

TP modélisation par homologie avec pymod

1 Préambule

1.1 Terminal

Pour ouvrir un Terminal sous Ubuntu : **Ctrl+Alt+T**

Commandes de base en Linux / Terminal :

1. Montrer le path du répertoire actuel : **pwd**

2. Changer de répertoire :

cd relativePath/

cd /absolutePath/

cd .. (remonte le path)

cd ~ (home-directory ou « dossier personnel »)

3. Afficher le contenu d'un répertoire : **ls -l** ("list" avec l'option "long")

4. Lancer un programme qui est dans le répertoire actuel :

./programme

5. Compresser un dossier :

tar cvzf archive.tar.gz dossier/

6. Décompresser un archive dans le répertoire actuel :

tar xvzf archive.tar.gz

7. Copier :

cp source cible

1.2 pymol

pymol se lance à partir d'un Terminal. Il s'ouvre une fenêtre avec deux parties, nous utilisons principalement la grande partie noir en bas. En bas de cette fenêtre vous pouvez taper les commandes pymol. Pour obtenir de l'aide sur une commande taper `help <commande>`.

Comme dans un Terminal on peut revoir les anciennes commandes dans pymol avec les curseurs haut/bas du clavier. La touche "Echap" permet de basculer entre l'affichage 3D et le retour en texte des commandes pymol.

Avec le bouton gauche de la souris on tourne la protéine. Avec le bouton droit de la souris on peut zoomer et avec le bouton du milieu (=la molette) on peut déplacer la protéine.

Sinon vous avez aussi le wiki de pymol avec beaucoup d'exemples :

http://www.pymolwiki.org/index.php/Main_Page

http://www.pymolwiki.org/index.php/Practical_Pymol_for_Beginners

Essayer aussi mes TP d'intro sur pymol disponibles ici :

<http://www.impmc.upmc.fr/~stratmann/cours/proteinStructure/>

1.3 Démarrer pymod

pymod est un plugin pour pymol pour faire de la modélisation par homologie sous pymol en intégrant plusieurs outils, comme BLAST, Clustal et modeller. pymod est installé sur la machine virtuelle.

1. Dans la machine virtuelle lancer pymol depuis le Terminal en tapant simplement :
pymol
2. Aller dans le menu Plugin de pymol, il devrait maintenant y apparaître une entrée "PyMod 3". Cliquer dessus, puis une fenêtre devrait s'ouvrir avec le titre "PyMod 3".

Sur cette page

<http://schubert.bio.uniroma1.it/pymod/download.html>

vous pouvez télécharger le "User's Guide" de pymod et les fichiers d'installation si vous voulez l'installer chez vous.

2 Premier exemple avec pymod

2.1 Les projets pymod

On peut sauvegarder et charger ses projets pymod via le menu **File->Projects** de pymod. Pensez à faire une sauvegarde de temps en temps pendant que vous avancez dans le TP. Pour sauvegarder plusieurs points d'avancement, utilisez des numéros, par exemple projet01.pmse, projet02.pmse, etc.

Par défaut les projets sont enregistrés sous pymod/projects/

2.2 Importer la séquence cible (= "target sequence")

Ce premier exemple est pris du User's Guide de pymod (chapitre 5 Usage Examples). Il s'agit de générer des modèles par homologie d'une seule chaîne, ici la "dihydrofolate reductase of Mycobacterium avium".

1. Récupérer la séquence en format FASTA depuis le site www.uniprot.org avec l'identifiant O30463. Une fois sur la page de la séquence en question il faut cliquer sur le bouton bleu "Format" (en dessous du titre) puis cliquer droit sur "FASTA (canonical)", puis "Enregistrer la cible du lien sous...".
2. Sur cette même page d'uniprot chercher s'il y a déjà des structures 3D résolues pour cette protéine et noter leur code PDB. Cliquer aussi sur le lien vers le "ProteinModelPortal" et noter avec quel template 3D le modèle a été obtenu.
3. Encore sur la page d'uniprot chercher la partie "Sequence" puis noter la longueur de la séquence.
4. Dans la fenêtre pymod importer la séquence via **File -> Sequences and Structures -> Open from File**

2.3 Recherche et sélection de template 3D

5. Dans la fenêtre pymod : Sélectionner la séquence cible précédemment importé en cliquant sur son nom à gauche. La couleur devrait changer de rouge en vert. Sélectionner maintenant le menu **Tools -> Database Search -> BLAST -> Local BLAST**. Cette interface permet de changer quelques options de BLAST, puis de faire une recherche avec BLAST sur la base de données locale déjà téléchargée dans le dossier pymod. Ici on ne change pas les options, on doit juste sélectionner la base de données : il y a deux options **pdbaa** et **swissprot**, la première étant les séquences d'acides aminés (AA) des structures de la PDB et la deuxième un ensemble de séquences plus large avec des séquences sans structure PDB associée. On choisit **pdbaa** ici, car on cherche des templates 3D pour notre séquence requête. Si vous faites

un BLAST directement sur le serveur de la NCBI, alors la banque "nr" serait sélectionné par défaut, c'est-à-dire une très grande banque de séquences de protéines pour la plupart sans structure PDB associée. Cliquer sur "Submit" pour lancer BLAST avec votre séquence cible. Une fenêtre "BLAST Results" s'ouvre et affiche la liste des séquences similaires trouvées (agrandir pour voir toutes les colonnes).

6. La première entrée a une identité de séquence de 100%, ce qui vient du fait qu'une structure 3D a déjà été résolue pour notre séquence cible. On n'utilisera bien sûr pas cette structure comme template, mais on pourra l'utiliser pour vérifier nos modèles générés. Pour comprendre la signification des colonnes "Query span" et "Subject span" on refait ce run de BLAST via le site de la NCBI directement.
7. Aller sur le serveur de BLAST de la NCBI et choisir un protein-BLAST. Pour faire une recherche on peut donner directement le code uniprot de notre séquence, puis il faut faire attention de bien choisir la bonne "database". Sur la page des résultats on voit d'abord des informations sur notre séquence "query", comme sa longueur et sa décomposition possible en domaines. En dessous il y a un graphique avec des lignes horizontales colorées. Chaque ligne correspond à une séquence similaire au query. Qu'est-ce qu'on peut dire sur la partie C-terminal de notre séquence (les dix derniers résidus environ)? En cliquant sur une ligne l'alignement qui se trouve plus bas de la page est affiché. Chercher l'entrée 4KL9. En regardant cet alignement que veut dire "Subject span" dans la fenêtre pymol? Aller sur la RCSB PDB puis chercher l'entrée 4KL9, puis choisir l'onglet "Sequence". Est-ce que la numérotation des résidus dans la structure correspond au "Subject span"?
8. Retour sur la fenêtre "BLAST Results" de pymol : Dans la longue liste de templates possibles on choisit de prendre deux templates de deux espèces différentes : *Bacillus anthracis* et *Moritella profunda*. Pour chaque espèce on prend la première entrée de la liste, c'est-à-dire les PDBs 3JW3 et 2ZZA. Cocher la case de ces deux entrées de la liste, puis cliquer sur "Submit". L'alignement de séquences s'affiche dans fenêtre de pymol. Regarder cet alignement tout le long de la séquence en utilisant la barre de défilement horizontal. Noter quelle partie de la séquence cible est alignée avec la séquence de 3JW3 et de 2ZZA, respectivement. A quoi correspondent les symboles *, . et : (voir cours)? Colorier les séquences avec **Display -> Color all Sequences -> By residue properties -> Polarity**.
9. Importer les structures 3D des templates 3JW3 et 2ZZA en faisant un clique-droit sur le nom de chaque séquence, puis choisir **Structure -> Fetch PDB file** dans le menu local. On doit choisir si on veut importer tout le fichier PDB, c'est-à-dire toutes les chaînes avec ligands et cofacteurs ou seulement la partie qui est alignée avec la séquence query. Ici on choisit **Import in PyMod the structure of every chain of the PDB files**, comme on voudrait créer un modèle avec le cofacteur NADP. Si on aurait des templates multi-domaines, alors il aurait fallu choisir l'autre option pour ne garder que la partie des templates qui correspond au domaine à modéliser. Une fois les deux templates 3JW3 et 2ZZA importés on voit que les séquences sont remplacées par les séquences des fichiers PDB sans alignement avec notre séquence cible. Chaque PDB a deux chaînes, A et B. Est-ce qu'il s'agit de dimères biologiques ici? Pour répondre regarder l'annotation sur la RCSB PDB pour les deux templates. Pour mieux comprendre la différence entre *Asymmetric Unit* et *Biological Assembly* lire ceci : pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies Comme on veut créer un monomère comme modèle, on peut effacer les deux chaînes B avec le menu local **Sequence -> Delete sequence** en faisant un clique-droit sur leur nom.
10. Dans la fenêtre pymol on peut voir les structures 3D des deux chaînes, coloriées par la couleur de la séquence dans la fenêtre pymol. Arrangez les deux fenêtres pour qu'elles s'affichent en même temps, par exemple en utilisant les raccourcis clavier "Touche Windows" + "flèche

gauche ou droite” pour diviser l’écran en deux parties. Pour centrer la vue 3D sur une chaîne, il suffit de cliquer sur son nom dans la fenêtre pymod pour l’activer (le nom devient vert) puis faire un clique avec le bouton du milieu (= la molette) de la souris. Pour cacher une chaîne on utilise aussi le bouton du milieu, mais sur un nom de chaîne désactivé, c’est-à-dire en rouge. Pour sélectionner un résidu d’une chaîne, on clique avec le bouton du milieu sur le résidu en question dans la fenêtre pymod. Pour avoir une vue de l’ensemble des structures 3D on peut utiliser la commande **zoom** dans l’invite de commande de **pymol** (pas pymod!).

2.4 Alignement

Après la sélection des templates, il faut encore les aligner correctement avec notre séquence cible. Cet alignement peut être guidé par un profil généré en superposant en 3D les templates.

11. Générer ce profil “structural “ sélectionner les deux chaînes 3JW3_Chain_A et 2ZZA_Chain_A (elles doivent être en vert) puis choisir le menu **Tools -> Alignment Tools -> Structural alignment -> SALIGN (Structural alignment)**. Attention de ne pas choisir SALIGN (Sequence alignment)! Laisser les options par défaut et cliquer sur “Submit”. Les structures vont être superposées dans la fenêtre pymol et l’alignement qui en résulte au niveau séquence est montré dans la fenêtre pymod. Les symboles *, : et . se réfèrent maintenant à l’alignement de deux séquences sans notre séquence cible. Inspecter dans la fenêtre pymol l’alignement structural. Quelles sont les zones qui sont moins bien superposées? Pour répondre à cette question activer l’affichage des séquences dans pymol avec le menu de pymol : **Display -> Sequence**. En sélectionnant un résidu dans la structure, celui-ci s’affiche aussi dans les séquences et inversement.
12. Pour aligner notre séquence cible à l’alignement des deux templates, on utilise des méthodes d’alignement séquence-profil, comme celles qui sont implémenté dans pymod : ClustalW, SALIGN ou Clustal Omega. Le “profil” est ici simplement l’alignement obtenu lors de la question précédente. On sélectionne donc notre séquence cible et les deux séquences des templates (tout doit être en vert en dessous de “SEQUENCE ID”). Puis, on choisit ici Clustal Omega avec le menu **Tools -> Alignment Tools -> Profile alignment -> Clustal Omega**. Laisser les options par défaut, puis cliquer sur “Submit”. Sauvegarder l’alignement obtenu en ouvrant le menu local (clique-droit sur “Alignment”) et en sélectionnant **Edit Cluster -> Save Alignment to File**. Sauvegarder l’alignement dans le dossier de votre projet pymod avec un nom de fichier avec l’extension “.fasta”, par exemple “align1.fasta”. Charger l’alignement sauvegardé pour dupliquer l’alignement, afin de le modifier manuellement tout en gardant “l’original “ : Choisir **File -> Alignment -> Open From File** puis choisir le fichier qu’on vient de sauvegarder.
13. Inspecter l’alignement obtenu sur l’alignement original (“pas imported”). *ATTENTION : On peut manuellement changer l’alignement avec la souris : on sélectionne un résidu avec la souris et on reste appuyé avec le bouton gauche, tous les résidus à partir de ce résidus sont déplacés avec la souris. Faire ces opérations uniquement sur l’alignement dupliqué (“imported”)* Quelles zones ont une faible identité de séquence? Pour quelles zones de notre séquence cible on n’a aucune information structurale à partir des templates? Vérifier les positions des gaps dans la structure, soit avec clique-milieu sur les résidus dans pymod, soit en changeant la couleur de ces résidus dans pymol, par exemple avec cette commande dans l’invite de commande pymol :
color blue, resi 51-52
Est-ce que cela sera facile d’insérer les résidus de notre séquence cible sur ces deux structures?

14. **(Avancé)** Essayer d'améliorer manuellement l'alignement avec la souris sur l'alignement dupliqué ("imported") en gardant l'alignement original inchangé.
15. Les derniers 13 résidus de notre séquence cible n'ont aucune correspondance dans l'alignement avec les templates. Ces résidus seront trop difficile à modéliser, pour cela on préfère de les enlever : Cliquez-droit sur le nom de notre séquence cible, puis **Sequence -> Edit Sequence**. Effacer les 13 derniers résidus. Faire cette opération pour les deux alignements (original et imported).

2.5 Génération du modèle

En prenant l'alignement original cible <-> templates obtenu par Clustal Omega, on génère un modèle avec MODELLER.

16. Sélectionner la séquence cible dans pymod (sous "Alignment 1 (merged)") puis choisir le menu **Tools -> Modeling -> MODELLER (HOMOLOGY Modeling)**. Dans les options on sélectionne d'abord les deux templates dans **Use as Template**, puis pour 3JW3 on sélectionne **Do not use any heteroatomic residue** et on n'inclut pas les molécules d'eau de 3JW3. Pour 2ZZA par contre on sélectionne **Select single heteroatomic residues**, puis on coche "NAP". On coche aussi pour les 157 molécules d'eau à inclure. Voir le "User's Guide" de pymod pour plus de détail sur ces options (partie 5.1.4 Model building). Avant de lancer Modeller, on choisit encore quelques options dans l'onglet "Option" : On met le "Optimization Level" à "High" et on choisit le "DOPE Score" dans "Color models by". Après vérification de toutes les options on clique sur "Submit". Le calcul peut prendre quelques minutes.

2.6 Analyse du modèle

Une fois la génération du modèle terminé correctement, les résultats seront affichés dans pymod et pymol. Dans la fenêtre pymol on voit le modèle qui est coloré par le score DOPE. Le modèle est superposé au templates qui sont également évalués par le score DOPE. Ces scores par résidu sont aussi affichés dans une fenêtre à part. En cliquant sur un endroit de la courbe, le résidu est sélectionné sous pymol.

17. Identifier les zones avec grandes valeurs du score DOPE, c'est-à-dire les zones potentiellement mal prédites. Regarder ces zones dans les structures. Comparer ces zones par rapport à votre inspection de l'alignement obtenu avec Clustal Omega (position des gaps, etc).
18. Comparer le modèle obtenu avec la structure expérimentale de notre cible : Taper dans pymol :
fetch 2W3W
Importer la séquence de cette structure dans pymod avec le menu "File -> Sequences -> Import PyMOL Objects". Puis pour superposer cette structure de référence avec notre modèle, on sélectionne le modèle ("model_1_1_chain_A") et la séquence 2W3W dans pymod et on choisit "Tools -> Alignment Tools -> Structure alignment -> SALIGN (Structural alignment)". Utiliser dans pymol la commande **zoom** pour centrer la vue sur les deux protéines.
Dans pymol n'afficher que le modèle et 2W3W, puis inspecter l'accord. Est-ce que le score DOPE a bien prédit les zones mal prédites?
19. **Pour aller plus loin** : revenez sur l'étape d'alignement cible <-> templates et essayer de trouver un alignement qui améliore les zones mal prédites. C'est-à-dire pour chaque nouveau alignement il faut régénérer un modèle avec Modeller, puis le comparer avec la référence.

3 Deuxième exemple avec pymod

Le "User's Guide" de pymod contient un deuxième exemple comme tutoriel : la modélisation par homologie d'un complexe de protéines. Suivre les instructions (en anglais) du "User's Guide", partie 5.2.