Avant propos

Ce diaporama correspond à une leçon donnée Le 11 mai 2011 par Henri Grosjean, Directeur de recherche émerite du CNRS et ex-Prof de l'Université de Bruxelles en Belgique, dans le cadre d'une série de cours et de conférences organisées par le Prof. Marc Fontecave au Collège de France (chaire de Chimie des Processus Biologiques, année académique 2010-11)

Sur le thème:

Chimie Biologique Radicalaire: de l'origine de l'ADN au métabolisme d'aujourd'hui

Collège de France

Chaire de Chimie des Processus Biologiques

Modification et Edition Post-transcriptionelles des Acides Nucléiques (ADN mais surtout ARN)

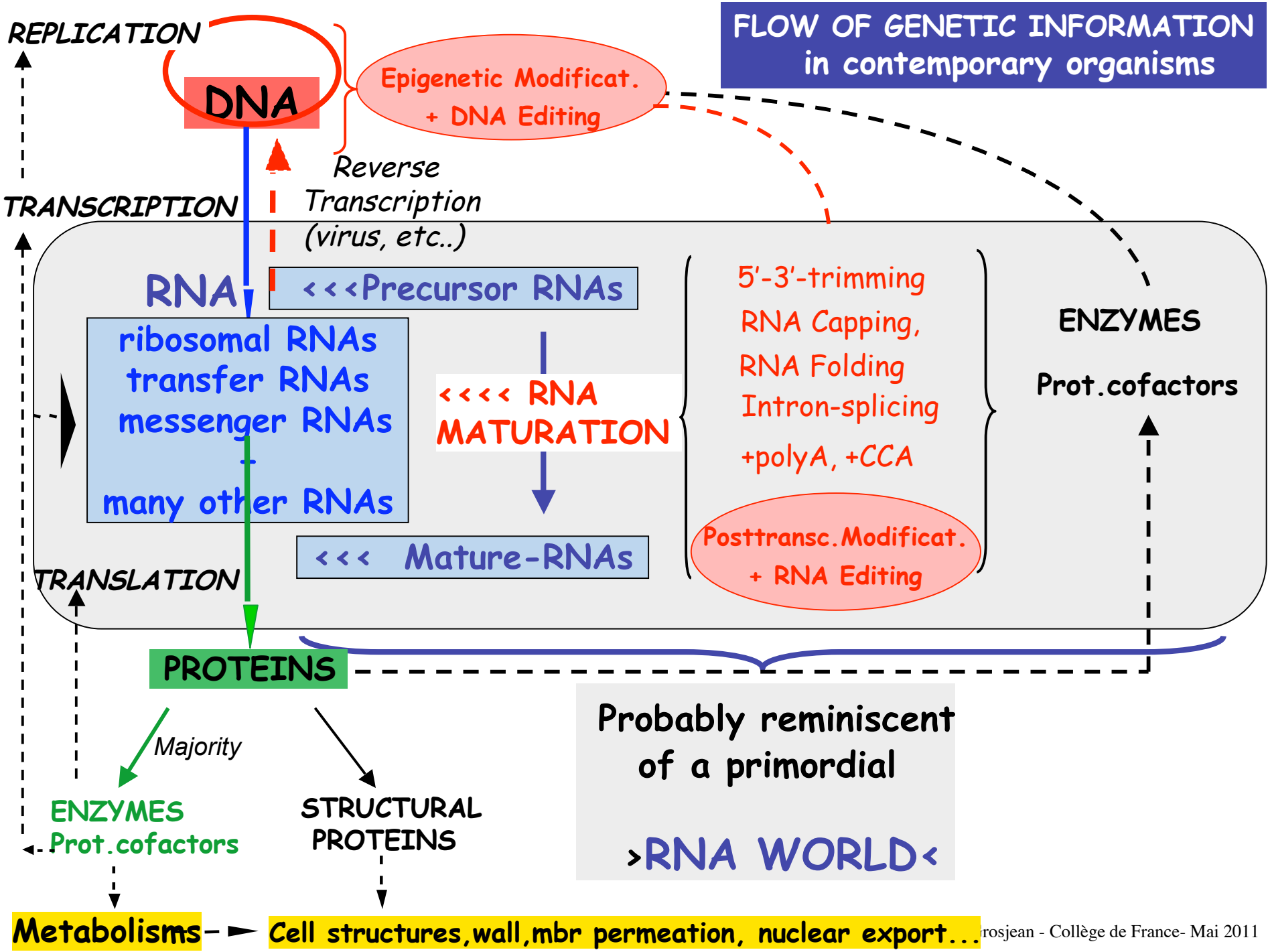
Henri GROSJEAN - 11 mai 2011

henri.grosjean@igmors.u-psud.fr ou henri4g@me.com

L'exposé a été présenté en deux parties:

1- Généralités sur les nucléotides modifiés dans les ARNs (et ADN), ainsi que leurs enzymes de modification correspondants

2- Importance de ces nucléotides modifiés pour la traduction génétique - en particulier pour l'émergence d'un code génétique 'quasi universel' tel que nous le connaissons actuellement


FLOW OF GENETIC INFORMATION
in contemporary organisms
REPLICATION
DNA
Epigenetic Modificat.
+ DNA Editing
Reverse
Transcription
(virus, etc..)
TRANSCRIPTION
RNA
<<<Precursor RNAs
ribosomal RNAs
transfer RNAs
messenger RNAs
-
many other RNAs
<<<< RNA
MATURATION
<<< Mature-RNAs
5'-3'-trimming
RNA Capping,
RNA Folding
Intron-splicing
+polyA, +CCA
Posttransc.Modificat.
+ RNA Editing
ENZYMES
Prot.cofactors
TRANSLATION
PROTEINS
Majority
ENZYMES
Prot.cofactors
STRUCTURAL
PROTEINS
Probably reminiscent
of a primordial
>RNA WORLD<
Metabolisms
Cell structures,wall,mbr permeation, nuclear export...

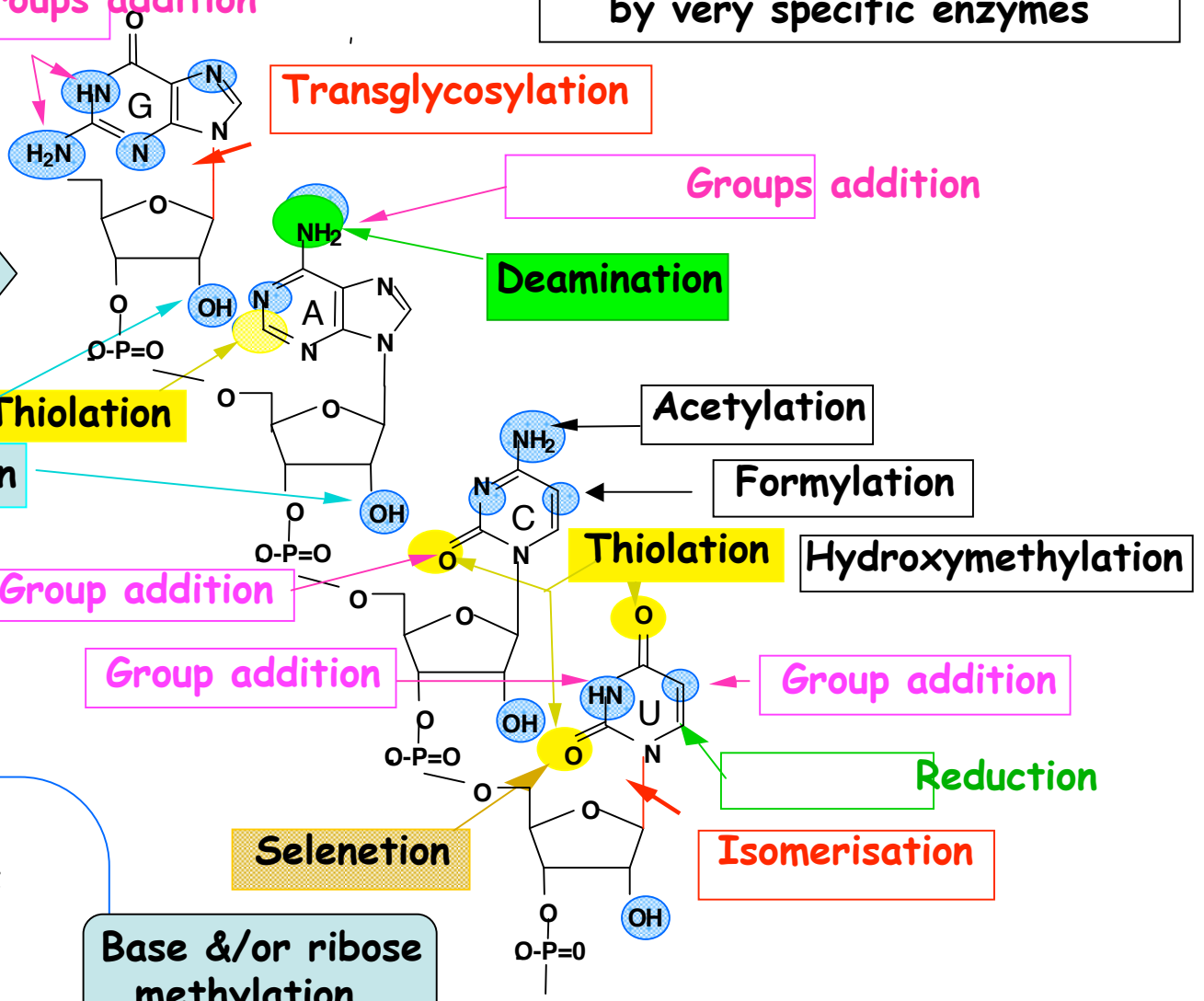
Post-transcriptional modifications of bases and ribose in RNA

= chemical reactions catalysed by very specific enzymes

Transcription from DNA genes

Polymerase machinery including many co-factors

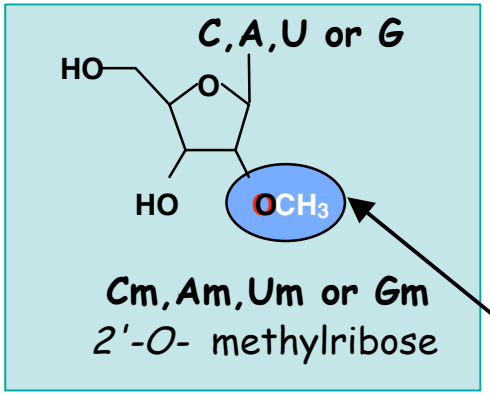
ATP+GTP+ CTP + UTP



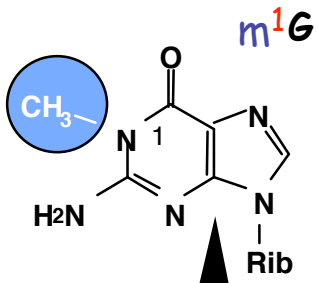
Together with many other processes of RNA maturation,

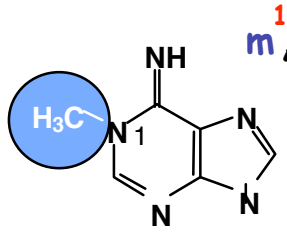
- The most diversified type of modifications is RNA methylation

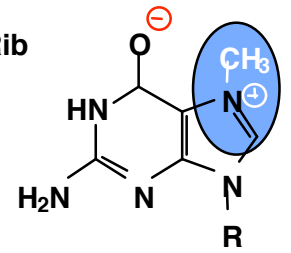
Examples of base &/or ribose methylation in RNA

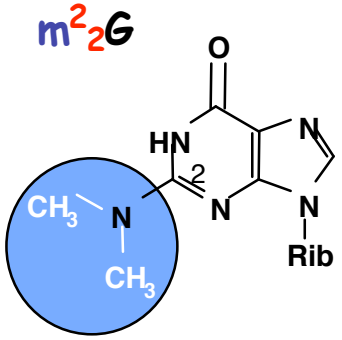


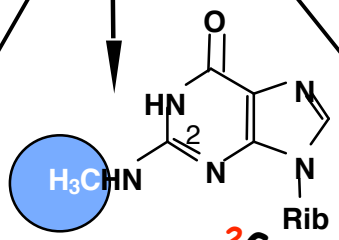
Several METHYLASES

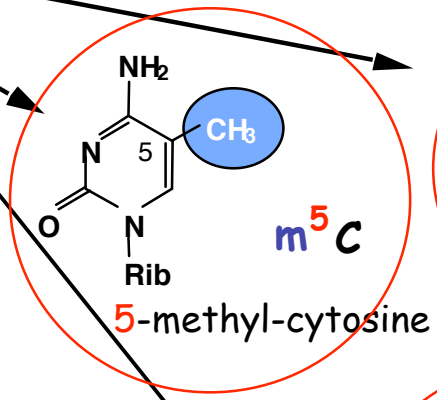


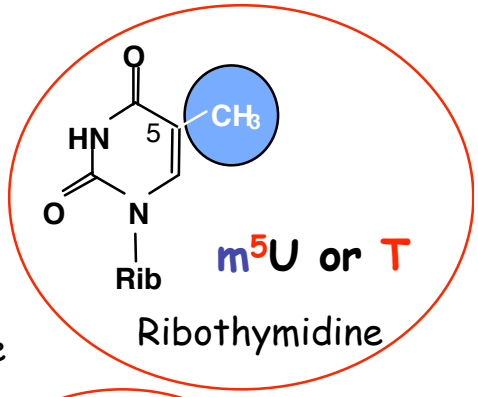


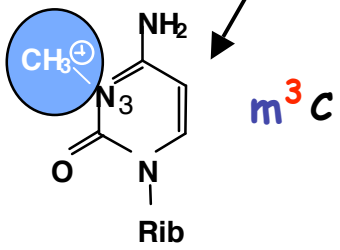


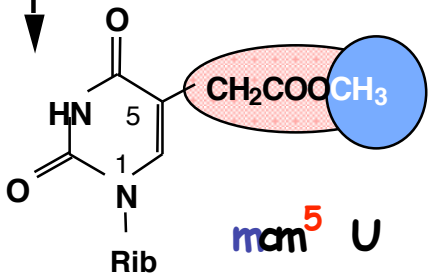


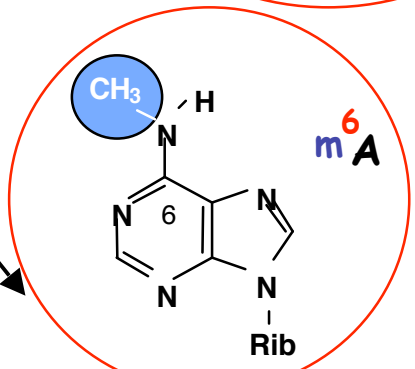




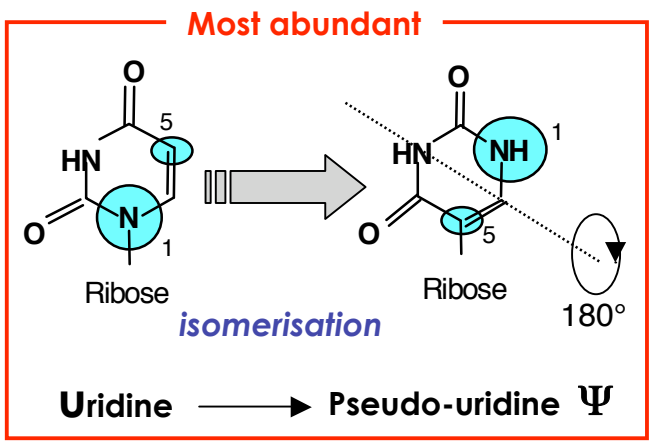


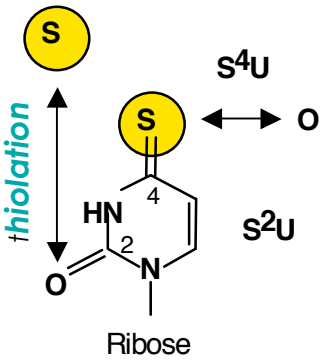


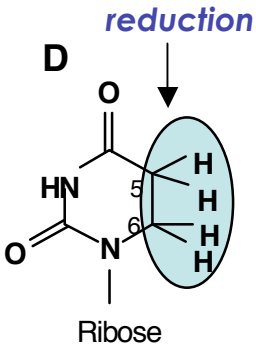


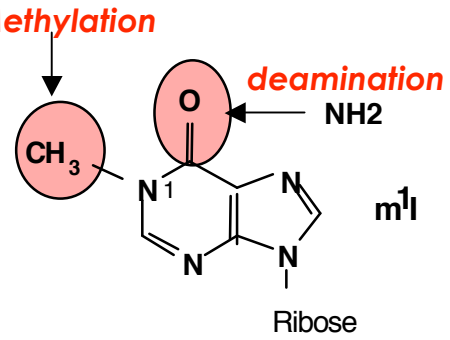


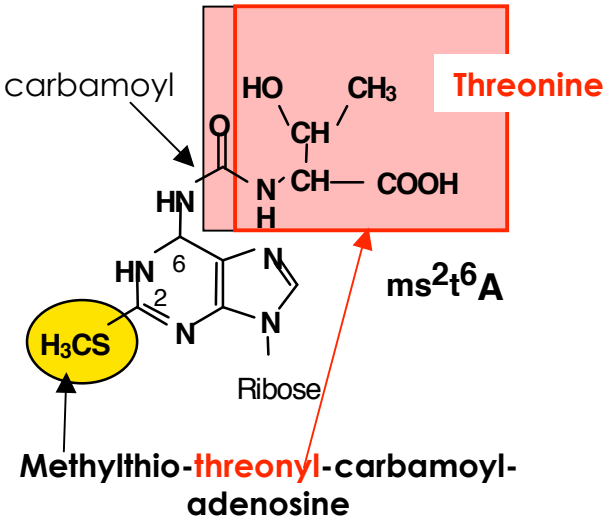
Other 'simple' Modifications in RNA

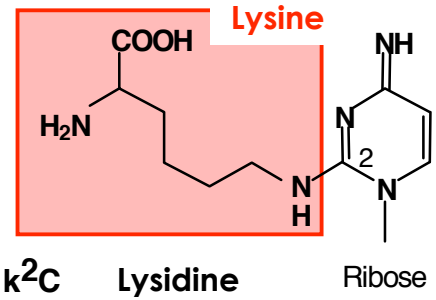


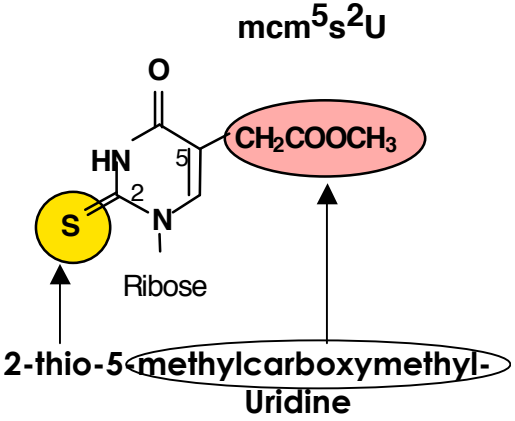




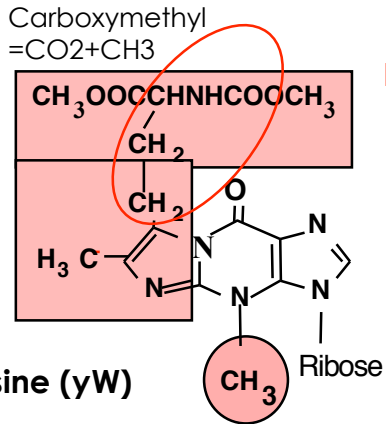








Hyper-modifications



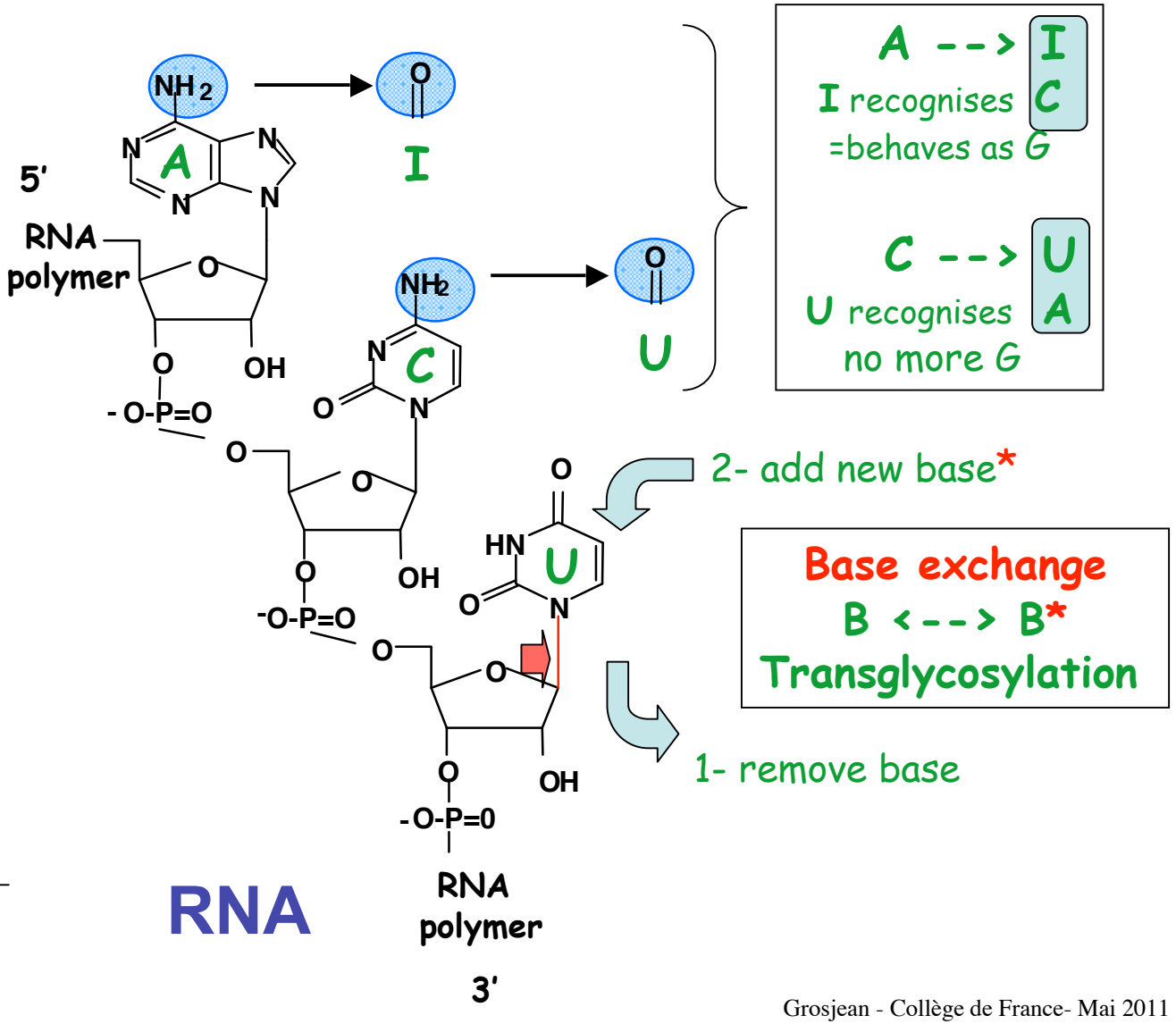
From methionine (acp)

More 'complex' modifications in RNA

RNA 'EDITING'

= BASE CONVERSION

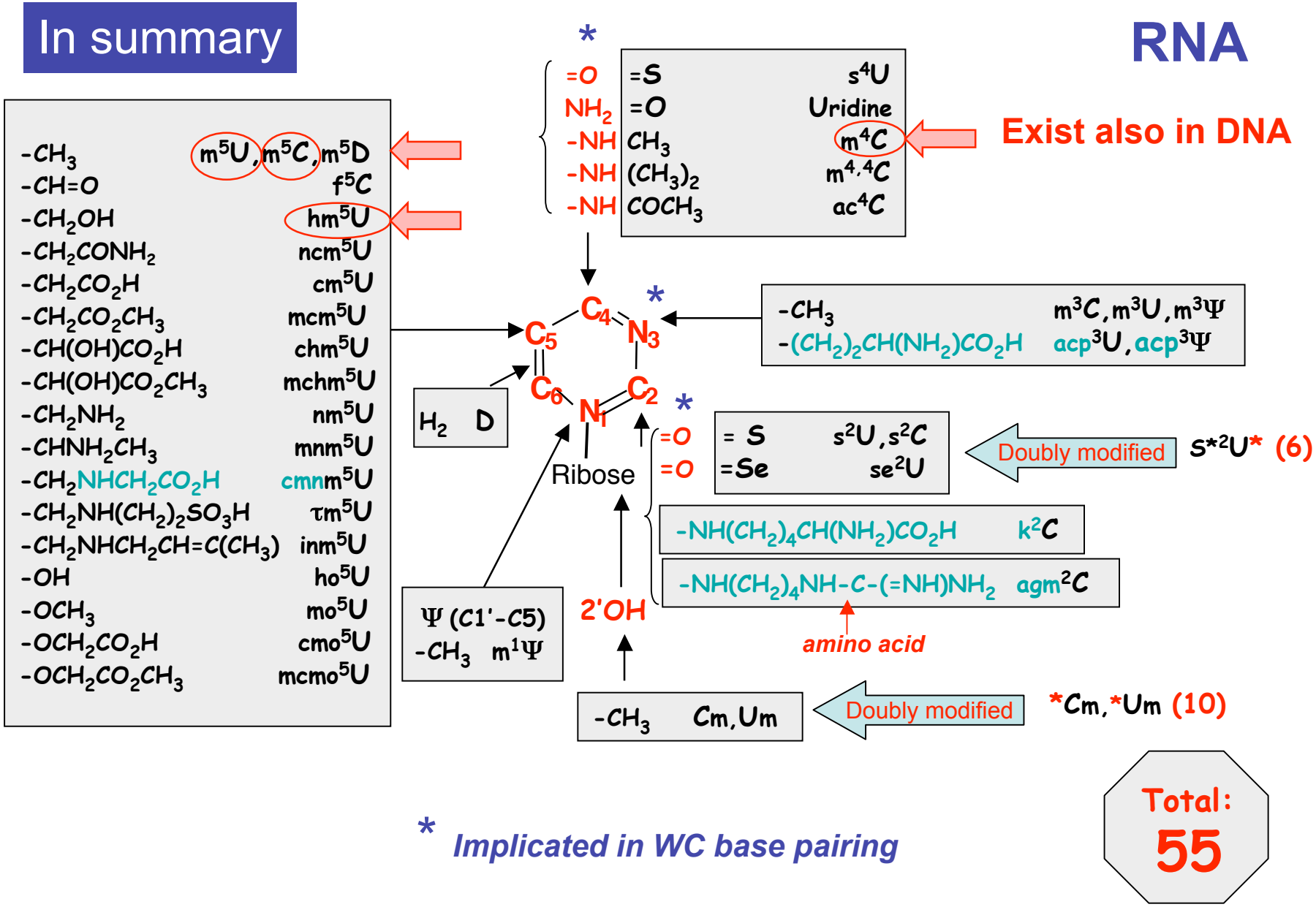
Deamination enzymatique = RNA modification



Same types of base conversion also exist for DNA but require distinct enzymes

DNA 'EDITING'

In summary



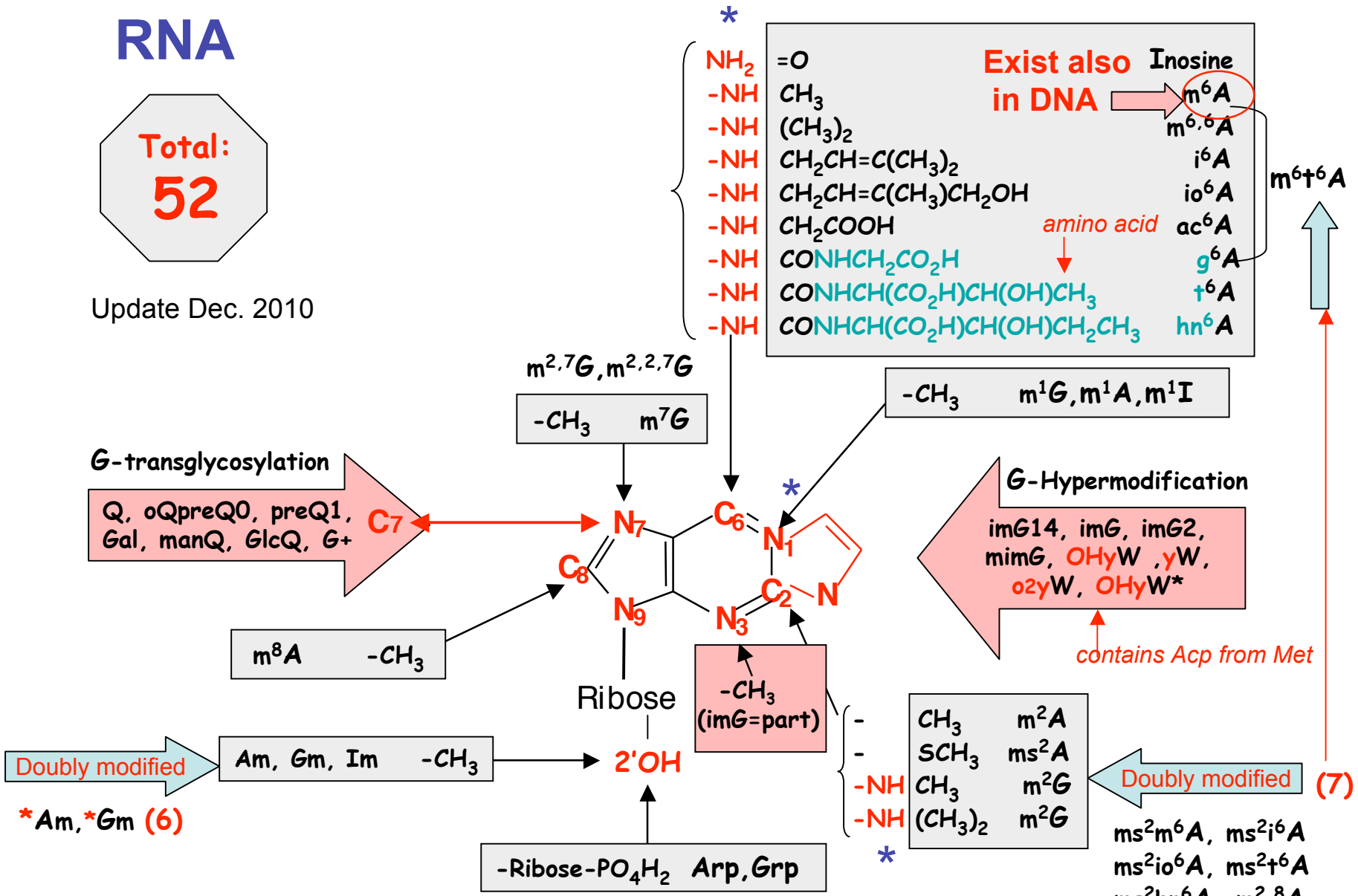
Pyrimidine derivatives in RNA

Update Dec. 2010

RNA

Total: 52

Update Dec. 2010



Purine derivatives in RNA

* *Implicated in WC base pairing*

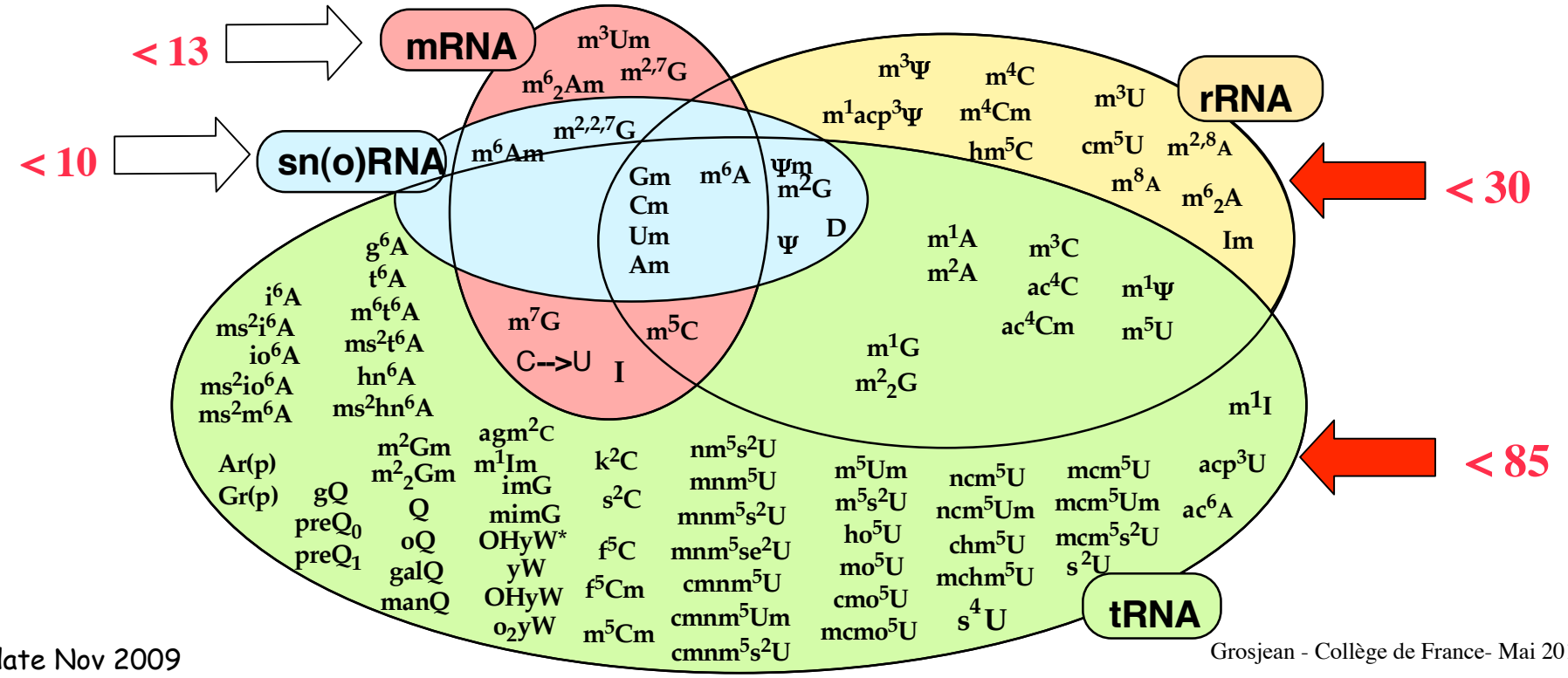
Total of **107** distinct Mod Nucl.

Occurrence of modified nucleosides according to the type of cellular RNAs

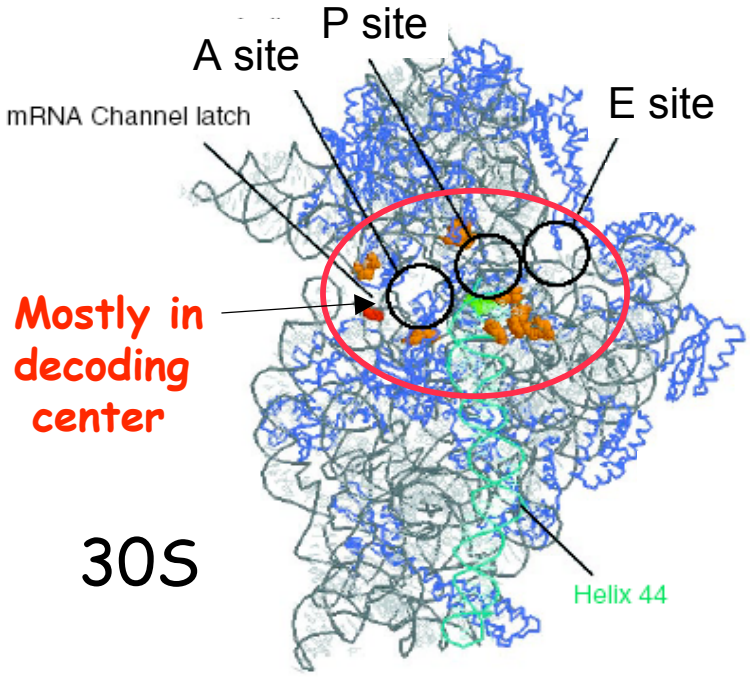
RNA

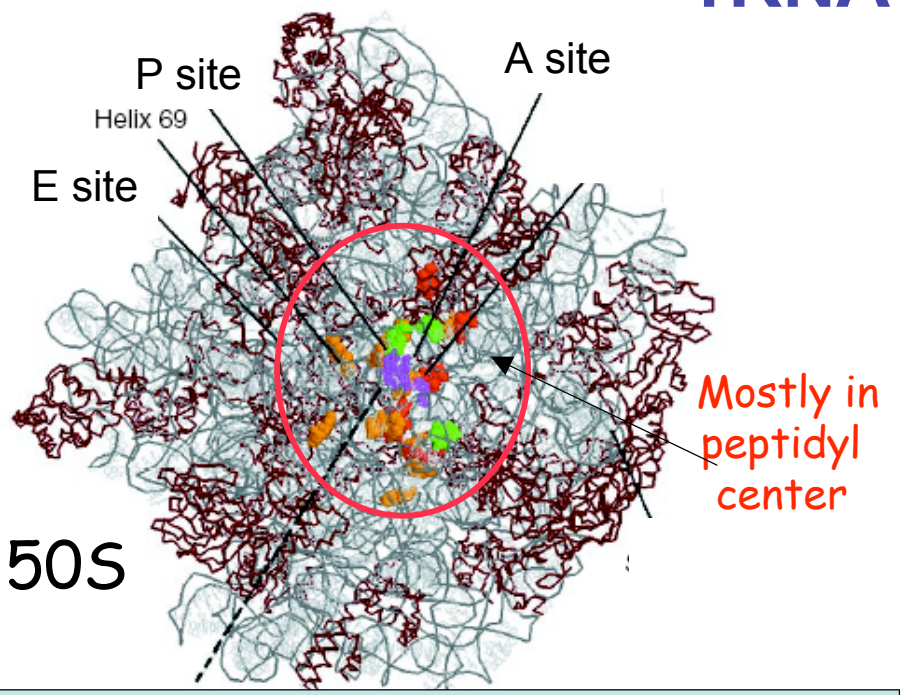
RNA type	tRNA 75-95nt	sn(o)RNA 70-1200nt	rRNA 1600-4000nt	mRNA 100-3000nt
Content	3-20 nt	0-13 nt	65-200 nt	0-5 nt
	4-26%	0-10%	1-3%	<1%
Diversity	85	10	30	13

Majority are exclusively present in tRNAs

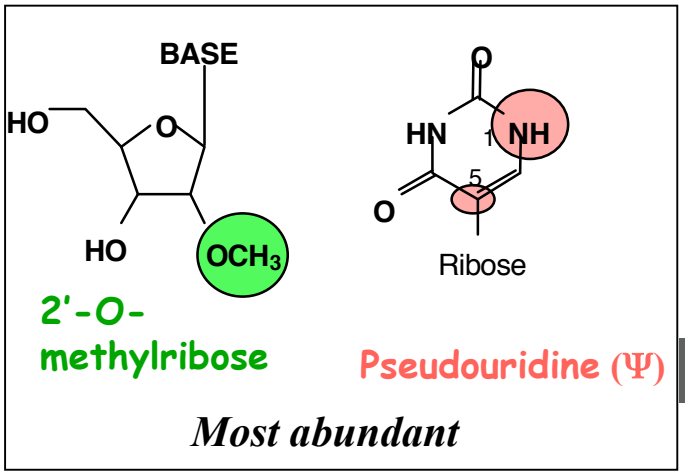


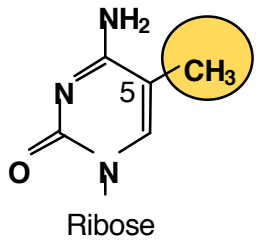
E. coli rRNA

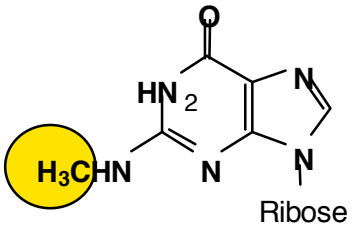


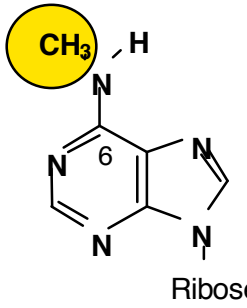


⇨ They are mostly located in functional regions of RNAs molecules



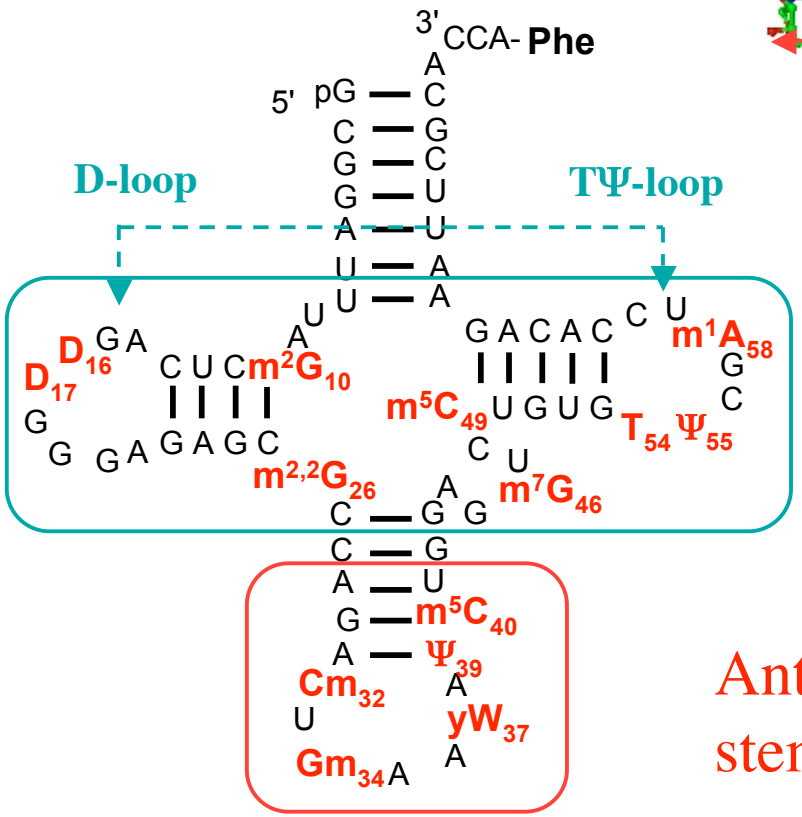






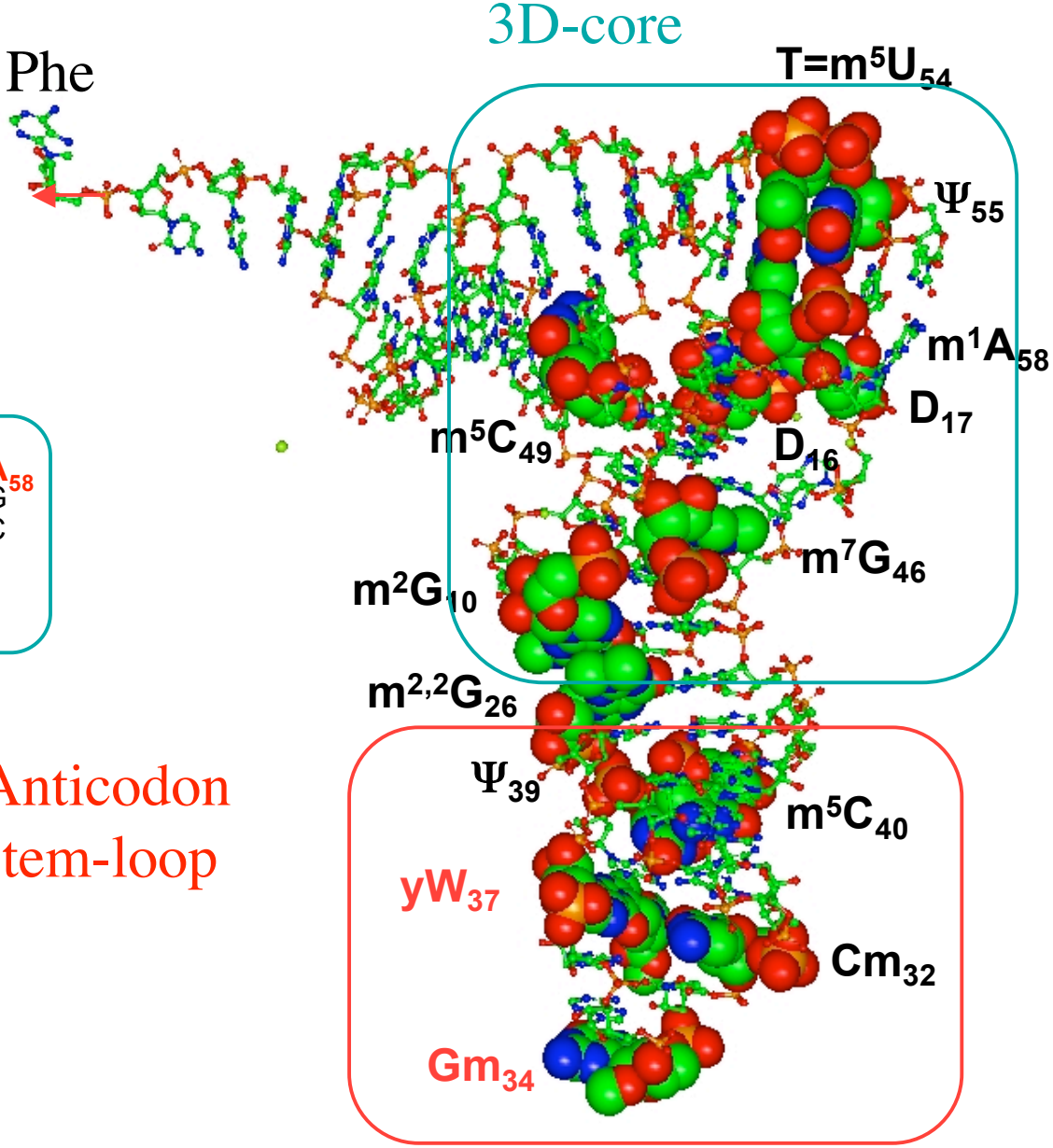
5-methyl-*C* (m^5C) 2-methyl-*G* (m^2G) 6-methyl-*A* (m^6A)

Fully mature, functional tRNA^{Phe} from Yeast



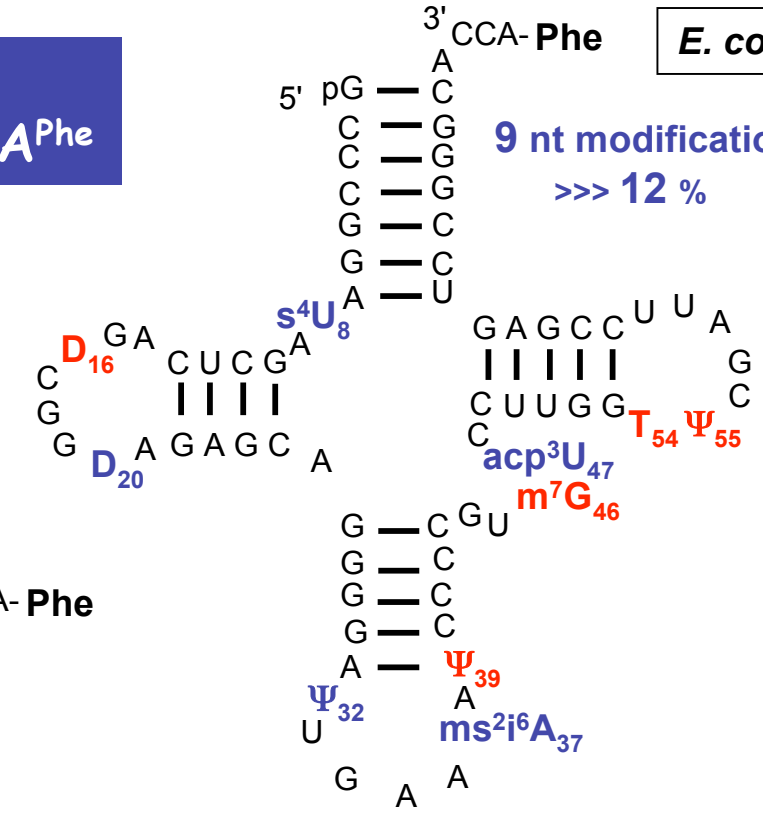
14 different Modifications = 19%

Anticodon stem-loop



Adapted from Phizicky et al, 2002

Fully mature, functional tRNA^{Phe}

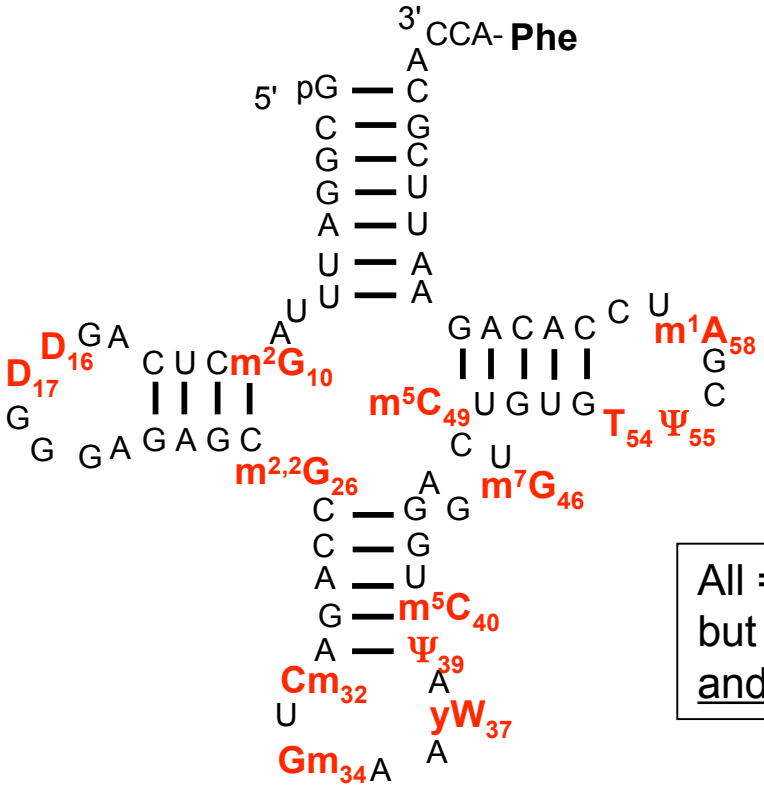


E. coli

9 nt modifications
>>> 12 %

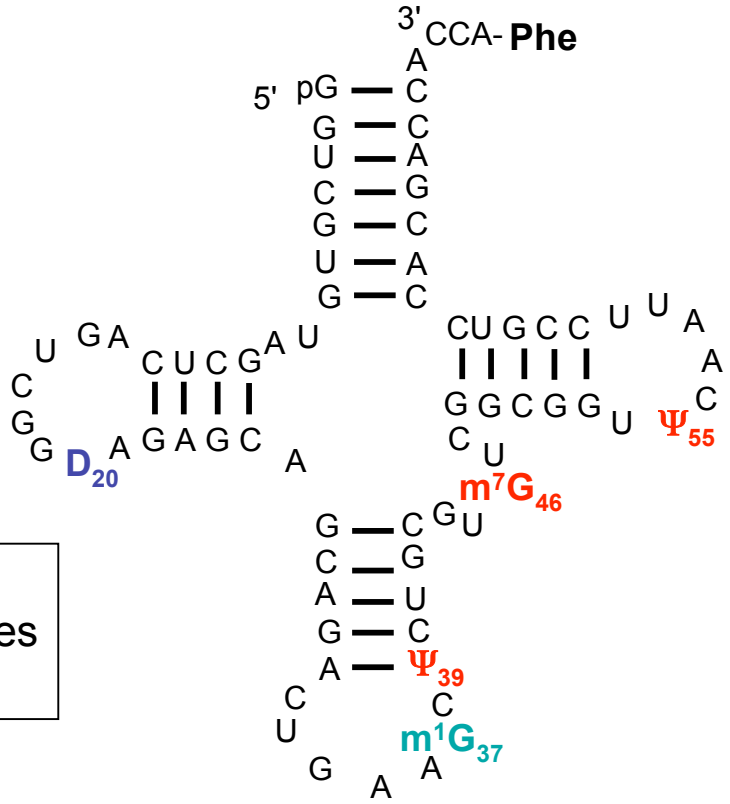
S. cerevisiae

14 nt modifications
>>> 19 %



Mycoplasma capricolum

5 nt modifications
>>> 7 %

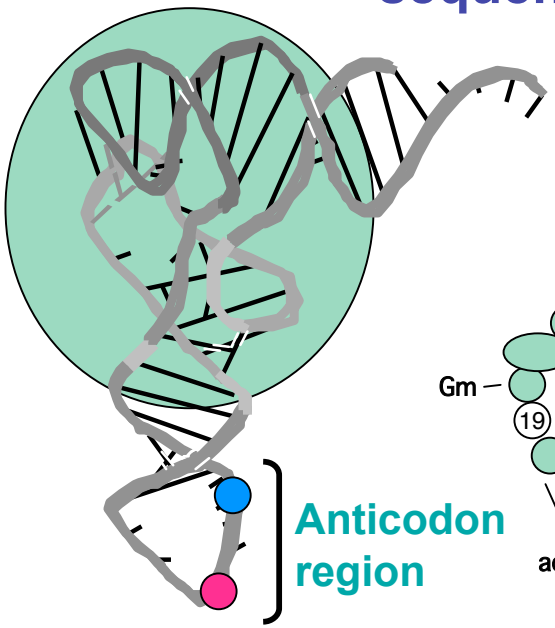


All = 76 nucl. in length
but of different sequences
and of pattern of mods

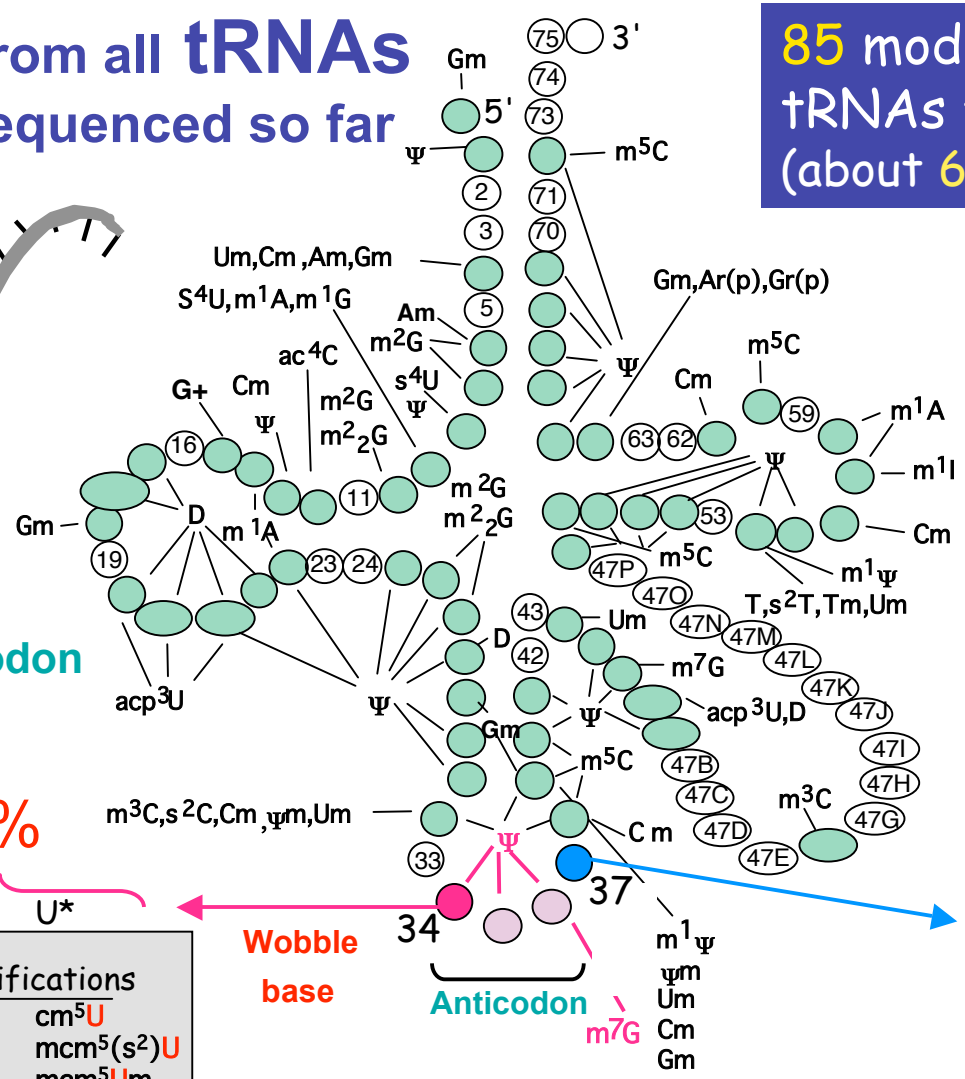
3D-core

From all tRNAs sequenced so far

85 modified nucleotides in tRNAs from various origins (about 600 sequenced so far)



Anticodon region



25 nt = 29%

mostly to stabilize 3D-structure of tRNA and/or to favor/forbid interactions with selected proteins/Enz

39 nt = 44%

A*/G*/C* U* U*

34 - wobble modifications

I	Um Ψ	cm5U
Gm	s2U	mcm5(s2)U
(o)Q	cmnm5(s2)U	mcm5Um
galQ	cmnm5Um	ho5U
manQ	mnm5(s2)U	mo5U
GluQ	mnm5Um	cmo5U
f5C(m)	nm5(s2)U	(m)cmo5U
m5C	mnm5seU	(m)chm5U
Cm		τm5(s2)U
ac4C(m)		
k2C, agmt2		

Majority = U-derivatives

62 nt = 71%

mostly involved in translation process

23 nt = 27%

G* A*

37- modifications

m1G	m1I
imG-14	m2A
(m)imG	m6A
imG2	(ms2)m6A
yW	(ms2)t6A
o2yW	(m)t6A
OHyW(*)	g6A
	(ms2)i6A
	io6A
	(ms2)hn6A

Grosjean
Mai 2011

Functions of modified nucleotides in RNAs

>> Mostly **'tuning roles'**, probably **'collective actions'**
> thus extremely difficult to demonstrate a precise function of a given modified nucleotide in RNA

(same as for mods in DNA)

Nevertheless

- **Some have structural/stabilization roles**

- **Others favor or forbid selected RNA/protein interactions = molecular screens**

- **Many of them (in t+rRNAs) play essential roles in translation process (efficiency/accuracy/regulation)**

Les enzymes de modification des ARN et leur évolution

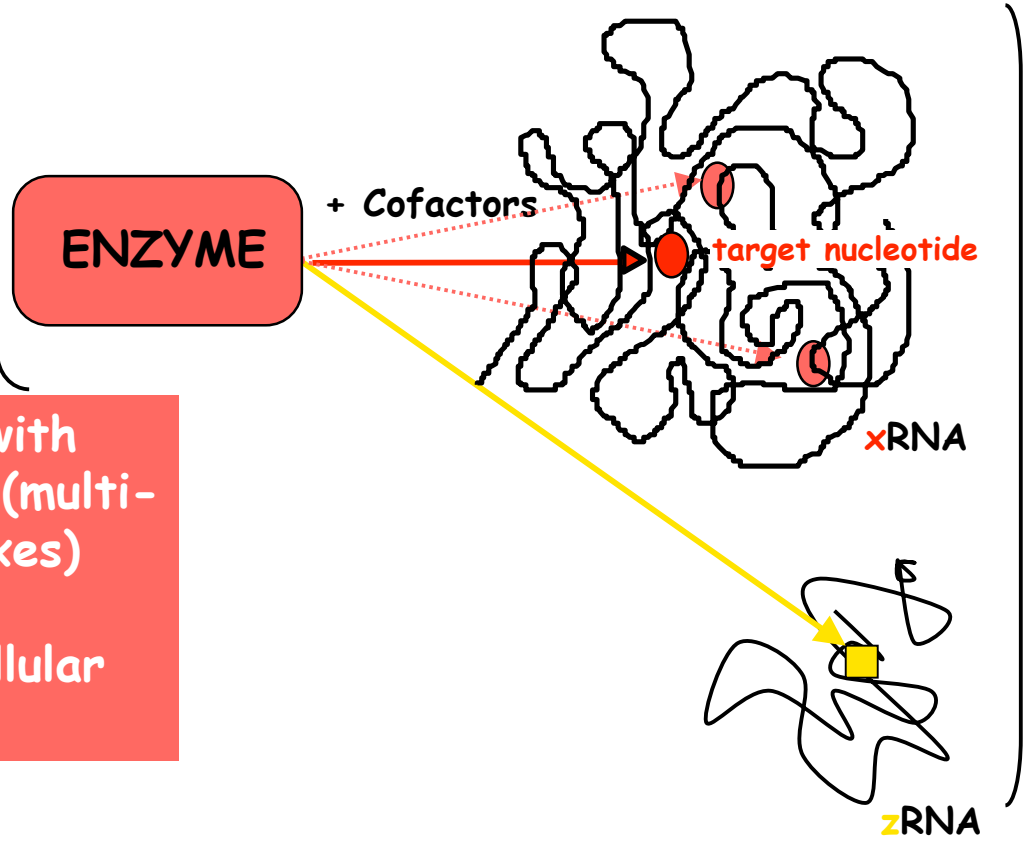
Most are single protein
α2, α4…

Other are heteromers
αβ, α2β2, …

or associated with other proteins (multi-protein complexes)

or bound to cellular substructures

'RNA substrates'
(generally bigger than the enzyme itself)

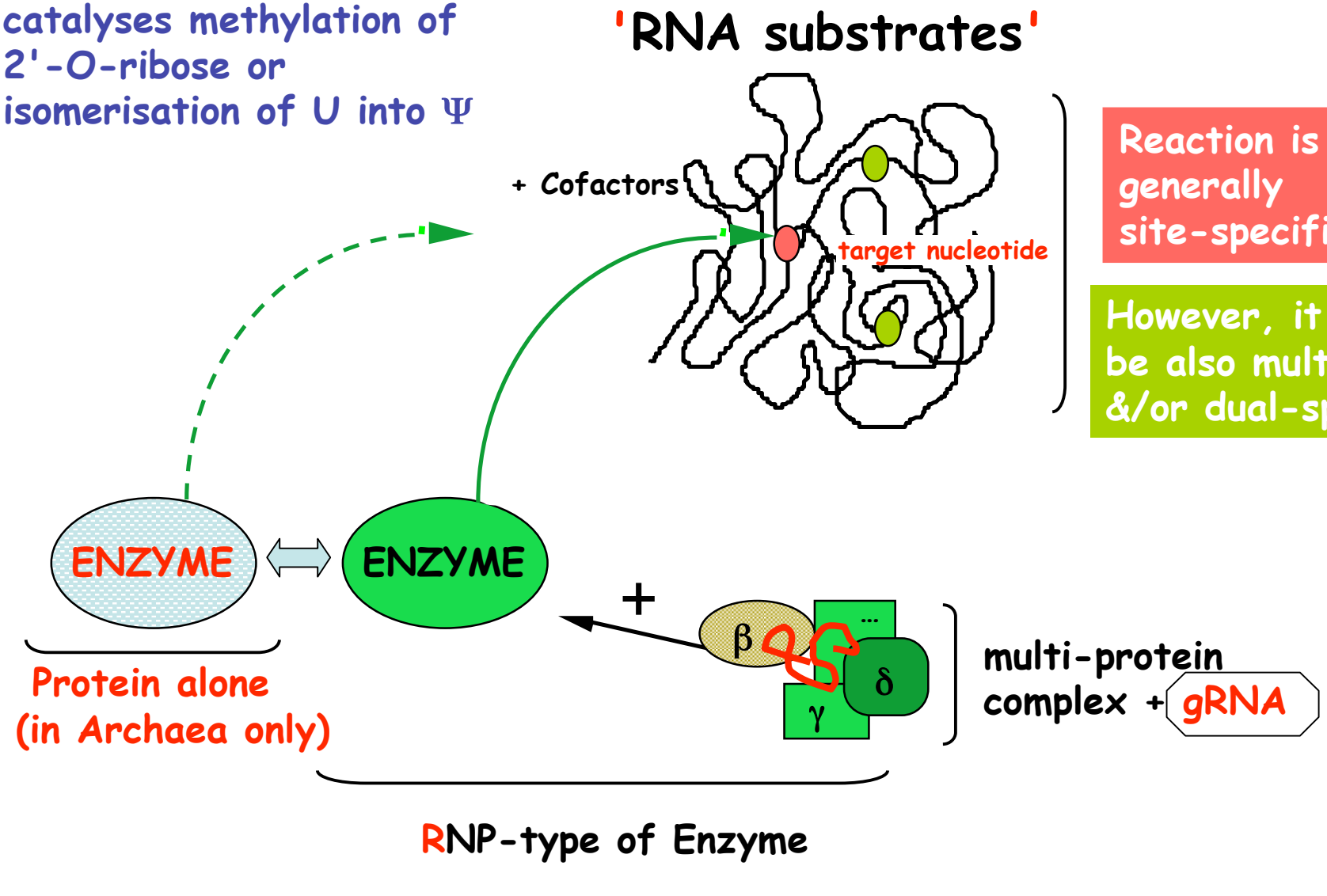


Most are:
site-specific

Other are:
multisite- or
dual-specific

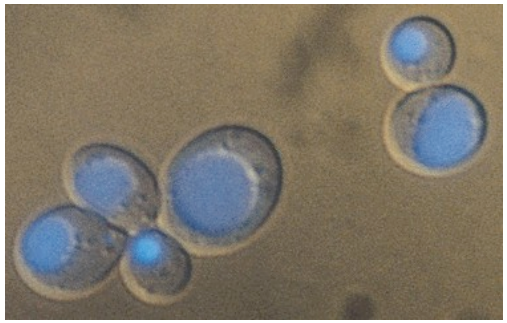
THESE TYPES OF ENZYMES CORRESPOND TO MAJORITY OF THE RNA/DNA MODS ENZYMES IDENTIFIED SO FAR

Such types of machineries catalyses methylation of 2'-O-ribose or isomerisation of U into Ψ



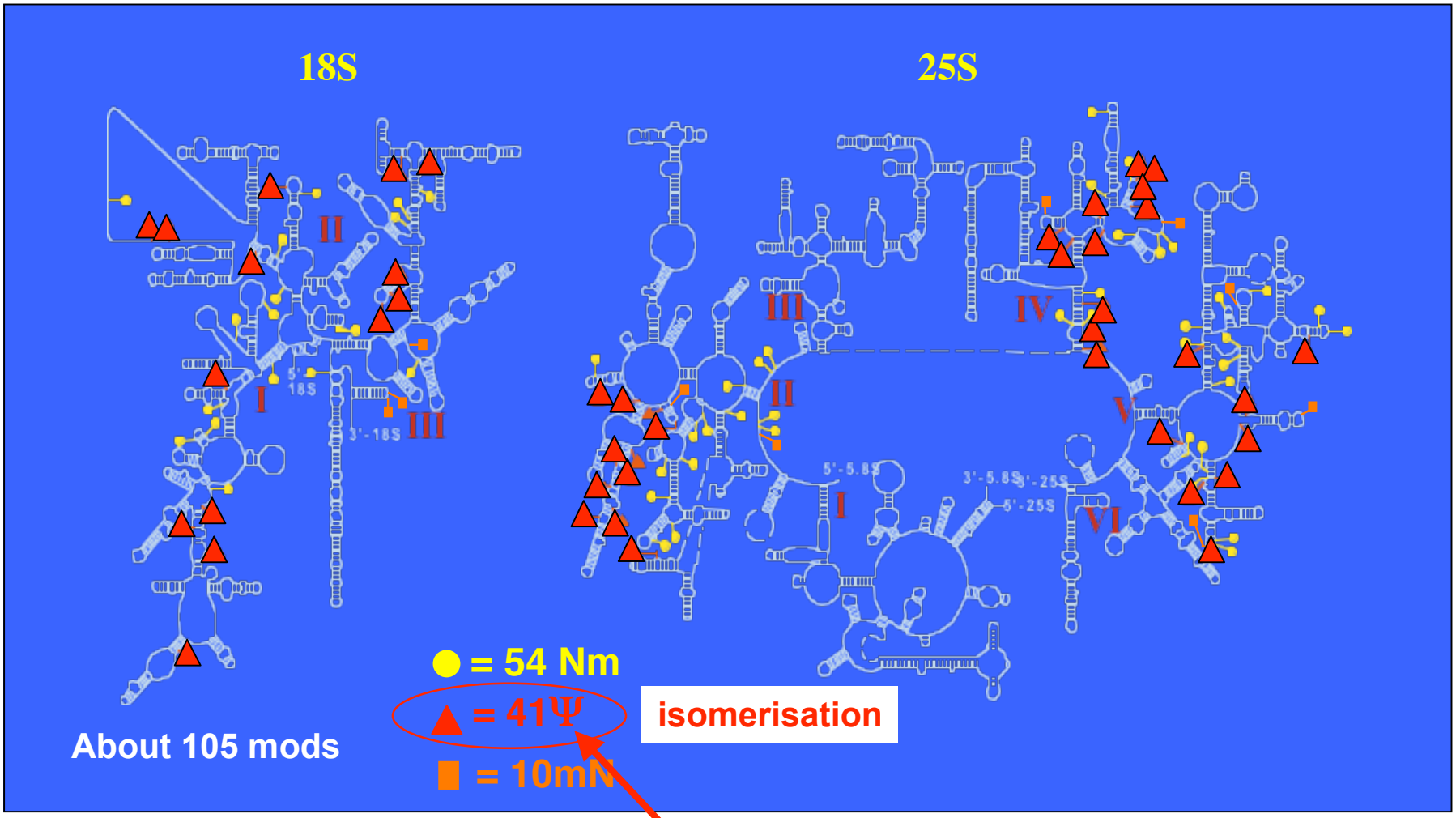
The specificity of the enzyme depends on the sequence of the RNA-guide
> thus here one common enzyme + various gRNAs

HOW MANY RNA MODIFICATION ENZYMES EXIST IN A GIVEN CELL TO ACCOUNT FOR ALL MODIFICATIONS IN THE DIFFERENT TYPES OF CELLULAR RNAs ?



Saccharomyces cerevisiae (a model system)

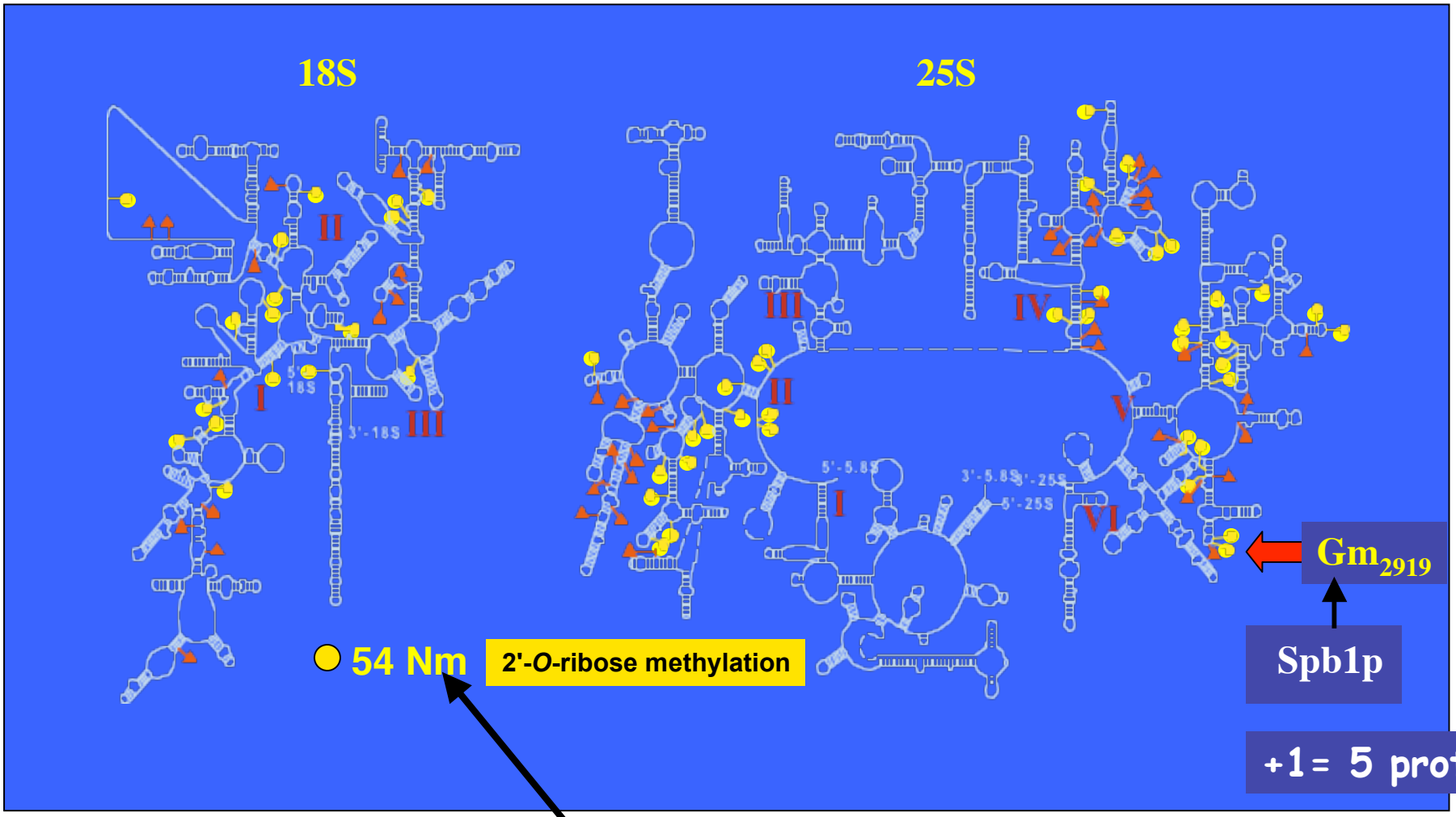
Modification map of yeast rRNAs



Catalysed by snoRNP complex
containing Cbf5p (enzyme) + Nhp2 + Nop10 + Gar1 = 4 proteins
and about 38 guide-RNA

Adapted from
Fournier et al, 2002

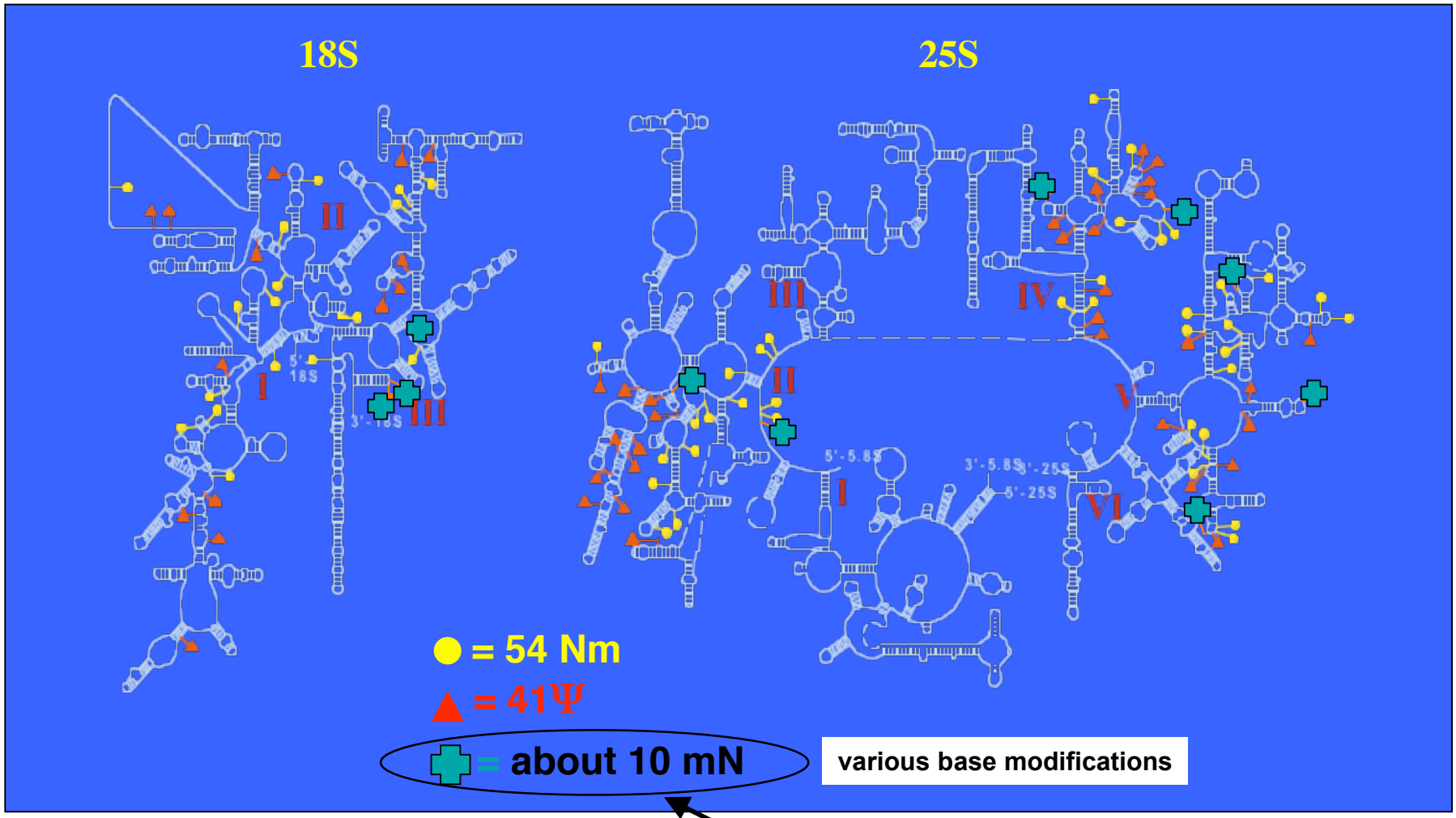
Modification map of yeast rRNAs



All, except one are catalysed by snoRNP complex
containing Nop1p (enzyme) + Nop56p+Nop58p+Snu13p = 4 prot
and about 50 guide-RNAs

Adapted from
Fournier et al, 2002

Modification map of yeast rRNAs



Catalysed by about 10 base modification enzymes

Adapted from
Fournier et al, 2002

tRNAs

S.cerevisiae

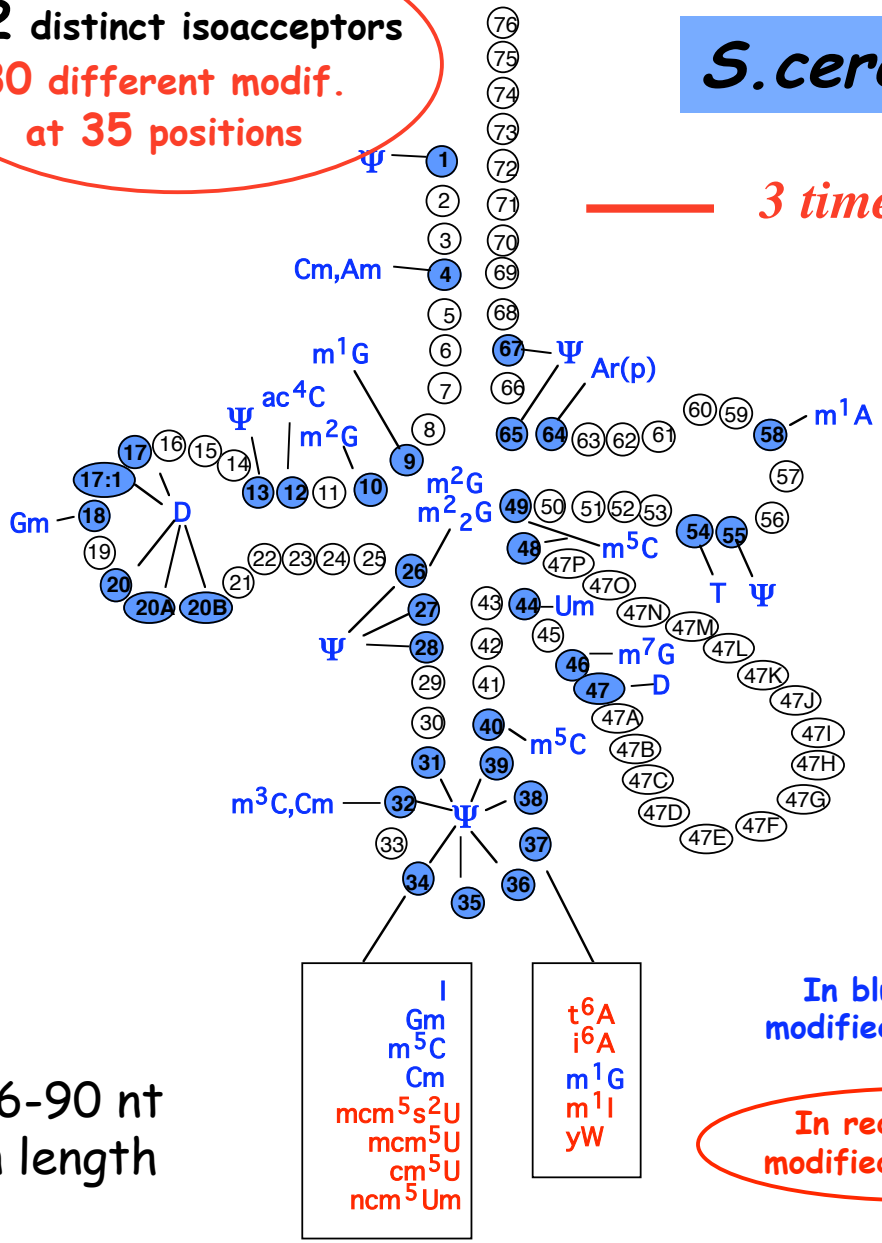
Cytoplasmic tRNAs

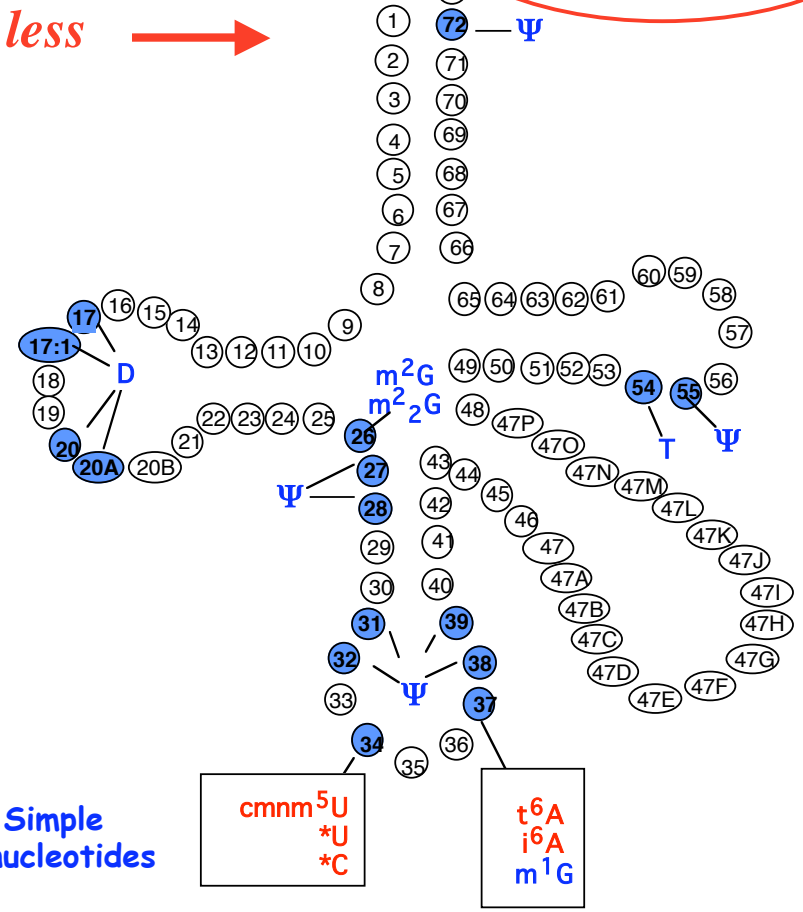
42 distinct isoacceptors
30 different modif.
at 35 positions

Mitochondrial tRNAs

24 distinct isoaccept.
10 different modif.
at 15 positions

3 times less



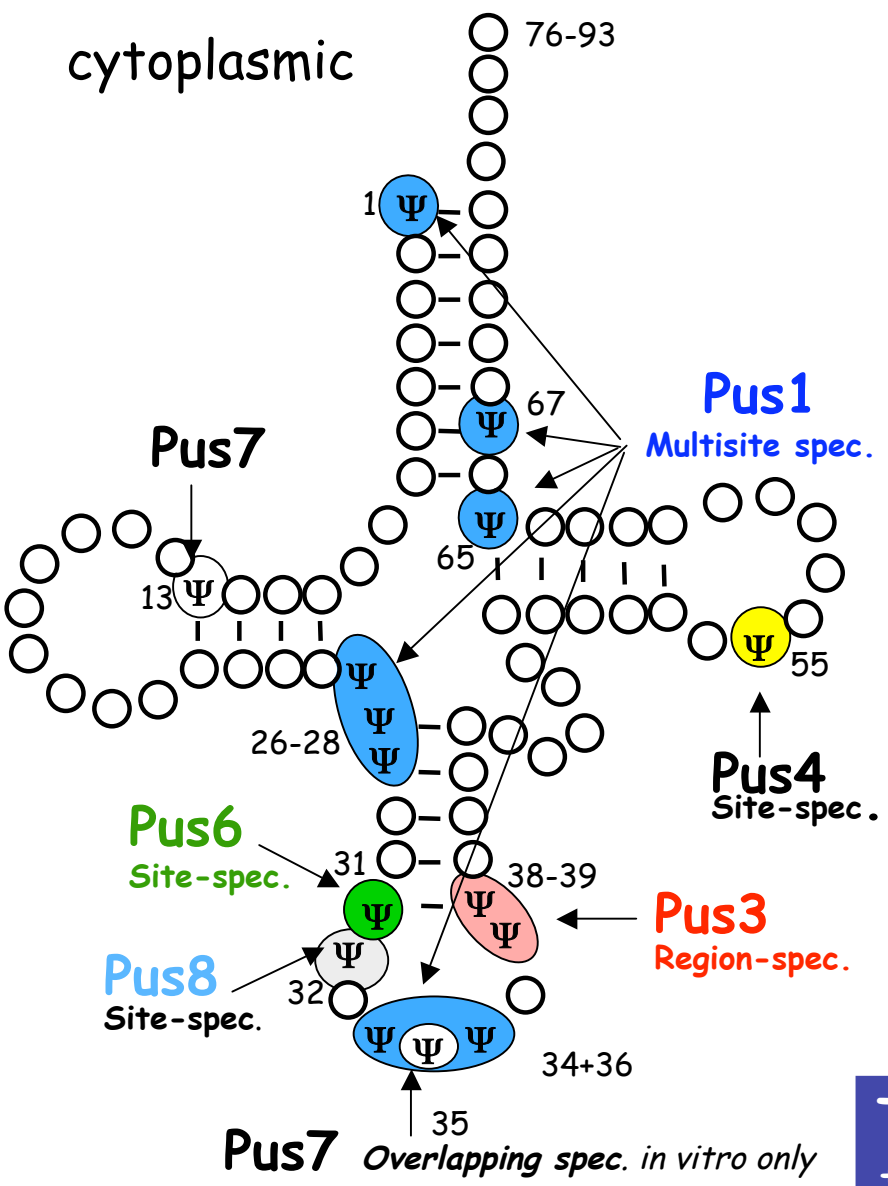


In blue:Simple
modified nucleotides

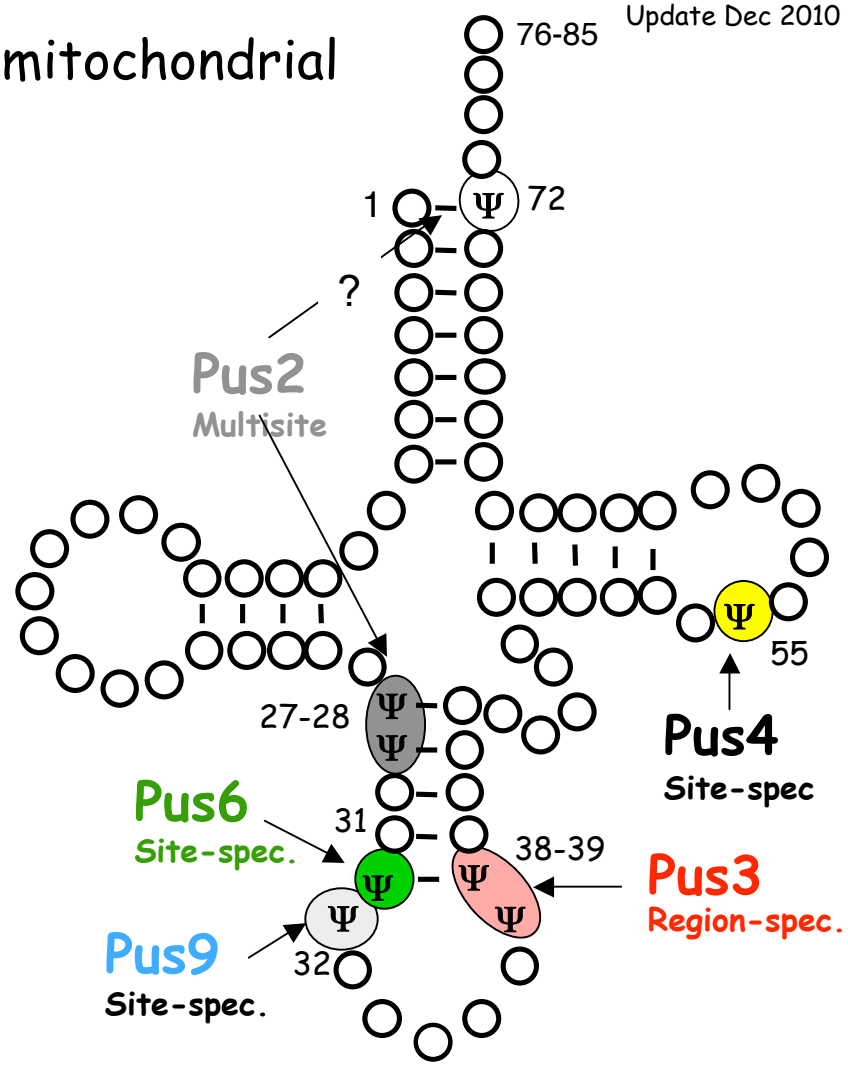
In red:Complex
modified nucleotides

76-90 nt
in length

Enzymatic formation of pseudouridine in yeast tRNAs



Pus5: pos. 2819 in mit 21S rRNA

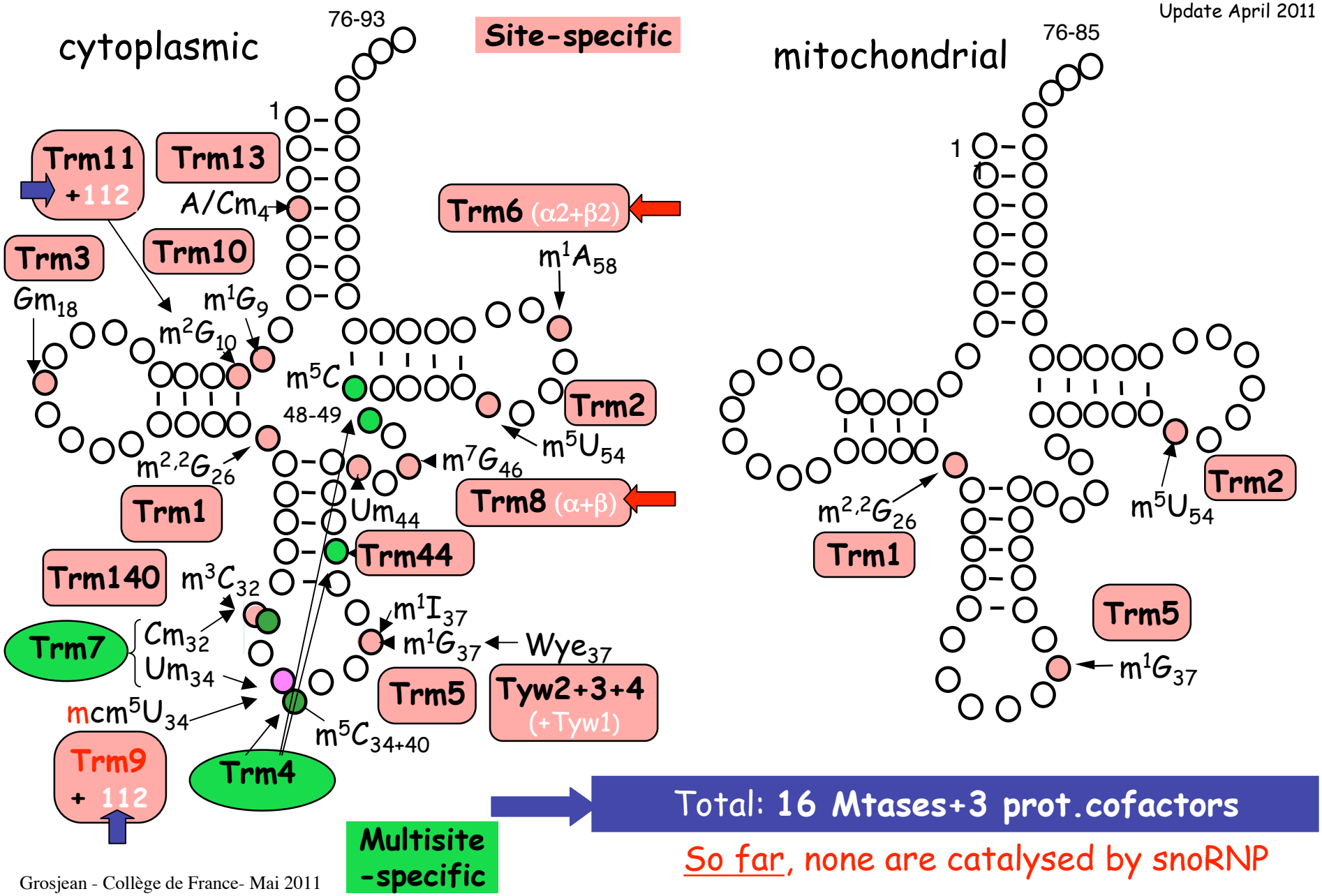


TOTAL: 8 tRNA:Ψ-synthases
+ one=Pus5->the only Ψ in mit.in rRNAs

So far, none are catalysed by snoRNP

Enzymatic formation of methylated nucleotides in ytRNAs

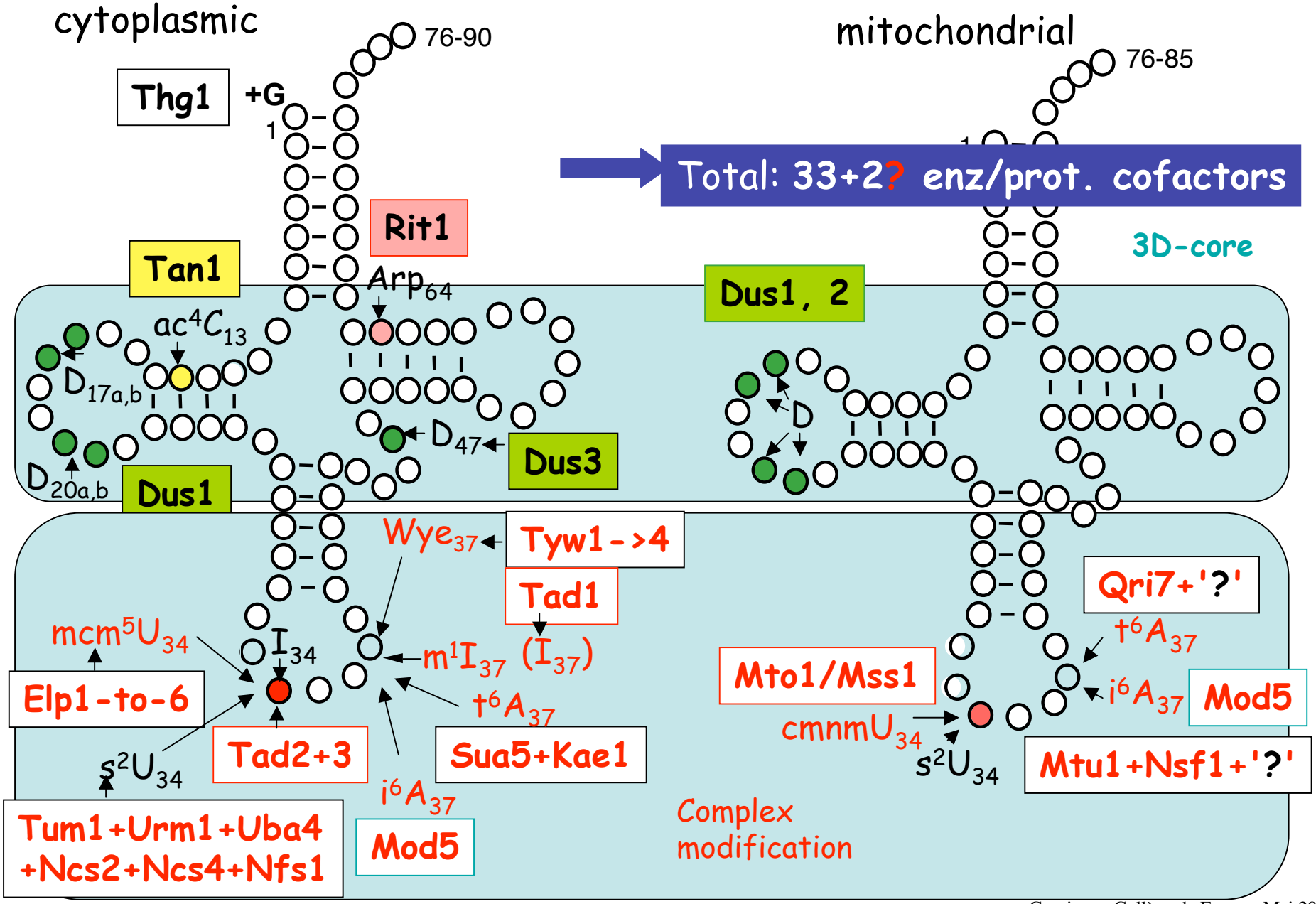
Update April 2011



Total: **16 Mtases+3 prot.cofactors**

So far, none are catalysed by snoRNP

Other types of base-modification enz. in *S.cerevisiae*



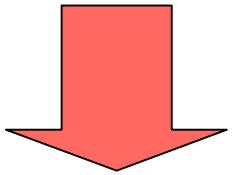
S. cerevisiae (mito+cyto)	Enzymes (including Prot. factors) / ORFs
tRNAs (31 distinct mods/36 pos)	min 62 - 65 MAX
rRNAs (9 distinct mods/105 pos)	18 - 25 + MANY gRNAs
mRNAs, snRNAs, snoRNAs (6 distinct mods/about 15 pos)	10 - 15
Other types of RNAs	(?)
TOTAL:	90 - 105

Not including all other 'maturation' enzymes (intron splicing, CCA addition, 5'+ 3' trimming...)

Assuming an average MW of 60 kDa (about 550 aa) per enzyme
(as well as for the auxiliary protein factors)
---> **90-100 PROTEINS** correspond to **MORE** than 1% of the
S.cerevisiae genome / i.e more than the genetic information required
for coding the structural genes for the pre-t+rDNAs (about 0.35%).

Same conclusion (about 1%) was reached by Glenn Björk in the case of E.coli

The evolutionary aspect of RNA modification machinery in the different types of organisms



Methylases
Isomerases
Deaminases
Thiolases
Acetylases
Formylases
Oxydases
Reductases
Glycosylases
Ribosylases
Transferases
+ many others...

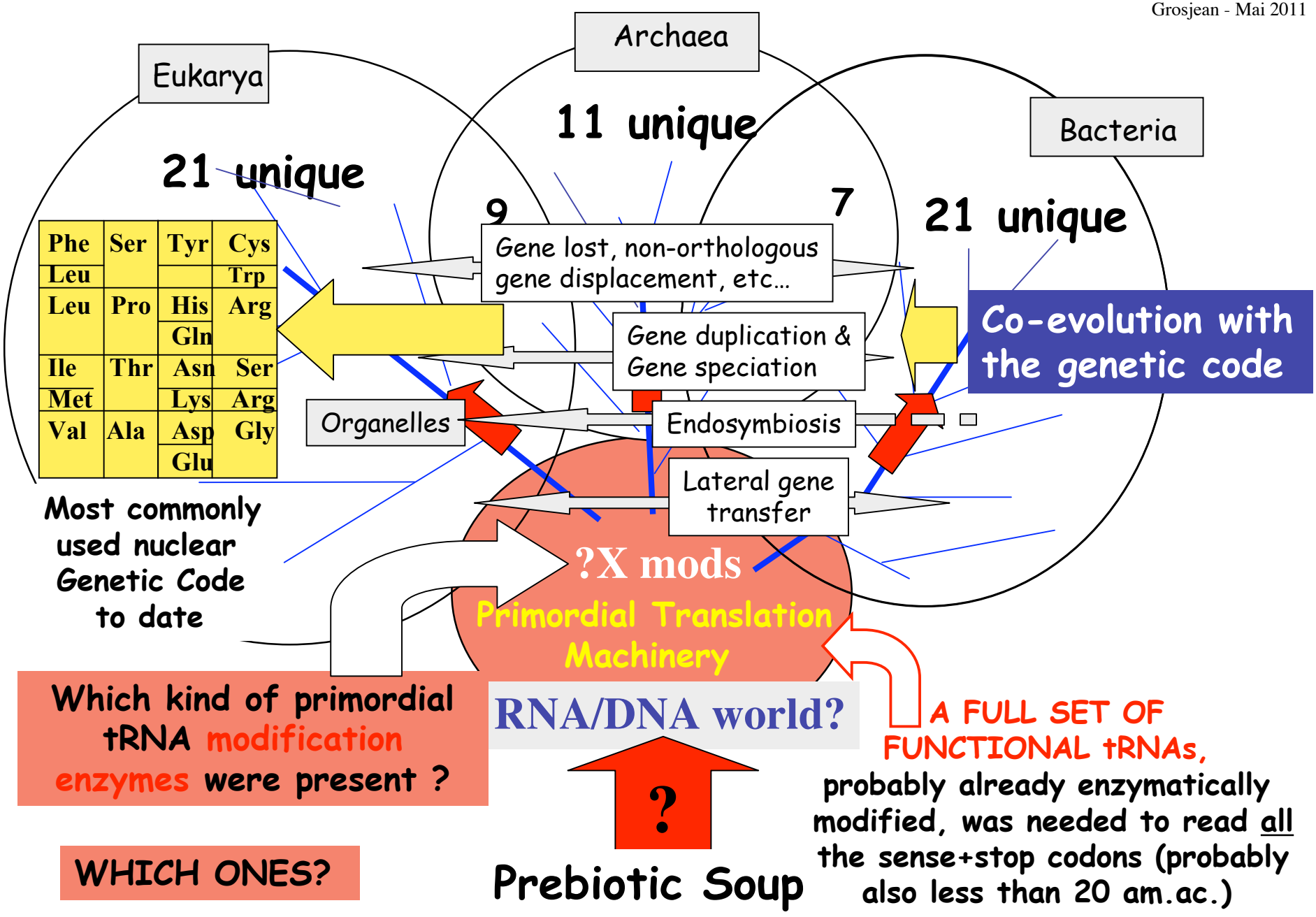
More than 100

1) They certainly have been acquired **progessively** during cellular evolution

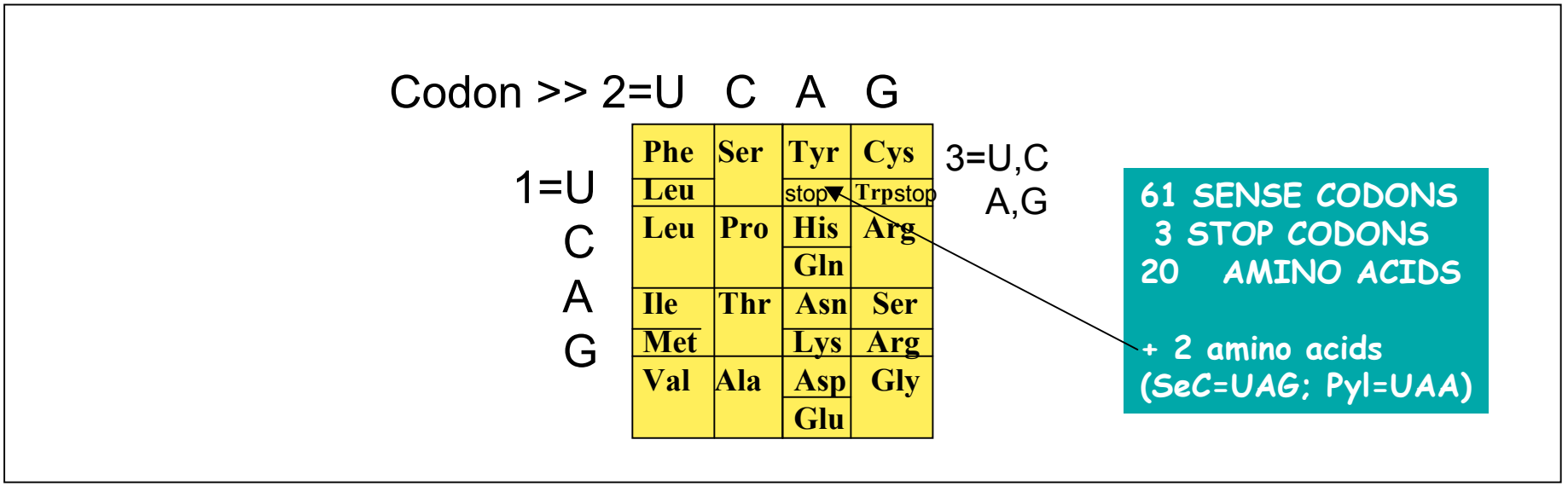
2) Inspection of their amino acid sequences and their 3D-architecture (when available) should allow to shed light on their emergence during cellular evolution

The emergence of RNA modification enzymes and the genetic code

Grosjean - Mai 2011



The 'almost universal' Genetic Code



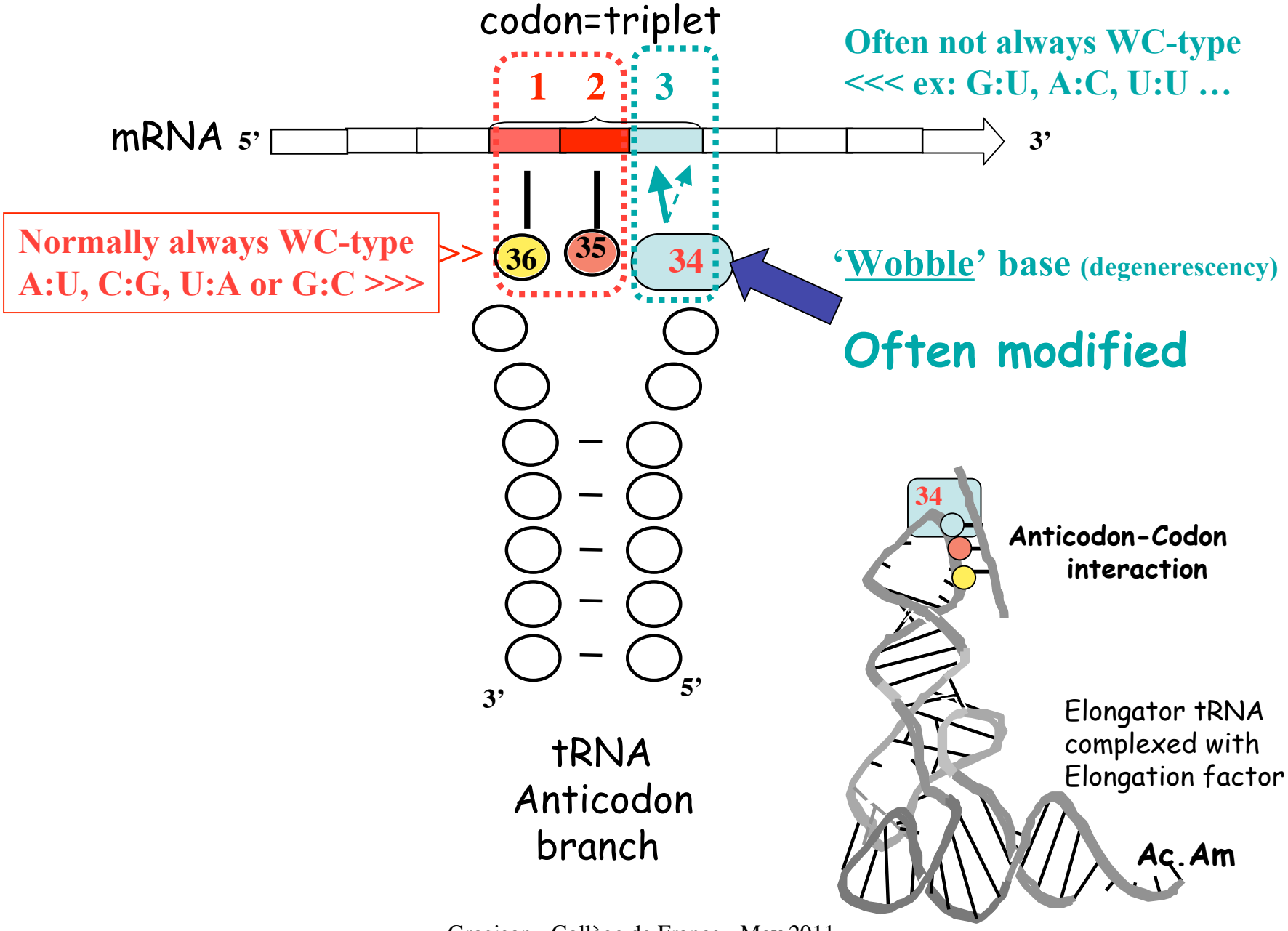
Questions ?

1- Does it results entirely from a very early 'frozen accident' (Francis Crick hypothesis)?

2- Or from a laborious progressive molecular tinkering during Evolution ?

3- How important was the progressive acquisition of the many RNA modification enzymes in either of these two, not necessarily exclusive alternative hypothesis ?

GENETIC TRANSLATION ON THE RIBOSOME (A-site)



Distribution of (hyper)modified nucleosides in anticodon of tRNAs

in position 34 (wobble base)

out of 37

10 blk

= 9U*

4-codon boxes>>

Eukarya

Organelles

Bacteria

Archaea

ho^{5}U
f^{5}C (m)
τm^5(s^2)U
cmnm^{5}U
k^{2}C

m^{5}C
galQ
manQ
cm^{5}U
mcm^5(s^2)U(m)
(m)hm^{5}U
ncm^{5}U(m)
Psi

oQ
GluQ
preQ1
(m)cmo^{5}U
(cm)nm^{5}U(m)
(cm)nm^5(s^2)U
mnm^5(s^2)U

I
Q

Gm
Cm
Um
s^{2}U

ac^{4}C
mnm^5(se^2)U
mo^{5}U

s^{2}Um
U*?,
ac^{4}Cm
agm^{2}C

Phe / Leu	Ser	Tyr / stop	Cys / Trp
Leu	Pro	His / Gln	Arg
Ile / Met	Thr	Asn / Lys	Ser / Arg
Val	Ala	Asp / Glu	Gly

< 2-codon boxes

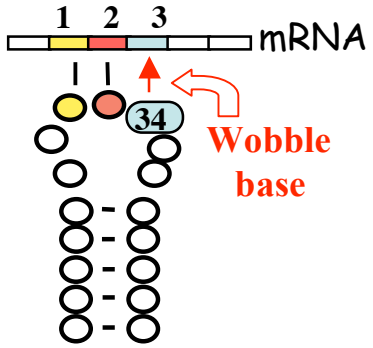
> (Hyper)modifications are late acquisition during evolution

out of 37

27 red

= 16U* = 43%

are found in tRNAs of the **'split' decoding boxes**



Decoding pattern of *Mycoplasma capricolum* (minimalist Bacterium)

*Two tRNA-Met (e+i) + special tRNA-Ile(C*AU)*

codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon
UUU Phe UUC	GAA	UCU UCC Ser UCA UCG	UGA	UAU Tyr UAC	GUA	UGU Cys UGC	GCA
UUA Leu UUG	cmnm5 UmAA CmAA			UAA UAG	Stop	UGA Trp UGG	cmnm5 UmCA CmCA
CUU CUC CUA Leu CUG	UAG	CCU CCC CCA Pro CCG	UGG	CAU His CAC	GUG	CGU CGC CGA Arg CGG	ICG <Rare codon Unassigned
				CAA Gln CAG	cmnm5 s2UUG		
AUU AUC Ile AUA Met	GAU k2CAU	ACU ACC Thr ACA ACG	AGU UGU	AAU Asn AAC	GUU	AGU Ser AGC	GCU
AUG e+i	CAUe			AAA Lys AAG	cmnm5 s2UUU CUU	AGA AGG	cmnm5 UCU
GUU GUC GUA Val GUG	UAC	GCU GCC GCA Ala GCG	UGC	GAU Asp GAC	GUC	GGU GGC GGA Gly GGG	UCC
				GAA Glu GAG	cmnm5 s2UUC		

CAUi

>>> TOTAL : 28 tRNA with distinct anticodons

Decoding pattern of *Escherichia coli*

*Two tRNA-Met (e+i) + special tRNA-Ile(C*AU)*

codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon
UUU Phe UUC	GAA	UCU UCC Ser UCA UCG	GGA cmo^{5}UGA CGA	UAU Tyr UAC	QUA	UGU Cys UGC	GCA
UUA Leu UUG	cmnm^5 UmAA CmAA			UAA UAG Stop		UGA UGG Trp	Stop CmCA
CUU CUC CUA Leu CUG	GAG cmnm^5 UAG CAG	CCU CCC CCA Pro CCG	GGG cmo^{5}UGG CGG	CAU His CAC	QUG	CGU CGC CGA Arg CGG	ICG Rare codon CCG
				CAA CAG Gln	mnm^{5}s^{2}UUG CUG		
AUU AUC Ile	GAU k^{2}CAU	ACU ACC Thr	GGU cmo^{5}UGU	AAU Asn AAC	QUU	AGU Ser AGC	GCU
AUA Ile		ACA		AAA Lys	mnm^{5}s^{2}UUU	AGA	mnm^5 UCU
AUG Met (CAUi)	ac^{4}CAUe	ACG	CGU	AAG		AGG	CCU
GUU GUC GUA Val GUG	GAC cmo^{5}UAC	GCU GCC GCA Ala GCG	GGC cmo^{5}UGC	GAU Asp GAC	gluQUC	GGU GGC GGA Gly GGG	GCC mnm^5 UCC CCC
				GAA Glu GAG	mnm^{5}s^{2}UUC		

>>> TOTAL: 41 tRNA with distinct anticodons

Decoding pattern of *S. cerevisiae (cytoplasmic)*

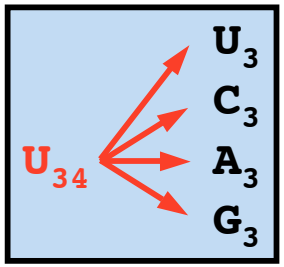
Two tRNA-Met (e+i); NO special tRNA-Ile(UAU)

codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon	codon aminoacid	anticodon
UUU Phe UUC	GmAA	UCU UCC Ser UCA UCG	IGA ncm⁵UGA CGA	UAU Tyr UAC	GΨA	UGU Cys UGC	GCA
UUA Leu UUG	ncm⁵UmAA m⁵CAA			UAA UAG Stop		UGA UGG Trp	Stop CmCA
CUU CUC CUA Leu CUG	GAG UAG	CCU CCC CCA Pro CCG	IGG ncm⁵UGG	CAU His CAC	GUG	CGU CGC CGA Arg CGG	ICG CCG
				CAA Gln CAG	mcm⁵s²UUG CUG		
AUU AUC Ile AUA Ile	IAU ΨAΨ	ACU ACC Thr ACA ACG	IGU ncm⁵UGU CGU	AAU Asn AAC	GUU	AGU Ser AGC	GCU
AUG Met	CAUe			AAA Lys AAG	mcm⁵s²UUU	AGA AGG	mcm⁵UCU CCU
GUU GUC GUA Val GUG	GAC ncm⁵UAC CAC	GCU GCC GCA Ala GCG	IGC ncm⁵UGC	GAU Asp GAC	GUC	GGU GGC GGA Gly GGG	GCC mcm⁵UCC CCC
				GAA Glu GAG	mcm⁵s²UUC		

CAUi

>>> TOTAL : 42 tRNA with distinct anticodons

in 4-(unsplit) codon set

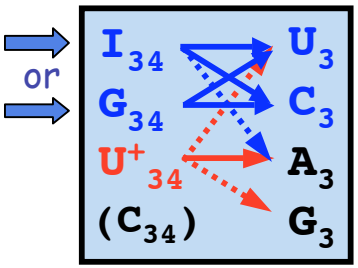


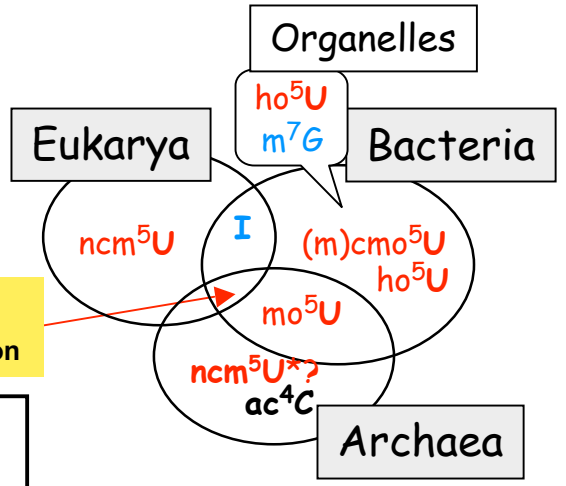
U+

EVOLUTION

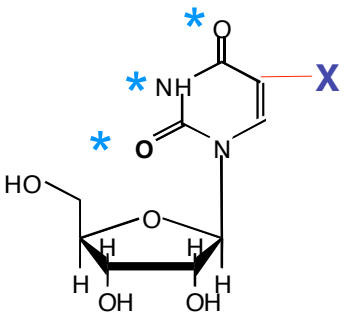
in 4-(unsplit) codon set

or





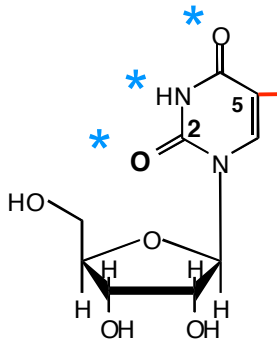
NOTHING in common



Xo^5U_{34}

<> X = OH (ho^5U)
X = OCH_3 (mo^5U)
X = OCH_2COOH (cmo^5U)
X = OCH_2COOCH_3 ($mcmo^5U$)

in Bacteria



Xcm^5U_{34}

<> X = NH_2 (ncm^5U)

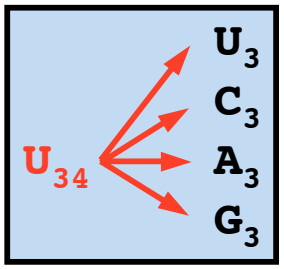
in Eukarya/Archaea

* Positions involved in WC base pairing

U+

need several enzymes distinct in Bacteria as in Eukarya/Archaea

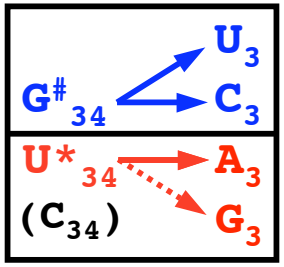
in 4-(unsplit) codon set

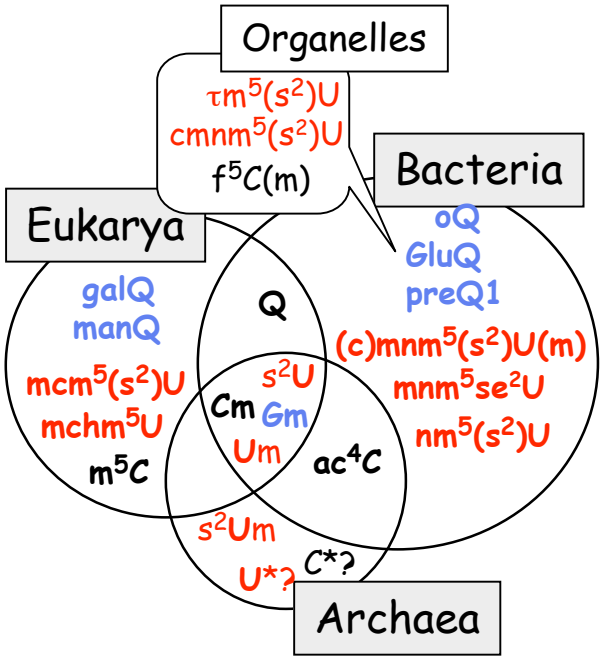


U*

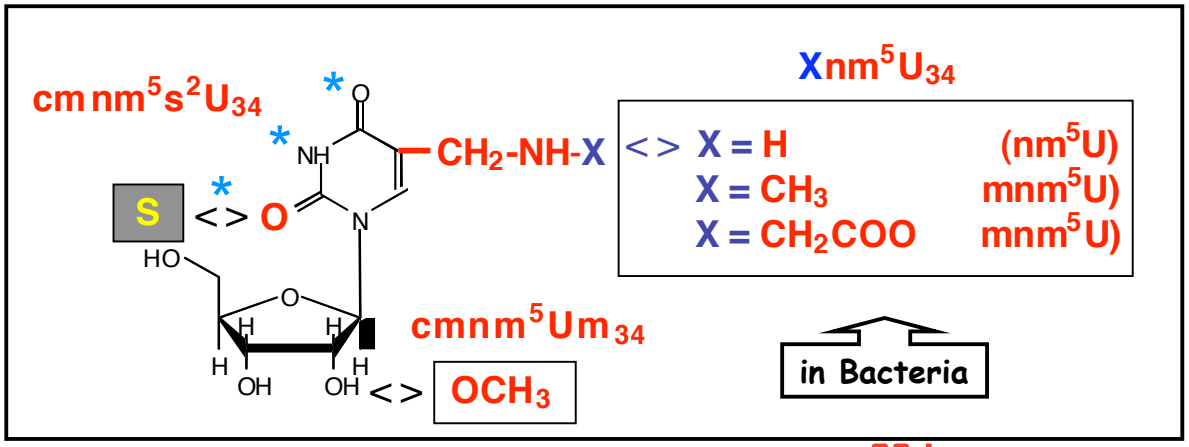
EVOLUTION

in 2-(split) codon set

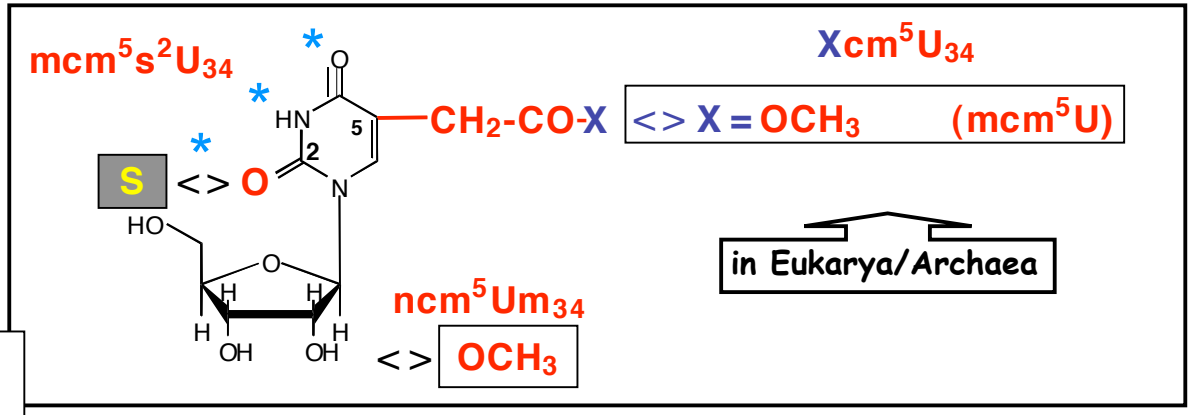




need several enzymes
= distinct as for B/E/A

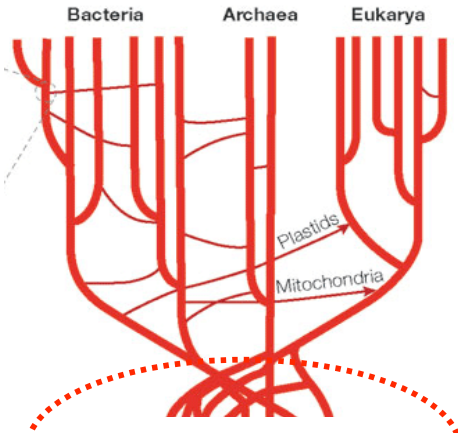


U*34

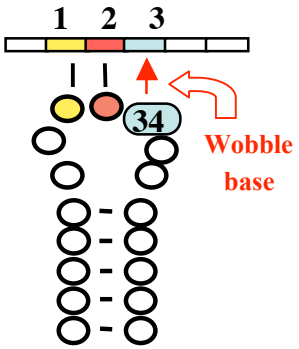


* Positions involved in WC base pairing

Evolutionary implications



Obviously, progressive introduction of modified U*/G*/C* at the wobble position-34 AND addition of new tRNAs allow the introduction of more amino acids within the 'split' decoding boxes



From 'Luca's / Commonote'

>>>>> Bacteria # Archaea # Eukarya > Evolution still goes on

4-codon boxes

MOD ENZ.

4 + 2-codon boxes

4 + 2 + 1 codon boxes

U	C	A	G
U: ?	Ser	?	?
C: Leu	Pro	Glu	Arg
A: Ile	Thr	Asp	Ser
G: Val	Ala	Asp	Gly

?

Primordial frozen accident?

Phe Leu	Ser	Tyr	Cys
Leu	Pro	His Gln	Arg
Ile	Thr	Asn Lys	Ser Arg
Val	Ala	Asp Glu	Gly

Selenocysteine

Pyrolysine

Trp

Met

'Latest' introduction

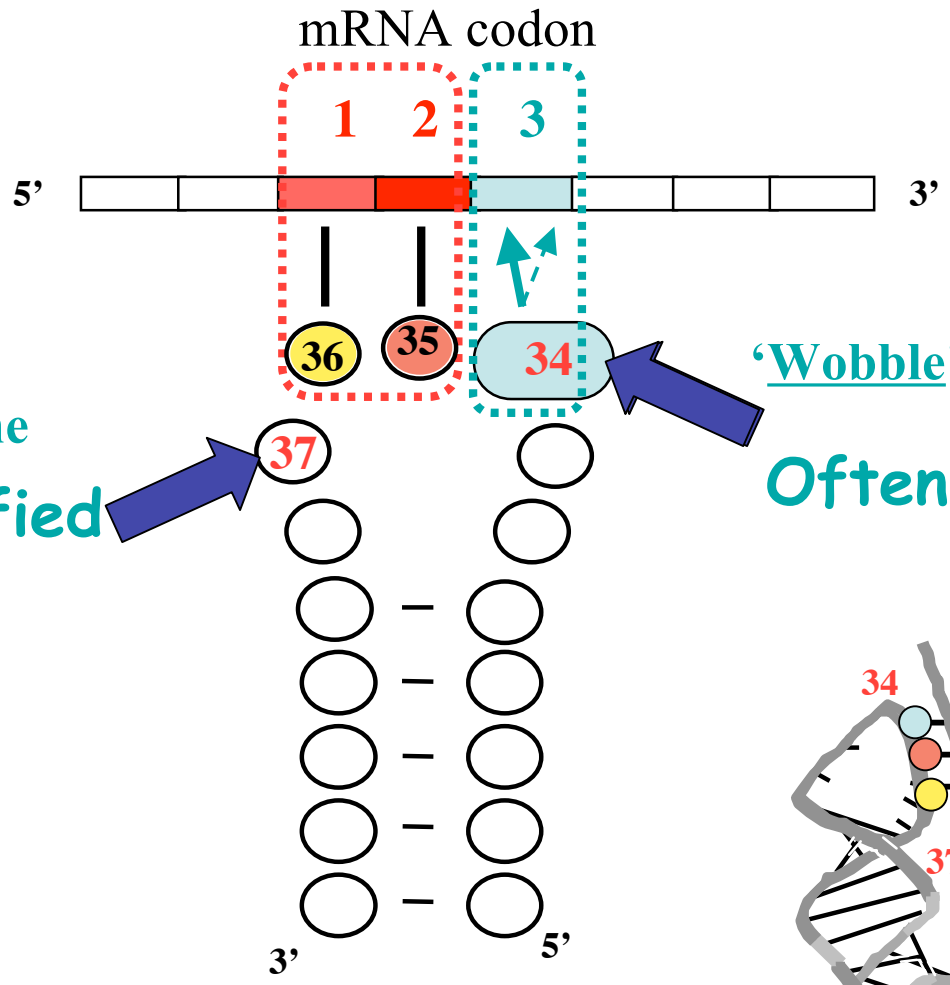
Few am.ac. (which ones ?)

18 am.ac.

22 am.ac.

20 am.ac.

GENETIC TRANSLATION ON THE RIBOSOME (A-site)



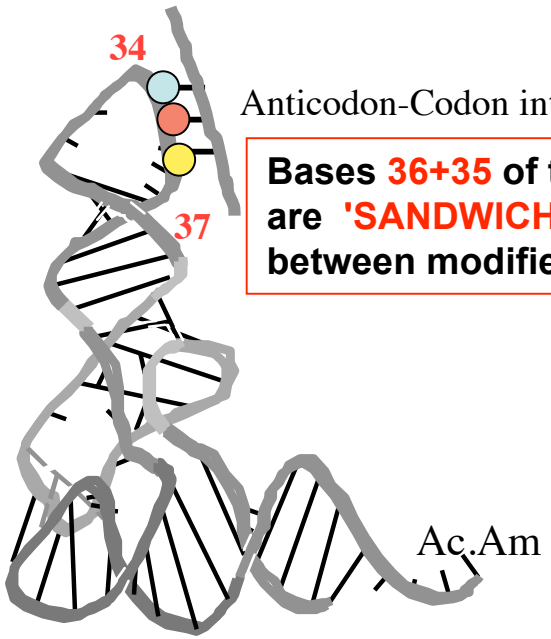
'Dangling' purine

Often modified

'Wobble' base (degenerescency)

Often modified

tRNA Anticodon branch

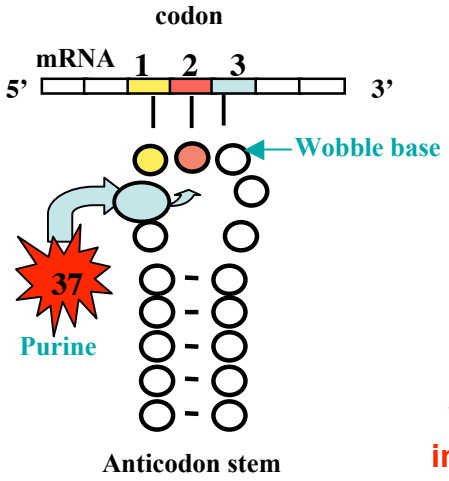


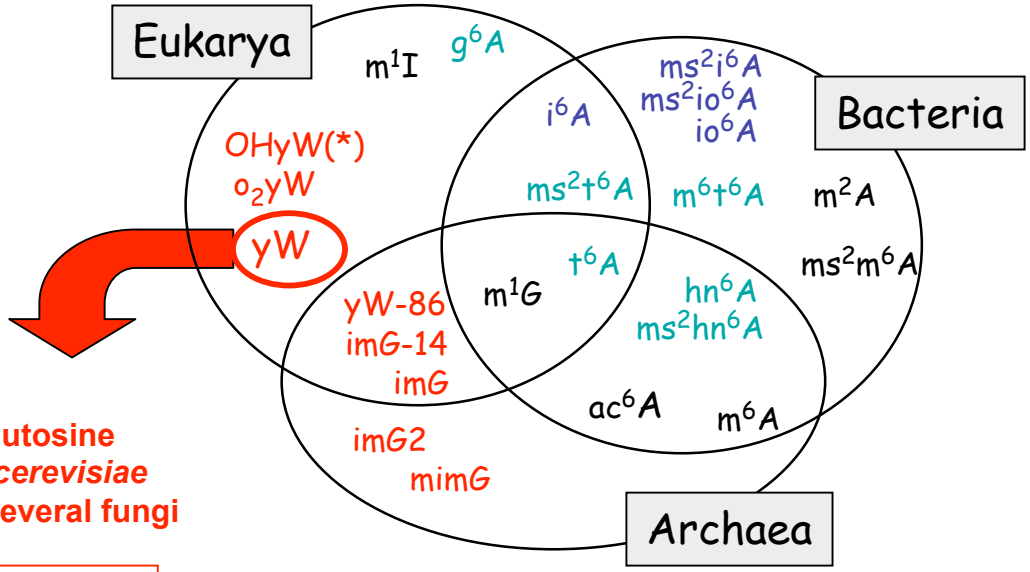
Anticodon-Codon interaction

Bases 36+35 of tRNA are 'SANDWICHED' between modified nts

Modification of Purine-37

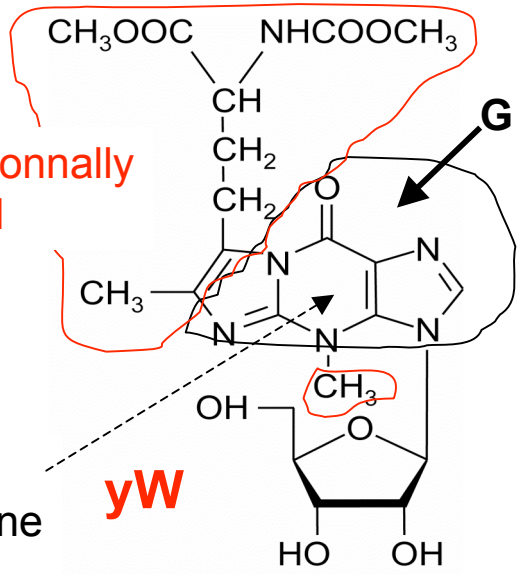
→ Also Species specific





Wybutosine in *S. cerevisiae* and several fungi

Post-transcriptionnally modified

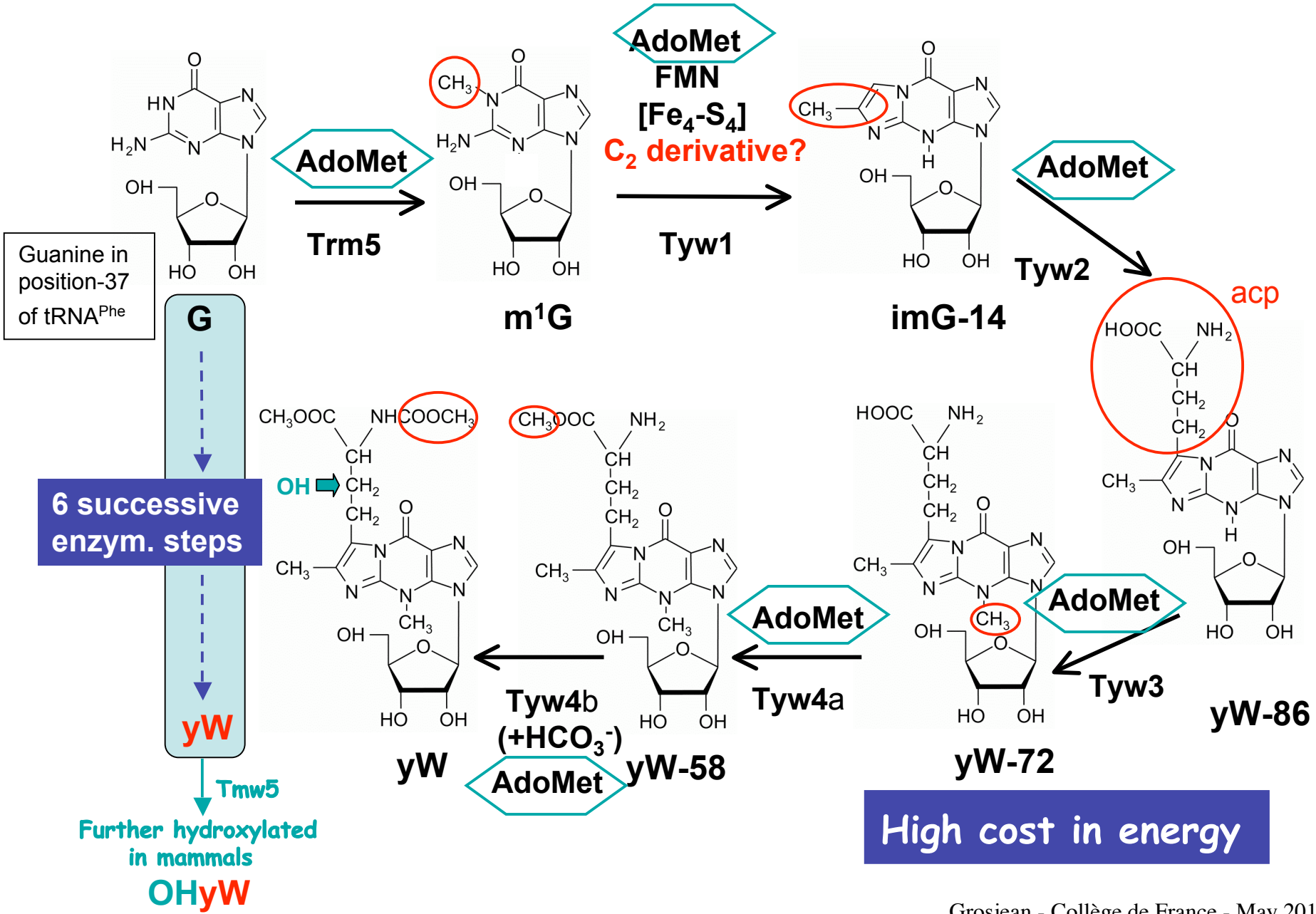


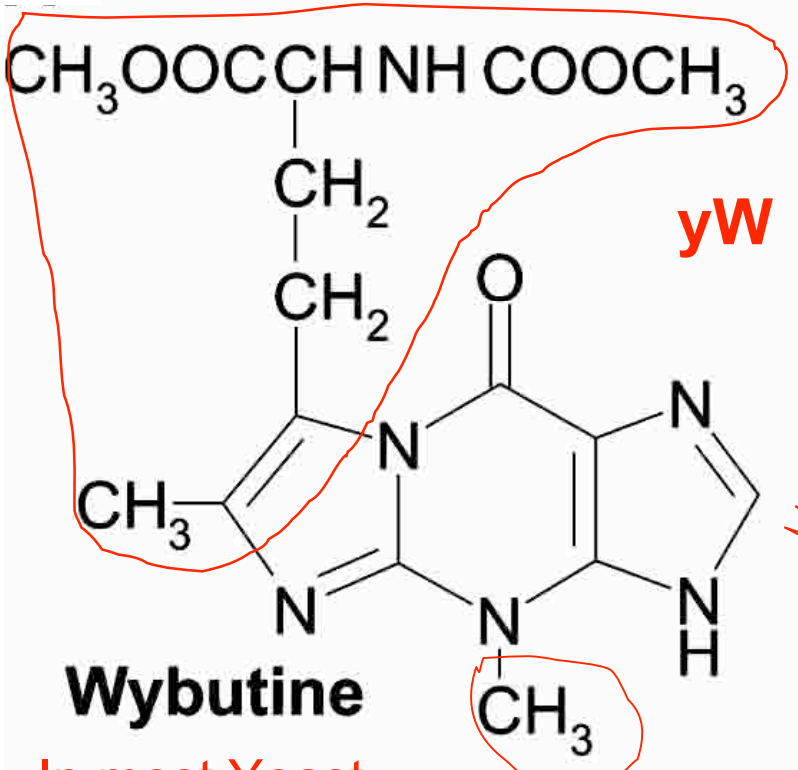
Tricyclic Imidazo-purine ring

G = encoded (gene level)

Phe	Ser	Tyr	Cys
Leu		stop	Trp
Leu	Pro	His	Arg
		Gln	
Ile	Thr	Asn	Ser
Met		Lys	Arg
Val	Ala	Asp	Gly
		Glu	

Biosynthesis of Wyosine derivatives in Eukarya

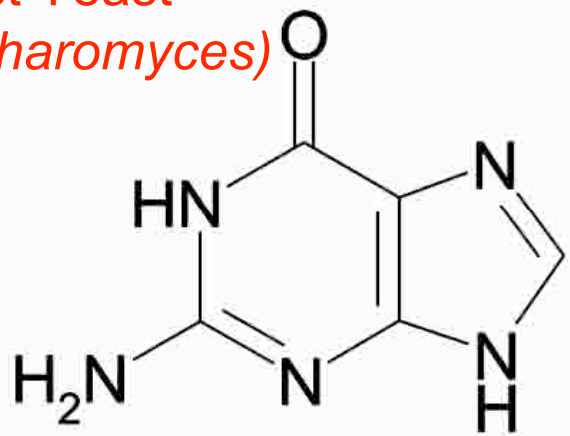




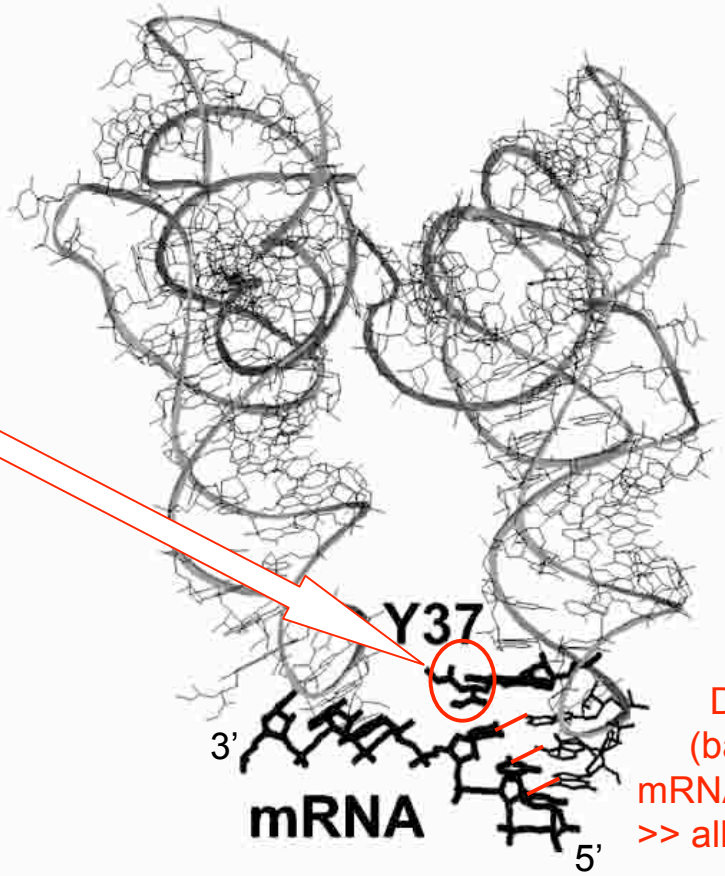
yW

Wybutine

In most Yeast
(S.saccharomyces)



Guanine



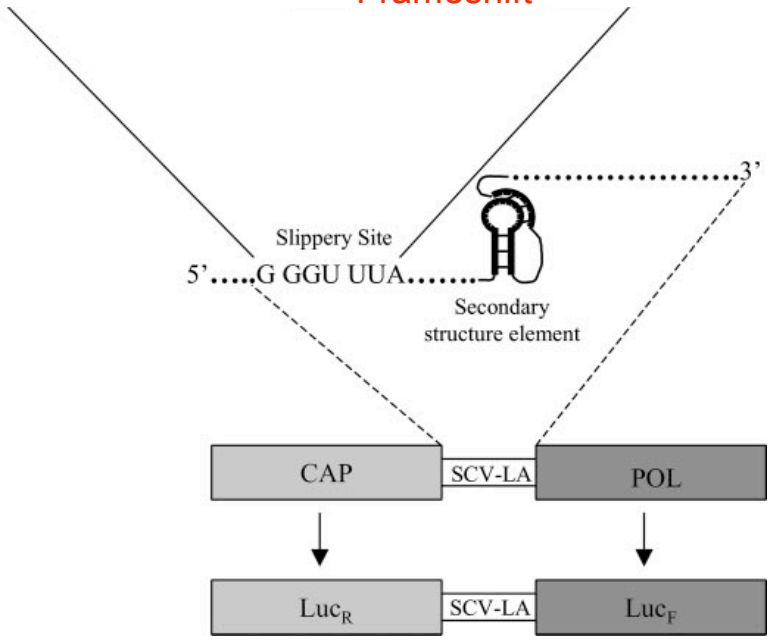
Dangling end
(base stacking)
mRNA-U_3/tRNA-A_{34}
>> allows stabilization

Stabilizes codon/anticodon interaction at the A-site and allows to avoid frameshifting during mRNA translation

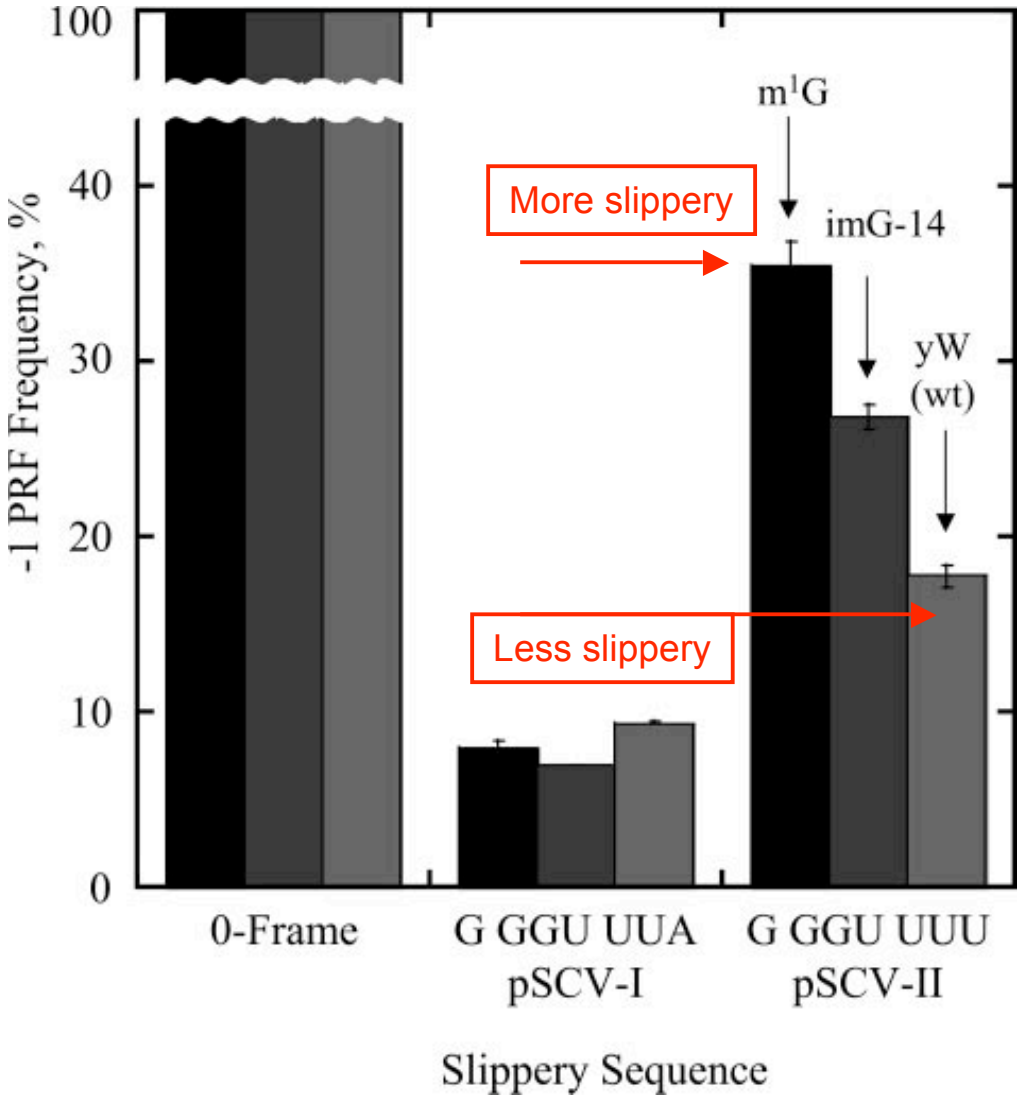
Frameshifting avoidance during translation : importance of modified base in tRNA

Construct	Slippery site variation	
	Pre-slippage	Post-slippage
pSCV-I	G GGU UUA	GGG UUU A
pSCV-II	G GGU UUU	GGG UUU U

Slippery
= Frameshift

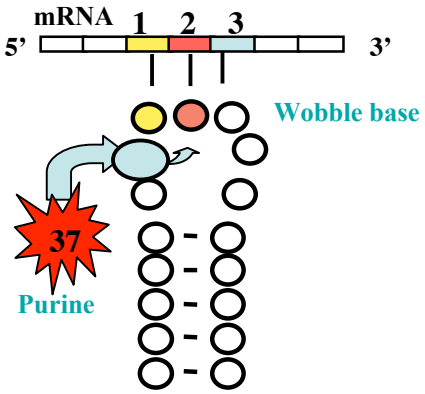


Frameshift frequency evaluated from the ratio of Luc-R/Luc-F fluorescence



Adapted from William F. Waas et al (Paul Schimmel's lab) JBC 282, 2007l

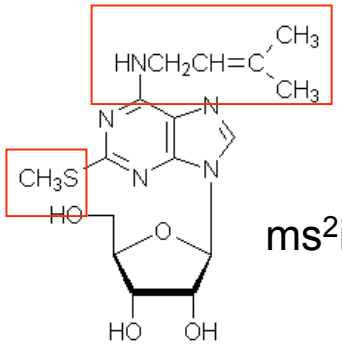
Main function of Purine-37 during translation: a molecular 'Anti-slip'



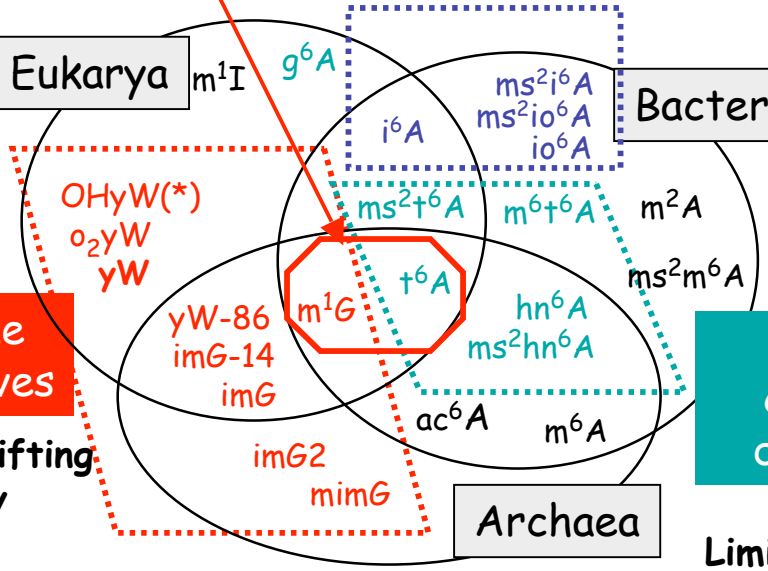
Probably the most ancient 'anti-slip'

Limit frameshifting frequency

Isopentenyl derivatives



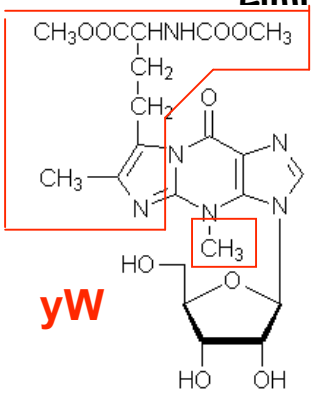
ms²i⁶A



Wyosine Derivatives

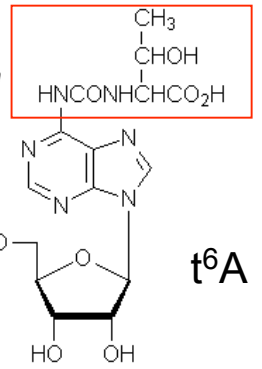
Threonyl-carbamoyl derivatives

Limit frameshifting frequency



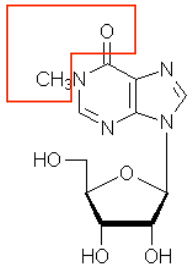
yW

Limit frameshifting frequency

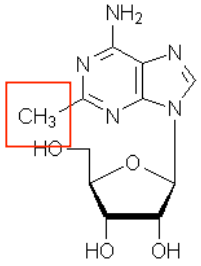


t⁶A

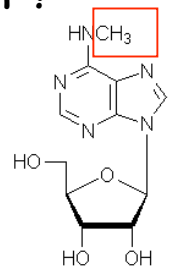
These probably also limit frameshift ?



m¹I



m²A



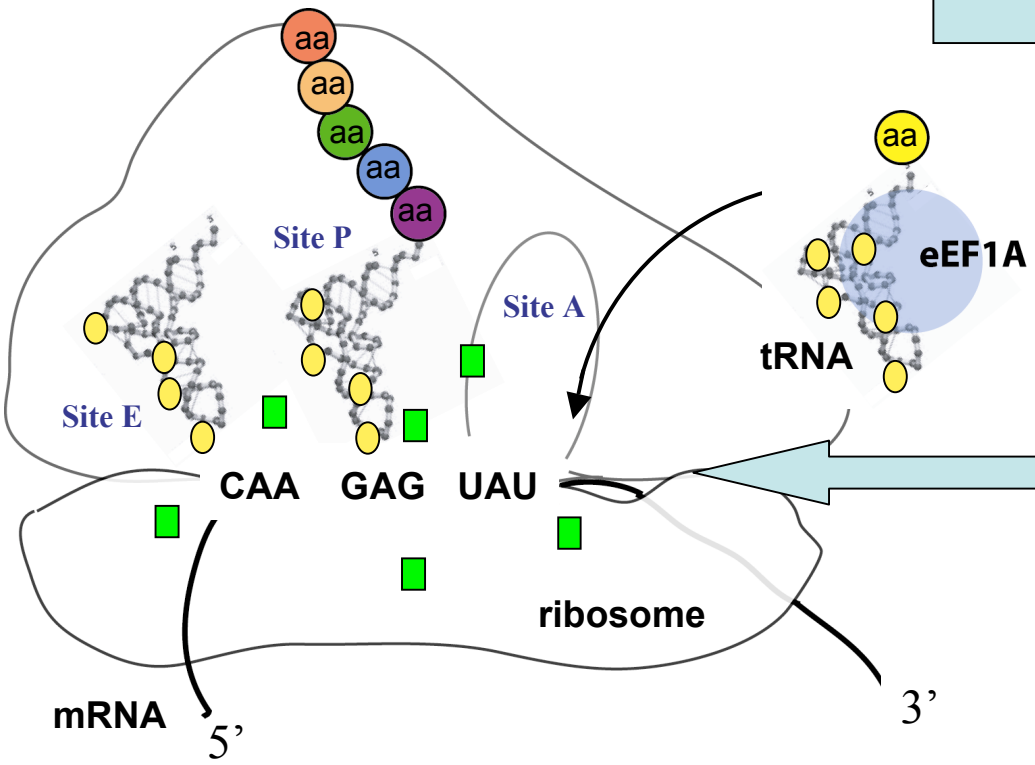
m⁶A

Grosjean - May 2011

CONCLUSION

Nucleotides modification in tRNAs *(also in rRNAs-not discussed here)* allows:

More accurate in-frame translation process



Phe	Ser	Tyr	Cys
Leu		stop	Trp
Leu	Pro	His	Arg
		Gln	
Ile	Thr	Asn	Ser
Met		Lys	Arg
Val	Ala	Asp	Gly
		Glu	

During cellular evolution, the progressive acquisition of genes coding for t+rRNA modification enzymes have obviously allowed to build-up the GC as it is to date

In other words, the GC has never been completely 'frozen' and it is still co-evolving with the emergence of new modification enzymes

For more information, consult the following book:

DNA and RNA Modification Enzymes:
Structure, Mechanism, Function and Evolution

Published by LANDES BioSciences , 2009
Editor: Henri Grosjean

and :

Fine-Tuning of RNA Functions by Modification and Editing

Published by Springer-Verlag, 2005
Editor: Henri Grosjean

These books are available at the Library of University Paris-11 - Orsay

Also :

- a Review paper recently published in FEBS Letters (2010), vol 584, pp 252-264, entitled:

Deciphering synonymous codons in the three domains of Life: Co-evolution with specific tRNA modification enzymes

By H. Grosjean, V. de Crécy-Lagard and C. Marck

and :

- IUBMB Life (2007), vol 59, pp 634658

Comparative Rnomics and Modomics in Mollicutes: prediction of gene function and evolutionary implications

By de Crécy-Lagard V, Marck C, Brochier-Armanet C and Grosjean H.

PDF available upon request to HG
henri.grosjean@igmors.u-psud.fr