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 organismes 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 augmente la rigidité du squelette phospho-ribose et donc stabilise la structure ARN. La réaction d’isomérisation peut être catalysée par un complexe ribonucléique (sRNP H/ACA) qui permet l’interaction entre un sRNA à boîte H/ACA et un ARN cible. Ce complexe RNP est constitué de 5 domaines (aCBF5, PUA, aGAR1, aNOP10 et L7Ae) et reconnaît le sRNA par son motif K-turn.

Ainsi le sRNA sert de guide et se lie par complementarité à l’ARN cible à modifié, permettant de cibler le U à isomeriser. La littérature montre que L7Ae reconnaît spécifiquement le motif K-turn du sRNA et que l’activité enzymatique ARN-pseudouridyle synthase est porté par aCBF5.

## But

On veut ici créer un programme qui nous permettra

1. De comprendre le rôle de chacun des domaines dans l’interaction avec l’ARN
2. d’identifier les résidusde de chaque domaine de la proteine jouant un rôle clé dans l’interaction avec l’ARN
3. mesurer la flexibilité des différents domaines de la protéine lors de l’interaction avec l’ARN.

# L’Etude

## Les Données

Nous disposons de deux fichiers PDB contenant la structure de 4 domaines protéiques et du sRNA chez le complexe sauvage (référence). Ces fichiers sont visualisables sous PyMol (Figure 1).

1. Le premier fichier contient la structure de référence du complexe lié au sRNA.
2. le second contient 500 conformations de ce complexe +sRNA issues d’une étude de la dynamique moléculaire sur 10ns.
3. 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 la dynamique moléculaire avec un mutant. Ainsi des résidus clé ont été identifié. Nous allons ici le temps de contact entre ces résidus chez le sauvage.

## 

Figure : Structure du complexe visualisé sous PyMol. Chaque domaine est représenté d’une couleur différente. Entre parenthèse les identifiants des domaines dans les fichiers PDB fournis.

Bleu : aCBF5 (A1). Rouge : PUA (A2). Vert : aNOP10 (A3). Jaune : L7Ae (A4). Orange : le sRNA guide.

## Méthode : Comparaison et superposition de structure

Utiliser pour regrouper les protéines par similarité structurale (classification),on peut aussi estimer l’importance de certains résidus dans la structure (en terme de stabilité ou de mutabilité) si on compare la même structure au cours du temps (ici 10ns pour voir les différents changement de conformation globale ou locale).

Pour cela on calcul l’ecart quadratique moyen global des distances entre les N atomes (i) en correspondance dans 2 structures

### RMSD

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

On superpose donc la structure 1 à la structure 2, et on traite de manière identique tous les atomes (bien qu’ils soient plus flexible en surface). La signification du RMSD dépend de la taille des protéines mais ici on regarde l’évolution de 1 seule protéine. La valeur du RMSD dépend du choix des atomes (ici on regarde par rapport aux champs de masse, mais il est possible de choisir le carbone alpha des résidus comme point de référence dans le programme).

Les structures sont considéré comme identiques si le RMSD est égal à 0 et distantes pour un RMSD > 4 angstrom.

# Etude des changements conformationnels

## Changement conformationnel de la protéine

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

Pour une étude des changements conformationnels globaux on cherche le maximum d’atome alignés avec le plus faible écart.

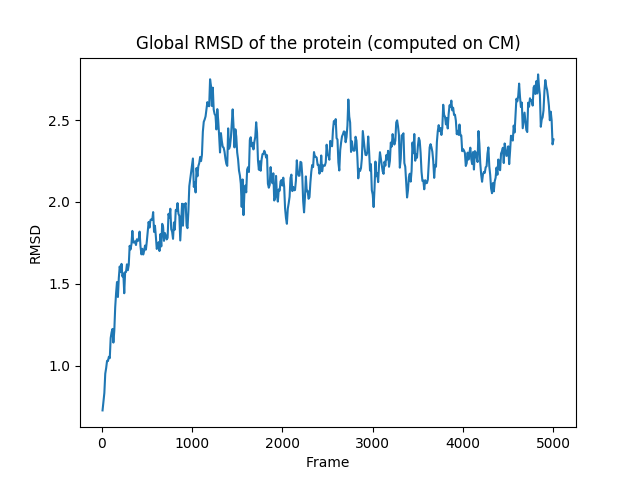
Le RMSD global de la protéine a été calculé entre chacune des 500 conformations de la dynamique et la structure de référence (Figure 2).

Figure : RMSD Global.

RMSD de la protéine calculé sur les centres de masse des résidus, au cours du temps. Chaque frame, numéroté de 0 à 5000 par pas de 10, correspond à une conformation de la dynamique.

On observe une augmentation progressive du RMSD pour atteindre un plateau entre 2 et 2.5 A (aux alentours de la conformation frame 1000). Des changments conformationnels par rapport à la structure de référence sont donc détectés au cours de la dynamique moleculaire, lorsque celle-ci est lié au sRNA. Bien que l’on observe un pic à plus de 2.5 A, le RMSD ne varie que peu au cours de la dynamique.Donc dans son ensemble le complexe ne subit que peu de modification lors de son contact avec le sRNA.

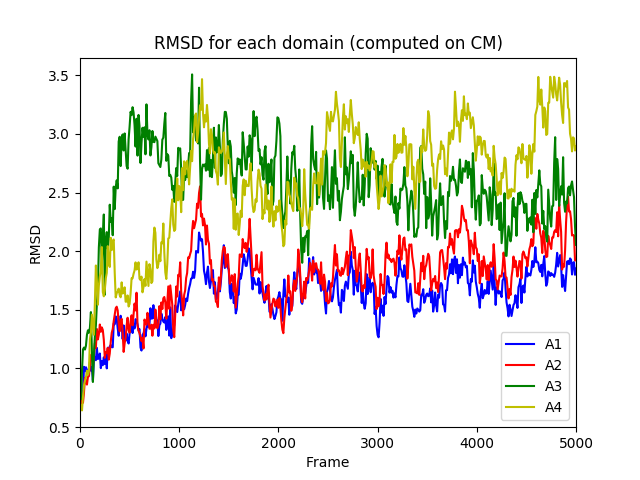
Changements conformationnels des domaines protéiques :

Figure : RMSD par domaine du complexe sRNP. RMSD calculé sur les centres de masse des résidus pour chaque domaine au cours du temps. Bleu : aCBF5 (A1). Rouge : PUA (A2). Vert : aNOP10 (A3). Jaune : L7Ae (A4).

Le RMSD a été calculé pour chacun des 4 domaines protéiques A, A2, A3 et A4 (Figure 3)

Pour les 4 domaines, on observe une augmentation progressive du RMSD jusqu’à un palier. Ce qui est cohérent avec le RMSD global (Figure 2). Des changements de conformation progressifs jusqu’à une « stabilisation » de la conformation.

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

Les domaines aCBF5 et PUA varient de concert et sont moins « flexibles » que les deux autres domaines. Ce qui est en accord avec la littérature où PUA est considéré comme faisant partie du domaine aCBF5 et joue un rôle dans la stabilité de aCBF5.

1. Au début de la dynamique, le domaine A3(aNOP10) subit le plus de changements conformationnels.
2. à la fin c’est A4.

Les domaines les plus « flexibles » sont ceux qui interagissent le plus avec l’ARN.

## Conclusion

Nous avons ici observé les résultats de sortie du programme permettant d’étudier si la proteine subit des changements conformationnels majeurs en solution via l’évolution au cours du temps du RMSD.

Ces résultats permettent de dire que la structure du complexe en solution est très proches de la structure de référence. Elle peut donc servir de base d’étude pour évaluer l’impact de mutation sur le complexe en solution.

De plus on a pu observer que chaque domaine de la proteine à sa propre évolution conformationnelle au cours du temps. On peut donc se dire que chaque domaine est necessaire à l’activité du complexe. Ce qui peut-être confirmé via des mutations spécifiques et une étude du RMSD résultant et de l’activité enzymatique de aCBF5.

Analayse des changements conformationnels locaux (cf fichier *interfaceFreq.txt*)

## Méthode

### Matrice de distance

Dans un premier temps, nous avons voulu visualiser des distance entre les résidus de la protéine et les résidus de l’ARN dans la structure de référence.

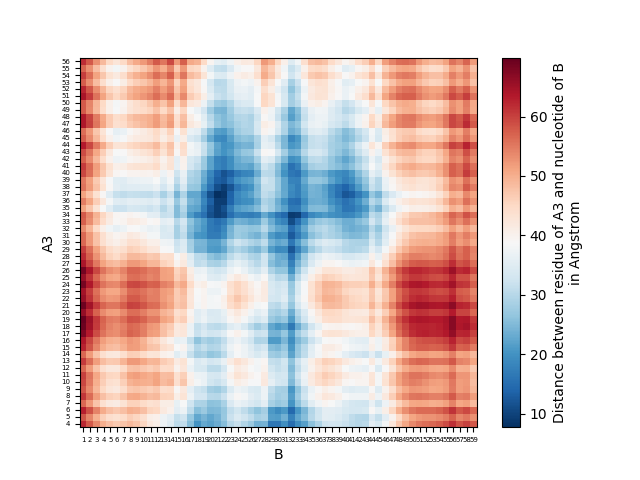
Ainsi, pour chaque domaine, nous avons calculé une matrice de distance : pour chaque résidus du domaine protéique, sa distance à tous les résidus de l’ARN est calculée. Les calculs de distance ont été réalisé entre les centres de masse des résidus, mais le programme permet de choisir le mode de calcul (centre de masse ou distance minimal entre les atomes des deux molécules). Les matrices de distances ont ensuite été représenté sous forme de heatmap (Figure 4 et Annexe : heatmap entre aNOP10 et L7Ae).

Figure Heatmap entre les résidus de L7Ae (A3) et l'ARN (B). matrice des distances entre les domaines A3 et B. Les zones bleu foncés indiquent les résidus proches (distance inférieur à 10 A).

Les résidus 34 à 38 de aNOP10 semblent prochent des résidus 21 et 22 de l’ARN. Ce qui est confirmé grâce à une visualisation des résidus sous PyMol.

On peut ainsi regarder la position de chaque résidus des différents domaines et ainsi vérifier leur appartenances à des zones clé dans l’interaction avec l’ARN.

### Fréquence d’appartenance à l’interface protéine/ARN

Dans un second temps, nous avons calculé les fréquences d’appartenance à l’interface protéine/ARN de chacun des résidus au cours de la dynamique moléculaire. Les résidus dont la frèquence est non-nulle figurent dans le fichier ‘interfacFreq.txt’.

1. Soit N le nombre de fois qu’un résidus appartient à l’interface protéine/ARN (au maximum N=500).
   1. Pour chaque conformation, on calcul pour un résidus sa distance par rapport à tous les nucléotides de l’ARN. On garde la plus petite distance entre ce résidus et un nucléotide.
   2. Si cette distance est inférieure à un seuil fixé par l’utilisateur (par défuat, seuil =9.0 A) alors on incrémente N de 1.
2. Une fois toutes les conformations parcourues, la fréquence dappartenance est calculée



Figure : Résidus appartenant à l'interface avec l'ARN pour la première conformation. Visualisation PyMol

### Temps de contact entre résidus

Enfin nous avons calculé les temps de contacts entre différent paires de résidus clé. En effet, en amont de ce projet, une dynamique moléculaire de 10ns a été réalisée chez un mutant connu pour altérer l’activité du complexe. Les temps de contact ente 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’indentifier des résidus jouant un rôle clé dans l’interaction avec l’ARN. (Figure 6).

Pour chaque paire de résidus choisies, nous avons calculé le nombre de conformations pour lesquelles les résidus sont en contact, c’est-à-dire pour lesquelles la distance entre les centres de masse des molécules et inférieur à 9.0 A. Formule ???

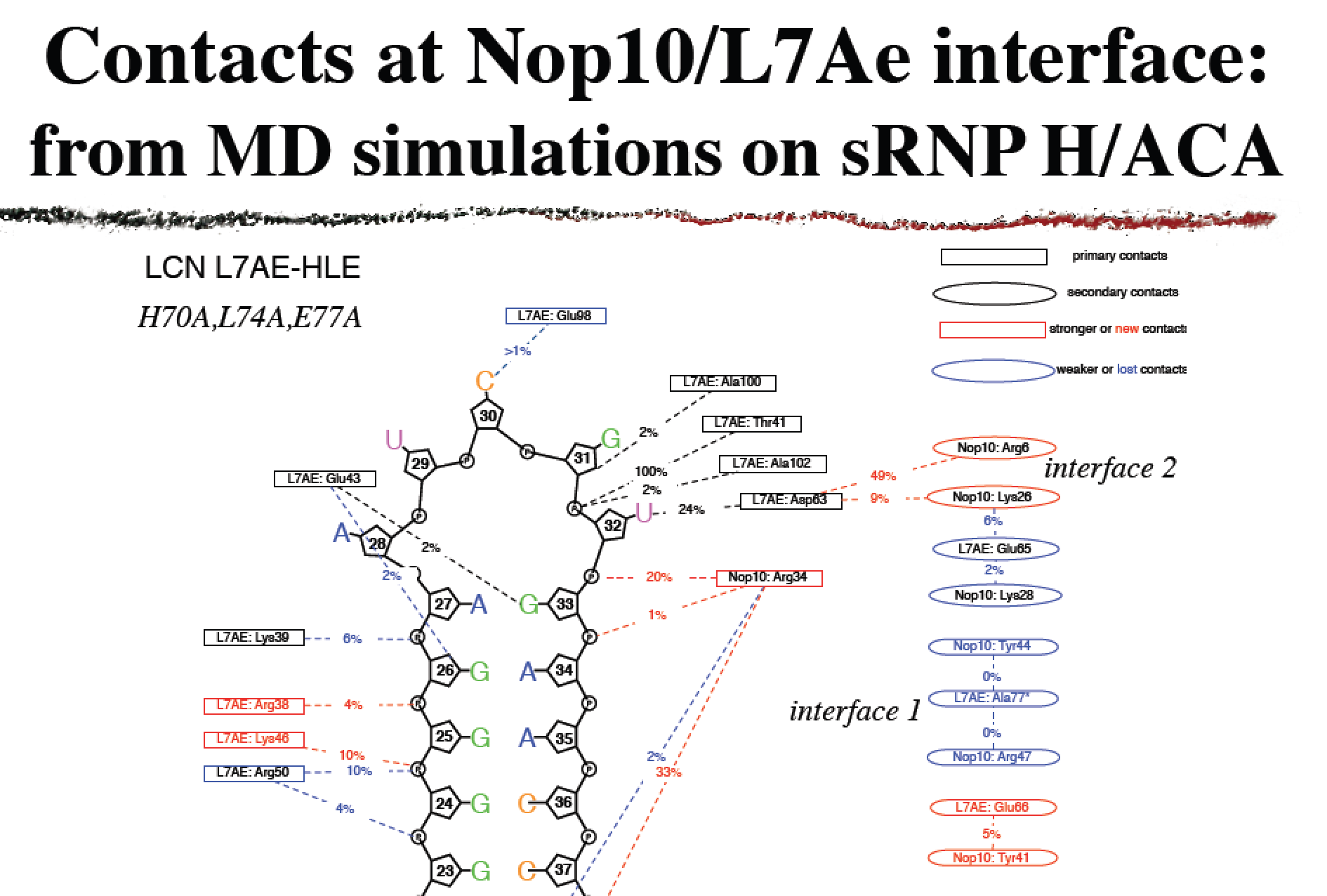


Figure : Contacts chez le mutant entre résidus clés pour l'interaction

## Résultats

Tous . interaction avec tige apicale de sRNA, domaine K-turn (aCBF5, aNOP10, L7Ae) permet de localiser la poche de pseudo sur le site catalytique de aCBF5. Si interaction modifié alors cela affecte l’activité catalytique : soit car ne reconnaît plus le K-turn ou alors il y a un défaut de fixation (décalage ou alors trop stringent et le complexe ne peut se défaire donc cela limite le nombre de pseudoU possible).

Pour que l’ARN se fixe à L7Ae il faut que celui-ci subissent de faibles changement de conformation car la poche de fixation du nucléotide U visé existe avant la fixation du sRNA donc trop de changement entraine une déréglement de cette poche et donc affecte l’activité catlytique de aCBF5 car on ne peut fixer le substrat.

Arginine 34 de NOP10 interaction avec sRNA (nucléotide 33) : nouvelle liaison

Lysine 46 de L7Ae avec nucl 25 nouvelle liaison

Alanine100 de L7Ae avec 31

Tyr41 de NOP10 interagit avec L7Ae =🡺iinteragit avec nucl 32

# Conclusion

Nous pouvons avec la méthode explicitée ici, analyser différente mutation du complexe sRNP afin de mettre en évidence les domaines nécessaires et plus précisément les résidus clés au maintien de l’activité catalytique du complexe sRNP H/ACA. De plus nous pouvons via l’étude du RMSD dans une étude de dynamique moléculaire évaluer l’impact de chaque mutation sur la protéine globale et sur les différents domaines, révelant ainsi l’impact qu’ils ont sur chacun.

On voit que chaque domaine donc maintenir une certaine intégrité et qu’ils sont tous peu flexibles. On peut donc penser que le coplexe peut devenir « inefficace » facilemnent.

En effet le sRNA doit s’apparier par complémentarité avec l’ARN cible (le plus souvent ribosomal) dans une poche créer via l’interaction de chaque domaine de la protéine, si le sRNA est détecté.

L’intégrité de L7Ae est donc primordiale car c’est le domaine qui reconnaît spécifiquement le motif k-turn du sRNA.