

Dynamique moléculaire du complexe sRNP H /AxCa en interaction avec l'ARN

Arnaud BEGUE et Taylor VINGADASSALON

M1 BIBS - 2017

Le complexe H/ACA est une petite nucléoprotéine composée de cinq domaines, dont un ARN guide. Une précédente étude réalisée chez un mutant a montré l'existence de résidus jouant un rôle clé dans l'interaction avec l'ARN (Li et Ye, 2006). On se propose de mesurer les temps de contact entre ces résidus ainsi que la flexibilité des différents domaines de la protéine lors de l'interaction avec l'ARN mais cette fois-ci, dans la structure sauvage.

Introduction

Pyrococcus abyssi est une bactérie de la famille des archaea, vivant à des températures élevées (~90°C).

Chez ces bactéries, le complexe sRNP H/ACA, constituée de cinq domaines distincts, permet l'isomérisation de l'uridine en pseudo-uridine.

Cette transformation chimique, catalysée par la protéine Cbf5 du complexe, est la plus abondante dans les cellules.

Elle intervient dans plusieurs processus structuraux et fonctionnels tels que la stabilisation des doubles-hélices d'ARN (et donc des structures secondaires de l'ARN) ou encore la résistance à la dégradation.

Au cours de cet étude, nous cherchons comprendre le rôle de chacun de ces domaines dans l'interaction avec l'ARN.

Pour cela, nous disposons d'une dynamique moléculaire de 10ns du complexe RNP H/ACA sauvage (non muté).

À l'aide d'une API python que nous développerons dans le cadre de cette étude, les informations d'intérêt tels que les temps de contact entre les résidus dans la structure sauvage ainsi que la flexibilité des différents domaines de la protéine lors de l'interaction avec l'ARN, seront mis en évidence.

Méthode

Dans un premier temps, nous avons développé un outil capable de parser les fichiers PDB (de protéine) et de mémoriser les informations sur les atomes constituant la protéine : domaine auquel appartient l'atome, chaîne, conformation, résidu et position dans

l'espace (x,y,z) de l'atome. La structure de données choisie afin de stocker ces informations est un dictionnaire contenant autant de dictionnaires qu'il y a de type d'informations à mémoriser.

Secondairement, nous avons développé une fonction permettant de calculer les RMSD (Root-mean-square deviation) entre deux conformations de la protéine, c'est à dire la distance moyenne entre les atomes de protéines préalablement alignées. Les atomes choisis pour calculer les RMSD sont les atomes constitutifs du squelette peptidique de la protéine (pour les domaines protéiques) carbone-alpha (C α), oxygène (O) et azote (N), et les carbones 5' (C5) pour le domaine ARN guide.

Les RMSD ont été calculés pour toute la dynamique de la protéine (figure 1) et pour la dynamique de chaque domaine de la protéine (figure 2 à 6).

Afin de calculer ce RMSD pour les 5 domaines, on a choisi de diviser les données en sous-fichiers (1 fichier par domaine), afin d'optimiser l'utilisation de l'espace mémoire. Ces sous-fichiers seront utilisés dans le reste de l'analyse (projet).

Troisièmement, nous avons réalisé une fonction permettant de déterminer les résidus appartenant à l'interface, c'est-à-dire les résidus en contact avec l'ARN (dans notre cas). Pour cela on calcule le centre de masse (cdm) à partir de tous les atomes constituant le résidu, ainsi que la distance entre le centre de masse et l'atome le plus éloigné du résidu donné. On a donc un résidu qui est décrit par une sphère

autour de laquelle d'autres résidus seraient potentiellement en contact. On a choisi de ce fait d'identifier une interaction entre un résidu de la protéine et un résidu de l'ARN : lorsque la distance entre le centre de masse du résidu de l'ARN et de le centre de masse du résidus de la protéine est inférieur ou égale à la somme du rayon des deux résidus plus la distance d'interaction d'une liaison d'hydrogène (2,2 Å). On fait tourner cette fonction sur les différents domaines de la protéine à différents temps afin de mettre en évidence la fréquence d'interaction des différents résidus avec l'interface.

Résultats & Discussions

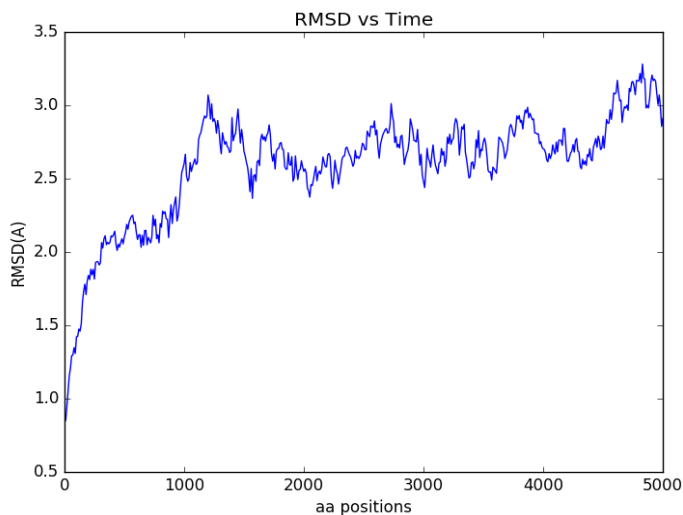


Figure 1: Dynamique de la protéine H/ACA en solution (10 ns)

Après mise en solution, la distance moyenne entre les atomes des résidus des squelette peptidique et ribonucléiques des conformations à différents temps augmente d'environ 2,6 ångström (Å) par rapport à la conformation de référence. Cela nous indique d'une part que la mise en solution de cette protéine entraîne des changement au sein de la conformation de la protéine. D'autre part, que ces changement de conformations se répercutent sur les squelettes protéiques et ribonucléique.

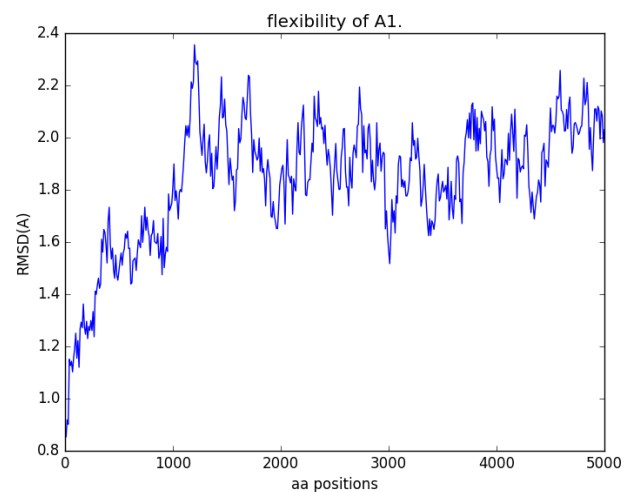


Figure 2: Dynamique du domaine A1 de la protéine H/ACA en solution.

Par rapport au profil observé à la figure 1, le profil du domaine A1 varie peu au cours de la dynamique par rapport à la dynamique de la protéine de référence. Ce qui sous-entend que le domaine est peu flexible.

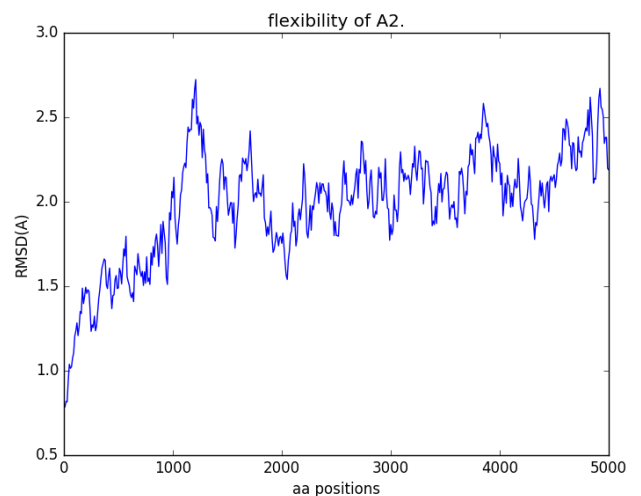


Figure 3: Dynamique du domaine A2 de la protéine H/ACA en solution.

Le domaine A2 semble peu flexible tout comme le domaine A1. Car il varie d'environ 2 Å par rapport à la configuration de référence.

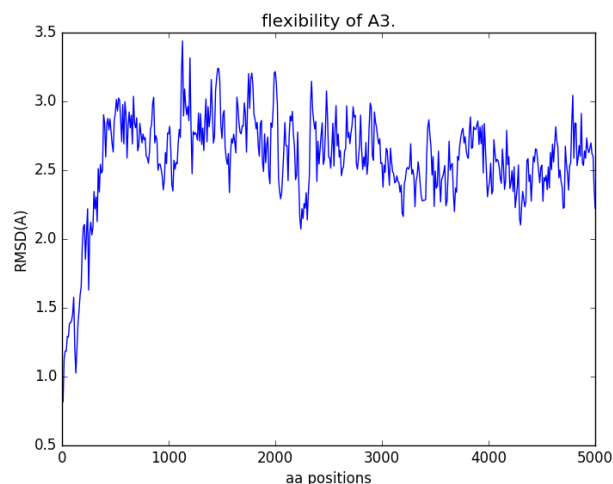


Figure 4: Dynamique du domaine A3 de la protéine H/ACA en solution.

Le domaine A3 montre une variation de la valeur du RMSD beaucoup plus importante que celle des domaines A1 et A2 par rapport à la configuration de référence. Cette augmentation laisse penser que ce domaine est plus flexible.

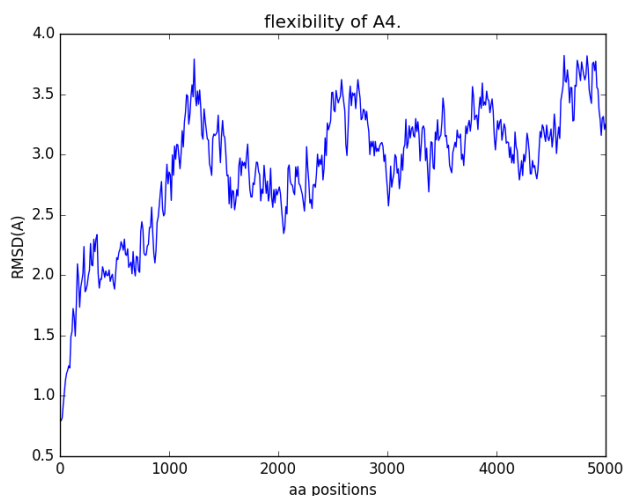


Figure 5: Dynamique du domaine A4 de la protéine H/ACA en solution.

On remarque que le RMSD de ce domaine varie d'environ 3Å par rapport à la configuration de référence. Cette variation suppose que ce domaine est lui aussi flexible.

Parmi les 4 domaines présentés, les domaines A1 et A2 semblent moins flexibles que les domaines A3 et A4, cependant globalement, il n'y a pas de variation importante du RMSD de ces domaines au cours de la dynamique et donc de la conformation de la protéine.

L'analyse de l'interface avec l'ARN montre que 128 résidus sont en contact avec l'ARN et tous ces résidus appartiennent aux chaînes A3 et A4. De plus, 108 de ces résidus (84 %) appartiennent aux résidus clés trouvés chez le mutant (résidus 23 à 37). Une de ces interactions est représentée sur la figure 6.

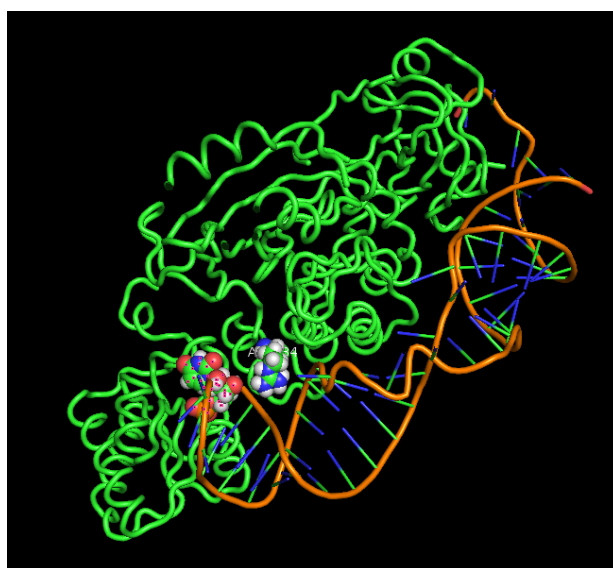


Figure 6: Résidu 34 de la chaîne A3 en interaction avec le résidu 32 de la chaîne B.

On constate que certains couples de résidus de l'interface ont une fréquence proche ou égale à 1, cela signifie qu'à cours de la dynamique étudiée, ces résidus restent fortement en interaction avec les résidus 23 à 37 de l'ARN.

D'autre part la majorité des 108 résidus trouvés restent en contact avec ces résidus clés, pendant tout le temps de la dynamique (10ns). On en conclut que ces résidus jouent un rôle majeur dans l'interaction avec l'ARN.

Conclusion

Au cours de cette étude, nous avons remarqué que certains domaines (A3 et A4) semblent plus flexibles que d'autres et interagissent avec l'ARN pendant toute la durée de la dynamique. De plus, ces domaines semblent contenir les résidus permettant l'interaction avec l'ARN (84 %).

Les domaines flexibles de la protéine semblent correspondre aux zones de fixation de la protéine à l'ARN.

Dans le cadre de cette étude, on voit bien que les résidus 23 à 37 jouent un rôle important dans l'interaction Protéine-ARN pour la protéine sauvage.