

Des ARNs circulaires possibles chaînons de l'évolution de la traduction médiée par les ARNt : les affinités codons-acides aminés augmentent en fonction de l'ancienneté des ARN circulaires

Comptes Rendus Biologies 343 (2020) 111-122

Jacques Demongeot ; Hervé Seligmann

(traduction post-éditée par N. Bacaër, suggestions d'amélioration : nicolas.bacaer@ird.fr)



Résumé

Les affinités nucléotidiques des interactions non covalentes avec les acides aminés produisent des associations entre les ARNm et les peptides apparentés, régulant potentiellement la traduction ribosomale. Les corrélations entre les affinités nucléotidiques et l'hydrophobicité des résidus sont explorées pour 25 ARNs circulaires théoriques d'une longueur minimale de 22 nucléotides, conçus *in silico* pour coder chaque acide aminé une fois et une seule après trois tours de traduction et former des structures « hairpin » (épingle à cheveux) maximales. Cette conception imite vraisemblablement les premiers ARNs de la vie. Les ARNs circulaires ressemblent à des ARNt consensuels, suggérant une fonction de proto-ARNt, mettant en jeu un anticodon prédit et un acide aminé apparenté. Les 25 ARNs circulaires et leur ordre d'évolution présumé, déduits de l'ordre d'intégration dans le code génétique de l'acide aminé apparenté à leur anticodon prédit, montrent des associations remarquables avec plusieurs propriétés anciennes du mécanisme de traduction de la cellule. Ici, nous les utilisons pour explorer l'évolution des corrélations entre hydrophobicité d'un résidu et affinité à son codon, en supposant qu'elles reflètent des propriétés de la traduction pré-ARNt et pré-ribosomique. Cette hypothèse suppose que les corrélations diminuent avec les ordres d'inclusion dans le code génétique du codon apparenté à un ARN circulaire. Les associations entre les affinités des nucléotides et les hydrophobicités des résidus ressemblent à celles existant actuellement entre gènes et protéines. Les forces d'association diminuent avec les rangs d'inclusion dans le code génétique des acides aminés apparentés au proto-ARNt. La conception des ARNs circulaires minimaux ne tient pas compte des affinités entre ceux-ci et les peptides qu'ils codent. Cependant, les interactions entre les ARNs circulaires et leurs peptides apparentés ressemblent à celles observées pour les gènes naturels actuels. Cette propriété est plus forte pour les ARNs circulaires anciens et plus faible pour les récents, couvrant une période au cours de laquelle la traduction moderne à base d'ARNt et de ribosomes a probablement évolué. Les résultats indiquent que la traduction usant d'adaptateurs fondés sur les affinités codon-acide aminé et sur le code génétique primitif a préexisté à une traduction médiée par l'ARNt. Les ARNs circulaires minimaux théoriques apparaissent ainsi comme des modèles prébiotiques pour la transition entre les traductions à médiation pré-ARNt et ARNt.

1. Introduction

Les attributions de codon -acides aminés selon le code génétique ne sont pas aléatoires [1] car de nombreuses propriétés de codons et d'anticodons sont corrélées avec les propriétés d'acides aminés apparentés [2 , 3 , 4]. De nombreuses analyses indiquent que le code génétique minimise les impacts des erreurs pour divers processus et propriétés: mutations sur la structure des protéines [5 , 6], comment les protéines se replient [7 , 8 , 9], mauvaise charge de l'ARNt par les ARNt synthétases [10 , 11 , 12 , 13], les décalages de trames pendant la traduction avant et après ceux-ci (avant l'apparition [14 , 15 , 16 , 17], après l'événement [18 , 19 , 20 , 21 , 22 , 23]). Le code génétique maximise la diversité des propriétés

physicochimiques des acides aminés codés [24 , 25] et optimise l'erreur en fonction de plusieurs propriétés à la fois [1].

1.1. Évolution du code génétique

De nombreuses hypothèses sur les ordres d'inclusion des acides aminés dans le code génétique ont été déduites de différentes propriétés et expériences. Ces hypothèses prédisent les ordres selon lesquels les acides aminés ont été attribués aux codons [26 , 27]. La plupart de ces hypothèses indépendantes prédisent des ordres congruents, justifiant un ordre consensuel d'inclusion d'acides aminés dans le code génétique [26 , 27]. Des analyses ultérieures ont produit des ordres d'inclusion de code génétique qui convergent avec ces rangs d'inclusion d'acides aminés. La position moyenne des types d'acides aminés dans les protéines modernes est en corrélation avec ces rangs d'inclusion. La position moyenne des acides aminés anciens est plus proche de la protéine C-terminale, les acides aminés récemment inclus sont en moyenne plus proches de la protéine N-terminale, qui correspond à l'extrémité 5 'du gène avec le codon d'initiation / start [28].

1.2. Attributions codon-acide aminé du code génétique stéréochimique

Cela montre que les structures des biomolécules modernes intègrent des informations fossilisées sur les origines de la vie. En effet, les structures cristallines des rayons X du ribosome montrent que les contacts entre les triplets de nucléotides d'ARN ribosomique et les résidus de protéines ribosomiques favorisent les attributions d'acides aminés-codons / anticodons. Les acides aminés qui ont intégré tôt le code génétique favorisent les contacts avec leurs codons, les acides aminés tardifs favorisent les contacts avec leurs anticodons [29]. Cela montre deux étapes différentes pendant la formation du code génétique: (1) une traduction sans adaptateur / ARNt où les codons interagissent directement avec les acides aminés, et (2) une traduction basée sur l'ARNt. Les deux résultats confirment, au niveau de la structure ribosomale, l'hypothèse selon laquelle (au moins une partie) des attributions du code génétique résultent d'affinités nucléotides-acides aminés dues à des interactions stéréochimiques. L'importance des interactions stéréochimiques n'est pas surprenante, car elles ont également déterminé l'homochiralité des principales molécules de la vie, l'ARN droitier (ARN-R) et les acides aminés gauchers (acides aminés L), qui interagissent mieux lorsque ceux-ci ont une main opposée [30 , 31 , 32 , 33 , 34]. La première étape de l'évolution du code génétique déduite des structures cristallines ribosomales suggère des interactions stéréochimiques directes codon-acide aminé. Ceci est conforme aux observations selon lesquelles dans la plupart des ARNm, les affinités nucléotidiques pour les acides aminés sont en corrélation avec les besoins polaires des acides aminés [35 , 36 , 37]. Ces observations impliquent que les fréquences des contacts observés entre les nucléotides et les acides aminés, à partir desquels les affinités nucléotides-acides aminés sont déduites, sont approximées par l'hydrophobicité des acides aminés, car ces interactions sont des interactions hydrophobes au sens large. Cela implique que les attributions de codes génétiques résultaient d'interactions stéréochimiques avant que les proto-ARNt ne passent de leur fonction d'initiateurs de polymérisations ARN / ADN [38 , 39] à leur principale fonction moderne de traduction. Les traductions par interactions directes codon-acide aminé s'inscrivent dans un monde peptide-ARN primordial, plutôt qu'avec un monde ARN primordial [40].

1.3. Anneaux d'ARN minimaux théoriques comme modèles pour le monde primitif de l'ARN

Nous étudions ici les associations entre les affinités des codons pour leur acide aminé apparenté et l'hydrophobicité de cet acide aminé, spécifiquement dans le contexte de 25 anneaux d'ARN minimaux théoriques [41 , 42 , 43 , 44 , 45]. Ces ARN longs de 22 nucléotides ont été conçus pour coder pour un codon de début et un codon d'arrêt et une fois pour chacun des 20 acides aminés du code génétique par trois cycles de traduction consécutifs (figure 1). Cela garantit une capacité de codage maximale sur la longueur la plus courte possible. Leur conception a également été contrainte de former une épingle à cheveux en boucle. Des analyses indépendantes de la structure secondaire ont confirmé que la plupart des anneaux d'ARN minimaux théoriques ressemblent aux proto-ARNt [44 , 45]. Un acide aminé apparenté, au sens d'acylation d'ARNt, peut être attribué à chaque anneau d'ARN, sur la base de leur séquence anticodon prédite [46]. Les anneaux d'ARN théoriques correspondent également aux prédictions selon lesquelles les ARN circulaires ont joué un rôle critique dans la formation précoce des protogénomes [47]. Plus important encore, leur conception suppose la préexistence du code génétique à la traduction assistée par ARNt. Il est à noter que

certaines structures secondaires du cycle ARN ressemblent plus aux ARNr qu'aux ARNt [46], et que la maturation de certains ARNt modernes comprend la circularisation [48].

Ribosomal translation of RNA

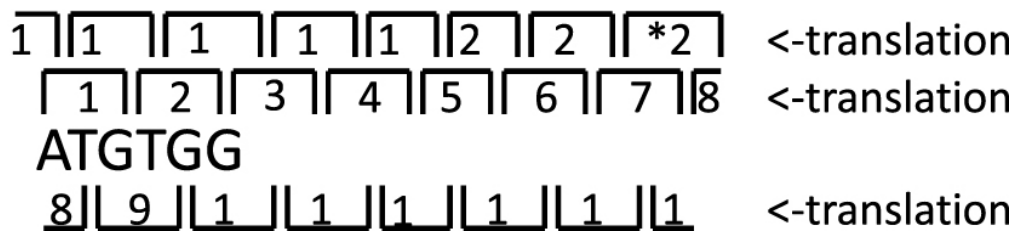


Figure 1. Traduction en trois tours de cycle d'ARN minimal théorique. Les nombres indiquent l'ordre des codons. * indique le codon d'arrêt.

La conception des anneaux d'ARN minimaux théoriques minimise la longueur de séquence et maximise la diversité des acides aminés codés. Les ARNt circularisés sont des intermédiaires dans les modifications post-transcriptionnelles de certains ARNt [40]. Les ARN circulaires régulent l'expression des gènes dans les cellules modernes [51 , 52 , 53 , 54 , 55 , 56 , 57] et sont parfois traduits [58]. Ils codent pour des protéines ayant différentes fonctions, fréquemment dans les muscles des mammifères [59]. L'hypothèse d'un codage par chevauchement est également réaliste: des gènes conçus artificiellement codant sur trois cadres de chevauchement pour trois protéines prédéfinies provenant de domaines protéiques différents sont relativement faciles à produire [60]. Cela correspond aux observations suggérant que le codage par chevauchement est une propriété commune des gènes [61 , 62 , 63 , 64 , 65 , 66 , 67 , 68 , 69]. En effet, le code génétique est optimisé pour le codage après des décalages de cadres [1 , 67 , 68 , 69].

Nous nous attendons à ce que les associations entre les affinités des codons de l'anneau ARN et l'hydrophobicité des acides aminés apparentés dans les peptides codés par les anneaux d'ARN ressemblent à celles observées dans les gènes modernes naturels. Deux scénarios évolutifs existent dans ce contexte. (1) Cette association est apparue spontanément et, dans les gènes modernes, est un report d'un mécanisme de traduction antérieur dépourvu d'ARNt et de ribosome. Dans ce cas, sa force devrait diminuer avec l'ordre d'inclusion du code génétique des acides aminés apparentés des anneaux d'ARN. (2) Cette association résulte d'un processus évolutif et reflète des processus qui se produisent encore dans les cellules modernes. Selon ce scénario, la force de cette association devrait augmenter des anneaux d'ARN précoces à tardifs.

2. Matériels et méthodes

Nous avons traduit les 25 peptides codés par les anneaux d'ARN minimaux théoriques (tableau 1) selon le code génétique standard et estimé la force des associations entre leur affinité codon-acide aminé et l'hydrophobicité de leurs acides aminés apparentés formant le peptide. L'affinité codon-acide aminé est calculée en additionnant l'affinité des nucléotides uniques formant le codon avec l'acide aminé codé par le codon [35]. Ces affinités nucléotidiques uniques pour les acides aminés sont estimées à partir des fréquences de contact nucléotide-acide aminé dans les structures cristallines des complexes ARN-protéine en interaction [35].

RNA AA	CDs	Peptide	R1	R2
1 F	AATTCATGCCAGACTGGTATGA	NSCQTGMKFMFDWYIEIHARLV*	-13.0	-13.7
2 M	CATGCCAGAAATTCTGGTATGA	HARNSGMTQCQKFWYDMPEILV*	-10.7	-11.8
3 S	ATGGTGCCACTATTCAAGATGA	MVPLFKMNGATIQDEWCHYSR*	-20.9	-15.0
4 Pyl	ATGCTATTACCAAGATGGTGA	MLFTKMVNIAHQDGECSYSPRW*	-13.4	-8.3
5 V	ATGGTGCTACCATTCAAGATGA	MVLPFKMNGATIQDEWCYHSR*	-16.8	-9.7
6 N	ATGGCCTATTACAAGATGTGA	MAYSQDVNGLFTRCEWPIHKM*	-19.1	-16.8
7 D	ATGCCACTGGTATTCAAGATGA	MPLVFKMNATGIQDECHWYSR*	-19.8	-14.1
8 R	ATGTGGCCTACATTCAAGATGA	MWPTFKMNVAYIQDECGLHSR*	-22.6	-18.6
9 Sec	ATGCCAAGATGGTATTCACTGA	MPRWYSLNAKMVFTECQDGIH*	-19.8	-14.1
10 L	GCAATGTTTATGGAGACCATAA	AMFMETISNVYGDHKQCLWRP*	-25.8	-19.8
11 P	TATGTTTGGAGACCAAGCATAA	YVWRPSMICLETKHNMGDQA*	-24.4	-18.9
12 E	TTCATGCCAGAACTGGTATGA	FMPETGMIHARNWYDSCQKLV*	-24.7	-19.7
13 G	ATGGTACTGCCATTCAAGATGA	MVLPFKMNGTAIQDEWYCHSR*	-23.6	-16.0
14 I	GCAGAATGTTTATGGACCATAA	AECLWTISRMFMDHKQNVYGP*	-29.0	-24.2
15 Q	AATATGTTTGGACCAAGCATAG	NMFGPSIEYVWTKHRICLDQA*	-35.7	-34.5
16 L	TATGTTTGGAAAGCCAGACATAA	YVWKPDIICLEARHNMFSGT*	-20.8	-22.7
17 K	TACATTTGGAAGCCAGATGTAA	YIWKPDVMHLEARCNTFGSQM*	-27.1	-27.0
18 S	ACAATGTTTATGGAAGCCATAG	TMFMEAIDNVYGSHRQCLWKP*	-32.9	-37.8
19 A	ATGGAAGCCATTACAATGTAG	MEAIYNVDGSHLQCRWKPFTM*	-38.4	-38.3
20 W	TACAGATGGAAGCCATTTGTAA	YRWKPFVMQMEAICNTDGSHL*	-27.1	-27.0
21 C	AACATGCCAGATTCTGGTATGA	NMPDSGMKHARFWYETCQILV*	-24.8	-17.7
22 H	TGCCAGAAACATTCTGGTATGA	CQKHSGMMPETFWYDARNILV*	-21.8	-17.7
23 T	TATGGTCTGCAAGAACCATGA	YGSARTMIWFCKNHDMVLQEP*	-28.2	-17.2
24 Y	TACCATTCTGCAAGAATGGTGA	YHSARMVMPFCKNGDTILQEW*	-21.7	-13.5
25 G	ATGGTGCCATTCAAGACTATGA	MVPFKTMNGAIQDYEWCHSRL*	-19.9	-14.5

Tableau 1. Anneaux d'ARN minimaux et séquences peptidiques correspondantes selon les traductions régulières. Dans la colonne 1, r indique des structures secondaires ressemblant à plus d'ARNr, les restants ressemblent davantage à de l'ARNt, les majuscules indiquent l'acide aminé apparenté présumé du cycle ARN s'il est considéré comme un proto-ARNt selon son anticodon prédit. * indique les codons d'arrêt. Les colonnes 4 et 5 indiquent les coefficients de corrélation de Pearson $r \times 100$ entre l'affinité codon-acide aminé [35] et l'hydrophobicité des acides aminés estimée à partir des équilibres de distribution entre la vapeur et la solution aqueuse (R1, $\log(RH_{vap} / RH_{water})$, R2, potentiel d'hydratation, [73])

Ces sommes linéaires d'affinités de nucléotides uniques avec des acides aminés produisent la même affinité trinuéclotidique pour les trinuéclotides N1N2N3, N1N3N2, N2N1N3, N2N3N1, N3N1N2 et N3N2N1, bien que dans la plupart des cas, ces permutations ne codent pas pour le même acide aminé. Par conséquent, cette somme d'affinités nucléotidiques uniques doit être considérée comme une approximation des affinités trinuéclotide-acide aminé. Néanmoins, les biais pour les contacts codon-acide aminé observés dans les structures cristallines ribosomales [29] sont en corrélation avec ces affinités de codons calculées en additionnant les affinités nucléotidiques uniques (figure 2). Cette observation préliminaire justifie la somme linéaire des affinités nucléotidiques uniques pour approximer les affinités codon-acide aminé.

Position	Log	Pot
Round 1		
1	0.100*	0.101*
2	-0.070*	-0.002*
3	-0.292*	-0.231*
4	-0.290*	-0.265*
5	-0.508*	-0.331*
6	-0.618*	-0.602*
7	-0.697*	-0.697*
Round 2		
8	-0.535*	-0.612*
9	-0.333*	-0.321*
10	-0.074*	-0.043*
11	0.006*	-0.067*
12	-0.607*	-0.544*
13	-0.599*	-0.137*
14	-0.063*	-0.260*
Round 3		
15	-0.775*	-0.459*
16	-0.116*	-0.116*
17	-0.055*	-0.047*
18	-0.375*	-0.333*
19	-0.380*	-0.433*
20	-0.028*	-0.019*
21	-0.210*	-0.199*
22-stop	n.d.	n.d.

Tableau 2. Coefficients de corrélation de Pearson r entre l'affinité des codons et l'hydrophobicité des acides aminés apparentés (comme pour le tableau 1 , [73], Log est pour log (RHvap / RHwater) et Pot pour le potentiel d'hydratation utilisé comme estimations d'hydrophobicité) pour chacun des 21 acides aminés- codant des codons dans les 25 anneaux d'ARN minimaux théoriques du tableau 1 . Les analyses sont effectuées à travers les 25 anneaux d'ARN, en gardant constante la position du codon. Les astérisques indiquent $P < 0,05$. Les valeurs P sont bilatérales

Les coefficients de corrélation de Pearson r ont été calculés entre les affinités des codons et plusieurs estimations des besoins polaires en acides aminés, des besoins polaires expérimentaux et calculés [70 , 71 , 72], deux estimations des proportions de surfaces d'acides aminés accessibles et enfouies dans les protéines [71 , 72], et plusieurs estimations des affinités des chaînes latérales d'acides aminés pour le solvant eau [73 , 74], d'autres solvants [75 , 76] et pour l'eau à différentes températures (à 25°C contre 100°C [78]).

3. Résultats et discussion

3.1. Corrélations d'affinité des codons-hydrophobicité des acides aminés à des codons spécifiques

Notre première analyse considère chaque codon sur les 25 anneaux d'ARN minimaux théoriques et calcule la corrélation entre l'affinité du codon et l'hydrophobicité pour l'acide aminé codé par chaque codon. Cela produit 21 coefficients de corrélation Pearson r pour chaque estimation d'hydrophobicité, pour les 21 positions de codon codant pour un acide aminé, et alignés comme indiqué dans le tableau 2. Ces corrélations devraient refléter celles observées pour les ARNm réguliers et leurs protéines apparentées. Les affinités sont estimées à partir des fréquences de contact nucléotide-résidu et log-transformées pour correspondre aux enthalpies de contact. Si les affinités sont proportionnelles aux hydrophobies des acides aminés observées dans les gènes réguliers, les corrélations avec les estimations des affinités transformées en log devraient produire des corrélations négatives avec les hydrophobies des acides aminés.

Parmi dix-sept échelles d'hydrophobicité différentes testées, le plus grand nombre de positions pour lesquelles des associations négatives entre l'affinité des codons et l'hydrophobicité des acides aminés est de 20 sur 21 positions, pour les échelles basées sur la distribution des équilibres entre la solution aqueuse diluée et la phase vapeur à température ambiante ($\log(RH_{vap} / RH_{water})$, potentiel d'hydratation [74]). Notez que Pro est absent de ces échelles. Par conséquent, le modèle observé dans les ARNm et les protéines réguliers (corrélations négatives) est également observé pour la plupart des positions dans les anneaux d'ARN théoriques selon ces échelles d'hydrophobicité. Cette majorité est statistiquement significative ($P = 0,00001$, test de signe bilatéral) et le reste également après avoir considéré la multiplicité du test, même selon la correction surconservatrice de Bonferroni ($P = 0,05/17 = 0,002941$, ce qui est toujours supérieur à $P = 0,00001$). La tendance globale est plus faible si l'on considère d'autres échelles d'hydrophobicité, notamment celles obtenues selon des équilibres similaires entre solvants, mais autre que l'eau, solvant naturel des systèmes biologiques. Ceux-ci ne sont pas rapportés dans ce qui suit car les résultats étaient systématiquement les plus forts pour les échelles présentées dans [74].

Cette analyse produit également des corrélations statistiquement significatives à des codons spécifiques dans les anneaux d'ARN. Les six corrélations statistiquement significatives sont négatives comme prévu des ARNm et protéines modernes et sont situées à la transition entre chaque cycle de traduction, pour les acides aminés codés par les 6e, 7e, 8e, 12e, 15e et 19e codons dans les anneaux d'ARN. Cela suggérerait que les peptides interagissent avec leur anneau ARN matrice à chaque nouveau cycle de traduction. Ce modèle émergent des anneaux d'ARN théoriques ne fait pas partie de leur conception et pourrait refléter les propriétés de la traduction pré-ARNt à base d'ARN.

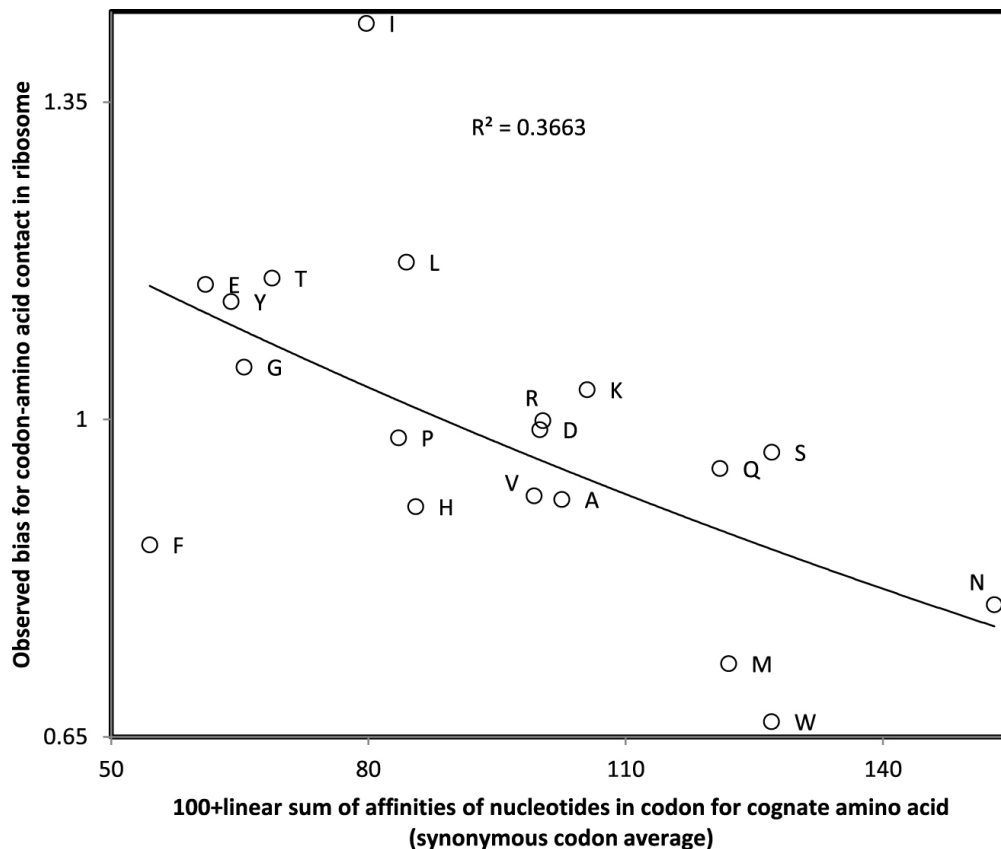


Figure 2. Biais observé pour les contacts entre le codon et l'acide aminé apparenté dans la structure cristalline du ribosome (Johnson et Lee 2010) en fonction de la somme linéaire des affinités nucléotidiques uniques pour l'acide aminé apparenté en moyenne sur tous les codons synonymes (les affinités fortes correspondent à de faibles valeurs sur l'axe x). La corrélation ($r = -0,605$, P unilatéral = 0,003) justifie l'utilisation des sommes des affinités nucléotidiques uniques pour approximer les affinités des codons.

Les corrélations obtenues pour les échelles d'hydrophobicité estimées à partir des équilibres de distribution entre la solution neutre et le cyclohexane à 25°C et 100°C [77] produisent des schémas moins clairs. Les modèles sont légèrement plus forts à 100°C qu'à 25°C, ce qui suggère que les associations codon-acide aminé que nous observons dans les organismes modernes ont évolué à l'origine à des températures élevées,

conformément à certaines considérations antérieures sur les origines de la vie et le code génétique [78 , 79 , 80 , 81 , 82].

3.2. Corrélations d'affinité des codons-hydrophobicité des acides aminés pour des anneaux d'ARN spécifiques

Notre deuxième analyse teste les corrélations entre l'affinité des codons et l'hydrophobicité des acides aminés (équilibres de distribution entre la solution aqueuse diluée et la phase vapeur à température ambiante [74]) à travers les codons, pour chaque anneau ARN spécifique (Tableau 1). Les 25 corrélations sont négatives ($P < 0,05$ pour l'anneau ARN 19), une majorité statistiquement significative de cas ($P = 0,00000003$, test de signe bilatéral). Cela signifie que l'association entre l'affinité des codons et l'hydrophobicité des acides aminés pour les 25 anneaux d'ARN ressemble à celle observée dans les gènes modernes.

3.3. Évolution des associations entre l'affinité des codons et l'hydrophobicité des acides aminés

Chacun des 25 anneaux d'ARN minimaux théoriques, en raison de leur similitude avec les séquences d'ARNt consensus, peut être attribué à un acide aminé apparenté en fonction de leur séquence anticodon prédite, en supposant une fonction de type ARNt. Chacun de ces acides aminés apparentés à l'ARNt se voit attribuer un rang d'inclusion dans le code génétique, selon diverses hypothèses sur l'histoire d'inclusion de code génétique. Fait intéressant, et malgré que ces nombreuses hypothèses aient des origines très différentes, de mathématiques (par exemple le code circulaire naturel, [14], le modèle algébrique [84]), physicochimique (par exemple les rendements en acides aminés dans des expériences mimant l'état de la terre primitive [85], complexité de la taille des acides aminés [86]) à physiologiques (par exemple l'hypothèse de coévolution métabolique [87], le modèle autoréférentiel [88]), ils tendent à prédire les histoires globales congruentes des attributions de code génétique d'acides aminés (40 ordres d'inclusion hypothétiques examinés dans [26 , 27]). Certaines hypothèses sont dérivées des règles d'aminoacylation des petites hélices d'ARN [89 , 90] et des ARNt [91], affectant le repliement des protéines [9 , 92], permettant de définir un code primordial dans les tiges accepteurs d'ARNt et un ordre associé d'intégration des acides aminés dans le code génétique.

Les rangs d'inclusion des acides aminés apparentés prédits du cycle ARN sont supposés refléter l'histoire du système organique théorique basé sur ces 25 anneaux ARN théoriques et leurs peptides apparentés (le terme apparenté ici est utilisé dans le sens du peptide traduit du cycle ARN comme un ancien modèle pour la synthèse des protéines). Nous avons examiné les associations des coefficients de corrélation de Pearson entre l'affinité des codons et l'hydrophobicité des acides aminés pour les 25 anneaux d'ARN et les rangs d'inclusion de leur acide aminé apparenté (le cognate est utilisé ici dans le sens de l'acide aminé vraisemblablement aminoacylé à l'anneau d'ARN en relation à sa fonction de type ARNt présumée). Les anneaux d'ARN avec les anticodons prédits correspondant aux acides aminés apparentés sélénocystéine et pyrrolysine ont reçu l'avant-dernier et le dernier rang d'inclusion de code génétique.

Le scénario 1 suppose que les corrélations entre l'affinité des codons et l'hydrophobicité des acides aminés reflètent des associations spontanées qui étaient la base physico-chimique de la traduction préexistante de la traduction moderne sans ARNt et sans ribosome. Le long de ce scénario, les associations d'affinité des codons et d'hydrophobicité des acides aminés observées aujourd'hui seraient des vestiges de ce mécanisme de traduction antérieur. Par conséquent, les coefficients de corrélation de Pearson r entre l'affinité des codons et l'hydrophobicité des acides aminés des anneaux d'ARN devraient augmenter (devenir plus positifs) avec le rang d'inclusion de l'acide aminé attribué comme proto-ARNt apparenté pour cet anneau d'ARN.

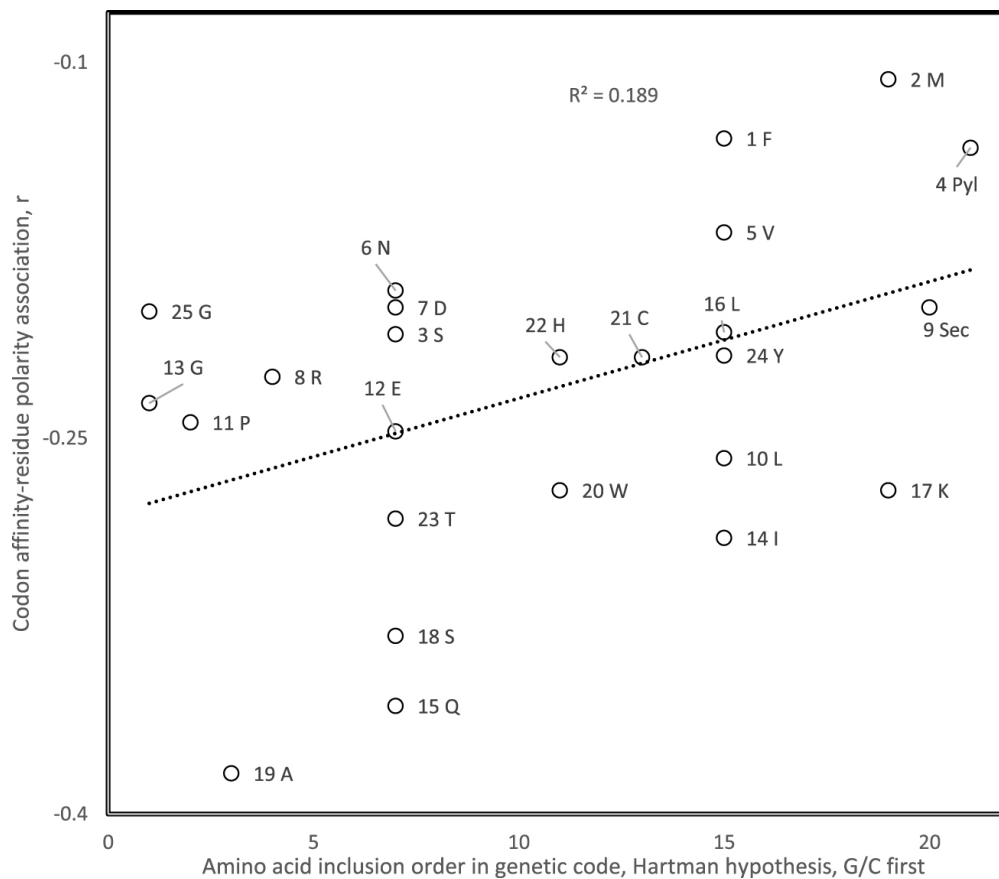


Figure 3. Coefficient de corrélation de Pearson d'association entre l'exigence polaire de l'acide aminé et l'affinité du codon pour cet acide aminé pour 25 anneaux d'ARN minimaux théoriques de 22 nucléotides codant après trois cycles de traduction pour un codon de départ et d'arrêt et un codon par acide aminé, en fonction de la génétique ordre d'inclusion du code de l'acide aminé apparenté présumé de cet anneau d'ARN s'il est considéré comme un proto-ARNt. L'acide aminé apparenté à l'ARNt est déterminé par le triplet de nucléotides à la position anticodon prédite. L'ordre d'inclusion du code génétique est conforme au code d'un nucléotide primordial de Hartman, en supposant que G et C étaient les premiers [93 , 94 , 95 , 96 , 97].

Le scénario 2 s'attend à ce que les corrélations aient évolué pour devenir plus similaires à ce qui est connu pour les gènes et protéines existants, ce qui est généralement une corrélation négative. Par conséquent, les coefficients de corrélation de Pearson r entre l'affinité des codons et l'hydrophobicité des acides aminés des anneaux d'ARN devraient diminuer (devenir plus négatifs) avec le rang d'inclusion de l'acide aminé attribué comme proto-ARNt apparenté pour cet anneau d'ARN. Les coefficients de corrélation r obtenus pour les associations entre les affinités des codons et l'hydrophobicité estimées à partir des équilibres de distribution entre la vapeur et la solution aqueuse correspondent au scénario 1 pour 38 des 40 hypothèses d'évolution du code génétique des ordres d'inclusion d'acides aminés, une majorité significative ($P = 8 \times 10^{-8}$, test du signe à deux queues). La corrélation avec l'ordre d'inclusion des acides aminés est statistiquement significative à $P < 0,05$ pour huit hypothèses spécifiques pour les ordres d'inclusion d'acides aminés. L'association positive la plus forte entre l'ordre d'inclusion présumé de l'acide aminé apparenté du cycle ARN et la corrélation affinité-hydrophobicité des codons rest pour l'hypothèse de Hartman qui suppose que le code génétique a commencé comme un code à une lettre, puis est devenu un code à deux puis à trois lettres, avec G et C comme nucléotides primordiaux ($r = -0,43$, unilatéral $P = 0,008$, figure 3) [93 , 94 , 95 , 96]. Par conséquent les résultats sont en accord avec le scénario 1 qui suppose que les associations codon-amino hydrophobicité acide dans des paires de protéines de gènes modernes sont des restes d'un ancien tRNA- et le mécanisme de translation libre de ribosome sur la base de ces interactions physico - chimiques [98], qui de facto aurait produit les attributions de codon-acides aminés de code génétique [99].

4. Conclusions

Les résultats des analyses de peptides codés par des anneaux d'ARN et des codons d'ARN indiquent que les associations codon-acide aminé affinité-acide aminé hydrophobicité étaient les plus fortes / les plus négatives

pour les anneaux d'ARN supposés les plus anciens. En conséquence, de telles associations observées dans les gènes modernes et leurs protéines seraient des vestiges d'un mécanisme de traduction plus ancien dépourvu d'ARNt et de ribosome, basé sur des interactions physico-chimiques directes entre le codon et l'acide aminé. Les attributions codon-acide aminé dans le code génétique auraient résulté de ces interactions directes codon-acide aminé [99 , 100], ce qui expliquerait également les biais pour les contacts codon-acide aminé dans le ribosome ([29], et ici la figure 2) .

On ne sait pas comment la conception des anneaux d'ARN minimaux théoriques implique des propriétés évolutives complexes telles que des associations entre l'affinité des codons et l'hydrophobicité des acides aminés apparentés. Cependant, ce système imite en fait des propriétés étonnamment nombreuses et complexes du code génétique [99 , 101], des ARNt [43], des gènes codant pour les protéines [102 , 103], de leurs protéines [104] et des origines de répliation [105]. Ce système de cycle ARN permet de récupérer des scénarios d'évolution probables d'ARNt et d'ARNr [104 , 106 , 107 , 108 , 109 , 110 , 111], et l'attribution de AUG comme codon d'initiation universel [112]. La conception des anneaux d'ARN implique que le code génétique préexistait à l'émergence des anneaux d'ARN. Par conséquent, parce que les anneaux d'ARN ressemblent aux proto-ARNt, le code génétique devait exister avant les ARNt et les proto-ARNt. L'association de l'affinité des codons avec l'hydrophobicité pour les anneaux d'ARN indique que la traduction selon le code génétique préexistait à la traduction médiée par l'ARNt, potentiellement le long d'une structure de codon différente où les nucléotides aux positions 1 et 2 des codons étaient commutés [113 , 114], mais aussi évolué vers la situation observée dans les gènes modernes lors de l'évolution de ce système d'ARN primordial basé sur un anneau d'ARN [49 , 50 , 83]. Les anneaux d'ARN minimaux théoriques sont des modèles réalistes pour le monde de l'ARN lors de la transition de la traduction basée sur l'affinité des codons pour les acides aminés et manquant d'adaptateurs de type ARNt à la traduction médiée de type ARNt.

Remerciements

Merci à Charles W. Jr Carter pour ses conseils importants sur les estimations d'hydrophobicité.

Références

1. B. Kumar; S. Saini Analysis of the optimality of the standard genetic code, *Mol. Biosys.*, Volume 12 (2016), pp. 2642-2651
2. C. R. Woese; D. H. Dugre; S. A. Dugre; M. Kondo; W. C. Saxinger On the fundamental nature and evolution of the genetic code, *Cold Spring Harb. Symp. Quant. Biol.*, Volume 31 (1966), pp. 723-736
3. S. R. Pelc; M. G. E. Welton Stereochemical relationship between coding triplets and amino-acids, *Nature*, Volume 209 (1966), pp. 868-870
4. H. Seligmann; G. N. Amzallag Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code, *Die Naturwissenschaften*, Volume 89 (2002), pp. 542-551
5. D. H. Ardell On error minimization in a sequential origin of the standard genetic code, *J. Mol. Evol.*, Volume 47 (1998), pp. 1-13
6. S. J. Freeland; L. D. Hurst The genetic code is one in a million, *J. Mol. Evol.*, Volume 47 (1998), pp. 238-248
7. A. Guilloux; J. L. Jestin The genetic code and its optimization for kinetic energy conservation in polypeptide chains, *Biosystems*, Volume 109 (2012), pp. 141-144 | arXiv:2020.03.31.016048
8. B. Caudron; J. L. Jestin Sequence criteria for the anti-parallel character of protein beta-strands, *J. Theor. Biol.*, Volume 315 (2012), pp. 146-149
9. H. Seligmann; G. Warthi Genetic code optimization for cotranslational protein folding: codon directional asymmetry correlates with antiparallel betasheets, tRNA synthetase classes, *Comput. Struct. Biotechnol. J.*, Volume 15 (2017), pp. 414-424
10. H. Seligmann Do anticodons of misacylated tRNAs preferentially mismatch codons coding for the misloaded amino acid?, *BMC Mol. Biol.*, Volume 11 (2010), 13473, 41 pages
11. H. Seligmann Error compensation of tRNA misacylation by codon-anticodon mismatch prevents translational amino acid misinsertion, *Comput. Biol. Chem.*, Volume 35 (2011), pp. 81-95 | Zbl 1403.92081
12. H. Seligmann Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes, *Curr. Genomics*, Volume 13 (2012), pp. 37-54

13. R. M. Barthélémy; H. Seligmann Cryptic tRNAs in chaetognath mitochondrial genomes, *Comput. Biol. Chem.*, Volume 95 (2016), pp. 119-132
14. D. G. Arquès; C. J. Michel A complementary circular code in the protein coding genes, *J. Theor. Biol.*, Volume 182 (1996), pp. 45-58 | arXiv:2020.02.22.20026500
15. A. Ahmed; G. Frey; C. J. Michel Frameshift signals in genes associated with the circular code, *Silico Biol.*, Volume 7 (2007), pp. 155-168
16. N. El Houmami; H. Seligmann Evolution of nucleotide punctuation marks: from structural to linear signals, *Front. Genet.*, Volume 8 (2017), 36 pages
17. G. Warthi; H. Seligmann Transcripts with systematic nucleotide deletion of 1–12 nucleotide in human mitochondrion suggest potential non-canonical transcription, *PLoS ONE*, Volume 14 (2019) (e0217356)
18. H. Seligmann; D. D. Pollock The ambush hypothesis: hidden stop codons prevent off-frame gene reading, *DNA Cell. Biol.*, Volume 23 (2003), pp. 701-705
19. H. Seligmann Cost minimization of ribosomal frameshifts, *J. Theor. Biol.*, Volume 249 (2007), 1007188, pp. 162-167
20. H. Seligmann The ambush hypothesis at the whole-organism level: off frame, ‘hidden’ stops in vertebrate mitochondrial genes increase developmental stability, *Comput. Biol. Chem.*, Volume 34 (2010), 12818, pp. 80-85 | Zbl 1366.92050
21. H. Seligmann Localized context-dependent effects of the “ambush” hypothesis: more off-frame stop codons downstream of shifty codons, *DNA Cell. Biol.*, Volume 38 (2019), pp. 786-795
22. X. Wang; Q. Dong; G. Chen; J. Zhang; Y. Liu; J. Zhao; H. Peng; Y. Wang; Y. Cai; X. Wang; C. Yang; M. Lynch (2016 “The universal genetic code, protein coding genes and genomes of all species were optimized for frameshift tolerance”, bioRxiv, posted Aug 25, 2016, <http://dx.doi.org/10.1101/067736>)
23. R. Geyer; M. A. Madany On the efficiency of the genetic code after frameshift mutations, *PeerJ*, Volume 6 (2018) (e4825)
24. G. K. Philip; S. J. Freeland Did evolution select a nonrandom “alphabet” of amino acids?, *Astrobiology*, Volume 11 (2011), pp. 235-240
25. M. Ilardo; M. Meringer; S. Freeland; B. Rasulev; H. J. Cleaves II Extraordinarily adaptive properties of the genetically encoded amino acids, *Sci. Rep.*, Volume 5 (2015), 9414 pages
26. E. N. Trifonov Consensus temporal order of amino acids and evolution of the triplet code, *Gene*, Volume 261 (2000), pp. 139-151
27. E. N. Trifonov The triplet code from first principles, *J. Biomol. Struct. Dyn.*, Volume 22 (2004), pp. 1-11
28. H. Seligmann Protein sequences recapitulate genetic code evolution, *Comput. Struct. Biotech. J.*, Volume 16 (2018), pp. 177-199
29. D. B. F. Johnson; L. Wang Imprints of the genetic code in the ribosome, *Proc. Natl. Acad. Sci. USA*, Volume 107 (2010), pp. 8298-8303
30. R. Root-Bernstein Simultaneous origin of homochirality, the genetic code and its directionality, *Bioessays*, Volume 29 (2007), pp. 689-698
31. X. Han Da; H. Y. Wang; Z. L. Ji; A. F. Hu; Y. F. Zhao Amino acid homochirality may be linked to the origin of phosphate-based life, *J. Mol. Evol.*, Volume 70 (2010), pp. 572-582
32. D. G. Blackmond The origin of biological homochirality, *Phil. Trans. R. Soc. Lond. B*, Volume 366 (2011), pp. 2878-2884
33. R. Breslow The origin of homochirality in amino acids and sugars on prebiotic earth, *Tetrahedron Lett.*, Volume 52 (2011), pp. 4228-4232
34. C. J. Michel; H. Seligmann Bijective transformation circular codes and nucleotide exchanging RNA transcription, *Biosystems*, Volume 118 (2014), pp. 39-50
35. A. A. Polyansky; B. Zagrovic Evidence of direct complementarity interactions between messenger RNAs and their cognate proteins, *Nucl. Acids Res.*, Volume 41 (2013), pp. 8434-8443
36. L. Bartonek; B. Zagrovic mRNA/protein sequence complementarity and its determinants: the impact of affinity scales, *PLoS Comput. Biol.*, Volume 13 (2017) (e1005648)
37. B. Zagrovic; L. Bartonek; A. A. Polyansky RNA-protein interactions in an unstructured context, *FEBS Lett.*, Volume 592 (2018), pp. 2901-2916
38. N. Maizels; A. M. Weiner Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation, *Proc. Natl. Acad. Sci. USA*, Volume 91 (1994), pp. 6729-6734
39. H. Seligmann Swinger RNA self-hybridization and mitochondrial non-canonical swinger transcription, transcription systematically exchanging nucleotides, *J. Theor. Biol.*, Volume 399 (2016), pp. 84-91

40. P. R. Wills; C. W. Carter Jr Insuperable problems of the genetic code initially emerging in an RNA world, *Biosystems*, Volume 165 (2018), pp. 155-166
41. J. Demongeot Sur la possibilité de considérer le code génétique comme un code à enchaînement, *Rev. Biomaths*, Volume 62 (1978), pp. 61-66
42. J. Demongeot; J. Besson Genetic-code and cyclic codes, *C. R. Acad. Sci. III*, Volume 296 (1983), pp. 807-810 | arXiv:2020.03.09.20033118
43. J. Demongeot; J. Besson Genetic code and cyclic codes II, *C. R. Acad. Sc. III*, Volume 319 (1996), pp. 520-528
44. J. Demongeot; A. Elena; G. Weil Potential-Hamiltonian decomposition of cellular automata. Application to degeneracy of genetic code and cyclic codes III, *C. R. Biol.*, Volume 329 (2006), pp. 953-962
45. J. Demongeot; A. Moreira A possible circular RNA at the origin of life, *J. Theor. Biol.*, Volume 249 (2007), pp. 314-324
46. H. Seligmann; D. Raoult Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses, *Curr. Opin. Microbiol.*, Volume 31 (2016), pp. 1-8
47. H. Seligmann; D. Raoult Stem-loop RNA hairpins in giant viruses: invading rRNA-like repeats and a template free RNA, *Front. Microbiol.*, Volume 9 (2018), 101 pages
48. G. Soslau Circular RNA (circRNA) was an important bridge in the switch from the RNA world to the DNA world, *J. Theor. Biol.*, Volume 447 (2018), pp. 32-40
49. J. Salzman; C. Gawad; P. L. Wang; N. Lacayo; P. O. Brown Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types, *PLoS ONE*, Volume 7 (2012) (e30733)
50. E. Lasda; R. Parker Circular RNAs: diversity of form and function, *RNA*, Volume 20 (2014), pp. 1829-1842
51. S. P. Barrett; J. Salzman Circular RNAs: analysis, expression and potential functions, *Development*, Volume 143 (2016), pp. 1838-1847
52. Y. Zhang; W. Xue; X. Li; J. Zhang; S. Chen; J. L. Zhang; L. Yang; L. L. Chen The biogenesis of nascent circular RNAs, *Cell Rep.*, Volume 15 (2016), pp. 611-624
53. S. Huang; B. Yang; B. J. Chen; N. Bliim; U. Ueberham; T. Arendt; M. Janitz The emerging role of circular RNAs in transcriptome regulation, *Genomics*, Volume 109 (2017), 104787, pp. 401-407
54. Y. Zhong; Y. Du; X. Yang; Y. Mo; C. Fan; F. Xiong; D. Ren; X. Ye; C. Li; Y. Wang; F. Wei; C. Guo; X. Wu; X. Li; Y. Li; G. Li; Z. Zeng; W. Xiong Circular RNAs function as ceRNAs to regulate and control human cancer progression, *Mol. Cancer*, Volume 17 (2018), 79 pages
55. B. P. Nicolet; S. Engels; F. Agliandolo; E. van den Akker; M. von Lindern; M. C. Wolkers Circular RNA expression in human hematopoietic cells is widespread and cell-type specific, *Nucl. Acids Res.*, Volume 46 (2018), pp. 8168-8180
56. N. R. Pamudurti; O. Bartok; M. Jens; R. Ashwal-Fluss; C. Stottmeister; L. Ruhe; M. Hanan; E. Wyler; D. Perez-Hernandez; E. Ramberger; S. Shenzis; M. Samson; G. Dittmar; M. Landthaler; M. Chekulaeva; N. Rajewsky; S. Kadener Translation of circRNA, *Mol. Cell*, Volume 66 (2017), pp. 9-21
57. I. Legnini; G. Di Timoteo; F. Rossi; M. Morlando; F. Briganti; O. Sthandler; T. Santini; A. Andronache; M. Wade; P. Laneve; N. Rajewsky; I. Bozzoni Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis, *Mol. Cell*, Volume 66 (2017), pp. 22-37
58. V. Opuu; M. Silvert; T. Simonson Computational design of fully overlapping coding schemes for protein pairs and triplets, *Sci. Rep.*, Volume 7 (2017), 15873 pages
59. E. Faure; L. Delaye; S. Tribolo; A. Levasseur; H. Seligmann; R. M. Barthélémy Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene, *Biol. Direct*, Volume 6 (2011), 56 pages
60. H. Seligmann Two genetic codes, one genome: frameshifted primate mitochondrial genes code for additional proteins in presence of antisense antitermination tRNAs, *Biosystems*, Volume 105 (2011), pp. 271-285
61. H. Seligmann An overlapping genetic code for frameshifted overlapping genes in *Drosophila* mitochondria: antisense antitermination tRNAs UAR insert serine, *J. Theor. Biol.*, Volume 298 (2012), pp. 51-76
62. H. Seligmann Overlapping genetic code for overlapping frameshifted genes in *Testudines*, and *Lepidochelys olivacea* as special case, *Comput. Biol. Chem.*, Volume 41 (2012), pp. 18-34 | Zbl 1365.92038

63. H. Seligmann Directed mutations recode mitochondrial genes: from regular to stopless genetic codes, *Mitochondrial DNA-New Insights* (H. Seligmann; G. Warthi, eds.), IntechOpen, 2018
64. S. Itzkovitz; U. Alon The genetic code is nearly optimal for allowing information within protein-coding sequences, *Genome Res.*, Volume 17 (2007), pp. 405-412
65. X. Wang; X. Wang; G. Chen; J. Zhang; Y. Liu; C. Yang The shiftability of protein coding genes: the genetic code was optimized for frameshift tolerating, *PeerJ Preprints*, Volume 3 (2015) (e806v1)
66. R. Geyer; M. A. Madany On the efficiency of the genetic code after frameshift mutations, *PeerJ*, Volume 6 (2018) (e4825)
67. C. R. Woese Evolution of the genetic code, *Die Naturwissenschaften*, Volume 60 (1973), pp. 447-459
68. R. Wolfenden; L. Andersson; P. M. Cullis; C. C. B. Southgate Affinities of amino acid side chains for solvent water, *Biochemistry*, Volume 20 (1981), pp. 849-855
69. D. C. Mathew; Z. Luthey-Schulten On the physical basis of the amino acid polar requirement, *J. Mol. Evol.*, Volume 66 (2008), pp. 519-528
70. C. Chothia The nature of accessible and buried surfaces in proteins, *J. Mol. Biol.*, Volume 105 (1976), pp. 1-12
71. J. Janin Surface and inside volumes in globular proteins, *Nature*, Volume 277 (1979), p. 491-492
72. R. Wolfenden; P. M. Cullis; C. C. B. Southgate Water, protein folding, and the genetic code, *Science*, Volume 206 (1979), pp. 575-577
73. A. Radzicka; R. Wolfenden Comparing polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution, *Biochemistry*, Volume 27 (1988), pp. 1664-1670
74. R. Wolfenden Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins, *J. Gen. Physiol.*, Volume 129 (2007), pp. 357-362
75. M. Di Giulio The late stage of genetic code structuring took place at a high temperature, *Gene*, Volume 261 (2000), pp. 189-195
76. R. Wolfenden; C. A. Lewis; Y. Yuan; C. W. Carter Jr Temperature dependence of amino acid hydrophobicities, *Proc. Natl. Acad. Sci. USA*, Volume 112 (2015), pp. 7484-7488
77. M. Di Giulio The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence, *J. Theor. Biol.*, Volume 221 (2003), pp. 425-436 | Zbl 07196981
78. M. Di Giulio The universal ancestor and the ancestor of bacteria were hyperthermophiles, *J. Mol. Evol.*, Volume 57 (2003), pp. 721-730
79. S. Akanuma; Y. Nakajima; S. Yokobori; M. Kimura; N. Nemoto; T. Mase; K. Miyazono; M. Tanokura; A. Yamagishi Experimental evidence for the thermophilicity of ancestral life, *Proc. Natl. Acad. Sci. USA*, Volume 110 (2013), pp. 11067-11072
80. N. H. Sleep Geological and geochemical constraints on the origin and evolution of life, *Astrobiology*, Volume 18 (2018), pp. 1199-1219
81. J. E. M. Hornos; Y. M. M. Hornos Algebraic model for the evolution of the genetic code, *Phys. Rev. Lett.*, Volume 71 (1993), pp. 4401-4404
82. S. L. Miller Production of amino acids under possible primitive earth conditions, *Science*, Volume 117 (1953), p. 528-529
83. M. J. Dufton Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins?, *J. Theor. Biol.*, Volume 187 (1997), pp. 165-173
84. J. T. F. Wong A co-evolution theory of the genetic code, *Proc. Natl. Acad. Sci. USA*, Volume 72 (1975), pp. 1909-1912
85. R. C. Guimarães Self-referential encoding on modules of anticodon pairs - roots of the biological flow system, *Life (Basel)*, Volume 7 (2017), 16 pages
86. C. Francklyn; K. Musier-Forsyth; P. Schimmel Small RNA helices as substrates for aminoacylation and their relationship to charging of transfer RNAs, *Euro. J. Biochem.*, Volume 206 (1992), pp. 315-321
87. C. W. Carter Jr; R. Wolfenden tRNA acceptor-stem and anticodon bases form independent codes related to protein folding, *Proc. Natl. Acad. Sci. USA*, Volume 112 (2015), pp. 7489-7494
88. C. W. Carter Jr; P. R. Wills Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases, *Nucl. Acids Res.*, Volume 46 (2018), pp. 9667-9683
89. C. W. Carter Jr; R. Wolfenden Acceptor-stem and anticodon bases embed amino acid chemistry into tRNA, *RNA Biol.*, Volume 13 (2016), pp. 145-151
90. H. Hartman Speculations on the evolution of the genetic code, *Orig. Life Biosph.*, Volume 6 (1975), pp. 423-427

91. H. Hartman Speculations on the origin and evolution of metabolism, *J. Mol. Evol.*, Volume 4 (1975), pp. 359-370
92. H. Hartman Speculations on the evolution of the genetic code II, *Orig. Life Biosph.*, Volume 9 (1978), pp. 133-136
93. H. Hartman Speculations on the origin of the genetic code, *J. Mol. Evol.*, Volume 40 (1995), pp. 541-544
94. E. N. Trifonov Elucidating sequence codes: three codes for evolution, *Ann. N.Y. Acad. Sci.*, Volume 870 (1999), pp. 330-338
95. M. Yarus The genetic code and RNA-amino acid affinities, *Life (Basel)*, Volume 7 (2017), 13 pages
96. J. Demongeot; H. Seligmann Spontaneous evolution of circular codes in theoretical minimal RNA rings, *Gene*, Volume 705 (2019), pp. 95-102
97. H. Seligmann First arrived, first served: competition between codons for codon-amino acid stereochemical interactions determined early genetic code assignments, *Naturwissenschaften*, Volume 107 (2020), 20 pages
98. J. Demongeot; H. Seligmann Pentamers with non-redundant frames: bias for natural circular code codons, *J. Mol. Evol.*, Volume 88 (2020), pp. 194-201
99. J. Demongeot; H. Seligmann More pieces of ancient than recent theoretical minimal proto-tRNA-like RNA rings in genes coding for tRNA synthetases, *J. Mol. Evol.*, Volume 87 (2019), pp. 1-23
100. J. Demongeot; H. Seligmann Bias for 3'-dominant codon directional asymmetry in theoretical minimal RNA rings, *J. Comput. Biol.* (2019) (doi:10.1089/cmb.2018.0256)
101. J. Demongeot; H. Seligmann Theoretical minimal RNA rings recapitulate the order of the genetic code's codon-amino acid assignments, *J. Theor. Biol.*, Volume 471 (2019), pp. 108-116 | Zbl 07057653
102. J. Demongeot; H. Seligmann Theoretical minimal RNA rings designed according to coding constraints mimic deamination gradients, *Naturwissenschaften*, Volume 106 (2019), 44 pages
103. J. Demongeot; H. Seligmann The uroboros theory of life's origin: 22-nucleotide theoretical minimal RNA rings reflect evolution of genetic code and tRNA-rRNA translation machineries, *Acta Biotheor.*, Volume 67 (2019), pp. 273-297
104. J. Demongeot; H. Seligmann Evolution of tRNA into rRNA secondary structures, *Gene Rep.* (2019) (100483)
105. J. Demongeot; H. Seligmann Accretion history of large ribosomal subunits deduced from theoretical minimal RNA rings is congruent with histories derived from phylogenetic and structural methods, *Gene*, Volume 738 (2020) (144436)
106. J. Demongeot; H. Seligmann Comparisons between small ribosomal RNA and theoretical minimal RNA ring secondary structures confirm phylogenetic and structural accretion histories, *Sci. Rep.*, Volume 10 (2020), 7693 pages
107. J. Demongeot; H. Seligmann RNA rings strengthen hairpin accretion hypotheses for tRNA evolution: a reply to commentaries by Z. F. Burton and M. Di Giulio, *J. Mol. Evol.*, Volume 88 (2020), pp. 243-252
108. J. Demongeot; H. Seligmann The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters, *BMC Genet.*, Volume 21 (2020), 7 pages
109. J. Demongeot; H. Seligmann Why is AUG the start codon?: theoretical minimal RNA rings: maximizing coded information biases 1st codon for the universal initiation codon AUG, *Bioessays* (2020) (e1900201)
110. H. Seligmann; J. Demongeot Codon directional asymmetry suggests swapped prebiotic 1st and 2nd codon positions, *Int. J. Mol. Sci.*, Volume 21 (2020) (e347)
111. J. Demongeot; H. Seligmann Deamination gradients within codons after 1 ? 2 position swap predict amino acid hydrophobicity and parallel ?-sheet conformational preference, *Biosystems*, Volume 191-192 (2020) (104116)
112. A. Soma; A. Onodera; J. Sugahara; A. Kanai; N. Yachie; M. Tomita; F. Kawamura; Y. Sekine Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*, *Science*, Volume 318 (2007), pp. 450-453
113. T. Pan; R. R. Gutell; O. C. Uhlenbeck Folding of circularly permuted transfer RNAs, *Science*, Volume 254 (1991), pp. 1361-1364
114. A. Gutfraind; A. Kempf Error-reducing structure of the genetic code indicates code origin in non-thermophile organisms, *Orig. Life Evol. Biosph.*, Volume 38 (2008), pp. 75-85