

# ÉTUDE DE LA DYNAMIQUE MOLÉCULAIRE – APPLICATION A LA BARSTAR

Pierre Merckaert – Ophélie Da Silva

## INTRODUCTION

---

La modélisation de complexes protéiques peut être réalisée à partir des formes non liées de protéines obtenues par cristallographie. Ces structures rigides isolées peuvent alors présenter des distorsions au niveau des interfaces, comme la présence de chaînes latérales mal orientées, ce qui diminuerait l'affinité entre les partenaires, ainsi que leur dynamique d'association (*Kimura et al., 2001*). Les interactions entre protéines sont permises par les changements conformationnels qui conduisent à leur complémentarité de forme. Or, les contraintes de distance, telles que les liaisons covalentes et non covalentes, conditionnent la rigidité des molécules. Les zones flexibles peuvent alors être déterminées via l'identification des liaisons limitant les mouvements dans la molécule.

La dynamique moléculaire permet de suivre et de comprendre en détail la structure et la dynamique à l'échelle atomique. Cette méthode simule le mouvement naturel des molécules biologiques en solution au cours du temps (*Freddolino et al., 2008*) pour prédire l'évolution de leur structure (*Lindahl, 2008*). Des études ont montré l'efficacité de cette méthode pour identifier les régions flexibles à partir de l'analyse des trajectoires obtenues (*Gohlke et al., 2004*).

La barstar est un polypeptide intracellulaire inhibiteur de la barnase (ribonucléase extracellulaire de *Bacillus amyloliquefaciens*) avec qui elle forme un complexe très serré. Le but de cette étude est d'étudier la dynamique des chaînes latérales de la surface de la barstar qui jouent un rôle important dans la formation du complexe protéique.

## MATÉRIELS ET MÉTHODES

---

### 1. DONNÉES

Les changements de flexibilité de la barstar vont être étudiés à partir de simulations de dynamique moléculaire. Les conformations ont été échantillonnées toutes les 10 et 100 ns (fichiers PDB contenant respectivement 2000 et 200 conformations). La conformation de référence est celle obtenue par cristallographie.

### 2. ROOT MEAN SQUARE DEVIATION, RMSD

L'écart quadratique moyen ou RMSD mesure de la proximité structurale entre les conformations issues de la dynamique moléculaire et la conformation du ligand dans la structure cristalline (conformation de référence). Il est défini comme :

$$RMSD_{résidu} = \sqrt{1/N \sum_{résidu \in [1,N]} [(x_{résidu} - x_{ref})^2 + (y_{résidu} - y_{ref})^2 + (z_{résidu} - z_{ref})^2]}$$

Le RMSD rend compte de la qualité de la superposition entre la conformation de référence et celle considérée. Les valeurs de RMSD croissantes traduisent des conformations de plus en plus divergentes par rapport à la conformation de référence, *i.e.* la superposition est de moins en moins bonne qualité.

### **3. CENTRE DE MASSE**

Le centre de masse de la conformation correspond au barycentre des résidus la constituant. Il est défini comme la position moyenne de l'ensemble des résidus :

$$x_{CM} = 1/N \sum_{i \in [1,N]} x_{résidu\ i}, y_{CM} = 1/N \sum_{i \in [1,N]} y_{résidu\ i}, z_{CM} = 1/N \sum_{i \in [1,N]} z_{résidu\ i}$$

Dans notre étude, le centre de masse des résidus peut être calculé de deux manières différentes : soit en calculant le barycentre des atomes du résidu considéré, soit en approxinant le centre de masse par la position du carbone  $\alpha$ .

### **4. DISTANCE**

Pour chaque résidu, sa distance au centre de masse de la conformation a été calculée telle que :

$$d_{résidu} = \sqrt{(x_{résidu} - x_{CM})^2 + (y_{résidu} - y_{CM})^2 + (z_{résidu} - z_{CM})^2}$$

Plus la distance associée à un résidu est grande, plus ce résidu est en périphérie. Au contraire, de petites distances relatives sont caractéristiques des résidus enfouis. Les mesures de distances traduisent donc la compacité d'une protéine.

### **5. RAYON DE GIRATION**

Le rayon de giration est défini pour une conformation comme la distance maximale entre un résidu et le centre de masse de la conformation.

$$r_{giration} = \max\{d_{résidu} \mid résidu \in [1, N]\}$$

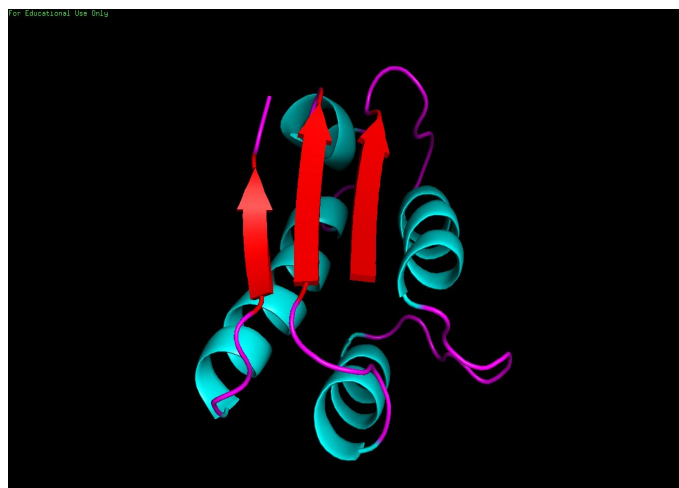
### **6. ÉCHELLES D'ÉTUDE**

Les fluctuations atomiques ont été étudiées à deux échelles spatiales et temporelles. Au niveau de la protéine, les analyses réalisées au niveau des conformations permettent de les comparer de manière globale. Celles faites au niveau des résidus permettent d'identifier les positions atypiques du point de vue des caractères étudiés. La dynamique moléculaire a également été réalisée pour des intervalles de temps différents (10 et 100 ns), ce qui permet de comparer l'information contenue pour ces résolutions.

## RÉSULTATS

### PYMOL

Résidus	Structure II
1-6	Brin $\beta$
13-24	Hélice $\alpha$
34-43	Hélice $\alpha$
48-54	Brin $\beta$
56-61	Hélice $\alpha$
66-80	Hélice $\alpha$
84-87	Brin $\beta$



**Table 1.** Tableau récapitulatif des enchaînements d'acides aminés impliqués dans des structures secondaires et type de structures II associé.

**Figure 1.** Représentation tridimensionnelle de la barstar. Les hélices  $\alpha$  en bleu, les brins  $\beta$  en rouge et les boucles en violet.

Parmi les 89 résidus qui composent la protéine, 53 sont inclus dans des structures II de type hélice  $\alpha$  ou brin  $\beta$ . Les trois brins  $\beta$  sont associés en un feuillet parallèle.

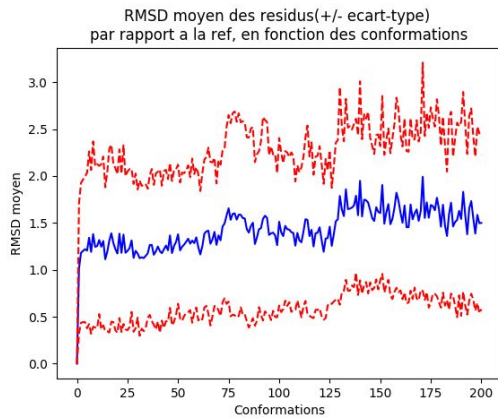
### 1. ÉCHELLE GLOBALE

#### RMSD

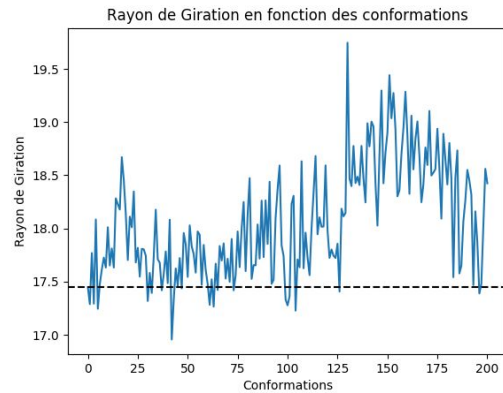
La conformation 0 a un RMSD moyen de 0. Elle correspond à la structure de référence. Elle est parfaitement alignée avec elle-même. Les conformations 0 à 75 présentent des RMSD inférieurs à 1.5, celles 75 à 125 de l'ordre de 1.5 et de 125 à 200, elles ont un RMSD d'environ 1.7 (**figure 2**). Globalement, le RMSD augmente avec le temps. Les structures sont de plus en plus divergentes de celle de référence (**figure 2**).

#### Rayon de Giration

Deux régimes se distinguent (**figure 3**). Les conformations 0 à 125 présentent un rayon de giration relativement petit par rapport aux suivantes. Jusqu'aux conformations 180, il est globalement 8 % plus grand, bien que de nombreuses fluctuations sont observables. Les conformations 180 à 200 ont un rayon de giration comparable à la conformation de référence.



**Figure 2.** RMSD moyen en bleu et son écart-type en rouge.

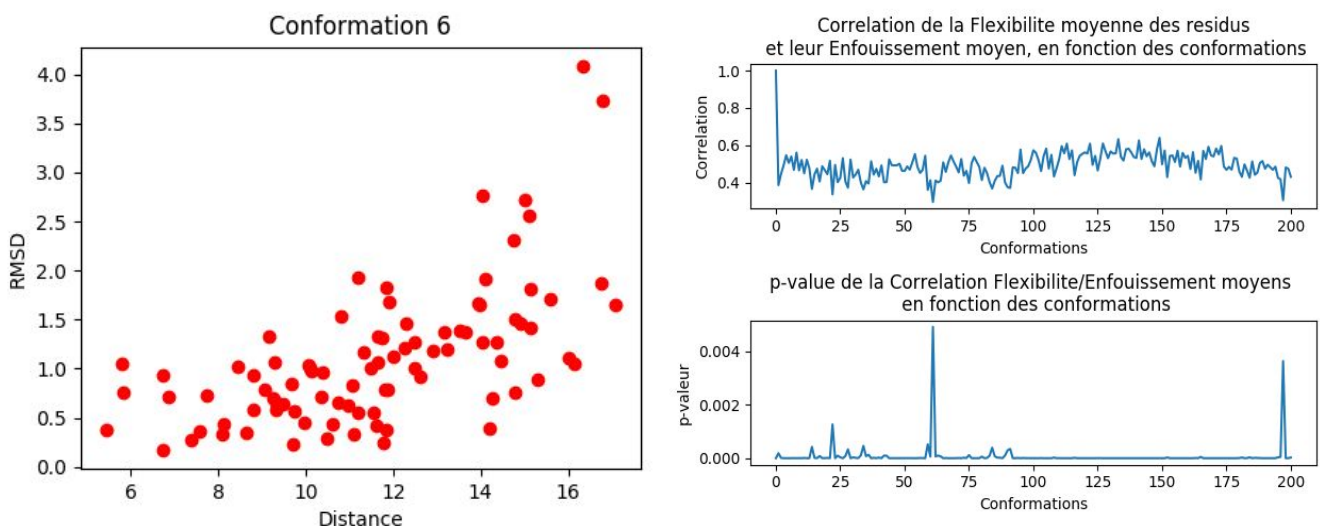


**Figure 3.** Rayon de Giration des conformations en bleu. Rayon de Giration de la référence en pointillé noir.

### Corrélation RMSD-Distance

Il y a globalement une corrélation de 0.4 pour l'ensemble des conformations (**Figure 4b**). La valeur du RMSD augmente avec la distance. Plus le résidu est éloigné du centre de masse de la protéine, plus il est mobile. Il y a un lien négatif entre l'enfouissement d'un résidu et sa flexibilité.

Toutefois, la corrélation de Pearson est basée sur une relation linéaire entre deux métriques. Or, la **figure 4a** laisse supposer une relation de type exponentielle, ce qui explique les valeurs peu élevées de corrélation, bien que constantes.



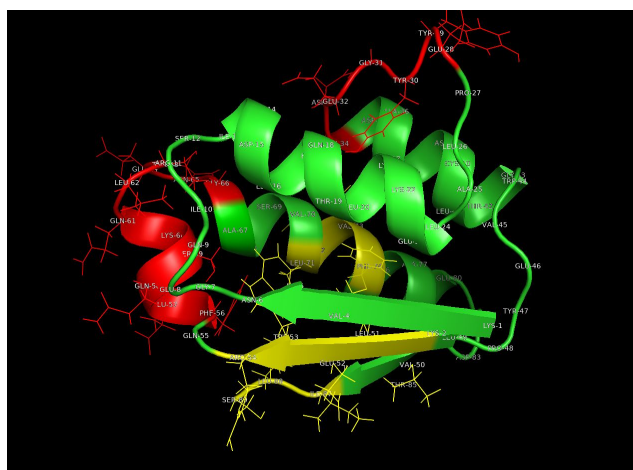
**Figure 4.** (gauche) RMSD en fonction de la distance pour la conformation 6. (haut droit) Corrélation entre RMSD et distance pour chaque conformation et (bas droit) p-valeur associées (pearson).

## 2. ÉCHELLE LOCALE

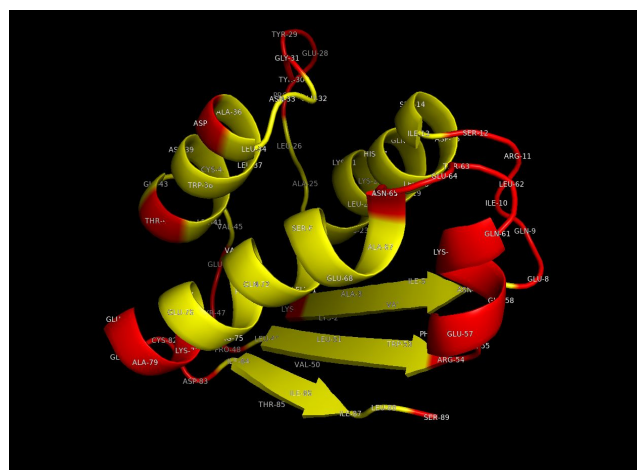
### RMSD

Deux régions se distinguent des autres par des valeurs de RMSD relatives beaucoup plus importantes. Il s'agit des résidus 28 à 34 (boucle en surface) et 56 à 66 (boucle+hélice en surface) (**Figure 6**). Au contraire, d'autres régions présentent des successions d'acides aminés aux RMSD moyens relativement faibles (50-54, 71-74 et 87-89), ces régions ont également une distance au centre de masse plus faible.

De plus, la région 56-66 présente les écarts-types les plus importants. Donc d'une conformation à une autre, les positions de ces résidus fluctuent de manière plus importante (**Figure 5**).



**Figure 5.** Représentation 3D de la barstar. Les zones en rouge correspondent aux résidus associés à des grandes valeurs relatives de RMSD et ceux en jaune aux valeurs les plus faibles.



**Figure 6.** Représentation 3D de la barstar. Les zones en rouge sont celles les plus éloignées du centre de masse de la conformation (distance relative grande) et celles en jaune sont les plus proches (distance relative faible).

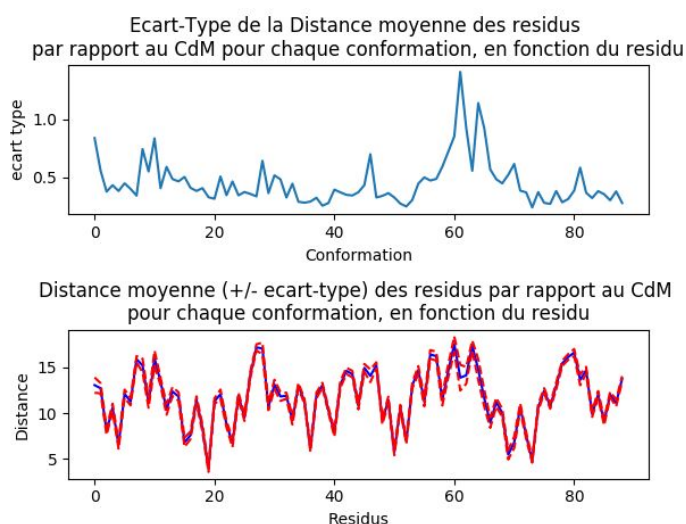
### Distance

Les résidus peuvent être classés dans deux catégories suivant leur distance moyenne au centre de masse. La valeur seuil a été fixée à 12Å et la classification en résidus en distance faible et élevée est présentée dans la **Table 2**. A l'exception des résidus 54 à 66 qui se situent dans une hélice  $\alpha$  éloignée du centre de masse, les résidus caractérisés par une distance faible au centre de masse sont présents dans des structures II.

De plus, l'écart-type associé à la mesure de distance est le plus grand pour les résidus 54 à 66 (**figure 7**). L'état d'enfouissement de ces résidus est très variable d'une conformation à une autre.

Distance faible	2-7, 13-26, 32-34, 36-41, 45, 49-53, 67-77, 84-88
Distance élevée	1, 8-12, 28-31, 35, 42-44, 46-48, 54-66, 78-83, 89

**Table 2.** Résidus situés à faible ou grande distance du centre de masse de la protéine. Seuil calculé à partir de la distance moyenne des résidus au centre de masse = 11.7Å (+/- 0.08).



**Figure 7.** Écart-type de la distance moyenne des résidus par rapport au centre de masse, en fonction du résidu.

**Figure 8.** Distance moyenne (bleu) +/- écart-type (rouge) des résidus par rapport au centre de masse, en fonction du résidu.

Les zones proches du centre de masse des conformations sont caractérisées par une composition riche en leucine, isoleucine, valine et alanine qui sont des acides aminés hydrophobes (environ 60% de résidus hydrophobes). Les zones moins enfouies au coeur de la protéine sont composées à 30% de résidus hydrophobes. Les compositions en acides aminés sont décrites dans la **Table 3**.

### Corrélation RMSD-Distance

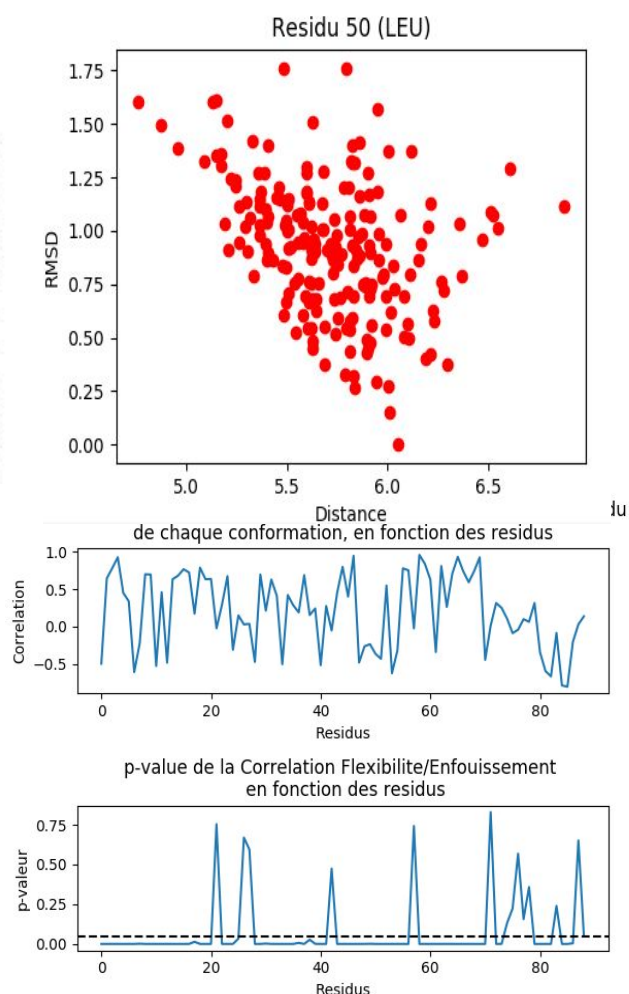
La corrélation entre le RMSD d'un résidu et sa distance au centre de la protéine est très variable (**Figure 9b**). Il n'y a pas de tendance claire entre ces deux métriques pour un acide aminé donné. En effet, la distribution des couples distance-RMSD pour un résidu (**Figure 9a**) n'affiche pas de tendance linéaire et présente une forme en entonnoir. Lorsque le RMSD est égal à 0, la distance entre le résidu et le centre de masse est celle de la conformation de référence. Toutes les conformations pour lesquelles la distance est inférieure à cette valeur (environ 7,4 Å) sont dans un état plus compact pour ce résidu que la conformation de référence. Inversement, une plus grande distance marque une dilatation. De ce fait, contrairement à l'échelle globale, le lien entre flexibilité (RMSD) et enfouissement (distance) n'est pas direct.

### 3. COMPARAISON 10 ns – 100 ns

Globalement, les trajectoires observées sont très similaires (**figure 10**). Bien que les simulations de résolutions 10 ns présentent de nombreuses fluctuations additionnelles, il n'y a pas de transition supplémentaire observée, ce qui laisse penser que le niveau de détail obtenu en 100 ns est suffisant pour étudier la dynamique de la protéine.

De façon cohérente, sur l'échantillon de 2000 conformations la p-value des tests de corrélation est plus souvent sous le seuil 5%.

	Nature	Type	% d faible	% d grande	% hélice $\alpha$
ALA	hydrophobe		9,8	2,7	0
ARG	hydrophile	charge +	2	5,4	0
ASN	hydrophile	polaire	3,9	2,7	9,1
ASP	hydrophile	charge -	3,9	5,4	0
CYS	hydrophobe		2	2,7	0
GLN	hydrophile	polaire	3,9	10,8	18,2
GLU	hydrophile	charge -	9,8	16,2	18,2
GLY			2	10,8	9,1
HIS	hydrophile	charge +	2	0	0
ILE	hydrophobe		9,8	0	0
LEU	hydrophobe		21,6	5,4	9,1
LYS	hydrophile	charge positive	5,9	8,1	9,1
PHE	hydrophobe		2	2,7	9,1
PRO	hydrophobe		0	5,4	0
SER	hydrophile	polaire	3,9	8,1	9,1
THR	hydrophile	polaire	3,9	2,7	9,1
TRP	hydrophobe		3,9	2,7	0
TYR	hydrophobe		0	8,1	0
VAL	hydrophobe		9,8	0	0



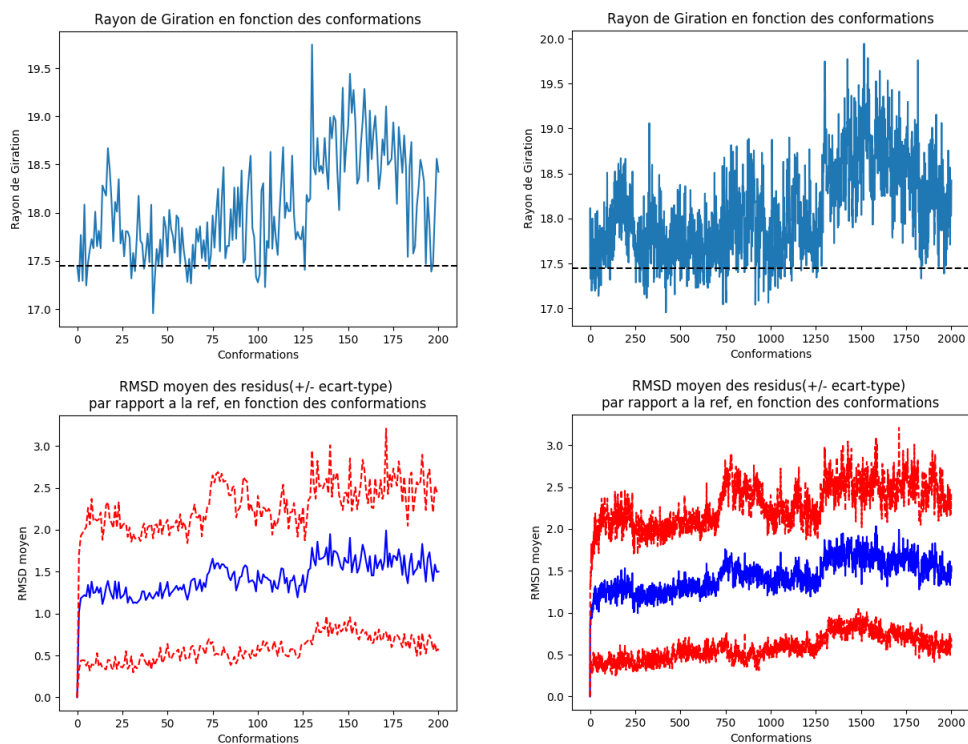
**Table 3.** Composition en acides aminés des différentes zones identifiées selon la distance au centre de la protéine.

**Figure 9.** (a) RMSD en fonction de la distance pour la Leucine 50. (b) Corrélation entre RMSD et distance pour chaque résidu et (c) p-valeur associées (pearson).

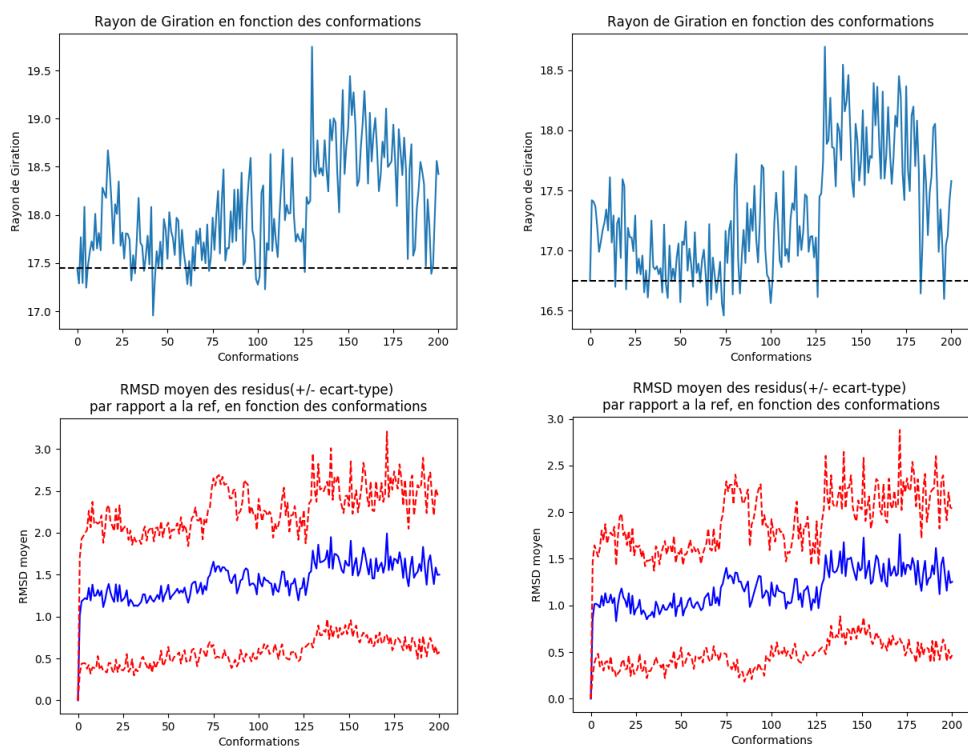
#### 4. COMPARAISON CA-ALL

Le rayon de giration et le RMSD sont plus élevés en calculant le centre de masse des résidus qu'en considérant leur carbone  $\alpha$  (figure 11). Ce qui est cohérent car selon la chaîne latérale du résidu considéré, le centre de masse peut être très éloigné du carbone  $\alpha$ . Cependant les tendances observées sont très semblables.





**Figure 10.** Comparaison des graphes obtenus en fonction du nombre de conformations considérées : 200 conformations (gauche) ou 2000 conformations (droite).



**Figure 11.** Comparaison des graphes obtenus avec le calcul du centre de masse des résidus (gauche) ou en considérant simplement la position du Carbone  $\alpha$  (droite).



## DISCUSSION

---

Les protéines peuvent présenter une grande diversité de conformations. Au sein des structures, les hélices  $\alpha$  et les feuillets  $\beta$ , deux motifs très ordonnés, peuvent être distingués. Les principales échelles de temps rencontrées en dynamique moléculaire dans des systèmes biologiques couvrent une gamme de 16 ordres grandeurs (de la flexion de liaison  $10^{-14}$  au repliement de protéine  $10^2$ s) (Poulain, 2006). Dans le cas de la barstar, l'information contenue dans les simulations de dynamique moléculaire 100 ns semble comparable à celle à 10 ns. De plus, approximer la position d'un résidu par celle de son carbone  $\alpha$  n'impacte pas les tendances observées.

Le RMSD permet de suivre les phénomènes de flexion et leur amplitude. Couplé aux mesures de distance, il informe sur l'état plus ou moins compact des conformations. Le rayon de giration varie dans le même sens. En effet, si une conformation est plus dilatée, la distance maximale entre son résidu le plus éloigné et son centre va globalement augmenter.

Globalement, le rayon de giration et le RMSD varient de manière simultanée. Les conformations présentant des distances résidus/centre de masse les plus élevées sont celles qui ont les plus grandes valeurs de RMSD et de rayon de giration. Ces conformations, allant de la 125 à la 200, sont donc les conformations les moins compactes par rapport à la référence. De plus, le RMSD et rayon de giration de ces conformations sont éloignés de ceux de la référence. Il est ici bien mis en évidence qu'en solution la barstar subit d'importantes variations conformationnelles impactant l'ensemble de sa structure.

Au niveau local, les résidus appartenant aux structures II sont en général situés à des distances faibles du centre de masse de la conformation. Ce sont donc majoritairement des résidus enfouis. De plus, leurs valeurs de RMSD sont faibles, ce qui implique peu de flexibilité. Ces résidus, majoritairement hydrophobes, sont au coeur de la structure et conditionnent le repliement.

La zone comprenant les résidus 56 à 66 se distingue de ce cas général. Les résidus forment une hélice  $\alpha$  située en périphérie de la molécule. Les résidus qui la constituent sont largement polaires et chargés permettant la mise en place d'interactions de type ligand-récepteur. De plus, c'est une zone hautement flexible. Les fluctuations atomiques pourraient conditionner les interactions de la protéine.

L'analyse des changements conformationnels de la barstar a permis d'identifier trois types de conformations. Une étude précédente de Kimura et al. (2001) en mettait en évidence deux. Ils ont étudié en particulier un résidu (ASP39), supposé comme jouant un rôle dans l'interaction barnase-barstar. Or, s'agissant d'un résidu enfoui, les contraintes de mouvement peuvent être plus importantes que celles appliquées en moyenne sur l'ensemble des résidus.

## CONCLUSION

---

Les mouvements atomiques engendrent des différences conformationnelles. Les variations de structure de la barstar ont pu être mises en relation avec des changements de flexibilité. Le cœur de la protéine présente globalement une structure stable, hormis certains résidus enfouis flexibles a priori impliqués dans les interactions de la barstar, contrairement à la zone périphérique. Ce sont les mouvements atomiques de surface qui impactent majoritairement la structure et qui sont à l'origine des fluctuations de compacité de la protéine. Ces régions plus flexibles peuvent moduler l'accessibilité du substrat. En particulier, des régions ne présentant pas de structure II ont été mises en évidence comme zones de plus faible stabilité, ainsi qu'une région correspondant à une hélice  $\alpha$  identifiée comme point clé de la fluctuation structurale de la barstar.

## BIBLIOGRAPHIE

---

Freddolino, P. L., Liu, F., Gruebele, M., & Schulten, K. (2008). Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophysical journal*, 94(10), L75-L77.

Gohlke, H., Kuhn, L. A. and Case, D. A. (2004), Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins*, 56: 322–337. doi:10.1002/prot.20116

Kimura, S. R., Brower, R. C., Vajda, S., & Camacho, C. J. (2001). Dynamical view of the positions of key side chains in protein-protein recognition. *Biophysical journal*, 80(2), 635-642.

Lindahl, E. R. (2008). Molecular dynamics simulations. *Molecular modeling of proteins*, 3-23.

Poulain, P. (2006). Structure et dynamique de protéines isolées : approches statistiques. Bio-physique [physics.bio-ph]. Université Claude Bernard - Lyon I.