

Conversion d'une séquence protéique d'un fichier PDB en un fichier Anagène

Publié le 10.04.03

Auteurs : Marie-Jeanne Pellerin et Gilles Furelaud

Table des matières

1. Introduction
 2. Récupération de la séquence dans le fichier PDB
 3. Récupération de la macro AA3.dot (site SVT de l'académie de Lyon)
 4. Conversion par AA3.dot
 5. Ouverture dans Anagène
 6. Sauvegarde pour une réutilisation ultérieure
-

1. 1. Introduction

Les fichiers PDB (.pdb) permettent la visualisation d'une molécule en trois dimensions grâce à des logiciels tels que Rasmol, Rastop ou Chime. Ces fichiers PDB comportent donc toutes les informations concernant la structure de la molécule : position des atomes, des liaisons, etc. Dans le cas d'une protéine, le fichier PDB comprend aussi la séquence de cette protéine.

Or de très nombreux sites sur Internet proposent des fichiers PDB à visualiser et à télécharger. L'enseignant de SVT dispose ainsi d'un large accès à ces fichiers moléculaires. D'un point de vue pédagogique, il est souvent utile de fournir aux élèves, en plus de la structure tridimensionnelle d'une protéine (structure tertiaire), sa séquence (structure primaire). De même, l'usage d'Internet montre qu'il est souvent plus facile de trouver dans les sites académiques (et autres) le fichier PDB d'une protéine plutôt que sa séquence...

Il apparaît donc qu'il est particulièrement utile de pouvoir "récupérer" une séquence protéique exploitable à partir de ces fichiers PDB.

Le logiciel le plus couramment utilisé dans les classes de lycée pour l'exploitation des séquences de protéines est le logiciel Anagène. Ce document présente donc une procédure permettant, en utilisant une macro développée par le site SVT de l'académie de Lyon, de convertir la séquence protéique récupérable dans un fichier PDB en un fichier utilisable sous Anagène. Cette procédure est décrite sous Word 97 et sous Word XP (en absence de précision, les procédures sont identiques).

Ce protocole reprend et adapte la page publiée par le [site académique SVT de Lyon](#), en y rajoutant les procédures propres à Office XP développées par Marie-Jeanne Pellerin.

2. 2. Récupération de la séquence dans le fichier PDB

Le comprenant les coordonnées tridimensionnelles de la protéine permet de le visualiser en trois dimensions, grâce à Rasmol, Rastop, Chime, Protein Explorer, MOLusc, etc. Ce fichier est un fichier .pdb, .ent, .mol, ou encore .xyz. Dans tous les cas, la séquence de la protéine y est présente.

Pour retrouver cette séquence, ouvrir le fichier dans un traitement de texte (Word, Wordpad, etc.). Après les caractéristiques de la molécule et de nombreuses remarques ('REMARK'), on trouve la-dite séquence d'acides aminés.

Elle apparaît sous la forme suivante (des légères différences peuvent exister selon les types de fichiers) :

SEQRES 1 A 298 MET GLU ASN PHE GLN LYS VAL GLU LYS ILE GLY GLU GLY

1FIN 100

SEQRES 2 A 298 THR TYR GLY VAL VAL TYR LYS ALA ARG ASN LYS LEU THR

1FIN 101

SEQRES 3 A 298 GLY GLU VAL VAL ALA LEU LYS LYS ILE ARG LEU ASP THR

1FIN 102 *etc.*

Explications :

- SEQRES : indique qu'il s'agit de la séquence
- 1 : numéro de ligne
- A : précise qu'il s'agit de la "chaîne" A
- 298 : nombre total de résidus

De plus, le nom du fichier (ici 1FIN) peut apparaître en plus à la fin de chaque ligne

Ceci n'est pas interprétable par Anagène. Il va falloir éliminer toutes les informations "parasites", grâce à une macro Word, c'est à dire un petit programme qui va effectuer une série de commandes prédéterminées sous Word.

A l'aide de la souris, sélectionner l'ensemble de la séquence, dans le fichier PDB (attention : si plusieurs protéines sont présentes, ne sélectionner que la séquence de la molécule voulue, identifiée par sa "chaîne"). Copier cette séquence.

Pour pouvoir poursuivre, il faut avoir récupéré la macro Word en question...

3. 3. Récupération de la macro AA3.dot (site académique SVT de Lyon)

Le fichier AA3.dot est en fait un fichier modèle de Word 97, intégrant la macro. Vous pouvez télécharger la macro AA3.dot en cliquant sur [ce lien](#), et la sauvegarder dans votre ordinateur :

Sous Word 97	Sous Word XP
Vous devez copier le fichier AA3.dot dans le répertoire Modèles de Microsoft Office , lui-même dans le répertoire Program Files .	Vous pouvez copier AA3.dot où vous voulez. Il faut juste que cela soit un endroit facile à vous rappeler ou à retrouver.

4. 4. Conversion par AA3.dot

Si ce n'est déjà fait, **ouvrir Word**.

Sous Word 97	Sous Word XP
<p>Choisir Fichier puis Nouveau, et sélectionner AA3.dot.</p> <p>Dans la page vierge qui s'affiche, coller la séquence copiée précédemment sur le fichier PDB.</p> <p>Choisir Outil, puis Macro..., et enfin Macros. Dans la boîte de dialogue qui s'affiche, choisir AA3 et cliquer sur Exécuter : la macro est alors exécutée par Word 97.</p>	<p>Ouvrir le fichier AA3.dot (Fichier, puis Ouvrir, et sélectionner AA3.dot là où vous l'avez sauvegardé).</p> <p>Si besoin il faut alors activer les macros : choisir Outils, puis Macros, et enfin Sécurité. Positionner alors cette sécurité sur faible, afin de permettre l'exécution des macros.</p> <p>Dans le document (AA3.dot), coller la séquence copiée précédemment sur le fichier PDB.</p> <p>Choisir Outils, puis Macro et enfin Macros. Dans la fenêtre, choisir AA3 et cliquer sur Exécuter : la macro est alors exécutée par Word XP.</p>

La macro élimine les chiffres, les espaces, le mot SEQRES, et converti le code 3 lettres en code 1 lettre.

Si la séquence comportait le nom du fichier (1FIN dans notre exemple), celui-ci n'est pas éliminé par la macro ! Il faut alors enlever manuellement de tels nom subsistant. Pour éviter toute fausse manipulation, il est préférable de réaliser cette "élimination" avant d'exécuter la macro...

Choisir enfin **Fichier**, puis **Enregistrer sous**. Dans la fenêtre, taper un nom de fichier de votre choix, et sélectionner **texte brut** comme type de fichier : le fichier est alors enregistré sous le nom "votrenom.txt". Il est préférable d'enregistrer de fichier dans le répertoire "Sauve" d'Anagène.

Ouvrir l'explorateur de fichiers, et modifier le nom du fichier de "votrenom.txt" en "votrenom.pro". Ne pas tenir compte d'un éventuel message d'alerte.

Fermer Word.

5. 5. Ouverture dans Anagène

Ouvrir Anagène, choisir **Fichier**, puis **Ouvrir**. Sélectionner le fichier que vous venez de renommer "votrenom.pro".

Anagène devrait alors vous signaler que ce fichier n'a pas été créé par Anagène, et vous demander : "Voulez-vous l'ouvrir comme fichier de séquence ?". Cliquer alors sur **OK**. (Il peut être nécessaire de préciser qu'il s'agit bien d'une séquence protéique)

6. 6. Sauvegarde pour une réutilisation ultérieure

Maintenant qu'Anagène a ouvert le fichier "votrenom.pro" comme une séquence protéique, vous pouvez l'enregistrer par Fichier, Enregistrer. Vous pouvez alors choisir le format Anagène qui vous convient le mieux...

Le fichier est maintenant ouvrable directement par Anagène simplement en double-cliquant dessus.