TP 1 : ARN

Introduction

Pour le premier tp, vous allez construire des classes permettant la manipulation de chaîne d'ARN (acide ribonucléique). Le projet sera composé de trois classes : `Nucleotide`, `AcideAmine` et `AcideRibonucleique` (ARN). Aussi, une classe d'exception sera utilisée : `NucleotideARNInvalide`. Du code de base pour ces classes vous est donné, il vous reste à le compléter.

Description

Nucléotide

Les nucléotides sont des molécules de base pour la construction ARN. Il existe quatre nucléotides : l'adénine (**A**), la cytosine (**C**), la guanine (**G**) et l'uracile (**U**). Chaque nucléotide à un complément. **A** est le complément de **U**, **U** est le complément de **A**, **C** est le complément de **G** et **G** est le complément de **C**.

Codes IUPAC

L'IUPAC (Union Internationale de Chimie Pure et Appliquée) a défini des codes pour regrouper des ensembles de nucléotide. Par exemple, la lettre **R** représente un ensemble contenant une adénine et une guanine. La table suivante contient ces codes avec leur ensemble et le code représentant l'ensemble des compléments.

Code	Ensemble	Complément
R	{A, G}	Y
Y	{C, U}	R
S	{C, G}	S
W	{A, U}	W
K	{G, U}	M
M	{A, C}	K
B	{C, G, U}	V
D	{A, G, U}	H
H	{A, C, U}	D
V	{A, C, G}	B
N	{A, C, G, U}	N

Remarque : S, W et N sont leurs propres ensembles de complément.

Acide Aminé

Dans une chaîne de nucléotide (ARN), les éléments sont divisés en séquence de trois appelé codon. Chaque codon représente une acide aminé protéinogène. La table suivante contient les 22 acide aminé possibles avec

les différentes séquences qu'ils représentent. Remarquez que certains acides aminés sont représenté par plusieurs séquences.

Abréviation	Codon(s)	Nom
Ala	GCU, GCC, GCA, GCG	Alanine
Arg	CGU, CGC, CGA, CGG, AGA, AGG	Arginine
Asn	AAU, AAC	Asparagine
Asp	GAU, GAC	Aspartate
Cys	UGU, UGC	Cystéine
Glu	GAA, GAG	Glutamate
Gln	CAA, CAG	Glutamine
Gly	GGU, GGC, GGA, GGG	Glycine
His	CAU, CAC	Histidine
Ile	AUU, AUC, AUA	Isoleucine
Leu	UUA, UUG, CUU, CUC, CUA, CUG	Leucine
Lys	AAA, AAG	Lysine
Met	AUG	Méthionine
Phe	UUU, UUC	Phénylalanine
Pro	CCU, CCC, CCA, CCG	Proline
Pyl	UAG	Pyrrolysine
Sel	UGA	Sélénocystéine
Ser	UCU, UCC, UCA, UCG, AGU, AGC	Sérine
Thr	ACU, ACC, ACA, ACG	Thréonine
Trp	UGG	Tryptophane
Tyr	UAU, UAC	Tyrosine
Val	GUU, GUC, GUA, GUG	Valine

Il manque un codon : UAA. Ce codon a pour rôle d'indiquer la fin d'une séquence. Les codons UAG et UGA peuvent aussi indiquer la fin d'une séquence en plus d'être des acides aminés. Les codons AUG, UUG et GUG sont aussi utilisés pour représenter le début d'une séquence. Les nucléotides et les acides aminés vont être représentés par un type énuméré.

Acide ribonucléique

Votre classe pour l'ARN va hériter de la classe `ArrayList<Nucleotide>`. Il vous faudra ajouter des méthodes spécialisées pour l'ARN.

Méthodes :

1. `boolean estValide()` : vérifie que la séquence est valide. Une séquence valide à une taille divisible par 3, commence par un codon de début (AUG, UUG, GUG) et termine par un codon de fin (UAA, UAG, UGA). Aussi, si le codon UAA est présent, alors il est à la fin.
2. `AcideAmine getAcideAmine( int position ) throws IndexOutOfBoundsException` : cette méthode cherche les trois nucléotide à partir de l'indice '`position`' multiplié par 3. Ensuite l'acide aminé correspondant à ces trois nucléotides est retourné.

S'il n'y a pas trois nucléotides à l'indice indiqué, alors l'exception '`IndexOutOfBoundsException`' est lancée.

3. `AcideRibonucleique complement()` : cette méthode retourne une ARN qui contient les compléments de l'ARN de base. Cette transformation est faite indice par indice en utilisant la troisième colonne de la table à la première page de l'énoncé.
4. **`boolean`** `decrit( AcideRibonucleique decrite )` : Les deux ARNs sont comparés indice par indice. Il faut tester si la lettre dans la séquence '**`this`**' décrit correctement la lettre dans le séquence en argument '`decrite`'. La table suivante indique les combinaisons correctes. Par exemple, la séquence 'AUGNKGARUUGA' décrit correctement la séquence 'AUGWGGARUUGA'. Si les séquences n'ont pas la même taille, alors la méthode retourne '**`false`**'.

Les lettres suivantes	décrivent cette letter.
A, R, W, M, D, H, V, N	A
C, Y, S, M, B, H, V, N	C
G, R, S, K, B, D, V, N	G
U, Y, W, K, B, D, H, N	U
R, D, V, N	R
Y, B, H, N	Y
S, B, V, N	S
W, D, H, N	W
K, B, D, N	K
M, H, V, N	M
B, N	B
D, N	D
H, N	H
V, N	V
N	N

Les lettres de la colonne de droite représentent des ensembles qui sont inclues dans ceux de la colonne de gauche.

5. `AcideRibonucleique generalise( AcideRibonucleique arn )` : Cette méthode construit une séquence à partir de deux sequences ('**`this`**' et '`arn`'). La séquence résultante va décrire les deux séquences en entrées. Cette construction est faite indice par indice. La table suivante donne le résultat pour l'application entre deux codes. Les cases de cette table ont été construite en prenant l'union des ensembles représentés par les lettres. Par exemple, Y = { C, U } et W = { A, U } donnent H = { A, C, U }.

	A	C	G	U	R	Y	S	W	K	M	B	D	H	V	N
A	A	M	R	W	R	H	V	W	D	M	N	D	H	V	N
C	M	C	S	Y	V	Y	S	H	B	M	B	N	H	V	N
G	R	S	G	K	R	B	S	D	K	V	B	D	N	V	N
U	W	Y	K	U	D	Y	B	W	K	H	B	D	H	N	N
R	R	V	R	D	R	N	V	D	D	V	N	D	N	V	N
Y	H	Y	B	Y	N	Y	B	H	D	H	B	N	H	N	N
S	V	S	S	B	V	B	S	N	B	V	B	N	N	V	N
W	W	H	D	W	D	H	N	W	D	H	N	D	H	N	N
K	D	B	K	K	D	D	B	D	K	N	B	D	N	N	N
M	M	M	V	H	V	H	V	H	N	M	N	N	H	V	N
B	N	B	B	B	N	B	B	N	B	N	B	N	N	N	N
D	D	N	D	D	D	N	N	D	D	N	N	D	N	N	N
H	H	H	N	H	N	H	N	H	N	H	N	N	H	N	N
V	V	V	V	N	V	N	V	N	N	V	N	N	N	V	N
N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Directives

1. Le tp est à faire seul.
2. Vous pouvez ajouter des méthodes aux classes. Vous ne pouvez pas modifier la signature des méthodes déjà données.
3. Commentaire :
    a. Commentez l'entête de chaque classe et méthode. Ces commentaires doivent contenir la description de la méthode et le rôle de ces paramètres.
    b. Une ligne contient soit un commentaire, soit du code, pas les deux.
    c. Utilisez des noms d'identificateur significatif.
    d. Utilisez le français.
4. Code :
    a. Pas de `goto`, `continue`.
    b. Les `break` ne peuvent apparaitre que dans les `switch`.
    c. Un seul `return` par méthode.
5. Indentez votre code. Assurez-vous que l'indentation est faite avec des espaces.

Remise

Remettre le tp par l'entremise de Moodle. Placez vos fichiers `'*.java'` dans un dossier compressé de Window, vous devez remettre l'archive. Le tp est à remettre avant le 13 octobre 23 :59.

Évaluation

- Fonctionnalité (7 pts) : des tests partiels vous serons remis. Un test plus complet sera appliqué à votre tp.
- Structure (1 pt) : veillez à utiliser correctement le mécanisme d'héritage et de méthode.
- Lisibilité (3 pts) : commentaire, indentation et noms d'identificateur significatif.