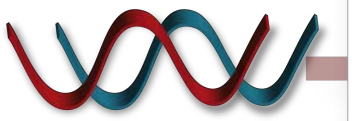
Origine et évolution du code génétique
&
Biologie Synthétique

Hubert Becker
IPCB, UMR7156 4ème étage

http://gmgm.unistra.fr/index.php?id=3634

Nirenberg, M. W. & Matthaei, J. H. (1961). Proc. Natl. Acad. Sci. U. S. A. 47,1588-1602.



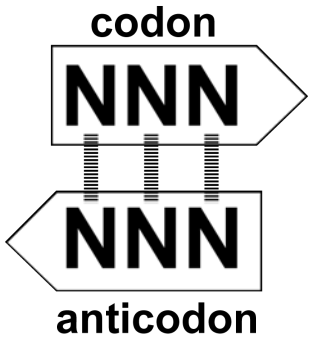
Genes
(storage)

First Codon Letter	Second Codon Letter: U	Second Codon Letter: C	Second Codon Letter: A	Second Codon Letter: G	Third Codon Letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG	UAG Stop	UGG Trp	G
C	CUU	CGU Pro	CAU His	CGU Arg	U
	CUC	CGC	CAC	CGC	C
	CUA	CGA	CAA Gln	CGA	A
	CUG	CGG	CAG	CGG	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met	ACG	AAG	AGG	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu	GGA	A
	GUG	GCG	GAG	GGG	G

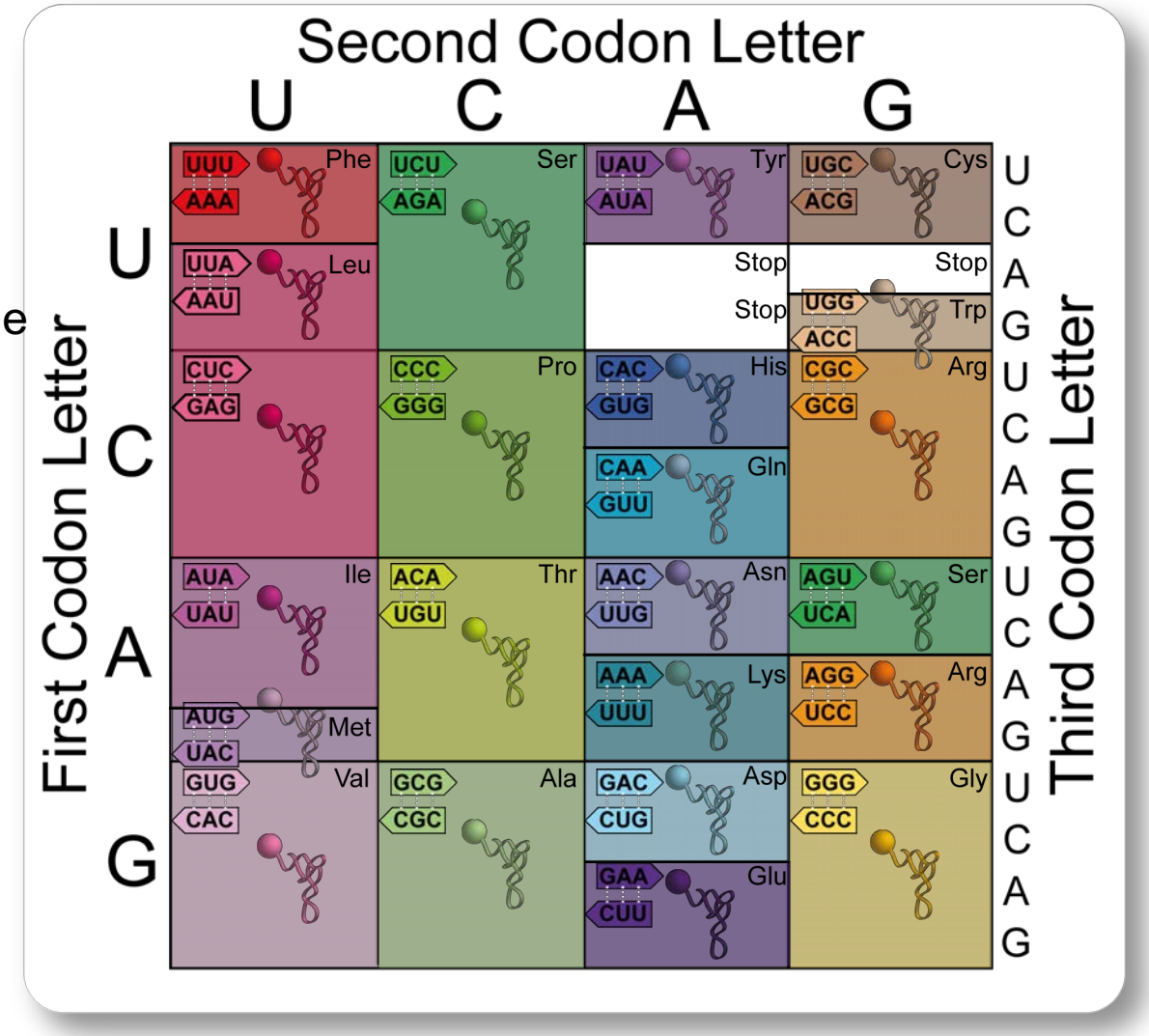
Proteins
(function)

Nirenberg, M. W. & Matthaei, J. H. (1961). Proc. Natl. Acad. Sci. U. S. A. 47,1588-1602.

Theoritycal Code



Physical Code



20 Adaptors = 20 aa-tRNAs … and 20 aminoacyl-tRNA synthetases (aaRS)

Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. 12, 138–163.ad. Sci. U. S. A. 47,1588-1602.
Söll, D. and Schimmel, P.R. (1974).The Enzymes 10, 489–538.

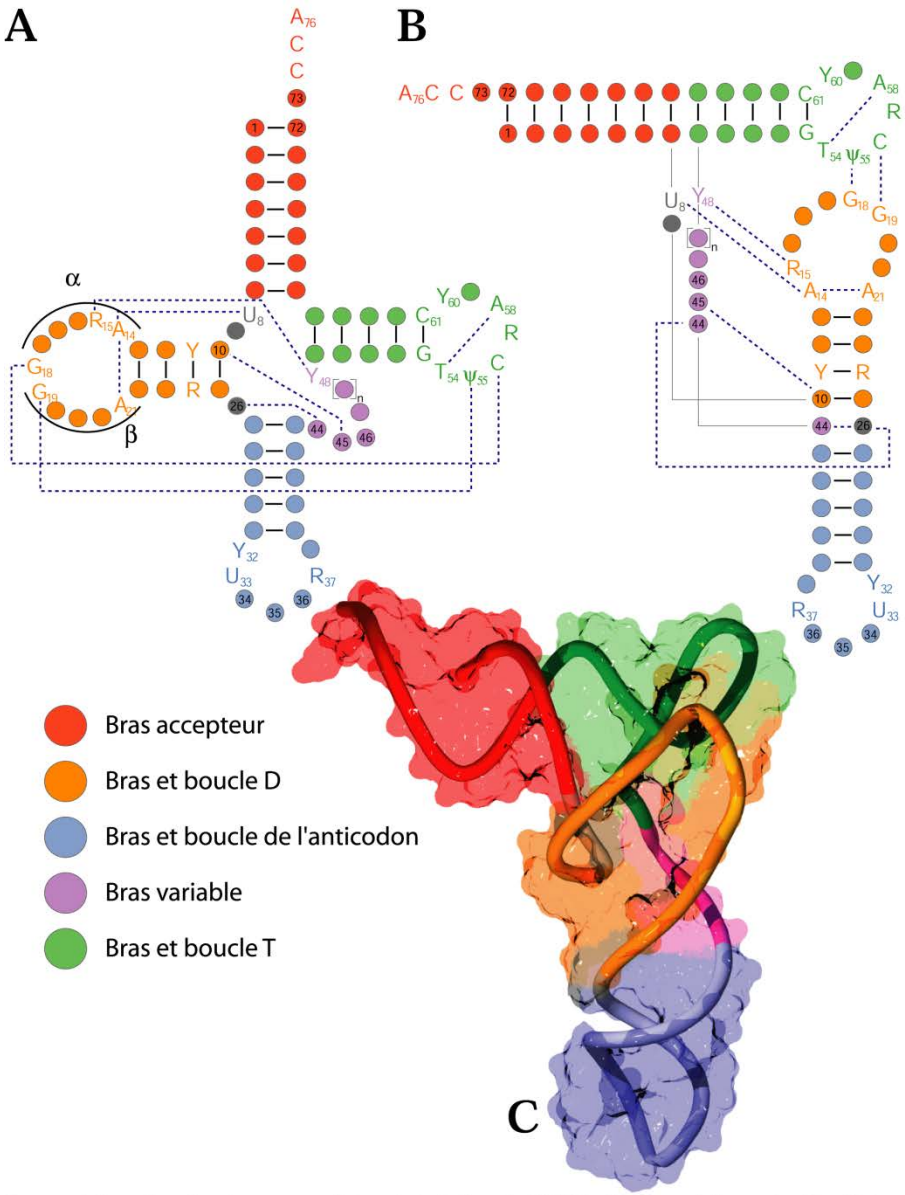
The adaptor hypothesis

Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. 12, 138–163.

A theory by Francis Crick published in 1958, which prdicted how protein synthesis proceeds and what molécules are involved while none where characterized at the time.

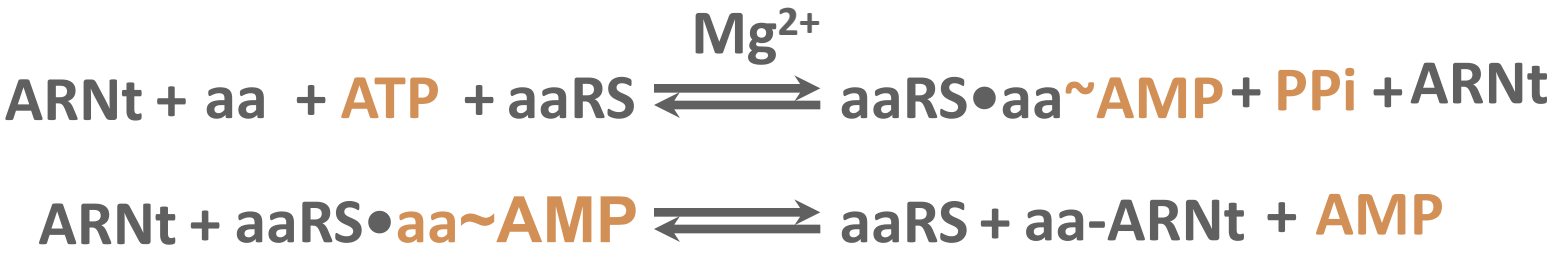
'… RNA from microsomial particules constitutes the matrix'	Messenger RNA
'… whatever the molecule that binds the matrix, it binds with specificity and through hydrogen bonds'	codon-anticodon interactions
'… the amino acid is brought by an adaptor molecule'	aminoacyl-tRNA
'… this adaptor... Very likely contains nucleotides'	tRNA
'… a particular enzyme is required to bind each amino acid to each corresponding adaptor'	aminoacyl- tRNA synthetases (aaRS)
'… the specificity requires... To distinguish valine from isoleucine is carried by these enzymes'	aaRS's editing activity

Reaction substrates: tRNA

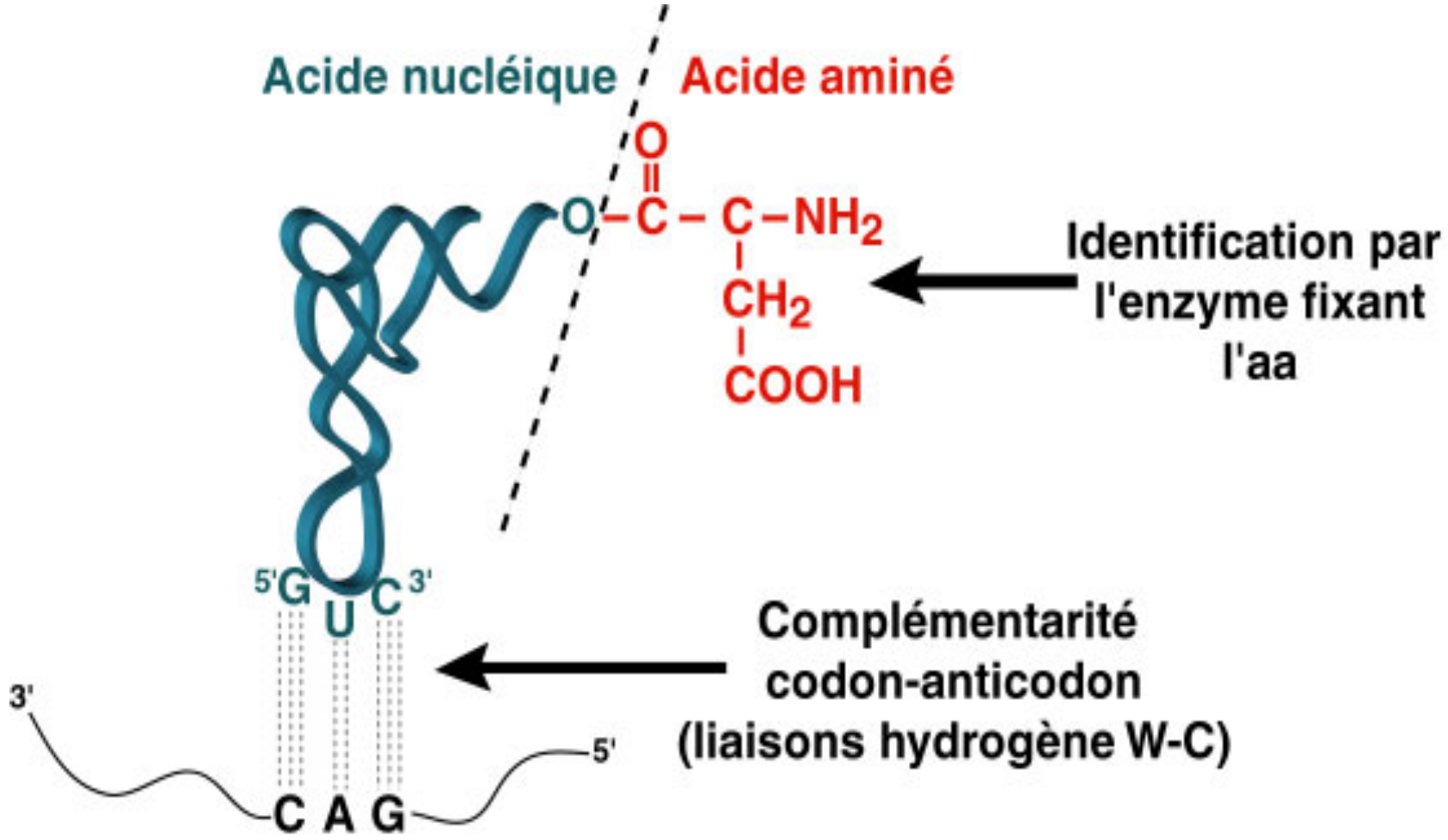


Aminoacylation reaction, components, peculiarities

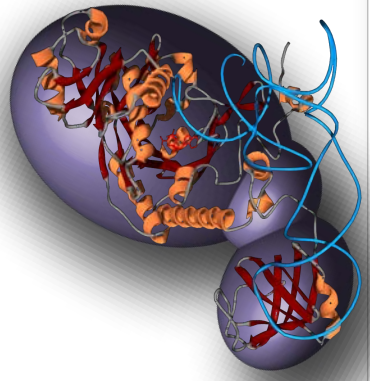
- La réaction



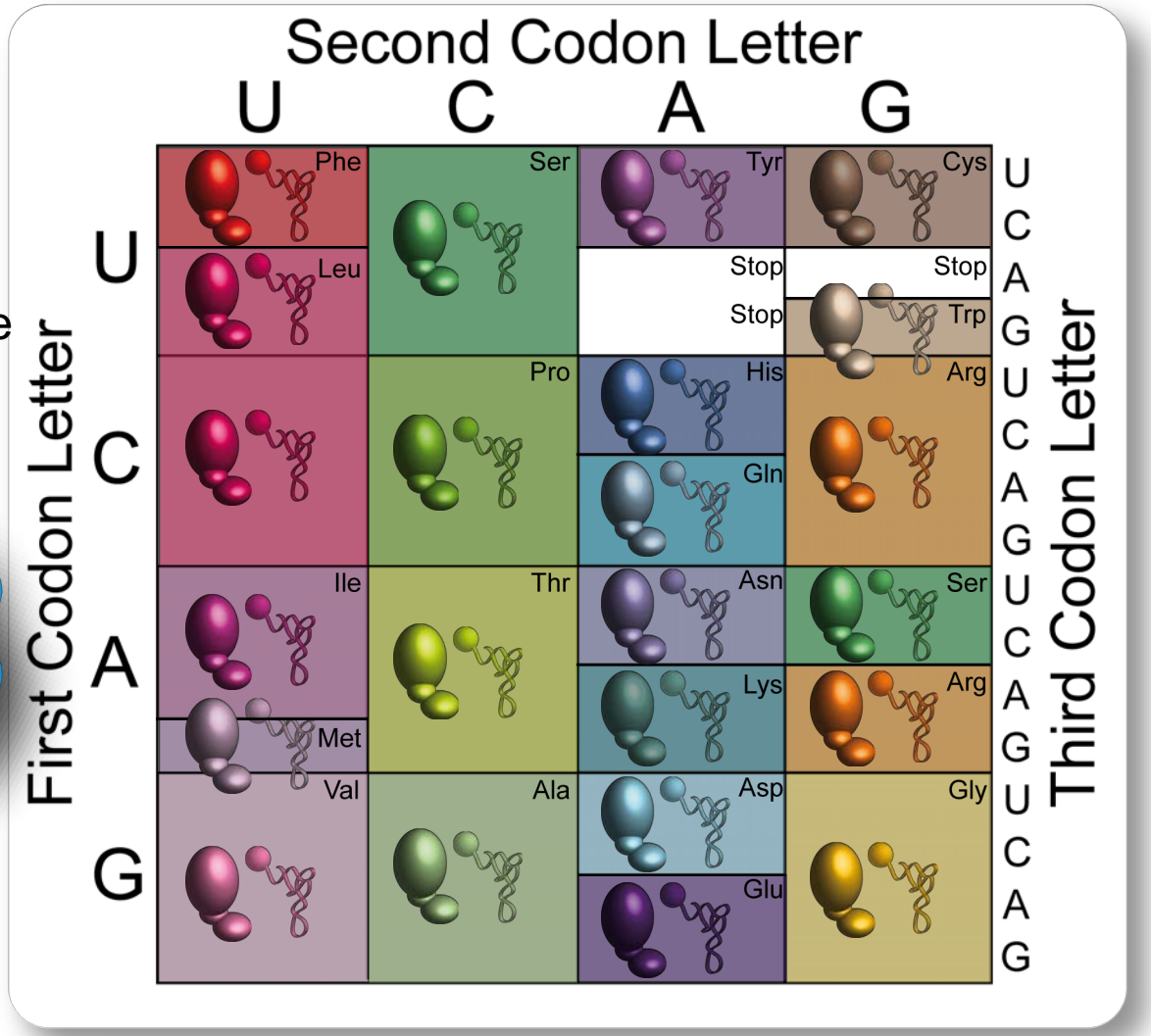
Aa-tRNA dual identity



Theoritycal Code



Physical Code



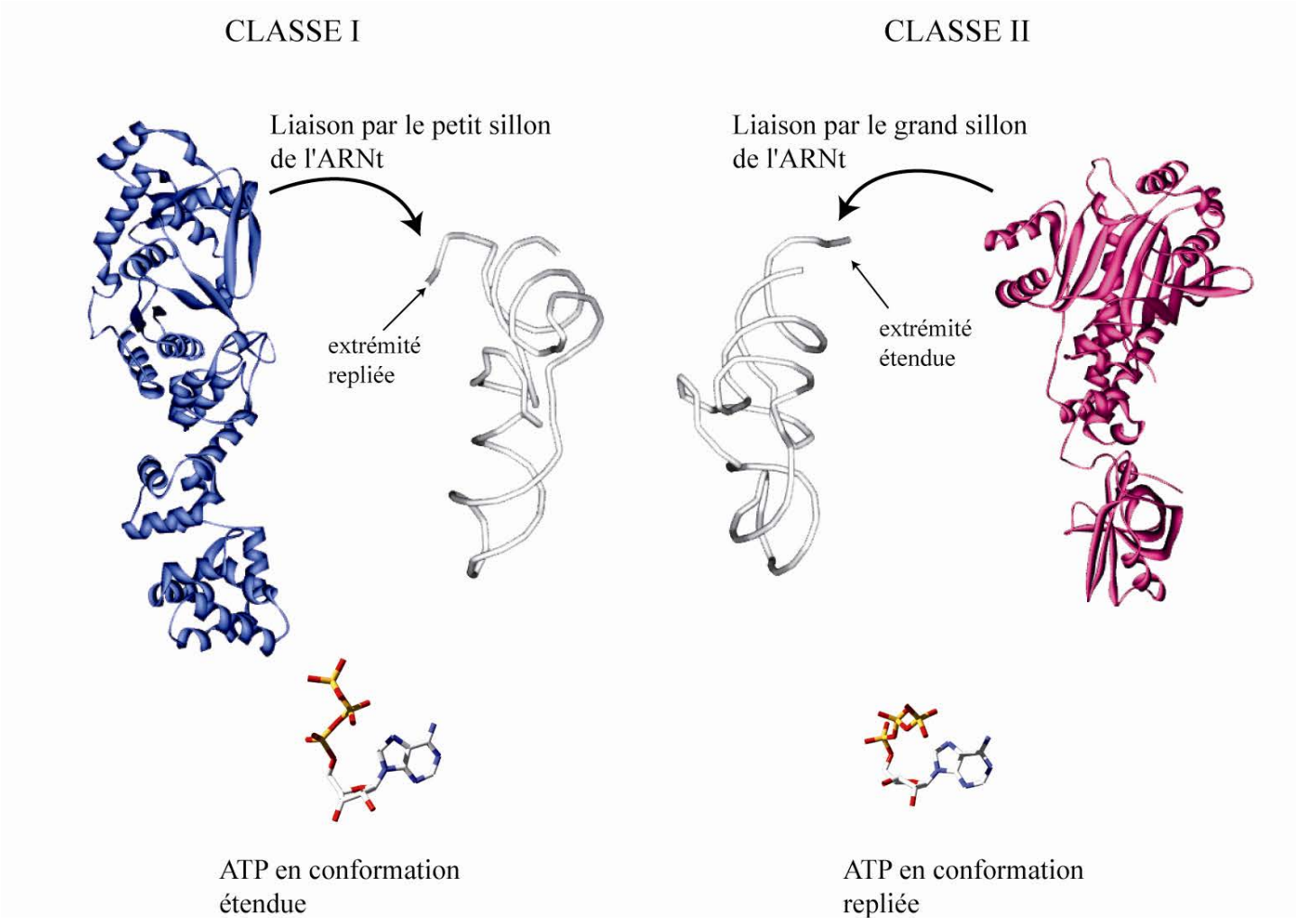
20 Adaptors = 20 aa-tRNAs … and 20 aminoacyl-tRNA synthetases (aaRS)

Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. 12, 138–163.ad. Sci. U. S. A. 47,1588-1602.
Söll, D. and Schimmel, P.R. (1974).The Enzymes 10, 489–538.

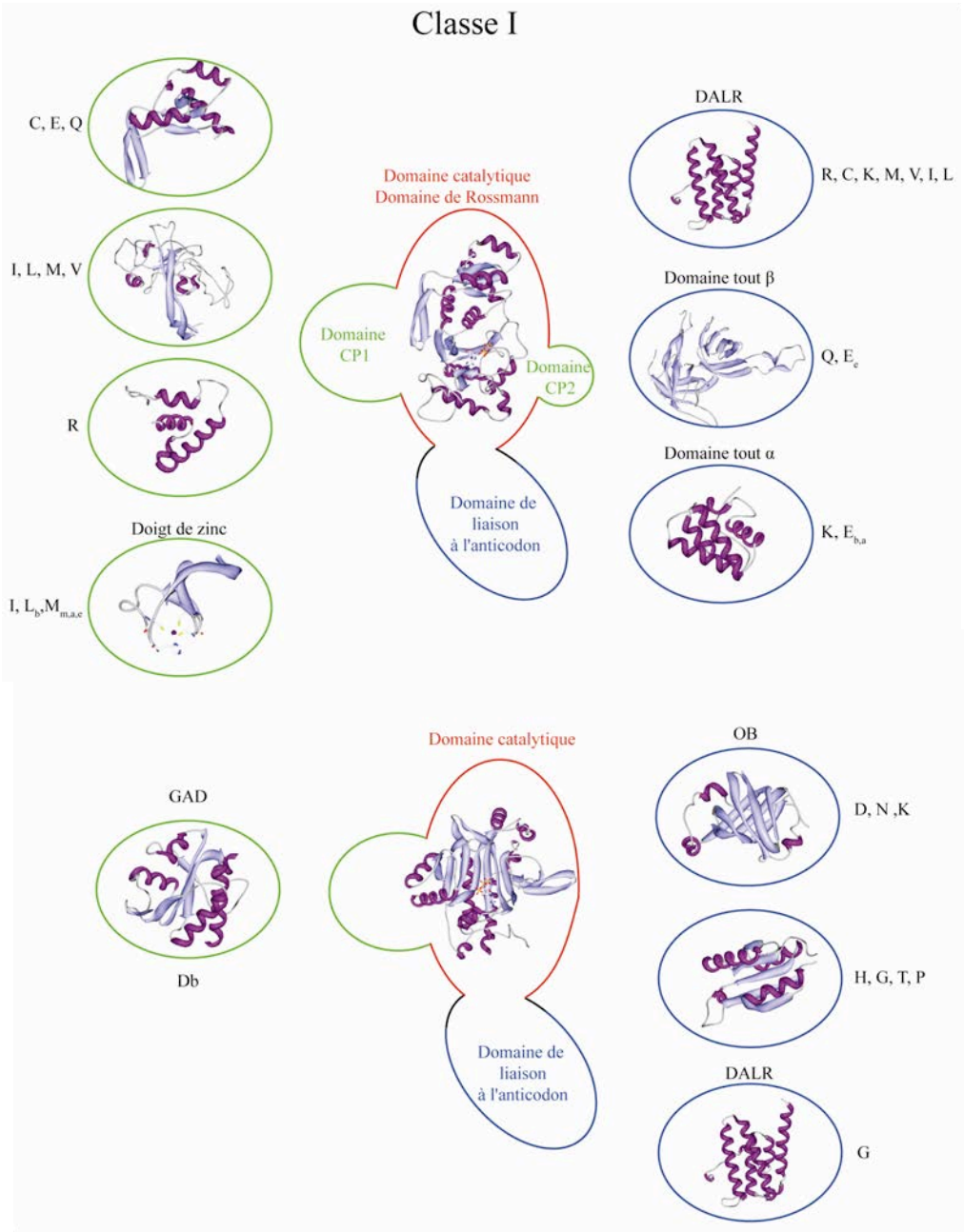
The enzymes: Class I and class II aaRSs

		Classe I	Classe II
Séquences signatures		HIGH et KMSKS	Motifs consensus 1, 2, 3: Motif 1: ϕXXX+GϕXXϕXXPϕϕ Motif 2: +ϕϕXϕXXX(F/H/Y)RX(E/D) 4 à 12 X (R/H)ϕX-FXXX-ϕXϕϕ Motif 3: λXϕGϕGϕGϕERϕϕϕϕϕ 7 à 12 X ϕP
Aminoacylation de l'ARNt		2'OH	3'OH
Structure du domaine catalytique		Domaine de Rossmann 5 brins b parallèles	7 brins b antiparallèles
Sous-classes et structure oligomérique	a	Leu a Ile* a Val* a Cys a Met a_2, a Arg a	Ser a_2 Thr* a_2 His a_2 Pro a_2 Gly a_2, a_2
	b	Glu a Gln a	Asp a_2 Asn a_2 Lys II a_2
	c	Tyr a_2 Trp a_2	Phe* a_2b_2, (a, mitochondriale) Gly a_2b_2 Ala a_2, a,
Exceptions		Lys I a, a	Pyl Sep

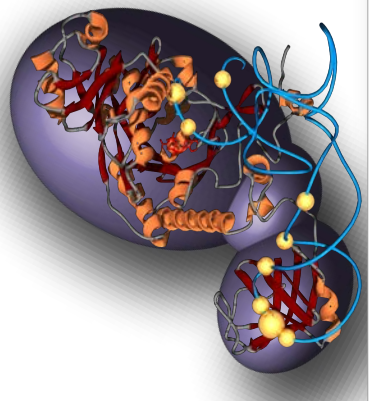
Structural & functional characteristics of the 2 Class



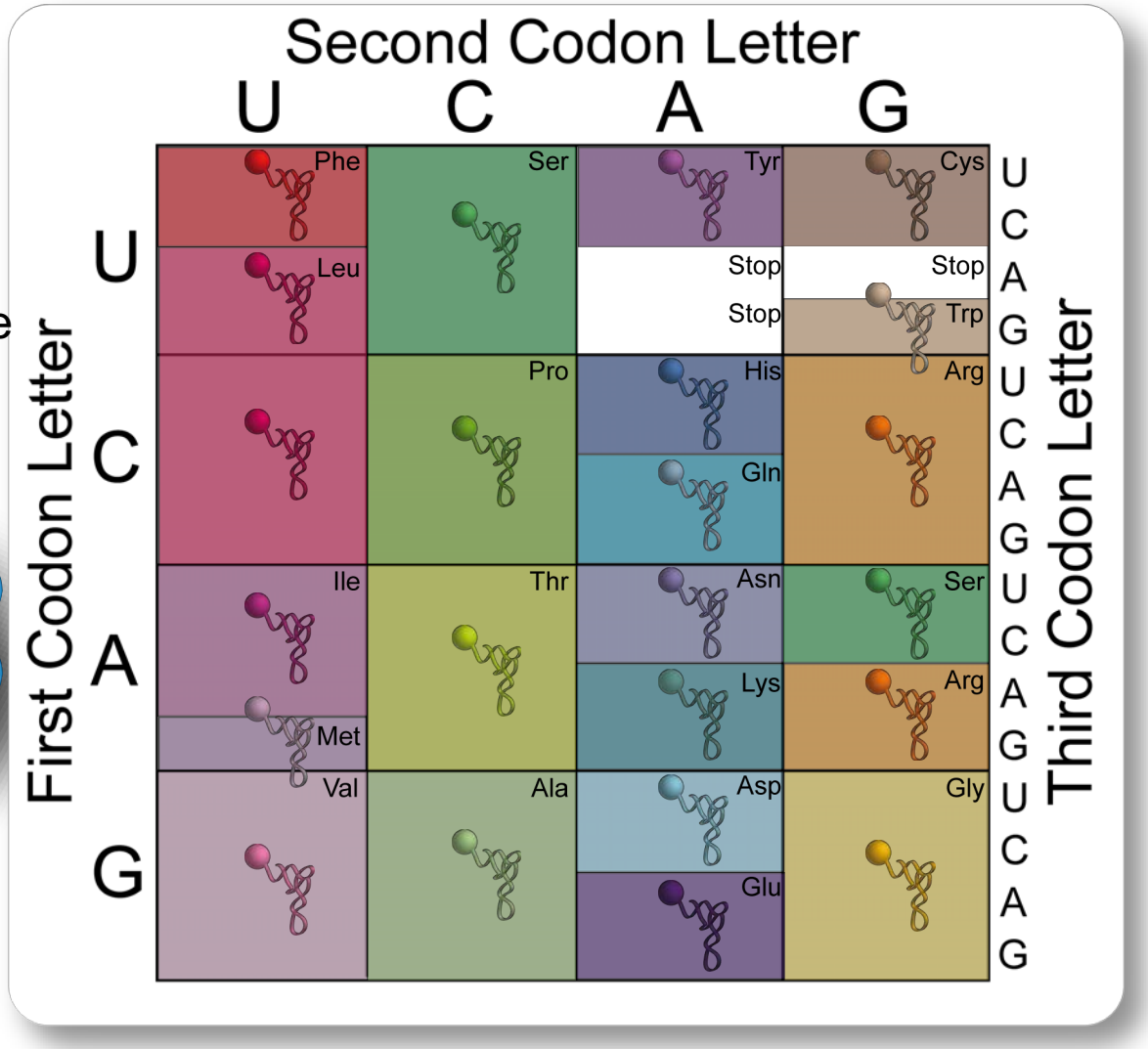
Modularity of the aaRSs



Theoritycal Code



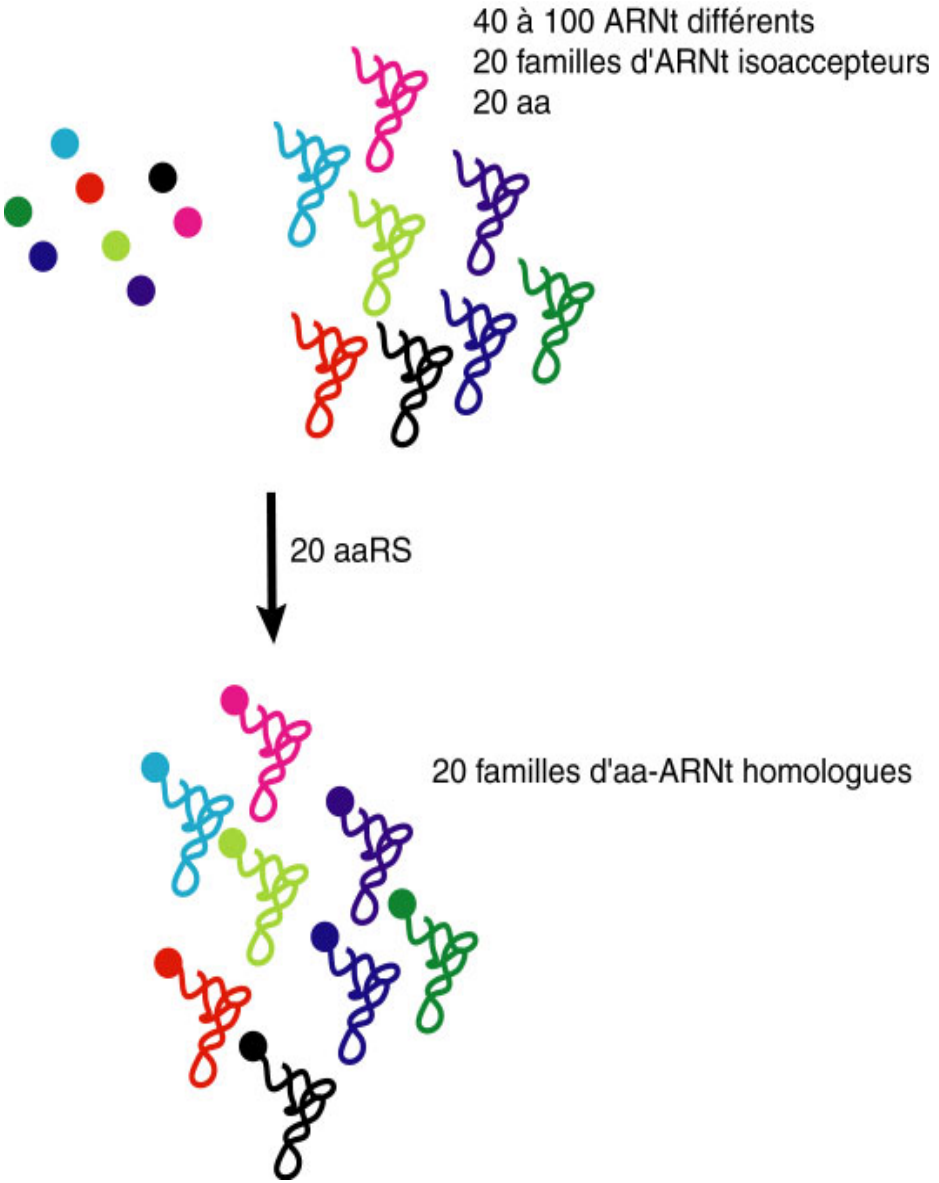
Physical Code

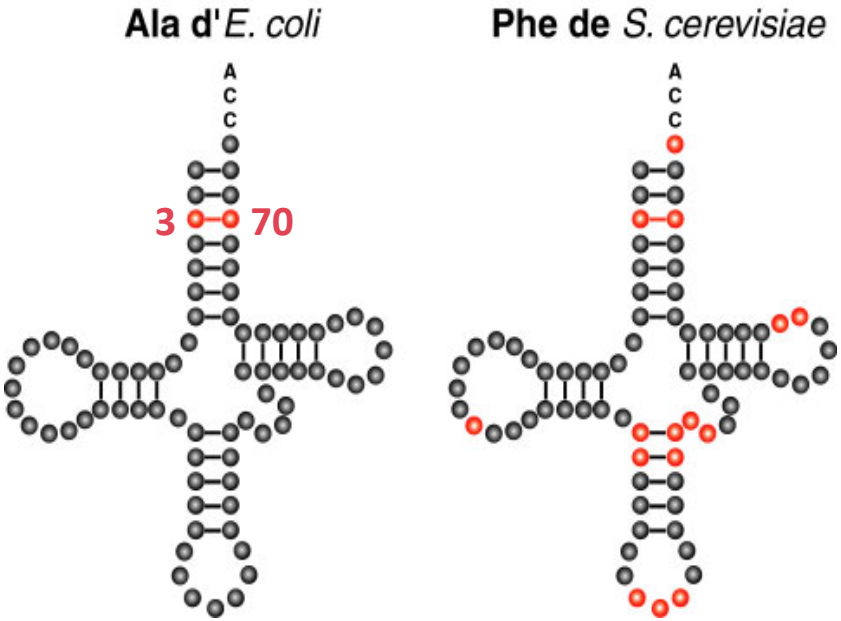


20 Adaptors = 20 aa-tRNAs … and 20 aminoacyl-tRNA synthetases (aaRS)

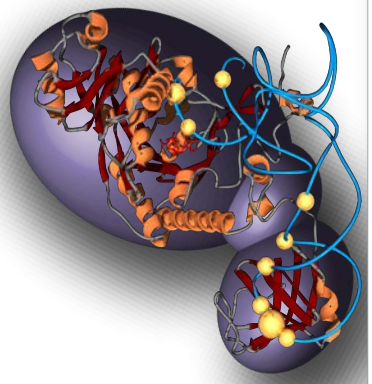
Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. 12, 138–163.ad. Sci. U. S. A. 47,1588-1602.
Söll, D. and Schimmel, P.R. (1974).The Enzymes 10, 489–538.

tRNA identity elements

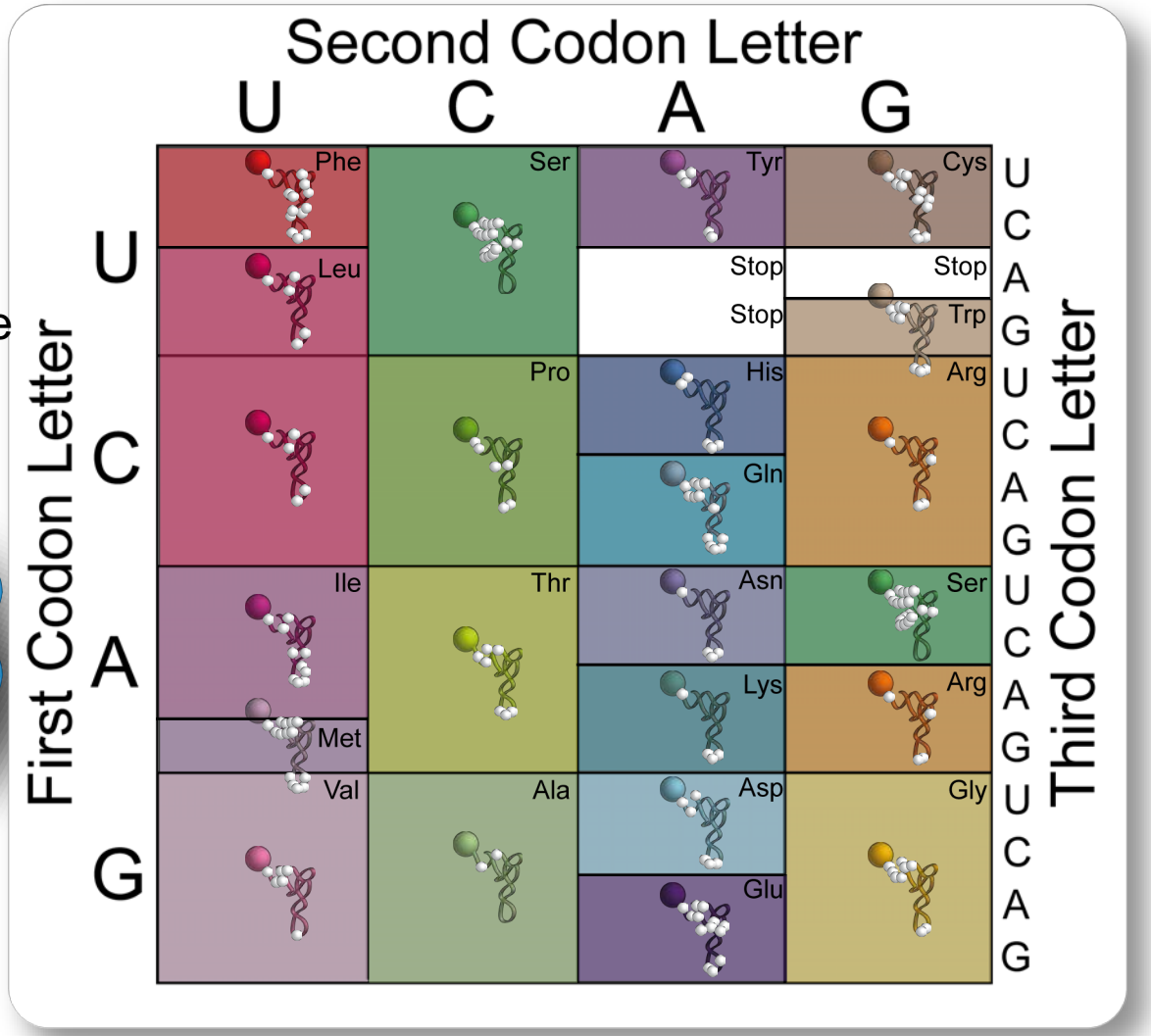




Theoritycal Code

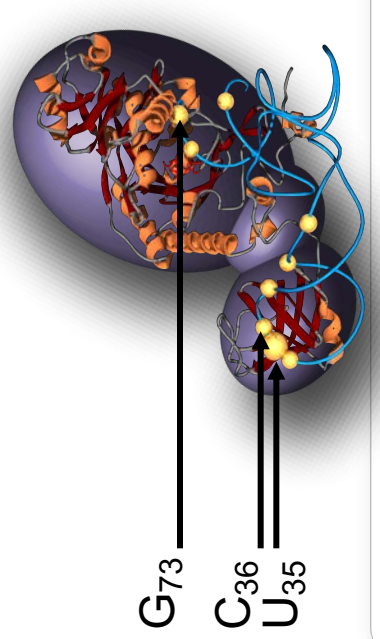


Physical Code



20 Adaptors = 20 aa-tRNAs … and 20 aminoacyl-tRNA synthetases (aaRS)

Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. 12, 138–163.ad. Sci. U. S. A. 47,1588-1602.
Söll, D. and Schimmel, P.R. (1974).The Enzymes 10, 489–538.



Second Codon Letter

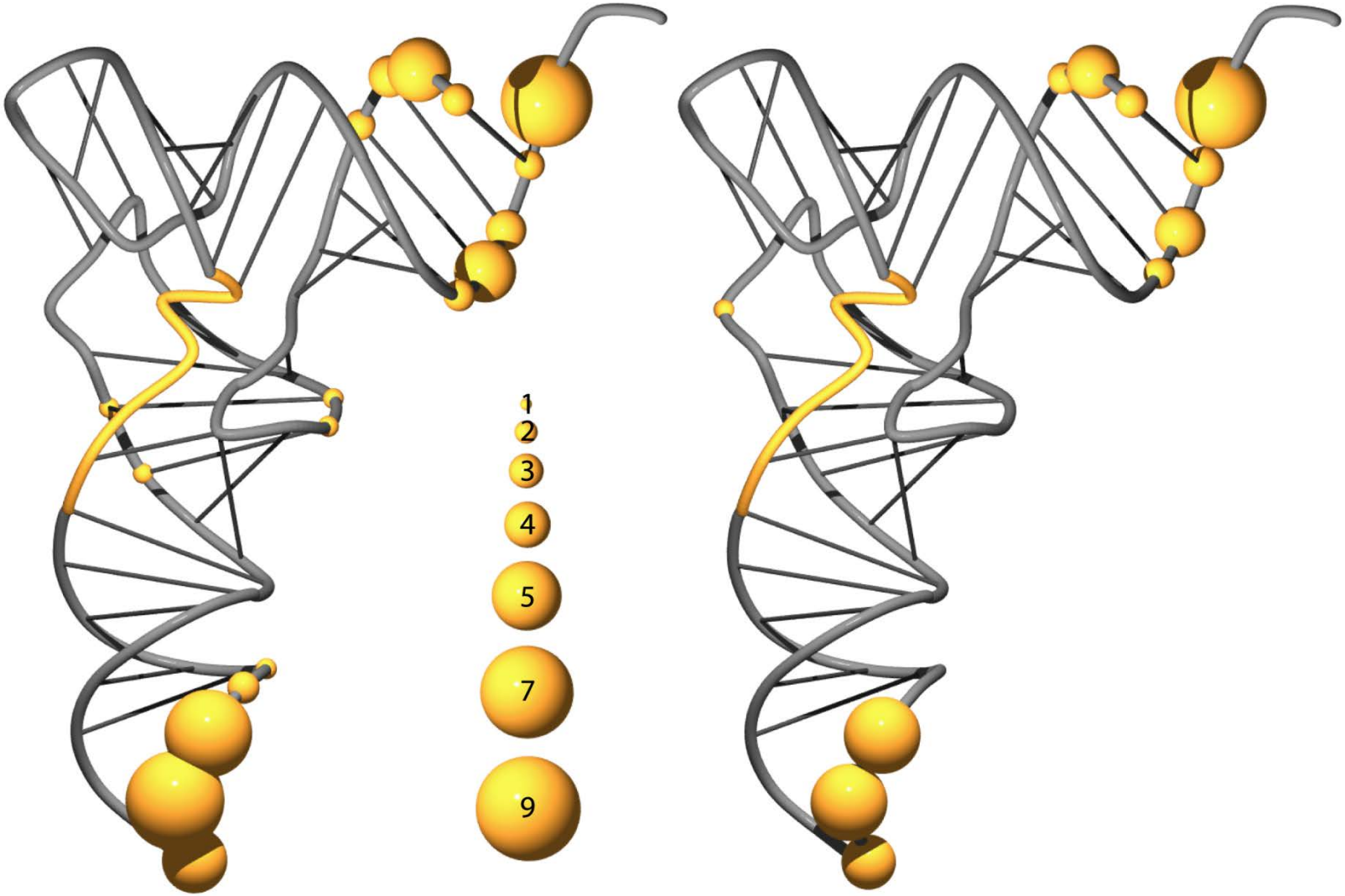
First Codon Letter	U	C	A	G	Third Codon Letter
U	A_{73} G_{34},A_{35},A_{36} U_{20},G_{44},U_{45},$U_{59\text{-}60}$ G_{27}-C_{41},G_{28}-C_{42}	A_{73},C_{72} G_2-C_{71},R_4-Y_{69} C_{11}-G_{24},V-loop	A_{73} U_{35}	U_{73} G_2-C_{71},C_3-G_{70} G_{34},C_{35},A_{36}	U
	A_{73} U_8•A_{14}			G_{73} A_1-U_{72},C_3-G_{70} G_{34},A_{35},U_{36}	C
C	A_{73} U_8•A_{14}	A_{73},G_{72} G_{35},G_{36} G_{15}•U_{48}	C_{73} G_{-1} G_{34},U_{35},G_{36}	A/G_{73} G_{35},U/G_{36} A_{20}	A
			G_{73} U_1-A_{72},G_3-C_{70} Y_{34}-U_{38} G_{10}		G
A	A_{73} C_4-G_{69} L_{34}-U_{38} U_{12}-A_{23},G_{29}•G_{41}	G_1-C_{72},C_2-G_{71} C_{34},G_{35},U_{36}	G_{73} G_{34},U_{35},U_{36}	G_{73},C_{72} G_2-C_{71},R_4-Y_{69} C_{11}-G_{24},V-loop	U
	A_{73} G_2-C_{71},A_5-U_{68} C_{32}-A_{37}		A_{73} s^4U_{34},U_{35},U_{36}	A/G_{73} G_{35},U/G_{36} A_{20}	C
G	A_{73} G_3•U_{70} U_4-A_{69}	A_{73} G_3•U_{70} G_{20}	G_{73} G_2-C_{71} G_{34},U_{35},C_{36},C_{38}	U_{73} G_1-C_{72},G_3-C_{70} C_{35},C_{36}	A
			U_{73} G_1-C_{72},C_2-G_{71} s^4U_{34},U_{35},A_{37} U_{11}-A_{24},U_{13}•G_{22} A_{46}, Δ_{47}		G

Eiler, S., et al. (1999) EMBO J. 18, 6532-6541.

Giegé, R., et al. (1998). Nucl. Acids Res. 26, 5017-5035.

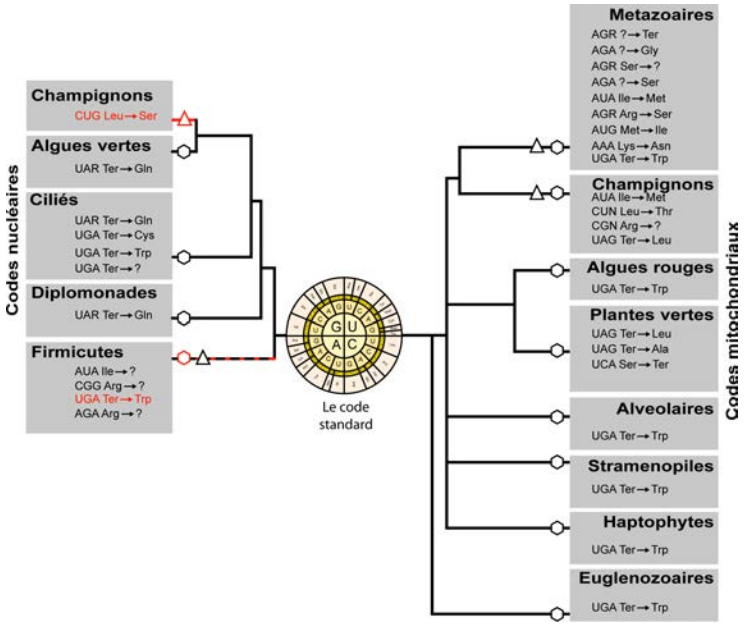
Localization of the identity elements

- Class I aaRS
- Class II aaRS



Genetic code evolution

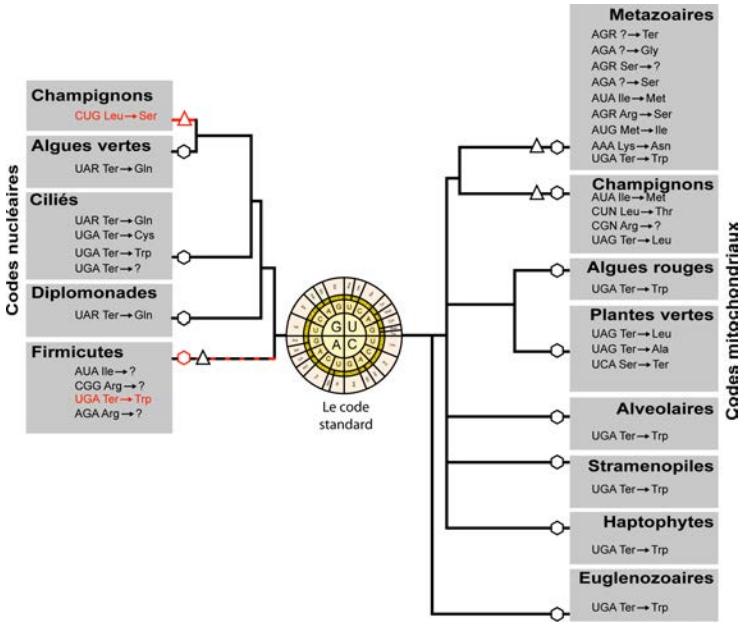
There are alternate codes – which means that the genetic code is not universal – which means has evolved



The genetic code is not a *frozen accident* of evolution

Genetic code evolution

There are alternate codes – which means that the genetic code is not universal – which means has evolved



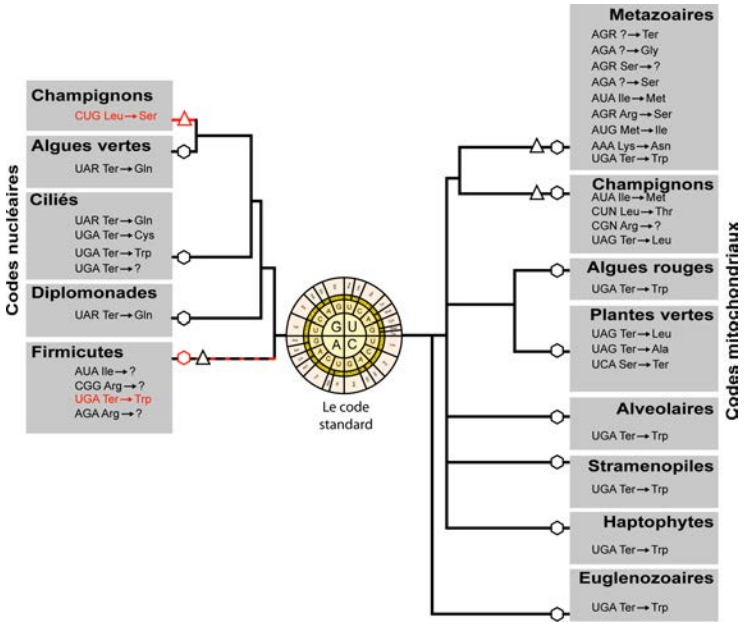
The genetic code is not a *frozen accident* of evolution

Is there a logic in codon assignements ?

Première lettre	Deuxième lettre U	Deuxième lettre C	Deuxième lettre A	Deuxième lettre G	Troisième lettre
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

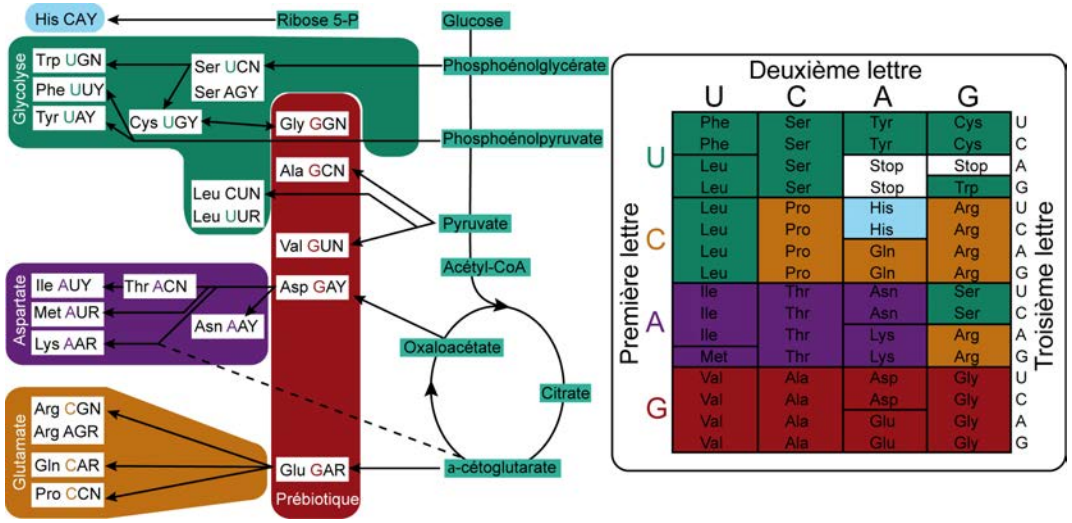
Genetic code evolution

There are alternate codes – which means that the genetic code is not universal – which means has evolved



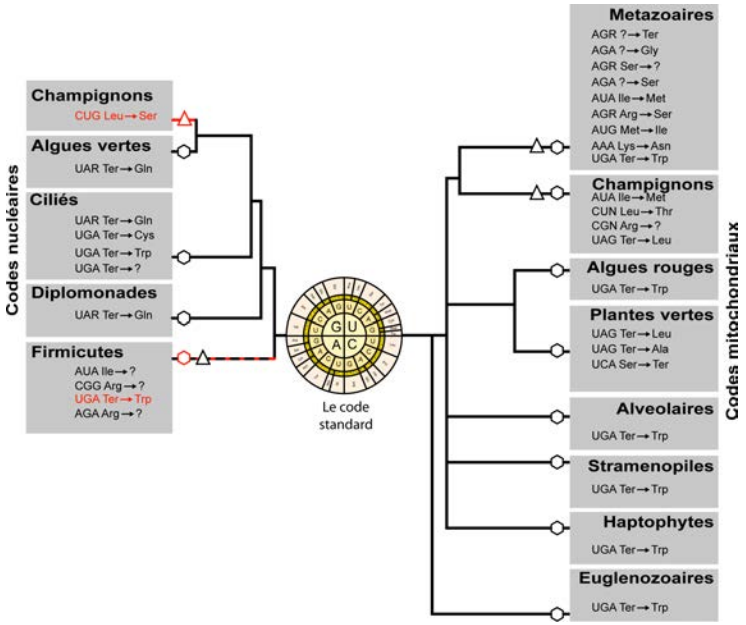
The genetic code is not a *frozen accident* of evolution

Is there a logic in codon assignements ? 1st letter of codons = aa biosynthesis pathway



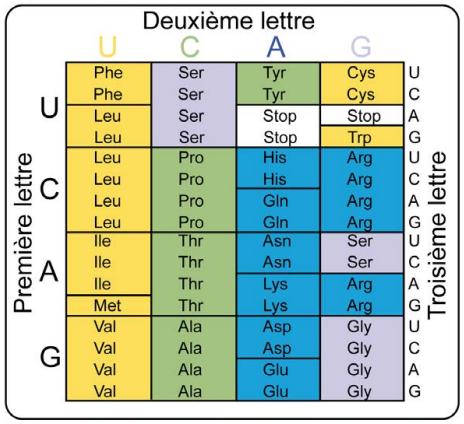
Genetic code evolution

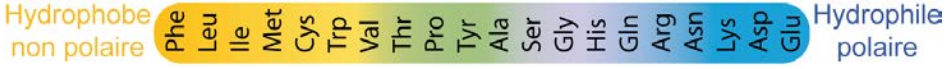
There are alternate codes – which means that the genetic code is not universal – which means has evolved



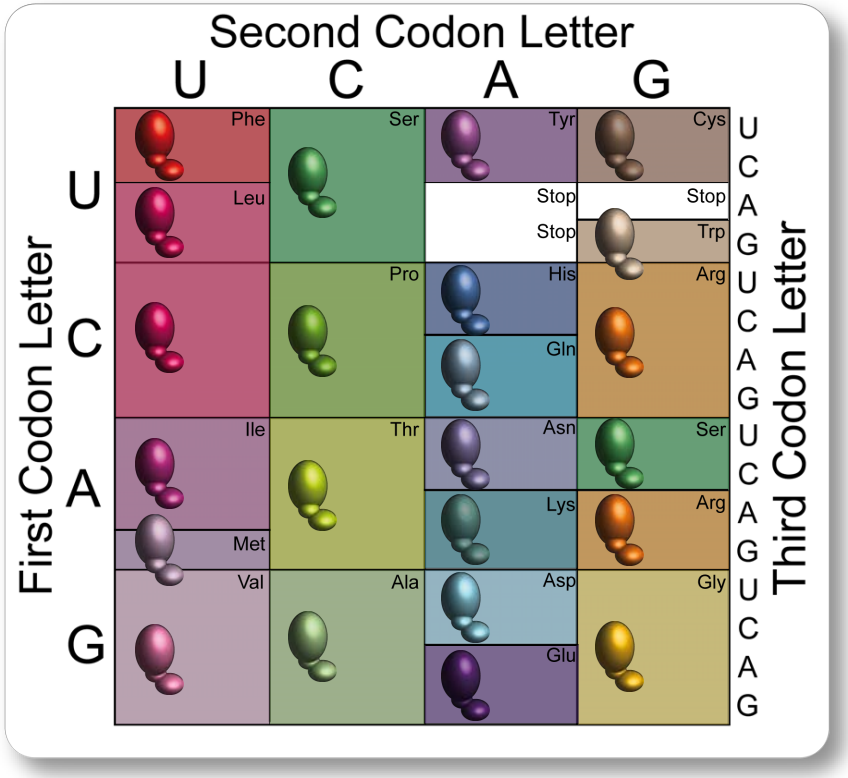
The genetic code is not a *frozen accident* of evolution

Is there a logic in codon assignements ? 2nd letter of codons = aa chemical properties

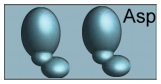




Genetic code nomalies and molecular fossils to study evolution

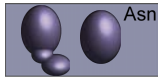


Anomalies in the pool of aaRS



Duplication of aaRS
(>50% procaryotes)

Absence of a given aaRS
(>50% procaryotes)

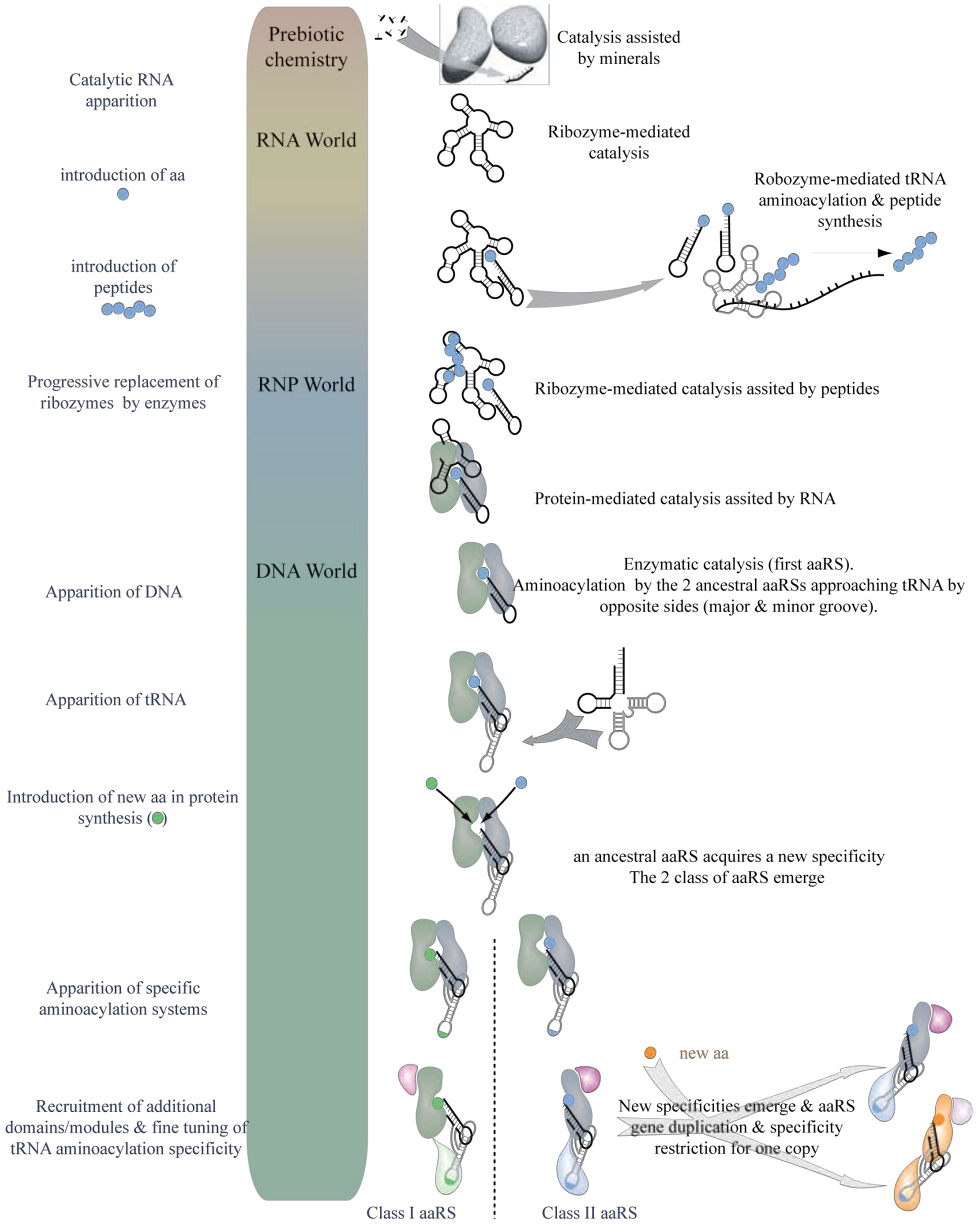


Presence of pieces of aaRS
(>50% procaryotes)

Study molecular fossils to understand
Genetic Code Evolution

In 2005, from the **240** procaryotic genomes that had been sequenced only **18** encoded a **full and unique set of 20 aaRS**

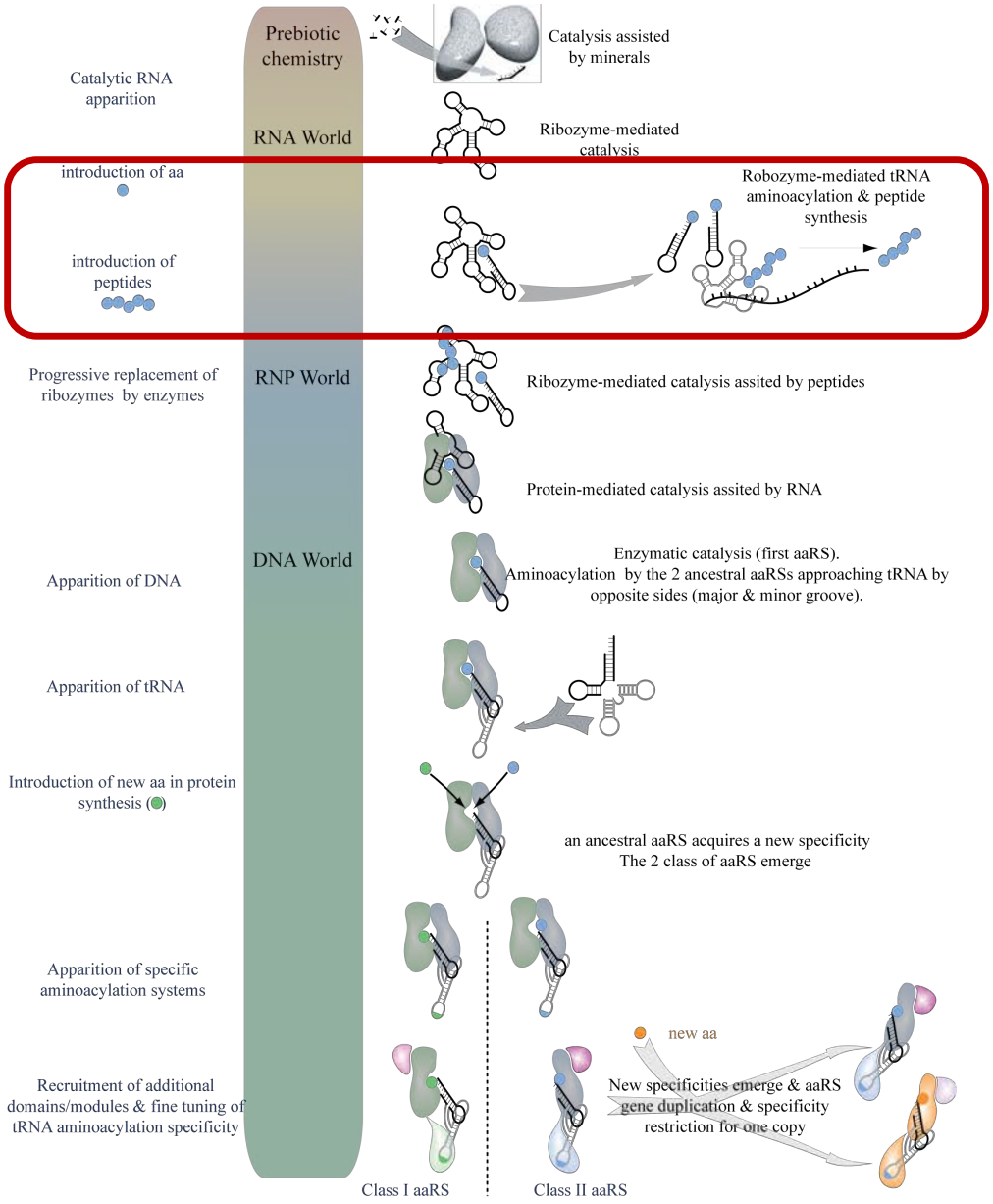
Origin of the Genetic code & RNA world

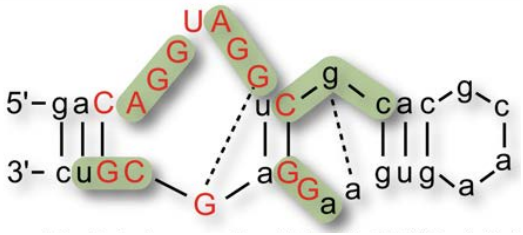


- The genetic code is pre-biotic (-3,5 billion years)

- Understanding how the genetic code appeared = how an >RNA piece was selected to carry an aa

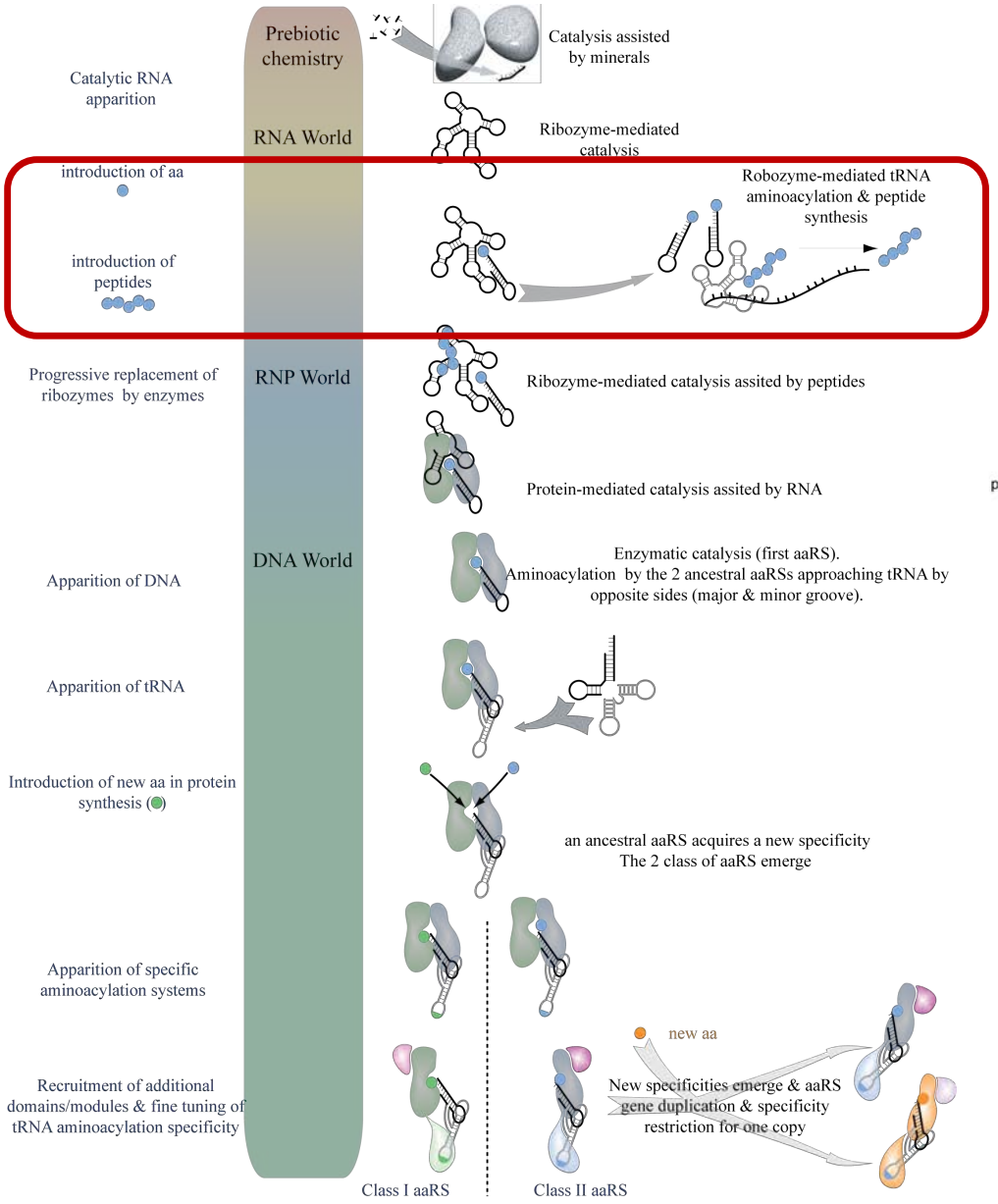
Origin of the Genetic code & RNA world

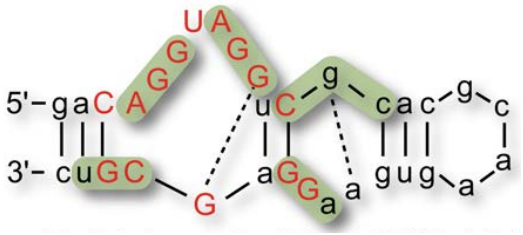




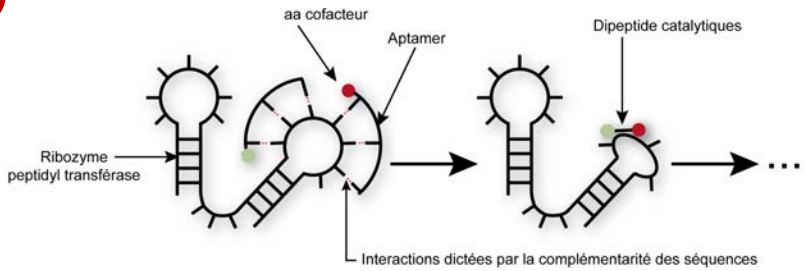
Secondary stucture of the TAR aptamer (knight & Landweber (1998). Ntides in Red are Arg-binding site, Arg codons are on grey background; dotted lines: non WC base pairs

Origin of the Genetic code & RNA world



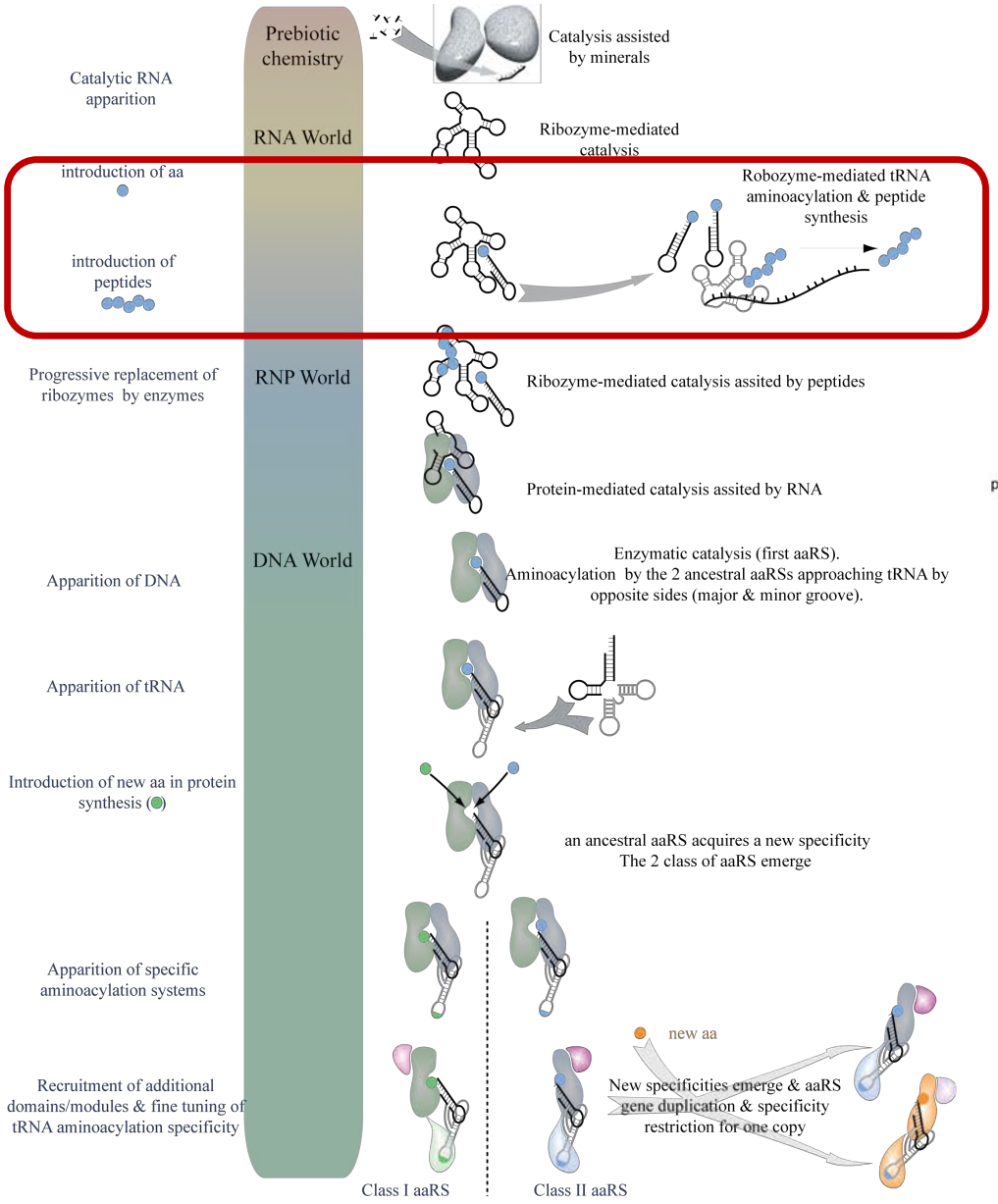


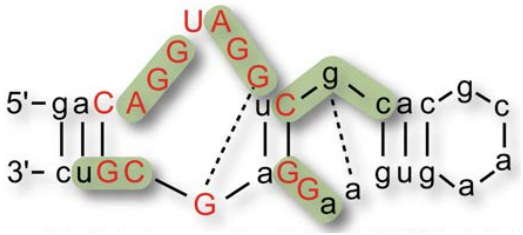
Secondary stucture of the TAR aptamer (knight & Landweber (1998). Ntides in Red are Arg-binding site, Arg codons are on grey background; dotted lines: non WC base pairs



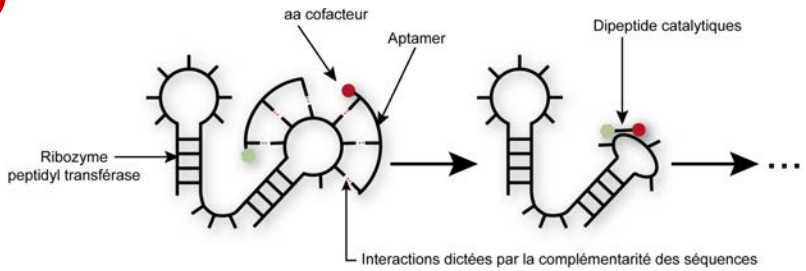
Protein synthesis catalyzed by ribozymes using aa-tRNAs as cofactors (Szathmary; 1999). Peptidyl-transferase ribozyme that bind in a sequence-specific manner RNAs with aa co-factors. The first catalytic dipeptides are made which can bind to other aptamers or assist ribozymes increasing the composition and size of the peptides that are made

Origin of the Genetic code & RNA world



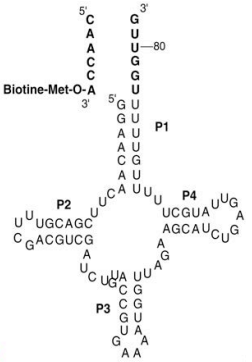


Secondary stucture of the TAR aptamer (knight & Landweber (1998). Ntides in Red are Arg-binding site, Arg codons are on grey background; dotted lines: non WC base pairs

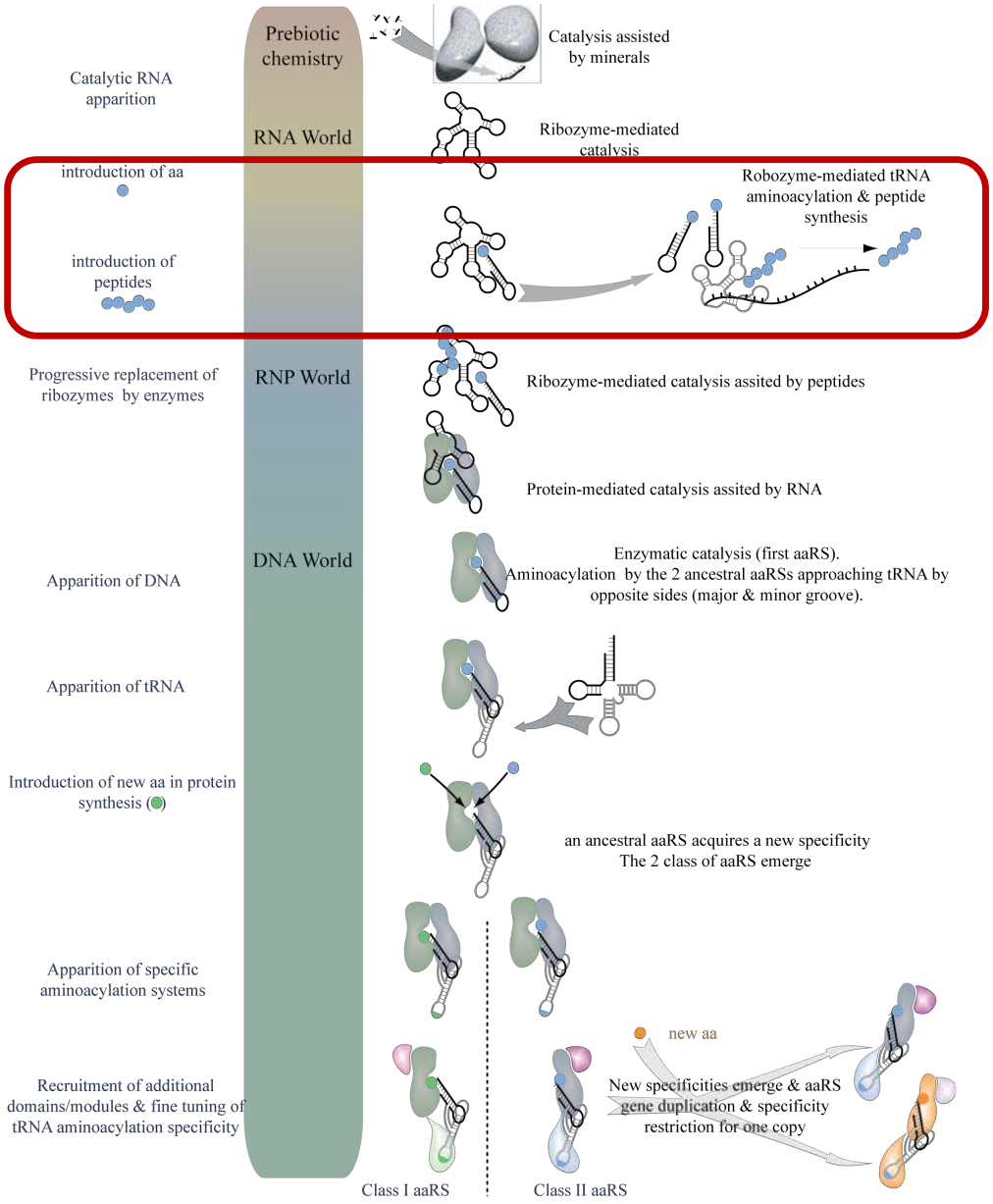


Protein synthesis catalyzed by ribozymes using aa-tRNAs as cofactors (Szathmary; 1999). Peptidyl-transferase ribozyme that bind in a sequence-specific manner RNAs with aa co-factors. The first catalytic dipeptides are made which can bind to other aptamers or assist ribozymes increasing the composition and size of the peptides that are made

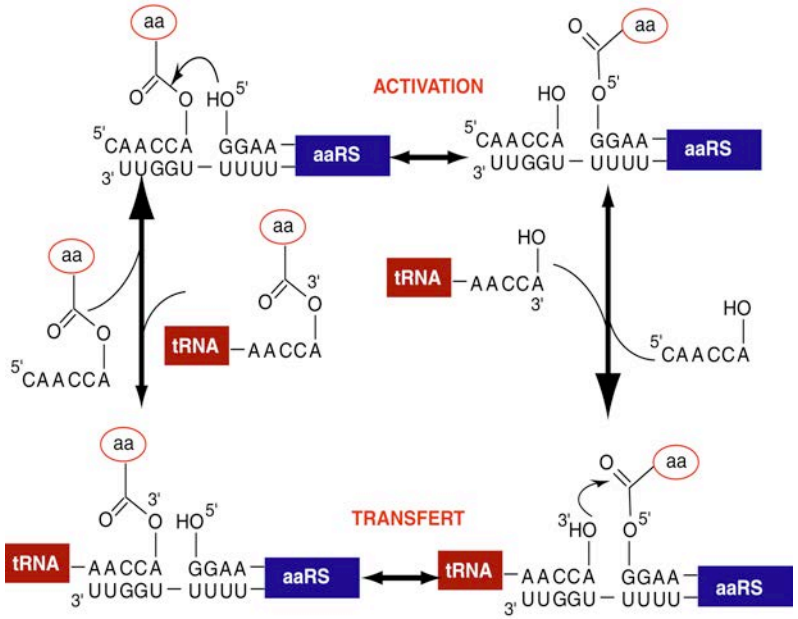
Flexizyme: a ribozyme that aminoacylates tRNAs (work by H. Suga's group)



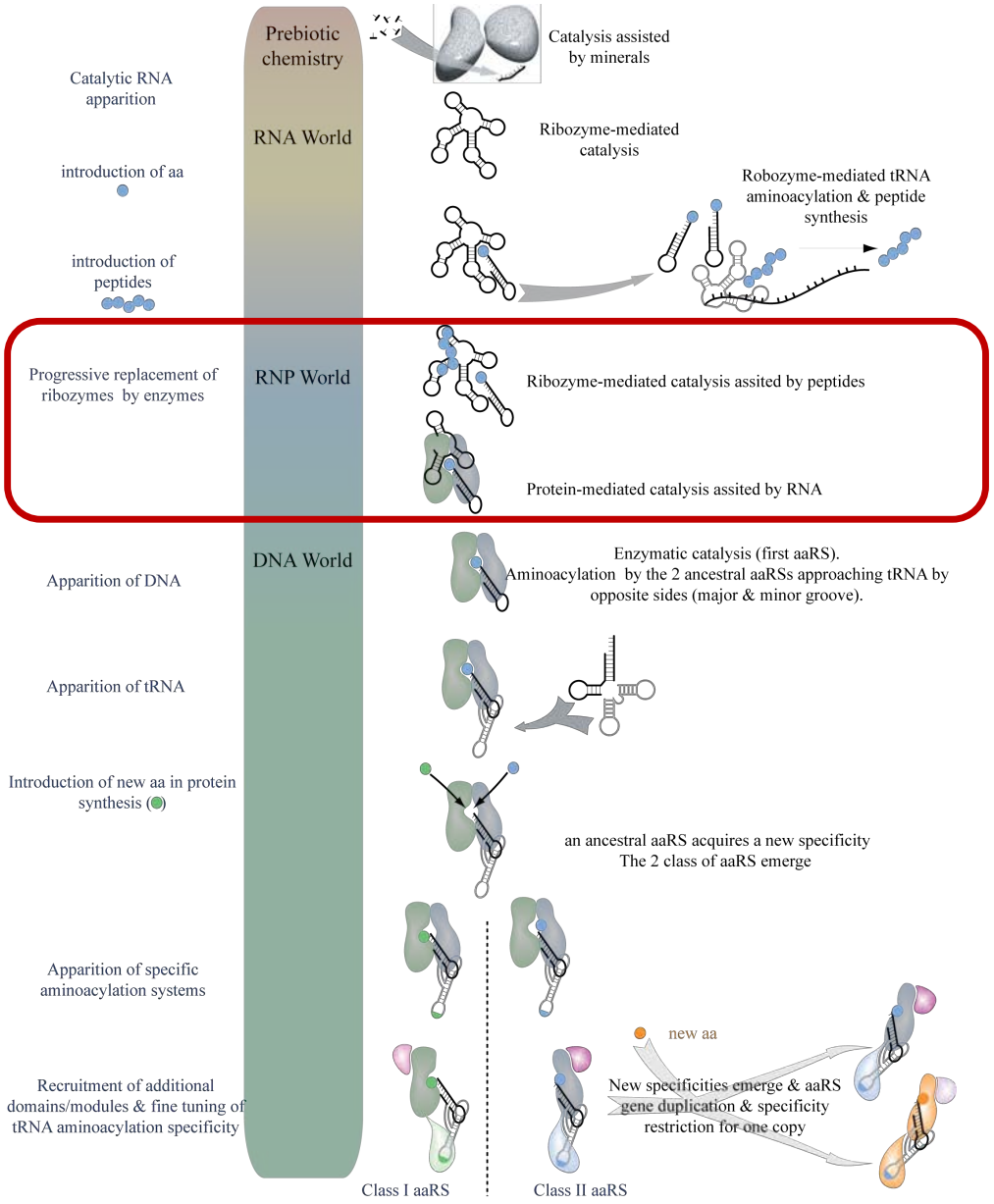
Origin of the Genetic code & RNA world



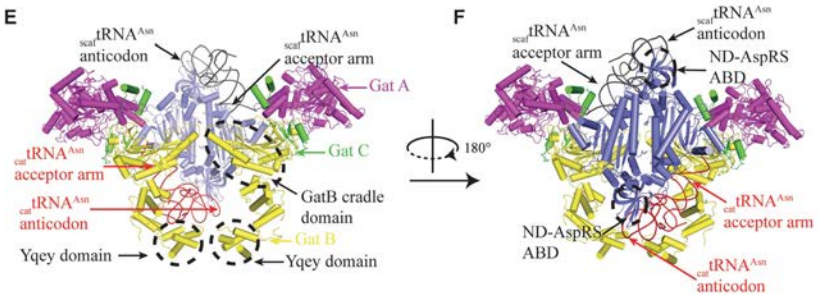
Flexizyme's catalytic mechanism



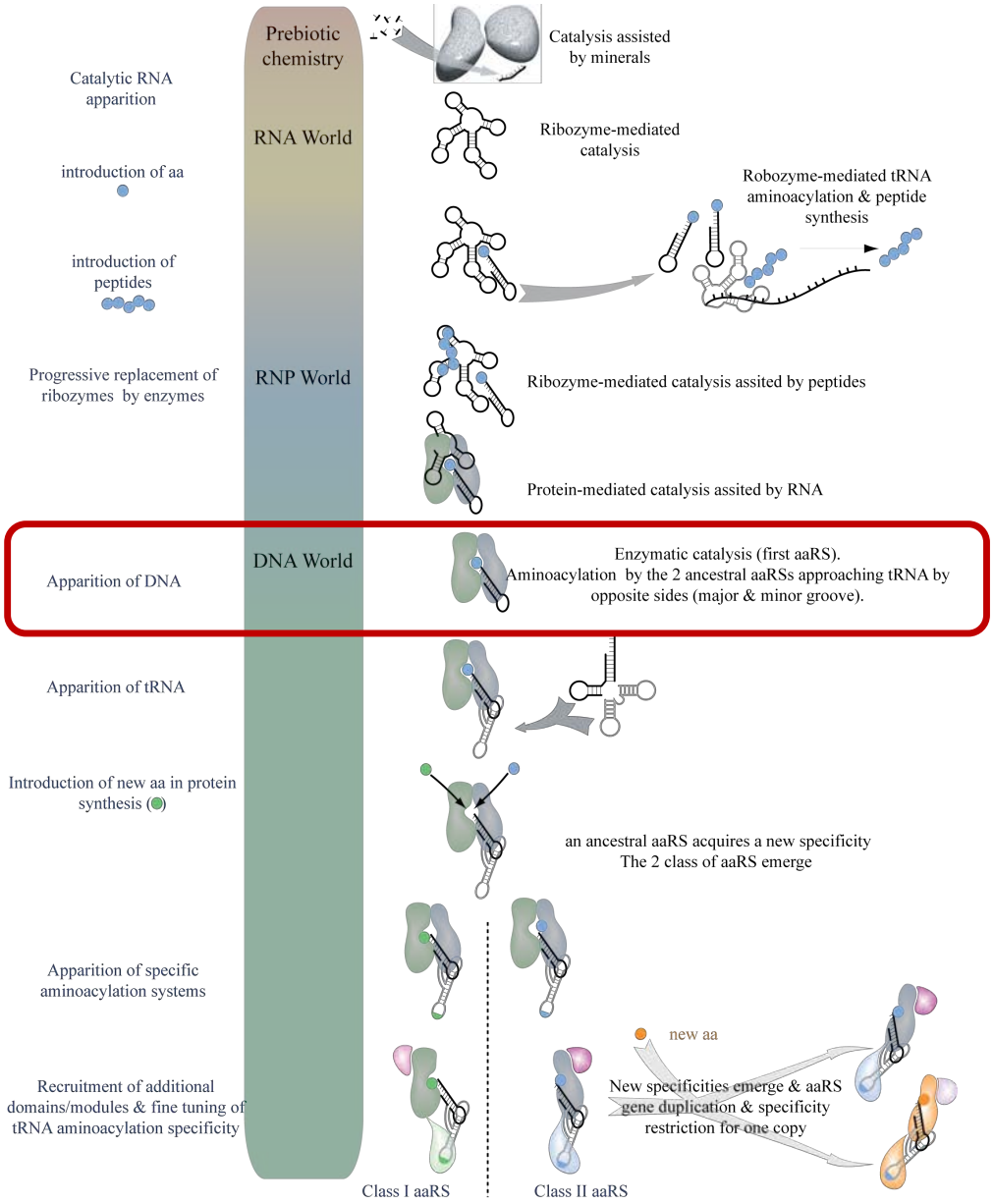
Origin of the Genetic code & RNA world



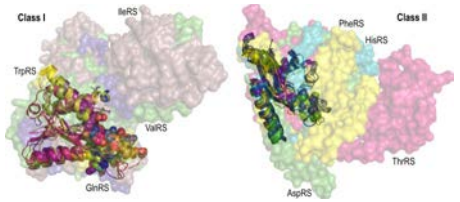
See part on introduction of Asn into the genetic code for which one step was RNP-mediated Asn-tRNAAsn synthesis catalyzed by a RNP in which tRNA is a scaffolding molecule

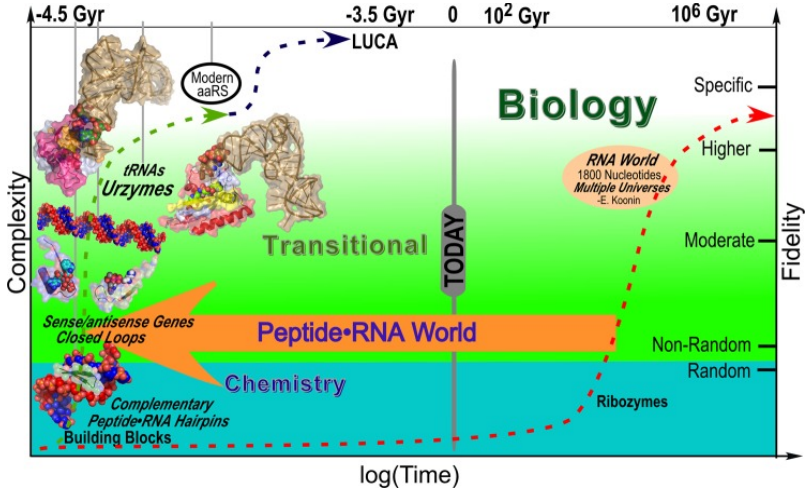


Origin of the Genetic code & RNA world



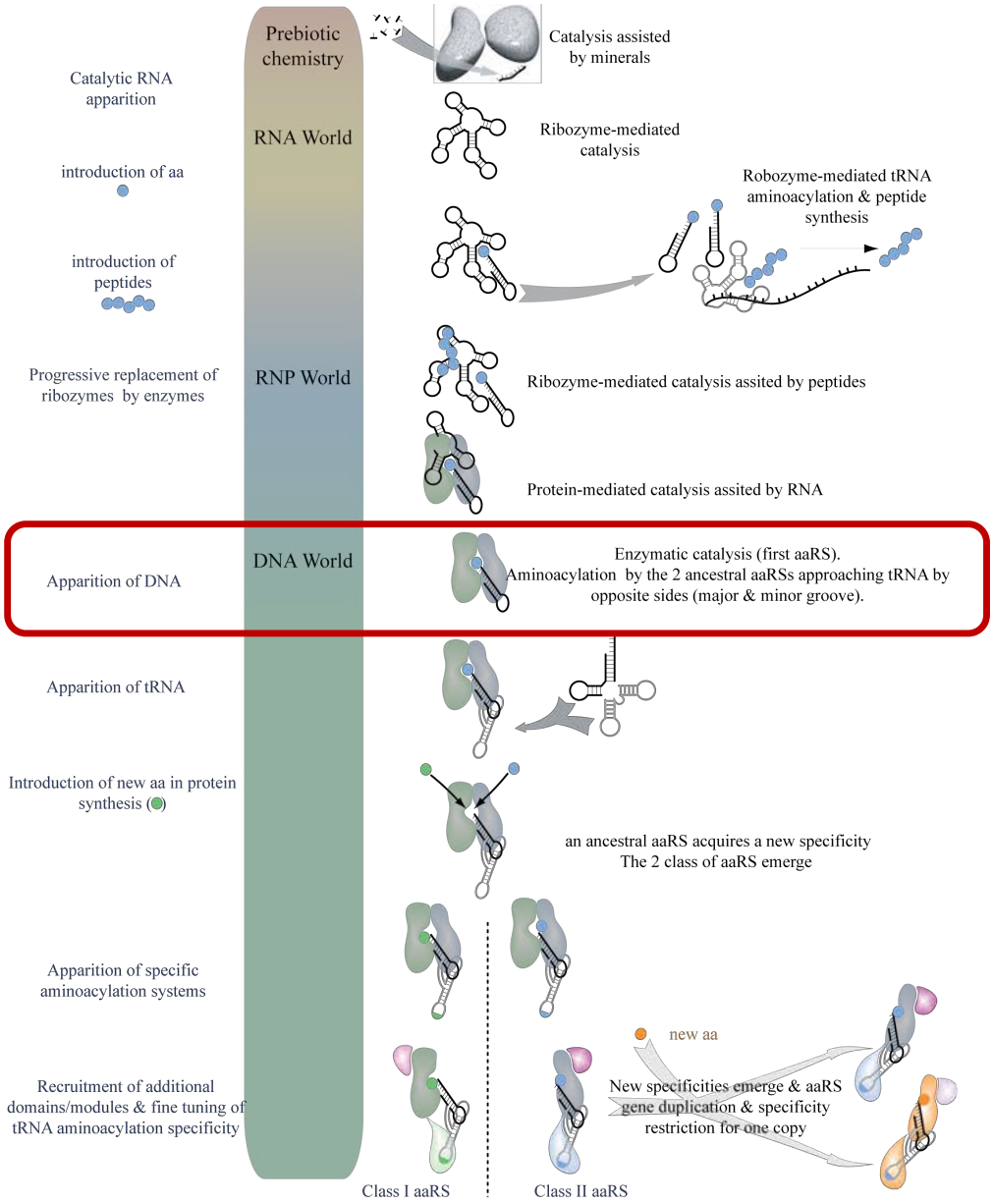
Less than the catalytic core of modern aaRSs « Urzymes » capable of aminoacylating ancestral tRNAs composed of the acceptor arm of modern tRNAs



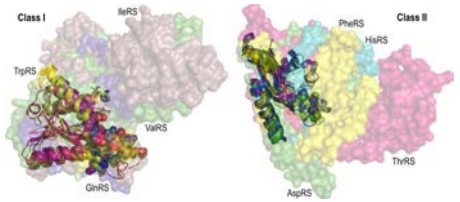


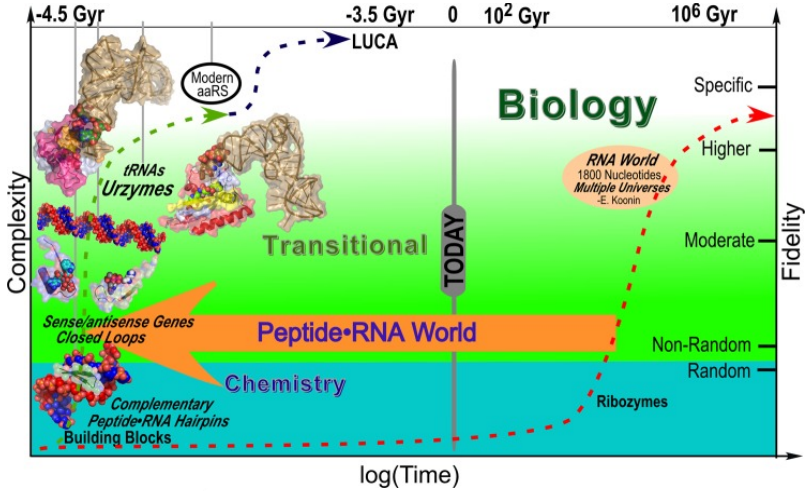
Franklyn & Carter J Biol Chem. 2013 Sep 13; 288(37): 26856–26863.

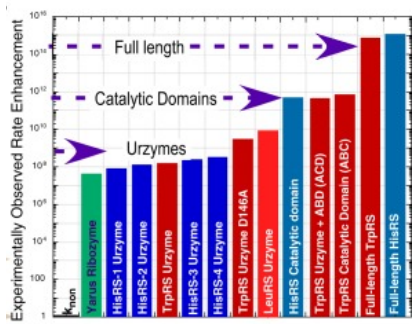
Origin of the Genetic code & RNA world



Less than the catalytic core of modern aaRSs « Urzymes » capable of aminoacylating ancestral tRNAs composed of the acceptor arm of modern tRNAs

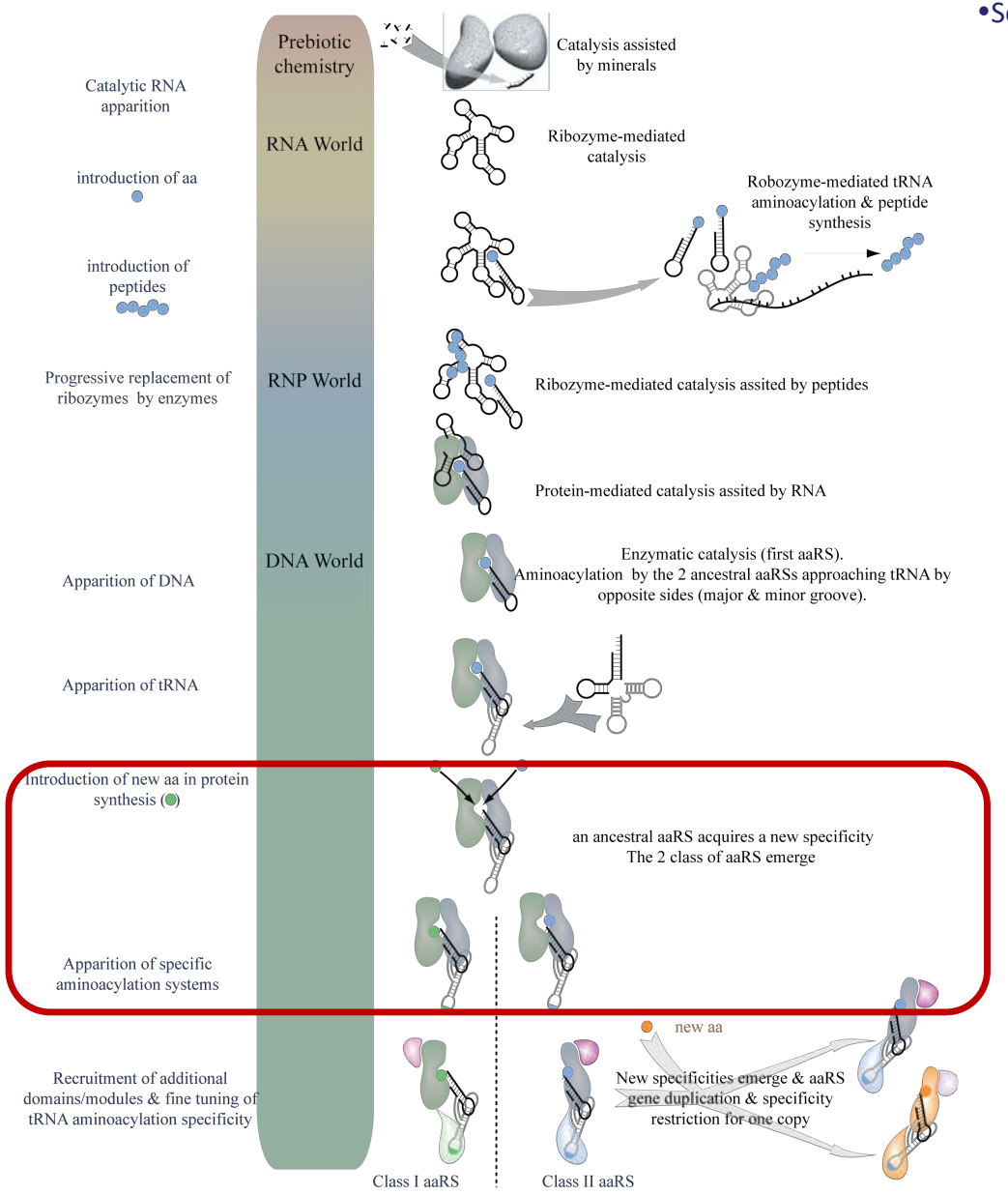






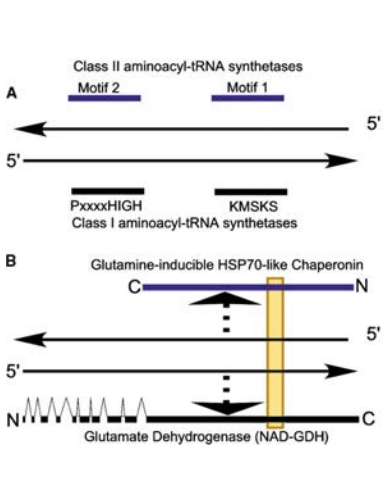
Franklyn & Carter J Biol Chem. 2013 Sep 13; 288(37): 26856–26863.

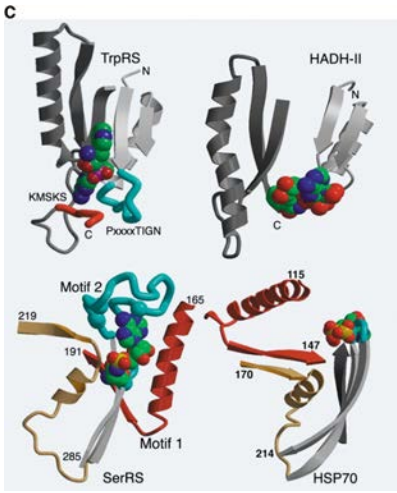
Origin of the Genetic code & RNA world



•Sense-Antisense Relationships and the aaRS Class Distinction

freshwater mold *Achlya klebsiana*



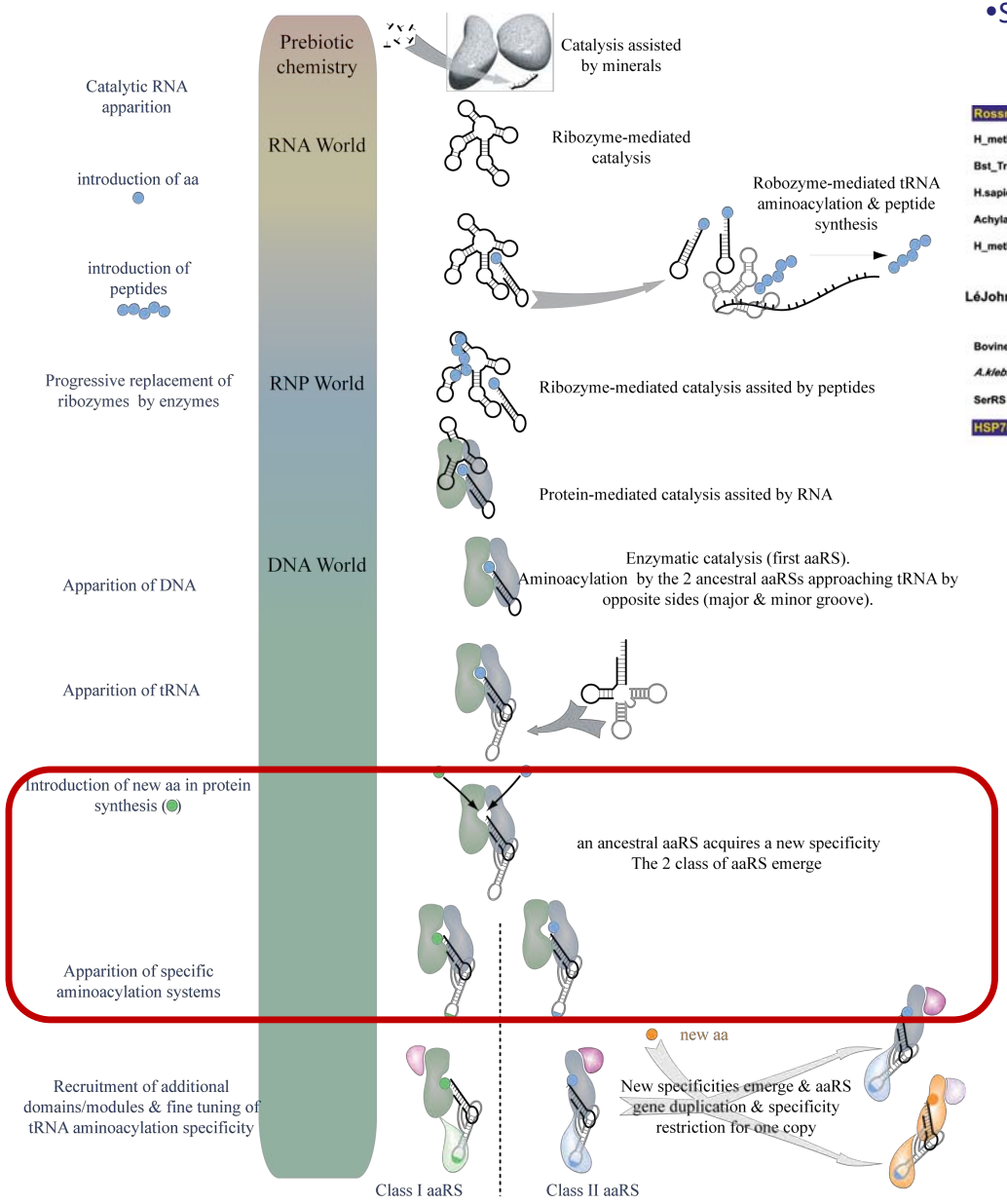


(A) Antisense coding of class I (PxxxxHIGH; KMSKS) and class II (motifs 1 and 2) aaRS catalytic motifs Rodin and Ohno 1995 S.N. Rodin and S. Ohno, *Orig. Life Evol. Biosph.* **25** (1995), pp. 565–589.

(B) Contemporary proteins coded by in-frame, antisense sequences (LéJohn et al., 1994b). The beige box identifies sequences involved in structural superpositions with class I and class II aaRS.

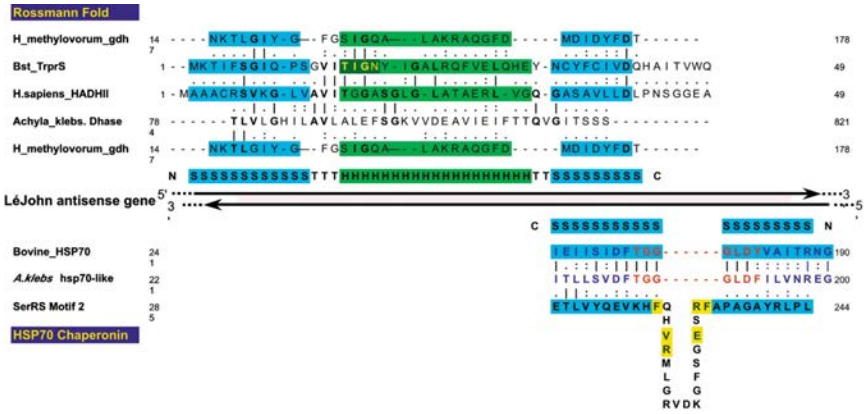
(C) Nucleotide binding sites in models of the two contemporary sense-antisense proteins (right) and corresponding fragments of classes I and II aaRS (left). Superimposed fragments (CDSFIT [CCP4, 1991]) are light gray; aaRS ATP binding signatures are cyan (motif 2 and TIGN, the TrpRS variant of HIGH) and red (motif 1 and KMSKS). The class IIa TxE signature that orients the α-amino group is in the gold-colored turn connecting the β strand and α helix.

Origin of the Genetic code & RNA world



•Sense-Antisense Relationships and the aaRS Class Distinction

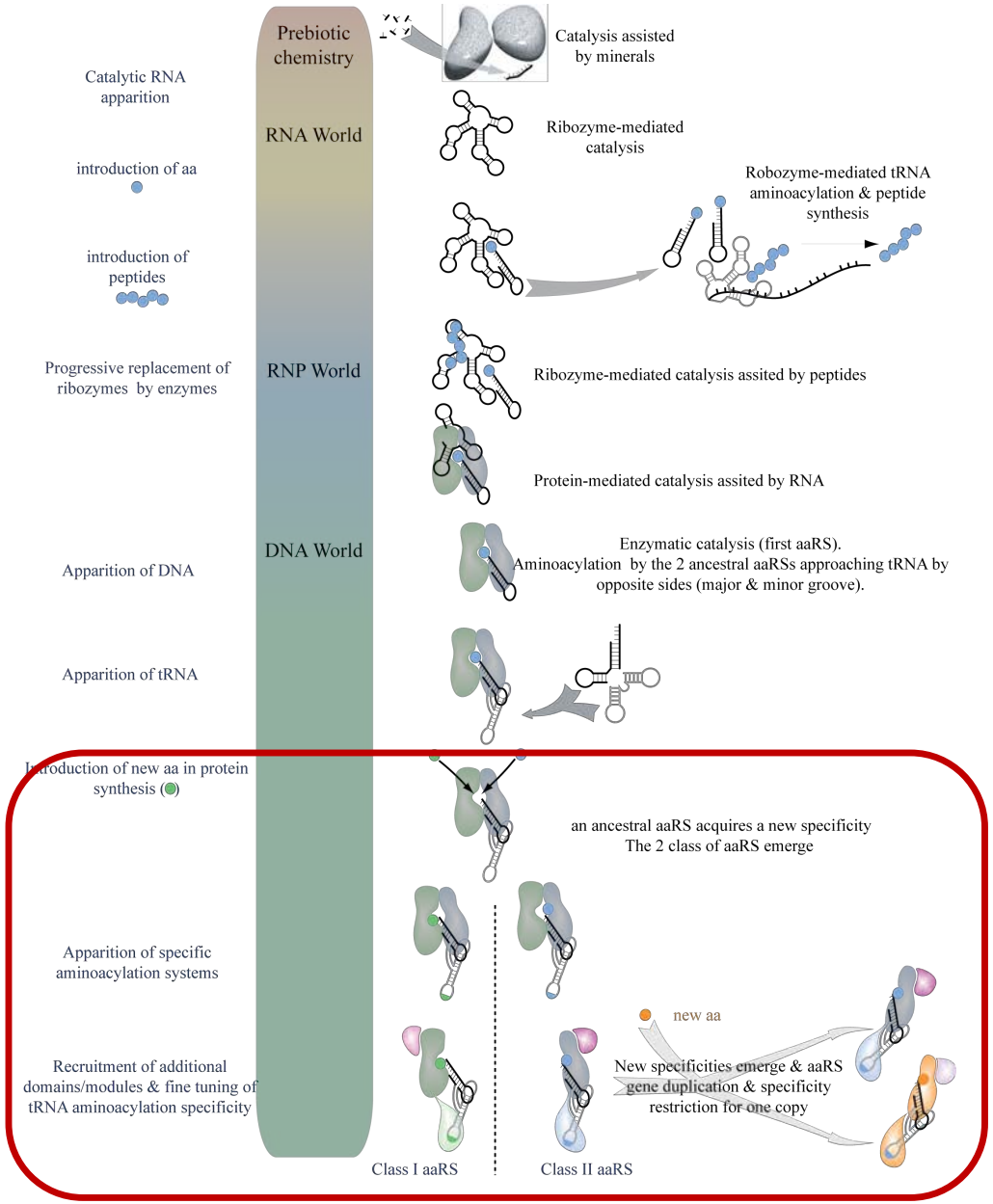
freshwater mold *Achlya klebsiana*



The *A. klebsiana* Sense-Antisense Gene Region Highlighted in Figure 1C

Secondary structures are blue for β strand and green for α helix. The DLGGGT HSP70 signature is highlighted by red letters. TIGN and motif 2 sequences are dark green and yellow, respectively. Alignments were performed using EMBOSS (Rice et al., 2000) and CLUSTALX Thompson et al. 1997 J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, *Nucleic Acids Res.* **24** (1997), pp. 4876–4882.

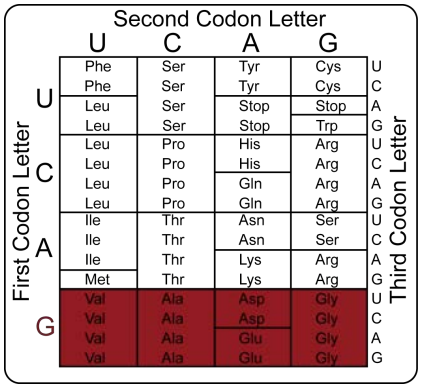
Origin of the Genetic code & RNA world



Genetic code Evolution from the aa-tRNA forming point of vue

Origin of the Genetic from the aa-tRNA forming point of vue

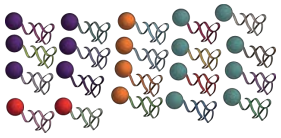
Wu, HL., Bagby, S. et van den Elsen J. M. (2005). J. Mol. Evol. 61, 54-64.



prebiotic aa

4-5 aa-tRNAs

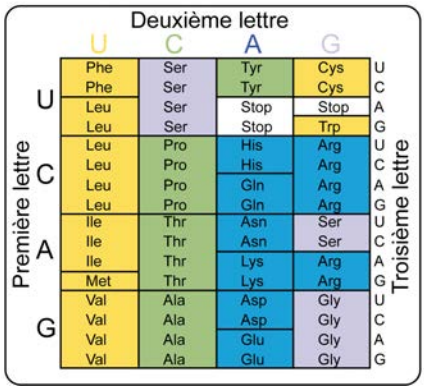
Urzymes aaRSs
Acc-arm tRNAs

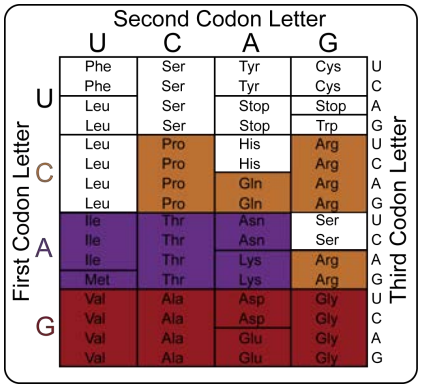


One-letter code

Origin of the Genetic from the aa-tRNA forming point of vue

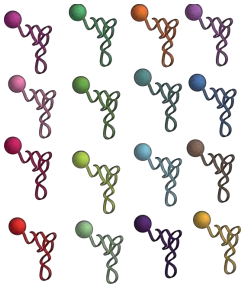
Wu, HL., Bagby, S. et van den Elsen J. M. (2005). J. Mol. Evol. 61, 54-64.

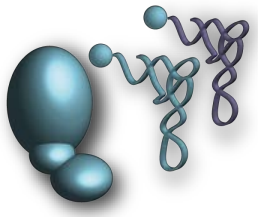




15 aa

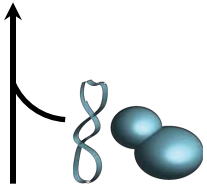
15 aa-tRNA + 1 STOP



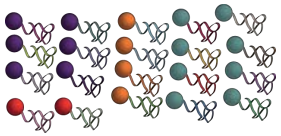


Binary code

additional domains

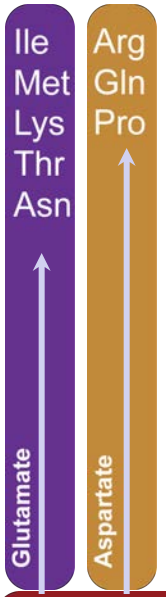


4-5 aa-tRNAs



Urzymes aaRSs
Acc-arm tRNAs

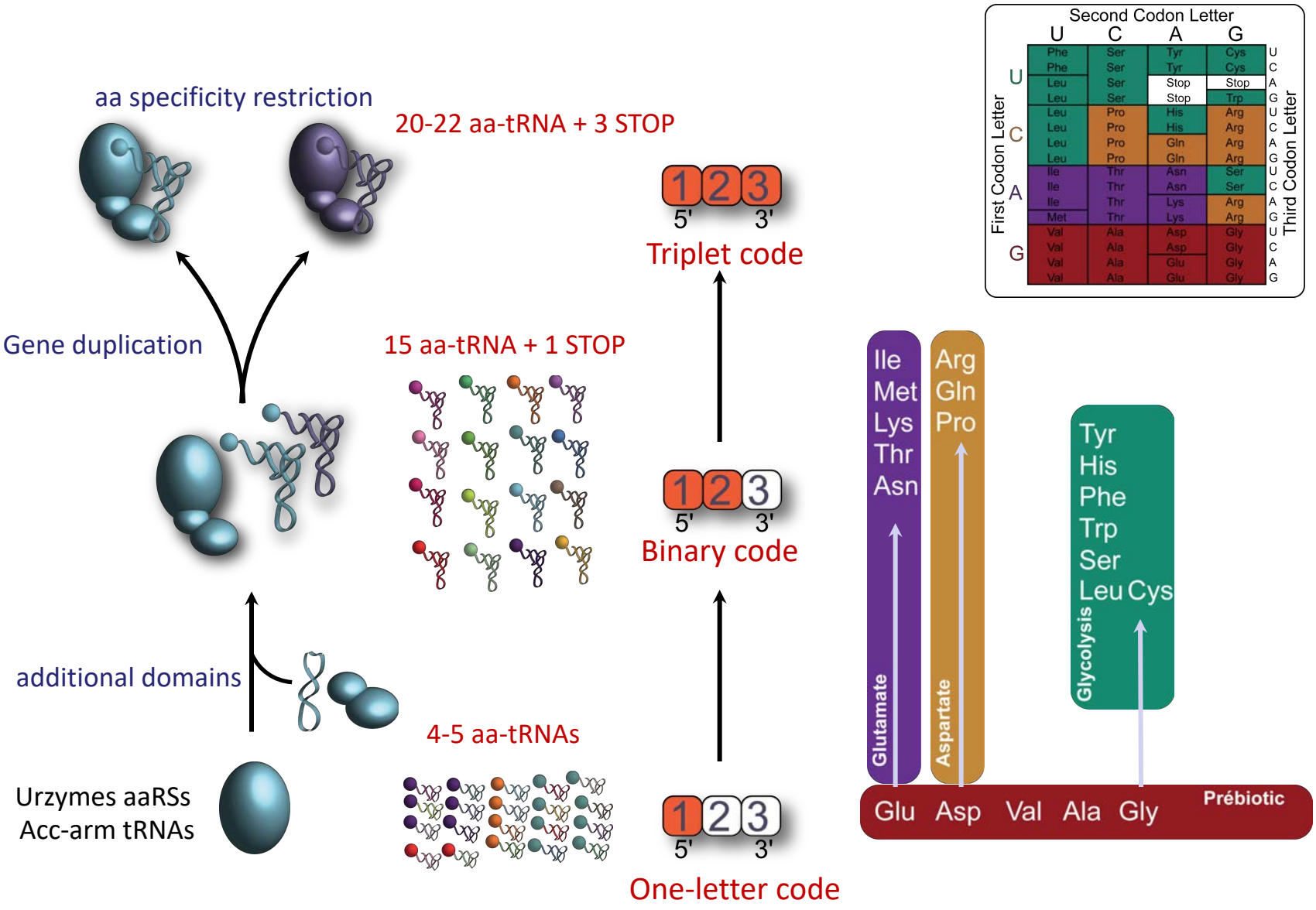
One-letter code



Limlited pool of prebiotic aa

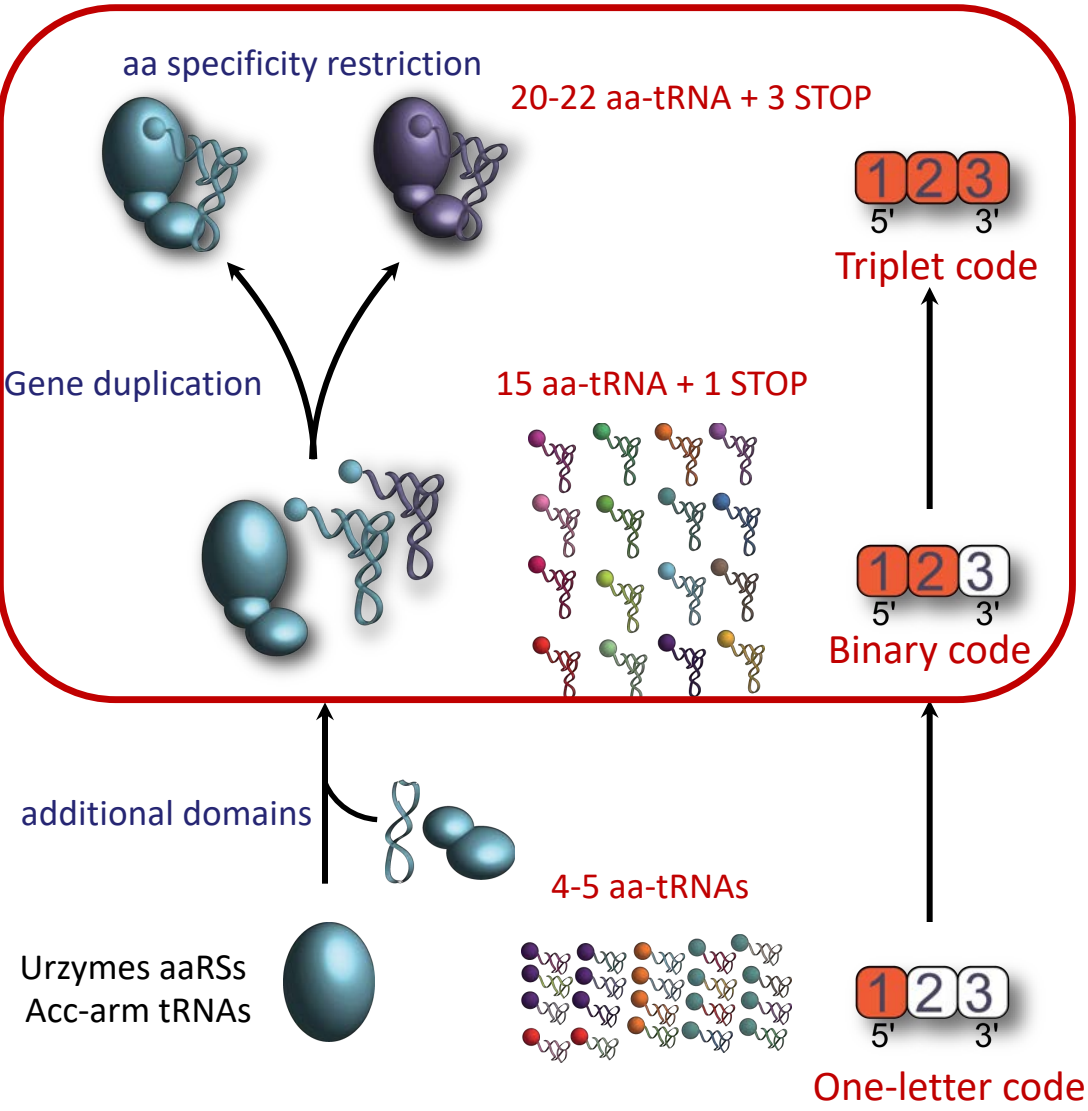
Origin of the Genetic from the aa-tRNA forming point of vue

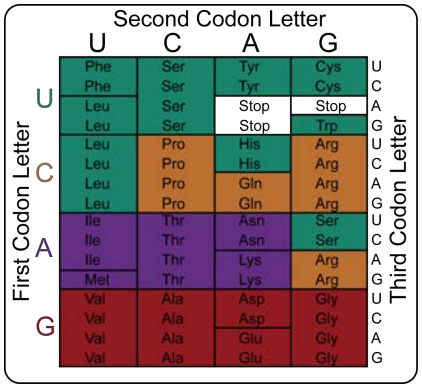
Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M. & Mirande, M. (1994). Proc. Natl. Acad. Sci. U.S.A. 91, 8670-8674.



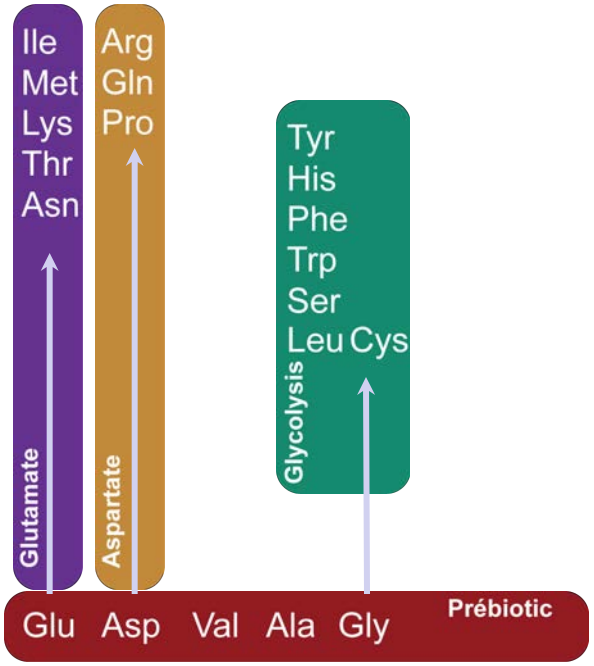
Origin of the Genetic from the aa-tRNA forming point of vue

Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M. & Mirande, M. (1994). Proc. Natl. Acad. Sci. U.S.A. 91, 8670-8674.



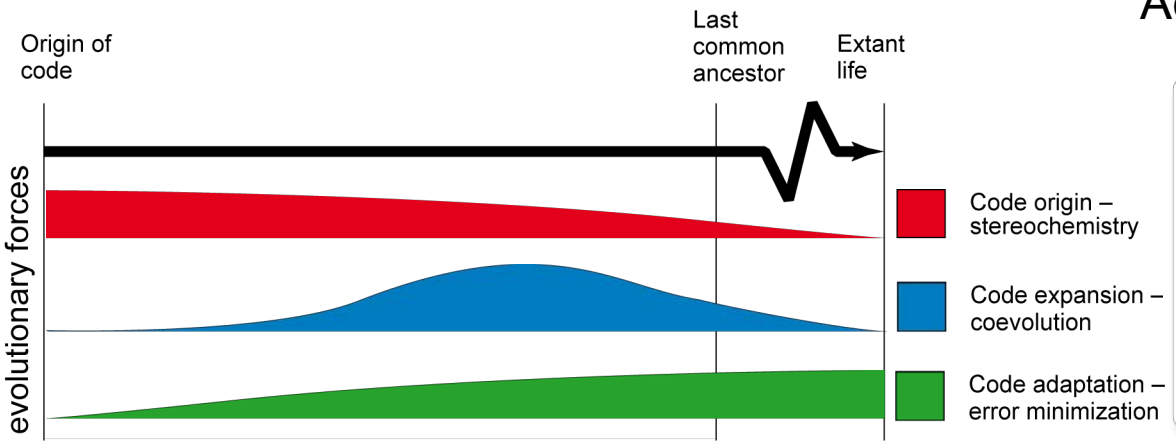


20-22 aa

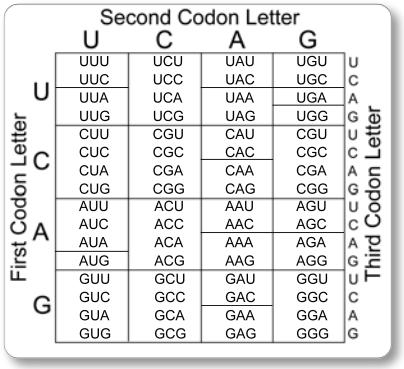


Evolutionary forces that shaped the GC – co-evolution theory

Simplier Code (One letter Code)



Actual Triplet Code

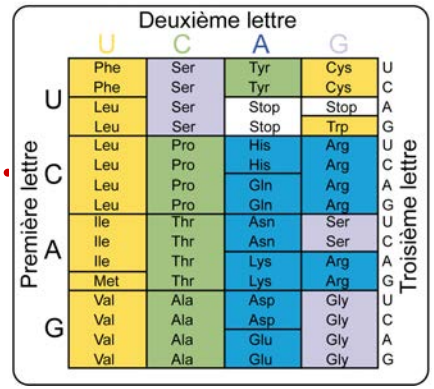


Expansion of the Genetic Code through coevolution led to codon reassignment

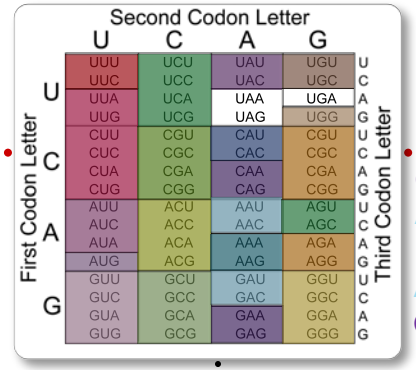
Knight, R., et al. (1999). *TIBS.* **24,** 241-247.

Evolutionary forces that shaped the GC – co-evolution theory

Few aa encoded



18 aa encoded

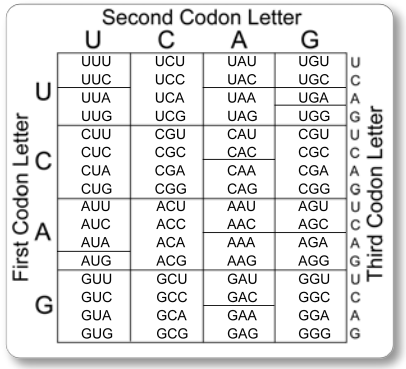


Glu
Asp

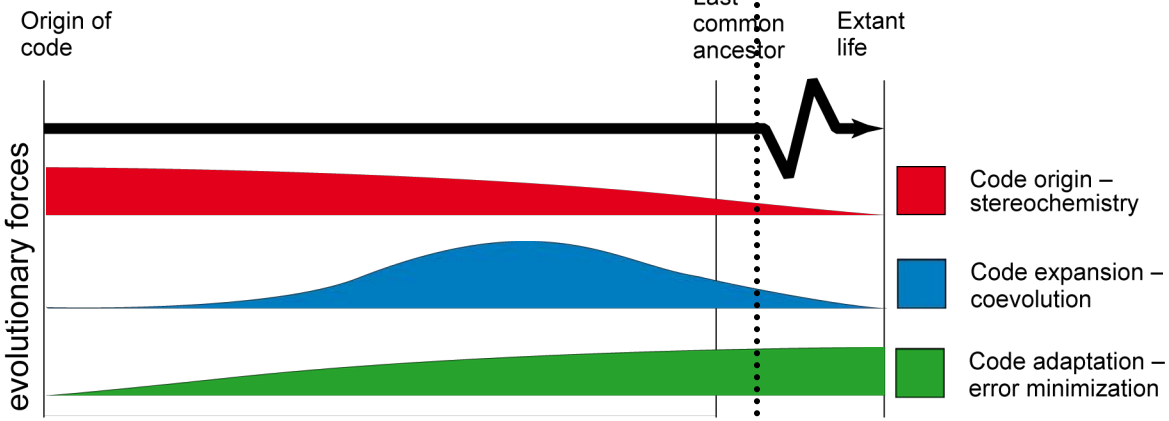
Asp
Glu

20 aa encoded

Actual Triplet Code



Simplier Code (One letter Code)



Expansion of the Genetic Code through coevolution led to codon reassignment

Knight, R., et al. (1999). *TIBS.* **24,** 241-247.

Evolutionary forces that shaped the GC – co-evolution theory

Few aa encoded

18 aa encoded

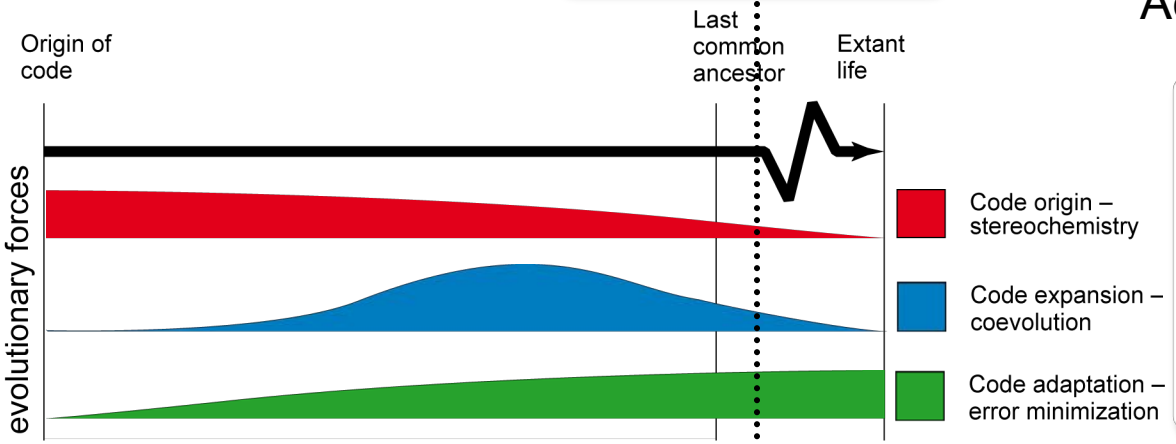
20 aa encoded

Second Codon Letter

First Codon Letter	U	C	A	G	Third Codon Letter
U	UUU	UCU	UAU	UGU	U
	UUC	UCC	UAC	UGC	C
	UUA	UCA	UAA	UGA	A
	UUG	UCG	UAG	UGG	G
C	CUU	CGU	CAU	CGU	U
	CUC	CGC	CAC	CGC	C
	CUA	CGA	CAA	CGA	A
	CUG	CGG	CAG	CGG	G
A	AUU	ACU	AAU	AGU	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA	AGA	A
	AUG	ACG	AAG	AGG	G
G	GUU	GCU	GAU	GGU	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA	GGA	A
	GUG	GCG	GAG	GGG	G

Gln
Asn
Asp
Glu

Actual Triplet Code



Simplier Code (One letter Code)

Expansion of the Genetic Code through coevolution led to codon reassignment

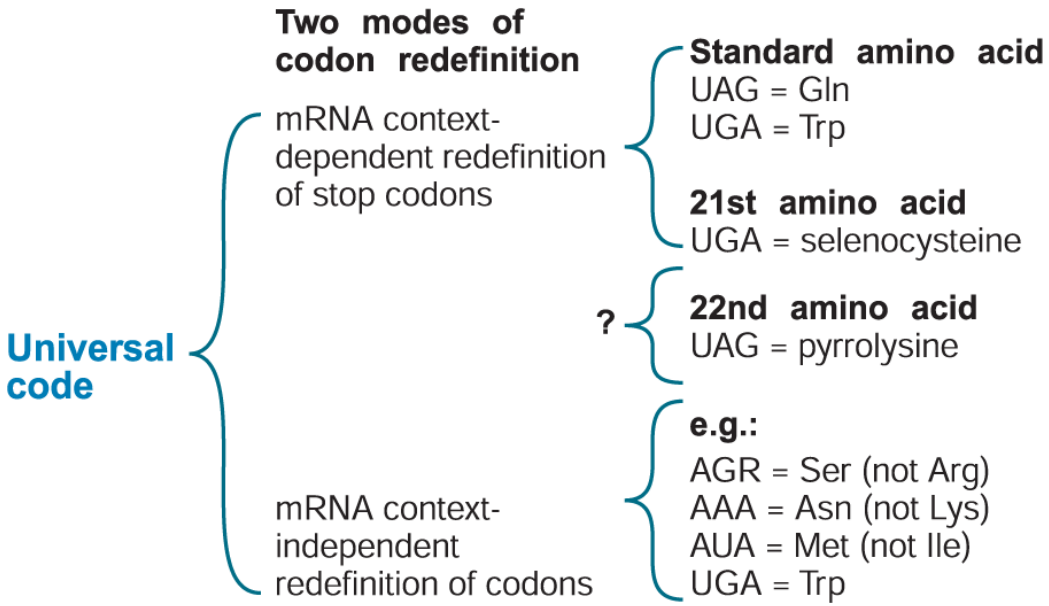
Knight, R., et al. (1999). *TIBS.* **24,** 241-247.

Natural introduction of a new aa into the Genetic Code

1- By infiltration of the coding capacity of other codons

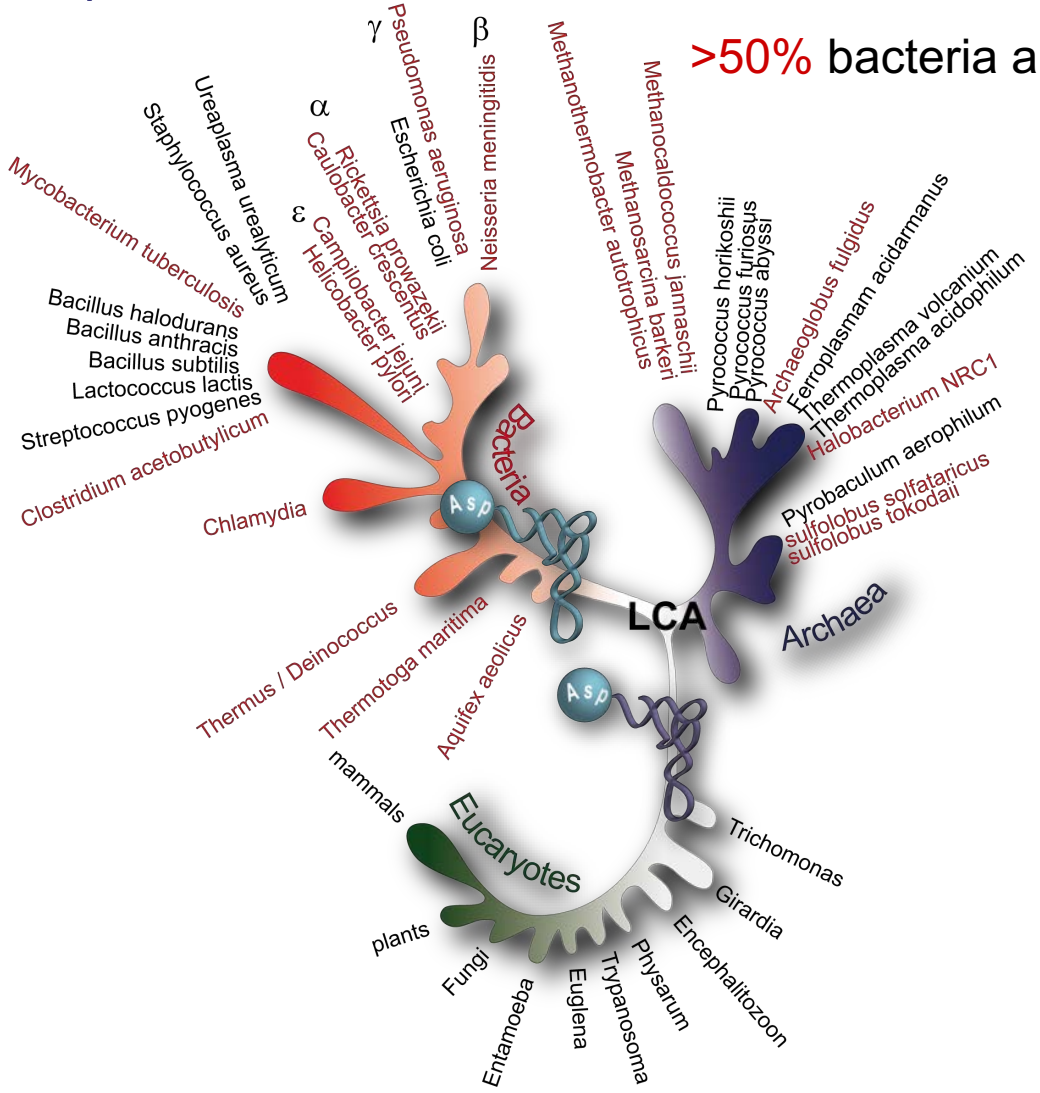
2- By redifinition of a STOP codon

- Context-independent redefinition
- Context-dependent redefinition



By infiltration of the coding capacity of other codons: Asp → Asn

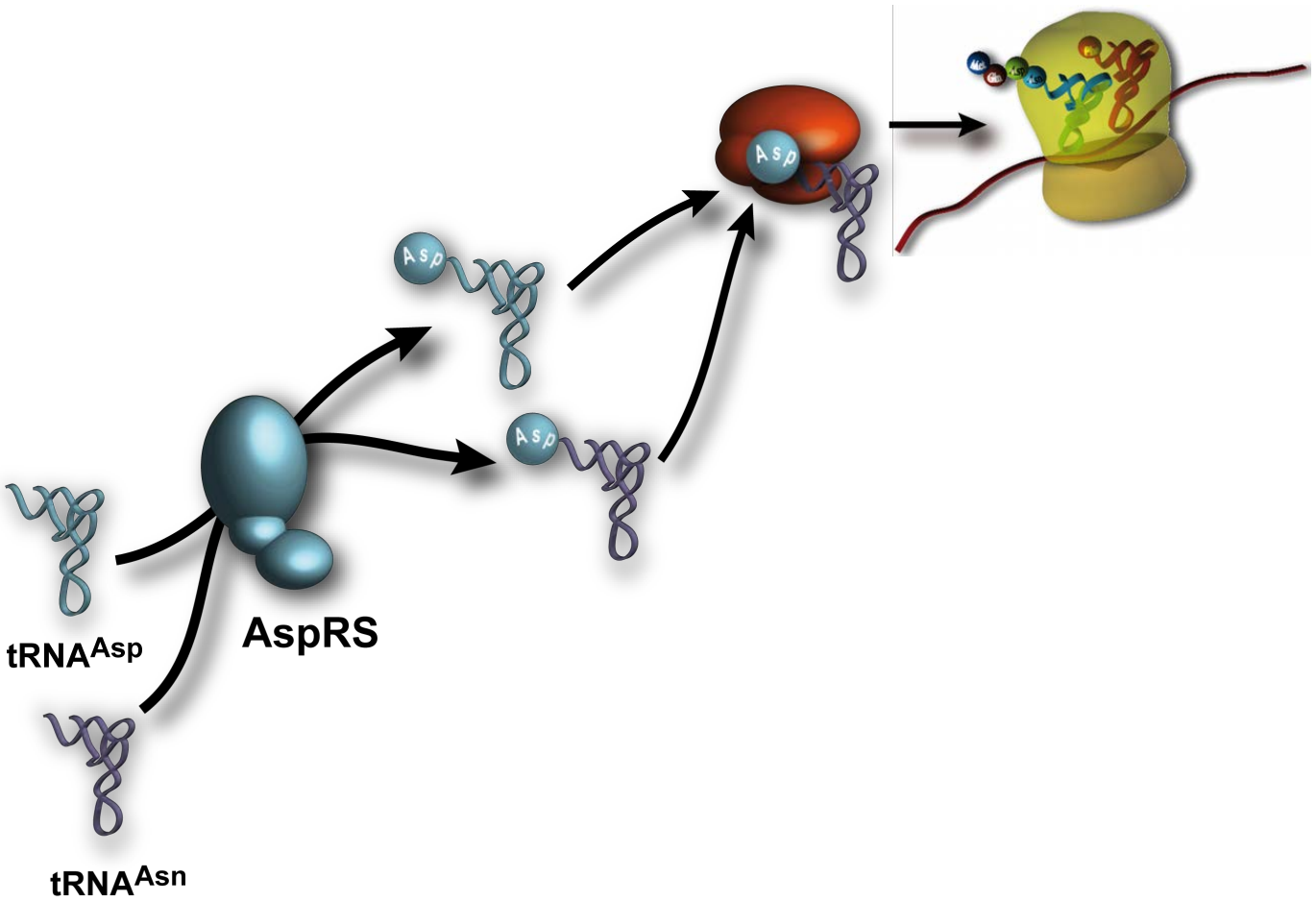
Step 1: AAU and AAC codons were translated by Asp



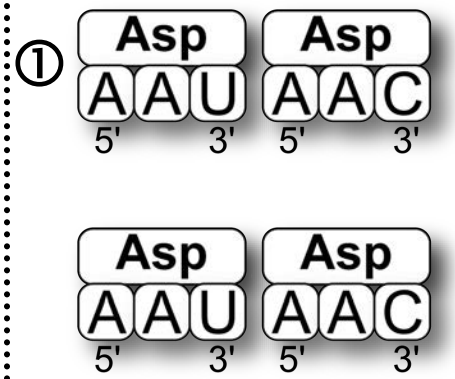
Step 1: AAU and AAC codons were translated by Asp

>50% bacteria are lacking AsnRS

They all display a Nondiscriminating AspRS



AAU and AAC Coding Capacity



Becker, H. D. & Kern, D. (1998). Proc. Natl. Acad. Sci. U.S.A. 95, 12832-12837.

Step 1: AAU and AAC codons were translated by Asp

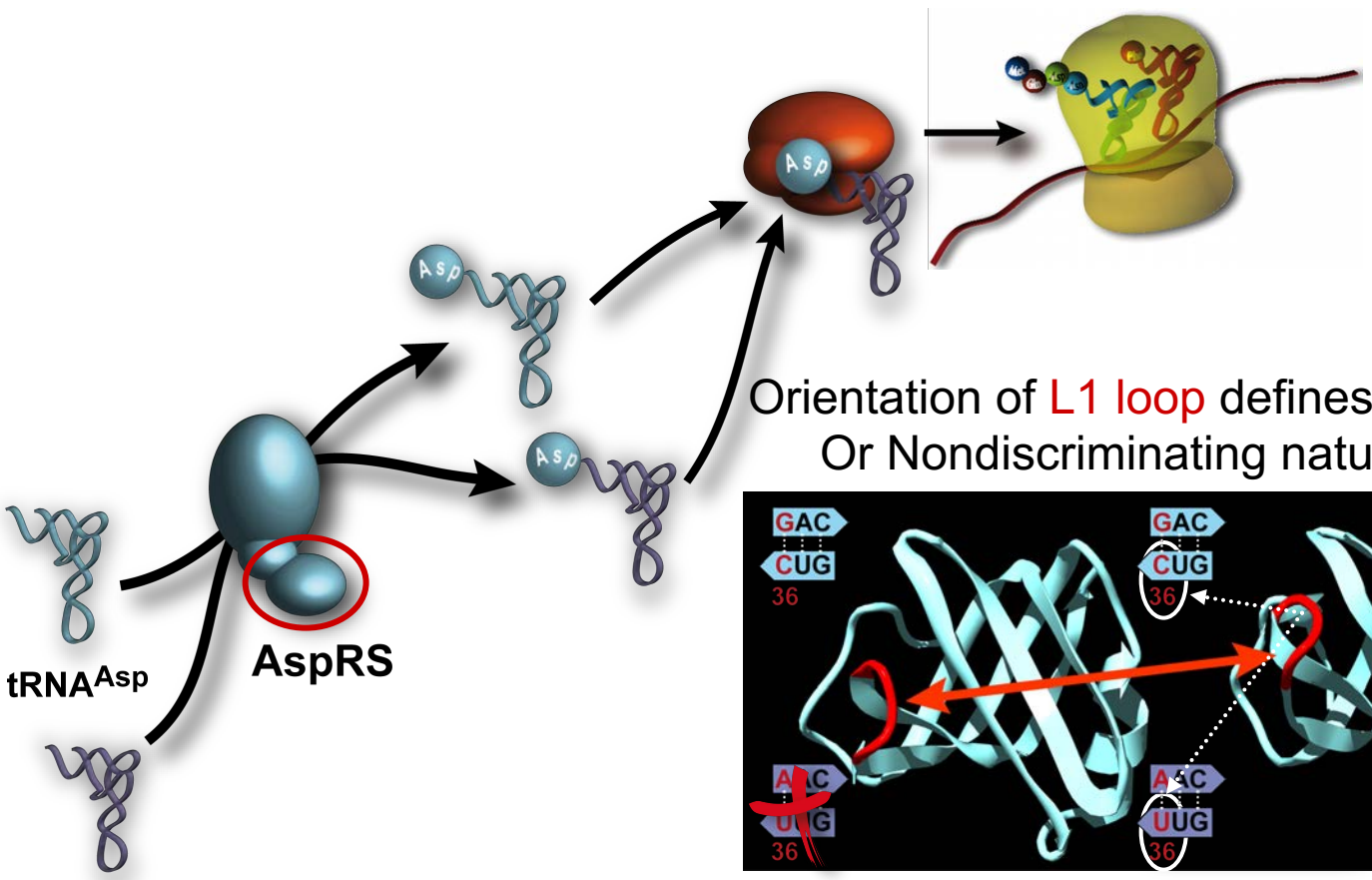
>50% bacteria are lacking AsnRS

They all display a Nondiscriminating AspRS

AAU and AAC Coding Capacity

① Asp Asp
A A U A A C
5' 3' 5' 3'

Asp Asp
A A U A A C
5' 3' 5' 3'



Orientation of L1 loop defines the Discriminating Or Nondiscriminating nature of the AspRS

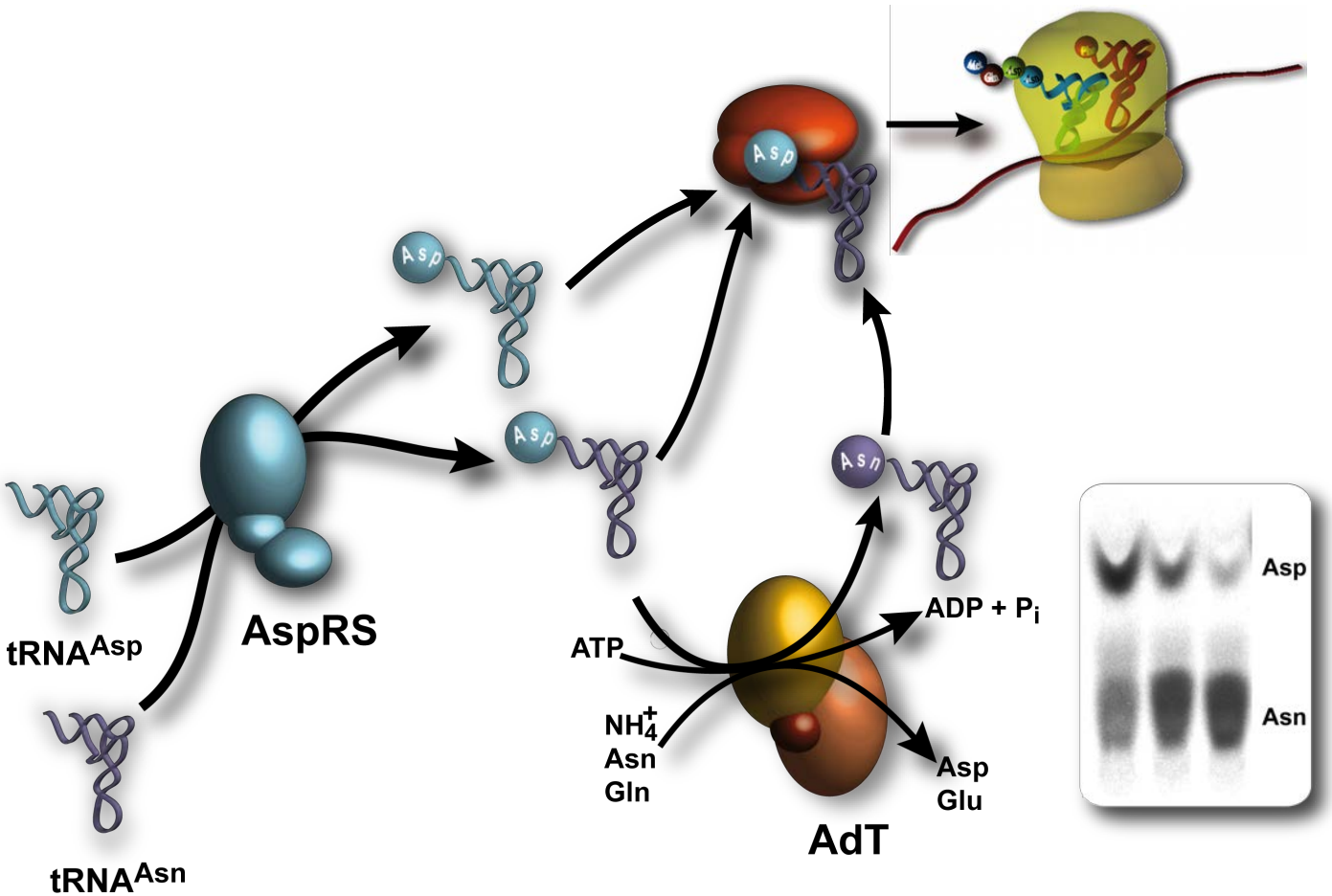
Charron, C. et al. (2003) EMBO J. 22, 1632-1643

Becker, H. D. & Kern, D. (1998). Proc. Natl. Acad. Sci. U.S.A. 95, 12832-12837.

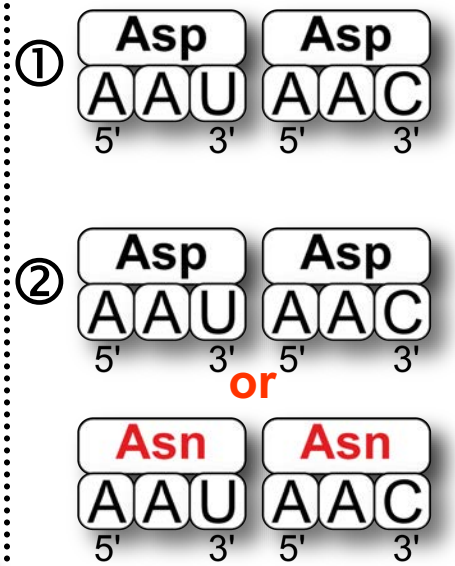
Step 2: AAU and AAC codons were translated by Asp or Asn

>50% bacteria are lacking AsnRS

They all display a Nondiscriminating AspRS

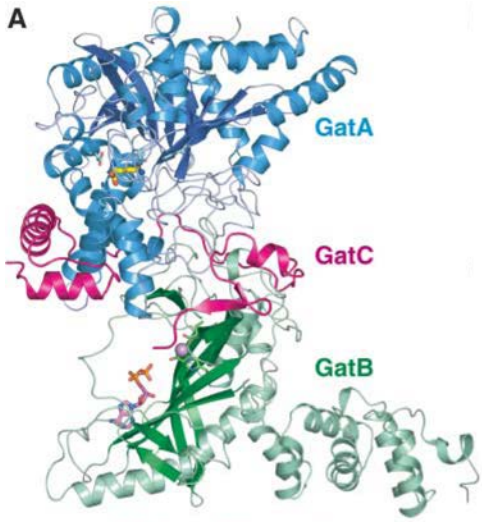


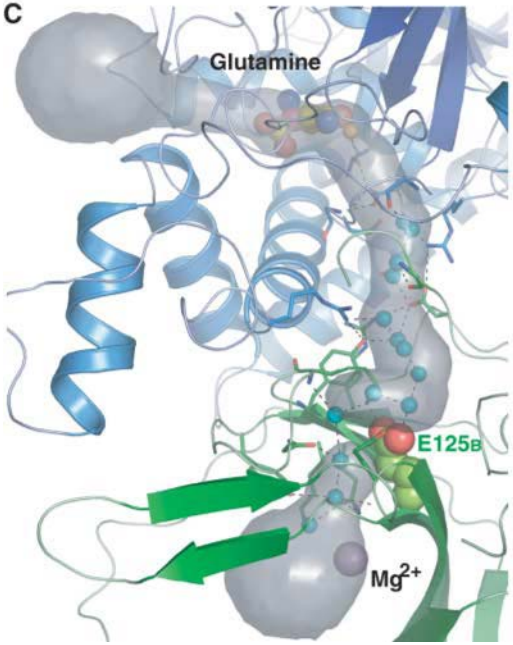
AAU and AAC Coding Capacity



Curnow, A. W., Ibba, M. & Söll, D. (1996). Nature 382, 589-590.
Becker, H. D. & Kern, D. (1998). Proc. Natl. Acad. Sci. U.S.A. 95, 12832-12837.

Step 2: tRNA-dependent amidotransferases (AdT)

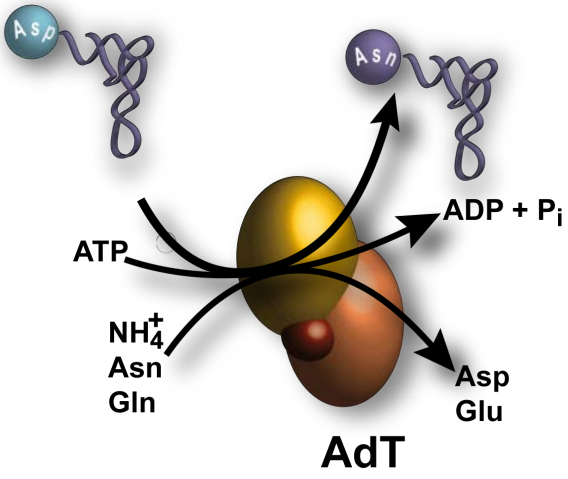


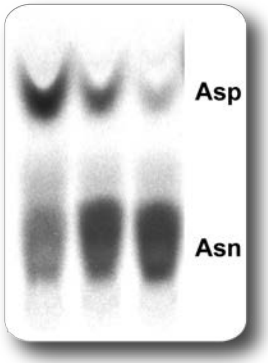


Gat Subunit function

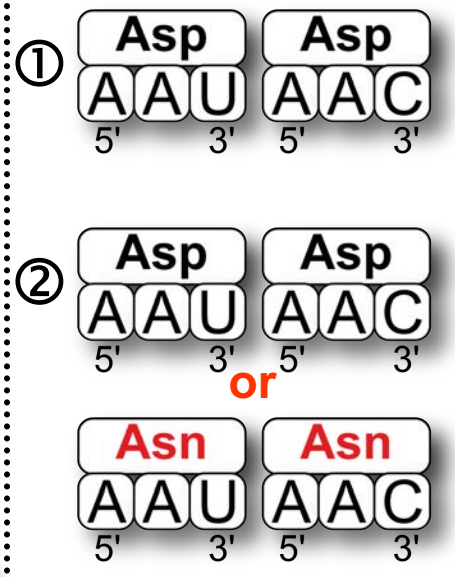
GatA	GGSSGGSAAAhA.....hGpDTGGShRpEA (Amidases signature); G....GKS (P-Loop)	1/2 Catalytic subunit; Generates NH_2
GatB	hD.NR.G.PIhEIVp.P (pet-112 signature)	1/2 Catalytic subunit; Binds aa-tRNA
GatC		Chaperone?

- GatA: generates ammonia that enters a 30 Å-long channel to reach the transamidation site located in GatB
- GatC: a ring that keeps GatA connected to GatB



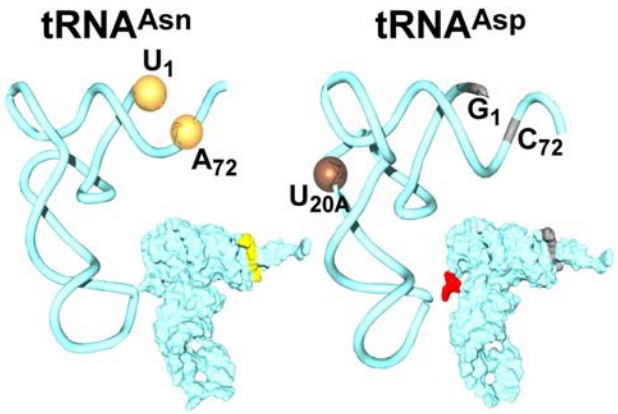


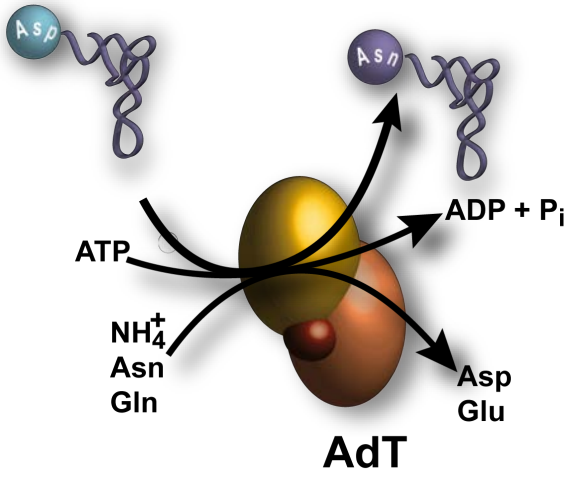
AAU and AAC Coding Capacity



Nakamura, A. et al. (2006). *Science* **312,** 1954-1958

Step 2: tRNA-dependent amidotransferases (AdT)





AAU and AAC Coding Capacity

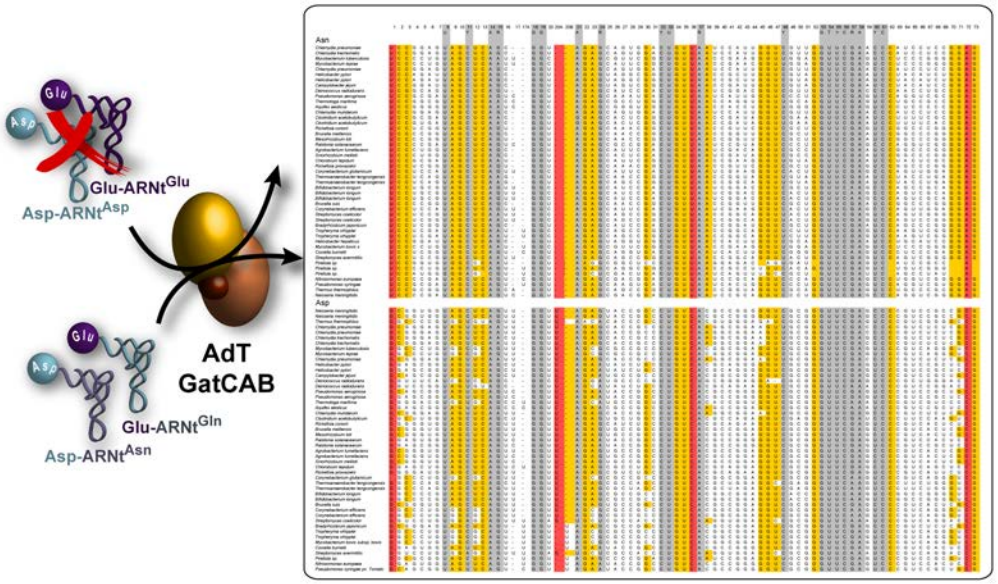
① Asp Asp
A A U A A C
5' 3' 5' 3'

② Asp Asp
A A U A A C
5' 3' 5' 3'
or
Asn Asn
A A U A A C
5' 3' 5' 3'

How does the AdT distiguish between Asp-tRNA^{Asn} and Asp-tRNA^{Asp} ?

Nakamura, A. et al. (2006). *Science* **312,** 1954-1958

Step 2: tRNA-dependent amidotransferases (AdT)



AAU and AAC Coding Capacity

① Asp AAU 5' 3' Asp AAC 5' 3'

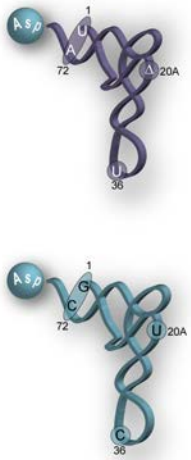
② Asp AAU 5' 3' Asp AAC 5' 3'

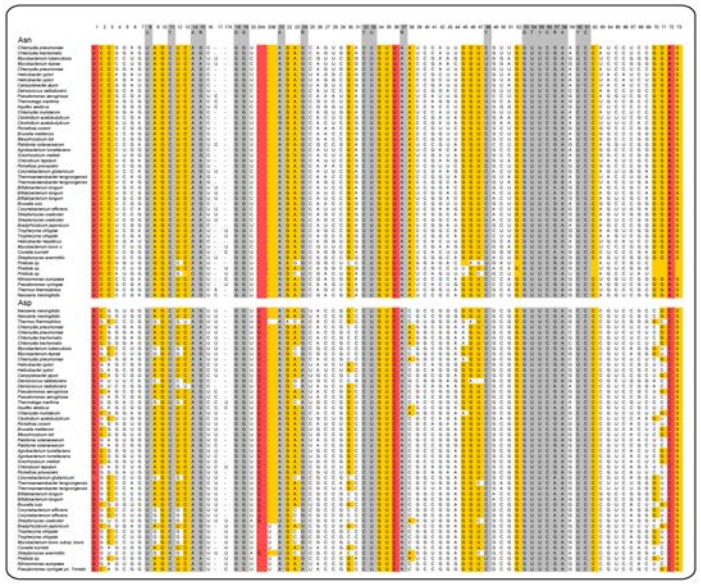
or

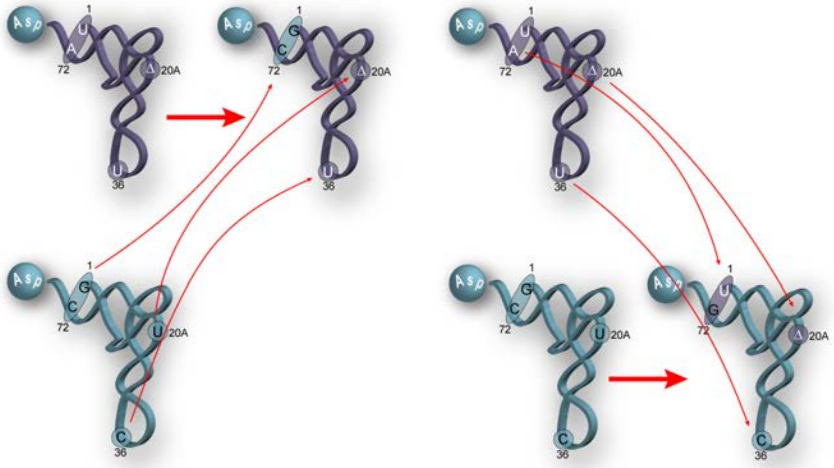
Asn AAU 5' 3' Asn AAC 5' 3'

How does the AdT distiguish between Asp-tRNA^{Asn} and Asp-tRNA^{Asp} ?

Step 2: tRNA-dependent amidotransferases (AdT)







AAU and AAC Coding Capacity

① Asp AAU 5' 3' Asp AAC 5' 3'

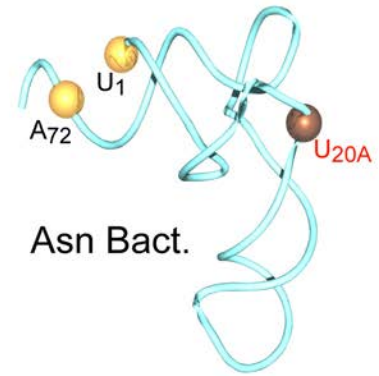
② Asp AAU 5' 3' Asp AAC 5' 3'

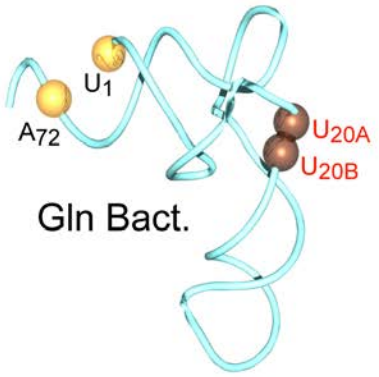
or

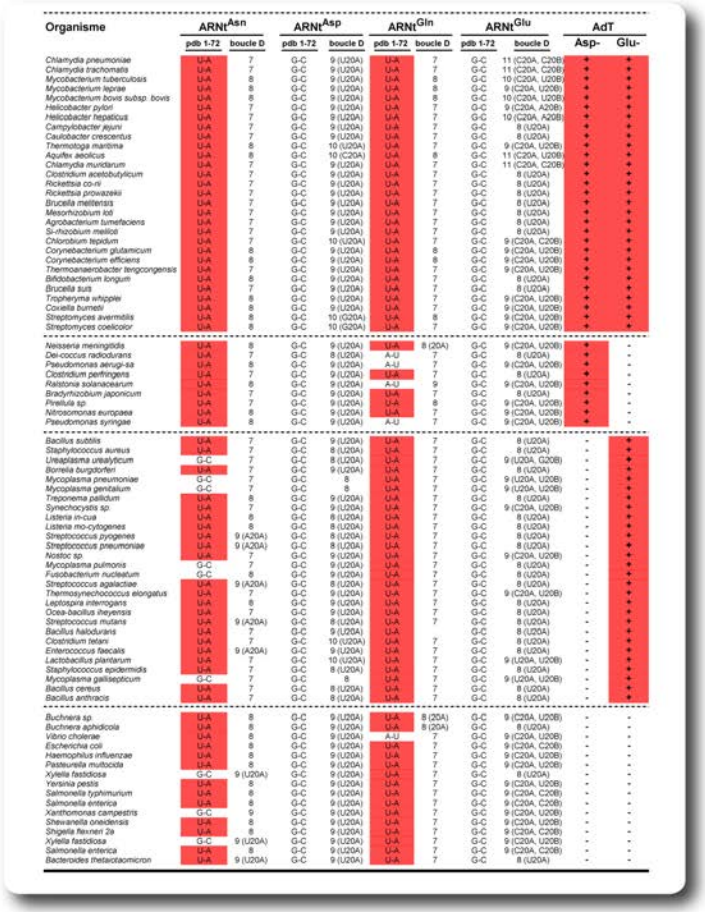
Asn AAU 5' 3' Asn AAC 5' 3'

How does the AdT distiguish between Asp-tRNA^{Asn} and Asp-tRNA^{Asp} ?

Step 2: tRNA-dependent amidotransferases (AdT)

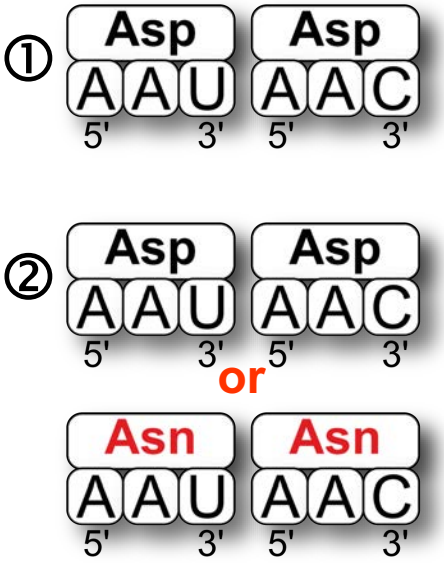






Organisme	ARNt^{Asn}		ARNt^{Asp}		ARNt^{Gln}		ARNt^{Glu}		AdT	
	pdb 1-72	boucle D	pdb 1-72	boucle D	pdb 1-72	boucle D	pdb 1-72	boucle D	Asp-	Glu-
Chlamydia pneumoniae	U-A	7	G-C	9 (U20A)	U-A	7	G-C	11 (C20A, C20B)	+	+
Chlamydia trachomatis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	11 (C20A, C20B)	+	+
Mycobacterium tuberculosis	U-A	8	G-C	9 (U20A)	U-A	8	G-C	10 (C20A, U20B)	+	+
Mycobacterium leprae	U-A	8	G-C	9 (U20A)	U-A	8	G-C	9 (C20A, U20B)	+	+
Mycobacterium bovis subsp. bovis	U-A	8	G-C	9 (U20A)	U-A	8	G-C	10 (C20A, U20B)	+	+
Helicobacter pylori	U-A	7	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, A20B)	+	+
Helicobacter hepaticus	U-A	7	G-C	9 (U20A)	U-A	7	G-C	10 (C20A, A20B)	+	+
Campylobacter jejuni	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Caulobacter crescentus	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Thermotoga maritima	U-A	8	G-C	10 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	+
Aquifex aeolicus	U-A	8	G-C	10 (C20A)	U-A	8	G-C	11 (C20A, U20B)	+	+
Chlamydia muridarum	U-A	7	G-C	9 (U20A)	U-A	7	G-C	11 (C20A, C20B)	+	+
Clostridium acetobutylicum	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Rickettsia conorii	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Rickettsia prowazekii	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Brucella melitensis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Mesorhizobium loti	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Agrobacterium tumefaciens	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Sinorhizobium meliloti	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Chlorobium tepidum	U-A	7	G-C	10 (U20A)	U-A	7	G-C	9 (C20A, C20B)	+	+
Corynebacterium glutamicum	U-A	8	G-C	9 (U20A)	U-A	8	G-C	9 (C20A, U20B)	+	+
Corynebacterium efficiens	U-A	8	G-C	9 (U20A)	U-A	8	G-C	9 (C20A, U20B)	+	+
Thermoanaerobacter tengcongensis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	+
Bifidobacterium longum	U-A	8	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Brucella suis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	+
Tropheryma whipplei	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	+
Coxiella burnetii	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	+
Streptomyces avermitilis	U-A	8	G-C	10 (U20A)	U-A	8	G-C	9 (C20A, U20B)	+	+
Streptomyces coelicolor	U-A	8	G-C	10 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	+
Neisseria meningitidis	U-A	8	G-C	9 (U20A)	U-A	8 (20A)	G-C	9 (C20A, U20B)	+	-
Deinococcus radiodurans	U-A	7	G-C	8 (U20A)	A-U	7	G-C	8 (U20A)	+	-
Pseudomonas aeruginosa	U-A	8	G-C	9 (U20A)	A-U	7	G-C	9 (C20A, U20B)	+	-
Clostridium perfringens	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	-
Ralstonia solanacearum	U-A	8	G-C	9 (U20A)	A-U	9	G-C	9 (C20A, U20B)	+	-
Bradyrhizobium japonicum	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	+	-
Pirellula sp.	U-A	7	G-C	9 (U20A)	U-A	8	G-C	9 (C20A, U20B)	+	-
Nitrosomonas europaea	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	+	-
Pseudomonas syringae	U-A	8	G-C	9 (U20A)	A-U	7	G-C	9 (C20A, U20B)	+	-
Bacillus subtilis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Staphylococcus aureus	U-A	7	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Ureaplasma urealyticum	G-C	7	G-C	8 (U20A)	U-A	7	G-C	9 (U20A, U20B)	-	+
Borrelia burgdorferi	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Mycoplasma pneumoniae	G-C	7	G-C	8	U-A	7	G-C	9 (U20A, U20B)	-	+
Mycoplasma genitalium	G-C	7	G-C	8	U-A	7	G-C	9 (U20A, U20B)	-	+
Treponema pallidum	U-A	8	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Synechocystis sp.	U-A	7	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	+
Listeria in-cua	U-A	8	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Listeria mo-cytogenes	U-A	8	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Streptococcus pyogenes	U-A	9 (A20A)	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Streptococcus pneumoniae	U-A	9 (A20A)	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Nostoc sp.	U-A	7	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	+
Mycoplasma pulmonis	G-C	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Fusobacterium nucleatum	G-C	8	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Streptococcus agalactiae	U-A	9 (A20A)	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Thermosynechococcus elongatus	U-A	7	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	+
Leptospira interrogans	U-A	8	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Ocea-bacillus iheyensis	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Streptococcus mutans	U-A	9 (A20A)	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Bacillus halodurans	U-A	7	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Clostridium tetani	U-A	7	G-C	10 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Enterococcus faecalis	U-A	9 (A20A)	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Lactobacillus plantarum	U-A	7	G-C	10 (U20A)	U-A	7	G-C	9 (U20A, U20B)	-	+
Staphylococcus epidermidis	U-A	7	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Mycoplasma gallisepticum	G-C	7	G-C	8	U-A	7	G-C	9 (U20A, U20B)	-	+
Bacillus cereus	U-A	7	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Bacillus anthracis	U-A	7	G-C	8 (U20A)	U-A	7	G-C	8 (U20A)	-	+
Buchnera sp.	U-A	8	G-C	9 (U20A)	U-A	8 (20A)	G-C	9 (C20A, U20B)	-	-
Buchnera aphidicola	U-A	8	G-C	9 (U20A)	U-A	8 (20A)	G-C	8 (U20A)	-	-
Vibrio cholerae	U-A	8	G-C	9 (U20A)	A-U	7	G-C	9 (C20A, U20B)	-	-
Escherichia coli	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, C20B)	-	-
Haemophilus influenzae	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Pasteurella multocida	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Xylella fastidiosa	G-C	9 (U20A)	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	-
Yersinia pestis	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Salmonella typhimurium	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, C20B)	-	-
Salmonella enterica	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, C20B)	-	-
Xanthomonas campestris	G-C	9	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Shewanella oneidensis	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Shigella flexneri 2a	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, C20B)	-	-
Xylella fastidiosa	G-C	9 (U20A)	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, U20B)	-	-
Salmonella enterica	U-A	8	G-C	9 (U20A)	U-A	7	G-C	9 (C20A, C20B)	-	-
Bacteroides thetaiotaomicron	U-A	9 (U20A)	G-C	9 (U20A)	U-A	7	G-C	8 (U20A)	-	-

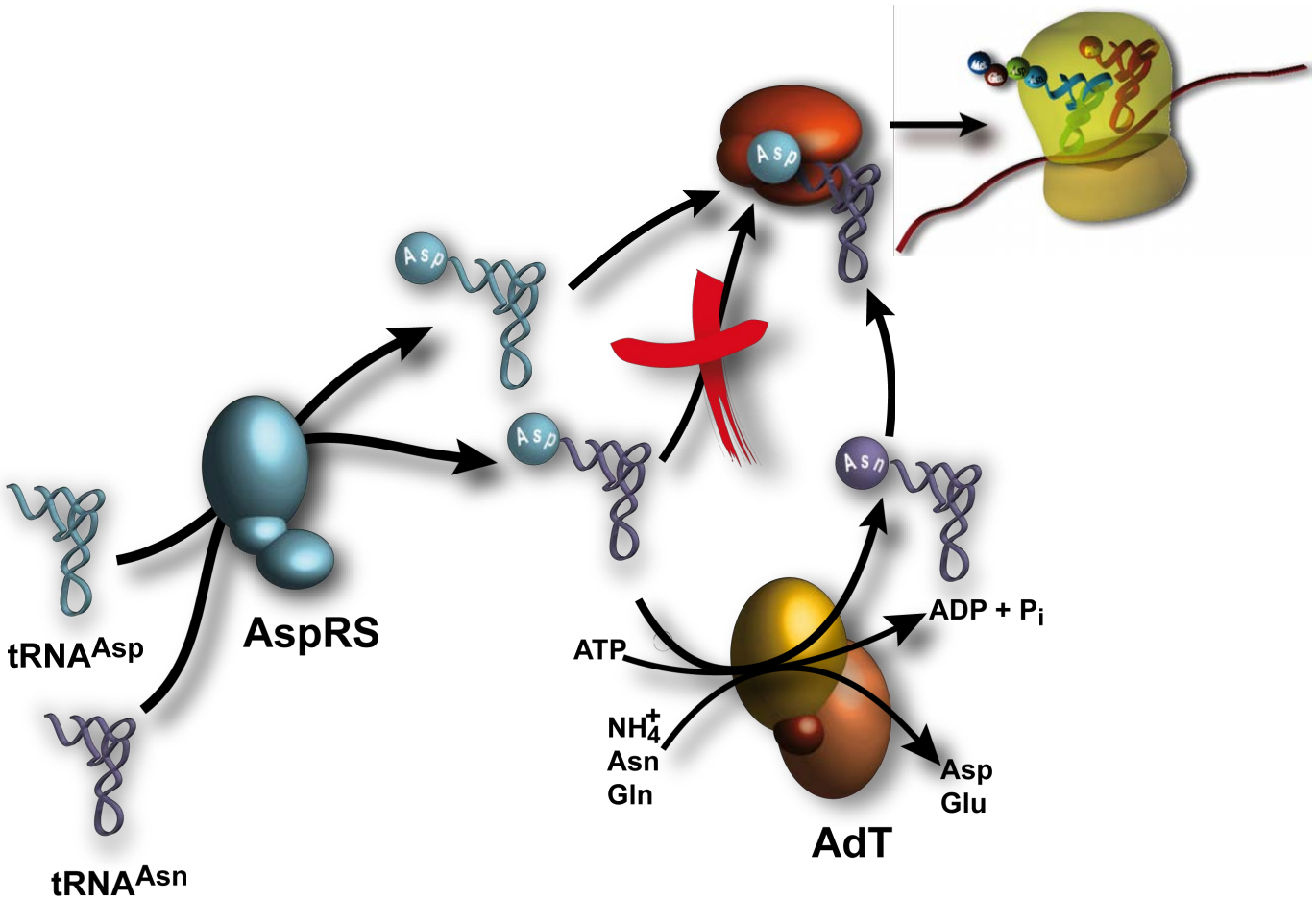
AAU and AAC Coding Capacity



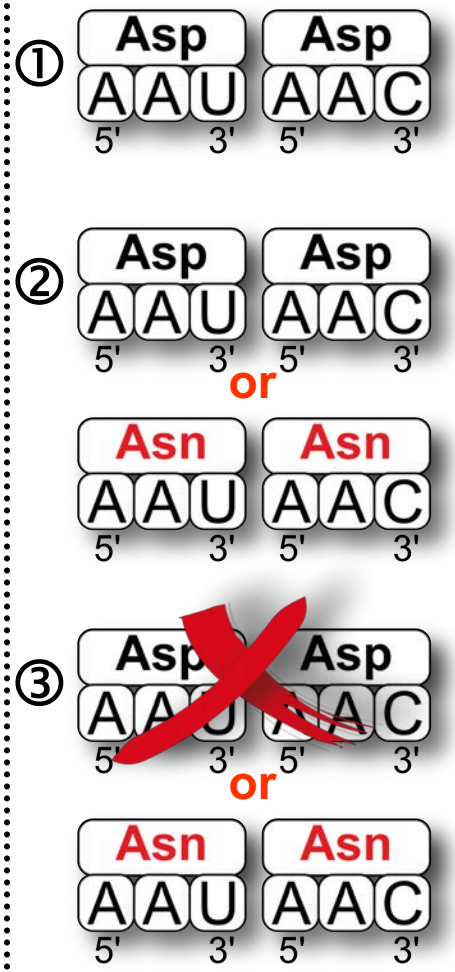
How does the AdT distiguish between Asp-tRNA^{Asn} and Asp-tRNA^{Asp} ?

Step 3: Suppression of AAU and AAC codon ambiguity

EF-Tu – mediated suppression of Codon ambiguity

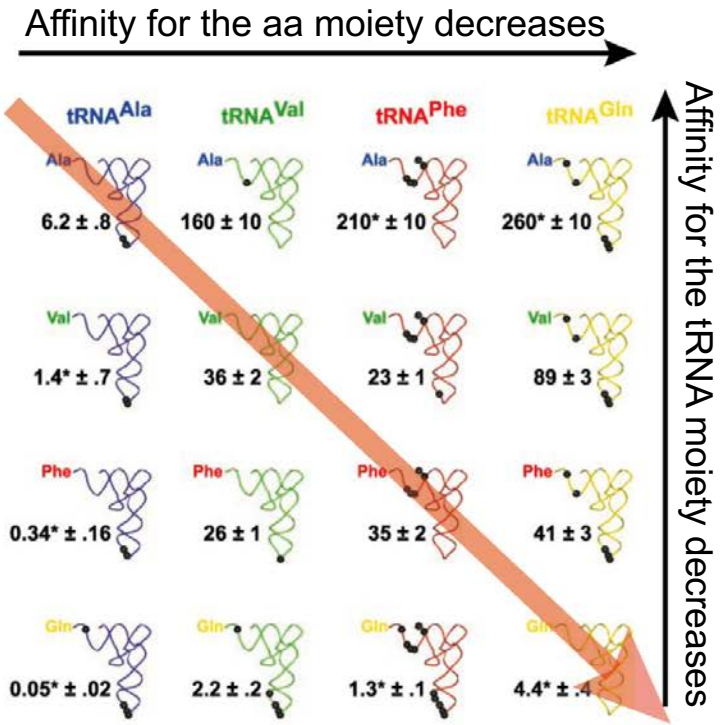


AAU and AAC Coding Capacity



Step 3: Suppression of AAU and AAC codon ambiguity

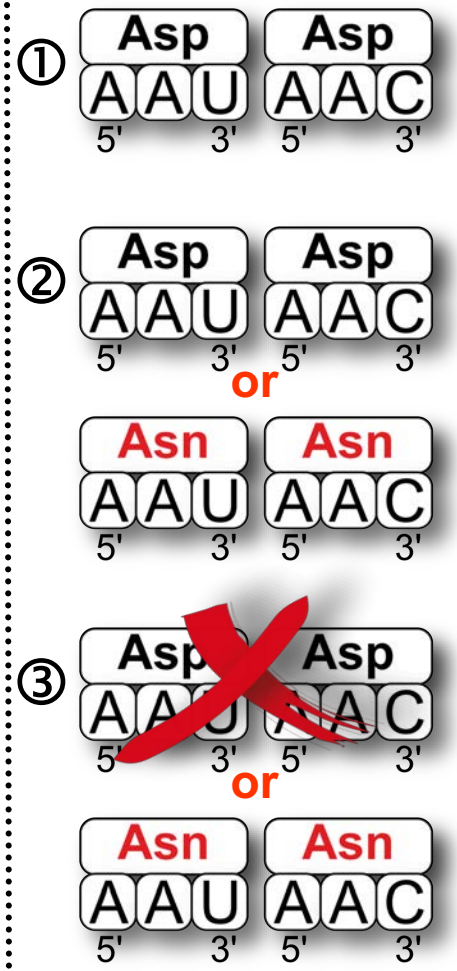
EF-Tu – mediated suppression of Codon ambiguity



Thermodynamic compensation of aa & tRNA binding affinities

EF-Tu allows only weak-strong or strong-weak couples to enter protein synthesis

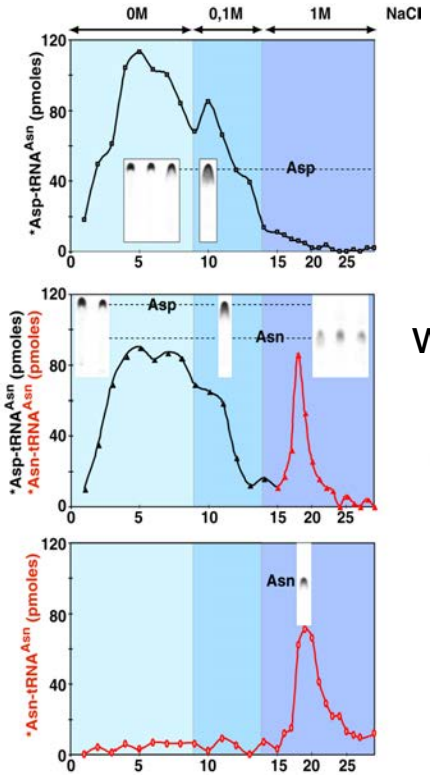
AAU and AAC Coding Capacity



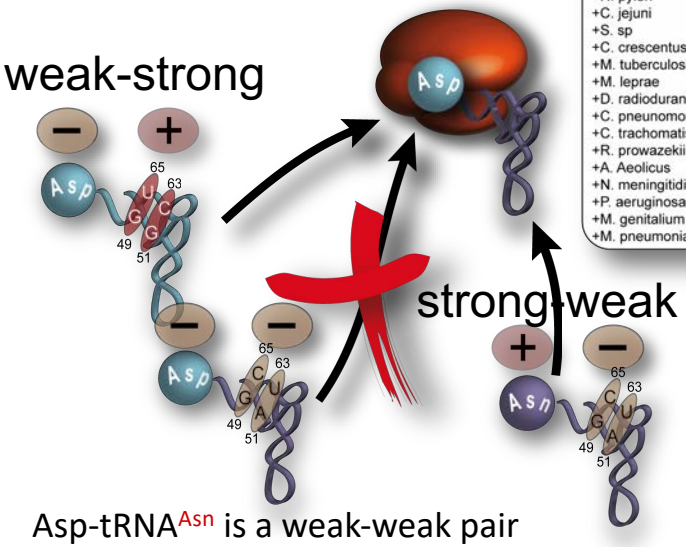
Larivière, F. J. et al. (2001). *Science* **294,** 165-168

Step 3: Suppression of AAU and AAC codon ambiguity

Two acceptor-arm base-pairs define the strength and weakness of the tRNA moiety

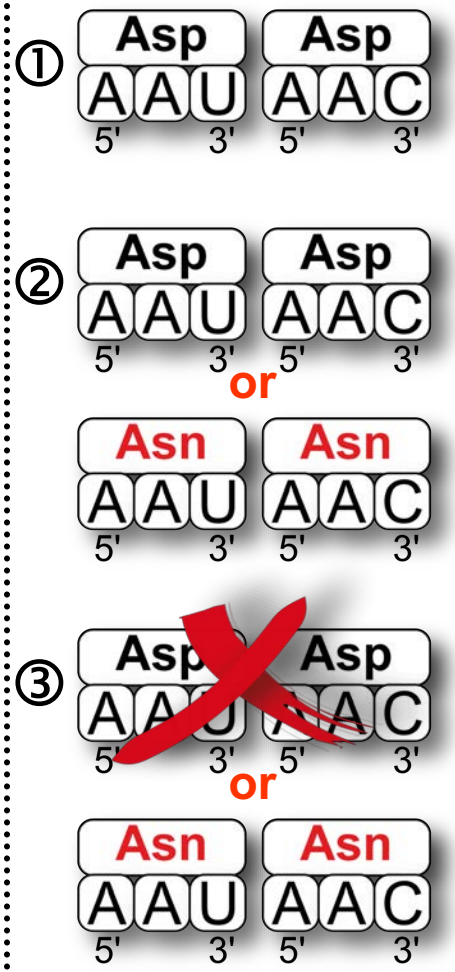


	tRNA^{Asp} 49:65	tRNA^{Asp} 51:63	tRNA^{Asn} 49:65	tRNA^{Asn} 51:63
-T. pallidum	GU	GC	CG	UG
-E. coli	GU	GC	AU	UA
-B. aphidicola	GU	GC	AU	UA
-H. influenzae	GU	GC	GC	UA
-V. cholerae	GU	GC	GC	UA
-P. multocida	GU	GC	GC	UA
-X. fastidiosa	GU	GC	GC	UA
-U. urealitycum	GU	GC	GC	AU
-B. subtilis	GU	GC	GC	AU
-B. halodurans	GU	GC	GC	AU
-S. aureus	GU	GC	GC	AU
-B. burgdorferi	GU	GC	GC	AU
+T. thermophilus	GU	GC	GC	AU
+T. maritima	GU	GC	GC	AU
+H. pylori	GU	GC	GC	AU
+C. jejuni	GU	GC	GC	AU
+S. sp	GU	GC	GC	AU
+C. crescentus	GU	GC	GC	AU
+M. tuberculosi	GU	GC	GC	AU
+M. leprae	GU	GC	GC	AU
+D. radiodurans	GU	GC	GC	UA
+C. pneunomonia	GU	GC	GC	UA
+C. trachomatis	GU	GC	GC	UA
+R. prowazekii	GU	GC	GC	UG
+A. Aeolicus	GU	GC	GC	GC
+N. meningitidi	GU	GC	CG	UA
+P. aeruginosa	GU	GC	CG	UA
+M. genitalium	GU	GC	AU	AU
+M. pneumoniae	GU	GC	AU	AU



Asp-tRNA^{Asn} is a weak-weak pair excluded from protein synthesis

AAU and AAC Coding Capacity

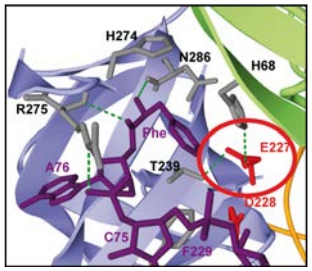


Roy, H. et al. (2007). *Nucl. Acids Res.* **35,** 3420-3430

Step 3: Suppression of AAU and AAC codon ambiguity

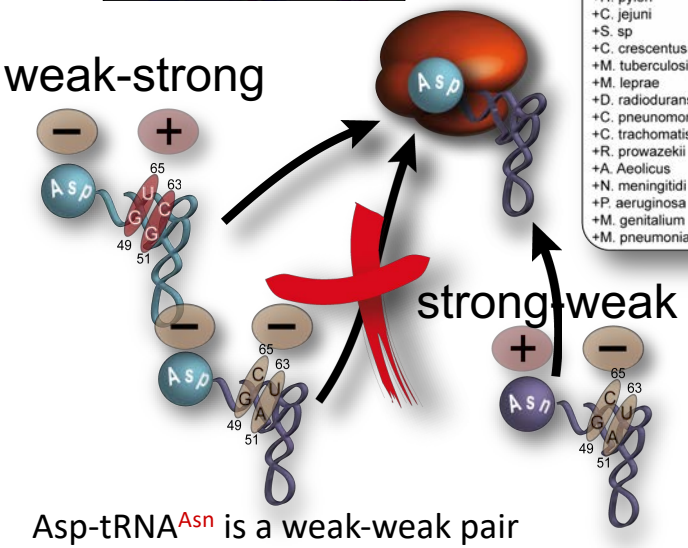
Two acceptor-arm base-pairs define the strength and weakness of the tRNA moiety

	222	230
-T. pallidum	FLLSIEDVYT	
-E. coli	FLLPIEDVFS	
-B. aphidicola	FLLPIEDVFS	
-H. influenzae	FLLPIEDVFS	
-V. cholerae	FLMPIEDVFS	
-P. multocida	FLLPIEDVFS	
-X. fastidiosa	FLMPVEDVFS	
-U. urealitycum	FLLAIEDVFT	
-B. subtilis	FMMPVEDVFS	
-B. halodurans	FMMPVEDVFS	
-S. aureus	FMMPVEDVFS	
-B. burgdorferi	FLLAVEDVFS	
+T. thermophilus	FLMPVEDVFT	
+T. maritima	FLMPIEDVFS	
+H. pylori	FLMPVEDVFS	
+C. jejuni	FLMPIEDVFS	
+S. sp	FLMAVEDVFS	
+C. crescentus	FLMPVEDVFS	
+M. tuberculosis	FLMPVEDVFT	
+M. leprae	FLMPVEDVFT	
+D. radiodurans	FLMPVEDVFT	
+C. pneunomonia	FLMPIEDVFS	
+C. trachomatis	FLMPIEDVFS	
+R. prowazekii	FLMPIEDVFS	
Cons. Bacteria	FLMPIEDVFS	
Cons. Archaea	LRIPIQDVYK	
Cons. Eukarya	LRLPIQDVYS	



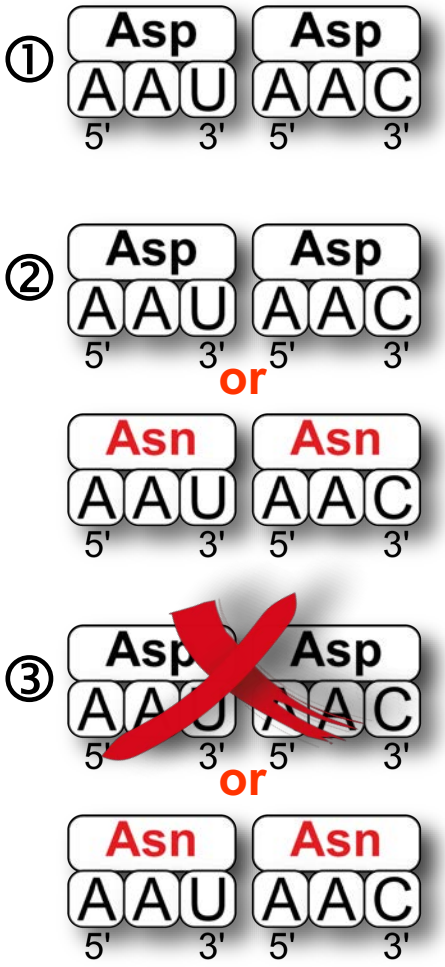
	tRNA^{Asp} 49:65	tRNA^{Asp} 51:63	tRNA^{Asn} 49:65	tRNA^{Asn} 51:63
-T. pallidum	GU	GC	CG	UG
-E. coli	GU	GC	AU	UA
-B. aphidicola	GU	GC	AU	UA
-H. influenzae	GU	GC	GC	UA
-V. cholerae	GU	GC	GC	UA
-P. multocida	GU	GC	GC	UA
-X. fastidiosa	GU	GC	GC	UA
-U. urealitycum	GU	GC	GC	AU
-B. subtilis	GU	GC	GC	AU
-B. halodurans	GU	GC	GC	AU
-S. aureus	GU	GC	GC	AU
-B. burgdorferi	GU	GC	GC	AU
+T. thermophilus	GU	GC	GC	AU
+T. maritima	GU	GC	GC	AU
+H. pylori	GU	GC	GC	AU
+C. jejuni	GU	GC	GC	AU
+S. sp	GU	GC	GC	AU
+C. crescentus	GU	GC	GC	AU
+M. tuberculosi	GU	GC	GC	AU
+M. leprae	GU	GC	GC	AU
+D. radiodurans	GU	GC	GC	UA
+C. pneunomonia	GU	GC	GC	UA
+C. trachomatis	GU	GC	GC	UA
+R. prowazekii	GU	GC	GC	UG
+A. Aeolicus	GU	GC	GC	GC
+N. meningitidi	GU	GC	CG	UA
+P. aeruginosa	GU	GC	CG	UA
+M. genitalium	GU	GC	AU	AU
+M. pneumoniae	GU	GC	AU	AU

EF-Tu	K_A (10^6 M^{-1})	
wt	3.8	nm
D228A	5.8	nm
E227A	13.4	1.9
E227Q	15.4	nm



Asp-tRNA^{Asn} is a weak-weak pair excluded from protein synthesis

AAU and AAC Coding Capacity



Mutation in Ala allows Asp-tRNA^{Asn} to bind to EF-Tu

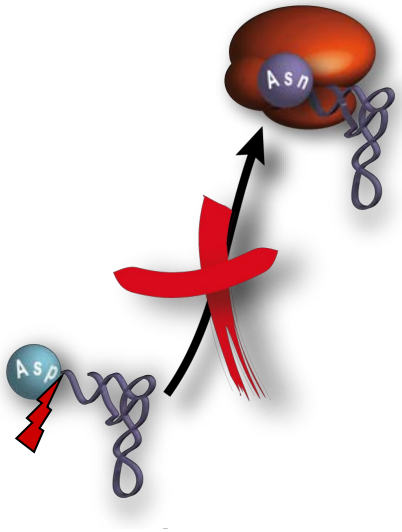
One EF-Tu Glu residue modulates the strength and weakness of the aa moiety

Roy, H. et al. (2007). *Nucl. Acids Res.* **35,** 3420-3430

Step 2-3: RNP-mediated Asn-tRNA^{Asn} synthesis

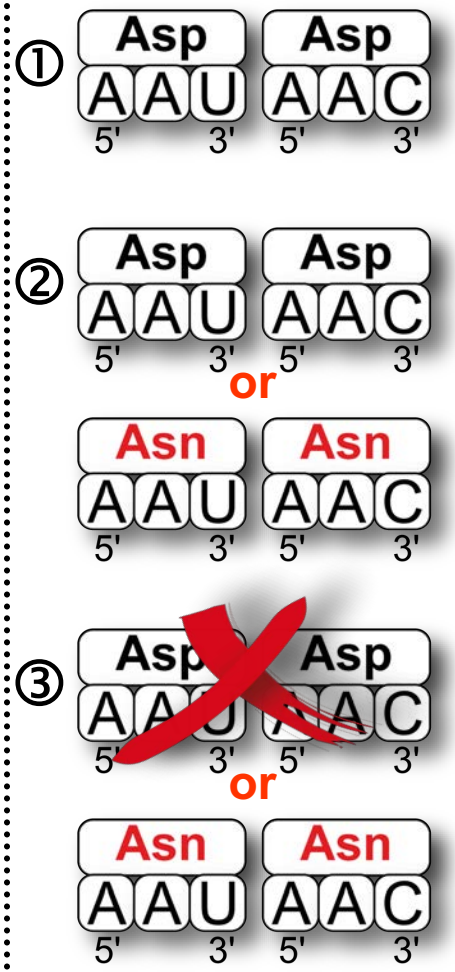
How is Asp-tRNA^{Asn} protected against hydrolysis ?

Asn-tRNA^{Asn} is protected against hydrolysis because it binds EF-Tu



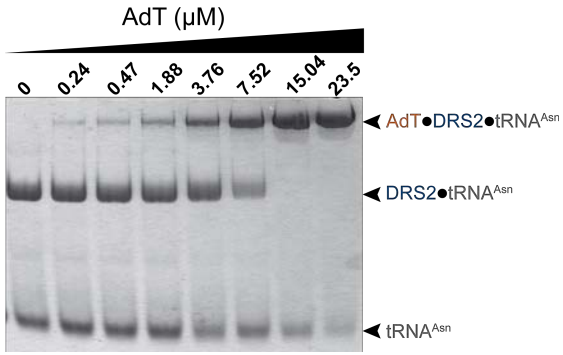
A tRNP, the transamidosome, is assembled prior to any aa attachment and modification

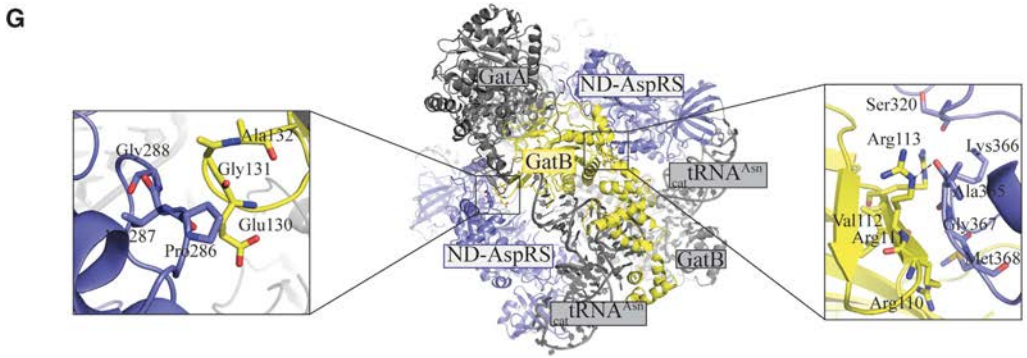
AAU and AAC Coding Capacity

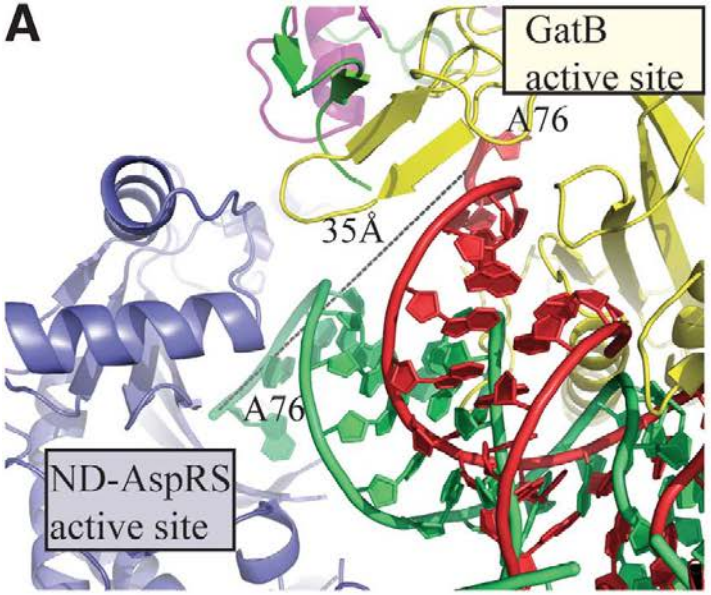


Bailly, M et al. (2007) *Molecular cell.* **28,** 228-239

Step 2-3: RNP-mediated Asn-tRNA^{Asn} synthesis



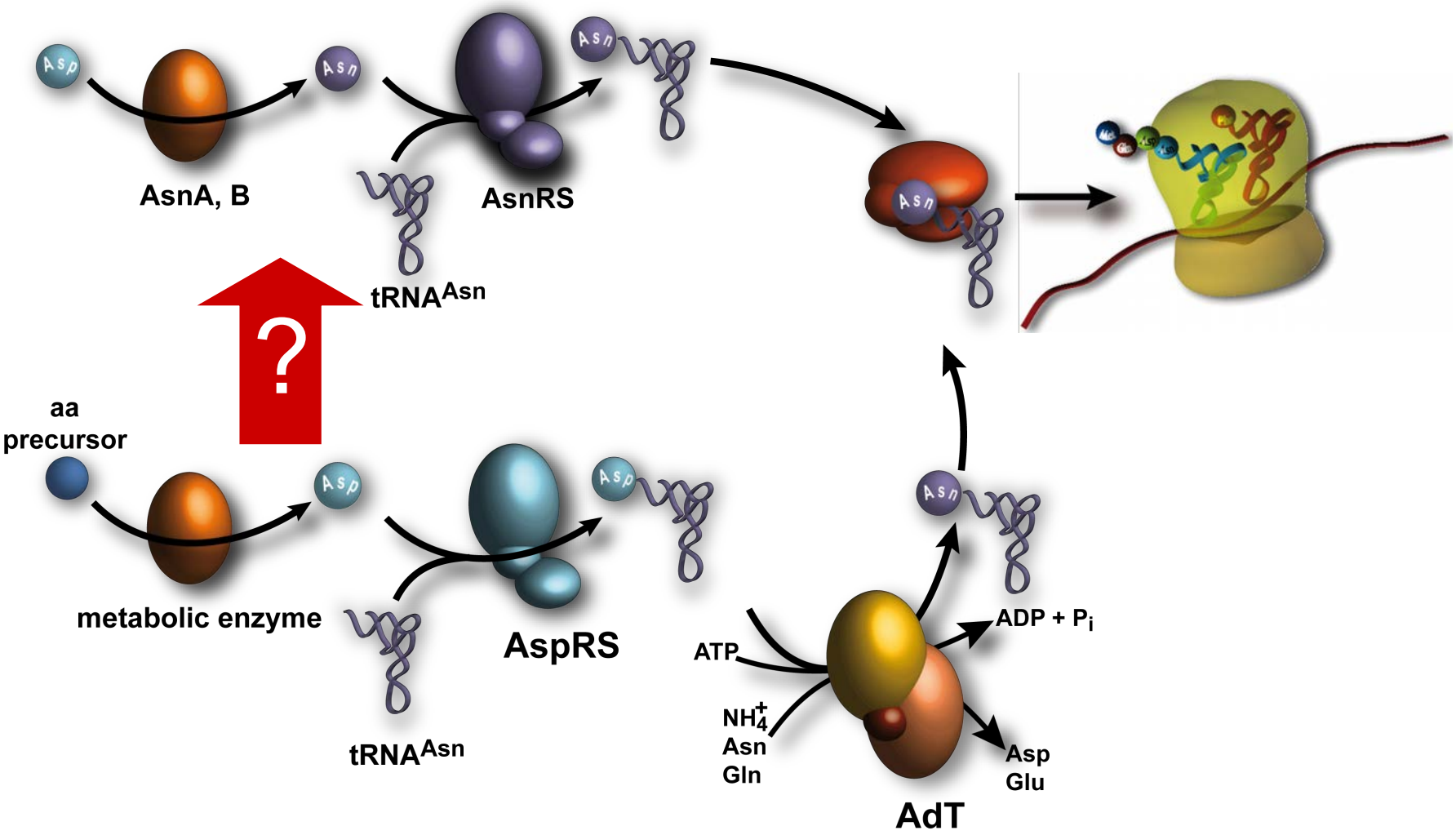




Asp-tRNA^{Asn} is not released prior to amidation

Bailly, M et al. (2007) *Molecular cell.* **28,** 228-239

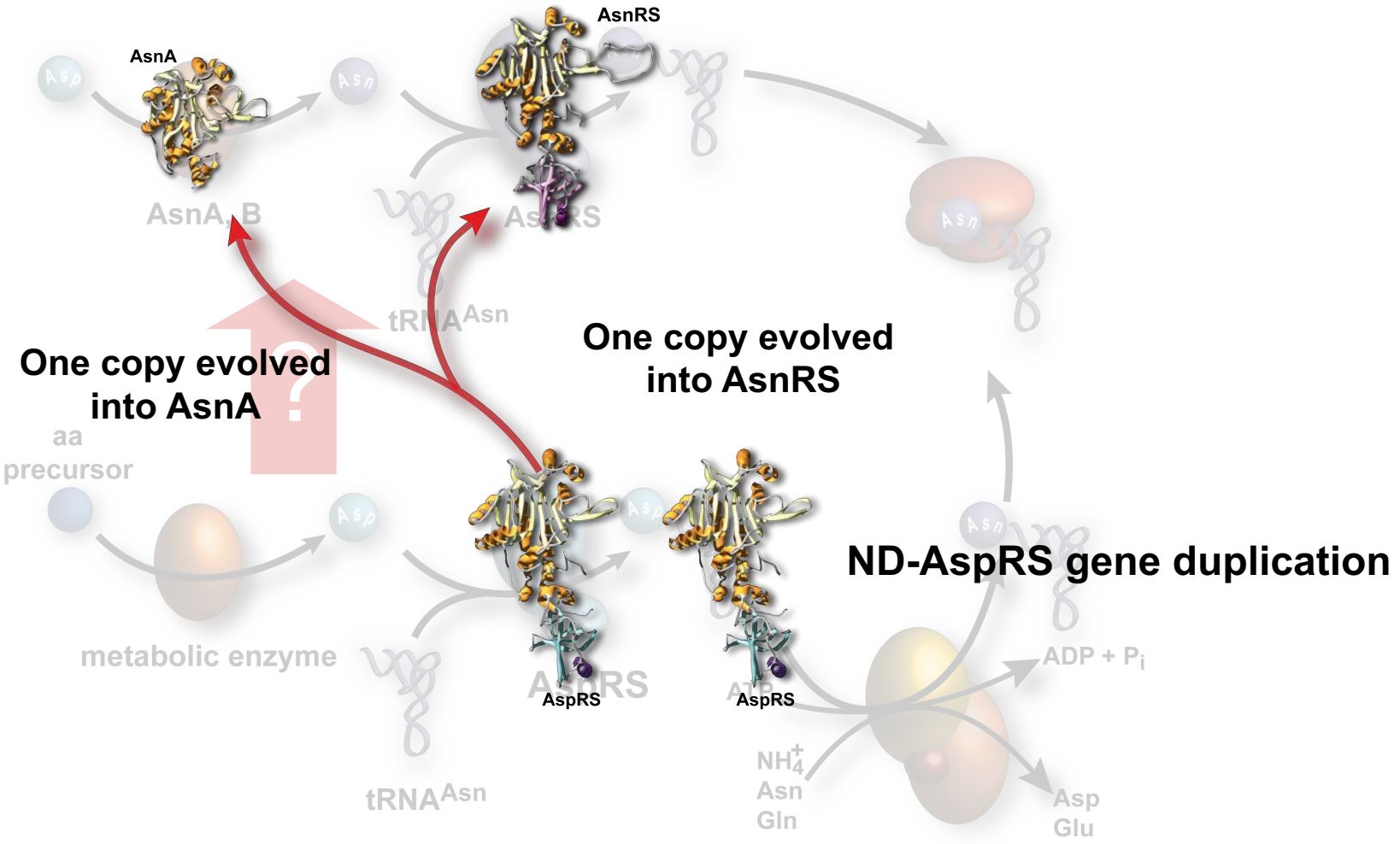
Step 4: The ND-AspRS evolved both AsnRS and AsnA



The ancestral AspRS was the structural scaffold for evolving the new aminoacyl-tRNA synthetase and the new metabolic enzyme

Roy, H. et al. (2003). *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9837-9842.

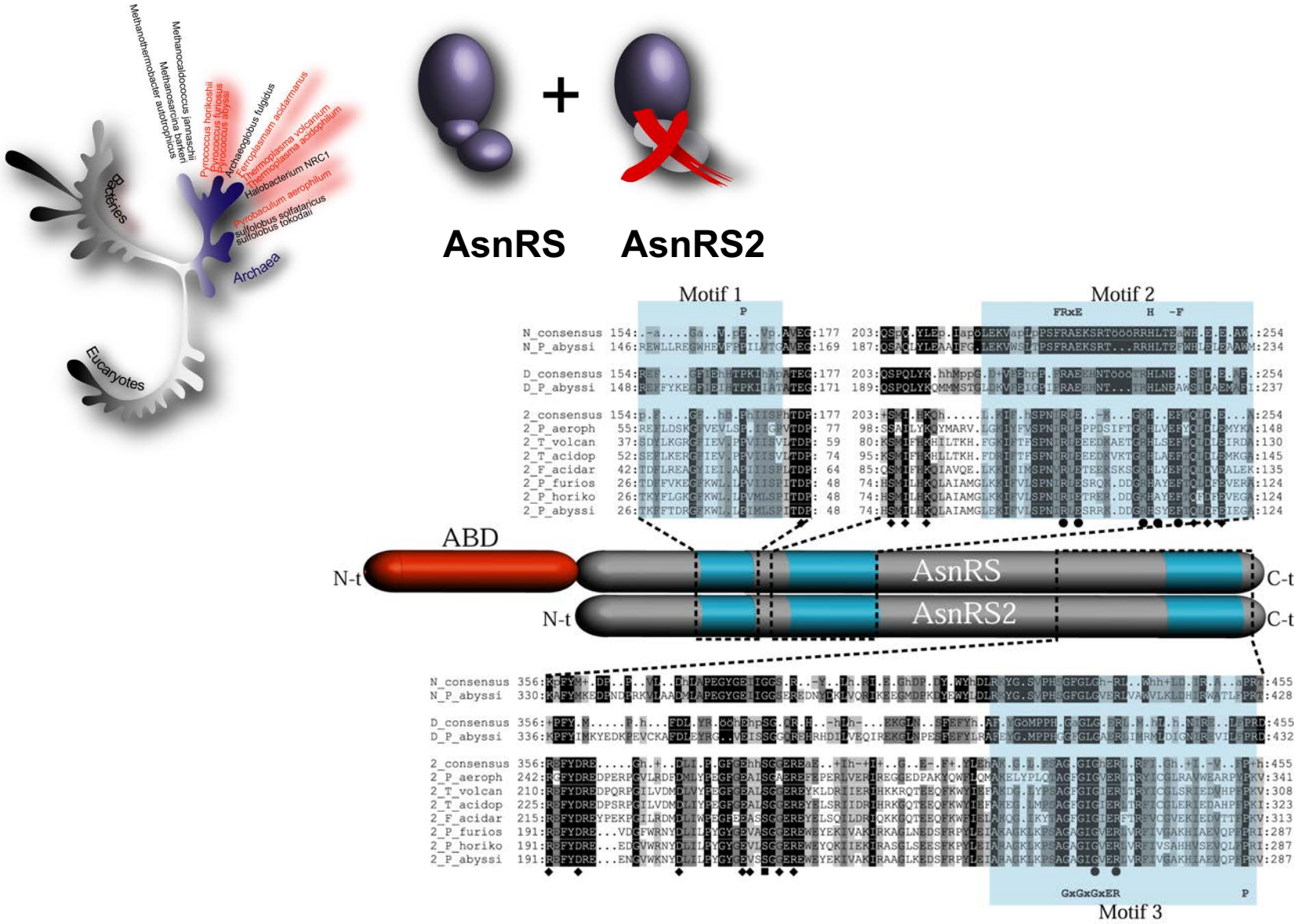
Step 4: The ND-AspRS evolved both AsnRS and AsnA



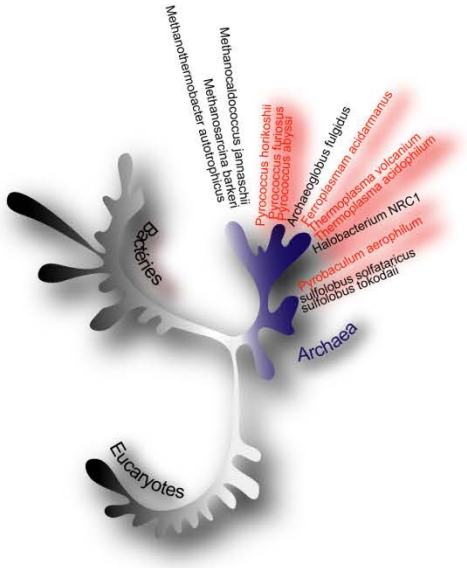
The ancestral AspRS was the structural scaffold for evolving the new aminoacyl-tRNA synthetase and the new metabolic enzyme

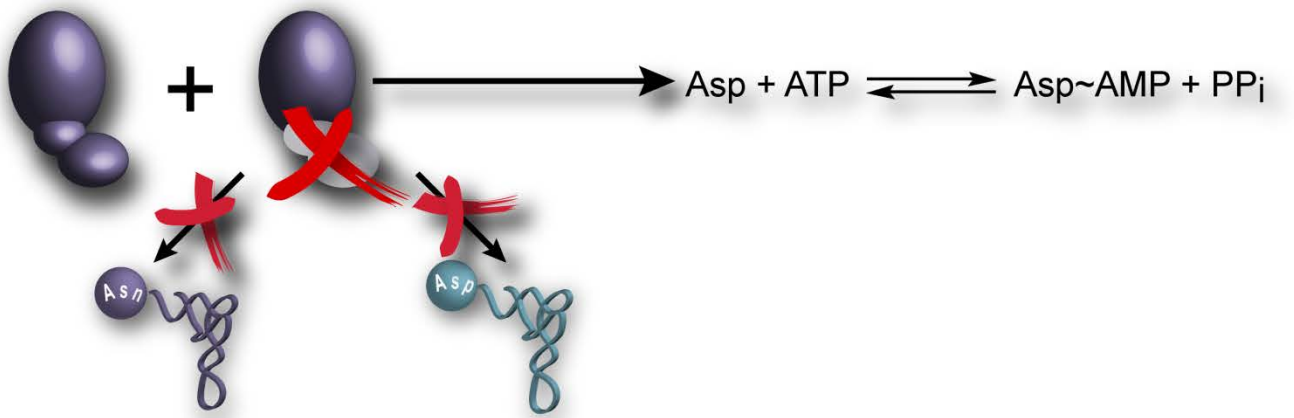
Roy, H. et al. (2003). *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9837-9842.

Step 4: The ND-AspRS evolved both AsnRS and AsnA

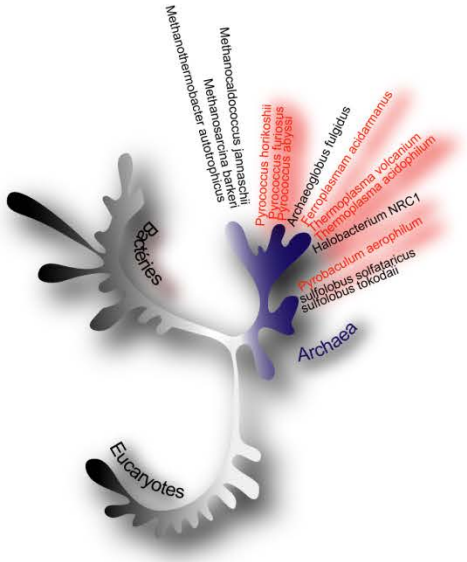


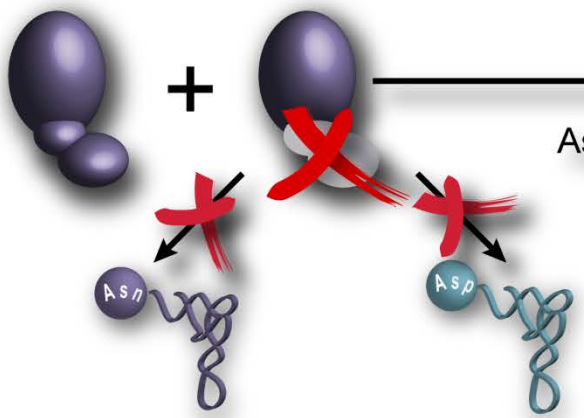
Step 4: The ND-AspRS evolved both AsnRS and AsnA





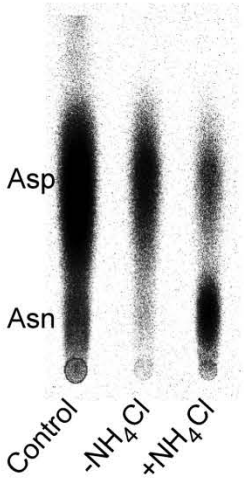
Step 4: The ND-AspRS evolved both AsnRS and AsnA





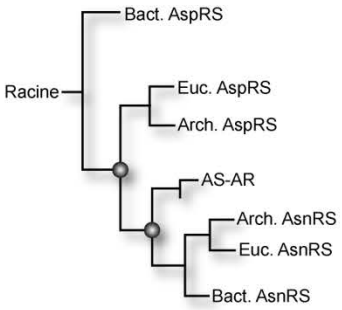
$$Asp + ATP \rightleftharpoons Asp\sim AMP + PP_i$$

$$Asp\sim AMP + NH_4^+ \rightleftharpoons Asn + AMP$$

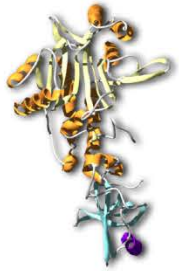


Roy, H., Becker, H. D., Reinbolt, J. & Kern, D. (2003). Proc. Natl. Acad. Sci. U. S. A. 100, 9837-9842.

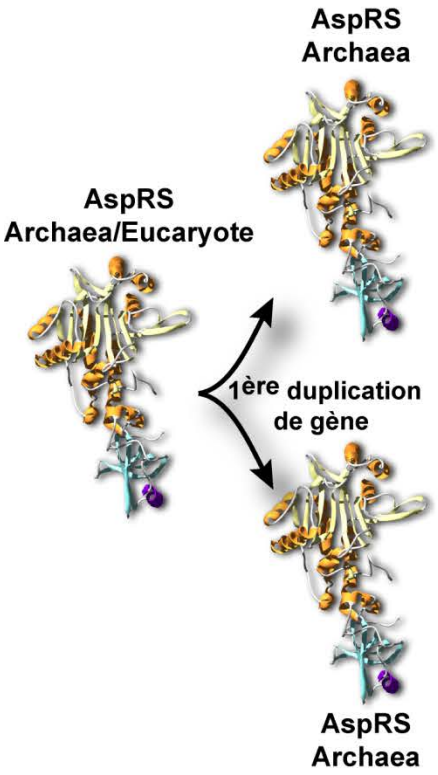
Step 4: The ND-AspRS evolved both AsnRS and AsnA

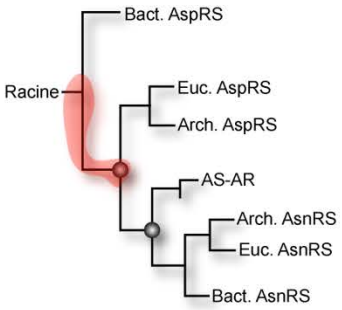


AspRS
Archaea/Eucaryote

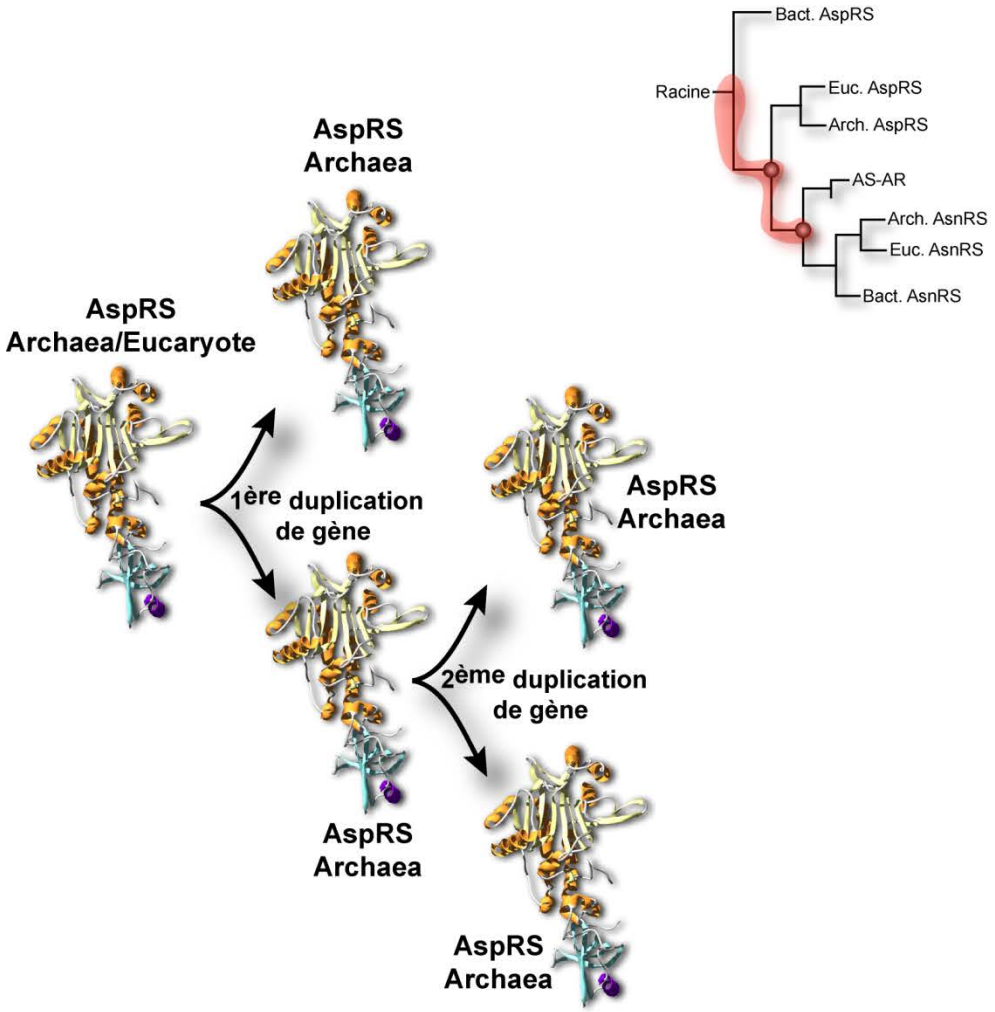


Step 4: The ND-AspRS evolved both AsnRS and AsnA

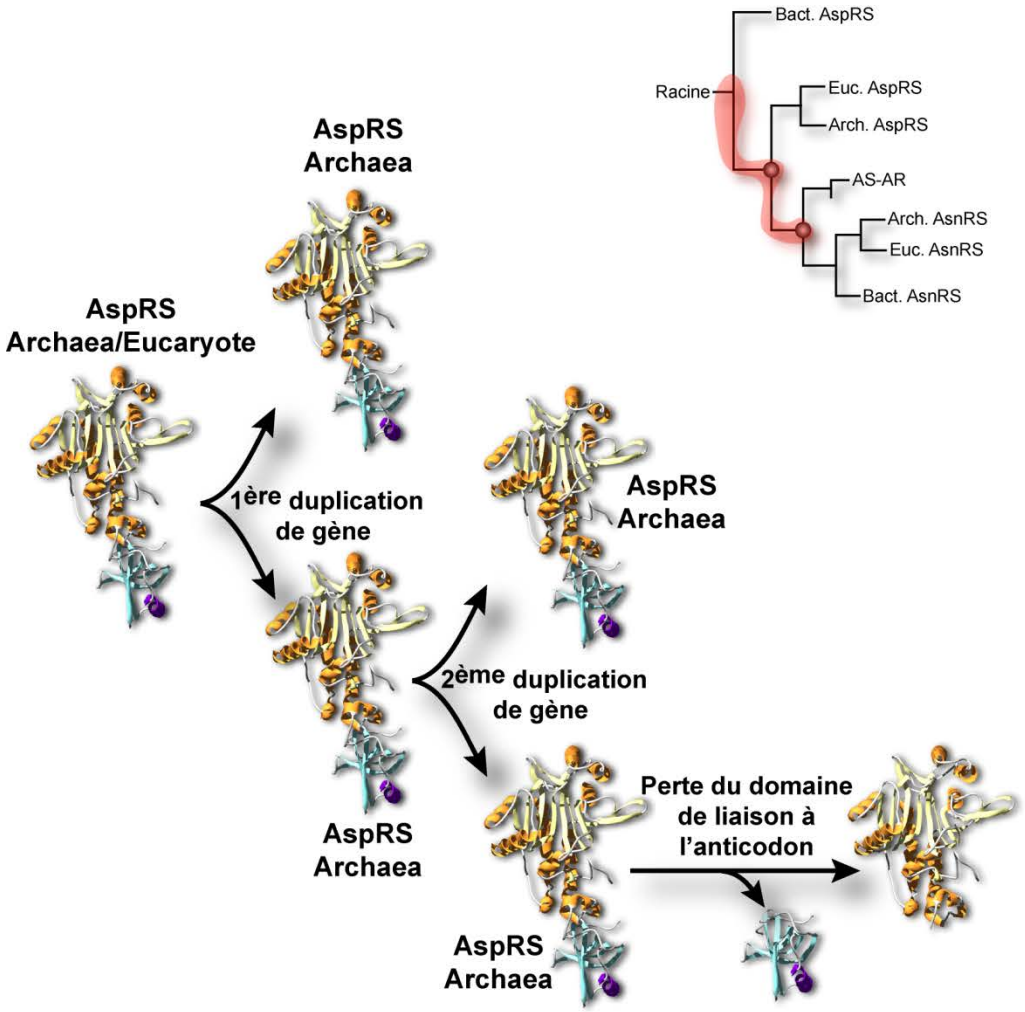




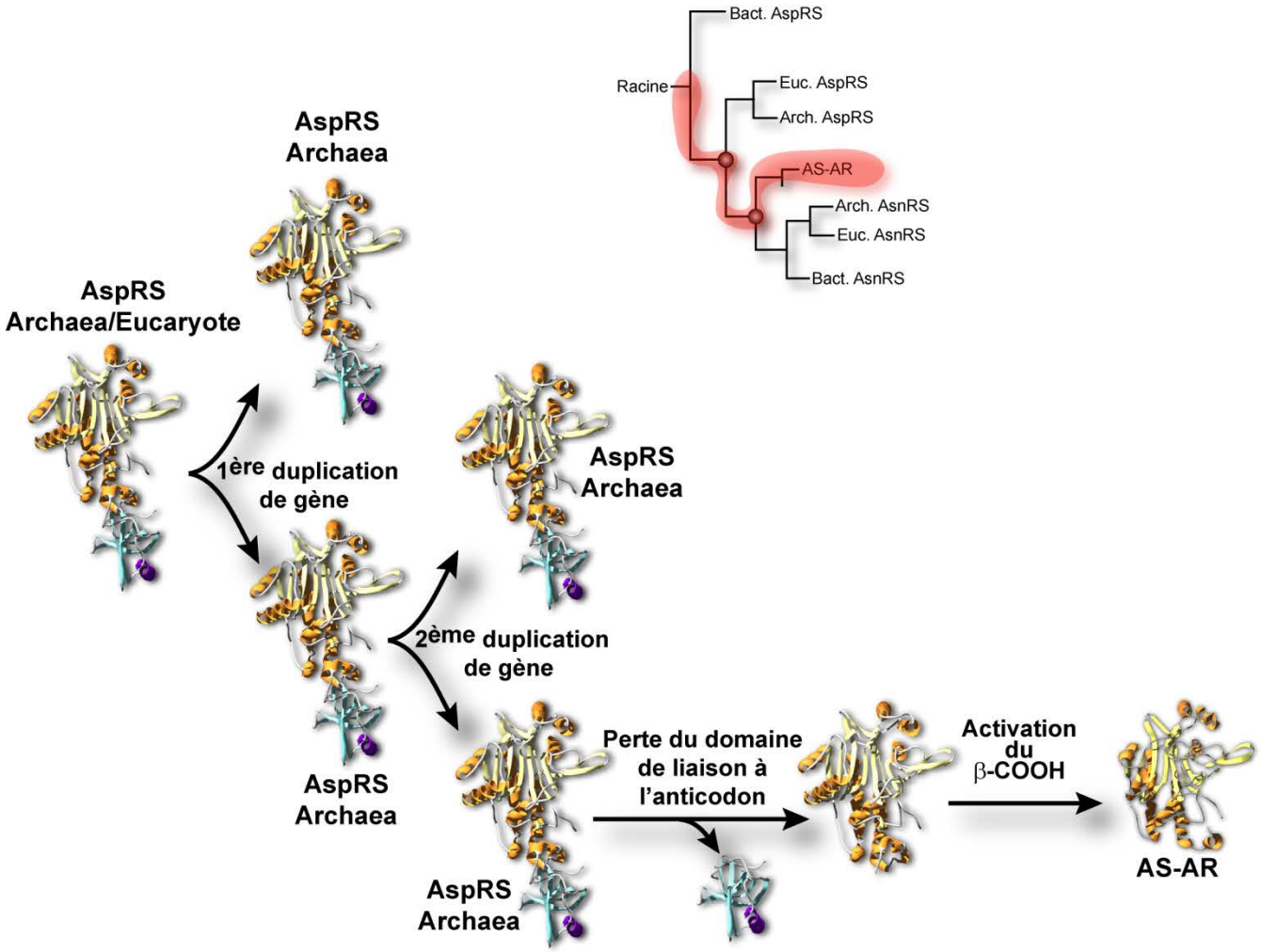
Step 4: The ND-AspRS evolved both AsnRS and AsnA



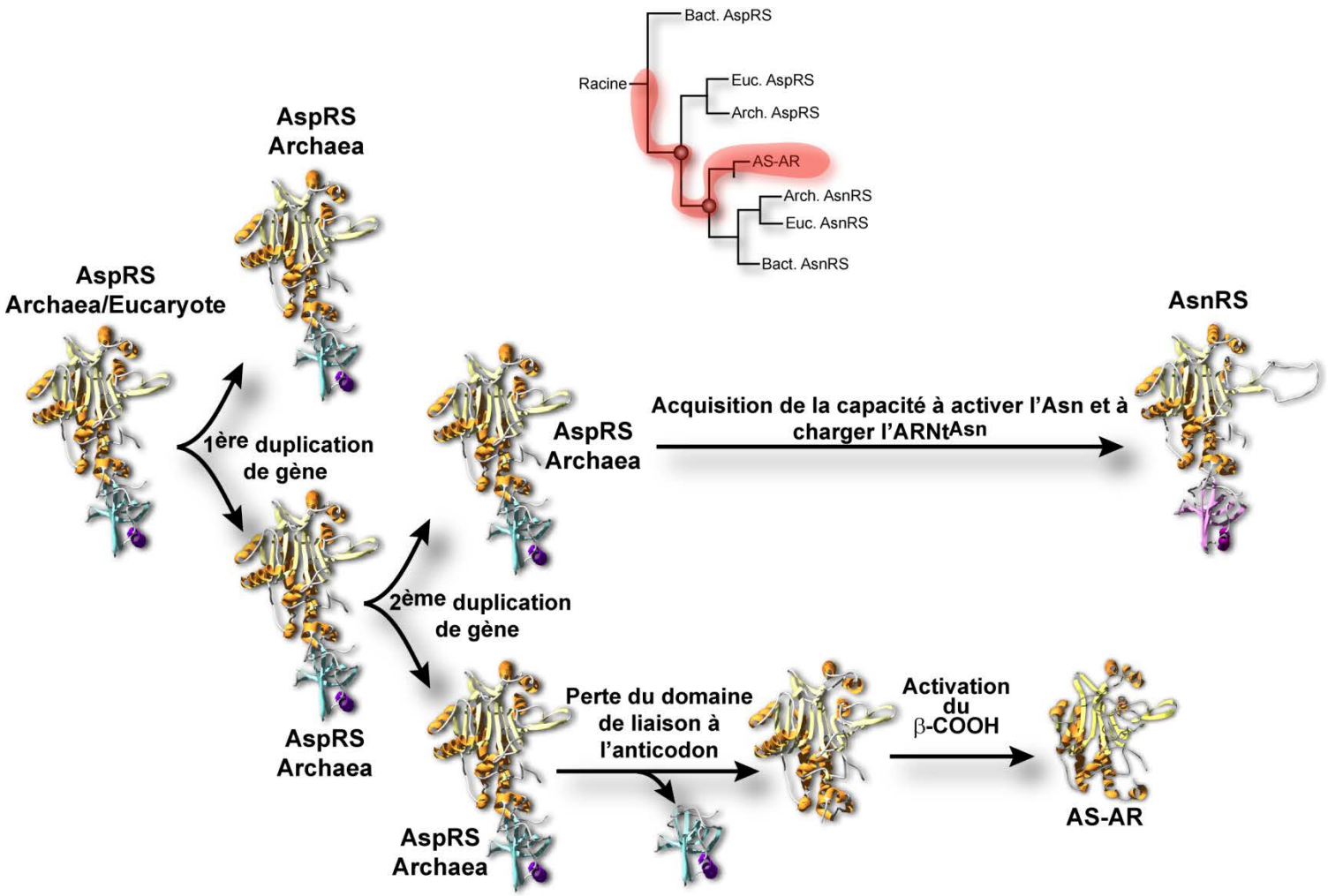
Step 4: The ND-AspRS evolved both AsnRS and AsnA



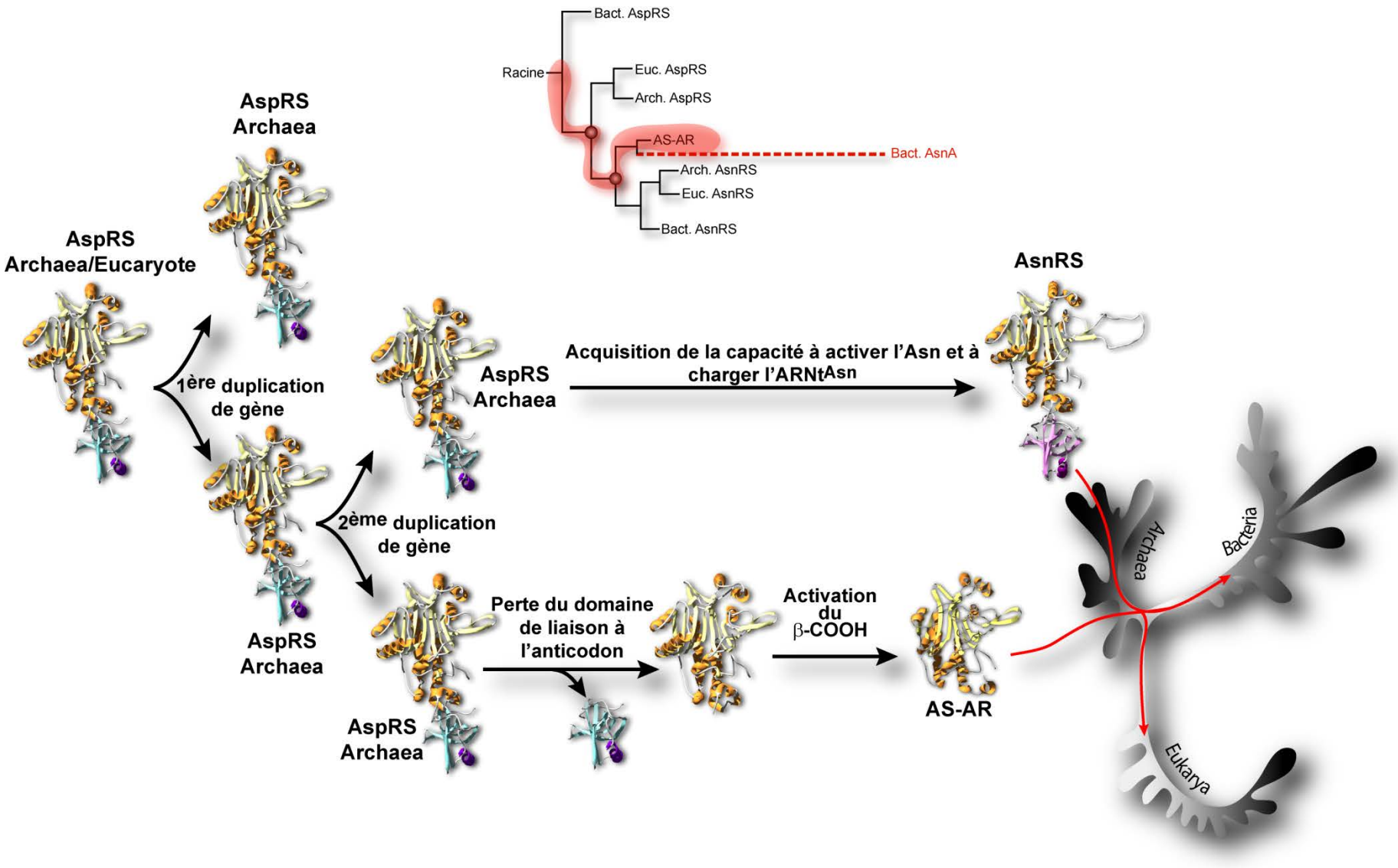
Step 4: The ND-AspRS evolved both AsnRS and AsnA



Step 4: The ND-AspRS evolved both AsnRS and AsnA



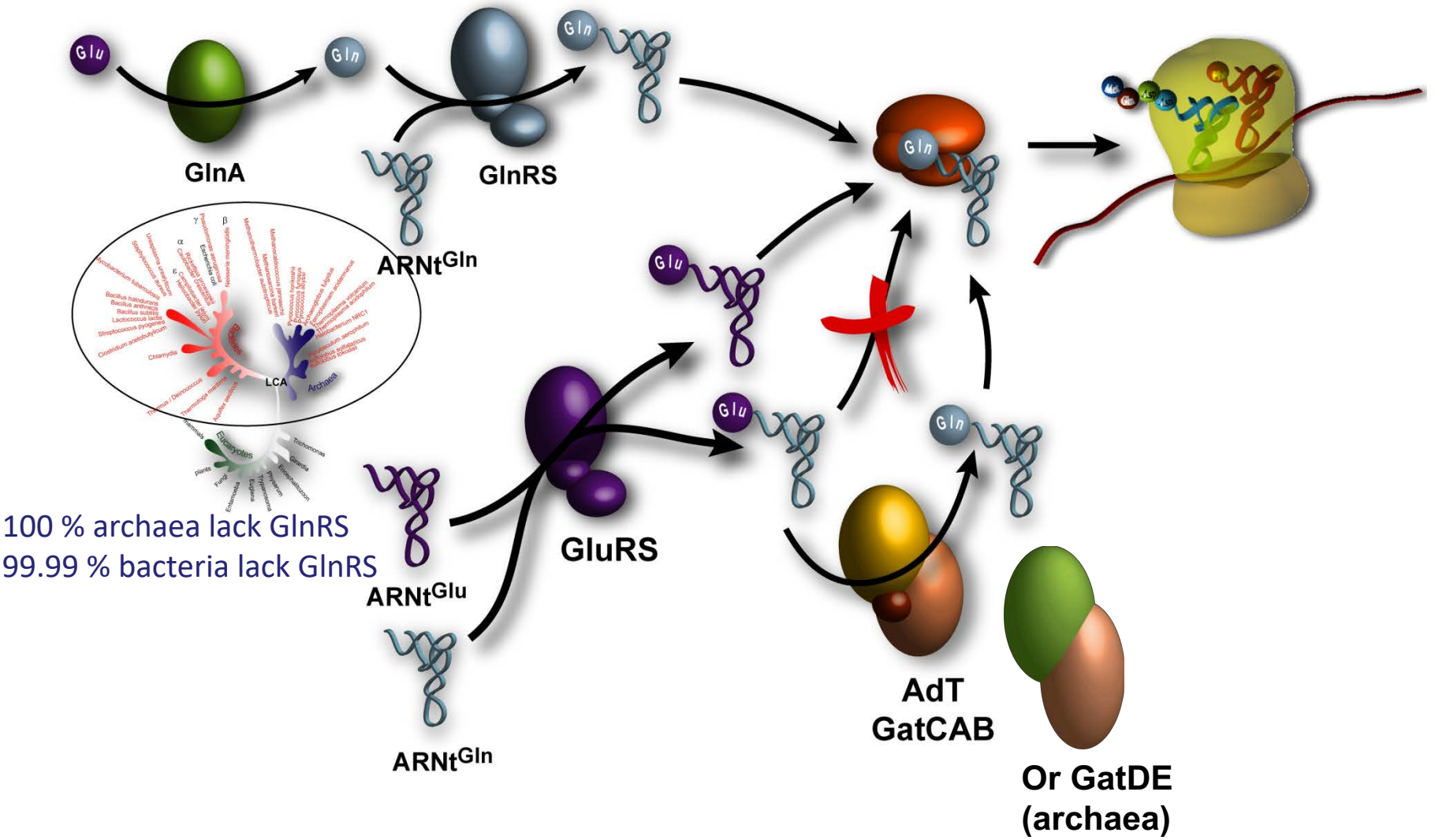
Step 4: The ND-AspRS evolved both AsnRS and AsnA



Roy, H., Becker, H. D., Reinbolt, J. & Kern, D. (2003). Proc. Natl. Acad. Sci. U. S. A. 100, 9837-9842.

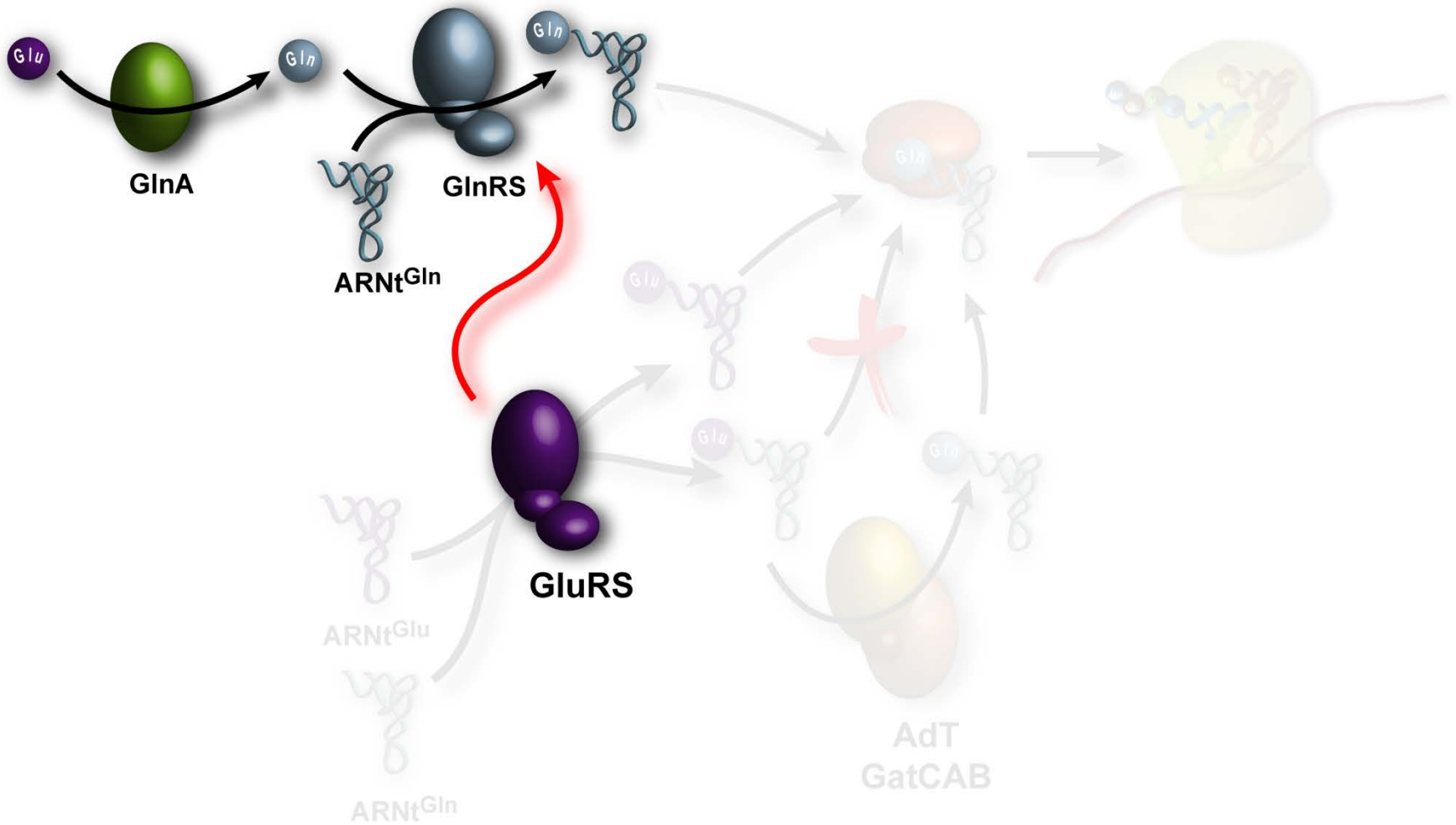
Given the important functions of lipids (5) and sterol modifications (6, 7) in fungal signalling and pathogenesis, we hypothesized that ASs may have similar functi

Introduction of Gln into the GC by infiltration of CAA & CAG Glu codons



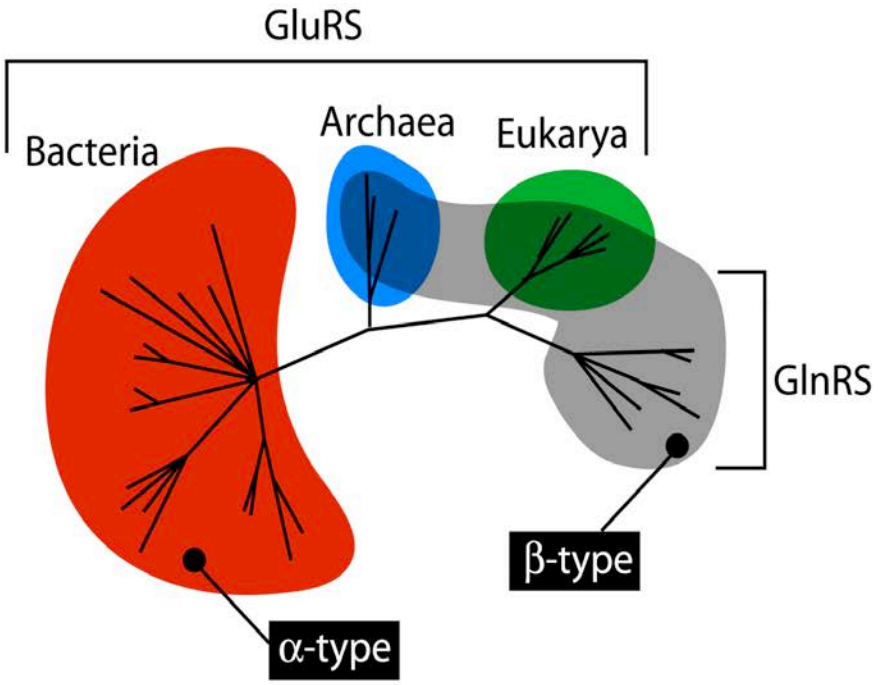
100 % archaea lack GlnRS
99.99 % bacteria lack GlnRS

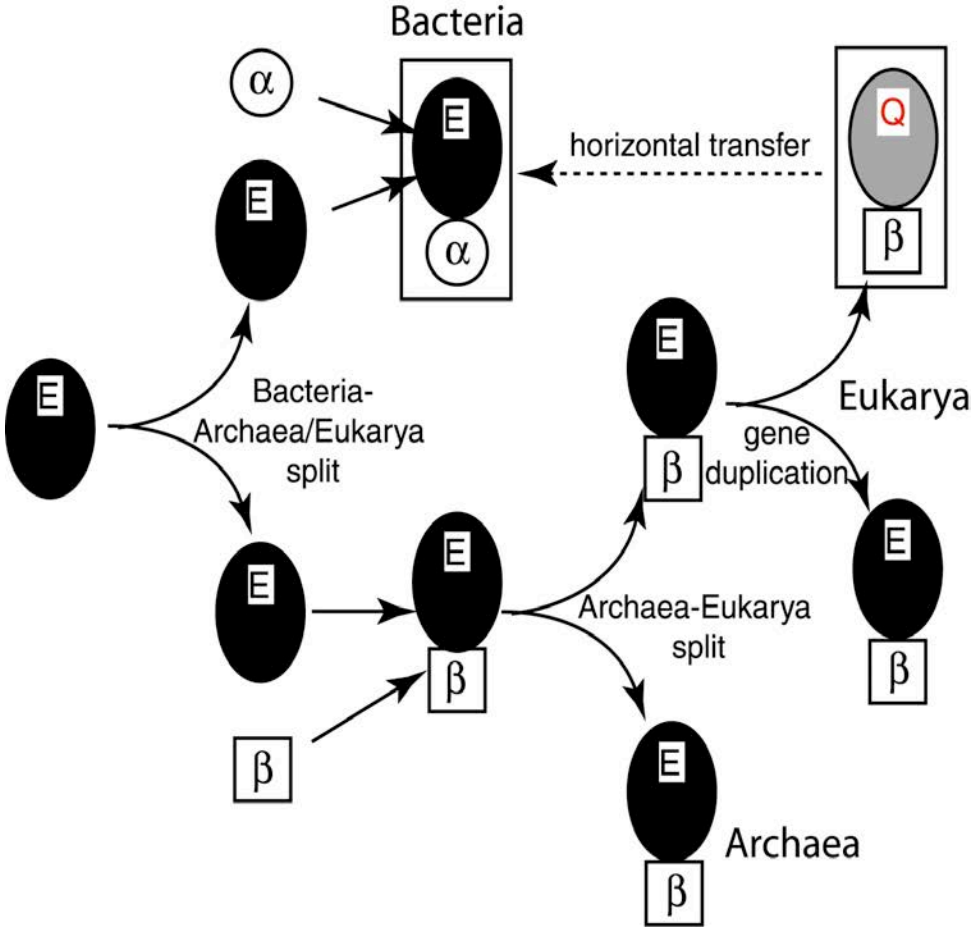
Introduction of Gln into the GC by infiltration of CAA & CAG Glu codons



Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M. & Mirande, M. (1994). Proc. Natl. Acad. Sci. U.S.A. 91, 8670-8674.

Introduction of Gln into the GC by infiltration of CAA & CAG Glu codons



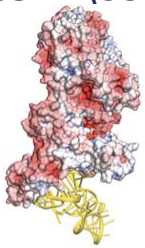


Introduction of Gln into the GC by infiltration of CAA & CAG Glu codons

In Eukaryotic Organelles

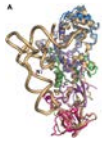
Kingdom-specific pathways for Gln-tRNA^{Gln} synthesis

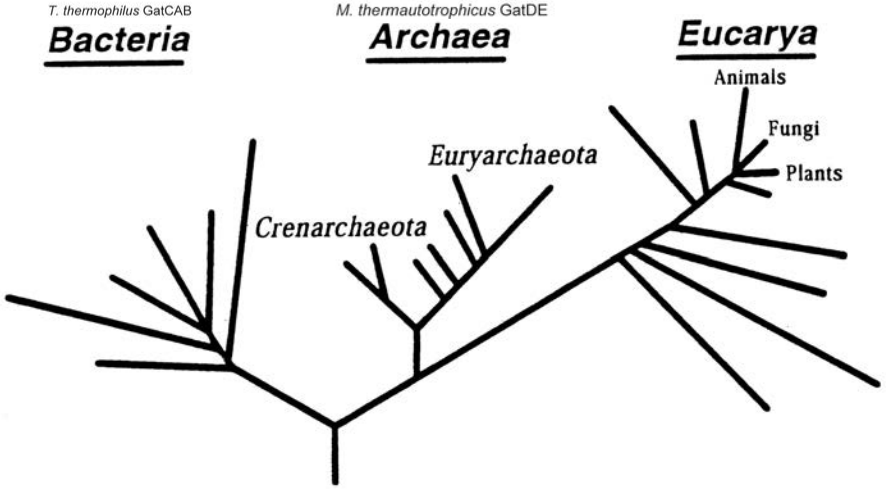
GatCAB (99.99 %) **GatDE** (100 %) **GlnRS** (100 %)

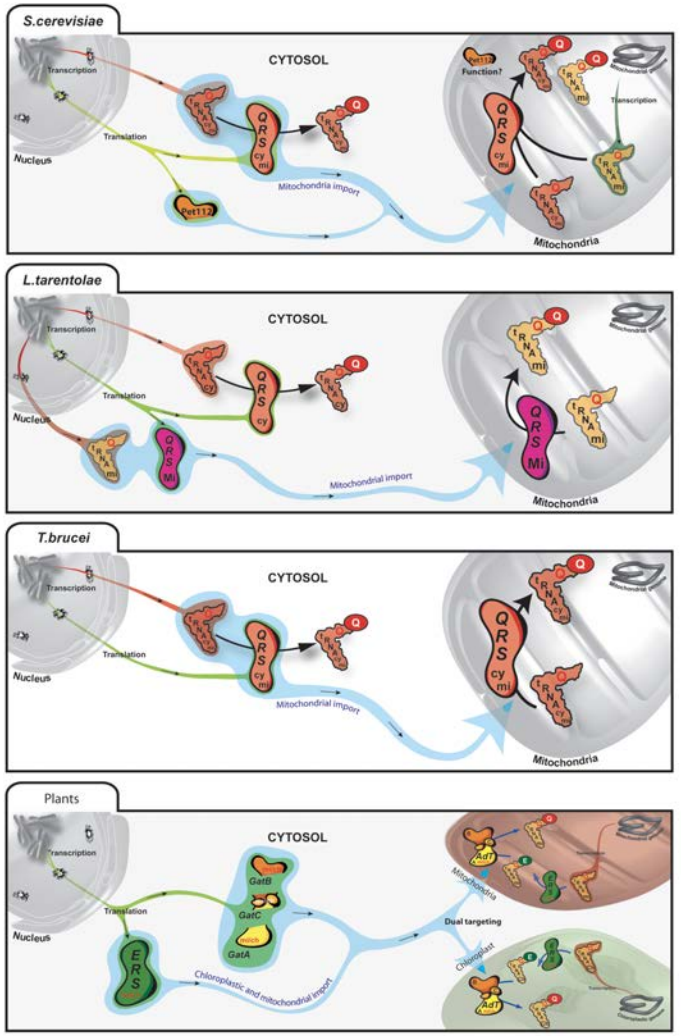


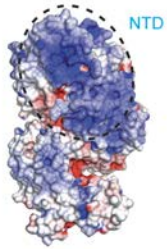
T. thermophilus GatCAB

M. thermautotrophicus GatDE









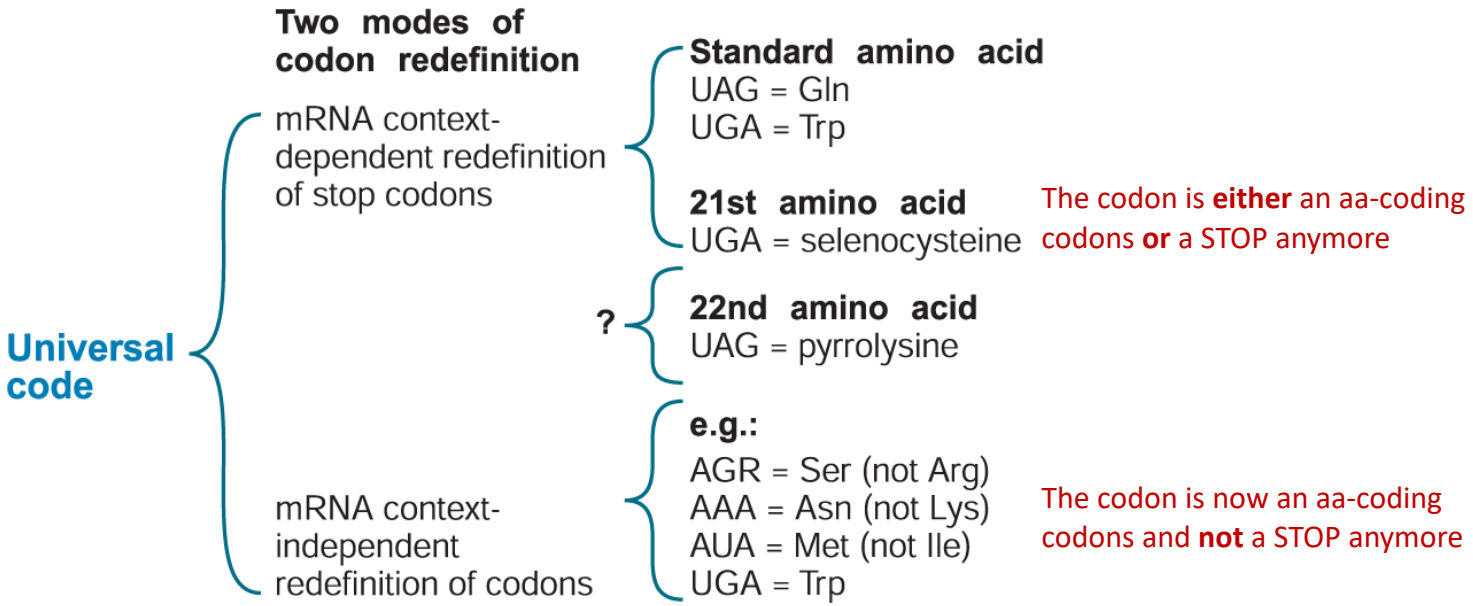
S. cerevisiae GatFAB

Natural redefinition of a STOP codon

1- By infiltration of the coding capacity of other codons

2- By redifinition of a STOP codon

- Context-independent redefinition
- Context-dependent redefinition

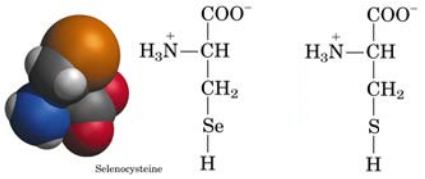


Natural redefinition of a STOP codon example of Sec insertion

Requirements

1. a new aa to add to the GC

1.



Sec
UGA
(Opale)

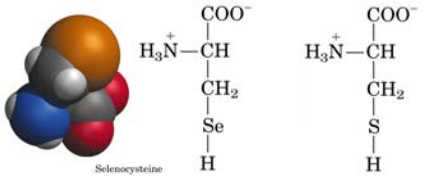
Cys
UGU
UGC

Natural redefinition of a STOP codon example of Sec insertion

Requirements

1. a new aa to add to the GC
2. a suppressor tRNA that will specify this new aa

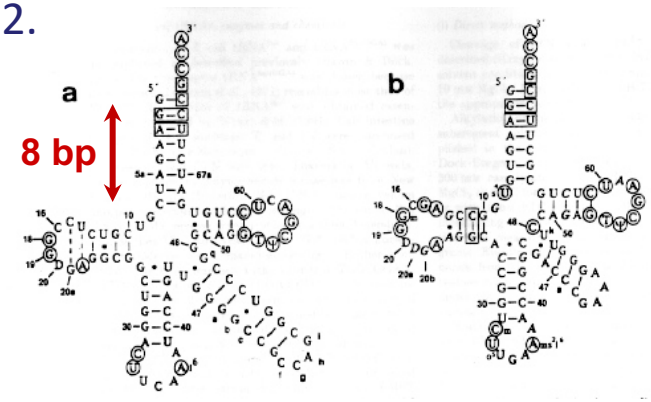
1.



Sec
UGA
(Opale)

Cys
UGU
UGC

2.

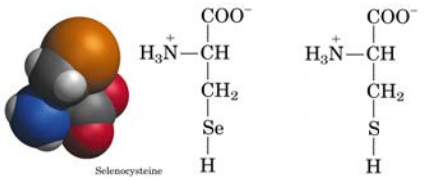


Natural redefinition of a STOP codon example of Sec insertion

Requirements

1. a new aa to add to the GC
2. a suppressor tRNA that will specify this new aa
3. an aaRS activating this new aa and charging the $tRNA^{Sup}$

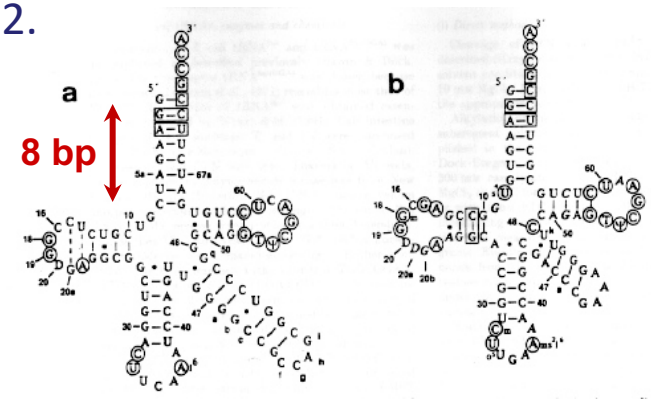
1.



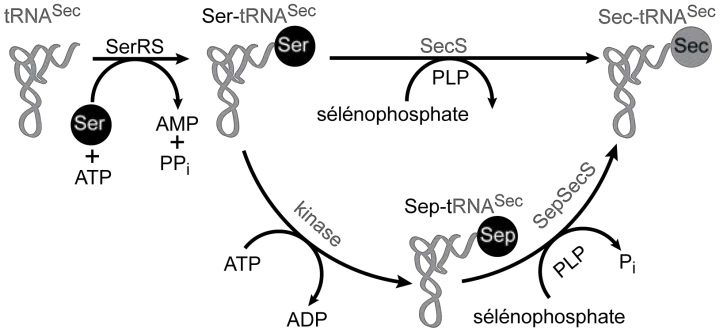
Sec
UGA
(Opale)

Cys
UGU
UGC

2.



3.



BACTERIA

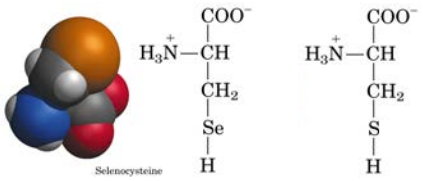
ARCHEA & EUKARYOTES

Natural redefinition of a STOP codon example of Sec insertion

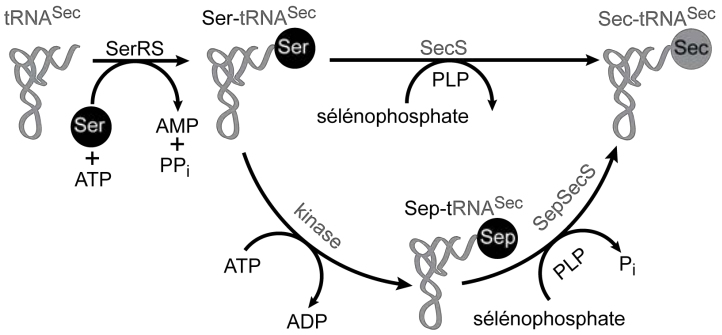
Requirements

1. a new aa to add to the GC
2. a suppressor tRNA that will specify this new aa
3. an aaRS activating this new aa and charging the $tRNA^{Sup}$
4. an elongation factor that will bring the new aa-$tRNA^{Sup}$ to the ribosome (optional)
5. a message on the mRNA indicating to the ribosome that this STOP codon is not a STOP but encodes the new aa

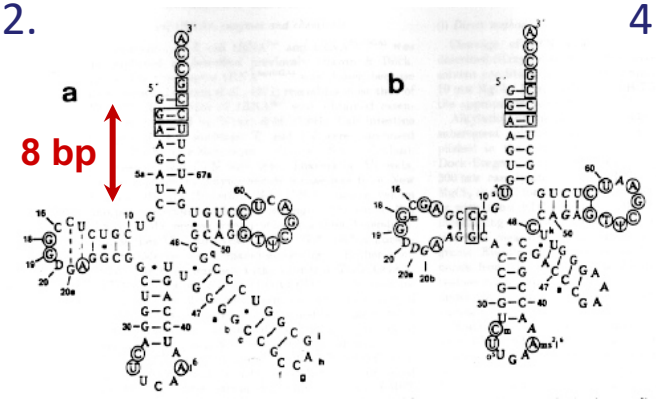
1.



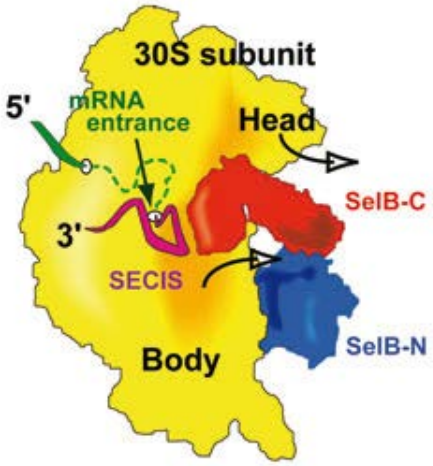
3.



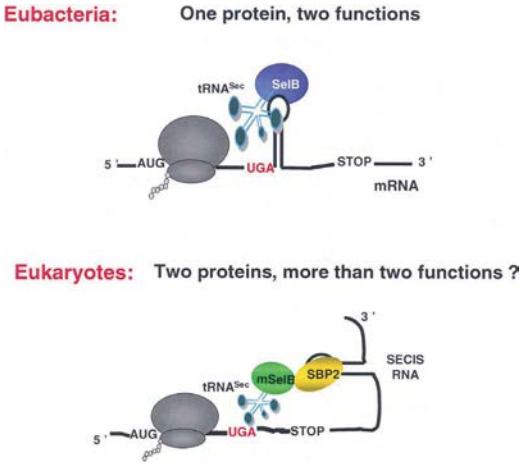
2.



4.



5.



Synthetic redefinition of a STOP codon example of Sep insertion

MJ# Gene description

Translation

0160 PET112 prot

Amino acyl tRNA synthetases

0564 alanyl-tRNA Sase
0237 arginyl-tRNA Sase
1555 aspartyl-tRNA Sase
1377 glutamyl-tRNA Sase
0228 glycyl-tRNA Sase
1000 histidyl-tRNA Sase
0947 isoleucyl-tRNA Sase
0633 leucyl-tRNA Sase
1263 methionyl-tRNA Sase
0487 phenylalanyl-tRNA Sase, alpha sub
1108 phenylalanyl-tRNA Sase, beta sub
1238 prolyl-tRNA Sase
1197 threonyl-tRNA Sase
1415 tryptophanyl-tRNA Sase
0389 tyrosyl-tRNA Sase
1007 valyl-tRNA Sase
1077 seryl-tRNA Sase

Missing

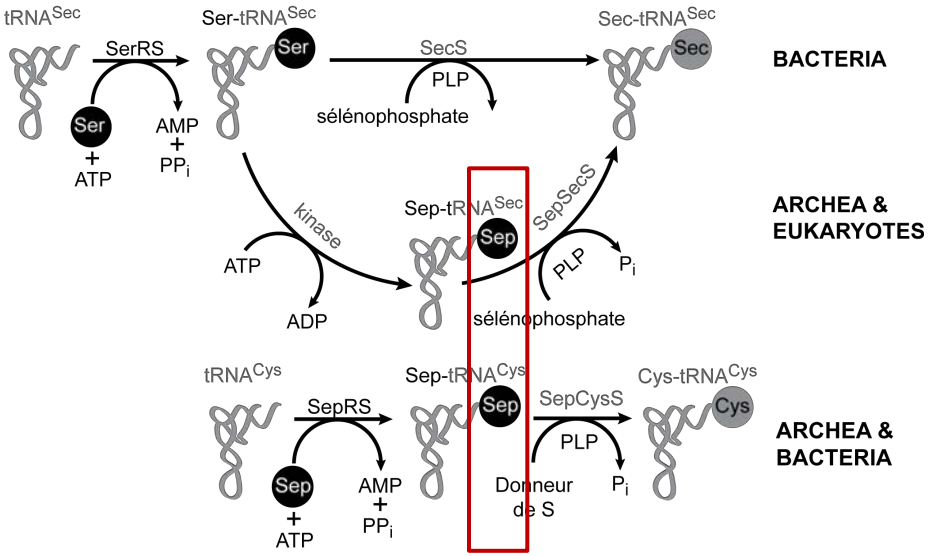
-AsnRS

-GlnRS

-LysRS → A Euryarchaeal Lysyl±tRNA Synthetase: Resemblance to Class I Synthetases

Michael Ibba, Susan Morgan, Alan W. Curnow, David R. Pridmore, Ute C. Vothknecht, Warren Gardner, Winston Lin, Carl R. Woese, Dieter Soll*

-CysRS



SCIENCE VOL 307 25 MARCH 2005

RNA-Dependent Cysteine Biosynthesis in Archaea

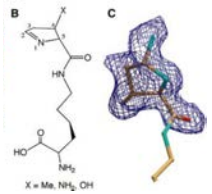
Anselm Sauerwald, [1] Wenhong Zhu, [3] Tiffany A. Major, [4] Hervé Roy, [5] Sotiria Palioura, [1] Dieter Jahn, [6] William B. Whitman, [4] John R. Yates 3rd, [3] Michael Ibba, [5] Dieter Söll [1,2]*

Natural redefinition of a STOP codon example of Pyl insertion

Requirements

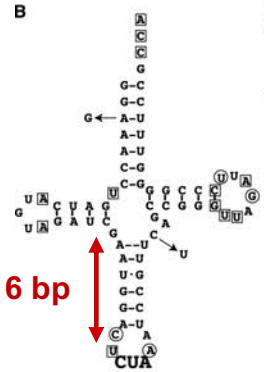
1. a new aa to add to the GC
2. a suppressor tRNA that will specify this new aa
3. an aaRS activating this new aa and charging the $tRNA^{Sup}$
4. an elongation factor that will bring the new aa-$tRNA^{Sup}$ to the ribosome (optional)
5. a message on the mRNA indicating to the ribosome that this STOP codon is not a STOP but encodes the new aa

1.



Pyl
UAG
(Amber)

2.

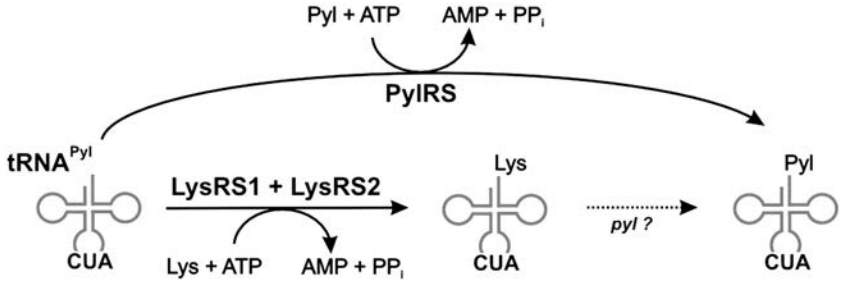


A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase

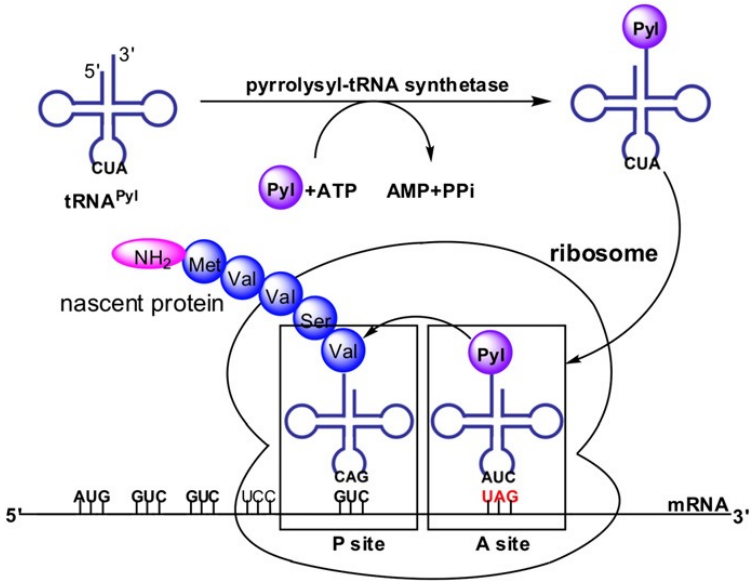
Bing Hao,[1] Weimin Gong,[1] Tsuneo K. Ferguson,[2] Carey M. James,[2] Joseph A. Krzycki,[2] Michael K. Chan[1]

24 MAY 2002 VOL 296 SCIENCE

3.



4 & 5.



Synthetic Biology

Designing and engineering of biological systems that aren't found in nature

Aims:

1. To understand natural systems.

By changing a system or making new ones or related ones you gain insights into understanding a system

To be able to predict the outcome when we change the system

1. To produce new biological systems for useful purposes.

Essentially for Biotechnology & Biomedicine

Synthetic Biology

Designing and engineering of biological systems that aren't found in nature

Aims:

1. To understand natural systems.

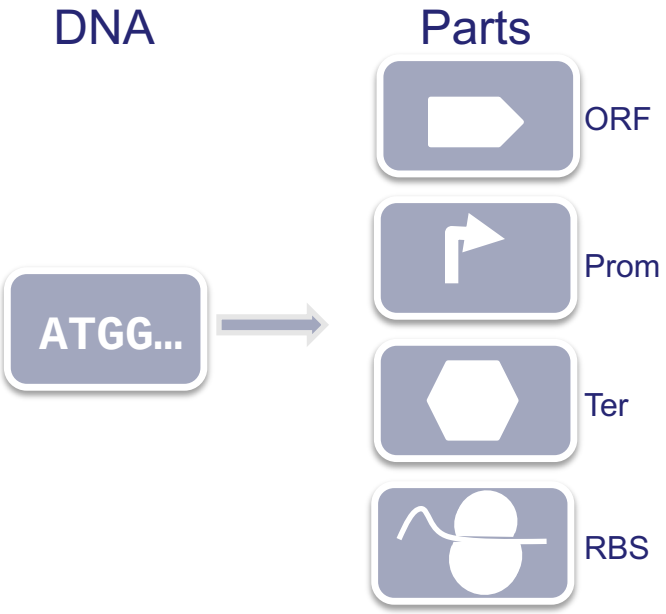
By changing a system or making new ones or related ones you gain insights into understanding a system

To be able to predict the outcome when we change the system

1. To produce new biological systems for useful purposes.

Essentially for Biotechnology & Biomedicine

Strategy: combining building blocks "Biobricks"



Synthetic Biology

Designing and engineering of biological systems that aren't found in nature

Aims:

1. To understand natural systems.

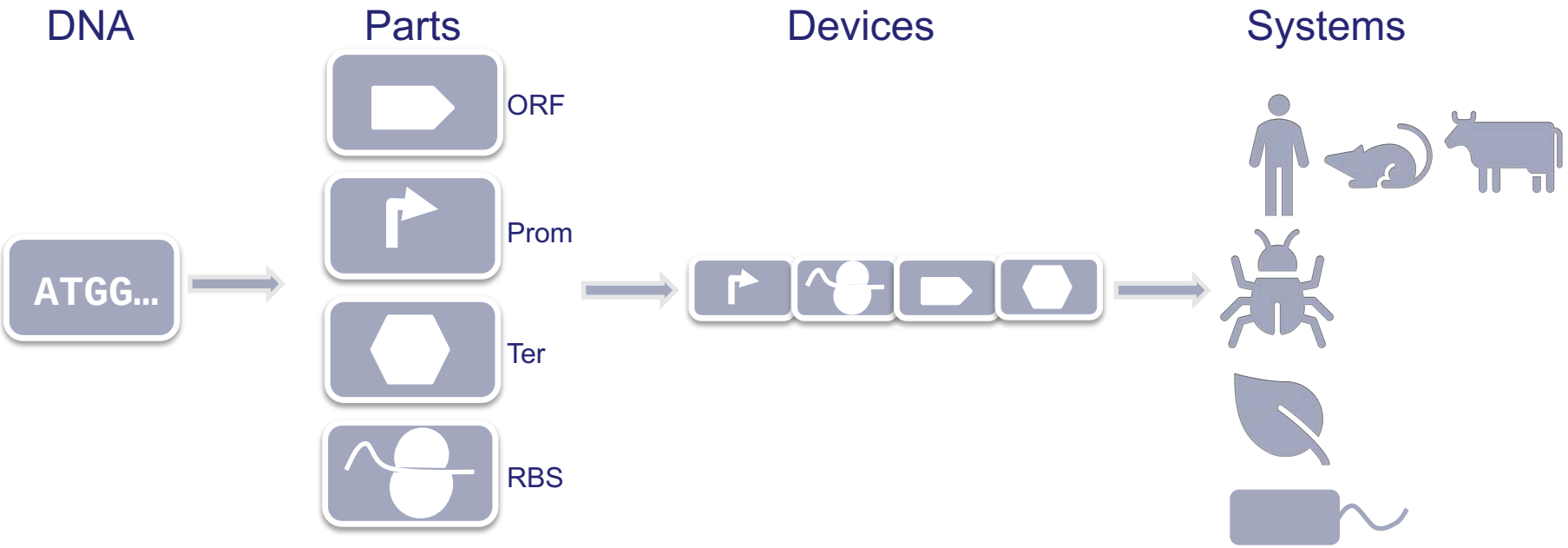
By changing a system or making new ones or related ones you gain insights into understanding a system

To be able to predict the outcome when we change the system

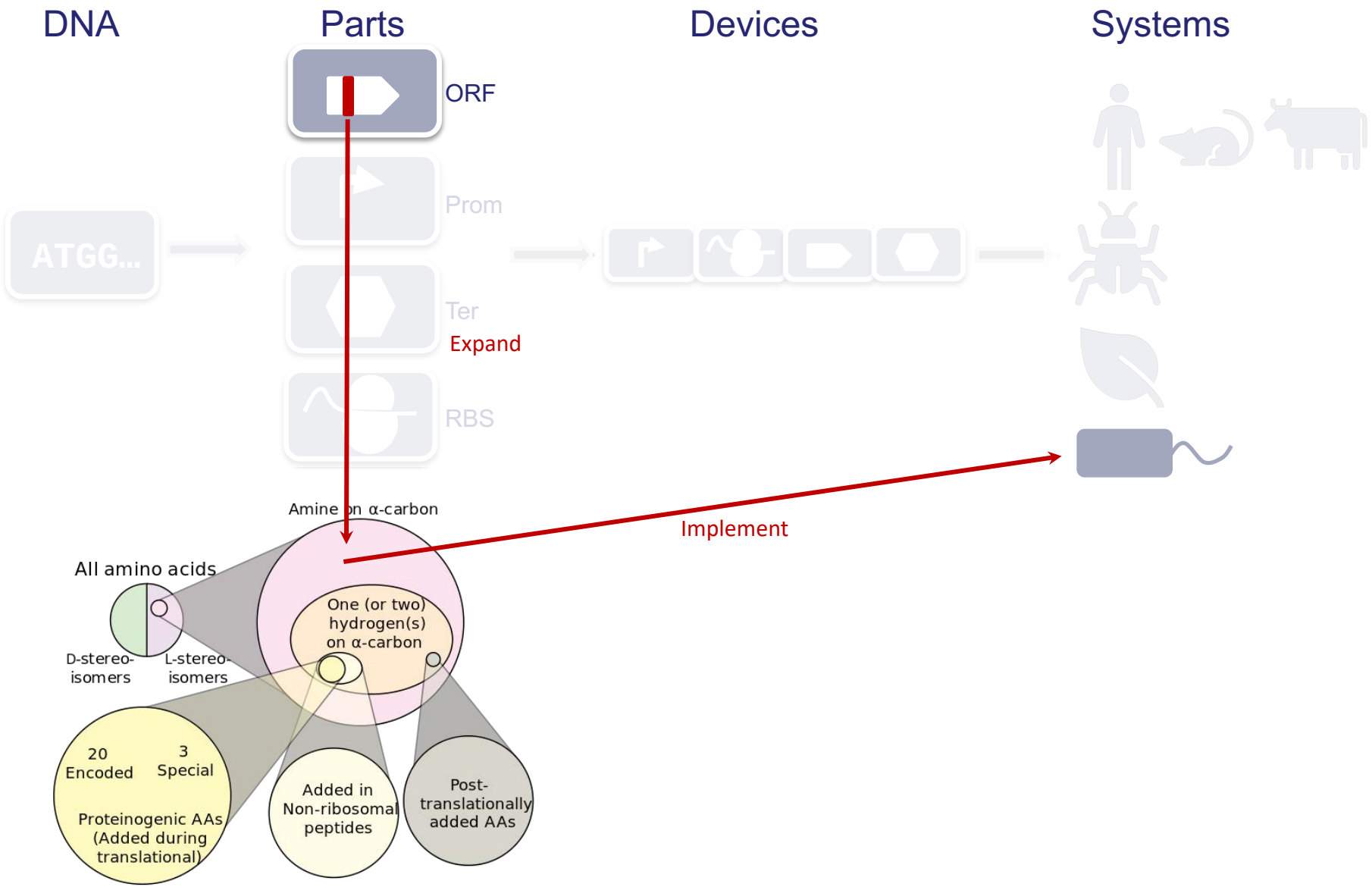
1. To produce new biological systems for useful purposes.

Essentially for Biotechnology & Biomedicine

Strategy: combining building blocks "Biobricks"



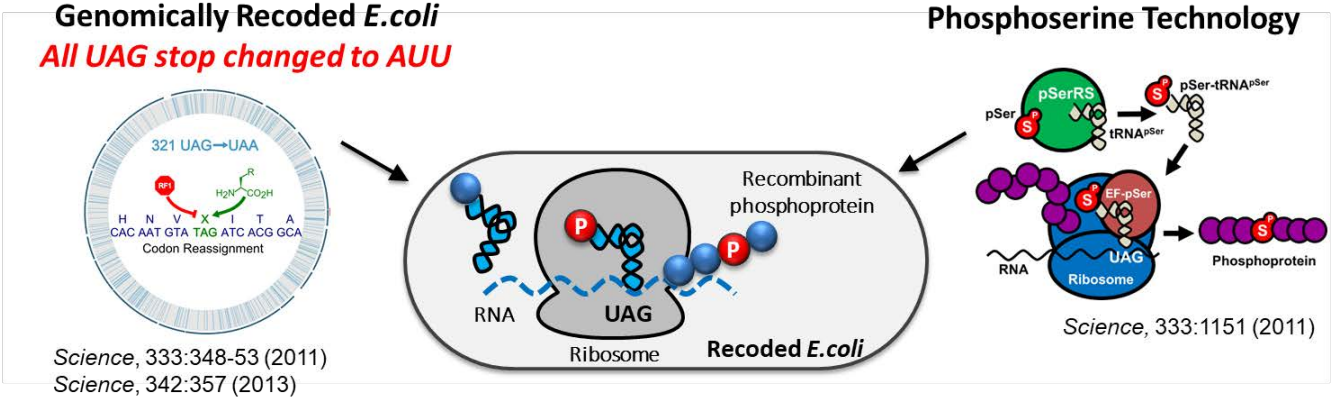
Synthetic Biology: expanding artificially the genetic code



Over 140 amino acids are known to occur naturally in proteins and thousands more may occur in nature or be synthesized in the laboratory.

Using SepRS for synthetic biology

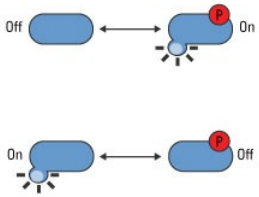
Phosphorylation *On Demand*



Science, 333:348-53 (2011)
Science, 342:357 (2013)

Science, 333:1151 (2011)

Gene Encodes
Phosphoserine

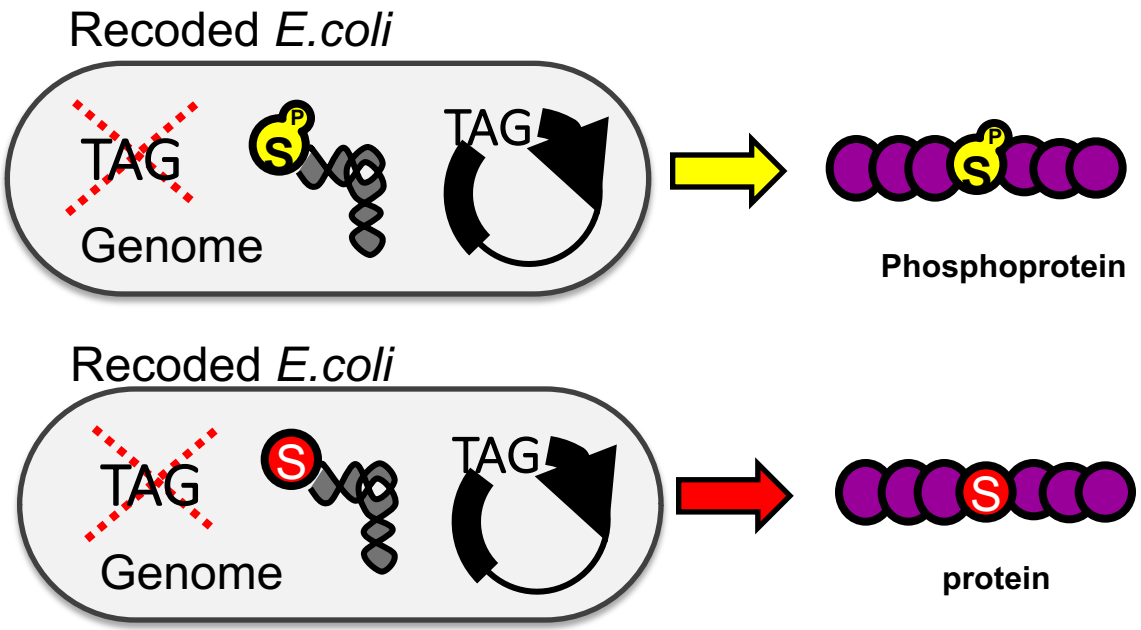


Pirman et al., ***Nature Commun.*** 2015

Oza et al., ***Nature Commun.*** 2015

Amiram et al. ***Nature Biotechnology*** *2015*

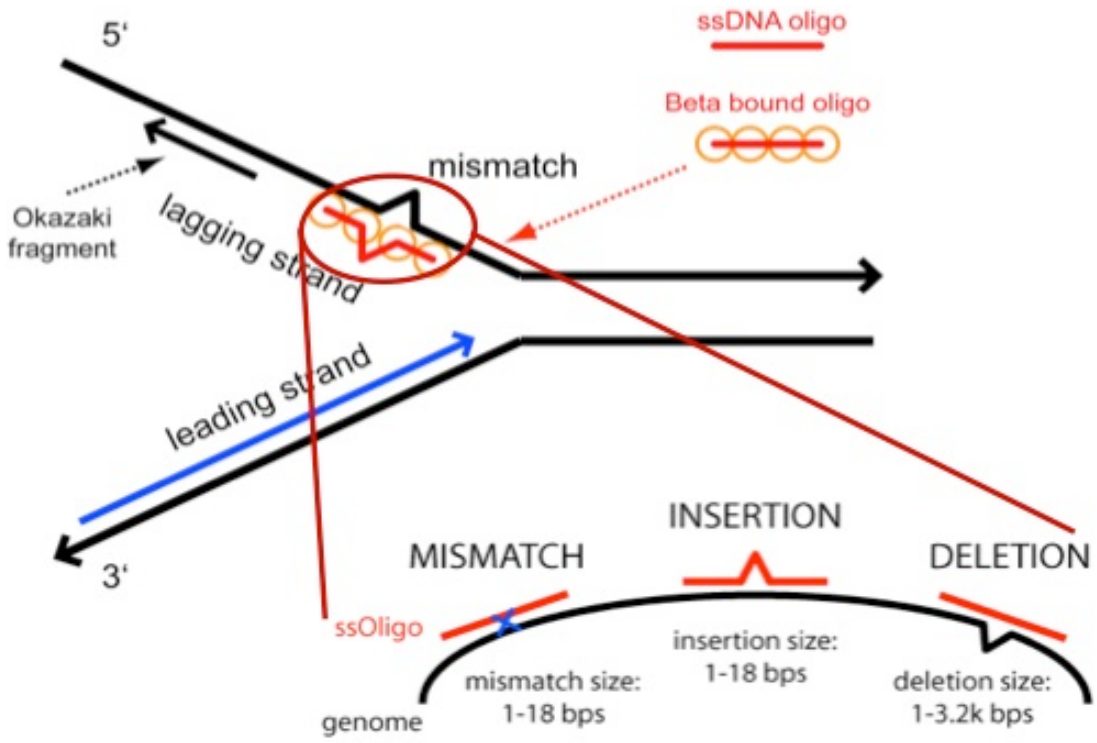
Using SepRS for synthetic biology



Using SepRS for synthetic biology

Whole genome engineering to delete RF1

Multiplex Automated Genome Engineering (MAGE)

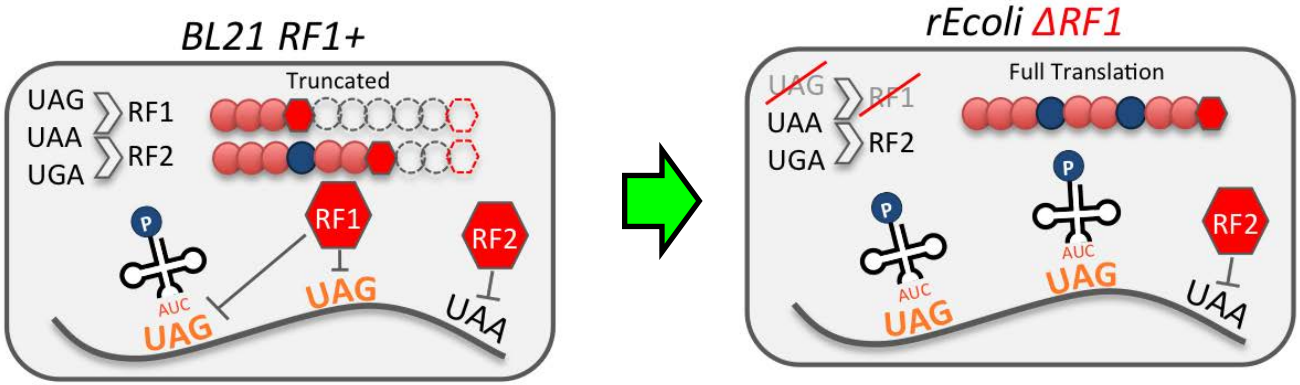


(Isaacs F, et al. 2011 *Science*.)
(Wang HH, Isaacs FJ, et al. 2009 *Nature*.)

The Rinehart Lab

Using SepRS for synthetic biology

Removing the amber STOP codon (UAG)

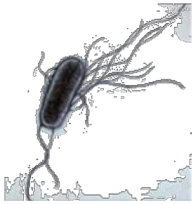


(Park *et al. Science* 2011)

(Lajoie *et al. Science* 2013)

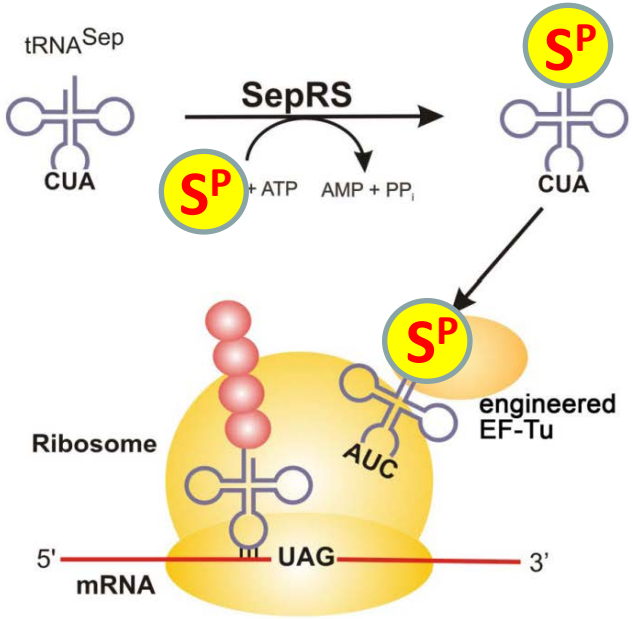
- **321 UAG stop changed to UAA**
- **RF1 deletion with no fitness defect**
- **No UAG in the genome**
- **UAG is now a dedicated sense codon**

Using SepRS for synthetic biology



Expanding the Genetic Code of *Escherichia coli* with Phosphoserine

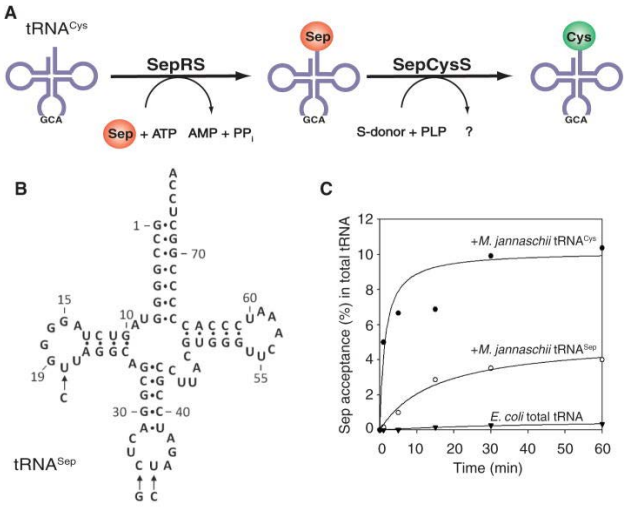
Hee-Sung Park,[1*†] Michael J. Hohn,[1*] Takuya Umehara,[1] Li-Tao Guo,[1] Edith M. Osborne,[2‡] Jack Benner,[2] Christopher J. Noren,[2§] Jesse Rinehart,[3,4§] Dieter Söll[1,5§]



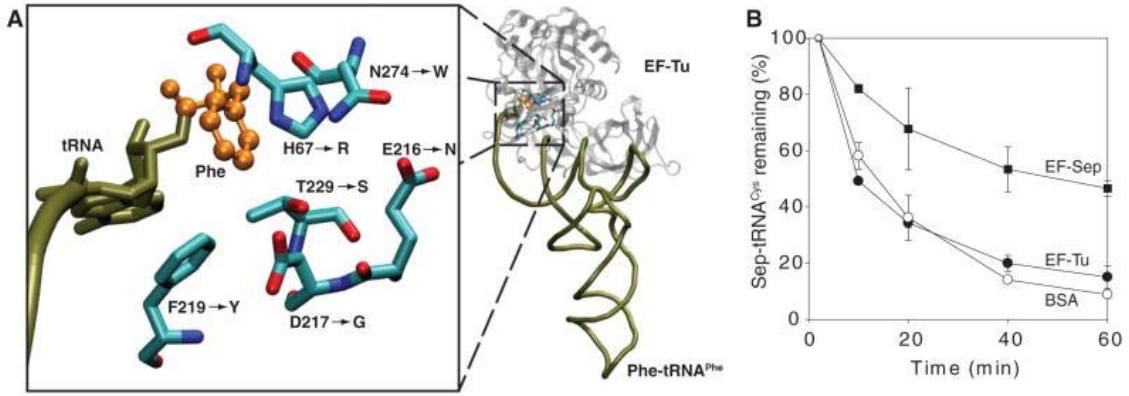
(Park *et al. Science* 2011)

U.S.S.N. 11/396,439 "Site Specific Incorporation of Phosphoserine into Polypeptides Using Phosphoseryl-tRNA Synthetase" by Dieter Söll and Jesse Rinehart

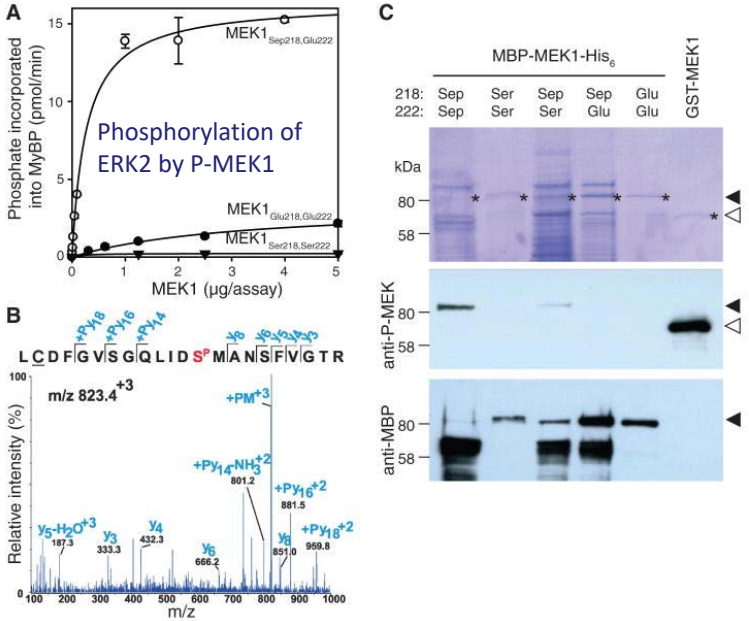
Using SepRS for synthetic biology



1. Transformation of $tRNA^{Sep}$ into $tRNA^{Cys}$



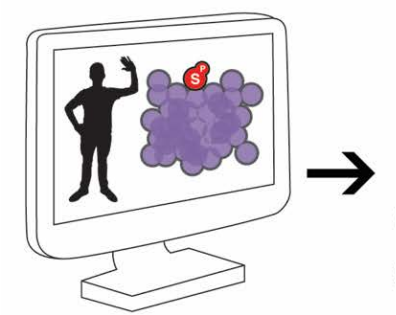
2. Transforming EF-Tu into EF-Sep that binds $Sep\text{-}tRNA^{Cys}$



3. Producing human Phosphorylated MEK1 *in E. coli*

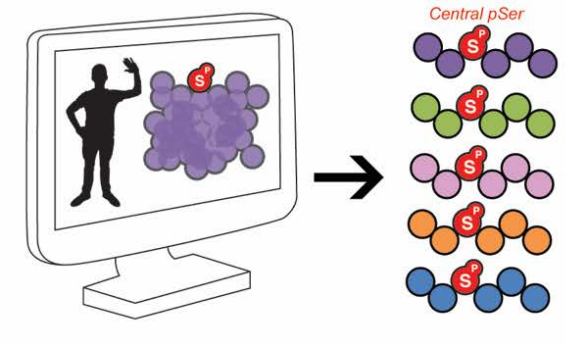
Using SepRS for synthetic biology

Designing phosphosites from human sequences

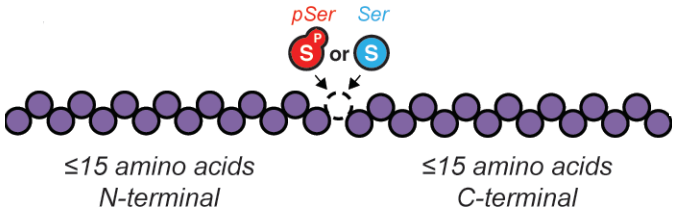


Downloaded all previously observed instances
of pSer observed in human proteins

(110,139 sites, by mass spectrometry)

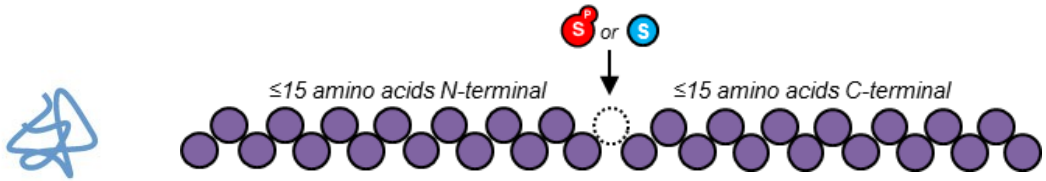


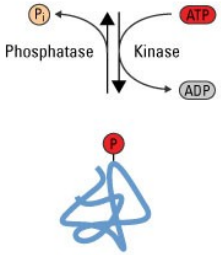
"Phosphosite" design



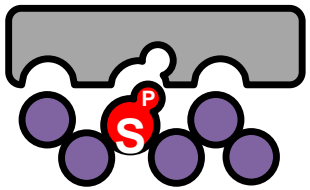
Using SepRS for synthetic biology

Designing phosphosites from human sequences

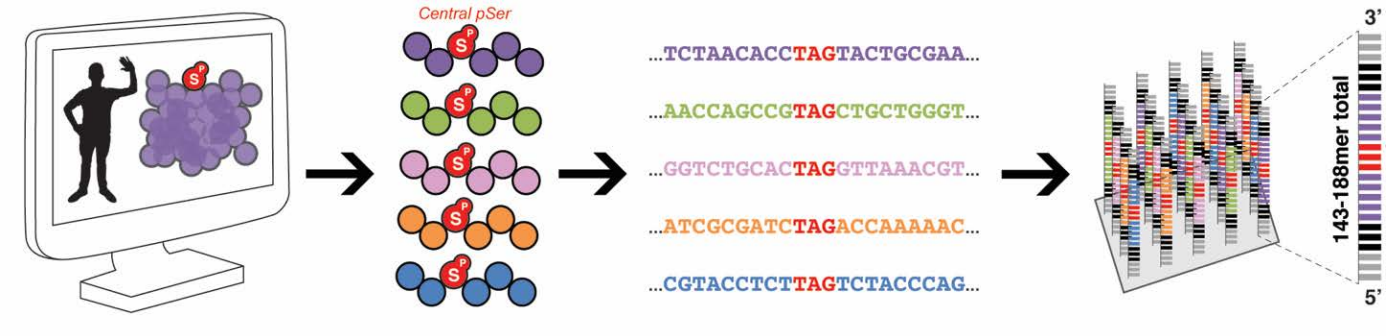




- **Sufficient for kinase & phosphatase recognition**
- **Sufficient for protein-protein interactions**

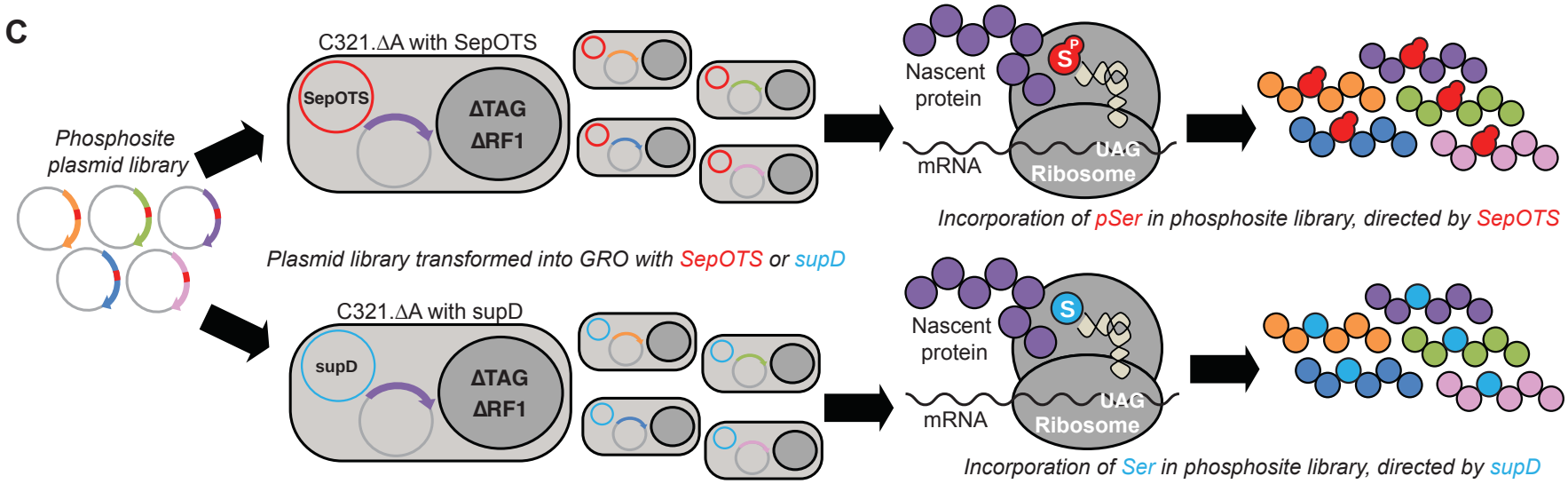


*Central **TAG** for pSer or Ser incorporation*



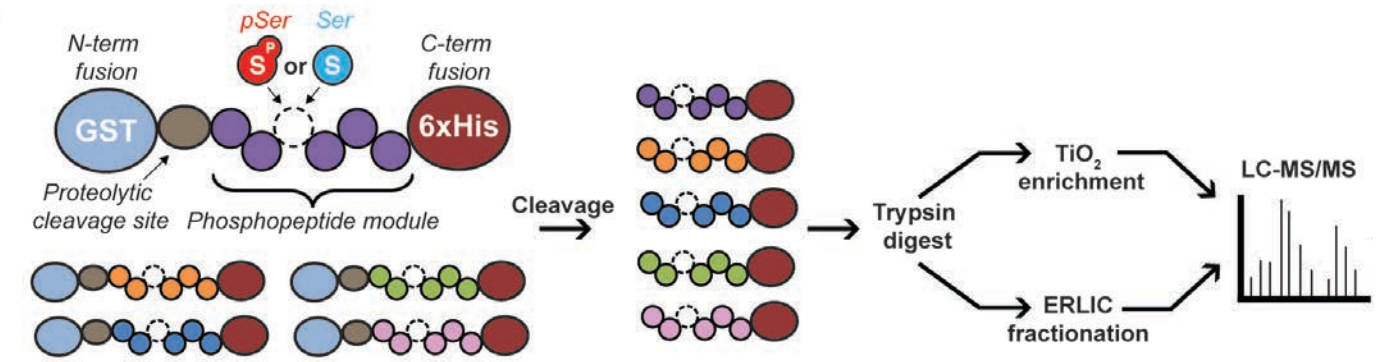
Using SepRS for synthetic biology

Creating the libraries of Sep or Ser peptides



Using SepRS for synthetic biology

Creating the libraries of Sep or Ser peptides



- *Direct evidence for >**36,000** phosphosites expressed in a single experiment*
- Direct evidence for **>71,000** proteins expressed

Using SepRS for synthetic biology

Application: validation of phospho-antobodies

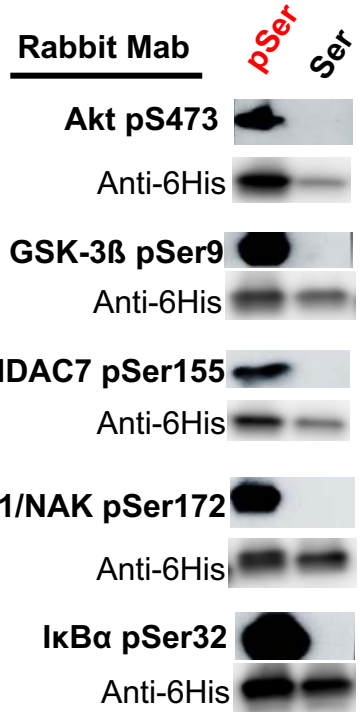
MECVDSERRPHFPQF**pS**YSASGTA

MSGRPRTT**pS**FAESCKPVQQPSAFG

LPSDPPEHFPLRKTV**pS**EPNLKLRYKPKKSLE

DFGAARELEDDEQFV**pS**LYGTEEYLHPDMYER

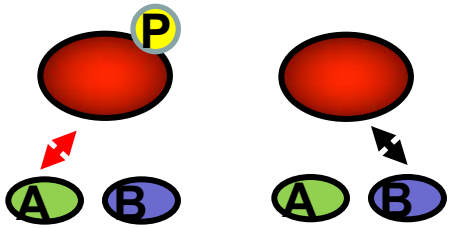
RDGLKKERLLDDRHD**pS**GLDSMKDEEYEQMVK



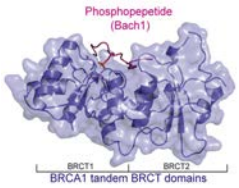
Using SepRS for synthetic biology

Application: decoding the phospho-interactome

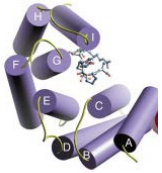
Protein-Protein Interactions



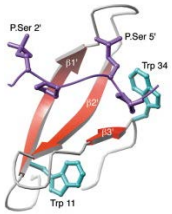
BRCT



14-3-3



WW



Bioplex interactome database

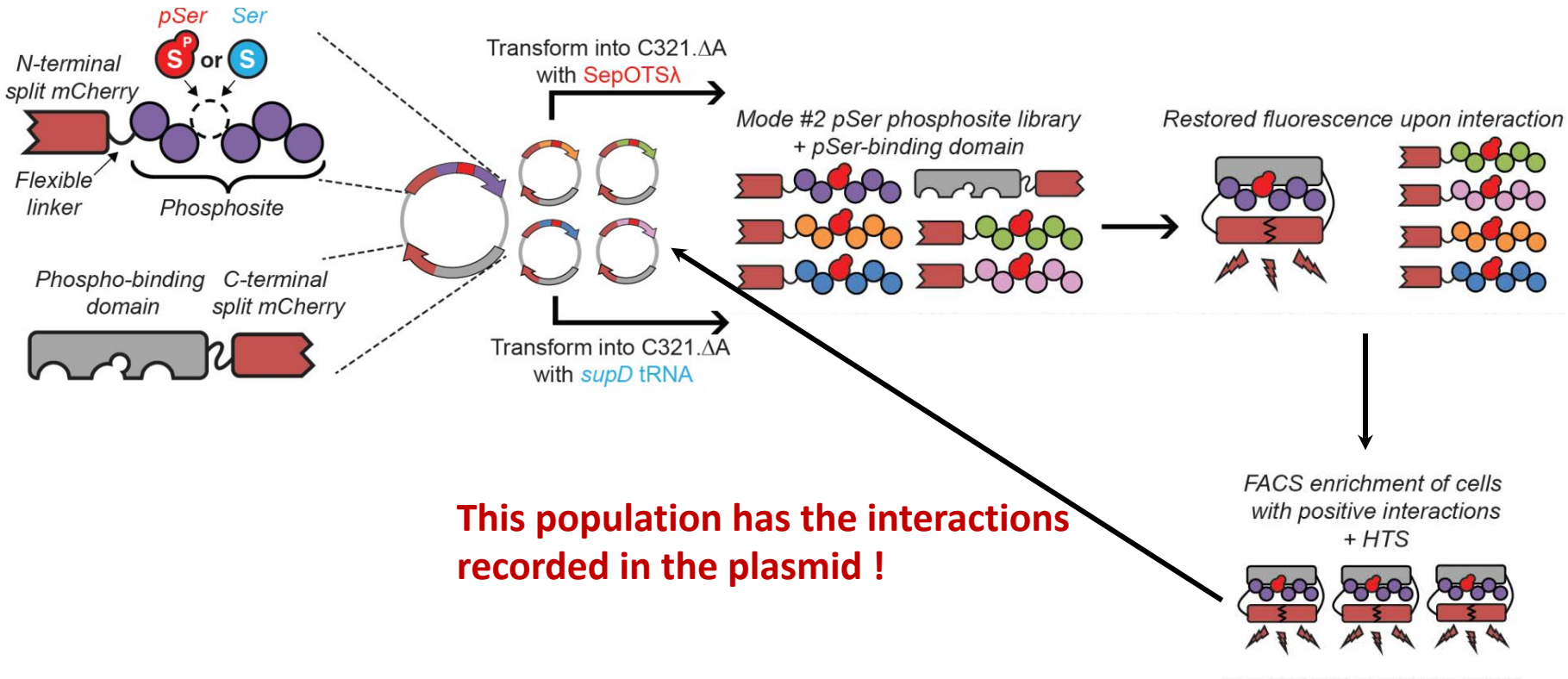
- **Huttlin, *Nature* 2017**
- **Huttlin, *Cell* 2015**

The Rinehart Lab

Using SepRS for synthetic biology

Application: decoding the phospho-interactome

Hi-P for proteome wide analysis of phosphorylation dependent protein-protein interaction

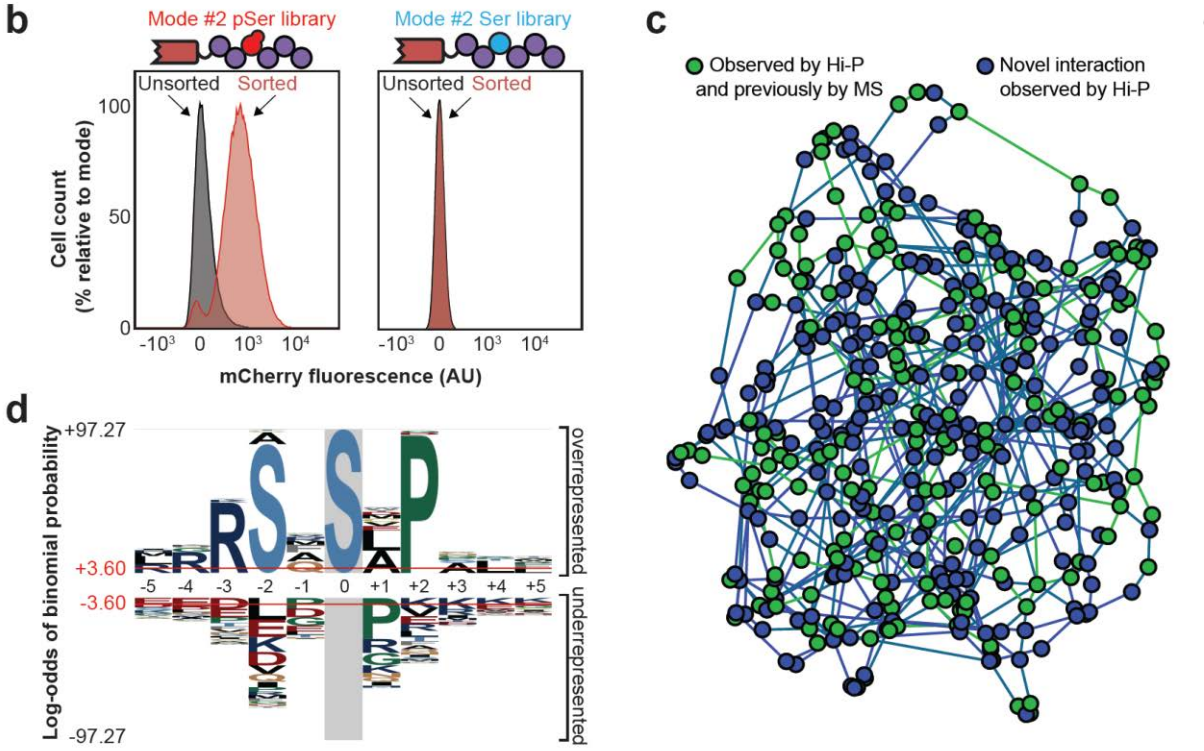


Using SepRS for synthetic biology

Application: decoding the phospho-interactome

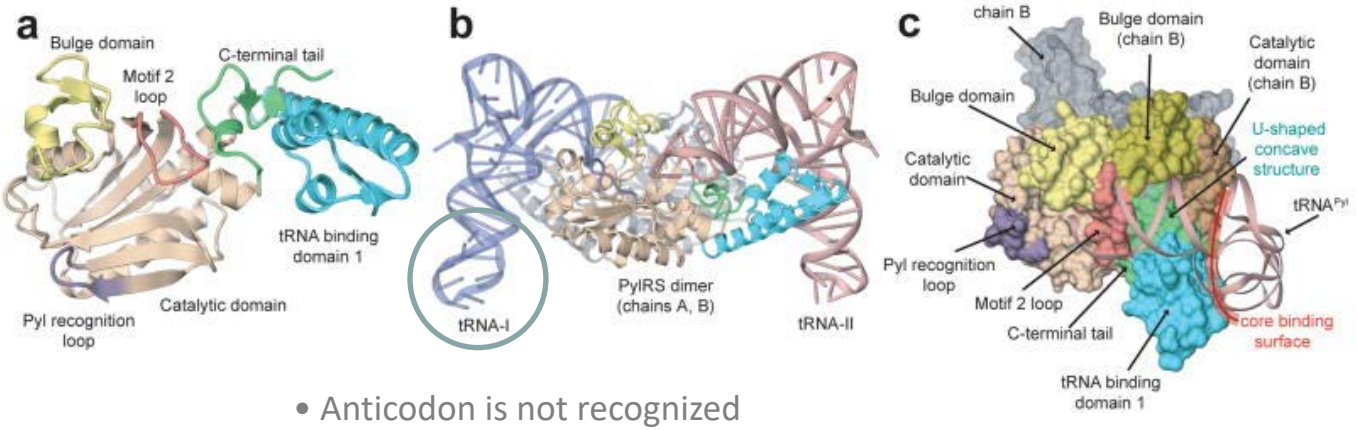
Hi-P for proteome wide analysis of phosphorylation dependent protein-protein interaction

Hi-P analysis of the 14-3-3β interactome



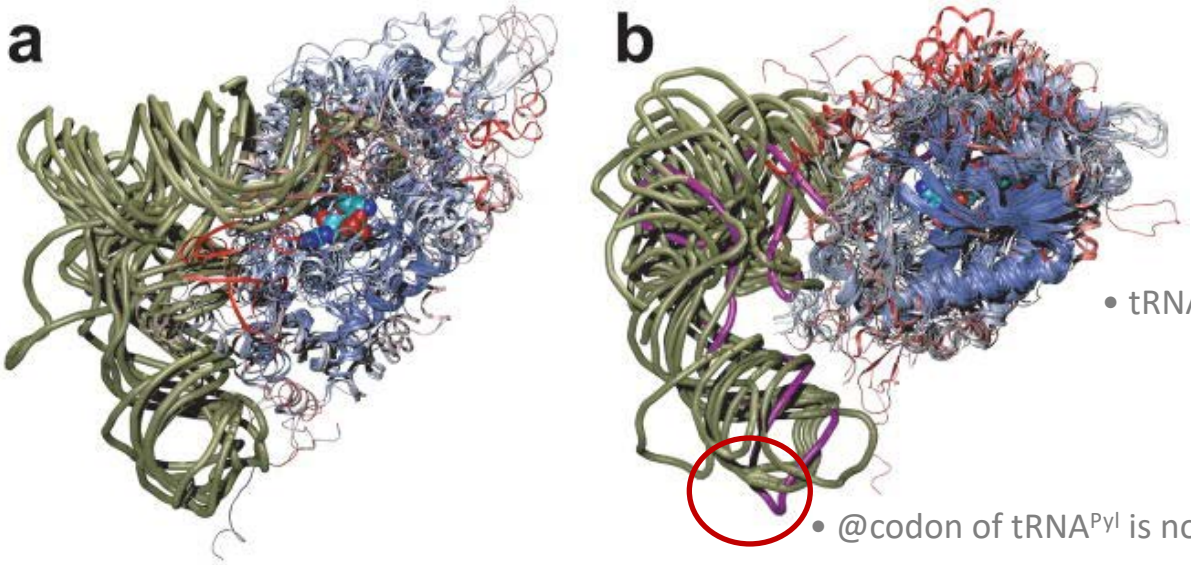
Orthogonality of the PylRS/tRNAPyl pair

An 'orthogonal' tRNA synthetase-tRNA pair is an engineered aminoacyl-tRNA synthetase (aaRS) specifically and exclusively acylates the orthogonal tRNA with a non-canonical amino acid.



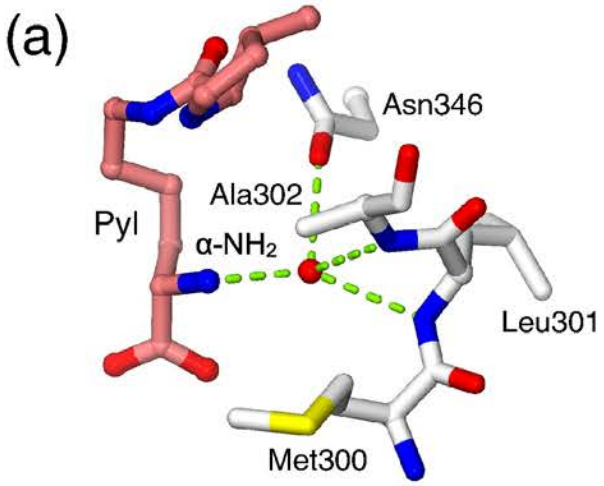
- Anticodon is not recognized
- Allows engineering of $tRNA^{Pyl}$ binding to UAG, UGA and UAA STOP codons

- Class I aaRS/tRNA surperimpositions
- Class II aaRS/tRNA surperimpositions

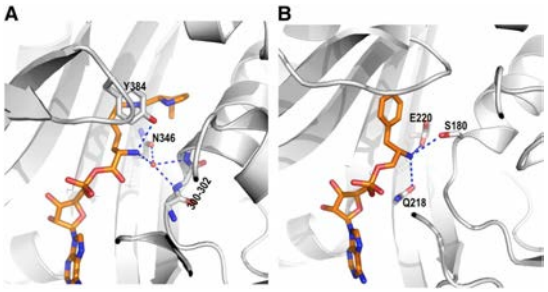


- $tRNA^{Pyl}$ (purple) recognition is unique
- @codon of $tRNA^{Pyl}$ is not facing PylRS

Orthogonal PylRS/tRNAPyl pair allows incorporation of NNAAs



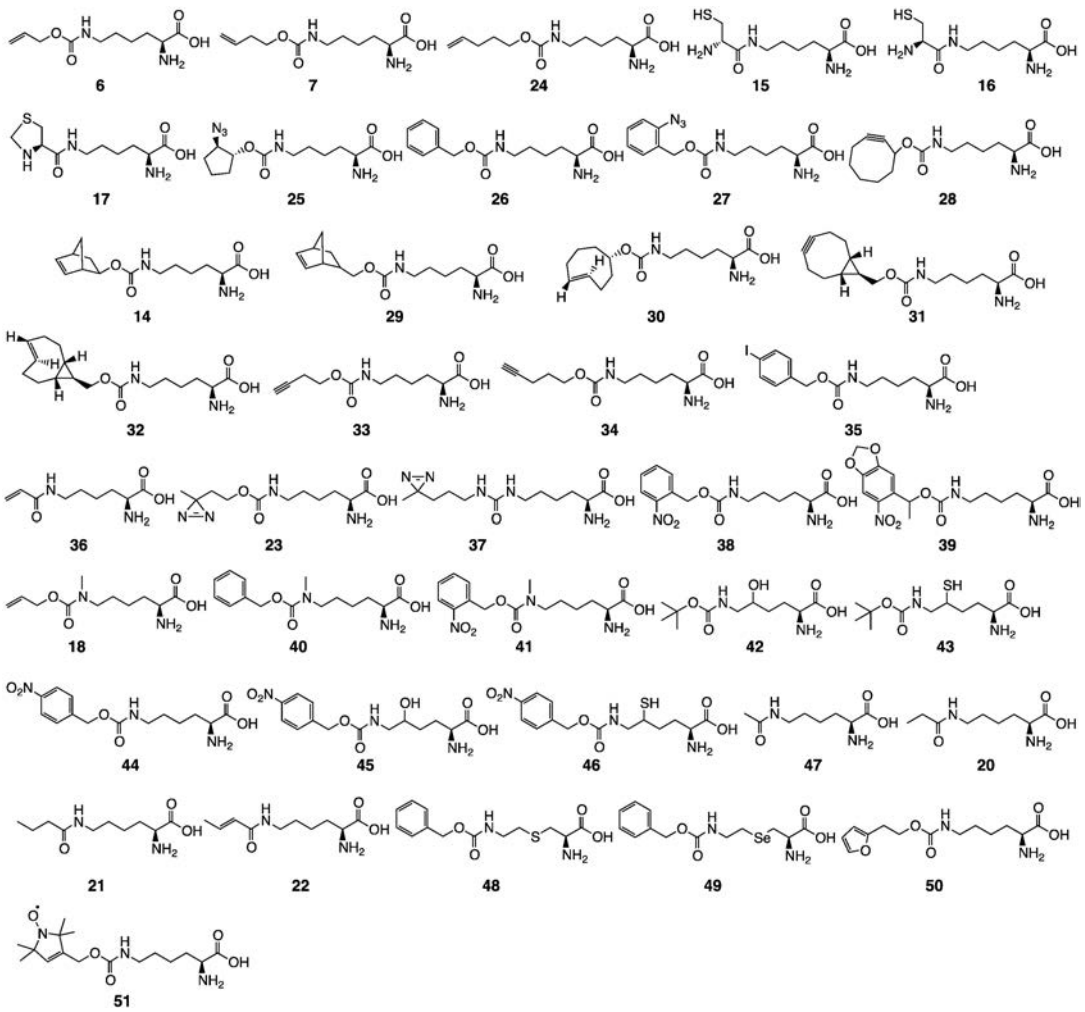
• essentially α-NH2 recognized



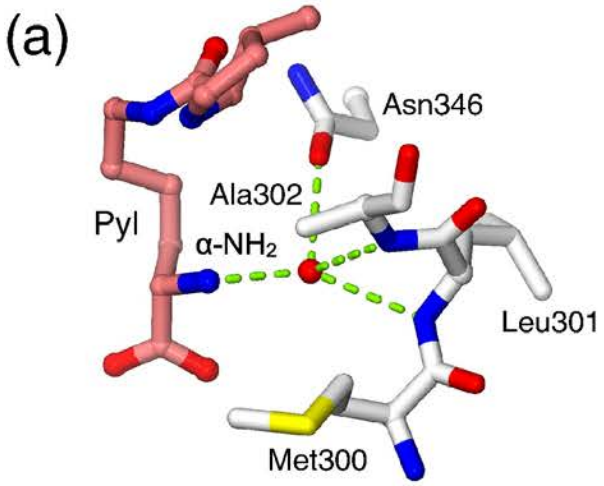
• Pyl derivatives recognition

• Phe derivatives recognition

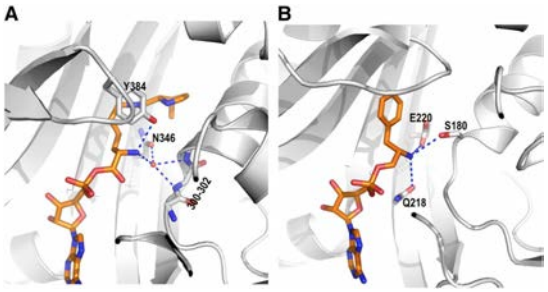
• **Lys** derivatives incorporated *in vivo*



Orthogonal PylRS/tRNAPyl pair allows incorporation of NNAAs



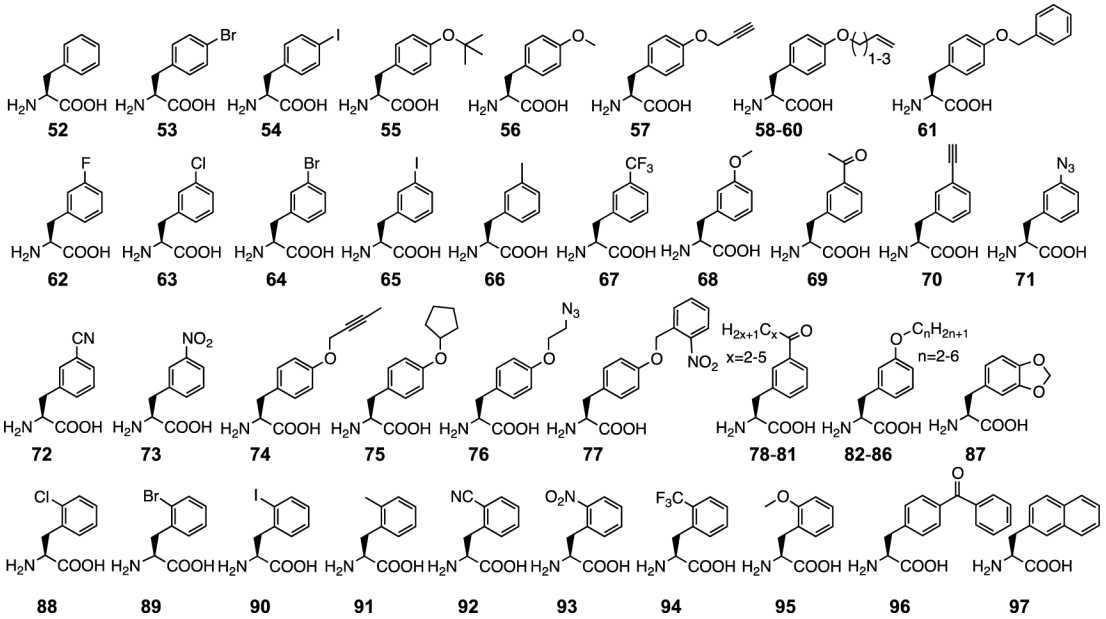
• essentially α-NH2 recognized



• Pyl derivatives recognition

• Phe derivatives recognition

• **Phe** derivatives incorporated *in vivo*



Orthogonal PylRS/tRNAPyl pair allows incorporation of NNAAs

Table 1
Sequence information of PylRS mutants that recognize a variety of NCAAs as substrates.

Substrate	Sequence												
	M276	L301	A302	L305	Y306	L309	N346	C348	M350	Y384	V401	W417	M. mazei [a]
	M241	L266	A267	L270	Y271	L274	N311	C313	M315	Y349	V367	W383	M. barkeri [a]
6, 7, 14, 25, 98–101						A		A		F			M. barkeri
						A		V		F			M. barkeri
15, 16, 17			S					V	F				M. barkeri
17								V					M. barkeri
14, 26–31, 50					A					F			M. mazei
26						A		V					M. mazei
29					G					F			M. mazei
31, 32					M	G		A					M. barkeri
30, 32					A	M		A					M. barkeri
33, 34, 35, 37						A		A		F			M. barkeri
						A		S		F			M. barkeri
36, 20		M			L	A		F		W			M. mazei
23						M		A		F			M. barkeri
38					M	A		A		F			M. mazei
					I	A		A		F			M. mazei
39	F		S		C	M		C					M. barkeri
18										F			M. barkeri
26, 38, 40, 41, 48, 49					M	A		T					M. mazei
					M	A		C					M. mazei
					M	P		C					M. mazei
41					I	M		A					M. barkeri
42, 43										W			M. barkeri
44					M	G		A					M. barkeri
45, 46					M	G		A		W			M. barkeri
20, 36, 47, 102		M			L	A		F					M. mazei
		M			L	L		S					M. mazei
		V		I	F	A		F					M. barkeri
		L		I	L	A		F					M. barkeri
		M		I	F	A		F					M. barkeri
51					A	M				F			M. mazei
21, 22										W			M. mazei
						A		F		F			M. barkeri
52							A	L					M. mazei
							A	K					M. mazei
53, 54			L		M		S	L		L			M. mazei
			F		L		T	F		L			M. mazei
				M	L	S	S	M					M. mazei
55–95							A	A					M. mazei
56			T				V	W		F	L		M. mazei
77				F	M		G	G					M. barkeri
96			T				T	T					M. mazei
97			T				G	T			I	Y	M. mazei

[a] Native enzyme origins.