

Octobre 2002

Feuille d'exercices n° 2

- 1 – En biologie, l'ADN est codé par une suite de nucléotides. On définit les nucléotides comme des éléments pris dans l'alphabet $V = \{ A, T, C, G \}$. Un triplet de nucléotides (ex. AAT) est appelé codon. Pour détecter une protéine dans un brin d'ADN on doit être capable de reconnaître son commencement et sa fin. La fin d'une protéine est marquée par un codon STOP codé par **TAA**, **TAG**, ou **TGA**. Le début d'une protéine est déterminé par de nombreux facteurs : pour simplifier nous considérerons qu'il n'est marqué que par le codon **ATG**.

Nous considérerons qu'il n'existe qu'une seule phase pour lire un brin d'ADN donné (la séquence AATGCCA... peut se lire AAT GCC A... (phase 1) ou ...A **ATG** CCA ... (phase 2) ou ...AA TGC CA... (phase 3)) : toute séquence sera donc lue en phase 1. Une protéine est donc du type :

CCA **ATG** CTA CTC CCA ACT GCC **TAA** CTT... (TAA pouvant être remplacé par TAG ou TGA).

Donnez une **grammaire** définissant les protéines.

- 2 – Définissez une **grammaire** associée au langage des expressions arithmétiques.
- 3 – Définissez une **grammaire formelle** permettant d'engendrer les identificateurs de variables dans un langage de programmation tel que le C. Les identificateurs débutent tous par une lettre, puis sont composés de lettres, de chiffres ou du caractère « _ ».
- 4 – On définit un langage de marquage proche du HTML. Une page de code PseudoHTML est composée de tags permettant de mettre en forme du texte. Il existe un tag de départ `<tag>` et un tag de fin `</tag>` qui est le nom du tag de départ inversé. Un tag de départ peut recevoir un nombre illimité de paramètres ; il est de la forme : `<tag param1=chaîne1 ... paramn=chaînen>`. Pour simplifier, nous considérerons qu'un tag et un paramètre sont des identificateurs (voir 3) et qu'une chaîne est une suite finie contenant n'importe quel caractère imprimable.

La structuration d'une page PseudoHTML est donc la suivante :

```
<pseudohtml>
  <body>
    <tag1 param1=chaîne1> chaîne </tag1>
    ...
    <tagn> chaîne </tagn>
  </body>
</pseudohtml>
```

Donnez une **grammaire** vérifiant la syntaxe d'un fichier PseudoHTML.

- 5 – Pour les langages sur $X = \{a, b, c\}$ définis ci-dessous, définissez une grammaire les engendrant. Vous prouverez qu'il y a bien équivalence entre les langages considérés et les langages engendrés. Aussi, il vous faut trouver les grammaires qui facilitent au mieux ces preuves. Enfin, vous donnerez le type des différentes grammaires définies.

- $L_1 = \{ a^n b^n / n \geq 0 \}$
- $L_2 = \{ a^n b^p c^q / n \geq 1, p \geq 1, q \geq 1 \}$
- $L_3 = \{ a^n b^p / n > p > 0 \}$
- $L_4 = \{ a^n b^m / n \neq m, n, m \geq 0 \}$
- $L_5 = \{ a^n b^p c^q d^n / n, p, q > 0 \} \cup \{ a^p b^n c^q d^n / n, p, q > 0 \}$
- $L_6 = \{ a^n b^p c^q / n, q \geq 0, p \geq n+q \}$
- $L_7 = \{ a^n b^p / n \neq p+3 \}$
- $L_8 = \{ a^n b^n c^p / n, p \geq 1 \}$
- $L_9 = \{ a^n b^n c^n / n \geq 1 \}$