02/05/2017

De Tollenaere Maud & Liegeois Séverine

Etude du complexe sRNP H/ACA

de *Pyrococcus abyssi*

Plan du Compte-Rendu

Introduction

But

L’Etude

Les Données

Méthode : Comparaison et superposition de structure

RMSD

Etude des changements conformationnels

Changement conformationnel de la protéine

Conclusion

Analyse des changements conformationnels locaux

Méthode

Matrice de distance

Fréquence d’appartenance à l’interface protéine/ARN

Temps de contact entre résidus

Résultats

Conclusion et Perspective

Annexes

# 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 (). Par rapport à U,  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.

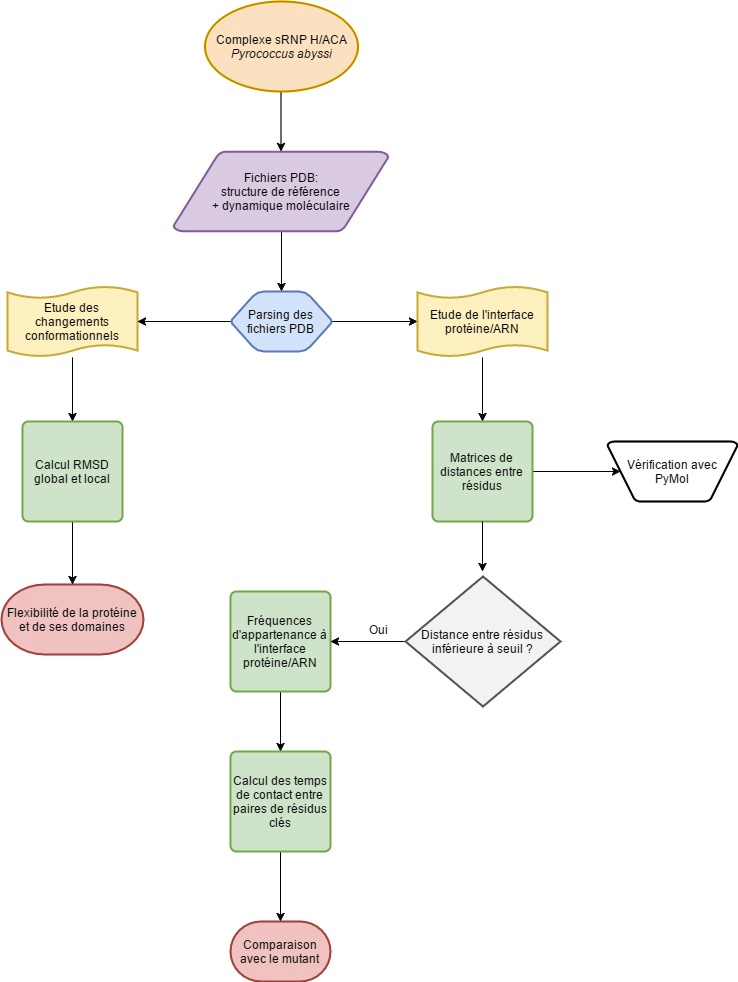
Ainsi le sRNA sert de guide et se lie par complementarité à l’ARN cible à modifier, permettant de cibler le U à isomeriser. La littérature montre que L7Ae reconnaît spécifiquement le motif K-turn du sRNA et que l’activité enzymatique ARN-pseudouridyle synthase est porté par aCBF5.

## 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. Mesurer la flexibilité des différents domaines de la protéine lors de l’interaction avec l’ARN.
3. D’identifier les résidus de chaque domaine de la proteine jouant un rôle clé dans l’interaction avec l’ARN.

# La méthodologie : flowchart



# L’Etude

## Les Données

Le projet porte sur l’étude du complexe sRNP H/ACA chez l’archée *Pyrococcus* abyssi. Nous disposons de deux fichiers PDB contenant la structure de 4 domaines protéiques et du sRNA chez un individu sauvage. Ces fichiers sont visualisables sous PyMol (Figure 1).

Le premier fichier contient la structure de référence du complexe protéine-sRNA.

Le second contient 500 conformations de ce complexe issues d’une dynamique moléculaire de 10ns.

Les temps de contact entre plusieurs résidus de la protéine et certains résidus de la protéine avec l’ARN ont été mesurés au cours d’une dynamique moléculaire avec un mutant. Ainsi des résidus clé ont été identifiés. Nous allons mesurer le temps de contact entre ces résidus chez le sauvage.

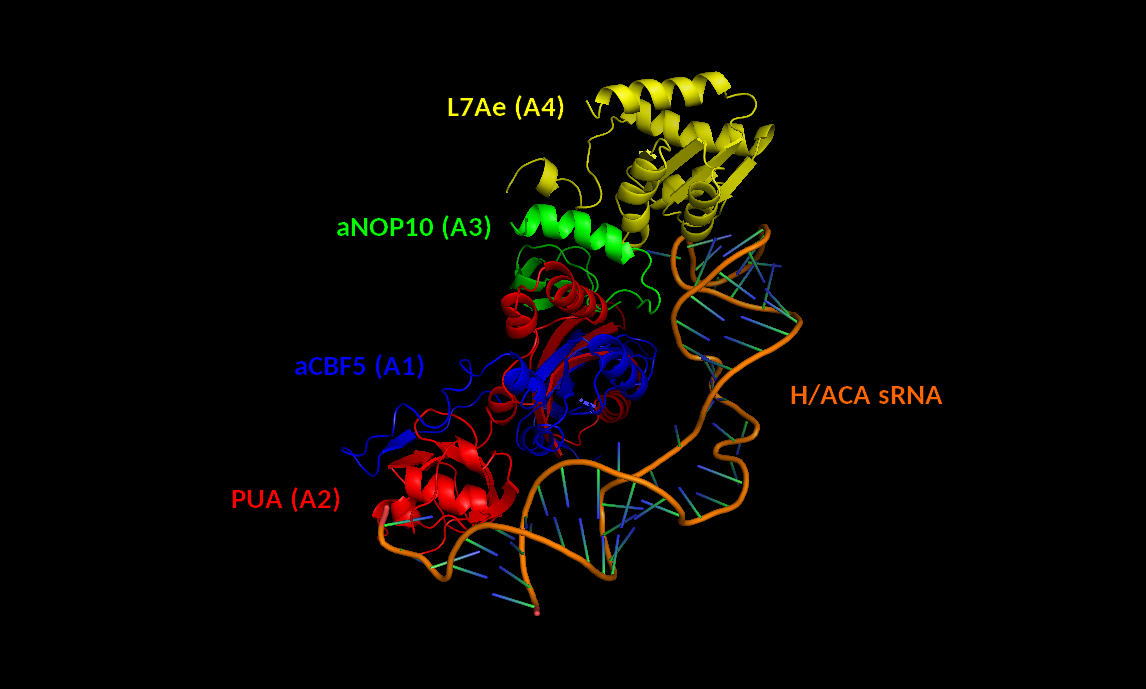


Figure 1: Structure du complexe visualisée sous PyMol.

Chaque domaine est représenté par une couleur différente. Entre parenthèses figurent 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

On estime ici 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 changements de conformation globale ou locale).

Pour cela on calcul l’écart quadratique moyen global des distances entre les atomes en correspondance dans 2 structures.

### RMSD

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 (un atome de la structure 1 et l’atome correspondant de 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.

La signification du RMSD dépend de la taille des protéines mais ici on regarde l’évolution d’une seule protéine. La valeur du RMSD dépend du choix des atomes (ici on regarde par rapport aux centres de masse, mais il est possible de choisir le carbone alpha des résidus comme point de référence dans le programme).

Les structures sont considérées comme identiques si le RMSD est égal à 0 et distantes pour un RMSD supérieur à 4 Å.

# Etude des changements conformationnels

## Changement conformationnel de la protéine

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

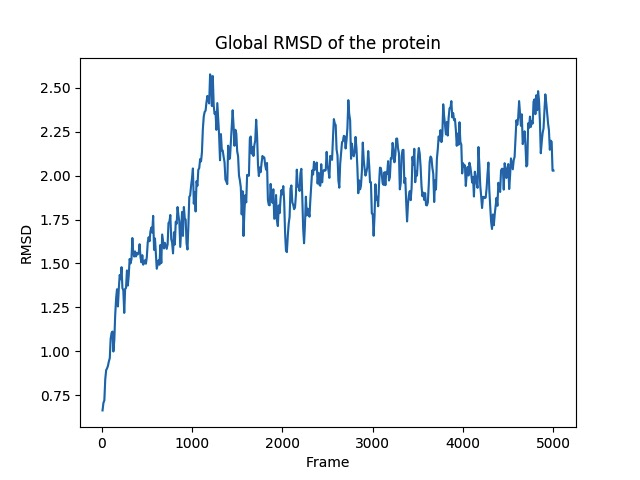
Le RMSD global de la protéine a été calculé entre chacune des 500 conformations de la dynamique et la structure de référence (Figure 2).

Figure 2: RMSD Global.

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 progressive du RMSD pour atteindre un plateau entre 2 et 2.5 Å aux alentours de la conformation (frame 1000). Des changements conformationnels par rapport à la structure de référence sont donc détectés au cours de la dynamique moléculaire. Bien que l’on observe un pic à plus de 2.5 Å, le RMSD ne varie que peu au cours de la dynamique. Dans son ensemble la protéine ne subit donc que peu de modifications au cours du temps.

Changements conformationnels des domaines protéiques

Le RMSD a été calculé pour chacun des 4 domaines protéiques A1, 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.

On remarque que les domaines A1 (aCBF5) et A2 (PUA) varient de la même façon, avec un RMSD moyen 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 de 2.5 Å.

## 

Figure 3: 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).

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

Au début de la dynamique, le domaine A3 (aNOP10) subit le plus de changements conformationnels, à la fin il s’agit de A4 (L7Ae).

Les domaines les plus « flexibles » sont ceux qui interagissent le plus avec le sRNA.

## Conclusion

Nous avons ici observé les résultats de sortie du programme permettant d’étudier si la protéine 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 proche de la structure de référence. Elle peut donc servir de base d’étude pour évaluer l’impact de mutations sur le complexe en solution.

De plus nous avons pu observer que chaque domaine de la protéine a sa propre évolution conformationnelle au cours du temps. On peut donc supposer 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.

# Analyse des changements conformationnels locaux

## Méthode

### Matrice de distance

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

Figure 4 : Matrice des distances entre les résidus de aNOP10 (A3) et l'ARN (B). Les zones bleu foncé indiquent les résidus proches (distance inférieur à 10 Å).

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 (centre de masse ou distance minimale entre les atomes des deux molécules). Les matrices de distances ont ensuite été représentées sous forme de *heatmaps*. La Figure 4 montre les distances entre les résidus du domaine A3 et l’ARN. Les *heatmaps* des autres domaines figurent en Annexe 1.

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 (Figure 5).

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

### 

Figure 5 : Représentation en 3D des résidus 34 à 38 (en rouge) du domaine A3 et des résidus 21 et 22 (en cyan) de 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 protéiques au cours de la dynamique moléculaire. Les résidus dont la frèquence est non-nulle figurent dans le fichier ‘interfaceFreq.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 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.
   2. Si cette distance est inférieure à un seuil fixé par l’utilisateur (par défaut, seuil = 9.0 Å) alors on incrémente N de 1.
2. Une fois toutes les conformations parcourues, la fréquence d’appartenance est calculée :

.

En 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 pour la première conformation, en rouge. Visualisation sous PyMol.

### Temps de contact entre résidus

Enfin nous avons calculé les temps de contacts entre différentes 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 7).

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 est inférieure à 9.0 Å. Le temps de contact T est alors donné par la formule :

## Résultats

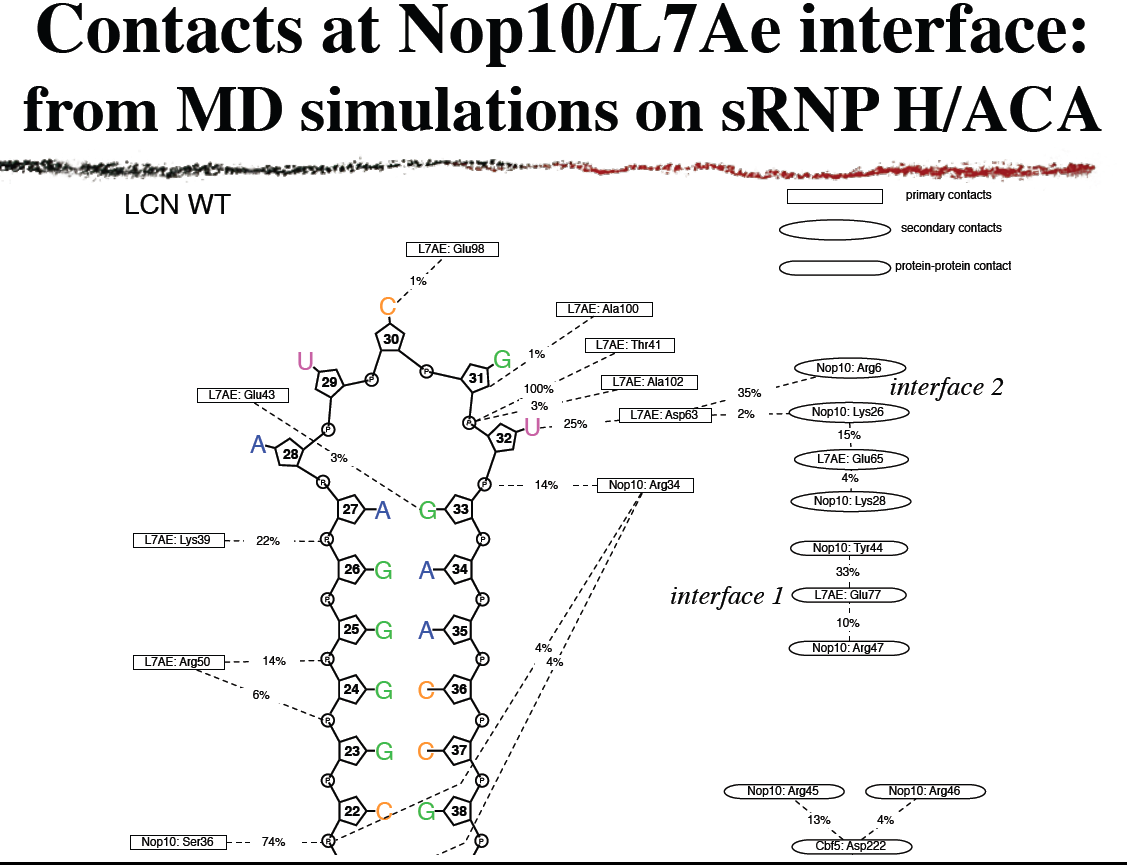
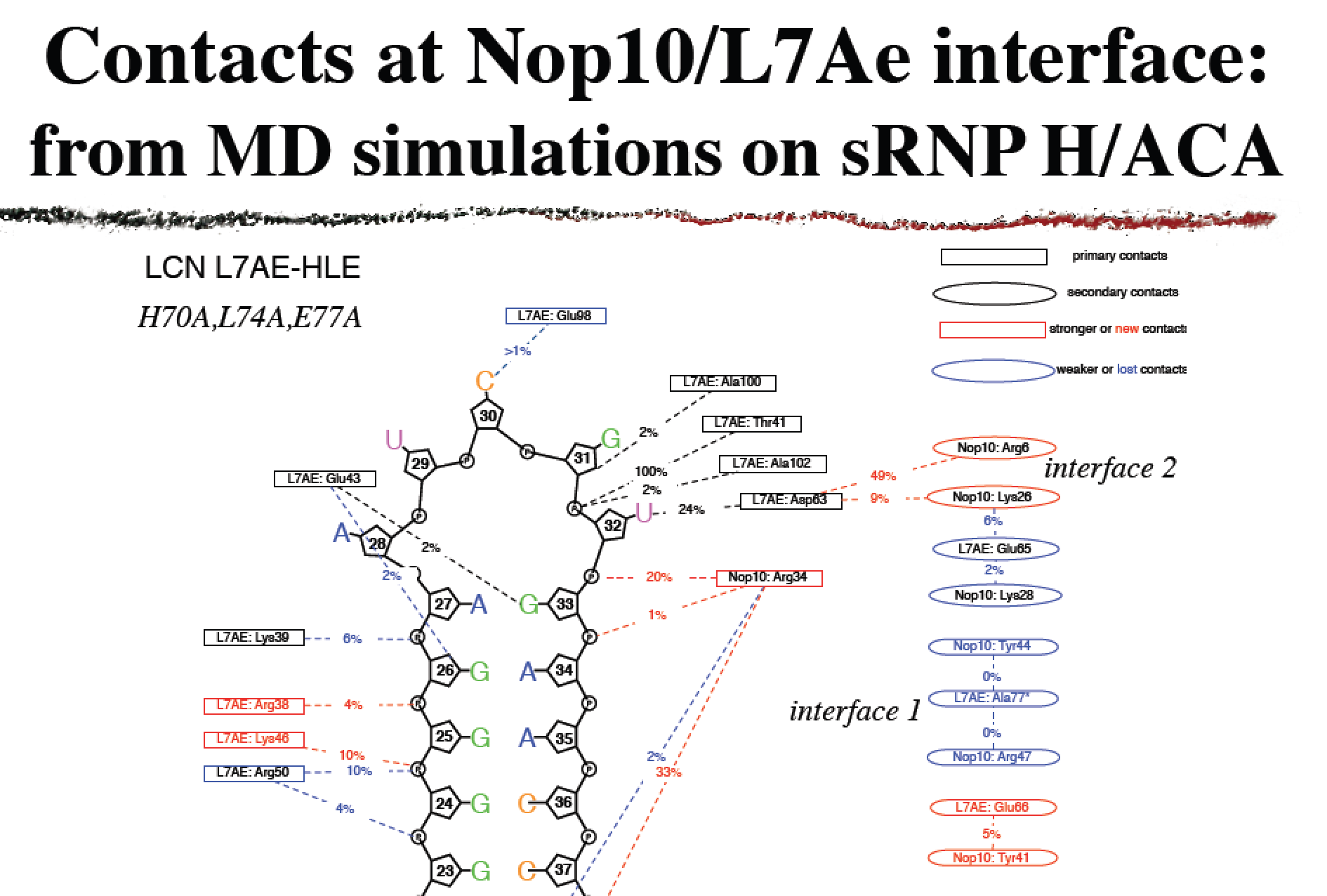
Nous avons choisi de calculer les temps de contacts pour les paires de résidus suivantes :  
L7Ae 41 – ARN 32 L7Ae 100 – ARN 31 L7Ae 46 – ARN 25 aNOP10 34 – ARN 33  
aNOP10 6 – L7Ae 63 L7Ae 66 – aNOP10 41 L7Ae 98 – ARN 30 aNOP10 26 – L7Ae 65

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

Pour que l’ARN se fixe à L7Ae il faut que celui-ci subisse de faibles changements de conformation car la poche de fixation du nucléotide U visé existe avant la fixation du sRNA donc trop de changements entrainent un dérèglement de cette poche et donc affectent l’activité catalytique de aCBF5 car il ne peut fixer le substrat.

Tableau : Temps de contact pour les résidus clé chez le complexe sauvage

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Couple de résidus étudié | Pourcentage de liaison chez le mutant | Pourcentage de liaison chez le sauvage | Temps de contact chez le sauvage (ns) | Effet de la mutation  (ici je conclu par rapport au diapo du prof) |
| L7Ae 41 – ARN 32 | 100 | 100 | 10 | Pas d’effet |
| L7Ae 100 – ARN 31 | 2 | 1 | 10 | Renforcement de la liaison |
| L7Ae 46 – ARN 25 | 10 | inexistant | 9.88 | Création d’une nouvelle liaison |
| aNOP10 34 – ARN 33 | 20 | 14 | 0 | Renforcement de la liaison |
| aNOP10 6 – L7Ae 63 | 49 | 35 | 9.72 | Renforcement de la liaison |
| L7Ae 66 – aNOP10 41 | 5 | inexistant | 6.8 | Création d’une nouvelle liaison |
| L7Ae 98 – ARN 30 | >1 | 1 | 9.7 | Affaiblissement de la liaison |
| aNOP10 26 – L7Ae 65 | 6 | 15 | 8.28 | Affaiblissement de la liaison |



(a) (b)

Figure 7 : Contacts chez le mutant (a) et chez le sauvage (b) entre résidus clés pour l'interaction.

# Conclusion et Perspective

L’étude menée ici permet de voir que chaque domaine doit maintenir une certaine intégrité et qu’ils sont tous peu flexibles. On peut donc penser que le complexe peut devenir « inefficace » facilement.

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

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

Nous pouvons, avec la méthode explicitée ici, analyser différentes mutations du complexe 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.

Avec le programme élaboré nous pouvons donc mettre en œuvre une étude plus poussée des mutations menant à un défaut d’activité du complexe sRNP H/ACA et mieux définir le rôle de chaque domaine.

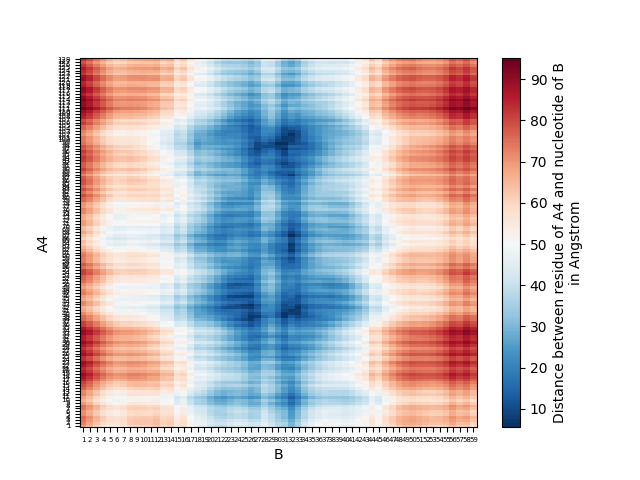
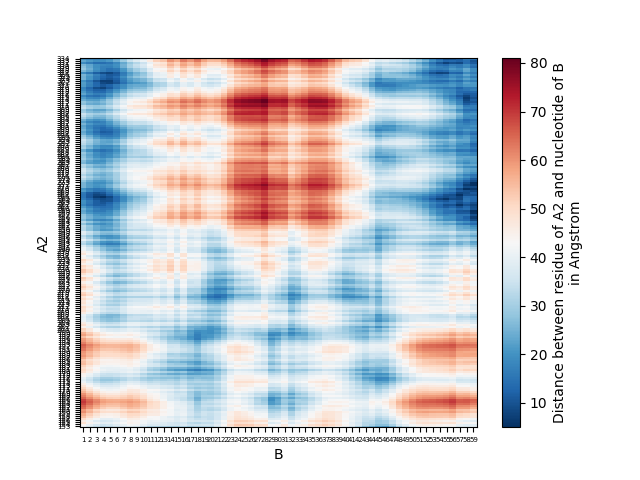
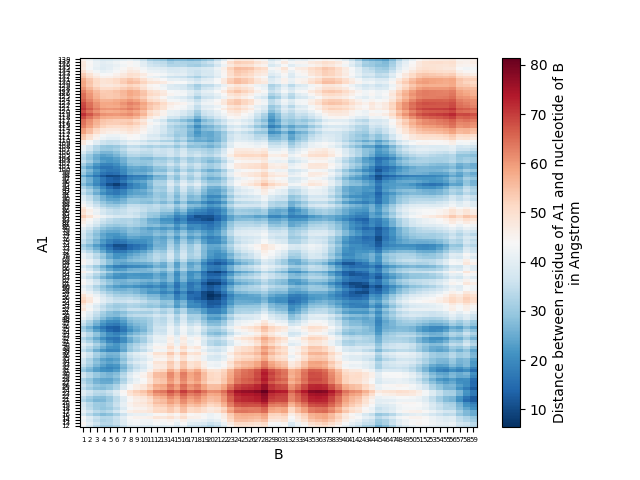
# Annexes

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

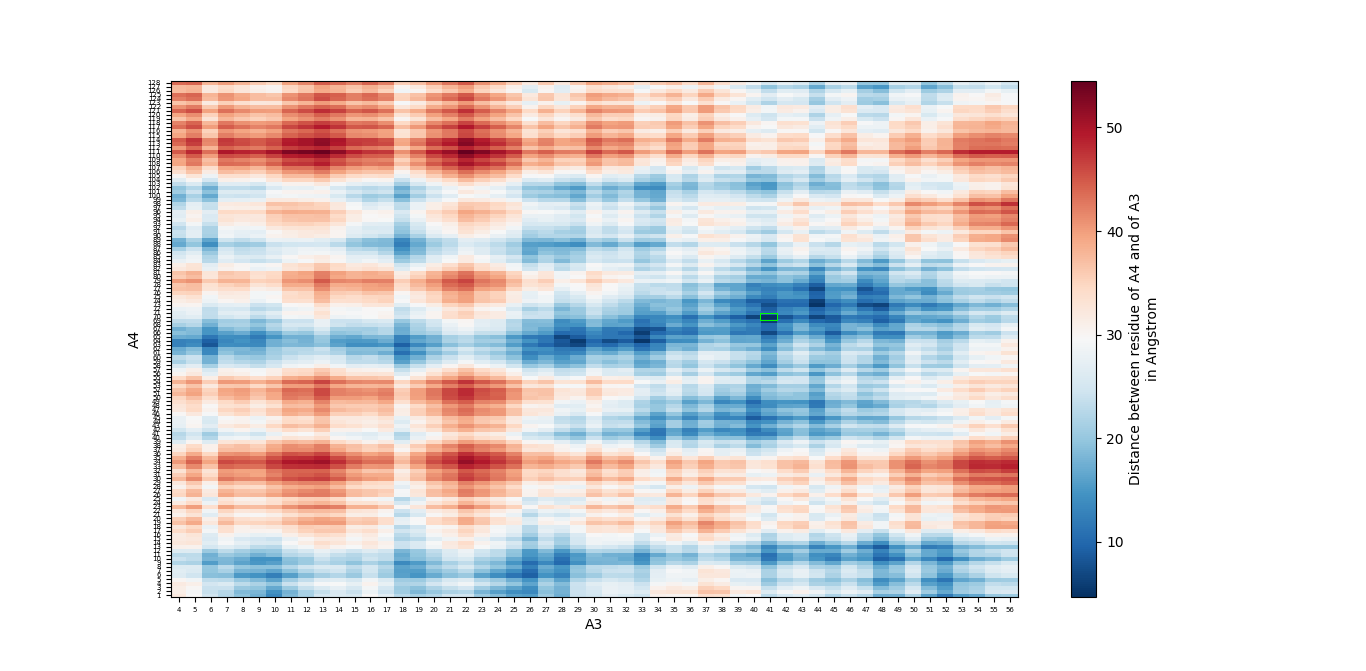
**Annexe 2**: Matrice des distances représentée sous forme de *heatmap* entre les domaines A3 et A4.

**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.