Scientific Research Publishing

Open Journal of Biophysics, 2024, 14, 265-329
https://www.scirp.org/journal/ojbiphy
ISSN Online: 2164-5396
ISSN Print: 2164-5388

Origin and Basic Mechanism of Life

Jan Helm

Department of Electrical Engineering, Technical University Berlin, Berlin, Germany
Email: jan.helm@alumni.tu-berlin.de

How to cite this paper: Helm, J. (2024) Origin and Basic Mechanism of Life. *Open Journal of Biophysics*, **14**, 265-329.
https://doi.org/10.4236/ojbiphy.2024.143011

Received: April 20, 2024
Accepted: June 3, 2024
Published: June 6, 2024

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. We present here a calculation method for biochemical reactions based on the available reaction data base, and using this method, we calculate precise scenarios for the first life cycle, and for the first stages of terrestrial biological evolution.

Keywords

Proto-Life-Cycle, Prebiotic Evolution, PNA World, Genetic Proto-Code, LUCA

1. Introduction

The evolution of life is known roughly down to LUCA (Last Universal Common Ancestor). Our knowledge about LUCA and the later life evolution is summarized in Chap. 2.11.

We have little evidence about the two preceding stages of life evolution: the prebiotic chemo-evolution and the proto-life cycle.

Important steps in the research concerning these two stages of life evolution can be summarized as follows.

1) Oparin-Haldane hypothesis

The earliest was the Oparin-Haldane hypothesis, which suggests that life arose from inorganic molecules, with basic amino acids and nucleobases forming first and building polymers within self-replicating lipid membranes [1].

2) Miller-Urey experiment

The Miller-Urey experiment provided the first evidence that some amino acids and nucleobases could be formed in water from a reducing atmosphere, under heating and electric discharge [1].

3) RNA-world

The RNA world hypothesis suggests that the first life was self-replicating primitive RNA [2] [3].

4) PNA-world

PNA (peptide nucleic acid, *i.e.* peptides with nucleosides) has been proposed as a precursor of RNA and as a catalyzer and information carrier for peptide synthesis [2] [3]. This ansatz is shown in Chap. 2.6 and Chap. 3.2 as the most probable precursor of RNA-world and LUCA (PNA proto-cell), supported by theoretical and experimental evidence.

5) Wächtershäuser's iron-sulfur world

Wächtershäuser traces the proto-biotic synthesis of amino acids and nucleobases back to simple, surface-mediated autotrophic reactions, driven by metal catalysts, mainly iron ions ([2]). Wächtershäuser's pyrite cycle is shown in Chap. 2.6 to be the central energy cycle of the PNA proto-cell.

6) Life from HCN

Patel demonstrated that ribonucleotides, lipid-precursors and amino acids, can be formed from CH_4, NH_3, H_2S in water under UV irradiation ([2]). This ansatz is proved to be successful as the basis for the numerically calculated model of PNA proto-cell in Chap. 2.6 and Chap. 3.2.

The tasks required to gain knowledge about these preceding stages can be summarized in the following scheme.

1) Prebiotic chemo-evolution

Here the first goal is to deduce the reactions and components of the chemo-evolution in space of amino acids and nucleobases.

The second goal is to find out the precursors, energy cycles, the environment, and the basic amino acids and nucleobases, which creates the conditions for the proto-life-cycle.

2) The evolution and the structure of the proto-life-cycle

Here it is required to find the minimal genetic code, which is consistent with the modern genetic code, and the reactions with the corresponding enzymes (peptides) and transferases.

The reaction network should be verified by calculation or by experiment or both.

In order to achieve this, the following methods are available

1) Prebiotic soup experiments

2) Reconstruction of primeval genetic code

This can be done by tracking down the oldest common genes, and by finding minimal sets of genes for life functionality.

3) Analysis of RNA code

Here the goal is to find the structure of early stages of genetic code.

4) Calculation of biological reaction networks

Here the goal is to find a reliable reaction model for peptide-catalyzed synthesis of poly-nucleobases, amino acids, and peptides.

For a given reaction

$$A + B\left(k_{+} \rightarrow\right) C + D + \Delta E \quad \text{resp.} \quad A + B\left(\leftarrow k_{-}\right) C + D + \Delta E$$

with the mass action law

$$\frac{\mathrm{d}C}{\mathrm{d}t} + \frac{\mathrm{d}D}{\mathrm{d}t} = k_{+}\left[A\right]^{n_A}\left[B\right]^{n_B}$$

we have the Arrhenius law for the reaction constant

$$k_r = \left(1/t_0\right)\exp\left(-E_a/kT\right)$$

The goal is here to calculate the fundamental constants char. Time t_0 and activation energy E_a in dependence of the structure parameters (bonds, geometry) of the involved molecules *A*, *B*, *C*, *D*

$$t_0 = t_0\left(A,B,C,D\right), \quad E_a = E_a\left(A,B,C,D\right)$$

In Chap. 2, we develop a mathematical model with method 4, based on the current biochemical data base.

We use this model for precise calculation of the late prebiotic chemo-evolution and the proto-life cycle under plausible conditions, based on HCN and H_2S as precursors, PNA-controlled peptide synthesis, with genetic proto-code with 5 genes, and pyrite reaction as energy cycle.

We develop plausible scenarios for the later evolution up to the LUCA stage, using method 2 and method 3.

Based on this, we present in Chap. 3 a complete scenario of terrestrial life evolution, and set up principles for biological life in general.

2. Principles and Evolution of Life Chemistry

2.1. 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 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, vesicle-based chemistry is a precondition for a full life-cycle with bio-matter production and proliferation (see Chap. 2.8.4). 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.

In extrapolation of these basics, we can say that life in general has 6 principal features (Figure 1):

- **synthesis** of biological matter (including basic compounds) in a liquid (terrestrial: water H_2O) catalyzed by enzymes (terrestrial: proteins), where basic compounds (terrestrial: 20 amino acids, 4 nucleobases A, U, G, C) are synthesized from anorganic precursors (terrestrial primary: HCN, H_2S)

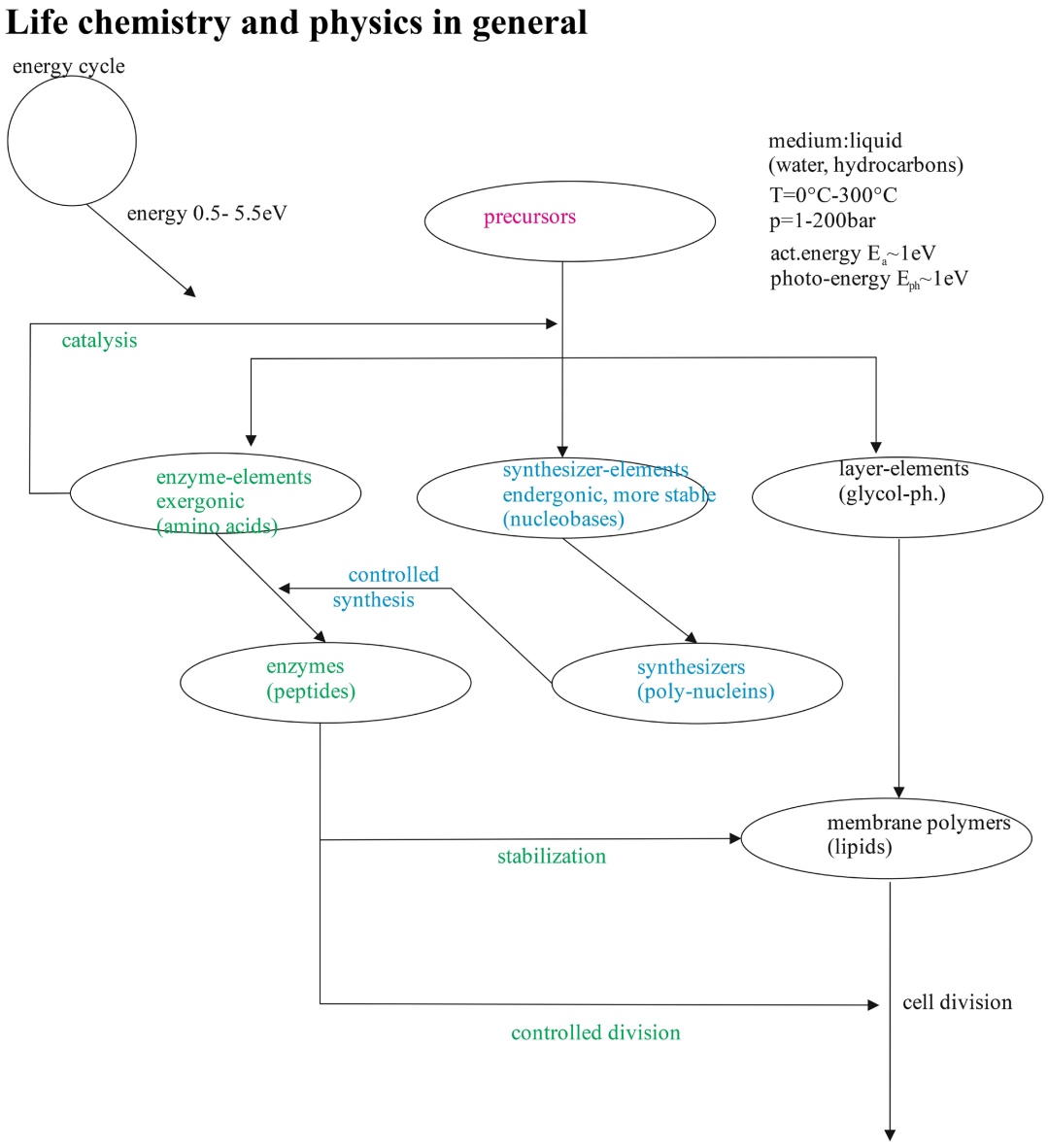


Figure 1. Schematic life in general.

- use of an **energy cycle** with a continuous energy flow against entropic decay (terrestrial primary cycle: Wächtershäuser's iron-sulfur cycle, terrestrial secondary cycle: photo-synthesis)
- life chemistry based on catalyzer **enzymes** (terrestrial: peptides) and **synthesizers** (genes, terrestrial: poly-nucleobases) built from basic compounds, and running within a cell-membrane, where synthesizers are more stable than enzymes; synthesizers serve as a pattern for enzyme structure, catalyze the enzyme synthesis, and preserve and copy the pattern
- spontaneous emergence of a **proto-lifecycle**, which starts reproduction and evolution through selection
- **replication** with copying of genes, and cell division within a fixed time interval (terrestrial: ca. 1000 s)

- **evolution** by mutation and adaptation to the environment with the goal of survival

We can formulate some boundary conditions for life:

- The only common element, which supports a large variety of molecules, and polymers, is carbon.

Terrestrial life is *based on carbon chemistry*, and presumably the same is true for most forms of life in the universe.

- The *fluid carrier medium* in terrestrial life is *water*, and water is the only common molecule with a fluid phase around $T = 300$ K.
- Biological carbon chemistry has *activation energies* in the range (25 kJ/mol, 130 kJ/mol), corresponding to per-molecule energy range (0.25 eV, 1.3 eV).
- Star light as an energy source has an approximate energy range ($E(\lambda = 0.5\ \mu\text{m}) = 2.5$ eV, $E(\lambda = 1\ \mu\text{m}) = 1.2$ eV), with $E(\lambda) = \frac{hc}{\lambda}$, center wavelength $\lambda = 0.5\ \mu$m for sun-like stars, center wavelength $\lambda \approx 1\ \mu$m for Red Dwarves, which is *feasible as energy source for carbon chemistry*.
- The thermal energy $E_{th}(T) = k_B T$ needed to overcome the activation energy in carbon chemistry is approximately (2σ above mean) $E_{th,c} e^2 = E_{a,min}$, where $E_{a,min} = 0.25$ eV for carbon chemistry, *i.e.* $E_{th,c} = 0.034$ eV, with *corresponding minimum temperature* $T_c = 390$ K, which lies within *the water fluidity temperature range*.

We can extrapolate general features of the four life-functional molecular families.

- **Precursors, solvents, energy cycle molecules**

These molecules are synthesized in space, or belong to the primeval molecules on planets. Among them, there is at least one molecule, which is exothermic (supplies energy), and one which is the (fluid) solvent.

In terrestrial biology, precursors are: HCN, H_2S, H_2, CO, H_3PO_4, and the solvent is of course water H_2O. Among these, HCN is the energy supplier and HCN + H_2O are the atom suppliers, and determine the characteristic atom content ratio $(H:C:N:O) = (3:1:1:1)$ in basic amino acids and nucleobases.

The energy cycle molecules participate in the basic energy cycle reaction, in the terrestrial biology they are: H_2S, FeS which form the fundamental pyrite reaction $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$.

- **Enzymes**

The enzymes are the catalyzers, which control the life synthesis reactions. They are chains consisting of basic molecules bound by one kind of chemical bond.

In terrestrial biology, they are peptides, which are linear chains of amino acids bound by the amino-bond with energy 8 - 16 kJ/mol.

- **Synthesizers**

The synthesizers are the coding molecules for the enzymes, they are more stable than the enzymes, and therefore are ring molecules (rings are normally more stable than their linear counterparts). Their diameter is smaller than the length

of the linear enzymes by a factor 1/2 to 1/3. Therefore they form codons from 1, 2 or 3, which translate into one enzyme (1-bit, 2-bit, 3-bit code).

In terrestrial biology, they are poly-nucleobases (PNB's), which are linear chains of nucleobases

complementary adenine Ade, guanine Gua: 5-ring + 6-ring

complementary thymine Thy, uracil Ura: 6-ring.

PNB's are linked by hydrogen-bonds, in RNA by phosphodiester-bonds.

- **Layer-bilders**

These molecules form the protective semi-permeable membrane of the life vesicle (biological cell).

The membrane controls the influx and outflux of molecules, and the vesicle is capable of self-recreation.

In terrestrial biology, they are phospho-lipids, namely in the proto-cell glycerol-phosphate.

- **Reaction network**

Enzymes, synthesizers, and phospho-lipids are all synthesized in the life self-supporting reaction network catalyzed by enzymes.

Precursors, solvent, and energy cycle molecules are supplied outside the vesicle and enter resp. leave it through the membrane.

Alternative life-models

Carbon-nitrogen-hydrogen chemistry in water (terrestrial model)

$T = 0^{\circ}\text{C}...150^{\circ}\text{C}$ (P = 1 - 100 bar)

activation energy range: (25 kJ/mol, 130 kJ/mol), per molecule (0.25 eV, 1.3 eV)

basic materials: $HCN + FeS + H_2S + H_3PO_4$

enzyme polymers: peptides,

synthesizers (genes): poly-nucleins

solvent: water

energy cycle: $FeS + H_2S$, photosynthesis

location: water-carrying planets in habitable zone, in contact with mineral rock

This is the terrestrial life model, mainly dealt with in this paper. The spontaneous emergence of a proto-lifecycle is proved numerically based on chemical reaction data, and partly supported by direct observation (see Chap. 2.8.3, Chap. 3.2).

The proto-lifecycle runs with

- 3 amino acids Gly, Pro. Cys
- 2 nucleobases Gua, Cyt
- 5 peptides (enzymes) Gly2, Gly3, Pro, Pro2, GlyPro
- 1 lipid Gly1ph
- 5 PNB's (genes) Gua2, Gua3, Cyt2, Cyt, GuaCyt
- 5 PNA's GlyGua2, GlyGua3, ProCyt2, ProCyt, GlyGuaProCyt2

Carbon-hydrogen-oxygen chemistry in water

$T = 0^{\circ}C...150^{\circ}C$ ($P = 1$ - 100 bar)

activation energy range: (25 kJ/mol, 130 kJ/mol), per molecule (0.25 eV, 1.3 eV)

basic materials: $CO_2 + CH_4 + NH_3$, no iron or sulfur

enzyme polymers: poly-vinyl-alcohols, poly-ethylene-glycols, poly-acryl-amide

synthesizers (genes): aliphatic amines

solvent: water

energy cycle: photosynthesis, $CO + H_2O$

location: water-carrying planets in habitable zone, no contact with minerals

This is a tentative life model for water environments without minerals, iron, and sulfur.

Aliphatic amines like tetra-methyl-ethylene-diamine function as catalyzers for poly-acryl-amides.

Carbon-hydrogen-nitrogen chemistry fluid hydrocarbons

$T = 0^{\circ}C...300^{\circ}C$ ($P = 1$ - 100 bar)

basic materials: $CH_4 + H_2S + NH_3$, no oxygen

enzyme polymers: hydrocarbon linear-branched polymers with substitutions N, S

synthesizers (genes): polycyclic aromatic hydrocarbons (PAH) with substitutions N, S

solvent: fluid hydrocarbons

energy cycle: photosynthesis

location: waterless planets in habitable zone with dense atmosphere with hydrocarbons and nitrogen

This is a hypothetic life model based on hydrocarbon solvent, where the spontaneous emergence of proto-lifecycle with proto-genes from PAH's and enzymes from poly-hydro-carbons with N and S substitutions is still to be proved.

This model is arguably the only viable alternative to water-based life models, it supports also temperatures up to 300˚C.

PAH's are known to be common in the cosmic matter, and they are also rather stable and can function as catalyzers for hydrocarbons. PAH's can polymerize under UV radiation [4].

Silicon chemistry in molten rock

$T = 1000^{\circ}C...1500^{\circ}C$

basic materials: $SiO_2 + Al_2O_3 + MgO$

polymers: polysilicates

solvent: molten silicates

energy cycle: radiation

location: hot Earth planets

The temperatures in this model are far outside the life-favorable range, therefore it is very probably no viable model for life.

Carbon + salts compounds in liquid ammonia

reactions probably too slow for life evolution

$T = -78 \ldots -33^{\circ}C$ ($P = 1$ bar)

$T = -77$ C...98˚C ($P = 60$ bar)

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: carbon-hydrogen chemistry?

location: Titan-like planets

The temperatures in this model are far below the life-favorable range, therefore it is very probably no viable model for life.

Carbon compounds in liquid ethane-methane

reactions probably too slow for life evolution

$T = -183 \ldots -89^{\circ}C$ ($P = 1$ bar)

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

The temperatures in this model are far below the life-favorable range, therefore it is very probably no viable model for life.

2.2. Principles of Terrestrial Life Chemistry

In the case of the terrestrial life, the process took place probably from the precursor molecules {HCN, H_2} 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 **Figure 2** "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 [5] [6]: that is where the Darwinian evolution mechanism sets in. The amino acids and nucleobases also degrade, although at higher temperatures ($T > 185^{\circ}C$), so they, too, must be replenished [7]. 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.

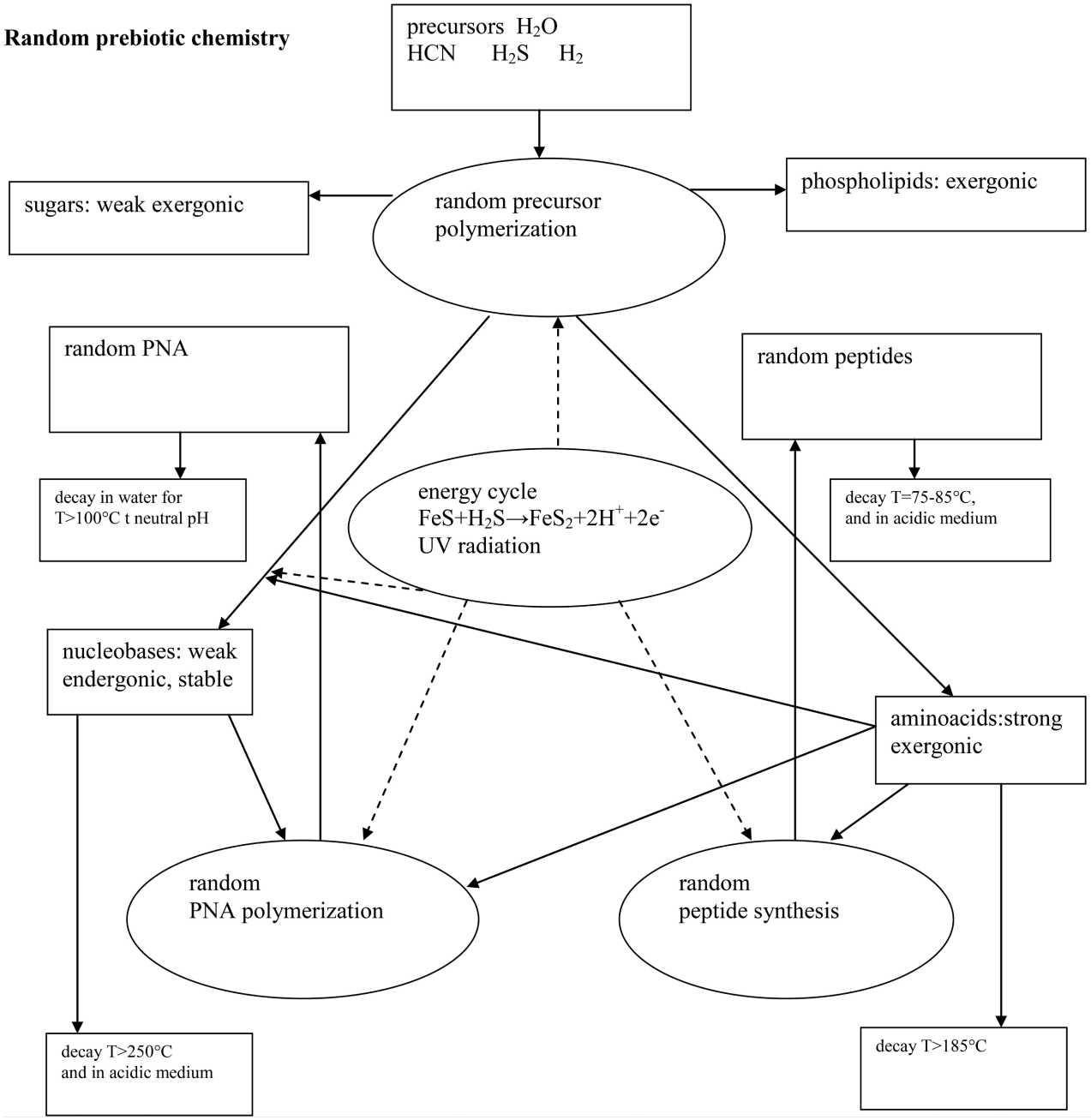


Figure 2. Schematic random prebiotic chemistry.

The result for the **proto-life cycle** (Chap. 3.2) is a **model1 3-bit binary (G/C) 5 × 2 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 **Figure 3** "proto-life chemistry").

This model is 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 [8] [9]. Furthermore, the binary genetic code

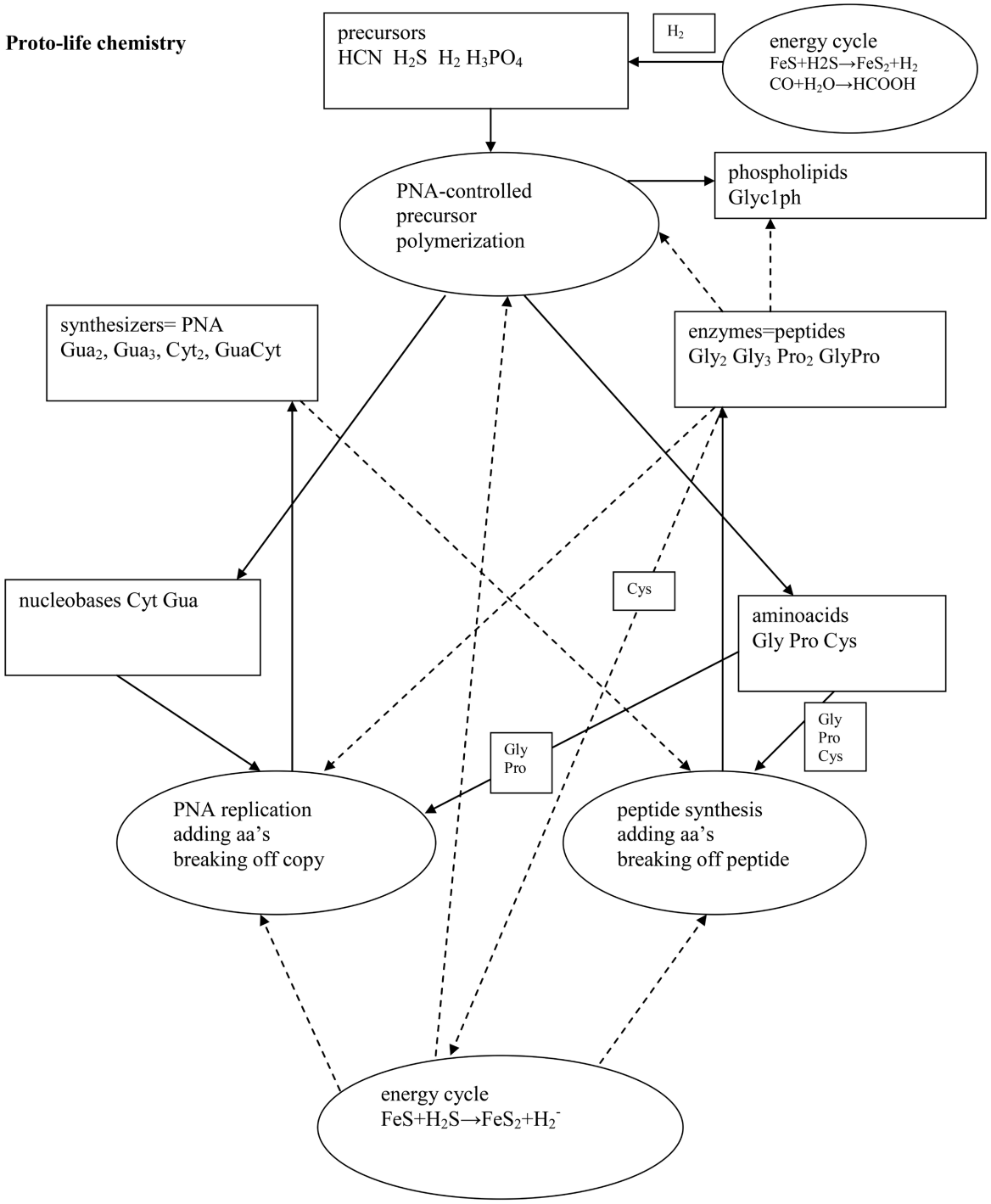


Figure 3. Schematic terrestrial proto-life chemistry.

has been shown to be the probable original mechanism of PNA-RNA-coding and protein synthesis [10] [11]. 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, H_2O, H_2) (see Chap. 2.2, Chap. 2.4, Chap. 2.5 and [12]). Finally, digital simulation based on a simplified model of the HCN chemistry shows that molecular evolution selects indeed a binary genetic code like the one

proposed in [10] [11].

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.

The **present stage** of terrestrial life evolution (chap. 3.5) represents the DNA-cell (bacteria, archaea, eukaryotes), which uses the full DNA genetic code with 20 amino acids, and corresponding 20 aaRNA's for peptide synthesis, 4 nucleotides: cytosine, thymine, adenine, guanine (thymine replaces uracil in RNA-coding). The pairs A-G and C-T are complementary in the DNA-double-helix. The self-sustainable energy cycle is mostly the photosynthesis (plants, cyanobacteria). The self-sustainable energy cycle is mostly photosynthesis (plants, cyanobacteria).

2.3. 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.

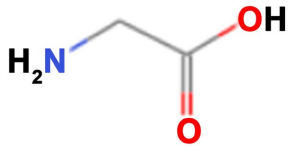
There are two kinds of data, which are needed for a structure-based numerical simulation of chemical evolution.

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 ΔE of a given reaction $A + B\left(k_+ \rightarrow\right) C + \Delta E$.

Second, we need the reaction rate $\frac{\mathrm{d}C(t)}{\mathrm{d}t}$ 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).

2.3.1. Empirical Model of Molecular Energy

The *molecular forming free enthalpy* ΔH_f has been measured for hundreds of molecules, values can be found in [13] [14] [15] [16]. It is defined as the forming energy from a natural state at normal conditions (T = 300 K, P = 1 bar). The *molecular forming energy from atoms* H_f is calculated as $H_f = \Delta H_f - E(\text{gas})$, where E(gas) is the forming energy of gaseous components (H_2, O_2, N_2) in the molecule, e.g. for glycine we get



formula $C_2H_5NO_2$

$\Delta H_f = -390.5$ kJ/mol [13]

$$E(\text{gas}) = E(2.5 * \text{HH}, \text{O=O}, 0.5\ \text{N}\equiv\text{N}) = 2.5 * 432 + 494 + 0.5 * 942 = 2045$$

$$H_f = \Delta H_f - E(\text{gas}) = -2045 - 390.5 = -2435.5\ \text{kJ/mol}$$

where the bond energies of HH = 432 kJ/mol in H_2, (O=O) = 494 kJ/mol in O_2, (N≡N) = 942 kJ/mol in N_2 [16].

The values of H_f are calculated in [16].

A molecule can be characterized by a list of bonds, e.g. for glycine we get {OH, CO, C2O, CC, 2CH, CN, 2NH}, where C2O is the double bond C=O [16]. The values of the bond energies are given in data tables like [13] [14]. 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 [12] [16].

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., P2O = 544., PO = 335.}

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

{CH = 126.369, C3N = 596.779, CO = 345.13, NH = 381.107, OH = 341.591, C2O = 551.761, CC = 34.60, CN = 167.601, C2C = 60.20, C3C = 612.333, C2N = 313.895, CS = 214.021, SH = 36.30, NN = 317.912, C2S = 420.200, NO = 402.000, N2O = 637.217, P2O = 533.905, PO = 457.687}

with mean relative error merr = 0.056

example: for glycine measured (absolute) $H_{fm} = 2435.5$

calculated $H_f = \sum_i f_{b,i} E_{b,i} = 2499.9$, with the relative error

rerr(H_f) = 64.4/2435.5 = 0.026.

2.3.2. Empirical Model of Reaction Rate

In physical chemistry, the reaction rate of a reaction is described by the mass action law [17].

$$A + B\ (k_+ \rightarrow)\ C + D + \Delta E \quad \text{resp.} \quad A + B\ (\leftarrow k_-)\ C + D + \Delta E \tag{1}$$

$k_+ [A]^{n_A} [B]^{n_B} = k_- [C]^{n_C} [D]^{n_D}$ at equilibrium, where $[A]$ is concentration and n_A is the multiplicity of A, and ΔE is the reaction energy, the reaction is exothermic, when $\Delta E > 0$.

The time-dependent form of the mass action law is

$$\frac{\mathrm{d}C}{\mathrm{d}t}+\frac{\mathrm{d}D}{\mathrm{d}t}=k_{+}\left[A\right]^{n_A}\left[B\right]^{n_B} \tag{2}$$

The Arrhenius law for the reaction constant states that [17]

$$k_r = A_0 \exp\left(-E_a/kT\right) \tag{3}$$

where E_a is the activation energy, A_0 is the pre-exponential factor.

The reaction energy is the difference of the sum of the bond energies of input and output:

$$\Delta E = \sum_k Eo_{b,k} - \sum_k Ei_{b,k} \tag{4}$$

When concentration are measured in (dimensionless) *relative mole* (and not 1/mole as usual), the constant A_0 has the dimension 1/s, *i.e.* $A_0 = \frac{1}{t_0}$, 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_0 is only weakly temperature-dependent.

The rate constant A_0 depends on the diffusion constants and the critical length (mean free path) λ of the liquid $\left(A_{0+}\right)^{-1} = t_0 = \lambda^2/6\left(D_A + D_B\right)$ [17], where D is the diffusion constant.

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

For linear molecules, we can approximate r_0 by $r_0 = \sqrt{L_m d_m}$ [18], 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 $D = \frac{D_0}{\sqrt{\sum_k R_{b,k}}}$ [18], 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 than the arithmetic mean in the formula for t0 is a good model:

$$t_0 = \frac{\lambda^2}{3\sqrt{D_1 D_2}} \tag{5}$$

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

$$\lambda\left(\text{water}\right) = 1.62 \text{ nm}, \quad D_0\left(\text{water}\right) = 2.52\times10^{-9} \text{ m}^2/\text{s}$$

The resulting bond length in A are

H1HO = 0.0268, H3N1H = 10.10, CH = 0.19027, C3N = 0.0116, HH = 0.0102, CO = 0.0143, C1OH = 0.020, NH = 0.0101, OH = 0.3861, C2O = 0.0120, CC = 0.01540, CN = 0.0147, C2C = 0.0134, C2N = 0.0129, CS = 6.18268, NO = 0.0140, N2O = 0.01210, O2O = 12.10, N3N = 0.010, OO = 0.2257, D0 = 0.7681, λ = 2.04029

with mean relative error merr = 0.073.

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 corrected bond energies $\tilde{E}_{b,k} = f_{b,k} E_{b,k}$ in kJ/mol

{H1HO = 112.775, H3N1H = 130., HO1H = 292.07, CH = 58.5568, C3N = 939.947, HH = 2406.09, CO = 211.088, C1OH = 0.142, NH = 322.903, OH = 87.8274, C2O = 207.482, CC = 307.252, CN = 327.516, C2C = 622.483, C2N = 495.348, CS = 2221.8, SH = 3159.27, NO = 917.996, N2O = 204.21, O2O = 302.165, N3N = 1210.54, OO = 737.965}

with mean relative error merr = 0.066.

2.4. 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 with 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 **Figure 4** [19].

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 $r_n = r(x_n + \Delta x/2)$, $x_n = n\Delta x$, Δx 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:

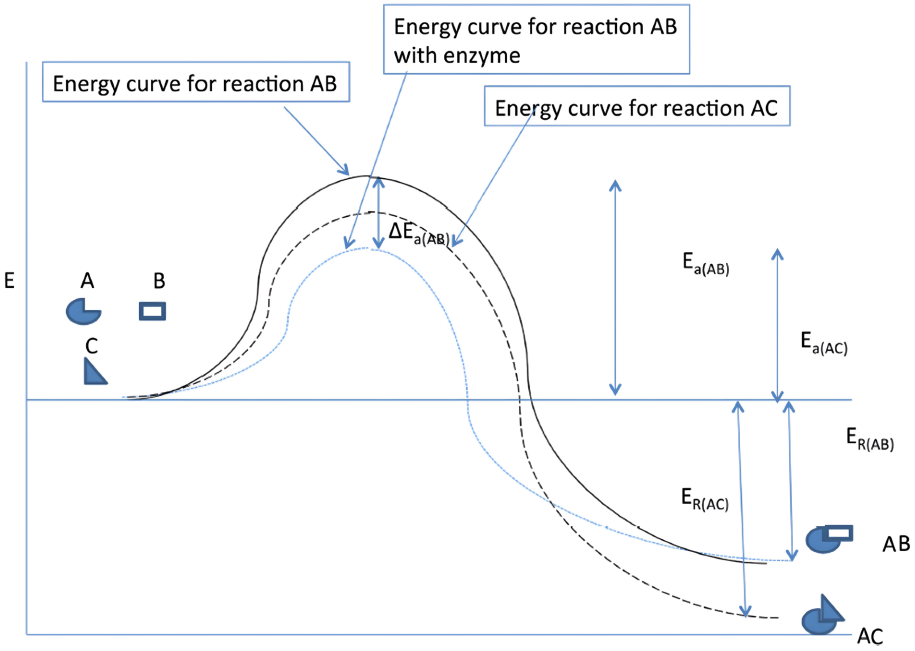


Figure 4. Information in living systems manifests through "temporal gradients". Here the system contains initially two substrates and one enzyme. In the absence of the enzyme, reaction $C \rightarrow G + H$ will proceed more rapidly because it has both lower final free energy and lower activation energy. However, the enzyme lowers the E_a for reaction $B \rightarrow E + F$. 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 $C \rightarrow G + H$.

$$H_{KL}(p \| r) = \sum_{n=1}^{N} p_n \ln\left[\frac{p_n}{r_n}\right] = \min \tag{6}$$

$\sum_n p_n = 1$, 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 $\sum_n r_n = 1$, with bond energies E'_n and the molecular energy $\sum_n E'_n = E'_b$, and $r_n = \frac{E'_n}{E'_b}$, E'_b is the molecular energy of the enzyme.

With Lagrange-multipliers we get the minimization problem

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

Differentiation ∂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 $\Lambda_1 = 1 + \ln\left(\frac{p_N}{r_N}\right)$ as 2 equations for p_N and Λ_1 and $p_n = r_n \exp(\Lambda_1 - 1)$

In this case, differentiation yields the (approximate) 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** [12]

$$2\mathrm{HCN} + 2\mathrm{H_2O} + 2\mathrm{H} \xrightarrow[\text{diglycine}]{} \text{glycine} + \mathrm{NH_3},$$

$$\text{where glycine} = \mathrm{OH{-}CO{-}CH_2{-}NH_2} \text{ and}$$

$$\text{diglycine} = \mathrm{OH{-}CO{-}CH_2{-}NH{-}CO{-}CH_2{-}NH_2}$$

We formulate the substrate-enzyme interaction pattern in the input as follows

HOH HCN HH $\mathrm{H_2O}$ HCN substrate $2\mathrm{HCN} + 2\mathrm{H_2O} + 2\mathrm{H}$

OH C=O HCH NH C=O HCH HNH enzyme diglycine

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

E_{in} = {H1HO, H1HO, CH, C3N, HH, H1HO, H1HO, CH, C3N}
{143.694, 143.694, 143.694, 165.472, 658.049, 143.694, 931.529, 143.694, 143.694, 143.694, 143.694, 165.472, 658.049}

enzyme bond energies in kJ/mol

E'_{in} = {C2O, C2O, CH, CH, NH, C2O, C2O, CH, CH, NH, NH}
{110.436, 110.436, 165.472, 165.472, 381.609, 110.436, 110.436, 165.472, 165.472, 381.609, 381.609}

The KL-weights are

p_{in} = {0.000751014, 0.990651, 0.00112518, 0.00112514, 0.0025965, 0.000750763, 0.000750619, 0.00112519, 0.00112512}

enzyme density values $r_{in} = E'_{in} \Big/ \sum_n E'_{in}$, {0.0743559, 0.0743559, 0.111411, 0.111411, 0.256934, 0.0743559, 0.0743559, 0.111411, 0.111411}

We formulate the substrate-enzyme interaction pattern in the output correspondingly

OH C=O HCH NH−C=O $\mathrm{CH_2{-}NH_2}$ enzyme diglycine

OH C=O HCH $\mathrm{NH_2}$ $\mathrm{NH_3}$ $\text{glycine} = \mathrm{OH{-}CO{-}CH_2{-}NH_2} + \mathrm{NH_3}$

E_{on} = {OH, CO, C2O, CC, CH, CH, CN, NH, NH, NH, NH, NH}
{257.109, 290.849, 110.436, 157.643, 165.472, 165.472, 35.7925, 381.609, 381.609, 381.609, 381.609, 381.609}

E'_{on} = {OH, CO, C2O, CC, CH, CH, CN, NH, C2O, CH, CH, NH}
{257.109, 290.849, 110.436, 157.643, 165.472, 165.472, 35.7925, 381.609, 110.436, 165.472, 165.472, 381.609}

enzyme density values $r_{on} = E'_{on} \Big/ \sum_n E'_{on}$, {0.107695, 0.121828, 0.0462586, 0.066032,

0.0693113, 0.0693113, 0.0149924, 0.159845, 0.0462586, 0.0693113, 0.0693113, 0.159845}

The KL-weights are

p_{on} = {0.000213839, 0.000242668, 0.000090888, 0.000129961, 0.00013647, 0.00013647, 0.998051, 0.000317807, 0.0000907091, 0.00013647, 0.00013647, 0.000317807}

The mean bond energies original and original activation energy are

$$E_{ibm} = \left(\sum_n E_{in}\right) \Big/ n_i = 350.37$$

Calculation of biochemical reactions

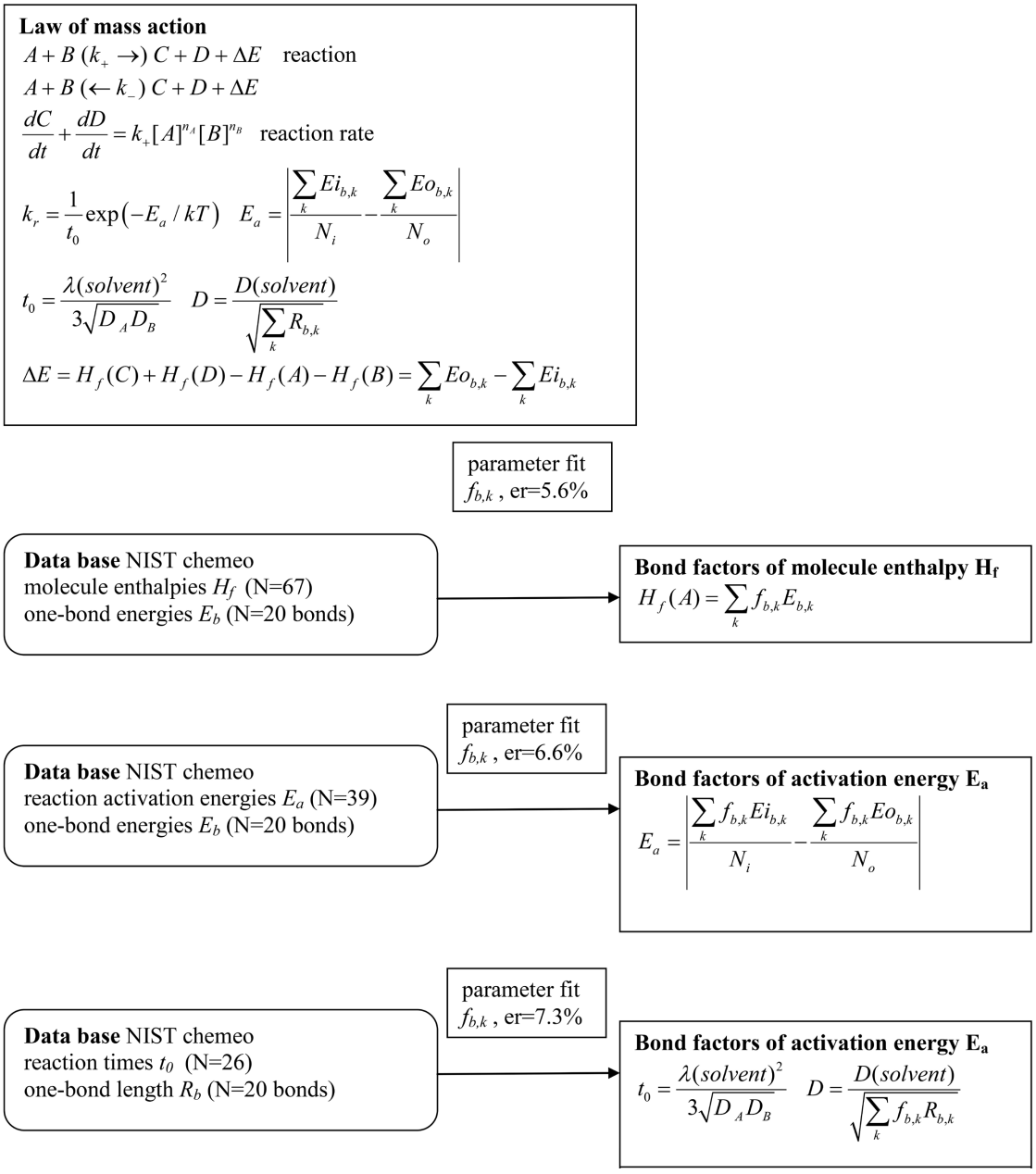


Figure 5. Schematic calculation of biochemical reactions ($N = n$ (elements)).

$$E_{obm} = \left(\sum_n E_{on}\right) \Big/ n_o = 257.57$$

$$E_a = E_{obm} - E_{ibm} = 92.80$$

The mean bond energies catalyzed, and catalyzed activation energy are

$$E'_{ibm} = \left(\sum_n p_{in} E_{in}\right) \Big/ n_i = 11.10$$

$$E'_{obm} = \left(\sum_n p_{on} E_{on}\right) \Big/ n_o = 2.79$$

$$E'_a = E'_{obm} - E'_{ibm} = 8.31$$

so the original activation energy is $E_a = 92.80$, and the catalyzed activation energy is $E'_a = 8.31$, and the attenuation ratio is $f_a = \frac{E_a}{E'_a} = 11.17$.

The calculation of biochemical reaction rates from 2.3.2 is described below in Figure 5.

The calculation of lock-and-key enzyme model from 2.4 is described below in Figure 6.

Catalytic action of enzyme lowers the activation energy E_a

Lock-and-key model of enzyme action

reacting molecule $\sum_n E_n = E(molec)$, bond energy E_n

enzyme $\sum_n r_n E'_n = E'(enzyme)$

weights molecule p_n variable, weights enzyme r_n,

$\sum p_n = 1 \quad \sum r_n = 1 \quad r_n = \frac{E'_n}{E'}$ fixed

minimization problem

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

$\min(p_n, \Lambda_1)$

weights p_n, Λ_1, mean bond energy:

$$\frac{1}{N}\sum_n p_n E_n = \frac{1}{N} E_1(molec) < \frac{1}{N} E(molec)$$

Example: glycine synthesis with enzyme diglycine

$$2HCN + 2H_2O + 2H \xrightarrow[diglycine]{} glycine + NH_3$$

input

	HOH	HCN	HH	$H2O$	HCN		substrate $2HCN + 2H_2O + 2H$
OH	$C=O$	HCH	NH	$C=O$	HCH	HNH	enzyme diglycine

output

OH	$C=O$	HCH	$NH-C=O$	CH_2-NH_2	enzyme diglycine
OH	$C=O$	HCH	NH_2	NH_3	$glycine = OH - CO - CH_2 - NH_2 + NH_3$

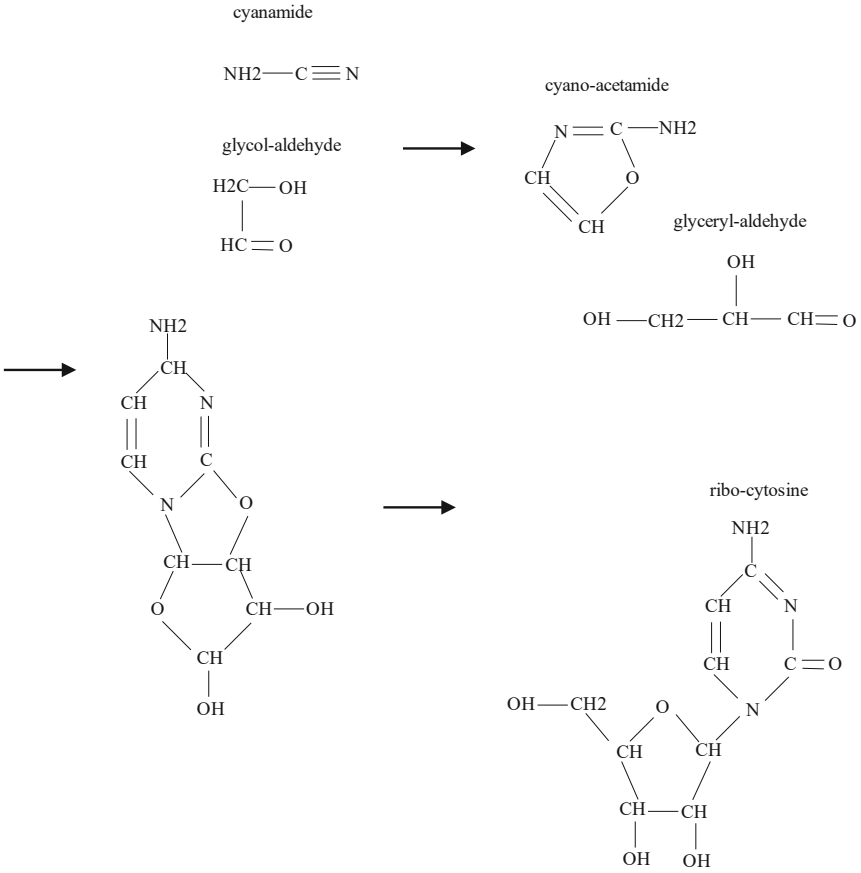
mean bond energy input $E_{bim} = 350.37\ kJ/mole$ output $E_{bom} = 257.57$
act.energy $E_a = E_{bim} - E_{bom} = 92.81$
catalyzed mean bond energy input $E'_{bim} = 11.10\ kJ/mole$ output $E'_{bom} = 2.79$
catalyzed act.energy $E'_a = E'_{bim} - E'_{bom} = 8.31$

Figure 6. Schematic catalytic action of enzyme.

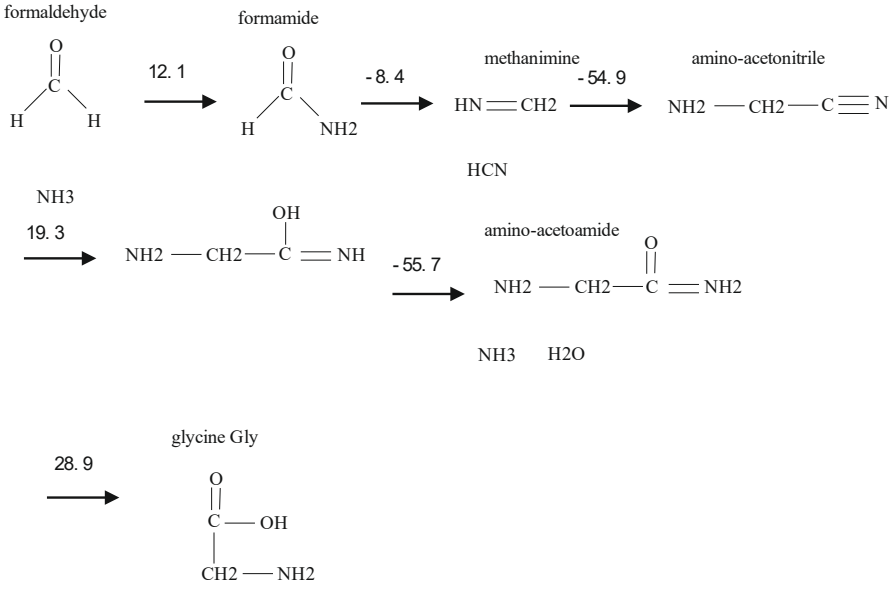
2.5. Synthesis Reactions in Prebiotic HCN-Chemistry

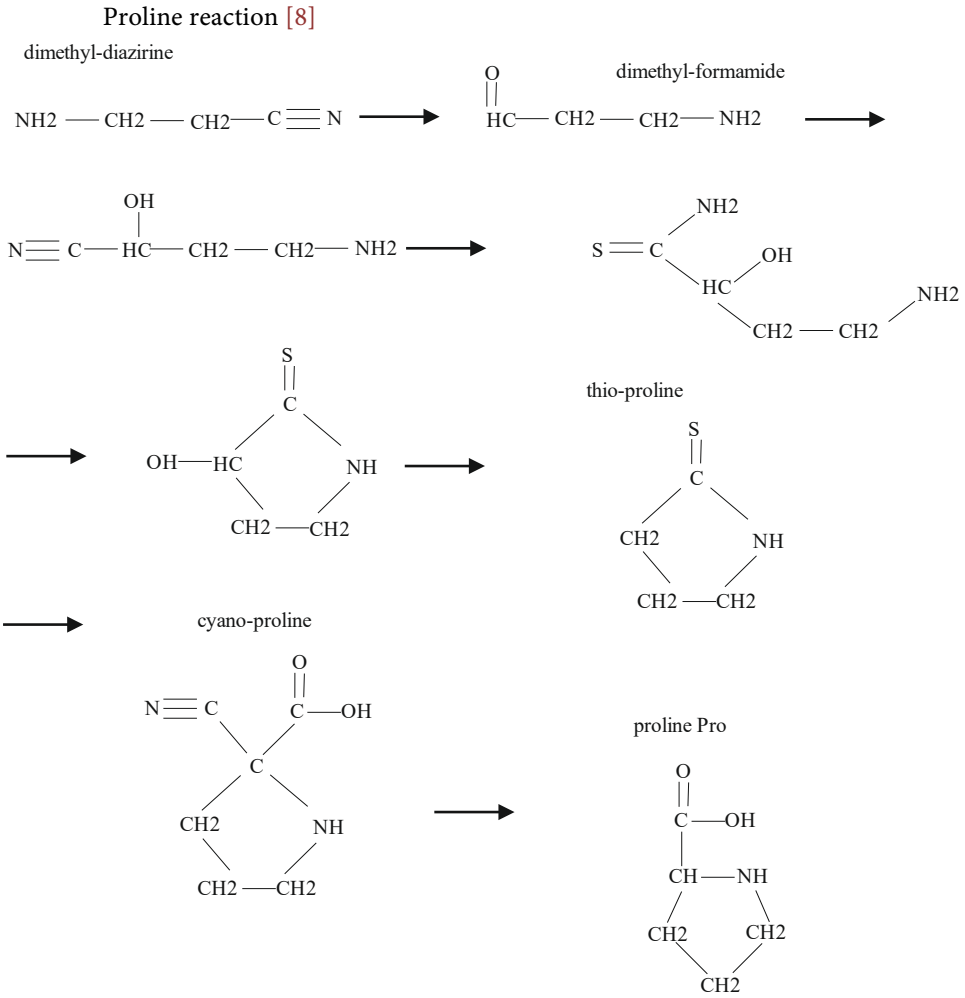
We present here three prebiotic reactions of the HCN-chemistry, which generate important amino acids and nucleobases of the proto-life scenario described in the following section [20]. These reactions have been carried out in the lab by Patel *et al.* [8] and Das *et al.* [9].

Cytosine reaction [8]



Glycine reaction [9], including energy output in kJ/mol (negative = released)





2.6. 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 [8] [9], the binary genetic proto-code proposed by Carter-Wills [10] and Rodriguez *et al.* [11], and the pyrite (iron-sulfur) energy cycle proposed by Wächtershäuser [21].

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 xHCN + $2H_2O$ + xH of the HCN chemistry with hydrogen provided by the energy cycle.
- the maximum activation energy E_a of the spontaneous HCN-chemistry is 40 kJ/mol [8], which fits very well with the reaction energy ΔE = 41 kJ/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 compo-

nents.

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 cysteine: 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 max-4-bit-codons (guanine Gua or cytosine Cyt) coding for 2 amino acids (glycine Gly or proline Pro). It is a 1-1 code, *i.e.* a gene has only one codon, so there is no need for several steps of "sequencer" tRNA's as in the full DNA-code, the *peptide synthesis in the proto-life-cycle is a 1-step process*. The codon = gene e.g. GG for Gly is represented by the corresponding poly-nucleobase PNB GuaGua, which "grabs" the corresponding amino acids (in this case Gly) via H-bonds (consuming energy) and forms the corresponding "stacked" poly-nucleic-acid (PNA), which after completion spontaneously splits-off the complete peptide (releasing energy).

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 with max-3-bit codons, where 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 NH_2-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 and 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 is available) to CO

2.6.1. 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 [22]) and 2 PNA's for 5 amino acids, 1 lipid and 2 nucleobases.

glycine gene GlyGly, coded GG, produces glycine Gly

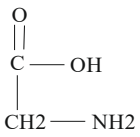
synthesis reaction

$$\left(\mathrm{HCN}+2\mathrm{H_2O}\right)+\left(\mathrm{HCN}+2\mathrm{H}\right)\rightarrow \mathrm{Gly}+\mathrm{NH_3}$$

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 NH_2	GlyGly



cysteine gene GlyGlyGly, coded GGG, produces cysteine Cys (catalyzes pyrite reaction)

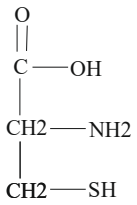
synthesis reaction

$$\left(\mathrm{HCN}+2\mathrm{H_2O}\right)+\left(\mathrm{HCN}+5\mathrm{H}\right)+\left(\mathrm{HCN}+\mathrm{H_2S}\right)\rightarrow \mathrm{Cys}+2\mathrm{NH_3}$$

molecule enthalpy H_f(Cys) = −2780.1 kJ/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 NH_2	GlyGlyGly



cysteine gene GlyGlyGly, coded GGG, produces glycerol-1-phosphate with phosphorous acid (Glyc1Ph forms membranes)

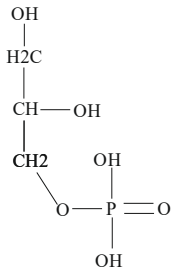
synthesis reaction

$$3\mathrm{HCN}+2\mathrm{H_2O}+8\mathrm{H}+\mathrm{H_3PO_4}\rightarrow \mathrm{Glyc1Ph}+3\mathrm{NH_3}$$

molecule enthalpy H_f(Glyc1Ph) = −5033.8 kJ/mol

enzyme-substrate alignment

HH OH-H HCN HH OH-H HCN HH HPO_4 HCN 2HH	precursors
OH CO HCH NH CO HCH NH CO HCH NH_2	GlyGlyGly



proline gene ProPro, coded CC, produces proline Pro

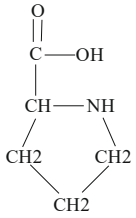
synthesis reaction

$$\left(\mathrm{HCN}+2\mathrm{H_2O}\right)+4\mathrm{HCN}+12\mathrm{H}\rightarrow \mathrm{Pro}+4\mathrm{NH_3}$$

molecule enthalpy H_f(Pro) = −3275.2 kJ/mol

enzyme-substrate alignment (N-CO is the peptide bond between amino acids, the penta-ring is split at CH-CH_2)

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

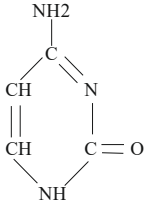
synthesis reaction

$$4\mathrm{HCN}+\mathrm{H_2O}+2\mathrm{H}\rightarrow \mathrm{Cyt}+\mathrm{NH_3}$$

molecule enthalpy H_f(Cyt) = −2799. kJ/mol

enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2, 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

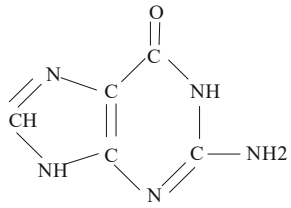
synthesis reaction

$$5\mathrm{HCN}+\mathrm{H_2O}\rightarrow \mathrm{Gua}+\mathrm{H_2}$$

molecule enthalpy H_f(Gua) = −3753.3 kJ/mol

enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2, marked ---)

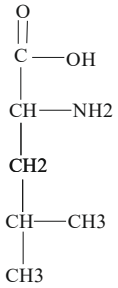
OH-H HCN OH-H HCN HCN HCN HCN precursors
OH CO HCH NH CO CH NH HCH HCH HCH GlyPro



leucine gene GlyGlyGlyGly, coded GGGG, produces leucine Leu
synthesis reaction

$$\left(\mathrm{HCN}+2\mathrm{H_2O}\right)+5\mathrm{HCN}+18\mathrm{H}\rightarrow \mathrm{Leu}+5\mathrm{NH_3}$$

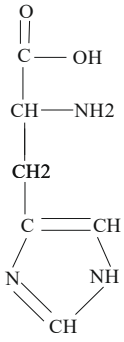
molecule enthalpy H_f(Leu) = −4181.7 kJ/mol
enzyme-substrate alignment
HH OH-H HCN 2HH OH-H HCN 2HH HCN HCN 2HH HCN HCN 2HH
precursors
OH CO HCH NH CO HCH NH CO HCH NH CO HCH NH_2 GlyGlyGlyGly



histidine gene ProGlyGly, coded CGG, produces histidine His
synthesis reaction

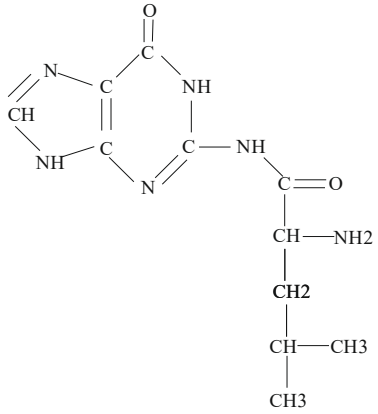
$$\left(\mathrm{HCN}+2\mathrm{H_2O}\right)+5\mathrm{HCN}+8\mathrm{H}\rightarrow \mathrm{His}+3\mathrm{NH_3}$$

molecule enthalpy H_f(His) = −4259.7 kJ/mol
enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2)
2HH OH-H HCN HCN HCN HCN precursors
OH CO CH N HCH HCH HCH Pro
CO HCH NH CO HCH NH_2 GlyGly
2HH HCN 2HH OH-H HCN 2HH precursors



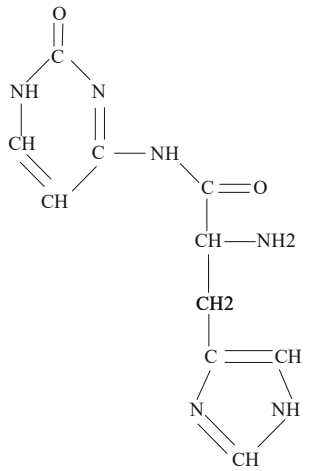
The **ligase and PNA for Gly** is Leu-Gua bound by the peptide bond –NH-C=

O-



$$\text{Leu} + \text{Gua} \rightarrow \text{LeuGua} + H_2O$$

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



$$\text{His} + \text{Cyt} \rightarrow \text{HisCyt} + H_2O$$

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 ad-joining 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 model 1 (see reaction table below),

$$\text{Gly} + \text{Gua} \rightarrow \text{GlyGua} + H_2O\,,\ \Delta E = 339\ \text{kJ/mol},\ E_a = 14\ \text{kJ/mol}$$

$$\text{Pro} + \text{Cyt} \rightarrow \text{ProCyt} + H_2O\,,\ \Delta E = -44.3\ \text{kJ/mol},\ E_a = 12.6\ \text{kJ/mol}$$

the Pro-reaction is endothermic (needs energy).

In the corresponding model2-reactions

$$\text{Leu} + \text{Gua} \rightarrow \text{LeuGua} + H_2O\,,\ \Delta E = 339\ \text{kJ/mol},\ E_a = 10\ \text{kJ/mol}$$

$$\text{His} + \text{Cyt} \rightarrow \text{HisCyt} + H_2O\,,\ \Delta E = 109.4\ \text{kJ/mol},\ E_a = 11.9\ \text{kJ/mol}$$

the His-reaction is exothermic, and both activation energies are lower than with their model1-counterparts.

2.6.2. Simplified Model1 Proto-Code with 3 Amino Acids

One gets a simpler version of the proto-code if 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 binary 3-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 [22]) 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

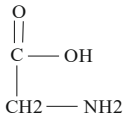
$$\left(\text{HCN} + 2H_2O\right) + \left(\text{HCN} + 2\text{H}\right) \rightarrow \text{Gly} + NH_3$$

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 NH_2 GlyGly



cysteine gene enzyme GlyGlyGly, coded GGG, produces cysteine Cys (catalyzes pyrite reaction)

synthesis reaction

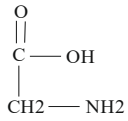
$$\left(\text{HCN} + 2H_2O\right) + \left(\text{HCN} + 4\text{H}\right) + \left(\text{HCN} + H_2S\right) \rightarrow \text{Cys} + 2NH_3$$

molecule enthalpy ΔH(Cys) = −534.1 kJ/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 NH_2 GlyGlyGly



cysteine gene enzyme GlyGlyGly, coded GGG, produces glycerol-1-phosphate with phosphorous acid (Glyc1Ph forms membranes)

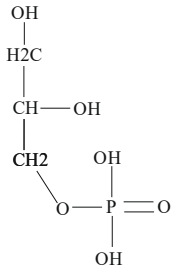
synthesis reaction

$$3HCN + 2H_2O + 8H + H_3PO_4 \rightarrow Glyc1Ph + 3NH_3$$

molecule enthalpy ΔH(Glycerol) = −577.9 kJ/mol

enzyme-substrate alignment

HH OH-H HCN HH OH-H HCN HH HPO_4 HCN 2HH	precursors
OH CO HCH NH CO HCH NH CO HCH NH_2	GlyGlyGly



proline gene enzyme ProPro, coded CC, produces proline Pro

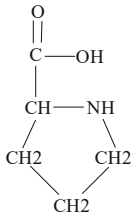
synthesis reaction

$$\left(HCN + 2H_2O\right) + 4HCN + 12H \rightarrow Pro + 4NH_3$$

molecule enthalpy ΔH(Pro) = −366.2 kJ/mol

enzyme-substrate alignment (N-CO is the peptide bond between amino acids, the penta-ring is split at CH-CH_2)

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

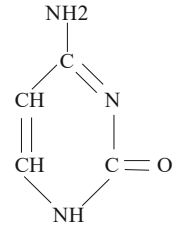
synthesis reaction

$$4HCN + H_2O + 2H \rightarrow Cyt + NH_3$$

molecule enthalpy ΔH(Cyt) = −221.3 kJ/mol

enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2, 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

$$5\mathrm{HCN} + \mathrm{H_2O} \rightarrow \mathrm{Gua}$$

molecule enthalpy ΔH(Gua) = −183.9 kJ/mol

enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2, 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

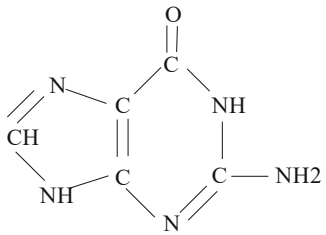
synthesis reaction

$$5\mathrm{HCN} + \mathrm{H_2O} \rightarrow \mathrm{Gua}$$

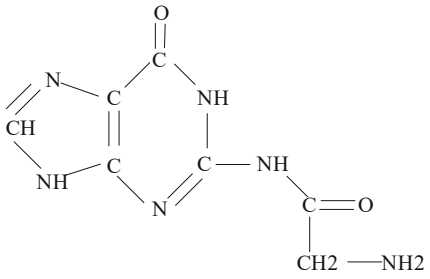
molecule enthalpy ΔH(Gua) = −183.9 kJ/mol

enzyme-substrate alignment (the penta-ring of Pro is split at CH-CH_2, marked ---)

HCN HCN HCN HCN	precursors
CH N HCH HCH HCH	ProGly
CO HCH HNH	precursors
OH-H HCN	ProGly

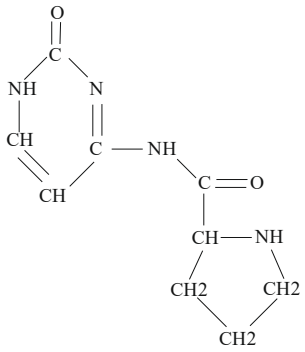


The **ligase and PNA for Gly** is GlyGua bound by the peptide bond –NH-C=O-



$$\mathrm{Gly} + \mathrm{Gua} \rightarrow \mathrm{GlyGua} + \mathrm{H_2O}$$

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



$$\mathrm{Pro} + \mathrm{Cyt} \rightarrow \mathrm{ProCyt} + \mathrm{H_2O}$$

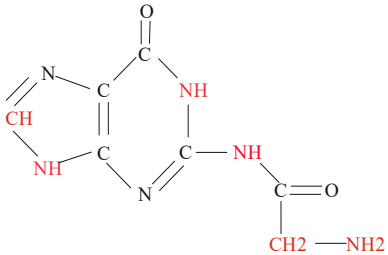
The peptide bond, which forms poly-peptides, has a bond energy of 8 - 16 kJ/mol.

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.

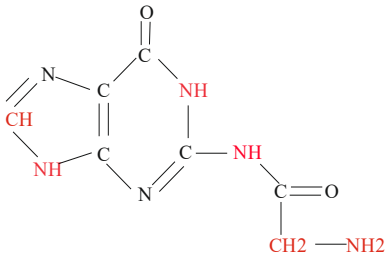
During **polymerization of PNA** H-bonds form in those sites, which have H-atoms.

GlyGuaGlyGua has 6 vertical H-bonds (in red)



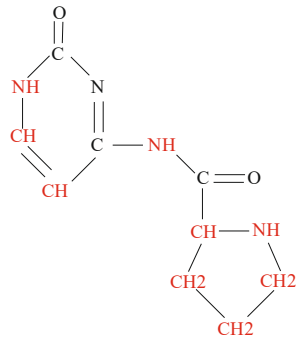
$$2\mathrm{GlyGua} \rightarrow \mathrm{GlyGuaGlyGua} + 6(\mathrm{H}-)$$

GlyGuaGlyGuaGlyGua has 12 vertical H-bonds (in red)



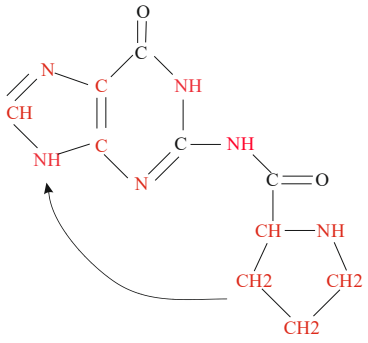
$$3\mathrm{GlyGua} \rightarrow \mathrm{GlyGuaGlyGuaGlyGlua} + 12(\mathrm{H}-)$$

ProCytProCyt has 9 vertical H-bonds (in red)



$$2\text{ProCyt} \rightarrow \text{ProCytProCyt} + 9(\text{H}-)$$

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



$$\text{GlyGua} + \text{ProCyt} \rightarrow \text{GlyGuaProCyt} + 13(\text{H}-)$$

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.

$$\text{GlyGuaGlyGua} + \Delta E \rightarrow \text{GlyGly} + \text{GuaGua} + \text{pb}(\text{GlyGly}) - 2\text{pb}(\text{PNA}) - 3(\text{H}-)$$

$$\text{ProCytProCyt} + \Delta E \rightarrow \text{ProPro} + \text{CytCyt} + \text{pb}(\text{ProPro}) - 2\text{pb}(\text{PNA}) - 6(\text{H}-)$$

$$\text{GlyGuaProCyt} + \Delta E \rightarrow \text{GlyPro} + \text{GuaCyt} + \text{pb}(\text{GlyPro}) - 2\text{pb}(\text{PNA}) - 9(\text{H}-)$$

$$\text{GlyGuaGlyGuaGlyGua} + \Delta E \rightarrow \text{GlyGlyGly} + \text{GuaGuaGua} + 2\text{pb}(\text{GlyGly}) - 3\text{pb}(\text{PNA}) - 6(\text{H}-)$$

2.7. Reactions of Model1 Proto-Code

The reactions of the model1-proto-code fall into 9 categories (**Table 1**, **Table 2**)

1) Catalyzed synthesis of basic compounds and energy reaction (pyrite reaction)

The basic compounds (amino acids and nucleobases) are synthesized from precursors HCN, H_2O, H_2, H_2S, 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*:

Table 1. Reactions of model1 proto-code [12] (ΔE reaction energy (kJ/mol), E_a original activation energy (kJ/mol), E_c catalyzed activation energy (kJ/mol), reaction time t_0 (ns), n_B number of bonds).

Reaction	Type	Enzyme	ΔE	E_a	E_c	t_0	n_B
Catalyzed basic compound synthesis							
$2HCN + 2H_2O + H_2 \rightarrow Gly(COOH{-}CH_2NH_2) + NH_3$	1	GlyGly	223	92.8	8.3	0.70	25
$5HCN + 2H_2O + 6H_2 \rightarrow Pro(C_5H_9NO_2) + 4NH_3$	1	ProPro	741.7	256.1	0.53	0.47	53
$3HCN + 2H_2O + 2H_2 + H_2S \rightarrow Cys(COOH{-}CHNH_2{-}CH_2{-}SH) + 2NH_3$	1	GlyGlyGly	1232.4	144.3	0.94	0.74	39
$4HCN + H_2O + H_2 \rightarrow Cyt(C_4H_5N_3O) + NH_3$	1	Pro	403.5	140.2.	17.5	0.76	28
$5HCN + H_2O \rightarrow Gua(C_5H_5N_5O) + H_2$	1	GlyPro	505.	145.9	6.56	0.99	36
$5HCN + H_2O \rightarrow Gua(C_5H_5N_5O) + H_2$	1	ProGly	505.	145.9	27.6	0.99	36
$FeS + H_2S \rightarrow FeS_2 + H_2$	1	Cys	41.	160.6	0.02	3.45	16
$Cyt(C_4H_5N_3O_3) + HCN + 6H_2 \rightarrow Pro(C_5H_9NO_2) + H_2O + 2NH_3$	1		582.9	21.7		1.02	16
$3HCN + 2H_2O + 4H_2 + H_3PO_4 \rightarrow Glyc1ph + 3NH_3$	1	GlyGlyGly	395.5	139.5	6.6	0.99	27
simple PNA polymerization							
$Gly + Gly + GuaGua \rightarrow GlyGuaGlyGua(6H-) + 2H_2O$	5		−94.2	16.4		1.53	53
$Gly + Gly + Gly + Gua3 \rightarrow GlyGua3(12H-) + 3H_2O$	5		−109.8	16.5		1.60	81
$Pro + Pro + CytCyt \rightarrow ProCytProCyt(9H-) + 2H_2O$	5		692.8	5.9		2.26	63
$Gly + Pro + GuaCyt \rightarrow GlyGuaProCyt(13H-) + 2H_2O$	5		−146.5	15.1		1.89	61
$Gly + Pro + CytGua \rightarrow ProCytGlyGua(13H-) + 2H_2O$	5		−146.5	15.1		1.89	61
$Gly + Pro + GuaCyt \rightarrow GlyCytProGua(13H-) + 2H_2O$	5		−146.5	15.1		1.89	61
PNB polymerization							
$Cyt + Cyt \rightarrow CytCyt(3H-)$	2		491.	5.0		1.31	26
$Gua + Gua \rightarrow GuaGua(3H-)$	2		705.3	2.9		1.14	34
$Gua + Gua + Gua \rightarrow Gua3(6H-)$	2		1089.5	3.7		1.14	51
$Gua + Cyt \rightarrow GuaCyt(6H-)$	2		661.2	6.9		1.22	30
$Cyt + Gua \rightarrow CytGua(6H-)$	2		661.2	6.9		1.22	30
peptide formation							
$GlyGuaGlyGua + 2H_2O \rightarrow GuaGua(3H-) + GlyGly + H_2O$	6		661.6	12.5		2.46	58
$ProCytProCyt + 2H_2O \rightarrow CytCyt(3H-) + ProPro + H_2O$	6		566.1	7.0		3.61	62
$GlyGuaProCyt + 2H_2O \rightarrow GuaCyt(6H-) + GlyPro + H_2O$	6		308.6	9.2		3.05	58
$ProCytGlyGua + 2H_2O \rightarrow CytGua(6H-) + ProGly + H_2O$	6		1008.	3.0		3.05	58
$GlyGuaGlyGuaGlyGua + 3H_2O \rightarrow GuaGuaGua(8H-) + GlyGlyGly + 2H_2O$	6		640.6	8.1		1.13	90
multiple PNA polymerization							
$GlyGua + GlyGua \rightarrow GlyGuaGlyGua(6H-)$	4		126.	3.0		1.66	48
$3GlyGua \rightarrow (GlyGua)3(12H-)$	4		252.	3.9		1.66	72
$ProCyt + ProCyt \rightarrow ProCytProCyt(9H-)$	4		189.	4.3		1.66	48
$GlyGua + ProCyt \rightarrow GlyGuaProCyt(13H-)$	4		273.	5.80		1.66	48
$GlyGua + ProCyt \rightarrow ProCytGlyGua(13H-)$	4		273.	5.80		1.66	48

Continued

PNA building					
Gly + Gua → GlyGua + H_2O	3	339.4	14.6	1.38	26
Gly + Cyt → GlyCyt + H_2O	3	1156.	17.2	1.47	22
Pro + Cyt → ProCyt + H_2O	3	−44.3	12.6	1.79	30
Pro + Gua → ProGua + H_2O	3	62.8	11.1	1.67	34
Leu + Gua → LeuGua + H_2O	3	339.	10.	1.79	38
His + Cyt → HisCyt + H_2O	3	109.4	11.9	1.70	32
diverse synthesis and energy cycles					
CO_2 + NH_3 + H_2 → HCN + $2H_2O$	7	−90.9	74.2	3.45	6
CO_2 + NH_3 + SH_2 → HCN + $2H_2O$ + S	7	−111.5	6.5	3.45	7
FeS + H_2S → FeS_2 + H_2	7	41.	160.6	3.45	3
CO_2 + H_2 → CO + H_2O	7	90.1	115.8	0.23	3
CO + H_2O → HCOOH	7	548.8	73.3	0.20	3
Gly + H_2O → Gle(HCO−HCOH−H_2COH) + NH_3	7	346.2		1.62	9
basic compounds decay by hydrolysis C=O					
Gly(COOH−CH_2NH_2) + H_2O → OH−CH_2NH_2 + HCOOH	8	−335.9	9.0	1.66	11
Cys(COOH−CHNH_2−CH_2−SH) + H_2O → OH−CHNH_2−CH_2−SH + HCOOH	8	−157	6.6	5.29	15
Pro(COOH−ring(−NH−)) + H_2O → OH−ring() + HCOOH	8	−612.4	5.2	2.45	19
Cyt(ring(−C=O−)) + H_2O → ring() + HCOOH	8	−89.4	20.5	1.30	15
Gua(ring(−C=O−ring(−NH−CH−N−))) + H_2O → ring(−ring(−NH−CH−N−)) + HCOOH	8	−15.1	5.2	1.14	19
direct peptide polymerization					
Gly + Gly → GlyGly + H_2O	9	−117.7	64.8	1.66	18
Gly + Gly + Gly → GlyGlyGly + $2H_2O$	9	−99.5	83.8	1.66	27
Pro + Pro → ProPro + H_2O	9	−1097.7	27.9	2.45	34
Gly + Pro → GlyPro + H_2O	9	−258.	43.7	2.02	26
Gly + Pro → ProGly + H_2O	9	−258.	43.7	2.02	26
GlyGly + Gly → GlyGly + H_2O	9	+18.2	48.	1.77	24

Table 2. Model1 genetic code.

code	product	enzyme	RNA code
GG	Gly	Gly2	GGx
GGG	Cys	Gly3	UGx
CC	Pro	Pro2	CCx
C	Cyt	Pro	
GC	Gua	GlyPro	

$2HCN + 2H_2O + H_2 \rightarrow Gly(COOH{-}CH_2NH_2) + NH_3$, enzyme GlyGly, ΔE = 223 kJ/mol, E_a = 94.8 kJ/mol, E_c = 7.6 kJ/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:

$FeS + H_2S \rightarrow FeS_2 + H_2$, enzyme Cys, ΔE = 41 kJ/mol, E_a = 160.6 kJ/mol, E_c = 0.02 kJ/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

$t_r = t_0 \exp\left(\frac{E_a}{kT}\right) = 3.45\times10^{-9}\times1.37\times10^{21} = 5.32\times10^{12}\text{ s} = 1.69\times10^{5}\text{ a}$).

The synthesis reactions have high reaction energy (ΔE = 100...1200 kJ/mol), high activation energy (E_a = 100...200 kJ/mol), and do not run in water without enzymes.

The "self-catalyzing" synthesis of proline via cytosine (ring) has a very low activation energy

$Cyt(C_4H_5N_3O_3) + HCN + 6H_2 \rightarrow Pro(C_5H_9NO_2) + H_2O + 2NH_3$, ΔE = 582.9 kJ/mol, E_a = 21.7 kJ/mol, n_B = 16

The basic compounds synthesis reactions are all exothermic, and have high activation energy E_a = 92...145 kJ/mol (using cytosine-proline synthesis). when running without catalysis, the maximum reaction time is (*Gua*)

$t_r = t_0 \exp\left(\frac{E_a}{kT}\right) = 0.99\times10^{-9}\times1.59\times10^{19} = 1.57\times10^{10}\text{ s} = 0.50\times10^{3}\text{ a}$

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 H_3PO_4, catalyzed by triglycine *GlyGlyGly*

$3HCN + 2H_2O + 4H_2 + H_3PO_4 \rightarrow Glyc1ph + 3NH_3$, ΔE = 395.5 kJ/mol, E_a = 139.5 kJ/mol, E_c = 6.6 kJ/mol, n_B = 27.

2) Nucleobase polymerization

Nucleobases form stacks bound by H-bonds (poly-nucleobase = PNB), these reactions are spontaneous, strongly exothermic (ΔE ~600 kJ/mol) and have low activation energy (E_a = 5…10 kJ/mol).

$Cyt + Cyt \rightarrow CytCyt(3H-)$ CytCyt is bound by 3 H-bonds, ΔE = 491 kJ/mol, E_a = 5 kJ/mol

3) Simple PNA formation

Simple peptide-nucleic-acid (PNA) form from an amino acid and a nucleobase

$Gly + Gua \rightarrow GlyGua + H_2O$, ΔE = 339 kJ/mol, E_a = 14 kJ/mol

$Pro + Cyt \rightarrow ProCyt + H_2O$, ΔE = −44.3 kJ/mol, E_a = 12.6 kJ/mol

These reactions are low exo- or low endothermic and have low activation energy.

4) Direct PNA polymerization

Simple PNA's form chains

$$\mathrm{GlyGua} + \mathrm{GlyGua} \rightarrow \mathrm{GlyGuaGlyGua}\left(6\mathrm{H}-\right),\ \Delta E = 126\ \mathrm{kJ/mol},\ E_a = 3\ \mathrm{kJ/mol}$$

5) PNA formation from amino acids and nucleobases

Two amino acids and a simple pNB form a PNA

$$\mathrm{Gly} + \mathrm{Gly} + \mathrm{GuaGua} \rightarrow \mathrm{GlyGuaGlyGua}\left(+2\mathrm{H}-\right) + 2\mathrm{H_2O},\ \Delta E = 468\ \mathrm{kJ/mol},\ E_a = 10\ \mathrm{kJ/mol}$$

These reactions are mostly endothermic and have a low activation energy.

6) Peptide forming

$$\mathrm{GlyGuaGlyGua} + 2\mathrm{H_2O} \rightarrow \mathrm{GuaGua}\left(+4\mathrm{H}-\right) + \mathrm{GlyGly} + \mathrm{H_2O},\ \Delta E = 682\ \mathrm{kJ/mol},\ E_a = 11.3\ \mathrm{kJ/mol}$$

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), whereas 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] Chap. 3.4), there is evidence of the presence of HCN, along with methane CH_4 and ammonia NH_3 in the early Earth atmosphere. But there are also cyanide synthesis reactions, which could produce HCN under volcanic pools or in hydrothermal vents.

$$\mathrm{CO_2} + \mathrm{NH_3} + \mathrm{H_2} \rightarrow \mathrm{HCN} + 2\mathrm{H_2O},\ \Delta E = -90.9\ \mathrm{kJ/mol},\ E_a = 74.2\ \mathrm{kJ/mol}$$

$$\mathrm{CO_2} + \mathrm{NH_3} + \mathrm{SH_2} \rightarrow \mathrm{HCN} + 2\mathrm{H_2O} + \mathrm{S},\ \Delta E = -111.5\ \mathrm{kJ/mol},\ E_a = 6.5\ \mathrm{kJ/mol}$$

They have as reducing component H_2 or SH_2, 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 SH_2-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:

$$\mathrm{Gly} + \mathrm{H_2O} \rightarrow \mathrm{Gle}\left(\mathrm{HCO{-}HCOH{-}H_2COH}\right) + \mathrm{NH_3},\ \Delta E = 346.2\ \mathrm{kJ/mol},\ E_a = 14\ \mathrm{kJ/mol}$$

As for energy cycles, there is the mentioned pyrite reaction

$$\mathrm{FeS} + \mathrm{H_2S} \rightarrow \mathrm{FeS_2} + \mathrm{H_2},\ \Delta E = 41\ \mathrm{kJ/mol},\ E_a = 160.6\ \mathrm{kJ/mol}$$

but also the CO_2-reduction by hydrogen, which was present in the prebiotic Earth atmosphere

$$\mathrm{CO_2} + \mathrm{H_2} \rightarrow \mathrm{CO} + \mathrm{H_2O},\ \Delta E = 90.1\ \mathrm{kJ/mol},\ E_a = 115.8\ \mathrm{kJ/mol}$$

and oxidation of carbon monoxide CO in water, which is still used by methanogen bacteria [23]

$\mathrm{CO}+\mathrm{H_2O}\rightarrow\mathrm{HCOOH}$, $\Delta E=548.8$ kJ/mol, $E_a=33.3$ kJ/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 $T>185^{\circ}\mathrm{C}$ 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 [24].

The decay via hydrolysis of the basic compounds functions via break-off of the C = O radical, which all of them contain, with the 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:

$\mathrm{Gly}\left(\mathrm{COOH}-\mathrm{CH_2NH_2}\right)+\mathrm{H_2O}\rightarrow\mathrm{OH}-\mathrm{CH_2NH_2}+\mathrm{HCOOH}$, $\Delta E=-335.9$ kJ/mol, $E_a=9.0$ kJ/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):

$\mathrm{Gly}+\mathrm{Gly}\rightarrow\mathrm{GlyGly}+\mathrm{H_2O}$, $\Delta E=-117.7$ kJ/mol, $E_a=64.8$ kJ/mol

$\mathrm{Pro}+\mathrm{Pro}\rightarrow\mathrm{ProPro}+\mathrm{H_2O}$, $\Delta E=-1097.7$ kJ/mol, $E_a=27.9$ kJ/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

$\left(\Delta E\left(\mathrm{Gly}+\mathrm{Gly}\rightarrow\mathrm{GlyGly}+\mathrm{H_2O}\right)\approx 3E\left(\mathrm{FeS}+\mathrm{H_2S}\rightarrow\mathrm{FeS_2}+\mathrm{H_2}\right)=123\ \mathrm{kJ/mol}\right)$, *i.e.* very slowly even with a catalyzed energy reaction.

The opposite reactions like

$$\mathrm{GlyGly}+\mathrm{H_2O}\rightarrow\mathrm{Gly}+\mathrm{Gly}$$

are exothermic *decay reactions of peptides* with a decomposition temperature T_d, where $t_0\exp\left(E_a/kT_d\right)=1\,\mathrm{s}$, so for $t_0\approx 10^{-9}\,\mathrm{s}$ we get approximately for Gly: $kT_d=E_a/\log\left(10^9\right)$ and $T_d=350\ \mathrm{K}=77^{\circ}\mathrm{C}$, which is in good agreement with the known decomposition temperature of poly-peptides (75˚C - 85˚C).

Therefore the PNA-controlled peptide formation sequence type2 → type5 → type6 is really necessary for the molecular evolution of life from the HCN-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 H_2S, CO_2-reduction or CO-oxidation.

With seed-concentration for the basic compounds of $c=10^{-6}$, and remaining concentration from [23] given below, we get the energy production of the 3

energy cycles [12]

$E_p(\mathrm{H_2S}) = 0.0045\ \mathrm{kJ/s \cdot mol}$, $E_p(\mathrm{CO_2}) = 5.3\times10^{-7}\ \mathrm{kJ/s \cdot mol}$,
$E_p(\mathrm{CO}) = 0.00062\ \mathrm{kJ/s \cdot mol}$,

and using this, for the effective build-up time of $\mathrm{Gly + Gly \rightarrow GlyGly + H_2O}$ with the H_2S-pyrite reaction

$t_c(2\mathrm{Gly}) = \dfrac{t_r(2\mathrm{Gly})}{c(\mathrm{Gly})}\dfrac{\Delta E(2\mathrm{Gly})}{t_r(2\mathrm{Gly})}\dfrac{1}{E_p} = 1.0\times10^9\ \mathrm{s} = 31.7\ \mathrm{a}$ with the effective reaction time $t_r = t_0 \exp\left(\dfrac{E_a}{kT}\right)$.

2.8. Numerical Simulation of Terrestrial Proto-Life Evolution

2.8.1. Diffusion and Convection

Molecular flow plays 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 (δ_θ 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 μm, and the time scale is $t \sim 1$ s [25], here diffusion is a realistic mechanism.

For diffusion, we have Fick's law [25]

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2} \tag{7}$$

For the gradient-driven **osmosis through a membrane**, the following relation holds [26]:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{A}{dV}KD\Delta c \tag{8}$$

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, Δc is the difference in relative concentrations at the membrane. K depends on the molecule and the material of the membrane, and has values (for lipids and octanol) in the range log K_{ow} = −4...6.5 [24].

The **thermal convection** flow, driven by a thermal difference of $\Delta T \sim 80$ K is much faster [25] [27].

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 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 $Nu(L = 10\text{ m}, \Delta T = 80\text{ K}) = 160$ $\delta_\theta = 6.3$ cm [25] [27] 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: ρ density, ΔT temperature difference, Δc concentration difference, L edge length of the reaction region, ν kinematic viscosity of water (m^2/s), g gravitational acceleration constant, β water volume expansion coefficient (1/K).

Resulting viscosity force density is

$$F_b - F_v = \Delta\rho g - \rho\frac{\mathrm{d}\left(\frac{\nu^2}{x^3}\right)}{\mathrm{d}x}\Delta x = \Delta\rho g - \rho\frac{3\nu^2}{x^4}\Delta x$$

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

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

With real values: $\nu(\text{water}, T = 20^\circ\text{C}) = 1.0\times10^{-6}\ \text{m}^2/\text{s}$

$\beta(\text{water}, T = 20^\circ\text{C}) = 2.1\times10^{-4}\ \text{K}^{-1}$

$x = \delta_\theta = 6.3$ cm, $\Delta T = 80$ K, $t_c = 1\Big/\sqrt{\frac{80\text{ K}}{300\text{ K}}\frac{9.81}{1} - \frac{3\times1.0\times10^{-12}}{\left(6.3\times10^{-2}\right)^4}} = 0.61\text{ s}$ and

$F_v \ll F_b$, the viscosity force is negligible against the buoyancy force component, and the convection velocity v_c becomes $v_c = \left(\frac{\mathrm{d}t_c}{\mathrm{d}x}\right)^{-1} = \frac{2\sqrt{\frac{T'(x)}{T(x)}g}^3}{g\left(\left(\frac{T'(x)}{T(x)}\right)^2 - \frac{T''(x)}{T(x)}\right)}$,

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

$$\frac{\partial c(t,x)}{\partial t} = -\frac{\partial c(t,x)}{\partial x}\Big/\frac{\mathrm{d}t_c}{\mathrm{d}x} = -\frac{\partial c(t,x)}{\partial x}v_c \tag{9}$$

2.8.2. Scenario1: Hydrothermal Vent with Spontaneous Synthesis of Basic Components

As was outlined in ([2] Chap. 3.4), the most plausible scenario for the origin of the first life-cycle is hydrothermal vents in submarine volcanic rocks or volcanic pools. We carried out calculation based on this scenario, and on the reaction table in Chap. 2.7.

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

$$\frac{\partial c(t,x)}{\partial t} = k c_1(t,x)^{k_1} c_2(t,x)^{k_2} \cdots c_n(t,x)^{k_n},$$

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 Chap. 2.8.1)

diffusion $\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}$ (10a)

membrane osmosis $\frac{\partial c(t,x_1)}{\partial t} = \frac{A}{dV} KD\left(c(t,x_1) - c_0\right)$ the boundary x_1 (10b)

convection $\frac{\partial c(t,x)}{\partial t} = -\frac{\partial c(t,x)}{\partial x} v_c$ (10c)

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 (**Figure 7**).

Scenario1 represents basically the primordial amino acid and peptide synthesis cycle (**Table 3(a)**, **Table 3(b)**).

Scenario1 starts with realistic primordial concentrations (init1) of basic compounds (amino acids and nucleobases) $c = 10^{-6} = 1$ ppm, low concentration of peptides $c = 10^{-8} = 0.01$ ppm, and $c = 0.001$ for precursors.

A second simulation starts with higher "enriched" concentrations (init2) of $c = 10^{-4} = 100$ ppm for basic compounds, $c = 10^{-6} = 1$ ppm for peptides, and $c = 0.01$ for precursors.

The reaction network consists of spontaneous and peptide-catalyzed basic-compound-synthesis (t1), which uses the H_2S energy cycle as a 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.

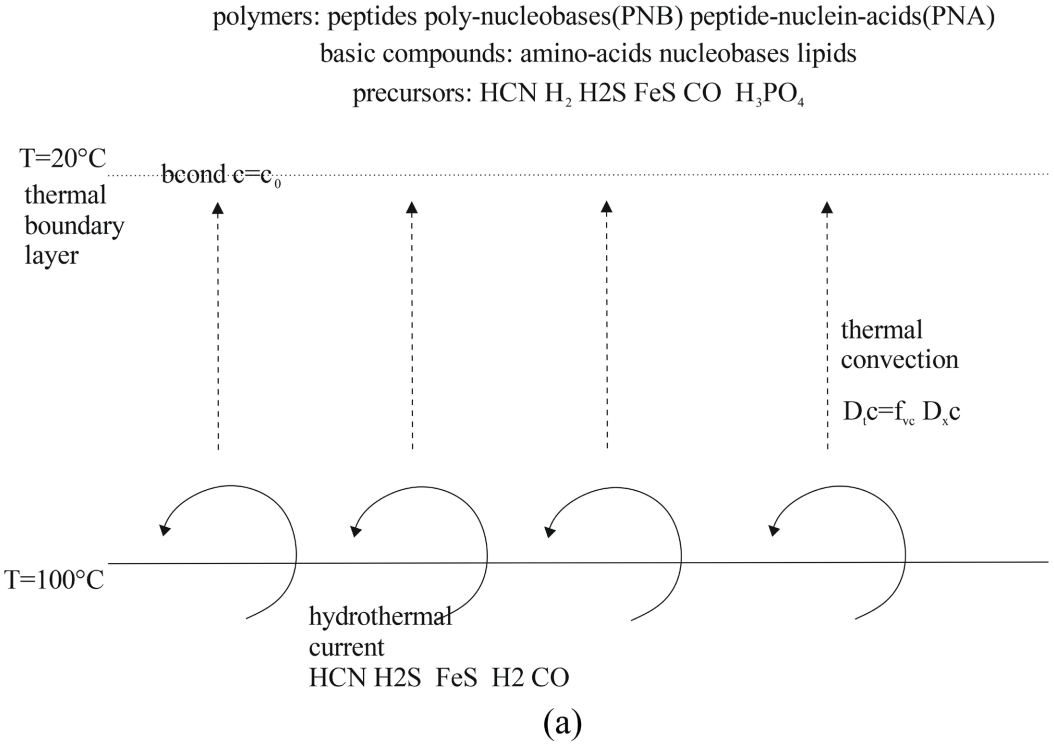
The peptides decay thermally (t8) above decomposition temperature T_c.

The result [12] is an enrichment of amino acids, but no significant enrichment of nucleobases and peptides, as shown in **Table 3** below.

2.8.3. Scenario2: Hydrothermal Vent with Proto-Lifecycle

Scenario2 is the full proto-life cycle from the reaction table in Chap. 2.7 (**Figure 8**), in the physical environment of a hydrothermal vent. It serves as an amplifier process for all involved molecule classes (amino acids, nucleobases, peptides,

Scenario1/2: model



Scenario1: reactions

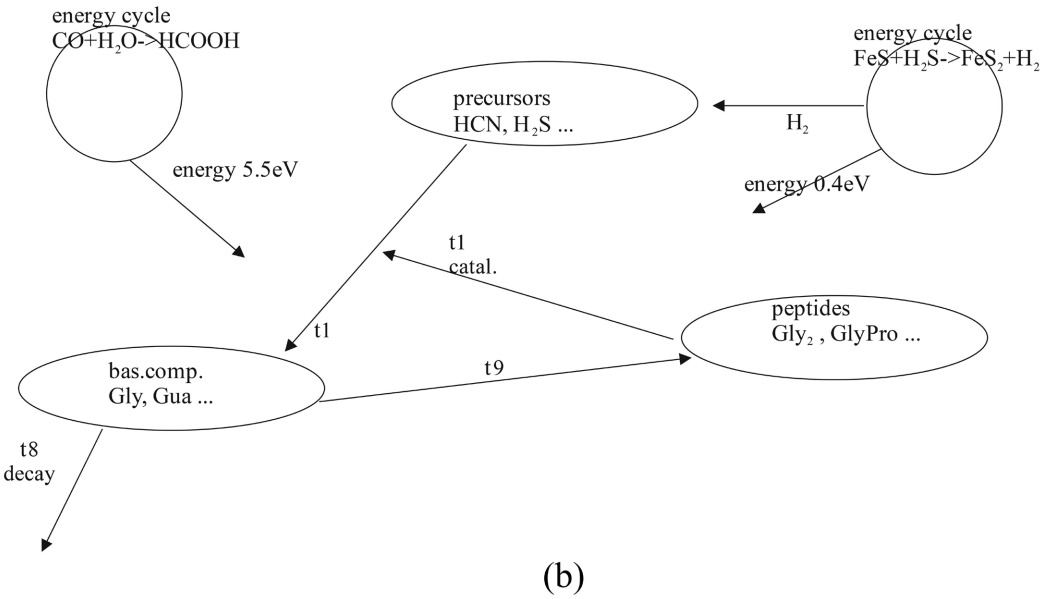


Figure 7. (a) Schematic scenario1/2 model; (b) Schematic scenario1 reactions.

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.

Table 3. (a) Scenario1 reactions simul1; (b) Scenario1 reactions simul2.

(a)					
Molecule	Initial value (ppm)	Final value	Behavior	Time (s)	f_{vc} (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	$1. \times 10^{-8}$	constant		
Pro2	0.01	6×10^{-12}	decreasing		
GlyPro	0.01	3×10^{-10}	decreasing		
(b)					
Molecule	Initial value	Final value	Behavior	Time (s)	f_{vc} (m/s)
Gly	$1. \times 10^{-4}$	$23. \times 10^{-4}$	asymptotic	0.7×10^4	0.5
Pro	$1. \times 10^{-4}$	$60. \times 10^{-4}$	asymptotic		
Cys	$1. \times 10^{-4}$	$1. \times 10^{-4}$	constant		
Gua	$1. \times 10^{-4}$	$1. \times 10^{-4}$	constant		
Cyt	$1. \times 10^{-4}$	1.04×10^{-4}	constant		
Gly2	$1. \times 10^{-6}$	0.94×10^{-6}	constant		
Gly3	$1. \times 10^{-6}$	0.1×10^{-6}	decreasing		
Pro2	$1. \times 10^{-6}$	0.01×10^{-6}	decreasing		
GlyPro	$1. \times 10^{-6}$	0.26×10^{-6}	decreasing		

Scenario2 starts with enriched concentrations (init2) of $c = 10^{-4} = 100$ ppm for basic compounds, $c = 10^{-6} = 1$ ppm 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 H_2S 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 $f_{vc} = 0.5$m/s are given in the following **Table 4(a)**.

This can be formulated concisely in the following scheme (**Table 4(b)**):

time constant: $t_r \approx 500$ s

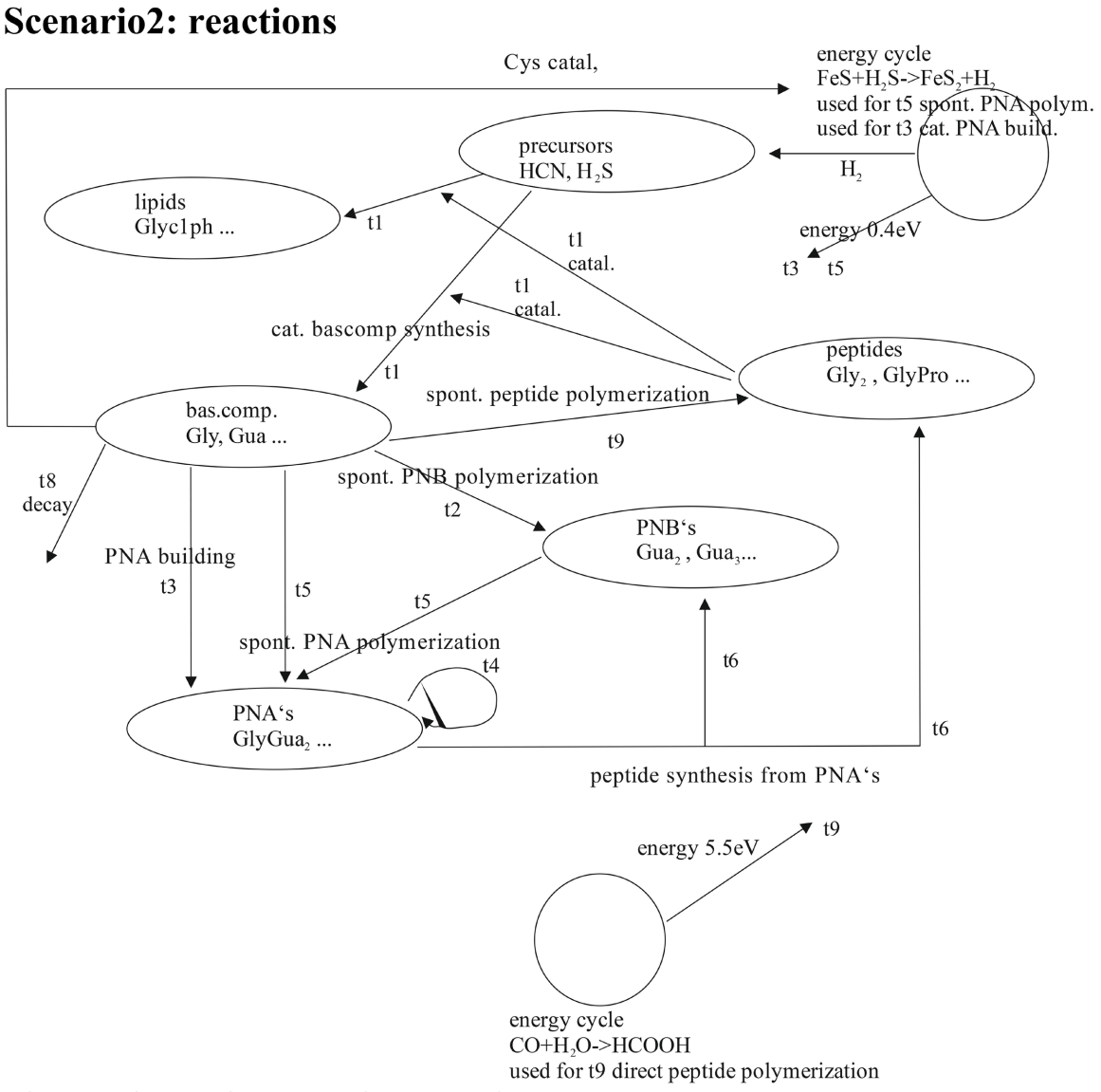


Figure 8. Schematic scenario2 reactions.

Table 4. (a) Scenario2 reactions [12]; (b) Scenario2 amplification factor.

(a)

Molecule	initial value	final value	behavior	time (s)	f_{vc} (m/s)
Gly	$1. \times 10^{-4}$	$0.0025 \rightarrow 9.2 \times 10^{-4}$	peak	500	0.5
Pro	$1. \times 10^{-4}$	$0.008 \rightarrow 52. \times 10^{-4}$	peak		
Cys	$1. \times 10^{-4}$	$1. \times 10^{-4}$	constant		
Gua	$1. \times 10^{-4}$	0.7×10^{-4}	decreasing		
Cyt	$1. \times 10^{-4}$	1.1×10^{-4}	decreasing		
Gly2	$1. \times 10^{-6}$	$0.008 \rightarrow 1. \times 10^{-6}$	peak		
Gly3	$1. \times 10^{-6}$	$0.0008 \rightarrow 4. \times 10^{-6}$	peak		
Pro2	$1. \times 10^{-6}$	$1.2 \times 10^{-5} \rightarrow 8.5 \times 10^{-6}$	peak		

Continued

GlyPro	$1. \times 10^{-6}$	1.3×10^{-5} ->> $49. \times 10^{-6}$	peak
Gua2	$1. \times 10^{-6}$	0.004 -> $36. \times 10^{-4}$	peak
Gua3	$1. \times 10^{-6}$	0.009 -> $41. \times 10^{-4}$	peak
Cyt2	$1. \times 10^{-6}$	3.4×10^{-4}	increase
GuaCyt	$1. \times 10^{-6}$	$15. \times 10^{-4}$	increase

(b)

	typical amplification factor
amino acids (Gly, Pro)	25, 80
peptides (Gly2)	8000
PNB's (Cyt2)	300

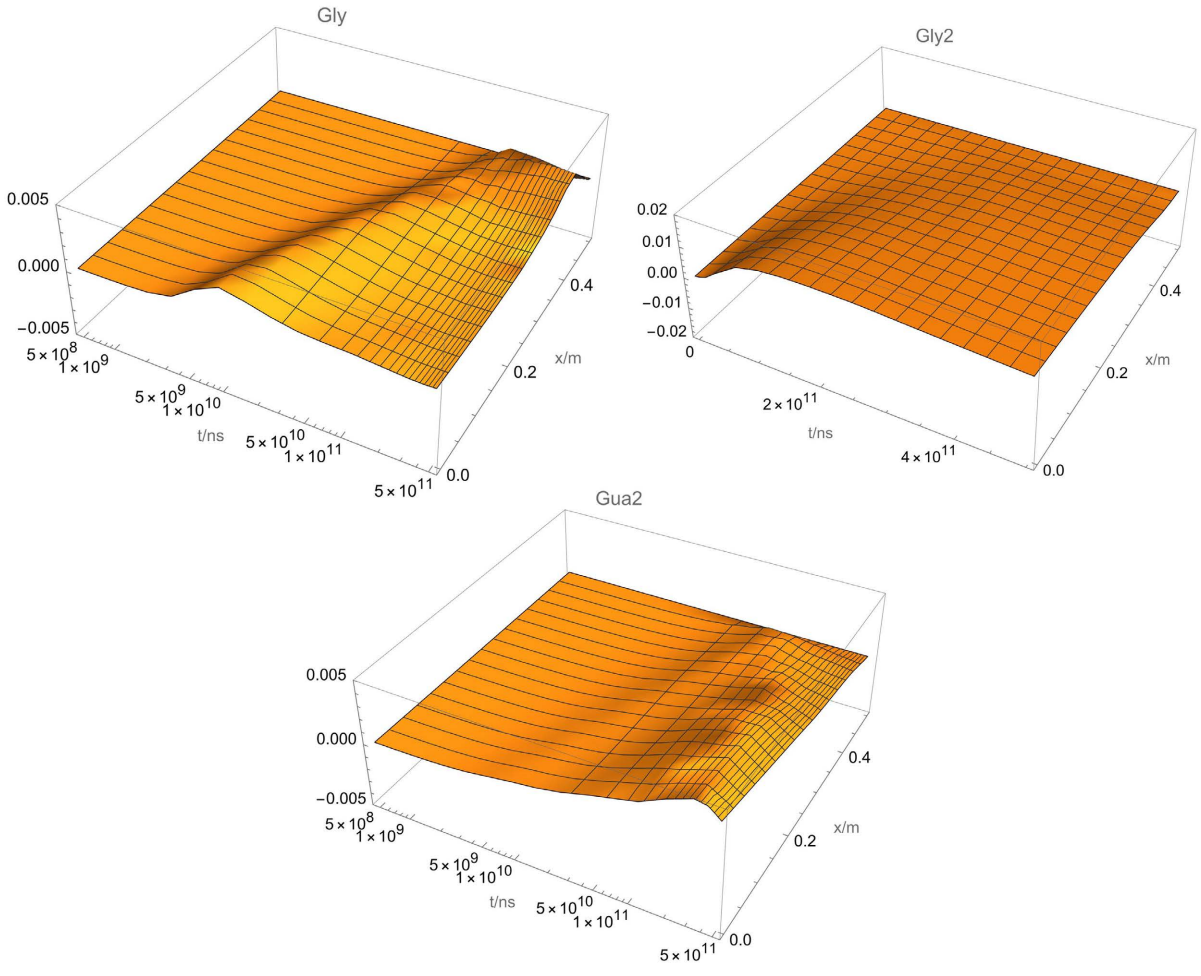


Figure 9. Scenario2 concentration buildup.

A typical concentration build-up for the amino acid Gly, the peptide Gly2 and the peptide Gua2 is shown below [12] (**Figure 9**).

2.8.4. 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] Chap. 3.4), 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 (**Figure 10**, **Figure 11**).

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 PNB's become now the *proto-genes*, which are stable in the cycle.

The basic components synthesis runs under catalytic peptide control, the pyrite energy cycle is catalyzed by Cys, PNA building from amino acids and PNB's takes energy from the energy cycle, the PNA's split into peptides and PNB's releasing energy, and the cycle runs anew.

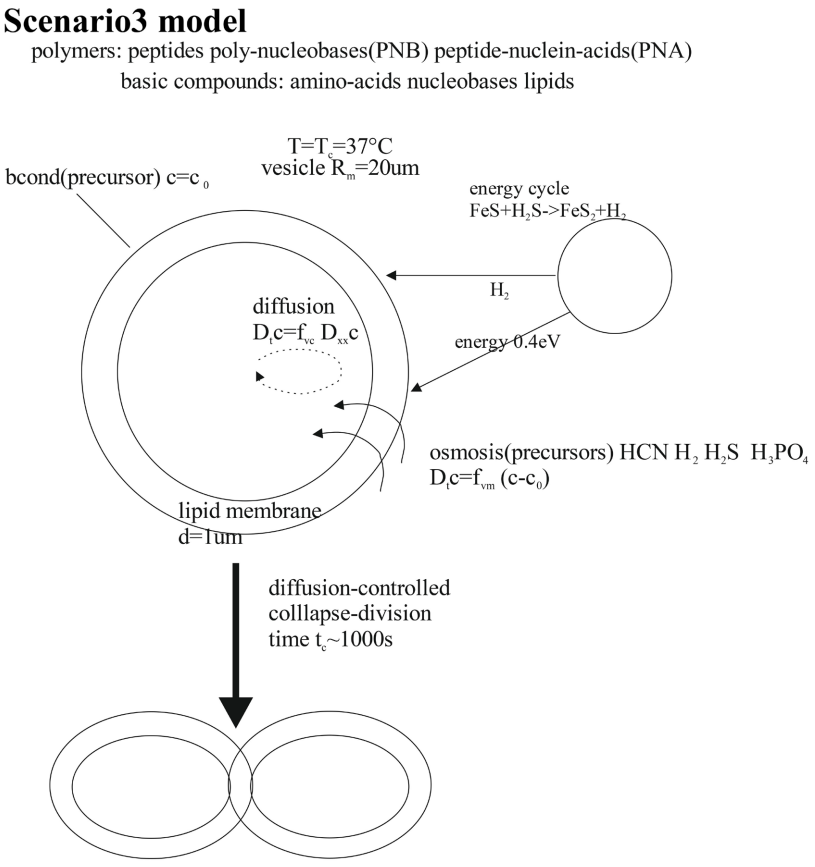


Figure 10. Scenario3 model.

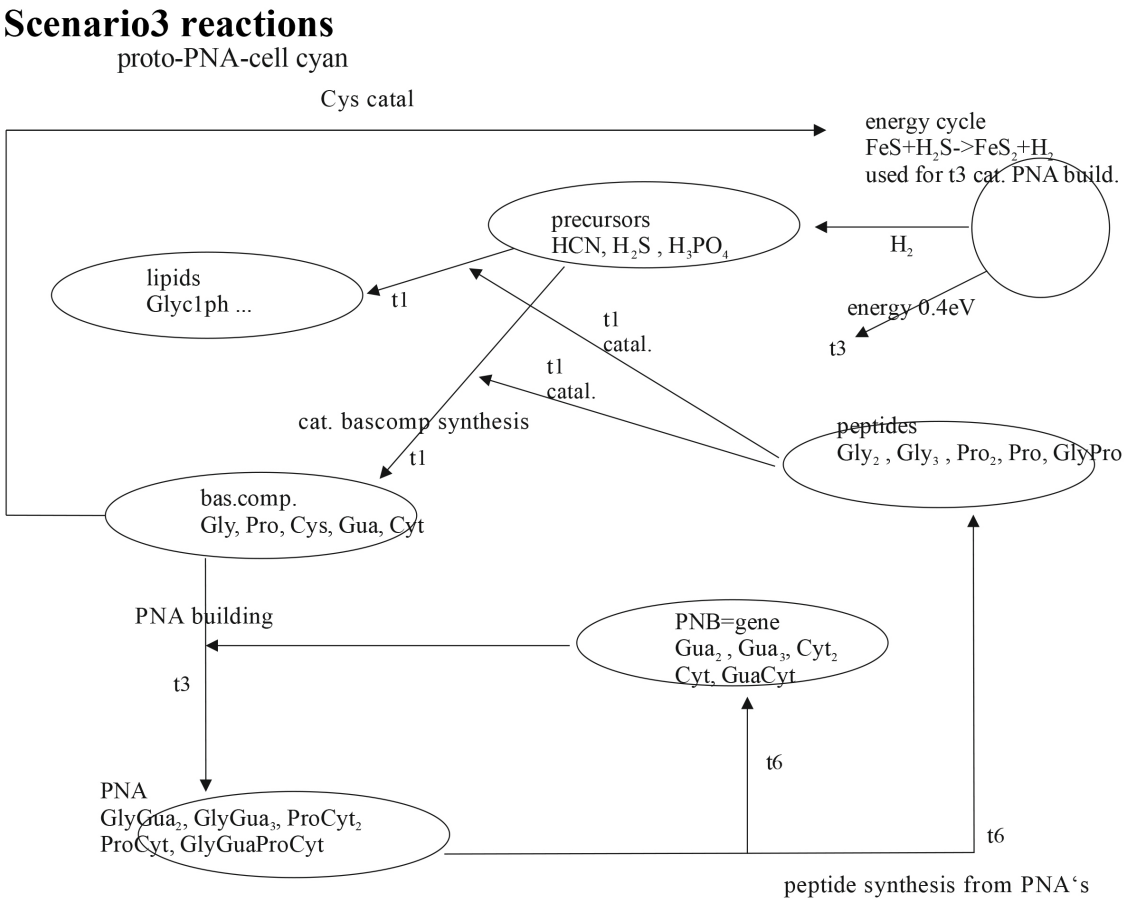


Figure 11. Schematic scenario3 reactions.

The pyrite energy reaction delivers energy and H_2 at the membrane, catalyzed by Cys, the insoluble FeS_2 falls out in colloidal form.

The current of precursors through the membrane is governed by osmosis:

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

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 1000 s, which agrees well with the observed division periods of self-reproducing lipid-amino-acid vesicles and also of bacteria.

The results for osmosis constant $f_{vm} = 0.3\ s^{-1}$ and diffusion constant $f_{vd} = 0.01\ m^2 \cdot s^{-1}$ are given in the following Table 5.

Table 5. Scenario3 reactions [12].

Molecule	Initial value	Final value	Behavior	Time (s)	f_{vm} (1/s) f_{vd} (m^2/s)
Gly	$1. \times 10^{-4}$	0.0016	increasing	880	0.3, 0.01
Pro	$1. \times 10^{-4}$	0.0034	plateau		
Cys	$1. \times 10^{-4}$	$81. \times 10^{-6}$	plateau		
Gua	$1. \times 10^{-4}$	7.6×10^{-7}	decreasing		
Cyt	$1. \times 10^{-4}$	11.9×10^{-7}	decreasing		
Gly2	$1. \times 10^{-6}$	34×10^{-6}	increasing		
Gly3	$1. \times 10^{-6}$	82×10^{-6}	plateau		
Pro2	$1. \times 10^{-6}$	1.4×10^{-6}	peak-plateau		
GlyPro	$1. \times 10^{-6}$	$22. \times 10^{-6}$	plateau		
Gua2	$1. \times 10^{-6}$	$19. \times 10^{-5}$	increasing		
Gua3	$1. \times 10^{-6}$	$34. \times 10^{-5}$	increasing		
Cyt2	$1. \times 10^{-6}$	3.5×10^{-5}	plateau		
GuaCyt	$1. \times 10^{-6}$	$10. \times 10^{-5}$	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 [12] (**Figure 12**).

2.9. The Complete Scenario of the Terrestrial Life Origin with Proto-PNA Genetic Code

The result of the simulation is a completely realistic scenario of the origin of terrestrial life based on HCN and H_2S as precursors, and the H_2S energy cycle. It consists of two phases:

Phase1 (scenario2) is the enrichment of amino acids, nucleobases and peptides in a hydrothermal pool with amplification factors of 25 to 8000 and a time constant of 500s, the process is guided by thermal convection and proceeds via spontaneous peptide polymerization, which needs considerable energy taken from the more powerful CO-energy cycle.

Phase2 (scenario3) is the build-up of a genuine self-catalytic life-proto-cycle in a self-assembling lipid membrane.

The underlying genetic code is a proto-code with 3 amino acids (Gly, Pro, Cys) and 2 nucleobases (Gua, Cyt) with 5 proto-genes (GG, GGG, CC, C, GC) coding for 5 peptides (Gly2, Gly3, Pro2, Pro, GlyPro), which catalyze the synthesis of 5 basic components (Gly, Cys, Pro, Cyt, Gua). The peptide synthesis is carried out by peptide-nucleid-acids (PNA's) generated in the proto-cycle, which correspond to the modern RNA-transferases. The process is guided by diffusion, and the calculation shows, that there is a periodic instability, *i.e.* a collapse-division time of the proto-cell of roughly 1000 s, which is about the same as the cell-division cycle time in modern bacteria. The PNB-proto-genes remain stable in the cycle and play the role of information carriers.

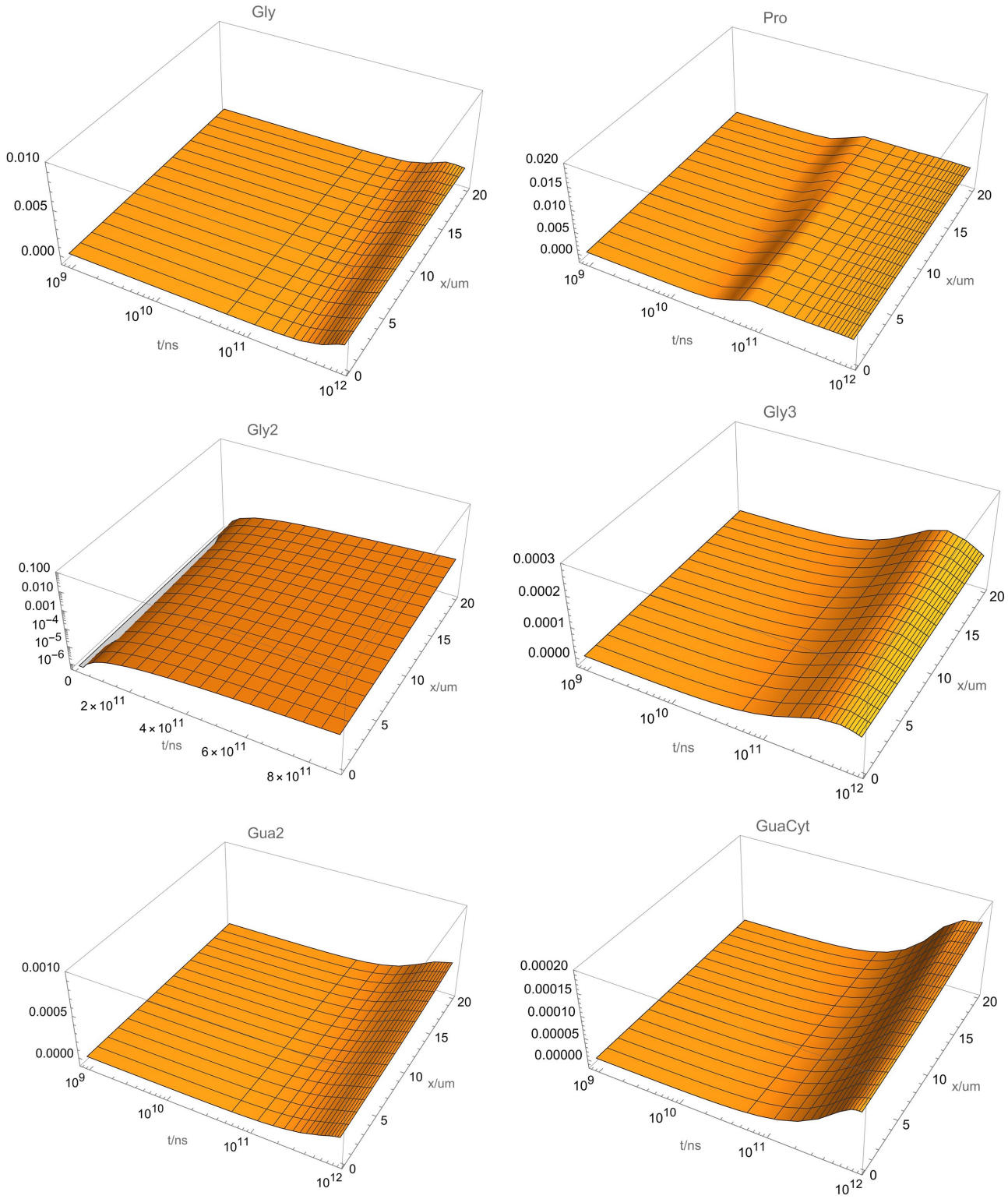


Figure 12. Scenario3 concentration buildup.

The process described above uses *precise* reactions and proto-code.

On the other hand, it is unrealistic that the proto-code described above is the only alternative. The PNA building is a spontaneous process, which is not very

accurate. There will be other PNA's, which will produce other similar peptides. Those peptides will survive, if they can at least partly catalyze a basic reaction,

That means, there will be not one gene PNB = GuaGua with the corresponding PNA = $(GlyGua)_2$ for glycine synthesis, but a family of genes (*quasi-gene* in the terminology of Manfred Eigen). The set of quasi-genes forms a quasi-species, where the variable genes adapt to the changing environment.

The evolution leads to an increasing number of more and more complex genes and peptides, where the catalysis is more and more *specific*: the *quasi-species* evolves into a *genuine species*.

Considering all this, we get the following life-origin quasi-species scenario (Figure 13(a), Figure 13(b)).

Complete genesis of proto-life cycle reaction scenario

phase1: red, phase2: black

In phase1, there are additional spontaneous polymerization reactions fed by the additional energy cycle, which are replaced by peptide-catalyzed reactions in phase2.

Complete genesis of proto-life cycle model

phase1: red, phase2: black

Here is included in red the phase1 open hydrothermal pool, whose reactions yield the enrichment in basic compounds and peptides necessary for the start of phase2.

We have the following Table 6 of genetic proto-code (model1) of the proto-lifecycle, discussed in full in Chap. 3.2.

2.10. Advanced PNA Genetic Code with Non-Cyan Precursors

The model1 binary proto-life-cycle uses a binary 1-1-code, *i.e.* a 2-letter-code (G, C) with 1 codon per gene. It has 2 coding amino acids (Gly for G, Pro for C) and 3 precursors (HCN, H_2S, H_2).

HCN has a relatively low molecular energy (=free enthalpy H_f) of $|H_f| = 551.9$, so the life-cycle needs energy only for PNA-building, which is supplied by the pyrite energy cycle (ΔE = 41 kJ/mole), whereas the basic-compound-synthesis runs exothermally without energy input.

HCN forms by electric discharge or UV-light from CO_2 and NH_3, and is also destroyed by UV-light, so it was only available early in the Earth history. Later (after ~100 My) the life-cycle had to adapt to other, more stable precursors. The

Table 6. Genetic proto-code model1.

Code	Product	Enzyme	RNA code
GG	Gly	Gly2	GGx
GGG	Cys	Gly3	UGx
CC	Pro	Pro2	CCx
C	Cyt	Pro	
GC	Gua	GlyPro	

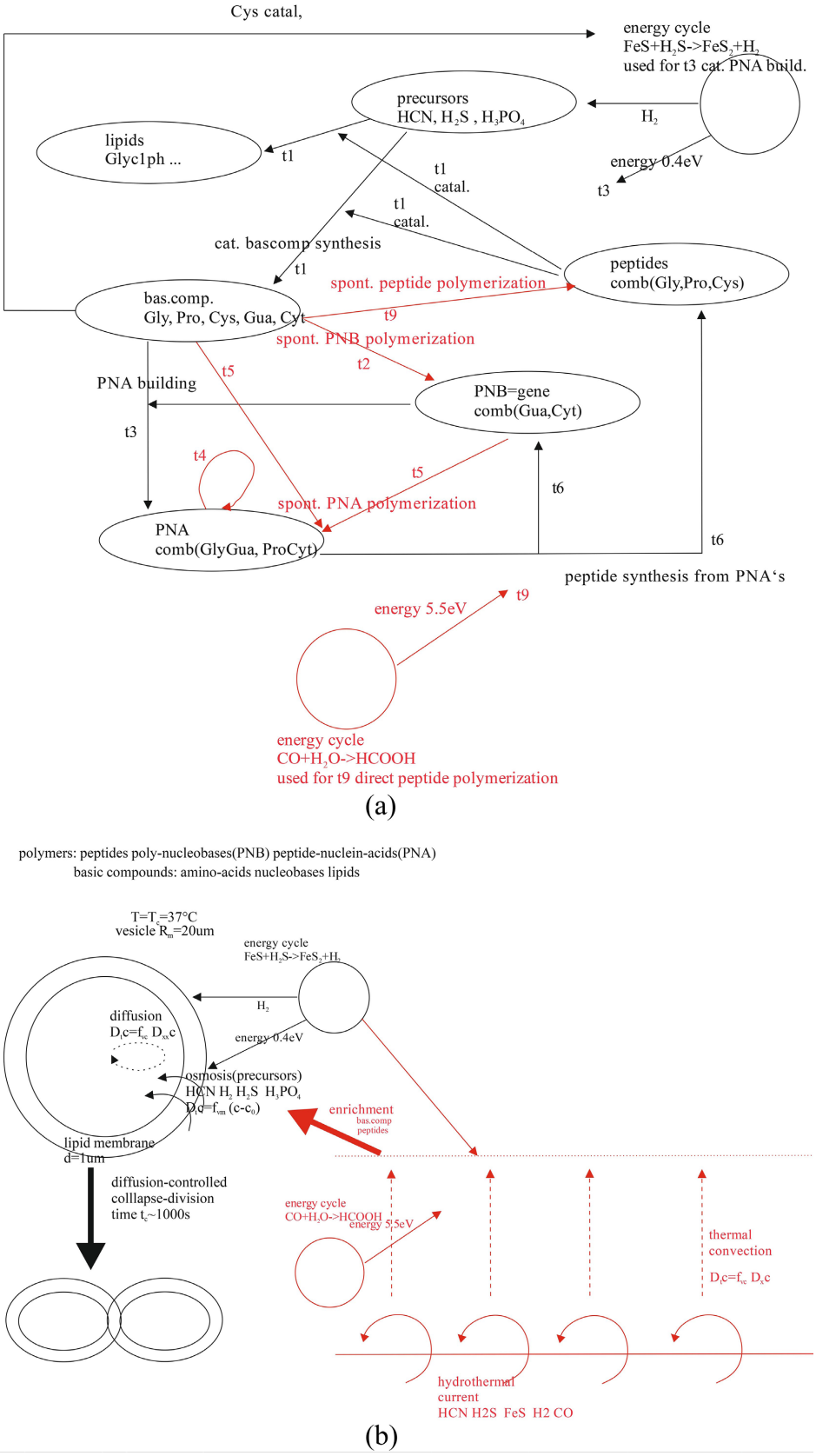


Figure 13. (a) Schematic complete genesis of proto-life cycle reactions; (b) Complete genesis of proto-life cycle model.

most plausible and least energy-consuming alternative is (CO, NH_3, H_2S, H_2), where the energy needed for the basic-compound-synthesis is in the range 140 - 400 kJ/mol, and can be supplied by multiple pyrite energy cycle (or single CO energy cycle).

The synthesis of the 2 basic amino acids (Gly, Pro) and the 2 nucleobases (Gua, Cyt) runs according to the scheme ([16] p. 94)

$$2CO + NH_3 + H_2 \rightarrow Gly\left(C_2H_5NO_2\right),\ \Delta E = -139 \text{ kJ/mole (endothermal)}$$

$$5CO + NH_3 + 6H_2 \rightarrow Pro\left(C_5H_9NO_2\right) + 3H_2O,\ \Delta E = -163.3 \text{ kJ/mol}$$

$$4CO + 3NH_3 + H_2 \rightarrow Cyt\left(C_4H_5N_3O\right) + 3H_2O,\ \Delta E = -340.5 \text{ kJ/mol}$$

$$5CO + 5NH_3 \rightarrow Gua\left(C_5H_5N_5O\right) + 4H_2O + H_2,\ \Delta E = -400 \text{ kJ/mol}$$

Model3 proto-PNA code

In order to produce enzymes for these reactions in addition to the existing ones, more coding amino acids and a larger code with 4 nucleobases is needed. A minimalistic version is model3: the extended model1 proto-PNA ternary code with 4 letters (G, C, A, U), 2-bit-codons, 12 coding amino acids and 4 nucleobases, and 1 codon per gene (1-1-code).

We start with the PNA 2-bit code (Table 7)

We take model 2 with (Gly, Pro, Cys, His, Leu) as a basic configuration, and the code of model1 includes the enzymes.

The coding is the same as in the RNA-code except, when modified because of a nucleobase instead of an amino acid: Gua (GC) replaces Ala, Cyt (CG) replaces

Table 7. PNA genetic 2-bit code.

am.acid	PNA code
Phe	UU
Leu	CU
Ile/Start/Met	AU
Val	GU
Ser	UC
Pro	CC
Thr	AC
Ala	GC
Tyr/Stop	UA
His/Gln	CA
Asn/Lys	AA
Asp/Glu	GA
Cys/Stop/Trp	UG
Arg	CG
Ser	AG
Gly	GG

Arg. Furthermore, we replace Tyr/Stop (UA) by Uracil Ura, and we replace Asn/Lys (AA) by Adenine Ade (the two U-A-codons with A at the end).

For the enzymes, we accept the extended correspondence from model1

G↔Gly, C↔Pro, A↔His, U↔Leu

In case of double-coding like Asp/Glu (GA) or Asn/Lys (AG), we take the simpler (=older) amino acid.

The nucleobase-pairing rule in PNA is G↔G, C↔C, A↔U.

Now we have the following code table **Table 8**.

Now, we have 4 basic-enzyme-coding amino acids (Gly, Pro, His, Leu), plus 8 enzyme-coding amino acids (Cys, Phe, Ile, Val, Ser, Thr, Asp, Asn), plus 4 nucleobases (Gua, Cyt, Ura, Ade).

This code is a plausible configuration, not a calculated solution, as in model1. Still, it can be regarded as a "basic" code within the quasi-species of the 2-bit ternary 1-1 PNA-proto-code.

The compounds in the reaction network are:

- 12 amino acids (Gly, Pro, His, Leu, Cys, Phe, Ile, Val, Ser, Thr, Asp, Asn)
- 4 nucleobases (Gua, Cyt, Ura, Ade)
- 16 peptides (enzymes) (Gly2, LeuGly, Pro2, ProGly2, ..., GlyHis, HisGly)
- 1 lipid Gly1ph
- 16 poly-nucleobases PNB's (genes) (GG, UG, CC, CG, ..., AA, GA, AG)
- 16 poly-nucleobase-aminoacids PNA's

Table 8. Model3 genetic code.

code mod1	code mod3	product	enzyme	full RNA code	PNA
GG	GG	Gly	Gly2	GG	GlyGua2
GGG	UG	Cys	LeuGly	UG	LeuAdeGlyGua
CC	CC	Pro	Pro2	CC	ProCyt2
C	CG	Cyt	ProGly2	CG(Arg)	ProCytGlyGua2
GC	GC	Gua	GlyPro	GC(Ala)	GlyGuaProCyt
	CA	His	ProHis	CA	ProCytHisUra
	CU	Leu	ProLeu	CU	ProCytLeuAde
	UU	Phe	Leu2	UU	LeuAde2
	AU	Ile	HisLeu	AU	HisUraLeuAde
	GU	Val	GlyLeu	GU	GlyGuaLeuAde
	UC	Ser	LeuPro	UC	LeuAdeProCyt
	AC	Thr	HisPro	AC	HisUraProCyt
	UA	Ura	LeuHis	UA(Tyr)	LeuAdeHisUra
	AA	Ade	His2	AA(Asn, Lys)	HisUra2
	GA	Asp	GlyHis	GA	GlyGuaHisUra
	AG	Asn	HisGly	AG	HisUraGlyGua

GlyGua2, LeuAdeGlyGua, ProCyt2, ProCytGlyGua2,
GlyGuaProCyt, ProCytHisUra, ProCytLeuAde
LeuAde2, HisUraLeuAde, GlyGuaLeuAde, LeuAdeProCyt, HisUraProCyt,
LeuAdeHisUra, HisUra2, GlyGuaHisUra, HisUraGlyGua
The precursors are CO, NH_3, H_2S, H_3PO_4
The energy cycle: pyrite reaction $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$
We have the following model3 scenario **Figure 14**.

2.11. LUCA with RNA Genetic Code

The model3: the extended model1 proto-PNA ternary code with 4 letters (G, C, A, U), 2-bit-codons, 8 coding amino acids and 4 nucleobases, and 1 codon per gene (1-1-code).

A plausible transition from model3 proto-PNA code to the proto-RNA code of LUCA proceeds by full use of the 3-bit-code, by increasing the number of amino acids to 20, and by extending the 1-codon genes to multi-codon genes with the introduction of start- and stop-codons.

Furthermore, the PNB-PNA synthesis is replaced by tRNA-synthesis with one tRNA for every amino acid, with the addition of 20 codons coding for the 20 tRNA's.

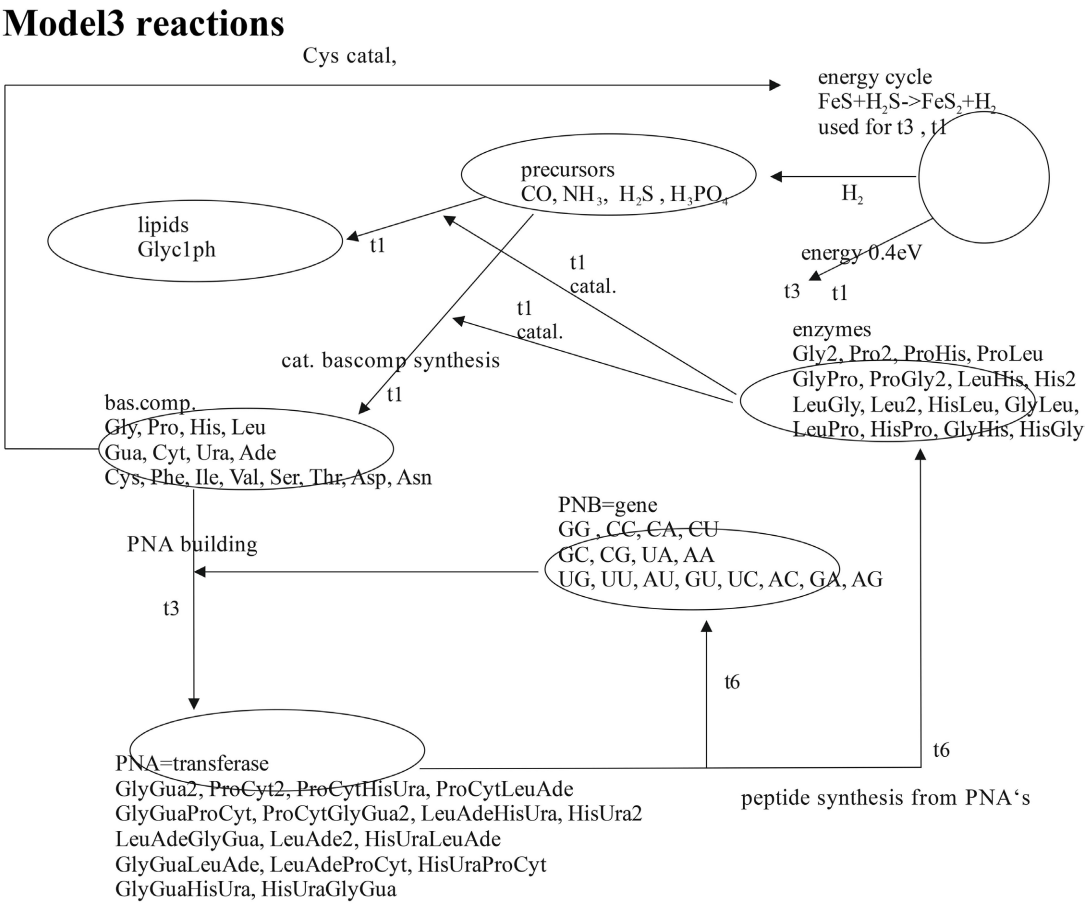


Figure 14. Schematic model3 (advanced PNA) reactions.

The PNB's are replaced by a single RNA-strand with the supporting skeleton of ribose and phosphate radicals, which makes the RNA much more stable than its fore-runner PNB's. The RNA is exactly copied by the enzyme RNA-polymerase (which is coded in a dedicated gene). Now, the RNA carries fixed genes, and is not a collection of quasi-genes like the PNB's.

LUCA was living in hydrothermal vents and/or hot volcanic pools.

LUCA's environmental conditions were: T ~80˚C, pH = 9, intermediate pressure.

LUCA used as energy cycle the acetogenesis (Wood-Ljungdahl)

$$H_2 + CO_2 \rightarrow \text{formate} \rightarrow \begin{matrix} CH_3\ \text{protein} \\ CO \rightarrow \end{matrix} \rightarrow \text{acetylCoA} \rightarrow \text{acetate} + \text{ATP}$$

The precursors were: H_2, CO_2, NH_3, H_2S, PO_4^{3-} ion

Now, we have the following reaction scenario Figure 15.

3. New Picture of the Terrestrial Life Evolution

We present now a completely modified model of terrestrial life evolution based on current knowledge ([2] Chap. 3) and on numerical simulation results from Chap. 2 (Figure 16).

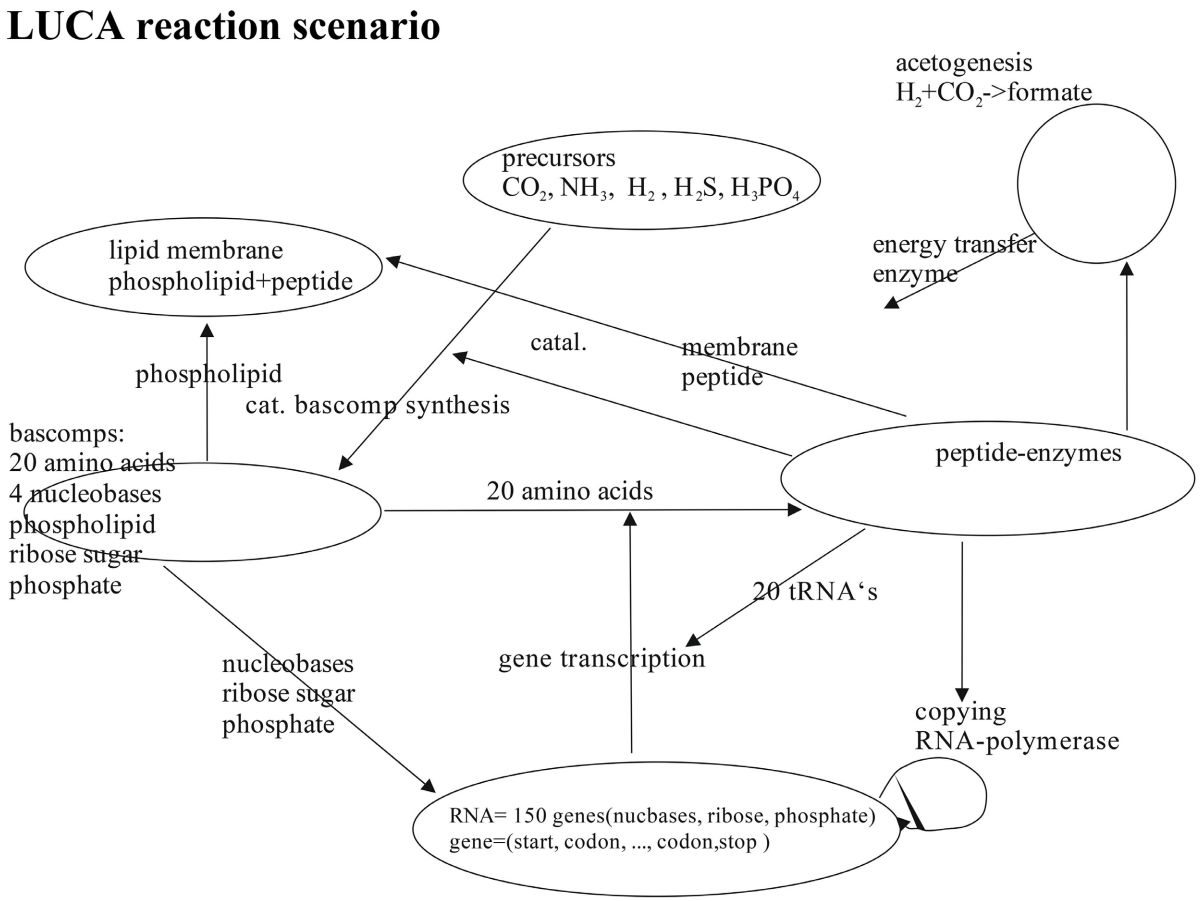


Figure 15. Schematic LUCA reactions.

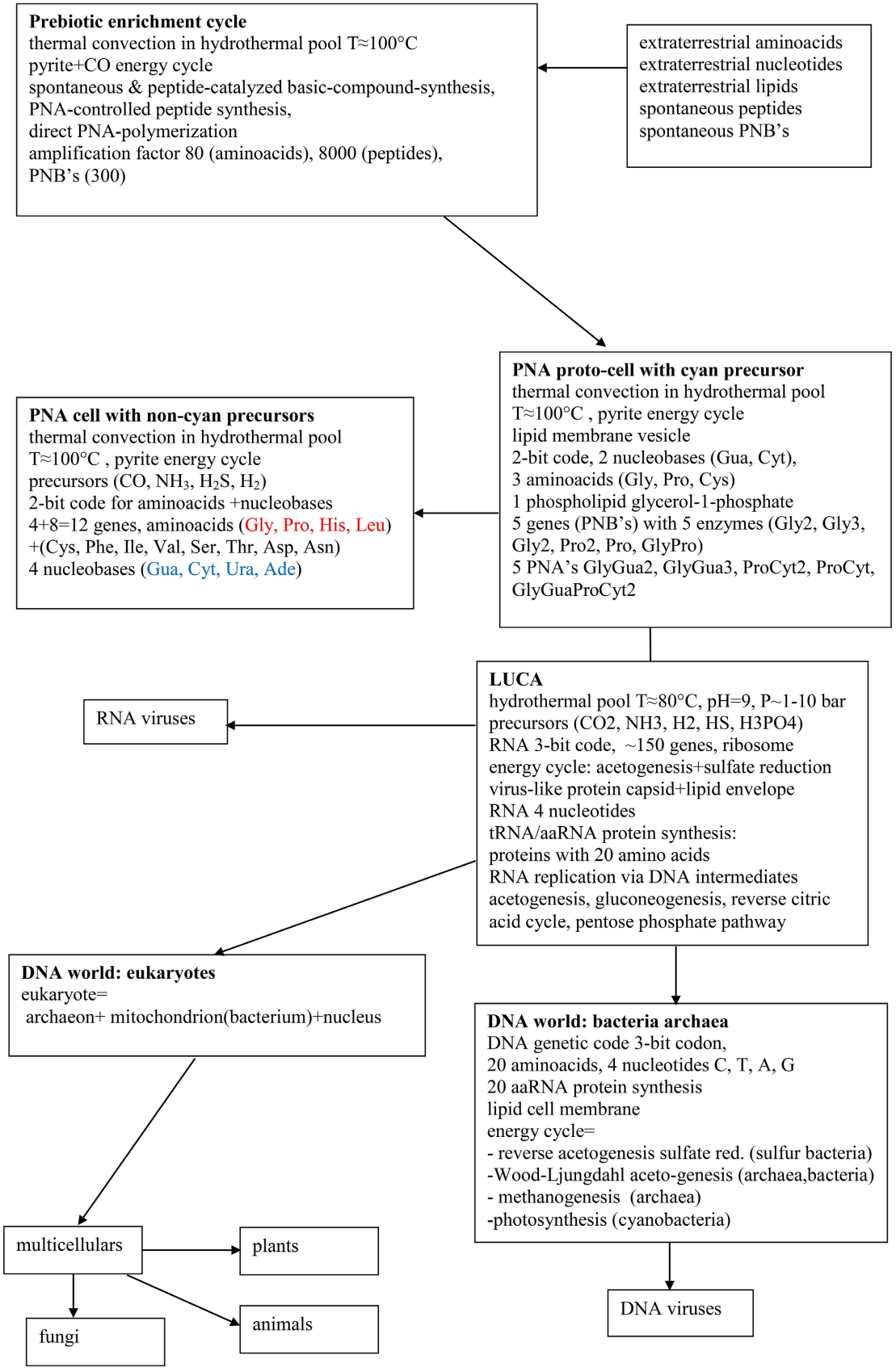


Figure 16. Terrestrial life evolution.

3.1. Prebiotic Enrichment Cycle

Scenario2 Chap. 2.8 is the full proto-life cycle from the reaction table in Chap. 2.7, 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 starts with enriched concentrations (init2) of $c = 10^{-4} = 100$ ppm for basic compounds, $c = 10^{-6} = 1$ ppm for peptides, and $c = 0.01$ for precursors.

The reaction network consists of spontaneous and peptide-catalyzed basic-compound-synthesis, with the H2S (pyrite) energy cycle and the CO energy cycle, then the PNA-controlled peptide synthesis, and direct PNA-polymerization.

The PNA-controlled peptide synthesis works also without the initial presence of peptides (*i.e.* initial peptide concentration $c = 0$), whereas the direct peptide polymerization requires high rates of CO energy cycle to counteract the reverse reaction (direct peptide decay).

The results of the enrichment reactions are given in the following Table 9.

Table 9. (a) Pre-biotic enrichment cycle; (b) Pre-biotic enrichment cycle amplification.

(a)

Molecule	Initial value	Final value	Behavior	Time (s)	f_{vc} (m/s)
Gly	$1. \times 10^{-4}$	$0.0025 \rightarrow 9.2 \times 10^{-4}$	peak	500	0.5
Pro	$1. \times 10^{-4}$	$0.008 \rightarrow 52. \times 10^{-4}$	peak		
Cys	$1. \times 10^{-4}$	$1. \times 10^{-4}$	constant		
Gua	$1. \times 10^{-4}$	0.7×10^{-4}	decreasing		
Cyt	$1. \times 10^{-4}$	1.1×10^{-4}	decreasing		
Gly2	$1. \times 10^{-6}$	$0.008 \rightarrow 1. \times 10^{-6}$	peak		
Gly3	$1. \times 10^{-6}$	$0.0008 \rightarrow 4. \times 10^{-6}$	peak		
Pro2	$1. \times 10^{-6}$	$1.2 \times 10^{-5} \rightarrow 8.5 \times 10^{-6}$	peak		
GlyPro	$1. \times 10^{-6}$	$1.3 \times 10^{-5} \rightarrow >49. \times 10^{-6}$	peak		
Gua2	$1. \times 10^{-6}$	$0.004 \rightarrow 36. \times 10^{-4}$	peak		
Gua3	$1. \times 10^{-6}$	$0.009 \rightarrow 41. \times 10^{-4}$	peak		
Cyt2	$1. \times 10^{-6}$	3.4×10^{-4}	increase		
GuaCyt	$1. \times 10^{-6}$	$15. \times 10^{-4}$	increase		

(b)

	Typical amplification factor
amino acids (Gly, Pro, Cys)	25, 80, 1
nucleobases (Gua, Cyt)	1, 1
peptides (Gly2)	8000
PNB's (Cyt2)	300

This can be formulated concisely in the following scheme:

time constant: about 500 s

Yaman & Harvey [28] calculate minimal reaction time for amino acid synthesis catalyzed by dinucleotide $t_r\left(E_a = 20\ \text{kcal/mol}\right) = \exp\left(E_a/kT\right) 1\ \text{ns} \approx 200\ \text{s}$.

3.2. PNA Proto-Cell with Proto-Code for 3 Amino Acids and Cyan Precursors

The proto-code consists of binary 3-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 [22]) and 2 PNA's for 3 amino acids, 1 lipid and 2 nucleobases is now as follows (**Table 10**, **Figure 17**).

The scenario (=scenario3 Chap. 2.8) of the PNA proto-cell is the full proto-life cycle, confined in the interior of a lipid vesicle. As was reported in ([2] Chap. 3.4), amino acids within a lipid membrane stabilize the lipid layer, and the lipids

Table 10. Genetic proto-code.

code	product	enzyme	RNA code
GG	Gly	Gly2	GGx
GGG	Cys	Gly3	UGx
CC	Pro	Pro2	CCx
C	Cyt	Pro	
GC	Gua	GlyPro	

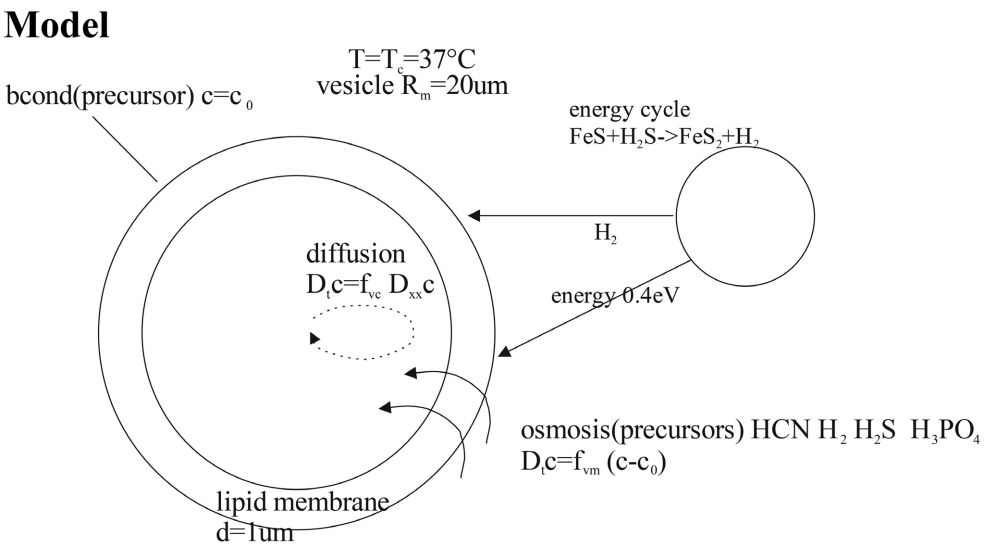


Figure 17. PNA proto-cell model.

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 (**Figure 18(a)**):

- 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
- proliferation through bio-matter production and vesicle division.

The compounds in the reaction network are:

- 3 amino acids Gly, Pro. Cys
- 2 nucleobases Gua, Cyt
- 5 peptides (enzymes) Gly2, Gly3, Pro, Pro2, GlyPro
- 1 lipid Gly1ph
- 5 PNB's (genes) Gua2, Gua3, Cyt2, Cyt, GuaCyt
- 5 PNA's GlyGua2, GlyGua3, ProCyt2, ProCyt, GlyGuaProCyt2

The precursors are HCN, H_2S, H_3PO_4

The energy cycle: pyrite reaction $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$

This 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 PNB's become now the *proto-genes*, which are stable in the cycle.

The basic components synthesis runs under catalytic peptide control, the pyrite energy cycle is catalyzed by Cys, PNA building from amino acids and PNB's takes energy from the energy cycle, the PNA's split into peptides and PNB's releasing energy, and the cycle runs anew.

The pyrite energy reaction delivers energy and H_2 at the membrane, catalyzed by Cys, the insoluble FeS_2 falls out in colloidal form.

The current of precursors through the membrane is governed by osmosis:

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

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.

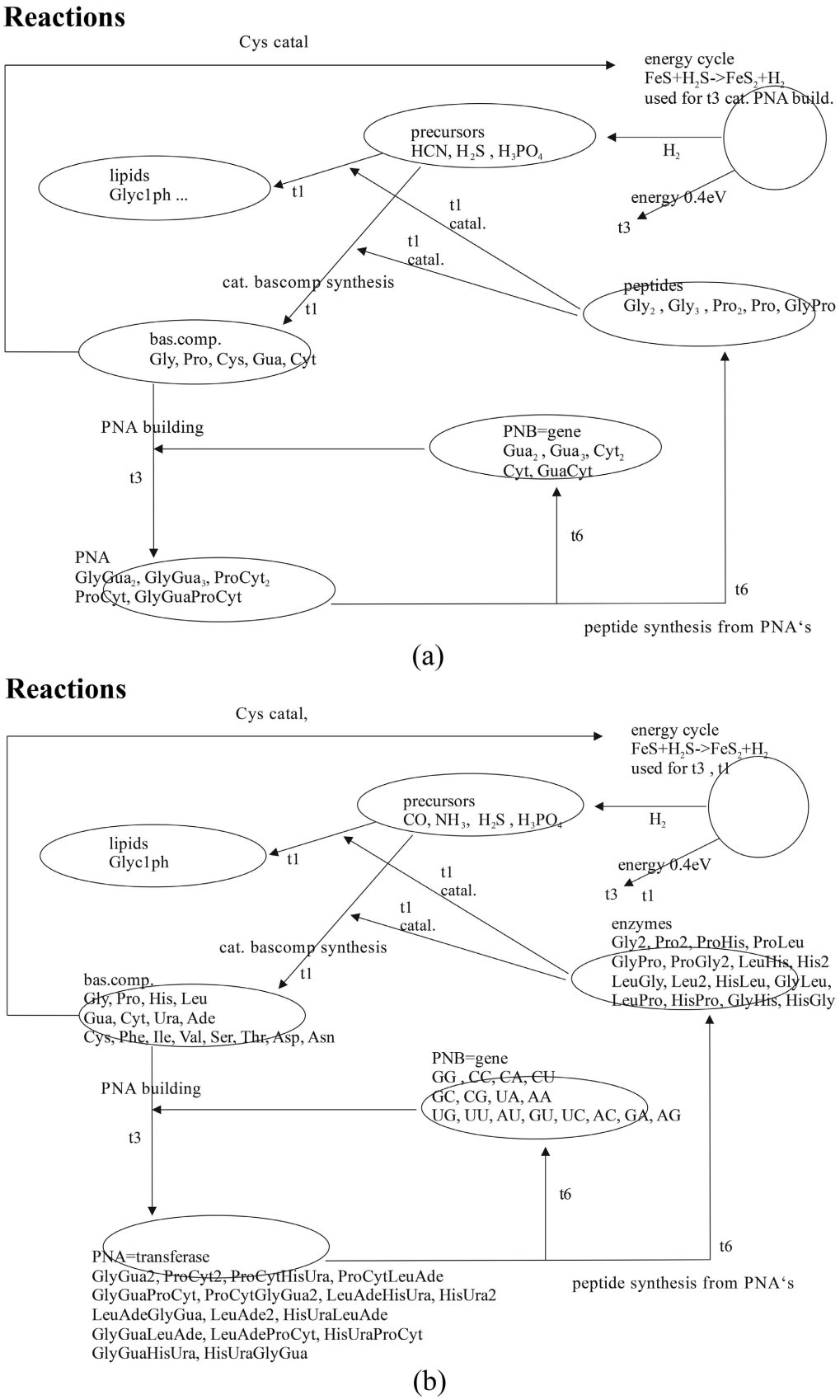


Figure 18. (a) Schematic PNA proto-cell reactions; (b) Schematic advanced PNA-cell reactions.

Yaman & Harvey [28] presented a simplified proto-genetic code with 9 two-letter codons from (A, G, C) for 5 amino acids and 3 precursors, which runs purely abiotically by catalysis with the corresponding dinucleotide (**Table 11**).

With only two letters (G, C) and 4 codons, it reduces to the table that matches the model1 table for the entries CC and GG, whereas the entries CG and GC catalyze indirectly arginine resp. alanine, which does not participate directly in the proto-lifecycle of model1.

3.3. Advanced PNA Cell with Non-Cyan Precursors

After ~100 My, the life-cycle had to adapt to other, more stable precursors than HCN.

The most plausible and least energy-consuming alternative precursors (CO, NH_3, H_2S, H_2).

We have the following model3 code in **Table 12**.

We have here a full 2-bit code for aminoacids + nucleobases, specifically, we have 4 basic-enzyme-coding amino acids (Gly, Pro, His, Leu), plus 8 enzyme-coding amino acids (Cys, Phe, Ile, Val, Ser, Thr, Asp, Asn), plus 4 nucleobases (Gua, Cyt, Ura, Ade) (**Figure 18(b)**).

This code is a plausible configuration, not a calculated solution, as in model1. Still, it can be regarded as a "basic" code within the quasi-species of the 2-bit ternary 1-1 RNA-proto-code.

3.4. LUCA with Proto-RNA Genetic Code

The model3 is the extended model1 proto-PNA ternary code with 4 letters (G, C, A, U), 2-bit-codons, 8 coding amino acids and 4 nucleobases, and 1 codon per gene (1-1-code).

A plausible transition from model3 proto-PNA code to the proto-RNA code of LUCA proceeds by full use of the 2-bit-code, by increasing the number of amino acids to 20, and by extending the 1-codon genes to multi-codon genes with the introduction of start- and stop-codons.

Table 11. (a) Yaman & Harvey genetic proto-code; (b) Yaman & Harvey simplified genetic proto-code.

(a)			
	A adenine	G guanine	C cytosine
A adenine	asparagine Asn	diaminobutyric acid Dab	homoserine Hsr
G guanine	aspartic acid Asp	glycine Gly	alanine Ala
C cytosine	glutamine Gln	ornithine Orn	proline Pro

(a)		
	G guanine Gua	C cytosine Cyt
G guanine Gua	glycine Gly	alanine Ala
C cytosine Cyt	ornithine Orn → arginine Arg	proline Pro

Table 12. Advanced PNA cell genetic code.

code mod1	code mod3	product	enzyme	full RNA code
GG	GG	Gly	Gly2	GG
GGG	UG	Cys	LeuGly	UG
CC	CC	Pro	Pro2	CC
C	CG	Cyt	ProGly2	CG(Arg)
GC	GC	Gua	GlyPro	GC(Ala)
	CA	His	ProHis	CA
	CU	Leu	ProLeu	CU
	UU	Phe	Leu2	UU
	AU	Ile	HisLeu	AU
	GU	Val	GlyLeu	GU
	UC	Ser	LeuPro	UC
	AC	Thr	HisPro	AC
	UA	Ura	LeuHis	UA(Tyr)
	AA	Ade	His2	AA(Asn, Lys)
	GA	Asp	GlyHis	GA
	AG	Asn	HisGly	AG

Furthermore, the PNB-PNA synthesis is replaced by tRNA-synthesis with one tRNA for every amino acid, with the addition of 20 genes coding for the 20 tRNA's.

The PNB's are replaced by a single RNA-strand with the supporting skeleton of ribose and phosphate radicals, which makes the RNA much more stable than its fore-runner PNB's. The RNA is exactly copied by the enzyme RNA-polymerase (which is coded in a dedicated gene). Now, the RNA carries fixed genes, and is not a collection of quasi-genes like the PNB's.

The full RNA code is shown in 1.1.

The precursors are CO_2, NH_3, H_2, H_2S, H_3PO_4

The energy cycle is the aceto-genesis from H_2 and CO_2

$$H_2 + CO_2 \rightarrow \text{formate} \rightarrow \begin{matrix} CH_3\ \text{protein} \\ CO \rightarrow \end{matrix} \rightarrow \text{acetylCoA} \rightarrow \text{acetate} + \text{ATP}$$

The biosynthetic reaction network consists of ≈150 basic reactions, from which 95% were exergonic [29]. The four key reactions are acetogenesis (Wood-Ljungdal), gluconeogenesis, reverse citric acid cycle, pentose phosphate pathway. From the 26 key intermediate cofactors the most important 7 are: ATP, GTP, THF, NAD, NADP, CoA, SAM.

Now, we have the following reaction scenario (Figure 19).

3.5. DNA cell with Full DNA Genetic Code

The DNA-cell (bacteria, archaea, eukaryotes) uses the full DNA genetic code (Table 13) with 20 amino acids, and corresponding 20 aaRNA's for peptide synthesis,

Reactions

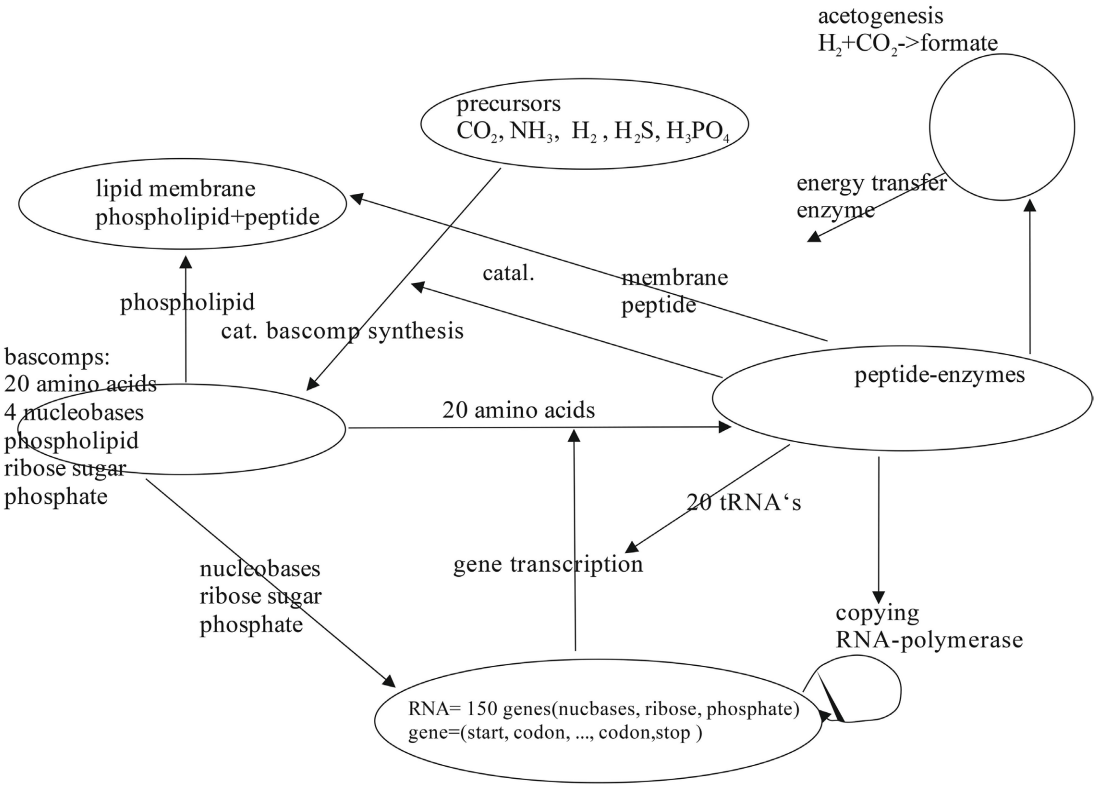


Figure 19. LUCA reactions.

Table 13. DNA-cell genetic code.

	T		C		A		G		
T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T
	TTC		TCC		TAC		TGC		C
	TTA	Leu	TCA		TAA	Stop	TGA	Stop	A
	TTG		TCG		TAG	Stop	TGG	Trp	G
C	CTT		CCT	Pro	CAT	His	CGT	Arg	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	Gln	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	Lys	AGA	Arg	A
	ATG	Met	ACG		AAG		AGG		G
G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	Glu	GGA		A
	GTG		GCG		GAG		GGG		G

4 nucleotides: cytosine, thymine, adenine, guanine (thymine replaces uracil in RNA-coding). The pairs A-G and C-T are complementary in the DNA-double-helix.

The cell membrane consists of ester and ether lipids (bacteria, eukaryotes) or diether lipids (archaea).

The (self-sustainable) energy cycle is widely diverse:

- reverse acetogenesis with sulfate reduction (sulfur bacteria)
- Wood-Ljungdahl aceto-genesis (archaea and bacteria in hydrothermal vents and volcanic pools)
- methanogenesis from carbon dioxide and hydrogen (archaea)
- photosynthesis (cyanobacteria)

4. Conclusions

Chap. 2 presents the mathematical formulation and numeric simulation of the genetic proto-code model based on an empirical reaction model developed for life chemistry.

The result of the simulation is a *complete realistic scenario of the origin of terrestrial life* based on HCN and H_2S as precursors, and the H2S energy cycle. It consists of two phases:

Phase1 is the *enrichment* of amino acids, nucleobases and peptides in a hydrothermal pool with amplification factors of 25 to 8000 and a *time constant of* 500 s, the process is guided by *thermal convection*.

Phase2 is the build-up of a genuine self-catalytic *life-proto-cycle* in a self-assembling lipid membrane based on a *genetic proto-code* with 3 amino acids (Gly, Pro, Cys) and 2 nucleobases (Gua, Cyt) with 5 proto-genes (GG, GGG, CC, C, GC) coding for 5 peptides (Gly2, Gly3, Pro2, Pro, GlyPro), which catalyze the synthesis of 5 basic components (Gly, Cys, Pro, Cyt, Gua). The peptide synthesis is carried out by peptide-nucleid-acids (PNA's) generated in the proto-cycle, which correspond to the modern RNA-transferases. The process is guided by *diffusion*, and the calculation shows, that there is a fundamental *collapse-division time* of the proto-cell of roughly 1000 s, which is about the same as the cell-division cycle time in modern bacteria.

Chap. 3 presents a detailed modified picture of terrestrial life evolution based on the models derived in the preceding chapter.

In this picture, there are 4 basic stages of genetic complexity.

- Stage1 enrichment cycle

This is the enrichment cycle without membrane, for 3 amino acids, 2 nucleobases, peptides, poly-nucleobases (PNB)

	typical amplification factor
amino acids (Gly, Pro, Cys)	25, 80, 1
nucleobases (Gua, Cyt)	1, 1
peptides (Gly2)	8000
PNB's (Cyt2)	300

- Stage2 PNA proto-cell

This is the proto-lifecycle in the proto-cell within lipid vesicle with cyan precursors.

The compounds in the reaction network are:

- 3 amino acids Gly, Pro. Cys
- 2 nucleobases Gua, Cyt
- 5 peptides (enzymes) Gly2, Gly3, Pro, Pro2, GlyPro
- 1 lipid Gly1ph
- 5 poly-nucleobases PNB's (genes) Gua2, Gua3, Cyt2, Cyt, GuaCyt
- 5 poly-nucleobase-aminacids PNA's GlyGua2, GlyGua3, ProCyt2, ProCyt, GlyGuaProCyt2

The precursors are HCN, H_2S, H_3PO_4

The energy cycle: pyrite reaction $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$

The genetic code table (model2) is as follows.

code	product	enzyme	RNA code
GG	Gly	Gly2	GGx
GGG	Cys	Gly3	UGx
CC	Pro	Pro2	CCx
C	Cyt	Pro	
GC	Gua	GlyPro	

- Stage3 advanced PNA-cell

After ~100 My there is a more advanced cell with more stable precursors.

We have here a full 2-bit code for aminoacids +nucleobases, specifically, we have 4 basic-enzyme-coding amino acids (Gly, Pro, His, Leu), plus 8 enzyme-coding amino acids (Cys, Phe, Ile, Val, Ser, Thr, Asp, Asn), plus 4 nucleobases (Gua, Cyt, Ura, Ade).

The genetic code table (model3) is as follows.

code mod1	code mod3	product	enzyme	full RNA code	PNA
GG	GG	Gly	Gly2	GG	GlyGua2
GGG	UG	Cys	LeuGly	UG	LeuAdeGlyGua
CC	CC	Pro	Pro2	CC	ProCyt2
C	CG	Cyt	ProGly2	CG(Arg)	ProCytGlyGua2
GC	GC	Gua	GlyPro	GC(Ala)	GlyGuaProCyt
	CA	His	ProHis	CA	ProCytHisUra
	CU	Leu	ProLeu	CU	ProCytLeuAde
	UU	Phe	Leu2	UU	LeuAde2
	AU	Ile	HisLeu	AU	HisUraLeuAde
	GU	Val	GlyLeu	GU	GlyGuaLeuAde

Continued

	UC	Ser	LeuPro	UC	LeuAdeProCyt
	AC	Thr	HisPro	AC	HisUraProCyt
	UA	Ura	LeuHis	UA(Tyr)	LeuAdeHisUra
	AA	Ade	His2	AA(Asn, Lys)	HisUra2
	GA	Asp	GlyHis	GA	GlyGuaHisUra
	AG	Asn	HisGly	AG	HisUraGlyGua

The compounds in the reaction network are:

- 12 amino acids (Gly, Pro, His, Leu, Cys, Phe, Ile, Val, Ser, Thr, Asp, Asn)
- 4 nucleobases (Gua, Cyt, Ura, Ade)
- 16 peptides (enzymes) (Gly2, LeuGly, Pro2, ProGly2, ..., GlyHis, HisGly)
- 1 lipid Gly1ph
- 16 poly-nucleobases PNB's (genes) (GG, UG, CC, CG, ..., AA, GA, AG)
- 16 poly-nucleobase-aminoacids PNA's
  GlyGua2, LeuAdeGlyGua, ProCyt2, ProCytGlyGua2,
  GlyGuaProCyt, ProCytHisUra, ProCytLeuAde,
  LeuAde2, HisUraLeuAde, GlyGuaLeuAde, LeuAdeProCyt, HisUraProCyt,
  LeuAdeHisUra, HisUra2, GlyGuaHisUra, HisUraGlyGua
  The precursors are CO, NH_3, H_2S, H_3PO_4
  The energy cycle: pyrite reaction $FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-$
- Stage4: LUCA
  LUCA lived about 4.3 Gy ago and had

- RNA genome ~150 genes
- ribosome
- virus-like protein capsid + lipid envelope
- RNA 4 nucleotides
- proteins with 20 aminoacids
- RNA replication via DNA intermediates
- DNA polymerase
- Class I: 10 aaRS (D, Cys, Gly)
- Class II: 10 aaRS (A, B, Pro)
- 15 families of small subunit proteins, 18 families of large subunit proteins
- synthesis of RNA using DNA templates
- acetogenesis energy cycle from H_2
- temperature ~80˚C, pH = 9, pressure intermediate ~1 - 10 bar
- LUCA used as energy cycle the acetogenesis (Wood-Ljungdahl)

$$H_2 + CO_2 \rightarrow \text{formate} \rightarrow \begin{matrix} CH_3\ \text{protein} \\ CO \rightarrow \end{matrix} \rightarrow \text{acetylCoA} \rightarrow \text{acetate} + \text{ATP}$$

- The precursors were: H_2, CO_2, NH_3, H_2S, PO_4^{3-} ion

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Ruse, M., *et al.* (2009) Evolution: The First Four Billion Years. Harvard University Press, Cambridge.

[2] Helm, J. (2021) Origin and Basic Mechanism of Life. https://www.researchgate.net/

[3] Lehman, N. (2015) The RNA World: 4,000,000,050 Years Old. *Life*, **5**, 1583-1586. https://doi.org/10.3390/life5041583

[4] Allamandola, L. *et al.* (2016) Cosmic Distribution of Chemical Complexity. NASA

[5] Lindahl, T. and Karlstrom, O. (1973) Heat-Induced Depyrimidination of Deoxyribonucleic Acid in Neutral Solution. *Biochemistry*, **12**, 5151-5154. https://doi.org/10.1021/bi00749a020

[6] Locke, W. (1997) Introduction to Molecular Orbital Theory. ICSTM Department of Chemistry. https://www.ch.ic.ac.uk/vchemlib/course/mo_theory/main.html

[7] Weiss, I.M., Muth, C., Drumm, R. and Kirchner, H.O.K. (2018) Thermal Decomposition of the Amino Acids Glycine, Cysteine, Aspartic Acid, Asparagine, Glutamic Acid, Glutamine, Arginine and Histidine. *BMC Biophysics*, **11**, Article No. 2. https://doi.org/10.1186/s13628-018-0042-4

[8] Patel, B. *et al.* (2015) Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nature Chemistry*, **7**, 301-307. https://doi.org/10.1038/nchem.2202

[9] Das, T. *et al.* (2019) Insights into the Origin of Life: Did It Begin from HCN and H2O? *ACS Central Science*, **5**, 1532-1540. https://doi.org/10.1021/acscentsci.9b00520

[10] Carter Jr., C.W. and Wills, P. (2017) Interdependence, Reflexivity, Fidelity, Impedance Matching, and the Evolution of Genetic Coding. *bioRxiv*. https://doi.org/10.1101/139139

[11] Martinez-Rodriguez, L., Erdogan, O., Jimenez-Rodriguez, M., *et al.* (2015) Functional Class I and II Amino Acid-Activating Enzymes Can Be Coded by Opposite Strands of the Same Gene. *Journal of Biological Chemistry*, **190**, 19710-19725. https://doi.org/10.1074/jbc.M115.642876

[12] Helm, J. (2021) Code EnzymeCalc.nb. https://www.researchgate.net/publication/358271046_Reaction_constants_in_organic-chemical_reactions

[13] NIST Chemistry WebBook. https://webbook.nist.gov/chemistry/

[14] Chemeo Molecular Database. https://www.chemeo.com/

[15] NIST Chemical Kinetics Data Base. https://kinetics.nist.gov/kinetics/

[16] Helm, J. (2021) Chemical Data Base. https://www.researchgate.net/publication/343182078_ChemicalData_JH0420doc

[17] Chasnov, J. (2009) Mathematical Biology. Lecture in Hong Kong University, Hong Kong.

[18] de With, G. (2013) Liquid State Physical Chemistry: Fundamentals, Modeling, and Applications. Wiley-VCH, Weinheim. https://doi.org/10.1002/9783527676750

[19] Gatenby, R. and Frieden, B. (2016) Investigating Information Dynamics in Living Systems through the Structure and Function of Enzymes. *PLOS ONE*, **5**, e0154867. https://doi.org/10.1371/journal.pone.0154867

[20] Biro, J. (2006) Amino Acid Size, Charge, Hydropathy Indices and Matrices for Protein Structure Analysis. *Theoretical Biology and Medical Modeling*, **3**, Article No.

15. https://doi.org/10.1186/1742-4682-3-15

[21] Wächtershäuser, G. (2000) Life as We Don't Know It. *Science*, **289**, 1307-1308. https://doi.org/10.1126/science.289.5483.1307

[22] NIST Chemistry Webbook. https://www.chemeo.com/

[23] Zahnle, K. (2010) Earth's Earliest Atmospheres. Cold Spring Harbor Laboratory Press, Long Island. https://doi.org/10.1101/cshperspect.a004895

[24] Helm, J. (2021) Biochemistry Basics: A Compendium of Models and Data. https://www.researchgate.net/publication/343182129_BiochemistryBasics_JH0420doc

[25] Helm, J. (2021) Chemical Thermodynamics: A Compendium of Models and Data. https://www.researchgate.net/publication/343181993_ThermodynChemical_JH0420doc

[26] Lodish, H., Berk, H., *et al.* (2000) Molecular Cell Biology. W. H. Freeman, New York.

[27] Landau, L.D. and Lifshitz, E.M. (1987) Fluid Mechanics. 2nd Edition, Pergamon Press, Oxford.

[28] Yaman, T. and Harvey, J.N. (2021) Computational Analysis of a Prebiotic Amino Acid Synthesis with Reference to Extant Codon-Amino Acid Relationships. *Life*, **11**, Article 1343. https://doi.org/10.3390/life11121343

[29] Wimmer, J.L.E., Xavier, J.C., Vieira, A.D.N., *et al.* (2021) Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last Universal Common Ancestor (LUCA). *Frontiers in Microbiology*, **12**, Article 793664. https://doi.org/10.3389/fmicb.2021.793664