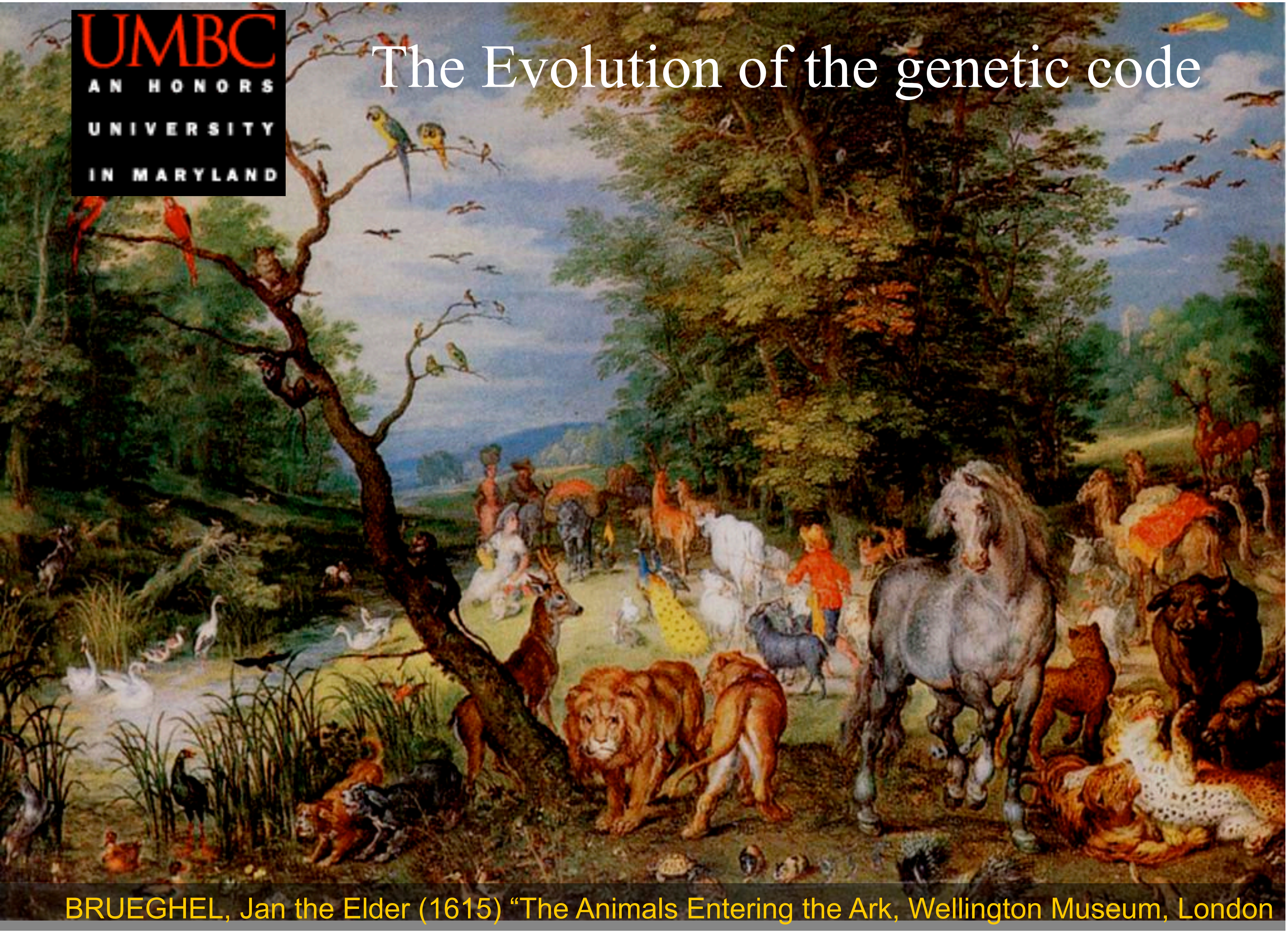
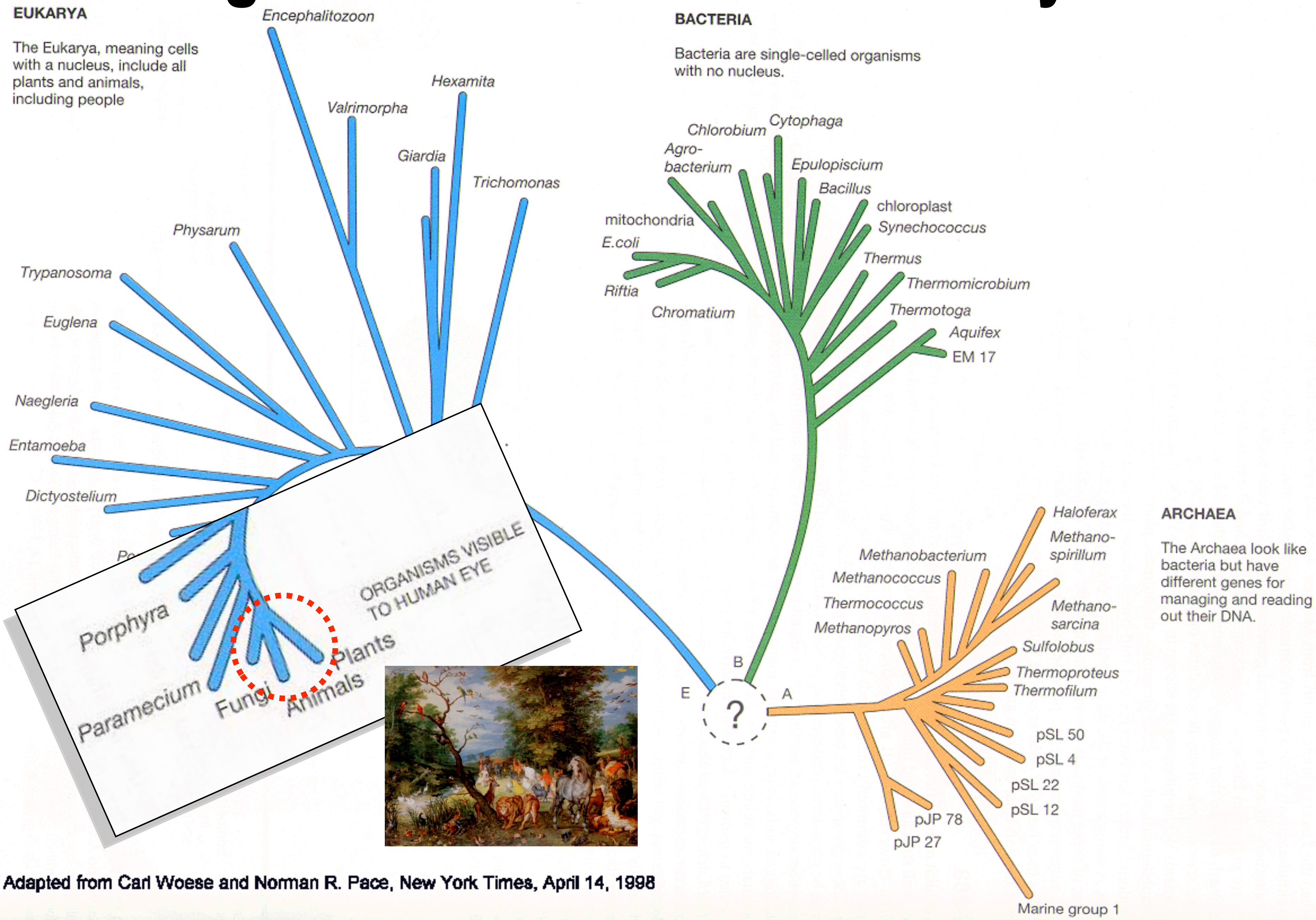
UMBC
AN HONORS UNIVERSITY IN MARYLAND

The Evolution of the genetic code



BRUEGHEL, Jan the Elder (1615) “The Animals Entering the Ark, Wellington Museum, London

◆ A genetic tree of life's diversity...



EUKARYA

The Eukarya, meaning cells with a nucleus, include all plants and animals, including people

BACTERIA

Bacteria are single-celled organisms with no nucleus.

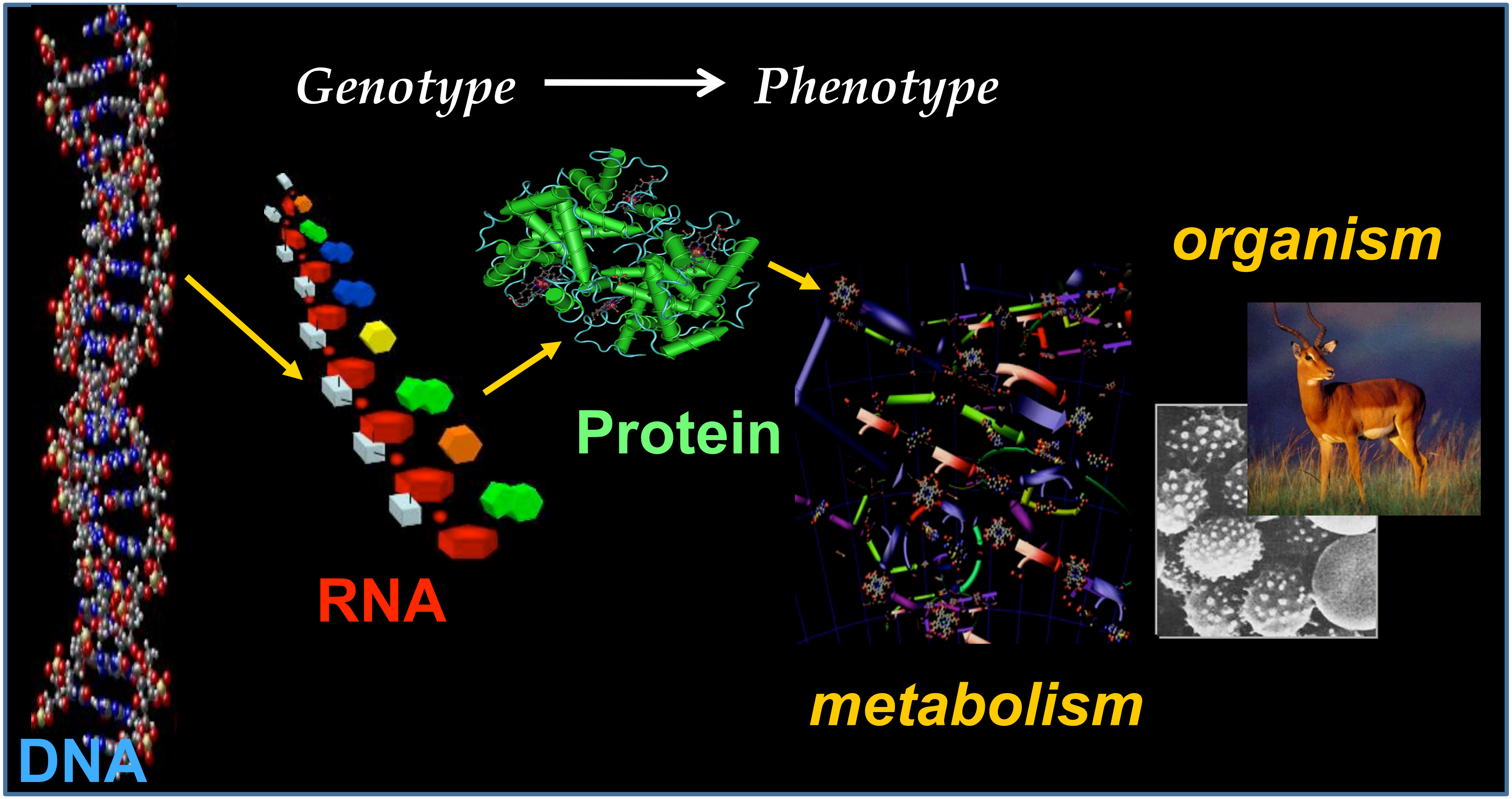
ARCHAEA

The Archaea look like bacteria but have different genes for managing and reading out their DNA.

Adapted from Carl Woese and Norman R. Pace, New York Times, April 14, 1998

Unifying principles of biology

In all organisms, all of the time, biological information is being transformed from genotype (information) into phenotype (living, breathing organisms)



Genotype ⟶ *Phenotype*

DNA

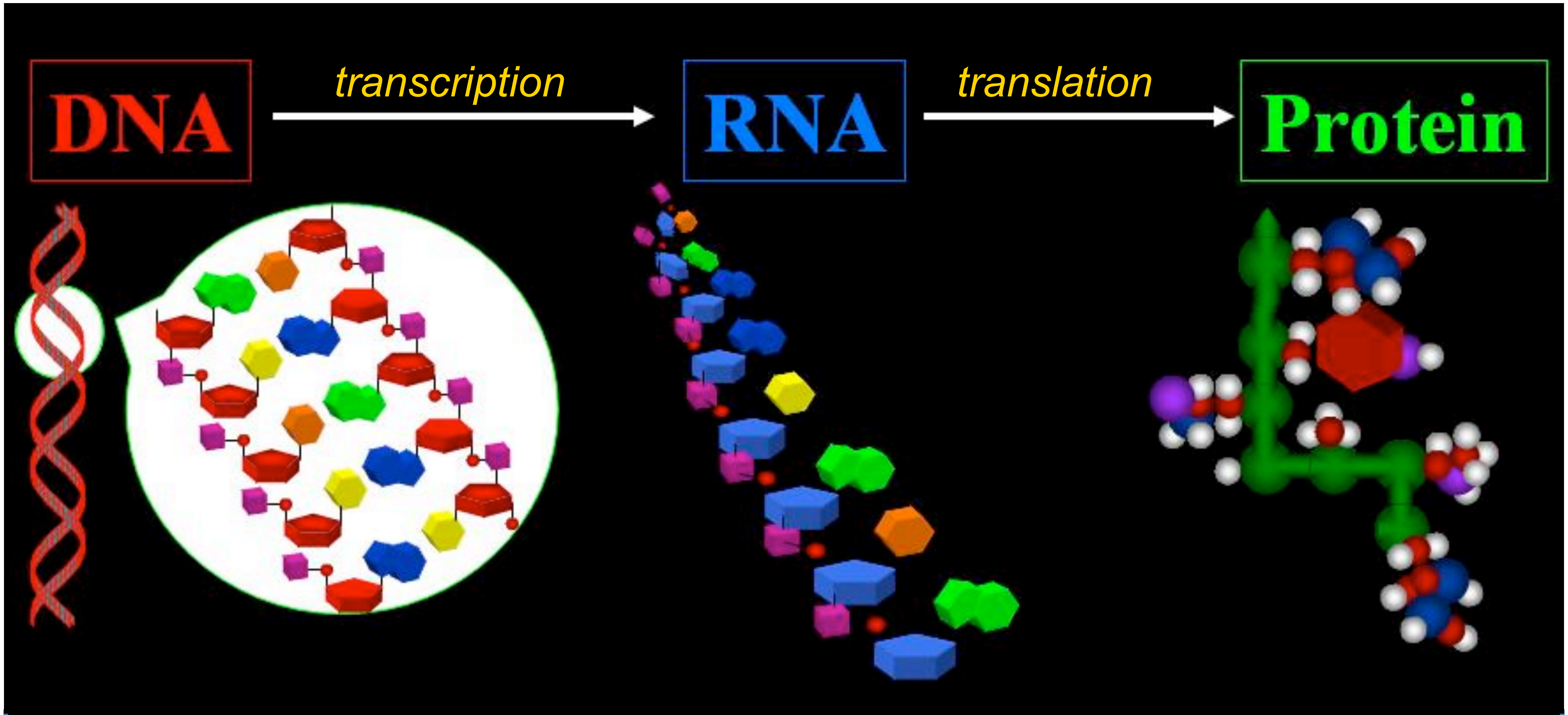
RNA

Protein

metabolism

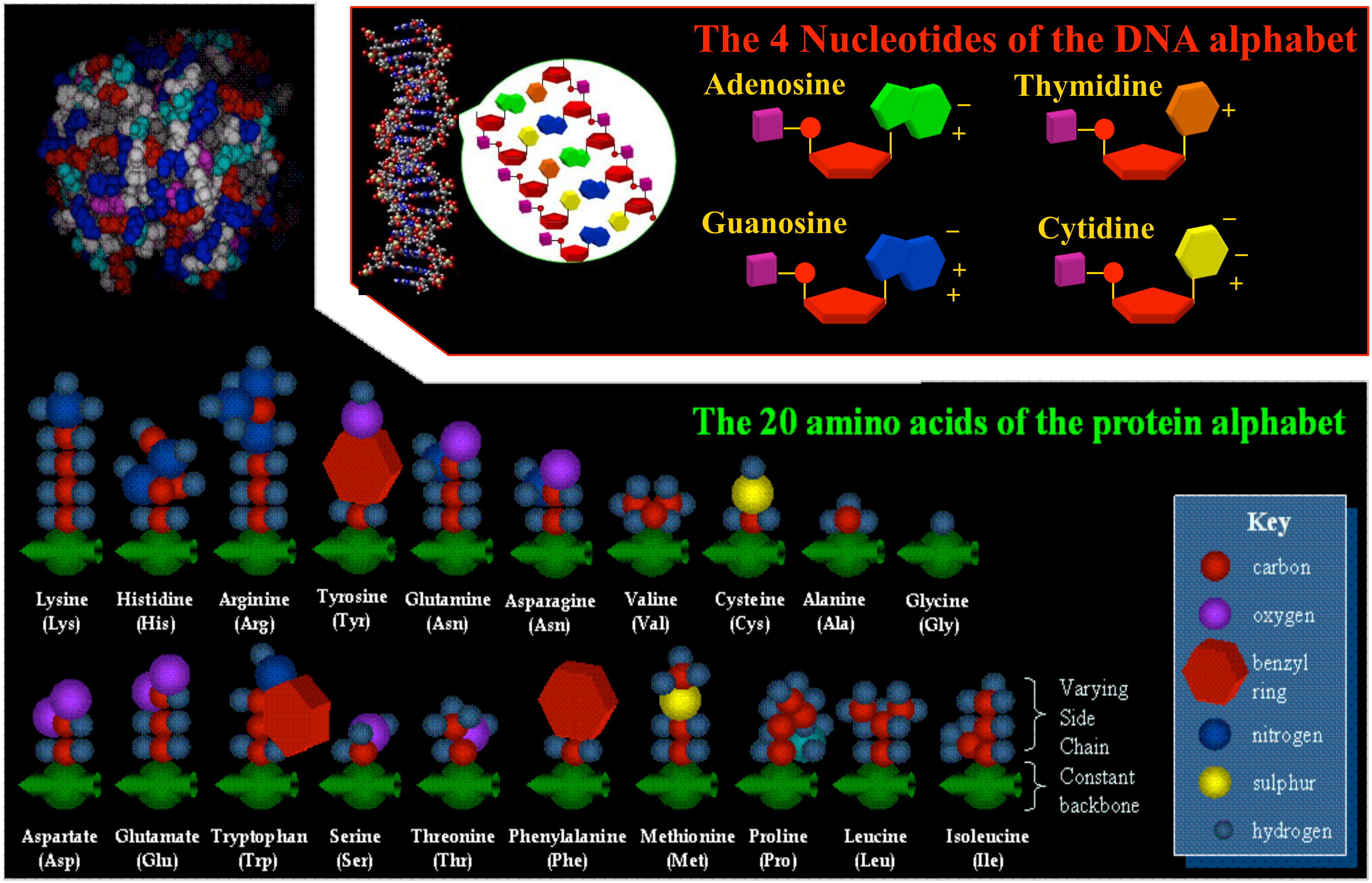
organism

The "central dogma" of molecular biology



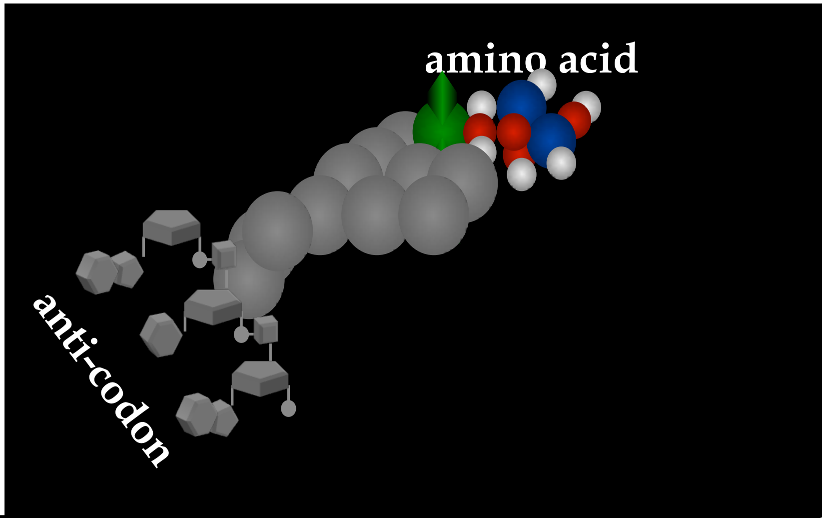
- the double helix of **DNA** is <u>transcribed</u> into (chemical sister language) **RNA**, and the resulting messenger RNA's are then <u>translated</u> into **Protein**

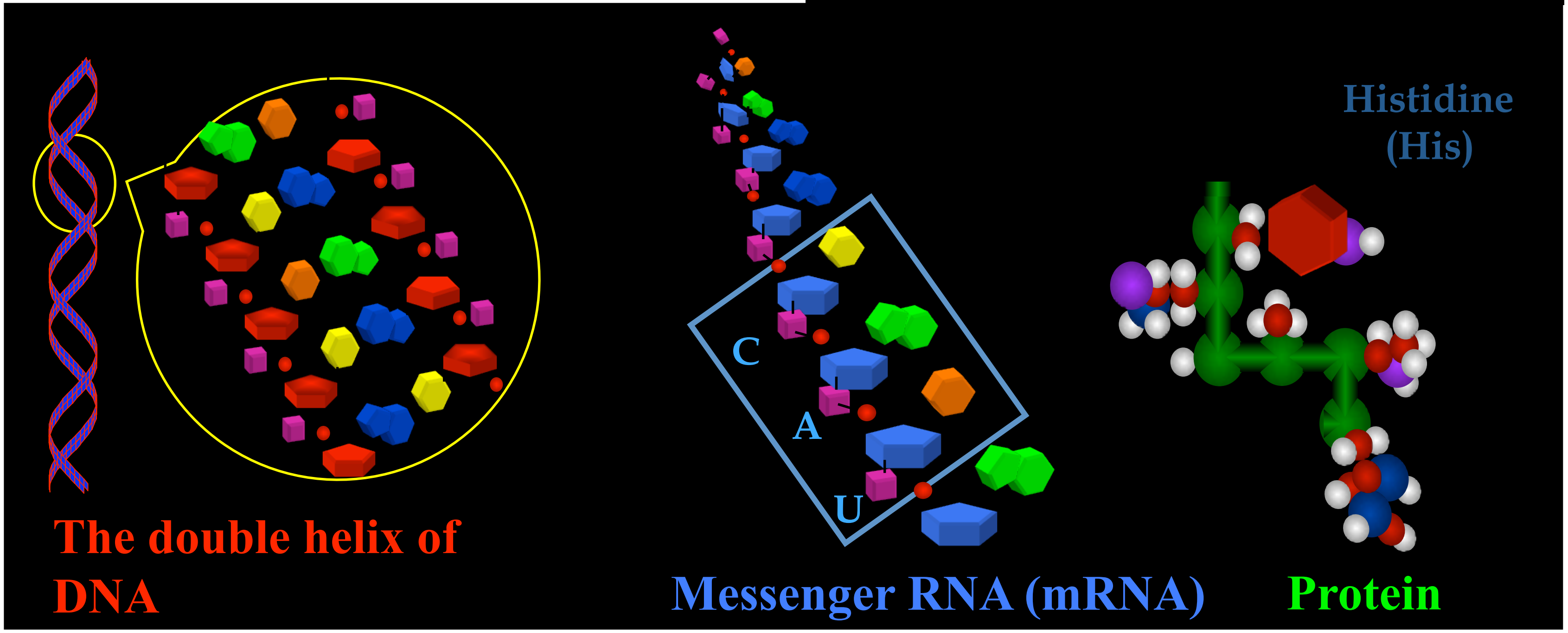
The "central dogma" of molecular biology – involves two *utterly different* chemical languages



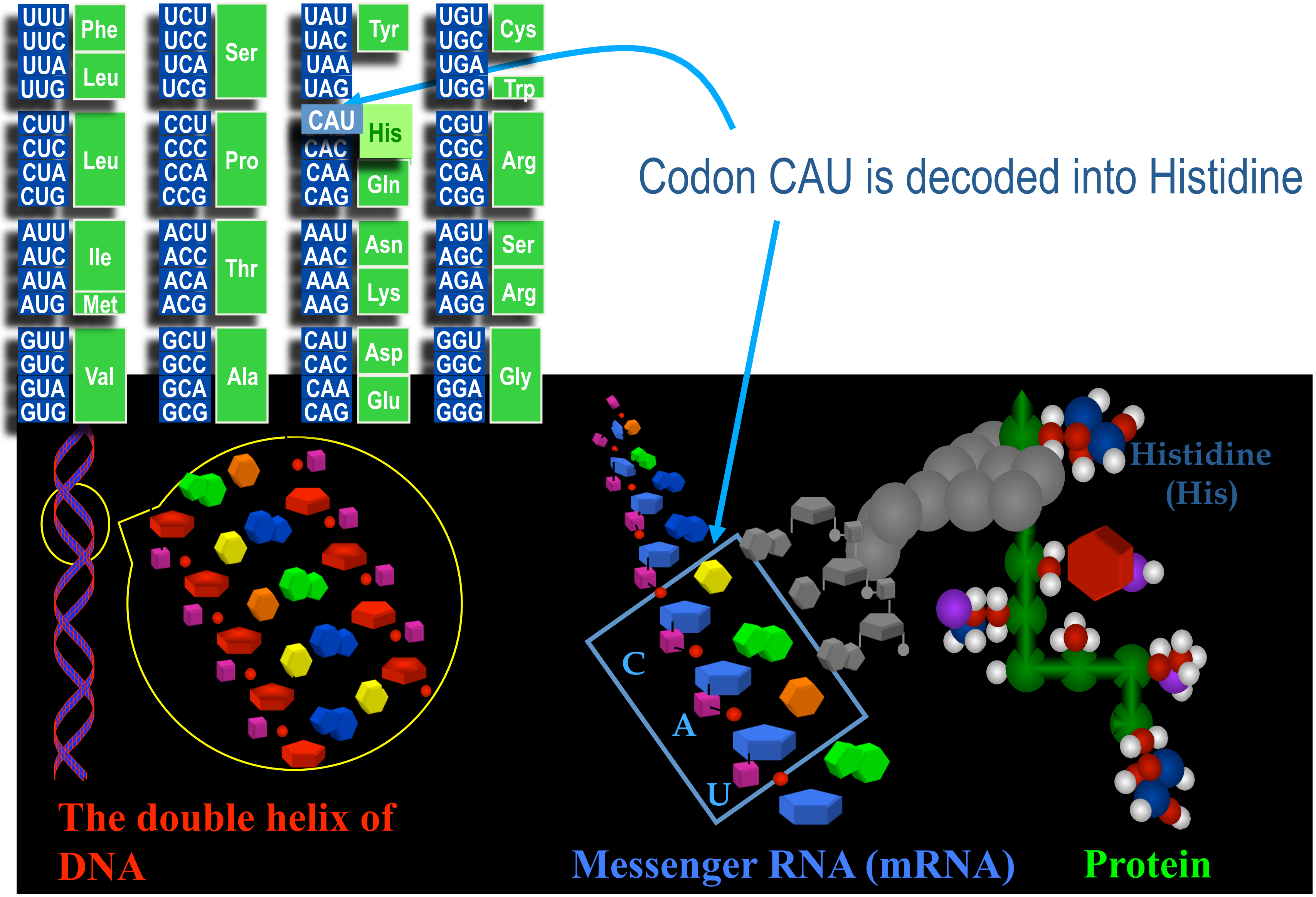
tRNA’s bridge between these two languages

Nucleotides of mRNA are read in triplets by tRNA’s (adaptor molecules) that translate this *codon* into a corresponding amino acid





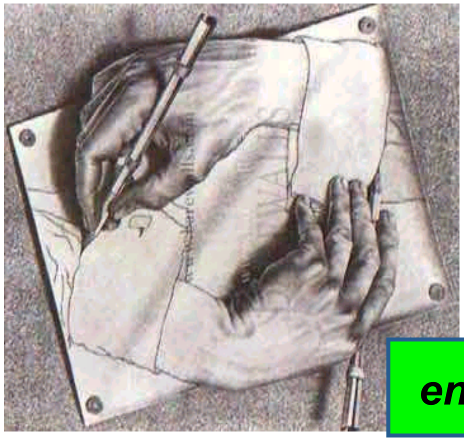
The Genetic Code is the set of rules describing what amino acid each possible codon is translated into

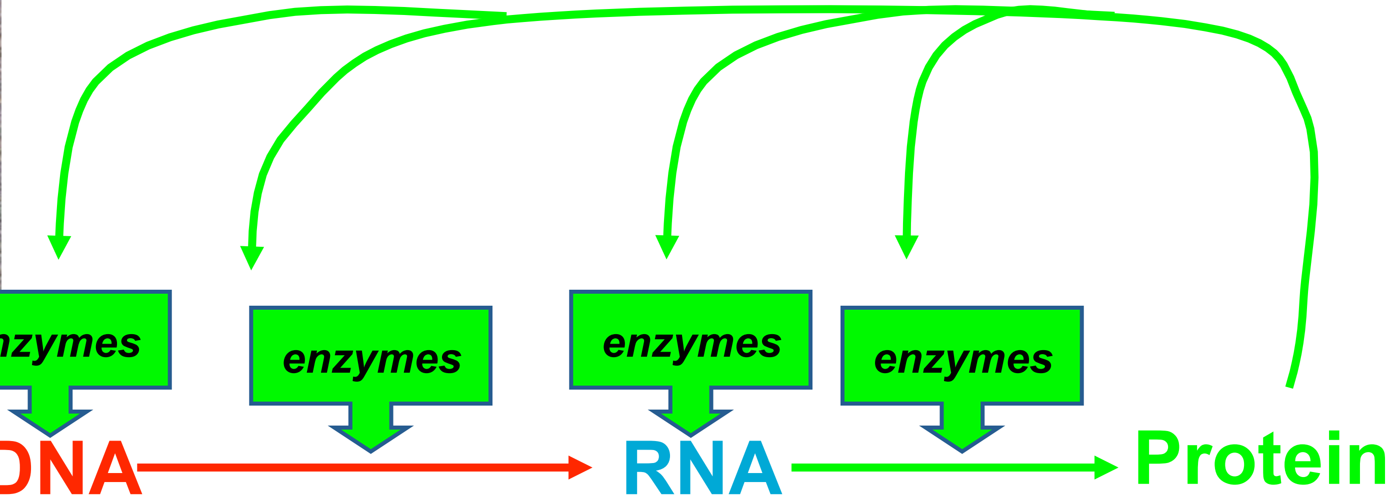


UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA		UGA	
UUG		UCG		UAG		UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lys	AGA	Arg
AUG	Met	ACG		AAG		AGG	
GUU	Val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

Codon CAU is decoded into Histidine

The central paradox of early evolution…





enzymes

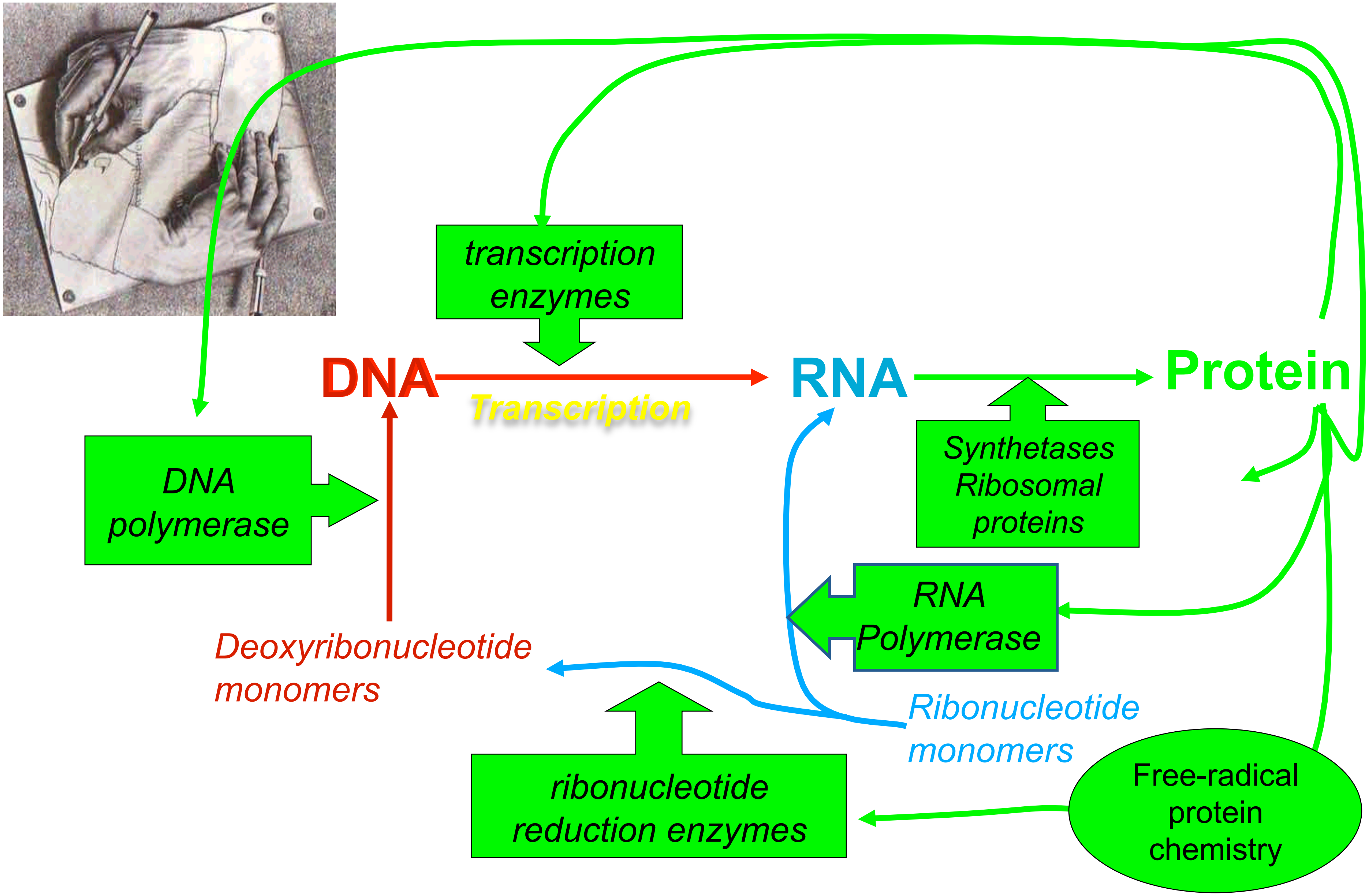
enzymes

enzymes

enzymes

DNA → **RNA** → **Protein**

The central paradox of early evolution…



DNA

transcription enzymes

Transcription

RNA

Protein

DNA polymerase

Synthetases Ribosomal proteins

RNA Polymerase

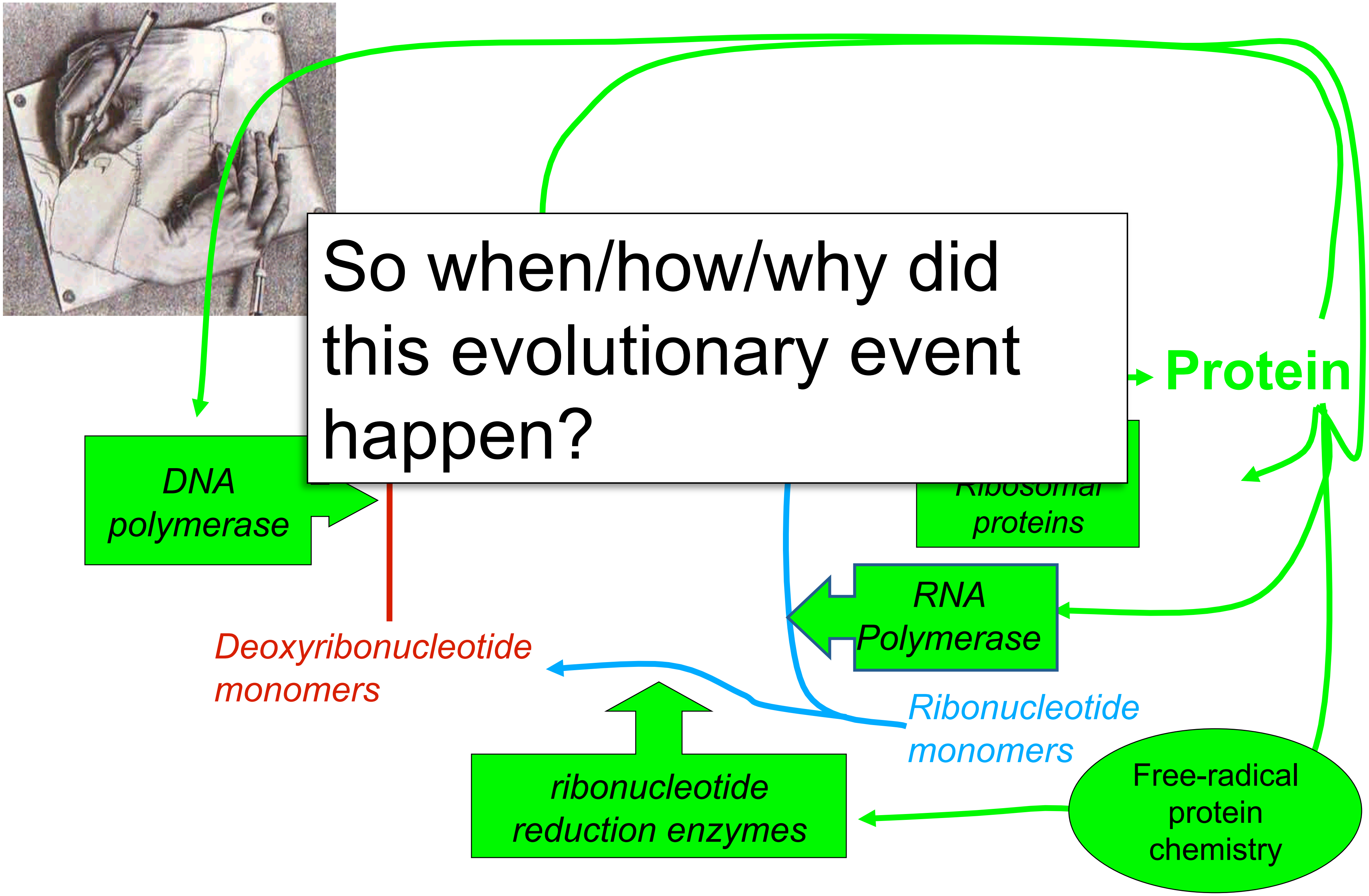
Deoxyribonucleotide monomers

Ribonucleotide monomers

ribonucleotide reduction enzymes

Free-radical protein chemistry

The central paradox of early evolution…

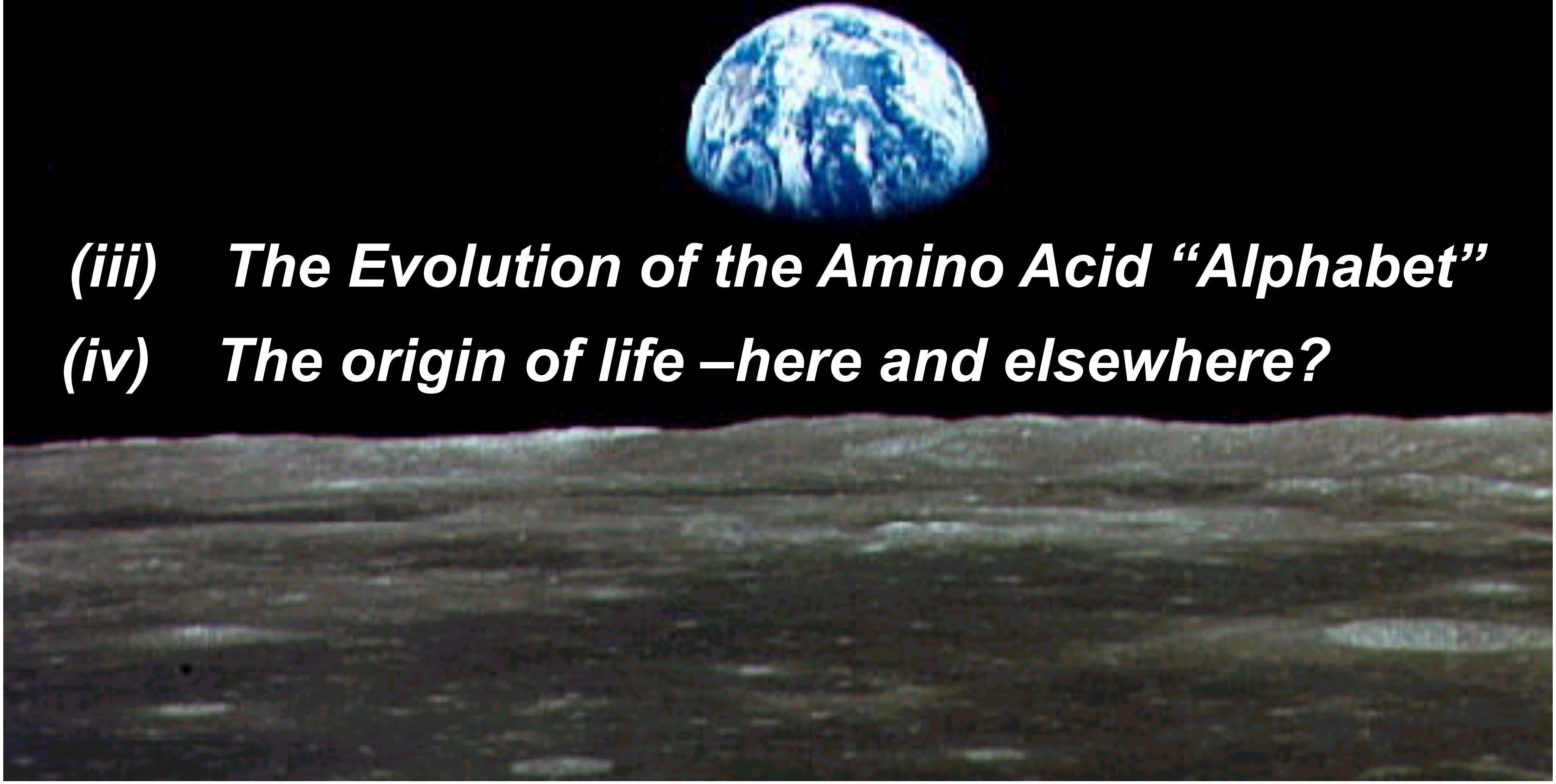


So when/how/why did this evolutionary event happen?

4 lectures…

(i) A History of Life on earth

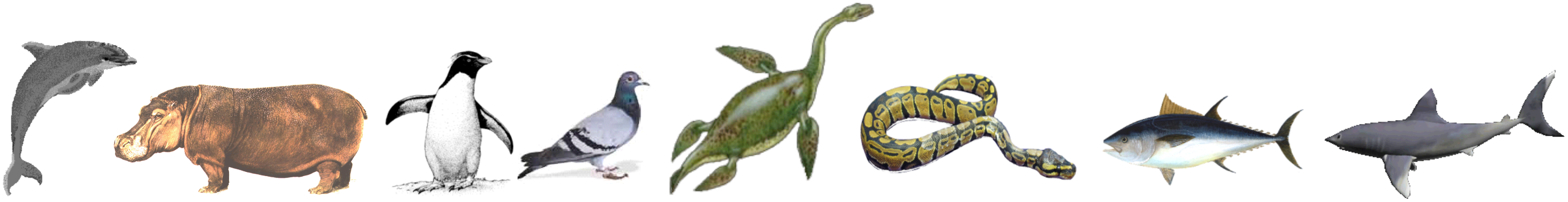
(ii) The Evolution of the Genetic Code



(iii) The Evolution of the Amino Acid “Alphabet”

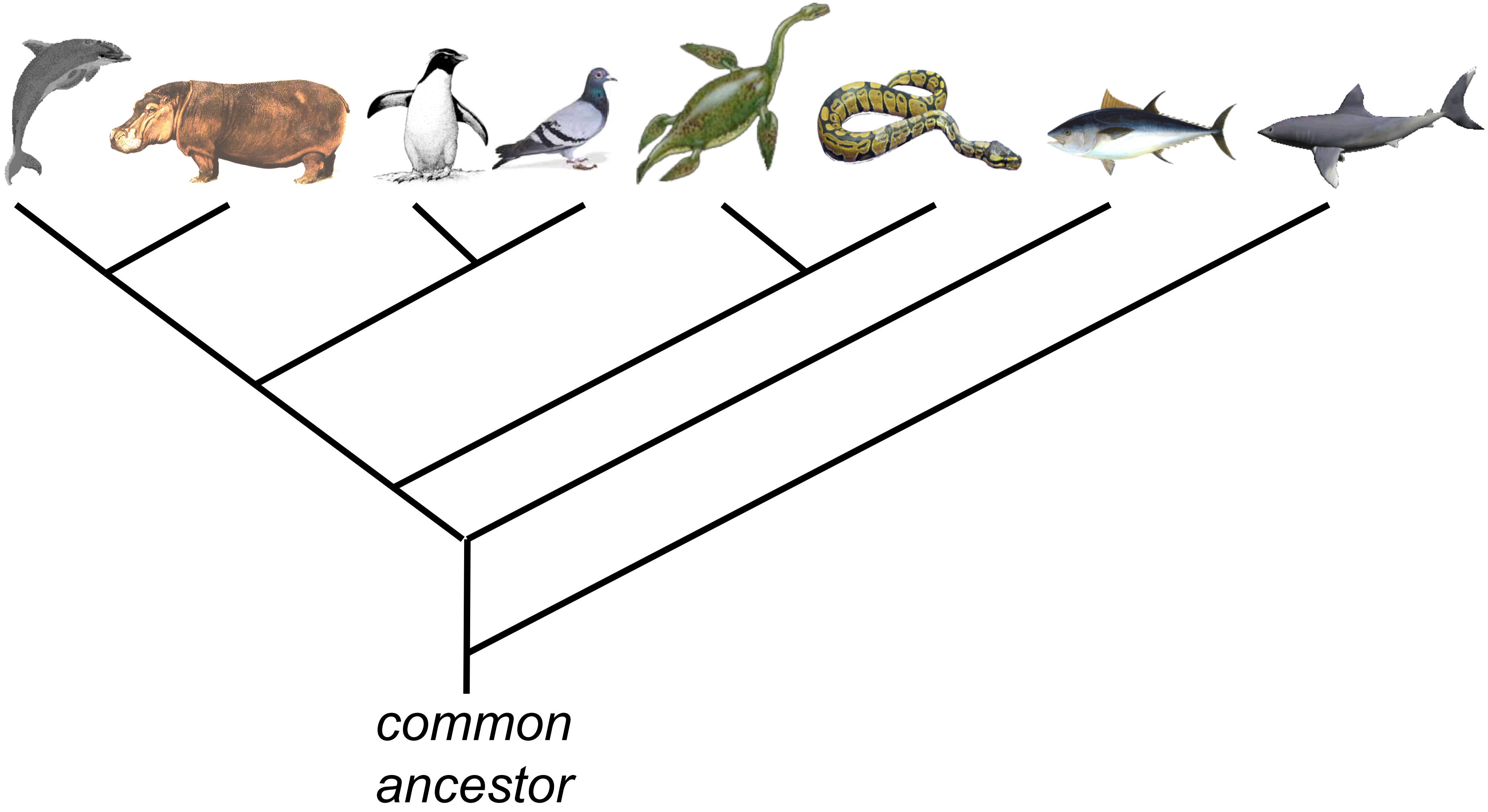
(iv) The origin of life –here and elsewhere?

“normal” methods for reconstructing evolution



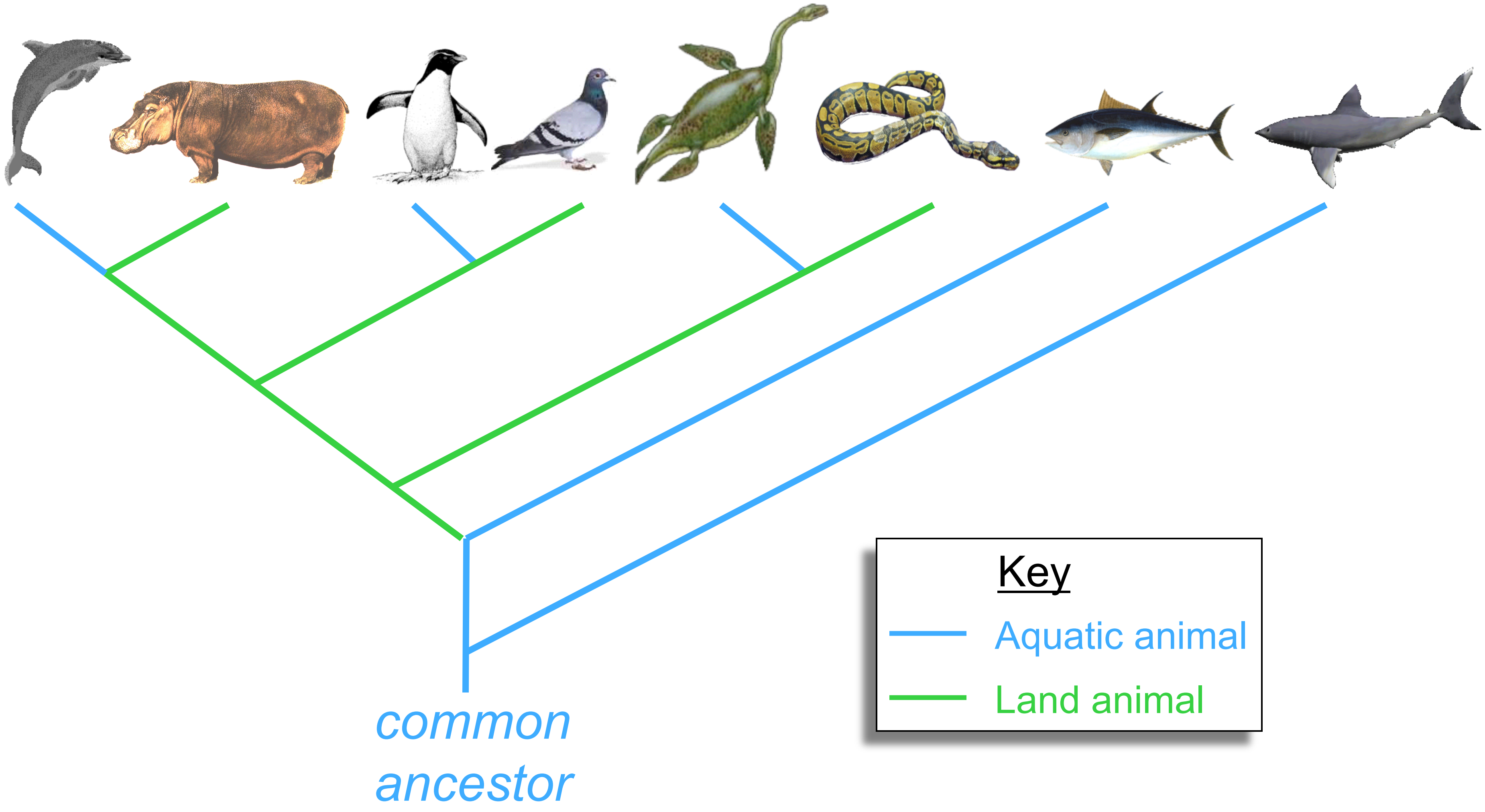
1. Collect and organize as much information as possible about their characteristics

“normal” methods for reconstructing evolution



1. Collect and organize as much information as possible about their characteristics
2. Construct an evolutionary tree, assuming most similarities = closest relatives

“normal” methods for reconstructing evolution



common ancestor

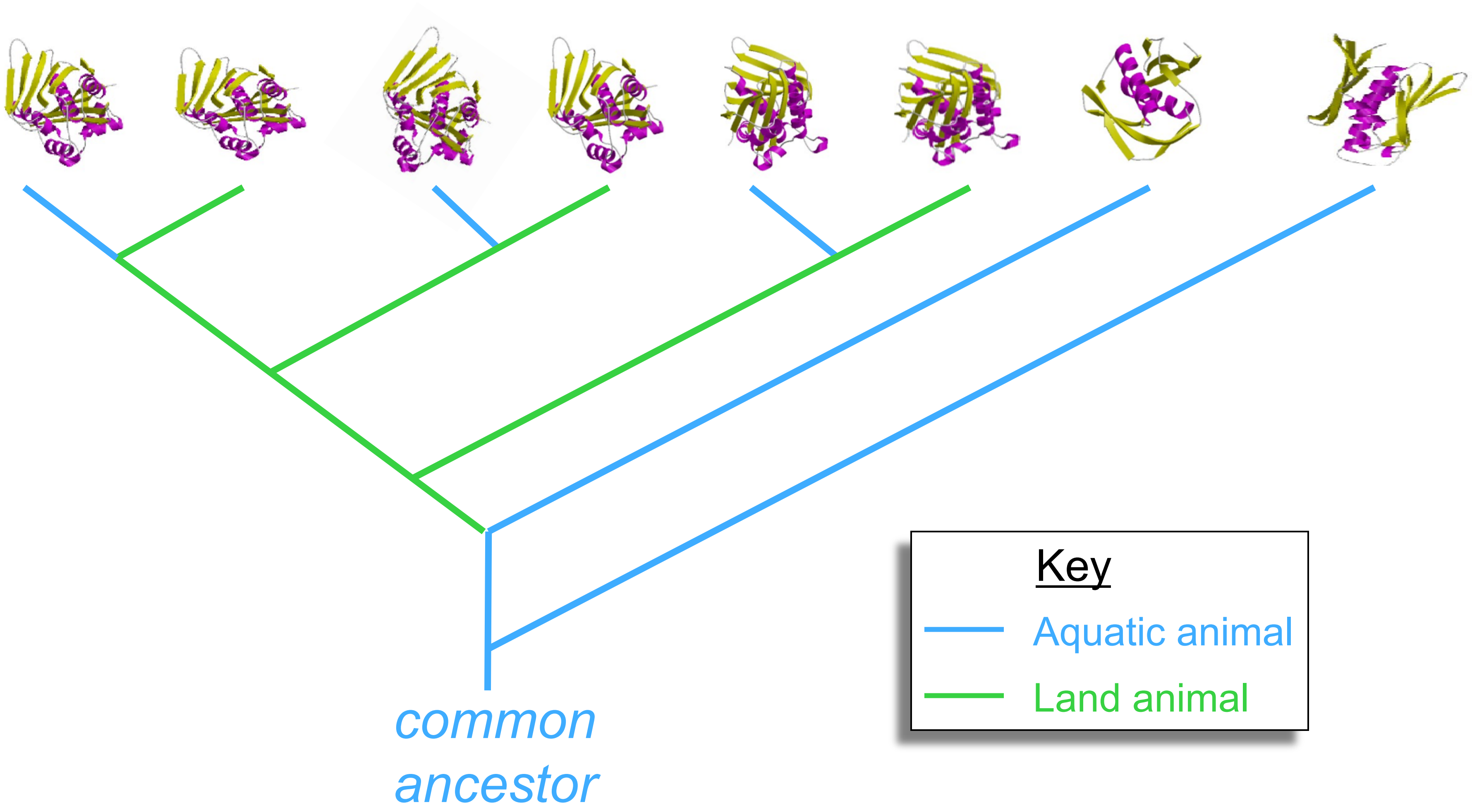
Key

— Aquatic animal

— Land animal

1. Collect and organize as much information as possible about their characteristics
2. Construct an evolutionary tree, assuming most similarities = closest relatives
3. Infer the order in which characteristics evolved

(It is increasingly common to use molecular data)



1. Collect and organize as much information as possible about their characteristics
2. Construct an evolutionary tree, assuming most similarities = closest relatives
3. Infer the order in which characteristics evolved

(It is increasingly common to use molecular data)

Whale

vhltgeeks gltalwakvn veeiggealg rllvvypwtq rffehfgdls tadavmknpk vkkhgqkvla

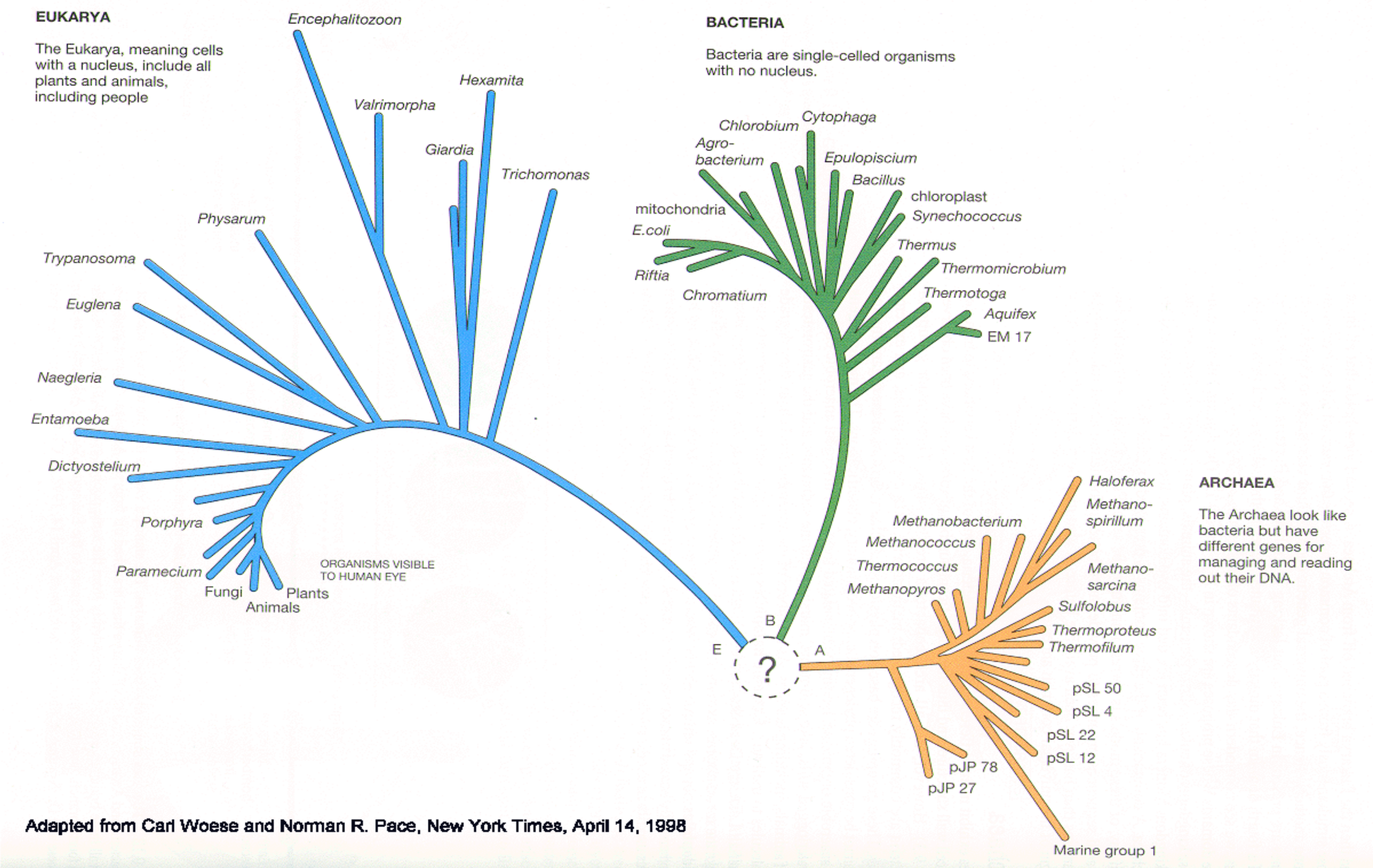
Human β-hemoglobin

mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk vkahgkkvlg

Rat β-hemoglobin

mvhltdaeka avnglwgkvn pddvggealg rllvvypwtq ryfdsfgdls sasaimgnpk vkahgkkvin

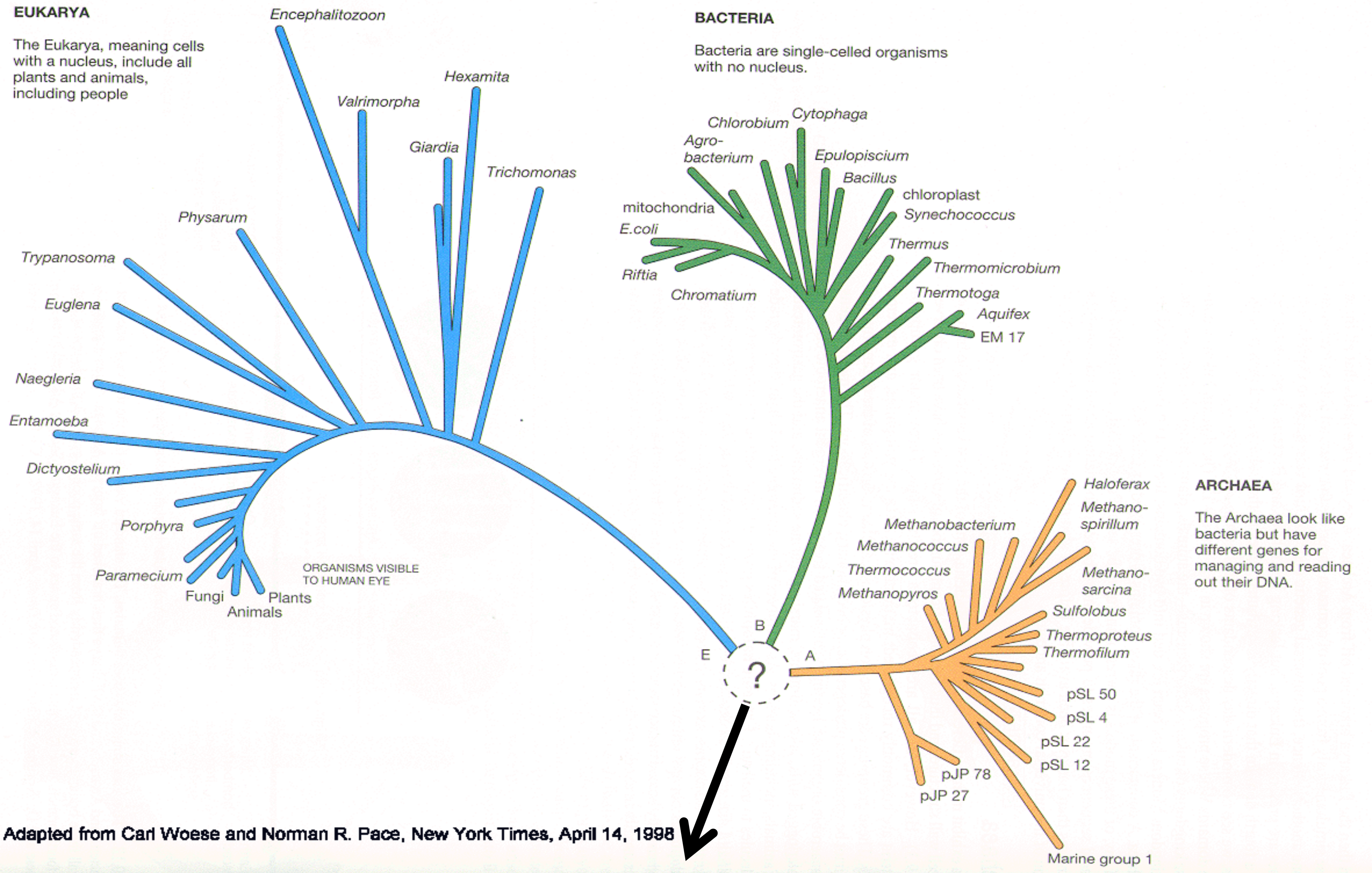
(It is increasingly common to use molecular data)



Adapted from Carl Woese and Norman R. Pace, New York Times, April 14, 1998

e.g. all species have their own ribosomes, so by comparing the DNA sequences of these ribosomes, we may construct a “tree of life”

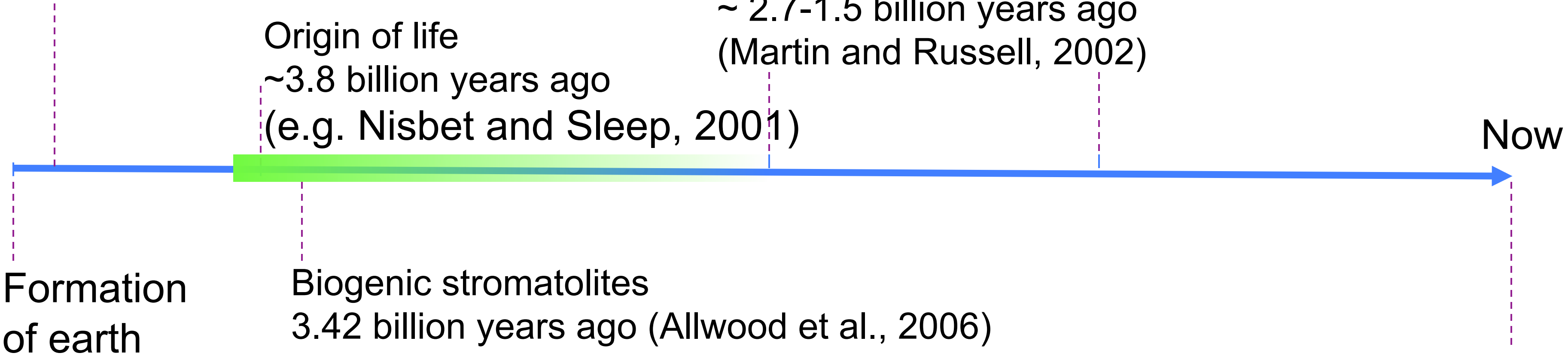
But before the last universal common ancestor (LUCA)…



Adapted from Carl Woese and Norman R. Pace, New York Times, April 14, 1998

…most fundamental biochemistry was already fixed – including the Central Dogma

When did the standard genetic code evolve?



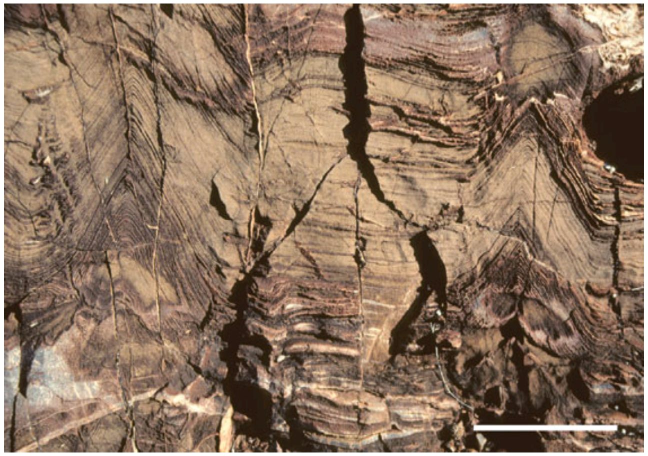
Origin of life
~3.8 billion years ago
(e.g. Nisbet and Sleep, 2001)

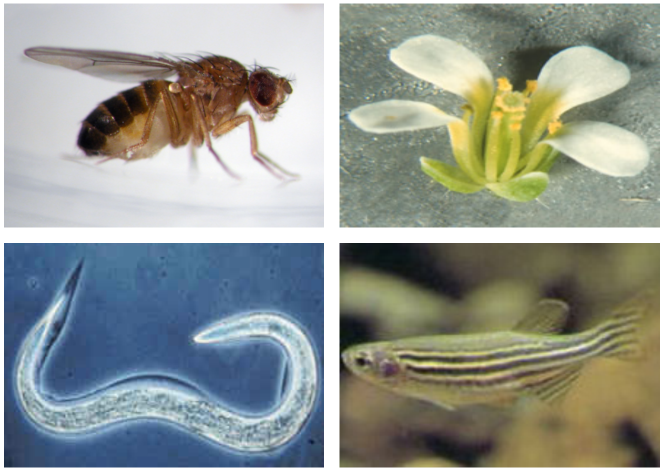
Divergence of eukaryotes
~ 2.7-1.5 billion years ago
(Martin and Russell, 2002)

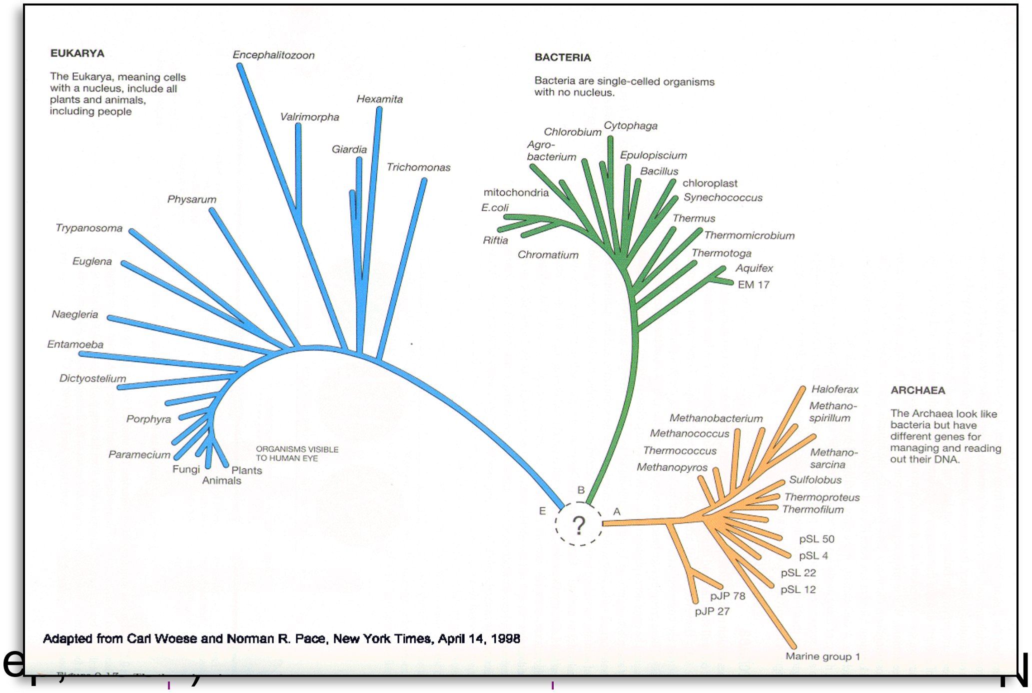
Now

Formation of earth
~4.6 billion years ago

Biogenic stromatolites
3.42 billion years ago (Allwood et al., 2006)







EUKARYA

The Eukarya, meaning cells with a nucleus, include all plants and animals, including people

BACTERIA

Bacteria are single-celled organisms with no nucleus.

ARCHAEA

The Archaea look like bacteria but have different genes for managing and reading out their DNA.

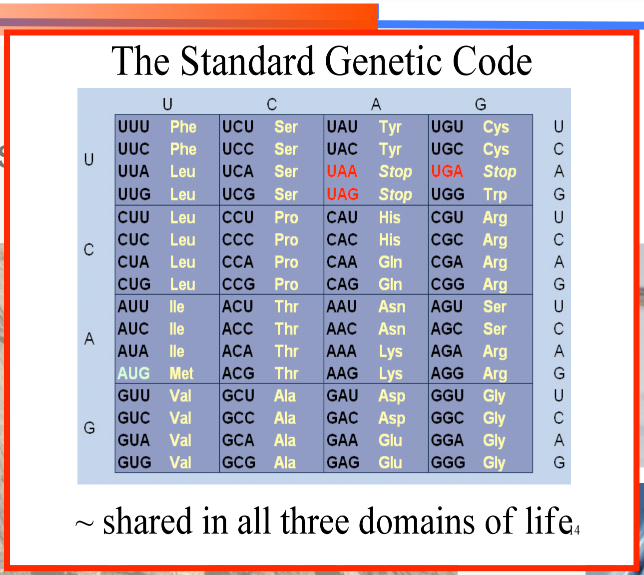
Adapted from Carl Woese and Norman R. Pace, New York Times, April 14, 1998

Origin of life
~3.8 billion years ago
(e.g. Nisbet and Slee

Now

Formation of earth
~4.6 billion years ago

Biogenic s
3.42 billio

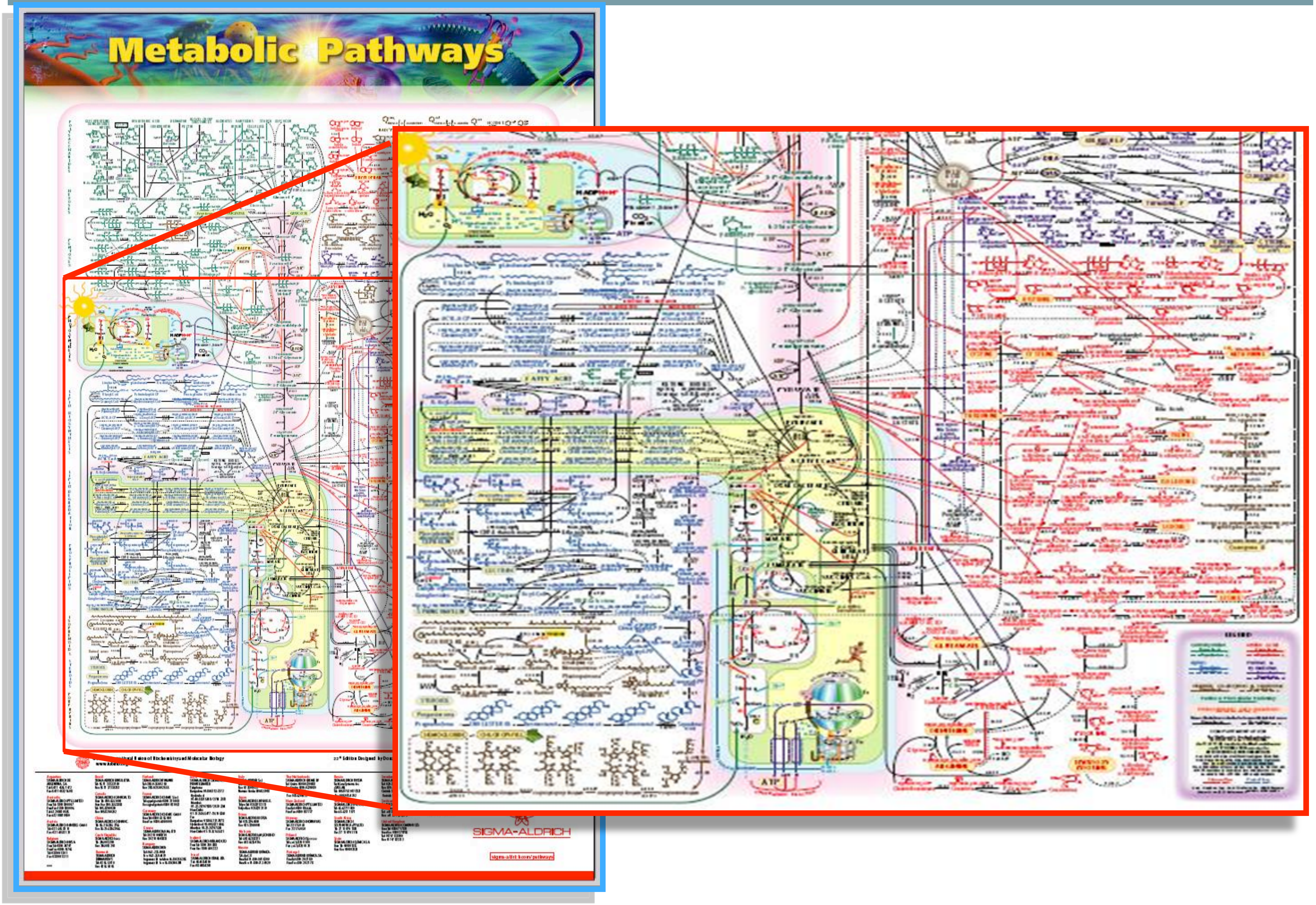


The Standard Genetic Code

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

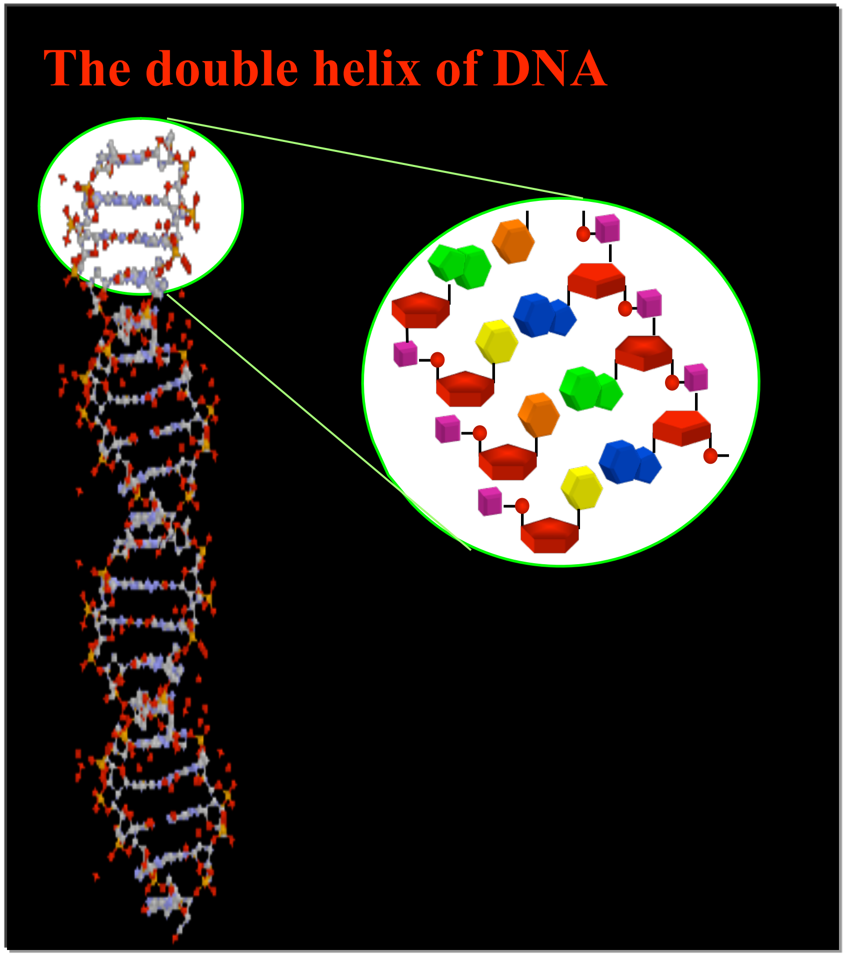
~ shared in all three domains of life

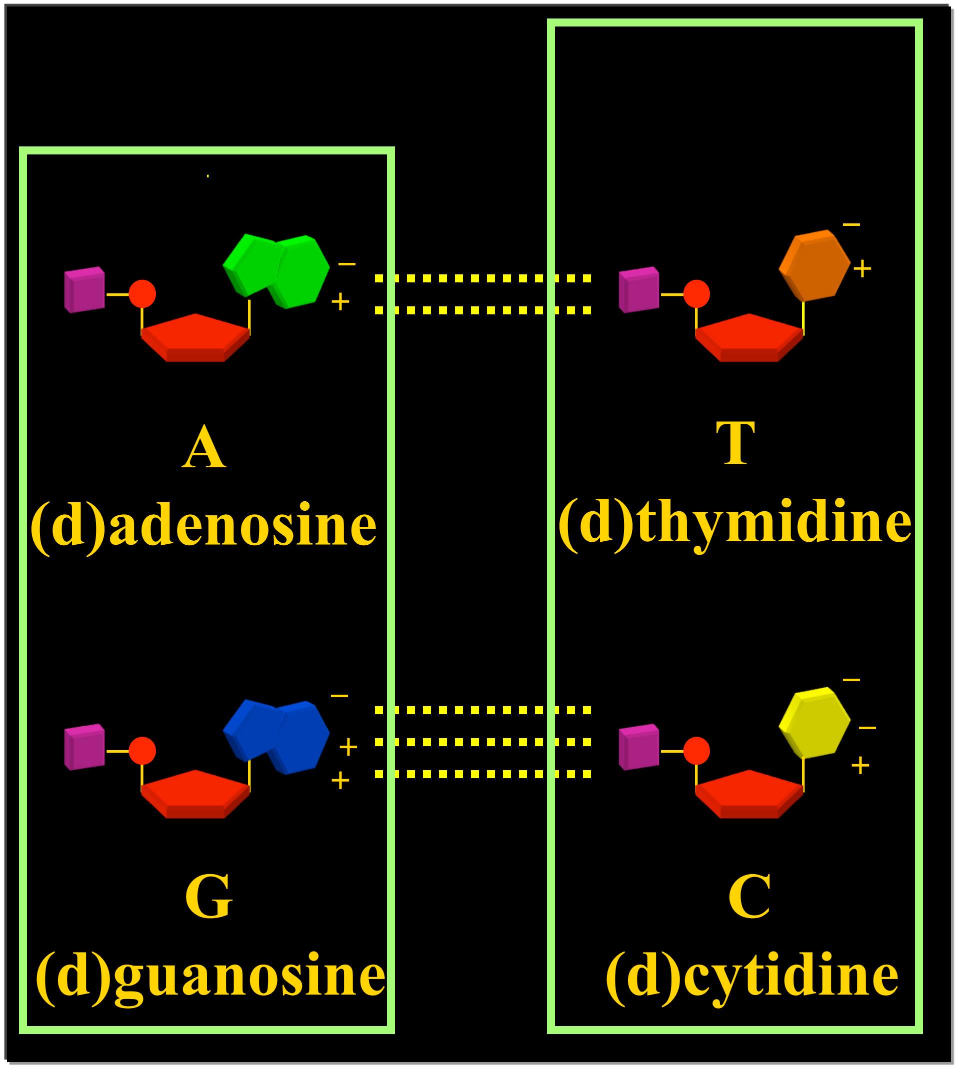
Theme I: Modern metabolism carries a footprint of ancient evolutionary history



(deoxy)Nucleotides – the building blocks of DNA

There are 4 ‘letters’ (nucleotides)

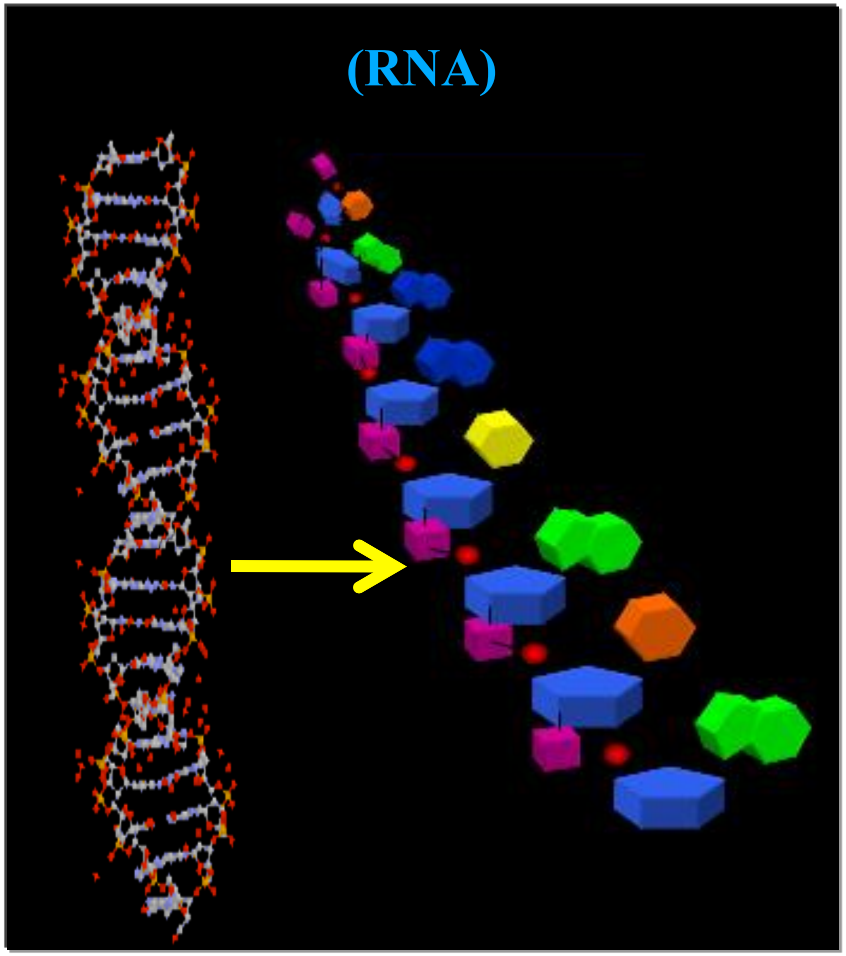

The double helix of DNA

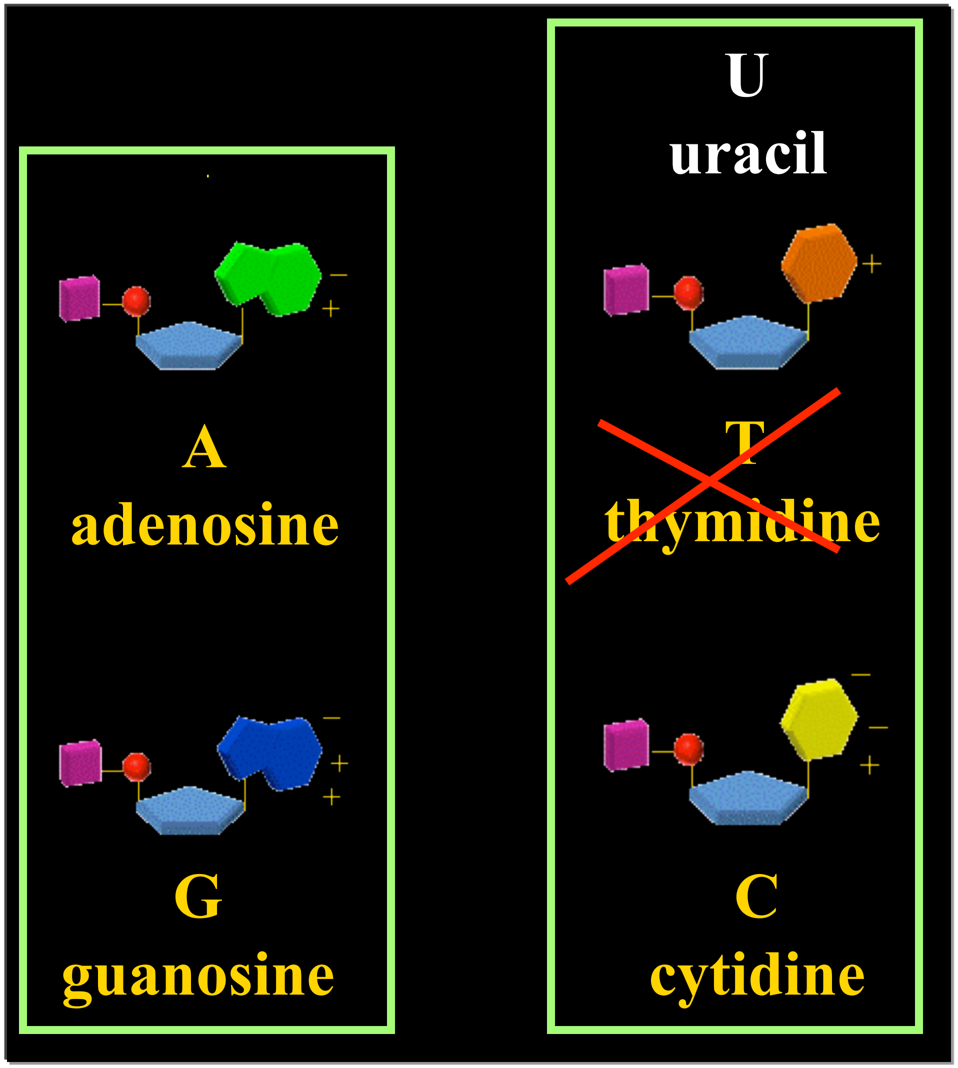


Purines **Pyrimidines**

(deoxy)Nucleotides – the building blocks of DNA

There are 4 'letters' (nucleotides)

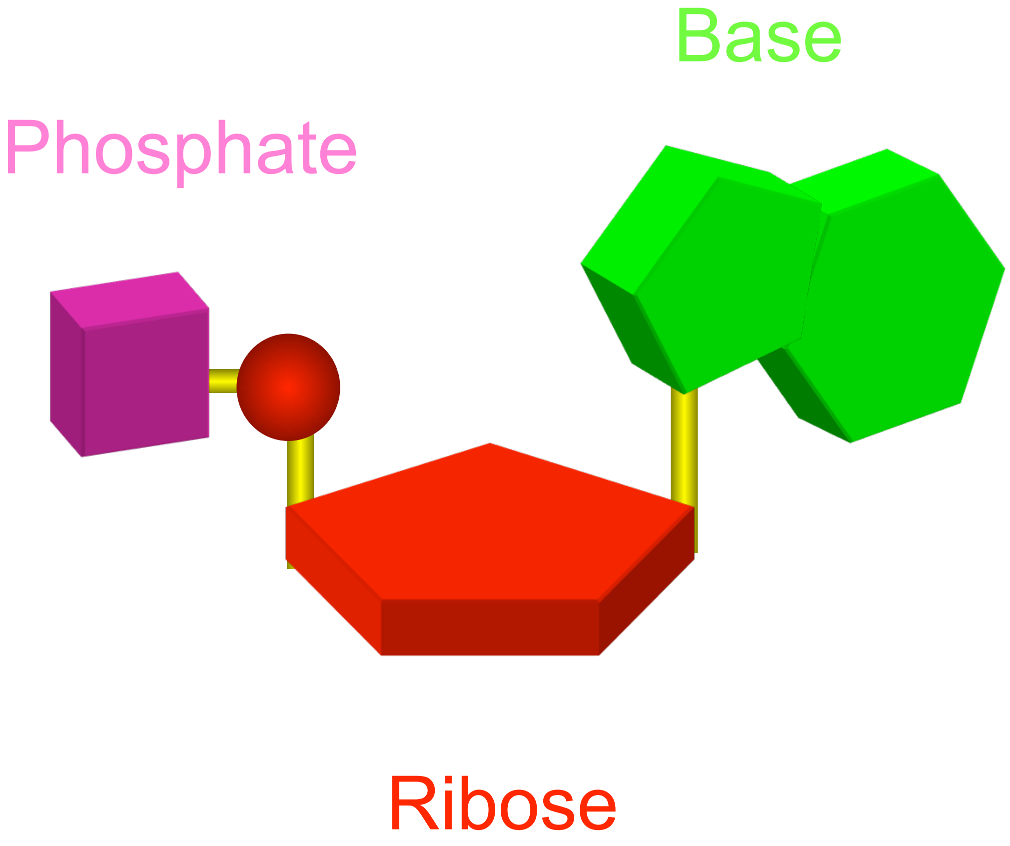




Purines **Pyrimidines**

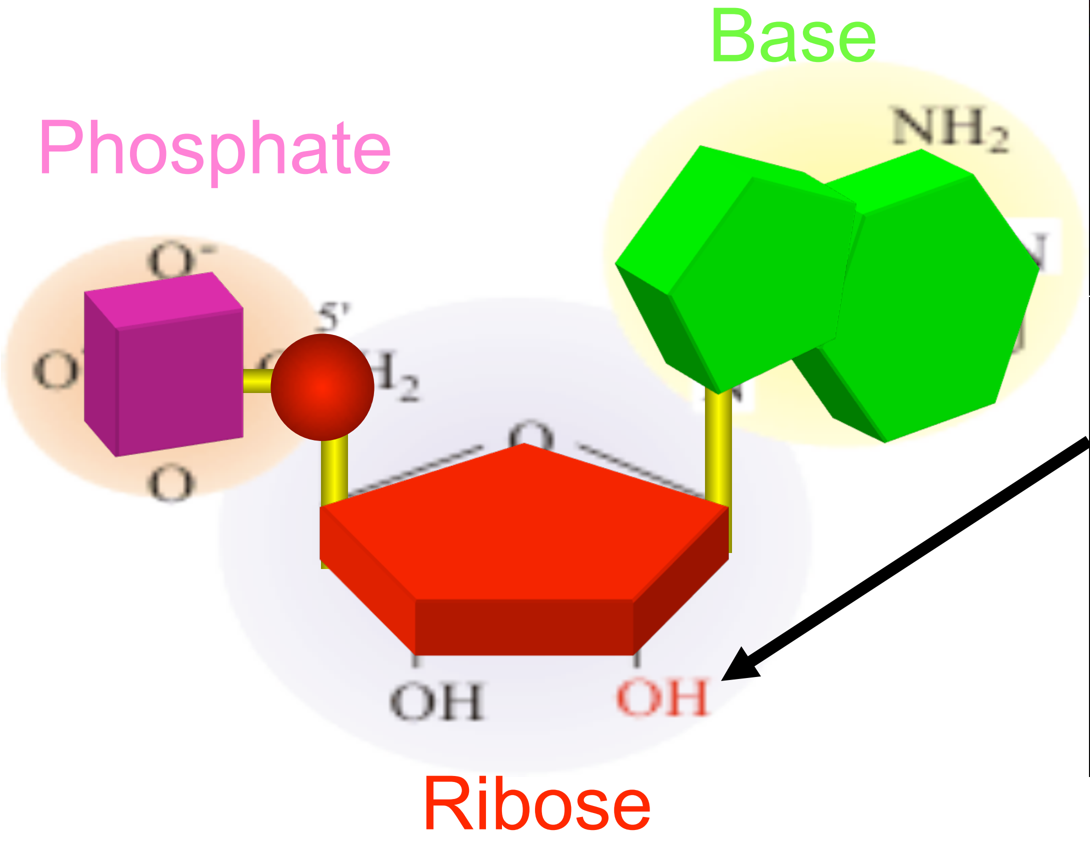
Guanosine

Anatomy of a nucleotide

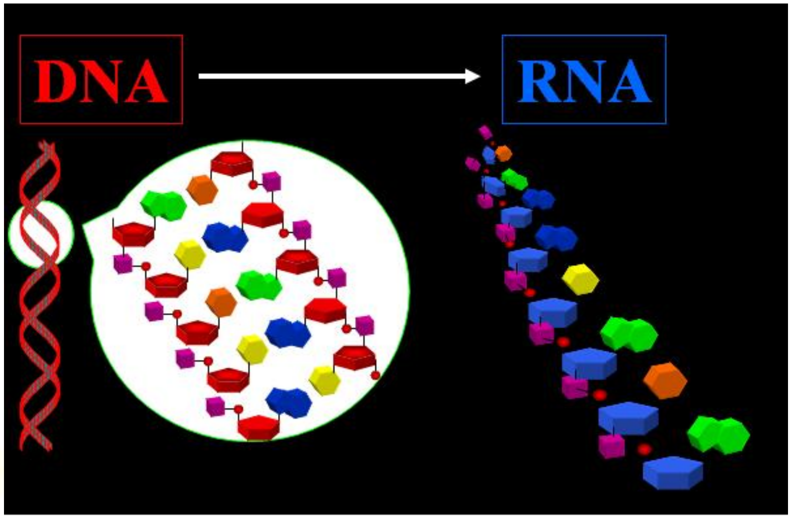


nucleotide

Anatomy of a nucleotide

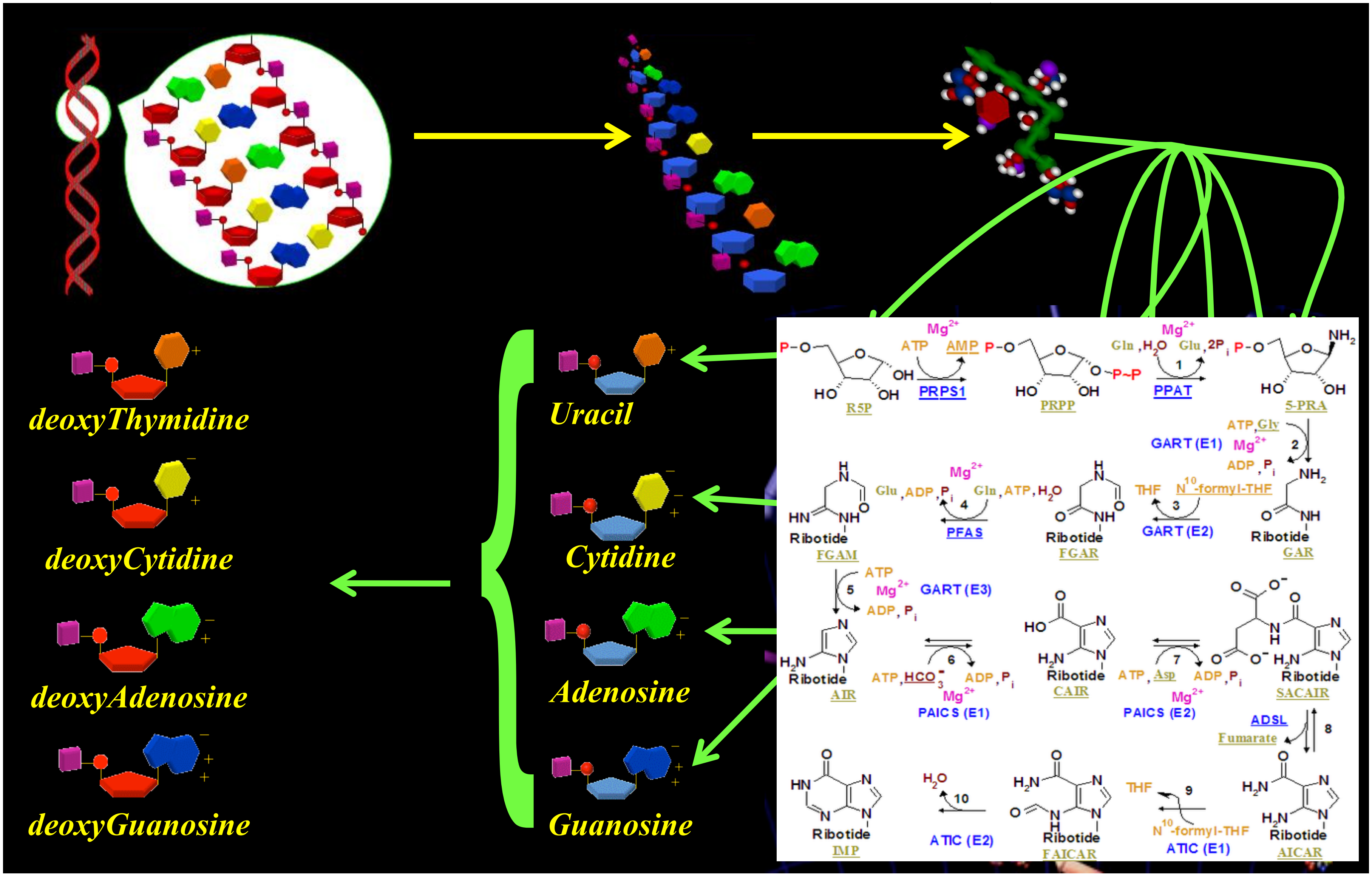


(deoxy)nucleotide

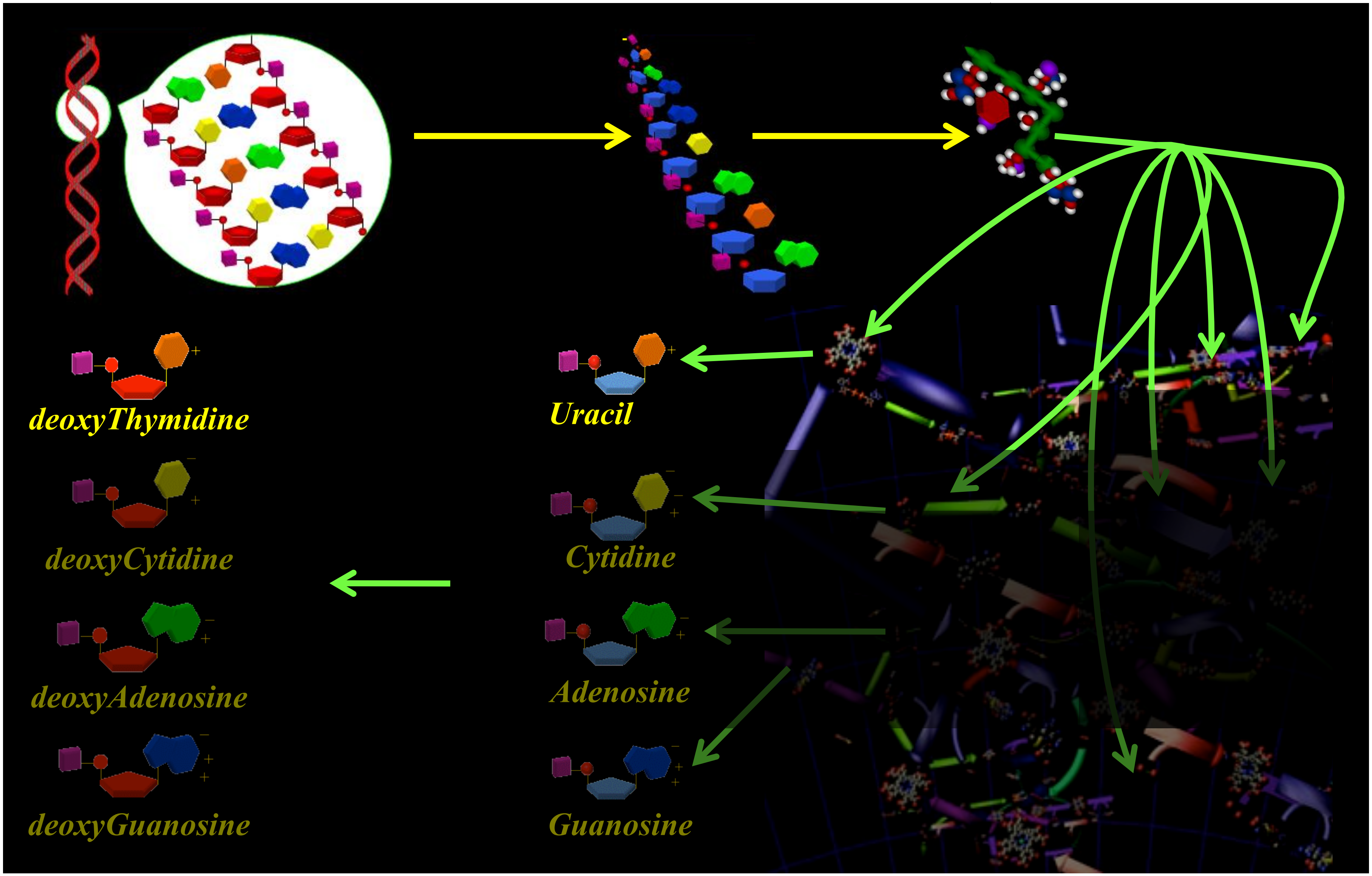


If this **OH** is reduced to **H** then ribose (used in **RNA**) becomes deoxyribose (used in **DNA**); this is the source of DNA's increased stability, relative to RNA

(deoxy)Nucleotides are formed by reducing nucleotides



d(Thymidine) is formed by manipulating Uracil



deoxyThymidine

Uracil

deoxyCytidine

Cytidine

deoxyAdenosine

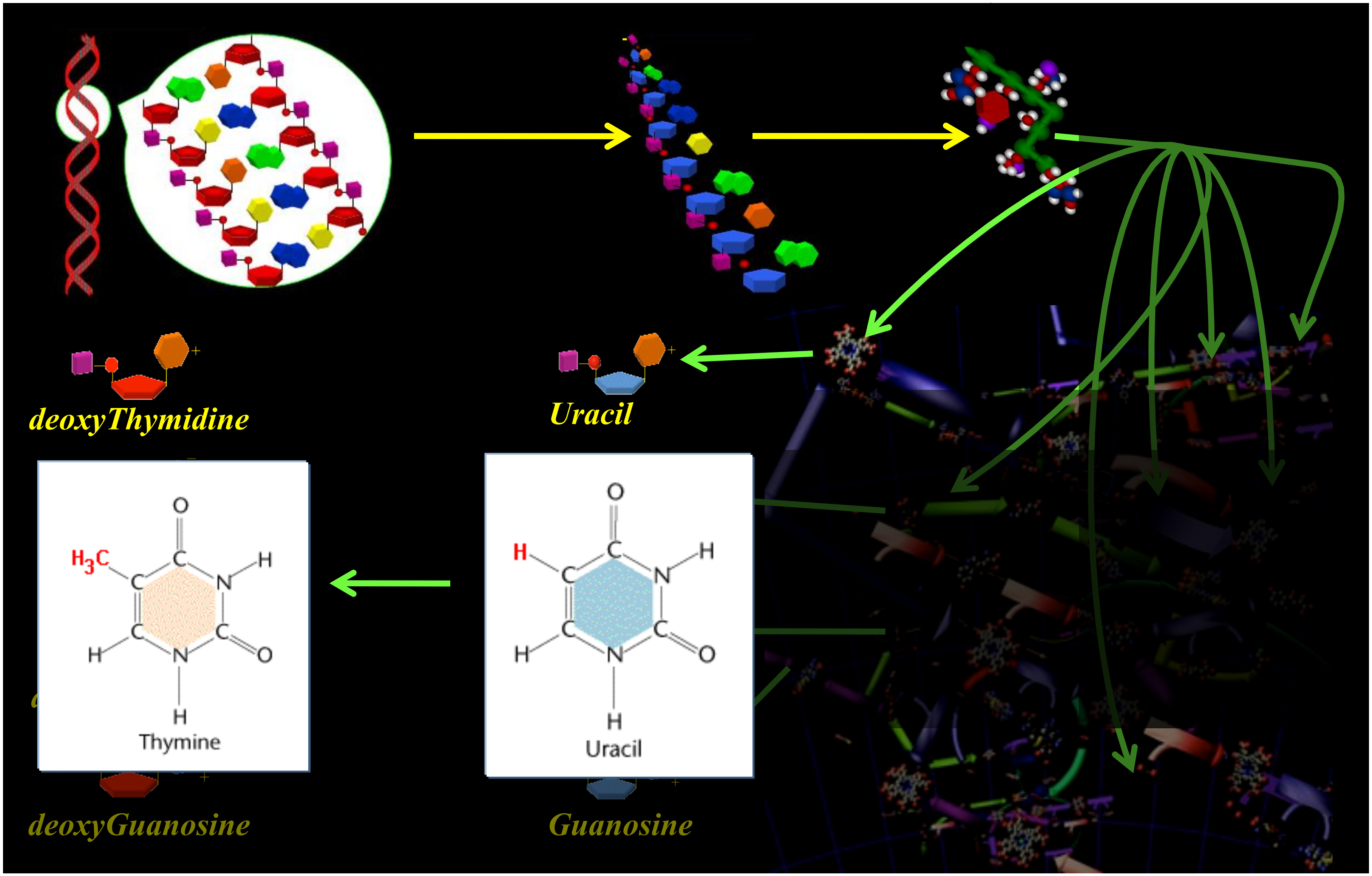
Adenosine

deoxyGuanosine

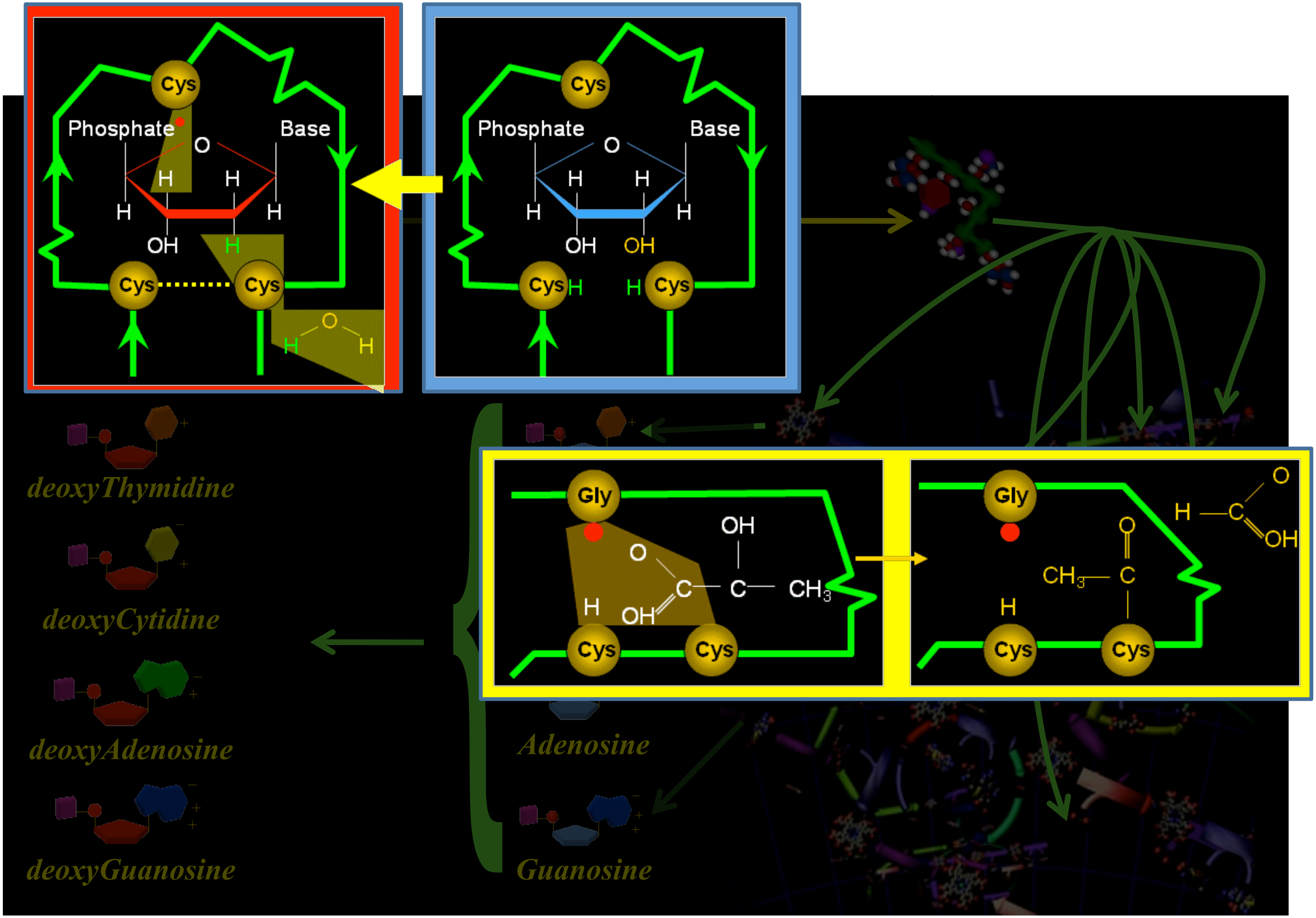
Guanosine

d(Thymidine) is formed by manipulating Uracil

…it would be parsimonious to assume that thymine had evolved before the machinery for transforming it into uracil!



(deoxy)Nucleotides are formed by reducing nucleotides

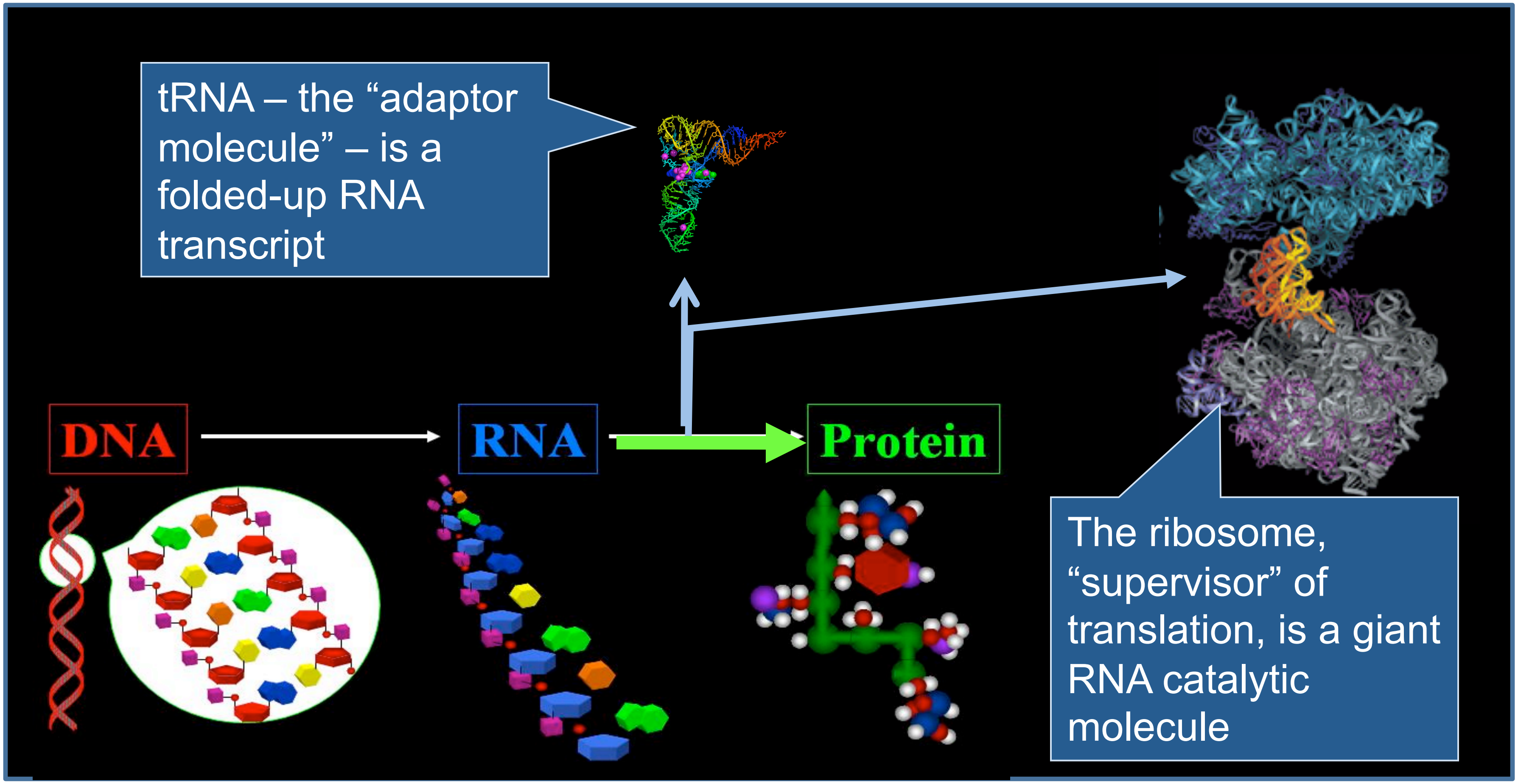


RNA evolved before DNA

1. **DNA is derived from RNA (Uracil is derived from Thymine; DeoxyRibose is derived from Ribose)**

   A. Lazcano, R. Guerrero, L. Margulis, J. Oro, J. Mol. Evol. 27, 283-290 (1988).

Some DNA genes are transcribed into RNA but never translated into proteins…



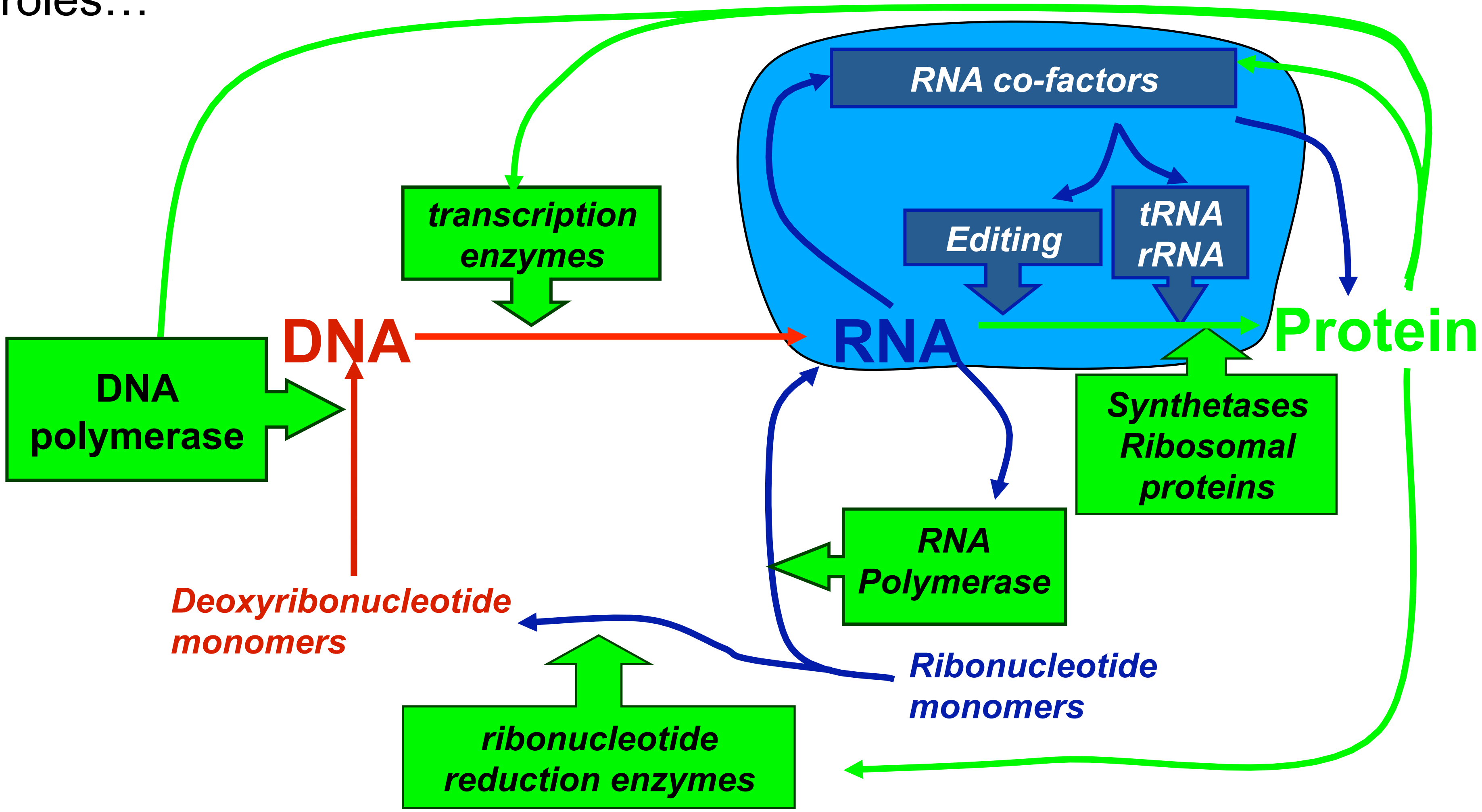
tRNA – the “adaptor molecule” – is a folded-up RNA transcript

The ribosome, “supervisor” of translation, is a giant RNA catalytic molecule

RNA can also fold into a complex shapes with non-informational (e.g. catalytic) properties…

Metabolic RNA as "molecular fossils"

'natural metabolic RNA' can be elegantly explained as molecular fossils of a truly ancient time when RNA performed both genetic and catalytic roles…



RNA evolved before DNA or proteins

1. **DNA is derived from RNA (Uracil is derived from Thymine; DeoxyRibose is derived from Ribose)**

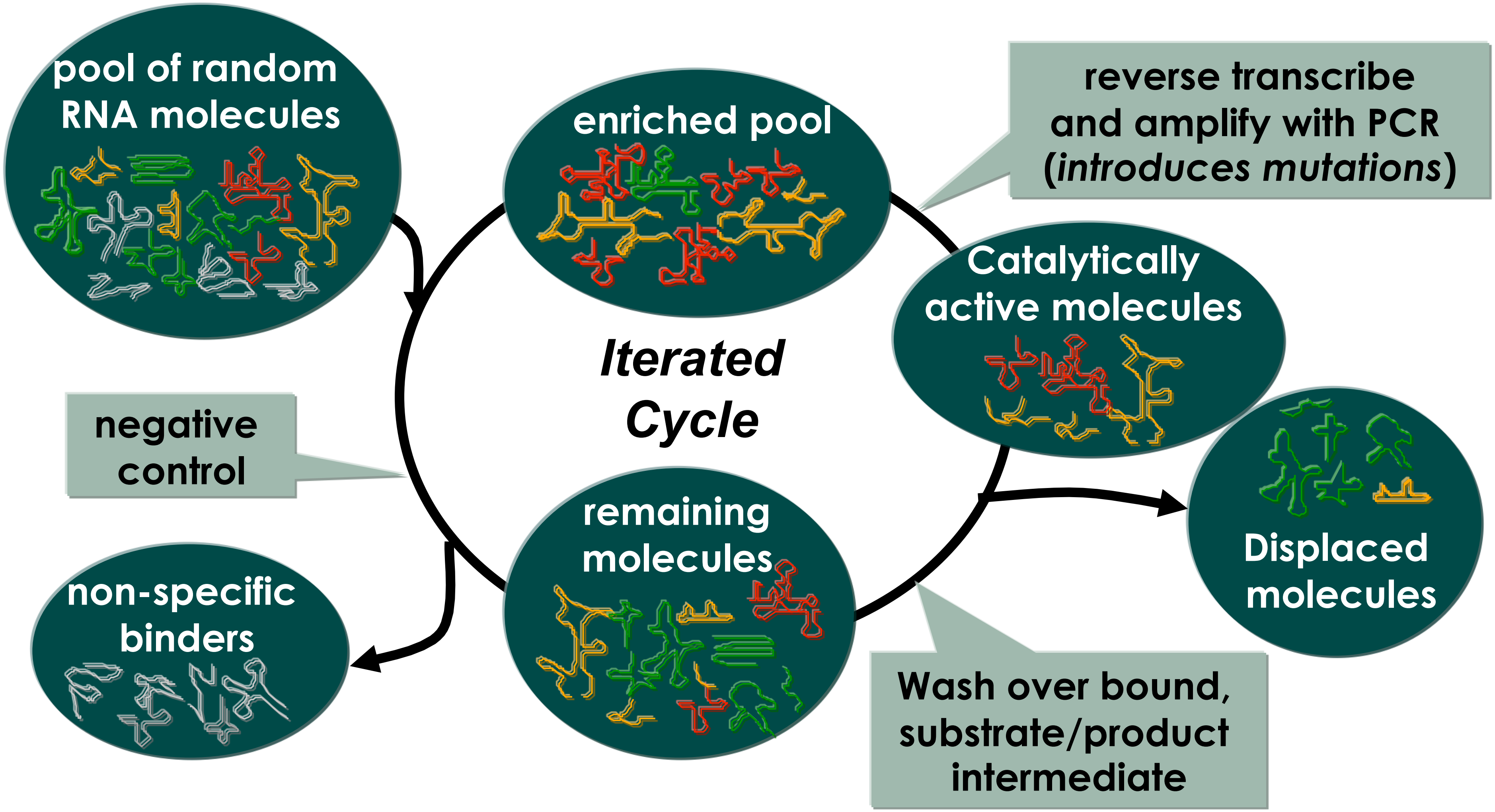
   A. Lazcano, R. Guerrero, L. Margulis, J. Oro, J. Mol. Evol. 27, 283-290 (1988).

2. **"Metabolic" RNA molecules (tRNAs, rRNA, ribonuclease P and enzyme co-factors) can be elegantly explained as molecular fossils** *H. B. White, J. Mol. Evol. 7, 101-104 (1976).*

3. **Looking carefully at 1 + 2, all the evidence suggests that RNA evolved "protein translation" before DNA arose**

   Freeland, Knight and Landweber, Science (2000)

In Vitro selection of *ribozymes*



pool of random RNA molecules

enriched pool

reverse transcribe and amplify with PCR (*introduces mutations*)

Catalytically active molecules

Iterated Cycle

negative control

Displaced molecules

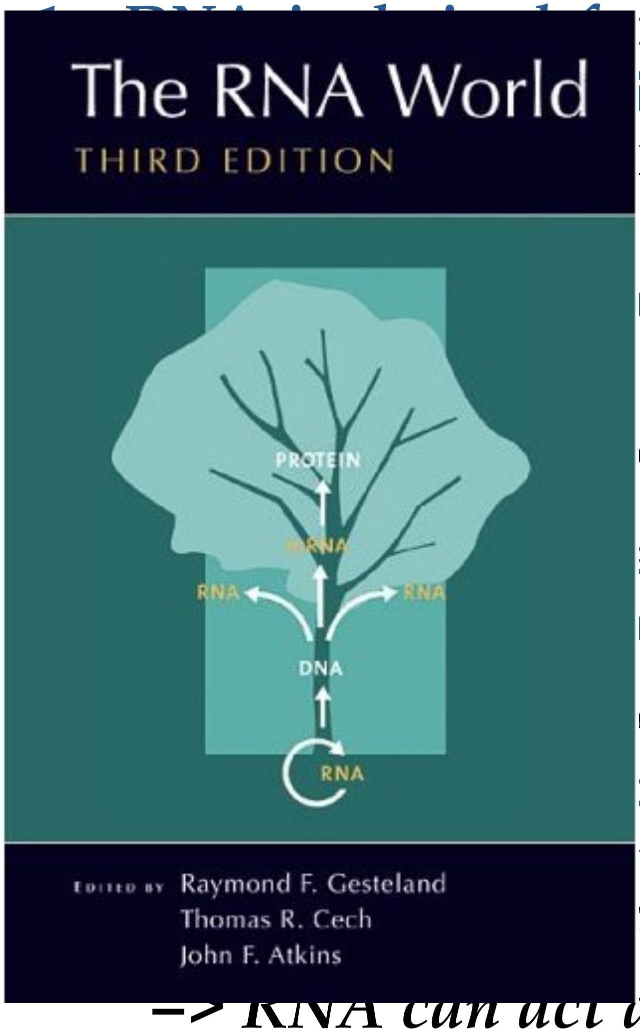
remaining molecules

non-specific binders

Wash over bound, substrate/product intermediate

Laboratory experiments have now created RNA molecules that perform a wide range of basic metabolic (catalytic) functions: these are ribozymes

Foundations of the RNA-world hypothesis


The RNA World

THIRD EDITION

EDITED BY Raymond F. Gesteland
Thomas R. Cech
John F. Atkins

om RNA (Uracil is derived from Thymine;
ived from Ribose)

. Margulis, J. Oro, J. Mol. Evol. 27, 283-290 (1988).

nolecules (tRNAs, rRNA, ribonuclease P and
can be elegantly explained as molecular

ol. Evol. 7, 101-104 (1976).

at 1 + 2, all the evidence suggests that RNA
nslation" before DNA arose

and Landweber, Science (2000)

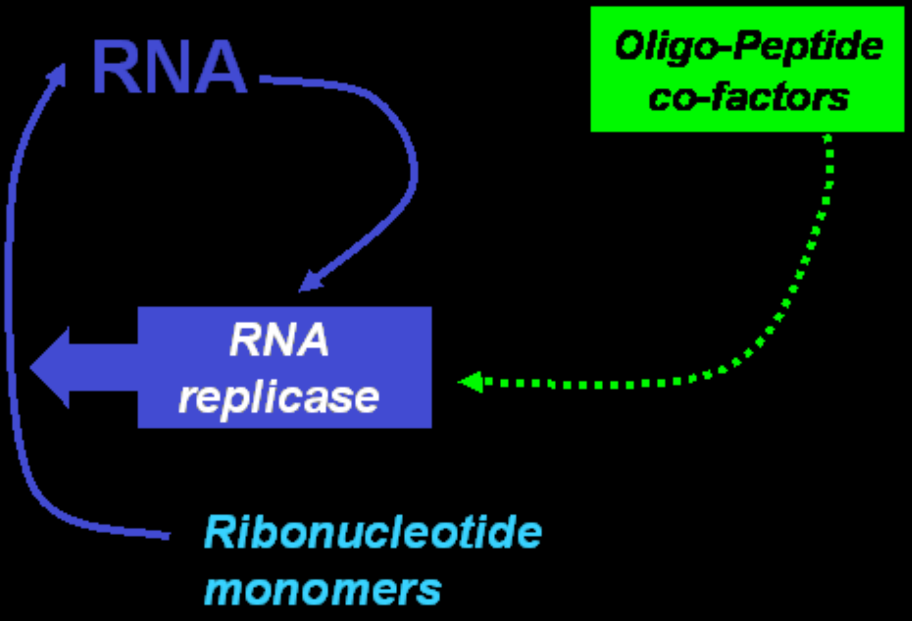
on experiments demonstrate an
tication for catalytic RNA

ds., The RNA World, 2nd ed (1999) Cold Spring Harbor Lab. Press

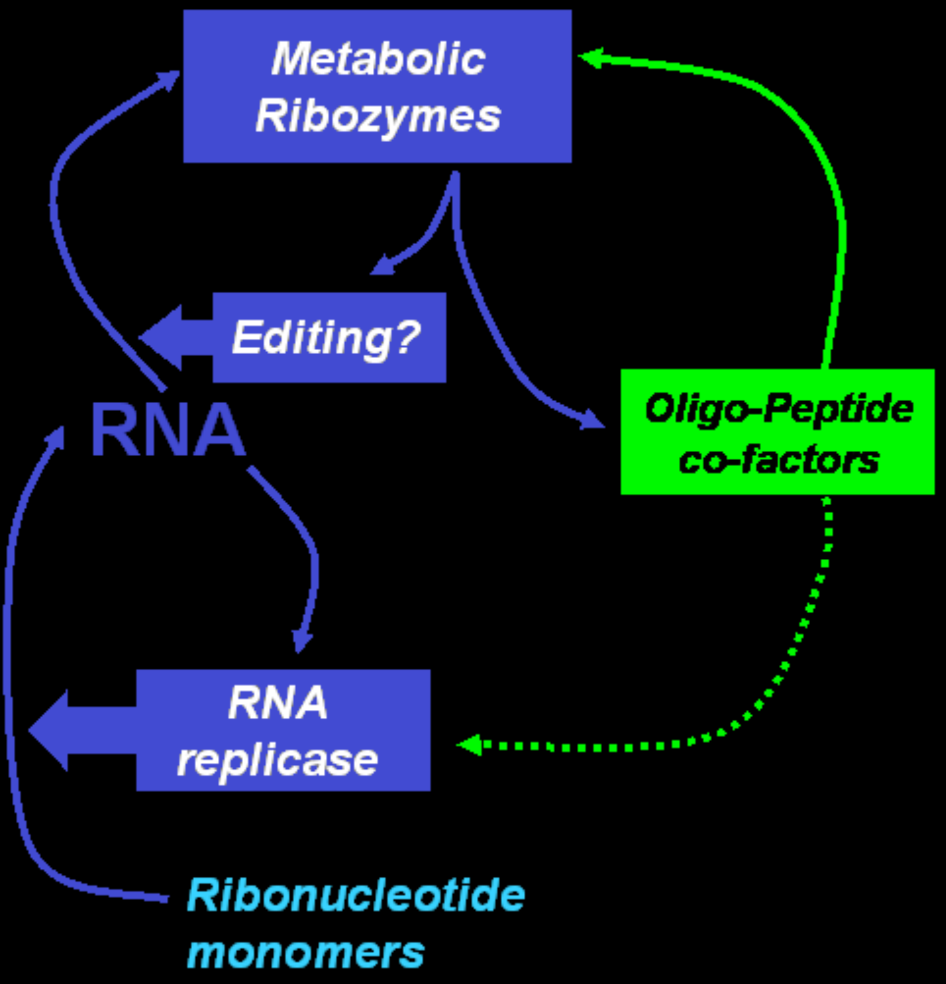
–> RNA can act as catalyst and information template!

(despite over 20 years of attempts, there is no evidence
for template-driven replication of proteins!)

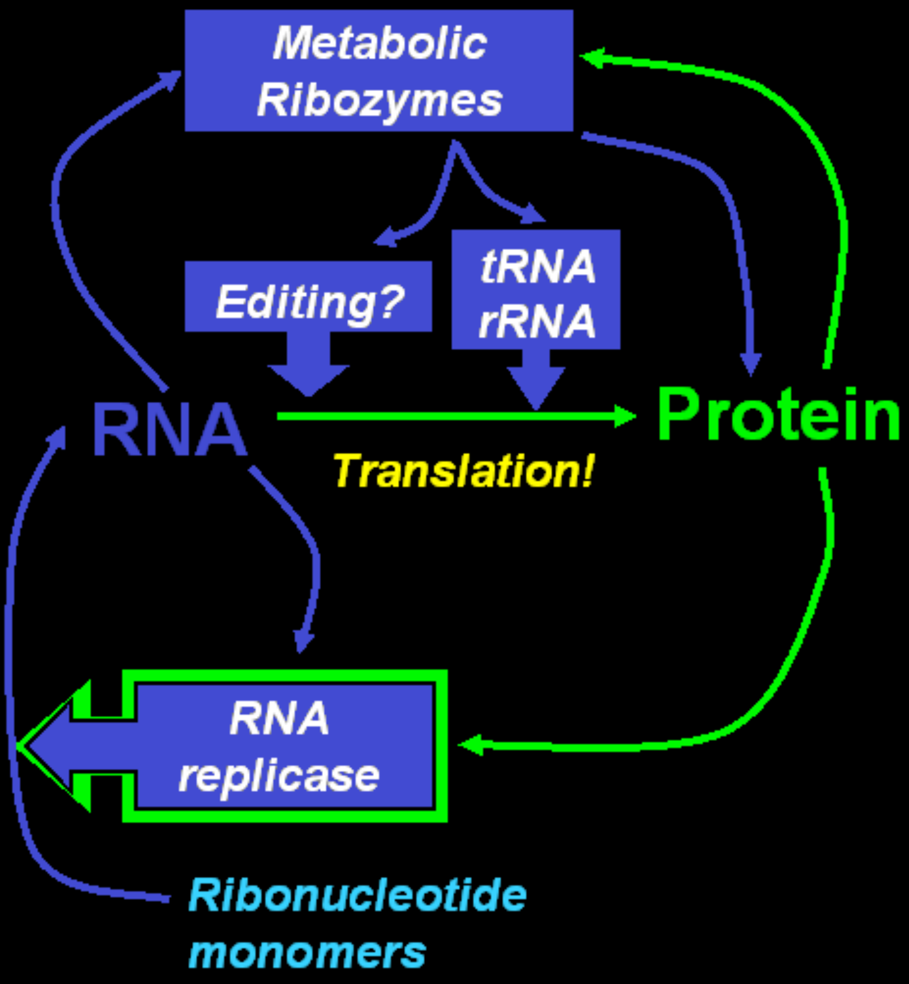
Solving the Puzzle of Primordial Self-Replication



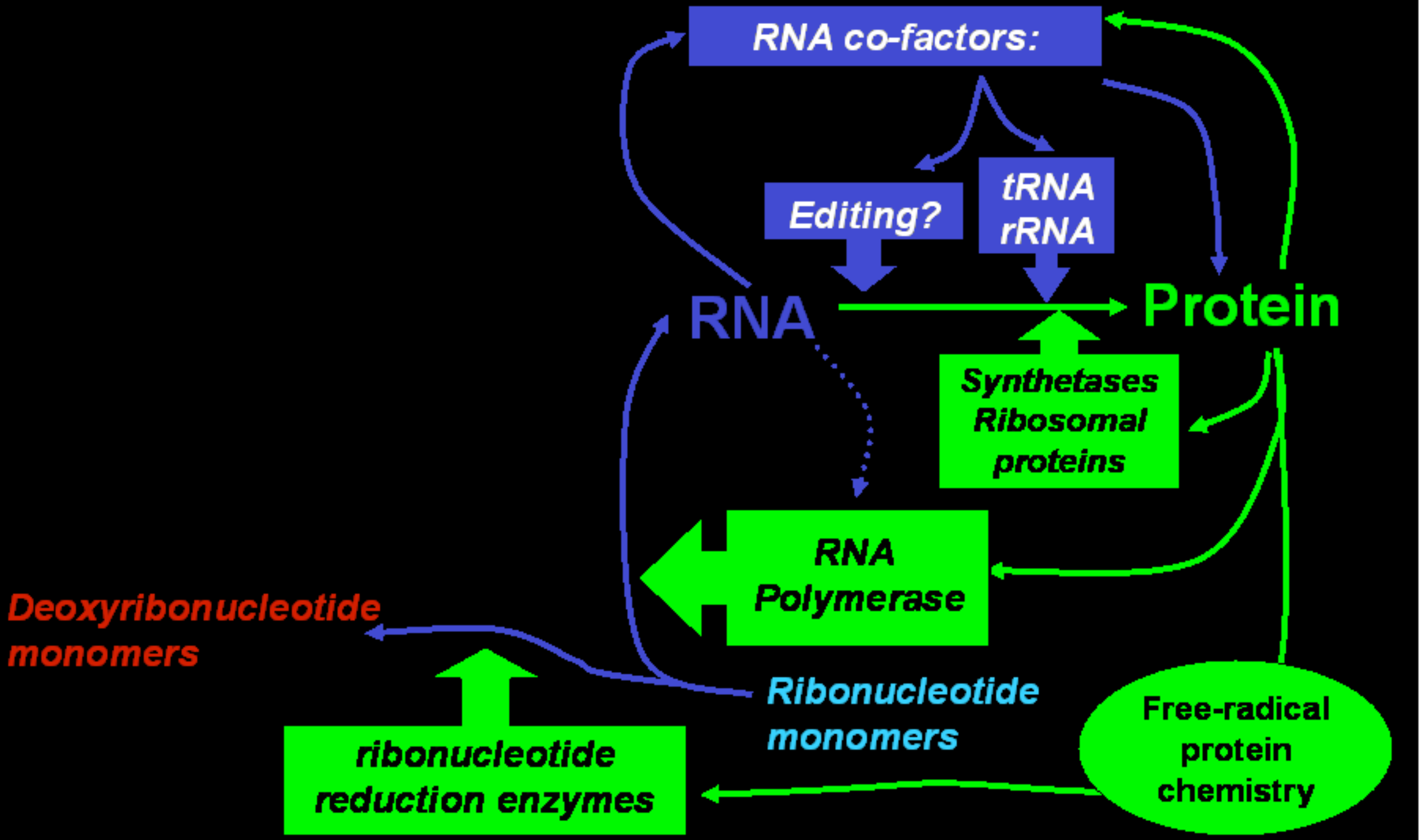
Solving the Puzzle of Primordial Self-Replication



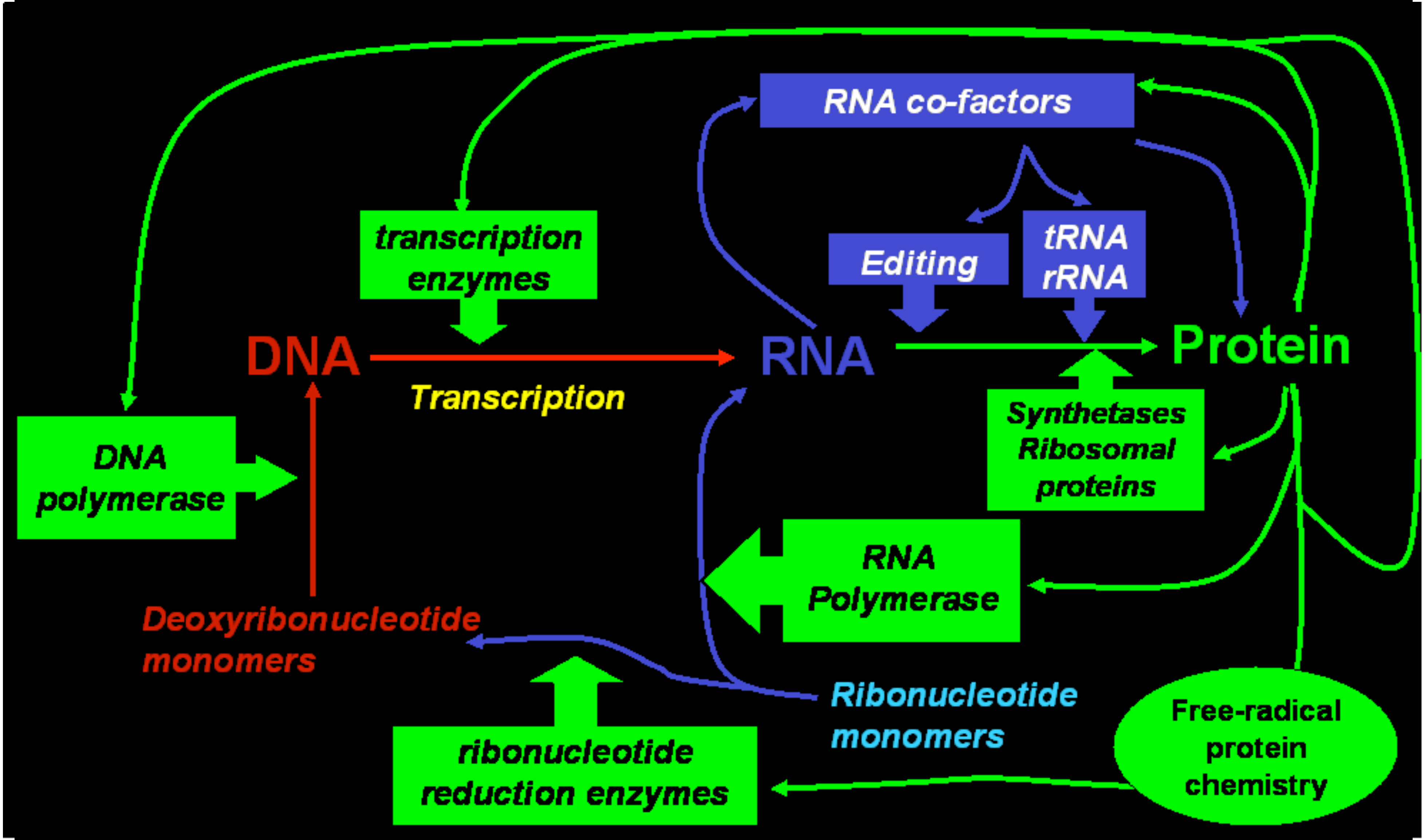
Solving the Puzzle of Primordial Self-Replication



Solving the Puzzle of Primordial Self-Replication

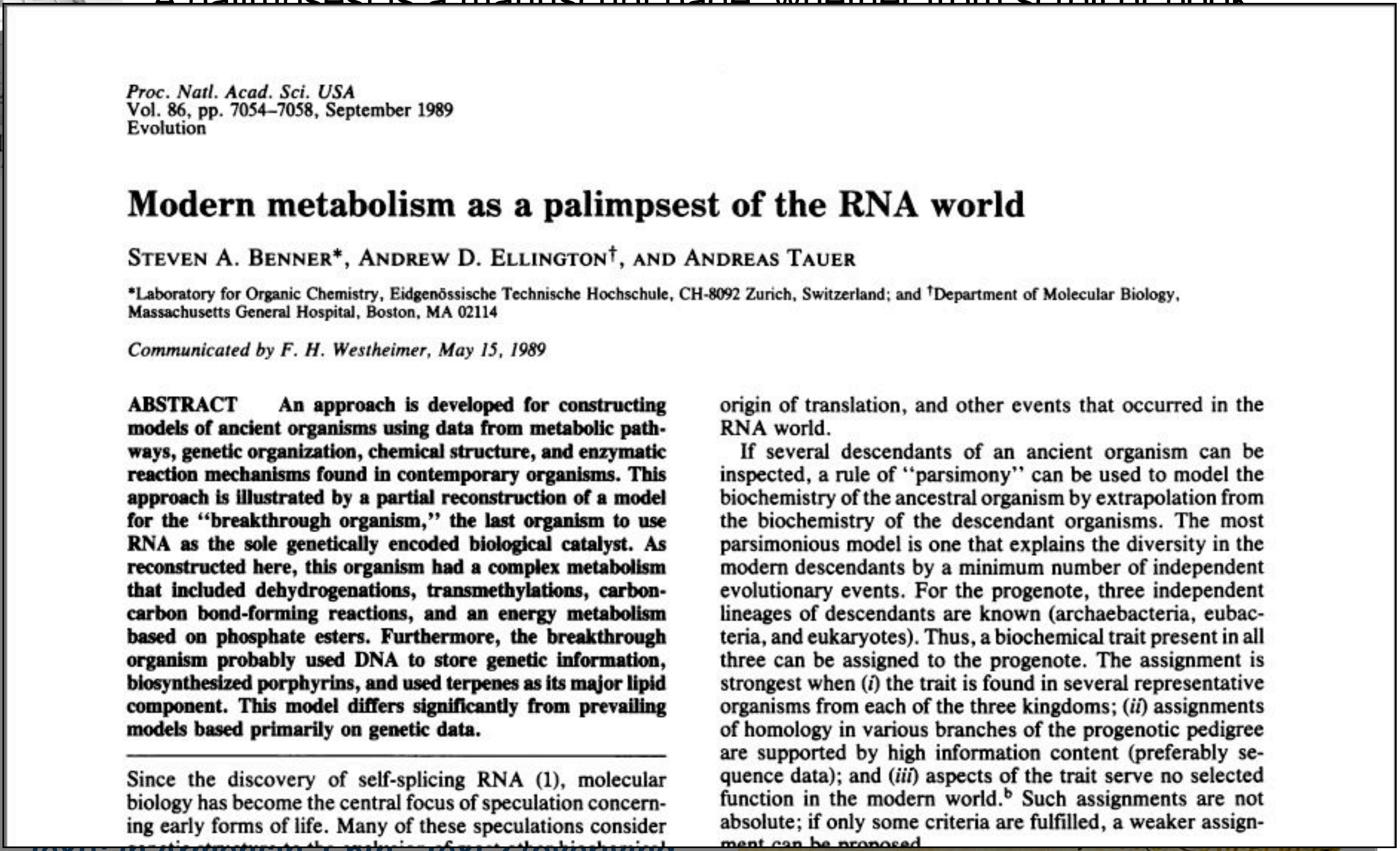


Solving the Puzzle of Primordial Self-Replication



the *palimpsest* view of metabolism

“A palimpsest is a manuscript page, whether from scroll or book



Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 7054–7058, September 1989
Evolution

Modern metabolism as a palimpsest of the RNA world

STEVEN A. BENNER*, ANDREW D. ELLINGTON†, AND ANDREAS TAUER

*Laboratory for Organic Chemistry, Eidgenössische Technische Hochschule, CH-8092 Zurich, Switzerland; and †Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

Communicated by F. H. Westheimer, May 15, 1989

ABSTRACT An approach is developed for constructing models of ancient organisms using data from metabolic pathways, genetic organization, chemical structure, and enzymatic reaction mechanisms found in contemporary organisms. This approach is illustrated by a partial reconstruction of a model for the “breakthrough organism,” the last organism to use RNA as the sole genetically encoded biological catalyst. As reconstructed here, this organism had a complex metabolism that included dehydrogenations, transmethylations, carbon-carbon bond-forming reactions, and an energy metabolism based on phosphate esters. Furthermore, the breakthrough organism probably used DNA to store genetic information, biosynthesized porphyrins, and used terpenes as its major lipid component. This model differs significantly from prevailing models based primarily on genetic data.

Since the discovery of self-splicing RNA (1), molecular biology has become the central focus of speculation concerning early forms of life. Many of these speculations consider

origin of translation, and other events that occurred in the RNA world.

If several descendants of an ancient organism can be inspected, a rule of “parsimony” can be used to model the biochemistry of the ancestral organism by extrapolation from the biochemistry of the descendant organisms. The most parsimonious model is one that explains the diversity in the modern descendants by a minimum number of independent evolutionary events. For the progenote, three independent lineages of descendants are known (archaebacteria, eubacteria, and eukaryotes). Thus, a biochemical trait present in all three can be assigned to the progenote. The assignment is strongest when (*i*) the trait is found in several representative organisms from each of the three kingdoms; (*ii*) assignments of homology in various branches of the progenotic pedigree are supported by high information content (preferably sequence data); and (*iii*) aspects of the trait serve no selected function in the modern world.[b] Such assignments are not absolute; if only some criteria are fulfilled, a weaker assign-

text): In Armenian, C6th. Text 2 (overlying

text): In Syriac, 1st half of C10th.

From the Schoyen Collection, Section 5 (Patristic literature)
http://www.schoyencollection.com/patristic.htm

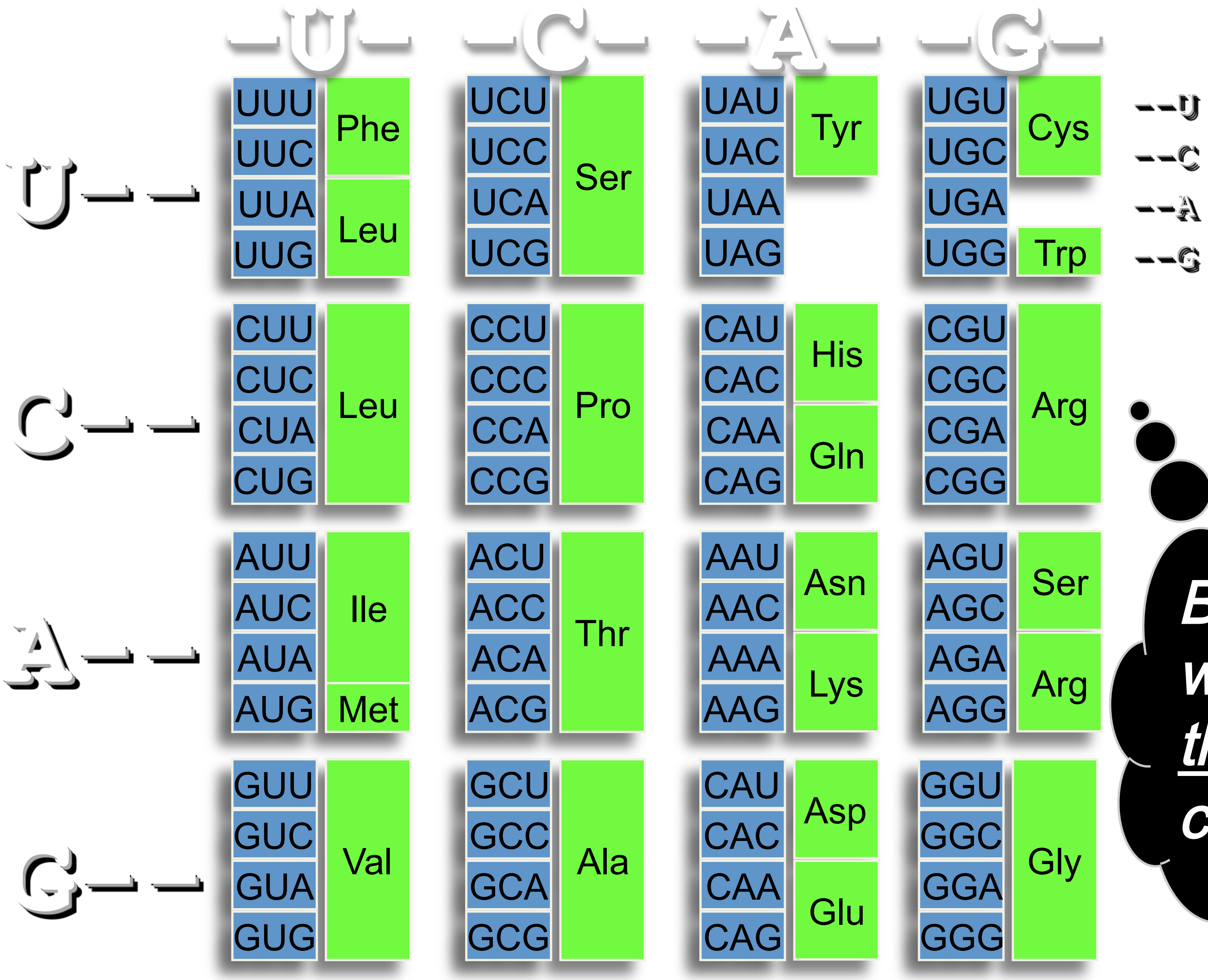
(P.S. note: Crick got here in 1968)

Francis Crick, 1953, posing with his co-solved structure of DNA (reported in Nature volume 171 : pages 737-738 & 964-967)

"In looking at … present-day … protein synthesis, once is struck by the considerable involvement of non informational nucleic acid. The ribosomes are mainly made from RNA and the adaptor molecules (tRNA) are exclusively RNA …Why is this?...we cannot help feeling that…rRNA and tRNA…were part of the primitive machinery for protein synthesis…In fact, as has been noted elsewhere, tRNA looks like Nature's attempt to make RNA do the job of a protein (Crick, 1966).

If indeed rRNA and tRNA were essential parts of the primitive machinery, one naturally asks how much protein, if any, was then needed. It is tempting to wonder if … the primitive machinery had no protein at all and consisted entirely of RNA"

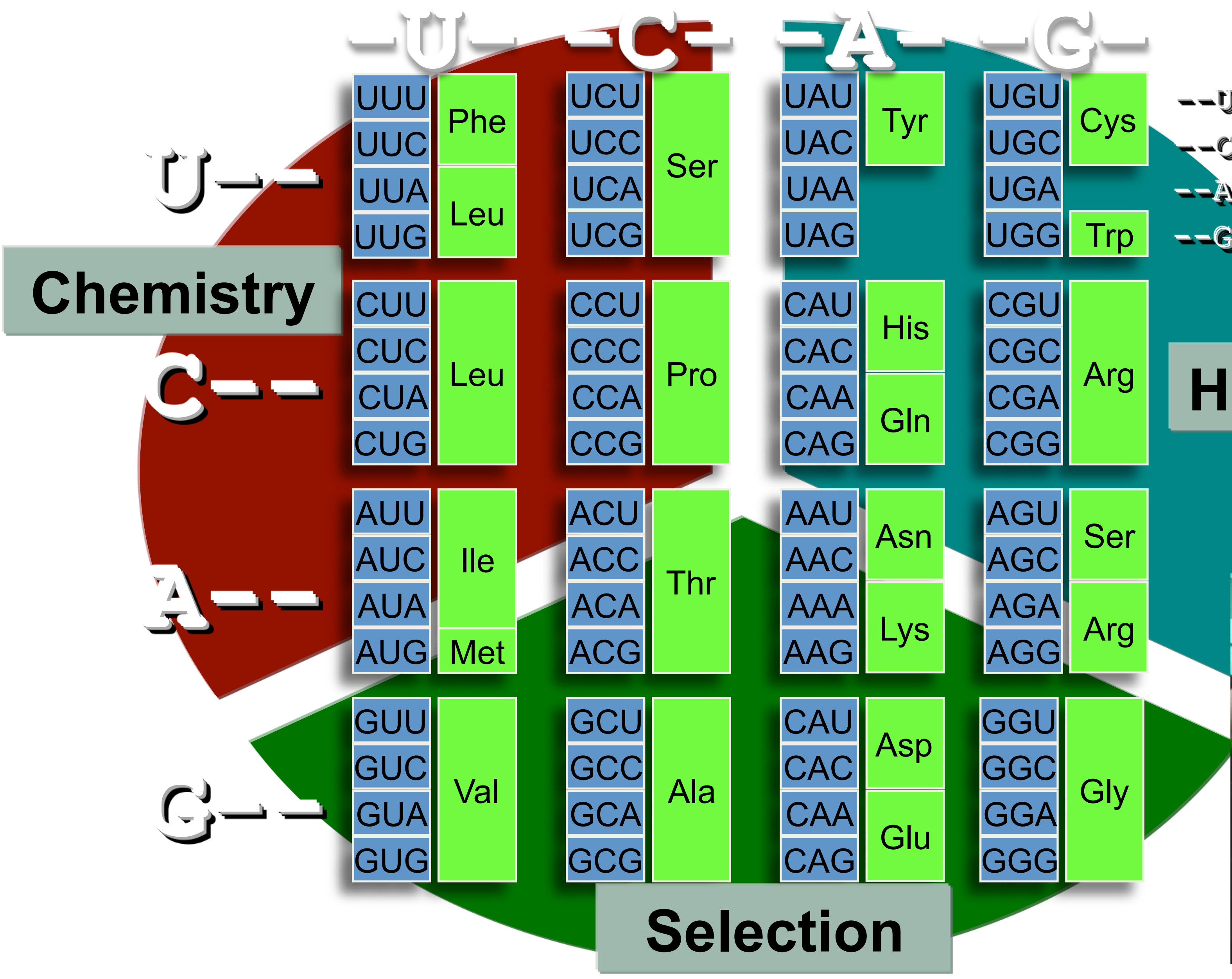
Crick, F.H.C. (1968) (page 371)



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

But why this code?

The 3 Faces of the Genetic Code

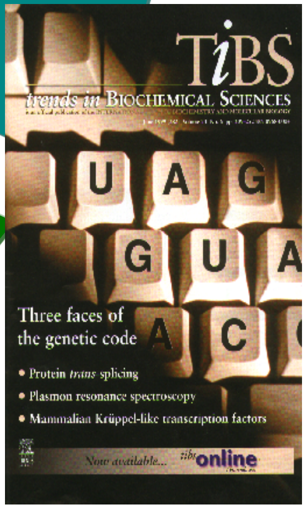


	–U–		–C–		–A–		–G–		
U– –	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	--U
	UUC		UCC		UAC		UGC		--C
	UUA	Leu	UCA		UAA		UGA		--A
	UUG		UCG		UAG		UGG	Trp	--G
C– –	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Gln	CGA		
	CUG		CCG		CAG		CGG		
A– –	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	AUC		ACC		AAC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

Chemistry

History

Selection



TiBS

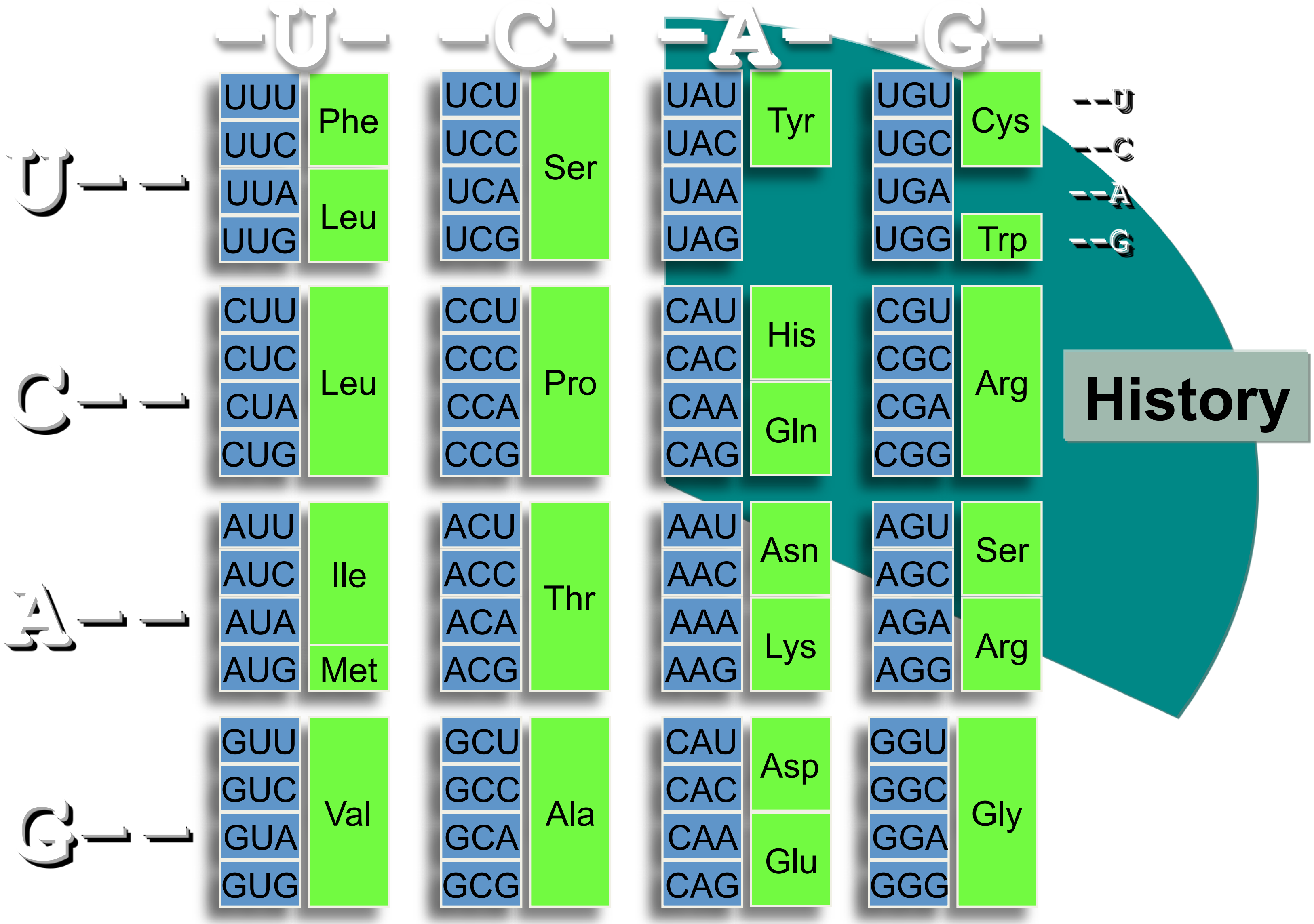
trends in BIOCHEMICAL SCIENCES

Three faces of the genetic code

- Protein *trans* splicing
- Plasmon resonance spectroscopy
- Mammalian Krüppel-like transcription factors

Now available... tibsonline

A code that reveals its own evolutionary history…



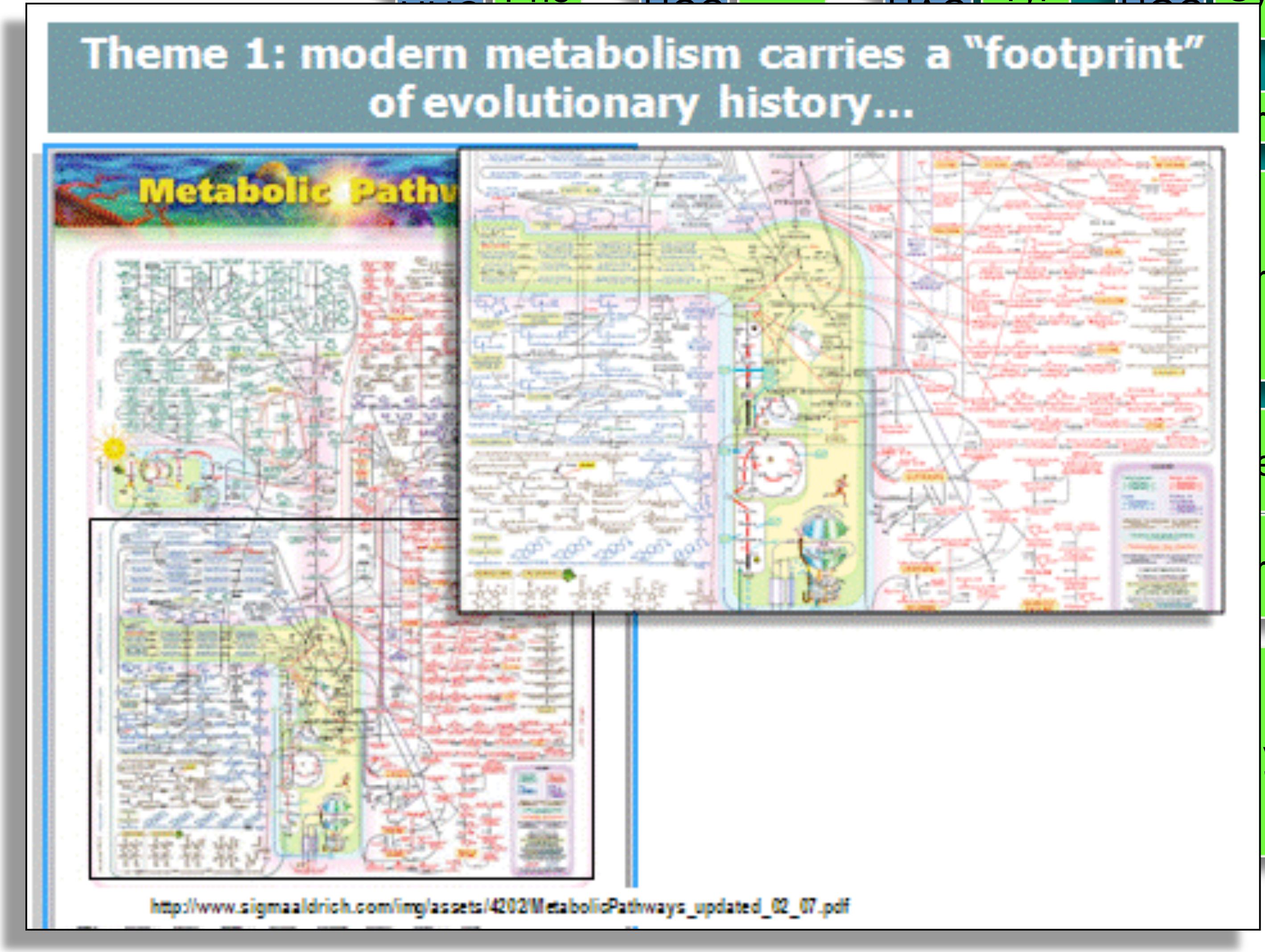
	-U-		-C-		-A-		-G-		
U--	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	--U
	UUC		UCC		UAC		UGC		--C
	UUA	Leu	UCA		UAA		UGA		--A
	UUG		UCG		UAG		UGG	Trp	--G
C--	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Gln	CGA		
	CUG		CCG		CAG		CGG		
A--	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	AUC		ACC		AAC	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

A code that reveals its own evolutionary history…

–U–		–C–		–A–		–G–	
UUU	Phe	UCU		UAU	Tyr	UGU	Cys
							rp
							rg
							er
							rg
							y

--U
--C
--A
--G

History

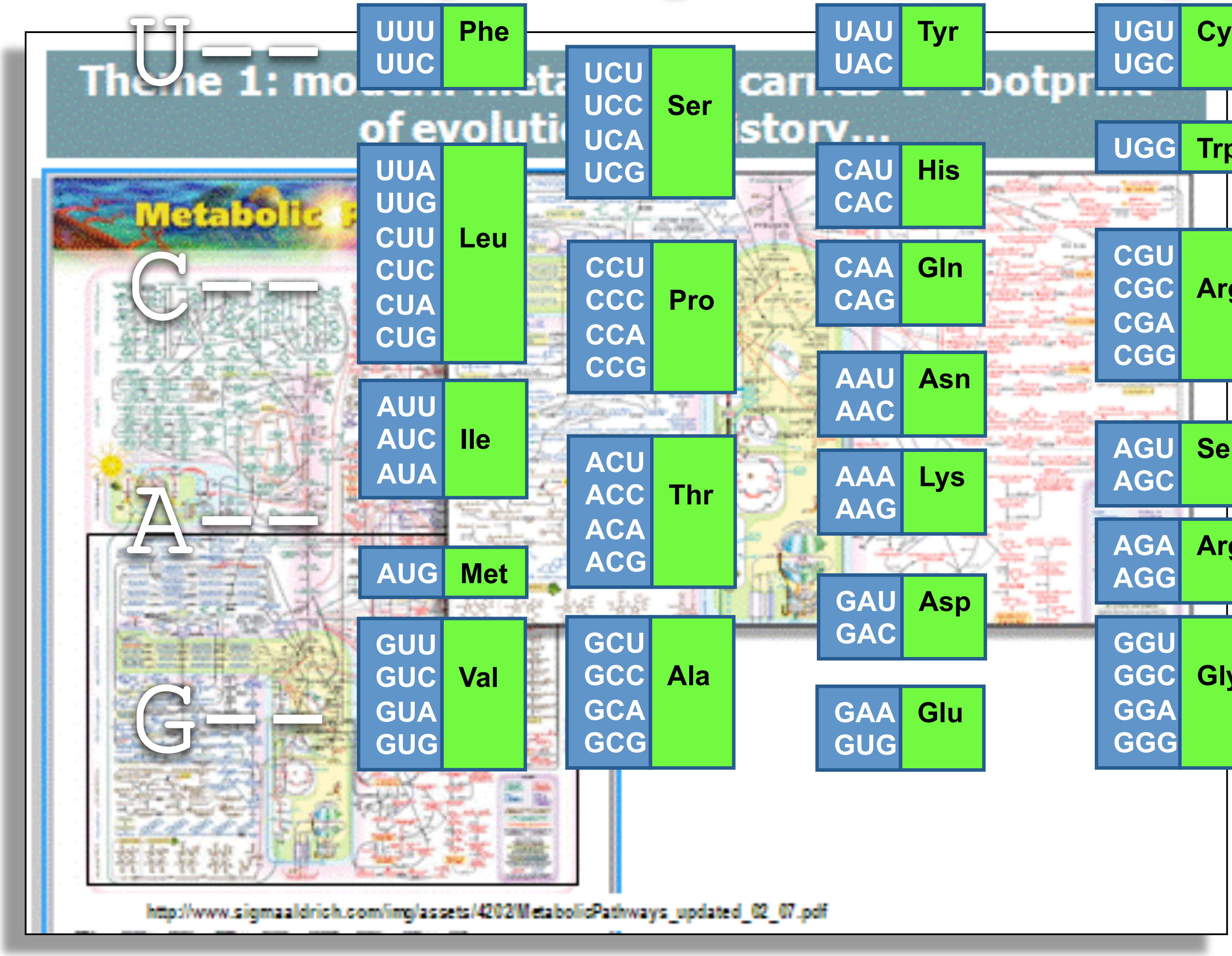


Theme 1: modern metabolism carries a "footprint" of evolutionary history...

Metabolic Pathw

http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_updated_02_07.pdf

Amino acid metabolism & codon assignments

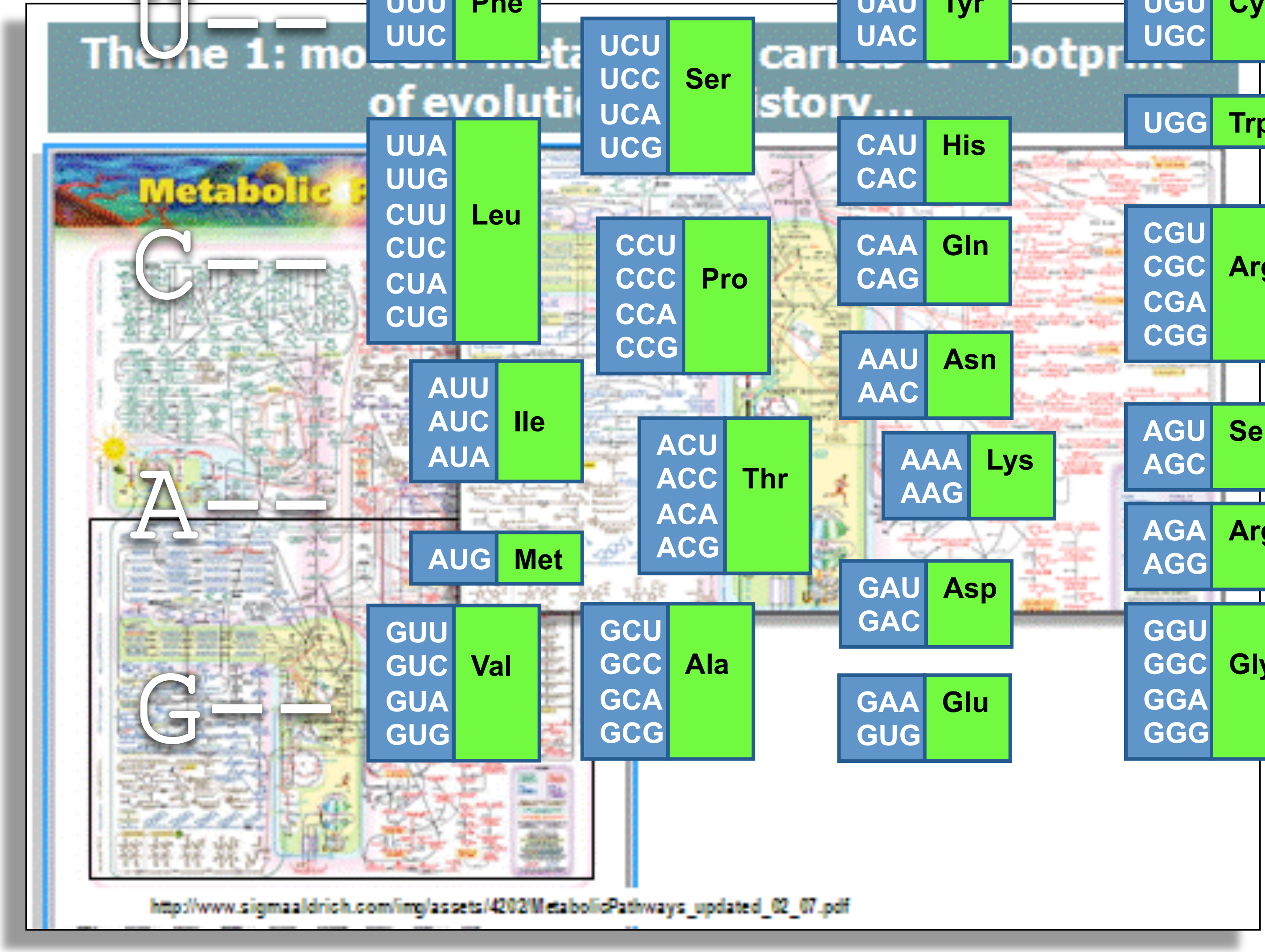


Theme 1: mo meta carri otpr of evoluti istory...

http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_updated_02_07.pdf

	–U–		–C–		–A–		–G–		
U– –	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	– –U
	UUC		UCC		UAC		UGC		– –C
	UUA	Leu	UCA						– –A
	UUG		UCG				UGG	Trp	– –G
C– –	CUU		CCU	Pro	CAU	His	CGU	Arg	– –U
	CUC		CCC		CAC		CGC		– –C
	CUA		CCA		CAA	Gln	CGA		– –A
	CUG		CCG		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		GUG		GGG		– –G

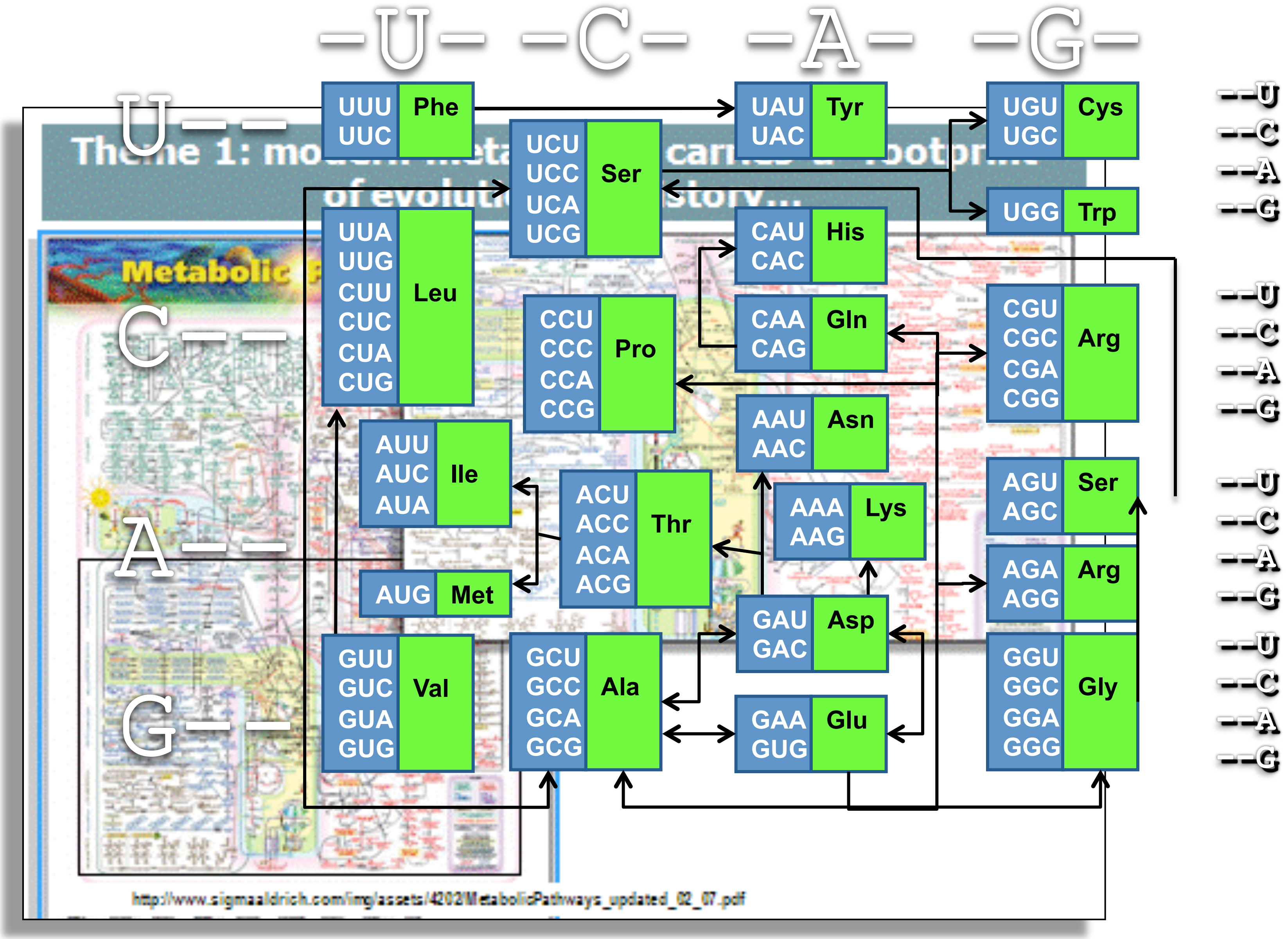
Amino acid metabolism & codon assignments



http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_updated_02_07.pdf

	–U–	–C–	–A–	–G–	
U– –	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	– –U
	UUC	UCC	UAC	UGC	– –C
	UUA Leu	UCA			– –A
	UUG	UCG		UGG Trp	– –G
C– –	CUU	CCU Pro	CAU His	CGU Arg	– –U
	CUC	CCC	CAC	CGC	– –C
	CUA	CCA	CAA Gln	CGA	– –A
	CUG	CCG	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	GUG	GGG	– –G

Amino acid metabolism & codon assignments



–U– –C– –A– –G–

U– –

UUU UUC Phe

UUA UUG Leu

UCU UCC UCA UCG Ser

UAU UAC Tyr

UGU UGC Cys

UGG Trp

C– –

CUU CUC CUA CUG Leu

CCU CCC CCA CCG Pro

CAU CAC His

CAA CAG Gln

CGU CGC CGA CGG Arg

A– –

AUU AUC AUA Ile

AUG Met

ACU ACC ACA ACG Thr

AAU AAC Asn

AAA AAG Lys

AGU AGC Ser

AGA AGG Arg

G– –

GUU GUC GUA GUG Val

GCU GCC GCA GCG Ala

GAU GAC Asp

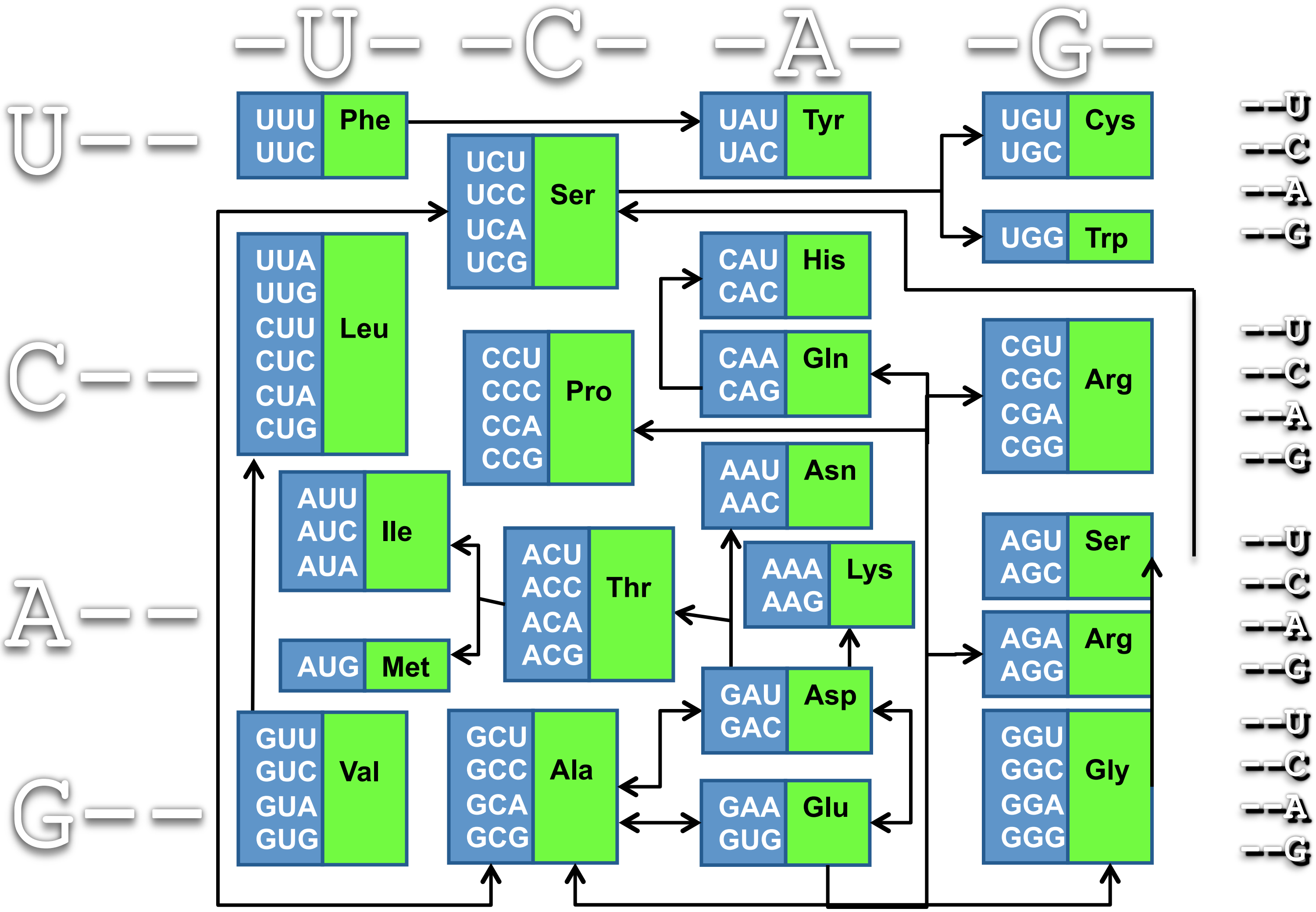
GAA GAG Glu

GGU GGC GGA GGG Gly

– –U
– –C
– –A
– –G

http://www.sigmaaldrich.com/img/assets/4202/MetabolicPathways_updated_02_07.pdf

Amino acid metabolism & codon assignments



–U– –C– –A– –G–

U– –

UUU UUC Phe

UCU UCC UCA UCG Ser

UAU UAC Tyr

UGU UGC Cys

UGG Trp

C– –

UUA UUG CUU CUC CUA CUG Leu

CCU CCC CCA CCG Pro

CAU CAC His

CAA CAG Gln

CGU CGC CGA CGG Arg

A– –

AUU AUC AUA Ile

AUG Met

ACU ACC ACA ACG Thr

AAU AAC Asn

AAA AAG Lys

AGU AGC Ser

AGA AGG Arg

G– –

GUU GUC GUA GUG Val

GCU GCC GCA GCG Ala

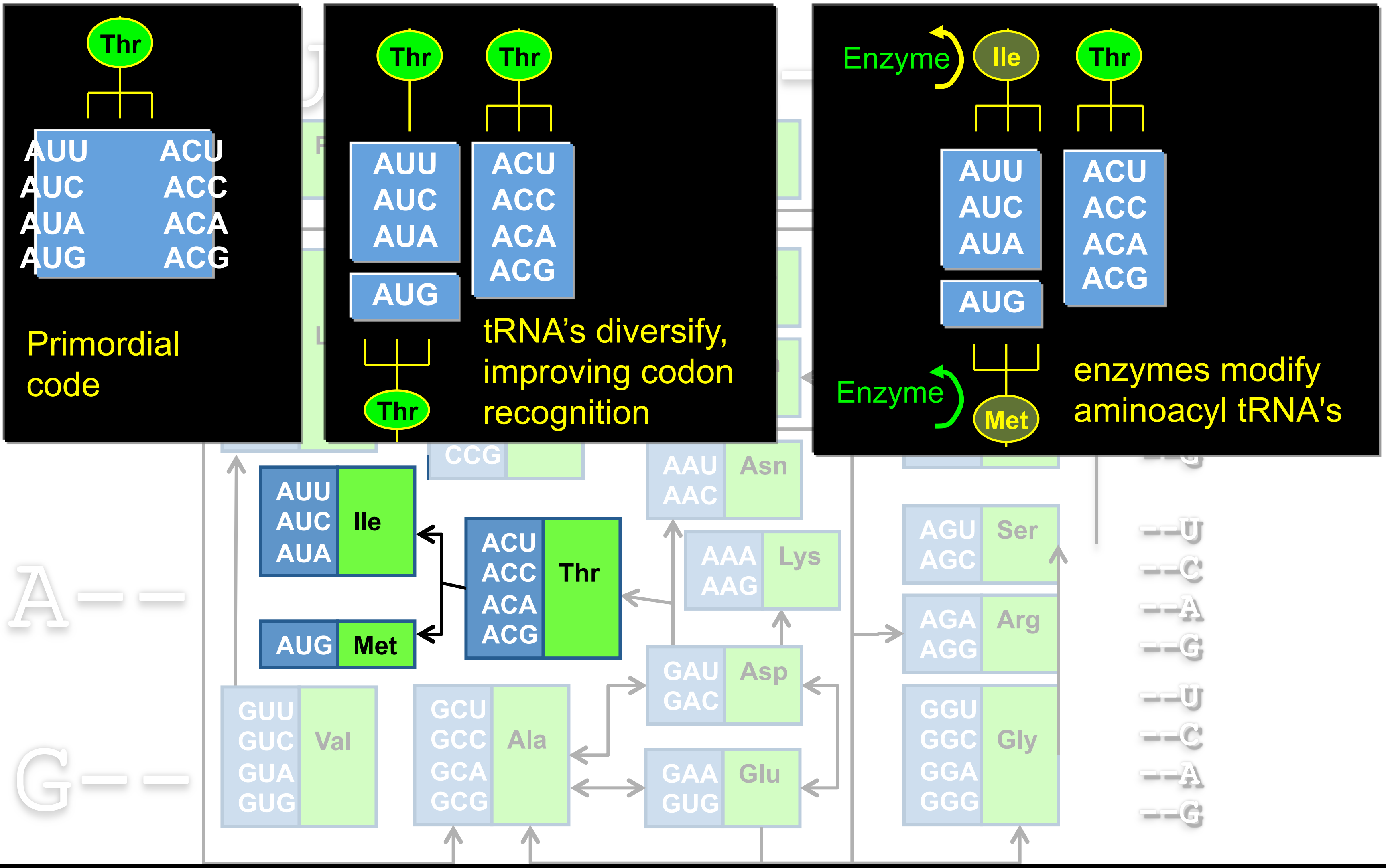
GAU GAC Asp

GAA GAG Glu

GGU GGC GGA GGG Gly

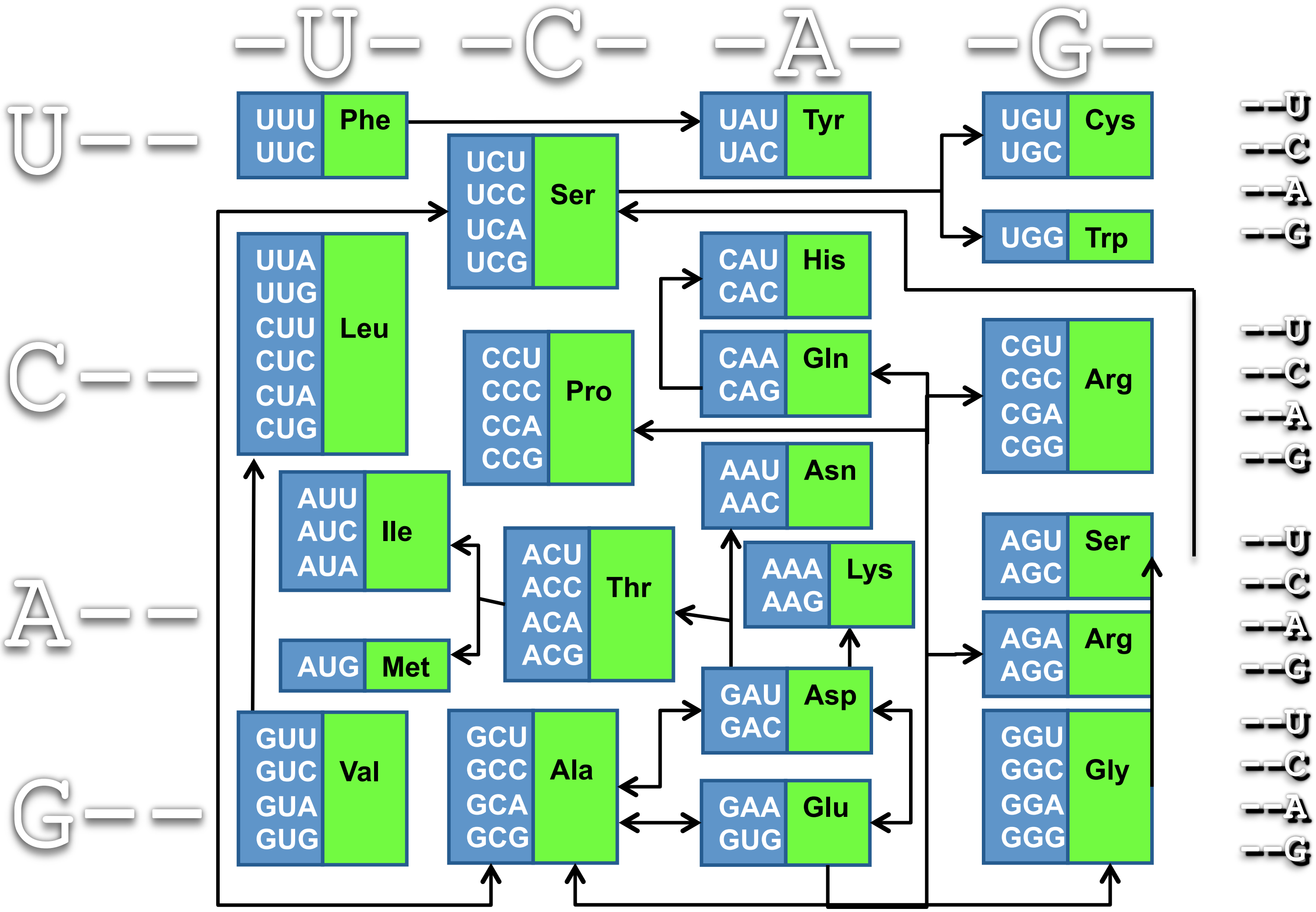
– –U
– –C
– –A
– –G

Amino acid metabolism & codon assignments



amino acid biosynthetic derivatives may have become incorporated into the code by modifying precursor amino acids on the tRNA ~'novel' amino acids would thus capture a sub-set of the codons belonging to their pre-cursors

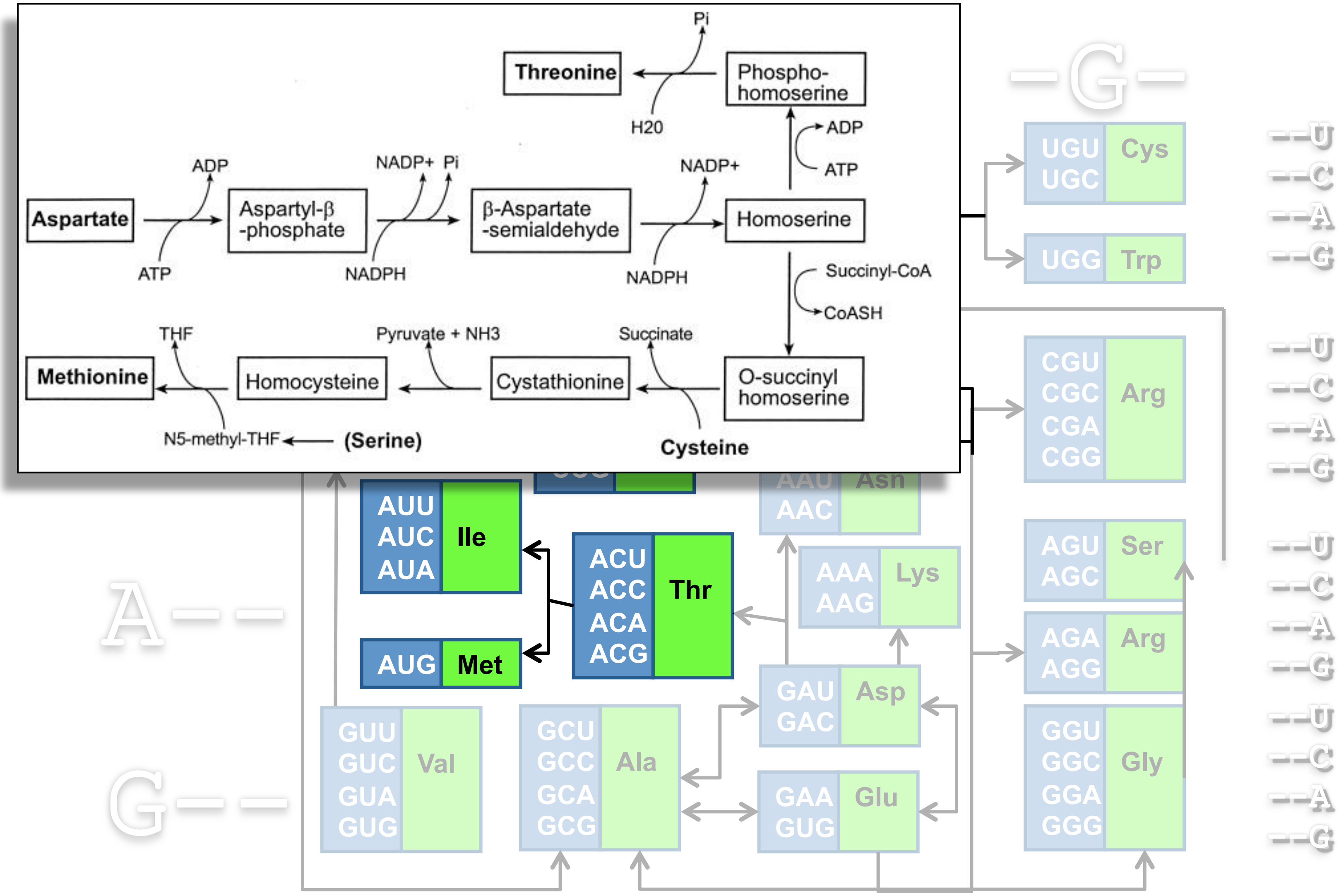
The Genetic Code “coevolution” hypothesis



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

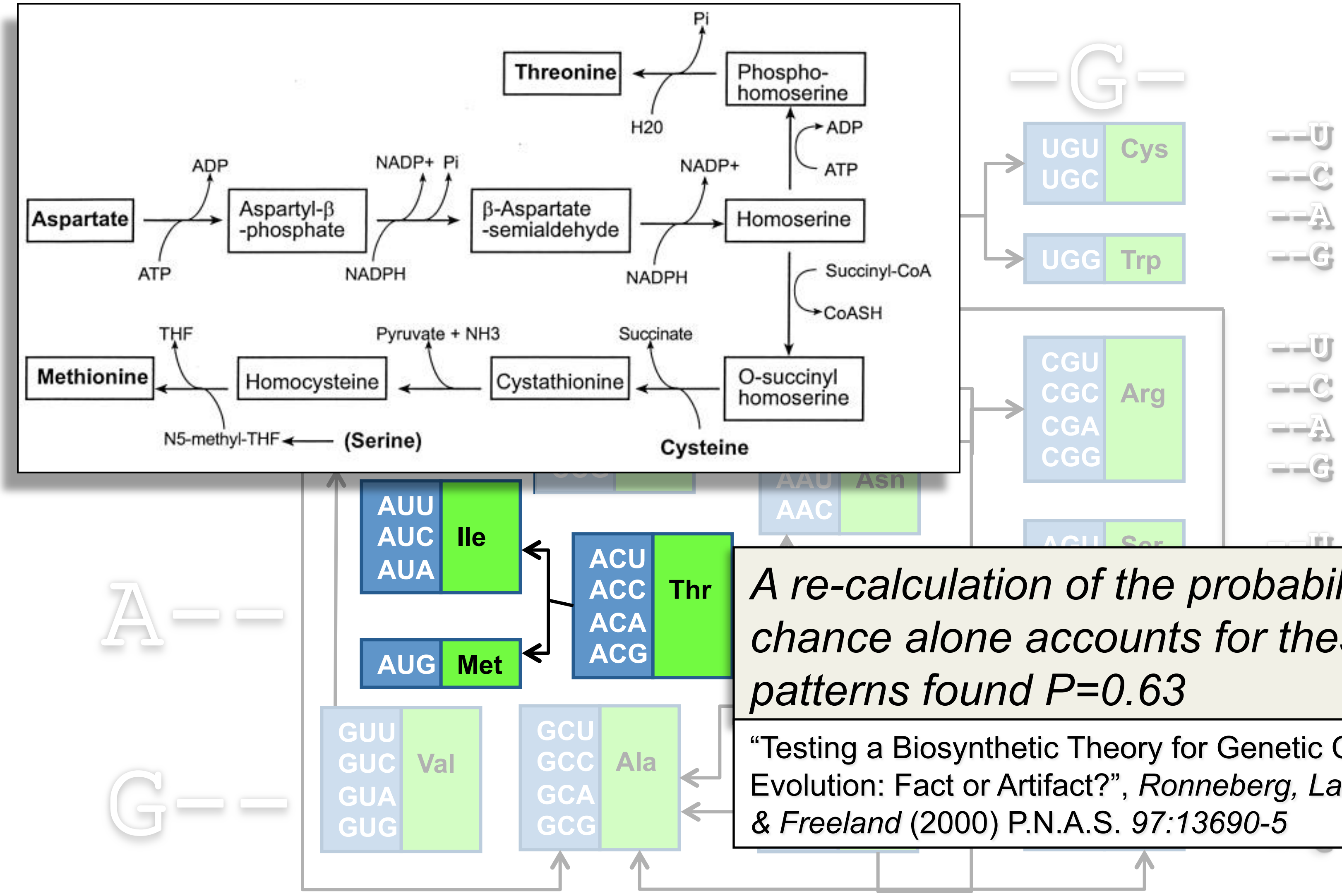
Wong’s (1975) analysis calculated a very low probability that this many precursor-product pairs would fit the expected pattern by chance alone

but it’s not that straightforward…



Several *‘precursors’* don’t actually precede their *‘products’* in extant metabolic pathways: rather, both amino acids are derived from a common, non-protein, precursor.

but it's not that straightforward…



A re-calculation of the probability that chance alone accounts for these patterns found P=0.63

"Testing a Biosynthetic Theory for Genetic Code Evolution: Fact or Artifact?", *Ronneberg, Landweber & Freeland* (2000) P.N.A.S. *97:13690-5*

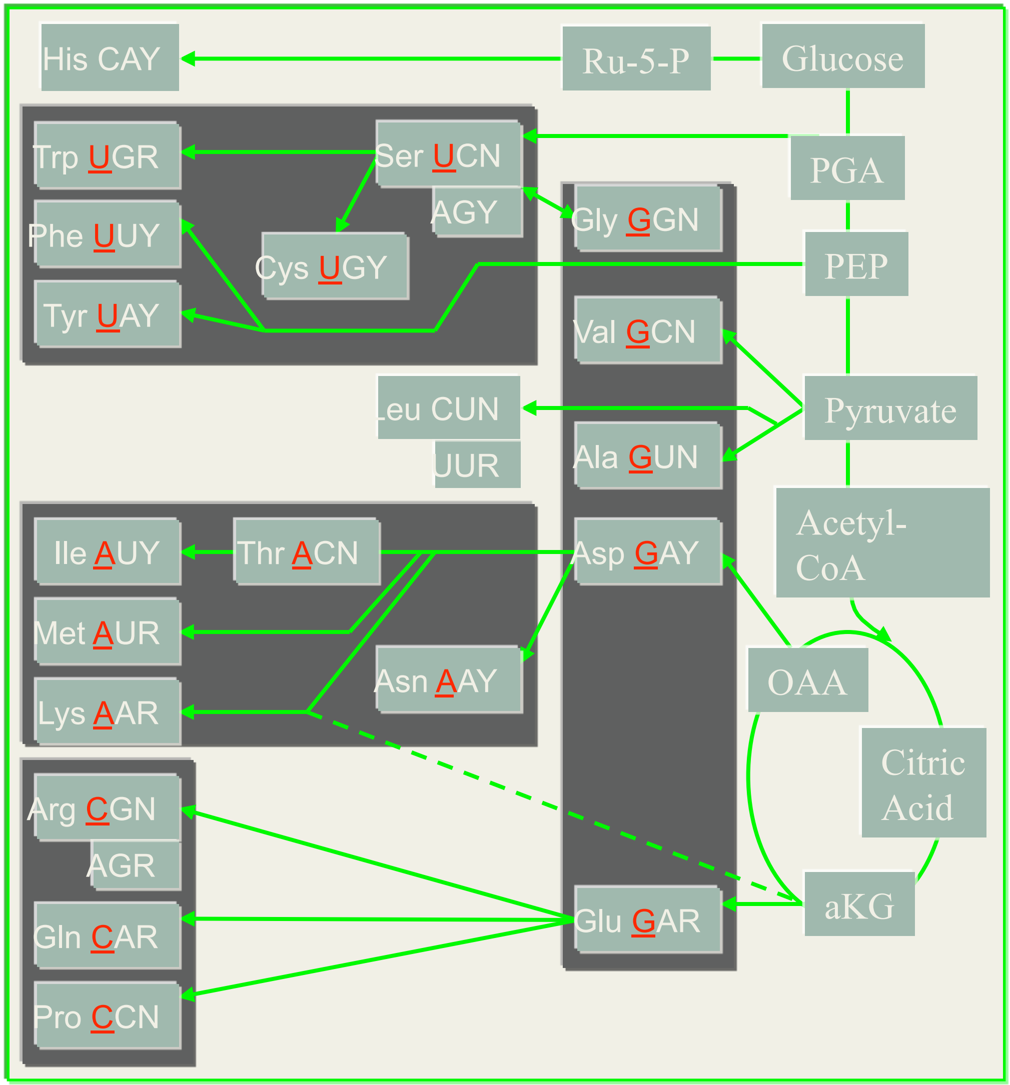
Several *'precursors'* don't actually precede their *'products'* in extant metabolic pathways: rather, both amino acids are derived from a common, non-protein, precursor.

A Consensus Incorporation Order

- Trifonov has compiled more than 60 incorporation orders from various genetic code studies and derived several “consensus” orders *(2000, 2004).*
- These orders mainly reflect the idea that small, chemically simple amino acids were encoded before big, complex amino acids
- According to Trifonov, the following one “has a merit of a best guess” based on computer analysis *(Table V in Trifonov, 2004)*:

  <u>G, A, D, V, P, S, E, T, L, R, N, I, Q, H, K, C, F, Y, M, W</u>

Simple, general patterns of biosynthetic relatedness



His CAY
Ru-5-P
Glucose
Trp UGR
Ser UCN
AGY
PGA
Phe UUY
Gly GGN
Cys UGY
PEP
Tyr UAY
Val GCN
Leu CUN
UUR
Pyruvate
Ala GUN
Ile AUY
Thr ACN
Asp GAY
Acetyl-CoA
Met AUR
OAA
Asn AAY
Lys AAR
Citric Acid
Arg CGN
AGR
aKG
Glu GAR
Gln CAR
Pro CCN

~ Amino acids from the same biosynthetic pathway tend to be assigned codons with the same first base.

~ The most likely prebiotic amino acids are all assigned to GNN codons

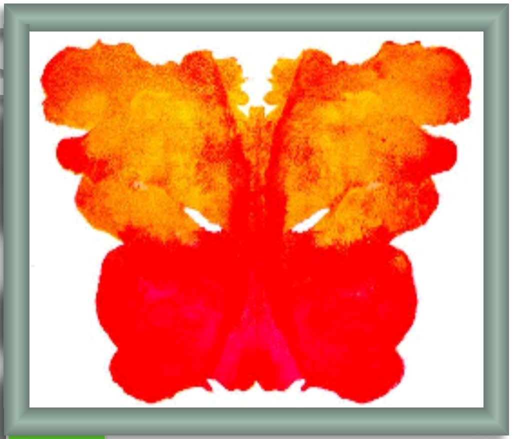
Taylor & Coates (1989) *BioSystems* 22, 177 - 187.

Miseta, A. (1989) *Physiol. Chem. Phys. Med. NMR* 21, 237 - 242.

The significance of such patterns remains unknown!

We must be careful not to overinterpret!

	U		C		A			
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	
	UUC		UCC		UAC		UGC	
U	UUA	Leu	UCA		UAA		UGA	
	UUG		UCG		UAG		UGG	
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC		CCC		CAC		CGC	
C	CUA		CCA		CAA	Gln	CGA	
	AUC	Ile	ACC	Thr	AAC		AGC	er
A	AUA		ACA		AAA		AGA	g
	GUC	Val	GCC	Ala	CAC		GGC	Gly
G	GUA		GCA		CAA	Glu	GGA	
	GUG		GCG		CAG		GGG	

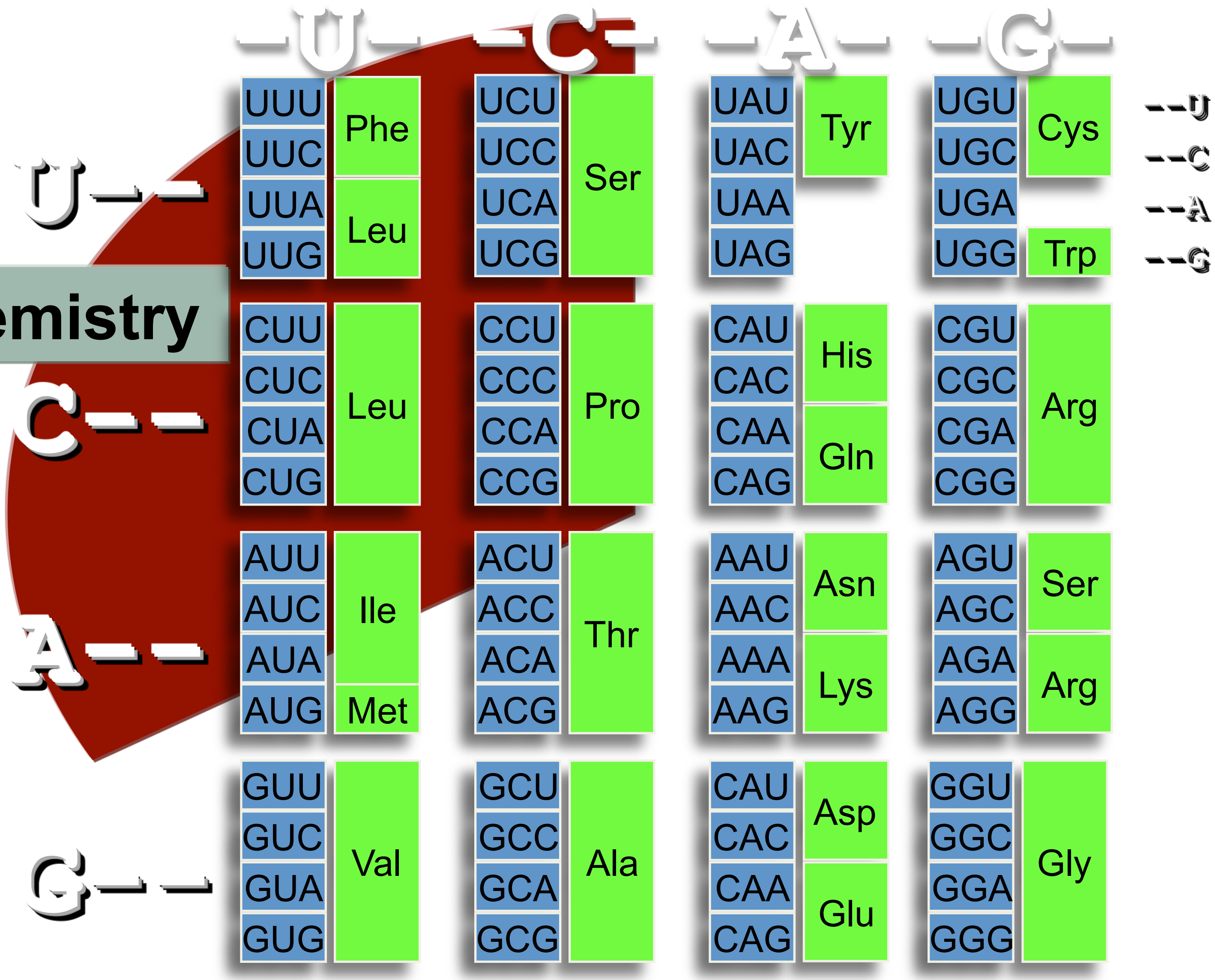


How did such different chemical languages start interacting?

Why did these particular amino acids become involved?

Several *'precursors'* don't actually precede their *'products'* in extant metabolic pathways: rather, both amino acids are derived from a common, non-protein, precursor.

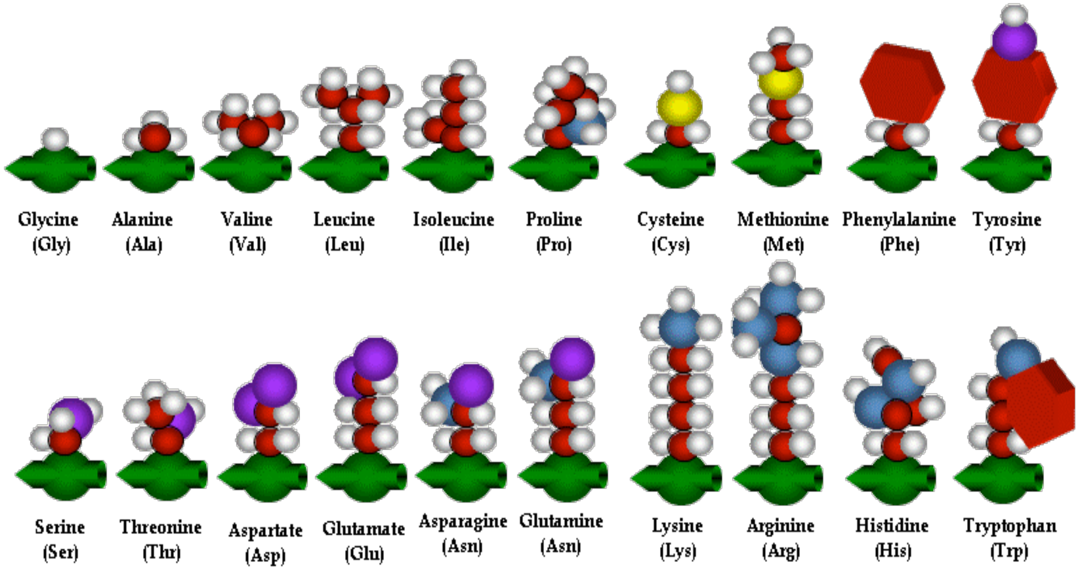
A stereochemically determined genetic code

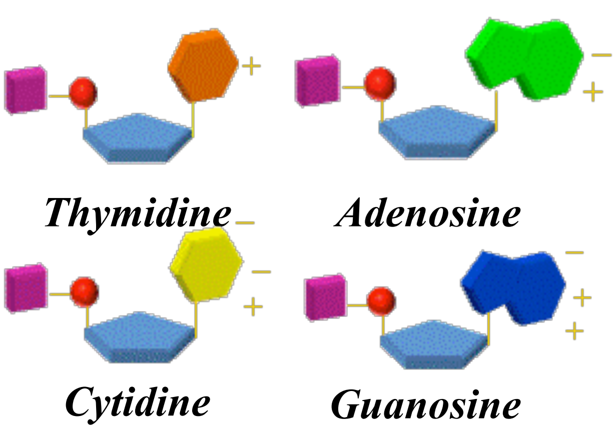


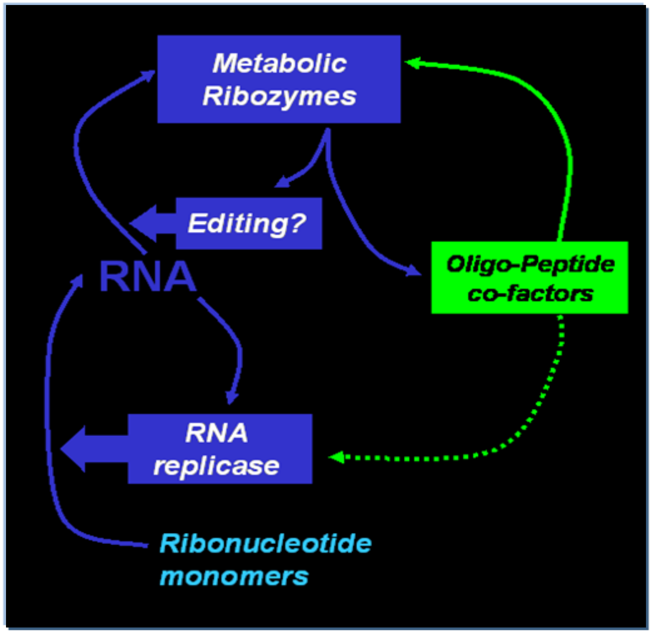
Chemistry

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

Amino acids and RNA evolved to interact







Metabolic Ribozymes

Editing?

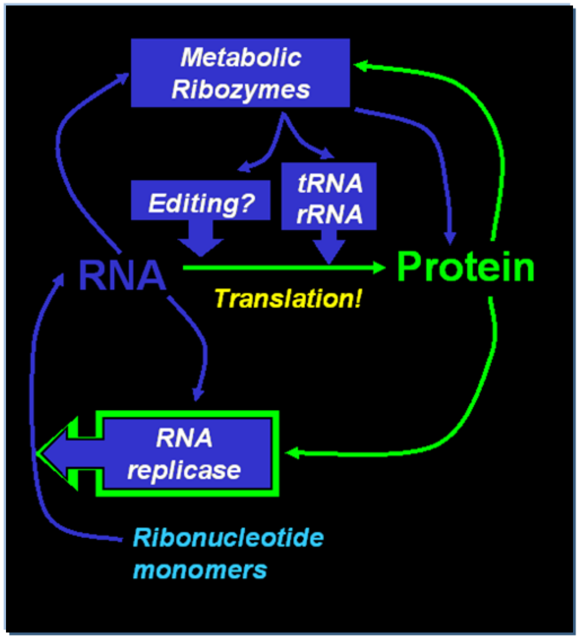
RNA

Oligo-Peptide co-factors

RNA replicase

Ribonucleotide monomers

Implies the evolution of some sort of chemical interaction



Metabolic Ribozymes

Editing?

tRNA rRNA

RNA

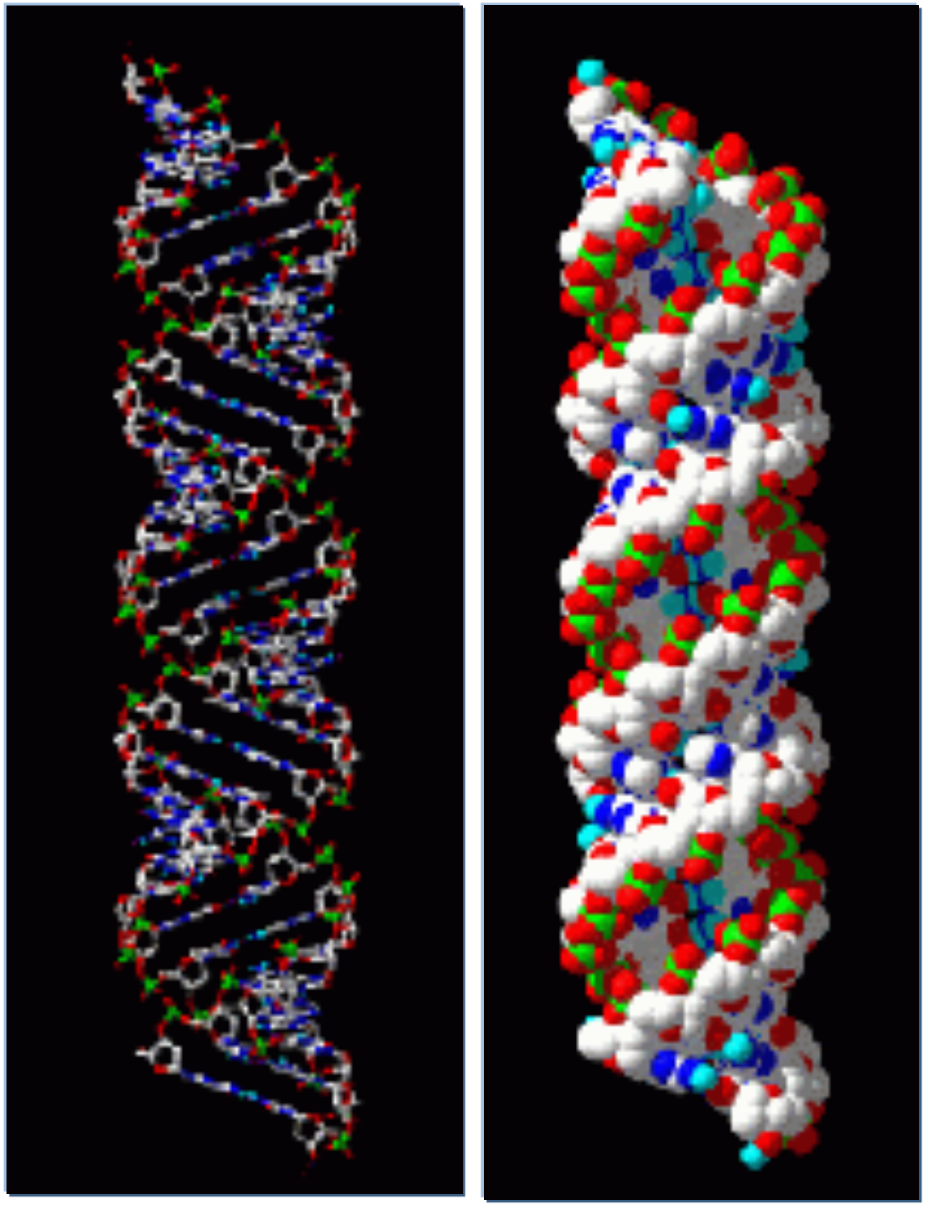
Protein

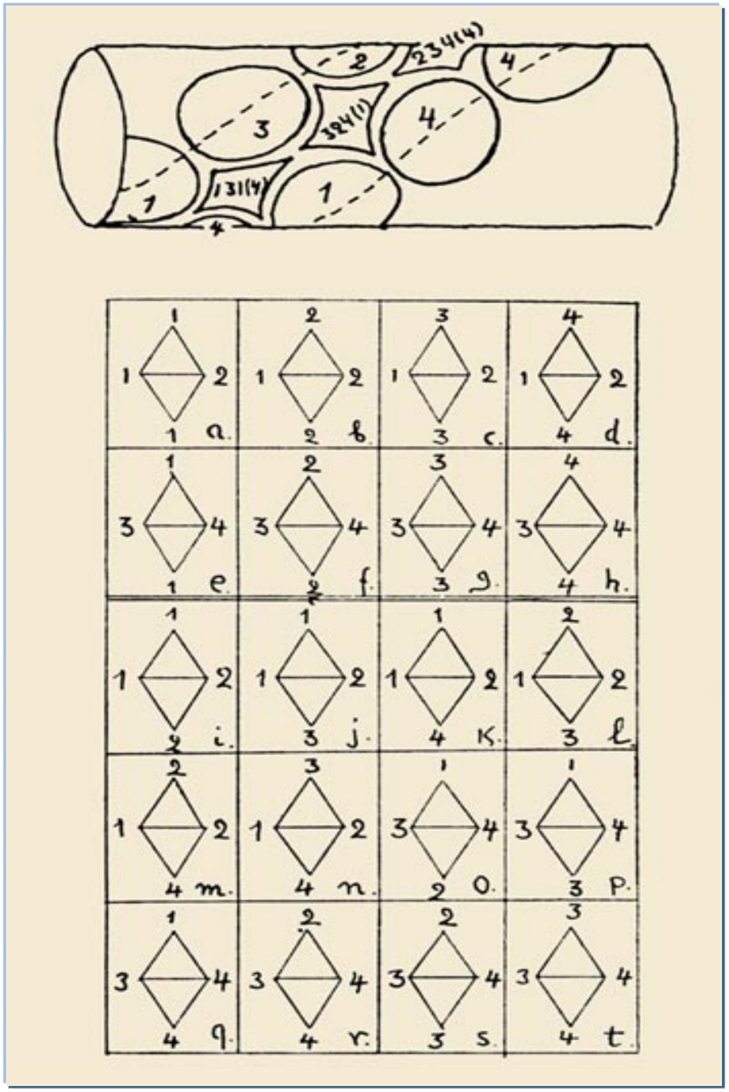
Translation!

RNA replicase

Ribonucleotide monomers

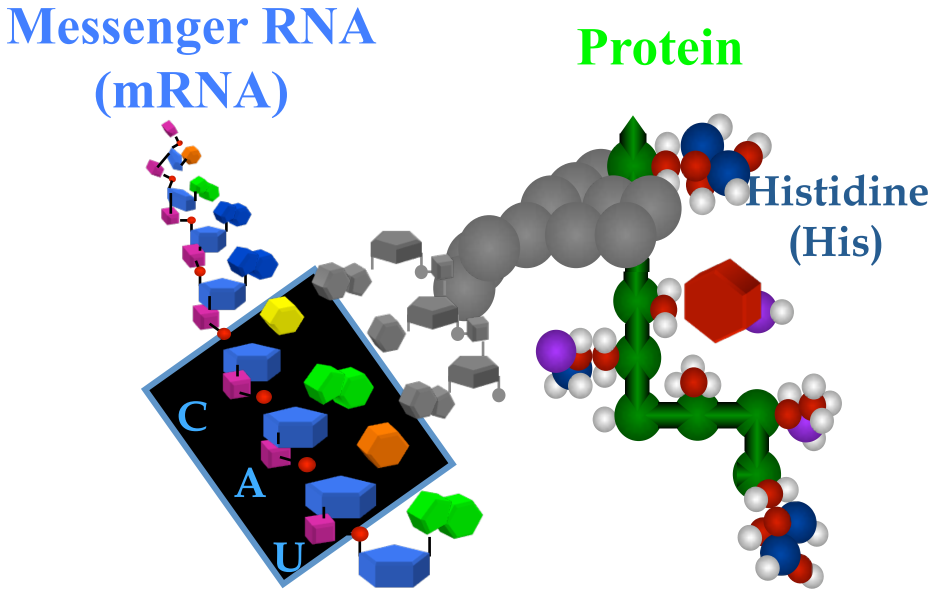
(old ideas for steric “fits”)





e.g. George Gamow’s “diamond code”

Molecular models of steric “matches”



Models ‘prove’ that amino acids fit sterically with:

Dunhill (1966) tRNA anticodons

Melcher (1974) anti-codon doublet

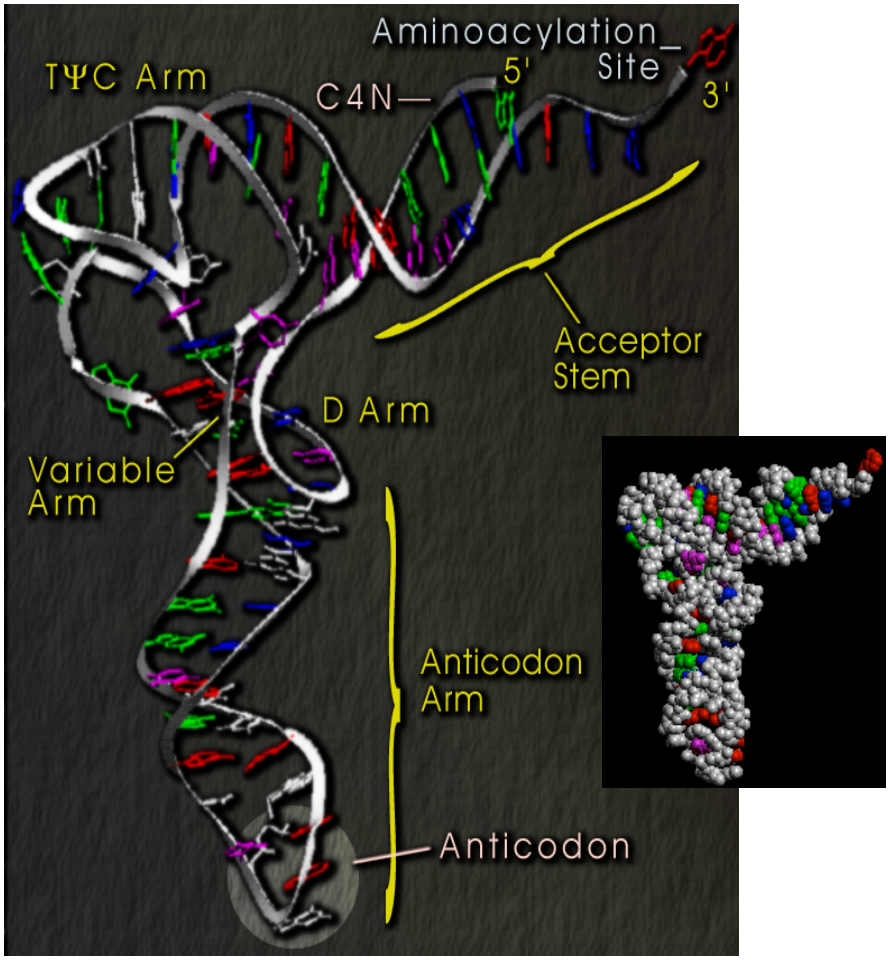
Hendry et al. (1979) double-stranded RNA

Hendry et al. (1981) cavity in B-DNA

Shimizu (1982) C4N complex

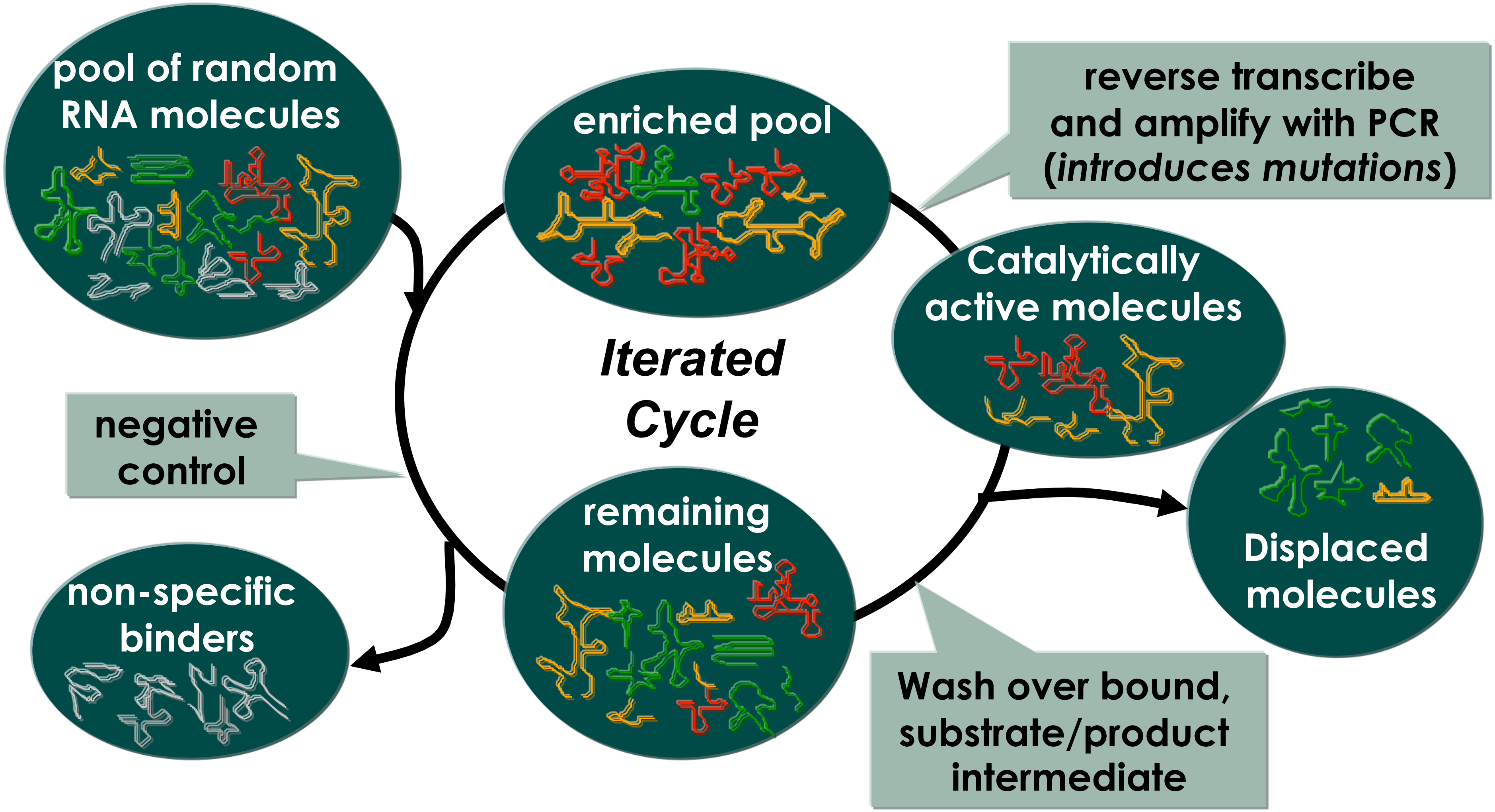
Root-Bernstein (1982) ‘backwards’ codons

…and many more!



~ Model building is not a reliable approach!

In Vitro selection of *ribozymes*



pool of random RNA molecules

enriched pool

reverse transcribe and amplify with PCR (*introduces mutations*)

Catalytically active molecules

Iterated Cycle

negative control

Displaced molecules

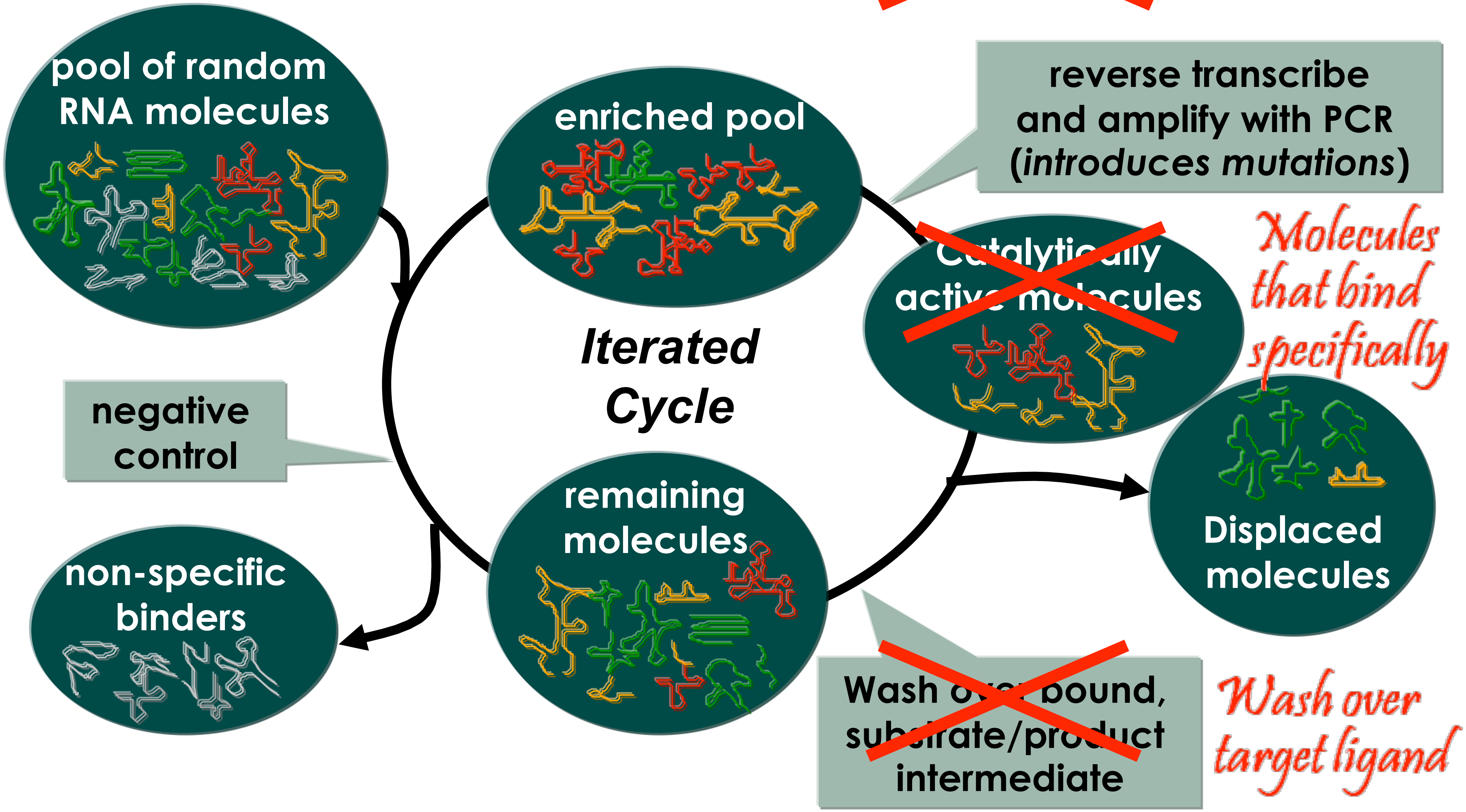
non-specific binders

remaining molecules

Wash over bound, substrate/product intermediate

Laboratory experiments have now created RNA molecules that perform a wide range of basic metabolic (catalytic) functions: these are ribozymes

In Vitro selection of ~~*ribozymes*~~ *aptamers*



pool of random RNA molecules

enriched pool

reverse transcribe and amplify with PCR (*introduces mutations*)

~~Catalytically active molecules~~

Molecules that bind specifically

Iterated Cycle

negative control

Displaced molecules

remaining molecules

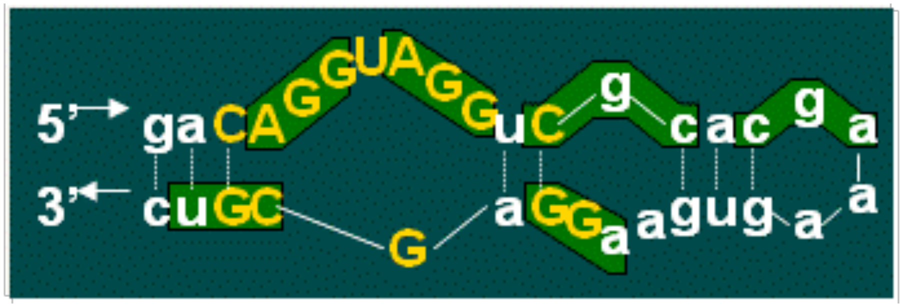
non-specific binders

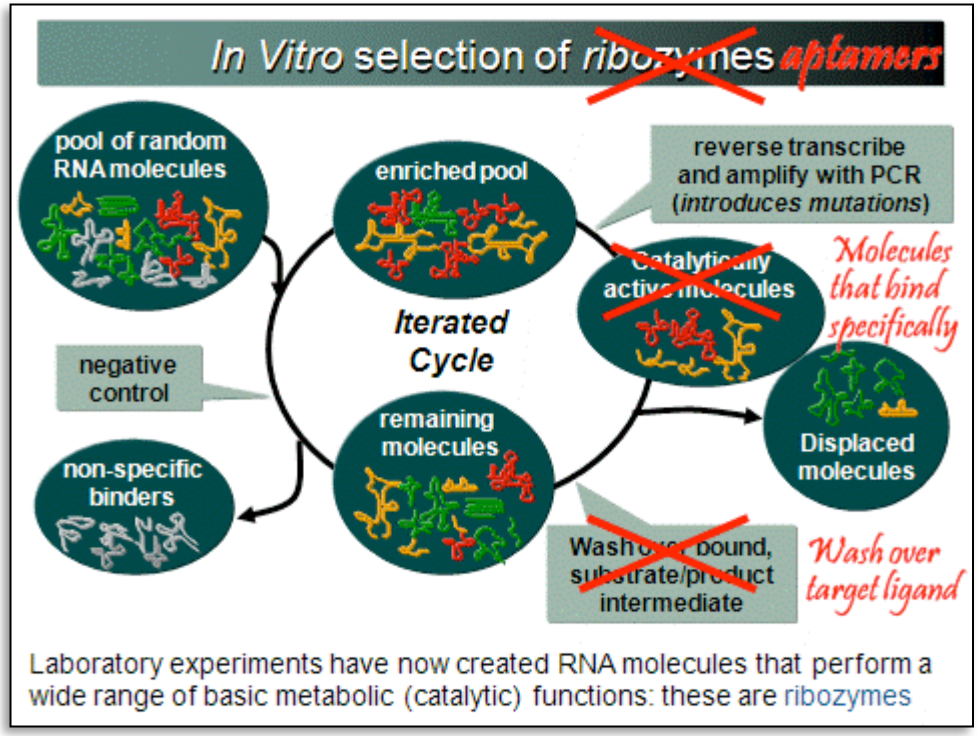
~~Wash over bound, substrate/product~~ intermediate

Wash over target ligand

Laboratory experiments have now created RNA molecules that perform a wide range of basic metabolic (catalytic) functions: these are ribozymes

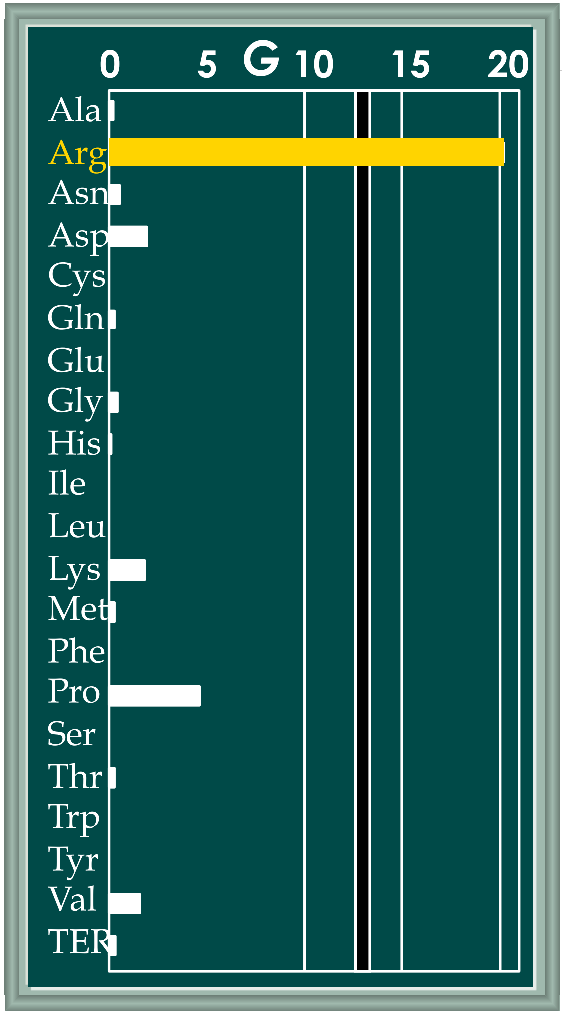
In Vitro selection of RNA aptamers for Arginine





In Vitro selection of ~~ribozymes~~ *aptamers*

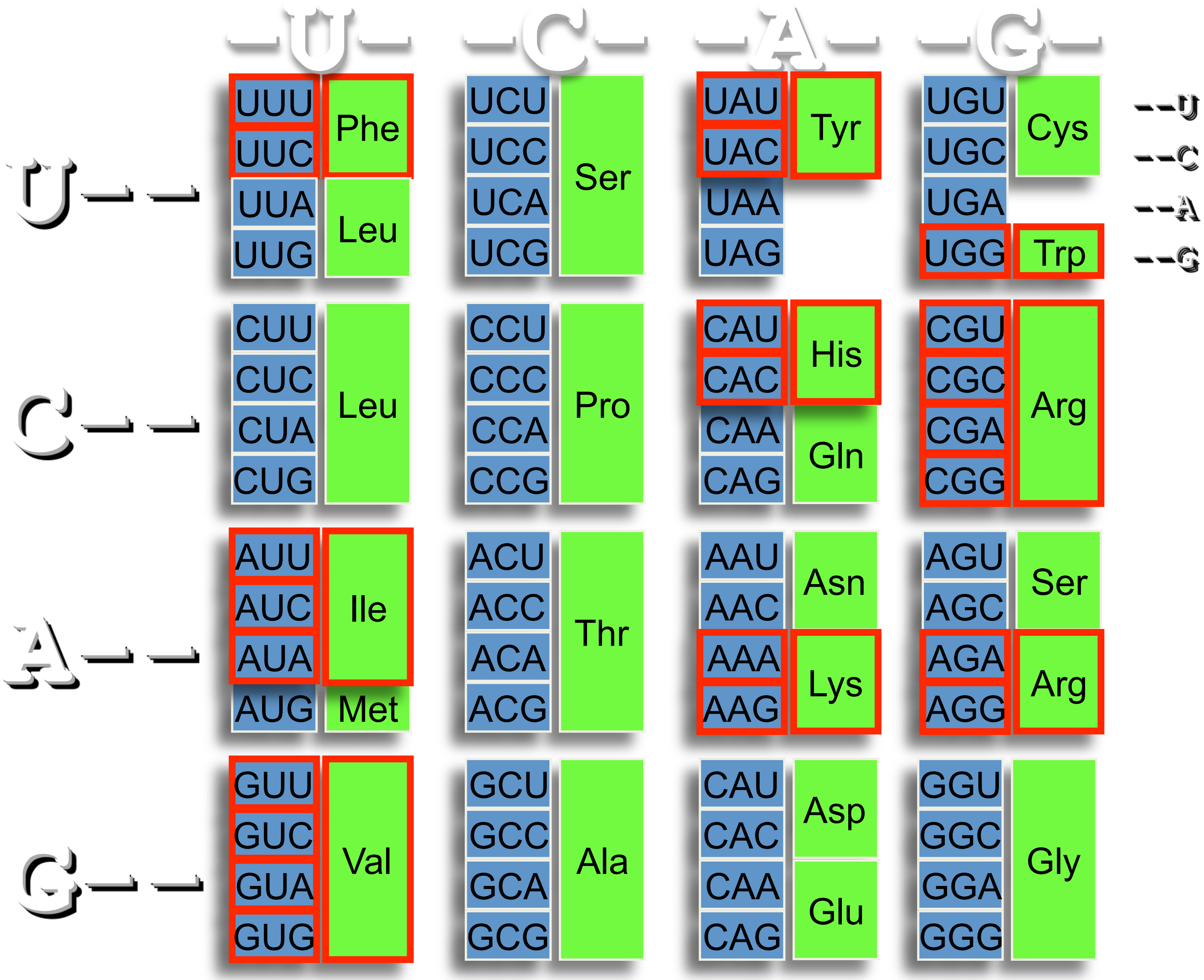
Laboratory experiments have now created RNA molecules that perform a wide range of basic metabolic (catalytic) functions: these are ribozymes



~ only Arg codons are significantly associated with Arg aptamers…

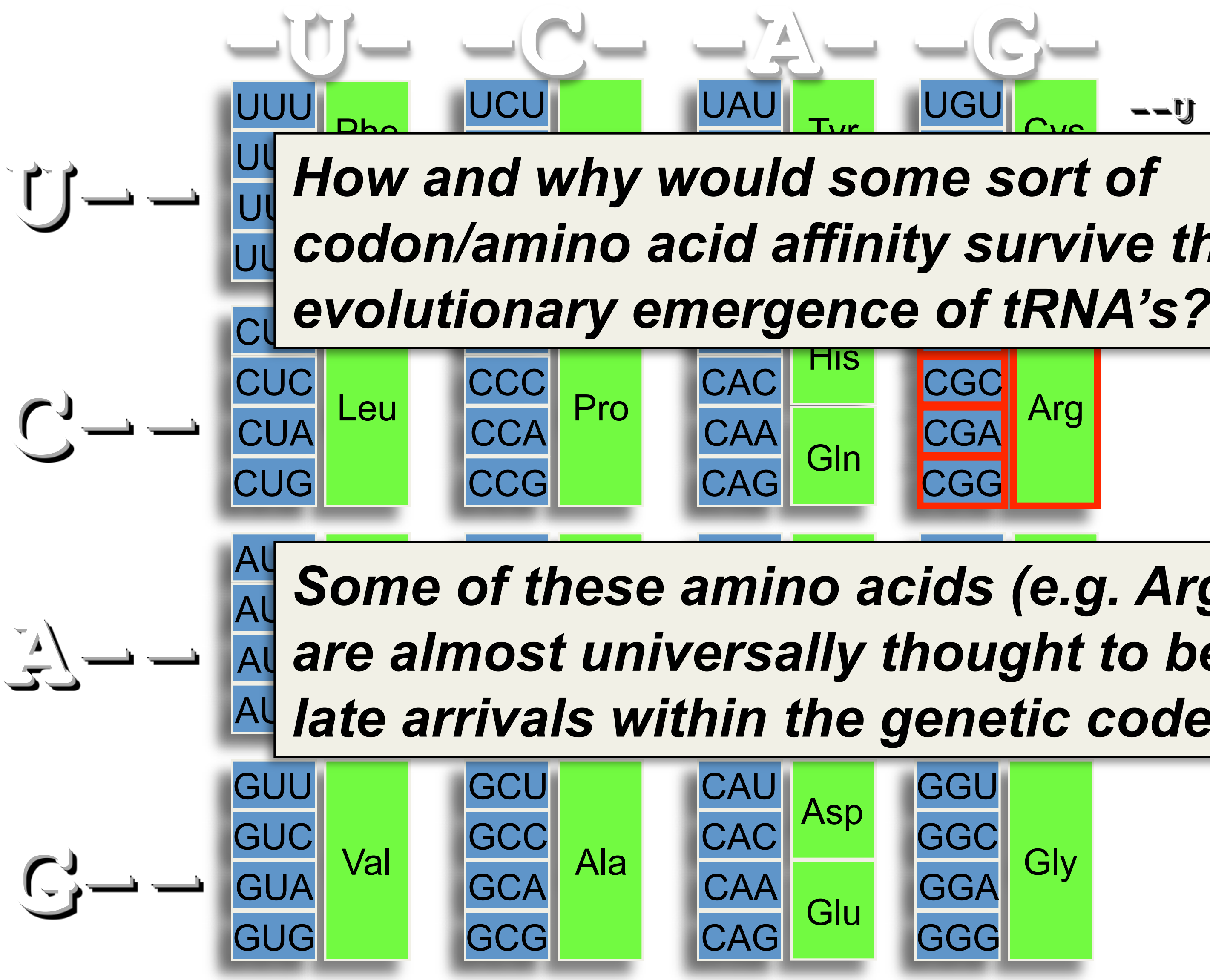
Knight and Landweber (1998) Chem. Biol. 5, pp. R215 - R220

Current state of stereochemical claims…



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

Unanswered questions…



How and why would some sort of codon/amino acid affinity survive the evolutionary emergence of tRNA's?

Some of these amino acids (e.g. Arg) are almost universally thought to be late arrivals within the genetic code.

Unanswered questions…

How and why would some sort of codon/amino acid affinity survive the evolutionary emergence of tRNA's?

Q W E R T Y

changing the meaning of a codon…

Unanswered questions…

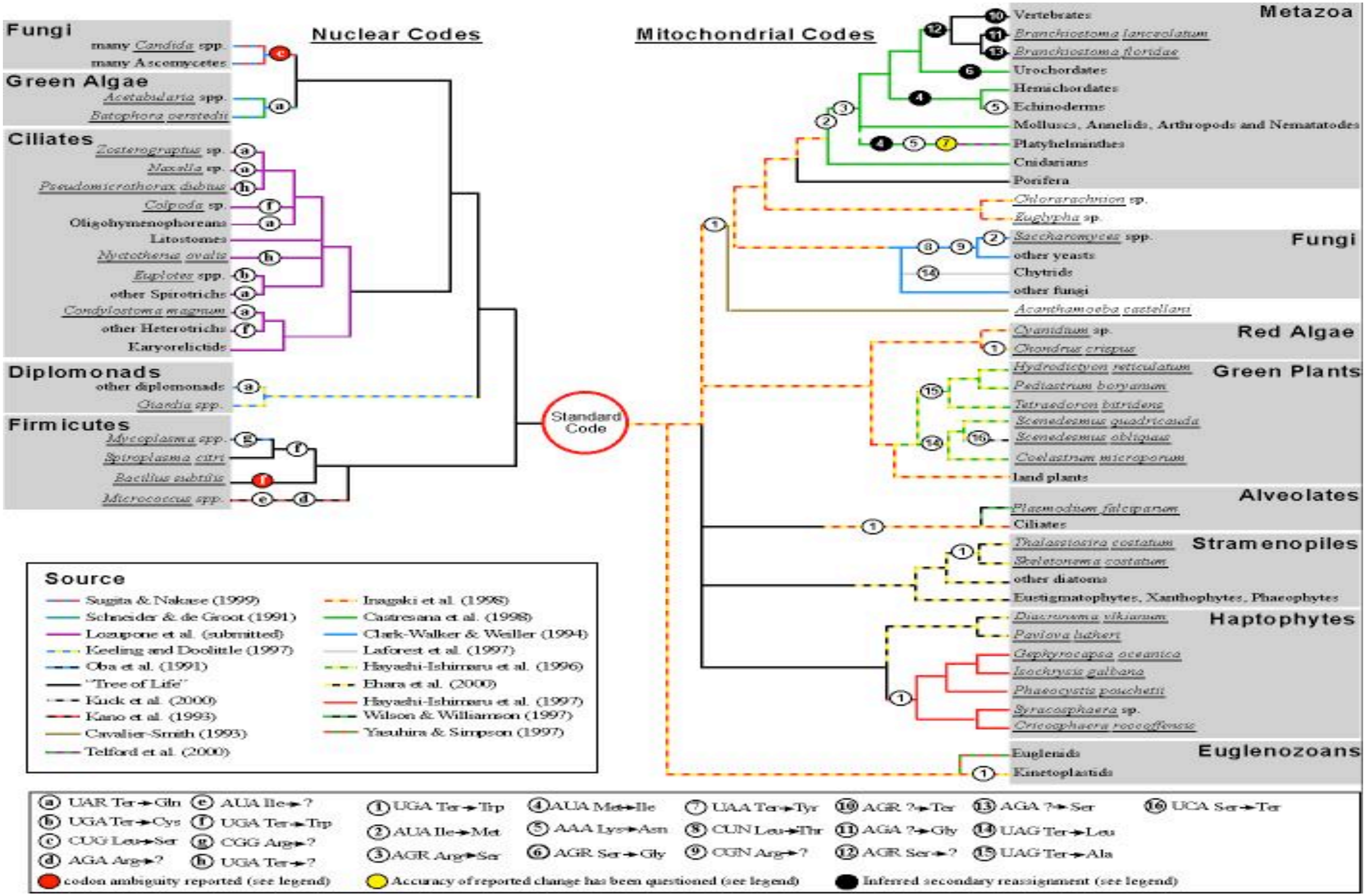
How and why would some sort of codon/amino acid affinity survive the evolutionary emergence of tRNA's?

Q W Z R T Y

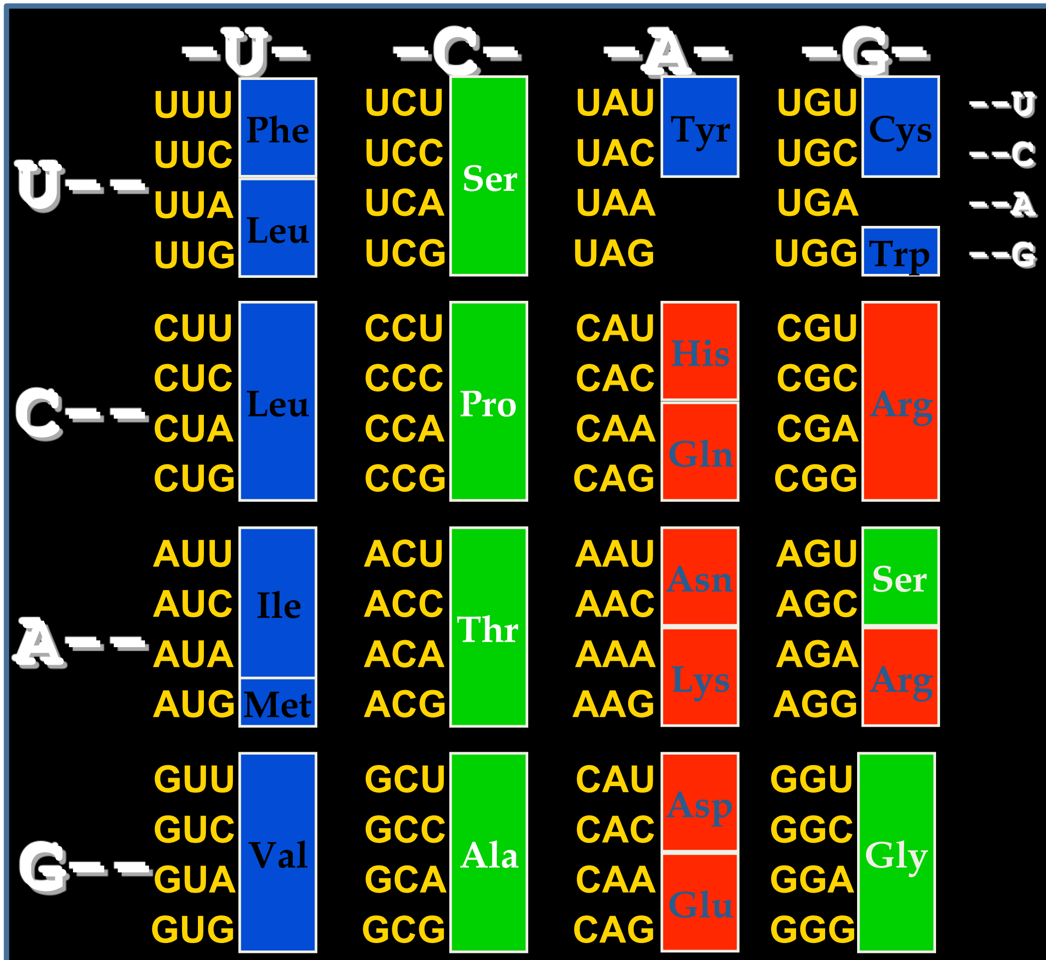
changing the meaning of a codon =
introducing countlzss mistakzs within a gznomz simultanzously.

As soon as it first appeared, the genetic code became locked into place because it is unchangeable (Crick 1968)

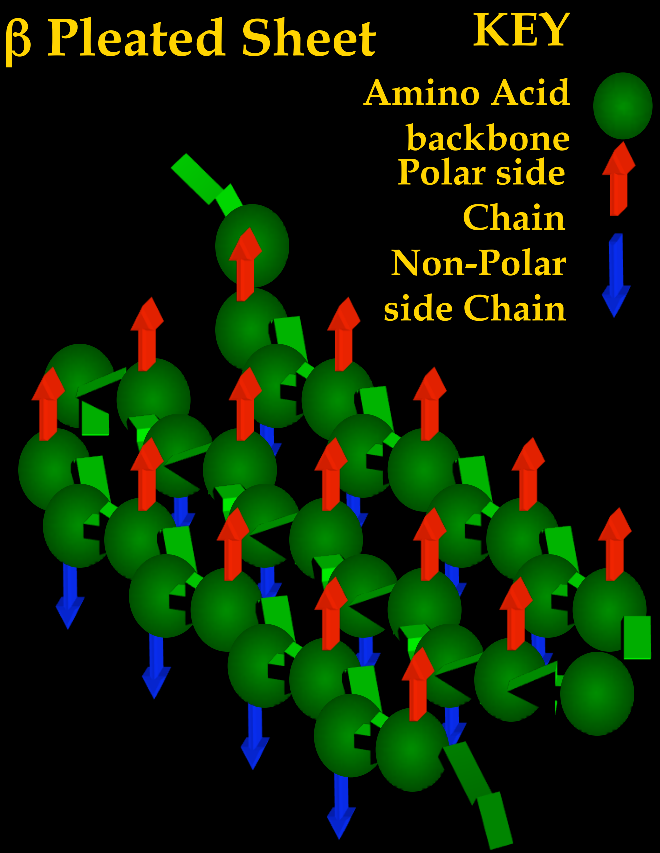
…and yet codon assignments can and do evolve



A Non-Random Genetic Code?

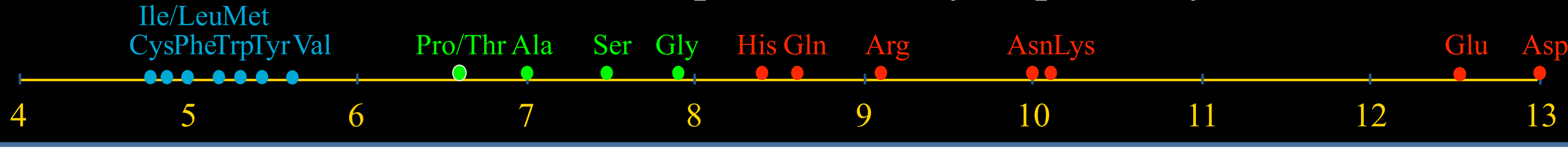


	-U-		-C-		-A-		-G-		
U--	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	--U
	UUC		UCC		UAC		UGC		--C
	UUA	Leu	UCA		UAA		UGA		--A
	UUG		UCG		UAG		UGG	Trp	--G
C--	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Gln	CGA		
	CUG		CCG		CAG		CGG		
A--	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	AUC		ACC		AAC	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



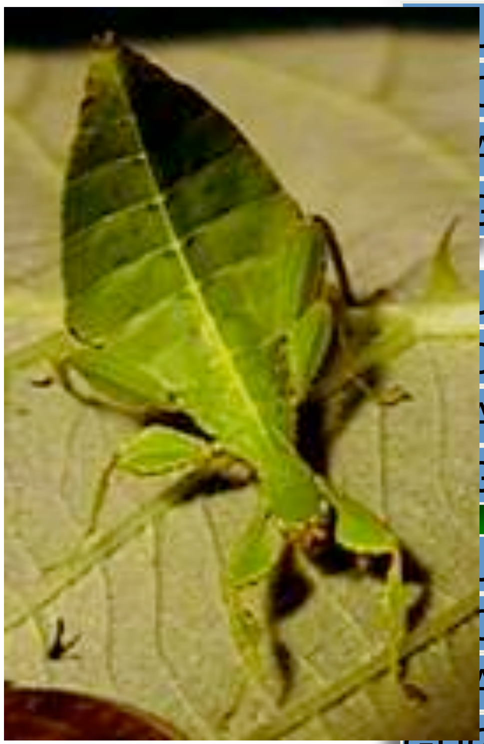
β Pleated Sheet

“Polar Requirement” (hydrophobicity)



Natural selection for an adaptive genetic code

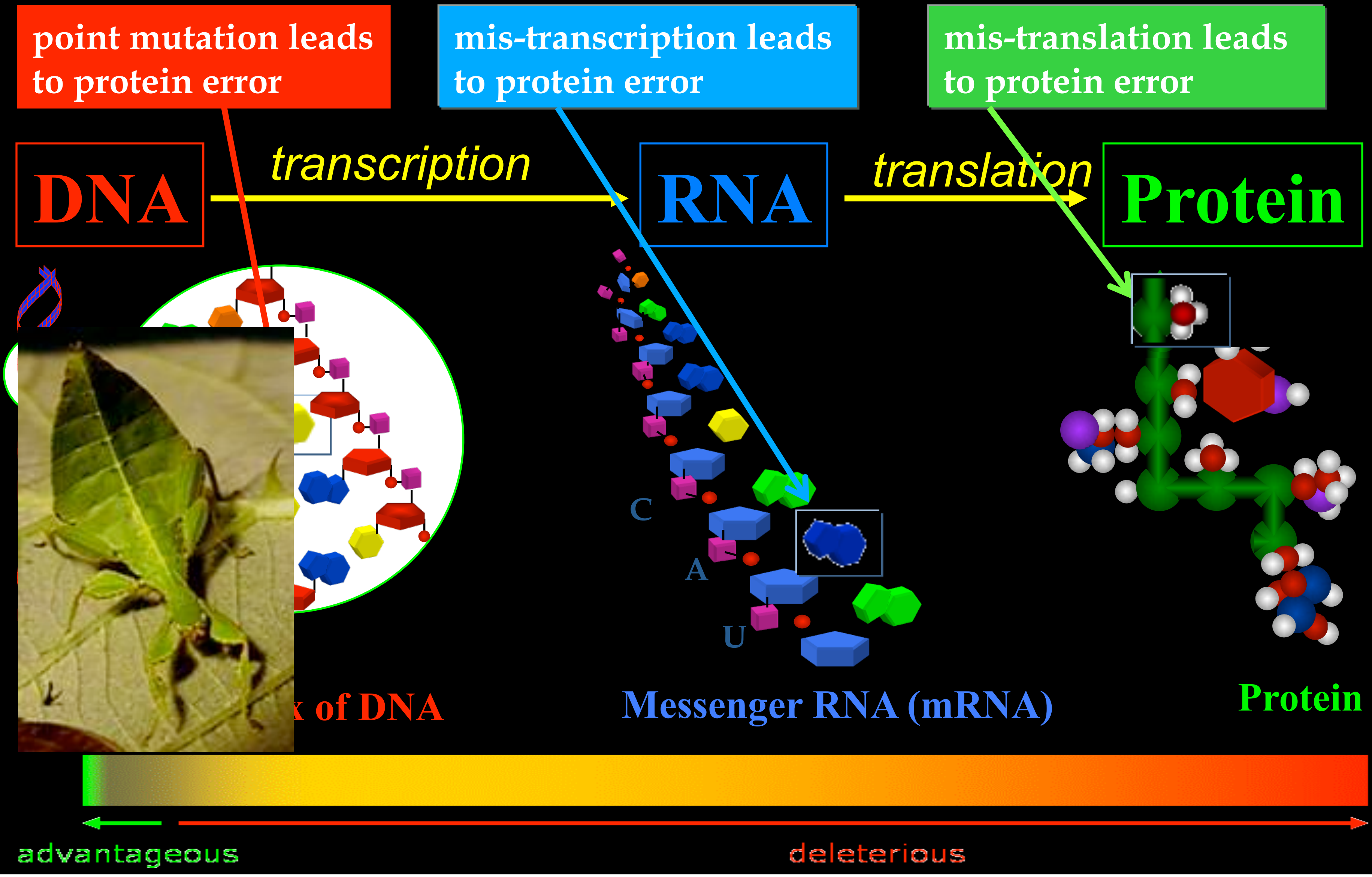
	–U–		–C–		–A–		–G–		
U– –	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	--U
	UUC		UCC		UAC		UGC		--C
	UUA	Leu	UCA		UAA		UGA		--A
	UUG		UCG		UAG		UGG	Trp	--G
	U	Leu	CCU	Pro	CAU	His	CGU	Arg	
	C		CCC		CAC		CGC		
	A		CCA		CAA	Gln	CGA		
	G		CCG		CAG		CGG		
	U	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	C		ACC		AAC		AGC		
	A		ACA		AAA	Lys	AGA	Arg	
	G	Met	ACG		AAG		AGG		
	U	Val	GCU	Ala	CAU	Asp	GGU	Gly	
	C		GCC		CAC		GGC		
	A		GCA		CAA	Glu	GGA		
	GUG		GCG		CAG		GGG		



Selection

What would make a code “good” or “bad”?

A code arranged such that genetic errors are silent, or lead to substitution of a similar amino acid will be more fit than alternatives in which this is not so

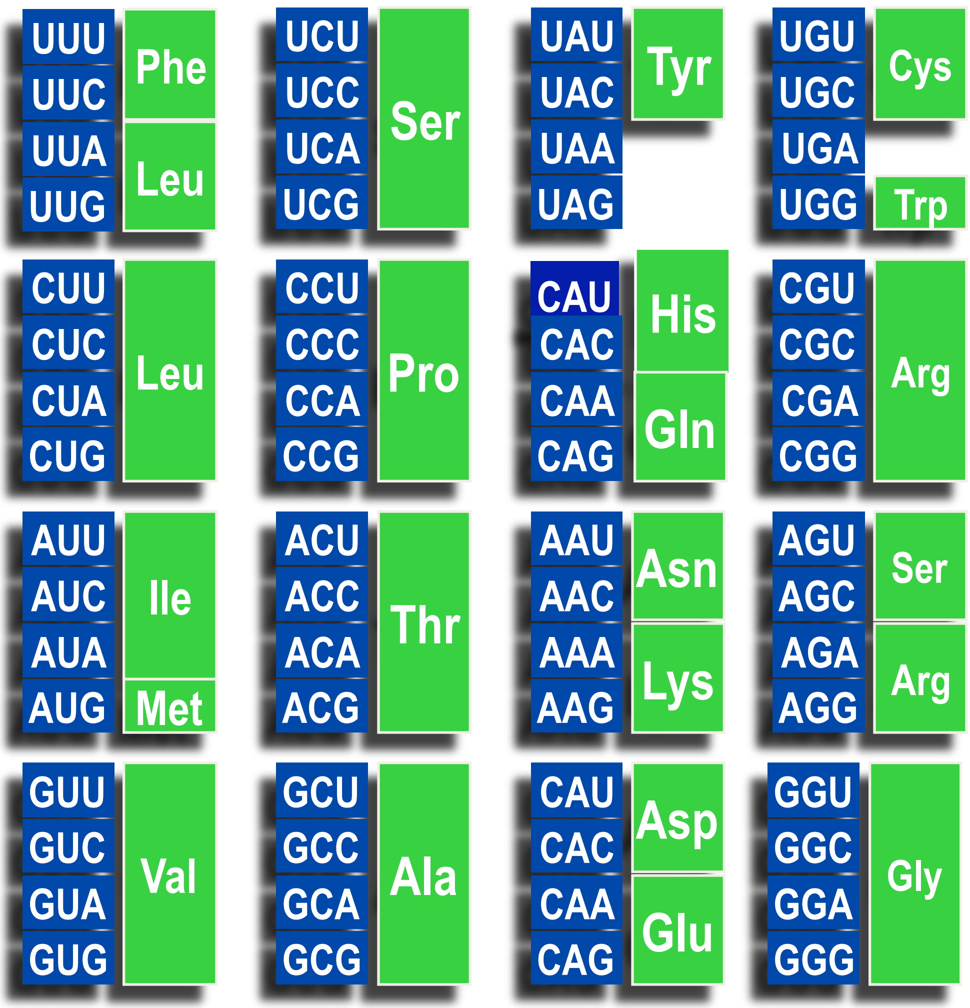


point mutation leads to protein error

mis-transcription leads to protein error

mis-translation leads to protein error

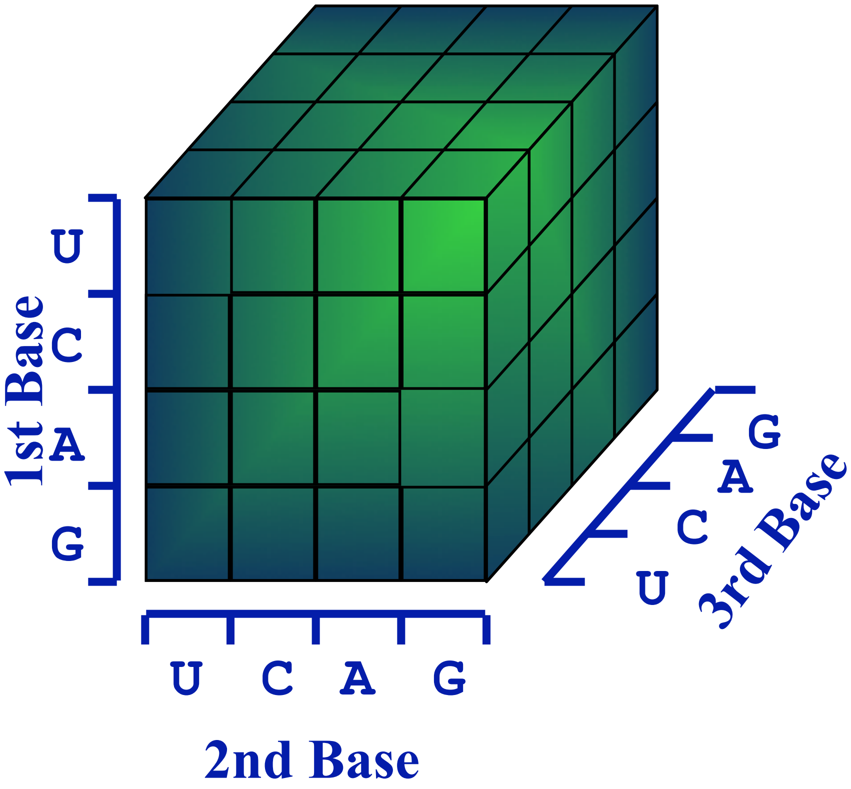
The Code as a 3D Matrix



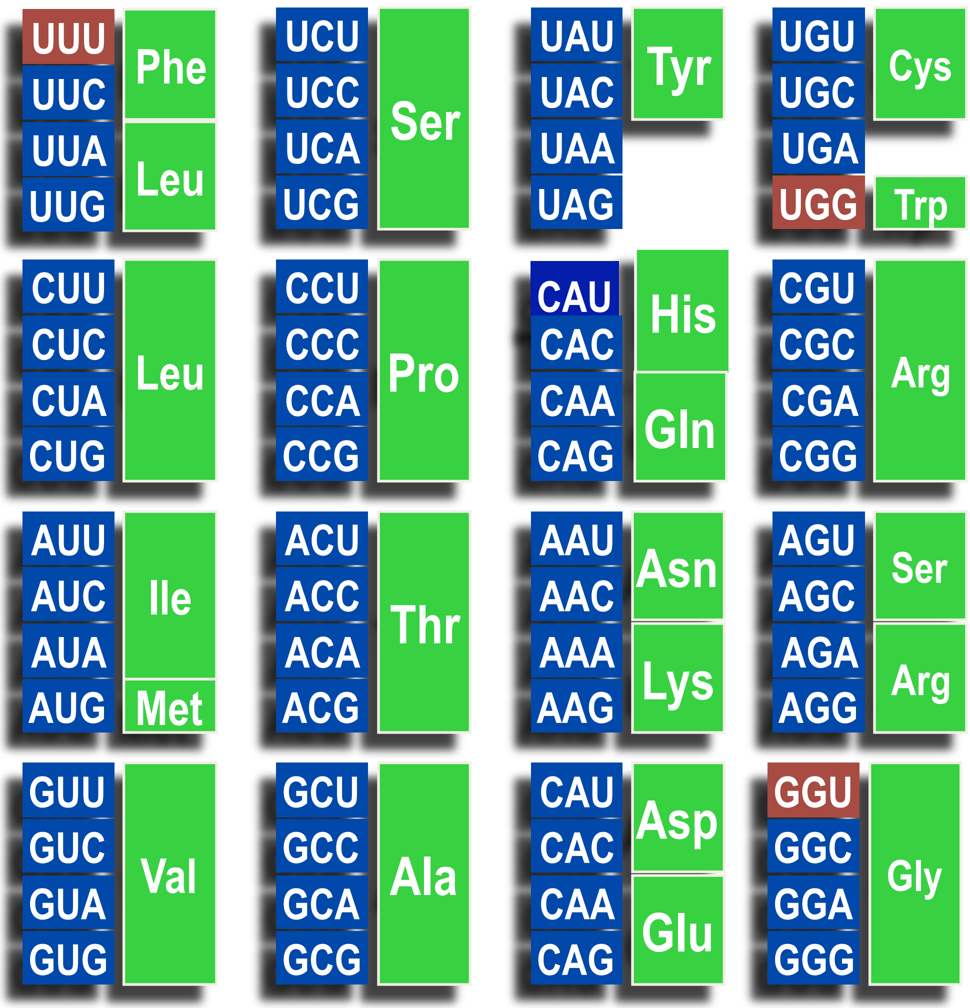
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA		UGA	
UUG		UCG		UAG		UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lys	AGA	Arg
AUG	Met	ACG		AAG		AGG	
GUU	Val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

The Code as a 3D Matrix

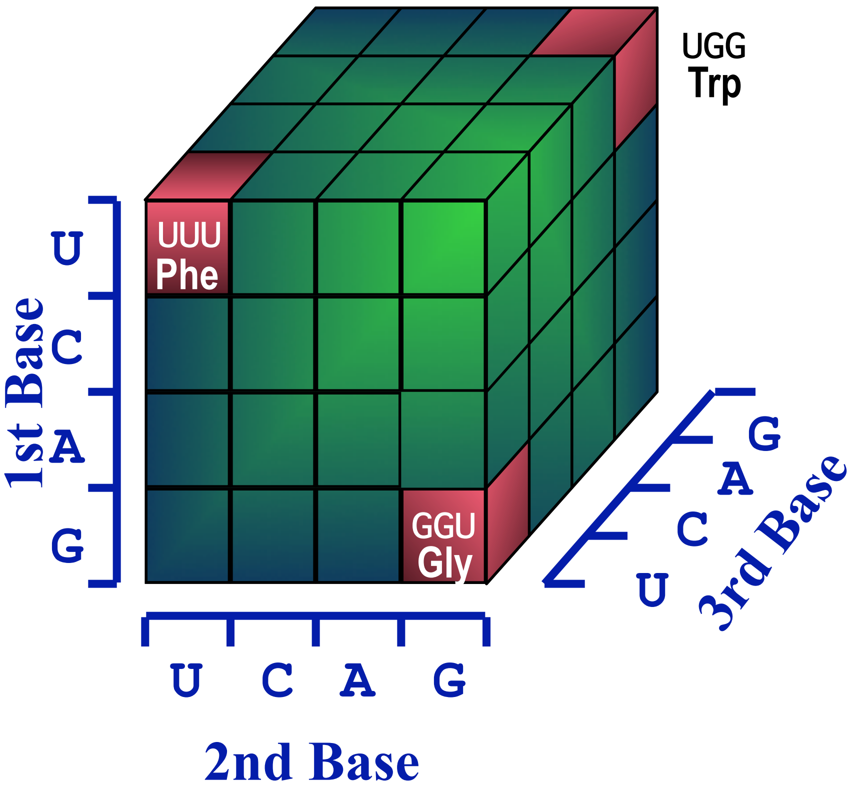
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA		UGA	
UUG		UCG		UAG		UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lys	AGA	Arg
AUG	Met	ACG		AAG		AGG	
GUU	Val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



The Code as a 3D Matrix

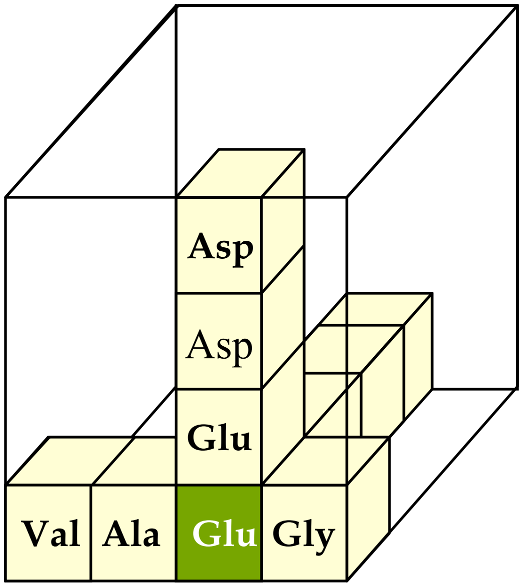


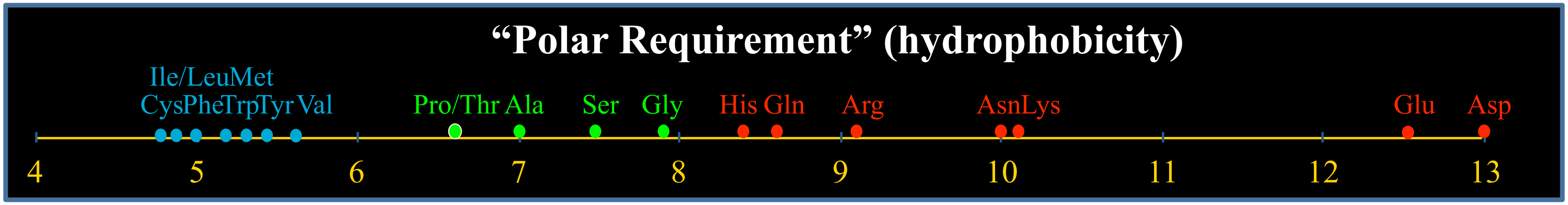
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC		UCC		UAC		UGC	
UUA	Leu	UCA		UAA		UGA	
UUG		UCG		UAG		UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lys	AGA	Arg
AUG	Met	ACG		AAG		AGG	
GUU	Val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



A codon can undergo 9 point mutations ...

Point mutations may be silent or may substitute an incorrect amino acid, leading to consequences of varying severity, depending on amino acid similarity:




“Polar Requirement” (hydrophobicity)

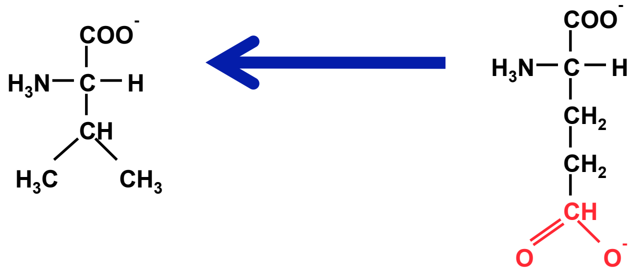
A codon can undergo 9 point mutations ...

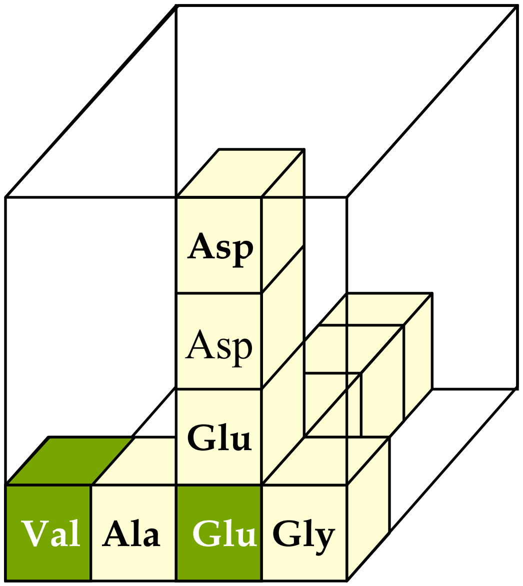
Point mutations may be silent or may substitute an incorrect amino acid, leading to consequences of varying severity, depending on amino acid similarity:

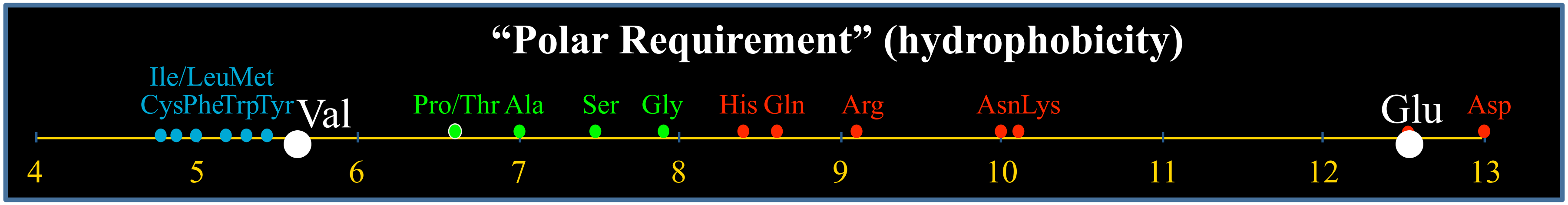
For example, sickle cell anemia results from a single point mutation, in the 6th codon of the β-chain of hemoglobin:

GUG (Val)

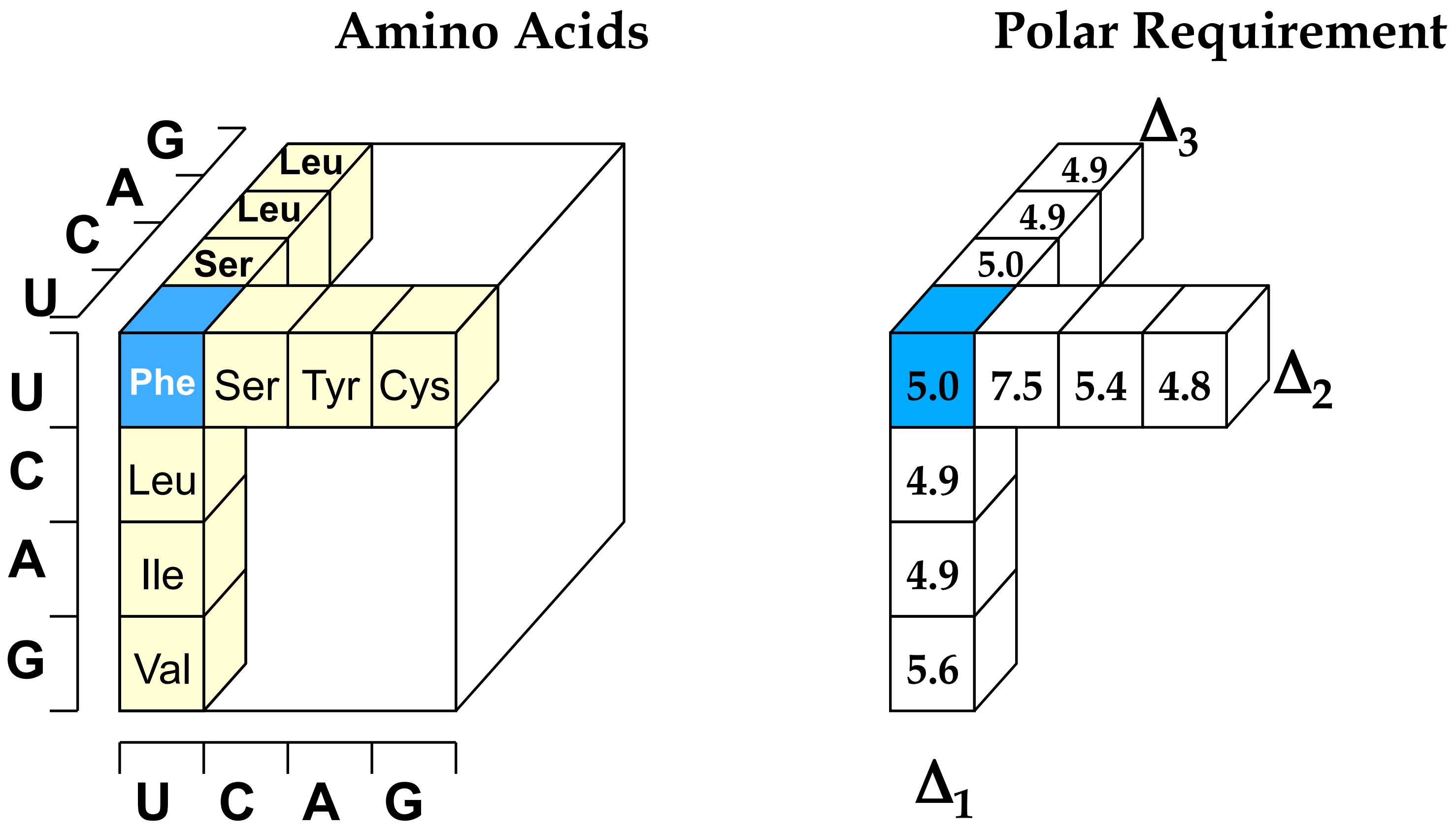
GAG (Glu)





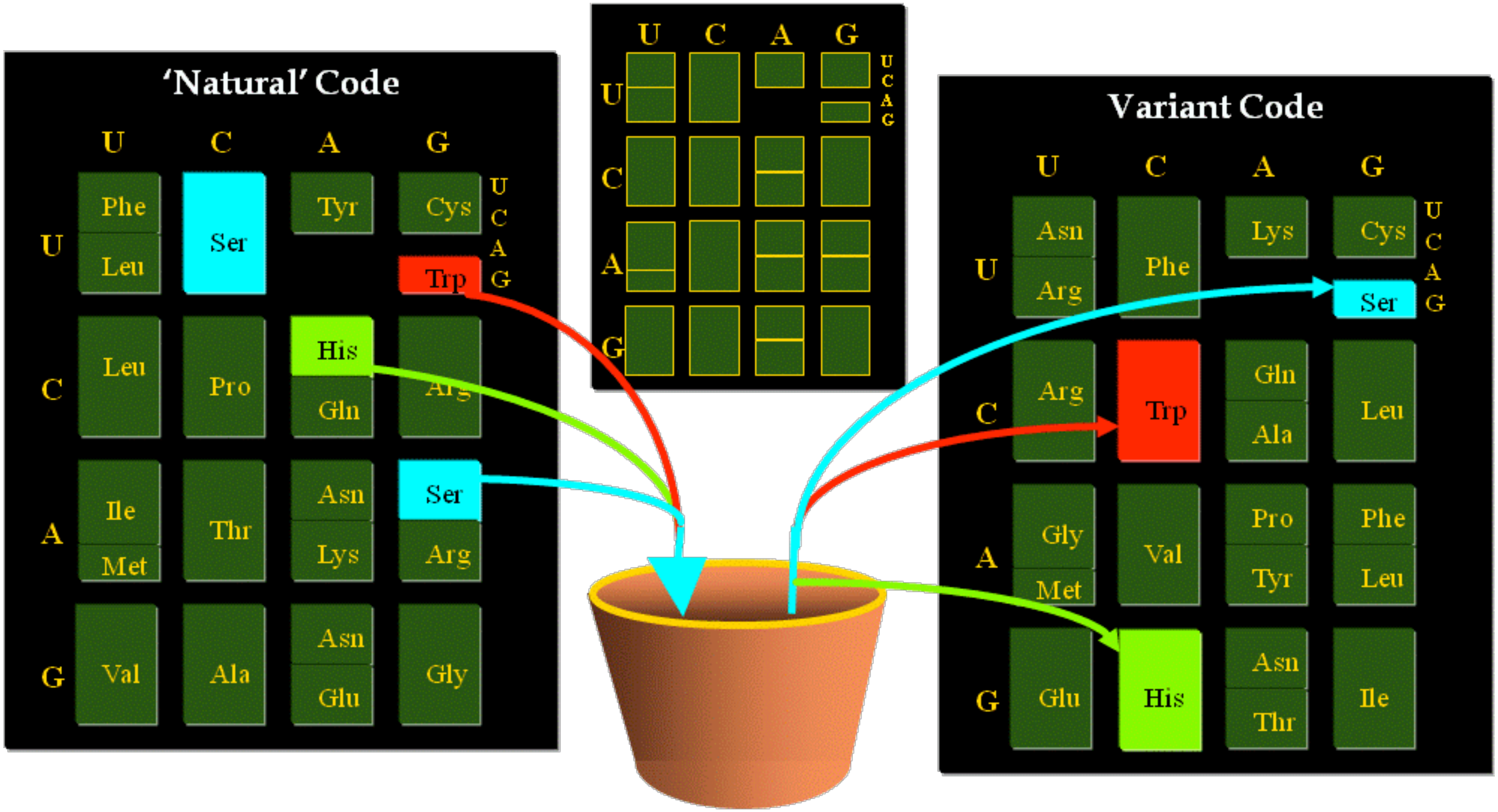

“Polar Requirement” (hydrophobicity)

Calculating the 'error value' (Δ) of a code

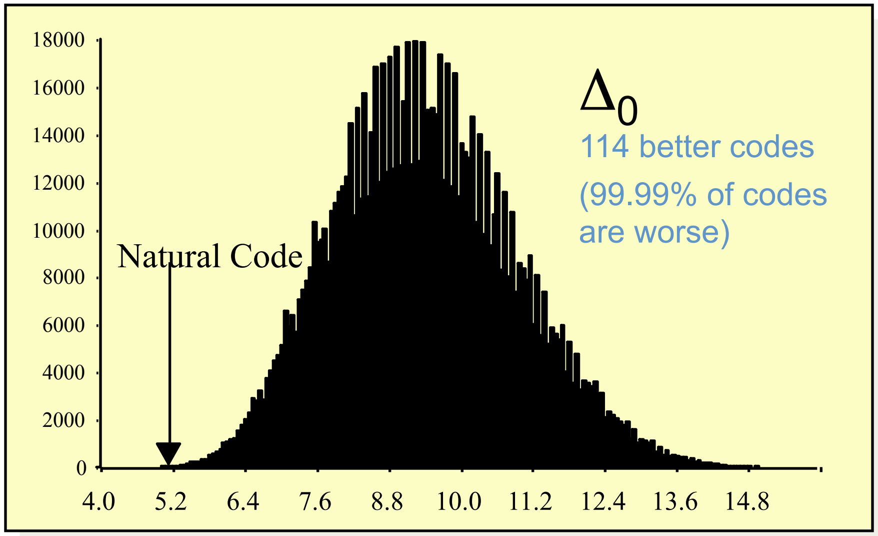


$$\Delta_1 \text{ for Codon UUU} = \frac{(5.0 - 4.9)^2 + (5.0 - 4.9)^2 + (5.0 - 5.6)^2}{3}$$

Generating a random variant code



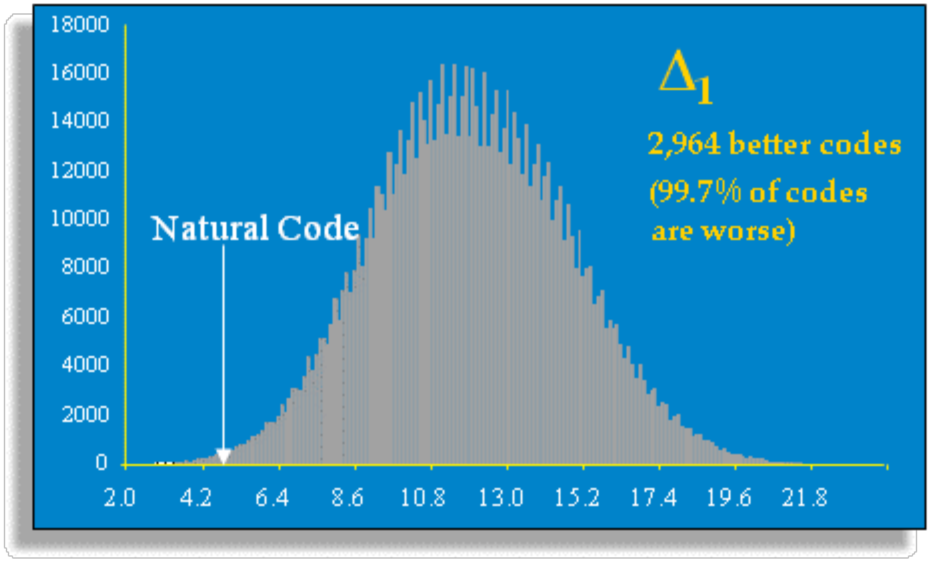
The ‘natural’ code compared to 1,000,000 random alternatives



Δ_0

114 better codes

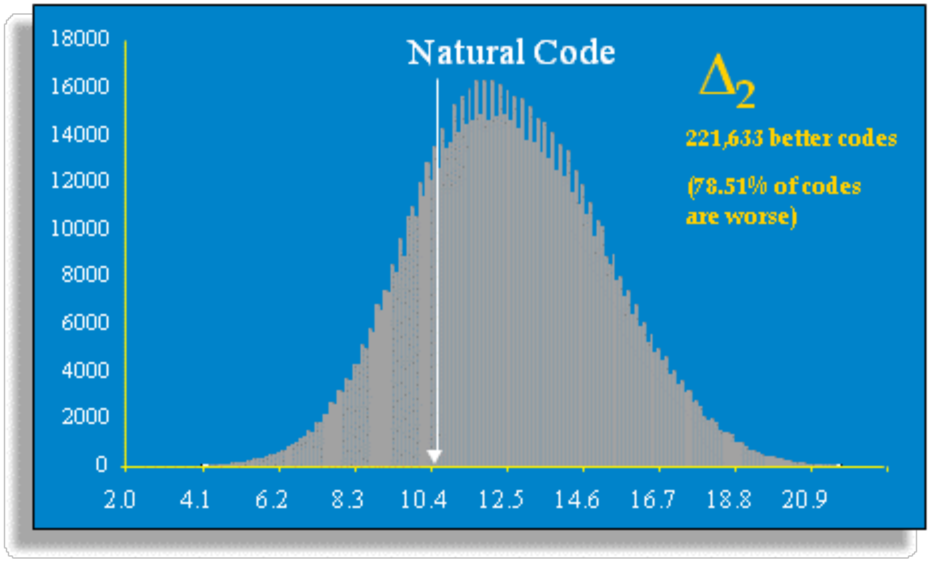
(99.99% of codes are worse)



Δ_1

2,964 better codes

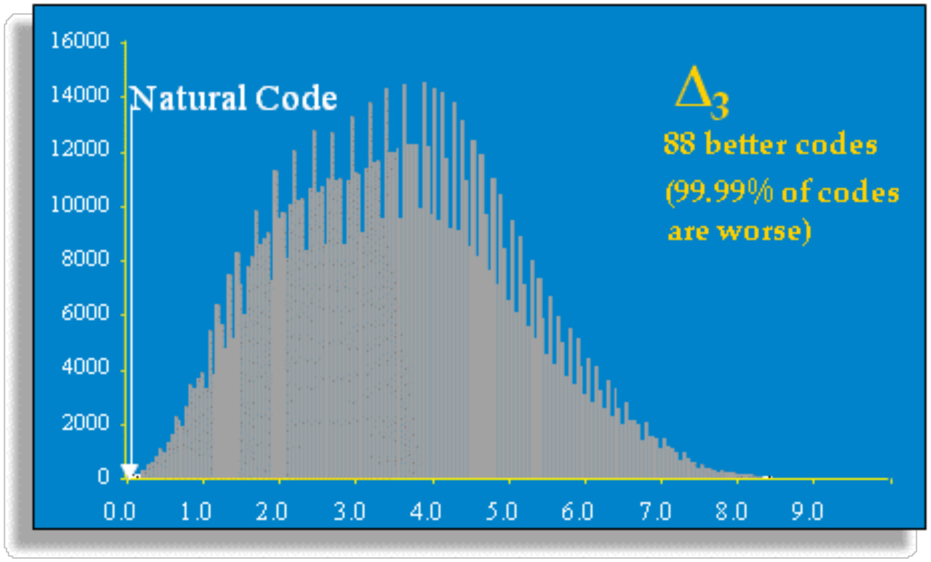
(99.7% of codes are worse)



Δ_2

221,633 better codes

(78.51% of codes are worse)

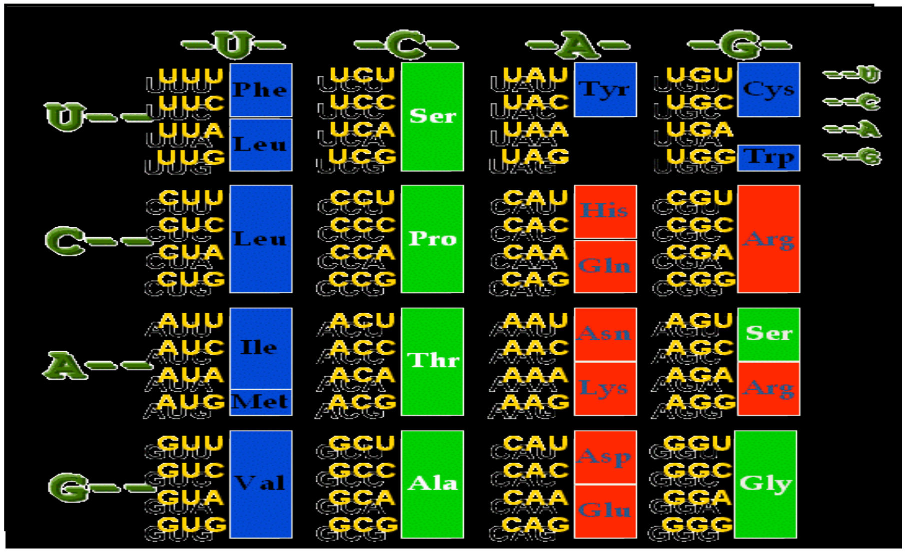


Δ_3

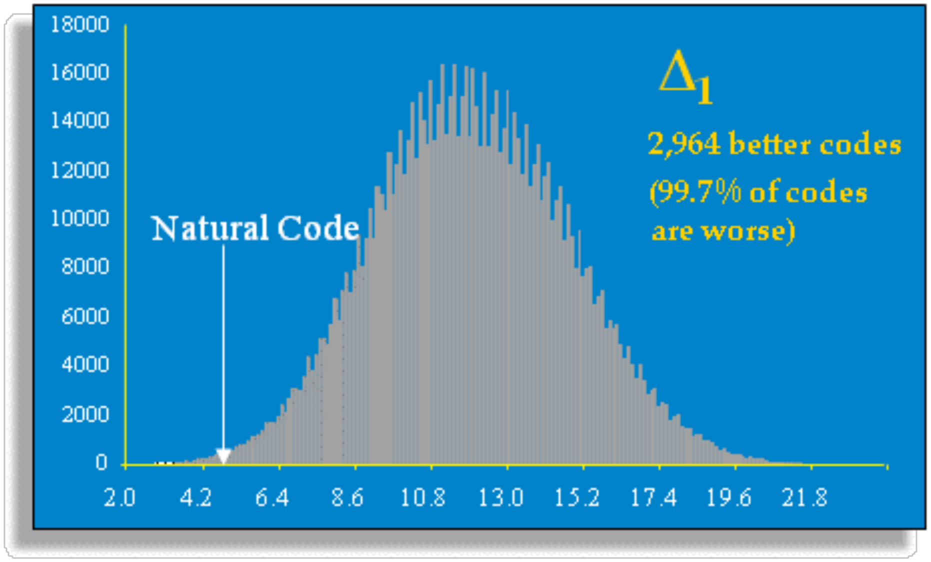
88 better codes

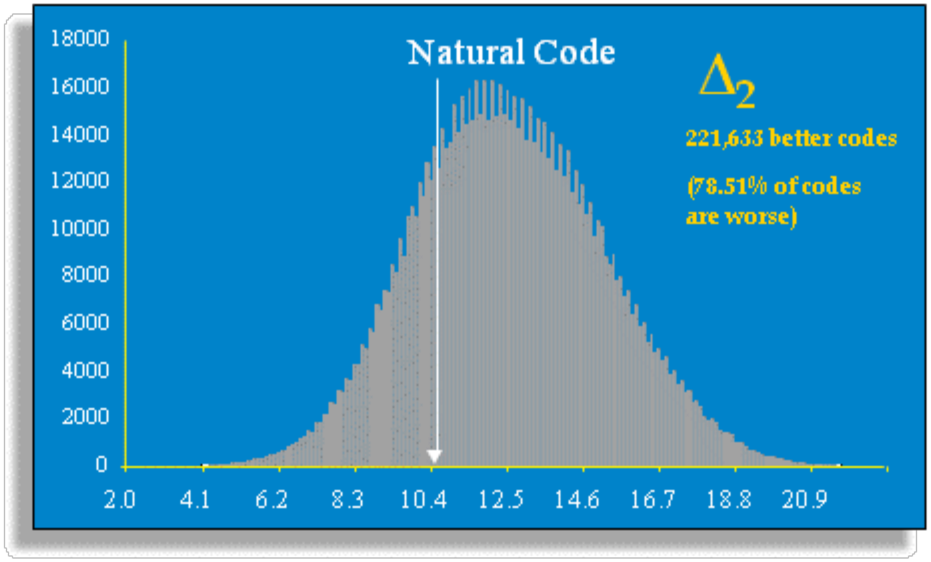
(99.99% of codes are worse)

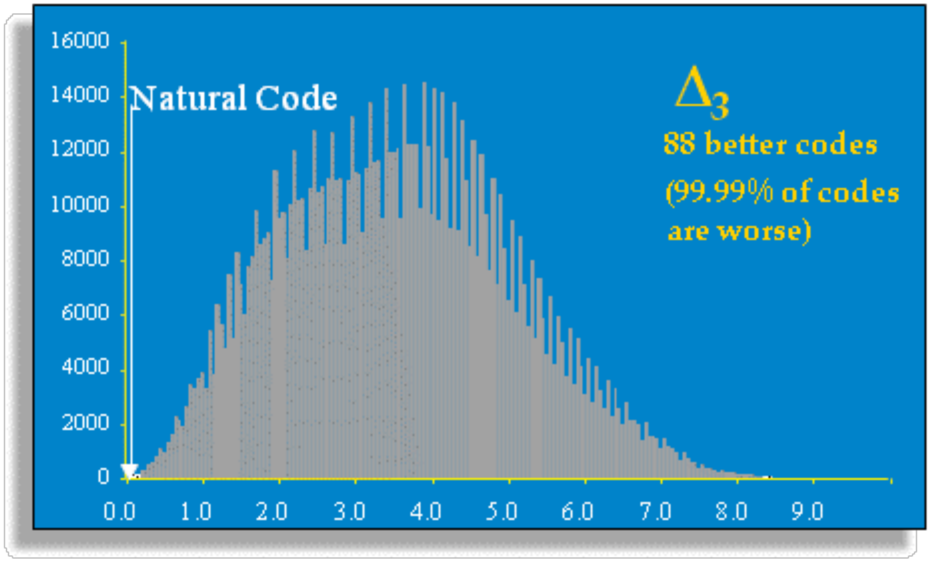
The ‘natural’ code compared to 1,000,000 random alternatives



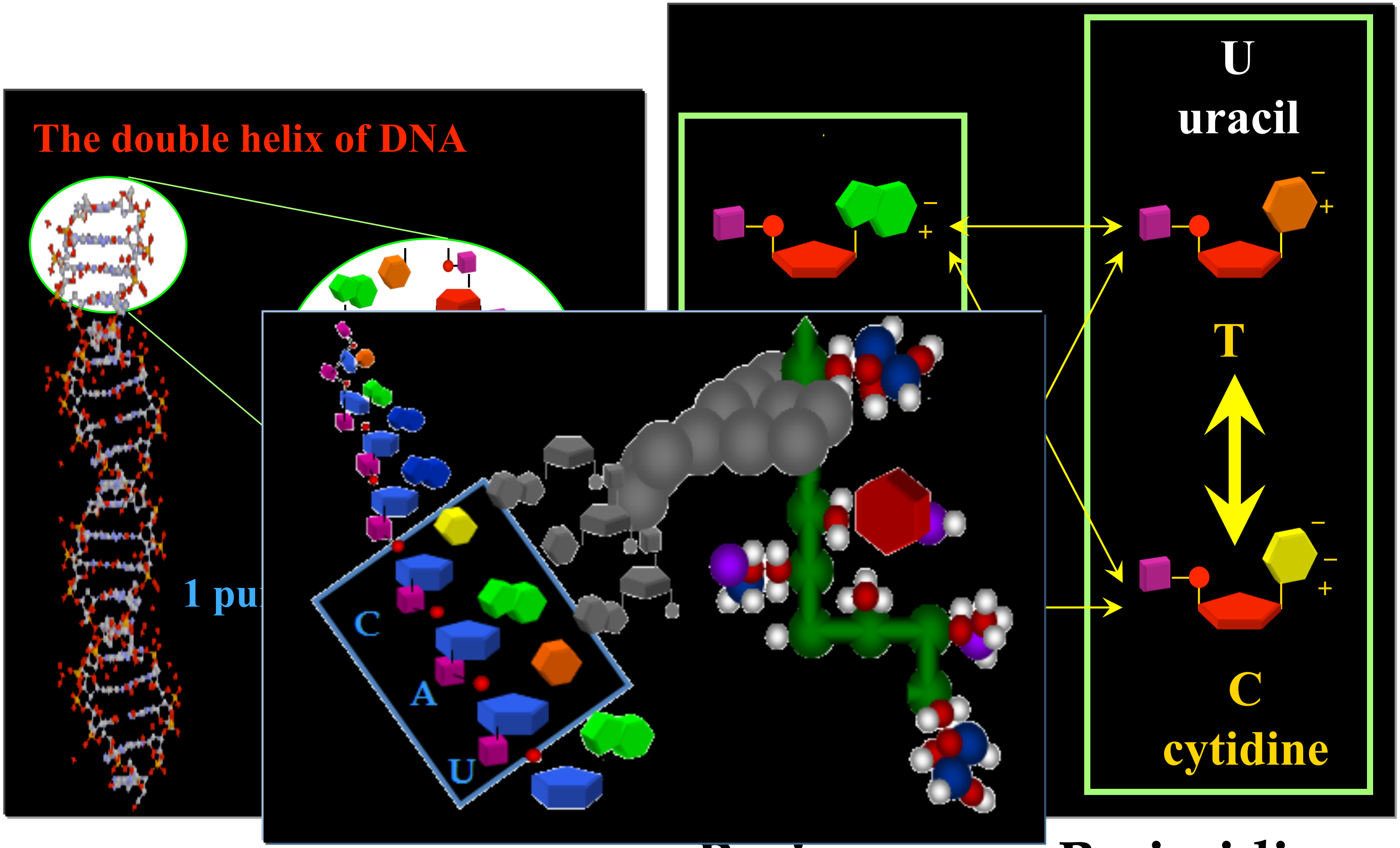
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA		UGA		A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Gln	CGA		
	CUG		CCG		CAG		CGG		
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	AUC		ACC		AAC		AGC		
	AUA		ACA		AAA	Lys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





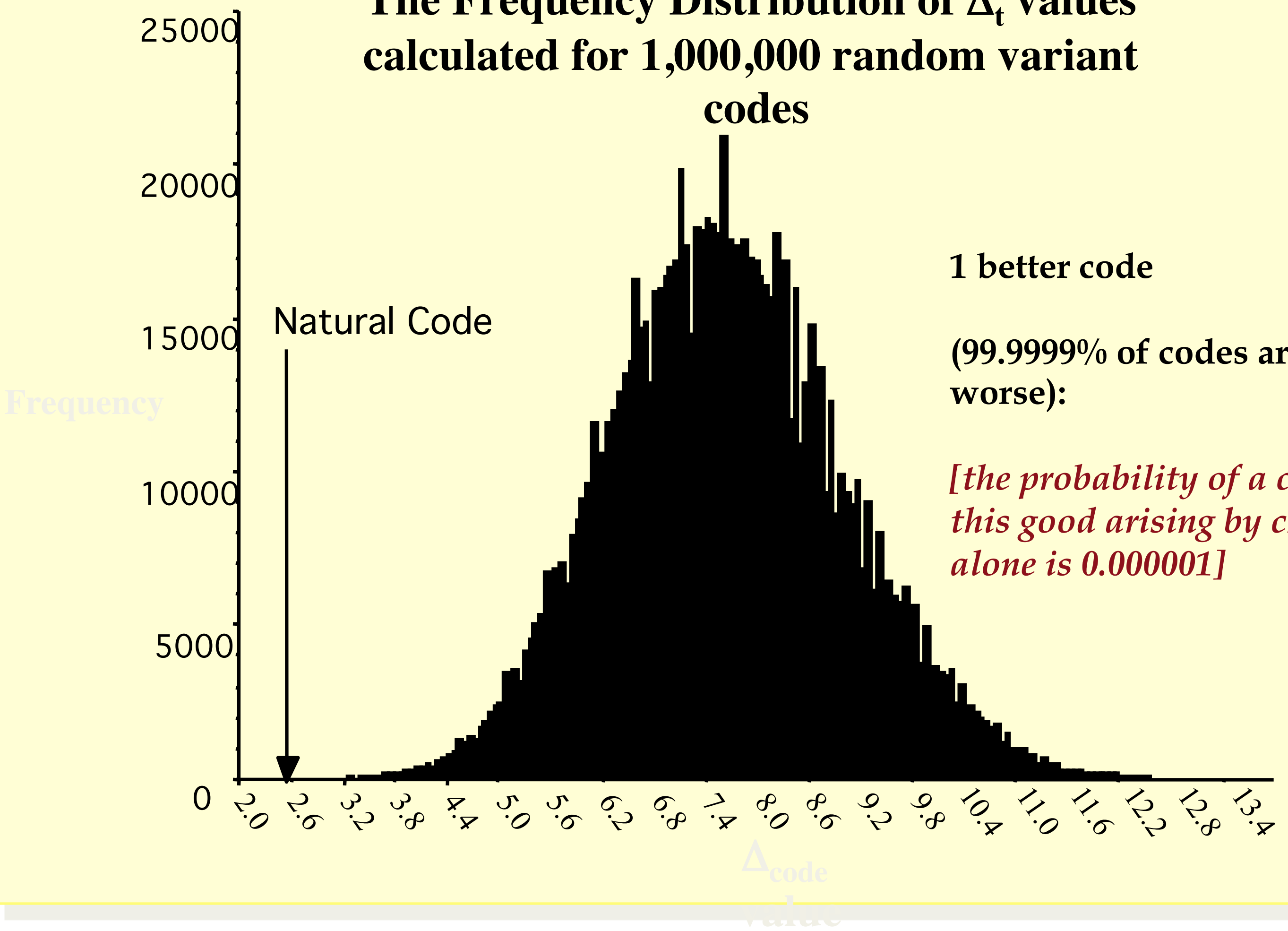


More subtle patterns of genetic error



Purines Pyrimidines

The ‘natural’ code compared to 1,000,000 random alternatives [incorporating patterns of translation error into Δ value calculations]



The Frequency Distribution of Δ_t values calculated for 1,000,000 random variant codes

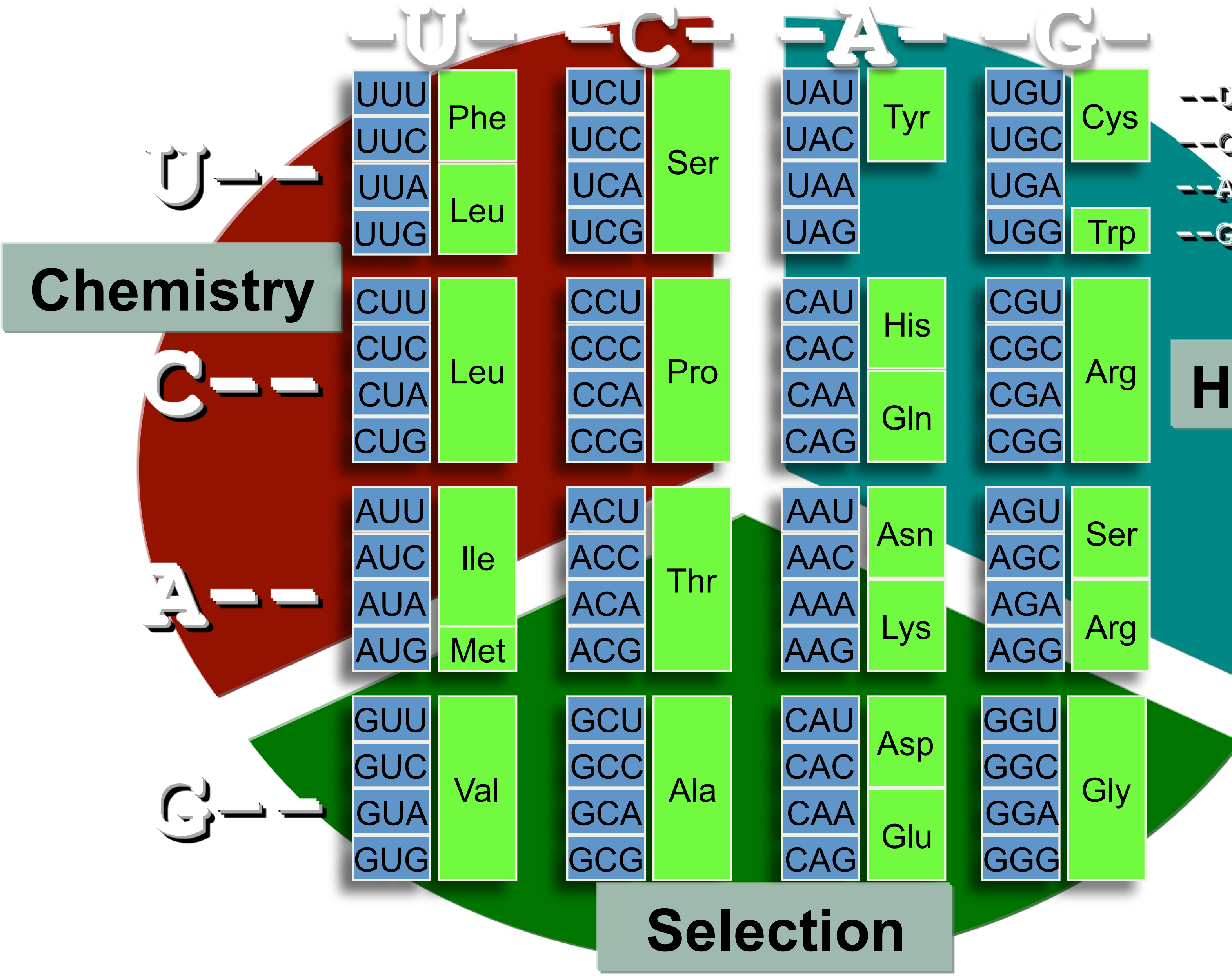
1 better code

(99.9999% of codes are worse):

[the probability of a code this good arising by chance alone is 0.000001]

“The Genetic Code is 1 in a Million”, *Freeland & Hurst* (1998) JME 47: 238-248

The 3 Faces of the Genetic Code

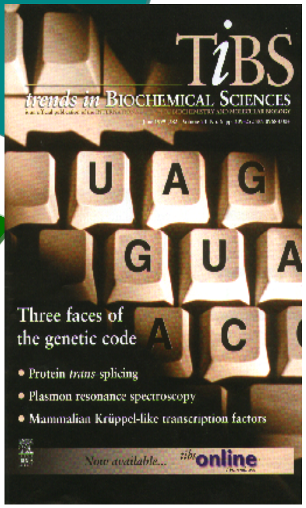


	–U–		–C–		–A–		–G–		
U– –	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	– –U
	UUC		UCC		UAC		UGC		– –C
	UUA	Leu	UCA		UAA		UGA		– –A
	UUG		UCG		UAG		UGG	Trp	– –G
C– –	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Gln	CGA		
	CUG		CCG		CAG		CGG		
A– –	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
	AUC		ACC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

Chemistry

History

Selection



TiBS

trends in Biochemical Sciences

Three faces of the genetic code

- Protein *trans* splicing
- Plasmon resonance spectroscopy
- Mammalian Krüppel-like transcription factors

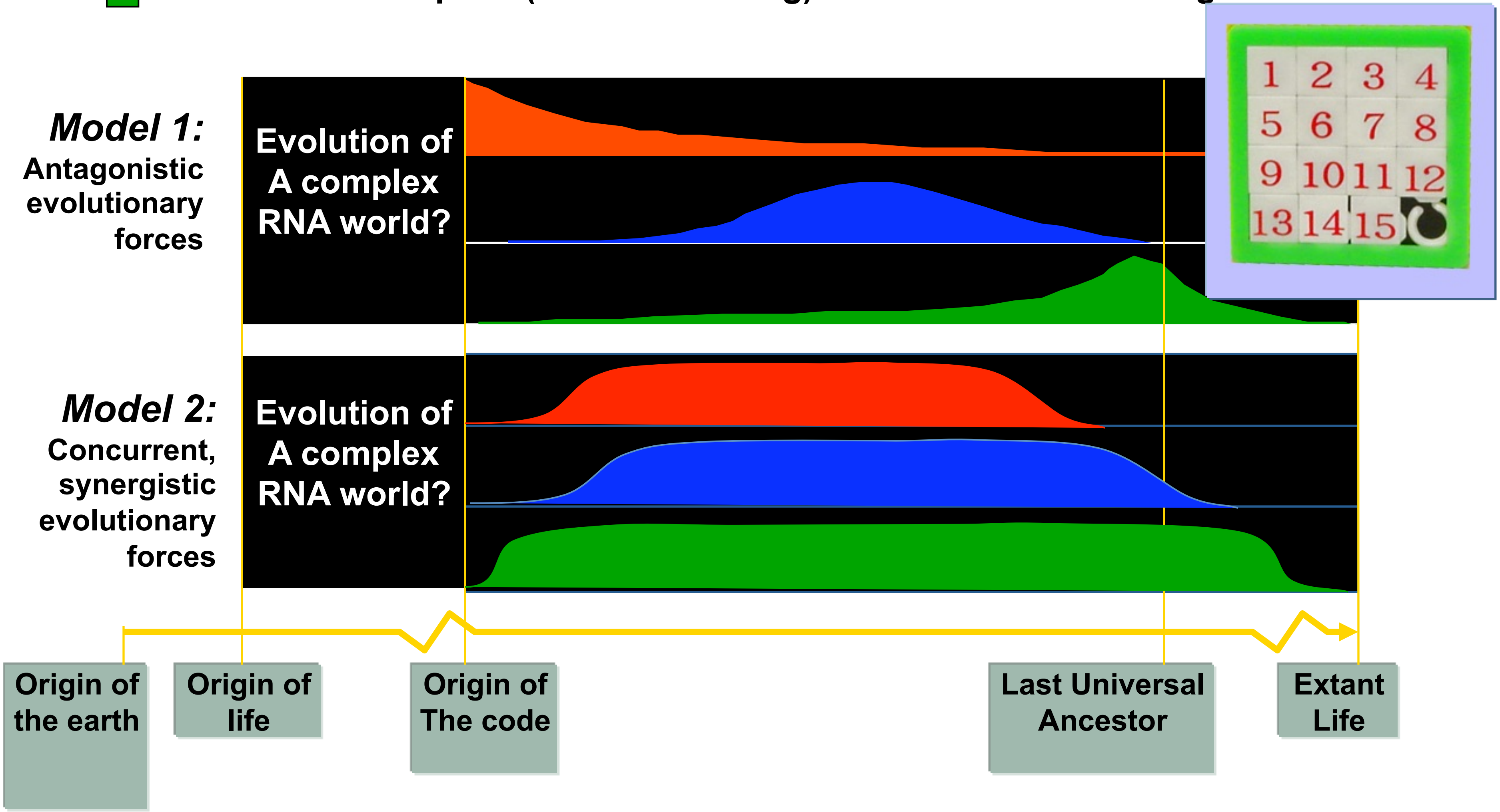
Now available... tibsonline

The 3 Faces : Antagonistic or Synergistic?

🟧 **Chemistry: steric code origins**

🟦 **History: biosynthetic code expansion**

🟩 **Selection: adaptive (error minimizing) “choice” of codon assignments**



Model 1:
Antagonistic evolutionary forces

Evolution of A complex RNA world?

Model 2:
Concurrent, synergistic evolutionary forces

Evolution of A complex RNA world?

Origin of the earth

Origin of life

Origin of The code

Last Universal Ancestor

Extant Life