European Review, Vol. 25, No. 2, 231–245 © 2016 Academia Europæa
doi:10.1017/S1062798716000570

Understanding Life: A Bioinformatics Perspective

NATALIA SZOSTAK[1,3], SZYMON WASIK[1,2,3] and JACEK BLAZEWICZ[1,2,3]

[1]Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland.
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznan, Poland. E-mail: swasik@cs.put.poznan.pl
[3]European Centre for Bioinformatics and Genomics, Piotrowo 2, 60-965 Poznan, Poland.

According to some hypotheses, from a statistical perspective the origin of life seems to be a highly improbable event. Although there is no rigid definition of life itself, life as it is, is a fact. One of the most recognized hypotheses for the origins of life is the RNA world hypothesis. Laboratory experiments have been conducted to prove some assumptions of the RNA world hypothesis. However, despite some success in the 'wet-lab', we are still far from a complete explanation. Bioinformatics, supported by biomathematics, appears to provide the perfect tools to model and test various scenarios of the origins of life where wet-lab experiments cannot reflect the true complexity of the problem. Bioinformatics simulations of early pre-living systems may give us clues to the mechanisms of evolution. Whether or not this approach succeeds is still an open question. However, it seems likely that linking efforts and knowledge from the various fields of science into a holistic bioinformatics perspective offers the opportunity to come one step closer to a solution to the question of the origin of life, which is one of the greatest mysteries of humankind. This paper illustrates some recent advancements in this area and points out possible directions for further research.

Introduction

Research to date has shown that our understanding of the origins of life should not only be the domain of biology and biochemistry, but should also include other very diverse fields of science. The scientific fields that can significantly impact our understanding of the beginning of life include mathematics, quantum physics, computer science and bioinformatics. Biology and chemistry require the patient and careful design of experiments, which are very sensitive to the quantity of molecules and environmental conditions. However, detailed knowledge of the geophysical

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

conditions on prebiotic Earth can only be estimated from the limited evidence available. This limited evidence is why, even if laboratory experiments can provide detailed answers, they may not represent an adequate explanation of the processes that underpin the origins of life. Moreover, they often cannot provide a broad enough treatment of the problem, as it is difficult to conduct multi-step and complex empirical studies because of an overwhelming number of variables. This is where mathematics helps because it enables the building and analysis of general models of the phenomena occurring on prebiotic Earth based on uncertain biological and geophysical knowledge. Computer simulations extend the mathematical models and make it possible to verify a wide range of proposed scenarios. Even if neither mathematics nor computer science can distinguish whether something is biologically possible, they can still provide critical insight into the preliminary verification of hypotheses related to the origins of life, because they enable the exploration of various configurations of the uncertain conditions that were prevalent on prebiotic Earth. Due to these advantages, many experiments have been conducted that tried to verify *in silico* hypotheses related to the origins of life.[1–5] Altogether, the wide spectrum of knowledge available from different fields of research can help to explain how simple inorganic compounds could have evolved into life.

It is assumed that life started on Earth 3.5 billion years ago, which is more than 10 billion years after the formation of the Universe, and around 1 billion years after the formation of Earth. Evolution began at this time. Primordial life had to face hard environmental conditions, such as an atmosphere without oxygen and the lack of an ozone layer. Another 1 billion years had to pass before photosynthesis evolved, which led to the Great Oxidation Event when oxygen entered the Earth's atmosphere. This gave a great boost to evolution and cells with nuclei evolved 2 billion years ago. The major kingdoms of life on Earth were established under these conditions, giving rise to the enormous diversity of life forms existing today. Of these, *Homo sapiens* is probably the only one to ask questions about the beginnings of the Universe, life and the process of evolution.

What is Life?

The definition of life is a far-reaching issue that affects not only biology and biochemistry, but also has an impact on searching for life in the Universe. The answer to the fundamental question, 'what is life?' has turned out to be one of the hardest to deliver.[6] Consequently, many definitions have been proposed that can be summarized by 'what you see depends on where you stand'. Below, we present some of these definitions.

The best-known definitions from a theoretical physics perspective come from Erwin Schrödinger and the Polish philosopher and cosmologist, Michał Heller (these and some other definitions were cited during a symposium of the Polish Academy of Science[7]). The first stated that living systems self-assemble against nature's tendency toward disorder, or entropy. Similarly, Heller said that life processes low-entropy solar energy by changing it into order and releasing disorder (heat) into space. On the

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

other hand, biologists and biochemists pay more attention to the internal processes of cells. For example, Andrzej Legocki claims that living systems are characterized by self-processed metabolism (acquiring energy) and autoreplication ability, and Włodzimierz Sedlak stated that life is a compound set of chemical reactions and electron processes in a semiconductive environment of proteins. From a molecular biology perspective, Jan Barciszewski very briefly defines life as a minimal set of important genes. One of the best-recognized definitions from the evolutionist perspective is a National Aeronautics and Space Administration of United States working definition that states that life is a self-sustaining chemical system capable of Darwinian evolution. Evolution is underlined by Jan Kozłowski, who stated that complex biochemical structures may be considered alive if they are capable of evolution via natural selection. Finally, from a computer science perspective, life can be defined as a system carrying and processing information that is capable of replicating itself without the help of other systems that do not belong to the same type.

The multitude of definitions of life presented above clearly shows that we are still lacking a general theory of living systems. The definition of life is different depending on the field of science that utilizes it. Therefore, wide-ranging multidisciplinary research will be required before a single, commonly accepted definition can be formulated. However, regardless of the definition that we choose, we can try to find processes for the beginning of life that has evolved to the form that we observe nowadays.

Probabilistic and Deterministic Hypotheses

In one of his seminal papers, 'A Mathematical Theory of Communication', the American mathematician, electronic engineer and cryptographer Claude Shannon linked the concepts of information and uncertainty (measured by probability).[8] Shannon noted that the amount of information conveyed (and the amount of uncertainty reduced) in a series of symbols is inversely proportional to the probability of occurrence of a particular event or symbol. For example, the predicted outcome of rolling a six-sided die is more improbable than the outcome of flipping a coin; therefore, it conveys more information. Shannon's theory also implies that information increases as a sequence of characters grows but it cannot distinguish functional or message-bearing sequences from random or useless ones. The same finding applies to the biological sequences of molecules. The longer the sequence and the less probable occurrence of each of its elements, the more information can be stored in biomolecules. However, it should be kept in mind that despite carrying quantitative information, which was considered by Shannon, biomolecules should also carry qualitative information (functional ones, called specificity). Therefore, their analysis requires much more complicated steps than only considering the ordering of symbols in macromolecules.

From the perspective of this probability, we can formulate two different hypotheses of the beginning of life: the random origin (chance alone) hypothesis and the

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

origin driven by laws of nature (deterministic) hypothesis. The hypothesis of the random origin of life assumes that some random processes have led to the beginning of life. Therefore, we can calculate the probability of the formation of life and verify whether the calculated probability is reasonable according to our knowledge of biochemistry and physics.

According to the work of Axe we can assume that the shortest functional protein is built of 150 amino acids.[9] Meyer calculated the probability of forming such a protein in completely random processes.[10] He claims that if we consider that functioning proteins tolerate only left-handed (L-form) amino acids, and given that the production of the right-handed (D-form) and left-handed isomers were roughly of equal frequency in abiotic conditions, the probability of randomly attaining a protein consisting only of 150 L-form amino acids is:

$$p_{\text{L-form}} = 2^{-150} \approx 10^{-45}$$

Moreover, he found that the probability of creation of proper peptide bonds between any two amino acids is approximately equal to half, so the chance of connecting all of them in the chain of 150 molecules is equal to:

$$p_{\text{pept}} = 2^{-149} \approx 10^{-45}$$

Thus, starting from mixtures of D-forms and L-forms, the probability of building a 150-amino-acid chain at random in which all bonds are peptide bonds and all amino acids are L-form is:

$$p_{\text{150aa chain}} = p_{\text{L-form}} \times p_{\text{pept}} \approx 10^{-45} \times 10^{-45} \approx 10^{-90}$$

assuming that all events are independent.

This probability describes only the complexity of a molecule in Shannon's sense. This chance is even lower if a particular variation of the 150-amino-acid chain that performs a specific function is required. Taking into consideration that there are 20 amino acids, its probability is equal to:

$$p_{\text{particular aa comb.}} \approx 20^{-150} \approx 10^{-195}$$

In fact, because some mutations in amino acid chains do not change the protein function, the cassette mutagenesis technique showed that the probability of producing any functional 150-amino-acid chain is:[9,11]

$$p_{\text{particular aa comb. by cassette mut.}} \approx 10^{-74}$$

Summing up, the probability that a 150-amino-acid compound assembled by random interactions in a prebiotic soup would be a functional protein (event X) can be estimated as:

$$p_{\text{event } X} = p_{\text{150aa chain}} \times p_{\text{particular aa comb. by cassette mut.}} \approx 10^{-164}$$

again assuming that all events are independent.

To answer whether this is a probable event or not, we have to consider whether there are enough probabilistic resources in the Universe (Ω). We can use the universal probability bound (UPB), which is a method of objectively measuring the plausibility

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

of any probabilistic hypothesis.[12–14] The UPB addresses the question of how many chances particles have had to interact since the Big Bang. The value of Ω is calculated by taking into account three factors:

(1) The number of seconds that have elapsed since the Big Bang, which, based on the 14 billion-year-old Universe, is assumed to be 10^{17} seconds.
(2) The number of possible state changes of particles per second which, based on the minimum amount of time required for light to traverse the Planck length, called Planck time, is assumed to be the inverse of Planck time which is 10^{43}.[10,12–14]
(3) An estimation of the number of elementary particles in the observable Universe, which is 10^{80}.[15]

Summing up, with the assumption that all events are independent, we have:

$$\Omega = 10^{80} \times 10^{43} \times 10^{17} = 10^{140}$$

which is the number of possible observable events in the Universe. This is an equivalent to the measure of the probabilistic resources that have ever existed in the entire observable Universe. This value is much too small compared with the computed chance 1 out of 10^{164} of obtaining the functional protein of 150 amino acids. Such calculations have persuaded many scientists to recognize the random origin hypothesis as being very improbable.

Next to the random origin hypothesis, there are deterministic hypotheses. They assume that processes that lead to the beginning of life could be driven by laws of nature: either by a biochemical predestination leading to self-organization,[16] or by external factors that could also lead to self-organization. In physics, self-organization refers to a spontaneous increase of the order of the system due to some natural processes, forces or laws, e.g. a vortex forms in a bathtub as the water swirls down the drain. Self-organization theories of the origins of life try to attribute the origins of life to physical or chemical forces or processes (laws of nature). In this case, simple monomers (amino acids, bases and sugars) arose from simpler atmospheric gases and energy. Polymers (proteins, DNA and RNA) arose from monomers and then primitive membranes formed around these polymers. Finally, primitive metabolism emerged inside these membranes as various polymers interacted chemically with one another. All these processes were driven by forces of chemical necessity, which were also responsible for the origins of life.

Prigogine and Nicolis turned their attention to the role of an external energy flowing into primitive living systems that may have played a role in the origins of biological organization.[17] However, their theory explains only the order of molecules (in the sense of symmetrical or repeating patterns), but cannot explain the information coding found in DNA or RNA. This approach fails to explain the issue because there is no example in nature that shows that information can be created *ab initio*.

Alternatively, Kauffman proposes a new model, which assumes that the emergence of metabolism preceded coding of information.[18] Kauffman's model states that a self-reproducing metabolic system might emerge directly from a set of

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

'low-specificity' catalytic peptides and RNA molecules in a prebiotic soup, which is a set of catalytic polymers in which no single molecule reproduces itself, but the system as a whole does. However, this model also fails to explain the origins of the specified complexity of biological molecules.

Necessity alone fails to explain the origins of life, but what about chance and necessity together? This is the most probable scenario because physical and chemical forces can enhance the probability of origins of life.

RNA World Hypothesis

One of the most recognized hypotheses of the origins of life is the RNA world theory that assumes that the first forms of life were made of RNA and their synthesis was catalysed by RNA polymerase, which was also an RNA molecule.[19] Therefore, the RNA world hypothesis is based on the assumption that RNA molecules were capable of catalysis and replication. This concept was independently proposed by a few researchers in the 1960s based on the fact that RNA can form complex secondary structures,[20–23] which suggests that RNA can potentially exhibit catalytic properties.[24] Nevertheless, the term 'RNA world' was first used in 1986 by the Chemistry Nobel Prize winner Walter Gilbert, in a commentary on how recent observations of the catalytic properties of various forms of RNA fit with this hypothesis. The RNA world theory is a combination of the random origin hypothesis and the origin driven by laws of nature hypothesis. Chance would have played a greater role at the beginning, because suitable environmental conditions were necessary prior to initiation of the laws of nature.

Since the suggestion was first made that RNA constituted the first forms of life, more and more biological and biochemical findings seem to confirm this thesis.[25,26] The first strong evidence came from the discovery of a catalytic activity of RNA molecules, which was found in a Tetrahymena group I intron by Thomas Cech and, independently, by Sydney Altman in the RNAse P complex. For these discoveries, both researchers were awarded the Nobel Prize in Chemistry in 1989.[27] The next evidence was the discovery of an RNA molecule motif, a hammerhead ribozyme, which is capable of a self-catalytic cleavage.[28] Later on, Johnston et al. showed that RNA polymerase can catalyse its own synthesis, which also is considered as strong evidence of the RNA world hypothesis.[29] In the same year, the ribosome was claimed to be a ribozyme after the discovery that the catalytic site of the ribosome is composed of RNA. The ribosome was named a living fossil of the RNA world. Moreover, many co-factors of enzymatic reactions are composed of RNA, especially these that are considered as being evolutionarily old. This finding suggests the central role of RNA in current and past biochemical pathways. Additional evidence supporting the RNA world hypothesis is the existence of viroids.[30] Viroids are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Based on their characteristics, viroids can also be considered as RNA world remnants.[31]

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

Recently, many researchers have conducted a wide range of laboratory experiments[32] to support assumptions of the RNA World theory. Moreover, mathematicians[33,34] and computer scientists[1–5] have tried to model and analyse systems that are based upon interacting RNA chains. Laboratory experiments, as well as mathematical and computer science approaches, with special attention given to the relationship between their results, are described in the next section. However, despite extensive study, it is not clear how all the pieces of the RNA world puzzle fit together, and some pieces are still missing.

Levels of Early Pre-life Organization

We propose dividing the processes of the origins of life with special attention to the RNA world hypothesis into four levels of complexity, α, β, γ, δ, and define the transitions between them that denote the change of the level of complexity. α is the level of simple inorganic compounds that were plausibly accessible on prebiotic Earth. β is the level of nucleotides. γ is the level of polynucleotide polymers. δ is the level of interacting chains of nucleotides that are either compartmentalized or unbounded. The transitions between them are as follows: $\alpha \rightarrow \beta$ denotes formation of nucleotides, $\beta \rightarrow \gamma$ denotes polymerization, and $\gamma \rightarrow \delta$ denotes self-organization. The levels of complexity and transitions between them are presented in Figure 1.

As for the $\alpha \rightarrow \beta$ transition, well-known Miller–Urey experiments have shown that, from simple inorganic compounds under conditions that mimic the early Earth environment, it is possible to create amino acids.[35,36] In addition to amino acids, fatty acids have been demonstrated to be easily accessible under primordial conditions by a Fisher–Tropsch type reaction.[37] Moreover, they were found in Carbonaceous chondrite meteorites.[38] However, neither amino acids nor lipids possess properties of inheritance because they cannot serve as a template for the next generation of molecules.

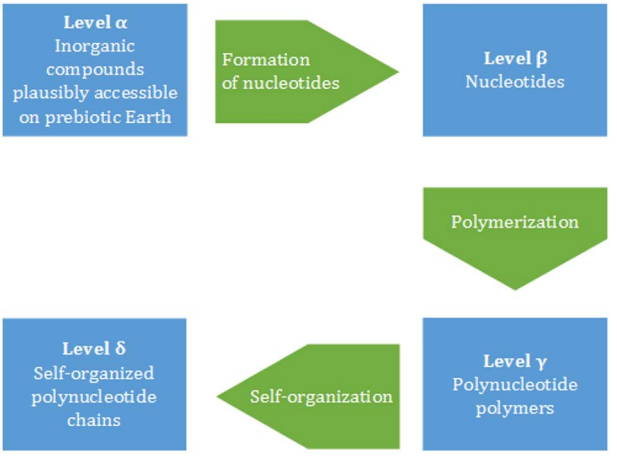


Figure 1. The levels of complexity and the transitions between them.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

Deriving ribonucleotides in a set of prebiotic conditions is a more challenging task. Juan Oro derived adenine by mixing hydrogen cyanide and ammonia in aqueous solution.[39] Later, the Sutherland Group demonstrated experimentally unconventional high yielding routes to pyrimidine ribonucleotides through amino-oxazoline intermediates.[40] The Saladino group chose another reaction scheme and concentrated on formamide as the key component of the system, which under various catalysts leads to all five nucleotides (A, G, C, T and U for RNA) and their analogues.[41] Moreover, formamide has been recently identified as one of the most common carbon-containing molecules in the Universe.[42,43] The reaction kinetic analysis of adenine in formamide formation has been studied with density functional theory, revealing that this mechanism is energetically favourable.[44] There are many possible chemical routes leading to ribonucleotides that are described in the literature. Nevertheless, the scientific community has not agreed on their plausibility in primordial conditions. What is more, intermediate states and transitions that have been recognized as the most probable have not yet been fully investigated.

With regard to the $\beta \rightarrow \gamma$ transition, the possibility of self-polymerization of nucleotide chains has been experimentally demonstrated by various researchers.[45–48] Researchers have also suggested that minerals have an essential role in the polymerization of polynucleotides.[46] Potential mechanisms for linking shorter RNA chains via ligation have also been studied.[47] The Nobel Prize winner in Physiology or Medicine, Jack Szostak, and his group concentrated on the polymerization of nucleotide chains and the biochemistry of fatty acids, which can constitute pre-cellular compartments.[49] They have shown that a selective advantage for protocells with the addition of phospholipids can be observed even at that simple level of organization.[50] Many researchers have also conducted experiments to develop ribozymes with various functions: peptidyl transferase ribozyme,[51] ligase[52,53] and nucleotide synthetase.[54] An interesting trans-aminoacylator was described in 2013,[55] having five nucleotides sufficient for reaction and being at the same time the smallest ribozyme yet discovered. Fatty acids have been shown to spontaneously form bilayer structures that are similar to a membrane consisting of phospholipids.[56–58]

The γ level itself has been studied extensively by Manfred Eigen (also a Nobel Prize winner in Chemistry) and Peter Schuster, whose biomathematics approaches led to the formulation of a quasi-species mathematical model describing the population of self-replicating polynucleotide chains.[59] Given imprecise replication, Eigen has also put forward the idea that there is an upper threshold, named the error threshold, for the length of the polynucleotide chains which, while preserved, guarantees that the information can be sustained in the quasi-species.[60] This length has been estimated to be around 100 nucleotides, which has been considered insufficient to code the polymerase with an increased accuracy of replication. This has led to Eigen's paradox and it has posed a problem for increasing the amount of information stored in the system.

Apart from biological experiments, there is no detailed approach to study the non-template-directed polymerization reaction of ribonucleotides. Such an approach could help to reveal the crucial aspects of the polymerization mechanism and suggest

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

how to overcome obstacles on the way to efficient polymerization before the formation of the polymerases.

Mathematical analysis regarding the $\gamma \rightarrow \delta$ transition has shown that there is a particular structure of self-organization, which is achievable theoretically in the quasi-species population. It can allow the system to cross the error threshold. This structure is called a hypercycle.[33] Over the years, the hypercycle theory has experienced many reformulations and methodological approaches, including application of bioinformatics. The most notable are applications of partial differential equations,[61] cellular automata[3,61–64] and the stochastic formulation of Eigen's problem.[65] In 2012, Vaidya *et al*. published the first experimental demonstration of the emergence of a cooperative network among fragments of ribozymes, which have the ability to self-assemble. Moreover, they showed their advantage over self-replicating cycles.[66] Several authors have argued that the traditional hypercycle model formulation using ordinary differential equations is vulnerable to parasites.[3,62–65,67,68] Spatial organization has been suggested to solve this problem.[3,61,64] At the same time, Takeuchi and Hogeweg showed, by a bioinformatics approach, that the replicase-parasite system is not only resistant to parasites but, more importantly, that parasites are responsible for the formation of traveling wave patterns and they are necessary to ensure the evolutionary stability of the system.[3] The work of Vaidya *et al*.[66] suggests that selfish subsystems can be integrated with cooperative subsystems and can support their growth. However, this conclusion differs substantially from results obtained in evolutionary dynamics.[69] According to evolutionary dynamics theory, selfish molecules dominate the system, even if the growth rate of a selfish subsystem in isolation is lower than cooperative subsystems. As can be seen, despite an extensive study of the role of parasites for cooperation, a different formulation of the problem can lead to completely contrasting conclusions. Moreover, regardless of the methodology that has been used to study hypercycles, the emergence of hypercyclic structure from quasi-species has not yet been explained fully. Additionally, the explanation of the emergence of the traveling waves pattern provided by Takeuchi and Hogeweg lacks mathematical formulation.[3] The methodology applied and the employed algorithms may also influence the pattern formation. Simple rules of cellular automata may result in a pattern that differs from the results obtained from simulations performed with the use of more complex methods. The dimensionality of the system may also influence pattern formation and properties.

Another issue at this level of organization is the need for membranes. Life, as we typically define it, is cellular, and the questions are why and what kind of processes led to this. The behaviour and properties of free-floating RNA chains as well as compartmentalized RNAs have been studied by wet-lab experiments,[49,50,70] mathematically with group theory,[2] and with the use of cellular automata.[62] Mansy *et al*. have shown that lipid bilayers are permeable for nucleotides.[70] But the real breakthrough came with the work of Szostak's laboratory.[49] Zhu and Szostak[71] have shown a pathway for coupled vesicle growth and division. Budin and Szostak have shown that the addition of phospholipids into a protocell's membrane gains selective advantages for the protocell.[50] Moreover, protocells can grow as a consequence of osmotic

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

movement of solvent molecules caused by the difference in osmotic potential between the interior of the protocells and their environment. This also suggested the existence of a direct link between vesicle growth and replication of internal molecules.[72]

With his group's theoretical approach, Nowak has proven that boundaries are advantageous for the system.[2] However, he does not consider the spatial aspect of the problem. In this context, the crowded state in a cell requires special attention as it can alter diffusion, leading to the different behaviour of the system. Hogeweg and Takeuchi explored the crucial role of the spatiality of a replicating system by simulating both free and compartmentalized RNAs, which showed the different trajectories of these two simulations.[62] However, they do not conclude whether one is better than the other, and if so, why. The advantages of the trade-off between compartmentalized and unbounded systems have not yet been investigated fully.

Conclusions and Future Directions

Despite great efforts put into solving the hurdles of the origins of life, there are still unsolved mysteries at every level of the complexity of life. A leading chemist in the field of the origins of life, John Sutherland, commented on recent findings: 'The big picture is it is not an RNA world, a peptide world, a lipid world. It only works if everything is connected.'[73] In our opinion, the case is even more complicated. Although the origins of life problem has its roots in biology, chemistry and biophysics, it also intersects with the fields of mathematics and computer science. It demands not only putting various molecules together in wet-lab experiments, but, more importantly, it demands the linkage of knowledge from diverse research approaches. Determining the exact linkage between various components is a major scientific challenge, as is the coupling of various levels of complexity. Making use of the data obtained at the lower level of complexity and relating it to the higher-level characteristics, as well as considering horizontal and vertical connections and information transfer, is still a demanding issue. Only in this way will we be able to infer the collective behaviour of the system and observe the emerging global organization that cannot be derived from the behaviour of its elements. Only in this way can we understand evolution and its beginnings. To achieve this goal, it will be necessary to take a truly interdisciplinary approach. It seems that bioinformatics may provide the perfect tools to model and test various scenarios of the origins of life.[74–80] Using a holistic bioinformatics approach to the problem, we will be able to unify the complex scientific landscape into a coherent model. Whether we can find answers *in silico* is still an open question, but it seems likely that linking efforts and knowledge from the various fields of science into the holistic bioinformatics view will be able to take us one step closer to understanding the origins of life.

Acknowledgements

This paper was partly supported by the grant 09/91/DSPB/0585.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

References

1. J. Wang, J. Gu, M.T. Nguyen, G. Springsteen and J. Leszczynski (2013) From formamide to adenine: a self-catalytic mechanism for an abiotic approach. *Journal of Physical Chemistry B*, **117**, pp. 14039–14045.
2. A.J. Markvoort, S. Sinai and M.A. Nowak (2014) Computer simulations of cellular group selection reveal mechanism for sustaining cooperation. *Journal of Theoretical Biology*, **357**, pp. 123–133.
3. N. Takeuchi and P. Hogeweg (2012) Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. *Physics of Life Review*, **9**, pp. 219–263.
4. J.A. Shay, C. Huynh and P.G. Higgs (2015) The origin and spread of a cooperative replicase in a prebiotic chemical system. *Journal of Theoretical Biology*, **364**, pp. 249–259.
5. W. Ma and J. Hu (2012) Computer simulation on the cooperation of functional molecules during the early stages of evolution. *PloS One*, **7**, e35454.
6. S.A. Benner (2010) Defining life. *Astrobiology*, **10**, pp. 1021–1030.
7. *Fenomen życia w ujęciu interdyscyplinarnym: teksty wykładów wygłoszonych na sympozjum naukowym zorganizowanym przez Oddział Polskiej Akademii Nauk i Wydział Teologiczny UAM w Poznaniu dnia 2 grudnia 2003 roku*. (Ośrodek Wydawnictw Naukowych, 2004).
8. C.E. Shannon (1948) A mathematical theory of communication. *Bell Systems Technology Journal*, **27**, pp. 379–423.
9. D.D. Axe (2000) Estimating the prevalence of protein sequences adopting functional enzyme folds. *Journal of Molecular Biology*, **341**, pp. 1295–1315.
10. S.C. Meyer (2010) *Signature in the Cell: DNA and the Evidence for Intelligent Design* (San Francisco: HarperOne).
11. J.U. Bowie, J.F. Reidhaar-Olson, W.A. Lim and R.T. Sauer (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science*, **247**, pp. 1306–1310.
12. W.A. Dembski (2006) *The Design Inference: Eliminating Chance through Small Probabilities* (Cambridge, UK: Cambridge University Press).
13. W.A. Dembski (2006) *No Free Lunch: Why Specified Complexity Cannot Be Purchased Without Intelligence* (Lanham, Maryland, Stany Zjednoczone: Rowman & Littlefield).
14. D.L. Abel (2009) The Universal Plausibility Metric (UPM) & Principle (UPP). *Theoretical Biology and Medical Modelling*, **6**, p. 27.
15. A.S. Eddington (2005) *The Nature of the Physical World* (Whitefish, Montana, USA: Kessinger Publishing, LLC).
16. D.H. Kenyon (1969) *Biochemical Predestination* (New York: McGraw Hill Text).
17. G. Nicolis and I. Prigogine (1977) *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (New York: Wiley).
18. S.A. Kauffman (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford, UK: Oxford University Press).
19. T.R. Cech (2012) The RNA Worlds in Context. *Cold Spring Harbor Perspectives on Biology*, 4, a006742.
20. F.H. Crick (1968) The origin of the genetic code. *Journal of Molecular Biology*, **38**, pp. 367–379.
21. L.E. Orgel (1968) Evolution of the genetic apparatus. *Journal of Molecular Biology*, **38**, pp. 381–393.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

22. S.H. Boyer (1968) The genetic code: the molecular basis for genetic expression. *American Journal of Human Genetics*, **20**, pp. 403–404.
23. R.F. Gesteland, T. Cech and J.F. Atkins (2006) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World* (New York: Cold Spring Harbor Laboratory Press).
24. C.R. Woese (1967) *The Genetic Code: the Molecular Basis for Genetic Expression* (New York: Harper & Row).
25. R.F. Gesteland (1993) *The RNA World: The Nature of Modern Rna Suggests a Prebiotic RNA World* (New York: Cold Spring Harbor Laboratory Press).
26. M. Neveu, H.-J. Kim and S.A. Benner (2013) The 'strong' RNA world hypothesis: fifty years old. *Astrobiology*, **13**, pp. 391–403.
27. The Nobel Foundation (2015) The Nobel Prize in Chemistry 1989. *Nobelprize.org*. at < http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1989/>.
28. A.C. Forster and R.H. Symons (1987) Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell*, **49**, pp. 211–220.
29. W.K. Johnston, P.J. Unrau, M.S. Lawrence, M.E. Glasner and D.P. Bartel (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, **292**, pp. 1319–1325.
30. T.O. Diener (1971) Potato spindle tuber 'virus': IV. A replicating, low molecular weight RNA. *Virology*, **45**, pp. 411–428.
31. R. Flores, S. Gago-Zachert, P. Serra, R. Sanjuán and S.F. Elena (2014) Viroids: Survivors from the RNA World? *Annual Review of Microbioogy*, **68**, pp. 395–414.
32. K. Ruiz-Mirazo, C. Briones and A. de la Escosura (2014) Prebiotic systems chemistry: new perspectives for the origins of life. *Chemical Reviews*, **114**, pp. 285–366.
33. M. Eigen and P. Schuster (1979) *The Hypercycle* (Berlin, Heidelberg: Springer)
34. P. Schuster (2011) Mathematical modeling of evolution. Solved and open problems. *Theory in Biosciences*, **130**, pp. 71–89.
35. S.L. Miller (1953) A production of amino acids under possible primitive earth conditions. *Science*, **117**, pp. 528–529.
36. S.L. Miller and H.C. Urey (1959) Organic compound synthesis on the primitive earth. *Science*, **130**, pp. 245–251.
37. T.M. McCollom, G. Ritter and B.R. Simoneit (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Origins of Life and Evolutions of Biospheres*, **29**, pp. 153–166.
38. J.G. Lawless and G.U. Yuen (1979) Quantification of monocarboxylic acids in the Murchison carbonaceous meteorite. *Nature*, **282**, pp. 396–398.
39. J. Oro and A.P. Kimball (1961) Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Archives of Biochemistry and Biophysics*, **94**, pp. 217–227.
40. M.W. Powner, B. Gerland and J.D. Sutherland (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, pp. 239–242.
41. R. Saladino, C. Crestini, S. Pino, G. Costanzo and E. Di Mauro (2012) Formamide and the origin of life. *Physics of Life Reviews*, **9**, pp. 84–104.
42. D. Despois, J. Crovisier, D. Bockelée-Morvan and N. Biver (2002) Comets and prebiotic chemistry: the volatile component. In: H. Lacoste (Ed.), *Proceedings of the First European Workshop on Exo-Astrobiology*, ESA SP-518, Noordwijk, the Netherlands, pp. 123–127.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

43. G.R. Adande, N.J. Woolf and L.M. Ziurys (2013) Observations of interstellar formamide: availability of a prebiotic precursor in the galactic habitable zone. *Astrobiology*, **13**, pp. 439–453.
44. J. Wang, J. Gu, M.T. Nguyen, G. Springsteen and J. Leszczynski (2013) From formamide to adenine: a self-catalytic mechanism for an abiotic approach. *Journal of Physics and Chemistry B*, **117**, pp. 14039–14045.
45. H.S. Zaher and P.J. Unrau (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA New York*, **13**, pp. 1017–1026.
46. J.P. Ferris, A.R. Hill, R. Liu and L.E. Orgel (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, **381**, pp. 59–61.
47. G. Costanzo, S. Pino, F. Ciciriello and E. Di Mauro (2009) Generation of Long RNA Chains in Water. *Journal of Biological Chemistry*, **284**, pp. 33206–33216.
48. K. Adamala and J.W. Szostak (2013) Non-enzymatic template-directed RNA synthesis inside model protocells. *Science*, **342**, pp. 1098–1100.
49. J.W. Szostak, *et al*. (2015) Szostak's Lab Publications. http://molbio.mgh.harvard.edu/szostakweb/publications.html.
50. I. Budin and J.W. Szostak (2011) Physical effects underlying the transition from primitive to modern cell membranes. *Proceedings of the National Academy of Science*, **108**, pp. 5249–5254.
51. B. Zhang and T.R. Cech (1997) Peptide bond formation by in vitro selected ribozymes. *Nature*, **390**, pp. 96–100.
52. M.P. Robertson, J.R. Hesselberth and A.D. Ellington (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA New York*, **7**, pp. 513–523.
53. N. Paul and G.F. Joyce (2002) A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences*, **99**, pp. 12733–12740.
54. P.J. Unrau and D.P. Bartel (1998) RNA-catalysed nucleotide synthesis. *Nature*, **395**, pp. 260–263.
55. R.M. Turk, N.V. Chumachenko and M. Yarus (2010) Multiple translational products from a five-nucleotide ribozyme. *Proceedings of the National Academy of Sciences USA*, **107**, pp. 4585–4589.
56. J.M. Gebicki and M. Hicks (1973) Ufasomes are stable particles surrounded by unsaturated fatty acid membranes. *Nature*, **243**, pp. 232–234.
57. J.M. Gebicki and M. Hicks (1976) Preparation and properties of vesicles enclosed by fatty acid membranes. *Chemical Physics Lipids*, **16**, pp. 142–160.
58. W.R. Hargreaves and D.W. Deamer (1978) Liposomes from ionic, single-chain amphiphiles. *Biochemistry (Mosc.)*, **17**, pp. 3759–3768.
59. M. Eigen and P. Schuster (1977) The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, **64**, pp. 541–565.
60. M. Eigen (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, **58**, pp. 465–523.
61. M.C. Boerlijst and P. Hogeweg (1995) Spatial gradients enhance persistence of hypercycles. *Physics of Nonlinear Phenomena*, **88**, pp. 29–39.
62. P. Hogeweg and N. Takeuchi (2003) Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Origins of Life and Evolution of the Biospheres*, **33**, pp. 375–403.
63. N. Takeuchi and P. Hogeweg (2009) Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization. *PLoS Computational Biology*, **5**, e1000542.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

64. M.C. Boerlijst and P. Hogeweg (1991) Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physics D*, **48**, pp. 17–28.
65. E. Szathmáry and L. Demeter (1987) Group selection of early replicators and the origin of life. *Journal of Theoretical Biology*, **128**, pp. 463–486.
66. N. Vaidya, *et al*. (2012) Spontaneous network formation among cooperative RNA replicators. *Nature*, **491**, pp. 72–77.
67. R.M. May (1991) Hypercycles spring to life. *Nature*, **353**, pp. 607–608.
68. M. Boerlijst and P. Hogeweg (1991) Self-structuring and selection: spiral waves as a substrate for prebiotic evolution. In: C.G. Langton, C. Taylor, J.D. Farmer, S. Rasmussen (Eds), *Artificial Life II*, Vol. 2, (Boston, USA: Addison-Wesley), pp. 255–276.
69. M.A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap Press, 2006).
70. S.S. Mansy, *et al*. (2008) Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, **454**, pp. 122–125.
71. T.F. Zhu and J.W. Szostak (2009) Coupled growth and division of model protocell membranes. *Journal of the American Chemical Society*, **131**, pp. 5705–5713.
72. J.P. Schrum, T.F. Zhu and J.W. Szostak (2010) The origins of cellular life. *Cold Spring Harbor Perspectives on Biology*, 2, a002212.
73. R.F. Service (2013) The life force. *Science*, **342**, pp. 1032–1034.
74. E. Gelenbe, E. Seref and Z. Xu (2001) Simulation with learning agents. *Proceedings of the IEEE*, **89**, pp. 148–157.
75. E. Gelenbe (2007) Dealing with software viruses: a biological paradigm. *Information Security Technical Report*, **12**, pp. 242–250.
76. T.L. Ören, S.K. Numrich, A.M. Uhrmacher, L.F. Wilson and E. Gelenbe (2000) Agent-directed simulation: challenges to meet defense and civilian requirements. in *Proceedings of the 32nd conference on Winter simulation* (Society for Computer Simulation International, San Diego, CA, USA), pp. 1757–1762.
77. S. Wasik, P. Jackowiak, M. Figlerowicz and J. Blazewicz (2014) Multi-agent model of hepatitis C virus infection. *Artificial Intelligence Medicine*, **60**, pp. 123–131.
78. S. Wasik, T. Prejzendanc and J. Blazewicz (2013) ModeLang – experts-friendly language for describing viral infection models. *Computational and Mathematical Methods in Medicine*, **8**.
79. S. Wasik, *et al*. (2010) Towards prediction of HCV therapy efficiency. *Computational and Mathematical Methods in Medicine*, **11**(2), pp. 185–199.
80. M.J. Pietal, N. Szostak, K.M. Rother and J.M. Bujnicki (2012) RNAmap2D–calculation, visualization and analysis of contact and distance maps for RNA and protein-RNA complex structures. *BMC Bioinformatics*, **13**, p. 333.

About the Authors

Jacek Blazewicz is a deputy director of the Institute of Computing Science at Poznan University of Technology. His research interests include algorithm design and complexity analysis of algorithms, especially in bioinformatics, as well as in scheduling theory. He has published widely in the above fields (over 350 papers) in many outstanding journals, among others in *IEEE Transactions on Computers*, *IEEE Transactions on Communications*, *Discrete Applied Mathematics, Discrete Mathematics*, *Parallel Computing*, *Acta Informatica*, *Performance Evaluation*, *Journal of*

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press

Computational Biology, *Bioinformatics*, *Computer Applications in Biosciences*, *Operations Research*, *European Journal of Operational Research*, *Information Processing Letters*, *Operations Research Letters*, *BMC Bioinformatics*, *ACM/IEEE Trans. on Computational Biology and Bioinformatics*, *Computational Biology and Chemistry*. He is also the author and co-author of 14 monographs. Dr Blazewicz is also an editor of the International Series of Handbooks in Information Systems (Springer Verlag) as well as a member of the Editorial Boards of ten scientific journals. His science citation index exceeds 3650 and h_index = 27 (according to ISI). In 1991 he was awarded the EURO Gold Medal for his scientific achievements in the area of operations research. In 2002 he was elected as a Member of the Polish Academy of Sciences. In 2006 he was awarded Dr.the H.C. degree from the University of Siegen. In 2012, the Copernicus Prize was bestowed upon him by DFG (D) and FNP (PL), while in 2013 he was elevated to the position of IEEE Fellow.

Szymon Wasik is a Postdoc and Assistant Professor at the Poznan University of Technology. His doctoral dissertation presented novel computing and mathematical methods for modelling of viral infections. In 2014, for the design and development of these methods, he was honoured with an Award of the Ministry of Science and Higher Education. Currently he is designing new modelling techniques in collaboration with Imperial College London and the Luxembourg Centre for System Biomedicine. His excellent algorithmic skills have been recognised both by winning many international competitions during his studies and currently by coaching representatives of Poznan University of Technology who compete in team programming contests.

Natalia Szostak is a PhD student at the Poznan University of Technology under the supervision of Professor Jacek Blazewicz. Her MSc dissertation prepared under the supervision of Professor Janusz Bujnicki presented a new tool for calculation, visualization and analysis of RNA contact and distance maps. Her main interests are computational methods for a prediction of RNA structure and the RNA world hypothesis. For some of her research results she has been awarded by the Polish Bioinformatics Society.

https://doi.org/10.1017/S1062798716000570 Published online by Cambridge University Press