Control over Translation: The Third Genetic code

E. Westhof
Institut de biologie moléculaire et cellulaire du CNRS
Université de Strasbourg, France
2018

The genetic code is **redundant quasi-universal.**

But the representation of the codon table is arbitrary

From
« FRANCIS CRICK, Hunter of Life's Secrets »
Robert Olby
CSHLP, 2009.

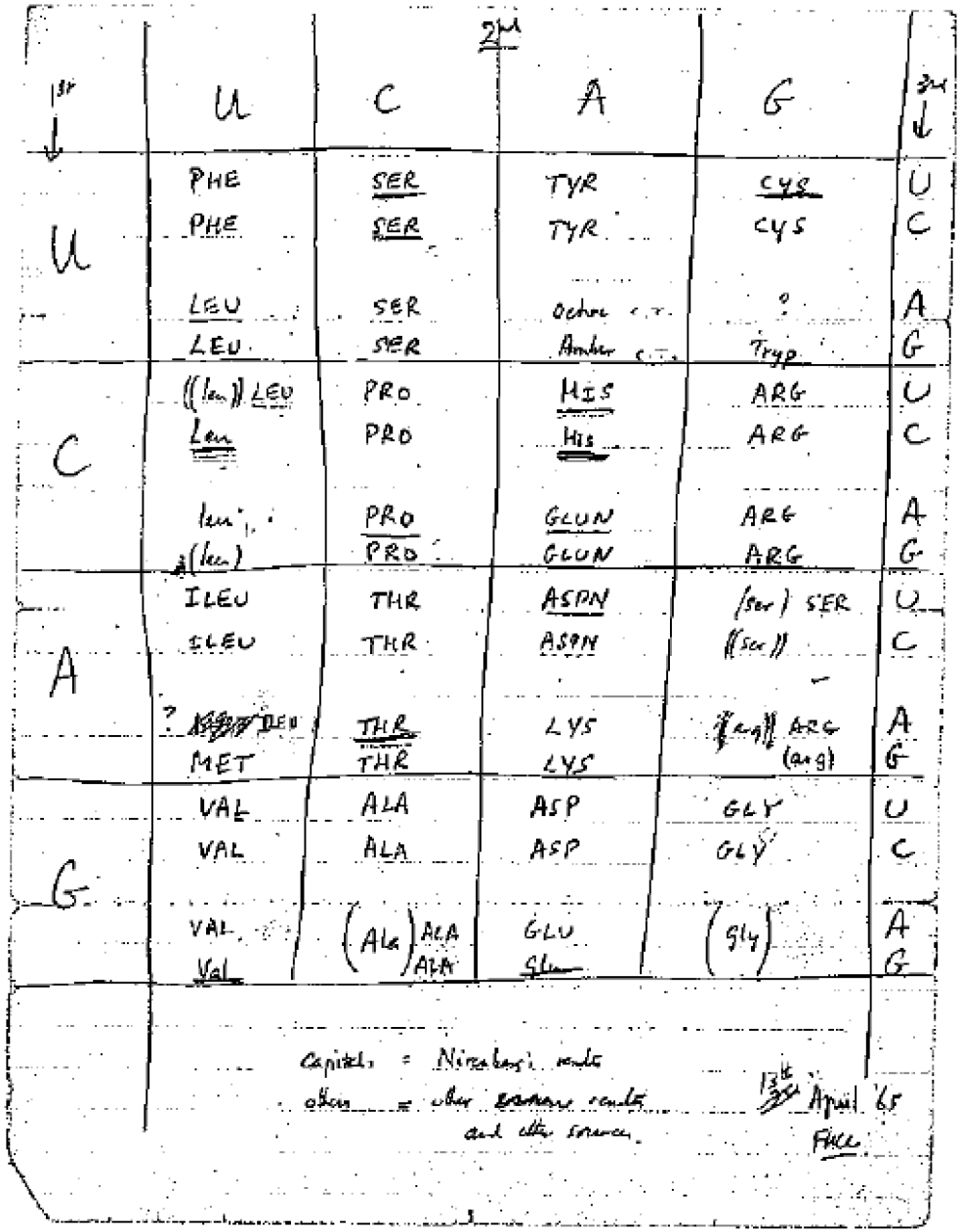
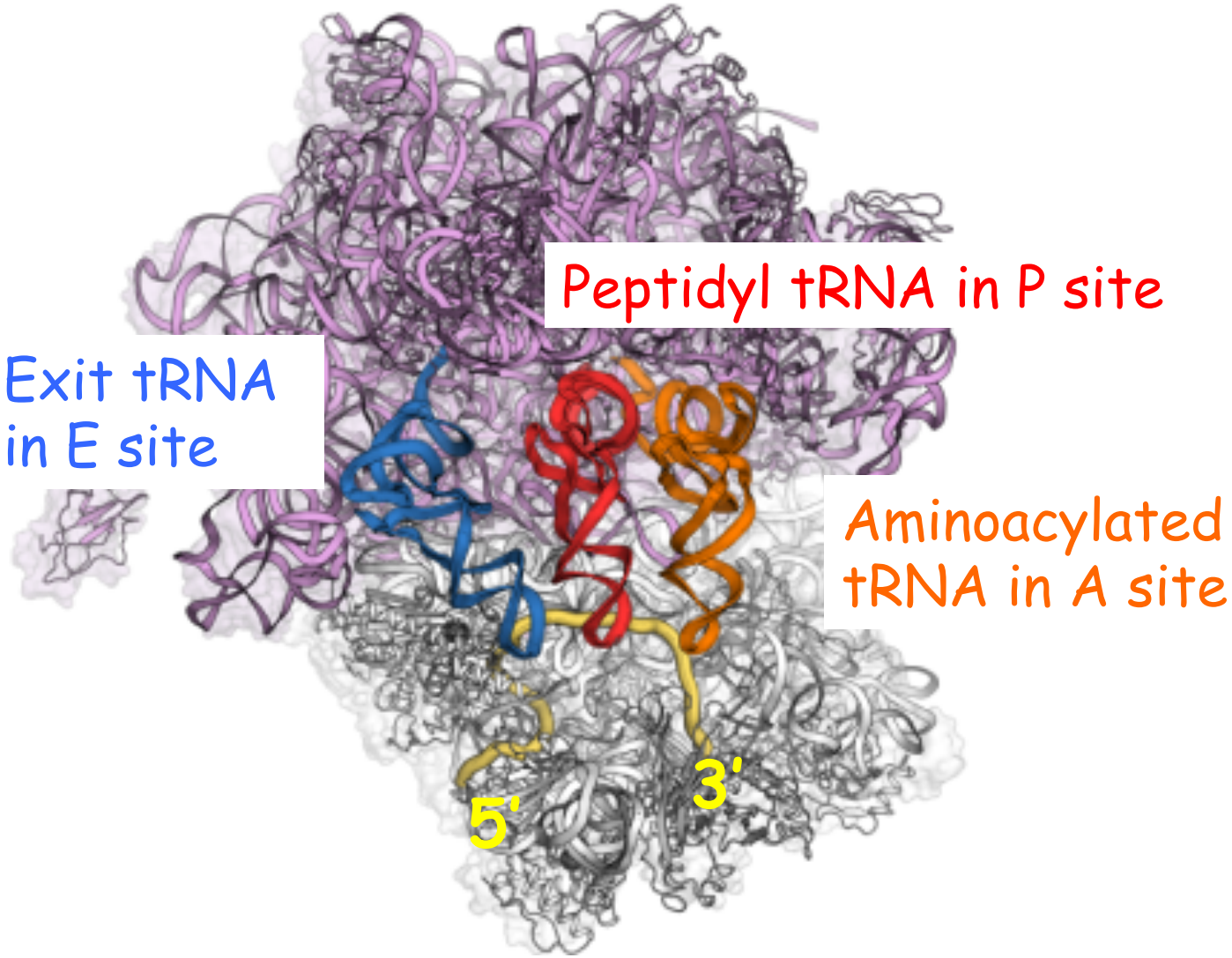


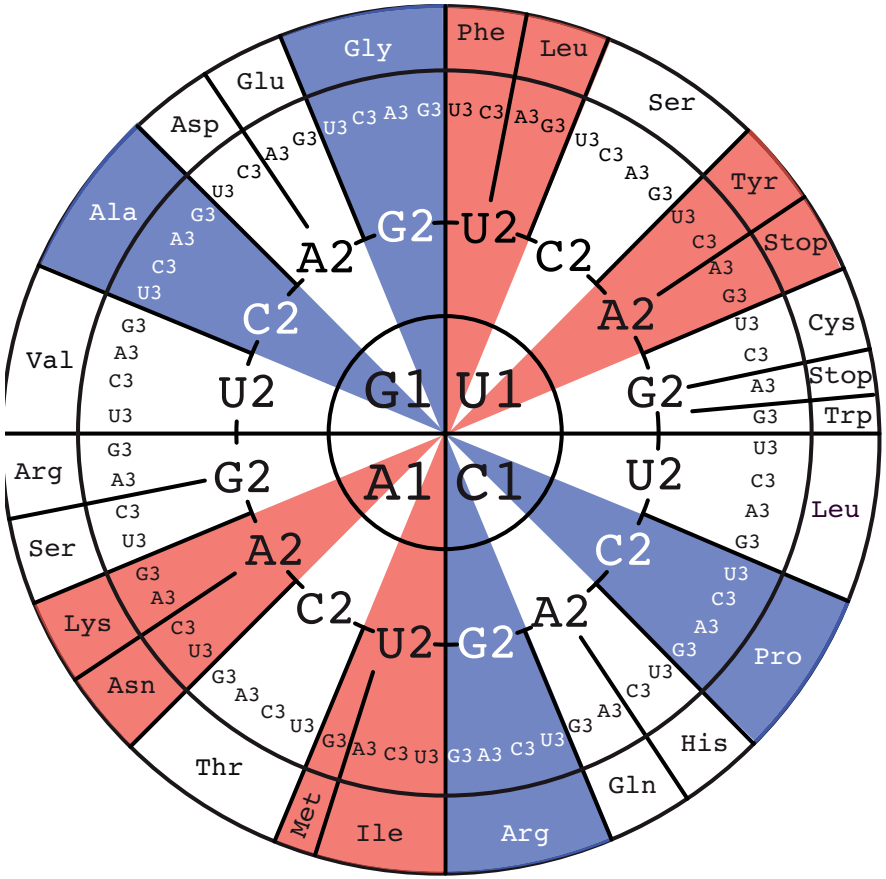
Figure 15.2 Crick's rough sketch of his checkerboard showing the stage reached in solving the genetic code in April 1965.

Only A-U pairs at 1st & 2nd

Only G=C pairs at 1st & 2nd

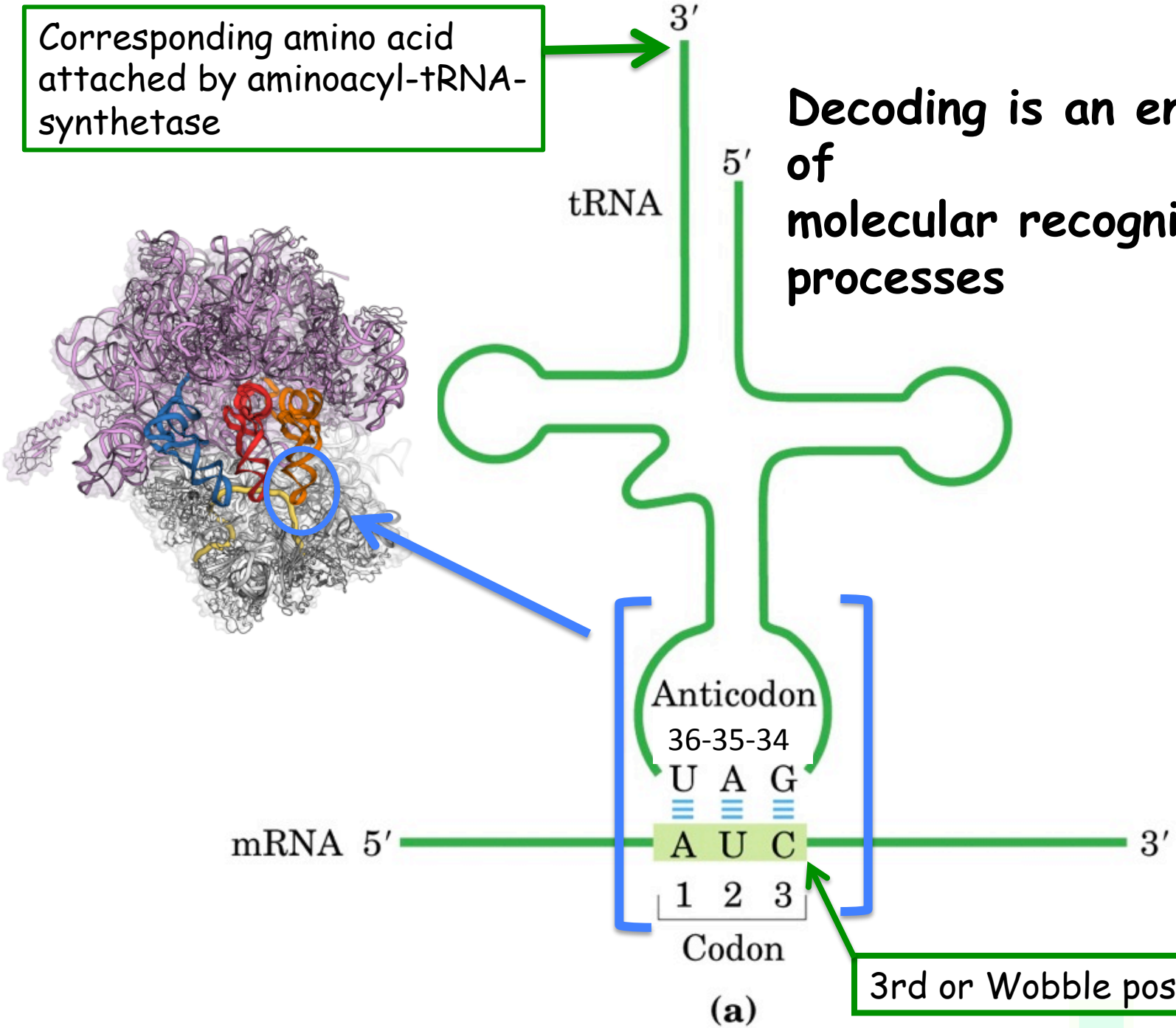
U		C		A		G	
UUU UUC	Phe	UCU UCC UCA UCG	Ser	UAU UAC	Tyr	UGU UGC	Cys
UUA UUG	Leu			UAA UAG	*Stop*	UGA	*Stop*
						UGG	Trp
CUU CUC CUA CUG	Leu	CCU CCC CCA CCG	Pro	CAU CAC	His	CGU CGC CGA CGG	Arg
				CAA CAG	Gln		
AUU AUC AUA	Ile	ACU ACC ACA ACG	Thr	AAU AAC	Asn	AGU AGC	Ser
AUG	Met			AAA AAG	Lys	AGA AGG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



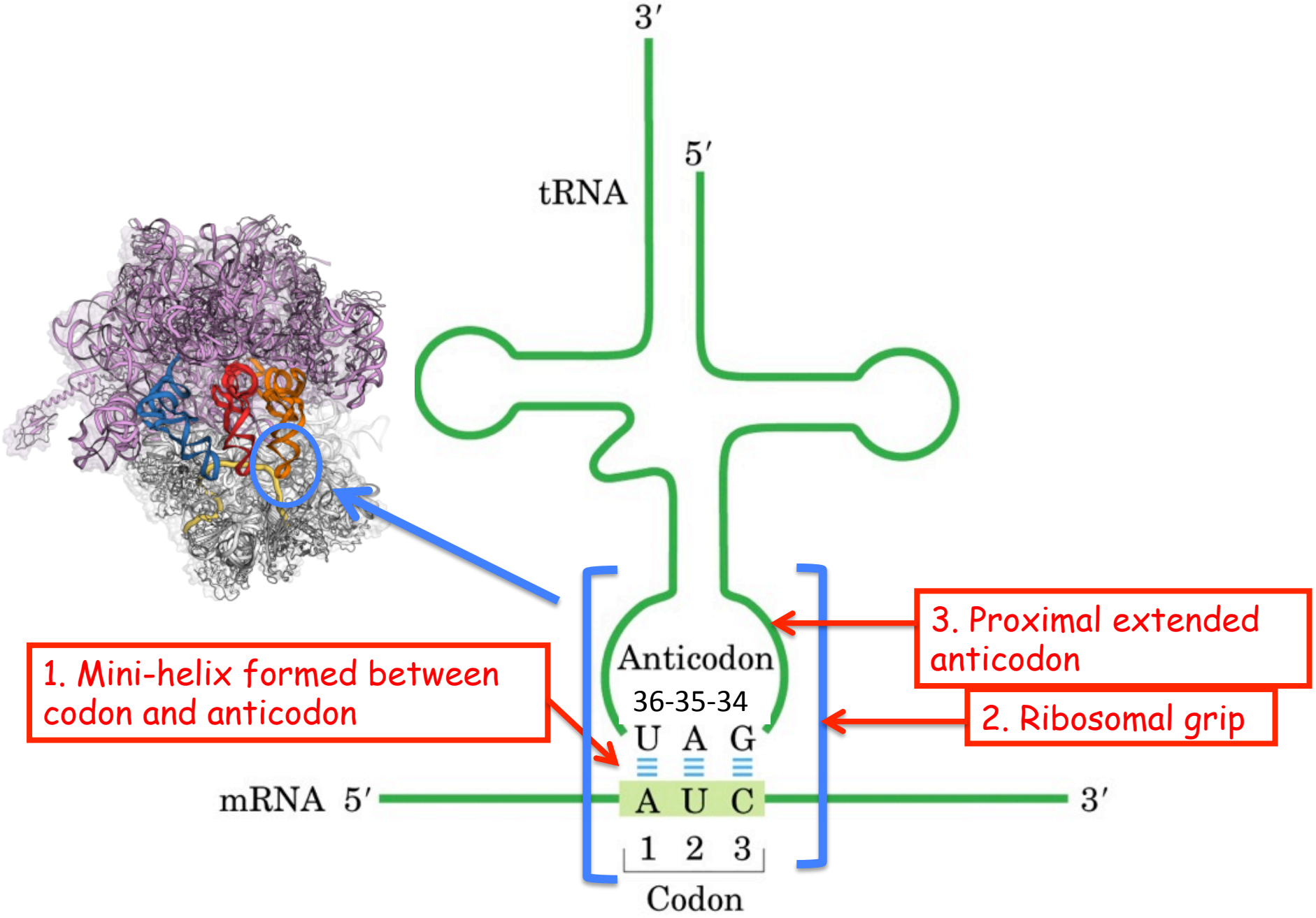

G1
U1
A1
C1
G2
U2
A2
C2
Gly
Phe
Leu
Ser
Tyr
Stop
Cys
Stop
Trp
Leu
Pro
His
Gln
Arg
Ile
Met
Thr
Asn
Lys
Ser
Arg
Val
Ala
Asp
Glu
U3 C3 A3 G3

Corresponding amino acid attached by aminoacyl-tRNA-synthetase

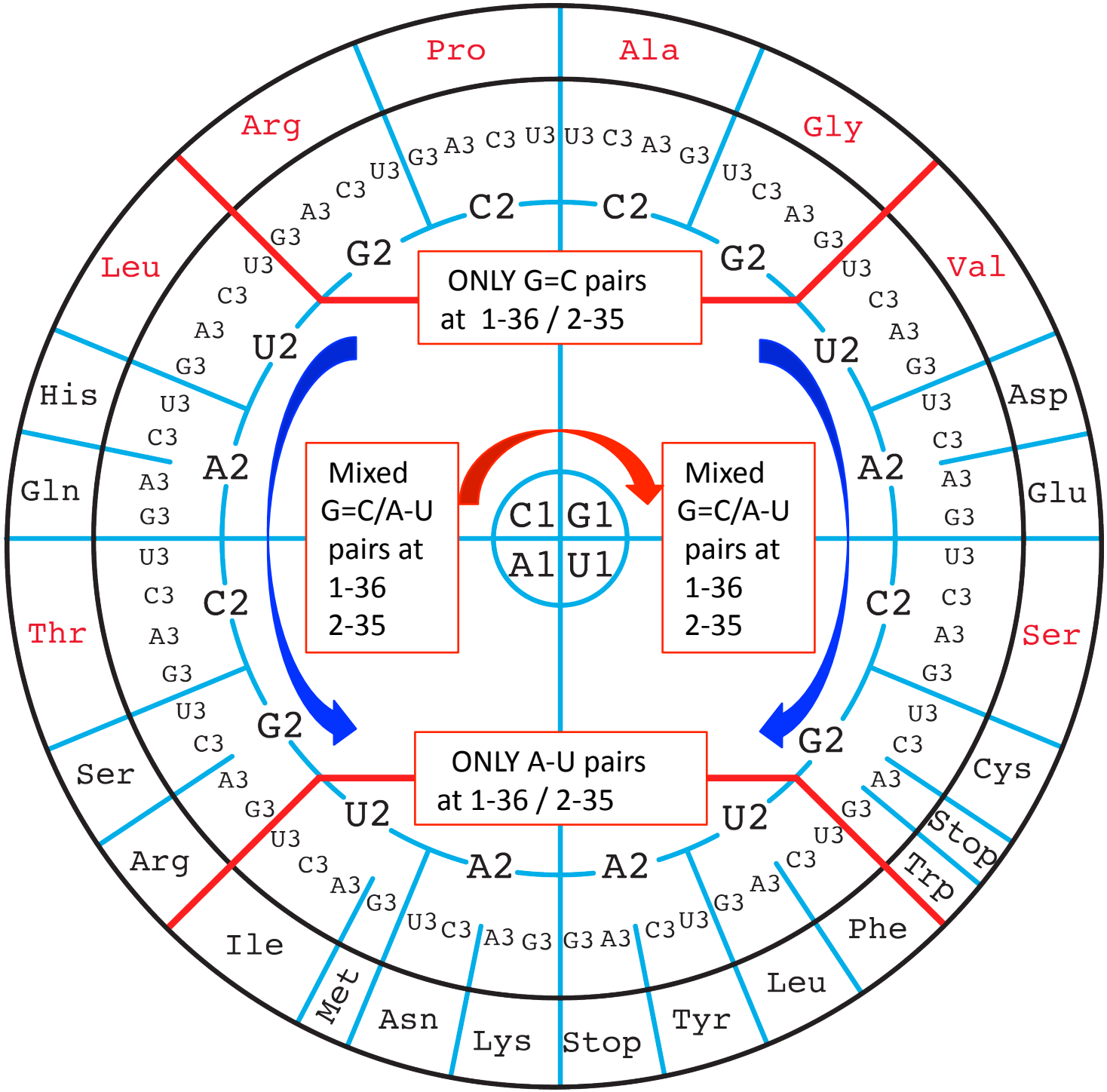
Decoding is an ensemble of molecular recognition processes



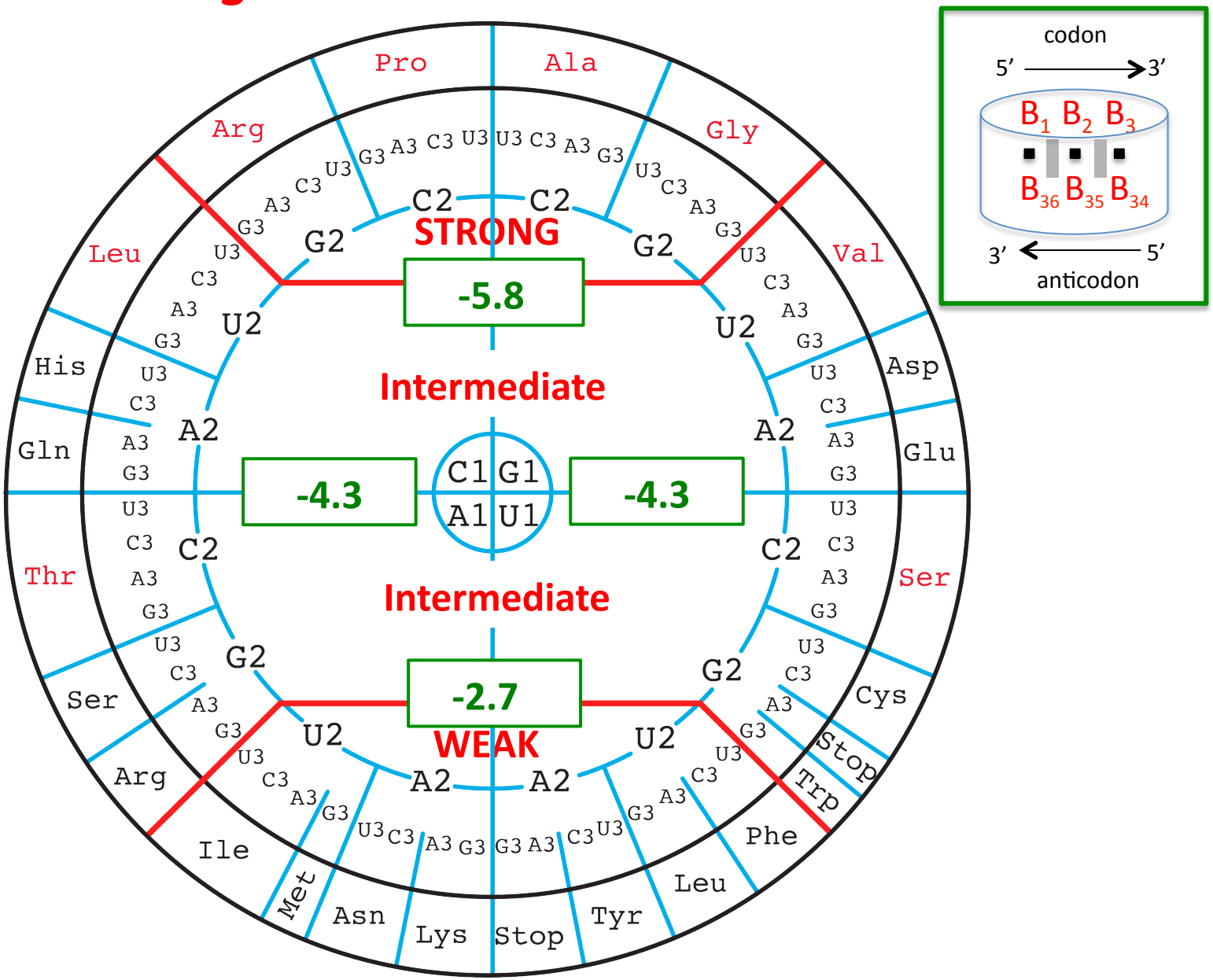
3rd or Wobble position (Crick)

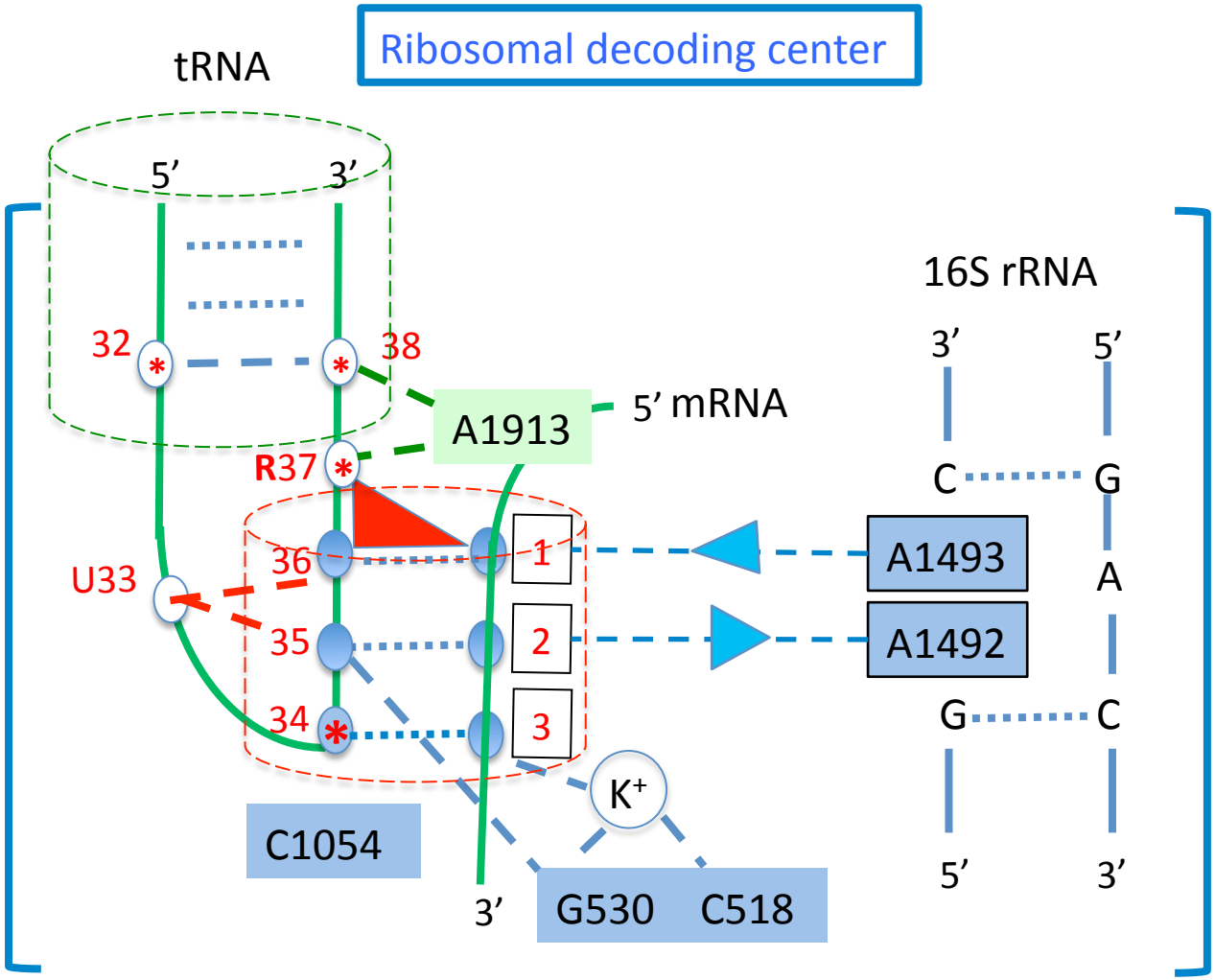

3′
5′
tRNA
Anticodon
36-35-34
U A G
A U C
1 2 3
Codon
mRNA 5′
3′
1. Mini-helix formed between codon and anticodon
2. Ribosomal grip
3. Proximal extended anticodon

(a)


ONLY G=C pairs
at 1-36 / 2-35
Mixed
G=C/A-U
pairs at
1-36
2-35
Mixed
G=C/A-U
pairs at
1-36
2-35
ONLY A-U pairs
at 1-36 / 2-35
C1 G1
A1 U1
Pro
Ala
Arg
Gly
Leu
Val
His
Asp
Gln
Glu
Thr
Ser
Ser
Cys
Stop
Trp
Arg
Phe
Ile
Leu
Met
Asn
Lys
Stop
Tyr
C2
G2
U2
A2

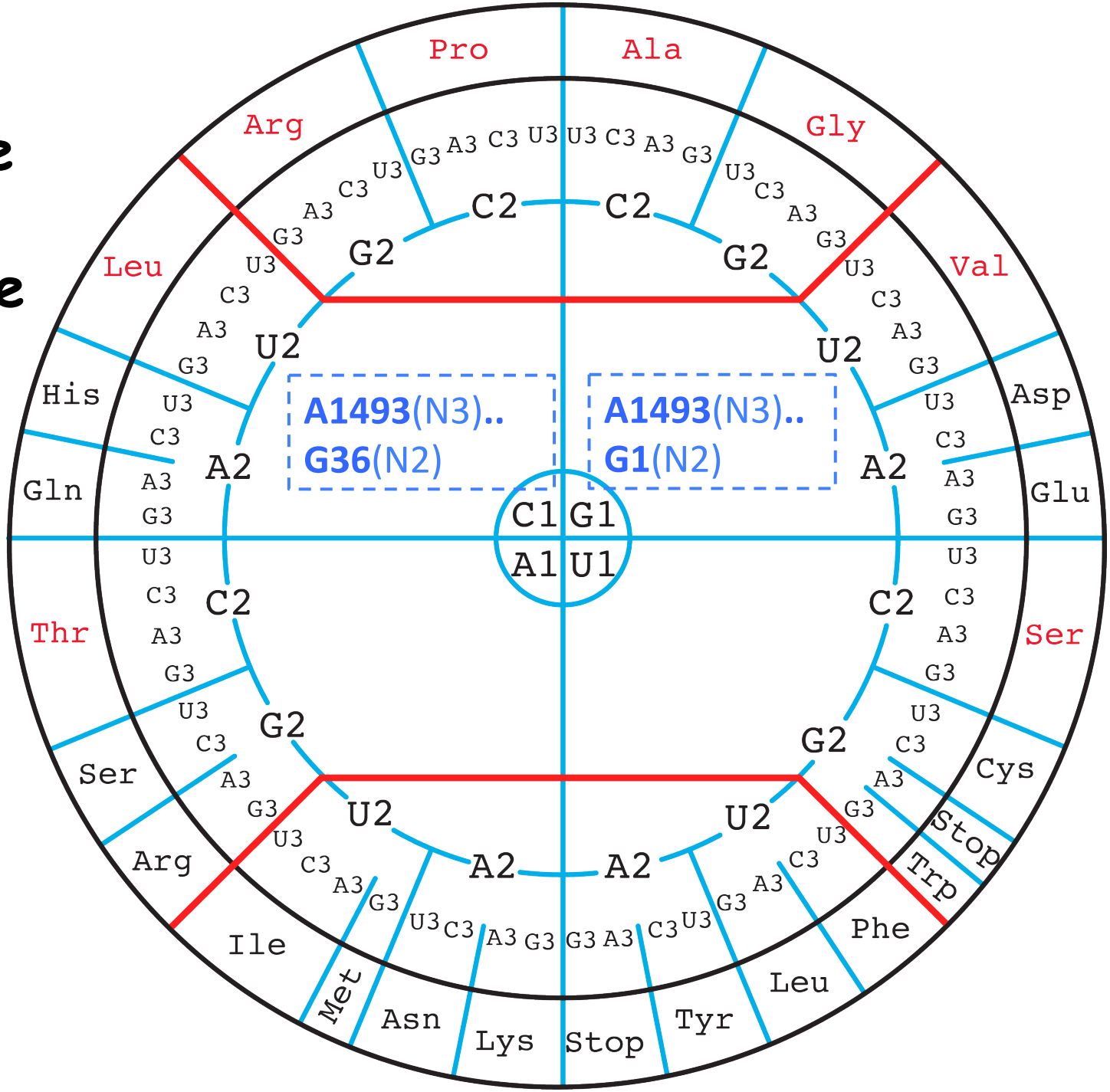
1. Calculated energies of the AntiCodon-Codon mini-helix



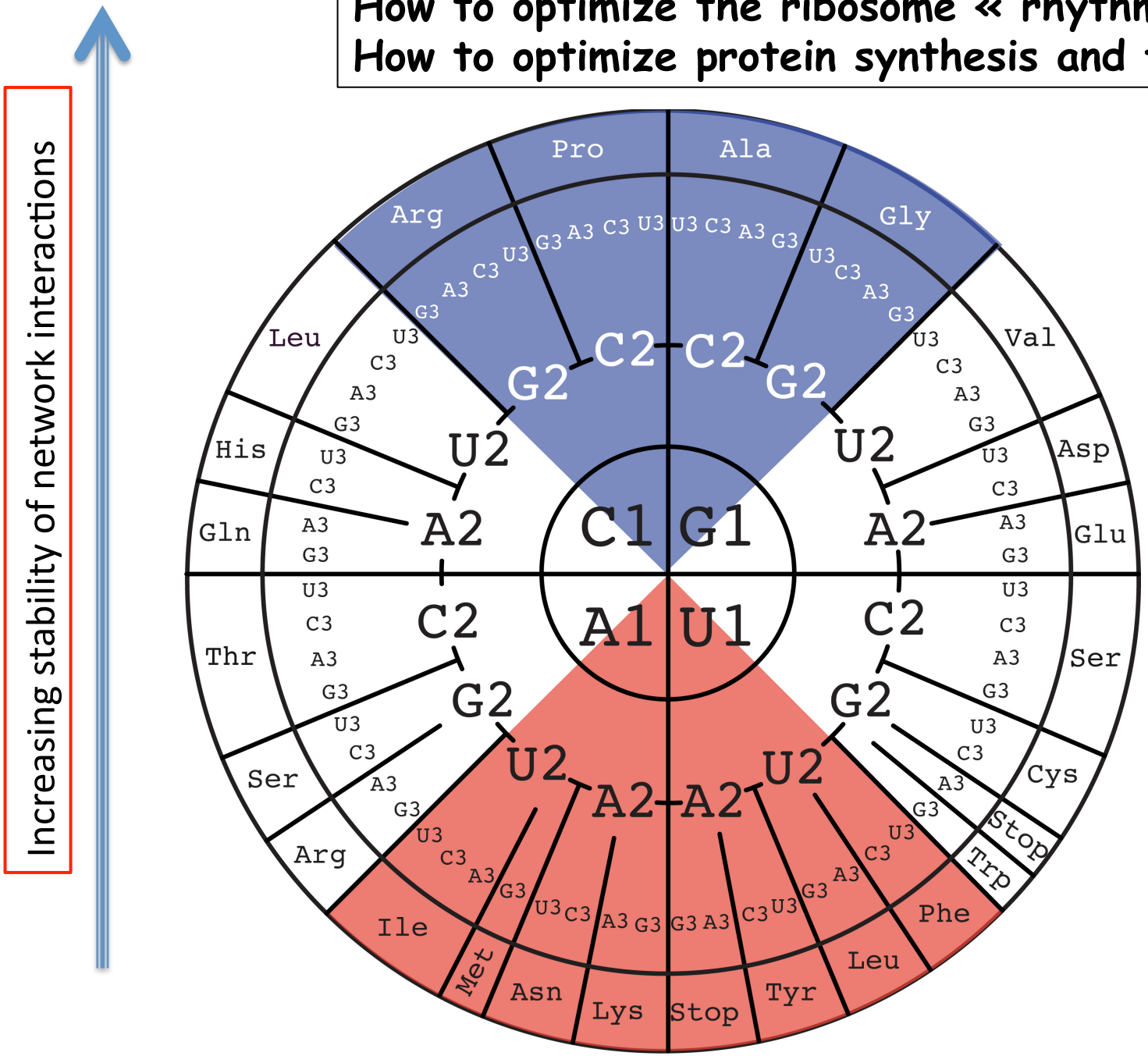


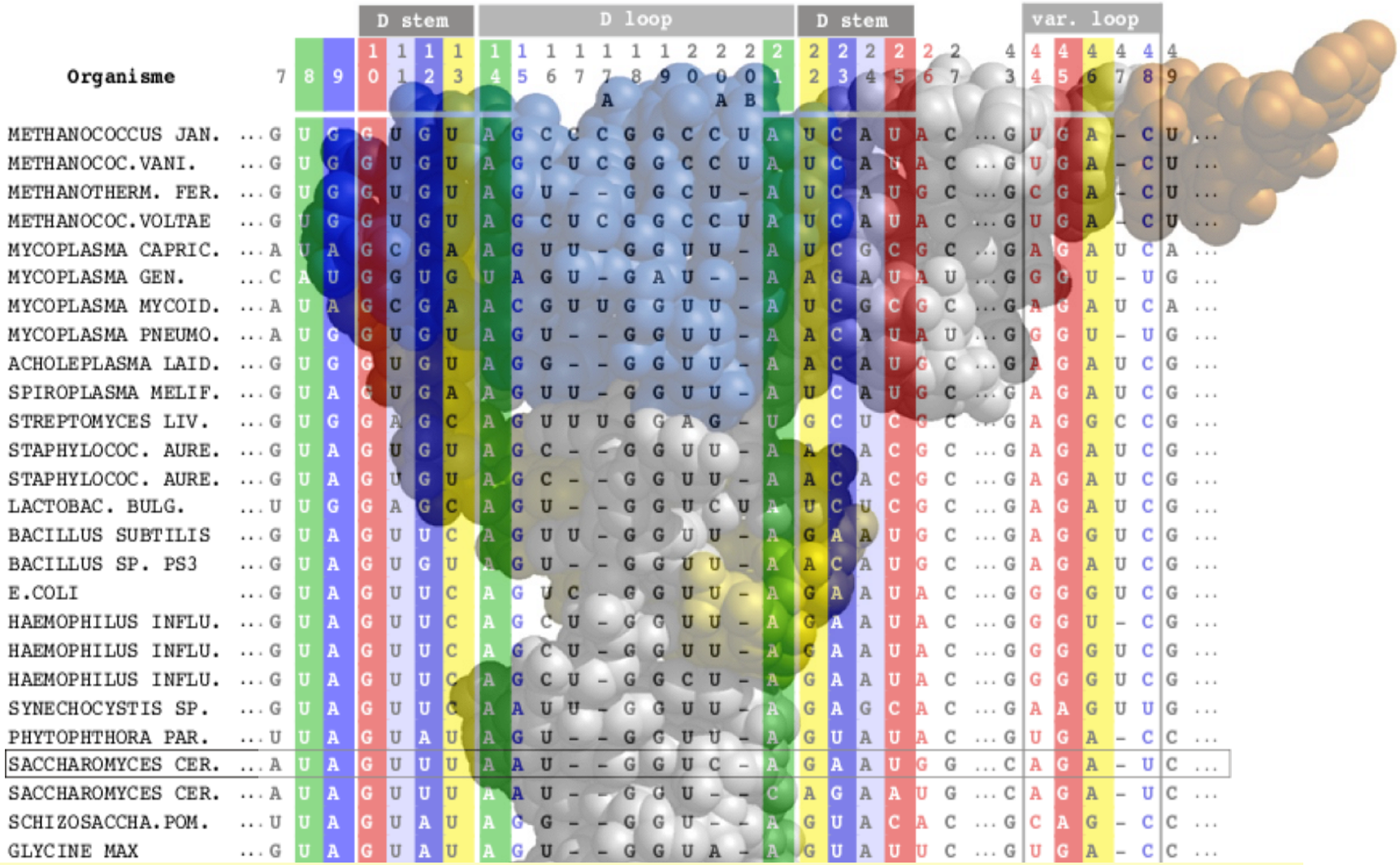
* < means often modified

Most of the contacts are invariant except those



How to optimize the ribosome « rhythm »?
How to optimize protein synthesis and folding?

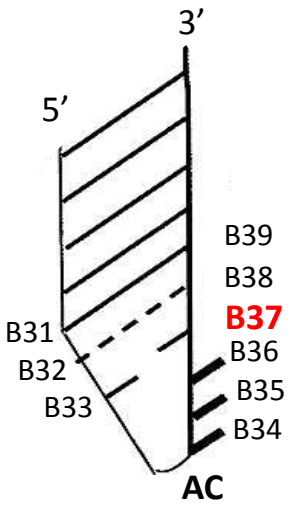




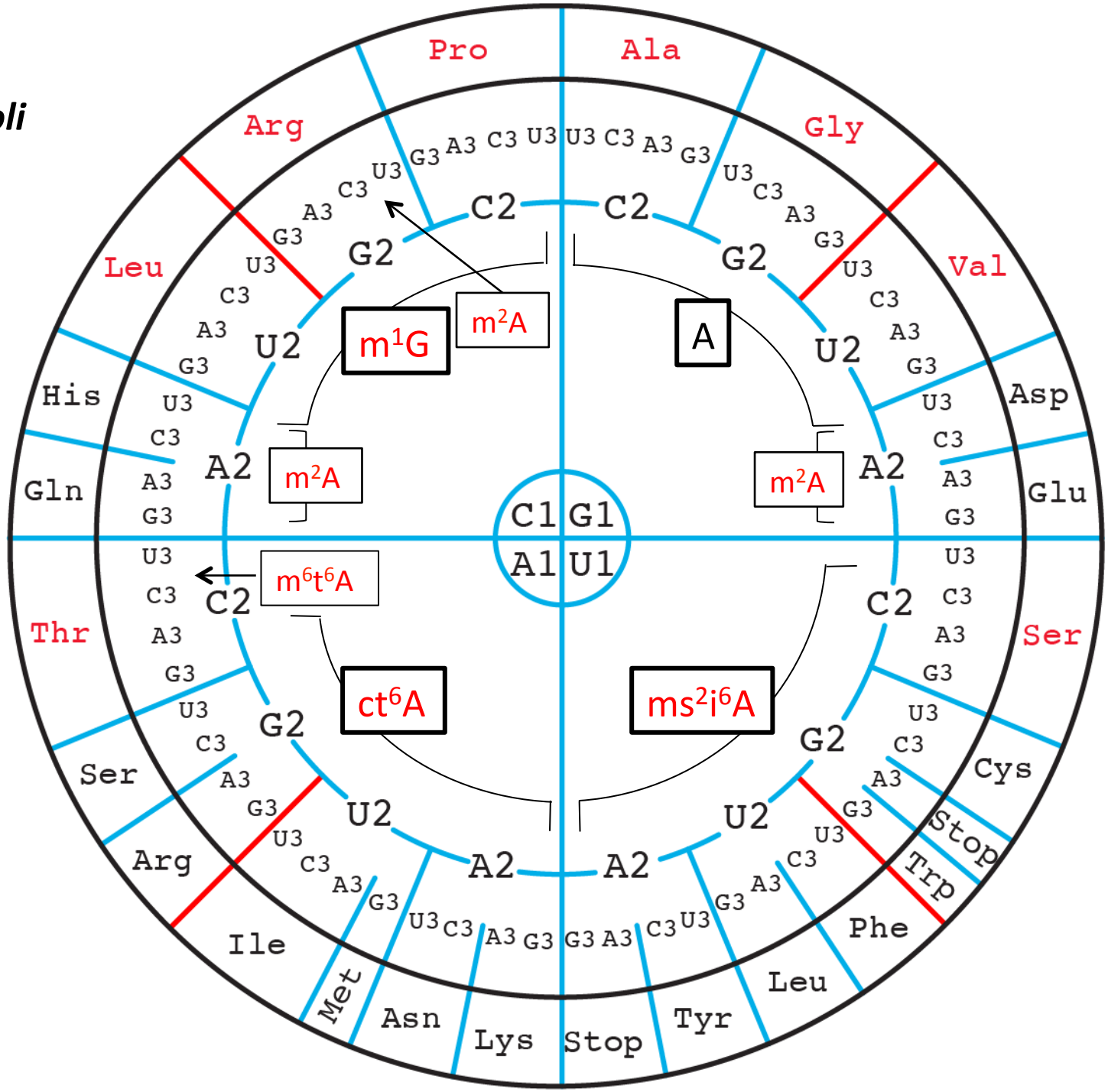
Organisme	7	8	9	D stem 10	11	12	13	D loop 14	15	16	17	17A	18	19	20	20A	20B	21	D stem 22	23	24	25	26	27	43	var. loop 44	45	46	47	48	49
METHANOCOCCUS JAN.	...G	U	G	G	U	G	U	A	G	C	C	C	G	G	C	C	U	A	U	C	A	U	A	C	...G	U	G	A	-	C	U ...
METHANOCOC.VANI.	...G	U	G	G	U	G	U	A	G	C	U	C	G	G	C	C	U	A	U	C	A	U	A	C	...G	U	G	A	-	C	U ...
METHANOTHERM. FER.	...G	U	G	G	U	G	U	A	G	U	-	-	G	G	C	U	-	A	U	C	A	U	G	C	...G	C	G	A	-	C	U ...
METHANOCOC.VOLTAE	...G	U	G	G	U	G	U	A	G	C	U	C	G	G	C	C	U	A	U	C	A	U	A	C	...G	U	G	A	-	C	U ...
MYCOPLASMA CAPRIC.	...A	U	A	G	C	G	A	A	G	U	U	-	G	G	U	U	-	A	U	C	G	C	G	C	...G	A	G	A	U	C	A ...
MYCOPLASMA GEN.	...C	A	U	G	G	U	G	U	A	G	U	-	G	A	U	-	-	A	A	G	A	U	A	U	...G	G	G	U	-	U	G ...
MYCOPLASMA MYCOID.	...A	U	A	G	C	G	A	A	C	G	U	U	G	G	U	U	-	A	U	C	G	C	G	C	...G	A	G	A	U	C	A ...
MYCOPLASMA PNEUMO.	...A	U	G	G	U	G	U	A	G	U	-	-	G	G	U	U	-	A	A	C	A	U	A	U	...G	G	G	U	-	U	G ...
ACHOLEPLASMA LAID.	...G	U	G	G	U	G	U	A	G	G	-	-	G	G	U	U	-	A	A	C	A	U	G	C	...G	A	G	A	U	C	G ...
SPIROPLASMA MELIF.	...G	U	A	G	U	G	A	A	G	U	U	-	G	G	U	U	-	A	U	C	A	U	G	C	...G	A	G	A	U	C	G ...
STREPTOMYCES LIV.	...G	U	G	G	A	G	C	A	G	U	U	U	G	G	A	G	-	U	G	C	U	C	G	C	...G	A	G	G	C	C	G ...
STAPHYLOCOC. AURE.	...G	U	A	G	U	G	U	A	G	C	-	-	G	G	U	U	-	A	A	C	A	C	G	C	...G	A	G	A	U	C	G ...
STAPHYLOCOC. AURE.	...G	U	A	G	U	G	U	A	G	C	-	-	G	G	U	U	-	A	A	C	A	C	G	C	...G	A	G	A	U	C	G ...
LACTOBAC. BULG.	...U	U	G	G	A	G	C	A	G	U	-	-	G	G	U	C	U	A	U	C	U	C	G	C	...G	A	G	A	U	C	G ...
BACILLUS SUBTILIS	...G	U	A	G	U	U	C	A	G	U	U	-	G	G	U	U	-	A	G	A	A	U	G	C	...G	A	G	G	U	C	G ...
BACILLUS SP. PS3	...G	U	A	G	U	G	U	A	G	U	-	-	G	G	U	U	-	A	A	C	A	U	G	C	...G	A	G	A	U	C	G ...
E.COLI	...G	U	A	G	U	U	C	A	G	U	C	-	G	G	U	U	-	A	G	A	A	U	A	C	...G	G	G	G	U	C	G ...
HAEMOPHILUS INFLU.	...G	U	A	G	U	U	C	A	G	C	U	-	G	G	U	U	-	A	G	A	A	U	A	C	...G	G	G	U	-	C	G ...
HAEMOPHILUS INFLU.	...G	U	A	G	U	U	C	A	G	C	U	-	G	G	U	U	-	A	G	A	A	U	A	C	...G	G	G	G	U	C	G ...
HAEMOPHILUS INFLU.	...G	U	A	G	U	U	C	A	G	C	U	-	G	G	C	U	-	A	G	A	A	U	A	C	...G	G	G	G	U	C	G ...
SYNECHOCYSTIS SP.	...G	U	A	G	U	U	C	A	A	U	U	-	G	G	U	U	-	A	G	A	G	C	A	C	...G	A	A	G	U	U	G ...
PHYTOPHTHORA PAR.	...U	U	A	G	U	A	U	A	G	U	-	-	G	G	U	U	-	A	G	U	A	U	A	C	...G	U	G	A	-	C	C ...
SACCHAROMYCES CER.	...A	U	A	G	U	U	U	A	A	U	-	-	G	G	U	C	-	A	G	A	A	U	G	G	...C	A	G	A	-	U	C ...
SACCHAROMYCES CER.	...A	U	A	G	U	U	U	A	A	U	-	-	G	G	U	-	-	C	A	G	A	A	U	G	...C	A	G	A	-	U	C ...
SCHIZOSACCHA.POM.	...U	U	A	G	U	A	U	A	G	G	-	-	G	G	U	-	-	A	G	U	A	C	A	C	...G	C	A	G	-	C	C ...
GLYCINE MAX	...G	U	A	G	U	A	U	A	G	U	-	-	G	G	U	A	-	A	G	U	A	U	U	C	...G	U	G	A	-	C	C ...

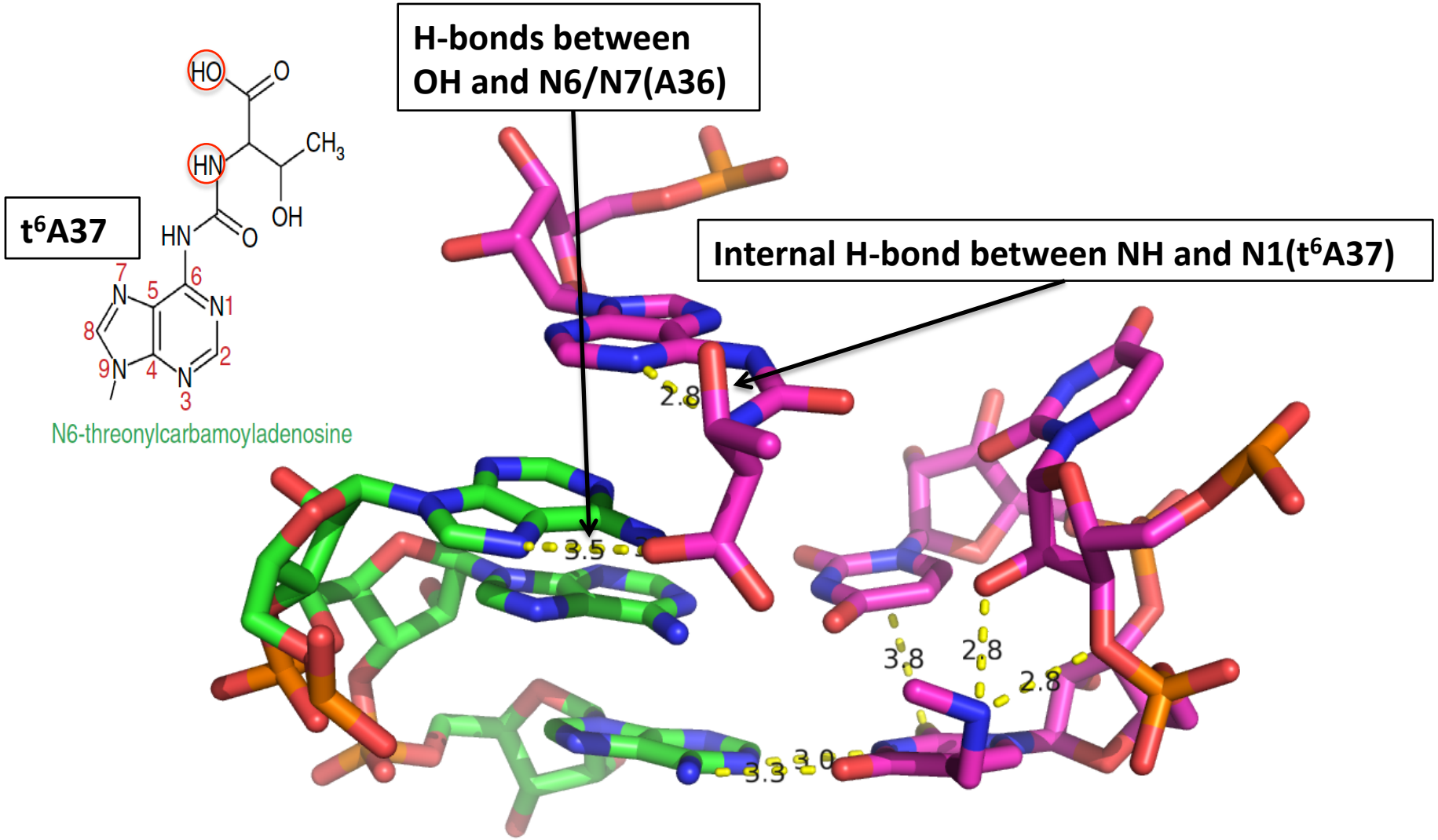
tRNAs are very active partners during ribosomal translation

Escherichia coli



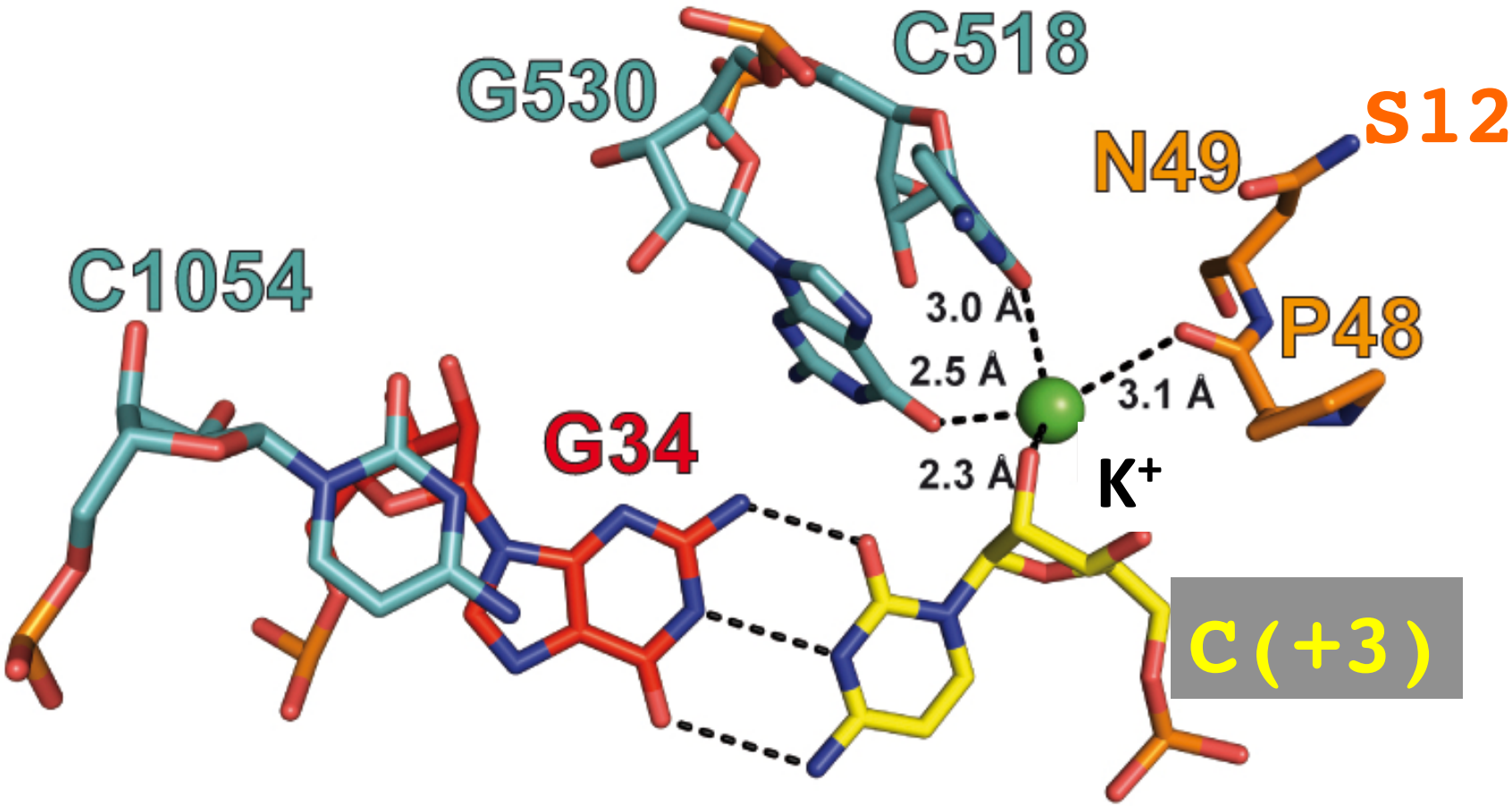
B37

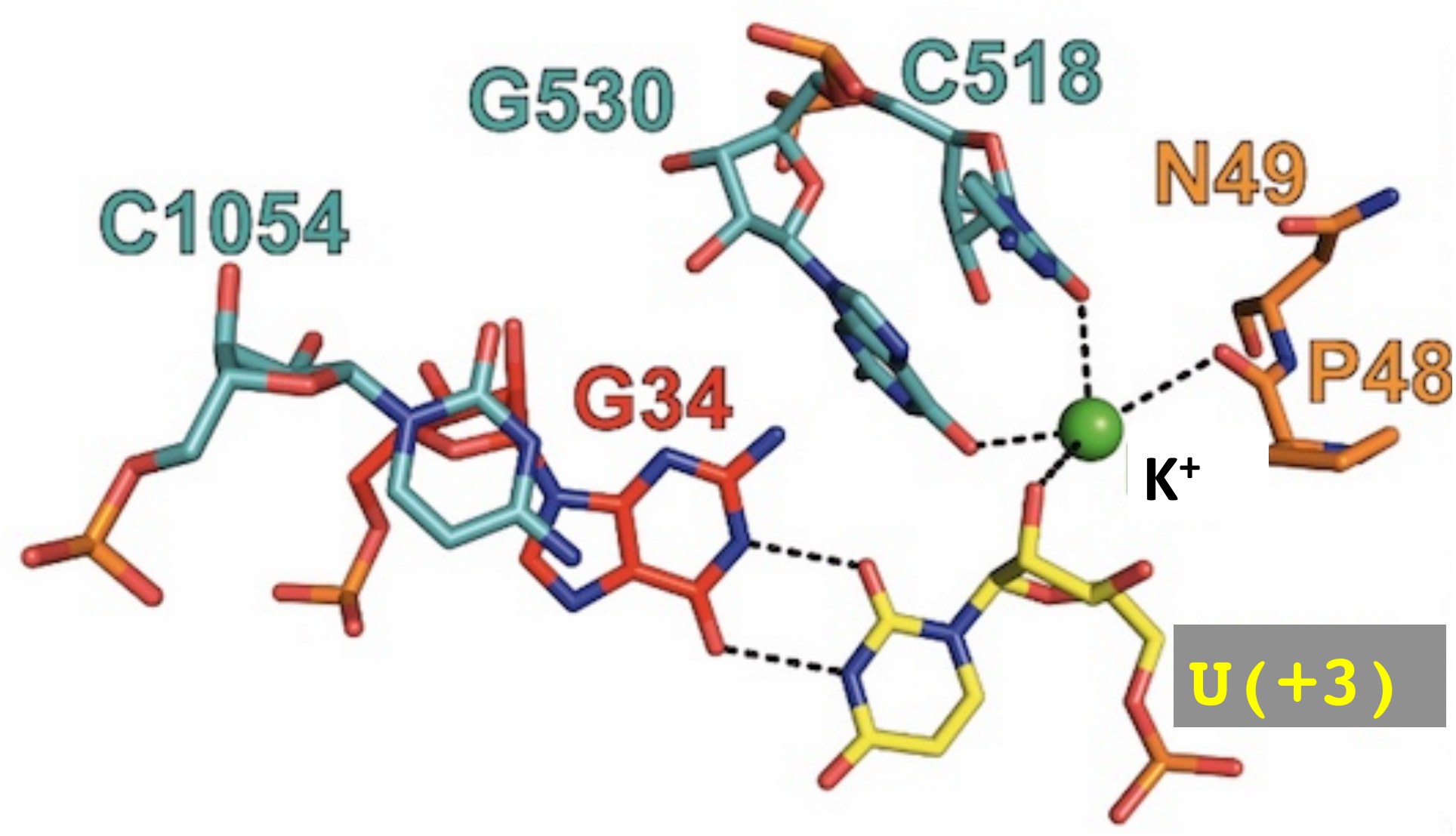




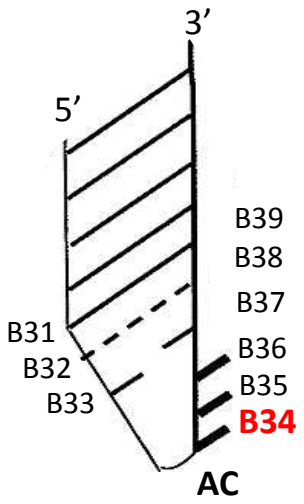
Rozov, et al. Nature Comm. 7, 10457(2016)

Around the third base pair



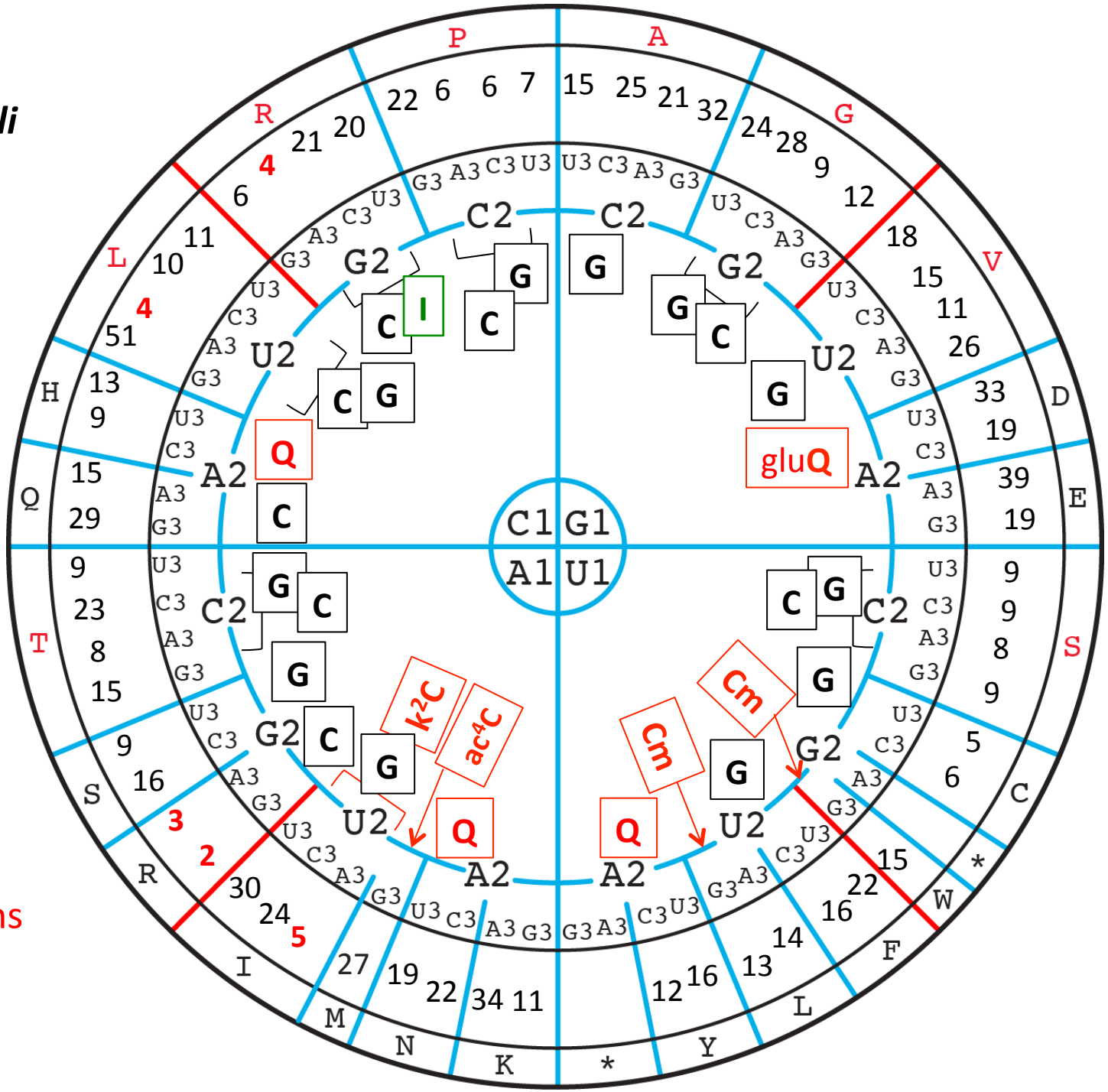

G530
C518
N49
C1054
P48
G34
K+
U(+3)

Escherichia coli



B34
except
(U34)

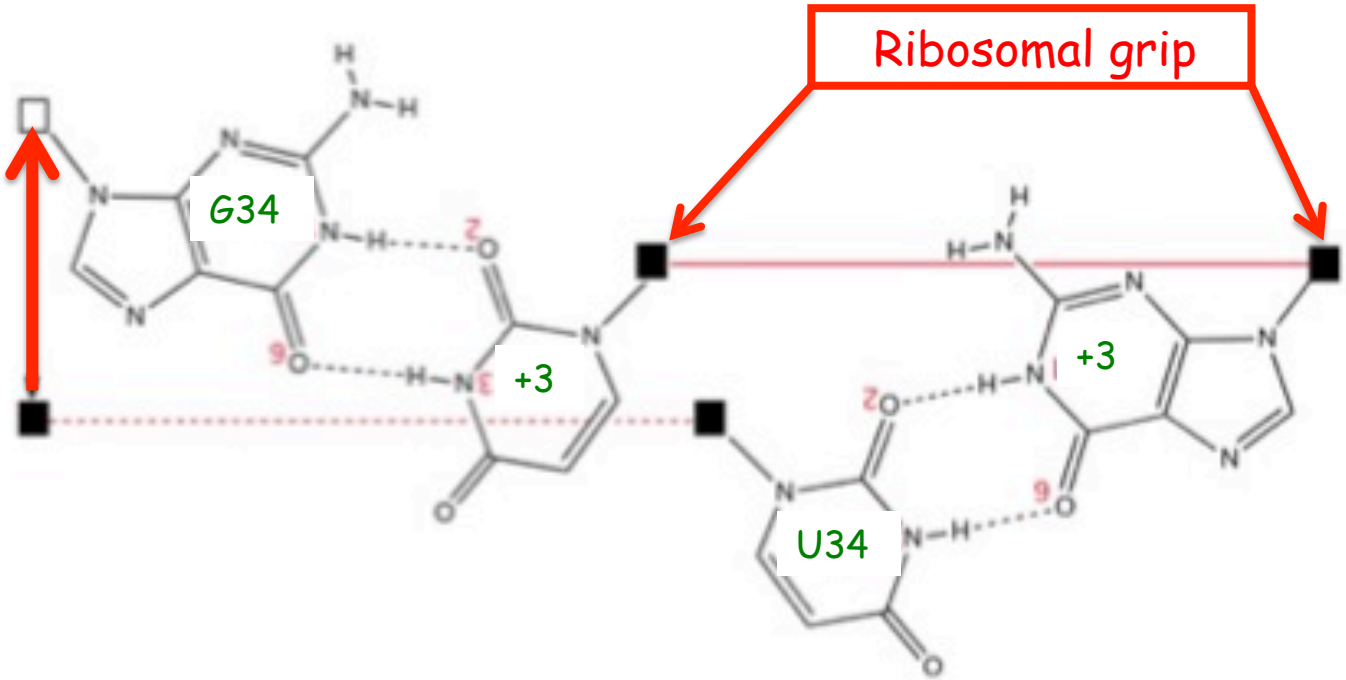
Y3-ending codons



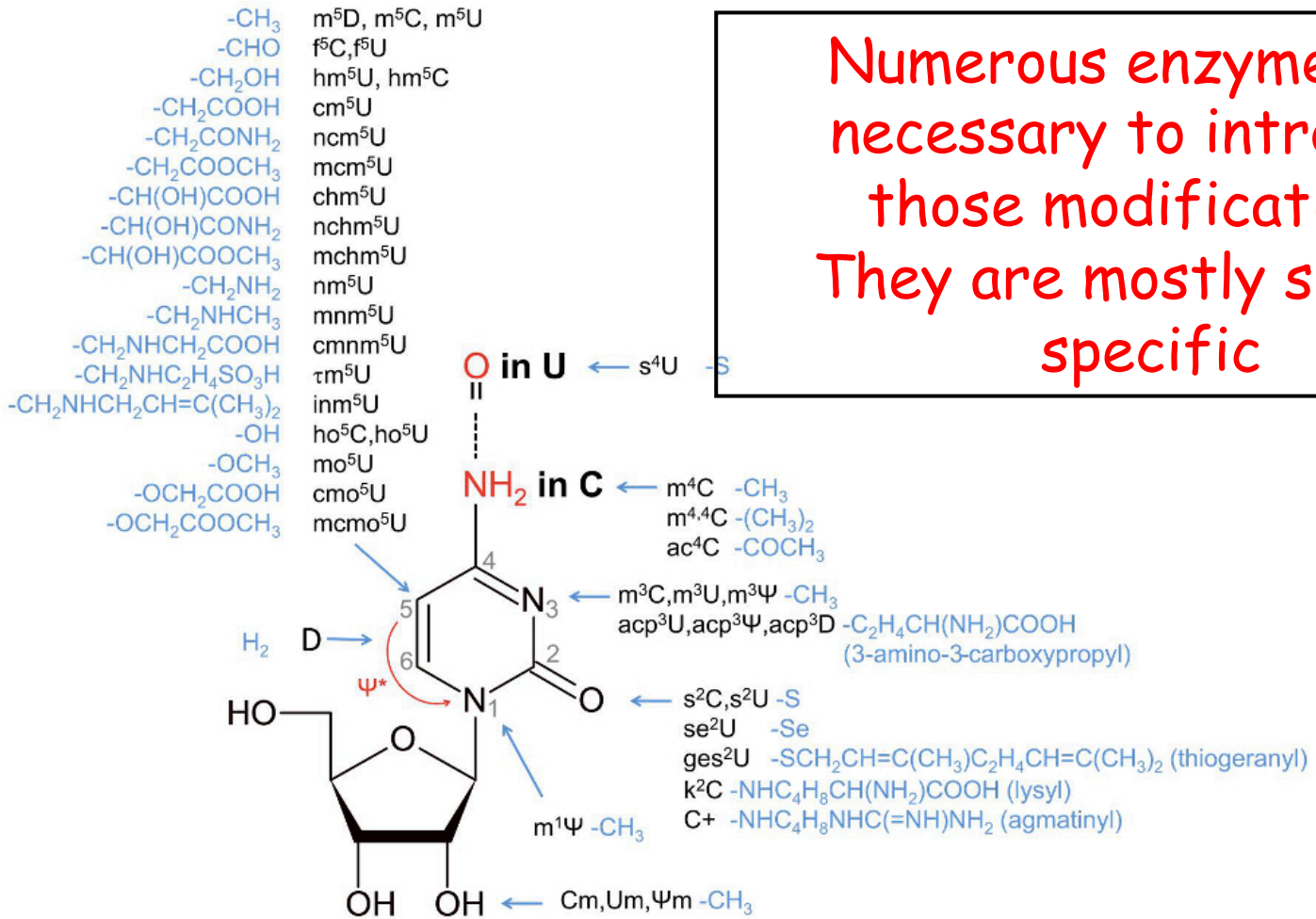
Uridine in the mRNA G34oU(+3)

versus

Uridine in the anticodon U34oG(+ 3)



More than 110 modified nucleotides in transfert RNA



Numerous enzymes are necessary to introduce those modifications. They are mostly species specific

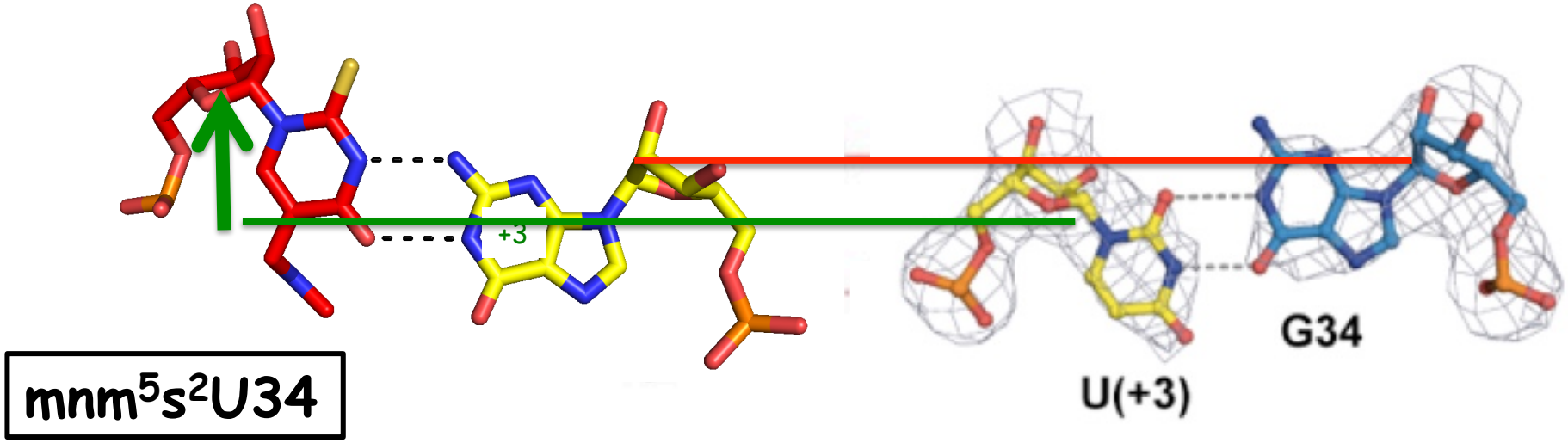
From Modomics database 2013 (http://modomics.genesilico.pl/)

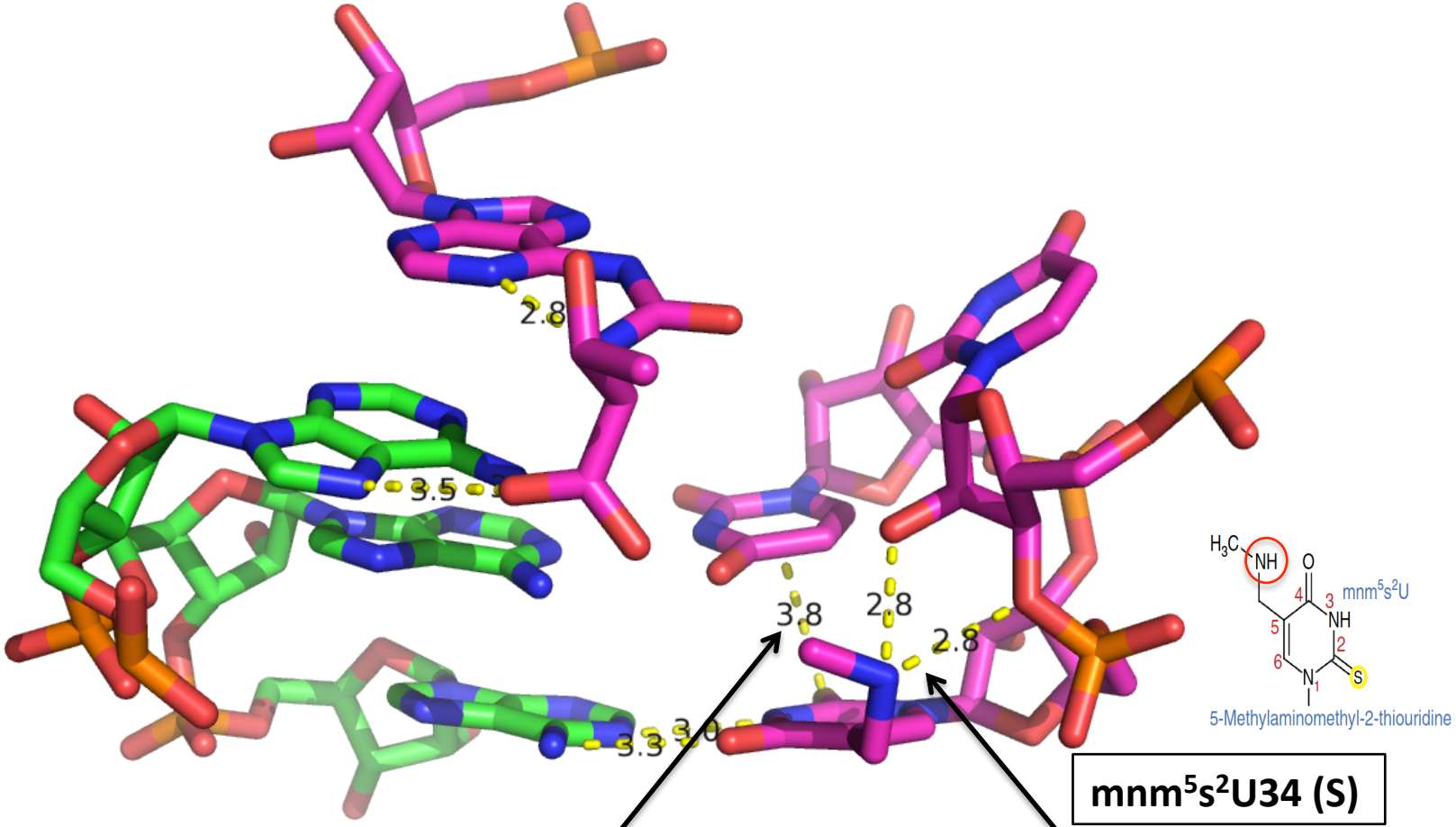
Modified Uridine $mnm^5s^2U34oU(+3)$

versus

Non-modified Uridine U34oG(+ 3)

in the anticodon





mnm^{5}s^{2}U34 (S)

S^2 stacking below U35

H-bonds between NH and U33 sugar

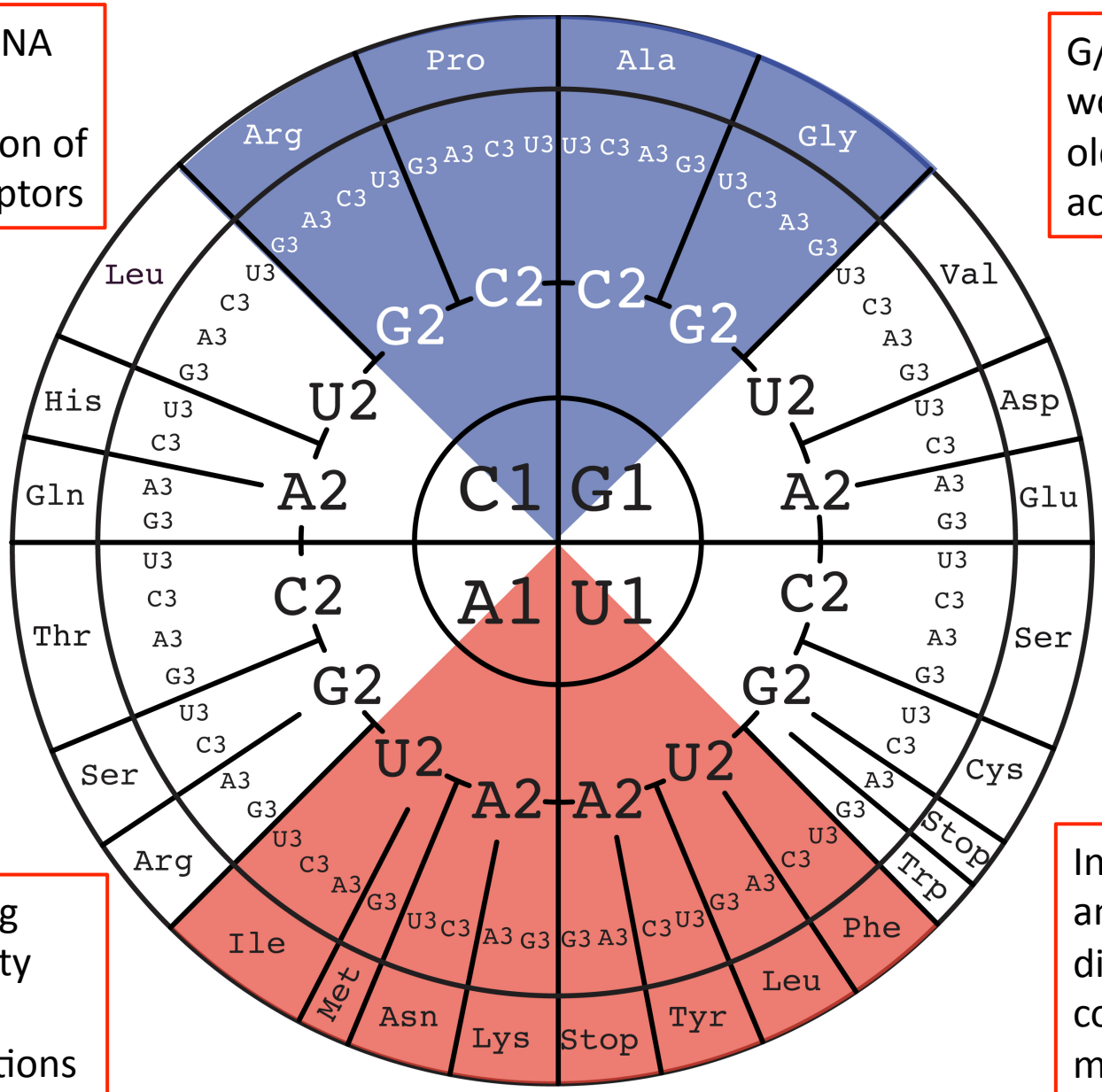
Rozov, et al. Nature Comm. 7, 10457(2016)

Coevolution of the genetic code with metabolic pathways

Increasing stability of network interactions

Strong tRNA binding
Modulation of helix adaptors

Increasing complexity of modifications at 34 & 37

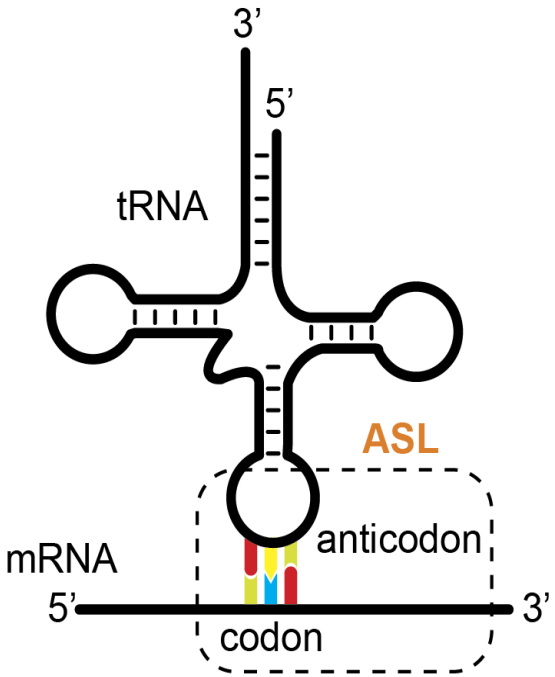


G/C-rich RNA world with oldest amino acids

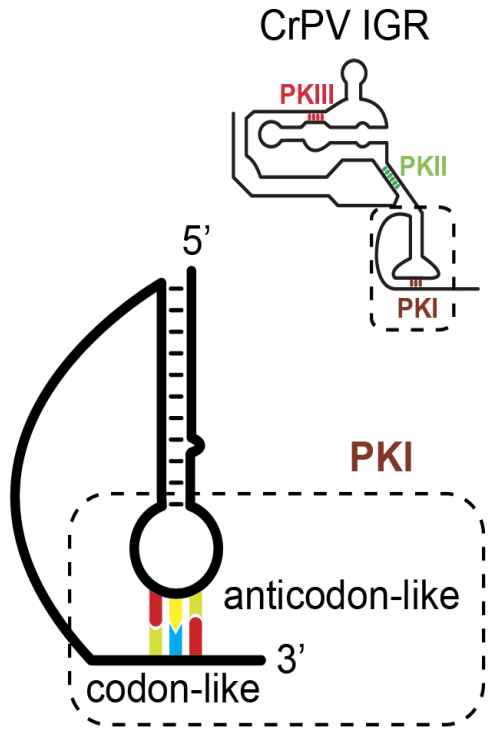
Increase in A/U and amino acid diversity coupled with metabolic enzymes

A microfluidic-based assay recapitulates eukaryotic ribosomal translation

Model system uses the intergenic IRES of Cricket Paralysis Virus coupled with the GFP as reporter

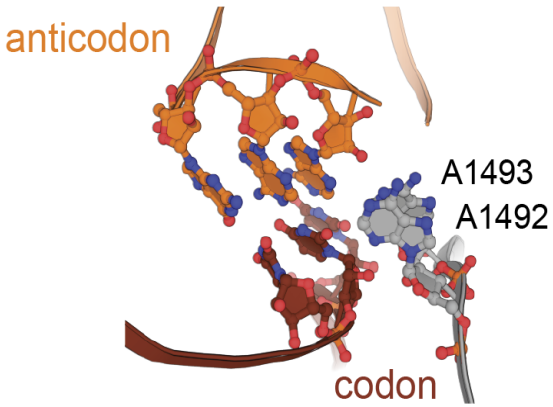


INTERmolecular
(2 molecules)

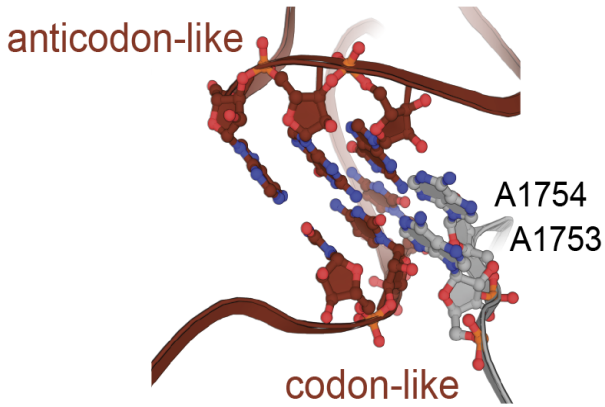


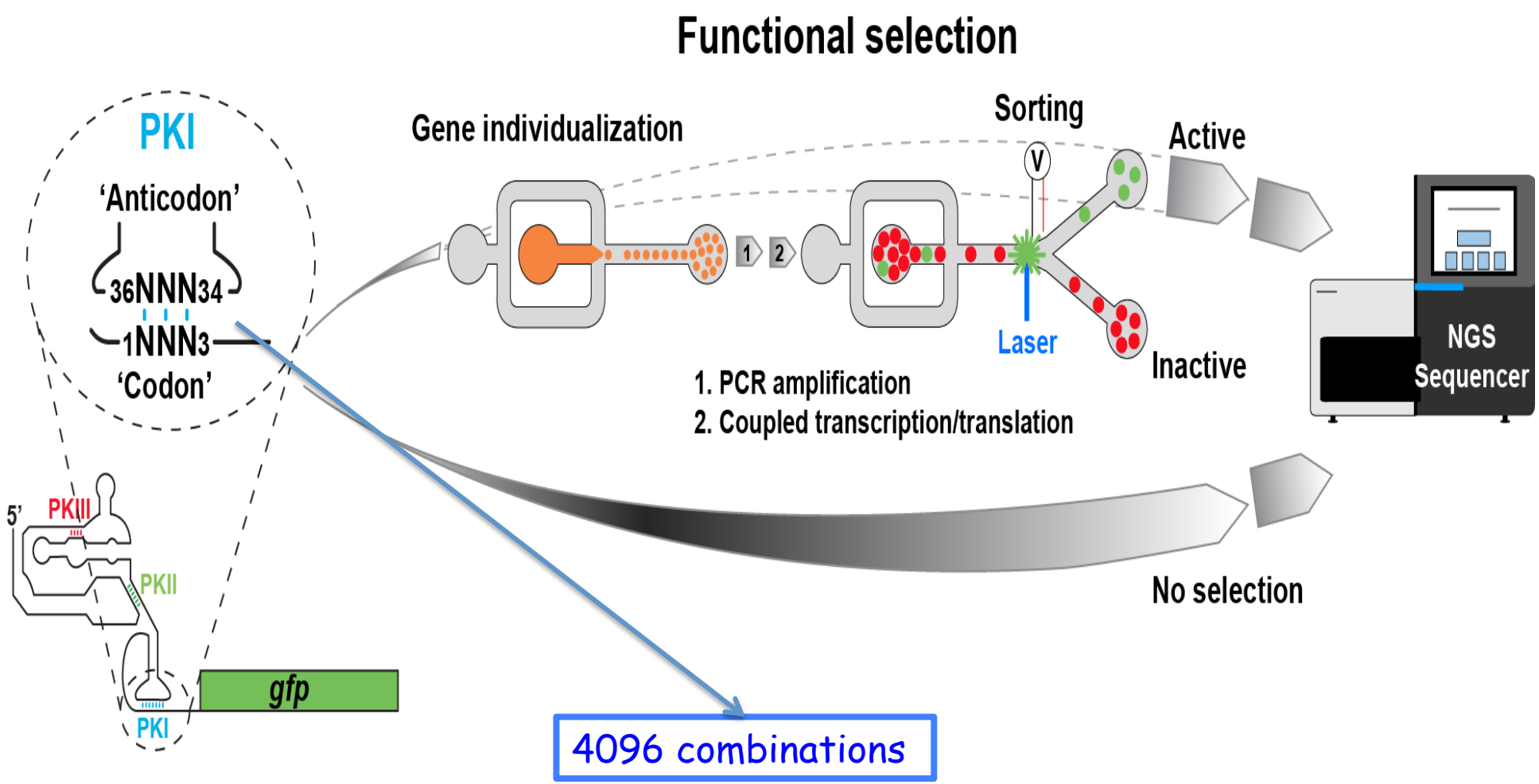
INTRAmolecular
(1 molecule pseudoknot)

ASL

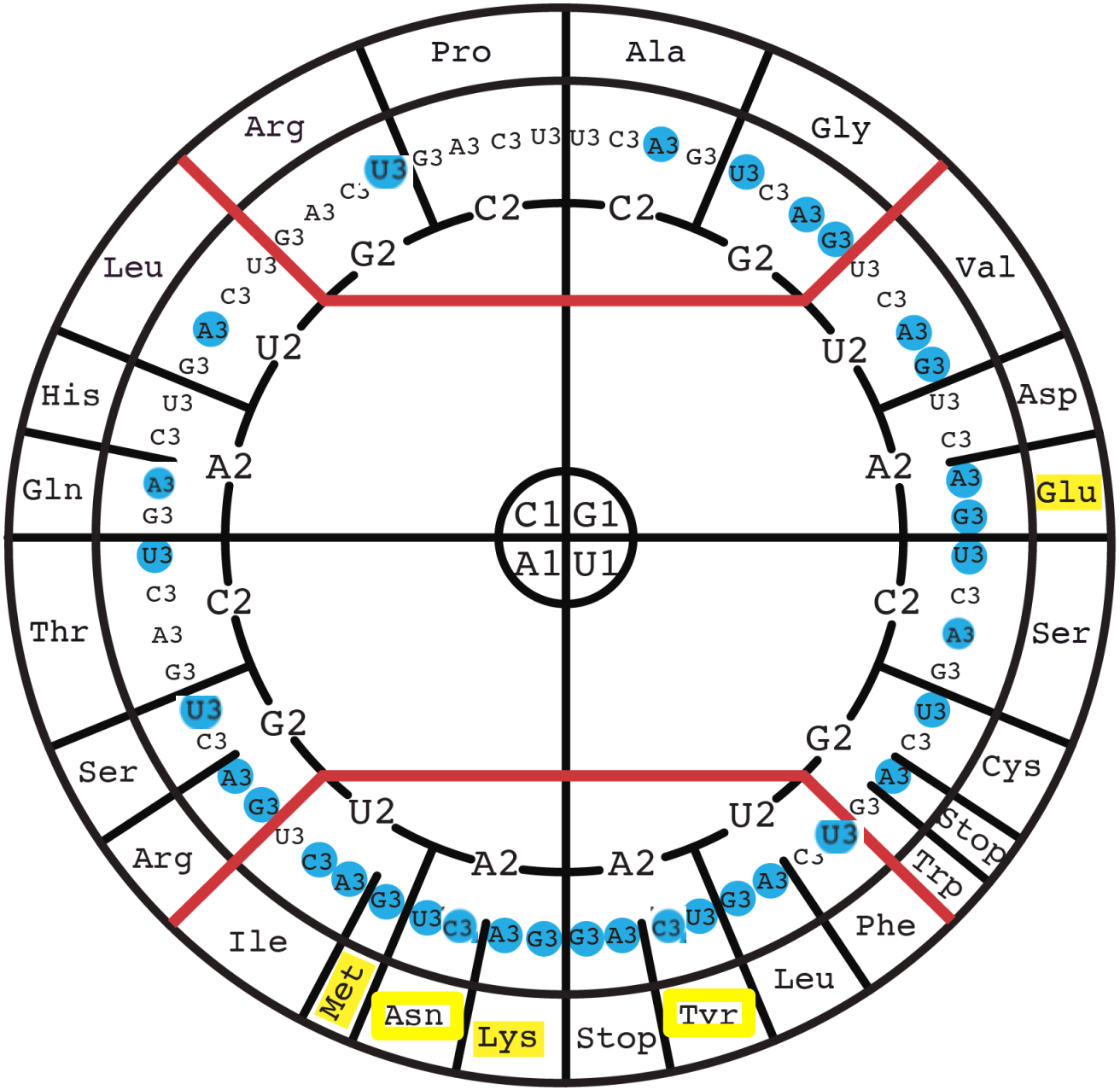


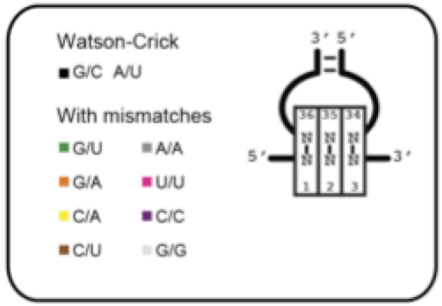
PKI




Functional selection
PKI
'Anticodon'
36NNN34
1NNN3
'Codon'
5'
PKIII
PKII
PKI
gfp
Gene individualization
1
2
Sorting
V
Laser
Active
Inactive
1. PCR amplification
2. Coupled transcription/translation
NGS
Sequencer
No selection
4096 combinations

BLUE = Absent codon/anticodon combination

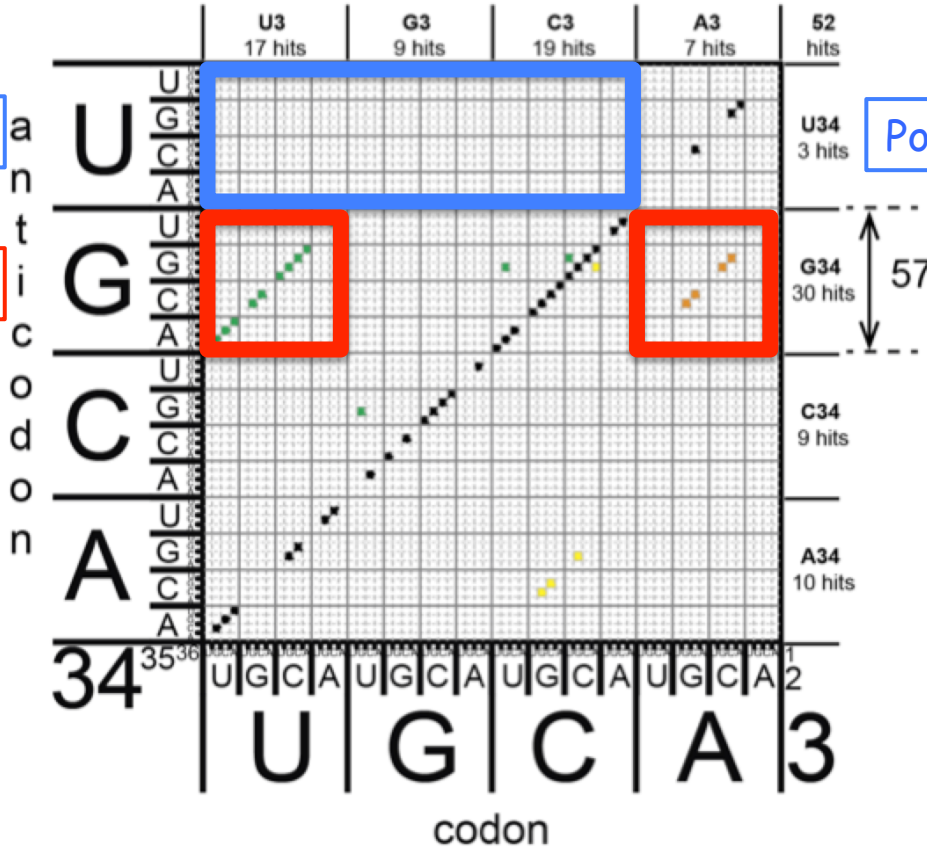




‘Stringent’ selection

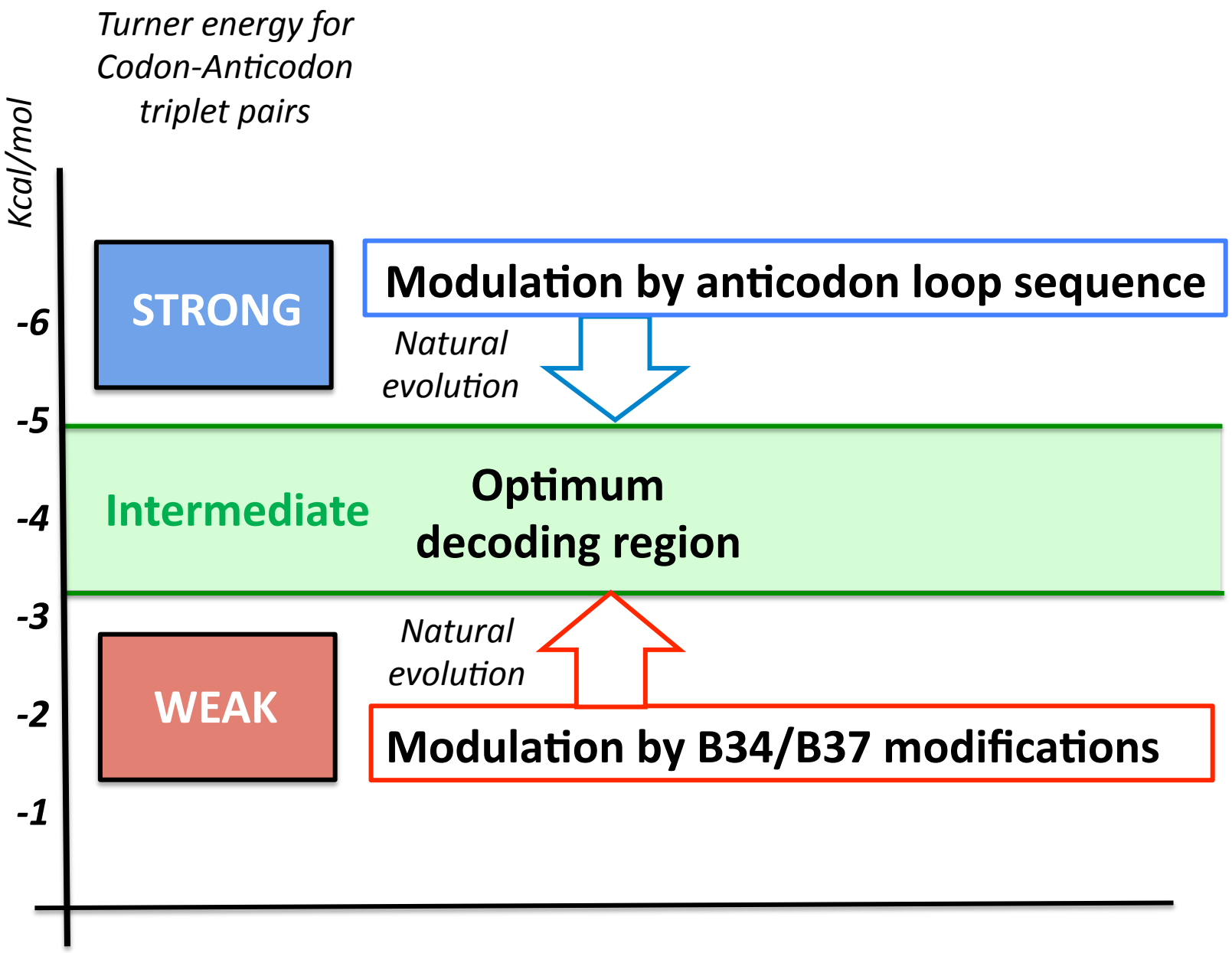
U34oG3 NOT present

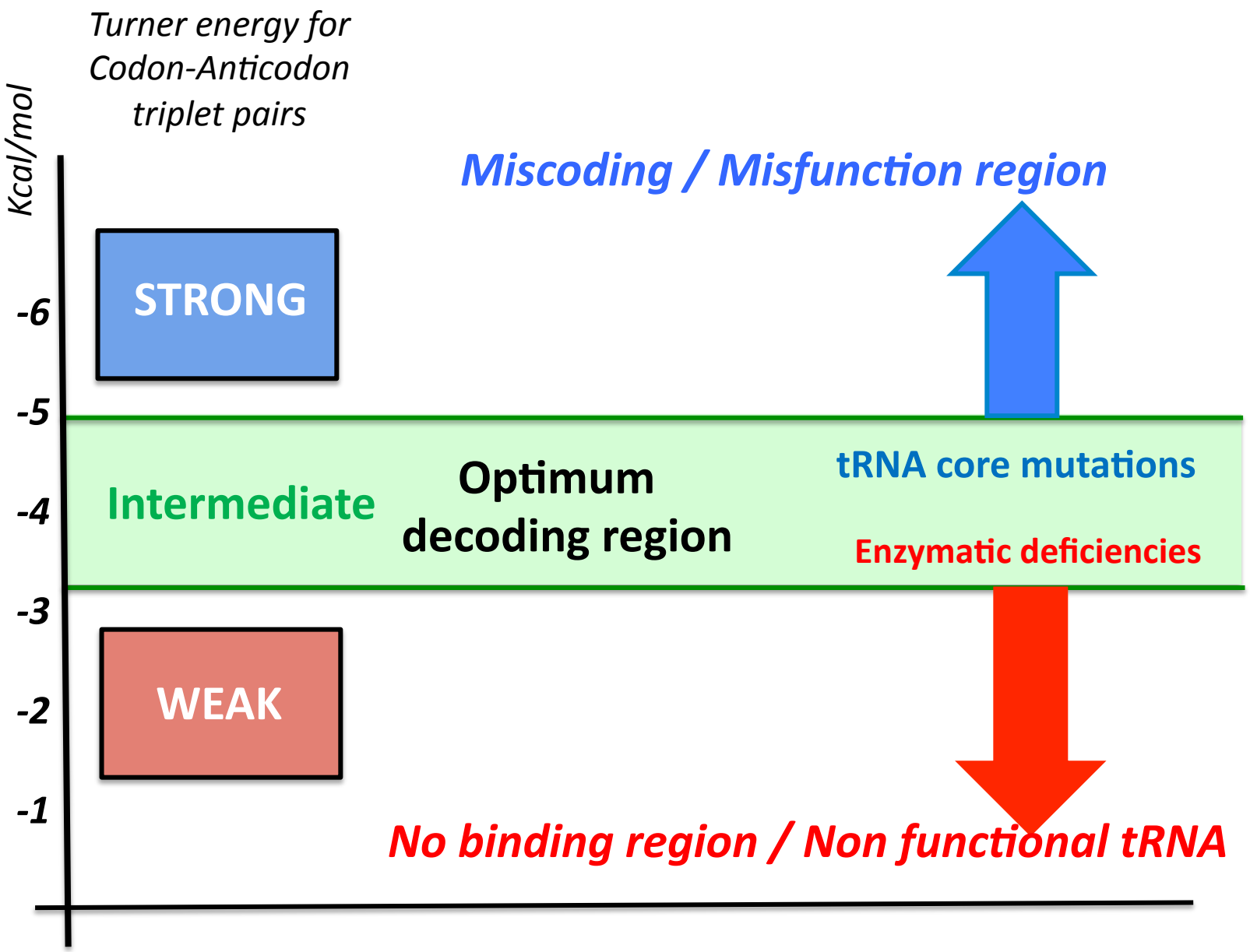
G34oU3 present



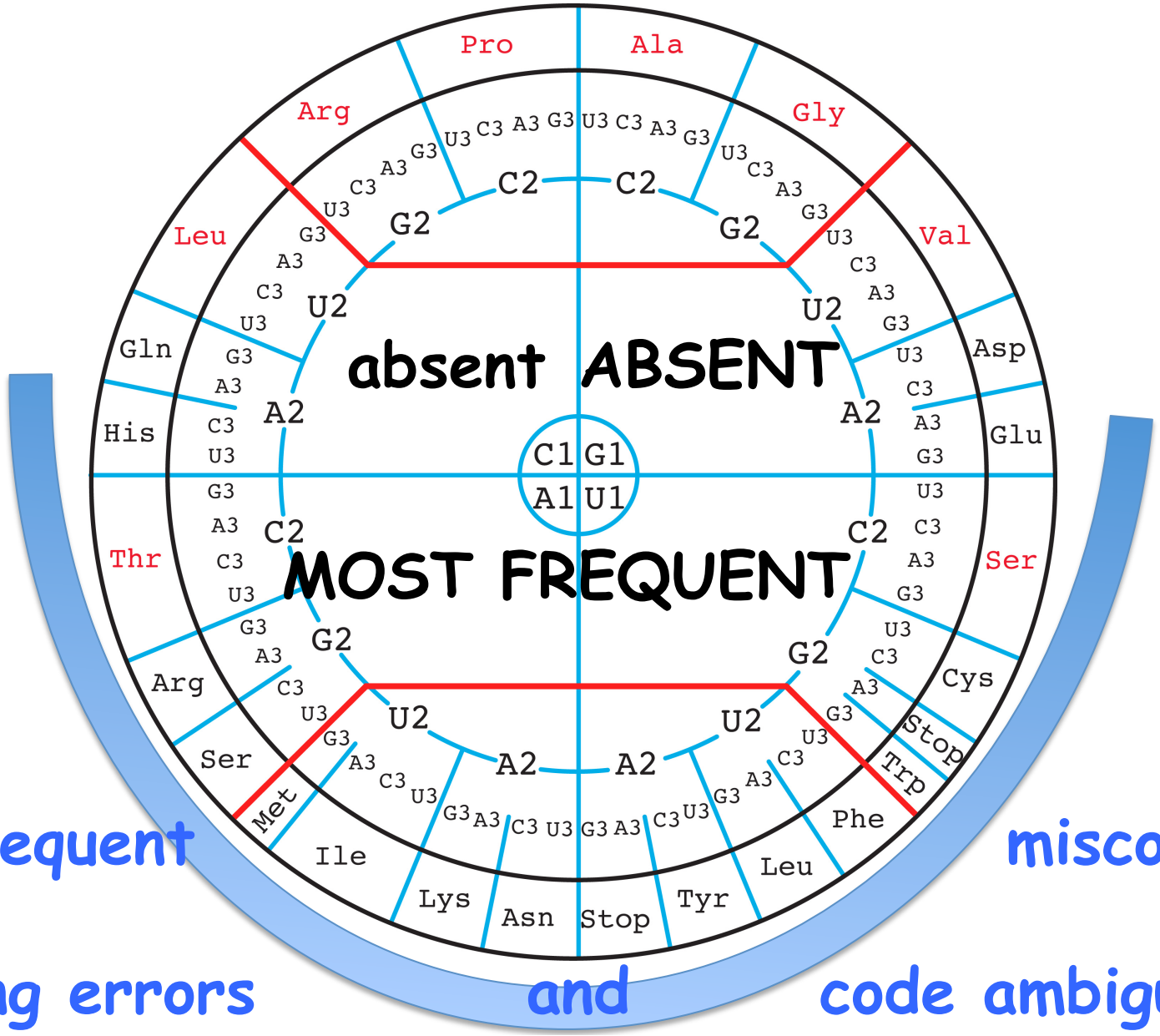
Poor decoding with U34

G34oA3 present


Turner energy for
Codon-Anticodon
triplet pairs
Kcal/mol
-6
-5
-4
-3
-2
-1
STRONG
Modulation by anticodon loop sequence
Natural
evolution
Intermediate
Optimum
decoding region
Natural
evolution
WEAK
Modulation by B34/B37 modifications

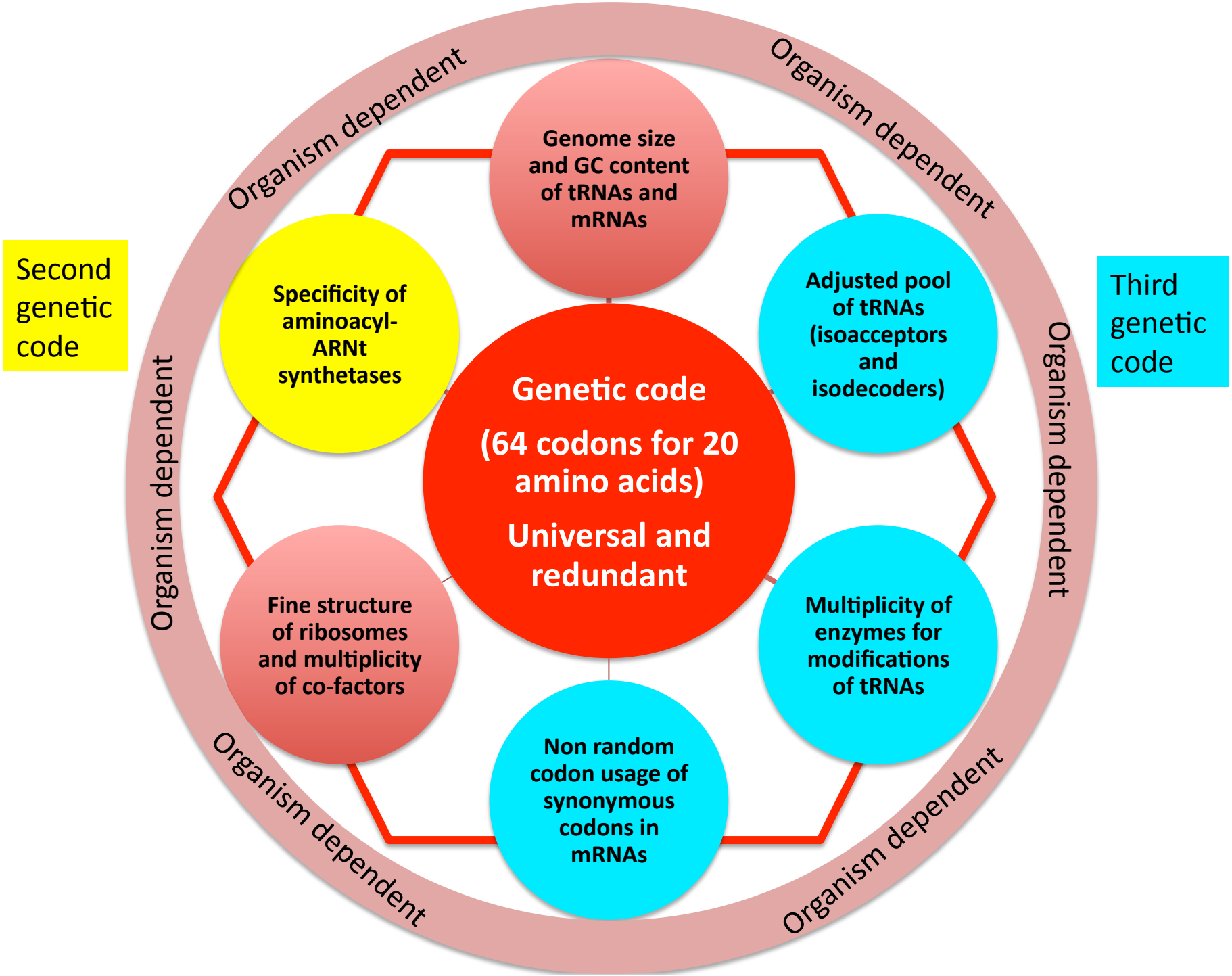

Turner energy for
Codon-Anticodon
triplet pairs
Kcal/mol
-6
-5
-4
-3
-2
-1
STRONG
Miscoding / Misfunction region
Intermediate
Optimum
decoding region
tRNA core mutations
Enzymatic deficiencies
WEAK
No binding region / Non functional tRNA

Code deviations, recoding, codon reassignments



Most frequent miscodings

(decoding errors and code ambiguities)


Second genetic code
Third genetic code
Organism dependent
Organism dependent
Organism dependent
Organism dependent
Organism dependent
Organism dependent
Genome size and GC content of tRNAs and mRNAs
Adjusted pool of tRNAs (isoacceptors and isodecoders)
Multiplicity of enzymes for modifications of tRNAs
Non random codon usage of synonymous codons in mRNAs
Fine structure of ribosomes and multiplicity of co-factors
Specificity of aminoacyl-ARNt synthetases
Genetic code
(64 codons for 20 amino acids)
Universal and redundant

UPR 9002 du CNRS,
Architecture et Réactivité de l'ARN,
Institut de Biologie Moléculaire et Cellulaire,
Université de Strasbourg

Henri Grosjean (Gif)

Gulnara & Marat Yusupov (IGBMC)

N. Demeshkina, L. Jenner, A. Rozov (IGBMC)

M.Ryckelynck, C. Rick, K. Pernod, F. Martin (IBMC)