This is the accepted version of the chapter “The Origin of Nucleic Acids”, published in: ***Prebiotic Chemistry and the Origin of Life***, eds. A. Neubeck, & S. McMahon (Springer, 2021). ISBN: 978-3-030-81038-2. DOI: 10.1007/978-3-030-81039-9

The Origin of Nucleic Acids

Frank Trixler

Abstract Nucleic acids are more than classical chemistry, but not quite biology. They are a manifestation of the transition from self-assembly (highly ordered) to self-organization (functional complexity) in the course of prebiotic chemical evolution. We might still be far away from conceiving the big picture of this epochal transition during the prebiotic era, but a closer look at the puzzle pieces we got so far from different research disciplines can already shape the view on nucleic acids, their origin, and the emergence of life fundamentally. This chapter sheds light on some of these puzzle elements and their relations in order to give an insight into the multi-faceted character of nucleic acids emergence and the interdisciplinary approach required to find insightful answers and new good questions.

Introduction

What are nucleic acids? The answer is easy to give and difficult at the same time. This is because it depends from which point of view we look at nucleic acids and ask that question.

From the perspective of biochemistry, we watch living systems and focus on how they work at the nanoscale. The answer to the initial question can thus easily be given by referring to the current understanding of what nucleic acids are with respect to their structure, function, and role in the intracellular network of functional biomolecules [1-3]. In structural terms nucleic acids (RNA, DNA) are the largest organized molecules known so far. They are natural polymers. Four different species of a certain type of molecules (nucleotides) act as structural building blocks for RNA and DNA respectively. When these monomeric units polymerize into a chain of nucleotides they form polynucleotides. Nucleic acids are examples of such polynucleotides. The nucleotide sequence in nucleic acids is of organized complexity. That means that its complexity is based on encoded information. The information that defines the sequence is transferred via replication – a capability of nucleic acids which is outstanding among all other known polymers. In the context of a living system the information-coded, organized complexity is functional: it enables nucleic acid sequences to provide all instructions that encode the assembly of biomolecular nanomachines and the structures of cells, organs, and organisms. In sum, nucleic acids can be defined as biopolymers with the characteristics of organized functional complexity that has the ability to replicate. This unique combination of properties makes nucleic acids to one of the most intriguing molecules and places them within the fuzzy zone between chemistry and biology.

When we change our point of view on nucleic acids and try to understand them from the perspective of prebiotic chemistry things become weird along all central aspects of their existence. Intriguing answers given by molecular biology and biochemistry are replaced by absorbing questions asked in astrobiology (Table 1). First, let us take a look at the building blocks: we see four nucleic acid nucleotides in RNA and DNA respectively. Decades of progress towards the abiotic synthesis of canonical and non-canonical building blocks and their subunits provided deep insights into possible prebiotic synthesis pathways and conditions [4,5] but many uncertainties remain with respect to prebiotic plausibility and there are numerous open issues still unsolved. Among the most puzzling aspects are the questions: why are there exactly four nucleotide species for a given type

of nucleic acids? And why is adenosine monophosphate – a subunit of one of these five building blocks – also found in many other functions and molecular structures all over the biochemistry of cellular life? Coming to the aspect of polymerization of nucleotides in water we see that it is immersed in a nagging paradox – the so called "water problem". The functional complexity of nucleic acid sequences also appears strange from the perspective of prebiotic chemistry because it is still mostly in the dark what gave birth to it initially. Biochemistry revealed that the sequence guides the assembly of living systems, but what makes a sequence functional, and what is that encoded sequence about, basically? The aspect of copying poses another paradox: similar to the question of what came first – chicken or egg – the rise of the close cooperation between nucleic acids (crucial in the copying and assembly of proteins) and proteins (crucial in the copying and assembly of nucleic acids) it is still an unsolved problem in prebiotic chemistry. But copying is more than just making more complex molecules and living systems – it is the expansion of life deep into the dimension of time. Now, what does that mean for the definition and evolution of life?

The scope of this chapter is to dive deeper into these questions. It aims to provide a clearer understanding of why it is so difficult to define what nucleic acids are when looking at them from the viewpoint of prebiotic chemistry. But exactly the understanding of this difficulty can be the starting point for scientists from various disciplines who take up the challenge and try to assemble the big picture of nucleic acids prebiotic origin and the emergence of life.

Table 1 Aspects of nucleic acids from the viewpoints of biochemistry and prebiotic chemistry

	Biochemistry	**Prebiotic Chemistry**
Building Blocks	Four nucleotide species for a given type of nucleic acids	Why just four? Why is AMP so common?
Polymerization	Via polymerase enzymes	How to solve the water paradox?
Sequence	For protein synthesis, directing the functions of living systems	What makes a sequences functional? Encoding for what, basically?
Copying	Replication of genetic information	How to solve the DNA-protein paradox? Why copying at all?

The Strange Ubiquity of AMP

If we look at biochemistry from a broader perspective and avoid getting lost in too many details something remarkable becomes apparent: in many very different biochemical processes which are not directly related to each other, a specific molecule appears as a crucial element: adenosine monophosphate (AMP) [6]. This nucleotide consists of the nucleobase adenine, a pentose, and one phosphate group (Fig. 1). But what is so special about AMP? The question is still open so far, but we can survey its multi-faceted role in biochemistry and try to find properties that at least correlate with its outstanding role.

When we shed light on nucleic acids in our survey we see that AMP is one of its canonical building blocks. But there is more: AMP is also highly relevant for the process of gene expression in the form of terminating poly(A) tails. Such tails consist of multiple AMPs. Comparative genomics suggests that polyadenylation (the addition of poly(A) tails) is very common in all known life, so the last universal common ancestor (LUCA) probably already had such a polyadenylation system [7]. Beyond these tasks for nucleic acids AMP has many more jobs in the molecular factory of a cell. It is involved in the form of AMP residues in various biomolecules which are of central importance for metabolism and signal transduction. Such biomolecules are, for example, adenosine triphosphate (ATP: energy current of the cell, also involved in intra- and extracellular signalling and many other biochemical processes), cofactors such as nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD) and acetyl conenzyme A (Acetyl-CoA) or molecules for intracellular signal transduction and other functions such as cyclic adenosine monophosphate (cAMP) and diadenosine tetraphosphate (Ap4A) (Fig. 1).

Why is AMP such an all-rounder in the biochemistry of all known life? The multi-functional role of AMP suggests that it has been selected very early in the prebiotic chemical evolution towards the origin of cellular life. Now, what property makes AMP so outstanding during this era? There is no definitive answer so far, but we have some indications which wait to further be explored.

One indication arises in experiments in which peptidyl RNAs are synthesized under conditions established for genetic copying [6]. These conditions include a mixture of a heterocyclic catalyst and a condensing agent. When RNAs and amino acids are added to this mixture in the absence of chemical preactivators, mineral surfaces, and enzymes a spontaneous formation of peptidyl RNA is reported. What is relevant for

this discussion is the observation that AMP turned out to be of outstanding reactivity in these synthesis processes. In addition, AMP reacts under the same experimental conditions with added NMN+, FMN and the tetrasodium salt of inorganic pyrophosphate to NAD^+, FAD, and ATP, respectively. All these results might indicate that AMP has favorable reactive properties for the generation of a variety of biomolecules which have a central role in cellular processes.

Another indication for an outstanding property of AMP comes from experiments that investigated a possible interplay between nucleotides and nanofluidic phenomena emerging in aqueous suspensions of particles where temporal nanoconfinements of water are formed. These experiments initially aimed to increase the efficiency of a certain interfacial effect (organic solid/solid wetting) for nanotechnological applications [8]. In the context of the experiments added phosphates turned out to be efficient catalysts for the solid/solid wetting behavior of molecular assemblies. However, when turning to biomolecular phosphates, AMP surprised with having an outstanding catalytic effect. The catalytic activity is of very high statistical significance not only in comparison to all tested inorganic phosphates but also with respect to other nucleotides and phosphate-residue containing biomolecules [9]. As the dense aqueous suspension of particles in these experiments can be regarded as an inorganic version of the crowded intracellular environment where dispersed macromolecules with nano-scale separation also create nanofluidic effects on water between them these findings have prebiotic relevance: the results suggest that catalytic reactions in temporal nanoconfinements of water may have played an important part in the prebiotic chemical evolution with AMP as a crucial actor. But which property can be responsible for the described outstanding catalytic role of AMP? An interesting correlation appears when comparing the stacking equilibrium constants of nucleoside monophosphates with the corresponding organic solid/solid wetting results. It turns out that AMP not only increases the extent of nanofluidic phenomena in these experiments more than any other used nucleoside monophosphates but also has the highest self-stacking constant among them [10] and the ability to form indefinite stacks [11].

These examples indicate that the strange ubiquity of AMP in biochemistry could be a message from the era of prebiotic chemistry transmitted by the conservative nature of evolution. Further research on that issue could thus help us to better develop and select possible scenarios about the origin of nucleic acids.

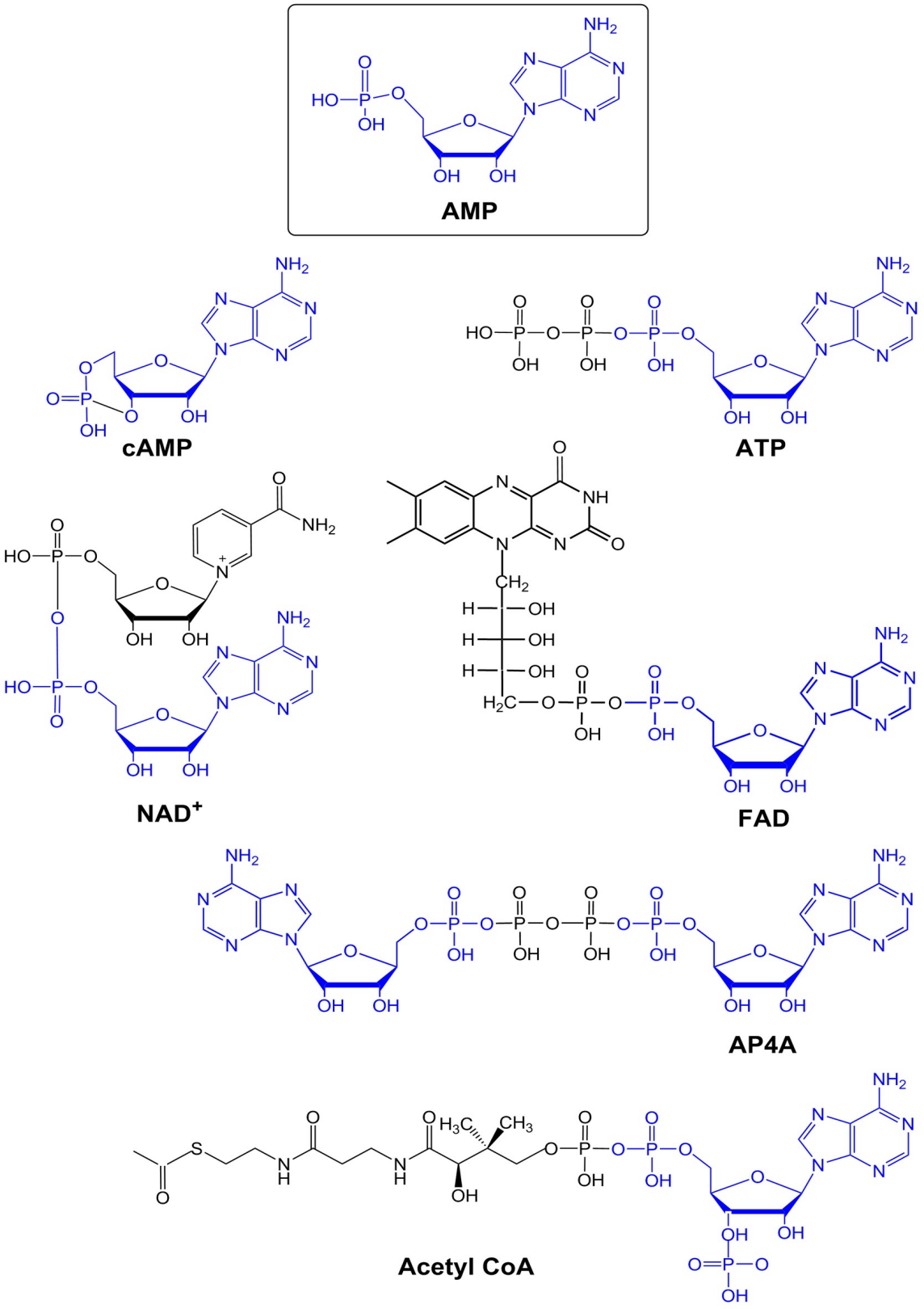

Fig. 1 AMP residues (blue) in various biomolecules which are crucial for metabolism and signal transduction.

Nucleotides and the Creative Power of Standardization

Why are there exactly four nucleotide building blocks life on Earth uses to build RNA or DNA respectively? Couldn't there be more? There is a definite "yes", but also an obvious "no".

Let us first focus on the "yes": according to results from synthetic biology there are indeed a lot more building blocks possible: the diversity of nucleic acid analogues that have been synthesized so far in order to explore the variety of structural and chemical parameters for information storage, heredity, and in vitro evolution reveals that the chemical space for synthetic genetic polymers is even tremendous [12,13]. The progress in the field of xenobiology has shown that nucleic acid analogues can successfully be created by replacing all three moieties of nucleotides with non-natural counterparts. Such modifications have been reported with respect to nucleobases, the sugar moiety (xeno-nucleic acids, termed XNAs [14]) or the phosphodiester linkage. Even combined modifications are possible, as e.g. in the form of peptide nucleic acids (PNAs) [15] where a synthetic peptide backbone replaces the natural sugar phosphate backbone (Fig. 2). Apart from replacing structural elements of nucleic acids approaches to extend their number in a genetic system have also successfully been demonstrated. For example, hachimoji RNA and DNA represent a synthetic genetic system based on eight different nucleotide species each instead of four [13]. All these results demonstrate that the structural scope of functional nucleic acids is huge. But for some reason this scope is not tapped by life as we know it.

This leads us to the "no" as the answer to the initial question. The huge structural scope of genetic polymers makes it very likely that a complex mixture of both non-canonical and canonical nucleic acids may have existed during the prebiotic era. As a consequence, some selection processes must have occurred which resulted in natural "standardization" and led to the few canonical versions of nucleic acids that are now used by all known life on Earth. What could possibly have led to such a selection before LUCA went on stage?

We can first take a look at this question from the perspective of biochemistry. Several possible processes and mechanisms have been discussed so far [16] which may have played a role in the transition from a heterogeneous mixture of genetic polymers to the homogeneous set used by known life. For example, by shedding light on the effect of UV photolysis a selective impact on nucleotides becomes apparent: experiments indicate that

canonical nucleobases have a higher stability against degrading photochemical reactions in comparison to non-canonical ones. Selective effects can also occur with respect to the sugar moiety and the phosphodiester linkage. The effects are based on steric constraints and differences in chemical reactivity which may have had an impact on the composition of prebiotic nucleic acids. There are also indications that selective pressure for superior function might have played a role in the process of natural "standardization" during the prebiotic era. This was derived from observations which revealed that mixed RNA/DNA oligonucleotides are of lower functionality compared to the known homogeneous ones. Another selective effect has experimentally been demonstrated with regard to non-enzymatic template directed primer extension reactions. Such reactions turned out to be much less efficient when running with a non-canonical nucleotide. Though being less efficient, the incorporation of a non-canonical nucleotide was reported to be possible, but, however, it inhibits continued primer extension reactions. Investigations also revealed that chimeric oligonucleotides which were formed by both canonical and non-canonical nucleotides turned out to be effective templates for RNA synthesis. These results suggest that during the prebiotic era untemplated polymerization initially generated chimeric oligonucleotides but that the formation of RNA was then favored by template copying chemistry [16].

But biochemistry is not the only discipline we can ask what possibly led to the few canonical versions of nucleotides before the wake of LUCA. Complexity science has something to add. Let us start with the genetic system: it consists of multiple, diverse elements that are highly interconnected in a coordinative and cooperative way with having the

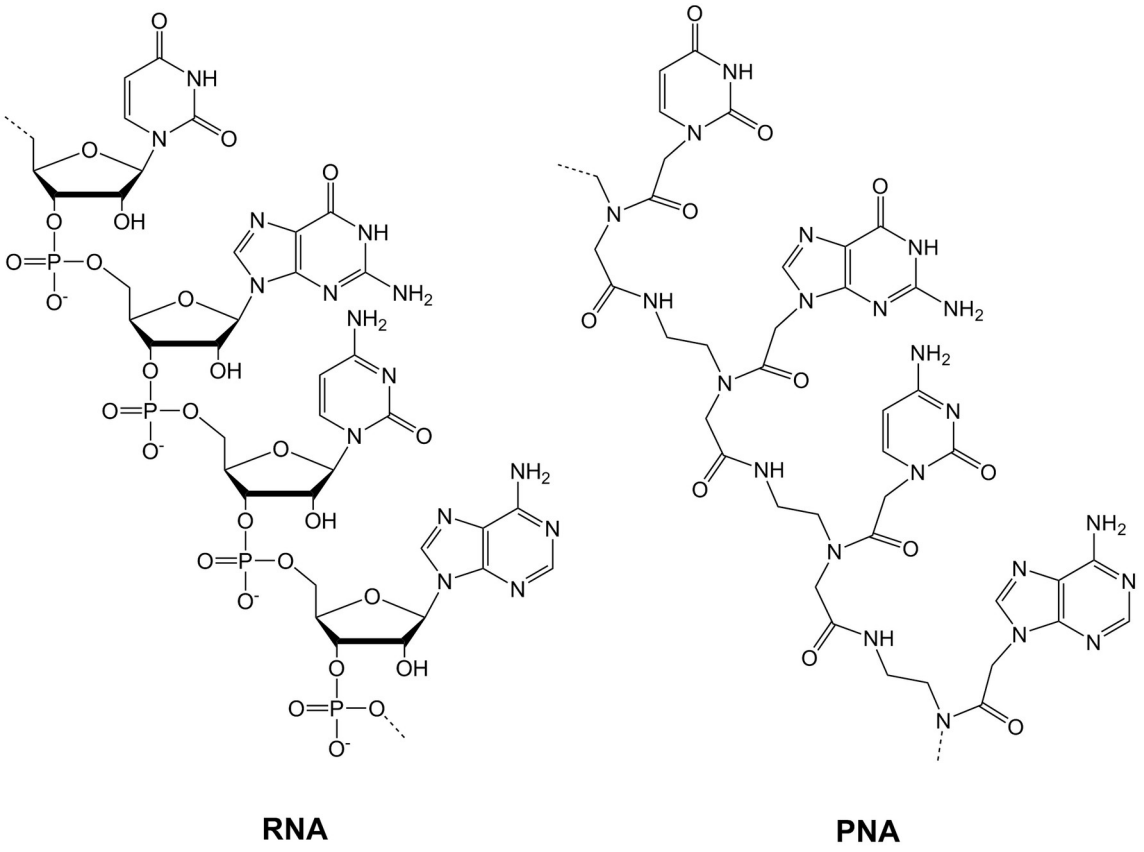

Fig. 2 Structural differences between RNA and PNA (Peptide Nucleic Acid). In PNA the ribose-phosphate-backbone of RNA is replaced by a synthetic peptide backbone.

autonomy to adapt. The sum of all these properties classifies this system to be a complex one. It is a system with functional complexity. Nucleic acids, as being elements of the genetic system, are by themself complex systems. They are a manifestation of a special category of complexity termed "Functional Sequence Complexity". We will see later in this chapter that functional sequence complexity can only exist in a narrow section of the complexity spectrum and that this section is closer to randomness than to order. What is relevant for this discussion is the relation between novelty and standards within this narrow section. If a complex, evolutionary system is either too sensitive or too resistant to variations and mutations its adaptability is corrupted. This is where standardization becomes relevant [17]. From the viewpoint of standardization, the selection of just a few versions of nucleotides from their broad chemical space may have been provided an essential contribution to poise the complexity of nucleic acids at the functional and adaptable part of the complexity spectrum, thus ensuring that the genetic system is neither too resistant nor too sensitive to mutations and thus well prepared for coping with genetic innovations. A comparable situation can be found in some economic systems where

technical innovations with high market relevance can lead to competing technical standards defined by different suppliers. This can result in so called “format wars” [18]. They usually end when one format catches on and becomes a general standard. This leads to a much better cooperative and integrative interconnection of the new technological elements within the complex socio-economical system, thus catalysing innovations enabled by the defined technical standards.

In sum, recent results from different disciplines such as xenobiology, biochemistry, and systems theory support the view on the canonical set of nucleotides as being just a small fraction of a much broader chemical space available for this class of molecules. The existence of this restriction can be interpreted as information about conditions during the prebiotic era. This message, if properly deciphered and better understood, could reveal aspects of the origin of nucleic acids that were of crucial importance for prebiotic chemistry before LUCAs descendants took the helm in evolution. The few, very basic insights we achieved so far are already stunning, but there is much work ahead to get an idea of the big picture about this facet of nucleic acids origin.

A Paradox Falls into Water

Earth is a very dry planetary body. Although is seems to be a water world, appearances are deceptive: the oceans are an extremely thin varnish with respect to Earth's diameter. A census of all water both on its surface and within the lithosphere [19] reveals that it is just about 0.04 % of Earth's total mass [20]. But despite water is a very rare compound on Earth, it is of central importance for the origin of terrestrial life: it has a highly destructive potential for prebiotic chemistry as it impedes the formation and weakens the stability of dissolved biopolymers such as nucleic acids. Nevertheless, life is spread all over the Earth. And it is just water-based.

Being attracted by paradoxes helps a lot when trying to understand life's origin. Among the amazing zoo of paradoxes [21] in the field of abiogenesis is the so-called "water problem" [22]. It is one of the most nagging but also fascinating ones.

The problem with water comes from the fact that it hampers the formation and stability of biopolymers such as nucleic acids [23]. If we take a look at phosphodiester bonds in nucleic acids we see that these bonds are formed in the process of a condensation-dehydration reaction. Such a reaction releases water. However, a product of a reaction is difficult to form when being at the same time the solvent. Thus, when occurring within an aqueous solution, a condensation-dehydration reaction is highly unfavorable. As a consequence, a spontaneous, water releasing formation of nucleic acids is prohibited by water. Even if RNA has been formed under some conditions it is spontaneously corroded by the effect of hydrolysis when being dissolved in water. RNA hydrolysis [24] means that the phosphodiester bonds of the polymer get spontaneously ruptured by water molecules.

We thus have an intriguing paradox: water inhibits the synthesis of essential biopolymers such as RNA and corrode them via hydrolysis, but, at the same time, is essential for running the biochemistry of all life as we know it.

Several methods have been proposed so far on how to overcome the "water problem" for the abiotic formation of oligonucleotides. The focus of these studies is mostly set by the scientific community on RNA. This focus is motivated by the influence of the popular RNA-World Hypothesis which is based on the assumption that before the wake of LUCA RNA was the main actor, playing a dual role not only as an information carrier but also as a catalyst for critical biochemical reactions [25]. Approaches that have successfully produced RNA-like oligonucleotides are playing with

parameters such as temperature, activating dehydration or reducing water activity. Such methods are based on just heating dissolved nucleotides at 160 °C in the presence of phosphates, adding condensing agents such as cyanamide to support dehydration reactions in water, tapping the eutectic phase between water ice crystals, using multi-component solvents such as a mixture of ammonium formate, urea, and water in order to achieve a solvent with low water activity, or even using water-free solvents such as formamide [5, 23, 26]. Another approach which is widely discussed is the application of wet-dry-cycles [27]. This method is also based on reducing water activity in the dry phase in which polymerization occurs. Although hydrolysis breaks some of the bonds during the wet phase, repositioning of the molecules enable the formation of new bonds in the subsequent dry phase, so each cycle increases the length of nucleotide polymers.

However, in view of these achievements which give valuable insights into the spectrum of possible pathways to circumvent the water problem, some weak spots appear: high temperatures far outside the known tolerance spectrum of terrestrial life, drying events, non-aqueous solvents, or the presence of condensing agents which are toxic even for primitive lifeforms are of limited prebiotic plausibility when taking evolutionary conservatism into account – the principle that evolution builds on existing solutions. In addition, as terrestrial life is happy with just using water for its biochemistry the water problem is obviously soluble in watery solutions.

A possible approach to achieve condensation reactions under aqueous, mild conditions and without needing condensing agents is based on diamidophosphate (DAP) as a crucial agent. DAP is very efficient in phosphorylating various species of prebiotic building blocks including nucleotides but the same reaction conditions turned out to be also suitable to generate oligonucleotides [28]. DAP is water-soluble and a possibly prebiotic relevant pathway to its synthesis is known which involves iron phosphate minerals [29]. In the next section, we will see that minerals have much more to say in finding prebiotically plausible solutions for the water paradox.

The Crystalline Womb

The huge diversity of life is accompanied by a huge diversity of minerals on Earth. As far as we know it, no other celestial body in the solar system presents such a tremendous variety of mineral species and morphologies [30, 31]. Are the presence of life and the highly diverse mineralogy on Earth just a coincidence or is there more?

We have seen that water destabilizes biopolymers and hampers condensation reactions. Life seems not only to cope with it but even use water as an essential element for running its complex biochemical system. Such an aqueous interplay between destruction and construction can also be observed with respect to geology: when astrochemistry and orbital dynamics work together and cast liquid water to the geology of a planetary body we see a massive impact to the geology in causing vast geological transformations. Water transformed the geology of Earth into a highly dynamic geo-*system* with multiple feedback loops. At the same time, water transformed the geochemistry on early Earth partly into *systems chemistry* and finally into biology – a system with organized complexity. Both geology and chemistry were told by water to become increasingly complex, thus gaining new possibilities of creation from destruction.

When focusing on mineral surfaces, we note that they adsorb much attention from the prebiotic chemistry community due to their considerable set of interactions with molecular building blocks of life. Mineral surfaces have been shown to be able to select and concentrate relevant organic molecules, increase their thermal stabilities and enable chiral symmetry breaking [32].

Within our context of nucleotide polymerization, the clay mineral montmorillonite deserves special attention. According to results from numerous experiments montmorillonite shows catalytic effects on the polymerization of activated nucleotides [33]. In some experiments oligomers of up to 40 nt have been detected. The catalytic effects can be traced back to the adsorption of nucleotides on the clay layers, mediated by Van-der-Waals interactions between the nucleobase groups of the nucleotides and the montmorillonite layers. After the adsorption nucleotides are oriented in such a way that the formation of 3´,5´-phosphodiester bonds becomes favorable. The catalysed polymerization reactions seem to occur only at specific active sites within the clay interlayers. But montmorillonite can do more: it is able to accelerate the formation of lipid vesicles. Such formed vesicles have often been observed to incorporate the clay particles during the process. This observation points to a possible prebiotic scenario

in which nucleotide oligomers are not only formed catalytically via clay particles but also subsequently get encapsulated within a lipid vesicle.

All these approaches are focusing on the catalytic *surfaces* of minerals. But there has recently been described a completely different way to polymerize nucleotides into nucleic acids with the help of mineral particles. This approach uses the *interspace* between particles suspended in water [34]. Due to nanofluidic phenomena emerging within temporal, nanoscopic interspaces between suspended particles the activity of water gets reduced. This is relevant for nucleotide polymerization due to thermodynamic reasons: in the previous section we have seen that the formation of phosphodiester bonds during a condensation reaction releases water, thus making nucleotide polymerization in water highly unfavorable. However, a significant reduction of water activity, caused by an interplay of various nanofluidic phenomena emerging within the nanoscopic particle interspaces can significantly reduce the thermodynamic barrier for the condensation reaction, thus making nucleotide polymerization favorable. This possibility to polymerize nucleotides is not restricted to a specific mineral species as comparative experiments with different, but highly common minerals such as graphite, magnetite, and quartz have shown. Aqueous suspensions of such minerals are geologically ubiquitous.

The interplay between minerals and organic molecules seems to be of significant help in trying to solve the water paradox in prebiotic chemistry. This help from minerals might be an indication that the presence of life and the highly diverse mineralogy on Earth is probably not just a coincidence.

Informed Molecules Chain Order with Chaos

Where should we search for life in the Universe? An obvious answer would focus on celestial bodies where indications for liquid water exist. The selection of targets for exobiology research such as Mars, dwarf planets, or icy moons in the outer solar systems can be seen against the background of a common strategy: "Follow the water" [35]. The conservative nature of this strategy is a good starting point, but we should be aware of a selection bias as we would only find water-based life. If a herd of dragons would lurk in the vast methane seas of Saturn's moon Titan [36] we would simply miss them. But there is another strategy which may be termed as "follow the waste": from the viewpoint of physics life can be regarded as a dissipative chemical system of organized complexity which is maintained by a constant flux of energy. The intake of energy is needed to locally reduce entropy, but the reduction is at the cost of an entropy increase of the surroundings or, in other words, the production of "waste". For example, oxygen in Earth's atmosphere is waste from this viewpoint as it is a byproduct of oxygenic photosynthesis. Due to its high reactivity, it would not be stable in our atmosphere without a constant resupply via lifeforms running oxygenic photosynthesis. As a consequence, oxygen on Earth can be considered as being a biomarker. We may have a similar situation on Mars with respect to methane. Thus, robotic missions that are in search for atmospheric methane on Mars can be regarded as missions that "follow the waste". The advantage of this strategy: it is not restricted to a specific chemical species but is based on a criterion of a more general nature. Another example of a more general strategy is based on thermodynamic issues and is termed "follow the energy" [35]. In the overall view, the purpose of all these strategies is to scout possible extraterrestrial locations which meet our assumptions about habitability.

But there is also a non-geographical location where habitable conditions exist. When we cast a light on the theory of complex systems in the search for life we find it thriving within a narrow section of the transition zone between order and disorder [37]. This brings us back to nucleic acids and their origin.

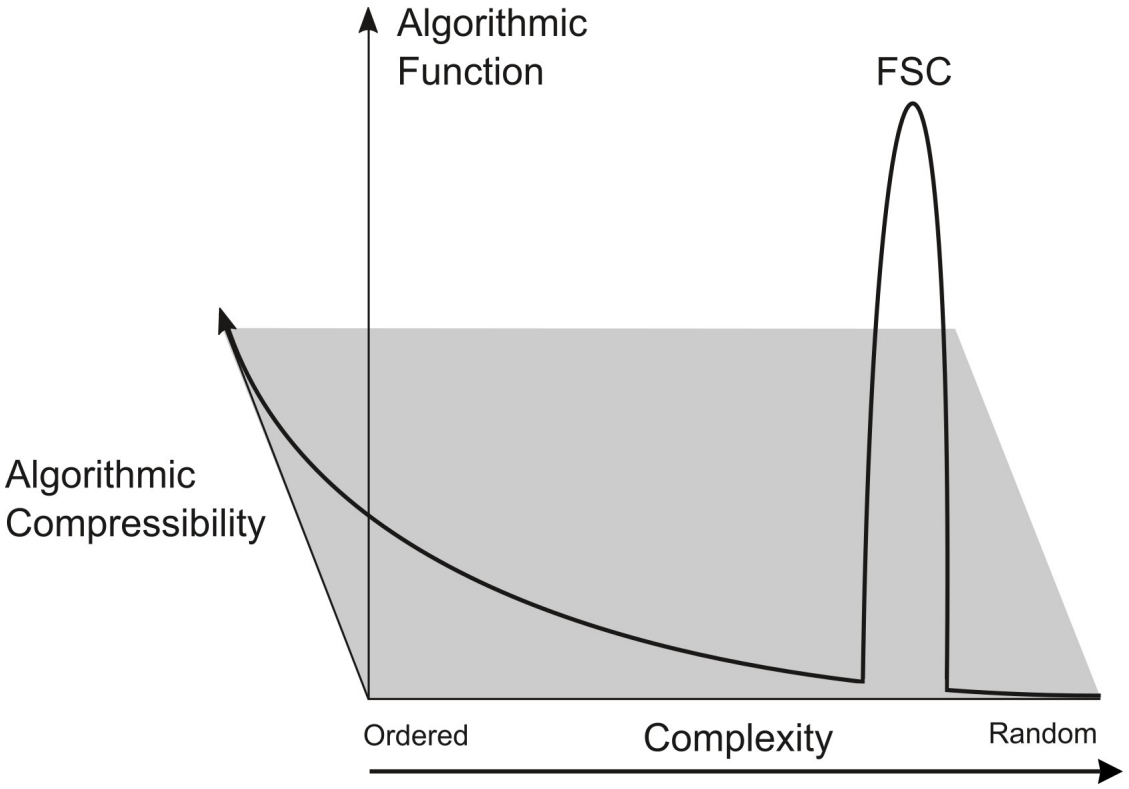

Fig. 3 Illustration of the relationship between complexity, algorithmic compressibility, and algorithmic function. FSC: Functional Sequence Complexity. Adapted from [36].

The outstanding character of nucleic acids among all other known molecules is the fact that they contain messages with meanings. The messages are addressed to the gene expression and regulation system. They are based on information which is encoded within the sequence of nucleic acids in a linear, segregated way. If we look at the genetic code from the viewpoint of information theory we see a unit of digital information: the two nucleotide triplets in a codon-anticodon pairing represent a six-bit byte of digital information [38]. So we have information stored in nucleic acids that is linear, segregated, and digitally encoded. But what does that mean for the overall nucleotide sequence? To answer this question we should have a look at complexity.

The complexity of a nucleotide sequence can be described in relation to aspects such as order, algorithmic compressibility, and algorithmic function [39]. A highly ordered sequence with a simple, repeating motive such as a poly(A) tail of mRNA has the lowest possible complexity. When complexity gradually increases it reaches the maximum at randomness where no order exists at all. Thus, the degree of complexity of a given

sequence can be described in relation to two extrema: order (minimal complexity) and randomness (maximum complexity) [40].

In addition to order/disorder, the aspect of algorithmic compressibility can be related to sequence complexity. What does algorithmic compressibility mean within this context? Let us take the example of a poly(A) tail of mRNA with the sequence "AAAAA". An algorithmic compression program for this ordered simple repetitive sequence can be very short by just reading "5A". This tells us that a sequences with low complexity (high degree of order) has high algorithmic compressibility. When the sequence complexity is increased its algorithmic compressibility decreases.

But there is even more about sequence complexity. It can also be related to a third aspect: algorithmic function. It is an aspect that refers to what a sequence can do itself (and not what can be done to it as in the case of algorithmic compression) and thus indicates the *functionality* of a sequence. Functionality means in this context the ability of a sequence to be instructional. When including the dimension of functionality in the description of sequence complexity something very remarkable becomes apparent: the ability of a sequence to be instructional (and thus functional) turns out to only be possible within a rather narrow part of the complexity spectrum. This part appears much closer to randomness than to the ordered end of complexity and opens up a new category of sequence complexity. This category is termed "*Functional Sequence Complexity* [39]. The relevant aspect of this discussion is that it enables algorithmic instructions. Outside the functional part of the complexity spectrum sequences of various degrees of complexity exist of course. But they cannot be instructive due to a very low level of algorithmic function (Fig. 3).

In sum, the theory of complex systems enables us to describe nucleic acids as polymers with a unique type of complexity – the functional sequence complexity. Its characteristic degree of aperiodicity opens up a window to algorithmic programming of biological organization. Thus, functional sequence complexity is of central importance for understanding the origin, function, and evolution of nucleic acids and can be regarded as the outcome of the nonrandom selection of prescriptive sequences during chemical and biological evolution [39]. As being much closer to randomness than to order the peculiar degree of aperiodicity of functional sequence complexity enables nucleic acid sequences to be *instructional* and *adaptive* to various environmental influences, their fluctuations and overall transformations.

After finding nucleic acids in a narrow section of the transition zone between order and randomness where algorithmic programming is possible

we can pick up this functional feature and turn to the initial question: where should we search for life in the Universe? The characteristic of nucleic acids gives us the hint to include the complexity vector in our toolbox which allows us to search for functional complexity. This category of complexity can be materialized in any form of aperiodicity with algorithmic programming that ultimately leads to organisation instead of just self-ordering (e.g. crystals) or randomness (the "asphalt problem" [22]) (Fig. 4).

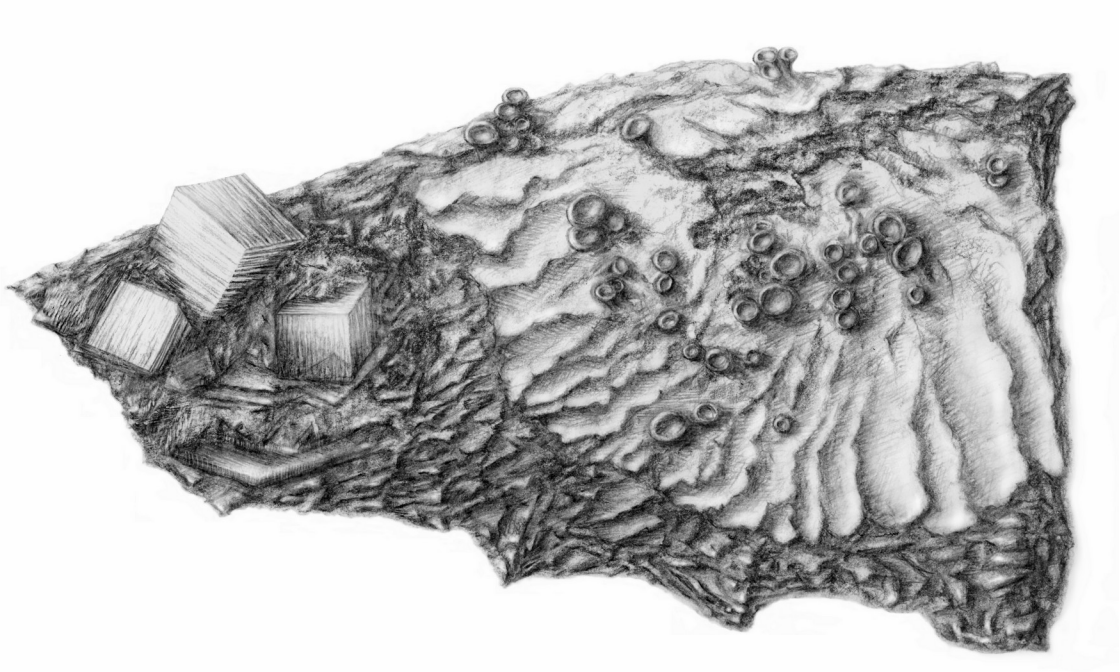

Fig. 4 Illustration of different complexity categories. A lichen (upper right part) representing *functional complexity* grows on a rock representing nearly *random complexity* and is accompanied by cubic crystals (left part of the rock) which represent *ordered complexity*. The functional complexity of the lichen is located between order and randomness, but closer to the random side. Its functional complexity is algorithmically programmed by nucleic acids which itself represent *functional sequence complexity*. Credit: H. Turhyt.

What is the Code About, Basically?

That question is easy to answer – until someone asks. Of course, we know from biochemistry that the genetic code of nucleic acids provides all instructions that encode the assembly and regulation of subcellular entities and the structures of cells, organs, and organisms. But why taking the trouble to run an encoded, organized assembly and regulation system? What motivated prebiotic chemistry to gradually become biology?

An analogy may be of some help to approach an answer to the initial question. We can ask: what is the construction plan of a house about, basically? Of course, to build a house. But that answer is only halfway to the core: what is the purpose of a house? It is to give a stable, protective environment for its inhabitants. This is usually needed because the builders want to settle in an environment that differs from those in which their distant ancestors initially have been evolved. So, the construction of a house, designed in adaption to its environment, is basically to allow humans to exist in locations where they haven't been adapted physically but want to stay for work by tapping the local resources for maintaining their existence. The house has to be designed in such a way that it is an open system, having a constant throughput of energy (electricity, heat, waste) and an intake of water in order to remain stable. Thus, when looking at a construction plan from this perspective, aspects known from evolutionary biology and thermodynamics become apparent. Against the background of this analogy, we now can approach the initial question via a combination of evolutionary biology and thermodynamics [41, 42].

A genetic code contains all instructions to assemble a living system. From the viewpoint of thermodynamics, a living system is a dissipative structure with functional complexity. A dissipative structure is a structure far from thermodynamic equilibrium. This means that it requires a net flow of energy and matter to maintain its dynamically ordered structure via degrading free energy. The energy degrading kinetic pathways in living systems are genetically instructed. An exchange of energy and matter with the environment is essential as it reduces the entropy of the system so a dynamically ordered structure can exist. Due to the essential exchange of energy and matter with the environment living systems are ecological entities that are dependent on their environment. Selection favors those living systems which manage either to increase the energy throughput or to use a given one more efficiently for both maintaining their internal organization (avoiding decay) and for replication.

From the perspectives of thermodynamics and evolutionary biology it seems that the genetic code is for instructing a dissipative, complex chemical system not to passively wait for resupply of free energy (as e.g. a candle flame – a non-complex dissipative structure – does) but to use mutation and selection for computing the local environment. This is done to actively explore and tap new thermodynamic disequilibria of any kind so the system can expand its dissipative structure via autopoiesis and replication deep into the dimension of time over billions of years.

Who Holds the Copyright? The Case Nucleic Acids versus Amino Acids

Where to start? Answering that question is problematic for prebiotic chemists, but obviously not for prebiotic chemical evolution when it comes to the DNA-protein paradox.

The research on the emergence of life runs again into a severe paradox when facing the strong interdependence between nucleic acids and proteins. The design of possible scenarios for the origin of this close linkage causes considerable headaches as the interdependence is of strict circular nature: DNA contains all instructions for algorithmically programming the organisation of the complex polypeptide sequence that defines the function of proteins while, on the other side, the replication, processing and repair of DNA are almost exclusively made by proteins. In other words: nucleic acids are essential for the synthesis of proteins, and proteins are essential for the synthesis of nucleic acids. One side cannot come into being without the other side.

It is this DNA-protein paradox that gave rise to the "RNA-World Hypothesis" [25]. The hypothesis is mainly based on the assumption that a primitive ribosome could have been existed in the prebiotic era solely based on RNA. Or, in other words: RNA came first by having a self-replicating capability. This idea was boosted when ribozymes have been discovered with some catalytic function e.g. in terms of self-splicing [43]. Further support comes from the fact that the energy currency of the cell – ATP (Fig. 1) – is a building block of RNA. Experimental observations also reveal that the core of ribosomes consists solely of RNA and that the activity of protein enzymes often requires RNA-like cofactors [44]. However, the hypothesis generates its own numerous paradoxes [45] and has extensively been challenged so far [44]. In sum, it appears that persisting on an "RNA first" scenario didn't really help to solve the circular RNA-protein paradox in a plausible way and that many unsolved questions remain regarding the origin of an RNA-World.

Thus, models which try to circumvent a decision on what came first become increasingly attractive. Such models are based on the assumption that a prebiotic chemical evolution of polypeptides and nucleic acids was closely connected right from the start. We can define this start as the emergence of prebiotic synthesis pathways resulting in the simultaneous generation of nucleotides and amino acids. Such pathways have been discovered and described in recent years [46]. Against this background an

experimental finding [34] could give a hint to how the initial connection between nucleotides and amino acids may have been started: water within nanoscopic confinements shows various nanofluidic phenomena which alter its properties compared to the bulk. Experimental results suggest that such altered thermodynamic properties of water allow not only to circumvent the water paradox for the polymerization of nucleotides into RNA. Moreover, when an amino acid was added to the aqueous solution of nucleotides it significantly catalyses nucleotide polymerization under nanofluidic conditions. This may be indicative of a very early interplay between the building blocks of nucleic acids and proteins.

When it comes to the complexity level of biopolymers some basic interactions between RNA and peptides have been revealed in which one partner benefits in terms of aspects such as functional structure, reactivity, or stability [46]. However, in living systems, we see mutual benefits between both biopolymer species. In this light, the relevance of recent studies becomes apparent that provide indications for a very early co-evolution of polypeptides and polynucleotides starting with mutual stabilizations of RNA and proto-peptides [46].

Despite such substantial achievements in prebiotic chemistry research that uncover some basic functional interactions between both types of biopolymers the big picture needed to solve the DNA-protein paradox remains unclear. There is still a wide-open field for further research for all who take up the challenge and try to solve this paradox from a broader perspective.

A New Form of Stability Arises

While the oceanic crust on Earth is constantly recycled, supercontinents form and break up again and mountains rise and fall over hundreds of millions of years, the presence of nucleic acids as one of the most vulnerable molecular species persists all over the aeons despite being constantly attacked by hydrolysis, radiation, high temperature and many other destructive influences from the environment. But it is exactly the extreme fragility of nucleic acids that makes their presence on Earth stable over billions of years. To comprehend this seemingly paradoxical issue and to understand how it has been solved conceptually, we have to pay a visit to one of the most influential research institutions of ancient times.

The great library of Alexandria stored thousands of written documents, mainly in the form of rolls of papyrus and parchment. Environmental conditions in that geographical area such as high humidity at the borderline of the Nile delta have led to a comparatively vast decomposition of the rolls. To cope with it, a huge effort was required to constantly replicate the library stocks. This needed a permanent source of funding. Although the Museion, including its great library, was attacked several times in the course of its history, it seems that the very end came when the funding was cut.

Today we can look at this situation from the viewpoint of physics and chemistry: instructional information was stored via fragile biopolymers (the main constituent of papyrus and parchment) as the main constituent of a metastable medium (rolls). The fragility of the medium was compensated by replication (copyists of the Museion). Replication needed a constant flux of energy (salary) by tapping thermodynamic disequilibria (funding). So the stability of the ancient data storage device was not given due to low chemical reactivity but rather due to another kind of stability raised by replication. What survived as information from the ancient documents usually comes from a chain of copies of copies. By far the most amount of rolls from ancient times didn't survive over time, but their content did when having been repeatedly replicated.

This brings us back to nucleic acids: most probably not a single genetic polymer snippet survived the billions of years since LUCA explored the early archean world, but for life, this is not the point. Living systems compensate for the fragility of their components and even their whole system via replication. The emergence of replicative chemistry changed everything in the prebiotic world and can be regarded as being a "phase transition" in prebiotic chemical evolution. This is because a replication

reaction is dissimilar to every other chemical reaction. It seems that replicative chemistry started with the emergence of nucleic acid species integrated into a prebiotic chemical system. The rise of replicative chemistry opened up a whole new world beyond chemical stability: it enables a molecular species such as nucleic acids to persist over time despite a constant degradation of individual entities. We thus have a new kind of stability which is termed "Dynamic Kinetic Stability (DKS)" [47].

What is the characteristic of DKS? A system based on DKS remains in a steady state as long as its replication rate is equal to or higher than the decay rate. It is opposite to thermodynamic (or "chemical") stability but includes it at the same time. A chemically (meta)stable system has reached a (local) minimum of Gibbs free energy. This can occur spontaneously via a downhill reaction (release of energy). Such a system remains stable as long as there is no intake of energy. In contrast, a system with DKS needs to be continuously reactive via an intake of energy (metabolism) in order to maintain replication. Its elements have to be of low chemical stability in order to overall maintain dynamic kinetic stability.

DKS has many fundamental implications for prebiotic chemistry and biology [48]. What is important for our discussion is the consequence of DKS for nucleic acids: it makes the presence of this molecular species with replication capability extremely stable over billions of years despite – or even because – of extremely low chemical stability.

Summary and Outlook

In this chapter, we took a spot on nucleic acids and discussed what happens when we change the perspective on these biopolymers from biochemistry to prebiotic chemistry: the clear biochemical definition of what nucleic acids are became fuzzy from this perspective. What we have seen were paradoxes and messages from the prebiotic era waiting to be fully deciphered. We can view these aspects as puzzle pieces which, in sum, give us a deeper insight into the multi-faceted character of nucleic acids and their origin.

But these instructional polymers tell us more: we can stitch the collected puzzle pieces together and see if something takes shape from a broader perspective.

The ubiquity of AMP in biochemistry comes along with a uniquely high self-association tendency among other nucleoside monophosphates. This points to strong supramolecular interactions which can form assemblies with high thermodynamic (chemical) stability. This kind of stability also becomes relevant within the context of "standardization" of nucleic acids building blocks: canonical nucleobases are more stable against photolysis than non-canonical ones. We also see an important contribution of thermodynamic stability within the context of the "water problem": DNA is used by living systems for long-term genetic information storage due to its higher chemical stability against hydrolysis compared to RNA [24]. A certain kind of stability takes also shape on the puzzle piece that refers to the aspect of standardization of nucleotides: selection during prebiotic chemical evolution poised the complexity of nucleic acids at the functional and adaptable part of the complexity spectrum. Thus, standardization of nucleic acids building blocks created and still maintains functional stability as well as the stability of adaptability. Furthermore, the replication ability of nucleic acids causes another kind of stability that enables these polymers to be persistent despite their fragility: Dynamic Kinetic Stability.

In sum, when viewed from a broader perspective, nucleic acids seem to be optimized by prebiotic chemical and biological evolution to work with different types of stability at the same time: we see thermodynamic stability with respect to their building blocks and the long-term genetic information storage. There is also functional stability and the stability of being adaptable – both maintained by functional sequence complexity. And we see dynamic kinetic stability maintained by the ability to replicate. It is easy to run into crystallization (order) or asphaltization (randomness), but how prebiotic chemical evolution managed to reach and combine these different kinds of stability in one molecular entity (nucleic acids) and balance them at the

narrow peak of functional complexity could only be revealed by achieving an even bigger picture. But this needs more puzzle elements, waiting for being discovered and collected by those who take the challenge.

References

1 S. Neidle: *Principles of Nucleic Acid Structure* (Elsevier, 2008)
2 G.M. Blackburn, M.J. Gait, D. Loakes, D.M. Williams: *Nucleic acids in chemistry and biology* (Royal Society of Chemistry, Cambridge, 2006)
3 S. Doonan: *Nucleic acids* (Royal Society of Chemistry, Cambridge, 2004)
4 D. M. Fialho, T. P. Roche, N. V. Hud, *Chem. Rev.* **120**, 4806−4830 (2020)
5 M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **120**, 4766−4805 (2020)
6 M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **54**(48), 14564 (2015)
7 V. Anantharaman, E. V. Koonin, L. Aravind, *Nucleic Acids res.* **30**(7), 1427 (2002)
8 A. Eberle, T. Markert, F. Trixler, *J. Am. Chem. Soc.* **140**(4), 1327 (2018)
9 A. Greiner, F. Trixler, Adenosine Monophosphate in Temporal Nanoconfined Water Catalyzes Molecular Self-Assembly, in: *Astrobiology – Life in the Context of Cosmic Evolution, Conference Abstracts* (2016), p. 46
10 J. Norberg, L. Nilsson, *J. Am. Chem. Soc.* **117**(44), 10832 (1995)
11 R. Tribolet, H. Sigel, *Biophys. Chem.* **27**(2), 119 (1987)
12 H.J. Cleaves, II, Ch. Butch, P.B. Burger, J. Goodwin, M. Meringer, *J. Chem. Inf. Model.* **59**, 4266 (2019)
13 P. Nie, Y. Bai, H. Mei, *Molecules* **25**(15), 3483 (2020)
14 J.C. Chaput, P. Herdewijn, *Angew. Chem. Int. Ed.* **58**(34), 11570 (2019)
15 P.E. Nielsen, D.H. Appella (Eds.): P*eptide nucleic acids: methods and protocols (*Springer Science & Business Media, 2014)
16 S.C. Kim, L. Zhou, W. Zhang, D.K. O'Flaherty, V. Rondo-Brovetto, J.W. Szostak, *J. Am. Chem. Soc.* **142**(5), 2317 (2020)
17 B.D. Higginbotham, *The Standardization of Standardization: The Search for Order in Complex Systems*. PhD Thesis, George Mason University, Fairfax, USA (2017).
18 N. Anscombe, *Nature Photon.* **2**, 412 (2008)
19 D. Pearson, F. E. Brenker, F. Nestola, J. McNeill, L. Nasdala, M. T. Hutchison, S. Matveev, K. Mather, G. Silversmit, S. Schmitz, B. Vekemans, L. Vincze, *Nature* **507**, 221 (2014)
20 M. Williams, *Universe Today*, 12/1 (2014) (https://www.universetoday.com/65588/what-percent-of-earth-is-water/)
21 S.A. Benner, *Orig. Life Evol. Biosph.* **44**, 339 (2014)
22 S.A. Benner, H. J. Kim, M. A. Carrigan, *Acc. Chem. Res.* **45**(12), 2025 (2012)
23 C. Lang, J. Lago, M. A. Pasek: *Phosphorylation on the Early Earth: The Role of Phosphorous in Biochemistry and its Bioavailability*, in: Handbook of Astrobiology, V. M. Kolb, Ed. (CRC Press, 2019), chap. 5.8.
24 D. Voet, J.G. Voet, Ch.W. Pratt: *Fundamentals of Biochemistry – Life at the Molecular Level* (Wiley, 2019), p 854
25 K. Le Vay, H. Mutschler, *Emerg. Top. Life. Sci.* **3**(5), 469 (2019)
26 A. Lazcano, *ACS nano* **12**(10), 9643 (2018)
27 P. G. Higgs, *Life* **6**(2) 24 (2016)
28 C. Gibard, S. Bhowmik, M. Karki, E.-K. Kim, R. Krishnamurthy, *Nat Chem.* **10**(2), 212–217 (2018)
29 C. Gibard, I. B. Gorrell, E. I. Jiménez, T. P. Kee, M. A. Pasek, R. Krishnamurthy, *Angew. Chem.* **131**(24), 8235-8239 (2019)

30 R.M. Hazen, D. Papineau, W. Bleeker, R.T. Downs, J.M. Ferry, T.J. McCoy, D. A. Sverjensky, H. Yang, *American Mineralogist* **93**, 1693 (2008)
31 E.G. Grosch, R.M. Hazen, *Astrobiology* **15**(10), 922 (2015)
32 R.M. Hazen, D.A Sverjensky, *Cold Spring Harbor perspectives in biology* **2**(5), a002162 (2010)
33 N. Kitadai, S. Maruyama, *Geosci. Front.* **9**, 1117 (2018)
34 A. Greiner, F. Trixler, Prebiotic Reaction Vessels – RNA Formation in Nanoconfinements of water. *19th EANA Astrobiology Conference*, Orléans, France (2019).
35 X.C. Abrevaya, R. Anderson, G. Arney, D. Atri, A. Azúa-Bustos, J.S. Bowman, W.J. Brazelton, G.A. Brennecka, R. Carns, A. Chopra, J. Colangelo-Lillis, Ch.J. Crockett, J. DeMarines, E.A. Frank, C. Frantz, E. de la Fuente, D. Galante, J. Glass, D. Gleeson, Ch.R. Glein, C. Goldblatt, R. Horak, L. Horodyskyj, B. Kaçar, A. Kereszturi, E. Knowles, P. Mayeur, S. McGlynn, Y. Miguel, M. Montgomery, C. Neish, L. Noack, S. Rugheimer, E.E. Stüeken, P. Tamez-Hidalgo, S.I. Walker, T. Wong, *Astrobiology* **16**(8) 561 (2016)
36 J.D. Hofgartner, A.G. Hayes, J.I. Lunine, H. Zebker, B.W. Stiles, C. Sotin, J. W. Barnes, E.P. Turtle, K.H. Baines, R.H. Brown, B.J. Buratti, R.N. Clark, P. Encrenaz, R.D. Kirk, A. Le Gall, R.M. Lopes, R.D. Lorenz, M.J. Malaska, K.L. Mitchell, P.D. Nicholson, P. Paillou, J. Radebaugh, S.D. Wall, C. Wood, *Nature Geoscience* **7**(7) 493 (2014)
37 J. Hidalgo, J. Grilli, S. Suweis, M.A. Muñoz, J.R. Banavar, A. Maritan, *PNAS* **111**(28), 10095 (2014)
38 H.P. Yokey: *Information theory, evolution, and the origin of life* (Cambridge University Press, New York, 2005)
39 D. L. Abel, J.T. Trevors, *Theor. Biol. Med. Model.* **2**(1), 29 (2005)
40 K.K. Durston, D.K. Chiu, D.L. Abel, J.T. Trevors, *Theor. Biol. Med. Model.* **4**(1), 47 (2007)
41 S. Black, *Persp. Biol. Med.* **21**(3), 348 (1978)
42 J.S. Wicken, *Syst. Res.* **15**(5), 365 (1998)
43 K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, T.R. Cech, *cell* **31**(1), 147 (1982)
44 E.J. Hayden, N. Lehman, P.J. Unrau, *RNA and Ribozymes in the Development of Life,* in: Handbook of Astrobiology, V. M. Kolb, Ed. (CRC Press, 2019), chap. 6.1
45 M.P. Robertson, G.F. Joyce, *Cold Spring Harbor perspectives in biology* **4**(5), a003608 (2012)
46 M. Frenkel-Pinter, J. W. Haynes, A. M. Mohyeldin, M. C, A. B. Sargon, A. S. Petrov, R. Krishnamurthy, N. V. Hud, L. D. Williams, L. J. Leman. *Nat. Commun.* **11**(1), 1-14 (2020)
47 A. Pross, *Curr. Org. Chem.* **17**(16), 1702 (2013).
48 A. Pross: *What is Life? How chemistry becomes biology* (Oxford University Press, Oxford, 2012)