# Origin and basic mechanism of life

Jan Helm
Technical University Berlin
Email: jan.helm@alumni.tu-berlin.de

**Abstract**

This paper presents in a concise way the main characteristics of life from the physical point of view and the most successful theories of biogenesis, together with a mathematical formulation and simulation of proto-biogenesis.

Part 1 is a state-of the-art report on the genetic foundations (genetic code, DNA, RNA, protein synthesis), the building blocks (amino acids, nucleobases), the structure and hierarchy of life (cells, tree of life) and the chemistry of life (metabolic pathways).

Part 2 describes the evolution of terrestrial life according to the current knowledge. For the first three stages (extra-terrestrial prebiotic synthesis, prebiotic evolution, proto-PNA-RNA world) plausible scenarios are presented, for the next two stages (LUCA, DNA world) the evolution is described according to our current knowledge.

Part 3 presents the mathematical formulation and simulation of the genetic proto-code model based on an empirical reaction model developed for life chemistry. Here the transition from the prebiotic chemistry to the first life cycle is calculated for two realistic scenarios: hydrothermal vent and lipid-bubbles.

**Contents**

# 1 Main characteristics of terrestrial life

Terrestrial life can be roughly **subdivided into 3 levels**, with growing complexity

**Molecular life**

*Quasi-species* (proposed by Manfred Eigen) are molecular assemblies of related macro-molecules which consume a constant energy flow (e.g. iron-sulfur-cycle), evolve into more complex and more adaptable systems with survival-evolution as the driving force, e.g. Otto's self-replicating aminoacid-nucleobase-compounds.
It is assumed that pre-biotic life functioned in this way, presumably contained in some sort of membranes, e.g. self-replicating Dworkin-bubbles

**Encoded-replication life**

*Species* of cellular (bacteria, archaea) and non-cellular (viruses) living beings have their protein-synthesis encoded in DNA (for retro-viruses in RNA). The RNA-controlled protein production generates proteins, which control life functions: food intake and energy production, reproduction.

**Multi-cellular life**

A multicellular organism consists of an assembly of genetically identical cells cooperating under central control.
The basic life functions are carried out in the cells, but there is a central chemical coordination (enzymes), a central food intake and energy-carrier production (normally sugars), a central reproduction apparatus, and some sort of a nervous system (sensory input, goal-oriented data processing with goal=individual survival, species survival, actor-output=movement and chemical output=enzymes).

**Central life functions**

energy conversion(catabolic reaction)
build-up of biomolecules with energy consumption (anabolic reaction)

catabolic reaction: e.g. reverse acetogenesis with sulphate reduction, catalyzed by AMP=adenine monophosphate


anabolic reaction: e.g. photosynthesis of sugars, catalyzed by NADP

**Biomolecules** are built in a hierarchical structure with increasing complexity

■ *precursors* (M.W.=30±15), such as CO2, H2O, NH3 obtained from the environment

■ *metabolic intermediates* (M.W. =150±100), such as acetates, keto acids, methanol, simple esters

■ *building block* biomolecules (M.W. =300±150) : mono-sugars(glucose), amino acids, fatty acids, isoprenes, nucleotides

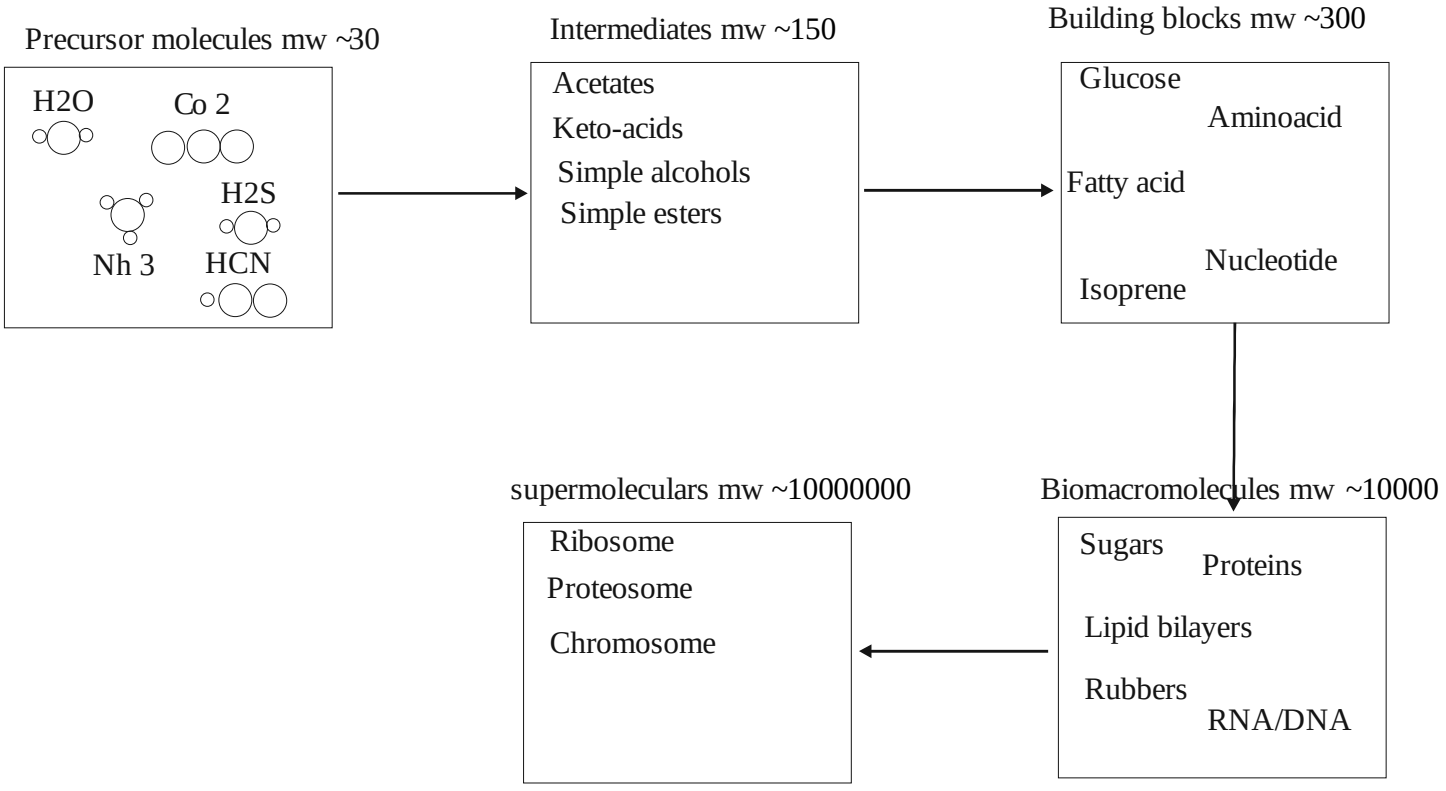
■ *biomacromolecules* (M.W. ~$10^4$ ), formed by linkage via covalent bonds from building blocks, they fall into *five distinct classes*: polysaccharides, lipid bilayers, terpenes(rubber), proteins, nucleic acids

■ *supramolecular structures* (particle weight $10^7$ ) such as proteosome, ribosomes, and chromosomes.

**Fundamental properties of biomolecules**

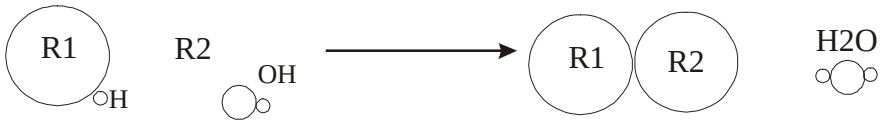
Energy and distance in biomolecule bonds

| chemical force | description | energy(kJ/mol) | distance(nm) |
|---|---|---|---|
| van der Waals | induced electronic | 0.4 - 4.0 | 0.2 |
| hydrogen bond | covalent bonded H | 12 - 38 | 0.15 - 0.30 |
| ionic bond | charge attraction | ~20 | 0.25 |
| hydrophobic | sticking together in water | ~25 | |

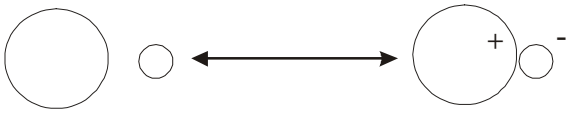


**Biochemical reactions** are enzyme-catalyzed and anabolic (building of biomolecules with energy consumption) and catabolic (energy production like photosynthesis, reverse acetogenesis)
They fall into one of six general categories:
-- oxidation and reduction,
-- functional group transfer,
-- hydrolysis,
-- reaction that forms or breaks carbon—carbon bond,
-- reaction that rearranges the bond structure around one or more carbons,
-- reaction in which two molecules condense with elimination of water.

Dehydration

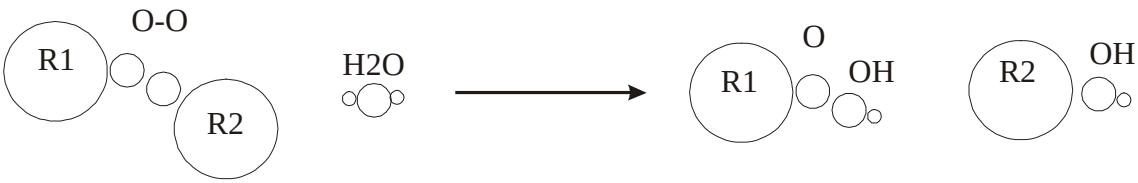



Oxidation/reduction

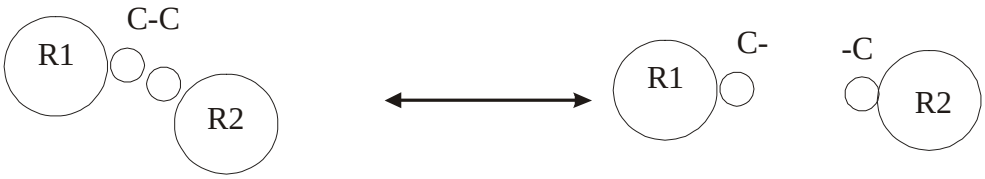



Group transfer

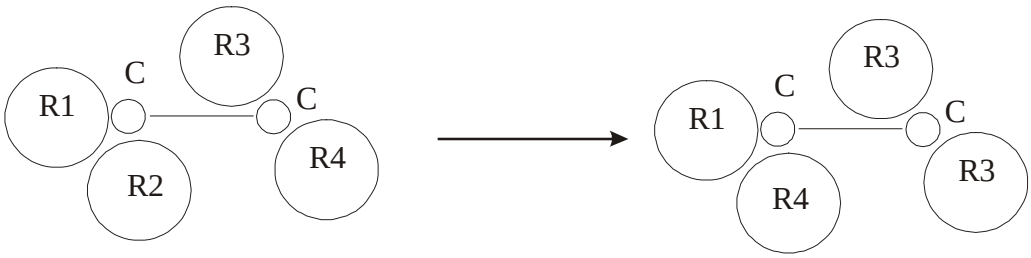

Hydrolysis



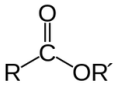

C-C bond

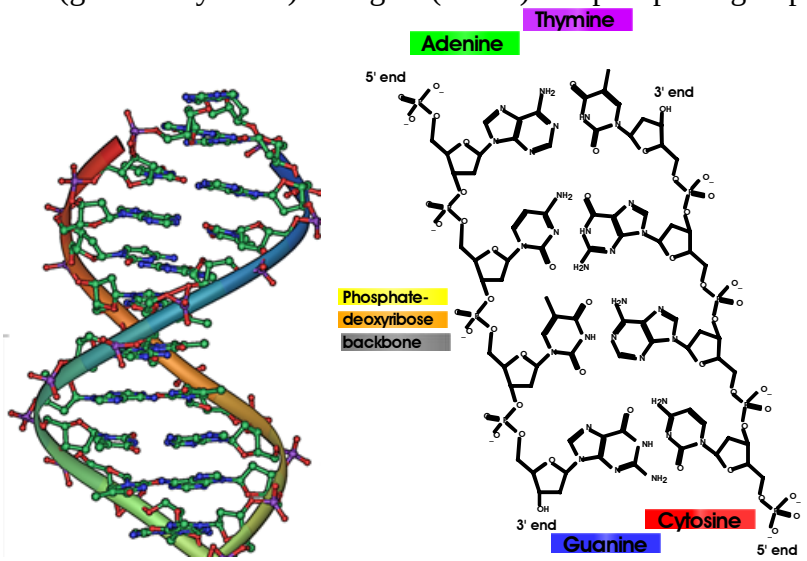


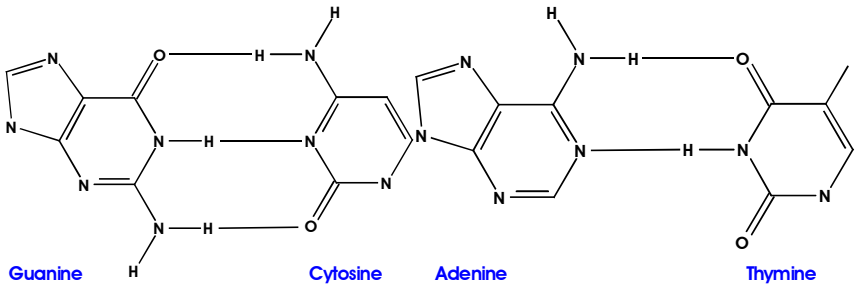

Carbon rearrangement



## 1.1 DNA

DNA(desoxyribonucleid acid) is a double helix of paired nucleotides A-T (adenine-thymine) and G-C(guanine-cytosine) of sugars (ribose) and phosphate groups joined by ester bonds .  [1]




The genetic code consists of three-letter 'words' called *codons* formed from a sequence of three nucleotides (e.g. ACT, CAG, TTT).
These encode the twenty standard aminoacids, giving most aminoacids more than one possible codon.
In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. This RNA copy is then decoded by a ribosome (eukaryote organelle) that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA, which carries aminoacids.
In DNA replication, a DNA-dependent DNA polymerase makes a DNA copy of a DNA sequence.

### Genetic code

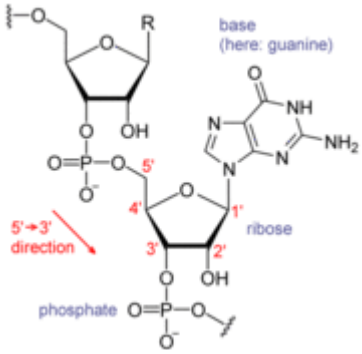
The genetic code is the set of rules used by living cells to translate information encoded within genetic material (DNA or mRNA sequences of nucleotide triplets, or codons) into proteins.
It is a 4x3-code: each of the 3 positions is occupied by one of the 4 nucleotides, so the code has $4^3$=64 possibilities.
The translation of an mRNA begins with the codon AUG, and a special tRNA is required to start translation.
The end of the protein-coding message (gene) is signaled by the presence of one of three stop codons (UAA, UAG, or UGA).

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

### 1.2 RNA

RNA (ribonucleid acid) is, like DNA, a polymer of nucleotides. Each nucleotide consists of a nucleobase , a ribose sugar, and a phosphate [1].
RNA is very similar to DNA, but differs in several important points:
■RNA is usually single stranded, while DNA is usually double stranded
■RNA nucleotides contain ribose while DNA contains desoxyribose (ribose lacking one oxygen), and in RNA uracil U substitutes for thymine T.
■The hydroxyl group makes RNA less stable than DNA: it is more prone to hydrolysis.

Each nucleotide in RNA contains a ribose sugar, with carbons numbered 1' through 5'. A nucleobase is attached to the 1' position: A, C, G, or U. A phosphate group is attached to the 3' position of one ribose and the 5' position of the next.

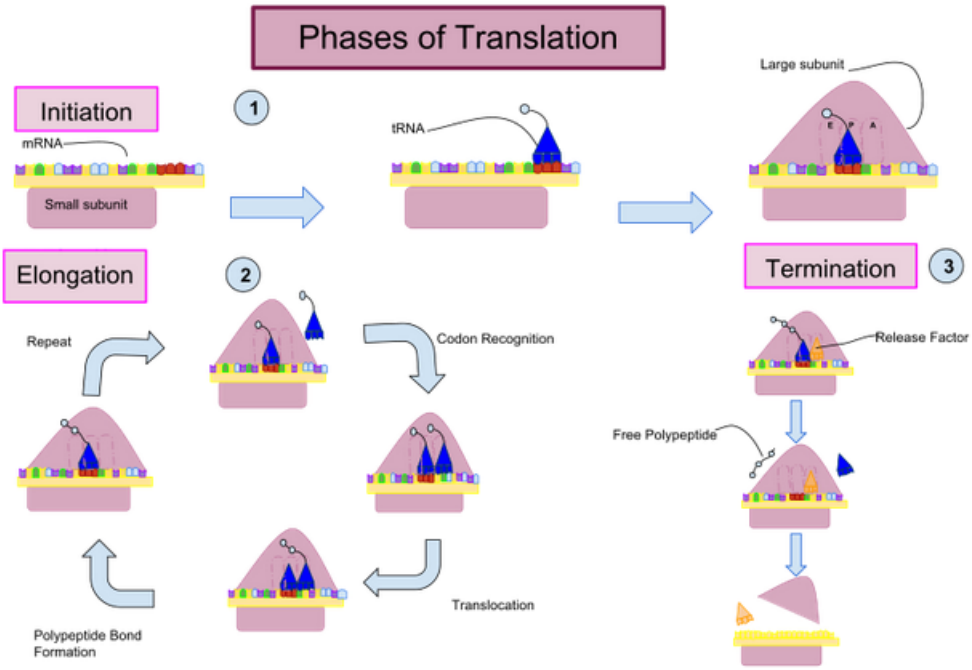



RNA is transcribed from DNA by enzymes RNA polymerases .
RNA is controls the translation of genes into proteins. A type of RNA called *messenger RNA* (mRNA) carries gene information from DNA to ribosomes. The ribosomes are complexes of proteins and *ribosomal RNA* (rRNA), which controls the protein synthesis. The protein synthesis is carried-out by *transfer RNA* (tRNA), which transfers a specific aminoacid to a growing protein chain.
*Micro RNA* (miRNA) are found in eukaryotes and act through RNA interference: they block or activate gene translation for specific genes.

### 1.3 Gene translation

Translation is the process in which ribosomes in the cytoplasm or endoplasmic reticulum synthesize proteins after the process of transcription of DNA to RNA .

8
Translation proceeds in three phases:
Initiation: The ribosome assembles around the target mRNA. The first tRNA is attached at the start codon.
Elongation: The tRNA transfers an aminoacid to the tRNA corresponding to the next codon. The ribosome then moves (*translocates)* to the next mRNA codon to continue the process, creating an aminoacid chain.
Termination: When a peptidyl tRNA encounters a stop codon, then the ribosome folds the polypeptide into its final structure.




[1]

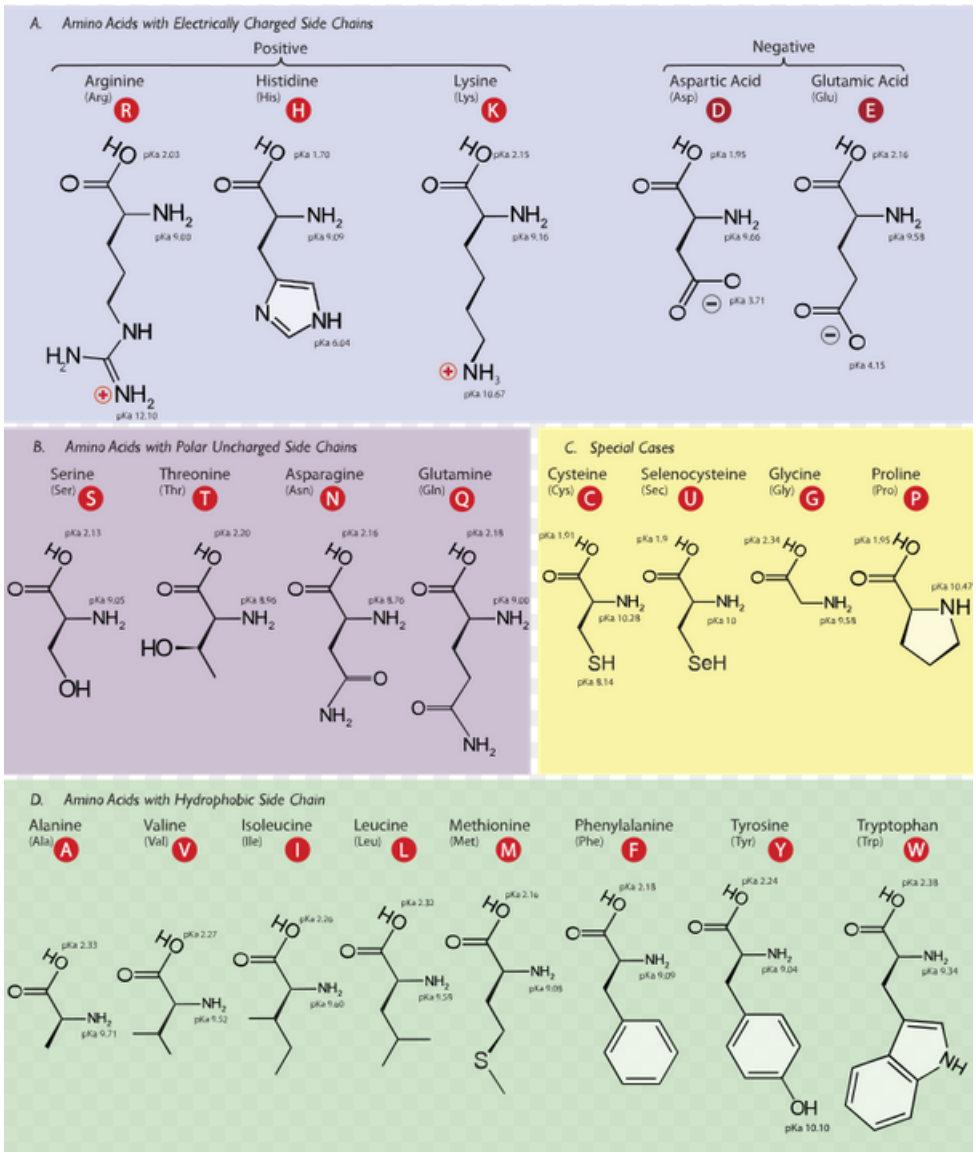
The actual coupling between tRNA and its corresponding aminoacid is carried out by the corresponding enzyme aaRS (aminoacyl-tRNA synthetase or tRNA-ligase).

## 1.4 Amino acids and nucleotides: molecular fundament of life

### 22 aminoacids

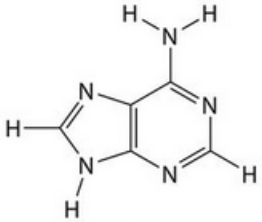
20 proteinogenic aminoacids are used in eukaryotes and coded in DNA/RNA
essential aminoacids are not synthesized in humans and must be supplied in food
2 non-essential aminoacids are used in archaea and bacteria only and coded by the stop-codons UAG and UGA

| Essential | Conditionally essential | Non-essential |
|---|---|---|
| Histidine (H) | Arginine (R) | Selenocysteine (U) |
| Isoleucine (I) | Cysteine (C) | Pyrrolysine* (O) |
| Leucine (L) | Glutamine (Q) | |
| Lysine (K) | Glycine (G) | |
| Methionine (M) | Proline (P) | |
| Phenylalanine (F) | Tyrosine (Y) | |
| Threonine (T) | Alanine (A) | |
| Tryptophan (W) | Aspartic acid (D) | |
| Valine (V) | Asparagine (N) | |
| | Glutamic acid (E) | |
| | Serine (S) | |

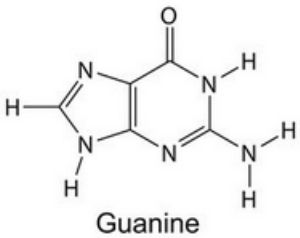



### 5 nucleotides: 2 purines A G , 3 pyrimidines U C T

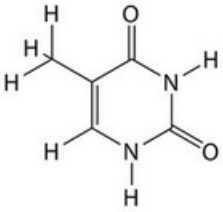
10



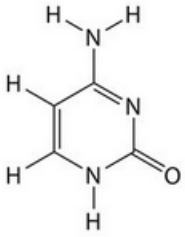
Adenine



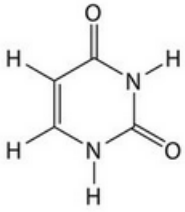
Guanine



Thymine



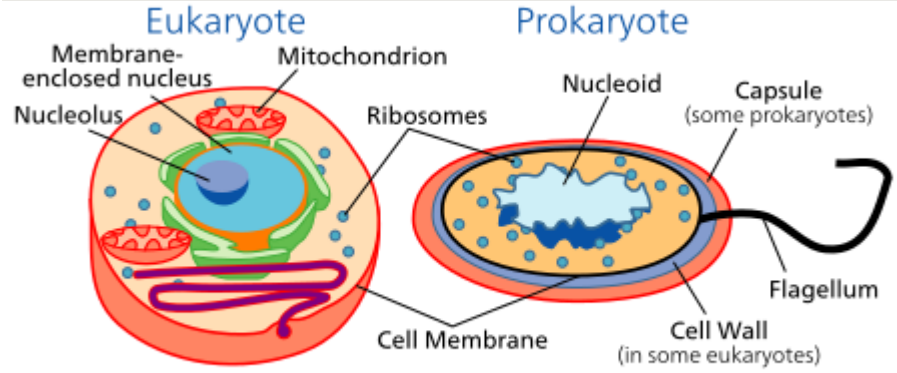
Cytosine



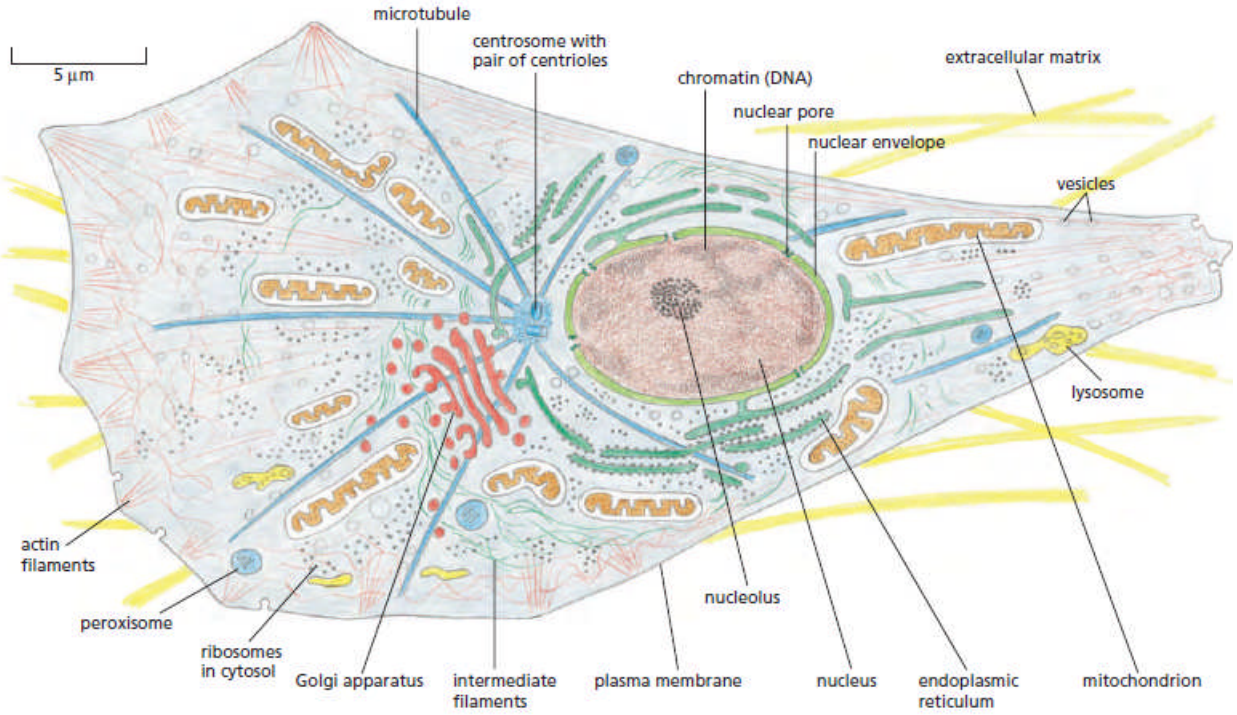
Uracil

## 1.5 Cells

The cell  is the basic structural, functional, and biological unit of all living (terrestrial) organisms, except virusses. Cells are the smallest unit of life that can replicate independently.



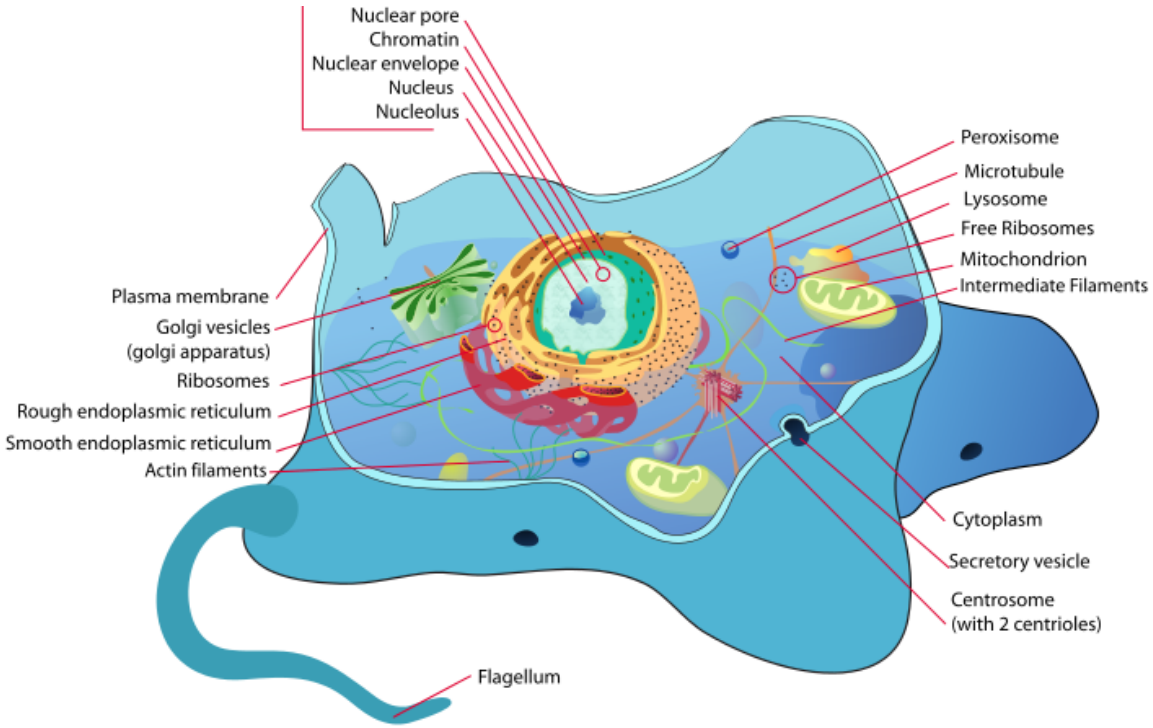

[2]




**Major features of eukaryotic cells** [3]

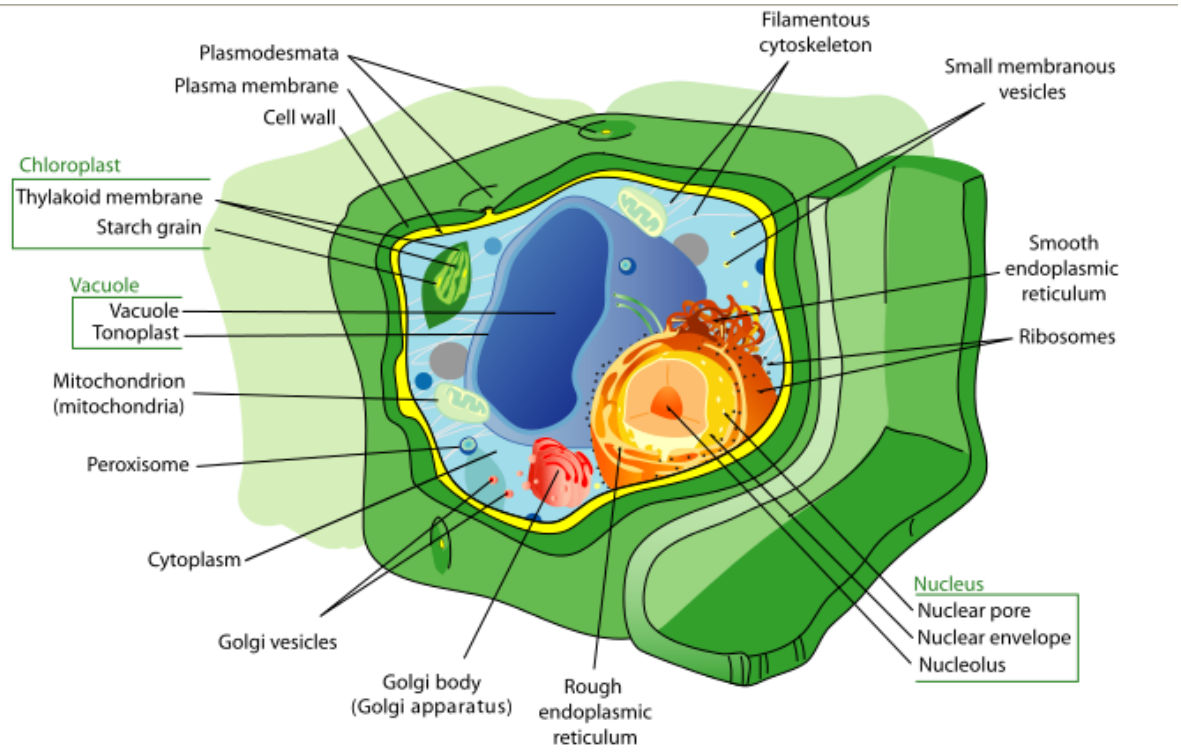
Eukaryotic cells contain a nucleus enclosed within membranes and membrane-bound organelles such as mitochondria and the Golgi apparatus, and in addition, some cells of plants and algae contain chloroplasts (). Unlike unicellular archaea and bacteria, eukaryotes may also be multicellular and include organisms consisting of many cell types forming different kinds of tissue. Animals and plants are the most familiar eukaryotes. Eukaryotes can reproduce both asexually through mitosis and sexually through meiosis and gamete fusion. In mitosis, one cell divides to produce two genetically identical cells. In meiosis, DNA replication is followed by two rounds of cell division to produce four haploid daughter cells. These act as sex cells (gametes). Each gamete has just one set of chromosomes, each a unique mix of the corresponding pair of parental chromosomes resulting from genetic recombination during meiosis.

12

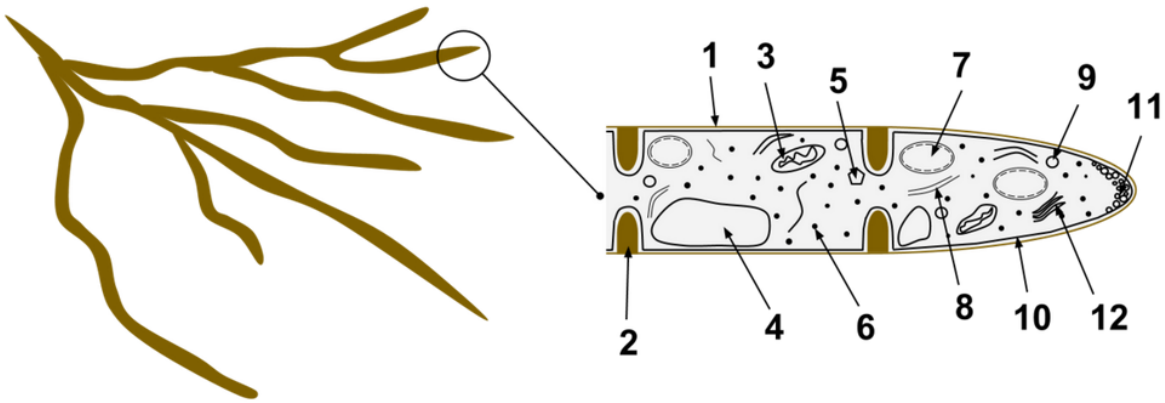



[2]

Scheme of an animal cell




Scheme of a plant cell



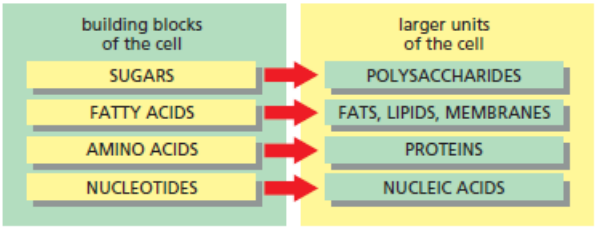

Scheme of a fungal  hyphae cell: 1 – hyphal wall, 2 – septum, 3 – mitochondrion, 4 – vacuole, 5 – ergosterol crystal, 6 – ribosome, 7 – nucleus, 8 – endoplasmic reticulum, 9 – lipid body, 10 – plasma membrane, 11 – spitzenkörper, 12 – Golgi apparatus

All cells share common features:

■ the genetic code is stored in the DNA, a codon (3 nucleobases) codes one aminoacid, a gene codes (basically) one protein

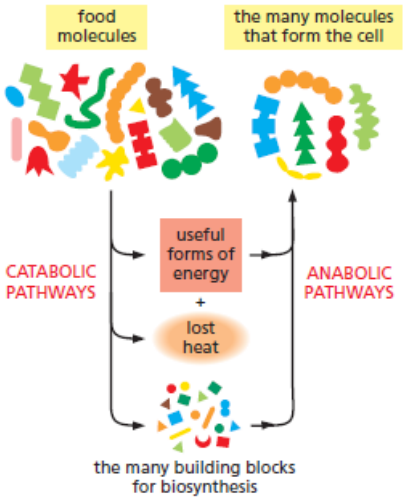
■ cells replicate their genetic information: the DNA is copied by *DNA-polymerase*

■ cells use proteins (enzymes) as catalysts of chemical reactions to create their main constituents: membranes, proteins, nucleid acids




(A) lysozyme

■ cells use energy sources (food) to generate energy (*catabolism*) in order to create proteins, to move, and to replicate (*anabolism*)




■ cells translate genes from the DNA into proteins using different RNA's

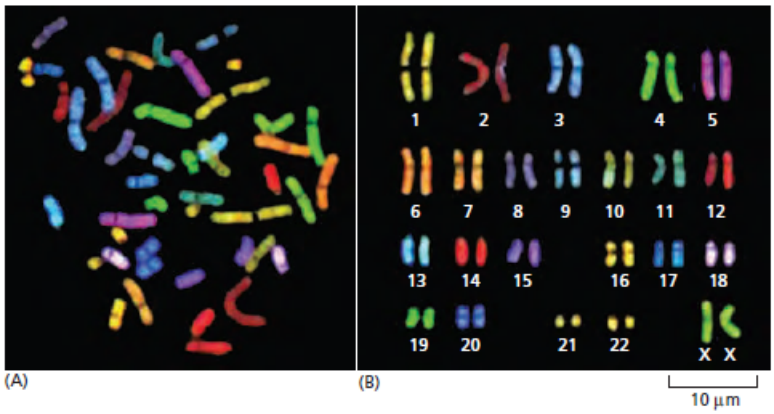
■ Genes are located on chromosomes

The number of genes is minimum 500 and up to 40000
virusses have 4-80 genes
bacteria and archea have 1000-6000 genes
eukaryotes 6000-40000
humans 21000

| | Organism | # of protein-coding genes | # of genes naïve estimate: (genome size /1000) | BNID |
|---|---|---|---|---|
| viruses | HIV 1 | 9 | 10 | 105769 |
| | *Influenza A* virus | 10-11 | 14 | 105767 |
| | Bacteriophage λ | 66 | 49 | 105770 |
| | Epstein Barr virus | 80 | 170 | 103246 |
| prokaryotes | *Buchnera sp.* | 610 | 640 | 105757 |
| | *T. maritima* | 1,900 | 1,900 | 105766 |
| | *S. aureus* | 2,700 | 2,900 | 105500 |
| | *V. cholerae* | 3,900 | 4,000 | 105760 |
| | *B. subtilis* | 4,400 | 4,200 | 111448 |
| | *E. coli* | 4,300 | 4,600 | 105443 |
| eukaryotes | *S. cerevisiae* | 6,600 | 12,000 | 105444 |
| | *C. elegans* | 20,000 | 100,000 | 101364 |
| | *A. thaliana* | 27,000 | 140,000 | 111380 |
| | *D. melanogaster* | 14,000 | 140,000 | 111379 |
| | *F. rubripes* | 19,000 | 400,000 | 111375 |
| | *Z. mays* | 33,000 | 2,300,000 | 110565 |
| | *M. musculus* | 20,000 | 2,800,000 | 100308 |
| | *H. sapiens* | 21,000 | 3,200,000 | 100399, 111378 |
| | *T. aestivum* (hexaploid) | 95,000 | 16,800,000 | 105448, 102713 |

The number of chromosomes varies from 1 circular (prokaryotes) to several dozens (eukaryotes)

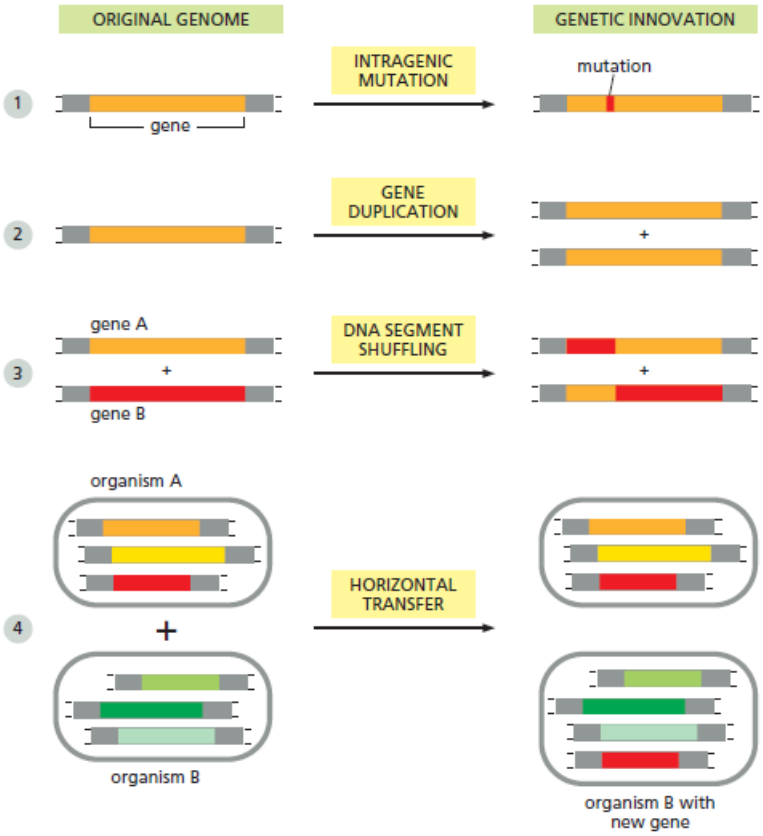
| Species | Large Chromosomes | Intermediate Chromosomes | Microchromosomes |
|---|---|---|---|
| *Trypanosoma brucei* | 11 | 6 | ≈100 |
| Ciliated protozoa | 10 | | |
| Domestic pigeon | 18 | – | 59–63 |
| Chicken[61] | 8 | 2 sex chr. | 60 |
| Rye (diploid) | 14 | | |
| Fruit fly | 8 | | |
| Humans | 23 | 2 sex chr. | |




The 23 human chromosomes

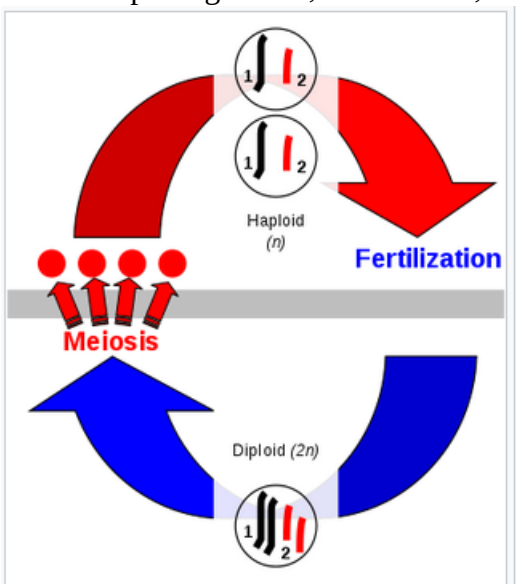
■ genes mutate and form new genes (genetic innovation)
Genetic innovation

16
1. *Intragenic mutation*: an existing gene can be randomly modified by changes in its DNA sequence
2. *Gene duplication*: an existing gene can be accidentally duplicated so as to create a pair of initially identical genes within a single cell; these two genes may then diverge in the course of evolution.
3. *DNA segment shuffling*: two or more existing genes can break and rejoin to make a hybrid gene consisting of DNA segments that originally belonged to separate genes.
4. *Horizontal (intercellular) transfer*: a piece of DNA can be transferred from the genome of one cell to that of another—even to that of another species.




[3]

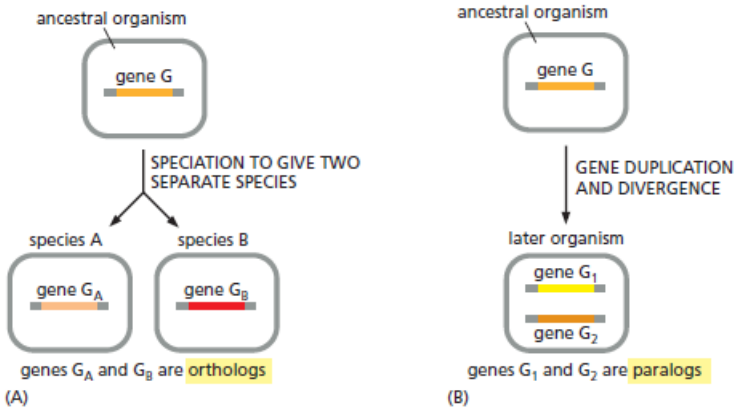
■ cells reproduce sexually or asexually
sexual: haploid gametes, fertilization, meiosis




asexual: fission, budding(prokaryotes), vegetative(plants)

■ cells evolve through gene evolution within a species and group into new species

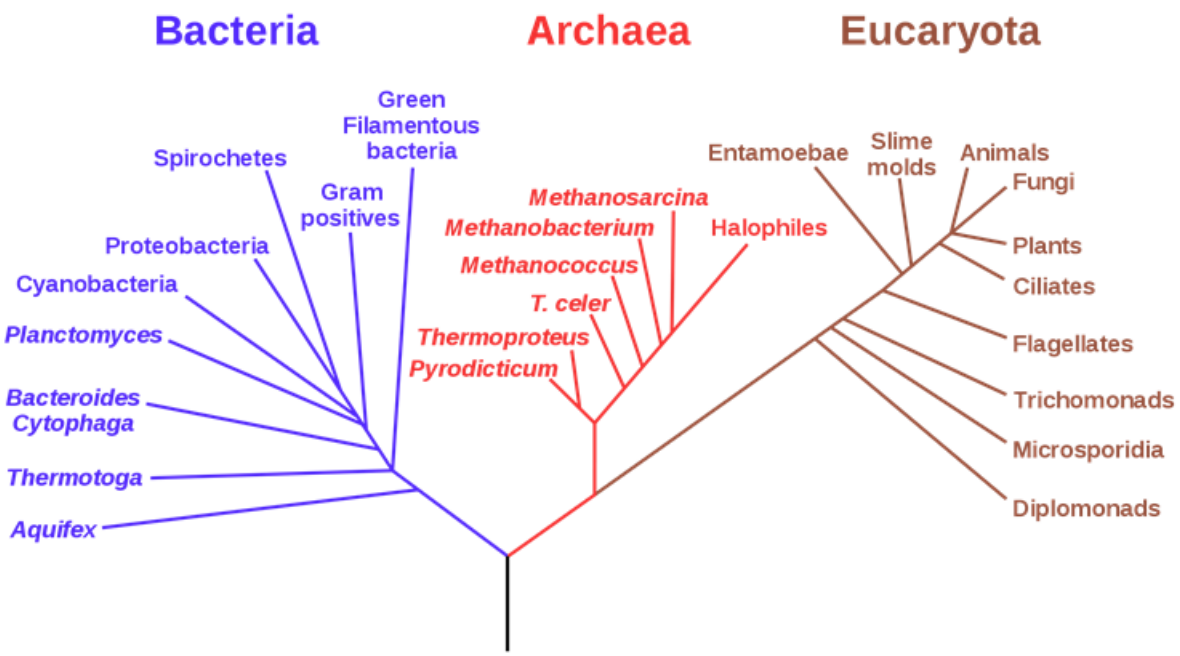
17


ancestral organism
gene G
SPECIATION TO GIVE TWO SEPARATE SPECIES
species A
species B
gene $G_A$
gene $G_B$
genes $G_A$ and $G_B$ are orthologs
(A)
ancestral organism
gene G
GENE DUPLICATION AND DIVERGENCE
later organism
gene $G_1$
gene $G_2$
genes $G_1$ and $G_2$ are paralogs
(B)

[3]

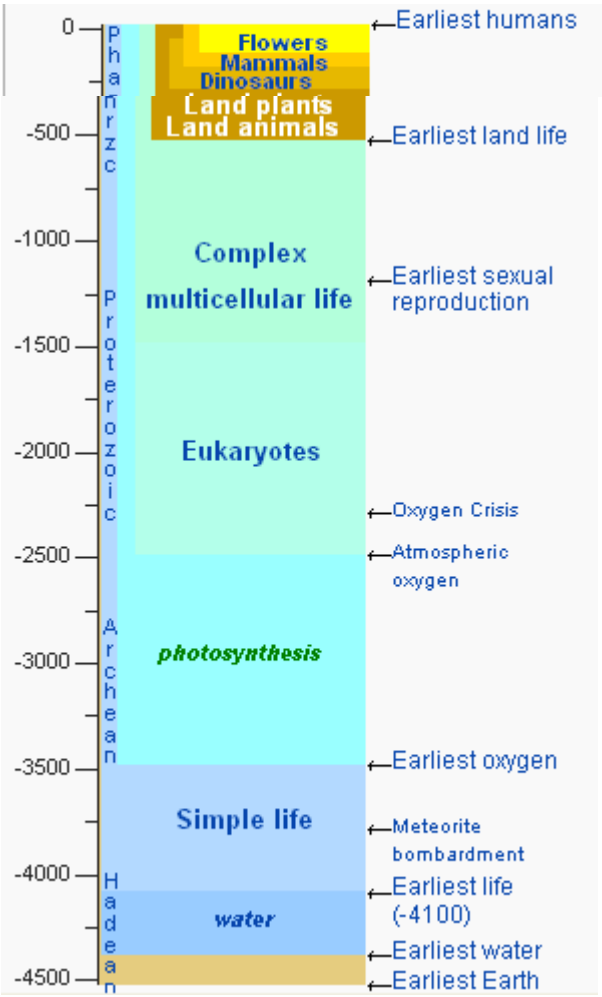
### 1.6 Tree of life

The non-viral organisms separate into bacteria, archaea , eukarytes




Phylogenetic tree of life (based on rDNA) [4]

### Timeline of life




| |
|---|
| CT extinction -65 |
| Permian extinction -250 |
| |
| first land plants & animals -450 |
| vertebrates -525 animals -580 |
| |
| |
| |
| |
| algae on land -1200 |
| |
| |
| |
| |
| multicellulars -1700 |
| eukaryotes -1850 |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| |
| earliest oxygen -3500 methane sulfur cyano bacteria -3500 |
| |
| |
| late heavy bombardment -4000 ...-3800 |
| bacteria & archaea -3800 |
| LUCA -4000 |
| RNA proto-ribosome -4250 |
| crust and oceans form -4400 |
| Moon forms from impact -4500 |

Crucial events: [4,5,6]
planet Earth formed -4540 My
Moon formed from impact -4500My

oceans and crust formed -4400 My
RNA , proto-ribosome .-4250 My (biogenic 13C/12C ratio in zircon)
LUCA -4000 My
bacteria & archaea -3800 My
methane- sulfur- cyano-bacteria -3500 My
eukaryotes -1850 My
multicellulars -1700 My
animals -580My
vertebrates -525 My
land plants & animals -450 My

**Research methods**

1 *Archaeo-biology*: evaluation of fossils, biochemical and isotope analysis of micro-fossils, especially of zirconites, also analysis of extraterrestrial materials (meterorites)
2 *Chemical simulation of prebiotic processes*: prebiotic soups with different composition and physical conditions
3 *Analysis of metabolic cycles* in the living world with the goal of finding the most ancient and basic energy supply mechanisms
4 *DNA and RNA analysis*: determination of evolution relationships, the timeline can be calculated via known mutation rates, determination of elementary genotypes (e.g. LUCA)
5 *Chemical simulation of protein synthesis* with different RNA variants, also with non-biological RNA and amino acids

### 1.7 Metabolic pathways of life

The pathways described here are of 3 types: prebiotic (pre-life generation of macro-molecules like aminoacids, nucleotides, sugars), anabolic (production of macro-molecules from building blocks with energy input) and catabolic (breakdown of nutrient molecules with energy production).
Catabolic pathways involve the breakdown of nutrient molecules or harmful molecules into usable forms (building blocks). In this process, energy is either stored in energy molecules for later use, or released as heat. Anabolic pathways build new molecules, and these pathways typically use energy. The new molecules built via anabolic pathways (macro-molecules) are useful for building cell structures and maintaining the cell.

**Anabolic, catabolic pathways**
**Wood-Ljungdahl** (aceto-genesis)

$$H^+ + CO_2 \rightarrow formate \rightarrow \begin{matrix} CH_3 protein \\ CO \rightarrow \end{matrix} \rightarrow acetylCoA \rightarrow acetate + ATP$$

This ana-catabolic reaction reduces carbon dioxide and H-ions or hydrogen in boiling-hot water to acetyl-coenzymeA (*acetyl-CoA*) and then to acetate and ATP (adenosine tri-phosphate, an energy-carriers in life-chemistry). It runs in hydrothermal vents and in volcanic pools and caverns. It was the energy-cycle used by LUCA (Last Common Ancestor). [9,7]

**Citric acid cycle**
This catabolic cycle consumes acetate (in the form of acetyl-CoA) and water, reduces $NAD^+$ to NADH, and produces carbon dioxide, energy is transferred via NADH (nicotinamide adenine dinucleotide, an energy carrier) to oxidative phosphorylation (electron transport) pathway. [9]

$$acetyl - CoA + H_2O + NAD^+ \rightarrow NADH + CO_2$$

The citric acid cycle together with the oxidative phosphorylation is the main energy cycle running in the eukaryotic mitochondria.

**Glycolysis**
One of the primary sources of *acetyl-CoA* is from the breakdown of sugars by (catabolic) glycolysis which yield pyruvate (CH3COCOOH) that in turn is decarboxylated by the pyruvate dehydrogenase complex generating acetyl-CoA according to the following reaction scheme:

$$pyruvate^- + HS\,CoA + NAD^+ \rightarrow acetylS\,CoA + NADH + CO_2$$

**Oxidative phosphorylation**
This catabolic reaction is summarized as
$1/2\ O_2 + NADH + H^+ \rightarrow H_2O + NAD^+ + 1.14eV$
It is the main way of energy production in prokaryotes and in eukaryotic mitochondria.

**Reverse acetogenesis with sulfate reduction**

$$4H_2 + SO_4^{2-} + H^+ + AMP \rightarrow HS^- + 4H_2O + ATP + 2P$$

This catabolic reaction reduces sulfur dioxyde to hydrogen sulfide and transfers energy via AMP/ATP cycle [9].
It is used by sulfur bacteria in hydrothermal vents.

**Photosynthesis**
This important anabolic reaction using photonic energy is the energy source in plants and cyanobacteria.
The general equation for photosynthesis is

$$CO_2 + 2H_2A + photons \rightarrow [CH_2O] + 2A + H_2O$$

carbon dioxide + electron donor + light energy → carbohydrate + oxidized electron donor + water
In green plants and cyanobacteria the energy-generating part of the reaction is [3]

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

**Prebiotic pathways**
**Formose reaction**
The prebiotic formose reaction produces higher sugars (ribose, pentose) from formalehyde catalyzed by calcium ions in an alkaline medium. [9]

$$formaldehyde \rightarrow \overset{\rightarrow}{Ca(OH)_2} \rightarrow ribose$$

It is a prebiotic reaction involved in RNA synthesis.

**Pyrite reaction**

The pyrite (*FeS*) reaction yields energy and free electrons

$$FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^- \ +41\text{kJ/mol}$$

and has been demonstrated to run under volcanic pool and hydrothermal vent conditions.

It is a prebiotic purely anorganic reaction, which has been proposed as the primeval energy source in the iron-sulfur-world hypothesis of Wächtershäuser.

**Methanogenesis from carbon dioxide**

The methane synthesis from carbon dioxide by hydrogenation

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \ +165\text{kJ/mol}$$

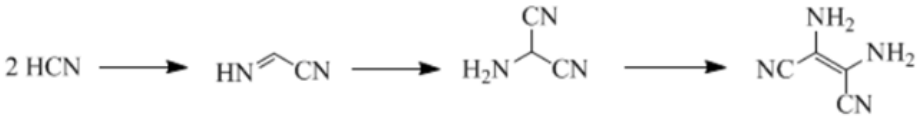
(Sabatier reaction) is used at high pressure, 400°C and with nickel catalyzer in the industry for methane production.

Archeans in the “deep biosphere” (subterranean rocks, sub-marine crust) use this reaction as energy source [54].

**Prebiotic nucleotide synthesis**

**Synthesis of purines** (A,G): [9,15,16]

From cyan HCN diaminomaleodinitrile (DAMD) is formed:




From this adenine and guanine is created

$$DAMD + UVlight + HCN \rightarrow adenine$$

$$DAMD + UVlight + HCN + H_2O \rightarrow guanine$$

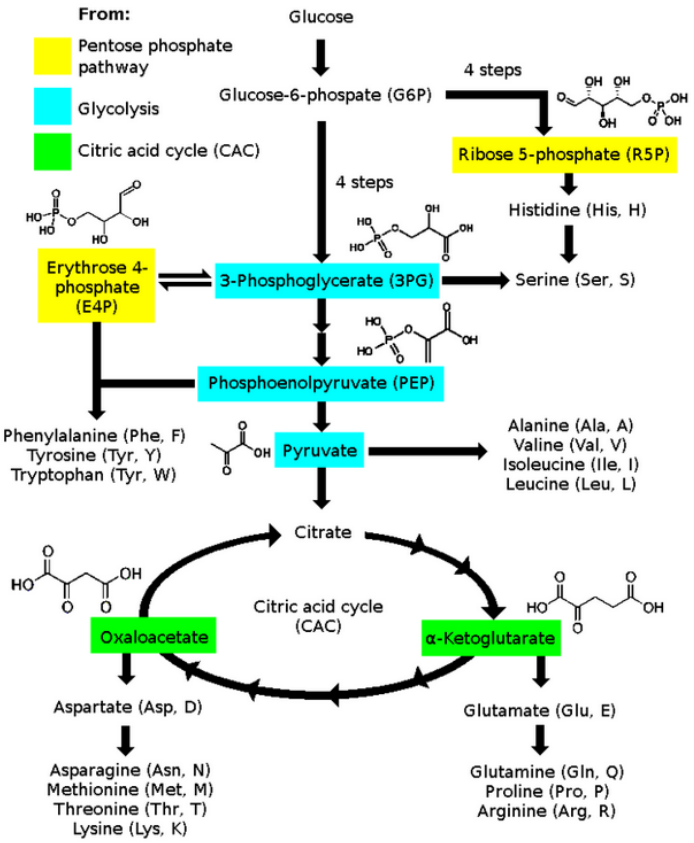
**Synthesis of pyrimidines** (U,C) [14,15]

The pyrimidines are formed from simple prebiotic cyan- and glycol-carrying molecules and phosphate, which are plausible prebiotic feedstock molecules, and the conditions of the synthesis are consistent with potential early-Earth geochemical models

$$cyanamide + cyanacetylene + glycolaldehyde + glyceralaldehyde + phosphate \rightarrow uracil + cytosine$$

**Prebiotic aminoacid synthesis** [17,24]

The 20 basic aminoacids can be formed from simple organic compounds under prebiotic conditions.


From:
Pentose phosphate pathway
Glycolysis
Citric acid cycle (CAC)
Glucose
Glucose-6-phospate (G6P)
4 steps
Ribose 5-phosphate (R5P)
Histidine (His, H)
4 steps
Erythrose 4-phosphate (E4P)
3-Phosphoglycerate (3PG)
Serine (Ser, S)
Phosphoenolpyruvate (PEP)
Phenylalanine (Phe, F)
Tyrosine (Tyr, Y)
Tryptophan (Tyr, W)
Pyruvate
Alanine (Ala, A)
Valine (Val, V)
Isoleucine (Ile, I)
Leucine (Leu, L)
Citrate
Citric acid cycle (CAC)
Oxaloacetate
α-Ketoglutarate
Aspartate (Asp, D)
Glutamate (Glu, E)
Asparagine (Asn, N)
Methionine (Met, M)
Threonine (Thr, T)
Lysine (Lys, K)
Glutamine (Gln, Q)
Proline (Pro, P)
Arginine (Arg, R)

# 2 Biogenesis

## 2.1 Fundamental life functions

We can summarize our knowledge about extraterrestrial organic chemistry and prebiotic and early life biosynthesis in the following scheme of fundamental life entities.

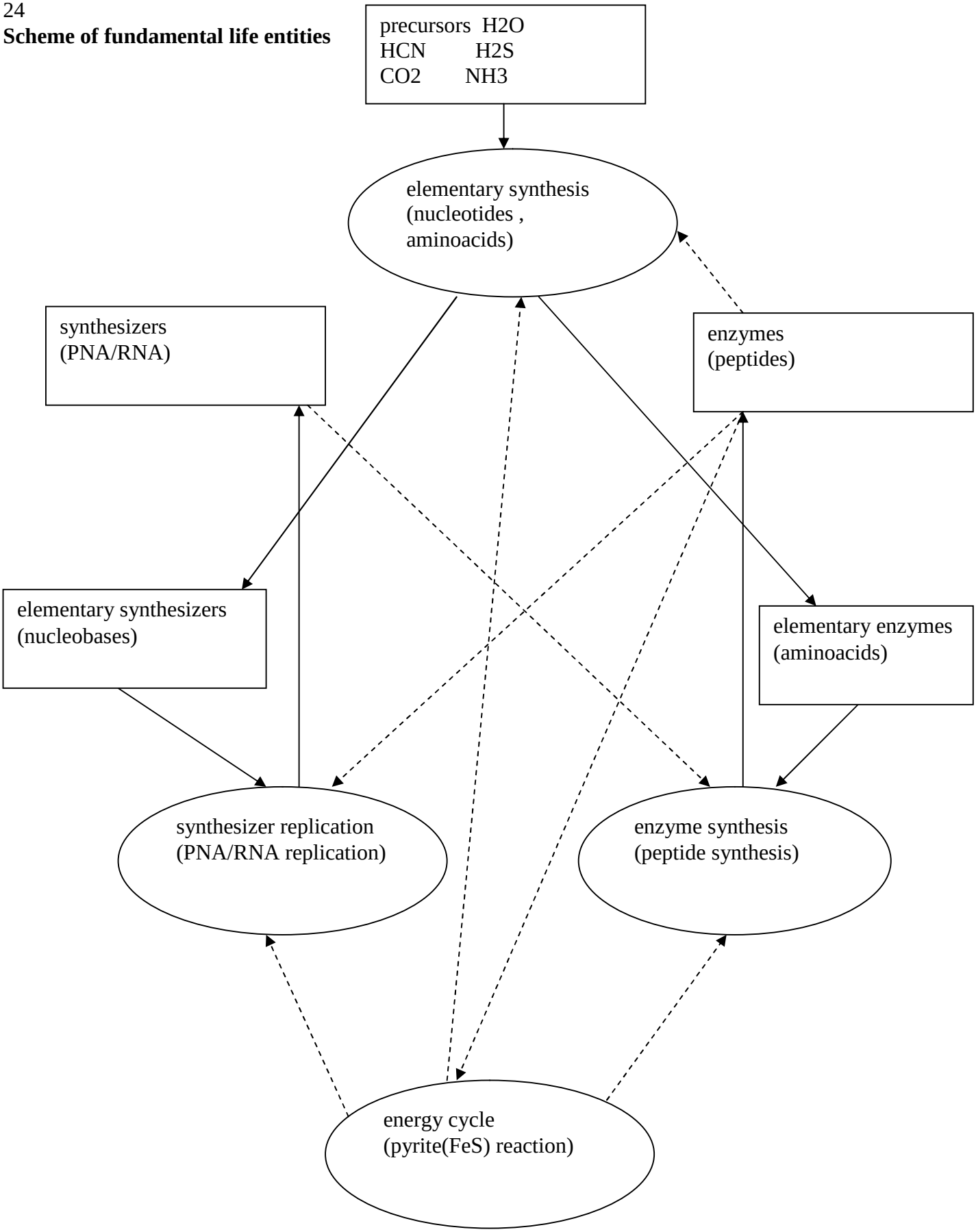
**1 Energy cycle**
Energy cycle, which generates a continuous energy flow, enabling stable biochemical systems with Darwinian evolution, competing for survival.
Terrestrial realization: pyrite reaction, later sulfate reduction and photosynthesis.

**2 Enzymes**
A set of potentially self-replicating **polymer enzymes** (terrestrial: **peptides**) built of elementary units (terrestrial: aminoacids), which catalyze the energy cycle, the synthesis of other life-molecules, and replication.
Life function: catalysis of **biosynthesis of non-enzymes** and **replication** (terrestrial: replication of PNA/RNA)

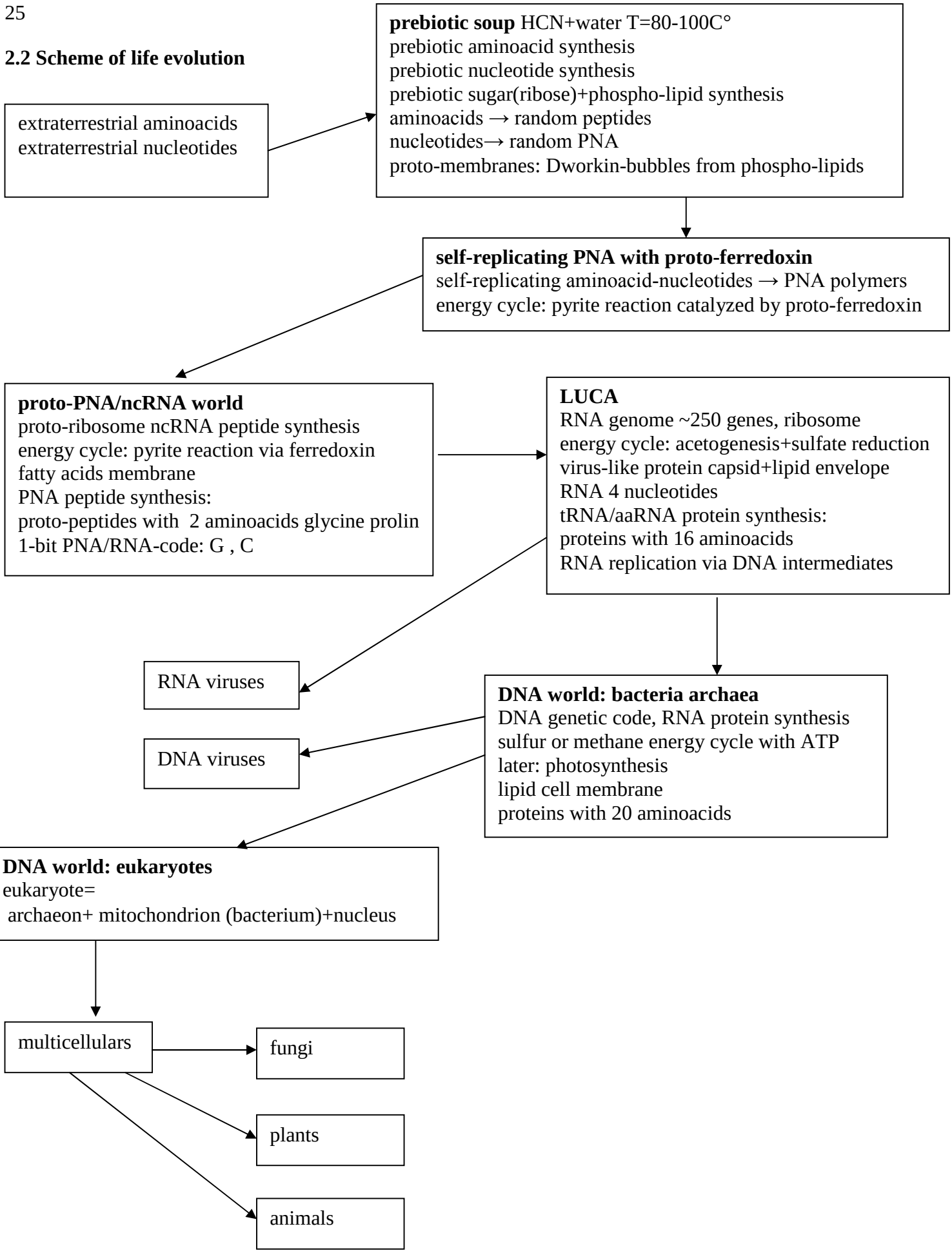
**3 Synthesizers**
A set of **synthesizer polymers** (terrestrial: **PNA/RNA**), built of elementary units (terrestrial: nucleotides), which synthesize the enzyme, based on (encoded by) the synthesizer structure (terrestrial: PNA/RNA genetic code).
Life function: **biosynthesis of enzymes** (terrestrial peptide synthesis)

**4 Mediator-enzymes**
One or several **mediator-enzymes** (terrestrial: proto-ferredoxins), capable of initial catalysis of the energy cycle and initial starting of synthesizer polymer replication (terrestrial: mediate **self-replicating aminoacid-nucleotide polymers**).

**Scheme of fundamental life entities**



**2.2 Scheme of life evolution**



### 2.3 Extra-terrestrial prebiotic synthesis

5 natural *aminoacids* have been found in meteorites: glycine, alanine and glutamic acid as well as unusual ones like isovaline and pseudoleucine, probably *all 20 natural aminoacids* were brought to the early Earth by asteroids and chondritic meteorites, they are *mostly left-handed*. [25]
As for *nucleotides*, all of the purines (adenine, guanine, hypoxanthine, and xanthine) and the one pyrimidine (uracil) are reported in meteorites [26].

### 2.4 Prebiotic evolution

**Miller-Urey experiment** 1953 created the first **prebiotic soup** , modified 2011: abiotic creation of amino-acids [13,24]

conditions: $\xrightarrow[100°C,\,discharge]{}$

$$H_2O + CH_4 + NH_3 + H_2 \rightarrow HCN + formaldehyde(CH_2O) + acetylene(C_2H_2) + cyanoacetylene(C_3HN)$$

$$\rightarrow 22 aminoacid \xrightarrow[formose]{} ribose + pentose$$

Extraterrestrial conditions: electric discharge replaced by UV

**Oro, Sutherland experiments**: prebiotic nucleotides synthesis
Oro [14] . Sutherland [16] and Ritson [52] showed in experiments that hydrogen cyanide (HCN) and cyanoacetylene(CHCNC) are the two major precursors for the prebiotic synthesis of purines and pyrimidines under UV irradiation.

**Life from HCN: Patel's synthesis of ribonucleotides, lipids and aminoacids**
Patel et al. (2015) [35] demonstrated that the building blocks of life, ribonucleotides (via precursors glycolaldehyde and glyceraldehyde), lipid-precursors (glycerol-diphosphate) and aminoacids, can be formed from $CH_4$, $NH_3$, $H_2S$ in water under UV irradiation in a "one pot" synthesis. HCN (hydrogen cyanide) and cyanoacetylene is produced from methane $CH_4$ and nitrogen $N_2$ under prebiotic conditions from spark discharge [41].

$$HCN + H_2S + cyanamide + cyanoacetylene + phosphate + H_2O \xrightarrow[UV,100°C]{}$$

$$aminoacids + ribonucleotides + phospho-lipids$$

input materials: hydrogen cyanide (HCN), hydrogen sulfide ($H_2S$), cyanamide ($CNNH_2$ , can form from hydrogen cyanide and ammonia), cyanacetylene (CHCNC, can form from acetylene and ammonia). phosphate, water.
scenarios: hydrothermal vents, volcanic pools under UV-radiation, volcanic pools with freezing/heating cycles

**Life from HCN: aminoacids, sugars and nucleotide precursors form from HCN in water**
Das et al.(2019) [41] carried out Hartree-Fock chemical simulations and experiments in a nanoreactor with HCN and water at T=80-100°C . They showed that under these condition without UV important life molecule precursors arise: the simplest amino acid glycine, and the RNA precursors the sugar glycoaldehyde (COH-CH2-COOH), cyanamide (a precursor in the Patel synthesis) and cyan-amine-oxazole (purine precursor, penta-ring CH-N-CNH2-CNC-O).
input materials: hydrogen cyanide (HCN), water T=80-100°C, possibly with pyrite reaction as energy source
scenarios: hot volcanic pools, volcanic pools with freezing/heating cycles

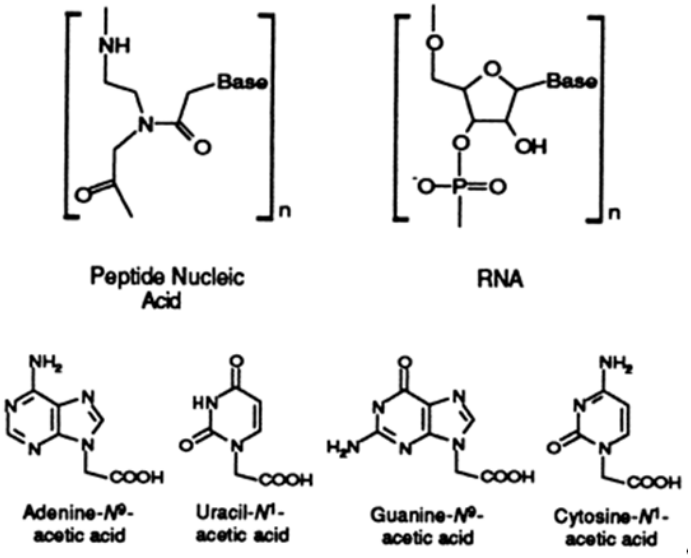
**Iron-sulfur world (Wächtershäuser)**
Wächtershäuser [38] proposed that early life may have formed on the surface of iron sulfide minerals, where the primary source of energy is the pyrite ($FeS_2$) reaction yielding energy and free electrons [40]

$$FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^- + 41\text{kJ/mol}$$

which has been demonstrated under volcanic pool and hydrothermal vent conditions.
Possible biosynthetic products are
thioacetic acid ($CH_3$-CO-SH) as a progenitor of *acetyl-CoA*

*aminoacids* glycolate/glycine, lactate/alanine, glycerate/serine as well as *pyruvic acid* (energy source in glycolysis).
Recently it has been shown that this process is coupled to the reduction of $CO_2$ to $CH_4$ (metanogenesis) and used in the deep sea biosphere as bio-energy source by at least one archaeon and five bacterial species [40].
In presence of CO2 at 75°C under anaerobic conditions in water this reaction yields also thiols, CS2 and dimethyldisulfide [63].

**PNA nucleotides (Miller 2000)**
PNA in the form of (A,G,U,C)-acetate *nucleotides* with the *amino-acid N*-(2-aminoethyl)glycine (AEG) replacing ribose-phosphate in RNA can *bind to peptides and to DNA*, and *polymerizes readily* at 100°C, can form a double-strand and is stable against hydrolysis (as opposed to RNA).[27]
AEG is produced by cyanobacteria: it is still present in the life chemistry.
Furthermore, under electric discharge $NH_4CN$ polymerization for ethylenediamine (ED) and AEG is possible, so both AEG and ED are prebiotic materials.

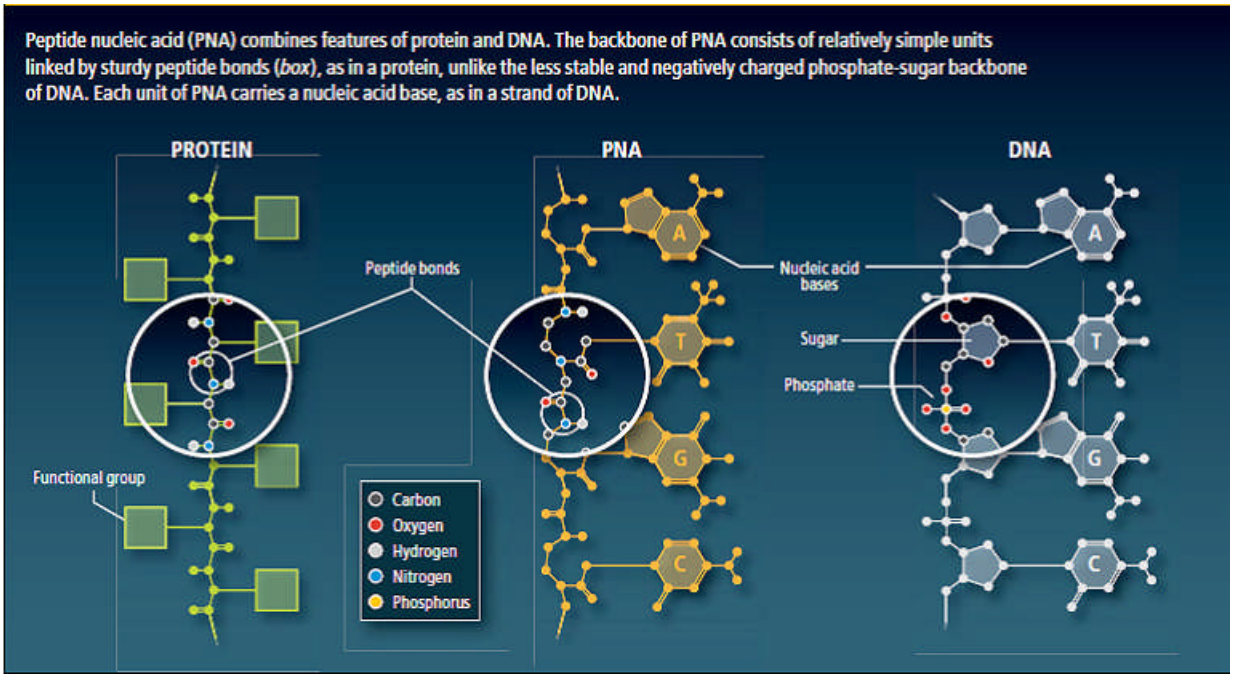



The basic AEG-amino-acid *glycine* initiates the synthesis of purine-acetates and its derivate hydantoic acid initiates the synthesis of pyramidine-acetates: [27]

$$NH_3 + HCN + glycine \xrightarrow[discharge]{} A-acetate + G-acetate$$

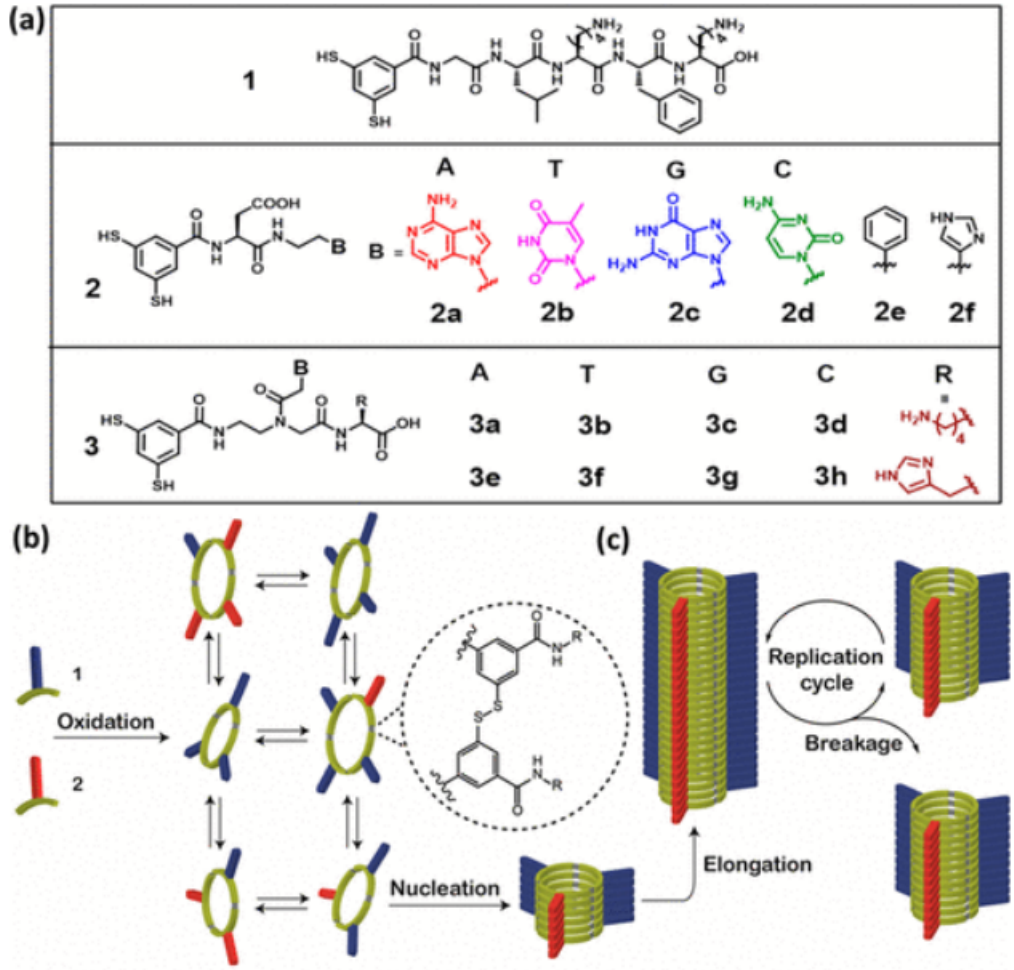
$$hydantoicac + cyanoacetaldehyde + urea \xrightarrow[discharge]{} U-acetate + C-acetate$$

PNA binds to DNA-oligomers and form PNA-DNA double-strands, which bind selectively to the corresponding mRNA-sequences and interfere in the peptide synthesis [12]. PNA-DNA-complexes form peptide-bonds with proteogenic aminoacids, depending on the associated nucleotide.




So basically, PNA-DNA or PNA-RNA double-strands can serve as vehicles for controlled peptide-synthesis from the 22 basic proteogenic amino acids.
Furthermore, PNA-DNA-complexes reproduce within lipid-bubbles, forming a sort of synthetic life.

**Self-replicating aminoacid-nucleotide polymers (Otto et al. 2020)**
In a seminal experiment the amino acids and nucleobases linked together into ring-shaped molecules [37]. Sets of these rings then assembled into cylindrical stacks.
Otto et al. have constructed molecular networks based on thiol–disulfide chemistry. Three families of relatively simple dithiol building blocks were prepared featuring peptides or amino-acid nucleobase conjugates.
A thiol is any organosulfur compound of the form *R-SH* , where R is an alkyl, i.e. *C-H* ring or chain.
Oxidizing (mixtures of) these dithiols generates mixtures of macrocyclic disulfides with different compositions and ring sizes, which interconvert through reaction of the disulfides with residual thiolate anion.
The stacks could copy themselves by somehow encouraging other rings to stack together in the same way. The mixture had to be shaken for this to work, but otherwise Otto’s team didn’t have to intervene.



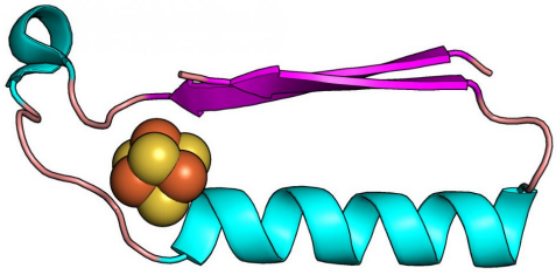

(a) Building block Structures; (b) Oxidation of these dithiol building blocks gives rise to dynamic combinatorial libraries of macrocyclic disulfides; (c) Assembly of a specific ring size into stacks leads to the autocatalytic formation of more of these rings through a fiber growth-breakage cycle

**Prebiotic proto-PNA based on ferredoxin (Raanan 2020)**

Recently, Raanan et al [39] showed that there is a simple prebiotic precursor peptide of *ferredoxins*, which bind to iron-sulfur compounds (i.e. can catalyze the *pyrite reaction* of the iron-sulfur world) and the so called *Rossmann-fold-peptides*, which can bind to nucleotides (i.e. can *form a PNA*).

Ferredoxins are small proteins containing FeS-clusters, they act as an electron transfer agent in biological redox reactions: in anaerobic bacteria e.g. in air nitrogen fixation, and in photosynthesis.

Consequently, Raanan's *proto-ferredoxin* is capable of catalyzing a pre-biotic energy cycle (the pyrite reaction) and can bind to nucleotides and amino-acids, i.e. it can presumably serve as building block in *self-replicating aminoacid-nucleotide polymers* like Otto's dithiol.



**Chirality**

The extraterrestrial aminoacids are predominantly left-handed. In terrestrial life chemistry, aminoacids are left-handed, nucleotides are right-handed (in its sugar part). It is well-known that right-handed tRNA handles left-handed aminoacids and vice versa. Therefore, it is plausible that peptide chemistry on Earth was initiated by

extraterrestrial aminoacids, and the left-handed variants gradually eliminated the less common right-handed variants.

**Membranes**

**Dworkin-bubbles**

A 2001 experiment led by Jason Dworkin [31] subjected a frozen mixture of water, methanol, ammonia and carbon monoxide to UV radiation, mimicking conditions found in an extraterrestrial environment or freezing-thawing cycles in terrestrial volcanic pools.

$$H_2O + NH_3 + CO \xrightarrow[UV,\ freezing]{} amphiphilic\ bubbles$$

This combination yielded large amounts of organic amphiphilic compounds (like aliphatic hydrocarbons in the Murchison meteorite) that self-organized to form bubbles or micelles when immersed in water. The bubbles produced in these experiments were between 10 to 40 um, or about the size of red blood cells. Remarkably, the bubbles fluoresced, when exposed to UV light, so they could have been a precursor to primitive photosynthesis.

**Proto-membranes from fatty acids**

The membranes of the first protocells on the early Earth were likely self-assembled from prebiotic extraterrestrial fatty acids, but these mebranes are unstable in salty water. Caitlin et al. showed 2019 that a set of unmodified, prebiotic amino acids binds to prebiotic fatty acid membranes and that a subset stabilizes membranes in the presence of salt and $Mg^{2+}$ [32]. Especially effective is here the hydrophobic amino acid leucine.

**Co-evolving and proliferating lipid vesicles and peptides**

A recent experiment by Mayer and Schreiber [62] , which mimics the conditions in volcanic rock cavities, shows that under pressure (100bar) and high temperature (120°C) in presence of CO2 in water, spontaneous formation and co-evolution of lipid vesicles and interacting oligopeptides takes place. The membranes form from the lipids of octadecylamine and octanoic acid, and peptide chains with up to 8 molecules build-up from added amino acids.

Castro et al. [67] found that lipid vesicles with a diameter of 10-20μm periodically bud and divide within a time of t=1300-1500s, under supply of precursor lipids and in presence of a copper catalyzer.

Kudella et al. [68] showed that lipid vesicles with a diameter of d~50μm divide under thermal heating by convection with periods around t=350s .

**Prebiotic conditions on Earth**

Early prebiotic soup experiments (e.g. Miller- Urey) postulated a reducing early Earth atmosphere rich in CH4 and NH3 . Recent research based on ancient zircons (~4.4Gy) showed that the **early atmosphere** consisted of ~90% N2 and ~10% CO2, with low contents of CH4, NH3 and HCN [49]. However, outgassing from chondrite meteorites must have released large quantities of CH4 , NH3 and also HCN [50]. These gasses are destroyed by UV radiation, but Ranjan et al. [51] showed that 10% CO2 was sufficient to protect CH4 and HCN from photolysis at the Earth surface, furthermore both are easily solvable in water. So in water, there was enough CH4, NH3 and HCN to support a prebiotic evolution. At temperatures around 0°C and pH~9 , steady-state HCN concentration of $2\ 10^{-6}$ in the early ocean has been estimated [41].

Furthermore hydrothermal vents were probably a powerful source of aquatic and atmospheric NH3, CO, CH4, and H2 on early Earth.[53]

There is evidence from ancient zircons [50, 51] that the surface temperature was at ~4.4Gy low enough for liquid water, and that **Earth was basically a water planet** before the begin of plate tectonics, with volcanic islands emerging from the ocean. As the young Sun produced much less radiation than at present, the moderate temperature on early Earth also supports a high CO2 content of ~10%.

The temperature in volcanic surface water were 80°-100°C, as today, which enable HCN chemistry with energy barrier around 40kcal/mol (167kJ/mol) [41]. Furthermore, pH values in the early oceans are estimated to 8..9, which allows HCN to stay undissociated ($pK_a$=9.31) and stable in solution [41].

Research on effects of lightning and bolide impacts on early Earth has shown that in these events **HCN is produced in large quantities** from CH4-NH3 and also from CO-N2-H2O (here by a factor 1000 less) [50].

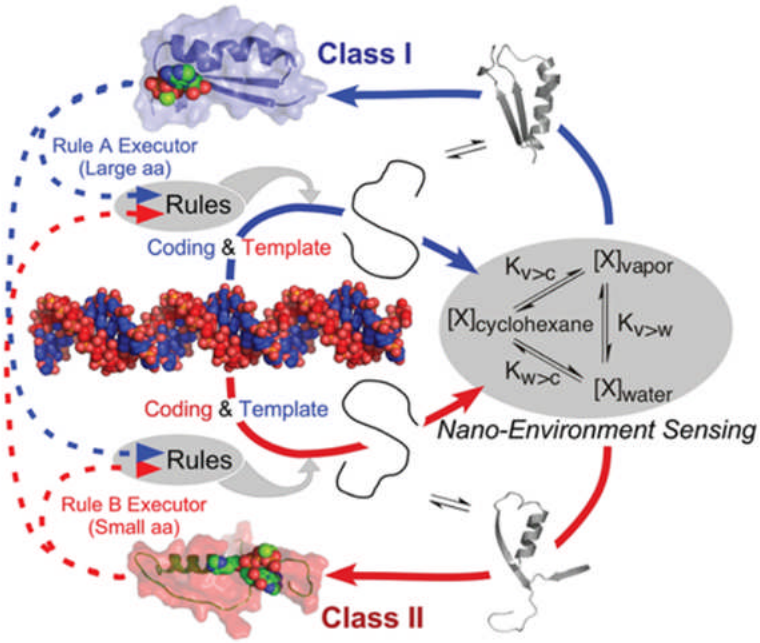
Zahnle [50] showed that HCN can be produced in high yields photochemically from EUV photolysis of N2 in CO2-N2 atmospheres provided that methane is present in significant amounts. These results support the existence of HCN prebiotic chemistry, which is currently the best candidate for life precursors.

## 2.5 Proto-PNA-RNA world

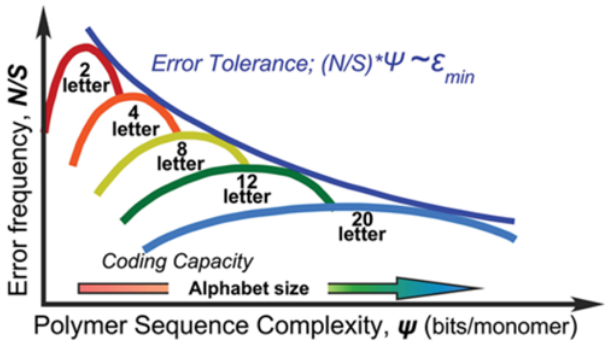
**PNA peptide synthesis based on a proto-genetic code with 1-bit-codon (Carter-Wills 2017)**

The 20 aaRS (tRNA-ligase) are central to the protein synthesis. In a seminal paper in 2017 Carter & Wills [29] analyzed their structure and found that they separate into two distinct classes, which descend from proto-leucine-RS and proto-histidine-RS for classI and classII respectively.
Correspondingly, the 20 proteinogenic aminoacids separate into classI (8 D-amino-acids, cysteine, glycine) with large side chains and classII (5 A-amino-acids, 4 B-amino-acids, proline) with small side chains.
We get a proto-code with 1-bit-codon, occupied by *glycine- guanine-PNA* (=AEG-G) and coding for classI-aminoacids and leucine-RS (*leucine-guanine-PNA* proto-ligase), or its code-complementary *proline- cytosine-PNA* coding for classII-aminoacids and histidine-RS (*histidine-cytosine-PNA* proto-ligase) . The proto-ligases build up the PNA-ladder, which splits at the end of the enzyme-synthesis, releasing the finished enzyme.
In the RNA-code glycine is coded GGx, and proline is coded by the code-complementary CCx .
The *energy transfer* was carried out probably by GTP (guanine-triphosphate) instead of the ubiquitous ATP in today's organisms, as only G and nucleotides were involved in protein synthesis.
The *protozyme-gene* for peptide synthesis via proto-RS was a *bidirectional gene*, which coded for proto-leucine-RS and its complementary version coded for proto-histidine-RS . The classI ancestral aminoacid was glycine and classII ancestral aminoacid was proline. [30]
The peptide production and reproduction for classI and classII peptides influenced each other and formed a *hypercycle*.
The complexity of aaRS is about 250 aminoacids, the calculated complexity of the protozyme gene is about 40 aminoacids. [29]

**Evolution of full RNA genetic code from PNA proto-genetic code**




Process simulation shows, that the translation error frequency decreases strongly with the number of letters(aminoacids), so natural selection favored increasing number of aminoacids, and also the complexity of generated peptides, which improved their functionality in enzymes.

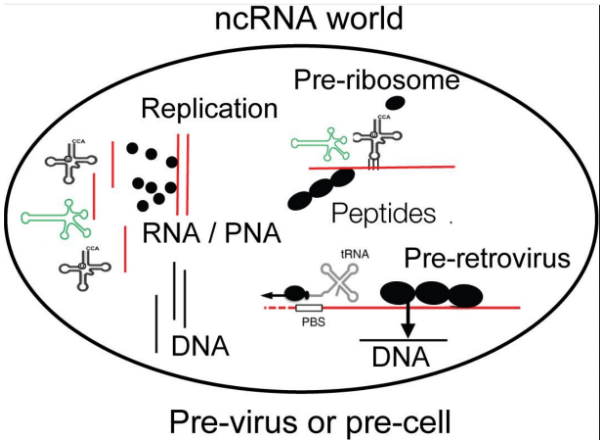



Rodriguez et al. showed 2015 [30] how the PNA 2x1 genetic code with 2 code letters (glycine-guanine-PNA, proline-cytosine-PNA), 1 protozyme-gene with proto-leucine-RS and proto-histidine-RS as translation

enzymes, 2 aminoacids (glycine, proline) , evolved into the full 4x3 RNA genetic code with 4 nucleotides (A,G,U,C) 20 aminoacids, 20 aaRS coded by 20 rRNA genes and ribosome peptide synthesis.

| | |
|---|---|
| PNA code 2x1<br>1 bit codon<br>2 letters Gly-G-PNA Pro-C-PNA<br>2 aminoacids Gly Pro<br>2 enzymes Leu-RS His-RS<br>1 bidirectional gene, length ~40 bases | → full RNA code 4x3<br>3 bit codon<br>4 letters A G U C<br>20 aminoacids<br>20 enzymes aaRS<br>20 genes for aaRS, length ~750 bases |

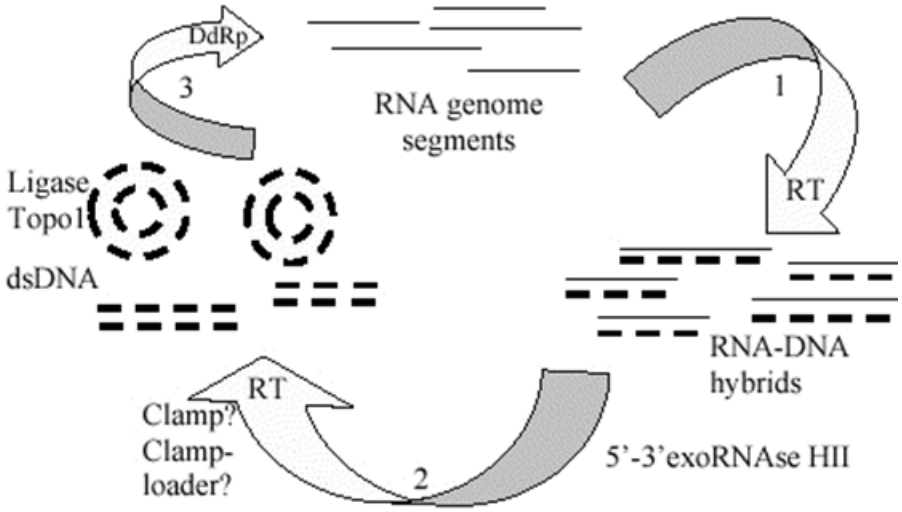
**proto-ribosome=ribozyme: ncRNA world**
The ribozyme was a ss-circular RNA, it could form tRNA and other ribozymes:it was a a quasi-species .[34]
The ribozyme could make peptides without genetic coding, only by structural information, it could also perform reverse transcription of RNA to DNA like the reverse viruses.
Manfred Eigen demonstrated 2013 that a mixture containing no RNA to start with but only ribonucleotides and the Qβ replicase can under the right conditions in a test tube spontaneously generate self-replicating ncRNA.[34] .



**2.6 LUCA**

LUCA (Last Universal Common Ancestor of the three super-kingdoms of life) had: [36]
■RNA genome ~250 genes
■ribosome
■virus-like protein capsid+lipid envelope
■RNA 4 nucleotides
■proteins with 20 aminoacids
■RNA replication via DNA intermediates
■DNA polymerase
■Class I: 9 aaRS
■Class II: 7 aaRS → only 16 aminoacids
■15 families of small subunit proteins, 18 families of large subunit proteins
■synthesis of RNA using DNA templates

LUCA cells was using the reverse acetogenesis pathway with sulfate reduction, it was living in hydrothermal vents and/or hot volcanic pools.[6,10]
LUCA's genome consisted of multiple segments of RNA, which replicated via DNA intermediates in a retrovirus-like replication cycle. This accounts both for the lack of conservation of several central components of modern replication systems and for the presence of some other conserved components, such as, for example, the sliding clamp, clamp-loader ATPase, and RNAse H, as well as enzymes of DNA precursor biosynthesis, and the basal transcription machinery. It also explains, why the DNA-polymerases in bacteria and in archaea are different.




**RNA viruses**

dsRNA(double-strand) viruses evolved from +RNA viruses on at least two independent occasions. [33].
Analysis based on the RdRp (RNA-dependent RNA polymerase) suggests that the last common ancestors of the major branches of RNA viruses encoded only the RdRp and a single jelly-roll capsid protein.
+RNA-viruses (transcription in the same direction as template) originated from the ribozyme.
Genetic studies suggest that RNA viruses evolved from LUCA and are older than bacteria and archaea.

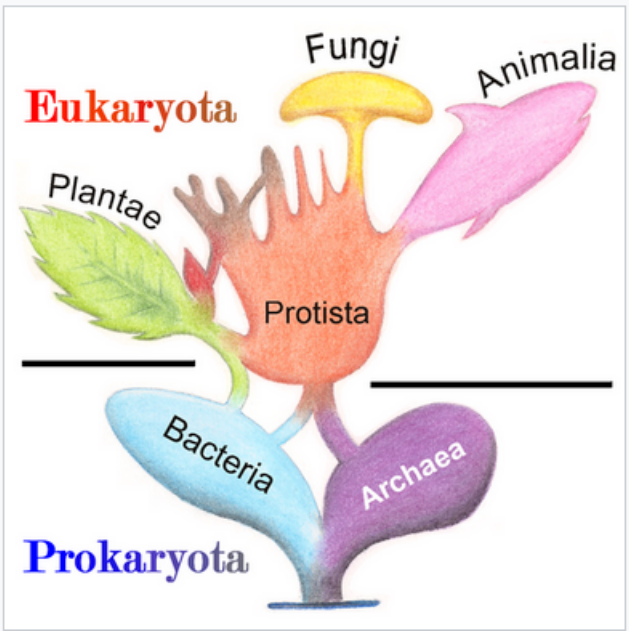
## 2.7 DNA world

### Bacteria and archaea

Bacteria and archaea evolved from LUCA probably 3800 million years ago. Archaea populated volcanic pools and hydrothermal vents. Bacteria populated coastal waters.
Both developed a DNA-genome, but with different DNA polymerases. Both acquired a lipid membrane and cell wall, but with different structure. Archaea extended the number of aminoacids (and aaRS) to 22, bacteria to 20. The RNA replication functions via RNA polymerases (as opposed to LUCA's RNA-DNA hybrids), but the functionality is different.

Archaea**:** cell membrane contains ether linkages; cell wall lacks peptidoglycan; genes and enzymes behave more like eukaryotes; have three RNA polymerases like eukaryotes.
Bacteria**:** cell membrane contains ester bonds; cell wall made of peptidoglycan; have only one RNA polymerase.

### Eukaryotes




Eukaryotes: protists, fungi, plants, animals

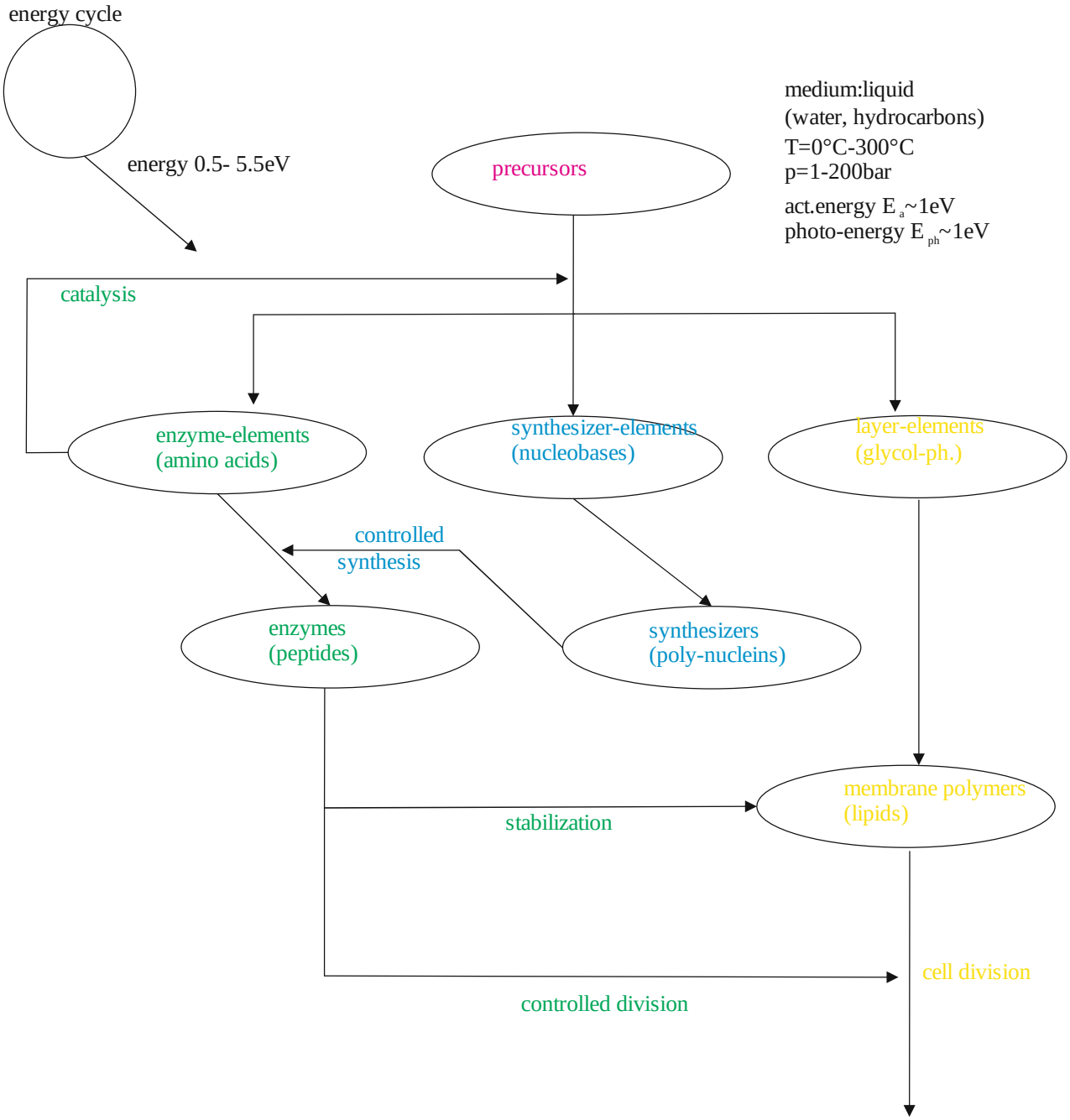
Eukaryotic protists emerged from a symbiosis between an archaean with a bacterium (mitochondrion). In a similar process, eukaryotic plants acquired chloroplasts by symbiosis with a cyanobacterium.

# 3 Principles and evolution of life chemistry

## 3.1 Chemical base of life in general

Based on current experimental and theoretical evidence, we can conclude that the proto-biotic life chemistry in general (not only on Earth) develops in a polar liquid from one or several exothermic molecules (precursors) by selection-evolution via two parallel prebiotic random polymerization processes from basic compounds to catalyzers (enzymes) and synthesizers (polymer molecules which carry out template-based enzyme synthesis based on their sequence, i.e. genetic code) fed by an energy cycle. In terrestrial life chemistry, a vesicle-based chemistry is a precondition for a full life-cycle with bio-matter production and proliferation (see 3.8.4). Therefore it is plausible to assume that in general there is a third class of polymers (layer-builders) produced by the life chemistry, which form a membrane and a self-dividing vesicle.

### Life chemistry and physics in general



**Alternative life-models**

**Carbon-nitrogen-hydrogen chemistry in water**
T=0°C...150°C (P=1-100bar)
basic materials: $HCN + FeS + H_2S$
polymers: peptides, poly-nucleins
solvent: water
energy cycle: $FeS + H_2S$ , photosynthesis
location: water-carrying planets in habitable zone

**Carbon-hydrogen-oxygen chemistry in water**
T=0°C...150°C (P=1-100bar)
basic materials: $CO_2 + CH_4 + NH_3$
polymers: poly-alcohols, poly-ethylene-glycols, poly-ethylene-imines
solvent:water
energy cycle: photosynthesis, $CO + H_2O$
location: water-carrying planets in habitable zone

**Carbon-hydrogen-nitrogen-sulfur chemistry in heavy hydrocarbons**
T=0°C...300°C (P=1-100bar)
basic materials: $CH_4 + CO_2 + H_2S + NH_3$
polymers: hydrocarbons, aromatic hydrocarbons (benzene-derivates), cyclo-hydrocarbons, amino-hydrocarbons, thio-hydrocarbons
solvent: pentane and heavy hydrocarbons $C_nH_{2n+2}$
energy cycle: photosynthesis ?, electrical discharge ?
location: planets with medium-range temperatures with dense atmosphere (Venus-like)

**Silicon chemistry in molten rock**
T=1000°C...1500°C
basic materials: $SiO_2 + Al_2O_3 + MgO$
polymers: polysilicates
solvent: molten silicates
energy cycle: ?
location: hot Earths

**Carbon+salts compounds in liquid ammonia**
T= −78 ... −33 C (P=1bar)
T=−77 C ...98 C (P=60bar)
solvent: liquid ammonia
basic materials: hydrocarbons, salts, organic metal complexes, organic carbon compounds $C - H - N$
polymers: complex hydrocarbons with oxyl- , amino- and metal-radicals
energy cycle: photosynthesis ?, tidal heating ?
location: Titan-like planets

**Carbon compounds in liquid ethane-methane**
T= −183 ... −89 C (P=1bar)
solvent: liquid ethane-methane
energy cycle: photosynthesis ?, tidal heating ?
basic materials: organic carbon compounds $C - H - N$
polymers: complex hydrocarbons with oxyl- , amino-radicals
membrane (computer-modeled in February 2015): acrylonitrile similar to a phospholipid bilayer
location: Titan-like planets

**3.2 Principles of terrestrial life chemistry**

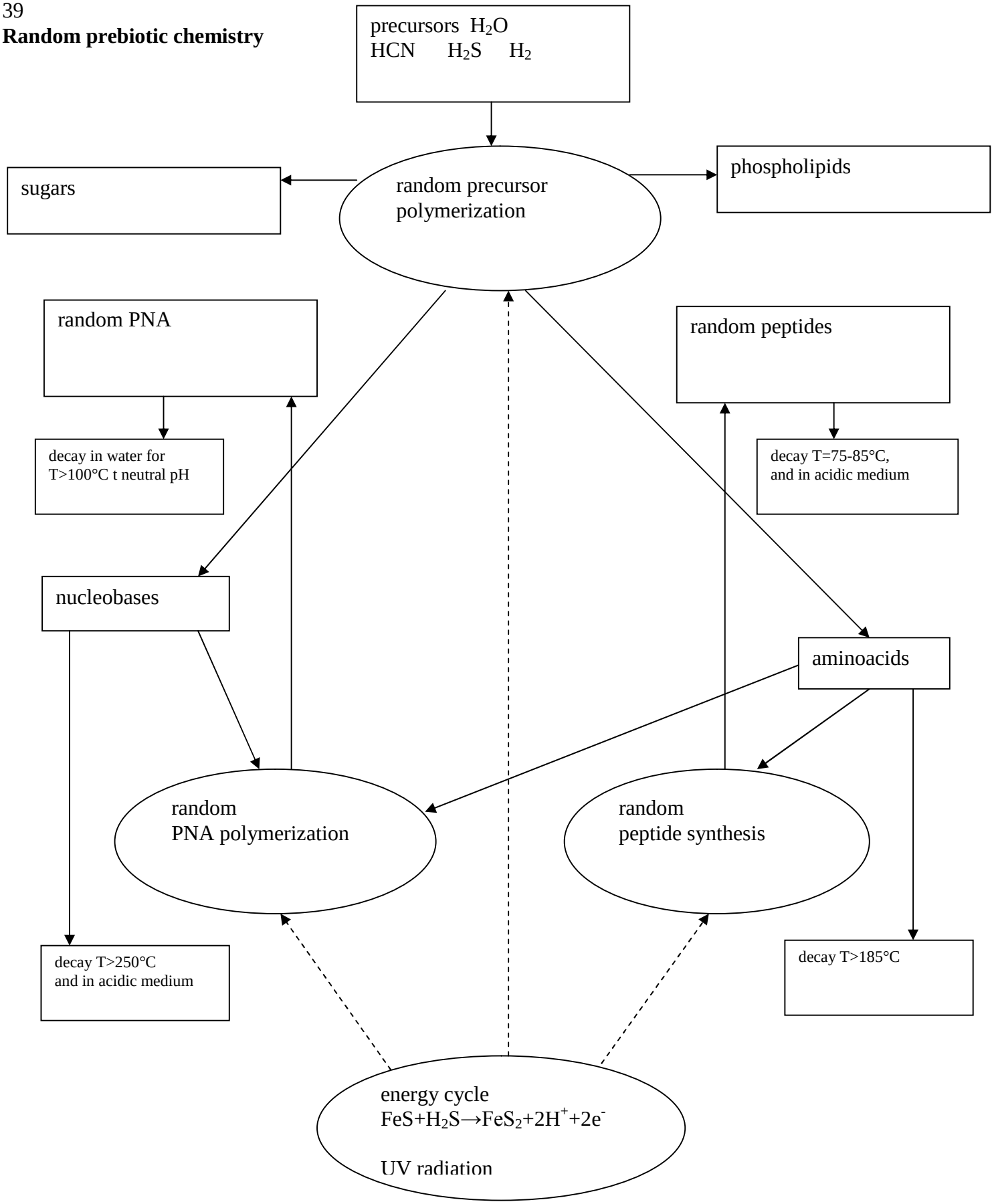
In the case of the terrestrial life, the process took place probably from the precursor molecules {HCN , H2} in water via prebiotic synthesis of amino acids and nucleobases under UV-radiation, and then parallel polymerization to catalyzers (poly-peptide enzymes) and synthesizers (peptide-nucleotides PNA) fed by the pyrite energy cycle (see flowchart 'random prebiotic chemistry').
The polymers poly-peptides and PNA are not very stable, especially in acidic water and temperatures around boiling point 100°C, so they must be continually reproduced in order to survive [43, 44]: that is where the Darwinian evolution mechanism sets in. The amino acids and nucleobases also degrade, although at higher temperatures (T>185°C), so they, too, must be replenished [42]. On the other hand, the enzymes are continuously in contact with precursors molecules and with the energy cycle, so they degrade relatively quickly, so they must be reproduced as quickly. Simple copying is not as efficient as replication based on a stable pattern molecule: this is the role of the synthesizers (PNA).
The evolution selects a coupling between peptides and PNA-segments (genes), where the PNA-segments build-up the corresponding peptide from amino acids, and in turn each of the peptides catalyzes one of the required metabolic processes: synthesis of required amino acids and nucleobases from precursors, the energy cycle, synthesis of phospholipids and the build-up of the protective membrane.

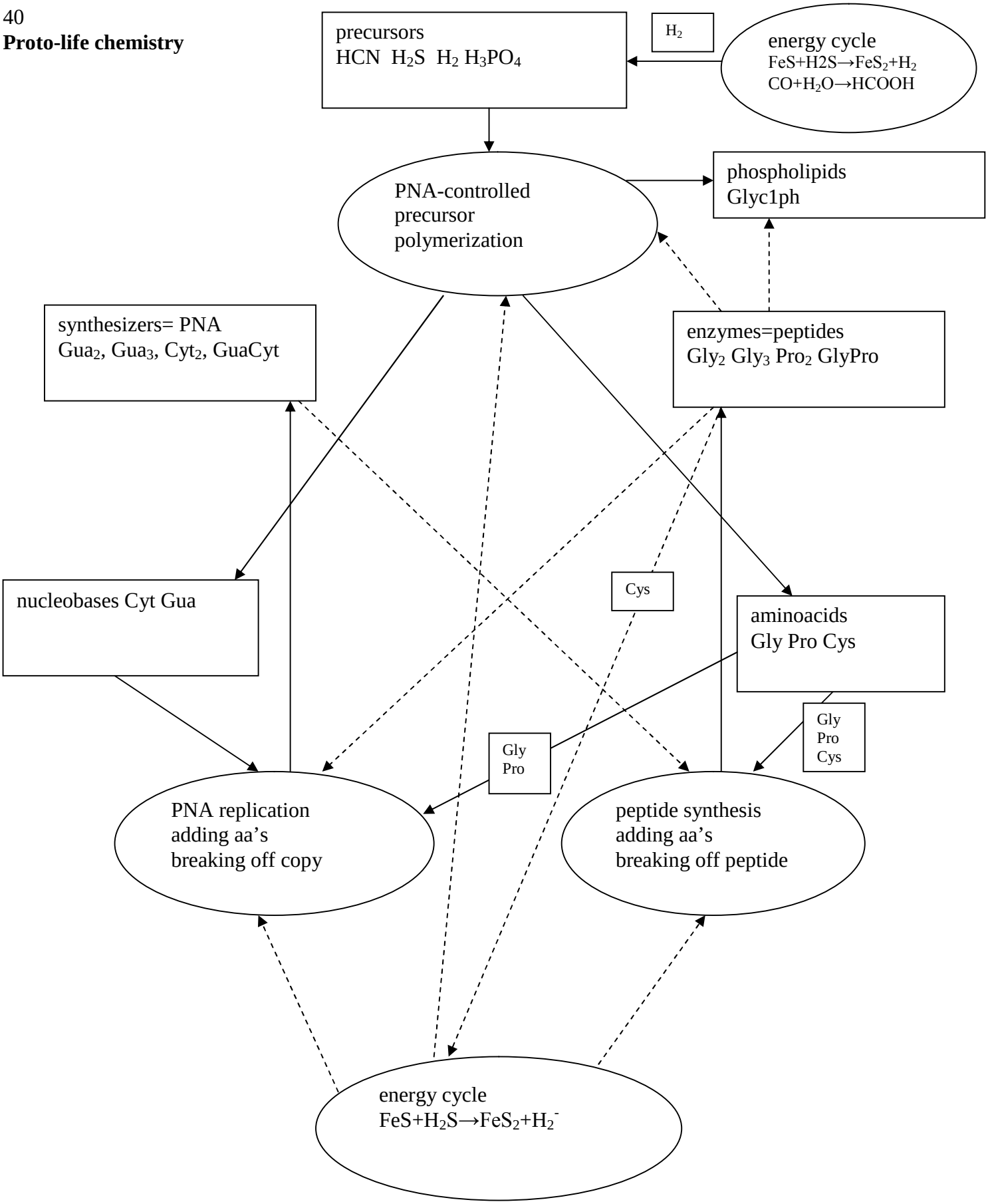
The result is a **model1 1-bit genetic code** (Gua-Cyt nucleobase pair) stored in the PNA (Leu-Gua and His-Cyt), coding for Gly and Pro amino acid sequences, which in turn catalyze the synthesis of **5 needed amino acids** (Gly, Pro, Leu, His, Cys) , nucleobases (Cyt, Gua) and phospholipid glycerol-phosphate , and the catalysis of the energy cycle via Cys with SH-radical (see flowchart 'proto-life chemistry').
Model1 has 7 genes coding for 8 biomolecules Gly, Pro, Leu, His, Cys, Cyt, Gua, glycerol-phosphate .
In the **enzyme synthesis** Leu-Gua functions as the ligase for Gly, i.e. during the synthesis process Gly docks to Leu (resp. Pro docks to His), the ligase separates from the growing enzyme, and the next step begins. The enzyme is completed, when the last amino acid in the gene sequence is added. We can assess the selectivity of the ligases by comparing the *hydrophobe compatibility index* HI [48] of the ligands:
HI(Pro, Pro)= HI(Gly, Gly)= 20.0 , HI(Pro, Gly)=18.3, HI(Pro, Leu)=16.5 , HI(Pro, His)=17.5 ,
HI(Gly, Leu)=17.3 , HI(Gly, His)=15.8
The selectivity of Leu-Gua between Gly and Pro is ΔHI= HI(Gly, Leu)- HI(Pro, Leu)= 0.8 ,
the selectivity of His-Cyt between Pro and Gly is ΔHI= HI(Pro, His)- HI(Gly, His)= 1.7 .

In the **model2 genetic code**, the enzyme synthesis can be simplified by replacing Leu-Gua by Gly-Gua and His-Cyt by Pro-Cyt. In fact, Leu = Gly + CH-CH-CH3
-CH3 inserted at the end
His = Pro + N-CH-NH breaking the 5-ring at NH and inserting the 3-chain to close the ring .
The selectivity of Gly-Gua between Gly and Pro is ΔHI= HI(Gly, Gly)- HI(Gly, Pro)= 1.7 ,
the selectivity of Pro-Cyt between Pro and Gly is ΔHI= HI(Pro, Pro)- HI(Gly, Pro)= 1.7 .
So the selectivity of the ligases is comparable to model1. On the other hand, the binding enthalpies of Leu and His are considerably higher than those for Gly and Pro [47]:
$\Delta H_f$(Gly)=-390.5kJ/mol , $\Delta H_f$(Pro)=-366.2kJ/mol, $\Delta H_f$(Leu)=-648.9kJ/mol, $\Delta H_f$(His)=-441.8kJ/mol ,
so the model1-PNA's are more stable than in model2, therefore model1 should be considered as an evolutionary improvement of model1.
With this simplification, model2 works with **3 needed amino acids** (Gly, Pro and Cys) and 2 nucleobases (Cyt, Gua), with enzymes built from (Gly , Pro) and PNA-ligases Gly-Gua and Pro-Cyt .
Model1 has 5 genes coding for 6 biomolecules Gly, Pro, Cys, Cyt, Gua, glycerol-phosphate .
With Gly-Gua and Pro-Cyt as PNA's and ligases, the enzyme synthesis works by binding Gly to Gly and Pro to Pro side-by-side by peptide-bond and then breaking the finished enzyme at the peptide bonds away by energy (i.e. electron) transfer from the energy cycle. The PNA-replication works by selectively binding Gly to Gly and then Gua to Gly, and by binding Pro to Pro and then Cyt to Pro; at the end the PNA-copy breaks away at the peptide bond again by energy transfer.

These two models are of course at first only a plausible scenario, but it is supported by the known prebiotic HCN-chemistry able to produce all amino acids, all nucleobases, phospholipids and sugars [35, 41].
Furthermore, the 1-bit genetic code has been shown to be the probable original mechanism of PNA-RNA-coding and protein synthesis [29, 30] . Then, one can show by enzyme calculations, that Gly-Pro-peptides

indeed act as enzymes in the synthesis of amino acids and nucleobases from (HCN, H2O , H2) (see 3.2 , 3.4 , 3.5 and [45]) . Finally, digital simulation based on a simplified model of the HCN chemistry show that molecular evolution selects indeed a 1-bit genetic code like the one proposed in [29, 30].
A general consequence of the validity of this scenario is that the elementary enzymes (amino acids) and elementary synthesizers (nucleobases) are *selected by evolution based on the precursor chemistry*, i.e. here of the HCN-water chemistry, and not vice versa.


precursors $H_2O$
HCN $H_2S$ $H_2$
sugars
random precursor polymerization
phospholipids
random PNA
random peptides
decay in water for T>100°C t neutral pH
decay T=75-85°C, and in acidic medium
nucleobases
aminoacids
random PNA polymerization
random peptide synthesis
decay T>250°C and in acidic medium
decay T>185°C
energy cycle
$FeS+H_2S \rightarrow FeS_2+2H^+ +2e^-$
UV radiation


precursors
HCN $H_2S$ $H_2$ $H_3PO_4$
$H_2$
energy cycle
$FeS+H2S \rightarrow FeS_2+H_2$
$CO+H_2O \rightarrow HCOOH$
PNA-controlled
precursor
polymerization
phospholipids
Glyc1ph
synthesizers= PNA
$Gua_2$, $Gua_3$, $Cyt_2$, GuaCyt
enzymes=peptides
$Gly_2$ $Gly_3$ $Pro_2$ GlyPro
nucleobases Cyt Gua
Cys
aminoacids
Gly Pro Cys
Gly
Pro
Gly
Pro
Cys
PNA replication
adding aa's
breaking off copy
peptide synthesis
adding aa's
breaking off peptide
energy cycle
$FeS+H_2S \rightarrow FeS_2+H_2^-$

### 3.3 Empirical models of molecular energy and reaction rate

In principle, it is possible with methods of quantum chemistry (e.g. Hartree-Fock calculation) to calculate the chemical evolution in time of a given mixture of molecules in liquid solution, given the initial concentration, and physical conditions (temperature, pressure, external flow of matter, energy supply by radiation and spark discharge), and based solely on the structure of initially given and of emerging molecules. In reality, in order to calculate a realistic scenario in this way, the needed computational power exceeds by far the performance of today's supercomputers.

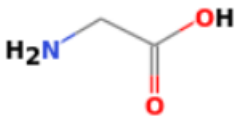
There are two kinds of data, which are needed for a structure-based numerical simulation of chemical evolution. First, we need the molecular formation energy $H_f$ (molecular energy for short) for a given molecule, as a function of its bond structure. $H_f$ is required for the calculation of the reaction energy $\Delta E$ of a given reaction

$A + B\ (k_+ \rightarrow)\ C + \Delta E$

Second, we need the reaction rate $\frac{dC(t)}{dt}$ for such a reaction as a function of the bond structure of the reacting molecules A and B (here $C(t)$ denotes the concentration of the compound $C$).

#### 3.3.1 Empirical model of molecular energy

The *molecular forming free enthalpy* $\Delta H_f$ has been measured for hundreds of molecules, values can be found in [55, 56, 58] . It is defined as the forming energy from natural state at normal conditions (T=300K, P=1 bar). The *molecular forming energy from atoms* $H_f$ is calculated as $H_f = \Delta H_f$-E(gas) , where E(gas) is the forming energy of gaseous components (H2, O2, N2) in the molecule, e.g. for glycine we get




formula C2H5NO2
$\Delta H_f$=-390.5 kJ/mol [55]
E(gas)=E(2.5xHH,O=O, 0.5 N≡N)=2.5x432+494+0.5*942=2045
$H_f = \Delta H_f$-E(gas)=-2045-390.5=-2435.5 kJ/mol
where the bond energies of HH=432kJ/mol in H2, (O=O)=494kJ/mol in O2, (N≡N)=kJ/mol in N2 [58].
The values of $H_f$ are calculated in [58].
A molecule can be characterized by a list of bonds, e.g. for glycine we get
{ OH, CO, C2O, CC, 2 CH, CN, 2 NH }, where C2O is the double bond C=O [58]. The values of the bond energies are given in data tables like [55, 56]. The sum of bond energies is in most cases considerably higher than the molecular energy $H_f$ , the actual bond energy $E_b$ in a molecule is mostly lower than the measured breaking energy of a bond. The bond factor $f_b$ depends predominantly on the bond, and, much more weakly, on the neighboring bonds in the molecules.
The empirical formula with the bond factors becomes $H_f = \sum_i f_{b,i} E_{b,i}$ ,
where the bond factors $f_{b,i}$ are fitted on a large set of 67 molecules, among them amino acids, nucleobases and biochemical building block molecules [58, 45].
The original bond energies $E_{b,i}$

```
{H1HO → 21., H3N1H → 13., HO1H → 138., CH → 411., C3N → 887., H1CN → 120., CN1CN → 535., HH → 432., CO → 358.,
 C1OH → 14.2, NH → 314., OH → 459., C2O → 799., CC → 346., CN → 305., C2C → 602., C3C → 835., C2N → 615., CS → 272.,
 SH → 363., NN → 167., C2S → 573, NO → 201., N2O → 607., O2O → 494., N3N → 942., OO → 142., FeS → 315., Fe2S → 330.}
```

are factorized with the $f_{b,i}$ and modified

```
{CH → 133.583, C3N → 464.17, CO → 338.768, NH → 385.499, OH → 341.877, C2O → 527.8, CC → 34.5999, CN → 160.685,
 C2C → 60.2, C3C → 83.5, C2N → 329.712, CS → 203.504, SH → 36.3, NN → 313.921, C2S → 282.831, NO → 402., N2O → 623.75}
```

with mean relative error merr= 0.0552
example: for glycine measured (absolute) $H_{fm} = 2435.5$
calculated $H_f = \sum_i f_{b,i} E_{b,i} = 2467.2$ , with the relative error $rerr(Hf) = 31.7/2435.5 = 0.013$ .

### 3.3.2 Empirical model of reaction rate

In physical chemistry, the reaction rate of a reaction is described by the law of mass action [61].

$A + B\ (k_+ \rightarrow)\ C + D + \Delta E$ resp. $A + B\ (\leftarrow k_-)\ C + D + \Delta E$

where $\Delta E$ is the reaction energy , the reaction is exothermic, when $\Delta E > 0$ .

The Arrhenius law for the reaction constant states that [26]

$k = A\exp(-E_a / kT)$ , where $E_a$ is the activation energy, $A$ is the pre-exponential factor .

When concentration are measured in (dimensionless) *relative mole* (and not *1/mol* as usual), the constant $A$ has the dimension *1/s*, i.e. $A = \frac{1}{t_0}$ , where $t_0$ is the interaction reaction time, in biochemical reactions under normal conditions in water, $t_0 \sim 10^{-9} s$ . The constant $A$ is only weakly temperature-dependent.

The rate constant $A$ depends on the diffusion constants and the critical length (mean free path) $\lambda$ of the liquid $(A_+)^{-1} = t_0 = \lambda^2 / 6(D_A + D_B)$ [26], where $D$ is the diffusion constant.

In liquids, $D$ is described by the Einstein formula $D = \frac{kT}{b\pi\, r_0 \eta}$ ,where $\eta$ is the liquid viscosity, $r_0$ the molecule radius, $b=6$ for large molecules.

For linear molecules, we can approximate $r_0$ by $r_0 = \sqrt{L_m d_m}$ , where the molecule is described by a cross section area of length $L_m$ and diameter $d_m$ .

So we can describe D by the formula $D = \frac{D_0}{\sqrt{\sum_k R_{b,k}}}$ , where $R_{b,k}$ are the bond lengths in $A$ and $D_0$*(solvent)* is a diffusivity constant of the solvent, here water.

Fitting measurement data with these models shows that the geometric mean of the $D_i$ rather the arithmetic mean in the formula for t0 is a good model:

$$t_0 = \frac{\lambda^2}{3\sqrt{D_1 D_2}}$$

Fitting the measured $t_0$ with these models with the bond-factors $f_{b,k}$ , the diffusivity $D_0$ and the mean free path $\lambda$ yields the values

$\lambda$*(water)=1.62 nm* , $D_0$*(water)= 2.52* $10^{-9}$ $m^2/s$

The resulting bond length in $A$ are

```
{H1HO → 0.982011, H3N1H → 10.1, CH → 4.01377, C3N → 0.0407441, HH → 1.60278, CO → 0.0160739, C1OH → 0.162053,
 NH → 0.0251487, OH → 1.0567, C2O → 0.0119996, CC → 0.0155343, CN → 0.0945953, C2C → 0.13584, C2N → 0.210149, CS → 3.51452,
 NO → 0.0139999, N2O → 0.0504201, O2O → 5.71509, N3N → 0.0109996, OO → 5.58425, D0x → 2.52539, lamx → 1.62113}
```

In the same way, fitting the activation energy $E_a$ with the bond-factors $f_{b,k}$ of the bond energies $E_{b,k}$ in the model ,

$$E_a = \left| \frac{\sum_k f_{b,k} Ei_{b,k}}{N_i} - \frac{\sum_k f_{b,k} Eo_{b,k}}{N_o} \right|$$ , where $Ei_{b,k}$ resp. $Eo_{b,k}$ are the input resp. output bond energies

yields the following values for $E_{b,k}$ in kJ/mol

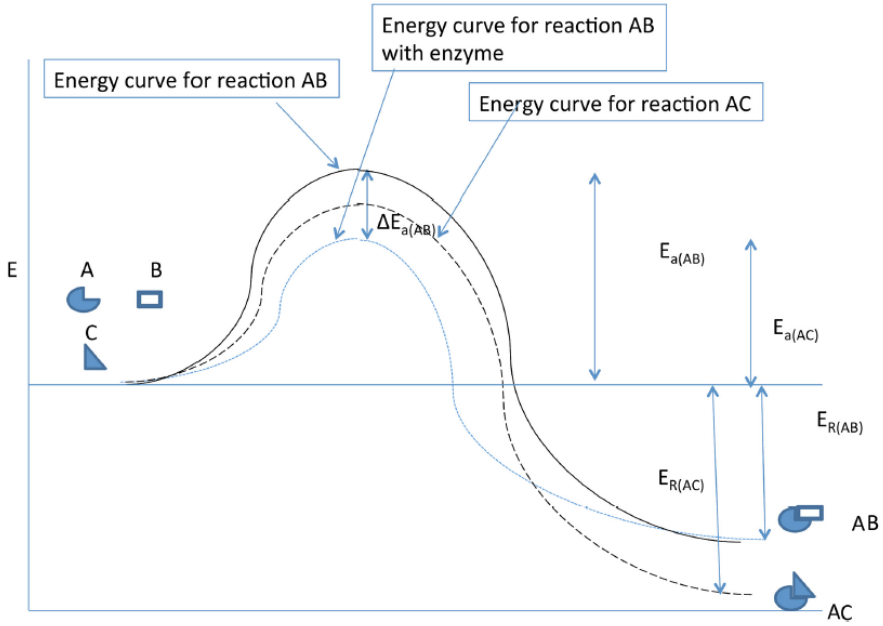
```
{H1HO → 138.548, H3N1H → 76.5232, HO1H → 243.782, CH → 190.452, C3N → 735.774, HH → 940.299,
 CO → 275.043, C1OH → 142., NH → 410.903, OH → 226.86, C2O → 61.3247, CC → 212.022, CN → 146.484, C2C → 530.371,
 C2N → 248.433, CS → 2.75897, SH → 289.709, NO → 280.681, N2O → 6.06966, O2O → 265.101, N3N → 743.278, OO → 90.7048}
```

### 3.4 Theory of enzymes: lock-and-key theory

Enzymes (in terrestrial life peptides) are the key element of life in general. They accelerate specific reactions $10^4$-$10^6$ fold and, coupled to synthesizers (PNA/RNA), they enable the reproduction and survival of all life molecules, including themselves.

The specific action of an enzyme with a single substrate (precursor molecules) can be explained using a lock-and-key analogy. In this analogy, the lock is the enzyme and the key is the substrate. Only the correctly sized key (substrate) fits into the key hole (active site) of the lock (enzyme).

Below, we present the mathematical formulation of the action of an enzyme: the substrate (with site bonds $E_n$) allocates along the enzyme (with site bonds $E'_n$), the substrate sites are bound to the substrate sites with bond energy $F_n$ . The substrate bonds are then weighted with local density $p_n$ , and the resulting mean bond energy $E_{bc} = \sum_n p_n E_n$ is considerably lower than the original mean bond energy $E_{bm} = \sum_n E_n / N$ : this is threshold-reducing action of an enzyme, illustrated in the graphics below [46].




Information in living systems manifest through “temporal gradients”. Here the system contains initially two substrates and one enzyme. In the absence of the enzyme, reaction C→G + H will proceed more rapidly because it has both lower final free energy and lower activation energy. However, the enzyme lowers the $E_a$ for reaction B→E + F. The information in the enzyme produces an observable gradient over time as the concentrations of E and F are increased and B is decreased when compared to an uncatalyzed system. In contrast, because of its specificity, the enzyme has no effect on the temporal evolution of the substrate and product concentrations of reaction C→G+H.

Initially assuming a well-mixed distribution of enzymes and substrate of equal concentration ,we view the “lock” as constantly-spaced enzyme molecules of density profile $r_n = r(x_n + \Delta x/2)$, $x_n = n\Delta x$, $\Delta x$ small, with density values $r_n$.
The substrate (or “key”) molecules are particle pairs having a local density profile $p_n = p(x_n)$ at positions $x_n = n\Delta x$. Each enzyme-substrate ‘complex’ locally lowers the activation energy of the reaction so that overall activation energy is maximally lowered when all key particles are closest (Kullback-Leibler-distance) to the corresponding lock particles. The KL-distance is a $r_n$ -weighted Boltzmann entropy of the profile $p_n$ , which makes it a plausible ansatz from the viewpoint of thermodynamics.

The enzyme KL-model is formulated as follows.
Minimization problem:

$$H_{KL}(p||r) = \Sigma_{n=1}^{N} p_n \ln\left[\frac{p_n}{r_n}\right].$$ =min

$\Sigma_n p_n = 1,$ constraint: normalization substrate,
with bond energies $E_n$ , $\sum_n E_n = E_b$ , $E_b$ is the molecular energy of the substrate.

Enzyme density values $r_n$ are normalized $\Sigma_n r_n = 1,$ with bond energies $E'_n$ and the molecular energy $r_n = \frac{E'_n}{E'_b}$ , and $\sum_n E'_n = E'_b$ , $E'_b$ is the molecular energy of the enzyme.
With Lagrange-multipliers we get the minimization problem

$$\sum_{n=}^{N} p_n \ln\left(\frac{p_n}{r_n}\right) - \Lambda_1\left(\sum_{n=}^{N} p_n - 1\right) = \min$$

Differentiation $\partial p_n$ and $\partial \Lambda_1$ gives $N+1$ equations for $N+1$ variables $\{p_n, \Lambda_1\}$

$$1 + \ln\left(\frac{p_n}{r_n}\right) - \Lambda_1 = 0 \quad \text{and} \quad \sum_{n=1}^{N} p_n = 1$$

we get $1 - \sum_{n=1}^{N-1} p_n = p_N$ and $\Lambda_1 = \left(1 + \ln\left(\frac{p_N}{r_N}\right)\right)$ as 2 equations for $p_N$ and $\Lambda_1$

and $p_n = r_n Exp(\Lambda_1 - 1)$

In this case, differentiation yields the local minimum $p_i=r_i$ with $min=0$, numerical minimization yields global minimum with negative values, the solution has one large value $p_{i0}$ close to 1, at the index of the smallest $r_i$ ( $r_{i0} = Min(r_i)$ ) , and the rest of the $p_i$ is small. The enzyme coupling "selects" the weakest bond and preserves it, and attenuates the energy of the other bonds: this is the catalyzer mechanism.

**Example lock-and-key theory**

We illustrate the Kullback-Leidler enzyme model by calculation of a concrete example: the **prebiotic synthesis of glycine from HCN**, **water and hydrogen catalyzed by diglycine** [45]

$2HCN + 2H_2O + 2H \xrightarrow[diglycine]{} glycine + NH_3$ , where $glycine = OH - CO - CH_2 - NH_2$ and

$diglycine = OH - CO - CH_2 - NH - CO - CH_2 - NH_2$

We formulate the substrate-enzyme interaction pattern as follows

```
OH-H   HCN   HH   OH-H HCN
OH-CO HCH   NH     CO   HCH   NH2
```

substrate $2HCN + 2H_2O + 2H$

enzyme diglycine

with substrate bond energies in kJ/mol (H1HO is HH bond in water)

$E_n$= {H1HO,H1HO,CH,C3N,HH,H1HO,H1HO,CH,C3N}

```
{138.548, 138.548, 190.452, 735.774, 940.299, 138.548, 138.548, 190.452, 735.774}
```

enzyme bond energies in kJ/mol

$E'_n$= {C2O,C2O,CH,CH,NH,C2O,C2O,CH,CH,NH,NH}

```
{61.3247, 61.3247, 190.452, 190.452, 410.903, 61.3247, 61.3247, 190.452, 190.452, 410.903, 410.903}
```

enzyme density values $r_n = E'_n / \sum_n E'_n$ ,

the molecular energies are: bond sum $E_b = 3346.9$ , mean bond energy $E_{bm}=E_b/N = 371.9$ *kJ/mol=3.61 eV* for the substrate (as *100kJ/mol=1.03eV*)

bond sum $E'_b = 2239.8$, mean bond energy $E'_{bm}=E'_b/N = 203.6$ for the enzyme

The results are: $p_n$=

```
{0.000252084, 0.000252165, 0.000782118, 0.000782124, 0.00169157, 0.000250671, 0.994426, 0.000782172, 0.000782115}
```

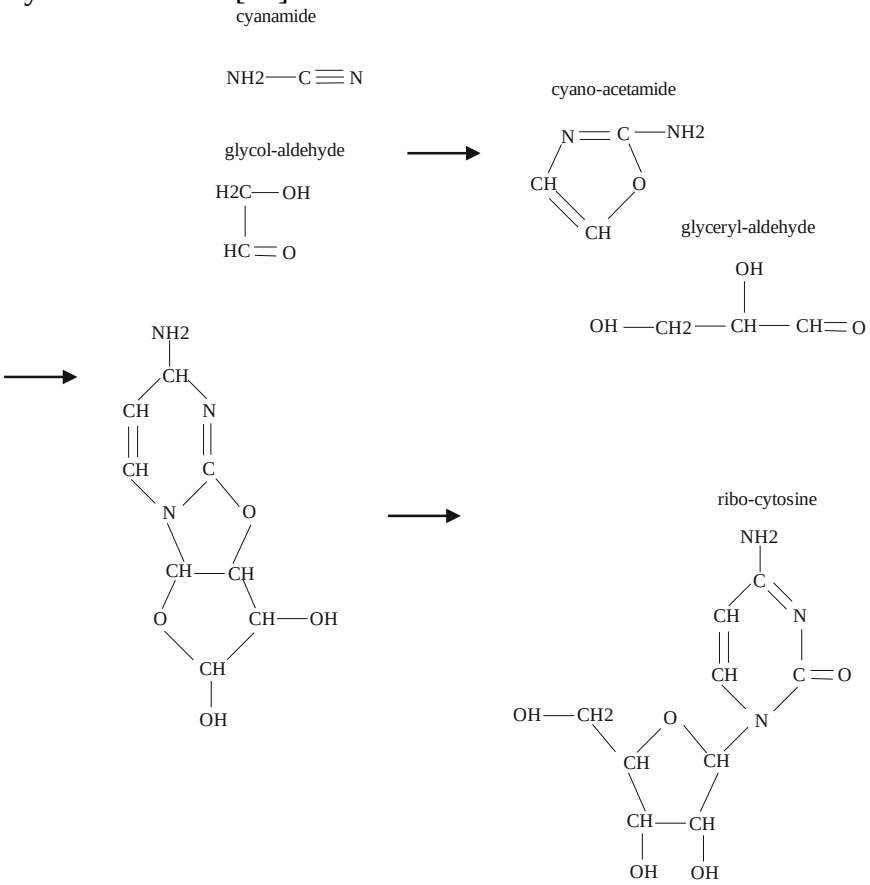
and the catalyzed mean bond energy $\sum_n p_n E_n / N = E_{bmc} = 15.66$ *kJ/mol=0.152 eV* , so the energy factor

$f_c = E_{bm}/E_{bmc} = 23.7$ , which is a very strong energy reduction.

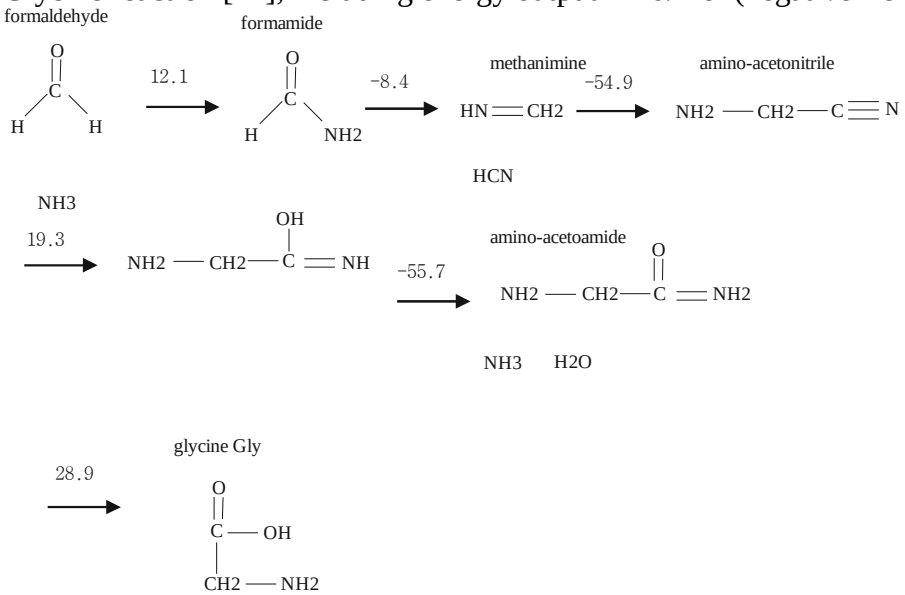
### 3.5 Synthesis reactions in prebiotic HCN-chemistry

We present here three prebiotic reactions of the HCN-chemistry, which generate important amino acids and nucleobases of the proto-life scenario described in the following section. These reactions has been carried out in the lab by Patel et al. [35] and Das et al. [41] .

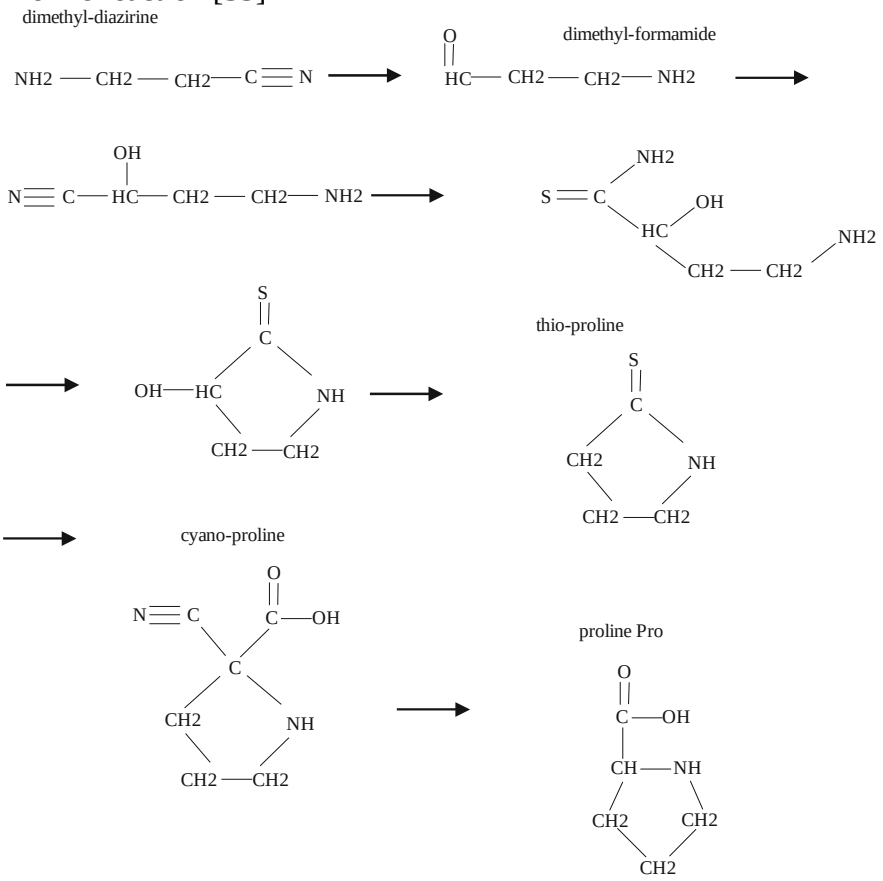
Cytosine reaction [35]




Glycine reaction [41], including energy output in kJ/mol (negative=released)



Proline reaction [35]



### 3.6 Models of proto-code

We present here a plausible scenario of terrestrial proto-life based on the HCN-chemistry according to the Patel-Das model [35, 41], the one-bit genetic proto-code proposed by Carter-Wills [29] and Rodriguez et al [30], and the pyrite (iron-sulfur) energy cycle proposed by Wächtershäuser [38].

There are several basic features of the terrestrial life chemistry, which support this scenario
-the basic components, amino acids and nucleobases, have roughly the composition scheme $C_xN_xH_{2x}O_2$ , which corresponds to the precursor input *x(HCN)+2H2O+xH* of the HCN chemistry with hydrogen provided by the energy cycle.
-the maximum activation energy $E_a$ of the spontaneous HCN-chemistry is *40kJ/mol* [35], which fits very well to the reaction energy *ΔE=41kJ/mol* of the iron-sulfur-cycle, the iron-sulfur-cycle and the reverse acetogenesis, both involving sulfur, are the two main non-photogenetic energy cycles (apart from the Wood-Ljungdahl cycle) used in the terrestrial life chemistry.
-both basic component families can be traced back to fundamental components.
For the amino acids it is the special group C of amino acids, set apart from the remaining groups A (electrically charged side chain), B ( polar uncharged side chain) and D (hydrophobic side chain). The special group C consists of the simplest linear amino acid glycine, the simplest ring-amino-acid proline with its penta-ring and the simplest sulfur-amino acid cystein: these are the 3 components of the model1-proto-code.
For the nucleobases it is cytosine, the hexa-ring nucleobase, and the coupled nucleobase guanine, which is a double hexa-penta-ring molecule, these are the 2 nucleobases of the model1 proto-code.

The model2 proto-code consists of 1-bit-codons (guanine Gua or cytosine Cyt) coding for 2 amino acids (glycine Gly or proline Pro). The enzymes are Gly-Pro-sequences, which catalyze the synthesis of the 8 needed compounds (5 amino acids Gly, Pro, Leu, His, Cys, 2 nucleobases Gua, Cyt, 1 phospholipid glycerol-1-phosphate) from the precursors (hydrogen cyanide HCN, $H_2O$, H, hydrogen sulfide $H_2S$, phosphoric acid $H_3PO_4$).
The model1 proto-code is a simpler version , where the the 2 additional amino acids Leu and His in the PNA are replaced by the enzyme-building amino acids Gly and Pro.
This is *a minimalistic version* of a proto-code: at least 2 amino acids are needed as elements of enzymes (here Gly, Pro), a third (here Cys) containing sulfur is needed for catalyzing the energy cycle and stabilizing the lipid membrane.

Glycine is the simplest amino acid and is a linear molecule with 2 carbon-nodes (COOH-head and C-node with a NH2-radical), the corresponding nucleobase guanine is a double penta-hexa-ring with 5 C- and 4 N-nodes.
Proline consists of the COOH-head and a penta-ring with 4 C-nodes an one N-node, the corresponding nucleobase cytosine is a hexa-ring with 4 C-nodes and 2 N-nodes .
Therefore, it is obvious that enzymes, which catalyze linear molecules like glycine must contain glycine, and those which catalyze ring-molecules like proline or the nucleobases must contain proline.
During catalysis, the precursors align along the enzyme, so the enzyme must have at least as many nodes (C, N) as the resulting compound. The precursors for the ring-parts of the compound align along the ring-part of the enzyme. The precursor sequence is contiguous.
The alignment rules are:
- HCN binds to HCH or HCN
- water OH-H binds to CO
- HH resp. 2HH binds to NH or OH or (if none available) to CO

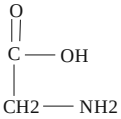
#### 3.6.1 Model2 proto-code with 5 amino acids

According to the above rules, we set up the **model 2 proto-code** with 7 proto-genes for 8 compounds (enthalpy values from [47]) and 2 PNA's for 5 amino acids, 1 lipid and 2 nucleobases .

**glycine gene** GlyGly, coded GG, produces glycine Gly
synthesis reaction
$$(HCN + 2H_2O) + (HCN + 2H) \rightarrow Gly + NH_3$$

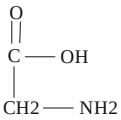
molecule enthalpy $H_f(Gly)$= *-2435.5 kJ/mol*
enzyme-substrate alignment (NH-CO is the peptide bond between amino acids)

| | |
|---|---|
| *OH-H HCN HH OH-H HCN* | *precursors* |
| *OHCO HCH NH CO HCH NH2* | *GlyGly* |



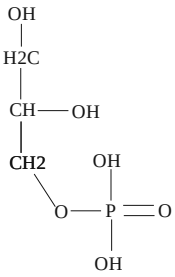
**cysteine gene** GlyGlyGly, coded GGG, produces cysteine Cys (catalyzes pyrite reaction)
synthesis reaction
$(HCN + 2H_2O) + (HCN + 5H) + (HCN + H_2S) \rightarrow Cys + 2NH_3$
molecule enthalpy $H_f(Cys)$= *-2780.1kJ/mol*
enzyme-substrate alignment

| | |
|---|---|
| *OH-H HCN HH OH-H HCN HH SH-H HCN H* | *precursors* |
| *OHCO HCH NH CO HCH NH CO HCH NH2* | *GlyGlyGly* |



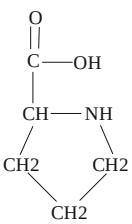
**cysteine gene** GlyGlyGly, coded GGG, produces glycerol-1-phosphate with phosphorous acid (Glyc1Ph forms membranes)
synthesis reaction
$3HCN + 2H_2O + 8H + H_3PO_4 \rightarrow Glyc1Ph + 3NH_3$
molecule enthalpy $H_f(Glyc1Ph)$= *-5033.8kJ/mol*
enzyme-substrate alignment

| | |
|---|---|
| *HH OH-H HCN HH OH-H HCN HH HPO4 HCN 2HH* | *precursors* |
| *OH CO HCH NH CO HCH NH CO HCH NH2* | *GlyGlyGly* |



**proline gene** ProPro, coded CC, produces proline Pro
synthesis reaction
$(HCN + 2H_2O) + 4HCN + 12H \rightarrow Pro + 4NH_3$
molecule enthalpy $H_f(Pro)$= *-3275.2kJ/mol*
enzyme-substrate alignment (N-CO is the peptide bond between amino acids, the penta-ring is split at CH-CH2)

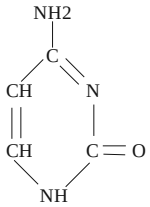
| | |
|---|---|
| *HH OH-H HH HH HCN HCN HCN* | *precursors* |
| *OH CO CH N HCH HCH HCH* | *Pro* |
| *CO CH NH HCH HCH HCH* | *Pro* |
| *OH-H HH HH HCN HCN HH* | *precursors* |



**cytosine gene** Pro, coded C, produces the nucleobase cytosine Cyt
synthesis reaction
$4HCN + H_2O + 2H \rightarrow Cyt + NH_3$
molecule enthalpy $H_f(Cyt) = -2799.kJ/mol$
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2, marked ---)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *OH-H* | *HCN* | *HH* | *HCN* | *HCN* | *HCN* | *precursors* |
| *OH* | *CO* | *CH* | *NH* | *HCH* | *HCH* | *HCH* | *Pro* |

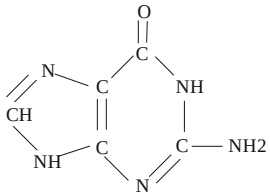
---




**guanine gene** GlyPro, coded GC, produces the nucleobase guanine Gua
synthesis reaction
$5HCN + H_2O \rightarrow Gua + H_2$
molecule enthalpy $H_f(Gua) = -3753.3kJ/mol$
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2, marked ---)

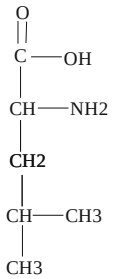
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *OH-H* | *HCN* | *OH-H* | *HCN* | *HCN* | *HCN* | *HCN* | *precursors* |
| *OH* | *CO* | *HCH* | *NH* | *CO* | *CH* | *NH* | *HCH* | *HCH* | *HCH* | *GlyPro* |

---




**leucine gene** GlyGlyGlyGly, coded GGGG, produces leucine Leu
synthesis reaction
$(HCN + 2H_2O) + 5HCN + 18H \rightarrow Leu + 5NH_3$
molecule enthalpy $H_f(Leu) = -4181.7kJ/mol$
enzyme-substrate alignment

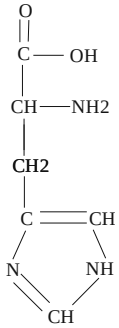
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *HH* | *OH-H* | *HCN* | *2HH* | *OH-H* | *HCN* | *2HH* | *HCN* | *HCN* | *2HH* | *HCN* | *HCN* | *2HH* | *precursors* |
| *OH* | *CO* | *HCH* | *NH* | *CO* | *HCH* | *NH* | *CO* | *HCH* | *NH* | *CO* | *HCH* | *NH2* | *GlyGlyGlyGly* |



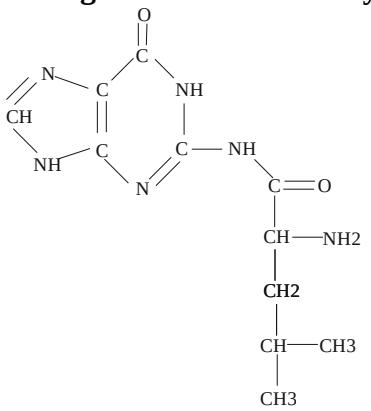

**histidine gene** ProGlyGly, coded CGG, produces histidine His
synthesis reaction
$(HCN + 2H_2O) + 5HCN + 8H \rightarrow His + 3NH_3$
molecule enthalpy $H_f(His) = -4259.7kJ/mol$
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *2HH* | *OH-H* | *HCN* | *HCN* | *HCN* | *HCN* | *precursors* |
| *OH* | *CO* | *CH* | *N* | *HCH* | *HCH* | *HCH* | *Pro* |

*CO HCH NH CO HCH NH2 GlyGly*
*2HH HCN 2HH OH-H HCN 2HH precursors*

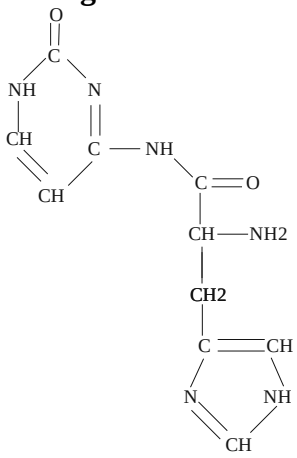


The **ligase and PNA for Gly** is Leu-Gua bound by the peptide bond *–NH-C=O-*



$$Leu + Gua \rightarrow LeuGua + H_2O$$

The **ligase and PNA for Pro** is His-Cyt bound by the peptide bond *–NH-C=O-*



$$His + Cyt \rightarrow HisCyt + H_2O$$

The peptide bond, which forms poly-peptides, has a bond energy of 8–16 kJ/mol .
The hexa-rings of the PNA's are stacked on each other by H-bonds, and the amino acids do not form peptide bonds, because they are not aligned properly, they form H-bonds: the resulting structure is a poly-PNA built from Leu-Gua and His-Cyt. During the synthesis or the replication process however, the adjoining acid forms a peptide-bond with its twin, and at the end the peptide bond breaks, because it is weaker than the several H-bonds between stacked amino acids of the PNA-chain: the complete enzyme (resp. PNA-copy) breaks off.
The model2-version poly-PNA from Leu-Gua and His-Cyt elements is more stable than the model1-version from Gly-Gua and Pro-Cyt, its H-bonds are stronger, because the active linear part of the amino acid is twice as long: this is an advantage of the 5-gene-code. In the simple PNA formation reactions of model1 (see reaction table below),
*Gly+Gua→GlyGua+H2O , ΔE=339kJ/mol , $E_a$=14kJ/mol*
*Pro+Cyt→ProCyt+H2O , ΔE=-44.3kJ/mol , $E_a$=12.6kJ/mol*
the Pro-reaction is endothermic (needs energy).
In the corresponding model2-reactions

*Leu+Gua→LeuGua+H2O , ΔE=339kJ/mol , $E_a$=10kJ/mol*
*His+Cyt→HisCyt+H2O , ΔE=109.4kJ/mol , $E_a$=11.9kJ/mol*
the His-reaction is exothermic, and both activation energy are lower than with their model1-counterparts.

### 3.6.2 Simplified model1 proto-code with 3 amino acids

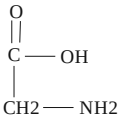
One gets a simpler version of the proto-code if the the 2 additional amino acids Leu and His in the PNA are replaced by the enzyme-building amino acids Gly and Pro, as described above .

The proto-code consists now of 1-bit-codons (guanine Gua or cytosine Cyt) coding for 2 amino acids (glycine Gly or proline Pro). The enzymes are Gly-Pro-sequences, which catalyze the synthesis of the 6 needed compounds (3 amino acids Gly, Pro, Cys, 2 nucleobases Gua, Cyt, 1 phospholipid glycerol-1-phosphate) from the precursors (hydrogen cyanide HCN, $H_2O$, H, hydrogen sulfide $H_2S$, phosphoric acid $H_3PO_4$).

The simplified **model1 proto-code** with 5 proto-genes for 6 compounds (enthalpy values from [47]) and 2 PNA's for 3 amino acids, 1 lipid and 2 nucleobases is now as follows.

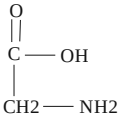
**glycine gene** enzyme GlyGly, coded GG, produces glycine Gly
synthesis reaction
$(HCN + 2H_2O) + (HCN + 2H) \rightarrow Gly + NH_3$
molecule enthalpy *ΔH(Gly)= -390.5 kJ/mol* (negative=energy released in synthesis)
enzyme-substrate alignment (NH-CO is the peptide bond between amino acids)

| | |
|---|---|
| *OH-H HCN HH OH-H HCN* | *precursors* |
| *OHCO HCH NH CO HCH NH2* | *GlyGly* |



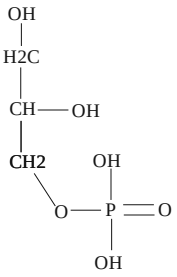

**cysteine gene** enzyme GlyGlyGly, coded GGG, produces cysteine Cys (catalyzes pyrite reaction)
synthesis reaction
$(HCN + 2H_2O) + (HCN + 4H) + (HCN + H_2S) \rightarrow Cys + 2NH_3$
molecule enthalpy *ΔH(Cys)= -534.1kJ/mol*
enzyme-substrate alignment

| | |
|---|---|
| *OH-H HCN HH OH-H HCN HH SH-H HCN H* | *precursors* |
| *OHCO HCH NH CO HCH NH CO HCH NH2* | *GlyGlyGly* |



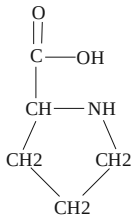

**cysteine gene** enzyme GlyGlyGly, coded GGG, produces glycerol-1-phosphate with phosphorous acid (Glyc1Ph forms membranes)
synthesis reaction
$3HCN + 2H_2O + 8H + H_3PO_4 \rightarrow Glyc1Ph + 3NH_3$
molecule enthalpy *ΔH(Glycerol)= -577.9kJ/mol*
enzyme-substrate alignment

| | |
|---|---|
| *HH OH-H HCN HH OH-H HCN HH HPO4 HCN 2HH* | *precursors* |
| *OH CO HCH NH CO HCH NH CO HCH NH2* | *GlyGlyGly* |



**proline gene** enzyme ProPro, coded CC, produces proline Pro
synthesis reaction
$(HCN + 2H_2O) + 4HCN + 12H \rightarrow Pro + 4NH_3$
molecule enthalpy *ΔH(Pro)= -366.2kJ/mol*
enzyme-substrate alignment (N-CO is the peptide bond between amino acids, the penta-ring is split at CH-CH2)

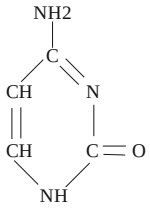
| | |
|---|---|
| *HH OH-H HH HH HCN HCN HCN* | *precursors* |
| *OH CO CH N HCH HCH HCH* | *Pro* |
| *CO CH NH HCH HCH HCH* | *Pro* |
| *OH-H HH HH HCN HCN HH* | *precursors* |




**cytosine gene** enzyme Pro, coded C, produces the nucleobase cytosine Cyt
synthesis reaction
$4HCN + H_2O + 2H \rightarrow Cyt + NH_3$
molecule enthalpy *ΔH(Cyt)= -221.3kJ/mol*
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2, marked ---)

| | |
|---|---|
| *OH-H HCN HH HCN HCN HCN* | *precursors* |
| *OH CO CH NH HCH HCH HCH* | *Pro* |

---




**guanine gene** enzyme GlyPro, coded GC, produces the nucleobase guanine Gua (alignment and catalyzing not so good as ProGly)
synthesis reaction
$5HCN + H_2O \rightarrow Gua$
molecule enthalpy *ΔH(Gua)= -183.9kJ/mol*
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2, marked ---)

| | |
|---|---|
| *HCN OH-H HCN HCN HCN HCN* | *precursors* |
| *OH CO HCH NH CO CH NH HCH HCH HCH* | *GlyPro* |

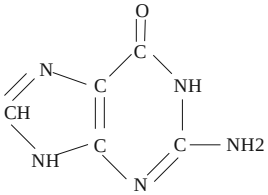
---

**guanine gene** alternative enzyme ProGly, coded CG, produces the nucleobase guanine Gua
synthesis reaction
$5HCN + H_2O \rightarrow Gua$
molecule enthalpy *ΔH(Gua)= -183.9kJ/mol*
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH2, marked ---)

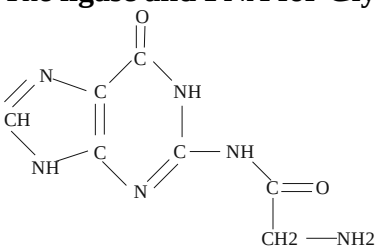
| | |
|---|---|
| *HCN HCN HCN HCN* | *precursors* |
| *CH N HCH HCH HCH* | *ProGly* |

---

| | |
|---|---|
| *CO HCH HNH* | *precursors* |
| *OH-H HCN* | *ProGly* |

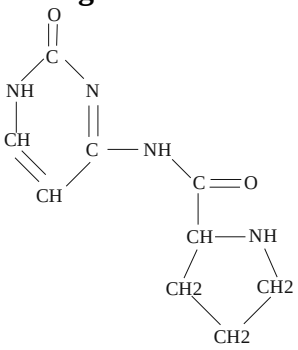


The **ligase and PNA for Gly** is GlyGua bound by the peptide bond *–NH-C=O-*



$$Gly + Gua \rightarrow GlyGua + H_2O$$

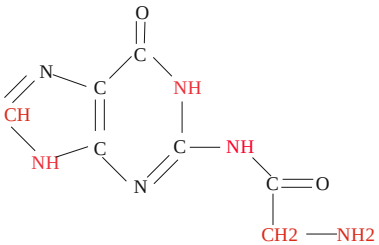
The **ligase and PNA for Pro** is ProCyt bound by the peptide bond *–NH-C=O-*



$$Pro + Cyt \rightarrow ProCyt + H_2O$$

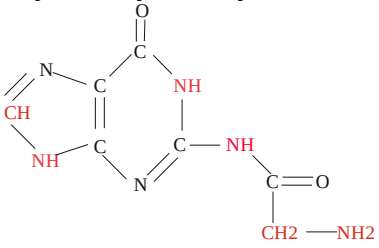
The peptide bond, which forms poly-peptides, has a bond energy of 8–16 kJ/mol .
The hexa-rings and the penta-rings of the PNA's are stacked on each other by H-bonds, and the amino acids do not form peptide bonds, because they are not aligned properly, they form H-bonds: the resulting structure is a poly-PNA built from Gly-Gua and Pro-Cyt.
During the synthesis or the replication process however, the adjoining acid forms a peptide-bond with its twin, and at the end the PNA peptide bond and the vertical H-bonds break, with energy input: the complete enzyme (resp. PNA-copy) breaks off.

During **polymerization of PNA** H-bonds form in those sites, which have H-atoms.
GlyGuaGlyGua has 6 vertical H-bonds (in red)



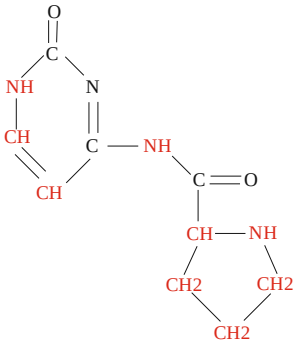
$$2GlyGua \rightarrow GlyGuaGlyGua + 6(H\text{-})$$

GlyGuaGlyGuaGlyGua has 12 vertical H-bonds (in red)



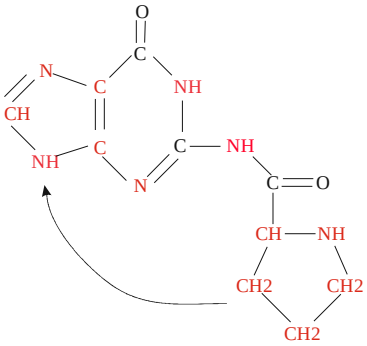
*3 GlyGua → GlyGuaGlyGuaGlyGlua + 12(H-)*

ProCytProCyt has 9 vertical H-bonds (in red)



*2ProCyt → ProCytProCyt + 9(H-)*

GlyGuaProCyt has 13 vertical H-bonds (in red), the arrow marks the alignment of the penta-rings



*GlyGua + ProCyt → GlyGuaProCyt + 13(H-)*

The **poly-peptide (enzyme) formation** works by breaking-off of the amino-acid-stack from the PNA-stack under energy input *ΔE*.
During enzyme formation the peptide bond *pb(PNA)* breaks, as well as some of the H-bonds, and *pb(peptide)* is formed instead.
*GlyGuaGlyGua + ΔE → GlyGly + GuaGua + pb(GlyGly) – 2 pb(PNA) - 3(H-)*
*ProCytProCyt + ΔE → ProPro + CytCyt + pb(ProPro) – 2 pb(PNA) - 6(H-)*
*GlyGuaProCyt + ΔE → GlyPro + GuaCyt + pb(GlyPro) – 2 pb(PNA) - 9(H-)*
*GlyGuaGlyGuaGlyGua + ΔE → GlyGlyGly + GuaGuaGua + 2 pb(GlyGly) – 3 pb(PNA) - 6(H-)*

### 3.7 Reactions of model1 proto-code

The reactions of the model1-proto-code fall into 9 categories

1 Catalyzed synthesis of basic compounds and energy reaction (pyrite reaction)
The basic compounds (amino acids and nucleobases) are synthesized from precursors HCN, H2O, H2, H2S, catalyzed by enzymes controlled by the corresponding genes.
Example: (*ΔE* reaction energy, $E_a$ original activation energy, $E_c$ catalyzed activation energy, $n_B$ number of bonds)
Direct glycine synthesis catalyzed by *GlyGly* :
*2HCN+2H2O+H2→Gly(COOH−CH2NH2)+NH3* , enzyme *GlyGly*, $\Delta E=223kJ/mol$ , $E_a=94.8kJ/mol$, $E_c=7.6kJ/mol$ , $n_B=25$
A key role plays the pyrite reaction, as energy provider, but foremost as the source of hydrogen for the synthesis of basic compounds:
*FeS+H2S→FeS2+H2* , enzyme *Cys*, $\Delta E=41kJ/mol$ , $E_a=160.6kJ/mol$, $E_c=0.02kJ/mol$ , $n_B=16$

The pyrite reaction has a low energy output, but a high activation energy, so it runs practically only with catalysis (uncatalyzed reaction time is $t_r = t_0\, exp\left(\frac{E_a}{kT}\right) = 3.45*10^{-9}\, 1.37*10^{21} = 5.32*10^{12}\, s = 1.69*10^{5}\, a$ ).
The synthesis reactions have high reaction energy (*ΔE=100...1200kJ/mol*), high activation energy ($E_a$*=100...200kJ/mol*), and do not run in water without enzymes.
The "self-catalyzing" synthesis of proline via cytosine(ring) has a very low activation energy
*Cyt(C4H5N3O3)+HCN+6H2→Pro(C5H9NO2)+H2O+2NH3 , ΔE=582.9kJ/mol , $E_a$=21.7kJ/mol, $n_B$=16*
The basic compounds synthesis reactions are all exothermic, and have high activation energy $E_a$*=92...145kJ/mol* (using cytosine-proline synthesis). when running without catalysis , the maximum reaction time is (*Gua*) $t_r = t_0\, exp\left(\frac{E_a}{kT}\right) = 0.99*10^{-9}\, 1.59*10^{19} = 1.57*10^{10}\, s = 0.50*10^{3}\, a$
Apart from amino acid and nucleobase synthesis, the third important basic compounds are the phospholipids, which form the proto-membrane, here the synthesis of the lipid precursor glycerol-1-phosphate *Glyc1ph* from the basic precursor phosphoric acid *H3PO4* , catalyzed by triglycine *GlyGlyGly*
*3HCN+2H2O+4H2+H3PO4→Glyc1ph+3NH3 , ΔE=395.5kJ/mol , $E_a$=139.5kJ/mol, $E_c$=6.6kJ/mol , $n_B$=27*

2 Nucleobase polymerization
Nucleobases form stacks bound by H-bonds (poly-nucleobase=PNB), these reactions are spontaneous, strongly exothermic (*ΔE~600kJ/mol*) and have low activation energy ($E_a$*=5..10kJ/mol*).
*Cyt+Cyt→CytCyt(3H−)  CytCyt* is bound by 3 H-bonds, *ΔE=491kJ/mol , $E_a$=5kJ/mol*

3 Simple PNA formation
Simple peptide-nucleic-acid (PNA) form from an amino acid and a nucleobase
*Gly+Gua→GlyGua+H2O , ΔE=339kJ/mol , $E_a$=14kJ/mol*
*Pro+Cyt→ProCyt+H2O , ΔE=-44.3kJ/mol , $E_a$=12.6kJ/mol*
These reactions are low exo- or low endothermic and have low activation energy.

4. Direct PNA polymerization
Simple PNA's form chains
*GlyGua+GlyGua→GlyGuaGlyGua(6H−)  , ΔE=126kJ/mol , $E_a$=3kJ/mol*

5 PNA formation from amino acids and nucleobases
Two amino acids and a simple pNB form a PNA
*Gly+Gly+GuaGua→GlyGuaGlyGua(+2H−)+2H2O, ΔE=468kJ/mol , $E_a$=10kJ/mol*
These reactions are mostly endothermic and have a low activation energy.

6 Peptide forming
*GlyGuaGlyGua+2H2O→GuaGua(+4H−)+GlyGly +H2O, ΔE=682kJ/mol , $E_a$=11.3kJ/mol*
A peptide "breaks-off" from a PNA-stack, the reactions are exothermic and have low activation energy.
The sequence type2→type5→ type6 is the PNA-controlled peptide synthesis, in contrast to type9, the spontaneous peptide synthesis (see below). This sequence is exothermic (the endothermic reactions type5 use the energy from the preceding reaction type2), wheres the reactions of type9 are endothermic, and their activation energy $E_a$ is mostly lower than the corresponding $E_a$ in type 9 .

7 Diverse synthesis and energy cycles
The HCN chemistry as the basis for the proto-code reactions depends on a sufficient supply of HCN. As discussed in 2.4, there is evidence of presence of HCN, along methane CH4 and ammonia NH3 in the early Earth atmosphere. But there are also cyanide synthesis reactions, which could produce HCN under volcanic pools or in hydrothermal vents.
*CO2+NH3+H2→HCN+2H2O , ΔE=-90.9kJ/mol , $E_a$=74.2kJ/mol*
*CO2+NH3+SH2→HCN+2H2O +S, ΔE=-111.5kJ/mol , $E_a$=6.5kJ/mol*
They have as reducing component *H2* or *SH2*, which are supplied by the pyrite reaction.
They are both endothermic, so they depend upon energy supply, e.g. from the pyrite reaction.

Their activation energy $E_a$ is relatively low, especially for the SH2-reaction: this one can run purely thermally, without enzymes.
There is the synthesis of the proto-sugar glyceraldehyde *Gle*, a precursor of ribose needed for RNA :
*Gly+H2O→Gle(HCO-HCOH-H2COH)+NH3, ΔE=346.2kJ/mol , $E_a$=14kJ/mol*
As for energy cycles, there is the mentioned pyrite reaction
*FeS+H2S→FeS2+H2 , ΔE=41kJ/mol , $E_a$=160.6kJ/mol*
but also the CO2-reduction by hydrogen, which was present in the prebiotic Earth atmosphere
*CO2+H2→CO+H2O, ΔE=90.1kJ/mol , $E_a$=115.8kJ/mol*
and oxidation of carbon monoxide *CO* in water, which is still used by methanogen bacteria [50]
*CO+H2O→HCOOH, ΔE=548.8kJ/mol , $E_a$=33.3kJ/mol*

8 Decay of basic compounds
Under present biotic conditions on Earth, the basic compounds are long-lived, amino acids decompose in water thermally at temperature T>185°C at neutral pH, nucleobases have in water a half-life of 20-200 days at 100°C, and they degrade at 250C at neutral pH [59].
The decay via hydrolysis of the basic compounds functions via break-off of the C=O radical, which all of them contain, with formation of formic acid. All decays are endothermic (run only with energy input), and have a low activation energy.
A typical decay reaction is the hydrolysis of glycine:
*Gly(COOH−CH2NH2)+H2O→OH−CH2NH2+HCOOH, ΔE=-335.9kJ/mol , $E_a$=9.0kJ/mol*

9 Direct peptide polymerization
These are the direct merging reactions of amino acids into peptides, without the interaction of PNA's (see peptide formation):
*Gly+Gly→GlyGly+H2O , ΔE=-117.7kJ/mol , $E_a$=64.8kJ/mol*
*Pro+Pro→ProPro+H2O , ΔE=-1097.7kJ/mol , $E_a$=27.9kJ/mol*
These reactions are endothermic, with moderate activation energy, they run spontaneously only at high temperature and with a multiple energy input from the energy reaction
(*ΔE(Gly+Gly→GlyGly+H2O)≈3 E(FeS+H2S→FeS2+H2)=123kJ/mol*), i.e. very slowly even with a catalyzed energy reaction.
The opposite reactions like
*GlyGly+H2O→Gly+Gly*
are exothermic *decay reactions of peptides* with a decomposition temperature $T_d$ , where
$t_0 Exp(E_a / kT_d) = 1s$ , so for $t_0 \approx 10^{-9} s$ we get approximately for *Gly* :
$kT_d = E_a / Log(10^9)$ and $T_d = 350K = 77°C$ , which is in good agreement with the known decomposition temperature of poly-peptides (75-85°C).

Therefore the PNA-controlled peptide formation sequence type2→type5→ type6 is really necessary for the molecular evolution of life from the HCN-chemistry.
On the other hand, in the prebiotic chemistry on Earth there was a seed concentration of amino acids from meteorites (rough estimation for the concentration $c=10^{-6}$ ), but no polymers, i.e. no enzymes, and therefore no organic catalysis. In the initial prebiotic period before the onset of the model1 proto-code life cycle, peptide polymerization was spontaneous, with the support of an energy cycle: *Cys*-catalyzed pyrite reaction of *H2S* , *CO2*-reduction or *CO*-oxidation .
With seed-concentration for the basic compounds of $c=10^{-6}$ , and remaining concentration from [50] given below , we get for the energy production of the 3 energy cycles [45]
*$E_p$(H2S)= 0.0045kJ/s mol , $E_p$(CO2)= 5.3 $10^{-7}$kJ/s mol , $E_p$(CO)= 0.00062kJ/s mol* ,
and using this, for the effective build-up time of *Gly+Gly→GlyGly+H2O* with the *H2S*-pyrite reaction
$t_c(2Gly) = \frac{t_r(2Gly)}{c(Gly)} \frac{\Delta E(2Gly)}{t_r(2Gly)} \frac{1}{E_p} = 1.0\ 10^9 s = 31.7a$ with the effective reaction time $t_r = t_0\, exp\left(\frac{E_a}{kT}\right)$

**Table: Reactions of model1 proto-code**

($\Delta E$ reaction energy ($kJ/mol$), $E_a$ original activation energy ($kJ/mol$) , $E_c$ catalyzed activation energy ($kJ/mol$), $t_0$ ($ns$), $n_B$ number of bonds)

| reaction | type | enzyme | ΔE | $E_a$ | $E_c$ | $t_0$ | $n_B$ |
|---|---|---|---|---|---|---|---|
| catalyzed basic compounds synthesis | | | | | | | |
| 2HCN+2H2O+H2→Gly(COOH−CH2NH2)+NH3 | 1 | GlyGly | 223 | 92.8 | 8.3 | 0.70 | 25 |
| 5HCN+2H2O+6H2→Pro(C5H9NO2)+4NH3 | 1 | ProPro | 741.7 | 256.1 | 0.53 | 0.47 | 53 |
| 3HCN+2H2O+2H2+H2S→ Cys(COOH−CHNH2−CH2−SH)+2NH3 | 1 | GlyGlyGly | 1232.4 | 144.3 | 0.94 | 0.74 | 39 |
| 4HCN+H2O+H2→Cyt(C4H5N3O)+NH3 | 1 | Pro | 403.5 | 140.2. | 17.5 | 0.76 | 28 |
| 5HCN+H2O→Gua(C5H5N5O)+H2 | 1 | GlyPro | 505. | 145.9 | 6.56 | 0.99 | 36 |
| 5HCN+H2O→Gua(C5H5N5O)+H2 | 1 | ProGly | 505. | 145.9 | 27.6 | 0.99 | 36 |
| FeS+H2S→FeS2+H2 | 1 | Cys | 41. | 160.6 | 0.02 | 3.45 | 16 |
| Cyt(C4H5N3O3)+HCN+6H2→Pro(C5H9NO2) +H2O+2NH3 | 1 | | 582.9 | 21.7 | | 1.02 | 16 |
| 3HCN+2H2O+4H2+H3PO4→Glyc1ph+3NH3 | 1 | GlyGlyGly | 395.5 | 139.5 | 6.6 | 0.99 | 27 |
| simple PNA polymerization | | | | | | | |
| Gly+Gly+GuaGua→GlyGuaGlyGua(6H−)+2H2O | 5 | | -94.2 | 16.4 | | 1.53 | 53 |
| Gly+Gly+Gly+Gua3→GlyGua3(12H−)+3H2O | 5 | | -109.8 | 16.5 | | 1.60 | 81 |
| Pro+Pro+CytCyt→ProCytProCyt(9H−)+2H2O | 5 | | 692.8 | 5.9 | | 2.26 | 63 |
| Gly+Pro+GuaCyt→GlyGuaProCyt(13H−)+2H2O | 5 | | -146.5 | 15.1 | | 1.89 | 61 |
| Gly+Pro+CytGua→ProCytGlyGua(13H−)+2H2O | 5 | | -146.5 | 15.1 | | 1.89 | 61 |
| Gly+Pro+GuaCyt→GlyCytProGua(13H−)+2H2O | 5 | | -146.5 | 15.1 | | 1.89 | 61 |
| PNB polymerization | | | | | | | |
| Cyt+Cyt→CytCyt(3H−) | 2 | | 491. | 5.0 | | 1.31 | 26 |
| Gua+Gua→GuaGua(3H−) | 2 | | 705.3 | 2.9 | | 1.14 | 34 |
| Gua+Gua+Gua→Gua3(6H−) | 2 | | 1089.5 | 3.7 | | 1.14 | 51 |
| Gua+Cyt→GuaCyt(6H−) | 2 | | 661.2 | 6.9 | | 1.22 | 30 |
| Cyt+Gua→CytGua(6H−) | 2 | | 661.2 | 6.9 | | 1.22 | 30 |
| peptide formation | | | | | | | |
| GlyGuaGlyGua+2H2O→GuaGua(3H−)+GlyGly+H2O | 6 | | 661.6 | 12.5 | | 2.46 | 58 |
| ProCytProCyt+2H2O→CytCyt(3H−)+ProPro+H2O | 6 | | 566.1 | 7.0 | | 3.61 | 62 |
| GlyGuaProCyt+2H2O→GuaCyt(6H−)+GlyPro+H2O | 6 | | 308.6 | 9.2 | | 3.05 | 58 |
| ProCytGlyGua+2H2O→CytGua(6H−)+ProGly+H2O | 6 | | 1008. | 3.0 | | 3.05 | 58 |
| GlyGuaGlyGuaGlyGua+3H2O→GuaGuaGua(8H−) +GlyGlyGly+2H2O | 6 | | 640.6 | 8.1 | | 1.13 | 90 |
| multiple PNA polymerization | | | | | | | |
| GlyGua+GlyGua→GlyGuaGlyGua(6$H$−) | 4 | | 126. | 3.0 | | 1.66 | 48 |
| 3GlyGua→(GlyGua)3(12$H$−) | 4 | | 252. | 3.9 | | 1.66 | 72 |
| ProCyt+ProCyt→ProCytProCyt(9$H$−) | 4 | | 189. | 4.3 | | 1.66 | 48 |
| GlyGua+ProCyt→GlyGuaProCyt(13H−) | 4 | | 273. | 5.80 | | 1.66 | 48 |
| GlyGua+ProCyt→ProCytGlyGua(13H−) | 4 | | 273. | 5.80 | | 1.66 | 48 |
| PNA building | | | | | | | |
| Gly+Gua→GlyGua+H2O | 3 | | 339.4 | 14.6 | | 1.38 | 26 |
| Gly+Cyt→GlyCyt+H2O | 3 | | 1156. | 17.2 | | 1.47 | 22 |
| Pro+Cyt→ProCyt+H2O | 3 | | -44.3 | 12.6 | | 1.79 | 30 |
| Pro+Gua→ProGua+H2O | 3 | | 62.8 | 11.1 | | 1.67 | 34 |
| Leu+Gua→LeuGua+H2O | 3 | | 339. | 10. | | 1.79 | 38 |
| His+Cyt→HisCyt+H2O | 3 | | 109.4 | 11.9 | | 1.70 | 32 |
| diverse synthesis and energy cycles | | | | | | | |
| CO2+NH3+H2→HCN+2H2O | 7 | | -90.9 | 74.2 | | 3.45 | 6 |
| CO2+NH3+SH2→HCN+2H2O+$S$ | 7 | | -111.5 | 6.5 | | 3.45 | 7 |
| FeS+H2S→FeS2+H2 | 7 | | 41. | 160.6 | | 3.45 | 3 |
| CO2+H2→CO+H2O | 7 | | 90.1 | 115.8 | | 0.23 | 3 |
| CO+H2O→HCOOH | 7 | | 548.8 | 73.3 | | 0.20 | 3 |
| Gly+H2O→Gle(HCO-HCOH-H2COH)+NH3 | 7 | | 346.2 | | | 1.62 | 9 |
| basic compounds decay by hydrolysis C=O | | | | | | | |
| Gly(COOH−CH2NH2)+H2O→OH−CH2NH2+HCOOH | 8 | | -335.9 | 9.0 | | 1.66 | 11 |
| Cys(COOH−CHNH2−CH2−SH)+H2O→ OH−CHNH2−CH2−SH+HCOOH | 8 | | -157 | 6.6 | | 5.29 | 15 |
| Pro(COOH−ring(−NH−))+H2O→OH−ring()+HCOOH | 8 | | -612.4 | 5.2 | | 2.45 | 19 |
| Cyt(ring(−C=O−))+H2O→ring()+HCOOH | 8 | | -89.4 | 20.5 | | 1.30 | 15 |
| Gua(ring(−C=O−ring(−NH−CH−N−)))+H2O→ ring( −ring(−NH−CH−N−))+HCOOH | 8 | | -15.1 | 5.2 | | 1.14 | 19 |
| direct peptide polymerization | | | | | | | |
| Gly+Gly→GlyGly+H2O | 9 | | -117.7 | 64.8 | | 1.66 | 18 |
| Gly+Gly+Gly→GlyGlyGly+2H2O | 9 | | -99.5 | 83.8 | | 1.66 | 27 |
| Pro+Pro→ProPro+H2O | 9 | | -1097.7 | 27.9 | | 2.45 | 34 |
| Gly+Pro→GlyPro+H2O | 9 | | -258. | 43.7 | | 2.02 | 26 |
| Gly+Pro→ProGly+H2O | 9 | | -258. | 43.7 | | 2.02 | 26 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GlyGly+Gly→GlyGly+H2O | 9 | | +18.2 | 48. | | 1.77 | 24 |

## 3.8 Numerical simulation of terrestrial proto-life evolution

### 3.8.1 Diffusion and convection

Molecular flow pays an important part in biochemical reactions and must be taken into account in the simulation model.
In prebiotic chemistry, there are strong thermal and concentration gradients.
The purely **diffusion-driven flow** is very slow in water : the diffusivity constants are in the range of $10^{-9}$ $m^2/s$ , so in the length scale of $L = \delta_\theta \sim 10\ cm$ ( $\delta_\theta$ is the width of the thermal boundary layer, see below) , we get a time scale of $t \sim 10^7 s$ for the thermal vent scenario, which is too slow. For the lipid-bubble scenario the length scale is the bubble radius $R_m = 20\mu m$ , and the time scale is $t \sim 1s$ , here diffusion is a realistic mechanism.
For diffusion, we have Fick's law [65]

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}$$

For the gradient-driven **osmosis through a membrane** the following relation holds [66]:

$\frac{dc}{dt} = \frac{A}{d\,V} KD\,\Delta c$ , where $V$ is the volume, $D$ is the diffusion coefficient of the molecule, $K$ is the (dimensionless) partition coefficient, $A$ is the membrane area, $d$ is the membrane thickness, $c$ is the relative concentration, $\Delta c$ is the difference in relative concentrations at the membrane. $K$ depends on the molecule and on the material of the membrane, and has values (for lipids and octanol) in the range $\log K_{ow}$= -4... 6.5 [59].

The **thermal convection** flow, driven by thermal difference of ΔT~80K is much faster [64, 65].
In thermal convection, there are two opposite forces: the buoyancy force $F_b$, driven by the thermal density gradient, and the viscosity force $F_v$ driven by viscous friction.
We get for the buoyancy force density $F_b = \beta \Delta T \rho_0 g$

$Gr = \frac{F_b}{F_v} = g\beta^* \Delta c \frac{\delta_\theta^{\,3}}{\nu^2}$ , where $L = \delta_\theta$ is the width of the thermal boundary layer, here $\delta_\theta = \frac{L}{Nu}$ ,

here $Nu(L=10m, \Delta T=80K) = 160$ $\delta_\theta = 6.3cm$ [64, 65]

and $\beta^* = -\frac{1}{\rho}\frac{\partial \rho}{\partial c}$ and $\beta^* c \approx \frac{\Delta \rho}{\rho}$

then follows $F_v = \frac{F_b}{Gr} \approx \frac{\Delta \rho}{\Delta \rho / \rho}\frac{\nu^2}{\delta_\theta^{\,3}} = \rho\frac{\nu^2}{\delta_\theta^{\,3}}$

with the denominations: $\rho$ density, $\Delta T$ temperature difference, $\Delta c$ concentration difference, $L$ edge length of the reaction region, $\nu$ kinematic viscosity of water ($m^2/s$), $g$ gravitational acceleration constant, $\beta$ water volume expansion coefficient ($1/K$) .

resulting viscosity force density is $F_b - F_v = \Delta\rho g - \rho\frac{d\left(\frac{\nu^2}{x^3}\right)}{dx}\Delta x = \Delta\rho g - \rho\frac{3\nu^2}{x^4}\Delta x$

so we get the acceleration $\Delta a(x) = \frac{\Delta\rho}{\rho}g - \frac{3\nu^2}{x^4}\Delta x = \Delta x\left(\frac{\rho'(x)}{\rho(x)}g - \frac{3\nu^2}{x^4}\right)$

and the convection time $t_c = 1/\sqrt{\left(\frac{\rho'(x)}{\rho(x)}g - \frac{3\nu^2}{x^4}\right)}$ or for thermal convection

$t_c = 1/\sqrt{\left(\frac{T'(x)}{T(x)}g - \frac{3\nu^2}{x^4}\right)}$ ,

With real values: $\nu(water, T = 20°C) = 1.010^{-6}\ m^2/s$ $\beta(water, T = 20°C) = 2.1\ 10^{-4}\ K^{-1}$

$x = \delta_\theta = 6.3cm$ , $\Delta T = 80K$, $t_c = 1/\sqrt{\left(\frac{80K}{300K}\frac{9.81}{1} - \frac{3*1.0\,10^{-12}}{(6.3\,10^{-2})^4}\right)} = 0.61s$ and $F_v << F_b$ , the viscosity force is negligible against the buoyancy force component, and the convection velocity $v_c$ becomes

$$v_c = \left(\frac{dt_c}{dx}\right)^{-1} = \frac{2\sqrt{\left(\frac{T'(x)}{T(x)}g\right)^3}}{g\left(\left(\frac{T'(x)}{T(x)}\right)^2 - \frac{T''(x)}{T(x)}\right)}$$ , with the given data, $v_c$= *1.07 m/s*

If we have a concentration profile *c(x)*, then the thermal convection driven concentration flow will be

$$\frac{\partial c(t,x)}{\partial t} = -\frac{\partial c(t,x)}{\partial x} / \left(\frac{dt_c}{dx}\right) = -\frac{\partial c(t,x)}{\partial x} v_c$$

### 3.8.2 Scenario1: hydrothermal vent with spontaneous synthesis of basic components

As was outlined in 2.4, the most plausible scenario for the origin of the first life-cycle are hydrothermal vents in submarine volcanic rocks or volcanic pools. We carried out calculation based on this scenario, and on the reaction table in 3.7 .

In this calculation, the reaction is described by a differential equation for the corresponding law of mass action according to the scheme

$\frac{\partial c(t,x)}{\partial t} = k\, c_1(t,x)^{k_1} c_2(t,x)^{k_2} ... c_n(t,x)^{k_n}$ , where $c_i$ are the concentrations of the reaction participants with multiplicities $k_i$ , and with the reaction constant

$k = \frac{\exp(-E_a / kT)}{t_0}$ , in time $t$ and location $x$ .

Furthermore, we have terms for three possible transport mechanisms (see 3.8.1)

diffusion $\frac{\partial c(x,t)}{\partial t} = D \frac{\partial^2 c(x,t)}{\partial x^2}$

membrane osmosis $\frac{\partial c(t,x_1)}{\partial t} = \frac{A}{d\,V} KD\,(c(t,x_1) - c_0)$ the boundary $x_1$

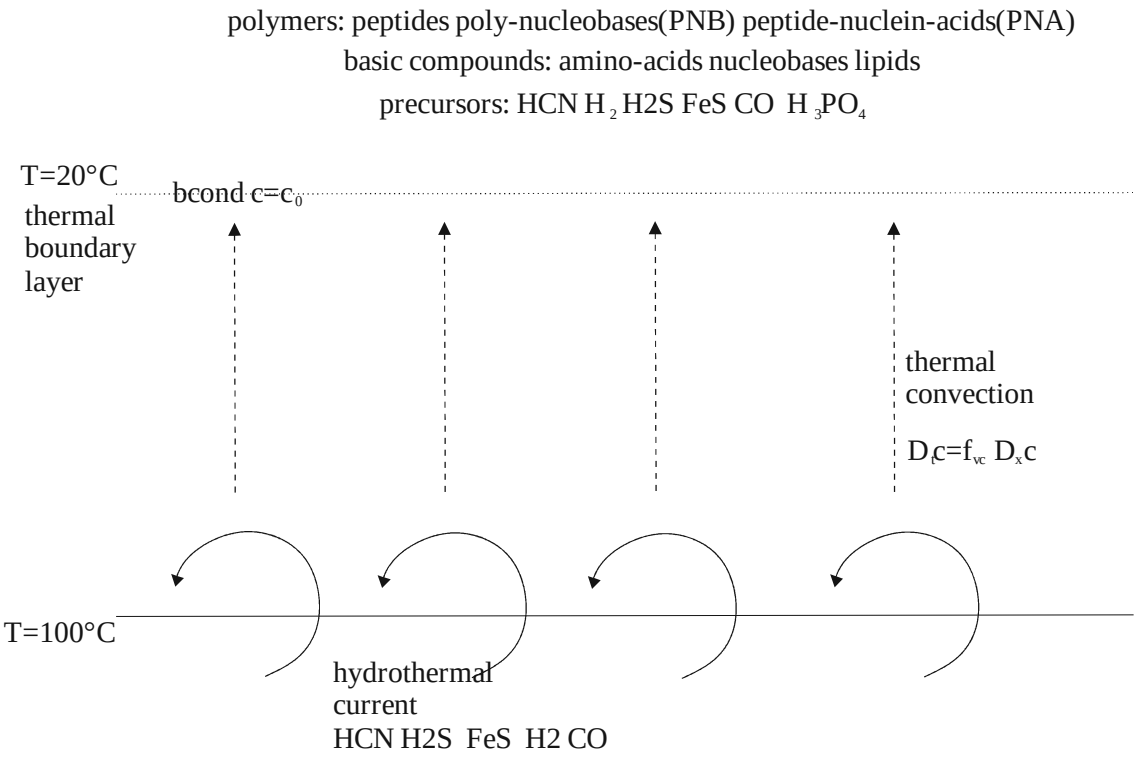
convection $\frac{\partial c(t,x)}{\partial t} = -\frac{\partial c(t,x)}{\partial x} v_c$

So in general, we have coupled algebraic (non-linear) partial differential equations of degree 1 in $t$, and of degree 1 or 2 in $x$ .

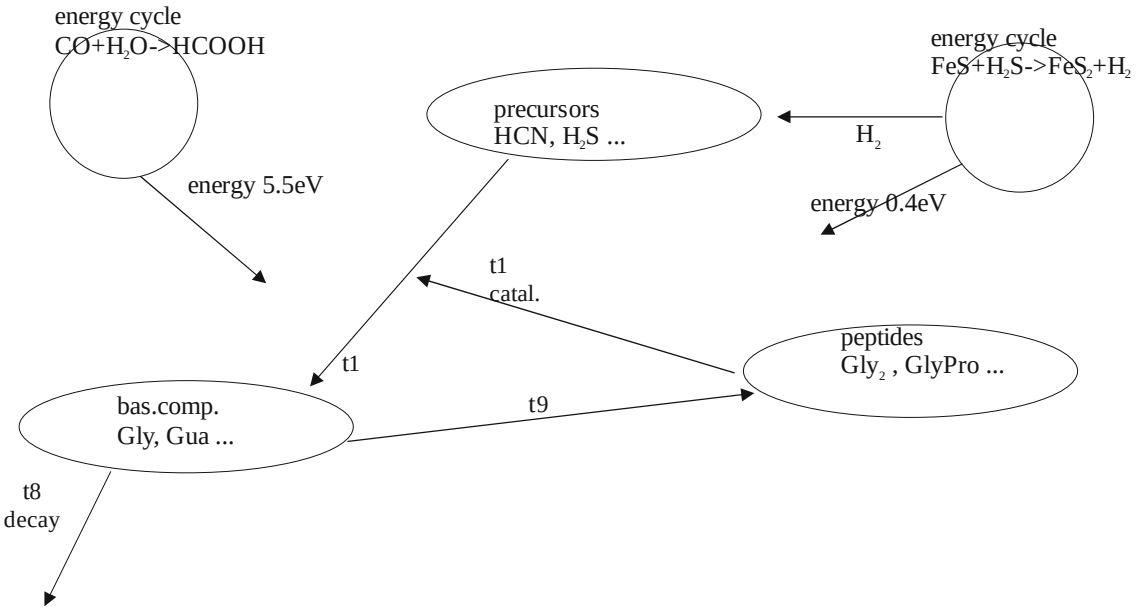
We impose boundary conditions $c(t,x_1) = c_0$ at the external boundary $x_1$, and initial conditions in the form $c(0,x) = c_b(x)$

The parameters of scenario1 and its reactions can be described by the following scheme.

**Scenario1: model**




**Scenario1: reactions**




Scenario1 represents basically the primordial amino acid and peptide synthesis cycle.
Scenario1 starts with realistic primordial concentrations (init1) of basic compounds (amino acids and nucleobases) $c=10^{-6}$=1ppm, low concentration of peptides $c=10^{-8}$=0.01ppm, and c=0.001 for precursors.
A second simulation starts with higher "enriched" concentrations (init2) of $c=10^{-4}$=100ppm for basic compounds, $c=10^{-6}$=1ppm for peptides, and c=0.01 for precursors.
The reaction network consists of spontaneous and peptide-catalyzed basic-compound-synthesis (t1), which uses the H2S energy cycle as source of energy and hydrogen, and spontaneous peptide polymerization from amino acids (t9), which uses the CO energy cycle with its high energy yield, because it is highly exothermic.
The peptides decay thermally (t8) above decomposition temperature $T_c$ .
The result is an enrichment of amino acids, but no significant enrichment of nucleobases and peptides, as shown in the table below.

| molecule | initial value(ppm) | final value | behavior | time(s) | fvc(m/s) |
|---|---|---|---|---|---|
| Gly | 1 | 14 ppm | asymptotic | $10^4$ | 0.5 |

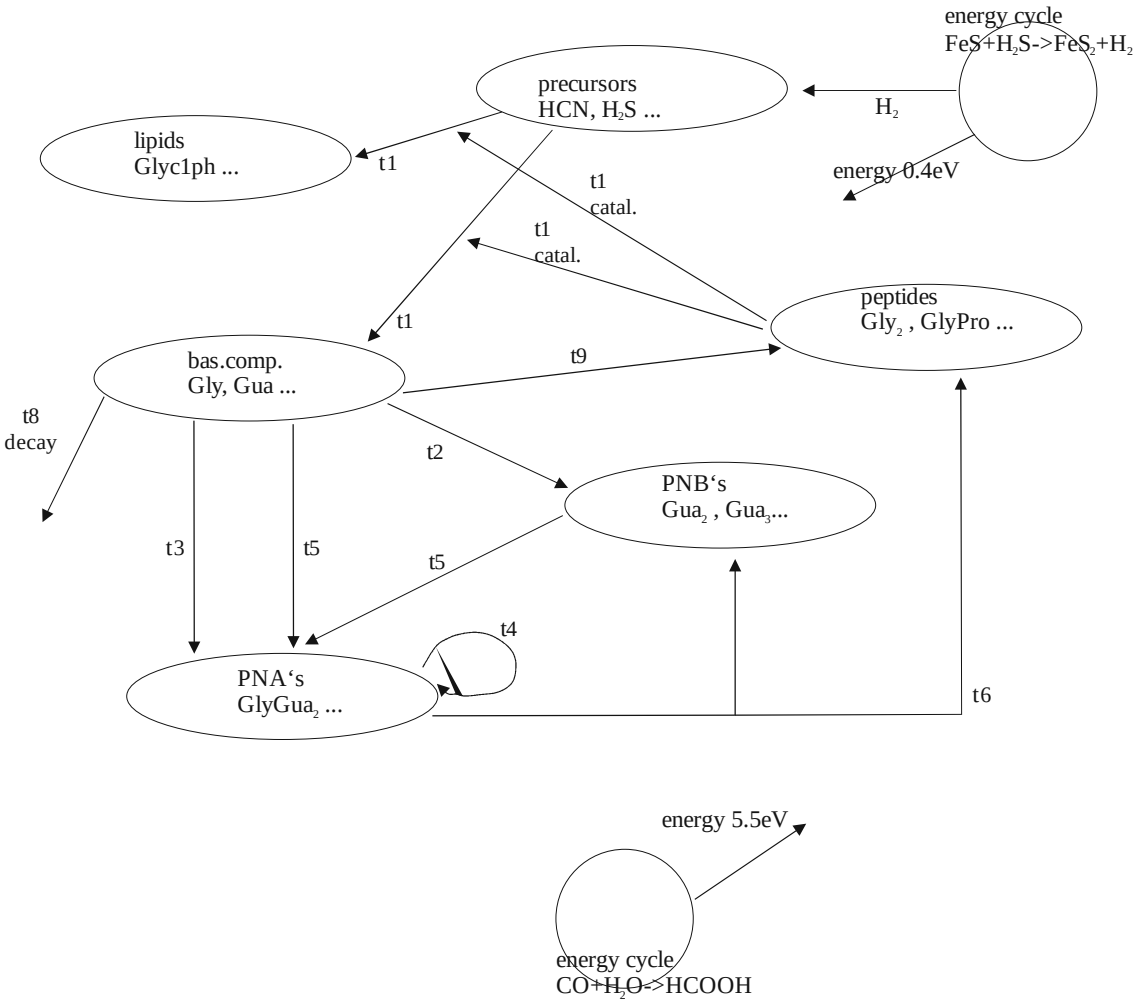
| | | | | | |
|---|---|---|---|---|---|
| Pro | 1 | 47 ppm | asymptotic | | |
| Cys | 1 | 1 ppm | constant | | |
| Gua | 1 | 1 ppm | constant | | |
| Cyt | 1 | 1 ppm | constant | | |
| Gly2 | 0.01 | $0.93*10^{-8}$ | constant | | |
| Gly3 | 0.01 | $1.*10^{-8}$ | constant | | |
| Pro2 | 0.01 | $6*10^{-12}$ | decreasing | | |
| GlyPro | 0.01 | $3*10^{-10}$ | decreasing | | |

| molecule | initial value | final value | behavior | time(s) | fvc(m/s) |
|---|---|---|---|---|---|
| Gly | $1.*10^{-4}$ | $23.*10^{-4}$ | asymptotic | $0.7*10^{4}$ | 0.5 |
| Pro | $1.*10^{-4}$ | $60.*10^{-4}$ | asymptotic | | |
| Cys | $1.*10^{-4}$ | $1.*10^{-4}$ | constant | | |
| Gua | $1.*10^{-4}$ | $1.*10^{-4}$ | constant | | |
| Cyt | $1.*10^{-4}$ | $1.04*10^{-4}$ | constant | | |
| Gly2 | $1.*10^{-6}$ | $0.94*10^{-6}$ | constant | | |
| Gly3 | $1.*10^{-6}$ | $0.1*10^{-6}$ | decreasing | | |
| Pro2 | $1.*10^{-6}$ | $0.01*10^{-6}$ | decreasing | | |
| GlyPro | $1.*10^{-6}$ | $0.26*10^{-6}$ | decreasing | | |

### 3.8.3 Scenario2: hydrothermal vent with proto-lifecycle

Scenario2 is the full proto-life cycle from the reaction table in 3.7, in the physical environment of a hydrothermal vent. It serves as an amplifier process for all involved molecule classes (amino acids, nucleobases, peptides, poly-nuclein-acids PNA, lipids), which are distributed in surroundings by thermal convection, keeping the concentration of bio-molecules in the thermal boundary layer stable and relatively low.

**Scenario2: reactions**

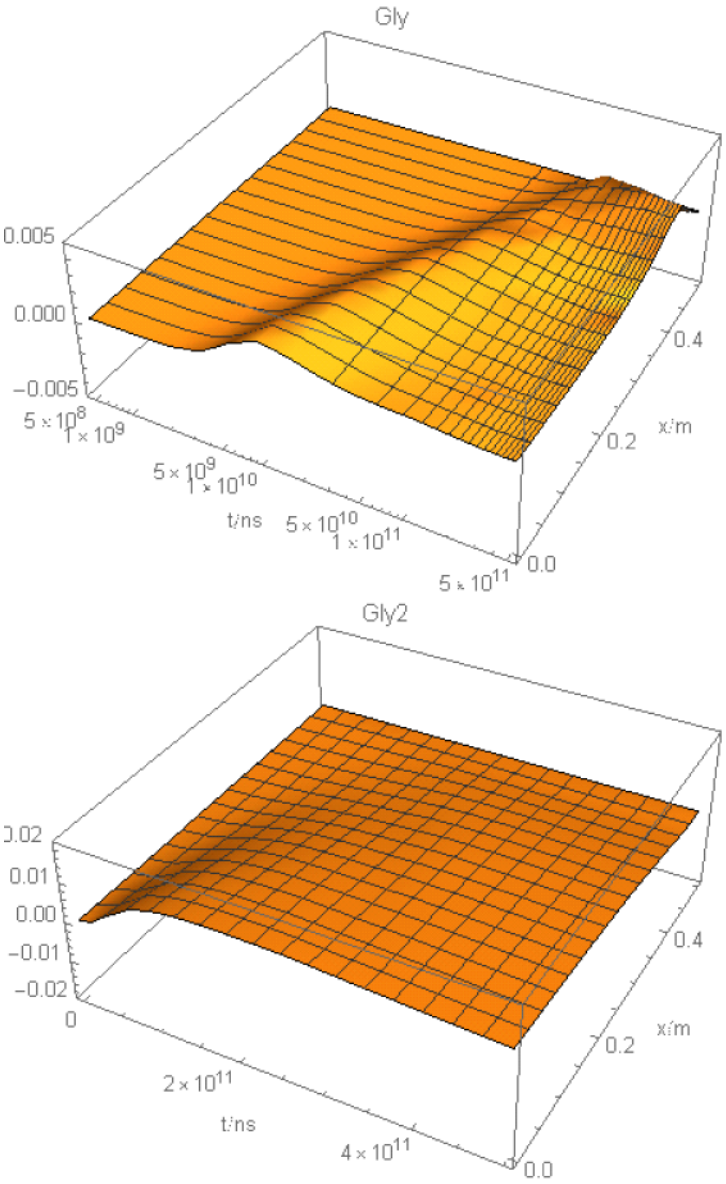


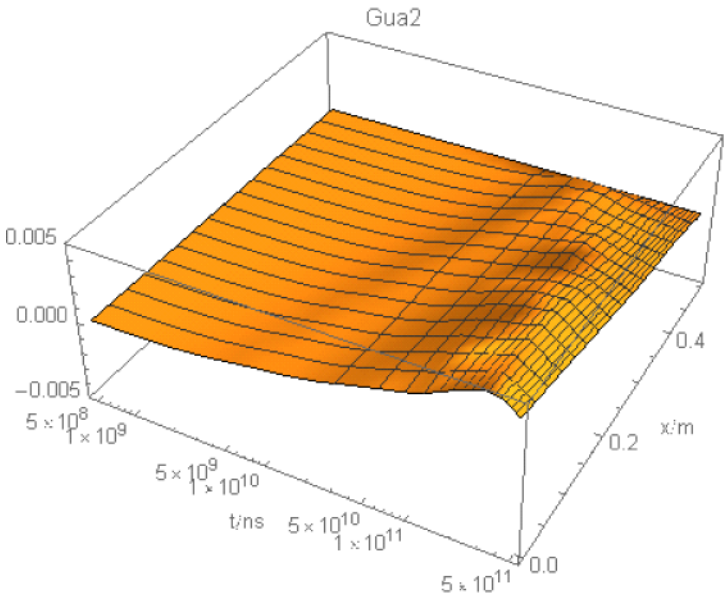

Scenario2 starts with enriched concentrations (init2) of $c=10^{-4}$=100ppm for basic compounds, $c=10^{-6}$=1ppm for peptides, and c=0.01 for precursors.
The reaction network consists of spontaneous and peptide-catalyzed basic-compound-synthesis (t1, t9. t8), with the H2S energy cycle and the CO energy cycle, then the PNA-controlled peptide synthesis (t2, t5, t6, t4) , and direct PNA-polymerization t3.
The PNA-controlled peptide synthesis works also without initial presence of peptides (i.e. initial peptide concentration c=0), whereas the direct peptide polymerization t9 requires high rates of CO energy cycle to counteract the reverse reaction (direct peptide decay), which is exothermal and has the same activation energy $E_a$ , although it has a little larger reaction times $t_0$ and therefore is a little slower.
The results for convection velocity fvc=0.5m/s are given in the following table.

| molecule | initial value | final value | behavior | time(s) | fvc(m/s) |
|---|---|---|---|---|---|
| Gly | $1.*10^{-4}$ | $0.0025->9.2*10^{-4}$ | peak | 500 | 0.5 |
| Pro | $1.*10^{-4}$ | $0.008->52.*10^{-4}$ | peak | | |
| Cys | $1.*10^{-4}$ | $1.*10^{-4}$ | constant | | |
| Gua | $1.*10^{-4}$ | $0.7*10^{-4}$ | decreasing | | |
| Cyt | $1.*10^{-4}$ | $1.1*10^{-4}$ | decreasing | | |
| Gly2 | $1.*10^{-6}$ | $0.008->1.*10^{-6}$ | peak | | |

| | | | | | |
|---|---|---|---|---|---|
| Gly3 | $1.*10^{-6}$ | $0.0008->4.*10^{-6}$ | peak | | |
| Pro2 | $1.*10^{-6}$ | $1.2*10^{-5}->8.5*10^{-6}$ | peak | | |
| GlyPro | $1.*10^{-6}$ | $1.3*10^{-5}->>49.*10^{-6}$ | peak | | |
| Gua2 | $1.*10^{-6}$ | $0.004->36.*10^{-4}$ | peak | | |
| Gua3 | $1.*10^{-6}$ | $0.009->41.*10^{-4}$ | peak | | |
| Cyt2 | $1.*10^{-6}$ | $3.4*10^{-4}$ | increase | | |
| GuaCyt | $1.*10^{-6}$ | $15.*10^{-4}$ | increase | | |

A typical concentration build-up for the amino acid Gly, the peptide Gly2 and the poly-nuclein Gua2 is shown below.




Gua2
0.005
0.000
−0.005
5 × 10^8
1 × 10^9
5 × 10^9
1 × 10^10
t/ns
5 × 10^10
1 × 10^11
5 × 10^11
0.0
0.2
0.4
x/m

### 3.8.4 Scenario3: lipid vesicles with proto-lifecycle and proliferation

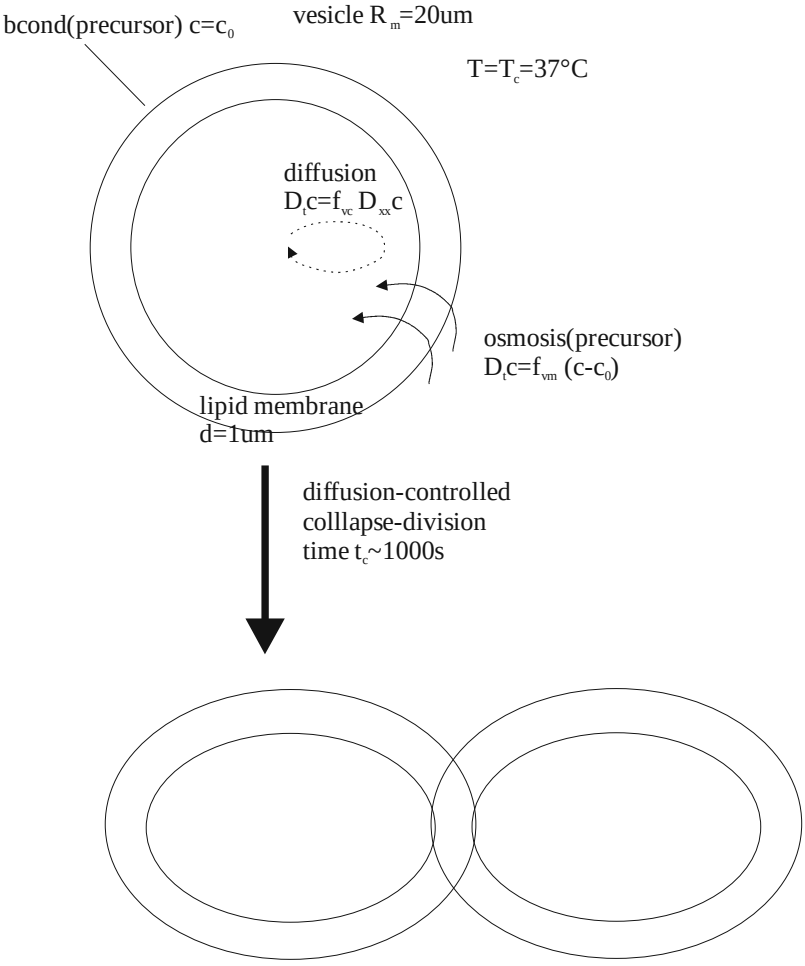
Scenario3 is the full proto-life cycle, confined in the interior of a lipid vesicle. As was reported in 2.4, amino acids within a lipid membrane stabilize the lipid layer, and the lipids serve under certain conditions as catalyzer for peptide polymerization. Precursor molecules enter the vesicle via osmosis, but bio-molecules are enclosed in it, and are not carried away by convection, there is diffusion only within the vesicle. Therefore the concentration of bio-molecules can increase until it reaches a critical level, and the system becomes unstable, the vesicle divides and proliferation takes place. The reaction cycle now has **all attributes of life**: synthesis of basic compounds from precursors ("food") using an energy cycle, self-regulation by catalysis through peptides-enzymes, PNA-gene-controlled enzyme production from amino acids, and proliferation through bio-matter production and vesicle division.

**Scenario3 model**

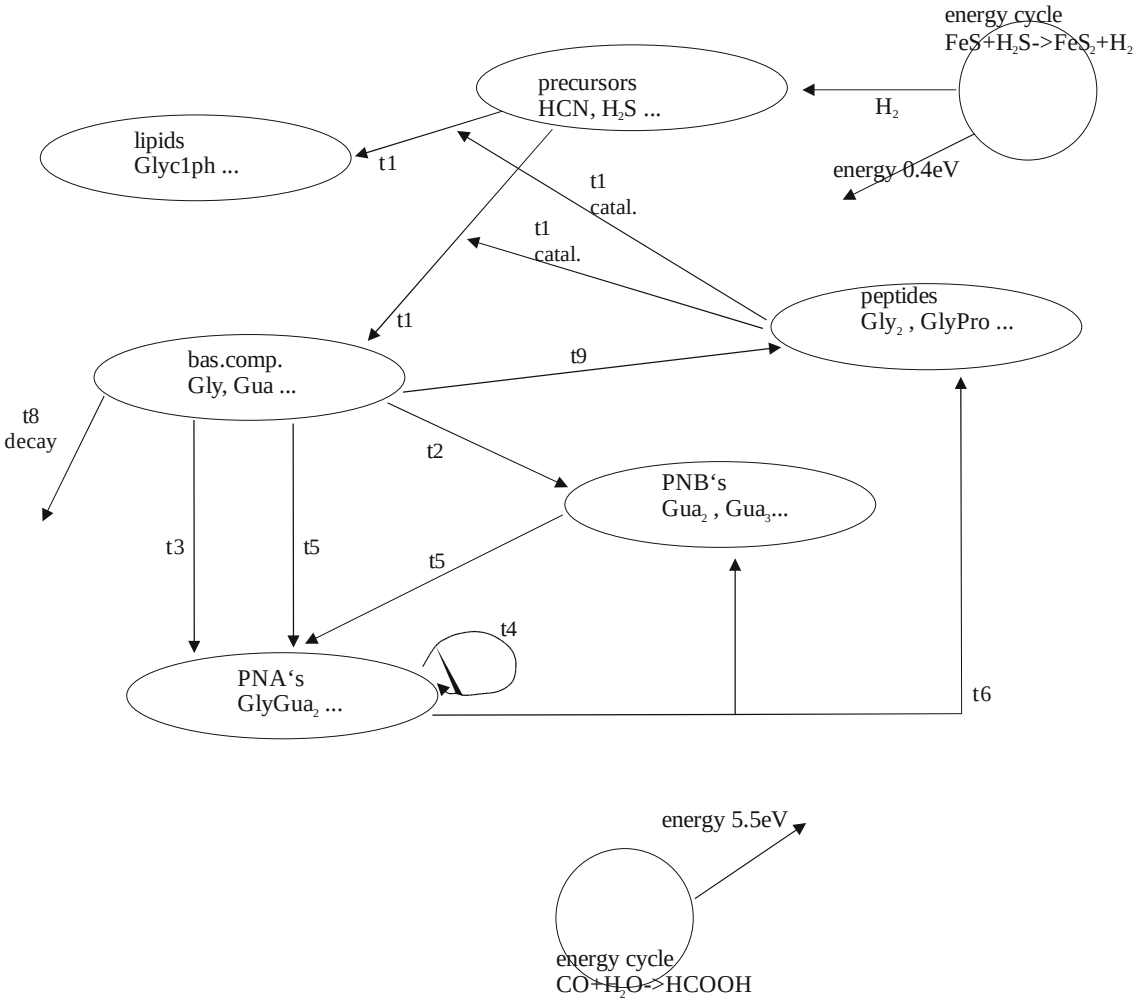
polymers: peptides poly-nucleobases(PNB) peptide-nuclein-acids(PNA)

basic compounds: amino-acids nucleobases lipids

precursors: HCN $H_2$ H2S FeS CO $H_3PO_4$



**Scenario3 reactions**




Scenario3 reaches significantly higher concentrations than scenario2, because the bio-molecules cannot leave the vesicle. Only precursor molecules can pass the membrane, for them the boundary condition $c=c_0$ is valid at the membrane. The current of precursors through the membrane is governed by osmosis:

$$\frac{\partial c(t,x_1)}{\partial t} = f_{vm}\,(c(t,x_1) - c_0)$$

and within the vesicle there is diffusion for all molecules

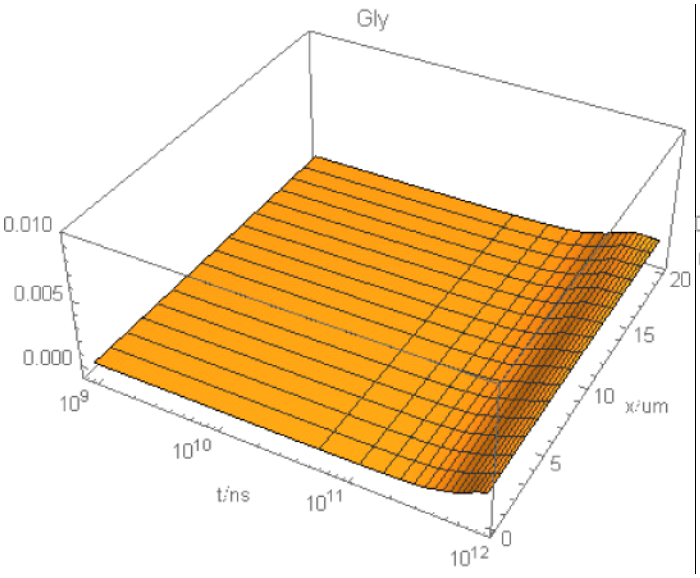
$$\frac{\partial c(x,t)}{\partial t} = f_{vd}\,\frac{\partial^2 c(x,t)}{\partial x^2}$$

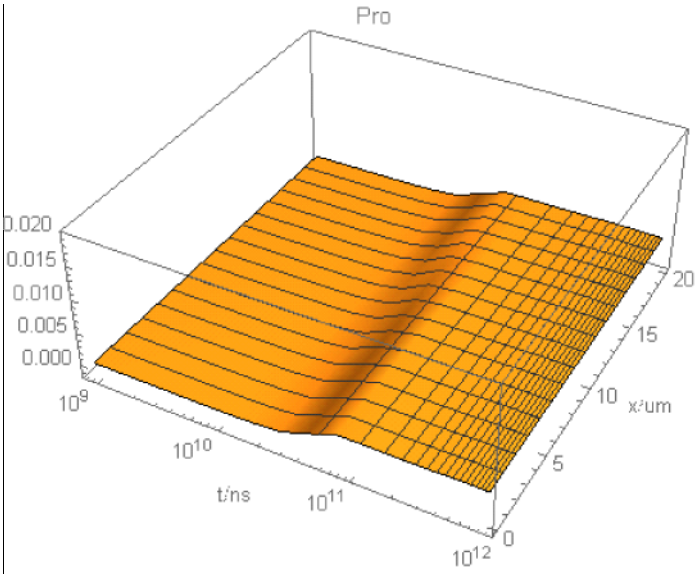
If $f_{vd}=0$, i.e. without diffusion, the system reaches an equilibrium, and there is a solution for all times.
The diffusion introduces a definite critical time $t_c$ , where the system becomes unstable, i.e. where the solution of the differential equations ceases to exist: the vesicle divides, separates into two, and the reaction cycle starts again. This critical time depends on the system parameters and the initial concentration, and has values around 1000s, which agrees well with the observed division periods of self-reproducing lipid-amino-acid vesicles and also of bacteria.

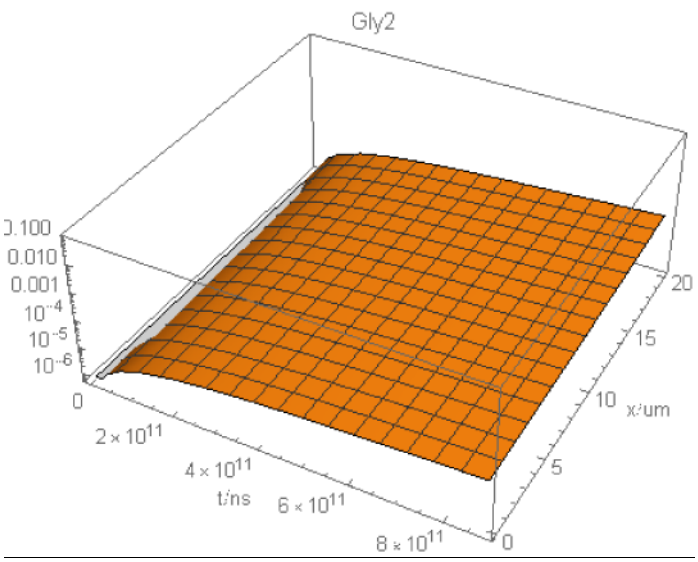
The results for osmosis constant $f_{vm}=0.3s^{-1}$ and diffusion constant $f_{vd}=0.01\ m^2s^{-1}$ are given in the following table.

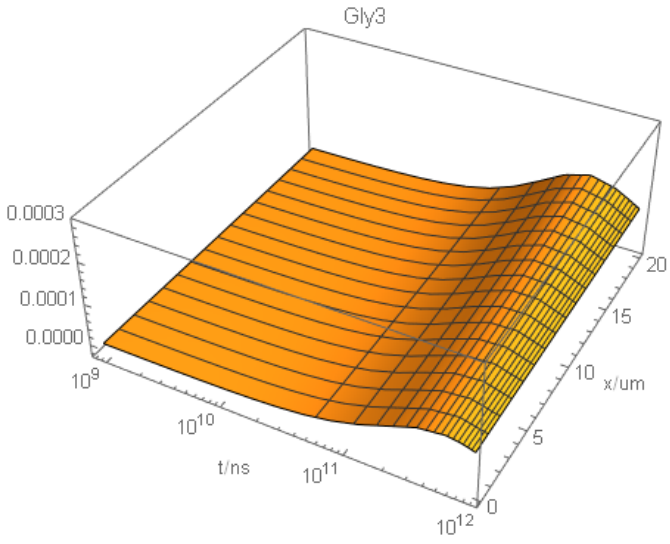
| molecule | initial value | final value | behavior | time(s) | fvm(1/s) fvd(m2/s) |
|---|---|---|---|---|---|
| Gly | $1.*10^{-4}$ | 0.0016 | increasing | 880 | 0.3 , 0.01 |
| Pro | $1.*10^{-4}$ | 0.0034 | plateau | | |
| Cys | $1.*10^{-4}$ | $81.*10^{-6}$ | plateau | | |
| Gua | $1.*10^{-4}$ | $7.6*10^{-7}$ | decreasing | | |
| Cyt | $1.*10^{-4}$ | $11.9*10^{-7}$ | decreasing | | |
| Gly2 | $1.*10^{-6}$ | $34*10^{-6}$ | increasing | | |
| Gly3 | $1.*10^{-6}$ | $82*10^{-6}$ | plateau | | |
| Pro2 | $1.*10^{-6}$ | $1.4*10^{-6}$ | peak-plateau | | |
| GlyPro | $1.*10^{-6}$ | $22.*10^{-6}$ | plateau | | |
| Gua2 | $1.*10^{-6}$ | $19.*10^{-5}$ | increasing | | |
| Gua3 | $1.*10^{-6}$ | $34.*10^{-5}$ | increasing | | |
| Cyt2 | $1.*10^{-6}$ | $3.5*10^{-5}$ | plateau | | |
| GuaCyt | $1.*10^{-6}$ | $10.*10^{-5}$ | plateau | | |

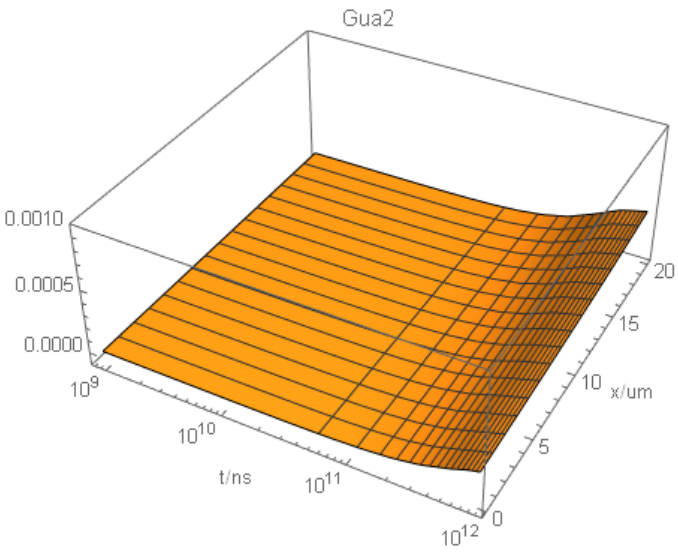
The concentration *c(t,x)* for the amino acids Gly, Pro, the peptides Gly2, Gly3 and the PNA's Gua2, GuaCyt is shown below.

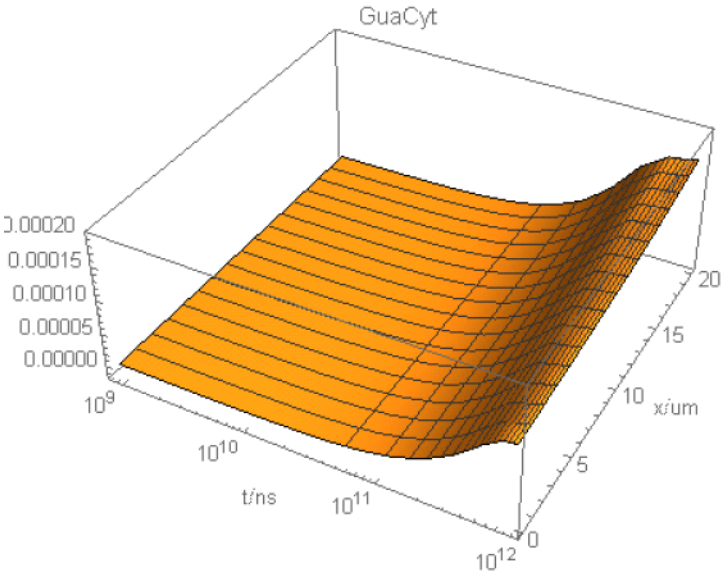



Gua2
0.0010
0.0005
0.0000
$10^9$
$10^{10}$
t/ns
$10^{11}$
$10^{12}$
0
5
10
15
20
x/um


GuaCyt
0.00020
0.00015
0.00010
0.00005
0.00000
$10^9$
$10^{10}$
t/ns
$10^{11}$
$10^{12}$
0
5
10
15
20
x/um

**References**

[1] wikipedia 2019, DNA RNA
[2] wikipedia 2019, Cell
[3] B. Alberts et al., Molecular biology of the cell, Garland Science, 2015
[4] wikipedia 2019, Evolutionary history of life
[5] C. Ouzounis et al, A minimal estimate for the gene content of the LUCA , Res Microbiol., 2006 Jan-Feb
[6] A. Poole, The LUCA, actionbioscience.org, 07/2002
[7] M. Russell et al., The rocky roots of the acetyl-CoA pathway, Trends in Biochemical Science 29, 2004
[8] W. Martin M. Russell, On the origin of cells, Phil. Trans. R. Soc. Lond. B, 2002
[9] L. Orgel, Self-organizing biochemical cycles, Nat. Acad. of Sciences, 2000
[10] B. Holmes, The very first living thing (LUCA), New Scientist 1 November 2017
[11] B. Holmes, First glimpse at the viral birth of DNA, NewScientist 18 April 2012
[12] P. Nielsen, A new molecule of life (PNA), Sciam 12/2008
[13] R. Shapiro, A Simpler Origin for Life, Sciam 12 February 2007
[14] J. Sutherland et al., Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, Nature volume 459 2009
[15] T. Carrel, A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway, Science 13 May 2016
[16] J. Oro, Prebiotic syntheses of purines and pyrimidines, Advances in Space Research Volume 4 Issue 12 1984
[17] wikipedia 2019, Amino acid synthesis
[18] S. Harrison, Life as a guide to prebiotic nucleotide synthesis, Nature December 2012
[19] L. Geddes , Spark of life: Metabolism appears in lab without cells (PNA/RNA world), New Scientist 25 April 2014
[20] J, Cepelewicz , The End of the RNA world is near (PNA/RNA world), quanta magazine 19 December 2017
[21] C. Carter , Why a Peptide/RNA Partnership Merits Renewed Experimental Attention (PNA/RNA world), Life 2015 *5*(1)
[22] M. Le Page , Meet your maker: Homing in on the ancestor of all life (LUCA), New Scientist 12 August 2014
[23] M. Marshall, First life: The search for the first replicator (RNA-clay-world ), New Scientist 15 August 2011
[24] E. Parker et al., Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment, . Proceedings of the National Academy of Sciences. 108, 2011
[25] J. Cronin S. Pizzarello, Amino acids in meteorites, Advances in Space Research Volume 3 Issue 9, 1983
[26] J. Dworkin et al., Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases, PNAS, 2011
[27] S. Miller et al., Peptide nucleic acids rather than RNA may have been the first genetic molecule, PNAS 11 April 2000 97 (8), 2000
[28] M. Marshall, Life in the hodge-podge world, NewScientist 08 January 2012
[29] C. Carter P. Wills, Interdependence, Reflexivity, Fidelity, Impedance Matching, and the Evolution of Genetic Coding, Molecular Biology and Evolution, 2017
[30] L. Rodriguez et al., Functional Class I and II Amino Acid-activating Enzymes Can Be Coded by Opposite Strands of the Same Gene, Journal of Biological Chemistry, 2015
[31] J. Dworkin et al., Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices, PNAS, 2001
[32] C. Caitlin et al., Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes, PNAS August 27, 2019
[33] Y. Wolf et al., Origins and Evolution of the Global RNA Virome, mBio, November 2018
[34] K. Moelling F. Broecker, Viruses and Evolution – Viruses First? A Personal Perspective, Front. Microbiol., March 2019
[35] B. Patel et al., Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat. Chem. 7, 2015
[36] E. Koonin, Comparative genomics and new evolutionary biology , Boston, Kluwer Academic; 2003.

[37] O. Sijbren et al., Spontaneous Emergence of Self-Replicating Molecules Containing Nucleobases and Amino Acids, American Chemical Society, February 2020
[38] G. Wächtershäuser, ORIGIN OF LIFE: Life as We Don't Know It, Science 2000 289 (5483)
[39] H. Raanan, Origin of the building blocks of life, Science 03/2020
[40] J. Thiel et al., Pyrite formation from FeS and H2S is mediated through microbial redox activity, PNAS April 2019
[41] T. Das et al., Did it begin from HCN and H2O?, ACS Cent. Sci. 2019-5
[42] I. Weiss, Thermal decomposition of the amino acids, BMC Biophys. 2018-11
[43] T. Lindahl, Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution, Biochemistry 1973-12
[44] Hydrogen bond, Peptide bond, Phosphodiester bond, wikipedia 2020
[45] Mathematica-notebook EnzymeCalc.nb, www.janhelm-works.de
[46] R. Gatenby B. Frieden, Investigating information dynamics in living systems, Plos One, May 2016
[47] NIST Chemistry Webbook, www.chemeo.com
[48] J. Biro, Amino acid size, charge, hydropathy indices, Theoretical Biology and Medical Modeling 2006-3
[49] D. Trail et al., The oxidation state of Hadean magmas and implications for early Earth's atmosphere, Nature 480(7375) 2011-12
[50] K. Zahnle, Earth’s earliest atmospheres , Cold Spring Harbor Laboratory Press 2010
[51] S. Ranjan D. Sasselov, Influence of the UV Environment on the Synthesis of Prebiotic Molecules
[52] D. Ritson et al., Mimicking the surface and prebiotic chemistry of early Earth using flow chemistry, Nature Communications 2018-9
[53] H. J. Cleaves , Prebiotic Chemistry: What We Know, What We Don't, Evo Edu Outreach 2012-5
[54] K. Pedersen S. Ekendahl, Distribution and activity of bacteria in deep granitic groundwaters.., Microbial Ecology 1990-20
[55] NIST Chemistry WebBook, webbook.nist.gov/chemistry/
[56] Chemeo molecular database, www.chemeo.com
[57] NIST chemical kinetics data base, kinetics.nist.gov
[58] Jan Helm, Chemical data base, ChemicalData_JH0220.pdf, www.janhelm-works.de
[59] Jan Helm, Biochemistry basics: a compendium of models and data, BiochemistryBasics_JH0420.pdf, www.janhelm-works.de
[60] G. de With, Liquid state physical chemistry, Wiley-VCH, Weinheim 2013
[61] J. Chasnov, Mathematical biology, Lecture Hongkong university, 2009
[62] C. Mayer U. Schreiber et al., Molecular Evolution in a Peptide-Vesicle System, Life 2018-8 16
[63] W. Heinen A. Lauwers , Organic Sulfur Compounds Resulting From the Interaction of Iron Sulfide..., Orig Life Evol Biosph 1996-04
[64] L. Landau E. Lifshitz, Fluid mechanics, Elsevier, 1987
[65] Jan Helm, Chemical thermodynamics: a compendium of models and data, ThermodynChemical_JH0420, www.janhelm-works.de
[66] H. Lodish H. Berk et al. , Molecular Cell Biology, W. H. Freeman; New York 2000
[67] J. M. Castro et al., Budding and division of giant vesicles..., Nature scientific reports, 2018-01-17
[68] P. W. Kudella et al., Fission of lipid-vesicles by membrane phase transitions in thermal convection, Nature scientific reports, 2018-9