

Modélisation en chimie: utilisation de VMD pour l'analyse de sites enzymatiques

Cours de Master M1

Isabelle Navizet

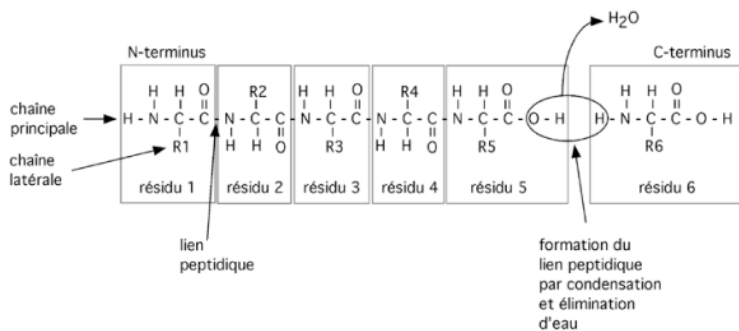
Isabelle.navizet@u-pem.fr

Laboratoire Modélisation et Simulation Multi-Echelle
(MSME)

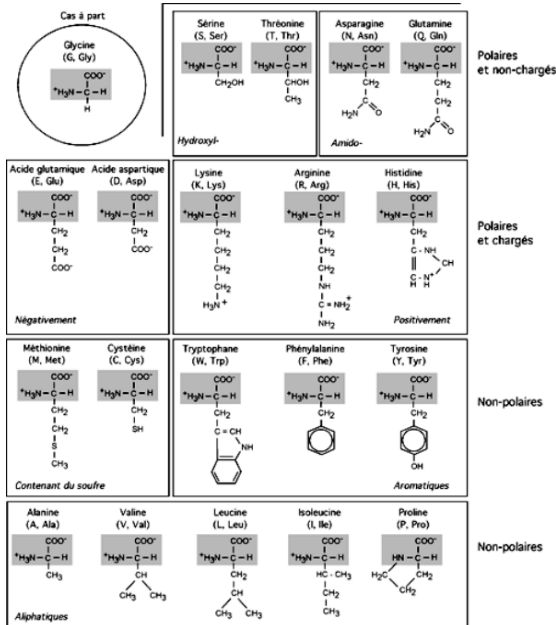
Les macromolécules du vivant: les protéines

- <https://fr.wikipedia.org/wiki/Protéine>
- 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, et de [peptide](#) jusqu'à quelques dizaines de résidus.

Les macromolécules du vivant: les protéines



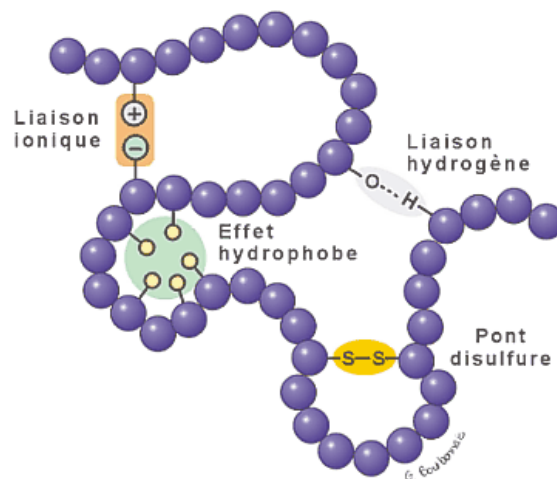
20 acides aminés de propriétés chimiques spécifiques



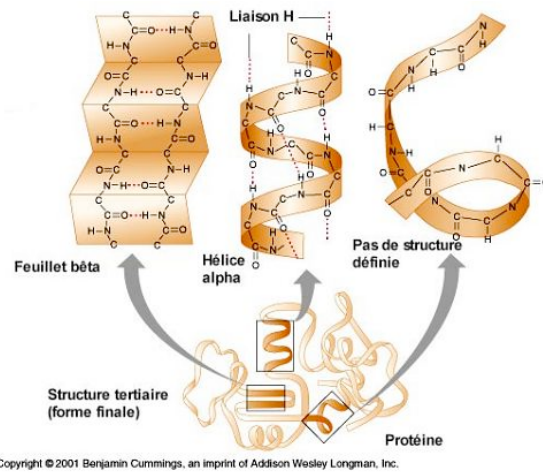
Structure primaire

- **Exemple: Le lysozyme est une protéine formée de l'assemblage, dans un ordre bien précis, de 130 acides aminés. Chacun des mots de trois lettres de cette liste représente un acide aminé.**

LYS VAL PHE GLU ARG CYS GLU LEU ALA ARG THR LEU LYS ARG LEU
 GLY MET ASP GLY TYR ARG GLY ILE SER LEU ALA ASN TRP MET CYS LEU
 ALA LYS TRP GLU SER GLY TYR ASN THR ARG ALA THR ASN TYR ASN
 ALA GLY ASP ARG SER THR ASP TYR GLY ILE PHE GLN ILE ASN SER ARG
 TYR TRP CYS ASN ASP GLY LYS THR PRO GLY ALA VAL ASN ALA CYS HIS
 LEU SER CYS SER ALA LEU LEU GLN ASP ASN ILE ALA ASP ALA VAL ALA
 CYS ALA LYS ARG VAL VAL ARG ASP PRO GLN GLY ILE ARG ALA TRP VAL
 ALA TRP ARG ASN ARG CYS GLN ASN ARG ASP VAL ARG GLN TYR VAL
 GLN GLY CYS GLY VAL



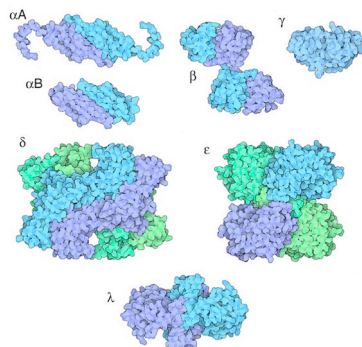
Structures secondaires et tertiaires



Structure quaternaire

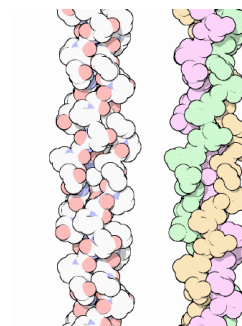
Exemple: Crystallines

<http://www.rcsb.org/pdb/101/motm.do?momID=127>



Collagène

<http://www.rcsb.org/pdb/101/motm.do?momID=4>



Où trouver les structures 3D des protéines ?

- La protéine data Bank: <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>
- 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
- Données expérimentales: RMN, cristallographiques : Coordonnées des atomes
- Visualisation des structures: logiciel VMD: <http://www.ks.uiuc.edu/Research/vmd/>
- Tutoriel de VMD : http://gvallver.perso.univ-pau.fr/doc/tut_vmd.pdf

- 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
- 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