

## 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

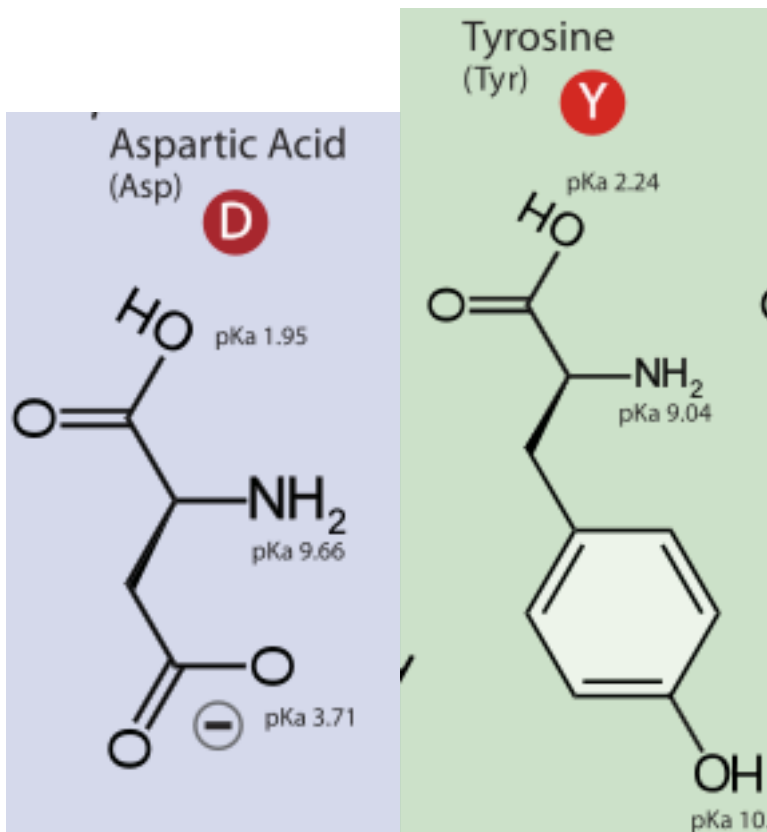
La séquence du gène est la suivante :

>TDR

```
MGRGDHLLMKNSNAAAAA AVNNGGTS LDAALRPLVGS DGWDYCIYWRLS
PDQRFLEMTGFCCSSELEAQVS ALLDLPSSIPLDSSSIGMHAQALLSNQP
IWQSSSEEEEEADGGGGAKTRLLVPVAGGLVELFASRYMAEEQQMAELVMA
QCGGGGAGDDGGGQAWPPPETPSFQWDGGADAQRLMYGGSSLNLFDAAAA
DDDPFLGGGGDAVGDEAAAAGAWPYAGMAVSEPSVAVAQEQMQHAAGGG
VAESGSEGRKLHGGDPEDDGDGEGRS GGAKRQQCKNLEAERKRRKKNLGH
LYKLRS LVPNITKMDRASILGDAIDYIVGLQKQVKELQDELEDNHVHHKP
PDVLIDHPPPASLVGLDNDDASPPNSHQQPPLAVSGSSRRSNKDPAMT
DDKVGSGGGGGHRMEPQLEVRQVQGNELFVQVLWEHKPGGFVRLMDAMNA
LGLEVINVNVTTYKTLVNLVFRVMVRDSEVAVQADRVRDSLLEVTTRETYP
GVWPSPOEEDDAKFDGGDGGQAAAAAAAAGGEHYHDEVGGGYHQHLHYLA
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 ?

Graphical overview

Color code for identity 0-100% =

Accession	Entry name	0Query hit552	0Match hit (sqrt scale)805	Name (Organism)
<input type="checkbox"/> Query	20120217607NDI19K			
<input type="checkbox"/> Q6YUS3	Q6YUS3_ORYSJ			BHLH protein-like (Oryza sativa subsp. japonica)
<input type="checkbox"/> B8AGK6	B8AGK6_ORYSI			Putative uncharacterized protein (Oryza sativa subsp. indica)
<input type="checkbox"/> F2EFZ7	F2EFZ7_HORVD			Predicted protein (Hordeum vulgare var. distichum)
<input type="checkbox"/> C0PKL7	C0PKL7_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> C0PHL6	C0PHL6_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> B4G0A3	B4G0A3_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> C5XSZ6	C5XSZ6_SORBI			Putative uncharacterized protein... (Sorghum bicolor)
<input type="checkbox"/> B6THL7	B6THL7_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> B4FB38	B4FB38_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> C0BP8	C0BP8_MAIZE			Putative uncharacterized protein (Zea mays)
<input type="checkbox"/> B6TPA8	B6TPA8_MAIZE			Helix-loop-helix DNA-binding domain c... (Zea mays)
<input type="checkbox"/> D7L887	D7L887_ARALL			Putative uncharacterized protein (Arabidopsis lyrata subsp. lyrata)
<input type="checkbox"/> B955L3	B955L3_RICCO			Transcription factor, putative (Ricinus communis)
<input type="checkbox"/> D85Q00	D85Q00_SELML			Putative uncharacterized protein AMS-1 (Selaginella moellendorffii)
<input type="checkbox"/> D8T8N9	D8T8N9_SELML			Putative uncharacterized protein AMS-2 (Selaginella moellendorffii)
<input type="checkbox"/> F6HHT7	F6HHT7_VITVI			Putative uncharacterized protein (Vitis vinifera)
<input type="checkbox"/> Q2VX2	AMS_ARATH			Transcription factor ABORTED MICROSPORES (Arabidopsis thaliana)
<input type="checkbox"/> Q84WX1	Q84WX1_BRANA			BHLH transcription factor (Brassica napus)
<input type="checkbox"/> F6I0Z6	F6I0Z6_VITVI			Putative uncharacterized protein (Vitis vinifera)
<input type="checkbox"/> A9T3F9	A9T3F9_PHYPA			Predicted protein (Physcomitrella patens subsp. patens)
<input type="checkbox"/> E5GBJ2	E5GBJ2_CUCME			BHLH transcription factor (Cucumis melo subsp. melo)
<input type="checkbox"/> B955L4	B955L4_RICCO			Putative uncharacterized protein (Ricinus communis)
<input type="checkbox"/> A9RSC1	A9RSC1_PHYPA			Predicted protein (Physcomitrella patens subsp. patens)
<input type="checkbox"/> B95HA3	B95HA3_RICCO			Putative uncharacterized protein (Ricinus communis)
<input type="checkbox"/> A5C879	A5C879_VITVI			Putative uncharacterized protein (Vitis vinifera)
<input type="checkbox"/> B955L8	B955L8_RICCO			Putative uncharacterized protein (Ricinus communis)

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 ...

Cliquez sur submit ... prenez un café

**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](#)  
[Paiblast](#)

[Pipred, 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:  
MPSHRNSNLKFCCTVCASNRRSMESHKVLQEAGYNVSSYGTGSAVRLPGLSIDKPNVYSFGT  
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 & Lobesse - Nucleic Acids Research, Web Server Issue, 2009](#)

L'étape suivante consiste à valider un ou plusieurs templates pour lancer la dénégation 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](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

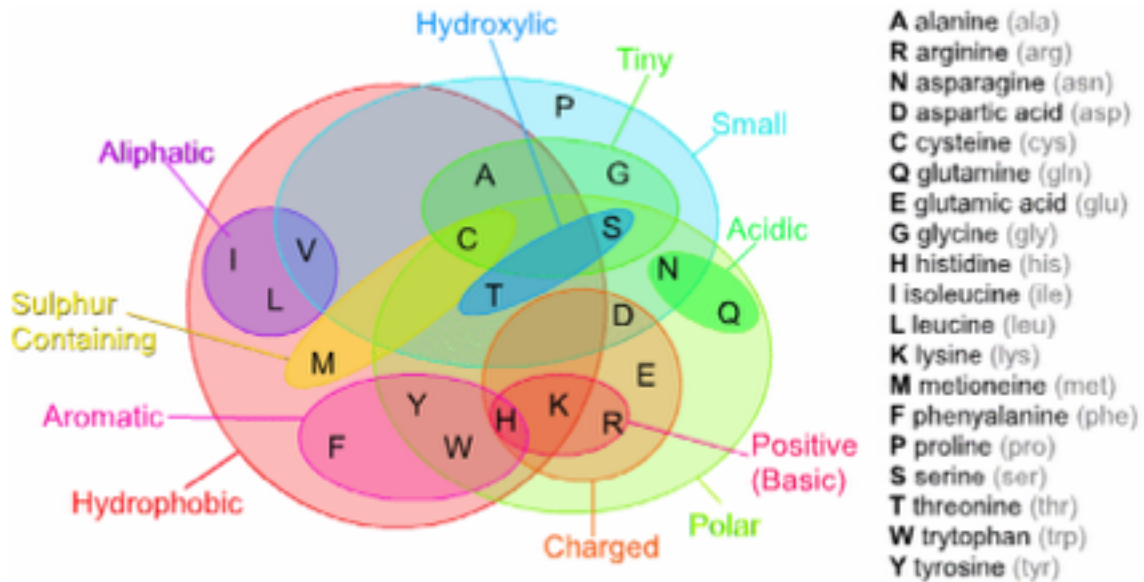
10.1007/s11032-013-9972-3

*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](mailto: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 ...



Lien pour terminer la modélisation

**MERCI DE NE PAS ACTIVER CE LIEN !**

<http://atome.cbs.cnrs.fr/AT2B/TEMP/1364313/atome.html>