

**Sur l'origine du code génétique : un hypothétique précurseur de 27-codons  
d'un intermédiaire intriqué de 64 codons a façonné le code moderne.**

**Bernard Dujon \***

Institut Pasteur, Dept. Genomes and Genetics, CNRS (UMR3525) et Sorbonne Université (UFR927),  
Paris, France

**Matériel supplémentaire : 2 tables et 3 figures**

**Tableau supplémentaire S1 : Analyse de la table des codons dans les formes standard et variantes utilisées dans ce travail.**

Nucleotide à la position 1	A	C	G	U
Familles de codons non scindées	ACN > Thr	CCN > Pro CGN > Arg CUN > Leu	GCN > Ala GGN > Gly GUN > Val	UCN > Ser
Familles scindées de codons	AGN * > Ser + Arg (Gly) AUN > Ile + Met AAN > Asn + Lys	CAN > His + Gln	GAN > Asp + Glu	UGN > Cys + Trp UUN > Phe + Leu UAN * > Tyr + (Gln)

Nucleotide à la position 2	A	C	G	U
Familles de codons non scindées		GCN > Ala CCN > Pro ACN > Thr UCN > Ser	GGN > Gly CGN > Arg	GUN > Val CUN > Leu
Familles scindées de codons	GAN > Asp + Glu CAN > His + Gln AAN > Asn + Lys UAN * > Tyr + (Gln)		AGN * > Ser + Arg (Gly) UGN > Cys + Trp	AUN > Ile + Met UUN > Phe + Leu

Le tableau regroupe en colonnes les familles de codons du code moderne selon les nucléotides en première ou deuxième position des codons, avec indication de leur signification sous les formes standard ou variantes (parenthèses) du code. \* Gln remplace stop pour les codons UAR de la famille UAN ; Gly remplace Arg pour les codons AGR de la famille AGN. Code de couleur des a.a.s. comme dans le tableau 1.

Notez que tous les codons avec C à la deuxième position appartiennent à des familles non scindées (boîtes entièrement dégénérées de 4 codons) alors que tous les codons avec A à la même position appartiennent à des familles scindées (boîtes partiellement dégénérées de 2, exceptionnellement 3, codons). Les codons avec G ou U à la deuxième position sont répartis de manière égale entre les familles scindées et non scindés. Il n'existe pas de biais similaire pour la première position des codons. En raison de l'interaction des anticodons avec la poignée ribosomique (*ribosomal grip*), les codons des familles non scindées ont toujours soit une paire C-G en position centrale des duplex codon-anticodon, soit une paire C-G / G-C en première position, alors que les codons des familles scindées ont toujours soit une paire A-U en position centrale, soit une paire U-A / A-U en position 1 [39].

Notez que tous les codons avec C en deuxième position codent des acides aminés chargés par des aaRS de classe II [43, 44] alors que presque tous les codons avec U à cette position codent des acides aminés chargés par des aaRS de classe I. Les codons avec U en deuxième position codent des acides aminés hydrophobes, ceux avec C codent des petits acides aminés neutres et ceux avec A codent des acides aminés hydrophiles et des acides aminés neutres plus grands [42].

Enfin, notez que la deuxième position du codon est également la plus sensible aux erreurs potentielles de mutation [70].

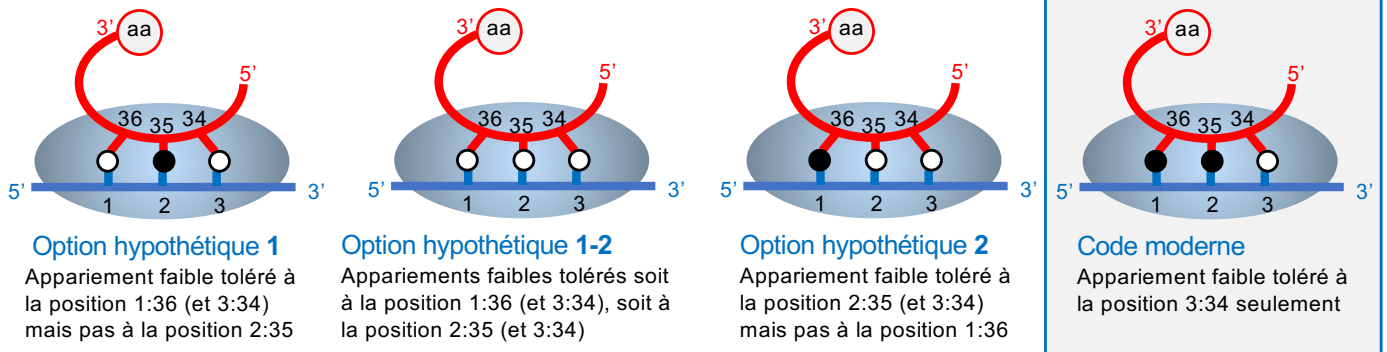
**Tableau supplémentaire S2 : Intrication prédite des codons selon l'option d'appariement 1 et ses traces dans le code moderne.**

Type d'anticodon partagé	Familles intriquées		Signification		Coïncidence observée	Remarques
	Famille n°1	Famille n°2	Famille n°1	Famille n°2		
NGG	<b>CCN</b>	<b>UCN</b>	Pro ( <i>IIA</i> )	Ser ( <i>IIA</i> )	Même sous-classe (IIA) d'aaRS	Possible ambiguïté ancienne Pro/Ser de décodage (voir <a href="#">Figure 1</a> ).
NGU	<b>GCN</b>	<b>ACN</b>	Ala ( <i>IIC</i> )	Thr ( <i>IIA</i> )	Même classe (II) d'aaRS	Probable ambiguïté ancienne Ala/Thr de décodage (voir <a href="#">Figure 2</a> ).
NAG	<b>CUN</b>	<i>UUN</i>	Leu ( <i>IA</i> )	Leu ( <i>IA</i> ) Phe ( <i>IIC</i> )	Même a.a. (Leu)	Les familles CUB et UUB étaient non-codantes dans le code précurseur à 27 codons.
NAU	<b>GUN</b>	<i>AUN</i>	Val ( <i>IA</i> )	Ile ( <i>IA</i> ) Met ( <i>IA</i> )	Même sous-classe (IA) d'aaRS	Val et Ile sont les plus similaires entre eux de tous les a.a.s protéinogéniques. La famille GUB était non-codante dans le code précurseur à 27 codons. Probable ambiguïté Val / Ile de décodage dans le code intermédiaire.
NCG	<b>CGN</b>	<i>UGN</i>	Arg ( <i>ID</i> )	Cys ( <i>IB</i> ) Trp ( <i>IC</i> )	Même classe (II) d'aaRS	Pour Arg et Cys, voir <a href="#">Discussion</a> .
NCU	<b>GGN</b>	<i>AGN</i>	Gly ( <i>IIA IIC</i> )	Ser ( <i>IIA</i> ) Arg ( <i>ID</i> ) Gly <sup>(2)</sup> ( <i>IIA IIC</i> )	Même sous-classe (IIA) d'aaRS	Coïncidence accidentelle probable de Gly dans un code mitochondrial variant.
NUG <sup>(1)</sup>	<i>CAN</i>	<i>UAN</i>	His ( <i>IIA</i> ) Gln ( <i>IB</i> )	Tyr ( <i>IC</i> ) Gln <sup>(3)</sup> ( <i>IB</i> )	Même a.a. (Gln)	Coïncidence significative probable de Gln dans les codes nucléaires variants.
NUU <sup>(1)</sup>	<i>GAN</i>	<i>AAN</i>	Asp ( <i>IIB</i> ) Glu ( <i>IB</i> )	Asn ( <i>IIB</i> ) Lys ( <i>IIB IE</i> )	Même sous-classe (IIB) d'aaRS	Possible ambiguïté ancienne Asp/Asn de décodage ou remplacement tardif de Asp par Asn (voir processus de transamidation).

Le tableau énumère toutes les paires intriquées de familles de codons (colonnes 2 et 3, familles non scindées en caractères gras, familles scindées en *italiques*) prédites par l'option d'appariement 1 avec leur type d'anticodon partagé (première colonne) et leur signification de codage (colonnes 4 et 5) dans le code moderne (la couleur des a.a.s se réfère à la classe des aaRS comme dans le [tableau I](#), les sous-classes indiquées entre parenthèses sont conformes à [64]). Pour chaque paire intriquée prédite de familles de codons, la coïncidence observée dans le code moderne est notée dans la colonne 6 avec des remarques dans la colonne 7. Notes : <sup>(1)</sup> Type d'anticodon inactif dans le code précurseur à 27 codons (voir [figure 1](#)). <sup>(2)</sup> Gly se trouve à la place de Arg dans le code mitochondrial des Urochordés (voir [texte](#)). <sup>(3)</sup> Gln est codé par les codons UAR dans le code nucléaire de nombreux eucaryotes unicellulaires (voir [texte](#)).

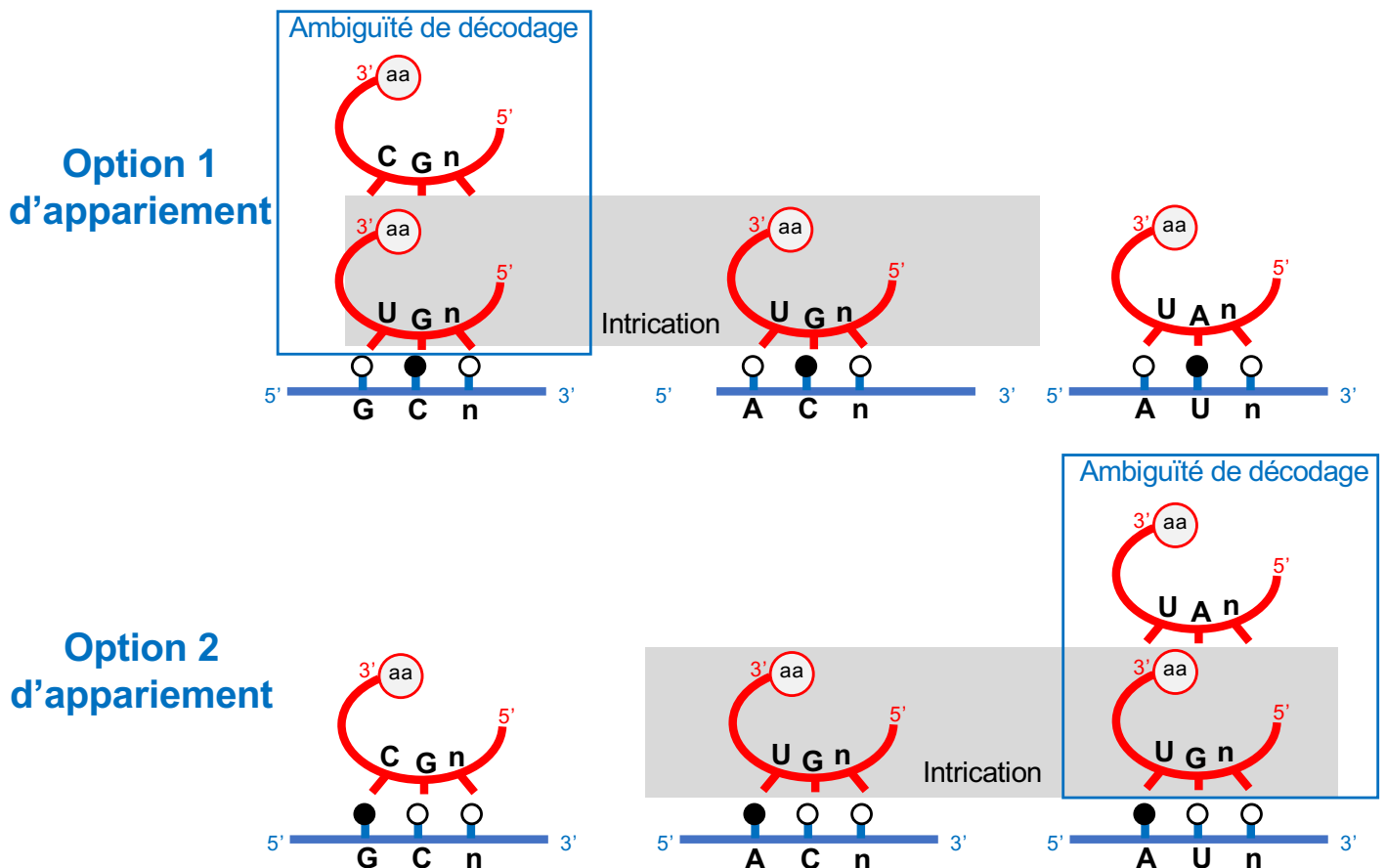
# Figure supplémentaire S1

## A: Options d'appariement



La figure illustre les 3 options hypothétiques d'appariement codon-anticodon examinées dans ce travail, par rapport aux règles d'appariement du code moderne (insert de droite). Les ovales gris symbolisent la machinerie de décodage; lignes bleues: codons; lignes rouges: anticodons; numérotation des nucléotides selon [39]. Cercles: interactions purine-pyrimidine tolérées à chaque position des duplex codon-anticodon : noir : paires Watson-Crick uniquement ; blanc : même paires plus paires faibles (ex. G•U ou U•G). Notez que dans le cadre de l'option 1-2, les paires faibles ne sont pas tolérées simultanément aux positions 1:36 et 2:35.

## B: Exemples d'ambiguïté de décodage et d'intrication des codons prédites par les options d'appariement 1 et 2



La figure illustre les conséquences des options d'appariement 1 ou 2 sur les codes hypothétiques en utilisant comme exemple les familles de codons GCN, ACN et AUN. Selon l'option choisie, une famille donnée de codons peut être reconnue par plusieurs types d'anticodons, ce qui crée une ambiguïté potentielle de décodage (encadré), et des familles distinctes de codons peuvent partager un même type d'anticodon (fond gris), ce qui crée une assignation commune (désignée ici par "intrication"). Ces deux propriétés sont des caractéristiques intrinsèques de toutes les matrices codon-anticodon explorées dans ce travail. Les codons affectés dépendent de l'option d'appariement choisie, comme illustré.

**Figure supplémentaire S2 : Matrice d'interaction codon-anticodon dans un code précurseur hypothétique G, A, U à 27 codons et conséquences.**

<b>A</b>			Familles de codons (5' – 3')								
			RRn				RYn		YRn		YYn
			GGD	GAD	AGD	AAD	GUD	AUD	UGD	UAD	UUD
nRR	5a	DGG	-	-	-	-	-	-	-	-	-/-/
	5b	DGA	-	-	-	-	-	-	-	-	-/+/+
	5c	DAG	-	-	-	-	-	-	-	-	+/+/-
	5d	DAA	-	-	-	-	-	-	-	-	++
nRY	6a	DGU	-	-	-	-	-/-/	-/+/+	-	-	-
	6b	DAU	-	-	-	-	+/+/-	++	-	-	-
nYR	7a	DUG	-	-	-	-	-	-	-/-/	+/+/-	-
	7b	DUA	-	-	-	-	-	-	-/+/+	++	-
nYY	8a	DUU	-/-/	+/+/-	-/+/+	++	-	-	-	-	-

<b>B</b>		Ambiguïté de décodage et intrication des codons								
		GGD	GAD	AGD	AAD	GUD	AUD	UGD	UAD	UUD
Pairing option 1		-	8a	-	8a	6b	6b	-	7a 7b	5c 5d
Pairing option 1-2		-	8a	8a	8a	6b	6a 6b	7b	7a 7b	5c 5c 5d
Pairing option 2		-	-	8a	8a	-	6a 6b	7b	7b	5b 5d

<b>C</b>	Amino acids correspondants		
	Option d'appariement 1	Option d'appariement 1-2	Option d'appariement 2
DGG	-	-	-
DGA	-	?	?
DAG	?	?	-
DAA	?	?	?
DGU	-	?	?
DAU	Val, ?	Val, ?	?
DUG	?	?	-
DUA	?	?	?
DUU	?	?	?

Matrice d'interaction codon-anticodon pour les molécules d'ARN primitives composées de deux purines hypothétiques (**G** et **A**) et d'une seule pyrimidine (**U**). Comparer à la [figure 1](#). Même légende, sauf pour les symboles : ++ appariement actif indépendant de l'option choisie (deux paires **A-U** ou **U-A** aux positions 1 et 2 des codons) ; + appariement actif dépendant de l'option choisie (une paire **A-U** ou **U-A** plus une paire **G-U** ou **U-G** aux positions 1 et 2 des codons) ; - aucun appariement (toute autre combinaison). **D** = non **C**.

**Figure supplémentaire S3 : Matrice d'interaction codon-anticodon dans un code intermédiaire à 64 codons G', C, U, A' et conséquences.**

A	Anciennes familles de codons (5' – 3')										Nouvelles familles de codons (5' – 3')								
	RRn		RYn		YRn		YYn				RRn		RYn		YRn				
	G'G'N	G'CN	G'UN	CG'N	UG'N	CCN	CUN	UCN	UUN	A'A'N	G'A'N	A'G'N	A'CN	A'UN	CA'N	UA'N			
nRR	1'a	NG'G'	-	-	-	-	-	-	-	++	-/+/+	+/-/+	-/-/-	-	-	-	-	-	-
nRY	2'a	NG'C	-	++	-/+/+	-	-	-	-	-	-	-	-	+/-/+	-/-/-	-	-	-	-
	2'b	NG'U	-	+/-/+	-/-/-	-	-	-	-	-	-	-	-	++	-/+/+	-	-	-	-
nYR	3'a	NCG'	-	-	-	++	+/-/+	-	-	-	-	-	-	-	-	-/+/+	-/-/-	-	-
	3'b	NUG'	-	-	-	-/+/+	-/-/-	-	-	-	-	-	-	-	-	++	+/-/+	-	-
nYY	4'a	NCC	++	-	-	-	-	-	-	-	-	-	-/-/-	-/+/+	+/-/+	-	-	-	-
	4'b	NCU	+/-/+	-	-	-	-	-	-	-	-	-	-/+/+	-/-/-	++	-	-	-	-
	4'c	NUC	-/+/+	-	-	-	-	-	-	-	-	-	+/-/+	++	-/-/-	-	-	-	-
	4'd	NUU	-/-/-	-	-	-	-	-	-	-	-	-	++	+/-/+	-/+/+	-	-	-	-
nRY	2'c	NA'C	-	-/+/+	++	-	-	-	-	-	-	-	-	-	-	-/-/-	+/-/+	-	-
	2'd	NA'U	-	-/-/-	+/-/+	-	-	-	-	-	-	-	-	-	-	-/+/+	++	-	-
nYR	3'c	NCA'	-	-	-	+/-/+	++	-	-	-	-	-	-	-	-	-/-/-	-/+/+	-	-
	3'd	NUA'	-	-	-	-/-/-	-/+/+	-	-	-	-	-	-	-	-	+/-/+	++	-	-
nRR	1'b	NA'G'	-	-	-	-	-	-/+/+	++	-/-/-	+/-/+	-	-	-	-	-	-	-	-
	1'c	NG'A'	-	-	-	-	-	+/-/+	-/-/-	++	-/+/+	-	-	-	-	-	-	-	-
	1'd	NA'A'	-	-	-	-	-	-/-/-	+/-/+	-/+/+	++	-	-	-	-	-	-	-	-

C	Amino-acids correspondants		
	Option d'appariement 1	Option d'appariement 1-2	Option d'appariement 2
	Anti-codons		
NG'G'	Pro, Ser	Pro, Ser, Leu	Pro, Leu
NG'C	Ala, Thr	Ala, Val	Ala, Val
NG'U	Ala, Thr	Ala, Thr, ?	Thr, ?
NCG'	Arg, ?	Arg, ?	Arg, ?
NUG'	?	Arg, ?	Arg, ?
NCC	Gly, ?	Gly, ?	Gly, ?
NCU	Gly, ?	Gly, ?	?
NUC	?	Gly, ?	Gly, ?
NUU	?	?	?
NA'C	Val	Val, Ala, ?	Val, Ala, ?
NA'U	Val, ?	Val, Thr, ?	Thr, ?
NCA'	Arg, ?	Arg, ?	?
NUA'	?	?	?
NA'G'	Leu, ?	Leu, Pro, ?	Leu, Pro
NG'A'	Ser, Pro	Ser, Pro, ?	Ser, ?
NA'A'	Leu, ?	Leu, Ser, ?	Ser, ?

B	Ambiguïté de codage et intrication des codons															
	G'G'N	G'CN	G'UN	CG'N	UG'N	CCN	CUN	UCN	UUN	A'A'N	G'A'N	A'G'N	A'CN	A'UN	CA'N	UA'N
Option d'appariement 1	4'a	2'a	2'c	3'a	3'a	1'a	1'b	1'a	1'b	4'c	4'c	4'a	2'a	2'c	3'b	3'b
	4'b	2'b	2'd	3'c	3'c	1'c	1'd	1'c	1'd	4'd	4'd	4'b	2'b	2'd	3'd	3'd
Option d'appariement 1-2	4'a	2'a	2'a	3'a	3'a	1'a	1'a	1'a	1'b	4'b	4'a	4'a	2'a	2'b	3'a	3'b
	4'b	2'b	2'c	3'b	3'c	1'b	1'b	1'c	1'c	4'c	4'c	4'b	2'b	2'c	3'b	3'c
Option d'appariement 2	4'c	2'c	2'd	3'c	3'd	1'c	1'd	1'd	1'd	4'd	4'd	4'd	2'd	2'd	3'd	3'd
	4'a	2'a	2'a	3'a	3'c	1'a	1'a	1'c	1'c	4'b	4'a	4'b	2'b	2'b	3'a	3'c
	4'c	2'c	2'c	3'b	3'd	1'b	1'b	1'd	1'd	4'd	4'c	4'd	2'd	2'd	3'b	3'd

Matrice d'interaction codon-anticodon pour des molécules d'ARN hypothétiques composées de deux purines hypothétiques (symbolisées par G' et A') avec des préférences d'appariement opposées pour les deux pyrimidines (C et U), c'est-à-dire G'-C ou C-G' > G'•U ou U•G' et A'-U ou U-A' > A'•C ou C•A'. Comparer avec la figure 2. Même légende, sauf pour les symboles : ++ appariement actif indépendant de l'option choisie (deux paires G'-C ou C-G', deux paires A'-U ou U-A', ou une paire G'-C ou C-G' plus une paire A'-U ou U-A' aux positions 1 et 2 des codons) ; + appariement actif dépendant de l'option choisie (une paire G'-C, C-G', A'-U ou U-A' plus une paire G'•U, U•G', A'•C ou C•A' aux positions 1 et 2 des codons) ; - aucun appariement (toute autre combinaison). Notez que la partie C n'est présentée ici que pour faciliter la comparaison avec la figure 2, en supposant que les deux purines, G' et A', sont équivalentes à G et A, respectivement.