

## HLIN608 – algorithmique du texte – exercices sur l'alignement

**Exercice 1 :** L'algorithme de programmation dynamique pour chercher l'alignement optimal permet également de trouver les plus longues sous-séquences communes à deux séquences, selon les modifications suivantes :

- les substitutions de caractères différentes ne sont plus utilisées
- deux caractères identiques apportent un score de 1
- un indel ne coûte rien et n'apporte rien.

Utiliser cette méthode pour trouver la longueur des plus longues sous-séquences communes entre AGCTGA et CAGATCAGAG, puis extraire ces sous-séquences.

**Exercice 2 :** L'algorithme AA suivant construit un automate.

AUTO-ALIGN-OPT( $x, m, y, n, T$ )

```
1   $M \leftarrow$  NOUVEL-AUTOMATE()
2  initialiser  $E$ 
3   $E[-1, -1] \leftarrow initial[M]$ 
4   $E[m - 1, n - 1] \leftarrow$  NOUVEL-ÉTAT()
5   $terminal[E[m - 1, n - 1]] \leftarrow$  VRAI
6  AA( $m - 1, n - 1$ )
7  retourner  $M$ 
```

AA( $i, j$ )

```
1  si  $i \neq -1$  et  $j \neq -1$ 
   et  $T[i, j] = T[i - 1, j - 1] + Sub(x[i], y[j])$  alors
2      si  $E[i - 1, j - 1] = NIL$  alors
3           $E[i - 1, j - 1] \leftarrow$  NOUVEL-ÉTAT()
4          AA( $i - 1, j - 1$ )
5           $Succ[E[i - 1, j - 1]] \leftarrow$ 
              $Succ[E[i - 1, j - 1]] \cup \{(x[i], y[j]), E[i, j]\}$ 
6  si  $i \neq -1$ 
   et  $T[i, j] = T[i - 1, j] + Dél(x[i])$  alors
7      si  $E[i - 1, j] = NIL$  alors
8           $E[i - 1, j] \leftarrow$  NOUVEL-ÉTAT()
9          AA( $i - 1, j$ )
10      $Succ[E[i - 1, j]] \leftarrow Succ[E[i - 1, j]] \cup \{(x[i], \epsilon), E[i, j]\}$ 
11  si  $j \neq -1$ 
   et  $T[i, j] = T[i, j - 1] + Ins(y[j])$  alors
12     si  $E[i, j - 1] = NIL$  alors
13          $E[i, j - 1] \leftarrow$  NOUVEL-ÉTAT()
14         AA( $i, j - 1$ )
15      $Succ[E[i, j - 1]] \leftarrow Succ[E[i, j - 1]] \cup \{(\epsilon, y[j]), E[i, j]\}$ 
```

A quoi sert cet automate ? Construire l'automate correspondant à l'exercice 1 de la feuille de TD précédente.