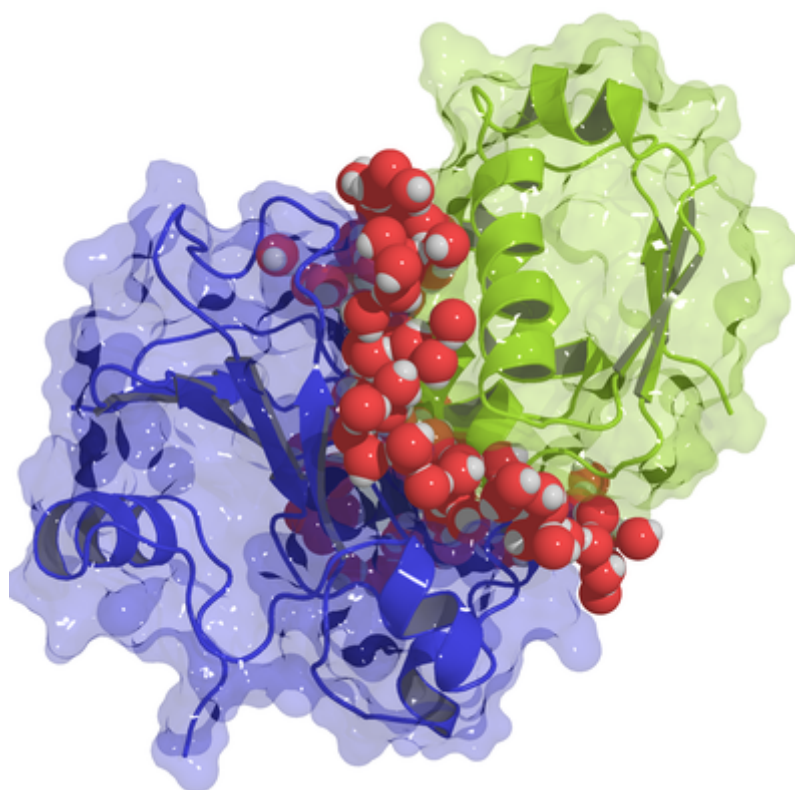


Étude de la BARSTAR



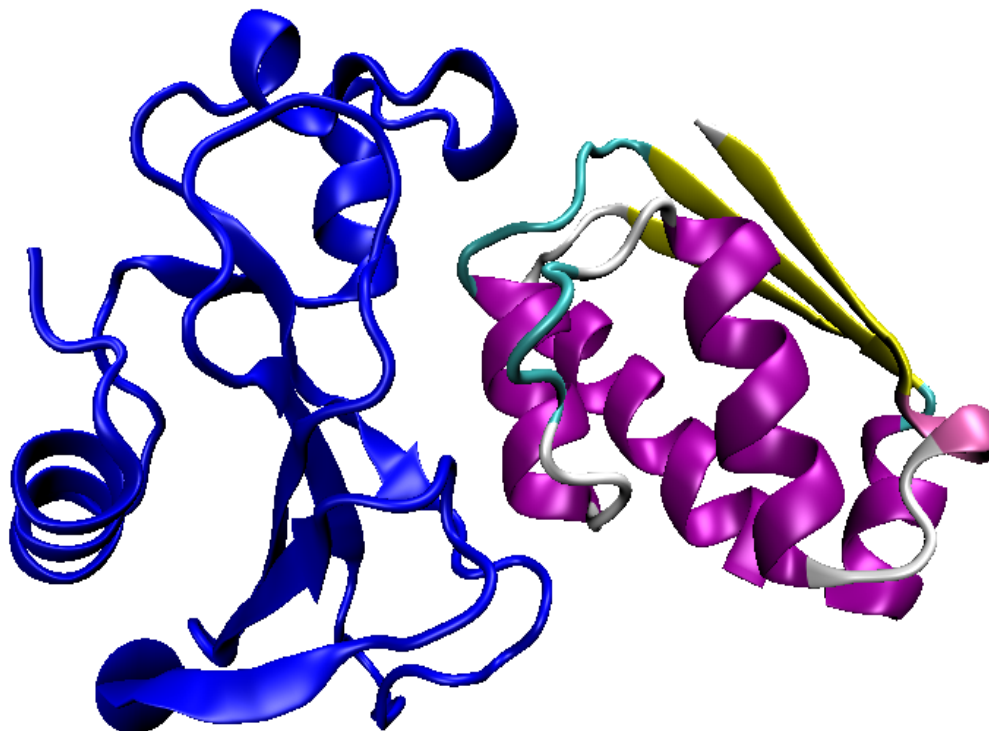
1. INTRODUCTION ET CONTEXTE	3
2. ORGANISATION DU PROJET	4
3. OBSERVABLES STRUCTURALES DES PROTEINES	4-5
3.1 RAYON DE GIRATION	4
3.2 DISTANCE BOUT-A-BOUT	4
3.3 RMSD	4-5
4. PROGRAMME	5-6
4.1 CONSTRUCTION DU PROGRAMME	5
4.2 FONCTIONS PRINCIPALES	5-6
5. RESULTATS	6-8
5.1 ANALYSE DES CHANGEMENTS CONFORMATIONNELS GLOBAUX	6
5.2 ANALYSE DES CHANGEMENTS CONFORMATIONNELS LOCAUX	5
6. CONCLUSION	8-9
6.1 COMPARAISON DES RESULTATS AVEC D'AUTRES DONNEES	8-9
6.2 COMPARAISON DES RESULTATS AVEC UN ARTICLE	9
7. DISCUSSION	10

1. Introduction et contexte

Dans le domaine de la modélisation moléculaire, une méthode appelée **amarrage** (en anglais, *docking*), permet de prédire l'orientation d'une molécule par rapport à une autre. Cette méthode calcule l'orientation préférée d'une molécule vers une seconde lors de la formation d'un complexe et permet ainsi de prédire la conformation la plus optimale (c'est-à-dire la plus stable) du complexe étudié. Cependant, cette technique a ses limites et un grand nombre d'algorithmes de docking ne prennent pas en compte un paramètre important pour ces prédictions, la mobilité. Pour s'emboîter parfaitement, les protéines doivent subir des changements conformationnels au moment du docking, appelés « *induced fit* », qui sont définis comme des changements conformationnels liés à l'interaction entre deux molécules.

En effet, les molécules n'étant pas des structures rigides, le fait de ne considérer qu'une seule forme tri-dimensionnelle nous permet de conclure uniquement sur cette forme spécifique et non sur l'ensemble des conformations possibles. Afin de corriger ce biais, l'algorithme de docking doit donc prendre en compte un maximum de formes différentes des mêmes molécules afin de représenter leur dynamique moléculaire. Ce projet a donc pour objectif de représenter la dynamique moléculaire de ces molécules.

Dans ce projet, nous nous intéressons à la BARNASE et la BARSTAR. Ces deux protéines sont issues de l'espèce *Bacillus amyloliquefaciens*, une bactérie à Gram positif. La BARNASE est une ribonucléase qui forme un complexe très étroit avec la BARSTAR, qui est elle-même un inhibiteur intracellulaire de la BARNASE.



2. Organisation du projet

Nous allons donc chercher à réaliser un programme de *docking* qui permet, à partir de fichiers .pdb, d'analyser les meilleurs assemblages entre ces deux protéines au cours du temps en considérant la mobilité entre résidus.

Pour ce faire, nous devons analyser un grand nombre de conformations à des intervalles de temps courtes. Ici, deux jeux de données seront utilisés. Un premier ensemble de 2000 conformations avec une intervalle de 10ps entre chacune d'entre-elles et un second ensemble de 200 conformations avec 100ps d'intervalle. Nous pourrions ainsi comparer nos résultats sur deux jeux de données différents.

Une fois l'ensemble des données analysées avec ces jeux de données et ces indicateurs, nous analyserons nos résultats à l'aide de ceux obtenus dans l'article « Dynamical View of the Positions of Key Side Chains in Protein-protein Recognition » S.roy Kimura et al.

3. Observables structurales des protéines

3.1 Rayon de giration

Le rayon de giration d'une protéine est défini comme la distance à l'axe de rotation à laquelle il faut placer un point de masse égale à celle du corps pour qu'il aie le même moment d'inertie que le corps. Mathématiquement, cela correspond à :

$$Rg = \sqrt{\frac{\sum_i m_i (r_i - r_{CM})^2}{\sum_i m_i}},$$

où r_{CM} est la position du centre de masse, r_i est la position de l'atome i et m_i sa masse.

3.2 Distance bout-à-bout

La distance bout-à-bout permet de mesurer l'étirement d'une chaîne peptidique. Elle peut se définir comme la distance entre l'azote du N-terminal et l'oxygène du C-terminal ou entre les carbones alpha du N-terminal et du C-terminal.

3.3 RMSD

En bioinformatique, le RMSD « *Root-Mean-Square Deviation of atomic positions* ») permet de comparer des superpositions de structures de protéines. Le calcul du RMSD permet de déterminer la qualité de l'alignement entre deux structures. On peut ainsi calculer la distance spatiale entre deux atomes qui sont sensés être superposés. On va passer au carré, calculer la moyenne des distances, appliquer la racine carrée et obtenir une mesure de la qualité de la superposition.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2]}$$

Le calcul du RMSD global permet la superposition de deux conformations de structures afin de trouver le maximum d'atomes alignés avec le plus faible écart.

Le calcul du RMSD local permet de trouver les régions les plus communes entre deux conformations.

4. Programme informatique

4.1 Construction du programme

Le programme que nous avons réalisé est divisé en deux parties distinctes : un fichier `fonctions.py` qui regroupe les différentes fonctions utilisées dans le programme et un fichier `projet_barstar.py`, le main de notre projet composé des données de sorties et de l'affichage graphique. Le programme est exécuté de la façon suivante :

`./projet_barstar.py <fichier d'entrée au format .pdb>`

4.2 Fonctions principales

La première fonction du programme est ***parserPDB***, elle permet de récupérer les coordonnées de chaque atome de chaque acide aminé de chaque conformation de la protéine. Elle prend en entrée le nom du fichier `.pdb` et renvoie un dictionnaire contenant toutes les informations nécessaires au programme. Ensuite, une autre fonction, ***parserPDB_CA***, permet de récupérer les mêmes informations que précédemment, sauf qu'elle ne concerne uniquement les informations relatives aux carbones alpha présents. Elle prend en entrée le nom du fichier `.pdb` et renvoie un dictionnaire contenant toutes les informations sur ces carbones alpha.

La procédure ***rmsd_global*** permet de calculer la distance moyenne entre les atomes de deux conformations. Elle prend en entrée le dictionnaire des carbones alpha et modifie ce dictionnaire en y ajoutant le RMSD entre un modèle de référence et le modèle étudié pour chaque conformation.

La procédure ***rmsd_local*** permet de calculer l'ensemble des RMSD entre des paires d'atomes. Elle prend en entrée un dictionnaire (soit le dictionnaire complet de `parserPDB`, soit celui de `parserPDB_CA`) et y ajoute les RMSD entre chaque paire d'atomes.

La fonction ***rmsd_moyen*** permet de calculer le RMSD local moyen de chaque acide aminé au cours du temps. Elle prend en entrée un dictionnaire et renvoie la liste de ces RMSD pour chaque conformation.

La procédure ***rayon_giration*** permet de calculer la distance entre le centre de masse d'une molécule et son atome le plus éloigné. Le calcul du centre de masse est à faire en amont de l'utilisation de cette fonction. Elle prend en entrée le dictionnaire obtenu avec `parserPDB` auquel on a ajouté les calculs de centre de masses et le modifie en ajoutant le rayons de giration de chaque conformation.

La procédure ***distance*** permet de calculer la distance entre chacun des centres de masse des résidus et la conformation étudiée. Elle prend un dictionnaire en entrée et le modifie et y ajoute les distances pour chacun des résidus.

La fonction ***distance_moyenne*** permet de calculer la moyenne des distances au cours du temps entre le centre de masse d'un résidu et le centre de masse de la protéine. Elle prend un dictionnaire en entrée et renvoie la distance moyenne de chaque résidu.

La procédure ***centerMassOfConf*** permet de calculer le centre de masse de chaque conformation en tenant compte de tous les atomes de chaque résidus. Elle prend en entrée le dictionnaire obtenu avec `parserPDB` et le modifie en y ajoutant les coordonnées du centre de masse pour chaque conformation.

La procédure *centerMassOfRes* permet de calculer le centre de masse de chaque résidu en tenant compte de tous leurs atomes. Elle prend en entrée le dictionnaire obtenu avec *parserPDB* et le modifie en y ajoutant les coordonnées du centre de masse de chaque résidu.

5. Résultats

A l'aide du programme décrit précédemment, les changements conformationnels de la BARSTAR seront analysés au cours du temps. Pour cela, nous allons séparer en deux parties distinctes l'étude de cette protéine : une analyse de la dynamique moléculaire au niveau global et au niveau local.

Nous utiliserons ainsi des jeux de données différents : le premier est constitué de 200 mesures de coordonnées d'atomes de la BARSTAR à 100 ps d'intervalle et le deuxième est constitué de 2000 mesures à 10 ps d'intervalle.

Nous analyserons les résultats conformationnels sur le fichier à 2000 mesures puis nous les comparerons ensuite à ceux obtenus sur 200 mesures. Ce choix est porté sur le fait qu'un plus grand nombre d'informations est contenu dans le plus gros fichier et les mesures seront ainsi plus précises.

5.1 Analyse des changements conformationnels globaux

Les deux indicateurs que nous utilisons dans cette étude globale sont le rayon de giration et le RMSD. Observons leur évolution au cours du temps pour 2000 modèles.

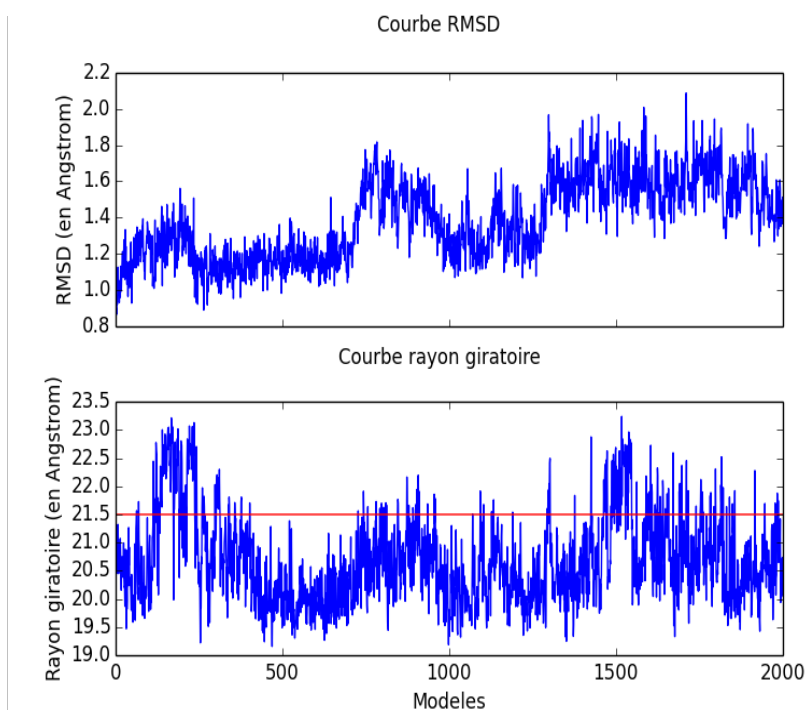


Figure 1: RMSD Global et Rayon de giration pour 2000 modèles

La figure des RMSD globaux met en avant la variabilité des conformations. En effet, on peut observer la différence entre le premier modèle (modèle de référence) et les autres. Plus le RMSD des modèles est différent à celui de référence, plus la différence conformationnelle entre ces modèles est élevée.

La courbe des rayons de giration permet également de mettre en avant une variabilité de cette mesure au cours du temps. Ces informations nous permettent de dire que la protéine subit des changements conformationnels au cours du temps, elle n'est pas fixe.

En comparant ces deux graphiques, on observe une corrélation entre les variations des deux mesures. Le RMSD représentant une distance globale (moyenne) entre deux atomes de deux conformations et le rayon de giration représentant l'encombrement stérique de la protéine, cette corrélation était attendue.

5.2 Analyse des changements conformationnels globaux

Au niveau local, nous allons comparer l'évolution du RMSD moyen de chacun des résidus aux données d'enfouissement des résidus afin de déterminer s'il y a un lien entre la mobilité et l'enfouissement d'un résidu.

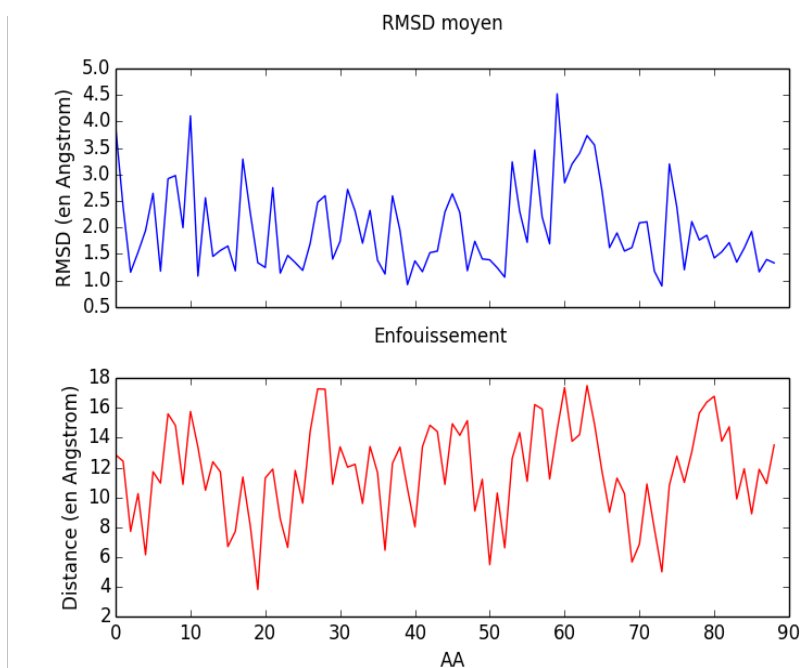


Figure 2 : RMSD moyen et enfouissement de chaque résidu

Nous pouvons distinguer les deux régions les plus flexibles de la protéine en observant les résultats du RMSD moyen : entre les acides aminés 5 à 10 et entre les acides aminés 50 à 65. on peut également remarquer que ces régions correspondent à des distances très élevées des résidus. On observe ainsi une corrélation entre un RMSD élevé et un enfouissement faible du résidu. Nous pouvons ainsi nous demander si les acides aminés qui possèdent un grand RMSD sont toujours en surface de la protéine. De plus, nous voulons également regarder si les acides aminés avec un enfouissement profond (ceux qui sont au cœur de la protéine) peuvent avoir un fort RMSD. Pour vérifier cela, nous pouvons analyser les résultats de RMSD d'un résidu en surface et ceux d'un résidu enfoui. De plus, nous pourrions observer les valeurs de RMSD de résidus qui semblent avoir des comportements atypiques, c'est-à-dire ne correspondant pas à un enfouissement faible.

Afin d'observer des comportements atypiques des acides aminés, nous avons recherché, graphiquement, des résidus présentant un grand enfouissement dans la protéine (proche du cœur de la protéine) et un grand RMSD global ou alors une combinaison opposée. Après analyse des résultats obtenus, nous n'avons pas pu mettre en évidence d'acide aminé répondant à ces critères, nous n'avons donc pas trouvé de contre-exemple dans cette étude.

6. Conclusion

6.1 Comparaison des résultats avec d'autres données

Nous avons pour le moment analysé nos données sur le fichier de 2000 conformations séparées par 10ps. Nous souhaitons désormais confirmer ou infirmer nos résultats avec le second fichier de 200 modèles. Si les résultats sont les mêmes, cela signifie que nos résultats conduisent à une même conclusion. Cela signifiera également que le fichier contenant 200 conformations est suffisant pour traiter notre sujet. Si les résultats sont différents, nous pourrions nous demander si nous pouvons réellement apporter une conclusion à notre projet et ensuite rechercher les raisons qui mènent à ces différences.

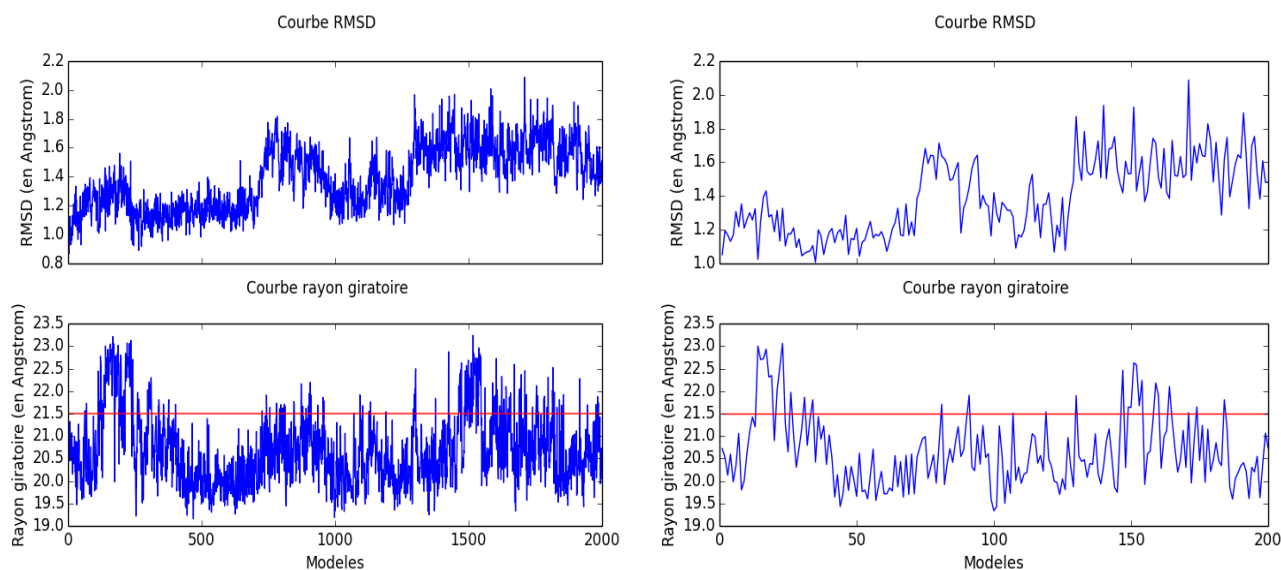


Figure 3 : RMSD Global et Rayon de giration pour 2000 modèles à gauche et 200 à droite

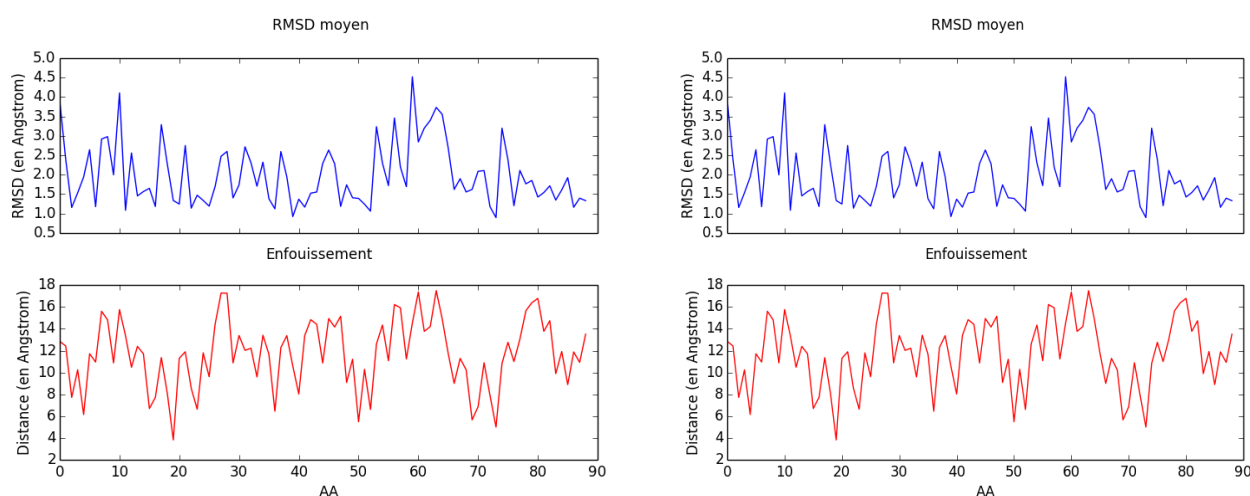


Figure 4 : RMSD Moyen et enfouissement pour 2000 modèles à gauche et 200 à droite

On remarque ainsi que les graphiques obtenus sont parfaitement superposables. Quel que soit le jeu de données, les variations des mesures sont les mêmes. Cela permet ainsi de dire que l'étude des conformations sur 200 mesures est suffisamment précise pour les calculs du RMSD et du rayon giratoire au cours du temps. Pour cette étude, 200 mesures sont donc suffisantes et nous permettent ainsi de gagner du temps d'attente des résultats. Néanmoins, dans un souci de précision des résultats, un grand nombre de mesures est toujours intéressant.

6.2 Comparaison des résultats avec Kimura et al

Afin de répondre aux questions posées en 5.2, nous allons étudier les acides aminés 39 aspartate et 76 glutamate afin de comparer les résultats obtenus. L'étude de ces acides aminés a permis de donner les résultats suivants :

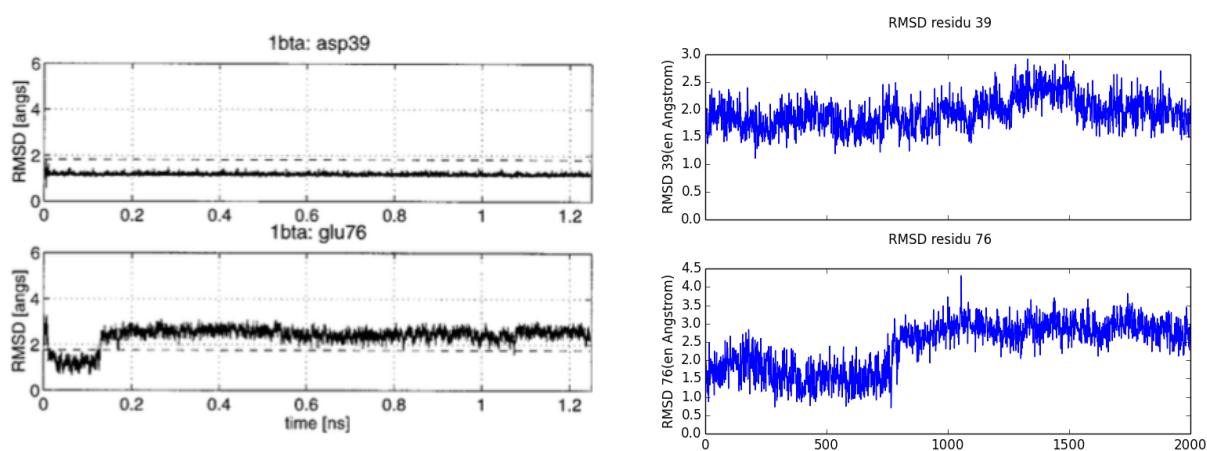


Figure 4 : Variation du RMSD de 39ASP et 76GLU au cours du temps. Résultats de Kimura et al à gauche et résultats obtenus sur 2000 modèles à droite

On remarque que les deux graphiques sont similaires, les valeurs des RMSD se situent dans les mêmes ordres de grandeur. Cela nous permet donc de dire que les résultats obtenus semblent concorder avec ce qui a été obtenu dans l'article, et donc que nos résultats semblent être validés.

7. Discussion

Avec les résultats obtenus sur le RMSD et le rayon giratoire, nous pouvons confirmer que la structure de la BARSTAR est mobile et donc qu'elle subit des changements conformationnels en solution. De plus, les intervalles de ces résultats (entre 1 et 2 Angström pour le RMSD et entre 19 et 23 Angström pour le rayon de giration) nous montrent que ces changements sont importants à l'échelle de la protéine.

De plus, l'observation des résultats liés à la position d'un résidu dans la protéine permet de soumettre l'hypothèse que cette position joue un rôle dans la mobilité du résidu considéré. En effet, pour le cas de l'acide aminé 39, on observe que le résidu est enfoui dans la structure et l'intervalle des valeurs des RMSD est de 1 à 3 Angström. En revanche, pour l'acide aminé 76, par contre, la

position de surface a un intervalle de valeurs de RMSD de 0.6 à 4.2 Angström. Cela nous permet ainsi de conclure que les résidus enfouis dans la protéine sont moins mobiles que ceux qui sont présents à la surface de celle-ci.