

Compte rendu Projet Barstar

Team : Les Mugiwaras

Auteur : ARBES Hugo et HENRIQUES Adrien

M1 BIBS

Introduction

La Barstar est une protéine synthétisée par une bactérie (*Bacillus amyloliquefaciens*) afin d'inhiber l'activité de la ribonucléase de la Barnase en formant un complexe avec celle-ci. En formant ce complexe, la Barstar couvre le site actif de la Barnase empêchant ainsi la Barnase de dégrader l'ARN dans le cytoplasme de la cellule.

Dans cette étude, nous allons nous intéresser à la protéine Barstar et notamment à ses changements conformationnel lors de l'interaction Barstar-Barnase. En effet, la majorité des algorithmes de docking considèrent que les deux protéines testées sont rigides et ne subissent donc pas de changement conformationnels. Le but est donc de créer un algorithme qui va nous permettre d'étudier la stabilité de la structure de la protéine Barstar au cours du temps ainsi que ses changements conformationnels. Pour cela, nous allons donc dans un premier temps s'intéresser aux changements conformationnels globaux, puis nous nous intéresserons aux changements conformationnels locaux.

PEUT ETRE DIRE CE QU'ON POURRA VOIR DANS LES GLOBAUX ET LOCAUX

Manuel d'utilisation

Dans cette partie, nous allons aborder l'utilisation du programme et principalement des fichiers d'entrée ainsi que des pré-requis nécessaires pour pouvoir exécuter le code.

Pré-requis

Le programme est composé de plusieurs fonctions pour calculer, afficher, récupérer des données. Ce programme est créé en langage Python (Python 2.7), pour que celui-ci puisse être compris par l'ordinateur, il est nécessaire d'avoir installé Python. Pour installer Python il faut ouvrir un terminal et selon votre debian/système d'exploitation, il faut alors taper :

-Unix : `$ sudo apt-get install python`

Si vous êtes sous Windows, il vous faudra alors télécharger la dernière version de python et l'installer (<https://www.python.org/downloads/>).

Une fois installé, certaines de ses fonctions utilisent des fonctions graphiques telles que Matplotlib. Pour pouvoir les utiliser il est donc nécessaire d'avoir installé Matplotlib. Pour installer Matplotlib, cela dépend une fois encore du système d'exploitation. Pour cela, ouvrez un terminal et tapez :

- Unix : `$ sudo apt-get install python-matplotlib`
- Fedora : `$ sudo yum install python-matplotlib`
- Windows : `python -m pip install --U steuptools`
`python -m pip install matplotlib`

Ainsi pour exécuter le code, il nous suffira de taper la commande :

`$ Python BarStar.py`

qui exécutera le code en appelant les différentes fonctions.

Fichiers d'entrée

Les fichiers qui vous seront demandés lors de l'exécution du programme doivent absolument suivre une structure spéciale. En effet, ceux-ci doivent être au format PDB suivant :

Colonne	Data Type	Field	Definition
1 - 6	Strings	"ATOM" ou "MODEL"	Type de données de la ligne
7 - 11	Integers	Atome	Identifiant de l'atome
13-16	Strings	Atome	Type d'atome
17 - 20	Strings	Type résidu	Type du résidu
21 - 26	Integers	Résidu	N° du résidu
32 - 38	Integers	Coordonnées	Position X
40 - 46	Integers	Coordonnées	Position Y
48 - 54	Integers	Coordonnées	Position Z
5 - 14	Integers	Model	N° du modèle

Si les fichiers d'entrée ne suivent pas ce format, le processus de parsing ne se fera pas correctement et le programme risque d'avoir des erreurs ou de ne pas donner les bonnes sorties.

De plus il est nécessaire de suivre les indications qui s'affichent dans le terminal et notamment lorsque l'on écrit les fichiers, il est nécessaire de respecter l'ordre d'entrée des fichiers. Tout d'abord on entre le fichier qui contiendra la Barstar ayant la conformation référent, puis dans un deuxième temps, le fichier contenant les différentes conformations de la Barstar dans le temps.

Les tests

Le programme a été testé sous plusieurs angles. En effet, nous l'avons testé en entrant mal un des 2 fichiers pour tester si la reconnaissance d'erreur est détectée. De plus, nous avons testé les parties plus calculatoires (RMSD, Rayon de giration et calcul de distance) sur des jeux de données restreint pour pouvoir vérifier les résultats et ainsi la robustesse des calculs. Et enfin nous avons testé ce programme avec les données fournies afin de voir si celui-ci fonctionne.

PEUT ETRE PARLER DES SORTIES

Calcule du rayon de giration et des distances

Calcul de RMSD

Dans cette partie nous allons détailler le calcul de la RMSD afin de voir comment celui-ci est obtenu et à quoi peut-il servir.

La RMSD (Root Mean Square Deviation) correspond à la racine de la moyenne des distances élevées au carré. En effet, nous disposons de la Barstar de référence ainsi que d'un autre fichier avec les

différentes conformations de la Barstar, dans lesquels les séquences ont été alignées. Le principe consiste donc à comparer la Barstar de référence à ses conformations en les superposant et à calculer la distance entre toutes les paires d'atomes.

Calcul de RMSD globale

La RMSD globale d'une protéine correspond à la distance spatiale globale entre les 2 protéines. Pour la calculer nous avons donc pour chaque pair d'atome calculé la RMSD pour ensuite additionner cette RMSD de chaque pair d'atomes pour avoir la RMSD de chaque résidu. Ensuite ces RMSD de chaque résidu sont additionnés pour avoir la RMSD de chaque conformation.

$$RMSD_{glob} = \sum RMSD_{residu} = \sum \sum RMSD_{pairatome}$$

$$\text{Avec } RMSD_{pairatome} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta^2 i}$$

Avec N le nombre de pair d'atome a chaque résidu.

La RMSD obtenu ainsi nous permet de voir quelle conformation s'écarte beaucoup de la conformation de référence et aussi pouvoir déterminer si cette protéine est stable car si la RMSD change beaucoup, cela signifierait que la distance entre la Barstar de référence et la conformation change et donc que la protéine est instable.

Pour faciliter l'analyse, nous avons intégré une fonction graphique pour pouvoir observer la RMSD globale en fonction des conformations et donc en fonction du temps (car les conformations possèdent des identifiants et sont rangées dans l'ordre croissant) (voir figure 1).

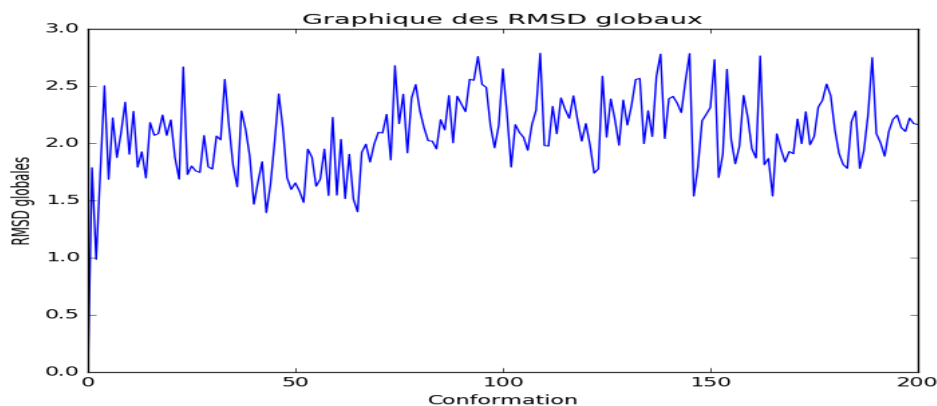


Figure1 : Exemple de graphique RMSD globaux

Nous pouvons ainsi avoir une vision globale des RMSD obtenue et pouvoir interpréter plus facilement. Nous avons aussi fait une fonction pour récupérer les données générées après l'analyse globale, dans un fichier texte pour pouvoir avoir une analyse plus approfondie.

Calcul de RMSD locales.

La RMSD locale d'une protéine correspond quand a elle aux RMSD moyennes à chaque résidu. Pour la calculer, il nous faut donc calculer la RMSD pour chaque pairs d'atomes puis sommer la RMSD a

chaque pair d'atomes pour avoir la RMSD a chaque résidu. Ainsi nous obtenons la RMSD pour chaque conformation de chaque résidu et ensuite nous avons additionné ces RMSD à chaque conformation pour obtenir la somme de toutes les RMSD de chaque position. Enfin nous avons divisé chaque RMSD par le nombre de conformations pour obtenu une RMSD moyenne a chaque positions.

$$RMSD_{locale} = \frac{RMSD_{residu}}{nb \text{ de conformation}}$$

Avec $RMSD_{residu}(j, i) = \sum_{j=0}^{Nbres} RMSD_{pairatome}(i)$

Avec Nbres correspondant au nombre de résidu dans les conformations.

FORMULE UN PEU BANCALE JE PENSE

La RMSD ainsi obtenu nous permet de voir la variation moyenne de RMSD dans la protéine. Ainsi nous pourrions déterminer quelle partie de la protéine est mieux conservé et quelles parties ne sont au contraire pas conservées. Pour faciliter l'analyse de cette partie, nous avons introduit une fonction graphique nous permettant d'observer la variation d'RMSD moyenne en fonction des positions et pouvoir ainsi voir visuellement cette variation (voir figure 2). Nous avons aussi créé une fonction nous permettant de récupérer les données générées de l'analyse locale dans un fichier texte pour ainsi avoir une analyse plus approfondie.

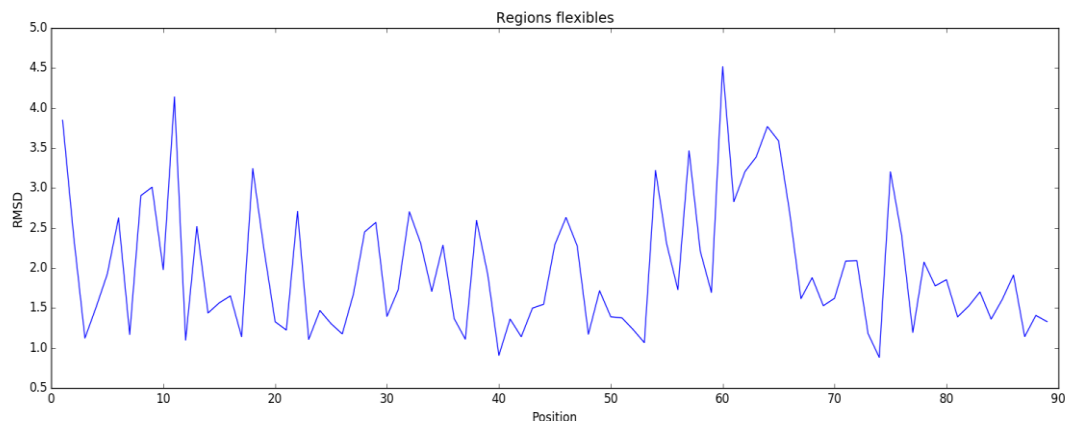


Figure 2 : Exemple de graphique RMSD local

Résultats biologiques et interprétation

Grace à l'exécution de ce programme, nous avons obtenues plusieurs résultats que nous allons analyser au cours de cette partie.

Résultats globaux

Tout d'abord analysons la sortie de RMSD globales. Tout d'abord, afin d'avoir un aperçu générale de la RMSD, nous pouvons regarder le graphiques des RMSD globales par rapport aux conformations. Nous pouvons voir que celle-ci est comprise majoritairement entre 1.4 à 3.5 A selon le fichier, avec énormément de fluctuation entre les conformations. Nous pouvons aussi voir que seule la première conformation à une RMSD très basse et donc cette conformation est très proche de la protéine de

référence. Cela montre donc que la protéine subit bien des changements conformationnels au cours du temps.

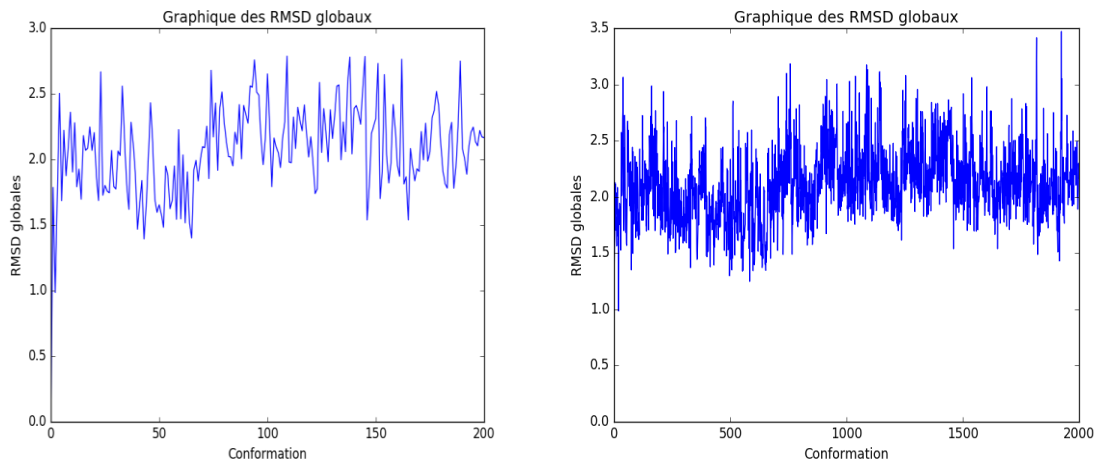


Figure3 : Graphiques des RMSD Globales en fonction des conformations (1) avec le fichier md_prot_only_skip100 (2) avec le fichier md_prot_only_skip10

Ces changements peuvent être dus à des mouvements de boucles d'assez grande amplitude pour provoquer des différences de distances spatiales entre les protéines et donc une augmentation de la RMSD. Cela est visible quelque soit le nombre de conformations et en augmentant le nombre de conformations, nous n'arrivons donc pas à une RMSD globale stable qui pourrait montrer conformation stable après un certain temps. Or nous ne distinguons pas de tendance ni de valeurs stable dans ces valeurs de RMSD mais nous pouvons tout de même remarquer que les fluctuations semblent diminuer entre les conformations 1900 et 2000 vers une RMSD de 2.1 Å.

Pour vérifier à quoi sont dus ces mouvements, nous allons donc regarder les résultats de rayon de giration des conformations afin de déterminer si ces mouvements sont dus à un phénomène de compaction ou de dépliement de la protéine.

En observant ces résultats, on remarque que le rayon de giration varie énormément entre 17 et 20 Å. Le rayon de giration correspond à la distance entre le résidu le plus éloigné et le centre de la protéine. Sachant cela, comme nous observons beaucoup de variation, cela signifie que le résidu le plus éloigné s'éloigne et se rapproche du centre de la protéine. Cela décrit à la fois un mouvement de compaction mais aussi un mouvement de dépliement de la protéine. Lorsque le rayon de giration augmente, le mouvement est un mouvement de dépliement, alors que lorsque le rayon de giration diminue, le mouvement est un mouvement de compaction. De plus, en observant les résultats sur les différents graphiques, il n'y a pas de valeur de rayon de giration stable même après 2000 conformations. Cela montre donc que la protéine est en perpétuel mouvement de repliement et de compaction.

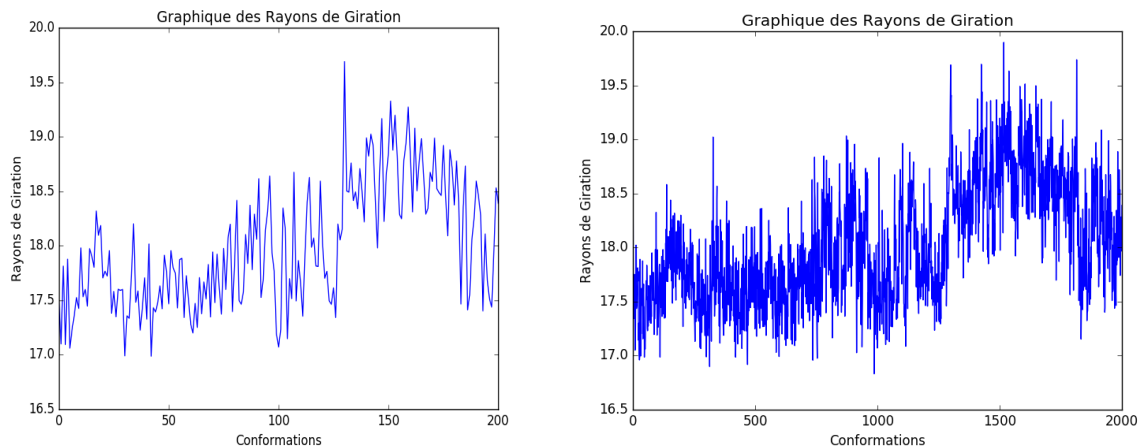


Figure4 : Graphiques des Rayons de giration en fonction des conformations (1) avec le fichier md_prot_only_skip100 (2) avec le fichier md_prot_only_skip10

Cela est cohérent avec les observations précédentes car du fait que la protéine soit en perpétuelle mouvement de compaction/dépliage, ce qui explique que la RMSD globale fluctue énormément aussi avec la compaction et le dépliement des boucles qui font alors un mouvement de grande amplitude.

Nous allons désormais observer les résultats locaux pour observer ces modifications au niveau des résidus afin de voir quels sont les résidus impliqués dans ces changements conformationnels mais aussi où sont placés ces résidus.

Résultats locaux

Dans cette partie, les résultats ont été obtenus grâce à des moyennes, en comparant les résultats avec les fichiers md_prot_only_skip100 et le fichier md_prot_only_skip10, nous nous apercevons que nous obtenons les mêmes résultats car les 200 premières conformations suffisent pour faire une moyenne et les 1800 autres conformations contenues dans le fichier md_prot_only_skip10 ne changent pas ces moyennes étant donné que les valeurs sont globalement similaires. Dans cette partie, nous allons donc prendre uniquement les données de md_prot_only_skip100 mais ces résultats et interprétations sont tout aussi bien applicables au fichier md_prot_only_skip10.

Dans un premier temps, observons les résultats de RMSD locaux obtenus en fonction du résidu. Nous pouvons ainsi remarquer que la RMSD varie entre 0.98 et 4.5 Å. Cependant nous pouvons observer des régions où les variations sont moins grandes (par exemple les résidus 48 à 53) alors qu'à d'autres endroits, la RMSD est plus grande (par exemple les résidus 61 à 68). C'est ainsi que nous avons pu déterminer que les régions suivantes subissent le plus de variation car leur RMSD est grande :

- 0 à 3 (ce qui est normal car c'est l'extrémité C-ter)
- 9 à 12 : succession de pics
- 27 à 39 : succession de pics
- 43 à 46 : large pic
- 54 à 72 : large succession de pics

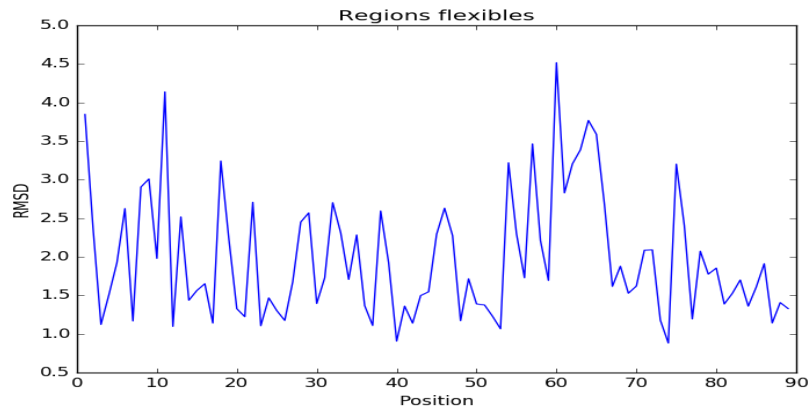


Figure5 : Graphiques des RMSD locales en fonction des résidus avec le fichier md_prot_only_skip100

Nous avons donc déterminé les régions qui semblent plus flexibles que d'autres et désormais nous allons observer l'enfouissement de ces régions afin de déterminer si elles sont enfouies ou non dans la protéine.

Pour cela, nous disposons des résultats de calcul de l'éloignement du centre en fonction des résidus. Plus la valeur de l'éloignement est grande et moins le résidu est enfouie dans la protéine et donc plus il est en surface.

Dans ce graphique nous pouvons observer des fluctuations, ce qui est normal car la protéine est repliée sur elle-même et donc certains résidus sont enfouies et d'autres en surface. Ce qui nous intéresse donc est de savoir si les régions déterminées avant sont enfouies. Nous regardons donc les résidus :

- 9 à 12 : la valeur de l'éloignement varie entre 11 et 16 ce qui est plutôt élevé, donc la région est en surface
- 27 à 39 : les valeurs varient entre 11 et 17 donc la région est aussi en surface
- 43 à 46 : les valeurs varient entre 12 et 15 donc la région est en surface
- 54 à 72 : les valeurs varient entre 6 et 17 il y a donc une partie en surface et une partie enfouie.

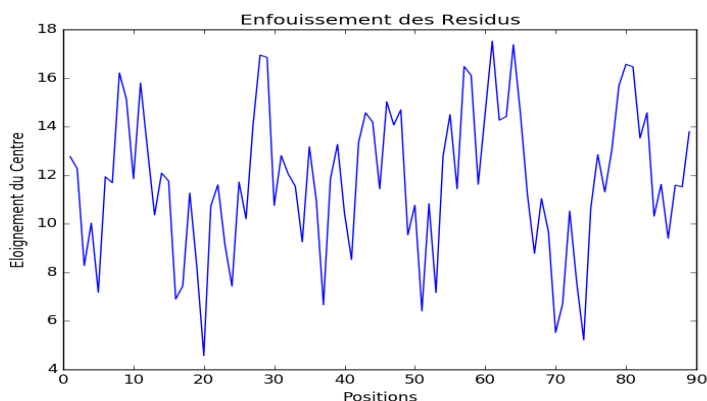


Figure5 : Graphiques de l'enfouissement du centre en fonction des résidus avec le fichier md_prot_only_skip100

Nous avons donc vu que la plus part des résidus qui subissent le plus de changement sont situés en surface mais qu'il y a aussi des parties plus internes qui changent mais cela reste une minorité.

Nous avons donc vu que la protéine Barstar subit des changements conformationnels au cours du temps et ces changements ont lieu pour la majeure partie sur des résidus de surface mais il arrive que certains résidus enfouis subissent des changements.

Pour appuyer ces résultats, nous avons modélisés cette protéine via le logiciel PyMol, ainsi que marquer les régions précédemment retenues comme étant les régions à forte flexibilités, et nous avons superposé des conformations de la Barstar au dessus de la protéine Barstar de référence.

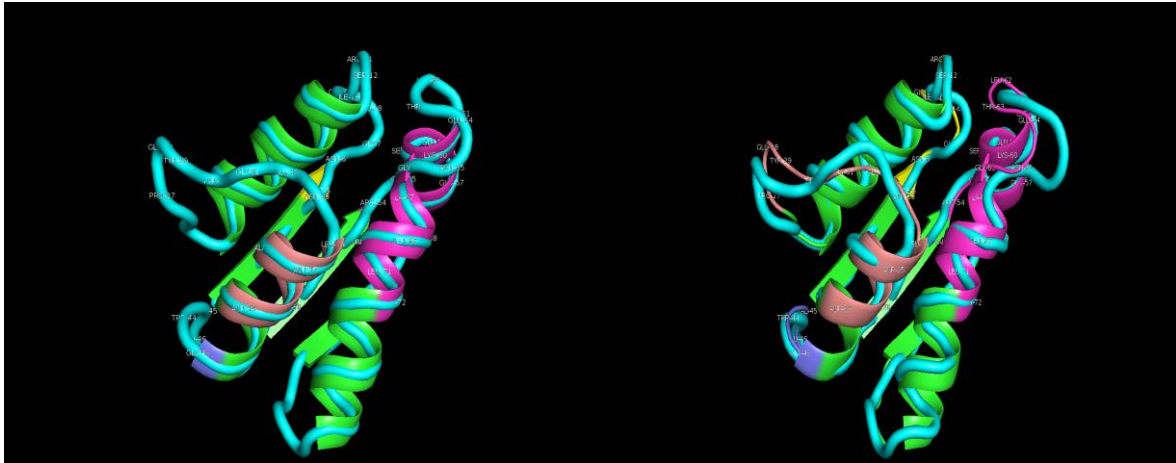


Figure6 : Photographie PyMol de la Barstar de référence (en vision structure 3D) et de l'une des ses conformation (structure bleu).

Nous pouvons donc grâce à ces photographies, voir que sur la photographie de droite, la conformation de la Barstar concorde parfaitement avec la structure de la Barstar, mais que après quelque temps, la conformation ayant changée, celle-ci ne concorde plus à certains endroits et principalement aux endroits colorées correspondant aux régions supposées très flexibles.

Conclusion