

## **DECOUVERTE DU CODE GENETIQUE ET DES ARN MESSAGERS**

## L'élucidation du code génétique

### A. L'hypothèse de l'adaptateur

Dans un article daté de 1953, WATSON et CRICK écrivaient que « la séquence précise des bases est le code qui porte l'information génétique » sans qu'aucun d'eux eût une vision claire de l'existence d'un quelconque code génétique. En 1954 George GAMOW, après avoir eu connaissance des résultats sur la structure en double hélice de l'ADN, publia un article proposant une correspondance directe entre les bases de l'ADN et les acides aminés. Dans cet article il suggérait qu'il existait dans l'ADN des cavités (au nombre de 20) pouvant constituer des réceptacles pour les acides aminés qui seraient alors reliés ultérieurement par des liaisons covalentes.

Cet article eut pour effet de stimuler la réflexion sur une correspondance possible entre les nucléotides de l'ADN et les acides aminés. En 1957 Francis CRICK et Leslie ORGEL proposaient l'existence d'un code à trois bases, sans interruption, non chevauchant, dans lequel 20 des 64 codons correspondraient aux 20 acides aminés principaux présents dans les protéines.

Parallèlement, dès 1954, Francis CRICK avait émis l'hypothèse de l'existence d'un **intermédiaire** entre le gène et la protéine. Même si certains chercheurs, parmi lesquels GAMOW, avaient proposé des mécanismes de biosynthèse des protéines dans lesquels les acides aminés interagiraient directement avec l'ADN, le fait que des acides aminés radioactifs pouvaient se lier de façon transitoire avec des ARN de faible poids moléculaire a permis de tester l'hypothèse de l'adaptateur, ultérieurement appelé ARN de transfert. Il devait donc exister, selon CRICK, **20 ARN adaptateurs** correspondant aux **20 acides aminés** majeurs présents dans les protéines et **20 enzymes** capables de lier spécifiquement un acide aminé à son ARN adaptateur. Une fois cette réaction effectuée, l'ARN adaptateur devait pouvoir aller s'apparier à une matrice d'acide nucléique.

### B. Elucidation du code

En 1960, Paul ZAMECNIK mit au point un système *in vitro* de biosynthèse des protéines à partir d'extraits bactériens. Amélioré par Alfred TISSIERE, ce système allait permettre rapidement le décryptage du code génétique.....

En 1961, Marshall NIRENBERG et Heinrich MATTHAEI démontrent que l'addition de **polyU**, un homopolymère synthétique, dans un système acellulaire de biosynthèse des protéines (S30) contenant des acides aminés radioactifs, se traduit par la synthèse de **polyphénylalanine** radioactive, et donc que le codon UUU correspond, contre toute attente, à la phénylalanine. En effet le modèle de CRICK et ORGEL prévoyait que, dans l'ADN, les triplets constitués d'un seul nucléotide (AAA, CCC, GGG et TTT soit UUU dans l'ARN) étaient des codons non-sens ne correspondant à aucun des 20 acides aminés, et il semble que cette expérience avait été incluse comme **contrôle négatif** dans le protocole expérimental.

Grâce à une enzyme de la bactérie *Azotobacter vinelandii*, la Polynucléotide Phosphorylase ou PNPase purifiée par Marianne GRUNBERG-MANAGO et Sévero OCHOA, de nombreux homo- ou copolymères ont été synthétisés et ont permis d'élucider partiellement le code génétique et de confirmer les observations sur la colinéarité entre gène et protéine. Ainsi, un polyA dirige la synthèse de polylysine et un polyC celle de polyproline.

En 1962, François CHAPEVILLE réalise une expérience de biosynthèse des protéines avec un copolymère alterné, le poly-UG. Dans un système acellulaire contenant des acides aminés radio-actifs, il montre que cette matrice dirige la synthèse d'un peptide alternant des résidus Val et Cys (codons GUG et UGU respectivement). Chimiquement, il est possible de réduire la cystéine en alanine tout en maintenant cet acide aminé fixé sur son ARN soluble en faisant passer l'extrait sur une colonne de nickel de Raney. Si cet extrait est ajouté à un système acellulaire contenant un poly-UG comme matrice, on constate que les acides aminés incorporés sont la valine et l'alanine. Cette expérience démontre que le seul moment où l'**acide aminé joue un rôle** au cours de la synthèse protéique est lors de l'**amino-acylation** de l'ARN soluble ou ARN de transfert, et que c'est l'ARN de transfert qui reconnaît le codon sur l'ARN messenger, confirmant ainsi l'hypothèse de l'**adaptateur** de Francis CRICK.

En 1964, Marshall NIRENBERG démontre que des trinuécléotides synthétiques particuliers permettent la fixation d'aminoacyl-tRNA sur des ribosomes. Ainsi le trinuécléotide pUUU facilite la fixation de Phe-tRNA sur des ribosomes, le trinuécléotide pAAA celle de Lys-tRNA et le trinuécléotide pCCC celle de Pro-tRNA. Grâce à cette approche, 50 des 64 triplets seront rapidement caractérisés. Toujours en 1964, Gobind KHORANA développe la synthèse chimique de polynucléotides et permet de caractériser les codons UAA, UAG et UGA comme codons stop. L'élucidation complète du code est achevée en 1966. On voit que tous ces travaux se sont déroulés en même temps que ceux qui ont abouti à la découverte de l'ARN messenger et il est bien difficile de présenter tous ces faits de façon chronologique.

## La découverte de l'ARN messenger

### A. Le concept d'ARN messenger

Comment passer d'une information codée par 4 lettres (A,T,G,C) aux macromolécules que sont les protéines, et constituées de 20 acides aminés différents ? En 1958, Francis CRICK publie un papier intitulé « A propos de la synthèse des protéines » dans lequel il propose que le repliement des protéines est un **phénomène spontané** et que la conformation finale d'une protéine ne dépend que de l'enchaînement des acides aminés qui la composent. De même, selon lui, dans la séquence d'un acide nucléique, l'enchaînement des bases constituent le **code** qui détermine l'ordre des acides aminés dans les protéines. Il y énonce le **dogme central de la biologie moléculaire** : l'information génétique peut passer d'un acide nucléique à un acide nucléique, d'un acide nucléique à une protéine, mais ne peut aller ni d'une protéine à une autre protéine, ni d'une protéine à un acide nucléique. Il y rappelle également l'hypothèse émise quelques années plus tôt sur l'existence d'adaptateurs et évoque le problème d'un code génétique. Cependant, rien ou presque n'est dit sur une **éventuelle relation entre ADN et ARN** : « La synthèse d'au moins quelques-uns des ARNs microsomiques (*entendez associés aux ribosomes*) doit être sous le contrôle de l'ADN du noyau. »

Si cette relation entre ADN et ARN fut la dernière à être clairement établie, en particulier grâce aux travaux de François JACOB et Jacques MONOD que nous verrons plus loin, on avait déjà un certain nombre de preuves que l'ARN était impliqué dans la synthèse des protéines. Torbjörn CASPERSSON avait montré dès 1939 que le cytoplasme des cellules eucaryotes était riche en acides ribonucléiques et que ce contenu était proportionnel à l'activité métabolique de la cellule et donc à son niveau de synthèse protéique. Les expériences les plus démonstratives furent réalisées en 1955 par Jean BRACHET. Il pratiqua des expériences d'énucléation d'œufs de *Xenopus laevis* et montra que dans les œufs énucléés, la synthèse protéique se poursuivait pendant plusieurs jours après l'opération.

Parallèlement, à la fin des années 40, suite à la description par Albert CLAUDE de particules apparaissant dans les cellules consécutivement à une infection virale et appelées **microsomes**, George PALADE précisa la structure de ces particules microsomales. Ces particules d'un diamètre de 150 angströms, associées ou non à des membranes, étaient composées à parts égales d'ARN et de protéines. En 1950 Henry BORSOOK montra que la synthèse protéique prenait place sur les microsomes, ce que confirma Paul ZAMECNIK un peu plus tard grâce au système acellulaire qu'il avait développé.

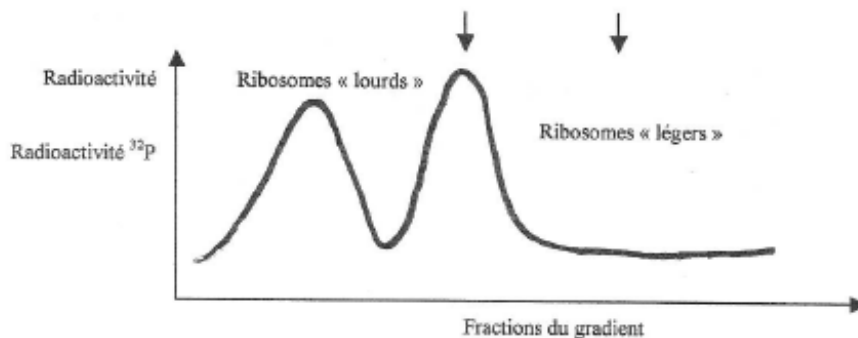
En 1960, à la suite de travaux sur le contrôle de la synthèse protéique chez *E. coli*, grâce au modèle de l'opéron lactose en particulier, François JACOB et Jacques MONOD émettent l'hypothèse d'un messenger structural qui serait un intermédiaire entre le gène et les particules microsomales. La nature de ce messenger est alors inconnue, mais il doit posséder certaines propriétés :

- C'est **probablement un ARN**
- Il est **synthétisé et dégradé très rapidement**
- Il est **différent** des ARNs microsomiques qui eux, sont stables

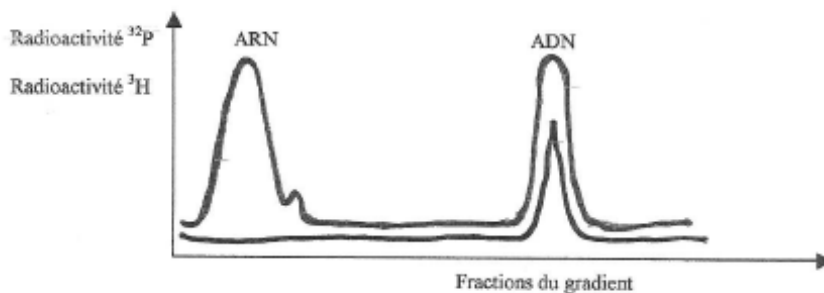
Ces résultats étaient à rapprocher de ceux obtenus en 1956 par Elliot VOLKIN et Lazarus ASTRACHAN restés jusque là inexplicés. En étudiant la répllication des **bactériophages T2 et T7** dans la bactérie *Escherichia coli*, ils avaient constaté qu'un **ARN de composition en bases identique à l'ADN du phage**, était synthétisé peu de temps après l'infection. Ils avaient proposé que cet ARN était un précurseur de l'ADN viral lors de l'étape de répllication. François JACOB, Sydney BRENNER et Francis CRICK, lors d'une discussion conclurent que l'ARN mis en évidence par VOLKIN et ASTRACHAN était un **ARN messenger** semblable à celui identifié par JACOB et MONOD, et commandant la synthèse des protéines virales nécessaires à la répllication du phage T2. L'ARN contenu dans les particules microsomales n'est pas l'ARN messenger dirigeant la synthèse des protéines. Ce dernier, de taille hétérogène, ne représente qu'une **petite fraction** des ARNs cellulaires et était passé **inaperçu** en raison de sa **courte durée de vie**.

Sydney BRENNER, François JACOB et Matthew MESELSON vont apporter la preuve expérimentale du rôle d'intermédiaire de l'ARN messenger dans les processus de l'expression génétique. Ils utilisent pour cela le modèle de l'infection d'*Escherichia coli* par le bactériophage T2. Dans un premier temps, les bactéries sont cultivées dans un milieu riche en azote <sup>15</sup>N et en carbone <sup>13</sup>C pendant plusieurs générations. Après **infection par le phage T2** les bactéries sont cultivées dans un **milieu normal** contenant du <sup>32</sup>P capable de marquer les acides nucléiques, ou du soufre <sup>35</sup>S capable de s'incorporer dans les protéines. En présence de <sup>32</sup>P, si on fractionne l'extrait cellulaire selon sa densité par centrifugation sur un gradient de chlorure de césium, on constate que les ARNs radioactifs néosynthétisés s'associent aux ribosomes lourds et sont de demi-vie courte. Par contre en présence de <sup>35</sup>S, les protéines apparaissent

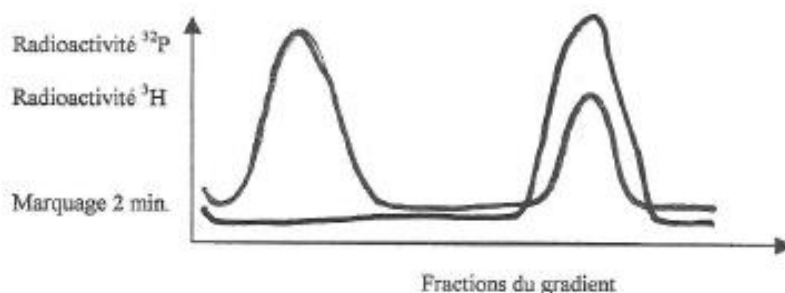
associées de façon transitoire aux ribosomes. Les ribosomes ne jouent donc qu'un rôle passif dans la synthèse des protéines du phage. Ce sont donc des structures non spécialisées qui synthétisent à un moment donné la protéine codée par l'ARNm qui leur est associé.

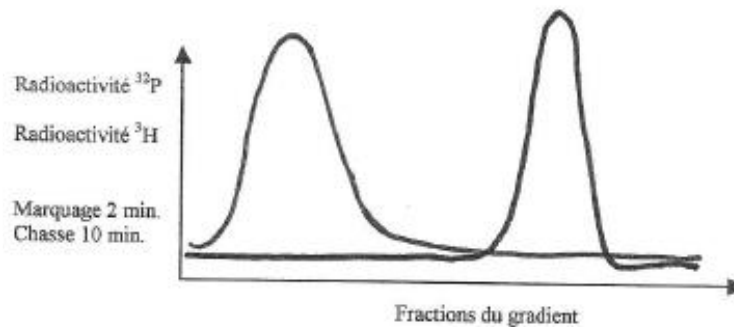


De plus, des expériences d'hybridation réalisées la même année dans le laboratoire de Sol SPIEGELMAN démontrent que cet ARN qui apparaît après l'infection peut s'hybrider spécifiquement avec de l'ADN du bactériophage T2 préalablement marqué au tritium. Deux autres classes d'ARN, de demi-vie beaucoup plus longue, peuvent également être mises en évidence dans cette expérience : les ARN ribosomiaux et les ARN dits solubles, ultérieurement assimilés aux ARN de transfert, mais ils ne s'hybrident pas avec l'ADN du phage T2.



Dans le même temps, François GROS va utiliser une autre approche pour démontrer l'existence des ARNs messagers. Cette expérience fait appel à des bactéries normales cultivées pendant plusieurs générations en présence de tritium  $^3\text{H}$ . Les bactéries sont alors transférées dans un milieu normal auquel on ajoute du phosphore  $^{32}\text{P}$  pendant deux minutes suivies ou non d'une chasse par du phosphore non radioactif pendant dix minutes (**pulse-chase**). Par centrifugation en gradient de saccharose, on peut alors isoler une fraction d'ARN de taille hétérogène marqué au  $^{32}\text{P}$  différente de l'ARN microsomique, et dont la demi-vie est courte. Ces ARNs sont cependant associés aux ribosomes pré-existants et ont donc toutes les caractéristiques des ARNs messagers.





L'ARN messager ainsi caractérisé possède les propriétés suivantes:

- C'est un polyribonucléotide
- Sa composition en base reflète celle de l'ADN qui le spécifie
- Il est de taille hétérogène
- Il est synthétisé et dégradé très rapidement
- Il est associé transitoirement aux ribosomes

## B. Colinéarité entre gène et protéine

Pour passer d'un enchaînement de 4 nucléotides à un polymère composé de 20 acides aminés, il est nécessaire de déchiffrer la séquence polynucléotidique par groupes de 3 nucléotides successifs.

4 nucléotides lus individuellement correspondraient à 4 acides aminés, et lus deux à deux ils ne coderaient que 16 acides aminés. Lus trois à trois ces 4 nucléotides permettent 64 possibilités, ce qui signifie que certains triplets sont inutilisés ou que le code génétique est **dégénéré** ou **redondant**.

En 1961, grâce à l'utilisation d'agents intercalants tels que la proflavine ou l'acridine orange, Sydney BRENNER et Francis CRICK vont démontrer que le code génétique est un code à 3 lettres et qu'il a **colinéarité** entre un gène et le produit de ce gène. Le modèle utilisé est celui du bactériophage T4 de la bactérie *Escherichia coli*. La proflavine induit des mutations par insertion ou délétion d'une base et en particulier la mutation *FC0* à l'intérieur du cistron *rIIB*. Ce mutant est capable de pousser sur un hôte dit « permissif » la souche *E. coli* B, mais ne peut se multiplier sur la souche « restreinte » *E. coli* K12 ( $\lambda$ ). Dans certains cas, la multiplication du mutant *FC0* dans la souche *E. coli* B se traduit par l'apparition de phages ayant un phénotype **sauvage**, c'est à dire capable de se multiplier dans la souche « restreinte » *E. coli* K12 ( $\lambda$ ). Cependant des expériences de recombinaison avec un phage sauvage démontrent qu'il ne s'agit pas d'une réversion mais d'une deuxième mutation appelé *FC1*. Les mutants *FC0* et *FC1* sont donc des mutants **suppresseurs intragéniques**.

En appliquant le même protocole expérimental à *FC1*, ils vont isoler un mutant *FC2* suppresseur intragénique de *FC1*, puis les mutants *FC3*, *FC4*, *FC5*, etc... dans lesquels *FC(n)* est le suppresseur intragénique de *FC(n-1)*.

Grâce à des expériences de recombinaisons entre ces différents mutants, CRICK et BRENNER tirent un certain nombre de conclusions :

- Si la mutation *FC0* est une **insertion**, alors *FC1* est une **délétion**, *FC2* est une insertion, *etc...* et inversement.
- Le gène est lu de façon séquentielle à partir d'un point fixe. L'insertion ou la délétion d'un nucléotide change le cadre de lecture de l'ARN messager qui est décodé sous forme de codons successifs.
- Chaque insertion ou délétion se traduisant par une mutation, cela signifie qu'il n'y a pas de ponctuation entre les codons.
- Le code est constitué de 64 triplets
- La grande majorité de ces triplets correspondent à des acides aminés, ce qui signifie que le code est dégénéré ou autrement dit redondant.

En réalité la démonstration définitive de la colinéarité entre gène et protéine viendra des travaux de Charles YANOFSKY sur la cartographie de mutants de la tryptophane synthétase, le produit du gène *trpA* de *E. coli*.

### Les propriétés remarquables du code génétique

Pour résumer ce que nous venons de voir, nous pouvons dire que le code génétique possède les caractéristiques suivantes :

- **Le code est constitué de 64 triplets**
- **Le code est non chevauchant et sans ponctuations**
- **Le code est universel et dégénéré**

La propriété la plus remarquable du code génétique est que la correspondance entre triplets et acides aminés ne se fait pas au hasard et que le code génétique a évolué de manière à **minimiser** les effets des mutations.

Lorsqu'on construit un tableau à trois entrées où sont reportés dans 16 carrés les 64 codons en utilisant l'ordre U, C, A, G, on constate plusieurs choses :

		Deuxième position				
Première position	U	C	A	G	Troisième position	
U	Phe	Ser	Tyr	Cys	U	
	Phe	Ser	Tyr	Cys	C	
	Leu	Ser	Stop	Stop	A	
	Leu	Ser	Stop	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	

- La plupart des acides aminés se retrouvent dans le même carré, sauf Arg, Leu et Ser qui ont 6 codons.
- XYU et XYC correspondent toujours au même acide aminé.
- XYA et XYG correspondent toujours au même acide aminé, sauf pour Met et Trp qui ne possèdent qu'un codon.
- Les codons ayant un U comme deuxième base correspondent à des acides aminés hydrophobes.
- Les codons ayant un A comme deuxième base correspondent à des acides aminés hydrophiles.

## Origine et évolution du code génétique

### A. Origine

Avec les ARN et les protéines, on est réellement en présence du problème de la poule et de l'œuf. Qui des deux a précédé l'autre ?

Les travaux les plus récents privilégient l'hypothèse selon laquelle l'ARN aurait prévalu à l'origine. Ces ARN auraient été doués d'auto-réplication et auraient pu catalyser la formation de liens peptidiques entre des acides aminés présents dans la « soupe primordiale » grâce à des interactions spécifiques entre codons et acides aminés. Les travaux menés par l'équipe de Laura LANDWEBER ont montré que lorsque de l'arginine radioactive est incubée avec des polynucléotides, des interactions se créent entre les deux types de molécules. Dans 28 à 30 % des cas l'arginine se lie à l'ARN de façon non spécifique, mais dans 70 à 72 % des cas l'arginine se lie à un codon arginine. L'utilisation d'aptamères va permettre d'explorer cette voie plus en détail avec d'autres acides aminés.



## B. Evolution

Ceci ne permet pas d'expliquer comment le code génétique a pu évoluer. L'hypothèse proposée par Edward TRIFONOV est que le code a évolué à partir du codon GCU qui correspond à l'alanine. Nous avons indiqué que les ARN pouvaient être doués de la capacité d'autoréplication. Le codon GCU, correspond au triplet GCT dans l'ADN, triplet que l'on retrouve **amplifié** dans plusieurs maladies génétiques telles que la dystrophie myotonique et la maladie de **Huntington** qui peut provoquer un glissement de la DNA polymérase lors de la réplication de l'ADN. Si l'on suppose que le triplet GCU puisse avoir eu les mêmes propriétés dans un monde dominé par l'ARN, sa présence aurait eu pour conséquence de favoriser l'apparition de molécules d'ARN plus longues, capables de se replier et d'interagir avec des acides aminés et d'autres molécules.

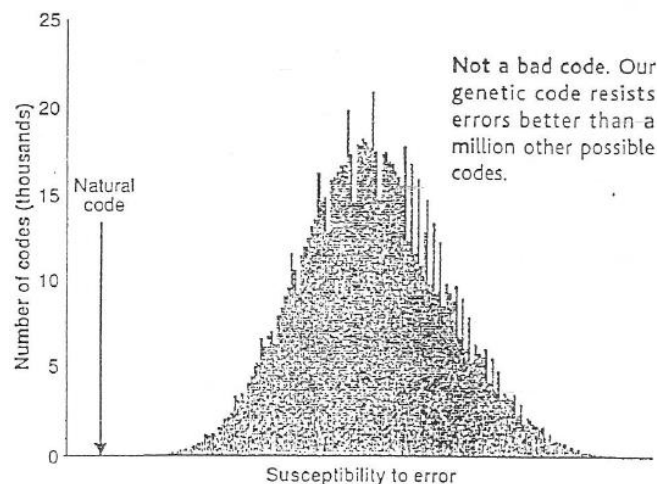
A partir du codon GCU, on peut par une mutation ponctuelle générer des codons correspondant à 6 **acides aminés**, puis par une deuxième des codons correspondant à 6 autres acides aminés, comme cela est présenté dans le tableau suivant :

	Nombre de changements			
	0	1	2	3
<b>Acides aminés</b>	Alanine	Acide aspartique Proline Thréonine Valine Serine Glycine	Asparagine Leucine Arginine Isoleucine Lysine Acide glutamique	Méthionine et autres acides aminés

Si l'on fait la liste des acides aminés les plus anciens, c'est à dire ceux qu'on aurait retrouvés dans la soupe primordiale, elle se superpose avec celle de la **colonne 1** du tableau précédent. Voir à ce sujet les travaux de MILLER et UREY en 1953.

L'évolution du code génétique va de pair avec son optimisation. Nous présentons différents modèles qui reflètent les possibilités en termes d'évolution du code génétique au cours du temps, pour arriver à ce qu'il est aujourd'hui.

Nous avons dit précédemment que le code avait évolué de manière à minimiser les effets des mutations. En 1991, Laurence HURST et Davis HAIG avaient montré, grâce à des calculs mathématiques, que parmi tous les codes possibles construits à partir de 4 bases et 20 acides aminés, le code dit « universel » est le **plus fiable** (1 sur  $10^6$ ) en terme de protection contre les mutations.



### C. Les codes alternatifs

Qu'il s'agisse de l'évolution du code universel ou de l'apparition de codes alternatifs, deux mécanismes assez proches ont été proposés :

#### 1. Le modèle d'Osawa-Jukes

Dans ce modèle, certains codons disparaissent en raison de la pression de sélection due au changement de la composition en bases AT / GC du génome de l'organisme. En conséquence, les ARNt correspondant à ces codons disparaissent également. Lorsque la pression de sélection s'inverse, l'absence d'ARNt pour décoder ces codons inhibe la traduction. Toute mutation qui va permettre la lecture de ces codons sera favorisée. Par exemple, la duplication d'un gène d'ARNt suivie d'une mutation dans l'anticodon créera une nouvelle spécificité si l'ARNt est toujours reconnu par son aminoacyl-tRNA synthétase.

#### 2. Le modèle de Schultz-Yarus

Le mécanisme proposé est assez semblable, mais ici, la disparition complète du codon n'est pas nécessaire. La duplication du gène d'ARNt s'accompagne d'une mutation dans l'anticodon et/ou d'une modification de la spécificité vis à vis des aminoacyl-tRNA synthétases.

La liste des variants nucléaires ou mitochondriaux s'allonge chaque année un peu plus. Les tableaux suivants résument l'état actuel des connaissances dans ce domaine.

VARIANTS NUCLEAIRES			
Codon	Code universel	Code alternatif	Espèce
UAA, UAG	Ter	Gln	Acetabularia Tetrahymena Paramécie
UGA	Ter	Cys Trp  Sélénocystéine	Euplotes Mycoplasmes Spiroplasma Mammifères Bactéries
CGG	Arg	Ter	Mycoplasmes Spiroplasma
AGA	Arg	Ter	Micrococcus
CUG	Leu	Ser	Candida Saccharomyces
AUA	Ileu	Ter	Micrococcus

VARIANTS MITOCHONDRIAUX			
Codon	Code universel	Code alternatif	Espèce
UAA	Ter	Tyr	Plathelminthes
UAG	Ter	Leu Ala	Chloroplastes Chloroplastes
UGA	Ter	Trp	Dictyostelium Plantes
CGU	Arg	Ter	Levures
CGG	Arg	Ter	Candida Prototheca (algue)
AGA	Arg	Ser Ter Gly	Bilateria Tuniciés Vertébrés
AAA	Lys	Asn	Echinodermes Plathelminthes
AUA	Ileu	Met	Levures Echinodermes
CUG	Leu	Thr	Levures