

Modélisation et design d'un motif structural pour l'incorporation de pyrrolysine

Travail à faire en monôme ou binôme. A la fin du TP, envoyer par mail à alain.denise@lri.fr et yann.ponty@lix.polytechnique.fr un fichier (format OpenOffice, LaTeX ou pdf) contenant les résultats de vos expérimentations. Ils pourront être donnés en partie sous forme de copies d'écran commentées.

On sait maintenant qu'aux vingt acides aminés « classiques » on peut en ajouter deux : la Sélénocystéine et la Pyrrolysine. La pyrrolysine, le « 22ème acide aminé », est présente dans des gènes de certaines *archae* méthanogènes (. Une particularité de cet acide aminé est que son codon est aussi un codon stop : UAG. Dans les CDS (Coding DNA sequences) considérés, il y a donc un codon stop UAG en phase. Au cours de la traduction, ce codon peut soit être considéré comme un codon stop « normal » par la machinerie de traduction, auquel cas la traduction s'arrête là, soit considéré comme codant pour une pyrrolysine, auquel cas celle-ci est incorporée dans le peptide en cours de construction et la traduction continue jusqu'au stop suivant.

Selon certains auteurs, une structure secondaire particulière en 3' du codon UAG favoriserait, dans ces *archae*, l'incorporation de la pyrrolysine. Cette hypothèse a été testée dans *A. Théobald-Dietrich, R. Giegé, J. Rudinger-Thirion, Evidence for the existence in mRNA of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins, Biochimie 87 (2005) 813-817*. Dans ce travail, une structure en épingle à cheveux a été déterminée en combinant des expériences de *probing* et des alignements manuels des sites considérés. La figure ci-dessous résume les résultats du *probing*.

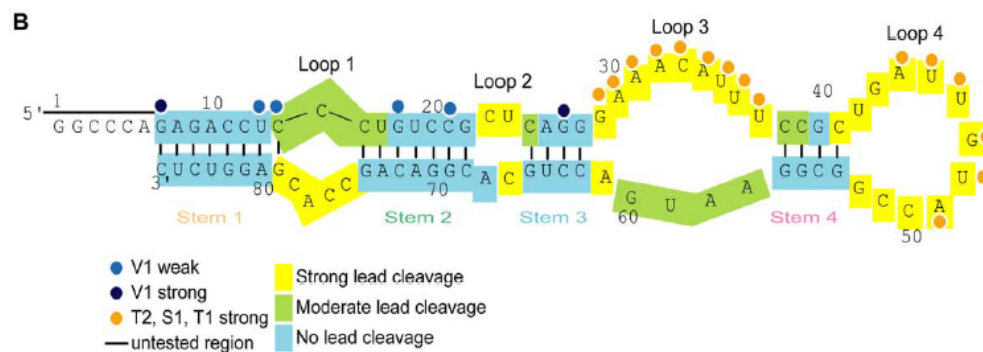


Fig. 1. Probing of PYLIS transcript. (A) Autoradiogram of 12% denaturing polyacrylamide gels from probing experiments on 5'-labeled PYLIS transcript. Probing experiments with Pb^{2+} are in the left part, those with nucleases T1, T2, V1 and S1 are in the right part. Final concentrations of Pb^{2+} are indicated in mM. Control incubations (C) were run in parallel. Lane-S1 checks the effect of $ZnCl_2$ present in nuclease S1 buffer. Lanes AH represent alkaline ladders and lanes G denaturing nuclease T1 ladders. Numbering defines the position of G residues. The double-stranded regions are highlighted in colors. Spontaneous degradations are indicated by d. (B) Summary of the probing experiments displayed on the PYLIS sequence. Nucleotides cut by nucleases S1, T1 and T2 are indifferently indicated by orange dots whereas those cut by nuclease V1 are indicated by blue dots. Yellow areas correspond to residues strongly cut by Pb^{2+} , green areas to nucleotides moderately cut by Pb^{2+} . Regions resistant to Pb^{2+} cleavage are in blue.

La structure qu'ils proposent, commune à au moins six CDS de trois organismes différents, est la sui-

tuelles liaisons non canoniques dans la prédiction. Attention dans ce cas le délai de réponse peut être très long. Laissez la fenêtre du calcul en cours ouverte et rafraîchissez-la de temps en temps.

4. *Prédiction comparative sans alignement préalable.* Allez sur le site de [CARNAC](#). Lisez la présentation, puis choisissez la version en ligne (web server). Lancez l'application sur l'ensemble des cinq séquences. Comparez le consensus trouvé à la structure réelle. Essayez maintenant [RNAshapes](#), [LocARNA](#).
5. *Prédiction comparative avec alignement préalable.* Contrairement à ces programmes, [RNAalifold](#) requiert que les séquences soient préalablement alignées, par exemple avec [CLUSTAL](#). Utilisez-le aussi.
6. *Prédiction comparative (bis).* Remplacez les deux séquences de *M. acetivorans* par celles-ci :


```
>Chimere1
UAGGGCCAGAGACCUCGAGUGACCGUCCGAAAUUUUGGACUGACUGCGCCGGUCCAAUGAGGUCCAGUCACUCCACGAGGUCUCGCAACUAAACG
>Chimere2
UAGGGCCAGAGACCUCGUGUGAGCUGAGCGAAAUUUUCACUGACUGCGCCGGUGAAUGAGCUCACACCCACGAGGUCUCGCAACUAAACG
```

 et relancez les programmes de la question précédente. Y a-t-il des différences dans les prédictions ? Si oui, comment l'expliquez-vous ?

7. *Design naïf de site PYLIS.* Le problème du design d'ARN est le problème inverse de celui de la prédiction de structure : la donnée est une structure d'ARN dont on ne connaît pas forcément la séquence, et on recherche une séquence qui est susceptible de se replier conformément la structure donnée. Nous allons voir ici une approche très naïve, mais dont le principe de base – la modélisation par une grammaire non contextuelle et la génération aléatoire – est utilisé dans de nouvelles approches beaucoup moins naïves de ce problème.

Le logiciel [GenRGenS](#) est capable d'engendrer des séquences structurées, en utilisant des grammaires non contextuelles. Lisez la documentation à l'url :

<http://www.lri.fr/~genrgens/manual/GRGs-manual-html>

et inspirez-vous notamment de [cet exemple](#) pour engendrer 10 séquences dont la structure prédite est proche de celle du site PYLIS. Repliez ces séquences avec [RNAfold](#) et [RNAalifold](#), et comparez le résultat au modèle de structure utilisé pour la génération.

Répétez l'expérience à partir de [NUPack](#), une autre méthode pour le design qui cherche à maximiser la probabilité des séquences engendrées de se replier en la structure cible. Commentez la diversité de 10 séquences engendrées. Qu'en concluez-vous ?

N'oubliez pas d'envoyer par mail à alain.denise@lri.fr et yann.ponty@lix.polytechnique.fr un fichier (format OpenOffice, LaTeX ou pdf) avec les résultats de vos expérimentations.