

Université Paris-Sud 11

DUT Informatique 2011-2012

Bioinformatique

Travaux pratiques 1 : structure de l'ARN

On sait maintenant qu'aux vingt acides aminés « classiques » on peut en ajouter deux : la Sélénocystéine et la Pyrrolysine. La pyrrolysine, le « 22^{ème} 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 dependant 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*.

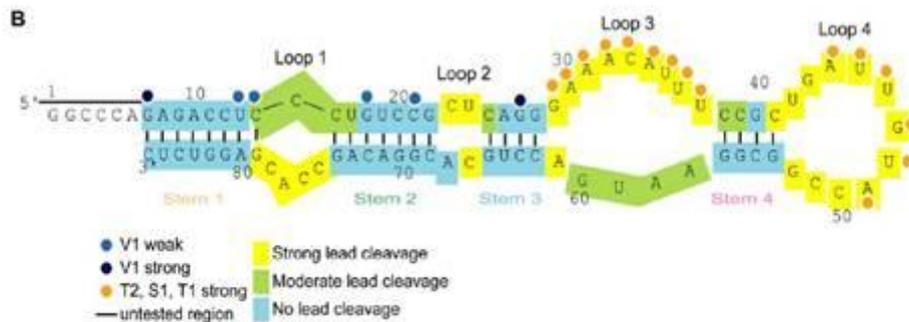
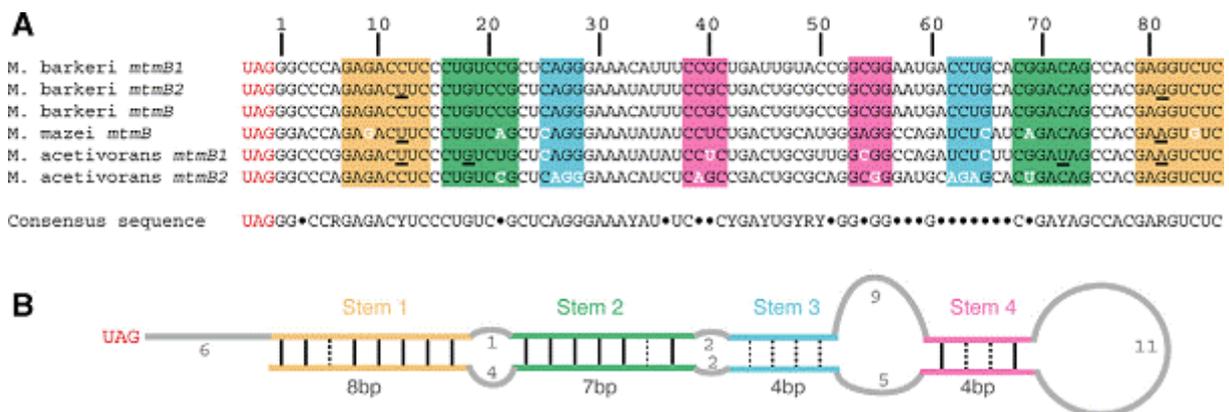


Fig. 1. Probing of PYLIS transcript. (A) Autoradiogram of 12% denaturing polyacrylamide gels from probing experiments on 5'-labeled PYLIS transcript. Probing experiments with Pb^{2+} are in the left part, those with nucleases T1, T2, V1 and S1 are in the right part. Final concentrations of Pb^{2+} are 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 suivante [1] :



Votre travail va consister notamment à tenter de retrouver cette structure en utilisant exclusivement des outils bioinformatiques. On notera que vous ne travaillerez pas vraiment en conditions réelles: en général, on ne connaît pas à l'avance la structure que l'on recherche !

Voici cinq des séquences correspondant à celles-ci-dessus, en format FASTA. (La séquence de *M. Barkeri* mtmB est exactement la même que celle de *M. barkeri* mtmB1, c'est pourquoi il est inutile de la considérer.)

```
>M_barkeri_mtmB1|readthrough
TAGGGCCAGAGACCTCCCTGTCCGCTCAGGGAAATATTTCCGCTGACTGCGCCGGCGGAATG
ACCTGCACGGACAGCCACGAGGTCTCGCAACTAA
>M_barkeri_mtmB2|readthrough
```

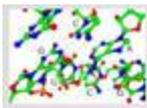
```

TAGGGCCCAGAGACCTCACTGTCCGCCAGGGAAACATTTCCGCTGACTGTGCCGGCGGAATG
ACCTGTACGGACAGTCACGAGGTCTCGCAACTCTG
>M_mazei_mtmB|gene|full_len=1377|start_pos=1
TAGGGACCAGAGACTTCCCTGTCTGCTCAGGGAAATATATCCTCTGACTGCATGGGAGGCCAG
ATCTCATCAGACAGCCACGAAGTGTACAGCTCAA
>M_acetivorans_mtmB1|readthrough
TAGGGCCCGGAGACTTCCCTGTCTGCTCAGGGAAATATATCCTCTGACTGCGTTGGCGGCCAG
ATCTCTTCGGATAGCCACGAAGTCTCCAGCTCAA
>M_acetivorans_mtmB2|readthrough
TAGGGCCCAGAGACCTCCCTGTCCGCTCAGGGAAACATCTCAGCCGACTGCGCAGGCGGGATG
CAGAGCACTGACAGCCACGAGGTCTCGCAACTTAA

```

Tout au long du travail, vous pouvez utiliser VARNA, une applet Java téléchargeable à <http://varna.lri.fr>, pour visualiser les structures secondaires que vous obtiendrez.

1. Prédiction comparative de structure, phase 1. Allez sur le site de CARNAC : <http://bioinfo.lifl.fr/RNA/carnac/>. Lisez la présentation, puis choisissez la version en ligne (web server). Lancez l'application sur l'ensemble des cinq séquences. Téléchargez sur le même site le programme RNAfamily. Il utilise les fichiers au format CT (voir image ci-dessous) générés par CARNAC, pour visualiser les structures communes trouvées dans les 5 séquences. Vous remarquez des éléments de structure communs dans les cent premiers nucléotides (environ) en aval de l'UAG.
2. Prédiction ab initio. Allez sur le serveur de mfold (<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>). Repliez les cinq séquences avec mfold. Comparez les résultats, visuellement, à la structure recherchée et aux structures données par CARNAC (vous pouvez utiliser RNAfamily).
3. Comparaison de structures. Pour comparer plus objectivement les structures obtenues à la structure recherchée, utilisez RNAForester. (<http://bibiserv.techfak.uni-bielefeld.de/rnaforester/submission.html>). Il vous faudra pour cela écrire à la main la structure recherchée (points et parenthèses), suivant le dessin **B** ci-dessus.
4. Prédiction comparative. Allez sur le site de CARNAC : <http://bioinfo.lifl.fr/RNA/carnac/>. Lisez la présentation, puis choisissez la version en ligne (web server). Lancez l'application sur l'ensemble des cinq séquences. Téléchargez sur le même site le programme RNAfamily. Il utilise les fichiers au format CT (voir image ci-dessous) générés par CARNAC, pour visualiser les structures communes trouvées dans les 5 séquences. Utilisez aussi RNASHAPES (<http://bibiserv.techfak.uni-bielefeld.de/rnashapes/submission.html>) puis RNAALIFOLD (<http://rna.tbi.univie.ac.at/cgi-bin/alifold.cgi>), en faisant varier les paramètres et l'ensemble des séquences (par exemple, supprimez une ou plusieurs séquences de l'ensemble). Notez que RNAALIFOLD requiert que les séquences soient préalablement alignées, avec CLUSTAL W par exemple (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Comparez chacun de vos résultats à la structure vraie.



carnac

[home](#) [web server](#) [help](#) [examples](#) [retrieve result with an ID](#)

- bioinfo.lifl.fr**
 - Sequoia
- mreps**
- YASS**
- Magnolia**
- Proteins**
 - Path
 - Protea
 - ReBLOSUM
- RNA**
 - Carnac

Results for job Mar_05_2010_10_35_52_20006

- M_mazei_mtmB|readthrough|full_len=774|start_pos=1
CT file , PS file , JPEG file, [bracket notation](#), list of stems
- M_acevivorans_mtmB2|readthrough|full_len=774|start_pos=1
CT file , PS file , JPEG file, [bracket notation](#), list of stems
- M_barkeri_mtmB2|readthrough|full_len=774|start_pos=1
CT file , PS file , JPEG file, [bracket notation](#), list of stems
- M_acevivorans_mtmB1|readthrough|full_len=774|start_pos=1
CT file , PS file , JPEG file, [bracket notation](#), list of stems
- M_barkeri_mtmB1|readthrough|full_len=774|start_pos=1
[CT file](#) , PS file , JPEG file, [bracket notation](#), list of stems



[1] On remarquera cependant que, bien qu'il y ait de fortes présomptions sur le fait que la structure proposée existe et est commune aux six CDS, elle n'a pas été totalement validée.

[2] Il est possible que certaines séquences diffèrent entre la base et la figure ci-dessus car les données ne proviennent pas des mêmes sources.