

# Informatique – Biologie L2 – TP3 : Programmation Python

Jean-Baptiste LAMY (jiba@tuxfamily.org)

28 février 2007

Nous allons travailler sur les séquences d'ADN suivantes (brin codant, à télécharger sur le campus virtuel, adresse <http://www-limbio.smbh.univ-paris13.fr/campusvirtuel/>, fichier tp3\_données.py; ce fichier contient aussi les fonctions de calcul du pourcentage de GC, de transcription et de traduction réalisées lors du TP2 et qui seront réutilisées) :

```
adn1 = "atgagtgaacgtctgagcattaccccgctggggccgtatatcggcgca"
adn2 = "atgagtgaacgtctgagcattaccccgctggggccgtatatcggcgcacataacgg"
adn3 = "atgagtgaacgt"
adn4 = "atgagt"
adn5 = "atgagcgagaggctgagcattaccccgctggggccgtatatcggcgcacataacccaaggt"
```

## 1 Calcul du point de fusion (Tm) de l'ADN

Plusieurs formules existent pour calculer le point de fusion (Tm) d'une chaîne d'ADN.

### 1.1 Règle de Wallace

Cette règle fonctionne sur des oligonucléotides très courts ( $\leq 15$  bases). La formule est la suivante :

$$Tm = 8 + 2 \times (nbAT) + 4 \times (nbGC)$$

où nbAT est le nombre de AT, et nbGC le nombre de GC.

Écrire une fonction pour calculer le point de fusion d'un ADN avec la règle de Wallace. Écrire une fonction pour calculer le point de fusion d'un ADN avec la formule de Howley. Utiliser votre fonction pour calculer le point de fusion de l'ADN 3.

### 1.2 Calcul par le contenu en GC (Howley *et al.*)

Cette formule est valable pour les longs oligonucléotides ( $> 10$  bases). La formule est la suivante :

$$Tm = 67,5 + (0,34 \times \%GC) - (395/nbbases)$$

où %GC est le pourcentage de GC, et nbbases le nombre de base dans la séquence.

Écrire une fonction pour calculer le point de fusion d'un ADN avec la formule de Howley. Utiliser votre fonction pour calculer le point de fusion de l'ADN 3.

### 1.3 Méthode du plus proche voisin

Cette méthode est valable pour les séquences entre 20 et 60 nucléotides. Elle s'appuie sur les propriétés thermodynamiques des dinucléotides (paires de nucléotides voisins).

$$Tm = \frac{1000 \times dH}{A + dS + R \times \ln(\frac{C}{4})} - 273,15 + 16,6 \times \log(cK^+)$$

où :

- dH est la somme des enthalpies de chaque dinucléotide
- dS est la somme des entropies de chaque dinucléotide
- A est l'entropie de la formation de l'hélice de l'ADN (constante -10,8 cal)
- R est la constante des gaz parfaits (1,984 cal/grad x mol)
- C est la concentration en oligonucléotide (nous prendrons 250 pmol/l)
- cK<sup>+</sup> est Concentration en ions potassium dans la solution (nous prendrons 50 mmol/l)

Les enthalpies et entropies des dinucléotides sont données par les dictionnaires suivants :

```
enthalpies_des_dinucléotides = {
    "aa" = -9.1,
    "ag" = -7.8,
    "ac" = -6.5,
```

```

...
}
entropies_des_dinucleotides = {
    "aa" = -24.0,
    "ag" = -20.8,
    "ac" = -17.3,
    ...
}

```

Comment faire pour lister les dinucléotides d'une séquence d'ADN ?

Écrire une fonction pour calculer le point de fusion d'un ADN avec la méthode du plus proche voisin. Pour calculer le log et le ln, il faut :

- placer "import math" au début de votre fichier
- ln s'obtient alors avec "math.log"
- log s'obtient alors avec "math.log10"

Utiliser votre fonction pour calculer le point de fusion de l'ADN 1.

## 1.4 Fonction générique

Écrire une fonction qui calcule le point de fusion d'un ADN, en utilisant automatiquement la formule la plus appropriée. En utilisera en priorité la méthode du plus proche voisin, puis la formule de Howley, puis celle de Wallace. Utiliser cette fonction pour calculer le point de fusion des ADN 1, 2, 3 et 4.

## 1.5 Fonction générique (2)

Modifier la fonction précédente pour faire la moyenne des formules de Howley et de Wallace, dans le cas où les deux peuvent être utilisées. Utiliser cette fonction pour calculer le point de fusion des ADN 1, 2, 3 et 4.

# 2 Recherche de motifs au sein d'une protéine

## 2.1 Motif [ST]-x-[RK]

Dans une chaîne protéique, le motif [ST]-x-[RK] correspond à un site de phosphorylation par la protéine-kinase C. Dans le motif, "x" correspond à n'importe quel acide aminé, "S" correspond à la Sérine (et de même pour les autres acides aminés), et "[ST]" correspond à une Sérine OU une Thréonine. Les tirets "-" sont utilisés pour séparer les acides aminés.

Écrire une fonction pour recherche le motif [ST]-x-[RK] dans une protéine. La fonction doit retourner l'indice de l'acide aminé où commence le motif, ou -1 si le motif est absent.

Transcrire et traduire les ADN 2 et 4, et rechercher le motif [ST]-x-[RK] dans les protéines ainsi obtenues.

## 2.2 Motifs génériques

Écrire une fonction pour recherche un motif dans une protéine, fonctionnant pour n'importe quel motif. On pourra s'aider du squelette suivant (remplacer les \_\_\_ par le code approprié) :

```

def recherche_motif(_____):
    aas = motif.split(_____)
    for i in range(_____):
        for j in range(len(aas)):
            aa = aas[j]
            if aa[0] == "[":
                if not proteine[_____] in aa:
                    break
            elif aa != "x":
                if not proteine[_____] == aa:
                    break
        else:
            return i
    return _____

```

Utiliser cette fonction pour rechercher dans les protéines codés par les ADN 2 et 4 les motifs "[ST]-x-[RK]", "S-x-x-P" et "P-[PY]-x-G-[KA]".