

# Etude de la Barstar

Delphine DEYRIS - Narjes NALOUTI

# PLAN

## Introduction

## I Description globale du programme

## II Analyse des résultats

### 1 . Analyse des changements conformationnels globaux

### 2 . Analyse des changements conformationnels locaux

## III Conclusion

## INTRODUCTION

Le *docking*, ou amarrage protéine-protéine, est une problématique de la bioinformatique structurale, qui, à partir des structures tridimensionnelles des protéines, ainsi que leurs propriétés physico-chimiques, consiste à prédire comment les protéines s'associent entre-elles pour former des complexes fonctionnels au niveau de la cellule. La connaissance de la structure des complexes protéine-protéine permet de comprendre les mécanismes clé mis en jeu lors d'une interaction entre protéines, fondamentale pour l'étude de la majorité des processus biologiques. De telles informations proviennent généralement de l'étude des interactions telles que des complexes antigène-anticorps, ou encore protéase-inhibiteur. Le cas du complexe barstar-barnase constitue aujourd'hui un modèle largement étudié.

La barnase est une endoribonucléase bactérienne, synthétisée et sécrétée par *Bacillus amyloliquefaciens*, qui catalyse le clivage de l'ARN. Son action est létale pour la cellule lorsque celle-ci est exprimée en l'absence de son inhibiteur, la barstar. Les deux protéines forment alors un complexe, faisant appel à une interaction protéine-protéine particulièrement forte, tant que la barnase se situe dans le cytosol. La barstar se positionne ainsi sur le site actif de la barnase, empêchant cette dernière de dégrader l'ARN intracellulaire, ce qui a pour conséquence de conserver l'intégrité de la cellule. La particularité de la barnase est que celle-ci ne nécessite pas de passage par une structure tertiaire, et ne fait pas appel à des groupes prosthétiques pour adopter la conformation de son état natif, c'est-à-dire sa conformation fonctionnelle.

Le principe d'une expérience de dynamique moléculaire consiste en la simulation du mouvement d'une molécule dans une « boîte » de solvant. En effet, étant donné que le milieu intracellulaire est essentiellement composé d'eau, il s'agit de modéliser la dynamique d'une molécule dans les conditions les plus proches de la réalité. L'avantage de faire appel à une expérience de dynamique moléculaire réside également dans la possibilité de générer une réelle trajectoire au cours du temps, et donc dans la possibilité d'enregistrer des changements conformationnels, ce qui est un inconvénient dans les algorithmes de docking, qui considèrent que les structures de deux protéines testées sont rigides, c'est-à-dire que les formes liées et non liées des protéines sont les mêmes. Cela n'est en aucun cas représentatif de la réalité, puisque dans la cellule, lors d'une interaction protéine-protéine, des changements conformationnels ont lieu. De plus, il existe des acteurs protéiques, les protéines chaperonnes, qui participent à aider la protéine à adopter la conformation souhaitée. L'objectif de cette étude est donc d'analyser les résultats d'une expérience de dynamique moléculaire de la protéine barstar, qui, avec la barnase, forme un complexe, caractérisé par une interaction protéine-protéine d'une intensité remarquable.

Dans un premier temps, les différentes fonctions du programme seront décrites, puis, une analyse ainsi qu'une interprétation des résultats seront faites. Une comparaison avec les résultats de l'article sera faite, afin de mieux mettre en lumière l'adéquation des résultats obtenus.

## I – Description globale du programme

Le programme se compose en deux fichiers. Le premier 'Fonctions.py' comporte un parser de fichier PDB, les fonctions de calcul des RMSD locaux et globaux, les calculs giratoires locaux et globaux ainsi que le calcul du centre de masse. Le deuxième fichier Main.py faisant appel au premier comporte les fonctions gérant les sorties .txt et l'affichage graphique. L'appel du programme dans le terminal se fera à l'aide de la commande suivante : `python Main.py fichier_entree.pdb`.

Les fichiers d'entrées sont « md\_prot\_only\_skip100.pdb » et « md\_prot\_only\_skip10.pdb ».

Ces deux fichiers sont le résultat de l'acquisition des trajectoires (c'est-à-dire la liste des coordonnées des différents atomes composant la protéine), au cours d'une expérience de dynamique moléculaire. L'expérience a été réalisée sur un temps de 20ns. Dans le premier fichier, les trajectoires ont été enregistrées toutes les 20ps, ce qui a généré un total de 2000 conformations, et dans le second fichier, les trajectoires ont été enregistrées toutes les 200ps, ce qui a généré un total de 200 conformations. La nécessité de réaliser ces différents enregistrements réside dans le besoin de collecter des données à des temps différents afin de mieux apprécier les changements conformationnels subis par les atomes, et donc par la protéine, au cours du temps.

La première fonction **ParserPDB** définie dans le fichier **Fonctions.py** est un parser de fichier PDB. Elle consiste à extraire des données d'un fichier .pdb et à renvoyer un dictionnaire comprenant le numéro du modèle, les atomes des résidus correspondants

ainsi que leurs coordonnées spatiales (x,y,z), du fichier pdb d'entrée. Ce dictionnaire va servir de référence dans toutes les fonctions de calcul.

Le RMSD (Root Mean Square Deviation) mesure la distance moyenne entre les atomes des différentes conformations que prend la protéine. Ainsi, il est possible d'étudier la similitude des conformations des protéines et donc la structure de la même protéine dans le temps. La formule du RMSD est la suivante :

$$\sqrt{\frac{1}{N} \sum_{i=1}^N |x_i - y_i|^2}$$

Où N correspond au nombre de paires d'atomes entre deux conformations de la protéine et  $x_i, y_i$  représente la distance entre la position de l'atome dans la structure de référence et la position de l'atome au temps t.

La fonction **RMSDGlobal** renvoie une liste **result\_glob** du calcul des RMSD de tous les atomes entre les différentes conformations (modèle 1 à n, selon le fichier d'entrée) par rapport à la conformation de référence (modèle 0), du fichier d'entrée.

La fonction **RMSDlocal** renvoie une liste **result\_loc** du calcul de RMSD des résidus entre les différentes conformations.

La fonction **gir\_global** prend en entrée la fonction **centredemasse** et renvoie une liste **lst\_memoire** des rayons giratoires maximaux. Cette fonction calcule la distance entre tous les atomes et le centre de masse de la protéine et sélectionne le rayon maximal de chaque modèle.

La fonction **distance** prend en entrée la fonction **centredemasse** et calcule la distance entre les résidus de chaque modèle et le barycentre de la protéine. Elle renvoie une liste **dist**.

Pour étudier les changements conformationnels de la protéine dans le temps, on calculera dans un premier temps le RMSD et le rayon giratoire de chaque atome sur 2000 modèles (md\_start\_prot\_only\_skip10.pdb). Dans un second temps, on s'intéressera aux changements locaux en calculant les distances et RMSD des résidus par modèle. Enfin, on comparera ces résultats au 200 modèles (md\_start\_prot\_only\_skip100.pdb).

## II Analyse des résultats

### 1. Analyse des changements conformationnels globaux

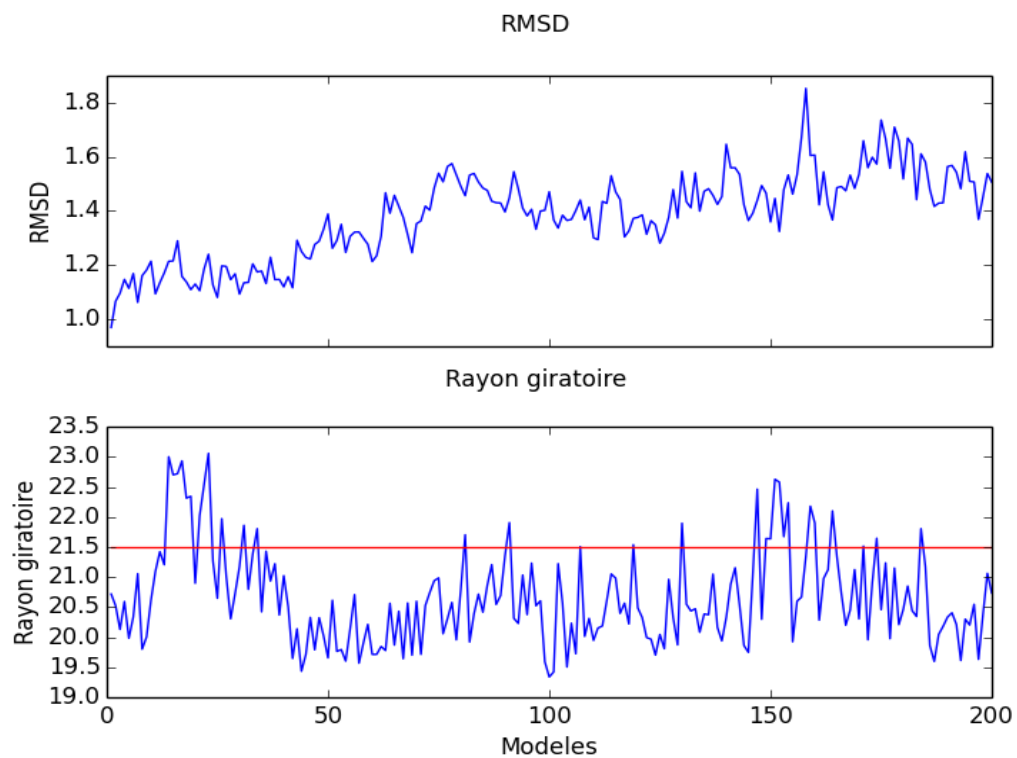
La dynamique globale de la molécule est ici évaluée grâce à deux paramètres :

- Le rayon de giration, qui correspond à la distance entre le centre de masse de la protéine, et l'atome le plus éloigné de celle-ci.
- Le RMSD, qui mesure la déviation entre la conformation de référence, celle correspondant au modèle 0, et la conformation enregistrée au temps t.

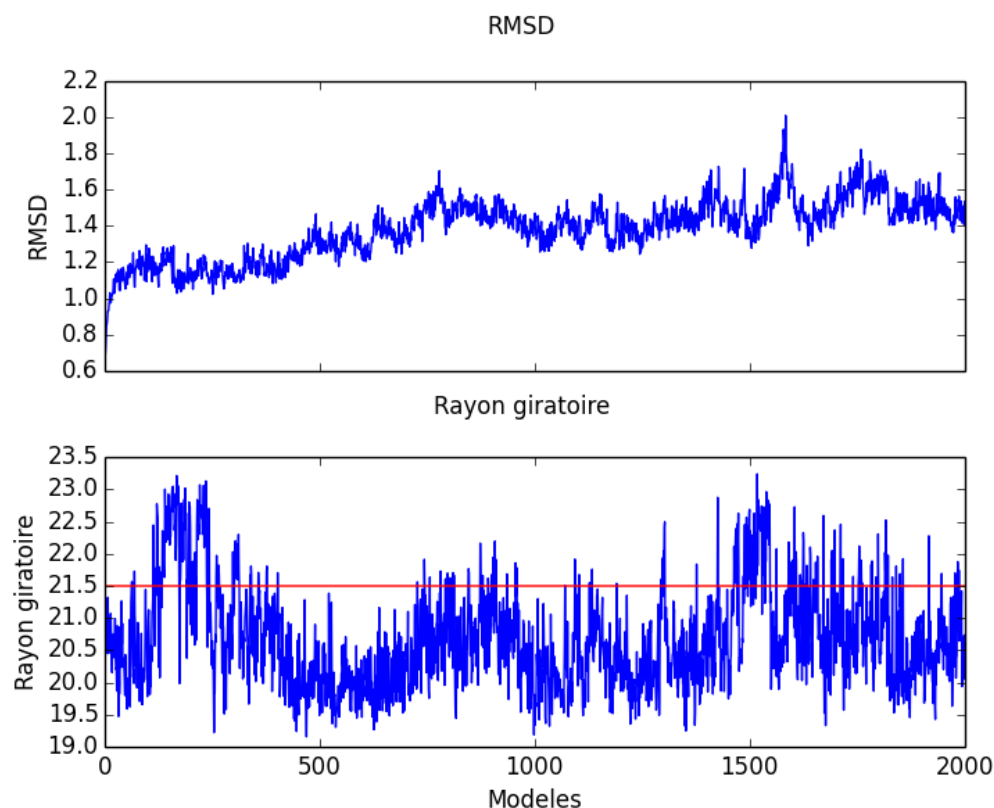
L'étude des RMSD et des rayons giratoires génère les graphes suivants :

- Figure 1.a : Evolution du RMSD en fonction du rayon giratoire (200 conformations)
- Figure 1.b : Evolution du rayon giratoire au cours du temps (200 conformations)

- Figure 2.a : Evolution du RMSD en fonction du rayon giratoire (2000 conformations)
- Figure 2.b : Evolution du rayon giratoire au cours du temps (2000 conformations)



**Figure 1.a) Evolution du RMSD en fonction de rayon de giration b) Evolution du rayon giratoire en fonction du temps (200 conformations)**



**Figure 2.a) Evolution du RMSD en fonction de rayon de giration b) Evolution du rayon giratoire en fonction du temps (2000 conformations)**

En examinant la courbe des RMSD globaux (Figure 1), on observe une différence entre la conformation de la structure de référence et les 2000 modèles dans le temps. Plus la valeur du RMSD est grande, plus elle signifie une différence de conformation. En effet, un RMSD égal à 0 signifie que les structures sont superposées, et qu'il n'y a pas eu de changement de trajectoire. Le premier constat est donc que la structure de la Bastar évolue dans le temps, sa conformation de référence n'est pas figée et constante dans le temps.

La courbe du rayon giratoire en fonction des modèles (Figure 2) quant à elle, révèle que la protéine subit des transformations spatiales. La ligne rouge représente le rayon de giration de la conformation de référence. La ligne rouge horizontale représente le rayon giratoire de la conformation de référence de la Barstar. La courbe bleue représente l'évolution au cours du temps du rayon de giration. Lorsque la courbe bleue se situe en dessous de la ligne rouge, c'est-à-dire que le rayon de giration est inférieur au rayon de giration de la structure de référence, cela indique que la Barstar est plus repliée que la conformation de référence. A l'inverse, lorsque la courbe bleue se situe au-dessus de la ligne rouge, la Bastar présente alors une conformation plus dépliée que la structure de référence. L'analyse de la figure 2 permet alors de déduire que la Barstar est majoritairement plus repliée que sa conformation initiale.

## 2. Analyse des changements conformationnels locaux

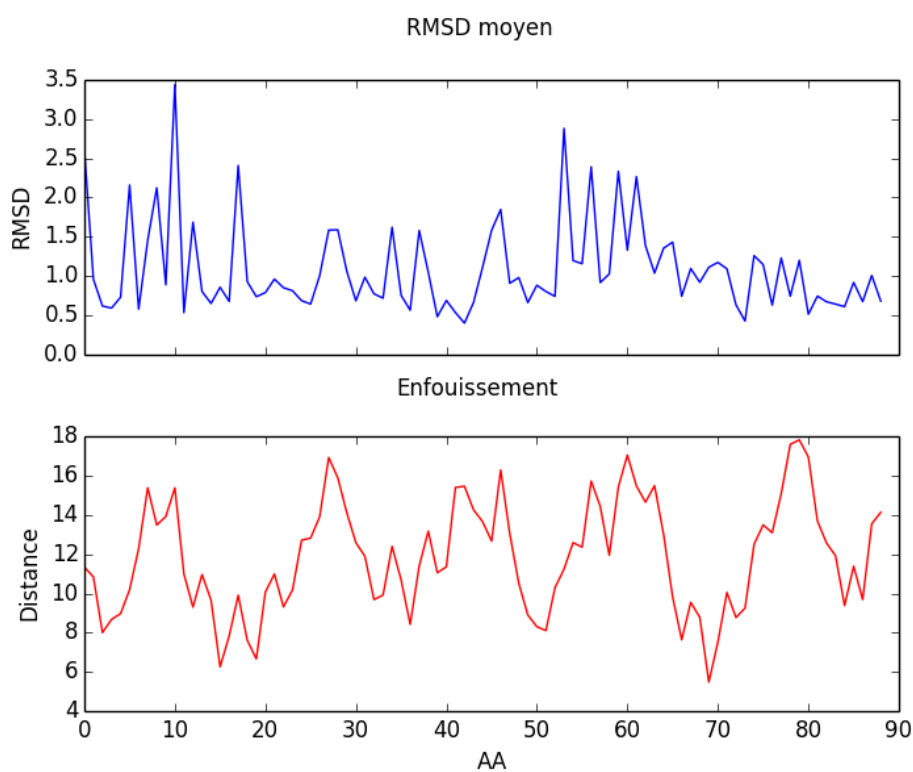
Dans un second temps, l'analyse de la dynamique se fait à l'échelle locale, c'est-à-dire au niveau des acides aminés de la protéine. Dans cette partie, nous calculons le RMSD moyen de tous les acides aminés de la protéine par



rapport à la conformation de référence. Ce calcul permettra alors de rendre compte des régions flexibles de la protéine, c'est-à-dire de mettre en évidence plus précisément les zones sujettes au mouvement. Pour chaque acide aminé, nous calculerons également la distance moyenne séparant chaque acide aminé du centre de masse au cours du temps, afin d'avoir une idée de l'enfouissement. L'objectif est de mettre en évidence s'il existe un lien entre l'enfouissement d'un résidu et la flexibilité. De manière intuitive, nous avons tendance à penser que les résidus les plus enfouis, situés au cœur de la protéine, seront situés dans des zones de plus faible flexibilité.

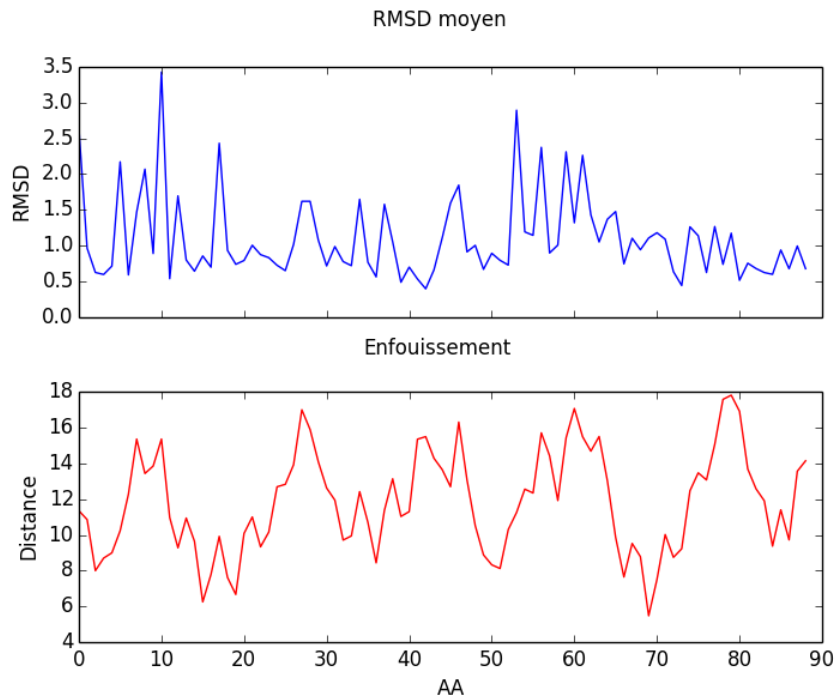
Le programme a généré les graphes suivants :

- Figure 3.a : RMSD en fonction de l'enfouissement (200 conformations)
- Figure 3.b : Distance moyenne séparant chaque acide aminé du centre de masse de la protéine (200 conformations)
- Figure 4.a : RMSD en fonction de l'enfouissement (2000 conformations)
- Figure 4.b : Distance moyenne séparant chaque acide aminé du centre de masse de la protéine (2000 conformations)



**Figure 3.a) RMSD en fonction de l'enfouissement b) Distance moyenne séparant chaque acide aminé du centre de masse de la protéine (200 conformations)**

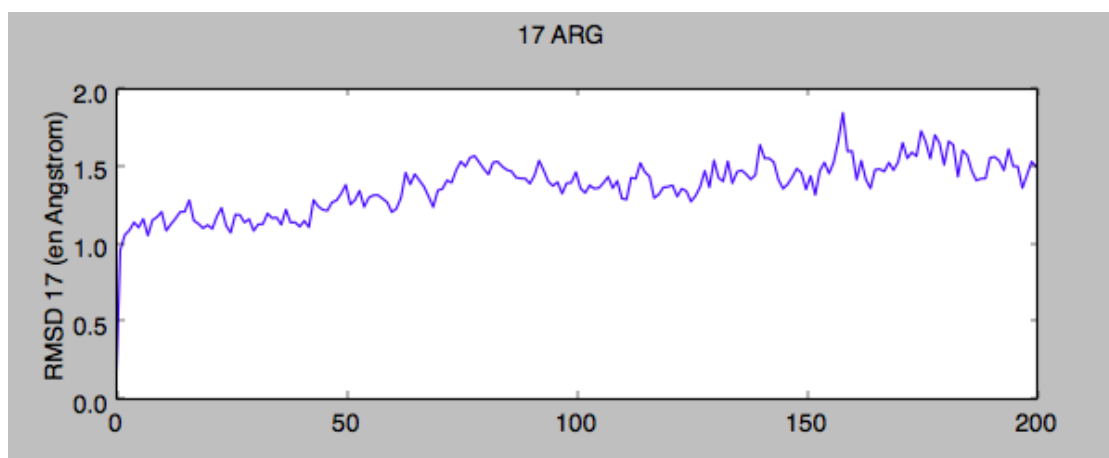


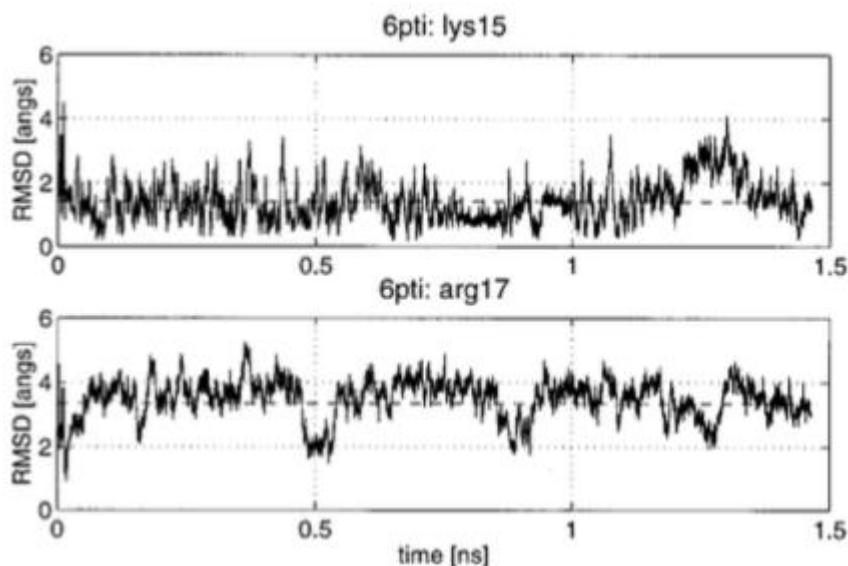


**Figure 4.a) RMSD en fonction de l'enfouissement b) Distance moyenne séparant chaque acide aminé du centre de masse de la protéine (2000 conformations)**

L'étude du RMSD moyen de chaque acide aminé permet de mettre en évidence les régions flexibles qui correspondent aux différents pics de la courbe : d'après la figure 3.b, elles se situent entre le 5ème et le 10ème acide aminé, entre le 25ème et le 30ème acide aminé, entre le 40ème et le 45ème, entre le 60ème et le 50ème, et enfin entre le 75 et le 80ème acide aminé.

Ces résultats nous invitent alors à nous demander si les acides aminés présentant un RMSD élevé sont toujours en surface, et si les acides aminés enfouis, c'est-à-dire présents au cœur de la protéine, peuvent présenter un RMSD élevé. Afin de répondre à ces questions, nous avons sélectionné l'acide aminé 17 ARG, afin de comparer les résultats obtenus dans l'article de Kimura et al, avec ceux que nous obtenons.





Les différents résultats nous permettent alors de répondre aux différentes questions posées, en validant ainsi notre intuition. En effet, plus le RMSD est élevé, plus l'acide aminé est flexible, en revanche, il faut que l'acide aminé se situe nécessairement à la surface de la protéine, car lorsqu'il est enfoui, les pressions d'ordre structurelles sont élevées, et cela limite l'amplitude du mouvement résiduel. En revanche, lorsque le RMSD d'un acide aminé est faible, qu'il soit situé en surface ou au cœur de la protéine, sa flexibilité sera moindre.

### III Conclusion

Afin d'effectuer les analyses précédentes, nous nous sommes basées sur les résultats obtenus avec le fichier contenant 200 conformations (figures 1.a. et 1.b, figure 3.a et 3.b.) . Nous allons à présent comparer si le nombre de conformation à exploiter est important dans l'analyse des résultats.

Nous constatons alors que les graphes obtenus dans la première partie d'analyse des changements conformationnels globaux, pour le calcul du RMSD et l'évolution du rayon giratoire, avec 200 conformations et 2000 conformations sont parfaitement superposables. Il en est de même pour la seconde partie qui concerne l'analyse des changements conformationnels locaux, les graphes obtenus sont superposables, les conclusions tirées à l'issue des graphes en fonction du nombre de fichier sont identiques.

Nous pouvons donc conclure que le fichier comportant 200 conformations est suffisant dans le cadre de cette étude, et il est inutile de générer trop de données. L'acquisition d'un grand nombre de données est importante lorsque les expériences menées présentent des coûts importants, ou qu'il est difficile de se procurer les échantillons en question.

Il aurait été néanmoins intéressant d'effectuer la partie de simulation en amont, c'est-à-dire l'expérience de dynamique moléculaire en utilisant l'outil GROMACS, ce qui nous aurait permis de mieux appréhender la problématique biologique.

