

IFT 1015 – Programmation 1

TP 1

- À faire en groupe de **deux** étudiants
- Remise le 28.03 à **23 :59** au plus tard

1. Objectif du TP

Le but du TP est de partir d'un brin d'ADN pour arriver aux protéines codées par les gènes contenus dans ce brin d'ADN. Pour cela il faudra effectuer la transcription : passer de l'ADN à l'ARN. Ensuite il faudra effectuer la traduction : passer de l'ARN à la protéine.

2. Introduction

L'ADN est composé de deux brins complémentaires composés d'une suite de nucléotides : A, T, C et G.

Sur les deux brins complémentaires, il y a toujours un A en face d'un T et un C en face d'un G.

Exemple :

```
ATCAATTGCAAGTTCC
TAGTTAACGTTCAAGG
```

Les deux brins se lisent en sens inverse. Ainsi le brin supérieur se lit de gauche à droite et le brin inférieur de droite à gauche.

Les deux brins peuvent avoir des gènes.

Le brin fourni dans cet exercice est celui se lisant de gauche à droite (le brin supérieur).

Vous devrez générer le brin complémentaire pour vérifier s'il y a des gènes dessus. N'oubliez pas que deux étapes seront nécessaires :

- Écrire le brin complémentaire
- Le lire en sens inverse ou le retourner pour le lire de gauche à droite

Un gène débute toujours par la même séquence de trois nucléotides : TAC.

Il y a ensuite un nombre de nucléotides quelconque mais toujours un multiple de trois. On trouve ensuite une séquence permettant de terminer le gène, cette séquence peut être ATT, ATC ou ACT.

Un gène est donc TAC (xxx) ATT ou TAC (xxx) ATC ou TAC (xxx) ACT. (xxx) étant un nombre quelconque de nucléotides mais toujours un multiple de trois.

Une fois le gène isolé, il doit être transformé en ARN. L'ARN est une suite de nucléotides aussi mais ce sont A, C, G et U (qui remplace T). Donc si le gène isolé est ATGAAACCCGGGTTTAA, l'ARN correspondant sera complémentaire :

ATGAAACCCGGGTTTAA

UACUUUGGCCCAAATT

À partir de l'ARN, les nucléotides sont lus trois par trois. Chaque combinaison de trois correspond à un acide aminé. Les acides aminés sont ensuite chainés les uns à la suite des autres.

Les correspondances entre les acides aminés et les combinaisons de trois nucléotides de l'ARN sont fournies dans le fichier Python de départ sous forme d'un tableau associatif.

À la fin d'un gène se trouve toujours trois nucléotides signifiant l'arrêt de la transcription (codon STOP). Ces trois nucléotides ne représentent pas un acide aminé mais signalent seulement l'arrêt. Il y a trois codons signalant l'arrêt, ils sont aussi dans le tableau associatif.

Vous devrez donc créer une chaîne de caractères représentant la protéine (chaîne d'acides aminés) puis l'afficher dans la console sous la forme :

Méthionine (Start)-Proline-Thrénine-Aspartate-Arginine-
Sérine-Aspartate-Arginine-Glutamine-Leucine-Alanine-Sérine-
Sérine-Aspartate-Arginine-Proline-Thrénine-Aspartate-
Arginine

Notez qu'il y a un « - » entre chaque acide aminé mais qu'il n'y en a pas à la fin et que le codon STOP ne s'affiche pas non plus.

Une fois les protéines créées, vous devrez les dessiner avec la Tortue. Chaque acide aminé peut être représenté sous la forme d'une lettre (voir deuxième tableau associatif dans le fichier de départ). Vous devrez dessiner les protéines comme dans l'exemple ci-dessous (15 acides aminés par ligne, puis retour à la ligne et saut de ligne entre chaque protéine).

M	G	R	S	V	G	R	F	G	R	S	A	M	G	R
S	V	G	R	C	R	S	V	G	R	S	L	R	S	V
G	S	R	S	V	G	R	S	L	D	A	S	C	R	S
L	R	S	V	G	L	R	S	V	G	R	S	L	R	S
V	G	Y	K	S	V	G	R	S	L	V	R	S	C	R
S	L	R	S	V	G	Y	K	S	V	G	R	S	L	R
L														
M	G	R	S	V	G	R	C	R	S	V	G	R	S	L
R	S	V	G	S	R	S	V	G	R	S	L	D	A	S
C	R	S	L	R	S	V	G	L	R	S	V	G	R	S
L	R	S	V	G	Y	K	S	V	G	R	S	L	V	R
S	C	R	S	L	R	S	V	G	Y	K	S	V	G	R
S	L	R	S	L										
M	G	R	S	V	G	R	F	G	R	S	A	N	G	R
S	V	G	R	L	T	L	A	V	G	R	F	G	R	S
A	N	G	R	S	V	G	R	L	S	R	R	K	L	S
V	A	S	V	G	R	L	T	V	G	R	S	V	A	S
V	G	R	L	T	V	G	R	S	V	A	N	G	R	S
V	G	R	F	G	R	S	A	N	G	R	S	V	G	R
L	T	L	A	V	G	R	F	G	R	S	A	Y	G	R
S														
M	P	T	D	R	S	D	R	Q	L	A	S	S	D	R
P	T	D	R											

Votre code devra être commenté correctement, et chacune des fonctions devra avoir ses tests unitaires configurés.

3. En résumé

- Créer le brin complémentaire au brin d'ADN fourni.
- Lire les deux brins (dans le bon sens) à la recherche de gènes (trouvez la position du codon de départ et celle du codon stop le plus proche mais situé à un multiple de 3 nucléotides).
- Transcrire chacun des gènes en ARN (seul le gène doit être transcrit, pas tout le brin d'ADN).
- Traduire l'ARN en protéine à l'aide du premier tableau associatif et afficher les protéines sous forme de chaîne de caractères (sans le codon stop à la fin).
- Afficher les protéines à l'aide de la tortue en utilisant le deuxième tableau associatif, chaque acide aminé étant représenté par une seule lettre.

4. Spécifications

4.1 Fonction **antisens (brinAdn)**

Part de brin ADN fourni et renvoie le brin d'ADN complémentaire.

4.2 Fonction **trouveDebut (brinAdn)**

Recherche tous les codons de départ sur un brin d'ADN et renvoie un tableau contenant les positions du premier nucléotide de chacun des codons. Ainsi si TAC se trouve aux positions 3, 67 et 89 (ces trois valeurs étant les positions du T de TAC) il renverra le tableau suivant : [3 , 67 , 89] .

4.3 Fonction **trouveFin (brinAdn)**

Même chose que la fonction précédente mais renvoie un tableau avec les positions de tous les codons de terminaison (attention, il y a trois possibilités de codons de terminaison).

4.4 Fonction **trouveGene (debut, fin)**

Prend en paramètre un tableau contenant les positions de tous les codons de départ et un autre tableau contenant les positions de tous les codons de terminaison pour un brin d'ADN et renvoie un tableau de tuples contenant la liste des gènes (début et fin) trouvés sur un brin.

Ainsi, s'il y a trois gènes sur un brin, le tableau renvoyé ressemblera à :

```
[ (debutGene1, finGene1) , (debutGene2, finGene2) ,  
(debutGene3, finGene3) ]
```

FinGene doit être supérieur à debutGene et finGene doit être situé à un multiple de trois nucléotides de debutGene.

4.5 Fonction **transcrire (brinAdn)**

Prend en paramètre la sous-chaine de caractère du brin d'ADN débutant au début du gène et se terminant à la fin du gène et renvoie le brin d'ARN correspondant sous forme d'une chaîne de caractères.

4.6 Procédure **traduire (brinArn)**

Prend en paramètre un brin d'ARN (chaîne de caractères) et affiche la protéine sous forme d'une chaîne de caractères et la dessine à l'aide de la tortue.

4.7 Procédure **carre(longueur, nombre)**

Prend deux entiers en paramètre (taille du côté du carré et l'indice du carré à dessiner) et trace un carré à l'aide de la tortue.

5. Évaluation

Indiquez clairement vos noms dans les commentaires au début du code.

Remettez le fichier **bioInfo.py**. La remise doit se faire sur le site Studium du cours.

Voici les critères d'évaluation du travail :

- L'exactitude (respect de la spécification)
- L'élégance et la lisibilité du code
- La présence de commentaires explicatifs lorsque nécessaire
- Le choix des identificateurs
- La décomposition fonctionnelle et le choix des tests unitaires

Indications :

- La performance de votre code doit être raisonnable.
- Chaque fonction devrait avoir un bref commentaire pour indiquer ce qu'elle fait.
- Il devrait y avoir des lignes vides pour que le code ne soit pas trop dense.
- Les identificateurs doivent être bien choisis pour être compréhensibles (éviter les noms à une lettre à l'exception de i, j, ... pour les variables d'itération des boucles).
- Vous devez respecter le standard de code pour ce projet (noms de variables en camelCase).
- Vous pouvez créer des fonctions intermédiaires dont vous pourriez avoir besoin.