**Etude de la Barstar**

**Introduction**

Contexte générale :

Les interactions protéiques participent et jouent un rôle important dans de nombreux processus biologiques, comme le métabolisme, la signalisation cellulaire, l’expression des gènes et la réponse immunitaire. Comprendre le fonctionnement de ces interactions est donc d'une très grande importance.

La conformation de protéine peut être différente lorsqu’elle est seule et lorsqu’elle est en complexe avec son partenaire. Dans un couple de protéines en interactions, certains résidus sont essentiels pour une induction de la formation du complexe, ou une inhibition de la formation du complexe, ou un maintien de la conformation en complexe, et que ces rôles sont en étroites relations avec la position, la flexibilité ainsi que la propriété physico-chimique de ces résidus.

Une étude focalise sur ces changements conformationnels peut apporter des bénéfices aux nombreuses méthodes d'analyses *in silico* développées de nos jours pour une étude de l’interaction protéique.

Objectif de notre étude :

On cherche à étudier au cours du temps la stabilité d’une structure protéique, la BARSTAR-BARNASE, et les changements conformationnels qu’elle subit.

La Barstar est une protéine synthétisée par des bactéries (Bacillus) dont le but est d’inhiber l’activée de la barnase (qui est une ribonucléase). Ensemble, ces deux protéines forment un complexe protéine-protéine. L’inhibition de la barnase nécessite une interaction entre les deux protéines afin que la barstar couvre le site actif de la barnase, ce qui implique vraisemblablement des changements conformationnels.

Nous allons, pour cela, nous intéresser aux changements conformationnels globaux afin de savoir si il y a des mouvements de boucles de grandes amplitudes, des mouvements allostériques, du dépliement ou encore de la compaction.

Nous nous intéresserons également aux changements conformationnels locaux dans le but d’identifier des petits mouvements de boucles ou encore des mouvements de chaîne latérales.

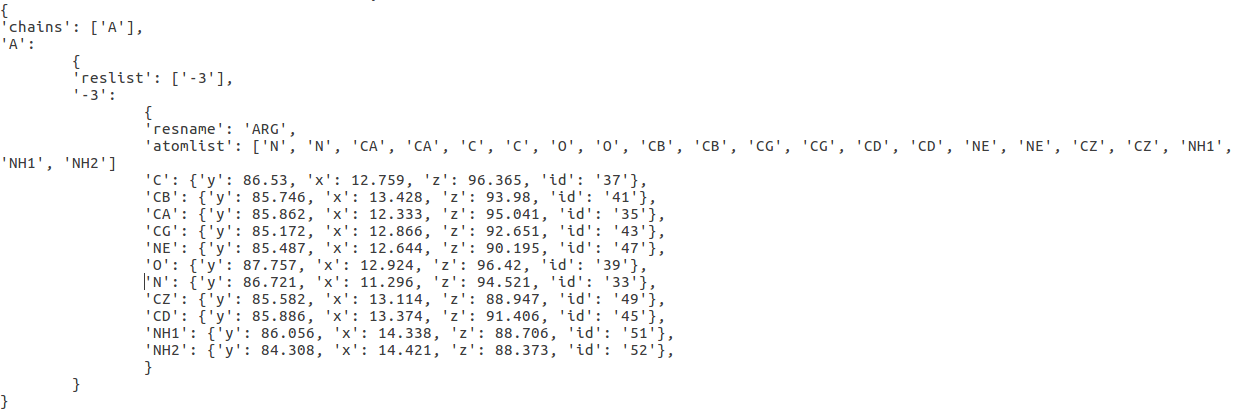
Nous avons réaliser l’ensemble de nos analyses sur 2 fichiers de données : l’un avec des données extraites toutes les 10 ps et l’autre avec des données extraites toutes les 100 ps.

**Méthodes**

Nous avons un script permettant d’analyser l’ensemble des fichiers pdb en lui fournissant le chemin de répertoire contenant des fichiers pdb et le type d’analyse souhaité (global ou local ou both). Un fichier start\_prot\_only.pdb contenant la structure d’origine de la protéine est nécessaire d’être présente dans le répertoire. A la fin de programme, un dossier appelant « PythonProgResults » contenant tous les fichiers de sortie sera créé dans le répertoire que l’utilisateur avait fourni.

Afin de réaliser une analyse globale et une analyse locale à partir d’un fichier pdb. Plusieurs fonctions basiques sont créées :

* ParsingPDB pour la création d’un dictionnaire à partir d’un fichier pdb. La sortie est pratiquement similaire à celle que nous avons vu en TD, à la différence que l’on a rajouté le numéro de modèle comme clé. Un exemple de la sortie de présenté ci-dessous :



}

{MODEL N°

* « CM », « Distance », « RMSD » sont des fonctions de base pour le calcul de centre de masse, de distance dans l’espace 3D et le calcul de RMSD respectivement.

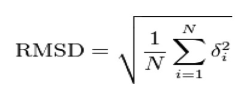
Le calcul du centre de masse en négligeant la masse atomique :

Où x, y, z sont l’abscisse, l’ordonnée et la cote respectivement, n est le nombre de points participés.

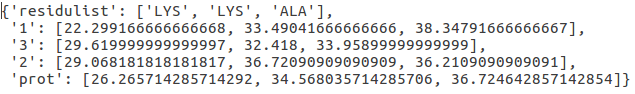
Le calcul de distance dans l’espace 3D :

Distance=

Le calcul de RMSD :



* « CMglob » pour la création d’un dictionnaire contenant le centre de masse des résidus (Residu Sequence Number en clé) d’une protéine ainsi que le centre de masse d’une protéine. Un exemple de sortie est présenté ci-dessous :



La clé « residulist » permet d’avoir une liste de résidu dans l’ordre, dans cet exemple, on peut voir que 1 est une lysine, 2 est une lysine, 3 est une alanine. Les listes correspondent aux coordonnées [x,y,z] du centre de masse.

* « createClass » pour le classement des valeurs en certain nombre de classes.
* « Temps » et « graph » pour tracer les graphes.

Analyse globale

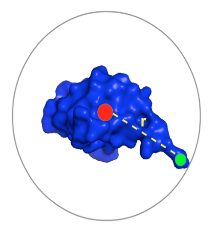
RMSD global

On a créé une fonction « RMSDglobal » pour calculer le RMSD global de chaque conformation en comparant avec la structure d’origine de la protéine. Le RMSD global permettant de rendre compte de la déviation structurale entre deux protéines alignées. Cette fonction calcule la distance entre le carbone alpha d’un résidu dans la conformation numéro N et celui dans la structure d’origine. Ces distances sont stockées dans une liste appelant « liste\_delta » dans la fonction. Ensuite, le RMSD global est calculé. A la sortie, on obtient un dictionnaire ayant le numéro de conformation comme clé et le RMSD global comme valeur associée. Un exemple de sortie est donné ci-dessous :

C:\Users\BEI\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot from 2017-04-30 18-48-08.png

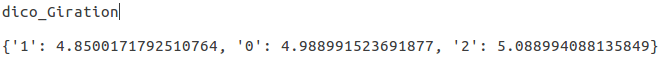
Rayon de giration

On a créé une fonction « giration » permet de calculer le rayon de giration de chaque conformation. Le rayon de giration permet de rendre compte du rayon de la protéine correspondant à la distance entre le centre de masse de la protéine et le résidu le plus éloigné du centre de masse de la protéine.



*Figure : le rayon est représenté en jaune, le centre de masse de la protéine est représenté en rouge, le centre de masse du résidu le plus éloigné est représenté en vert*

Cette fonction calcule la distance entre le centre de masse de chaque résidu et le centre de masse de la protéine. Ces distances sont stockées dans une liste, et le maximum de cette liste correspond au rayon de giration. A la sortie, elle retourne un dictionnaire ayant le numéro de conformation comme clé et le rayon de giration de cette conformation comme valeur associée. Un exemple de la sortie est donné ci-dessous :



Analyse locale

On a créé une fonction « Enfouissement » pour calculer la distance entre le centre de masse de chaque résidu et celui de la structure d’origine (pour chacune des conformations). Elle crée un dictionnaire permettant d’associer les distances calculées pour chaque conformation au numéro du résidu concerné. La moyenne des valeurs du distance pour chaque résidu est la moyenne d’enfouissement. Pour chaque résidu, la moyenne de l’enfouissement rangée dans un dictionnaire qui possède le numéro de résidu comme clé et l’enfouissement moyen comme valeur associée.

On a créé une fonction « RMSDlocal » pour calculer le RMSD de chaque résidu et le RMSD moyen de chaque résidu pour l’ensemble des conformations. Le RMSD local permet d’identifier les résidus présents dans les régions flexibles. Cette fonction calcule la distance entre l’atome d’un résidu dans la conformation N et celui dans la structure d’origine. Ces distances d’un résidu donné sont stockées dans une liste pour pouvoir ensuite calculer le RMSD pour chaque résidu sur l’ensemble des conformations.

Un dictionnaire comprenant en clé le numéro de résidus avec pour valeurs ses RMSD pour chaque conformation est alors crée.

Pour chaque résidu, nous avons ensuite calculé la moyenne des RMSD et nous l’avons stocké dans un nouveau dictionnaire (numéro de résidu comme clé et RMSD moyen comme valeur associée).

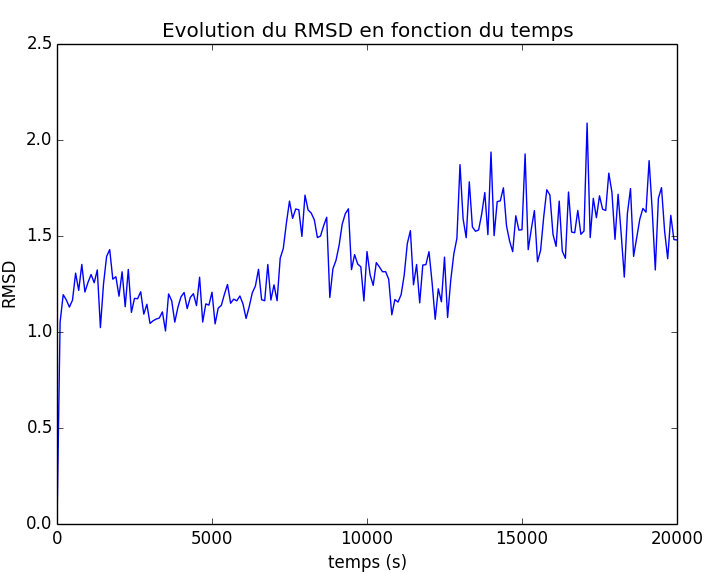
En sortie, on obtient un tuple : un dictionnaire de RMSD pour chaque résidu et un dictionnaire de RMSD moyen.

**Résultats et Interprétation**

1. **Analyse des changements conformationnels globaux**

Le but ici est d’étudier si la Barstar subit des changements conformationnels majeurs en solution.

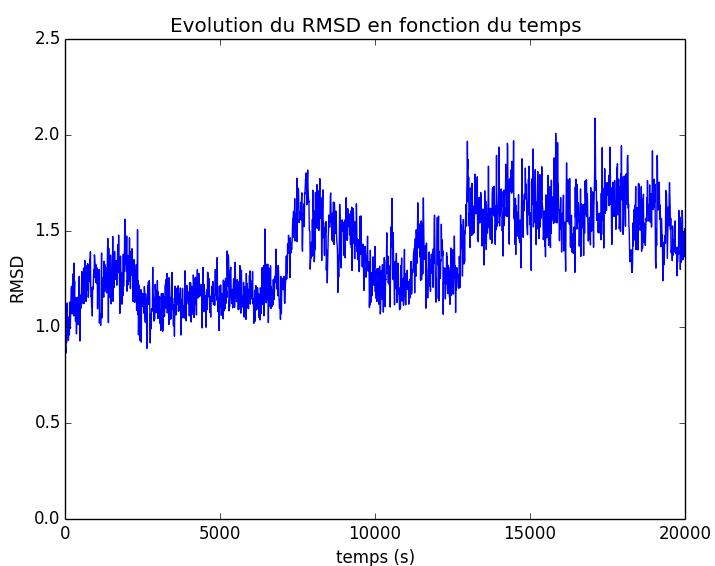
Pour cela, nous avons regardé l'évolution du RMSD et du rayon de giration en fonction du temps dans un premier temps et en fonction de la conformation dans un deuxième temps.



On sait que le RMSD (Root Mean Square deviation) rend compte de la ressemblance entre 2 structures (et donc de la déviation) : plus le RMSD est petit, plus les 2 structures comparées se ressemblent.

Les mesures du RMSD sont ici en Angstrom.

Dans la figure 1, le RMSD est représenté au cours du temps. Les rayons de giration sont ici en Angstrom.

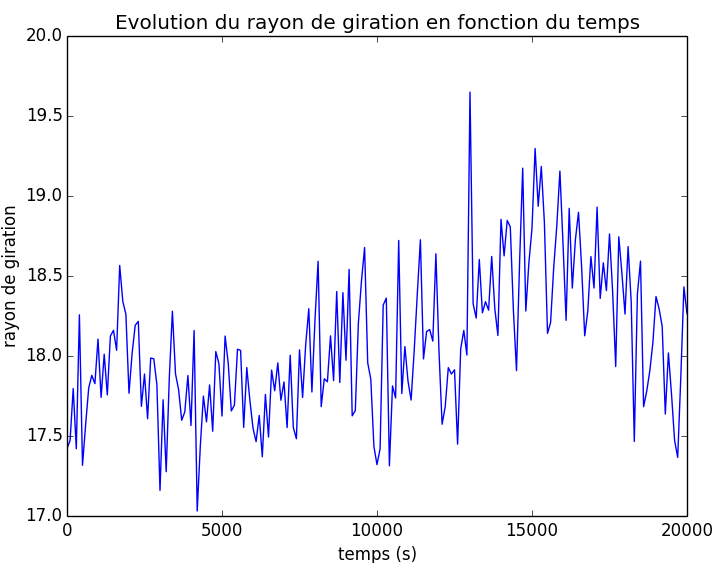
Quand on représente l’évolution du RMSD au cours temps, on se rend compte que celui-ci varie entre 0 et 2. On observe également une légère augmentation du RMSD au cours du temps.

Données extraites toutes les 100 ps

Les variations sont fréquentes et correspondent à des changements de conformationnels. Il y a donc des changements de conformations globaux de la protéine

.

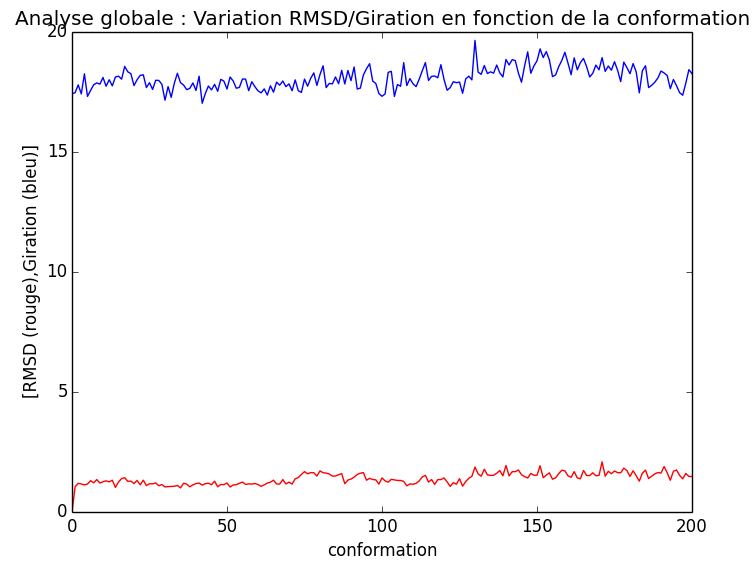
Données extraites toutes les 10 ps

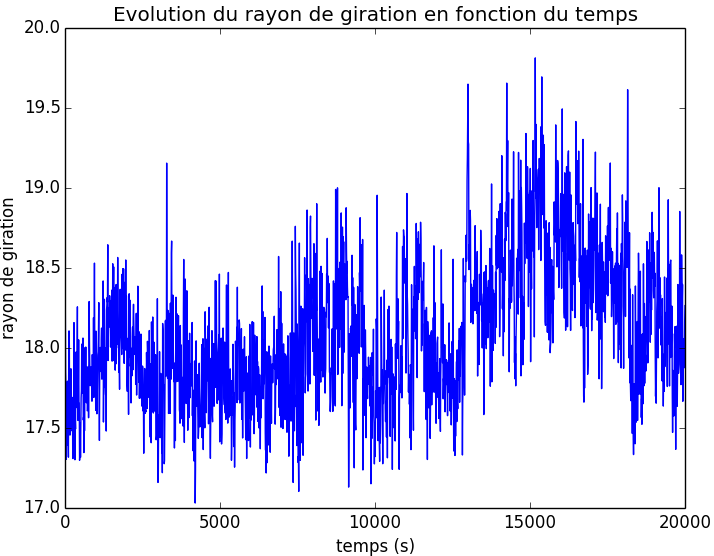


Le rayon de giration varie entre 17 et 20 Angstrom et on observe des variations fréquentes du rayon de giration pour les deux graphiques. Comme le rayon de giration rend compte du rayon de la protéine, cela signifie que le rayon de la protéine change fréquemment (augmente et diminue). Cela peut donc correspondre à des mouvements de dépliements et de compaction de la protéine.

Données extraites toutes les 10 ps

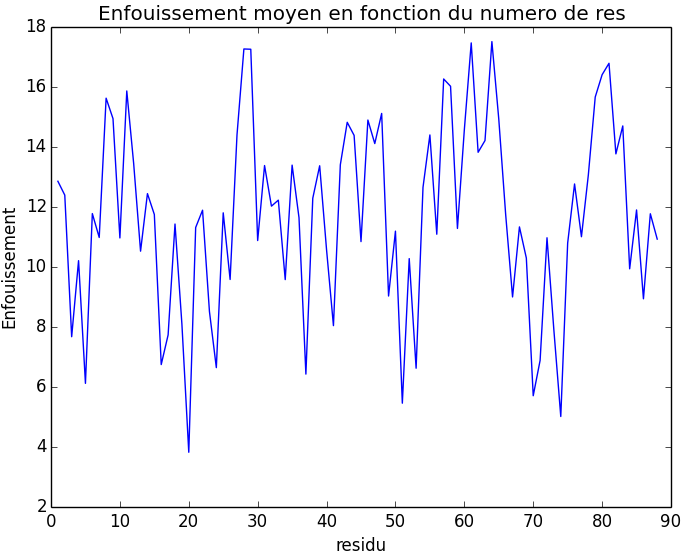
Pour les 2 mesures réalisées, nous avons choisis de représenter les variations sur 2 jeux de données différents (toutes les 10 ps et toutes les 100ps). Cela permet de vérifier la robustesse des observations. Pour les deux mesures, on voit que les tracés sont similaires pour les deux jeux de données : l’information extraite des 200 conformations est donc suffisante pour effectuer l’analyse globale.



On peut comparer comment ces 2 métriques varient par rapport à la structure d'origine pour chacune des conformations (=traduction du fichier de sortie): pour les 2 métriques, on retrouve bien les variations décrites plus haut. Il est normal de retrouver ces variations car le fichier contient autant de mesures de temps que de conformations.

D’après les résultats précédents, on peut donc dire que la protéine subit des changements conformationnels globaux en solution. Ces changements correspondent sûrement à des mouvements de dépliements /compaction. Le but est maintenant d’identifier les résidus qui sont impliqués dans ces changements de conformations (analyse des changements conformationnels locaux).

1. **Analyse des changements conformationnels locaux**

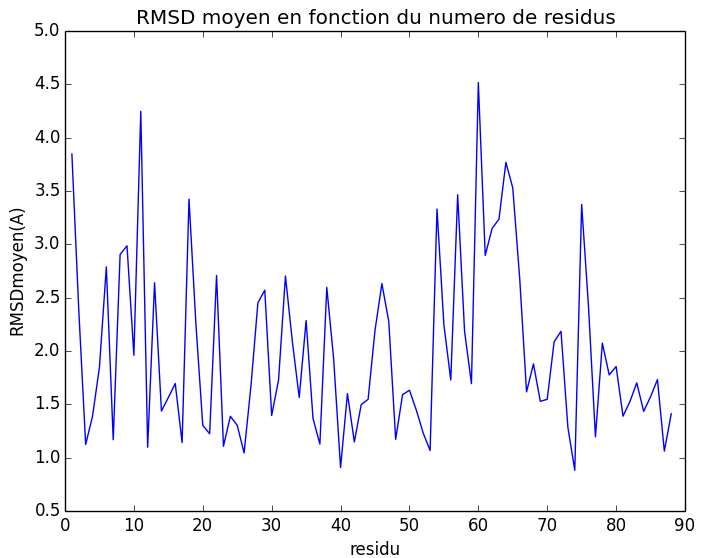
Le but ici est d’identifier les régions les plus flexibles de la protéine qui pourraient jouer un rôle dans l’interaction.

Nous voulons dans un premier temps localiser les résidus sur la protéine afin d’identifier ceux situés en surface et ceux enfouis. Pour cela, nous avons mesuré l’enfouissement de chaque résidu.

Les résidus les plus enfouis sont ceux qui possèdent une faible valeur d’enfouissement alors que les résidus de surface présentent un enfouissement élevé.

On peut ainsi facilement identifier la position de quelques résidus en les classant dans deux groupes différents :

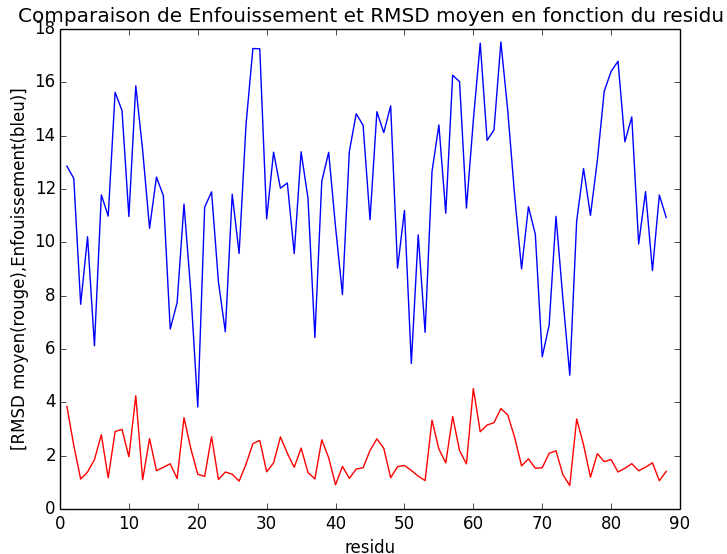
* Résidus les plus enfouis :
* Ce sont des résidus qui ont des valeurs d’enfouissement inféreieures à 10 A. On peut par exemple citer , et .
* Résidus de surface :
* Ce sont les résidus qui ont des valeurs d’enfouissement supérieures à 10A. Dans le fichier de sortie, les résultats significatifs ont des valeurs proche de 13/14A. Ces résidus sont plus à la surface que les autres résidus de surface.

Dans un deuxième temps, nous voulons identifier les résidus situés dans les régions les plus flexibles de la protéine. Nous avons alors représenté pour chaque résidu le RMSD moyen qui lui est associé.

On peut observer que certains résidus ont des valeurs moyennes de RMSD particulièrement élevées : il s’agit par exemple des résidus LYS60, 64, le 11 et le 1.

Comme ces résidus ont des valeurs de RMSD élevées, ils sont situés dans des régions flexibles.

On remarque également que les résidus présents dans les régions flexibles sont également à surface de la protéine.

Ce qui nous amène à une question : est-ce que les régions les plus flexibles sont le plus souvent à la surface ou enfouies ?

D’après le graphique, on voit que ce sont les régions les plus flexibles qui se trouvent le plus souvent à la surface des protéines. En effet, quand l’enfouissement moyen est élevé, le RMSD moyen est également élevé.

Cette observation est logique puisque la protéine a besoins à sa surface d’être flexible pour pouvoir interagir avec les autres constituants moléculaires. Ces régions flexibles permettent de connecter entre eux des éléments réguliers comme les brins ou les hélices.

Nous avons représenté sous Pymol la structure de la protéine. Nous avons décidé de représenter les résidus qui sont en contact entre les deux protéines du complexe. Les résidus en interface sont les résidus GLU 76, ASP39, ASP35, THR42, TRP38, VAL73 HIS102…

Quand on regarde les propriétés de ces résidus, on constate qu’ils ont tous des valeurs d’enfouissement autour de 10/11A et qu’ils n’ont pas de RMSD significatifs.

* Une partie résidus sont situés en surface mais partiellement.
* De plus, les valeurs RMSD de ces résidus montrent que ces derniers ne sont pas situés dans des régions flexibles. Cela peut s’expliquer facilement : la zone de contact entre les 2 protéines permet à la Barstar d’inhiber l’activité de la barnase quand cela est nécessaire. Cette région ne doit pas être flexible car sinon l’attachement serait trop faible.

Pour la suite de l’étude, nous avons sélectionné les résidus  60, 11, 1 (qui ont la particularité d’être dans des régions flexibles et à la surface), les résidus 39 et 17 (qui sont dans des régions enfouies de la protéine) et le résidu 15 (qui est à la surface mais pas vraiment et qui n’est pas dans une région flexible).

On souhaite à présent regarder l’évolution du RMSDlocal au cours du temps pour l’ensemble de résidus sélectionnés

Analyse des résultats flexibles et surface :

Ces 3 résidus ont été identifiés comme étant situés dans des régions flexibles et à la surface de la protéine. On constate pour ces 3 résidus des variations importantes de RMSD au cours du temps. Cela laisse penser que ces résidus changent rapidement et souvent de conformation (et la conformation est différente de celle de la structure d’origine). En effet, le RMSD prend en compte la distance séparant deux molécules. Si la conformation change, la distance va être modifiée et le RMSD va varier.

* pour le résidu 60, la conformation au départ est celle pour un RMSD proche de 0.5 A. Ce résidu reste toujours éloigné de la structure de départ et a quelques conformation pour un RMSD de 1.4 A.
* 1 : reste toujours éloigné de la conformation de départ avec quelques rares conformations pour un RMSD de 2
* 11 : idem

🡺Comme sont à la surface, on peut en déduire que ces conformations sont induites par l’effet du solvant ou par l’interaction des 2 protéines entre elles.

Analyse des résultats enfouis :

* 17

Pour ces 2 résidus, les valeurs de RMSD sont plus faibles (normal car pas dans des régions flexibles).

Les variations semblent plus rapides que pour les résidus 60, 65 et 57. Les amplitudes de variations sont plus faibles que pour les résidus précédents. De plus, les mêmes conformations sont souvent visitées : par exemple, la conformation à 1.5 A est souvent visitée par le résidu 39. =>ce phénomène est plus fréquent que pour les 3 d’avant

Résidus surface et non flexible :

* 39
* 76

Ces deux résidus ont était identifié comme étant situés à la surface de la protéine dans des régions à priori non flexibles.

Pour glu76 : Jusqu’au temps 7000, il semble qu’il y ait une conformation privilégiée. Après, on observe un changement brutal de la valeur du RMSD. Il semble à présent qu’une autre conformation soit privilégiée.

La conformation au début reste donc proche de celle de la structure de départ. Mais s’en éloigne radicalement à partir de 7000ps.

*Est-ce que ce pourrait être l’action de la barstar sur la barnase pour inhiber l’activité de la barnase*?

**Conclusion**

# à ajouter la conclusion par rapport à notre étude

Ouverture : De nombreuses méthodes d'analyses *in silico* ont été développées dans le but d’identifier des interactions protéiques, notamment le docking moléculaire. Une modélisation précise sur les résidus, surtout les résidus dans l’interface de l’interaction, peut améliorer la sortie de ces méthodes d’analyses.