**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.

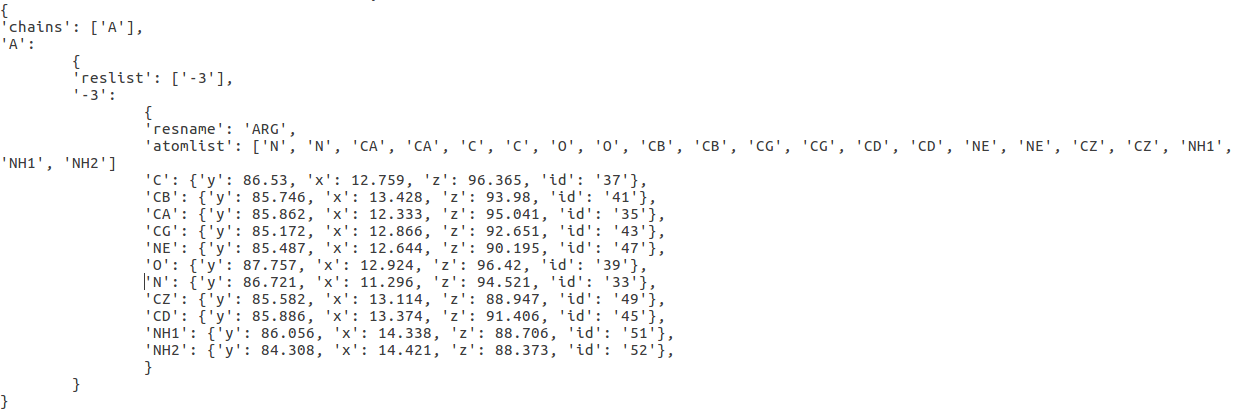
Nous avons réaliser l’ensemble de nos analyses sur 2 fichiers de données : l’un avec des données extraites toutes les 10 ps et l’autre avec des données extraites toutes les 100 ps.

**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.

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

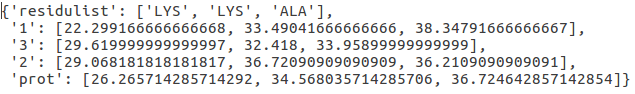
* ParsingPDB pour la création d’un dictionnaire à partir d’un fichier pdb. La sortie est pratiquement similaire à celle que nous avons vu en TD, à la différence que l’on a rajouté le numéro de modèle comme clé. Un exemple de la sortie de présenté ci-dessous :



}

{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.
* 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.

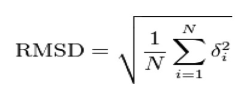
* « createClass » pour le classement des valeurs en certain nombre de classes.
* « Temps » et « graph » pour tracer les graphes.

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 de façon ci-dessous :

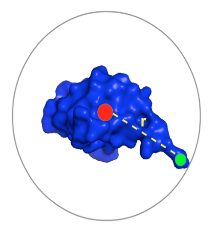


A la sortie, on obtient un dictionnaire ayant le numéro de conformation comme clé et le RMSD global comme valeur associée. Un exemple de sortie est donné ci-dessous :

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

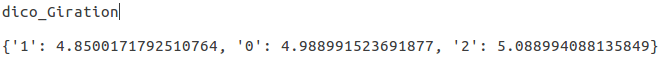
Rayon de giration

On a créé une fonction « giration » permet de calculer le rayon de giration de chaque conformation. Le rayon de giration permet de rendre compte le 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

* une fonction Enfouissement : le centre de masse de chaque résidu pour chacune des conformations a été calculé et rangé dans un dictionnaire. Pour chaque conformation, nous avons ensuite calculé entre le résidu et le centre de masse, ce qui nous donne pour chaque résidu une valeur d’enfouissement pour chaque conformation.

Pour chaque résidu, la moyenne de l’enfouissement est calculée et rangée dans un dictionnaire qui possède le numéro de résidu comme clé et l’enfouissement moyen comme valeur associée.

* une fonction « RMSD local » 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 sont stockées dans une liste appelant « liste\_delta » dans la fonction. Le RMSD est ensuite calculé de la même manière que dans l’analyse globale.

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

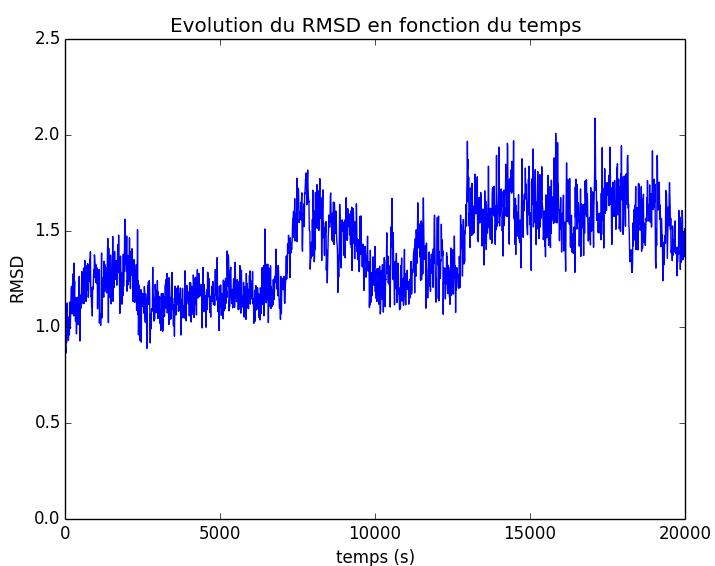
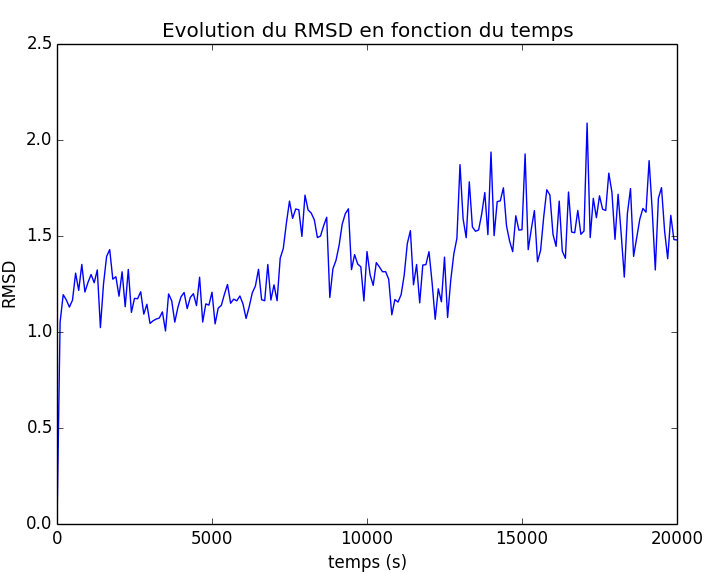
**Résultats et Interprétation**

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 et en fonction de la conformation.

Le RMSD (Root Mean Square deviation) rend compte de la ressemblance entre 2 structures (et donc de la déviation). En effet, plus le RMSD est petit, plus les 2 structures comparées se ressemblentLes mesures du RMSD sont ici en Angstrom.

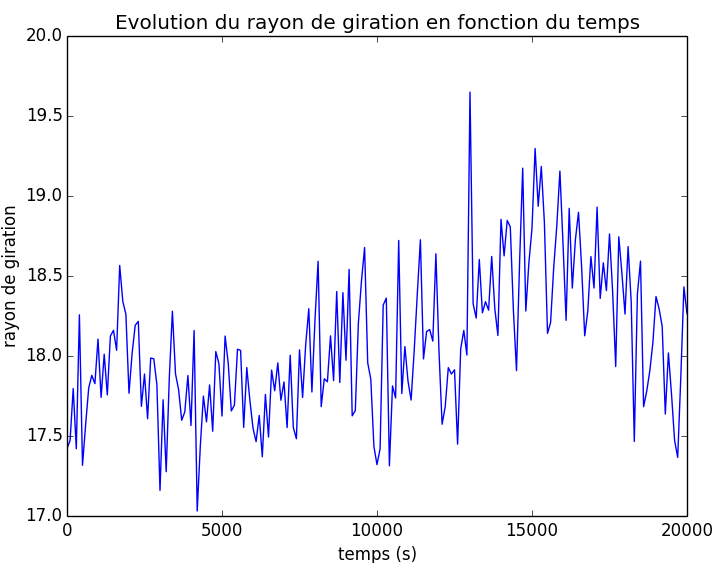


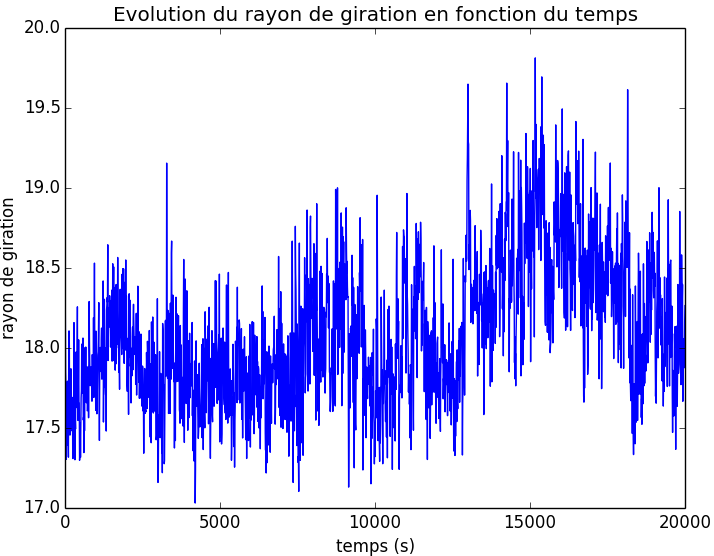
Données extraites toutes les 10 ps

Données extraites toutes les 100 ps

Quand on représente l’évolution du RMSD au cours temps, on se rend compte que celui-ci varie entre 0 et 2. Au cours du temps la valeur du RMSD reste donc petite, ce qui nous amène dans un premier temps à penser qu’il n’y a pas de grande différence entre les structures comparées aux différents temps (pas de changements conformationnels globaux majeurs)

Les rayons de giration sont ici en Angstrom.





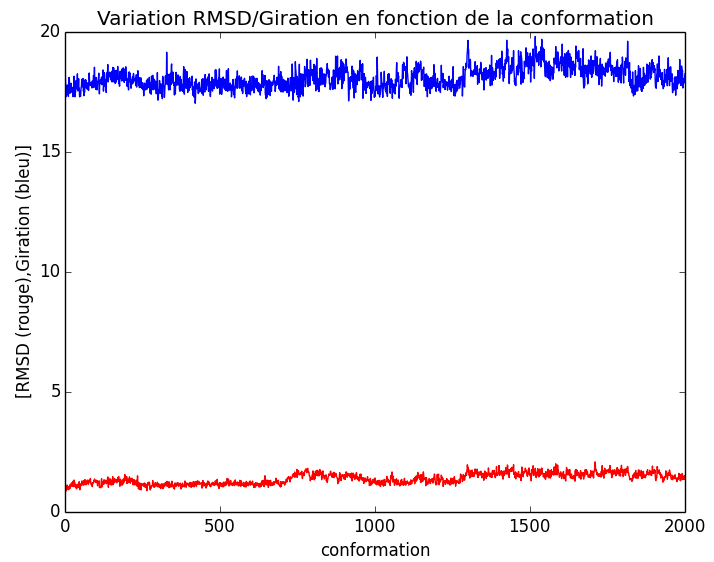
Données extraites toutes les 100 ps

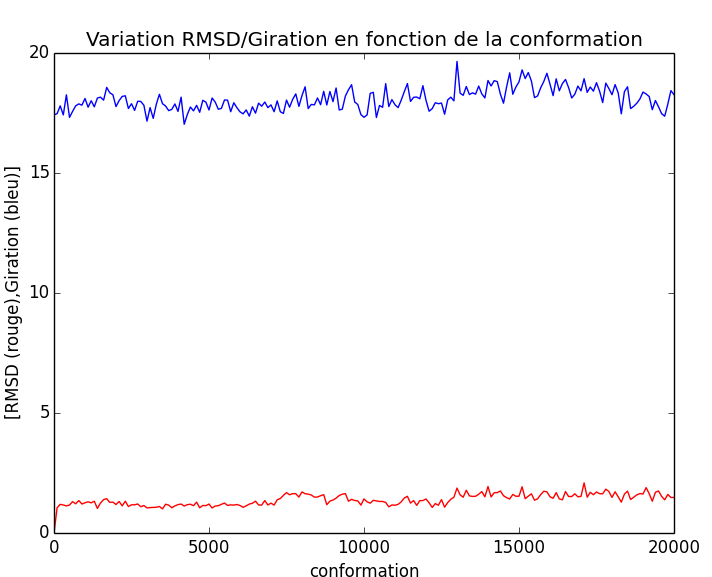
Données extraites toutes les 10 ps

Le rayon de giration est en moyenne élevée (varie entre 17 et 20 A).

Une fois ces deux mesures calculées, on peut facilement estimer comment ces 2 métriques varient par rapport à la structure d'origine.

Comme il est difficile d'interpréter la sortie (format texte), nous avons décidé de représenter les résultats sous forme d'un graphique :





Données extraites toutes les 10 ps

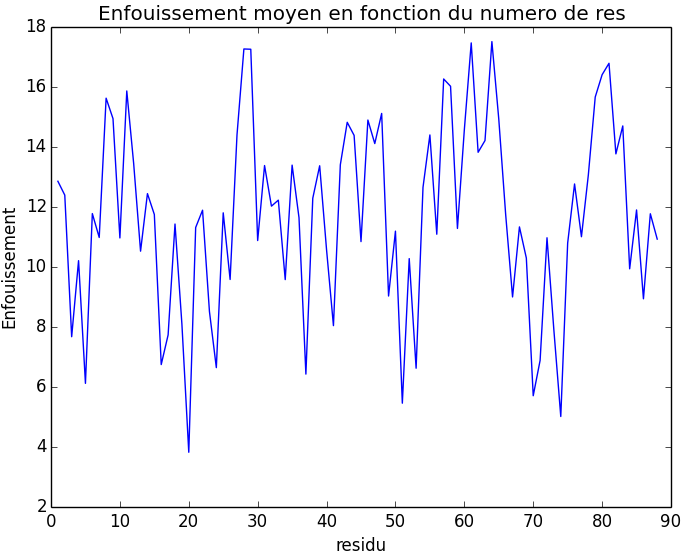
Données extraites toutes les 100 ps

On peut remarquer que pour ces 3 mesures, les variations ne sont pas très grandes. De plus, pour chacune des conformations, le RMSD est petit et le rayon de giration est grand.

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.

Enfouissement de chaque résidu. Les résidus les plus enfouis présente un enfouissement plus petit que les résidus de surface



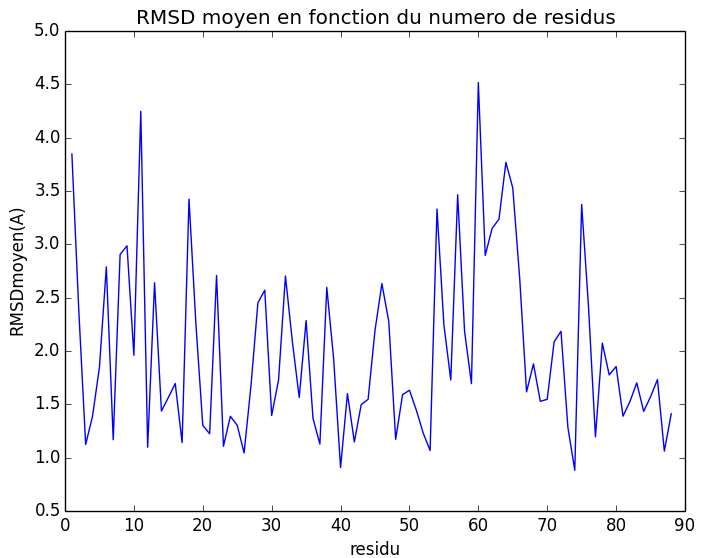
Les résidus les plus enfouis sont ceux qui possède une faible valeur d’enfouissement alors que les résidus de surface présentent un enfouissement élevé.

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

* Résidus les plus enfouis :
* Le résidu 17 possède un enfouissement de 5 environ
* Le résidu 3 avec un enfouissement proche de 4 A
* Le résidu 82
* Le résidu 39 avec un enfouissement proche de 7 A
* Résidus de surface :
* Le résidu 46 avec un enfouissement de 14 A
* Le résidu 76 avec un enfouissement de 12.7 A
* ASN65
* Enfouissement moyen :
* Le résidu 15  avec un enfouissement proche de 11A.
* Le résidu 45
* Le résidu 68
* LYS22
* GLN18
* ARG75

A partir de cette classification, on peut donc facilement dire que la GLU46 et la GLU76 sont présents à la surface de la protéine et que ARG17, ARG82 et ARG39 sont plus enfouis. Pour ceux qui possèdent un enfouissement moyen, on décide qu’ils sont présents à la surface(choix arbitraire mais fortement influencé par la publi).

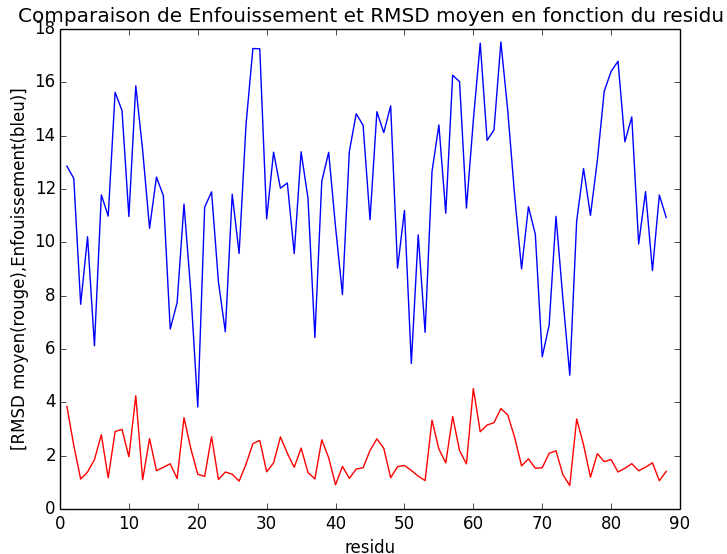
On cherche maintenant à identifier les résidus situés dans les régions les plus flexibles :



On peut remarquer que certains résidus ont des valeurs moyennes de RMSD particulièrement élevées : il s’agit par exemple des résidus LYS60, ASN65, ARG75, GLN18 ou LYS22.

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

On peut remarquer 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.

Pour la suite de l’étude, nous avons sélectionné les résidus : LYS15, GLU76, ARG17, ARG45, ASP39, ARG68.

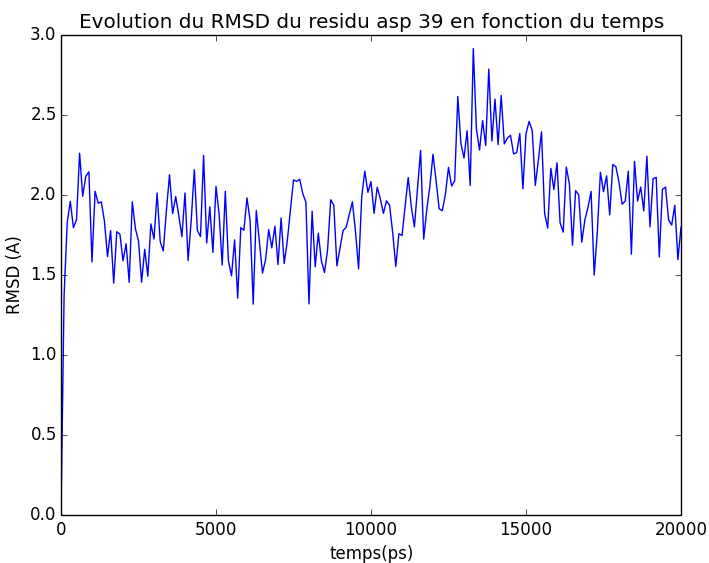
On souhaite à présent regarder l’évolution du RMSDlocal au cours du temps pour un ensemble de résidus.

Nous savons d’après les conclusions précédentes que les résidus GLU76, ARG17 et ARG45 sont exposés à la surface du complexe.

🡺Je pense qu’il faut combiner ces observations avec ceux de pymol pour conclure convenablement

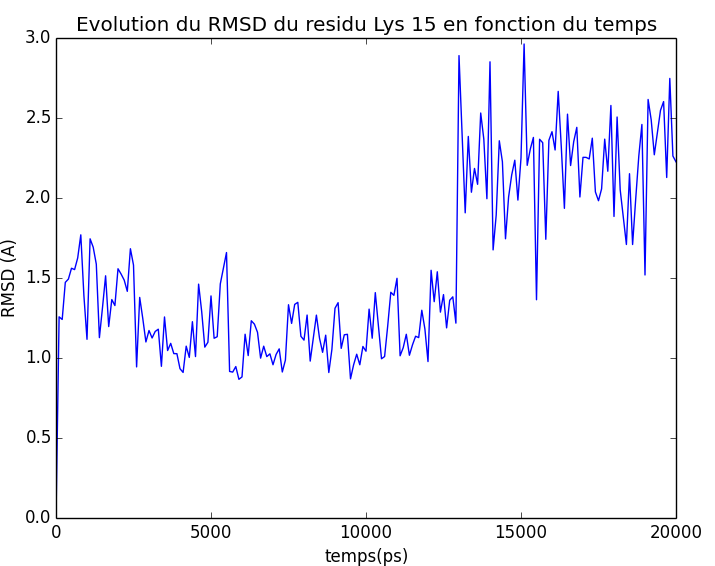
🡺regarder ou sont situés les résidus

Asp 39 : on a quelques fluctuations mais le RMSD reste faible (inférieur à 2.5 A la plupart du temps, sauf au temps 250000).

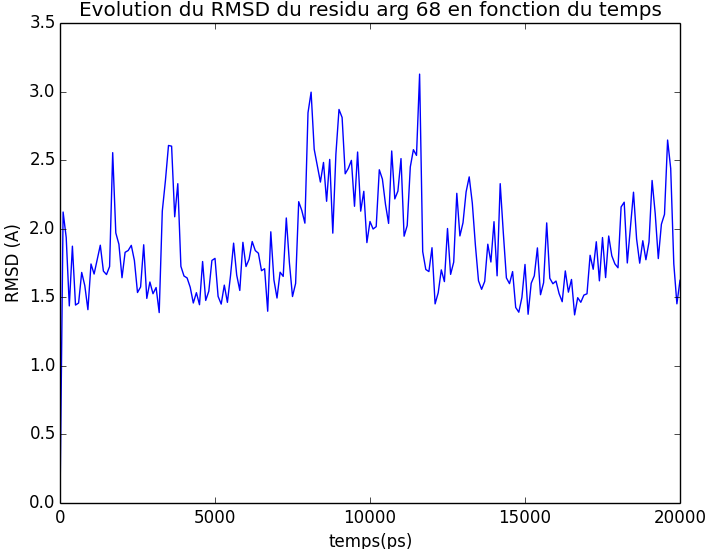


Lys 15 : on remarque plus de fluctuations pour ce résidus que pour Asp 39 :

* Les fluctuations sont surtout entre 1 et 1.5 A RMSD les 14000 première ps.
* A partir de 15000 ps, les fluctuations ont une plus grande amplitude (1.5 à 3 A).

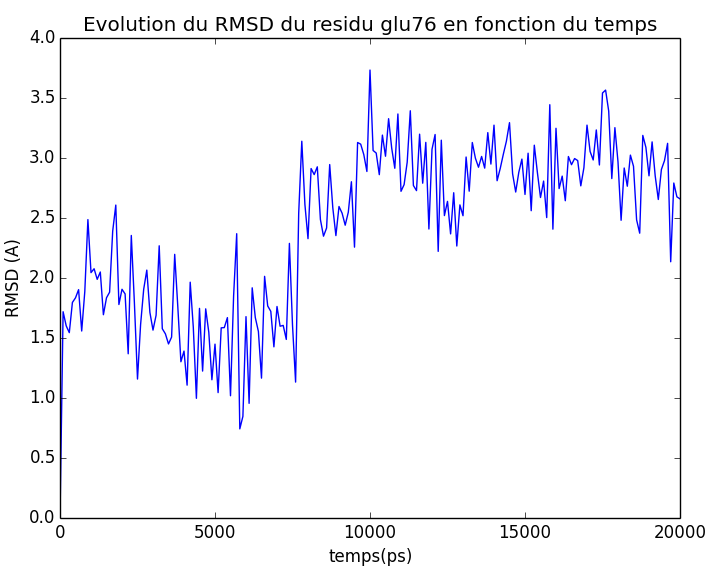


Arg 68 : Les fluctuations se font entre 0 et 3 A et sont régulières



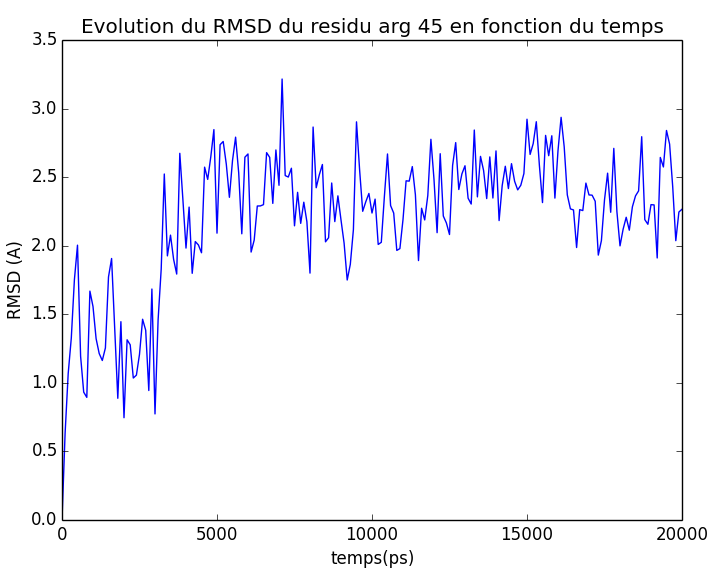
Glu 76 : valeurs plus grandes du RMSD puisque la plupart du temps, le RMSD est supérieur à 2.

* les 20000 première ps, le RMSD reste faible (inférieur à 2.5 ps)
* puis augmentation du RMSD pour atteindre 3.5 A par moments



Arg 45 :

* RMSD faible pour les 5000 premières ps
* Puis augmentation du RMSD



**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.