

# Note complémentaire concernant l'isomorphisme entre la manière dont les codons codent les acides aminés et celle dont les nombres entiers se regroupent autour des nombres premiers

20, 21 ou 22 acides aminés ?

19, 20, 21 ou 22 nombres premiers entre 0 et 63 ?

(Alain Bruyère 28 janvier 2015)

Suite à l'interpellation d'une participante à la journée d'hommage à Xavier Sallantin le 17 janvier 2014 à propos de ma présentation, « Xavier le biologiste », je souhaite clarifier quelques points afin de répondre aux questions liées à l'isomorphisme présenté entre la manière dont les 64 codons codent les 20, 21 ou 22 acides aminés et celle dont les 64 premiers nombres entiers (0 à 63) se regroupent autour des 19, 20, 21 ou 22 premiers nombres premiers.

Dans ma présentation, basée sur l'approche de Xavier Sallantin, j'indique en effet qu'il y a 22 acides aminés et 22 nombres premiers entre 0 et 63. Mais d'aucuns disent : « ce n'est pas correct il n'y a que 20 acides aminés ». En ce qui concerne les nombres premiers, afin d'obtenir 22 nombres premiers, comme Xavier Sallantin, je prends 0 et 1 de même que 9 et 27. A nouveau, d'aucuns disent : « ce n'est pas correct il n'y a que 18 nombres premiers entre 0 et 63 ».

Voyons cela de plus près en commençant par les acides aminés.

## Les acides aminés

Les acides aminés, pour rappel, sont enchaînés les uns aux autres selon une séquence définie par le code génétique. L'ARN messager qui se constitue en une série de triplet de bases ou codons parmi les 4 bases ACGU (ex ACU, UGC, AUC, ...) à partir de l'ADN (transcription), « passe » dans le ribosome (l'ARN messager est donc produit lui-même par transcription à partir de l'ADN). Le ribosome recherche alors l'ARN de transfert possédant le triplet complémentaire (ou l'anti codon) de l'ARN messager (traduction) et l'acide aminé transporté par l'ARN de transfert, se détache alors de cet ARN de transfert et est alors ajouté à la chaîne des acides aminés qui se construit petit à petit pour former la protéine.

Il y a 64 codons vu qu'il y a 4 valeurs possibles pour les bases qui sont assemblées 3 à 3 =>  $4*4*4 = 64$ .

Pour comprendre le mécanisme global et détaillé, voir :

[http://fr.wikipedia.org/wiki/Code\\_g%C3%A9n%C3%A9tique](http://fr.wikipedia.org/wiki/Code_g%C3%A9n%C3%A9tique)

[http://fr.wikipedia.org/wiki/Acide\\_amin%C3%A9](http://fr.wikipedia.org/wiki/Acide_amin%C3%A9)

<http://fr.wikipedia.org/wiki/Codon-stop>

Que nous disent ces articles ?

Tout d'abord

« Les acides aminés nécessaires à la synthèse des protéines sont au nombre de 21 chez les eucaryotes (animaux dont les hommes, les champignons, les plantes, .. regroupe tous les organismes unicellulaires ou pluricellulaires qui se caractérisent par la présence d'un noyau et de mitochondries dans leurs cellules,) et la plupart des procaryotes (êtres vivants unicellulaires sans noyau, ex bactéries et archées) et de 22 chez certaines archées méthanogènes (ce sont des micro-organismes unicellulaires procaryotes à métabolisme méthanogène qui produisent du méthane comme sous-produit métabolique de la vie en conditions anoxiques. Ces archées, ou *Archaea* (du grec ancien ἀρχαῖος, « originel, primitif »), encore appelées archéobactéries, sont donc des êtres vivants constitués d'une cellule unique qui ne comprend ni noyau ni organites, à l'instar des bactéries.».

« 20 acides aminés dits standards sont directement spécifiés par un codon d'ADN ou d'ARN. Les 2 restants — la sélénocystéine et la pyrrolysine — sont insérés lorsqu'un codon STOP particulier est associé à une *séquence d'insertion* en épingle à cheveux, ou « tige-boucle », appelée élément SECIS et élément PYLIS respectivement ».

« Le nombre de séries de trois lettres prises dans un alphabet de quatre lettres étant de  $4^3$ , le code génétique comporte 64 codons différents, codant directement les vingt acides aminés standard et les signaux de fin de la traduction, appelés codons-stop ou codons de terminaison. Ces 2 acides aminés mineurs, la sélénocystéine et la pyrrolysine sont donc insérés de façon spécifique au niveau de certains codons-stop, dont le recodage est obtenu grâce à des structures spécifiques sur l'ARN messager. »

Contrairement à la plupart des acides aminés rares, qui n'appartiennent pas à la série des 20 acides aminés codés directement par le code génétique, la sélénocystéine n'est pas formée après la traduction de l'ARNm, mais incorporée directement lors de la synthèse de la chaîne polypeptidique par le ribosome.

La pyrrolysine quant à lui est un acide aminé naturel possédant une correspondance dans le code génétique utilisé par certaines archées méthanogènes. Il est principalement rencontré dans les enzymes faisant partie du métabolisme responsable de la production de méthane ».

Ou encore

« Un codon-stop ou codon de terminaison ou encore codon Non-Sens (codon NS) est l'un des trois codons (parmi les 64 codons du code génétique) qui marquent la fin de la traduction d'un gène en protéine. Il n'est en général jamais traduit car il n'existe pas d'ARN de transfert correspondant.

Le codon UGA peut cependant, dans les cas exceptionnels des sélénoprotéines, coder une sélénocystéine ; cet acide aminé rare est alors incorporé à l'aide de signaux supplémentaires présents sur l'ARN messager et d'un ARN de transfert particulier.

De la même manière, il a été montré que chez certaines archées méthanogènes, le codon UAG peut également coder un acide aminé, la pyrrolysine. Une mutation ponctuelle qui donne naissance à un codon d'arrêt est appelé une mutation non-sens. Les trois codons-stop sont :

- UAA (ocre)
- UAG (ambre), pouvant coder la pyrrolysine
- UGA (opale ou ombre), pouvant coder la sélénocystéine

Enfin

Le premier acide aminé incorporé (déterminé par le codon-Start AUG) est une méthionine chez la plupart des eucaryotes, plus rarement une valine (chez certains eucaryotes), et une formyl-méthionine chez la plupart des procaryotes. En outre, ce codon d'initiation est parfois GUG ou GUU chez certains procaryotes.

L'on a donc le tableau suivant :

<u>Acide aminé</u>	<u>Codons</u>
<u>Alanine</u>	<u>Ala A</u> GCU, GCC, GCA, GCG.
<u>Arginine</u>	<u>Arg R</u> CGU, CGC, CGA, CGG ; AGA, AGG.
<u>Asparagine</u>	<u>Asn N</u> AAU, AAC.
<u>Acide aspartique</u>	<u>Asp D</u> GAU, GAC.
<u>Cystéine</u>	<u>Cys C</u> UGU, UGC.
<u>Glutamine</u>	<u>Gln Q</u> CAA, CAG.
<u>Acide glutamique</u>	<u>Glu E</u> GAA, GAG.
<u>Glycine</u>	<u>Gly G</u> GGU, GGC, GGA, GGG.
<u>Histidine</u>	<u>His H</u> CAU, CAC.
<u>Isoleucine</u>	<u>Ile I</u> AUU, AUC, AUA.
<u>Leucine</u>	<u>Leu L</u> UUA, UUG ; CUU, CUC, CUA, CUG.
<u>Lysine</u>	<u>Lys K</u> AAA, AAG.
<u>Méthionine</u>	<u>Met M</u> AUG.
<u>Phénylalanine</u>	<u>Phe F</u> UUU, UUC.
<u>Proline</u>	<u>Pro P</u> CCU, CCC, CCA, CCG.
<u>Pyrrrolysine</u>	<u>Pyl O</u> UAG, après séquence PylIS.
<u>Sélénocystéine</u>	<u>Sec U</u> UGA, après séquence SecIS.
<u>Sérine</u>	<u>Ser S</u> UCU, UCC, UCA, UCG ; AGU, AGC.
<u>Thréonine</u>	<u>Thr T</u> ACU, ACC, ACA, ACG.
<u>Tryptophane</u>	<u>Trp W</u> UGG. (UGA)
<u>Tyrosine</u>	<u>Tyr Y</u> UAU, UAC.
<u>Valine</u>	<u>Val V</u> GUU, GUC, GUA, GUG.
<u>START</u>	AUG. (UUG, GUG)
<u>STOP Ambre</u>	UAG.
<u>STOP Ocre</u>	UAA.
<u>STOP Opale</u>	UGA.

L'on voit donc sur ce tableau que ce sont de manière très précise, en ce qui concerne les protéines de la vie de manière large,

- que ce sont 63 codons qui codent les 22 acides aminés et les codons STOP Ambre et Opale et
- 1 codon UAA qui code le codon STOP Ocre

En effet le codon UAA ne code aucun acide aminé stricto sensu, mais un codon STOP non associé à un acide aminé. Par contre, les 2 autres codons STOP que sont UGA et UAG codent respectivement

l'acide aminé Sélénocystéine et l'acide aminé Pyrrolisine dans certains cas précis comme indiqué plus haut.

Selon le modèle présenté dans le cadre de l'isomorphisme dont question entre codons/acides-aminés et nombres entiers/nombres premiers, les correspondances suivantes sont faites :

- UGA code à la fois lié à l'acide aminé Sélénocystéine et est un signe de ponctuation STOP, c'est-à-dire, un codon NS
- UGG code l'acide aminé Tryptophane
- AUG code à la fois lié à l'acide aminé Méthionine et est le signe de ponctuation START, c'est-à-dire, un codon START
- UAG code à la fois lié à l'acide aminé Pyrrolisine et est un signe de ponctuation STOP, c'est-à-dire, un codon NS
- Et UAA est juste un signe de ponctuation STOP, c'est-à-dire un codon NS

Xavier Sallantin a regroupé les 2 codons UAA et UAG et les a reliés à l'acide aminé Pyrrolisine. C'est là que peut être critiquée cette association étant donné que UAA ne code pas la Pyrrolisine, c'est UAG qui la code. Cependant, les 2 codons sont des codons STOP, ce qui donne quand même une certaine légitimité à ce regroupement (regroupement qui n'aurait pas pu être fait avec le codon UGA qui lui code un autre acide aminé, la Sélénocystéine). Cependant il aurait pu lier UAA à UGA au lieu de UAA à UAG. Pourquoi n'a-t-il pas choisi cette solution étant donné que la Sélénocystéine serait plus proche du vivant humain que la Pyrrolisine ? Quoiqu'il en soit, ceci ne changerait rien à l'isomorphisme démontré, un peu forcé, serait-on tenté de dire, par Xavier Sallantin.

## Les nombres premiers

La définition d'un nombre premier est la suivante :

« Un nombre premier est un entier naturel qui admet exactement deux diviseurs distincts entiers et positifs (qui sont alors 1 et lui-même). Ainsi, 1 n'est pas premier car il n'a qu'un seul diviseur entier positif; 0 non plus car il est divisible par tous les entiers positifs ».

Les 18 nombres premiers inférieurs à 64 sont donc

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61

Manifestement, 9 n'est pas premier pas plus que 27. Or Xavier Sallantin considère ces 4 nombres, 0, 1, 9, 27 comme premiers. Assurément c'est là une erreur. Personnellement je considère ces nombres comme semi-premiers.

En effet 0 et 1 sont des nombres tout à fait particuliers puis que 0 ne divise aucune nombre et que 1 divise tous les nombres. En quelque sorte ils sont donc particuliers et par une extension osée, je les considère, en tout cas Xavier Sallantin fait le pas sans sourciller, comme semi-premiers.

En ce qui concerne 9 et 27, là aussi Xavier Sallantin les considère comme premiers du fait que ce sont des puissances de 3 et que donc comme puissance de 3, ils sont tout particulièrement remarquables. En effet le nombre 3 pour Xavier Sallantin revêt une signification toute particulière (logique trialectique notamment). De là à les considérer comme premiers, il y a un pas qu'il franchit allègrement. Je suis plus prudent et les prends en compte mais en spécifiant bien qu'il s'agit de nombres que je qualifie également de semi-premiers.

# Isomorphisme

On peut alors produire le tableau suivant en parlant des 22 acides aminés et des 22 bio-nombres ou nombres semi-premiers

22	acides aminés	22	19	9	8	3	3	64	22	bionombres/ nombres semi- premiers	22	19	9	8	3	3	64			
1NS	1 Sélénocystéine	UGA							0	1 n°1	0									
1 TRY	2 Tryptophane	UGG	3 combinaisons de 1						1	2 n°2	1	3 combinaisons de 1								
1 MET	3 Méthionine	AUG	3						2	3 n°3	2	3								
2NS	4 Pyrrolysine	UAA UAG	10 combinaisons de 2						3	4 n°4	3	4	10 combinaisons de 2							
2CYS	5 Cystéine	UGC UGU	20						4	5 n°5	5	6	20							
2TYR	6 Tyrosine	UAC UAU							5	6 n°6	7	8								
2GLU	7 Acide glutamique	GAA GAG							6	7 n°7	9	10								
2ASP	8 Acide aspartique	GAC GAU							7	8 n°8	11	12								
2HIS	9 Histidine	CAC CAU							5	9 n°10	17	18								
2GLN	10 Glutamine	CAA CAG							9	10 n°13	27	28								
2PHE	11 Phénylalanine	UUC UUU							10	11 n°14	29	30								
2LYS	12 Lysine	AAA AAG							11	12 n°17	41	42								
2ASN	13 Asparagine	AAC AAU							12	13 n°21	59	60								
3ILE	14 Isoleucine	AUA AUC AUU	1 combinaison de 3						13	14 n°22	61	62	63	1 combinaison de 3						
4THR	15 Thréonine	ACA ACC ACG ACU	5 combinaisons						14	15 n°9	13	14	15	16	5 combinaisons					
4PRO	16 Proline	CCA CCC CCG CCU	de 4						15	16 n°11	19	20	21	22	de 4					
4ALA	17 Alanine	GCA GCC GCG GCU	20						16	17 n°12	23	24	25	26	20					
4VAL	18 Valine	GUA GUC GUG GUU							17	18 n°16	37	38	39	40						
4GLY	19 Glycine	GGA GGC GGG GGU							18	19 n°18	43	44	45	46						
6LEU	20 Leucine	UUA UUG CUA CUC CUG CUU	3 combinaisons						19	20 n°15	31	32	33	34	35	36	3 combinaisons			
6ARG	21 Arginine	AGA AGG CGA CGC CGG CGU	de 6						20	21 n°19	47	48	49	50	51	52	de 6			
6SER	22 Sérine	AGC AGU UCA UCC UCG UCU						18	21	22 n°20	53	54	55	56	57	58	18			

1NS	1	Sélénocystéine	UGA					
1 TRY	2	Tryptophane	UGG					
1 MET	3	Méthionine	AUG					
2NS	4	Pyrrolysine	UAA	UAG				
2CYS	5	Cystéine	UGC	UGU				
2TYR	6	Tyrosine	UAC	UAU				
2GLU	7	Acide glutamique	GAA	GAG				
2ASP	8	Acide aspartique	GAC	GAU				
2HIS	9	Histidine	CAC	CAU				
2GLN	10	Glutamine	CAA	CAG				
2PHE	11	Phénylalanine	UUC	UUU				
2LYS	12	Lysine	AAA	AAG				
2ASN	13	Asparagine	AAC	AAU				
3ILE	14	Isoleucine	AUA	AUC	AUU			
4THR	15	Thréonine	ACA	ACC	ACG	ACU		
4PRO	16	Proline	CCA	CCC	CCG	CCU		
4ALA	17	Alanine	GCA	GCC	GCG	GCU		
4VAL	18	Valine	GUA	GUC	GUG	GUU		
4GLY	19	Glycine	GGA	GGC	GGG	GGU		
6LEU	20	Leucine	UUA	UUG	CUA	CUC	CUG	CUU
6ARG	21	Arginine	AGA	AGG	CGA	CGC	CGG	CGU
6SER	22	Sérine	AGC	AGU	UCA	UCC	UCG	UCU