4. Life on Earth

- Basic biochemical structure
- Origin and Evolution of Life

Basic biochemical structure built on C-chain molecules:

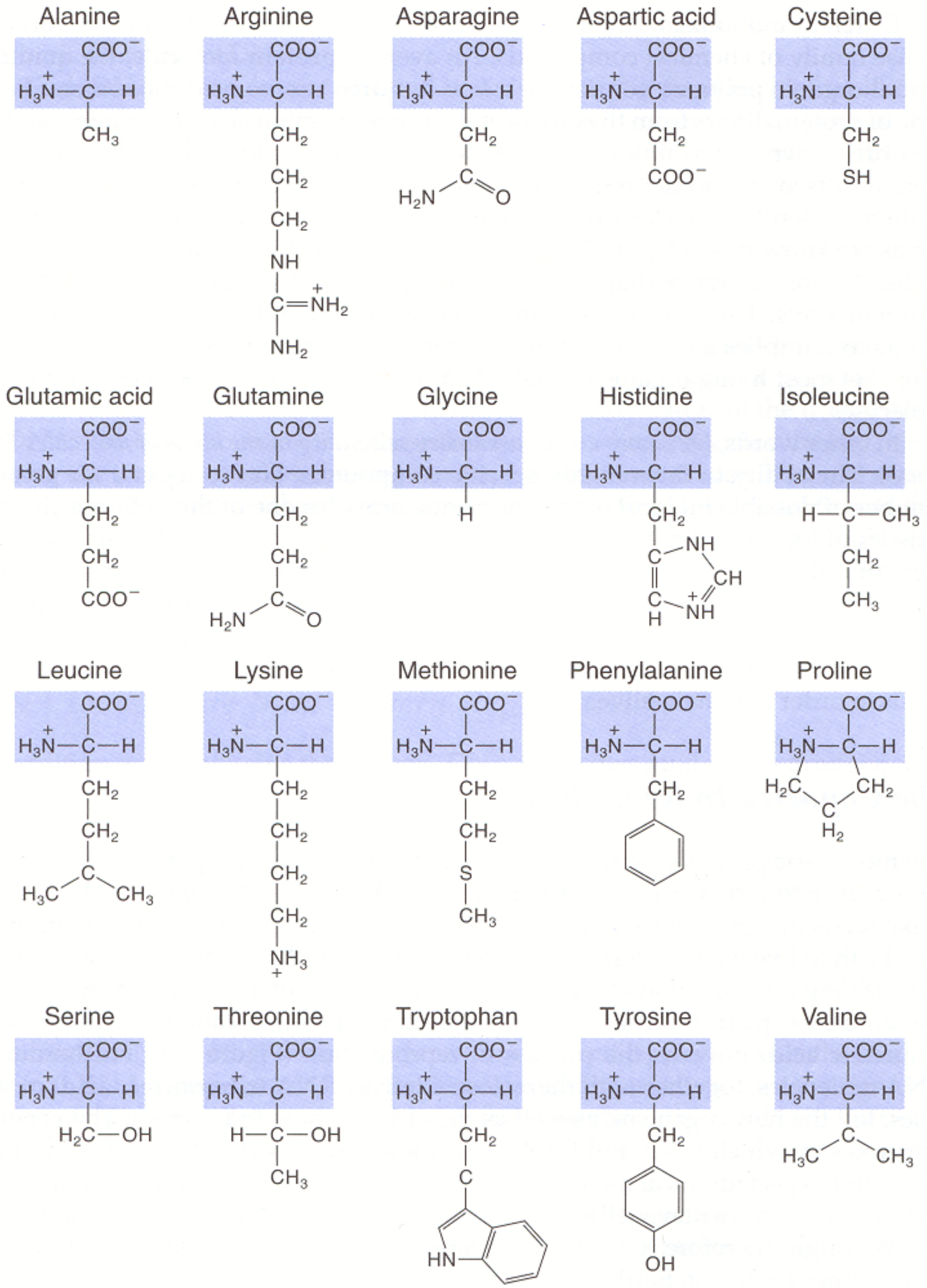
Simplicity:	Complexity:
Limited number of monomer building blocks	Large number of polymer molecules
20 Amino acids	→ proteins
4 Nucleotides	→ genetic code (DNA cross-links)
Sugars	→ structural and energy storing molecules

Distinguish between

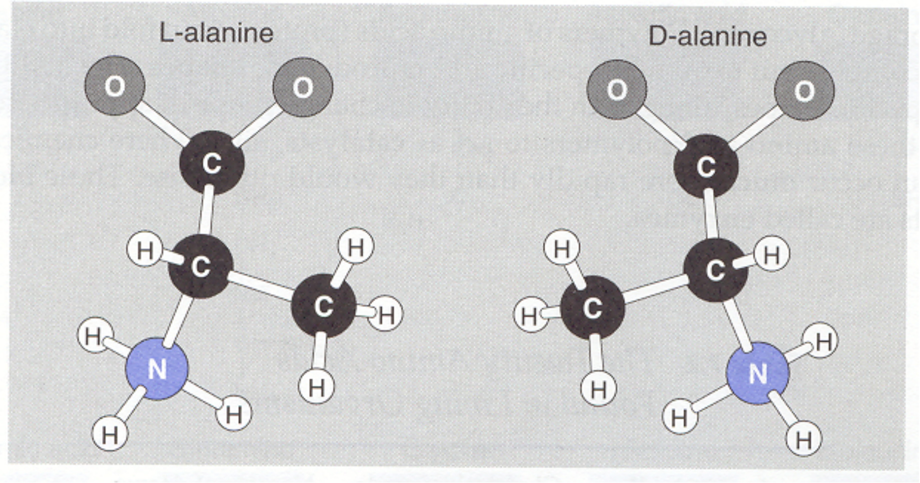
- Polymers repeating without ordering (e.g. cellulose, glycogen)
- Polymers with specific ordering of monomers → complex shapes and highly specific functionality (e.g. proteins)

20 amino acids → 10,000 proteins << “theoretical number” (see Lecture 1)

20 amino acids in terrestrial life



The "handedness" of terrestrial Life



All amino-acid monomers found in terrestrial life are "left handed" (in sense of arrangement of the bonds to CO_2, NH_3, H and the 4th branches)

Note: glycine is actually ambidextrous because of having an H atom as the 4th branch

- System of right-handed monomers would "work" equally well
- Having only one handedness certainly allows greater structural definition of long-chain molecules
- Evidence that amino-acids in the Murchison meteorite are 2-9% enhanced in Left vs. Right (not terrestrial contamination) – see Science Feb 14, 1997

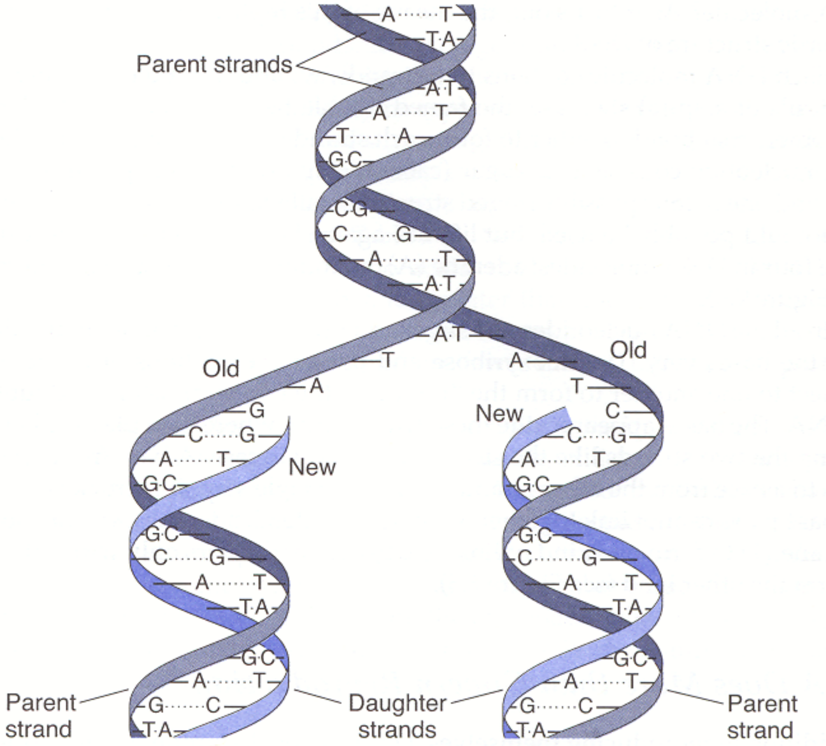
Role of DNA in replication and protein synthesis

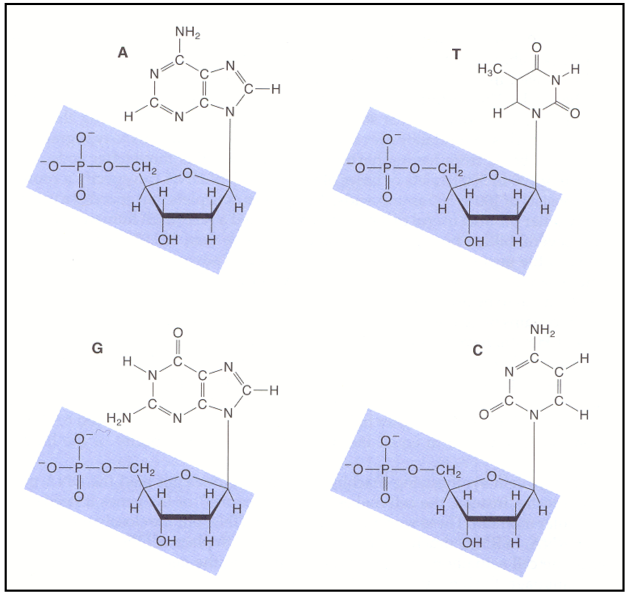
Double helix structure of DNA with ladder of sugar+phosphate and “rungs” composed of nucleotide pairs:

Adenine (A) + Thymine (T)
Cytosine (C) + Guanine (G)

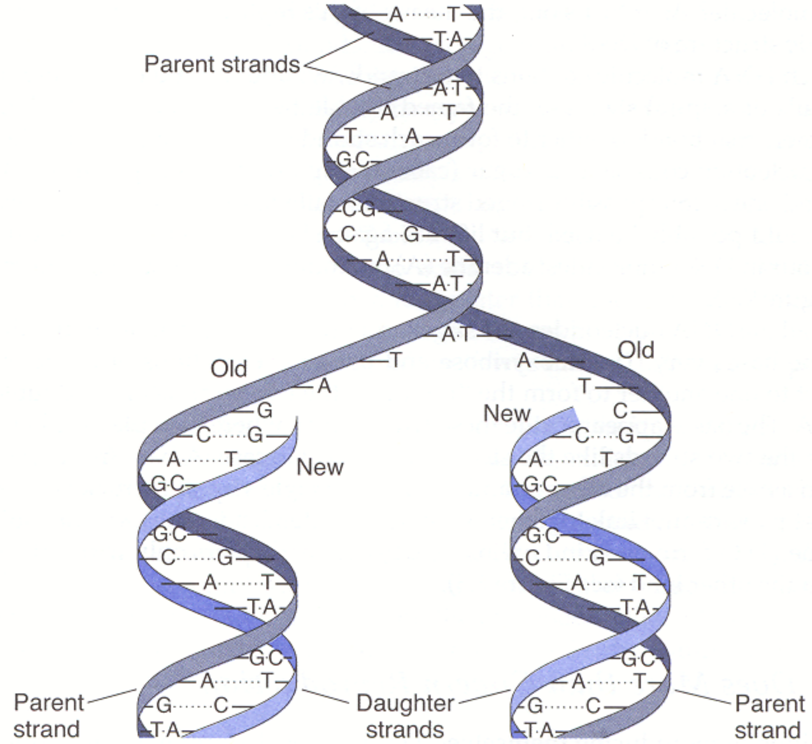
When DNA splits in half:

- each half contains all the required information to replicate the missing half
- each single strand can code for amino-acid chains (RNA)

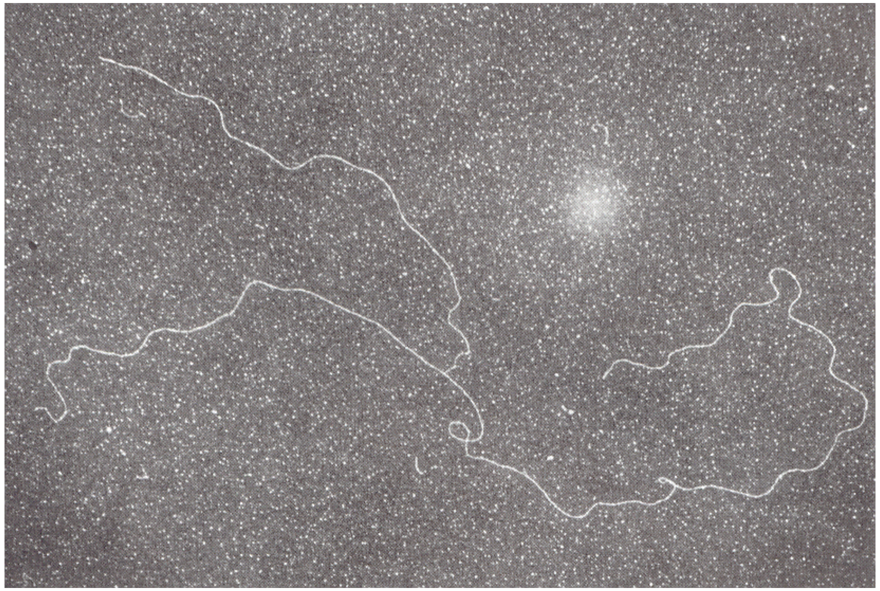




Role of DNA in replication and protein synthesis:

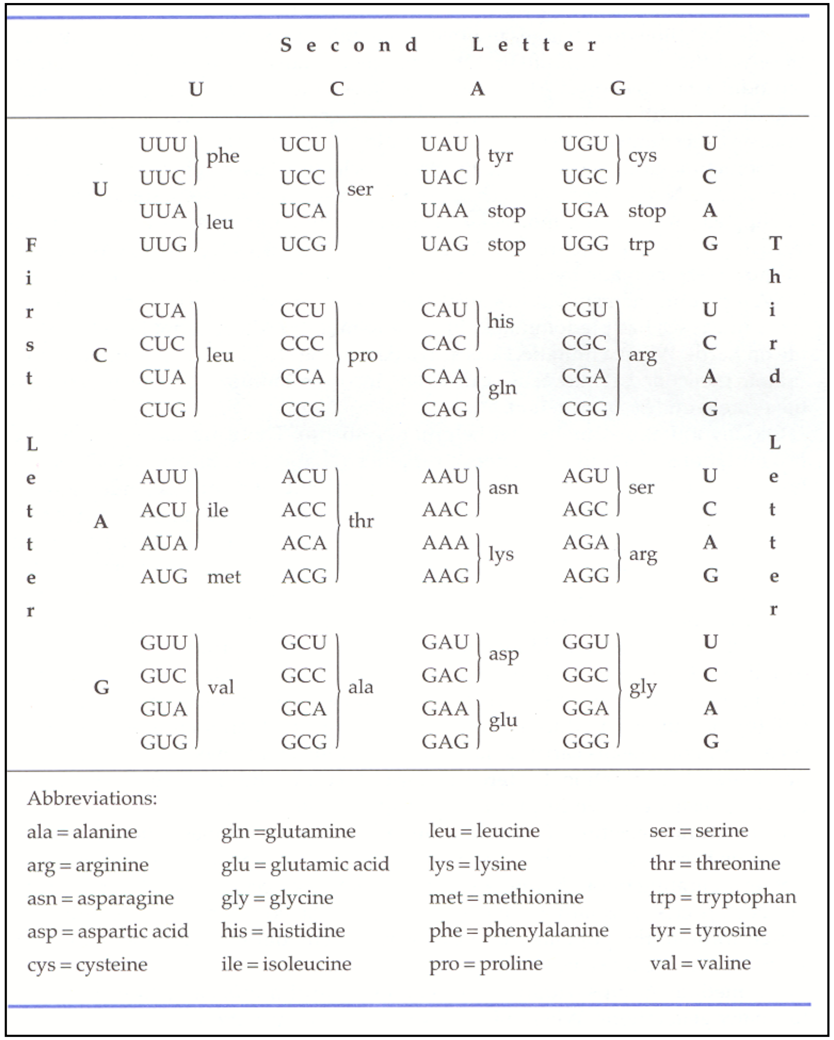


Replication through splitting now seen directly in electron microscopy



Electron micrograph of DNA in process of replication

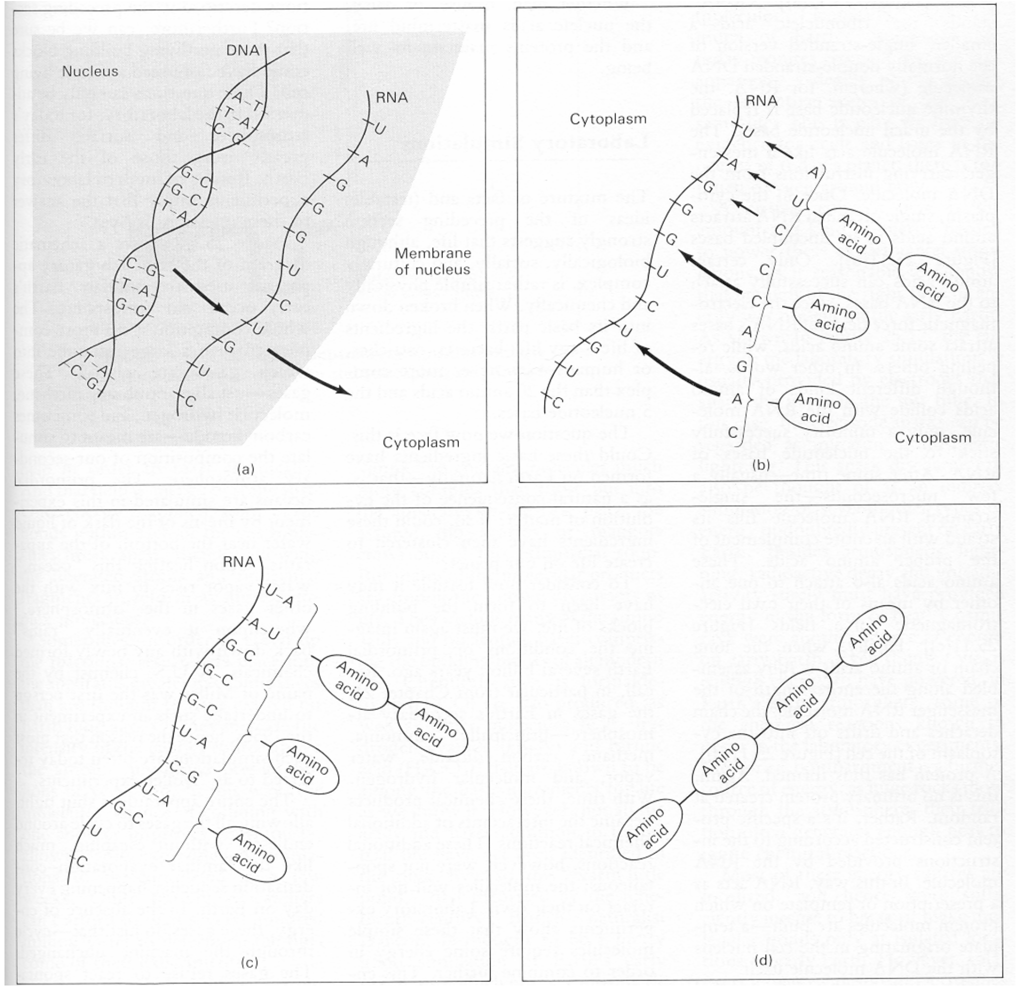
Triplets of bases on RNA code for individual aminoacids in protein chain (with some redundancy)



First Letter	Second Letter U		Second Letter C		Second Letter A		Second Letter 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	CUA	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
	ACU		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

Abbreviations:

ala = alanine
arg = arginine
asn = asparagine
asp = aspartic acid
cys = cysteine
gln = glutamine
glu = glutamic acid
gly = glycine
his = histidine
ile = isoleucine
leu = leucine
lys = lysine
met = methionine
phe = phenylalanine
pro = proline
ser = serine
thr = threonine
trp = tryptophan
tyr = tyrosine
val = valine



Note: 4 bases give $4^3 = 64$ triplets (> 20 amino acids + 1 “stop”) but only $4^2 = 16$ pairs.
2 bases would give only 8 triplets.

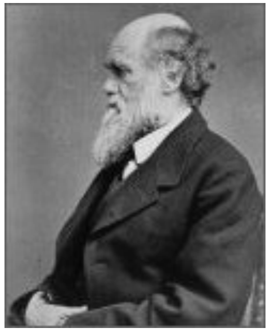
A key feature of replication is that it is not perfect

Inaccuracies in replication of DNA (mutations) due to:

- Radiation (cosmic rays or radioactivity)
- Chemical agents
- Rare errors of DNA copying machinery

Differential propagation of the resulting “trials”

- “natural selection” in one form or another



Charles Darwin

Thomas Malthus

Gregor Mendeleev

Energy for Life

- Chemical energy from surroundings (very small amounts)
- Geothermal energy
- Sunlight
- Other life products (sugars)

Origin of Life on Earth:

0. Earth's evolving atmosphere:

- Initially highly reducing atmosphere (NH_3, H_20, CH_4)
- Early: photo-dissociation by solar ultraviolet and H loss → less reducing atmosphere (CO, CO_2, N_2 and H_20)
- Much Later: Production of O_2 by Life itself (photosynthesis) → highly oxidizing atmosphere (N_2, O_2)

0.1 What about Panspermia ("Life" brought from elsewhere)?

- Long duration interstellar travel hazardous (γ-radiation, CR)
- "Martian" meteorites suggest interplanetary travel certainly not impossible, recent evidence for interstellar asteroids.
- Just defers the question of the "origin" to somewhere else (note: only increases the available time by x3)
- Complex organic molecules are observed in gas in space and in primitive meteorites in Solar System (see later).

Aside: “Complex” molecules detected in interstellar space (in giant molecular gas clouds)

- Detected with characteristic spectral lines (generally in mm waveband)

# atoms	Molecule	Name	Mass
10	$(CH_3)_2CO$	Acetone	58
	$(CH_2OH)_2$	Ethylene glycol	62
	CH_3CH_2CHO	Propanol	58
	CH_3OCH_2OH	Methoxymethanol	62
	CH_3C_5N	Methyl-cyano-diacetylene	89
	CH_3CHCH_2O	Propylene oxide	58
11	HC_8CN	Cyanotetra-acetylene	123
	C_2H_5OCHO	Ethyl formate	74
	CH_3COOCH_3	Methyl acetate	74
	CH_3C_6H	Methyltriacetylene	88
12	C_6H_6	Benzene	78
	C_3H_7CN	N-Propyl cyanide	69
	$(CH_3)2CHCN$	Iso-Propyl cyanide	69
13	C_6H_5CN	Benzonitrile	104
	$HC_{10}CN$	Cyanopenta-acetylene	147
60	C_{60}	Buckminsterfullerene	720
70	C_{70}	C_{70} fullerene	840

1. Chemical synthesis of organic chemicals:

- Requires reducing atmosphere and absence of free oxygen
- Miller-Urey experiments (highly reducing atmosphere) do make amino-acids, sugars, nucleotide bases etc: e.g. glycine

$$\mathrm{NH_3 + 2\,CH_4 + 2\,H_2O + energy \rightarrow C_2H_5O_2N + 5\,H_2}$$

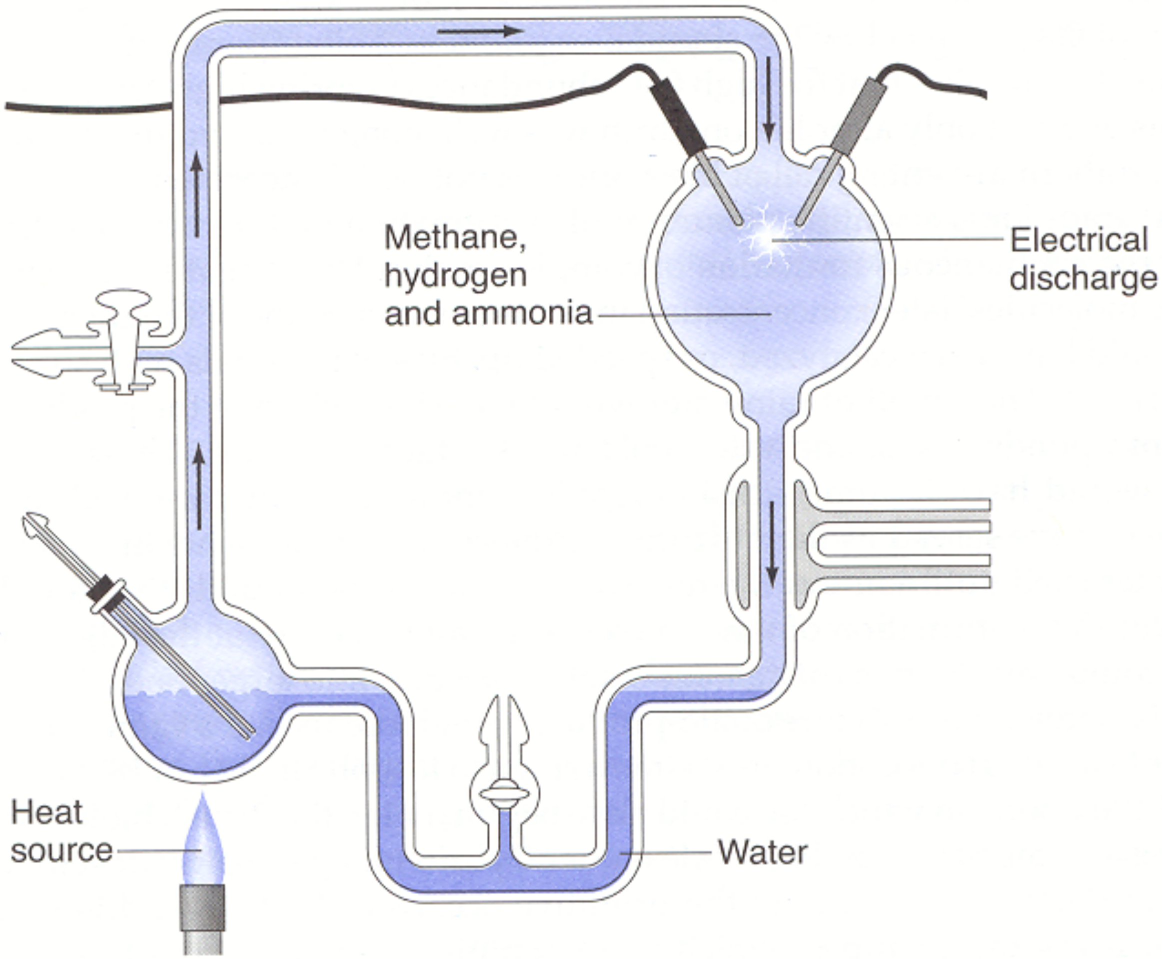
- Experiments with less reducing atmospheres also work:

$$\mathrm{3\,HCN + 2\,H_2O + energy \rightarrow C_2H_5O_2N + CN_2H_2}$$

- Chemical reactions in small surface liquid pools are attractive: preserve reagents, utilize solar energy etc., compared to undersea vents which have short duration, severe dilution problems, extreme conditions.
- Role of impacts on the young Earth? External origin of complex organic molecules? – e.g. amino-acids found in meteorites. These impacts can also certainly sustain the reducing atmosphere by replenishing volatile materials (CH_4, NH_3, H_20 etc).

e.g. Murchison meteorite contains 14,000 organic compounds including 70 amino acids, with a small excess of Left- over Right- chirality.

Miller-Urey-type experiments



Miller-Urey experiment (1953)
One of several similar conducted at the time.

Debate over precise composition of atmosphere (e.g. effect of volcanoes, CO_2 vs CH_4, H-content etc) but basic conclusions remain.

Analysis of sealed vials indicate results better than earlier analysis indicated: >25 amino acids produced. Most of the natural amino acids, hydroxyacids, purines, pyrimidines, and sugars have been made in variants of the experiment.

Bottom line: it is relatively easy to make basic organic building blocks of complex molecules in the kind of reducing atmosphere(s) that are expected on early Earth.

Amino acids found in Miller-Urey, Murchison and used in proteins

Amino acid	Miller-Urey	Murchison meteorite	Used in proteins
glycine	****	****	✔
alanine	****	****	✔
α-amino-N-butyric acid	***	****	✗
α-aminoisobutyric acid	****	**	✗
valine	***	**	✔
norvaline	***	***	✗
isovaline	**	**	✗
proline	***	*	✔
pipecolic acid	*	*	✗
aspartic acid	***	***	✔
glutamic acid	***	***	✔
β-alanine	**	**	✗
β-amino-N-butyric acid	**	**	✗
β-aminoisobutyric acid	*	*	✗
γ-amino-butyric acid	*	**	✗
sarcosine	**	***	✗
N-ethylglycine	**	**	✗
N-methylalanine	**	**	✗

2. Polymerization (how to make the initial strand of DNA)

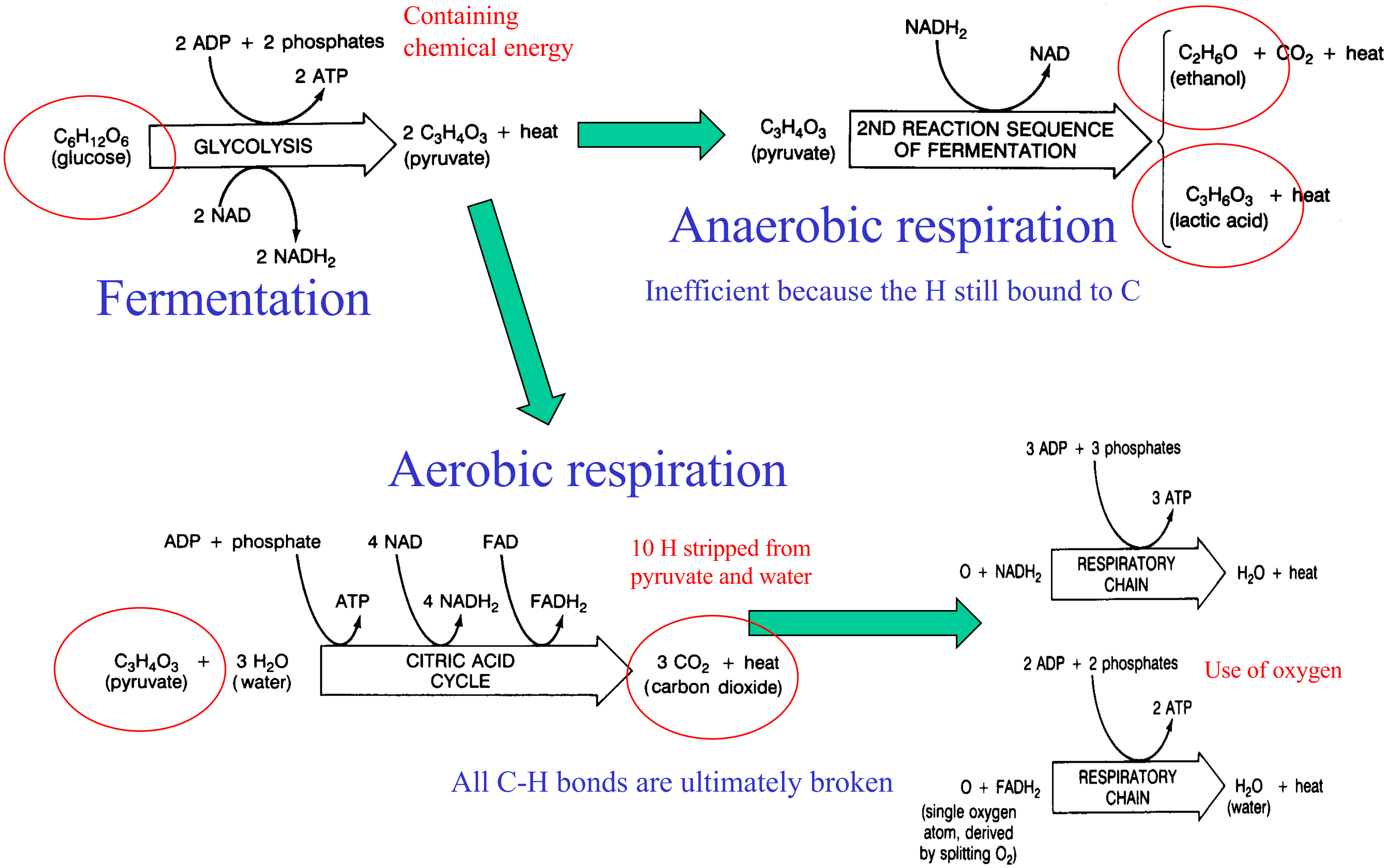
- Today, the production of polymers (e.g. proteins) from DNA (i.e. from RNA) itself depends on catalysis by other proteins hat are themselves made from RNA. Chicken and egg problem?
- Some sequences of RNA can perform catalysis function → "RNA-World": how to make the RNA?
- Interest in clays as structures to promote polymerization
  - Regular structures with high surface area
  - Organic compounds attracted to surface
  - Water is absorbed
  - Located near standing water

- Idea of possible earlier, simpler, but now <u>disappeared</u>, non-DNA "Life" as a kind of "scaffolding" on which DNA-life was constructed?

3. Emergence of prokaryotic cells (~3.5 Gyr ago)

- Proto-cells as droplets in solution?
- Simplest cells: Prokaryotes (without nucleus): Archaea and Bacteria (differ on basis of RNA production). Single long strand of DNA (several 1000 genes).
- Archaea likely developed first – still found as extremophiles (extreme conditions) and anaerobic systems. Depend on (small amounts of) local chemical energy.
- Source of free organic compounds gradually declined as atmosphere became less reducing
- 2-3 billion years ago, some prokaryotes develop photosynthesis (earliest form may have used H_2S) – producing O_2 and sugars (e.g. glucose) out of sunlight energy. This enabled more complex Life using more efficient aerobic respiration. First really good fossil evidence of complex Life.

Digression: Energy for cells - fermentation and respiration



$C_6H_{12}O_6$ (glucose)

GLYCOLYSIS

2 ADP + 2 phosphates → 2 ATP

2 NAD → 2 $NADH_2$

2 $C_3H_4O_3$ + heat (pyruvate)

Containing chemical energy

Fermentation

$C_3H_4O_3$ (pyruvate)

2ND REACTION SEQUENCE OF FERMENTATION

$NADH_2$ → NAD

C_2H_6O + CO_2 + heat (ethanol)

$C_3H_6O_3$ + heat (lactic acid)

Anaerobic respiration

Inefficient because the H still bound to C

Aerobic respiration

$C_3H_4O_3$ (pyruvate) + 3 H_2O (water)

CITRIC ACID CYCLE

ADP + phosphate → ATP

4 NAD → 4 $NADH_2$

FAD → $FADH_2$

3 CO_2 + heat (carbon dioxide)

10 H stripped from pyruvate and water

All C-H bonds are ultimately broken

O + $NADH_2$

RESPIRATORY CHAIN

3 ADP + 3 phosphates → 3 ATP

H_2O + heat

O + $FADH_2$ (single oxygen atom, derived by splitting O_2)

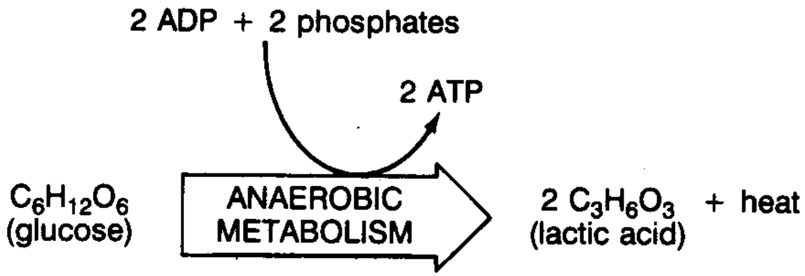
RESPIRATORY CHAIN

2 ADP + 2 phosphates → 2 ATP

H_2O (water) + heat

Use of oxygen

Net effect of “burning” glucose in a controlled way: the effect of oxygen



2 ADP + 2 phosphates

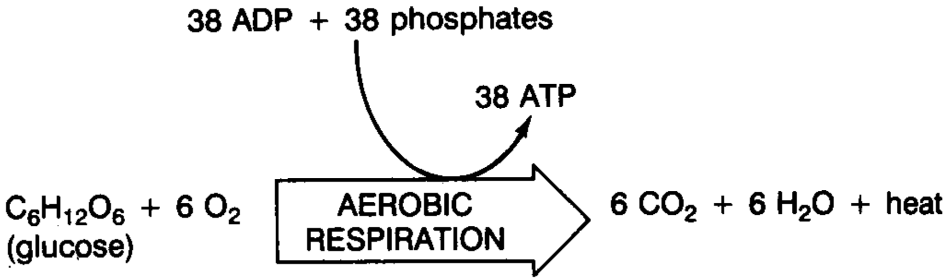
2 ATP

$C_6H_{12}O_6$ (glucose)

ANAEROBIC METABOLISM

2 $C_3H_6O_3$ (lactic acid) + heat

2% energy “efficiency”
(energy in *useful* form)



38 ADP + 38 phosphates

38 ATP

$C_6H_{12}O_6$ + 6 O_2 (glucose)

AEROBIC RESPIRATION

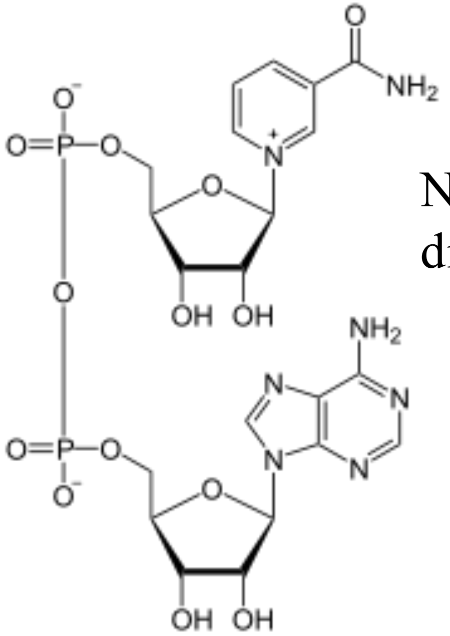
6 CO_2 + 6 H_2O + heat

40% energy “efficiency”

(c.f. 0% in a wood fire – burning uncontrollably – and 30% in modern combustion engine)

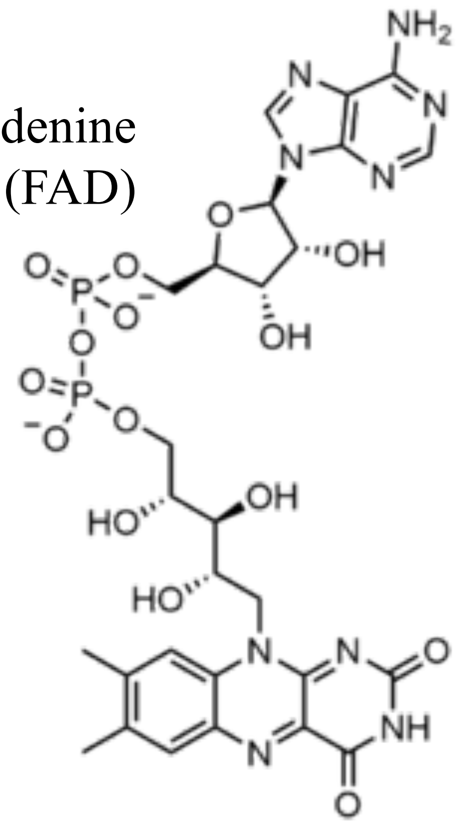
Availability of oxygen allows vastly more efficient extraction of energy

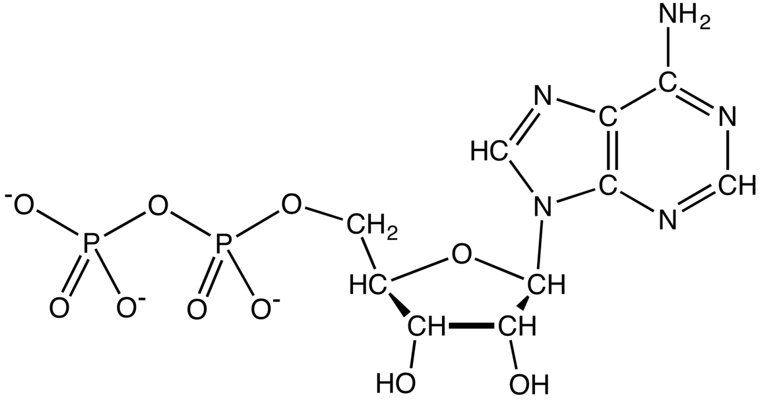
These energy-transporting molecules are very complicated...



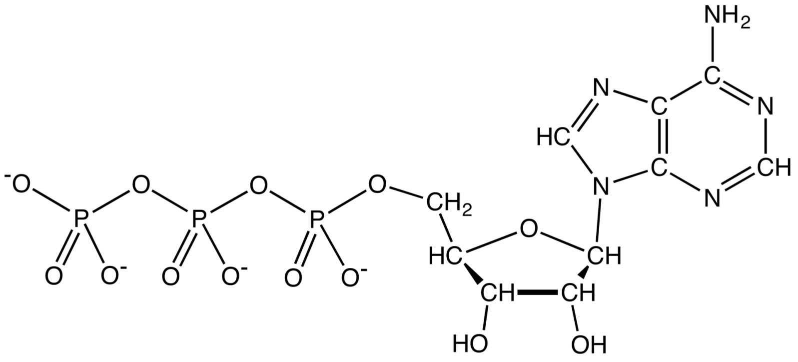
Nicotinamide adenine dinucleotide (NAD)

Flavin adenine dinucleotide (FAD)





Adenosine diphosphate (ADP)



Adenosine triphosphate (ATP)

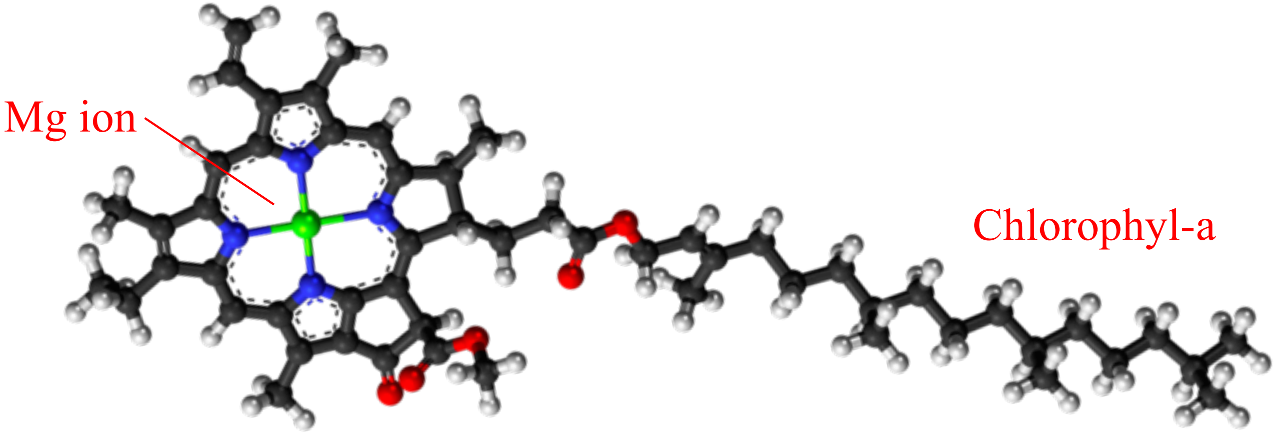
Photosynthesis: extraction of energy from sunlight to make glucose and oxygen

Easy to write the basic reaction, but the details are extremely complex:

Photosynthesis:

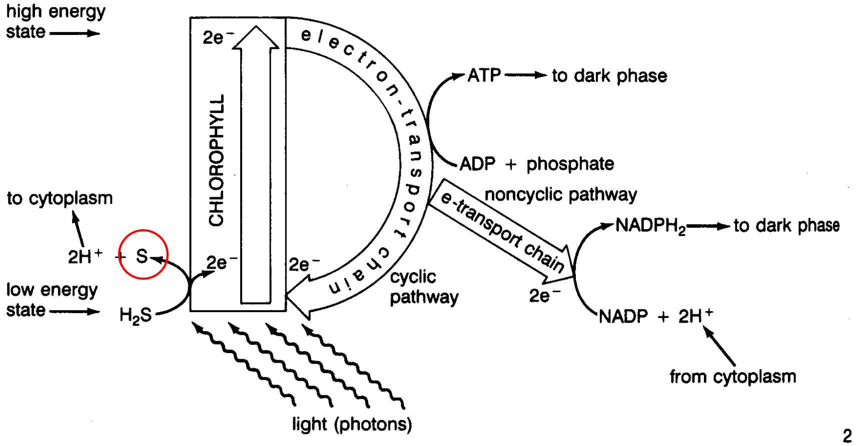
(some bacteria use S not O)

$$\boxed{6\ CO_2 + 6\ H_2 0 + \text{light} \rightarrow C_6H_{12}O_6 + 6\ O_2}$$

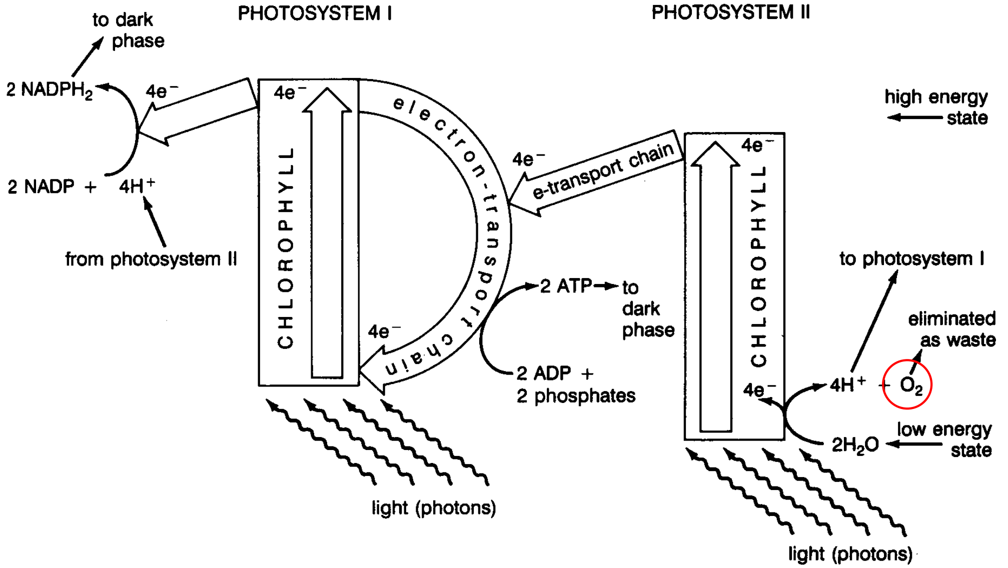


Photosynthesis: extraction of energy from sunlight to make glucose

Two forms: both produce glucose to be used in fermentation and respiration. The most efficient form also liberates O_2 as waste.



Anaerobic photosynthesis (light phase) liberating e.g. sulphur from H_2S



Aerobic photosynthesis (light phase) liberating oxygen from H_20.

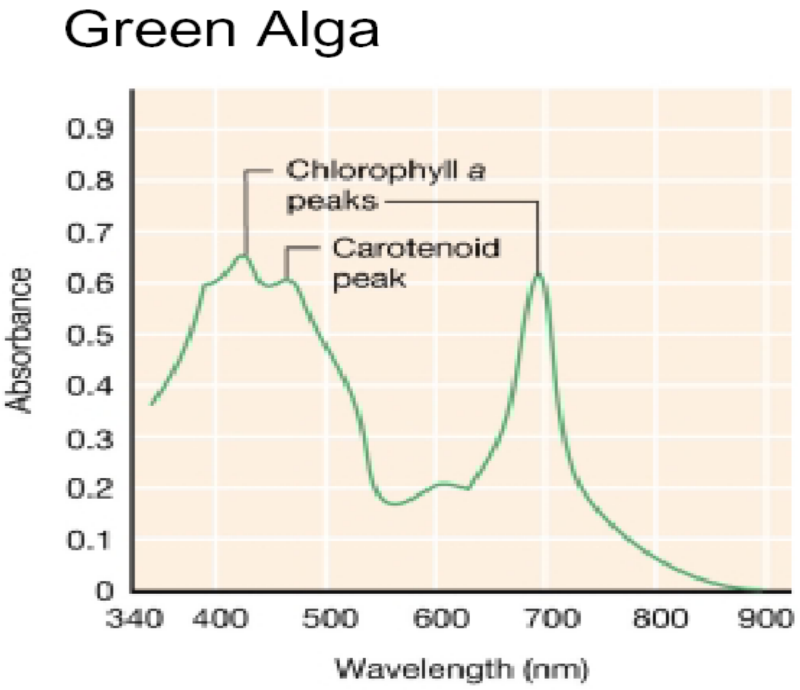
Adds new Photosystem II producing oxygen to Photosystem I from anaerobic.

High energy photons needed to split (abundant) water

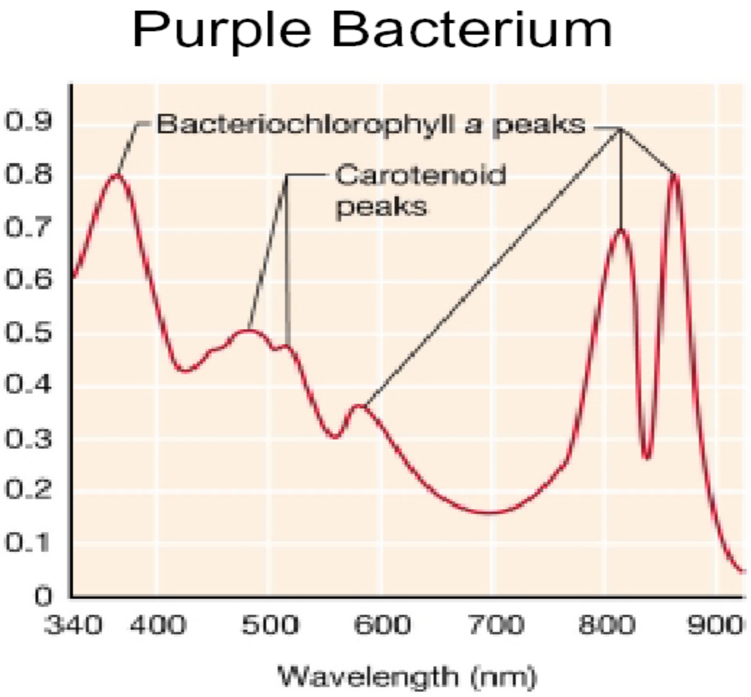
Common dark phase: ATP, $NADPH_2$ synthesize glucose out of CO_2

Multiple electronic transitions in chlorophyl produce a complex absorption spectrum of light (which would be detectable remotely!)

Oxygenic photosynthesis under aerobic conditions


Green Alga

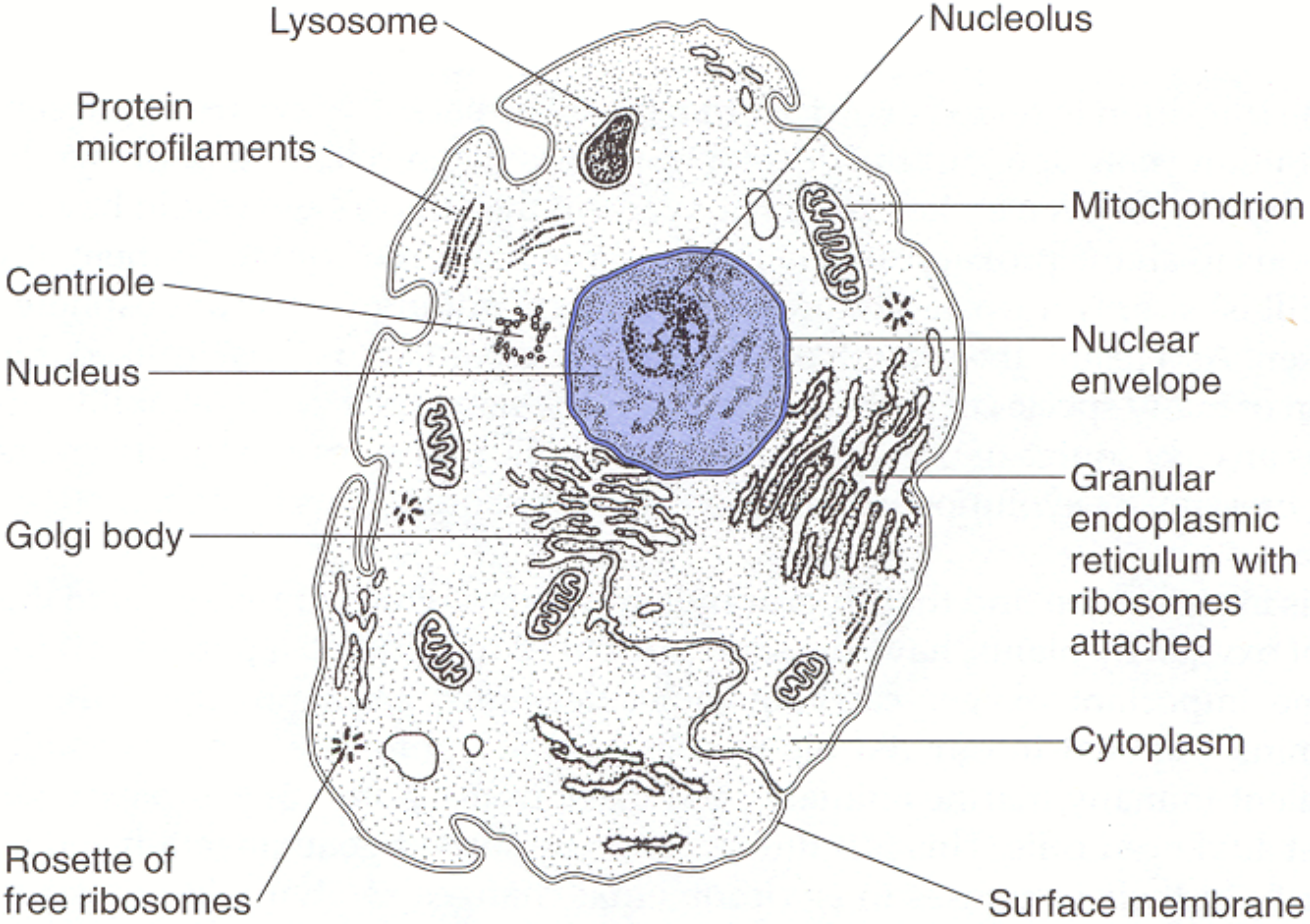
Anoxygenic photosynthesis under anerobic conditions


Purple Bacterium

Origin of Life on Earth (continued)

4. Emergence of Eukaryotic cells (~2.7 Gyr ago)
    - Single cells with distinct cell nucleus, plus specialized organelles (mitochondria, chloroplasts etc.)
    - Contain 10-1000 times more DNA than in bacteria
    - Incorporation of prokaryotes into proto-Eukaryote (similar gene systems to bacteria)?
    - Development of sexual reproduction (meiosis), producing vastly increased genetic sampling (about 1 Gyr ago?)
    - Development of multi-cellular organisms (multiple development)
        - → e.g. Jellyfish (<u>700 Myr ago</u>)***

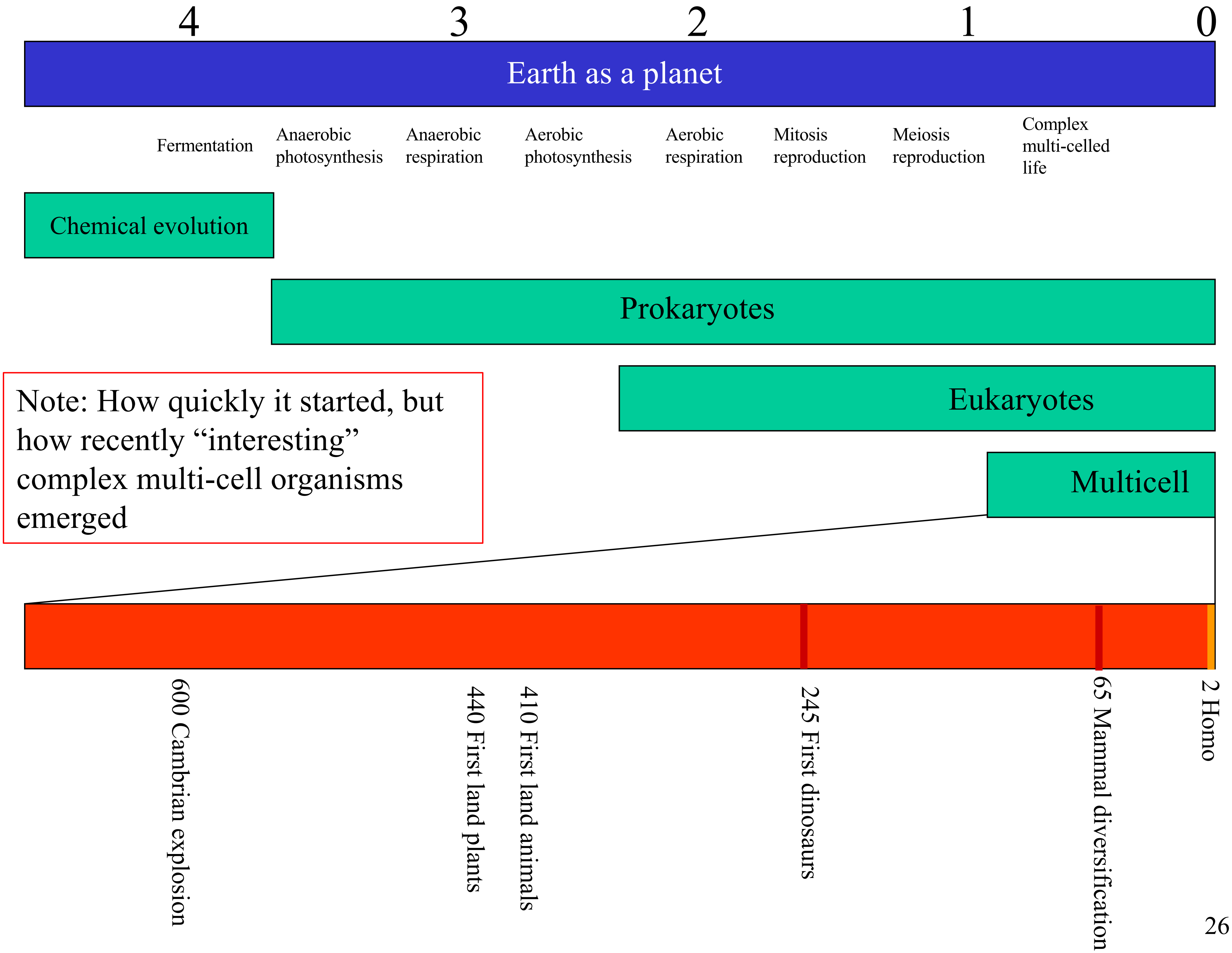
*** note that the Earth was already 85% of present age


Lysosome
Nucleolus
Protein microfilaments
Mitochondrion
Centriole
Nuclear envelope
Nucleus
Granular endoplasmic reticulum with ribosomes attached
Golgi body
Cytoplasm
Rosette of free ribosomes
Surface membrane

Origin of Life on Earth (continued)

5. Emergence of advanced animals

- Cambrian explosion (600 Myr ago) → anatomically modern animals (two eyes, legs, shells/carapaces/skeletons, etc.)
- Colonization of land by plants and arthropods (440 Myr ago)
- Appearance on land of vertebrates, amphibians, insects, seed plants (410 Myr ago)
- Diversification of reptiles (285 Myr ago)
- Emergence of dinosaurs and reptiles (245 Myr) after a mass extinction
- Diversification of mammals (65 Myr) after another mass extinction
- Appearance of Homo (2 Myr)….



4 3 2 1 0

Earth as a planet

Fermentation

Anaerobic photosynthesis

Anaerobic respiration

Aerobic photosynthesis

Aerobic respiration

Mitosis reproduction

Meiosis reproduction

Complex multi-celled life

Chemical evolution

Prokaryotes

Eukaryotes

Multicell

Note: How quickly it started, but how recently “interesting” complex multi-cell organisms emerged

600 Cambrian explosion

440 First land plants

410 First land animals

245 First dinosaurs

65 Mammal diversification

2 Homo

Short diversion on statistics:

What if anything can be inferred from the apparently rapid emergence of the most basic forms of Life on Earth?

Life is easy and common in the Universe?? Not necessarily!

Interesting point on conditional probabilities….

Suppose we have a sequence of "steps" in a process, each of which has a certain chance p_i of happening per unit time. Generally, we would expect to have to wait a time $t_i \sim p_i^{-1}$ for a given step to occur.

A step with $p_i > t_{avail}^{-1}$ is "easy" and is likely to happen. A "difficult step" has $p_i < t_{avail}^{-1}$ and is unlikely to happen in the available time.

Clearly, if we have one or more difficult steps in the process, the chance that all the steps successfully take place will be small.

But, if we look only at just those systems that were, by great good luck, able to complete *all* the steps, what is the distribution of the time that was spent "waiting" for each step in the chain to occur?

A perhaps counter-intuitive answer:

- For “easy” steps, the average wait, $<t_i> \sim p_i^{-1}$ (as expected)
- For “difficult” steps, $<t_i>$ is *independent* of the difficulty p_i of that step, and is instead given just by the number N of of difficult steps $<t_i> \sim t_{avail} /(N+1)$.

Example: Run a simulation with a total time interval $t_{avail} = 10^6$ units, and eight “steps” with individual probabilities of occurring per unit time of 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4},10^{-5} (“easy”) and then 10^{-6}, 10^{-7}, 10^{-8} (increasingly hard steps) – the precise ordering does not matter.

The chance of them all happening in 10^6 time steps is very small, of order $10^{-2} \times 10^{-1} \times 1 \sim 10^{-3}$ (actually $\sim 3.3 \times 10^{-4}$). The distribution of $<t_i>$ in the few lucky ones that complete all steps is as follows:

	“easy” steps, $<t_i> \sim p_i^{-1}$					“difficult” steps, $<t_i> \sim t_{avail} (N+1)^{-1}$			
Step p_i^{-1}	10	100	1000	10^4	10^5	10^6	10^7	10^8	Left at end
$<t_i>$	9.8	95.8	989	9,810	68,385	207,920	257,860	254,910	220,015
$\sigma(t_i)$	9.3	99.5	837	11,236	60,212	164,590	179,680	189,212	161,118

Note that the dispersion in the individual step lengths $\sigma(t_i)$ is large, $\sigma(t_i) \sim <t_i>$

The explanation of this perhaps counter-intuitive result is actually straightforward:

If we know that all the hard steps were completed, then the probability that we have a particular configuration of t_i is given by the product of the probabilities that each step completed in its t_i, which is $p_i t_i$ (for $p_i t_i << 1$).

$$\boxed{P = \prod_i t_i p_i = \prod_i p_i \prod_i t_i}$$

The Πp_i is constant, so P is maximized when Πt_i is maximized, subject to the constraint that $\Sigma t_i + t_{end} = t_{avail}$, where t_{end} is the time after the last step completed before the end of the experiment. It is easy enough to show that Πt_i is maximized (subject to this constraint) when all the t_i and t_{end} are equal, i.e.

$$\boxed{t_i = \frac{t_{avail}}{(N+1)}}$$

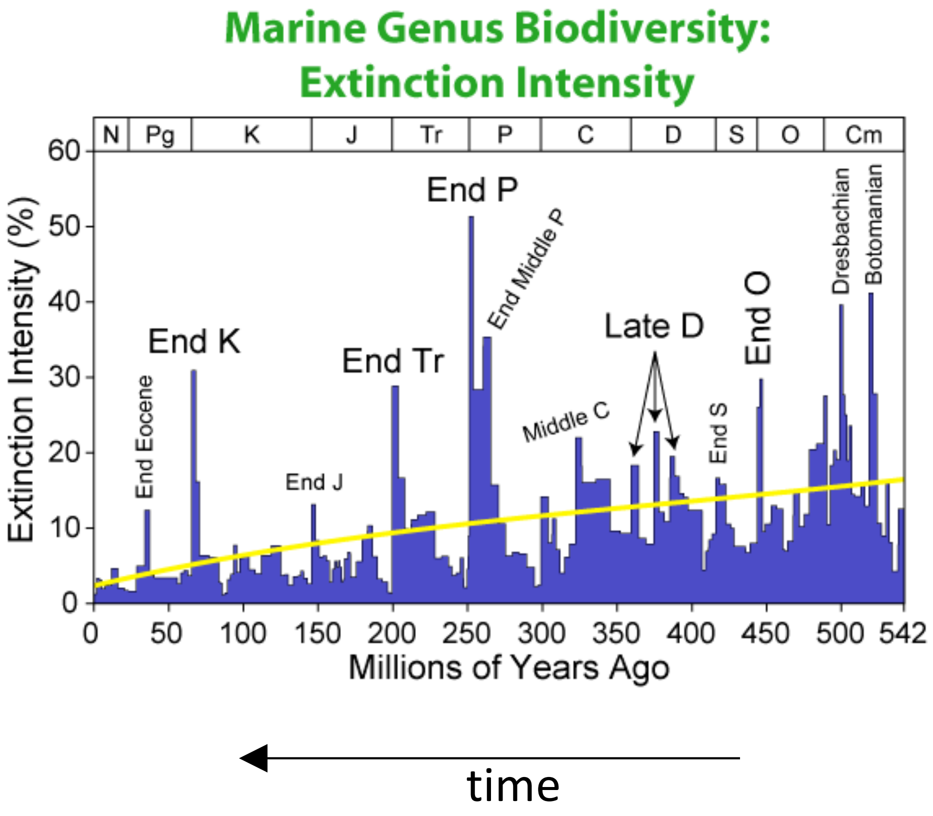
On Earth $t_1/t_{avail} \sim 0.1$. This may tell us nothing at all about the value of p_1. It is just as consistent with all the p steps being very unlikely (all p_i being very small) but having $N \sim 10$.

Continuing interactions of Life on planets with the astronomical environment

1. “Nearby” explosions in the Galaxy, e.g. γ-ray bursts

2. Late epoch impacts leading to mass extinctions

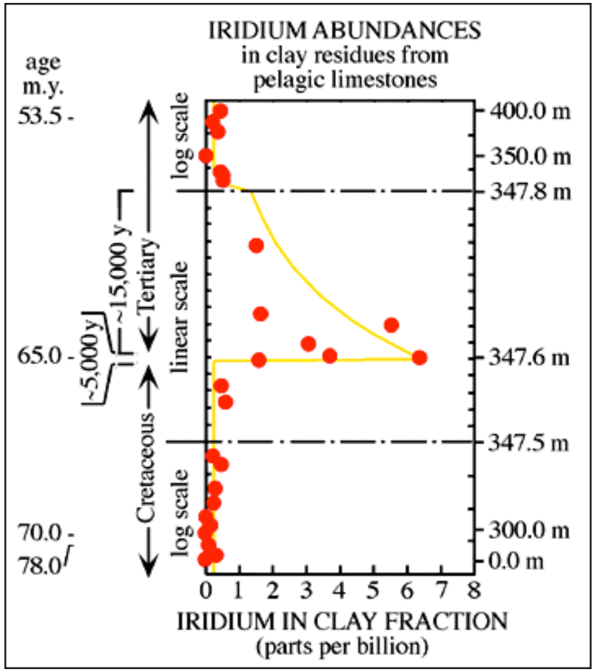
- The rate at which species (families etc) become extinct shows wide variations, leading to “sudden” loss of large numbers (e.g. Permian/Triassic 50% of marine families = 95% of marine species, 70% of land species)
- Easiest to see in marine fossil record, but the effect is distributed across environments and across the planet (synchronous?)
- Almost impossible to estimate the duration if less than 10^4-10^5 years due to the limited resolution of the fossil record.

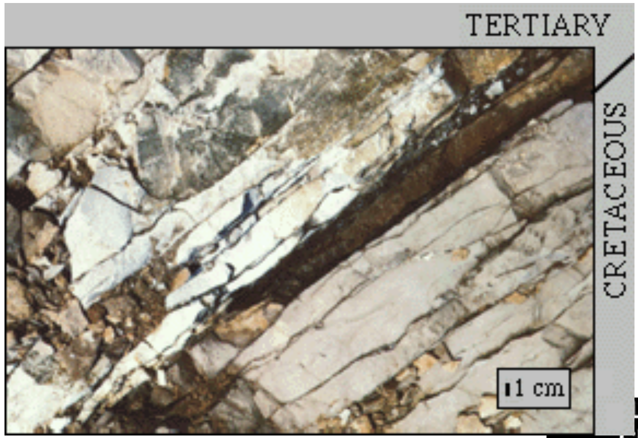


Marine Genus Biodiversity: Extinction Intensity

Causes are debated but there is rather good evidence that the K/T extinction 65 million years ago (dinosaur extinction) was associated with a major impact (10-20 km asteroid). The KT boundary has:

- Iridium anomaly (Alvarez et al 1980) --> leading to estimate of 10 km asteroid.
- Beads of “shocked quartz”, especially near the Caribbean region.
- Isotope anomalies in the Iridium and Chromium that are similar to those found in meteoritic material.
- The major impact crater Chicxulub (Yucatan) accurately dated to the KT event (200 km diameter)





Key Point: Large impacts do happen

There will be a broad spectrum of impacts:
Small ones happen often, large ones happen rarely.

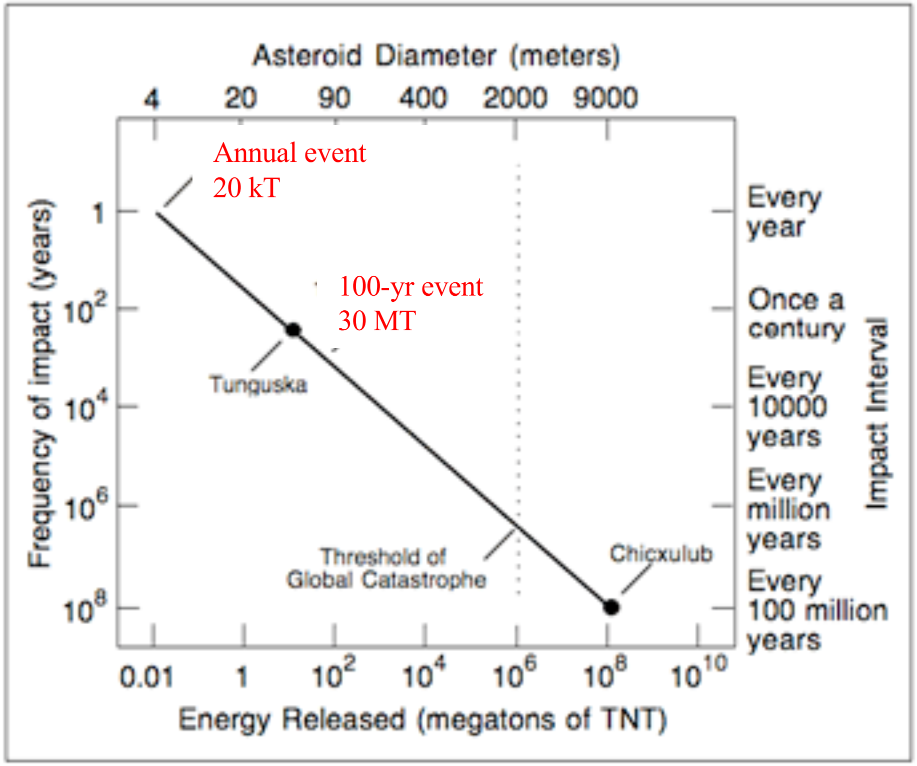
Empirical relation

$$\frac{\Delta t}{yr} \sim \left(\frac{E}{0.01\,\mathrm{MT}}\right)^{0.8}$$

e.g. 2013 Chelyabinsk meteorite (~20m diameter, 500kT deposited energy) was likely the largest impact since Tunguska in 1908 (estimated to be ~100m, with ~ 50 MT energy).

Resulting impact craters are produced by release of gas (vapourized rock) that is produced by the impactor tunneling through rock.

Impacts are expected to have global effect for diameters above about 1 km (10^6 MT), which occur of order once per million years.



Expected effects:

- Dust in atmosphere blocks out sunlight
- Tsunamis swamp coastal regions
- Blasts of hot gas
- Fall-out of hot rock

Note: A 10km object contains enough material to deposited a 2mm layer over entire surface of the Earth

Key Point: Large impacts do happen

$$\boxed{\frac{\Delta t}{yr} \sim \left(\frac{E}{0.01\ \mathrm{MT}}\right)^{0.8}}$$

Interesting question:

What is the worst impact that has been "witnessed":

- by "historical people" in the last 10,000 years?
  1000 MT
- by Homo in the last one million years
  300,000 MT

Compare these with the largest man-made explosion:

- Tsar Bomba RDS-220 hydrogen bomb
  which was 50 MT

5. Other chemical systems than C-N-H-O based Life

How strange could life be?

Terrestrial Life is based on organic molecules (long C-C-chains) with water as a solvent.

- Carbon's ability to form long hydrogenated chains: carbon with H up to $C_{90}H_{84}$, compared with oxygen (H_2O, H_2O_2) or nitrogen (NH_3, N_2H_2)

Could well imagine different "details" (e.g. other handedness, use of other amino acids, etc.).

But what about other elements and/or solvents?

- e.g. Silicon-based instead of Carbon-based?
- e.g. NH_3 or CH_3OH instead of water as solvent?

What about Silicon-based Life?

Si is the next valence-4 element below C in Periodic Table). BUT:

- Si-Si bond has half the strength of the C-C bond.
- Chain-ending Si-H and Si-O are actually stronger than Si-Si (c.f. C-O ~ C-H ~ C-C).
- Si_mH_n is unstable for $m > 3$.
- SiO polymers do exist (silicones) but are essentially completely inert.
- Si has extreme affinity for O: SiH_4 only exists at $T > 1000$ K
  - e.g. Jupiter has hydrogenated NH_3, PH_3, AsH_3, CH_4, GeH_4, H_2S, H_20 in atmosphere but no detectable SiH_4 (all Si is in the form of SiO_2)
- Si found with other elements as Silicates (e.g. many rock types on Earth!)

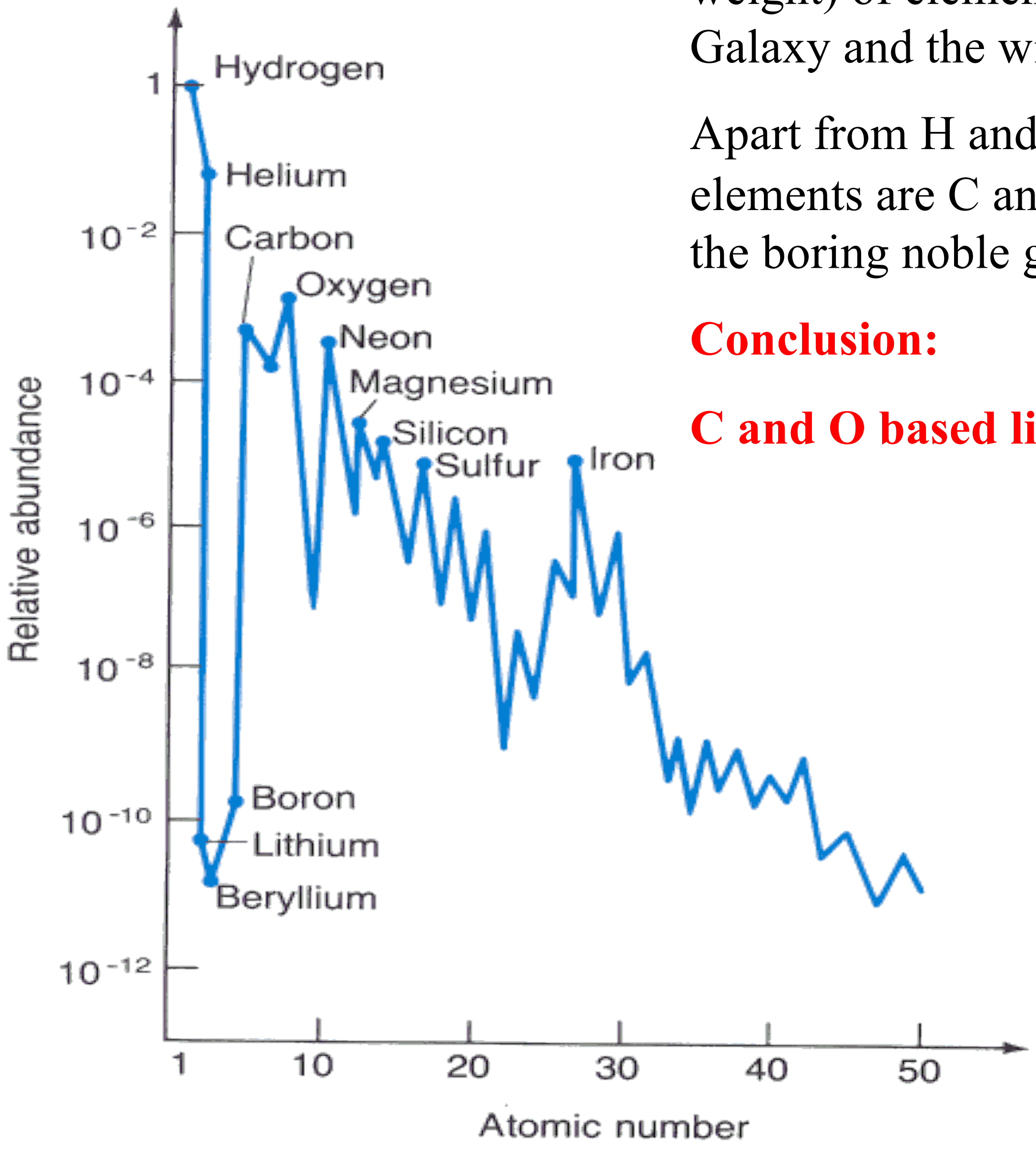
CO_2	SiO_2	CH_4	SiH_4
Gas at > 200K	Gas at > 2300K	Exists in oxidising atmosphere (e.g. 10 yr on Earth)	Bursts spontaneously into flame with oxygen
Soluble in H_2O	Insoluble (except in HFl)		
→ C + O in uv	Stable (quartz)		

What would be an ideal solvent?

- Solvent should be a liquid over a range of temperature (better than a gas for concentration reasons) which:
  - is high enough for chemical reactions to occur.
  - is low enough for survival of complex molecules.
  - is wide enough to give tolerance of environmental variations (e.g. on planets).
- Should have high heat capacity and latent heats to facilitate active temperature control.

	H_2O	NH_3	CH_3OH
Liquid	0-100 C	-78 – -33 C	-94 – +65 C
SHC	1	1.2	0.6
LHV	600	300	290
Surface tension	1	0.5	0.33
“Solvent”	1	0.5	0.5

Conclusion: H_20 is the best solvent, but NH_3 and CH_3OH are not bad



Look again at the relative abundances (by weight) of elements in the Sun (typical of the Galaxy and the wider Universe):

Apart from H and He, the <u>most abundant</u> elements are C and O with N next (excluding the boring noble gas Ne).

Conclusion:

C and O based life is natural and easy!

Key points

- Hierarchy of structure and function in terrestrial Life
- Complexity gives extreme efficiencies for energy conversion
- Organic molecules were naturally produced on and around young Earth
- There were multiple steps in the emergence of advanced Life on Earth
- Little can be said from the early emergence of Life on Earth
- There is interaction of Life with the extraterrestrial environment via impacts etc.
- C-life is extraordinarily much easier than Si-life!
- Other solvents may be possible, but why not H_2O, since O is so common?