratoire. Grâce à des expériences de puces à ADN, un ARNm de 783bp a été identifié. Il est très fortement surexprimé dans le cas pathologique.

Ci après, la séquence de cet ARNm traduite (Vous pouvez aussi télécharger cette séquence à l'adresse : <http://www.ibpc.fr/~gelly/DOC/seq.fasta>) :

```
>translated mRNA
LSAGRELPVSFHPGRFRSETPSSRPGGPAAPADRRLLPWLHARPAPRPGRLRAAPVGGGLPGTMEWV
WALVLLAALGSARAEAEQVSNFQVKKNFDKARFAGTWYAMIKKDPEGLILQDNIVAQEFSDEN
GRMSATAKGRVRLINSWDVCADMVGTFTDTEPAKFKMKYWGVSFLQKGNDDHWIIDTDYDTY
AIQYSCRILNLDGTCADSYSFVFSRDPNGLPLEAQKIVRQRQEELCLARQYRLIVHNGYCDGKS
EPNTL
```

Votre laboratoire vous confie la tâche d'analyser cette séquence par les méthodes bio-informatiques usuelles.

2) Analyse de la séquence protéique

2-1) Recherche de séquences

Le site du NCBI permet d'accéder à de nombreux outils dont les plus utilisés sont les programmes de recherche de similarités tels que BLAST.

Recherchez des séquences similaires à la séquence identifiée grâce au programme BLAST (manipulez les séquences sous le format FASTA plus facile à utiliser).

Utilisez la banque RefSeq pour vos recherches. Quelles sont les caractéristiques de cette banque ?

Les méthodes d'alignements multiples permettent de mettre en évidence des conservations et des relations de parenté de manière beaucoup plus évidente que les programmes de recherche de similarité. Vous utiliserez le programme CLUSTALW (à télécharger sur

<http://www.ebi.ac.uk/clustalw/>) qui est un des programmes les plus utilisés pour ce type de tâche et réaliserez un alignement multiple entre les séquences. Quels sont les résidus conservés ? Quelles hypothèses pouvez-vous formuler sur la fonction de ces résidus conservés ?

L'outil SRS, accessible sur le site de l'EBI (European Bioinformatics Institute) à l'adresse <http://srs.ebi.ac.uk> permet d'interroger de manière élaborée de nombreuses banques. En utilisant SRS, déterminez la ou les fonctions des *PRBP*.

2-2) Recherche de motifs et de domaines

Comme vous avez pu le remarquer lors de vos recherches précédentes, les *PRBP* appartiennent à une famille plus large dont les différents membres sont présents dans de nombreux organismes. Quelle est cette superfamille ?

CLUSTALW vous a permis de mettre en évidence des acides aminés conservés au sein des *PRBP*. Ces acides aminés caractérisent un motif spécifique à la famille des *PRBP*. La superfamille de protéines à laquelle appartiennent les *PRBP* est aussi caractérisée par un motif, partagé par tous ses membres. En utilisant la banque PROSITE (<http://www.expasy.org/prosite/>), une banque de motif, déterminez le motif caractérisant cette superfamille. Comparez ce motif à celui mis en évidence par l'alignement multiple.

3) Analyse de structures

L'étude de la structure des protéines est essentielle pour comprendre la fonction des protéines et leur rôle physiologique au niveau moléculaire.

3-1) Analyse des structures protéiques

Les structures tridimensionnelles donnent des informations sur les mécanismes et la fonction moléculaire des protéines. La première structure d'un membre de la famille des *PRBP* a été élucidée par cristallographie et diffraction de rayon X en 1984. Depuis de nombreuses autres membres de la famille ainsi que des *PRBP* liées ou non liées à leur ligand ont été obtenues.

- Visualisation de la structure

RASMOL est un logiciel dédié à la visualisation de la structure des protéines à partir des coordonnées cartésiennes de leurs atomes. La *Protein Data Bank (PDB)* est la base de données dédiée au stockage de la structure des protéines obtenues expérimentalement (par RMN, cristallographie et diffraction de RX, et par microscopie électronique). Cette base est accessible à l'adresse <http://www.pdb.org> . Actuellement elle contient plus de 40000 fichiers de structures de protéines, d'acides nucléiques et sucres. Le format des fichiers est le format *PDB* qui contient, en outre, les coordonnées des atomes de la macromolécule permettant sa représentation.

description	→	HEADER	ENDODEOXYRIBONUCLEASE	08-MAR-91	1ATN
Coordonnées z	→	COMND	DEOXYRIBONUCLEASE I COMPLEX WITH ACTIN		
Coordonnées y	→	SOURCE	ACTIN FROM RABBIT (ORYCTOLAGUS CUNICULUS) MUSCLE,		
Coordonnées x	→	SOURCE	2 DEOXYRIBONUCLEASE I FROM BOVINE (BOS STAURUS) PANCREAS		
n° de l'acide aminé	→	AUTHOR	W. KABSCH, H. G. MANNHERZ, D. SUCK, E. PAI, K. C. HOLMES		
chaîne	→	REVDAT	1 15-JUL-92 1ATN 0		
type d'acide aminé	→	JRNL	AUTH W. KABSCH, H. G. MANNHERZ, D. SUCK, E. PAI, K. C. HOLMES		
type d'atome	→	JRNL	TITL ATOMIC STRUCTURE OF THE ACTIN: /IN\$ASE I COMPLEX		
n° de l'atome	→	JRNL	REI NATURE V. 347 37 1990		
	→	JRNL	REVW ASYM NATUAS UK ISSN 0028-0836		
Champs	→	ATOM	1 CA ACE A 0 105.046 51.546 40.626 1.00 72.72		
atom	→	ATOM	2 C ACE A 0 105.314 50.822 41.951 1.00 72.72		
	→	ATOM	3 O ACE A 0 105.220 51.451 43.013 1.00 72.56		
	→	ATOM	4 N ASP A 1 105.665 49.507 41.867 1.00 71.64		

Figure 1: Organisation d'un fichier PDB

Téléchargez sur le site de la *PDB* la structure de la *PRBP* humaine (code *PDB* [1brp](#)). Ouvrez le fichier et observez son organisation, puis répondez aux questions suivantes :

Par quelle méthode cette structure a elle été obtenue ? Que représentent les atomes HOH du champ HETATM ? Quelle est l'autre molécule présente ?

Téléchargez RASMOL, à l'adresse <http://openrasmol.org/> . L'interface de RASMOL est constituée d'un menu, d'une fenêtre de visualisation et d'une fenêtre supplémentaire permettant de taper des commandes dans un langage spécifique. Des exemples d'utilisation de ce langage et des descriptions sur les effets des différents mots clés sont disponibles sur le site <http://www.umass.edu/microbio/rasmol/seleccmd.htm>. Vous trouverez une aide plus complète à l'adresse : <http://openrasmol.org/doc/rasmol.html>.

Ouvrez le fichier 1brp avec RASMOL puis visualisez les structures secondaires avec la commande strand ribbon ou cartoon. Testez d'autres types de visualisation (sticks, backbone...). Que peuvent apporter ces différents modes de visualisation ?

Observez le positionnement du rétinol dans sa cavité. Comment le rétinol se fixe à cet endroit ?

Quels résidus sont en interaction avec le rétinol ? Mettez en évidence les résidus conservés identifiés par alignement multiple réalisé précédemment. Que constatez-vous ?

Y a-t-il d'autres molécules prises dans la cavité et pourquoi ?

- **Alignement de structures**

L'alignement de structure permet la comparaison de deux à plusieurs structures afin de mettre en évidence des parties structurales similaires. Il existe de nombreuses applications à ce type d'expérience ; par exemple la comparaison, chez une même protéine, entre une forme fixée au ligand et une forme non fixée donnera des informations sur les mouvements moléculaires provoqués par la fixation du ligand, la comparaison d'une même famille de protéine chez plusieurs organismes génétiquement éloignés mettra en évidence la conservation au cours de l'évolution d'un cœur commun et d'éléments essentiels au repliement, à la stabilité et à la conservation de la fonction.

DALILITE (<http://www.ebi.ac.uk/DaliLite>) est un programme de comparaison structurale. Ce programme utilise la carte de contacts Carbone α – Carbone α de chaque protéine à aligner pour produire le meilleur alignement structural possible.

En utilisant DALILITE, superposez les structures de [1brq](#) et [1rbp](#). Existe-t-il beaucoup de différences entre ces structures ? Détaillez ces différences et tentez de déduire le mécanisme fonctionnel de la *PRBP* ?

Téléchargez d'autres membres de la famille des *PRBP* et superposez-les à [1brp](#). Mettez en évidence au sein des différentes structures de protéines les motifs identifiés par PROSITE.

4) Conclusion

Proposez des hypothèses moléculaires et fonctionnelles sur la pathologie des tigres de Sibérie.