



Review Article

Crucial steps to life: From chemical reactions to code using agents



Guenther Witzany*

Telos-Philosophische Praxis, Vogelsangstr. 18c, 5111 Buermoos, Austria

ARTICLE INFO

Article history:

Received 18 November 2015
 Received in revised form 5 December 2015
 Accepted 7 December 2015
 Available online 23 December 2015

Keywords:

Non-coding RNAs
 Repeat sequences
 Genetic parasites
 Natural code users

ABSTRACT

The concepts of the origin of the genetic code and the definitions of life changed dramatically after the RNA world hypothesis. Main narratives in molecular biology and genetics such as the “central dogma,” “one gene one protein” and “non-coding DNA is junk” were falsified meanwhile. RNA moved from the transition intermediate molecule into centre stage. Additionally the abundance of empirical data concerning non-random genetic change operators such as the variety of mobile genetic elements, persistent viruses and defectives do not fit with the dominant narrative of error replication events (mutations) as being the main driving forces creating genetic novelty and diversity. The reductionistic and mechanistic views on physico-chemical properties of the genetic code are no longer convincing as appropriate descriptions of the abundance of non-random genetic content operators which are active in natural genetic engineering and natural genome editing.

© 2015 Elsevier Ireland Ltd. All rights reserved.

Contents

1. Introduction.....	49
2. Deficits of the old main narratives in molecular biology and genetics	50
3. The agents of natural genetic engineering and natural genome editing.....	50
3.1. Non-coding repeat syntax vs. coding non-repeat syntax.....	50
3.1.1. Genome fragmentation and assembly and the role of repeat syntax.....	51
3.2. Natural genome editors differentiate between self and non-self.....	51
3.3. RNA world agents: where the genotype is the phenotype.....	51
4. Infectious parasite lifestyle results in immune systems	52
5. Crucial steps to life: from abiotic chemistry to biotic code usage	52
5.1. Interacting RNA groups generate the first natural code	53
5.2. Essential tools in RNA group building: ligase and endonuclease.....	53
5.3. A new domain of life: natural genetic identity producers.....	54
5.4. How to do things with sequences.....	54
6. DNA innovation: unstable RNA productivity and stable DNA storage medium	54
7. Conclusion	55
References	55

1. Introduction

While the RNA world hypothesis is strongly supported by an abundance of detailed results in RNA biology and virology the origin of biological selection in contrast to prebiotic chemistry remains unknown. Although several promising approaches investigate this borderline between life and non-life empirically and theoretically,

i.e. coherent and essential use of nucleic acid codes for genetic information generation, storage and heredity, a breakthrough with convincing explanatory power is lacking (Lehman, 2015). The complementary features of genetic phenomena that are clearly absent in prebiotics, such as usage of (i) combinatorial rules of nucleotide syntax, (ii) the context dependent (pragmatic) transcription that lead to (iii) the final function (semantics) of the transcription products are not part of quantitative analyses in molecular biology and genetics. Therefore it is less helpful to simply ignore these features and to propose a gradual emergence out of prebiotic chemistry into biology (Eigen and Schuster, 1982; Szathmari and Maynard

* Tel.: +43 62746805; fax: +43 62746805.
 E-mail address: witzany@sbg.at

Smith, 1997; Mayr, 2004; Goldenfeld and Woese, 2011; Koonin, 2011; Yarus, 2011; Wächtershäuser, 2014). This review will therefore look at essential features of natural codes that may lead to a new definition of life.

2. Deficits of the old main narratives in molecular biology and genetics

For decades it was assumed that if the genetic code represents a context-free universal grammar determined exclusively by natural laws, copying events can form only identical copies of the wild type. If there is any change in such copies it is correct to term them as replication errors. Error replications (mutations) are a key element in explanatory models of the evolutionary origin of genetic diversity, being the ultimate resource for biological selection.

More detailed, error-prone RNAs are a resource for generating genetic diversity followed by selection out of cumulative series of error replications. If the error rate is too high, then genetic content gets lost, followed by quasi-species collapse. Thus the genetic code looks like an emergent property of self-organizing matter and information that underlies natural laws such as thermodynamics, kinetics, atomic structure and their binding forces (Eigen and Schuster, 1977, 1978, 1982). Therefore the natural genetic code as well as human language can be fully investigated by information theory, i.e. a mathematical theory of communication, game theory or related mathematical concepts of language, because formalizable procedures are the best tools for depicting the logic of material reality (Eigen and Winkler, 1981). This logic of material reality determines the physics and chemistry of natural codes (Schuster, 2011) as well as natural code-generating systems like cells, organs (e.g. the human brain) and organisms (Eigen and Winkler, 1981).

If we look at the genetic code and its characters, the rules of combination detected by Erwin Chargaff which were later termed “Watson-Crick base-pairing” (Forsdyke and Mortimer, 2000), and the information-storing function essential for heredity, we can confidently support Manfred Eigen’s claim that the genetic code represents a real language and not just a metaphor. “All the words of the molecular language compose a meaningful text that can be divided up into sentences” (Eigen and Winkler, 1981, p. 272). “Nature’s two great evolutionary processes – the development of all forms of life and the evolution of the intellect – both depended on the existence of language” (Eigen and Winkler, 1981, p. 282).

Yet the concept of language and communication Eigen used is based on Turing and von Neumann’s machine concepts, Shannon and Weaver’s information theory and Chomsky’s universal grammar concept. Current empirical knowledge about language and communication contradicts these assumptions fundamentally (Witzany, 1995, 2014a). Real lifeworld communication by living agents cannot be depicted by mathematical equations because communicating living agents do not behave mechanistically and the syntactical structure of natural codes does not represent its meaning, which in most cases is not available in the superficial but in the hidden deep grammar (Witzany, 1995). In such outdated concepts a natural code like the genetic code is reduced to the sum of its measurable and superficial attributes which can be computed in a digital and algorithm-based procedure.

The coherent generation of *de novo* sequences by competent agents in real-life situations cannot be part of algorithm-based analyses (Witzany, 2010a, 2012), because such non-random processes of competent productivity cannot be explained by quantitative analysis. Nevertheless, Manfred Eigen and Peter Schuster used this concept to explain the evolution of quasi-species and life in general (Eigen and Schuster, 1977, 1978, 1982). Because their theoretical assumptions on the essential features of natural languages used in communication processes do not match the current empirical data about consortial interacting RNA agents,

the question arises how a new concept could explain the origin of the genetic code more coherently by integrating the role of viruses, mobile genetic elements and non-coding RNAs in parallel.

Also, more recent approaches that identify the key players in the emergence of biological information such as polypeptides, RNA-like polymers, and lipids and the known parameters such as viable cores, connectivity kinetics, information control, scalability, resource availability, and compartmentalization which may guide the aggregate evolution of collectively autocatalytic sets remain on the molecular stage and cannot explain how the genetic code molecules emerged into a real natural code (Nghe et al., 2015; Shay et al., 2015) with three inherent levels of rules (syntax, pragmatics, semantics). Mathematical analyses of network formation cannot explain the rule-based real-life interactions of social groups that share genetic identities within historically grown ecospheres (Witzany, 2005, 2014b).

3. The agents of natural genetic engineering and natural genome editing

Since Barbara McClintock’s far-reaching detection of mobile genetic elements, the RNA world hypothesis followed by the rise of RNA biology and the comeback of virology, we have a variety of terms for RNA agents that play essential roles in natural genetic engineering and natural genome editing (Witzany, 2010a,b, 2015a). Crucial steps in systematizing the available empirical data that show an abundance of different pathways by which cells generate genetic diversity and gene regulations constituted the concept of natural genetic engineering outlined by Shapiro (2009, 2016). An abundance of investigations on non-random genetic content operators such as transposons, retrotransposons, long terminal repeats, non-long terminal repeats, various forms of introns, phages and plasmids is described in his publications (Shapiro, 1992, 2005, 2011).

Of equal importance was the comeback of virology which coherently explained sub-cellular genetic content organization by persistent viruses, infectious DNA or RNA, genetic parasites, selfish DNA, and the techniques of insertion, deletion, restriction, modification, counterbalancing toxin and anti-toxin “addiction modules” that re-regulate their counterpart according to contextual needs (Villarreal, 2005, 2009a, 2012). Accordingly the concept of natural genetic engineering had to be completed by that of natural genome editing (Witzany, 2010a).

Interestingly, all of these agents involved in natural genetic engineering share a common feature in genetic syntax: they represent repeated nucleotide composition (Shapiro and von Sternberg, 2005; von Sternberg and Shapiro, 2005; Jurka et al., 2007). Additionally we find this phenomenon in very conserved DNA genome content such as telomeres and centromeres (Witzany, 2008; Alliegro, 2011).

3.1. Non-coding repeat syntax vs. coding non-repeat syntax

One of the key elements that provides RNA content to be translated into DNA is reverse transcriptase, which has polymerase, endonuclease and RNase H abilities, also represented by repeated sequence syntax. Therefore we should remember this feature because it contrasts all the protein-coding nucleic acid sequence syntax, which is coded in non-repetitive nucleotide sequences. In contrast with the concepts of former decades it is now acknowledged that, together with the large variety of non-coding RNAs, RNA agents represented by repeat sequences are key to all cellular and genetic regulatory processes (Shapiro, 2009; Mattick, 2009; Qureshi and Mehler, 2012; Moelling and Broecker, 2015). This means RNA agents share a repeat sequence syntax whereas the

nucleotide sequence of DNA that codes for proteins is characterized by non-repeat syntax.

Recently it has been acknowledged that repetitive nucleotide syntax is a pre-requisite to recombination and rearrangements, i.e. the module-like tool function of the genetic code depends somehow on the repeat grammar of nucleotide sequence order (Sciamanna et al., 2009; Deng et al., 2012). The reason for this is currently unknown, but if we look at cellular life constituted by protein-based cells and coded by non-repeat syntax and its regulatory (repeat syntax) RNA agents we can differentiate a level of genetic information which contains competent regulation relevant for coded proteins by a non-repeat nucleotide syntax.

In the early 1990s a research team from Boston University investigated quantifiable aspects of coding and non-coding genetic sequences such as statistical measurements of nucleotides of the molecular alphabet. While coding DNA is predominantly short-range correlated, the order in non-coding DNA is largely determined by long-range correlations. The authors suggested that their “results are consistent with the possible existence of one (or more than one) structured biological language(s) present in non-coding DNA sequences” (Mantegna et al., 1994, p. 3172).

From the action theoretical perspective a side result is more interesting: The tests revealed that non-coding DNA more closely resembles everyday language, i.e. concrete speech acts, while the structure of coding DNA more closely resembles that of artificial (e.g. scientific) languages. Current knowledge about the functions of these two language domains suggests that everyday language is the medium of getting socialized within group interactions by getting the competence to correctly combine linguistic utterances with the meaning function of words. In contrast to this artificial languages are based on special conventions determining the meaning of certain words and narratives in limited (such as scientific) language communities.

Whereas the first one is essential for any biotic group building as a primary tool for co-ordinating and organizing real-life situations, the latter is a useful tool for constructing blueprints for complex artificial objects (Witzany, 2010b). This fits with the role of non-coding (repeat) sequences representing key elements of interaction, recombination, regulation and innovation and the coding regions (for proteins) as conserved blueprints for complex protein bodies.

3.1.1. *Genome fragmentation and assembly and the role of repeat syntax*

One interesting phenomenon is how genes that code for proteins in cellular organisms are dispersed throughout the genome. In alternative splicing pathways the separating (repeat syntax) introns are removed, and the remaining (non-repeat) exons that represent genes that code for proteins are assembled into a coherent line-up of coding regions to form an mRNA that is transported out of the nucleus to the tRNAs for translation into proteins.

As mentioned recently by Ricardo Flores, some RNA viruses behave similarly to mRNAs (Minoia et al., 2015). They move towards the ribosome of a cell and provide instruction to the ribosome to synthesize proteins. Interestingly, this ability is essential to many RNA agents for synthesizing new RNA molecules de novo using the RNA-dependent RNA polymerase (McDonald, 2013). This may indicate the evolutionary pathway from early RNA infections of cellular host organisms to its new use as exapted tool in host messenger RNAs.

Especially the eukaryotic genome full of repeat elements derived from former infectious RNA (Villarreal, 2005), that serve as crucial regulatory tools in a variety of non-coding RNAs of both long non-coding RNAs and micro RNAs (Delihias, 2011; Cox and Sullivan, 2014). Why are such intronic regions or gene coding regions so much dispersed throughout the genome? Or is this not restricted to

coding regions? Are there also long non-coding RNAs with important regulatory functions dispersed throughout the host genome (Wang et al., 2011)? What is the point of genome fragmentation as an evolutionary conserved technique? Is it for immunity against genetic parasites, in that an attack would probably reach only small fractions of the whole genetic identity? Then genetic fragmentation would be related to adaptive immunity systems such as VDJ and even CRIPRS/Cas systems (Sommer, 2005; Mojica et al., 2005). Is it because of energy optimization or maybe because the several parts may serve several functions at different loci?

3.2. *Natural genome editors differentiate between self and non-self*

How did prebiotic chemistry start life? From the pragmatic action theoretical perspective the variety of RNA agents that share the competence to act on nucleic acid sequences by several techniques (de novo generation, ligation/degradation, insertion/deletion, silencing, amplification, epigenetic markings, etc.) is coherent with the general assumption that no natural code codes itself but needs some competent agents that act on this code (Witzany, 2011, 2012, 2014b; Witzany and Baluska, 2012; Salmena et al., 2011).

Now we can ask how these agents evolve, how they interact, co-operate or compete, or even both, at different times, depending on the varying real life context. All such interactions depend on an ability to identify appropriate interaction partners and therefore differentiate between self and non-self, which means differentiating identity-sharing agents of proponents representing non-identity, which must be fought or precluded (Marraffini and Sontheimer, 2010). Others may represent co-operation partners; some are actually interaction partners. Others remain as silenced enemies which could destroy identity but may be useful tools for later co-opting or ex-apting processes. At different times formerly fought non-identity proponents may be useful tools for generating new partnership (Villarreal, 2011a,b, 2012, 2014). Several DNA and RNA degrading processes are not investigated under the assumption that they serve as useful resource for re-use and identity-building modules (Villarreal, 2014).

Additionally important and a key feature of this agent based co-operativity are the fast-changing identities which lead to the question how a former enemy becomes a new co-operation partner. What does this mean for former co-operative partners which may turn into enemies, the fast-changing functional partnerships and the concurrent changes in organization and syntax order? As these interaction processes start from the beginning of RNA life (Hayden and Lehman, 2006; Mattick, 2011; Yarus, 2011; Higgs and Wu, 2012; Higgs and Lehman, 2015), i.e. life without unicellular organisms, in abundant ecospheres and additionally in indefinite processes, conserved DNA hereditary pathways may be used successfully to reproduce, so to speak, “selected works of RNA agency.” If we subtracted all these RNA-based processes would cellular life function?

3.3. *RNA world agents: where the genotype is the phenotype*

With the split into a RNA world and a DNA world that codes for proteins with a non-repeat genetic syntax the split between the genotype and the phenotype emerged. The genetic code now also codes for proteins and the protein enzymes form the protein bodies, which we know as cells, cellular parts that organize metabolism, tissues, organs and organisms, all of which we term life from the most primitive last universal common ancestor to human beings. The percentage of coding for such proteins in the human 3 billion base pairs is only 1.5%, however. All the other DNA contains non-coding RNAs (up to 45%) and non-coding areas with unknown

functions (55%) which I suggest represent non-coding RNAs of regulatory functions. This means in all organisms we can find the genotype and the encoded phenotype. The genotype contains the whole blueprint for the phenotype. The interesting thing is that the variety of parts that constitute the phenotype throughout its developmental stages as well as its regulation is coded in the genotype together with epigenetic markings. Epigenetics is an effective tool for expanding the informational content of a given nucleotide sequence syntax with multiple meanings/functions (Slotkin and Martienssen, 2007; Mattick, 2010; Conley and King Jordan, 2012) representing experiences in historical contexts of real life-worlds (Donkin et al., 2016).

We know that epigenetic markings silence or increase the various outcomes of genetic information: a good example is the phenotype of brain organs and their synaptic neuronal networks (Barry and Mattick, 2012). This means a single genotype may have several reading patterns which result in slightly different phenotypes.

In contrast, in the previous pure RNA world the RNA world agents served as a template or active catalyst with several functional domains such as RNA binding, protein-binding, DNA-binding, conformational switch and the important function of scaffolding into a coordinated ribonucleoprotein function (Mercer and Mattick, 2013). In the early RNA world proteins served to stabilize RNA structures only not as building blocks of cellular bodies (Wolf and Koonin, 2007). According the early emergence of the self/non-self differentiation competence in RNA groups we have to think on the evolution of compartmentalization by membranes that serve as essential tools for the emergence of identity (self/non-self). Such membrane encoding genes (*env*-genes) later on served as exapted regulatory elements in host organisms as documented in the various roles of repetitive elements of endogenous retroviruses such as LTRs and non-LTRs (Villarreal, 2005).

The double function of genotype and phenotype is unique in biology, where we can find most organisms divided into genetic coding and protein-based bodies within their real lifeworld interactions. In former times it was thought that the information transfer was a one-way road from genotype to phenotype but with epigenetics and RNA biology it became obvious that the reverse direction of information flow plays a crucial role in evolution, development and adaptational flexibility (Mattick, 2010).

4. Infectious parasite lifestyle results in immune systems

One main assumption of the current investigations of the pre-conditions of life is self-organized replication, which means that a molecular-based entity that is coded in hereditary material can replicate in other entities than itself (Cheng and Unrau, 2010). Without ribozymes that are able to replicate themselves the RNA world would not have prevailed. In this context we can differentiate between a “dead” state controlled by prebiotic chemistry and a “living” state controlled by autocatalytic replication (Larson et al., 2012; Vaidya et al., 2013).

Interestingly, a new perspective has arisen in which passive RNA serves as a template for active RNA that acts as a catalyst with different fitness values. RNA-based parasites that are predominant destroy the system (error threshold) because their error rate is too high to be conserved (leading to a (quasi-)species collapse). In contrast, parasites that are not predominant enrich the system because they constitute nucleotide sequence novelty for adaptive purposes, such as efficient immune response tools in the host. The infectious parasite lifestyle uses and manipulates other agents than itself for its purposes, which may lead to beneficial co-evolutionary results as documented in the several immune systems of cellular organisms (Villarreal, 2009a,b, 2012; Bushman, 2003). Especially this agent-based perspective will replace former narratives such as

“autocatalytic replications” (Kauffman, 2014) that reduced biological agent-behaviour to systems theoretical properties.

Co-operation, Darwinian evolution, behaviour: are these biotic or prebiotic terms? Some authors suggest they are prebiotic (Pross, 2011) and others that they are clearly biotic (Larson et al., 2012). A chimaera seems to be the mixture of both such as “behavioural chemistry.” Yet if behaviour is an integral feature of life and all organisms are built of molecules and atoms every biotic behaviour is intrinsically molecular behaviour as every organism replicates like an ensemble of chemicals (Cello et al., 2002.). However, do such generalizations lead to a convincing description of e.g. human poems by physics, chemistry and mathematics?

In contrast to former concepts with the prebiotic “soup” and a central role of “hot steamers” that serve as candidates for the transition of prebiotic chemistry to life more recently it has been assumed Earth at the origin of biotic behaviour was an ice ball. This makes sense theoretically, because cold water and ice are ideal preconditions for RNA chemistry (Vlassov et al., 2004; Attwater et al., 2010; Yarus, 2012; Lehman, 2013), which facilitates nucleotide pairing and stabilizes RNA structures which otherwise would be destroyed.

5. Crucial steps to life: from abiotic chemistry to biotic code usage

Short RNA nucleotide sequences have an intrinsic ability to fold back on themselves to form the tools essential for life: the RNA hairpin with its typical base-paired stems and the loop which is constituted by a single-stranded RNA nucleotide sequence (Moore, 1999; Hendrix et al., 2005; Gwiazda et al., 2012; Martin et al., 2015) both constituted in a repeat syntax. While the double-stranded stems are not very active in binding to RNAs that are not part of the stem, the loops are open for binding events to other RNA such as loops from both self and non-self RNAs (Smit et al., 2006).

Most interestingly, RNAs that are able to fold back by themselves can co-operate (Muller et al., 2012; Petkovic and Muller, 2013), i.e. build cooperative groups with identity sharing as well as identity precluding abilities – switching back and forth even in short time events with far-reaching consequences: the RNA genotype may behave as a phenotype similar to viroids (Flores et al., 2012, 2014). A coherent genotype is possible because of strict adherence to the rules of molecular syntax, the base-pairing rules, which include that if the rules are not followed base pairing does not happen and loop building with non-base pairing nucleotide sequences occurs. This is at the core of identity building of RNA groups because this mix of two different forms (double-stranded stems and single-stranded loops) of nucleotide molecular syntax in the genotype is the essence of the phenotype that may interact as RNA consortium with other RNA consortia or even with DNA sequences or proteins (Mercer and Mattick, 2013; Behrouzi et al., 2012). Yet the crucial feature is not the sequence syntax but the function of the entire RNA stem loop and its tertiary folding, which means varying syntax combinations may lead to the same function (Doudna et al., 1989; Doudna and Rath, 2002; Atkins et al., 2011).

If the interactions of individual RNA stem loops lead to groups of RNA consortia we can look at the crucial event of their becoming more than chemistry, i.e. real life. Exclusively the group interactions of RNA stem loops make the genotype evolve into semiotic biology, which means the emergence of natural code rules (syntax, pragmatics, semantics). The characters get their language-like feature only through the social interactions of RNAs. Because the binding relates not only to physical-chemical parts but also nucleotides they have an inherent opportunity to serve as semiotic signs, characters for a language/code-like property. This feature cannot be substantiated by quantifiable analyses because it is a socially interacting feature. As mentioned above social interaction patterns of living organisms can be measured by statistical procedures,

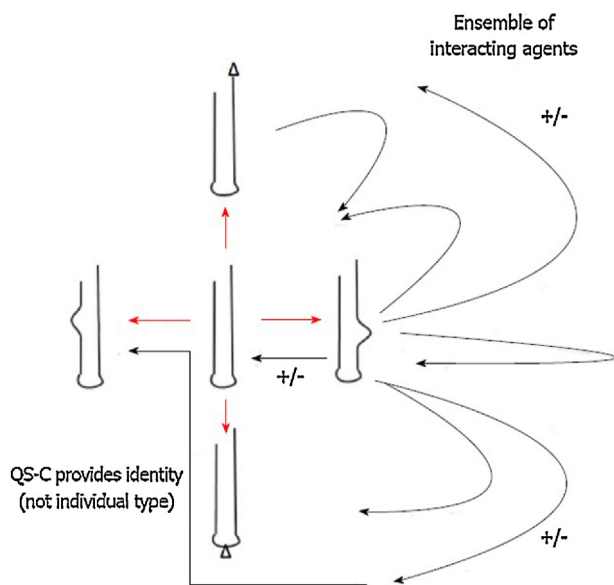


Fig. 1. If base-paired stems of an RNA stem loop build bulges, the whole identity of the stem loop and the group of which this stem loop is part can change rapidly. Neither the fittest types nor selfish agents are successful early replicators but the group which is the result of cooperative RNA stem loops (quasispecies-consortia; QS-C). However, in the early RNA world they changed their identity rather fast and constantly. Through spontaneous single bulge loop productivity the whole identity may change and with it the interactional behaviour against related structures, possible cooperators and enemies that do not fit into the realm as well as against formerly degraded agents (that may serve as appropriate tools for later group ensembles) and against real genetic parasites (from: Villarreal and Witzany, 2013b).

but such methods are less helpful for understanding consortial interacting motifs and techniques because such real life motifs of biological agents are not fully computable by algorithm based procedures (Witzany, 1995, 2010b). No other physico-chemical entities share this property, only nucleic acid molecules. The social interaction of RNAs turns this property into a coding capacity.

In both cases, single RNA stem loops lacking biotic behaviour and RNA groups with clear biotic behaviour by several cooperating RNA stem loops we have the same chemicals, suggesting the step from prebiotic chemistry to real life is a gradual one as suggested by several of the authors cited earlier. Yet the emergence of biotic behavioural interactions is not gradual; it is fundamental because such behavioural patterns are completely absent in abiotic nature. Additionally RNA group behaviour is essential not only for the origin of life but is the key feature of gene regulation and cellular processes. If we stopped this behaviour no life would function as we know it (Fig. 1).

5.1. Interacting RNA groups generate the first natural code

To escape the constant flux of unstable RNA sequences, DNA innovation was the next important step in generating a fully natural code. However, let us recall that the single RNA stem loop does not represent a natural code but solely the physico-chemical assemblies of base-pairing molecules. If we look at the nucleotide sequences of the single RNA stem loop from a purely physico-chemical perspective the genetic code begins as abiotic chemistry. From a biocommunication perspective no co-operativity is found, no biological selection occurred (Villarreal and Witzany, 2013a,b).

Without RNA stem loop group building and group identity, however, the chemicals do not become a natural code with its inherent syntactic, pragmatic and semantic rules but remain on the abiotic stage determined by natural laws of physics and chemistry. Single RNA nucleotides or a single RNA stem loop - although built up of the characters of the nucleic acid language - does not represent a

genetic code, but solely chemical molecules. Although we can identify all characters of the code, the genetic code does not function as a code but solely as chemical molecules.

The genetic code with its typical language-like features (characters adopt syntactic, semantic, pragmatic rules) take the stage of a real natural language with the interactional group building of various RNA stem loops. The interactional and group building cooperativity of the RNA stem loops constitute the genetic code as a real natural code, not its physico-chemical key characters alone. The RNA stem loops are part of an identity which co-operates or precludes and by compartmentalization (within lipid membranes?) gets the same first stages of conservation. Until DNA was invented conservation would have been a rare event instead of a constant evolutionary innovation process of maximum productivity of nucleotide sequence novelties, which in former concepts has been termed error catastrophe.

With the emergence of the first natural language/code, the participating RNA agents get a tool which is open for indefinite generation of new sequences, information processing and storing.

The biotic agents of RNA stem loop groups display their behavioural motifs through the third level of real natural language/code: pragmatics, the context-dependent use of syntax structure that emerges semantic meanings, i.e. concrete functionality, that is not restricted to the syntax structure but determined by the group behavioural pattern in a real-life context with its historically derived background. In the case of pure chemical assemblage of RNA nucleotides without RNA group building and group identity, only the syntax structure can determine meaning, but this is prebiotic, not biotic, not life. Life starts with the real pragmatic use, i.e. the interaction in group building.

This is not the primary goal of interacting RNAs, however. The primary goal is the interaction to become part of cooperative networks that may also be competing and restricting (Popovic et al., 2015), as documented in the predominant ligase function in contrast to the nuclease function (Hayden and Lehman, 2006; Díaz Arenas and Lehman 2010). This means the turn to life is also a behavioural pattern between becoming part of an identity of RNA groups that reject those that do not fit into this identity. Accompanying this is continued productivity in that new sequences are generated which form stems and single-stranded RNA sequences that base pair to others.

Therefore the origin of life is a complementary building up of cooperative behaviour by RNA stem loop groups that may generate and change the identity of RNA groups.

5.2. Essential tools in RNA group building: ligase and endonuclease

The basic RNA stem loop is solely a physical-chemical entity without any biotic behaviour lacking natural selection. But within several others it can use its loop to base pair to non-self agents and form groups (“gangen”). The loop is the biocommunicative interaction tool. A group is built up of different members that represent a group identity (like gangs), as suggested by Villarreal (2014).

One main behavioural motif here is the “kissing loop” motif. The single-stranded loop RNA is open for interactions with other (self or even non-self) single-stranded RNAs via base pairing according to base-pairing syntax rules (Brunel et al., 2002; Cao and Chen, 2011). This may have highly dynamic aspects such as the loop-loop interaction within an RNA stem loop group (as represented by editosomes, spliceosomes, ribosomal subunits and viruses) or between RNA stem loops with non-related RNA loops present within the cell, tissue, organ or organism. Some of them may be remnants of formerly degraded RNAs. RNA recycling is an important resource for building and rebuilding RNA identity

groups, i.e. formerly foreign (non-self) RNA group members may become relevant parts of a new (self) identity.

The essential tool for building groups or dissolving groups into parts is the ligase/endonuclease module which counter-regulates group building in early and current RNA worlds (Ohmori et al., 2011; Stoddard, 2011; Villarreal, 2014). The group buildings initiate a new property of behavioural patterns of RNA groups that does not exist in former pre-group stages of RNA worlds: the self/non-self differentiation behaviour which is the appropriate tool for co-operating with groups that fit into group identity and rejecting RNAs that do not fit into these identities (Battle and Doudna, 2002). Later on many such RNA stem-loop groups were conserved in DNA storage medium, ready to be transcribed from DNA into RNA agents that enacted cellular replication, recombination, repair and immunity (Guilliam et al., 2015).

Other examples include the two halves of tRNAs (Maizels and Weiner, 1993; Maizels et al., 1999), the two important subunits of the ribosome that emerged out of polymerase ribozymes (Harish and Caetano-Anolles, 2012; McGary and Nudler, 2013; Root-Bernstein and Root-Bernstein, 2015; Caetano-Anolles and Sun, 2014). We think also of recombination tools that evolved out of repair abilities, not forgetting the several parts that diversified within cellular life such as group II introns that evolved into LTR regulatory tools (Lambowitz and Zimmerly, 2011; Zimmerly and Semper, 2015) or even into the important spliceosome that functions as a splicezyme (Matlin and Moore, 2007; Amini et al., 2014). A similar cooperative unit is the editosome in RNA editing (Witzany, 2011). Interestingly consortial interacting RNAs outcompete selfish genetic elements which contradicts the selfish gene narrative fundamentally (Vaidya, 2012; Vaidya et al., 2012).

5.3. A new domain of life: natural genetic identity producers

Genetic parasites, parasite replicators, mobile genetic elements, introns, viruses, and several other entities that are involved in competent genetic content arrangements and natural genome editing are known today. Where, how and when did they evolve? Do they share common features inherited from common ancestors? Their role has been investigated within the last three decades and a variety of essential features are common to them, but about their origins there is no consensus.

For decades it has been assumed that viruses escaped from cellular organisms representing genetic parasites. There are several good arguments why viruses evolved before cellular agents (Villarreal, 2005, 2009a,b; Villarreal and Witzany, 2010). Some argue that they co-evolved (Koonin, 2010; Forterre and Prangishvili, 2010; Koonin and Dolja, 2014; Rohwer et al., 2014).

In this review it is argued that the RNA consortia that led to viruses are older than cellular life. Viruses seem to represent the first more complex genetic identities built up by co-operative RNA stem loops. Only in the virosphere we can find both, RNA groups without membranes such as viroids and with membranes, some of them with essential regulatory roles in cellular host organisms such as LTRs/non-LTRs (Ryan, 2009). Only the infectious behaviour is shared by both RNA sequences and viruses. However, is it really infection or is it primarily interaction which may lead to both a predator manipulating prey behaviour and also to co-operative behaviour beneficial for both interaction partners?

This means what we define as a virus may be older than cellular life forms, although then we should not term them viruses at all, but RNA consortia (Villarreal, 2014) later on followed by viruses, viral derived parts or defectives such as SINEs LINEs ALUs, LTRs non-LTRs or more commonly genetic parasites, parasite replicators, mobile genetic elements, introns. All of them are the result of RNA stem loop co-operations, with several conserved genetic identities, all highly able to edit host genomes in various ways, all essential tools

of the different agents and sub-populations (Villarreal and Witzany, 2015).

We may term these agents now as genetic identity producers, to describe their common competence to differentiate self from non-self, to co-operate, combine and recombine, to generate coherent nucleotide sequences de novo, insert or delete them into non-self genomes serve as exaptable tools for host cellular needs (Bushman, 2003; Witzany, 2009). This may lead to the acknowledgement as an essential domain of life affecting all known three domains of life fundamentally (archaea, prokarya, eukarya).

Their identity and specificity changed dramatically with the invention of DNA and protein-based cell morphologies and the success story of co-evolution between RNA stem loop identities (commonly genetic parasites, parasite replicators, mobile genetic elements, introns) and cellular life forms.

5.4. How to do things with sequences

The RNA stem loop groups not only generate and constitute nucleotide sequences that serve as information storage media, but primarily interact (Nicholson and White, 2014). The generation of RNA stem loop consortia is real entities not just genetic syntax, and they are active in contributing to the identity of such groups and rejecting agents that do not fit this identity. In several motifs the interactions depend on the context in which an interaction takes place (Przybilski and Hammann, 2007; Popovic et al., 2015). This may be evolutionary conserved in DNA as demonstrated in the co-operative behaviour of the two ribosomal subunits. The activity is crucial for the result of the co-operation (Doudna and Rath, 2002).

In all these natural genetic identity producers (predator RNAs, infectious RNAs, viruses, genetic parasites and defectives) the common feature is the reverse transcriptase (RT) which copies the RNA nucleotide sequence into a DNA strand. Interestingly, defectives, i.e. infectious agent-derived RNAs that remain as exapted and coopted RNA elements in their host genome (Brosius and Gould, 1993; Cox and Sullivan, 2014) may serve for host gene regulation together with counterparts that regulate the regulatory tool. For example, this happened with the group II introns that remain as splicing only elements in mitochondria and chloroplasts: these parts lost their RT as documented in a variety of group II introns (Zimmerly and Semper, 2015). The abundance of retrovirus-derived LTRs are excellent examples of exapted gene regulation tools as are the Alu sequences of which we find a million copies within the human genome and are recently identified as key player to build up the nucleolus as important place for ribosomal subunit assembly (Carmo-Fonseca, 2015). Another excellent example are the telomere/telomerase tools for the genome maintenance and therefore for conservation of the whole genetic information of the cell which is reduced through the billion replications through ageing (Witzany, 2008). In all these examples genetic sequences of natural genome editors do something, with the result that they act as vital agents within cellular organisms.

6. DNA innovation: unstable RNA productivity and stable DNA storage medium

The innovation of DNA by the reverse transcriptase opened a new window in the evolution of life, i.e. completing unstable productivity with stable conservation. Additionally, the overwhelming number of interactions of RNA groups in the RNA world meant every transfer into DNA served as a kind of immune function; escape out of the continued possibility of information collapse in the RNA world into a world of conserved information. So DNA also became a habitat for RNA groups to hide in for some time or constantly (Witzany, 2009). There are good reasons why DNA arose

together with protein bodies, cells or cellular ensembles limited by a membrane that protected them against invading parasites. This was the next step after proteins served only as stabilization tools for RNA sequence structures.

The evolution of cellular morphology is another secret of life which has not been investigated in this review, but is of importance if we want a satisfactory answer to the question of what life is. One main pathway is the translation process, the transfer of genetically encoded information into protein language, its regulation of all developmental processes (Wolf and Koonin, 2007). The initiation of the translation process is also regulated by long-range RNA–RNA interactions, i.e. the interactions of RNA stem loops within one organism, triggered by internal and external factors, leads to the translation of genetic information into proteins that constitute cellular bodies.

Additionally, the new pathway of evolution, cellular protein-based phenotypes with their own behavioural patterns and success stories, represents a much more comfortable habitat for the RNA world agents, not forgetting the variety of new RNPs and even RNA groups that emerged in a co-evolutionary way with cellular organisms, and expanded the matrix for colonization geometrically (Witzany, 2009, 2015b).

The emergence of DNA and protein-based cellular life created a new biosphere with real cell-based organisms (Takeuchi et al., 2011). Their own identity group building, evolution and development from an RNA-only perspective opened up a completely new dimension of genome-bearing entities as appropriate resources for infectious/cooperative RNA settlements. Although RNAs function at the bottom of all genetic processes that regulate protein-based life, all protein bodies interact independently through an abundance of phenotypic signalling processes (biocommunication) within and between related and non-related organisms. The protein bodies of cellular organisms generate and constitute a planetary lifeworld which by itself functions as the complementary part to feedback information and co-evolution to current RNA world agents (Witzany, 2015b).

7. Conclusion

As long as the DNA storage medium took centre stage in molecular biology, genetics and evolution theory genetic changes that may lead to genetic diversity were explained by replication errors. The abundance of empirical knowledge about non-random genetic content operators together with the rise of RNA Biology and the parallel comeback of virology took RNA at centre stage. RNAs can co-operate, build communities, generate nucleotide sequences de novo and can insert/delete them into host genetic content. In this sense they determine host genetic identities throughout all kingdoms and the planetary virosphere and represents a basic vertical domain of life. The natural genome editing competences of RNA consortia contradict the random-like error replication narrative fundamentally.

This knowledge about the generation of new sequences that are not the result of error replication rejects all attempts by mathematical biology to fully explain the origin of the genetic code and the evolution of life as well as the behavioural motifs of the core agents of life. Although it is possible to measure the physico-chemical properties of nucleotide sequence molecules, cell biology, genetics, computing their statistical features, essential properties such as de novo generation without gene duplication or error replication will have non-predictable, incomputable features in principle. This is the reason why mathematically and computation derived explanations of the origin of life remain unsatisfactory. In parallel no creationist explanatory model is needed to explain and understand the empirical data that represent the interactional motifs in

RNA group building. Both narratives of this traditional controversy are outdated now and a third way to explain biological evolution seems to have more explanatory power than the previous ones.

The emergence of the genetic code is not the emergence of the nucleotides which constitute the code. The origin of the genetic code as a natural language occurs exclusively within co-operative interacting RNAs. So we can define life as an emergent property of social interacting RNAs. From this perspective life is coherent with the emergence of communication, i.e. interactions based on a natural code that underlies syntactic, pragmatic and semantic rules.

References

- Alliegro, M.C., 2011. The centrosome and spindle as a ribonucleoprotein complex. *Chromosome Res.* 19, 367–376.
- Amini, Z.N., Olson, K.E., Muller, U.F., 2014. Spliceozymes: ribozymes that remove introns from pre-mRNAs in trans. *PLOS ONE* 9 (7), e101932, <http://dx.doi.org/10.1371/journal.pone.0101932>.
- Atkins, J.F., Gesteland, R.F., Cech, T.R. (Eds.), 2011. *RNA Worlds. From Life's Origin to Diversity in Gene Regulation*. Cold Spring Harbor Laboratory Press, New York.
- Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A., Holliger, P., 2010. Ice as a protocellular medium for RNA replication. *Nat. Commun.* 1, 76.
- Barry, G., Mattick, J.S., 2012. The role of regulatory RNA in cognitive evolution. *Trends Cogn. Sci.* 16, 497–503.
- Battle, D.J., Doudna, J.A., 2002. Specificity of RNA–RNA helix recognition. *PNAS* 99, 11676–11681.
- Behrouzi, R., Roh, J.R., Kilburn, D., Briber, R.M., Woodson, S.A., 2012. Cooperative tertiary interaction network guides RNA folding. *Cell* 149, 348–357.
- Brosius, J., Gould, S.J., 1993. Molecular constructivity. *Nature* 365, 102.
- Brunel, C., Marquet, R., Romby, P., Ehresmann, C., 2002. RNA loop–loop interactions as dynamic functional motifs. *Biochimie* 84, 925–944.
- Bushman, F.D., 2003. Integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115, 135–138.
- Caetano-Anolles, G., Sun, F.J., 2014. The natural history of transfer RNA and its interactions with the ribosome. *Front. Genet.* 5, 1–5.
- Cao, S., Chen, S.J., 2011. Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA* 17, 2130–2143.
- Carmo-Fonseca, M., 2015. Assembly of the nucleolus: in need of revision. *EMBO J.* 34, 2731–2732.
- Cello, J., Paul, A.V., Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* 297, 1016–1018.
- Cheng, L.K.L., Unrau, P.J., 2010. Closing the circle: replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* 2 (10), a002204.
- Conley, A.B., King Jordan, I., 2012. Endogenous retroviruses and the epigenome. In: Witzany, G. (Ed.), *Viruses: Essential Agents of Life*. Springer, Dordrecht, pp. 309–323.
- Cox, J.E., Sullivan, C.S., 2014. Balance and stealth: the role of noncoding RNAs in the regulation of virus gene expression. *Annu. Rev. Virol.* 1, 89–109.
- Delihias, N., 2011. Impact of small repeat sequences on bacterial genome evolution. *Genome Biol. Evol.* 3, 959–973.
- Deng, Z., Wang, Z., Lieberman, P.M., 2012. Telomeres and viruses: common themes of genome maintenance. *Front. Oncol.* 2, 201.
- Díaz Arenas, C., Lehman, N., 2010. Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube. *BMC Evol. Biol.* 10, 80.
- Donkin, I., Versteijhe, S., Ingerslev, L.R., Qian, K., Mechta, M., et al., 2016. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metabolism* 23, 1–10.
- Doudna, J.A., Cormack, B.P., Szostak, J.W., 1989. RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron. *Proc. Natl. Acad. Sci. U. S. A.* 86, 7402–7406.
- Doudna, J.A., Rath, V.L., 2002. Structure and function of the eukaryotic ribosome: the next frontier. *Cell* 109, 153–156.
- Eigen, M., Schuster, P., 1977. A principle of natural self-organization. *Naturwissenschaften* 64, 541–565.
- Eigen, M., Schuster, P., 1978. The hypercycle. A principle of natural self-organization. Part B: the abstract hypercycle. *Naturwissenschaften* 65, 7–41.
- Eigen, M., Schuster, P., 1982. Stages of emerging life—five principles of early organization. *J. Mol. Evol.* 19, 47–61.
- Eigen, M., Winkler, R., 1981. *The Laws of the Game: How the Principles of Nature Govern Chance*. Princeton University Press, Princeton.
- Flores, R., Gago-Zachert, S., Serra, P., Sanjuan, R., Elena, S.F., 2014. Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* 68, 395–414.
- Flores, R., Serra, P., Minoia, S., Di Serio, F., Navarro, B., 2012. Viroids: from genotype to phenotype—still relying on RNA sequence and structural motifs. *Front. Microbiol.* 3, 217.
- Forsdyke, D.R., Mortimer, J.R., 2000. Chargaff's legacy. *Gene* 261, 127–137.
- Forster, P., Prangishvili, D., 2010. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. In: Witzany, G. (Ed.), *Natural Genetic Engineering and Natural Genome Editing*. John Wiley & Sons, Hoboken, pp. 65–77.

- Goldenfeld, N., Woese, C., 2011. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.* 2, 375–399.
- Guilliam, T.A., Keen, B.A., Brissett, N.C., Doherty, A.J., 2015. Primase–polymerases are a functionally diverse superfamily of replication and repair enzymes. *Nucleic Acids Res.* 43, 6651–6664.
- Gwiazda, S., Salomon, K., Appel, B., Muller, S., 2012. RNA self-ligation: from oligonucleotides to full length ribozymes. *Biochimie* 94, 1457–1463.
- Harish, A., Caetano-Anolles, G., 2012. Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* 7, e32776.
- Hayden, E.J., Lehman, N., 2006. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* 13, 909–918.
- Hendrix, D.K., Brenner, S.E., Holbrook, S.R., 2005. RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.* 38, 221–243.
- Higgs, P.G., Lehman, N., 2015. The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* 16, 7–17.
- Higgs, P.G., Wu, M., 2012. The importance of stochastic transitions for the origin of life. *Orig. Life Evol. Biosph.* 42, 453–457.
- Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V., 2007. Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8, 241–259.
- Kauffman, S.A., 2014. Prolegomenon to patterns in evolution. *Biosystems* 123, 3–8.
- Koonin, E.V., 2010. On the origin of cells and viruses: primordial virus world scenario. In: Witzany, G. (Ed.), *Natural Genetic Engineering and Natural Genome Editing*. John Wiley & Sons, Hoboken, pp. 47–64.
- Koonin, E.V., 2011. The Logic of Chance: The Nature and Origin of Biological Evolution. FT Press, Upper Saddle River.
- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* 78, 278–303.
- Lambowitz, A.M., Zimmerly, S., 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003616.
- Larson, B.C., Jensen, R.P., Lehman, N., 2012. The chemical origin of behavior is rooted in abiogenesis. *Life* 2, 313–322.
- Lehman, N., 2013. Cold-hearted RNA heats up life. *Nat. Chem.* 5, 987–989.
- Lehman, N., 2015. The RNA world: 4,000,000,050 years old. *Life* 5, 1583–1586.
- Maizels, A., Weiner, A.M., 1993. The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), *The RNA World*, 3rd ed. Cold Spring Harbor Laboratory Press, New York, pp. 577–602.
- Maizels, N., Weiner, A.M., Yue, D., et al., 1999. New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tRNA substrates. *Biol. Bull.* 196, 331–334.
- Mantegna, R.N., Buldyrev, S.V., Goldberger, A.L., Havlin, S., Peng, C.K., Simons, M., Stanley, H.E., 1994. Linguistic features of noncoding DNA sequences. *Phys. Rev. Lett.* 73, 3169–3172.
- Marraffini, L.A., Sontheimer, E.J., 2010. Self versus non-self discrimination during CRIPR RNA-directed immunity. *Nature* 463, 568–571.
- Martin, L.L., Unrau, P.J., Muller, U.F., 2015. RNA synthesis by in vitro selected ribozymes for recreating an RNA world. *Life* 5, 247–268.
- Matlin, A.J., Moore, M.J., 2007. Spliceosome assembly and composition. In: Blencowe, B.J., Graveley, B.R. (Eds.), *Alternative Splicing in the Postgenomic Era*. Landes BioScience and Springer Science Business Media, New York, pp. 14–35.
- Mattick, J., 2009. Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann. N. Y. Acad. Sci.* 1178, 29–46.
- Mattick, J.S., 2010. RNA as a substrate for epigenome-environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning and cognition. *Bioessays* 32, 548–552.
- Mattick, J.S., 2011. The double life of RNA. *Biochimie* 93 (November (11)), viii–ix, [http://dx.doi.org/10.1016/S0300-9084\(11\)00355-5](http://dx.doi.org/10.1016/S0300-9084(11)00355-5).
- Mayr, E., 2004. *What Makes Biology Unique?* Cambridge University Press, New York.
- McDonald, S.M., 2013. RNA synthetic mechanisms employed by diverse families of RNA viruses. *WIREs RNA* 4, 351–367.
- McGary, K., Nudler, E., 2013. RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* 16, 112–117.
- Mercer, T.R., Mattick, J.S., 2013. Structure and function of long non-coding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300–307.
- Minoia, S., Navarro, B., Delgado, S., Di Serio, F., Flores, R., 2015. Viroid RNA turnover: characterization of the subgenomic RNAs of potato spindle tuber viroid accumulating in infected tissues provides insights into decay pathways operating in vivo. *Nucleic Acids Res.* 43, 2313–2325.
- Moelling, K.I., Broecker, F., 2015. The reverse transcriptase-RNase H: from viruses to antiviral defense. In: Witzany, G. (Ed.), *DNA Habitats and Their RNA Inhabitants*. Wiley & Sons, Hoboken, pp. 126–135.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182.
- Moore, P.B., 1999. The RNA folding problem. In: *The RNA World*, 2nd ed. Cold Spring Harbor Laboratory Press, New York.
- Muller, S., Appel, B., Krellenberg, T., Petkovic, S., 2012. The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA. *IUBMB Life* 64, 36–47.
- Nicholson, B.L., White, K.A., 2014. Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* 12, 493–504.
- Nghe, P., Hordijk, W., Kauffman, S.A., Walker, S.I., Schmidt, F.J., Kemble, H., Yeates, J.A., Lehman, N., 2015. Prebiotic network evolution: six key parameters. *Mol. Biosyst.* 1, 3206–3217.
- Ohmori, R., Saito, H., Ikawa, Y., Fujita, Y., Inoue, T.T., 2011. Self-replication reactions dependent on tertiary interaction motifs in an RNA ligase ribozyme. *J. Mol. Evol.* 73, 221e229.
- Petkovic, S., Muller, S., 2013. RNA self-processing: formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett.* 587, 2435–2440.
- Popovic, M., Fliss, P.S., Ditzler, M.A., 2015. In vitro evolution of distinct self-cleaving ribozymes in diverse environments. *Nucleic Acids Res.* 43 (14), 7070–7082.
- Pross, A., 2011. Toward a general theory of evolution: extending Darwinian theory to inanimate matter. *J. Syst. Chem.* 2, 1.
- Przybilski, D.M., Hammann, C., 2007. The tolerance to exchanges of the Watson–Crick base pair in the hammerhead ribozyme core is determined by surrounding elements. *RNA* 13, 1625–1630.
- Qureshi, I.A., Mehler, M.F., 2012. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.* 13, 528–541.
- Rohwer, F., Youle, M., Maughan, H., Hisakawa, N., 2014. *Life in Our Phage World*. Wholon, San Diego, USA.
- Root-Bernstein, M., Root-Bernstein, R., 2015. The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* 367, 130–158.
- Ryan, F., 2009. *Virology*. Harper Collins, London, UK.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., Pandolfi, P.P., 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358.
- Schuster, P., 2011. Mathematical modeling of evolution. Solved and open problems. *Theory Biosci.* 130, 71–89.
- Sciamanna, I., Vitullo, P., Curatolo, A., Spadafora, C., 2009. Retrotransposons, reverse transcriptase and the genesis of new genetic information. *Gene* 448, 180–186.
- Shapiro, J.A., 1992. Natural genetic engineering in evolution. *Genetica* 86, 99–111.
- Shapiro, J.A., 2005. A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* 345, 91–100.
- Shapiro, J.A., 2009. Revisiting the central dogma in the 21st century. *Ann. N. Y. Acad. Sci.* 1178, 6–28.
- Shapiro, J.A., 2011. *Evolution: A View from the 21st Century*. FT Press, Washington, DC.
- Shapiro, J.A., 2016. The basic concept of the read-write genome: mini-review on cell-mediated DNA modification. *BioSystems*, <http://dx.doi.org/10.1016/j.biosystems.2015.11.003> (in press).
- Shapiro, J.A., von Sternberg, R., 2005. Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250.
- Shay, J.A., Huynh, C., Higgs, P.G., 2015. The origin and spread of a cooperative replicase in a prebiotic chemical system. *J. Theor. Biol.* 364, 249–259.
- Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.
- Smit, S., Yarus, M., Knight, R., 2006. Natural selection is not required to explain universal compositional patterns in rRNA secondary structure categories. *RNA* 12, 1–14.
- Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* 2, 16.
- Stoddard, B.L., 2011. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19, 7–15.
- Szathmari, E., Maynard Smith, J., 1997. From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* 187, 555–571.
- Takeuchi, N., Hogeweg, P., Koonin, E.V., 2011. On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS Comput. Biol.* 7, e1002024.
- Vaidya, N., (PhD dissertation) 2012. *Spontaneous Cooperative Assembly of Replicative Catalytic RNA Systems*. Portland State University, Portland.
- Vaidya, N., Manapat, M.L., Chen, I.A., et al., 2012. Spontaneous network formation among cooperative RNA replicators. *Nature* 491, 72–77.
- Vaidya, N., Walker, S.I., Lehman, N., 2013. Recycling of informational units leads to selection of replicators in a prebiotic soup. *Chem. Biol.* 20, 241–252.
- Villarreal, L.P., 2005. *Viruses and the Evolution of Life*. ASM Press, Washington.
- Villarreal, L.P., 2009a. *Origin of Group Identity: Viruses, Addiction and Cooperation*. Springer, New York.
- Villarreal, L.P., 2009b. The source of self: genetic parasites and the origin of adaptive immunity. *Ann. N. Y. Acad. Sci.* 1178, 194–232.
- Villarreal, L.P., 2011a. Viruses and host evolution: virus-mediated self identity. In: Lopez-Larrea, C. (Ed.), *Self and Non-Self*. Landes Bioscience and Springer Science Business Media, Austin, pp. 185–217.
- Villarreal, L.P., 2011b. Viral ancestors of antiviral systems. *Viruses* 3, 1933–1958.
- Villarreal, L.P., 2012. The addition module as a social force. In: Witzany, G. (Ed.), *Viruses: Essential Agents of Life*. Springer, Dordrecht, pp. 107–145.
- Villarreal, L.P., 2014. Force for ancient and recent life: viral and stem-loop RNA consortia promote life. *Ann. N. Y. Acad. Sci.* 1341, 25–34.
- Villarreal, L.P., Witzany, G., 2010. Viruses are essential agents within the stem and roots of the tree of life. *J. Theor. Biol.* 262, 698–710.
- Villarreal, L.P., Witzany, G., 2013a. The DNA habitat and its RNA inhabitants: at the dawn of RNA sociology. *Genomics Insights* 6, 1–12.
- Villarreal, L.P., Witzany, G., 2013b. Rethinking quasispecies theory: from fittest type to cooperative consortia. *World J. Biol. Chem.* 4, 79–90.
- Villarreal, L.P., Witzany, G., 2015. When competing viruses unify: evolution, conservation, and plasticity of genetic identities. *J. Mol. Evol.* 80, 305–318.

- Vlassov, A.V., Johnston, B.H., Landweber, L.F., Kazakov, S.A., 2004. Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world. *Nucleic Acids Res.* 32, 2966–2974.
- von Sternberg, R., Shapiro, J.A., 2005. How repeated retroelements format genome function. *Cytogenet. Genome Res.* 110, 108–116.
- Wächtershäuser, G., 2014. The place of RNA in the origin and early evolution of the genetic machinery. *Life* 4, 1050–1091.
- Wang, X., Song, X., Glass, C.K., Rosenfeld, M.G., 2011. The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. *Cold Spring Harb. Perspect. Biol.* 3, a003756.
- Witzany, G., 1995. From the “logic of the molecular syntax” to molecular pragmatism. Explanatory deficits in Manfred Eigen’s concept of language and communication. *Evol. Cogn.* 1, 148–168.
- Witzany, G., 2005. Natural history of life: history of communication logics and dynamics. *SEED J.* 5, 27–55.
- Witzany, G., 2008. The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance. *Biosemiotics* 1, 191–206.
- Witzany, G., 2009. Noncoding RNAs: persistent viral agents as modular tools for cellular needs. *Ann. N. Y. Acad. Sci.* 1178, 244–267.
- Witzany, G. (Ed.), 2010a. *Natural Genetic Engineering and Natural Genome Editing*. John Wiley & Sons, Hoboken.
- Witzany, G., 2010b. *Biocommunication and Natural Genome Editing*. Springer, Dordrecht.
- Witzany, G., 2011. The agents of natural genome editing. *J. Mol. Cell Biol.* 3, 181–189.
- Witzany, G., 2012. From molecular entities to competent agents: viral infection derived consortia act as natural genetic engineers. In: Witzany, G. (Ed.), *Viruses: Essential Agents of Life*. Springer, Dordrecht, pp. 407–419.
- Witzany, G., 2014a. Language and communication as universal requirements for life. In: Kolb, V. (Ed.), *Astrobiology. An Evolutionary Approach*. CRC Press, Boca Raton.
- Witzany, G., 2014b. Pragmatic turn in biology: from biological molecules to genetic content operators. *World J. Biol. Chem.* 5, 279–285.
- Witzany, G. (Ed.), 2015a. *DNA Habitats and Their RNA Inhabitants*. Wiley & Sons, Hoboken.
- Witzany, G., 2015b. Life is physics and chemistry and communication. In: Witzany, G. (Ed.), *DNA Habitats and Their RNA Inhabitants*. Wiley & Sons, Hoboken, pp. 1–9.
- Witzany, G., Baluska, F., 2012. Life’s code script does not code itself. The machine metaphor for living organisms is outdated. *EMBO Rep.* 13, 1054–1056.
- Wolf, Y.I., Koonin, E.V., 2007. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol. Direct.* 2, 14.
- Yarus, M., 2011. *Life from an RNA World: The Ancestor Within*. Harvard University Press, Harvard.
- Yarus, M., 2012. Darwinian behavior in a cold, sporadically fed pool of ribonucleotides. *Astrobiology* 12, 870–883.
- Zimmerly, S., Semper, C., 2015. Evolution of group II introns. *Mob. DNA* 6, 7, <http://dx.doi.org/10.1186/s13100-015-0037-5>.