

Dock@Grid : Échantillonnage
Conformationnel et Docking Moléculaire
Position et Modélisation du problème

Benjamin Parent

6 février 2007

Table des matières

1	Description d'une molécule	2
2	Position du problème	3
3	Un champ de forces	4
3.1	Principales contributions	4
3.2	Note sur le champ de forces	5
3.3	Paramétrage du champ de force	6
4	Le docking moléculaire	7

1 Description d'une molécule

Une *molécule* est une réunion d'*atomes* maintenus ensemble par des forces de différents types. La mise en commun de certaines *orbitales électroniques* crée en particulier ce qu'on appelle des *liaisons de covalence* qui permettent d'assimiler la molécule à un **réseau** d'atomes. La position de chacun de ces atomes dans l'espace tridimensionnel génère autant de degrés de liberté et permet de modéliser la *géométrie* de la molécule, appelée *conformation*.

Comme de nombreux systèmes dynamiques, on sait que la molécule présente un ou plusieurs *états d'équilibre* dits *stables* autour desquels elle peut éventuellement orbiter. Cela se traduit par des vibrations de plus ou moins grandes amplitudes et plus ou moins rapides.

En particulier, on sait que la déformation des *longueur de liaisons* interatomiques sont très faibles et interviennent *a priori* peu dans le repliement moléculaire global. De même, les angles entre liaisons incidentes à un atome commun (dits *angles de valence*), bien que plus flexibles et intervenant probablement dans le processus de repliement, restent relativement rigides vis-à-vis des déformations globales de la molécule. La flexibilité repose donc, en dernière analyse, essentiellement sur la rotation des fragments moléculaires autour des liaisons de valence. Ces rotations sont repérées par leurs *angles de torsion*, et constituent les seuls degrés de liberté que nous considérerons dans la suite du projet (Cf. figure 1 et 2).

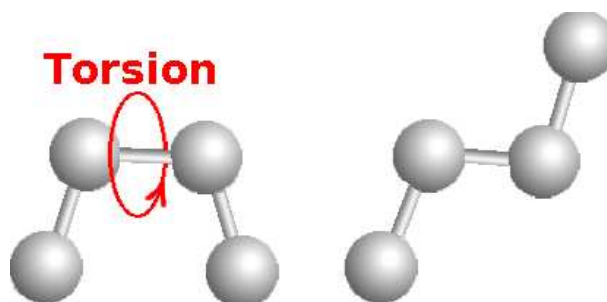


FIG. 1 – Les rotations autour des liaisons de covalence permettent à la molécule d'adopter différentes conformations.

Remarquons enfin que certaines molécules présentent des *cycles* dans leur graphe moléculaire... Pour permettre l'étude de telles structures dans le cadre général de la rotation des sous-fragments autour des liaisons covalentes,

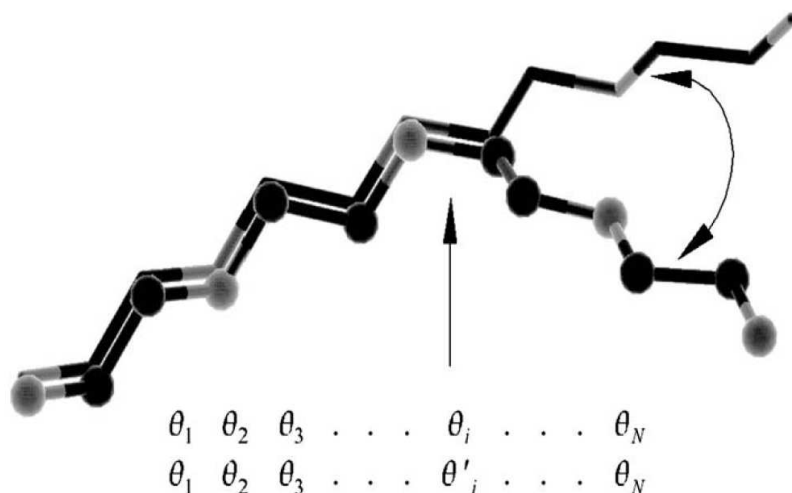


FIG. 2 – Codage des torsions de la molécule dans un vecteur

on « coupe » formellement chaque cycle d'intérêt pour explorer l'ensemble des géométries.

2 Position du problème

La vie d'une molécule se résume donc à l'exploration de son espace conformationnel ; exploration dirigée par le jeu des forces interatomiques, des chocs avec l'environnement, et de sa propre énergie cinétique ; on dit que la molécule *échantillonne* son espace de phase. Ce jeu se prolonge jusqu'à ce que la molécule subisse une modification dite covalente et change de structure primaire, la faisant alors sortir du cadre de la présente étude (réactions chimiques).

Nous étudions ici les modifications conformationnelles et plus généralement le repliement des molécules indépendamment de considérations particulières relatives à la structure de la molécule (protéine, molécule organique, molécule cyclique, etc.). Cette approche repose sur l'hypothèse de l'existence d'un *champ de forces* « universel », c'est-à-dire dont le domaine de validité couvre une gamme relativement large de molécules. L'ensemble de ces types de problèmes est regroupé sous le terme général d'*échantillonnage conformationnel*.

De plus, contrairement à certaines approches, nous ne cherchons pas à décrire la dynamique de la molécule qui dépend de conditions initiales, des chocs stochastiques avec les molécules environnantes et donc de la température. Nous nous intéressons au cas statique, dans lequel on perd l'information des transitions entre états métastables (franchissement de barrières de potentiel), mais avec la possibilité, *a posteriori*, de décrire le comportement de la molécule de manière statistique autour de ses points d'équilibre.

3 Un champ de forces

L'existence de points d'équilibre (et de spectres vibrationnels) émanent de la présence de *forces*, au sein de la molécule, et entre la molécule et son environnement. Il est possible d'approximer chacune de ces forces (c'est du moins l'hypothèse fondamentale des approches par champs de forces) en considérant les différents phénomènes électromagnétiques qui peuvent se jouer sur la scène moléculaire (il s'agit ici du champ de force CVFF, voir Hagler et al. (1974) et Hagler et Lifson (1974))...

3.1 Principales contributions

Par exemple, l'attraction de deux ions de charges opposées ou la répulsion de deux ions de même charge occasionne un apport énergétique répondant à la *loi de Coulomb*¹ (équation 1).

$$E_C = \frac{Q_i Q_j}{d_{ij}} \quad (1)$$

où $Q_{i,j}$ représente la charge des atomes et d_{ij} la distance entre eux. De même, il existe des interactions ioniques de type dipôles-dipôles (permanents ou induits) qui donnent lieu à des contributions de type Van der Waals² et dans lesquelles on inclut généralement également un terme de répulsion entre atomes trop proches (équation 2).

$$E_{\text{vdw}} = \frac{A_{i,j}}{d_{ij}^{12}} - \frac{B_{i,j}}{d_{ij}^6} \quad (2)$$

¹Charles Augustin Coulomb (1736 1806) : officier, ingénieur et physicien français.

²Johannes Diderik Van der Waals (1837 1923) : physicien et prix Nobel néerlandais.

Enfin, tout repliement moléculaire est synonyme de compactage et donc d'exclusion du solvant. Dans un solvant polaire comme l'est l'eau, isoler les parties hydrophobes est bénéfique et doit donc être pris en compte par un terme particulier. Par contre, la présence de solvant entre deux atomes peut modifier leur interaction. . . Un modèle de solvant continu basé sur des travaux de recherche antérieurs (Horvath (1997)) donne un comportement du type de l'équation (3)

$$E_{\text{desolv}} = \frac{K_{i,j}}{d_{ij}^4} \quad (3)$$

Enfin, dans le cas de l'échantillonnage des cycles, la solution retenue de couper formellement des liaisons impose l'intégration d'une pénalité énergétique de type harmonique (par approximation au deuxième ordre) lorsque la distance entre les atomes impliqués n'est pas la distance nominale. De même, les angles de valence dans ces liaisons coupées doivent être calculés afin d'ajouter une pénalité pour les conformations ne respectant pas la géométrie des orbitales électroniques (pénalité de type harmonique également).

Le dernier terme est une contribution directement issue des angles de torsion et n'a pas de justification théorique évidente. La validation et le calibrage d'un champ de forces tel qu'il est décrit ici est fait par confrontation avec des données expérimentales ; cette confrontation a donné naissance à ce dernier terme torsionnel et qui favorise les géométries observées expérimentalement pour chaque type de torsion (quadruplet d'atomes liés).

3.2 Note sur le champ de forces

Malheureusement, les dimensions de la molécule font d'elle un objet quantique tombant sous le diktat de l'équation de Schrödinger³ ; en particulier, les atomes ne sont pas des corps homogènes telles des billes dont la masse et la charge sont concentrées en son barycentre, mais bien des nuages électroniques délocalisés autour d'un noyau, et éventuellement perturbés par leur environnement. Ceci nous oblige donc à modifier les paramètres théoriques afin de tenir compte de phénomènes qui échappent à la description newtonnienne de la molécule. D'autres constantes dans le champ de force ne sont pas

³Erwin Rudolf Josef Alexander Schrödinger (1887 1961) : physicien autrichien, prix Nobel de physique, célèbre pour son équation d'évolution des systèmes quantiques faisant intervenir un cadre probabiliste.

non-plus déductibles immédiatement des équations de la physique, d'autres encore découlent d'approximations faites pour préserver le temps de calcul.

Ce champ de forces ainsi défini, façonne alors un paysage d'énergie potentielle (ou hypersurface) dans lequel la molécule équilibre son bilan énergétique entre énergie potentielle, énergie cinétique et son environnement (interactions électromagnétiques et agitation thermique). Dans le cadre d'une étude statique, nous nous restreignons donc au calcul de l'énergie potentielle ; les solutions les plus stables étant alors données par les niveaux d'énergies les plus bas. C'est ce que résume l'équation de Boltzmann⁴, issue de la physique statistique, et qui définit en terme de densité de probabilité, les niveaux de peuplement sur l'espace de phases (équation 4) :

$$\text{Pr}(\text{État d'énergie } E) \propto \exp\left(-\frac{E}{k_B T}\right) \quad (4)$$

où T est la température en Kelvin⁵ et k_B est la constante de Boltzmann.

Le champ de force est l'outil par lequel les problèmes d'échantillonnages conformationnels peuvent être interprétés en terme d'optimisation de fonction coût et donne accès à toute la gamme des algorithmes de recherche dédiés aux problèmes de grandes dimensions. . .

L'équation de Boltzmann est cependant particulièrement importante car elle modifie considérablement le problème : toute tentative de description de la conformation d'une molécule doit nécessairement prendre en compte l'aspect multimodal de son paysage d'énergie potentielle. En effet, les conformations d'énergies similaires donnent lieu à des niveaux de peuplement similaires, et tous les minima relevant de l'espace de phases sont *a priori* nécessaires pour pouvoir déduire des propriétés statistiques de la molécule.

3.3 Paramétrage du champ de force

Comme nous l'avons vu, de nombreux paramètres du champ de force sont estimés expérimentalement, sur la base de confrontations avec un ensemble de données nécessairement restreint. En particulier, l'étalonnage de certaines constantes repose sur un ensemble de *petites* molécules, qui permettent de

⁴Ludwig Eduard Boltzmann (1844 1906) : physicien autrichien, père de la physique statistique et un des premier à croire à l'hypothèse de l'atome.

⁵Sir William Thomson, alias Lord Kelvin (1824 1907) : physicien britannique qui a postulé l'existence d'un zéro absolu dans l'échelle des températures.

limiter les effets de bords et le nombre d'interactions entre atomes. Le champ de forces CVFF, bien que plus détaillé que certains champs de forces spécifiquement dédiés aux protéines (CHARMM⁶, AMBER⁷), possède donc un champ d'application limité aux petites molécules organiques et aux mini-protéines, ce qui est particulièrement adapté à l'objectif final de docking dans un site flexible.

4 Le docking moléculaire

Dans le cadre du docking moléculaire, nous nous intéressons, plus qu'à l'évolution d'une seule molécule, à l'évolution de deux molécules entrant en interaction. Plus particulièrement, nous étudions la fixation et l'affinité d'une petite molécule généralement organique (le *ligand*) dans un *site actif* d'une protéine ou d'un complexe. Cette description se veut aussi proche que possible de la réalité, c'est à dire prenant en compte la flexibilité du ligand et de la protéine réceptrice.

Notre approche du docking repose sur l'hypothèse suivante : « la description des phénomènes électromagnétiques et/ou quantiques doit pouvoir se formuler de la même manière, que l'on traite une molécule ou plusieurs ». Il s'ensuit que le docking ne doit être perçu que comme une généralisation de l'échantillonnage conformationnel où sont considérés

- les degrés de liberté du ligand,
- les degrés de liberté du site actif,
- les degrés de liberté relatif au positionnement spatial du ligand dans le site.

Pour atteindre ce but, il est prévu d'échantillonner en premier lieu le ligand et le site actif de manière séparée, puis de proposer une estimation rapide des meilleures paires candidates site-ligand de manière rigide avant de passer à l'étape finale de docking pleinement flexible.

Références

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., , et Karplus, M. (1983). Charmm : a program for macromole-

⁶Brooks et al. (1983)

⁷Cornell et al. (1995)

- cular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2) :187–217.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., et Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of American Chemical Society*, 117 :5179–5197.
- Hagler, A. T., Huler, E., et Lifson, S. (1974). Energy functions for peptides and proteins. i. derivation of a consistent force field including the hydrogen bond from amide crystals. *Journal of American Chemical Society*, 96(17) :5319–5327.
- Hagler, A. T. et Lifson, S. (1974). Energy functions for peptides and proteins. ii. the amide hydrogen bond and calculation of amide crystal properties. *Journal of American Chemical Society*, 96(17) :5327–5335.
- Horvath, D. (1997). A virtual screening approach applied to the search for trypanothione reductase inhibitors. *Journal of Medicinal Chemistry*, 40(15) :2412–2423.