

TD5 : Prédiction de la structure tridimensionnelle d'une protéine Modélisation moléculaire

Vous aurez besoin des programmes suivant :
d'un éditeur de séquence
d'un visualiseur de structure 3D (PyMOL)

Avant-propos

Le fonctionnement d'un système biologique implique des cascades d'interactions. Les protéines en sont les principaux acteurs.

La capacité d'une protéine à interagir dépend de sa structure tridimensionnelle (3D). Connaître cette dernière permet donc de mieux comprendre son mode d'action (activité enzymatique, transport, signalisation, liaison avec un ligand, un récepteur, une membrane, etc).

Les méthodes expérimentales de détermination de la structure 3D des protéines sont lourdes et coûteuses en temps et ressources, voire inapplicables (cas des protéines non solubles, par exemple).

Les méthodes prédictives dites *in silico* proposent une alternative rapide et bon marché. Elles sont basées sur un ensemble de lois physiques, statistiques et biologiques. On convient généralement qu'il en existe 2 grandes classes : les méthodes dites de « modélisation comparative » et les méthodes « *ab initio* ». Les premières dépendent de l'existence de protéines homologues dont les structures ont été déterminées expérimentalement. Les secondes méthodes ne se basent que sur des lois physiques et statistiques. Les algorithmes utilisés par ces dernières sont très gourmands en temps de calcul et les résultats obtenus progressent avec les progrès en informatique. A l'heure actuelle, et malgré les immenses progrès des méthodes *ab initio*, les méthodes comparatives sont encore celles qui proposent les meilleures prédictions.

Lorsqu'on s'intéresse à la fonction d'une protéine, il est utile d'avoir recours à la modélisation afin de tester des hypothèses ou d'orienter des expériences.

Enfin, la modélisation moléculaire bénéficie de tous les avantages de l'informatique qui la rendent compatible avec les analyses à grande échelle.

Nous allons tenter de poser une hypothèse sur le rôle d'un SNP dans un gène d'intérêt

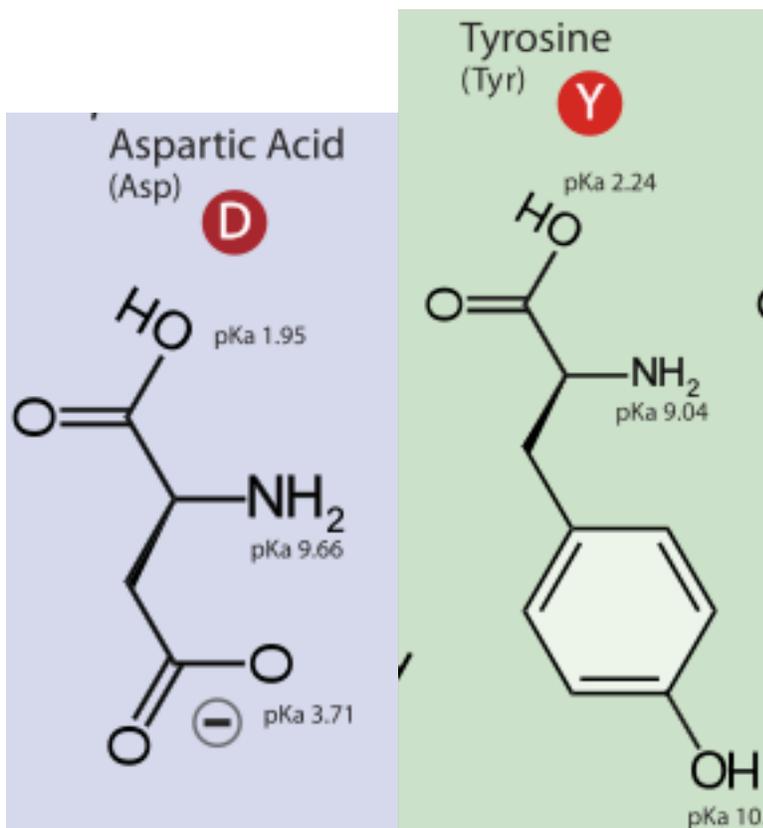
La séquence du gène est la suivante :

>TDR

```
MGRGDHLLMKNSNAAAAA AVNNGGTS LDAALRPLVGS DGWDYCIYWRLS
PDQRFLEMTGFCCSSELEA QVSALLDLPSSIPLDSSSIGMHAQALLSNQP
IWQSSSEEEEADGGGGAKTR LLVPVAGGLVELFASRYMAEEQQMAELVMA
QCGGGGAGDDGGGQAWPPP ETPSFQWDGGADAQRLMYGSSLNLFDAAAA
DDDPFLGGGGDAVGDEAAA AAGAWPYAGMAVSEPSVAVAQEQMQHAAGGG
VAESGSEGRKLHGGDPEDD GDGEGRS GGAKRQQCKNLEAERKRRKKNLGH
LYKLRS LVPNITKMDRASILG DAIDYIVGLQKQVKELQDELEDNHVHHKP
PDVLIDHPPPASLVGLDNDD ASPNSHQQPPLAVSGSSRRSNKDPAMT
DDKVGSGGGGGHRMEPQLE VRQVQGNELFVQVLWEHKPGGFVRLMDAMNA
LGLEVINVNVTTYKTLV LNVFRVMVRDSEVAVQADRVRDSLLEV TREYTP
GVWPSPOEEDDAKF DGGDGGQAAAAAAAAGGEHYHDEVGGGYHQHLHYLA
FD
```

En rouge, focus sur le SNP: D chez le sauvage devient Y chez le mutant
ITKMDRASIL devient ITKYRASIL

Un résidu chargé et hydrophile remplacé par un acide aminé neutre et hydrophobe
Quel peut être l'impact de cette mutation ponctuelle sur la fonction de la protéine ?



Récupération d'une séquence protéique d'intérêt :

Un simple copier coller depuis ce document ou sur le site de la formation
Sequence_TD5

Recherche de données sur la séquence :

Chercher des protéine similaires dans la base de données Expasy Swiss-Prot (<http://www.uniprot.org>).

Recherche de similarité de séquences avec *BlastP* :

Soumettre la séquence d'intérêt à Blast:

Choisir l'onglet *Blast* (blastP)

Copier la séquence ou charger le fichier puis sélectionner

Il est possible de sélectionner une base de données en particulier (laissez les réglages standards)

Lancer le Blast (peut nécessiter de 2 à 20 minutes ...) puis observer les *hits* retournés :

Que sont les protéines en question (fonction, poids moléculaire ?...)

Quelle est la plus grande couverture d'alignement et quel est le plus grand pourcentage d'identité obtenus ? Est-ce le même *hit* qui présente les deux ?

Exemple de résultat de requête :

The screenshot displays the UniProt BLAST search interface. At the top, there's a search bar and navigation tabs. Below, a color scale indicates identity percentages from 0 to 100. The 'Filter by' section on the left lists various categories like 'Reviewed (1)', 'Unreviewed (99)', and 'Organisms'. The 'Overview' section shows a table of search results:

Entry	Protein names	Match hit	Identity
Q6YU53	BHLH protein-like [Oryza sativa subsp. japonica]	100	100.0%
B8AGK5	Putative uncharacterized protein [Oryza sativa subsp. indica]	300	99.0%
T2B0H9	Uncharacterized protein [Oryza sativa subsp. indica]	300	99.0%
I1NWN6	Uncharacterized protein [Oryza glaberrima]	400	97.0%

The 'Alignments' section below shows detailed views for the top three hits, including alignment diagrams and statistics like E-value, Score, and Ident. For example, the first alignment for Q6YU53 shows an E-value of 0.0, a score of 2,811, and 100.0% identity.

Sélectionnez Q6YUS3

Explorer les différents types d'informations

Privilégiez l'ouverture des liens dans un nouvel onglet

Regardez particulièrement

Protein Attributes

General Annotations

Dans l'item « *Cross-references* :

Lancez une requête ModBase

Prêtez attention à DataSet Information

Ouvrir les liens PFAM et Protein Model Portal

PFAM

Consultez les informations

Protein Model Portal

Ouvrir le lien HLH (nouvel onglet)

Appréciez la qualité du model

Existe-t-il déjà une structure de cette protéine = Y a-t-il un numéro d'accès à la Protein DataBank (PDB) pro cette protéine ? Si oui, chercher la référence sur la PDB (<http://rcsb.org/pdb/>).

Est-ce une structure déterminée expérimentalement (Résonance Magnétique Nucléaire, Cristallographie aux rayons X) ?

Si oui, rentrer chez soi : un modèle expérimental vaut toujours mieux qu'un modèle théorique.

Si c'est un modèle théorique, essayer d'en produire un meilleur !

Modélisation :

Editez la séquence de départ pour ne prendre que la partie repérée avec Protein Model Portal soit la région de 250 à 360

Sequence_TD5_EDIT

atome2

<http://atome.cbs.cnrs.fr/AT2B/meta.html>

Donnez un titre

Collez la séquence

Laissez une adresse mail

Cliquez toutes les cases de Search for homologous ...

Warning: Access to @TOME2 server is restricted to academic use only !

This website is free and open to all academic users and there is no login requirement.

This website is optimized for Firefox.

[Documentation](#)

Embedded tools

[HHsearch](#)
[Fugue, PS3](#)
[Psiblast](#)

[Paired, P-Sea](#)

[Tito, Scwrl](#)
[Modeller](#)

[QMean](#)
[Verify3D](#)
[Eval23D](#)

[T-Coffee](#)
[Clustalw](#)
[Muscle](#)
[Jalview](#)

[MatCluster](#)
[Proft, Matt](#)

[Plants](#)
[MedanaScore](#)
[XScore](#)
[DSX](#)

[Pat](#)
[JMol](#)
[Babel](#)
[Fconv](#)
...

Title :

Primary Sequence: [One letter code] eg:
MPSHRNSNLKFPCTVCASNNRSMESHKVLQEAGYNVSSYGTGSAVRLPGLSIDKPNVYSFGT
PYNDIYNDLLSQSADRYKSNGLLQMLDRNRRLKKAPEKWQESTKVDFVFTCEERCDFAVCE
DLMNRGGKLNKIVHVINVDIKDDDENAKIGSKAILELADMLNDKIEQCEKDDIPFEDCIMI
LTEWQSSHSQPLSLYAPSY

Email :

Search for Homologous Sequences & Comparative Modeling

Select tools :

Psi-Blast (PDB) Options...

HHSearch (PDB) Options...

Fugue (Homstrad) Options...

SP3 Options...

Global Options:

Hide redundant structural alignments (>98%) in Atome selection

Ligands Selection & Complexes Prediction by Comparative Docking

Modules

Ref: [Pons & Labesse - Nucleic Acids Research, Web Server Issue, 2009](#)

Cliquez sur submit ... prenez un café

L'étape suivante consiste à valider un ou plusieurs templates pour lancer la génération du modèle 3D

Téléchargez le ou les modèles

Téléchargez le template utilisé pour la génération des modèles

Visualisation :

Lancer PyMOL et charger le modèle :

La structure générale est-elle cohérente ?

Essayez plusieurs types de représentation

Renommez le modèle

Chargez le template (1HLO.pdb)

Aligner les structures : *align prot1, prot2*

Explorer les fonctionnalités de PyMOL, et au besoin, consulter [PyMOL Wiki](http://www.py-molwiki.org/index.php/Main_Page) (http://www.py-molwiki.org/index.php/Main_Page).

Applications :

D'après la qualité du modèle obtenu, quelles applications sont et ne sont pas envisageables ?

Articles afférents

PMB

10.1007/s11103-013-0166-5

Molecular Breeding

10.1007/s11032-013-9972-3

Si vous avez des questions relatives à ce TD ou à cette thématique de recherche, n'hésitez pas à nous contacter : lamotte@supagro.inra.fr .

NB

Il est important de garder à l'esprit que les degrés de similitudes / différences entre acides aminés sont bien plus complexes que pour les nucléotides.

Les différences portent sur la taille, la rigidité du squelette, la charge, l'hydrophobicité ...

