Bioinformatique Structurale

# Notions de base (rappels)

### Concepts de la biologie

caract majeure : extreme complexité.

Facteurs de complexité : nombreux et antinomiques avec la modélisation.

* immensité des variables
* intéraction avec les variables
* variabilité du comportement
* échelles des phénomènes
* perception dynamique

modélisation : on décide de laisser des choses decoté pour aller a l’essentiel. (drug design)

simulation : Expliquer comment un médicament rentre dans un site actif …(dynamique moléculaire).

A quel moment s’éloigne –t-on trop de la réalité ? A force d’éliminer des paramètres, on élimine bcp de petits mouvements moléculaires qui, tous additionnés, forment un mouvement qu’on oublie…

### Conséquences de cette accélération technologique

Seule la cristallogénèse permet d’accéder à la structure.

Définition du modèle :

Lorsqu’on simplifie un modèle, on perd de l’information. Il faut utiliser des paramètres pertinants.

Selection de paramètres  explications, prédictions : C’est le cas du drug design.

Tester des hypothèses, les vérifier.

Qques expl de modélisations : modèle symbolique : séquences linéaires de lettres, 1D, pour l’ADN. modéles géométriques, modèles énergétiques, modèle statique ou dynamique.

Ere du haut débit : réponse au gigantismes : 3 bases de données énormes à disposition.

Depuis 82, le nb de bases dans genBank double env. ts les 18mois. Ce n’est pas le cas du nb de structures dispos.

BdD pour les small molecules : BD de cambridge.

# PDB

2 types d’info :

* cristallographie : grosse prots
* RMN : fragments de prot, peptides… (on peut pas dépasser 100 ou 200 positions atomiques aujourd’hui avec cette technique).

Bcp de modélisation de prots membranaires, mais tjrs un risque (car la base n’est pas structurellement défini). Plsrs fctions biologiques peuvent partager la meme architecture (un repliement…). Mais s’il existe des sites qui prédisent le repliement 3D à partir de la séquence, rien n’est moins sûr : ça marche svt par homologie.

Si les infos viennent de la cristallo : le constexte est figé, mais meme dans ces structures, il y qques infos sur la dynamique de la molécule dans son activité biologique (surtout grace aux molecules d’eau). Le repliement se fait avant que la cristallisation se fasse et non l’inverse. Donc les infos concernant la dynamique existant au moment du repliement sont normalement conservées.

Faire attention au critère de qualité.

Nuages électroniques de diffraction : c’est de la que viennent les coordonnées atomiques. Ces nuages sont relativement imprécis.

Résolution d’une structure n’est pas homogène. Il y a des zones où la densité electronique est trop flou  les positions ne sont pas données apr le cristallographe. La première chose à faire dans la modélisation est de rajouter les zones manquantes (meme de 4 ou 5 aa), avant de faire la minimisation.

Faire de la modélisation : nécessaire de partir d’une structure fiable. Puis il faut la minimiser, à cause des imprécisions des positions atomiques.

RMN : on mesure des couplages proton – proton  reconstitution géométrique à partir d’un tableau de couplages de distances (protons-protons).

Représentation topologique des repliements : 1 coup sur 2 c’est pas bon.

Les coordonnées cartésiennes (dispos dans les fichiers PDB.txt) n’ont aucunes relation avec les coordonnées internes (issues de la physique de la molecule (quantique), voir cours suivants).

Angles dièdres : projection de deux plans.

Chiralité : elle conditionne la chimie de ts les syteme vivants.

Des labos testent des aa en configuration R alors qu'ils n'existent qu'en S dans la nature → permet de tester des pptés qui n'ont pas encore été testées par la nature (intéressant en drug design, au niveau de l'affinité p.ex).

# Architecture, fonctions organiques et MM

Les molécules sont biochimiques. Elles sont chimiques avant d'etre Biologiques.

On va parler d'alcanes : construisent des architectures complexes.

Facile à modéliser ?

Leur rôle ?

Apport énergétique ?

Alcanes :

Constituées uniquement de chaines carbonées saturées (atomes C et H), en sp3.

Des pseudo-alcanes : alcanes qui ont subit dans substitutions (O, N, S …) : relativmt facil à faire d'un point de vue chimique (car toujours en sp3).

4grandes classes d'alcanes :

* chaines linéaires : assez peu rigides : linéaires, angles dièdres → rotation assez libre autour des C – C.
* alcanes ramifiées : compacité + importante → ridité + importante.
* Alcanes cycliques : très rigide. CnH2n. Pour le tordre, ça nécessite des conditions de contraintes très particulière.

Dès que le nombre de C augmente, les combinaisons sont très importantes.

Notion de carbones primaires, secondaires, tertiaires ou quaternaires.

A masse égale, plus c'est ramifié, plus le point d'ébullition est bas.

Pourquoi ?

La T°C d'ébullition indique l'energie à fournir pour dissocier les molécules → plus c'est compacte, plus il faut fournir d'energie. Forces attractive (van der waals) et forces dispersives (London).

Les plus linéaires s'enchevêtrent beaucoup entre molécules voisines, ils faut donc bcp d'énergie pour les dissocier.

Les plus compactes ont bcp moins d'interpénétration avec les autres → il faut moins d'énergie pour les dissocier de ses molécules voisines.

Energie de rupture de liaison : Dc-c = 350kJ.mol-1 et DC-H : 400kJ.mol-1

Très grande stabilité, donc très peu réactifs.

En modélisation moléculaire, on peut pas prendre en compte la dissociation d'une liaison C-C, car les energie ne sont pas de l'odrdre de grandeur qui nous intéresse. On peut en revanche mesurer la déformation de ces géométries.

C'est un biais car si on chauffe ces sytemes pdt qu'on fait de la dynamique moléculaire, l'énergie fournit devrait dissocier les liaisons, mais en fait on peut pas considérer ça.

Comme c'est stable, quand on les dissocie, on fournit de l'énergie → les alcanes servent principalement de carburant.

Analyse conformationnelle des alcanes non-cycliques : hybridation sp3, orientation spatiale type tétraèdre, angle de valence 109,47°. Quand on connecte les C les un aux autres, plusieurs configuration différentes sont disponibles. Une fois qu'on en impose une pour la modélisation, on ne peut plus revenir en arrière.

Quand on chauffe un système à haute energie, la molécule pourrait changer de configuration → il faut mettre un critère pour dire qu'on impose à la molécule de ne pas changer de conformation.

Analyse conformationnelle :

presque-liberté rotationnelle autour de C–C. Presque pq ça coute qd meme un peu d'energie, à T° ambiante, il suffit de 3 ou 4 kCal.mol-1 (le systeme peut très bien l'absorber du milieu extérieur).

D'un point de vue énergétique, la position éclipsée est privilégiée par rapport à la positon décalée (explications géométriques + quantiques).

Quand on fait de la minimisation d'energie, suivant le point de départ, on peut obtenir différent minima → déjà une ambiguité dès ce stade.

Les différentes structures de cyclo-alcanes :

triangle : angle de valence 60° : très grande tension → on les trouve presque pas.

Avec 4 atomes : 90° → idem.

Avec 6 atomes : 120° → c'est pas acceptable en sp3. → déformation en sortant du plan pour tendre vers les 109,47°.

Cyclopentane : quasi pas de déformation : assez proche (naturellement) de 109,47°. Cependant, ce sont pas les + fréquents. Ce sont les cycloalcanes à 6 et à 7 C qui sont les + fréquents.

Conformation chaise → **très** stable. Peut s'agencer en réseau.

Conformation bateau : mois stable, si on remplace les H par des C, on ne peut pas faire de réseau...

Strucures polycycliques : la bases d'architectures qui possèdent plusieurs fonctions organiques. Dans le détail, on voit que certains C sont asymétriques. Vérifier que les position R et S des C modélisés correspondent bien à la réalité (si c'est pas le cas, c'est qu'on est pas en train de voir la molécule naturelle).

Hétérocycle : remplacement d'un (ou plusieurs) C par un hétéroatome (O, N ou S). Ces systemes gardent l'hybridation sp3. N : 3liaisons, O : deux liaisons.

C-C : 1,54 C-O : 1,45 → énergies de liaisons un peu différentes mais on reste globalement dans une conformation type chaise.

Substitution C par O → pyranose : très stable.

Eléctronégativité : Plus un atome sera électronégatif, plus il pourra attirer des électrons.

Alkyles

Hormis la glycine, tous les aa sont naturellement en conformation S pour le C central.

En prenant en compte uniquement des groupements alkyles dans la chaîne latérale des AA.

Alanine : différence d'encombrement par rapport à Gly. Mais ça reste quasi la mm chose. Si on veut faire abstraction de la taille des chaines latérale, on peut muter ts les aa en alanines, ça nous donne le volume disponible pour les chaines latérales.

Si l'aa en question est trop volumineux, on le remplace par qqchose de plus petit → glycine ou alanine : on gagne de la place → Voir ensuite si ça modifie les pptés de la protéine.

Valine : encombrement + important. Asymétrie de volume (mais sans carbone asymétrique). Leucine et Isoleucine aussi. En les tilisant, on a la volonté de boucher un volume quelconque. Si on veut montrer qu'un canal d'eau est fonctionnel : on remplace un truc par isoleucine → le canal est bouché → normalement plus (-) efficace → conséquence.

On a une prot qui fonctionne : on veut couper le canal pour empecher la circulation d'eau. En général ça marche pas. Il y a des chances pour que la nature ait testé cette alternative, et donc ait changé sa conformation pour qu'une seule mutation comme ça ne mette pas HS la protéine. Pour jouer à ce jeu avec la nature, il faut tester des combinaisons de mutations.

Déporter le groupement qui nous intéresse plus ou moins loin de la chaine principale : en mutant différent aa (voir diapo 29 poly2, en mettant que la fonction d'intérêt est la double liaison (C=O ou C=NH)).

Alcools

qIl y a pas libre rotation autour de la liason C-O, mais il y a 3 configurations privilégiées.

27-09

-calcul énergétique pas très bien fait sur pyMol, mais il existe une option qui permet de faire comme si on faisait de la minimisation.

-Une énergie de liaison : ~ 400kJ (ordre de grandeur qu'on peut pas prendre en compte dans nos calculs)

-dès qu'on a 2 carbones en hybridation sp3 (ou deux autres atomes) → une courbe de variation d'énergie en fonctions de l'angle dièdre → c'est ça qui explique qu'il y ait des changmts de conformation.

→ la liaison H est difficulté supp : il y aura une multitude de conformations possibles autour de cette liaison. C'st aussi un facteur de stabilité (les molec. Vont s'attirer entre elles, dans des facteurs bien spécifiques).

- On est pas trop d'accord pour dire où commence et où finit une liaison H. Mais il y a une E optimale.

- Env 2,8 à 3Anstreum : distance ~optimale

-liaison non-covalente → peut se dissocier, se reformer … → fluctuation.

-bifurcation : un H hésite entre une liaison et une autre → il n'arrete pas de passer de l'une à l'autre. Comme nous on fait une photo → il est lié aux 2.

-Dans les champs de forces, certains prennent en compte spécifiquement la liaison H.

-suivant où se font les liaisons H, des réseaux différents de liaison H sont créés. Des réseaux sont présents dans les feuillets beta et les hélices alpha p.ex.

-en cristallo, on ne voit pas les liaisons H.

Un liaison H coûte env. 4 à 5kCal.mol-1 .

Quand on fait (4 à 5Kcal) \* (nb de liaisons H dans le réseau) : ça fait bcp.

Dans l'expression mathémat de liaison H, il faut tenir compte

Sérine ou thréonine : on peut s'en servir pour créer des liaisons H.

Dénaturer une prot : remplacer des aa par des groupements thiols → formation de ponts disulfures un peu partout. Stabilité S-S bien meilleure que liaison O-O (énergies de formation : 65 kcal.mol-1 et 35 kcal.mol-1 respectivement).

La cystéine contient une fonction thiol (le seul aa) → si on le met en bout de chaine, il peut créer un pont S-S et ainsi faire une boucle dans la protéine (dans 95 % des cas).

Dérivé d'un grpmt thiol ds la méthionine : un grpmt méthyl après le S empèche la formation S-S → ça a pas vraiment de fction pour la modélisation. C'est juste gros et assez linéaire.

Les Amines

Dérivés de l'amoniac : NH3. Primaire, Secondaire, Tertiaire → c'est lié au nb de substitutions de l'Azote et non pas du Carbone.

N → susceptible de faire des liaisons H.

Histidine : Possibilité de faire une liaison H, mais surtout la cyclisation la met surtt dans le groupe des aa pouvant faire de la délocalisation.

Lysine : fction amine Iaire en bout de chaine linéaire → svt prise pour des mutations à cause de sa taile et de sa forme allongée.

Qd on fait de la substitution : on a la possibilité de faire plusieurs conformations, pas uniquement la + probable.

→ quand on fait une mutation, il faut aussi en mesurer les conséquences ds l'environnement. Attention à ne pas oublier que ce qu'on voit à l'écran n'est pas ce qu'il se passe vraiement.

Svt, ces lysines sont mises à l'extérieur de la protéine. Au cœur de la prot, ça peut completement déstructurer la protéine.

Arginine : un peu comme la lysine. Très flexible. Énorme potentiel de liaisons H.

Les Alcènes

double liaison : hybridation type sp2.

Liaison π lieu de réactivité: forte densité électronique et électrons facilement accessibles car la molécule est plane (angle de valence théorique de 120 ) : ceux qui nous intéressent.

Carboxyle

La liaison C=O nous intéresse plus spécifiquement.

Qd on veut faire de la mutation in silico pour faire du stacking et qu'on a de la place → tryptophane ! Si moins de place : histidine.

Bilan Stabilisation molécule :

liaisons covalentes, liaisons H, stacking, forces de Van der Walls.