

# Révisions - TD

Les exercices s'appuient sur la notion d'[adn](#). On supposera dans l'ensemble des exercices qu'un adn est codé sous la forme d'une chaîne de caractères contenant uniquement les lettres A, T, G, C correspondant aux différentes bases (adénine, cytosine, guanine ou thymine).

## 1 Exercice 1

Définir la fonction `est_base` prenant en paramètre un caractère et retournant `True` si ce caractère correspond à une base (est un des caractères A, T, G, C), et `False` sinon.

Quelle est la complexité de cette fonction ?

## 2 Exercice 2

Définir la fonction `est_adn` prenant en paramètre une chaîne de caractères et retournant `True` si la chaîne correspond à un adn (est constituée uniquement des caractères A, T, G, C), et `False` sinon.

Quelle est la complexité de cette fonction ?

## 3 Exercice 3

Définir la fonction `pourcentages_bases` prenant en paramètre un adn et retournant un dictionnaire contenant comme information le pourcentage de chaque base. Par exemple, pour l'adn AAATCGTT, la fonction doit renvoyer le dictionnaire :

```
{ 'A': 37.5, 'T': 37.5, 'C': 12.5, 'G': 12.5 }.
```

Quelle est la complexité de cette fonction ?

## 4 Exercice 4

Écrire un programme demandant à l'utilisateur de saisir une séquence d'adn (saisie répétée tant que l'utilisateur n'a pas saisi un brin d'adn avec au minimum 8 bases) et affiche le pourcentage de chaque base dans la saisie.

## 5 Exercice 5

La recherche de caractère génétique (recherche de maladie génétique ou recherche de lien de parenté) consiste à déterminer si un adn contient une séquence de bases (correspondant à la signature d'une maladie génétique par exemple).

Définir la fonction `contient_sequence` prenant en paramètre un adn et une séquence de bases (également codée sous la forme d'une chaîne de caractères contenant uniquement les caractères A, T, G, C). La fonction retourne `True` si l'adn contient cette séquence, et `False` sinon.

Quelle est la complexité de cette fonction ?

## 6 Exercice 6

En pratique, il peut y avoir des erreurs lors de la réplication de l'adn. Cela signifie que quelques bases peuvent changer. Pour la recherche de séquence, il faut tenir compte de cette possibilité.

On dit qu'un adn "contient" une séquence s'il existe une portion de l'adn qui est identique à 80% minimum à la séquence recherchée. Par exemple, on dit que AAATCGTT "contient" la séquence AACGT puisque l'adn contient ATCGT qui est identique à AACGT excepté une lettre (80% des lettres sont identiques).

Définir la fonction `contient_sequence_similaire` prenant en paramètre un adn et une séquence de bases et retourne `True` si l'adn contient une portion identique à au moins 80% à la séquence, et `False` sinon.

Quelle est la complexité de cette fonction ?

## 7 Exercice 7

*Remarque :* Cet exercice est difficile et n'est à faire que si le reste est compris et terminé.

On veut rechercher si un adn contient une portion identique à une séquence donnée. Cependant, on suppose que la portion peut être découpée en deux endroits distincts de l'adn (la première partie doit se trouver avant la deuxième partie).

Par exemple, l'adn AAATGCCGTAAATC contient AATGAAATC en deux parties AATG puis AAATC.

Définir la fonction `contient_deux_parties` prenant en paramètre un adn et une séquence et retournant `True` si l'adn contient cette séquence en une ou deux parties, et `False` sinon.

Quelle est la complexité de cette fonction ?