UNIVERSITÄT
DUISBURG
ESSEN

From simple molecules to the world of LUCA (Last Universal Common Ancestor)

Ulrich Schreiber, Christian Mayer

EANA – Poster contribution

At the joint conference of German and European astrobiology (DAbG, EANA resp., 24.- 28. Sept. 2018), the pathway for a development of a cell from simple organic molecular building blocks was presented for the first time under realistic conditions. The poster outlines six stages of development which can be shown on fracture zones of the continental crust. The fractures are filled with fluids and gases. The focus is the depth of 700 to 1000 meters, where in CO_2/N_2 mofettes the transition from supercritical to subcritical gases takes place. Due to pressure fluctuations caused by earth tides or by eruptions of cold water geysers the depth of phase transition varies.

Poster explanations

The first stage (**Phase I**) is characterized by formation of organic molecules under hydrothermal conditions, in part comparable to Fischer/Tropsch synthesis conditions. In addition to water, the starting materials are CO_2, CO, H_2, N_2, NH_3, HCN, PO_4 from dissolved apatites, sulfur, boron compounds, alkalis, alkaline earths and dissolved metals. At different depths of the entire crust, molecules are formed under varying pressure and temperature conditions and at varying pH values. The pH depends on the N_2 concentration in the supercritical phase. The molecules are transported by ascending supercritical gases (droplet form) into the 1000 meter region. In micro autoclaves, small pockets of water and supercritical gas, they can react or precipitate when the pressure decreases to such an extent that the phase transition to the gas takes place.

In **Phase II**, three different developments are running parallel.

II.1 In supercritical CO_2/N_2 and at the phase boundary with water, organic bases, ribose and phosphate react to produce random RNA strands. Temperatures can vary over a wider range from the average assumed temperature of 50 °C, depending on the water access from the sides or hot inflow from the depth. Longer RNA molecules have the advantage that they can catalyze their own duplication (ribozyme) as in the case with enzymes. Furthermore, catalysis to varying, new RNA strands is possible. From the multitude of possibilities a primeval transport RNA is developed. Their basic characteristic has been preserved until today. The stability of the RNA is optimal under the given pH values and temperatures of the upper continental crust. From a certain lengths double strands will be formed which are no longer available for duplication. They must be separated (melted) by higher temperatures. Here the pressure changes during the geyser eruptions play an important role. During the start of an eruption, water is pressed out of the fault zone by rising gas, which reduces the pressure in the depth. As the result, the boundary zone of $scCO_2$/gas shifts to the depth. Due to the pressure release the gas expands abruptly. Thus, the temperature in the reaction chambers decreases in the short term (up to 20 °C). This allows complementary RNA strands to be added to double strands. After the eruption, returning water quickly restores the original pressure. As a result, the existing gas is compressed to the supercritical phase. The process is associated with a temperature rise that reaches the melting temperature of the double-stranded RNA. The geyser eruptions thus cause a cyclic replication of RNA strands.

II.2 With the same eruption cycles of the geysers, the formation of vesicles in the limit of 1000 meters of crust depth can be explained. An experiment in a high-pressure chamber shows a cycle in which water, CO_2 and a mixture of long-chain amines and fatty acids (green in the middle Fig.) are put under elevated pressure and temperatures. With a controlled pressure loss, the previously supercritical CO_2 changes to the gaseous state (steps 1 and 2 in Fig. from Mayer et al., 2018). Immediately, the water dissolved in the supercritical CO_2 forms a mist of countless water droplets and collects all organic molecules that can no longer remain in the gas. The amines and fatty acids form a shell on the outer skin (lipid film) and slowly sink to the interface of the lower water body (step 3). Here, too, amines and fatty acids have accumulated in a characteristic orientation and form a continuous layer on the interface. Upon contact of the sinking water droplets with the interface, the incoming droplets form a second shell (step 4). The resulting vesicle resembles a cell in its membrane structure. It consists on the inside of water with low ionic and elevated organic content. On the outside, a lipid bilayer exists, a structure that can be considered as the basis of a proto cell. Clear detection of the vesicle structure was achieved using nuclear magnetic resonance (NMR) spectroscopy. In addition, the first measurements showed evidence of concentration gradients, which are important in a later development of the vesicle as an energy source: The water droplets collect a large number of organic molecules during the condensation in the CO_2 gas. After sinking into the water, the concentration of these molecules in the droplet is orders of magnitude higher than that of the surrounding water. On the other hand, no salts are dissolved in the condensed water droplets. However, the host water in the hydrothermal fault zones is significantly enriched. From these gradients a first proto-cell could have drawn the energy for a simple metabolism. Steps 5 and 6 of the figure indicate the degradation of the vesicles and the restart of the cycle.

In addition to RNA, the formation of amino acids under hydrothermal conditions is possible and already proven. However, only about ten different of the more than twenty canonical amino acids are formed here. It is plausible for physicochemical reasons that there must be a marked imbalance in the formation of the ten hydrothermal amino acids. The simplest representatives will be by far the most common formations. Investigations by Mayer et al. 2018 have shown that under the described conditions of the continental crust, peptides form, which undergo a selective interaction with the envelope walls of the vesicles. Although cyclic disintegration of the vesicles and replacement of the peptides into the shells causes constantly new variations of the peptide sequences. From the multiplicity of the formations, groups with similar properties can be expected. A crucial situation is shown in the upper third of Phase II. Peptides with enzymatic properties form groups with similar contact possibilities to some of the most abundant amino acids. At the same time they come into contact with RNA molecules that are loaded with the amino acid. This process is perfected today in the highly specific loading of the tRNA with an amino acid by the aminoacyl-tRNA synthetase.

Phase III

With **Phase III** a lengthy process of randomly combining "enzymes" begins with amino acids loaded onto proto-tRNAs. In the microautoclaves of the fracture zones there are random sequences of loaded tRNAs. This sequence can occur independently or already with catalytic support of further RNAs from the ribozyme development (today in the ribosome). This process demonstrates the importance of the tRNA molecule. It has on the one hand an amino acid, which can now link to an adjacent one, on the other structural side is a base triplet (now anti-codon), which can be used as an information unit. This triplet can be easily varied by mutations, so that a relatively different base combinations are available relatively early. At this time, however, no information content is present on the tRNAs, since the amino acid species, on the other hand, has no specific association with an enzyme. Nevertheless, the contiguous anti-codons can be used for a template, to which complementary bases attach and combine to form an RNA.

Phase IV

The random ordering of the loaded tRNAs results in a large number of peptides, all of which have a different sequence. Assuming that the two simplest amino acids glycine and alanine predominate, peptides are commonly (over a hundred thousand years) composed only of these two species. The associated anti-codons of the transporting tRNAs are only weakly specific. It results from the newly formed enzymes of Phase II, which can only be distinguished into a few groups. The low specificity results more from a preference of certain tRNAs to a group than an exclusive assignment. At this point, it takes a long time to complete the next steps. The templates provided by the anti-codons may be available for the formation of RNAs.

Phase V

In phase V, the decisive step is taken that reveals the first principles of life. Phase IV develops peptides that consist of only the two simplest amino acids. With infinite combinations over long periods of time, there are two that take on enzymatic functions after folding, in the way they have synthetases. That is, one specifically charges one tRNA with glycine, the other specifically with alanine. The RNA formation from the phase IV corresponds to today's messenger RNA, which is the template for the code of the peptides. Even though the specificity of the "mRNA" is still low at this point in time, the many combinations enable the subsequent production of the first synthetases of phase V over and over again. The higher specific binding of synthetases S1 and S2 in phase V results in feedback to mRNA of phase IV to an ever-increasing specific assignment. The crucial breakthrough comes when the "mRNA" consists of specific codons of the two synthetases S1 and S2 as well as the two RNA strands of the tRNAs (which have attached themselves). It can be arbitrarily copied and thereby secured and with the use as a template for the attachment of the tRNAs, the synthetases S1 and S2 are constantly replicated. **This is the first development that guarantees the preservation of the information system of the two synthetases and their multiplication**.

Parallel to this, the free combination of loaded tRNAs continues. Each time, peptide formation also allows the storage of information about the sequence of the chain in an RNA at the same time. This means that a variety of combinations are created, which are almost 100% unusable but still have documentation in an RNA. Some of these RNA strands can be linked to those already existing. Maybe they will be usable later. What definitely useful is is a possible third synthetase from one of the glycine/alanine chains. This third synthetase, whose sequence is stored in the RNA, involves a third amino acid that is specifically loaded onto a third tRNA. Leading to new opportunities in the free combination. With three amino acids, a much higher combination of amino acids can be accommodated in the peptides, all again storable in a simultaneous emerging RNA. After some time, a fourth synthetase can be formed consisting of the combination of three amino acids which specifically charges a fourth amino acid. This can be continued until the last available amino acid.

Phase VI

Until phase V, all postulated reactions have taken place in an open system of fracture zones, with any exchange and supply of molecules and removal of obstructive substances. In addition to the synthetases, the information stores (the various RNAs) have arisen and a multitude of molecules that have enzymatic functions but no area of application. The process of vesicle formation in Phase II has experimentally demonstrated the pathway of incorporation of diverse molecules into the vesicles. It is a process that is constantly taking place to this day in depth. By the time all the molecules necessary for LUCA were present in large numbers, it was a matter of time that a large number of these molecules would simultaneously occupy one sheath in one cycle. These were already proteins that allowed a transport of substances through the cell wall and enzymes that were catalytically active for the proliferation of lipids and other molecules. The result was a growth of the shell and the proliferation of functional molecules. Above a certain size, the turbulence in the water column from a

geyser outbreak was sufficient to exert shear forces on the cell. The result was a physical division after it was lengthened to a critical length. Due to the high number of all molecules in the starting cell, there were still sufficient molecules in both newly formed subcellular cells that maintained the replication of the cell components. The process of the first division then continued in the daughter cells.
The purely physical division process is hardly comparable with today's cell divisions. The fine-tuned and highly efficient mechanism of the cell, which today causes the division, developed later in the course of evolution.

A few thoughts on how it could have gone on.

The cells were able to ascend through continuous gas transport and reach the surface of the earth with each burst of geyser. Here, new selection mechanisms by UV radiation, low temperatures and fresh water occurred. Organic films could develop from dead cells, providing new conditions. The descendants of these first cells on the surface of the land could have justified the development of the bacteria.
Another line is to be expected by submarine or coastal leakage of the cells, through which contact with Black Smokers was possible. Black Smokers provide similar nutrient conditions as those present in the fracture zones. The special conditions i.a. high temperature gradients, high metal cation load, and low pH levels may have been the cause of the special adaptation and development of these cells on the archaeal line.
Ocean currents caused a rapid spread of the Black Smoker cells (BS cells), the number of hot water outflows at the bottom of the oceans was much higher than today. With each impact of a larger meteorite into the oceans, large areas of the young continents were flooded. This led to a spread of BS cells to other continental fracture zones, whose geysers were virtually vaccinated with the offspring from the deep sea. As a result, completely independent cell lines could develop over periods of many millions of years, which had no contact with the cells of the distant continents. But at a certain point plate tectonics began and the continents began to migrate towards each other. What happened when the fusion of small continents into a larger one occurred? How did the two "endemic" cell types communicate with each other when they suddenly came into masses? Was this one of the causes that led to the process of endosymbiosis, the uptake of parts of one cell type by the other?

Where do the viruses in the hypothetical model come from?

Viruses are "part-time beings". They need other cells for their multiplication, which provide them with the tools to replicate their genome. It is unclear when the viruses first appeared and what functions they had in the history of the development of life. In the compilation of the six developmental phases, they can be assumed early in the transition from phase V to phase VI (development in the free non-cellular area but contact with the first cells).

The problem of chirality

One of the big questions in the development of life is the cause and timing of the handedness of molecules selected in biochemistry. The amino acids occur almost exclusively in the L-form, the ribose of the RNA/DNA in the D-form. The supply of chiral molecules in the initial phase of life development was exactly the same with regard to the enantiomers, the respective L and D forms. How and when did it come to the definition of today's existing forms? There are far-reaching discussions. Based on the development phases shown in the poster, an answer can be identified.

First two comments:

In the free linkage of amino acids to peptides, both enantiomers are incorporated in equal parts in the peptides in a balanced supply of L and D forms. As a result, the peptides cannot optimally reach structure (e.g., alpha helix) by folding. It takes place only with enantiomerically pure chains. The enantiomeric purity of the chain presumably promotes rapid folding, which abruptly increases the stability of the peptide. This applies equally to L-peptides such as D-peptides. Both have completely identical properties, provided that the sequence of the incorporated amino acid species is the same. Thus, if such folded peptides are present at a particular stage of development, only members of one form have been randomly linked.

Another selection option may be due to the interaction of amino acids and RNA. The linkage of an amino acid with a tRNA is always carried out today on the ribose of the last base (adenine) at the ACC end of the tRNA strand. This process must have been set very early in development. For steric reasons, the tRNA structure with D-ribose may favor an L-amino acid (as is the case today) and L-ribose, in turn, may prefer the D version of the amino acid.

The prerequisites described make it clear that there are separation processes for the mixture of the respective forms. But it is not apparent from this that, as a result, one or the other form has been preferred in the further development. Based on the development phases shown in the poster, this means that all steps that take place in phases I to IV are carried out in equal proportions with molecules of both forms in parallel. This changes abruptly with the phase V, in which for the first time the encoded protein synthesis (formation of synthetase) takes place. It is almost impossible that this stage has been reached by both lines of development simultaneously. One line must have reached the state rather than the other and was able to succeed in the future. It was ultimately the line of development with L-amino acids and D-ribose.

A suggestion for another solution of the chirality problem comes from Christian Mayer. He assumes that at the beginning of development for the formation of an RNA not necessarily the chiral molecule ribose was required, but that another achiral molecule (e.g. glycerol) could have taken on the same function in the formation of nucleotides. By linking the nucleotides to an RNA-analogous molecule, chirality arises, and the most readily reduplicating molecule is necessarily chiral. This enantiomer prevails over competitors and gradually imposes its chirality on everything else.

Literature

- Schreiber, U.; Locker-Grütjen, O., Mayer, C. (2012): Hypothesis: Origin of Life in Deep-Reaching Tectonic Faults. Prebiotic Chemistry. Origins of Life and Evolution of Biospheres 42(1) 47 – 54.
- Mayer, C., Schreiber, U., Davila, M.J. (2015): Periodic vesicle formation in tectonic fault zones – an ideal environment for molecular evolution, Orig. Life Evol. Biosph., Volume 45, Issue 1-2, pp 139-148.
- Mayer C., Schreiber U., Dávila M.J. (2017): Selection of Prebiotic Molecules in Amphiphilic Environments. Life 7, 3 doi:10.3390/life7010.
- Schreiber, U., Mayer, C., Schmitz, O.J., Rosendahl, P., Bronja, A., Greule, M., Keppler, F., Mulder, I., Sattler, T., Schöler, H.F. (2017): Organic compounds in fluid inclusions of Archean quartz – analogues of prebiotic chemistry on early Earth. PLOS ONE; https://doi.org/10.1371/journal.pone.0177570.
- Mayer, C., Schreiber, U., Dávila, M.J., Schmitz, O.J., Bronja, A., Meyer, M., Klein, J., Meckelmann, S.W. (2018): Molecular Evolution in a Peptide-vesicle System. Life 2018, 8, 16 (doi:10.3390/life8020016).