

Examen de théorie des langages

La clarté et la précision de la rédaction seront prises en compte dans l'évaluation.

Le barème est indicatif.

Aucun document autorisé.

Machines (ordinateurs et calculatrices) interdites.

Durée : 1h 30min

L'acide ribonucléique (ARN) est une molécule biologique qui permet l'expression du code génétique. Il est composé de 4 bases nucléiques qui sont l'adénine, la cytosine, la guanine et l'uracile que l'on notera respectivement **a**, **c**, **g** et **u**.

Dans la suite, on considèrera qu'un brin d'ARN est un mot sur l'alphabet $\{a, c, g, u\}$.

Phases ouvertes de lecture (2.5+2.5)

L'ARN est lu séquentiellement par groupes de 3 bases appelés triplets (ou codons). On admettra qu'une phase ouverte de lecture commence à un triplet **aug** (codon start) et qu'elle se poursuit jusqu'à un codon stop (**uaa**, **uag** ou **uga**). Les codons situés entre le codon start et le codon stop sont quelconques (excepté qu'ils ne sont pas des codons stop). Il peut éventuellement n'y avoir aucun codon entre le start et le stop.

- 1) Donner une grammaire lexicale qui décrit ce langage.
- 2) Donner une grammaire syntaxique.

Hélices (2+1+2+1+4)

Les propriétés chimiques des bases font que l'ARN se replie sur lui-même à cause de la complémentarité entre **a** et **u** d'une part et **c** et **g** d'autre part. On appellera ici hélice une séquence contiguë de **g** en relation avec une suite contiguë de **c** située en aval. On appellera les nucléotides non appariés (ici des **a** et des **u**) une boucle.

```
+-----+
| +-----+ |
| | +-----+ | |
| | |           | | |
+---+---+---+---+---+---+
g g g a u a u u c c c
```

- 1) Donner une grammaire pour le langage des mots qui contiennent exactement une hélice (avec au moins un g) et une boucle (avec au moins un a ou un u). (La boucle est donc "contenue" dans l'hélice.)
- 2) Donner le type de la grammaire.
- 3) Donner une machine abstraite qui reconnaît ce langage.
- 4) Rappeler les conditions pour appliquer l'algorithme CKY. Sont-elles vérifiées ici ?
- 5) Montrer que ggaucc appartient à ce langage en appliquant CKY.

Pseudo-noeuds (1+4)

On appelle pseudo-noeud un repliement particulier dans lequel les bases d'une boucle forment elles-mêmes une hélice avec des bases situées en aval.

On considèrera ici que la partie gauche de cette deuxième hélice est constituée de toutes les bases de la boucle et que celle-ci n'est composée que de a en relation avec des u.

Chacune des deux hélices peut éventuellement être vide.

```

+-----+
| +-----+ |
| | +-----+ | |
| | | | | |
+-----+
g g g a a c c c u u
      | | | | |
      | +-----+ |
      +-----+

```

- 1) Donner un argument (formel ou informel) qui montre que le langage des pseudo-noeuds n'est ni de type 3, ni de type 2.
- 2) Donner une machine de Turing qui le reconnaît.

