

Dynamique moléculaire: exercices

Marc Baaden

BPTI exo

Copyright © 2003 Marc Baaden

Introduction

La protéine BPTI (*bovine pancreatic trypsin inhibitor*) était la première biomolécule simulée par dynamique moléculaire en 1977 (1). Depuis elle a été abondamment étudiée par de nombreuses approches théoriques et l'est encore jusqu'à ce jour.

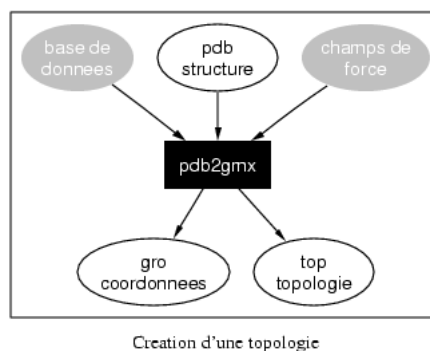
Au cours de cet exercice nous allons utiliser le logiciel Gromacs [<http://www.gromacs.org>] pour mettre en œuvre une simulation par dynamique moléculaire *in vacuo* et analyser une trajectoire dans l'eau.

Dynamique moléculaire: préparation et mise en œuvre

Simulation de BPTI *in vacuo*

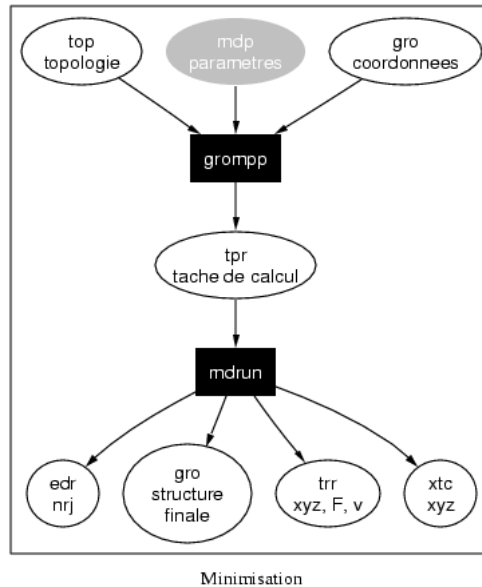
Nous allons reproduire une simulation de BPTI à partir de coordonnées déterminées par RMN. Les coordonnées ont été déposées sous le code 1LD5 à la *Protein Data Bank* [<http://www.rcsb.org/pdb/>], et nous avons déjà extrait une des 10 structures déterminées dans le fichier `bpti.pdb`. En plus de la structure, nous avons besoin de la topologie de la protéine. Celle-ci peut-être générée par l'outil `pdb2gmx` de Gromacs. Nous partons de la structure, choisissons un champs de force et `pdb2gmx` interroge sa base de données d'acides aminés pour construire la topologie correspondante faisant référence aux types d'atomes et paramètres du champs de force choisi. Utilisez le raccourci `make prep` pour appeler `pdb2gmx` avec les bons paramètres, et essayez de comprendre les sorties qui s'afficheront à l'écran.

Figure 1. Schéma de la génération de topologie.



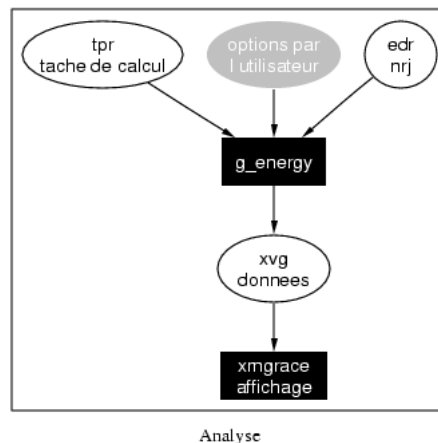
Nous allons maintenant effectuer une minimisation de la protéine dans le vide, ce qui permet de réduire d'éventuelles contraintes au sein de la molécule. Le résultat de cette minimisation dépend également fortement du champs de force utilisé. Utilisez le raccourci `make em` et observez.

Figure 2. Schéma de la génération de topologie.



Vérifiez si l'énergie de la molécule a bien été abaissée avec la commande **make emnrj**.

Figure 3. Schéma de l'analyse énergétique.



Passons à une simulation de dynamique moléculaire. Nous allons simuler la BPTI pendant 10 ps dans le vide. Pour préparer les fichiers de paramètres et lancer la dynamique, utilisez **make md**.

Vérifions ce qui s'est passé au cours de la dynamique. Le raccourci **make mdcheck** affichera température et énergie au cours de la simulation. Qu'en pensez vous ?

Etapes supplémentaires pour préparer une dynamique en solution

Nous allons simplement discuter les étapes supplémentaires qui s'imposent pour préparer un système en solution. Ces préparations ont déjà été effectuées pour vous, et nous partirons de la dynamique déjà calculée dans la prochaine section.

Pour effectuer une simulation dans un solvant, il faut dans un premier temps définir une boîte de solvant qui va englober la protéine BPTI et qui sera suffisamment grande pour éviter des artéfacts de périodicité. Cette boîte est créée avec la commande **editconf** et ensuite remplie d'eau en utilisant une

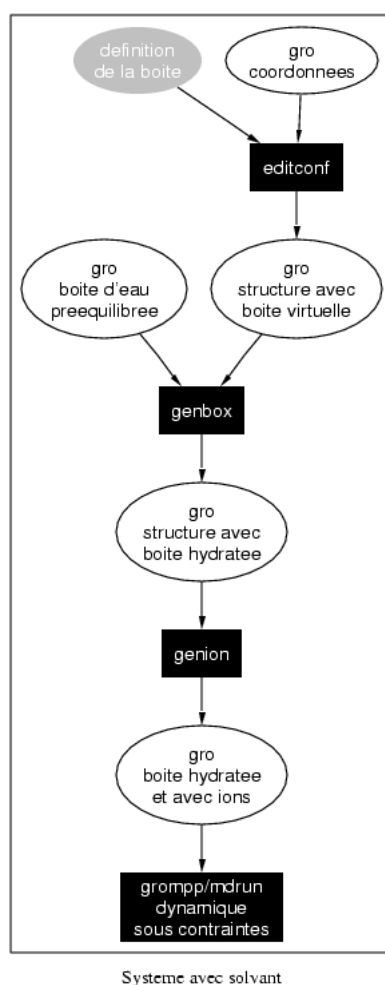
boîte de solvant prééquilibrée avec la command **genbox**. La BPTI est également chargée positivement comme vous avez du le remarquer à l'issu de la commande **grompp** qui avait indiquée:

```
[...]  
System has non-zero total charge: 6.0  
[...]
```

Il faudra donc ajouter au moins 6 contreions, par exemple des chlorures, avec la commande **genion**.

Le système complet ayant été assemblé, il est conseillé d'effectuer une équilibration du solvant et des ions en fixant la protéine et laissant le solvant libre. On relaxe ainsi d'éventuelles gênes stériques et permet à la pression de s'ajuster. Ensuite la dynamique dite «de production» est lancée.

Figure 4. Schéma de la préparation d'une dynamique en solution.



Analyse d'une trajectoire dans l'eau

Une première étape dans l'analyse d'une trajectoire est souvent la visualisation. Nous utiliserons le logiciel VMD [<http://www.ks.uiuc.edu/Research/vmd/>] pour examiner la trajectoire de BPTI dans l'eau. Utilisez le raccourci **make vis** pour lancer la visualisation. Que signifient les sauts des ions chlorures ?

Pour analyser des changements structuraux de la protéine BPTI nous regarderons d'abord son rayon de gyration en fonction du temps. Le rayon de gyration est obtenu en moyennant les distances de chaque

atome au centre de masse. Chaque contribution à la moyenne est pondérée en fonction de la masse de l'atome considéré.

$$r_g = \sqrt{\frac{\sum r^2 \cdot m}{\sum m}}$$

Pour lancer le calcul du rayon de gyration et visualiser le résultat de cette analyse, vous pouvez utiliser le raccourci **make gyr** qui va à son tour appeler les commandes **g_gyrate** et **xmgrace** pour l'analyse et la visualisation respectivement:

```
[/tmp] make gyr echo -e "1\n" \ | g_gyrate -f full.trr -s  
full.tpr -o gyrate.xvg :- ) G R O M A C S (-: [..] xg.pl  
gyrate.xvg [..]
```

Analysez le graphe qui apparaît à l'écran. Comment évolue le rayon de gyration au cours du temps ? Y-a-t-il des changements abrupts ?

La BPTI présente une section efficace de collision de $548 \pm 30 \text{ \AA}^2$ dans le vide (2). Le rayon de gyration permet également de calculer cette section à l'aide de la formule $5\pi r_g^2/3$. Calculez la section efficace pour la simulation dans l'eau à partir du rayon de gyration moyen et comparez à la valeur expérimentale *in vacuo*. Pour obtenir la valeur moyenne d'un résultat d'une analyse, utilisez la commande **g_analyze**. Dans le cas présent **g_analyze -f gyrate.xvg** donne les valeurs moyennes, l'écart type et d'autres valeurs statistiques. Quel est l'effet de l'eau sur la section efficace ?

Une deuxième analyse de l'évolution de la structure globale d'une protéine à laquelle on fait souvent appel est l'écart quadratique moyen entre la structure au cours de la simulation et la structure cristallographique ou de départ (RMSD du terme anglais *root mean square deviation*). Le RMSD prend en compte la distance scalaire entre atomes du même type comparant deux structures, après que ces structures aient été superposées, le plus souvent en minimisant la valeur du RMSD. Souvent on se limite aux RMSD des carbones α pour comparer les protéines. Nous allons faire de même. Dans Gromacs la valeur du RMSD entre deux structures 1 et 2 est calculée par:

$$\text{RMSD} = \sqrt{\frac{1}{M} \sum_i m_i \|r_{i,1} - r_{i,2}\|^2}$$

Utilisez le raccourci **make rms** pour calculer le RMSD de toute la protéine et des carbones α par rapport à la structure de départ en fonction du temps de simulation. Comparez les valeurs obtenues après 10 ps avec celles données dans l'article de McCammon *et al.* Que se passe-t-il par la suite ?

Des mesures de distance permettent également d'aborder le sujet de la structure secondaire et tertiaire d'une protéine. En établissant une matrice des distances minimales entre chaque résidu on distingue des motifs caractéristiques. Utilisez **make dmat** pour lancer le calcul et la visualisation d'une matrice correspondant à la structure moyenne pendant la simulation. Quels éléments de structure pouvez vous distinguer ?

La structure secondaire peut être directement analysée en fonction du temps en se servant du raccourci **make dssp** qui fait appel à la librairie DSSP (*dictionary for secondary structure of proteins*, (3)) pour déterminer la structure secondaire à partir de paramètres structuraux. Décrivez les changements observés.

La mobilité des résidus au sein de leur structure secondaire en fonction du temps peut être étudiée par le graphe de Ramachandran (4). Son évolution en fonction du temps permet entre autre d'identifier les résidus mobiles au sein de la structure secondaire. Nous allons choisir deux exemples, les résidus ARG20 et CYS38, et afficher leurs graphes de Ramachandran des angles ϕ/ψ pendant la simulation. Utilisez la commande **make rama**. Qu'est-ce qu'on peut dire sur ces deux exemples ?

Maintenant nous allons regarder l'ensemble du graphe de Ramachandran au cours de la simulation sous forme d'animation. La commande est **make rama2**.

Les fonctions de corrélation

L'utilisation des fonctions de corrélation permet une analyse plus poussée du «désordre» introduit dans le système au cours du temps. On peut illustrer ces fonctions comme fonctions de mémoire qui montrent le temps (Δt) nécessaire pour qu'une variable donnée (par exemple un angle) perde sa relation avec une autre variable (par exemple le même angle à un temps ultérieur). Un type de fonction de corrélation utilisé fréquemment est le produit scalaire entre deux vecteurs à deux instants distincts:

$$C(\tau) = \langle u(t) \cdot v(t + \tau) \rangle_t = \langle \cos\theta_{uv} \rangle_t$$

L'indice t représente la moyenne dans le temps (ou moyenne sur l'ensemble). Une telle fonction de corrélation qui établit le lien entre un objet et ce même objet à un temps ultérieur s'appelle fonction d'autocorrélation. Dans ce cas les vecteurs u et v sont les mêmes. Par exemple la fonction d'autocorrélation pour un angle dièdre correspondra au changement du cosinus de cet angle dièdre:

$$C_{\text{dièdre}}(\tau) = \langle \cos(\theta(t) - \theta(t + \tau)) \rangle_t$$

Nous allons maintenant calculer la fonction d'autocorrélation de quelques vecteurs N-H du squelette de la protéine et du pont disulfure C14-C38. Utilisez la commande **make corr**. Que constatez vous ?

Regardons maintenant un angle dièdre, celui autour de CYS38, de plus près. La commande **make corr2** calculera et visualisera l'angle moyen, la fonction d'autocorrélation et la distribution de l'angle.

Bibliographie

1. McCammon J.A., Gelin B.R., et Karplus M. (1977) *Nature* **267**, 585-590
2. Fischer S. et Verma C.S. (1999) *Proc.Natl.Acad.Sci.U.S.A.* **96**, 9613-9615
3. Kabsch W. et Sander C. (1983) *Biopolymers* **22**, 2577-2637
4. Ramachandran G.N., Ramakrishnan C., et Sasisekharan V. (1963) *J. Mol. Biol.* **7**, 95-99