

## MODELISATION EN CHIMIE

### Utilisation de VMD pour l'analyse de sites enzymatiques (séance 1)

#### Contexte

La visualisation des structures des macromolécules biologiques comme les protéines est un outil d'analyse très utilisé. Le nombre de structures cristallographiques des protéines est de plus en plus grand. Le graphisme moléculaire permet de mieux comprendre ces structures et de pouvoir prédire de nouveaux modèles.

#### Les protéines

Une **protéine** est une [macromolécule](#) biologique formée d'une ou de plusieurs chaînes [polypeptidiques](#). Chacune de ces chaînes est constituée de [résidus](#) d'[acides aminés](#) liés entre eux par des [liaisons peptidiques](#). On parle généralement de protéine au-delà d'une cinquantaine de résidus dans la molécule<sup>1</sup>, et de [peptide](#) jusqu'à quelques dizaines de résidus.

La nature des protéines est déterminée avant tout par leur séquence en acides aminés, qui constitue leur [structure primaire](#). Les acides aminés ayant des propriétés chimiques fort diverses, leur disposition le long de la chaîne polypeptidique détermine leur arrangement spatial. Celui-ci est décrit localement par leur [structure secondaire](#), stabilisée par des [liaisons hydrogène](#) entre résidus d'acides aminés voisins, et globalement par leur [structure tertiaire](#), stabilisée par l'ensemble des interactions entre les résidus — parfois très éloignés sur la séquence peptidique mais mis en contact spatialement par le [repliement de la protéine](#) — ainsi qu'entre la protéine elle-même et son environnement ; la [réticulation](#) de plusieurs chaînes peptidiques entre elles par des [ponts disulfure](#) entre résidus de [cystéine](#) est également décrite au niveau de la structure tertiaire de la protéine. Enfin, l'assemblage de plusieurs [sous-unités protéiques](#) pour former un complexe fonctionnel est décrit par la [structure quaternaire](#) de cet ensemble.

Voir <https://fr.wikipedia.org/wiki/Protéine>

Les 20 acides aminés

<b>Acide glutamique</b>	Glu	E	<b>Leucine</b>	Leu	L
<b>Acide aspartique</b>	Asp	D	<b>Lysine</b>	Lys	K
<b>Alanine</b>	Ala	A	<b>Méthionine</b>	Met	M
<b>Arginine</b>	Arg	R	<b>Phénylalanine</b>	Phe	F
<b>Asparagine</b>	Asn	N	<b>Proline</b>	Pro	P
<b>Cystéine</b>	Cys	C	<b>Sérine</b>	Ser	S
<b>Glutamine</b>	Gln	Q	<b>Thréonine</b>	Thr	T
<b>Glycine</b>	Gly	G	<b>Tryptophane</b>	Trp	W
<b>Histidine</b>	His	H	<b>Tyrosine</b>	Tyr	Y
<b>Isoleucine</b>	Ile	I	<b>Valine</b>	Val	V

## La Protein Data Bank (PDB)

Les structures cristallographiques des protéines obtenues par des méthodes de cristallographie de rayon X, une analyse RMN, par reconstruction 3D à partir de microscopie cryo-électron, ou bien par modélisation, sont répertoriées dans une banque de données la PDB (Protéine Data Bank). Le site internet de cette banque de donnée est : <http://www.rcsb.org>

Il y a plus de 100000 structures disponibles dans la PDB en 2015. Les statistiques peuvent se trouver sur la page :

<http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>

Les structures cristallographiques sont répertoriées avec un code PDB, alphanumérique de 4 caractères. L'identifiant PDB est publié par les auteurs qui ont déterminé la structure. Pour chaque identifiant, une page web donne les informations sur la structure et un fichier dit en format PDB donne l'ensemble des informations lisibles par des logiciels de modélisation.

Une protéine peut avoir plusieurs entrées dans la banque de données car elle peut être cristallisée par plusieurs groupes, ou avec des mutations, ou dans des conditions différentes (pH par ex), avec des ligands différents...Par exemple l'hémoglobine a plus de 400 entrées. Lorsqu'on utilise un structure PDB, on donne le code référence et on cite la publication origine de la structure.

## Format d'un fichier PDB

Les fichiers .pdb sont des fichiers de texte qui peuvent s'ouvrir avec un éditeur de texte (Wordpad, vi, textedit ...). Chaque ligne commence par un mot clé et a un format spécifique.

Le format des fichiers .pdb peut se trouver en cliquant sur le menu Learn /Understanding pdb data ou en allant directement :

[http://www.rcsb.org/pdb/101/static101.do?p=education\\_discussion/Looking-at-Structures/intro.html](http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/intro.html)

La première partie du fichier donne des informations générales et importantes sur la structure. Elle commence par le mot clé HEADER et finit lorsque le mot clé ATOM commence. On peut vérifier dans les rubriques HEADER, TITLE, COMPND et SOURCE qu'on a bien la structure souhaitée.

La partie commençant par ATOM/HETATM donne les coordonnées des atomes. ATOM est le mot clé devant les atomes d'une protéine ou d'un acide nucléique, le mot HETAM correspond aux autres molécules (ligand, ions, solvant ...). Le mot clé TER permet de définir la fin d'une chaîne.

Exemple :

```
ATOM 1 N HIS A 6 49.668 24.248 10.436 1.00 25.00
```

Cette ligne correspond à l'atome d'azote, qui est l'atome 1 du résidu Histidine numéro 6 de la chaîne A avec les coordonnées x =49.668 Angström y=24.248 Angström z=10.436 Angström. Taux d'occupation « Occupancy » 1 (il n'y a pas de conformations autres détectées, sinon ce nombre serait plus petit que 1), facteur de température ou B-factor (plus l'atome bouge, plus ce nombre est grand) de 25.0

## Applications

### Exercice 1 : Prise en main

Le but de cet exercice est de se familiariser avec la PDB et le logiciel VMD. Nous allons prendre comme exemple la protéine aquamarine CFP de la figure 1 de l'article « Lumières sur le vivant » de F. Mérola, H. Pasquier et M. Erard, L'actualité chimique juin-juillet 2015, page 23-28 (voir TP de Hulis).

- 1) Récupérer la structure : Dans le moteur de recherche de la PDB, taper Cyan Fluorescent Protein. Vous devez trouver 3 entrées : 1OXD.pdb, 1OXE.pdb et 1OXF.pdb
- 2) Télécharger 1OXD.pdb
- 3) Avec VMD, ouvrir 1OXD.pdb : on voit tous les atomes représentés sous un fond noir.
- 4) Dans la fenêtre Graphics puis Representation, choisir Drawing Method : New Ribbons (voir ce que les autres Drawing Method proposent comme représentation).
- 5) Un clic sur la touche du clavier R puis un clic gauche permet de faire tourner la molécule dans l'espace. Un clic sur la touche S puis clic gauche permet de zoomer ou dézoomer. Un clic sur la touche T puis clic gauche permet de déplacer la molécule.
- 6) Dans la fenêtre Graphical Representation, regarder ce que les choix de Coloring Method et de Material changent.
- 7) Nous allons chercher à faire la même figure que la figure 1 de l'article : il faut donc faire apparaître le chromophore. Dans un éditeur de texte, chercher dans le fichier pdb le code à 3 lettres du chromophore.
- 8) Dans Graphical Representation, cliquer sur Create Rep. Cela crée une nouvelle ligne. Dans la fenêtre Selected Atoms taper : « resname » suivi du code à 3 lettres que vous avez repéré dans le fichier pdb. Vous pouvez obtenir la même chose en allant sur l'onglet « selections » puis Keyword « resname » puis choix du code à 3 lettres. Rien ne change sur la figure si vous êtes en drawing Method NewRibbons. Choisir la bonne Drawing Method pour obtenir un dessin proche de la figure. Choisir un Coloring Method adapté.
- 9) Changer la couleur du fond en blanc: onglet dans VMD Main : Graphics Colors Display Background.
- 10) Sauver l'image : VMD main : onglet File Render ou bien en faisant un imprimer écran de Windows.
- 11) Pour sauver les changements faits sur VMD, on peut sauver un fichier .vmd dans File « Save State » et choisir un nom finissant par .vmd. On peut, après avoir fermé VMD, récupérer son travail en faisant File « load State » et choisir le fichier qui finit par .vmd

### Exercice 2 : superposer deux structures

On va ouvrir une des autres conformations 1OXE.pdb pour comparer avec 1OXD.pdb.

- 1) Si vous avez fermé la fenêtre de travail de l'exercice 1, ouvrez le fichier .vmd que vous avez sauvé avec la commande load State.
- 2) Ouvrir 1OXE.pdb : File, New Molecule , trouver l'emplacement du fichier, load.

Choisir la même drawing Method mais utiliser une couleur différente pour cette molécule. Les deux molécules en forme «New Ribbon» se superposent très bien. . Si besoin élargir un des rubans (thickness) pour enlever la double couleur.

3) Dans VMD Main, onglet Extensions, Analysis, RMSD Calculator, choisir de faire le calcul du RMSD (root mean square deviation) entre tous les atomes de la molécule : « all » dans la fenêtre et décocher Backbone only. Un clic sur le bouton RMSD donne la valeur du RMSD. Si on avait besoin de superposer les structures ce qui n'est pas le cas dans cet exemple, on aurait cliqué d'abord sur « align » puis sur « RMSD ».

4) Sur le graphe, repérer la boucle qui diffère entre les deux structures. En cliquant sur 1 puis sur le ruban avant et après la boucle, repérer le numéro des résidus concernant cette boucle.

On peut effacer les labels dans Graphics/Label/ sélectionner les labels et « Delete ».

Dans Graphical Représentation, créer une nouvelle représentation pour les deux molécules et sélectionner les numéros des résidus (avec « resid XXX to YYY », XXX et YYY étant les numéros trouvés. Vérifier que la sélection « residue XXX to YYY » donne un décalage vers le bas car la numérotation commence à 0 et est linéaire or il manque les résidus 65 et 67 dans le fichier pdb).

5) Faire apparaître le résidu TYR 145 avec le drawing method « licorice » pour les deux structures. Faire un fond blanc, imprimer et sauvegarder votre travail.

### 6) Questions à faire après avoir abordé l'exercice 3 :

a) Dans Analysis, RMSD Calculator, sélectionner les résidus des boucles (resid XXX to YYY) puis calculer le RMSD et faire align, calculer le RMSD, observer les deux structures. Pour revenir aux structures initiales, refaire un align avec une sélection « all ».

b) Regarder la distance entre la Tyrosine TYR 145 et le chromophore : Pour avoir le label de la mesure de la distance entre deux atomes, on clique sur 2 puis on clique sur les 2 atomes. Remarque : pour celui d'un angle, on clique sur 3 puis sur les 3 atomes. On peut effacer les labels dans Graphics/Label/ sélectionner les labels et « Delete ».

c) On peut sélectionner les atomes qui sont à proximité d'une région, par exemple pour avoir les atomes à moins de 5 angströms du résidu CFR, on peut taper dans la sélection:  
« within 5 of resid 145 ».

Faire apparaître les résidus à 5 Angströms du chromophore en rose (color ID numéro à trouver).

### Exercice 3 (bonus): visualisation d'une dynamique

Fermer VMD et le rouvrir. Dans le répertoire VMD/proteins, récupérer la structure pdb alanin.pdb via (File/New Molecule). Sinon, récupérer ce fichier sur le site :  
[alpha.univ-mlv.fr/MODELISATION-MASTER](http://alpha.univ-mlv.fr/MODELISATION-MASTER)

On va récupérer une trajectoire de dynamique, c'est à dire l'ensemble des structures lors d'une simulation de la molécule qui bouge. Avec un clic droit sur alanin.pdb écrit dans la fenêtre « VMD Main », choisir « Load data into molecule » VMD/proteins/alanin.dcd ou alanin2.dcd.

Maintenant, dans Frames, on doit voir 100 Frames. Dans VMD Main, on peut regarder les différentes conformations avec le curseur ou choisir un défilé automatique.

Faire apparaître et observer les liaisons H qui se forment et se cassent.

Dans Analysis, Hydrogen bond, faire l'analyse du nombre de liaison H par frame.

**Complément** Tutoriel de VMD : [http://gvallver.perso.univ-pau.fr/doc/tut\\_vmd.pdf](http://gvallver.perso.univ-pau.fr/doc/tut_vmd.pdf)