Syomie du code génétique

La voie vers les caractéristiques linguistiques réelles des protéines en cours de synthèse

P. P. Gariaev, E. A. Leonova-Gariaeva

SARL Institut de Génétique Quantique

Table de matière

Syomie du code génétique 1
Table de matière 1
Figures 1
Résumé 1
L'hypothèse indéfinie (wobble) de F. Crick 2
Les facteurs d'univocité et de dégénérescence du code de protéine de *E. Coli* 3
Comment se fait le choix d'acides aminés et des STOP positions dans le cas de la rencontre du ribosome et d'un codon homonyme de l'ARN messager 4
Rectification de l'hypothèse indéfinie de F. Crick 6
Pourquoi les STOP codons se trouvent dans les familles de siomies 7
Discussion 8
Bibliographie 11

Figures

Figure 1 : Tableau du code génétique 3
Figure 2: La binarité homonymo-synonymique du code génétique 5
Figure 3: La dégénération binaire homo-synonymique du code des triplets protéiniques 7

Résumé

Il est fourni une analyse du rôle joué par le 3ième nucléotide des codons, lors de la synthèse protéique, sur la base d'une analyse théorique et de travaux expérimentaux. Sa signification est comprise de façon élargie par rapport à la conception habituellement admise. Le 3ième nucléotide divise les familles des codons, fonctionnellement et symétriquement, en 32 codons synonymes et 32 codons hybrides SYnonymo-hOMmonymes (les SYOMs). Les syoms disposent de la fonction déclenchante de l'analyse ribosomique déportée de l'ARNm, en tant que contexte réel dans le langage de l'ADN.

Une telle analyse s'avère être naturellement indispensable, pour pouvoir choisir entre deux acides aminés, ou bien entre un acide aminé et une STOP position, dans la situation de la rencontre d'un ribosome et de codons syoms qui disposent d'une capacité de double codage.

Ceci avait été justifié théoriquement précédemment [1, 2, 3]. Le travail expérimental [4] est venu confirmer une telle théorie, quand il a été démontré que deux acides aminés distincts, la sélécystéine et la cystéine, sont codés par le même codon UGA pour l'infusoire *Euplotes Crassus*. Ce résultat ne jette pas un doute sur le dogme du codage univoque des acides aminés et des STOP positions, par les génomes des cellules, mais il implique d'introduire des rectifications dans le modèle du codage génétique.

Ces rectifications se fondent sur une conception élargie du rôle particulièrement signifiant du 3ième nucléotide dans les codons et sur l'acceptation d'une textualité véritable et non métaphorique des gènes des protéines (ARNm). Une telle interprétation des gènes (ARN messagers) comme un quasi-langage, et du rôle qu'y joue le 3ième nucléotide, dans les codons, nous amène à une simple conception de quasi-intelligence (biocomputing) du système qui synthétise la protéine, et à sa capacité d'évaluer le contexte (le sens) de l'ARNm, afin de faire le bon choix de l'acide aminé ou du STOP position, dans le cas d'une situation de syom, en partant du sens des textes des gènes de l'ARNm.

Mots clés : Information génétique, codons, syom, syomie.

L'hypothèse indéfinie (wobble) de F. Crick

On a beaucoup écrit sur l'hypothèse indéfinie (wobble) de Francis Crick, y compris l'auteur lui-même de cette hypothèse. Mais la plupart des jugements se fondent sur la formulation utilisée par F. Crick, dans son livre « Quelle recherche démente » 1988 [5].

Voici ces paroles fondatrices : « *Et bien que le code génétique applique des lois précises, il est important de noter, que dans de nombreux cas précisément deux bases codent pour un acide aminé, et la troisième ne joue aucun rôle. Dans le cas contraire la structure du code n'aurait aucun sens évident* »[1]

Mais il y a des faits significatifs et complémentaires qui découlent de ce bref postulat. C'est l'objet de la présente étude. Le code génétique « standard » des protéines a été obtenu par l'équipe de M. Nirenberg à partir de l'étude de la synthèse des protéines du bacille *E. Coli*. Le tableau du code génétique standard, est le bilan de ce travail. Il exprime les fonctions des gènes des protéines, comme une structure statique de codage, dans laquelle tous les codons codent EXPLICITEMENT les acides aminés et les STOP positions (les points d'arrêt de la synthèse de la protéine). Ce qui importe, en conformité avec la wobble hypothèse, c'est que la moitié des 64 codons connus, soit 32, sont en surnombre par rapporte aux 20 acides aminés connus. Pour ce qui est du 21ième acide aminé, la sélécystéine, voir plus bas. Les codons en excès sont des codons synonymes, qui avec une redondance variable, codent pour les mêmes acides aminés ou des STOP positions. Ce sont là les prémices de la position de M. Nirenberg et de F. Crick qui s'est associé à lui. Cette notion règne depuis 50 ans, depuis le jour de l'attribution du prix Nobel à M. Nirenberg pour son modèle, en 1968. De nos jours il s'est accumulé de

[1] Dans original: "An important point to notice is that although the genetic code has certain regularities—in several cases it is the first two bases that encode one amino acid, the nature of the third being irrelevant—its structure otherwise makes no obvious sense."

nombreux résultats théoriques et expérimentaux qui impliquent d'apporter certains compléments à cette façon d'interpréter le codage génétique des protéines. Voici ces compléments :

Les facteurs d'univocité et de dégénérescence du code de protéine de *E. Coli*

Le tableau du code standard est divisé fonctionnellement en deux parties où l'une est constituée des 32 codons qui codent uniquement des acides aminés, univoquement et de façon redondante, (ce sont les codons synonymes). L'autre moitié est constituée de 32 codons (non synonymes), qu'ont peut appeler homonymes [1, 2, 3], et qui codent NON SANS AMBIGÜITÉ les acides aminés et les STOP positions, et pas toujours en conformité avec le tableau du code standard. Plus précisément, chaque codon homonyme code simultanément pour deux acides aminés ou pour un acide aminé et une STOP position. Cela signifie qu'il est indispensable, pour une synthèse protéique exacte, qu'ait lieu un CHOIX entre deux acides aminés possibles, ou bien entre un acide aminé et une STOP position. Le décodage de l'acide aminé ou de la STOP position peut dans ces cas ne pas correspondre au tableau du code standard, car le ribosome identifie et choisit le bon acide aminé ou la STOP position, d'après les codons-homonymes, dans la dynamique de sa lecture de l'ARNm, appuyée par une analyse logique de son contexte.

Ceci est en contradiction avec le dogme de M. Nirenberg et de F. Crick concernant le codage univoque, que l'on a considéré comme indestructible jusqu'aux travaux [1, 2, 3] et la monographie [4], dans laquelle il est prouvé expérimentalement que le codon de l'acide aminé de la sélécystéine code également pour un autre acide aminé : la cystéine. Ceci a donné un motif de douter de l'évidence de ce dogme et de chercher une explication à ce phénomène, non comme destructeur du dogme du codage univoque, mais comme une confirmation d'un point de vue linguistique, du principe de l'homonymie, c'est à dire d'une réelle textualité (non métaphorique) des gènes de l'ARNm.

Ceci se produit dans la synthèse des protéines, comme l'opposé de son contraire, qui est le codage d'un acide aminé par divers codons, en situation de synonymie linguistique. Ce dernier correspond à la wobble-hypothèse de F. Crick, et il est prouvé expérimentalement, par la présence des ARNt iso-accepteurs. On pourrait définir de façon plus globale la synonymie et l'homonymie comme suit : la synonymie serait la représentation du codage de l'unité par la multitude diversifiée, alors que l'homonymie serait la représentation du codage par l'unité d'une multitude diversifiée.

Ceci est présenté de façon nouvelle dans la figure 1 du code génétique, où l'on voit la division fonctionnelle et symétrique du code, en codons synonymes et homonymes.

Le regroupement par familles est réalisé d'après le schéma de Lagerkvist [6], dans lequel le facteur familial est constitué par les deux premier nucléotides des codons (des triplets). Mais nous avons regroupé les familles autrement : selon le critère de synonymie ou d'homonymie

Figure 1 : Tableau du code génétique

En rouge= les codons mixtes : syoms

En bleu= les codons synonymes

	C	G	T(U)	A
T(U)	TCT Ser TCC Ser TCA Ser TCG Ser	TGT Cys TGC Cys TGA Stop TGG Trp	TTT Phe TTC Phe TTA Leu TTG Leu	TAT Tyr TAC Tyr TAA Stop TAG Stop
A	ACT Thr ACC Thr ACA Thr ACG Thr	AGT Ser AGC Ser AGA Arg AGG Arg	ATT Ile ATC Ile ATA Ile ATG Met	AAT Asn AAC Asn AAA Lys AAG Lys
C	CCT Pro CCC Pro CCA Pro CCG Pro	CGT Arg CGC Arg CGA Arg CGG Arg	CTT Leu CTC Leu CTA Leu CTG Leu	CAT His CAC His CAA Gln CAG Gln
G	GCT Ala GCC Ala GCA Ala GCG Ala	GGT Gly GGT Gly GGA Gly GGG Gly	GTT Val GTC Val GTA Val GTG Val	GAT Asp GAC Asp GAA Glu GAG Glu

Nous voyons que le tableau est partagé symétriquement, en codons synonymes et en codons homonymes. Ce tableau est emprunté dans la monographie [4].

Comment se fait le choix d'acides aminés et des STOP positions dans le cas de la rencontre du ribosome et d'un codon homonyme de l'ARN messager

Un tel CHOIX est fait par le ribosome au prix d'une prise en compte par lui-même (ou par la cellule entière) du contexte de l'ARNm présent. La manifestation de ce choix signifie automatiquement l'existence d'une intelligence du système qui procède à la synthèse de la protéine, l'existence de ses capacités de calculateur biologique (biocomputing) [7]. La quasi intelligence est présente parce que l'ARNm (copie du gène) est un texte, au sens propre du terme, et non au sens métaphorique [1, 2, 3].

La situation UNIVOQUE du codage par les SYNONYMES est déterminée par le fait que dans chacune des 8 familles de codons TOUS les TRIPLETS (les codons sont constitués de trois nucléotides chacun) sont différents. À cause de cela, dans ces familles, le codage implique les trois nucléotides constituants le triplet (codon), et dans chaque famille le triplet ne code que pour un seul acide aminé.

La situation NON UNIVOQUE PRIMAIRE du codage par les triplets HOMONYMES est présente dès le début (avant la lecture de l'ARNm par le ribosome) chez la moitié des codons non synonymes (en fait chez les homonymes). Cela dépend du fait que le 3ième nucléotide des codons, qui est un participant actif dans le fonctionnement du génome-biocalculateur de chaque cellule, avant l'acte de lecture de l'ARNm par le ribosome, il « ne fait pas de projections » sur l'état de la statique de participation au codage, et cela peut être potentiellement n'importe lequel des quatre nucléotides possibles. Je rappelle que F. Crick n'a pas commenté, d'aucune manière de telles situations du dynamisme du travail de ribosome. Mais dans deux des familles il se passe la chose suivante : Le doublet de la famille TA code deux fois pour la Tyrosine et un STOP, alors que dans les deux doublets TG, l'un des doublets code par deux fois pour la Cystéine, alors que dans le second doublet de la famille des TG l'une des paires code pour la Cystéine alors que l'autre pais code pour un STOP et une Tryptophane.

Cela signifie globalement, que dans ce cas également il y a présence du facteur d'homonymie, mais avec des caractéristiques complémentaires importantes. Ce phénomène avait été découvert par l'équipe de M. Nirenberg et de F. Crick, sur l'exemple de la familles des codons T(U)T(U) [8], lorsque le triplet UUU code simultanément pour la Phénylalanine et la Cystéine. Cette simultanéité n'avait pas été comprise par M. Nirenberg et de F. Crick, et ils ne l'avaient pas évaluée comme contradictoire avec le postulat du codage UNIVOQUE de tous les 64 codons, qu'ils avaient avancé. Ceci a duré jusqu'aux travaux de Touranov et d'autres [4], où ils ont démontré une simultanéité semblable pour la Sélénocistéine et la Cystéine, ce qui avait été découvert depuis longtemps pour le codon UUU [8] par M. Nirenberg et de F. Crick.

C'est précisément après le travail [4] où l'ambiguïté du codon UGA a été démontrée expérimentalement ainsi que la justification théorique de ce phénomène [2, 3, 4], que sont apparus les premiers doutes sur l'évidence du dogme du codage univoque des acides aminés et des STOP positions par les 64 codons.

Il est plus aisé d'apprécier cette information complémentaire et fondamentale concernant l'homonymie de la moitié des codons, grâce à la nouvelle présentation du code génétique (Figure 2). Elle est extraite du travail [1].

Figure 2: La binarité homonymo-synonymique du code génétique

synonymie

Asp	Glu	Lys	Gln	Gln	His	Leu	Phe	Ileu	Met
GA_C	GA_A	AA_C	AA_A	CA_A	CA_C	UU_A	UU_C	AU_A	AU_G
GA_U	GA_G	AA_U	AA_G	CA_G	CA_U	UU_G	UU_U	AU_C	
								AU_U	

Arg	Ser	Trp	Stop	Tyr	Stop
AG_A	AG_C	UG_G	UG_A	UA_C	UA_A
AG_G	AG_U	UG_T	UG_C	UA_U	UA_G

homonymie

La fonction biologique d'un tel dualisme synonymo-homonymique (dans la limite des familles mixtes des codons) serait une encore meilleure flexibilité du code. Une telle dualité signifie en fait une hybridation des possibilités du code dans les huit familles mixtes de codons. C'est pourquoi il serait plus facile de les dénommer des familles SYOM (De la fusion des mots SYnomyme et hOMomnyme), et d'appeler cette caractéristique du code, une SYOMIE.

Il est substantiel que le tableau des protéines du code de *E. Coli*, adopté par la communauté scientifique, soit STATIQUE, et n'exprime pas le facteur essentiel et emblématique du dynamisme de la synthèse biologique des protéines in vivo. C'est l'une des raisons de l'incompréhension partielle par une majorité, de la fonction clé du 3ième nucléotide dans les codons syoms, qui amènent le génome au niveau d'une véritable textualité (et non métaphorique) des structures de l'ADN et de l'ARN.

La syomie ouvre au système qui synthétise les protéines, des horizons sans limite (quasiment linguistique) du sens du codage. Cela est particulièrement significatif pour les fonctions des neurones du cerveau de l'être humain, où les actes de pensée-conscience se réalisent, probablement, par la voie d'une matérialisation des enchainements de textes

formés d'ADN-ARN-Protéine de courte durée de vie, sous forme de champs physiques de matérialisation d'équivalents des pensées. [9]

Nous observons de cette façon une dégénération synonyme-homonymique (une syomie) du code génétique des protéines, ce qui constitue une certaine rectification de l'ancien modèle du code.

Ce fait exprime l'unicité des antithèses des fonctions des codons synonymes et des codons syoms, où les synonymes apparaissent comme acteurs d'un codage excédentaire et précis, alors que les syoms représenteraient la souplesse et l'adaptabilité du code aux variations du milieu extérieur.

Rectification de l'hypothèse indéfinie de F. Crick

Dans le cas des synonymes, les quatre nucléotides (T, C, A, G) en troisième position dans les codons, peuvent se substituer réciproquement librement. Cela n'influe pas sur leur fonction de codage. Mais pour les syoms, cela ne correspond pas, par principe, au tableau du code standard officiel. Ceci ne veut pas dire un refus d'un codage univoque. L'univocité s'obtient par une orientation du ribosome, par le contexte de l'ARNm en cours. Dans le cas d'un décalage de la mire de lecture, des substitutions artificielles ou de mutation du 3ième nucléotide peuvent survenir, qui vont introduire des choix anormaux, liés au contexte, du codage des acides aminés ou/et des STOP positions, dans les syoms, au niveau du transcodage de l'ARNm en protéine. Ceci est possible, si lors du décalage de la mire de lecture, le 3ième nucléotide du syom venait à occuper la place du 1er ou du 2ième nucléotide dans le syom.

De telles substitutions sont aléatoires et ne sont pas indifférentes pour la synthèse protéique, comme l'exige l'hypothèse imprécise de F. Crick. Les significations des deux premiers nucléotides sont imposés par le contexte d'ARNm, alors que les rôles du 3ième nucléotide seraient double :

-participer au codage avec « une fonction déléguée »,

-en complément, renforcer purement physiquement la paire codon- anticodon sur le ribosome.

Qu'est ce qu'une « fonction déléguée » ? Comme l'ordre des nucléotides dans le gène de protéine, et donc dans l'ARNm, est rigide (héréditaire), alors, lors de la lecture de l'ARNm par le ribosome et la rencontre d'un doublet de codon syom, (avec les 1er et 2ième nucléotides du triplet) il apparaît une indétermination liée au 3ième nucléotide indéfini (wobbling) selon F. Crick (et conformément à la Nature du code). Quel est son rôle emblématique dans le texte du codon syom ? Il se réalise probablement par délégation de la fonction emblématique manquante du 3ième nucléotide du syom, selon le schéma des orientations par contexte. Voici un exemple pris dans la linguistique: voici deux propositions : « il a vu un rDt dans son lit», et « il a entendu un rNt dans le bus». Compte tenu du contexte, dans la première proposition, il faut dans le mot rDt déléguer la fonction de la lettre A, à la lettre D, et dans la seconde proposition, au mot rNt, il faut déléguer la fonction de la lettre O à la lettre N. Dès lors les doublets syom perdent leur indétermination de sens et dans le cadre d'un triplet, précédemment syom, reconstitué, ils acquièrent une sémantique précise et UNIQUE du triplet codant : le choix d'un acide aminé ou d'une STOP position. Le dualisme du syom est perdu et l'univocité est rétablie. Le codage univoque est acquis, mais dans la dynamique de l'acte de lecture de l'ARNm. Telle est la conséquence stratégique de la dégénérescence synonymo-homonymique

(syomie, dualité) du code des protéines. Le principe est simple, comme tous ce qui est génialement « conçu » par la Nature. Ceci est apparenté au transcodage des codons, connu depuis bien longtemps dans les situations de stress des systèmes biologiques (choc thermique, présence d'antibiotiques exogènes, carence d'acides aminés), mais leur relation aux doublets temporairement incompréhensibles, des codons syoms, est inexplicable pour l'instant. Il existe en linguistique un exemple canonique du rôle de la fin des mots dans la proposition (dans le contexte) pour déléguer le sens des mots incompréhensibles précédemment…la fin des mots délèguent un sens relativement précis. Il est probable que dans les textes ADN et ARNm un tel exemple soit employé. On peut voir la dualité synonymo-homonymique des codons syoms dans la figure 3, pour l'exemple de la famille des codons TA. Dans la famille TA, la synonymie par pair a bien lieu. Mais cela ne se propage qu'à moitié sur la famille TG : Il n'y a pas de synonymie entre TGA et TGG. On y code STOP et Trp. Ceci est exceptionnel. Toutes les autres familles de syoms, la synonymie par paires est respectée. De même que l'homonymie d'opposition par paires.

Dans les codons-syoms la substitution des 3^{ièmes} nucléotides induiront des choix liés au contexte du codage des acides aminés et des STOP positions. De telles substitutions sont aléatoires et ne peuvent se produire que dans le cadre de mutations radioactives, chimiques ou induites artificiellement. Seules ces mutations sont capables de substituer, ou plus exactement abimer, le 3ième nucléotide dans les codons- syoms, rigidement imposés par l'hérédité.

Il est possible de supposer l'existence de la règle suivante : Dans les codons-syoms, les 3^{ièmes} nucléotides adoptent le sens de l'un des quatre nucléotides A,U,G,C que leur attribue le nano calculateur du ribosome, dans son processus de lecture de l'ARNm. À son tour, un tel choix détermine quel complexe « acide aminé- ARNt- anticodon :codon-syom » sera activé, suivi de l'inclusion de l'acide aminé choisi dans la chaîne peptidique en construction.

Figure 3: La dégénération binaire homo-synonymique du code des triplets protéiniques

TAT TYR	TAC TYR	TAA STOP	TAG STOP
TAC TYR			
TAA STOP			
TAG STOP			

Synonymes couplés

Homonymes couplés antagonistes

Pourquoi les STOP codons se trouvent dans les familles de siomies

La terminaison, la fin de la synthèse de la protéine, se produit lorsque l'un des STOP codons, AUG, UAA, UGA apparaît sur le site A du ribosome. Le peptidile-ARNt reste lié au site P du ribosome, à cause de l'absence d'ARNt correspondant à ces codons. C'est là qu'entrent en jeu les protéines spécifiques RF1 ou bien RF2, qui catalysent le décrochage de la chaîne peptidique de l'ARNm, mais aussi RF3 qui provoque la dissociation de

l'ARNm du ribosome. RF1 va reconnaître dans le site A la présence d'UAA ou UAG, et RF2 va reconnaître UAA ou UGA.

Ceci est précédé d'un événement important : la décision par les trois codons STOP (syoms) d'arrêter la synthèse. Dans le cas présent « la décision » n'est pas une simple métaphore, mais le résultat du travail du nano biocalculateur, qui est probablement le système produisant la synthèse protéique [7]. C'est précisément le nano biocalculateur qui analyse le CONTEXTE des séquences de l'ARNm, et c'est qu'alors, et alors seulement que l'un des trois triplets syoms ambigus (soit un STOP, soit un acide aminé) acquière son sens soit d'un STOP, soit d'un acide aminé. Pourquoi est-ce exactement ainsi ?

Imaginez que les fonctions STOP aient appartenu à quelque codon synonyme. Dans ce cas la fonction stratégique d'analyse du sens du texte des gènes (de l'ARNm) serait perdue. En effet les synonymes codent les acides aminés de façon rigide, univoque et redondante, ce qui découle de l'invariabilité naturelle des textes natifs des gènes d'ARNm. À l'opposé du fonctionnement univoque des codons synonymes, les syoms-STOP vivent dans l'ARNm dans un régime d'attente du sens du contexte de l'ARNm (du gène). Selon ce sens, il est pris une décision sur la signification exacte du codon syom ambigu : devenir un codon d'acide aminé et continuer la synthèse ou bien l'arrêter comme l'indique la STOP position.

Discussion

Dans cette étude nous développons l'idée non contradictoire, que le ribosome ou le système complet synthétisant les protéines, « choisit » les acides aminés et les STOP positions, lors de la mise en présence du ribosome avec des codons non synonymes (des codons-syoms), en lisant eu en évaluant réellement le sens du contexte de l'ARNm. Le choix se fait entre deux anticodons identiques ARNt, porteurs de deux acides aminés distincts. Ces anticodons sont identifiés, évalués et sélectionnés par le complexe « ribosome+ codons-syoms dans le contexte e l'ARNm » en fonction du sens de l'ARNm. Le choix de la sémantique des syoms-codons et en conséquence de l'un des deux ARNt, porteur de deux acides aminés distincts ou d'un acide aminé et d'une STOP position, se fait selon l'orientation de sens du ribosome, dans le contexte de l'ARNm, en tant que nano biocalculateur. Cette situation, à première vue, semble contredire le canon de la génétique, concernant le codage génétique univoque de tous les acides aminés. Mais un tel « choix » ne rejette pas la thèse fondamentalement juste du codage univoque des acides aminés lors de la synthèse protéinique. L'apparente contradiction est levée par a fonction particulière du 3ième nucléotide, dans les 32 codons syoms, non synonymiques. L'importance stratégique de ce fait réside dans le déplacement du codage des protéines dans le domaine des lois de la linguistique, précédemment ignorées pour le travail du génome. La fonction du 3ième nucléotide dans les 32 codons syoms, réside dans un marquage sémantique distinctif des familles de triplets de nucléotides synonymiques ou synonymo-homonymiques (syoms) participants à la synthèse des protéines. Dans ce processus ont cours des lois linguistique, comparables à la parole, de la structuration des textes protéiques (la parole) à l'aide des lettres constituées d'acides aminés. Une telle représentation du fonctionnement du génome cesse d'être métaphorique et acquière un sens exact, fondé non seulement sur la logique mais aussi sur une preuve expérimentale du dualisme complexe et mélangé de la sémantique des codons-syoms [4].

La métaphore du « choix » des acides aminés et des STOP positions lors de la synthèse des protéines, cesse d'être une métaphore, et devient l'un des faits scientifiques

d'une Génétique de Mendel, et d'une biologie moléculaire plus avancées. Fondé sur une analyse théorique [1, 2, 3] et une vérification expérimentale [4] on analyse le rôle du 3ième nucléotide dans les codons syoms, pendant la synthèse protéique. Sa signification est élargie par rapport aux conceptions habituelles.

Le 3ième nucléotide partage symétriquement et fonctionnellement les familles des codons en 32 codons synonymes, et 32 codons syoms. Les codons syoms disposent de la fonction stratégique de participer au déclenchement de l'analyse délocalisée, par le ribosome, en tant que nano biocalculateur, du contexte réel de l'ARNm, dans le langage de l'ARNm. Une telle analyse est une nécessité naturelle permettant le choix d'un acide aminé parmi deux, ou du choix entre un acide aminé et une STOP position, dans le cas de la rencontre avec un codon syom qui possède la capacité d'un double codage. Ceci avait été justifié précédemment [1, 2, 3]. Le travail expérimentale [4] a conformé la théorie, lorsqu'on a montré que deux acides aminés distincts, la Sélécystéine et la Cystéine étaient codé par le même codon syom UGA chez le monocellulaire *Euplotes Crassus*, ce qui est en principe applicable, dans une certaine mesure, à l'être humain. Ce résultat ne met pas en question le dogme du codage univoque des acides aminés et des STOP positions, par le génome des cellules, mais il implique d'introduire des rectifications importantes dans le modèle accepté depuis longtemps et non remis en cause, du codage génétique. Ces rectifications sont basées sur une conception élargie du rôle emblématique particulier du 3ième nucléotide dans les codons et de l'acceptation de l'idée d'une textualité réelle (et non métaphorique) des gènes des protéines.

Une telle conception d'une similitude linguistique des gènes (ARNm) et du rôle du 3ième nucléotide dans les codons, nous entraine vers une acceptation simple d'une quasi intelligence (biocomputing) du système de la synthèse protéique et sa capacité à évaluer le contexte actuel, réel, (le sens) de l'ARNm pour prendre des décisions sur le choix des acides aminés et des STOP positions, en partant de la signification des textes des gènes(ARNm).

Pourquoi les caractéristiques complémentaires du code des protéines proposé ici est-il plus pragmatique que le modèle de M. Nierenberg et F. Crick, qui est en soi tactiquement juste mais stratégiquement immature ? Et pourquoi les tentatives de voir dans le code quelque chose de plus grand que ce qu'y ont vu ses créateurs, ont-elles eu si peu de succès ? Ces tentatives ont été faites dans les travaux de Lagerkvist et Roumer [10]. Lagerkvist s'est trompé en considérant que les codons[2] mixés (les syoms selon notre terminologie) n'apparaissaient que rarement dans l'ARNm, alors que Roumer a vu des symétries dans le code génétique, en le classant selon la force des liaisons hydrogène codon-anticodon, ce qui est assez proche du partage des familles de codons en synonymes et syoms. Mais ils n'ont pas vu que la synthèse des protéines serait correcte si les codons étaient divisés en deux groupes mutuellement complémentaires , comme ça l'est dans la réalité. L'un de ces deux groupes, les synonymes, assure l 'exactitude et la redondance du codage des acides aminés, alors que l'autre groupe, les syoms, permet la flexibilité et l'adaptabilité aux changements du milieu, des protéines synthétisées, au prix d'un changement composition du pool des acides aminés et de leurs séquences. C'est là que réside la sagesse du code génétique.

Dans le contexte de cette façon de voir, la publication de Lolle et d'autres est fort intéressante : https://uwaterloo.ca/biology/people-profiles/susan-j-lolle

concernant la génétique de restitution de certaines plantes [11]. Il est montré dans ce

[2] « codons mixés » est un terme introduit par Lagerkvist.

travail qu'il n'y a pas de distinction entre la séquence du gène Ler du type rustique et du gène HTH de l'Arabidopsis thaliana mutante, qui répondent à la liaison directe entre les propriétés biologiques de la cuticule, l'adhérences cellulaire et la prolifération d'Arabidopsis. Les auteurs écrivent : «In every case the sequence of the reverted HTH allele matched the Ler wild-type sequence exactly» (dans chaque cas l'allèle inversé, HTH coïncidait parfaitement avec la séquence Ler de l'espèce sauvage).

Cela signifie que Lolly et Priuitt avaient découvert chez Arabidopsis, l'effet de régénération d'une partie de la génétique vers l'ancêtre. Ce qui est fantastique dans ce fait, c'est que le gène sauvage et le gène mutant qui s'est reconstitué sont séquentiellement identiques, ce qui est inexplicable depuis l'ancienne position mendélienne. Mais c'est explicable du point de vue de la génétique ondulatoire et linguistique. Pourquoi le même gène donne-t-il des expressions phénotypiques différentes ?

Pour obtenir une réponse, dans le cadre de la rectification envisagée du modèle du code des protéines, il est indispensable de vérifier la colinéarité des ARNm et de leur production protéique chez les gènes sauvages et mutants. On peut prévoir que les séquences de acides aminés de ces gènes seront distinctes. Les séquences d'acides aminés vont diverger par la composition en acides aminés , car les séquences adjacentes d'ADN du 3ième et le 5ième codons des gènes sauvage et mutant, sont distincts, ce qui donne une variation du contexte et consécutivement une variation du sens des mêmes codons dans l'ARNm des gènes sauvage et mutant. Les auteurs écrivent que le haut niveau de régression du mutant vers le sauvage, donnait au niveau des nucléotides une copie exacte du gène sauvage, qui était observable dans les génération précédentes. Malheureusement les séquences des nucléotides des secteurs des génomes sauvage et mutant, que citent les auteurs, ne sont pas découpés en codons. Mais une autre chose est évidente : La séquence ADN des terminaisons 3 et 5 des deux gènes sont distincts, et en conséquence leur contenu contextuel d'ARNm est distinct. Cela permet de pronostiquer une divergence des séquences d'acides aminés des protéines produites par deux gènes « pseudo » identiques et évidemment des morphogénèses distinctes pour les parties de la plante codées par ces gènes.

Une analyse détaillée du travail de Lolly et d'autres [11], en relation avec ceux de Touranov et d'autres [4], sont intéressants en ce que leurs résultats fondamentaux induisent la génétique à des explorations nouvelles dans la stratégie du codage génétique des protéines. Nous voyons bien qu'il y a là encore beaucoup de choses à préciser. À partir de ces positions nouvelles en génétique, on peut subodorer des menaces possibles dues aux erreurs dans les technologies recombinatoires de l'hybridation artificielle de divers

gènes, qui induiraient à un abus sémantique au niveau des sens des ARNm, définissant le choix et l'exactitude du codage des acides aminés et des STOP positions par les codons – syoms.

Le paradoxe de la génétique réside dans le fait qu'après 50 d'existence du modèle du code génétique, universellement accepté, il n'ait pas été fait une vérification minutieuse et massive sur des centaines de protéines, et une vérification de colinéarité (protéines-ARNm) appuyée par la statistique. Si on découvre des divergences par rapport au tableau standard du code des protéines de *E Coli*, cela ne va pas contester le modèle de M. Nierenberg et F. Crick,. Cela signifiera l'insondabilité des principes du codage des protéines, particulièrement dans la direction linguistique, et quasiment textuelle.

Bibliographie

1. Gariaev P.P., 1997, Wave genetic code. Monograph. Institute of Control Sciences of the Russian Academy of Sciences. 107 p. (In Russian).

2. Gariaev P.P., 2009, Linguistic-wave genome. Theory and practice. Monograph. 216 p. (In Russian).

3. Gariaev P.P., 2015, Another Understanding of the Model of Genetic Code Theoretical Analysis. Open Journal of Genetics, 2015, 5, 92-109. Published Online June 2015 in SciRes. http://www.scirp.org/iournal/oigen http://dx.doi.org/10.4236/ojgen.2015.52008 Institute of Quantum Genetics LLC. Moscow, Russia.

4. Turanov A.A. et al., 2009, Genetic Code Supports Targeted Insertion of Two Amino Acids by One Codon. Published 9 January 2009, Science 323, 259 (2009). DOI: 10.1126/science.1164748

5. Crick F., 1988, what mad pursuit. A personal view of scientific discovery, p. 90.

6. Lagerkvist U., 1978, ―Two out of Three‖: an alternative method for codon reading. Proc. Natl. Acad. Sci. USA., 1978.V. 75. P. 1759-1762.

7. Gariaev P.P., Birshtein B.I., Iarochenko A.M., Marcer P.J., Tertishny G.G., Leonova K.A., Kaempf U., 2001, The DNA-wave bio computer. ―CASYS‖ – International Journal of

Computing Anticipatory Systems (ed. D.M. Dubois), Liege, Belgium, v.10, pp.290-310.

8. Crick F., Nierenberg M., 1964. The genetic code. Successes of physical sciences., v.LHHH11. In Russian.

9. Gariaev P., Leonova-Gariaeva E.A., Nonlocal functions of DNA-RNA proteins in brain and Consciousness-Thinking (is being prepared)

10. Rumer Yu.B., 1968, The codification of codons in the genetic code. Reports of the Academy of Sciences of the USSR. (1968) V.183, N 1, P.225-226.

11. Lolle S.J., Jennifer L.V., Young J.M., Pruitt R.E.

1.
inheritance of extra-genomic information in Arabidopsis, Nature, 434:505-09, March 24,

, 2005, Genome-wide non-mendelian

1.