The molecular origins of life

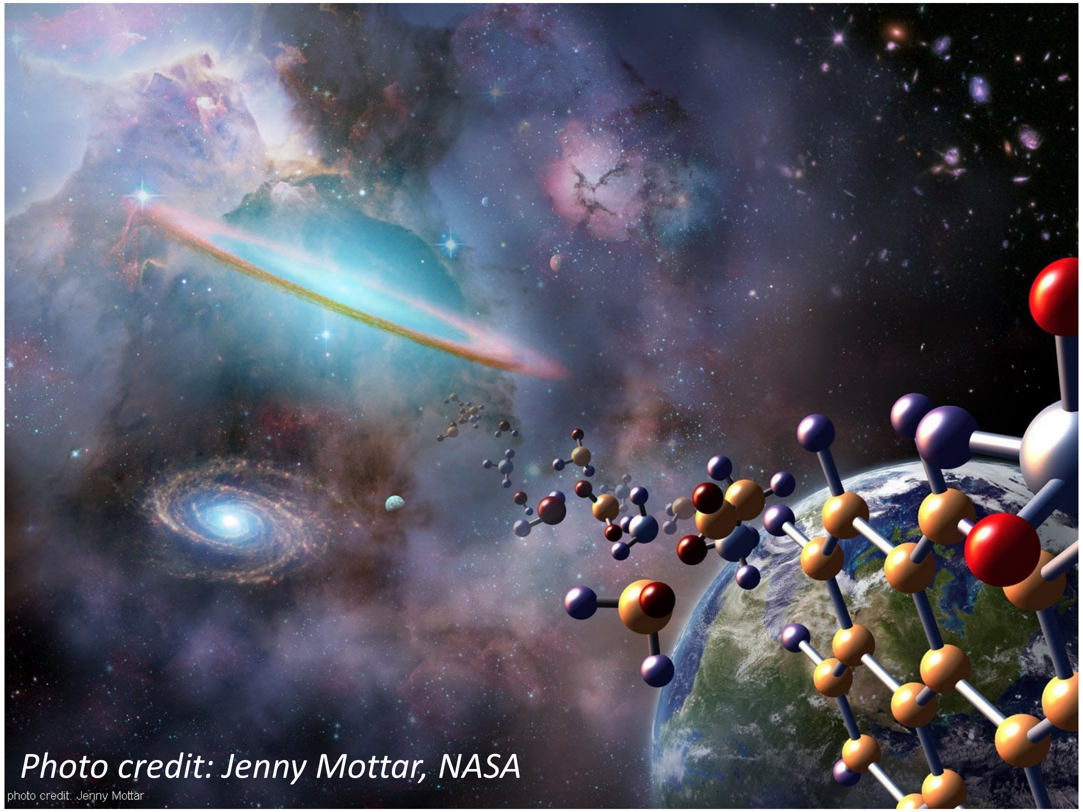


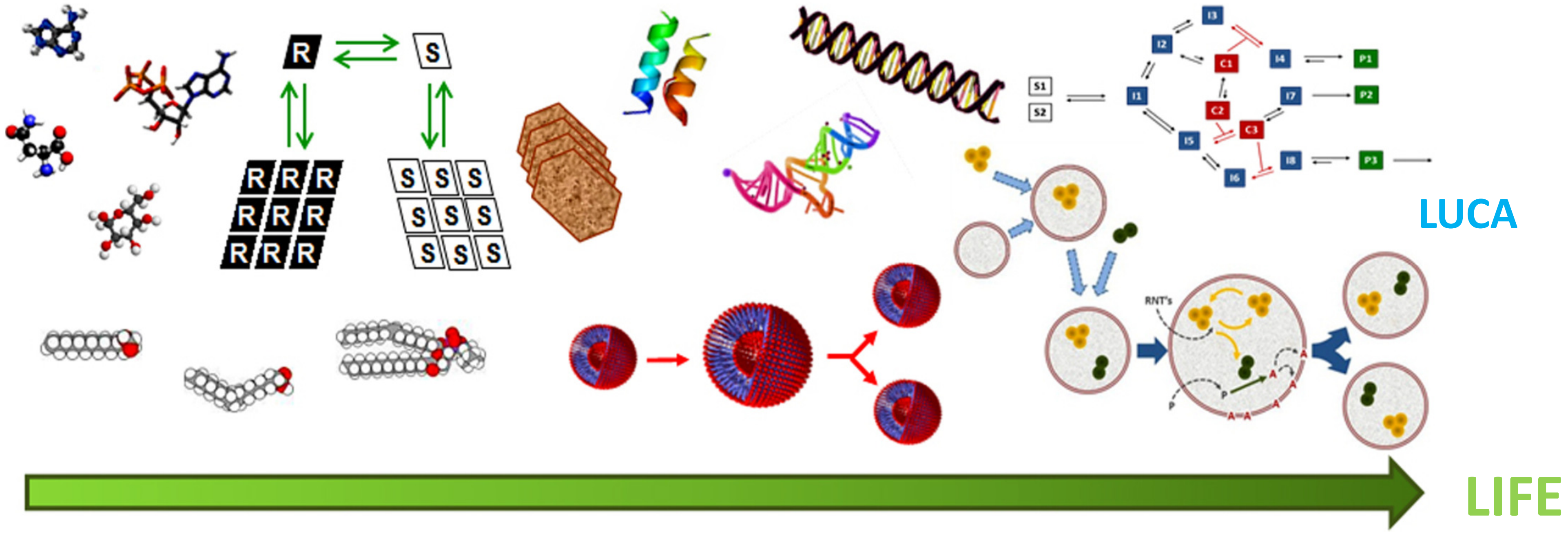
Photo credit: Jenny Mottar, NASA

photo credit: Jenny Mottar

Lecture 7, SoSe 2019 KIT

Zbigniew Pianowski

Self-organization of molecules and chemical reactions



Increasing complexity from molecules to systems

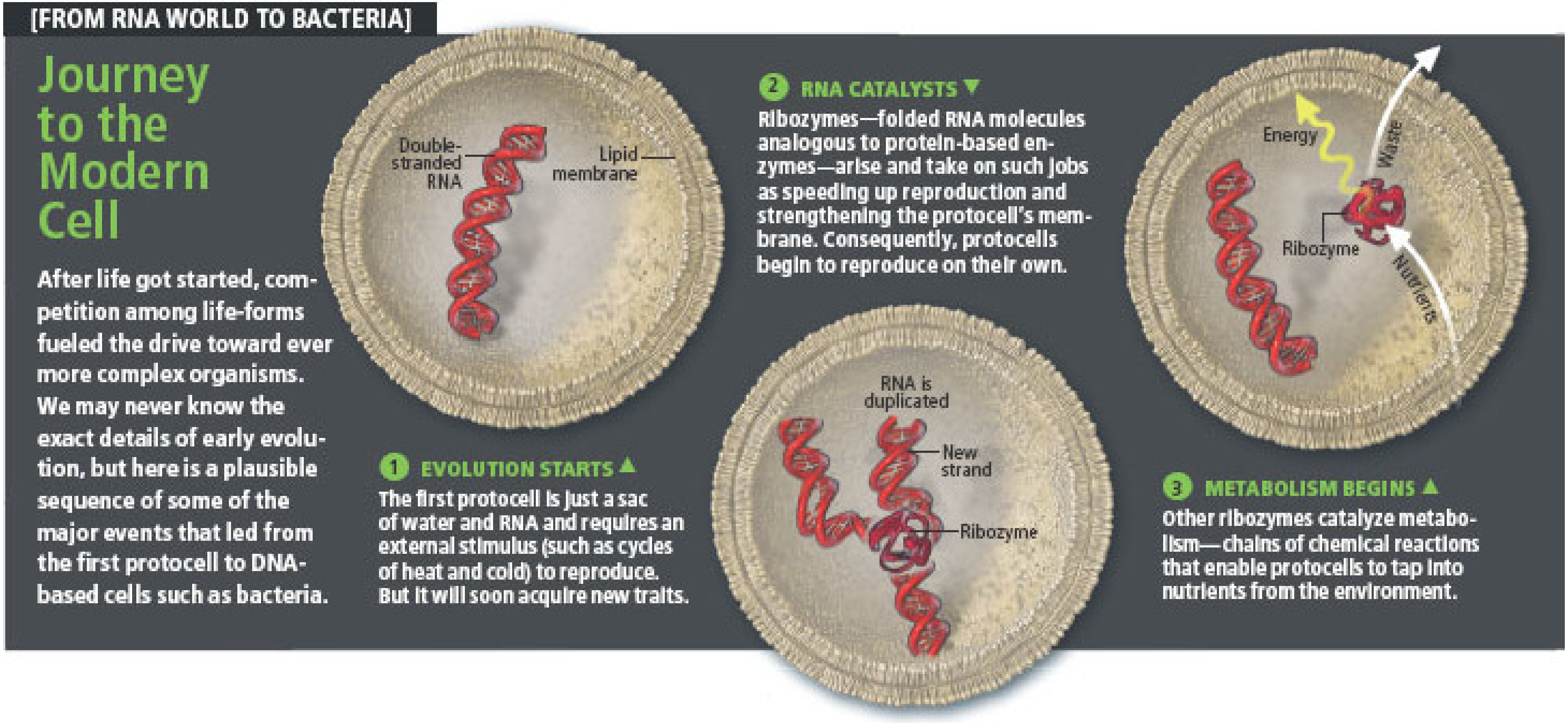
Origin of the Universe – stars, planets, elements

Origin of biorelevant monomers – primordial soup

Complex chemical processes on the way to living systems

Protocells and LUCA

From RNA world to bacteria



[FROM RNA WORLD TO BACTERIA]

Journey to the Modern Cell

After life got started, competition among life-forms fueled the drive toward ever more complex organisms. We may never know the exact details of early evolution, but here is a plausible sequence of some of the major events that led from the first protocell to DNA-based cells such as bacteria.

① EVOLUTION STARTS ▲

The first protocell is just a sac of water and RNA and requires an external stimulus (such as cycles of heat and cold) to reproduce. But it will soon acquire new traits.

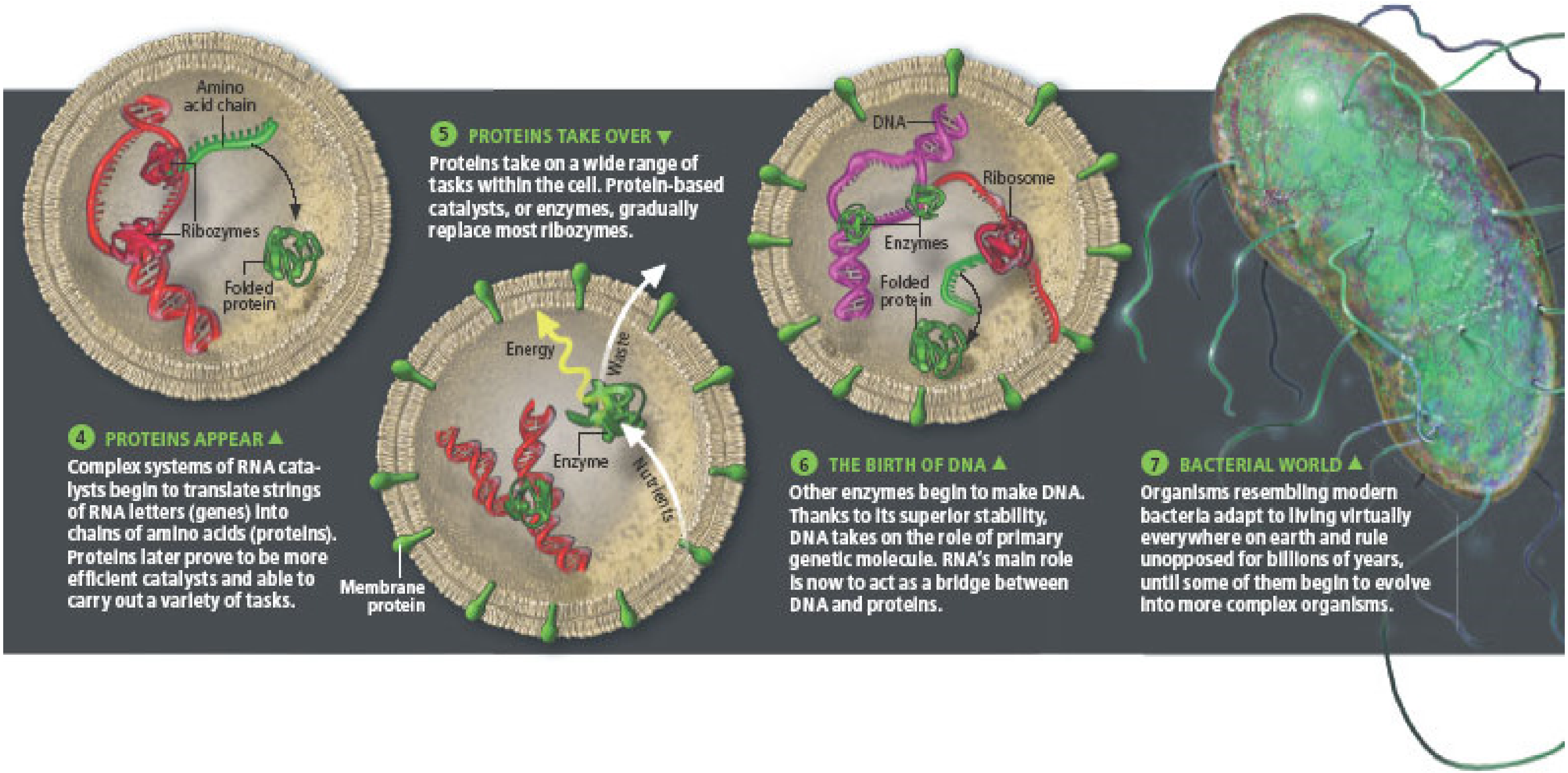
② RNA CATALYSTS ▼

Ribozymes—folded RNA molecules analogous to protein-based enzymes—arise and take on such jobs as speeding up reproduction and strengthening the protocell's membrane. Consequently, protocells begin to reproduce on their own.

③ METABOLISM BEGINS ▲

Other ribozymes catalyze metabolism—chains of chemical reactions that enable protocells to tap into nutrients from the environment.

From RNA world to bacteria



④ PROTEINS APPEAR ▲

Complex systems of RNA catalysts begin to translate strings of RNA letters (genes) into chains of amino acids (proteins). Proteins later prove to be more efficient catalysts and able to carry out a variety of tasks.

⑤ PROTEINS TAKE OVER ▼

Proteins take on a wide range of tasks within the cell. Protein-based catalysts, or enzymes, gradually replace most ribozymes.

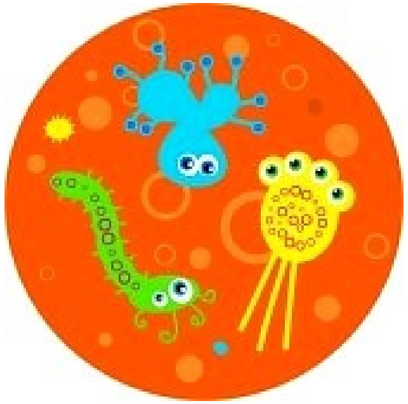
⑥ THE BIRTH OF DNA ▲

Other enzymes begin to make DNA. Thanks to its superior stability, DNA takes on the role of primary genetic molecule. RNA's main role is now to act as a bridge between DNA and proteins.

⑦ BACTERIAL WORLD ▲

Organisms resembling modern bacteria adapt to living virtually everywhere on earth and rule unopposed for billions of years, until some of them begin to evolve into more complex organisms.


LUCA
Last Universal Common Ancestor



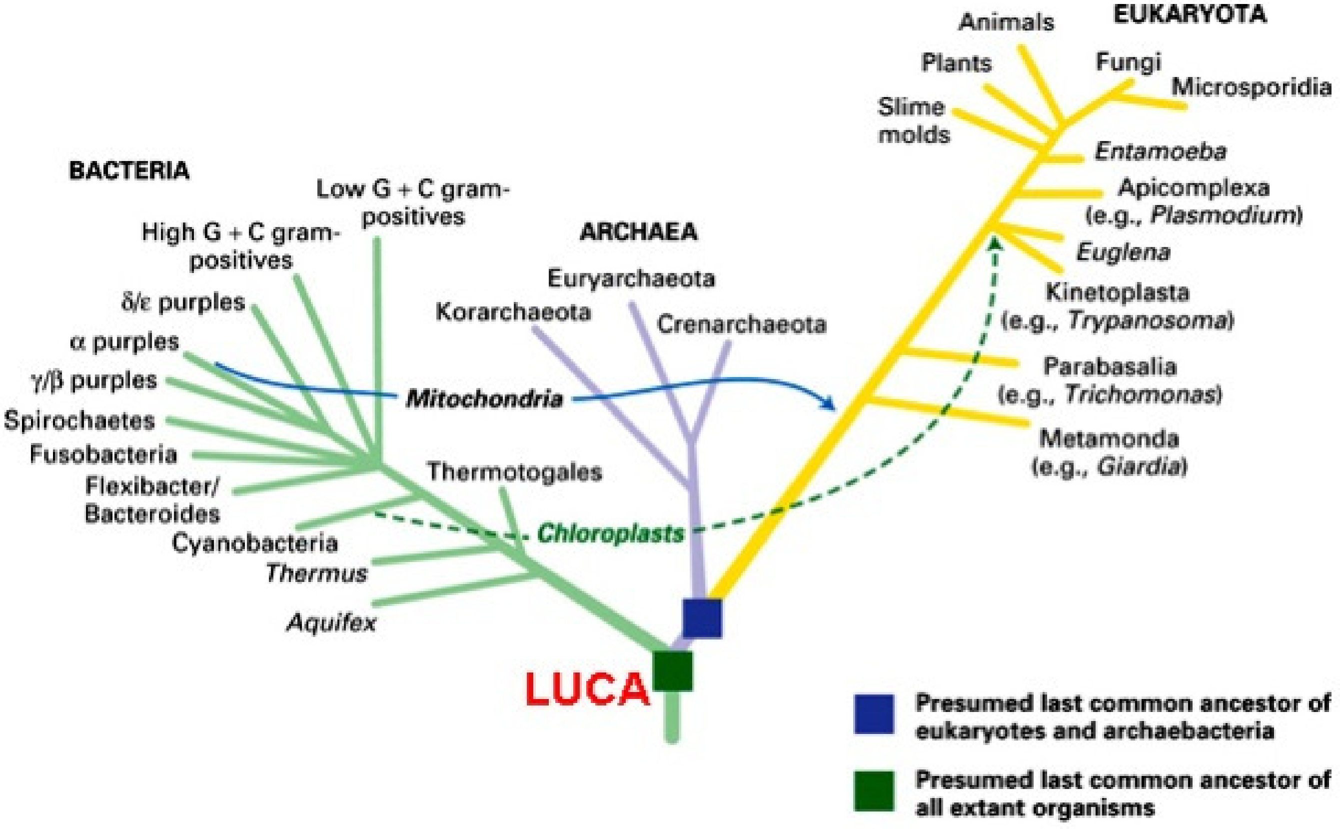
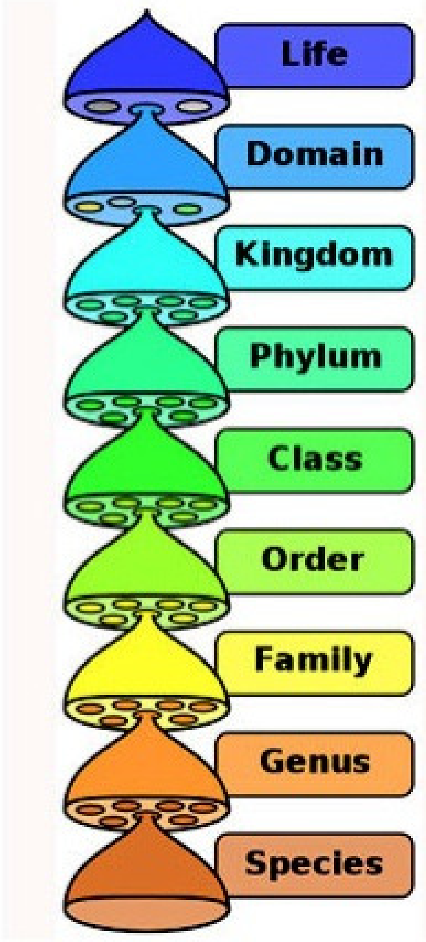
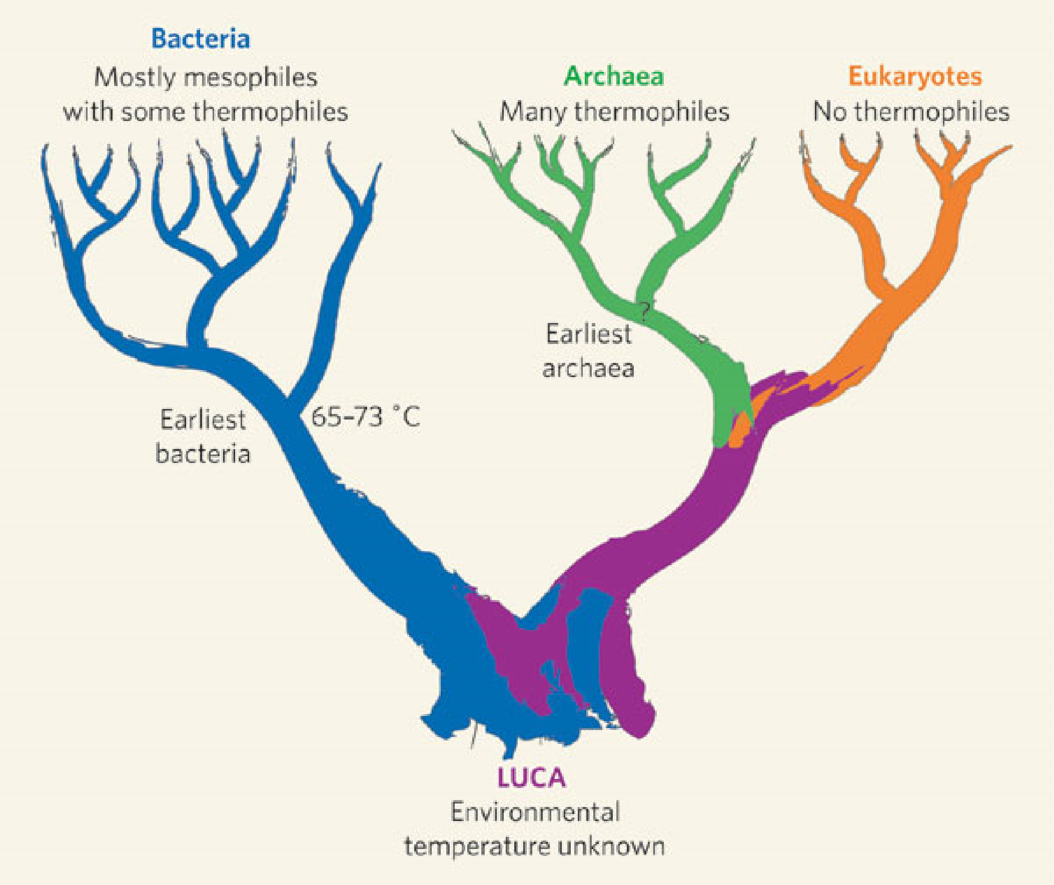


Figure 1.1a

Molecular Cell Biology, Seventh Edition




Bacteria
Mostly mesophiles with some thermophiles
Archaea
Many thermophiles
Eukaryotes
No thermophiles
?
Earliest archaea
Earliest bacteria
65–73 °C
LUCA
Environmental temperature unknown

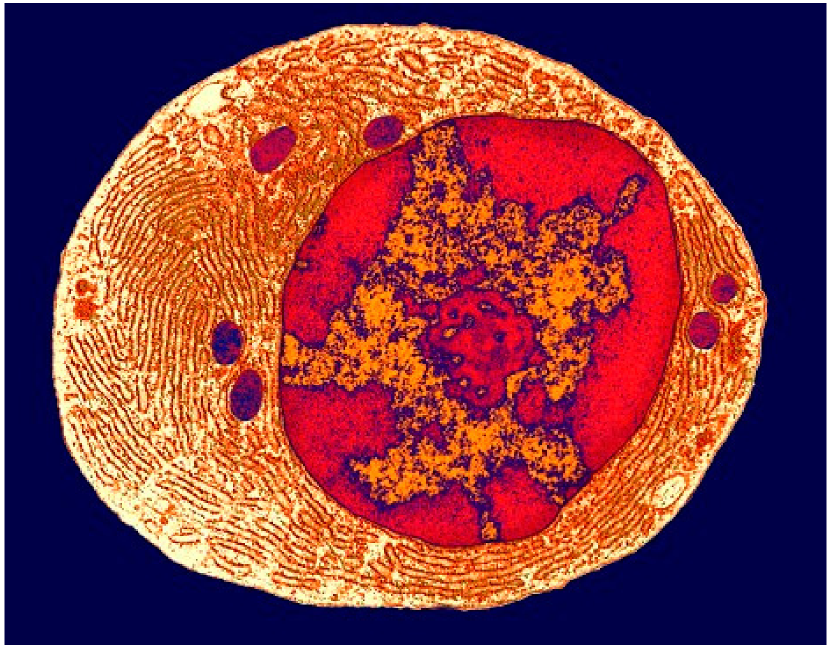
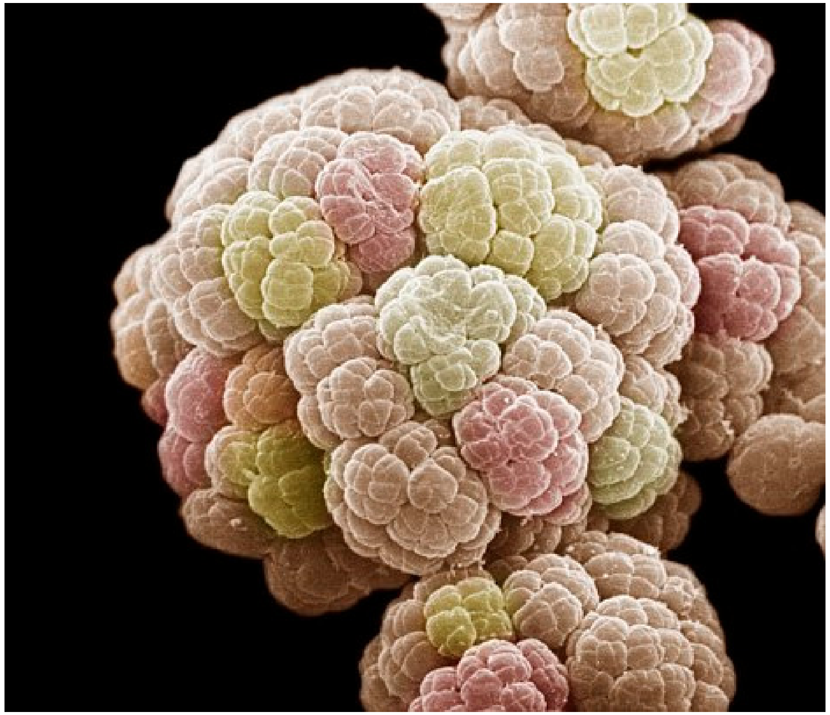
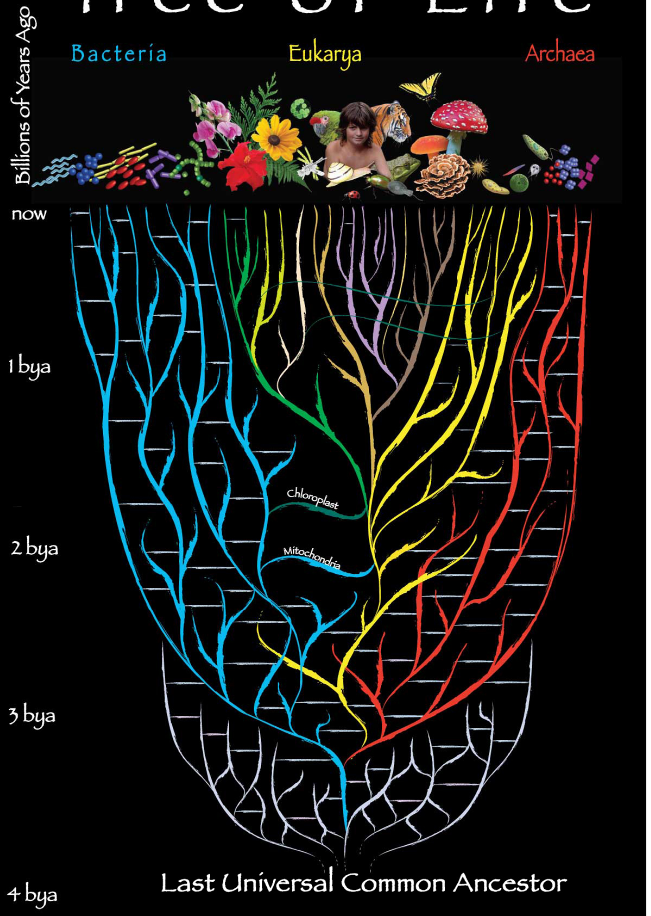


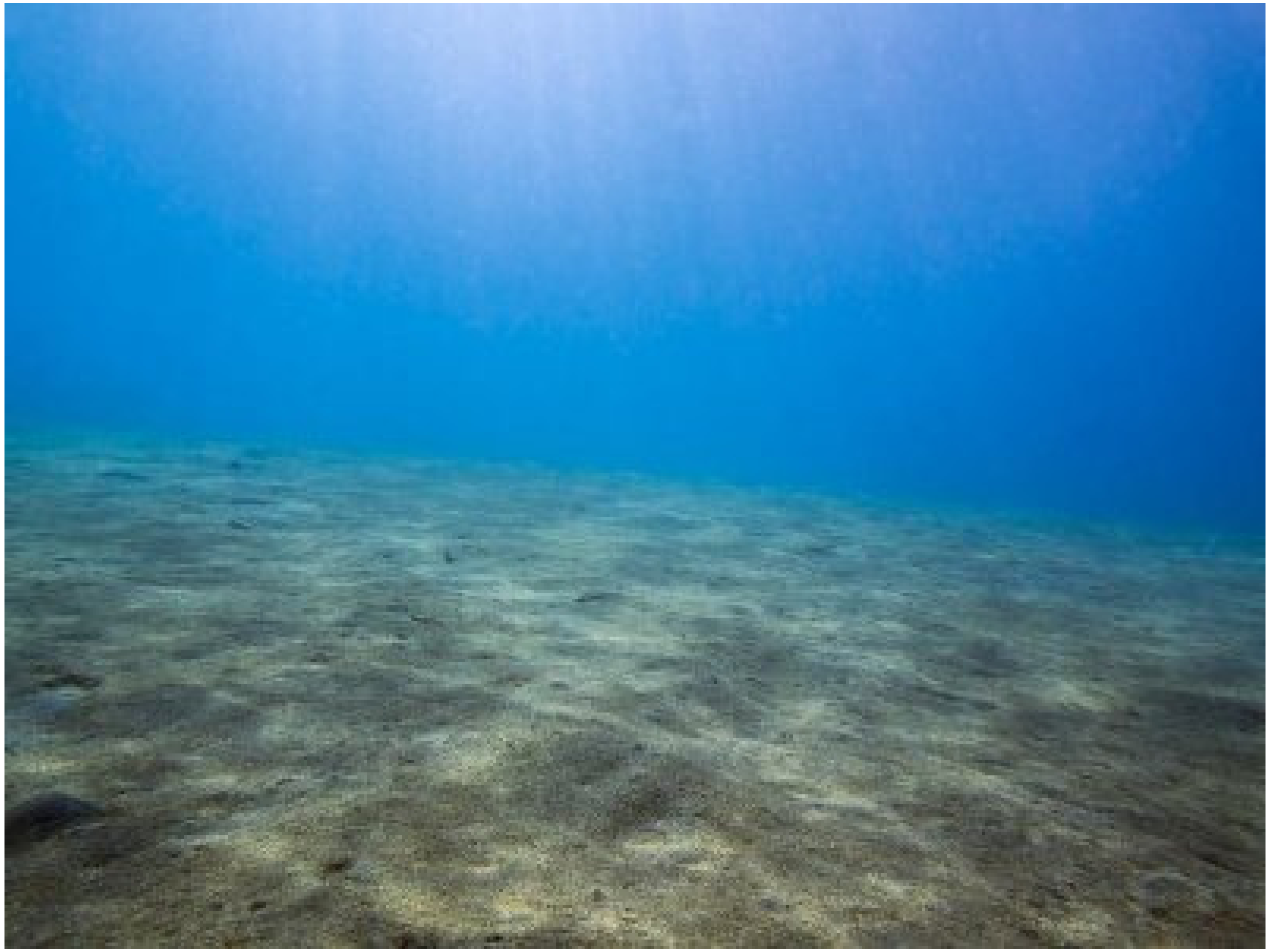
Image of a eukaryotic cell contains numerous organelles, which are now thought to be present in the last universal common ancestor



A colony of the archaea, which form one of the three lines of the tree of life in evolutionary history

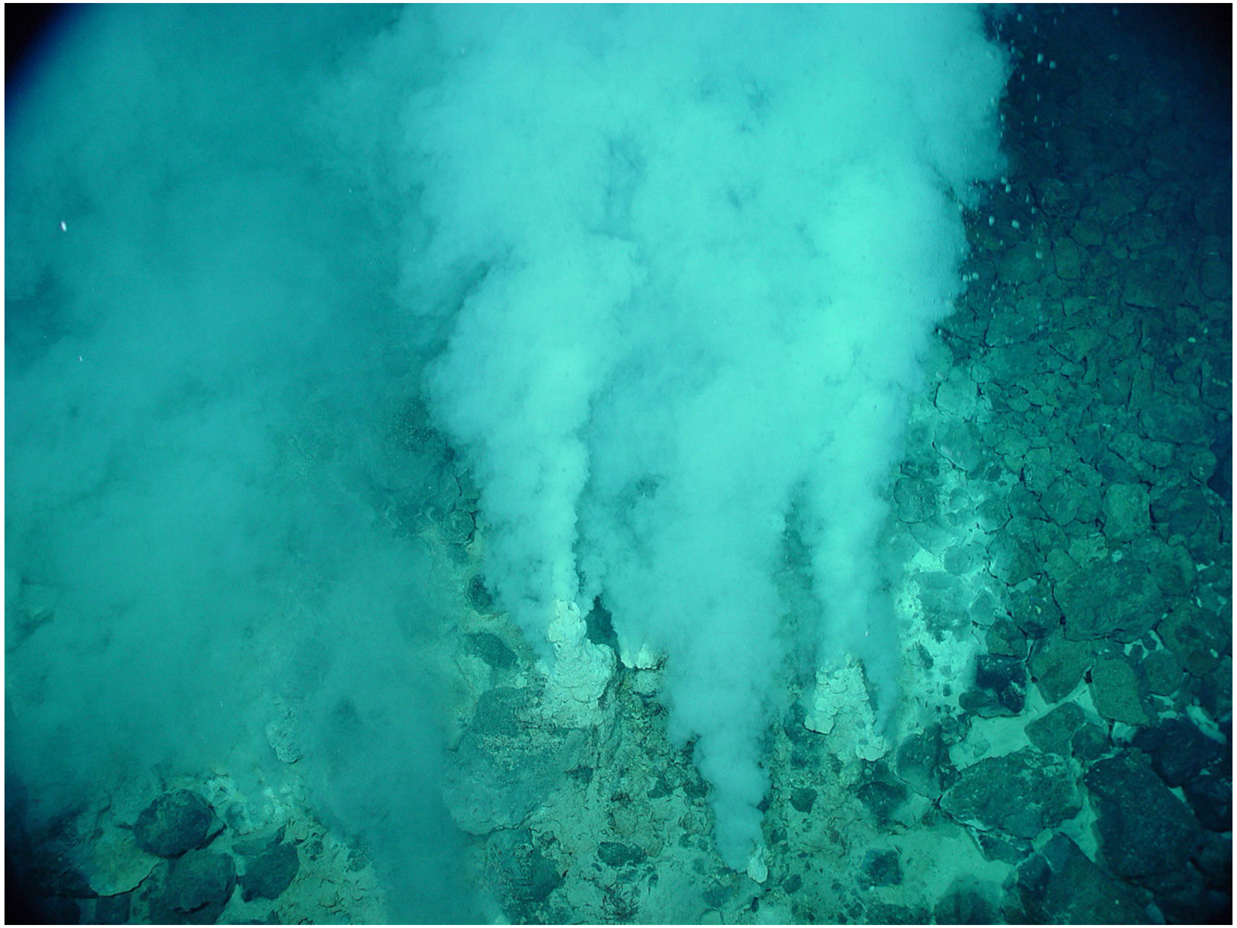
Tree of Life


Billions of Years Ago
Bacteria
Eukarya
Archaea
now
1 bya
2 bya
3 bya
4 bya
Chloroplast
Mitochondria
Last Universal Common Ancestor

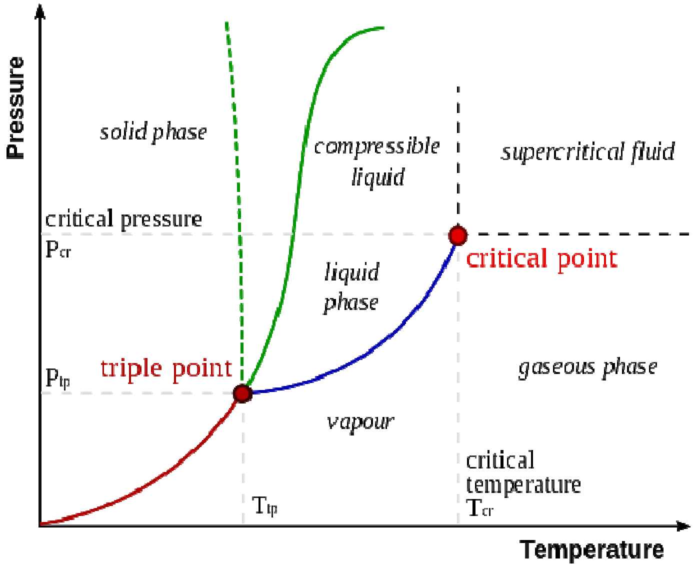


The Beginning or Origin of Life near Deep Sea Hydrothermal vents

Hydrothermal vents

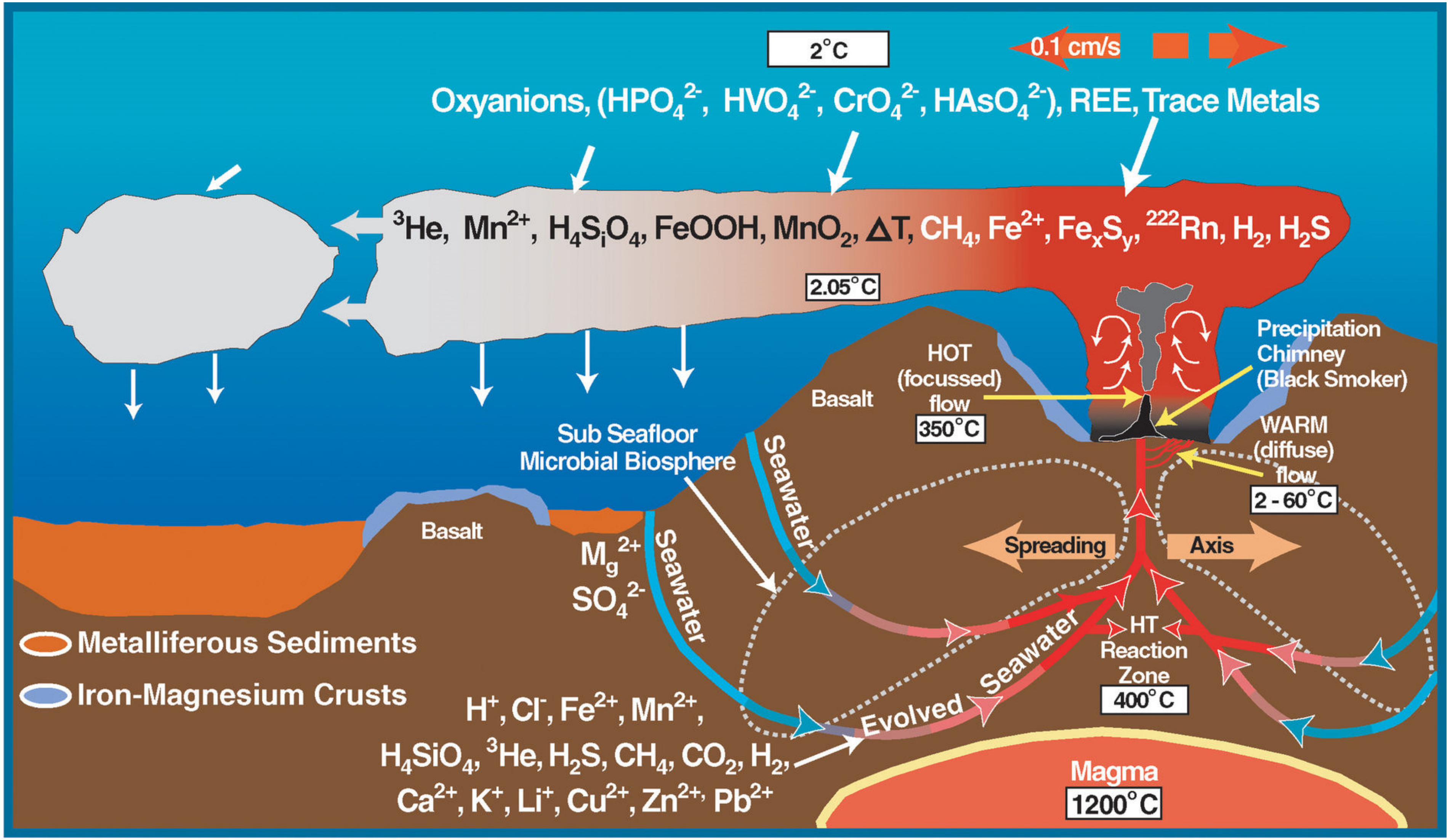


White flocculent mats in and around the extremely gassy, high-temperature (>100°C, 212°F) white smokers at Champagne Vent.

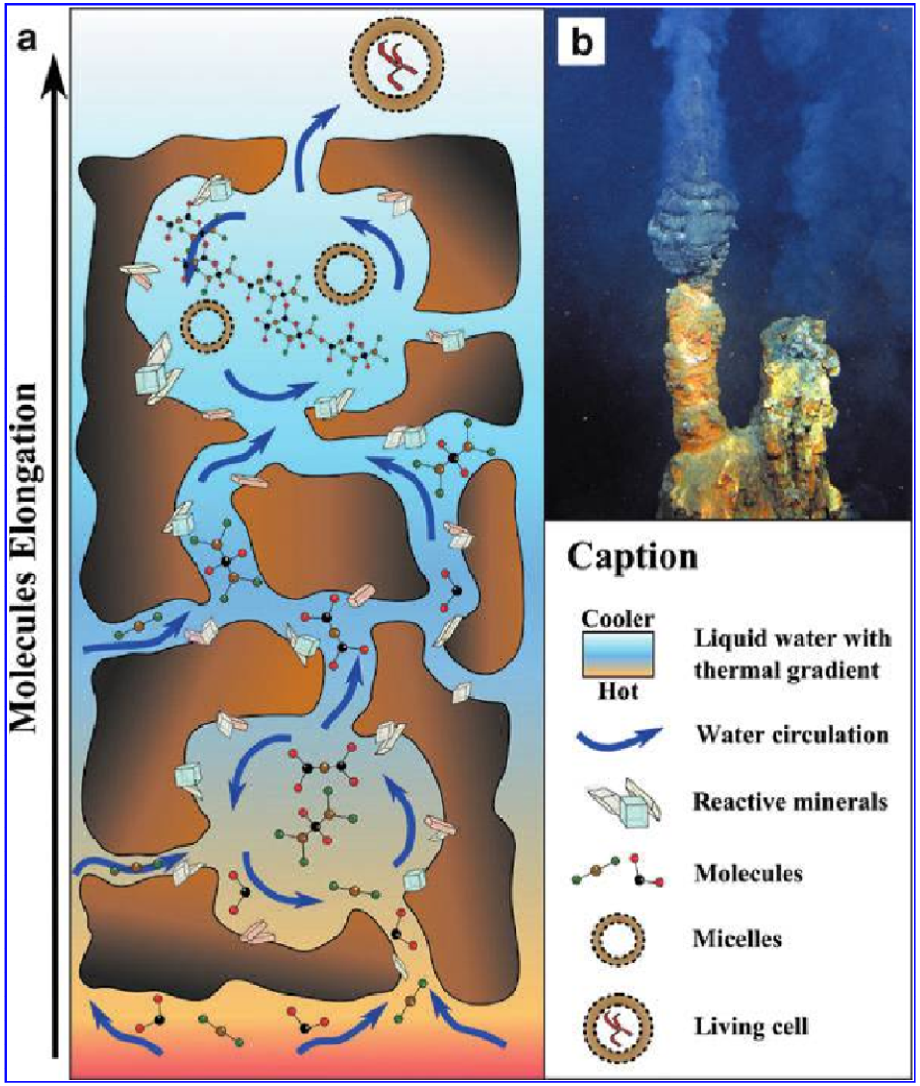


Alkaline hydrothermal vents consist of microscale caverns coated by thin membraneous metal sulfide walls → ‚Iron-sulfur world'

Deep sea vent biogeochemical cycle diagram



Deep sea vent origin of life



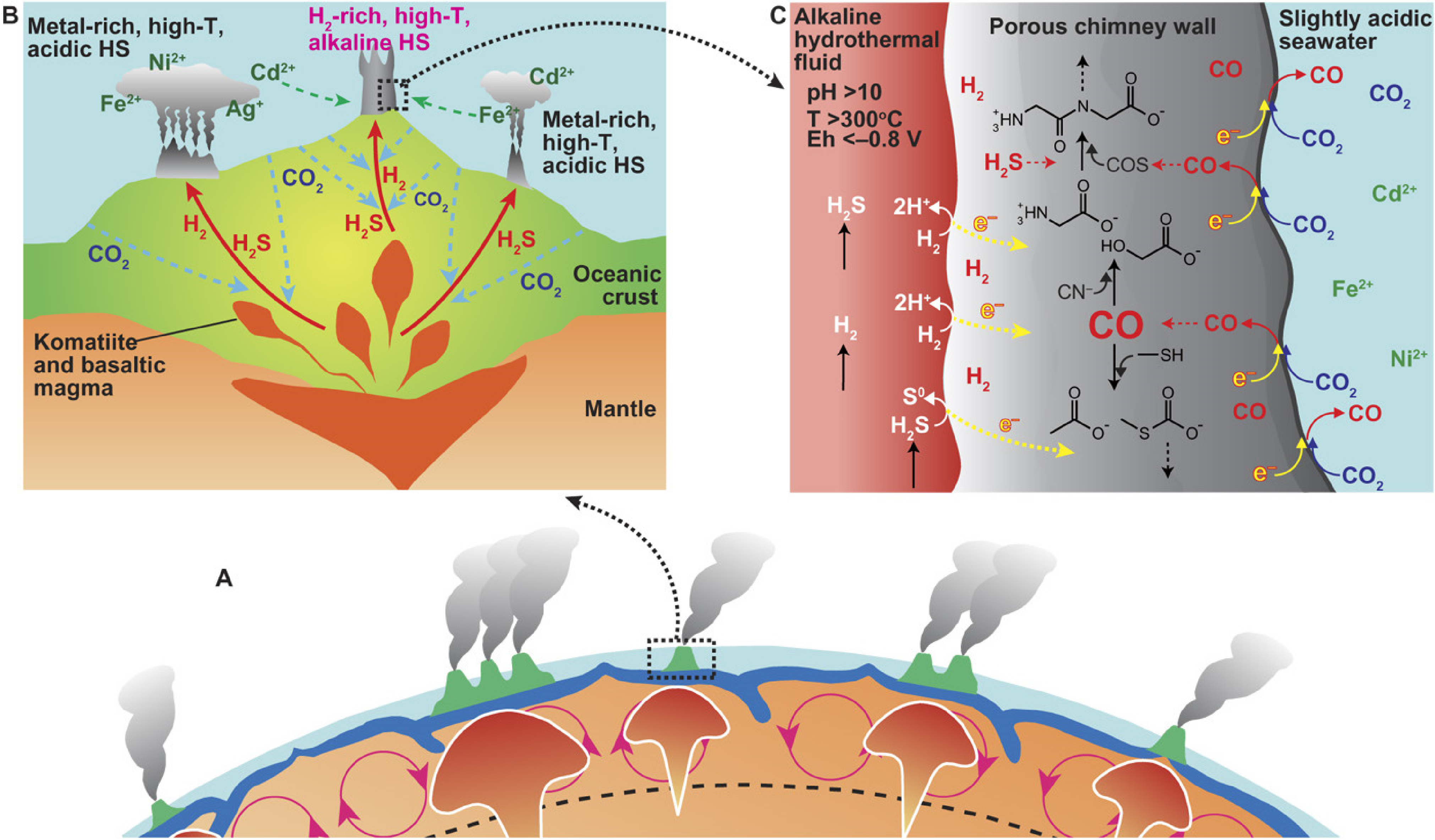
Possible origin of life in porous hydrothermal vents.

(a) Sketch showing a porous beehive structure where hydrothermal fluids and seawater can circulate, leading to the accumulation of organic molecules. The reduced mineral surfaces within the vent pores could be favorable locations for the structural organization of macromolecules. We hypothesize the formation of lipid micelles in these environments and the incorporation of information-transferring molecules within the micelles, perhaps due to moderate agitation of the hydrothermal effluent.

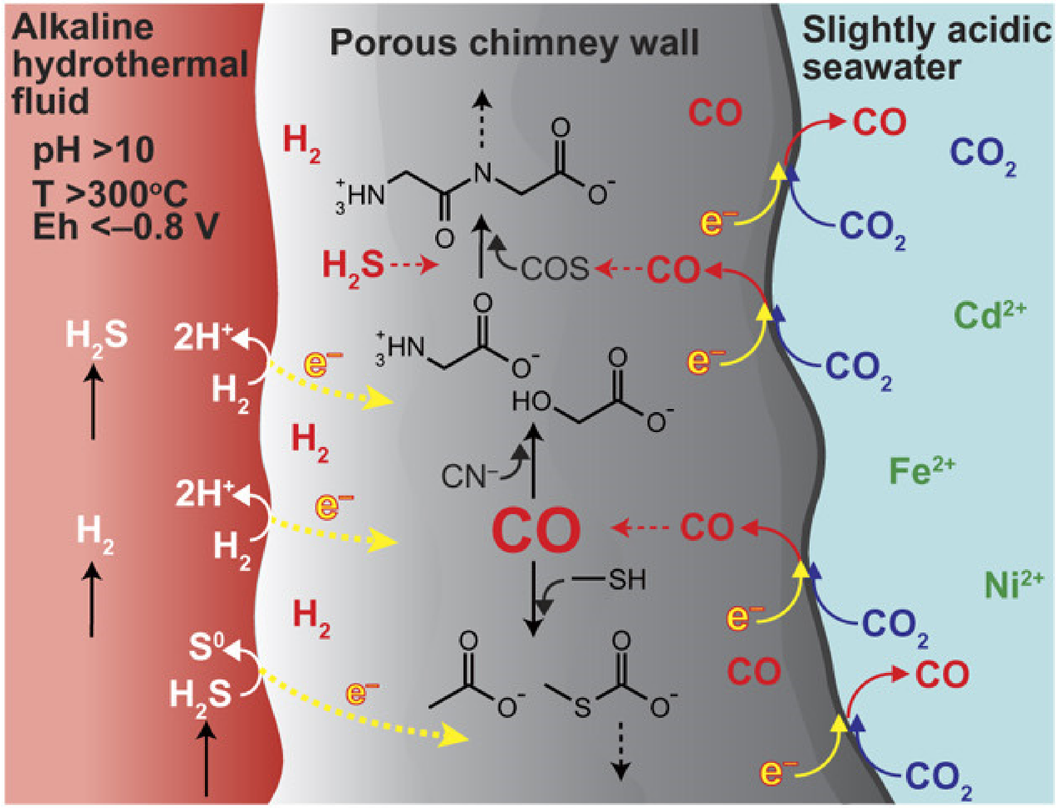
(b) Image of a modern black smoker

(image credit: National Oceanographic and Atmospheric Administration). Color images available online at www.liebertonline.com/ast

F. Westall et al., *Astrobiology* **2013**, *13(9)*, 887-897


B
Metal-rich, high-T, acidic HS
H_2-rich, high-T, alkaline HS
Ni^{2+}
Fe^{2+}
Ag^{+}
Cd^{2+}
Cd^{2+}
Fe^{2+}
Metal-rich, high-T, acidic HS
CO_2
H_2
H_2S
CO_2
H_2
H_2S
CO_2
H_2S
CO_2
Oceanic crust
Komatiite and basaltic magma
Mantle
C
Alkaline hydrothermal fluid
pH >10
T >300°C
Eh <–0.8 V
Porous chimney wall
Slightly acidic seawater
H_2
H_2S
COS
CO
CO
CO
CO_2
CO_2
H_2S
$2H^+$
H_2
e^-
CO_2
Cd^{2+}
H_2
CN^-
H_2
$2H^+$
H_2
e^-
CO
CO
Fe^{2+}
SH
H_2
S^0
H_2S
e^-
CO_2
Ni^{2+}
CO
CO
CO_2
e^-
A

Deep sea vent origin of life

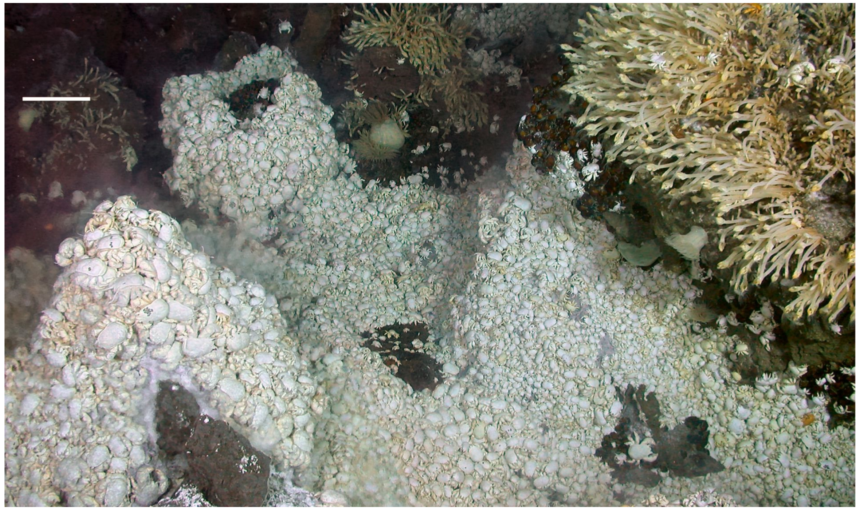


Abiotic carbon fixation in the primitive hydrothermal system.

On the ocean floor, mixing of the hydrothermal fluids and seawater generated sulfide-rich chimneys, and the potential gradient across the chimney drove a continuous electron flow. The electric potential at the chimney-seawater interface could reach less than –1 V (versus SHE) in alkaline hydrothermal vent environments. The low potential, in the presence of sulfides rich in Cd^{2+} and Ag^+, allowed the electrochemical CO_2 reduction to CO with the FE as high as dozens of percent, together with H_2 evolution. The produced CO served as a driving force for the subsequent abiotic organic synthesis that preceded the origin of life as schematically indicated in the figure

Kitadai et al., *Sci. Adv.* **2018**; 4: eaao7265

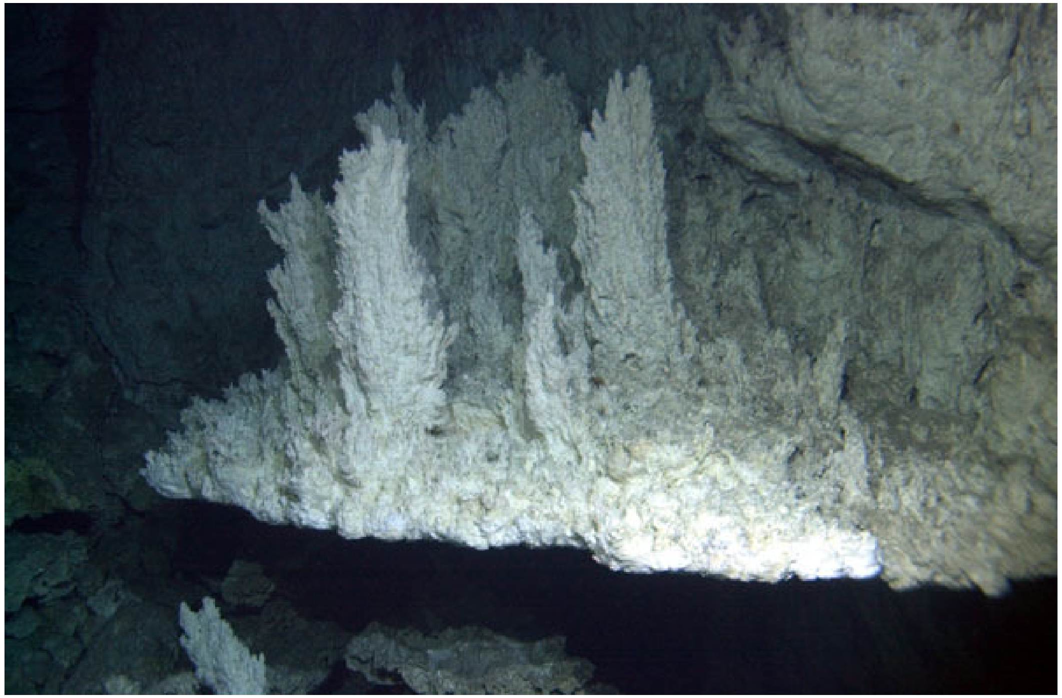
Deep sea vent fauna



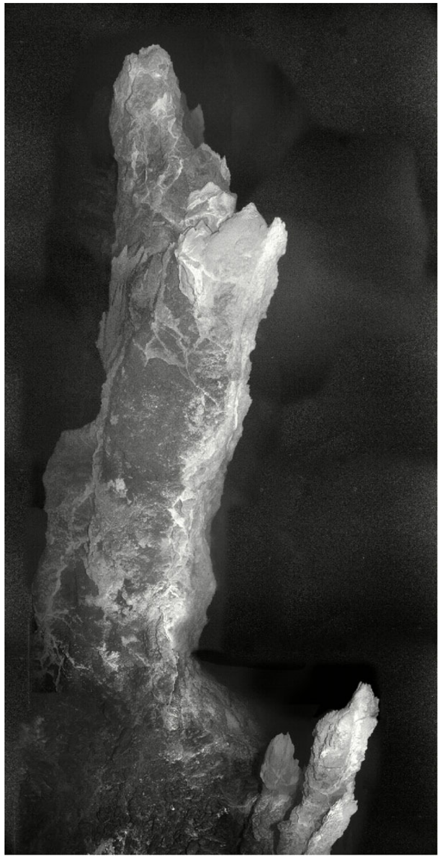
A dense fauna (*Kiwa* anomurans and *Vulcanolepas* like stalked barnacles) near East Scotia Ridge vents

Giant tube worms (*Riftia pachyptila*) cluster around vents in the Galapagos Rift

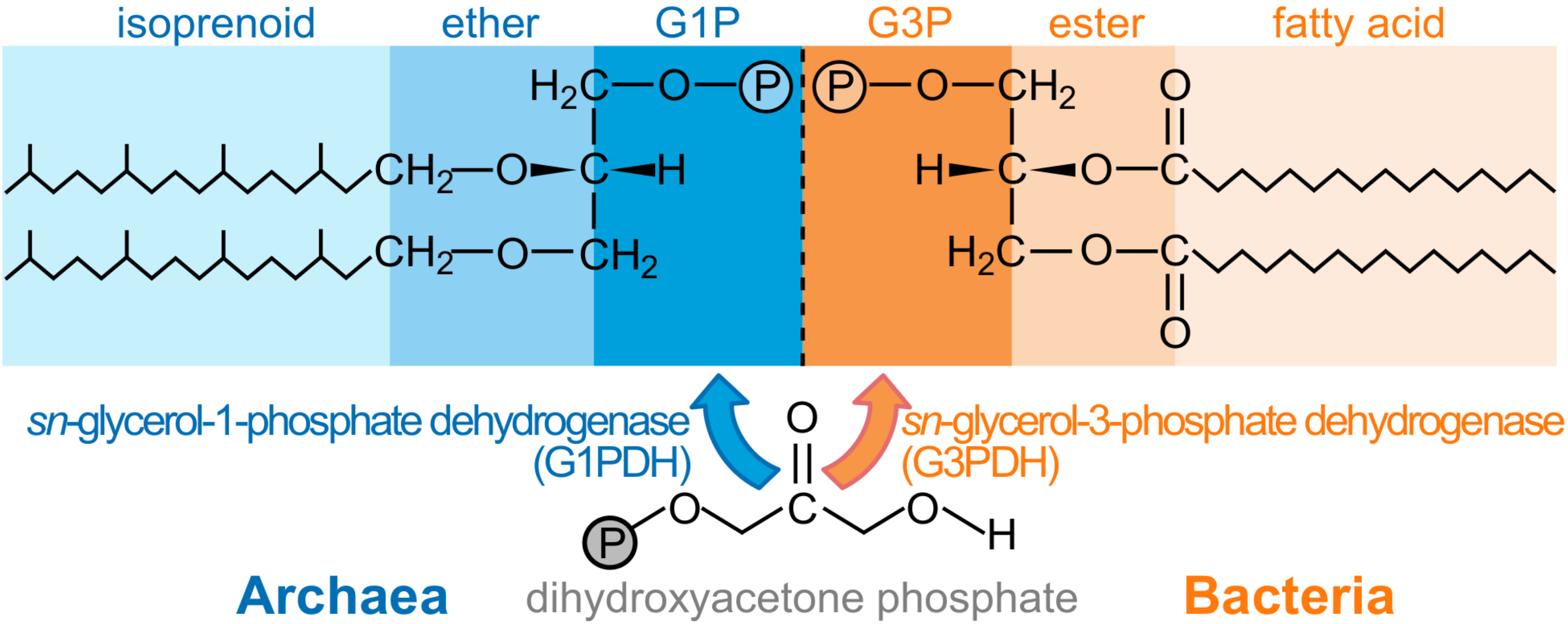
„Lost city" – white smokers: alkaline hydrothermal vents



A 1.5-meter-wide ledge on the side of a chimney is topped with dendritic carbonate growths that form when mineral-rich vent fluids seep through the flange and come into contact with the cold seawater.



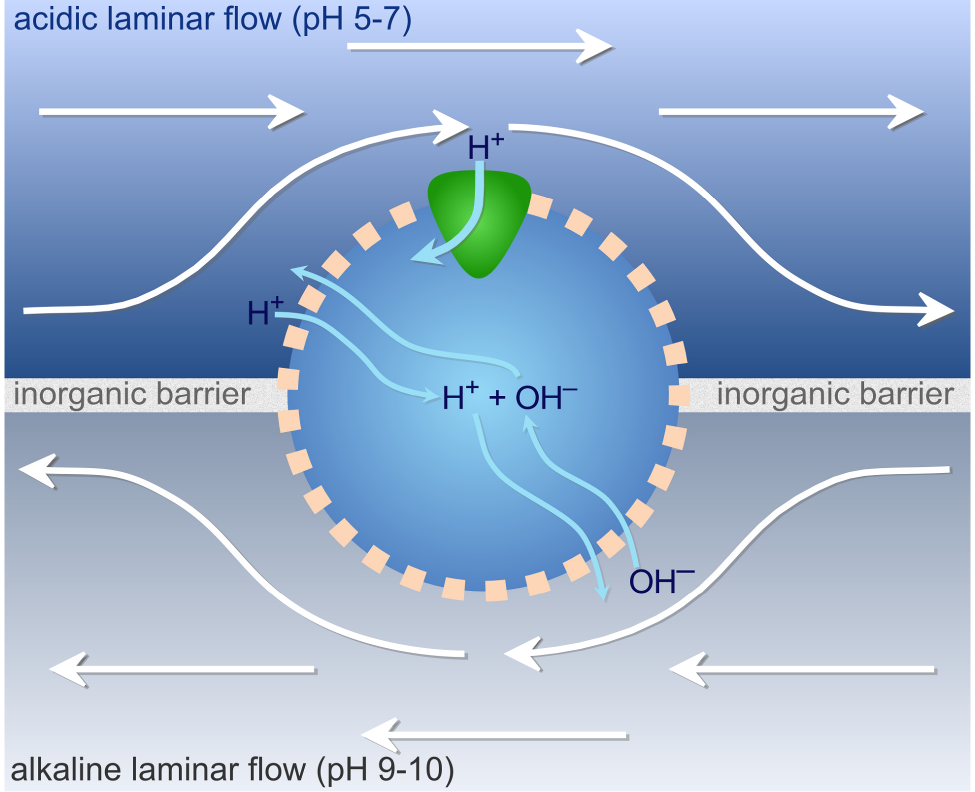
A carbonate chimney more than 9 meters (30 feet) in height. The white, sinuous spine is freshly deposited carbonate material. The top shows evidence of collapse and re-growth, as indicated by the small newly developed cone on its top



Archaeal lipids: isoprenoid chains + ether bonds + *sn*-glycerol-1-phosphate (G1P) backbone.
Bacterial lipids: fatty acids + ester linkage + *sn*-glycerol-3-phosphate (G3P) skeleton.

Despite widespread horizontal gene transfer, no bacterium has been observed with the archaeal enantiomer, or vice versa.
(ether linkages have been observed in bacterial membranes and isoprenoids are common to all three domains)

V. Sojo, A. Pomiankowski, N. Lane *PLOS Biology*, **2014**, *12(8)*, e1001926

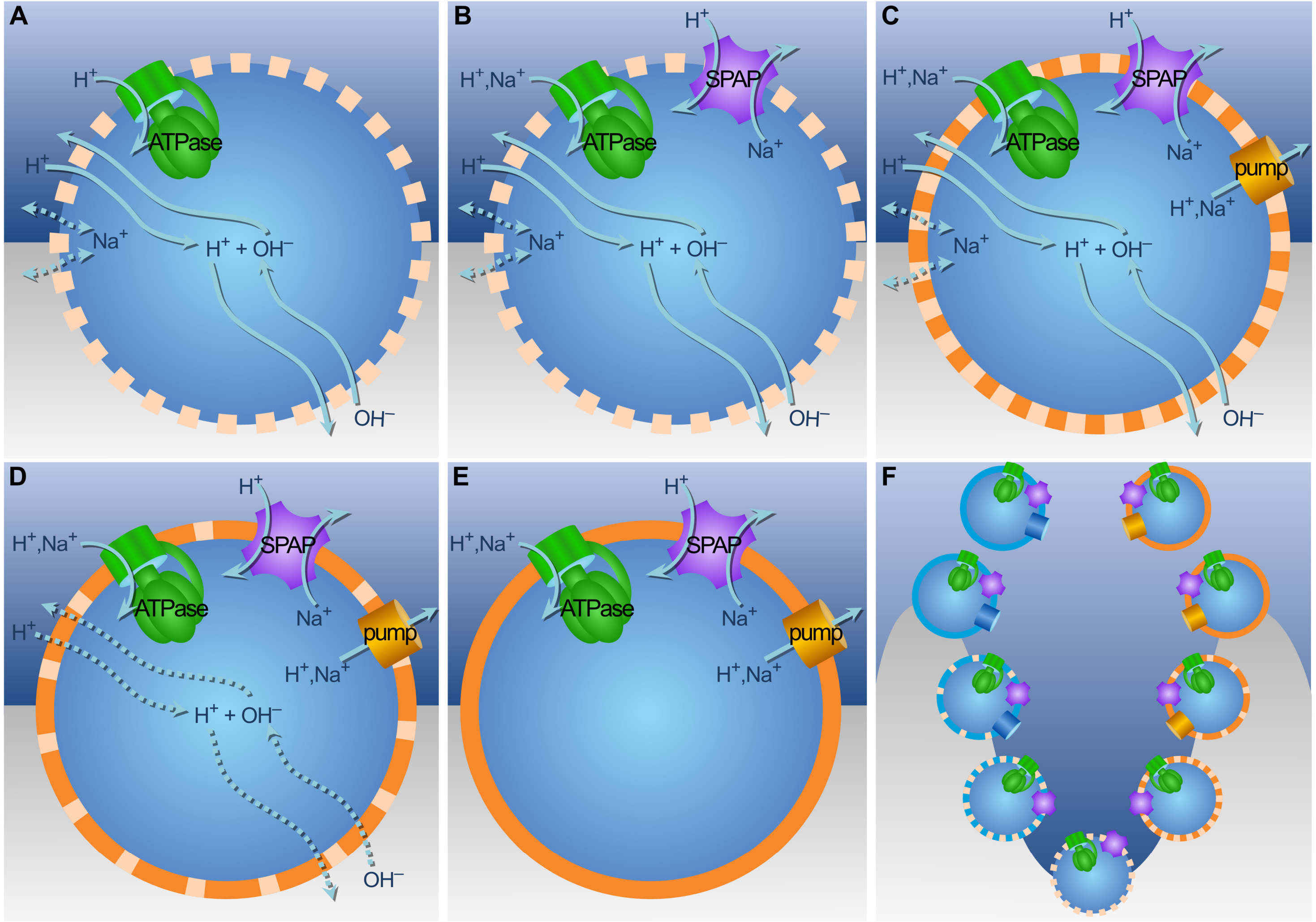


A cell with a semi-permeable membrane at the interface between an alkaline and an acidic fluid (separated elsewhere with an inorganic bareer. H^+, OH^-, Na^+, K^+, Cl^- and other ions flow according to their natural gradients.
Inside the protocell, H^+ and $OH^−$ can neutralize into water, or leave towards either side.

A protein capable of exploiting the natural proton gradient sits on the acidic side, allowing energy assimilation via ATP production, or carbon assimilation via CO_2 fixation.

V. Sojo, A. Pomiankowski, N. Lane *PLOS Biology,* **2014**, *12(8),* e1001926

The role of sodium-proton antiporter (SPAP)



A) H^+ gradient drives energy metabolism (ATPase) or carbon metabolism (Ech)

B) SPAP generates Na^+ from H^+ gradient

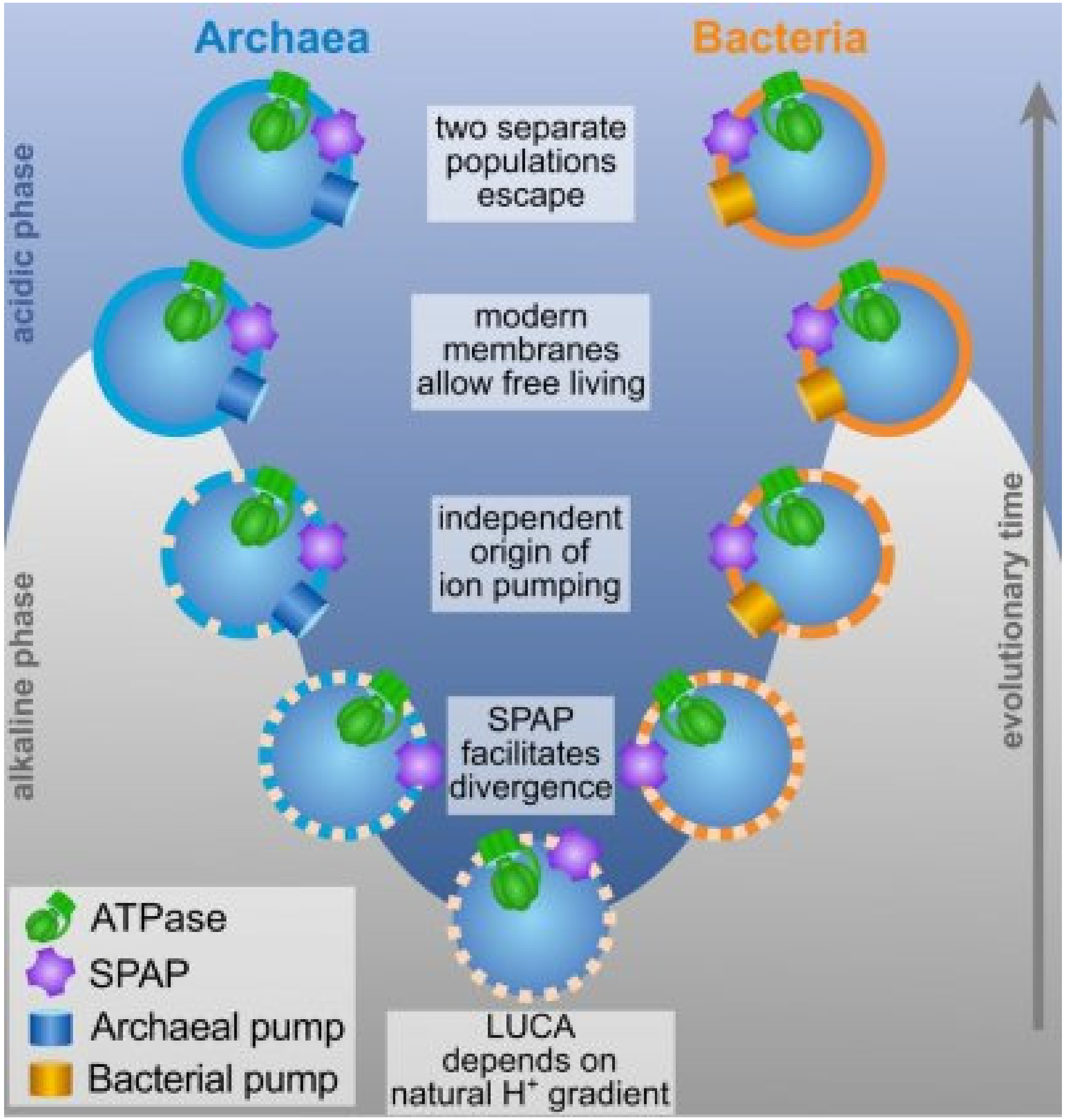
C) Membrane pumps secret H^+ and Na^+

D) Tighter membranes are now produced, to colonize less alkaline environments

E) Impermeable membranes → gradients created by proteins, independently from the natural environmental gradients

F) SPAP favors divergence, selection for active pumping and tighter membranes; independent evolution of archea and bacteria

V. Sojo, A. Pomiankowski, N. Lane
PLOS Biology, **2014**, *12(8)*, e1001926



Origin of autotrophy and development of cell membrane

Ion pumping and phospholipid membranes evolved independently in bacteria and archea.

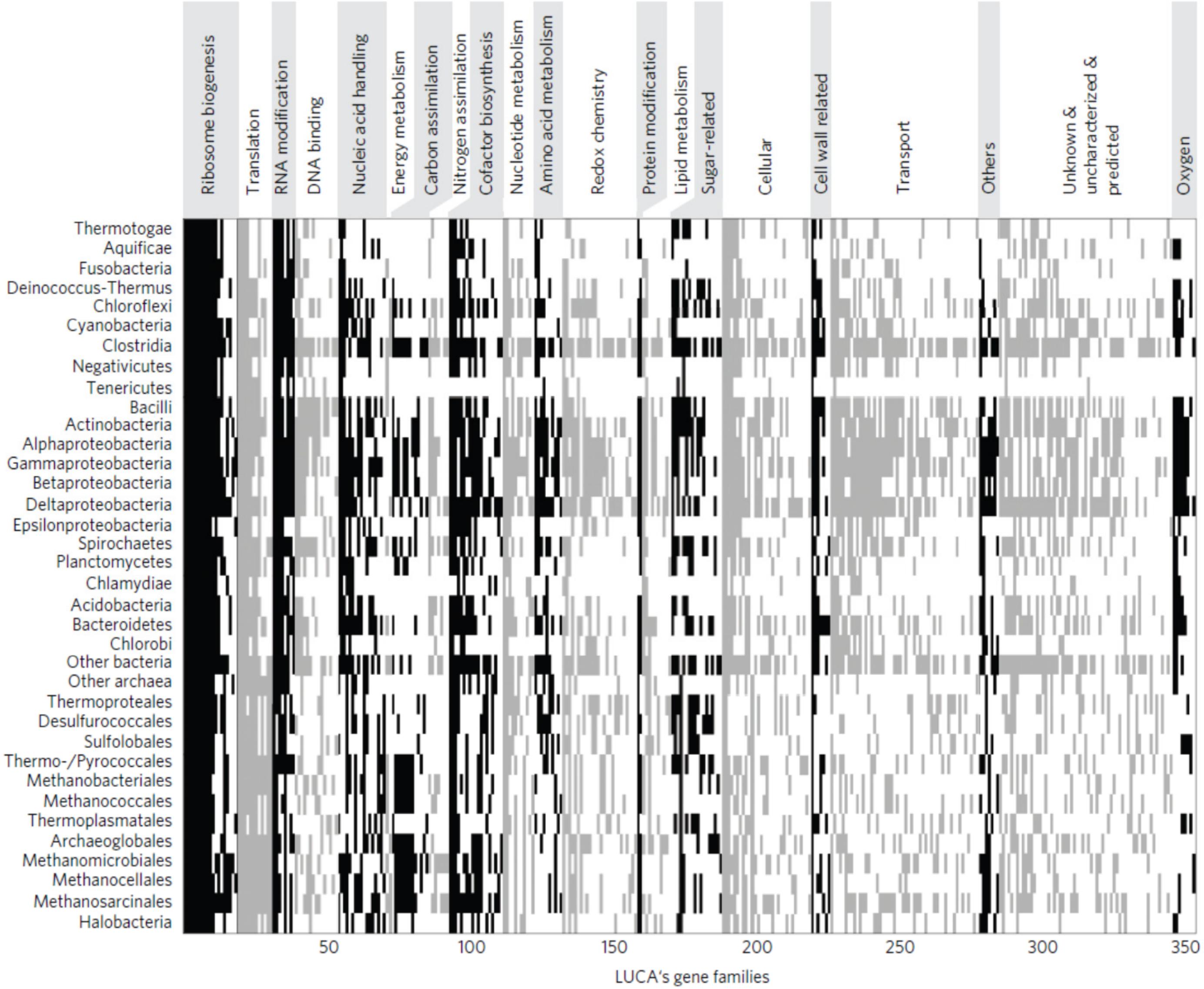
Energy to LUCA could have been delivered by the natural proton gradient in alkaline hydrothermal vents, if the membrane was much more leaky than contemporary ones.

Development of proton pumping allowed for escape from the vent environment.

sodium-proton antiporter (SPAP)

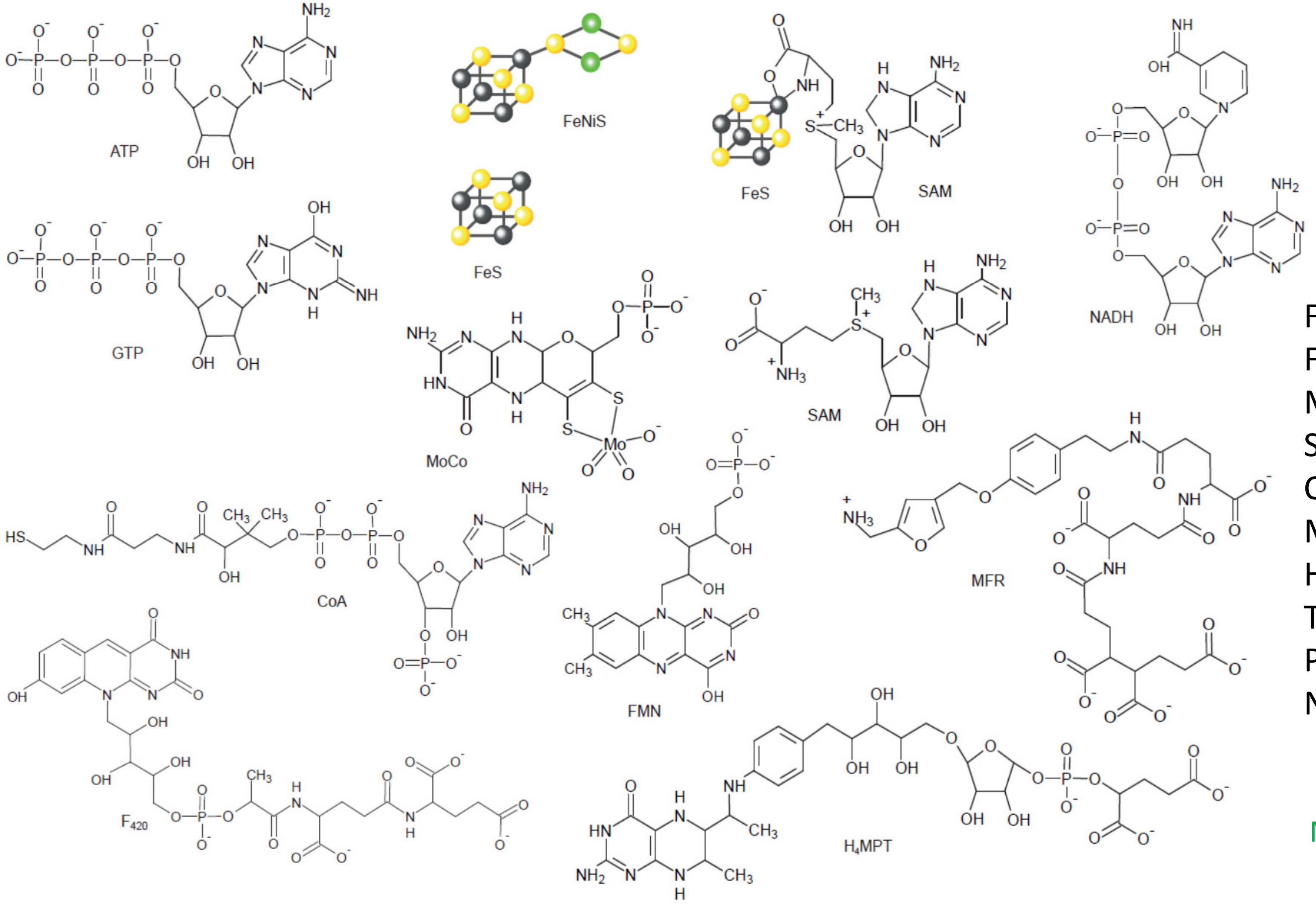
V. Sojo, A. Pomiankowski, N. Lane
PLOS Biology, **2014**, *12(8)*, e1001926

Taxonomic distribution of LUCA's genes grouped by functional categories



M.C. Weiss et al. *Nature Microbiology,*
2016, *Article 16116*

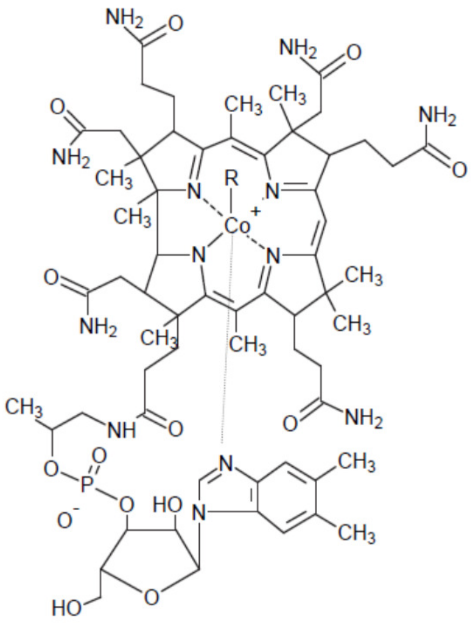
Structures of the cofactors found in LUCA's protein set.



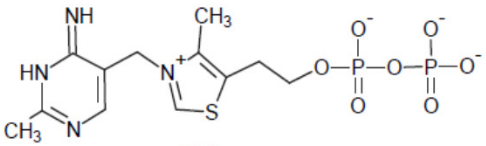
FeNiS – nickel-iron-sulfur cluster
FeS – iron-sulfur cluster
MoCo – molybdenum cofactor
SAM – S-adenosylmethionine
CoA – coenzyme A
MFR – methanofuran
H4MPT – tetrahydromethanopterin
TPP - thiamine pyrophosphate
PLP - pyridoxal phosphate
NTP – nucleoside triphosphate.

M.C. Weiss et al. *Nature Microbiology,*
2016, *Article 16116*

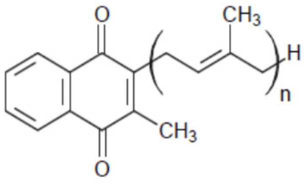
Structures of the cofactors found in LUCA's protein set.



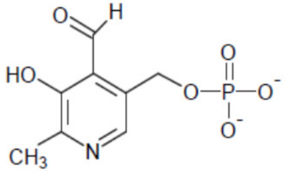
Cobalamin



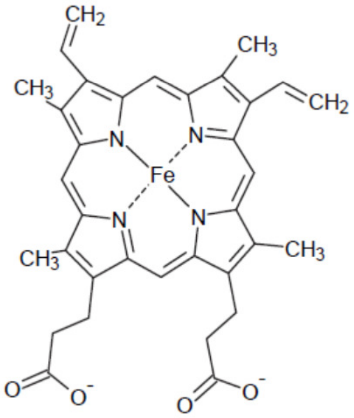
TPP



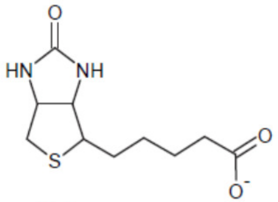
Menaquinone



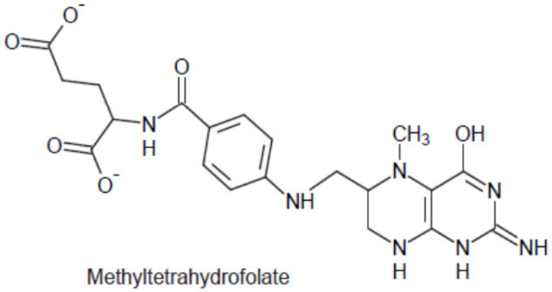
PLP



Heme b



Biotin



Methyltetrahydrofolate

Mononuclear metal centers (Fe and Cu) and the non-standard amino acid selenocysteine are not shown, nor are small protein electron carriers such as ferredoxin or rubredoxin. NTP is also listed as a cofactor, but not shown here as it stands for any of the nucleoside triphosphates in those cases when it's not known which one is bound by the enzyme, or when more than one nucleoside triphosphate can be used

M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*

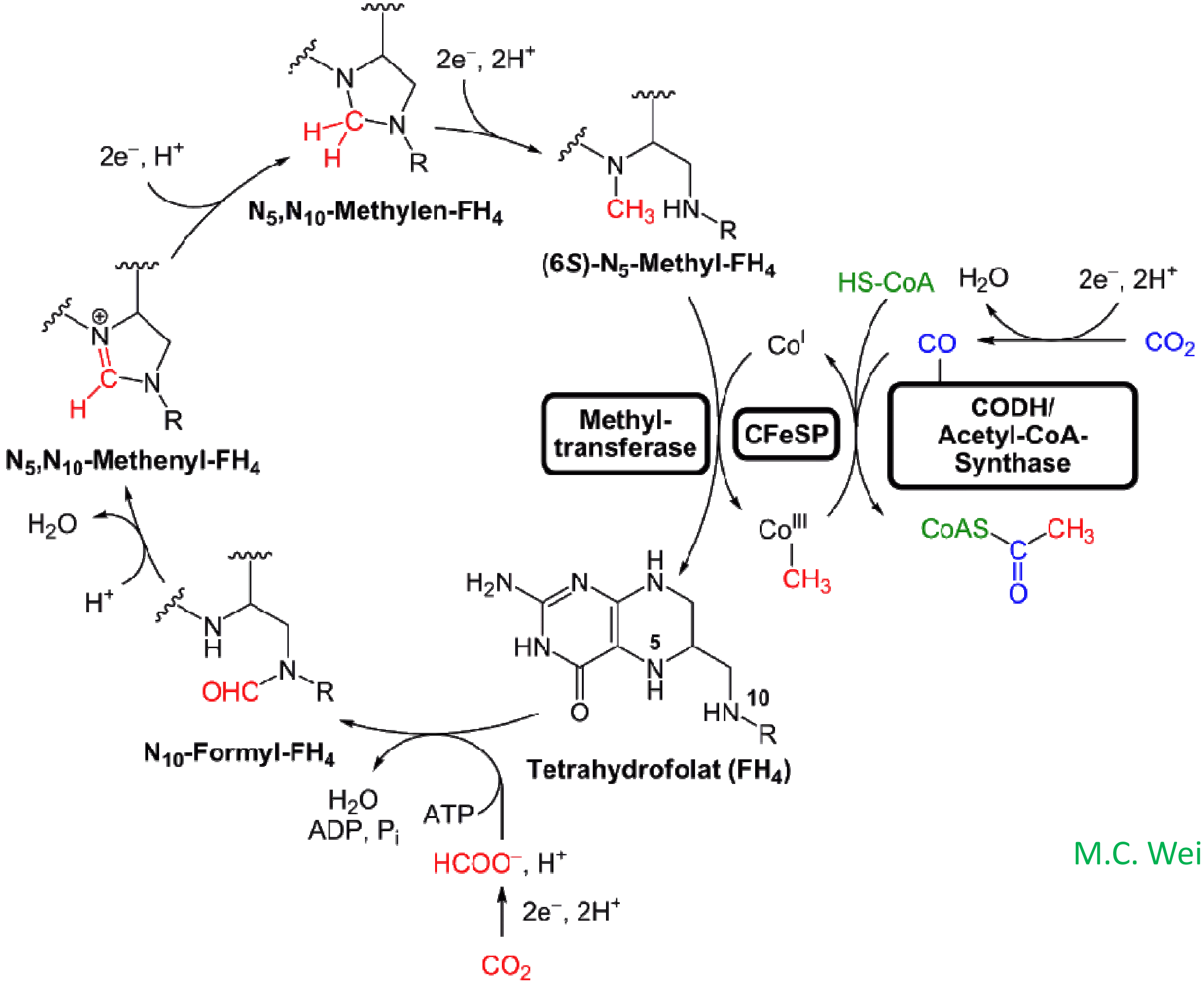
Phylogenetic identification of LUCA's proteome

355 protein families shared among contemporary *archaea* and *bacteria*, including:

- 19 proteins involved in ribosome biogenesis
- 8 aminoacyl tRNA synthethases
- proteins for carbon, energy, and nitrogen metabolism
- rotor-stator ATP synthase subunit (ion gradients were likely supplied geochemically)
- substrate-level phosphorylation (acetylphosphate from acetyl-CoA)
- reverse gyrase – specific for currently living hyperthermophilic organisms
- chemolitoautotrophy enzymes present (WL pathway), chemoorganoautotrophy enzymes absent

M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*

Wood-Ljungdahl (WL) anaerobic pathway of carbon fixation



M.C. Weiss et al. *Nature Microbiology,*
2016, *Article 16116*

A primitive metabolic pathway for carbon fixation, still used by some contemporary chemoautotrophic organisms

Metabolism of LUCA

Among six currently known pathways of CO_2 fixation, only WL pathway was present in LUCA:

The relevant enzymes are packed with FeS and FeNiS centres

They require cofactors: flavin, F_{420}, methanofuran, two pterins and corrins

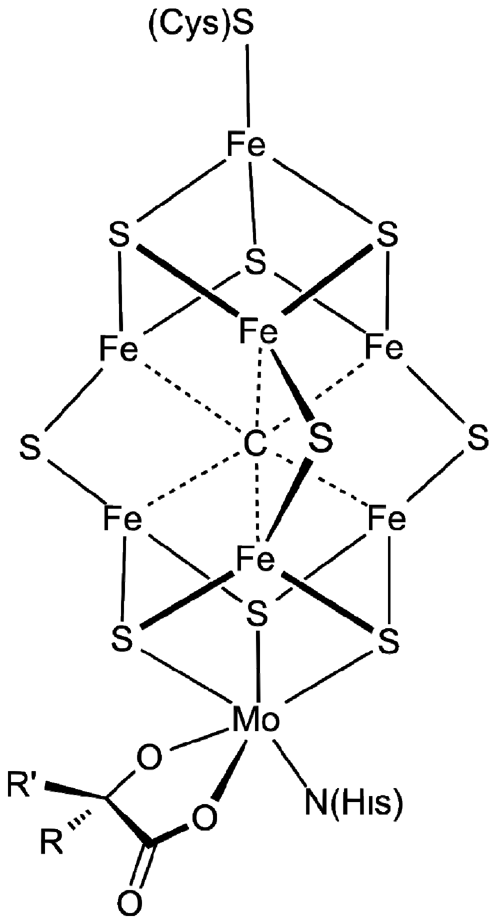
Hydrogenases also present in LUCA's genome → electrons likely obtained from hydrogen, as in modern microbes using the WL pathway

Nitrogenase and glutamine synthethase serve for nitrogen fixation

WL pathway, nitrogenase and hydrogenases are very oxygen-sensitive
→ LUCA was an anaerobic autotroph that could live from gases H_2, CO_2, and N_2.

M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*

Metabolism of LUCA

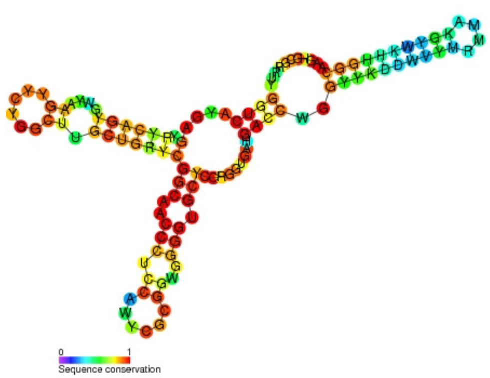


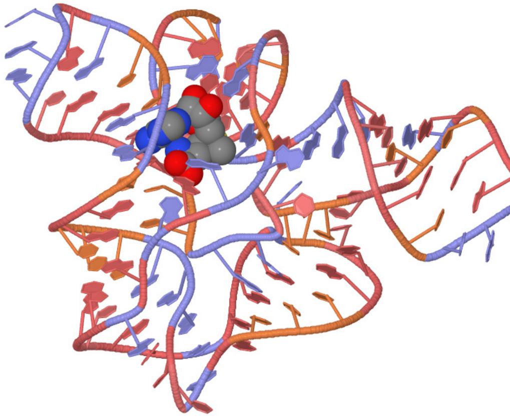
FeMo cofactor of nitrogenase

Enzymes for cofactor biosynthesis, including pterins, MoCo, cobalamin, siroheme, TPP, CoM and F_{420}, are also conserved.

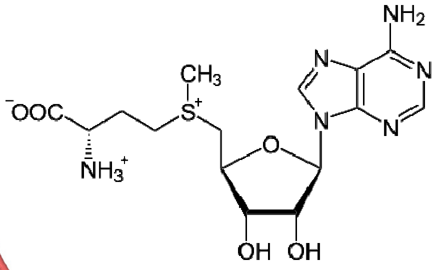
Many of them are S-adenosyl methionine(SAM)-dependent

SAM chemistry is based on oxygen-sensitive FeS-containing proteins that initiate radical-dependent methylations.

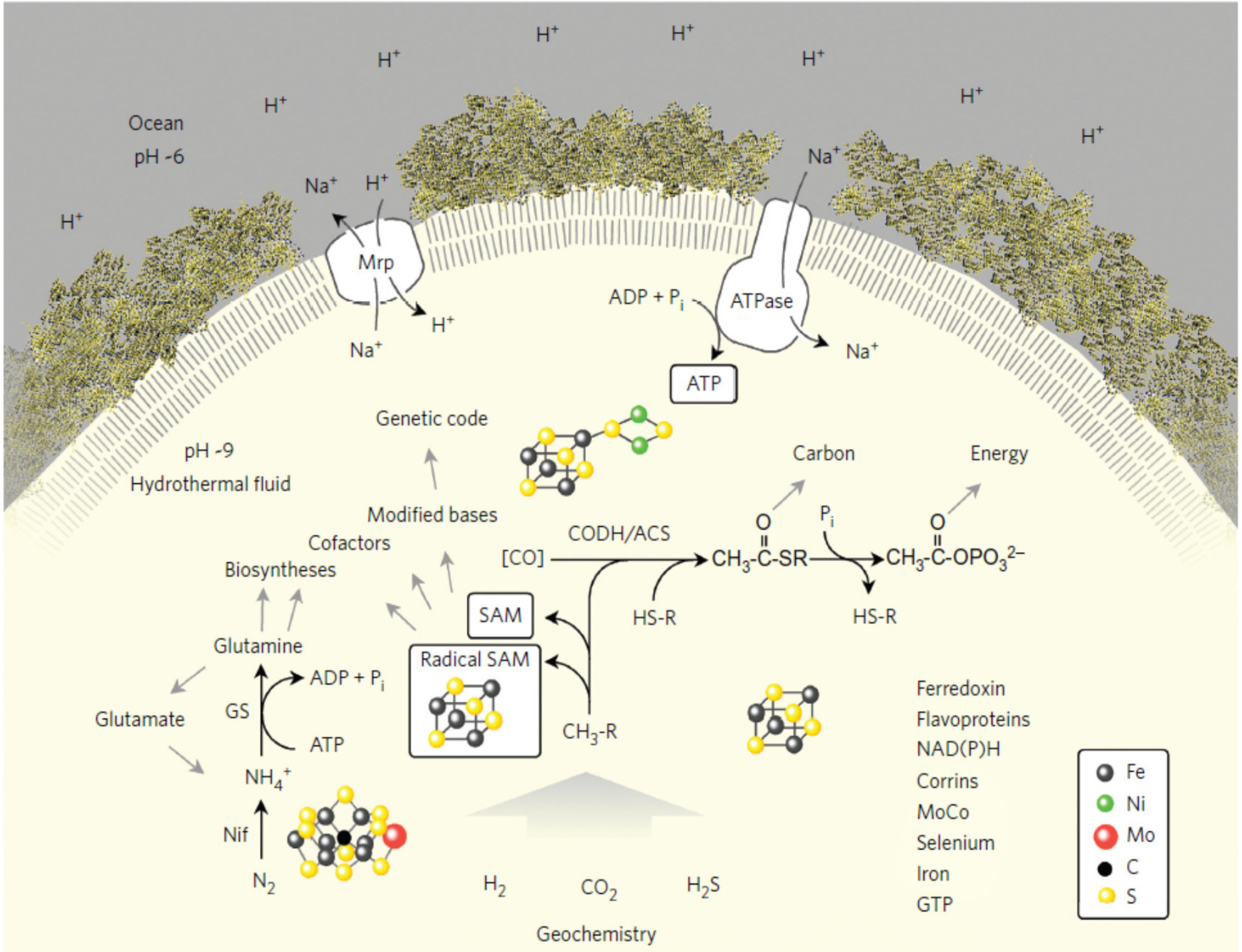




SAM-binding riboswitches



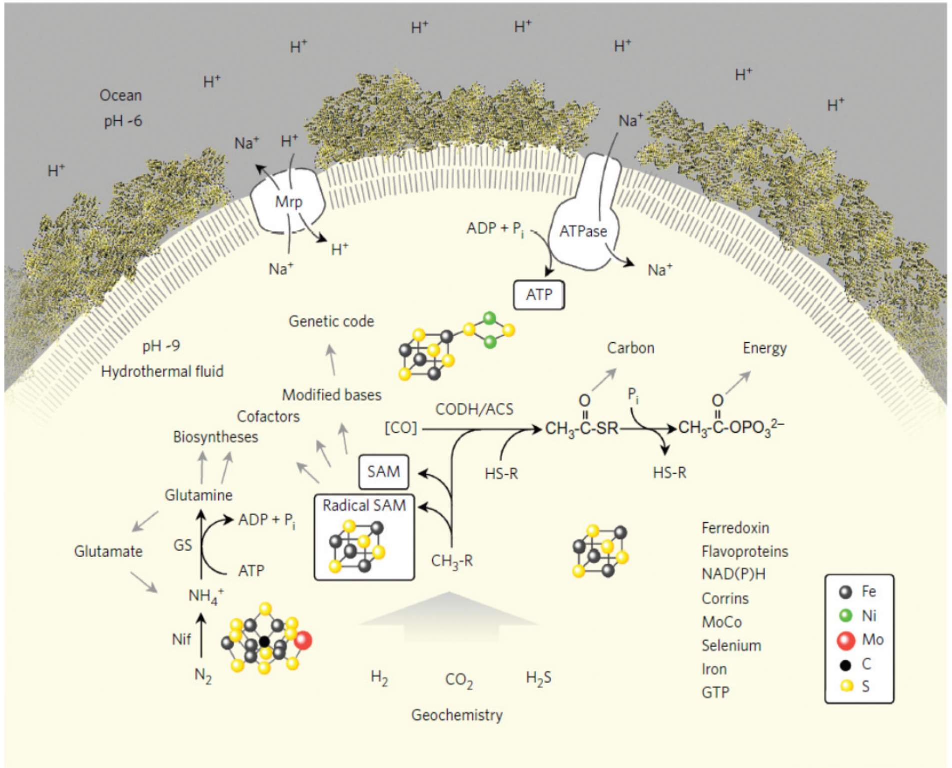
M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*



M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*

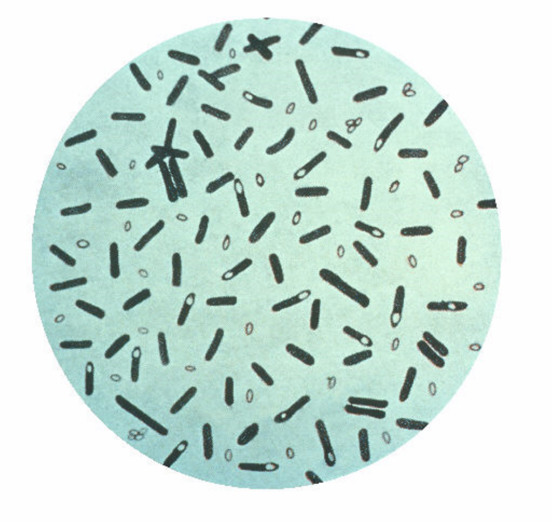
LUCA reconstructed from the genome data

Summary of the main interactions of LUCA with its environment, a vent-like geochemical setting as inferred from genome data.

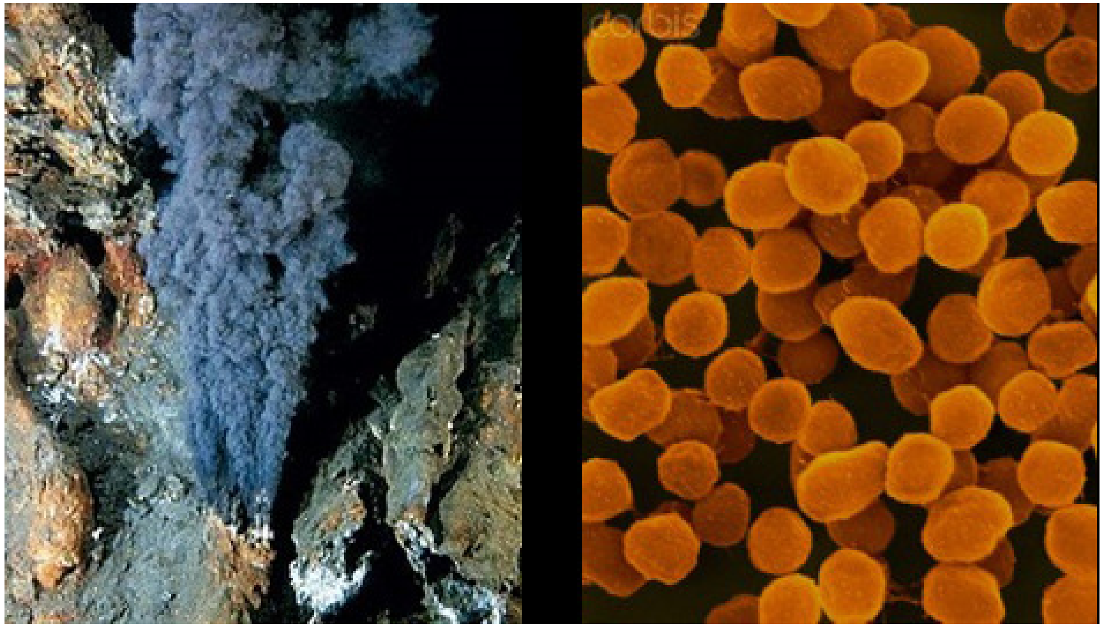


CO source unknown: In modern CODH/ACS complexes, CO is generated from CO_2 and reduced ferredoxin. In primordial metabolism, CO can appear uncatalysed *via* the gas water shift reaction or catalysed *via* transition metals.
A Na^+/H^+ antiporter could transduce a geochemical pH gradient (indicated on the left) inherent in alkaline hydrothermal vents into a more stable Na^+ gradient to feed a primordial Na-dependent ATP synthase.

Closest living relatives of LUCA



clostridia
anaerobic bacteria
(botulin, gangrene, tetanus)



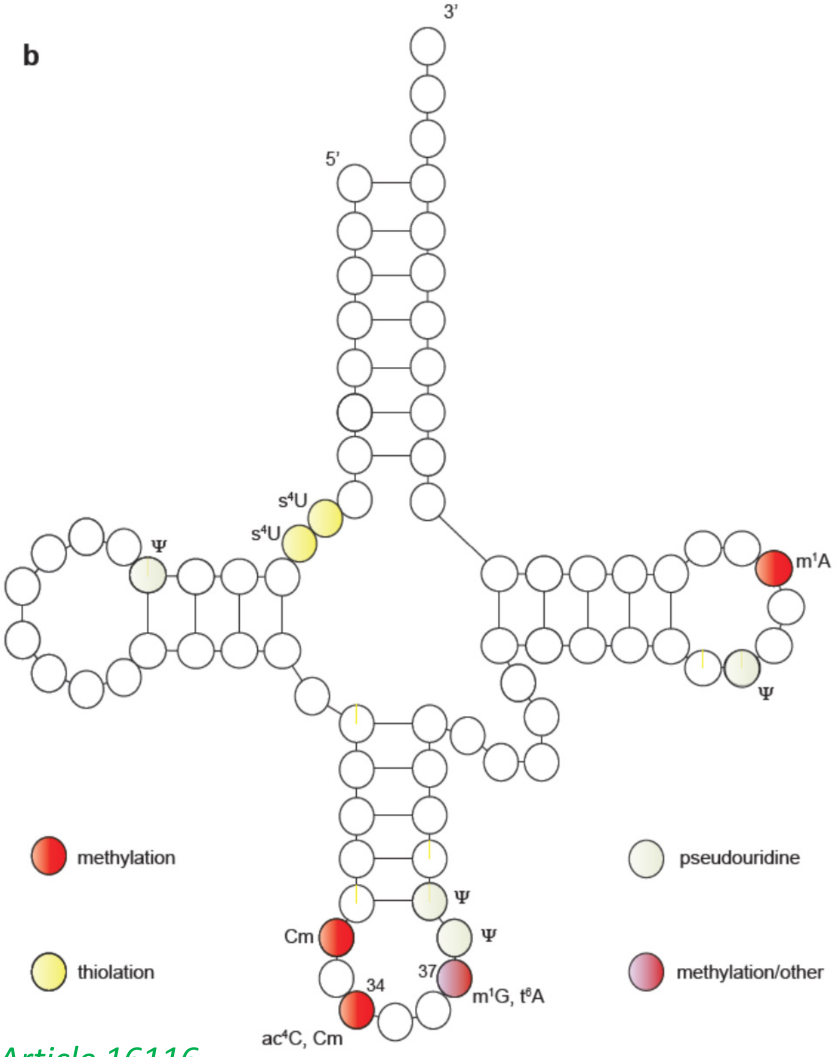
Deep ocean vent - home to the extremophilic archeon
Methanococcus jannaschii

They use the WL pathway, abundant also today, some species can live from methyl groups (methane gas on marshes andwetlands), and they depend on H_2 (from geology or H_2-producing fermentation)

Geological source of hydrogen: serpentinization (iron + hot water, anoxic) $Fe^{2+} + H_2O \rightarrow Fe_3O_4 + H_2$

Modified nucleosides and the genetic code

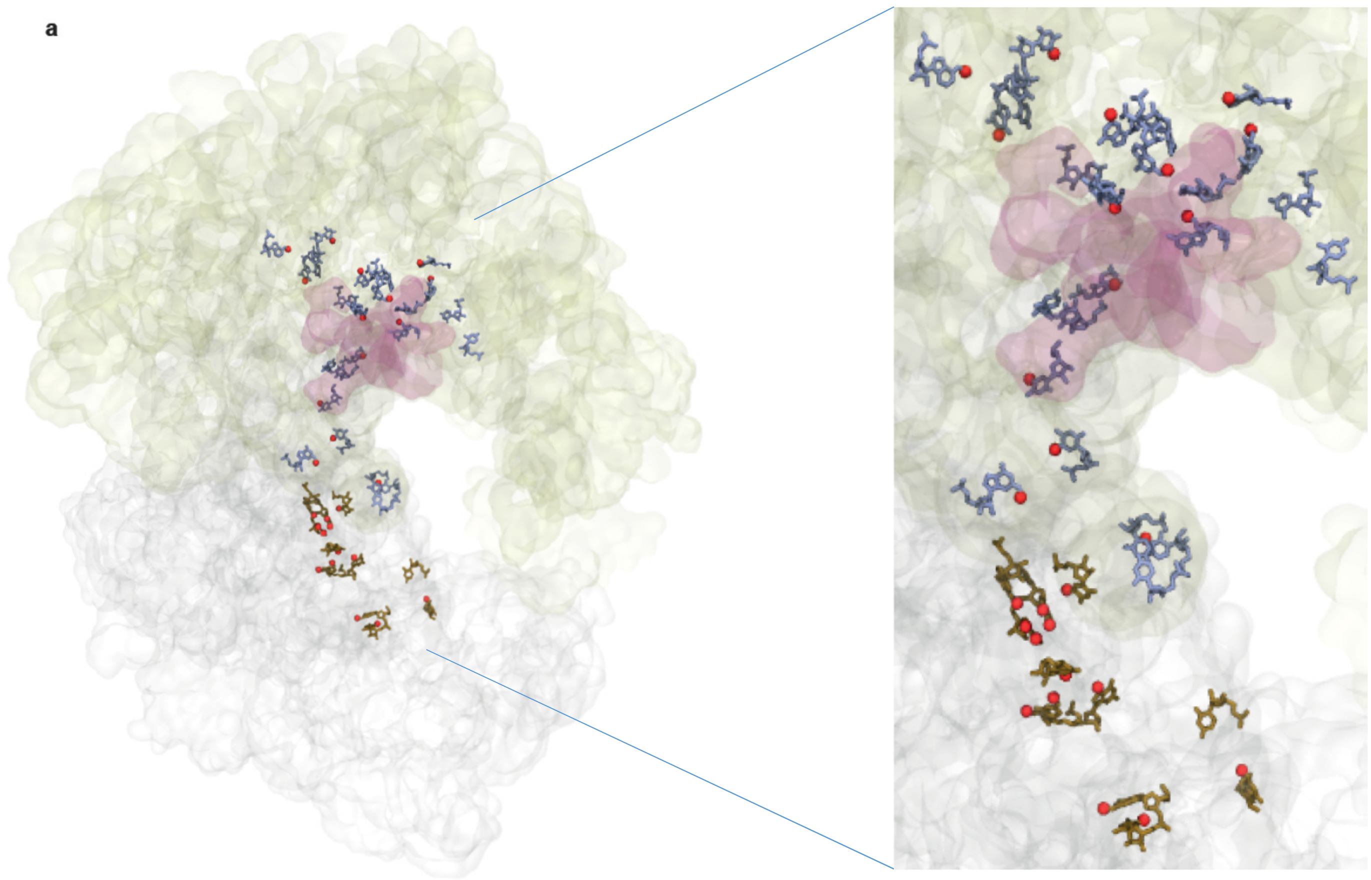
LUCA had also genes involved in RNA nucleoside modifications (mainly methylations and thiomethylations) still required today e.g. for the anticodon recognition process.



Cloverleaf secondary structure representation of tRNA showing only those posttranscriptional nucleoside modifications that are conserved among bacteria and archaea in both identity and position. (5-methoxyuridine at position 34 in archaea has been disputed).

M.C. Weiss et al. *Nature Microbiology,* **2016**, *Article 16116*

Modified nucleosides and the genetic code



M.C. Weiss et al. *Nature Microbiology,*
2016, *Article 16116*

Structure of the E. coli ribosome (PDB ID: 4YBB), with the large and small subunits shown in green and silver, respectively. The peptidyl-transferase site is shaded pink. The modified nucleosides of 23S rRNA are depicted in icy blue, while in 16S rRNA they are ochre. Modification of C2501 to 5-hydroxycytidine is not present in the structure. Methyl group carbons are shown as red balls.

Transition from the RNA world to LUCA

Ribozymes – self-acting → metabolic

Evolution of ribosome

Incorporation of aminoacids and peptides

The genetic code and archival storage

Enzyme-driven metabolism and membranes

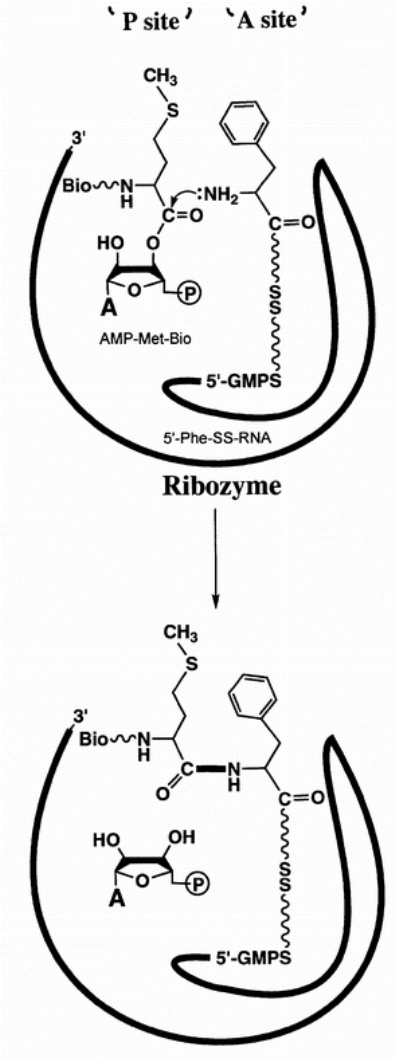
Ribozymes

Initially only self-processing ribozymes (introns, RNAses) discovered.

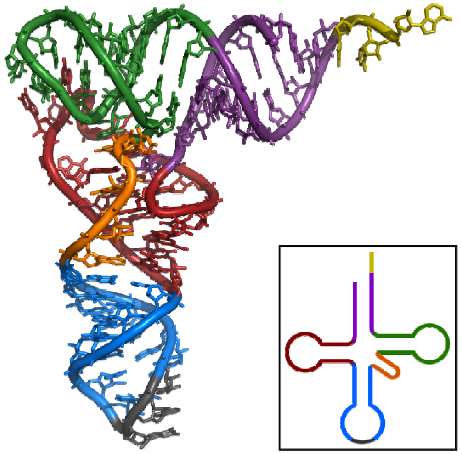
1992 – first ribozyme isolated capable to cleave the bond of methionine with its tRNA (also the reverse reaction – transacylation – is catalysed)

1995 (Yarus) – a random RNA sequence found capable of attaching an activated aminoacid to itself

1997 (Szostak) – an RNA sequence that transfers one aminoacid to another one, forming a dipeptide → analogue of the peptidetransferase center of the ribosome



Ribozymes



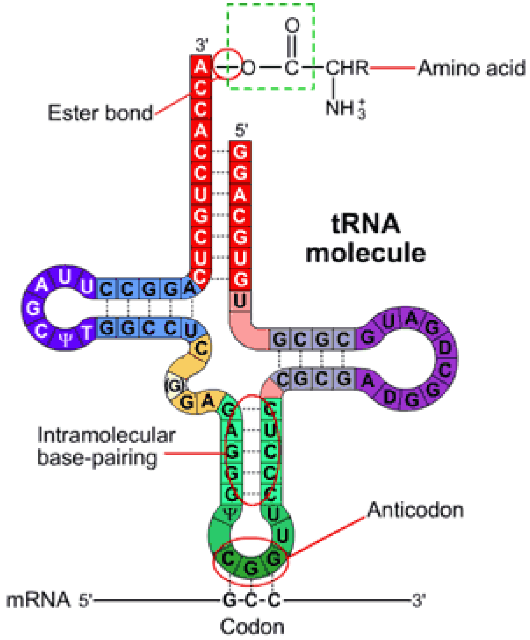
Ribozymes incorporate aminoacids to enhance their catalytic abilities

It opens ways to improved metabolism and provides evolutionary advantage in receiving energy from outside

Initially incorporation of aminoacids may have improved synthesis of nucleotides to produce more RNA

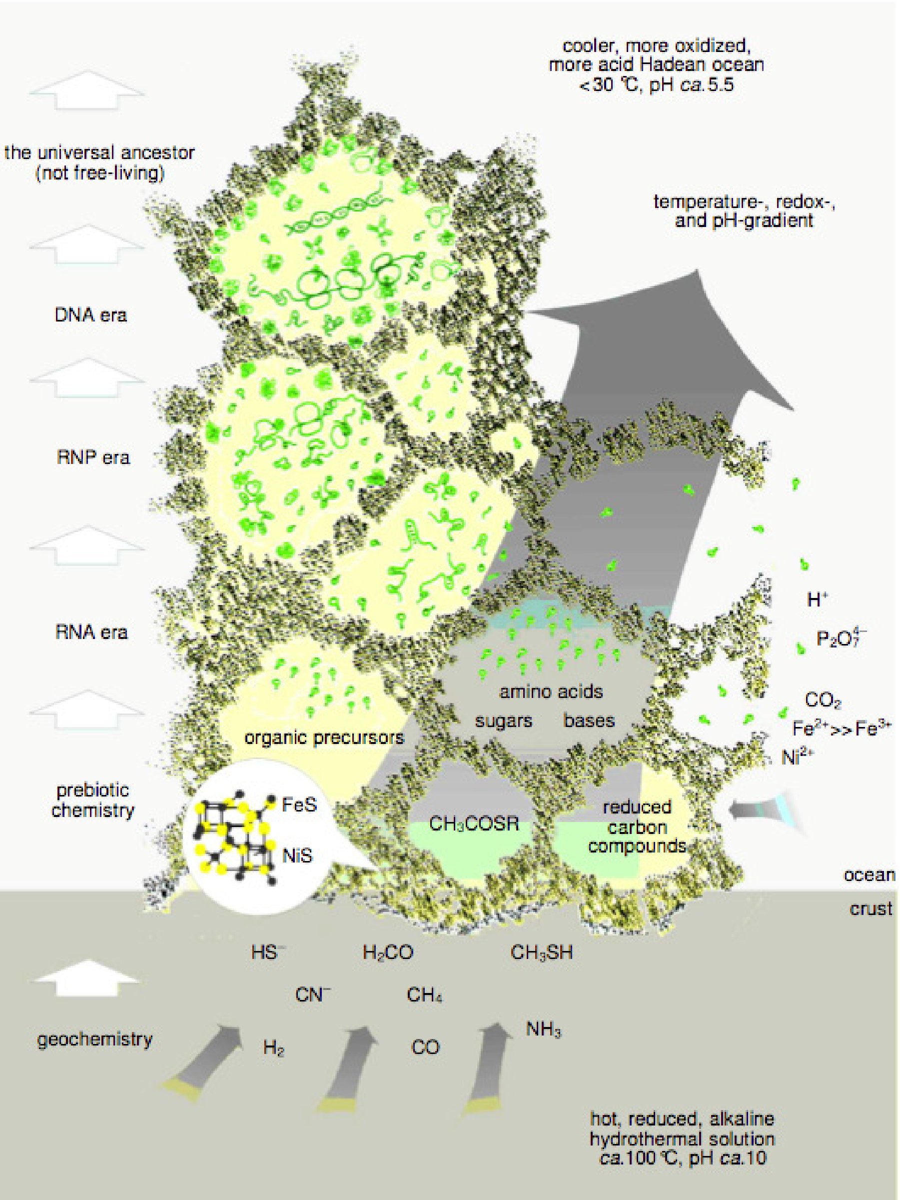
Primordial tRNAs were most likely self-charging, today special enzymes do it (tRNA synthethases)

Peptide chains increase in size, the RNA part decrease → non-covalent binding of nucleoside cofactors to contemporary enzymes



aminoacyl-tRNA

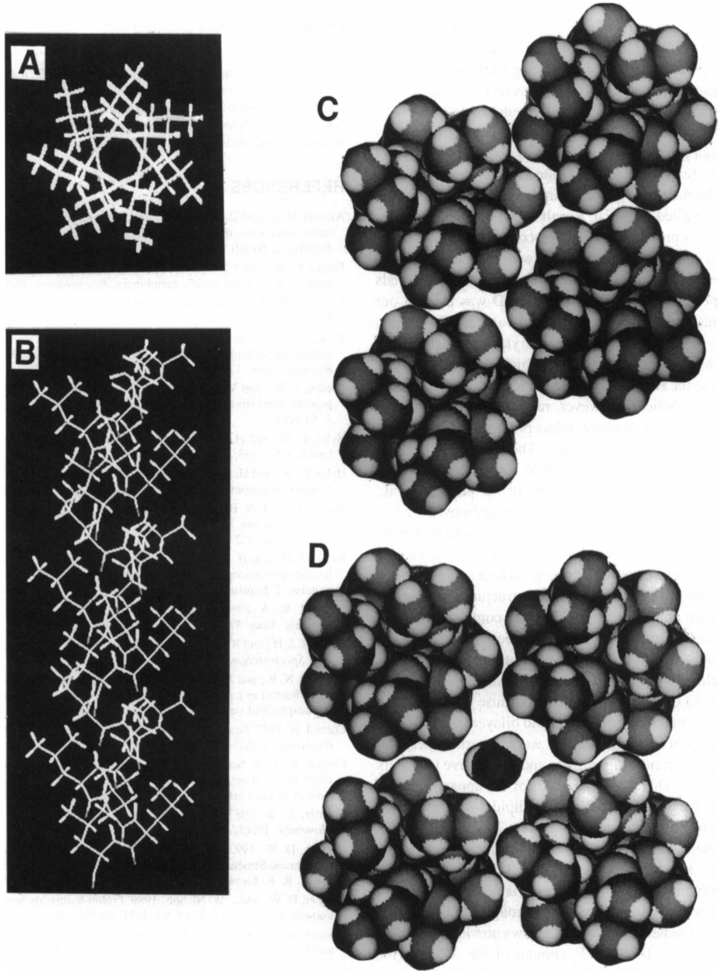
Increasing metabolic complexity



Complex metabolic machinery closed in the same compartment that genetic polymers (RNA) which generated it.

We don't see ribozyme-based metabolism today anymore, because protein catalysts (enzymes) for the same reactions are orders of magnitude faster than the ribozymes

Evolutionary advantage of proteins

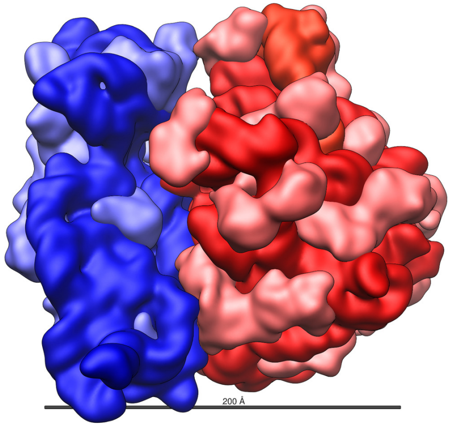


Poly-alanine and poly-leucine form ion channels that selectively transport protons across lipid bilayers (not Na^+ or K^+)

Short peptides with polar positively charged end (arginines) and unpolar Leu/Phe/Trp drive RNAs to membranes (Szostak)

A. E. Olivier, D. E. Deamer, *Biophysical Journal* **1994**, *66*, 1364-1379

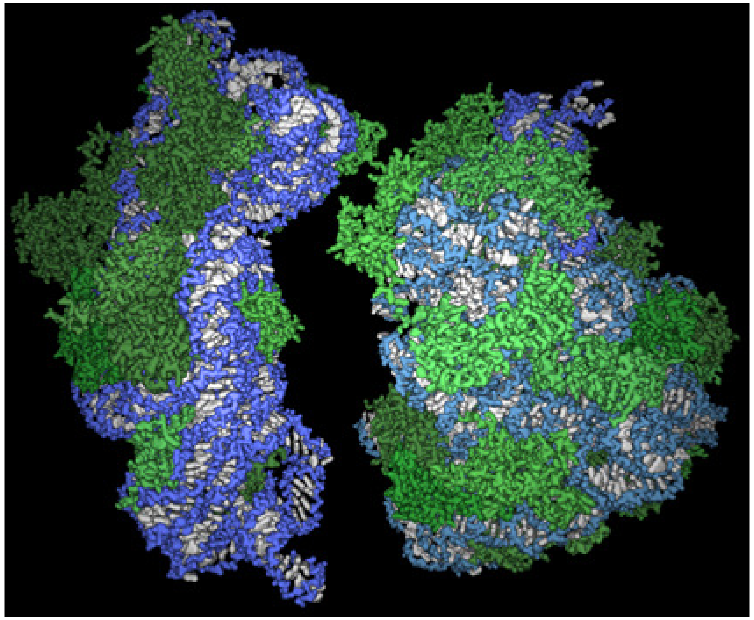
Ribosome



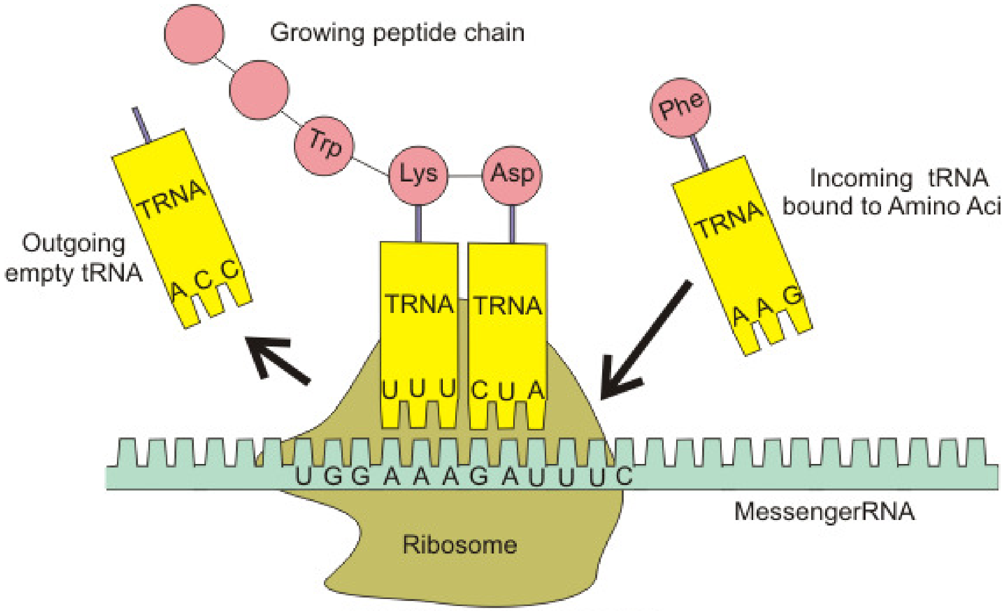
Every protein component of the ribosome can be removed without loosing the activity.

Sequence-specific synthesis of proteins was invented late

Initially large subunit catalysed transacylations, later the small subunit used another RNA strand to ‚guide' the new peptide growth in a sequence-specific manner (by codon-anticodon recognition). This strand (proto-mRNA) allowed tighter binding.

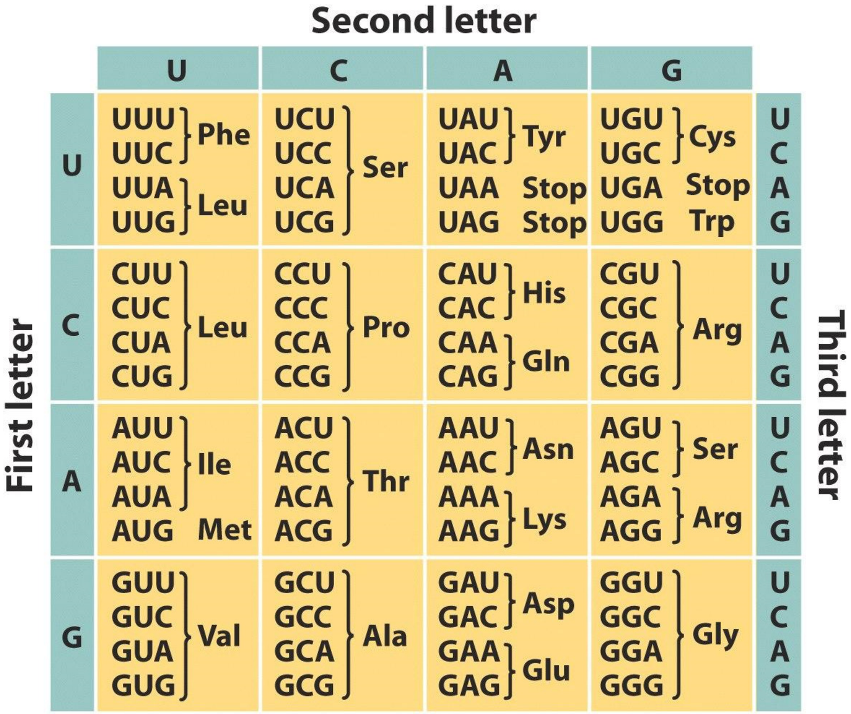


Ribosome: green - proteins, blue and white - RNA

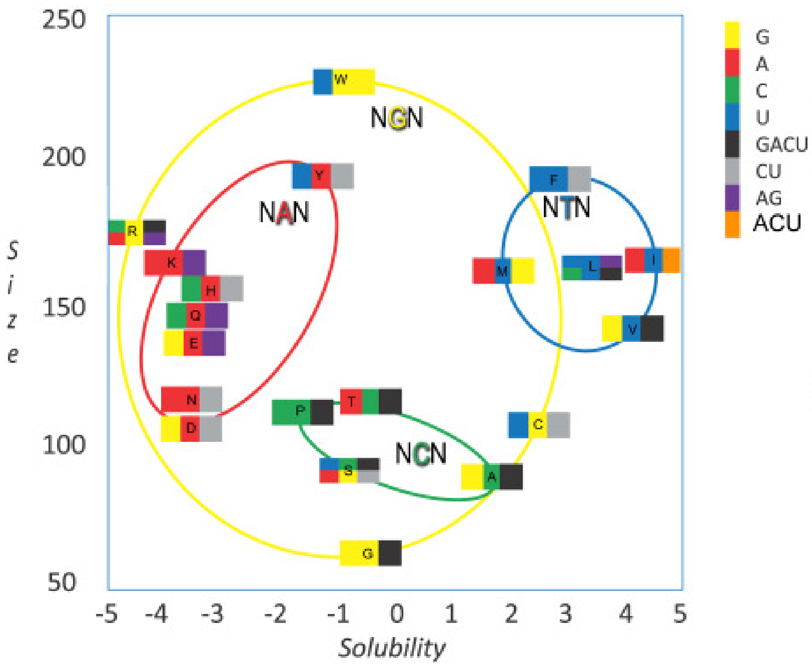


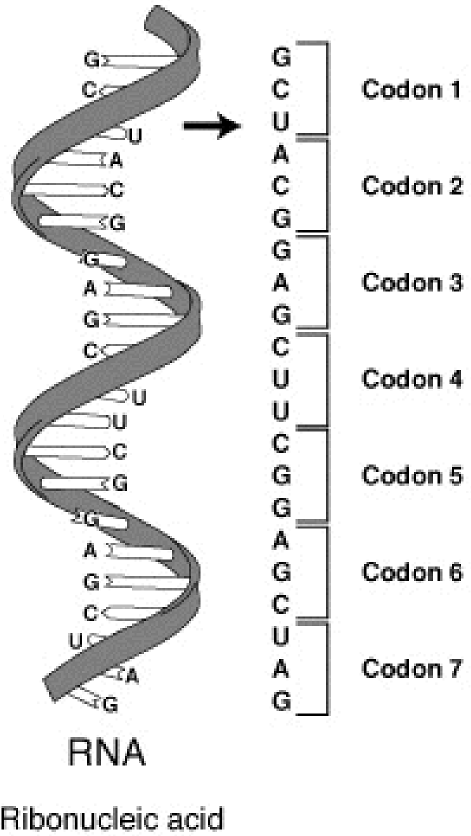
Peptide Synthesis

The genetic code

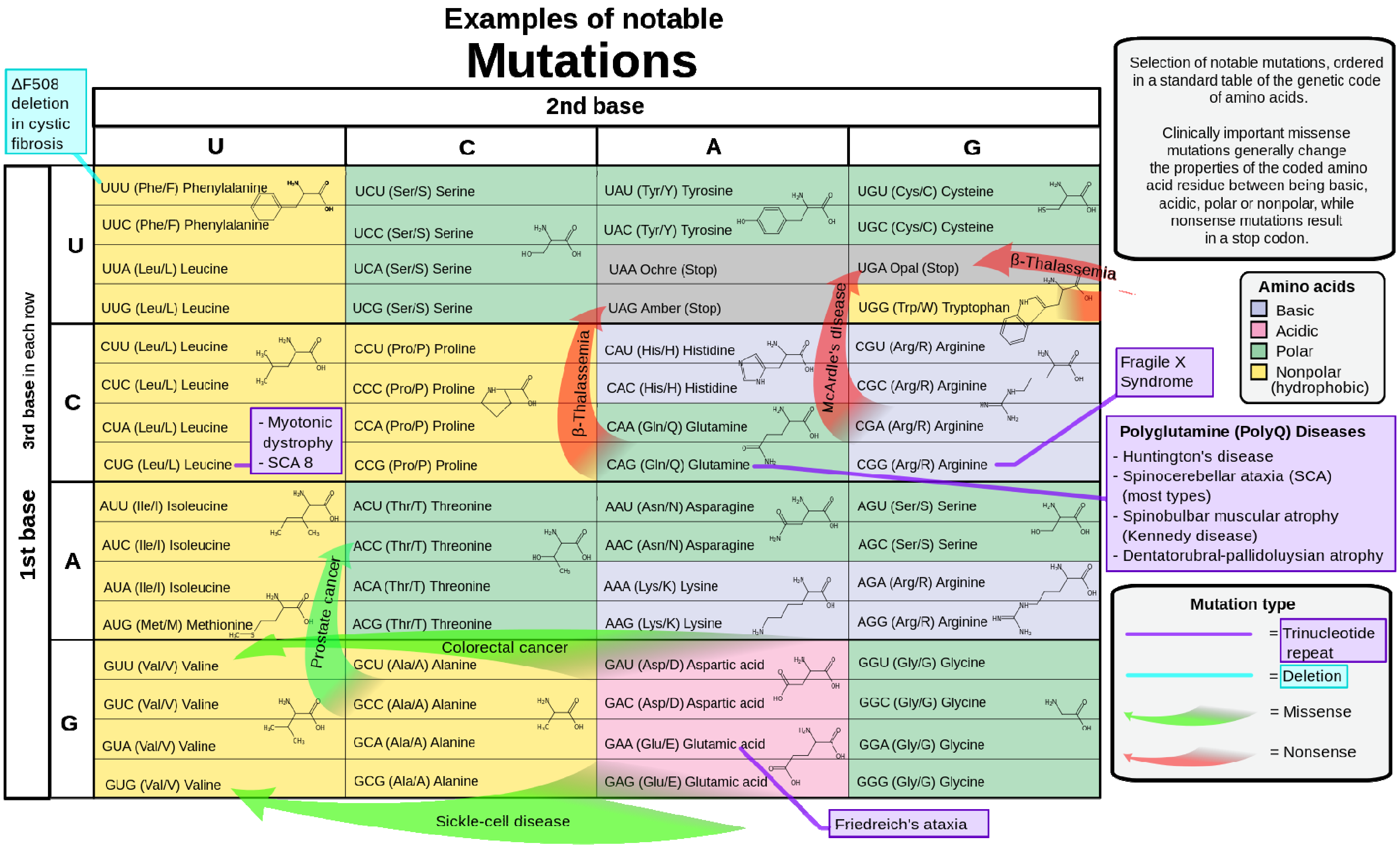


First letter	Second letter: U		C		A		G		Third letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	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		GAG		GGG		G





The genetic code



Examples of notable Mutations

1st base	3rd base in each row	2nd base U	2nd base C	2nd base A	2nd base G
U	U	UUU (Phe/F) Phenylalanine	UCU (Ser/S) Serine	UAU (Tyr/Y) Tyrosine	UGU (Cys/C) Cysteine
	C	UUC (Phe/F) Phenylalanine	UCC (Ser/S) Serine	UAC (Tyr/Y) Tyrosine	UGC (Cys/C) Cysteine
	A	UUA (Leu/L) Leucine	UCA (Ser/S) Serine	UAA Ochre (Stop)	UGA Opal (Stop)
	G	UUG (Leu/L) Leucine	UCG (Ser/S) Serine	UAG Amber (Stop)	UGG (Trp/W) Tryptophan
C	U	CUU (Leu/L) Leucine	CCU (Pro/P) Proline	CAU (His/H) Histidine	CGU (Arg/R) Arginine
	C	CUC (Leu/L) Leucine	CCC (Pro/P) Proline	CAC (His/H) Histidine	CGC (Arg/R) Arginine
	A	CUA (Leu/L) Leucine	CCA (Pro/P) Proline	CAA (Gln/Q) Glutamine	CGA (Arg/R) Arginine
	G	CUG (Leu/L) Leucine	CCG (Pro/P) Proline	CAG (Gln/Q) Glutamine	CGG (Arg/R) Arginine
A	U	AUU (Ile/I) Isoleucine	ACU (Thr/T) Threonine	AAU (Asn/N) Asparagine	AGU (Ser/S) Serine
	C	AUC (Ile/I) Isoleucine	ACC (Thr/T) Threonine	AAC (Asn/N) Asparagine	AGC (Ser/S) Serine
	A	AUA (Ile/I) Isoleucine	ACA (Thr/T) Threonine	AAA (Lys/K) Lysine	AGA (Arg/R) Arginine
	G	AUG (Met/M) Methionine	ACG (Thr/T) Threonine	AAG (Lys/K) Lysine	AGG (Arg/R) Arginine
G	U	GUU (Val/V) Valine	GCU (Ala/A) Alanine	GAU (Asp/D) Aspartic acid	GGU (Gly/G) Glycine
	C	GUC (Val/V) Valine	GCC (Ala/A) Alanine	GAC (Asp/D) Aspartic acid	GGC (Gly/G) Glycine
	A	GUA (Val/V) Valine	GCA (Ala/A) Alanine	GAA (Glu/E) Glutamic acid	GGA (Gly/G) Glycine
	G	GUG (Val/V) Valine	GCG (Ala/A) Alanine	GAG (Glu/E) Glutamic acid	GGG (Gly/G) Glycine

ΔF508 deletion in cystic fibrosis

- Myotonic dystrophy
- SCA 8

β-Thalassemia

McArdle's disease

β-Thalassemia

Fragile X Syndrome

Prostate cancer

Colorectal cancer

Sickle-cell disease

Friedreich's ataxia

Selection of notable mutations, ordered in a standard table of the genetic code of amino acids.

Clinically important missense mutations generally change the properties of the coded amino acid residue between being basic, acidic, polar or nonpolar, while nonsense mutations result in a stop codon.

Amino acids

🟦 Basic
🟥 Acidic
🟩 Polar
🟨 Nonpolar (hydrophobic)

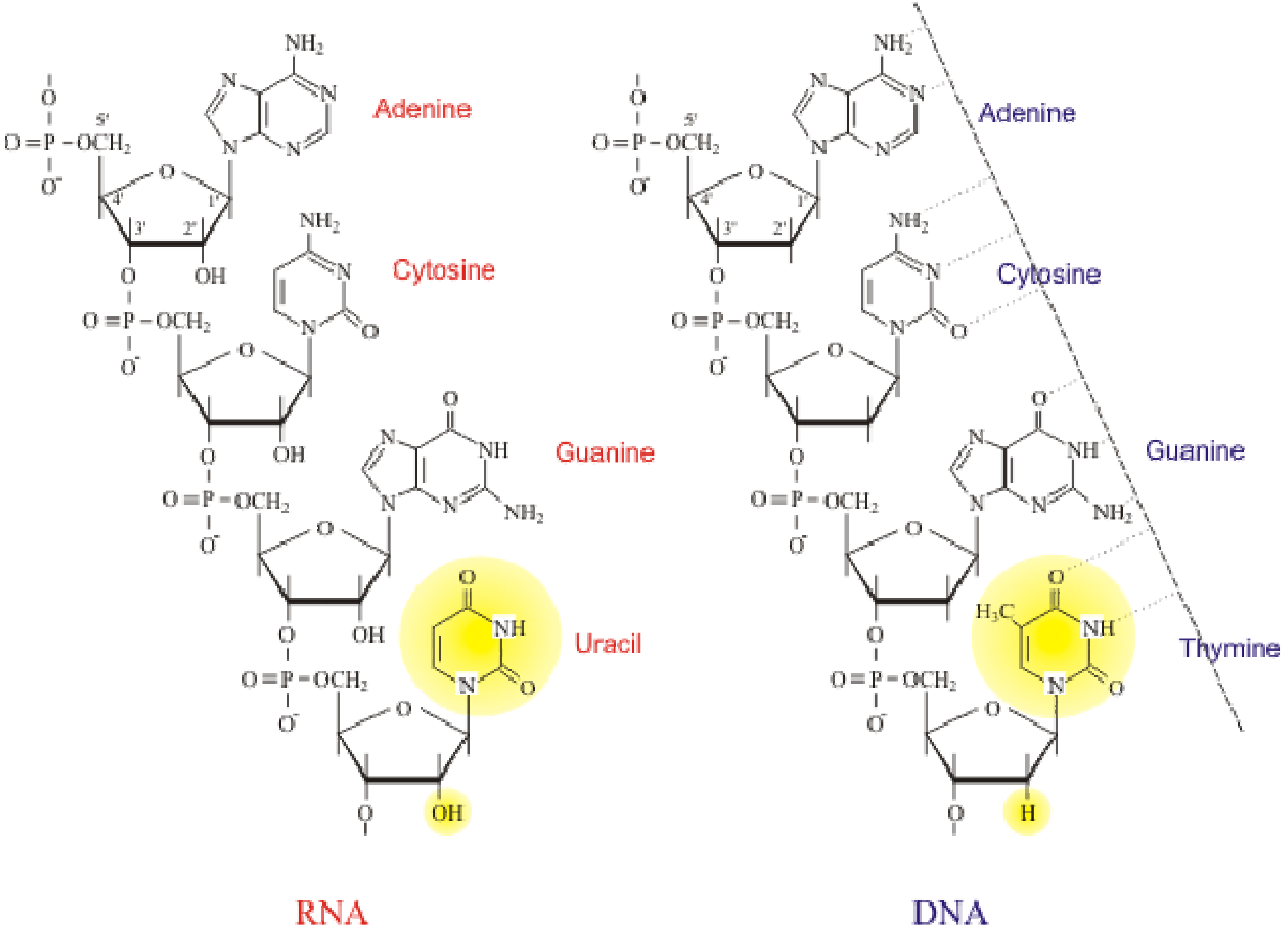
Polyglutamine (PolyQ) Diseases

- Huntington's disease
- Spinocerebellar ataxia (SCA) (most types)
- Spinobulbar muscular atrophy (Kennedy disease)
- Dentatorubral-pallidoluysian atrophy

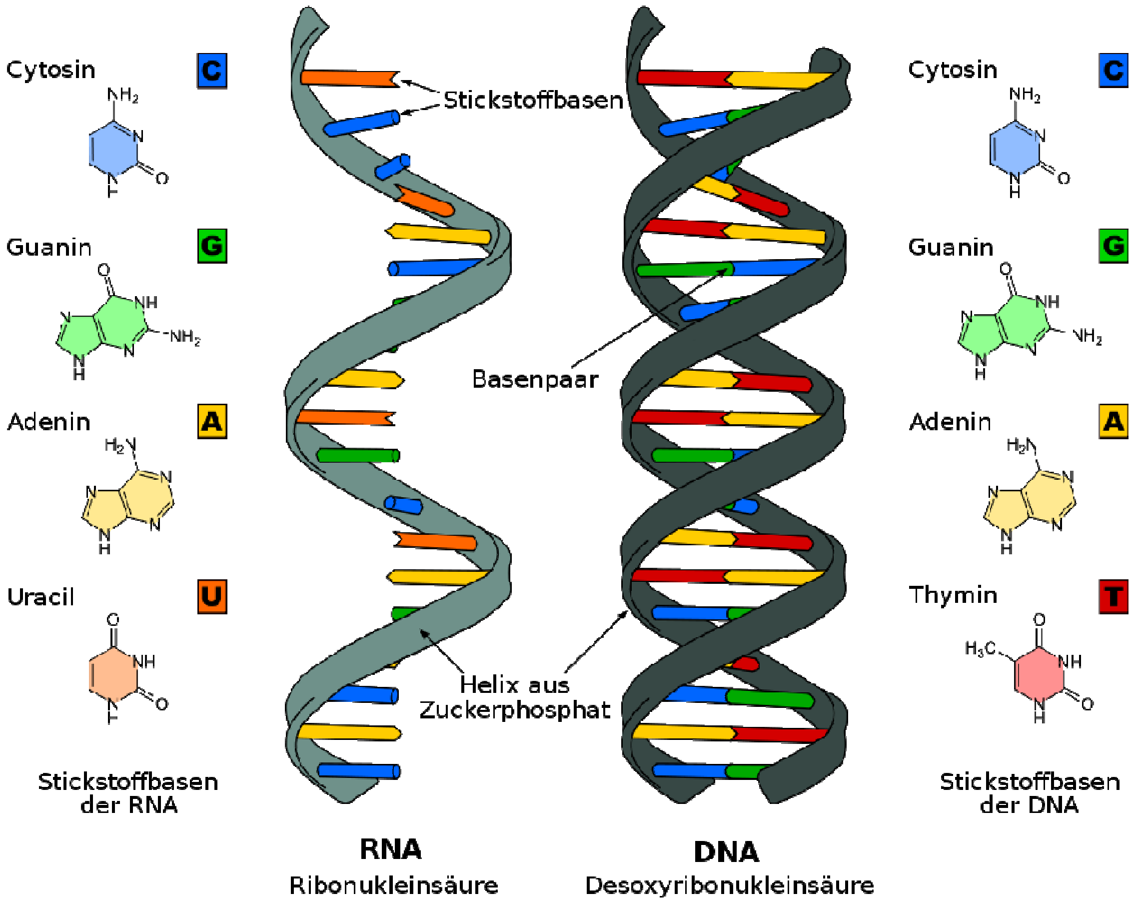
Mutation type

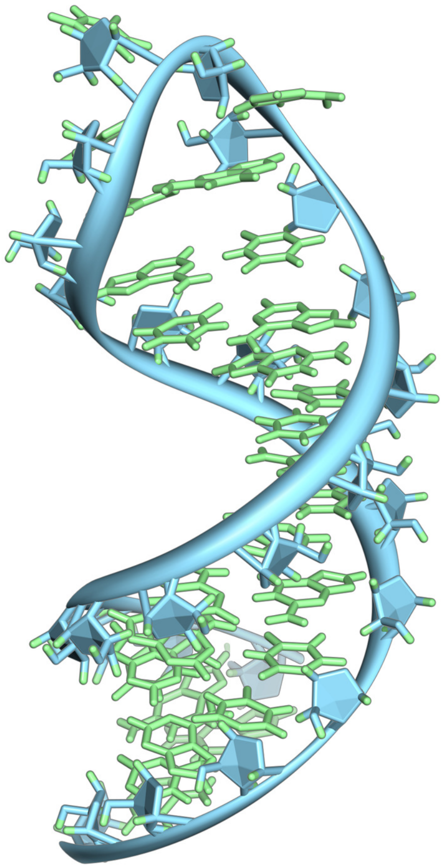
= Trinucleotide repeat
= Deletion
= Missense
= Nonsense

The origin of DNA

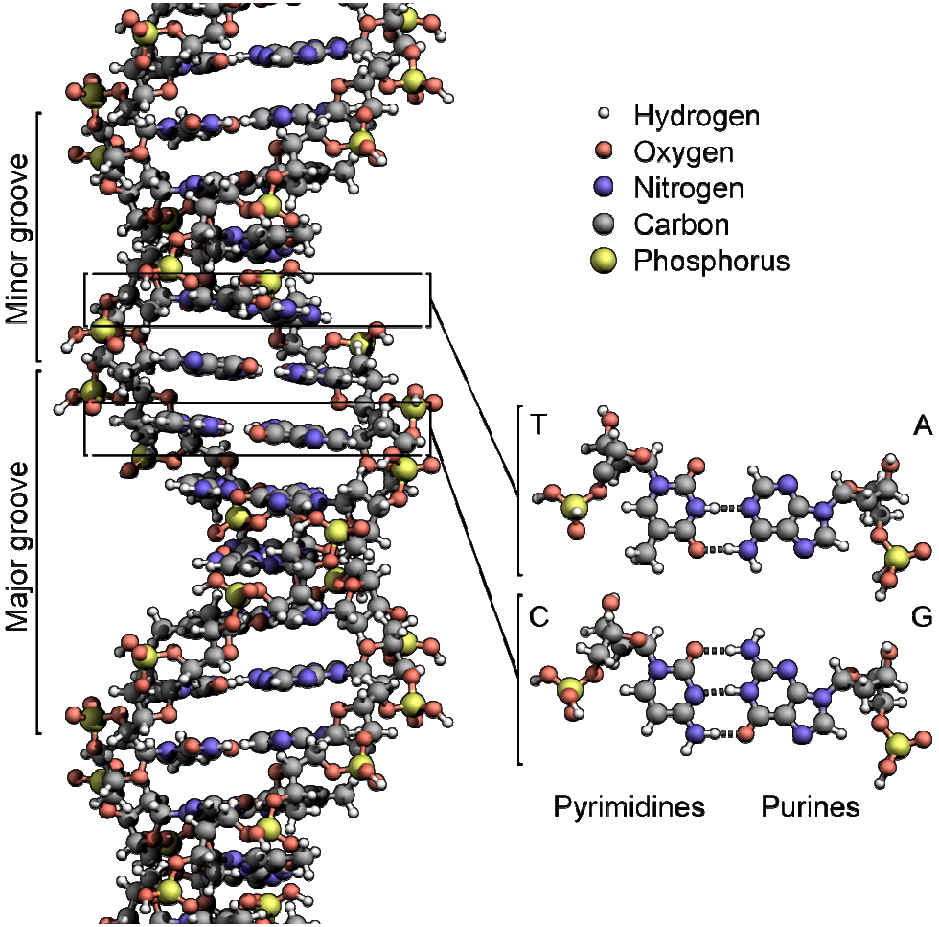


The origin of DNA





The origin of DNA



Maximal size of RNA-based genome: 3000-5000 bases (HIV, West Nile virus)

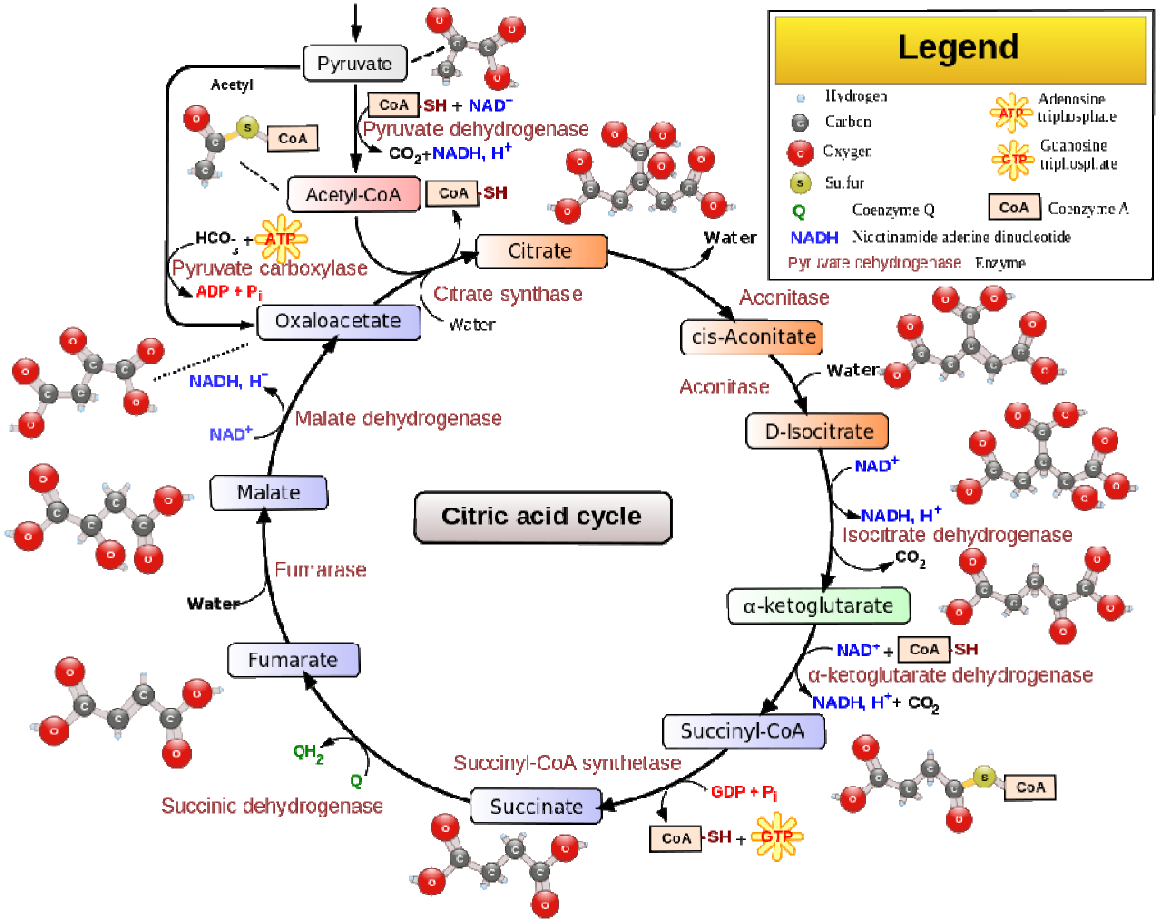
Reason: above that, statistically certain to generate at least one self-cleaving RNA sequence (ribozyme)

Maximal DNA size – unlimited

- no self-cleaving DNAzymes,
- tight storage as dsDNA,
- methylated uracil (thymine) → no accidental C-to-U mutations

Metabolic pathways

Citric acid cycle



Acetyl-CoA + 3 NAD$^+$ + GDP → CoA + 3 NADH + 2 GTP + 2 CO$_2$ + 3H$^+$

Oxidative, not present in Archaea, most likely absent in LUCA

Light harvesting - photosynthesis

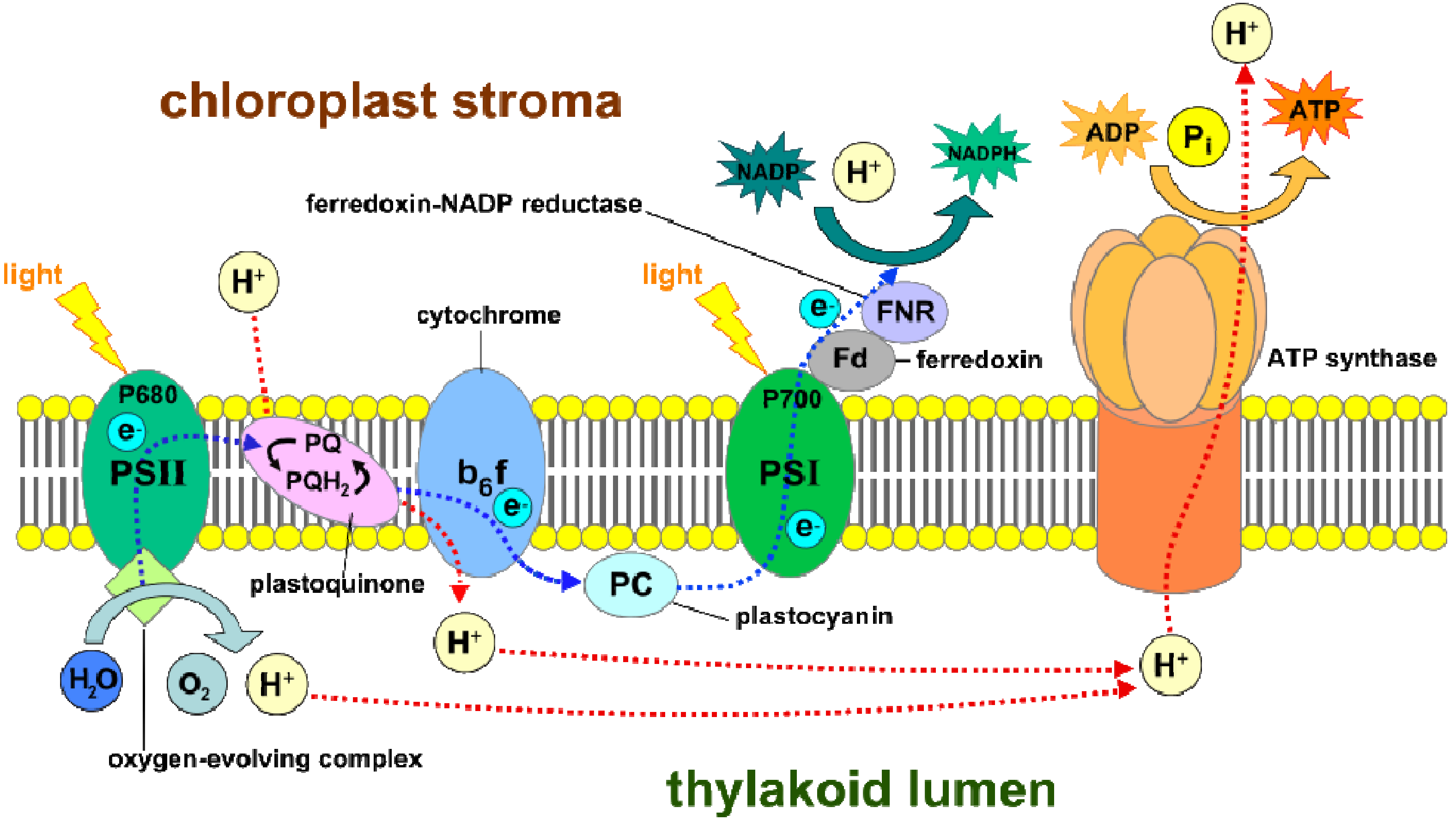
$$6\ CO_2 + 6\ H_2O \rightarrow \text{glucose}\ C_6H_{12}O_6 + 6\ O_2$$

PS I from green sulfur bacteria *Chlorobiaceae*

Light harvesting - photosynthesis

Light-dependent reactions of photosynthesis at the thylakoid membrane

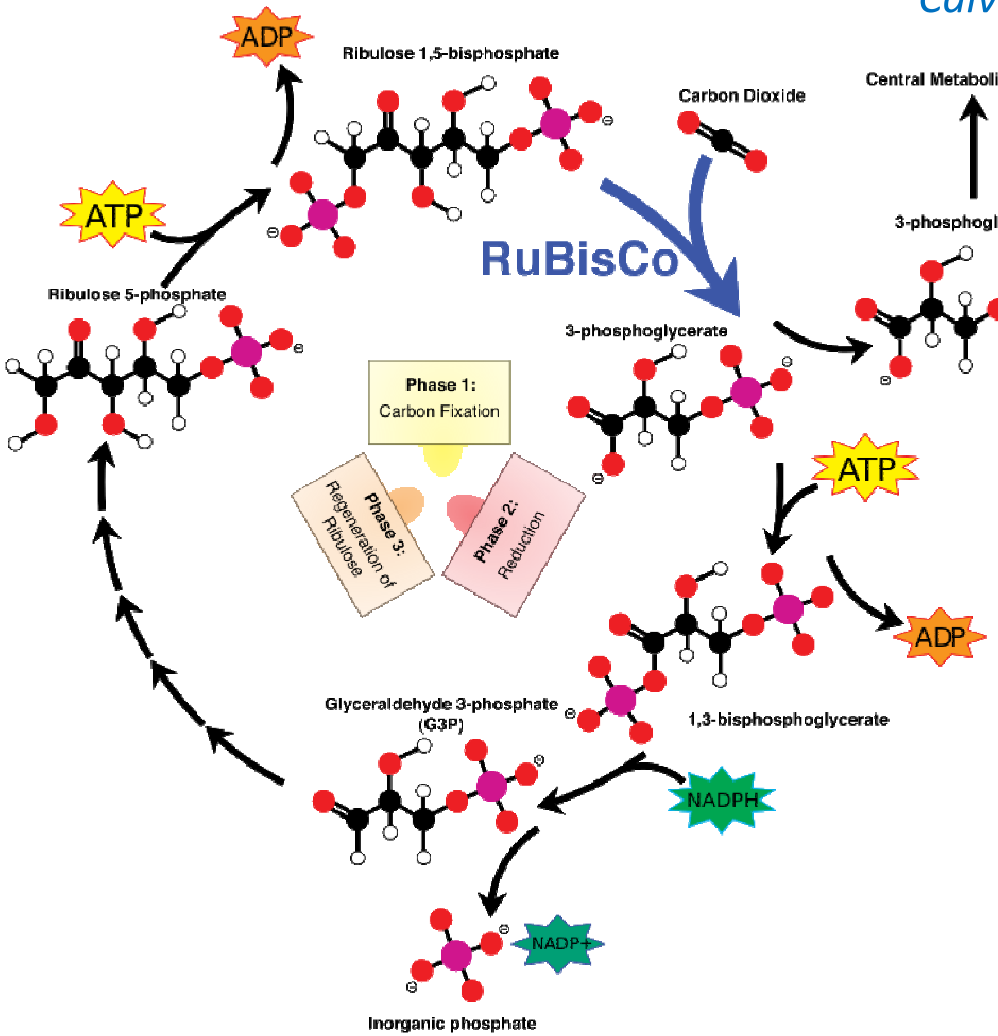
$2H_2O + 2NADP^+ + 3ADP + 3P_i \rightarrow O_2 + 2NADPH + 3ATP$

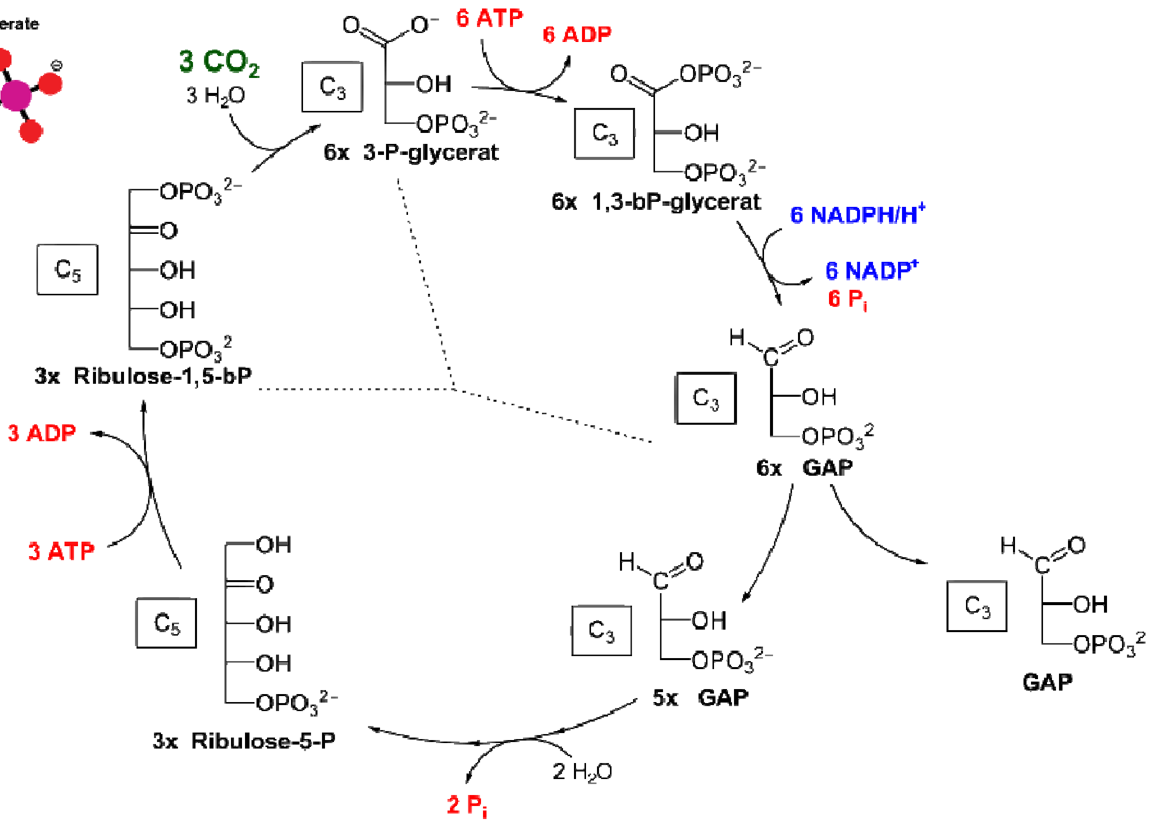


Water difficult to oxidize. Only combination of two photosystems provides enough electrochemical potential.

Metabolic pathways

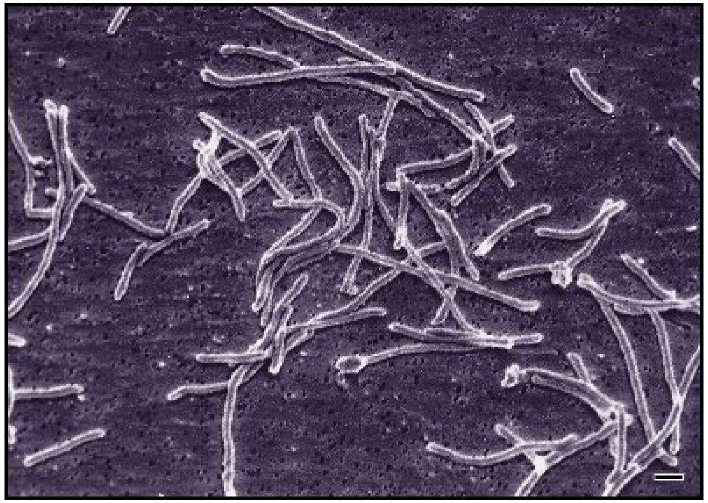
Calvin cycle



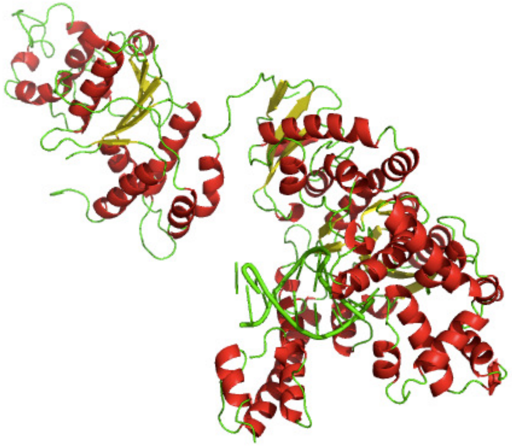


$3\ CO_2 + 6\ NADPH + 5\ H_2O + 9\ ATP \rightarrow$ glyceraldehyde-3-phosphate (G3P) $+ 2\ H^+ + 6\ NADP^+ + 9\ ADP + 8\ Pi$

Thermophiles



Thermus aquaticus

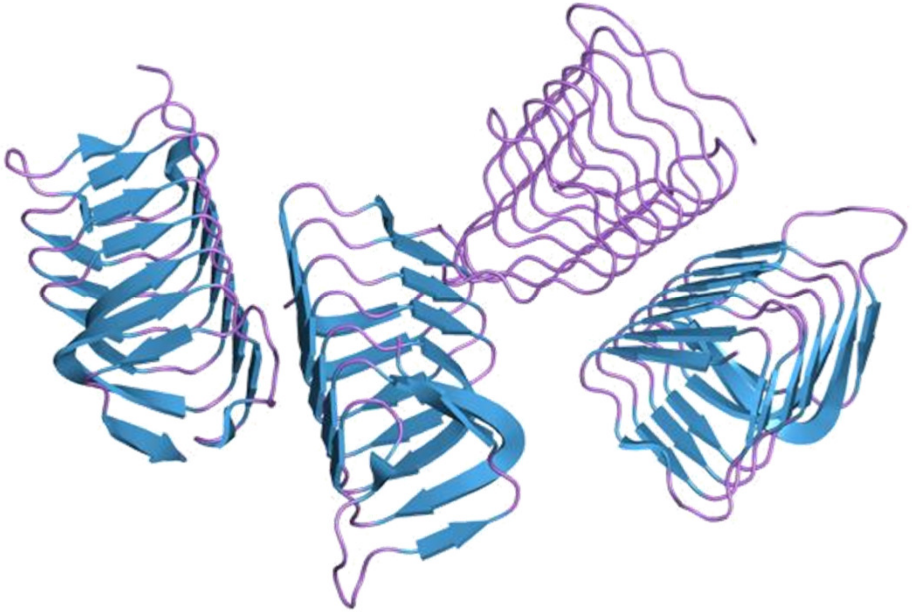


3D structure of Taq Polymerase.

Hot springs with algae and bacteria in Yellowstone National Park

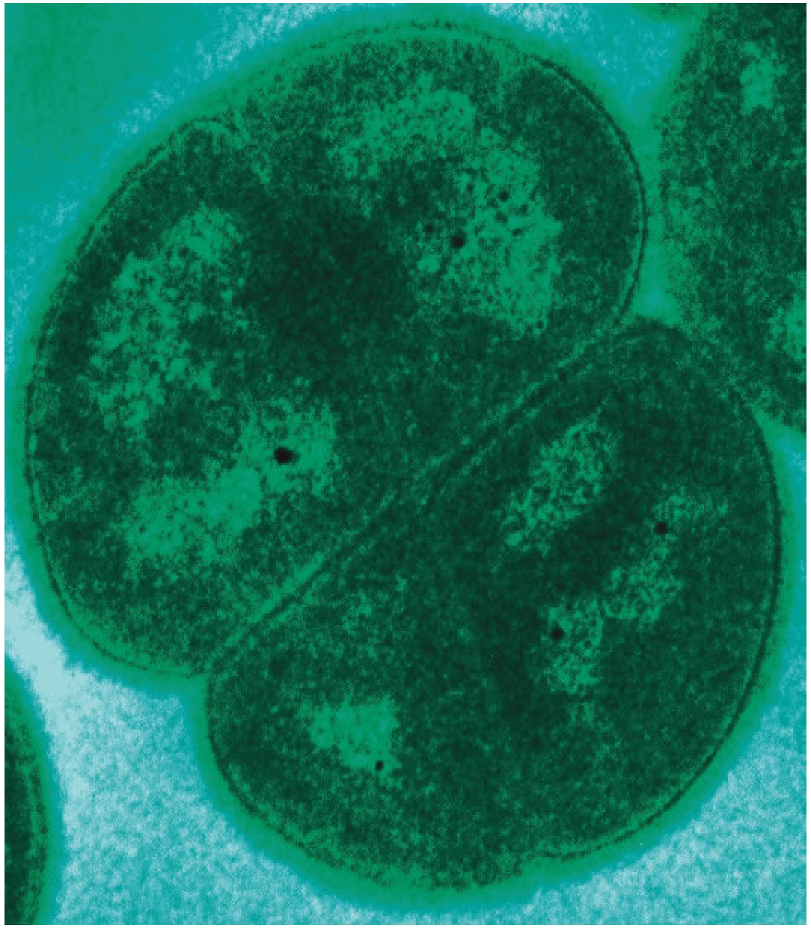
Cold adaptation

Structure of the *Tenebrio molitor* beta-helical antifreeze protein



Structure of *Choristoneura fumiferana* (spruce budworm) beta-helical antifreeze protein

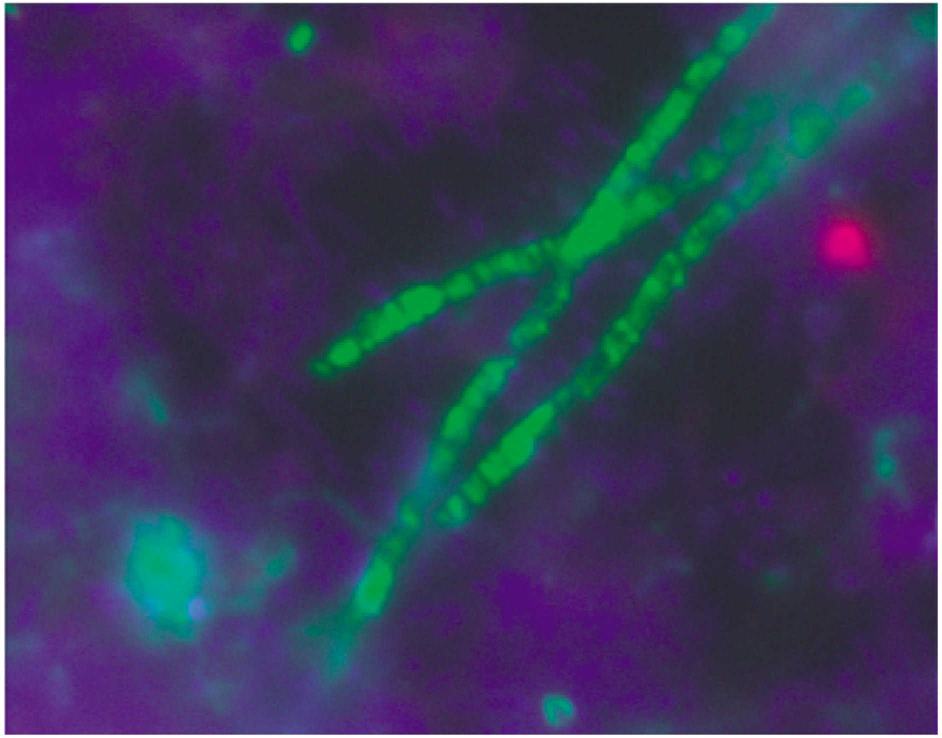
Drought, salinity, radiation



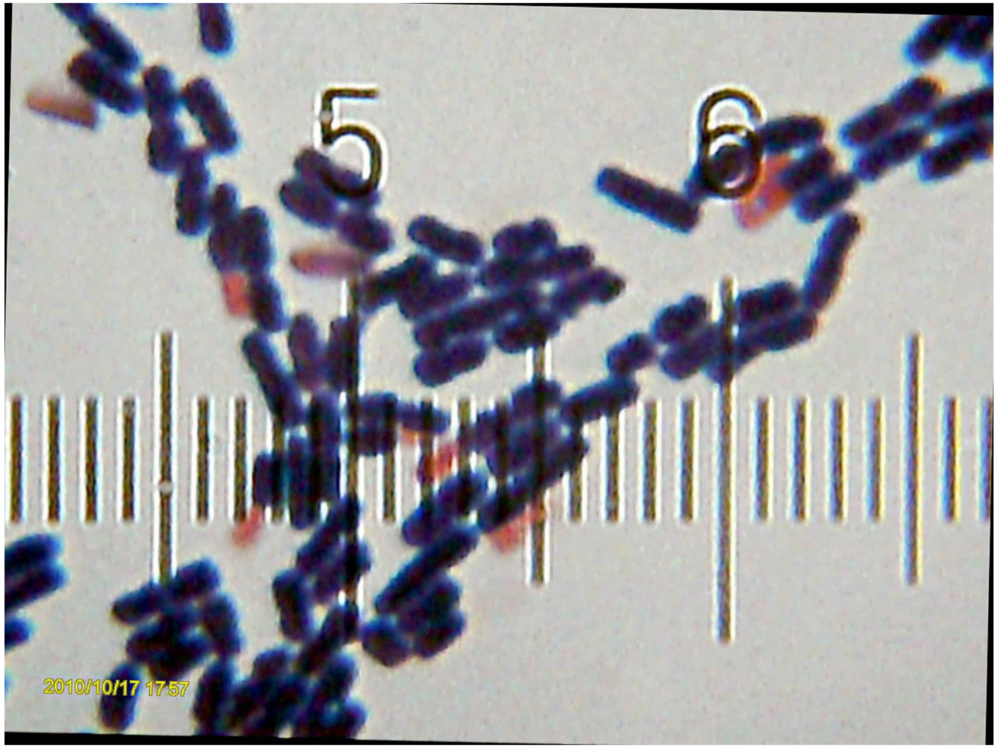
A tetrad of D. *radiodurans*

Efficient DNA damage repair,
Trehalose as the main sugar – glass solid, no crystallization

Acid, base



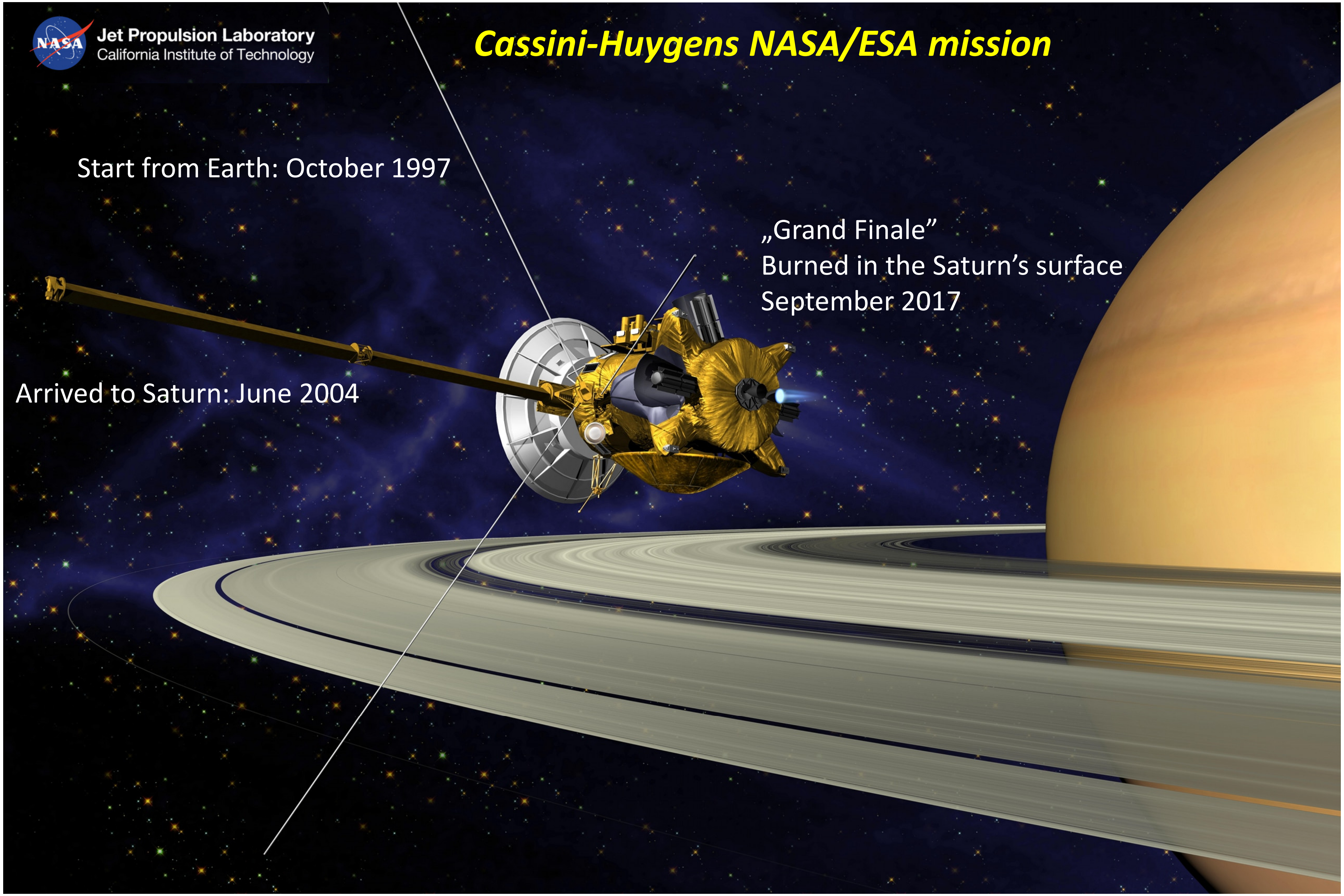
Acidobacterium



A typical *bacillus* culture. Many alkaliphiles possess a *bacillus* morphology

Jet Propulsion Laboratory
California Institute of Technology

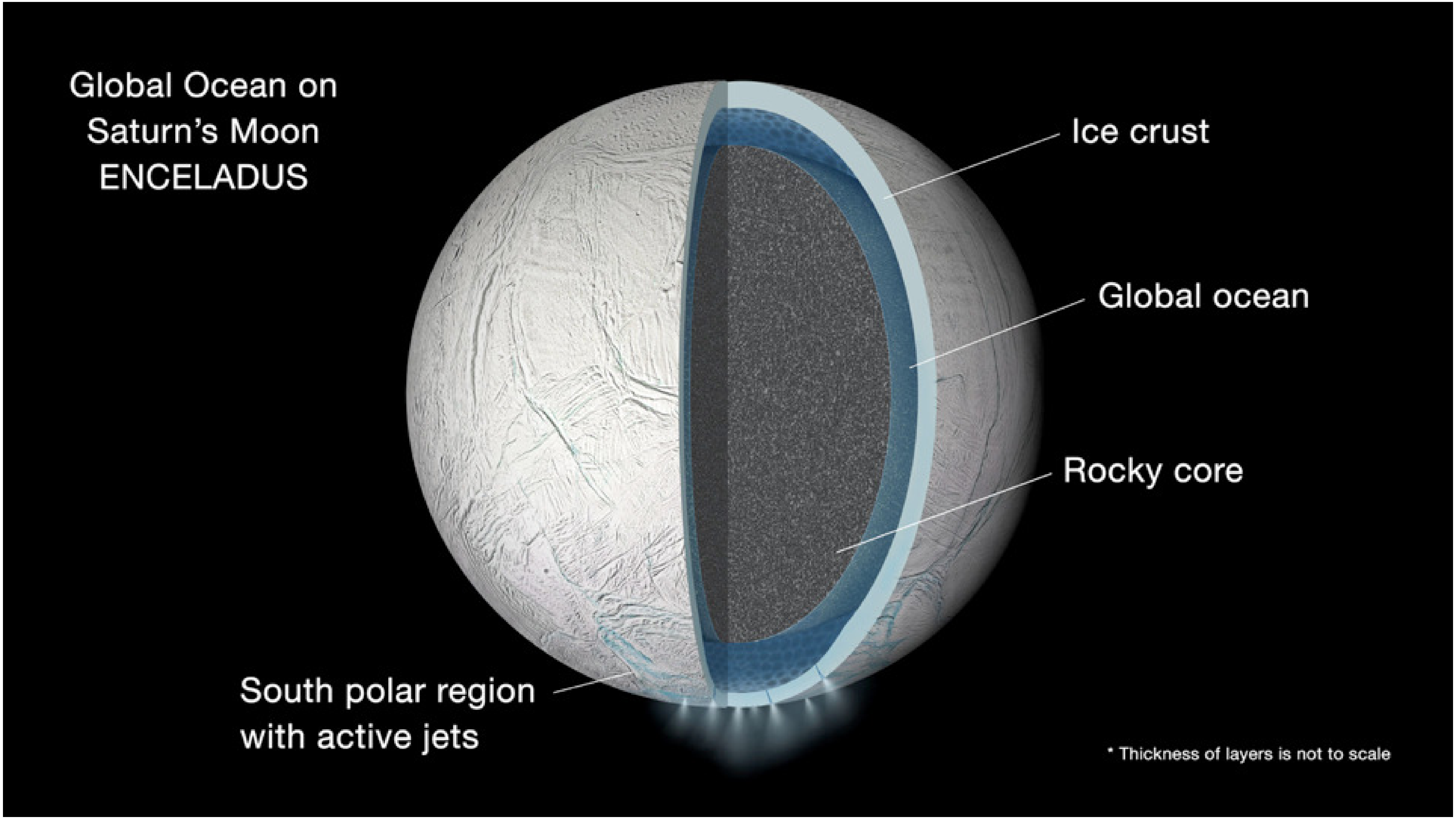
Cassini-Huygens NASA/ESA mission

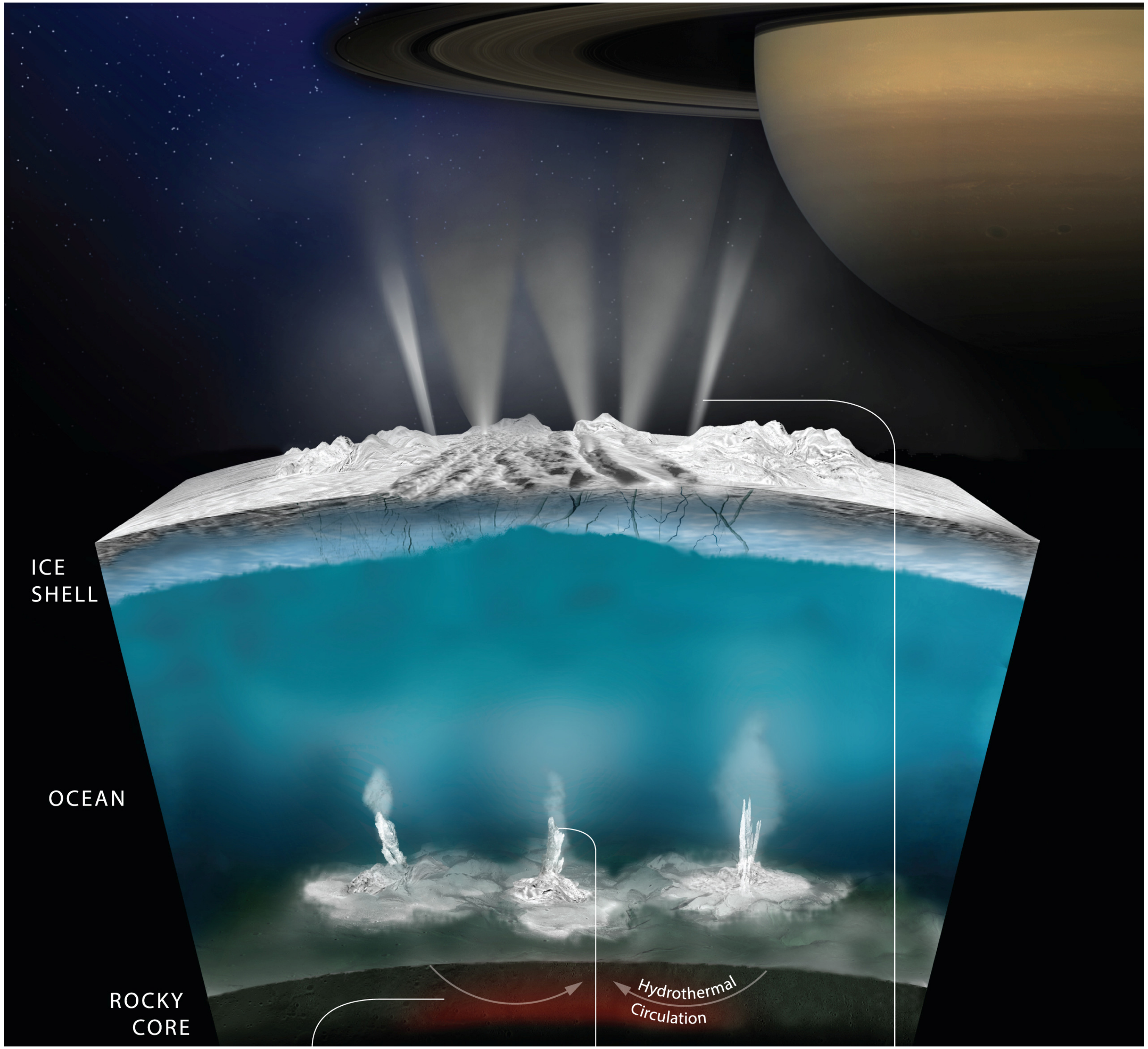


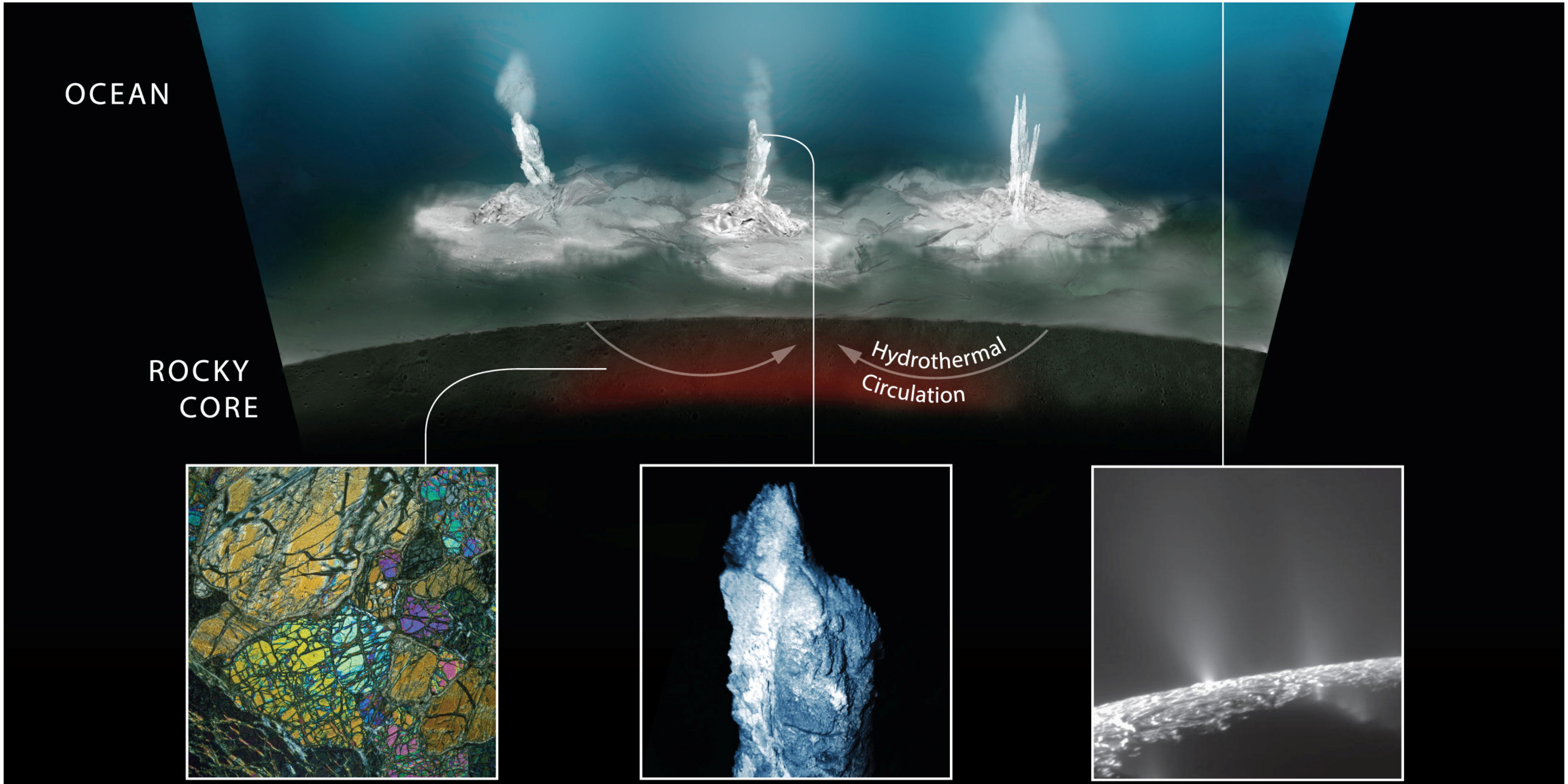
Start from Earth: October 1997

„Grand Finale”
Burned in the Saturn’s surface
September 2017

Arrived to Saturn: June 2004


Global Ocean on Saturn's Moon ENCELADUS
Ice crust
Global ocean
Rocky core
South polar region with active jets
* Thickness of layers is not to scale


ICE
SHELL
OCEAN
ROCKY
CORE
Hydrothermal
Circulation



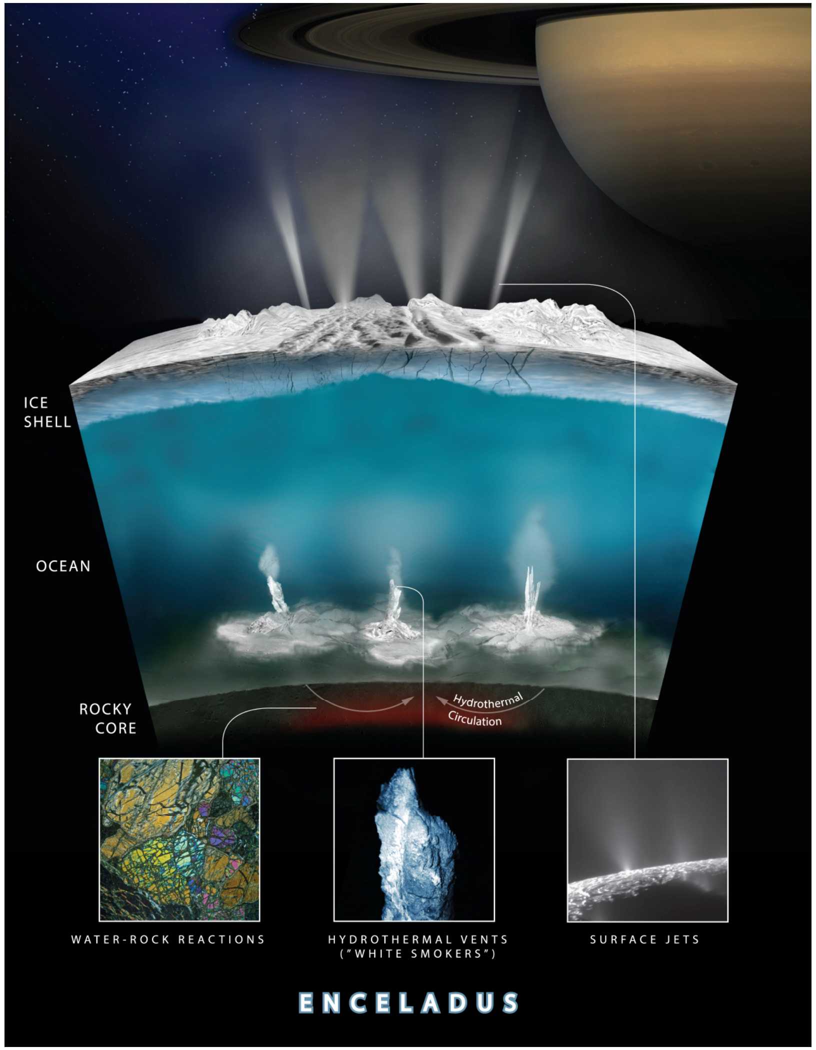
WATER-ROCK REACTIONS

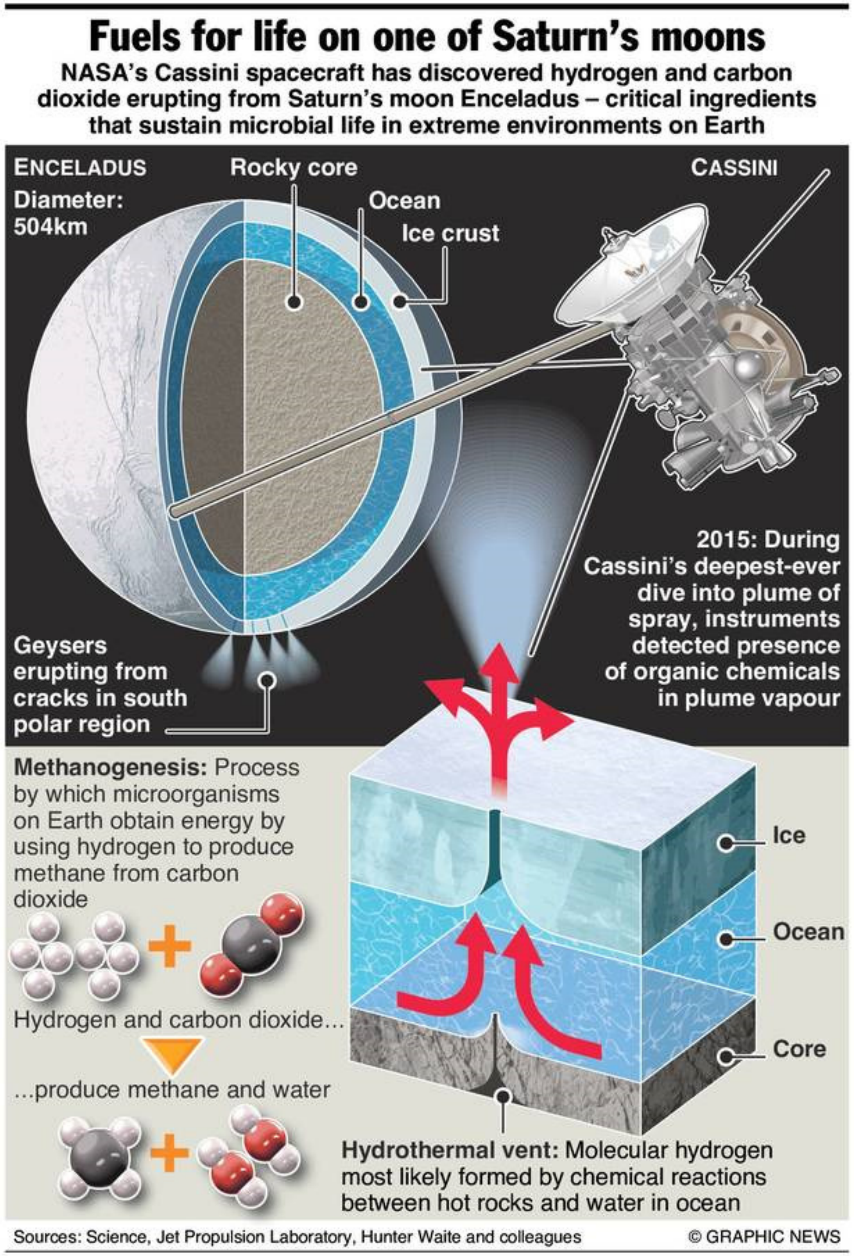
HYDROTHERMAL VENTS
("WHITE SMOKERS")

SURFACE JETS

ENCELADUS

Can (and does?) Enceladus host (microbial) life?





Fuels for life on one of Saturn's moons

NASA's Cassini spacecraft has discovered hydrogen and carbon dioxide erupting from Saturn's moon Enceladus – critical ingredients that sustain microbial life in extreme environments on Earth

2015: During Cassini's deepest-ever dive into plume of spray, instruments detected presence of organic chemicals in plume vapour

Geysers erupting from cracks in south polar region

Methanogenesis: Process by which microorganisms on Earth obtain energy by using hydrogen to produce methane from carbon dioxide

Hydrogen and carbon dioxide…

…produce methane and water

Hydrothermal vent: Molecular hydrogen most likely formed by chemical reactions between hot rocks and water in ocean

Sources: Science, Jet Propulsion Laboratory, Hunter Waite and colleagues

© GRAPHIC NEWS