

Opinion

A Peptide–Nucleic Acid Replicator Origin for Life

Bernard M.A.G. Piette^{1,3,*} and Jonathan G. Heddle^{2,3,*,@}

Evolution requires self-replication. But, what was the very first self-replicator directly ancestral to all life? The currently favoured RNA World theory assigns this role to RNA alone but suffers from a number of seemingly intractable problems. Instead, we suggest that the self-replicator consisted of both peptides and nucleic acid strands. Such a nucleopeptide replicator is more feasible both in the light of the replication machinery currently found in cells and the complexity of the evolutionary path required to reach them. Recent theoretical and mathematical work supports this idea and provide a blueprint for future investigations.

The Ancestral Replicator

Life as we understand it is cellular. The last universal common ancestor (LUCA) of all cells (not a single cell of course but a population) is understood in some detail; it possessed a cell membrane, DNA, the basic molecular machines for copying DNA (i.e., polymerase etc.), and a functional ribosome, among many more [1]. From this highly truncated list alone it is clear that LUCA was far too complex to spontaneously assemble. It must have evolved from simpler systems, themselves able to self-replicate with some tolerance for error (otherwise they would not be able to evolve). Indeed, it is difficult to imagine that anything recognisably a cell could have spontaneous origins. This means that they in turn must have evolved from even simpler self replicators; that is, molecular self-replicators. The first such replicator is referred to as the initial Darwinian ancestor (IDA) [2].

The identity of the IDA has been cause for much speculation over the years. Nonbiological replicators such as clay crystals [3] have been invoked but are unconvincing as they require a complete takeover of one substrate with another and lack a persuasive argument to show how this could have occurred. An IDA built from biological molecules is more convincing and necessitates physicochemical conditions compatible with their formation, and with the self-replicating reaction cycle of the IDA itself. The latter likely required relatively mild conditions approaching those of analogous biochemical reactions today. There is now ample evidence that biological building blocks such as amino acids, ribose, and deoxyribose, among others, were present on the early Earth [4–7]. Potential chemistries for building block synthesis have been demonstrated, with cyanosulfidic chemistry [8] showing great promise. Furthermore, recent work has shown convincing peptide ligation in prebiotic conditions [9]. Given these findings, it now seems reasonable to assume that the IDA was constructed of components highly similar or identical to those found in life today. Indeed, such a replicator must have occurred at some stage very early in evolution even if not at the very beginning. The scene being set for an IDA constructed of biological molecules to arise, our focus turns to the main issue of this work – deciding on its identity.

Three approaches will be useful to us in this endeavour. (i) Consideration of current self-replicating biological systems. By looking at how cells currently achieve self-replication we may be able to extrapolate into the past and deduce the composition of the start point. (ii) Consideration of the simplest system. Which IDA is most likely able to achieve self-replication and has the simplest

Highlights

The first ancestor to all life was a self-replicating entity capable of evolving but must have been much simpler than a cell; that is, a molecular replicator.

Recent work on RNA polymerase ribozymes supports the view that the molecular replicator was made of RNA, and no other molecule was required (the RNA World).

The RNA World is challenged by the fact that a self-replicating RNA polymerase ribozyme has yet to be demonstrated and that little evidence for its existence is seen in life today as well as problems in how the transition from RNA-only to RNA–protein world could have occurred.

A molecular replicator with two components –RNA and peptide –overcomes these problems and may be a better fit.

¹Department for Mathematical Sciences, Durham University, Durham, UK

²Bionanoscience and Biochemistry Laboratory, Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland

³These authors contributed equally to this work.

*Correspondence: jonathan.heddle@uj.edu.pl (J.G. Heddle).
 @Twitter: @HeddleLab (J.G. Heddle).



path to arrive at current self-replicating systems? (iii) Dynamical feasibility. It is not always obvious that a given self-replication scheme is actually dynamically feasible. For example, expected rates of substrate breakdown will determine if the mathematical model allows self-sustained self-replication. Where substrates break down more quickly than they can be synthesised, the mathematical model will rightly predict self-sustained self-replication to be unfeasible. A dynamically unfeasible scheme need not be considered further.

A number of suggestions for the IDA have been made over the years. These include protein or peptide alone, such as thiol-rich peptides; **amyloid** (see [Glossary](#)) inspired in part by our understanding of **prions** [10,11]; nucleic acid alone (mainly the RNA World [12–14]); and a mixture of both [15–18]. The most widely accepted is the RNA World which posits that the IDA was an RNA strand capable of folding into an active replicase. This is the main alternative to a **nucleopeptide replicator** considered in this work. The RNA World is immediately attractive when considering our three tests above. (i) RNA is intricately involved in current self-replicating systems ([Figure 1](#)). It carries the transcribed genetic information (and there is general agreement that the original substrate for the genetic code was RNA, not DNA) and it constitutes the majority of the ribosome, the heart of cellular self-replication, including the catalytic site itself [19]. In addition, cells do contain RNA sequences (ribozymes) that are able to catalyse reactions; a function more typically associated with proteinaceous enzymes. (ii) Considerations of simplicity make the RNA World attractive because it is able to house both encoding of sequence information and catalytic activity in a single molecule in a way which proteins (which are poor at information storage) and DNA (poor at catalysis) are not. (iii) In principle, single-component self-replicating systems (RNA or otherwise) are not dynamically or chemically forbidden and indeed rudimentary systems have been demonstrated on the macroscopic scale [20]. However, a closer look at the RNA-only IDA in the context of these three considerations reveals serious problems, some of which appear insurmountable. Comparison with a nucleopeptide IDA suggest the latter is a more likely candidate.

Current Self-Replicators as a Guide to the Identity of the IDA

The concept of using current cells and molecules as a guide to predict the features of earlier (now extinct) ones and indeed, even resurrect extinct proteins is well established [4]. Bioinformatics techniques [4] allow phylogenetic trees to be reconstructed [21] and predictions to be made of the identity of ancient ancestor molecules. The extreme chronological distance between the IDA and the current day would make this challenging but perhaps possible if a simple self-replicator was present in any cells today. In fact, no such single molecule replicator exists. Instead, replication is split between nucleic acids and proteins with the general rule being that nucleic acids encode and transfer information (DNA and mRNA), and proteins (enzymes) carry out catalytic functions including synthesising new copies of the nucleic acids ([Figure 1](#)). A notable exception being the ribosome, the catalytic centre of which is a ribozyme [22]. This means that a basic cross-catalytic symmetry is observed: RNA makes protein, protein makes RNA. Thus, in the spirit of Spiegelman's Monster ([Box 1](#)), we can conceive that given the correct conditions, a supply of energy and suitable chemical building blocks, a self-replicating system using components from current cells could function and would include DNA, DNA polymerase, ribosome, RNA polymerase, tRNA and **tRNA synthetases** among others. Indeed, recent progress has been made towards this in experiments using liposome-based synthetic cells. These contained DNA replication machinery from $\Phi 29$ and were capable of self-sustained DNA amplification [23]. However, such systems would not be indefinitely self-sustaining due to the eventual degradation of the components responsible for synthesising the protein machinery. Unsurprisingly, these preliminary functional self-replicating systems retain the nucleic acid–protein division of labour.

Glossary

Amyloid: protein aggregates made from multiple repeating proteins, typically forming β sheets.

Aptamers: nucleic acid or peptide molecules selected for specific binding to molecular targets of interest.

Entropy trap: A structure that is able to catalyse a chemical reaction between two components simply by constraining their spatial arrangement.

Horizontal gene transfer: transfer of genes between individuals not having a parent–offspring relationship.

Hypercycle: theoretical construct consisting of a cyclic, autocatalytic set of self-replicating molecules.

Ligase: enzyme that catalyses covalent bond formation between two molecules.

Nucleopeptide replicator: a self-replicating system that consists of both nucleic acid and peptide where each component can catalyse replication of the other.

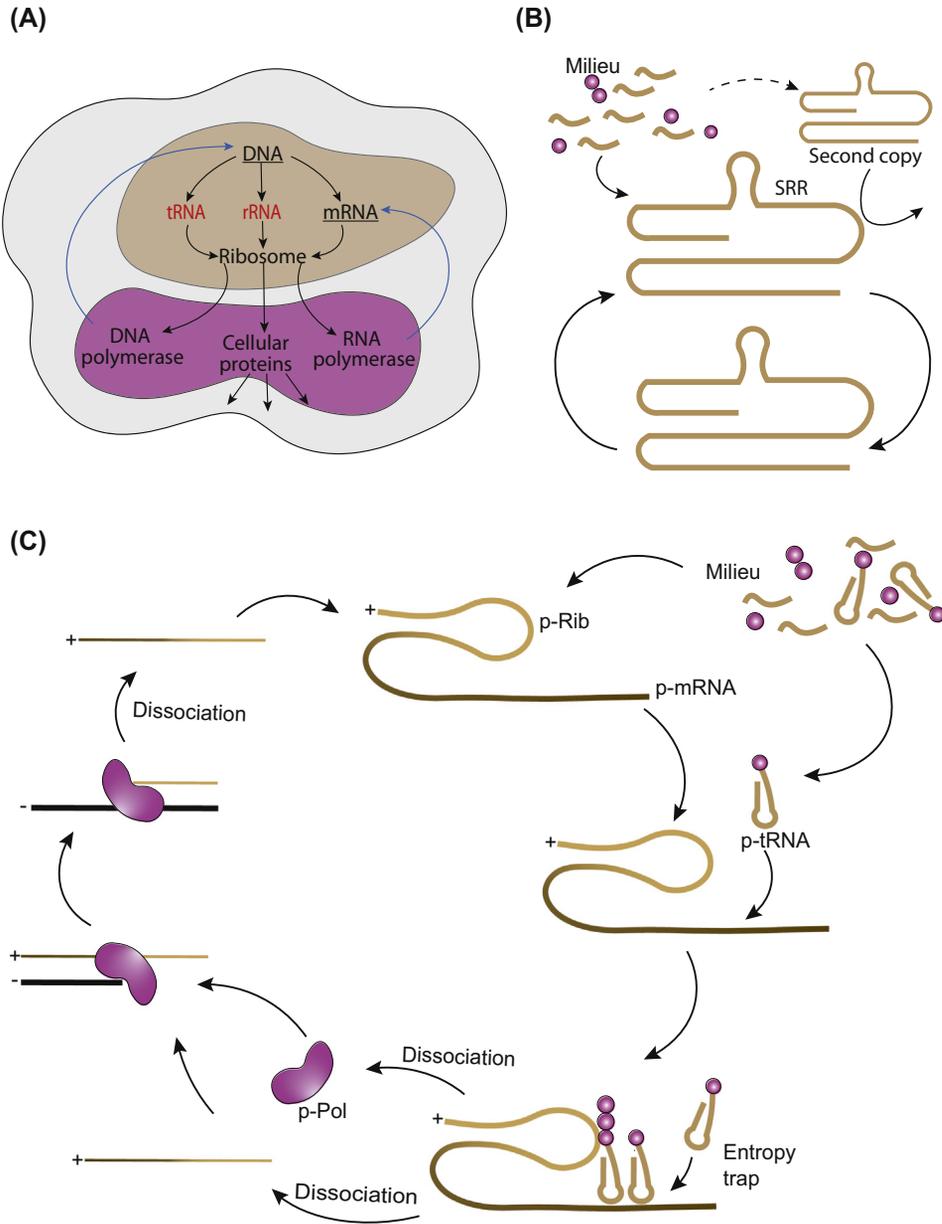
Prion: infectious amyloid proteins.

RNase P: ribozyme whose substrate is a tRNA precursor, which it cleaves.

tRNA synthetases: enzymes that catalyse ester bond formation between amino acids and their cognate tRNAs.

Turing instabilities: mechanism resulting from the combination of diffusion and some nonlinear reaction between reactants (chemicals or cells) leading to the formation of spatial patterns of varying concentration of the reactants.

Vertical gene transfer: transfer of genes from parent to offspring.



Trends in Ecology & Evolution

Figure 1. The Nucleopeptide Initial Darwinian Ancestor (IDA). (A) Minimal self-replication in extant life occurs at the cellular level and requires nucleic acids (beige) and proteins (purple). Underlined molecules act as stores of sequence information. Molecules acting as or carrying molecular building blocks are shown in red. Black arrows indicate flow of information, materials. Blue arrows indicate replication by polymerases. (B) A typical RNA World scenario begins with a milieu containing various components including short RNA strands (beige) and amino acids (purple). RNA strands spontaneously assemble into a longer strand with self-replicating ribozyme (SRR) functionality. The process repeats, forming a second, functionally identical molecule, able to copy the first (and vice versa). (C) A possible nucleopeptide IDA. The milieu is the same with additionally, short RNA strands bound to amino acids. An RNA strand is able to form a structure with primordial mRNA (p-mRNA) and primordial ribosome (p-Rib) functionality (+ strand). Some of the amino acid-bound RNA strands contain sequences complementary to sequences on the p-mRNA and so are able to act as primordial tRNA (p-tRNA). Alignment of such charged p-tRNA on the p-Rib/p-mRNA in an entropy trap results in formation of a peptide chain. For some sequences, dissociation of the components results in release of a peptide with primordial polymerase (p-Pol) activity and the + strand. The p-Pol acts on the + strand producing a complementary copy (- strand, black). The - strand is able to act as a template for the polymerase, producing the + strand. Dissociation releases all components and the + strand is again able to act as a p-Rib.

Box 1. Spiegelman's Monster and the Rise of *In Vitro* Evolution

Q β bacteriophage is an RNA virus whose genome is replicated by Q β replicase, an RNA-dependent RNA polymerase. In 1965, Spiegelman carried out an interesting experiment – he mixed together Q β replicase and the phage RNA together with RNA nucleotides. As a result he was able to observe *in vitro* replication of the genome [64] – a breakthrough at the time. Furthermore, the system was able to evolve [65], freed from constraints and requirements of functioning in cells and encoding other virus proteins. Over the course of 75 serial transfers, the RNA evolved to become shorter, eventually reducing in length to 550 nt from an original of 3600 nt. Subsequent experiments were able to isolate a 218 nt RNA; essentially the minimum required to function. The system did not encode or synthesize the Q β replicase, which was provided. Nevertheless, it showed the potential for simple, self-replicating systems to function and, importantly, to evolve outside of the cellular environment. Crucially, it showed that this could be achieved with a few constituents, comprising only building blocks, an RNA strand, and a protein. The spirit of Spiegelman's experiment lives on today as efforts continue to produce fully self-contained *in vitro* self-replicating systems [66].

As a thought experiment, we can simplify further the extant nucleopeptide replicator system so that only indispensable components remain and assume that instead of multiple examples of each class there was one able to carry out all the functions currently undertaken by that class. That is, there existed (i) a single nucleic acid sequence that encoded a single amino acid sequence and acted as its own mRNA and its own ribosome; and (ii) a peptide sequence that acted to copy the mRNA that encoded the amino acid sequence of the same peptide (Figure 1). This is the most fundamental ancestral self-replicator that maintains the existing functional split between nucleic acids and peptides.

What advantage does such a system have over the main alternative, the RNA World? Perhaps the most convincing answer comes when considering how they could have evolved into the current replication machinery. For the nucleopeptide IDA the path is straightforward. Each component simply maintains the same role, evolving greater specificity and efficiency over time, something that can be achieved by errors that would inevitably arise during replication. The part with mRNA functionality expands its code to encode for an increased number of amino acids. This leads to production of more efficient polymerases, thus ensuring better survival. The part with ribosome functionality separates from the mRNA portion and gains increasing catalytic efficiency, again increasing survival likelihood.

For an RNA-only replicator, the story is less convincing. A self-replicating RNA molecule seems at first glance feasible, and if such an RNA-only IDA existed, it is likely that it would be a self-replicating ribozyme polymerase (Figure 1) [24]. However, this is not what is seen in extant cells and it is difficult to understand the steps by which it gained mRNA functionality, and transferred polymerase functionality to peptides and information storage functionality to DNA. Of these, the first two present most difficulty as the ribozyme, being a self-sufficient self-replicating entity simply has no need for interaction with peptides or amino acids. This makes a gradual transfer of these functions to them difficult to envision. How could gain of mRNA functionality; that is, encoding of amino acid sequences in RNA occur in a ribozyme whose sequence is optimised for ribozyme self-copying functions? Any change to encode a helpful peptide would presumably decrease ribozyme effectiveness. It seems unlikely that by chance the sequence giving useful ribozyme functionality would also happen to encode a functional peptide that would increase the replication and survival of the ribozyme itself. If this did take place then it would consequently have required the ribozymal polymerase sequence to serve also as mRNA; encoding an amino acid sequence that could be translated into a functional peptide. We then come close to the nucleopeptide replicator outlined in Figure 1; the difference being that the system in Figure 1 has the advantage of not requiring the mRNA to also have polymerase functionality. Similar arguments pertain to transfer of polymerase functionality from RNA to peptides. Recent work shows that this is highly unlikely, suggesting such an RNA functionality never existed [25,26].

The Simplest Self-Replicating System

By asking which possible IDA is simplest we are considering which functional replicator is simple enough to be realistically feasible as the first, spontaneously occurring IDA. The simplest system does not have to be the real one but arguments for parsimony in biology are powerful [27].

The simplest RNA World theory requires only a self-replicating ribozyme. This could be an RNA strand with **ligase** activity; that is, self-templating using pre-existing large fragments of complementary sequences. Ligase (but not self-replicating ligase) ribozymes do exist in nature [28] and *in vitro* designed/evolved ribozyme ligases have been produced, beginning with the work of Bartel and Szostak [29]. Efforts have been made to produce minimal ligases, for example by Kurihara *et al.* [30] when they made an ~50 nt functional version of R3C ligase [31], similar in length to the small L1 ligase [32]. These ligases, however, do not self-replicate. Efforts to produce self-replicating ligases have borne fruit. Paul and Joyce, for example, modified R3C ligase ribozyme so that it could template two half copies of itself that it then ligated [33]. Difficulties arose because of substrate inhibition, overcome by a cross-catalytic approach [34,35], whereby two template strands each catalyse ligation of two halves of the other template strand. However, these systems still required at least one of the strands to be 50 nt or greater in length and in the first round require spontaneous assembly of the full length ribozyme (typically well over 100 nt). This may seem unlikely, but in fact, recent work suggests that functional ribozyme ligases can be produced spontaneously (i.e., nonenzymatically) from short building blocks more likely to have been present on the early Earth [36,37]. However, again, these ligases do not replicate themselves. This may be an insurmountable problem, as Wachowius *et al.* stated: 'Fundamentally, emergence of new functions when assembling long sequences is confounded by the nature of such activities: ligases use less information to choose substrates than is required to define the ligase activity itself, so cannot copy themselves (or other components) from sequences lacking that information, i.e. random sequence' [7].

A ribozyme acting as a polymerase therefore seems more promising. This could copy any template strand from only short nucleotide building blocks (Figure 1). The first designed ribozyme able to convincingly do this was R18 [38,39]. The tC9Y ribozyme made a breakthrough, being able to polymerise products slightly longer than itself [40]. The most recent advance is the 24-3 ribozyme that can copy RNA sequences having secondary structure, although this is still possible only for short sequences; that is, they cannot copy themselves. These recent ribozymes, at close to 200 nt in length, are likely too big to spontaneously arise, although other recent work [41] suggests that, in some cases, ribozymes may have been able to function as fragments working together. In summary, if a self-replicating RNA-only ribozyme polymerase does prove possible, it may be that it is too long and complex to have arisen as the IDA.

An alternative to an RNA-only IDA is that it consisted of both nucleic acid and peptide components, each able to catalyse polymerisation of the other. An early example of this idea was the Carter and Kraut model [18], which proposed that a short double-stranded (ds)RNA sequence would be able to catalyse formation of a short β hairpin structure that in turn would catalyse dsRNA polymerisation. With a potential for partial coding [42], sustained self-replication could be possible.

Our favoured nucleopeptide IDA model is a conceptual relative of the Carter and Kraut model (Figure 1). [43] It relies only on random production of short amino acids and RNAs (or possibly a mix of RNA, DNA, and other monomers since lost, known as XNA). Here, a short stretch of single-stranded (ss)RNA could act as both a primordial RNA (p-RNA) and a primordial ribosome (p-Rib), encoding a peptide sequence and catalysing its polymerisation. If the resulting

peptide was able to act as a primordial polymerase (p-Pol) and copy the ssRNA then an IDA would result. The first and most fundamental problem with this concept is how specific amino acids (or classes of amino acids) can be encoded and located at specific mRNA sequences (the coding problem; [Box 2](#)). This is troublesome as tRNA and certainly tRNA synthetases did not then exist. The simplest answer to this problem is hard stereochemical selection, where there is a direct interaction between amino acids and their codons or anticodons [44,45]. Such an interaction could bring specific amino acids into close proximity on the p-mRNA, increasing the probability of polymerisation via peptide bond formation. This is a so called **entropy trap** and has been suggested in other models of early biological replicators wherein RNA-based carriers of amino acids align on mRNA sequences via codon–anticodon interactions [46]. The release of the produced peptide could occur stochastically or through periodic environmental changes (e.g., changes in temperature). There is some experimental support for entropy traps [47,48] and indeed they may even play a role in peptide bond formation in the extant ribosome [49].

The key assumption of hard stereochemical theories – that codons can bind to the amino acids they encode – has been given some impetus in recent years such as by the demonstration that amino acid-binding **aptamers** can be evolved *in vitro* [50]. In addition, analysis of ribosome structures has shown that some amino acids are located proximal to their codons or anticodons with statistical significance [51]; perhaps reflecting the remnants of an ancient stereochemical interaction. Recent computational analyses have lent some support to this idea but doubts remain as to whether such apparently weak interactions (and the fact that they are for peptides rather than single amino acids) are robust enough to carry out the proposed function [52]. It also makes tRNA unnecessary, requiring it to be invented at a later date.

A more appealing, soft stereochemical solution was recently elaborated in which a particular amino acid is linked to its anticodon not directly but via binding to specific sequences on primordial tRNA (p-tRNA), which are distal from the anticodon [43,53] ([Figure 1](#)). Such a p-tRNA combined with the dual function p-Rib/p-mRNA and the encoded p-Pol would comprise a functional IDA with a clear path connecting it to present day replication in cells [43].

Box 2. How Did the Link between Codons and Amino Acids Arise?

In all life, triplet codons in mRNA code for specific amino acids that are brought to the ribosome as activated amino acids attached to a tRNA that contains the relevant, specific anticodon. That the correct amino acid is connected to the correct tRNA is ensured by enzymes called aminoacyl tRNA synthetases (aaRSs). These attach the amino acids to the 3' end of the tRNA distal from the anticodon. This is a form of symbolic coding; that is, there is no direct interaction between the anticodon and the amino acid. However, when the earliest IDA arose, there would have been no aaRS, so how was the link between the code and the amino acid made? Without an aaRS it is reasonable to think that the connection between tRNA and amino acid could have originally been chemically different, and not necessarily an ester bond. An obvious first answer is that the codon (or anticodon) binds specifically to the amino acid for which it codes via a direct physicochemical interaction, thus doing away with the need for an intermediary such as aaRS. This is known as the stereochemical hypothesis, first put forward several decades ago and which remains one of the main theories for how the genetic code first came about [52]. However there has been little compelling evidence to support this theory and it also raises practical questions (e.g., if the amino acid is bound to the anticodon how does the anticodon recognise the codon?).

One possible answer is that there was a physicochemical interaction between p-tRNA and amino acids but that the sequence of the interacting RNA was not the anticodon sequence and was located distal from the anticodon (as in current RNAs). In this scenario the code came about by simple chance and was 'frozen' in place, in line with the concept of the 'frozen accident' origin of the genetic code first put forward by Crick [67]. This could have been achieved if a particular combination of p-tRNA and amino acid allowed a self-replication system to come into being that included a polymerase able to replicate the p-tRNAs themselves. Such an idea was recently discussed [53]. One challenge to confirming this hypothesis is that once such a system was replaced by aaRS charging of tRNA, then the original amino acid recognition sequence may well have degraded to the point of being untraceable.

To qualify as a nucleopeptide IDA, both the peptide and the nucleic acid components are obligatory, but it does not necessarily follow that their contributions are equal. The weakest version of the hypothesis would consist of a polymerase in which the main catalytic role is carried out by RNA (a ribozyme) but where a contribution from a peptide is required to achieve replication 'escape velocity'; for example, by increasing the stability of the RNA. Such helper roles for peptides are evident in **RNase P**, in which the peptide component is required for activity in certain conditions [54]. Specifically, it was shown that a peptide consisting of 10 lysines could allow a polymerase ribozyme to function at low magnesium ion concentrations [55]. This is particularly relevant, as at high magnesium concentrations, RNA hydrolysis is accelerated, making a self-sustaining self-replication reaction kinetically less favourable. A nucleopeptide–RNA World hypothesis requiring only helper peptide functionality could just as easily be called a weak RNA World hypothesis. By contrast, the strongest nucleopeptide IDA hypothesis posits that the peptide produced is solely or largely responsible for the polymerase catalytic activity. This idea is attractive in that no change of polymerase activity from RNA to protein is required over the course of evolution. In such a system, RNA produces protein which produces RNA (Figure 1) and the cycle is complete, but is it dynamically viable?

Dynamical Feasibility of a Nucleopeptide IDA

Over 40 years ago, Eigen and Schuster [56] established a mathematical model of generic Darwinian evolution. Introducing the concept of **hypercycles**, they showed how the best genotypes can be progressively selected via a natural process of random evolution. Their model assumes an already existing machinery of genetic replication and does not explain how the DNA–polypeptide duplication system was generated in the first place. Yet, the fundamental mathematical properties of hypercycles are similar to models describing the occurrence of the RNA/polypeptide replication mechanism. More recently Bagley *et al.* [57] did study mathematically the evolution of metabolism, showing formally how polymers can be dynamically selected as a result of belonging to some auto-catalytic pathway.

Natural selection is usually defined [58] as resulting from variation among members of a population, differential reproduction, and heritability of traits important for survival. At the molecular level though, the population is made out of a large number of possible chains and the selection occurs when a subset of these chains is more stable but also duplicated in larger proportions. This was also argued by Wills and Carter [26] in a different context.

Recently, Arnoldt *et al.* [59] have presented a model to describe the switching between **horizontal gene transfer** (HGT) and Darwinian **vertical gene transfer**. While the system they describe already assumes a protein–DNA translation mechanism, the essence of the model can be applied to the system of interest to us here. A pool of random RNA and polypeptide chains that progressively favour a population that is more stable and self-replicating.

As outlined above, we can assume that before life appeared on Earth, there were mixtures of p-tRNA and amino acids that slowly, but spontaneously, polymerised to form short polypeptide and RNA chains. As argued by Banwell and colleagues [43], one of the polypeptide chains could have been a polymerase able to copy RNA chains. At the same time, each RNA chain was a template for a polypeptide chain generated from charged p-tRNA assembling along the RNA backbone. The end results were two families of polymers of various lengths that mutually facilitated their polymerisation.

By itself, this process would generate a large set of polymers, but if the polymerase and its RNA counterpart were such that their mutual cross-polymerisation rates were larger than for the other

Box 3. Turing Instabilities and Self-Replication

The first mathematical model of self-replication was proposed by John von Newman [61], who identified the need for a template as well as a machine that reads it to make a copy of the same machine. This captures the essentials of how ribosomes translate mRNA to proteins. What this does not answer though, is how the very first biological self-replicating machine came about.

A model that we favour is that short polymer chains such as polypeptide and polynucleotides form spontaneously at a relatively low rate, with all possible sequences being generated. We can imagine that some sequences will be favoured by dynamical mechanisms.

How this could work can perhaps be best understood by considering a well know mechanism for dynamical selection, namely Turing instabilities [62], which can explain, for example, the zebra's stripes. Imagine if in the zebra embryo, two molecules, an activator and an inhibitor are present. These are able to interact such that the activator promotes itself as well as the inhibitor while the inhibitor inhibits its own production and reduces the activator's activity. Coupled with diffusion, this would lead to a striped pattern in the concentration of the two molecules and consequently, to the stripes on the animal.

For the self-replication mechanism of the IDA, we observe that the mathematical equations describing the polymerisation of chains of different length are identical to the ones describing diffusion in Turing's model, except that instead of describing the chemical density at different locations the equations describe the concentrations of the chains of different sequences and length. The dynamical properties of peptide and nucleotide chains depend on their length and their actual sequence. Only certain sequences may allow useful properties such as catalysis. This means that the configuration space is not homogeneous and Turing instabilities would thus be expected to occur in some specific place of the configuration. Hence selecting families of peptide and nucleotide chains which mutually enhance their relative concentrations. This is a potential mechanism to explain how some simple polymers were dynamically selected for self-replication.

chains, their concentration would increase faster [43]. As the polymers are all depolymerising, the p-tRNA and amino acids would slowly be converted dominantly to the polymerase and its RNA counterpart, leading to a natural selection process for these two types of polymers. A simple, but realistic dynamical model was presented by Banwell and colleagues [43], where it was shown that given the best known estimates of all reaction rates, a polymerase can be dynamically selected from a pool of t-RNA and peptides.

Any realistic model needs to take into account the fact that shorter chains polymerise to longer chains and longer chains depolymerise to shorter ones. This cross-dependence combined with the dependence of the polymerase concentration and the finite amount of p-tRNA and amino acids lead to a nonlinear system of equations [43].

The equations describing the transformation between polymers of similar sequence and length are identical to the diffusion equation and we postulate that combined with the reactions between RNA and peptide chains this leads to **Turing instabilities** [59] and hence the selection process. Our argument is outlined in **Box 3** and is supported by the simulations performed by Füchslin *et al.* [60].

Concluding Remarks and Future Perspectives: Building a Nucleopeptide IDA

We have argued that the IDA was a mixture of peptides and RNA(-like) chains. A number of questions remain to be answered before this can be stated with greater certainty (see Outstanding Questions). The most important missing piece of information is a convincing demonstration of a p-tRNA showing hard or soft stereochemical selectivity. For example, by binding a cognate amino acid at a binding pocket distal from the anticodon. The second important demonstration will be that a short peptide is able to act as an RNA-dependent RNA polymerase. Deciding on the sequence of such a peptide will be challenging. Computational analyses of existing polymerases may yield simple core polymerases and, given the possible ancient origin of viral RNA-dependent RNA polymerases [63], they may offer a useful start point. This will be difficult given its likely low catalytic rate, but highly sensitive labelling techniques should allow detection of low product yields. Finally, it will be necessary to show that the simple RNA strand can act

Outstanding Questions

Can an RNA polymerase ribozyme be demonstrated? If so, it would support the RNA World theory over competing theories.

Conversely, can a short peptide having RNA-dependent RNA polymerase activity be demonstrated? Such a demonstration would lend support to the nucleopeptide IDA proposal.

Is there any remnant of the primitive RNA-dependent RNA polymerase in extant cells and if so is it possible to reverse engineer a primitive RNA-dependent RNA polymerase?

Can a convincing stereochemical interaction between a coding sequence and an encoded amino acid be demonstrated? Whether directly or via a tRNA like intermediate, this would 'close the loop' between the RNA code and the peptide copier of the code.

Can one derive a reaction diffusion model describing the interaction between polypeptides and RNA chains? If so, does it exhibit Turing instabilities with higher concentration of selected polypeptides and RNA chains?

as a mRNA/ribosome. This will require the charged p-tRNAs to bind codons on simple RNA strands and form peptide bonds driven by spatial localisation. If the peptide thus produced is the RNA dependent RNA polymerase sequence, then an IDA would result.

Acknowledgments

J.G.H. was funded by National Science Centre (NCN, Poland) grant no. 2016/21 / B / CC1 / 00274 (OPUS-11). B.M.A.G.P would like to thank the Bionoscience and Biochemistry Laboratory, Malopolska Centre of Biotechnology, Jagiellonian University, Krakow for its hospitality.

References

- Weiss, M.C. *et al.* (2016) The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1, 16116
- Yarus, M. (2011) Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harb. Perspect. Biol.* 3, a003590
- Cairns-Smith, A.G. (1982) *Genetic Takeover: And the Mineral Origins of Life*, Cambridge University Press
- Zaucha, J. and Hedde, J.G. (2017) Resurrecting the dead (molecules). *Comput. Struct. Biotechnol. J.* 15, 351–358
- Nam, I. *et al.* (2018) Abiotic synthesis of purine and pyrimidine ribonucleosides in aqueous microdroplets. *Proc. Natl. Acad. Sci. U. S. A.* 115, 36–40
- Bada, J.L. (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.* 42, 2186–2196
- Wachowius, F. *et al.* (2017) Nucleic acids: function and potential for abiogenesis. *Q. Rev. Biophys.* 50, e4
- Sutherland, J.D. (2016) The origin of life – out of the blue. *Angew. Chem. Int. Ed.* 55, 104–121
- Canavelli, P. *et al.* (2019) Peptide ligation by chemoselective aminonitrile coupling in water. *Nature* 571, 546–549
- Wiltzius, J.J. *et al.* (2009) Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* 16, 973
- Wickner, R.B. *et al.* (2014) Amyloid diseases of yeast: prions are proteins acting as genes. *Essays Biochem.* 56, 193–205
- Gilbert, W. (1986) The RNA world. *Nature* 319, 618
- Higgs, P.G. and Lehman, N. (2015) The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* 16, 7–17
- Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol.* 39, 99–123
- Fishkis, M. (2011) Emergence of self-reproduction in cooperative chemical evolution of prebiological molecules. *Orig. Life Evol. Biosph.* 41, 261–275
- Di Giulio, M. (1997) On the RNA world: evidence in favor of an early ribonucleopeptide world. *J. Mol. Evol.* 45, 571–578
- Carter, C. (2015) What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. *Life* 5, 294–320
- Carter Jr., C.W. and Kraut, J. (1974) A proposed model for interaction of polypeptides with RNA. *Proc. Natl. Acad. Sci. U. S. A.* 71, 283–287
- Nissen, P. *et al.* (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930
- Griffith, S. *et al.* (2005) Robotics: self-replication from random parts. *Nature* 437, 636
- Yang, Z. and Rannala, B. (2012) Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13, 303–314
- Moore, P.B. and Steitz, T.A. (2003) After the ribosome structures: how does peptidyl transferase work? *RNA* 9, 155–159
- van Nies, P. *et al.* (2018) Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat. Commun.* 9, 1583
- Horning, D.P. and Joyce, G.F. (2016) Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* 113, 9786–9791
- Carter Jr., C.W. and Wills, P.R. (2017) Interdependence, reflexivity, fidelity, impedance matching, and the evolution of genetic coding. *Mol. Biol. Evol.* 35, 269–286
- Wills, P.R. and Carter Jr., C.W. (2018) Insuperable problems of the genetic code initially emerging in an RNA world. *Biosystems* 164, 155–166
- Crisci, J.V. (1982) Parsimony in evolutionary theory: law or methodological prescription? *J. Theor. Biol.* 97, 35–41
- Vicens, Q. and Cech, T.R. (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nat. Chem. Biol.* 5, 97–99
- Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* 261, 1411–1418
- Kurihara, E. *et al.* (2014) Development of a functionally minimized mutant of the R3C ligase ribozyme offers insight into the plausibility of the RNA World Hypothesis. *Biology* 3, 452–465
- Rogers, J. and Joyce, G.F. (1999) A ribozyme that lacks cytidine. *Nature* 402, 323–325
- Robertson, M.P. *et al.* (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA* 7, 513–523
- Paul, N. and Joyce, G.F. (2002) A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12733–12740
- Lincoln, T.A. and Joyce, G.F. (2009) Self-sustained replication of an RNA enzyme. *Science* 323, 1229–1232
- Robertson, M.P. and Joyce, G.F. (2014) Highly efficient self-replicating RNA enzymes. *Chem. Biol.* 21, 238–245
- Zhou, L. *et al.* (2019) Assembly of a functional ribozyme from short oligomers by enhanced non-enzymatic ligation. *bioRxiv* 700229
- Wachowius, F. and Holliger, P. (2019) Non-enzymatic assembly of a minimized RNA polymerase ribozyme. *ChemSystemsChem* 1, 12–15
- Johnston, W.K. *et al.* (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292, 1319–1325
- Wochner, A. *et al.* (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332, 209–212
- Attwater, J. *et al.* (2013) In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* 5, 1011–1018
- Akoope, A. and Müller, U.F. (2016) Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Phys. Chem. Chem. Phys.* 18, 20118–20125
- Carter, C.W.J. (1975) Cradles for molecular evolution. *New Sci.* 27, 784–787
- Banwell, E.F. *et al.* (2018) Reciprocal nucleopeptides as the ancestral Darwinian self-replicator. *Mol. Biol. Evol.* 35, 404–416
- Gamow, G. (1954) Possible relation between deoxyribonucleic acid and protein structures. *Nature* 173, 318
- Woese, C.R. *et al.* (1966) On the fundamental nature and evolution of the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* 31, 723–736
- Carter Jr., C.W. and Wills, P.R. (2018) Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases. *Nucleic Acids Res.* 46, 9667–9683
- Tamura, K. and Schimmel, P. (2003) Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8666–8669
- Jacobsen, J.R. and Schultz, P.G. (1994) Antibody catalysis of peptide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5888–5892
- Sievers, A. *et al.* (2004) The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7897–7901

50. Yarus, M. (2017) The genetic code and RNA-amino acid affinities. *Life* 7, 13
51. Johnson, D.B. and Wang, L. (2010) Imprints of the genetic code in the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8298–8303
52. Koonin, E.V. and Novozhilov, A.S. (2017) Origin and evolution of the universal genetic code. *Annu. Rev. Genet.* 51, 45–62
53. Koonin, E.V. (2017) Frozen accident pushing 50: stereochemistry, expansion, and chance in the evolution of the genetic code. *Life* 7, 22
54. Guerrier-Takada, C. et al. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857
55. Tagami, S. et al. (2017) Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nat. Chem.* 9, 325–332
56. Eigen, M. and Schuster, P. (1977) A principle of natural self-organization. *Naturwissenschaften* 64, 541–565
57. Bagley, R.J. et al. (1992) Evolution of a metabolism. In *Artificial life II*, pp. 141–158, Addison-Wesley
58. Johnson, B.R. and Lam, S.K. (2010) Self-organization, natural selection, and evolution: Cellular hardware and genetic software. *Bioscience* 60, 879–885
59. Arnoldt, H. et al. (2015) Toward the Darwinian transition: switching between distributed and speciated states in a simple model of early life. *Phys. Rev. E* 92, 052909
60. Fuchsliin, R.M. and McCaskill, J.S. (2001) Evolutionary self-organization of cell-free genetic coding. *Proc. Natl. Acad. Sci. U. S. A.* 98, 9185–9190
61. von Newman, J. (1945) A model of general economic equilibrium. *Rev. Econ. Stud.* 13, C1
62. Turing, A. (1952) The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B* 237, 37–72
63. de Farias, S.T. et al. (2017) Origin and evolution of RNA-dependent RNA polymerase. *Front. Genet.* 8, 125
64. Haruna, I. and Spiegelman, S. (1965) Autocatalytic synthesis of a viral RNA *in vitro*. *Science* 150, 884–886
65. Mills, D.R. et al. (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U. S. A.* 58, 217
66. Li, J. et al. (2017) Cogenerating synthetic parts toward a self-replicating system. *ACS Synth. Biol.* 6, 1327–1336
67. Crick, F.H. (1968) The origin of the genetic code. *J. Mol. Biol.* 38, 367–379