The molecular origins of life

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

**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
© 2013 W.H. Freeman and Company
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.
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

- Oxyanions, (HPO₄²⁻, HVO₄²⁻, CrO₄²⁻, HAsO₄²⁻), REE, Trace Metals
- ³He, Mn²⁺, H₂SO₄, FeOOH, MnO₂, ΔT, CH₄, Fe²⁺, FeₓSᵧ, ²²²Rn, H₂, H₂S
- 2.05°C
- 2°C
- 0.1 cm/s

Sub Seafloor Microbial Biosphere

- WARM (diffuse) flow
- HOT (focussed) flow
- 350°C

Seawater

Seawater

H⁺, Cl⁻, Fe²⁺, Mn²⁺,
H₄SiO₄, ³He, H₂S, CH₄, CO₂, H₂,
Ca²⁺, K⁺, Li⁺, Cu²⁺, Zn²⁺, Pb²⁺

Evolved Seawater

Precipitation Chimney (Black Smoker)

Magma

1200°C

Spreading Axis

HT Reaction Zone

400°C

Magnetiferous Sediments

Iron-Magnesium Crusts

Basalt
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
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.
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 barrier. 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) \( \text{H}^+ \) gradient drives energy metabolism (ATPase) or carbon metabolism (Ech)

B) SPAP generates \( \text{Na}^+ \) from \( \text{H}^+ \) gradient

C) Membrane pumps secret \( \text{H}^+ \) and \( \text{Na}^+ \)

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

E) Impermeable membranes \( \rightarrow \) 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_
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
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), chemoorganooautotrophy enzymes absent

*M.C. Weiss et al. Nature Microbiology, 2016, Article 16116*
A primitive metabolic pathway for carbon fixation, still used by some contemporary chemoautotrophic organisms

M.C. Weiss et al. *Nature Microbiology*, 2016, Article 16116
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\(\rightarrow\) electrons likely obtained from hydrogen, as in modern microbes using the WL pathway

Nitrogenase and glutamine synthetase serve for nitrogen fixation

WL pathway, nitrogenase and hydrogenases are very oxygen-sensitive\(\rightarrow\) 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

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**

FeMo cofactor of nitrogenase

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 and wetlands), and they depend on H$_2$ (from geology or H$_2$-producing fermentation)

Geological source of hydrogen: serpentinization (iron + hot water, anoxic)  \[
\text{Fe}^{2+} + \text{H}_2\text{O} \rightarrow \text{Fe}_3\text{O}_4 + \text{H}_2
\]
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

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.

M.C. Weiss et al. *Nature Microbiology*, 2016, Article 16116
Transition from the RNA world to LUCA

Ribozymes – self-acting $\rightarrow$ 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 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 $\rightarrow$ non-covalent binding of nucleoside cofactors to contemporary enzymes.
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)

Ribosome

Every protein component of the ribosome can be removed without losing 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.
The genetic code

<table>
<thead>
<tr>
<th>First letter</th>
<th>Second letter</th>
<th>Third letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU, UUC, UUA, UUG</td>
<td>Leu, Phe, Ser, Stop</td>
</tr>
<tr>
<td>UC</td>
<td>UCU, UCC, UCA, UCG</td>
<td>Stop, Tyr, UAA, UAG</td>
</tr>
<tr>
<td>AC</td>
<td>CAU, CAC, CAA, CAG</td>
<td>Pro, His, Gln, Stop</td>
</tr>
<tr>
<td>AG</td>
<td>CGU, CGC, CGA, CGG</td>
<td>Arg, UCA, UCG, UAG</td>
</tr>
<tr>
<td>U</td>
<td>UUU, UUC, UUA, UUG</td>
<td>Leu, Phe, Ser, Stop</td>
</tr>
<tr>
<td>C</td>
<td>CCU, CUC, CUA, CUG</td>
<td>Leu, Pro, Ser, Stop</td>
</tr>
<tr>
<td>A</td>
<td>AUC, AUC, AUA, AUG</td>
<td>Ile, Thr, Met, Stop</td>
</tr>
<tr>
<td>G</td>
<td>GCU, GCC, GCA, GCG</td>
<td>Val, Ala, Asp, Glu</td>
</tr>
</tbody>
</table>

Codons 1-7:
- Codon 1: GCG, GCA, GCU, GGC
- Codon 2: UCG, UCA, UCU, UCC
- Codon 3: CGU, CGC, CGA, CGG
- Codon 4: GGU, GGA, GGG
- Codon 5: AUC, ACC, ACA, ACU
- Codon 6: UCU, UCC, UCA, UCG
- Codon 7: GUG, GCU, GCA, GCG

Ribonucleic acid

Solubility

Size
# The genetic code

## Examples of notable Mutations

<table>
<thead>
<tr>
<th>1st base</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUC (Val) Valine</td>
<td>UCU (Ser) Serine</td>
<td>UAA (Ter) Stop</td>
<td>UGA (Ter) Stop</td>
</tr>
<tr>
<td>UUC (Val) Valine</td>
<td>UCU (Ser) Serine</td>
<td>UAG (Ter) Stop</td>
<td>UGA (Ter) Stop</td>
<td></td>
</tr>
<tr>
<td>UUA (Leu) Leucine</td>
<td>UCA (Ser) Serine</td>
<td>UAA (Ter) Stop</td>
<td>UGA (Ter) Stop</td>
<td></td>
</tr>
<tr>
<td>UUG (Leu) Leucine</td>
<td>UCG (Ser) Serine</td>
<td>UAG (Ter) Stop</td>
<td>UGA (Ter) Stop</td>
<td></td>
</tr>
<tr>
<td>CUU (Leu) Leucine</td>
<td>UCU (Ser) Serine</td>
<td>CAA (Gln) Glutamine</td>
<td>CGA (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>CUC (Leu) Leucine</td>
<td>UCC (Ser) Serine</td>
<td>CAC (Glu) Glutamic acid</td>
<td>CGC (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>CUA (Leu) Leucine</td>
<td>UCA (Ser) Serine</td>
<td>CAA (Gln) Glutamine</td>
<td>CGC (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>CUG (Leu) Leucine</td>
<td>UCG (Ser) Serine</td>
<td>CAC (Glu) Glutamic acid</td>
<td>CGC (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>AUA (Ile) Isoleucine</td>
<td>ACA (Thr) Threonine</td>
<td>AAA (Lys) Lysine</td>
<td>AGA (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>AUC (Ile) Isoleucine</td>
<td>ACC (Thr) Threonine</td>
<td>AAA (Lys) Lysine</td>
<td>AGA (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>AUA (Ile) Isoleucine</td>
<td>ACA (Thr) Threonine</td>
<td>AAG (Lys) Lysine</td>
<td>AGG (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>AUC (Ile) Isoleucine</td>
<td>ACC (Thr) Threonine</td>
<td>AAG (Lys) Lysine</td>
<td>AGG (Arg) Arginine</td>
<td></td>
</tr>
<tr>
<td>GUA (Val) Valine</td>
<td>GCU (Ser) Serine</td>
<td>GAA (Glu) Glutamic acid</td>
<td>GAG (Glu) Glutamic acid</td>
<td></td>
</tr>
<tr>
<td>GUG (Val) Valine</td>
<td>GCU (Ser) Serine</td>
<td>GAG (Glu) Glutamic acid</td>
<td>GAG (Glu) Glutamic acid</td>
<td></td>
</tr>
</tbody>
</table>

### Polyglutamine (PolyQ) Diseases
- Huntington's disease
- Spinocerebellar ataxia (SCA)
- Spinobulbar muscular atrophy (Kennedy disease)

### Fragile X Syndrome

### Friedreich's ataxia

### Myotonic dystrophy

### Colorectal cancer

### Breast cancer

### 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) \(\rightarrow\) no accidental C-to-U mutations
Metabolic pathways

Citric acid cycle

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

Oxidative, not present in Archaea, most likely absent in LUCA
Light harvesting - photosynthesis

\[ 6 \text{CO}_2 + 6 \text{H}_2\text{O} \rightarrow \text{glucose C}_6\text{H}_{12}\text{O}_6 + 6 \text{O}_2 \]

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.
3 CO₂ + 6 NADPH + 5 H₂O + 9 ATP → glyceraldehyde-3-phosphate (G3P) + 2 H⁺ + 6 NADP⁺ + 9 ADP + 8 Pi
**Thermophiles**

*Thermus aquaticus*

Hot springs with algae and bacteria in Yellowstone National Park

3D structure of Taq Polymerase.
Cold adaptation

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

Structure of *Choristoneura fumiferana* (spruce budworm) beta-helical antifreeze protein
Drought, salinity, radiation

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

A tetrad of D. radiodurans
Acid, base

Acidobacterium

A typical *bacillus* culture. Many alkaliphiles possess a *bacillus* morphology
Start from Earth: October 1997

Cassini-Huygens NASA/ESA mission

Arrived to Saturn: June 2004

„Grand Finale”
Burned in the Saturn’s surface
September 2017
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
Can (and does?) Enceladus host (microbial) life?