Synthesis and Characterization of Informational Molecules of Early Earth

A thesis

Submitted in partial fulfillment of the requirements

Of the degree of

Doctor of Philosophy

By

Chaitanya Vinayak Mungi

20112005

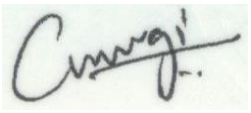
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH

PUNE

2018

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have not been properly cited or from whom proper permission has not been taken when needed.



Chaitanya Vinayak Mungi

Roll No. 20112005

Date: 19th July, 2018

CERTIFICATE

Certified that the work incorporated in the thesis entitled "**Synthesis and Characterization of Informational Molecules of Early Earth**", Submitted by Mr. Chaitanya Vinayak Mungi was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Dr. Sudha Rajamani

Thesis Supervisor

Date:19th July, 2018

ACKNOWLEDGEMENTS

Being a part of IISER Pune has been an exhilarating and transformative experience. Since joining IISER in 2011, I have met a lot of wonderful people and I would like to take this opportunity to thank them. I would like to express my profound gratitude to my thesis advisor Dr. Sudha Rajamani for giving me the opportunity to work with her. Her exceptional mentorship and joyful personality made the work presented here an enriching experience. She always gave me the freedom to explore new ideas and encouraged me to become an independent researcher for which I am really grateful. The writing and communication skills and the scientific approach I learned under her guidance is invaluable.

I would like to acknowledge my research advisory committee members Dr. Sayam Sengupta, Dr. Jeet Kalia and Dr. Thomas Pucadyil for the motivation and the recommendations throughout the project. I sincerely thank our collaborator Dr. Jeetender Chugh and his lab members for their contributions in NMR experiments. I also want to appreciate Dr. Yayoi Hongo for collaborating with us and carrying out MS analyses of some of the samples. The combined expertise of the aforementioned researchers has helped me tremendously in shaping this interdisciplinary project.

I would like to thank all my present COoL lab members and those who have worked with us in the past. I really appreciate Niraja, Anupam, Manesh, Shikha and Susovan for being a fun group to work with. A special thanks to Niraja for bearing with me for all these years and Anupam for his insightful suggestions and help with the synthesis aspects. I will genuinely miss the stimulating scientific discussions we had during lab meetings and journal clubs. I wish all of you the very best for your future scientific adventures.

My sincere thanks to the whole Biology department at IISER Pune for all the facilities. I would like to acknowledge the administrative, technical and non-technical staff, especially Mrinalini, Piyush, Kalpesh, Shabnam and Mahesh Rote for their support and Vijay for the help with microscopy facility. I really appreciate the hard work by past and current chair and co-chair of the department in making sure that we students get all the necessary help. I would like to thank Dr. Nagaraj Balasubramanian and Dr. Saikrishnan Kayarat for the opportunity and the training provided by them as mentors for my previous lab rotations. A special gratitude to Dr. L. S. Shashidhara for his continued support and encouragement during our tenure as iPhDs.

I thank the Chemistry department at IISER Pune for the opportunity to conduct experiments in various labs and allowing us to use their multiple facilities. Thanks Dr. Srivivas Hotha for help with the initial set up of the simulator. I also want to acknowledge Dr. Ajeet Singh from NCL Innovation Park for his help with MS analysis of some samples.

My sincere gratitude to Dr. H. James Cleaves, Dr. Ramnarayan Krishnamurthy, Dr. Bruce Damer, Dr. Dave Deamer and Dr. Kuhan Chandru for the discussions and expert suggestions on various occasions. The discussions with them immensely helped with filling up the intellectual void due to lack of an Origins of Life research community in India. Thanks to your openness, and your generosity with sharing ideas.

None of this would be possible without the never ending support from my entire family. I am thankful to my parents for their understanding and adjusting nature which has allowed to me strive for my goals without any worries. I thank my sister and her family for their encouragement and a special thanks to my nephew for bringing the joy of being a kid back in to my life. I would like to thank all of the innumerable friends I made during my time at IISER Pune. I thank Ketakee, Sayali, Roopali, Ajay and Shubhankar for being the most wonderful batch mates. I additionally wish to thank all my iPhD juniors, Drama Club friends, Manish, Sachin, Raunaq, Devika, Manasi, Sampada, Mahendra and Shraddha for their contributions in my life at a personal and professional level. I would like to thank all members of Spaceward Bound India for the wonderful expedition to Ladakh and an experience of a lifetime.

I greatly acknowledge IISER Pune for the institutional and financial support throughout the program, DST SERB grant for the funding for the research project and CSIR for my fellowship. Additionally, I thank Earth Origins network and staff for the long term visitor grant and the opportunity to work at ELSI, Japan.

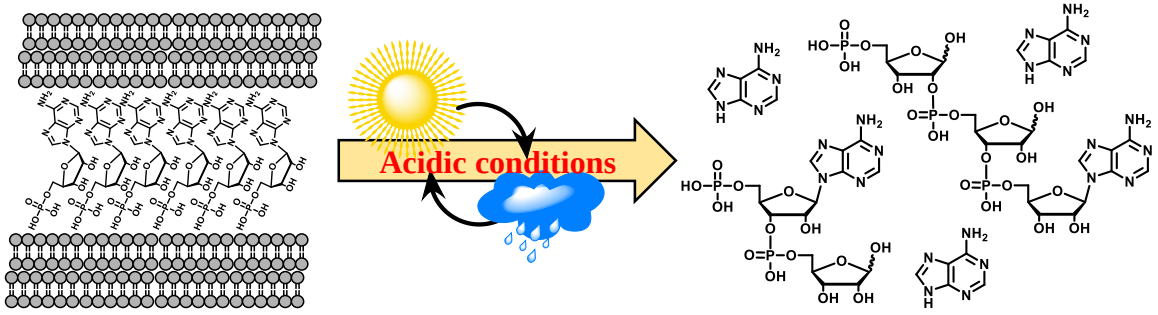
SYNOPSIS

Origin of life on Earth is one of the greatest scientific enigmas. A crucial step in this process involves the formation of informational molecules that would have been fundamental to the transition of chemistry to biology. Amongst the modern polymers, RNA is most pertinent as the "first" biopolymer due to its dual ability to store information, as well as carry out catalysis (in the form of ribozymes). Given this, the RNA World hypothesis has been one of the most advocated pathways for the origin of the first genetic material. Prior to RNA's emergence, polymerization had to be a nonenzymatic, chemically driven process. Studies in this context were historically focused on demonstrating oligomerization of various types of chemically activated nucleotides. However, the presence of activated nucleotides on prebiotic Earth, in large concentrations, is debatable. On the contrary, oligomerization of non-activated nucleotides is more prebiotically plausible and pertinent. The formation of RNA by nonenzymatic polymerization of non-activated has not been studied extensively. Few studies have looked at oligomerization of nucleoside 5'-monophosphates, in the presence of lipids, by performing dry-wet cycles.

Fluctuating environments, like dry-wet cycles, are thought to have promoted uphill reactions by facilitating formation of kinetic traps. Dehydration (DH) has shown to result in higher order structure formation such as multilamellar matrices from lipid vesicles. The aforementioned study hypothesized that the nucleotides get organized in a 2D array like arrangement within these multilamellar structures, resulting in concentration of the monomers, and thus favoring phosphodiester bond formation by condensation reaction. Subsequent rehydration (RH) results in vesicles that can encapsulate the newly formed oligomers. In the aforementioned manner, amphiphiles are thought to have played a role in, both, oligomerization, as well as subsequent encapsulation of the newly synthesized genetic material. Formation of RNA-like oligomers was reported by oligomerization of nucleoside monophosphates and product formation was demonstrated using radioactive gel analysis and Nanopore sequencing. Subsequent studies have confirmed the arrangement of monomers within lipid lamellae using techniques like neutron scattering. The exact chemical nature of these oligomeric products, however, could not be interpreted based on these studies. Towards this, the primary objective of our study was to

replicate the oligomerization reactions facilitated in prebiotic niches like volcanic geothermal pools, and systematic analysis of the resultant products using relevant biochemical techniques.

A prebiotic simulator was fabricated, using heating blocks and gas/liquid delivery systems, which allowed laboratory simulation of high temperatures, anaerobic conditions and low pH, as in terrestrial geothermal pools. Oligomerization of adenosine 5'-monophosphate (AMP) was carried out in the presence of phospholipids such as 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Effect of several conditions on the oligomerization reaction, including varying pH, temperatures, nature of the rehydrating agent etc., were characterized. Reactions were analyzed using analytical techniques like High Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS). Each of the parameters studied was found to affect certain aspects of the oligomerization reaction. Low pH and high temperatures favored greater oligomerization while strong mineral acids, like sulfuric acid (H_2SO_4), and moderate dehydration interval (1 hour), were found to be optimum for oligomerization. Presence of lipids was found to have a protective effect on oligomers but excess of lipid resulted in lower oligomeric yields. MS analysis of the AMP oligomerization reaction revealed abasic sites in the oligomers. The loss of nucleobase seen in these reactions can be attributed to acid-labile N-glycosidic linkages that are present in these nucleotides. Almost all the observed masses corresponded to a single base attached to abasic oligomers. The combination of low pH and high temperatures may have resulted in this breakdown seen in the products.



Schematic of lipid-assisted AMP oligomerization by DH-RH cycling and the possible chemical structures of the products obtained.

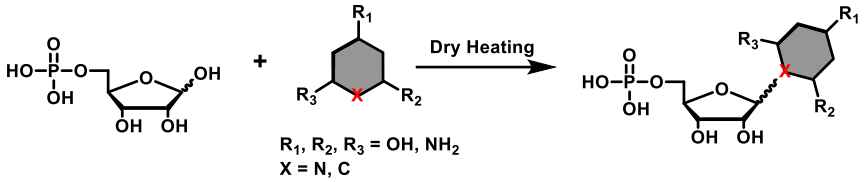
Similar results were obtained in reactions performed with other canonical nucleotides i.e. UMP, GMP and CMP. HPLC analysis of all the reactions showed oligomer formation and up to at least trimmers were observed in all of them. Significantly, the peaks for free bases were observed only in purine samples. Pyrimidine reactions did not show free base peaks on HPLC analysis, which

was consistent with observation of greater stability of pyrimidine nucleosides to acid catalyzed deglycosylation. MS analysis of these reactions confirmed the presence of abasic oligomers in the purine reactions. Reactions pertaining to pyrimidine nucleotides showed mass numbers for intact dimers and also for abasic dimers. The loss of base in case of pyrimidine reactions was attributed to fragmentation during ionization and not to the oligomerization reactions as free base peaks were detected in the MS analysis of pure monomer controls as well.

In related experiments, oligomerization reactions were also carried out with base pairing nucleotide combinations (AMP+UMP & GMP+CMP), and an equimolar mixture of all four nucleotides, to discern the potential for and extent of deglycosylation prevalent in these contexts. Oligomers were observed in all reactions, albeit with more variation in the products, as indicated by HPLC analysis. MS analysis of these reaction mixtures was, however, non-trivial due to the chemical complexity of the resultant products. Nonetheless, masses for abasic oligomers were detected in all the reaction combinations, with loss of bases primarily attributed to the purine components in the reaction. Given the prevalence of deglycosylation, it was characterized under acidic conditions for canonical ribonucleotides. AMP showed degradation to almost half of the starting concentration in ~6.35 hours, under the exact same conditions that also promoted optimum oligomerization (pH 2, 90°C). The presence of abasic sites and prominent deglycosylation of purines indicated a lower fitness for modern nucleotides under prebiotically plausible conditions. Other nucleobases or analogous structures may have instead acted as information storing moieties under harsher environmental conditions of the early Earth. These observations advocate the idea that modern nucleobases may be a product of chemical evolution, with chemical stability acting as a crucial early selection pressure.

As suggested by chemical evolution hypothesis, modern nucleobases could have been preceded by alternate heterocycles, which were likely present on early Earth. Canonical nucleobases do not readily form nucleosides with ribose under prebiotic conditions, an observation famously referred to as the ‘nucleoside problem’ in the field. To overcome this issue, few groups have demonstrated nucleoside formation with alternate heterocycles, in greater yields. Inspired by this idea, we synthesized nucleotides with non-canonical heterocycles, such as barbituric acid (BA), 2,4,6-triaminopyrimidine (TAP) and 2,4 diamino-6-hydroxypyrimidine (DAHP), under prebiotically pertinent conditions. Reactions were carried out with aforementioned alternate bases and ribose 5'-monophosphate (rMP), under dry heating conditions (90°C, 3 hours).

Formation of nucleotides was confirmed with analytical techniques such as HPLC and MS. In control reactions, canonical nucleotides did not result in nucleotides under the same conditions. Furthermore, TAP and DAHP also resulted in the formation of adducts containing more than one rMP attached to the nucleobase due to the presence of exocyclic amino groups on the latter. Detailed analysis was performed on the BA-nucleotides and NMR analysis revealed the formation of higher proportions of C-glycosides in the reaction. BA-nucleotides were also demonstrated to be stable under acidic conditions, as against canonical purine nucleotides that showed prominent deglycosylation.

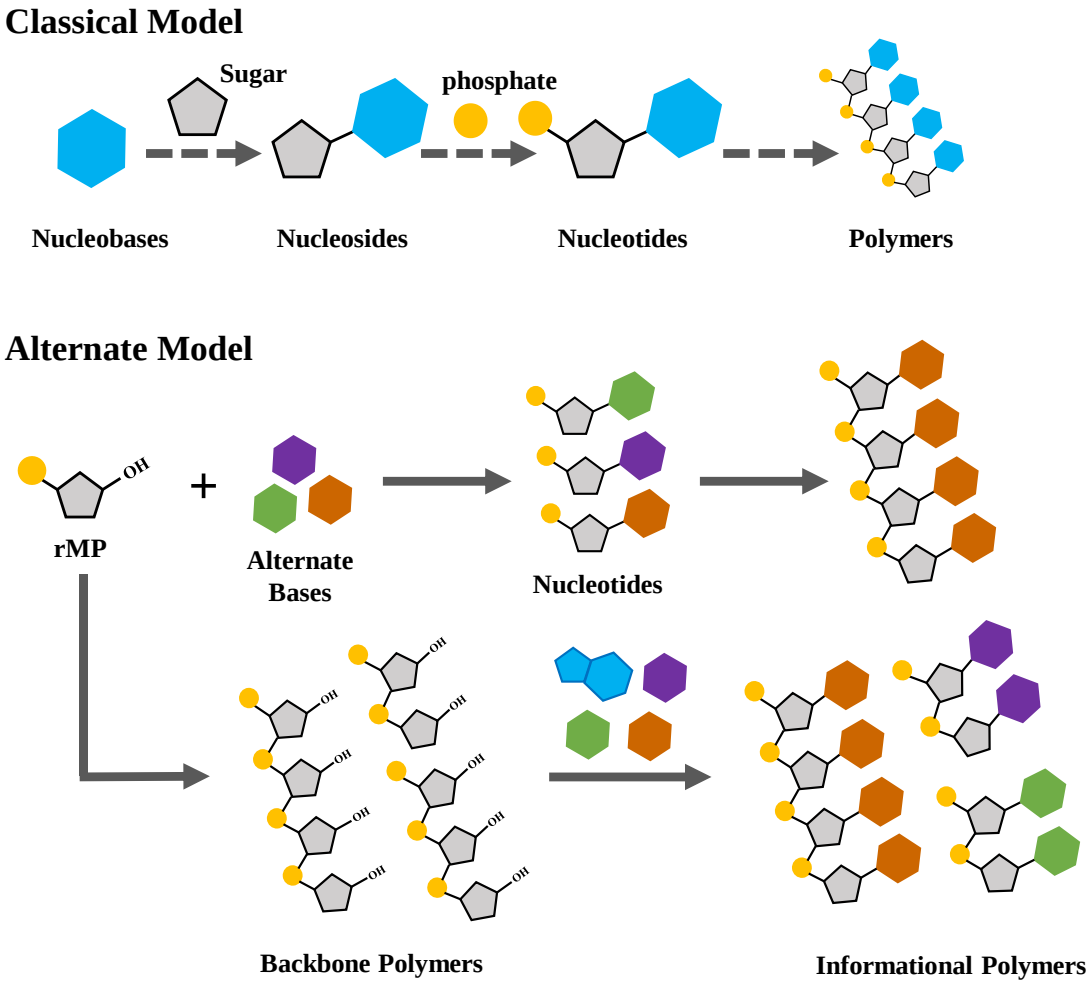


Ribose 5'-monophosphate **Alternate hetetocycles** **Acid stable Nucleotides**

Synthesis of nucleotides containing alternate heterocycles under prebiotic conditions

On subjecting to DH-RH cycles at low pH, the BA-nucleotides resulted in oligomers with intact bases. These observations implied that non-canonical nucleotides might be more suitable candidates for the formation of early informational molecules. The aforementioned synthesis of BA-nucleotides, and their subsequent oligomerization, was demonstrated for the first time by our studies. Yet another novel finding from this project was the oligomerization of rMP, under volcanic geothermal conditions. The oligomeric products from this reaction resemble the backbones of modern nucleic acids. A non-conventional model for the synthesis of informational polymers during the origin of life states that they may have formed by the fusion of a covalently linked backbone polymer and nucleobases. Towards this, we demonstrated the formation of such informational polymers by reacting preformed rMP oligomers and alternate bases, in a proof-of-principle experiment. Canonical nucleobases such as adenine did not react with these rMP oligomers under the conditions tested. Such reactions, involving sugar phosphate backbones and nucleobases, would have allowed for sampling of a large number of informational moieties prior to the emergence of a putative RNA World The traditional model for the emergence of informational molecules during the origin of life assumes a progression from nucleobases to oligomers, via nucleosides and nucleotides. Results from this study indicate that canonical

nucleotides, however, do not result in informational oligomer formation, via the aforementioned scheme. The N-glycosyl bonds are unstable under acidic conditions, resulting in abasic oligomers. On the contrary, alternate heterocycles readily form nucleotides, can undergo oligomerization under acidic conditions without loss of base, and can also form informational oligomers by fusion, with preformed backbones. Our findings highlight the drawbacks of conventional models of prebiotic polymerization and strengthen the hypothesis that modern biopolymers evolved from chemically distinct primitive informational polymers.



Formation of RNA by Classical model and Alternate model. The Latter is supported by several experiments in this project as shown in the schematic.

TABLE OF CONTENTS
(Hyperlinked)

LIST OF SYMBOLS AND ABBREVIATIONS

1,2-Dilauroyl-sn-glycero-3-phosphocholine	DLPC
1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine	POPC
1-phenyl-3-methyl-5-pyrazolone	PMP
2,4,6-Triaminopyrimidine	TAP
2,4-Diamino-6-hydroxypyrimidine	DAHP
2′,4′,6′-Trihydroxyacetophenone	THAP
5-β-Ribofuranosyl-2,4,6-Triaminopyrimdine	TARC
Absorbance at 260 nanometer	A_{260}
Acid dissociation constant	pK_a
Adenine	A
Adenosine 5'-monophosphate + Uridine 5'-monophosphate	AMP+UMP
Adenosine 5'-monophosphate	AMP
Ångström	Å
Barbituric acid (Sodium salt)	BA
Breakdown (in chromatograms)	(B)
Carbon dioxide	CO_2
Carbon-13 (NMR)	^{13}C
Change in enthalpy	ΔH
Change in Gibbs free energy	ΔG
Coenzyme A	CoA
Correlation spectroscopy (NMR)	COSY
Cyanuric acid	CA

Cycle	cyc
Cytidine 5'-monophosphate	CMP
Cytidine	C
Degree Celsius	°C
Dehydration	DH
Deoxyadenosine 5'-monophosphate	dAMP
Deoxyribonucleic acid	DNA
Deuterium oxide	D_2O
Dextrorotatory	D
Differential interference contrast	DIC
Diketopiperazine	DKP
Dimer (in chromatograms)	(D)
Earth Life Science Institute	ELSI
Electrospray ionization	ESI
Enhanced Green Fluorescence Protein	EGFP
Ethidium bromide	ETBR
Flexible nucleic acids	FNA
Formamide	CH_3NO
Glycerol nucleic acids	GNA
Glycine alanine aspartic acid valine	GADV
Guanine	G
Guanosine 5'-monophosphate + Cytidine 5'-monophosphate	GMP+CMP
Guanosine 5'-monophosphate	GMP
Half-live	$t_{1/2}$

Hertz	Hz
Heteronuclear multiple bond correlation (NMR)	HMBC
Heteronuclear single quantum correlation (NMR)	HSQC
High Performance Liquid Chromatography	HPLC
High-Resolution Mass Spectrometry	HRMS
Histidine	His
Hour	hr
Hydrochloric acid	HCl
Hydrogen cyanide	HCN
Kilocalorie per mole	kcal mol^{-1}
Last Universal Common Ancestor	LUCA
Levorotatory	L
Magnesium Chloride	$MgCl_2$
Mass spectrometry	MS
Matrix-Assisted Laser Desorption Ionization	MALDI
Measurement of hydrogen ion concentration	pH
Messenger Ribonucleic acid	mRNA
Methane	CH_4
Micrometer	μm
Millimolar	mM
Milliseconds	ms
Minute	min
Molecular ion	[M]
Monomer (in chromatograms)	(M)

Nanometer	nm
National Chemical Laboratory	NCL
Nitric acid	HNO_3
Nuclear magnetic resonance spectroscopy	NMR
Nucleoside 5' triphosphate	NTP
Nucleoside 5'-phosphorimidazolides	ImpN
Octadecyl carbon chain bound to silica	C18
Orthophosphoric acid	H_3PO_4
Parts per million	ppm
Peptide Nucleic Acids	PNA
Phosphodiester backbones	PDB
Polyadenylic acid	PolyA
Polycyclic aromatic hydrocarbons	PAH
Polyether ether ketone	PEEK
Polytetrafluoroethylene	PTFE
Polyuridylic acid	PolyU
Proton (NMR)	^1H
Reductive tricarboxylic acid cycle	rTCA
Rehydration	RH
Ribonucleic acid	RNA
Ribonucleoprotein	RNP
Ribose 5'-phosphate	rMP
Ribosomal Ribonucleic acid	rRNA
Rotating-frame Overhauser effect spectroscopy	ROESY

Salt induced peptide formation	SIPF
Second	s
Serine	Ser
Sodium Perchlorate	$NaClO_4$
Sulfuric acid	H_2SO_4
Tetramer (in chromatograms)	(Te)
Threose nucleic acids	TNA
Thymine	T
Time of flight	TOF
Total correlation spectroscopy (NMR)	TOCSY
Transfer Ribonucleic acid	tRNA
Tricarboxylic acid cycle	TCA
Trimer (in chromatograms)	(T)
Trizma Base	Tris
Ultraviolet	UV
Uracil	U
Uridine 5'-monophosphate	UMP
Wavelength	λ

LIST OF FIGURES
(Hyperlinked)

LIST OF TABLES
(Hyperlinked)

Chapter 1

Introduction

1.1 The Origin of Life Conundrum

Earth is the only planet known so far to harbor life. The phenomenon of origin of life, both, on Earth, and possibly in the universe, continues to be a fascinating scientific enigma. Decades of research has gone into hypothesizing and predicting the steps that may have led to the formation of first living organisms on Earth. Even then, recreating these steps in a laboratory environment continues to pose formidable challenges. It is universally accepted in the origins of life community that several environmental factors would have come into play while facilitating chemical reactions that eventually resulted in the molecules that we see in biology today. The transition from chemistry to biology is thought to have happened over millions of years by 'evolution' of these reaction systems. However, one of the major problems is the lack of evidence of this "molecular evolution" process, which is not preserved in modern biology.

Stromatolites, which are considered as fossil evidence of the oldest life forms, are the main proof for the existence of microbes, like cyanobacteria, about 3.8 billion years ago. Geological records, based on a study of zircons, indicate that liquid water was present on Earth as early as about 4.2 billion years ago (Bernhardt and Tate, 2012). Thus, between the periods when water first appeared on the planet, up to the period that allowed for the rise of complex microbes, quite a lot of prebiotic chemistry and minimal life processes must have ensued. This eventually would have allowed for the transition to biology, events for which fossil/geological records are hard to find. The problem with investigating these transitions without any geological fossils is being addressed by recreating the processes that may have eventually led to modern biochemistry. Consequently, research in the field of origin of life is focused on delineating chemical reactions that would have been possible on the early Earth, which are considered to have given rise to various complex molecules. These complex molecules are thought to have interacted with each other resulting in primitive metabolism, which, when encapsulated in amphiphilic membranes, would have formed structures that resemble protocells. Further evolution of these protocells with respect to the complexity of metabolism, the mechanism of energy utilization etc., would have resulted in the Last Universal Common Ancestor (LUCA) or progenote like entities. Our study mainly focused on the formation of informational polymers from monomers, as most functional units within the cells, such as nucleic acids and proteins, are polymers.

The term informational polymer used above refers to any polymer that is capable of storing information, which is reflected in the order of its monomers contained in the sequence. There may be emergent properties associated with the sequence, such as the formation of H-bonds, capacity to fold or the evolution of other functions, which are not being studied. These properties are not found at the level of the monomer and possibly arise due to formation of higher order structures in the polymers. This resultant increase in complexity is unique to the higher order system that is more than just the addition of the individual properties of the monomers. By this definition, both nucleic acids and proteins can be considered as informational molecules as they have an orderly arrangement of monomers and display aforementioned emergent properties.

Modern biology uses activated monomers (e.g. NTPs) and catalysts, in the form of protein complexes/ribosomes, to synthesize both these molecules. Nucleic acids are synthesized from nucleoside triphosphates with the help of polymerase enzymes, which make new strands based on the information stored in the template strand. Proteins, on the other hand, are synthesized by the process of translation in which ribosomes assemble aminoacylated tRNAs based on the sequence of nucleotides in the mRNA. This process of transfer of information from DNA to RNA to Proteins, also known as Central Dogma in biology, also highlights the interdependent nature of these molecules. For e.g. the synthesis of nucleic acids requires protein machinery while the translation process utilizes several modified RNAs and the information stored in the DNA to synthesize proteins. Which of these biomolecules originated first from the prebiotic soup? This is the biochemical version of 'the chicken and the egg causality dilemma.' One putative respite to this conundrum comes in the form of RNA molecules that can do both information storage and catalysis. This has given to rise to a popular hypothesis known as the 'RNA World hypothesis', which will be discussed in detail later. Other relevant ideas have also been put forward to conceptualize how early life may have used pertinent biomolecules to perform critical functions. In this context, several theories have been proposed to understand and explain the emergence of life-like systems on primitive Earth. Combined together, these theories are collectively referred to as the 'Many Worlds hypotheses', which deal with discerning the origin of specific biomolecules or systems from the prebiotic soup (Figure 1.1). We will take a look at the various putative worlds that have been hypothesized to have existed during the origin of life.

1.2 The ‘Many Worlds hypotheses’

Recreating the complexity of extant life by purely chemical means has proven to be a formidable task. The major problem is the lack of chemical fossils that can explain the rise of biochemistry in the absence of enzymes encoded by genes. Therefore, researchers have instead relied on tracing back the fundamental processes of life such as metabolism, reproduction, information storage, catalysis etc. Depending upon what types of molecules achieves these processes, many hypotheses have been proposed that explain certain aspects of life-like systems. These hypotheses encompass the emergence/evolution of early biomolecules under proposed geochemical/atmospheric conditions that may have been plausible on the early Earth. We will discuss some hypotheses which are relevant to the current study.

1.2.1 RNA World hypothesis

As the name suggests, this hypothesis considers RNA to be the most important biomolecule during the process of emergence of life. The original hypothesis was proposed by Alexander Rich (Neveu et al., 2013), which was reinforced by the discovery of ribozymes (catalytic RNA molecules) in the lab (Guerrier-Takada et al., 1983; Kruger et al., 1982). The term “RNA World” was coined by Gilbert (Gilbert, 1986) to describe a very early time in the evolution of life when RNA would have played a central role because of its ability to act as, both, the genetic material and as catalysts. Eventually, once a self-replicating RNA molecule emerged from the prebiotic soup, it is hypothesized to have had the ability to propagate its information (a primitive form of reproduction), and to also evolve by mutation of its fundamental sequence. Several other important discoveries, such as RNA viruses and the fact that the catalytic core of ribosomes is made exclusively from RNA moieties, have reinforced the aforementioned dual functional capability of RNA molecules. Furthermore, the various functions served by RNA molecules in modern biology, be it in the process of protein synthesis or as regulators of various biochemical processes, are considered to be remnants of a putative RNA World. Studies in this area basically look at how RNA molecules may have been synthesized on prebiotic Earth and how various functions that they perform can be acquired by the selective evolution of RNA molecules in the lab.

1.2.2 Ribonucleoprotein (RNP) World

As mentioned earlier, the example of ribosome has been considered a beacon for the RNA World hypothesis. Nonetheless, the catalytic core of the ribosome is actually a ribonucleoprotein wherein the intricate cooperation between the RNA and the protein counterparts is essential to facilitate peptide bond formation. Therefore, some have argued that RNA and proteins would have evolved simultaneously and not sequentially. Another important intermediate in protein synthesis, the peptidyl-tRNA, is considered a relic of this RNP World where covalently linked RNA and peptides/amino acids may have served as enzyme-coenzyme complex (Guerrier-Takada et al., 1983). This also has given rise to a "tRNA Core hypothesis", where proto-tRNAs are considered to be at the core of a proto-translation system and mRNA and rRNA are thought to have evolved later (de Farias et al., 2016). This sub-hypothesis tries to connect the RNA World to an RNP World, and the further evolution of fundamental biological processes, with tRNA as a central molecule. Some other researchers have proposed a peptide/RNA partnership during the early evolution of life diminishing the need for a single polymer to have both catalytic and informational properties(Carter, 2015). The RNP World ideas establish a role for peptides in the early stages of the origin of life and also link RNA World to modern biological processes.

1.2.3 Metabolism-first hypothesis

The RNA World hypothesis can also be considered as the replication-first hypothesis as it prioritizes the origin of self-replicating molecules. On the contrary, in the Metabolism-first hypothesis school of thought, life is conceptualized as a series of chemical reactions that is similar to metabolism and the origin of these early metabolic networks is considered crucial to the emergence of life. These self-perpetuating and evolving proto-metabolic networks would have comprised of simple organic molecules and preceded self-replicating genetic materials. This metabolism-first approach assumes that early life was a self-sustained chain of chemical reactions on mineral surfaces driven by environmental factors, with the genetic encoding of functions evolving much later (Bada and Lazcano, 2002). The primitive metabolism would have comprised of organic compounds synthesized by series of chemical reactions from simple 2 or 3 carbon molecules. A primitive type of reductive citric acid cycle (rTCA), in which the input is chemical energy, water and CO_2, and the output is a set of complex organic molecules, is often cited as a model to support this possibility (Trefil et al., 2009). Other studies have looked at

autocatalytic reactions that can give rise to the sustained production of certain biomolecules (Damiano and Luisi, 2010). Various studies pertaining to this school of thought investigate reactions such as the formose reaction, reverse TCA cycle, autocatalytic peptides etc. to decipher the origin of early metabolism.

1.2.4 The Iron-sulfur World

This idea was proposed by Gunter Wächtershäuser in the late 1980s and extends the idea of metabolism-first approach to include the role of specific minerals in the emergence of life. It suggests that early metabolism would have occurred on mineral surfaces such as pyrites. The iron-sulfur minerals have positive charges, which interact with negatively charged monomers, promoting the formation of polymers (Wächtershäuser, 1988). Some experiments have also shown the synthesis of pyruvate with iron-sulfate catalyst, using high pressure and temperature conditions (Cody et al., 2000). Other minerals such as Zn^{2+}, Cr^{3+} and Fe^{0} have shown to promote 6 out of 11 reactions of the rTCA cycle in aqueous acidic solutions (Muchowska et al., 2017). The same group has also demonstrated CO_2 fixation that results in acetate and pyruvate (products in acetyl-CoA pathway), using native transition metals that could have acted as C2 and C3 metabolites (Varma et al., 2018). Apart from the synthesis of several metabolic intermediates, iron-sulfide bubbles, containing alkaline and highly reduced hydrothermal solutions, are thought to have acted as early membranes allowing surface chemistry (Russell et al., 1994). This hypothesis also proposes the origin of the redox gradients, seen across modern cell membranes, as a result of evolution from iron-sulfide mounds in hydrothermal settings. The pH, temperature and redox gradients that existed between sulfide-rich hydrothermal fluid and iron (II) containing water of Hadean ocean floor may have acted as an energy source for early metabolism (Martin and Russell, 2003).

Apart from these hypotheses, several other ideas have also been proposed such as the Lipid World hypothesis (in which amphiphiles play a crucial role), Clay theory (first molecules of emerged and evolved on clay surfaces), GADV-protein World hypothesis [one letter genetic code giving rise to early proteins made of glycine (G), alanine (A), aspartic acid (D) and valine (V)] and Polycyclic Aromatic Hydrocarbons (PAH) World hypothesis (wherein PAH played a role in synthesis and evolution of RNA). Amongst all these worlds and hypotheses, the RNA World hypothesis is the most widely accepted one and several decades of research has been

focused on recreating prebiotic scenario that would have allowed for the emergence of RNA. RNA is a chemically complex molecule and the RNA World hypothesis obligates the presence of RNA polymers with catalytic functions. Smallest functional ribozymes evolved under laboratory conditions are still about 50 to 150 nucleotides long (Ferré-D'Amaré and Scott, 2010), which recapitulates the need for long polymers during RNA World, making polymerization a crucial step. Thus, one can broadly divide studies in RNA World hypothesis into three categories: (i) the formation of the RNA monomer, (ii) the polymerization of this monomer and (iii) the evolution of function in the nascent polymer. We will now briefly discuss each of these aspects and related studies in the following section.

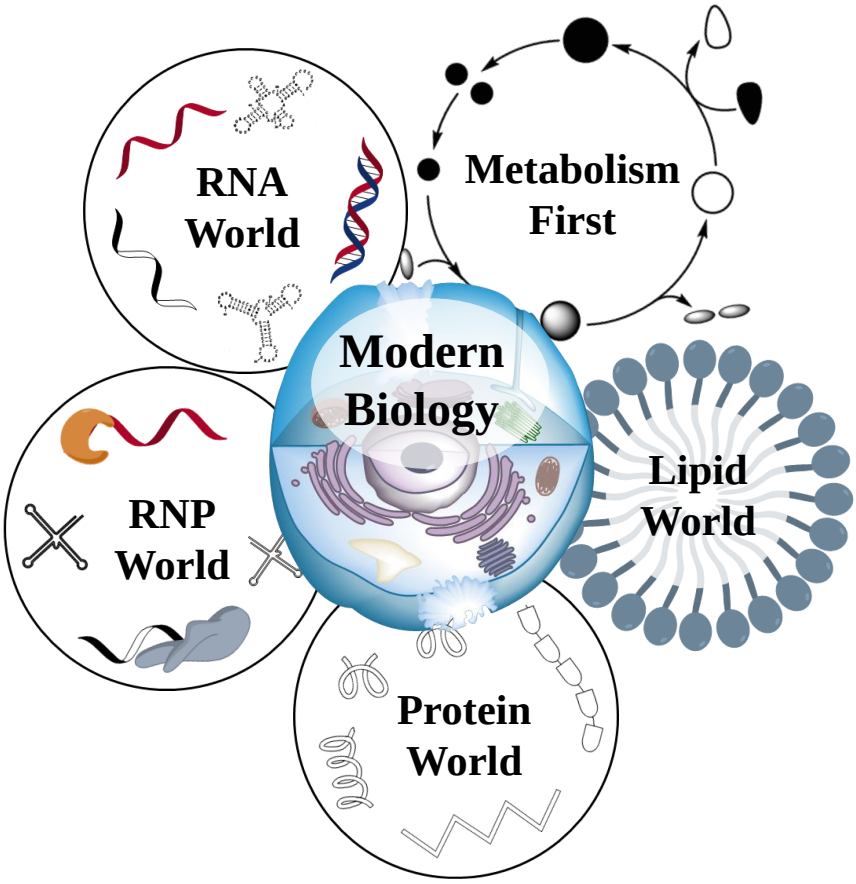


Figure 1.1: Many Worlds hypotheses regarding the origin of modern biology, and pictorial representation of the biomolecules or processes that are thought to have played major role in the respective hypothesis.

1.3 Defining the problem

Studies in the origin of life field address various levels of complexity at which this question operates. Astrobiology, the parent topic under which the origin of life studies falls, deals with understanding how life came about on our planet. Exobiology is the area of research that deals with discerning whether there is life in the universe. This includes detection of life elsewhere in the universe, studying celestial objects for the presence of bio-signatures and identifying planets that may have life-supporting conditions. A good body of research has revealed the various chemistries that could result in the abiotic synthesis of molecules that are usually associated with life as we know it. It assumes that a series of logical steps would have occurred during the process of life's evolution on Earth. Decades of studies pertaining to each part of this process has shed light on certain aspects of these steps which are discussed below.

1.3.1 Synthesis of biological monomers

Biomolecules are chemically complex structures with seemingly improbable chemical bonds. Even at the level of a monomer, molecules such as sugars, fatty acids, nucleobases and amino acids are heterogeneous multiatomic structures, which cannot be synthesized spontaneously. Scientists have thus proposed many chemical routes by which simple carbon compounds such as CO_2, HCN, CH_4, and CH_3NO etc. can give rise to more complex molecules. One of the pioneering studies was the Miller-Urey experiment that demonstrated the formation of amino acids and few other biomolecules by simulating early atmospheric conditions (Miller, 1953). A recent analysis of the same reaction vials, using more advanced analytical techniques, has revealed the formation of many more biomolecules, albeit in low concentrations (Johnson et al., 2008). Other synthesis pathways for biomolecules such as sugars and fatty acids have also been proposed. Formose reaction, a proposed pathway for prebiotic source of sugars, demonstrates the synthesis of sugars from formaldehyde, catalyzed by base and divalent cations (Orgel, 2000). As for the formation of amphiphiles such as fatty acids, Fischer-Tropsch-type synthesis from aqueous solutions of formic acid or oxalic acid, under hydrothermal conditions, has been shown to result in the formation of long chains of hydrocarbons (Mccollom et al., 1999). One of the most complex molecules in modern biology are nucleotides, which are monomers of nucleic acids. Synthesis of a complete nucleotide monomer, either in one step or involving sequential steps, has not been demonstrated so far. Synthesis of adenine was reported by heating the

solution of hydrogen cyanide at moderate temperatures (ORÓ, 1961). A plethora of biomolecules, including A, C, U and T nucleobases and their nucleosides, have been shown to be synthesized by meteorite-catalyzed proton irradiation of formaldehyde (Saladino et al., 2015). Furthermore, phosphorylation of nucleosides by mobilization of phosphate from minerals, in urea-based solvents, was recently described (Bradley et al., 2016). These are a few studies mentioned to highlight the formation of biologically relevant monomers and the various chemical processes that may have happened on the early Earth.

1.3.2 Going from monomers to polymers

In modern biology, polymerization is carried out using sophisticated enzymes and high energy molecules, in the form of nucleoside triphosphates and aminoacyl-tRNAs. Such high energy monomers enable uphill polymerization reactions, thus making it no longer a chance driven event due to the presence of specific protein catalysts. Fundamentally, polymerization reaction in biomolecules is chemically a condensation reaction where loss of a water molecule results in the formation of a bond between adjacent monomers. Both the significant classes of biopolymers, i.e. proteins and nucleic acids, have linkages which are a product of condensation. Origin of life researchers have, therefore, focused on understanding nonenzymatic polymerization of amino acids and nucleotides, to make peptides and nucleic acid oligomers, respectively. Oligomerization of amino acids such as glycine and alanine was demonstrated by Rode's group in a series of studies (Rode, 1999). This reaction is termed as salt-induced peptide formation (SIPF), which has been shown to be catalyzed by clay and other related catalysts like alumina and silica (Le Son et al., 1998). One major caveat of these studies was the short length of the resultant oligomers (typically up to trimers) due to the formation of diketopiperazine, a stable cyclic intermediate that acts as a 'dead end'. Recent reports have used one-pot dehydration-hydration as the driving force for the formation of peptides (Rodriguez-Garcia et al., 2015). This study showed the formation of glycine oligopeptides up to 20 amino acids long and also the incorporation of eight other amino acids by varying the reaction parameters. On parallel lines, oligomerization of α-hydroxy acids was also shown to take place by formation of ester bonds when subjected to multiple dry-wet cycles (Mamajanov et al., 2014). Similar reactions, when carried out with a starting mixture of α-hydroxy acids and α-amino acids, resulted in ester-mediated amide bond formation. The resultant 'depsipeptides' were up to 14 units long and contained a combination of ester and amide linkages, with a preference for the latter (Forsythe et

al., 2015). Oligomerization of nucleotides has also been studied extensively, with the majority of these studies using ribonucleotide monomers as motivated by the RNA World hypothesis. Since polymerization of nucleotides is also unfavorable in water, many environmental niches were considered as potential sites that support abiotic RNA synthesis. To increase the reactivity of the phosphate group, better leaving groups such as imidazole have been extensively used to modify the phosphates, such that these reactions can be followed on laboratory time scales. Studies pertaining to prebiotic oligomerization, resulting in RNA and RNA-like oligomers, have been discussed in detail in Section 1.4.

1.3.3 Structural and functional evolution

Non-Darwinian evolution of biomolecules must have taken place during the origin of life, before the emergence of genetic inheritance of characteristics. New properties can potentially emerge due to the formation of certain structures in the growing oligomers and these can impart selective advantage to the molecules in question. An argument, therefore, can be made for backtracking structure-function relationship seen in modern biological systems to prebiotic evolution. For example, in small oligomers, the ability to form secondary structures by H-bonding is an emergent property and can be considered as a function. On similar lines, the formation of vesicles by amphiphiles is an important emergent property that is considered to have resulted in early compartmentalization. Various studies have shown the formation of vesicles from mixed fatty acid systems, which have a selective advantage over simple fatty acid systems made of only a single type of amphiphile (Maurer et al., 2009; Namani and Deamer, 2008). Not just amphiphiles, even short peptides have been shown to form self-assemblies that resemble lipid membranes (Childers et al., 2009). Apart from structure formation, other functions such as catalysis must have also emerged early on. For e.g. catalytic properties of small heteropeptides were demonstrated for various kinds of reactions. L-dipeptides, mainly Serine-Histidine (Ser-His) dipeptide, has been shown to catalyze the formation of the phosphodiester bonds, resulting in short RNA chains from activated RNA monomers (Wieczorek et al., 2013). The same Ser-His dipeptide, when encapsulated in fatty acid vesicles, was shown to promote the competitive growth of vesicles analogous to predator-prey type interactions (Adamala and Szostak, 2013). The signature property of nucleic acids, i.e. information storage and replication, has been studied comprehensively by Orgel's group. Their studies on template-directed condensation of activated nucleotides have advanced our understanding of the feasibility and fidelity of nonenzymatic

replication (Joyce et al., 1984; Orgel, 1998; Wu and Orgel, 1992). To improve the fidelity of the replication process, changes to the chemical structure of the nucleobase, specifically 2-thio modification of U, have been shown to be very effective (Zhang et al., 2013). Other structural properties of RNA, such as the presence of unnatural 2'-5' linkages, have shown to lower melting temperatures and did not affect molecular recognition and ribozyme function (Engelhart et al., 2013). Overall, several properties of biomolecules that are dictated by their intrinsic structural features, have been studied, with a focus on those properties that are found in modern biological processes.

1.3.4 Prebiotic soup to protocells

The top-down approach for creating the 'simplest cell', by reducing the complexity to a bare minimum, has not yet yielded significant progress. Deconstructing the bacterial genome to its essential components, still leaves over 400 protein-coding genes in its minimal genome (Glass et al., 2017). Thus, a bottom-up approach to synthesizing protocells, primitive cell-like structures, has gathered much attention in recent past. A protocell is hypothesized to be consisting of a three-dimensional boundary structure (membrane) and an encapsulated genetic material (with replicating functions). Encapsulation of macromolecules such as DNA into lipid vesicles, under simulated prebiotic conditions, was demonstrated over 30 years ago (Deamer and Barchfeld, 1982). Aggregation of ribozymes with micelles has shown to improve polymerization yield of the product, thereby also suggesting a catalytic role for the amphiphiles and not just as boundary forming membranes (Bartel et al., 2008). Studies from Szostak's group have shown RNA adsorbed on clay particles getting encapsulated into fatty acid vesicles (Hanczyc et al., 2003). In another study, encapsulation of genetic material promoted growth of vesicles at the expense of other empty vesicles (Chen and Szostak, 2004). These growing vesicles were later shown to divide without the loss of the encapsulated content by application of modest shear forces (Zhu and Szostak, 2009). Furthermore, encapsulation of all components of a cell-free protein expression system in liposomes, which can transcribe and translate Enhanced Green Fluorescence Protein (EGFP), has also been successfully demonstrated (Murtas et al., 2007). The synthesis of EGFP in these liposomes was confirmed suggesting the possibility of facilitating complex biochemical reactions in semi-synthetic minimal cells. Although the attempts for recreating the processes that would have led to the first protocell like entities are still ongoing, our understanding of the individual pieces of this puzzle has improved considerably.

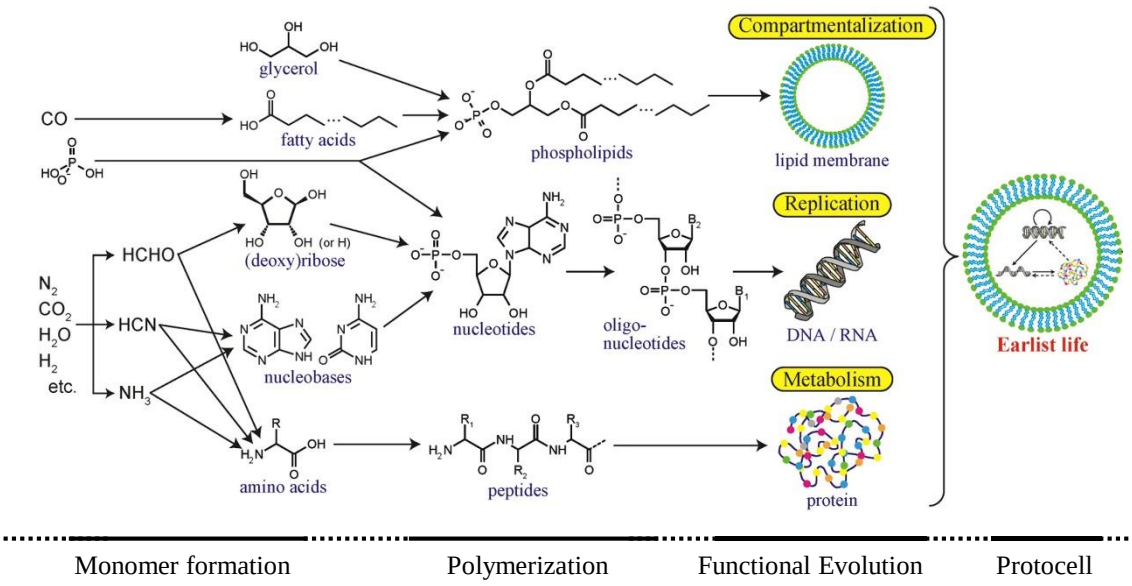


Figure 1.2: Overview of steps during the chemical origins of life on Earth. Progression from synthesis of monomers to formation of protocell is depicted as a discontinuous line in order to reflect the lack of direct evidence of this process. [Adapted from (Kitadai and Maruyama, 2017)]

Each arrow in the above figure represents complex reactions taking place in specific niches that can be studied extensively. However, in order to keep the questions well defined, we focused on delineating the processes pertaining to polymerization of nucleotides that result in oligomers. This allowed us to characterize the oligomerization reaction without having to deal with the complexities associated with the origin of monomers or the evolution of functions post oligomerization. Formation of RNA by nonenzymatic oligomerization may appear to be essentially a ‘simple’ condensation reaction. However, it has been shown to be a difficult task to reproduce under simulated conditions in the lab. This difficulty can be attributed to potential hydrolysis of the phosphodiester linkage that takes place in aqueous conditions. Also, the chance-driven nature of the polymerization reaction itself adds to the issue at hand. To increase the rate of this reaction, several catalysts, factors such as adsorption onto surfaces, and other concentrating mechanisms have been proposed. In the next section, we will discuss various ways in which RNA oligomerization reactions have been studied thus far.

1.4 Oligomerization of RNA monomers

1.4.1 Selection of RNA monomers for oligomerization

Formation of nucleic acids in water, by polymerization of nucleoside monophosphates, is quite improbable due to thermodynamic and kinetic constraints. The free energy (ΔG) required for the formation of the phosphodiester bond is estimated to be 5.3 kcal mol $^{-1}$ at 25°C, pH 7 (Dickson et al., 2000). Their half-life in oligonucleotides can vary in the range of 1 hour to 10 days at 100°C and pH 7 (Kawamura, 2004). Use of activating agents, such as cyanamide and water-soluble carbodiimide, as catalysts, has resulted in the formation of very short oligomers, or only dimers, respectively, in the aforementioned cases (Ferris et al., 1989; Ibanez et al., 1971). The rate of phosphodiester bond formation from nucleoside 5'-triphosphates (NTP), in the absence of enzymes, was found to be very low (5.4 x 10^{-7} per hour) (Bartel and Szostak, 1993). This reaction was shown to be accelerated in the presence of divalent cations (Mg^{2+}) and in vitro selected ribozymes, nevertheless negating the plausibility of spontaneous oligomerization of NTPs. Formation of phosphodiester bond, by condensation between two nucleotides (monophosphates), involves the loss of a water molecule, with the –OH on phosphate being the leaving group. In a modified nucleotide synthesis reaction the –OH was replaced by imidazole, thus creating nucleoside 5'-phosphorimidazolides (ImpN). In the case of these modified nucleotides, the phosphodiester bond forms by the elimination of imidazole, and because it is a better leaving group than –OH, the bond formation is much faster (Figure 1.3).

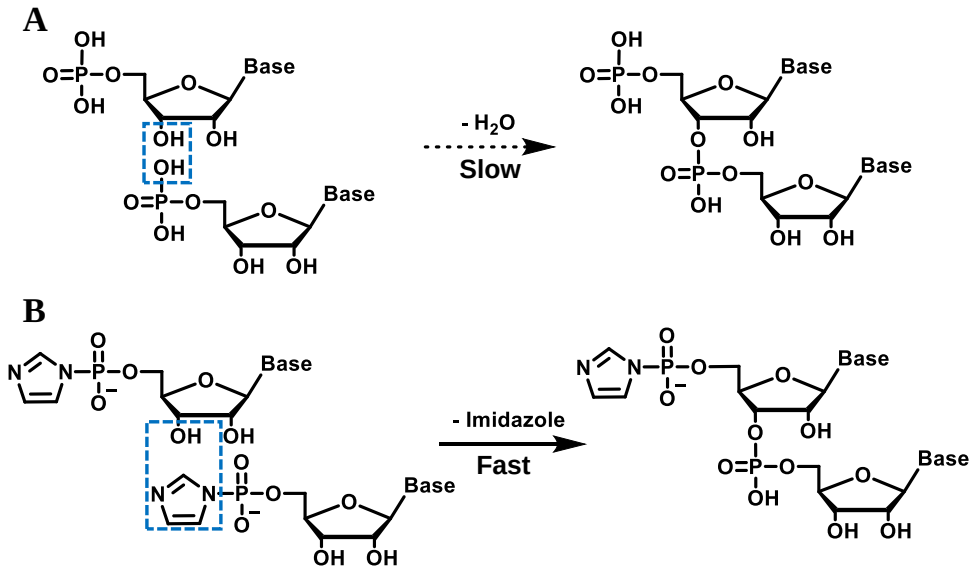


Figure 1.3: Condensation reactions between NMP monomers (A) and ImpN monomers (B), resulting in phosphodiester bond formation

Magnesium ion-catalyzed synthesis of adenosine 5'- phosphorimidazolides (ImpA) was demonstrated by heating adenosine 5'-polyphosphates and imidazole (Lohrmann, 1977). Formation of oligonucleotides from ImpA in aqueous solution was shown to be catalyzed by Zn^{2+} (Sawai and Orgel, 1975). Several other divalent cations (such as Pb^{2+}, Co^{2+} and 9 others) were also shown to promote oligomerization to varying extent, with Pb^{2+} being the most effective cation (Sawai, 1976). The oligomers observed from this reaction had varying types of internucleotide linkages, like 2'–5', 3'–5' & 5'–5' bonds, and even cyclic oligonucleotides. Follow up studies showed that certain cations, such as Uranyl cation (UO_2^{2+}), selectively catalyzed the formation of 3'–5' bonds but the presence of other cations in the reaction reduced this selectivity (Sawai et al., 1992). Alternate studies have used surface catalysts and other concentrating mechanisms for studying oligomerization of ImpNs, which are discussed below in detail.

1.4.2 Clay-catalyzed oligomerization

Nonenzymatic polymerization in bulk solutions is often a chance event as two monomers need to be in close proximity for the reaction to occur. The condensation does not take place very effectively due to the large diffusional mobility of the monomers in solution. Presence of adsorbing surfaces help in overcoming this hurdle and such surfaces also result in an increased local concentration of monomers, which, in turn, increases the chance of polymerization. In this regard, clay and mineral surfaces have been shown to act as catalysts in the formation of polymers. Clay would have been present in large abundance on prebiotic Earth (as a product of volcanic degassing), and has a high surface area to volume ratio due to its porous nature. Nucleotides have been shown to get adsorbed onto montmorillonite, a type of smectite clay, to varying extent. This adsorption process is found to be pH dependent and has been shown to cause ionic changes in the clay. For e.g., Fe^{2+} gets oxidized to Fe^{3+} on adsorption of nucleotides onto clay, thus the interaction is not merely a physical process (Carneiro et al., 2011). Ferris and coworkers have extensively investigated the effects of montmorillonite and other clays and have reported the synthesis of long oligonucleotides from ImpNs (Ferris and Ertem, 1992). Polymerization of activated nucleotides was found to be optimum at alkaline pH of around 7-9 and favored by high concentrations of salt in the solution (Miyakawa et al., 2006). Analysis of reaction products from clay-catalyzed oligomerization by mass spectrometry has revealed the formation of oligomers up to 40mers (Zagorevskii et al., 2006), but there is a chance of false positives in the analysis of such complex reaction mixtures. The catalysis by clay has been

shown to be regio and diastereo specific. When polymerization was carried out using a mixture of D-ImpA and L-ImpA, the homochiral dimers were formed in much greater amounts than heterochiral dimers. Also, the reaction was found to be more efficient in forming 3'–5' linkages between the monomers as against 2'–5' linkages (Urata et al., 2008). The former observation indicates that the two stereoisomers, due to a different geometry, might potentially undergo polymerization in a differential manner, thus resulting in more homochiral polymers. It has been proved that not all clays have similar catalytic properties due to the differences in their internal structure and ionic composition, which results in differences in the way the monomers can get adsorbed (Joshi et al., 2009). Clay-catalyzed oligomerization of ImpNs is one of the most extensively studied reaction and has contributed greatly towards our understanding of surface-catalyzed reactions resulting in RNA oligomers.

1.4.3 Eutectic phases as a concentrating mechanism

Freezing of water to form ice is caused by reduced thermal motion and formation of more number of hydrogen bonds between water molecules. Addition of salts to water reduces the overall freezing point of the solution and results in eutectic freezing. Eutectic phases in ice result from the differences in the freezing points of pure water versus that of the salt solution. As the temperature lowers, water molecules get associated with each other preferentially (due to H-bonding) and exclude solutes from the process of ice formation. At a particular temperature regime, there is significant ice formation in the solution, but the whole solution is not completely frozen. This selective exclusion of solute molecules from ice crystals results in increased concentration of that particular solute (e.g. nucleotides) in brine channels, consequently resulting in increased polymerization rates. Many of the previously reported studies, such as phosphorylation of nucleosides (Section 1.3.1) and the catalytic activity of histidine-containing dipeptides (Section 1.3.3), were observed in eutectic solvents and water-ice environments, respectively. Eutectic freezing can thus act as a concentrating mechanism for the molecules that are present in otherwise dilute solutions, thereby promoting oligomerization by increasing the proximity of the reacting monomers.

Experiments have also shown that various reactions resulting in the formation of nucleotides, from hydrogen cyanide and cyanoacetylene, can also take place in eutectic phases (Menor-Salván and Marín-Yaseli, 2012). Additionally, polymerization of ImpNs, to yield oligomers up

to 7-10mer in length, has been observed when a mixture of monomers and divalent cations was incubated at -18°C for several days (Kanavarioti et al., 2001) (Figure 1.4 A). Staining with nucleic acid specific stains (e.g., Acridine orange, ETBR) revealed that the fluorescence was confined to the eutectic phases surrounding the ice crystals (Figure 1.4 B). Polymerization under such conditions also seemed to favor the formation of 3'-5' phosphodiester bond as was indicated by enzymatic cleavage analysis.

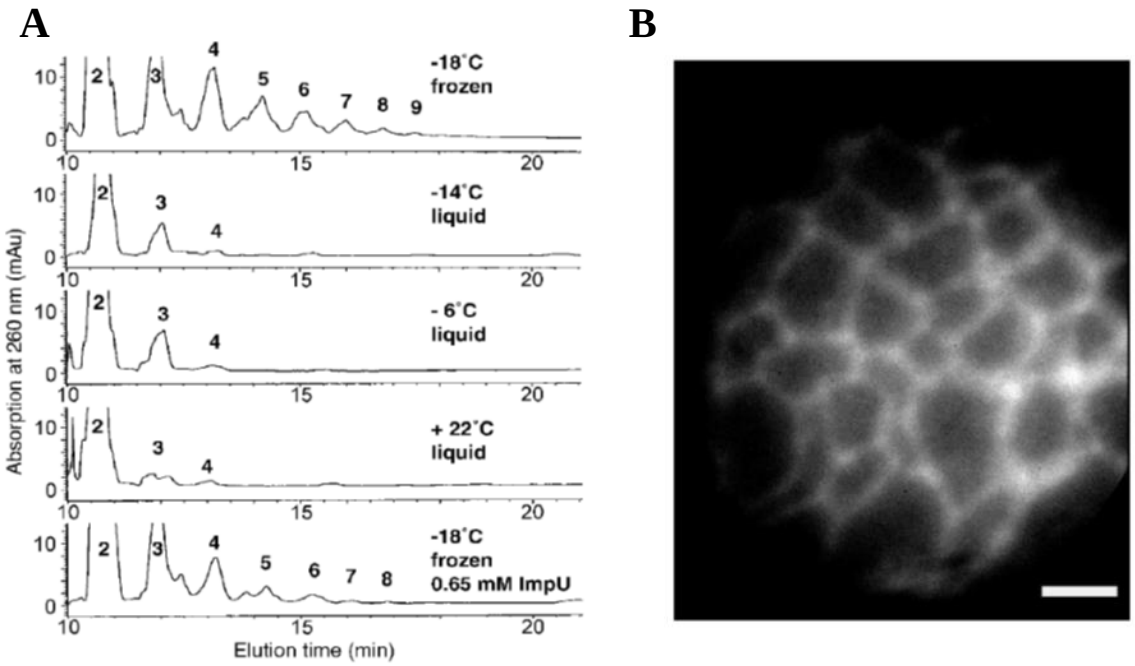


Figure 1.4: Temperature dependent oligomerization of uridine 5'-phosphoimidazolide (ImpU) was observed in reactions, wherein freezing of sample was necessary for higher oligomer formation (A). Staining of concentrated oligomers in frozen samples was observed as light outline around dark ice crystals by epifluorescence microscopy (B). Scale Bar = 26.7μm. [Adapted from (Kanavarioti et al., 2001)]

Another study showed that polymerization under eutectic conditions also resulted in comparable incorporation of pyrimidines as well, which generally is not the case in other polymerization scenarios (Monnard et al., 2003). Formation of mixed sequences is dependent on the starting ratio of all four nucleotides. The order of efficiency in which nucleotides undergo polymerization in these scenarios was found to be A > G ≈ U > C. However, the overall incorporation of both purines and pyrimidines with a ratio of 1.2-1.3 was observed in such reactions. Eutectic ice phases have also been shown to be conducive for metal-ion catalyzed templated-ligation of short

oligomers and many ribozymes have been shown to be functional at subzero temperatures in ice-water matrices (Monnard and Szostak, 2008; Vlassov et al., 2005). Importantly, low temperatures also have implications for reducing the hydrolysis rates of activated monomers and prolonging the half-life of ribozymes in the presence of divalent cations. (Attwater et al., 2010; Monnard and Ziock, 2008).

1.4.4 Oligomerization of cyclic nucleotides

Apart from ImpNs, another class of mononucleotides that have been hypothesized to be potential monomers for efficient synthesis of RNA is cyclic nucleotides. These nucleotides result from the formation of ring structures between the 5'-phosphate and 3' –OH groups or due to the presence of phosphate between 2' and 3' –OH groups. Such nucleotides are referred to as nucleoside 3',5'- and 2',3'- cyclic phosphates, respectively. The prebiotic plausibility of cyclic nucleotides has been explored in previous studies. Heating of nucleosides with KH_2PO_4, at high temperature in formamide produced cyclic 3',5'-phosphates (Saladino et al., 2009) and urea-catalyzed phosphorylation of nucleosides under dry heating conditions has shown to result in cyclic 2',3'-phosphates (Lohrmann and Orgel, 1971). A completely synthetic pathway, starting from simple molecules that also result in the formation of 2',3'- cyclic phosphates, was demonstrated by Sutherland and coworkers (Powner et al., 2009).

Oligomerization of purine 3',5'- cyclic phosphates was observed to be facilitated in water, at elevated temperatures (Costanzo et al., 2009). Long chains of oligomers (up to 25nt units) were claimed to have been formed in these reactions. Follow up studies have proposed a click chemistry like mechanism, catalyzed by the base in the reaction, for the oligomerization of 3',5'-cyclic phosphates (Costanzo et al., 2012). Oligomerization of adenosine cyclic 2',3'-phosphates was carried out in the presence of aliphatic amines as catalysts at alkaline pH, dry state and moderately high temperatures (Verlander et al., 1973). Oligomers up to hexamers were observed with an excess of 3'–5' bonds over 2'–5' bonds. Further analysis of the higher molecular weight products from similar reactions, which were carried on for about 40 days, revealed oligomers higher that hexamers in trace quantities (Verlander and Orgel, 1974).

1.4.5 Amphiphiles as organizing matrices

Amphiphiles such as fatty acids and their derivatives, and phospholipid-like molecules are thought to have been present on the early Earth. Fischer-Tropsch-type reactions may have

resulted in the synthesis of mixed fatty acids in deep vent niches while amphiphilic compounds have been shown to be present in meteorites and carbonaceous chondrites (Deamer and Pashley, 1989). The ability of amphiphiles to form higher order assemblies, such as membranes and vesicles, has fundamental implications for their role as plausible early compartments during the emergence of cellular life. Studies have shown the ability of vesicles to encapsulate hydrophilic solutes (similar to nutrient uptake in modern cells), and even facilitate catalytic reactions (Monnard and Deamer, 2002). Apart from their obvious role as compartments, lipids are also hypothesized to have helped play a role in RNA oligomerization. When solutions containing lipid vesicles are subjected to dehydration, the lipid molecules form multilamellar structures that have alternating hydrophilic and hydrophobic spaces (Deamer, 2012). Such arrangement results in the formation of matrix-like structures, which can be effective in organizing solute molecules in two-dimensional spaces while still permitting some diffusion as depicted in Figure 1.5.

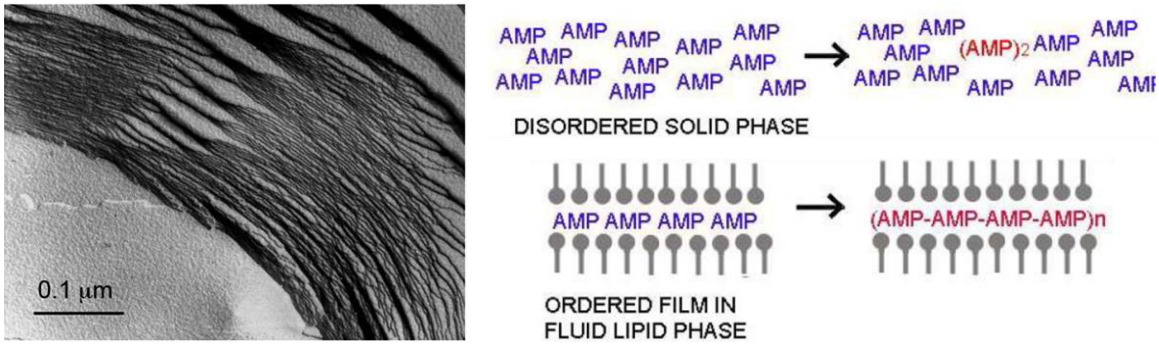


Figure 1.5: Freeze fracture microscopy of phospholipid vesicles revealed multilamellar structures with each line corresponding to a lipid bilayer about 5 nm thick. Monomers (such as AMP) were hypothesized to get organized within the bilayers as illustrated. [Adapted from (Deamer, 2012)]

An early study by Deamer's group looked at the oligomerization of non-activated nucleoside monophosphates, in the presence of lipid vesicles, by using dehydration-rehydration as a driving force for the reaction (Rajamani et al., 2008). Dry-wet cycles are considered to have been a recurring theme on prebiotic Earth, driven by day-night variations or seasonal fluctuations. The hypothesis is that during the dehydration phase, mononucleotides are trapped in the lipid matrix and polymerization takes place due to reduced activity of water. Subsequent rehydration enables vesicle formation from the multilamellar sandwiches and can result in encapsulation of these newly formed oligomers, resulting in the formation of protocell like entities. Although some

hydrolysis of phosphodiester bonds will take place during rehydration, repeated cycling acts as a kinetic trap, favoring uphill reactions and thus resulting in an accumulation of oligomers over multiple cycles. Detailed findings from this study, and other studies that have explored this idea further, are discussed in Section 2.1.

1.5 Objectives

The questions pertaining to the area of ribonucleotide polymerization that yields RNA-like oligomers are still expansive. Studying these nonenzymatic oligomerization reactions involves dissecting/narrowing down multiple components of the reaction, such as the type of monomers (mono/polyphosphates, imidazolites or cyclic phosphates etc.), conditions that facilitate polymerization, clay/mineral catalysis and the role of divalent cations. Most studies to date have looked at the polymerization of activated nucleotides due to their faster reaction rates. Although the synthesis of various types of monomers has been chemically demonstrated, the presence of large quantities of activated nucleotides on the prebiotic Earth is questionable. The problem of the short half-lives of activated nucleotides, under harsh conditions that were likely to have been present on the early Earth, further complicates their relevance as prebiotically viable monomers. Furthermore, most oligomerization related studies have looked at the mononucleotides in isolation, an unrealistic scenario given the heterogeneous nature of the prebiotic milieu, excepting a few studies that have used dipeptides or amphiphiles.

In this project, we have focused on discerning the oligomerization of nucleoside monophosphates in the presence of lipids, under dehydration-rehydration regimes. Dry-wet cycles are not only seen in modern environments but are also very feasible on early Earth. Nucleoside monophosphates lack high energy bonds or good leaving groups. Nonetheless, they can be prebiotically synthesized, which makes them a pertinent candidate for oligomerization under harsh volcanic geothermal conditions of high temperatures and low pH. To begin with, we carried out similar reactions as reported by Deamer group (Rajamani et al., 2008). Our aim was to standardize the synthesis of RNA-like oligomers that would allow for their systematic biochemical characterization. Towards this, optimum parameters for their oligomerization were determined by exploring the parameter space in terms of pH, temperature, duration of the dehydration phase etc. Subsequent analysis of the resultant oligomers was undertaken by various biochemical techniques to discern the chemical nature of these oligomers. The results from the

aforementioned studies and the implications for the presence of lipids in oligomerization reaction are discussed in the next chaper.

1.6 References

- Adamala, K., and Szostak, J.W. (2013). Competition between model protocells driven by an encapsulated catalyst. Nat. Chem. 5, 495–501.
- Attwater, J., Wochner, A., Pinheiro, V.B., Coulson, A., and Holliger, P. (2010). Ice as a protocellular medium for RNA replication. Nat. Commun. 1, 76.
- Bada, J.L., and Lazcano, A. (2002). Some Like It Hot, But Not the First Biomolecules. Science. 296, 1982 LP-1983.
- Bartel, D.P., and Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. Science. 261, 1411 LP-1418.
- Bartel, D.P., Muller, U.F., Müller, U.F., and Bartel, D.P. (2008). Improved polymerase ribozyme efficiency on hydrophobic assemblies. RNA 14, 552–562.
- Bernhardt, H.S., and Tate, W.P. (2012). Primordial soup or vinaigrette: did the RNA world evolve at acidic pH? Biol. Direct 7, 4.
- Bradley, B., Matthew, P., Maheen, G., J., C.B., Francisco, V., V., H.N., and César, M.-S. (2016). Darwin's Warm Little Pond: A One-Pot Reaction for Prebiotic Phosphorylation and the Mobilization of Phosphate from Minerals in a Urea-Based Solvent. Angew. Chemie 128, 13443–13447.
- Carneiro, C.E. a, Berndt, G., de Souza Junior, I.G., de Souza, C.M.D., Paesano, A., da Costa, A.C.S., di Mauro, E., de Santana, H., Zaia, C.T.B. V, and Zaia, D. a M. (2011). Adsorption of adenine, cytosine, thymine, and uracil on sulfide-modified montmorillonite: FT-IR, Mössbauer and EPR spectroscopy and X-ray diffractometry studies. Orig. Life Evol. Biosph. 41, 453–468.
- Carter, C.W. (2015). What RNA World? Why a Peptide/RNA Partnership Merits Renewed Experimental Attention. Life (Basel, Switzerland) 5, 294–320.
- Chen, I. a, and Szostak, J.W. (2004). A kinetic study of the growth of fatty acid vesicles. Biophys. J. 87, 988–998.
- Childers, W.S., Ni, R., Mehta, A.K., and Lynn, D.G. (2009). Peptide membranes in chemical evolution. Curr. Opin. Chem. Biol. 13, 652–659.

- Cody, G.D., Boctor, N.Z., Filley, T.R., Hazen, R.M., Scott, J.H., Sharma, A., and Yoder, H.S. (2000). Primordial Carbonylated Iron-Sulfur Compounds and the Synthesis of Pyruvate. Science. 289, 1337 LP-1340.
- Costanzo, G., Pino, S., Ciciriello, F., and Di Mauro, E. (2009). Generation of long RNA chains in water. J. Biol. Chem. 284, 33206–33216.
- Costanzo, G., Saladino, R., Botta, G., Giorgi, A., Scipioni, A., Pino, S., and Di Mauro, E. (2012). Generation of RNA Molecules by a Base-Catalysed Click-Like Reaction. ChemBioChem 13, 999–1008.
- Damiano, L., and Luisi, P.L. (2010). Towards an autopoietic redefinition of life. Orig. Life Evol. Biosph. 40, 145–149.
- Deamer, D. (2012). Liquid crystalline nanostructures: organizing matrices for non-enzymatic nucleic acid polymerization. Chem. Soc. Rev. 41, 5375–5379.
- Deamer, D.W., and Barchfeld, G.L. (1982). Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. J. Mol. Evol. 18, 203–206.
- Deamer, D.W., and Pashley, R.M. (1989). Amphiphilic components of the murchison carbonaceous chondrite: Surface properties and membrane formation. Orig. Life Evol. Biosph. 19, 21–38.
- Dickson, K.S., Burns, C.M., and Richardson, J.P. (2000). Determination of the Free-Energy Change for Repair of a DNA Phosphodiester Bond. J. Biol. Chem. 275, 15828–15831.
- Engelhart, A.E., Powner, M.W., and Szostak, J.W. (2013). Functional RNAs exhibit tolerance for non-heritable 2’-5’ versus 3’-5’ backbone heterogeneity. Nat. Chem. 5, 390–394.
- de Farias, T.S., Rêgo, G.T., and José, V.M. (2016). tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. Life. 6 (2). 15
- Ferré-D’Amaré, A.R., and Scott, W.G. (2010). Small self-cleaving ribozymes. Cold Spring Harb. Perspect. Biol. 2, a003574–a003574.
- Ferris, J.P., and Ertem, G. (1992). Oligomerization of ribonucleotides on montmorillonite: reaction of the 5’-phosphorimidazolide of adenosine. Science. 257, 1387 LP-1389.
- Ferris, J.P., Ertem, G., and Agarwal, V. (1989). Mineral catalysis of the formation of dimers of 5′-AMP in aqueous solution: The possible role of montmorillonite clays in the prebiotic synthesis of RNA. Orig. Life Evol. Biosph. 19, 165–178.

- Forsythe, J.G., Yu, S., Mamajanov, I., Grover, M.A., Krishnamurthy, R., Fernµndez, F.M., and Hud, N. V (2015). Ester-Mediated Amide Bond Formation Driven by Wet – Dry Cycles : A Possible Path to Polypeptides on the Prebiotic Earth. Angew. Chem. Int. Ed. 54, 9871–9875.
- Gilbert, W. (1986). The RNA world. Nature 319, 618.
- Glass, J.I., Merryman, C., Wise, K.S., Hutchison, C.A., and Smith, H.O. (2017). Minimal Cells—Real and Imagined. Cold Spring Harb. Perspect. Biol. 9.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35, 849–857.
- Hanczyc, M.M., Fujikawa, S.M., and Szostak, J.W. (2003). Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division. Science. 302, 618–622.
- Ibanez, J.D., Kimball, A.P., and Oro, J. (1971). Possible Prebiotic Condensation of Mononucleotides by Cyanamide. Science. 173, 444 LP-446.
- Johnson, A.P., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Lazcano, A., and Bada, J.L. (2008). The Miller Volcanic Spark Discharge Experiment. Science. 322, 404.
- Joshi, P.C., Aldersley, M.F., Delano, J.W., and Ferris, J.P. (2009). Mechanism of Montmorillonite Catalysis in the Formation of RNA Oligomers. J. Am. Chem. Soc. 131, 13369–13374.
- Joyce, G.F., Visser, G.M., van Boeckel, C.A., van Boom, J.H., Orgel, L.E., and van Westrenen, J. (1984). Chiral selection in poly(C)-directed synthesis of oligo(G). Nature 310, 602–604.
- Kanavarioti, A., Monnard, P.-A., and Deamer, D.W. (2001). Eutectic Phases in Ice Facilitate Nonenzymatic Nucleic Acid Synthesis. Astrobiology 1, 271–281.
- Kawamura, K. (2004). Behaviour of RNA under hydrothermal conditions and the origins of life. Int. J. Astrobiol. 3, 301–309.
- Kitadai, N., and Maruyama, S. (2017). Origins of building blocks of life: A review. Geosci. Front. 9, 1117–1153.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell 31, 147–157.
- Lohrmann, R. (1977). Formation of nucleoside 5′-phosphoramidates under potentially prebiological conditions. J. Mol. Evol. 10, 137–154.

- Lohrmann, R., and Orgel, L.E. (1971). Urea-Inorganic Phosphate Mixtures as Prebiotic Phosphorylating Agents. Science. 171, 490 LP-494.
- Mamajanov, I., Macdonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Ferna, F.M., Grover, M.A., Hud, N. V, et al. (2014). Ester Formation and Hydrolysis during Wet − Dry Cycles : Generation of Far-from-Equilibrium Polymers in a Model Prebiotic Reaction. Macromolecules 47, 1334–1343.
- Martin, W., and Russell, M.J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 358, 59 LP-85.
- Maurer, S.E., Deamer, D.W., Boncella, J.M., and Monnard, P.-A. (2009). Chemical Evolution of Amphiphiles: Glycerol Monoacyl Derivatives Stabilize Plausible Prebiotic Membranes. Astrobiology 9, 979–987.
- Mccollom, T.M., Ritter, G., and Simoneit, B.R.T. (1999). Lipid synthesis under hydrothermal conditions by fischer-tropsch-type reactions. Orig. Life Evol. Biosph. 29, 153–166.
- Menor-Salván, C., and Marín-Yaseli, M.R. (2012). Prebiotic chemistry in eutectic solutions at the water-ice matrix. Chem. Soc. Rev. 41, 5404–5415.
- Miller, S.L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. Science. 117, 528 LP-529.
- Miyakawa, S., Joshi, P.C., Gaffey, M.J., Gonzalez-Toril, E., Hyland, C., Ross, T., Rybij, K., and Ferris, J.P. (2006). Studies in the mineral and salt-catalyzed formation of RNA oligomers. Orig. Life Evol. Biosph. 36, 343–361.
- Monnard, P.-A., and Deamer, D.W. (2002). Membrane self-assembly processes: Steps toward the first cellular life. Anat. Rec. 268, 196–207.
- Monnard, P.-A., and Szostak, J.W. (2008). Metal-ion catalyzed polymerization in the eutectic phase in water-ice: a possible approach to template-directed RNA polymerization. J. Inorg. Biochem. 102, 1104–1111.
- Monnard, P., and Ziock, H. (2008). Eutectic Phase in Water-Ice: A Self-Assembled Environment Conducive to Metal-Catalyzed Non-Enzymatic RNA Polymerization. Chem. Biodivers. 5, 1521–1539.

- Monnard, P.-A.A., Kanavarioti, A., and Deamer, D.W. (2003). Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. J. Am. Chem. Soc. 125, 13734–13740.
- Muchowska, K.B., Varma, S.J., Chevallot-Beroux, E., Lethuillier-Karl, L., Li, G., and Moran, J. (2017). Metals promote sequences of the reverse Krebs cycle. Nat. Ecol. Evol. 1, 1716—1721.
- Murtas, G., Kuruma, Y., Bianchini, P., Diaspro, A., and Luisi, P. (2007). Protein synthesis in liposomes with a minimal set of enzymes.
- Namani, T., and Deamer, D.W. (2008). Stability of model membranes in extreme environments. Orig. Life Evol. Biosph. 38, 329–341.
- Neveu, M., Kim, H.-J., and Benner, S.A. (2013). The "Strong" RNA World Hypothesis: Fifty Years Old. Astrobiology 13, 391–403.
- Orgel, L.E. (1998). The origin of life--a review of facts and speculations. Trends Biochem. Sci. 23, 491–495.
- Orgel, L.E. (2000). Self-organizing biochemical cycles. Proc. Natl. Acad. Sci. 97, 12503 LP-12507.
- ORÓ, J. (1961). Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. Nature. 191, 1193.
- Powner, M.W., Gerland, B., and Sutherland, J.D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature. 459, 239–242.
- Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig. Life Evol. Biosph. 38, 57–74.
- Rode, B.M. (1999). Peptides and the origin of life. Peptides 20, 773–786.
- Rodriguez-Garcia, M., Surman, A.J., Cooper, G.J.T., Suárez-Marina, I., Hosni, Z., Lee, M.P., and Cronin, L. (2015). Formation of oligopeptides in high yield under simple programmable conditions. Nat. Commun. 6, 8385.
- Russell, M.J., Daniel, R.M., Hall, A.J., and Sherringham, J.A. (1994). A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life. J. Mol. Evol. 39, 231–243.

- Saladino, R., Crestini, C., Ciciriello, F., Pino, S., Costanzo, G., and Di Mauro, E. (2009). From formamide to RNA: the roles of formamide and water in the evolution of chemical information. Res. Microbiol. 160, 441–448.
- Saladino, R., Carota, E., Botta, G., Kapralov, M., and Timoshenko, G.N. (2015). Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. Proc. Natl. Acad. Sci. U. S. A. E2746–E2755.
- Sawai, H. (1976). Catalysis of internucleotide bond formation by divalent metal ions. J. Am. Chem. Soc. 98, 7037–7039.
- Sawai, H., and Orgel, L.E. (1975). Oligonucleotide synthesis catalyzed by the Zn^{2+} ion. J. Am. Chem. Soc. 97, 3532–3533.
- Sawai, H., Higa, K., and Kuroda, K. (1992). Synthesis of cyclic and acyclic oligocytidylates by uranyl ion catalyst in aqueous solution. J. Chem. Soc. Perkin Trans. 1 505–508.
- Le Son, H., Suwannachot, Y., Bujdak, J., and Rode, B.M. (1998). Salt-induced peptide formation from amino acids in the presence of clays and related catalysts. Inorganica Chim. Acta 272, 89–94.
- Trefil, J., Morowitz, H.J., and Smith, E. (2009). The Origin of Life: A case is made for the descent of electrons. Am. Sci. 97, 206–213.
- Urata, H., Fujimori, M., Aono, C., Yamakawa, T., Harada, E., and Akagi, M. (2008). Regio- and diastereo-selectivity of montmorillonite-catalyzed oligomerization of racemic adenosine 5'-phosphorimidazolide. Nucleosides. Nucleotides Nucleic Acids 27, 421–430.
- Varma, S.J., Muchowska, K.B., Chatelain, P., and Moran, J. (2018). Native iron reduces CO2 to intermediates and end-products of the acetyl-CoA pathway. Nat. Ecol. Evol. 2, 1019–1024.
- Verlander, M.S., and Orgel, L.E. (1974). Analysis of high molecular weight material from the polymerization of adenosine cyclic 2′, 3′-phosphate. J. Mol. Evol. 3, 115–120.
- Verlander, M.S., Lohrmann, R., and Orgel, L.E. (1973). Catalysts for the Self-Polymerization of Adenosine Cyclic 2 ', 3 ' -Phosphate. J. Mol. Evol. 6, 303–316.
- Vlassov, A. V, Kazakov, S. a, Johnston, B.H., and Landweber, L.F. (2005). The RNA world on ice: a new scenario for the emergence of RNA information. J. Mol. Evol. 61, 264–273.
- Wächtershäuser, G. (1988). Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452–484.

- Wieczorek, R.R., Dörr, M., Chotera, A., Luisi, P.L., and Monnard, P.-A.A. (2013). Formation of RNA Phosphodiester Bond by Histidine-Containing Dipeptides. ChemBioChem 14, 217–223.
- Wu, T., and Orgel, L.E. (1992). Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. J. Am. Chem. Soc. 114, 317–322.
- Zagorevskii, D. V, Aldersley, M.F., and Ferris, J.P. (2006). MALDI analysis of oligonucleotides directly from montmorillonite. J. Am. Soc. Mass Spectrom. 17, 1265–1270.
- Zhang, S., Blain, J.C., Zielinska, D., Gryaznov, S.M., and Szostak, J.W. (2013). Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. Proc. Natl. Acad. Sci. U. S. A. 110, 17732–17737.
- Zhu, T.F., and Szostak, J.W. (2009). Coupled Growth and Division of Model Protocell Membranes. J. Am. Chem. Soc. 131, 5705–5713.

Chapter 2

Standardization of Oligomerization using DH-RH Reactions

2.1 Introduction

Very few studies in the past have focused on the process of polymerization of non- activated nucleotides that result in nucleic acids. As discussed in the previous chapter, a study in 2008 discussed lipid-assisted polymerization of nucleotides under DH-RH conditions. The study reported polymerization of nucleotides by use of techniques such as Nanopore sequencing, RiboGreen assays and gel analysis (Rajamani et al., 2008). Samples from the reactions carried out showed blockades during Nanopore analysis, which were characteristic of an anionic polymer. Yields of the resultant polymers increased with cycling, higher temperatures and seemed to be affected by the lipid that was being used. Radiolabelled gel analyses resulted in polymer streaks which seemed to correspond to 25-75 nucleotide long oligomers. The gel analysis also confirmed the need for, both, lipids and cycling for the polymer formation to occur. Microscopy of the lipid vesicles revealed the stability of various lipids to cycling conditions and also showed potential encapsulation of oligomers in the vesicles. Preliminary HPLC analysis carried out in this study did reveal oligomerization. However, the authors claimed that the longer oligomers were too low in concentration and too variable in their composition to be detected by HPLC. The study also recommended larger scale preparations in order to generate oligomers in sufficient amounts for purification and detailed characterization. In all, this paper demonstrated for the first time the possibility of non-enzymatic polymerization of non-activated nucleotides in dehydrated lipid matrices.

Subsequent studies by Dr. Deamer's group presented proof-of-concept results pertaining to nonenzymatic information transfer in DNA, under dry-wet cycle conditions (Olasagasti et al., 2011). The products from this reaction were sequenced to determine the misincorporation rate and the fidelity of information transfer. The reaction resulted in low yields (about 0.5% with respect to the template), and in the absence of the template, the reaction products were below detection limit. The authors proposed a possible reaction mechanism for this acid-catalyzed esterification reaction, as showed in Figure 2.1A. Another study looked at the organization of nucleotide monomers in lipid crystalline matrices that are formed during the dehydration phase (Toppozini et al., 2013). When X-ray diffraction was carried out on the dried lipid phases, with nucleotides confined between the bilayers, the organization of monomers showed proximity of the phosphate groups to the 3' position of the ribose moiety. This manner of confinement of the AMP molecules was proposed to favor formation of RNA-like molecules that was seen in the

earlier studies. These findings were later confirmed in a study which reported two distinct organization of monomers (AMP, UMP and mixture), with 4.6 Å and 3.4 Å distances between consecutive nucleotides (Himbert et al., 2016). While the 4.6 Å distance was assigned to the formation of a disordered, glassy structure, the smaller 3.4 Å distance was thought to correspond to nucleobases that are stacked in an RNA-like configuration (Figure 2.1B). This kind of non-covalent arrangement was referred to as pre-polymers, which are linear stacks of monomers with phosphate groups and ribose in close proximity. When neutron scattering analysis was performed on AMP confined in lipid multilayers at very low levels of hydration, limited diffusion of AMP molecules was observed (Misuraca et al., 2017). This limited diffusion was thought to be the driving force in the formation of ordered structures that were observed in aforementioned studies. On increasing the hydration to about 35%, AMP molecules were able to diffuse freely within the bilayer and further hydration was thought to result in vesicle formation and encapsulation of the resultant oligomers in these vesicles.

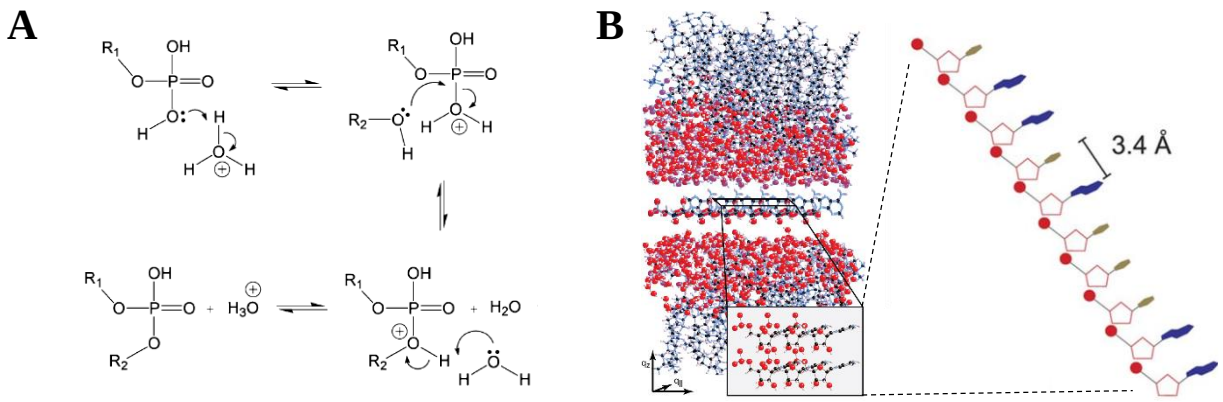


Figure 2.1: Possible mechanism for phosphodiester bond formation during lipid-assisted nonenzymatic polymerization (A). Disordered and ordered arrangement of monomers in dehydrated lipid bilayers as determined by X-ray scattering. (Adapted from Da Silva et. al., 2015, Toppozini et. al., 2013, and Himbert et.al., 2016 respectively.)

Another study in 2014 discussed the stability of duplexes under DH-RH cycles and the possibility for base pairing in the oligomers formed under these conditions (DeGuzman et al., 2014). Typically in DH-RH reactions, the rehydration step may induce hydrolysis of RNA-like oligomers and result in a loss of their structure forming ability. In order to verify the effects of cycling, intercalating dyes were used to check the stability of PolyA-PolyU duplexes under low pH (pH 3) and high temperature (80°C). Streaks were observed on electrophoresis after four DH-

RH cycles of 30 mins each, which was interpreted as an ability of the polymers to form duplexes even after cycling. This study also claimed oligomerization of AMP & UMP resulting in RNA-like polymers, which were analyzed by Nanopore sequencing and also using ethidium bromide containing gels. The authors also conducted hyperchromicity analysis of the products formed in the DH-RH reactions. Products form AMP+UMP reactions showed similar hyperchromicity curves to that of control PolyA+PolyU mixtures as seen in Figure 2.2. However, reannealing of these products was not observed on cooling of the samples, which was attributed to presence of dilute and possibly random sequences that might have been generated over the course of the DH-RH reaction.

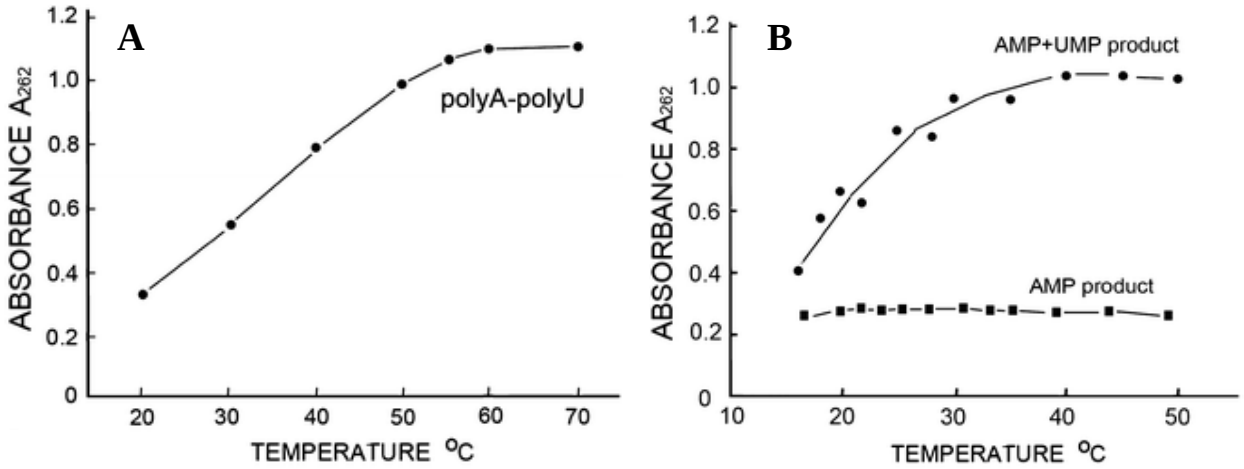


Figure 2.2: Hyperchromicity analysis of control PolyA+PolyU and reaction products from AMP and AMP+UMP reactions. (Adapted from Deguzman et. al 2014.)

Although all the aforementioned studies have explored the ability of monomers to undergo polymerization in the presence of lipids, most studies have used analytical techniques that do not give insights about the chemical nature of the resultant polymers. Furthermore, several side reactions can take place in the high temperature conditions, which may significantly alter the reaction products. Thermal degradation of sugars and the stability of various bonds in the nucleotides, at these high temperatures, are few of the side reactions that would diminish the pool of the starting nucleotides. These possibilities can be discerned by careful chemical analysis of the reaction products, which was missing in the previous studies. To facilitate and achieve this goal, this study involved the scaling up of these reactions, and standardizing the analysis of the products using rigorous biochemical techniques such as HPLC and mass spectrometry.

Additionally, we also systematically studied the effects of various parameters such as pH, temperature, drying time and lipids:nucleotide ratio, on the yield of the reaction and on the product distribution in these reactions.

2.2 Materials and Methods

2.2.1 Materials

Adenosine 5'-monophosphate (AMP) was purchased as disodium salt from Sigma-Aldrich (Bangalore, India) and used without further purification. Phospholipids, including 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC), were purchased from Avanti Polar Lipids Inc (Alabaster, AL, USA). All lipid preparations were extruded through 200nm polycarbonate membranes using Avanti mini extruder. Other chemicals used for analysis, such as acids, Tris buffer, $NaClO_4$ were of analytical grade and purchased from Sigma-Aldrich (Bangalore, India).

Methods

2.2.2 Prebiotic Simulator fabrication and standardizing reaction conditions

To carry out reactions under prebiotically relevant conditions, fabrication of a set up was needed that could be used to simulate high temperatures and an anaerobic environment. Since no suitable readymade instrumentation was available that could enable carrying out relevant reactions, we assembled our own set up by modifying a bench-top heating block that could maintain the requisite temperature. Figure 2.3 shows various parts of the set up and the complete set up, when in use. The heating block was fitted with an adaptor to accommodate several 20ml glass vials that contained the reaction mixtures. A typical reaction mixture consisted of mononucleotides and lipids that were kept under high temperature acidic conditions; the precise detail of each reaction has been mentioned in individual results. Typically, reaction vials were secured with caps that had PTFE septa in them, which were purchased from Chemglass (Vineland, NJ, USA). Through these, PEEK tubing of about 1–1.5 inches was inserted to deliver CO_2 into the vials. A second PEEK tube was inserted into the septum that served as an escape vent for air, initially, and then excess CO_2, and also for the water vapor that is generated during the reaction. The heating block was maintained at a specified temperature and the reaction mixtures were dried under CO_2. After the drying phase was completed, rehydration was

performed by introducing rehydrating agents through a third PEEK tubing that were attached to syringes. Multiple DH-RH cycles were performed to study the effect of different parameters on lipid-assisted polymerization reaction of ribonucleotides, and reaction mixtures were analyzed after several cycles using pertinent biochemical techniques. Various parameters such as pH, temperature, drying time etc. were varied, and the reactions were analyzed to see how these parameters would affect polymerization.

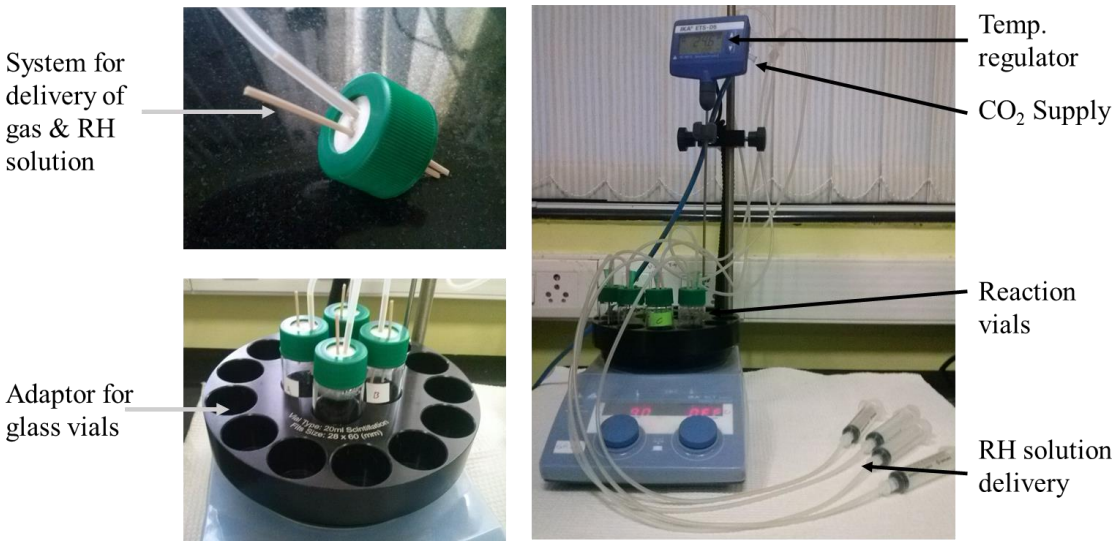


Figure 2.3: Image of prebiotic simulator used in the experiments.

2.2.3 Analysis of the reactions using High Performance Liquid Chromatography (HPLC)

Typically, the reaction samples contain both nucleotides and lipids. Since lipids could interfere with the analysis, the aqueous phase of the reaction mixtures was obtained by performing butanol-hexane extraction to remove the phospholipids from the samples (Rajamani et al., 2008). HPLC analysis was subsequently performed on the aqueous phase to analyze for oligomer formation. Chromatography was performed using an Agilent 1260 chromatography system (Agilent Technologies, Santa Clara, CA, USA) and DNAPac PA200 column (4 × 250 mm) from Dionex (now Thermo Scientific, Sunnyvale, CA, USA). Samples were analyzed with a linear gradient of $NaClO_4$ in 2 mM Tris buffer at pH 8, using a flow rate of 1 mL/min. All solvents, purchased from Sigma-Aldrich (Bangalore, India), were of HPLC grade and used after filtering through a 0.22-µm nylon filter, followed by degassing. Samples were detected using a high-sensitivity flow cell (60 mm path length) in a diode array detector. For the graphical representation of chromatograms, area under each peak was obtained using Agilent EZChrome

software and normalized such that addition of all the peaks totaled to100. Each peak was then assigned a normalized area based on what percentage of the total area calculated was represented by it. These numbers (the normalized percentage areas) were then plotted as a histogram to compare the product distribution in the different reactions. A minority of small peaks which did not correspond to any known product were neglected to avoid any overestimation of yields.

2.2.4 Mass analysis of the reaction mixtures

In order to confirm the chemical nature of the products, the reaction mixtures were also analyzed by mass spectrometry. Mass analysis was performed using Matrix-Assisted Laser Desorption Ionization (MALDI) and High-Resolution Mass Spectrometry (HRMS). MALDI was performed on aqueous phase samples, using 2′,4′,6′-Trihydroxyacetophenone (THAP) as the matrix, on 4800 Plus MALDI TOF/TOF™ Analyzer (AB SCIEX, Framingham, MA, USA). Prior to the analysis, attempts were made to reduce salt in the HPLC purified samples of the reaction mixtures by using C_{18} Zip-Tips (Merck KGaA, Darmstadt, Germany). HRMS analysis was carried out using Acquity UPLC+ system from Waters, with an Alltima C18 column (2 μm, 2.1 × 50 mm), using a water/acetonitrile gradient containing 0.1% formic acid. Mass determination was carried out in the positive ion mode with SYNAPT G2 Mass Spectrometry (Waters, Milford, CT, USA).

2.2.5 Microscopy

The organic phase, in which the lipids were extracted using the butanol-hexane procedure as mentioned in 2.2.3, was used for analyzing the stability of vesicles over multiple DH-RH cycles. This was investigated as the ability of these lipids to form vesicles could be lost due to oxidation or other adverse reactions that could potentially take place at high temperature and low pH. For this analysis, appropriate amounts of butanol-hexane phase extracted samples were dried on a clean grease-free microscopy slide. Subsequently, the dry spot was rehydrated using 8-10 μl of water, mixed thoroughly and a coverslip was placed gently on this drop without any air bubbles. The slide was then sealed to avoid evaporation. Differential Interference Contrast (DIC) microscopy was used to visualize and capture the images of vesicles formed by the extracted lipids on Zeiss Imager.Z1 microscope. Most of the images were captured at 40x magnification unless mentioned otherwise.

2.3 Results and Discussions:

2.3.1 HPLC analysis of pilot reactions

The aqueous phase samples were subjected to preliminarily analysis using HPLC. The anion exchange column DNAPac PA 200, which was used in this analysis, is reported to separate oligonucleotides (average 20mers) up to a single base resolution. The gradient of salt had to be standardized to separate small oligomers such as monomer, dimers etc. Mainly, AMP and hydrolyzed PolyA were used to standardize the salt gradient and to obtain well resolved individual peaks. A pilot cycling reaction was carried out with 5mM AMP and 1mM POPC, with low pH (~ 2) maintained using H_2SO_4. POPC was extruded prior to adding to the reaction mixture in order to obtain vesicles of uniform size. As mentioned in Section 2.2.1, multiple DH-RH cycles were carried out and samples were analyzed after 7 cycles to check for the oligomerization of AMP. Figure 2.4 shows a typical chromatogram of a sample, post removal of lipids, and as analyzed by HPLC. The black line in the chromatogram shows absorbance at 260nm and the green line shows the gradient of salt in the run. The first peak seems to elute in the dead volume (without any salt in the system) thus indicating a lack of interaction with the column matrix. This is possibly a breakdown product from the monomer (e.g. Adenine/ Adenosine) as it has an absorbance at 260 nm but no phosphate/charge that is crucial to interact with the column. The peak at 5.4 min corresponds to AMP as confirmed by control and co-injection runs. The subsequent peaks were that of higher oligomers as they eluted with higher salt concentration (confirmed subsequently using mass analysis as described in Section 2.3.5). The peaks for dimer and above showed clustering, which has been observed in other nonenzymatic oligomerization studies as well (Joshi et al., 2009). Such clustering indicates the presence of multiple isomeric species in each of these oligomers, which may differ in configuration or in the nature of their backbone linkages (2'-5' or 3'-5'). The resolution of peaks was lost after what corresponded to a possible tetramer peak, as shown in the figure inset, and this resulted in trailing of the HPLC signal. Reaction carried out in absence of any lipids showed smaller peaks for oligomers and increased breakdown (Figure 2.4, red trace). As several peaks were obtained in the gradient indicating oligomerization, we subjected the reactions mixtures to further characterization.

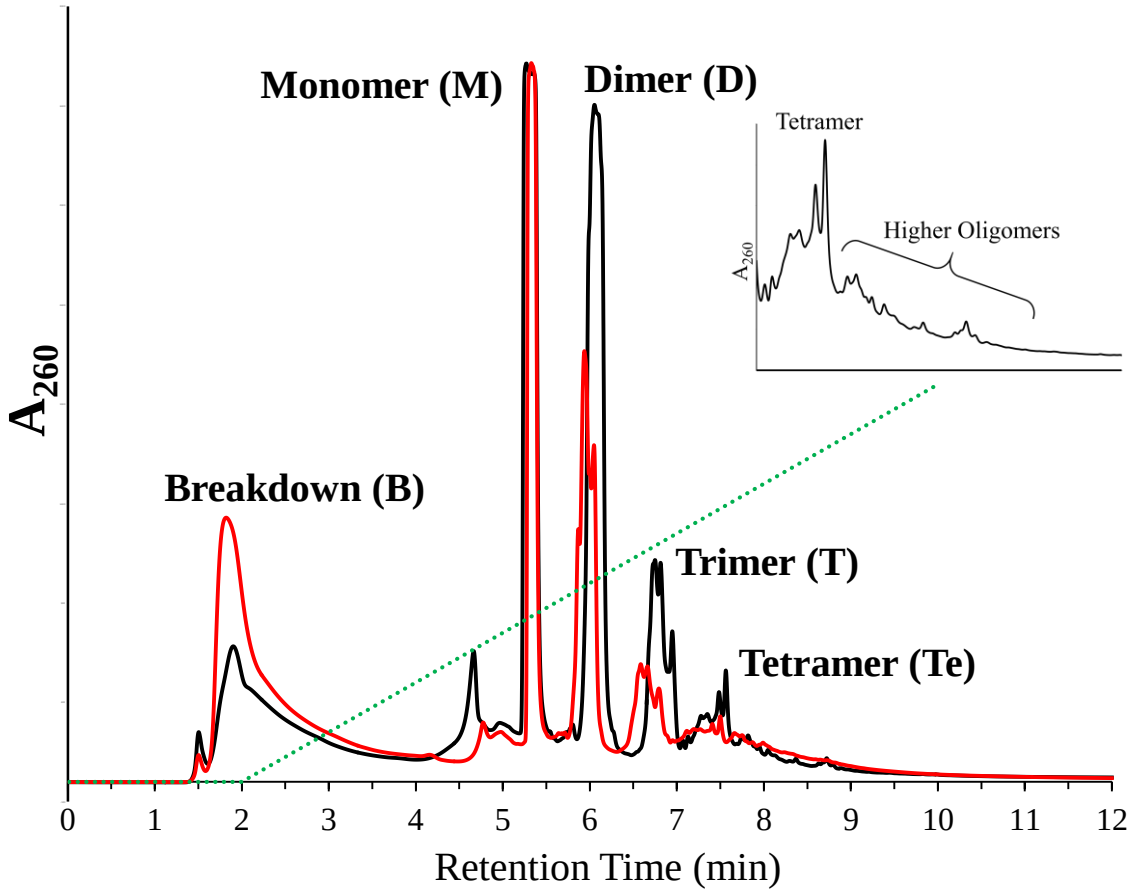


Figure 2.4: HPLC chromatogram of Cyc7 sample from a typical DH-RH reaction (at pH ~2) containing 5mM of 5'-AMP+ 1mM POPC (black trace) and 5mM of 5'-AMP reaction (red trace). The green trace indicates the gradient of salt that was used. The identities of the peaks have been indicated above each peak.

To check the effect of cycling on the reaction, multiple samples were collected after cycles 1, 3, 5 and cycle 7 and these were analyzed using HPLC, in a manner similar to the pilot reaction. Figure 2.5 A shows the comparison between the chromatograms from respective time points. As seen in the figure, oligomerization increases as the number of cycles is increased. However, the breakdown peak was also seen to increase substantially with increased cycling. In order to make sure that the higher oligomer peaks could be detected, more amount of aqueous sample had to be loaded on to the column. This, however, resulted in saturating peaks for, both, the Breakdown peak and the Monomer peak. Identity of the breakdown peak was uncertain as both adenine and adenosine tend to elute at the same time on the column (as was seen in control runs). As the extent of this loss of base/base + sugar was unclear, we decided not to calculate the yields of the

reactions from these chromatograms as it would always result in an underestimation of the same. Stability of the lipid component i.e. POPC, to the cycling conditions, was also analyzed using phase contrast microscopy. Samples were analyzed for the presence of vesicles as described in methods Section 2.2.5. As seen in Figure 2.5 B, all the cycles showed the presence of vesicles, which indicated the stability of POPC even under high temp. & low pH (~ 2) conditions. This was consistent with previous results, which showed that POPC and other longer phospholipids, with zwitter ionic head groups, were stable under DH-RH conditions (Rajamani et al., 2008). Further analysis was carried out to check the effects of various parameters on the reaction and the yield/distribution of the oligomers.

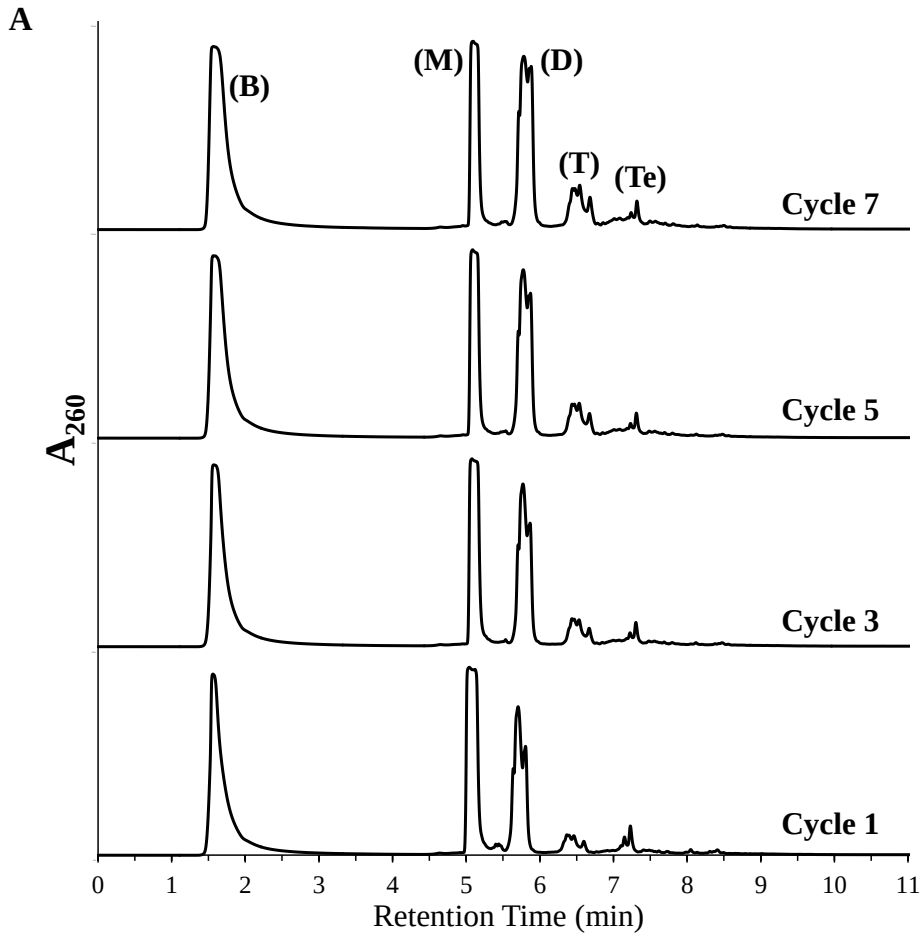
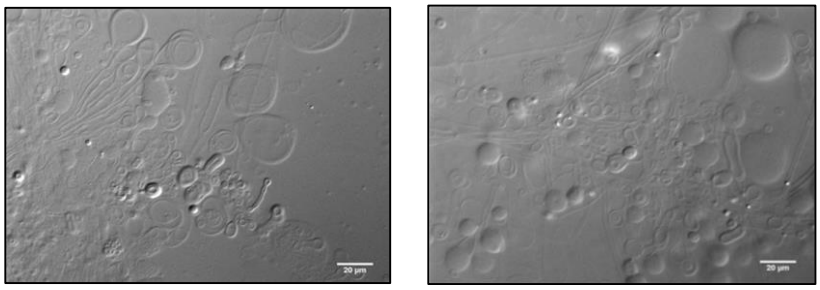


Figure 2.5 A: Formation of higher amounts of oligomers over cycles was observed using HPLC analysis, which corroborated previous results (Rajamani et al., 2008). Breakdown peak also increased with cycling suggesting that prolonged cycling did indeed promote unfavorable side reactions.

B



Cycle 1 **Cycle 7**

DIC Microscopy, 40x magnification; Scale bar = 20µm

Figure 2.5 B: Microscopy of the extracted lipids showed the presence of vesicles even after 7 DH-RH cycles (pH 2), indicating the stability of POPC under simulated prebiotic conditions.

2.3.2 Effect of variation of pH and rehydrating agent on the oligomerization process

Even though this reaction seemed to occur at low pH and high temperatures, in order to optimize the reaction for better oligomer yields, the effect of each of these parameters had to be delineated. We first looked at the optimum pH of the reaction and also how the acid component of the reaction affected the oligomer formation. Since the mechanism proposed for this type of polymerization is similar to that of acid-catalyzed esterification, we investigated the optimum pH conditions for polymerization of non-activated nucleotides. Mixtures of 5'-AMP, with and without lipids, were subjected to cycling at various pH values in the range of 1–4 and at intervals of 1 pH unit. Appropriate amount of H_2SO_4 was added to the reaction mixture to reach the required pH and subsequent rehydrations were carried out using Milli-Q water. Polymerization of 5'-AMP was assessed using HPLC analysis similar to what was described in the previous section. Figure 2.6 shows chromatograms that were obtained for cycling of 5'-AMP with DLPC at specified pH values. DLPC used in these reactions has similar chemical structure to POPC with slightly different fatty acid chains. Reactions whose pH was set above 2 did not result in good oligomer yields. Oligomerization was more efficient at a pH of 2 or below, as seen by the peaks corresponding to oligomers, which elute later along the salt gradient. A very low pH of 1, however, resulted in highest amount of breakdown peak that was generated. This was also evident from comparative analysis of these HPLC chromatograms at 280 nm (Figure 2.6, red traces). Comparison at this suboptimal wavelength was useful as the peaks were not saturated here and, therefore, were still useful in understanding the trends *per se*, which was the same at

both 260 and 280 nm. These results indicated pH 2 to be the optimum pH for polymerization of 5'-AMP, when all other conditions were kept the same. At this pH, the phosphate group in 5'-AMP is thought to be protonated, enabling nucleophilic attack by a neighboring nucleotide's 2'-OH or 3'-OH group, resulting in a phosphodiester bond (Olasagasti et al., 2011). A pH lower than 2 also facilitated this bond formation due to increased protonation, but was resulting in very high breakdown thus being counterproductive.

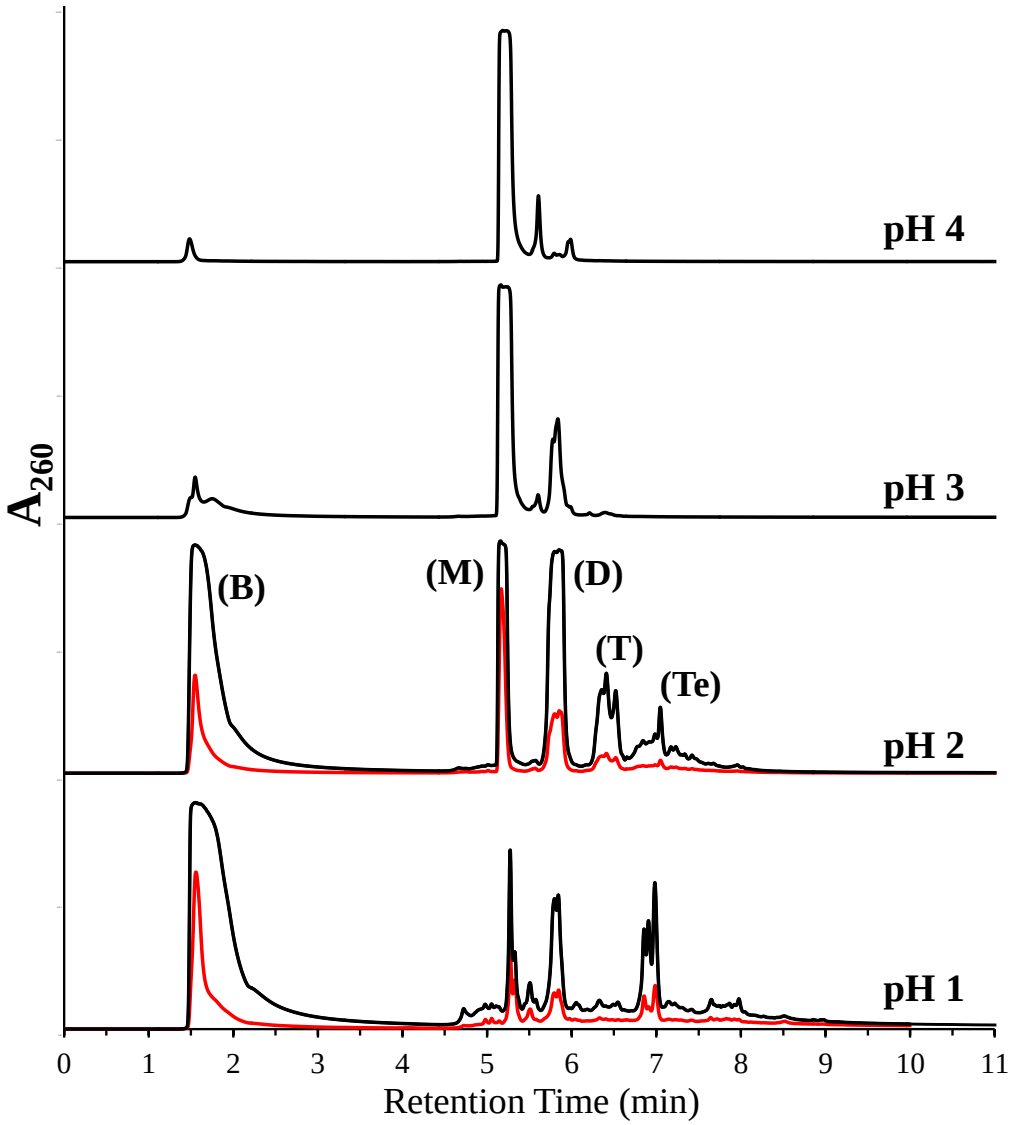


Figure 2.6: HPLC chromatograms obtained after seven DH-RH cycles of 5'-AMP (5mM) and DLPC (1mM), at specified pH values that were adjusted with H_2SO_4. Polymerization is seen in reactions at pH 2 and below. The red trace in pH2 and pH1 chromatograms indicates absorbance at 280 nm which shows increased breakdown at pH 1.

Since the previous experiments showed that pH 2 was optimum for polymerization, we next investigated if the acid used for lowering the pH had any effect on the polymerization process. In other words, is oligomerization favored by any acid that could bring the pH down to 2, or is it specific to only certain acids? To address this question, we carried out reactions with various mineral acids such as hydrochloric acid (HCl), sulfuric acid (H_2SO_4), nitric acid (HNO_3), orthophosphoric acid (H_3PO_4), and also with organic acids like formic acid and acetic acid. Some of these acids, e.g., nitric and formic acid, are known to be thermolabile and hence were replenished after every dehydration cycle in order to maintain the pH at ~2. As indicated in Figure 2.7, lesser polymerization was observed with HCl and HNO_3, whereas H_2SO_4 and H_3PO_4 both resulted in reasonable oligomer yields. However, the two organic acids tested failed to promote any polymerization despite repeated addition of these acids over multiple cycles. The ability of the acids tested to promote oligomerization seemed to be dependent on the nature of the acid. Thermolabile acids or acids with low boiling point (HNO_3, HCl, and the aforementioned organic acids) could not maintain low pH throughout the reaction, resulting in inefficient polymerization. H_2SO_4, which is also known to be a good dehydrating agent, was the most efficient in promoting oligomerization, followed by H_3PO_4. These results also indicated that acids may not play a direct role in the mechanism of polymerization but generally seemed to facilitate oligomerization by maintaining low pH that is necessary for the reaction to occur. Therefore, for all subsequent parameter characterizations, a pH of 2 was maintained using H_2SO_4 and rehydrations were carried out using water.

(Figure 2.7 is on the following page)

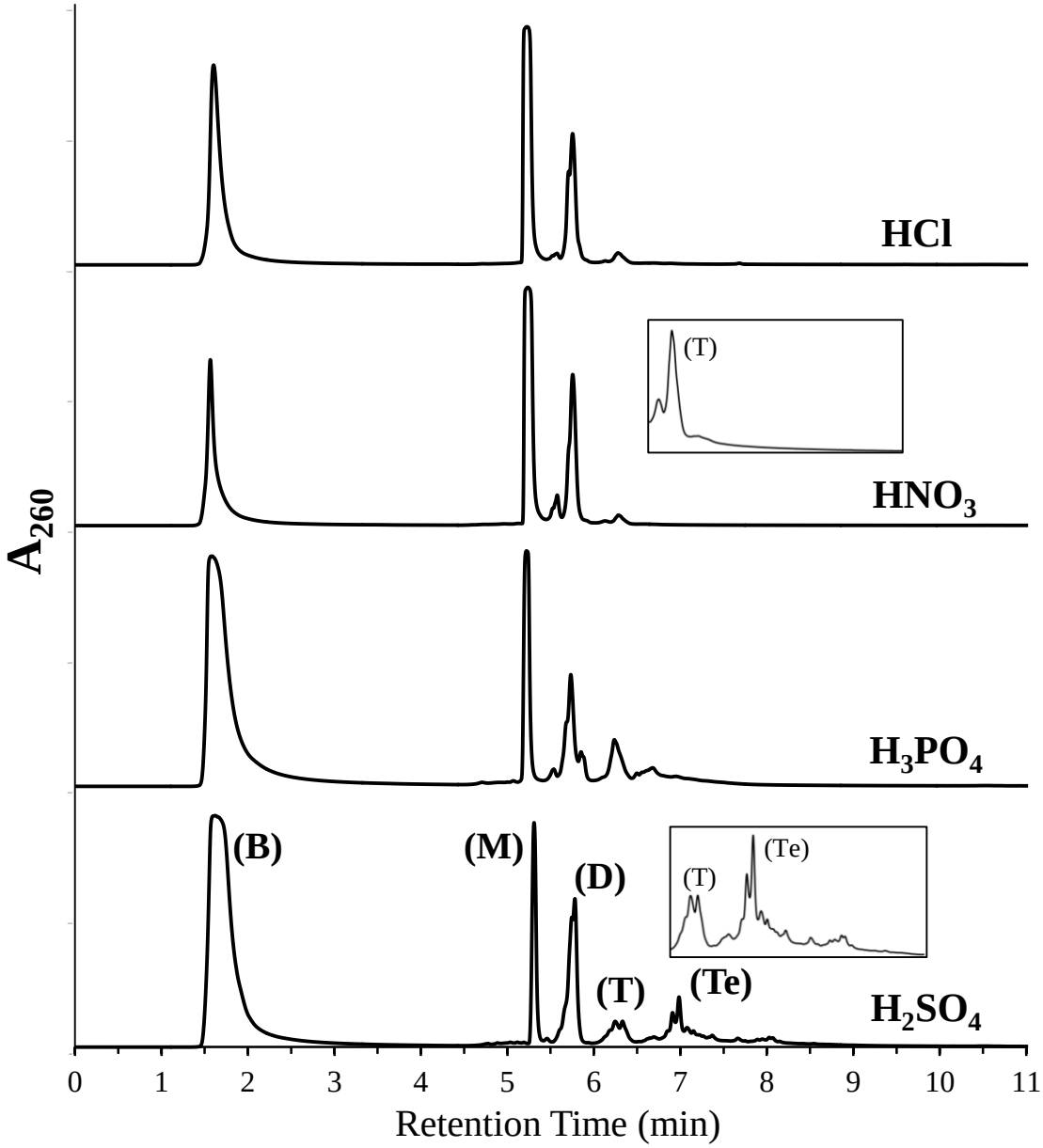


Figure 2.7: HPLC chromatograms obtained from Cycle 7 samples of 5'-AMP (5mM) and POPC (1mM) reactions performed with various acids, as indicated on each trace. Polymerization was observed in reactions with H_2SO_4 and H_3PO_4 while no products were formed when HCl, HNO_3 or organic acids were used as rehydrating agents.

2.3.3 Effect of temperature and dehydration time on oligomerization

There are two reactions, polymerization and hydrolysis, that are occurring alternatively during the cycling reaction. In these reactions, temperature plays a very important role in determining the rate of evaporation of water from the reaction. At lower temperatures, the water activity is more during the course of reaction as it is retained for a longer duration. Higher water activity will promote hydrolysis of newly formed oligomers, thus reducing oligomer yields. In contrast, high temperatures should result in higher yields as water will evaporate faster making hydrolysis less prevalent. It should, importantly, be noted that the starting material itself may undergo thermal decomposition at very high temperatures (e.g., charring of sugars). Therefore, there might be an optimum temperature range in which polymerization exceeds unfavorable reactions, resulting in better oligomer yields. To test this, cycling reactions were carried out at varying temperatures including at 60° C, 75°C, 90°C and 105°C. The resultant reaction mixtures from every alternate cycle (i.e. 1, 3, 5 and 7 cycles) were analyzed using HPLC. For effective visual comparison, the HPLC chromatograms are being represented graphically (Figure 2.8) by using the area as a proxy for oligomeric abundance in the different reactions (refer to Section 2.2.3). It is important to note that this area analysis does not reflect true oligomer yields due to technical concerns as discussed in Section 2.3.5.

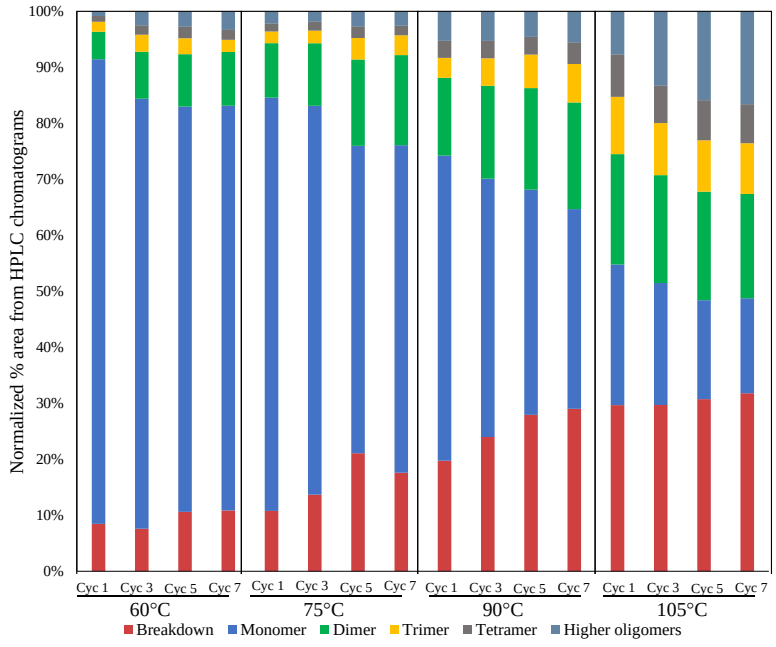


Figure 2.8: Comparison of polymerization reactions carried out at pH 2, over varying temperatures, using graphical representation of chromatograms obtained for the indicated DH-RH cycles. For all temperatures, more oligomerization was observed with higher cycles. Furthermore, greater extent of polymerization as well as more of the breakdown peak was observed with increase in temperature.

In 60 and 75°C reactions, overall less polymerization was observed, but the yields of oligomers still increased over cycles. In both cases, the breakdown peak was reduced but so was the extent of polymerization. The cycling reactions that were carried out at 90°C resulted in higher yields of oligomers even at initial cycles, which improved further with increasing number of DH-RH cycles. However, at the highest temperature of 105 °C, polymerization was observed as indicated by peaks after monomer but oligomer yields did not alter much with subsequent cycles. The breakdown peak observed was the highest here, possibly indicating loss of chemical integrity of the starting material. Comparison of cycle 7 from the DH-RH reactions at various temperatures (Figure 2.9) shows that higher temperatures indeed promoted better polymerization, but this was also accompanied by significant increase in the amount of the breakdown peak observed. Temperatures of 75–90 °C seemed optimal for multiple DH-RH cycles considering the balance between breakdown and oligomerization.

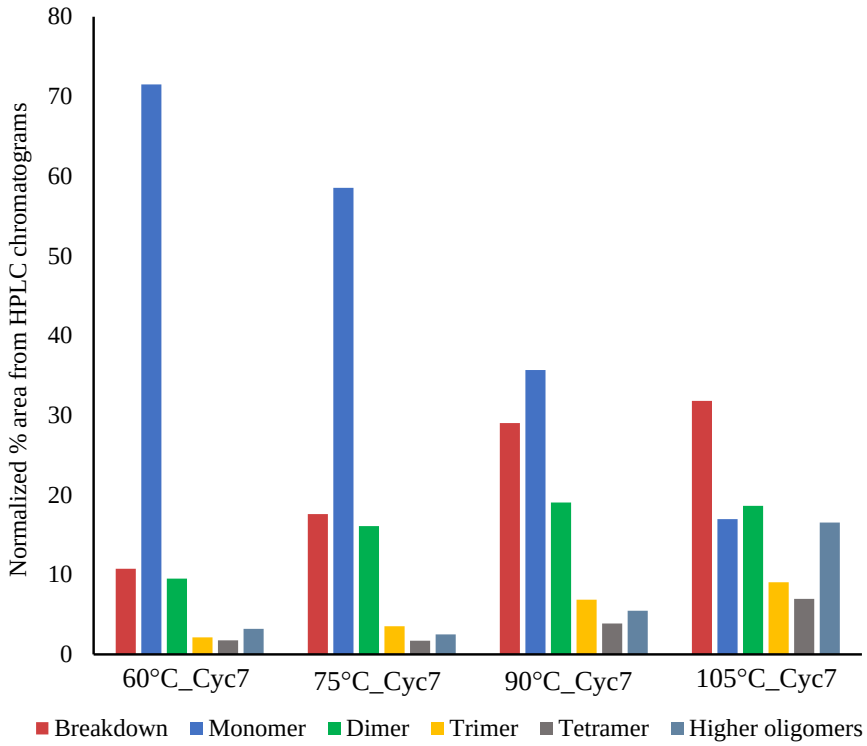


Figure 2.9: Graphical representation of HPLC chromatograms comparing all cycle 7 samples from reactions performed at various temperatures. Lower temperatures (60 and 75°C) show highest monomer peaks indicating low efficiency of the polymerization reaction whereas higher temperatures (90 and 105°C) promoted better polymerization but it was also accompanied by greater breakdown.

Condensation, which drives oligomerization, will take place during the dry phase of the DH-RH cycle, whereas hydrolysis is promoted during the rehydration phase. A longer drying time may, therefore, facilitate efficient polymerization but thermal breakdown of the starting reactants could also be higher over prolonged dehydration. Increasing the rehydration phase, in time or frequency, could result in reduced oligomer yields by promoting hydrolysis. However, these rehydration phases are also thought to be necessary as they will allow the formation of vesicles from dried lipid lamellae. Thus, encapsulation of newly formed oligomers inside vesicles is only possible in the wet phase. Rehydration phases are, hence, also important to allow reactant molecules and the various growing oligomers to get shuffled around, that will actually facilitate increased polymerization over several DH-RH cycles.

Given all these dynamics, there has to be an optimum regime of DH-RH that will balance all of the above mentioned processes and favor forward reaction of oligomerization over the backward reaction of hydrolysis, thus enabling efficient oligomerization. To identify such regime(s), various dehydration time scales, from 30 mins to 3 hours per cycle, were carried out and the reaction mixtures of pertinent cycles were analyzed. After initial dehydration, the rehydration of the reaction was performed after specified time intervals, and the reaction mixtures were subsequently allowed to dry completely. Therefore, each dry phase of the DH-RH cycle was of a specified duration (30 min, 1 hr, 2 hr and 3 hr) and each cycle count was considered from initial rehydration phase to the next one. It was observed that longer dehydration times indeed resulted in higher yields of oligomers, as shown in Figure 2.10. However, increasing the dehydration time to over 2 hours resulted only in a marginal increase in yields of higher oligomers from that of the 1 hour duration dehydration reactions. One to two hours of dehydration was found to be optimum considering the balance between the formation of oligomers and the breakdown product. This also indicated that a prolonged drying phase will not linearly increase oligomerization, but the relative efficiencies of forward and reverse reaction will mainly determine the distribution of products in a given reaction.

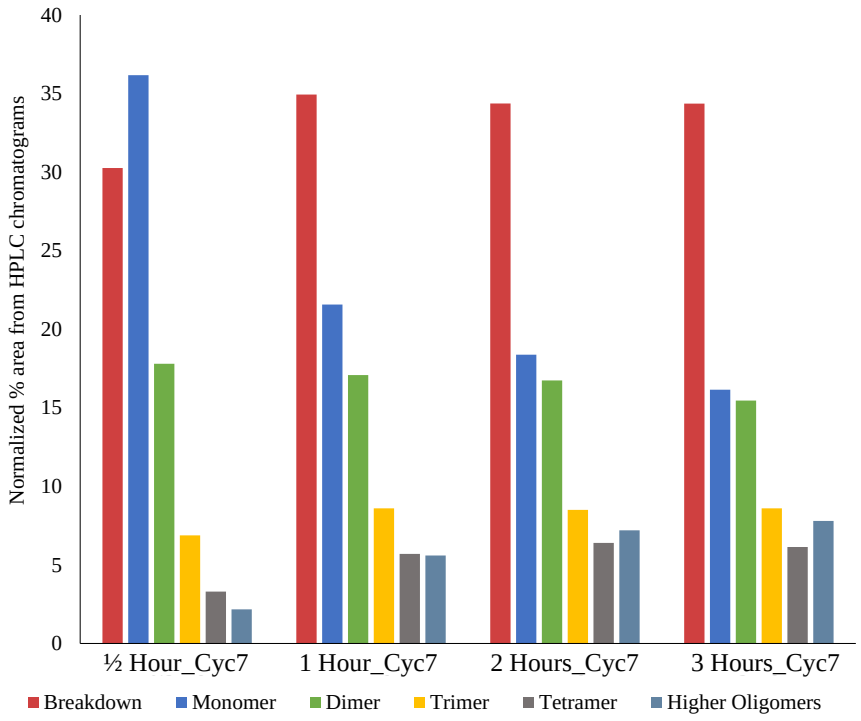


Figure 2.10: Graphical representation of HPLC chromatograms comparing cycle 7 samples of reactions carried out using varying dehydration time scales as mentioned. The optimum drying time for these reactions seem to be around 1 hour.

2.3.4 Variation of lipid to monomer ratios for achieving optimum polymerization

During the dehydration phase, lipids will form multi-lamellar films, which trap and organize the monomers within the bilayers. This results in increased concentration of monomers and also seems to impose an arrangement of their reactive groups in close proximity, thus facilitating oligomerization (Himbert et al., 2016; Toppozini et al., 2013). However, the ratio of lipid to monomer will determine the arrangement of monomers within a given dehydrated bilayer. For example, excess lipid may result in fewer monomer molecules being trapped per bilayer, thus limiting polymerization, whereas a smaller ratio of lipid to monomer might not confer the requisite organizational effect expected from lipids in the reaction. There may, once again, be an optimum ratio of lipids to monomers, which might facilitate greater polymerization. To test this hypothesis, reaction mixtures, with varying ratios of POPC: 5'-AMP were prepared. Monomer concentration was kept constant at 5mM so that maximum extent of polymerization is same in all cases. Amount of POPC added was adjusted to achieve 2:1, 1:1, 1:2, 1:5 and 1:10 ratios of

POPC:5'-AMP. These mixtures were subjected to seven cycles of DH-RH at pH 2, using H_2SO_4 at 90 °C. The reaction for each ratio was carried out in triplicates and normalized percentage area (used for graphical representation) was averaged for three reactions. Figure 2.11 shows the comparison of product distribution in reactions with varying ratios. More POPC than AMP in the reaction resulted in reduced overall oligomerization as seen in 2:1 reaction. Decreasing the amount of POPC in the reaction increased the oligomerization but the breakdown was also increased substantially.

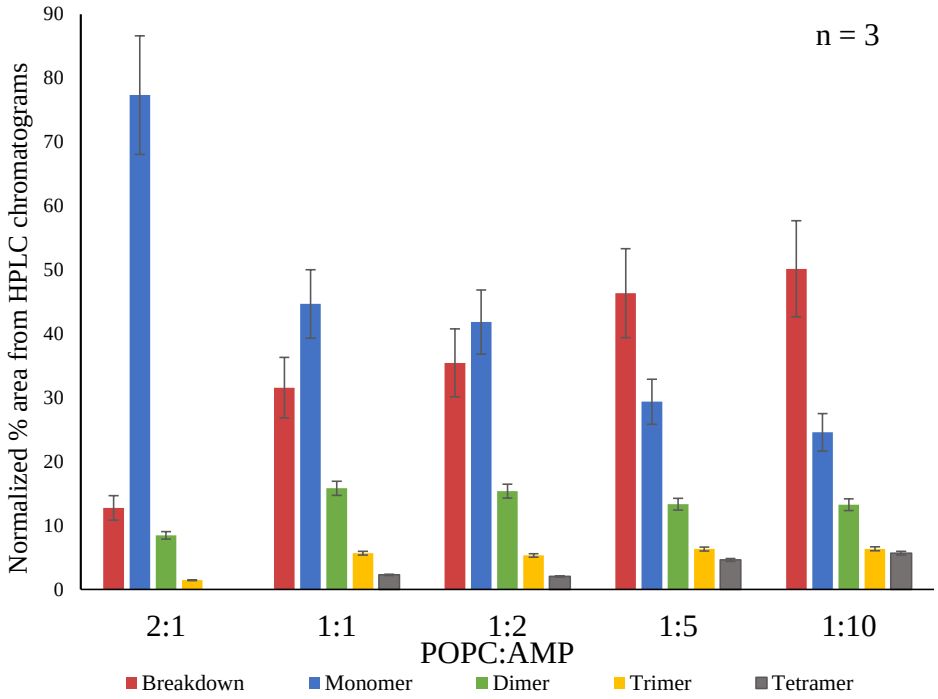


Figure 2.11: Graphical representations of reactions carried out in triplicate, using varying ratios of POPC:AMP. Presence of high concentrations of lipid (more than monomers) was found to be counterproductive for polymerization.

Since the polymerization reaction is completely chance driven, replicating the same at molecular level, even if all parameters are same, is not easily plausible. The error bars in the graphs were not scored for statistical significance as these do not represent the yields of oligomers. Importantly, this also was an evidence of the inherent variability in the reaction (unlike in controlled synthesis reactions) which makes it difficult to achieve exact replication. As for the oligomerization, ratio of 1:2 to 1:5 of POPC:5'-AMP was considered ideal as these reactions did not have excessive breakdown peaks and showed fairly higher amounts of oligomerization. A

less concentrated lipid scenario is also more realistic as prebiotic Earth did not have concentrated sources of any material and simple amphiphiles are thought to be synthesized in deep hydrothermal vents or resulted from exogenous delivery by meteorites. Dilute amphiphile solutions, which are still capable of vesicle formation, will result in fewer vesicles thus more oligomers can get encapsulated per vesicle. Interestingly, this could also favor interaction between oligomers, such as H-bonding or secondary structure formation that will further enhance functional selection.

However, we would like to acknowledge that this study uses phospholipids as a proxy for amphiphiles and does not make any claims of the prebiotic presence of these specific lipids. Attempts to replicate these results using fatty acids were undertaken but narrowing down mixed fatty acid systems that were stable, both, at high temperatures and under low pH, was non-trivial. This was because the fatty acid mixtures, which could form vesicles at high temperatures, were unable to tolerate low pH and vice versa. Further systematic characterization of the use of fatty acids to promote polymerization is one of the many open questions that have been raised by this study. The parameters that we have optimized thus far (pH, temperature etc.) will be important in narrowing down the combinations of fatty acids and their derivatives that could potentially be used to replace phospholipids in the aforementioned polymerization reactions.

2.3.5 Mass spectrometry of polymerization products

As demonstrated by HPLC, sufficient oligomers were formed during the DH-RH reactions. We, therefore, decided to perform mass analysis to further characterize these oligomers. Aqueous phase of samples from reactions with most optimal parameters (pH 2, 90°C, 1:5 ratio of POPC: 5'-AMP and 7 cycles of 1 hour each) was used for mass determination. MALDI was initially performed on the unpurified reaction mixtures. However, the material did not get ionized as expected. Possible reasons could be the presence of salt in the mixtures, as the starting material used was a sodium salt of 5'-AMP. Another problem was the use of anion exchange chromatography for achieving single nucleotide resolution, which resulted in introduction of more salt in the system, at a concentration that was much higher than the product yields. Individually collected peaks, therefore, could also not be analyzed by MALDI due to presence of excessive salt. Attempts to use C18 Zip-Tips (EMD Millipore) for reducing the salt content were not effective as it also resulted in the loss of shorter oligomers in the process. Therefore, HRMS

was employed for mass analysis of aqueous phase samples from the various reactions. The liquid chromatography aspect in this part of the analysis failed to resolve individual oligomeric peaks, but it did eliminate the salt from getting to the ionizer, resulting in mass profiles of whole reaction mixtures. Figure 2.12 shows the mass spectrum obtained from one of the reactions. Various masses observed in the spectrum were assigned to possible chemical species. Table 1 lists the possible chemical species, the expected mass for the various oligomeric species and the actual observed masses obtained from HRMS analysis.

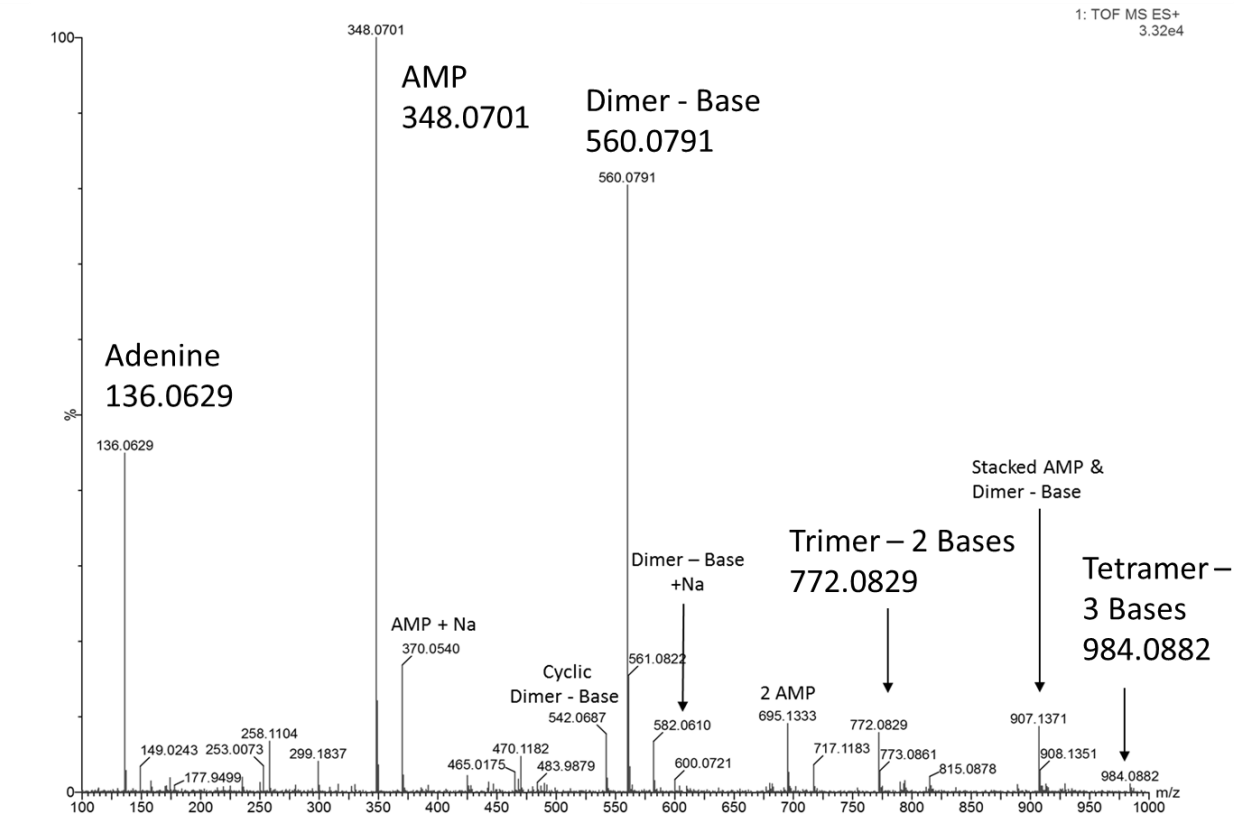


Figure 2.12: Mass spectrum of the unpurified AMP oligomerization reaction mixture. Various peaks observed were numbered and assigned chemical identities as listed in Table 2.1.

Table 2.1: Chemical identity assignment for the peaks from HRMS spectrum based on expected and observed mass numbers. Oligomers corresponding up to a tetramer were observed, albeit with abasic sites.

Chemical Species	Expected Mass	Observed Mass
Adenine	136.0617	136.0629
AMP	348.0703	348.0701
AMP + Na	370.0523	370.0540
Cyclic AMP Dimer - Base	542.0683	542.0687
AMP Dimer - Base	560.0789	560.0791
AMP Dimer - Base + Na	582.0608	582.0610
Stacked AMP	695.1334	695.1333
Stacked AMP + Na	717.1153	717.1183
AMP Trimer - 2 Bases	772.0875	772.0829
Stacked AMP & Dimer - base	907.1420	907.1371
Tetramer - 3 Bases	984.0961	984.0882

The observed masses clearly showed that the breakdown peak obtained in the chromatograms was that of adenine and not adenosine. Also, masses for intact oligomers were not observed in our multiple HRMS attempts. Instead, the observed numbers corresponded to oligomers with abasic sites. In an attempt to get a more precise measurement, HRMS analysis was performed on a purified dimer peak obtained by collection of just the dimer cluster from the HPLC profile. The peak was collected and lyophilized prior to subjecting it to HRMS analysis. The observed mass from the dimer peak confirmed the presence of a singly depurinated dimer (dimer minus a base). Surprisingly, the mass for intact dimer was not observed, even in lower abundance. Similarly, masses obtained for higher oligomers (trimer and above), as seen in Table 1, indicated that oligomers had multiple abasic sites. Loss of nucleobases in these oligomers may take place due to cleavage of the N-glycosidic bond. Some studies in the past have shown enhanced rate of depurination at low pH and high temperatures (Jordan and Niv, 1977; Suzuki et al., 1994; Zoltewicz et al., 1970). The proposed reaction mechanisms in these studies clearly show the

protonation of the nucleobases, which results in the subsequent generation of an abasic site at low pH (Figure 2.13 A). This side reaction seems to be facilitated by the same conditions that promote polymerization. Figure 2.13 B shows the possible structures of oligomers that have same masses as observed in the reaction and show abasic sites. The position of the intact base, which could be on the 5' or 3' end or on one of the internal sugars, could not be assigned based only on our results from mass spectrometry.

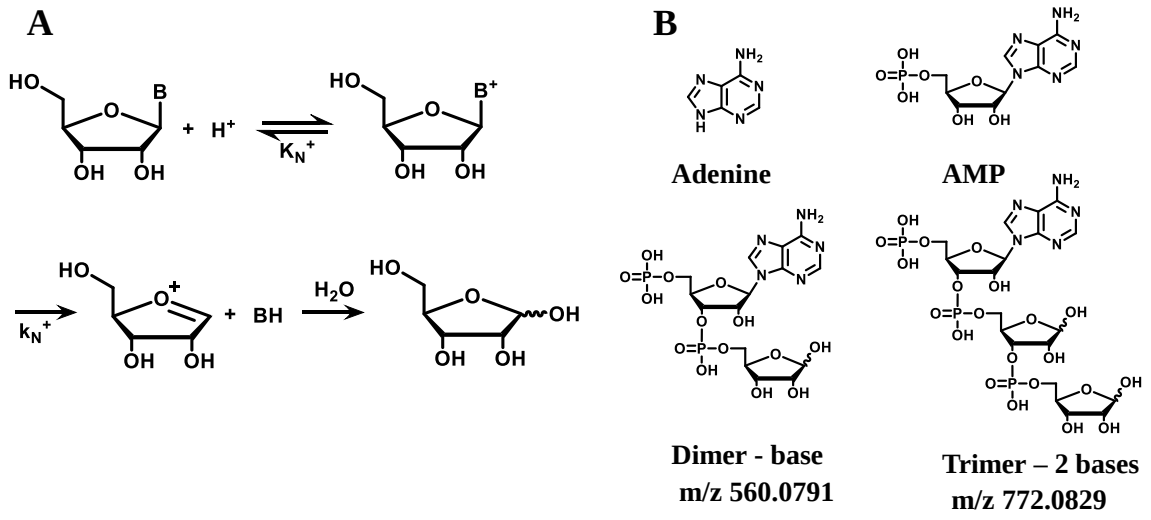


Figure 2.13: Proposed mechanism for loss of base under acidic conditions [modified from (Suzuki et al., 1994)] (A). Possible structures of oligomers, which have abasic sites due to depurination (B). Masses mentioned below each structure are the observed masses in the mass spectrum (Figure 2.12).

Formation of abasic sites during polymerization is undesirable as this will result in loss of the information containing moiety from the polymer. Such a polymer will not be able to store or transfer information and will also result in loss of structural properties. Since, both depurination and polymerization reactions occur at low pH and high temperatures, it is difficult to assign a causal relationship between them. Does loss of base take place after or during polymerization? Or, are these independent reactions that are facilitated under these reaction conditions? We found these questions very difficult to resolve due to the paradoxical nature of the monomers. One part of the molecule (phosphate) needs to be under low pH conditions for polymerization whereas the other part of the molecule (N-glycosidic bond) is unstable under acidic conditions, resulting in the loss of base. The reported glycosidic strengths of various nucleobases under harsh conditions

is not the same (Rios et al., 2015). Therefore, we decided to carry out reactions with other nucleotides as well, to determine if they might result in intact oligomers.

2.4 Summary

Lipid-assisted nonenzymatic polymerization of nucleoside 5'-monophosphates was replicated by performing DH-RH cycles at low pH and high temperatures. Several parameters that may affect this reaction were systematically varied in order to determine the optimum conditions for oligomerization. In addition, varying the different parameters also lead to a better understanding of the reaction kinetics, especially of the side reactions that take place during the DH-RH cycling process. As mentioned in the introduction, the reaction is hypothesized to have a mechanism similar to that of acid-catalyzed esterification of carboxylic acids (Olasagasti et al., 2011). This meant that reactions should be favored by low pH. Our results corroborated this as efficient polymerization was observed only when the reaction was carried out at a pH of 2 or lower. Chemical polymerization of non-activated nucleotides was likely to have been facilitated in niches which were at low pH, e.g., acidic geothermal pools. The rehydrating agent used for cycling should be able to maintain a low pH throughout the reaction in order to promote the polymerization process. Characterization of this parameter, using various acids as rehydrating agents, showed that thermolabile acids, organic acids with weak acidic strengths or acids with a lower boiling point such as HCL did not promote polymerization to a great extent. Sulfuric acid, when used to reduce pH, resulted in polymerization to the greatest extent in our study, followed by orthophosphoric acid. Both of these are strong mineral acids with high boiling points, which will be able to maintain low pH over the course of the reaction. Importantly, sulfuric acid is found in abundance in present day volcanic geothermal pools. High volcanic activity on the early Earth would have also resulted in the emission of sulfur dioxide, which when dissolved in water may have resulted in sulfurous acid. In the absence of oxygen, this would have favored reducing conditions and resulted in acidic environments rich in sulfuric acid.

Loss of water from reaction environments was found to play a key role in the condensation reaction that resulted in oligomerization. Both temperature and drying time were tested as these parameters will govern the rate of evaporation. As expected, the reactions carried out at higher temperatures of 90°C and 105°C yielded higher amounts of oligomers; however, 105 °C also resulted in high amount of breakdown products. Such high temperatures would have been more

common on the early Earth, in niches that supported extreme environments. Terrestrial volcanic/ geothermal pools may have likely been the geological niches that would have allowed for DH-RH cycles pertinent to our study. In general, lower temperatures (60–75 °C) resulted in lesser breakdown but also yielded in fewer oligomers. Increased drying time (duration of dehydrated phase) also promoted higher yields of oligomers. This can be attributed to condensation being favored in dry phase, thus resulting in oligomers. As these oligomers may undergo hydrolysis during rehydration, shorter wet phases and longer dry phases will eventually result in a dynamic accumulation of oligomers. Higher polymerization may also be achieved at lower temperatures by increasing the total number of DH-RH cycles or by having cycles of much longer duration, all of which are plausible on geological time scales but not feasible to undertake experimentally. Furthermore, even though phospholipids were used in this study, prebiotically relevant amphiphiles such as mixed fatty acids may have served a similar role in early stages of life's emergence. Interestingly, the presence of higher concentrations of amphiphiles was not only counterproductive in promoting polymerization but was also not very likely to be present in such niches by geological sources.

Abasic sites that were detected in the resultant oligomers can explain higher breakdown peaks observed in the different reactions. Both low pH and high temperature regimes favor loss of base even though they are necessary for polymerization as well. This scenario, therefore, results in the simultaneous occurrence of depurination and condensation in simulated DH-RH reactions. Oligomers with loss of base(s) will not run properly on gels, which may also explain the extensive streak like patterns that were observed in earlier studies. These oligomers will also have different absorption coefficients hindering their quantification. In light of these findings, we decided to perform oligomerization reaction with other nucleotides and binary mixtures of monomers capable of base pairing and also using a mixture all four nucleotides. Such mixed pool of monomers is more prebiotically realistic and could also result in heterogeneous oligomers with possibility of structure formation.

2.5 References

- DeGuzman, V., Vercoutere, W., Shenasa, H., and Deamer, D. (2014). Generation of Oligonucleotides Under Hydrothermal Conditions by Non-enzymatic Polymerization. J. Mol. Evol. 78, 251–262.

- Himbert, S., Chapman, M., Deamer, D.W., and Rheinstädter, M.C. (2016). Organization of nucleotides in different environments and the formation of pre-polymers. Sci. Rep. 6, 1–12.
- Jordan, F., and Niv, H. (1977). Glycosyl conformational and inductive effects on the acid catalysed hydrolysis of purine nucleosides. Nucleic Acids Res. 4, 697–709.
- Joshi, P.C., Aldersley, M.F., Delano, J.W., and Ferris, J.P. (2009). Mechanism of Montmorillonite Catalysis in the Formation of RNA Oligomers. J. Am. Chem. Soc. 131, 13369–13374.
- Misuraca, L., Natali, F., da Silva, L., Peters, J., Demé, B., Ollivier, J., Seydel, T., Laux-Lesourd, V., Haertlein, M., Zaccai, G., et al. (2017). Mobility of a Mononucleotide within a Lipid Matrix: A Neutron Scattering Study. Life 7, 2.
- Olasagasti, F., Kim, H.J., Pourmand, N., and Deamer, D.W. (2011). Non-enzymatic transfer of sequence information under plausible prebiotic conditions. Biochimie 93, 556–561.
- Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig. Life Evol. Biosph. 38, 57–74.
- Rios, A.C., Yua, H.T., and Tor, Y. (2015). Hydrolytic fitness of N-glycosyl bonds: Comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. J. Phys. Org. Chem. 28, 173–180.
- Suzuki, T., Ohsumi, S., and Makino, K. (1994). Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. Nucleic Acids Res. 22, 4997–5003.
- Toppozini, L., Dies, H., Deamer, D.W., and Rheinstädter, M.C. (2013). Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: insights into assembly of nucleic acid for primitive life. PLoS One 8, e62810.
- Zoltewicz, J.A., Clark, D.F., Sharpless, T.W., and Grahe, G. (1970). Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. J. Am. Chem. Soc. 92, 1741–1750.

Chapter 3

Oligomerization of Canonical Nucleotides using DH-RH Reactions

3.1 Introduction

Characterization of the reaction products from lipid-assisted polymerization of AMP suggested the presence of abasic sites in the resultant oligomers. This loss of base was attributed to the cleavage of the N-glycosidic bond caused by protonation due to the low pH of the reaction. Depurination has been studied in biological context in which abasic RNA was found to be significantly stable than abasic DNA (Leumann and Küpfer, 2007). The study looked at the rate of strand cleavage by β, δ-elimination and 2',3'- cyclophosphate formation and reported a 15-fold reduction in cleavage when abasic sites were present in RNA, as opposed to DNA. This greater tolerance towards strand cleavage could have potentially allowed some abasic sites in primitive RNA without significantly shortening the chain length. The mechanism for such loss of base, as discussed in the previous chapter, has been predominantly studied in DNA-based systems due to their biological relevance. Nonetheless, the insights from these studies can be used to identify the trends seen in loss of base reactions of nucleic acids in general. Typically, loss of base was found to be favored at low pH, especially below pH 2.4, which corresponded to the pK_a of the nucleobase that was studied (Suzuki et al., 1994). The N-glycosidic bond cleavage was found to take place by S_N1 mechanism and the contribution of 5'-phosphate group was negligible towards the rate of depurination (Jordan and Niv, 1977). Another study reported sharp increase in depurination with increase in temperature (Lindahl and Nyberg, 1972). The sigmoidal curve obtained for rate of depurination vs temperature showed a dramatic increase in slope due to loss of base above 85°C. Furthermore, depurination was found to proceed more rapidly in solvents with low ionic strength, and the presence of amines did not affect the rate. These studies further confirm the hypothesis that depurination will be favored under the same set of conditions that promote acid-catalyzed esterification.

A systematic kinetic analysis of the deglycosylation reaction was carried out by Tor's group for, both, modified and canonical nucleosides (Rios et al., 2015). This study reported that the protonation site (which was determined by pK_a), and the electronic nature of the heterocycles, were the most influential factors that resulted in the observed differences in the deglycosylation rates. Acidic conditions lowered the enthalpic activation parameter (ΔH) of deglycosylation, thus enhancing the ability of the leaving group that resulted in the loss of base, especially in the case of purines. pK_a of the nucleobases was found to be correlated with stability of the glycosidic bond; bases with higher pK_a (more basicity) also had higher ΔH, which resulted in lower rate

constants for deglycosylation. Heterocycles having electron-withdrawing groups (such as xanthine) possessed enhanced 'leaving group ability' as opposed to the ones with electron-donating ones (2,6-Diaminopurine). As seen in previous studies, DNA showed several fold higher rate of deglycosylation than RNA. Purine deoxyribonucleosides had half-lives ($t_{1/2}$) close to 15 minutes whereas their ribonucleoside counterparts showed half-lives of about 7-8 days under low pH (0.1M HCl) and physiological temperatures (37°C). When these values were calculated for higher temperatures of 50°C, $t_{1/2}$ for deoxyadenosine decreased even further to only 3.2 minutes, which confirms previous effect seen of temperature on the loss of base. One of the important claims of this study is that the canonical bases have the lowest rates of deglycosylation at physiological pH when compared to other modified bases. It has been argued that the stability of N-glycosidic linkages may have been one of the selection pressures in refining the genetic alphabet during transition from RNA World(s) to a DNA-based system of encoding information (Rios and Tor, 2012, 2013).

Given the aforementioned data and our observations with 5'-AMP, we decided to also carry out DH-RH reactions with the remaining three canonical nucleotides (i.e. 5'- CMP/UMP and GMP). We wanted to discern the potential for and extent of deglycosylation that might be prevalent in these contexts as well. In previous studies, native pyrimidines were observed to have really high stability ($t_{1/2}$ of over 400 days at pH 1 and 37°C (Shapiro and Danzig, 1972)), therefore making them better candidates for studying acid-catalyzed polymerization due to their higher glycosidic bond stability. Furthermore, the possibility of base pairing has been previously demonstrated in oligomers formed by DH-RH reactions (DeGuzman et al., 2014), which might also positively impact the yields of intact oligomeric. To check this possibility, mixtures of nucleotides capable of base pairing (i.e. AMP+UMP and GMP+CMP), and mixtures of all four nucleotides as well, were subjected to DH-RH reactions. It is pertinent to remember that complex product mixtures might result in these reactions as heteropolymers containing more than one type of nucleobase per strand, are expected to form. This requires calibration of analytical techniques and use of appropriate controls to avoid false positive results.

3.2 Materials and Methods

3.2.1 Materials

Adenosine 5'-monophosphate (AMP), uridine 5'-monophosphate (UMP), guanosine 5'-monophosphate (GMP), cytidine 5'-monophosphate (CMP), and ribose 5'-monophosphate (rMP) were purchased as disodium salts from Sigma-Aldrich (Bangalore, India) and used without further purification. The phospholipid, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), was purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA). All other reagents used were of analytical grade and purchased from Sigma-Aldrich (Bangalore, India).

Methods:

3.2.2 Dehydration-Rehydration Cycles

Oligomerization reactions were carried out in the same set up as mentioned in Section 2.2.2. The parameters that were optimized with AMP reactions were also used for carrying out reactions with the other three canonical nucleotides. Seven DH-RH cycles were carried out at 90°C, with 1 hour of drying time, and the pH of the reaction was lowered using H_2SO_4 and Milli-Q water was used as rehydrating agent for subsequent cycles. In order to check polymerization of other nucleotides, reactions were carried out with GMP, UMP and CMP in the same ratio as AMP:POPC (1:5 of lipid: nucleotide). Additionally, binary mixtures of nucleotides capable of base pairing i.e. AMP+UMP and GMP+CMP were also subjected to DH-RH cycles with 1:1 ratio of both nucleotides (for e.g., 2.5mM AMP + 2.5mM UMP). Finally, a reaction with all four nucleotides in equal ratio was also carried out to delineate oligomerization propensity of such a heterogeneous starting mixture.

3.2.3 Analysis of deglycosylation

Deglycosylation reaction analysis was studied mainly for AMP as it has been reported to have the highest N-glycosidic bond stability amongst canonical purines. AMP solution was maintained at pH 2 using H_2SO_4 and dried at 90°C. Separate reactions vials were used for individual time points and the reactions were analyzed for loss of base. Deglycosylation reactions were also carried out without dehydration of the starting reaction mixture by carrying out heating of the solutions in closed tubes. This was done in order to avoid oligomerization that also happens under these conditions. These samples were then analyzed by HPLC for analyzing

the breakdown in each sample and the percentage depurination was calculated as follows: (Area of breakdown peak/area of monomer peak) x 100. The time required for the breakdown of N-glycosidic linkage was calculated by plotting percentage depurination against time.

3.2.4. HPLC Analysis

Post butanol-hexane extraction to remove lipids, the aqueous phase of the various reactions was analyzed using HPLC in a manner similar to how the AMP reaction samples were analyzed (as mentioned in Section 2.2.3). Because some of the reactions contained more than one type of monomer, each nucleotide was injected separately as a control and the retention time of the respective peak was noted. In some cases more than one monomer peak eluted together (AMP and GMP) due to specificity of column to phosphates and not the actual bases. Since this column offered the best single-nucleotide resolution, we performed qualitative analysis using this column despite the aforementioned limitation of this technique, which did not allow any further optimization.

3.2.5 Mass Analysis

Initial mass analysis was carried out using similar techniques as mentioned in Section 2.2.4. HRMS was carried out for various reactions containing both single and multiple types or combinations of monomers in the starting mix. Some of the monomers (pyrimidines and mixtures thereof) could not be easily detected in mass spectrometric analysis, which may be due to poor ionization of these nucleobases under the conditions used during the analysis. For subsequent detailed analysis of nucleotide mixtures, we collaborated with Dr. Yayoi Hongo from Earth Life Science Institute (ELSI), Tokyo, Japan. Samples were lyophilized and shipped to ELSI at Tokyo Institute of Technology and further mass analysis was carried out there using Acquity UPLC+ system from Waters, with a CORTCES UPLC C18 column (1.6μm, 2.1 x 50 mm) using a water/acetonitrile gradient containing 0.1% trifluoroacetic acid. Mass determination was carried out in the positive ion mode with XEVO G2-XS QTof Mass Spectrometry.

3.3 Results and Discussions

3.3.1 Polymerization with other canonical nucleotides

Since formation of oligomers in lipid assisted DH-RH reactions with AMP could be successfully detected by chemical means (Section 2.3), reactions were also carried out with GMP, UMP and CMP. In this scenario, the mechanism of polymerization relies on the protonation of phosphate group and the subsequent nucleophilic attack of 2'/3' OH of a neighboring monomer, to result in a phosphodiester bond. This proposed mechanism does not involve any direct contribution form the nucleobase. Given this, in principle, all the four nucleotides should polymerize by the aforementioned acid-catalyzed esterification mechanism. However, AMP and GMP may have higher polymerization rates due to their ability for base stacking. This aids in better organization of the monomers, potentially facilitating efficient oligomerization.

In these set of reactions, the starting mixtures typically consisted of the nucleotide (AMP/ UMP/GMP/ CMP 5mM) and POPC (1mM), with the pH lowered to 2 with H_2SO_4. Seven DH-RH cycles were carried out and the reactions were analyzed by HPLC for analyzing the presence of oligomers and, to check for potential loss of base as well. Figure (3.1 A-D) shows the chromatograms from these reactions after completion of seven DH-RH cycles. Monomer and free base (as result of base loss) peaks were confirmed by comparing with control HPLC runs and doing co-injections for all reactions. AMP reaction was also repeated as a control/replicate under same conditions and similar results were observed as mentioned in the previous chapter (Figure 3.1 A). Reaction with GMP showed a chromatogram similar to what was obtained for the AMP reaction (Figure 3.1 B). Though higher oligomers were observed, as indicated by peaks after monomer, presence of free base was also observed as was evident from the presence of the breakdown peak. Subsequent mass analysis also confirmed the presence of abasic sites in the oligomers obtained from the GMP reaction. However, in the reactions with UMP and CMP, no peak was observed in the dead volume hinting at no loss of base in these reactions (Figure 3.1 C & D). Also, peaks were observed after monomers that matched the retention times for higher oligomers as was designated using relevant control oligomers. However, an interesting observation was that peaks were observed in both cases at the tetramer position despite the absence of a corresponding trimer peak. These peaks were designated based on corresponding retention times obtained for control oligomers containing AMP and this might possibly lead to

slight variation in elution times. It is, however, pertinent to remember that the separation of peaks on DNA PAC PA 200 columns is mainly based on the interaction of the backbone moiety with the column matrix and this is constant in all the nucleotides (a phosphate moiety). In all, different efficiencies in oligomerization were observed with the four nucleotides. However, due to the presence of abasic sites in some of the resultant oligomers, the reaction yields were not estimated for these reactions.

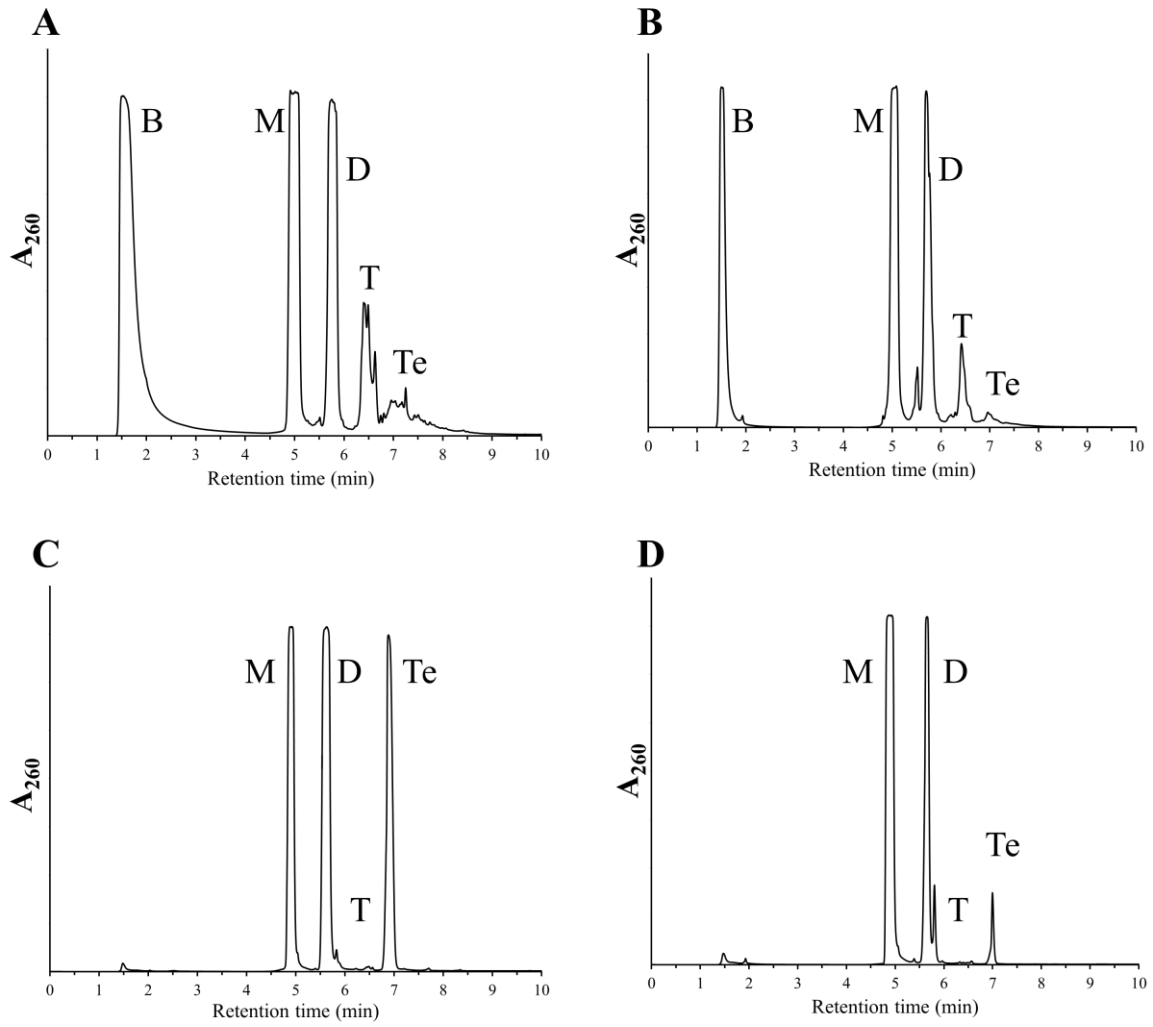


Figure 3.1: Chromatograms obtained from DH-RH reactions of AMP (A), GMP (B), UMP (C) and CMP (D). Within each chromatogram, peaks are indicated as breakdown (B), monomer (M), dimer (D), trimer (T) and tetramer (Te). These were tentative identities of peaks based on comparing retention times using related controls.

Initial mass spectrometric analysis was carried out for all the reactions using HRMS, just as it was done with the AMP reaction. Mass spectrum of the GMP reaction showed similar results to AMP reaction as depurinated oligomers were also detected in this case. Table 3.1 lists the chemical species, expected masses and observed masses from the GMP reaction. UMP and CMP reactions however did not result in mass spectrum that could be analyzed for potential oligomers. Even the control samples, which contained only the pure pyrimidine nucleotides, did not get ionized during mass analysis, resulting in spectrums with very low abundances. Due to this difficulty in analyzing the reactions with pyrimidines using the in-house machine, further analysis was carried out in collaboration with Dr. Yayoi Hongo, a mass analysis expert from ELSI at Tokyo Institute of Technology.

Table 3.1: List of identified peaks from GMP alone reaction mass analysis

Chemical Species	Expected Mass	Observed Mass
GMP	363.058	363.863
GMP Dimer - Base	575.06659	575.7672
GMP Trimer - 2 Bases	787.07518	787.6823
GMP Tetramer - 3 Bases	999.08377	999.3862

Mass spectrometry (MS) was performed at ELSI on individual monomer controls and reaction samples. Controls for all nucleotides showed the expected monomer peak but stacked higher oligomers were also present, which could result from association of molecules post ionization (Appendix Figure A.1-A.4). Fragmentation of the monomer was also taking place during MS as the control nucleotide samples showed peaks for free bases, ribose 5'-phosphate (rMP) and, in some cases, phosphate (iP) as well. Due to the possibility of such fragmentation during MS, loss of base was considered to take place during reaction only if it was observed in both HPLC as well as in the MS analysis.

Reaction samples from all four nucleotide reactions were also analyzed, including replicates for the AMP and GMP reactions, for maintaining uniformity in the analysis of results. All the mass numbers obtained from this analysis are summarized in Table 3.2. Both purine reactions (AMP and GMP) showed presence of mass numbers corresponding to the respective free bases, monomers, abasic dimers and abasic trimers. These results were consistent with previous observations (Table 2.1 and Table 3.1), which were performed on different mass analyzers. For

the pyrimidine reactions (UMP and CMP), mass numbers corresponding to the monomer and intact dimers were observed in both the cases. These were the only two reactions that showed the presence of intact dimers, which demonstrated a greater stability of the glycosidic bond in pyrimidines. However, mass numbers corresponding to free bases and abasic dimers were also observed in the pyrimidine reactions. This did not corroborate with the observations form the corresponding HPLC analysis, wherein no breakdown peaks were observed in these reactions (Figure 3.1 C & D). The presence of free base in the mass spectrum can thus be attributed to fragmentation during ionization in case of the pyrimidine only reactions. This was further confirmed by the presence of free bases in the mass spectrum of pyrimidine monomer controls too. Thus, pyrimidine reactions did not have obvious glycosidic bond cleavage during DH-RH cycles (based on combined evidences from HPLC and MS). Further reactions were carried out with nucleotide mixtures capable of H-bonding.

Table 3.2: Mass numbers observed in the reactions containing individual nucleotides

Chemical Species	Expected Mass	Observed Mass
AMP Reaction		
Adenine	136.0617	136.0635
AMP monomer	348.0703	348.0691
Abasic Dimer	560.0778	560.0764
Abasic Trimer	772.0874	772.0899
GMP Reaction		
Guanine	152.0566	152.057
GMP Monomer	364.0652	364.0627
Abasic Dimer	576.0737	576.0715
Abasic Trimer	788.0824	788.0803
UMP Reaction		
Uracil	113.0345	113.0346
UMP Monomer	325.0431	325.0435
Intact Dimer	631.0684	631.0682
Abasic Dimer	537.0517	537.0533
CMP Reaction		
Cytosine	112.0505	112.0494
CMP Monomer	324.0591	324.0596
Intact Dimer	629.1004	629.1017
Abasic Dimer	536.0677	536.0677

3.3.2 DH-RH reactions with nucleotide mixtures capable of hydrogen bonding

Hydrogen bonding (H-bonding) in the oligomers that resulted in DH-RH reactions of AMP and UMP was demonstrated in a previous study and this was based on hyperchromicity analysis of the products (DeGuzman et al., 2014). The study hypothesized that the possibility for H-bonding might have added to the organization of the reacting monomers, leading to better polymerization. However, as the previous results indicate (Table 3.2), loss of bases in the purine reactions, the number of H-bonds for a given length of oligomer might indeed be reduced than what is otherwise possible in an intact double stranded stretch. In order to check if starting with a mixture of base-pairing nucleotides could indeed affect the formation of oligomers because of H-bonding capability, and to assert the nature of oligomers in such mixed reactions, we carried out reactions with binary mixtures of nucleotides (AMP+UMP and GMP+CMP). Reactions consisting of 1:1 mixture of AMP:UMP or GMP:CMP (2.5mM each), along with 1mM POPC, were subjected to seven DH-RH cycles at pH 2 and 90°C. These samples were analyzed by similar methods as in the case of previous samples even though base specific separation of monomers or oligomers could not be achieved on the HPLC. Figure 3.2 shows chromatograms from analysis of these reactions, with peaks labelled in a manner similar to previous results. AMP+UMP reaction showed much more breakdown but also seemed to result in higher oligomeric peaks when compared to the GMP+CMP reaction. Some splitting of peak was observed in the peak corresponding to monomers (Figure 3.2 B, 5 mins) that could be attributed to base specific interactions, whereas higher oligomer peak clusters can be either due to mixed bases or previously mentioned variations possible in the resultant oligomers (Section 2.3.1). These results, as a proof-of-concept, indicated that even though oligomers were formed in mixed nucleotide reactions, abasic sites did result in these reactions as well. It was difficult to assign which bases, whether purines or pyrimidines, were being lost predominantly. However, results from the previous section (3.3.1) suggest loss of purines might be the major contributor to the observed breakdown peaks in these binary mixture reactions. Further characterization was carried out by mass analysis of the resultant oligomers.

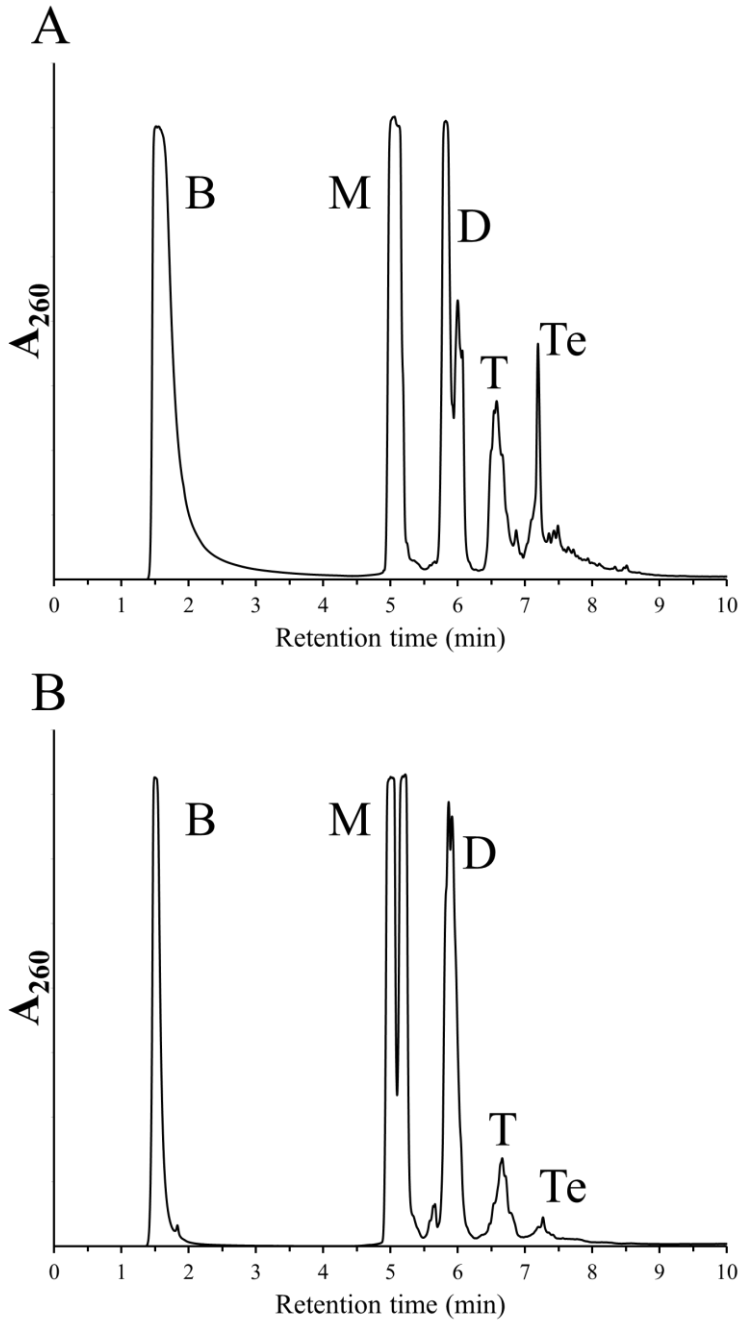


Figure 3.2: Chromatograms from AMP+UMP reaction (A) and GMP+CMP reaction (B) showed formation of higher oligomers but breakdown peaks were observed in both the reactions indicating loss of bases in these reactions too.

Our initial concern was that the reactions containing a mixture of all four nucleotides may result in oligomers that might be difficult to analyze. However, since the binary mixtures of nucleotides resulted in oligomers that could be resolved by HPLC, we decided to carry out a reaction

containing a mixture of all four nucleotides. Towards this, an AMP+UMP+GMP+CMP reaction mixture, containing 1.25mM of each nucleotide (5mM in total), and 1mM POPC, was subjected to 7 DH-RH cycles. HPLC analysis of this reaction revealed formation of similar oligomeric products in the mixed reaction as well (Figure 3.3). Due to potential competition within monomers, the yields of oligomers seemed somewhat reduced in this reaction as seen by a lower peak intensity for the oligomer peaks. Breakdown peak was observed in this reaction as well, which most likely corresponds to purine bases that might be lost due to cycling at low pH. The dimer peak resolved into multiple peaks, which can be attributed to high complexity in the dimer population, which results due to the possible incorporation of more than one nucleotide. For e.g. there are at least 10 types of intact dimers that can form in the reaction independent of the order of the bases. Apart from these, there will also be dimers with one intact base and one abasic site. Such hetero-oligomers were expected to result in this reaction and were characterized further with MS analysis.

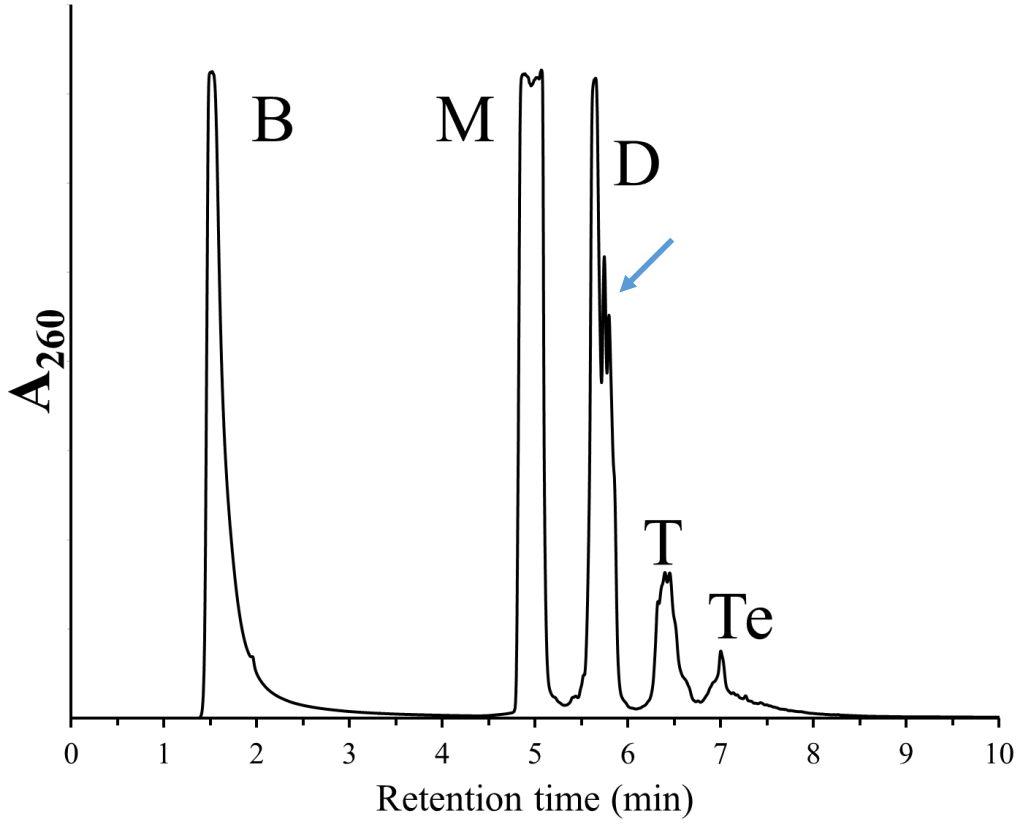


Figure 3.3: Reaction mixture containing all four nucleotides showed higher variation in oligomer peaks (as indicated by blue arrow), and loss of base as shown by breakdown peak.

Mass spectrometric analysis was carried out for, both, the binary and quaternary reaction mixtures. Though exact abundances of the peaks were different in these reactions, all of these reaction mixtures comprised of abasic oligomers with some purines and pyrimidines still left intact. These oligomers were similar to those observed in purine-only reactions that contained multiple abasic sites with a single intact base (as depicted in Figure 2.13 B). Intact dimers were observed only in G+C reaction but with a relatively poor abundance. Reaction containing all four nucleotides could not be analyzed in detail due to poor ionization of the sample. Table 3.3 lists all the mass numbers that could be identified from MS analysis of these mixed nucleotide reactions. Since this was not a quantitative MS analysis, we refrain ourselves from commenting on the yields of oligomers based on mass abundances. Species observed in the dimer and trimer populations mainly consisted of abasic oligomers with only one intact base. Completely abasic oligomers (such as rMP-rMP dimers) were not observed in these reaction possibly due to ligation of such products with other species resulting in higher oligomers (for e.g. formation of a species like AMP-rMP-rMP etc.). The loss of base seemed to be predominantly taking place during the DH-RH reaction, and not during MS analysis, as the breakdown peak was observed in the HPLC analyses of these reaction mixtures.

Table 3.3: Peaks observed in MS analysis of mixed nucleotide reactions

AMP+UMP Reaction			**GMP+CMP Reaction**		
Chemical Species	**Expected Mass**	**Observed Mass**	**Chemical Species**	**Expected Mass**	**Observed Mass**
Adenine	136.0617	136.0635	Guanine	152.0566	152.057
AMP monomer	348.0703	348.0691	GMP Monomer	364.0652	364.0627
Uracil	113.0345	113.0346	Cytosine	112.0505	112.0494
UMP Monomer	325.0431	325.0435	CMP Monomer	324.0591	324.0596
Abasic A Dimer	560.0778	560.0812	Intact CC Dimer	629.1004	629.1017
Abasic U Dimer	537.0517	537.0533	Intact CG Dimer	669.1065	669.1042
Abasic A Trimer	772.0874	772.0899	Abasic G Dimer	576.0737	576.0715
Abasic U Trimer	749.0603	749.0637	Abasic C Dimer	536.0677	536.0677
			Abasic G Trimer	788.0824	788.0803
			Abasic C Trimer	748.0763	748.0787

The exact chemical nature of these oligomers could not be resolved due to lack of purification methods that might have allowed for their evaluation by further analytical methods. Mass analysis of even the purified dimers was non-trivial due to presence of excessive salt in the purified fraction due to the use of ion-exchange chromatography. Other chromatographic methods (such as ion-paired reverse phase chromatography) did not yield sufficient resolution for purification. Nonetheless, these results strongly suggest that even in the reaction that contain all the four nucleotides, loss of base problem continued to persist. As stated earlier on in the introduction of this chapter, the rate of deglycosylation is variable amongst nucleobases, with purines being the most prone, thus indicating that the abasic sites are likely a result of mainly depurination. Furthermore, the polymerization of pyrimidines may be slower due to lack of base stacking, thus the formation of pyrimidine homopolymers may be difficult in the presence of other bases. Importantly, the observed abasic oligomers will not be able to store information or transfer it, which diminishes their significance as prebiotically relevant polymers. And, as both polymerization and deglycosylation reactions were being favored under similar conditions, we also decided to check for the degradation of nucleotides under DH-RH conditions.

3.3.3 Deglycosylation reactions during DH-RH cycles

As breakdown peak was predominantly observed in reactions involving purine nucleotides, either when used in isolation as singular monomers, or when present in mixtures, we followed a purine-based reaction at a smaller time frame. Since the AMP-based reaction showed the most breakdown peak intensity, we studied the depurination aspects of this reaction in detail. Since half- life ($t_{1/2}$) has previously been shown to decrease by about 5 times for dAMP, with increase in temperature from 37°C to 50°C (Rios et al., 2015), it was estimated that in the case of AMP this might decrease from days to hours/ minutes in reactions conducted at 90°C. Previous experiments have shown that DH-RH cycling was necessary for the formation of oligomers. Therefore, to minimize oligomerization in these experiments, we only carried out the reaction to look at several specified time points, without invoking rehydration. Samples were analyzed at 5, 10, 15, 30, 45 and 60 minutes after full dehydration of initial reaction mixture (Figure 3.4). HPLC analysis showed that the Breakdown peak was observed in as early as the 5 min sample, which indicated that some loss of base was taking place from very early-on. But in contrast with previous results (Rajamani et al., 2008), some oligomerization was also observed in these samples even in the absence of cycling. The dimer peak in the 5 min sample was about 3-4 % of

the monomer peak whereas the 60 min sample had much higher dimer, and even some trimer peak. These peaks would interfere with quantitation of deglycosylation as some of the monomers would also be utilized in the formation of oligomers.

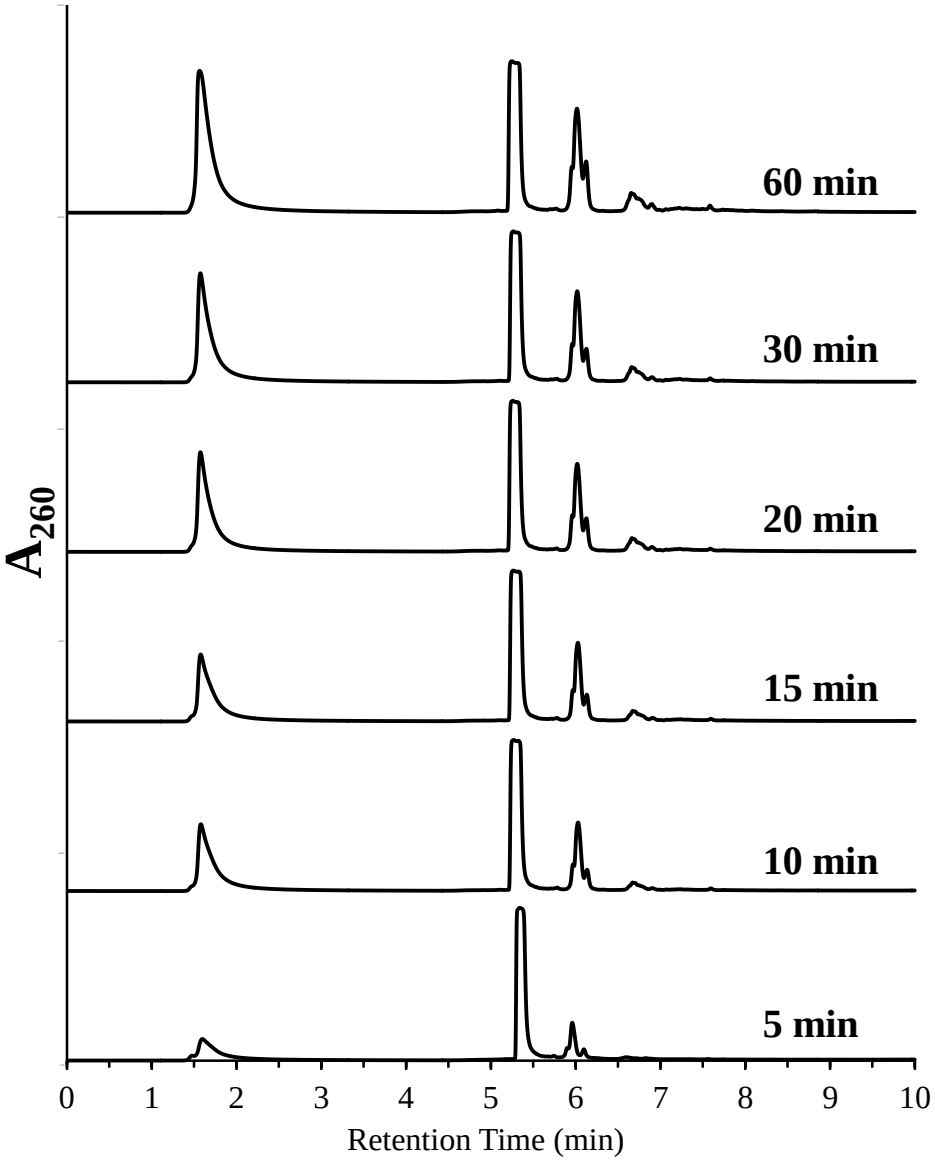


Figure 3.4: Breakdown and marginal oligomerization was observed in the reaction as early as 5 minutes.

Therefore, for calculating percentage depurination, we needed a set of conditions that could decouple oligomerization from deglycosylation reaction. Oligomerization requires complete dehydration as loss of water is not feasible in aqueous conditions. We, therefore, heated the AMP reaction mix at pH 2 and 90°C, under aqueous conditions in closed vials. Initially, similar time

points as mentioned before were taken for this reaction as well, but only about 6-7 % depurination was observed in 1 hour in these aqueous conditions. This reaction was then followed for 7 hours, which is equivalent to the duration of seven DH-RH cycles. In the pilot reaction, about 50% depurination was observed in the 7 hours sample and this was confirmed by repeating the reaction in triplicates. The average of percentage depurination (from 3 reaction replicates) was plotted vs time (Figure3.5), wherein the time taken for degradation of AMP to half of the starting concentration was found to be ~6.35 hours. This was about 30 fold reduction in the $t_{1/2}$ of AMP when compared to previous results of 8.2 days under pH 1 and 37°C. These results also confirmed that the rate of deglycosylation positively correlated with increase in temperature. Such rapid degradation of AMP at high temperatures poses serious challenge to the feasibility of long term reactions as a fixed amount of AMP will be lost in a fairly short time. Similar experiments were also carried out with the other canonical nucleotides and only GMP showed a high rate of breakdown similar to AMP. Both pyrimidines, UMP and CMP, did not show significant deglycosylation under similar reaction conditions even after 7 hours, which was in line with previous results that reported greater stability of pyrimidines (Garrett et al., 1966; Gates, 2009). However, pyrimidines did not yield oligomers with an efficiency that was comparable to purines, which somewhat undermines their glycosyl bond stability in the context of formation of informational molecules under prebiotic conditions.

(Figure 3.5 is on the following page)

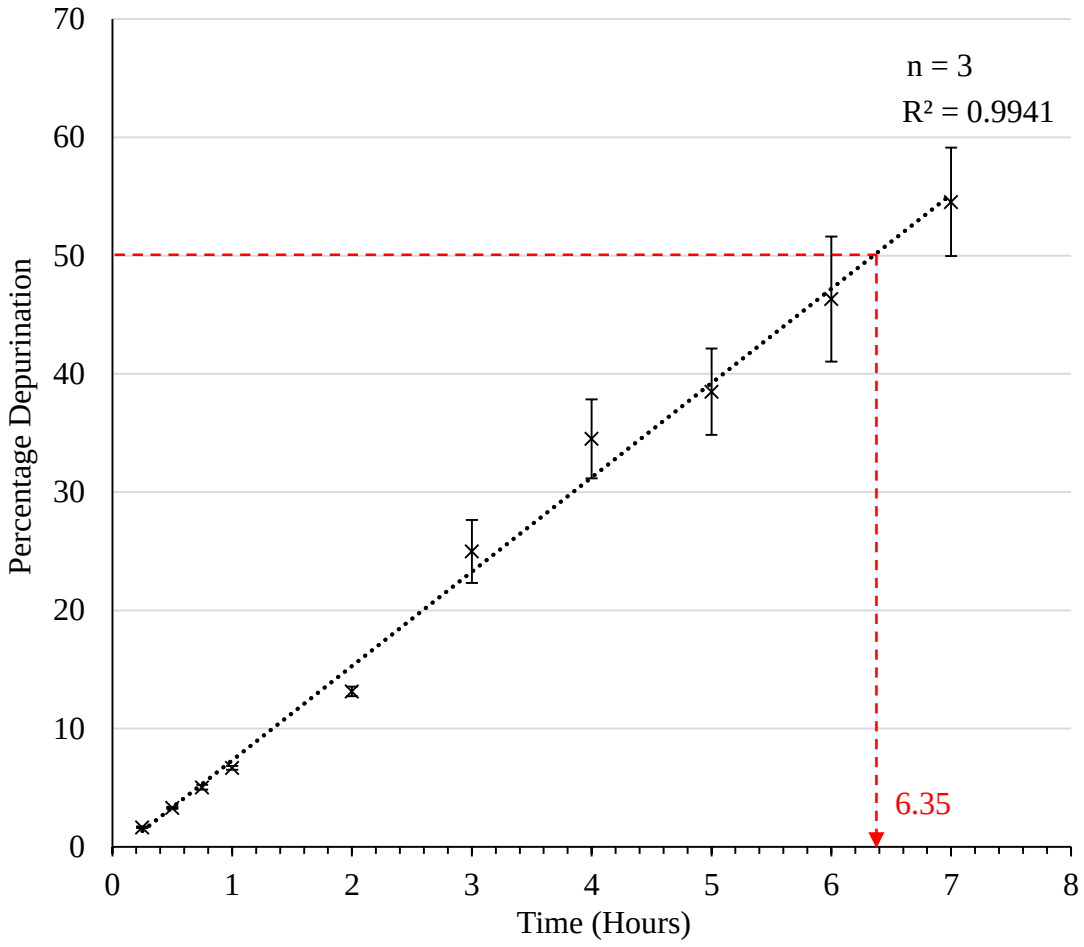


Figure 3.5: Percentage depurination (averaged over 3 replicates) has been plotted against time for estimating degradation of AMP under reaction conditions used for DH-RH reactions. AMP showed degradation to about half of the starting concentration in ~6.35 hr at pH 2 and 90°C.

3.4 Summary

Lipid assisted polymerization was carried out with canonical nucleotides, as either individual monomers, or binary mixtures capable of base pairing or as a mixture of all four nucleotides. Oligomerization was observed in reactions with CMP, UMP and GMP nucleotides, but the efficiency of the reactions varied depending on the nature of the nucleobase. The extent of oligomerization seen within nucleotides was as follows AMP ≥ GMP > UMP > CMP, based on reactions carried out with only individual nucleotides. Mass analysis of purine reactions was done at the in-house HRMS facility and showed presence of abasic sites in the GMP reaction products (similar to AMP). The pyrimidines however did not get sufficiently ionized and thus

could not be analyzed efficiently using HRMS. Further detailed analysis that was carried out in collaboration with Dr. Hongo at ELSI indicated the formation of oligomers in pyrimidine reactions as well. Intact dimers were observed in pyrimidine reactions along with dimers possessing abasic sites. These abasic sites were possibly an artifact of the analytical technique as fragmentation of even the control monomer was found to be taking place during ionization. Essentially, purines seemed to be more efficient at oligomerization but the oligomers had abasic sites, whereas pyrimidine only reactions showed formation of intact oligomers albeit with lower efficiency.

Reactions with binary and quaternary mixtures of nucleotides also resulted in oligomeric species while also resulting in loss of bases. Chromatograms from these reactions showed much higher variation in oligomers as they also comprise of hetero-oligomeric species consisting of more than one base. Mass analysis from these reactions revealed that the products consisted of oligomers with only one intact base (similar to Figure 2.12A). Such loss of base was very puzzling as one base was found to be intact in dimers and trimers, and this basically could be at any of the C1 positions in the resultant oligomers. However, the exact nature of these oligomers could not be delineated as the mass for a dimer with an intact base at the 3' end vs at the 5' end will be exactly the same. Selective fragmentation was not possible to identify the position of the intact base and it was hypothesized that such loss of base can happen at any position due to random deglycosylation. However, we also did not detect oligomers without any bases attached to them and this absence of fully abasic oligomers also was hard to explain. One possible explanation could be that once formed, they immediately react with intact monomer moieties that are present in much larger abundance, resulting in species like AMP-rMP-rMP etc.

Previous studies have reported that the phosphate group does not play any role in the mechanism of deglycosylation (Jordan and Niv, 1977). Also, the phenomenon of acid-catalyzed esterification might not promote further deglycosylation during oligomerization. Thus, there may not exist any causal relationship between the oligomerization and deglycosylation reactions, but both reactions are nevertheless found to be favored under the same set of conditions. The extent of deglycosylation was quantified by calculating the half-life for purine ribonucleotides under conditions used for DH-RH cycles. Earlier studies have suggested that AMP has one of the highest half-life among canonical purines (Rios et al., 2015) but under DH-RH conditions the apparent half-life of AMP was found to be only ~6.35 hours. Such high propensity of

degradation in purines raises imminent concerns about the stability of the glycosidic linkages in, both, monomers and the oligomers, especially under harsh prebiotic conditions. Pyrimidines showed higher half-life under similar conditions but in a realistic scenario, which consists of a mixture of nucleotides, the resultant oligomers will always tend to have many abasic sites. The loss of the purine bases will hamper the ability of such oligomers to store and transfer information as base pairing will be diminished.

Synthesis of canonical nucleosides, by formation of glycosidic bond between nucleobases and sugars, was found to be unfavorable under prebiotic conditions (Cafferty and Hud, 2014). Many studies in the past have looked at the formation of nucleosides, by the addition of canonical nucleobases to sugars, under dry conditions. However, such reactions did not result in expected products in high yields. Studies by Orgel and group, have reported synthesis of nucleosides with purines but the highest yields obtained for canonical nucleobases were less than 5-10% (W D Fuller et al. 1972a). Pyrimidines did not react with D-ribose at all thus no pyrimidine nucleosides could be synthesized in the aforementioned manner (W D Fuller et al. 1972b). Many other modified nucleobases, such as xanthine and hypoxanthine, showed slightly increased yields but canonical nucleosides were very difficult to synthesize prebiotically. This was famously termed the “nucleoside problem”, which highlights the intrinsic difficulty associated with the emergence of canonical nucleosides under prebiotic conditions. Combined with low stability of the glycosidic linkages under acidic conditions, it appears that modern canonical nucleobases may have been selected over time for their optimum fitness under physiological conditions (Rios and Tor, 2012). In fact, the canonical nucleobases may have been preceded by other heterocycles that were structurally similar while being more stable under harsh conditions. This idea has gained favor with a few origins of life research groups, including ours, and is being investigated in detail under the umbrella term “Chemical Evolution” (Hud et al., 2013). Based on the results we obtained from all the studies we carried out thus far with extant bases, we decided to study the feasibility of carrying out nucleotide synthesis with non-canonical nucleobases that may be more resistant to acid-catalyzed deglycosylation. Subsequently, successfully synthesized nucleotides containing alternate heterocycles were used to carry out oligomerization reaction under DH-RH conditions. We hypothesized that these reactions could result in the formation of informational oligomers that lacked abasic sites.

3.5 References

- Cafferty, B.J., and Hud, N. V. (2014). Abiotic synthesis of RNA in water: A common goal of prebiotic chemistry and bottom-up synthetic biology. Curr. Opin. Chem. Biol. 22, 146–157.
- DeGuzman, V., Vercoutere, W., Shenasa, H., and Deamer, D. (2014). Generation of Oligonucleotides Under Hydrothermal Conditions by Non-enzymatic Polymerization. J. Mol. Evol. 78, 251–262.
- Fuller, W.D., Sanchez, R.A., and Orgel, L.E. (1972a). Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. J. Mol. Biol. 67, 25–33.
- Fuller, W.D., Sanchez, R. a, and Orgel, L.E. (1972b). Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. J. Mol. Evol. 1, 249–257.
- Garrett, E.R., Seydel, J.K., and Sharpen, A.J. (1966). The Acid-Catalyzed Solvolysis of Pyrimidine Nucleosides1. J. Org. Chem. 31, 2219–2227.
- Gates, K.S. (2009). An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. Chem. Res. Toxicol. 22, 1747–1760.
- Hud, N. V., Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The origin of RNA and “my grandfather’s axe”. Chem. Biol. 20, 466–474.
- Jordan, F., and Niv, H. (1977). Glycosyl conformational and inductive effects on the acid catalysed hydrolysis of purine nucleosides. Nucleic Acids Res. 4, 697–709.
- Leumann, C.J., and Küpfer, P.A. (2007). The chemical stability of abasic RNA compared to abasic DNA. Nucleic Acids Res. 35, 58–68.
- Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry 11, 3610–3618.
- Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig. Life Evol. Biosph. 38, 57–74.
- Rios, A.C., and Tor, Y. (2012). Refining the Genetic Alphabet: A Late-Period Selection Pressure? Astrobiology 12, 884–891.
- Rios, A.C., and Tor, Y. (2013). On the Origin of the Canonical Nucleobases : An Assessment of Selection Pressures across Chemical and Early Biological Evolution. Isr. J. Chem. 53, 469–483.

- Rios, A.C., Yua, H.T., and Tor, Y. (2015). Hydrolytic fitness of N-glycosyl bonds: Comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. J. Phys. Org. Chem. 28, 173–180.
- Shapiro, R., and Danzig, M. (1972). Acidic hydrolysis of deoxycytidine and deoxyuridine derivatives. General mechanism of deoxyribonucleoside hydrolysis. Biochemistry 11, 23–29.
- Suzuki, T., Ohsumi, S., and Makino, K. (1994). Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. Nucleic Acids Res. 22, 4997–5003.

Chapter 4

Synthesis of Alternate Base Containing Nucleotides Under Prebiotic Conditions

4.1 Introduction

Synthesis of nucleotides has been a significant area of study concerning the origins of life research. The classical hypothesis regarding nucleotide synthesis has been as follows: Nucleobases can get synthesized from hydrogen cyanide or formamide under reducing conditions, or could be delivered by meteorites. Subsequently, these nucleobases form a glycosidic bond with ribose resulting in nucleosides and these nucleosides then get phosphorylated to form nucleotides (Kitadai and Maruyama, 2017). The main problem with this hypothesis has been the low efficiency of nucleoside formation with canonical nucleobases. As mentioned in the previous chapter (Section 3.4) this was given the term "nucleoside problem" and many studies have looked at the alternate strategies to synthesize nucleosides.

John Sutherland's group studied a ribose centric approach towards synthesizing nucleosides. The study reported the synthesis of pentose amino-oxazolines, an intermediate in the pathway, which was used in the successful synthesis of β-ribocytidine-2′,3′-cyclic phosphate (Anastasi et al., 2006). In their follow-up study, they also demonstrated conversion of oxazolines into β-ribouridine-2′,3′-cyclic phosphate by using photochemistry (Powner et al., 2009). A similar intermediate for the formation of purine nucleotides was also synthesized. However, the exact chemical pathway for synthesizing purine nucleotides, using the intermediate, remains to be demonstrated (Powner et al., 2010). Another study has used phosphorylated carbohydrates to test for prebiotically plausible synthesis of nucleotides (Kim and Benner, 2017). An alternate pathway for purine nucleoside synthesis using formamidopyrimidines, by using dry-wet cycles, was also demonstrated in other studies carried out by Carell's group (Becker et al., 2016, 2018). Some of the major criticisms concerning all the aforementioned studies has been the requirement for a highly coordinated process involving sequential set of steps (several in many cases) for synthesizing a said compound, and, importantly, the prebiotic plausibility of these reactions. These factors severely limit the application of these studies in discerning the actual origin of modern nucleotides, from the prebiotic soup.

One pertinent hypothesis that has emerged to overcome the nucleoside problem proposes the chemical evolution of related molecules (Hud et al., 2013). This hypothesis states that the chemical composition of nucleic acids may have changed over time and that the modern nucleotides would have been preceded by heterocycles that were different. Additionally,

studying the 'fitness' of canonical nucleobases also hints at the possibility of molecular evolution at play. Refinement of nucleobases is thought to have led to extant nucleotides that might have resulted post various selection pressures, which they may have encountered over a course of time (Rios and Tor, 2012). Harsh prebiotic conditions such as high temperatures, low pH and extensive UV radiation may have acted as selection criteria, allowing for the preference of specific heterocycles by virtue of their chemical stability under these conditions. Importantly, tolerance towards aforementioned harsh conditions would have not only been desirable, but also essential for the 'survival' of biomolecules in order to be functional. In this regard, few studies have looked at incorporation of non-canonical heterocycles, resulting in alternate nucleosides, as these may have been potential precursors of modern nucleobases.

Miller's group has reported the synthesis of nucleosides with a five-membered heterocyclic compound, which is isosteric to canonical pyrimidines (Kolb et al., 1994). Uracil, which is a six membered counterpart of urazole, does not form nucleosides with ribose under prebiotic conditions as has been shown in multiple studies. Nucleoside formation with Urazole was, however, found to take place in aqueous solution at moderate to high temperatures. The synthesized nucleoside was shown to form H-bonds with adenine, as demonstrated by NMR. The study proposes that urazole and guanazole may have acted as precursors to uracil and cytosine, respectively, in a pre-RNA world. Both of these precursors do not absorb in the UV range, which would have been an advantageous property as ozone layer would not have been present on the early Earth. Another interesting candidate for nucleoside formation was 2-pyrimidinone, which is an alternate pyrimidine with a chemical structure similar to uracil and cytosine around the 2nd carbon position. 2-pyrimidinone formed β-nucleoside (zebularine) under dry heating, in acidic solutions or at neutral pH, and with addition of Mg^{2+} (Bean et al., 2007). Zebularine was formed at much higher yields (~12%) under the conditions that were reported by Orgel for the synthesis of adenosine, whose yields under these same conditions was really low (<1%). The mechanistic insights obtained regarding nucleoside formation were as follows: The glycosidic bond formation under acidic conditions is facilitated by the cleavage of C1'-O_{OH} bond where the proton is donated by the protonated base and Mg^{2+} facilitates glycosylation by holding the ribose and the protonated base in proximity, and also lowers the energy of the transitions states. These studies have served as initial proof-of-concept that nucleosides can indeed be formed by alternate heterocycles, which can have relevant properties such as H-bonding.

Another study investigated glycosylation reaction with nucleobase analogues such as uracil with electron-donating substitutions at the 5^{th} or 6^{th} position (Kim and Benner, 2015). Amongst the various heterocycles studied, 6-aminouracil was found to result in nucleoside with more than 50% yield under dry heating conditions. Similar to previous studies, C-glycosides were formed with the sugar having either a furanose or pyranose structure, as confirmed by various analytical techniques. Acid catalyzed epimerization was carried out with purified nucleosides, confirming that it was C-nucleosides that were synthesized in the reaction. 6-aminouracil has two H-bonding patterns, one similar to uracil and another that is seen in purine analogues of an artificially expanded genetic information system. The authors of this study called this property being "biversal", which essentially would have allowed a nucleobase to have greater base pairing repertoire e in a pre-RNA world like scenario.

2,4,6-Triaminopyrimidine (TAP) is another interesting heterocycle, which was shown to form supramolecular assemblies with other heterocycles such as cyanuric acid (Cafferty et al., 2014). TAP has also been reported to form TAP-ribose conjugate with high yields in a one-pot like reaction under prebiotically plausible conditions (Chen et al. 2014). The major product of the reaction between TAP and ribose was a β-ribofuranoside, which was later found to be a C-glycoside by NMR analysis. The reaction resulted in about 60%-90% yield of nucleosides depending on the temperature, duration of incubation and the ratio of ribose to TAP, making it a robust reaction. Additionally, when purified TAP nucleosides (TARC) were mixed with cyanuric acid (CA), its base pairing partner, supramolecular assemblies consisting of π-stacked rosettes of H-bonded structures were observed. Similar assemblies were also observed on addition of CA to the crude mixture of nucleoside synthesis reaction. In this case, the size of the assemblies were almost half a micron in length, hinting at the formation of up to 1500 stacks of TARC-CA hexads. TAP, a potential precursor of modern nucleobases, could readily form a nucleoside and also select for its complementary heterocycles from a mixture of complex chemicals.

In a review article, Hud's group proposed that the base pairing seen in canonical nucleotides (i.e. Watson-Crick base pairing), could have been preceded by pyrimidine-pyrimidine base pairing (Cafferty and Hud, 2015). This hypothesis is supported by various observations such as lack of base pairing between Watson-Crick nucleobases in their monomer form, and the difficulty encountered in synthesizing canonical pyrimidine nucleosides. To narrow down potential nucleobase precursors, a theoretical chemical space of plausible proto-nucleobases was

constructed. This study considered 81 heterocycles, which have pertinent exocyclic groups such as O, H, or NH_2, out of which 27 were purines and 54 were pyrimidines. Each heterocycle was assessed based on several properties, including possible H-bonding with canonical bases, potential for nucleoside formation, presence in meteorites or propensity to be synthesized by model prebiotic reactions, and their ability to form self-assemblies with other heterocycles, as detailed in Figure 4.1. Several heterocycles showed more than one desirable property, which made them interesting candidates for future studies.

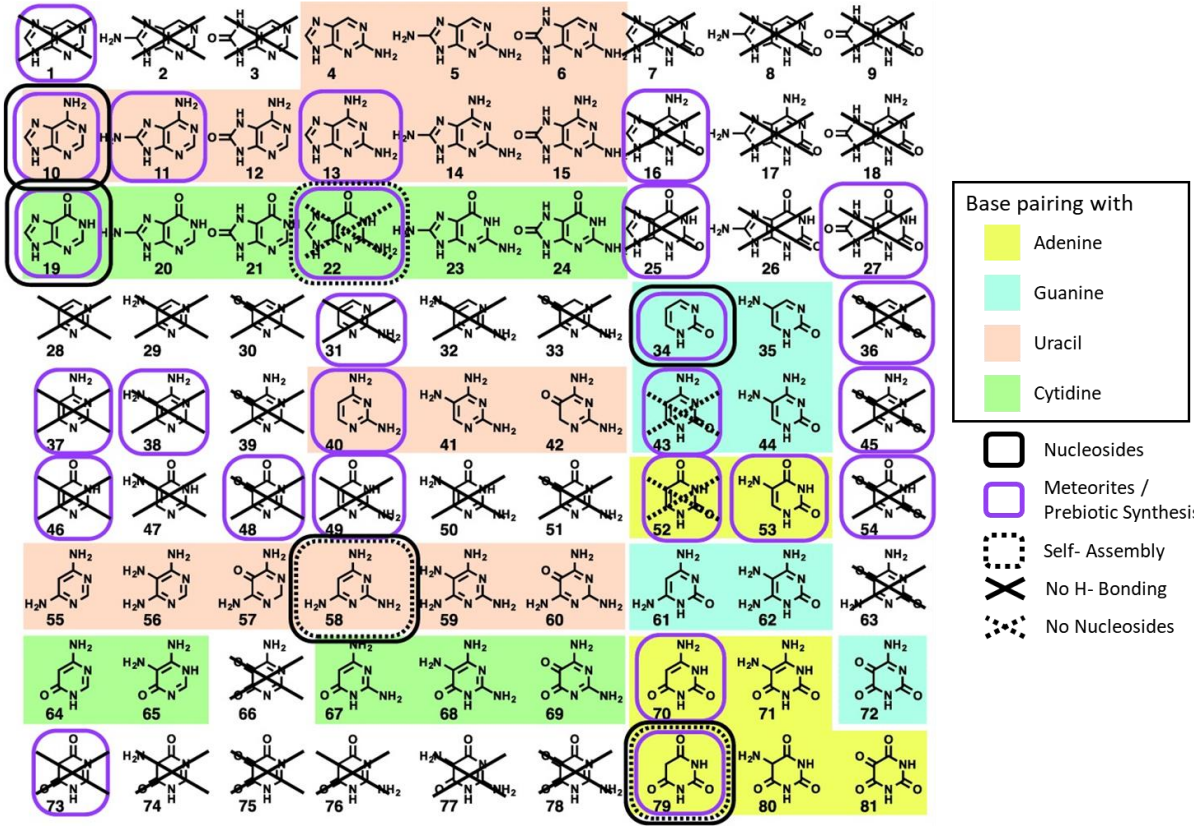


Figure 4.1 Figure showing selected candidates from 81 possible structures in chemical space. [Adapted from (Cafferty and Hud, 2015)]

Based on aforementioned selection criteria and other previous studies, we decided to carry out nucleotide synthesis with some of the alternate heterocycles. The primary goal of this study was to synthesize nucleotides that would be stable under acidic conditions, potentially aiding their chances to undergo polymerization, when subjected to DH-RH cycles under the same conditions. We performed the reactions with ribose 5'-phosphate (rMP) instead of ribose to, not only, eliminate the formation of pyranose epimers in the reaction, but also to have a preformed pool of

nucleotides that could readily undergo polymerization. Initially, we selected four candidates, including Cyanuric acid (CA), 2,4,6-triaminopyrimidine (TAP), barbituric acid (BA) and 2,4-diamino,6-hydroxypyrimidine, as alternate heterocycles for our nucleotide synthesis reactions. In order to keep the reactions prebiotically relevant, dry heating was performed with the aforementioned heterocycles and rMP, and the products were characterized further. Once nucleotide synthesis was successfully demonstrated, polymerization reactions were performed with these newly synthesized monomers to form informational molecules with alternate heterocycles.

4.2 Materials and Methods

4.2.1 Materials

Ribose 5'-monophosphate (disodium salt) (rMP), sodium barbiturate (BA), 2,4-diamino-6-hydroxypyrimidine (DAHP), 2,4,6-triaminopyrimidine (TAP), adenosine 5'-monophosphate (disodium salt) (AMP), deuterium oxide and sodium perchlorate were purchased from Sigma Aldrich (Bangalore, India) and were used without further purification. The phospholipid, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), was purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA). All other reagents and buffers that were used were of analytical grade and purchased from Sigma-Aldrich (Bangalore, India).

Methods

4.2.2 Nucleotide synthesis under prebiotic conditions

Nucleotide synthesis was performed by heating the aforementioned alternate heterocycles with rMP, under dry conditions. The set up described in Section 2.2.2 was used to simulate dehydration conditions at high temperature. Similar to experimental conditions that were reported by Orgel's group (Fuller et al., 1972), selected heterocycles were heated with rMP for 3 hours, at varying temperatures, with and without salts. The ratio of heterocycle to rMP was also varied in order to find the optimum ratio for maximizing the yield of the resultant nucleotide. Samples were dissolved in Milli-Q water after the completion of reaction and analyzed for nucleotide formation using analytical techniques that have been mentioned below.

4.2.3 HPLC Analysis

HPLC was performed using, both, reverse phase and ion-exchange chromatography for selected samples. Initially, C18 column (4.6 x 250 mm) from Agilent was used to standardize the separation of the free base from the resultant nucleotide(s). However, these solvents were found to interfere with subsequent analytical techniques, especially NMR, and hence ion-exchange chromatography was standardized for separation of the synthesized nucleotide(s). HPLC analysis was performed as detailed in Section 2.2.2. For analytical chromatography, DNAPac column was used with pH adjusted using Tris buffer. $NaClO_4$ was used for elution of the monomer at a flow rate of 1ml/min. The gradient of salt was adjusted for optimum resolution between the free base and the nucleotide. Area under the peaks for the starting material and the newly formed monomer were used for calculation of yields. Each nucleotide reaction was analyzed by measuring absorbance at the λ_{max} of the respective heterocycle used. Purification was subsequently carried out using DNAPac PA 200 column (9 x 250 mm). The solvents used were Milli-Q (Solvent A) and Milli-Q + $NaClO_4$ (Solvent B), and a flow rate of 2.5ml/min was used. Purified peaks were lyophilized using Labconco FreeZone 2.5 Benchtop Freeze Dryer and processed further for NMR analysis.

4.2.4 Acid stability of nucleotides containing alternate heterocycles

Canonical nucleotides showed lesser stability of glycosidic linkage under low pH and high temperature (Chapter 3). In order to compare the stability of nucleotides formed with non-canonical bases, the intact reaction mixture was heated under aqueous conditions at pH 2 (maintained by H_2SO_4) and 90°C, for 3 hours. Increase in the breakdown peak (that eluted in the void volume) was monitored as a proxy for deglycosylation.

4.2.5 Mass Analysis

Mass spectrometry was performed at the Center for Applications of Mass Spectrometry at the National Chemical Laboratory (NCL), Pune. Mass determination was carried out using the following set up: 6540 UHD Accurate Mass Q-TOF coupled with an Agilent 1260 Infinity Binary LC system (Agilent Technologies, Santa Clara, CA, USA); mobile phase: isocratic 50% acetonitrile/50% water containing 0.1% formic acid; Flow rate 0.3 ml/min without column.

4.2.6 NMR Spectroscopy

NMR spectroscopy was performed in collaboration with Dr. Jeetender Chugh's group in the Chemistry Department. Following are the details of the NMR analysis: All NMR experiments were performed on a Bruker Advance III Ascend NMR spectrometer operating at 14.1 T and equipped with a 5 mm quad-resonance (^{1}H, ^{31}P, ^{13}C, ^{15}N and ^{2}H lock) cryogenic probe. Samples were subjected to minimum 3 rounds of lyophilization/solubilization cycles in D_2O prior to measurement. Spectral acquisitions were performed with a sample volume of 300 µl in 5 mm Shigemi tubes (Shigemi Inc., PA, USA) matched for D_2O. Field frequency lock was achieved with D_2O. All 1D spectra with water suppression were acquired with a 90° flip angle (zgesgp in Bruker library) while 1D spectra without water suppression were acquired with a 30° flip angle (zg30 in Bruker library), a 12 ppm spectral width and a relaxation delay of 1 s; 32 transients were collected into 64 k data points with an acquisition time of 4.5 s and 4 dummy scans. All ^{1}H 1D spectra were referenced with respect to water signal (δ 4.7 ppm at 25 °C). The free induction decays (FIDs) were multiplied with an exponential weighing function corresponding to a line broadening of 0.5 Hz and zero fil led before Fourier transformation. The acquired spectra were manually phased and baseline corrected. ^{1}H-^{13}C 2D HSQC (hsqcgpph in Bruker library) were acquired with 1024x256 points, 160 transients, 16 dummy scans, relaxation delay of 1.5 s and an acquisition time of 85 ms.^{1}H-^{13}C 2D HMBC (hsqcgplpndqf in Bruker library) were acquired with 2048x64 points, 512 transients, 16 dummy scans, relaxation delay of 1.5 s and an acquisition time of 142 ms. 10 Hz of long range ^{1}H-^{13}C coupling was used to transfer coherence from ^{1}H to ^{13}C that are more than 2 bonds away. Carbon spectral width in both these experiments was 140 ppm with an offset at 100 ppm. ^{1}H-^{1}H TOCSY (mlevgpph19 in Bruker library) were recorded with 80 ms mixing time, 2048x512 points, 32 transients, 16 dummy scans, relaxation delay of 1 s and an acquisition time of 142 ms.^{1}H-^{1}H COSY (cosydfesgpph in Bruker library) were recorded 2048x512 points, 64 transients, 16 dummy scans, relaxation delay of 1.5 s and an acquisition time of 142 ms. Selective ^{1}H ROESY experiments (selrogp in Bruker library) were carried out by saturating H1' sugar protons in the molecule for 5 s.

4.3 Results and Discussion

4.3.1 Nucleotide synthesis with alternate heterocycles

For synthesizing nucleotides containing alternate bases, three heterocycles were narrowed down namely BA, TAP and DAHP. All of these are pyrimidines, which are structurally analogous to canonical pyrimidines. As mentioned in Section 4.1, nucleoside formation with TAP was demonstrated in an earlier study. In our study, we carried out TAP reaction with rMP by design, in order to form nucleotides. Pilot reactions for nucleotide synthesis were carried out by heating rMP and heterocycle at 90°C, under dry conditions for 3 hours. Ratio of the rMP to heterocycle used was kept at 5:1 (5mM rMP and 1mM base). A vast excess of rMP was used to facilitate complete incorporation of the base into the resultant product. Post three hours, the reaction mixture was analyzed for nucleotide formation using DNAPac column. Figure 4.2 shows chromatograms observed from each reaction (the respective heterocycle used in reaction is shown in the insert). Observation of peak(s) in the gradient indicated nucleotide formation and this was observed in all the reactions, albeit with varying efficiency. Multiple peaks observed in the gradient may correspond to varying types of mononucleotides (C/N glycoside or α/β isomer) that get synthesized in the reaction.

(Figure 4.2 is on the following page)

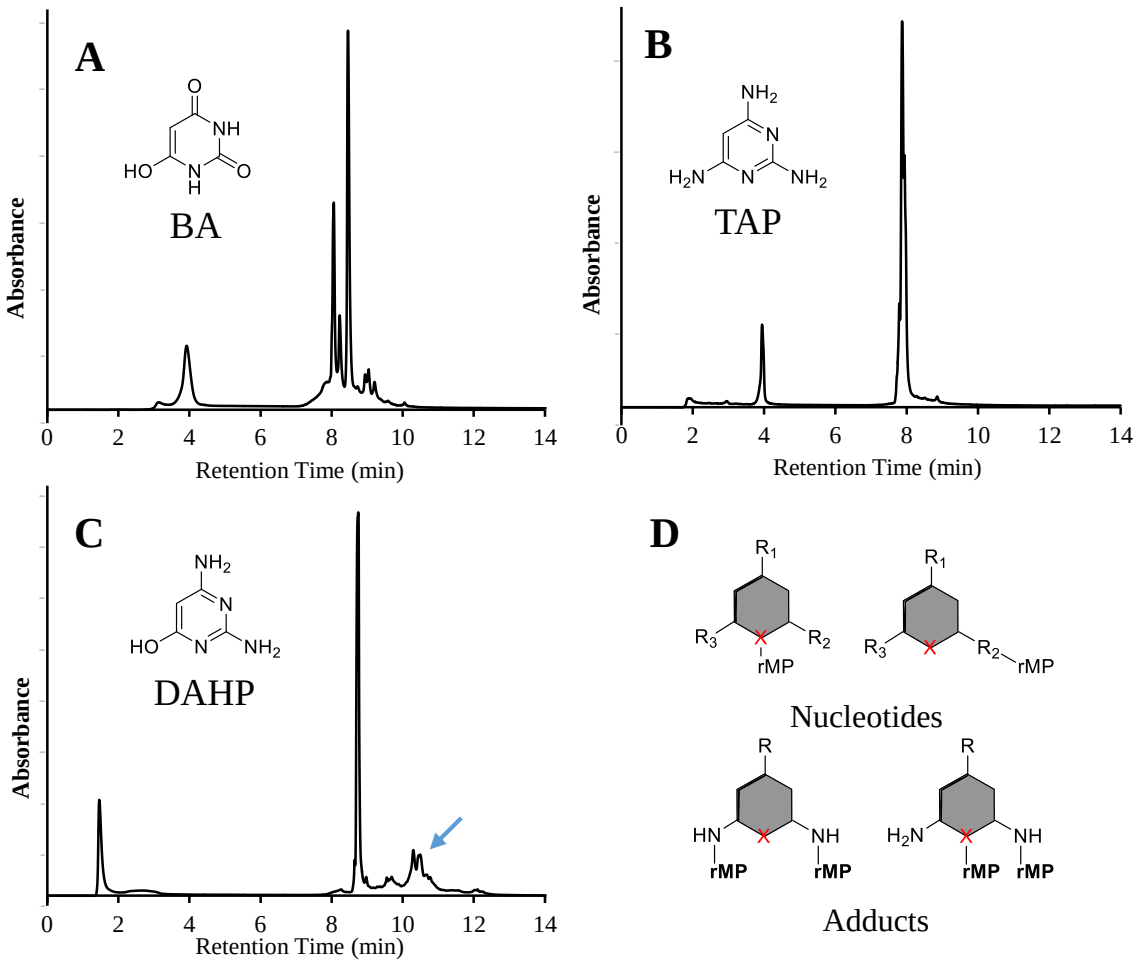


Figure 4.2: Synthesis of nucleotides with alternate heterocycles such as BA (A), TAP (B) and DAHP (C). The chromatograms showed peaks in the gradient, indicating the formation of nucleotides and plausible adducts, as mentioned in the description. Chemical structures of adducts was hypothesized based on previous reports (D).

Variability seen in the peaks that eluted in the gradient can be attributed to a variety of reasons including the type of glycosidic linkage that resulted in the monomers (C or N glycoside, α or β configuration), or the formation of adducts with a single base and multiple rMP moieties etc. Exocyclic amine groups present in TAP and DAHP will also act as nucleophiles. Therefore, more than one rMP could get attached to a single base moiety resulting in the aforementioned adducts. Plausible structures of these adducts are depicted in Figure 4.2 D. These adducts tend to elute later in the gradient (indicated by blue arrow in Figure 4.2) as they contain two sugars attached to one heterocycle. Interestingly, the BA reaction did not show such peaks as there are no exocyclic amine groups and the -OH group on the 5th carbon of BA does not act as a good

nucleophile. Such complexity in products affected the quantification of their yields. All three heterocycles indicated about 80% efficiency of incorporation of the respective base into nucleotides as roughly roughly by considering the area under the monomer peaks. Further confirmation of nucleotide formation was carried out by analyzing the reaction mixtures with mass spectrometry.

4.3.2 Mass analysis of products from nucleotide synthesis reactions

As most of the reactions indicated the formation of nucleotides, the reaction samples were analyzed with mass spectrometry in order to discern the chemical nature of nucleotides. Each reaction mixture was analyzed by mass spectrometry as mentioned in Section 4.2.5. Samples were injected directly into the mass spectrometer as the C18 column used could not separate the free base from the nucleotides. Mass numbers were recorded for each of the reaction mixtures. The expected masses for the products were calculated for all the reactions and the observed masses were compared to these numbers to designate the presence of a chemical species.

Reaction containing BA and rMP showed mass numbers corresponding to BA nucleotide and its sodium adducts. Expected mass for BA nucleotide is 340.0308 and the peak observed in the reaction was 340.0311, which confirmed the formation of the expected nucleotide. As sodium salts of starting materials was used for setting up the reactions, sodium adduct was observed in reaction in greater abundance (Expected mass: 363.0206; Observed mass: 363.0219). Similarly for the TAP reaction, the expected mass for the nucleotide was 337.0787 and the mass observed in the reaction was 337.0788, thus confirming the formation of the nucleotide. An adduct containing two sugar moieties and one TAP moiety was also observed in the reaction (Expected mass: 549.0873; Observed mass: 549.0862). Such adducts were also observed in previous studies (Chen et al. 2014) and were attributed to the sugar getting attached to exocyclic amines in addition to the formation of C- and/or N- nucleotides. Similar results were observed for DAHP reaction with the results showing the mass for the nucleotide (Expected mass: 338.0627; Observed mass: 338.0618) and for an adduct containing two sugars (Expected mass: 550.0713; Observed mass: 550.0695). All the masses discussed above and the chemical species they correspond to are summarized in Table 4.1.

Table 4.1: Chemical species observed during mass analysis of nucleotide synthesis reactions using alternate heterocycles.

Chemical Species	Expected Mass	Observed Mass	Mass Error (ppm)
BA-Nucleotide	340.0308	340.0311	1.07
BA-Nucleotide + Na	363.0206	363.0219	3.68
TAP-Nucleotide	337.0787	337.0788	0.21
TAP adduct	549.0873	549.0862	2.06
DAHP-Nucleotide	338.0627	338.0618	2.78
DAHP adduct	550.0713	550.0695	3.4

Interpretations from HPLC chromatograms obtained for nucleotide synthesis, and analysis of mass spectrometry data that resulted from these reaction mixtures, confirmed the formation of nucleotides in all the three reactions. Some adducts containing two sugar moieties were observed in case of DAHP and TAP, which was concurrent with previous reports. Detailed analysis of these reactions was found to be quite problematic as other chromatographic methods (such as reverse phase or ion-paired chromatography) could not effectively separate the adducts formed, from the main nucleotides in the reaction mix. Since BA reaction showed the least complexity in the resultant products, further conditions, such as variation in temperature and stoichiometry of the starting reactants, were characterized further using the BA reaction.

4.3.3 Optimizing conditions for nucleotide synthesis reaction

As the BA reaction seemed to yield only the nucleotide of interest, and no obvious adducts, we decided to use this as the base reaction to check the effects of various parameters on the reaction. These nucleotide synthesis reactions are essentially a condensation reaction between the 1'-OH of rMP and the C-H or N-H of barbituric acid (BA), with water acting as a leaving group. As it is a condensation reaction, we varied various parameters, such as presence of divalent cations, the reaction temperature etc., to see if they would allow for maximizing the yields. We also varied the ratio of BA to rMP to check its effects on the yields of the resultant monomer, the results from which have been discussed first. Towards this extent, reactions were performed with fixed concentration of BA (5mM), and to this rMP was added in increasing concentrations of 5, 10, 25 and 50mM, to achieve a stoichiometric ratio of 1:1, 1:2, 1:5 and 1:10, respectively. Previous

studies (Chen et al. 2014; Kim & Benner 2015) have reported the use of much higher concentrations of starting material. However, we performed these reactions at lower concentrations, as it is closer in semblance to what might have been plausible in a prebiotic scenario. The reaction mixtures were heated at 90°C for 3 hours, which is similar to conditions used in the pilot reactions, and they were subsequently analyzed by chromatography after re-suspension in Milli-Q water. Chromatograms from each of these reactions (Figure 4.3) showed efficient incorporation of BA into the nucleotide in, both, the 1:5 and 1:10 reactions. Other reactions also resulted in nucleotide formation, but the large peak that persisted in the dead volume indicated incomplete reaction. Higher concentration of rMP might be required as it seemed like significant amount of thermal degradation of the starting material (especially sugar) was taking place, as was evident from the browning of the reaction mixtures. Furthermore, the variations observed in the intensities of the peaks that eluted between 8 to 10 minutes indicated that the distribution of the resultant isomers could also be slightly different in the different reactions.

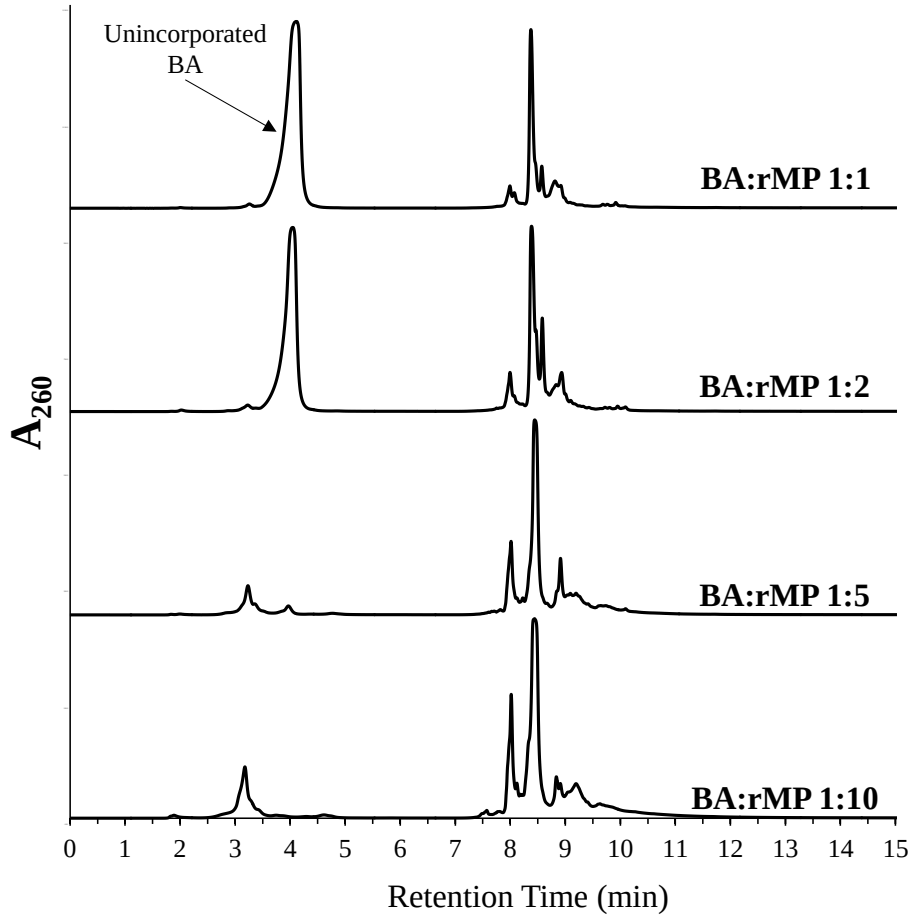


Figure 4.3: Variation of reactant ratios indicated efficient nucleotide formation in 1:5 and 1:10 ratio mixtures of BA:rMP while other ratios showed non-optimal nucleotide synthesis.

Additionally, other parameters like reaction temperature was also varied and the reaction was also carried out in the presence of Mg^{2+}. As mentioned earlier, nucleotide formation is the product of a condensation reaction and we hypothesized that the efficiency of the reaction might be proportional to the temperature used. This assumption was also based on our previous polymerization results (Mungi and Rajamani, 2015), where higher temperatures facilitated more oligomerization. Our results indicated that even though the nucleotide synthesis did take place at lower temperatures, the incorporation efficiency seemed to be reduced. For e.g., in the reaction that was carried out at 60°C, significantly lower yields were observed (Figure 4.4A) for the reaction with 1:5 of BA:rMP, but nucleotide formation was observed nonetheless. However, similar results have not been reported for canonical nucleobases, signifying that alternate heterocycles could potentially form nucleotides under a more diverse set of conditions. Particularly, the aforementioned result with BA indicated that this nucleotide's formation was feasible in environments with temperature in the regime of "Darwin's warm little pond", which may have been prevalent on prebiotic Earth. We also checked the effect of Mg^{2+} (in the form of $MgCl_2$) on the reaction. This was to enable comparison of results from previous studies where it was used (Fuller et al., 1972). Ratio of BA to rMP to Mg^{2+} in these reactions was kept at 1:2:2 and the mixture was heated at 90°C for 3 hours. As shown in Figure 4.4 B, significant increase in nucleotide yields was observed when Mg^{2+} was present in the reaction mixture. The black trace is for the control reaction and green trace represents reaction containing Mg^{2+}. Variations in peak intensities in the gradient changed in the presence of magnesium ions, indicating variation in the distribution of the resultant isomers in these reactions. This change of ratio was also observed in the case of 6-aminouracil nucleoside synthesis reactions that were carried out under similar conditions as reported in related studies (Kim and Benner, 2015).

(Figure 4.4 is on the following page)

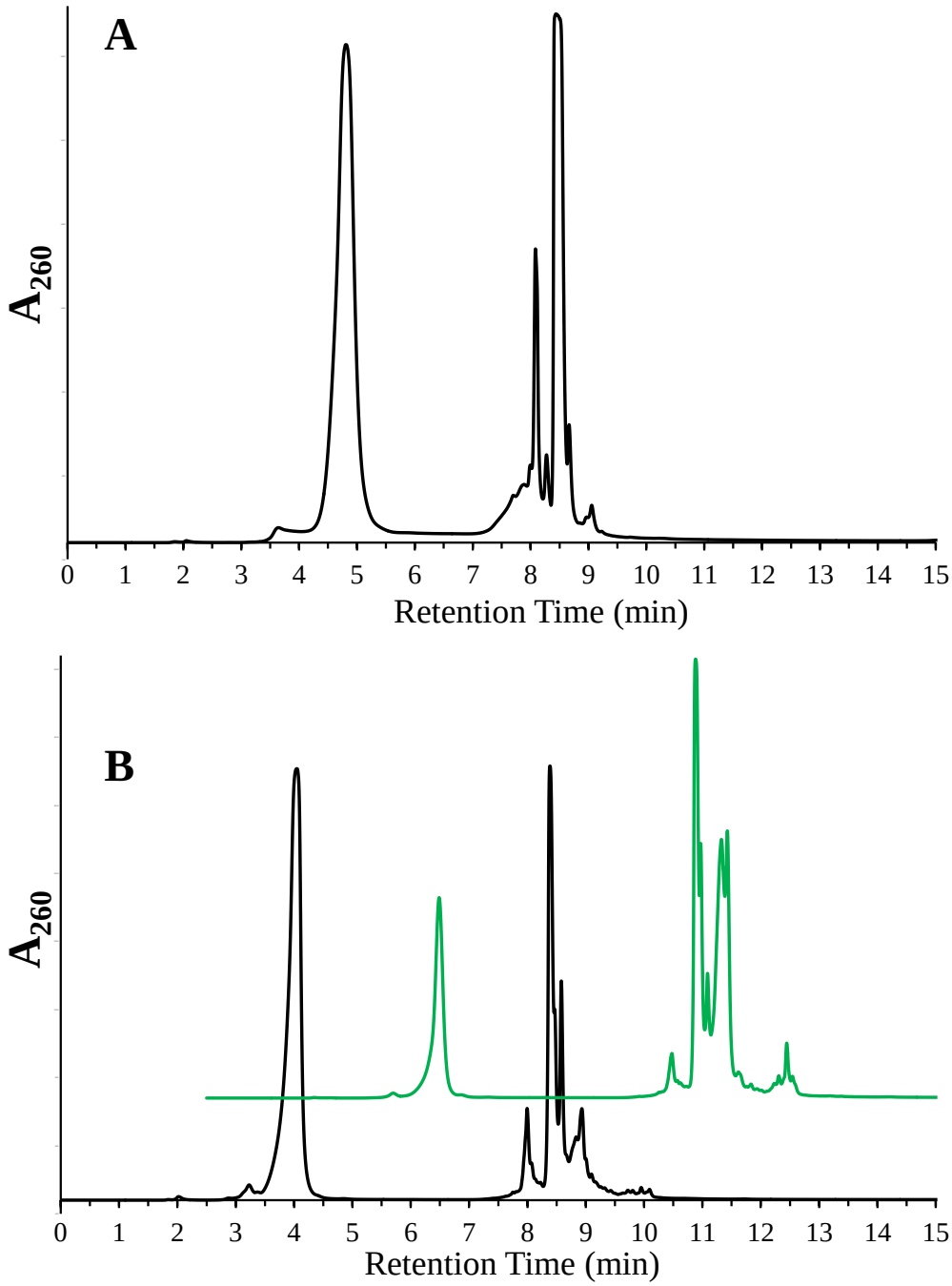


Figure 4.4: BA-Nucleotide reaction was found to be robust at low temperature and in the presence of divalent cations. Nucleotide synthesis with BA was found to be a robust reaction that could take place under varying conditions, albeit with lesser yields (compared to Figure 4.2 A) at lower temperature (A). It was also influenced by presence of divalent cations, such as Mg^{2+}, which seemed to increase the yields and also resulted in variation in the product distribution (B).

These results suggested that nucleotide formation with BA was a robust reaction and synthesis of such nucleotides containing alternate heterocycles, might have indeed been facilitated by prebiotic processes. However, the variation observed in the BA-nucleotide peaks, in terms of the nature of the resultant glycosidic linkage etc. was still unclear. Further analysis of BA-nucleotide products was performed using NMR to better understand the chemical nature of the monomer that was formed in this reaction.

4.3.4 NMR characterization of nucleotide synthesis reaction products.

For NMR characterization, BA-nucleotide was purified using semi-prep HPLC as mentioned in Section 4.2.3. Due to the formation of various adducts in the TAP and DAHP reactions, the peaks could not be purified readily by chromatography. Furthermore, solvents that were used in alternative purification methods that were attempted on the TAP and DAHP reactions interfered with the subsequent NMR characterization step, making it really challenging for precise interpretation. Given all the above constraints, only purified BA-nucleotide was subjected to NMR spectroscopy to identify the resultant predominant molecular species. NMR analysis was carried out as mentioned in Section 4.2.6 and Figure 4.5 shows the various NMR spectra obtained. ^{1}H-^{1}H COSY (A) and ^{1}H-^{1}H TOCSY (B) correlation spectra showed independent correlation between two sets (black and red) of protons. This revealed that two independent molecular species (A and B) were present in the purified HPLC peak (Figure 4.5 A and 4.4 B). Peaks at δ 5.23 ppm and δ 4.79 ppm have been assigned to two H1' protons corresponding to species A (black) and species B (red), respectively. All the correlations among ribofuranosyl protons have been seen from H1' to H5'/5". All the proton resonance assignments from COSY/TOCSY were transferred to respective carbons unambiguously in the ^{1}H-^{13}C 2D-HSQC spectrum (Figure 4.5 C). Observed C4' chemical shifts of δ 80.0 and δ 81.36 ppm hinted towards the presence of furanose form of ribose in both species A and B (as seen in previous studies[10]). ^{1}H-^{13}C 2D-HMBC experiment (optimized for 10 Hz) showed specific proton-carbon correlations suggesting the formation of both N- and C-nucleotide products (Figure 4.5 D). The chemical environments for C4 and C6 carbons were similar and were slightly upfield of C2 that is adjacent to two nitrogen atoms. Species A (black) showed correlation between H1' and C2 (δ 167.0 ppm) and not with C4 indicating it to be an N-nucleotide product. A proton at δ 5.03 ppm was seen correlating with C4/C6 at d 166.4 ppm and with another quaternary carbon at δ 88.5 ppm that could only result if either C4 or C6 undergo tautomerism to enolic form, further confirming the

N-nucleotide conformation of species A. Similarly, species B (red) showed correlation between H1' and C4/C6 as δ 166.6 ppm, and hence confirming the formation of a C-nucleotide product, and with another quaternary carbon at δ 85.4 ppm that could only result if either C4 or C6 showed tautomerism to enolic form. This tautomerism leaves C5 as quaternary carbon, thus making H5 unavailable for a correlation with H1' in COSY/TOCSY spectrum. Once formation of the BA-nucleotide was confirmed, ROESY was performed by saturating specific H1' to determine α and β configuration of the nucleotides. Saturation of H1' of the N-nucleotide at δ 5.23 ppm led to a through-space ROESY transfer of magnetization (negative intensity peak) to H4' at δ 4.0 ppm, shown by a green arrow, confirming β-conformation of H1'. Typically, selective ROESY experiment also contains positive intensity peaks that result from through-bond TOCSY transfer of magnetization, shown by cyan arrows, and has been seen from H1' to H2' at δ 4.17 ppm and H3' at δ 4.20 ppm (Figure 4.5 E). Saturation of H1' of C-nucleotide at δ 4.79 ppm led to a through-space ROESY transfer of magnetization (negative intensity peak) to H3' at δ 4.15 ppm and H5'/5" at δ 3.83 ppm, shown by a green arrow, confirming α-conformation of H1'. A TOCSY transfer of magnetization (positive intensity peak), shown by cyan arrow, was seen from H1' to H2' at δ 4.60 ppm (Figure 4.5 F). Thus, formation of the covalently linked product of BA and rMP was confirmed and the resultant BA-nucleotide was found to be a mixture of β-ribofuranosyl N-nucleotide (Figure 4.5 G) and α-ribofuranosyl C-nucleotide (Figure 4.5 H). Based on the NMR characterization C-nucleotide was the major product (~ 65%) whereas N-nucleotides was the minor product (~35%) from the nucleotide synthesis reaction.

(Figure 4.5 is on the following page)

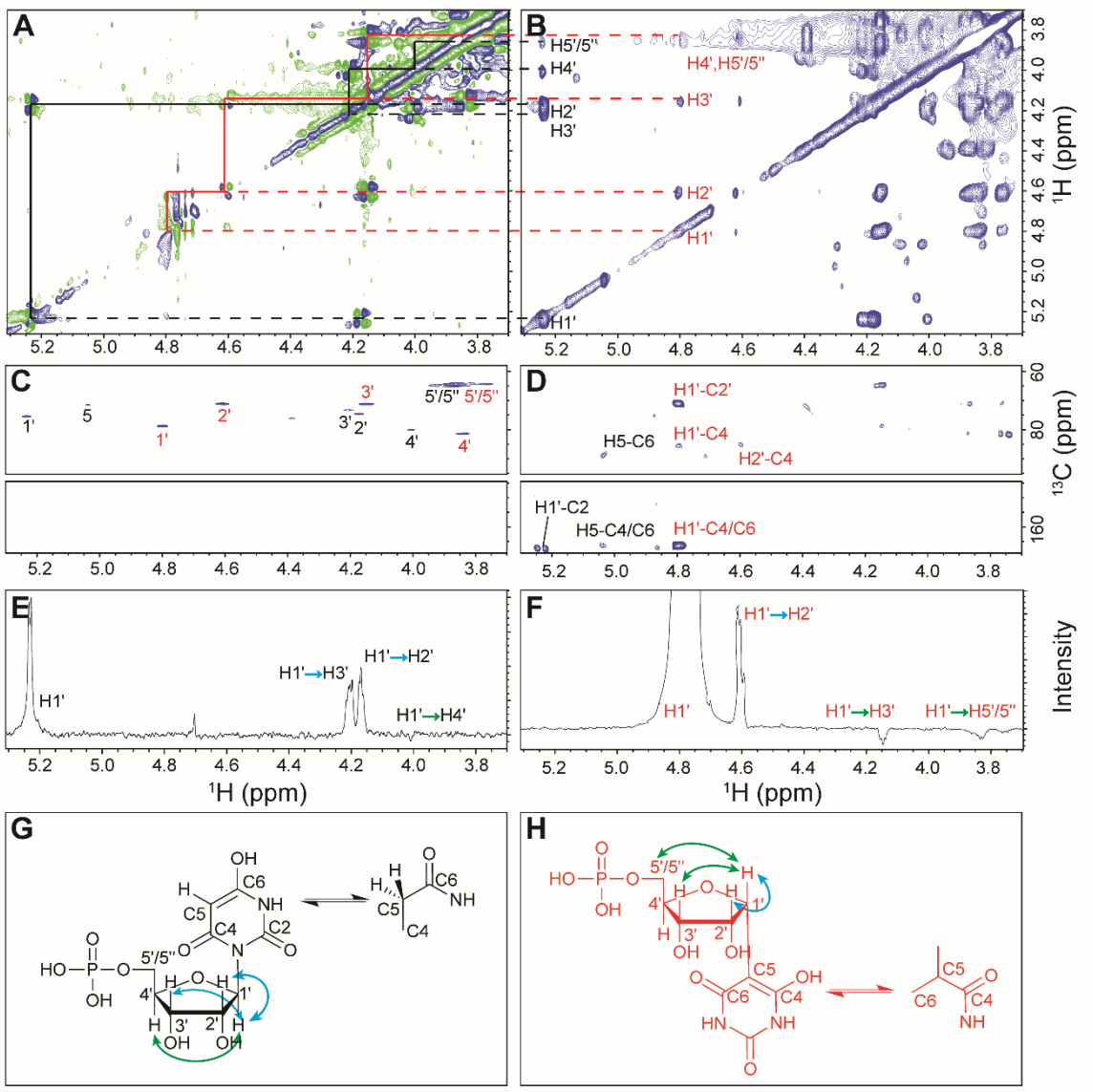


Figure 4.5 NMR analysis of BA-rMP reaction products: ^{1}H-^{1}H COSY (A) and ^{1}H-^{1}H TOCSY (B) correlation spectra indicates two sets of protons (black and red). (C) ^{1}H-^{13}C 2D-HSQC spectrum indicates the presence of furanose form of ribose. (D) ^{1}H-^{13}C 2D-HMBC spectrum confirmed the presence of N- and C- linked nucleotides. (E) ^{1}H selective ROESY spectrum where H1' of N-nucleotide at δ 5.23 ppm was saturated. (F) ^{1}H selective ROESY spectrum where H1' of C-nucleotide at δ 4.79 ppm was saturated. Structures of β-ribofuranosyl N-nucleotide (G) and α-ribofuranosyl C-nucleotide (H); green and cyan arrows show through-space and through-bond transfer of magnetization in selective ROESY experiment, respectively; and a plausible keto-enol tautomerism that BA might undergo has been depicted.

The plausible mechanism for the formation of nucleotides from the condensation of BA and rMP may be similar to those proposed for other heterocycles. Since BA is isosteric in part to 2-pyrimidinone, it may form N-glycoside in a manner similar to what was speculated for the formation of Zebularine or Urazole nucleoside (Bean et al., 2007; Kolb et al., 1994). In our reactions, the C5 carbon on BA was found to react with rMP to form C-glycoside, the mechanism of which will be similar to the previously reported mechanisms for TAP and 6-aminouracil (Chen et al. 2014; Kim & Benner 2015). The mechanism pertaining to the formation of, both, the C- and N-glycosides are shown in Figure 4.6, which proposes glycosidic bond formation analogous to that formed by TAP/ 6-aminouracil and 2-pyrimidinone/ Urazole, respectively. The mechanism also indicates that both C- and N-glycosides will be formed independently in the reactions, as was confirmed by NMR.

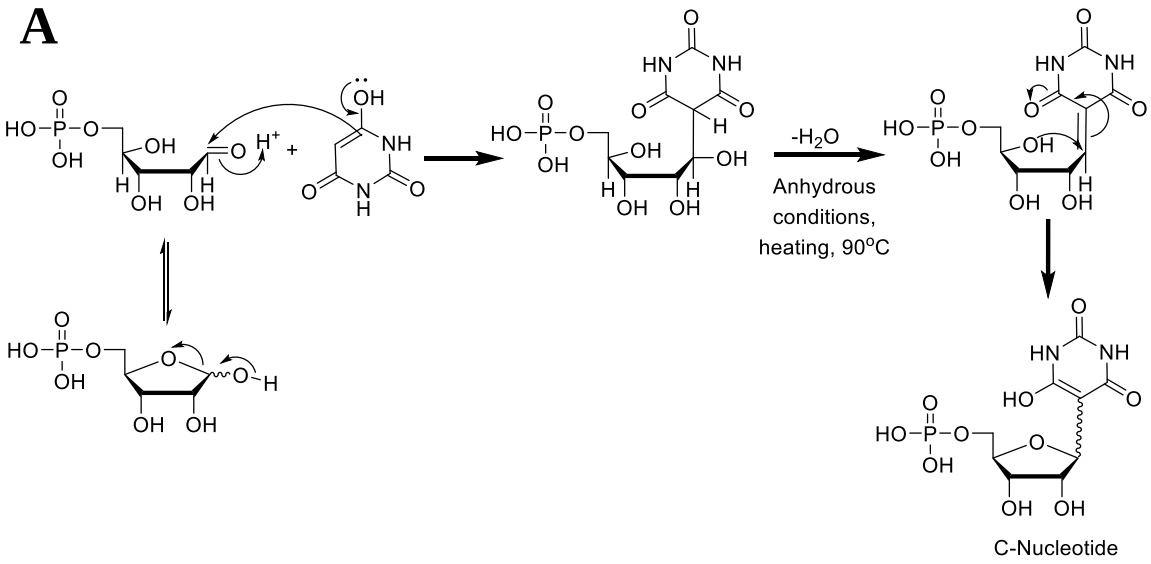


Figure 4.6 A: Possible mechanism for the formation of C-nucleotide from BA and rMP. The reaction mechanism has been adapted and modified based on previous literature (Kim and Benner, 2015)

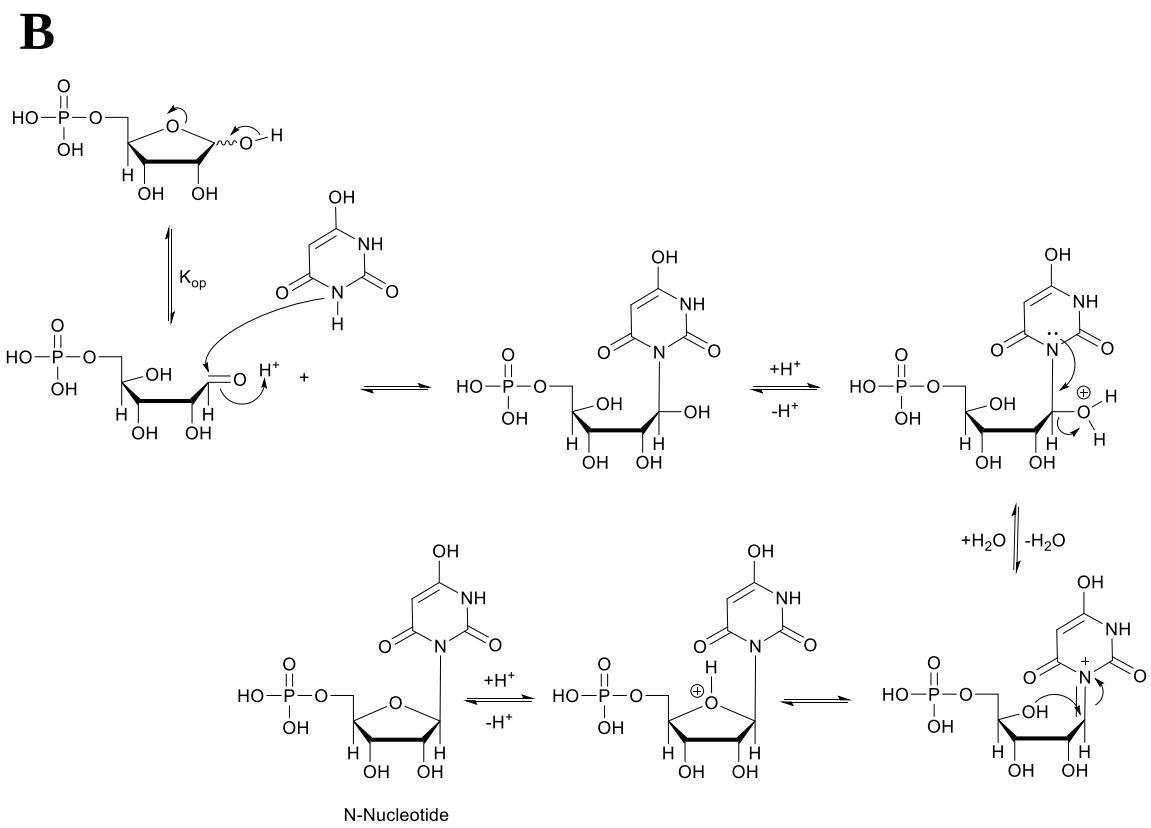


Figure 4.6 B: Possible mechanism for the formation of N-nucleotide from BA and rMP. The reaction mechanism has been adapted and modified based on previous literature (Bean et al., 2007).

4.3.5 Stability of non-canonical BA nucleotide under acidic conditions.

Loss of base under acidic conditions, due to cleavage of the N-glycosidic linkages, was the major problem with canonical nucleotides with regards to their stability under harsh prebiotically pertinent conditions. This led us to explore the possibility of synthesizing pre-RNA World nucleotides with alternate heterocycles. In this regards, we successfully synthesized the BA-nucleotide and used it to explore its potential role in the formation of pre-RNA World informational polymers. Interestingly, BA-nucleotide should not lose the informational moiety, i.e. the nitrogenous base, under acidic conditions as the BA-rMP reaction resulted in the formation of C-glycoside based nucleotides, which are known to be acid stable. To check this, the monomers from the reaction mixture were heated in the presence of a strong mineral acid. AMP was used as a positive control as it undergoes depurination under these high temperature and low pH conditions. The solution of monomers was subjected to heating at 90°C and pH 2 for

3 hours and the samples were analyzed using anion exchange chromatography. Figure 4.7 shows HPLC chromatograms of the reaction mixtures, pre- and post-heating, for both AMP (4.7 A) and BA-nucleotides (4.7 B), respectively. Analysis of AMP reaction showed a peak in the dead volume post-heating (red trace), which corresponds to the free adenine that is generated as a result of depurination. Whereas, the BA-nucleotide reaction did not show any significant change in the intensity of the peak that eluted in the dead volume. The distribution of peaks in the gradient, however, was slightly altered on heating with the acid. This can be attributed to acid-catalyzed epimerization of C-linked nucleotides, which is known to change the proportion of various isomers of the monomers as mentioned in Section 4.3.1.

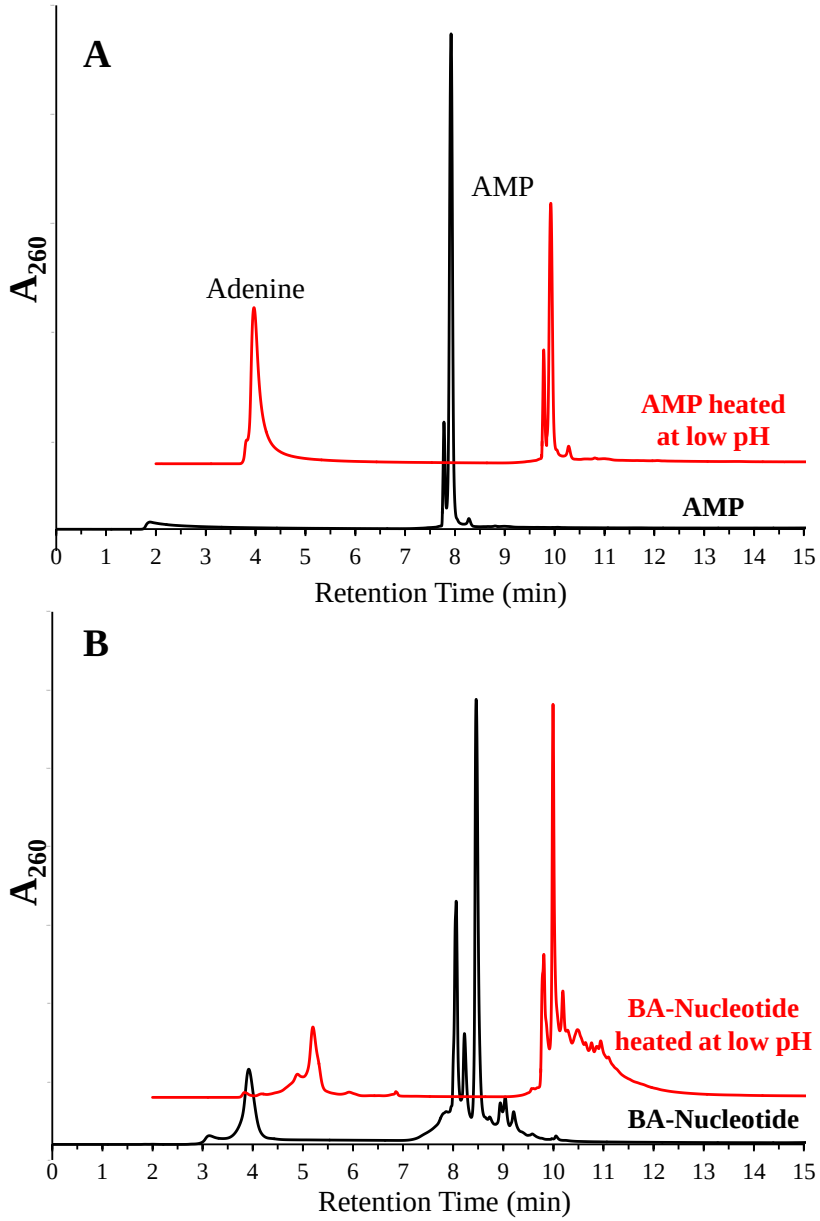


Figure 4.7 Stability of BA-nucleotide to acid catalysis. AMP (A) shows cleavage of the N-glycosidic linkage under acidic conditions, as demonstrated by peak in the dead volume corresponding to free adenine (red trace). BA-nucleotide reaction mixture (B) shows no noticeable change in the dead volume peak intensity even post heating, which indicates no obvious loss of the base moiety. Red traces in both the chromatograms is offset by 2 mins on the X axis and by about 10% on the Y axis for ease of visual comparison.

In conclusion, the BA-nucleotides were found to be more resistant to loss of base under acidic conditions, thus making them a more suitable candidate for polymerization under volcanic geothermal conditions. It has been argued that modern bases may have been selected under different selection pressures (Rios and Tor, 2012; Rios et al., 2015) at a later period during evolution, and that other bases like BA might have been more suited for incorporation into prebiotic polymers of, for e.g., a putative pre-RNA World (Hud et al., 2013).

4.4 Summary

Stability of nucleotides under potentially harsh prebiotic conditions is argued to have been a prominent selection pressure. The labile nature of N-glycosidic linkages in canonical nucleotides severely diminishes their stability under acidic conditions and high temperatures. Amongst modern nucleotides, purines were found to have a half-life of about 6-7 hours at pH 2. Such instability resulted in the formation of abasic sites in oligomers formed with acid-catalyzed esterification (Mungi and Rajamani, 2015). Persistent challenges in the formation of intact "informational" oligomers using modern nucleobases made us seriously consider the idea of chemical evolution as a plausible path that led to the formation of informational molecules of a putative RNA World. This idea posits that informational biomolecules may have evolved over time, with chemical stability under pertinent conditions acting as selection pressure that shapes this evolutionary path. Based on this hypothesis, few other studies have looked at alternate heterocycles that may result in nucleosides. Some of these nucleosides were proposed to be C-glycosides, which are known to be stable under acidic conditions. However, we were not aware of any studies when we started out that had looked at the formation of nucleotides using alternative heterocycles. We, therefore, decided to carry out nucleotide synthesis reactions with selected heterocycles under prebiotically relevant conditions i.e. dry heating. Successful synthesis of nucleotides was carried out with BA, TAP and DAHP heterocycles, as was demonstrated by analytical techniques such as HPLC and MS. Peaks obtained from HPLC indicated that multiple isomers (C/N or α/β nucleotides) were formed in all of these reactions. MS analysis confirmed the formation of nucleotides in all reactions. However, the reactions with TAP and DAHP resulted in adducts which possibly was made of one heterocycle and two rMP moieties. Due to formation of these sugar adducts (by virtue of their reaction with their exocyclic amine groups), only BA reaction was further analyzed in greater depth. In all, the formation of nucleotides with alternate heterocycles paralleled the observations made by other groups and

confirmed that non-canonical nucleobases could also form nucleotides and in relatively higher yields.

BA reaction was carried out under various conditions such as varying ratios of BA and rMP, varied reactions temperature and in the presence of divalent cations such as Mg^{2+}. The reaction was found to be robust as the nucleotide was found to form even at 60°C, albeit with lower yields, and almost complete incorporation of BA was observed at 1:5 (or higher) ratio of BA to rMP. Reactions carried out with Mg^{2+} resulted in greater yields of nucleotides and also seemed to change the ratio of isomers that was formed in the reaction. Extensive NMR characterization was carried out with purified BA-nucleotide mixture to assign chemical structure(s) to it. NMR confirmed the formation of both N- and C-glycosides with α-C-BA-nucleotide being the majority product (~65%) of the reaction. α-C-nucleotide formation, as seen in these reactions, was quite different from what is seen with modern nucleotides, which mostly result in β-N-nucleotides. BA-nucleotides were found to be stable under acidic conditions and minimal cleavage of glycosidic linkages was observed when heated at low pH and high temperature (unlike what is seen in AMP). This stability made BA-nucleotide a suitable candidate for oligomerization by DH-RH reactions as the informational moiety should potentially not get lost during the reaction. We performed oligomerization reactions with BA-nucleotide by subjecting it to similar conditions as discussed in Chapter 2 and characterized the products, the details of which are discussed in the next chapter.

4.5 References

- Anastasi, C., Crowe, M.A., Powner, M.W., and Sutherland, J.D. (2006). Direct assembly of nucleoside precursors from two- and three-carbon units. Angew. Chemie - Int. Ed. 45, 6176–6179.
- Bean, H.D., Sheng, Y., Collins, J.P., Anet, F.A.L., Leszczynski, J., and Hud, N. V. (2007). Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. J. Am. Chem. Soc. 129, 9556–9557.
- Becker, S., Thoma, I., Deutsch, A., Gehrke, T., Mayer, P., Zipse, H., and Carell, T. (2016). A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. Science. 352, 833 LP-836.

- Becker, S., Schneider, C., Okamura, H., Crisp, A., Amatov, T., Dejmek, M., and Carell, T. (2018). Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. Nat. Commun. 9, 163.
- Cafferty, B.J., and Hud, N. V (2015). Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs ? Insights from a Systematic Approach to the Origin of RNA. Isr. J. Chem. 55, 891–905.
- Cafferty, B.J., Avirah, R.R., Schuster, G.B., and Hud, N. V. (2014). Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK a matching of biomimetic monomers. Chem. Sci. 5, 9–10.
- Chen, M.C., Cafferty, B.J., Mamajanov, I., Gállego, I., Khanam, J., Krishnamurthy, R., and Hud, N. V (2014a). Spontaneous Prebiotic Formation of a β-Ribofuranoside That Self-Assembles with a Complementary Heterocycle. J. Am. Chem. Soc. 136, 5640–5646.
- Chen, M.C., Cafferty, B.J., Mamajanov, I., Gállego, I., Krishnamurthy, R., Hud, N. V., Khanam, J., Krishnamurthy, R., and Hud, N. V. (2014b). Spontaneous Prebiotic Formation of a β-Ribofuranoside That Self-Assembles with a Complementary Heterocycle. J. Am. Chem. Soc. 136, 5640–5646.
- Fuller, W.D., Sanchez, R. a, and Orgel, L.E. (1972). Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. J. Mol. Evol. 1, 249–257.
- Hud, N. V., Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The origin of RNA and “my grandfather’s axe”. Chem. Biol. 20, 466–474.
- Kim, H.-J., and Benner, S.A. (2017). Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. Proc. Natl. Acad. Sci. 114, 11315 LP-11320.
- Kim, H., and Benner, S.A. (2015). Prebiotic Glycosylation of Uracil with Electron-Donating Substituents. Astrobiology 15, 301–306.
- Kitadai, N., and Maruyama, S. (2017). Origins of building blocks of life: A review. Geosci. Front. 9, 1117–1153.

- Kolb, V.M., Dworkin, J.P., and Miller, S.L. (1994). Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. J. Mol. Evol. 38, 549–557.
- Mungi, C. V, and Rajamani, S. (2015). Characterization of RNA-Like Oligomers from Lipid-Assisted Nonenzymatic Synthesis: Implications for Origin of Informational Molecules on Early Earth. Life (Basel, Switzerland) 5, 65–84.
- Powner, M.W., Gerland, B., and Sutherland, J.D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459, 239–242.
- Powner, M.W., Sutherland, J.D., and Szostak, J.W. (2010). Chemoselective multicomponent one-pot assembly of purine precursors in water. J. Am. Chem. Soc. 132, 16677–16688.
- Rios, A.C., and Tor, Y. (2012). Refining the Genetic Alphabet: A Late-Period Selection Pressure? Astrobiology 12, 884–891.
- Rios, A.C., Yua, H.T., and Tor, Y. (2015). Hydrolytic fitness of N-glycosyl bonds: Comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. J. Phys. Org. Chem. 28, 173–180.

Chapter 5

Synthesis of Alternate Nucleobase Containing Oligomers using DH-RH Reactions

5.1 Introduction

Oligomerization of nucleotides to form RNA has been an extensively studied area, as mentioned in Chapter 1. The conventional models assume abiotic synthesis of nucleotides by chemical processes, and their subsequent oligomerization that results in nucleic acids. However, there are several issues with the formation of RNA by nonenzymatic oligomerization, especially when starting with non-activated nucleotides. Some of these are concerned with the formation of nucleotide monomers, as discussed in the previous chapter, while others impinge on the nature of polymers that are thought to have inhabited the RNA World. Modern RNA contains nucleobases that do not form H-bonds with each other in their monomeric state (Ts'o et al., 1963). If a nucleobase that is unable to form base pairing interactions (such as H-bonds) gets incorporated by oligomerization, it will result in an inability to transfer the encoded information. Selection based on H-bonding capability would not occur in this scenario as the monomers cannot be necessarily distinguished before oligomerization occurs. This will also result in the whole oligomer being obsolete as it could affect the folding of the molecule, which would be further compounded by the inability to remove that specific nucleotide (due to lack of enzymes). All these unfavorable processes would be detrimental to the process of structure formation and replication in an RNA World.

To address this issue, researchers have suggested the possibility that Watson-Crick base pairs might not have been the 'initial' nucleobases. Other combinations such as purine-purine or pyrimidine-pyrimidine base pairs could have preceded the current nucleobases (Engelhart and Hud, 2010). Non-canonical nucleobases are thought to be more suitable as informational moieties in a prebiotic context, as has been discussed in the previous chapter. Alternate heterocycles have been shown to form H-bonds with their base pairing partners even while in the monomeric form (Cafferty and Hud, 2014; Cafferty et al., 2013, 2014). BA and TAP, which have been used in this study, are shown to form higher order assemblies (hexads) by virtue of efficient H-bonding with melamine and cyanuric acid, respectively (Cafferty et al., 2016; Chen et al., 2014). As alternate bases seem to be more interesting in a pre-RNA world scenario, we looked at the oligomerization potential of mononucleotides containing BA as the nucleobase. DH-RH cycles were carried out with synthesized BA-nucleotides to check if BA containing oligomers could be synthesized and, if so, what might the nature of the products be.

Another model for the emergence of modern polymers is termed as the 'Polymer Fusion Model', as proposed in a perspective about the origin of RNA (Hud et al., 2013). The problem of lack of base pairing at monomer stage is circumvented in this model as it is proposed that non-covalent supramolecular assemblies of nucleobases readily result. Nucleobase analogues form higher order assemblies such as dyads or hexads, which are held together by H-bonds. In this model, the heterocycles are termed as recognition units (RU) as they perform the function of recognition of base pairing partners via H-bonding. This may result in higher order assemblies (hexads) as mentioned earlier. Another covalently linked polymer, consisting of an ionizable linker (similar to phosphates) and a trifunctional connector (like ribose sugar), acts as the backbone. If the spacing between two consecutive units in the stack of nucleobases (dyads/hexads) and the spacing of the trifunctional connector match, fusion of these can take place by formation of covalent bonds. This mechanism ensures that the nucleobases that are incorporated into the final informational polymer already have the ability to form H-bonds. Figure 5.1 shows the classical model (A) and the polymer fusion model (B) for polymerization as described in the review.

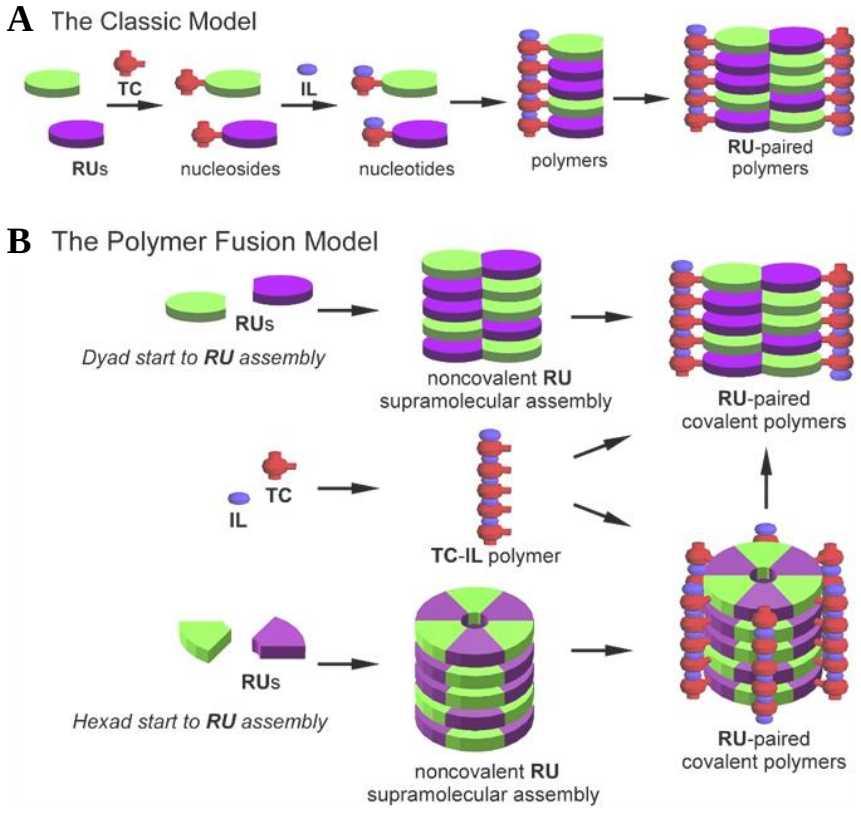


Figure 5.1: Two models for formation of informational oligomers on early Earth. (Adapted from Hud et al. 2013)

Formation of oligomers on the early Earth would have not been a uniform chemical reaction resulting in only one type of polymer. Furthermore, several polymers that resemble modern nucleic acids have been synthesized by chemical methods. These RNA analogues include Flexible Nucleic Acids (FNA), Glycerol Nucleic Acids (GNA), Threose Nucleic Acids (TNA) and Peptide Nucleic Acids (PNA), the structures of which are depicted in A. Some of these oligomers have been shown to form H-bonded structures with modern nucleotides (Engelhart and Hud, 2010). Therefore, the formation of informational molecules is not necessarily limited to the synthesis of molecules that are chemically identical to modern nucleic acids. A study has proposed a model for the emergence of polymers that can be produced from simple monomers, which also undergo degradation, resulting in prebiotically plausible repetitive cycles on the early Earth (Yakhnin, 2013). The study hypothesizes the formation of polymers by prebiotic processes, of which a fraction will be phosphodiester-linked polymers termed as phosphodiester backbones (PDB). The decay process of these PDB polymers, which can be fast or slow depending upon the nature of polymers, will replenish the monomer pool. The PDB polymers with slowest decay rates will accumulate over time and can also undergo transesterification type rearrangements, which may result in their improved stability (Figure 5.2 B). The study also hypothesizes prebiotic polymers containing various sugar moieties linked by phosphodiester linkages, which can get adsorbed onto mineral surfaces (Figure 5.2 C). Emergence of nucleic acid like polymers is hypothesized to be a result of evolution of these various PDB polymers.

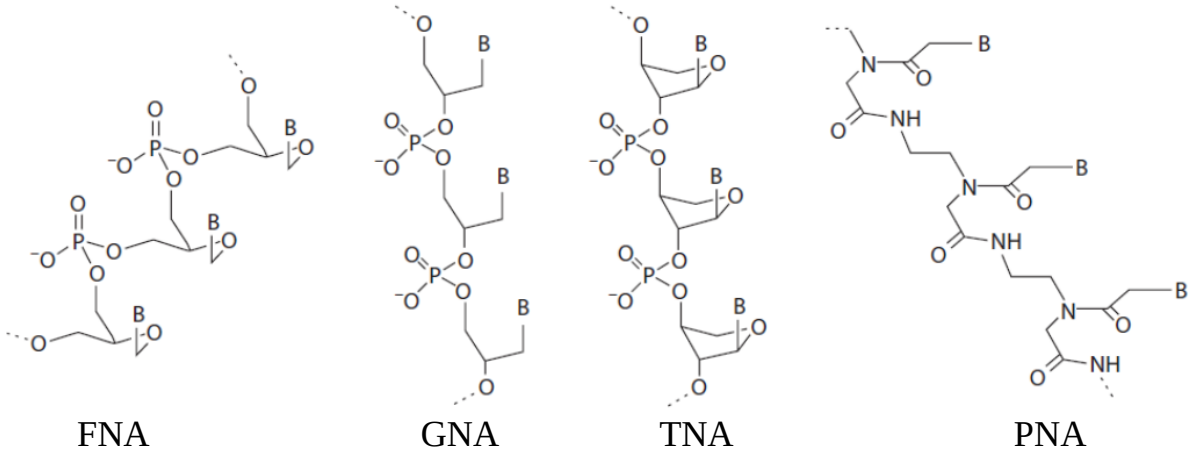


Figure 5.2 A: Structures of RNA analogues with different chemical backbones [Adapted from (Engelhart and Hud, 2010)].

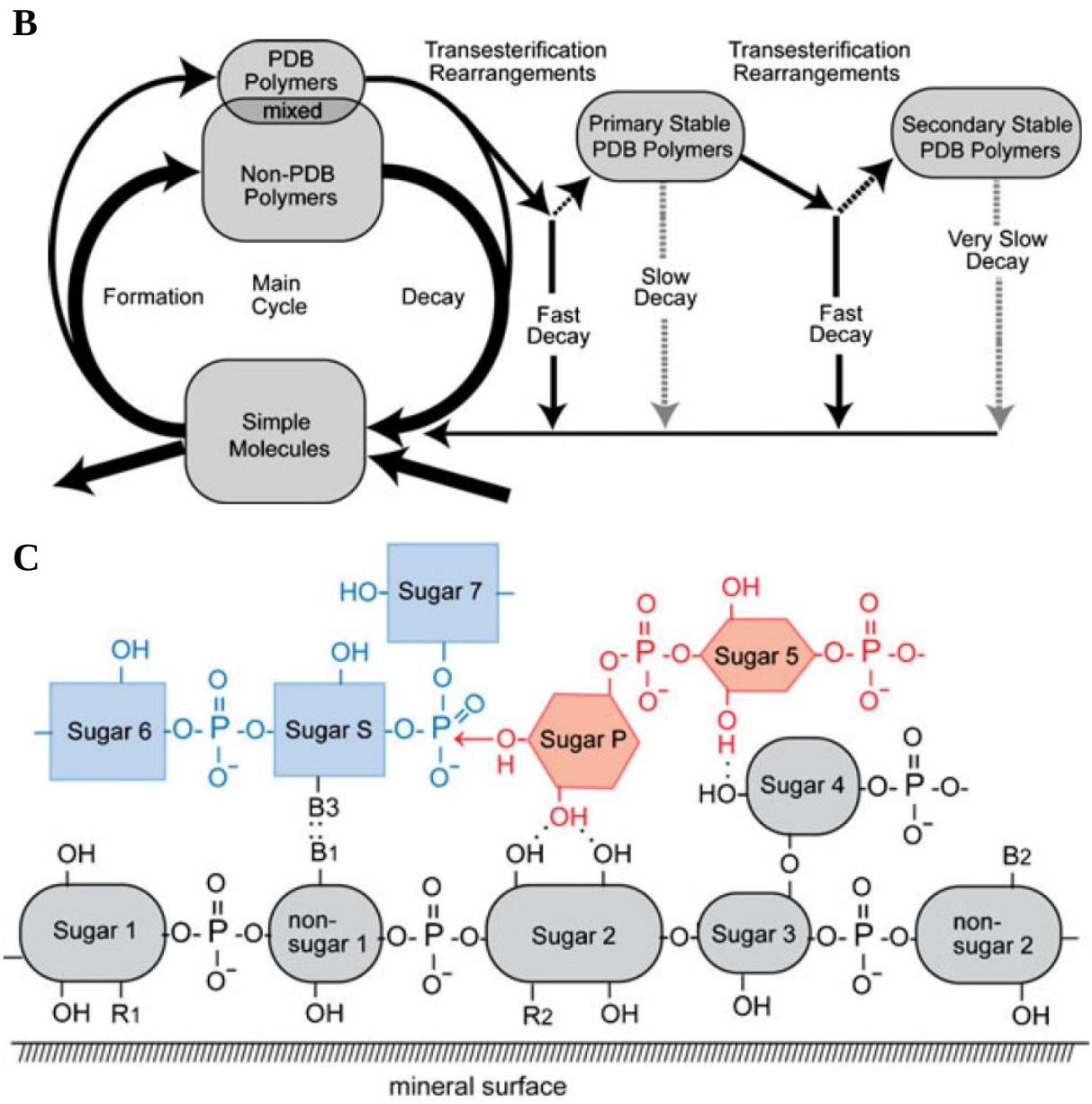


Figure 5.2: Models proposing the formation of stable PDB polymers based on rate of decay of the polymers (B); Hypothetical prebiotic polymer with diverse chemical structure, which may have acted as precursor to modern self-replicating molecules (C) [Adapted from (Yakhnin, 2013)].

In this chapter we present proof-of-concept results regarding attempts to synthesize oligomers with non-canonical nucleotide monomers, and the formation of polymers by the aforementioned alternate route.

5.2 Material and Methods

5.2.1 Materials

Ribose 5'-monophosphate (disodium salt) (rMP), sodium barbiturate (BA), 2,4-diamino-6-hydroxypyrimidine (DAHP), 2,4,6-triaminopyrimidine (TAP) and sodium perchlorate were purchased from Sigma Aldrich (Bangalore, India) and were used without further purification. The phospholipid, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), was purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA). All other reagents and buffers used for the studies were of analytical grade and purchased from Sigma-Aldrich (Bangalore, India).

Methods

5.2.2 Oligomerization of BA-Nucleotides by Dehydration-Rehydration (DH-RH) cycles

The BA-nucleotide synthesis reaction mixture, containing 5mM BA and 25mM rMP, was mixed with 1mM POPC and subjected to 7 DH-RH cycles of 1 hour each. These cycles were carried out in the set up mentioned in Section 2.2.2at 90°C, with the pH lowered to 2 using H_2SO_4, and Milli-Q water used as the rehydrating agent over the cycles. After 7 cycles, the dried sample was re-suspended in water and the lipids were separated using the butanol-hexane extraction protocol (as mentioned in Section 2.2.3). The aqueous phase was analyzed for the presence of oligomers by using HPLC and mass spectrometry as detailed below in Sections 5.2.5 and 5.2.6.

5.2.3 Oligomerization of ribose 5'-monophosphate (rMP)

Reaction mixture containing 5mM rMP and 1mM POPC was used for performing these backbone oligomerization experiments. pH of this reaction mixture was reduced to pH 2 using dilute H_2SO_4 and cycling was carried out at 90°C using Milli-Q water. Up to 10 DH-RH cycles, with 1 hour drying time were performed with the reaction mixtures, and samples were collected after 1, 3, 5, 7 and 10 cycles. Butanol-hexane extraction was performed on these samples and further analysis of the aqueous phase material was carried out using HPLC.

5.2.4 Reaction of alternate heterocycles with rMP oligomers.

We aimed to study the incorporation of alternate heterocycles on preformed sugar-phosphate oligomers that resulted from the formation of glycosidic bonds. Towards this goal, we carried out reactions with relevant bases and preformed rMP oligomers. Similar to previously discussed nucleotide formation experiments, alternate heterocycles like BA, TAP and DAHP were used to

check for the reaction of these bases with rMP oligomers. A similar reaction was also carried out with adenine as a control for comparison with canonical bases. Products from the 5mM rMP oligomerization reaction (post butanol- hexane extraction) were mixed with 1mM of the nucleobase, and the reaction mixture was dried at 90°C for 3 hours. After 3 hours, reaction mixtures were dissolved in Milli-Q water and analyzed by HPLC.

5.2.5 HPLC analysis

All the reaction mixtures were analyzed by HPLC using DNAPac PA 200 column, as mentioned in detail in Section 2.2.3. Analysis of BA-nucleotide oligomerization reaction was carried out by measuring the absorbance at 260nm with diode array detector (DAD). In case of analysis of reactions containing nucleobases and rMP oligomers, the DAD was set to the λ_{max} pertinent to the heterocycle that was used in the reaction. Since rMP does not have any absorbance in the UV range, the analysis of the rMP oligomerization reaction was carried out with the Fluorescence detector (FLD) module of the HPLC using λ_{ex} = 365 nm as the excitation wavelength and λ_{em} = 425 nm as the emission wavelength. This type of fluorescence was observed in the Maillard reaction in which sugars, when heated with other molecules, produced fluorescence with characteristic excitation maxima (340–370 nm) and emission maxima (420–470 nm) (Matiacevich et al., 2005).

5.2.6 Mass spectrometry

Mass analysis of BA-nucleotide oligomerization reaction was carried out at the Center for Applications of Mass Spectroscopy at the National Chemical Laboratory (NCL), Pune as detailed in Section 4.2.5. And, the mass spectrometry was performed on rMP oligomerization samples using the in-house HRMS facility as mentioned in Section 2.2.4. For all reactions, expected masses were calculated based on the most plausible chemical species and the ppm error was calculated for the observed masses and tabulated.

5.3 Results and Discussion

5.3.1 Oligomerization of BA-nucleotide by DH-RH reaction

BA-nucleotide that was synthesized under prebiotic conditions was found to be mainly α-C-glycosides on NMR analysis (Section 4.3.4). This, along with its stability analysis that was undertaken under acidic conditions, hinted at the likelihood of oligomerization of these nucleotides by DH-RH reactions, without the loss of base. To test this hypothesis, we subjected the BA-nucleotide mixture to oligomerization conditions, by lowering the pH to 2 using H_2SO_4, and performing dry-wet cycles at 90°C in the presence of POPC lipid. We used the reaction mixture as is, and not purified nucleotides, to simulate a one-pot reaction reminiscent of a real reaction niche. In this realistic scenario, progressive changes in the surrounding environmental conditions could have resulted in the synthesis of nucleotides under one condition, and their subsequent oligomerization in another condition, all facilitated in the same niche. Analysis of this reaction mixture after 7 DH-RH cycles showed multiple peaks in the gradient (Figure 5.3). These peaks were different from the peaks that were obtained for the isomers of BA-nucleotide. Interestingly, the retention times for some of the peaks in the gradient actually matched that of the oligomers that resulted from other related nucleotides (e.g. UMP) on the same gradient, indicating the formation of oligomers in the BA-reaction. No new peaks were observed in the dead volume indicating no breakdown of glycosidic bonds occurred during oligomerization.

(Figure 5.3 is on the following page)

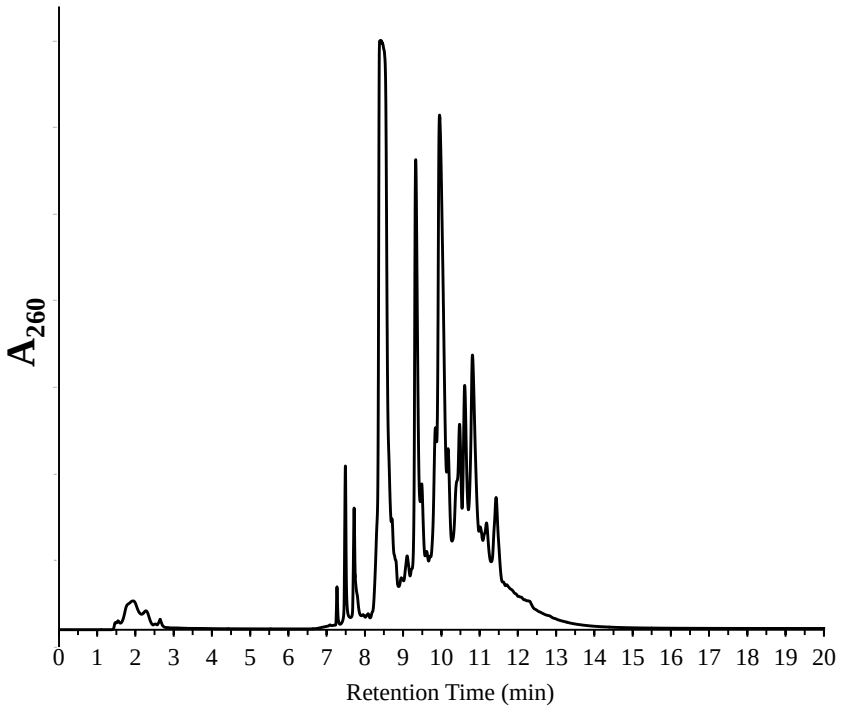


Figure 5.3: Oligomerization of BA-nucleotide using DH-RH reaction. Subjecting BA-nucleotide synthesis reaction mixture to seven dry-wet cycles, at low pH and high temperature, resulted in some higher oligomer formation as shown by HPLC.

Further analysis of the reaction sample by mass spectrometry confirmed the presence of oligomeric species. Apart from the monomers, mass numbers were observed for intact BA-nucleotide dimers and for BA-nucleotide-rMP dimer (abasic dimer). The masses observed in this analysis are summarized in Table 5.1. The abasic dimer seen in this reaction might not be due to loss of nucleobase because, in addition to no new peaks coming up in the dead volume, the mass for intact dimers were also observed. The presence of this abasic dimer population can actually be explained by the polymerization of a BA-nucleotide with unreacted rMP remaining from the original BA-nucleotide synthesis reaction. Another possible explanation could be the loss of base that could have occurred during ionization as observed in the mass analysis of UMP oligomerization reactions. These results, in many ways, were similar to the one obtained for canonical pyrimidines, but the enhanced stability of BA-nucleoside was apparent, which can be attributed to the presence of C-glycosidic bond.

Table 5.1: Mass numbers observed in the BA-nucleotide oligomerization reaction. The table summarizes the observed masses from the BA-nucleotide oligomerization reaction, with the details of the chemical species that these numbers corresponded to and the expected masses.

Chemical Species	Expected Mass	Observed Mass	Mass error (ppm)
BA-Nucleotide (BMP)	340.0308	340.0316	2.35
BMP Dimer	662.051	662.0518	1.21
BMP-rMP Dimer	552.0394	552.0405	1.99

The formation of oligomers with nucleotides containing alternate heterocycles by DH-RH reaction was demonstrated for the first time in the aforementioned experiments. Plausibility of oligomerization of non-canonical nucleotides suggests that a different class of covalently linked polymers could have been present on early Earth wherein alternate heterocycles could have been incorporated as the information carrying moiety. Another interesting observation was the potential oligomerization of rMP, resulting in the BA-nucleotide-rMP dimer. This suggested that acid-catalyzed ester bond formation of intact nucleotides could take place with rMP monomers, resulting in completely abasic oligomers. We tested this by performing DH-RH reactions with rMP monomers and analyzed the reaction products.

5.3.2 Oligomerization of ribose 5'-monophosphate (rMP)

Formation of phosphodiester bond in DH-RH reactions is hypothesized to take place by a mechanism that is independent of the nucleobase. This proposed mechanism involves the protonation of the phosphate group of one monomer, and nucleophilic attack by 2' or 3' OH group of a second monomer that results in the bond. We hypothesized that such reaction can take place with rMP monomers themselves, resulting in oligomers that resemble backbones of modern nucleic acids (i.e. ribose moieties linked by phosphodiester bonds). Crucially, formation of such oligomers by nonenzymatic means has implications for the synthesis of informational polymers by ‘Polymer Fusion Model’ (Hud et al., 2013), a very plausible ‘alternate’ model that has existed in the field to explain the emergence of a putative RNA World as discussed in introduction (Section 5.1). Towards this, DH-RH reactions were carried out with rMP as the starting reactants, using the same conditions that were standardized in Chapter 2, which were shown to promote oligomerization of AMP (pH 2 maintained with H_2SO_4, 90°C, 1mM POPC +

5mM rMP). The reaction products formed over various cycles were analyzed by HPLC using fluorescence detector (Figure 5.4). Green arrow indicates the retention time of nucleotide monomers on the same gradient (for comparison). Multiple peaks observed in the gradient suggested the formation of oligomers whose concentration seemed to increase with cycling. Chromatogram for cycle 0 (or rMP control) could not be generated as the fluorescence was only observed in samples that had undergone some heating. Dry heating of rMP, even for a very short duration, results in the formation of multiple products, as an isolated single peak was never observed in the gradient. This further complicated the assignment of the rMP monomer peak. Interestingly, oligomers of much longer lengths were observed in rMP reaction as opposed to reactions with canonical nucleotides, in which only up to tetramers were detected by HPLC (Section 2.3.5).

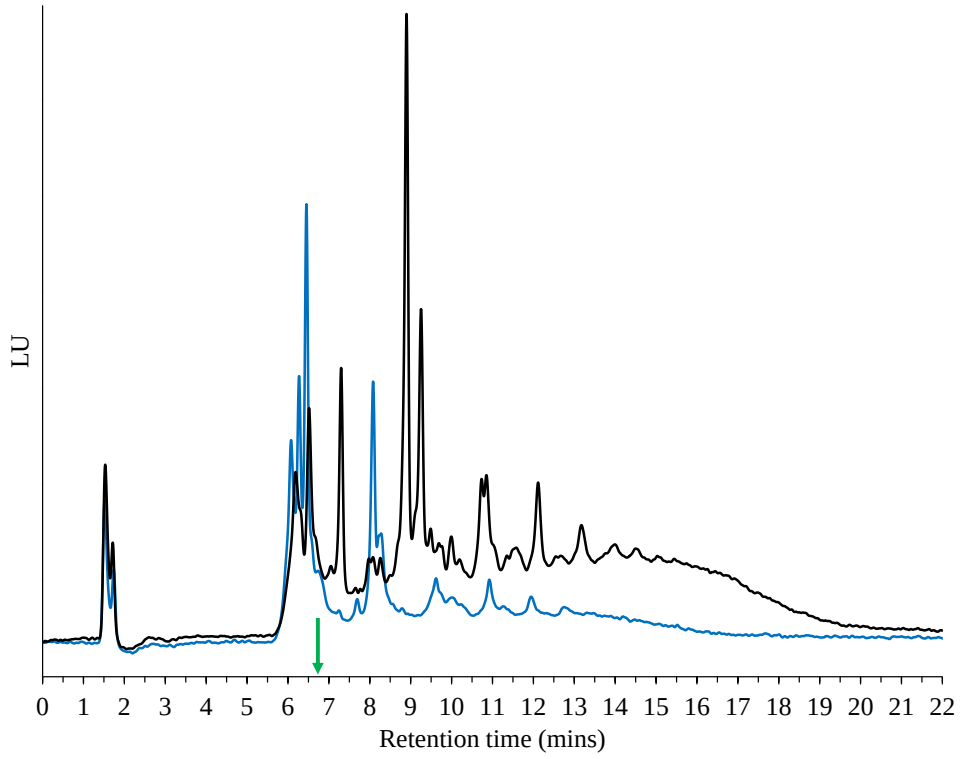


Figure 5.4: Oligomerization of ribose 5'-monophosphate (rMP) under DH-RH conditions. Chromatograms of cycle 1 (Blue) and cycle 10 (Black) of rMP oligomerization reaction as analyzed by HPLC. Oligomerization was seen to take place with just one DH-RH cycle, as seen in the blue chromatogram, which increased further with cycling (black chromatogram). Green arrow shows the retention time for 5'-AMP on the same gradient for comparison.

Mass analysis of rMP oligomerization reaction with HRMS showed peaks which corresponded to oligomers up to heptamers. Multiple peaks observed in cycle 10 reaction are shown in Figure 5.5, with clusters of peaks observed for each length of oligomer as designated. For each length of oligomer, apart from the linear isomer, mass numbers were observed for $[M - H_2O]^+$ in all reactions. These numbers usually correspond to loss of water during ionization but these may also correspond to cyclic oligomers. Cyclization of the oligomers may take place by a reaction between the phosphate on the 5' end and the 3'/2'-OH of the same oligomer. Such intramolecular strand cyclization is known to take place during nonenzymatic polymerization of nucleotides (Horowitz et al., 2010). Expected masses for oligomers of rMP and observed masses in these reactions have been summarized in Table 5.2.

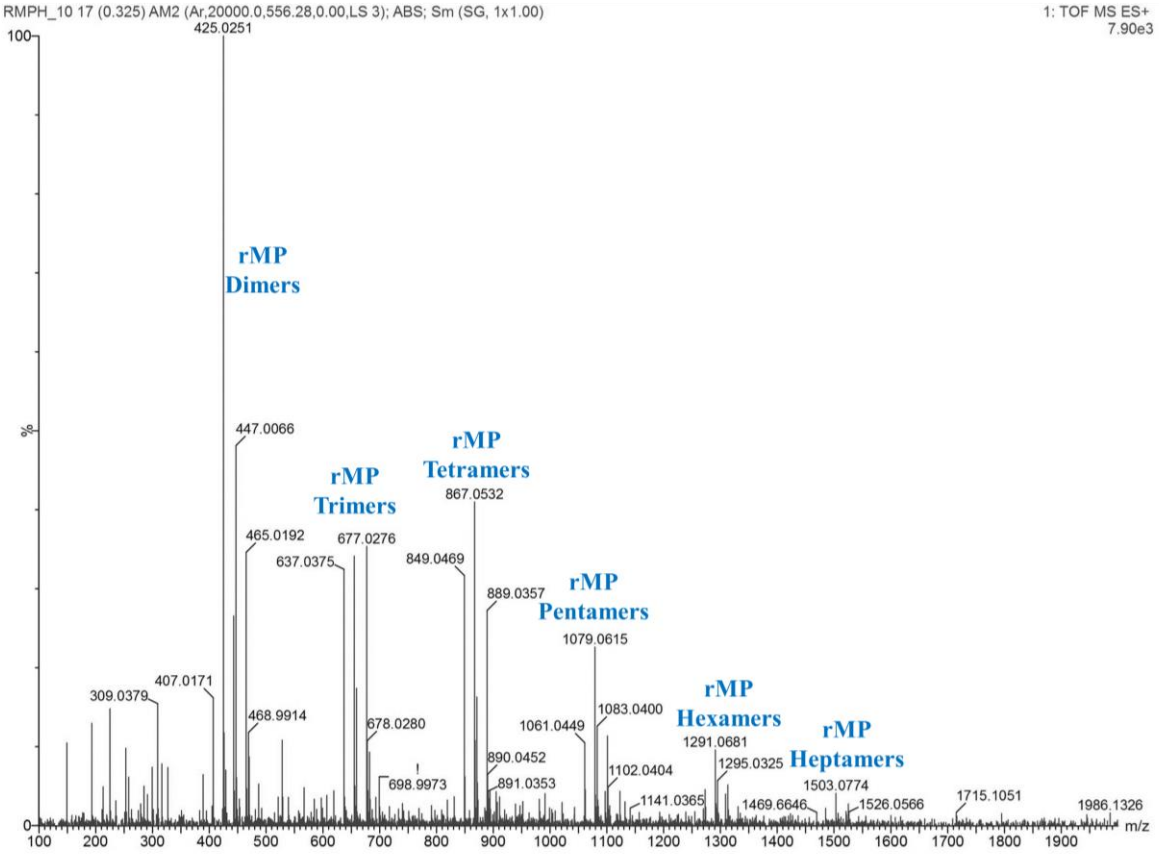


Figure 5.5: Mass profile of rMP oligomerization reaction (cycle 10). HRMS analysis of rMP reaction after ten DH-RH cycles shows clusters of mass numbers that correspond to longer oligomers, when compared to the ones seen in reactions with nucleoside monophosphate reactions.

Table 5.2: Mass numbers commonly observed in rMP reaction. Table summarizes chemical species, expected masses and observed masses for species observed in rMP oligomerization reaction when DH-RH cycles were performed at low pH and high temperature.

Chemical Species	Expected mass	Observed mass
Cyclic rMP Dimer	425.0244	425.0251
Linear rMP Dimer	443.0350	443.0363
Cyclic rMP Trimer	637.0330	637.0375
Linear rMP Trimer	655.0436	655.0424
Cyclic rMP Tetramer	849.0416	849.0469
Linear rMP Tetramer	867.0522	867.0532
Cyclic rMP Pentamer	1061.0502	1061.0449
Linear rMP Pentamer	1079.0608	1079.0615
Cyclic rMP Hexamer	1273.0588	1273.0591
Linear rMP Hexamer	1291.0694	1291.0681
Linear rMP Heptamer	1503.0780	1503.0774

Further characterization of these oligomers was not possible as the reaction products could not be obtained in quantities that were required for various analytical techniques. Low yields of the oligomerization reaction, and lack of a UV-absorbing moiety on these oligomers, resulted in unsuccessful attempts at further purification of the reaction products. Other experimental schemes involving derivatization of rMP for absorbance, using 1-phenyl-3-methyl-5-pyrazolone, or analysis using other techniques are under consideration. These preliminary results pertaining to oligomerization of rMP demonstrated the plausibility of formation of sugar-phosphate backbones on prebiotic Earth. rMP polymers could be considered a proxy for the covalently linked backbones as hypothesized in the ‘Polymer Fusion model’ of prebiotic informational molecule synthesis. Furthermore, the potential addition of nucleobases to these preformed oligomers, by the formation of a glycosidic bond between heterocycles and 1'-OH of rMP in the oligomers, was an interesting possibility. This idea was tested by performing experiments in which alternate heterocycles (e.g. BA) and preformed rMP oligomers were subjected to dry heating and the resultant products were analyzed.

5.3.3 Addition of alternate heterocycles on to preformed oligomers

Conventional models of nucleic acid formation assume the synthesis of individual monomers from their components, and subsequent polymerization of these monomers under pertinent conditions, resulting in oligomers such as RNA or DNA. As stated in the introduction (Section 4.1), synthesis of canonical monomers has been challenging due to factors such as limited availability of phosphates, difficulty in the selection of ribose over other sugars and the formation of N-glycosidic bond between ribose and nucleobases. This has led to a renewed focus on alternate models, one of which proposes the formation of nucleic acid and related molecules by the fusion of two or more components, one of these components being a covalently linked backbone polymer. However, to our knowledge, currently there are no studies that have tested this hypothesis by conducting experiments with preformed polymers and informational heterocycles. Our previous results showed that the polymerization of rMP results in oligomers that resemble the backbone of modern nucleic acids (Section 5.3.2). Additionally, alternate heterocycles, such as BA, DAHP etc., were also shown to form glycosidic linkages with rMP monomers (Section 4.3.1). Therefore, we investigated whether addition of these heterocycles was possible onto preformed rMP oligomers under dry heating conditions. In order to test this, the products from rMP oligomerization reaction (i.e. the aqueous phase of cycle 10 reaction) were mixed with alternate heterocycles (like BA, TAP and DAHP), and heated under dry conditions. Figure 5.6 shows the HPLC chromatograms of the reaction for all the three alternate heterocycles that we tested, as analyzed using a DNAPac column.

Free bases elute in the dead volume as they do not interact with the column due to lack of negative charge, while the rMP oligomers elute in the gradient but cannot be observed as they do not have any absorbance in the UV range. Hence, the peaks that could be observed in the gradient of these reactions can mainly be explained as a product that would have resulted from the covalent attachment of the heterocycles with the rMP oligomers. Formation of such oligomers occurred with varying efficiency in the three heterocycles tested, as seen from the HPLC profiles. BA reaction seems to have resulted in an efficient incorporation of bases amongst the reactions tested, and some of these peaks had retention times similar to the BA-nucleotide oligomers that were observed in the previous reactions (Section 5.3.1). This proof-of-concept result demonstrates the plausibility of informational molecules potentially being synthesized by the fusion of covalently linked backbone moieties and individual nucleobases.

Further analysis of these reactions has been challenging due to the as yet uncharacterized nature of the preformed rMP oligomers, which was used as one of the starting materials. This was further compounded by the fact that the rMP oligomerization reaction could not be performed in a controlled manner due to the complexity inherent to the reaction (stemming from multiple reactive groups), resulting in products that could not be isolated to collect dimers/trimers of rMP. Further experiments, to conduct targeted polymer fusion experiments, using specific synthetic backbone oligomers, are ongoing.

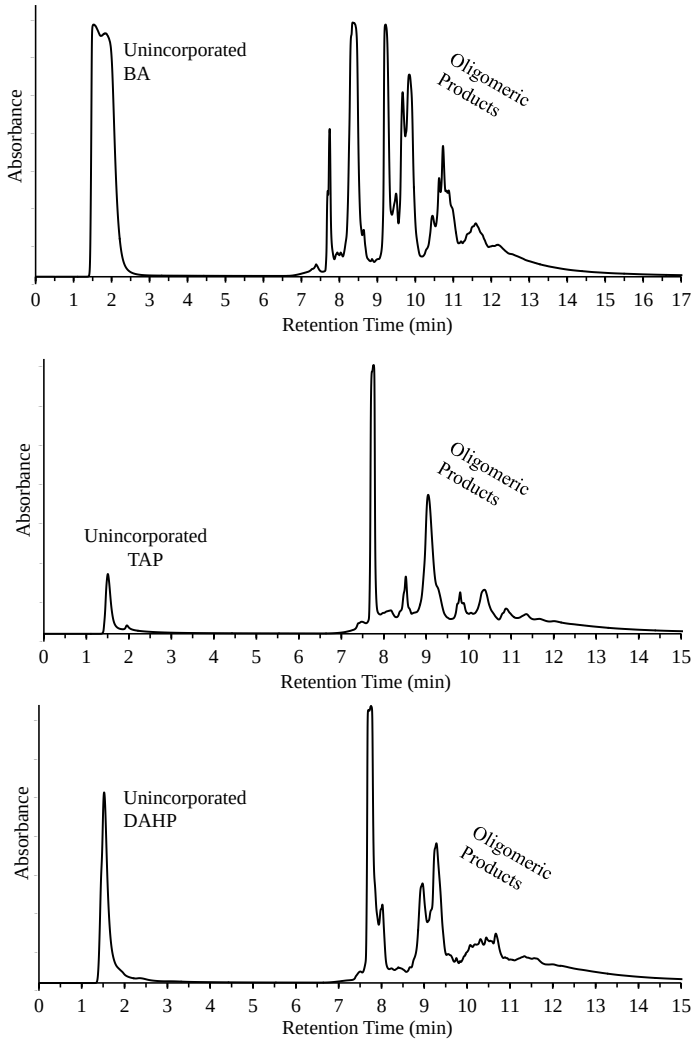


Figure 5.6 Reaction of alternate heterocycles with preformed rMP oligomers. Dry heating reactions were carried out with rMP oligomers and alternate heterocycles such as BA (A), TAP (B) and DAHP (C), and analyzed using ion-exchange HPLC. Presence of several peaks in the gradient indicates the formation of covalently linked UV-absorbing oligomeric products, with heterocycles incorporated in them.

5.4 Summary

Conditions that resulted in the oligomerization of non-activated canonical nucleotides were also found to promote oligomerization of the BA-Nucleotide. DH-RH reactions that were carried out with the BA-nucleotide synthesis reaction mixture, in the presence of lipids, at low pH and high temperatures, resulted in oligomers containing BA. Presence of these oligomers was demonstrated by both HPLC and Mass Spectroscopy (Section 4.3). The lack of a peak in the dead volume, in addition to the observed mass numbers from the sample, indirectly confirmed the formation of intact dimers in this reaction. The stability of the glycosidic linkage under acidic conditions could be partially attributed to the C-glycosides that were observed in case of the BA-nucleotides (discussed in in Section 4.3.5). Though canonical pyrimidines such as UMP also showed intact oligomers in DH-RH reactions, the formation of U containing nucleosides was not observed readily in Orgel's original reactions (Fuller et al., 1972). On the other hand, both the synthesis and oligomerization of BA-nucleotides, could be demonstrated under prebiotically plausible conditions in our reactions. This makes a strong case for alternate heterocycles being more interesting as candidates of informational carrying moieties in a pre-RNA World scenario.

As the mechanism of nonenzymatic polymerization of nucleotides by acid-catalyzed esterification, appears to be independent of the base, we hypothesized that an abasic monomer of nucleotide (rMP), could also undergo oligomerization by this mechanism. Towards this goal, oligomers that resemble the backbones of modern nucleic acids were synthesized by oligomerization of ribose 5'-monophosphate (rMP), under DH-RH reactions. Multiple peaks were observed in the DH-RH reactions carried out with rMP when analyzed with HPLC. Mass analysis of these reactions confirmed the presence of oligomers and also suggested possible cyclization in the newly formed oligomers. Synthesis of such oligomers was demonstrated for the first time by our study and has implications regarding the types of oligomers that may have been present in a pre-RNA world.

Finally, covalent attachment of alternate heterocycles to rMP oligomers was seen to take place in reactions under relatively simple conditions of dry heating. When preformed rMP oligomers were heated with BA, TAP and DAHP, multiple peaks were observed in the HPLC gradient. These peaks indicated the formation of chemical species, which have multiple phosphodiester bonds, with heterocycles attached to them as was evident due to the UV absorbance. This proof-

of-concept result implies the possibility that covalently linked backbone oligomers may have reacted with heterocycles present in the surrounding milieu, potentially resulting in primitive informational polymers. This model of synthesis of informational polymers eliminates the necessity for complete monomer synthesis that then subsequently undergoes polymerization (Classical model). Importantly, the Alternate model that we provide proof for, is also considered to have allowed for the 'sampling' of various heterocycles with prebiotically pertinent properties such as H-bonding. Taken together, our aforementioned set of results substantiates a plausible pathway for the emergence of molecules of a putative pre-RNA World(s). Covalently linked backbone oligomers could have incorporated alternate informational moieties (by way of formation of glycosidic linkages), resulting in primitive information polymers prior to the emergence of an RNA World.

5.5 Conclusions

Formation of RNA like molecules by lipid-assisted, acid-catalyzed oligomerization of nucleotides was studied in detail. We conducted oligomerization experiments with canonical nucleoside monophosphates (AMP, UMP, GMP and CMP) in the presence of lipids, by performing dehydration-rehydration cycles under prebiotically relevant volcanic geothermal conditions. Abasic sites were observed in the resultant oligomers due to acid-labile N-glycosidic linkages. We, therefore, synthesized nucleotides with alternate heterocycles, such as BA, TAP and DAHP, under simple dry heating conditions. Resultant nucleotides were characterized further using pertinent analytical techniques, with the structure of BA-nucleotide thoroughly ascertained. Subsequently, BA-nucleotides were subjected to oligomerization reactions and the oligomers were shown to retain the bases. Synthesis of oligomers that resemble backbones of modern nucleic acids was also performed under the aforementioned reaction regime. Alternate heterocycles reacted more efficiently with these backbone oligomers, as against canonical nucleobases. This implied a greater fitness for non-canonical nucleotides, qualifying them as candidates for formation of early informational polymers under the harsh conditions of early Earth. In conclusion, results from this body of work highlights drawbacks in the conventional model proposed for the emergence of polymers in an RNA World. Furthermore, it also strengthens the hypothesis that modern biopolymers might have evolved from chemically distinct primitive informational polymers that might have populated the pre-RNA World(s).

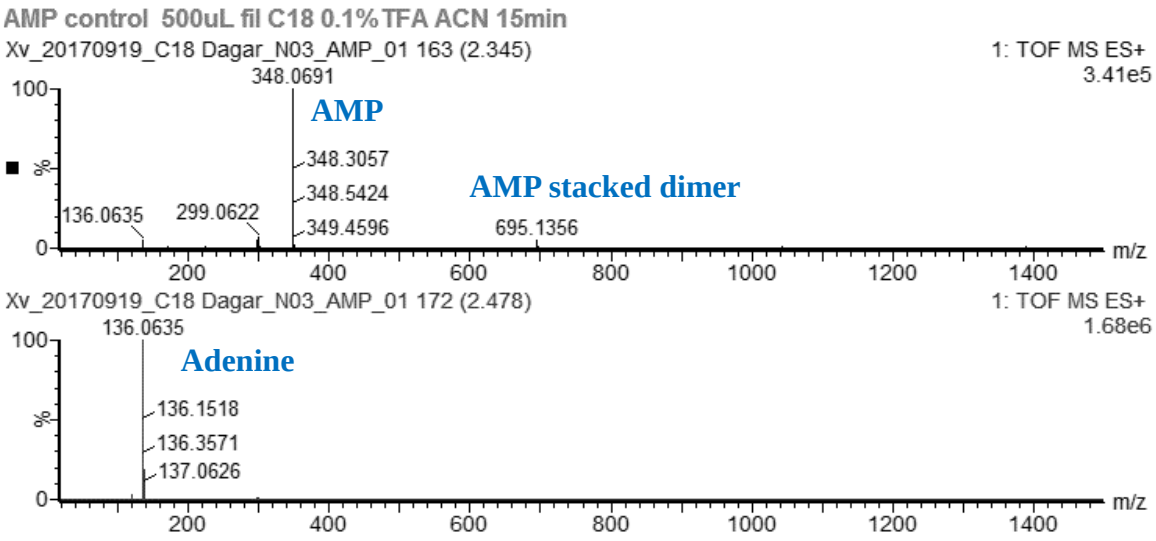
5.6 References

- Cafferty, B.J., and Hud, N. V. (2014). Abiotic synthesis of RNA in water: A common goal of prebiotic chemistry and bottom-up synthetic biology. Curr. Opin. Chem. Biol. 22, 146–157.
- Cafferty, B.J., Gállego, I., Chen, M.C., Farley, K.I., Eritja, R., and Hud, N. V. (2013). Efficient Self-Assembly in Water of Long Noncovalent Polymers by Nucleobase Analogues. J. Am. Chem. Soc. 135, 2447–2450.
- Cafferty, B.J., Avirah, R.R., Schuster, G.B., and Hud, N. V. (2014). Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK_a matching of biomimetic monomers. Chem. Sci. 5, 9–10.
- Cafferty, B.J., Fialho, D.M., Khanam, J., Krishnamurthy, R., and Hud, N. V. (2016). Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nat. Commun. 7, 11328.
- Chen, M.C., Cafferty, B.J., Mamajanov, I., Gállego, I., Khanam, J., Krishnamurthy, R., and Hud, N. V (2014). Spontaneous Prebiotic Formation of a β-Ribofuranoside That Self-Assembles with a Complementary Heterocycle. J. Am. Chem. Soc. 136, 5640–5646.
- Engelhart, A.E., and Hud, N. V. (2010). Primitive genetic polymers. C.S.H. Perspect. Biol. 2.
- Fuller, W.D., Sanchez, R. a, and Orgel, L.E. (1972). Studies in prebiotic synthesis: VII. Solid-state synthesis of purine nucleosides. J. Mol. Evol. 1, 249–257.
- Horowitz, E.D., Engelhart, A.E., Chen, M.C., Quarles, K. a, Smith, M.W., Lynn, D.G., and Hud, N. V (2010). Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. Proc. Natl. Acad. Sci. U. S. A. 107, 5288–5293.
- Hud, N. V., Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The origin of RNA and “my grandfather’s axe”. Chem. Biol. 20, 466–474.
- Matiacevich, S.B., Santagapita, P.R., and Buera, M.P. (2005). Fluorescence from the maillard reaction and its potential applications in food science. Crit. Rev. Food Sci. Nutr. 45, 483–495.
- Ts’o, P.O.P., Melvin, I.S., and Olson, A.C. (1963). Interaction and Association of Bases and Nucleosides in Aqueous Solutions. J. Am. Chem. Soc. 85, 1289–1296.
- Yakhnin, A. V (2013). A Model for the Origin of Life through Rearrangements among Prebiotic Phosphodiester Polymers. Orig. Life Evol. Biosph. 43, 39–47.

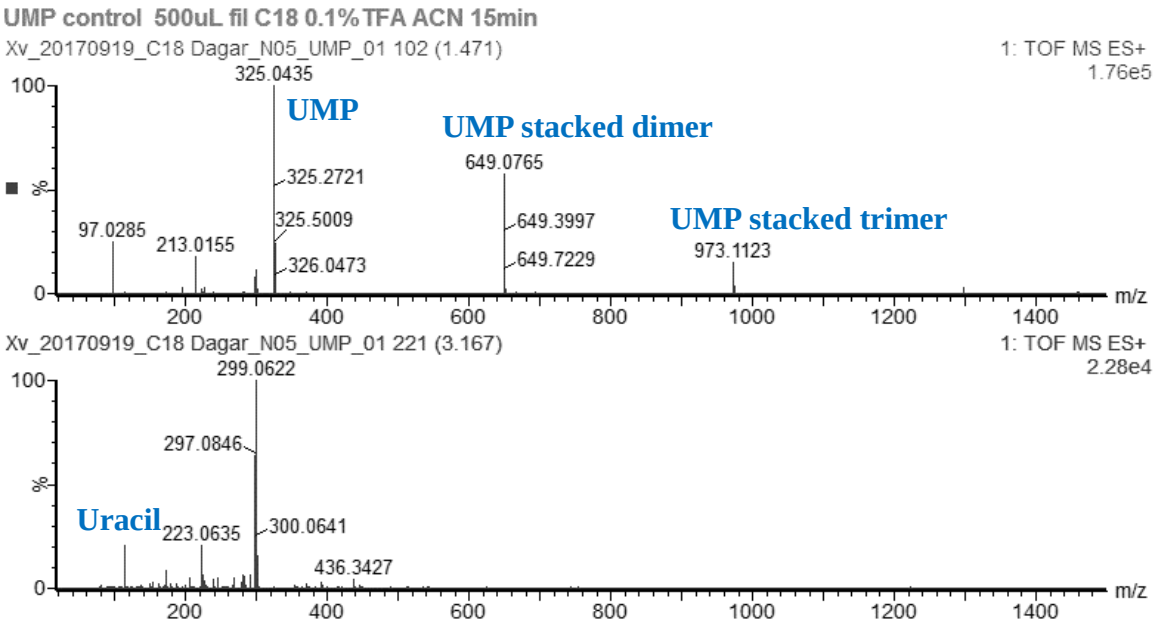
APPENDIX

Following mass spectra were acquired at ELSI, Japan in collaboration with Dr. Hongo. Samples were analyzed as described in Section 3.2.5.

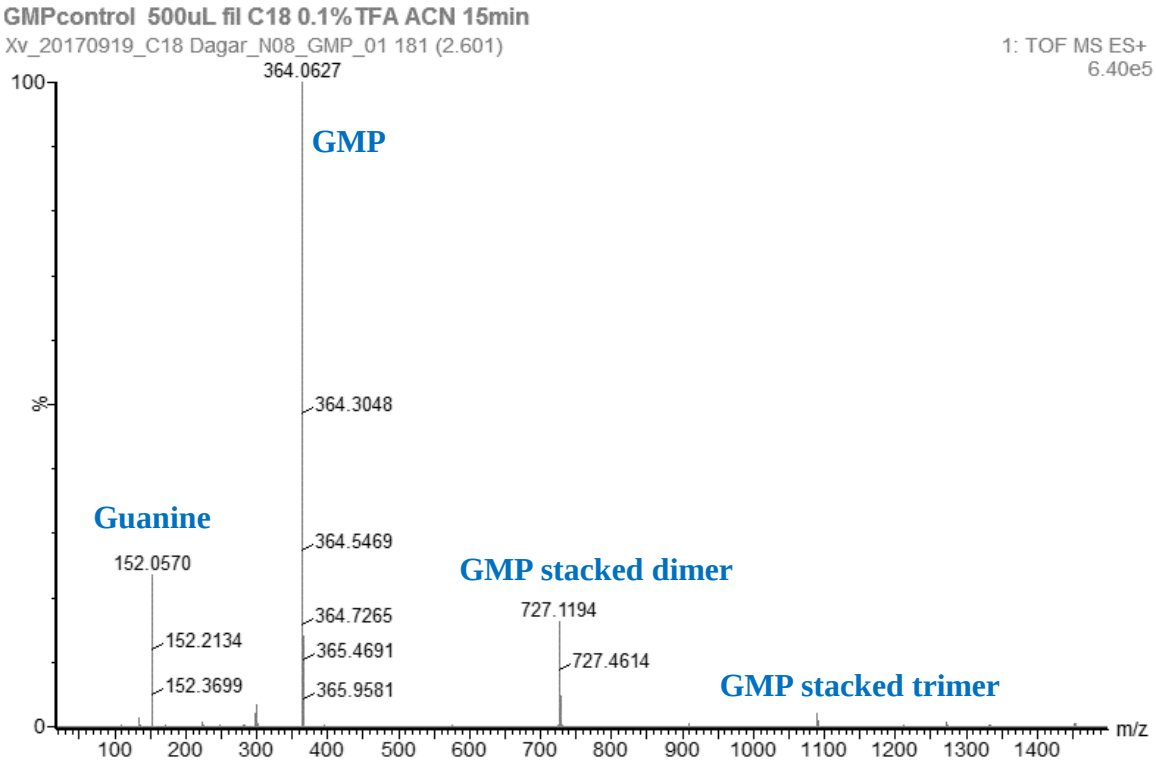
A.1: Mass spectrum of AMP control



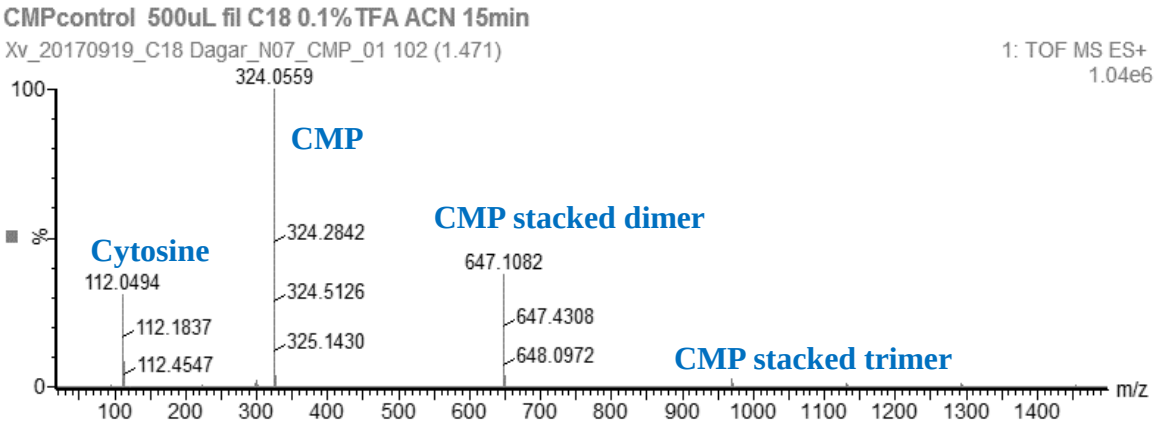
A.2: Mass spectrum of UMP control



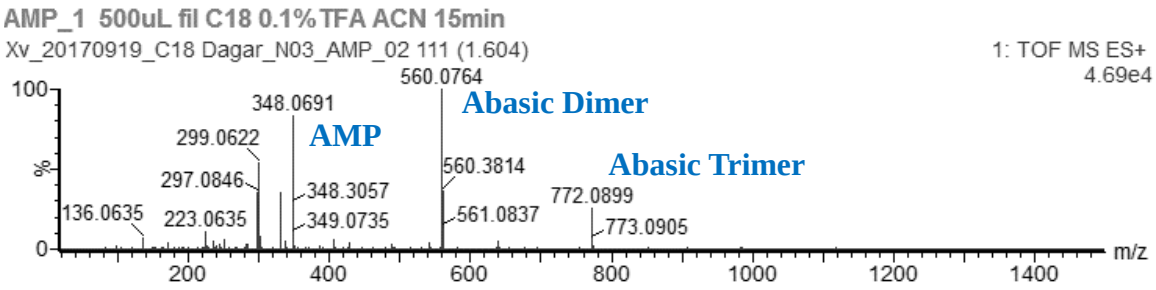
A.3: Mass spectrum of GMP control



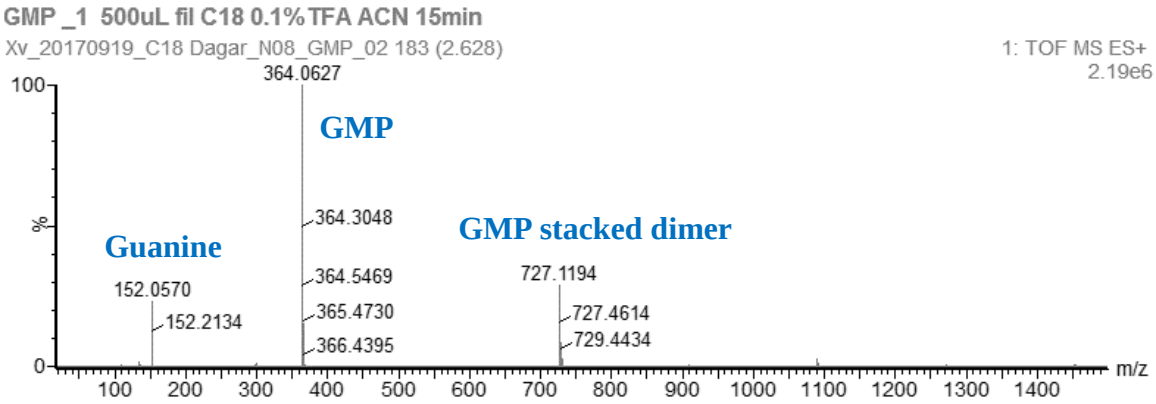
A.4: Mass spectrum of CMP control

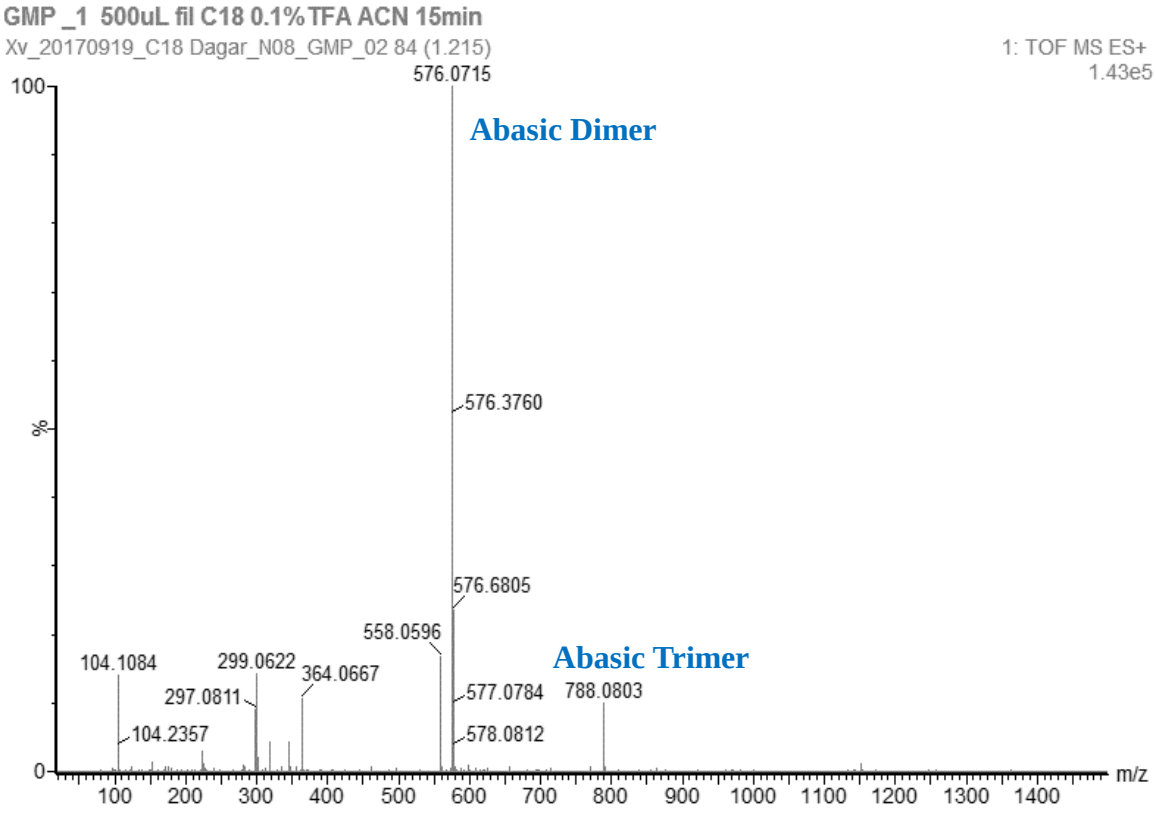


A.5: Mass spectrum of AMP oligomerization reaction

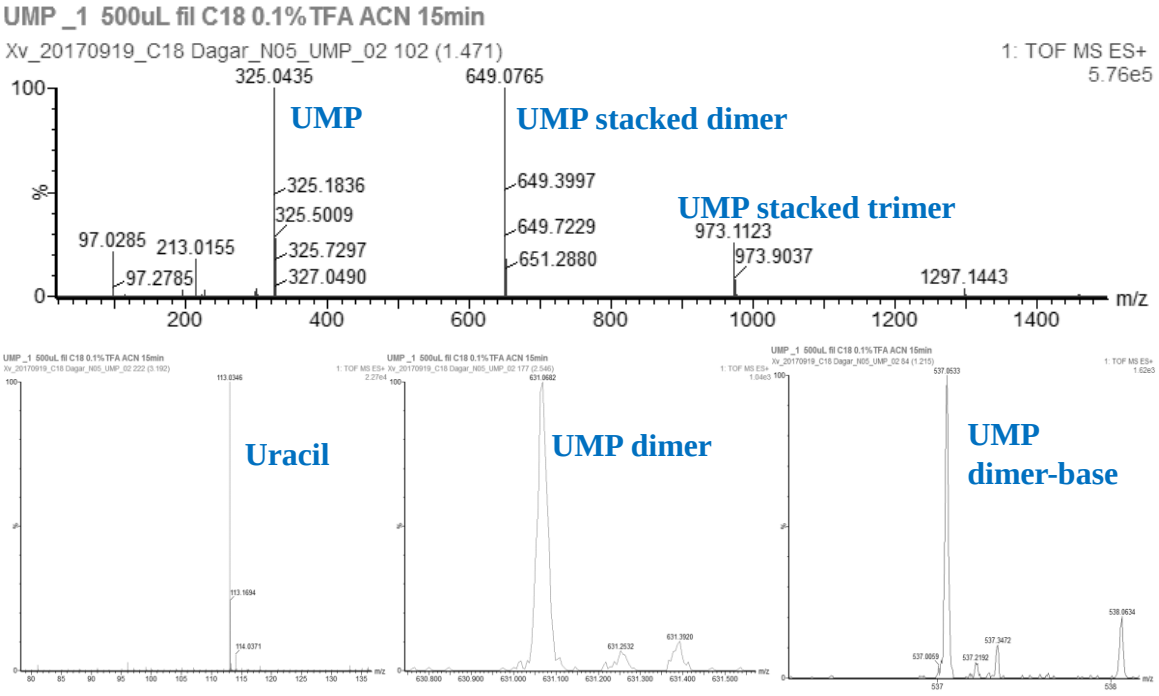


A.6: Mass spectrum of GMP oligomerization reaction

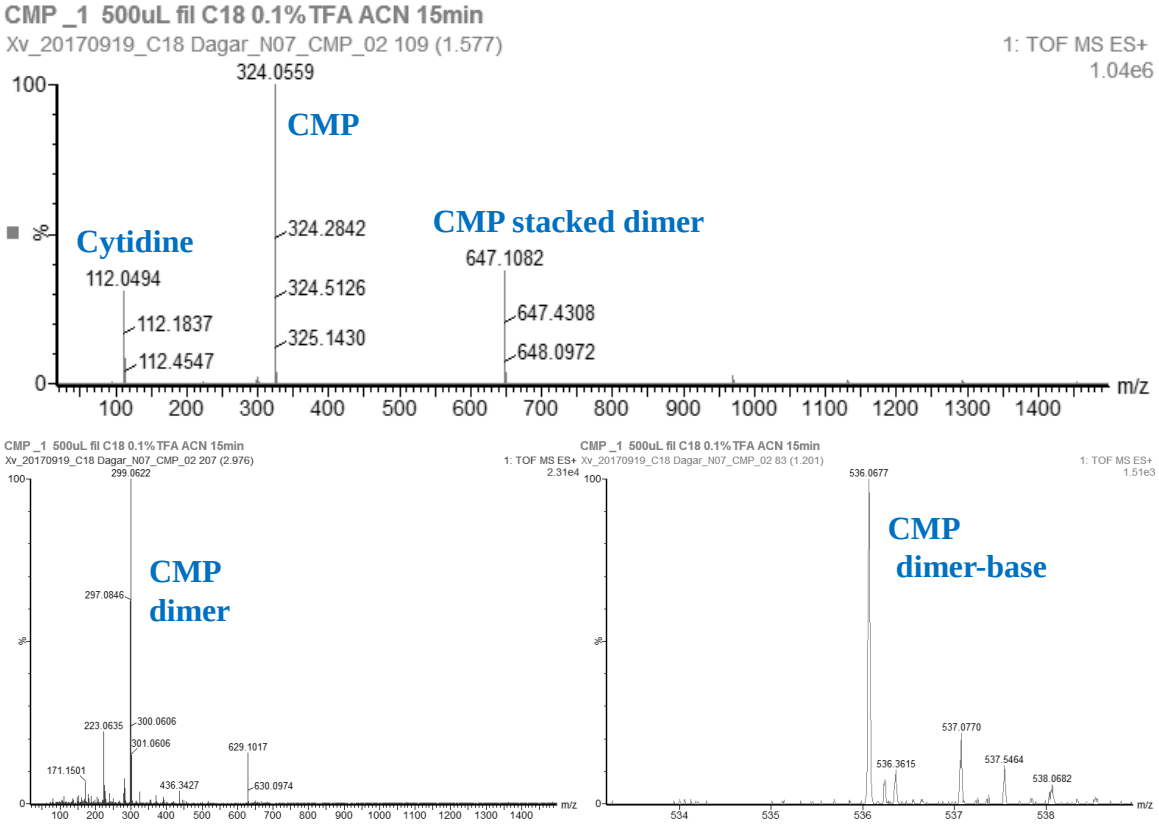




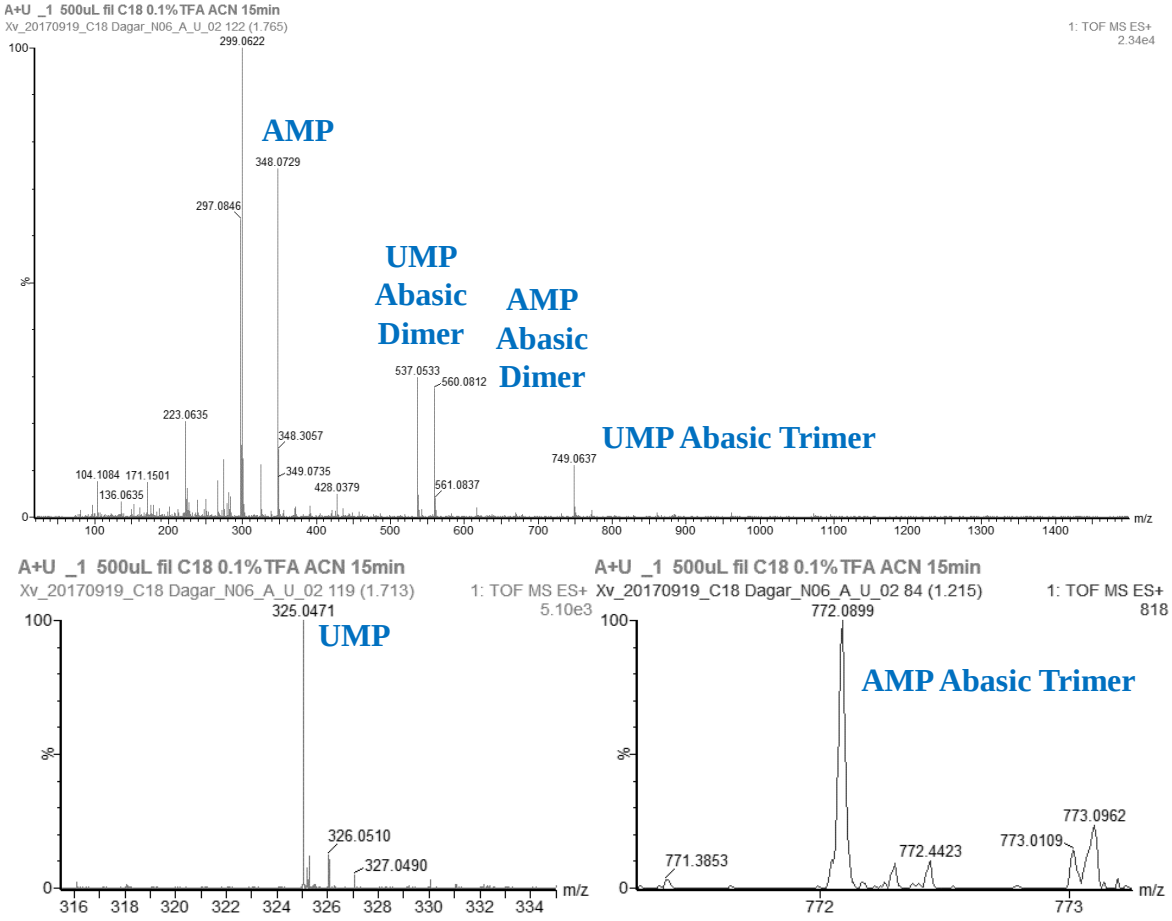
A.7 Mass spectrum of UMP oligomerization reaction



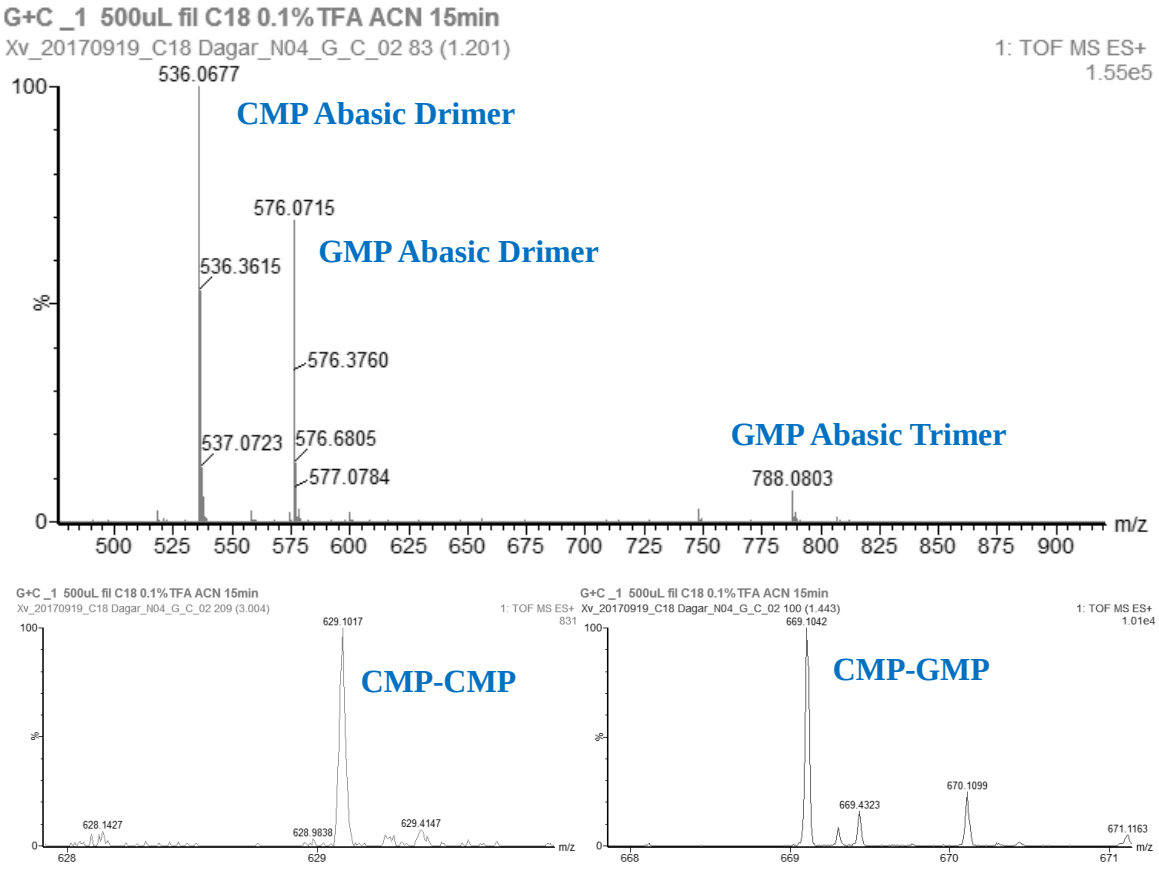
A.8 Mass spectrum of CMP oligomerization reactions



A.9: Mass spectrum of AMP+UMP reactions

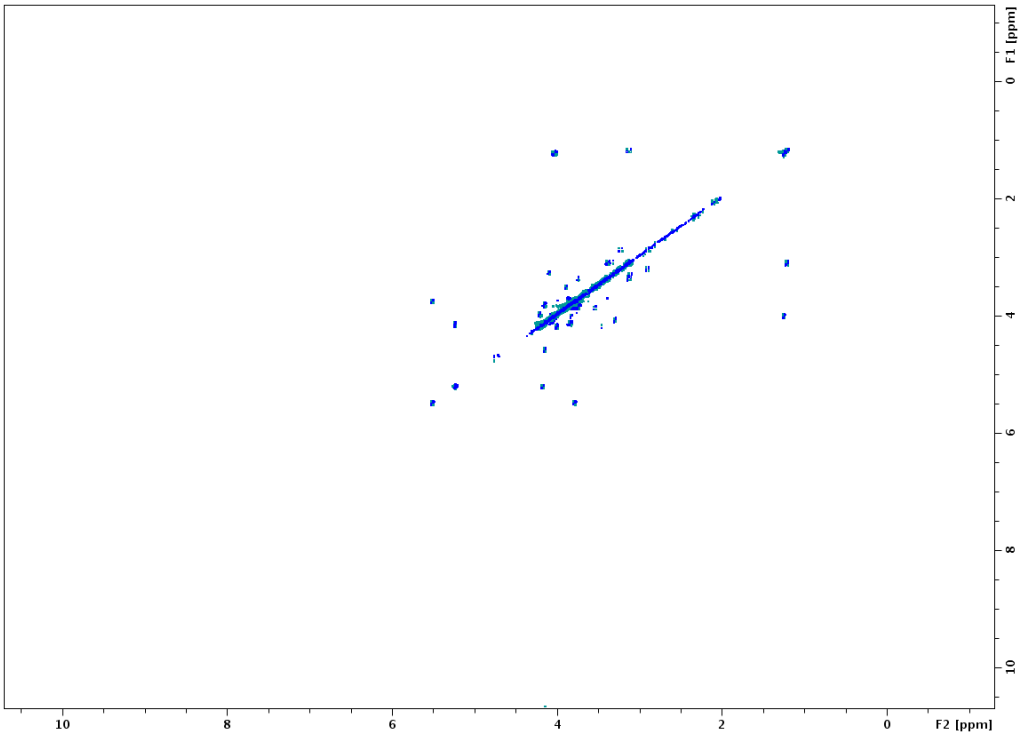


A.10: Mass spectrum of GMP+CMP reactions

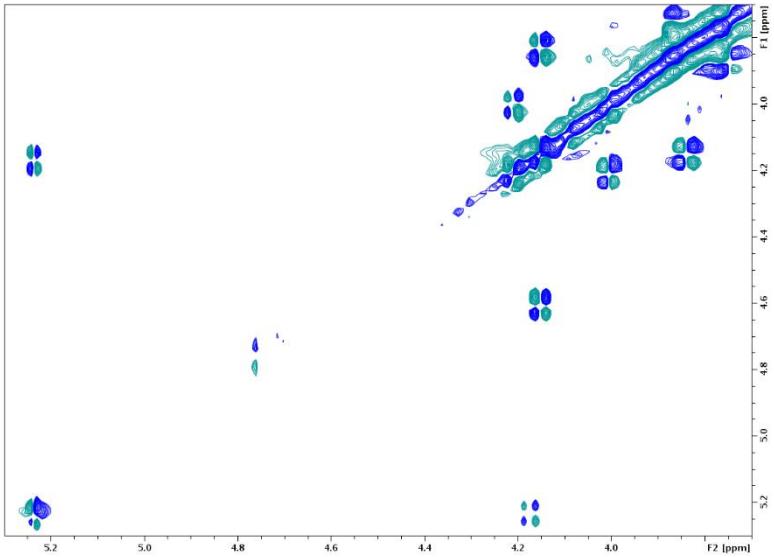


Following NMR spectra were acquired in collaboration with Dr. Jeetender Chugh (IISER Pune). The details of acquisition are mentioned in section 4.2.6. These represent full spectrums for the various analysis, of which selected part were used in Figure 4.5.

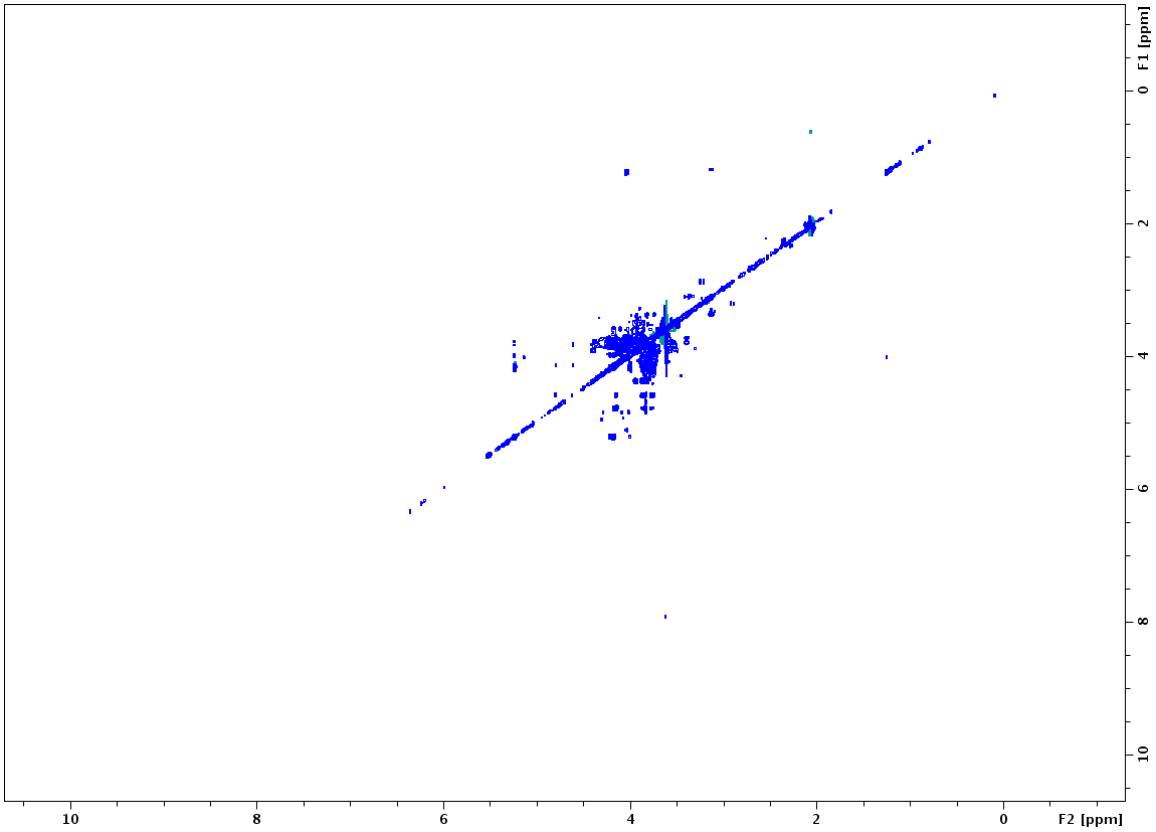
A.11: Full NMR Spectrum of ^{1}H-^{1}H COSY for synthesized and purified BA-Nucleotide (Figure 4.5 A).



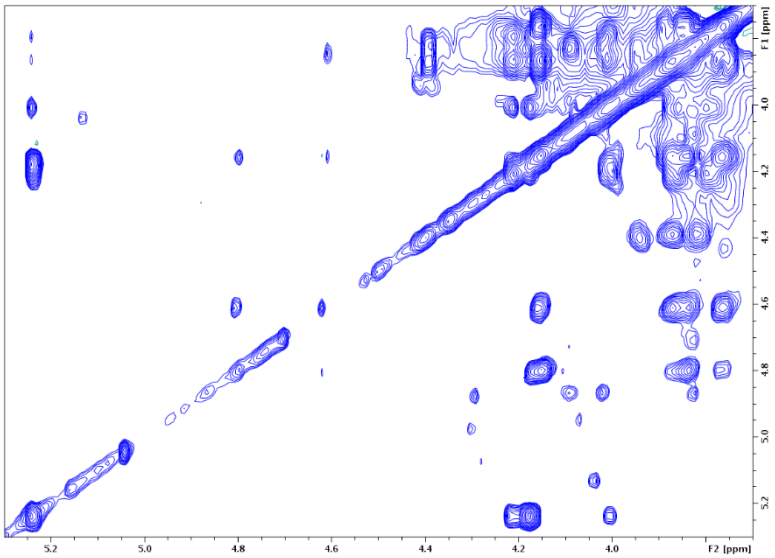
A.12: Zoomed in spectrum of ^{1}H-^{1}H COSY for synthesized and purified BA-Nucleotide (Figure 4.5 A).



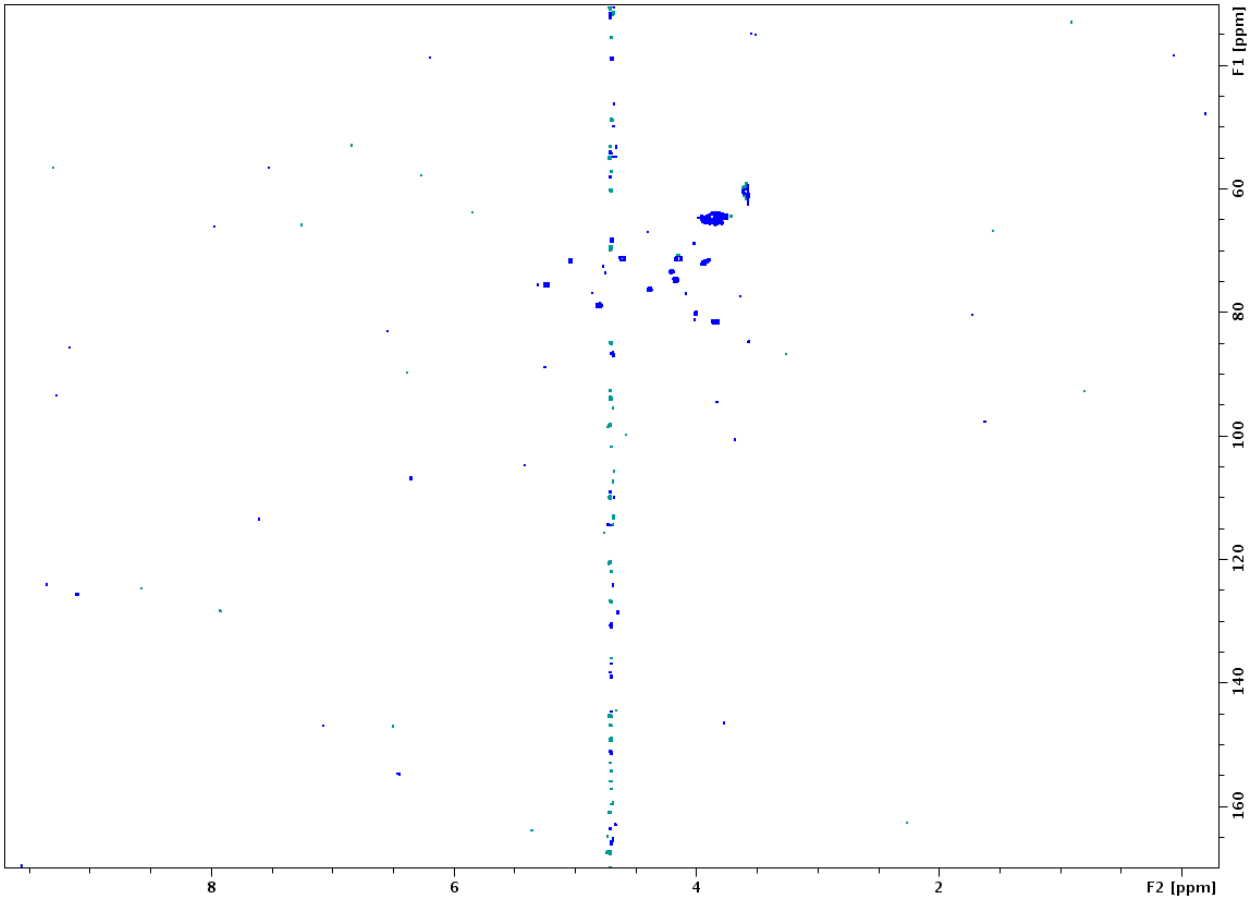
A.13: Full NMR Spectrum of ^{1}H-^{1}H TOCSY for synthesized and purified BA-Nucleotide (Figure 4.5 B)



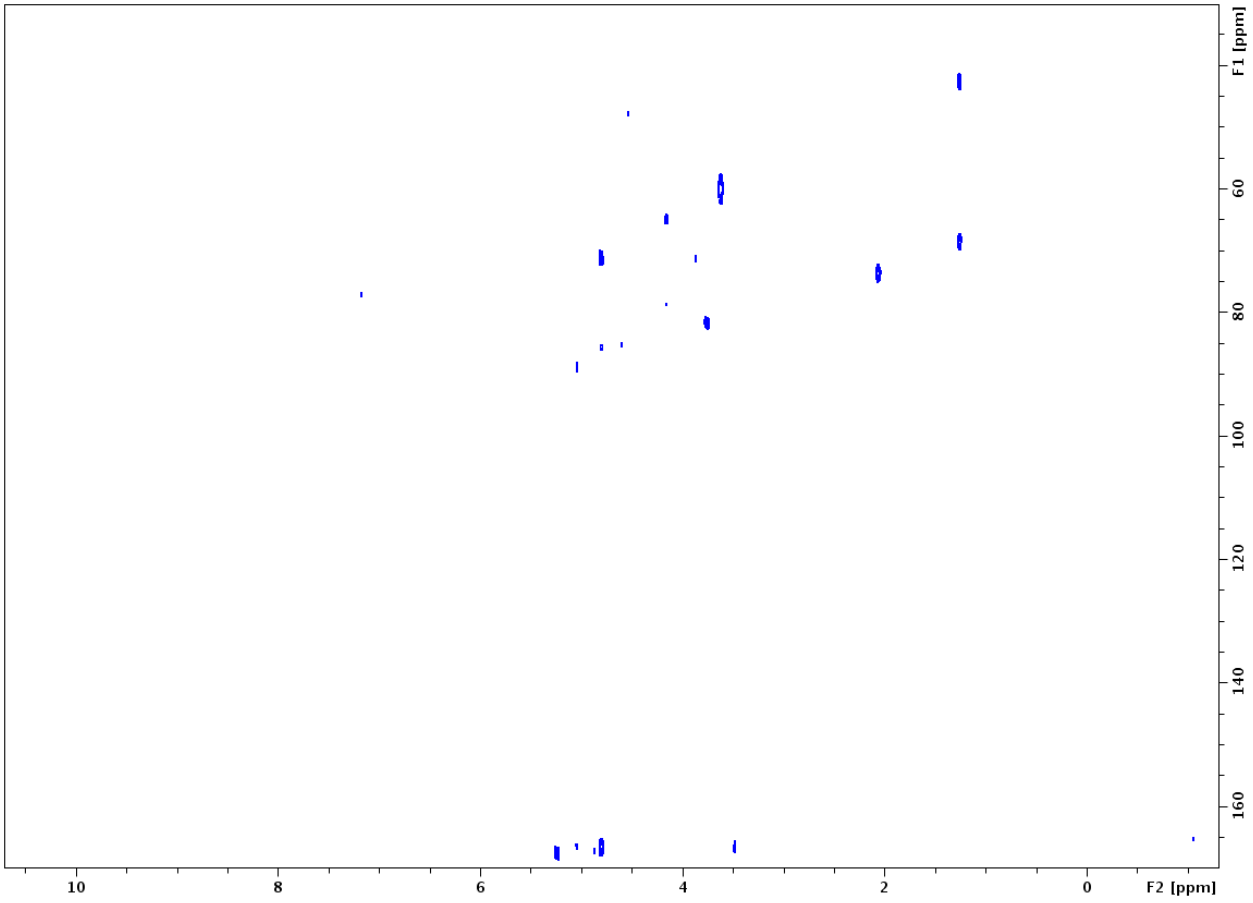
A.14: Zoomed in spectrum of ^{1}H-^{1}H TOCSY for synthesized and purified BA-Nucleotide (Figure 4.5 B)



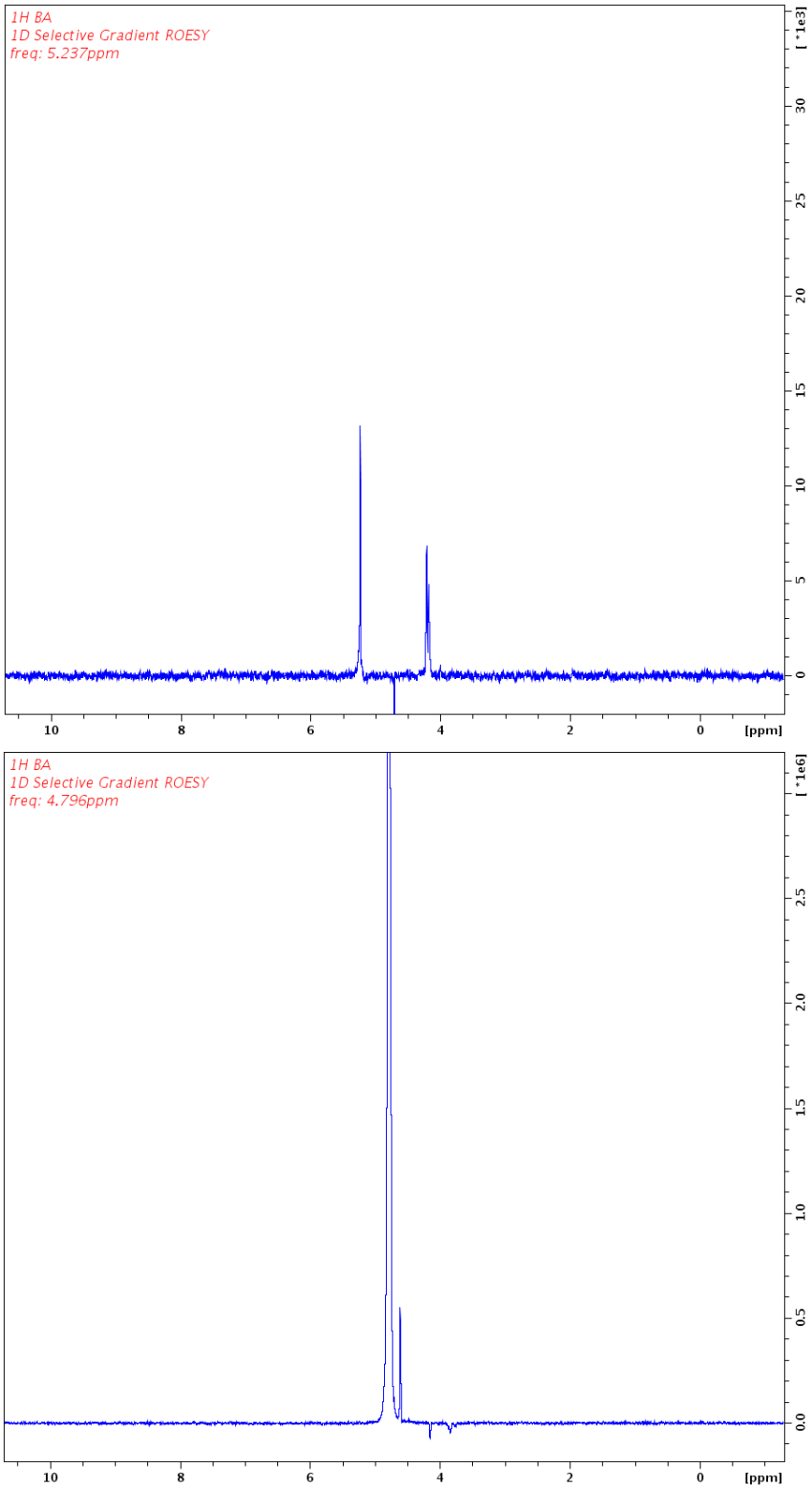
A.15: Full NMR Spectrum of ^{1}H-^{13}C 2D-HSQC for synthesized and purified BA-Nucleotide (Figure 4.5 C).



A.16: Full NMR Spectrum of ^{1}H-^{13}C 2D-HMBC for synthesized and purified BA-Nucleotide (Figure 4.5 D).



A.17: Full NMR Spectrum of ^{1}H selective ROESY for synthesized and purified BA-Nucleotide (Figure 4.5 E & F).



PUBLICATIONS

Mungi, C. V, and Rajamani, S. (2015). Characterization of RNA-Like Oligomers from Lipid-Assisted Nonenzymatic Synthesis: Implications for Origin of Informational Molecules on Early Earth. Life 5, 65–84.
http://www.mdpi.com/2075-1729/5/1/65
DOI: 10.3390/life5010065

Mungi, C. V, Singh, S.K., Chugh, J., and Rajamani, S. (2016). Synthesis of barbituric acid containing nucleotides and their implications for the origin of primitive informational polymers. Phys. Chem. Chem. Phys. 18, 20144–20152.
https://pubs.rsc.org/en/content/articlelanding/2016/cp/c6cp00686h
DOI: 10.1039/C6CP00686H