Etude du complexe sRNP H/ACA de *Pyrococcus abyssi*

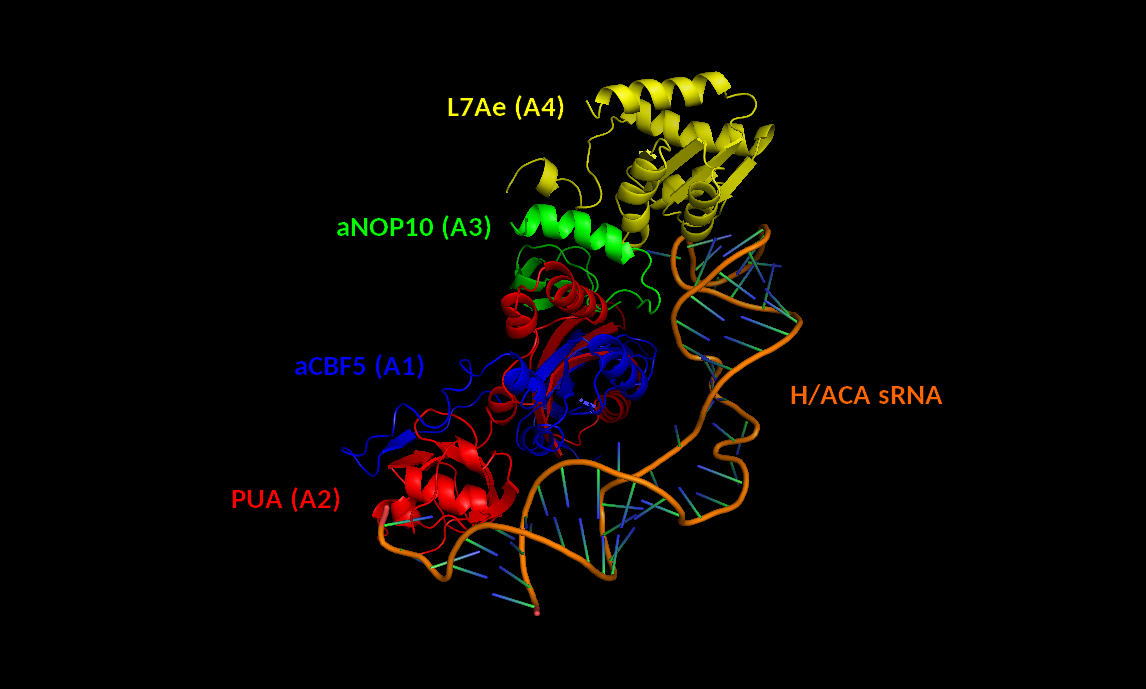
**Introduction**

La pseudouridylation est la modification post-transcriptionnelle la plus fréquente dans les ARN, et on la retrouve chez l’ensemble des orgnaismes vivants. Il s’agit de l’isomérisation d’un résidu uridine (U) en pseudouridine (Ψ). Par rapport à U, Ψ possède un groupe NH supplémentaire capable de former une liaison hydrogène, ce qui est utilisé par certains ARN pour stabiliser leur structure.

La réaction d’isomérisation est catalysée par un complexe ribonucléique (sRNP H/ACA), constitué d’une protéine à 5 domaines (aCBF5, PUA, aGAR1, aNOP10 et L7Ae) et d’un petit ARN à boîte H/ACA et possédant un motif K-turn. Le sRNA sert de guide et se lie par complémentarité à l’ARN cible qui Aesera modifié. Parmi les domaines protéiques du complexe, L7Ae reconnaît spécifiquement le motif K-turn. L’activité enzymatique ARN-pseudouridyle synthase est portée par aCBF5.

L’objectif de ce projet est d’étudier l’interaction entre les domaines protéiques et le sRNA du complexe sRNP H/ACA de l’archée *Pyroccocus abyssi*. Il s’agira de mesurer la flexibilité de chaque domaine ainsi que les temps de contact entre résidus chez un individu sauvage grâce à une dynamique moléculaire.

**Les données**

 Nous disposons de deux fichiers PDB contenant la structure de 4 domaines protéiques et du sRNA chez le sauvage. Ces fichiers sont visualisables sous PyMol (Figure 1). Le premier fichier contient la structure de référence du complexe, le second contient 500 conformations de ce complexe issues de la dynamique moléculaire de 10 ns.

***Figure 1****: Structure du complexe visualisée sous PyMol. Chaque domaine est représenté d’une couleur différente. Entre parenthèses : identifiants des domaines protéiques dans les fichiers PDB fournis.*

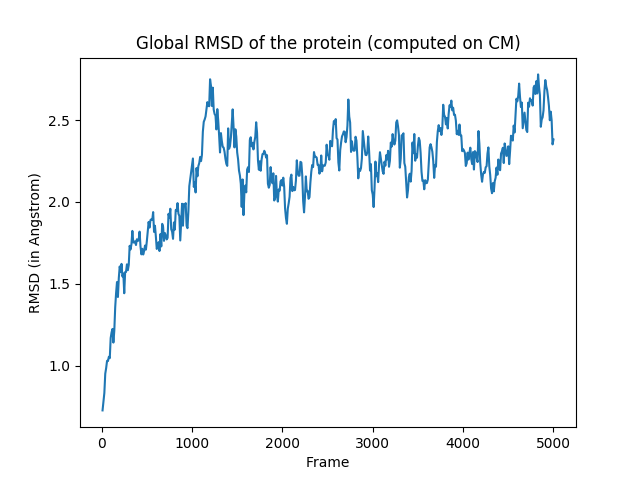
**Etude des changements conformationnels globaux**

Les résultats numériques sont disponibles dans le fichier ‘rmsd.txt’.

• Changements conformationnels de la protéine :

Le RMSD (*Root Mean Square Deviation*) est une mesure en Angström (Å) rendant compte de la déviation structurale entre deux structures protéiques alignées. Plus le RMSD est petit, plus les structures protéiques se ressemblent. Le RMSD se calcule sur des paires d’atomes, une paire comprenant un atome de la structure 1 et l’atome correspondant dans la structure 2.

avec N le nombre total de paires d’atomes utilisées pour le calcul, et la distance spatiale séparant les deux atomes d’une paire i.

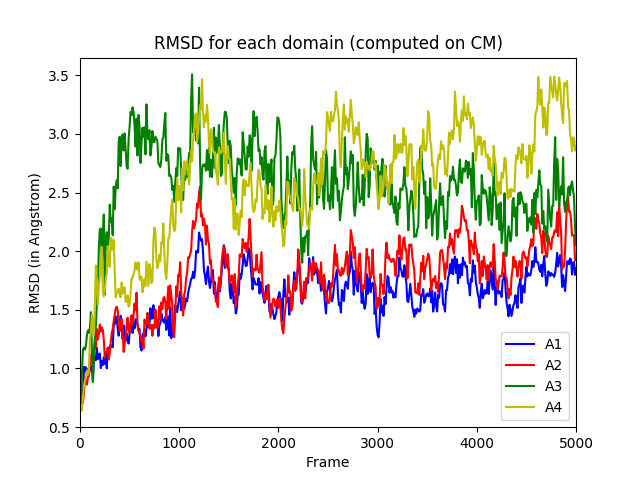
 Le RMSD global de la protéine a été calculé entre chacune des 500 conformations de la dynamique et la structure de référence (Figure 2). Nous avons choisi de le calculer sur les centres de masse (CM) des résidus, mais notre programme offre la possibilité de choisir le mode de calcul (sur les centres de masse, ou sur les carbones α des résidus).

***Figure 2****: RMSD de la protéine calculé sur les centres de masse des résidus, au cours du temps.  
Chaque frame, numérotée de 10 à 5000 par pas de 10, correspond à une conformation de la dynamique.*

On observe une augmentation du RMSD au début de la dynamique avec un pic à plus de 2.5 Å aux alentours de la conformation (frame 1000). Ensuite, le RMSD varie entre environ 2 et 2.5 Å, mais se stabilise globalement. Le RMSD reste inférieur à 3 Å tout au long de la dynamique moléculaire, on peut donc dire que la structure de la protéine change peu au cours du temps.

* C’est le RMSD sans ARN ? Il y a peu de variation du RMSD, la structure change peu au cours du temps, entre 1 et 3 les structures sont considérée comme similaires. C’est une protéine enzymatique chimérique (donc des sous-unités qui subissent leur propre modification). Là il n’y a rien à dire à mon avis c’est notre contrôle.
* Je ne suis pas sûre de comprendre ta question. La protéine est en contact avec le sRNA, mais le RMSD est calculé seulement sur la protéine. Je pense qu’il faut quand même commenter la figure vu que c’est demandé dans le sujet.

• Changements conformationnels des domaines protéiques :

 Le RMSD a été calculé (par rapport aux centres de masse des résidus) pour chacun des quatre domaines protéiques A1, A2, A3 et A4 (Figure 3).

***Figure 3****: RMSD calculé sur les centres de masse des résidus pour chaque domaine protéique,  
au cours du temps.*

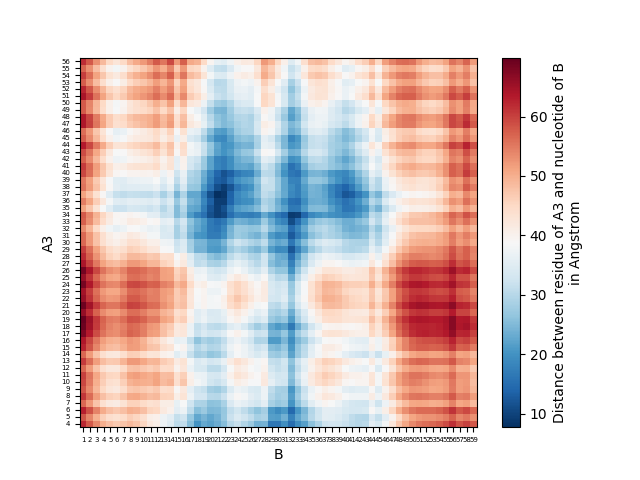
Pour les 4 domaines, on observe une augmentation avec un pic aux alentours de la conformation, puis une ‘stabilisation’ du RMSD, ce qui est cohérent avec les variations du RMSD global.

On remarque que les domaines A1 (aCBF5) et A2 (PUA) varient de la même façon, avec moyenne à environ 1.8 Å, et subissent moins de changements conformationnels que les domaines A3 (aNOP10) et A4 (L7Ae) dont le RMSD moyen est environ 2.5 Å. Les domaines A1 et A2 sont donc moins flexibles que A3 et A4.

Au début de la dynamique, le domaine subissant le plus de changements conformationnels est A3, à la fin il s’agit de A4. Il est intéressant de noter que les domaines les plus flexibles interagissent avec le sRNA.

* Conclusion sur le RMSD

**Etude des changements conformationnels locaux**

 Dans un premier temps, nous avons voulu visualiser des distances entre les résidus de la protéine protéique et les résidus de l’ARN dans la structure de référence. Ainsi, pour chaque domaine, nous avons calculé une matrice de distance : pour chaque résidu du domaine protéique, sa distance à tous les résidus de l’ARN est calculée. Les calculs de distance ont été réalisés entre les centres de masse des résidus, mais le programme permet de choisir le mode de calcul : entre les centres de masse, ou distance minimale entre les atomes des deux molécules. Les matrices de distance ont ensuite été représentées sous forme de *heatmap*. La Figure 4 montre les distances entre les résidus du domaine A3 et l’ARN. Les *heatmaps* des autres domaines figurent en Annexe 1.

***Figure 4****: matrice des distances entre le domaine A3 et l’ARN (B) représentée sous forme de heatmap.*

Les zones bleu foncé indiquent les résidus proches (distance inférieure à 10 Å). Les résidus 34 à 38 de aNOP10 semblent proches des résidus 21 et 22 de l’ARN. Ces résultats sont confirmés grâce à PyMol (Figure 5).

De même, une *heatmap* a été générée pour déterminer les résidus à l’interface entre aNOP10 et L7Ae (Annexe 2).

Dans un second temps, nous avons calculé les fréquences d’appartenance à l’interface protéine/ARN de chacun des résidus de la protéine au cours de la dynamique moléculaire. Les résidus dont la fréquence est non nulle figurent dans le fichier ‘interfaceFreq.txt’. Pour ce faire, nous avons procédé de la façon suivante.

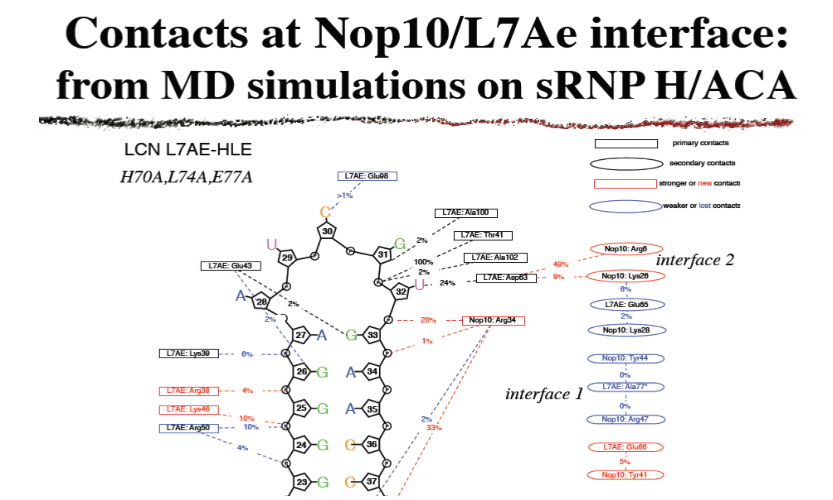
Soit N le nombre de fois qu’un résidu appartient à l’interface protéine/ARN (au maximum, N = 500 puisqu’il y a 500 conformations). Pour chaque conformation, on calcule pour un résidu sa distance par rapport à tous les nucléotides de l’ARN. On garde la plus petite distance entre ce résidu et un nucléotide. Si cette distance est inférieure à un seuil fixé par l’utilisateur (par défaut, seuil = 9.0 Å), alors on incrémente N de 1. Une fois toutes les conformations parcourues, la fréqence d’appartenance est .

7En modifiant le champ B-factor dans le fichier PDB, on peut visualiser avec PyMol les résidus apaprtenant à l’interface (Figure 6).

***Figure 6****: Résidus appartenant à l’interface avec l’ARN (en rouge) pour la première conformation (frame 10).*

Enfin, nous avons calculé les temps de contact entre différentes paires de résidus clés. En effet, en amont de ce projet, une dynamique moléculaire de 10 ns a été réalisée chez un mutant connu pour altérer l’activité du complexe. Les temps de contact entre plusieurs résidus de la protéine et certains résidus de la protéine avec l’ARN ont été mesurés au cours de cette dynamique, ce qui a permis d’identifier des résidus jouant un rôle clé dans l’interaction avec l’ARN (Figure 7).

Pour chaque paire de résidus choisis, nous avons calculé le nombres de conformations pour lesquelles les résidus sont en contact, c’est-à-dire pour lesquelles la distance entre les centres de masse des molécules est inférieure à 9.0 Å. Le temps de contact T est alors donné par la formule :



***Figure 7****: Contacts chez le mutant entre résidus clés pour l’interaction.*

Nous avons choisi de calculer les temps de contacts pour les paires de résidus suivantes :

* à choisir

Les résultats sont disponibles dans le fichier ‘contacts.txt’.

* interpréter les résultats

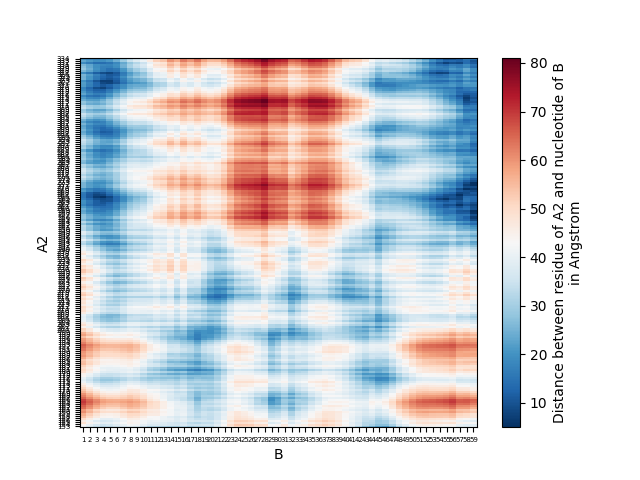
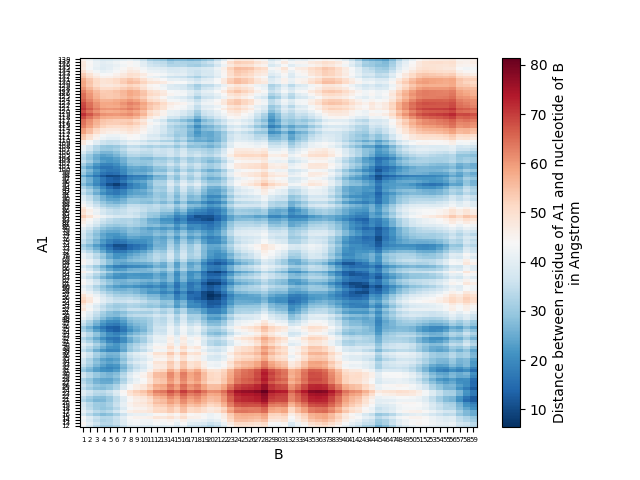
**Conclusion**

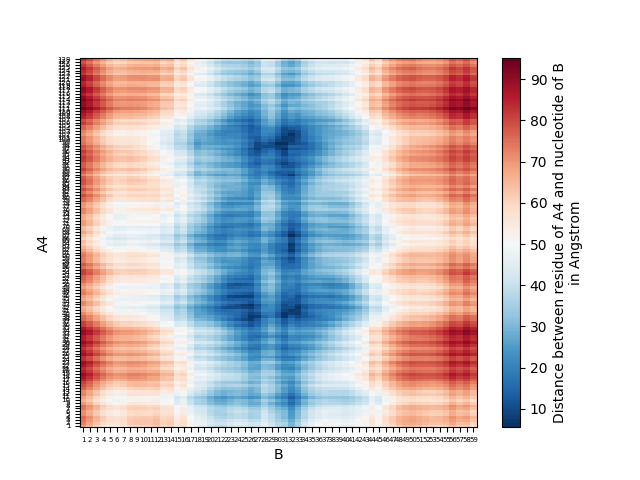
**Annexes**

**Annexe 1**: Matrices des distances représentées sous forme de *heatmaps* entre les domaines A1, A2 et A4 et l’ARN (B).

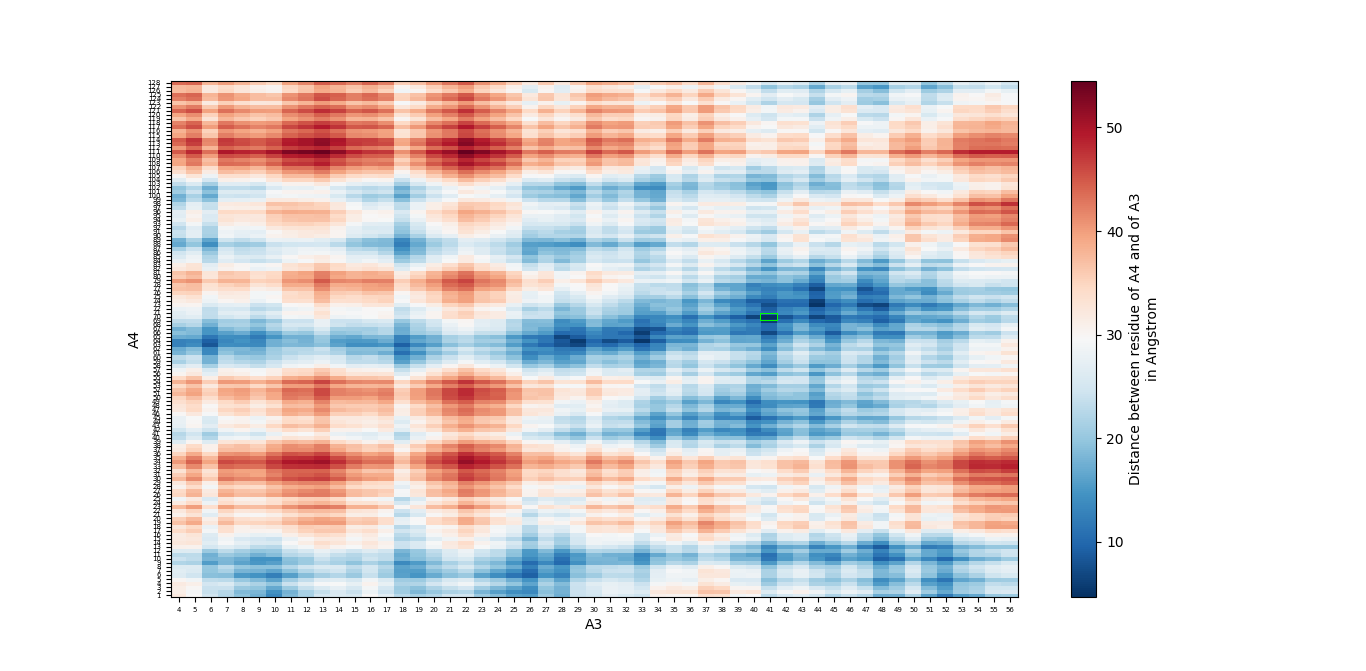
**Annexe 2**: Matrice des distances représentée sous forme de *heatmap* entre les domaines A3 et A4.

**Annexe 1**

Matrices des distances représentées sous forme de *heatmaps* entre les domaines A1, A2 et A4 et l’ARN (B).  
Les distances ont été calculées entre les centres de masse des résidus.



**Annexe 2**

****Matrice des distances représentée sous forme de *heatmap* entre les domaines A3 et A4.  
Les distances ont été calculées entre les centres de masse des résidus.

Visualisation sous PyMol des résidus 34 du domaine A3 et 70 du domaine A4 (rectangle vert sur la *heatmap*).  
Ces résidus sont en effet très proches.