
A prebiotic pathway to activate, polymerize and recycle ribonucleic acids

Juliette Langlais



München 2025

A prebiotic pathway to activate, polymerize and recycle ribonucleic acids

Juliette Langlais

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Juliette Langlais
aus Paris

München, den 30.10.2024

Erstgutachter: Prof. Dr. Dieter Braun
Zweitgutachter: Prof. Dr. Oliver Trapp
Tag der Einreichung zur Korrektur: 30.10.2024
Tag der mündlichen Prüfung: 13.12.2024

Zusammenfassung

Eine der wichtigsten Hypothesen für die Entstehung des Lebens auf der Erde ist, dass es aus einem System funktionalisierter RNA entstanden ist, das seine eigene Replikation katalysieren kann. Diese Ribozyme sind jedoch mindestens 50 Monomere lang, was die Frage aufwirft wie es abiotisch aus einem Pool von Nukleinsäuren de novo RNA-Stränge entstehen können, die lang genug sind, um aus demselben Pool von Nukleosiden/Tiden und einfachen Molekülen neue Stränge herzustellen? Es wurden mehrere Studien zu den verschiedenen Schritten durchgeführt, die zur Bildung der ersten Oligonukleotide führen konnten: Nukleotidsynthese, Phosphorylierung, Polymerisation, Ligation, Replikation, Selektion und Ribozymaktivität. Bis heute gibt es jedoch keinen Weg, auf dem alle diese Schritte unter denselben präbiotisch plausiblen Bedingungen effizient ablaufen.

In dieser Arbeit wird ein Weg zur Bildung der ersten kurzen RNA-Stränge vorgestellt, der auf der Fähigkeit der 2',3'-zyklischen Phosphatgruppe beruht, Diphosphoesterbindungen durch basenkatalysierte Transphosphoesterifikation unter milden, trockenen Bedingungen zu bilden. Im ersten Teil wurde die Polymerisation von 2',3'-zyklischem Phosphatguanosin bei alkalischem pH-Wert in Gegenwart von Kaliumsalzen, ohne weiteren Katalysator und durch Trocknung bei 140°C beobachtet. Es konnten Oligomere von bis zu 10mere nachgewiesen werden, was einer Gesamtausbeute der Polymerisation von etwa 3 % entspricht. A, C und U zeigten eine deutlich geringere Effizienz bei der Oligomerisierung. Bei gemischten Proben wurde der Einbau der anderen Nukleobasen in die G-Oligomere nachgewiesen, aber die Produkte wurden weiterhin stark von G dominiert. Die Reaktion fand zusätzlich auch an einer Wasser- Luft-Grenzfläche statt, die einem Temperaturgradienten ausgesetzt war. In dieser aus dem Gleichgewicht geratenen Umgebung durchlaufen die Moleküle Feucht-Trocken-Zyklen, und die Produkte weisen eine etwas geringere Tendenz zu G auf.

Im zweiten Teil wurde die gleiche Reaktion im manuellen Nass-Trocken-Verfahren bei Raumtemperatur durchgeführt. Diese Bedingungen verbessern die Polymerisationsausbeute für alle vier Nukleotide erheblich, insbesondere für A, C und U, für die 4-6mere nachgewiesen wurden. Der optimale Polymerisations-pH-Wert ist 10 für die Purinbasen G (70 %) und A (38 %) und 11 für die Pyrimidinbasen U (36 %) und C (39 %). Die Hydrolyse ist auch bei Raumtemperatur geringer, aber der Anteil an hydrolysierten Produkten nimmt mit höherer Zyklenzahl und höherem pH-Wert zu. Eine Tendenz zu G bleibt dennoch bestehen; Die Beobachtung des Trocknungsprozesses zeigt, dass die G-Monomere im getrockneten Zustand Flüssigkristalle und kristalline Phasen bilden. A, C und U bilden amorphe Phasen, obwohl für A einige kristalline Strukturen nachgewiesen werden konnten. Die Mischpolymerisation, die G einschließt, zeigt ausgeprägte amorphe und kristalline Phasen, die gemischten Produkte haben aber niedrige Konzentrationen, trotz einer ausgewogeneren Zusammensetzung bei pH 11.

Der letzte Teil konzentrierte sich auf die Bildung der 2',3'-zyklischen Phosphatgruppen unter Bedingungen, die mit der anschließenden Polymerisation vereinbar sind; Trocknung einer alkalischen Lösung bei niedriger Temperatur. Orthophosphat ist unter diesen Bedingun-

gen eher unreaktiv, aber Trimetaphosphat erzeugt eine gewisse Menge an 2',3'-zyklischen Phosphatprodukten durch Reaktion mit Nukleosid und 2'-/3'-Phosphat-Nukleotid bei alkalischen pH-Wert und einer Trocknungstemperatur von 60 °C. Die Ausbeuten liegen unter 1 %, aber einige Dimere können nachgewiesen werden. Die Anwesenheit anderer kleiner organischer Moleküle wie Harnstoff und Aminosäuren verbessert die Ausbeuten und beeinflusst die Regioselektivität der Reaktion. Die Phosphorylierung und Kondensation von Aminosäuren wird unter denselben Bedingungen beobachtet, was die parallele Bildung von RNA-Oligomeren und Peptiden plausibel macht.

Abstract

One of the most significant hypotheses for the emergence of life on Earth is that it started from a system of functionalized RNA able to catalyze its own replication. But those ribozymes are at least 50 monomers-long, which brings up the question: how to abiotically produce *de novo* RNA strands long enough from a pool of nucleosides/tides and simple molecules? Various works have been conducted on the different steps leading to the formation of the first oligonucleotides: nucleotide synthesis, phosphorylation, polymerization, ligation, replication, selection and ribozyme activity. But to this day, there is no pathway where all those steps are efficient in the same prebiotically-plausible conditions.

This thesis presents a pathway to the formation of the first short RNA strands based on the ability of the 2',3'-cyclic phosphate group to form diphosphoester bonds by base-catalyzed transphosphoesterification in mild, dry conditions. In the first part, the polymerization of 2',3'-cyclic phosphate guanosine was observed at alkaline pH in presence of potassium salts, without any other catalyst, and by drying at 40 °C. Oligomers up to 10mers could be detected, for a total yield of polymerization of around 3 %. A, C and U showed a notably lesser efficiency toward oligomerization. In the case of mixed samples, incorporation of the other nucleobases in the G oligomers was detected, but the products remained heavily dominated by G. The reaction also occurred at a water-air interface submitted to a temperature gradient. In this local out-of-equilibrium environment, the molecules go through wet-dry cycling, and the products have a slightly lesser bias toward G.

In the second part, the same reaction was conducted in manual wet-dry cycling at room temperature. The conditions improve considerably the polymerization yield for all four nucleotides, in particular A, C and U, for which 4-6mers were detected. The optimal polymerization pH is 10 for the purine bases G (70 %) and A (38 %), and 11 for the pyrimidine bases U (36 %) and C (39 %). Hydrolysis is also lower at room temperature, but the yield of hydrolyzed products still increases with higher numbers of cycles and pH value. A bias toward G remains: the observation of the drying process shows the G monomers form liquid crystals and crystalline phases in the dried state. A, C and U give amorphous phases although some crystalline structures could be detected for A. Mixed polymerization including G shows distinct amorphous and crystalline phases and the mixed products have low concentrations, in spite of a more balanced composition at pH 11.

The final part focused on the formation of the 2',3'-cyclic phosphate groups in conditions compatible with subsequent polymerization: drying of an alkaline solution at low temperature. Orthophosphate is rather unreactive in those conditions, but trimetaphosphate produces some amount of 2',3'-cyclic phosphate products by reaction on nucleoside and 2'-/3'-phosphate nucleotide, at alkaline pH, for a drying temperature of 60 °C. The yields are below < 1 %, but some dimers are already detected. The presence of other small organic molecules like urea and amino acids improve the yields and impact the regioselectivity of the reaction.

Amino acids phosphorylation and condensation are observed in the same conditions, which makes the parallel formation of RNA oligomers and peptides plausible.

List of Acronyms and Abbreviations

A Adenosine

ATP 5'-triphosphate adenosine

C Cytidine

DNA Deoxyribonucleic acid

EIC Extracted Ion Counts

eq equivalents

G Guanosine

G-2P 2'-monophosphate guanosine

G-3P 3'-monophosphate guanosine

G-5P 5'-monophosphate guanosine

G-5PP 5'-diphosphate guanosine

G-5PPP 5'-triphosphate guanosine

G-cP 2',3'-cyclic phosphate guanosine

G-noP G nucleoside

G₂-3P dimer G with a 3'-monophosphate ending

G₂-noP dimer G with no phosphate ending

G₃-3P trimer G with a 3'-monophosphate ending

G₃-noP trimer G with no phosphate ending

Gya gigayears ago

m/z Mass divided by charge

N-2/3/5P 2'/3'/5'-monophosphate nucleotide

N-cP 2',3'-cyclic phosphate nucleotide

P_i inorganic orthophosphate

RNA Ribonucleic acid

TMP Trimetaphosphate

U Uridine

Contents

Zusammenfassung	i
Abstract	iii
List of Acronyms and Abbreviations	v
1 Motivation	1
1.1 The synthesis of biomolecules from their building blocks	2
1.2 The RNA world hypothesis	3
1.3 Experimental constraints and prebiotic plausibility	4
1.4 Outline	4
2 RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides	7
2.1 Introduction	8
2.1.1 Prebiotic synthesis of RNA	8
2.1.2 2',3'-cyclic phosphate nucleotides oligomerization	8
2.1.3 Availability of 2',3'-cyclic phosphate nucleotides	10
2.2 Materials and methods	10
2.2.1 Reactants and standards	10
2.2.2 Sample preparation and experiment in dry phase	10
2.2.3 Experiment at a water-air interface	11
2.2.4 Ethanol precipitation	12
2.2.5 LC-MS instrument specifications and protocol	12
2.2.6 MS data analysis by custom-written LabVIEW program	13
2.2.7 Oligomers quantification	14
2.2.8 Determination of the phosphodiester linkage by Nuclease P1	16
2.2.9 ³¹ P NMR data acquisition	17
2.3 Results	17
2.3.1 G-cP dry polymerization: effect of the conditions	17
2.3.2 A-cP, U-cP, C-cP dry polymerization: effect of the nucleobase	19
2.3.3 Polymerization at water-air interface	22
2.3.4 Diphosphoester bond: 3'-5' versus 2'-5'	24
2.4 Discussion	24
Addendum 2	27
2.A Building of the thermal trap	27
2.B Spectral Browser workflow	28
2.C GA and GU mixed polymerization	29
2.D AUGC polymerization at water-air interface	30
2.E Enzymatic degradation results	31

3	Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature	33
3.1	Introduction	34
3.1.1	Wet-dry cycling	34
3.1.2	Liquid crystals	35
3.2	Materials and methods	35
3.2.1	Reactants and standards	35
3.2.2	Sample preparation and wet-dry-cycles protocol	35
3.2.3	LC-MS instrument specifications and protocol	36
3.2.4	MS data analysis by custom-written LabVIEW program	36
3.2.5	Quantification	38
	Monomers quantification	38
	Oligomers quantification	38
	Yield calculations	38
3.2.6	Microscopy observation	39
3.3	Results	39
3.3.1	Conditions of wet-dry cycling	39
3.3.2	Effect of wet-dry cycling on 2',3'-cyclic phosphate nucleotides	40
	Polymerization	40
	Dried phases	43
	Hydrolysis	44
3.3.3	Effect of wet-dry cycling on mixed polymerization	45
	Binary mix: adenosine and uridine	45
	Binary mix: guanosine and cytidine	47
	Quaternary mix	49
3.4	Discussion	51
	Addendum 3	53
3.A	Spectral Browser	53
3.B	Standards separation of the HPLC column	56
3.C	G polymerization on the pH range neutral to alkaline	57
4	Formation of 2',3'-cyclic phosphate nucleotides	59
4.1	Introduction	60
4.1.1	Prebiotic phosphorylation	60
4.1.2	Formation of 2',3'-cyclic phosphate nucleotide	61
	By direct phosphorylation	61
	Cyclization	62
4.1.3	Phosphorylating agent and their availabilities	62
4.2	Materials and methods	63
4.2.1	Reactants and standards	63
4.2.2	Sample preparation	63
	Direct phosphorylation	63
	By cyclization	63
	Wet-dry cycling	64
4.2.3	LC-MS instrument specifications	64
4.2.4	Standards and quantification	64
	Reactants quantification	64

Table of Contents

	Products quantification	65
	Valine-guanosine hybrid compound	66
	Yields calculations	66
4.3	Results	66
4.3.1	Direct phosphorylation of the nucleoside	67
	With orthophosphate	67
	With trimetaphosphate	69
	In presence of urea	70
	In presence of amino acids	71
4.3.2	Cyclization of the 2'- or 3'-phosphate nucleotide	73
4.3.3	Amino acid phosphorylation and polymerization	75
4.3.4	Formation of a covalent bond between amino acid and nucleotide . .	77
4.4	Discussion	78
	Addendum 4	81
4.A	Molecular structures	81
4.B	Effect of amino acids on phosphorylation of G nucleoside and cyclization of G monophosphate nucleotide	83
4.C	Phosphorylated amino acids and dipeptides	85
4.D	Amino-acid-guanosine hybrid compounds	86
5	Outlook	87
	Bibliography	91
	List of Figures	101
	List of Tables	102
	List of Publications	105
5.1	First Author publications	105
5.2	Contributed publications	105
	Acknowledgement	123

Chapter 1 | Motivation

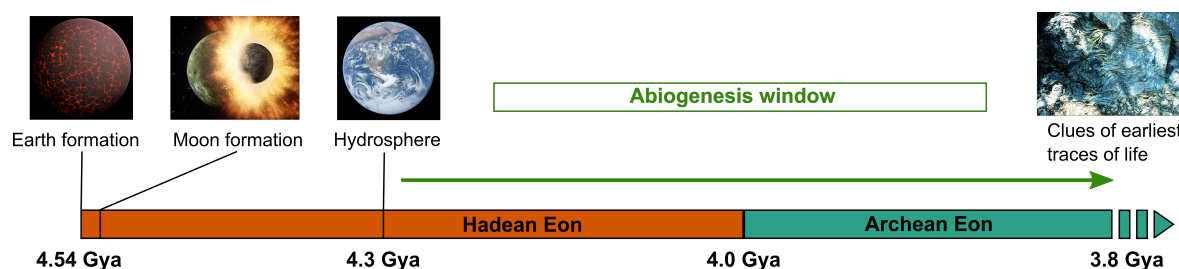


Figure 1.1: Abiogenesis window: Scheme of the chronology of the early Earth and plausible boundaries for the emergence of life on Earth. Images sources: [1–4]

The emergence of life on Earth remains an unsolved mystery to this day. All life forms show similar a composition and organization, prompting the idea that they all descend from a last universal common ancestor (LUCA), which is believed to have been similar in complexity to the modern bacteria [5, 6]. And, although some still advocate for the panspermia hypothesis [7], it is believed that life developed *de novo* on Earth [5]. As shown in Figure 1.1, the lower boundary of abiogenesis window is generally set around 4.3 Gya, after the moon-forming impact at 4.5 Gya and the subsequent cooling-down of the Earth until water could exist again on the planet [8]. The oldest traces of early life are mineral formations in rocks suggesting a biogenic origin, dated from 3.8 Gya [8, 9]. Knowing the origins of life would mean having a detailed account of the stages by which LUCA evolved during that abiogenesis window, going from simple chemicals at equilibrium in the conditions of the early Earth to highly-ordered and complex biomolecular systems of the extant life. However, several issues make that investigation rather complicated. The exact conditions in which life emerged are still unknown. Some studies explore the composition of the atmosphere [10], of the ocean [11] and the temperature [12] of the early Earth, as well as the potential environments [13, 14] and energy sources [15, 16] that existed at that time. Nonetheless, there are no rocks, fossils or samples left from that Eon to provide clues about the earliest stages of life's development. What's more, the experimental duration and limitations cannot match the realistic timescales and sample size of the early Earth. Therefore, the current studies have simpler, smaller and well-defined goals.

The present thesis will focus on three main axes:

1. Synthesis of the first biomolecules from their building blocks, salts and other small organic molecules.
2. Focus on one type of biopolymers in particular, RNA, as it is believed to have been able to form a first self-sustaining, self-replicating system.

3. The conditions of the reactions must be plausible in regards to what is expected of the conditions of the early Earth and the emergence of life.

1.1 The synthesis of biomolecules from their building blocks

In modern life, the synthesis of biomolecules is entirely guided and catalyzed by enzymes, which lower the activation barrier of the reaction and favor the formation of the correct molecules over the other potential products. Enzymes use the energy stored in the molecule of ATP to catalyze chemical reactions. However, neither enzymes nor ATP would have existed on the Hadean Earth before life arose. Then, there must have been one or several first abiotic pathways that produced biomolecules using other mechanisms and energy sources. But these synthesis are no longer used in extant life, although studies hypothesized that some characteristics of living organisms are traces of primitive life (such as the presence of sequences likely to form amyloids in older organisms [17] or the use of inorganic polyphosphate as energy storage [18, 19]). Therefore, apart from possible input from comets or other planets and moons, they can only be explored via experiments in the lab or simulations.

The abiotic synthesis of biomolecules has been explored very early, with the Urey-Miller [5, 20] experiment in 1952 and subsequent works proving that complex organic molecules such as amino acids [20–22] and nucleobases [23, 24] could be formed from electric discharges in a strongly reducing atmosphere (H_2 , NH_3 , CH_4 and additional compounds) in presence of liquid water. Although the conditions were later criticized as implausible [10], it was seen as a definitive proof that biomolecules could arise from inert chemical systems. Other experimental conditions have been proposed since [25], supported by experimental data, but there is no scientific consensus on a prebiotic synthesis of the building blocks of life.

Some of the most important biomolecules in extant life – proteins and nucleic acids – are polymers. Starting from their simplest monomers, amino acids and monophosphorylated nucleosides, the reaction leading to the formation of the polymer is a condensation, with the production of a molecule of water. This reaction is thermodynamically unfavorable, especially in an aqueous environment, whereas water is the solvent of life. Different strategies [26] have been explored to overcome that issue, either by using activated monomers (or activating them *in situ* by another reagent), very high temperatures (or moderately high temperatures associated with catalysts), or non-aqueous environments (dried conditions or using another solvent such as formamide) in which the monomer could be more reactive.

In this thesis, I will focus on the formation of short biopolymers starting from activated monomers which are not the building blocks of their biosynthesis by enzymes. I will also explore if those reactants can be produced under the same conditions as their polymerization. The experimental constraints, that were applied on these reactions in order to make the proposed pathway more plausible, are further discussed in the following section 1.3.

1.2 The RNA world hypothesis

1.2 The RNA world hypothesis

According to the central dogma of molecular biology, DNA is the molecule carrying genetic information. RNA is transcribed from DNA and RNA transfers the information to the ribosome to form the proteins, which take care of most of the cellular functions. DNA is the biopolymer storing the genetic information, including the coding exons used to produce proteins. But DNA needs proteins for its maintenance, replication, transcription etc. To that chicken and egg problem, a solution was proposed: RNA could be the molecule that started life on Earth [27]. Ekland and his co-workers [28] isolated RNA sequences capable of catalyzing the reaction of ligation, an activity similar to that of the extant DNA and RNA ligases. Many other ribozymes, enzymes made of RNA instead of proteins have been identified [29], and including in extant life [30, 31]. Thus, RNA carries the same genetic information as DNA, but can catalyze chemical reactions necessary for life. Furthermore, RNA (as DNA) possesses an inherent replication mechanism through hybridization between complementary strands.

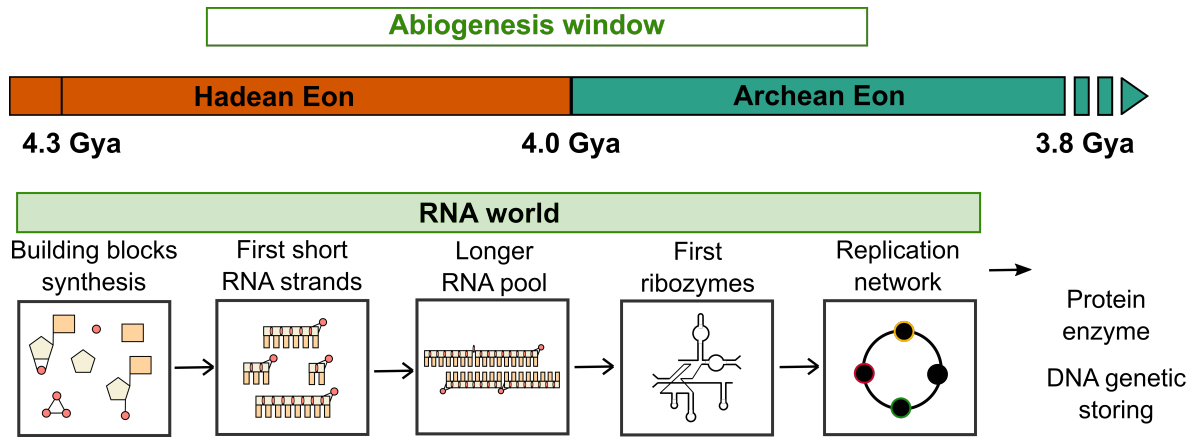


Figure 1.2: RNA world hypothesis: Description of the main steps for the development of the RNA world, leading to the emergence of modern life.

The *de novo* formation of the first RNA strands of the early Earth goes through several hallmarks, illustrated in Figure 1.2. The nucleotides are first synthesized from the chemicals available on the early Earth, then polymerize to form the first RNA strands. Through reactions of ligation and other mechanisms (such as adsorption on minerals for example, which would select longer strands against hydrolysis [32] or other selection pressures [33]), the pool of RNA trends toward longer sequences. Some of them are catalytically active and some of these ribozymes can favor RNA formation, ligation, replication and recombination [29]. Later, the chemical catalysis is overtaken by proteins and the more stable nucleic acid DNA becomes the carrier of genetic information.

This thesis will study the early steps of the RNA world, the synthesis of the first short RNA strands from their building blocks. Although this has been the focus of many studies in the field of origins of life (see section 2.1.1), the pathway from monomers to polymers is far from being established, especially if we take into account the constraints due to conditions existing on the early Earth.

1.3 Experimental constraints and prebiotic plausibility

There is no scientific consensus on what the exact conditions have been on the early Earth, and therefore on what exactly the conditions were for the prebiotic synthesis of biomolecules. The goal of this thesis is to explore the synthesis of RNA in a plausible scenario and that plausibility will be studied by following several experimental constraints.

The presence of water

Water plays an ambivalent role in the origins of life: it is a secondary product of the condensation reactions (by which most of the biopolymers form), making those reactions unfavorable. What's more, the presence of large quantities of water causes another major problem for the emergence of life, like dilution. Nonetheless, water is the universal solvent for the extant life on Earth, and its ubiquity might mean that it also played a role in its emergence [34]. Therefore, most of the significant prebiotic reactions are still attempted in aqueous solution, or at least in presence of water and the upper boundary for the habitability on Earth was established as the moment when the planet recovered a stable hydrosphere [8]. Furthermore, the presence of water is still the primary sign for life that astrobiologists track on exoplanets and moons. [35]

Thus, in this thesis, we will also explore the impact of water in our experiments.

Temperatures

Temperature plays a major role in both the kinetics of chemical reactions and the thermodynamic stability of the products. It is one of the major parameters studied for the synthesis of biomolecules. The temperatures that existed on the early Earth depended on different parameters that can only be extrapolated: the radiation from the young Sun, the composition of the atmosphere [36] etc. Through the analysis of minerals dated from the Archean Eon, liquid water is thought to have existed on Earth during the emergence of life [37], which, which still means a large range of temperature is possible. Some of the works on prebiotic chemistry rely on higher temperatures at the surface of the planet compared to today, while others on cold temperatures (some reactions and catalysis proposed happen only in ice). As of today, none of these scenarios can be completely excluded [38].

In these projects, we explored the emergence of RNA in a relatively cool environment or under moderate heating.

One pot-reactions

One of the important goals that have been pursued thorough the different projects was the uniformity of the conditions. Successive steps in the formation of RNA should happen in the same conditions. Reactions necessitating different environments would mean that the molecules have to be transferred from one to another, making the overall process less likely.

1.4 Outline

The topic explored by each chapter is represented on Figure 1.3.

Chapter 2 will explore a pathway for the non-enzymatic, non-templated polymerization of RNA from a mix of nucleotides. But instead of using the canonical 5'-phosphate nucleotide, the reactant here is the 2',3'-cyclic phosphate ribonucleotide (written as N-cP in the thesis).

1.4 Outline

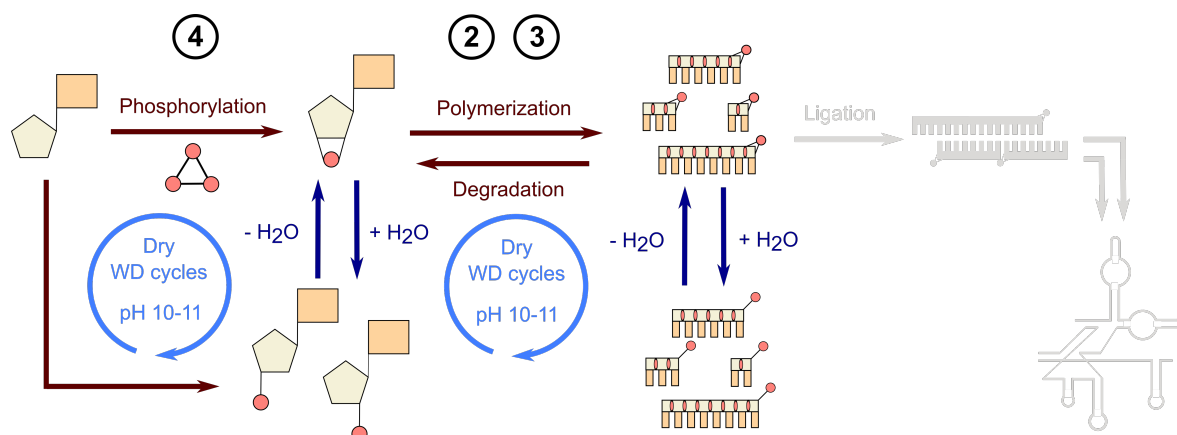


Figure 1.3: Chapters outline: Phosphorylation of the nucleoside to form a phosphorylated reactant and subsequent polymerization where conducted in drying or wet-dry-cycling experiment, at alkaline pH. Later step, ligation and the apparition of functional RNAs have been shown is similar conditions (the design of the ribozyme shown in the scheme correspond to the Sun Y ribozyme [29, 39]). The reaction of polymerization of 2',3'-cyclic phosphate nucleosides will be explored in Chapters 2 and 3. The phosphorylation and cyclization reactions leading to that reactant will be studied in Chapter 4.

It spontaneously polymerizes when dried from a mildly alkaline solution by moderate heating, in presence of cations such as Na^+ or K^+ . No additional activating agent or catalyst is required for the reaction to happen. Although the oligomerization cannot occur in aqueous solution beyond dimerization, the nucleotides can efficiently polymerize at a heated water-air interface.

In Chapter 3, the polymerization of cyclic phosphate nucleosides is optimized through wet-dry cycling at room temperature. The oligomers yields increase with the cycling, and so does the hydrolysis of the phosphate ring into unreactive 2'- or 3'-monophosphate nucleosides. G nucleotides, which polymerizes with the highest efficiency, is shown to form well-organized crystalline domains. In comparison, the less reactive A, C and U give only amorphous phases.

I will explore in Chapter 4 the formation of the 2',3'-cyclic phosphate ribonucleotide, in particular guanosine (G-cp), in conditions that do not deter the subsequent polymerization, by phosphorylation of nucleosides and cyclization of unreactive 2'- and 3'-monophosphate nucleotides.

Chapter 2

RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

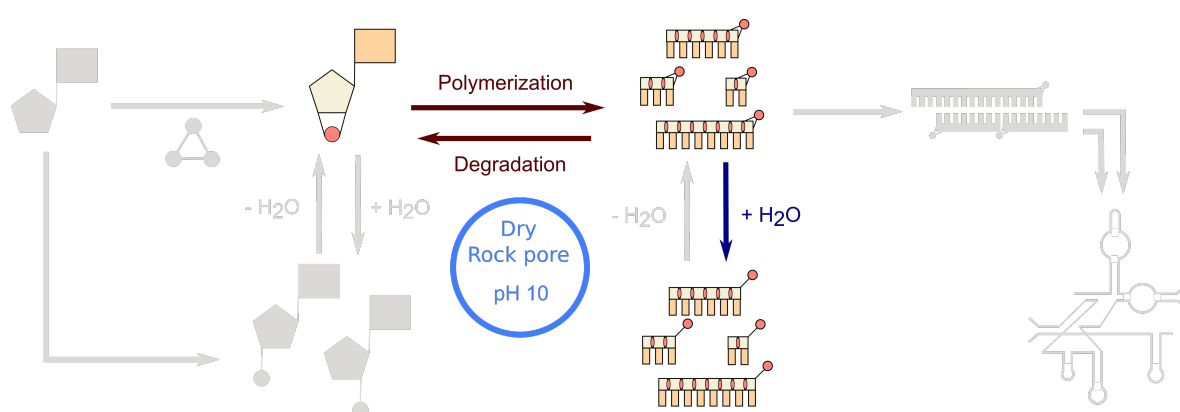


Figure 2.1: Chapter summary: Formation of short RNA strands by polymerization of the 2',3'-cyclic phosphate nucleotides by heat-drying of an alkaline solution either in a vial or at a water air interface, in a setting mimicking a rock pore.

Chapter summary

The *de novo* polymerization RNA from phosphate ribonucleotides, without the catalysis of ribozymes or the contribution of other complex organic molecules (such as a pre-synthesized template) is a challenging step. The chapter presents a possible pathway for the formation of short RNA strands, through the transphosphoesterification of 2',3'-cyclic phosphate ribonucleotides (N-cP). Oligomers can be formed by drying from an alkaline solution by moderate heating and do not require further activation. The analysis of the oligomers (2 to 10mers) is conducted by reversed-phase liquid chromatography, associated to negative-mode electrospray ionization mass spectrometry. However, those products are largely dominated by G, as A, C and U show a lesser oligomerization efficiency.

This project was developed with two colleagues from my lab: Sreekar Wunnava and Dr. Avinash Vicholous Dass.

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

2.1 Introduction

Efficient polymerization of a mixture of nucleotides in conditions compatible with the early Earth and in the presence of water is yet to be achieved. The canonical 5'-triphosphate and monophosphate nucleotides cannot polymerize without an enzymatic catalysis. Prebiotic synthesis of nucleic acids relies on activated nucleotides that are more reactive toward polymerization, such as 2',3'-cyclic phosphate nucleotides, molecules that can be obtained from the phosphorylation of nucleosides and the hydrolysis of already synthesized RNA strands.

2.1.1 Prebiotic synthesis of RNA

Different strategies have been deployed to form the first RNA strands, the main path studied being the activation of the 5'-phosphate group, in general by adding an imidazole derivative. It acts as a leaving group to make polymerization more efficient [32, 40–42]. Mineral surfaces were used to improve reaction yields and protect longer products from hydrolyzing by adsorption of phosphate groups on the surface [32, 43]. But the formation of the activated nucleotides *in situ*, in conditions compatible with further polymerization, is also a complex issue. This last step was achieved by using EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) as coupling reagent [44, 45]. Nonetheless, EDC is not considered to be prebiotically plausible, although other carbodiimides could have been formed by the isomerization of cyanamide on the early Earth [46, 47].

Šponer's team and collaborators worked on another reaction for the formation of RNA, through the oligomerization of 3',5'-cyclic monophosphate nucleotides in acidic and dry conditions [48–51]. Polymerization products of length up to 40mers could be detected in these conditions, but only for the nucleobase guanosine. No product beyond dimers could be observed for A, C, U. What's more, the 3',5'-cyclic phosphate nucleotides are very difficult to form from aqueous solutions [52, 53].

One common problem of those pathways is the formation of the activated monomers, especially in conditions compatible with the subsequent polymerization. The reaction also requires precise conditions and chemical composition, which are not necessarily prebiotically plausible.

2.1.2 2',3'-cyclic phosphate nucleotides oligomerization

As an alternative the 5'-phosphate activation, other studies focused on another reactant, for which the phosphate group is not attached to the canonical 5'-OH position, but on the 2'- and 3'-OH. The scheme of the polymerization mechanism can be seen in Figure 2.2: nucleophilic substitution (S_N2) occurs with the nucleophilic attack of the 5'-OH of the phosphate group of another nucleotide and the cleavage of either the 2'-O-P or the 3'-O-P bond. Therefore the reaction can yield either a canonical 3'-5' linkage or the unnatural 2'-5' one.

Orgel and his co-workers reported conditions for the oligomerization of 2',3'-cyclic monophosphate adenosine (A-cP) [54–56]. In 1971 [56], they observed the oligomerization of A-cP on a poly U template and in presence in Mg^{2+} and amine derivatives (spermine, spermidine,

2.1 Introduction

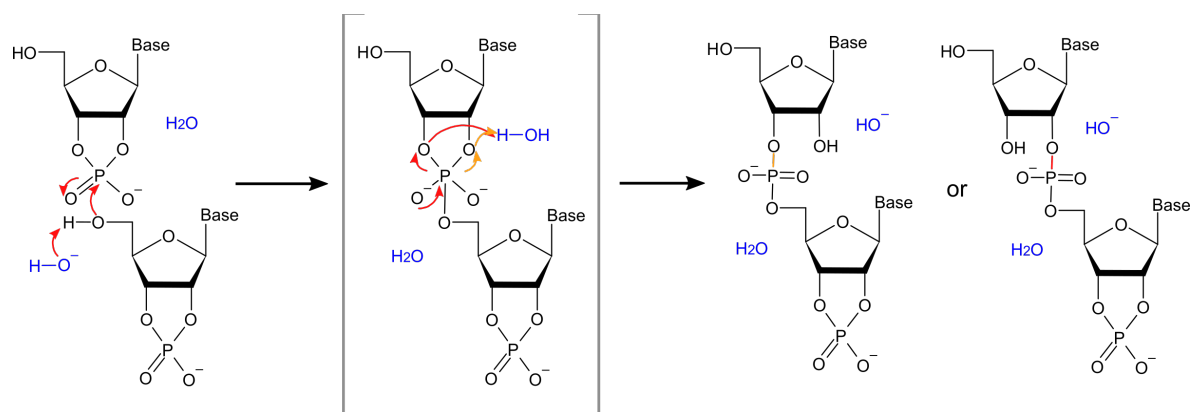


Figure 2.2: 2',3'-cyclic phosphate nucleotide (N-cP) polymerization mechanism: by transphosphoesterification at alkaline pH. The reaction proceeds by attack of the 5'-OH group of a monomer on the phosphorus center of another one. The phosphate ring can be opened on either 2' or 3', causing the formation of 3'-5' or 2'-5' linkage respectively. No water molecule are produced by this reaction.

1,4-diaminobutane, 1,2-diaminoethane...) in aqueous solution, during up to 50 days, for pH between 5.5 and 8.5. The main product detected was adenosine dinucleotides, with yields of 17 % from the starting material, obtained at pH 8.5, 25 days at 0 °C, in presence of 20 equivalents of ethylenediamine. Using enzymatic degradation that should cleave only the canonical 3'-5' linkage, they estimated that 97 % of the dinucleotides were 2'-5' linked. In 1973 [54], they worked on the non-templated polymerization of adenosine in dry conditions for 1-2 days at room temperature or heated up to 85 °C, in presence of amines compounds as catalysts. Their findings suggest that the self-polymerization of A-cP might be subjected to general base catalysis. 1,2-diaminoethane (5 equivalents) performs quite efficiently at room temperature, leading to a polymerization yield of 60 % (36 % for the dimers alone, the longer oligomers were not measured separately). Other catalysts such as imidazole needed higher temperatures for the oligomerization to work. Here linkages were found to be mostly 3'-5' (around 60 %). In a subsequent article [55], they focused on the analysis of the longer oligonucleotide products obtained for the non-templated oligomerization of A-cP in presence of 1,2-diaminoethane at alkaline pH, under drying conditions for several days at room temperature. After 3 days, they detected 13mers, which concentration corresponded to 0.67 % of the material.

Tapeiro and Nagyvary [57] also observed the oligomerization of C-cP triethyl ammonium salt under dry conditions at 138 °C for 48 h, with products of length up to hexamers. The yields for the dimers, trimers, tetramers and pentamers were respectively 16 %, 10 %, 4 % and 2 %. The proportion of 2'-5' linkages was estimated to approximately 50 % for the dimers and 70 % for the trimers.

The thermal phosphorylation of uridine nucleoside in presence of urea at 100 °C in dry nitrogen atmosphere for several days yielded significant amount of U-cP and polymeric material was also detected [58]. 50 % to 60 % of the internucleotide linkages were found to be the canonical 3'-5' bond.

Polymerization using N-cPs had been explored mostly with A and, and to a lesser extent, U

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

and C, but never with G. For A, U and C, the non-templated polymerization seems to require rather high temperatures and significant concentrations of amine derivatives as catalysts.

2.1.3 Availability of 2',3'-cyclic phosphate nucleotides

2',3'-cyclic nucleotide are the products of alkaline hydrolysis of RNA, through the nucleophilic attack of the 2'-OH group on the phosphodiester bond [59], as well as the products of enzymatic degradation by both protein enzyme [60], and by known ribozymes [30, 31]. Although nowadays they are not used as building blocks for the synthesis of RNA, N-cP are still present and used by modern life [61].

2',3'-cyclic monophosphate ribonucleotides are commonly reported products of prebiotic phosphorylation conditions, although often not the main focus of the study. High amounts of U-cP were produced by thermal phosphorylation of the nucleoside at 100 °C in urea [58]. The phosphorylating agent TMP produces also some amount of cyclic ribonucleotides [62]. The phosphorylation by diaminophosphate (DAP) developed by Krishnamurthy and his coworkers gives also very high yield of N-cP [44, 63, 64].

In the direct synthesis of the pyrimidine ribonucleotide by Sutherland et al., the final product is the beta-ribocytidine-2'3'-cyclic phosphate [65].

2.2 Materials and methods

2.2.1 Reactants and standards

2',3'-cyclic nucleotide monophosphates (N-cP): G-cP (monosodium salt), C-cP (monosodium salt), A-cP (monosodium salt), U-cP (monosodium salt). G-cP and U-cP were purchased from Biolog Life Science Institute GmbH & Co. KG. cAMP and cCMP were purchased from Sigma Aldrich. KOH, NaOH, HCl, MgCl₂, KCl and imidazole were purchased from Carl Roth. Oligomer standards for HPLC-Mass Spectrometer calibration G 2 to 10mers with a 3'-phosphate ending (and -OH at the 5' end) were purchased from biomers.net GmbH. To analyze the nature of the linkages, the enzyme Nuclease P1 produced from *Penicillium citrinum* was purchased from Sigma Aldrich. The 2'- 5' linked 5mers were purchased from biomers.net GmbH.

2.2.2 Sample preparation and experiment in dry phase

From stock solutions of 200 mM N-cP, 20 mM of each N-cP is prepared and the volume is adjusted to 100 µL. The pH adjustment is set using KOH and HCl. When reactions are carried out in the presence of cations K⁺, then KOH was used to adjust the pH, NaOH for Na⁺, LiOH for Li⁺. For Mg²⁺, KOH was used. The 100 µL sample is then put in 1.5 mL Eppendorf vials, which are then heated, open-lidded, in a heat block for 18 hours, unless the time is specifically mentioned. After drying the sample, the dried pellet is rehydrated with 100 µL of nuclease-free water.

2.2 Materials and methods

2.2.3 Experiment at a water-air interface

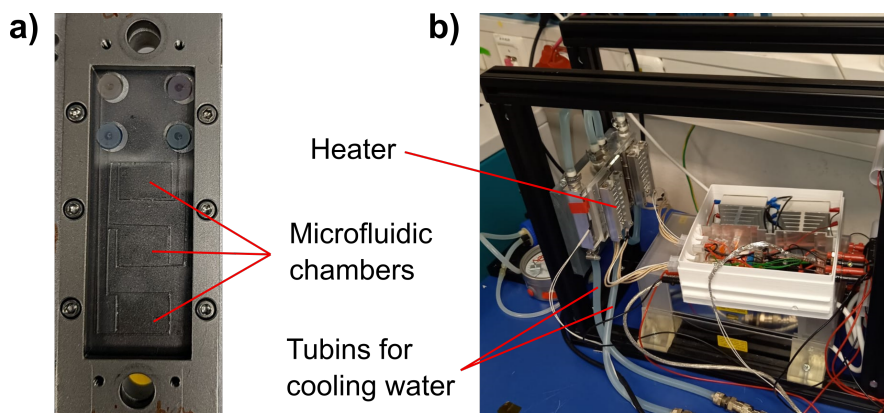


Figure 2.3: Thermal trap front view and installation: **a)** Front view of the microfluidic chambers, seen through the sapphire 5 in Figure 2.A.1. **b)** Picture of the full set up of the thermal trap.

The samples for the reaction at water-air interface proceeded without a microfluidic chamber (see Figure 2.3 **a)**) only partially filled with the reaction mix, subjected to a temperature gradient. The microfluidic device is referred as a "Thermal trap" in this chapter. The solution that reacts in the thermal trap is prepared as in section 2.2.2. Unless otherwise specified, the experiment inside the thermal trap lasts 18 hours, as in the dried state.

The thermal trap is assembled layer by layer, as indicated in Figure 2.A.1. Once the screws on the upper metal plate 6 are tightened, the chambers are closed so that the solution inside cannot leak outside, and there is no longer gas exchange with the external atmosphere: the water vapor produced by the heated remains inside the chamber. The solution is injected into the thermal trap using tubings attached to syringes and to the back of sapphire plate 2 in Figure 2.A.1 and flown in the three chambers cut in the Teflon layer 4. They all contain the same solution and run at the same time. The chambers are not completely filled, so there is a water-air interface; they contain roughly 30 μL each. The temperature of the metal support 1 is then cooled to 20 °C and the temperature of the heater attached to the top of 5 and 6 is set at 40, 60 or 80 °C. Inside the trap, in the three chambers, the reaction mixture is subjected to a temperature gradient. The exact boundary temperatures within the chamber were calculated by measuring the temperatures at the outer surface of the sapphires 2 and 5 and simulating the evolution of temperature between the different layers based on their thickness and the conductivities of the material. When the temperature of a water bath is set at 20 °C and the temperature of a heater at 60 °C, the extreme temperatures in the chambers will be 38 °C on the cold side and 54 °C on the hot side (see Figure 2.15 **a)** scheme). The aqueous solution is in contact with a Teflon layer on the cold side, a hydrophobic layer that will cause the water condensing there to form thicker droplets before they fall back into the bulk solution. On the hot side, the solution is in contact with the 2 mm-thick sapphire and monomers and polymers accumulate at the interface between water and air [66, 67]. The evaporation on the hot side and the recondensation on the cold side are what drives the wet-dry cycles in the trap.

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

Once the experiment is finished, the heater is stopped and the samples in the three chambers are extracted once the trap has cooled down. The solution is collected using syringes and the contents of the chambers are added together. The empty chambers are filled completely with nuclease-free water, in order to dissolve the potential remaining dried material, and this cleaning water is then added to the sample, to prepare it for analysis as in the sections below.

2.2.4 Ethanol precipitation

The precipitation is used to remove excess monomers which would otherwise saturate the HPLC column, yielding a robust method for the complex oligonucleotides mixture. 2.2.7 For the volume of 100 μ L, 2 μ L of 10 mg/mL of glycogen are added. Then 10 μ L of 5 M ammonium acetate are added. 250 μ L of cold 100 % ethanol are added and the sample is kept overnight at 4 °C. The sample is then centrifuged at 4 °C for 30 min at 21 000 rpm. The supernatant is discarded and 100 μ L of cold 70 % ethanol are added to the pellet. The sample is gently mixed by tapping on the vial and then centrifuged again at 4 °C for 30 min at 21 000 rpm. The supernatant is again discarded and the pellet is left to dried until most of the alcohol is removed. Then the pellet is rehydrated with 20 μ L nuclease-free water; the sample is homogenized by vortexing and 18 μ L of the volume are injected into the HPLC for analysis.

2.2.5 LC-MS instrument specifications and protocol

Due to the propensity of purines to form non covalent aggregates in mass spectrometry detection [68], a combination of HPLC and ESI-TOF techniques were used for detection and quantification of oligonucleotides. The non covalent stacked oligomers are discriminated by the HPLC retention times under the denaturing HPLC conditions [69–71]. The denaturing conditions of the HPLC column at 60 °C efficiently resolved covalently-bound oligo G without signs of aggregation, as shown in Figure 2.5. A definitive confirmation of the presence of diphosphoester bonds in the reacted mix was provided by Phosphorus NMR (see section 2.2.9).

The measurements were performed on a high performance liquid chromatography (HPLC Agilent 1260 Infinity II bioinert) with a G7115A 1260 Infinity II diode array detector and coupled to an electrospray ionization time-of-flight mass spectrometer (Agilent 6230B with dual AJS ESI). The column used was an Agilent Advance Oligonucleotide C18 Column (4.6 x 150 mm 2.7 μ m) heated at 60 °C with a pressure limit of 600 bar.

The oligomers were separated according to their length by ion-pairing reversed-phase HPLC. The eluent consisted of a mix of water (Bottle A) and methanol (Bottle B: 50 % water, 50 % methanol) containing each 8 mM trimethylamine (TEA) and 200 mM hexafluoroisopropanol (HFIP), with a gradient elution at a flow of 1 mL/min. The method started with 1 % of B for 5 min, followed by a gradient, increasing from 1 % to 30 % B over 22.5 min and then to 40 % for 15 min. Then, the column was flushed with 100 % B for 5 min before being returned to 1 % for 6 min, to re-equilibrate the column.

2.2 Materials and methods

Detection of eluted compounds and quantification of monomers is achieved by using a Diode Array Detector (DAD) WR (wavelength used: 260 nm).

The detection of eluted mononucleotides and quantification of eluted oligonucleotides are achieved using ESI-TOF in negative mode (employing specific source parameters: Gas temperature: 325 °C, Drying gas flow: 13 L/min, Sheath gas temperature: 400 °C, Sheath gas flow: 12 L/min, VCap: 3500 V, Nozzle Voltage: 2000 V). Reference masses are run in parallel to the sample run using standard reference and tuning mix recommended by Agilent (product number G1969-85000).

2.2.6 MS data analysis by custom-written LabVIEW program

The data files obtained from the mass spectrometer are converted to another format and then analyzed using a custom-written LabVIEW developed by my supervisor, Dieter Braun. To calculate the concentration of each product oligomer, the Spectral Browser takes the area of the peaks in the spectrum corresponding to the isotopic distribution(s) of the molecule. The concentration value is obtained from the ion counts of the isotopic distribution by comparison with a solution of standard oligomers of known concentrations.

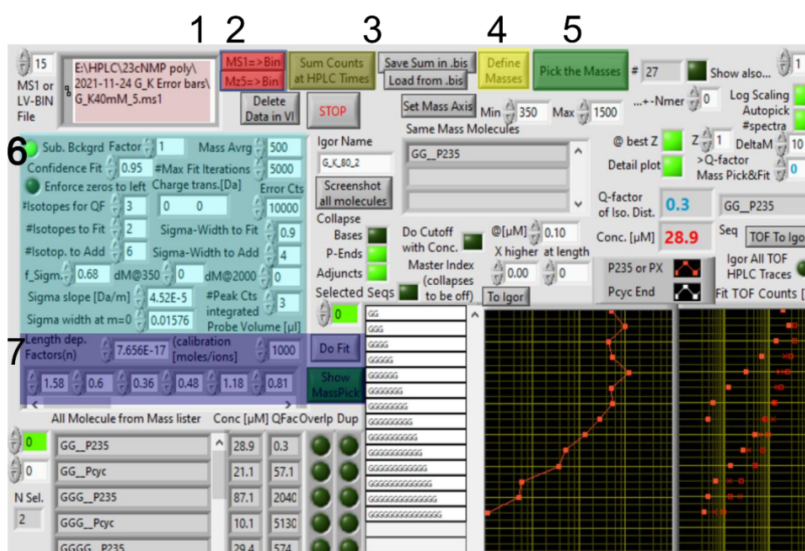


Figure 2.4: Representative screenshot of the workflow in the Spectral Browser

The Figure 2.4 shows, step by step (numbered 1 to 7), how the oligomers are quantified in the Spectral Browser, step by step. The MS1 data files are generated by a freely available program called MSConvert from ProteoWizard and dropped into 1. The MS1 file is converted into a binary file that is read by the Spectral Browser by function 2. The counts at defined cursor positions are accumulated by function 3. Oligomeric masses of 2-10mers with the cyclic phosphate or linear phosphate ends are defined by 4. Those masses are then processed, plotted and displayed in the program by clicking on 5. The 'Do Fit' function in the program is applied to resolve overlaps between distinct molecules' isotopic distributions and

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

the concentrations are calculated based on the fit qualities defined in 6. The concentrations of the oligomers are then calculated from the calibration coefficients defined in 7 (see how the coefficients were obtained in section 2.2.7).

2.2.7 Oligomers quantification

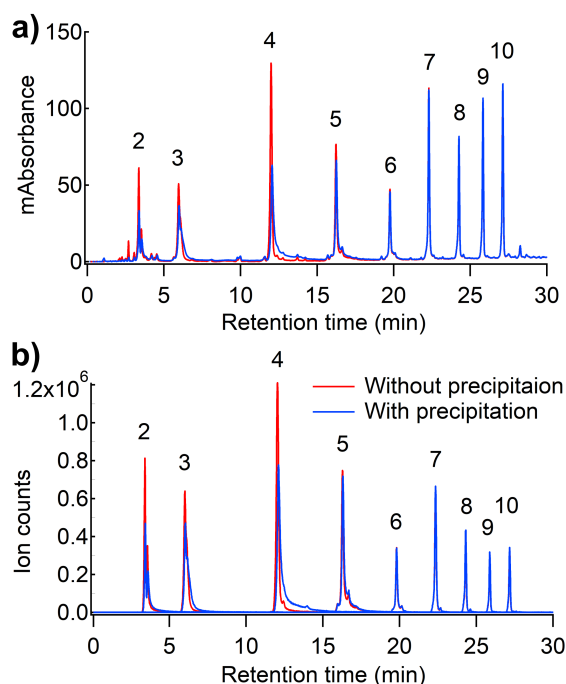


Figure 2.5: G oligomers standards 2 to 10mers with a 3'-P ending, 50 pmol each, both before and after an ethanol precipitation step. **a)** Efficient separation is observed with the HPLC-DAD (Diode Array Detector) at 260 nm. They are simultaneously injected into the ESI-MS after their elution from the HPLC column. **b)** The eluted peaks of each n-mer are then detected based on their masses and is confirmed by the Extracted Ion Counts (EIC) peaks. The y-axis displays the ion counts plotted against the corresponding HPLC retention times (x-axis) of the n-mers.

Oligomers from 2- to 15-mers were detected by HPLC-MS for G-cP oligomerization. For the quantification, only 2- to 10-mers were considered throughout the study.

For quantification, the HPLC retention times of the oligomer standards of G were first optimized on a RP C-18 HPLC column coupled to ESI-TOF. Figure 2.5 a) shows the HPLC chromatograms of 2- to 10-mers for oligo G standards with 5'-OH and 3'-phosphate endings (in presence of 1 eq of KCl) with their respective retention times. The ion counts of the n-mer with their HPLC retention times are shown in Figure 2.5 b). The high efficiency of the post-polymerization ethanol precipitation protocol and its negligible influence on the results were confirmed by comparing the ion counts of the oligomers detected before and after polymerization.

2.2 Materials and methods

All concentrations reported in the chapter are calculated for a volume of 100 μL . The total concentration of each n-mer is the sum of the oligomers of the same length ending with both the linear phosphate (2' or 3')-P and the cyclic phosphate (2',3')-cP on the terminus. Oligomers of the same length but with different endings are well discriminated by HPLC as oligomers with -cP endings are eluted before oligomers of the same length with -P endings (see 2.6). In most of the experiments in this chapter, the majority of the oligonucleotides consisted of -P endings.

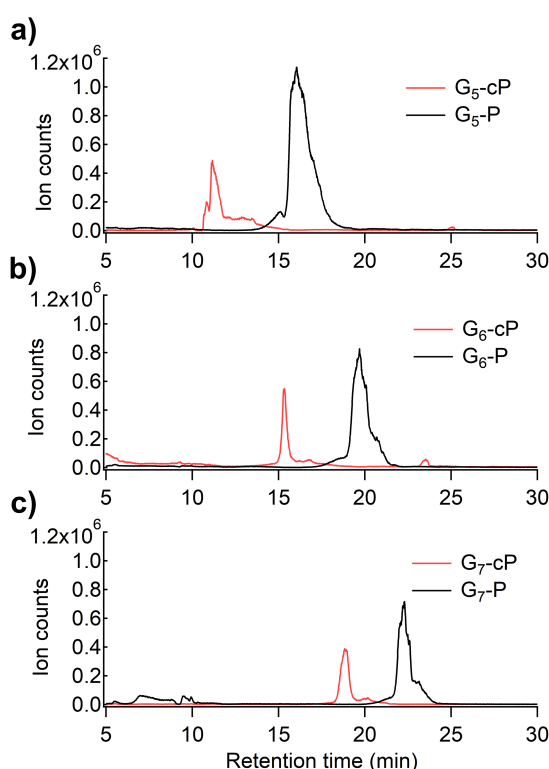


Figure 2.6: Comparison of the retention times of the -cP versus the -P ending oligomers: Products of the polymerization of G-cP from a solution of 20 mM at pH 10 dried at 40 °C for 18 hours in presence of 1 eq KCl: **a)** 5mers, **b)** 6mers, **c)** 7mers.

The calibration curves and coefficients (see the values in Table 2.1) were established for a range of quantities 1-500 pmol, obtained by injecting different volumes (1, 2, 5, 10, 25, 50 μL) of calibration mixes of concentrations 1 and 10 μL . The concentration of an oligomer was calculated using the same length-dependent coefficient, regardless of the composition and the ending (-cP or -P). The error bars were evaluated for the whole chapter based on 5 replicates performed under the same optimal G-cP polymerization conditions (see Figure 2.7). Products longer than 10mers were not quantified, but oligomers were clearly detected in the mass spectrometry analysis and the values shown here for 2-10mer, using moles/ions values in Table 2.1. The error bars are smaller than the markers, with a mean standard deviation of 2.95 μM for independent runs of the experiment. Therefore, the error bars will not be indicated in the plot of the section 2.3, as they would appear insignificant on the log scale.

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

Table 2.1: Coefficient values for the calibration of the standards $G_{2-10}-3'P$ Coefficient values by linear fit of the $EIC=f(\text{moles}) \pm$ one standard deviation, over the quantity range 0.5 to 50 pmol

Oligo length	Slope un-precipitated	Slope precipitated
2	13656 ± 259	10288 ± 87.7
3	21307 ± 453	20768 ± 276
4	38472 ± 1780	41287 ± 307
5	31293 ± 896	32831 ± 292
6	12173 ± 408	12407 ± 242
7	15117 ± 444	15601 ± 313
8	9866.8 ± 403	10429 ± 275
9	9806 ± 400	10404 ± 270
10	11048 ± 509	11551 ± 366

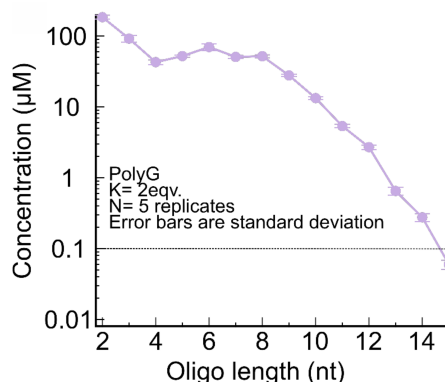


Figure 2.7: Error bars calculated as standard deviation for 5 replicates The 5 replicates were conducted in the same conditions: a 20 mM G-cP solution at pH 10 was dried with 2 eq K^+ at 40 °C for 18 hours. The dotted horizontal line indicates the quantification limit.

2.2.8 Determination of the phosphodiester linkage by Nuclease P1

Nuclease P1 digestion assay was conducted to assess the type of the phosphodiester linkages formed during the experiments. 3'-5' phosphodiester bonds are susceptible to Nuclease P1 while the 2'-5' bonds are not. This had previously been used to determine the 2'-5' linkages by ligation of the Peach Latent Mosaic Viroid [72]. For the assay, 100 μ L of 20 mM G-cP at pH 10 is oligomerized as in by drying at 40 °C for 24 hours. The samples are then precipitated (see section 2.2.2) and re-dissolved to contain 10 μ M of phosphodiester bonds.

Nuclease P1 from *Penicillium citrinum* (Sigma-Aldrich N8630) is dissolved in P1 storage buffer (25 mM Tris-HCL, 50 mM NaCl, 1 mM $ZnCl_2$, 50 % Glycerol, pH 7.2 at 25 °C) at a concentration below 0.5 U/ μ L. The digestion reaction is done in reaction buffer P1 (50 mM sodium acetate, pH 5.5) with oligomerized G-cP sample concentration of 1 μ M phosphodiester bond equivalents and 0.5 U enzyme in a volume of 100 μ L. For the control, 1 μ L of P1 storage buffer without the Nuclease P1 is used. The samples are incubated at 37 °C for 10, 30 or 240 min following which the enzyme is inactivated by heating at 75 °C for 10 min. Phenol-chloroform-isoamyl

2.3 Results

alcohol (Carl-Roth A156.3) extraction carried out by adding an equal volume to the sample followed by vigorous mixing and centrifugation at 15 000 rpm. The top aqueous phase is carefully removed and injected in the HPLC-MS.

For the HPLC-MS analysis, the EIC (Extracted Ion Counts) for the oligomers (whether -2',3'cP, -2'P, -3'P or no phosphate ends) are used and the concentration equivalent of the phosphodiester linkages is calculated for each sample and normalized to the estimated phosphodiester linkages remaining in the control experiment without enzymatic degradation (Figure 2.E.1).

2.2.9 ^{31}P NMR data acquisition

A sample of G-cP polymerized at pH 10 and dried at 40 °C for 18 h is rehydrated in a mix of 90 % water and 10 % deuterium oxide D_2O was measured on Bruker Avance III spectrometer and its ^{31}P NMR spectrum was recorded at 25 °C, for a Larmor frequency of 162 MHz. Sodium trimethylsilylpropane sulfonate (DSS) is used as an internal standard.

Several peaks corresponding to different phosphate groups were identified in Figure 2.8 **a**). The main peak between 21 and 20 ppm is attributed to the cyclic phosphate group of the monomer and the minor peak with a slightly lower chemical shift corresponds to the cyclic phosphate ending of oligomers. The linear phosphate groups 2'-P and 3'-P due to the hydrolysis of the cyclic phosphate give signals at 5-6 ppm [73]. Here again, the larger peaks are attributed to the linear phosphate group of the monomers and the smaller ones to the phosphate endings of the oligomers. The signals with a chemical shift between 0 and -1 ppm correspond to diphosphoesters [74].

2.3 Results

First, the physical-chemical parameters impacting the polymerization were explored, using G-cP as a model. Then, the reactivities toward polymerization of the four canonical nucleobases were compared: G is significantly more efficient than A, C and U. A possible explanation might be that G can self-organize in a supramolecular assembly before drying, which will be confirmed later in the chapter 3. Non-canonical 2'-5' and canonical 3'-5' linkages were detected in the newly formed oligomers.

2.3.1 G-cP dry polymerization: effect of the conditions

A 20 mM aqueous solution of sodium salt of G-cP was dried for 18 hours at 40 °C in the presence of 40 mM KCl. The presence of covalently bound oligomers is confirmed in Figure 2.8 by both **a**) analysis on HPLC-MS (see protocol in section 2.2.5) and **b**) detection by ^{31}P NMR. On the NMR spectrum, the peaks at 0/-1 ppm were attributed to diphosphoester groups. Figure 2.8 **a**) shows the Extracted Ion Counts (EIC) peaks corresponding to the mass divided by charge (m/z) of G oligomers with a linear phosphate -P ending. This phosphate group is due to the hydrolysis of the cyclic phosphate group. The method used for the HPLC does not discriminate between a -2'P or a -3'P, but the ^{31}P NMR spectrum in **b**) confirmed both

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

endings are present in the sample. Although less abundant, oligomers with cyclic phosphate -cP ending were also detected on HPLC-MS and NMR (minor peak around 20 ppm).

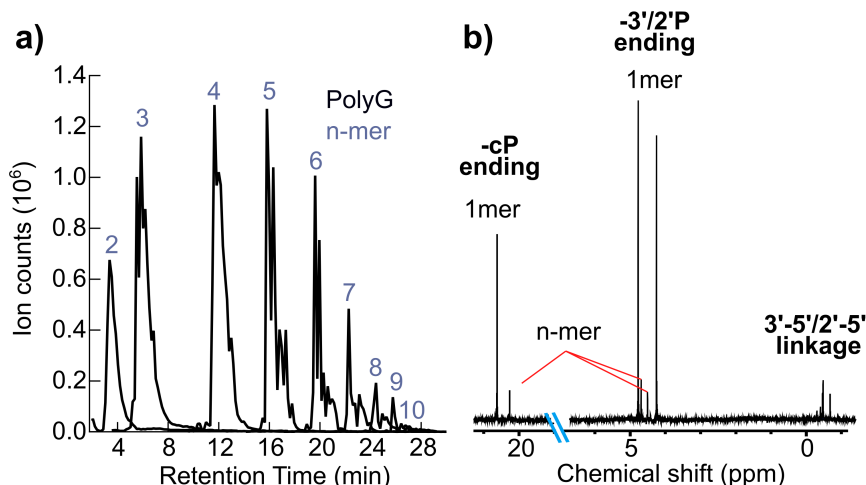


Figure 2.8: G polymerization: A 20 mM G-cP solution at pH 10 was dried with 2 eq K⁺ at 40 °C for 18 hours, under ambient pressure. **a)** Reconstructed chromatogram of G oligomers standards with -3'P ending. The ion counts from mass spectrometer are plotted against the retention time in HPLC for each length. **d)** ³¹P proton-decoupled NMR spectrum (10 % D₂O) of polymerized G sample: the signals corresponding to phosphodiester linkages (3'-5' and 2'-5') are between -0.8 to -1.1 ppm

Figure 2.9 compares the effect of pH on the length and concentrations of the oligomers, either in the presence of **a)** K⁺ cations or **b)** imidazole. Imidazole was often used as a catalyst for the polymerization by Orgel and his coworkers [54–56], as well as a nucleotide activation agent in other works [32, 40–42]. The optimal pH was found to be 10 for both conditions. Interestingly, the addition of imidazole seemed to have only a minor impact on the results, mostly at pH 7. For pH 10, no significant enhancement of the length and concentration of the products was observed compared to oligomerization in the presence of K⁺. Therefore, the efficiency of G-cP polymerization was further explored without the presence of a catalyst such as imidazole.

The impact of time and temperature on the reaction was studied in Figure 2.10 **a)**: the concentration of the products, especially of length > 3, increases over time until 24 hours. The aqueous control, in which the monomers G-cP were left to react for 24 hours at 40 °C in aqueous solution, proves that the drying step is essential for polymerization, as no polymers were detected. The efficiency of the reaction increases with decreasing drying temperature (see Figure 2.10 **b)**), the optimal one being 40 °C. However, one thing to note is that the 30 °C samples never fully dried for the duration of the experiment (18 h). The lower efficiency of the reaction at higher temperatures might be due to the increase of the hydrolysis of the phosphate ring.

2.3 Results

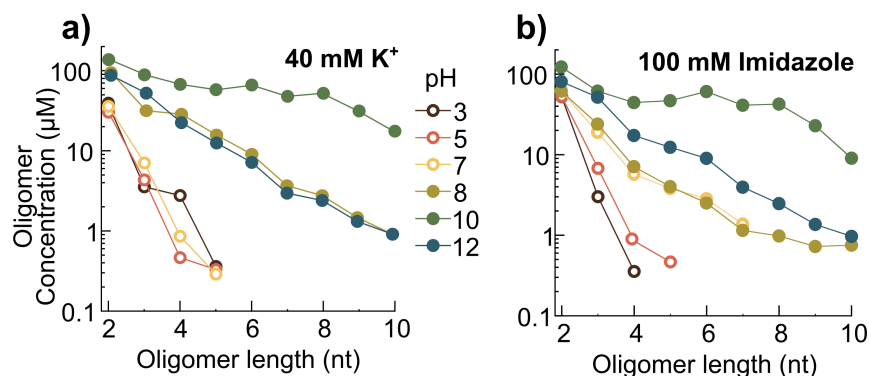


Figure 2.9: Effect of pH and the presence of a catalyst over the G-cP polymerization: A 20 mM G-cP solution was heat-dried with 2 eq K^+ at 40 °C for 18 hours, under ambient pressure. **a)** Screening over the pH range of 3-12 for each n-mer. Concentrations shown (for a volume of 100 μ L) are a sum of terminal 2',3'-cyclic phosphate and linear 2'- or 3'-phosphate. Polymerization was optimal at pH 10. **b)** Screening over a range of pH 3-12 in presence of 5 eq of imidazole under similar conditions: no significant increase in products concentration was observed, especially at pH 10, compared to the results shown in **a)**.

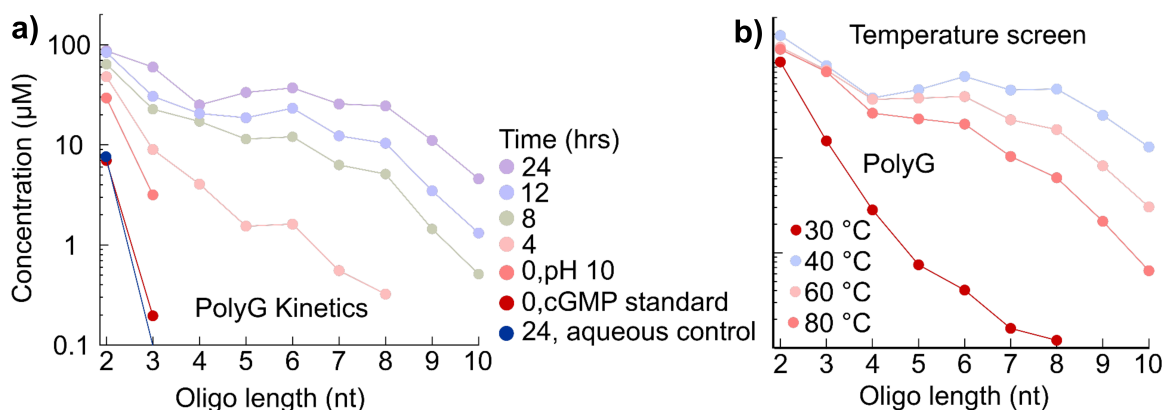


Figure 2.10: Kinetic and effect of the temperature over G-cP polymerization: Results of G-cP polymerization from a 20 mM solution at pH 10 in presence of 2 eq KCl. **a)** The solution was dried at 40 °C and different samples were left to react for different times: 0, 4, 8, 12 and 24 h. A aqueous control was conducted in the same conditions as the 24 h dried sample, except that the lid of the vial containing the control solution was closed, thus there was no drying. **b)** The solution was dried for 18 hours at different temperatures: 30, 40, 60 and 80 °C. The 30 °C sample never reach the fully dried state, which might have impacted the results.

2.3.2 A-cP, U-cP, C-cP dry polymerization: effect of the nucleobase

The polymerization tendencies of A-cP, U-cP and C-cP were also tested under the same heat-drying conditions: these monomers do not polymerize to the same length and concentrations as G-cP. Figure 2.11 shows that the polymerization trends decrease in the order G-cP > U-cP > A-cP > C-cP. The dominance of G polymerization prompted us to investigate in Figure 2.12 the copolymerization of these moderately reactive mononucleotides under the influence of the very reactive G-cP. A mixture of two or four different monomers was capable of gen-

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

erating mixed-sequence oligomers, but the majority of those oligomers were rich in G. We investigated if the oligomerization of G and C mixtures could reach levels where hybridization between strands could be possible.

Thus, a binary mixture of G-cP and C-cP of 20 mM each was oligomerized under heat-drying conditions at 40 °C in the presence of 80 mM KCl. Mixed GC oligomers were detected. The detailed sequence composition for mixed GC polymerization is seen in Figure 2.12 (the detailed sequence compositions of mixed GA and GU polymerization can be seen in Figure 2.C.1) and shows that the G homopolymers remain the main product of the reaction. Up to two C were incorporated into 4mers, one C was incorporated into 5mers. Similar observations can be made for the binary mixtures of GA and GU. At this stage, hybridization probably does not play any role, as is expected since the incorporation of C in polymers is so low.

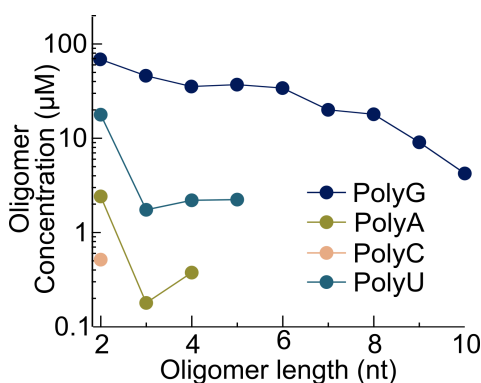


Figure 2.11: Homopolymers of G-cP, A-cP, C-cP and U-cP obtained from individual polymerization from 20 mM solution dried at 40 °C for 18 hours. Poly-G were formed in far higher concentrations than poly-C, poly-A and poly-U.

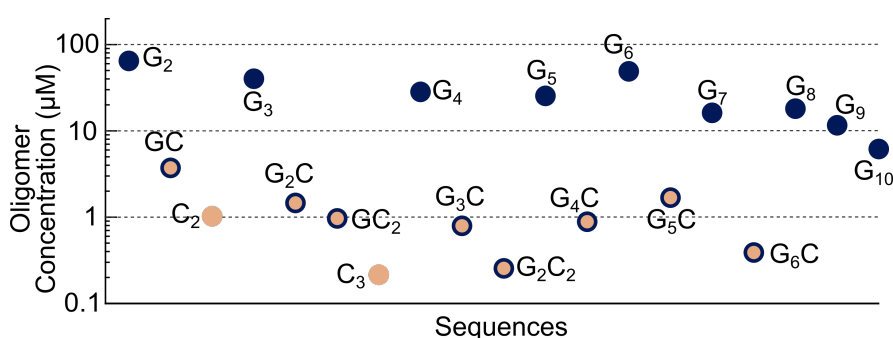


Figure 2.12: Mixed polymerization products composition: Sequence composition of the products synthesized from GC copolymerization with 2 eq K⁺. G-rich oligomers are significantly higher in concentration. G homopolymers are shown in the plot as dark blue dots. Homopolymers other than G are shown as colored dots. Mixed polymers are marked as colored dots surrounded by a blue circle.

GC mixed polymerization was favored by drying at 40 °C on the range 30-80 °C in Figure 2.13 **a**), similarly to G homopolymerization (see Figure 2.10 **b**)). Specific cations also influ-

2.3 Results

enced N-cP polymerization, as shown in Figure 2.13 **b**). K^+ ions yielded oligomers of higher concentrations and length in comparison to Na^+ ions at the same concentrations. The presence of Mg^{2+} ions in the mixture inhibited polymerization.

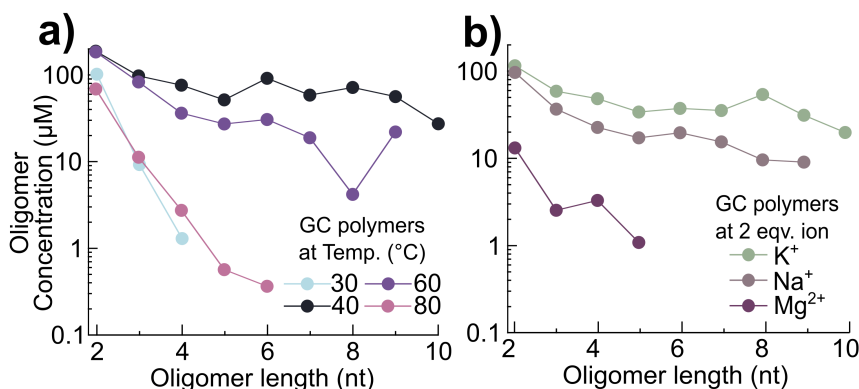


Figure 2.13: Conditions influencing the GC mixed polymerization: a) Temperature screening over a range of 30–80 °C for GC copolymerization. Reduced concentrations of n-mers longer than 3 nucleotides are observed for 80 °C, possibly due to degradation. **b)** Polymerization of GC in presence of different cations: 2 eq K^+ , 2 eq Na^+ , 2 eq Mg^{2+} . In all of those experiment, there also 1 eq of Na^+ due to the initial N-cP sodium salt.

Although all four nucleotides can be oligomerize through the reaction of the -2',3'-cP, G is significantly more reactive. A possible explanation might be that G can form, in aqueous solution, supramolecular assemblies where the position of the -5'OH group of a nucleotide and the 2',3'-cP group are in the right position to react, facilitating the formation of long strands. In comparison, A, C and U orientation would be entirely random when they dry, leading to only detectable 2mers and 3mers.

One possible type of assemblies would be stacks of the guanine bases, based on non covalent hydrophobic interactions of the nucleobases. Šponer and her collaborators [49] have proposed that a stack-assisted geometry triggers the oligomerization of G-3',5'-cP. Some of our collaborators explored the suitability of that mechanism applied to G-2',3'-cP by molecular dynamic simulations and modeling [75].

Another hypothesis would be that these supramolecular assemblies consist of G tetrads, in which four guanine nucleobases are bound to each other by hydrogen bonds, stacked on top of each other. They are stabilized by the presence of cations between each planar tetrad, as shown in Figure 2.14 **a**). Several clues point toward the plausibility of such structures forming under those conditions. G-quadruplexes can form from mononucleotides and oligomers [76] and the hydrogen bonds involved can form up to pH 10 (for higher pH values, the nucleobase is deprotonated, since G has a pK_a of 9–10). In addition, K^+ cations are known to stabilize these structures [77], much more than Na^+ , as they have the perfect radius. Mg^{2+} and especially Li^+ make G-quadruplexes extremely unstable [77, 78]. Polymerization of G-cP was tested in the presence of increasing quantities of Li^+ , compared to the same quantities of Na^+ in Figure 2.14 **b**). The concentrations and the length of the oligomer products decreased

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

with increasing concentrations of Li^+ .

It should be noted that it is difficult to distinguish between the effect of the interstacking of the nucleobases and the G-quadruplexes tetrad arrangements, based on the enhanced polymerization observed with K^+ .

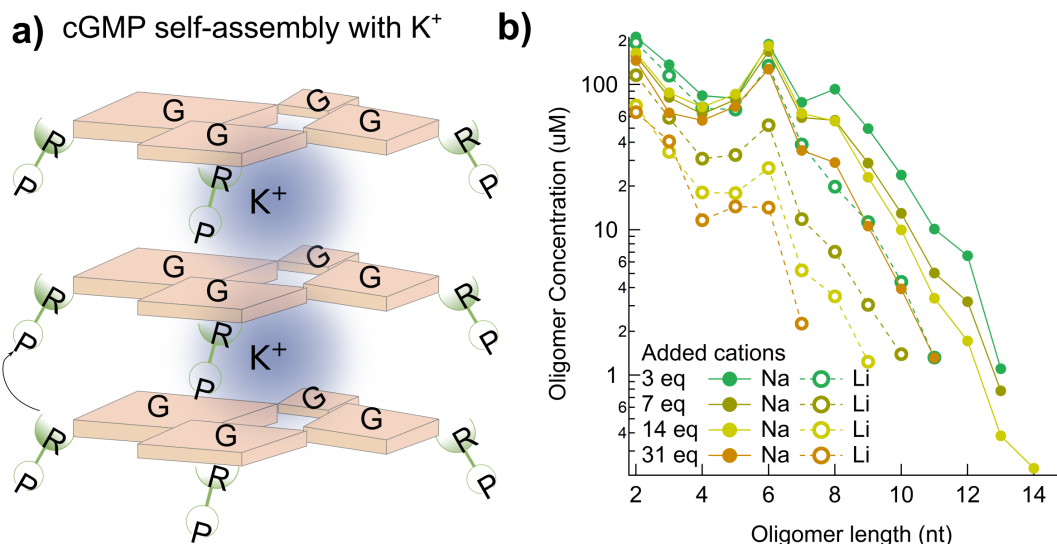


Figure 2.14: G quadruplexes: one possible supramolecular assembly: a) Scheme of G-quadruplexes stacks. The overall charge of this nucleotide assembly is known to be stabilized by K^+ cations, which could explain why K^+ promotes better polymerization than other cations (see Figure 2.13). **b)** Screening of Na^+ and Li^+ quantities for the polymerization of a solution of 20 mM G-cp at pH 10, dried for 18 hours at 40 °C. In addition to the added salts, there was already 1 eq Na^+ due to the 2',3'-cyclic phosphate guanosine sodium salt.

2.3.3 Polymerization at water-air interface

In Figure 2.10 **a)**, we show that the drying process is essential to polymerization and that N-cPs do not react if they remain in aqueous solution. However, we wanted to explore the possibility of making the reaction compatible with the presence of water, in an environment where both drying and rehydration occur, at the interface between the aqueous solution and the air. Previous works on local wet-dry cycling have led to the development of the thermal trap (see section 2.2.3). That microfluidic device mimicks a heated rock pore where air bubbles are trapped in contact with the aqueous solution. We tested the polymerization of N-cP in the thermal trap's chambers (see Figure 2.15), which were partially filled with a solution of N-cP and subjected to a temperature gradient. This thermal non-equilibrium drives constant wet-dry cycling within the chamber. Over time, the meniscus of the bulk solution receded and the material accumulated on the hot side of the wet-dry interface dried [66]. However, because of the formation of dew droplets on the cold side of the chamber and the fall of the droplets, a fraction of the dried material is redissolved.

Wet-dry cycles in surface-based geological environments are subjected to a drift in salt and

2.3 Results

pH conditions due to an imbalance caused by the evaporation of pure water and the rehydration by the salt-containing solution. Previous works reported important processes occurring within that setting, such as accumulation, phosphorylation, encapsulation, gelation, strand separation, enzymatic DNA replication (like an auto-PCR), crystallization and ribozyme activity [29, 66, 67, 79].

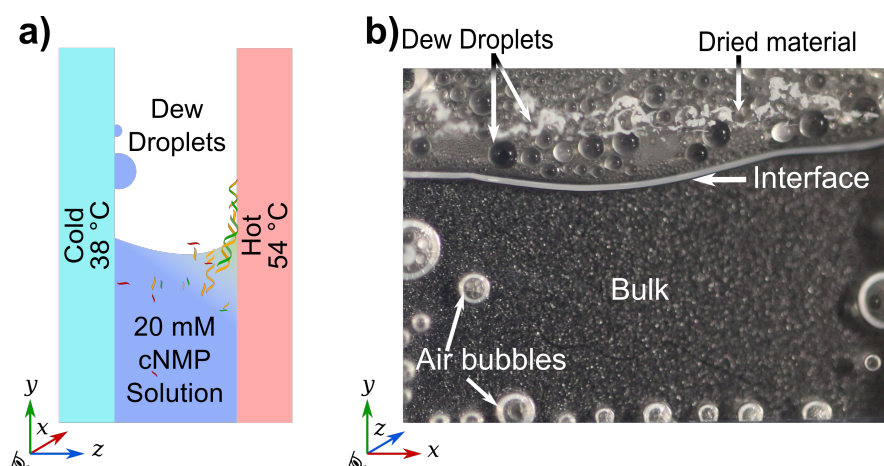


Figure 2.15: Microfluidic chamber under a temperature gradient, mimicking in a heated rock pore: **a)** Scheme illustrating the side view of the chamber filled with a solution of 20 mM N-cP adjusted at pH 10. The chamber is 500 μm thick and subjected to a heat flow with a temperature gradient of 38-54 $^{\circ}\text{C}$. **b)** Picture showing the front view of the chamber. The thermal gradient drives continuous evaporation and recondensation in the air bubbles, triggering accumulation and wet-dry cycling for the N-cP reactants. First, the molecules accumulated at the interface are dried from a receding interface, due to evaporation. Rehydration is provided by dew droplets forming on the colder side of the chamber, which then merge with the bulk solution.

The wet-dry cycling within the chamber went for 18 hours. Then the heating and cooling were stopped, therefore stopping the temperature gradient driving the cycles. After that, the thermal trap was dismantled and the remaining solution was collected along with the dried materials which was redissolved in water. The pH of the solution was lowered by a unit after reaction in the thermal trap. This could be due to the formation of acidic species, such as the 2'/3'-phosphate : at pH 10, one hydroxide anion is consumed to open the phosphate ring. It might also be caused by the dissolution of carbonic acid in the bulk solution, as its solubility is higher at alkaline pH.

For reactions at the water-air interface inside the thermal trap, a total concentration of 20 mM N-cP was used. For a mixed GC polymerization, the solution contained 10 mM G-cP and 10 mM C-cP, and for a AUGC experiment 5 mM each. A single chamber was filled with around 30 μL and the three chambers were run and extracted together (see section 2.2.3). The comparison between the results of experiments conducted under dried conditions and in the thermal trap, normalized to the same starting quantities of reactants, can be seen in Figure 2.16 for GC and 2.D.1 for AUGC experiments. Both experiments show comparable polymerization efficiency, but the product composition distribution is slightly different. For AUGC, all the dimers and most of the trimers were detected. However, the tetramers and pentamers

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

are predominantly G-rich sequences. Some differences were noted in the product distribution between the results in dried state and in the thermal trap for both AUGC and GC mixed polymerization. Although the G homopolymerization is lower in the thermal trap, the mixed and C products show higher concentrations. The product composition was more balanced for the reaction at the interface in the thermal trap.

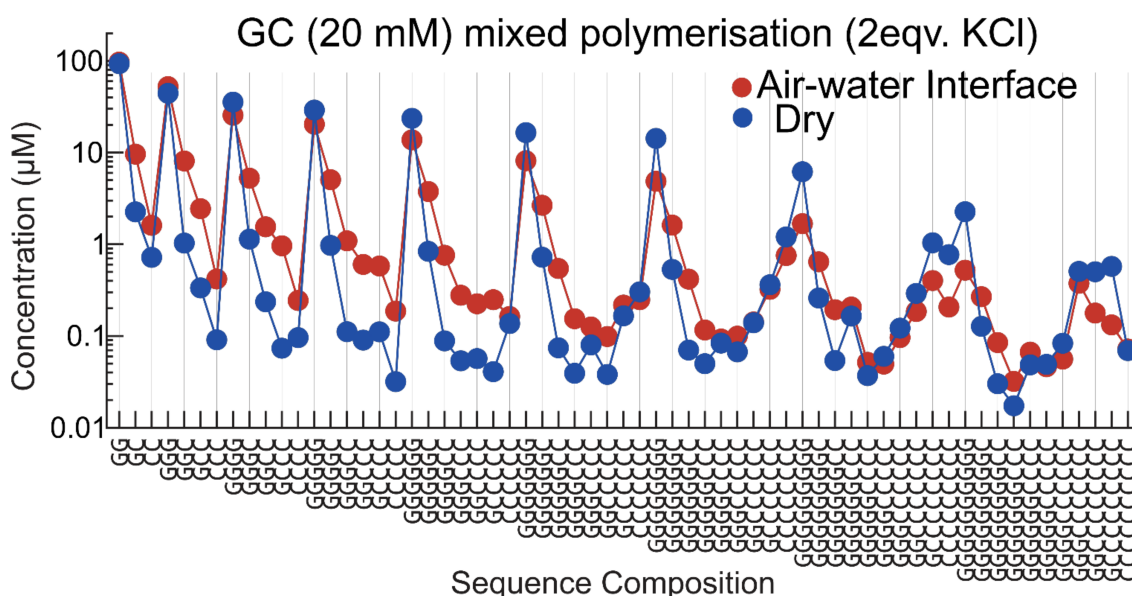


Figure 2.16: Comparison of the oligomerization efficiency between the reaction in dried state and experiment at water-air interface Oligomerization of mixed monomers: G-cp, C-cp 10 mM in presence of 40 mM KCl at pH 10 for 18 hours. Especially for the longer strands, the oligomerization in the simulated rock pore shows improved yields over the dry reaction.

2.3.4 Diphosphoester bond: 3'-5' versus 2'-5'

Our very preliminary digestion studies (in Figure 2.E.1) and ^{31}P NMR results (in Figure 2.8 **a**)) suggested a considerable backbone heterogeneity (3'-5' and 2'-5') within the oligomers. Figure 2.E.1 show the absorbance (at 260 nm) chromatogram of the control samples with annotations made to indicate the peaks shown in the corresponding **b**) to **i**) plots. The UV chromatograms show that some of the peaks disappear, and others remain identical after degradation by the enzyme. The nuclease degrades only the 3'-5' linkages, but does not interact with the 2'-5' ones. However, a full quantitative treatment is beyond the scope of this study.

2.4 Discussion

The data show that G-cp, which had not been tested in previous studies, easily oligomerizes in the dry state. The reaction occurs over a range of temperatures (40-80 °C) and pH (7-12),

2.4 Discussion

and does not require additional catalysts, making this reaction robust. In geological settings, dissolved gases, salts and the presence of some minerals can vary the pH in which the RNA formation occurs [38, 80]. Some environments on Earth naturally have local alkaline pH [14, 81]. Polymerization can also occur in a partially aqueous setting: the polymerization is driven there by wet-dry cycles at a heated interface, adding RNA synthesis to the pool of prebiotic processes possible within such a setting. The cyclic monomers undergo polymerization and ring opening (see Figure 2.8). It must be noted that the hydrolysis of the cyclic phosphate is the dominant reaction at the tested temperatures. Products with a length of up to 15 nucleotides could be detected (in Figure 2.7), for a total polymerization yield around 3% for the polymerization of G-cP at pH 10 and 40 °C. In mixed polymerization, the G oligomers incorporated some C, A, U monomers, albeit at lower concentrations and alone, C, A and U did not oligomerize significantly. To compare with the yields achieved by Verlander and Orgel when they tested the polymerization of A-cP in the presence of 1,2-diamine, around 0.4 % was detected for a G 6mer in 18 hours and they obtained 0.8 % poly A in 40 days.

An important feature of that polymerization is that at alkaline pH, the 2',3'-cyclic phosphate group form a diphosphoester bond by transphosphoesterification without *ex-situ* or *in-situ* activation mechanism or added catalyst, and under low salt conditions. This finding is a very good starting point for Darwinian evolution. Low salt condition are also interesting for RNA evolution and replication as they notably help strand separation and reduce RNA degradation [67]. To an extent, the polymerization of N-cP is found to be a relatively clean reaction under the tested conditions. In comparison, *in situ* EDC activation yields side products, especially at high temperatures [82]. I did not detect major side products with analysis by HPLC-MS, other than the hydrolysis products N-3'P and N-2'P, and the salt adducts of sodium and potassium.

The abiotic formation and recycling of N-cP monomers is feasible, as they are known to be produced by phosphorylation [62, 63, 83], nucleotide synthesis [65] and are common degradation products of RNA [60]. I have also explored the formation of the reactive phosphate ring in conditions compatible with its subsequent transphosphoesterification in Chapter 4. Thus, with subsequent studies, a cycle of reactions involving polymerization, oligomer extension, ligation, hydrolysis and recyclization of monomers under early Earth conditions become conceivable. Furthermore, recombination and templated ligation using the reactivity of the 2',3'-cyclic phosphate have already been observed [84, 85].

For our studies, my colleagues and I compared monovalent ions K^+ and Na^+ and divalent ions Mg^{2+} , Li^+ . They were chosen for their relevance in modern life and their abundance in the early Earth [86]. In addition, Mg^{2+} plays a key role in ribozyme activity [87]. K^+ cations increase more polymerization than Na^+ . The inhibition of the reaction by Mg^{2+} possibly occurs by a combination of base catalysis mechanism, the opening of -cP endings and degradation of the oligomer products. In spite of its role in ribozyme functionality, high concentrations of Mg^{2+} also deter RNA replication by favoring the formation of strong RNA duplexes. In addition, the presence of millimolar concentration of Mg^{2+} is enough to inhibit the membrane self-assembly of fatty acids and this has been considered an incompatible aspect for the co-emergence of RNA and fatty acid membranes [88, 89]. However, under the conditions studied, RNA formation and encapsulation within fatty membrane might be possible.

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

Our studies also show that the backbone's heterogeneity (2'-5' vs. 3'-5') within the polymers is significant. The heterogeneity of the linkage is a consequence of the vicinal –OH groups at 2'- and 3'- positions of the ribose (also their comparable nucleophilicity) and the equal probability of opening the ring from either side in the case of N-2',3'cP. It has been demonstrated that the presence of 2'-5' linkages allows efficient strand separation by reducing the melting temperature (T_m) of oligomers, which is relevant in the case of G-rich sequences observed in this polymerization. Lowering of T_m is therefore critical to replication of sequences [90, 91]. These studies also show that the presence of 2'-5' linkages allows the folding of RNA into three-dimensional structures, similar to native linkages. They do not hinder the evolution of functional RNAs.

Several type of self-assemblies might contribute to the polymerization of G-cP in the dry state. Based on the effect ions, my colleagues and I hypothesize that G-cP monomers form tetrads, and several such tetrads stack over each other with a central K^+ cation between the stacks and promote polymerization (see Figure 2.14). Such an arrangement is reminiscent of G-quadruplexes, which may exist under the reaction conditions tested. Moreover, as Li^+ cation are reported to have a destabilizing impact on the formation of G-quadruplexes, G-cP were dried in presence of increasing quantities of LiCl added to the reaction solution. The amount of polymers detected decreased with increasing amount of Li^+ . Some studies also point to self-assembly of G-5'P and G-3'P into helical stacks [92]. Individual cations such as Na^+ or K^+ could favor a specific arrangement; thus, explaining the differences in polymerization efficiencies as a function of the type of cation.

In conclusion, the formation of short RNA strands was observed under very simple conditions, alkaline pH and mild heat-drying, by reaction of N-cP. G-cp, which has never been tested before in literature, shows a particularly high efficiency, including in the absence of potential catalyst. In the next chapter, the polymerization results, especially for A, C and U, were improved by drying at room temperature and wet-dry cycling.

Addendum 2

2.A Building of the thermal trap

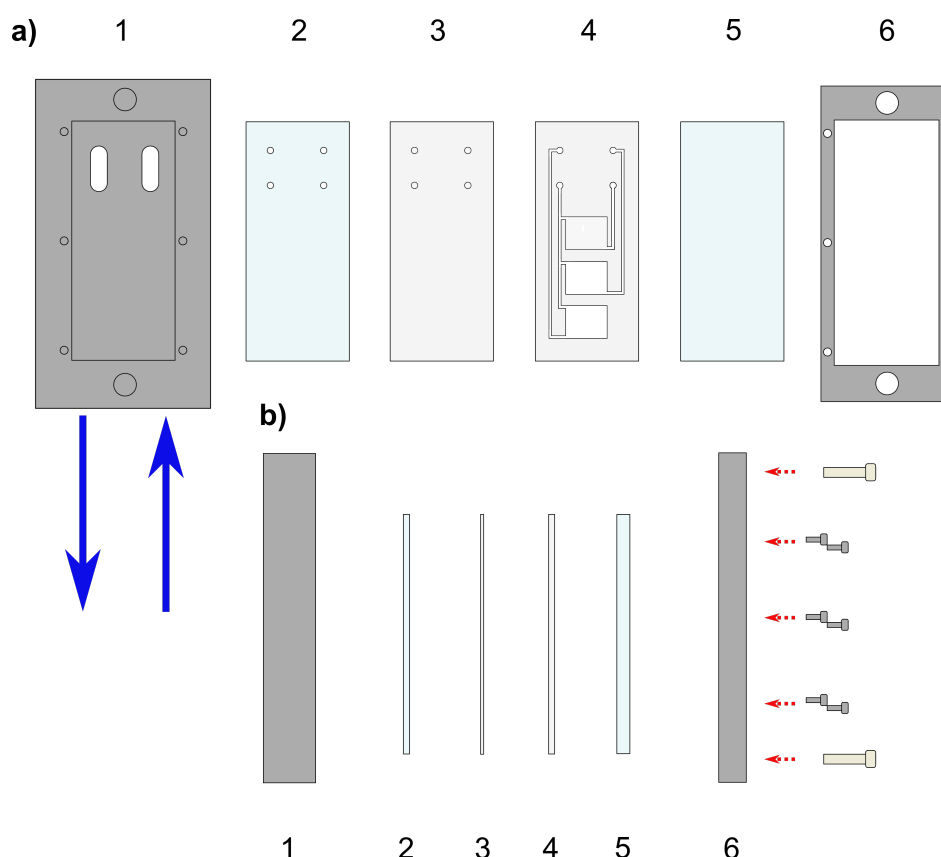


Figure 2.A.1: Thermal trap parts and building: **a)** Front view of the different parts of the thermal trap. **b)** Front view of the different parts of the thermal trap in the order in which they are put together for the experiment. **1** Metal support on which the thermal trap is attached which the two plastic screws shown in **b)**. The plate is connected to a water bath (the cooled down water input and output are represented as blue arrows) and cooled down to 20 °C. **2** 500 μm -thick sapphire in which four holes have been pierced. **3** 125 μm -thick Teflon layer in which four holes have been cut at the same place as **2**. **4** 500 μm -thick Teflon layer in which the full design of the microfluidic chambers, the channel and the feeding holes have been cut. Each chamber can contain 32 μL . **5** 2 mm-thick sapphire plate, which is the "wall" which closes the chambers on the other side. **6** Metal plate closing the thermal with the six small metal screws in **b)**. A metal heater will be attached on top of **5** and **6** to set the higher temperature. The reaction solution is introduced inside the chambers using tubins attached to the plate **1** and pressed against the four holes in **2**, **3**, **4** which are all connected. The solution flow through the channels cut in the Teflon layer **4** to the three chambers, which are run in parallel for one experiment.

2. RNA oligomerization without added catalyst from 2',3'-cyclic phosphate nucleotides

2.B Spectral Browser workflow

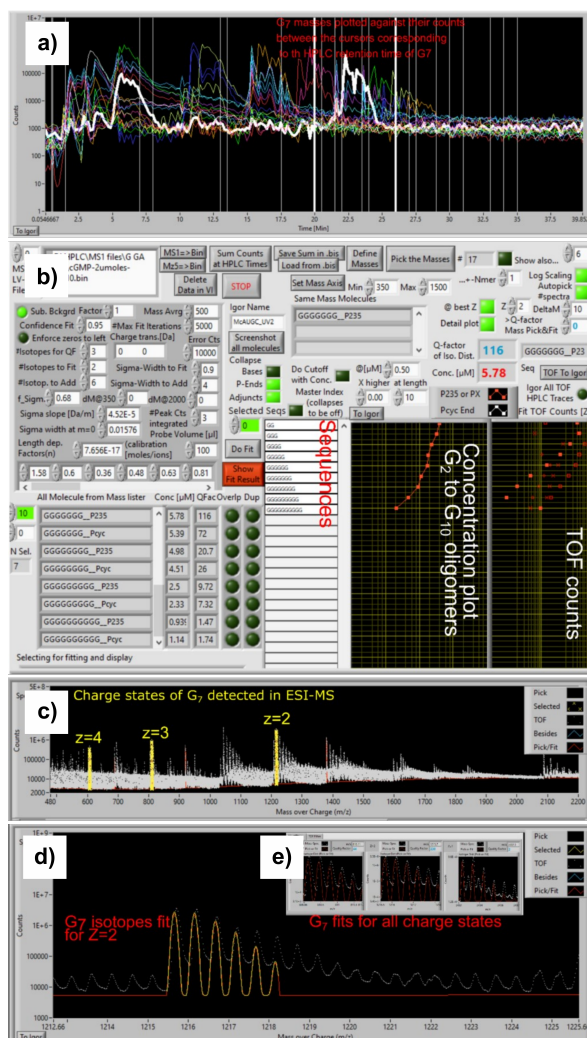


Figure 2.B.1: Example of a routine of G₇-3'P using the Spectral Browser: **a)** The masses of G₇ (sum of the counts of all the charged states) plotted against time on the x-axis. The peak of the G₇ standard is set between the two cursors in white. **b)** Concentrations of the oligomers detected the oligomers of G. On the right of concentration plot is the raw counts of the corresponding G oligomers. **c)** Combined spectrum resulting from the addition of all the spectrum taken between the cursors. The yellow traces correspond to all the charge states (z=2,3 and 4) detected in the ESI-MS for a G₇ oligo. Z=1 is a dormant charge state for a G₇ oligo and hence not observed. **d)** The fit results m/z = 1215.7 (corresponding to the second charged state). **e)** Inset are the fits for all the charge states of a G₇ oligomer.

2.C GA and GU mixed polymerization

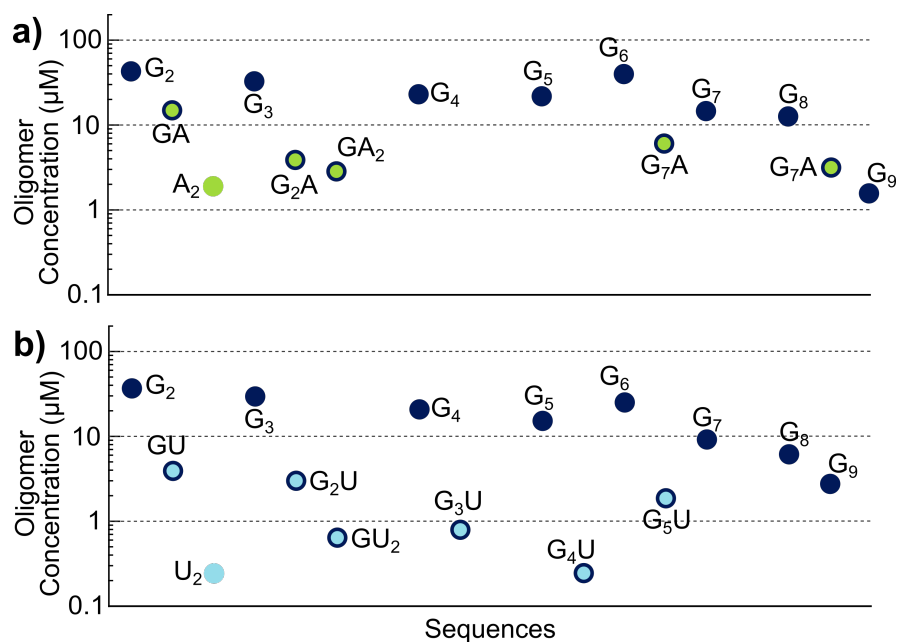


Figure 2.C.1: Mixed polymerization products composition. Sequence composition of the products synthesized from copolymerization with 2 eq K⁺: **a)** GA and **b)** GU. G-rich oligomers are significantly higher in concentration. G homopolymers are shown in the plot as dark blue dots. Homopolymers other than G are shown as colored dots. Mixed polymers are marked as colored dots surrounded by a blue circle.

2.D AUGC polymerization at water-air interface

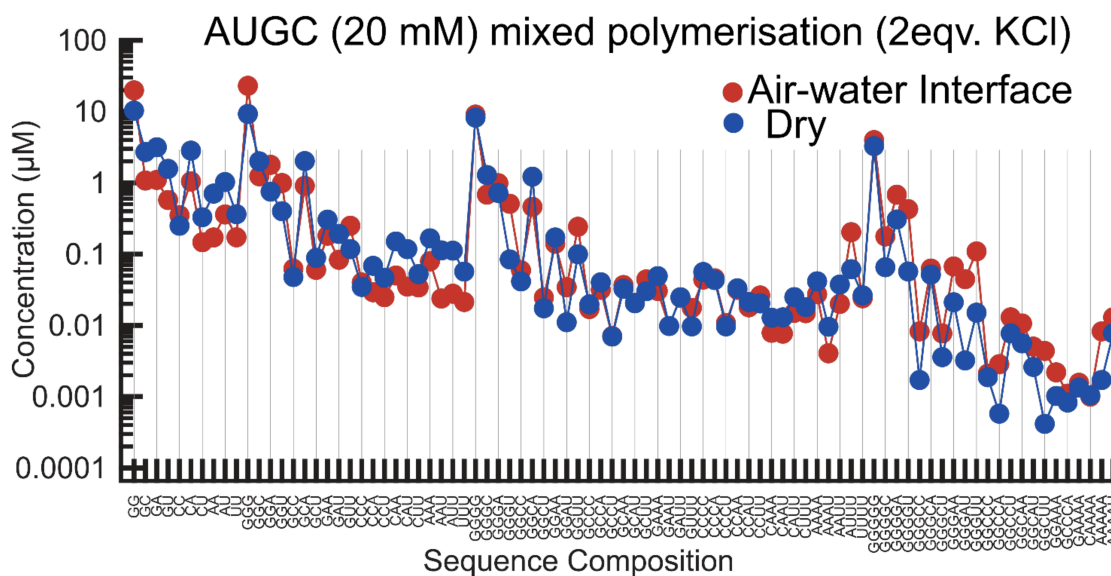


Figure 2.D.1: Comparison of the oligomerization efficiency between the reaction in dried state and inside the trap. Oligomerization of mixed monomers: G-cP, C-cP, A-cP, U-cP, 5 mM each, in presence of 40 mM KCl at pH 10 for 18 hours. Especially for the longer strands, the oligomerization in the simulated rock pore shows improved yields over the dry reaction.

2.E Enzymatic degradation results

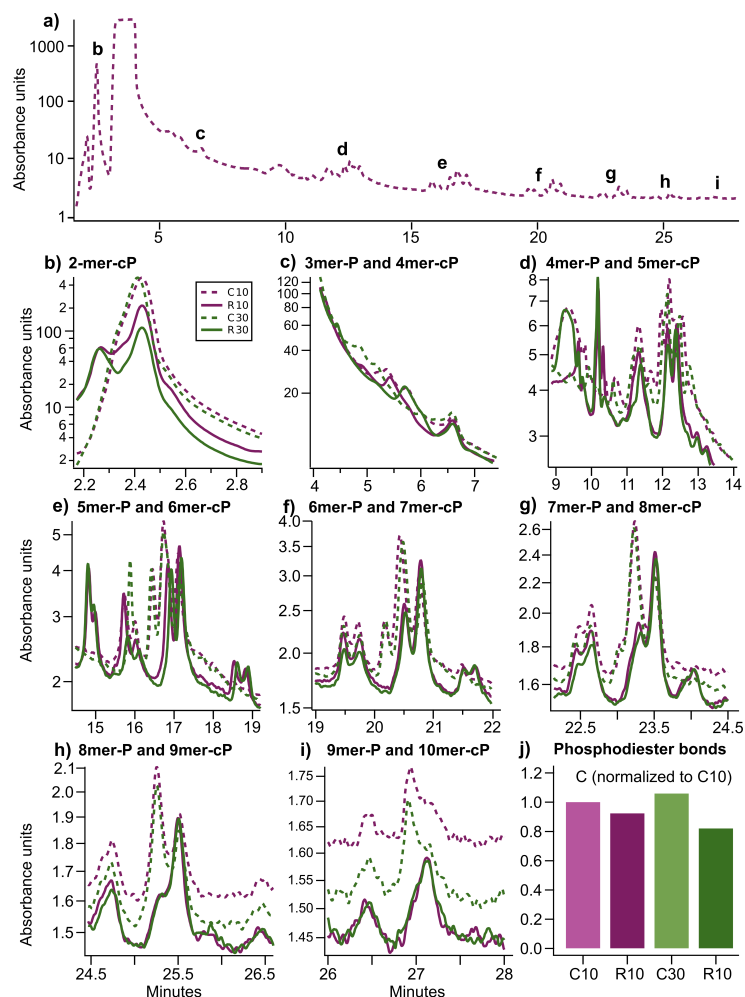


Figure 2.E.1: Nuclease P1 digestion conducted on polymerized samples of G-cP to determine the amount of linkages 3'-5' and 2'-5'. **a)** Absorbance (at 260 nm) chromatogram for the control sample C10 without enzyme left in the reaction medium for 10 min at 37°C. The peaks corresponding to the oligomers products are annotated by the letters b-i. Saturated peak at 4 min is due to the traces of phenol used in the extraction masking some oligonucleotide peaks. **b-i)** UV chromatograms for samples C10 (no-enzyme control, 10 min incubation, purple-dashed), R10 (digested by enzyme during 10 min, purple-full), C30 (no-enzyme control, 30 min incubation, green-dashed), R30 (digested by enzyme during 30 min, green-full). Intensity of some of the peaks is lowered in the enzymatic digest samples. **b.** G₂-P, **c.** G₃-P and G₄-cP, **d.** G₄-P and G₅-cP, **e.** G₅-P and G₆-cP, **f.** G₆-P and G₇-cP, **g.** G₇-P and G₈-cP, **h.** G₈-P and G₉-cP, **i.** G₉-P and G₁₀-cP. **j)** Concentration equivalent of phosphodiester bonds were calculated for each sample using the integrated EIC counts. Concentrations of the samples R10, C30 and R30 were normalized to that of C10.

Chapter 3

Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

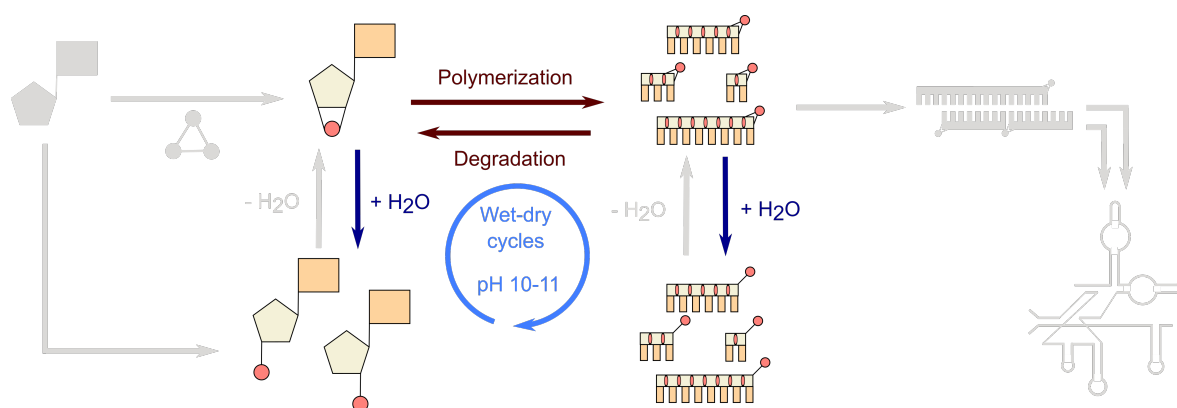


Figure 3.1: Chapter summary: Polymerization of 2',3'-cyclic phosphate nucleotides by wet-dry cycling at alkaline pH and room temperature, and hydrolysis and degradation in the same conditions.

Chapter summary: The work in this chapter was conducted in collaboration with co-workers for the University of Milano: Prof. Dr. Tommaso Bellini, Dr. Tommaso Fraccia, Dr. Federico Caimi and Francesco Fontana.

We further explored the effect of wet-dry cycling on polymerization of 2',3'-cyclic phosphate nucleotides, with collaborators from the university of Milano. The drying is now carried out at room temperature and completed within 8-10 h. Temperature plays an important role for polymerization and hydrolysis, which deactivates a significant portion the reactants. Those new conditions improve considerably the polymerization of all 4 nucleotides. Mixed samples still show a bias toward G, but it can be minimized by slightly more alkaline conditions. The drying of the monomers (and oligomers for a number of cycles > 1) different reactants was followed by microscopy. The aim was to detect the self-assembly of the reactants prior to the full drying, as it affects the efficiency of the polymerization and to the formation of long RNA strands. The formation of crystalline structures, resulting from supramolecular assemblies in aqueous solution, was confirmed for G and, to some extent, for A.

3.1 Introduction

In Chapter 2, we demonstrate the polymerization of 2',3'-cyclic phosphate ribonucleotides (N-cP) for all canonical nucleobases by drying of an alkaline solution, with moderate heating and without additional catalyst. However, the yields of polymerization were substantial only for G-cP, around 3-4 % at pH 10. The reaction of A-cP, C-cP and U-cP produced only dimers and trimers. When different N-cP were mixed together, it yielded mixed oligomers, but dominated by G. This lack of sequence diversity and heavy bias toward one of the nucleobases would be a problem for further evolution of RNA. It would inhibit hybridization and therefore templated ligation and replication. Also, homopolymers are not very useful in the genetic code [93].

The aim of the present work is to explore conditions and mechanisms that might drive the reaction further, improving both the yields and the length of the products and their composition. We already showed, in the previous Chapter, N-cP polymerization at a heated water-air interface. There, the polymerization was driven by a constant evaporation-rehydration cycling, and the bias toward G seemed slightly lower compared to the dry state. Thus, we decided to explore further the impact of wet-dry cycling on polymerization by performing manually the cycles and observing the results after each one.

The high polymerization for G is most likely due to self-assembly of the monomers before full drying. My collaborators monitored those potential assemblies in the aqueous phase and in the dried state by PTOM (Polarized Transmission Optical Microscopy) observation. PTOM can detect molecular ordering, including in liquid states, such as liquid crystals.

3.1.1 Wet-dry cycling

Cycling between dehydration and rehydration is used to drive a wide range of prebiotic reactions producing biomolecules and biopolymers: condensation of amino acids [94, 95], synthesis of nucleosides [96] and formation of oligonucleotides [97, 98]. Some researchers also fed the reaction with fresh reactants during rehydration, demonstrating enhanced yields for phosphorylation [98, 99].

In wet-dry cycling conditions, the reactants (and subsequent products) go from a diluted aqueous phase to a high-concentration aqueous phase and end up in a dry state. Compared to use of catalysts which only lower the energy activation barrier and improve the kinetics of a reaction, wet-dry cycling continuously changes the chemical potentials of the reactants and products, and thus the thermokinetic equilibrium of the system. The chemical potential of the reactants can be increased and the potential of the products can be lowered, making the reaction more favorable [100]. The wet-dry cycling also leads to the variation of other parameters such as the pH and salt concentrations, which are important factors in prebiotic reactions [67, 79].

Considering an early Earth scenario, the aqueous reservoir oscillations can be due to "weather conditions" such as temperature changes, humidity and deliquescent mineral [94, 101, 102]. Those oscillations of conditions for the reactants can also be obtained by non-equilibrium at water-air interfaces [103–105], as I explored in the previous Chapter.

3.2 Materials and methods

3.1.2 Liquid crystals

Liquid crystals are fluid phases, composed of partially oriented molecules showing positional ordering. Physically, they are mesophases between the isotropic liquid state and crystalline solids, a form of still-liquid self-organization. Thus, liquid crystal phases can easily be observed in polarized microscopy.

Nucleic acids liquid crystal were first observed from long DNA strands *in vitro* [106] and *in vivo* [107]. Later works showed that liquid crystals can be formed by smaller DNA sequences (6-20mers [108]). The crystalline micro-domains are composed of DNA duplexes, which possess an increased rigidity compared to single strands. Mononucleotides and nucleic acid single strands have only a weak propensity to stack, but the formation of hydrogen bonds reduces the water solubility of the compounds, which promotes aggregation and ordering. Thus, with the ability of the G monomers to form G-tetrads via non-Watson-Crick hydrogen bonds, they can form liquid crystals [109].

Liquid crystals are environments that could provide many advantages for the origins of life and starting evolution: protection by compartmentalization, selection of biopolymers based on their structures, enhancement of abiotic polymerization and elongation by ordering the reactants [110–112]. Inside the liquid crystals ordered domains, the reactive terminals of the oligomers are kept in close proximity to each other by packing and stacking forces [112]. In the case of the RNA world in particular, one of the fundamental questions is: how to synthesize *de novo* the first long RNA strands. The shortest ribozymes that are catalytically active in water are at least 50 monomers-long. To this day, there is no known abiotic pathway to oligomers of that length starting from monomers. Liquid crystals could provide a template that orders the mononucleotides and oligonucleotides and enhances the reactions of polymerization and ligation [113]. As an example, liquid crystals of nucleic acids reportedly improve the ligation of DNA oligomers by water-soluble carbodiimide EDC [114].

3.2 Materials and methods

3.2.1 Reactants and standards

Cyclic nucleotides (G-cP, A-cP, C-cP, U-cP) and the linear-ending monomers (G-2'P, G-3'P, A-2'P, A-3'P, C-2'P, C-3'P, U-2'P, U-3'P) were purchased from BioLog in their sodium salt form, as lyophilized samples.

Standard pre-synthesized homooligomers of G (2 to 10mers), C (2 to 10mers), A (2 to 4mers) and U (2 to 4mers) with a 3'-phosphate ending (and -OH at the 5' end) were purchased from Biomers as lyophilized samples.

3.2.2 Sample preparation and wet-dry-cycles protocol

The experiments were conducted by my collaborators from the university of Milano, Francesco Fontana and Dr. Federico Caimi.

Diluted solutions (10 mM) of 2',3'-cyclic monophosphate nucleotides species (A, U, G or C) and equimolar mixtures (AU, GC, AUGC) were prepared, portioned in 200 μ L Eppendorf tubes (1 μ mol of nucleotides in each tube) and then lyophilized overnight. The resulting powders

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

were stored at room temperature. Samples were prepared by dissolving each portion in 20 μ L of deionized water (the initial concentration is 50 mM) and the pH was adjusted using potassium hydroxide (KOH). Solutions were evaporated on a glass-bottomed multi-well plate (Corning, 4580) at room temperature during 24 hours.

For analysis by LC-MS, the samples were received in lyophilized state and then rehydrated in RNase-free water (from Thermo Fisher Scientific). 10 nmol (for single nucleotide samples) to 40 nmol (for AUGC samples) were injected in the HPLC for analysis, without further treatment for the majority of the samples.

3.2.3 LC-MS instrument specifications and protocol

Measurements were performed on a high-performance liquid chromatography (Agilent 1260 Infinity II bioinert) with a G7115A 1260 Infinity II diode array detector and coupled to an electrospray ionization time-of-flight mass spectrometer (Agilent 6230B with dual AJS EIS). The column used was an Agilent Advance Oligonucleotide C18 Column (4.6 x 150 mm 2.7 μ m) heated at 60 °C with a pressure rating of 600 bar.

The oligomers were separated by using ion-pairing reversed-phase HPLC. The eluent consisted of mixtures of water (Bottle A) and methanol (Bottle B: 50 % water, 50 % methanol) containing each 8 mM trimethylamine (TEA) and 200 mM hexafluoroisopropanol (HFIP), with a flow of 1 mL/min. The method started with 1 % of B for 5 min, followed by a gradient, increasing from 1 % to 30 % B over 22.5 min and then to 40 % for 15 min. Then, the column was flushed with 100 % B for 5 min before being returned to 1 % for 6 min, to re-equilibrate the column.

Detection of the eluted compounds is achieved by using a Diode Array Detector (DAD) WR (wavelength used: 260 nm), and a coupled mass spectrometer.

The quantification of eluted oligonucleotides was achieved using ESI-TOF in negative mode (employing specific source parameters: Gas temperature: 325 °C, Drying gas flow: 13 L/min, Sheath gas temperature: 400 °C, Sheath gas flow: 12 L/min, VCap: 3500 V, Nozzle Voltage: 2000 V). Reference masses are run in parallel to the sample run using standard reference and tuning mix recommended by Agilent (product number G1969-85000).

3.2.4 MS data analysis by custom-written LabVIEW program

The MS data acquired from HPLC ESI-TOF was converted to mz5 format using MSConvert, a component of ProteoWizard [115]. Subsequently, it was imported into a custom LabVIEW program (Spectral browser 3.58 or a newer version Spectral browser 4.63) for detailed analysis. These versions of the program were updated and modified compared to the one used in Chapter 2.

3.2 Materials and methods

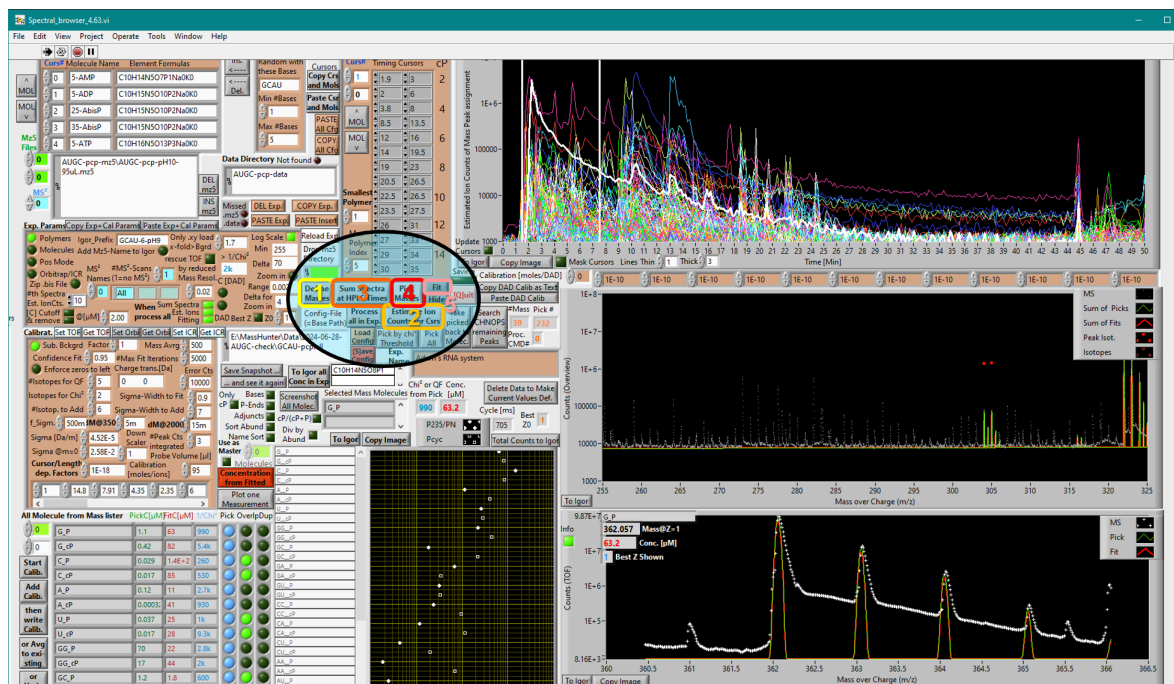


Figure 3.2: User Interface of Spectral Browser version 4.63: The MS data analysis pipeline involves a systematic process denoted by numbers (1-5), each step is color-coded (yellow to red) and described below.

Workflow for the Spectral Browser 4.63 (see Figure ??):

1. Mass List generation
2. Extracted Ion Counts chromatograms plotting
3. Summing of raw mass spectra
4. Pick and 5. Fit of summed isotope distribution and concentration determination.

1. Mass List generation (see Figure 3.A.1):

This list is then passed via ZeroMQ message transfer protocol to Python, where IsoSpecPy [116] retrieves theoretical isotope distributions (1-ii and 1-iii) for the corresponding chemical formulae.

2. Extracted Ion Counts (EIC) chromatogram plotting (see Figure 3.A.2):

The mz5 is loaded (folder in 2-A and file in 2-B) and the EIC chromatograms corresponding to the masses calculated in the previous step are plotted in 2-C. This step facilitates the determination of cursors position for different oligonucleotide lengths (2-D).

3. Summing of raw mass spectra (see Figure 3.A.3)

Raw mass spectra within the cursor positions are summed. Additionally, the program saves the summed spectra into another format (.data) with all the parameters for the calculs, optimizes memory usage and enhances data loading time for future analysis. This feature allows for efficient retrieval and utilization of the summed spectra in subsequent analyzes.

4. Pick and 5. Fit of summed isotope distributions (see Figure 3.A.4)

Theoretical isotope distributions for molecules between the cursor positions are assigned

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

to the corresponding m/z in the summed spectra using the most abundant isotope (4). Depending on the number of isotopes selected (4-A), their theoretical distribution is fit to the measured data (5) (and plotted in 4-B) to generate the TOF ion counts. The sum of TOF ion counts at different charge states is calculated for each molecule based on the fitted distributions. The concentration of the oligonucleotides (4-C) in the sample is determined for the volume (4-D) analyzed using the TOF counts. For each length of oligonucleotide, I calculated coefficients (moles/ion counts) (4-E) using oligonucleotide standards of 2-10 mer of known concentrations (see below).

3.2.5 Quantification

The separation in the HPLC column of the standards for each nucleobases can be seen in Figure 3.B.1.

Monomers quantification

The monomers are quantified using UV absorption for the wavelength 260 nm. The monomers can be detected by mass spectrometry but the quantities are too high for quantification: the function of ion counts depending on concentrations is not linear.

Oligomers quantification

The oligomers are quantified by their extracted ion count. For the concentrations I detected, the function of the extracted ion counts EIC depending on the concentration was linear. The calibration was done using the standard oligomer of the same length with a -3P ending (see Figure 3.B.1)

Yield calculations

For each molecule x , the yields are calculated over the total concentration of monomers A, U, C or G (remaining cyclic monomers + hydrolyzed monomers + monomers included in an oligomer). The total polymerization yield is defined as the total concentration of monomers included in an oligomer (cyclic and linear-ending) over the total concentration of monomers.

With $N = G, C, A$ or U , cP indicating a molecule with a cyclic phosphate ending, P a linear phosphate ending and j being the maximum length of the oligomers detected:

- the total concentration of monomers (in μM) is defined as:

$$C_{tot} = \sum_{i=1}^j \sum_N i * ([N_i - cP] + [N_i - P]) \quad (3.1)$$

with $[N_i - cP]$ the concentration of oligomer of length i and ending cP and $[N_i - P]$ of oligomer of length i and ending P .

- the total yield of polymerization for the products with a cyclic end (in %) is defined as:

$$Y_{cP}^{tot} = \sum_{i=2}^j \sum_N i * [N_i - cP] / C_{tot} * 100 \quad (3.2)$$

3.3 Results

- the total yield of polymerization for the products with a linear end (in %) is defined as:

$$Y_P^{tot} = \sum_{i=2}^j \sum_N i * [N_i - P] / C_{tot} * 100 \quad (3.3)$$

- the total yield of hydrolysis for the monomers (in %) is defined as:

$$Y_P^1 = \sum_N [N_1 - P] / C_{tot} * 100 \quad (3.4)$$

- the concentration of n-mer N_i with a defined AUGC composition reported for a theoretical starting concentration of 50 000 μM is defined as:

$$C_{N_i} = ([N_i - cP] + [N_i - P]) / C_{tot} * 50000 \quad (3.5)$$

to compensate for the potential loss of material during the lyophilization step.

- the total for a specific length i of oligomer (in μM) is defined as:

$$C_{tot}^i = \sum_N ([N_i - cP] + [N_i - P]) / C_{tot} * 50000 \quad (3.6)$$

3.2.6 Microscopy observation

The microscope imaging was performed by my collaborators. A Nikon TE 200 inverted microscope was used together with a set of Nikon magnification objective lenses (from 5x to 50x) and a TE-PSE100 Nikon lamp. Images were acquired with a Nikon DS-Fi3 camera. The system was connected to a TANGO 3 Desktop device (Marzhauser Sensotech) to control the stage mounted on the microscope. The automated image acquisition of samples evaporating overnight was managed using NIS-Elements BR software (exposure = 20 ms, gain = 1x, magnification = 5x). Polarizer and analyzer were crossed at 90°.

3.3 Results

We report the oligomerization of N-2',3'cP by dehydration-rehydration cycling at room temperature in presence of KOH. In a typical experiment, 20 μL of an aqueous solution of N-cPs (initial concentration 50 mM) and the pH of the solution is adjusted to values between 6 to 12 for the first wet-dry cycle. The results were analyzed by HPLC-MS (see Methods), and new peaks at later retention times are shown in Figure 3.3. Those peaks are attributed to hydrolyzed monomers and synthesized polymers by comparison with standards (see Figure 3.B.1) and detection with the mass spectrometer.

3.3.1 Conditions of wet-dry cycling

The evaporation was coupled with a slight reduction of the pH value, most likely due to a combination of phosphate ring opening and CO_2 (g) dissolution. We tested two different protocols and the results for G-cP at pH 10 are presented in Figure 3.3: for the first one (grey in the figure), the pH was readjusted to 10 by addition of KOH at each rehydration. For the

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

second one (black in the figure), the pH was set to 10 only for the first dehydration and each rehydration was done with pure water. The results presented in Figure 3.3 correspond to the 10th cycle.

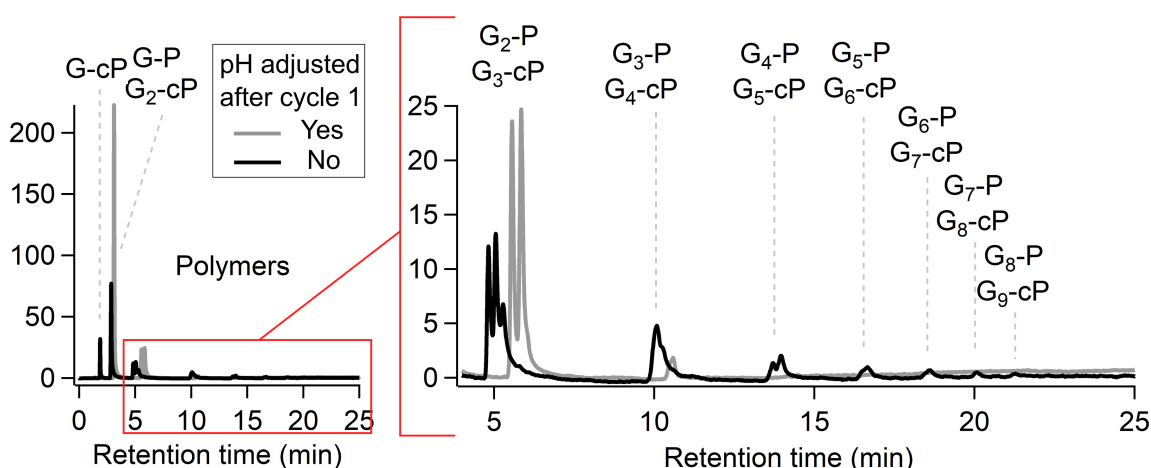


Figure 3.3: Effect of the pH adjustment during wet-dry cycling on the polymerization of G-cP: The starting solution of 50 mM G-cP was initially adjusted to pH 10. For one of the samples (in grey), the pH of the solution is adjusted to 10 after every rehydration. For the other one (in black), the pH is not adjusted.

The second protocol shows higher concentrations for the oligomers longer than 3, for a total polymerization yield of 70 %. For the first protocol with pH adjusted at every cycle, the peak corresponding to the remaining N-cP has entirely disappeared and most of the reactants – 67 % – are converted into hydrolyzed monomers N-2P or N-3P. Therefore, the lower polymerization results, 32 %, are due to base-catalyzed hydrolysis.

All the other experiments were conducted with the second protocol and the pH was not readjusted after the first drying.

3.3.2 Effect of wet-dry cycling on 2',3'-cyclic phosphate nucleotides

Polymerization

The drying at room temperature improved significantly the yields of polymerization compared to the drying at 40 °C (Chapter 2), for all four canonical nucleotides. After one dehydration at 40 °C, G polymerization reached 4 %, against 70 % at room temperature (see Figure 3.4). Previously, I could detect only dimers and trimers for the other three nucleotides, but for the drying at room temperature, the oligomers detected were as long as 4 to 6mers for the first cycle at pH 10 or 11 (see Figure 3.5) with total yields around 5 to 20 %.

3.3 Results

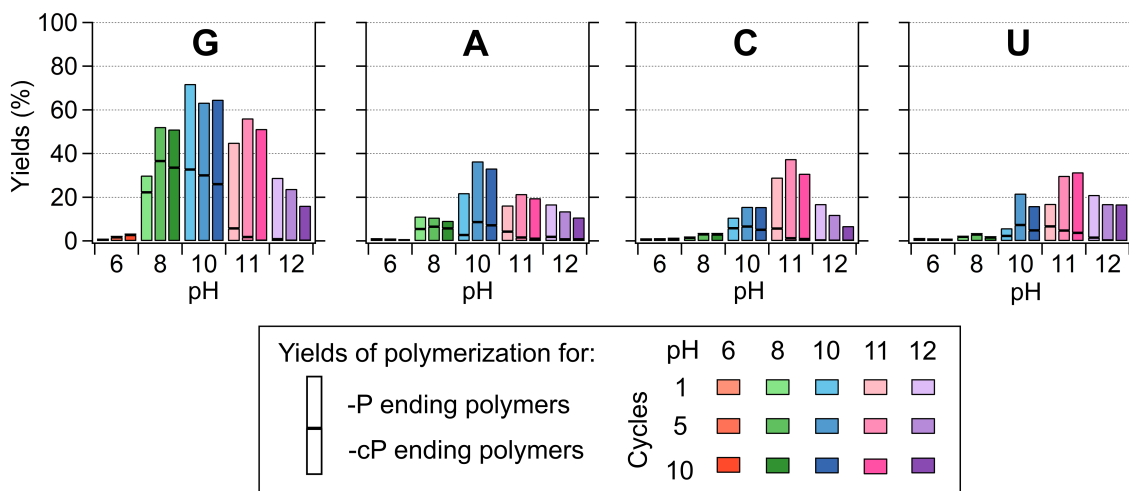


Figure 3.4: Polymerization yields for G, A, C and U: A solution of 50 mM N-cP (N=G, A, C or U) was adjusted to a starting pH of 6, 8, 10, 11 and 12. Total polymerization yields are presented after 1, 5 and 10 wet-dry cycles at room temperature. The bottom bar corresponds to the yields for the polymers ending with a 2',3'-cyclic phosphate -cP. The upper bar corresponds to the yields for the polymers ending with a linear 2'- or 3'-phosphate -P. The purine bases polymerize best at pH 10 and the pyrimidine at pH 11. G shows the highest efficiency but a significant fraction of A, U and C monomers also form oligomers by wet-dry cycling at room temperature.

The polymerization is highly pH dependent and is favored for a pH range going from neutral to alkaline, with variations between the different nucleotides (see Figure 3.4). G-cP efficiently polymerizes for a pH range from 7 to 11, with a significant drop in oligomer formation at pH 12, as observed in Figure 3.C.1. Like in Chapter 2, the optimal starting pH was found to be 10, for yields around 60-70 %. The optimal pH for the polymerization of A is also 10, with the formation of oligomers observed at pH 8 to 12, for yields around 9 to 38 %. The pyrimidine nucleobases have an optimal polymerization for pH 11. Oligomers formation is favored over the pH range [10-12], a little more alkaline than for the purine base, with polymerization yields for this pH between 7-38 % for C and 6-31 % for U.

HPLC-MS analysis confirmed that a significant amount of cyclic phosphate groups are still present in the sample after one dehydration (24 h), both for monomers (see Figure 3.6) and polymers (bottom bars in Figure 3.4). At pH 10, for the first cycle, there were still 20 % cyclic phosphate mononucleotides for G, 49 % for A, 68 % for C and 72 % for U. For pH 11, optimal for the polymerization of the pyrimidine bases, 2 % for G, 35 % for A, 23 % for C and 48 % for U. Thus, there were still reactants to continue the reaction for further cycles. The wet-dry cycling enhanced reaction yields at pH 10 up to 51 % for A, 45 % for C and 170 % for U. And longer oligomers also increased in concentration at pH 10, and at pH 11 for C and U. But at pH 12, the formed oligomers degraded after the 1st cycle, due to base-catalyzed degradation of the phosphate-sugar backbone and phosphate hydrolysis.

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

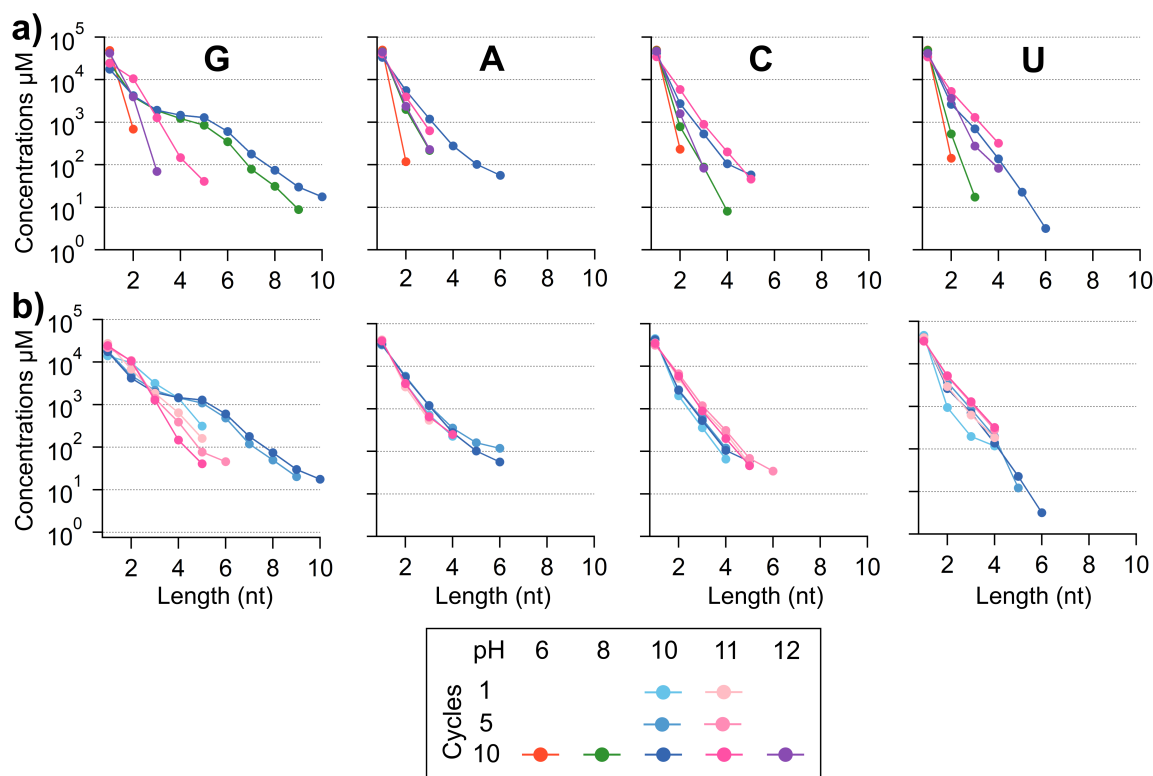


Figure 3.5: Polymers length for G, A, C and U: The results are from the same experiments as for Figure 3.4. A solution of 50 mM N-cP (N = G, A, C or U) was adjusted to a starting pH of 6, 8, 10, 11 and 12. The plots show the concentrations of the oligomer products per length (c-P and 2'-P and 3'-P endings added together) for: **a)** 10th cycle: pH 5, 8, 10, 11 and 12. **b)** 1st, 5th and 10th cycles at pH 10 and 11.

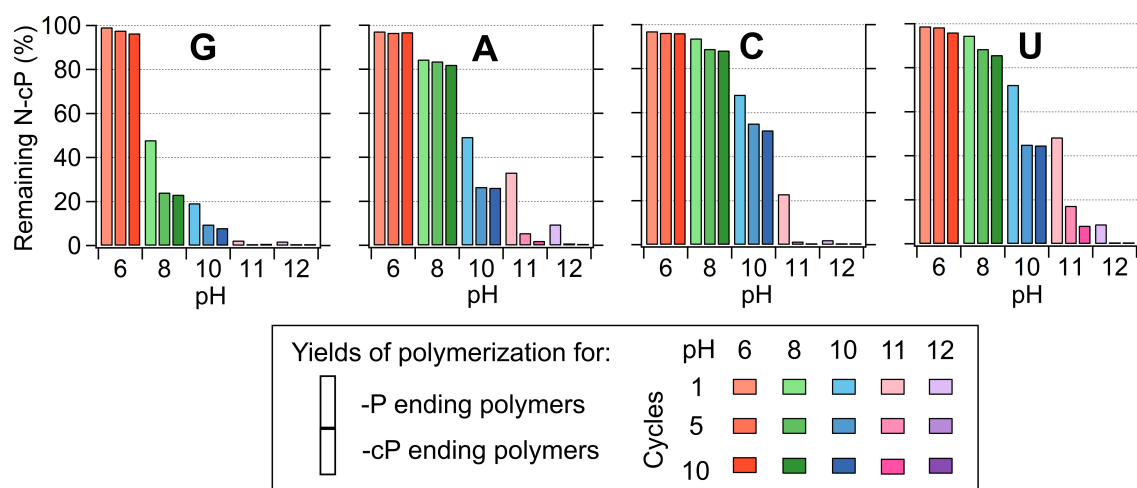


Figure 3.6: Remaining cyclic-ending reactants after reaction at pH 6, 8, 10, 11 and 12, after 1, 5 and 10 wet-dry cycles, for the homopolymerization of G-cP, A-cP, C-cP, U-cp. For pH 8, 10 and 11, a significant portion of the N-cP remain after the first cycle: polymerization can continue in the next cycles.

3.3 Results

Dried phases

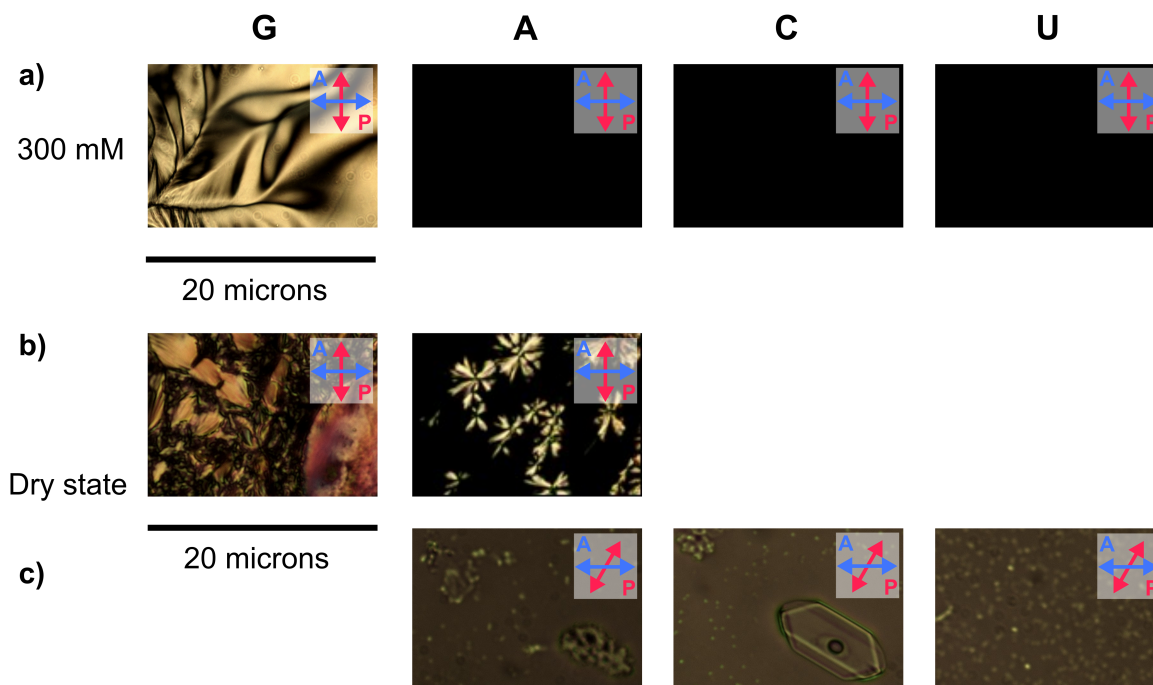


Figure 3.7: High concentration and dry state imaging by microscope: **a)** Microscope images of the samples just before the full drying. For the G samples, the concentration was estimated around 300 mM. Polarizer and analyzer were crossed at 90° . Birefringence textures were detected for G samples, but A, C and U showed only a black background. **b)** Microscope images of the samples G and A after they fully dried. Polarizer and analyzer were crossed at 90° . The C and U samples just showed black background. **c)** Microscope images of the samples A, C and U, fully dried, with the polarizer and analyzer uncrossed. The crystals that can be observed for the C samples are not composed of nucleotides.

Monitoring the drying process by PTOM (Polarized Transmission Optical Microscopy) observation showed that samples reached the full dry state after 8-10 hours (1 cycle is 24 hours), C and U samples remained completely homogeneous and isotropic during the evaporation: only a black background could be detected during (Figure 3.7 **a**) and after (Figure 3.7 **b**) dehydration when the samples were observed under polarized light. The crystals which are shown in **c**) for the C sample are not composed of nucleotides.

In the case of G (or G-containing samples), nucleation and growth of birefringent textures typical of liquid crystals was observed before the full drying. The birefringence textures vary smoothly, characteristic of still fluid liquid crystalline phases. The whole sample showed those textures 5 hours after the beginning of the experiment (8-10 hours for full drying). A transition to sharper high birefringence textures occurred once the G sample reached the fully dried state, a sign of lesser fluidity. In dried state, G nucleotides were in a highly ordered crystalline phase, resulting from the previous liquid crystals.

The images of the A samples at high concentration did not show any birefringence textures

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

that could indicate the presence of liquid crystals. However in the fully dried state, the presence of crystalline domains was confirmed, although it did not covered the full volume of the sample.

Birefringence is a good indicator of molecular ordering. For guanosine, the formation of G-quadruplexes and production of liquid crystal phases is well documented [109]. The assembly of those G tetrads was proposed as a mechanism explaining the enhanced polymerization of G. In that highly-ordered assembly, the functional groups involved in the polymerization would likely be kept in proximity, impacting the reaction efficiency. It would explain why G is not only producing higher yields, but also longer oligomers. As for the formation of A-cP crystals, it can be favored by higher stacking interaction of purine bases with respect to pyrimidines. Nonetheless, those crystalline domains do not seem to improve the polymerization of A compared to C and U as much as it did for G.

Even if the ordering happens already in liquid phase (in the case of G), no polymerization product could be observed before full drying, which is coherent with the results of the previous chapter.

Hydrolysis

The reaction of polymerization which produces diphosphoester bonds is in competition with hydrolysis, which causes phosphate ring opening, producing -2'P and -3'P endings. In Figure 3.4, we can see that already in the first cycle, a significant portion or most of the oligomers produced at pH 10 have a linear phosphate ending and that portion increases with each cycle. At pH 11 and for the 10th cycle, almost all of the oligomers have linear endings: 100 % for G, 94 % for A, 97 % for C and 88 %.

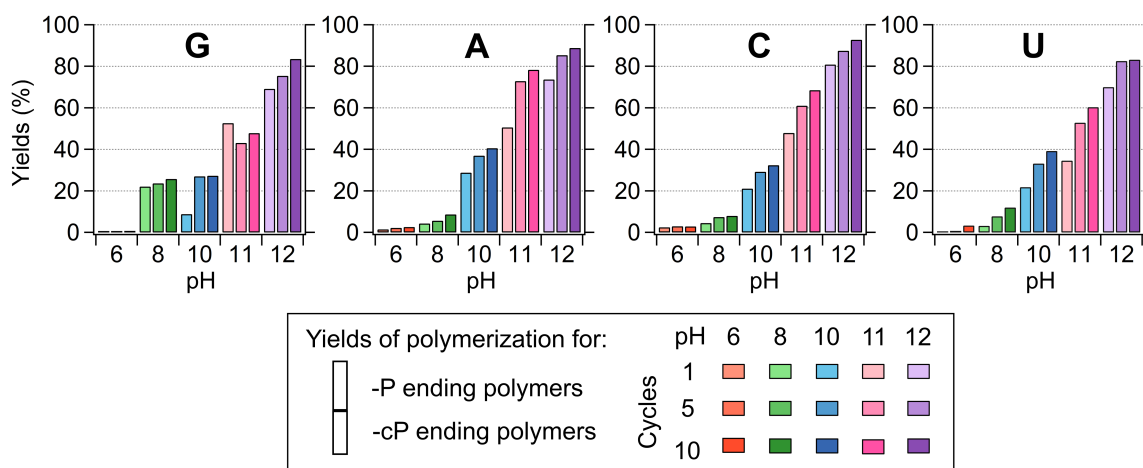


Figure 3.8: Hydrolysis yields for the monomers G, A, C and U: at pH 6, 8, 10, 11 and 12, after 1, 5 and 10 wet-dry cycles, for the homopolymerization of G-cP, A-cP, C-cP, U-cp. The hydrolysis yields of the monomers increase with higher value of pH and number of wet-dry cycles.

3.3 Results

The hydrolysis of the cyclic monomers into deactivated N-2'P or N-3'P was monitored by HPLC-MS measurements. This reaction was shown to be also highly pH-dependent and to increase with the pH and the number of wet-dry cycles conducted for all four nucleotides. The hydrolysis of the monomers was insignificant at pH 6, increased to 2-10 % of the initial monomers (20-25 % for G) at pH 8. For pH > 10, hydrolyzed monomers rose to (50-70 %), and inhibited the reaction of polymerization. For pH 12 in particular, we observed a decrease in polymerization yields and the length of the products, between the 1st, 5th and 10th cycle. We observed phosphate ring openings, but also the degradation of the phosphate-sugar backbone for the products formed in the first cycles.

3.3.3 Effect of wet-dry cycling on mixed polymerization

We performed experiments with mixtures of nucleotides: AU, GC and AUGC and observed the production of mixed products. Nonetheless, the incorporation of the different monomers depends highly on the nucleobase and the initial pH of the solution.

Binary mix: adenosine and uridine

AU mixed sample showed the lowest polymerization yields among the mixed experiments. They peak at 27 % for pH 11, after 10 cycles and oligomers as long as 5mers could be detected (see Figure 3.9).

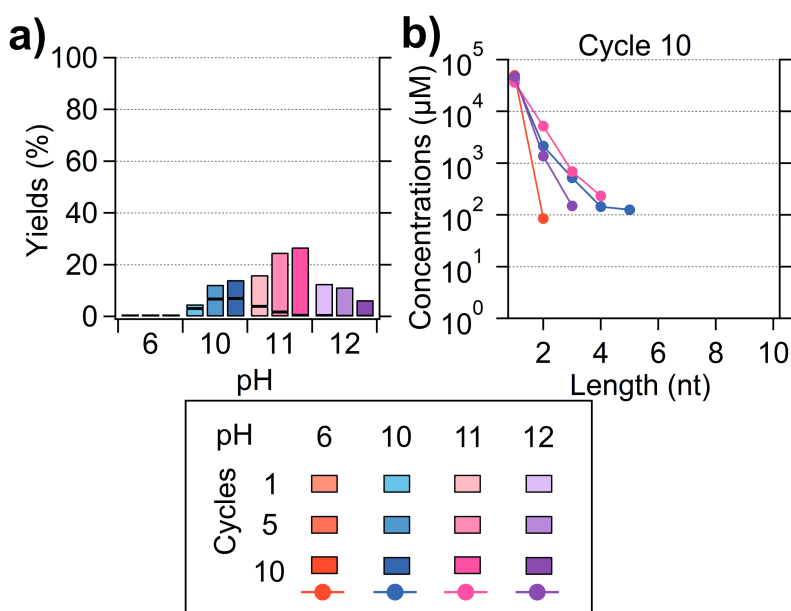


Figure 3.9: AU mixed polymerization yields and products length: A 25 mM A-cP and 25 mM U-cP solution was adjusted to pH 6, 10, 11 and 12 and went through several wet-dry cycles. **a)** Total polymerization yields (with distinction between -cP ending polymers and -P ending polymers). **b)** Concentrations of the oligomers products per length (different AU composition and endings summed up together).

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

I checked the different product compositions by analysis on the mass spectrometer; it must be noted that this analysis method does not discriminate between the different permutations of the same compositions (e.g. dimers AU and UA). Thus, in the theoretical case where both A and U polymerize as efficiently as each other and are randomly mixed together in the dry state, dimers of composition AU should be twice as abundant as AA or UU, trimers of composition AAU and AUU thrice as AAA or UUU, etc. This theoretical behavior is represented by the black line on Figure 3.10 **a)** and **b)**.

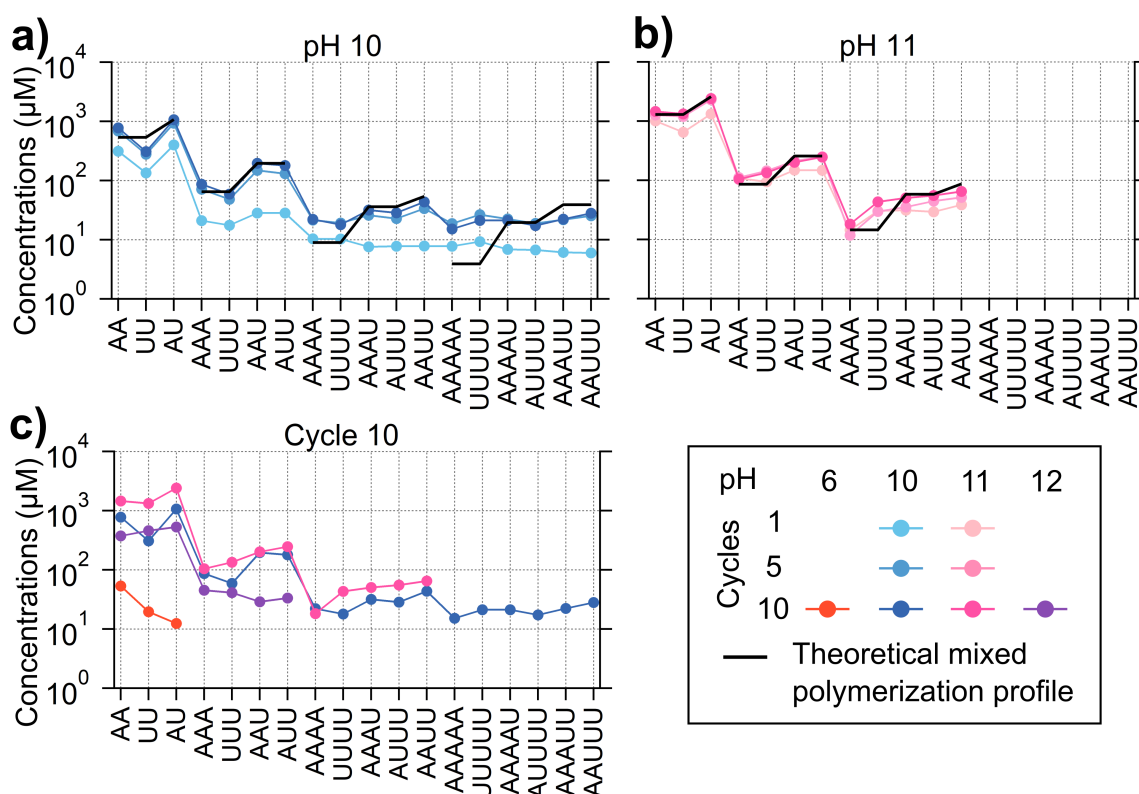


Figure 3.10: AU mixed polymerization products compositions: The oligomers concentrations are given for a starting N-cP concentration of 50 mM. **a)** At pH10, 1st, 5th and 10th cycle. **b)** At pH 11, 1st, 5th and 10th cycle. **c)** 10th cycle, for pH 6, 10, 11 and 12. In **a)** and **b)**, the black line represents the theoretical product composition distribution if A and U polymerize with comparable efficiency and in the same dry phase.

The concentrations per composition can be seen for different pH and cycles in Figure 3.10 **a)** for pH 10, **b)** for pH 11 and **c)** for the 10th cycle at different pH. At pH 10, the A-rich products are slightly favored and for pH 11, it is the U-rich products, which is expected in regard of the optimal pH of polymerization for A and U separately (see Figure 3.4). In spite of those biases, the AU sample shows a behavior close to the theoretical one. The concentrations of the heteropolymers are particularly enhanced by the wet-dry cycling compared to the homopolymers.

3.3 Results

The picture of the dry AU sample in Figure 3.13 shows only an amorphous phase, without any sign of crystalline structures.

Binary mix: guanosine and cytidine

GC polymerization peaks at pH 10-11 for a yield around 40 % and oligomers up to 10mers could be detected at pH 10 (see Figure 3.11).

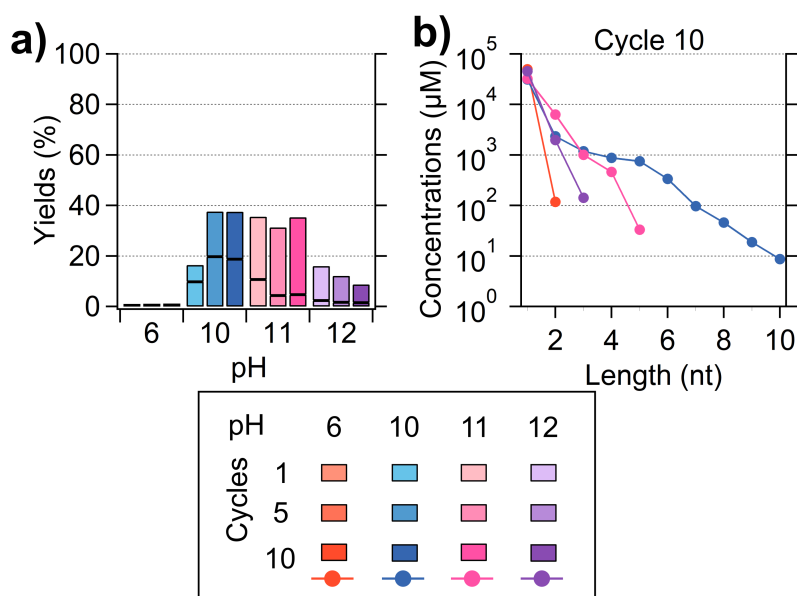


Figure 3.11: GC mixed polymerization yields and products length: A 25 mM G-cP and 25 mM C-cP solution was adjusted to pH 6, 10, 11 and 12 and went through several wet-dry cycles. **a)** Total polymerization yields (with distinction between -cP ending polymers and -P ending polymers). **b)** Concentrations of the oligomers products per length (different GC composition and endings are summed up together).

The product composition is significantly dominated by G (see Figure 3.12) at every pH. Mixed GC polymers are produced, but at lower concentrations than G homopolymers. Nonetheless, pH 11 and 12 make the reaction a bit more balanced between G and C. C polymerizes better at pH 11, and we detected a degradation of the G products after several wet-dry cycles, resulting in a flatter distribution between products of the same length in Figure 3.12 **b)**. The presence of highly ordered crystalline structures was confirmed by PTOM imaging. Unlike the G samples, the birefringent textures in Figure 3.13 do not cover the whole volume of the sample: the crystalline domains formed by G are intercalated with amorphous phases, most likely C-rich. Not only is there a difference in efficiency between G and C, but they also dry as separated phases, lowering the likelihood of mixed products. In Figure 3.12, some mixed products are at lower concentrations than pure C products. For pH 10 and the 10th cycle, GC dimers have a concentration of 317 μM against a concentration of 526 μM for CC. For trimers, I detected 158 μM of GGC, 75 μM of GCC and 126 μM of CCC.

3.3 Results

Quaternary mix

AUGC mixtures polymerize up to 40 % for an optimal pH of 11. However, longer oligomers can be detected at pH 8, 9, 10 (see Figure 3.14).

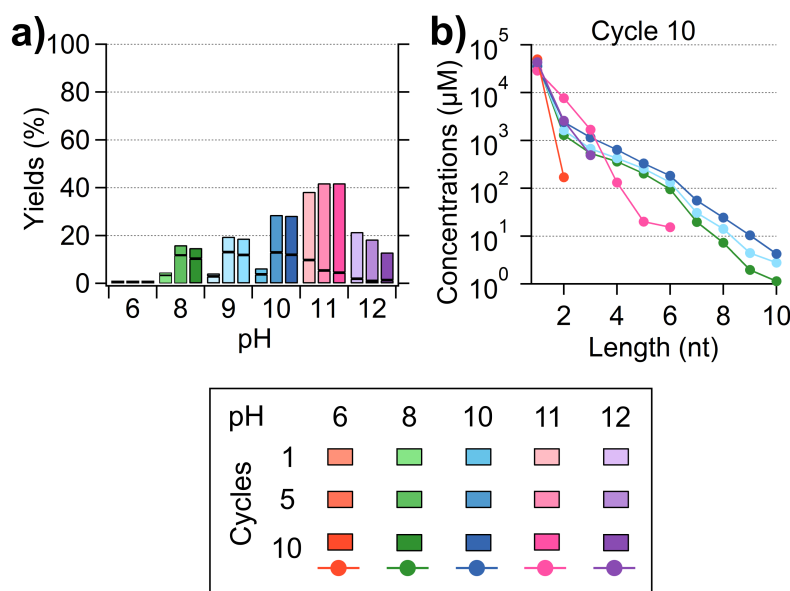


Figure 3.14: AUGC mixed polymerization yields and products length: A 12.5 mM A-cP, 12.5 mM U-cP, 12.5 mM G-cP and 12.5 mM C-cP solution was adjusted to pH 6, 10, 11 and 12 and went through several wet-dry cycles. **a)** Total polymerization yields (with distinction between -cP ending polymers and -P ending polymers). **b)** Concentrations of the oligomer products per length (different AUGC composition and endings summed up together).

For oligomers of a length > 4 , the products are dominated by G. In the 5 and 6mers, for the 10th cycle, up to two monomers other than G were incorporated, one in the 7 and 8mers. For 9 and 10mers, I detected almost only G homopolymers. The composition of 2mers and 3mers, in Figure 3.15 is a bit more diversified: I detected all of them. The distribution is flatter at pH 11 than 10, indicating a lesser bias in favor of G. As for the GC sample, the concentration of the G mixed polymers is lower than the mixed products including the other three nucleotides A, U and C. For the tenth cycle at starting pH 11, I measured a concentration of 728 μM for CA, 706 μM for CU, 957 μM for AU and for the pure dimers 288 μM for CC, 409 μM for AA, 577 μM for UU. In comparison, mixed dimers containing one G had concentrations of 346 μM , 386 μM for GA and 441 μM for GU. For trimers, the average concentration for GGX (X = A, C or U) is 54 μM , 61 μM for GXX and 88 μM for XXX.

The PTOM images of the dried AUGC samples at pH 10, 11 and 12 were recorded for the 6th cycle and are shown in Figure 3.16. Like for the GC dried sample, crystalline structures could be observed, although they do not occupy the whole surface. For pH 10 and 12, the crystalline and amorphous phases are clearly separated, but they are more mixed at pH 11. That more homogeneous phase and lower polymerization of G at pH 11 can explain the flatter product distribution.

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

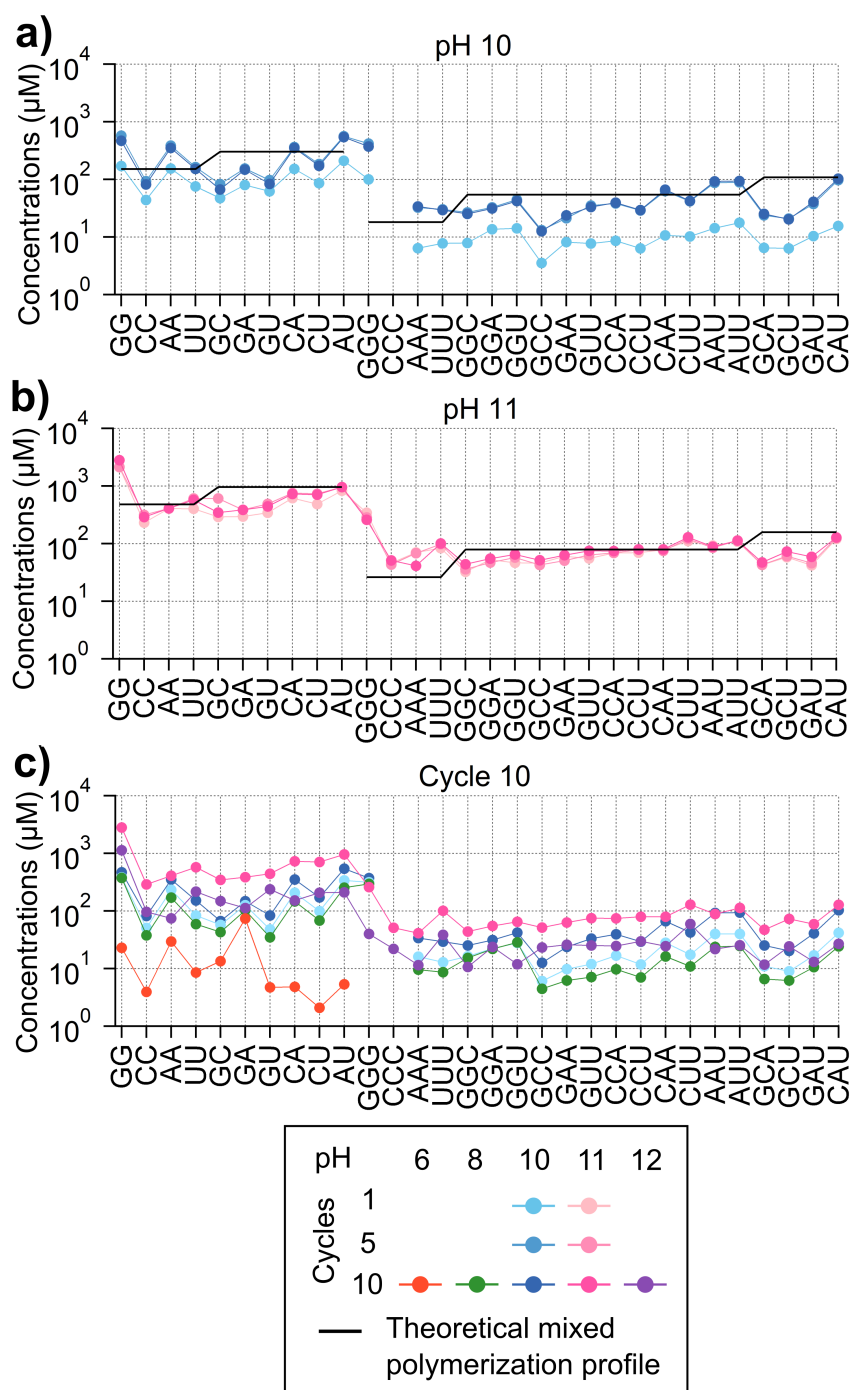


Figure 3.15: AUGC mixed polymerization products compositions: The oligomers concentrations are given for a starting N-cP concentration of 50 mM. **a)** At pH10, 1st, 5th and 10th cycles. **b)** At pH 11, 1st, 5th and 10th cycles. **c)** 10th cycle, for pH 6, 10, 11 and 12. In **a)** and **b)**, the black line represents the theoretical product composition distribution if G, C, A and U polymerize with identical efficiency and together.

3.4 Discussion

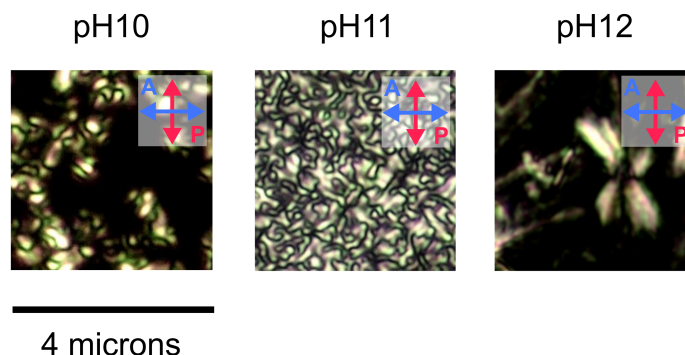


Figure 3.16: PTOM images for AU samples at pH 10, 11 and 12. The samples were observed in the dried state for the 6th cycle. The polarizer and analyzer are crossed at 90°. The birefringence textures indicating the presence of crystalline domains vary depending on the pH.

3.4 Discussion

The data reported in that chapter show the polymerization of G-cP, A-cP, C-cP and U-cP by drying an alkaline solution at room temperature, in the presence of potassium salts, and without any other catalyst. The reaction occurs for a pH between 8 to 12, with a significant decrease at pH 12. The results suggest that the oligomerization proceeds by base catalysis, but so does the hydrolysis. Hydrolysis of the cyclic phosphate groups leads to the more stable products (-2'P and -3'P). For $\text{pH} \leq 10$, the reaction is limited by the lower reactivity of the cyclic phosphate. At $\text{pH} > 10$, it is limited because of the competition of hydrolysis. Hydrolysis happens in aqueous solution, so a pathway to limit hydrolysis would be to lower the drying time, making the wet-dry cycling even more asymmetric in favor of the dry state. We further improve the polymerization by optimizing the frequency of the cycles. The work of Christoph Weber on the theoretical impact of wet-dry cycling on prebiotic reactions points to a "resonance frequency" at which the reaction proceeds with maximal efficiency [100].

I also observed a bias toward G in wet-dry cycling at room temperature: it polymerizes up to 70 % and forms longer oligomers compared to A, U and C. The bias toward G, against the pyrimidine bases, may result from intermolecular interactions such as stacking or hydrogen bonds. This phenomenon facilitates the geometric arrangements conducive to internucleotide phosphodiester bonds. By monitoring the samples by PTOM observation, my collaborators confirmed that the G monomers can self-assemble in a liquid phase at high concentration and form liquid crystals. After full drying, they form crystalline phases, facilitating the polymerization of G into long chains of oligomers. Higher pH, like 11, reduces slightly the bias, but it also reduces the polymerization efficiency of the purine base and increases considerably the hydrolysis.

Previous work has shown that liquid crystals can be formed from oligomers as small as 6-20mers in a random sequence pool where hybridization can occur between different strands

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

[108]. It would cause a liquid-liquid separation between duplexes-rich liquid crystalline domains and the isotropic aqueous solution containing mostly single strands [117]. Liquid crystals were also detected for solutions containing only four-base long self-complementary nano DNA: 5'-ATTA-3' and 5'-GCCG-3' [118]. In the case of AUGC samples obtained for this thesis, we confirmed the presence of crystalline structures. But it is difficult at this stage to verify the contribution of potential short oligomers duplexes compared to the G-quadruplexes. We observed a change in the textures of the crystalline structures of the mixtures at pH 11, in the conditions for which we had the highest concentrations of non-G rich products. But they remained short, 2mer and 3mers mostly, with some 4mers.

Wet-dry cycling at $\text{pH} < 12$ improves the polymerization yields for all nucleotides and the production of mixed samples. With a better incorporation of A, C and U and the production of long oligomers, hybridization, templated ligation and therefore replication become conceivable under conditions compatible with *de novo* polymerization. My colleagues Wunnava and Serrão [85] showed that ligation can already occur for oligomers as small as 8mers in aqueous solution at alkaline pH and low temperatures, by reaction of the 2',3'-cyclic phosphate group.

Addendum 3

3.A Spectral Browser

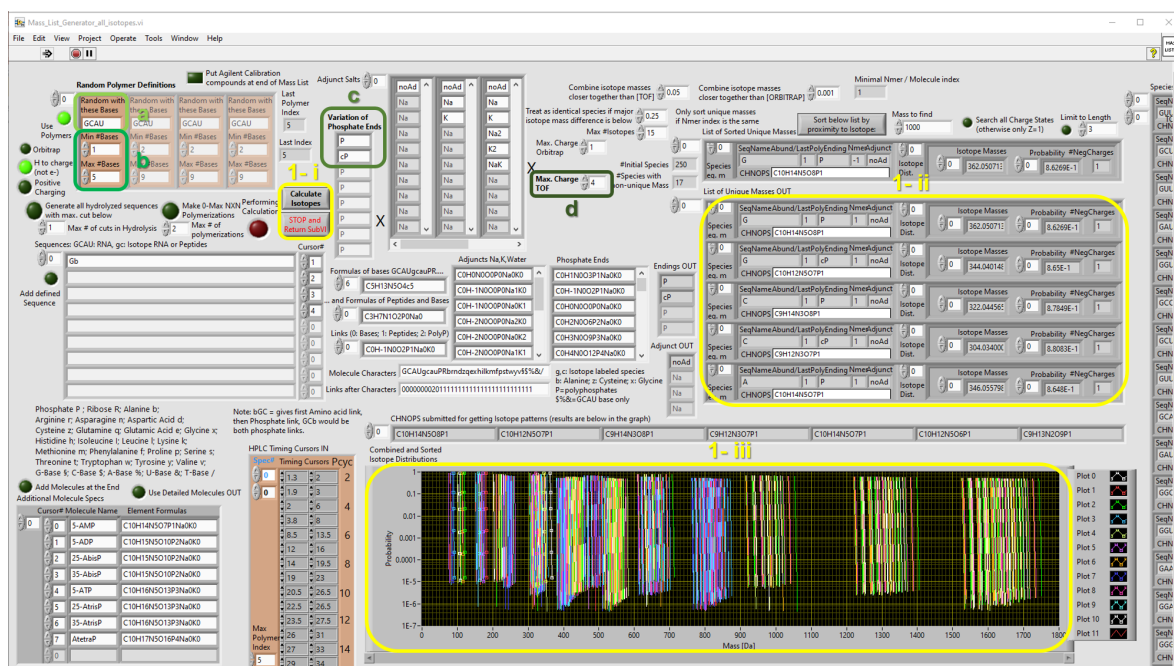


Figure 3.A.1: Masses calculation: A sub-VI called “mass list generator” takes inputs such as a) nucleobase combinations, b) length range, c) type of phosphate ends and charged states to generate (button 1-i) a list of all chemical formulae and the corresponding m/z listed in 1-ii and plotted in 1-iii.

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

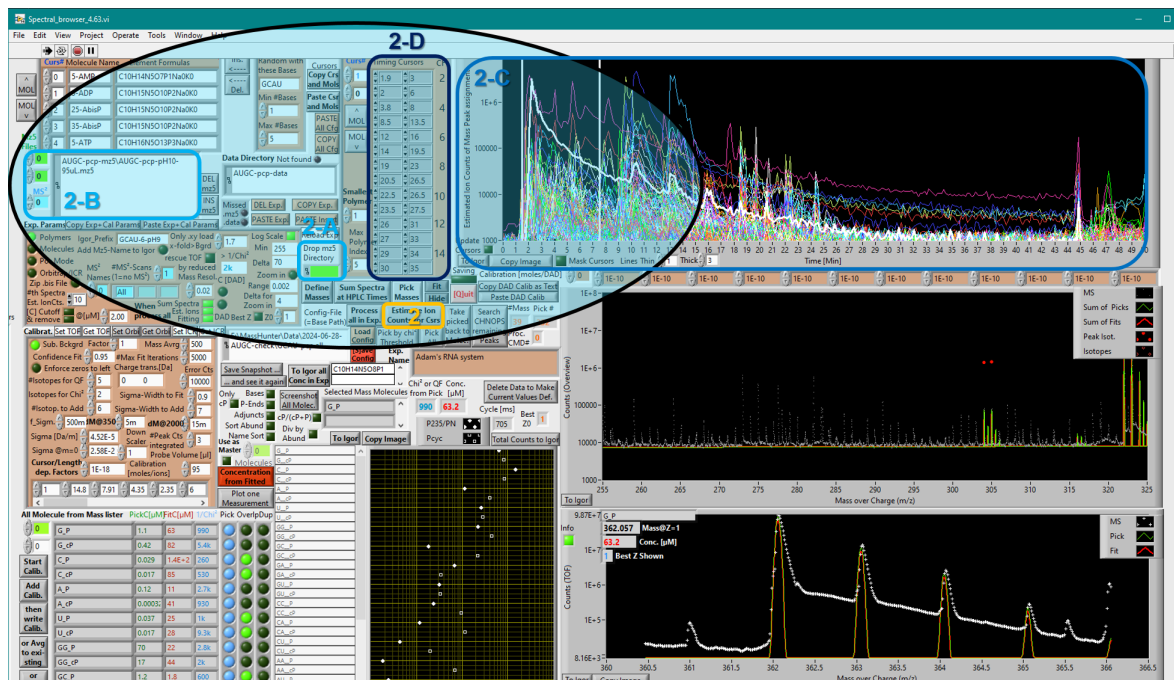


Figure 3.A.2: User interface of Spectral Browser version 4.63 highlighting the pathway of step 2

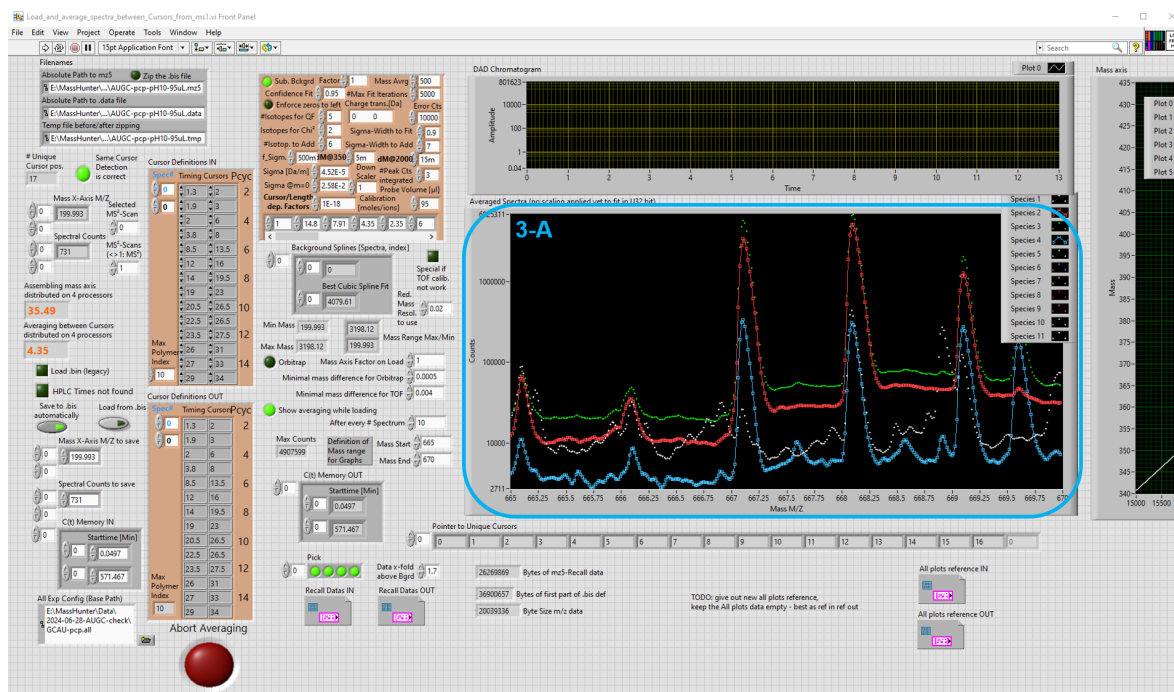


Figure 3.A.3: Front panel of the subVI "Load and average spectra between Cursors from ms1"

3.A Spectral Browser

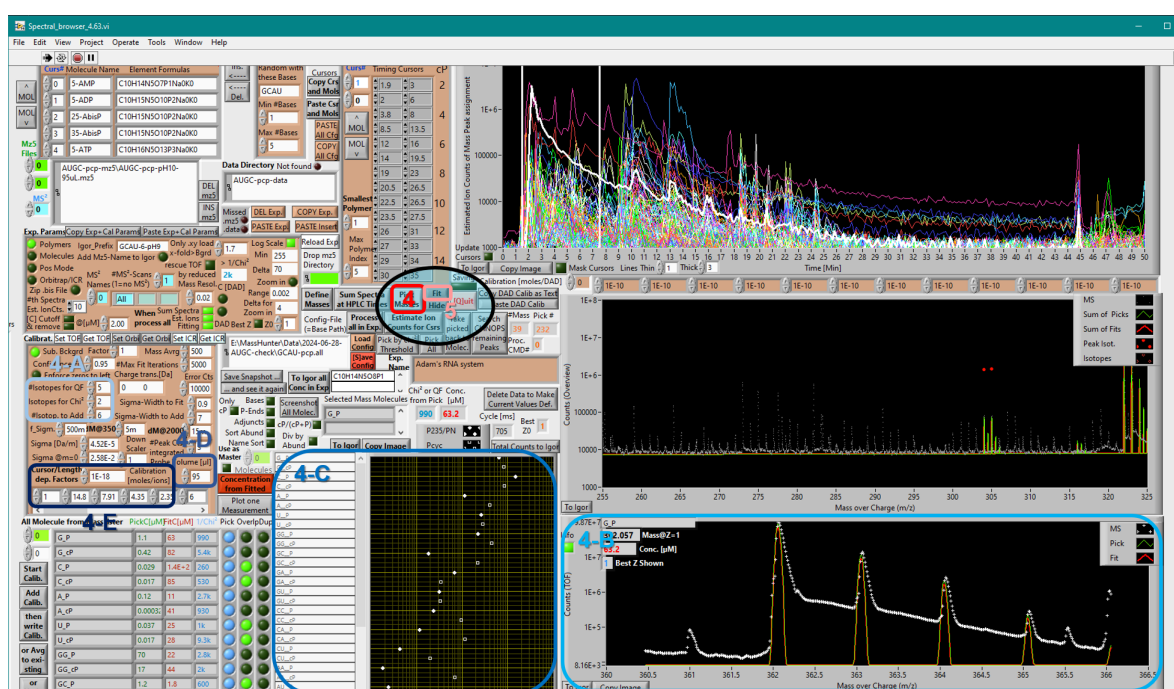


Figure 3.A.4: User interface of Spectral Browser version 4.63 highlighting the pathway of step 4-5

3. Polymerization for G, C, A and U 2',3'-cyclic phosphate nucleotides by wet-dry cycling at room temperature

3.B Standards separation of the HPLC column

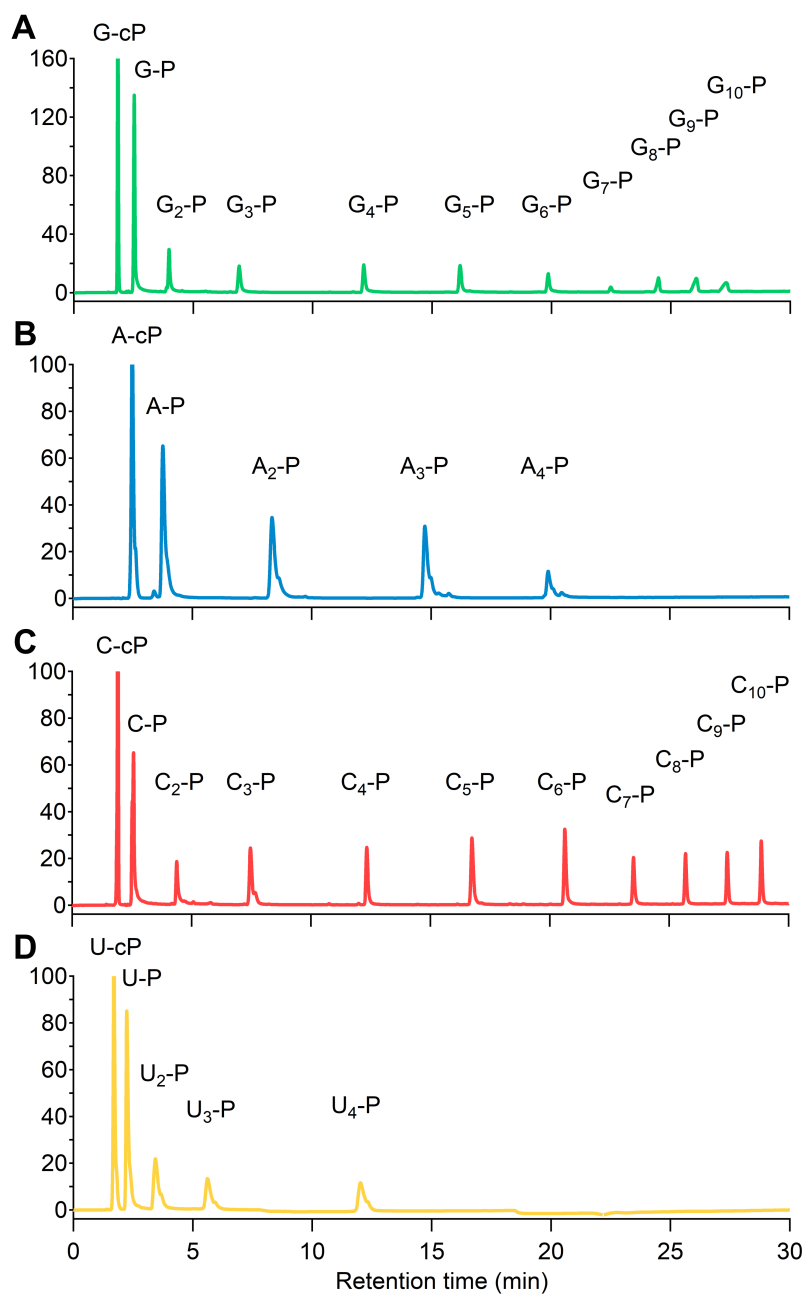


Figure 3.B.1: Calibration standards: UV absorption chromatograms at 260 nm of the four calibration mixes: **a)** G standards. **b)** A standards. **c)** C standards. **d)** U standards.

3.C G polymerization on the pH range neutral to alkaline

3.C G polymerization on the pH range neutral to alkaline

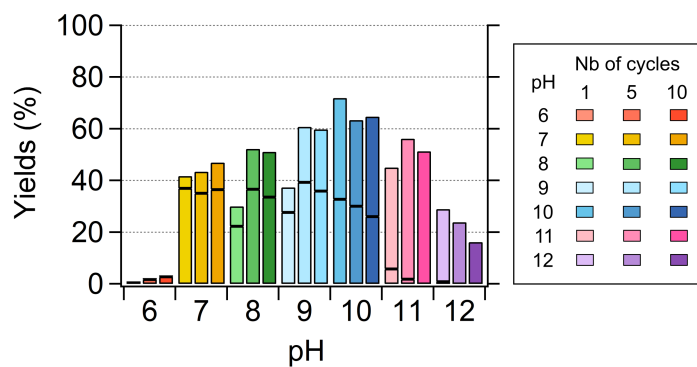


Figure 3.C.1: G polymerization yields for the full pH range 6, 7, 8, 9, 10, 11, 12:

Chapter 4

Formation of 2',3'-cyclic phosphate nucleotides

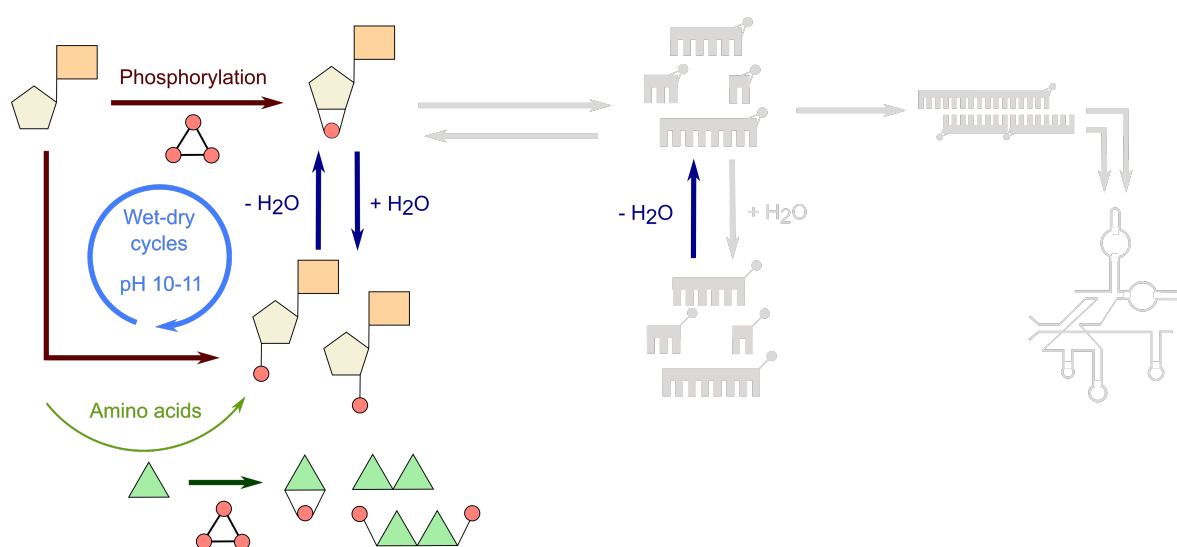


Figure 4.1: Chapter summary: Formation of 2',3'-cyclic phosphate nucleotides by phosphorylation and cyclization at alkaline pH, in presence of catalysts like amino acids

Chapter summary

The formation of N-cP was observed by reaction with trimetaphosphate under alkaline, dried conditions. The formation of dimers was observed under the same conditions. The reaction in aqueous conditions or in the presence of orthophosphate was significantly less efficient. Trimetaphosphate is selective for the 2',3'-diol group of the nucleoside. Cyclization of N-2'/3'P was also observed under these conditions. The yields remain rather low, below 0.4 % of the initial N-cPs. The presence of other small organic molecules, such as urea and amino acids, as well as salts (NH₄Cl) has an impact on the selective formation of the 2',3'-cyclic phosphate. In the case of amino acids, phosphorylated amino acids and peptides were detected in the same experiments, as well as the formation of amino acid-nucleotide hybrid compounds, making a common synthetic pathway for both RNA and oligopeptides plausible.

This project was conducted with the contribution of three Bachelor students: Francesco Testi, Nikolas Wetzel and Maik Krusche, enrolled at that time at the Fakultät für Physik of the LMU Munich.

4.1 Introduction

2',3'-cyclic phosphate nucleotides (N-cPs) are known to be products of enzymatic and non-enzymatic degradation of RNA. But before reaching the first RNA strands, 2',3'-cyclic phosphate groups had to be formed *de novo*. Sutherland and his collaborators [65] proposed a pathway to the pyrimidine nucleotides, for which the main product is the 2',3'-cyclic phosphorylated compound [65]. But in this part, the aim is to explore a pathway to the reactive cyclic phosphate that works for all 4 nucleotides, in particular the purine G. Although 2',3'-cyclic phosphate nucleotides often reported as products of phosphorylation, most studies focus on the canonical 5'-phosphate nucleotide.

Thus, in this third chapter, I explore the formation of the G-cP, most reactive toward polymerization, by direct phosphorylation of the guanosine nucleoside and cyclization of the 2'- and 3'-monophosphate guanosines, which are products of G-cP hydrolysis, under conditions that are compatible with subsequent polymerization.

4.1.1 Prebiotic phosphorylation

The phosphorylation of nucleosides, proteins or lipids is extremely important in extant life and is also one of the most thoroughly studied reactions in prebiotic chemistry. The phosphorylation of an organic substrate and an inorganic orthophosphate is a condensation reaction, which produces one molecule of water [119] (as for the reaction of cyclization of N-2'P and -3'P). To produce N-cP, two molecules of water are produced (see Figure 4.4). In aqueous solution medium, the equilibrium is extremely unfavorable to the products, according to Le Chatelier's principle. Different strategies have been implemented by researchers to overcome that issue, focusing either (or both) on the conditions of the reaction (indicated below by numbers **1-3**) and/or on the phosphorylating agent (indicated below by letters **a-d**). In general, three types of conditions have been explored:

- 1** Aqueous solutions: the reaction is made more efficient by using high concentrations of the reactants and condensing agents that reduce the activity of water.
- 2** Dry or wet-dry cycling conditions: the condensation becomes more thermodynamically favorable once the product in excess – water – is removed from the experiment.
- 3** Non-aqueous solvents.

The other parameter that is modified to improve the yields and/or the regioselectivity of the reaction is the phosphorylating agent:

- a** Inorganic orthophosphate, salts and minerals
- b** Polyphosphates (especially trimetaphosphate)
- c** Other activated phosphate compounds
- d** Reduced phosphorus compounds

4.1 Introduction

High energy phosphate compounds are activated by a functional side group that removes electrons from the phosphorus center, which also plays the role of a good leaving group, as it is much less abundant than water (NH_3 for DAP and pyrophosphate for trimetaphosphate) [120].

The impact of condensing agents and catalysts was explored for all of those conditions.

4.1.2 Formation of 2',3'-cyclic phosphate nucleotide

By direct phosphorylation

(3-a) Direct phosphorylation of nucleoside by orthophosphate salt or minerals in non-aqueous solvent, like formamide [121–123, 123] or semi-aqueous eutectic solvent [124] gives significant yields of the N-cP at high temperature. Different phosphate minerals heated at high temperature in a mixed solution of urea, ammonium formate and water, especially in the presence of ferrocyanide, give substantial amounts of N-cP, as well as other phosphorylated products, after 7 days [125, 126]. The reaction in other solvents makes the condensation easier. However, it seems unlikely that there were significant quantities of a solvent other than water on the early Earth and high temperature causes the degradation of organic solvents [119].

(2-a) N-cP was reported as a phosphorylating agent of orthophosphate, either salt or mineral, in dry conditions. In presence of high quantities of urea and ammonium salts NH_4Cl and NH_4HCO_3 , uridine is converted to 24 % U-cP after drying at 100 °C after 24 h [127]. At lower temperature (66 °C), the yield reaches only 4 %.

(1-b) The 2',3'-cyclic phosphorylated product was also obtained by reaction with TMP in aqueous medium, under very specific conditions. It is reported as a minor product of the four canonical nucleosides with TMP in the presence of 1 M of NaOH, but it quickly hydrolyzes into N-2/3P [128]. At neutral pH, Yamagata and his coworkers [62] obtained a conversion around 4 % after 1 day at 41 °C by phosphorylation by TMP (25 eq), in the presence of high concentrations of Mg^{2+} cations (10 eq).

(2-b) Under dried conditions and dehydration-rehydration cycling, trimetaphosphate gives generally higher yields. For neutral pH, a rate of conversion into A-cP increases up to 10 % in the presence of $\text{Ni}(\text{NO}_3)_2$ after wet-dry cycling (2 weeks) at 37 °C [83]. It was also reported as a side product (around 13 % for U) of the synthesis of 5'-triphosphate nucleotides from the reaction of nucleoside and trimetaphosphate in presence of urea and Ni(II) borate, by heat-drying at 90 °C [129].

(1/2-c) Other phosphorylating agents, such as diamido phosphate (DAP), also produced N-cP. The phosphorylation by DAP produces N-cP in high yields, both in aqueous, dry and wet-dry cycling – "paste" – conditions. In aqueous solution, the highest yields (25-30 % for G, C, A and U) were obtained for a pH between neutral to slightly acidic values, in the presence of metal cations, after several days at room temperature [44, 63]. The reaction also worked at water-air interface, in microdroplets [130].

Cyclization

(1) Cyclization from N-2'/3'P was also reported, especially in acidic conditions [57]. In the presence of acetic acid and the activating agent (aqueous solution) methyl isonitrile, A-3'P was converted to A-cP up to 56 % [131]. The yield was up to 70 % for the conversion of U-2'/3'P into U-cP at acid to neutral pH at 73 °C in the presence of several activating agents, the most efficient being a aqueous solution carbodiimide [132]. Mullen and Sutherland [133], reported an almost complete conversion in aqueous solution at pH 6 and 40 °C in presence of an isocyanide, an aldehyde and NH₄Cl salt.

(1-c) DAP also catalyzes the recyclization of N-cP in aqueous solution [63].

4.1.3 Phosphorylating agent and their availabilities

Orthophosphate is expected to have been the most abundant phosphate species on the early Earth as it is still nowadays, as insoluble minerals such as apatite [123, 134–137]. However, these minerals need non-aqueous solvents [125, 126] and to be in contact with a very acidic aqueous solution [138], where they are more soluble.

Different strategies have been tried to activate the phosphate: using condensing agent and/or by adding a better leaving group to the phosphate, such as DAP [44,63] or imidazole-phosphate [99]. However, the prebiotic plausibility of these reactants and their potential sources on the early Earth are unknown. Imidazole phosphate can be obtained in high concentrations of cyanate [99] or in the presence of carbodiimides [45]. Krishnamurthy and his coworkers [139–141] proposed a synthesis of DAP by reaction of TMP in high concentration of ammonia. At lower pH, DAP tends to convert back into TMP [63]. But the prebiotic feasibility of these syntheses is still being discussed.

Polyphosphates are also more reactive and cyclic trimetaphosphate is one of the main reagents used in prebiotic phosphorylations. It is also still used in extant life, notably as energy storage, and thus the role of inorganic polyphosphates in some modern organisms could be a indication of their role during the emergence of life [18, 19]. Polyphosphates also tend to be scarce under normal conditions [142]. Although naturally occurring minerals containing pyrophosphate have been discovered [143], pyrophosphate is not considered an effective phosphorylating agent [120, 144]. Some potential sources of TMP on the early Earth have been proposed: it has been detected in phosphorus volcanic material, produced by hydrolysis of P₄O₁₀ during the cooling process [145]. Condensed phosphates are stable in alkaline glasses containing some low amounts of Al₂O₃ [146,147]. Prebiotic synthetic pathways have also been proposed for the formation of polyphosphate and trimetaphosphate [148]. Pasek and his collaborators [149] reported the phosphorylation of nucleosides A and U by mild heating of the aqueous solution in presence of analogs of the mineral schreibersite found in meteorites. Schreibersite (Fe, Ni)₃P provides phosphorus (III) compounds, more reduced compared with the phosphorus (V) at the center of the phosphate. Those reduced phosphorus compounds also offer a synthetic path to polyphosphates and in particular, TMP [134, 150]. Trimetaphos-

4.2 Materials and methods

phate also presents one more considerable advantage over other phosphorylating agents: it shows regioselectivity toward the 2' and 3'-OH position in alkaline aqueous solution [151].

4.2 Materials and methods

The experiments are carried out by preparing solutions of different pH and by drying at different temperatures on glass slides. The samples are then collected by rehydration with water and then injected in a reversed-phase HPLC and the results are quantified by UV absorbance at 260 nm and mass spectrometry in negative mode.

4.2.1 Reactants and standards

G nucleoside and amino acids were purchased from Sigma-Aldrich.

Cyclic nucleotide G-cP and the linear-ending monomers G-2'P and G-3'P were purchased from BioLog in their sodium salt form, as lyophilized samples.

Standard pre-synthesized homopolymers of G, 2 and 3mer, with a 3'-phosphate ending (and -OH at the 5' end) and no phosphate ending were purchased from Biomers as lyophilized samples.

Glass slides were purchased from Carl Roth.

4.2.2 Sample preparation

Direct phosphorylation

A solution of 1 mM G nucleoside and 10 mM sodium salt of trimetaphosphate is prepared in RNase-free water, eventually with other organic molecules such as urea, amino acids, at different concentrations (1, 3, 10, 30, 100 mM) and the salt HN_4Cl . The pH of the solution is then adjusted to different pH using 1 M HCl and either 1 M KOH or NH_4OH . 50 μL of the solution are put on a glass slide with three 100 μL dips. The slides are put on a heating plate and protected by a cover. The solution is left to dry at different temperatures for 1 day, or undergoes several wet-dry cycles.

The dried material is solubilized in 2x50 μL . 20 μL of the collected sample are injected in HPLC-MS for analysis.

By cyclization

A solution of 1 mM equimolar mix of G-2'P and G-3'P, 10 mM sodium salt of trimetaphosphate is prepared in RNase-free water, eventually with other organic molecules such as urea, amino acids, at different concentrations (1, 3, 10, 30, 100 mM) and the salt HN_4Cl . The pH of the solution is then adjusted to different pH using 1 M HCl and either 1 M KOH and NH_4OH . 50 μL of the solution are put on a glass slide with three 100 μL dips. The slides are put on a heating plate and protected by a cover. The solution is left to dry at different temperatures for 1 day, or undergoes several wet-dry cycles.

4. Formation of 2',3'-cyclic phosphate nucleotides

The dried material is solubilized in 2x50 μ L. 20 μ L of the collected sample is injected in HPLC-MS for analysis.

Wet-dry cycling

The rehydration is done using 50 μ L pure water without pH adjustment. At the end of the last cycle, the dried material is solubilized in 2x50 μ L. 20 μ L of the collected sample is injected in HPLC-MS for analysis.

4.2.3 LC-MS instrument specifications

The measurements were performed on a high-performance liquid chromatography HPLC (Agilent 1260 Infinity II) coupled to an electrospray ionization time-of-flight mass spectrometer (Agilent 6230B with dual AJS EIS). The column used was a Agilent Advance Oligonucleotide C18 Column (4.6 x 150 mm 2.7 micron) heated at 30 °C and the nucleotides and oligomers were separated by length using ion-pairing reversed-phase HPLC. The eluent consisted of mixtures of water (Bottle A) and methanol (Bottle B: 50 % water, 50 % methanol) containing each 8 mM trimethylamine (TEA) and 200 mM hexafluoroisopropanol (HFIP), with a gradient elution at a flow of 0.6 mL/min. The method started with 1 % of B for 4 minutes, followed by several gradients increasing from 1 % to 4 % of B over 3 minutes, to 8 % over 3.2 min, to 15 % over 4.8 min, to 30 % over 5 minutes and then to 50 % for 5 minutes. Then, the column was flushed with 100 % B for 7 minutes before being returned to 1 % for 8 minutes, to re-equilibrate the column.

Detection of eluted mononucleotides and oligonucleotides was achieved using ESI-TOF in negative mode (employing specific source parameters: Gas temperature: 325 °C, Drying gas flow: 12 l/min, Sheath gas temperature: 400 °C, Sheath gas flow: 11 l/min, VCap: 3500 V, Nozzle Voltage: 2000 V) and Diode Array Detector (DAD) WR (wavelength used: 260 nm).

4.2.4 Standards and quantification

The standards used for the quantification of the products are shown in Figure 4.A.1 and they are well separated on the HPLC column (see Figure 4.2). A standards mixture is prepared and injected in the HPLC at different volumes: 1, 2, 3, 6, 10, 20, 30, 100 μ L.

Reactants quantification

- G-noP for phosphorylation experiments

Nucleosides are not easily ionized in the negative mode mass spectrometer. Thus, its extracted ion count is not linearly dependent on the quantity of nucleosides. There, I use the UV absorbance at 260 nm of the nucleobase to quantify the nucleoside. The HPLC separated well the nucleoside from the other compounds at 2.2 minutes, giving a clean peak which does not overlap with other molecules that also absorb at 260 nm. I calculated the area of the peak corresponding to the nucleoside using the Agilent MassHunter software Qualitative Analysis Navigator version B8.00 and found it to be linearly dependent on the quantity of the molecule.

- G-2/3P for cyclization experiments

4.2 Materials and methods

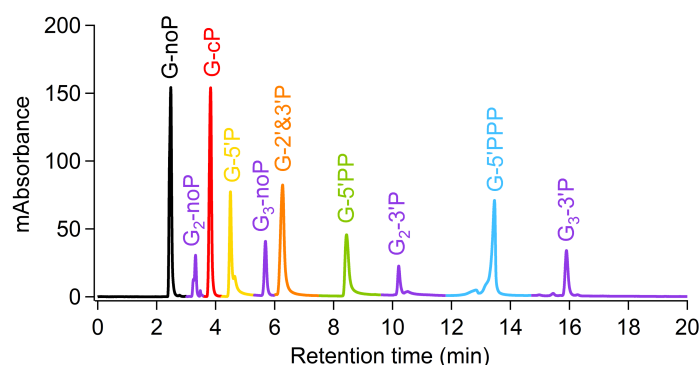


Figure 4.2: Separation of the standards on the column: UV chromatogram (absorption at 260 nm) of the standards used for identification and detection – G nucleoside (G-noP), dimer G with no phosphate ending (G₂-noP), 2',3'-cyclic phosphate nucleotide (G-cP), 5'-monophosphate nucleotide (G-5P), trimer G with no phosphate ending (G₃-noP), 2'/3'-monophosphate nucleotide (G-2/3P) are not separated with that HPLC method), 5'-diphosphate nucleotide (G-5PP), dimer G with a 3'-monophosphate ending (G₂-3P), 5'-triphosphate nucleotide (G-5PPP), trimer G with a 3'-monophosphate ending (G₃-3P). Their molecular structures can be seen in Figure 4.A.1.

For quantities > 1 nmol, the extracted ion counts of 2'/3'-monophosphate nucleotides are not linearly dependent on the quantity. Thus, I used the area of its UV absorbance (at 260 nm) peak corresponding to those compounds in the chromatogram. It should be noted that this LC method does not separate 2'- and 3'- monophosphate nucleotides.

Products quantification

The quantification was done using the same Spectral Browser as in the previous Chapter 3, therefore it was not described in this chapter.

- Standards compounds that are commercially available

For quantities below < 1 nmol, the extracted ion counts of the phosphorylated guanosine mononucleotides and oligomers are linearly dependent on the quantities injected in the HPLC. The calibration curves and quantities of the molecules are calculated using the Spectral Browser program (see Section 3.2.4).

- Compounds without any commercially available standards

Among the products of reactions, there were guanosine derivatives that were not commercially available: 2'/3'-diphosphate guanosine, 2'/5'- or 3'/5'-bisphosphate guanosine and the equivalent for the triphosphate. They are quantified in this chapter using the calibration coefficient of another standard (see table 4.1).

4. Formation of 2',3'-cyclic phosphate nucleotides

Table 4.1: Calibration standards for the products without commercially available standards

Compounds	Standards used for quantification
G-PP (2'/3' or bisphosphate)	G-5PP
G-PPP (2'/3' or and other isomers)	G-5PPP
G ₂ -cP	G ₂ -3P

Valine-guanosine hybrid compound

In the experiment of direct phosphorylation of G nucleoside by TMP in presence of valine put under wet-dry cycling, the quantity of phosphorylated guanosine covalently bound to a valine was estimated using their UV absorbance peak area at 260 nm. Quantification by mass spectrometry was not possible since the covalent bond between the amino acid and the phosphate changes the ionization of the compound.

Yields calculations

The yields are calculated over the initial concentration of G reactants (nucleoside for direct phosphorylation and monophosphorylated nucleotides for the cyclization).

With $[G]_0$ defined as the initial concentration of guanosine reactants:

- the total conversion yield of G (in %) is defined as:

$$Y_{tot} = ([G - cP] + [G - 2/3P] + [G - 5P] + [G - PP] + [G - PPP] + 2 * [G_2 - noP] + 2 * [G_2 - 3P] + 3 * [G_3 - noP] + 3 * [G_3 - 3P]) / [G]_0 \quad (4.1)$$

- the conversion yield (in %) for a phosphorylated mononucleotide G-X, with X representing the phosphate ending (G-cP, G-2/3P, G-5P, G-PP or G-PPP) is defined as:

$$Y_{G-X} = [G - X] / [G]_0 \quad (4.2)$$

- the polymerization yield (in %) of G (taken into account the main detected products G₂-noP, G₂-3P, G₃-noP and G₂-noP) is defined as:

$$Y_{pol} = (2 * [G_2 - noP] + 2 * [G_2 - 3P] + 3 * [G_3 - noP] + 3 * [G_3 - 3P]) / [G]_0 \quad (4.3)$$

For the experiments carried out in the presence of amino acids, the formation of amino acid-nucleotide hybrid compound was not taken into account in the total reaction of G (see Figure 4.8 and Figure 4.11). Their yields were estimated to be below 1 % of the initial concentration of guanosine (see Figure 4.14).

4.3 Results

I compared the two most prominent phosphorylation agents under dried conditions at different pH and different temperatures. While inorganic orthophosphate (P_i) remained unreactive, phosphorylation by trimetaphosphate (TMP) produced some amount of G-cP in acidic

4.3 Results

and alkaline pH and by heat drying. Dimers were detected at alkaline pH. The yields remain low, thus I tested several small organic molecules as condensing agents and catalysts, such as urea and amino acids. Urea and valine improved the yields of phosphorylation and in particular of the cyclic phosphate product.

Cyclization of the 2'/3'-monophosphate guanosine was also an important goal. In the previous Chapters, hydrolysis of the N-cP reactants deterred the polymerization reaction as it caused the deactivation of the cyclic phosphate group after several wet-dry cycles, even at room temperature. Under the same conditions as the phosphorylation, I observed some monophosphorylated nucleotides converted by TMP into cyclic phosphate, especially in the presence of lysine and proline.

4.3.1 Direct phosphorylation of the nucleoside

Based on the literature about the formation of 2'/3'-monophosphate guanosine (see section 4.1.2), I tested and compared two phosphorylating agents – inorganic phosphate P_i and trimetaphosphate TMP – under conditions compatible with further polymerization. A solution of 1 mM G nucleoside and 10 mM phosphorylating agent was prepared, adjusted to different pH and dried at different temperatures. From the results of the previous Chapters 2 and 3, I know that subsequent polymerization happens only in dried state and the literature about prebiotic phosphorylation shows that the reaction is globally more efficient in dried state than in aqueous solution. I tested the reaction of G nucleoside, first without any additional catalyst or condensing agent, to keep the conditions as simple as possible.

The phosphorylating agent needed to be both reactive and regioselective for the 2'/3'-OH position on the ribose. Theoretically, direct phosphorylation on the nucleoside can yield several products: 2'-phosphate, 3'-phosphate (2',3'-cyclic) or 5'-phosphate. In particular, I wanted to avoid the formation in large amount of the product G-5'P as it is a product which cannot react with G-cP to form polymers.

With orthophosphate

On early Earth, it is estimated that almost all of the phosphate were orthophosphate minerals. I tested P_i sodium salt first as a phosphorylating agent. A solution of 1 mM nucleoside and 10 mM Na_2HPO_4 was adjusted to either acidic, neutral or alkaline pH and was dried at temperatures on the range 25-100 °C. On Figure 4.3, I observed that orthophosphate is mostly unreactive under those conditions, without any condensing agent. At pH 3, monophosphorylated nucleotides are produced at very low yields (at 80 °C, 0.2 % G-5'P and G-2/3P, < 0.02 % G-cP). The yields increased with drying temperature, but remained below 1 %. Almost no phosphorylation product was detected at pH 10, regardless of the temperature.

The formation of G-cP by P_i produces two molecules of water (see Figure 4.4), which makes the reaction thermodynamically unfavorable as some water molecules most likely remained in the environment even after full drying. To make the reaction more efficient with orthophosphate, I would need much higher temperatures and large quantities of condensing agent. Both are rather incompatible with the subsequent polymerization of G-cP.

4. Formation of 2',3'-cyclic phosphate nucleotides

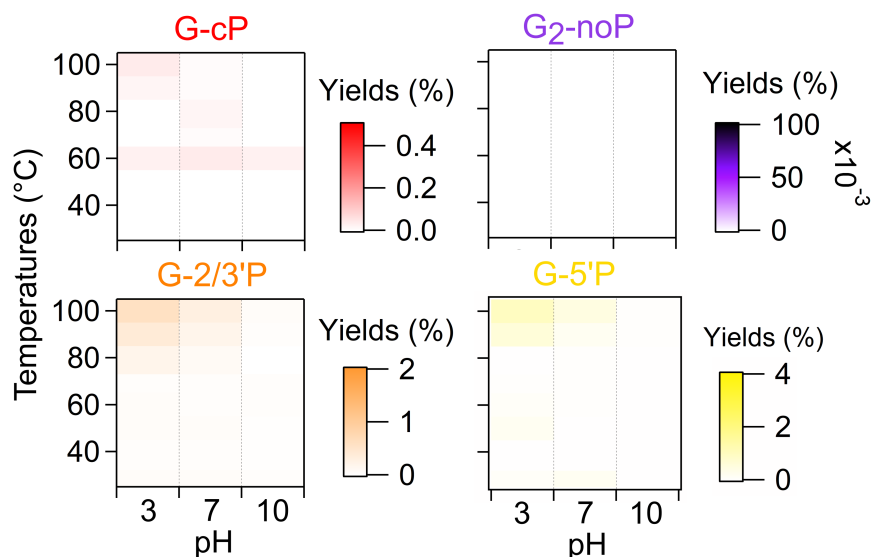


Figure 4.3: Phosphorylation of G nucleoside by orthophosphate, temperature and pH screening: A solution of 1 mM G nucleoside and 10 mM NaH₂PO₄ was adjusted either at pH acidic, neutral or alkaline by NaOH and HCl at different pH. The results are shown as yields for the four main products: the monophosphorylated products G-cP, G-2'/3'P, G-5'P and the first product of polymerization G₂-noP.

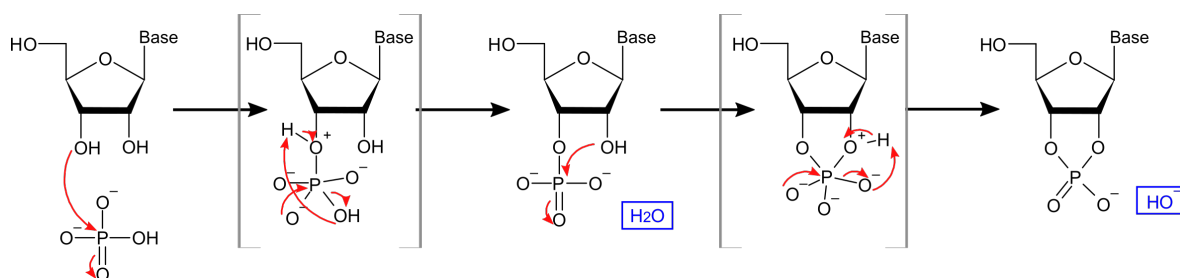


Figure 4.4: 2',3'-cyclic phosphate nucleoside (N-cP) synthesis by the phosphorylation of a nucleoside by orthophosphate at pH 10.

4.3 Results

With trimetaphosphate

Apart from inorganic orthophosphate, trimetaphosphate is the most studied phosphorylating agent in literature. It has been shown to be selective toward the 2'-/3'-OH positions on the ribose in highly alkaline aqueous solutions. In addition, its reaction on nucleoside (see Figure 4.5) does not produce any water molecule: the first phosphorylation causes the opening of the trimetaphosphate cycle, and the second one gives pyrophosphate as secondary product.

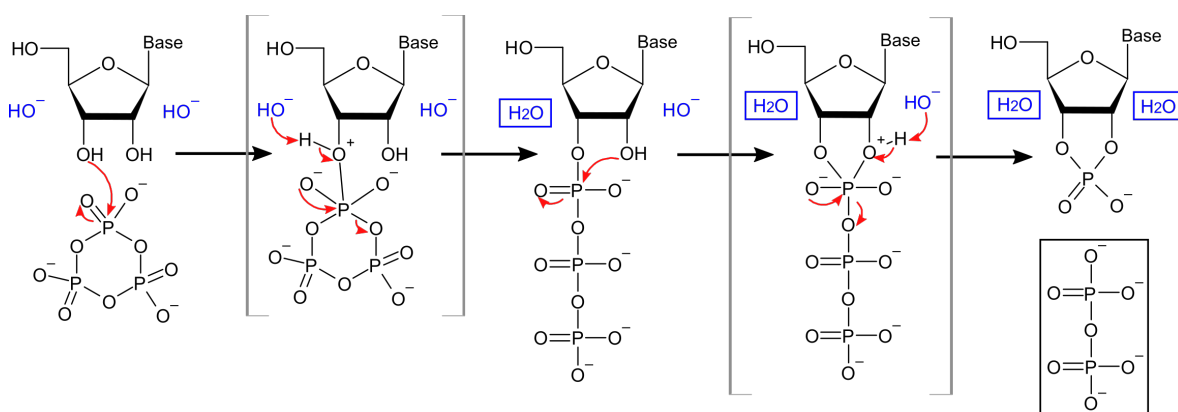


Figure 4.5: 2',3'-cyclic phosphate nucleoside (N-cP) synthesis by the phosphorylation of nucleoside by trimetaphosphate at pH 10 The reaction produces no water molecule but one molecule of pyrophosphate.

In the conditions tested in Figure 4.6, the phosphorylation yields are much higher. Trimetaphosphate is reactive at both acidic and alkaline pH. At pH 3 and 80 °C, monophosphorylated products yields were ten times higher than with orthophosphate (4 % G-5'P, 2 % G-2'/3'P, 0.4 % G-cP). Small amounts of di- and triphosphorylated products were also detected. At pH 10, the formation of G-5'P decreases significantly, below 0.5 %, while the yield of G-cP remained comparable. In addition, at alkaline pH, I detected the dimer GG-noP (without phosphate ending), which most likely results of the reaction between the produced G-cP and the remaining nucleoside.

In the following sections, I have used trimetaphosphate as a phosphorylating agent. Although more efficient than the reaction using orthophosphate, the phosphorylation by trimetaphosphate on G nucleoside gives low yields, below 0.5 % of G-cP. I tested several potential condensing agents and catalysts, such as urea and amino acids, to improve the yields.

4. Formation of 2',3'-cyclic phosphate nucleotides

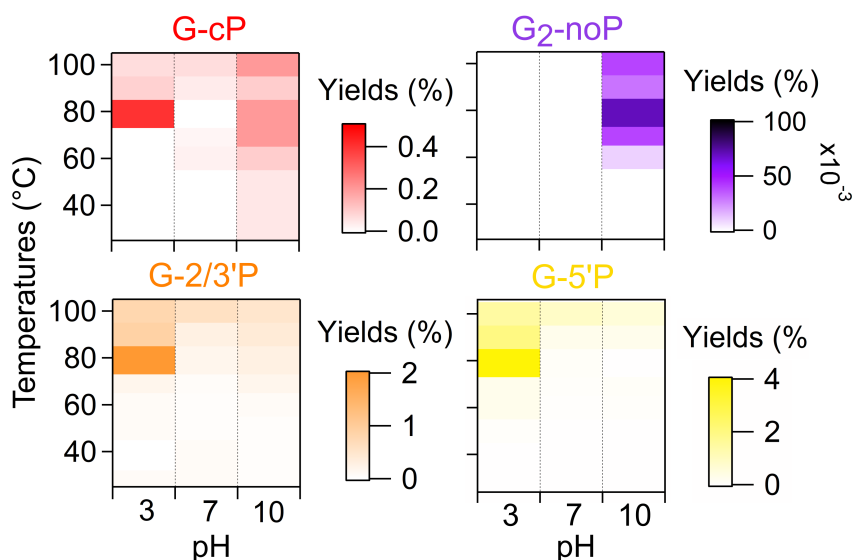


Figure 4.6: Phosphorylation of G nucleoside by trimetaphosphate, temperature and pH screening: A solution of 1 mM G nucleoside and 10 mM $\text{Na}_3\text{P}_3\text{O}_9$ was adjusted either at pH acidic, neutral or alkaline by NaOH and HCl at different pH. The results are shown as yields for the four main products: the monophosphorylated molecules G-cP, G-2'/3'P, G-5'P and the first product of polymerization G₂-noP.

In presence of urea

Urea is widely used as a condensing agent for prebiotic phosphorylation. Orgel and his co-workers [127] reported the production of N-cP in presence of orthophosphate, ammonium salt and high quantity of urea, by drying at rather high temperatures (66-100 °C). I tested used urea in presence of trimetaphosphate, and observed how it impacted the reaction for the different pH (see Figure 4.7).

At acidic and neutral pH, the presence of urea enhances 2', 3' and 5' linear phosphorylated products, as well as the di/bisphosphorylated products (see Figure 4.6).

At alkaline conditions, I detected higher yields of G-cP, 1.1 % at 60 °C and 1.7 % at 90 °C, for 10 mM urea. It is noticeable that higher amounts of condensing agent do not necessarily result in higher yields. When the sample is dried at 60 °C, G-cP is significantly enhanced by the presence of the condensing agent urea – around 20-fold higher compared to the control experiment without urea, against 11-fold enhancement for the G-2'/3'P and 3x for G-5'P. Polymerization product yield was also improved by addition of urea: the dimers GG-noP, GG-cP and GG-P and the trimers GGG-noP were detected for reaction at pH 11 with drying at 60 °C and 90 °C. The yield of G-cP is higher for 90 °C, but higher temperature also favors the linear product 2'/3'-monophosphate.

Even in presence of urea and at alkaline pH, TMP is rather unreactive at low temperature, which can be an issue as polymerization proceeds better at room temperature.

4.3 Results

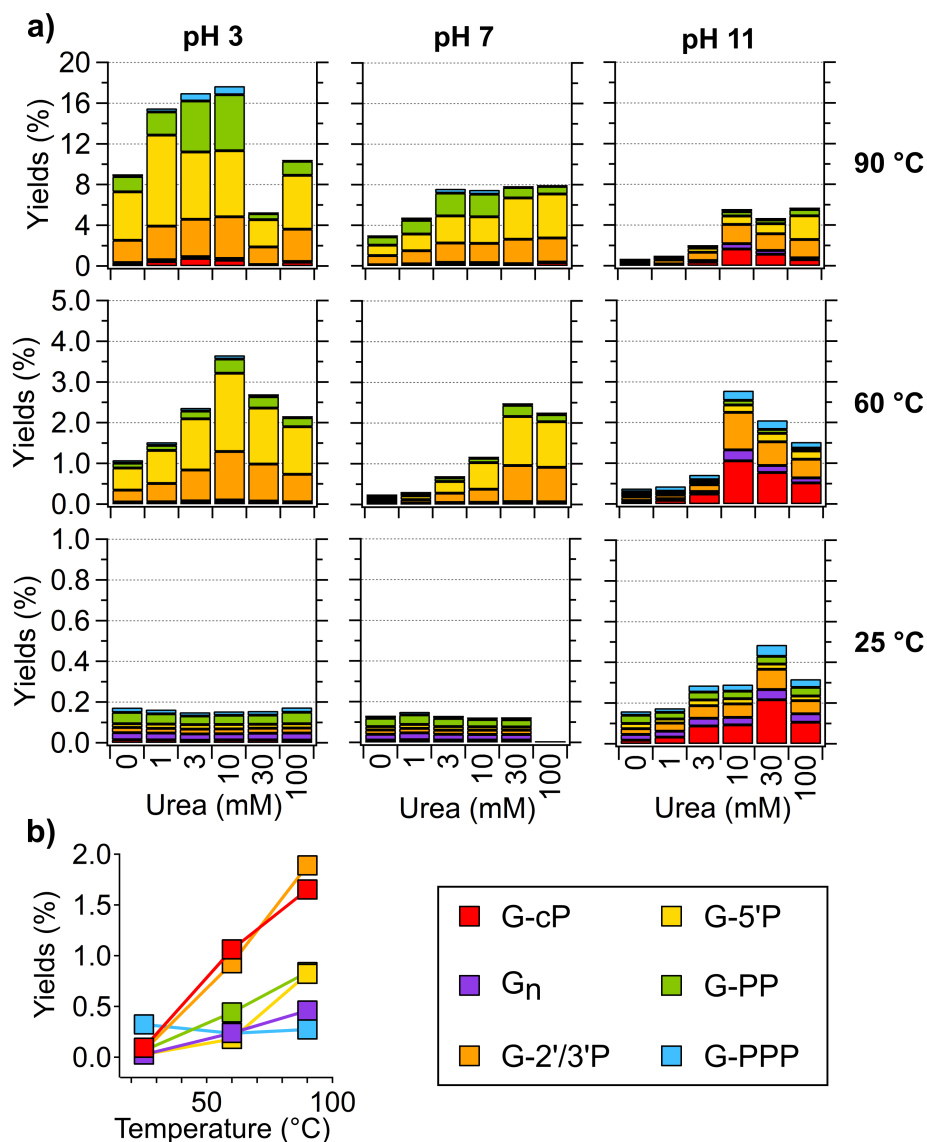


Figure 4.7: Screening of the condensing agent urea for the phosphorylation of nucleoside: at different pH, drying temperatures and quantities. Solutions of 1 mM guanosine nucleoside, 10 mM TMP, 10 mM NH_4^+ and different concentrations of urea (0, 1, 3, 10, 30 and 100 mM) are adjusted to pH 3, 7 and 10. 50 μL are dried at room temperature (25 °C, 60 °C, 90 °C). **a)** Yields of polymerization for each of the products and total yields of nucleoside reaction for each condition. **b)** Effect of temperature on the formation of the products at pH 11. The best conditions for the G-cP (and its subsequent dimerization) are 10 mM urea at pH 11 and drying at 60-90 °C.

In presence of amino acids

Amino acids have been tested in some studies as a catalyst for the nucleoside phosphorylation. Glycine, its dimers and trimers, as well as aspartic acid, were tested as additional catalysts for the synthesis of A-cP by reaction of nucleoside with trimetaphosphate in presence of Mg^{2+} in aqueous solution, but seemed to have little impact on the reaction [62]. The

4. Formation of 2',3'-cyclic phosphate nucleotides

work of my colleague Saroj Rout [152] shows that the presence of amino acids have an influence on the reaction of polymerization. In particular, hydrophobic amino acids such as valine improve the yields of A, U and C. More generally, the idea of amino acids/peptides and nucleotide/RNA catalyzing each others reactions and later coevolution during the emergence of life is attractive, as it would be one step further toward the organization of the extant life. What's more, both are expected to be found in the vicinity of each other, as amino acids and nucleobases are synthesized in similar conditions [153, 154].

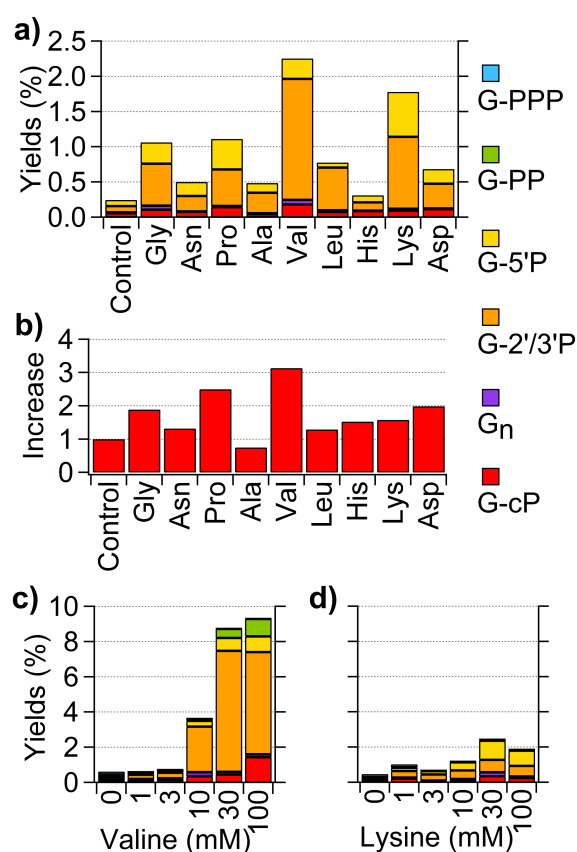


Figure 4.8: Effect of amino acids of the phosphorylation of G by trimetaphosphate: Solutions of 1 mM G nucleoside, 10 mM TMP, 10 mM amino acids for a) and b), adjusted to pH 10, was dried at 60 °C for 24 hours. The control experiment was done in the same conditions, but without any amino acid. **a)** Screening of different amino acids: yields of each products and total yield of G reaction. **b)** Increase on G-cP formation for each amino acid in comparison with the control experiment. **c)** Screening of different quantities of valine. **d)** Screening of different quantities of lysine. Valine and lysine have the highest impact on the total of G conversion. Valine enhanced the cyclic phosphate product the most.

Different amino acids were tested as catalysts for the phosphorylation of G nucleoside by trimetaphosphate at alkaline pH: glycine, asparagine, proline, hydrophobic amino acids (alanine, valine and leucine), basic amino acids histidine and lysine) and acidic aspartic acid (see Figure 4.A.2). They were all tested at concentration 10 mM at pH 10, for drying at 60 °C and the results can be see in Figure 4.8 **a)** and **b)**. With the exception of histidine, all the

4.3 Results

amino acids improved the total yield of phosphorylation, although to a lesser extent than urea. The reaction showed different regioselectivity depending on the amino acid used. The two best catalysts were found to be valine and lysine (see Figure 4.8 **c**) and **d**), but lysine increases best the product G-5'P, while valine enhanced mostly G-cP and G-2'/3'P. Of all the amino acids, valine is also the one that increases the most efficiently G-cP formation, up to 3 folds for 10 mM valine (see Figure 4.8 **b**)), and the yield of G-cP is even higher for higher valine quantities.

A mixture of 1 mM G nucleoside, 10 mM TMP, 10 mM valine, adjusted at pH 10 was put through several cycles of dehydration-rehydration at 60 °C, in order to observe the impact of wet-dry cycling on phosphorylation. The results are presented in Figures 4.9 and 4.B.1: wet-dry cycling has a strong impact on the total conversion yield of G nucleoside, which more than doubles between the 1st and the 6th cycles. Although the yield of G-cP does not increase after the first cycle, the polymers yields are enhanced by the wet-dry cycling. The produced G-cPs are probably converted into oligomers or hydrolyzed G-2'/3'P.

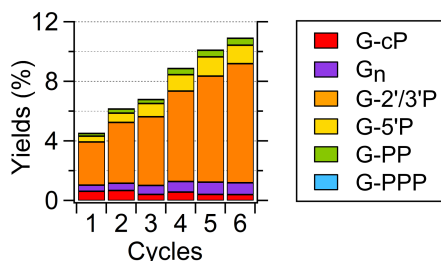


Figure 4.9: Effect of the wet-dry cycling on the phosphorylation of G by TMP in presence of valine: A solution of 1 mM G nucleoside, 10 mM TMP, 10 mM valine, adjusted to pH 10, was dried at 60 °C, during 24 hours for each cycles (except for the 5th and 6th cycle which lasted three days each). Each rehydration step was done using pure (RNase-free) water, without pH adjustment. Linear phosphorylated products and dimers were enhanced by the cycling.

4.3.2 Cyclization of the 2'- or 3'-phosphate nucleotide

In the previous experiments on the direct phosphorylation, I obtained higher yields of the linear products G-2'/3'P. Furthermore, polymerization starting from a pool of N-cP is also in competition with the hydrolysis of the reactants at alkaline pH, which also yields G-2'/3'P. The efficient cyclization of the linear monophosphate nucleotide is key to the synthesis of oligomers of increasing length. In literature, cyclization is achieved mostly under acidic conditions and in the presence of organic catalysts [57, 131, 132].

I tested the cyclization in similar conditions to direct phosphorylation by trimetaphosphate. TMP could act as a condensing agent by phosphorylating further the compound. The phosphate ring is formed by removing a molecule of linear triphosphate (see Figure 4.10). Other hydroxide groups of the ribose could be phosphorylated by TMP, resulting in bis/multiphosphorylated compounds as side products. In Figure 4.B.2, I detected 4-5 peaks corresponding to the mass of di/bis phosphate can be observed, showing that several distinct isomers are produced in the same conditions. Apart from G-5'PP, the standards of those isomers were

4. Formation of 2',3'-cyclic phosphate nucleotides

not commercially available, so I did not try to identify them separately.

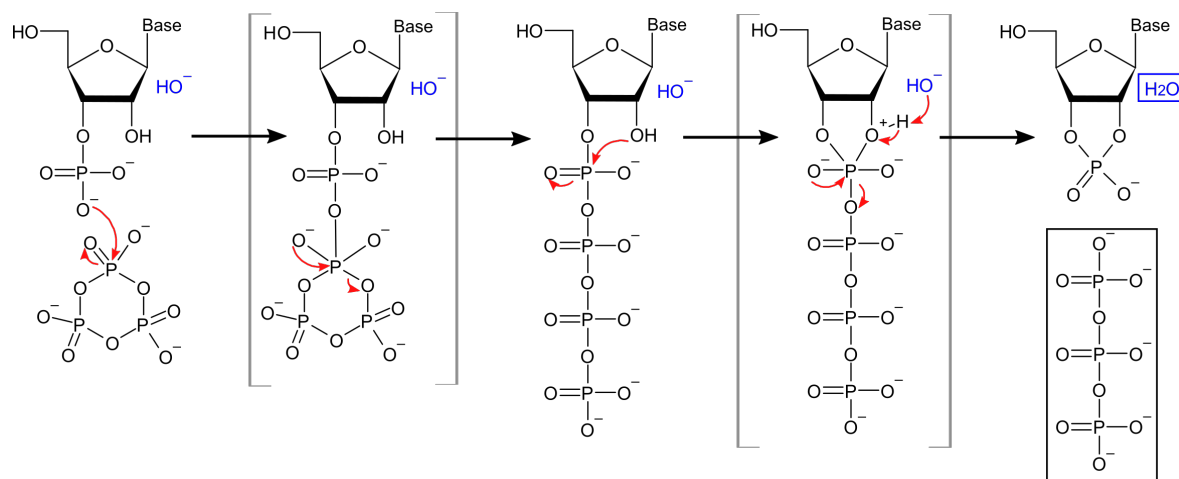


Figure 4.10: 2',3'-cyclic phosphate nucleoside (N-cP) synthesis by the recyclization of 2'- and 3'-phosphate nucleoside by trimetaphosphate at pH 10: The reaction produces no water molecule but one molecule of linear triphosphate.

In Figure 4.11, we can see that the reaction of TMP on G-2'/3'P produced G-cP, with higher yields than for the direct phosphorylation. From a solution of 1 mM G-2'/3'P and 10 mM trimetaphosphate adjusted to pH 10, dried at 60 °C without amino acids or other catalysts, a yield of 0.3 % G-cP was obtained. In the absence of any TMP, no cyclization was observed at alkaline pH. The presence of amino acids, in addition to the trimetaphosphate, impacts both the formation of the products. All of them enhanced the cyclization reaction to some extent; proline and lysine are the two most efficient catalysts, increasing significantly the reaction (9-10 folds). I tested different concentrations of lysine (see Figure 4.11) and observed that it improved not only the cyclization, but also the subsequent formation of the dimer GG-P, which is most likely the product of G-cP on G-2'/3'P. In comparison with direct phosphorylation, valine had a lower impact on the reaction of cyclization starting from an already phosphorylated nucleotide.

4.3 Results

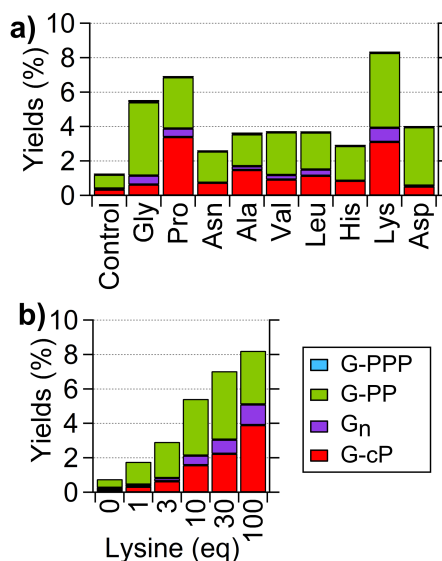


Figure 4.11: Effect of amino acids of the cyclization of G-P by trimetaphosphate: Solutions of 1 mM G 2'/3'-monophosphate, 10 mM TMP, 10 mM amino acids for a) and b), adjusted to pH 10, was dried at 60 °C for 24 hours. The control experiment was done in the same conditions, but without any amino acid. **a)** Screening of different amino acids: yields of each product and total yield of G reaction. **a)** Screening of different amino acids: yields of each product and total yield of G reaction. **b)** Screening of increasing quantities of lysine.

4.3.3 Amino acid phosphorylation and polymerization

In addition to the reaction on guanosine, I detected phosphorylated amino acids and peptide products resulting from the reaction between amino acids and trimetaphosphate, for each of the amino acids tested (see Figure 4.C.1). The evolution of each of those products for valine is detailed in the Figure 4.12 through the successive wet-dry cycles. The following analyses will remain purely qualitative: amino acids and their products after the reaction were not quantified.

The product which mass corresponds to the cyclic acyl-phosphoramidate is detected after the 1st dehydration and it decreases quickly with an increasing number of wet-dry cycles. I also detected masses corresponding to mono- and diphosphorylated products; it must be noted that analysis by mass spectrometry does not permit to know whether the phosphate group(s) is attached to the amide or carboxyl group of the amino acids. For valine, only one peak is observed, which might mean that only one of the isomers is produced. Orgel and his co-workers [155] estimated that the carboxyl group remains rather unreactive, so the main product of amino acid phosphorylation should be the *N*-phosphoramidate. Peptide formation is observed for each of the amino acids, phosphorylated or not, as can be seen in Figure 4.C.1 **b)** and **c)**. Unlike the phosphorylated amino acids, the quantities of dipeptide increase with wet-dry cycling (see Figure 4.12 **e)**, **f)**, **g)**, **h)** and **i)**).

The condensation of oligopeptides catalyzed by trimetaphosphate has already been re-

4. Formation of 2',3'-cyclic phosphate nucleotides

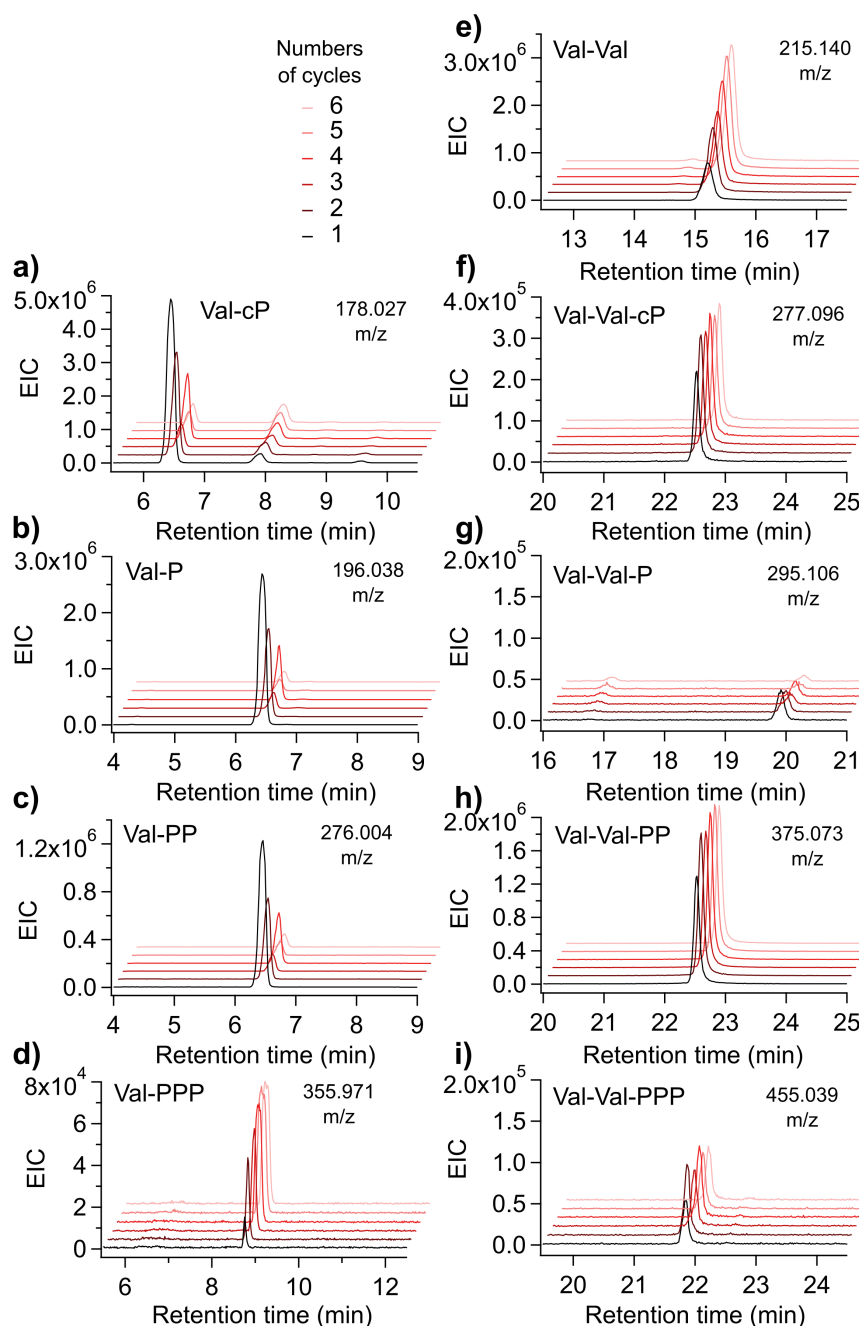


Figure 4.12: Reaction products between trimetaphosphate and valine A solution of 1 mM G nucleoside, 10 mM TMP, 10 mM valine is put through several wet-dry cycling at 60 °C (each cycle corresponds to 1 day, except for the 5th and 6th which lasted 3 days). EIC chromatograms corresponding to the detected products from the phosphorylation and oligomerization reaction of valine, for each cycle. **a)** Cyclic acyl phosphoramidate valine (Val-cP), **b)** monophosphorylated valine (Val-P), **c)** di/bisphosphorylated valine (Val-PP), **d)** triphosphorylated valine (Val-PPP), **e)** valine dipeptide (Val-Val), **f)** cyclic phosphate dipeptide valine (Val-Val-cP), **g)** monophosphorylated dipeptide (Val-Val-P), **h)** di/bisphosphorylated dipeptide (Val-Val-PP), **i)** triphosphorylated peptide (Val-Val-PPP). No tripeptide was detected for valine.

4.3 Results

ported in literature [155–161]. The reaction mostly occurs at alkaline pH and heat-drying, but it also occurs at very acidic pH (2-3) [162]. Two possible mechanisms were proposed for this reaction: Rabinowitz' mechanism [156] proposed as a first step the phosphorylation of the carboxylate group and dipeptide bond formation happens by nucleophilic attack of the amine group of another amino acid, producing a molecule of a linear triphosphate as a leaving group. Another mechanism was proposed later by Orgel and his coworkers [155] and can be seen in Figure 4.13. The first step of the reaction is initiated by phosphorylation of the amine group and gives as a reactive intermediate the acyl phosphoramidate. We detect this product for all amino acids. But in comparison with the mono and diphosphorylated product, we detected almost no triphosphoryl product for amino of the amino acids, although it is an intermediary product in the mechanism shown in Figure 4.13 for the formation of the cyclic acyl phosphoramidate (detected for all amino acids tested). It might be due to the molecule being highly reactive toward hydrolysis or cyclization.

In the condition of the dehydration-rehydration cycling in the presence of valine, the ion counts detected for the dipeptides increase with time, while the cyclic acyl phosphorymidate decreases, maybe partly due to the reaction of oligomerization.

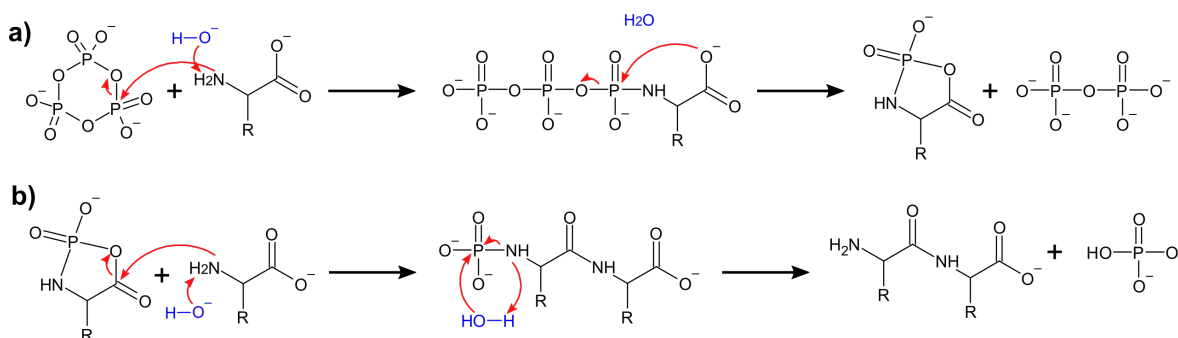


Figure 4.13: Proposed mechanism of dipeptide formation by TMP: for pH 10, adapted from the mechanisms proposed by Orgel and his coworkers and subsequent works [155,163–165].

a) Formation of the cyclic acyl-phosphoramidate: the nucleophilic attacks of the amine group on a phosphorus center of TMP, causing ring opening and production of a *N*-triphosphoryl amino acid, which then cyclized by removal of a molecule of pyrophosphate. **b)** Formation of the peptidic bond: the amine group of another amino acid attack the carbon center of the ester, causing the opening of the cyclic acyl-phosphoramidate: the product is a phosphorylated dipeptide. On a later step, hydrolysis causes the cleavage with the phosphate group.

4.3.4 Formation of a covalent bond between amino acid and nucleotide

In addition to the guanosine phosphorylation products reported in Figures 4.12, 4.B.1 and, I detected other peaks of UV absorption for the experiments with added amino acids. The mass of the compound corresponds to a mono or diphosphorylated nucleotide (G-P-aa or G-PP-aa) covalently bound to an amino acid, starting from either a nucleoside or a nucleotide (see Figure 4.D.1). In the later case, no G-P-aa was observed: the amino acid comes with the addition of one phosphate. This molecule could correspond to several isomers, for which the

4. Formation of 2',3'-cyclic phosphate nucleotides

phosphorylated amino acid is attached to different -OH group of the ribose, and the amino acid could be attached to the phosphate via either its carboxylate or amine group. In the case of G-P-aa, I observed mostly one peak, but for G-PP-aa, there are two to four peaks depending on the amino acid. It is possible that this molecule is the product of the reaction of the cyclic acyl phosphoramidate on the guanosine (nucleoside or nucleotide). In the case of valine, we can see that the UV absorption peak corresponding to the product G-P-Val increases with addition of wet-dry cycles. For the 6th cycle, its yield compared to the initial reactant of guanosine can be estimated to 6 %.

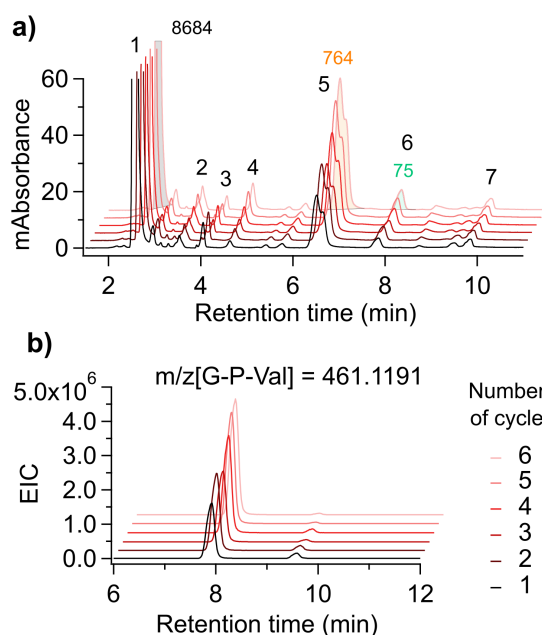


Figure 4.14: Valine-phosphorylated guanosine hybrid compound A solution of 1 mM G nucleoside, 10 mM TMP, 10 mM valine is put through several wet-dry cycling at 60 °C (each cycle corresponds to 1 day, except for the 5th and 6th which lasted 3 days). **a)** UV chromatogram (at 260 nm) for each cycle, showing the peaks corresponding to the nucleoside (**1**) and to the main products: dimer GG-noP (**2**), cyclic product G-cP (**3**), G-5P (**4**), G-2/3P (**5**). The peak **6** was attributed to the m/z value corresponding to valine-guanosine hybrid compound G-P-Val, but there several isomers possible for this product. The peak **7** was attributed to the mass of di/bisphosphate, and it can correspond to several isomers. The yield of formation of G-P-Val can be estimated to 0.8% for the 6th cycle, based on the comparison of the UV absorption peak area of G-noP (**8684**), of G-2/3P (**764**) and G-P-Val (**75**), shown on the Figure. **b)** EIC chromatograms corresponding to the m/z of G-P-Val, for every cycle: the products seem to be increased by the wet-dry cycling.

4.4 Discussion

Trimetaphosphate reaction on nucleoside and 2'-/3'-monophosphate guanosine gives G-cP as a product at alkaline pH. Those two conditions are quite compatible with the subsequent polymerization: we detect the presence of a dimer GG under the same alkaline conditions that produced G-cP. But trimetaphosphate does not react at low temperatures, so most of

4.4 Discussion

the experiments in this chapters were performed at 60 °C. The yields remain quite low, below 0.05 % for the direct phosphorylation and 0.3 % for the cyclization starting from 2'/3'-monophosphate guanosine. I tested several potential condensing agents and catalysts. Urea, a common condensing agents in literature, enhanced significantly – 20 fold – the cyclic product (for 10 mM urea), at alkaline pH.

I tested the impact of amino acids on those reactions. For direct phosphorylation, valine gave the highest yield – 3 fold increase for the formation of G-cP, for a valine concentration of 10 mM. Further wet-dry cycling wet-dry cycles did not result in a higher yield of G-cP, but they did improve the formation of G dimers. The cyclization was improved by the presence of lysine and proline – 10 fold increase. Valine and other amino acids tested also enhanced the G-cP. It is unclear at this point why those amino acids in particular give the best results. More generally, condensing agents work by diminishing the activity on water, bringing energy to the condensation reaction or/and reacting with the substrate to form a new, more reactive molecule [120]. Amino acids react with TMP to form, among other products, a cyclic acyl phosphoramidate. That molecule may produce G-cP, with one amino acid (or monophosphorylated amino acid) activated as the leaving group. The formation of a covalent bond between a phosphorylated nucleotide and an amino acid has been confirmed by mass spectrometry.

Rout et al [152] also observed that amino acids catalyze the polymerization of A-cP, U-cP and C-cP at room temperature and alkaline pH. However, under those conditions, there was no reaction between cyclic phosphate nucleotides and amino acids. The amine group of amino acids has a pK_a in general around 9 to 10, the pH used for the polymerization. Amino acids carrying a hydrophobic side chain, in particular valine, were found to be the most efficient for the polymerization.

The phosphorylation of amino acids by trimetaphosphate is reported in literature [166], with rather high yields (60-90 %). Condensation of amino acids via activation by a phosphate species, TMP [155–161] or DAP [63] is also one of the main known pathways for prebiotic peptide formation. Most of the literature focuses on glycine as a model amino acid [164]. With TMP, the reaction proceeds better at alkaline pH and higher drying temperature, similar conditions to those of the G-cP synthesis, with dipeptide as the main product [162]. Other amino acids have been tested, such alanine [162, 165, 167], but they tend to show a lower efficiency. Here, I detected the masses corresponding to dipeptides for every tested amino acid. Oligopeptides were also formed without TMP, but with a lower efficiency. This reaction can be improved by wet-dry cycling [94, 168]. Both pathways can be applied to our experiments, as we have drying and wet-dry cycles and we detected the phosphorylated intermediates for the dipeptide formation in the mechanism proposed by Orgel [155].

In addition, I detected, for all experiments conducted in presence of amino acids, a molecule corresponding to a nucleotide-amino acid hybrid compound. Several isomers are possible and the analysis by HPLC-MS was not enough to determine which one was produced in these experiments. Previous works established that the main products of phosphorylation of amino acids by trimetaphosphate are *N*-phosphoramidates, where the amino acid is attached to the phosphate group by the amine group [155]. In presence of a nucleoside, for example adenosine, trimetaphosphate and amino acid in alkaline solution, it was discov-

4. Formation of 2',3'-cyclic phosphate nucleotides

ered that the main product was the 2'-phosphorylated hybrid compound, when in comparison the 5'-phosphorylated product could not be generated in those conditions [169]. In this article, amino acid nucleotide hybrid compounds were detected for some of the biogenic amino acids and the found canonical nucleobases. Adenosine formed hybrid compounds with all the amino acids except proline, but guanosine, cytidine and uridine showed a lower efficiency. Glycine, aspartic acid, asparagine and glutamine were particularly difficult. In my experiments, I qualitatively detected amino-guanosine hybrid compounds in dried conditions for all the tested amino acids.

The 5'-aminoacyl adenylate (A-5'-P-aa, where the amino acid is attached to the 5'-phosphate group via the carboxyl group of the amino acid) are essential intermediates for protein biosynthesis, however it is not easily synthesized in prebiotic chemistry. Some deemed that the O-5'-P-aa might be a later evolutionary product, and another isomer, N-phosphoramidate could have been primitive activated intermediate for the prebiotic synthesis of peptides [170]. Phosphoramidate could be an activated compound for both the synthesis of oligopeptides and oligonucleotides [44, 171].

Addendum 4

4.A Molecular structures

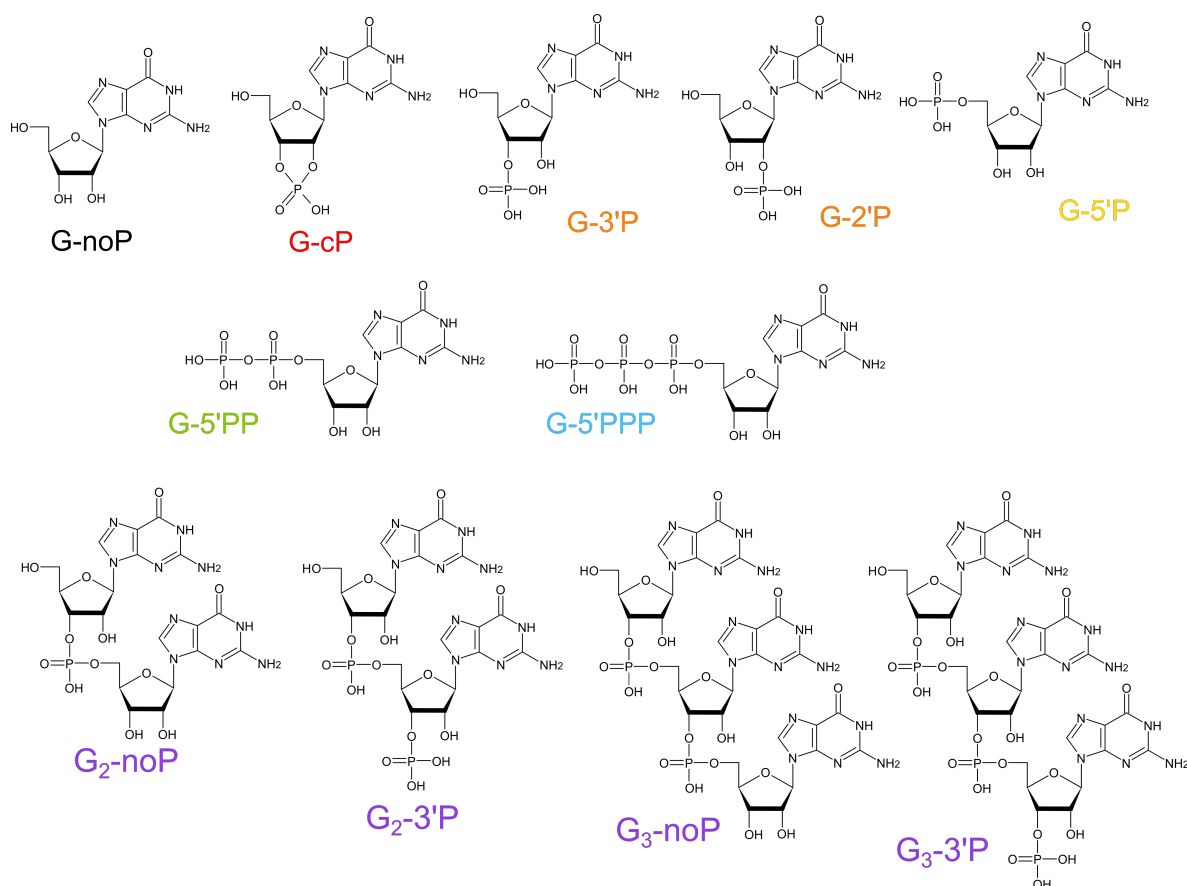


Figure 4.A.1: Standards used for quantification: G nucleoside (G-noP), 2',3'-cyclic phosphate G (G-cP), 3'-monophosphate G (G-3P), 2'-monophosphate G (G-2P), 5'-monophosphate G (G-5P), 5'-diphosphate G (G-5PP), 5'-triphosphate G (G-5PPP), G dimer with no phosphate ending (GG-noP), G dimer with 3'-phosphate ending (GG-3P), G trimer with no phosphate ending (GGG-noP), G trimer with 3'-phosphate ending (GGG-3P)

4. Formation of 2',3'-cyclic phosphate nucleotides

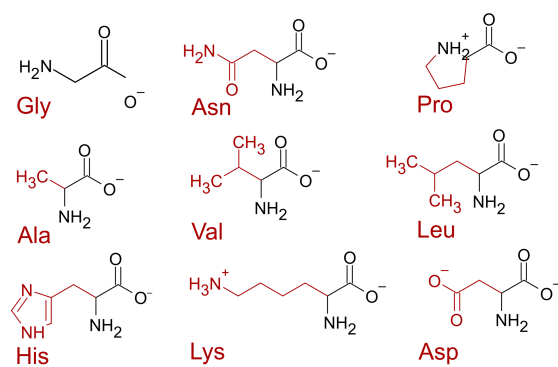


Figure 4.A.2: Amino acids structures: glycine, asparagine, alanine, valine, leucine, histidine, lysine, aspartic acid.

4.B Effect of amino acids on phosphorylation of G nucleoside and cyclization of G monophosphate nucleotide

4.B Effect of amino acids on phosphorylation of G nucleoside and cyclization of G monophosphate nucleotide

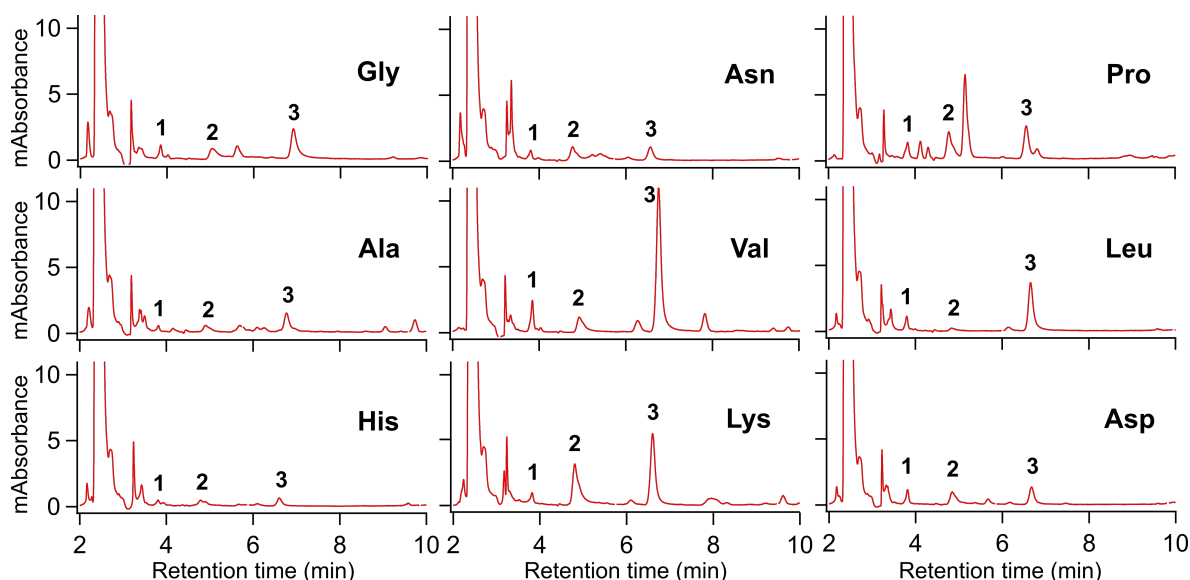


Figure 4.B.1: Effect of different of amino acids on the direct phosphorylation of G nucleoside: 50 μ L of a solution of 1 mM G nucleoside, 10 mM trimetaphosphate, 10 mM amino acids, adjusted of pH 10 using 1 M KOH were dried on a glass slide at 60 $^{\circ}$ C and left to react for 24 hours. The tested amino acids are glycine, asparagine, alanine, valine, leucine, histidine, lysine and aspartic acid.

The Figure shows the UV absorbance chromatogram (for 260 nm), focusing on the peaks of the main products of the experiment: **1** G-cP, **2** G-5'P, **3** G-2'/3'P (the unnumbered peak after 2 minutes corresponds to the nucleoside).

4. Formation of 2',3'-cyclic phosphate nucleotides

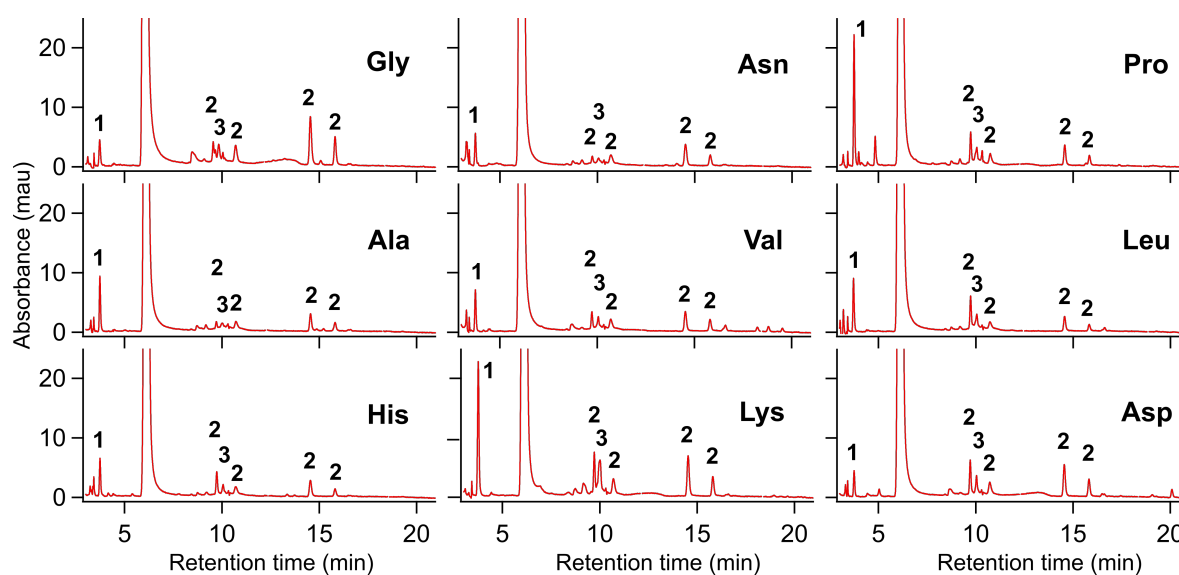


Figure 4.B.2: Effect of different amino acids on the cyclization and further phosphorylation of G monophosphate: 50 μ L of a solution of 1 mM G 2'/3'-monophosphate, 10 mM trimetaphosphate, 10 mM amino acids, adjusted of pH 10 using 1 M KOH were dried on a glass slide at 60 °C and left to react for 24 hours. The tested amino acids are glycine, asparagine, alanine, valine, leucine, histidine, lysine and aspartic acid.

The Figure shows the UV absorbance chromatogram (for 260 nm), focusing on the peaks of the main products of the experiment: **1** G-cP, **2** peaks attributed to the m/z corresponding to the mass of guanosine di/bisphosphate (bisphosphate: two monophosphates are attached to distinct hydroxide groups). The peaks were not identified one by one and they were quantified together as G-PP in Figure 4.11. **3** corresponds to both G-PP and the dimer GG-P. Proline and lysine gives the best results.

4.C Phosphorylated amino acids and dipeptides

4.C Phosphorylated amino acids and dipeptides

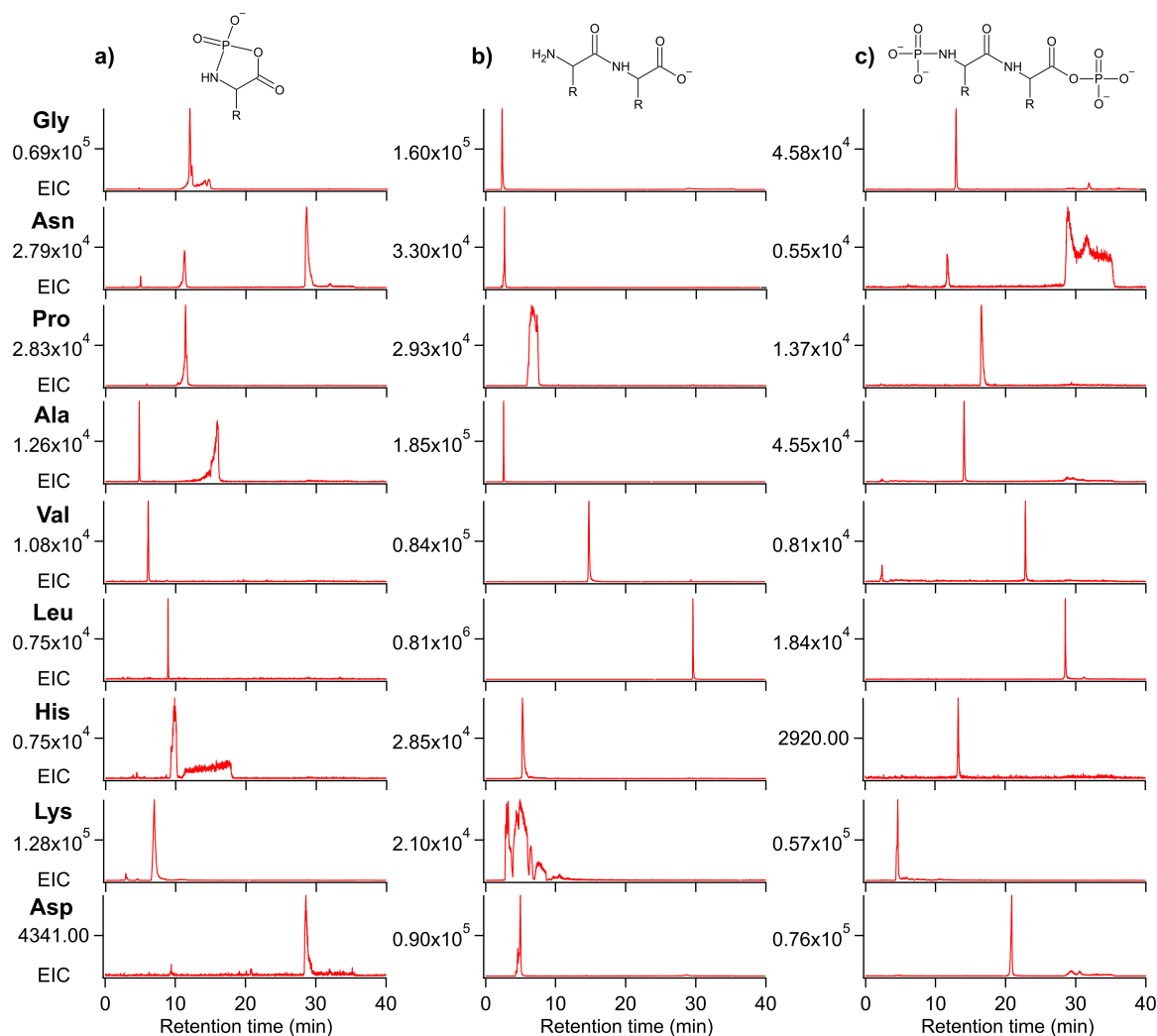


Figure 4.C.1: Phosphorylated amino acids and dipeptide products: The results shown correspond to the cyclization experiments, but the same products were detected also for the direct phosphorylation experiments. Conditions of the experiments: 1 mM G 2'/3'-monophosphate, 10 mM TMP, 10 mM amino acids, pH 10, 50 μ L dried at 60 $^{\circ}$ C and left to react for 24 hours. **a)** EIC chromatograms corresponding to the m/z of the cyclic acyl-phosphoramidate for each amino acid. **b)** EIC chromatograms corresponding to the m/z of the non-phosphorylated dipeptide. **c)** EIC chromatograms corresponding to the m/z of the di/bisphosphorylated dipeptide. Mono and di-phosphorylated amino acids were also detected, but not the triphosphorylated products. Monophosphorylated dimer was detected, but with lower ion count. Glycine was the only amino acid for which a trimer was detected.

4.D Amino-acid-guanosine hybrid compounds

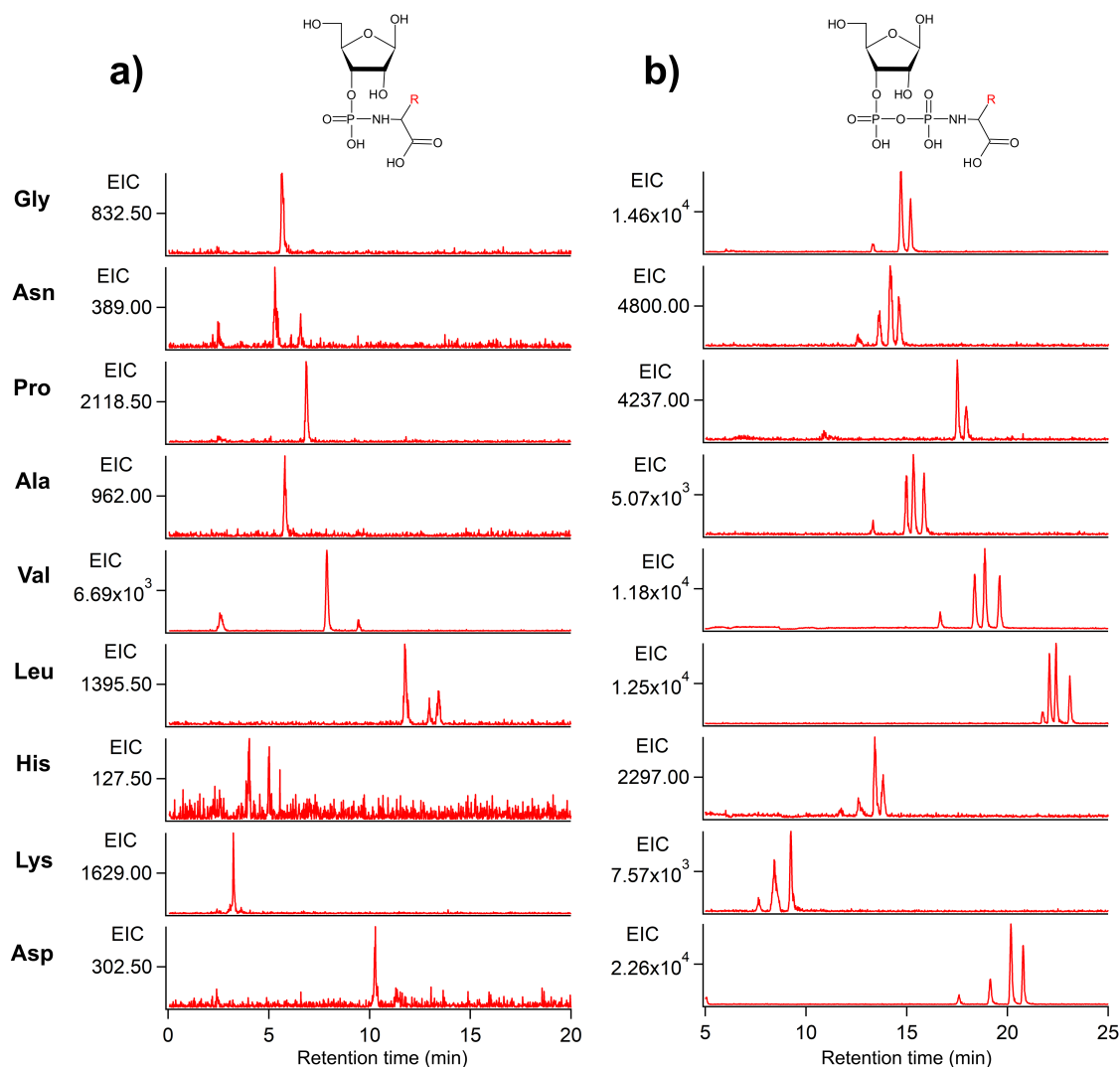


Figure 4.D.1: *N*-phosphoramidate amino acids guanosine: Extracted Ion Counts (EIC) chromatograms corresponding to the m/z of covalently bound G nucleotide and amino acids, detected for both phosphorylation and cyclization experiments. Those peaks were also observed on the UV absorption chromatogram. **a)** Conditions of the experiments: 1 mM G nucleoside, 10 mM TMP, 10 mM amino acids, pH 10, 50 μ L dried at 60 °C and left to react for 24 hours. For all amino acids, I detected a product corresponding to the monophosphorylated product covalently bound to one amino acid. **b)** Conditions of the experiments: 1 mM G 2'/3'-monophosphate G, 10 mM TMP, 30 mM amino acids, pH 10, 50 μ L dried at 60 °C. For all amino acids, I detected a product corresponding to the m/z of a di/bisphosphorylated product covalently bound to one amino acids. The monophosphorylated one was not detected in these experiments. Depending on the amino acid, the EIC chromatograms show 2 to 4 peaks: they might correspond to different isomers.

Chapter 5 | Outlook

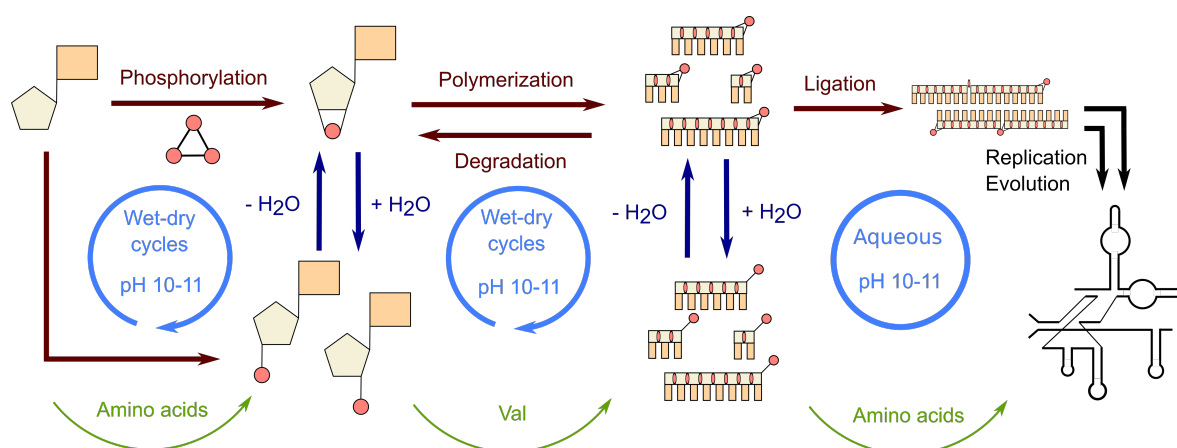


Figure 5.1: Outlook summary: The first steps of RNA synthesis – phosphorylation, cyclization, polymerization, as well as ligation and replication [85] – can be achieved in similar conditions, alkaline pH, aqueous solution or wet-dry cycling in presence of amino acids, by reaction of the 2',3'-cyclic phosphate group.

RNA world is one of the main hypotheses explaining the emergence of life. This is due to the ability of RNA to carry genetic information, to catalyze its own replication by hybridization between complementary strands and to form complex tridimensional structures that can catalyze chemical reactions. Nonetheless, the steps leading to the formation of ribozymes still need to be solved. How can we obtain long-enough RNA strands starting from mononucleosides/tides and simple organic molecule and in plausible conditions? I report here the formation of the first RNA via the opening of the 2',3'-cyclic phosphate ring (cP) at alkaline pH and by drying – or wet-dry cycling – at room temperature. A-cP, C-cP and U-cP polymerize around 20-40 % at pH 10-11 and gives products as long as 4-6mers after several wet-dry cycles. The imaging of their dry state shows only an amorphous phase. The yields are as high as 70 % G, with detection of products of length going from 2 to 10mers. None of the tested potential catalyst improved those yields, but using K^+ salts instead of Na^+ , Mg^{2+} or Li^+ increases the concentration of polymers. Microscope observation under polarized light shows the formation of liquid crystals, and after full drying of crystalline domains, resulting from a higher-order assembly between the G monomers and oligomers, which likely explain the enhanced polymerization for this nucleobase. This supramolecular assembly results likely from stackings of G-quadruplexes.

We tested the formation of the reactant N-cP starting from nucleosides and linear G-2'/3'P (0.05 % by direct phosphorylation and 0.5 % by cyclization). Drying an alkaline solution at 60 °C (or higher) gives the best results, with few side-products such as G-5P. However, un-

like polymerization, the reaction by TMP necessitates a drying process under higher temperatures to produce higher yields. The addition of small organic molecules such as urea or amino acids improves the formation of N-cP. In particular, valine enhances the reaction of direct phosphorylation. Lysine and proline improve the reaction of cyclization with the highest efficiency. The same experiments showed also the products of amino acids polymerization and phosphorylation by TMP.

Several major steps of RNA formation happen in very similar conditions: formation of the N-cP reactants, polymerization of those reactants into oligonucleotides and even the subsequent templated ligation [85]. They all proceed by opening of the phosphate cycle by base-catalysis. Polymerization and cyclic phosphate formation require drying, but they are not deterred by the presence of water, as long as dehydration happens. Ligation (and later ribozyme activity) happens in aqueous solution. The overall system could work in wet-dry cycling conditions. Polymerization and phosphorylation are enhanced by cycles of dehydration-rehydration and I expect ligation and elongation of RNA strands to be improved by self-assembly of the oligomers, such as liquid crystal, in highly-concentrated aqueous solutions.

Liquid crystals have been proposed as environment that could host the first steps of the RNA world. In detail: polymerization from monomers, selection of sequences based on their structure and composition, ligation, replication and apparition of catalytic functions [110]. Liquid crystals arise from molecular self-ordering, and they have been obtained from G-monomer solutions [109] which my collaborators observed, long DNA strands [106] and from shorter oligomers forming duplexes [108, 118]. Liquid crystals create a selection pressure toward the longer oligomers, as they fit better in the aligned environment, in which they are also protected from hydrolysis thanks to the lower water activity [112]. It also causes a separation between duplexes and simple helices (due to their difference in flexibility) from a solution of short random oligomers, with a very high number of distinct molecules [117, 172]. This could help to reduce the bias towards G. The phenomenon of liquid crystals formation from duplexes also enhances ligation by keeping the reactive ends in the vicinity of each other [114]. This was shown for the ligation driven by EDC, but could be applied for 2',3'-cyclic phosphate ligation. Liquid crystals were also proven to form short complementary strands with dangling endings, if that end find another complementary sequence. Thus, the linear aggregation and liquid crystal ordering are restored [173, 174]. This arrangement could favor the quick assembly of long oligomers via a series of splinted ligation like Wunnava and Serrão [85] demonstrated for the 2',3'-cyclic phosphate. Therefore, liquid crystal matrices could act as fluid and permeable microreactors for the formation of the first RNA strands. What's more, liquid crystals formation have been observed from DNA and peptide coacervates [175].

The different steps studied in this thesis preferably occur under mild alkaline conditions, around pH 10-11. The fact that they all need the same conditions improve the plausibility of that system. An alkaline pH is less likely to be found than a neutral pH on the early Earth. Nonetheless, some specific environments are known to locally have high pH. Hydrothermal

vents are considered potential sites, as natural electrochemical reactors [176–178]. In particular, the Lost City is a hydrothermal vent with a pH around 9 and 10 [13, 179, 180]. Some hydrothermal vents can also be found in shallow water. Wet-dry cycling could happen, as they can be exposed to air at low tide [181].

The exact temperatures at the surface of the early Earth are unknown. In most studies about prebiotic chemistry, lower temperatures are generally considered to be more plausible. The opening of the 2',3'-cyclic phosphate to form a diphosphoester bond (polymerization [75], ligation [85] and ribozyme-catalyzed ligation [182–184]) occurs at room temperature or even at lower temperatures, with slower kinetics. Higher temperatures tend to destabilize supramolecular assemblies and increase the kinetics of hydrolysis of the phosphate ring. Nonetheless, higher temperatures are needed for more efficient polymerization and cyclization by trimetaphosphate. In this thesis, I tested mild conditions by submitting the samples to a temperature of 60 °C. Maybe in later studies, the optimization of other parameters such as wet-dry cycling and the presence of efficient condensing agents will favor the reaction at lower temperatures.

Polymerization of G-cP (and ligation of short oligonucleotides) occurs without the presence of catalysts other than potassium salts. Nonetheless, amino acids such as valine improve the yield of A, U and C, making the reaction less biased toward G. They also improved the yields of cP formation. A more complex chemical composition makes it less likely to be found on the early Earth. But amino acids and nucleobases are believed to form in similar conditions [20–24, 154], so it is believable that they may have been in the same environments anyway. What's more, I observed amino acids phosphorylation and condensation also occurring in the same conditions: wet-dry cycling, alkaline solution, in presence of TMP. Recent works [185, 186] push the idea that peptides might have played a role in the RNA world hypothesis. Therefore the parallel development of both RNA and oligopeptides would be a major stepstone for the development of life on Earth. Nucleotide-amino acid hybrid molecules were detected: in those conditions, it is likely that I formed mostly the 2'-phosphoramidate guanosine rather than other isomers (5'-phosphoramidate guanosine, aminoacyl guanylate) [155, 169]. 5'-aminoacyl adenylate is used in extant Life as an intermediate for the biosynthesis of transfer RNA. But some other isomers might have played a similar role during the emergence of Life [44, 170, 171]. What's more, this reactive molecule could serve as an activated nucleotide or amino acid to enhance the formation of either peptides or RNA strands. In our experiments, its formation might partially be the reason of the enhancement of G-cP and dipeptide formation in the presence of most of the tested amino acids.

In conclusion, the results of this thesis show the first steps of the synthesis of RNA under simple, prebiotically likely conditions, based on the reactivity of the 2',3'-cyclic phosphate nucleotides. The phosphorylation (activation of the nucleotide), the polymerization and the cyclization (reactivation of the nucleotide) all occur by drying of an alkaline solution at room temperature or under mild heating. The presence of amino acids favors these reactions and

the amino acids themselves undergo phosphorylation (activation) and dimerization under the same conditions.

Bibliography

- [1] PublicDomainPictures. File:earth formation.jpg, 2013. <http://pixabay.com/en/mercury-venus-earth-mars-jupiter-163610>, Last accessed on 2024/09/26.
- [2] NASA/JPL-Caltech. File:artist's concept of collision at hd 172555.jpg, 2009. http://www.nasa.gov/multimedia/imagegallery/image_feature_1454.html, Last accessed on 2024/09/26.
- [3] NASA/Apollo 17 crew. File:the blue marble (remastered).jpg, 1972. <https://www.flickr.com/photos/projectapolloarchive/21081863984/>, Last accessed on 2024/09/26.
- [4] NPS National Park Service/P. Carrara. File:stromatolites.jpg, 1982. <https://commons.wikimedia.org/wiki/File:Stromatolites.jpg>, Last accessed on 2024/09/26.
- [5] Leslie E Orgel. The origin of life—a review of facts and speculations. *Trends in biochemical sciences*, 23(12):491–495, 1998.
- [6] Da-Fei Feng, Glen Cho, and Russell F Doolittle. Determining divergence times with a protein clock: update and reevaluation. *Proceedings of the National Academy of Sciences*, 94(24):13028–13033, 1997.
- [7] Mark J Burchell. Panspermia today. *International Journal of Astrobiology*, 3(2):73–80, 2004.
- [8] Frances Westall, André Brack, Alberto G Fairén, and Mitchell D Schulte. Setting the geological scene for the origin of life and continuing open questions about its emergence. *Frontiers in astronomy and space sciences*, 9:1095701, 2023.
- [9] Ben KD Pearce, Andrew S Tupper, Ralph E Pudritz, and Paul G Higgs. Constraining the time interval for the origin of life on earth. *Astrobiology*, 18(3):343–364, 2018.
- [10] James F Kasting. Earth's early atmosphere. *Science*, 259(5097):920–926, 1993.
- [11] Hisahiro Ueda and Takazo Shibuya. Composition of the primordial ocean just after its formation: constraints from the reactions between the primitive crust and a strongly acidic, co₂-rich fluid at elevated temperatures and pressures. *Minerals*, 11(4):389, 2021.
- [12] William R Kuhn and James F Kasting. Effects of increased co₂ concentrations on surface temperature of the early earth. *Nature*, 301(5895):53–55, 1983.
- [13] William Martin, John Baross, Deborah Kelley, and Michael J Russell. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*, 6(11):805–814, 2008.
- [14] Maximilian Weingart, Siyu Chen, Clara Donat, Vanessa Helmbrecht, William D Orsi, Dieter Braun, and Karen Alim. Alkaline vents recreated in two dimensions to study ph gradients, precipitation morphology, and molecule accumulation. *Science Advances*, 9(39):eadi1884, 2023.
- [15] RS Martin, TA Mather, and DM Pyle. Volcanic emissions and the early earth atmosphere. *Geochimica et Cosmochimica Acta*, 71(15):3673–3685, 2007.
- [16] Christopher Chyba and Carl Sagan. Electrical energy sources for organic synthesis on the early earth. *Origins of Life and Evolution of the Biosphere*, 21(1):3–17, 1991.
- [17] Saroj K Rout, Michael P Friedmann, Roland Riek, and Jason Greenwald. A prebiotic template-directed peptide synthesis based on amyloids. *Nature communications*, 9(1):234, 2018.
- [18] Arthur Kornberg, Narayana N Rao, and Dana Ault-Riche. Inorganic polyphosphate: a molecule of many functions. *Annual review of biochemistry*, 68(1):89–125, 1999.
- [19] SR Kornberg. Adenosine triphosphate synthesis from polyphosphate by an enzyme from escherichia coli. *Biochimica et biophysica acta*, 26(2):294–300, 1957.
- [20] Stanley L Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, 1953.
- [21] D Ring, Y Wolman, N Friedmann, and SL Miller. Proc. nat. acad. sci. USA, 69:765, 1972.

- [22] Nadav Friedmann and Stanley L Miller. Phenylalanine and tyrosine synthesis under primitive earth conditions. *Science*, 166(3906):766–767, 1969.
- [23] J Oró and AP Kimball. Synthesis of purines under possible primitive earth conditions. i. adenine from hydrogen cyanide. *Archives of biochemistry and biophysics*, 94(2):217–227, 1961.
- [24] James P Ferris, Robert A Sanchez, and Leslie E Orgel. Studies in prebiotic synthesis: Iii. synthesis of pyrimidines from cyanoacetylene and cyanate. *Journal of molecular biology*, 33(3):693–704, 1968.
- [25] Claudia Huber and Gunter Wächtershäuser. Activated acetic acid by carbon fixation on (Fe, Ni) s under primordial conditions. *Science*, 276(5310):245–247, 1997.
- [26] J Oro and E Stephen-Sherwood. Abiotic origin of biopolymers. *Origins of life*, 7:37–47, 1976.
- [27] Walter Gilbert. Origin of life: The rna world. *nature*, 319(6055):618–618, 1986.
- [28] Eric H Eklund, Jack W Szostak, and David P Bartel. Structurally complex and highly active rna ligases derived from random rna sequences. *Science*, 269(5222):364–370, 1995.
- [29] Annalena Salditt, Leonie Karr, Elia Salibi, Kristian Le Vay, Dieter Braun, and Hannes Mutschler. Ribozyme-mediated rna synthesis and replication in a model hadean microenvironment. *Nature Communications*, 14(1):1495, 2023.
- [30] Zasha Weinberg, Peter B Kim, Tony H Chen, Sanshu Li, Kimberly A Harris, Christina E Lünse, and Ronald R Breaker. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nature chemical biology*, 11(8):606–610, 2015.
- [31] Adrian R Ferré-D’Amaré and William G Scott. Small self-cleaving ribozymes. *Cold Spring Harbor perspectives in biology*, 2(10):a003574, 2010.
- [32] James P Ferris, Aubrey R Hill Jr, Rihe Liu, and Leslie E Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61, 1996.
- [33] Giacomo Bartolucci, Adriana Calaça Serrão, Philipp Schwintek, Alexandra Kühnlein, Yash Rana, Philipp Janto, Dorothea Hofer, Christof B Mast, Dieter Braun, and Christoph A Weber. Sequence self-selection by cyclic phase separation. *Proceedings of the National Academy of Sciences*, 120(43):e2218876120, 2023.
- [34] Andrey do Nascimento Vieira, Karl Kleiner-manns, William F Martin, and Martina Preiner. The ambivalent role of water at the origins of life. *FEBS letters*, 594(17):2717–2733, 2020.
- [35] Andrew Pohorille and Lawrence R Pratt. Is water the universal solvent for life? *Origins of Life and Evolution of Biospheres*, 42:405–409, 2012.
- [36] ANN Henderson-Sellers and AJ Meadows. Surface temperature of early earth. *Nature*, 270(5638):589–591, 1977.
- [37] John W Valley, William H Peck, Elizabeth M King, and Simon A Wilde. A cool early earth. *Geology*, 30(4):351–354, 2002.
- [38] James F Kasting and M Tazewell Howard. Atmospheric composition and climate on the early earth. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474):1733–1742, 2006.
- [39] Rachel Green and Jack W Szostak. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science*, 258(5090):1910–1915, 1992.
- [40] Stephanie J Zhang, Daniel Duzdevich, Dian Ding, and Jack W Szostak. Freeze-thaw cycles enable a prebiotically plausible and continuous pathway from nucleotide activation to nonenzymatic rna copying. *Proceedings of the National Academy of Sciences*, 119(17):e2116429119, 2022.
- [41] Travis Walton, Wen Zhang, Li Li, Chun Pong Tam, and Jack W Szostak. The mechanism of nonenzymatic template copying with imidazole-activated nucleotides. *Angewandte Chemie International Edition*, 58(32):10812–10819, 2019.
- [42] IA Kozlov and LE Orgel. Nonenzymatic template-directed synthesis of rna from monomers. *Molecular Biology*, 34:781–789, 2000.
- [43] Anatol Luther, Romana Brandsch, and G Von Kiedrowski. Surface-promoted replication and exponential amplification of dna analogues. *Nature*, 396(6708):245–248, 1998.
- [44] Megha Karki, Clémentine Gibard, Subhendu Bhowmik, and Ramanarayanan Krishnamurthy. Nitrogenous derivatives of phosphorus and the origins of life: plausible prebiotic phosphorylating agents in water. *Life*, 7(3):32, 2017.

Bibliography

- [45] Bradley T Burcar, Mohsin Jawed, Hari Shah, and Linda B McGown. In situ imidazole activation of ribonucleotides for abiotic rna oligomerization reactions. *Origins of Life and Evolution of Biospheres*, 45:31–40, 2015.
- [46] Fabrice Duvernay, Thierry Chiavassa, Fabien Borget, and Jean-Pierre Aycard. Experimental study of water- ice catalyzed thermal isomerization of cyanamide into carbodiimide: Implication for prebiotic chemistry. *Journal of the American Chemical Society*, 126(25):7772–7773, 2004.
- [47] N Aylward. A plausible route to a prebiotic synthesis of l-histidine. *WSEAS Trans. Biol. Biomed*, 9:10–23, 2012.
- [48] Giovanna Costanzo, Samanta Pino, Anna Maria Timperio, Judit E Šponer, Jiří Šponer, Olga Novakova, Ondrej Šedo, Zbyněk Zdráhal, and Ernesto Di Mauro. Non-enzymatic oligomerization of 3', 5'-cyclic amp. *PLoS One*, 11(11):e0165723, 2016.
- [49] Sreekar Wunnavu, Christina F Dirscherl, Jakub Vřravský, Aleš Kovařík, Roman Matyášek, Jiří Šponer, Dieter Braun, and Judit E Šponer. Acid-catalyzed rna-oligomerization from 3', 5'-cgmp. *Chemistry—A European Journal*, 27(70):17581–17585, 2021.
- [50] Matthias Morasch, Christof B Mast, Johannes K Langer, Pierre Schilcher, and Dieter Braun. Dry polymerization of 3', 5'-cyclic gmp to long strands of rna. *ChemBioChem*, 15(6):879–883, 2014.
- [51] Judit E Sponer, Jiri Sponer, Alessandra Giorgi, Ernesto Di Mauro, Samanta Pino, and Giovanna Costanzo. Untemplated nonenzymatic polymerization of 3', 5' cgmp: a plausible route to 3', 5'-linked oligonucleotides in primordia. *The Journal of Physical Chemistry B*, 119(7):2979–2989, 2015.
- [52] M Smith, GI Drummond, and HG Khorana. Cyclic phosphates. iv. 1 ribonucleoside-3', 5'-cyclic phosphates. a general method of synthesis and some properties. *Journal of the American Chemical Society*, 83(3):698–706, 1961.
- [53] Rémi Coulon, Barbora Papoušková, Elmira Mohammadi, Michal Otyepka, Sreekar Wunnavu, Jiří Šponer, and Judit E Šponer. Prebiotic synthesis of 3', 5'-cyclic adenosine and guanosine monophosphates through carbodiimide-assisted cyclization. *ChemBioChem*, 24(24):e202300510, 2023.
- [54] MS Verlander, R Lohrmann, and LE Orgel. Catalysts for the self-polymerization of adenosine cyclic 2', 3'-phosphate. *Journal of Molecular Evolution*, 2:303–316, 1973.
- [55] MS Verlander and LE Orgel. Analysis of high molecular weight material from the polymerization of adenosine cyclic 2', 3'-phosphate. *Journal of molecular evolution*, 3:115–120, 1974.
- [56] M Renz, R Lohrmann, and LE Orgel. Catalysts for the polymerization of adenosine cyclic 2', 3'-phosphate on a poly (u) template. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 240(4):463–471, 1971.
- [57] Claude M Tapiero and Joseph Nagyvary. Prebiotic formation of cytidine nucleotides. *Nature*, 231(5297):42–43, 1971.
- [58] R Österberg, LE Orgel, and R Lohrmann. Further studies of urea-catalyzed phosphorylation reactions. *Journal of molecular evolution*, 2:231–234, 1973.
- [59] Yingfu Li and Ronald R Breaker. Kinetics of rna degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *Journal of the American Chemical society*, 121(23):5364–5372, 1999.
- [60] Ronald Breslow. How do imidazole groups catalyze the cleavage of rna in enzyme models and in enzymes? evidence from" negative catalysis". *Accounts of chemical research*, 24(11):317–324, 1991.
- [61] Megumi Shigematsu, Takuya Kawamura, and Yohei Kirino. Generation of 2', 3'-cyclic phosphate-containing rnas as a hidden layer of the transcriptome. *Frontiers in genetics*, 9:562, 2018.
- [62] Y Yamagata, H Inoue, and K Inomata. Specific effect of magnesium ion on 2', 3'-cyclic amp synthesis from adenosine and trimeta phosphate in aqueous solution. *Origins of Life and Evolution of the Biosphere*, 25:47–52, 1995.
- [63] Clémentine Gibard, Subhendu Bhowmik, Megha Karki, Eun-Kyong Kim, and Ramarayanan Krishnamurthy. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nature Chemistry*, 10(2):212–217, 2018.
- [64] Eddy I Jiménez, Clémentine Gibard, and Ramarayanan Krishnamurthy. Prebiotic phosphorylation and concomitant oligomerization of deoxynucleosides to form dna. *Angewandte Chemie International Edition*, 60(19):10775–10783, 2021.

- [65] Matthew W Powner, Béatrice Gerland, and John D Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244):239–242, 2009.
- [66] Matthias Morasch, Jonathan Liu, Christina F Dirscherl, Alan Ianeselli, Alexandra Kühnlein, Kristian Le Vay, Philipp Schwintek, Saidul Islam, Mérina K Corpinot, Bettina Scheu, et al. Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nature Chemistry*, 11(9):779–788, 2019.
- [67] Alan Ianeselli, Christof B Mast, and Dieter Braun. Periodic melting of oligonucleotides by oscillating salt concentrations triggered by microscale water cycles inside heated rock pores. *Angewandte Chemie*, 131(37):13289–13294, 2019.
- [68] Judit E Šponer, Jiří Šponer, Aleš Kovařík, Ondřej Šedo, Zbyněk Zdráhal, Giovanna Costanzo, and Ernesto Di Mauro. Questions and answers related to the prebiotic production of oligonucleotide sequences from 3′, 5′ cyclic nucleotide precursors. *Life*, 11(8):800, 2021.
- [69] Andreas Premstaller and Peter J Oefner. Denaturing hplc of nucleic acids. *Lc Gc Europe*, 15:410, 2002.
- [70] Andreas Premstaller and Peter J Oefner. Denaturing high-performance liquid chromatography. *Single Nucleotide Polymorphisms: Methods and Protocols*, pages 15–35, 2003.
- [71] Phillip B Danielson, Richard Kristinsson, Robert J Shelton, and Gregory S LaBerge. Separating human dna mixtures using denaturing high-performance liquid chromatography. *Expert Review of Molecular Diagnostics*, 5(1):53–63, 2005.
- [72] Fabien Côté, Dominique Lévesque, and Jean-Pierre Perreault. Natural 2′, 5′-phosphodiester bonds found at the ligation sites of peach latent mosaic viroid. *Journal of Virology*, 75(1):19–25, 2001.
- [73] Helen R Palmer, Jennifer J Bedford, John P Leader, and Robin AJ Smith. 31p and 1h nmr studies of the effect of the counteracting osmolyte trimethylamine-n-oxide on interactions of urea with ribonuclease a. *Journal of Biological Chemistry*, 275(36):27708–27711, 2000.
- [74] Sebastian Motsch, Peter Tremmel, and Clemens Richert. Regioselective formation of rna strands in the absence of magnesium ions. *Nucleic Acids Research*, 48(3):1097–1107, 2020.
- [75] Avinash Vicholous Dass, Sreekar Wunnava, Juliette Langlais, Beatriz von der Esch, Maik Krusche, Lennard Ufer, Nico Chrisam, Romeo CA Dubini, Florian Gartner, Severin Angerpointner, et al. Rna oligomerisation without added catalyst from 2′, 3′-cyclic nucleotides by drying at air-water interfaces. *ChemSystemsChem*, 5(1):e202200026, 2023.
- [76] Jeffery T Davis. G-quartets 40 years later: from 5′-gmp to molecular biology and supramolecular chemistry. *Angewandte Chemie International Edition*, 43(6):668–698, 2004.
- [77] Eric Largy, Jean-Louis Mergny, and Valérie Gabelica. Role of alkali metal ions in g-quadruplex nucleic acid structure and stability. *The alkali metal ions: Their role for life*, pages 203–258, 2016.
- [78] Jing You, Hui Li, Xi-Ming Lu, Wei Li, Peng-Ye Wang, Shuo-Xing Dou, and Xu-Guang Xi. Effects of monovalent cations on folding kinetics of g-quadruplexes. *Bioscience reports*, 37(4):BSR20170771, 2017.
- [79] Alan Ianeselli, Miguel Atienza, Patrick W Kudella, Ulrich Gerland, Christof B Mast, and Dieter Braun. Water cycles in a hadean co2 atmosphere drive the evolution of long dna. *Nature Physics*, 18(5):579–585, 2022.
- [80] Joshua Krissansen-Totton, Giada N Arney, and David C Catling. Constraining the climate and ocean ph of the early earth with a geological carbon cycle model. *Proceedings of the National Academy of Sciences*, 115(16):4105–4110, 2018.
- [81] Jonathan D Toner and David C Catling. A carbonate-rich lake solution to the phosphate problem of the origin of life. *Proceedings of the National Academy of Sciences*, 117(2):883–888, 2020.
- [82] Evgeniia Edeleva, Annalena Salditt, Julian Stamp, Philipp Schwintek, Job Boekhoven, and Dieter Braun. Continuous nonenzymatic cross-replication of dna strands with in situ activated dna oligonucleotides. *Chemical Science*, 10(22):5807–5814, 2019.
- [83] Changmei Cheng, Chang Fan, Rong Wan, Chunyuan Tong, Zhiwei Miao, Jing Chen, and Yufen Zhao. Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions. *Origins of Life and Evolution of the Biosphere*, 32:219–224, 2002.

Bibliography

- [84] AV Lutay, EL Chernolovskaya, MA Zenkova, and VV Vlasov. Nonenzymatic template-dependent ligation of 2', 3'-cyclic phosphate-containing oligonucleotides catalyzed by metal ions. In *Doklady Biochemistry & Biophysics*, volume 401, 2005.
- [85] Adriana Calaça Serrão, Sreekar Wunnava, Avinash V Dass, Lennard Ufer, Philipp Schwintek, Christof B Mast, and Dieter Braun. High-fidelity rna copying via 2', 3'-cyclic phosphate ligation. *Journal of the American Chemical Society*, 146(13):8887–8894, 2024.
- [86] Sarah Maurer. The impact of salts on single chain amphiphile membranes and implications for the location of the origin of life. *Life*, 7(4):44, 2017.
- [87] Tai-Sung Lee, Carlos Silva López, George M Gambaşu, Monika Martick, William G Scott, and Darrin M York. Role of mg^{2+} in hammerhead ribozyme catalysis from molecular simulation. *Journal of the American Chemical Society*, 130(10):3053–3064, 2008.
- [88] Irene A Chen, Kourosh Salehi-Ashtiani, and Jack W Szostak. Rna catalysis in model protocell vesicles. *Journal of the American Chemical Society*, 127(38):13213–13219, 2005.
- [89] Katarzyna Adamala and Jack W Szostak. Nonenzymatic template-directed rna synthesis inside model protocells. *Science*, 342(6162):1098–1100, 2013.
- [90] Jia Sheng, Li Li, Aaron E Engelhart, Jianhua Gan, Jiawei Wang, and Jack W Szostak. Structural insights into the effects of 2'-5' linkages on the rna duplex. *Proceedings of the National Academy of Sciences*, 111(8):3050–3055, 2014.
- [91] Aaron E Engelhart, Matthew W Powner, and Jack W Szostak. Functional rnas exhibit tolerance for non-heritable 2'-5' versus 3'-5' backbone heterogeneity. *Nature chemistry*, 5(5):390–394, 2013.
- [92] Martin Gellert, Marie N Lipsett, and David R Davies. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences*, 48(12):2013–2018, 1962.
- [93] Gabrielle Leveau, Daniel Pfeffer, Bernhard Altaner, Eric Kervio, Franziska Welsch, Ulrich Gerland, and Clemens Richert. Enzyme-free copying of 12 bases of rna with dinucleotides. *Angewandte Chemie*, 134(29):e202203067, 2022.
- [94] Thomas D Campbell, Rio Febrian, Jack T McCarthy, Holly E Kleinschmidt, Jay G Forsythe, and Paul J Bracher. Prebiotic condensation through wet-dry cycling regulated by deliquescence. *Nature communications*, 10(1):4508, 2019.
- [95] Noam Lahav, D White, and S Chang. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science*, 201(4350):67–69, 1978.
- [96] Sidney Becker, Christina Schneider, Hidenori Okamura, Antony Crisp, Tynchtyk Amatov, Milan Dejmek, and Thomas Carell. Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nature communications*, 9(1):163, 2018.
- [97] Sudha Rajamani, Alexander Vlassov, Seico Benner, Amy Coombs, Felix Olasagasti, and David Deamer. Lipid-assisted synthesis of rna-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres*, 38:57–74, 2008.
- [98] Laura Da Silva, Marie-Christine Maurel, and David Deamer. Salt-promoted synthesis of rna-like molecules in simulated hydrothermal conditions. *Journal of molecular evolution*, 80:86–97, 2015.
- [99] Oliver R Maguire, Iris BA Smokers, and Wilhelm TS Huck. A physicochemical orthophosphate cycle via a kinetically stable thermodynamically activated intermediate enables mild prebiotic phosphorylations. *Nature communications*, 12(1):5517, 2021.
- [100] Ivar Svalheim Haugerud, Pranay Jaiswal, and Christoph A Weber. Nonequilibrium wet-dry cycling acts as a catalyst for chemical reactions. *The Journal of Physical Chemistry B*, 128(7):1724–1736, 2024.
- [101] Armen Y Mulkidjanian, Andrew Yu Bychkov, Daria V Dibrova, Michael Y Galperin, and Eugene V Koonin. Origin of first cells at terrestrial, anoxic geothermal fields. *Proceedings of the National Academy of Sciences*, 109(14):E821–E830, 2012.
- [102] Bruce Damer and David Deamer. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life. *Life*, 5(1):872–887, 2015.
- [103] Alan Ianeselli, Damla Tetiker, Julian Stein, Alexandra Kühnlein, Christof B Mast, Dieter Braun, and T-Y Dora Tang. Non-equilibrium

- conditions inside rock pores drive fission, maintenance and selection of coacervate protocells. *Nature chemistry*, 14(1):32–39, 2022.
- [104] T Matreux, Kristian Le Vay, A Schmid, P Aikkilä, L Belohlavek, AZ Çalışkanoğlu, E Salibi, A Kühnlein, C Springsklee, B Scheu, et al. Heat flows in rock cracks naturally optimize salt compositions for ribozymes. *Nature Chemistry*, 13(11):1038–1045, 2021.
- [105] Emre Tekin, Annalena Salditt, Philipp Schwintek, Sreekar Wunnavu, Juliette Langlais, James Saenz, Dora Tang, Petra Schwille, Christof Mast, and Dieter Braun. Prebiotic foam environments to oligomerize and accumulate rna. *Chem-BioChem*, 23(24):e202200423, 2022.
- [106] V Luzzati and A Nicolaieff. Etude par diffusion des rayons x aux petits angles des gels d'acide désoxyribonucléique et de nucléoprotéines:(note préliminaire). *Journal of Molecular Biology*, 1(2):127–IN5, 1959.
- [107] Y Bouligand. Twisted fibrous arrangements in biological materials and cholesteric mesophases. *Tissue and Cell*, 4(2):189–217, 1972.
- [108] Michi Nakata, Giuliano Zanchetta, Brandon D Chapman, Christopher D Jones, Julie O Cross, Ronald Pindak, Tommaso Bellini, and Noel A Clark. End-to-end stacking and liquid crystal condensation of 6-to 20–base pair dna duplexes. *Science*, 318(5854):1276–1279, 2007.
- [109] Jeffery T Davis and Gian Piero Spada. Supramolecular architectures generated by self-assembly of guanosine derivatives. *Chemical Society Reviews*, 36(2):296–313, 2007.
- [110] Tony Z Jia, Tommaso Bellini, Noel Clark, and Tommaso P Fraccia. A liquid crystal world for the origins of life. *Emerging Topics in Life Sciences*, 6(6):557–569, 2022.
- [111] Gordon T Stewart. Liquid crystals in biology ii. origins and processes of life. *Liquid crystals*, 31(4):443–471, 2004.
- [112] Tommaso Pietro Fraccia et al. Exploring the role of liquid crystal ordering of dna oligomers in the prebiotic synthesis of nucleic acids. 2014.
- [113] Tommaso P Fraccia, Gregory P Smith, Giuliano Zanchetta, Elvezia Paraboschi, Youngwooo Yi, David M Walba, Giorgio Dieci, Noel A Clark, and Tommaso Bellini. Abiotic ligation of dna oligomers templated by their liquid crystal ordering. *Nature communications*, 6(1):6424, 2015.
- [114] Tommaso P Fraccia, Giuliano Zanchetta, Valeria Rimoldi, Noel A Clark, and Tommaso Bellini. Evidence of liquid crystal–assisted abiotic ligation of nucleic acids. *Origins of Life and Evolution of Biospheres*, 45:51–68, 2015.
- [115] Darren Kessner, Matt Chambers, Robert Burke, David Agus, and Parag Mallick. Proteowizard: open source software for rapid proteomics tools development. *Bioinformatics*, 24(21):2534–2536, 2008.
- [116] Mateusz K Łacki, Dirk Valkenborg, and Michał P Startek. Isospec2: ultrafast fine structure calculator. *Analytical chemistry*, 92(14):9472–9475, 2020.
- [117] Giuliano Zanchetta, Michi Nakata, Marco Buscaglia, Tommaso Bellini, and Noel A Clark. Phase separation and liquid crystallization of complementary sequences in mixtures of nanodna oligomers. *Proceedings of the National Academy of Sciences*, 105(4):1111–1117, 2008.
- [118] Tommaso P Fraccia, Gregory P Smith, Lucas Bethge, Giuliano Zanchetta, Giovanni Nava, Sven Klusmann, Noel A Clark, and Tommaso Bellini. Liquid crystal ordering and isotropic gelation in solutions of four-base-long dna oligomers. *ACS nano*, 10(9):8508–8516, 2016.
- [119] Maheen Gull. Prebiotic phosphorylation reactions on the early earth. *Challenges*, 5(2):193–212, 2014.
- [120] Matthew A Pasek. Thermodynamics of prebiotic phosphorylation. *Chemical reviews*, 120(11):4690–4706, 2019.
- [121] Allen M Schoffstall, Robert J Barto, and David L Ramos. Nucleoside and deoxynucleoside phosphorylation in formamide solutions. *Origins of life*, 12:143–151, 1982.
- [122] Allen M Schoffstall. Prebiotic phosphorylation of nucleosides in formamide. *Origins of life*, 7(4):399–412, 1976.
- [123] Giovanna Costanzo, Raffaele Saladino, Claudia Crestini, Fabiana Cicciello, and Ernesto Di Mauro. Nucleoside phosphorylation by phosphate minerals. *Journal of biological chemistry*, 282(23):16729–16735, 2007.
- [124] Jennifer L Lago, Bradley T Burcar, Nicholas V Hud, Rio Febrian, Christopher Mehta, Paul J Bracher, Zachary D Atlas, and Matthew A Pasek. The prebiotic provenance of semi-aqueous solvents. *Origins of Life and Evolution of Biospheres*, 50:1–14, 2020.

Bibliography

- [125] Bradley Burcar, Alma Castañeda, Jennifer Lago, Mischaël Daniel, Matthew A Pasek, Nicholas V Hud, Thomas M Orlando, and César Menor-Salván. A stark contrast to modern earth: Phosphate mineral transformation and nucleoside phosphorylation in an iron-and cyanide-rich early earth scenario. *Angewandte Chemie*, 131(47):17137–17143, 2019.
- [126] Bradley Burcar, Matthew Pasek, Maheen Gull, Brian J Cafferty, Francisco Velasco, Nicholas V Hud, and César Menor-Salván. Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angewandte Chemie International Edition*, 55(42):13249–13253, 2016.
- [127] R Lohrmann and LE Orgel. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science*, 171(3970):490–494, 1971.
- [128] R Saffhill. Selective phosphorylation of the cis-2', 3'-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. *The Journal of Organic Chemistry*, 35(9):2881–2883, 1970.
- [129] Hyo-Joong Kim and Steven A Benner. Abiotic synthesis of nucleoside 5'-triphosphates with nickel borate and cyclic trimetaphosphate (ctmp). *Astrobiology*, 21(3):298–306, 2021.
- [130] AD Castañeda, Z Li, T Joo, K Benham, BT Burcar, R Krishnamurthy, CL Liotta, NL Ng, and TM Orlando. Prebiotic phosphorylation of uridine using diamidophosphate in aerosols. *Scientific Reports*, 9(1):13527, 2019.
- [131] Ziwei Liu, Long-Fei Wu, Jianfeng Xu, Claudia Bonfio, David A Russell, and John D Sutherland. Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nature chemistry*, 12(11):1023–1028, 2020.
- [132] R Lohrmann and LE Orgel. Prebiotic synthesis: phosphorylation in aqueous solution. *Science*, 161(3836):64–66, 1968.
- [133] Lee B Mullen and John D Sutherland. Simultaneous nucleotide activation and synthesis of amino acid amides by a potentially prebiotic multi-component reaction. *Angewandte Chemie-International Edition*, 46(42):8063–8066, 2007.
- [134] Matthew A Pasek, Terence P Kee, David E Bryant, Alexander A Pavlov, and Jonathan I Lunine. Production of potentially prebiotic condensed phosphates via phosphorus redox chemistry. 2008.
- [135] Alan W Schwartz. Prebiotic phosphorylation-nucleotide synthesis with apatite. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, 281(4):477–480, 1972.
- [136] G Arrhenius, B Sales, S Mojzsis, and T Lee. Entropy and charge in molecular evolution—the case of phosphate. *Journal of theoretical biology*, 187(4):503–522, 1997.
- [137] KB Föllmi. The phosphorus cycle, phosphogenesis and marine phosphate-rich deposits. *Earth-Science Reviews*, 40(1-2):55–124, 1996.
- [138] Thomas Matreux, Almuth Schmid, Mechthild Rappold, Daniel Weller, Ayse Zeynep Caliskanoglu, Kelsey R Moore, Tanja Bosak, Donald B Dingwell, Konstantin Karaghiosoff, Francois Guyot, et al. Heat flows solubilize apatite to boost phosphate availability for prebiotic chemistry. *bioRxiv*, pages 2024–08, 2024.
- [139] E Thilo. Zur strukturchemie der kondensierten anorganischen phosphate. *Angewandte Chemie*, 77(23):1056–1066, 1965.
- [140] Walter Feldmann and Erich Thilo. Zur chemie der kondensierten phosphate und arsenate. xxxviii. amidotriphosphat. *Zeitschrift für anorganische und allgemeine Chemie*, 328(3-4):113–126, 1964.
- [141] Abdulakeem Osumah and Ramanarayanan Krishnamurthy. Diamidophosphate (dap): A plausible prebiotic phosphorylating reagent with a chem to biochem potential? *ChemBioChem*, 22(21):3001–3009, 2021.
- [142] Anthony D Keefe and Stanley L Miller. Are polyphosphates or phosphate esters prebiotic reagents? *Journal of molecular evolution*, 41:693–702, 1995.
- [143] Roland C Rouse, Donald R Peacor, and Robert L Freed. Pyrophosphate groups in the structure of canaphite, $\text{Ca}_2\text{P}_2\text{O}_7 \cdot 4\text{H}_2\text{O}$; the first occurrence of a condensed phosphate as a mineral. *American Mineralogist*, 73(1-2):168–171, 1988.
- [144] ALAN Schwartz and Cyril Ponnampertuma. Phosphorylation of adenosine with linear polyphosphate salts in aqueous solution. *Nature*, 218(5140):443–443, 1968.
- [145] Y Yamagata, H Watanabe, M Saitoh, and T Namba. Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature*, 352(6335):516–519, 1991.

- [146] Michael J Toplis, Guy Libourel, and Michael R Carroll. The role of phosphorus in crystallisation processes of basalt: an experimental study. *Geochimica et Cosmochimica Acta*, 58(2):797–810, 1994.
- [147] Nils G Holm. Glasses as sources of condensed phosphates on the early earth. *Geochemical Transactions*, 15:1–4, 2014.
- [148] R Osterberg and LE Orgel. Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *Journal of molecular evolution*, 1:241–248, 1972.
- [149] Maheen Gull, Mike A Mojica, Facundo M Fernández, David A Gaul, Thomas M Orlando, Charles L Liotta, and Matthew A Pasek. Nucleoside phosphorylation by the mineral schreibersite. *Scientific reports*, 5(1):17198, 2015.
- [150] Arthur Omran, Josh Abbatiello, Tian Feng, and Matthew A Pasek. Oxidative phosphorus chemistry perturbed by minerals. *Life*, 12(2):198, 2022.
- [151] AW Schwartz. Specific phosphorylation of the 2'-and 3'-positions in ribonucleosides. *Journal of the Chemical Society D: Chemical Communications*, (23):1393a–1393a, 1969.
- [152] Dieter Braun, Saroj Rout, Sreekar Wunnavu, Matthew Powner, Judit Šponer, and Christof Mast. Amino acids catalyze rna formation under ambient alkaline conditions. 2024.
- [153] Yoshihiro Furukawa, Hiromoto Nakazawa, Toshimori Sekine, Takamichi Kobayashi, and Takeshi Kakegawa. Nucleobase and amino acid formation through impacts of meteorites on the early ocean. *Earth and Planetary Science Letters*, 429:216–222, 2015.
- [154] Yasuhiro Oba, Yoshinori Takano, Hiroshi Naraoka, Naoki Watanabe, and Akira Kouchi. Nucleobase synthesis in interstellar ices. *Nature communications*, 10(1):4413, 2019.
- [155] NM Chung, R Lohrmann, LE Orgel, and J Rabinowitz. The mechanism of the trimetaphosphate-induced peptide synthesis. *Tetrahedron*, 27(6):1205–1210, 1971.
- [156] Joseph Rabinowitz, Jose Flores, Rita Krebsbach, and George Rogers. Peptide formation in the presence of linear or cyclic polyphosphates. *Nature*, 224(5221):795–796, 1969.
- [157] Hideko INOUE, Yoshinobu BABA, Tomoko FURUKAWA, Yasuyo MAEDA, and Mitsutomo TSUHAKE. Formation of dipeptide in the reaction of amino acids with cyclo-triphosphate. *Chemical and pharmaceutical bulletin*, 41(11):1895–1899, 1993.
- [158] Moran Frenkel-Pinter, Mousumi Samanta, Gonen Ashkenasy, and Luke J Leman. Prebiotic peptides: Molecular hubs in the origin of life. *Chemical reviews*, 120(11):4707–4765, 2020.
- [159] Joseph Rabinowitz. Peptide and amide bond formation in aqueous solutions of cyclic or linear polyphosphates as a possible prebiotic process. *Helvetica Chimica Acta*, 53(6):1350–1355, 1970.
- [160] Grégoire Danger, Laurent Boiteau, Hervé Cottet, and Robert Pascal. The peptide formation mediated by cyanate revisited. n-carboxyanhydrides as accessible intermediates in the decomposition of n-carbamoylamino acids. *Journal of the American Chemical Society*, 128(23):7412–7413, 2006.
- [161] J Rabinowitz and A Hampai. Quantitative polyphosphate-induced “prebiotic” peptide formation in h₂o by addition of certain azoles and ions. *Journal of molecular Evolution*, 21:199–201, 1985.
- [162] Izabela Sibilska, Yu Feng, Lingjun Li, and John Yin. Trimetaphosphate activates prebiotic peptide synthesis across a wide range of temperature and ph. *Origins of Life and Evolution of Biospheres*, 48:277–287, 2018.
- [163] Junpei Yamanaka, Katsuhiko Inomata, and Yukio Yamagata. Condensation of oligoglycines with trimeta-and tetrametaphosphate in aqueous solutions. *Origins of Life and Evolution of the Biosphere*, 18:165–178, 1988.
- [164] Hayley Boigenzahn and John Yin. Glycine to oligoglycine via sequential trimetaphosphate activation steps in drying environments. *Origins of Life and Evolution of Biospheres*, 52(4):249–261, 2022.
- [165] Izabela K Sibilska, Bingming Chen, Lingjun Li, and John Yin. Effects of trimetaphosphate on abiotic formation and hydrolysis of peptides. *Life*, 7(4):50, 2017.
- [166] Feng Ni, Shuting Sun, Chao Huang, and Yufen Zhao. N-phosphorylation of amino acids by trimetaphosphate in aqueous solution—learning from prebiotic synthesis. *Green Chemistry*, 11(4):569–573, 2009.
- [167] Mitsutomo Tsuhako, Akemi Nakajima, Tohru Miyajima, Shigeru Ohashi, Hiroyuki Narai, and Itaru Motooka. The reaction of

Bibliography

- cyclo-triphosphate with l- α - or β -alanine. *Bulletin of the Chemical Society of Japan*, 58(11):3092–3098, 1985.
- [168] David Ross and David Deamer. Prebiotic oligomer assembly: what was the energy source? *Astrobiology*, 19(4):517–521, 2019.
- [169] Jianxi Ying, Songsen Fu, Xin Li, Liubin Feng, Pengxiang Xu, Yan Liu, Xiang Gao, and Yufen Zhao. A plausible model correlates prebiotic peptide synthesis with the primordial genetic code. *Chemical Communications*, 54(62):8598–8601, 2018.
- [170] Jean-Philippe Biron, Alastair L Parkes, Robert Pascal, and John D Sutherland. Expeditionary, potentially primordial, aminoacylation of nucleotides. *Angewandte Chemie*, 117(41):6889–6892, 2005.
- [171] Helmut Griesser, Maren Bechthold, Peter Tremmel, Eric Kervio, and Clemens Richert. Amino acid-specific, ribonucleotide-promoted peptide formation in the absence of enzymes. *Angewandte Chemie International Edition*, 56(5):1224–1228, 2017.
- [172] Tommaso Bellini, Giuliano Zanchetta, Tommaso P Fraccia, Roberto Cerbino, Ethan Tsai, Gregory P Smith, Mark J Moran, David M Walba, and Noel A Clark. Liquid crystal self-assembly of random-sequence dna oligomers. *Proceedings of the National Academy of Sciences*, 109(4):1110–1115, 2012.
- [173] G Zanchetta, M Nakata, M Buscaglia, NA Clark, and T Bellini. Liquid crystal ordering of dna and rna oligomers with partially overlapping sequences. *Journal of Physics: Condensed Matter*, 20(49):494214, 2008.
- [174] Itay Budin and Jack W Szostak. Expanding roles for diverse physical phenomena during the origin of life. *Annual review of biophysics*, 39(1):245–263, 2010.
- [175] Tony Z Jia and Tommaso P Fraccia. Liquid crystal peptide/dna coacervates in the context of prebiotic molecular evolution. *Crystals*, 10(11):964, 2020.
- [176] John B Corliss, Jack Dymond, Louis I Gordon, John M Edmond, Richard P von Herzen, Robert D Ballard, Kenneth Green, David Williams, Arnold Bainbridge, Kathy Crane, et al. Submarine thermal springs on the galapagos rift. *Science*, 203(4385):1073–1083, 1979.
- [177] Ja Baross, SE Hoffman, et al. An hypothesis concerning the relationships between submarine hot springs and the origin of life on earth. In *Oceanologica Acta, Special issue*. Gauthier-Villars, 1981.
- [178] Matthew S Dodd, Dominic Papineau, Tor Grenne, John F Slack, Martin Rittner, Franco Pirajno, Jonathan O’Neil, and Crispin TS Little. Evidence for early life in earth’s oldest hydrothermal vent precipitates. *Nature*, 543(7643):60–64, 2017.
- [179] Deborah S Kelley, Jeffrey A Karson, Donna K Blackman, Gretchen L Fruh-Green, David A Butterfield, Marvin D Lilley, Eric J Olson, Matthew O Schrenk, Kevin K Roe, Geoff T Lebon, et al. An off-axis hydrothermal vent field near the mid-atlantic ridge at 30 n. *Nature*, 412(6843):145–149, 2001.
- [180] Barry Herschy, Alexandra Whicher, Eloi Camprubi, Cameron Watson, Lewis Dartnell, John Ward, Julian RG Evans, and Nick Lane. An origin-of-life reactor to simulate alkaline hydrothermal vents. *Journal of molecular evolution*, 79:213–227, 2014.
- [181] Laura M Barge and Roy E Price. Diverse geochemical conditions for prebiotic chemistry in shallow-sea alkaline hydrothermal vents. *Nature Geoscience*, 15(12):976–981, 2022.
- [182] Hannes Mutschler, Aniela Wochner, and Philipp Holliger. Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry*, 7(6):502–508, 2015.
- [183] Emilie Yeonwha Song, Eddy Ivanhoe Jiménez, Huacan Lin, Kristian Le Vay, Ramanarayanan Krishnamurthy, and Hannes Mutschler. Prebiotically plausible rna activation compatible with ribozyme-catalyzed ligation. *Angewandte Chemie International Edition*, 60(6):2952–2957, 2021.
- [184] Robert Hieronymus, Simon Peter Godehard, Darko Balke, and Sabine Müller. Hairpin ribozyme mediated rna recombination. *Chemical Communications*, 52(23):4365–4368, 2016.
- [185] Charles W Carter Jr. What rna world? why a peptide/rna partnership merits renewed experimental attention. *Life*, 5(1):294–320, 2015.
- [186] Felix Müller, Luis Escobar, Felix Xu, Ewa Węgrzyn, Milda Nainytė, Tynchtyk Amatov, Chun-Yin Chan, Alexander Pichler, and Thomas Carell. A prebiotically plausible scenario of an rna-peptide world. *Nature*, 605(7909):279–284, 2022.

List of Figures

1.1	Abiogenesis window	1
1.2	RNA world hypothesis	3
1.3	Chapters outline	5
2.1	Chapter summary	7
2.2	2',3'-cyclic phosphate nucleotide (N-cP) polymerization mechanism	9
2.3	Thermal trap front view and installation	11
2.4	Representative screenshot of the workflow in the Spectral Browser	13
2.5	G oligomers standards 2 to 10mers	14
2.6	Comparison of the retention times of the -cP versus the -P ending oligomers	15
2.7	Error bars calculated as standard deviation for 5 replicates	16
2.8	G polymerization	18
2.9	Effect of pH and the presence of a catalyst over the G-cP polymerization	19
2.10	Kinetic and effect of the temperature over G-cP polymerization	19
2.11	Homopolymers of G-cP, A-cP, C-cP and U-cP	20
2.12	Mixed GC polymerization products composition	20
2.13	Conditions influencing the GC mixed polymerization	21
2.14	G quadruplexes: one possible supramolecular assembly	22
2.15	Microfluidic chamber under a temperature gradient, mimicking in a heated rock pore	23
2.16	Comparison of the oligomerization efficiency between the reaction in dried state and inside the trap	24
2.A.1	Thermal trap parts and building	27
2.B.1	Example of a routine of G ₇ -3'P using the Spectral Browser	28
2.C.1	Mixed polymerization products composition	29
2.D.1	Comparison of the oligomerization efficiency between the reaction in dried state and inside the trap	30
2.E.1	Nuclease P1 digestion conducted on polymerized samples of G-cP to determine the amount of linkages 3'-5' and 2'-5'	31
3.1	Chapter summary	33
3.2	User Interface of Spectral Browser version 4.63	37
3.3	Effect of the pH adjustment during wet-dry cycling on the polymerization of G-cP	40
3.4	Polymerization yields for G, A, C and U	41
3.5	Polymers length for G, A, C and U	42
3.6	Remaining cyclic-ending reactants after reaction	42
3.7	High concentration phase and dry state imaging by microscope	43
3.8	Hydrolysis yields for the monomers G, A, C and U	44
3.9	AU mixed polymerization yields and products length:	45
3.10	AU mixed polymerization products compositions	46

3.11 GC mixed polymerization yields and products length:	47
3.12 GC mixed polymerization products compositions	48
3.13 Microscope images of the binary mixtures AU and GC	48
3.14 AUGC mixed polymerization yields and products length:	49
3.15 AUGC mixed polymerization products compositions	50
3.16 PTOM images for AUGC samples in the fully dried state at pH 10, 11 and 12 . .	51
3.A.1 Masses calculation	53
3.A.2 User interface of Spectral Browser step 2	54
3.A.3 Front panel of the subVI "Load and average spectra between Cursors from ms1"	54
3.A.4 User interface of Spectral Browser steps 4-5	55
3.B.1 Calibration standards	56
3.C.1 G polymerization yields for the full pH range 6, 7, 8, 9, 10, 11, 12	57
4.1 Chapter summary	59
4.2 Separation of the standards on the HPLC column	65
4.3 Phosphorylation of G nucleoside by orthophosphate, temperature and pH screen- ing	68
4.4 2',3'-cyclic phosphate nucleoside (N-cP) synthesis by the phosphorylation of nucleoside by orthophosphate at pH 10	68
4.6 Phosphorylation of G nucleoside by trimetaphosphate, temperature and pH screening	70
4.7 Effect of increasing urea quantities on the phosphorylation of nucleoside . .	71
4.8 Effect of amino acids on the phosphorylation of G by trimetaphosphate . . .	72
4.9 Effect of the wet-dry cycling on the phosphorylation of G by TMP in presence of valine	73
4.10 2',3'-cyclic phosphate nucleoside (N-cP) synthesis by the recyclization of 2'- and 3'-phosphate nucleoside by trimetaphosphate at pH 10	74
4.11 Effect of amino acids on the cyclization of G-P by trimetaphosphate	75
4.12 Reaction products between trimetaphosphate and valine	76
4.13 Proposed mechanism of dipeptide formation by TMP	77
4.14 Valine-phosphorylated guanosine hybrid compound	78
4.A.1 Standards used for quantification	81
4.A.2 Amino acids structures	82
4.B.1 Effect of different amino acids on the direct phosphorylation of G nucleoside	83
4.B.2 Effect of different amino acids on the cyclization and further phosphorylation of G monophosphate	84
4.C.1 Phosphorylated amino acids and dipeptide products	85
4.D.1 N-phosphoramidate amino acids guanosine	86
5.1 Outlook summary	87

List of Tables

2.1	Coefficient values for the calibration of the standards G_{2-10}	16
4.1	Calibration standards for the products without commercially available standards	66

List of Publications

5.1 First Author publications

List of (shared-) first publications published during the course of the doctoral studies:

Avinash Vicholous Dass*, Sreekar Wunnava*, **Juliette Langlais***, Beatriz von der Esch, Maik Krusche, Lennard Ufer, Nico Chrisam, Romeo C. A. Dubini, Florian Gartner, Severin Angerpointner, Christina Dirscherl, Petra Rovó, Christof B. Mast, Judith E. Šponer, Christian Oschsenfeld, Erwin Frey, Dieter Braun; RNA Oligomerisation without Added Catalyst from 2',3'-Cyclic Nucleotides by Drying at Air-Water Interfaces, *ChemSystemsChem* (**2022**), doi: 10.1002/syst.202200026

Manuscript in preparation (working title)

Federico Caimi*, **Juliette Langlais***, Francesco Fontana*, Sreekar Fontana, Rout Saroj Kumar, Tommaso Bellini°, Dieter Braun°, Tommaso Fraccia°; High-yield abiotic RNA polymerization from mixtures of all four 2',3'-cyclic nucleotides under mildly alkaline wet-dry cycling.

5.2 Contributed publications

List of contributed publications published during the course of the doctoral studies:

Emre Tekin, Annalena Salditt, Philipp Schwintek, Sreekar Wunnava, **Juliette Langlais**, James Saenz, Dora Tang, Petra Schwillle, Christof Mast, Dieter Braun; Prebiotic Foam Environments to Oligomerize and accumulate RNA, *ChembioChem* (**2022**), doi:10.1002/cbic.202200423

Manuscript in preparation

Christine M. E. Kriebisch, **Juliette Langlais**, Brigitte A. K. Kriebisch, ANton S. Meier, Bernhard Rieger, Dieter Braun, Job Boekhoven°; The Multifunctional Role of Templates in chemically Fueled Dynamic Combinatorial Libraries.

* Equal contribution

° Corresponding authors

RNA Oligomerisation without Added Catalyst from 2',3'-Cyclic Nucleotides by Drying at Air-Water Interfaces**

Avinash Vicholous Dass^{+, [a]}, Sreekar Wunnava^{+, [a]}, Juliette Langlais^{+, [a]}, Beatriz von der Esch,^[b] Maik Krusche,^[a] Lennard Ufer,^[a] Nico Chrisam,^[a] Romeo C. A. Dubini,^[d] Florian Gartner,^[f] Severin Angerpointner,^[f] Christina F. Dirscherl,^[a] Petra Rovó,^[d, e] Christof B. Mast,^[a] Judit E. Šponer,^[g] Christian Ochsenfeld,^[b, c] Erwin Frey,^[f] and Dieter Braun^{*, [a]}

For the emergence of life, the abiotic synthesis of RNA from its monomers is a central step. We found that in alkaline, drying conditions in bulk and at heated air-water interfaces, 2',3'-cyclic nucleotides oligomerised without additional catalyst, forming up to 10-mers within a day. The oligomerisation proceeded at a pH range of 7–12, at temperatures between 40–80 °C and was marginally enhanced by K⁺ ions. Among the canonical ribonucleotides, cGMP oligomerised most efficiently. Quantification was performed using HPLC coupled to ESI-TOF by fitting

the isotope distribution to the mass spectra. Our study suggests a oligomerisation mechanism where cGMP aids the incorporation of the relatively unreactive nucleotides C, A and U. The 2',3'-cyclic ribonucleotides are byproducts of prebiotic phosphorylation, nucleotide syntheses and RNA hydrolysis, indicating direct recycling pathways. The simple reaction condition offers a plausible entry point for RNA to the evolution of life on early Earth.

Introduction

The central and multifunctional role of RNA within biology points towards RNA as a chief informational biopolymer for the onset of molecular evolution.^[1] Polymerisation involving more than a single type of canonical nucleotide, generating a varied pool of RNA strands, has not been achieved under aqueous conditions.^[2–6] Chemical activation strategies are deployed to trigger RNA polymerisation^[3,7,8] and template-directed primer extension of sequences.^[9,10] In the earliest self-replicating systems, the formation of complementary strands for replication and transfer of genetic information by non-enzymatic processes is believed to be important and homopolymers are not considered very useful as genes.^[11] Short RNA strands, especially from dimers^[11] to tetramers^[12,13] have been shown to enhance the copying of mixed-sequence templates in comparison to

monomers. Thus, it is necessary to have a oligomerisation mechanism that is able to generate short mixed-sequences that later function as primers and templates for copying of longer sequences.

We base this study on 2',3'-cyclic mononucleotides (cNMP) which (a) possess an intrinsically activated phosphate; (b) are products of several prebiotic phosphorylation and nucleotide syntheses;^[14–18] and (c) are products of neutral to alkaline chemical and enzymatic hydrolyses of RNA.^[19–23] In comparison, the dry oligomerisation of 3',5'-cGMP^[24–26] did not foster the oligomerisation of the other ribonucleotides.^[26] Orgel and coworkers, reported conditions for 2',3'-cAMP oligomerisation by drying for 40 days with a 5-fold excess of ethane-1,2-diamine and yields up to 0.67 % of 14-mers.^[4,6] Other catalysts such as imidazole or urea required temperatures up to 85 °C and offered lower yields.^[4,6]

[a] Dr. A. V. Dass,⁺ S. Wunnava,⁺ J. Langlais,⁺ M. Krusche, L. Ufer, N. Chrisam, C. F. Dirscherl, Dr. C. B. Mast, Prof. Dr. D. Braun
Faculty of Physics, Systems Biophysics
Ludwig-Maximilians-Universität München
Amalienstraße 54, 80799, Munich, Germany
E-mail: dieter.braun@lmu.de

[b] B. von der Esch, Prof. Dr. C. Ochsenfeld
Chair of Theoretical Chemistry, Department of Chemistry,
Ludwig-Maximilians-Universität München
Butentandstraße 5–13, 81377, Munich, Germany

[c] Prof. Dr. C. Ochsenfeld
Max Planck Institute for Solid State Research
Heisenbergstr. 1, 70569 Stuttgart, Germany

[d] R. C. A. Dubini, Dr. P. Rovó
Faculty of Chemistry and Pharmacy,
Ludwig-Maximilians-Universität München
Butentandstraße 5–13, 81377, Munich, Germany

[e] Dr. P. Rovó
Institute of Science and Technology Austria
c/o NMR Facility, Am Campus 1, 3400 Klosterneuburg, Austria

[f] Dr. F. Gartner, S. Angerpointner, Prof. Dr. E. Frey
Faculty of Physics, Statistical and biological physics,
Ludwig-Maximilians-Universität München,
Theresienstraße. 37, D-80333, Munich, Germany

[g] Dr. J. E. Šponer
Institute of Biophysics Academy of Sciences of the Czech Republic
Královopolská 135, 61265 Brno, Czech Republic

[*] These authors contributed equally to this work.

[**] A previous version of this manuscript has been deposited on a preprint server (<http://doi.org/10.26434/chemrxiv-2022-zwh2-t-v2>).

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/syst.202200026>

© 2022 The Authors. ChemSystemsChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

We found that 2',3'-cGMP oligomerised spontaneously under alkaline (pH 7–12) drying conditions (40–80 °C), within a day. The other canonical cNMP were relatively inert under similar conditions. Our observations of cAMP and cUMP forming up to trimers are consistent with literature.^[27,28] The oligomerisation is demonstrated in the presence of bulk water at the air-water interface, within a microfluidic thermal chamber. The chamber mimics conditions of a heated, water filled volcanic rock pore that includes a gas bubble.

In an oligomerisation mixture of cNMP, we observed oligomers rich in G nucleotides, but with C, A and U incorporated at lower concentrations. Computational and modelling results suggest that the oligomers of cGMP form a self-assembled scaffold in the dry state, which could incorporate the nucleotides C, A and U to form short mixed-sequence oligomers.

Results

Polymerisation of cGMP

An aqueous solution of the sodium salt of 2',3'-cGMP (20 mM) was dried for 18 hours at 40 °C in the presence of 40 mM KCl. Since the monomers are monosodium salts, there was an equal concentration of Na⁺ ions when in solution (20 mM). All the reported concentrations throughout the article are calculated for a volume of 100 µL. The total concentration of each n-mer (oligomer) is a sum of oligomers containing the linear-phosphate (-P) and the cyclic-phosphate (-cP) on the n-mer terminus. Both endings are well discriminated by HPLC as the n-mer-cP is eluted before a n-mer-P of the same length (S2d). Typically, about 90% of the n-mers consisted of -P endings (S5d). Due to propensity of purines to form non-covalent aggregates in mass spectrometry detection,^[29] a combination of HPLC and ESI-TOF techniques were used for detection of oligonucleotides. The non-covalent stacked n-mers (eg. two 4-mers) are discriminated from covalent n-mers (eg. an 8-mer) due to the higher mass of the stacked n-mers by one H₂O in the MS and the corresponding HPLC retention times of n-mers under denaturing HPLC conditions.^[30–32]

The denaturing conditions of the HPLC column at 60 °C efficiently resolved synthetic oligoG n-mers without signs of aggregation, as shown in Figure 1c. It must be noted that an n-mer-cP and a cyclised n-mer of the same length would have the same mass, but are unlikely not to be discriminated by the HPLC retention times. The presence of n-mer-cP is established from the ³¹P NMR peak at ~20 ppm in Figure 1d. Oligomers from 2- to 15-mers (S8a) were detected by HPLC-MS for cGMP oligomerisation. For quantification, only 2- to 10-mers were considered throughout the study.

The error bars can be estimated based on plots of cGMP oligomerisation (5 replicates) in S8a, with a mean standard deviation of 2.95 µM between independent runs of the experiment. The error bars are not indicated in the figures as they would appear insignificant on the log scale. For quantification, the HPLC retention times of the oligomer standards of G were

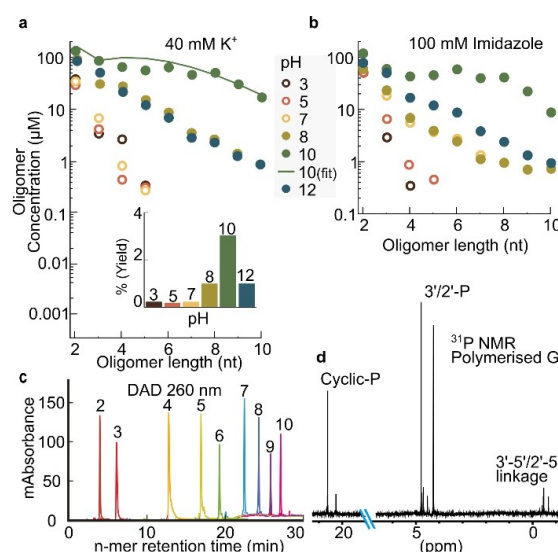


Figure 1. Oligomerisation of Guanosine-2',3'-cyclic monophosphate (cGMP-Na). A 20 mM cGMP-Na solution was heat-dried with 40 mM K⁺ at 40 °C for 18 hours, under ambient pressure in 100 µL volume. (a) Polymerisation was screened over a range of pH 3–12. The reported concentrations were the sum of terminal cyclic (-cP) and linear phosphate (-P) containing oligomers. Oligomers without terminal phosphates were not detected. Polymerisation was optimal at pH 10 with total oligomer yields of ~3.5% (inset). The solid line shows results of the polymerisation model based on stacked assembly (S19). (b) pH screen with 100 mM imidazole under similar conditions. No significant increase in oligomerisation was found by adding imidazole. (c) Diode array detector (DAD) absorbance at 260 nm for 50 µM oligoG standards (-P endings) and 100 µM KCl used for confirming HPLC separation and the determination of retention time for quantification with ion counts. (d) ³¹P proton-decoupled NMR spectrum (10% D₂O, pH 10), of oligomerised G sample: the signals corresponding to phosphodiester linkages for both 3'-5' and 2'-5' are between -0.8 and -1.1 ppm.

first optimised on an RP C-18 HPLC column coupled to ESI-TOF. Figure 1c shows the HPLC chromatogram of 2- to 10-mers for oligoG standards (with 1 eqv. of KCl) with their respective retention times. We found efficient separation and no evidence for the formation of aggregates. The ion counts of the n-mer with their HPLC retention times are shown in S2b. By comparing the ion counts, we confirmed the high efficiency of the post-polymerisation ethanol precipitation protocol and its negligible influence (S3). However, the precipitation was used to remove excess monomers which would otherwise saturate the HPLC column, yielding a robust method for the quantification of the complex oligonucleotide mixtures (S2c).

The calculated isotope probabilities of the n-mers in the various charge states were fitted to the raw mass spectra using a self-written LabView program. This allowed us to identify salt adducts formed in the mass spectrometer and to fit overlapping isotope patterns. The retention times of the oligoG standards were used to obtain time-brackets to sum the mass spectra. Further details on the calibration used for the quantification within the program and the functional modes of the program are elaborated in S1–S6. Based on preliminary enzymatic digestion experiments, we estimated that the formed G

oligomers were linked by 2'-5' and 3'-5' phosphodiester linkages in about 1:1 ratio (S21–S23). The linkage type in the oligomerisation was also confirmed by ^{31}P NMR (Figure 1d) and the peaks were assigned based on the literature values.^[33,34]

Figures 1a and 1b compare the effect of pH on the lengths and concentrations of the n-mers formed by drying with K^+ (Cl^-) and imidazole respectively. We determined the optimal reaction temperature to be 40 °C (S5, S8b). Imidazole and its derivatives are used in the literature as nucleotide activation agents for templated primer extension reactions,^[9] as a buffering agent and a catalyst for oligomerisation.^[4] The addition of imidazole did not enhance the length and concentration of n-mers in comparison to oligomerisation with K^+ .

Polymerisation from cNMP

We also tested the polymerisation tendencies of cAMP, cUMP and cCMP under the same heat-drying conditions and found that these monomers did not polymerise to the same lengths and concentrations as cGMP. Figure 2a shows that the polymerisation trend decreases in the order $\text{cGMP} > \text{cUMP} > \text{cAMP} > \text{cCMP}$. The dominance of G-polymerisation prompted us to investigate the copolymerisation of these moderately reactive mononucleotides under the influence of the well oligomerising cGMP. We found that a mixture of two or four different monomers was capable of generating mixed sequence oligomers, where the majority of the mixed oligomers were rich in G. We probed if the oligomerisation of a G and C mixture could reach levels where hybridisation between strands could be possible. Thus, we oligomerised a binary mixture of cGMP and cCMP (20 mM each), under heat-drying conditions (40 °C) in the presence of 40 mM KCl. Comparing quantities of C_2 in Figure 2a and 2b, the concentration of C_2 is enhanced 2 fold and C_3 became detectable; besides the fact that mixed GC oligomers are formed (Figure 2b). The detailed sequence composition for GC mixed polymerisation is seen in Figure 2e, showing that the G_2 to G_{10} contribute to the bulk of the oligomers formed in the polymerisation mixture. Up to two C's were incorporated into oligomers ≤ 4 -mers, one C is incorporated into 5-mers and none were detectable beyond them. A similar analysis of GA and GU binary mixtures is available in S9a, b.

GC mixed polymerisation was favoured at temperatures ranging from 40 °C to 80 °C (Figure 2c), similar to cGMP (S5b, S8b). It must be noted that in reactions at 30 °C for 18 hours, the drying was incomplete within the polypropylene tubes used for the experiment and the reaction kinetics in the dry state was reduced. Higher temperatures on the other hand possibly contributed to the degradation of the monomers (S4c) and the formed oligomers as seen in the trace comparisons under 80, 60 and 40 °C in Figure 2c.

Specific cations also influenced cNMP oligomerisation. We found that K^+ ions yielded higher concentrations and lengths of the oligomers in comparison to Na^+ ions at the same concentrations. The presence of Mg^{2+} ions in the reaction mixture inhibited polymerisation (Figure 2d). The dependence

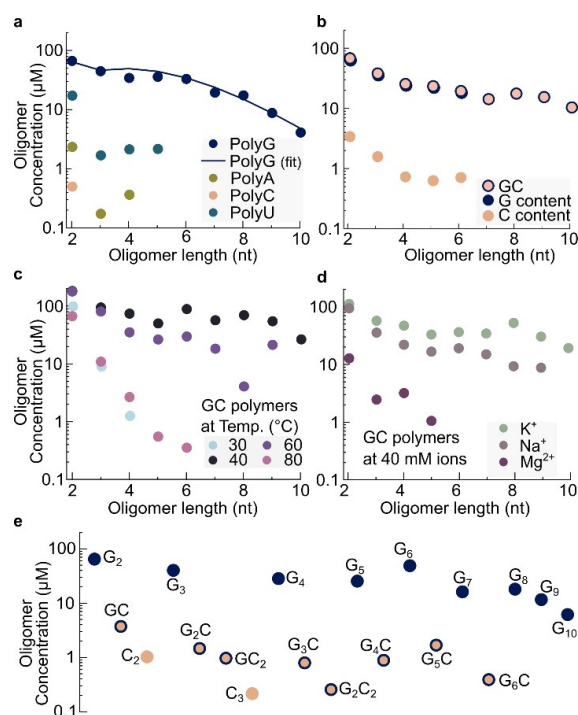


Figure 2. Oligomerisation of mixed Nucleotide 2',3'-cyclic monophosphate (cNMP). (a) Homooligomers of cGMP-Na, cAMP-Na, cCMP-Na, cUMP-Na were individually produced from a 20 mM solution at 40 °C for 18 hours. Oligomers of G were formed in far higher concentrations than polyC, A and U. The solid line shows results of the polymerisation model based on stacked assembly (S18). The approximately 3x lower yield for oligoG compared to Figure 1a is attributed to the lack of K^+ ions. (b) Oligomerisation of cGMP and cCMP at 40 °C with 40 mM K^+ . The base C is incorporated into the sequences in the presence of cGMP while only dimers were detected without it. (c) Temperature screening over a range of 30–80 °C for GC oligomerisation. Reduced concentration of n-mers > 3 is observed for 80 °C, possibly due to ring opening of the cyclic phosphate monomers (S4c). (d) The presence of 40 mM K^+ increased the concentration of n-mers while added Mg^{2+} quenched polymerisation (S10a). (e) Sequence composition of cGMP and cCMP mixed oligomers at 40 mM K^+ . Oligomers show G-rich n-mers and suggest the presence of all possible combinations in trimer sequences.

of polymerisation on K^+ , Na^+ and Mg^{2+} salt concentrations is shown in S10, indicating that 1–3 eqv. of the same cation display similar results, but the type of cation affected the efficiency of oligomerisation.

Polymerisation of cNMP in a heated rock pore mimic

Wet-dry cycles in surface-based geological settings are subjected to a drift in salt and pH conditions due to the imbalance caused by the evaporation of pure water and the rehydration of the fluid that contains salt. Wet-dry cycling can also occur in a closed chamber, subjected to a temperature gradient.^[35] The water that evaporates on the warm side re-enters the fluid on its cold side. This causes interface shifts and the dew droplet dynamics on the cold side, offering wet-dry cycles under

constant pH and salt conditions. The geological analogues of such a setting would be volcanic rock pores which are partially filled with fluids and are subjected to a thermal gradient. We have previously reported prebiotically important processes such as accumulation, phosphorylation, encapsulation, gelation, strand separation, enzymatic DNA replication and crystallisation within such settings.^[35–37]

For the polymerisation within this setting, we started with 20 mM total monomers (5 mM each of cG, cC, cA and cU). After the chamber was loaded with the monomer solution, a thermal gradient was applied which drove continuous wet-dry cycles just above the air-water interface inside the chamber (Figure 3a). Over time, the meniscus of the bulk liquid receded in an oscillatory manner depending on how many dew droplets formed above the interface; and dried material precipitated on the warm side as a consequence (Figure 3b). The dew droplets grew at the cooler side of the chamber by surface-tension driven fusion and made contact with the warm side, rehydrating the dried material and transporting it back into the bulk.^[37] This phenomenon was allowed to continue for 18 hours, after which the setup was dismantled and the remaining bulk liquid

and the dried flakes (after dissolution) were sampled for analysis.

The pH of the samples at the end the reaction was found to be lowered by a pH unit, indicating the formation of acidic species in the reaction mixture. A likely cause of the pH drop is the acidification by the ring opening of the cyclic phosphate in the mononucleotides and the oligomers (S4e, S5c, d). At higher temperatures, the pH drop was 1.5 to 2 pH units (S4e).

Despite the presence of bulk water, the oligomerisation inside the simulated volcanic-rock pore showed comparable yields as that of the heat-dried conditions. This indicates that the heated interface can access conditions favourable for polymerisation similar to bulk dried polymerisation conditions. The constant feeding of monomers from the bulk fluid could also be an important factor. A length-selective enzymatic DNA replication was reported recently within this setting, indicating the possible continuity of prebiotic chemistry in such a setting.^[37]

We observed all the dimer sequence combinations and most of the trimers (Figure 3c). However, the tetramers and pentamers are predominantly sequences rich in G. The length selectivity of the HPLC allowed the detection of longer sequences. However, the isotopic fit to the raw mass spectra provided by our LabView-based analysis showed that longer species with concentrations lower than 0.2 μM were lost in the background noise of the mass spectra. Moreover, different oligomers can have similar masses (eg. Table S3 and S4), so to avoid false positives, sequences with mass overlaps were not included here. This is in addition to the rigid selection criteria, based on fitting of the isotopic distribution (S12) and only considering mass spectra within the optimised n-mer retention times of the HPLC. A full sequence composition analysis for GC and GCAU mixtures with comparison between dry polymerisation and simulated rock-pore polymerisation is provided in S11. In comparison, CAU reaction mixture yielded only dimers (S9c), indicating again the central role of G in the copolymerisation process.

Computational study of the proposed intercalated stacked arrangement

Based on the hypothesis that a stack-assisted geometry is triggering the oligomerisation of 3',5'-cGMP,^[26] we studied the suitability of intercalated stack arrangements for the oligomerisation of 2',3'-cNMP. We explored the stability of the stack arrangements, and the incorporation of cNMP monomers into polymerised cGMP scaffold, based on minimum energy structures and molecular dynamics simulations (Figure 4a–c and S24–34).

To investigate the suggested intercalated stack arrangement for several possible species, we have computed the stacking interaction energies and evaluated the minimum energy geometries obtained at ω B97 M-V/def2-TZVPD level of theory.^[38–41] All systems were studied in the gas phase as well as with implicit solvation (C-PCM).^[42] The quantum mechanical

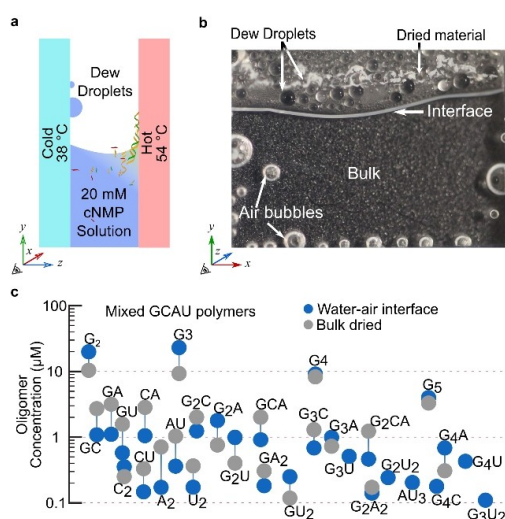


Figure 3. RNA oligomerisation in the vicinity of air inclusions in a heated simulated rock pore. (a) Side view. The chamber is 500 μm in depth and subjected to a heat flow with a temperature gradient of 38–54 $^{\circ}\text{C}$. (b) Front view. The thermal gradient drives continuous evaporation and recondensation in the air inclusions, triggering accumulation and wet-dry cycles. Molecules accumulated at the interface are dried from a receding interface due to evaporation. Rehydration is provided by dew droplets on the cold side which merge with the bulk solution due to surface tension.

(c) Oligomerisation of four canonical monomers: cGMP, cCMP, cAMP and cUMP, 5 mM each, 40 mM KCl at pH 10 for 18 hours. Especially for the longer strands, the oligomerisation in the simulated rock pore shows improved yields over the dry reaction. The physically triggered wet-dry cycling and length selectivity in this environment has been shown to drive efficient replication and selection cycles,^[37] making the finding of oligomerisation to provide the raw material for templated ligation very interesting. Moreover, this shows that oligomerisation under simulated geological conditions is possible without the need for arid conditions on early Earth. The trends show a rich set of mixed short sequences when all four nucleotides are mixed together for oligomerisation.

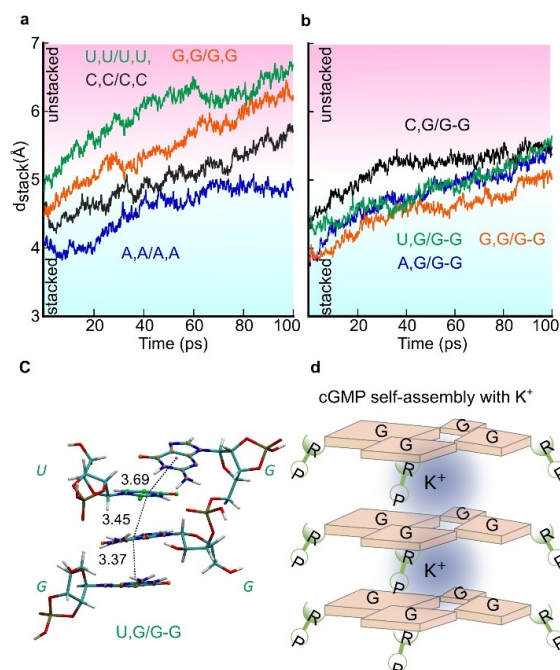


Figure 4. Possible supramolecular assemblies facilitating polymerisation of cNMP. Molecular dynamics simulations suggest polymerisation of A, U and C by intercalating into stacks of oligoG. (a) Distances between bases in the complex of unpolymerised nucleotides show that polymerisation is disfavoured due to drifting away of the complex. (b) The assembly formed when the bases are templated by covalently linked G–G (2′–5′) dimers, forms the most stable complexes and make possible the incorporation of the G, C, A, and U within the n-mers observed in our experiments. (c) Snapshot after energy minimization of stacking interaction between a oligomerised scaffold of cGMP (G–G) to template the cGMP and cUMP monomers (U,G). The dotted lines mark distances between the bases used to evaluate the stability of the complex. (d) However, the self-oligomerisation of oligoG could also be based on stacks of G-tetrads, stabilised by inner K^+ ions, coinciding with the promotion of oligoG formation by K^+ ions. R denotes the ribose of RNA.

computations were performed using FermiONs + γ ^[43–45] in combination with Chemshell.^[46]

These computations were complemented by GFN-FF molecular dynamics simulations,^[47] for the systems encapsulated in an explicit water sphere using xtb.^[48] The stacking of homogeneous monomers were tested (N,N/N,N) with N = A, U, G, C and the incorporation of monomers into a dimer and trimer scaffold of G was probed (N,N/G–G or N,N/G–G–G). The 3′–5′ linked G–G and G–G–G accommodate A, U and G monomers into the scaffold providing a stable arrangement for the initiation of polymerisation. For C an alternate arrangement involving hydrogen bonding with a G within the scaffold is observed (S28). We found that a 2′–5′ oligoG scaffold seemed to enhance the alignment (Figure 4a, b, c), confirmed both by static and dynamic computations (S30, S31).

Theoretical model of cGMP polymerisation

Additional evidence supporting a stacked polymerisation mechanism comes from the observed non-exponential length distribution of the oligomer concentrations. This supports the idea that the formation of dimers is the rate limiting step: the concentration drop from monomers to dimers was most significant. For the cGMP oligomerisation in Figure 1a, the 20 mM monomer concentrations drop to 0.15 mM for G_2 , then forming a flat concentration plateau, in contrast to the typical exponential length distribution in homogeneous polymerisation.^[49]

To test this idea, we fit the concentration distribution of G homooligomers with a stacked polymerisation model (solid line Figure 1a and 2a). The model assumed a three-step polymerisation reaction: i) a monomer of length i and an oligoG scaffold k can stack together with rate ν , ii) the de-stacking rate $\delta_{k,i}$ decreases exponentially with the number of stacked bases $n_{k,i}$, iii) another monomer of length j can stack to the complex. If the stacks persist long enough, the polymerisation reaction ligates the two monomers with rate ρ (see for details S18–S19). The model fits the experimental data, suggesting a rate limiting step for the formation of short oligomers due to the required mutual alignment. It should be noted that it is difficult to distinguish between inter-base stacking or a plausible G-tetrad arrangement suggested based on the enhanced polymerisation observed with K^+ (Figure 4d).

Discussion

Our data suggests that cGMP oligomerises in dry state at moderate temperatures and pH. The oligomerisation occurs over a range of temperatures (40–80 °C) and pH (7–12) and does not require additional catalysts, making this reaction robust. Dissolved gases and salts could adjust the pH of the environment, making RNA formation more probable under early Earth models.^[50,51] We also showed polymerisation in the wet-dry cycling environment at a heated air–water interface, adding RNA polymerisation to the pool of prebiotic processes possible within such a setting.^[35–37] The tested conditions of wet-dry cycles at an air–water interface or direct drying keep the reaction out of equilibrium. The cyclic monomers undergo polymerisation and ring-opening (Figure 1d), of which the ring-opening is still the dominant product at the tested temperatures (S4). Under the tested conditions, the reaction yielded oligomers up to 15-mers. The formed oligoG incorporated cCMP, cAMP and cUMP monomers, albeit in lower concentration, which did not homooligomerise significantly. As a rough comparison to the yields achieved by Verlander and Orgel with homooligomers of cAMP in the presence of ethane-1,2-diamine, we observed ~0.35% for a 6-mer of oligoG in 18 hours compared to 0.81% for polyA in 40 days.^[6]

An important feature of this oligomerisation is that the 2′,3′-cyclic phosphate group, under alkaline pH, is sufficient to trigger oligomerisation without ex-situ or in-situ activation mechanism or added catalysts, and under low salt conditions.

The finding that the oligomerisation starts without added catalysts – and that the reaction site is not yet blocked by a catalyst – is a very good starting point for Darwinian evolution to speed up this reaction rate. Low salt conditions are interesting for RNA evolution since they notably help strand separation and reduce RNA degradation.^[36]

cNMP oligomerisation is found to be a relatively clean reaction under the tested conditions. In comparison, *in situ* EDC activation yields side products, especially at high temperatures.^[52] We did not detect any major side products with ESI-TOF, other than the salt adducts of sodium and potassium.

The abiotic formation and recyclability of the cNMP monomers is feasible, as they are known to be produced under several phosphorylation conditions, nucleotide syntheses and are common degradation products of RNA.^[19–23] Thus, with the likelihood of finding catalytic boosts for this found reaction mode, a cycle of reactions involving polymerisation, oligomer extension, polymer hydrolysis and reactivation of monomers under early Earth conditions becomes conceivable. Furthermore, recombination and templated ligation involving 2',3'-cyclic ending oligomers^[53] have been observed.

For our studies, we compared two monovalent ions (K^+ , Na^+) and one divalent cation (Mg^{2+}). They were chosen for their relevance in contemporary cytosolic media, their abundance on the early Earth^[54] and for the role of Mg^{2+} in ribozyme activity.^[55] Polymerisation is enhanced in the presence of K^+ in comparison to Na^+ ions. The inhibition by Mg^{2+} ions possibly occurs by a combination of base catalysis mechanism, the deactivation of -cP ends of the reactant, products and enhanced oligomer hydrolysis. Despite its role in ribozyme functionality, at high concentrations Mg^{2+} inhibits RNA replication by creating strong RNA duplexes, limiting thermal denaturation and enhancing temperature dependent hydrolysis.^[56] It is also known that the presence of ~1.5 mM Mg^{2+} is sufficient to inhibit the membrane self-assembly of fatty acids and this has been considered an incompatible aspect for the co-emergence of RNA and fatty acid membranes.^[57,58] However, under the discussed reaction conditions of cNMP oligomerisation, RNA formation and encapsulation with fatty acids might be conceivable within freshwater locations on the primordial Earth. Moreover, we have shown that efficient strand separation can be achieved by low sodium concentrations, triggered by microscale water cycles within heated rock pores.^[36,37]

Our very preliminary digestion studies and ^{31}P NMR results suggest a considerable backbone heterogeneity (2'-5' and 3'-5') within the oligomers. However, a full quantitative treatment is beyond the scope of this study. It has been demonstrated that the presence of 2'-5' linkages allow efficient strand separation by reducing the melting temperature (T_m) of oligomers, which is pertinent in the case of G-rich sequences that are observed in this oligomerisation.^[59] Lowering of T_m is critical to replication of sequences.^[59,60] These studies also show that the presence of 2'-5' linkages allow the folding of RNA into three-dimensional structures, similar to native linkages and do not hinder the evolution of functional RNAs, such as ribozymes. The susceptibility towards enhanced hydrolysis of the 2'-5' over the 3'-5'

linkages could select the latter in wet-dry cycling conditions, similar to the reported backbone selection of RNA and DNA.^[59,61,62]

Mechanistically, molecular dynamics studies indicated that cGMP oligomerisation could be due to the formation of intercalated stacks of cGMP as a consequence of hydrophobic interactions between the guanine bases. On attaining a stable intermolecular arrangement, the 5'-OH of a nucleotide can attack the cyclic phosphate of the neighbouring nucleotide. This could allow the formation of oligomeric G-scaffolds (Figure 4c, d, S25). However, the formation of tetrad stacks over one another with a central K^+ ion between the stacks could also promote oligomerisation (Figure 4d).

The notion of multi-molecular assemblies is supported by the presence of slow-diffusing species observed in 1H , ^{31}P diffusion ordered spectroscopy (DOSY) of cGMP-KCl solution (S15, S16). Reports in literature point to self-assembly of 5'-GMP and 3'-GMP into helical stacks.^[63] The presence of several slow-diffusing species indicate a range of molecular environments, making it impossible to identify a single type of self-assembly by NMR. It has also been reported that G-quadruplex structures could be stable up to a pH of ~10.8 at ambient temperatures.^[64,65]

The formation of dimers appears to be a limiting step in the oligomerisation. Such a threshold behaviour is known to be an optimal control strategy for self-assembly processes.^[66] With this, monomers remain available in high concentration, leading to long-tailed, non-exponential polymer distributions. This limits the total efficiency of the polymerisation but favours the formation of the oligomers, important for downstream reactions such as templated replication.

It should be noted that an efficient generation of very long and random RNA sequences would make hybridisation and replication inconceivable. At this point, the generated G-rich sequences might not seem optimal for hybridisation and replication. However, a biased pool of short oligomers (10- to 15-mers) further constrains the sequence space, favouring selectivity and making templated replication plausible.^[11,67,68] We think that the findings are a first step to provide oligonucleotides for templated ligation and the emergence of an evolutionary dynamics with RNA.

Conclusion

We report the oligomerisation of canonical nucleotides that produced RNA of mixed sequences under drying conditions in bulk and at heated air-water interface. A wide range of temperatures (40–80 °C) and pH (7–12) promoted oligomerisation. Best yields were reported by mild heating (40 °C) of monomers at low salt concentrations and under alkaline drying conditions (pH 10). The reaction proceeded best at 1–2 equivalents of K^+ and Na^+ , while Mg^{2+} ions inhibited it. In an equal mixture of four nucleotides, equal incorporation of all four was not observed and the mixed sequences were dominated by G. However, 2',3'-cGMP fostered the incorporation of the otherwise scarcely reactive C, A, and U, generating short, mixed

sequences. This reaction under the tested temperatures, pH and salt conditions provide a novel route to fresh water oligomerisation towards short RNA strands, an important intermediate step towards providing the raw materials for an RNA-based emergence of life.

Acknowledgements

We would like to thank Ulrich Gerland, Tobias Göppel, Joachim Rosenberger and Bernhard Altaner for their helpful remarks and discussions; Thomas Matreux, Alexandra Kühnlein, Noël Yeh Martin and Maximilian Weingart for comments on the manuscript. The authors thank J. Kussmann (LMU Munich) for providing a development version of the FermiONs++ program package. Financial support was provided by the European Research Council (ERC Evotrap, grant no. 787356, the Simons Foundation (grant no. 327125), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 364653263 – TRR 235 (CRC 235), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311, and the Center for NanoScience. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: RNA · Polymerisation · Prebiotic chemistry · Non-equilibrium · Air-water interfaces

- [1] H. S. Bernhardt, *Biol. Direct* **2012**, *7*, 1–10.
- [2] G. Costanzo, S. Pino, A. M. Timperio, J. E. Šponer, J. Šponer, O. Nováková, O. Šedo, Z. Zdráhal, E. Di Mauro, *PLoS One* **2016**, *11*, 1–14.
- [3] J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59–61.
- [4] M. S. Verlander, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 303–316.
- [5] C. V. Mungi, N. V. Bapat, Y. Hongo, S. Rajamani, *Life* **2019**, *9*, 1–11.
- [6] M. S. Verlander, L. E. Orgel, *J. Mol. Evol.* **1974**, *3*, 115–120.
- [7] A. Luther, R. Brandsch, G. Von Kiedrowski, *Nature* **1998**, *396*, 245–248.
- [8] S. J. Zhang, D. Duzdevich, D. Ding, J. W. Szostak, *Proc. Nat. Acad. Sci.* **2022**, *119*, 2021.09.07.459201.
- [9] T. Walton, W. Zhang, L. Li, C. P. Tam, J. W. Szostak, *Angew. Chem. Int. Ed.* **2019**, *58*, 10812–10819; *Angew. Chem.* **2019**, *131*, 10926–10933.
- [10] E. Kervio, M. Sosson, C. Richert, *Nucleic Acids Res.* **2016**, *44*, 5504–5514.
- [11] C. Richert, G. Leveau, D. Pfeffer, B. Altaner, E. Kervio, U. Gerland, F. Welsch, *Angew. Chem. Int. Ed.* **2022**, 202203067, 1–6.
- [12] L. Li, N. Prywes, C. P. Tam, D. K. O'Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, J. W. Szostak, *J. Am. Chem. Soc.* **2017**, *139*, 1810–1813.
- [13] D. K. O'Flaherty, N. P. Kamat, F. N. Mirza, L. Li, N. Prywes, J. W. Szostak, P. Sheeringa, *J. Am. Chem. Soc.* **1997**, *119*, 1–5.
- [14] E. I. Jiménez, C. Gibard, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2020**, *60*, 19, 1077–10783, DOI 10.1002/anie.202015910.
- [15] Z. Liu, L. F. Wu, J. Xu, C. Bonfio, D. A. Russell, J. D. Sutherland, *Nat. Chem.* **2020**, *12*, 3–10.
- [16] H. J. Kim, S. A. Benner, *Astrobiology* **2021**, *21*, 298–306.
- [17] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.
- [18] Y. Yamagata, H. Inoue, K. Inomata, *Origins Life Evol. Biospheres* **1995**, *25*, 47–52.
- [19] R. Breslow, *Acc. Chem. Res.* **1991**, *24*, 317–324.
- [20] Y. Li, R. R. Breaker, *J. Am. Chem. Soc.* **1999**, *121*, 5364–5372.
- [21] H. Peng, B. Latifi, S. Müller, A. Lupták, I. A. Chen, *RSC Chem. Biol.* **2021**, *2*, 1370–1383.
- [22] A. M. Pyle, *Science* **1993**, *261*, 709–714.
- [23] S. I. Nakano, D. M. Chadalavada, P. C. Bevilacqua, *Science* **2000**, *287*, 1493–1497.
- [24] M. Morasch, C. B. Mast, J. K. Langer, P. Schilcher, D. Braun, *ChemBioChem* **2014**, *15*, 879–883.
- [25] J. E. Šponer, J. Šponer, A. Giorgi, E. Di Mauro, S. Pino, G. Costanzo, *J. Phys. Chem. B* **2015**, *119*, 2979–2989.
- [26] S. Wunna, C. F. Dirscherl, J. Výravský, A. Kovařík, R. Matyášek, J. Šponer, D. Braun, J. E. Šponer, *Chem. A Eur. J.* **2021**, *27*, 17581–17585, DOI 10.1002/chem.202103672.
- [27] C. M. Tapiero, J. Nagyvary, *Nature* **1971**, *231*, 42–43.
- [28] S. Dagar, S. Sarkar, S. Rajamani, *RNA* **2020**, *26*, 756–769.
- [29] J. E. Šponer, J. Šponer, A. Kovařík, O. Šedo, Z. Zdráhal, G. Costanzo, E. Di Mauro, *Life* **2021**, *11*, 1–13.
- [30] A. Premstaller, P. J. Oefner, *LCGC Eur.* **2002**, *15*, 7, 410–422.
- [31] A. Premstaller, P. J. Oefner, *Denaturing High-Performance Liquid Chromatography*, Humana Press, New Jersey, n.d.
- [32] P. B. Danielson, R. Kristinsson, R. J. Shelton, G. S. LaBerge, *Expert Rev. Mol. Diagn.* **2005**, *5*, 53–63.
- [33] H. R. Palmer, J. J. Bedford, J. P. Leader, R. A. J. Smith, *J. Biol. Chem.* **2000**, *275*, 27708–27711.
- [34] S. Motsch, D. Pfeffer, C. Richert, *ChemBioChem* **2020**, *21*, 2013–2018.
- [35] M. Morasch, J. Liu, C. F. Dirscherl, A. Ianeselli, A. Kühnlein, K. Le Vay, P. Schwintek, S. Islam, M. K. Corpinot, B. Scheu, D. B. Dingwell, P. Schwill, H. Mutschler, M. W. Powner, C. B. Mast, D. Braun, *Nat. Chem.* **2019**, *11*, 779–788.
- [36] A. Ianeselli, C. B. Mast, D. Braun, *Angew. Chem. Int. Ed.* **2019**, *58*, 13155–13160; *Angew. Chem.* **2019**, *131*, 13289–13294.
- [37] A. Ianeselli, M. Atienza, P. W. Kudella, U. Gerland, C. B. Mast, D. Braun, *Nat. Phys.* **2022**, *18*, 579–585, DOI 10.1038/s41567-022-01516-z.
- [38] N. Mardirossian, M. Head-Gordon, *J. Chem. Phys.* **2016**, *144*, DOI 10.1063/1.4952647.
- [39] O. A. Vydrov, T. Van Voorhis, *J. Chem. Phys.* **2010**, *133*, DOI 10.1063/1.3521275.
- [40] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- [41] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [42] M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681.
- [43] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *138*, DOI 10.1063/1.4796441.
- [44] J. Kussmann, C. Ochsenfeld, *J. Chem. Theory Comput.* **2015**, *11*, 918–922.
- [45] J. Kussmann, C. Ochsenfeld, *J. Chem. Theory Comput.* **2017**, *13*, 3153–3159.
- [46] P. Sherwood, A. H. De Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *J. Mol. Struct.* **2003**, *632*, 1–28.
- [47] S. Spicher, S. Grimme, *Angew. Chem.* **2020**, *132*, 15795–15803; *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.
- [48] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, 1–49.
- [49] C. B. Mast, S. Schink, U. Gerland, D. Braun, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8030–8035.
- [50] J. F. Kasting, M. T. Howard, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1733–1741.
- [51] J. Krissansen-Totton, G. N. Arney, D. C. Catling, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4105–4110.
- [52] E. Edeleva, A. Salditt, J. Stamp, P. Schwintek, J. Boekhoven, D. Braun, *Chem. Sci.* **2019**, *10*, 5807–5814.
- [53] A. V. Lutay, E. L. Chernolovskaya, M. A. Zenkova, V. V. Vlasov, *Dokl. Biochem. Biophys.* **2005**, *401*, 163–166.
- [54] S. Maurer, *Life* **2017**, *7*, DOI 10.3390/life7040044.
- [55] T. S. Lee, C. S. López, G. M. Giambaşu, M. Martick, W. G. Scott, D. M. York, *J. Am. Chem. Soc.* **2008**, *130*, 3053–3064.

- [56] A. Salditt, L. M. R. Keil, D. P. Horning, C. B. Mast, G. F. Joyce, D. Braun, *Phys. Rev. Lett.* **2020**, *125*, 48104.
- [57] I. A. Chen, K. Salehi-Ashtiani, J. W. Szostak, *J. Am. Chem. Soc.* **2005**, *127*, 13213–13219.
- [58] K. Adamala, J. W. Szostak, *Science* **2013**, *342*, 1098–1100.
- [59] A. E. Engelhart, M. W. Powner, J. W. Szostak, *Nat. Chem.* **2013**, *5*, 390–394.
- [60] J. Sheng, L. Li, A. E. Engelhart, J. Gan, J. Wang, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3050–3055.
- [61] S. Bhowmik, R. Krishnamurthy, *Nat. Chem.* **2019**, *11*, 1009–1018.
- [62] A. Mariani, J. D. Sutherland, *Angew. Chem. Int. Ed.* **2017**, *56*, 6563–6566; *Angew. Chem.* **2017**, *129*, 6663–6666.
- [63] M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
- [64] Y. Y. Yan, J. H. Tan, Y. J. Lu, S. C. Yan, K. Y. Wong, D. Li, L. Q. Gu, Z. S. Huang, *Biochim. Biophys. Acta Gen. Subj.* **2013**, *1830*, 4935–4942.
- [65] R. Del Villar-Guerra, R. D. Gray, J. B. Chaires, *Curr. Protoc. Nucleic Acid Chem.* **2017**, *2017*, 17.8.1–17.8.16.
- [66] F. M. Gartner, I. R. Graf, E. Frey, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, DOI 10.1073/pnas.2116373119.
- [67] P. W. Kudella, A. V. Tkachenko, A. Salditt, S. Maslov, D. Braun, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, DOI 10.1073/pnas.2018830118.
- [68] S. Toyabe, D. Braun, *Phys. Rev. X* **2019**, *9*, 11056.

Manuscript received: July 29, 2022
Accepted manuscript online: September 19, 2022
Version of record online: October 11, 2022

Prebiotic Foam Environments to Oligomerize and Accumulate RNA

Emre Tekin,^[a] Annalena Salditt,^[a] Philipp Schwintek,^[a] Sreekar Wunnava,^[a] Juliette Langlais,^[a] James Saenz,^[b] Dora Tang,^[c] Petra Schwillie,^[d] Christof Mast,^[a] and Dieter Braun^{*[a]}

When water interacts with porous rocks, its wetting and surface tension properties create air bubbles in large number. To probe their relevance as a setting for the emergence of life, we microfluidically created foams that were stabilized with lipids. A persistent non-equilibrium setting was provided by a thermal gradient. The foam's large surface area triggers capillary flows and wet-dry reactions that accumulate, aggregate and oligomerize RNA, offering a compelling habitat for RNA-based early life as it offers both wet and dry conditions in direct neighborhood. Lipids were screened to stabilize the foams. The prebiotically more probable myristic acid stabilized foams over

many hours. The capillary flow created by the evaporation at the water-air interface provided an attractive force for molecule localization and selection for molecule size. For example, self-binding oligonucleotide sequences accumulated and formed micrometer-sized aggregates which were shuttled between gas bubbles. The wet-dry cycles at the foam bubble interfaces triggered a non-enzymatic RNA oligomerization from 2',3'-cyclic CMP and GMP which despite the small dry reaction volume was superior to the corresponding dry reaction. The found characteristics make heated foams an interesting, localized setting for early molecular evolution.

Introduction

The emergence of Darwinian evolution demands that already the earliest form of life must be capable of *replication* and *mutation* to then be *selected* by its environment in order to adapt and develop further.^[1] A candidate for life's first polymer is proposed by the RNA world hypothesis.^[2] RNA can store information and fold into catalytic structures, also known as ribozymes. A ribozyme replicator could therefore have been a suitable starting point for molecular evolution and could have predated the otherwise necessary simultaneous emergence of functional proteins and DNA as information storage.^[3]

For entropic reasons, the prebiotic synthesis of RNA polymers from monomers had to be driven by non-equilibrium boundary conditions, such as temperature, pH or salt gradients. On the early earth, for example, hydrothermal vents or springs, volcanic eruption sites or heated ponds would offer temperature gradients at various scales. For replication to keep sequence information despite various degradation channels, a local reduction of entropy would be a central pre-requisite for Darwinian processes,^[4] notably enhancing the concentration of the participating molecules against diffusion. One example of such a recently discussed scenario is the convection of water in pores of heated volcanic rocks.^[5] Such thermal gradients were a likely setting at early earths hot crust^[6] and were shown to drive polymerase-based DNA replication towards increasing length.^[7]

Various studies have been recently published on the role of thermal gradient systems for the emergence of life. For example, a pointed heat source led to highly efficient RNA-catalyzed amplification of short RNA,^[8] the circulation of lipid vesicles in a thermal gradient gave rise to fission by membrane-phase transitions^[9] or by cycling temperatures of a random pool of 12 nucleotide long DNA, selective oligomer structures emerged by templated ligation.^[10] Within this context, gas-water interfaces were discussed, as they drive a continuous enrichment of molecular structures under a temperature gradient.^[5] Morasch *et al.* focused on the effects of gas-water interfaces, that were created with the help of single gas-filled cavities surrounded by a liquid. Here, we introduced foam into the reaction compartments and could significantly increase the interfacial area. Also, the addition of prebiotic more probable surfactants created additional structures at the interfaces.


Foams are heterogeneous inclusions of gas in liquids stabilized by surface active reactants, i.e., surfactants. Their very high surface to volume ratio promotes phase boundary processes such as gelation, accumulation and phosphorylation.


[a] E. Tekin, A. Salditt, P. Schwintek, S. Wunnava, J. Langlais, Dr. C. Mast, Prof. Dr. D. Braun
Systems Biophysics
Center for Nano-Science and
Origins Cluster Initiative Department of Physics
Ludwig-Maximilians-Universität München
Amalienstrasse 54, 80799 München (Germany)
E-mail: dieter.braun@lmu.de

[b] Dr. J. Saenz
Center for Molecular Bioengineering
Technische Universität Dresden
Helmholtzstrasse 10, 01069 Dresden (Germany)

[c] Dr. D. Tang
Dynamic Protocellular Systems
Max-Planck Institute for Molecular Cell Biology and Genetics
Pfotenhauerstrasse 108, 01307 Dresden (Germany)

[d] Prof. Dr. P. Schwillie
Cellular and Molecular Biophysics
Max-Planck Institute of Biochemistry
Am Klopferspitz 18, 82152 Martinsried (Germany)

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202200423>

 © 2022 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Shallow hydro-thermal and volcanic settings may have led to ascending gas that passed through numerous porous rocks and formed micro scale foam environments. Alkaline hydrothermal vents are hypothesized to be a favorable environment for the emergence of the first protocell.^[11] Studies have shown that in hydro-thermal fields (e.g. in Mount Lassen, California), myristic acid added to warm spring water gave rise to the self-assembly of foams and membranous compartments.^[12]

We base this study on 2',3'-cyclic mononucleotides (cNMP) which (a) possess an intrinsically activated phosphate; (b) are products of several prebiotic phosphorylation and nucleotide syntheses^[13–17] and (c) are products of neutral to alkaline chemical and enzymatic hydrolyses of RNA.^[18–22] In comparison, the dry polymerization of 3',5'-cGMP^[23–25] did not foster the polymerization of the other ribonucleotides.^[25] Orgel and coworkers, reported conditions for 2',3' cAMP polymerization by drying for 40 days with a 5-fold excess of ethane 1,2 diamine and yields up to 0.67% of 14-mers.^[26,27] Other catalysts such as imidazole or urea required temperatures up to 85 °C and offered lower yields.

In a highly diluted, hadean ocean, around 4.4 billion years ago,^[28] environments that promote high concentrations were crucial for triggering polymerization and fast reactions kinetics. Moreover, the localization of reaction products is deemed central for evolution to proceed. In this work, we showed that foams, in comparison to environments with low surface to bulk ratio, could improve the yield of reactions by the presence of heated gas-water interfaces. A non-enzymatic polymerization reaction was examined in both environments and produced significantly higher yield for the polymerization of RNA.

Results and Discussion

Foam creation

In order to create reproducible and prebiotically more probable foam environments, we imitated micrometer sized pores and cracks in rocks with the help of shaped FEP foils clamped between two sapphires. These 150 µm–500 µm thick chambers can be filled with liquid and gas, temperature gradients can be applied, and fluorescent as well as bright light microscopy readout is possible. Figure 1 shows a schematic sketch of the cross-section of a surfactant stabilized bubble in a thermal gradient. Inside the gas bubble water evaporates at the warm side and condenses at the cold side, forming pure water droplets.

For foam fabrication by the co-flow method, a controlled supply of surfactant solution (e.g., 1.5 mM – 4 mM SDS or Tween20) and gas pressure was required. The surfactant solution flowed in from two opposite channels and periodically pinched off gas coming in from the mid channel (Figure 2a). The gas pressure and liquid flow could be varied from 180 mbar to 400 mbar and 10 µL s^{−1} to 40 µL s^{−1}, which allowed for control over the average bubble sizes. Similar to other publications^[29,30] that studied micro fluidic foams, we found $d \sim p$ and $d \sim q^{-1}$, with the bubble diameter d , gas pressure p and liquid flow q

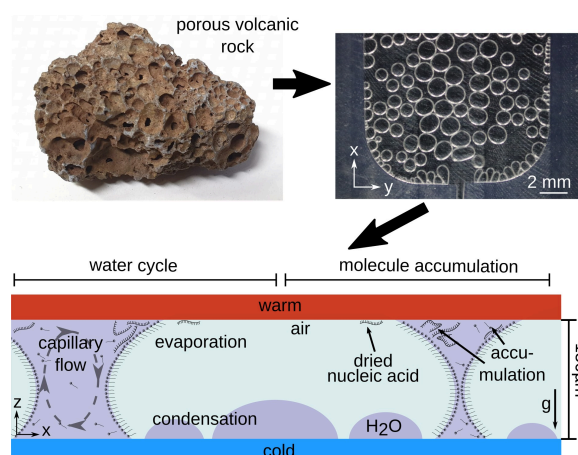


Figure 1. Foam in heated rock pores. Gas bubbles in heated porous volcanic rocks enable a prebiotic scenario of microfluidic foam systems, triggered by a thermal gradient. The scenario was recreated by foam filled 150 µm thick FEP cut-outs between heated sapphires. The surfactant solution of super-saturated sodium myristate created large foam structures as the surfactant adsorbed at gas-water interfaces and stabilized the gas bubbles. A temperature gradient triggered a continuous water cycling by evaporative capillary flow at the bubble interface of the foam that accumulated nucleotides at the warm side of the foam bubbles (grey arrow). The water vapor recondensed at the cold side and was shuttled by bubble aggregation driven by the surface tension back into the solution. This release of condensed water droplets into the bulk solution caused the gas bubble to fluctuate in size, transporting the dry accumulated nucleotides back into solution, establishing a localized, continuous wet-dry cycle similar to a heated air-water interface.^[5]

(Figure 2b). As pressure increased and flow decreased, the bubble sizes grew. Likely, due to its chemical structure and resulting packaging shape sodium myristate did not allow for precise control over the average bubble size, contrary to Tween20 or SDS.

We tested several lipids to create stable foams under the thermal gradient. In Figure 2c we plot the tested lipids' foam lifetimes over five orders of magnitude and plotted it against their corresponding hydrophilic lipophilic balance values. The lifetime was defined by the time half of the initial bubbles vanished in a 150 µm thick chamber while being exposed to a thermal gradient of $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 62 \text{ °C} - 25 \text{ °C} = 37 \text{ K}$.

For the ion induced stabilization of the foam in temperature gradients we used 200 mM dipotassium phosphate buffer. Phosphorus is a vital component of life, as it is essential for phosphorylation reactions for RNA and phospholipids. Carbonate-rich lakes can concentrate phosphorus up to more than 1 M.^[34] Due to the prebiotic plausibility and long lifetime a mixture of ~3.2 mM supersaturated sodium myristate and 200 mM dipotassium phosphate buffer was used for further experiments.

Fatty acids, such as myristic acid, can be synthesized under prebiotically more probable conditions^[31,32] and are therefore considered reasonable candidates for our foam experiments. We found that charged fatty acids are well suited for foam fabrication as their solubility is higher compared to their

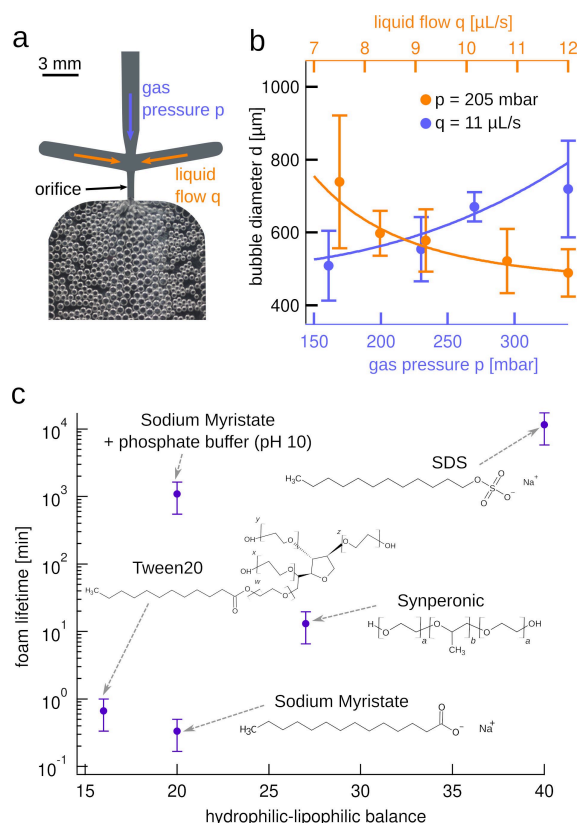


Figure 2. Foam formation and lifetime. a) Foam was generated in a microfluidic setup using the co-flow method. Gas pressure was applied to the central channel (purple arrow). The lateral inflow of the surfactant solution (orange arrows) allowed the pinch off of gas bubbles, so that the chamber was filled with foam within approximately 4 seconds. b) Tuning the liquid flow q and the gas pressure p allowed to control the average bubble diameter d . The shown foam was created using a 4.5 mM Tween20 solution. Data for a constant pressure of $p = 205$ mbar by varying flow and a constant flow of $q = 11$ $\mu\text{L s}^{-1}$ per channel by varying pressure is plotted. We found an approximate proportionality $d \sim p$ and $d \sim q^{-1}$. As a guide to the eye, power functions of the form $d(x) = k_0 \times x^{k_1} + k_2$ were displayed. The error bars represent the standard deviation from the average bubble diameter. c) The choice of surfactant influenced the lifetime of the corresponding foam in a thermal gradient of $\Delta T = 37$ K across the 150 μm thick chamber. Surfactants with a high hydrophilic-lipophilic balance tended to increase the life span of the foam. The addition of 200 mM dipotassium phosphate buffer to the supersaturated sodium myristate solution drastically increased the lifetime from several seconds to 20.5 h. The error bars were estimated to be 50 % (see Figure S10).

uncharged counterparts. In addition, the lifetime of sodium myristate increased significantly when the solution was enriched with ions. The likely reason is that the addition of ions to the surfactant solution shields the headgroup repulsion of the lipid, which enables better packing at the interface.^[33] Consequently, the critical packaging shape improves, and bubble formation is enhanced.

Accumulation of nucleotides

After having established a reproducible and stable way to form foams from prebiotically more probable surfactants, we investigated how the foams accumulate oligonucleotides. We prepared a solution of 3.2 mM sodium myristate, 200 mM dipotassium phosphate buffer and 10 μM Cy5 labeled, 14 nucleotide long ssDNA as a fluorescently marked proxy for polymerized RNA (see Supporting Information Table S1 for details). We have established previously, that DNA and RNA in capillary flow driven air-water systems behave similarly.^[10]

To compare the behavior of the DNA in a high surface area system (foam) and at a low surface areas system (single gas-water interface), a thermal gradient was applied to each system (Figure 3a, b). In both systems the same solution was injected, so that the only difference was the surface area. into the second chamber in the form of foam. The injection process of foam and sample is described in the Supporting Information (Figure S8). When the 150 μm thick chambers were exposed to a thermal gradient (here, $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 52$ C–38 C = 14 K), similar wet-dry dynamics and liquid fluxes occurred at the gas-water interfaces of both systems.

In both systems, evaporation at the warm side triggered capillary flows and accumulated molecules at the warm side of the water-gas interface, accumulating DNA.^[5] The gaseous water is recondensed at the cold side either into the bulk or in form of dew droplets either at the open interface or inside the air-filled bubbles of the foam. They are invisible in the fluorescence images of Figure 3 since they consist of pure water without fluorescence molecules. Close to the interface, the pure droplets went back to solution by surface tension and temporarily reduced local ion and molecule concentrations, which is analogous to “rainfall” in the macroscopic water cycle.^[35] The constant water cycle induced a fluctuating water level in the low surface area system, but in foam it led to fluctuation of the gas bubble diameters (Figure S11). In both cases, the fluctuating gas-water interfaces captured the accumulated molecules in continuous wet-dry cycles.

During the buildup of the thermal gradient an initial phase of Ostwald ripening^[36] was found and spontaneous degassing led to small bubbles of the foam disappearing in favor of larger bubbles (see Movie S2). When the buildup of the thermal gradient equilibrated, only much fewer bubbles collapsed and created larger gas bubbles. These effects caused slight movement of gas bubbles. Molecules that were captured in wet-dry cycles at the interfaces were torn out of the continuous cycle and left in the dry as the contact line between air and water receded away from them.

In Figure 3c, the accumulation and distribution of DNA is shown before and after the application of a thermal gradient of $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 49$ C–37 C = 12 K in z -direction. Due to the temperature dependence of the surface tension, the thermal gradient contracted the bubble at the warm side and expanded it at the cold side, which induced rather a truncated cone-like structure of the bubble. The top view through the sapphire gave a picture, in which two nested rings were observed, similar to Figure 3b. The bright inner ring of accumulated material was

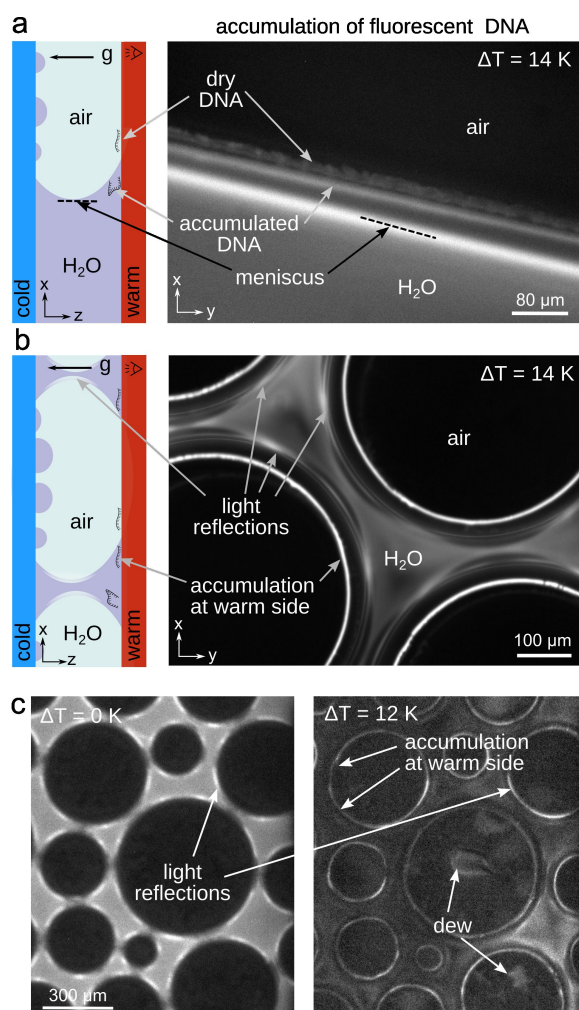


Figure 3. Capillary flow accumulates oligonucleotides at a gas-water interface and inside foams. a) Cross-section of a heated chamber filled with water and gas next to a fluorescent image of 14 nucleotide long FAM-labelled DNA at a single gas-water interface. DNA accumulated at the interface and eventually dried out. Oscillating water levels inside the foam bubbles redissolved the dried DNA and evoked a wet-dry cycle. Surface tension and temperature differences lead to a meniscus shaped interface, that produced two bright fluorescent lines with accumulated DNA. The sample was exposed to a temperature gradient of $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 52 \text{ }^{\circ}\text{C} - 38 \text{ }^{\circ}\text{C} = 14 \text{ K}$. b) Around the gas bubbles of the foam we found ring formed accumulations of DNA due to the capillary flow of evaporating water on the warm side. At the bubble surfaces light reflections appear, which are not to be confused with the direct fluorescence signal of concentrated DNA. c) Fluorescence image of foam before and after the application of a thermal gradient of $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 49 \text{ }^{\circ}\text{C} - 37 \text{ }^{\circ}\text{C} = 12 \text{ K}$. Water cycles inside the gas bubbles lead to luminous spots, which originated from growing dew droplets that spread from the cold to the warm side and thus rehydrating dry material inside the gas bubbles. Light reflexes can be distinguished from accumulation, as the brighter arcs appear also without the application of a thermal gradient. The experiments were performed in $150 \text{ }\mu\text{m}$ thick chambers with a solution of 200 mM dipotassium phosphate buffer, 3.2 mM supersaturated sodium myristate at pH 9.2 and $10 \text{ }\mu\text{M}$ Cy5 labeled DNA.

enclosed by an outer ring, which marked the air bubble boundary at the cold side.^[35] Light reflections in the right image

can easily be distinguished from high DNA concentrations, as they also occur for $\Delta T = 0 \text{ K}$ with $T = 24 \text{ }^{\circ}\text{C}$, where the equilibrium prevents an accumulation of DNA. The overall drop in fluorescence intensity was attributed to the temperature dependence of the dye.

For the emergence of life, the localization of DNA at each foam bubble effectively generated an inverted cell structure. The gas bubbles evoke several microenvironments of accumulated DNA at each of the bubbles. They offer slightly different local environmental conditions, based on the uneven distribution of DNA and the molecule transport between the foam bubbles is enhanced by the capillary flow and free diffusion, offering a complex spatial reaction setting, especially when the foam would be formed inside more complex rock pores. Inside the gas bubble, continuous evaporation and dew recondensation of water led to bright spots when dew encountered the warm side and thus redissolved previously dried fluorescent DNA there. Uneven arrangement of gas bubbles can lead to unevenly distributed concentrations in foam (Figure S12).

DNA-rich droplets accumulate at foam

Interestingly, highly concentrated structures appeared at the foam interfaces when the DNA was labelled with Cy5 (Figure 4a). To investigate these DNA-rich droplets further, a series of experiments with different DNA sequences and dyes were executed. We used a solution with $10 \text{ }\mu\text{M}$ fluorophore labelled DNA (see Table S1 for details), 3.2 mM supersaturated sodium myristate, 200 mM dipotassium phosphate buffer at pH 9.2. The chamber thickness was $150 \text{ }\mu\text{m}$ and the thermal gradient amounted $\Delta T = 12 \text{ K}$ with $T_{\text{warm}} = 49 \text{ }^{\circ}\text{C}$.

Cy5 labeled DNA showed liquid-liquid phase separated structures at the bubble interfaces. We observed this behavior for 14 nucleotide long and 84 nucleotides long Cy5 labelled DNA (Figure 4a, d). However, accumulation of 6-Fam labeled DNA at the gas bubble interface did not show a similar DNA-rich droplets (Figure 4b). Interestingly, an experiment with Cy5 dye without DNA, also gave rise to DNA-rich droplets at the gas bubble interfaces, indicating that the hydrophobic Cy5 dye as major driver of the effect, causing DNA-rich droplets in fatty acid solutions. Their resemblance to cellular structures could offer another level of complexity useful for localized prebiotic reactions.

We modified the DNA sample from Figure 4c by replacing Cy5 with 6-Fam and swapping the 6 bases next to the dye by poly G_6 . While the 6-Fam labeled DNA without integrated poly G_6 did show uniform accumulation at the interface, poly G_6 modified, 6-Fam labeled DNA gave rise to cumulated DNA structures. Thus, adding G-rich tails altered DNA-rich droplets at the interfaces. Inferred from the location in Figure 4, the GGGGGG-induced droplets are located in the bulk fluid of the interface. Given the fact that polymerization in the wet state is strongly reduced in its yield, we do not expect that this is the cause of the prominence of G in the polymerization product. Moreover, the polymerization of G is also enhanced in the dry state, indicating other mechanism for its preference, possibly

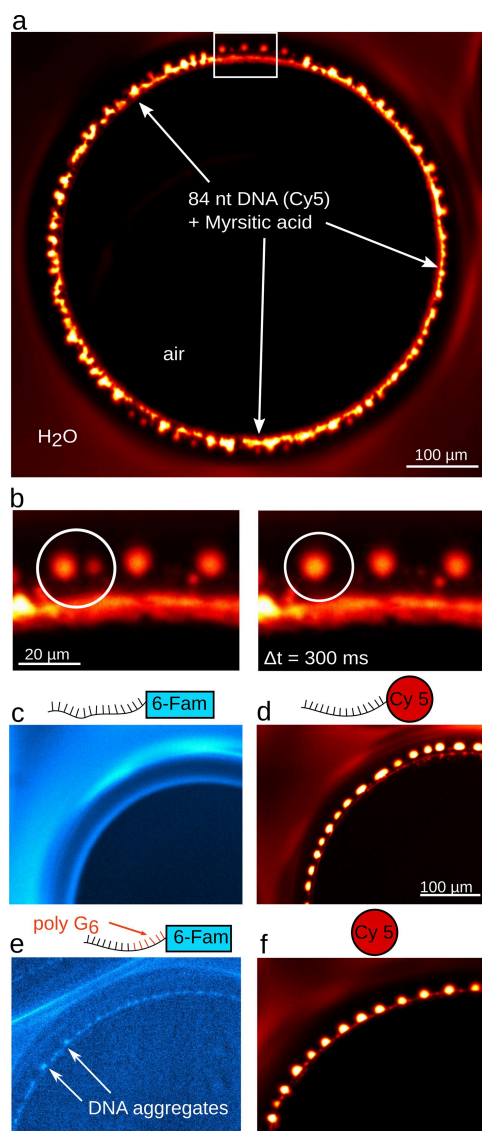


Figure 4. DNA-rich droplets form at heated foam interfaces. a) Cy5 labeled 84 nucleotide long DNA showed DNA-rich droplets at the accumulating, warm interface of gas bubbles (white arrows). b) Excerpt from white box (a), 300 ms apart. The DNA-rich droplets moved in the liquid phase at the warm interface, showing a fusion dynamic typical to oil droplets or coacervates. c, d) 6-Fam labeled 15 nucleotide long DNA did not show DNA-rich droplets while Cy5 labelled DNA of 14 nucleotides showed round structures, more homogeneous than the 84 nucleotide long DNA in (a). e) Replacing six bases of the 15 nucleotide long 6-Fam sequence by GGGGGG induced dot-like structures, indicating that the G-rich sequences also trigger the effect, confirming lipid interactions of G.^[37] f) Only Cy5 dye as a sample showed uniform DNA-rich droplets at the interface as well. All experiments applied a thermal gradient of $\Delta T = T_{\text{warm}} - T_{\text{cold}} = 49 \text{ }^{\circ}\text{C} - 37 \text{ }^{\circ}\text{C} = 12 \text{ K}$ in 150 μm thick vessels. The sample consisted of 10 μM fluorescent DNA or dye (f), 200 mM dipotassium phosphate buffer and 3.2 mM sodium myristate.

due to the pK_a of the base or its enhanced stacking interaction.^[38]

The observed DNA-rich droplets were either stuck at the warm interface or moved freely close to the interface (see Movie S3). The freely moving DNA-rich droplets are recognizable by their isolated circular form. The structures touched and fused (Figure 4f), which is typical for coacervates studied in a thermal gradient before^[39] or oil droplets.^[40] We would characterize the droplet structures rather as coacervates than vesicles as we measured them to be 3 μm –7 μm in size and larger than vesicles, which are up to 100 nm in size.^[41] Surfactants typically form micelles rather than vesicles.^[33]

Accumulation of sequence dependent DNA aggregates

Sequence-dependent large scale DNA aggregates can be obtained from self-binding of ssDNA.^[3,5,42–46] With the same experimental parameters as in Figure 4, a 36 nucleotide long 6-Fam labeled DNA sequence of G and C was designed to self-assemble into large DNA networks (Figure 5a). After the application of a thermal gradient ($\Delta T = 12 \text{ K}$ with $T_{\text{warm}} = 49 \text{ }^{\circ}\text{C}$) the accumulation of DNA aggregates became visible within minutes (Figure 5b) and grew to larger, ring-like assemblies. Sequences which were not self-binding did not create comparably large aggregates as seen in Figure 3 and 4).

This sequence dependent accumulation by gel formation^[3,25] gives RNA sequences a direct physical selection pressure: (i) the gel formation enhances the accumulation by the reduction in diffusion coefficient, making them more resistant to flow-based dilution away. (ii) The gelation enhances the formation of double stranded versus single stranded forms. This reduces the degradation of RNA by hydrolysis, leading to a sequence-dependent, multi-molecular selection pressure, adding a direct phenotype-genotype relationship for early Darwinian evolution.

The fluctuation of the gas bubble sizes caused sudden contractions of the gas-water interfaces, the aggregates could detach from one foam bubble and were transported by the capillary flows to another gas bubble (Figure 5c), demonstrating the sequence-dependent oligonucleotide transport in the foam. Compared to a single gas-water interface, foam offers more space for the formation of DNA aggregates^[5] and the exchange between bubbles offers complex exchange modes for self-similar genetic material.

GC polymerization

Foams are particularly interesting for reactions relying on wet-dry cycles and as seen, already short oligonucleotides show very diverse selective phenotypes. To approach prebiotic scenarios, we therefore explored the effectiveness of the wet-dry cycles for the polymerization of short RNA from prebiotic more probable 2',3'-cyclic nucleotides. Verlander et al. observed polyA oligomers up to a length of six nucleotides at alkaline pH conditions and temperatures from 20 $^{\circ}\text{C}$ to 70 $^{\circ}\text{C}$ with the help of amine catalysts.^[27] Under such conditions a transesterification^[47] of the phosphate group can occur, which results in the conversion of the 2'3'-cyclic phosphate group to a

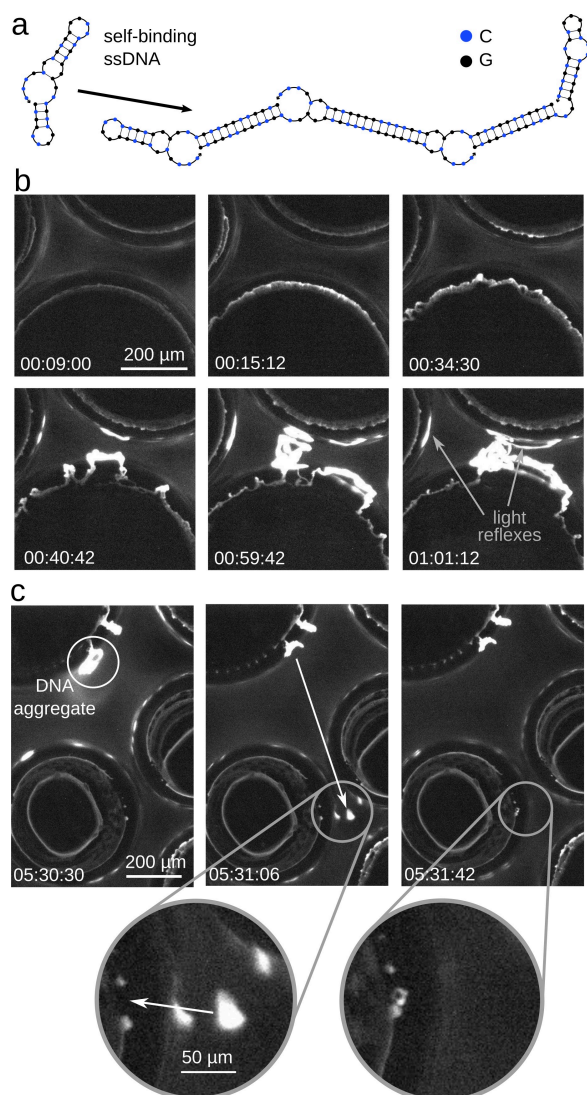


Figure 5. Aggregates of self-binding ssDNA form at the foam interfaces. a) Self-binding 36 nucleotide long ssDNA of G and C is designed to self-assemble into complex aggregates. b) Driven by a thermal gradient of $\Delta T = 12$ K with $T_{\text{warm}} = 49$ °C, 10 μM 6-Fam labeled ssDNA is forming aggregates that grew over several minutes, gaining size through fluctuations of the gas-water interface (Movie S4). c) Grown aggregates were detached from the interface and were transported through the bulk solution by diffusion and capillary flows. The circular excerpts show magnified images of the DNA aggregate docked at another bubble. The trap thickness was 150 μm . The solution included 3.2 mM sodium myristate and 200 mM dipotassium phosphate buffer.

3',5' or 2',5' linkage. In recent study, 2',3' cyclic nucleotides were shown to self-polymerize without catalysts under low salt conditions and elevated pH in the dry state. In case of mixed oligonucleotides polyG leads the copolymerization.^[38] Therefore, the wet-dry cycling inherent to the heated foam was tested.

We compared the polymerization from cGMP and cCMP in foam to polymerization in a system with significantly lower

interfacial area. We could keep the conditions like before, using a solution of 3.6 mM sodium myristate, 200 mM phosphate buffer, 25 mM cGMP and 25 mM cCMP each and KOH to adapt the pH to pH 9.2. After 19.5 hours the pH of the extracted sample dropped to 8.6.

In Figure 6a, the resulting polymers were quantified by HPLC-ESI-TOF mass spectrometry measurements. The high surface to bulk ratio in the foams lead to higher concentrations of polymerization products compared to the non-foam counterpart. The reaction in foam has progressed further (up to 8mers) than in non-foam (up to 5mers). The chamber thickness and

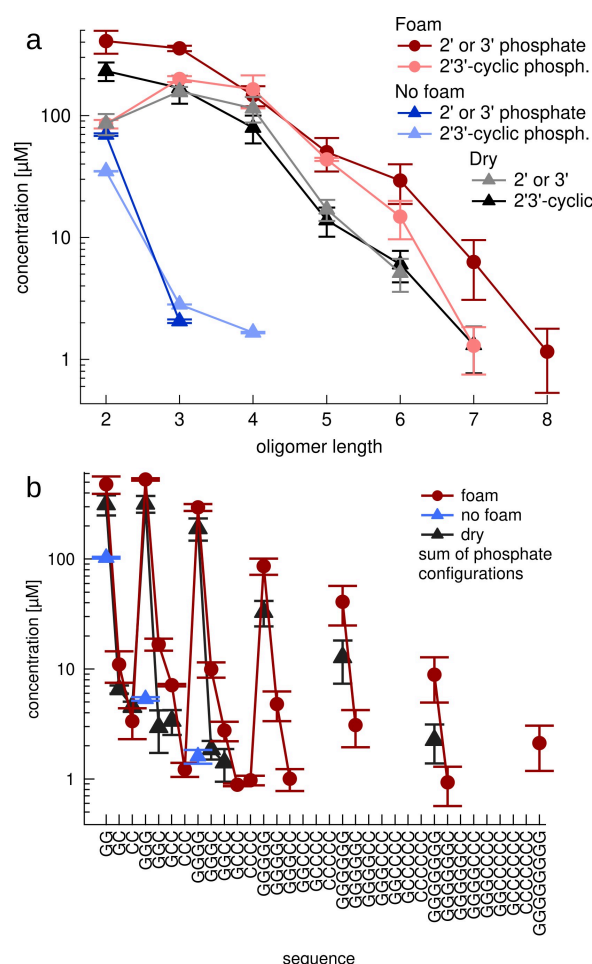


Figure 6. Foam induced polymerization of 2',3'-cGMP and 2',3'-cCMP. a) The wet-dry cycling conditions at the moving interfaces led to the oligomerization of RNA. The heated foam setting provided wet-dry cycles for polymerization (red), while a control experiments without bubbles did not oligomerize significantly (blue). The samples were exposed to a thermal gradient of $\Delta T = 25$ K with $T_{\text{warm}} = 59$ °C for $t = 19.5$ hours. When the whole solution was dried to T_{warm} the oligomerization yield was significantly lower (grey) despite subjecting all nucleotides to the reactive dry condition. Oligomers with linear phosphate ends were found, indicating significant hydrolysis towards the unreactive, open ended monophosphates. b) The sequences detected by mass spectrometry showed sequences that integrated the less reactive C base, but also confirmed that longer oligonucleotides were dominated by polyG.

temperatures in the polymerization experiments were 500 μm , $T_{\text{cold}} = 34\text{ }^{\circ}\text{C}$ and $T_{\text{warm}} = 59\text{ }^{\circ}\text{C}$. In b the sequence space of the same experiment can be seen. The sequences that occur in both experiments are much higher concentrated in the foam experiment. Without foam, the maximum length of the resulting polymers was G_4 , while polymerization in foam gave rise to G_8 at a higher concentration.

The foam brings together both the dry and wet state, offering increased flexibility for prebiotic reactions. It has been established previously that the oligomerization is driven in the dry state.^[23–25,38] In the foam setting, only a small fraction of the molecules are in the dry state at the air-water interface on the warm side. One could therefore expect that the oligomerization in the foam should be less effective than when the whole solution would be dried. However the opposite is true. As reported in Figure 6 with the grey data points, the oligomerization in the fully dry condition yields significantly less than for the reaction in the heated foam. At this point, we can only speculate that this might be due to the length dependent accumulation at the interface.^[7,38,39]

It is important to note that the shown oligomerization is inefficient in comparison to polymerization reactions known in chemistry. But if prebiotic chemistry would have provided very efficient polymerization for RNA, any competing replicative mechanism would have been overwhelmed by a large amount of non-replicated, polymerized RNA, making Darwinian evolution a difficult past. However our results can be compared to previously reported results by Verlander and Orgel.^[26,27] There, homo-oligomers of cAMP in the presence of 5-fold excess of ethane-1,2-diamine showed a yield of 0.81% for a 6mer of polyA after 40 days. In comparison, we observed here ~0.12% for a 6mer of oligoG without added catalysts after less than a day in a mixed GC setting. Adding other catalysts such as imidazole or urea required temperatures up to 85 $^{\circ}\text{C}$ and offered lower yields.^[26,27]

Clearly, the rather short mixed sequence oligomers reported are quite a bit away from the short oligomers required to for example run a templated ligation chain reaction. We hope that we will be able to better understand the degradation pathways of 2'3'-cyclic monomers with future experiments. But already at this stage, if an RNA world would operate with trimers, recently shown with a harder to reach triphosphate activation,^[48] the documented oligomerization could be a good source of raw materials for replication while at the same time offering strand separation functionality by the air-water interface.^[7]

Conclusion

For prebiotic reactions, heated liquid-gas interfaces offer dry and wet cycling conditions for the molecules that were attracted by the ensuing capillary flows. Here we maximized the surface to volume ratio by creating foams using prebiotic more probable lipids for stabilization. The result was an array of molecule-accumulating habitats for DNA or RNA that form an inverted cellular geometry.

Instead of insulating molecules by a bilayer, the molecules are concentrated at each bubble of the foam and are free to transfer between individual bubble habitats. At these interfaces, we found that the accumulation by the capillary flows triggered the formation of DNA-rich droplets and the sequence-dependent formation of large scale oligonucleotide aggregates, extending experiments at single air-water interfaces.^[5,39] The open setting also allows for thermophoretic accumulation effects, such as the accumulation of ions to boost RNA polymerization catalysis.^[49]

Since both accumulation and the formation of agglomerates and coacervation-like structures was triggered by oligonucleotides, it was intriguing to find that the heated foam was also capable to produce them by polymerizing RNA up to 8mers directly from the prebiotically more probable 2'3'-cyclic GMP and CMP. These combined findings show an interesting convergence of physical settings and chemical synthesis, establishing a new paradigm of compartmentalization for the selection, accumulation and oligomerization of informational polymers for early prebiotic evolution.

Experimental Section

Sample preparation: The reaction mixture was prepared by mixing 720 μL of the 5 mM supersaturated sodium myristate solution with 100 μL of a 2 M K_2HPO_4 solution, 50 μL of a 500 mM 2'3'-cGMP solution, 500 mM 2'3'-cCMP solution and 75 μL of RNase-free water. The pH was adjusted at 9.2 by adding 5 μL of 1 M KOH solution, resulting in 3.6 mM sodium myristate, 200 mM dipotassium phosphate, 25 mM 2'3'-cGMP and 25 mM 2'3'-cCMP.

Precipitation protocol: To provide best reproducibility, a precipitation protocol that was shown not to change oligomer concentrations down to 2nt significantly was developed. To an aqueous solution of the oligonucleotide sample 2 μL of 100 mg/mL of glycogen and 1/10th volume 5 M ammonium acetate was added. After mixing, 2–3 volumes of cold 100% ethanol was added. The final solution was incubated at 4 $^{\circ}\text{C}$ overnight. After incubation, the sample was centrifuged at 4 $^{\circ}\text{C}$ for 30 min at 15000 rpm. The remaining supernatant was carefully discarded and 10 μL of cold 70% ethanol was added. The sample was centrifuged again at 4 $^{\circ}\text{C}$ for 30 minutes at 15000 rpm. The supernatant was removed, and the dried pellet dissolved in 40 μL RNase-free water for HPLC analysis.

HPLC and mass spectrometry: The measurements were performed on Agilent G6230BA LC/TOF with G5654 A 1260 InfinityII bioinert HPLC using an Agilent AdvanceBio Oligonucleotide (4.6 \times 150 mm, 2.7 μm) column with a pressure rating of 600 bar, heated to 60 $^{\circ}\text{C}$. Oligomers were analyzed in a water-methanol solvent system with solvent A consisting of 200 mM HFIP and 8 mM TEA in water and solvent B a 50:50 methanol-water mixed with solvent A. After an initial flow of 1% B for first 5 minutes, followed by a gradient of 1% up to 30% B in 22.5 minutes and a gradient from 30% to 40% B in the next 15 minutes at 1 mL/minute. The TOF Mass spectrometer Agilent 6125 is equipped with a dual AJS ESI ion source, set to a gas temperature of 325 $^{\circ}\text{C}$, gas flow at 8 l/min, nebulizer at 45 psig, sheath gas temperature at 400 $^{\circ}\text{C}$ and sheath gas flow at 11 l/min. and run in negative mode. The samples were analyzed in

the negative ion mode. The scan source parameters were negative mode, Vcap=3000 V, nozzle voltage at 2000 V, fragmenter at 175 V, Skimmer at 65 and octapoleRFpeak value of 750.

Fitting TOF-MS data with LabView program: Raw spectra of the mass spectrometry was converted with MSConvert to a MS1 or MS2 format. The timing of the oligomers on the HPLC separation was determined with standards and to also determine the ionization efficiency. The raw mass spectra between each time bracket were integrated before fitting. Species of the polymers were defined and the isotope distribution calculated and used to fit the raw spectra before species of different charge are integrated and were converted to concentrations using the calibration from oligomer standards.

Thermal chamber and imaging: The thermal chamber was cooled by a water bath from the back and heated using 3D printer heating rods on the front side. The chamber was defined by Sapphire windows in between a the chamber geometry was defined by Teflon foil that was cut with a plotter (GRAPHTEC CE6000-40 Plus). An Arduino Mega 2560 was used to control temperature. The aluminum blocks and the backside sapphire provided openings for tubes or temperature sensors (Greisinger, GTH 1170). A heat camera (Seek Thermal, SQ-AAA) allowed temperature measurement on the front sapphire. The experiments were either monitored by a Canon90D camera with a 5× magnifying lens or a fluorescence microscope. Fluorescence imaging for Cy5 and 6-Fam was provided concurrently by an OptoSplit under LED illumination. Long distance 2×, 5× or 10× objectives were used in reflected light and imaged with a Stingray F145-B camera through a self-built fluorescence imaging setup using Thorlabs lens tube components.

Acknowledgements

This research was supported by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094-390783311; the European Research Council (ERC Evotrap, grant no. 787356 (D.B.)), the Simons Foundation (grant no. 327125 (D.B.)), the VW foundation – "Life" initiative (grant nos. 92857 (T.-Y.D.T.) and 94743 (C.B.M., D.B.)), the CRC 235 funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 364653263 – TRR 235. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: fatty acids · foam · nucleic acid structures · origin of life · polymerization

- [1] S. A. Benner, *Astrobiology* **2010**, *10*, 1021–1030.
- [2] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
- [3] A. Kühnlein, S. A. Lanzmich, D. Braun, *eLife* **2021**, *10*, 10.7554/eLife.63431.
- [4] D. Braun, A. Libchaber, *Phys. Biol.* **2004**, *1*, P1.
- [5] M. Morasch, J. Liu, C. F. Dirscherl, A. Ianeselli, A. Kühnlein, K. Le Vay, P. Schwintek, S. Islam, M. K. Corpinot, B. Scheu, D. B. Dingwell, P. Schwiller, H. Mutschler, M. W. Powner, C. B. Mast, D. Braun, *Nat. Chem.* **2019**, *11*, 779–788.
- [6] N. T. Arndt, E. G. Nisbet, *Annu. Rev. Earth Planet. Sci.* **2012**, *40*, 521–549.
- [7] A. Ianeselli, M. Atienza, P. W. Kudella, U. Gerland, C. B. Mast, D. Braun, *Nat. Phys.* **2022**, *18*, 579–585.
- [8] A. Salditt, L. M. R. Keil, D. P. Horning, C. B. Mast, G. F. Joyce, D. Braun, *Phys. Rev. Lett.* **2020**, *125*, 048104.
- [9] P. W. Kudella, J. Preißinger, M. Morasch, C. F. Dirscherl, D. Braun, A. Wixforth, C. Westerhausen, *Sci. Rep.* **2019**, *9*, 1–11.
- [10] P. W. Kudella, A. V. Tkachenko, A. Salditt, S. Maslov, D. Braun, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018830118.
- [11] N. Lane, W. F. Martin, *Cell* **2012**, *151*, 1406–1416.
- [12] D. Deamer, *Life* **2021**, *11*, 134.
- [13] E. I. Jiménez, C. Gibard, R. Krishnamurthy, *Angew. Chem. Int. Ed.* **2021**, *60*, 10775–10783.
- [14] Z. Liu, L. F. Wu, J. Xu, C. Bonfio, D. A. Russell, J. D. Sutherland, *Nat. Chem.* **2020**, *12*, 3–10.
- [15] H. J. Kim, S. A. Benner, *Astrobiology* **2021**, *21*, 298–306.
- [16] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.
- [17] Y. Yamagata, H. Inoue, K. Inomata, *Origins Life Evol. Biospheres* **1995**, *25*, 47–52.
- [18] R. Breslow, *Acc. Chem. Res.* **1991**, *24*, 317–324.
- [19] Y. Li, R. R. Breaker, *J. Am. Chem. Soc.* **1999**, *121*, 5364–5372.
- [20] H. Peng, B. Latifi, S. Müller, A. Lupták, I. A. Chen, *RSC Chem. Biol.* **2021**, *2*, 1370–1383.
- [21] A. M. Pyle, *Science* **1993**, *261*, 709–714.
- [22] S. I. Nakano, D. M. Chadalavada, P. C. Bevilacqua, *Science* **2000**, *287*, 1493–1497.
- [23] M. Morasch, C. B. Mast, J. K. Langer, P. Schilcher, D. Braun, *ChemBioChem* **2014**, *15*, 879–883.
- [24] J. E. Šponer, J. Šponer, A. Giorgi, E. Di Mauro, S. Pino, G. Costanzo, *J. Phys. Chem. B* **2015**, *119*, 2979–2989.
- [25] S. Wunna, C. F. Dirscherl, J. Výravský, A. Kovářik, R. Matyášek, J. Šponer, D. Braun, J. E. Šponer, *Chem. Eur. J.* **2021**, *27*, 17581–17585.
- [26] M. S. Verlander, L. E. Orgel, *J. Mol. Evol.* **1974**, *3*, 115–120.
- [27] M. S. Verlander, R. Lohrmann, L. E. Orgel, *J. Mol. Evol.* **1973**, *2*, 303–316.
- [28] S. A. Wilde, J. W. Valley, W. H. Peck, C. M. Graham, *Nature* **2001**, *409*, 175–178.
- [29] A. Huerre, V. Miralles, M. C. Jullien, *Soft Matter* **2014**, *10*, 6888–6902.
- [30] P. Marmottant, J. P. Raven, *Soft Matter* **2009**, *5*, 3385–3388.
- [31] D. W. Deamer, J. Oro, *BioSystems* **1980**, *12*, 167–175.
- [32] T. M. Mccollom, G. Ritter, B. R. T. Simoneit, *Origins Life Evol. Biospheres* **1999**, *29*, 153–166.
- [33] J. N. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed., Elsevier, **2011**.
- [34] J. D. Toner, D. C. Catling, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 883–888.
- [35] A. Ianeselli, C. B. Mast, D. Braun, *Angew. Chem.* **2019**, *131*, 13289–13294; *Angew. Chem. Int. Ed.* **2019**, *58*, 13155–13160.
- [36] J. Schmelzer, F. Schweitzer, **1987**, *12*, 255–270.
- [37] T. Czerniak, J. P. Saenz, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2119235119.
- [38] A. V. Dass, S. Wunna, J. Langlais, B. von der Esch, M. Krusche, L. Ufer, N. Chrisam, R. C. A. Dubini, F. Gartner, S. Angerpointner, C. F. Dirscherl, P. Rovó, C. B. Mast, J. Šponer, C. Ochsenfeld, E. Frey, D. Braun *ChemSystemsChem* **2022**, *4*, e202200026.
- [39] A. Ianeselli, D. Tetiker, J. Stein, A. Kühnlein, C. B. Mast, D. Braun, T.-Y. Dora Tang, *Nat. Chem.* **2021**, *1*–8.
- [40] T. Banno, S. Miura, R. Kuroha, T. Toyota, *Langmuir* **2013**, *29*, 7689–7696.
- [41] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, R. Nieuwland, *Pharmacol. Rev.* **2012**, *64*, 676–705.
- [42] M. Morasch, D. Braun, C. B. Mast, *Angew. Chem.* **2016**, *128*, 6788–6791; *Angew. Chem. Int. Ed.* **2016**, *55*, 6676–6679.

- [43] O. A. Saleh, B. J. Jeon, T. Liedl, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16160–16166.
- [44] B. J. Jeon, D. T. Nguyen, O. A. Saleh, *J. Phys. Chem. B* **2020**, *124*, 8888–8895.
- [45] D. T. Nguyen, O. A. Saleh, *Soft Matter* **2017**, *13*, 5421–5427.
- [46] D. T. Nguyen, B. J. Jeon, G. R. Abraham, O. A. Saleh, *Langmuir* **2019**, *35*, 14849–14854.
- [47] J. Otera, *Chem. Rev.* **2002**, *93*, 1449–1470.
- [48] J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, *eLife* **2018**, *7*.
- [49] T. Matreux, K. LeVay, A. Schmid, P. Aikkila, L. Belohlavek, A. Z. Çalışkanoğlu, E. Salibi, A. Kühnlein, C. Springsklee, B. Scheu, D. B.

Dingwell, D. Braun, H. Mutschler, C. B. Mast, *Nat. Chem.* **2021**, *13*, 1038–1045.

Manuscript received: July 26, 2022
Revised manuscript received: October 24, 2022
Accepted manuscript online: November 10, 2022
Version of record online: November 18, 2022

Acknowledgement

First and foremost, I would like to thank Dieter, for giving me the opportunity to do my PhD in your lab, and for your patience and understanding. You let me explore my ideas for the projects and you were there when I needed your help. It was a great experience of five years. Christof, the help and the support you provide to everyone in the lab is priceless. You always came up with great suggestions whenever I struggled and offered an understanding ear. That's the Christof effect: everything works better when you are there, experiments, instruments and people.

To the Braunies throughout the years, you are an amazing group. My doctorate would have been an very different experience had I ended up in another lab. You were always there, helping and cheering me on. And for that, thank you ! Sree, for facing with me the great terror of the mass spectrometer, with your resourcefulness and intuition. For listening to me calmly each time I needed to ramble. And the ginger beer ! Philipp, for helping me to get out of my shell and start exploring life in Munich outside of work. Thank you for having been there, especially during the lock down. Adriana, for the great discussions about science and about remaking the world! I wish you all the best for the one year summer plan, your determination is an inspiration for me. Martina, for your strength and your kindness and for supporting me when I was at my lowest point. Annalena, for introducing me to yoga! It saved my back and my sleep, and my mood during the writing of the thesis. Max, for being my teammate during five years of supervising the Praktikum and having been such a great help in everything: informatics, German and the writing of the thesis. Almuth, for the German coffee breaks, which helped me to improve my few skills in German and the help in organizing the Movie nights. Thomas, for the discussion in French sometimes late in the evening and helping me to feel at ease in the lab when I came first. Christina, as the original member of the MS team, and for coming to visit us at the occasion of the movie nights. Alex, I loved talking about space with you. Keeping looking at the stars! Félix, for your energy and enthusiasm, whether at work or in the mountains. Paula, for the more relaxing and enjoyable times in the Mensa, your parties and the vacation days in Spain. Saroj, for being such a great colleague and cook, and treating us with delicious Indian food. Noël, I will never forget your dance moves and your enthusiasm going forward in life. Alan, Alexandra, Patrick, Julian, my senior PhD colleagues, for guiding me in the beginning of my work in the lab and for always being available when I had questions. And I address my best wishes to my newer colleagues who just started, Riccardo and Zsófia have strength and courage! Bobby, for your kindness, your enthusiasm and the discussion in the coffee corner (including on the weekend). Katrina, for taking over the huge task of taking care of the organization of the Sommerfeld Keller. Emma, for your calm efficiency. Maer, for the sweets and the movie nights. My students – Joachim, Lennard, Maik, Patrick, Niko, Francesco – for your curiosity, your enthusiasm and your hard work: I believe you will be successful regardless of the path you choose to follow.

My collaborators, Iris, Avinash, Francesco, Federico, Tommaso F. and Tommaso B., I wish you all the best for the next steps.

My Master thesis supervisors, Ludovic and François, thank you for helping me to find my

5. Acknowledgement

PhD position.

My friends, Laure, Mickaël, Thibault, Manon, Clément, Jean-Vincent, Lucile, Corentin, Léon and Célia thank you for your loving support and your trust, for being there for me in vacation and on the phone, and especially when it is difficult. Mathieu, Rebecca, Gabriel, Romain, thank you for your support and for coming to see me in Munich. Vanamonde and Elenthya, our online sessions kept going during my whole thesis and even before. It would have been so much more difficult without you.

My parents, my brothers, my family, thank you for always believing in me, and making who I am.