

## EXPRESSION DE L'INFORMATION GÉNÉTIQUE

### Vue d'ensemble

#### **Les programmes**

Les documents d'accompagnement apportent quelques précisions sur l'application des programmes :

- En première S, il s'agit de mettre en place les deux grandes étapes de la synthèse : transcription et traduction. Le brin transcrit sert de matrice à la construction de l'ARN messager (ARNm), construction catalysée par l'ARN polymérase. La synthèse protéique est localisée dans le cytoplasme et décrite de manière très simple : les ribosomes sont présentés comme des outils de cette synthèse ; le codon initiateur oriente la lecture du message contenu dans l'ARNm et la traduction s'arrête au codon STOP. Les élèves apprennent à utiliser le code génétique. Ce code génétique est universel, non ambigu et dégénéré (le doublet initial peut être seul déterminant) ; la nature dégénérée du code a pour conséquence sa redondance.
- En première L et ES, il s'agit uniquement de montrer que la séquence de nucléotides dans l'ADN gouverne la séquence des acides aminés dans la protéine selon un système de correspondance, le code génétique.
- Les notions de gène morcelé, d'exon et d'intron ne sont pas au programme de première S ni de TS.

#### **Les données**

Dans cette perspective des programmes, on peut envisager l'étude de l'expression de l'information génétique en utilisant comme support les séquences des gènes codant pour les différentes chaînes de globine (alpha, bêta, delta et gamma) et les séquences d'acides aminés de ces chaînes. À vrai dire, ce sont les séquences des allèles les plus répandus dans les populations humaines qui sont ici retenues.

**Pour chaque gène** sont fournies :

- les séquences des deux brins de l'ADN (appelées brin1 et brin2) correspondant à un ARN messager commençant au codon d'initiation AUG et se terminant par un codon stop ; cet ARN également fourni est dénommé ARNm<sub>cod</sub> ;
- la séquence du gène (brin non transcrit, appelé ADN<sub>c</sub>) correspondant à l'ARN messager réellement présent dans le cytoplasme (sans la coiffe et le poly A), appelé ici ARNm ; cet ARNm comprend une région non codante en amont de AUG et une autre en aval du codon stop ;
- la protéine codée par ce gène.

D'autres séquences pourront être introduites par les utilisateurs grâce aux fonctions d'édition du logiciel.

#### **Les prérequis**

Les élèves doivent savoir au préalable que :

- l'ADN, localisé au niveau chromosomique, est le support de l'information génétique ;
- une molécule d'ADN renferme des milliers de gènes, la notion élémentaire du gène étant à ce stade un « bout » d'ADN ;
- la structure de l'ADN, son caractère informatif et son mode de réplication ;
- l'ADN régite le phénotype à travers la synthèse des polypeptides ;
- la fonction d'une protéine dépend de sa structure spatiale laquelle est conditionnée par la séquence des acides aminés.







## Convertir les séquences

- comparer le polypeptide résultant de cette traduction avec le polypeptide de la banque et constater ainsi leur identité ;

- inverser la séquence ; traduire la séquence inversée et constater que le polypeptide obtenu est très différent du polypeptide réel. À l'issue de ces opérations, l'unidirectionnalité de la traduction est corroborée ;
- traduire l'ARNm codant à partir des positions 4, 7, 10, etc. (en dupliquant la séquence et en supprimant le premier triplet, le second, etc.) et constater qu'on obtient toujours la même séquence polypeptidique diminuée à chaque fois d'un acide aminé. A contrario, la traduction de l'ARNm codant à partir du 2<sup>e</sup>, 3<sup>e</sup>, 5<sup>e</sup>, 6<sup>e</sup>... nucléotide, conduit à un polypeptide différent du polypeptide réel. Ainsi est confirmée l'idée d'une traduction de l'ARNm triplet par triplet, sans chevauchement.

## Le code génétique

### Une première approche

On peut en construire le tableau de façon théorique et voir émerger des questions relatives au nombre des triplets possibles par rapport au nombre d'acides aminés : des triplets différents peuvent-ils coder pour le même acide aminé ? Existe-t-il beaucoup de triplets ne codant pour aucun acide aminé ? Quelle est leur signification biologique ?

L'analyse de la séquence nucléique et du polypeptide mis en vis-à-vis (résultat de la conversion), permet de trouver des éléments de réponse à la première question : pour cela, il suffit de relever les triplets de l'ARNm codant pour chaque acide aminé.

À l'issue de ce relevé (qui peut être enrichi par le travail des élèves sur les ARNm codant de globines différentes) le caractère dégénéré du code apparaît, mais il reste des triplets (non utilisés dans les séquences étudiées) pour lesquels on ne dispose pas d'informations. Pour en savoir plus à leur sujet, on peut créer une séquence d'ARN (option **Créer** dans le menu **Fichier**) comportant la suite de ces triplets et en demander la conversion en polypeptide.



La création d'une séquence est facilitée par les icônes des bases : il suffit de cliquer dessus.

Les résultats seront variables en fonction de la séquence de triplets constituée, mais dans tous les cas, les élèves seront confrontés à la signification des triplets qui ne codent pour rien. Ils devraient être capables d'apporter la preuve qu'ils ont valeur de codons d'arrêt de la traduction. Ce travail est important pour la compréhension des conséquences d'une mutation entraînant l'apparition d'un codon stop anticipé (mutation non sens).

### Une deuxième approche

On peut également partir de la visualisation du code génétique proposé par *Anagène*. Le caractère univoque mais redondant du code génétique ressort avec évidence ; chaque acide aminé peut être désigné de manière active par la succession des trois bases qui le codent, et chaque triplet introduit montre l'acide aminé codé ce qui pourrait faciliter la compréhension et du code génétique.

### **AUG** Code génétique

Tableau du code génétique					
1ère base	2ème base				3ème base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	~~~	~~~	A
	Leu	Ser	~~~	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Fermer

**CAC**

His

Histidine

On est là aussi conduit à s'interroger sur les triplets qui ne codent pour aucun acide aminé. L'introduction de séquences d'ARN (option **Créer**, dans le menu **Fichier**) contenant l'un ou l'autre de ces triplets, puis la conversion de ces séquences en polypeptides permettent de tester les hypothèses formulées.

Au terme de ces activités, les notions suivantes sont établies :

- un ARN messager est synthétisé à partir d'un des brins du gène à la suite d'un processus basé sur la complémentarité des nucléotides ;
- la séquence de nucléotides de l'ARNm est la même que celle du brin non transcrit, mais les nucléotides à uracile remplacent ceux à thymine ;
- la traduction de l'ARNm en polypeptide est unidirectionnelle et séquentielle (triplet par triplet) ;
- un triplet de nucléotides de l'ARNm code pour un seul acide aminé ;
- plusieurs triplets codent pour le même acide aminé (ou tous les acides aminés sauf deux, tryptophane et méthionine, sont codés par plus d'un codon) ;
- trois triplets UAA, UAG, UGA ne codent pour aucun acide aminé, mais leur insertion dans une séquence d'ARNm entraîne un arrêt de la traduction : ce sont des codons Stop ;
- une délétion ou une addition d'un nombre de nucléotides qui n'est pas un multiple de 3 dans une séquence d'ADN ou d'ARN, entraîne un décalage dans la lecture et un changement presque complet du polypeptide.

Si on considère que les connaissances sur la machinerie impliquée dans la traduction (ARNt, enzymes aminoacyl-ARNt synthétase) ne sont pas au programme de première S, on voit que le logiciel permet d'atteindre des points significatifs de l'expression des gènes.

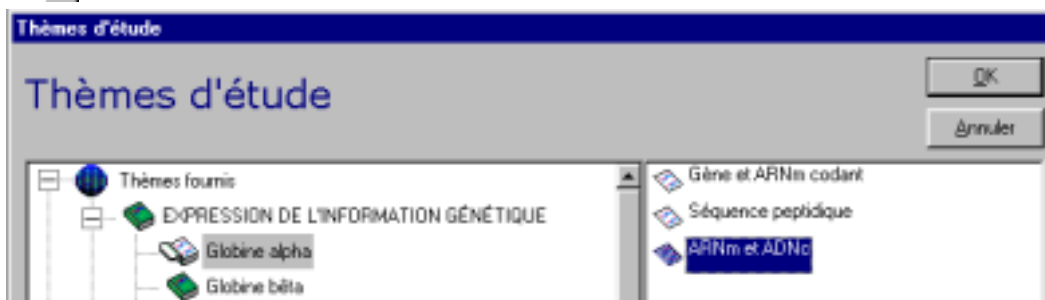
Fermer toutes les fenêtres



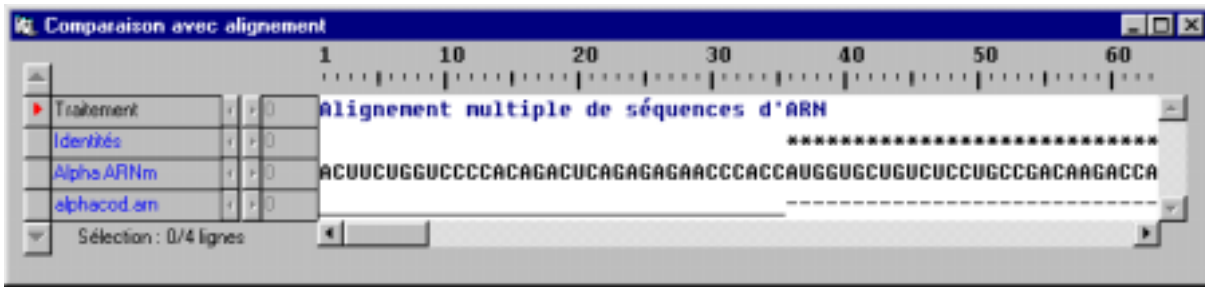
### ***Une vision un peu plus complexe du gène et de l'ARNm***

L'ARN messager réellement traduit dans la cellule est plus complexe que l'ARNm codant considéré jusqu'ici car il comprend en amont et en aval de la région strictement codante des parties non traduites. Il est possible de fournir cette vision plus complète en alignant (avec discontinuités) l'ARNm complet de alpha avec l'ARNm codant correspondant strictement au polypeptide traduit.

**Attention** : la comparaison doit se faire avec la séquence la plus longue (Alpha ARNm) en premier, sinon l'algorithme de comparaison place le premier A des deux AUG à l'extrémité gauche de l'alignement. Pour remonter une séquence, il faut la désigner d'abord en pointant la ligne puis utiliser le bouton ▲.



## ATGC -C-- Comparer les séquences



*Alignement avec discontinuité montrant au début de l'ARNm complet, avant le triplet AUG, une région ne codant pas pour la globine alpha.*

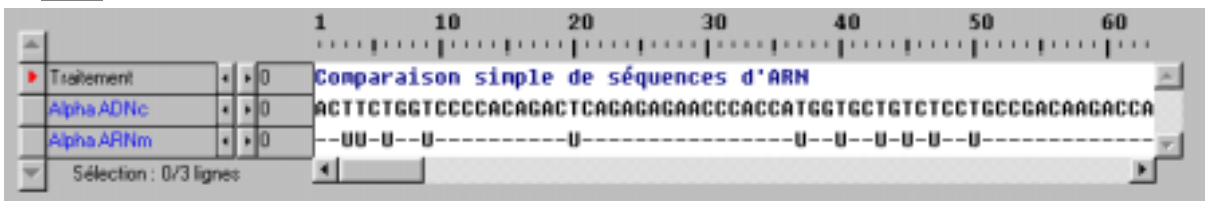
## ATGC -C-- Comparer les séquences



*Alignement avec discontinuité montrant à la fin de l'ARNm complet une région ne codant pas pour la globine alpha.*

Pour insister sur le fait que l'information génétique se trouve fondamentalement dans l'ADN, on pourra comparer aussi Alpha ADNc avec Alpha ARNm.

## ATGC -C-- Comparer les séquences



*L'ADNc et l'ARNm sont équivalents aux T - U près.*

La question du signal de fin de traduction étant déjà élucidée, il reste celle de l'initiation de la traduction. Si les groupes d'élèves travaillent sur des gènes différents (alpha, bêta, gamma, delta) la présence systématique d'un codon AUG en tête de la région traduite pourra être remarquée et sa signification précisée.

On peut également utiliser l'option de conversion appelée **Traduction des cadres ouverts de lecture**. Cette option implique que **la conversion commence par un codon AUG**. Le logiciel affiche les séquences peptidiques commençant par la méthionine dans les trois cadres de lecture possibles. La séquence convertie qui correspond à la protéine réelle est déterminée par le premier AUG rencontré.

Avec Alpha ARNm (donc l'ARNm complet) le résultat est le suivant :



### Convertir les séquences

*La conversion de Alpha ARNm dans les trois cadres de lecture possibles (en commençant par la première base, la seconde et la troisième), fait apparaître le premier AUG dans le troisième cadre de lecture. On peut vérifier que cette séquence correspond bien au polypeptide réel fourni dans la banque de données.*

La vision du gène et celle de son expression qui découlent de cette étude peuvent être schématisées de la manière suivante :

	Début	Codon d'initiation	Codon de terminaison	Fin
ADN	[ GENE	..... ATG .....	TAA (ou TGA) .....	.....
		..... TAC .....	ATT (ou ACT) .....	.....
	ARNm	..... AUG .....	UAA (ou UGA) .....	.....
POLYPEPTIDE		MET .....		

Il est utile de préciser que la méthionine mise en place en premier est ensuite éliminée, ce qui fait que les polypeptides matures ne commencent généralement pas par une méthionine.

Fermer toutes les fenêtres

