

Etude de la Barstar

Introduction

Contexte général :

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, une inhibition de la formation du complexe, ou un maintien de la conformation en complexe. Ces rôles sont en étroites relations avec la position, la flexibilité ainsi que la propriété physico-chimique des résidus.

Une étude focalisée 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 protéine, la barstar, et les changements conformationnels qu'elle subit. La barstar est impliquée dans un complexe protéique avec la barnase. Elle est synthétisée par des bactéries (*Bacillus*) et son but est d'inhiber l'activité de la barnase (qui est une ribonucléase). 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.

Dans ce projet, nous nous intéresserons à la dynamique moléculaire de la barstar. 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 allons 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.

Nous essaierons également d'identifier des résidus clés appartenant à l'interface de la barstar dans l'interaction avec la barnase.

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 du programme, un dossier s'appelant « PythonProgResults » contenant tous les fichiers de sortie sera créé dans le répertoire que l'utilisateur aura fourni.

Afin de réaliser une analyse globale et une analyse locale à partir d'un fichier pdb, plusieurs fonctions de bases 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 sortie est présenté ci-dessous :

{MODEL N°

```
{
  'chains': ['A'],
  'A': {
    'reslist': ['3'],
    '3': {
      'resname': 'ARG',
      'atomlist': ['N', 'N', 'CA', 'CA', 'C', 'C', 'O', 'O', 'CB', 'CB', 'CG', 'CG', 'CD', 'CD', 'NE', 'NE', 'CZ', 'CZ', 'NH1', 'NH2'],
      'C': {'y': 86.53, 'x': 12.759, 'z': 96.365, 'id': '37'},
      'CB': {'y': 85.746, 'x': 13.428, 'z': 93.98, 'id': '41'},
      'CA': {'y': 85.862, 'x': 12.333, 'z': 95.041, 'id': '35'},
      'CG': {'y': 85.172, 'x': 12.866, 'z': 92.651, 'id': '43'},
      'NE': {'y': 85.487, 'x': 12.644, 'z': 90.195, 'id': '47'},
      'O': {'y': 87.757, 'x': 12.924, 'z': 96.42, 'id': '39'},
      'N': {'y': 86.721, 'x': 11.296, 'z': 94.521, 'id': '33'},
      'CZ': {'y': 85.582, 'x': 13.114, 'z': 88.947, 'id': '49'},
      'CD': {'y': 85.886, 'x': 13.374, 'z': 91.406, 'id': '45'},
      'NH1': {'y': 86.056, 'x': 14.338, 'z': 88.706, 'id': '51'},
      'NH2': {'y': 84.308, 'x': 14.421, 'z': 88.373, 'id': '52'},
    }
  }
}
```

- « CM », « Distance », « RMSD » sont des fonctions de base pour le calcul du 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 :

$$CM = [x, y, z] = \left[\frac{\sum_i x_i}{n}, \frac{\sum_i y_i}{n}, \frac{\sum_i z_i}{n} \right]$$

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 :

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Le calcul de RMSD :

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

- « 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 :

```
{'residulist': ['LYS', 'LYS', 'ALA'],
 '1': [22.299166666666668, 33.490416666666666, 38.347916666666667],
 '3': [29.619999999999997, 32.418, 33.958999999999999],
 '2': [29.068181818181817, 36.72090909090909, 36.2109090909091],
 'prot': [26.265714285714292, 34.568035714285706, 36.724642857142854]}
```

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 » permet de récupérer les différentes mesures du temps et « graph » permet de tracer des graphiques

Analyse globale

RMSD global

On a créé une fonction « RMSDglobal » pour calculer le RMSD global de chaque conformation en la comparant avec la structure d'origine de la protéine. Le RMSD global permet 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 s'appelant « liste_delta ». Ensuite, le RMSD global est calculé.

En 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 :

dico_RMSD:

```
{'1': 0.8915903394122967, '0': 0.0, '2': 0.6756971708292576}
```

Rayon de giration

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

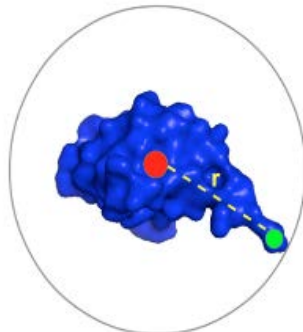


Figure 1. Rayon de giration

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. La fonction 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 :

dico_Giration|

```
{'1': 4.8500171792510764, '0': 4.988991523691877, '2': 5.088994088135849}
```

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 des distances pour chaque résidu correspond à l'enfouissement moyen. Pour chaque résidu, la moyenne de l'enfouissement est rangée dans un dictionnaire avec son numéro de résidus comme clé 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. Les 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 créé.

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

I. 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.

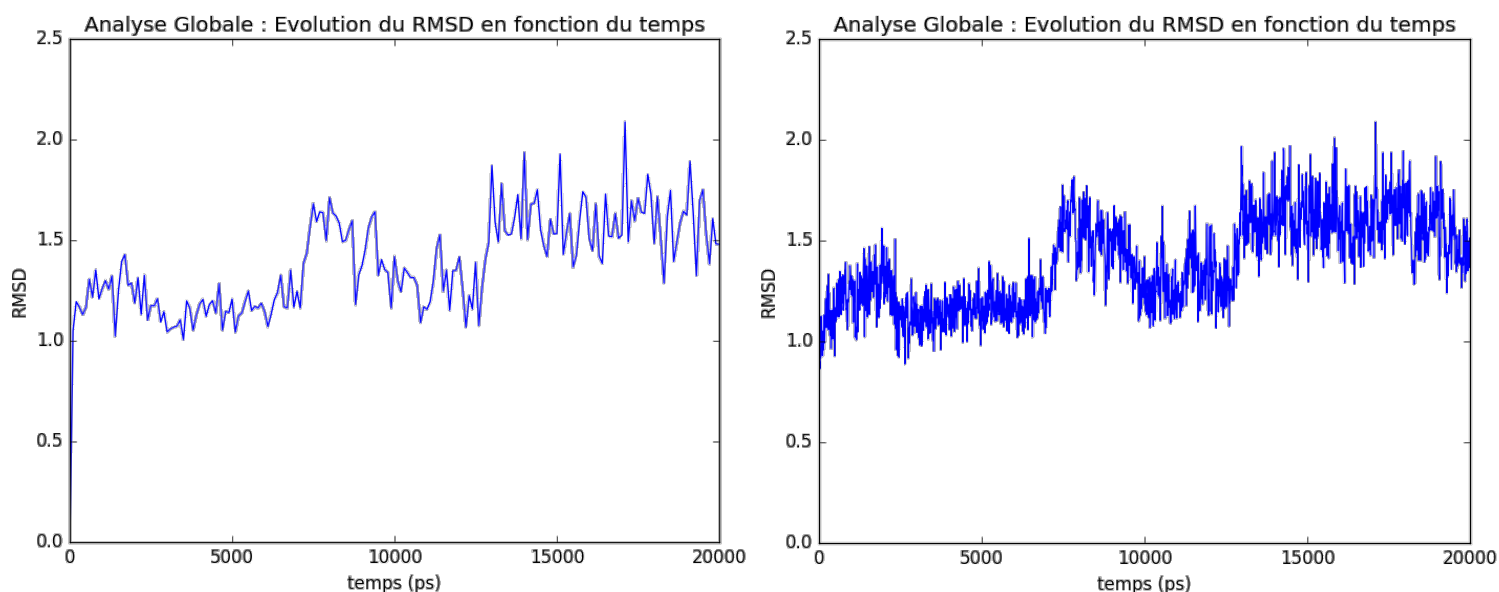


Figure 2. Représentation du RMSD en fonction du temps.

La figure à gauche représente la variation du RMSD (en Angstrom) de barstar au cours du temps (en ps) à partir des données extraites de toutes les 100 ps. La figure à droite représente la variation du RMSD (en Angstrom) de barstar au cours du temps (en ps) à partir des données extraites de toutes les 10 ps.

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.

D'après la figure 2, on se rend compte que le RMSD varie entre 0 et 2. On observe également une légère augmentation du RMSD au cours du temps.

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

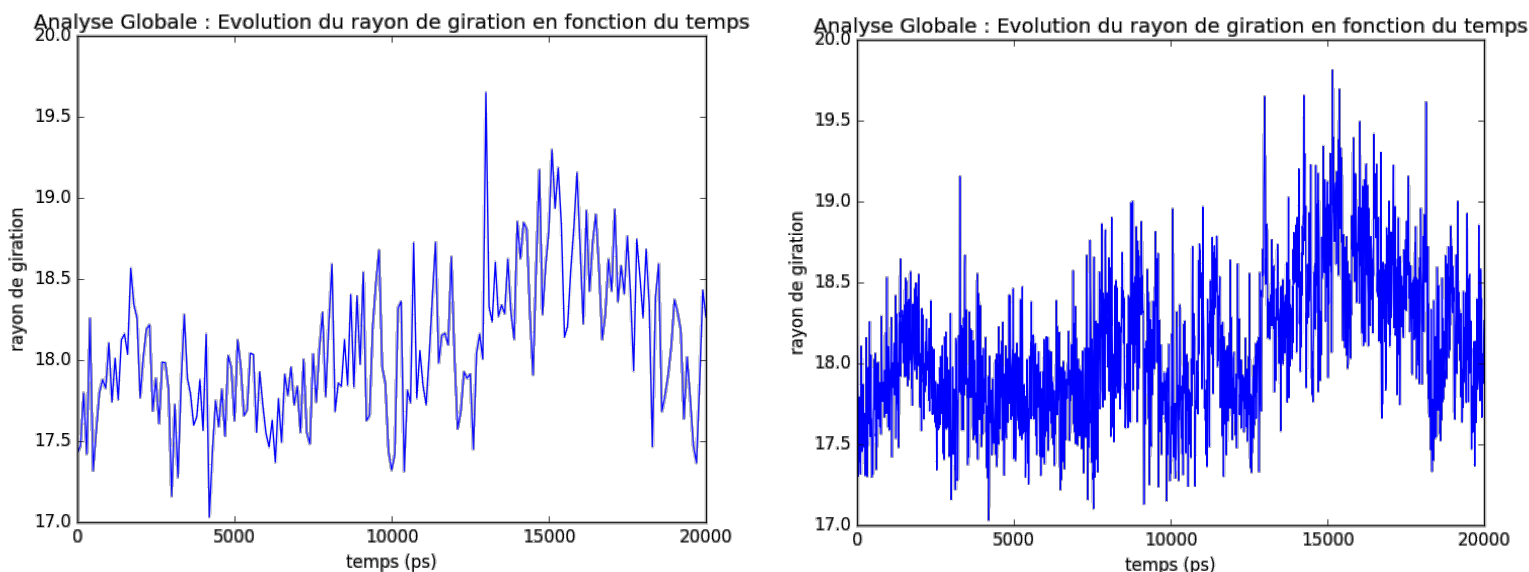


Figure 3. Représentation du rayon de giration en fonction du temps

La figure à gauche représente la variation du rayon de giration (en Angstrom) de barstar au cours du temps (en ps) à partir des données extraites de toutes les 100 ps. La figure à droite représente la variation du rayon de giration (en Angstrom) de barstar au cours du temps (en ps) à partir des données extraites de 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.

Pour les 2 mesures réalisées, nous avons choisis de représenter les variations sur 2 jeux de données différents (toutes les 10ps 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.

Analyse globale : Variation RMSD/Giration en fonction de la conformation

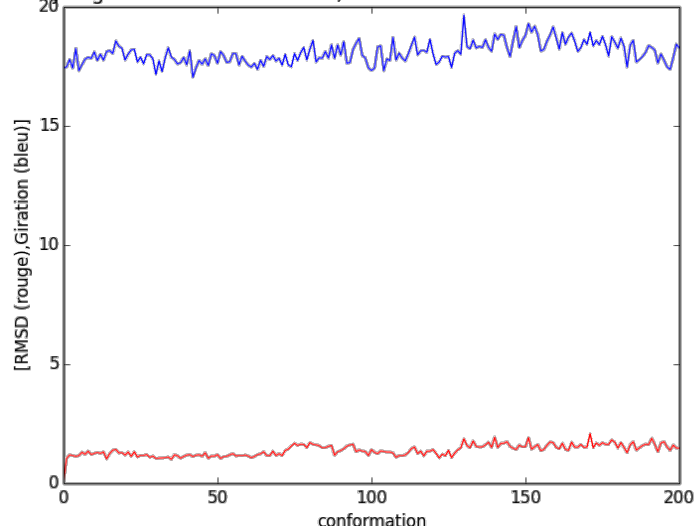


Figure 4. Représentation de la variation du RMSD/Giration en fonction de la conformation

La courbe en rouge correspond à la variation du RMSD. La courbe en bleu correspond à la variation du rayon de giration.

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).

I. 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. La valeur d'enfouissement calculée correspond à la moyenne de distance entre le centre de masse d'un résidu et celui de la protéine. Les résidus les plus enfouis sont ceux qui possèdent une faible valeur, cela signifie qu'ils sont plus proche du centre de masse de la protéine. Les résidus sont plus éloignés du centre de masse de la protéine donc ils ont une valeur élevée.

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 inférieures à 11 Å.
- Résidus de surface :
 - Ce sont les résidus qui ont des valeurs d'enfouissement supérieures à 11 Å. Dans le fichier de sortie, les résultats significatifs ont des valeurs proches de 13/14Å. On peut par exemple citer les résidus LYS 1, ARG 11, LYS60 et GLU 64.

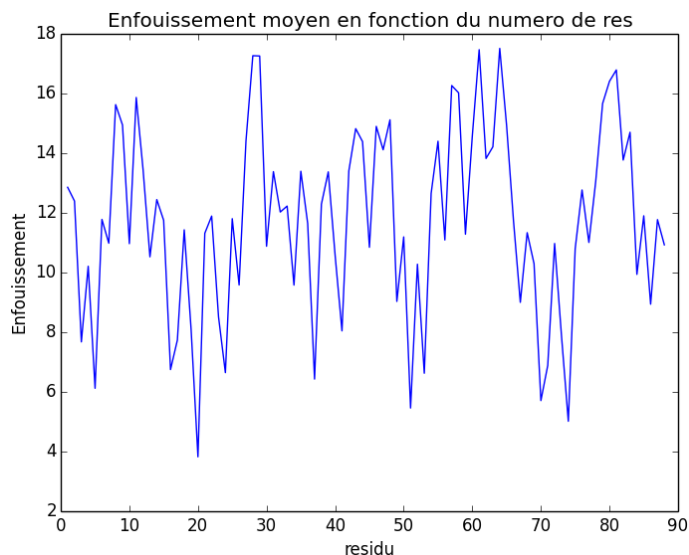


Figure 5. Représentation de la valeur d'enfouissement moyen de chaque résidu

L'enfouissement moyen est calculés pour chaque résidu. Les résidus sont identifiés par le numéro de résidu dans la séquence peptidique.

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 sont plus élevées que les autres : il s'agit par exemple des résidus LYS 1, ARG 11, LYS60 et GLU 64.

Ces résidus ayant des valeurs de RMSD les plus élevées par rapport aux autres résidus, ils sont donc particulièrement flexibles.

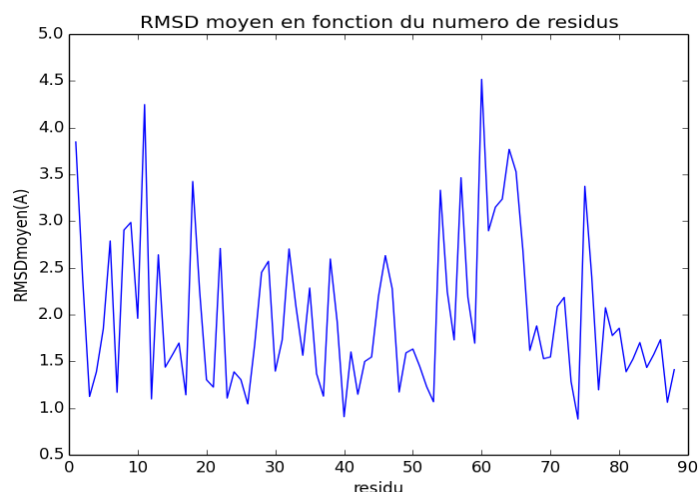


Figure 6. Représentation du RMSD moyen de chaque résidu

Le RMSD moyen (en Angstrom) est la moyenne des RMSD calculés pour chaque résidu sur l'ensemble des 200 conformations. Les résidus sont identifiés par le numéro de résidu dans la séquence peptidique.

On remarque que les résidus présents dans les régions flexibles sont également présents à la surface de la protéine. En effet, les résidus LYS 1, ARG 11, LYS60 et GLU 64 ont été identifié comme étant à la fois dans une région flexible et à la surface de la protéine. Cette observation est confirmée par la visualisation des résidus avec *Pymol* (figure 7).

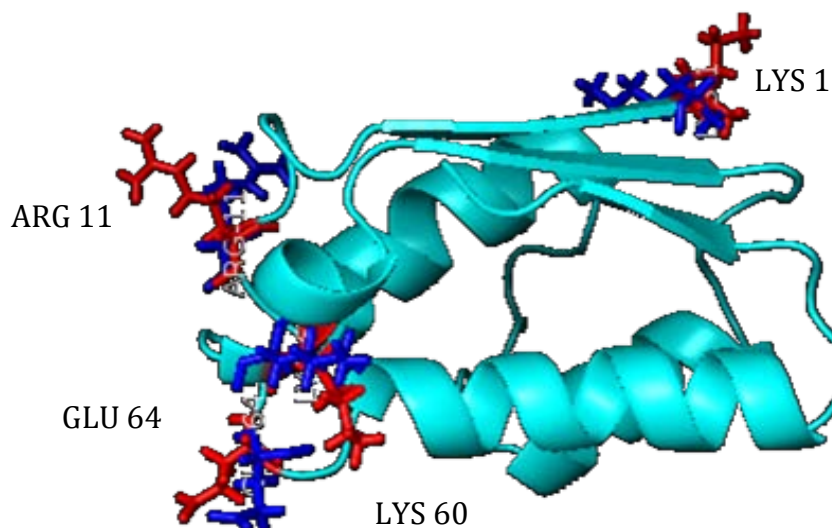


Figure 7. Représentation des résidus les plus flexibles Lys 1, Arg 11, Lys 60, Glu 64 sous Pymol

La protéine barstar est représentée en « cartoon », La chaîne latérale de Lys 1, de Arg 11, de Lys 60, de Glu 64 de la structure d'origine (en rouge) et de la protéine après un changement de conformation (model 200 dans le fichier md_prot_only_skip100.pdb) (en bleu) sont représentées en bâtons.

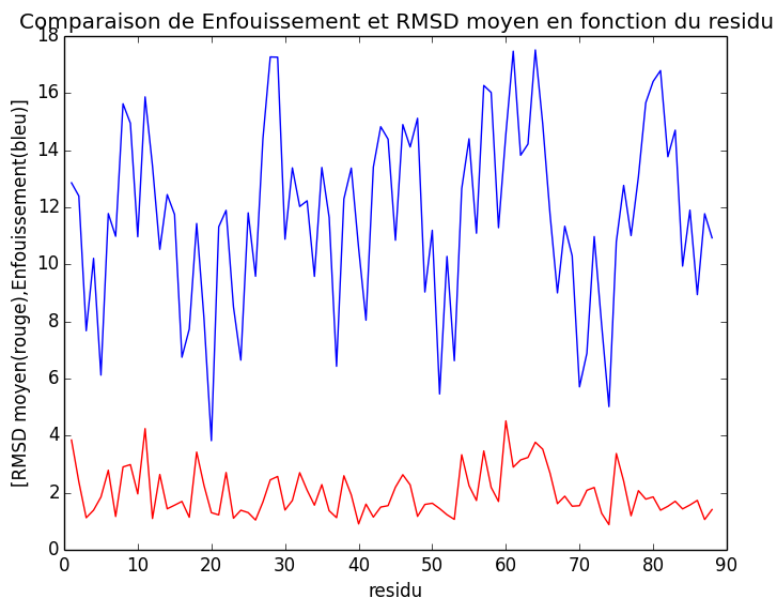


Figure 8. Représentation du RMSD/Enfouissement moyen de chaque résidu

La courbe rouge représente le RMSD moyen de chaque résidu. La courbe bleue représente la valeur d'enfouissement de chaque résidu. Les résidus sont identifiés par le numéro de résidu dans la séquence peptidique.

Ces dernières observations nous amènent 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 le RMSD moyen est élevé, l'enfouissement moyen est également élevé. Par contre, quand un résidu est situé en surface (enfouissement élevé), il n'est pas forcément dans une région très flexible.

Cette observation est logique puisque la protéine a besoins d'être flexible en surface pour pouvoir interagir avec les autres constituants moléculaires.

Nous avons décidé d'étudier les résidus qui sont en contact entre les deux protéines du complexe, ce qui correspond aux résidus ASP39, ASP35, TRP38, THR42 situés en interface, et les résidus Gln72, VAL73 et GLU76 situés en périphérie de cette interface (Lee-Peng Lee et Bruce Tidor. 2001).

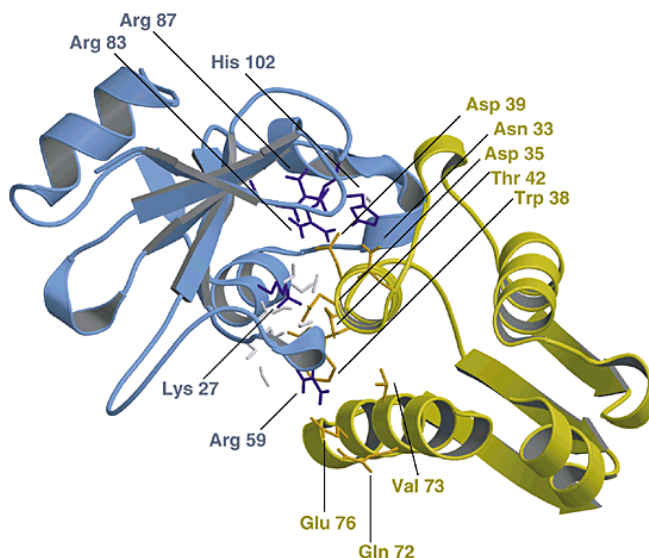


Figure 9. Structure du complexe Barnase-Barstar

Le complexe protéique est celui de la Barnase (en bleu) - Barstar (en jaune). Les résidus impliqués dans l'association sont indiqués.

Quand on regarde les propriétés de ces résidus, on constate qu'ils ont tous des valeurs d'enfouissement autour de 10/11A. Si on classe toutes les valeurs d'enfouissement en deux classes, ils ont un enfouissement significatif. On a alors tendance à dire que ces résidus sont situés à la surface de la protéine. La figure ci-dessus confirme ces observations. Cependant, ces résidus n'ont pas vraiment de RMSD très élevés par rapport à d'autres résidus ayant un RMSD moyen supérieur à 3 Å (résidus 1, 11, 60, 64). On remarque que les résidus 1, 11, 60 et 64 cités avant ne sont pas situés dans l'interface. Donc, les résidus se situant à l'interface peuvent être flexibles mais ils ne sont pas forcément les plus flexibles.

On souhaiterait savoir si ces résidus présents dans l'interface subissent des changements conformationnels majeurs et si ces résidus sont situés dans les mêmes régions de flexibilité. Nous avons donc visualisé deux résidus situés à l'interface correspondant aux résidus Asp39 et Thr42 et deux résidus situés en périphérie de l'interface correspondant aux résidus Gln72 et Glu76 sous *Pymol* au lieu de résidus 1, 11, 60, 64.

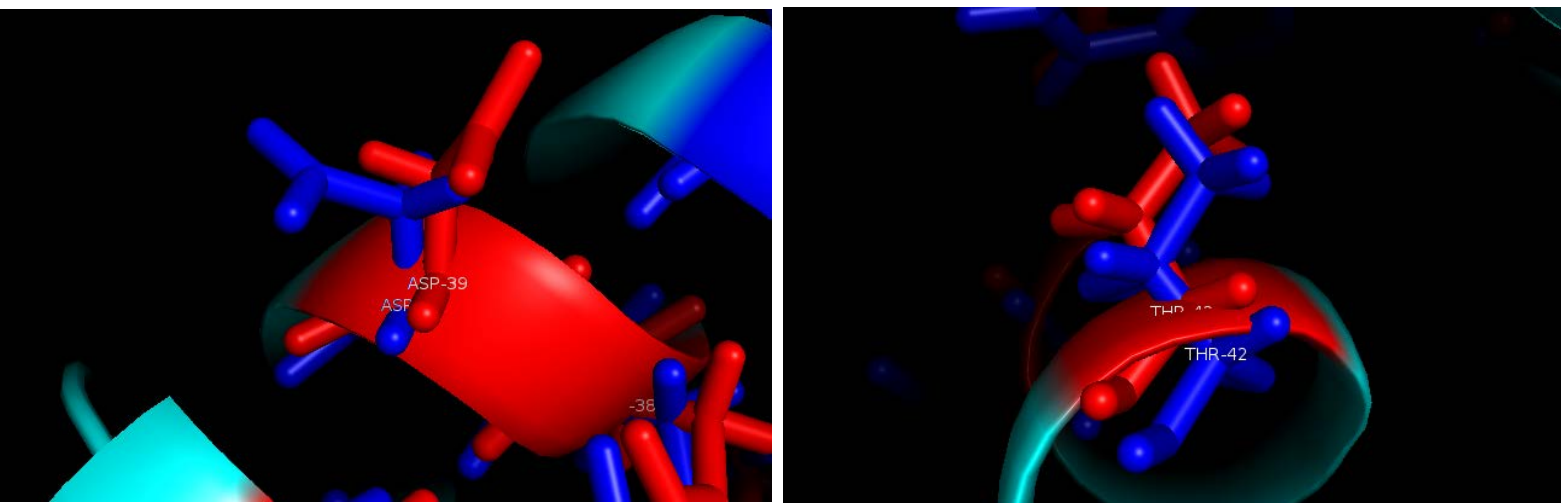
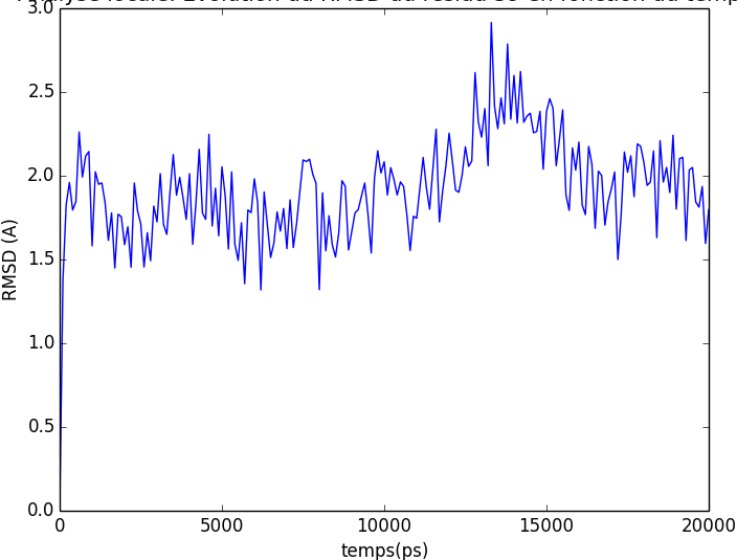


Figure 10. Représentation des résidus ASP 39 et THR 42 à l'interface

Le résidu en rouge est celui de la structure d'origine. Le résidu en bleu est celui de la structure après le changement conformationnel (model 200 sur le fichier md_prot_only_skip100).

Les deux résidus représentés ci-dessus sont à la surface et se situent dans l'interface (Figure 10). Il a été démontré que ASP 39 possède une propriété de clé pour identifier le bon partenaire afin d'induire une formation de complexe (S.Roy Kimura et al. 2001). On observe un faible changement de conformation dans cette région.

Analyse locale: Evolution du RMSD du residu 39 en fonction du temps



Analyse locale: Evolution du RMSD du residu 42 en fonction du temps

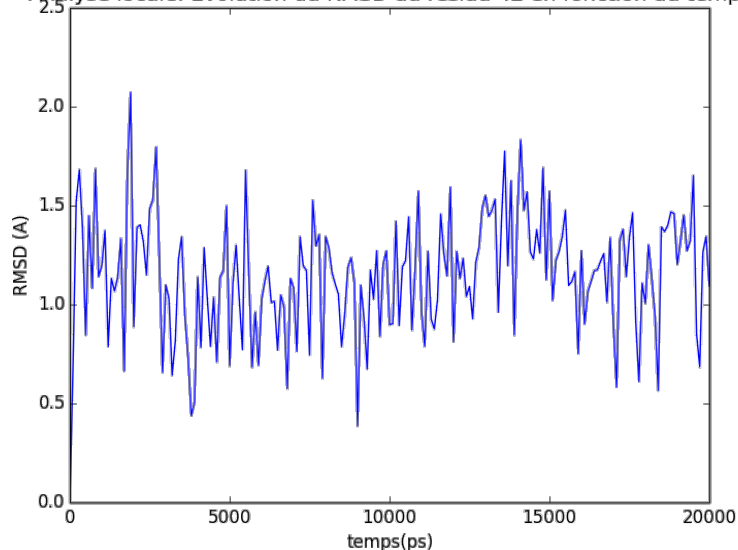


Figure 11. Evolution du RMSD du résidu ASP 39 et THR 42 en fonction du temps

La courbe représente la variation du RMSD (en Angstrom) du résidu ASP 39 (figure à gauche) et du résidu THR 42 (figure à droite) en fonction du temps (en ps).

Pour ces 2 résidus, on observe une variation fréquente de la valeur du RMSD. Il y a donc des changements de conformations, qui correspondent bien à ceux que l'on avait observé avec *Pymol*.

On remarque également que les variations sont souvent autour de 1.5 Å. Il n'y a donc pas vraiment d'importants changements conformationnels.

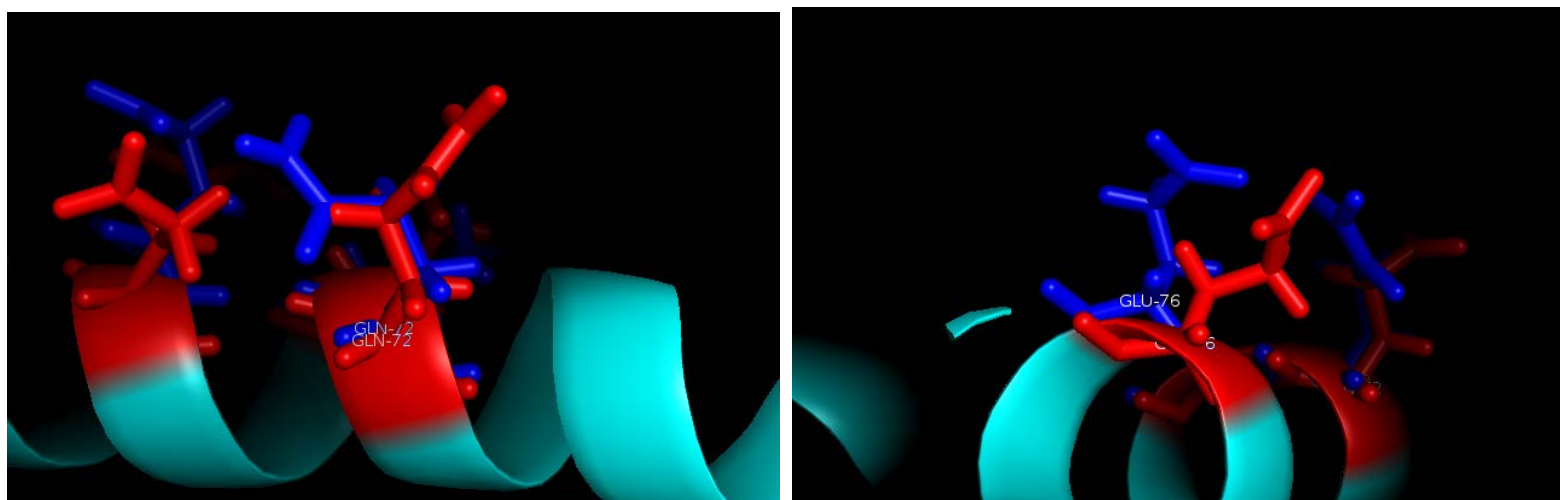


Figure 12. Représentation des résidus GLN 72 et GLU 76 en périphérie de l'interface

Le résidu en rouge est celui de la structure d'origine. Le résidu en bleu est celui de la structure après le changement conformationnel.

Les deux résidus représentés ci-dessus sont aussi à la surface et se situent en périphérie de l'interface dont GLU 76 a été démontré d'impliquer dans la stabilité du complexe formé (S.Roy Kimura et al. 2001).

Dans cette région, il y a également des changements conformationnels mais ces derniers semblent de plus grande amplitude (comparé à la figure 10).

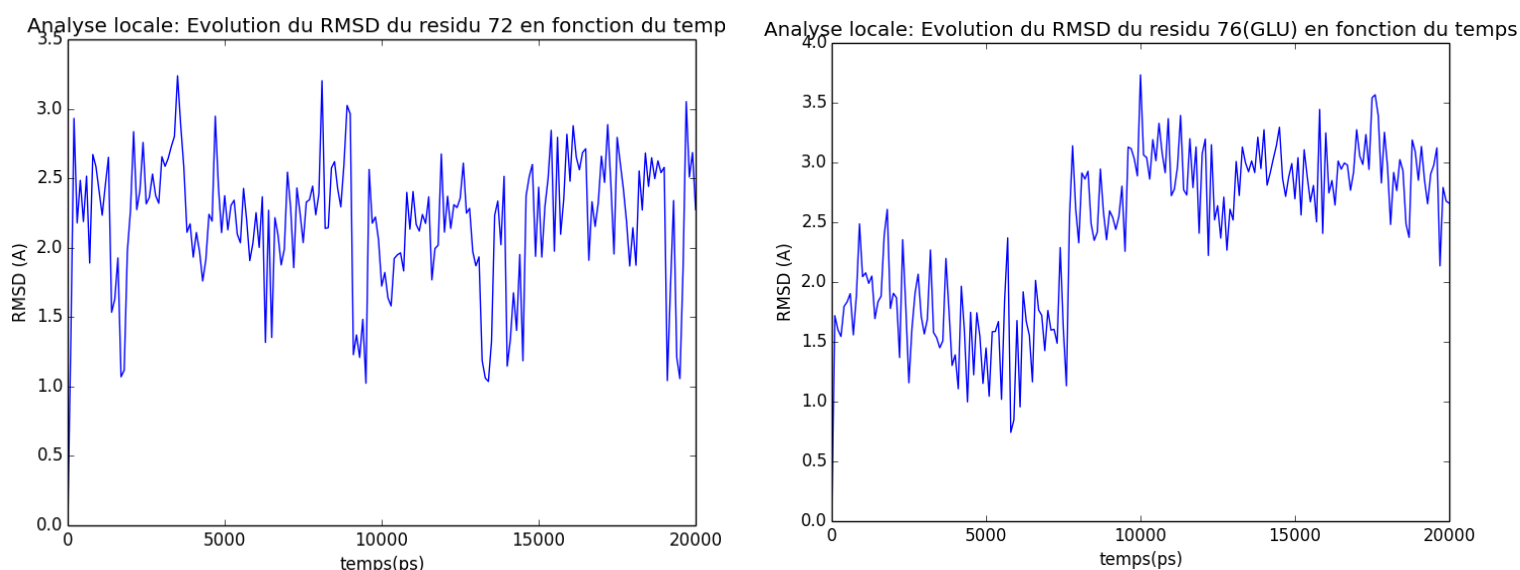


Figure 13. Evolution du RMSD de résidu GLN 72 et GLU 76 en fonction du temps

La courbe représente la variation du RMSD (en Angstrom) du résidu GLN 72 (figure à gauche) et du résidu GLU 76 (figure à droite) en fonction du temps (en ps).

Pour ces deux résidus, il y a également des changements de conformation (confirmé par les résultats *Pymol*). L'amplitude des variations de RMSD est plus élevée que pour les résidus de l'interface, ce qui nous laisse penser que les résidus situés en périphérie de l'interface sont dans des régions plus flexibles.

De plus, pour le résidu GLU 76, on remarque que pour les 7000 premières picosecondes, un ensemble de conformation est privilégié. A partir de 7000ps, on observe un changement conformationnel radical et on ne retrouvera plus les conformations initiales.

Le résidu GLN 72 ne fait pas de changement radical de conformation. En effet, il fait des allers-retours entre sa conformation de départ et d'autres conformations.

Ce phénomène peut s'expliquer par l'activité du complexe : la barstar a besoins de couvrir le site actif de la barnase pour inhiber son activité et de découvrir son activité quand il n'y a pas besoins d'inhiber cette activité. Ces phénomènes nécessitent d'importants changements conformationnels.

On remarque également un étrange phénomène pour le résidu LYS 38. En effet, la visualisation sous *Pymol* nous montre que ce résidu est dans l'interface du complexe protéique. Toutefois l'évolution de son RMSD au cours du temps montre un phénomène comparable à ceux observés pour les résidus situés en périphérie de l'interface : Comme pour le résidu GLU76, on observe un saut dans les valeurs de RMSD qui suggère un changement brutal de conformation. Ce phénomène est peut-être du à la taille plus importante du résidu qui l'empêche de bouger comme les autres.

Analyse locale: Evolution du RMSD du residu 38(LYS) en fonction du temps

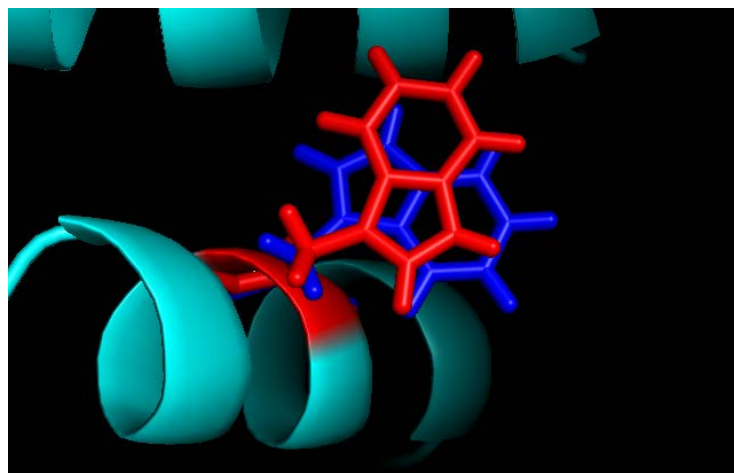
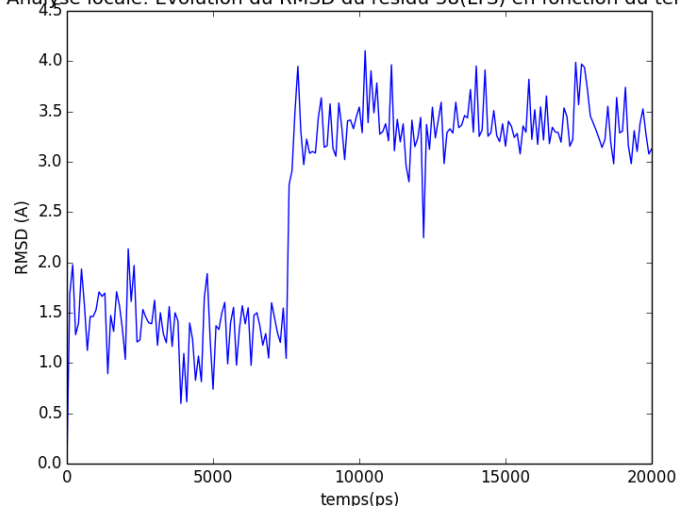


Figure 14. Evolution du RMSD de résidu LYS 38 en fonction du temps et sa représentation sous *Pymol*

La courbe représente la variation du RMSD (en Angstrom) du résidu LYS 38 (figure à gauche). L'image à gauche est la déviation du résidu LYS 38 entre sa position d'origine et sa position au bout de 20000ps.

Conclusion

Nous avons créé un programme qui permet d'identifier les changements conformationnels globaux et locaux d'une protéine. Ces changements de conformation sont importants, en particulier dans le cadre d'une interaction protéine-protéine puisqu'ils permettent à deux protéines de s'accrocher et de se détacher.

Nous avons mis en évidence que les régions les plus flexibles de la protéine étaient essentiellement situées à sa surface et qu'il n'y avait pas de régions flexibles à l'intérieur de la protéine. Ce phénomène est expliqué par le fait que la surface de la protéine est en contact avec le solvant et des ligands et a donc besoins de régions flexibles pour changer facilement sa conformation et ainsi pouvoir interagir avec ses partenaires.

Nous avons aussi étudié les changements de conformations des résidus situées à l'interface de la protéine avec la barcarasse et nous avons vu qu'ils n'étaient pas situés dans les régions les plus flexibles. Les résidus GLU76 et 38 sont remarquables par leur changement brutal de conformation. Ces changements de conformation se font au même moment. Quand la barstar décide d'inhiber la barnase, cela se fait donc à un moment précis et est suivis par un ensemble de changements de conformation pour plusieurs résidus pour stabiliser le complexe. 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.

Bibliographie:

S.Roy Kimura et al. Dynamical View of the Positions of Key Chains in Protein-Protein Recognition. *Biophysical Journal*. Volume 80. (2001)

Lee-Peng Lee et al. Barstar is electrostatically optimized for tight binding to barnase. *nature structural Biology* 8,73-76. (2001)

