The Big Questions in Origin-of-Life Research

1. How did the "primordial soup" acquire the simple monomeric building blocks essential for biochemistry and the production of information-containing polymers?

    - Was there ever such an open-water “soup”?
    - If not, where did life originate?

2. Are the many odd features of today’s cellular biochemistry frozen, historical reflections of this legacy?

3. Which came first – DNA, RNA, protein, or something else? Can a single polymer be capable of self-replication and information storage?

4. Compartmentalization (individualization) is necessary for self-recognition during replication and for preventing the diffusion of gene products. How did membrane lipids evolve?

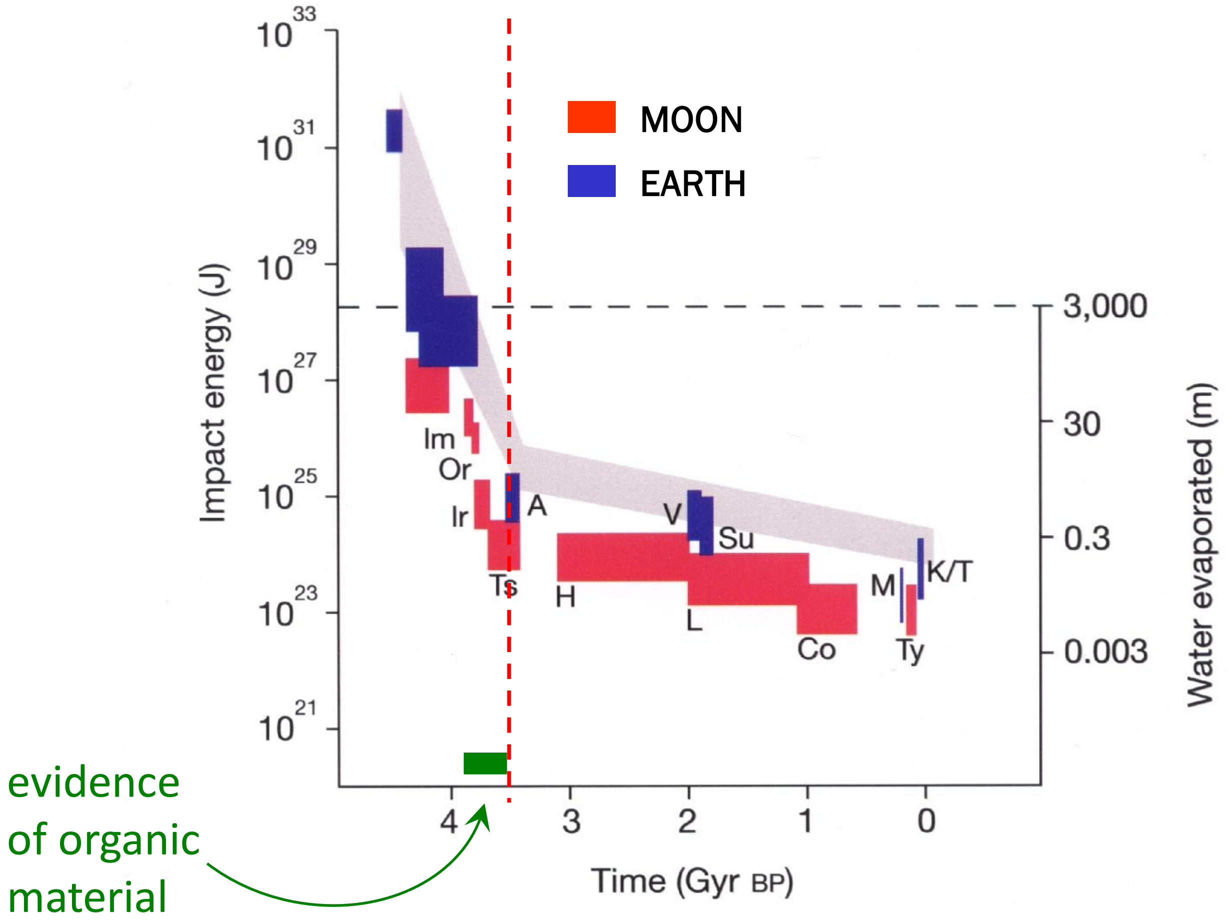
5. Can a simple self-replicating, membrane-bound cell be developed *de novo* in the laboratory?

The Basic Properties of Life Were Present on Earth ~3.5 to 4.0 Billion Years Ago

Vaporization of the Earth's Oceans Until ~3.8 BYA

Figure 11.14 **The history of large impacts on the Earth and Moon** The horizontal axis represents time (billions of years before present). The left vertical axis represents the energy of impact (joules). The right vertical axis shows the depth (m) to which the oceans would be vaporized by an impact with a given energy. The dashed line represents vaporization of the entire global ocean. Each box encloses the range of times during which a particular impact is estimated to have occurred, and the range of energies the impact is estimated to have fallen within. The open boxes are for the Moon; the filled boxes are for the Earth. Boxes with labels represent impacts documented by craters or other geological evidence. The unlabeled box at the upper left represents a large impact thought to be responsible for the formation of the Moon. The other unlabeled boxes are for hypothetical impacts. The gray band shows the largest impacts likely to have hit the Earth at any given time. Lunar craters: Im = Imbrium; Or = Orientale; Ir = Iridum; Ts = Tsiolkovski; H = Hausen; L = Langrenus; Co = Copernicus; Ty = Tycho. Terrestrial craters: A = Archaean spherule beds; V = Vredevort; Su = Sudbury; M = Manicougan; K/T = Cretaceous/Tertiary impact (crater located off the Yucatan peninsula). From Sleep et al. (1989).

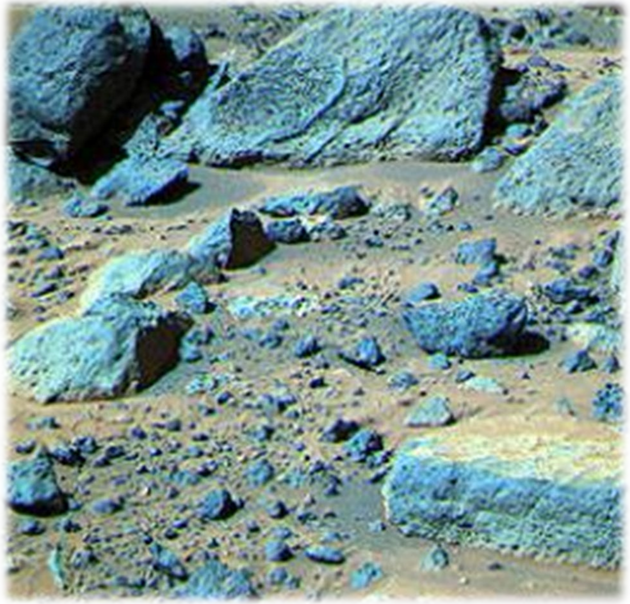
Fig. 14.10 The history of large impacts on the Earth and Moon



Evolutionary Analysis, 2nd Edition
Scott Freeman and Jon C. Herron

Evidence of Life on Mars?

Photos from Beagle 2 →



Are these biological microfossils? →

From: MacKay et al., 1996, Science.

Identifying potential scenarios under which life might have originated requires insight into energetic, geochemical, and physical opportunities in alternative settings:

- Accounting for the three major requirements for life:

  1) metabolism for resource acquisition;

  2) a genome for heritable transmission of genetic information;

  3) external membranes or a structured environment necessary for individuality.

- Carbon in the Earth's early atmosphere was dominated by oxidized forms of carbon (CO and CO_2), so some source of sustained energy would have been required to produce reduced-carbon compounds – the origins of metabolism.

- Without oxygen, there would have been no ozone shield, and the damaging effects of UV radiation would have been dramatically elevated – limitations on genome size.

The Urey-Miller (1953) Experiment: Spontaneous Production of a "Primordial" Soup

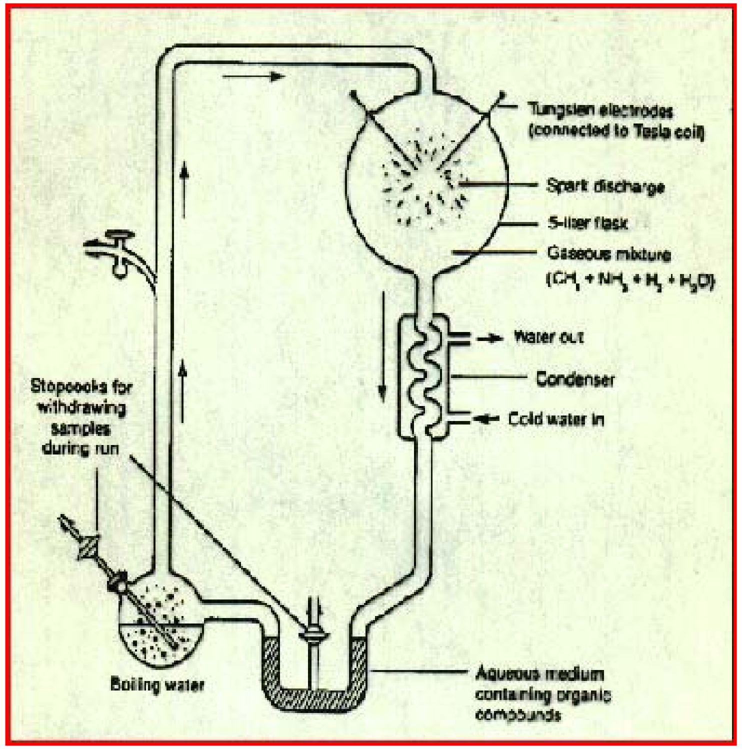


Table 2. Yields from Sparking a Mixture CH_4, NH_3, H_2O, and H_2.

Compound	Yield µmoles	Yield %
Glycine	630	2.1
Glycolic acid	560	1.9
Sarcosine	50	0.25
Alanine	340	1.7
Lactic acid	310	1.6
N-Methylalanine	10	0.07
α-Amino-*n*-butyric acid	50	0.34
α-Aminoisobutyric acid	1	0.007
α-Hydroxybutyric acid	50	0.34
β-Alanine	150	0.76
Succinic acid	40	0.27
Aspartic acid	4	0.024
Glutamic acid	6	0.051
Iminodiacetic acid	55	0.37
Iminoacetic-propionic acid	15	0.13
Formic acid	2330	4.0
Acetic acid	150	0.51
Propionic acid	130	0.66
Urea	20	0.034
N-Methyl urea	15	0.051
Total		15.2

The present yields are based on carbon; 59 mmoles (712 mg) of carbon was added as CH_4.

Many of the compounds necessary for life could be produced in a putatively "pre-biotic" atmosphere.

- H_2CO – formaldehyde
- HCN – hydrogen cyanide
- Amino acids
- Urea

Doubts About a “Heterotrophy-First” Hypothesis for Life’s Origin

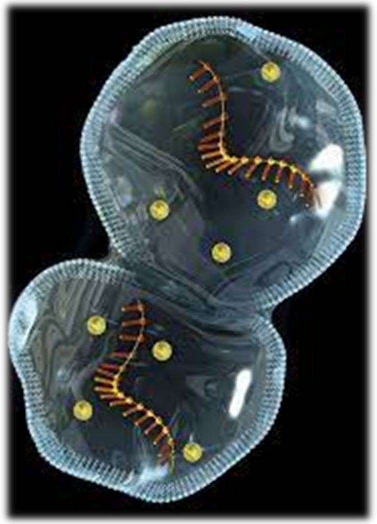
- The Earth’s early atmosphere was different than what was envisioned by Urey and Miller.

- Open water is counterproductive to the maintenance of organic aggregates essential to the nucleation of life.

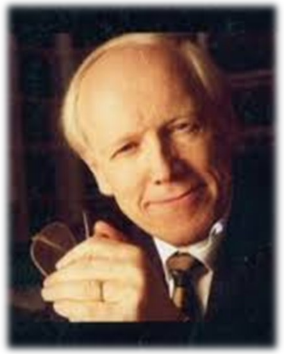
- How long could life have relied on the soup before encountering a resource-limitation crisis?

- Modern hypotheses attempt to identify early environmental settings that might have been conducive to the spontaneous emergence of life from inorganic materials.

- Of special interest – the peculiar sets of reactions, metal cofactors, and metabolic building blocks that life came to depend may be a reflection of the ancestral setting.



An "autotrophy-first" hypothesis: the chemoautotrophic origin of life at hydrothermal vents.



Günter Wächtershäuser
(chemist, turned lawyer)

- Vents = very high pH + H_2 (providing a natural source of electrons).

  Surrounding water = very high CO_2 (natural source of carbon, 1000x higher in the past, electron acceptor).

  Natural "chemiosmotic" state, resulting in an influx of H^+, providing a source of carbon and energy.

  Energy production in all of today's organisms relies on chemiosmosis, using a complex molecular machine known as ATP synthase.

- Metal catalysts, e.g., FeS, can be used to drive the reduction of CO_2 to formate, then acetate, then pyruvate.

  Pyruvate plays a pivotal role in energetics in all organisms.

  Fe and Ni sulfur clusters are the catalytic centers of many enzymes involved in catalysis.

- Life evolved in an anaerobic environment.

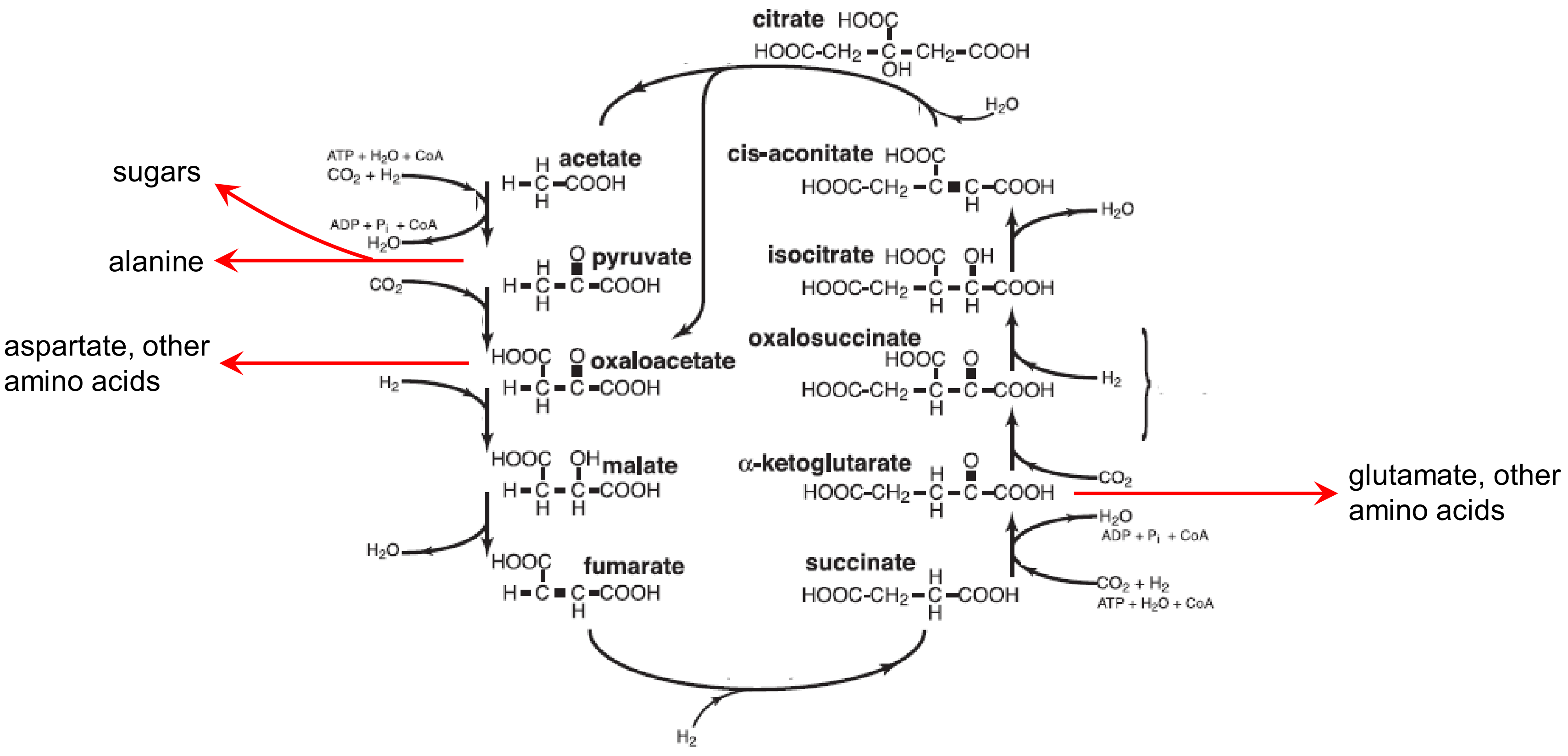
  Acetyl-CoA pathway, one of five known mechanisms for carbon fixation, uses oxygen-intolerant enzymes with FeS and FeNiS centers.

  The only CO_2 fixation pathway that yields energy while fixing carbon.

  Present in both prokaryotic groups, bacteria (acetogens) and archaea (methanogens), suggesting an ancient origin – parallel evolution?

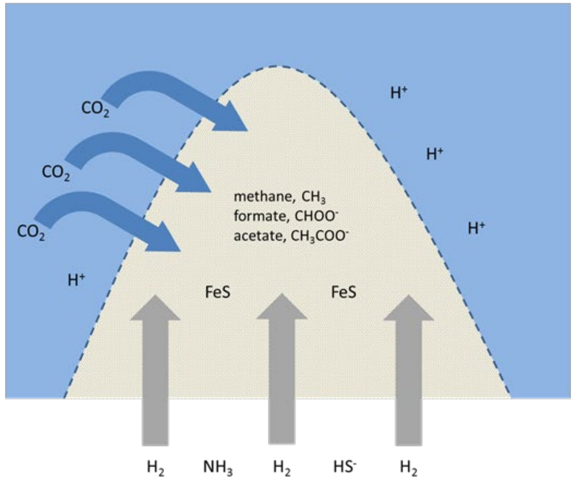
Reductive Citric-acid Cycle: in the absence of O_2, the citric-acid cycle may have run in reverse in early life.

- CO_2 is the input rather than the output.
- Many of the metabolites of the citric-acid cycle have been found in meteorites.



Alternative Geological Settings for an “Autotrophy-first” World

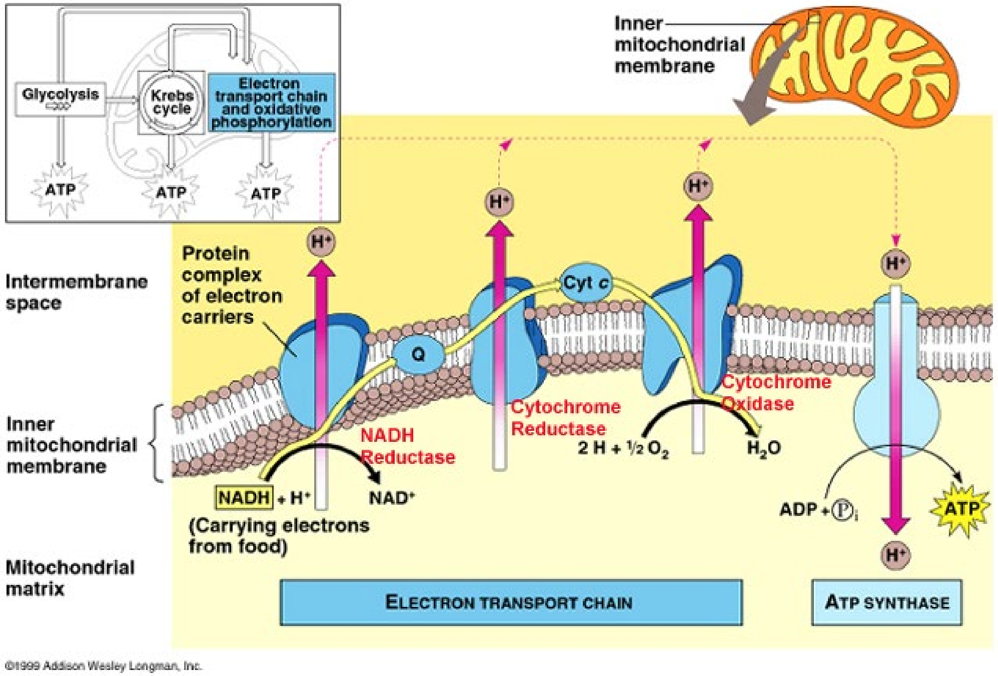
Alkaline Hydrothermal Vent

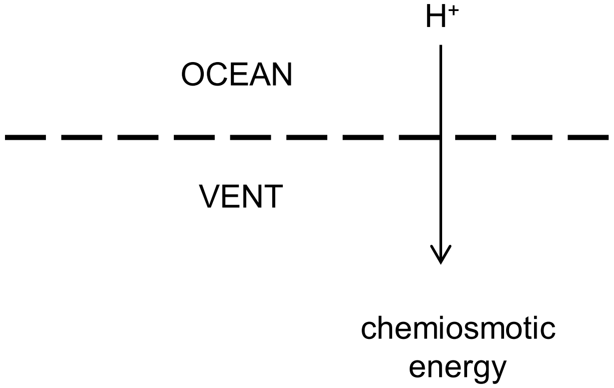


- Less extreme temperature than “hot smokers”.
- Support porous towers of calcium carbonate, which could serve as potential sites of compartmentalization.

Martin et al. 2008; Russell et al. 2010, 2014

The origin of metabolism: is today's universal mechanism of generating ATP by chemiosmosis a "molecular fossil,"
i. e., a simple consequence of "descent with modification" from the earliest stages of life?





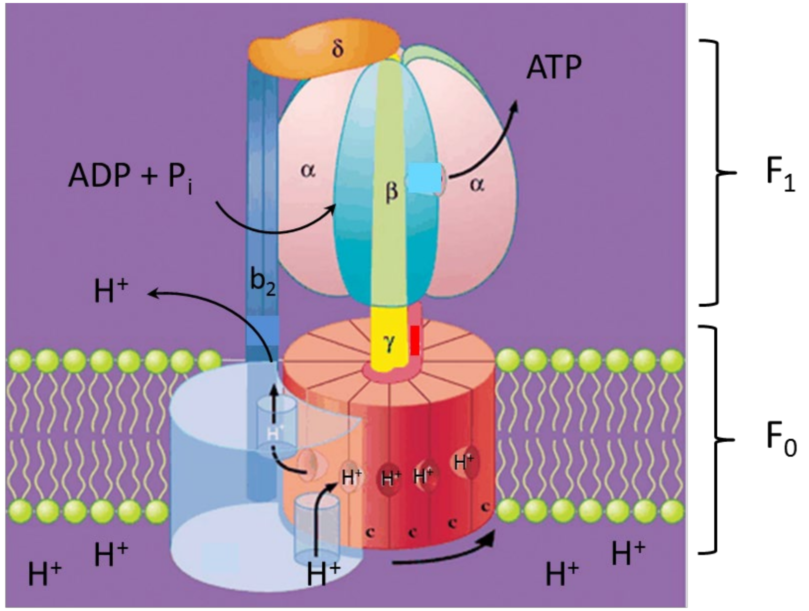
CO_2 → methane, formate, acetate

- The electron-transport chain releases H^+ to the outside of today's membranes, creating a proton gradient, resulting in **chemiosmosis** of protons back into the cell. This provides the energy for ATP production via ATP synthase, a molecular turbine.

- Under the vent hypothesis, the proton-motive force is maintained abiotically (for free).

ATP synthase: virtually all organisms use a hydrogen-ion gradient (the proton-motive force) to produce ATP.

- Hydrogen ions are pumped to the outside of cellular membranes only to reenter through an ATP synthase, where the energy from the resultant flow of protons is used to join ADP and Pi to produce ATP.



- Unlike a dam or a windmill, the initial source of energy comes from the cell itself via metabolic breakdown of food.

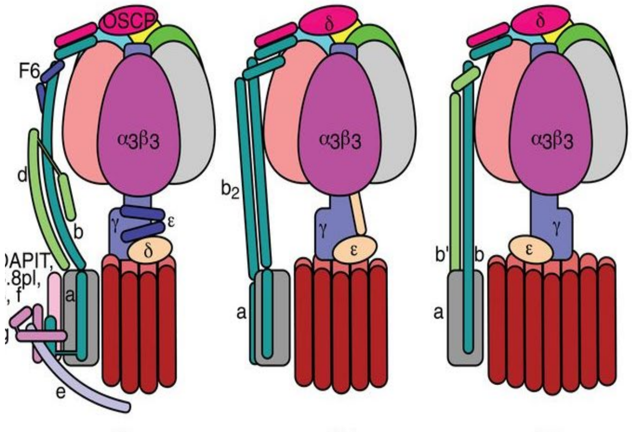
- Given its ubiquity, ATP synthase must have been present in LUCA (the last universal common ancestor).

- LUCA must have been membrane bound.

- Suggests life could not have inhabited open water until a membrane-bound, genome-encoded ATP synthase emerged.

Despite its essentiality for all of life, the structure of ATP synthase varies substantially among species.

- Bioenergetic efficiency is directly related to the structure of the membrane subunit:
  - each stalk rotation yields 3 ATP molecules;
  - number of protons required per rotation = the number of subunits in the rotating membrane ring (~10x faster than the wheels of a car at 60 mph).



Bovine mitochondrial F-ATP synthase | *E. coli* F-ATP synthase | Spinach chloroplast F-ATP synthase

- In the bacterium *E. coli*, there are ten subunits per ring:

  bioenergetic efficiency = 3 ATPs / 10 protons.

- Across the Tree of Life, the number of ring subunits varies from 8 to 15, implying a two-fold range of variation in efficiency.

- Numerous other structural variants are known for the stalk and stator subunits in ciliates, green algae, bacteria, etc.

- Major challenges in the origin of life would have involved the establishment of:

1) a catalytic capacity for biosynthesis;

2) a reliable means of genome transmission – membranes for individuality and accurate polymerases for replication fidelity.

The Origin of Information: Acquisition of an Evolvable, Information-bearing Polymer

DNA first?

Strong points:

- Use of multiple nucleotides provides the basis for a simple language.
- Double-strandedness provides an inherent template for replication – one strand specifies the other.

Weak point:

- DNA is almost completely lacking in catalytic activity.

Proteins first?

Strong points:

- Easy to synthesize amino acids and polymers under a variety of conditions.
- Even small peptides can exhibit catalytic activity.
- 20 amino acids provides for high information content.

Weak points:

- Globular structure and lack of complementarity preclude self-replication.
- Modern proteins cannot be produced without DNA, and vice versa.

The RNA-World Hypothesis

- Promotes the view that at an early stage of evolution, a single type of molecule (RNA) simultaneously provided the means for catalysis and information storage.

  - Genetic continuity was assured by the replication of an RNA-based genome, which specified a phenotype.

  - Proteins were not involved as catalysts, with all metabolic functions being carried out by RNA.

Molecular Fossils from the RNA World?

All of the major players in transcription and translation are RNAs:

- Messenger RNA carries genetic information to the ribosome.
- The catalytic heart of the ribosome consists of ribosomal RNAs.
- Transfer RNAs assist in the acquisition of amino acids for protein assembly.
- Small nuclear RNAs form the catalytic heart of the spliceosome.
- Micro RNAs are involved in gene regulation.
- Small nucleolar RNAs direct the modification of nucleotides in ribosomes.
- Primer RNAs direct the initiation of replication.

An RNA World?

Strong points:

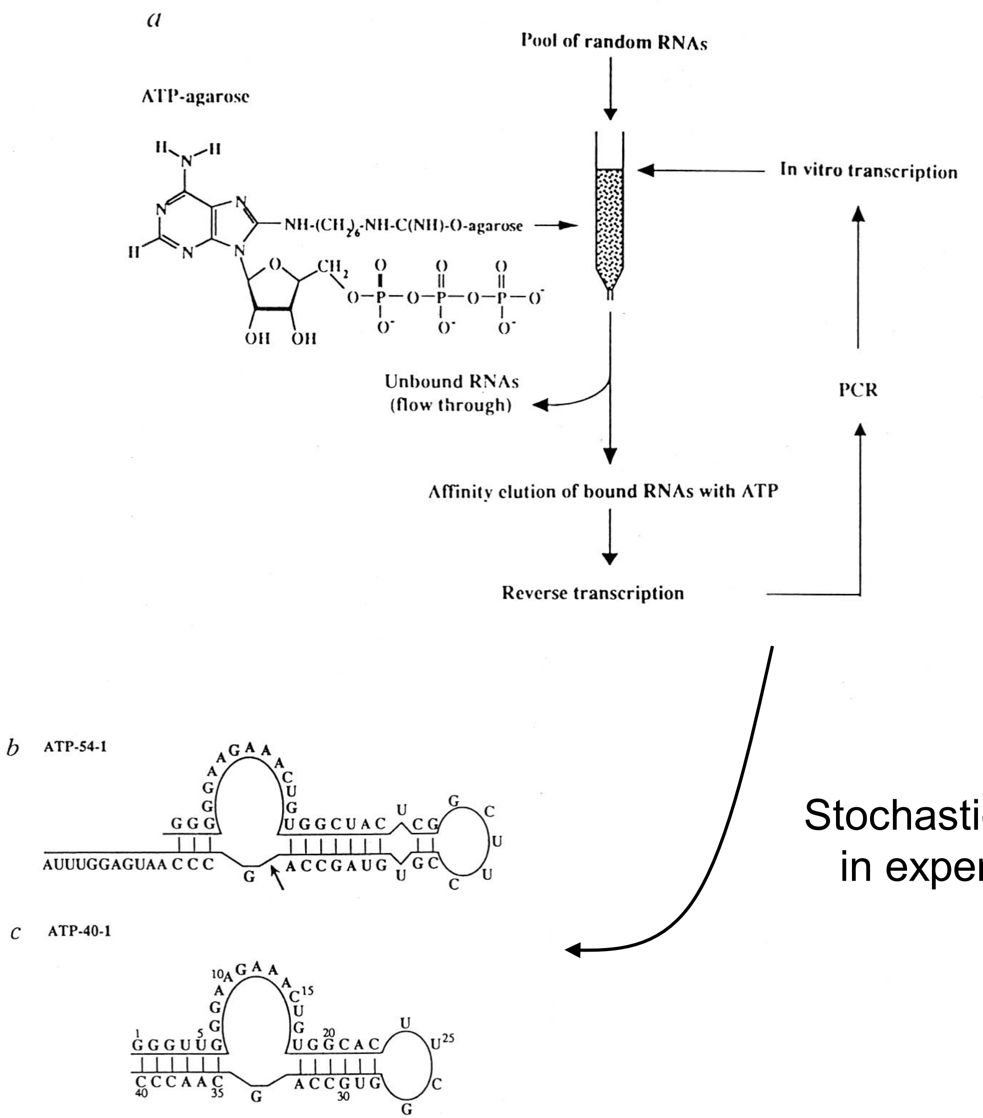
- As with DNA, multiple nucleotides provide the basis for a language, and the potential for double-strandedness provides a template for replication.
- Self-splicing introns proved that RNA can have catalytic properties – RNA is the only molecule known to be capable of both specifying a genotype and expressing a phenotype.
- RNA can bind amino acids, complement DNA, and has many other catalytic properties.
- The genomes of some viruses are entirely RNA based.
- *In vitro* experiments demonstrate the evolvability of RNA in simple systems.

Weak point:

- RNA lacks the ability to self-replicate.

In Vitro RNA Evolution Experiment

Example. Selection for an RNA motif that binds ATP.
M. Sassanfar and J. W. Szostak. 1993. Nature 364: 550-553.



Stochasticity and predictability
in experimental replicates

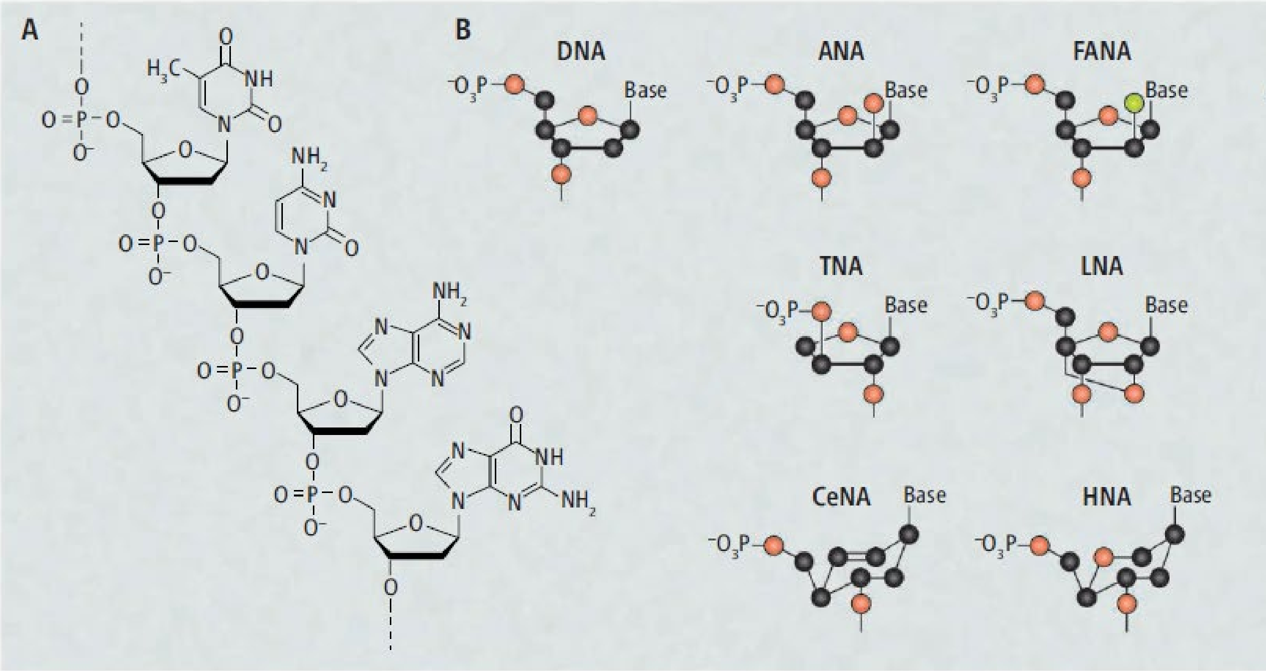
Arguments in favor of late-arriving DNA:

- An earlier RNA-Protein World would imply the existence of an RNA-World genetic code, consistent with the ubiquitous use of mRNAs, tRNAs, and rRNAs in transcription and translation.
- Modern cells derive deoxyribonucleotides from ribonucleotides.

Did DNA or its biosynthetic pathways evolve more than once?

- Putatively unrelated thymidylate synthases exist among different prokaryotic lineages.
- Apparently unrelated sets of DNA-replication proteins are used in bacteria vs. archaea.

Could There Have Been an Early Stage of Life Based on an Entirely Different Nucleic Acid?



XNA analogs of nucleic acids. DNA is composed of a deoxyribose–phosphate backbone and the standard four bases (**A**). Pinheiro *et al*. show that through transcription using engineered polymerases, DNA can be copied to various XNAs that are analogs of the DNA structure (**B**). Black, red, and green balls indicate carbon, oxygen, and fluorine atoms, respectively. These XNAs can in turn be reverse transcribed to DNA using other engineered polymerases. The current XNA amplification process relies on PCR amplification of a DNA intermediate (**C**).

- Numerous alternative forms of nucleic acids exist with altered sugar components, and/or with nucleoside analogs.
- Can be replicated with altered polymerases.
- Harbor potential value for the field of synthetic biology.

From: Joyce (2012, Science).

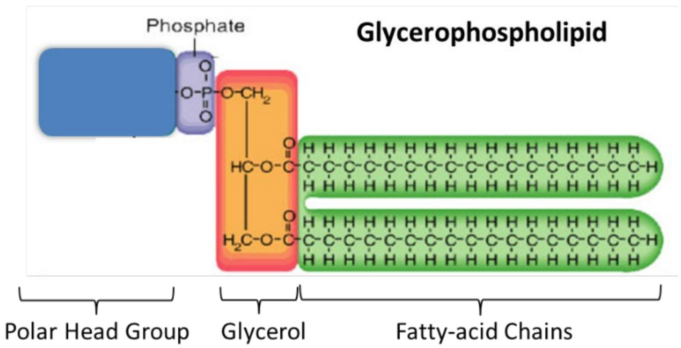
Membranes and the Origin of Individuality

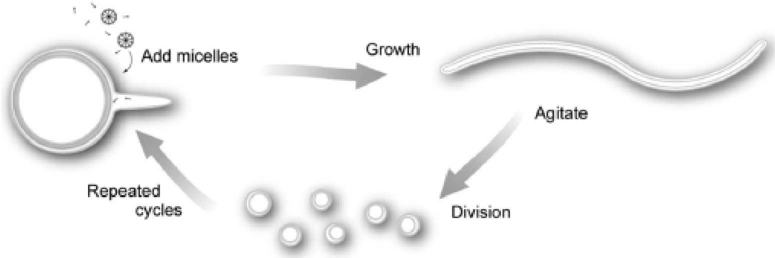
- Membranes: 1) enable cells to keep genomic products together at concentrations that promote intermolecular interaction;

  2) ensure that individuals reap the benefits of their own products, and avoid cheaters.

- Lipids were available: found in meteorites; and are produced when CO, CO_2, and H_2 gases interact with metal catalysts at high temperatures.

- Simple fatty acids spontaneously assemble into microspheres, owing to their amphiphilic nature (hydrophilic head groups, hydrophobic tails).





Spontaneous growth and fission of protocells:
as small micelles fuse with a larger vesicle, a surface to volume imbalance results in a filamentous protrusion that can be returned to smaller spheres by agitation.

A Hypothesis for Early Compartmentalization Without Membranes

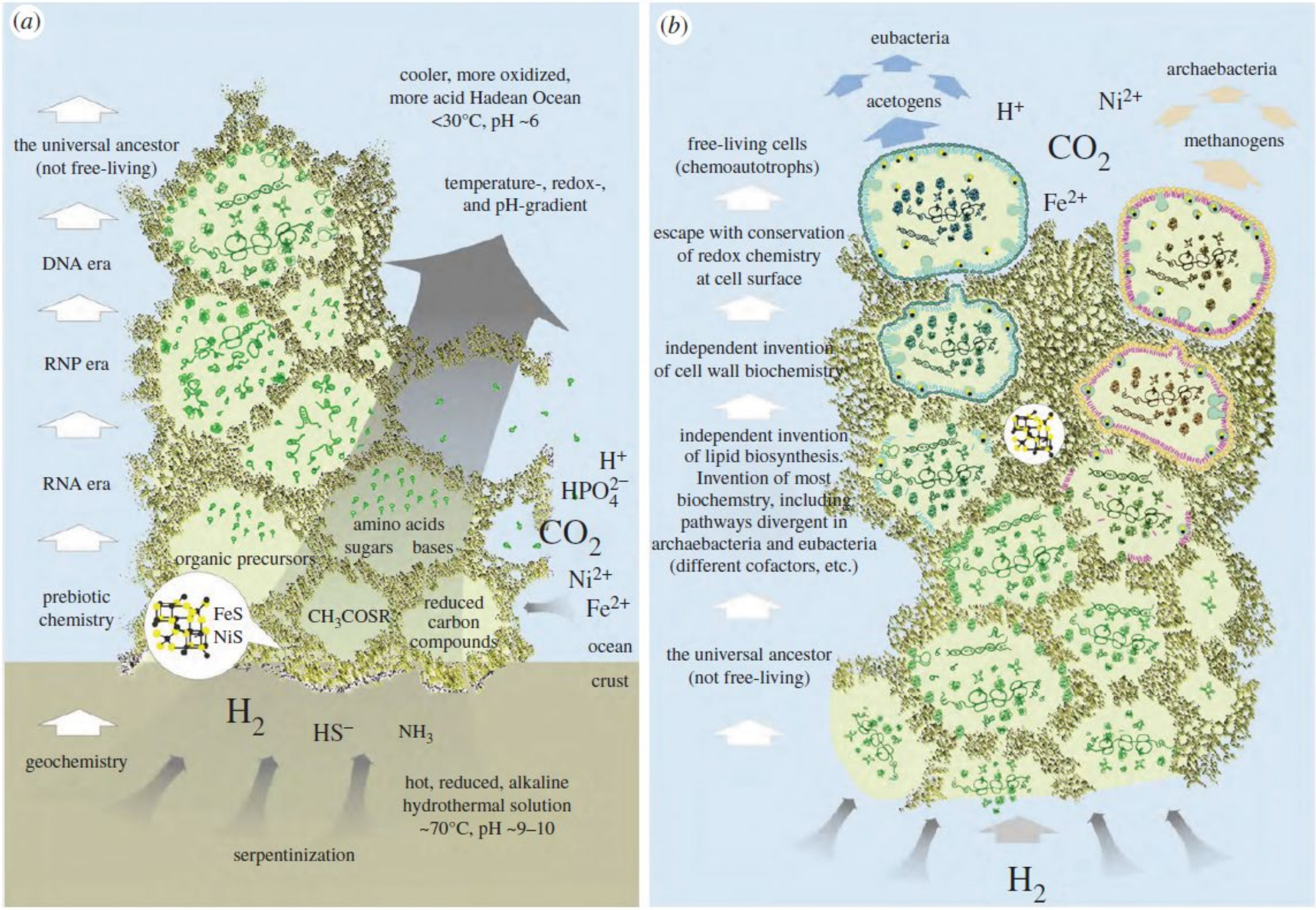
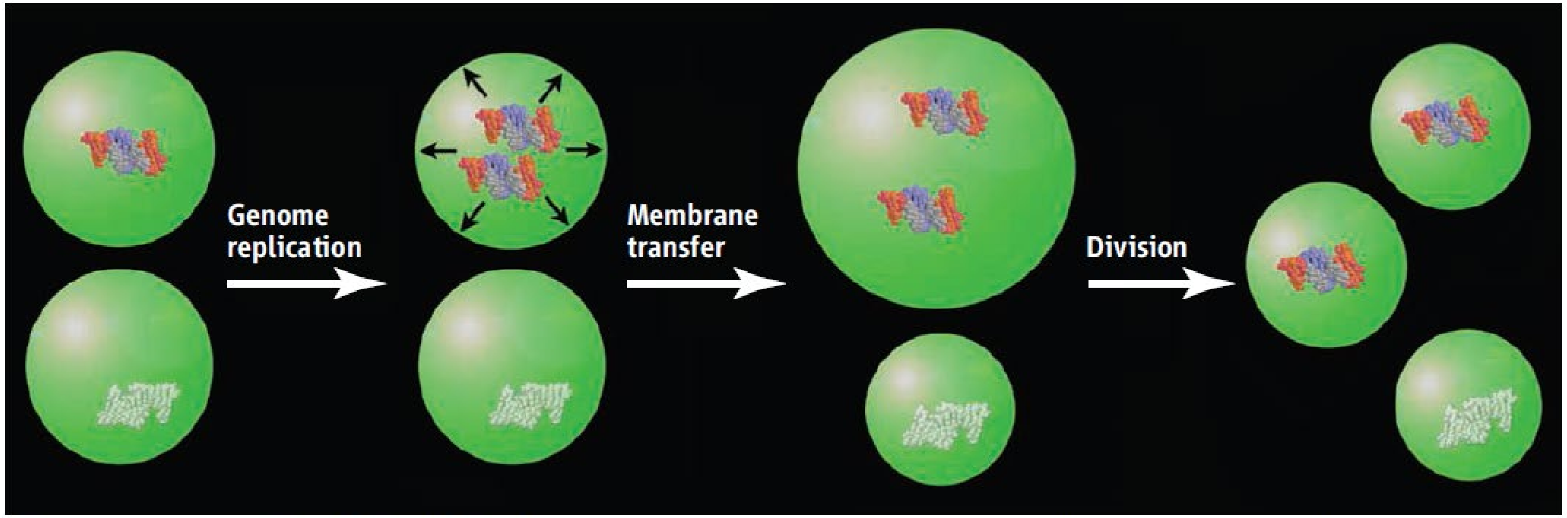


Figure 1. A scheme for the origin of cells [42,46].

From: Sousa et al. (2013, Phil. Trans. Roy. Soc. Lond. B).

The Intrinsic Capacity of an RNA-containing Protocell to Grow in the Absence of Any Encoded Mechanism for Growth

- Vesicles containing RNA experience osmotic stress that is relieved by recruiting lipids from empty vesicles, which reduces the local concentration.



The emergence of cellular behavior. Competition emerges as protocells containing replicating genomes steal membrane from protocells containing inactive molecules.

Chen (2006, Science)

Why the use of phospholipids in today's cells?

- Fatty-acid membranes are permeable to charged solutes, including charged nucleotides; phospholipid membranes present a much stronger barrier.

- Phospholipids aggressively compete for incorporation into simple lipid membranes – any cell predisposed to producing such compounds would assemble such membranes for purely physical reasons.

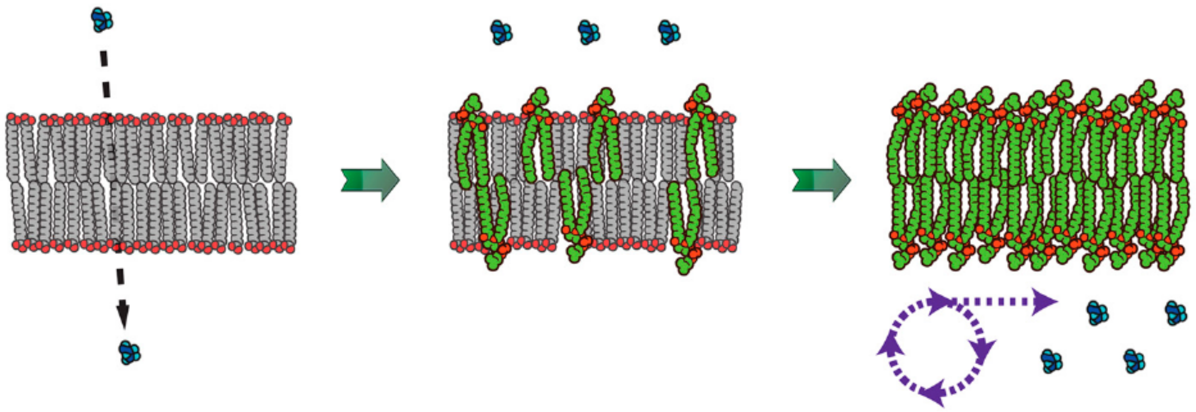


Fig. 5. Schematic for membrane-driven cellular evolution. The gradual transition from highly permeable primitive membranes (*Left*) to phospholipid membranes (*Right*) is driven by the selective growth advantage provided by increasing phospholipid content in the membrane. In turn, this transition in membrane composition imposes a selective pressure for the emergence of internalized metabolism to counter the reduced permeability of diacyl lipid membranes.

Budin and Szostak (2011, PNAS)

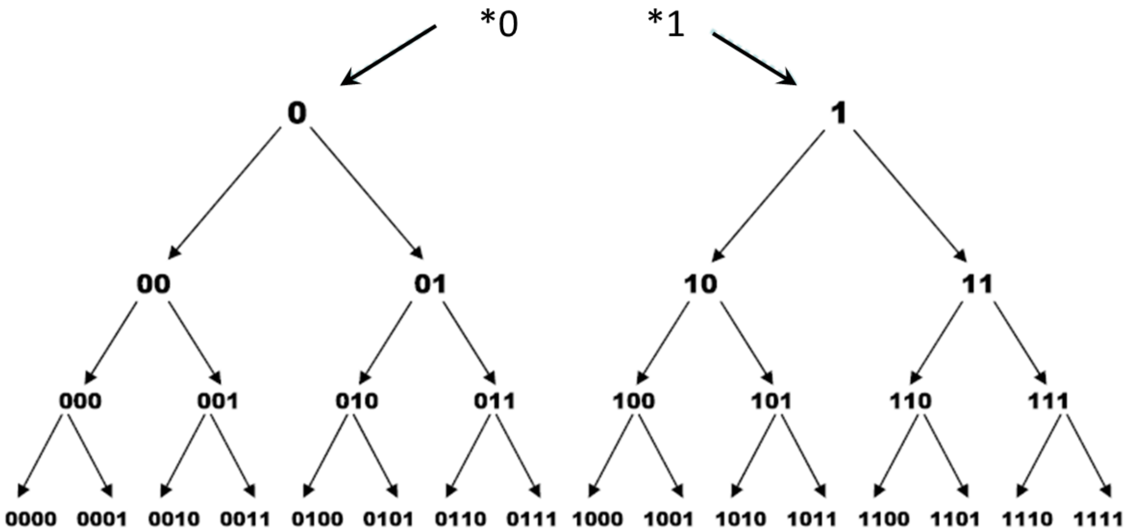
The Population-genetic Constraints on the Origin of Life

- Variation is essential to the operation of natural selection.

- A reliable mode of parent-offspring transmission of information is essential for maintaining the genotype-phenotype-fitness link critical to a productive response to selection.

- Replication fidelity and population sizes must high to avoid extinction by a mutational meltdown.

Evolution of Polymeric Diversity in the Absence of Self-replication / Selection



- Even prior to self-replication, a steady-state distribution of molecular types will be maintained, provided a continual supply of charged monomers.
- The relative frequency of different “strings” will depend on the relative attachment / detachment rates of 0 and 1.
- The population of molecules contains information in the terms of length and molecular sequence.

- If some sequences gained the capacity for self-replication, the new equilibrium distribution (and the success of the self-replicator) would depend on the rate of self- relative to natural-background replication and on the accuracy of replication.

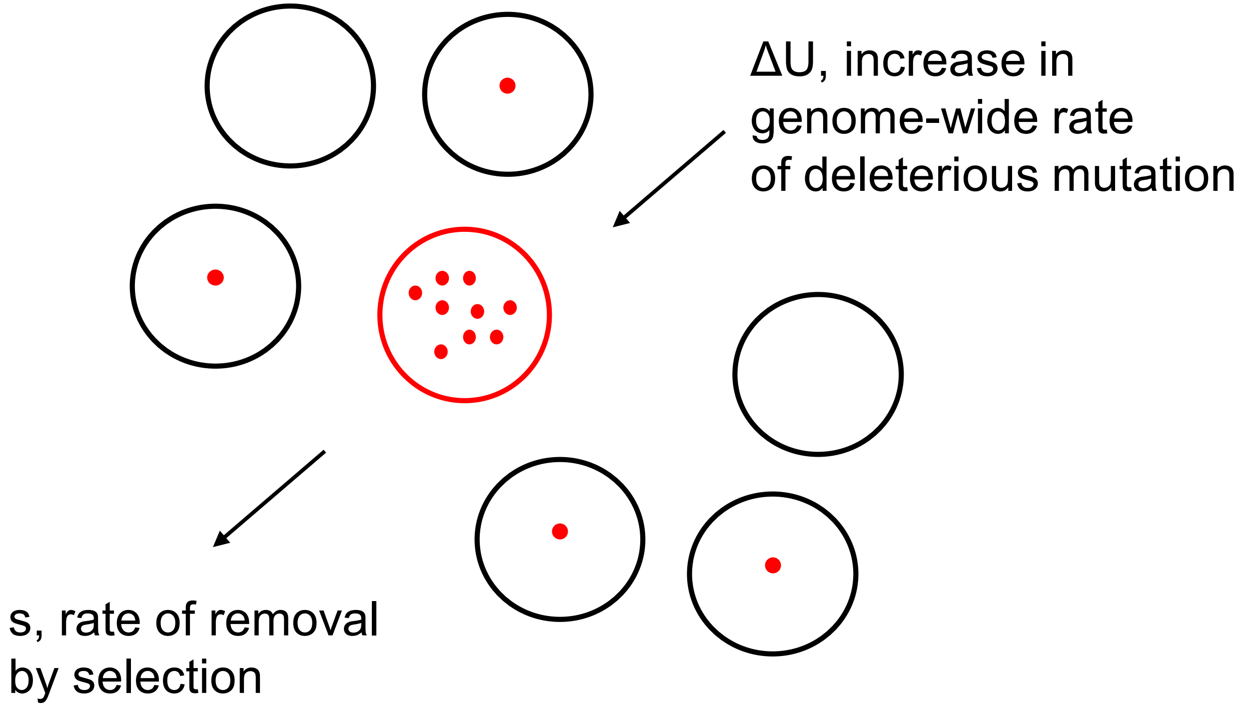
Population-Genetic Constraints on the Evolution of a Complex Genome

- Mutation creates variation in fitness among individuals, but most mutations are deleterious – if the genome is too large, there will be a negligible chance of producing a mutation-fee offspring.

- The degree to which natural selection can reduce the deleterious mutation rate is positively associated with the population size.

- Thus, there must be critical minimum population sizes and maximum genome sizes beyond which it becomes impossible to avoid extinction by a mutational meltdown.

The Magnitude of Selection Operating to Improve Replication Fidelity



Excess number of mutations at equilibrium = ΔU / s

↓

X

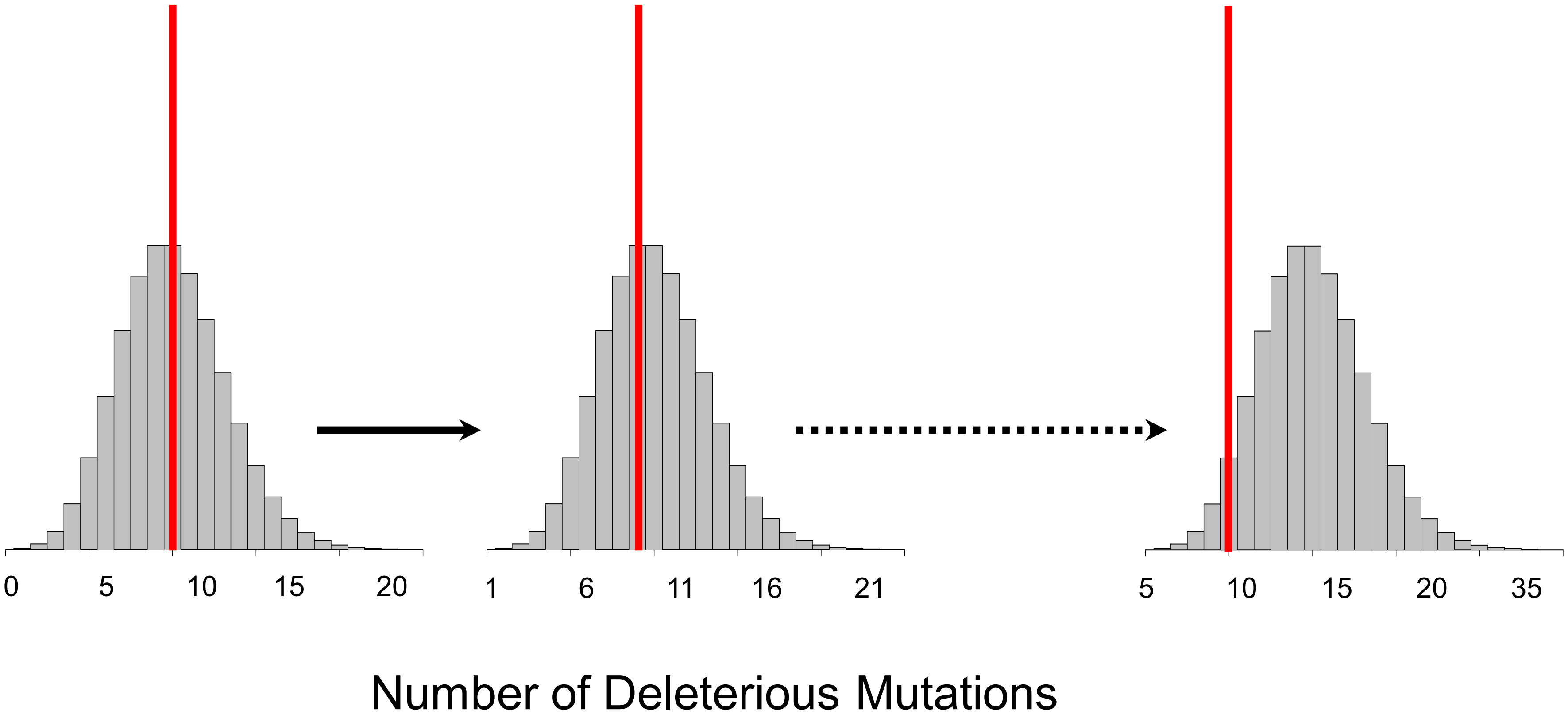
Effect / mutation = s

↓

Total effect on fitness = ΔU

- Selective advantage of an antimutator = decrease in genome-wide deleterious mutation rate.

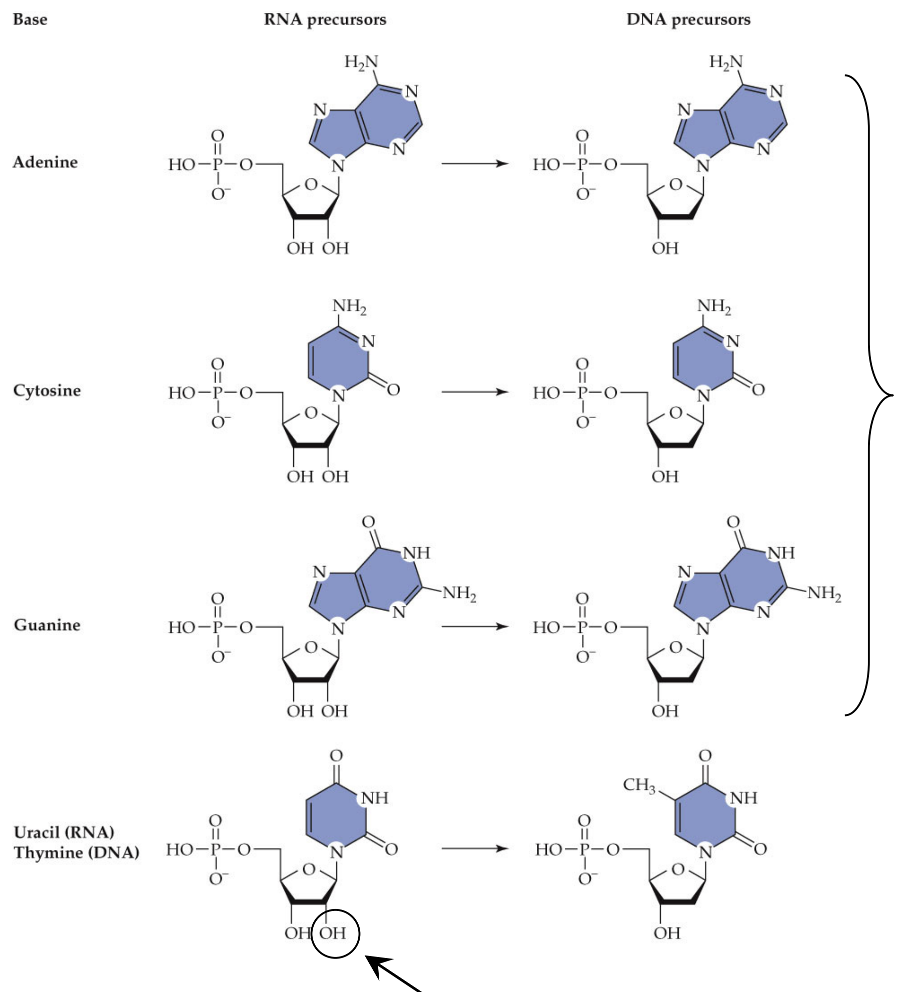
Muller’s Ratchet for the Loss of High-Fitness Classes



H. J. Muller

- The probability of loss of the best class increases with:
  - decreasing population size;
  - increasing genome size.

Entry Into the DNA World: reduced mutational vulnerability and a more permissive environment for genomic expansion.



Conversion by ribonucleotide reductases

Conversion by thymidylate synthase

One of the most common forms of spontaneous mutation – cytosine deamination

Structural instability resulting from second OH group

Most Viruses with Genomes > 30,000 bp are DNA-Based

Table 6.1. *Single-stranded RNA viruses*

Host	Virus	Number of nucleotides	Replicative units
Bacteria	Qβ	4500	1
	R17; MS2	4000	1
	ϕCb5	~ 4000	1
	PP7	~ 4000	1
Plants	Cucumber Mosaic	~ 3500	1
	Tobacco Mosaic	6800	1
	Como	~ 9000	1
Animals	Picorna (Polio)	~ 6000	1
	Toga (German Measles)	12000	Several
	Orthomyxo (Influenza)	12000	8
	Paramyxo (Measles)	22000	1
	Tumour	30000	3

- SARS-CoV-2 has a ~30 kilobase genome on a single chromosome, but has evolved a proof-reader domain on its polymerase.

Summary

- One conceptual problem with origin-of-life narratives is that they are just that.

  Demonstrations of the plausibility of single steps towards life often lead to increasingly confident adherence to specific views on the nature and order of events leading to the origin of life.

  This weaving of a series of low-probability events into a comprehensive scenario should be interpreted with caution.

- Putting aside these philosophical issues:

  Considerable experimental evidence suggests that rather than being improbable, the origin of life may be a nearly inevitable consequence of the geochemical environments on early Earth-like planets.