RNA/peptide editing in small soluble binding proteins, a new theory for the origin of life on Earth's crust

Jean-François Picimbon

QILU University of Technology, School of Bioengineering, Jinan 250353, Shandong, China

Corresponding Author:
Picimbon JF
QILU University of Technology, School of Bioengineering
Jinan 250353, Shandong, China
Email: jfpicimbon@gmail.com/jfpicimbon@163.com

Running Title: Evolutionary path from protocell to brain

Abstract

We remind about the dogma initially established with the nucleic acid double helix, *i.e.* the DNA structure as the primary source of life. However, we bring into the discussion those additional processes that were crucial to enable life and cell evolution. Studying chemosensory proteins (CSPs) and odor binding proteins (OBPs) of insects, we have found a high level of pinpoint mutations on the RNA and peptide sequences. Many of these mutations are found to be tissue-specific and induce subtle changes in the protein structure, leading to a new theory of cell multifunction and life evolution. Here, attention is given to RNA and peptide mutations in small soluble protein families known for carrying lipids and fatty acids as fuel for moth cells. A new phylogenetic analysis of mutations is presented and provides even more support to the pioneer work, i.e. the finding that mutations in binding proteins have spread through moths and various groups of insects. Then, focus is given to specific mechanisms of mutations that are not random, change α-helical profilings and bring new functions at the protein level. In conclusion, RNA and peptide mutations are not seen as representative of a multitude of diseases, but rather as an alternative way by which protocells developed to acquire multifunction and totipotency. This provides a basis for the theory of RNA/peptide mutations for birth and evolution of life on earth's crust proposed here.

Key Words

Insecta - Chemosensory protein - Odorant binding protein - Peptide mutation - Cell evolution - Abiogenesis

1. Introduction

The appearance and growth of living cells in an environment that constantly changes must not be an evasive subject. While humanity keeps searching for new ways to preserve itself, “How could life start on earth?” remains a key question for humans since all time. The origins of life are our own sets of problem and potential in answering key questions in biological diversity and global biodiversity protection [1].

Theory of biodiversity creation always came from various religious or scientific debates. Soon enough, Charles Darwin’s theory and way to see the world conflict the theory of an unchangeable divine creation from Richard Owen and colleagues [2-4]. The two main ideas from C. Darwin describing the world are as follows: 1) “life originates in simple forms and develops with time into more and more complex systems”. 2) “Species from reptiles to birds, plants to insects and mammals to human are not fixed all time but evolve as a result of natural selection” [2]. This new world made of constant changes is entirely opposed to the old world where no transmutation or change is possible [3-4]. Somehow, the intrinsic rigidity of DNA and the limits imposed by the genetic code on protein synthesis can be named an unchanged system for By, but even this system had to acquire some flexibility in order for the cell to be born and evolve [5-10]. Nucleotide modifications and dual-use codons are as important as cognate tRNAs to underlie flexible decoding process of DNA and its translation into protein, mainly in the recognition of stop codons [9-10].

But the question still remains, if nucleotides build life, how far back in history can we trace their origin and evolution? A recent discovery identifying base nucleotides on the crust of meteorite fragments allow us to trace the origin of DNA and RNA back to about 4 000 Mya, where a rain of meteorites impacts the earth [11-12]. So, space brought the basic molecular building blocks for DNA and RNA, but how life started on earth from these RNA blocks? How much was left to Earth’s crust to develop life from the nucleobases?

Earth at many various times of history has been well suited for a drastic, so huge and phenomenal event such as the appearance of life to happen. Earth's natural environment is a world of rather unstable interactions in an eternally unstable, constantly fluctuated and continously changing climate. There have been expansive forces on earth that broke up enough crust to separate a number of large continents and lead on several occasions to mass extinction in animal and plant species, much before human appearance [13]. Continents are still drifting, and every year rain stones and meteorites fall somewhere on Earth. Earth's sixth mass extinction has already begun. Over the last decade, more than 40% of the mammalian species have experienced more than 80% range shrinkage [14]. Global warming that affects the habitats of numerous species, including human, has severe negative effects on biodiversity, and the effects of global warming on biodiversity is rapidly growing in an uncontrolled way, but it remains that molecules of life and their polymerization into RNA occurred in some "warm" little ponds. Heating was crucial for abiogenesis and the birth of RNA as a polymer, prelude to the first cellular life form on Earth [15].

Importantly, the switch from RNA to cell could be explained by a recent discovery in the silkworm moth *Bombyx mori.* In *B. mori*, we have discovered that one gene or one single RNA strand encoding *chemosensory protein* (or *CSP*) does not produce one single type of protein, but rather an extremely huge diversity of variant isoforms thanks to RNA and/or peptide mutations expressed in a tissue-specific manner [16-19]. This suggests a new theory for life: RNA built in space, assembled on earth at one very long time in the most remote distant past, and was subject to mutations on Earth's crust many thousands Mya over a thermal shock and/or an irradiation peak to produce multifunctional proteins and thereby first protocells (Figure 1). The role of CSPs and/or RNA/peptide mutations in the origin of protocells, i.e. the first self-organized, endogenously ordered, spherical collection of lipids is based not only on their propensity or natural tendency to change structure and conformation, but also on the interaction capacities of this protein family with a grand variety of lipid and fatty acid molecules [16-20]. Small soluble proteins such as csps that can bind to specific lipids and fatty acids have certainly played a key role in the lipid bilayer formation, prelude to lipid membranes and universal cell-membrane structures [21].

2. The Chemosensory Protein (CSP) family

2.1 Definition

Evolutionary study of the CSP proteins traditionally refer to a family of small acidic soluble proteins of 110-120 amino acids with an average molecular weight of about 10-12 kDa, a typical structure that occurs not only in insects and arthropods, but also in the bacteria superkingdom [22-23]. They were first described in various biochemical and molecular studies aiming at characterizing developmental and/or olfactory protein genes in cockroaches and flies [24-27]. Their presence over the whole insect body in particular in dual organs such as antennae and legs was not an analytical artifact or a cross-contamination during tissue extraction, but rather the first indication of multifunction in this protein family [27-30].

Their expression in a wide range of insect and bacterial species, in many non-sensory tissues, and at many different developmental stages, as well as in response to exposure to xenobiotic chemicals and/or infectious microbial agents tackle further the notion of CSP as multifunction system which was introduced early in [22]. CSPs are highly expressed at the adult stage, but their expression peaks also in absence of sensory sructures at very early stages during leg development [22, 27-30]. In addition, CSPs are not found uniquely associated to insect olfactory, taste or sensory cells, but they are also found in many various tissues, including the wings, the gut, epidermis and fat body [27-34]. CSP knock out in bees results in the archaid develoment of the head [35]. CSPs (or pherokines) are even found in the hemolymph of flies in response to viral infection [36]. They are found to be expressed not only in insects and bacteria, but also in aquatic species such as shrimps and the most common species of water flea (NCBI Acc. Nb. ABH88166-ABH88167), showing their occurrence not only among terrestrial systems, but also among marine lives. Coincidentally, CSPs bind to very different types of chemical compounds. They attach to molecules as diverse as linoleic acid or cinnamaldehyde as found in a pionnering study of the chemosensory/immune system of the sweetpotato whitefly *Bemisia tabaci* [37-39]. Interestingly, alike *B. mori*, *CSPs* from *B. tabaci* are found to be subject to specific RNA editing or RNA base mutation in a biotype-specific manner, strongly

suggesting that RNA editing and protein recoding for new function is a general mechanism in the expression of CSPs [37-39].

2.2. Phylogenetic analysis of peptide mutations

Here, it is further shown that the editing of CSP proteins can be generalized to most insect species (Figure 2). An evolutionary analysis of the relationship between edited/mutated BmorCSP fragments, BmorCSPs and their counterparts in bacterial systems, as well as CSPs from very distantly related insect species such as *Aphis gossypii*, *Apolygus lucorum*, *Danaus plexippus, Drosophila melanogaster, Glossina morsitans morsitans, Papilio xuthus, Schistocerca gregaria*, *Sesamia inferens* and *Tribolium castaneum* shows a specific grouping of mutated BmorCSPs with other species-specific sequences.

These sequences have been selected on the basis of their high level of similarity with BmorCSPs. Edited/mutated peptides (p) correspond to CSP fragments sequenced by nano-liquid chromatography coupled to tandem mass spectrometry from a peptide library of the silkworm moth pheromone gland [16]. For parsimony analysis (Bootstrap), amino acid sequences were aligned using ClustalX1.8 and converted to NEXUS. The NEXUS alignment was used as template to build strict consensus trees (conditional bootstrap: 1000 replicates) in PAUP4.010b (Altivec) as described in Swofford [40]. Maximum parsimony (MP) trees (PAUP4.010b, Altivec) are with bacterial odor binding protein (OBP) sequences (WP_071209023, WP_007070378 and CCZ55282) as outgroup [41].

The present Bootstrap analysis in PAUP (56 taxa; 305 total characters) shows that the edited/mutated fragments from *Bombyx* are more closely related to CSPs from other insect species than to BmorCSPs and their counterparts in bacteria. Several mutated sequences form a specific orthology group with CSPs from aphid, beetle, butterfly, fly, locust, noctuid moth, papilio or true bug, but not with BmorCSPs (Figure 2).

The topology of the MP tree is such that the peptide fragment from *B. mori* p1, for instance, is more closely related to the flour beetle CSP sequence NP_001039285 (76%

bootstrap) than to BmorCSPs. Another peptide fragment, p2, is more closely related to locust AAC25400 (75% bootstrap) than to *Bombyx* counterparts. P4, a mutated form of BmorCSP6/WP_089438515 falls more closely related to AGY49271, a CSP from another moth species, *S. inferens* (52% bootstrap). P5, p6 and p7 fall closer to NP_001039288 from *T. castaneum*. P8 fell between BmorCSP10 and *Tribolium* NP_001039286, while p9 appears to be more closely related to Dipteran CBA11329/NP_524966 sequence although without forming any orthology group supported by bootstrap value (Figure 2).

Interestingly, P3 peptide sequence is attracted by a more distantly group of binding proteins (outgroup; OBPs) and falls outside the group of CSP sequences, perhaps suggesting that new families of binding proteins, particularly OBPs, were produced from specific peptide mutations in CSPs (Figure 2).

This new analysis of mutations (peptide variants) in the CSP gene family provides even more support to the previous finding, *i.e.* a high degree of similarity is possessed by these mutated BmorCSPs with CSPs from other insect species. Both percent identity and bootstrap value reflect relatedness between mutation sequences and CSPs from many various insects species from aphids to true bugs. Here is one more argument that CSPs may be regulated specifically on the protein level. These protein mutations seem to be rather very ancient. They are apparently found in a huge diversity of very distantly related species. Either the CSP mutation arose randomly (as most of all mutations do), but once it arose, it gave a distinct advantage to the insect populations, or it existed at a very distant time and gave the incredible possibility for most elementary protocells to come to life. These mutations observed in CSPs may represent an unique mechanism or phenomenon that is conserved and spread in all living organisms from bacteria and insects to most complex living systems, including mammals and human.

2.3 A new mechanism of peptide mutation

The detailed analysis of mutations in CSPs brings the notion of protein change not only at the level of RNA editing, but also later after the polypeptide chain elongation [19, 42]. This assumption comes from the finding of subtle pinpoint single point

mutations at the RNA level, but of peptide fragments with specific amino acid and/or amino acide motif at the protein level. For instance, frameshift mutations leading to shortened proteins but no triplet codons insertion are found for *CSPs* [16]. Sequences of CSP peptide fragments from libraries specific to the female sex pheromone gland reveal many various changes in amino acid motifs, that includes inversions, deletions and replacements of a complete sequence motif [16]. Such mutations or sequence changes cannot be explained by the activity of adenosine deaminase acting on RNA (ADAR), activation-induced cytidine deaminase (AID) or APOBEC enzyme, a mammalian cytidine deaminase known to act on DNA and RNA to convert cytidine to uridine in various metabolic tissues such as the intestine and the liver [43-49]. The number of RNA editing enzymes in insects is rather limited. They express only one type of ADAR, ADAR2 [50]. Even if ADAR2, AID or APOBEC enzymes are involved in the editing of *CSPs*, their accuracies cannot be high enough to induce such a drastic but precise change in the N-terminus of the CSP protein structure [16-20] (Figure 3). Intriguingly, each of these drastic mutations in the N-terminus did not result from a unique combination or re-arrangement of amino acids, but they resulted from a variety of combinations of different amino acid residues added in specific motif. Most of the CSP peptide mutations were found in the N-terminus (Figure 3). Interestingly, p10 peptide mutation overlapped with the signal peptide, which may be important in order to change the secretory export pathway in the cell and/or to extend the N-terminal region of the protein. Other peptide mutations had a precise position on the protein structure. P2 and P3 peptide mutations were precisely added near Cysteine at position 55, while p4 mutation was precisely located near Cys29. P1 and P8 mutations were found closer to the C-terminal end of the molecule (Figure 3). In addition, there was a load of supplementary cysteines (+Cys) added by mutation specifically in the N-terminus of the CSP structure. There were also a large number of CSP fragments characterized by insertion of glycine (+Gly) in the flank of pre-existing Cys and disulfide bridges at specific locations. +Gly enrolled Cysteine at position 29, 36 and/or 55, but never Cysteine at position 58 (Figure 3). Glycine fitted either on the left or on the right of Cysteine at position 55, but insertion of Gly near Cys29 and Cys36 rather oriented right. Insertion of Gly residue on the left of Cys36 was due to inversion mutation of 34-Gly-Pro-35 sequence. Insertion mutations involving residues other than Cysteine or Glycine were also observed. Phe-Val-Phe inserted in the flanck of Cys36, *i.e.* right after the first disulfide bridge. The insertion was always done on

the right at the level of Cysteine residue at position 36, regardless of the nature of the inserted amino acid residue (Figure 3) [16-20]. This would suggest a very precise cellular mechanism capable of inserting specific amino acid residues so precisely in the flanck of cysteines and/or disulfide bridge of the protein.

2.4 Effects of peptide mutations on CSP structure

The CSP protein folds into α-helical molecules made of four Cysteine residues at conserved position over the protein structure (Cysteine 29, 36, 55 and 58). This builds a prism or pyramid-like structure and a binding cavity particularly suitable to transport lipids or small simple aromatic rings [39, 52-54]. The CSP prism carries out six or seven α-helices depending on species [39, 52-54]. The position of α-helices also varies across different species, in particular in the N-terminus [55]. In addition, CSPs are characterized by a very typical breathing mechanism, that is conformational change and cooperativity upon ligand binding [56]. Therefore, both flexibility of α-helical profiling and structural adaptability upon ligand binding can serve multifunction in this class of proteins. These structural properties in CSPs may reflect the functional properties of ancient proteins, *i.e.* the ancestors of all modern proteins that contributed to the beginning of life and survived as molecular fossils. So, studying insect/bacteria CSPs and their structural malleability at both RNA and peptide levels may help understand which motifs or amino acids existed billion years ago and which ones were capable of forming foldable α-helical proteins in a specific (halophile) environment. Overall, this may bring a key step closer to understanding how life first emerged on Earth billions of years ago.

Modeling the protein structure encoded by CSP-RNA mutations cloned in various tissues (antennae, legs, head, wings and pheromone gland) from the silkworm *B. mori* earlier suggested that they can produce multiple kinds of variant CSP protein structures [16-20]. Importantly, +Gly insertion mutation led to removal of specific α-helical maillon (mainly α2) in the N-terminal region of the protein, which could be essential for protein evolution [16-19]. Single α-helices are usually stable, particularly in proteins that are conserved to maintain one and only one single function such as keratin or myosin motor proteins [57]. In this case, two or more α-helices wrap around

each other to form a highly stable supercoil structure particularly suitable for dimerization. However, the stability of α-helical motifs strongly depends not only on their amino acid composition (stable α-helices typically end with a charged amino acid Arg or Glu), but also on their size [58]. A typical stable α-helix contains about 10 amino acid residues (three maillons), which may represent the sufficient length to attain functionality, being therefore the common one-helix ancestor for prokaryotic and eukaryotic proteins [59-60]. The N-terminal α-helices of CSPs are very short; they are made of only a few residues (only 1-2 maillons), which may explain increased mutation rate and flexibility of the N-terminus in this particular protein family [16-20]. The flexibility of the N-terminus described here may be functionally extremely important not only for CSPs, but also for life's first proteins.

In this new analysis of mutations in CSPs, we inserted peptide sequence mutations (p1-p10) into soluble BmorCSP structure (BmorCSP1, 2jnt.1.A) [16-19, 52-53]. First, we performed multiple sequence alignment by CLUSTALW to precisely locate the p mutation on the BmorCSP1 sequence. Then, we used the p-variant CSP1 sequence as a target for modelling of the protein structures produced through or mediated via specific peptide mutations. Proteins structures were modelled in http://swissmodel.expasy.org using NMR structure of soluble BmorCSP1 (100% match) as a template as described in Xuan et al. [17]. Xuan et al. analyzed the structural components of CSP-RNA mutations [17]. Here, we analyzed the impact of CSP-peptide mutations on three dimensional structure and profiling of α-helices (Figure 4).

Interestingly enough, the modelling results described here show that peptide mutations affect the CSP structure and in particular the profiling of α-helices (mainly α2), similarly to CSP-RNA mutations [17] (Figure 4). All the protein models investigated fold into a prism or pyramide made of six α-helices. Therefore, the folding of CSP is not affected by peptide mutations (Figure 4). However, we find that specific peptide mutations enhance reconfiguration of the α-helical stretch, in particular in α2, α3 and α6 (Figure 4). P2, p4, p7, p8 and p10 mutations have no impact on configuration and profiling of α-helices. However, p1 peptide mutation clearly removes the last maillon of the C-terminal α-helix (α6), thereby increasing the

length of the free C-terminal tail (Figure 4). Inserting p3 in the CSP structure removes the last maillon of the central α-helix (α3) at the bottom of the prism. This structural change could affect protein stability, increase protein flexibility and provide new dynamic behaviors, including new protein conformational changes. P5, p6 and p9 peptide mutations drastically shorten the N-terminal α-helix α2, as found for most CSP-RNA mutations [17] (Figure 4), indicating that the N-terminal region and in particular α2 is rather unstable and heavily subject to RNA and/or peptide mutations to lead to a specific protein change, i.e. loss of α-helix [16-19] (Figure 4).

RNA/peptide mutation and the malleability of the α-helical distribution in specific binding protein structures may have been essential for the protein to acquire the ability for conformational changes and multifunctional properties. It is quite possible that both stable and unstable proteins contributed to the pool of functional proteins at the origin of life [60], but most likely conformational dynamics have been essential for protein evolution [61-64]. Therefore, the mechanisms of peptide mutations that we analyze here have certainly played a key role in the origins and evolution of life. Perhaps peptides can form without amino acids, but peptides cannot evolve and acquire new function without mutations [65]. This is probably true not only for enzymes, but also for the vast panoply of extreme multifunctional moonlighting proteins that originally possessed one single function but through evolution acquired multiple functions as diverse as catalytic cores, DNA binding sites, transcription factors, crystallins, protein transporters, chaperons and even receptor molecules [66-67]. Genetic polymorphism and protein conformational plasticity are two ways to promote multifunctionality in Calmodulins. Calmodulin and EF-hand calcium binding proteins can bind more than 300 different target proteins [68]. Genetic polymorphism, protein conformational plasticity and RNA + peptide mutation seem to be essential to promote evolution and multifunctionality in a binding protein superfamily such as CSPs, which might be extremely important and vital to recognize an enormously different set of target proteins or cognate ligands.

3. The Odorant Binding Protein (OBP) family

3.1 Definition

Odorant-binding proteins (OBPs) are small (14-16 kDa) soluble transporter proteins that are thought to increase the capture and the solubilization of odorant pheromone molecules in the aqueous medium or sensillar lymph surrounding the olfactory receptors [69-70]. They are often described in relation with the antennae, the maxillary palps, the salivae and the proboscis for insect olfaction and taste [71-72]. However, several OBPs were reported from many other various tissues and fluids such as legs, wings, venom gland, venom reservoir, the gut and the pheromone gland, which may highlight the flexible multifunctional unstable versatile nature of this binding protein family similar to csps [16-19, 22, 73-75]. In moths, the OBP superfamily is traditionally divided into pheromone-binding proteins (PBPs) and general-odorant binding proteins (GOBPs) dedicated to the recognition of pheromones and plant odors, respectively [76-77]. Both PBP and GOBP structures are characterized by six highly conserved cysteine residues corresponding to six/seven α-helices joined by three interlocked disulfide bridges to form a large central hydrophobic cavity highly suitable to transport fatty acid molecules with about 10-20 carbons such as moth pheromones [78-79].

3.2 Conformational change and RNA/peptide editing

Similarly to CSPs, OBPs are not stable proteins. Their structure exhibits specific flexibility or conformational change upon pH variation [80]. Interestingly, not only structural plasticity but also multiple mutations in RNA/peptide have been described in particular in *PBPs* and *GOBPs* [16-19], suggesting that protein conformational change and RNA/peptide editing are two main mechanisms of binding proteins/ soluble transport proteins for multifunction.

In moths such as the noctuid *Agrotis* species, different types of PBP coexist in the antennae [81-82]. A specific grouping of PBPs has been proposed, and the different groups of PBPs appear to be most variable at the N-terminal site. In addition, multiple mutations in the N-terminus have been found not only for the type 1, but also for the type 2 of PBPs from the black cutworm moth, *A. ipsilon* [81-82]. This emphasized about the importance of the N-terminal region for binding protein flexibility and

multifunction as found for some prion proteins [83]. RNA and peptide editing has been the focus of more detailed study in the analysis of moth PBPs. In *B. mori*, *PBP1* is found to be expressed not only in the antennae, but also in the legs. In both tissues, a huge amount of mutations or RNA/DNA differences have been found mainly in α4, α5 and α6 that form the functional binding pocket of silkworm PBP1 [16-19, 78]. Mutations are also found at the beginning of α3, which constitutes a ground receptacle for various functions in the binding pocket of PBPs [16-19, 78]. Specific mutations in the binding pocket of the PBP protein may mediate a switch of functions from antennae to legs.

A similarly large and diverse spectrum of mutations has been reported for many other *OBP* genes, including genes encoding GOBP, PBP-related protein, sericotropin or protein B1, suggesting that RNA and/or peptide editing is a universal mechanism to underlie multifunction in protein transporters [16-19]. Very interestingly, in contrast to CSPs and GOBPs, the mutations observed in PBPs are shown to result in the addition of an α-helical maillon in the functional structure of the protein [19], suggesting that the various binding protein families evolve and/or acquire multiple functions thanks to RNA editing and specific manipulation of α-helical profilings. If we wish to understand how the primary proteins evolved to multifunction thereby contributing to the appearance of life, perhaps the most important and challenging problem is to find the mechanism or phenomenon behind these subtle mutations in α-helix. As indicated in insects, single amino acid mutations or mutation in amino acid sequence motif would occur in specific regions of various binding protein molecules causing the protein molecule to lose or gain an α-helical maillon, an ancient dominant element of protein structure. This would depend on the binding protein family and the specialized functions of many different types of the original protein biomolecule. The capacity of the original biomolecule to change conformation and/or adopt a new α-helical pattern may have been essential to adapt to environmental conditions. Malleable binding proteins would have been able to provide new metabolic activities, a humble beginning to life on Earth's crust.

3.3. A general mechanism of RNA/peptide editing in living organisms

Changes in gene expression due to RNA editing have been described in other insect species such as the fruit fly *Drosophila melanogaster* and the German cockroach sodium channels [84]. However, mutations are not restricted to ion channel receptors, glutamate receptors and genes involved in neuronal excitability. RNA editing is necessary for a much larger number of protein gene families including genes involved in actin crosslinking, cell motility, circadian clocks, sleep, cytoplasmic protein interaction, ion homeostasis, signal transduction, Alu and KP transposable elements, so it is definitely not only limited to genes involved in neuronal excitability [84-91]. Both pyrimidine and purine conversion are known to be tissue-dependent, mainly expressing itself during brain development and immune response, as well as in the metabolic tract of mammals including humans [92-94]. Therefore, RNA editing seems to be particularly relevant for physiological systems that need to recognize and process thousands/millions different ligand molecules. This is true not only for the sensory system, but also for the immune system and metabolic pathways that need to recognize and degrade thousands/millions different foreign xenobiotic toxic molecules [42].

The diversity of peptide and RNA variants observed in several binding protein families may underlie functional impact of isoform diversity in olfactory structures of insects, *i.e.* detection of specific odor pheromone molecules among millions of odorants. Similar to human, insects have to discriminate more than a trillion olfactory/taste stimuli for adapted behavior [95]. Following mutation-driven evolution, peptide and RNA mutations on odor/taste receptors and/or binding proteins may be an explanation for the seemingly perfect fit of organisms to their environment. This view might apply to other sensory systems in many different types of animals. The *cSlo* gene expressed in chick's cochlear hair cells is known to produce through alternative splicing about five hundred different protein isoforms, all calcium-activated (BK) potassium channels, homologues of the *Drosophila* "slowpoke gene", each of them crucial to recognize a specific sound frequency [96]. RNA splicing and mutation might provide enough molecular variation that, when any new sound occurs, a species can adapt to it. Recognizing anything that can change and become crucial for the living cell might certainly help to overcome any environmental change. In flies, alternative splicing of *Dscam* (Down syndrome cell adhesion molecule) potentially generates about 38,016 different protein isoforms, all transmembrane proteins from the

immunoglobulin superfamily, each comprising one of about 19,008 alternative extracellular domains [97-98]. Perhaps there is even more diversity in cell-surface molecules for neuronal wiring, cell recognition and circuit assembly. It could be that both alternative splicing and RNA/peptide editing are required for production of thousands/millions cell surface molecules such as *Dscam*. An enormous repertoire of Dscams would have been a crucial advantage for protocells to recognize other cells and form diverse tissues, as well as specific organs and many various complex organisms. There are not thousands, but millions of different possible combinations leading to new functional protein when alternative splicing and RNA editing have a combined influence on the expression of a single gene. Because there are even more possible isoforms induced by peptide editing, it could be that the gland, the nervous tissue, the neuron and the gial cell, as well as the lymphocyte in the thymus and its earliest protobiont make use of all these various mechanisms from gene to protein to accomplish multiple tasks from lipid metabolism to excretion of a particular signal. Nucleotide swich in RNA strand and peptide mutation could be how cells interact and communicate with each other from most early forms of life.

4. Evolutionary applications of RNA and peptide mutations in cell growth (totipotency)

Each protocell and/or each totipotent cell in most modern organisms may use this complete extensive panoply of all possible combinations from RNA splicing, RNA editing and/or protein re-arrangement to differentiate into a given tissue in a specific physiological compartment at a precise time. This species response to a precise environmental change may be the conceptual model of a universal phenomenon of adaptation in a changing environment from unicellular protozoa to man through insects and microbes.

4.1 Results of studies of moth tissue

Subtle mutations such as specific amino acid substitution or complete replacement of protein motifs would be a complementary robust modification or recoding of the protein in addition of acetylation, glycosylation, methylation and phosphorylation processes, i.e all important factors regulating cell development and totipotency in many various organisms [99-101]. In particular, RNA and peptide editing may be a key step in the pathway to totipotency in eggs, oocytes and sperm [102]. What is found in *Bombyx* is much more than genomic signature and/or chromatin reprogramming to cell growth and totipotency. It is multiple translational recoding of the transporter/binding protein at RNA and peptide levels in a tissue-specific manner [16-19], which may be extremely useful for the protocells or stem cells that need to grow and develop in a complex environment (Figure 5).

The experiments were done in two families of small soluble binding proteins, i.e. CSPs and OBPs, because we observed numerous nucleotide substitutions and different N-terminal tails for the same protein in these two families in several insect species, including cockroaches and moths [16-19, 27-30, 81-82]. Later, in a biochemical and molecular study to tell about these mutations more precisely, we therefore set out to look out at the nucleotide switch/substitution and the protein isoform diversity in *CSPs* and *OBPs* at the level of individuals in the silkworm moth *B. mori*, an emblematic insect species. For each individual, we compared ten genomic DNA clones and ten cDNA clones from the antennal, leg, head, wing and pheromone gland tissue for each gene investigated (*BmorCSP1*, *BmorCSP2*, *BmorCSP4*, *BmorCSP14* and *BmorPBP1*). *BmorActin* was used as a control. Finally, we checked for the expression of variant isoforms using immunoblots and peptide sequencing by nanoLC-MS/MS [16]. The main finding is in *Bombyx* moth a surprising amount of DNA/RNA mismatches (or mutations) that could not just be some *in vitro* artifacts or the result of reverse transcriptase, PCR or sequencing errors. For most *in vitro* cDNA/RNA amplification, the number of incorporated errors during reverse transcription can only be negligible for one simple reason: the reverse transcription process does not amplify introduced errors in the cDNA strand. For the four *CSPs* and *BmorPBP1*, no mutations were found analysing the sequences from the cloned genomic DNA loci of the gene, but we found stop codon, deletion, insertion, very high level of A-to-I and C-to-U conversion, as well as U-to-A, U-to-G and C-to-I mutations expressed in a tissue-specific manner on the RNA sequence [16]. Interestingly, A-to-G mutations are found about every

dozen or two dozens base pairs in the RNA encoding CSP. This is quite similar to the C-to-U mutation pattern in *CSPs*. Eventually mutations can be found juxtaposed (G-to-A) in the central core of the gene and/or at the 5' end as found in *PBP1* (C-to-U). Mutations are often present after a GG, a CC or a AA double base pair [16]. Some of them had clear effects on CSP protein structures [17]. Most of the mutations were more closely related to edited RNA than to the genomic DNA sequence [18]. Furthermore, analysing peptide sequences, we found even more protein isoforms from a single gene not only in *CSPs*, but also in *OBPs* such as *GOBP2* and *OBP56*, suggesting that RNA and peptide editing are two successive independent events needed in order to make a panoply of new functional isoforms in a variety of binding protein families [16-19].

4.2 "One gene-multiple proteins"

All these findings bring the genetic conceit of protein synthesis in the ribosome very far from one gene producing one single protein. On the one hand, "one gene-one protein" may provide a cell structure perfectly adapted to its environment. However, on the other hand, "one gene-one protein" may not fit with an environment that constantly changes. This may take to have a much more supple and flexible mechanism in cellular systems. RNA and peptide editing as new sources for producing even more molecular variants are ideal associations with alternative splicing to generate a number of mutations that attempt to support multifunction in order to adapt to all possible environmental changes [103-104].

Multi-level gene recoding may be the most important mechanism for regulation of principal cellular processes in all living systems. High levels of RNA editing is not unique to insects or bacteria [16-19, 105]. Tens of thousands mutation sites are commonly found on the RNA from coleoid cephalopods, including squid, octopus and cuttlefish [106-107]. The preponderance of mutation sites in squids is not adaptive [108]. Similarly to squids and other cephalopods, RNA editing is found to be abundant not only in the moth tissue, but also in the bumblebee brain. In a social insect such as the bumblebee *Bombus terrestris*, high levels of RNA editing are found to associate with task or behavioral performance [109]. Following Nei [110], extensive RNA and

peptide editing in the bee brain, in the squid nervous system, in the moth pheromone gland or in the protocell at the most early stage of life argue for multifunction and mutation-driven evolution.

This goes much beyond Darwin's evolutionary theory and principle of natural selection. Not all mutations are rare, restricted and constraint-breaking. Numerous mutation processes at the RNA and protein level seem to precede the environmental change. The reasoning is that RNA and protein mutations can cause even more useful mutations to occur and selection can have a forward influence on these mutations. Thus, gene, RNA and peptide sequence can be changed and recomposed in a complex integration of cellular mechanisms without being lethal or harmful for adaptation and survival of the organism in the natural environment. The natural environment may not trigger and lead specific changes in the cell, eventually selecting only one positive unharmful single point mutation necessary to promote specific skill, adaptive behavior, new phenotype and/or species/organism evolution. The results obtained not only in moths and terrestrial insects, but also in squids and other sea animals show that many types of mutation can precede any changes in the environment, as a trade-off between behavioral output, phenotypic variation, cell potency, multifunction, transcriptome plasticity, protein diversity and genome evolution as proposed here for the protocells and the earliest life-forms, i.e. biomolecules and/or micro-organisms on Earth (Figure 5). Our results in moths show that the cell builds mutations of all sorts not only at the RNA level, but also at the level of the peptide alone, mainly in the α-helical profiling of small soluble binding proteins/transporters, eventually to sense and accurately blend with the environment, but before all to keep and accomplish a series of multiple functional processes tasks, sophisticated performances and primary types of specialized transport mechanisms [16-19, 42]. This is maybe how life arose from non life more than 3,5 billion years ago.

5. From protocell to brain: a new theory for the beginning of life

In the beginning, as newly born cells and living organisms shifted from the RNA world to the protein world, the number of proteins was not large and the enzymes if any probably had rather broad specificity. However, the levels of substrates and

ligands such as fatty acids and lipids were probably highly abundant on Earth's crust in the primordial world, which was already subject to constant environmental changes and drastic mutations. Life largely builds upon key families of chemicals such as sugars, amino acids, fatty acids and lipids (membrane cell walls, fats, phospholipids, general components of cells, fuel molecules, hormone signals and metabolic regulators), which are all a crucial part of it. Lipids as simple carbon and hydrogen molecules probably emerged 3,5 billions of years ago in the ocean water, not so very long after the formation of the Earth (4,54 Bya). Lipids are not mobile molecules in particular in water. It only needed an accumulation of other organic molecules near the amphiphiles in the ancient crust in the deep ocean for the formation of biochemical molecules, first interaction and spontaneous emergence of life (abiogenesis) [111]. With the formation of cells and DNA carrying the hereditary genetic code, it was the beginning of the natural process of selection and evolution that strongly needed some starting materials as crucial as lipid-protein interactions and protein conformational changes to give birth to a more complex world of life, *i.e.* biodiversity. Successful completion of the pivotal evolutionary events in the biodiversity on Earth is certainly the development of the brain, the most complex system ever built by nature.

On the basis of our studies of small soluble binding proteins in insects, CSPs and OBPs, conformational change propensity, their high flexibility not only in protein structure, but also in gene expression, extremely high levels of RNA editing and protein recoding, I speculate that multifunction in early transporter proteins through or mediated by a tremendously large number of mutations or combinations of mutations (RNA + Protein) have been crucial for the assembly of lipids, fatty acids and the first biomolecules in the earliest step of molecular evolution and emergence of life. Multifunction of binding proteins and multifunctional targets may have been crucial also in the ribosome assembly, prelude to the evolution of the genetic code and gene sequence expression [112]. Later, genetic mutations, genome formation, multi-RNA editing and variant peptide multi-function may have had a final impact on the process of edification and development of the brain (Figure 5).

I might, therefore, speculate further on the origins of cellular life on Earth. I propose that not only RNA editing, but also peptide editing and structural flexibility (mainly in α-helix), have been essential for adaptation, defence, reproduction and

thereby survival in changing and often hostile natural environment. Adaptation, evolution and protocell development may not be so different than adaptation, evolution and development of modern cells, in particular human totipotent stem cells. While pluripotent cells would develop mainly through or via RNA editing, a natural path to multiple diverse functional organs and tissues would involve totipotent cells characterized by multiple types of RNA editing combined with multiple types of peptide mutations (Glycine insertion near Cys, addition of Cys in the N and C-tails, new disulfide bond, alteration of the length in N- and/or C-terminus, loss or gain of α-helical links, amino acid inversion, expansion of specific motifs, single amino acid replacement and/or replacement of complete patterns) as described in the moth pheromone gland [16-19, 42]. Polymorphism in RNA and peptide editing mechanisms would be necessary in particular to influence epidermis colour evolution in mimicry, for instance. It would be crucial for immune, metabolic and sensory tissues whereby transporter proteins are exposed to an extremely huge diversity of potential ligands. However, there would be no such drastic RNA/peptide recoding for muscle tissue, for instance, because muscle is only used for locomotion or mastication, therefore mutation amounts a tissue such as muscle can be null without drastically affecting the evolutionary trail (Figure 5).

Charles Darwin's theory of evolution and natural selection is a tremendous life contribution and an essential part of crucial evolutionary concepts from the observation of the global distribution of organisms, shapes and behaviors. Darwin's theory explains the whole world in a biological sense, how species emerge, change, adapt or become extinct in response to environmental changes. It also explains the blossoming of biodiversity, i.e. when a rich variety of new species evolved, but the theory has a hole. It provides a gigantic multiple species comparative phenotype analysis and relay diversity on mutations, but it does not explain what happened before, *i.e.* before the tree of life and the emergence of rich biodiversity. That is probably left to us to analyse the cellular compartments and perform a gigantic multiple species comparative molecular analysis, addressing more particularly the existence of RNA splicing and editing in combination with peptide mutations from bacteria to most evolved mammals, including human.

In my view of the beginning of life, not only RNA mutations but also specific amino acid re-arrangements at the peptide level have played a central role for the genesis of proteins with entirely new functions, prerequisite for cell birth and evolution. In the debate about how life began from biomolecule, the finding about RNA mutations combined with peptide variations applauds to an RNA/peptide world theory in which RNA and α-helical molecules proliferated and mutated before the evolution of DNA. Both RNA and proteins were first, much before the appearance and evolution of DNA and genomes (Figure 5). It is known that RNA molecules are self-replicating entities that carry genetic information, act as catalyst forms and even adopt different conformations for specific enzymatic activity, in a way similar to the action of protein enzymes [113]. Although single-stranded RNA molecules can fold into highly elaborate structures, more powerful and highly specific catalytic functions need to be brought by proteins or peptide polymers, particularly in enzymatic reactions such as lipid or fatty acid desaturation (Figure 5). In our study of small soluble binding proteins in insects, we show that RNA strands differing by a few bases can already lead to an enormous repertoire of proteins, and this molecular repertoire can be drastically expanded thanks to high levels of peptide mutations [16-19]. When a RNA strand can cause the reaction to mutate and copy itself and when a few amino acids can assemble and make a few flexible α-helical motifs in a protoprotein capable of conformational change and/or structural mutation, there is no doubt that life had a chance to form on earth's crust even on a naked surface of quartz crystals or in the middle of iron minerals from an extraordinary change in the environment.

On the basis of RNA and peptide mutations, a plausible scenario for the appearance of life and brain development is shown on Figure 5. The "*theory of RNA and peptide mutations*" exposed here says that:

"*A key element for the appearance of life is that RNA not only produces a large number of 'perfect' copies of itself, but also an extremely large number of copies with subtle tiny mistakes or mutations in the base sequence. The amounts of mutant RNA sequences are such that RNA concentration is now enough so that replication can take place under any plausible abiotic condition. RNA diversity is now enough so that multiple proteins can be built and eventually start to interact with lipid and fatty acid chemicals to form membrane, cell or tissue under the same plausible abiotic condition.*

Meanwhile, a second key element for the appearance of life is that different sets of amino acids combine and permute to produce a large number of small soluble flexible α-helical binding proteins. Protein diversity is now enough so that multiple ligand-binding cavities and ligand-binding pockets can be built for the transport of an enormous repertoire of organic carbon compounds, ligands, chemicals or substances that need to be extracted from the environment under any plausible abiotic condition. It happened in a drop of water to solubilize the protein."

6. Conclusion: Origin and evolution of RNA and protein editing, the universal enigma

Bacteria, cephalopods and insects such as the silkworm moth *B. mori* make crucial models for stem cell research in human and brain evolution, as well as for a better understanding of the birth and origins of life. Featuring the protocells, there are some combined unusual mechanisms for playing with nucleotides and/or amino acids to recompose a gene product encoding transporter/binding protein in a tissue-specific manner in a great variety of modern organisms. How ancient protocells, progenitor cells, embryonic stem cells, T cells and brain neurons can express and regulate such a dense trafficking of RNA transcripts and peptide variants remains to be found. The discovery in the moth tissue is profound to understand a basic phenomenon of the cell and eventually rethink therapeutic medical strategies in genetic, neurological or metabolic diseases.

Analyzing mutations in two different insect binding protein families, CSPs and OBPs, brings the notion of RNA and protein editing throughout the whole universal cellular system linked to carbon compounds. CSPs and OBPs are proteins known to play an important role in lipid metabolism and thereby pheromone biosynthesis, immune regulation and regulation of tissue growth, renewal and development. In particular, they have been shown to bind long fatty acid lipid chains such as linoleic acid, which might serve as fuel in signalling pathways of modern organisms and membrane protocells. It is postulated that the first vesicles were formed from linoleic acid and other fatty acid molecules that would form vesicles highly stable at extreme temperature on the surface of ancient ocean sediments and crude oils in absence of

phosphate in prebiotic times. One key step in the origin of life on earth's crust was probably the self-assembly of lipids and fatty acids with the building blocks of RNA and protein, resulting in an important viable source of aggregate [114]. Binding proteins such as CSPs and OBPs probably took part in the assembly and organization of these various multimolecular complexes, building the block necessary not only for a stem cell to differentiate into a number of specialized cells and tissues, but also for a protocell to acquire and develop vital functions for birth and evolution.

In single cells, small networks, more complex circuit, and across the entire brain, RNA and peptide may have some private mutations, but it could be that many mutations such as A-to-I or +Gly are expressed in many different tissues in many various organisms. So, expression of RNA and peptide mutations is general and may confer a huge and significant advantage of evolution and particularly new protein functions. Most of the mutations described in moths are non-synonymous, numerous and pinpointed in the coding region of the gene, in crucial elements of the functional protein structure [16-19] (see Figures 2-5). This may demonstrate that RNA editing and protein recombination are the crucial mechanisms for the totipotent cells to differentiate into a specific cell type in a given tissue, for the lymphocytes to adapt to antigen in the thymus, for the chemosensory (taste) cells to respond to multiple chemical stimuli, and for the neurons to establish new connection or synapse in the brain, as well as for earliest protocells to adapt to environmental conditions. RNA and peptide editing may be a global view of the basic phenomenon of eukaryotic and prokaryotic cells and present a new perspective on life's origin.

RNA/peptide editing is a new mechanism to find. These types of mutations, ie., protein mutations, particularly Gly and Cys insertions, finely tuned to N-tails and hydrophobic pockets of molecules, were never observed previously. This is the first time such a flexibility in gene expression was identified, probing insects as models to study the epigenetic basis of protein synthesis and multifunction. In the case with multiple RNA editing and protein re-arrangement, as well as expression of a remarkable diversity of mutations at both RNA transcription and protein elongation is frequently seen in various kingdoms of organisms, there might be "hot spots"of mutation not only in the RNA sequence, but also in the protein structure. Perhaps the whole RNA or peptide sequence is hot spot and prone to mutations under specific

circumstances. This would be extremely relevant for a newly born cell that has yet to fully adapt to new environment and sustain not only development, but also natural evolution.

Ackowledgements
Overseas high-level talent and Taishan scholar (NO.tshw20091015).

References

1. Benton, M.J. Origins of Biodiversity. *PLoS Biol.* **2016**, 14, e2000724.
2. Darwin, C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, *Nature (London: John Murray)* **1859**, 5, 502.
3. Owen, R. Review of Darwin's Origin of Species. *Edinburgh Rev.* **1860**, 3, 487-532.
4. Müller, M. Darwinism tested by the science of language. *Nature* **1870**, 1, 256-259.
5. Watson, J.D., Crick, F.H.C. A structure for Deoxyribose Nucleic Acid. *Nature* **1953**, 171, 737-738.
6. Crick, F.H.C., Barnett, L., Brenner, S., Watts-Tobin, R.J. General nature of the genetic code for proteins. *Nature* **1961**, 192, 1227-1232.
7. Koonin, E.V., Novozhilov, A.S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life.* **2009**, 6, 99-111.
8. Geggier, S., Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc Natl Acad Sci USA*. **2010**, 107, 15421-15426.
9. Blanchet, S., Cornu, D., Hatin, I., Grosjean, H., Bertin, P., Namy, O. Deciphering the reading of the genetic code by near-cognate tRNA. *Proc Natl Acad Sci USA*. **2018**, 115, 3018-3023.
10. Babbitt, G.A., Coppola, E.E., Mortensen, J.S., Ekeren, P.X., Viola, C., Goldblatt, D., Hudson, A.O. Triplet-based codon organization optimizes the impact of synonymous mutation on nucleic acid molecular dynamics. *J Mol Evol.* **2018**, 86, 91-102.
11. Callahan, M.P., Smith, K.E., Cleaves, H.J.2nd, Ruzicka, J., Stern, J.C., Glavin, D.P., House, C.H., Dworkin, J.P. Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc Natl Acad Sci USA*. **2011**, 108, 13995-13998.
12. Black, R.A., Blosser, M.C., Stottrup, B.L., Tavakley, R., Deamer, D.W., Keller, S.L. Nucleobases bind to and stabilize aggregates of a prebiotic amphiphile, providing a viable mechanism for the emergence of protocells. *Proc Natl Acad Sci USA*. **2013**, 110, 13272-13276.
13. Feulner, G. Climate modelling of mass-extinction events: a review. *Int J Astrobiol.* **2009**, 8, 207-212.
14. Ceballos, G., Ehrlich, P.R., Dirzo, R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc Natl Acad Sci USA*. **2017**, 114, E6089-E6096.
15. Pearce, B.K.D., Pudritz, R.A., Semenov, D.A., Henning, T.K. Origin of the RNA world: the fate of nucleobases in warm little ponds. *Proc Natl Acad Sci USA*. **2017**, 114, 11327-11332.
16. Xuan, N., Bu, X., Liu, Y.Y., Yang, X., Liu, G.X., Fan, Z.X., Bi, Y.P., Yang, L.Q., Lou, Q.N., Rajashekar, B., Leppik, G., Kasvandik, S., Picimbon, J.F. Molecular

evidence of RNA editing in *Bombyx* chemosensory protein family. *PLoS ONE* **2014**, 9, e86932.
17. Xuan, N., Rajashekar, B., Kasvandik, S., Picimbon, J.F. Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*. *Agri Gene*. **2016**, 2, 53-58.
18. Xuan, N., Rajashekar, B., Picimbon, J.F. DNA and RNA-dependent polymerization in editing of Bombyx chemosensory protein (CSP) gene family. *Agri Gene*. **2019**, 2, 53-58.
19. Picimbon, J.F. Evolution of protein physical structures in insect chemosensory systems. In: Picimbon, J.F. (Ed.), Olfactory Concepts of Insect Control-Alternative to Insecticides, vol. 2, Springer Nature, Switzerland, **2019**, pp. 231-263.
20. Lartigue, A., Campanacci, V., Roussel, A., Larsson, A.M., Jones, T.A., Tegoni, M., Cambillau, C. X-ray structure and ligand binding study of a moth chemosensory protein. *J Biol Chem.* 2002, **277**, 32094-32098.
21. Deamer, D. The role of lipid membranes in life's origin. *Life (Basel)* **2017**, 7, 5.
22. Picimbon, J.F. Biochemistry and evolution of CSP and OBP proteins. *In*: Insect Pheromone Biochemistry and Molecular Biology - The Biosynthesis and Detection of Pheromones and Plant Volatiles. Ed.: G.J. Blomquist, R.G. Vogt, Elsevier Academic Press, London, San Diego, **2003**, pp. 539-566.
23. Liu, G.X., Yue, S., Rajshekar, B., Picimbon, J.F. Expression of Chemosensory Protein (CSP) structures in *Pediculus humanis corporis* and *Acinetobacter baumannii*. *SOJ Microbiol Infect Dis.* **2019**, 7, 1-17.
24. Nomura, A., Kawasaki, K., Kubo, T., Natori, S. Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *Int J Dev Biol.* **1992**, 36, 391-398.
25. Pikielny, C.W., Hasan, G., Rouyer, F., Rosbach, M. Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* **1994**, 12, 35-49.
26. McKenna, M.P., Hekmat-Scafe, D.S., Gaines, P., Carlson, J.R. Putative *Drosophila* pheromone-binding-proteins expressed in a subregion of the olfactory system. *J Biol Chem.* **1994**, 269, 16340-16347.
27. Picimbon, J.F., Leal, W.S. Olfactory soluble proteins of cockroaches. *Insect Biochem Mol Biol*. **1999**, 29, 973-978.
28. Picimbon, J.F., Dietrich, K., Breer, H., Krieger, J. Chemosensory proteins of *Locusta migratoria* (Orthoptera : Acrididae). *Insect Biochem Mol Biol.* **2000a**, 30, 233-241.
29. Picimbon, J.F., Dietrich, K., Angeli, S., Scaloni, A., Krieger, J., Breer, H., Pelosi, P. Purification and molecular cloning of chemosensory proteins from *Bombyx mori*. *Arch Insect Biochem Physiol.* **2000b**, 44, 120-129.
30. Picimbon, J.F., Dietrich, K., Krieger, J., Breer, H. Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). *Insect Biochem Mol Biol.* **2001**, 31, 1173-1181.
31. Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R., Pelosi, P. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur J Biochem.* **1999**, 262, 745-754.
32. Ozaki, M., Wada-Katsumata, A., Fujikawa, K., Iwasaki, M., Yokohari, F., Satoji, Y., Nisimura, T., Yamaoka, R. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science* **2005**, 309, 311-314.
33. Picimbon, J.F. Renaming *Bombyx mori* Chemosensory Proteins. *Int J Bioorg Chem Mol Biol.* **2014**, 2, 201-204.

34. Xuan, N., Guo, X., Xie, H.Y., Lou, Q.N., Bo, L.X., Liu, G.X., Picimbon, J.F. Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. *Insect Sci.* **2015**, 22, 203-219.
35. Maleszka, J., Forêt, S., Saint, R., Maleszka, R. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*). *Dev Genes Evol.* **2007**, 217, 189-196.
36. Sabatier, L., Jouanguy, E., Dostert, C., Zachary, D., Dimarcq, J.L., Bulet, P., Imler, J.L. Pherokine-2 and -3: Two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. *Eur J Biol.* **2003**, 270, 3398-3407.
37. Liu, G.X, Xuan, N., Chu, D., Xie, H.Y., Fan, Z.X., Bi, Y.P., Picimbon, J.F., Qin, Y.C., Zhong, S.T., Li, Y.F., Gao, Z.L., Pan, W.L., Wang, G.Y., Rajashekar, B. Biotype expression and insecticide response of *Bemisia tabaci* chemosensory protein-1. *Arch Insect Biochem Physiol.* **2014**, 85, 137-151.
38. Liu, G.X., Ma, H.M., Xie, H.Y., Xuan, N., Picimbon, J.F. Sequence variation of *Bemisia tabaci* Chemosensory protein 2 in cryptic species B and Q: new DNA markers for whitefly recognition. *Gene* **2016a**, 576, 284–291.
39. Liu, G.X., Ma, H.M., Xie, Y.N., Xuan, N., Xia, G, Fan, Z.X., Rajashekar, B., Arnaud, P., Offmann, B., Picimbon, J.F. Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: role of CSP in insect defense. *PLoS ONE* **2016b**, 11, e0154706.
40. Swofford, D.L. PAUP*: phylogenetic analysis using parsimony, version 4.0b10. Sinauer Associates, Sunderland, MA.
41. Liu, G.X., Picimbon, J.F. Bacterial origin of insect chemosensory odor-binding proteins. *Gene Transl Bioinform*. **2017**, 3, e1548.
42. Picimbon, J.F. A new view of genetic mutations. *Australas Med J.* **2017**, 10, 701-715.
43. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem.* **2002**, 71, 817-846.
44. Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., Neuberger, M.S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol.* **2005**, 22, 367-377.
45. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem.* **2010**, 79, 321-349.
46. Maas, S. Gene regulation through RNA editing. *Discov Med.* **2010**, 10, 379-386.
47. Savva ,Y.A., Rieder, L.E., Reenan, R.A. The ADAR protein family. *Genome Biol.* **2012**, 13, 252.
48. Wang, C., Zou, J., Ma, X., Wang, E., Peng, G. Mechanisms and implications of ADAR-mediated RNA editing in cancer. *Cancer Lett*. **2017**, 411, 27-34.
49. Burgess, D.J. Switching APOBEC mutation signatures. *Nat Rev Genet.* **2019**, 20, 253.
50. Keegan, L.P., McGurk, L., Palavicini, J.P., Brindle, J., Paro, S., Li, X., Rosenthal, J.J.C., O'Connell, M.A. Functional conservation in human and *Drosophila* of metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects. *Nucleic Acids Res.* **2011**, 39, 7249-7262.
51. Mosbah, A., Campanacci, V., Lartigue, A., Tegoni, M., Cambillau, C., Darbon, H. Solution structure of a chemosensory protein from the moth *Mamestra brassicae*. *Biochem J.* **2003**, 369, 39-44.

52. Jansen, S., Zídek, L., Löfstedt, C., Picimbon, J.F., Sklenar, V. 1H, 13C, and 15N resonance assignment of *Bombyx mori* chemosensory protein 1 (BmorCSP1). *J Biomol NMR* **2006**, 36, 47.
53. Jansen, S., Chmelik, J., Zídek, L., Padrta, P., Novak, P., Zdrahal, Z., Picimbon, J.F., Löfstedt, C., Sklenar, V. Structure of *Bombyx mori* Chemosensory Protein 1 in solution. *Arch Insect Biochem Physiol.* **2007**, 66, 135-145.
54. Tomaselli, S., Crescenzi, O., Sanfelice, D., Ab, E., Wechselberger, R., Angeli, S., Scaloni, A., Boelens, R., Tancredi, T., Pelosi, P., Picone, D. Solution structure of a chemosensory protein from the desert locust *Schistocerca gregaria. Biochemistry* **2006**, 45, 1606-1613.
55. Picimbon, J.F., Regnault-Roger, C. Composés semiochimiques volatils, phytoprotection et olfaction: cibles moleculaires pour la lutte intégrée. In: C. Regnault-Roger, B.J.R. Philogene, C. Vincent (Eds), Biopesticides d'Origine Vegetale. 2nd Edition, Lavoisier, Paris, 2008, pp. 383-415.
56. Campanacci, V., Lartigue, A., Hällberg, B.M., Jones, T.A., Giuici-Orticoni, M.T., Tegoni, M., Cambillau, C. Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proc Natl Acad Sci. USA* **2003**, 100, 5069-5074.
57. Simm, D., Hatje, K., Kollmar, M. Distribution and evolution of stable single α-helices (SAH domains) in myosin motor proteins. *PLoS ONE* **2017**, 12, e0177716.
58. Wolny, M., Batchelor, M., Bartlett, G.J., Baker, E.G., Kurzawa, M., Knight, P.J., Dougan, L., Woolfson, D.N., Paci, E., Peckman, M. Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability. *Sci Rep.* **2017**, 7, 44341.
59. Lim, S.A., Bolin, E.R., Marqusee, S. Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. *Elife* **2018**, 7, e38369.
60. Pohorille, A., Wilson, M.A., Shannon, G. Flexible proteins at the origin of life. *Life (Basel)* **2017**, 7, 23.
61. Segev, E., Grumbach, M., Gerber, R.B. Evolution of conformational changes in the dynamics of small biological molecules: a hybrid MD/RRK approach. *Phys Chem Chem Phys.* **2006**, 8, 4915-4923.
62. Jeon, J., Nam, H.J., Choi, Y.S., Yang, J.S., Hwang, J., Kim, S. Molecular evolution of protein conformational changes revealed by a network of evolutionary coupled residues. *Mol Biol Evol.* **2011**, 28, 2675-2685.
63. Maria-Solano, M.A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J., Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem Commun (Camb).* **2018**, 54, 6622-6634.
64. Petrovic, D., Risso, V.A., Kamerlin, S.C.L., Sanchez-Ruiz, J.M. Conformational dynamics and enzyme evolution. *J R Soc Interface* **2018**, 15, pii: 20180330.
65. Cavanelli, P., Islam, S., Powner, M.W. Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* **2019**, 571, 546-549.
66. Huberts, D.H., van der Klei, I.J. Moonlighting proteins: an intriguing mode of multitasking. *Biochim Biophys Acta* **2010**, 1803, 520-525.
67. Jeffery, C.J. Protein species and moonlighting proteins: very small changes in a protein's covalent structure can change its biochemical function. *J Proteomics* **2016**, 134, 19-24.
68. Ikura, M., Ames, J.B. Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality. *Proc Natl Acad Sci USA* **2006**, 103, 1159-1164.

69. Vogt, R.G., Riddiford, L.M. Pheromone binding and inactivation by moth antennae. *Nature* **1981**, 293, 161-163.
70. Steinbrecht, R.A, Ozaki, M., Ziegelberger, G. Immunocytochemical localization of pheromone binding protein in moth antennae. *Cell. Tissue Res.* **1992**, 270, 287-302.
71. Ozaki, M. Odorant-binding proteins in taste system: putative roles in taste sensation and behavior. Picimbon, J.F. Evolution of protein physical structures in insect chemosensory systems. In: Picimbon, J.F. (Ed.), Olfactory Concepts of Insect Control-Alternative to Insecticides, vol. 2, Springer Nature, Switzerland, **2019**, pp. 187-204.
72. Terrado, M., Pinnelli, G.R., Sanes, J., Plettner, E. Binding interactions, structure-activity relationships and blend effects in pheromone and host olfactry detection of herbivorous lepidoptera. In: Picimbon, J.F. (Ed.), Olfactory Concepts of Insect Control-Alternative to Insecticides, vol. 2, Springer Nature, Switzerland, **2019**, pp. 265-310.
73. Calvello, M., Brandazza, A., Navarrini, A., Dani, F.R., Turillazzi, S., Felicioli, A., Pelosi, P. Expression of odorant-binding proteins and chemosensory proteins in some Hymenoptera. *Insect Biochem Mol Biol.* **2005**, 35, 297-307.
74. Danneels, E.L., Rivers, D.B., de Graaf, D.C. Venom proteins of the parasitoid wasp *Nasonia vitripennis*: recent discovery of an untapped pharmacopee. *Toxins (Basel)* **2010**, 2, 494-516.
75. Benoit, J.B., Vigneron, A., Broderick, N.A., Wu, Y., Sun, J.S., Carlson, J.R., Aksoy, S., Weiss, B.L. Symbiont-induced odorant-binding proteins mediate insect host hematopoiesis. *Elife* **2017**, 6, pii: e19535.
76. Vogt, R.G., Prestwich, G.D., Lerner, M. Odorant-Binding Protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *J. Neurobiol.* **1991**, 22, 74-84.
77. Steinbrecht, R.A., Laue, M., Ziegelberger, G. Immunolocalization of pheromone-binding protein and general odorant-binding protein in olfactory sensilla of the silk moths *Antheraea* and *Bombyx. Cell Tissue Res.* **1995**, 282, 287-302.
78. Sandler, B.H., Nikonova, L., Leal, W.S., Clardy, J. Sexual attraction in the silkworm moth: structure of the pheromone-binding protein-bombykol complex. *Chem Biol.* **2000**, 7, 143-151.
79. Zhou, J.J., Robertson, G., He, X., Dufour, S., Hooper, A.M., Picket, J.A., Keep, N.H., Field, L.M., Characterization of *Bombyx mori* Odorant-binding proteins reveals that a general odorant-binding protein discriminates between sex pheromone components. *J Mol Biol.* **2009**, 389, 529-545.
80. Wojtasek, H., Leal, W.S. Conformational change in the pheromone-binding protein from *Bombyx mori* induced by pH and by interaction with membranes. *J Biol Chem.* **1999**, 274, 30950-30956.
81. Picimbon, J.F., Gadenne, G. Evolution in noctuid pheromone binding proteins: identification of PBP in the black cutworm moth, *Agrotis ipsilon*. *Insect Biochem Mol Biol.* **2002**, 32, 839-846.
82. Abraham, D., Löfstedt, C., Picimbon, J.F. Molecular characterization and evolution of pheromone binding protein genes in *Agrotis* moths. *Insect Biochem Mol Biol.* **2005**, 35, 1100-1111.
83. Béland, M., Roucou, X. The prion protein unstructured N-terminal region is a broad-spectrum molecular sensor with diverse and contrasting potential functions. *J Neurochem.* **2012**, 120, 853-868.

84. Song, W., Liu, Z., Tan, J., Nomura, Y., Dong, K. RNA editing generates tissue-specific sodium channels with distinct gating properties. *J Biol Chem.* **2004**, 279, 32554-32561.
85. Stapleton, M., Carlson, J.W., Celniker, S.E. RNA editing in *Drosophila melanogaster*: New targets and functional consequences. *RNA* **2006**, 12, 1922-1932.
87. Riedmann, E.M., Schopoff, S., Hartner, J.C., Jantsch, M.F. Specificity of ADAR-mediated RNA editing in newly identified targets. *RNA* **2008**, 14, 1110-1118.
88. Fukui, T., Itoh, M. RNA editing in P transposable element read-through transcripts in *Drosophila melanogaster*. *Genetica* **2010**, 138, 1119-1126.
89. Stulic, M., Jantsch, M.F. 2013 Spatio-temporal profiling of Filamin A RNA-editing reveals ADAR preferences and high editing levels outside neuronal tissues. *RNA Biol.* **2013**, 10, 1611-1617.
90. St Laurent, G., Tackett, M.R., Nechkin, S., Shtokalo, D., Antonets, D., Savva, Y.A., Maloney, R., Kapranov, P., Lawrence, C.E., Reenan, R.A. Genome-wide analysis of A-to-I RNA editing by single molecule sequencing in *Drosophila*. *Nat Struct Mol Biol.* **2013**, 20, 1333-1339.
91. Sinigaglia, K., Wiatrek, D., Khan, A., Michalik, D., Sambrani, N., Sedmik, J., Vukic, D, O'Connell, M.A., Keegan, L.P. ADAR RNA editing in innate immune response phasing, in circadian clocks and in sleep. *Biochim Biophys Acta Gene Regul Mech.* **2019**, 1862, 356-359.
92. Picardi, E., Horner, D.S., Pesole, G. Single cell transcriptomics reveals specific RNA editing signatures in the human brain. *RNA* **2017**, 23, 860-865.
93. Lamers, M.M., van den Hoogen, B.G., Haagmans, B.L. ADAR1: "editor-in-chief" of cytoplasmic innate immunity. *Front Immunol.* **2019**, 10, 1763.
94. Sapiro, A.L., Shmueli, A., Henry, G.L., Li, Q., Shalit, T., Yaron, O., Paas, Y., Li, J.B., Shohat-Ophir, G. Illuminating spatial A-to-I RNA editing signatures within the *Drosophila* brain. *Proc Natl Acad Sci USA* **2019**, 116, 2318-2327.
95. Bushdid, C., Magnasco, M.O., Vosshall, L.B., Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **2014**, 343, 1370-1372.
96. Ramanathan, K., Michael, T.H., Jiang, G.J., Hiel, H., Fuchs, P.A. A molecular mechanism for electrical tuning of cochlear hair cells. *Science* 1999, 283, 215-217.
97. Neves, G., Zucker, J., Daly, M., Chess, A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nature Genet.* **2004**, 36, 240-246.
98. Hattori, D., Demir, E., Kim, H.W., Viragh, E., Zipursky, S.L., Dickson, B.J. Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* **2007**, 449, 223-227.
99. Corley, L.S., Lavine, M.D. A review of insect stem cell types. *Semin Cell Dev Biol.* **2006**, 17, 510-517.
100. Ishiuchi, T., Torres-Padilla, M.E. Towards an understanding of the regulatory mechanisms of totipotency. *Curr Opin Genet Dev.* **2013**, 23, 512-518.
101. Baker, C.L. Pera, M.F. Capturing totipotent stem cells. *Cell Stem Cell*. **2018**, 22, 25-34.
102. Seydoux, G., Braun, R.E. Pathway to totipotency: lessons from germ cells. *Cell* **2006**, 127, 891-904.
103. Jones, A.K., Buckingham, S.D., Papadaki, M., Yokota, M., Sattelle, B.M., Matsuda, K., Sattelle, D.B. Splice-variant- and stage-specific RNA editing of the *Drosophila* GABA receptor modulates agonist potency. *J Neurosci.* **2009**, 29, 4287-4292.

104. Picimbon, J.F. RNA mutations: source of life. *Gene Technol.* **2014**, 3, 2.
105. Bar-Yaacov, D., Pilpel, Y., Dahan, O. RNA editing in bacteria: occurrence, regulation and significance. *RNA Biol.* **2018**, 15, 863-867.
106. Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J.C., Eisenberg, E. Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell* **2017**, 169, 191-202.
107. Wang, M.C., Mohlhenrich, E. Evolutionary analyses of RNA editing and amino acid recoding in cephalopods. *Cold Spring Harbor Lab.* **2019**, doi:https://doi.org/10.1101/714394.
108. Jiang, D., Zhang, J. The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. *Nat Commun.* **2019**, 10, DOI: 10.1038/s41467-019-13275-2.
109. Porath, H.T., Hazan, E., Shpigler, H., Cohen, M., Band, M., Ben-Shahar, Y., Levanon, E.Y., Eisenberg, E., Bloch, G. RNA editing is abundant and correlates with task performance in a social bumblebee. *Nat Commun.* **2019**, 10, DOI: 10.1038/s41467-019-09453-w.
110. Nei, M. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA* **2007**, 104, 12235-12242.
111. Lancet, D., Segrè, D., Kahana, A. Twenty years of "lipid world": a fertile partnership with David Deamer. Life (Basel) **2019**, 9, pii: E77.
112. Fox, G.E. Origin and evolution of the ribosome. *Cold Spring Harb Perspect Biol.* **2010**, 2, a003483.
113. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Molecular Biology of the Cell. 4th Edition, New York: Garland Science, **2002**. *The RNA world and the origins of life*.
114. Black, R.A., Blosser, M.C. A self-assembled aggregate composed of a fatty acid membrane and the building blocks of biological polymers provides a first step in the emergence of protocells. *Life (Basel)* **2016**, 6, 33.

Legends

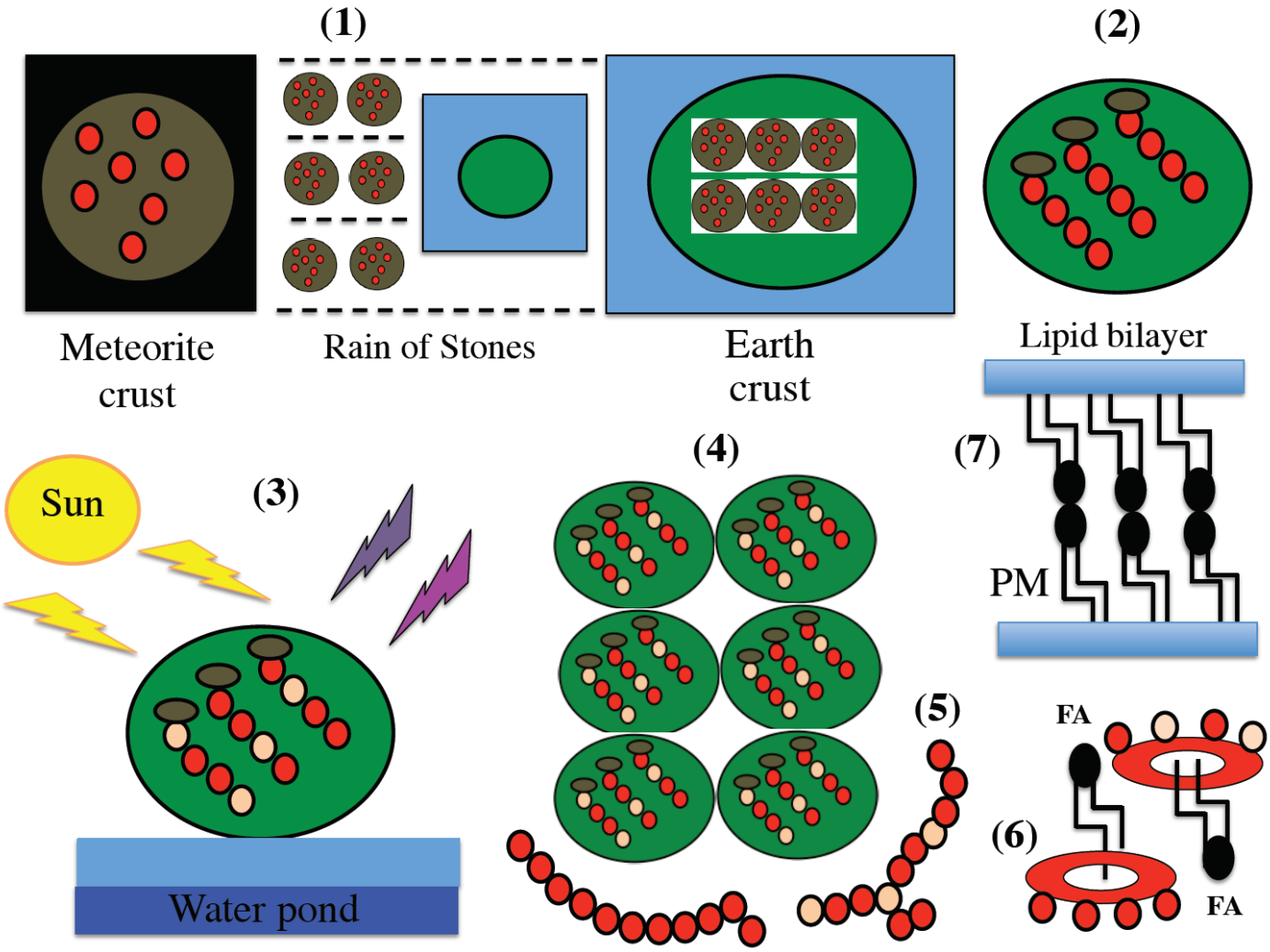

Figure 1: Emergence of membrane protocells from RNA on meteorite's crust.

Life emerged once from RNA building blocks found in meteorite that took shape in deep space. Space rocks and dust once rained onto Earth's crust billion years ago (Bya), "contaminating" the planet with RNA (1). From each piece of crust (lithosphere), the organic building blocks of RNA assembled, thereby producing the first fundamental prebiotic building blocks or polymers of life on Earth (2). Water pond was prior to proceeding to the prebiotic mutation that led to the origin of life. Water (in blue) was necessary to attract molecules and to mix them together for specific chemical reactions to happen. These chemical reactions, largely including mutations of the primitive RNA building blocks, took place under an extraordinary change in environment, with global warming and extreme ultraviolet radiation by sun (3). Then, RNA mutated and replicated on the mineral surface of Earth's crust. Compartmentalization of water ponds allowed RNA protomolecules and mutations to aggregate and self-organize into long genetic and catalytic polymers (4). The diversity of polymers with expansion and often elongation of

the RNA chain was such that some RNAs were faithful copies of the RNA protomolecule, while many others included mutations at specific locations in the polymer structure (5). In this prebiotic RNA world, the different primitive polymers folded in a way to attach fatty acids (FA) and keep them together at relatively high concentration near to each other (6). Nucleobases bound to and stabilized aggregates of fatty acids and long-chain lipids that soon enough assembled into a lipid bilayer that prefigured the structure of the cell plasma membrane (PM) (7). Red dots represent RNA building blocks. In light beige shows RNA building blocks mutated on Earth's crust.

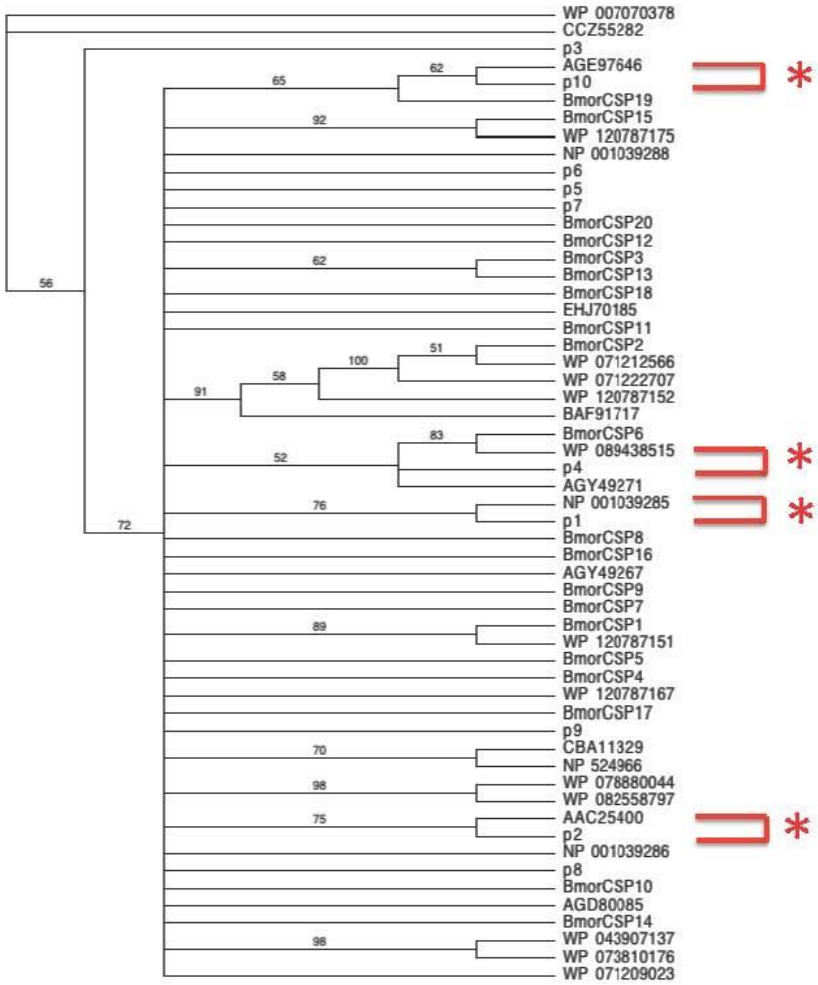

Figure 2: Phylogeny of mutated CSP peptides**.**

Phylogenetic analysis of Bombycidae *B. mori* CSPs (BmorCSPs), bacterial CSPs and mutated peptide sequences (p) with CSP-related sequences from other insect species: AAC25400 (Acrididae, *S. gregaria*), AGD80085 (Miridae, *A. lucorum*), AGE97646 (Aphididae, *A. gossypii*), AGY49267/AGY49271 (Noctuidae, *S. inferens*), BAF91717 (Papilionidae, *P. xuthus*), CBA11329 (Glossinidae, *G. morsitans morsitans*), EHJ70185

(Nymphalidae, *D. plexippus*), NP_001039285/ NP_001039286/NP_001039288 (Tenebrionidae, *T. castaneum*), NP_524966 (Drosophilidae, *D. melanogaster*). p1: YFESQKK; p2: LVPDALSNK; p3: ALSVEEDCAK; p4: KLLVPYLK; p5: WMAVDVACLTDPGYDNLDVDELLDQR; p6: SLGYESKYDNLDVEELLENR; p7: TQYSDVDELLENR; p8: VLRHLLDNKPEMWAK; p9: LLTNDRLFLNYFK; p10: TQLSRPEDVK [16, 23]. Maximum parsimony trees (PAUP4.010b, Altivec) are with *Acinetobacter* and *Dialister* OBP-like sequences (WP_071209023, WP_007070378 and CCZ55282) as outgroup [46]. Starting seed: 174083080, Tree Length: 2713, CI: 0.495, RC: 0.153, HI: 0.505, G-fit: -55.372 (number of parsimony-informative characters = 149). The asterisks (*) following the bars in red show the position of orthology groups between mutated BmorCSP peptide sequences (p) and CSP-related sequences from other insect species.

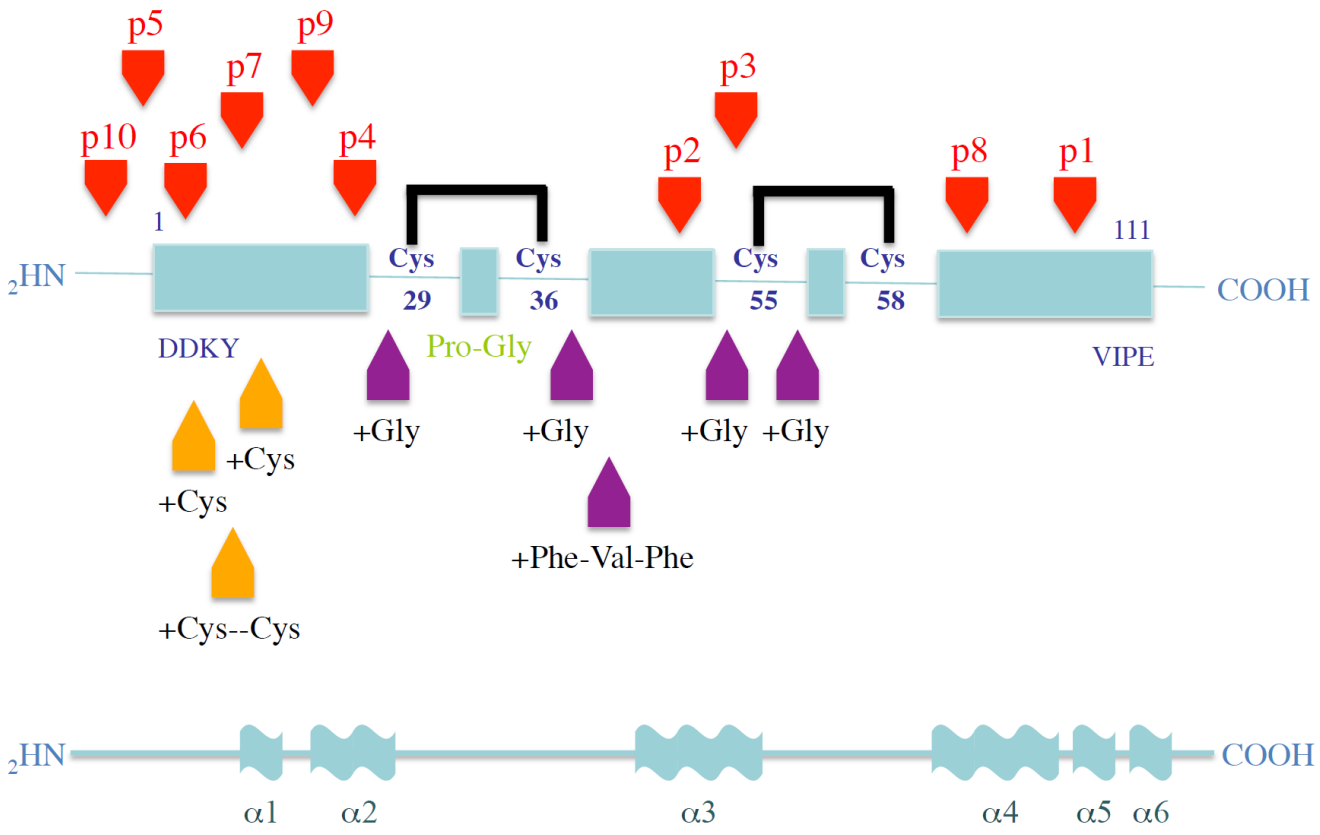

Figure 3: Localization of peptide mutation on CSP structure.

Typical primary structure of CSPs. 1-DDKY and VIPE-111 correspond to N- and C-terminus, respectively. The identity and the position of the first amino acid residue is from N-terminal sequencing of purified native soluble CSP protein [27, 29]. Cys29,

Cys36, Cys55 and Cys58 are four cysteines at conserved position. Disulfide double bonds (represented by bold black lines) involve adjacent cysteines. The α-helical profiling (six α-helices) is shown below the amino acid blocks (2jnt.1.A) [52-53]. Red arrows show the position of specific peptide mutations (p1-p10). p1: YFESQKK; p2: LVPDALSNK; p3: ALSVEEDCAK; p4: KLLVPYLK; p5: WMAVDVACLTDPGYDNLDVDELLDQR; p6: SLGYESKYDNLDVEELLENR; p7: TQYSDVDELLENR; p8: VLRHLLDNKPEMWAK; p9: LLTNDRLFLNYFK; p10: TQLSRPEDVK. Orange arrows show the position of single Cys insertion or double Cys insertion mutation. Purple arrows show the position of Glycine insertion or insertion of Phenylalanine-Valine-Phenylalanine motif near Cysteine at a specific position [16].

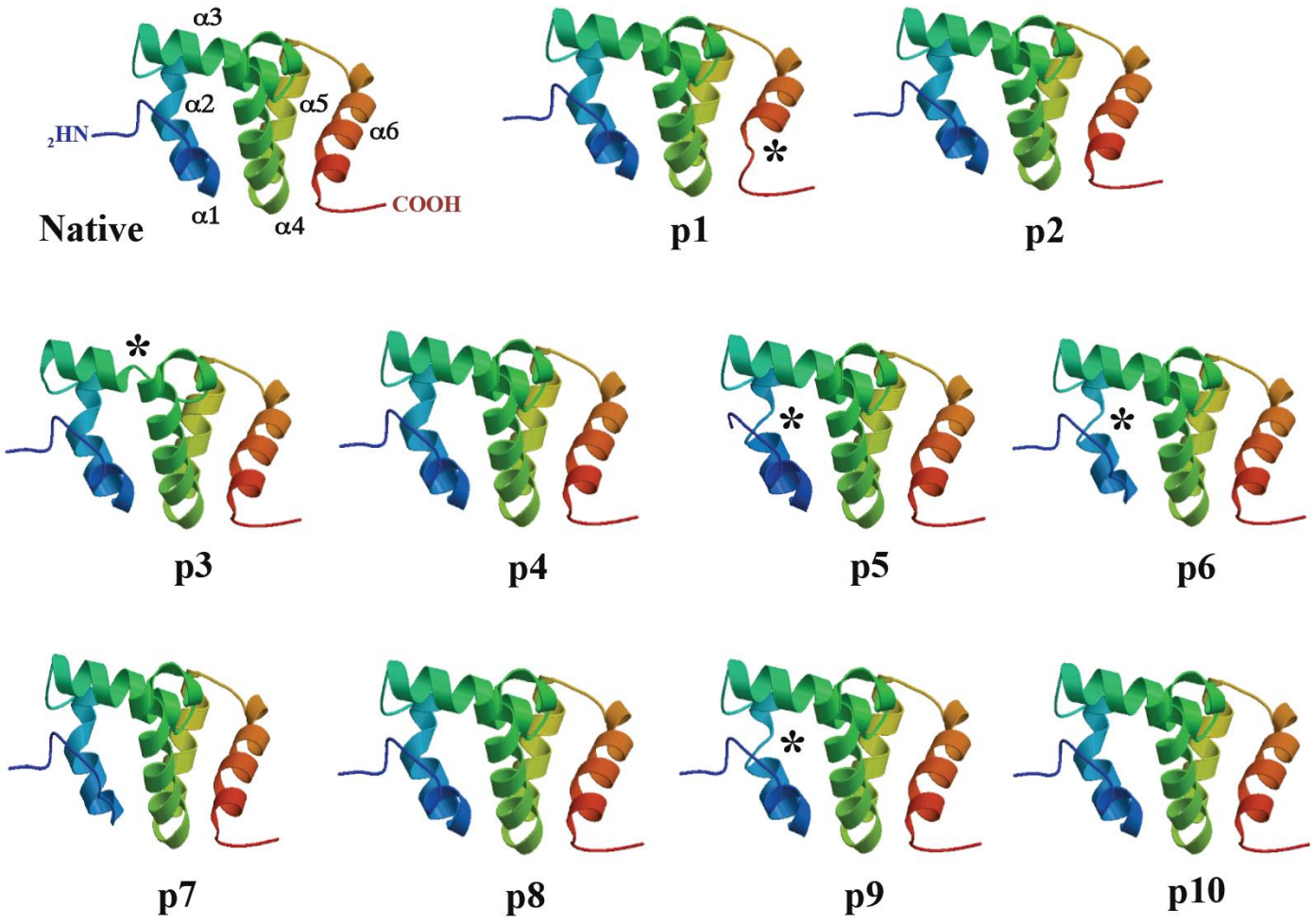

Figure 4: Modelling of protein structures produced through or mediated via CSP-peptide mutations. Comparison of the NMR structure of soluble BmorCSP1 (native; 2jnt1.A) [29, 52-53] and model structure of CSP1 with specific peptide mutations (p1-p10) [16-19]. p1: YFESQKK; p2: LVPDALSNK; p3: ALSVEEDCAK; p4: KLLVPYLK; p5:

WMAVDVACLTDPGYDNLDVDELLDQR; p6: SLGYESKYDNLDVEELLENR; p7: TQYSDVDELLENR; p8: VLRHLLDNKPEMWAK; p9: LLTNDRLFLNYFK; p10: TQLSRPEDVK. The identity and the position of the first amino acid residue is from N-terminal sequencing of purified native soluble BmorCSP1 protein [29]. Variant protein structures are modelled in http://swissmodel.expasy.org using NMR structure of soluble BmorCSP1 as a template (native, 2jnt1.A) [29, 52-53]. The asterisk (*) in black indicates specific change in the protein structure (loss of α).

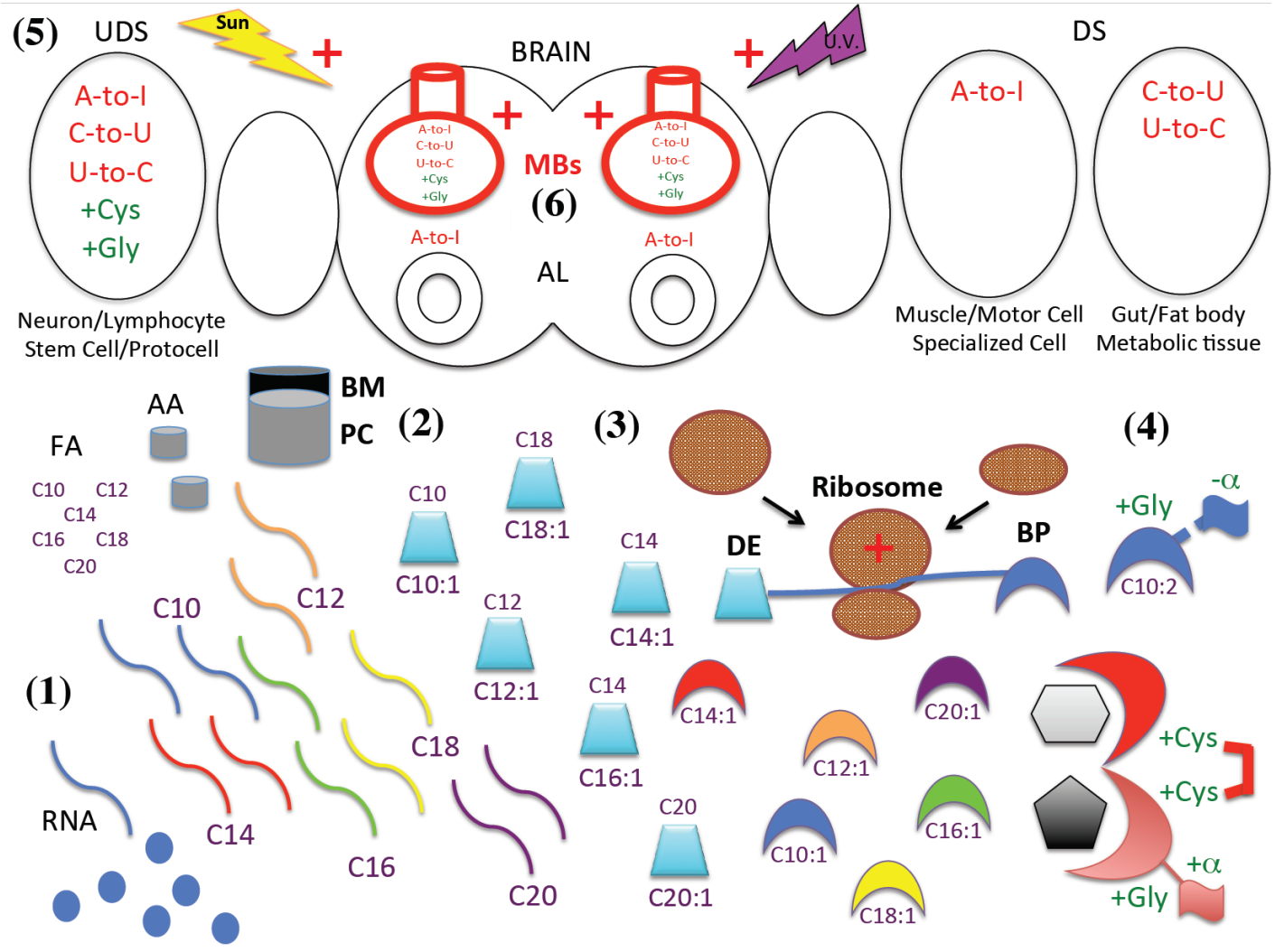

Figure 5: Theory of RNA and peptide mutations for the origin of life, cell differentiation and evolution.

Evolution from protomolecules of single RNA strand to specialized tissue and differentiation of stem cells into brain neurons. In the RNA world (about 4 Bya), RNA molecules (in blue) make copies of themselves from a mix of nucleotides (represented by blue dots). Pieces of RNA naturally stick to small fatty acids or FA (ten carbons, C10).

RNA has the ability to produce a large number of ‘perfect’ copies of itself, but also an extremely large number of copies with subtle tiny mistakes or mutations in the nucleotide sequence. Mutant pieces of RNA have the ability to interact with FAs differing by length (orange: C12, red: C14, green: C16, yellow: C18 and purple: C20). So, RNA-based proto-life-form replicates and binds to FA (1). This helps concentrate certain lipids with primitive protein bits (amino acids, AA), form small lipid bilayer membranes (BM) and construct early protocells (PC) (2). In the protocell, proto-RNAs and AAs combine to form the ancient proto-ribosomal sub-units that will evolve for protein biosynthesis. Soon enough, using proto-RNA and AA, the ribosome evolves the ability to synthesize diverse proteins that perform either enzymatic (DE) or transport (BP) functions. DE: Desaturase Enzyme, BP: Binding Protein. Specific FA-desaturases introduce one C=C double bond (:1) in saturated hydrocarbon chain of lipids. This helps increase the chemical diversity of lipids and FAs in the protocell. Meanwhile, the newly born ribosome makes small soluble BPs to transport specific FA chains in the different compartments of the protocell. Variant mutant forms or isoforms of BPs serve to transport the high diversity of lipids and fatty acids (3). Specific mutations such as Glycine insertion (+Gly) lead to structural change and/or loss of α-helix (-α), resulting in a new binding pocket more suitable to interact with fatty acid chains having two C=C double bonds or (:2). Other types of mutations including Cysteine insertion in N and C-tails of BPs lead to form cross-linkages and intermolecular disulfide bridges of critical importance to snap two BPs into a covalently linked heterodimeric complex. Heterodimers develop functional binding sites for new types of ligands or molecular shapes such as cyclic and ring compounds (4). Multiple RNA/peptide profiling and mutation pathways remain in elements of the undifferentiated system (neurons, lymphocytes, stem cells and protocells). Multi-RNA and multi-peptide mutations also remain in the integrative complex units of mushroom bodies (MBs). Relay centers such as the antennal lobe (AL) only exhibit RNA editing (A-to-I). Specific RNA mutations (A-to-I or C-to-U/U-to-C) also occur in differentiated systems (DS) such as the muscle tissue and the digestive tract (5). High temperature and ultraviolet radiation bring RNA and peptide editing mutation loads to brain and MBs for recognition and adaptation to global environmental change. The RNA/peptide mutation machinery is induced in the ribosome, allowing the newly born cell world to reach a high level of biochemical sophistication (+) (6).