

2I003 – Interrogation sur le Projet (20 min)

14 décembre 2017

1 Introduction et rappels.

Une séquence d'ADN est un mot sur l'alphabet de nucléotides $\mathcal{A} = \{A, C, G, T\}$. Les nucléotides se regroupent par paires $\{A, T\}$ et $\{C, G\}$. La structure secondaire S d'une séquence d'ADN $a = a_1 \dots a_n$ de taille n est l'ensemble des couples (i, j) , $1 \leq i < j \leq n$ tels que :

- $\forall 1 \leq i \leq n$, il existe au plus un couple de S contenant i ,
- il n'existe pas de noeud, c'est-à-dire de couples (i_1, j_1) et (i_2, j_2) tels que $i_1 < i_2 < j_1 < j_2$,
- tout couple $(i, j) \in S$ satisfait $\{a_i, a_j\} \in \{\{A, T\}, \{C, G\}\}$.

On note S_{ij} la structure secondaire de taille maximale de la sous-séquence $a_i \dots a_j$, et $E_{ij} = |S_{ij}|$. On suppose qu'il existe une fonction $e(x, y)$ qui renvoie 1 si (x, y) forme une paire de nucléotide, et 0 sinon. On donne enfin deux fonctions pour calculer la taille de la structure maximale. Une fonction récursive

```
def tailleMaxRec(a, i, j):
    if (i >= j):
        return 0
    m = tailleMaxRec(a, i+2, j-1) + e(a[i-1], a[j-1])
    for k in range(i+1, j+1): # toutes les valeurs de k entre i+1 et j
        val = tailleMaxRec(a, i, k-1) + tailleMaxRec(a, k, j)
        if val > m:
            m = val
    return m
```

et une fonction itérative (qui calcule *toutes* les valeurs des $E_{i,j}$ et les stocke dans un dictionnaire) :

```
def tailleMaxIter(a, n):
    E = dict() # on cree un dictionnaire E
    E[(1, 1)] = 0
    for i in range(2, n+1): # toutes les valeurs entre 2 et n
        E[(i, i)] = 0
        E[(i, i-1)] = 0
    for p in range(1, n): # toutes les valeurs entre 1 et n-1
        for i in range(1, n-p+1): # toutes les valeurs entre 1 et n-p
            m = E[(i+1, i+p-1)] + e(a[i-1], a[i+p-1])
            for k in range(i+1, i+p+1):
                val = E[(i, k-1)] + E[(k, i+p)]
                if val > m:
                    m = val
            E[(i, i+p)] = m
    return E
```

2 Questions.

Question 1. Soit $a = TCGGCTGCATTTCTGA$. Que valent $S_{1,15}$ et $E_{1,15}$?

Question 2. Soit a une séquence de taille n et $1 \leq i < j \leq n$. Quelle est la valeur maximale de E_{ij} ?

Question 3. Rappeler sans calcul les complexités des deux fonctions `tailleMaxRec` et `tailleMaxIter`.

Question 4. Quelle structure de données (liste simplement chaînée, liste doublement chaînée circulaire, tableau, arbre, etc.) utiliseriez-vous pour implanter la séquence a ? Argumentez votre réponse.

Question 5. Vous devez vérifier expérimentalement qu'une fonction a une complexité exponentielle en un paramètre n . Expliquez votre protocole en quelques lignes.

Question 6 (difficile). Dans un programme C , par défaut la pile d'exécution a pour taille maximale $1\text{Mo} = 2^{20}$ octets. Supposons que chaque appel récursif a une empreinte mémoire de 16 octets. Quelle taille maximale d'ADN pouvez-vous espérer traiter avec l'algorithme récursif `tailleMaxRec` (sans considération de temps) ?
Indication : de quel paramètre de l'arbre des appels récursifs déduit-on l'espace occupé sur la pile d'exécution ?