Role of co-solutes in nonenzymatic RNA replication on prebiotic Earth

A Thesis

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

By

Niraja Vijay Bapat

20133242

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

2018

Indian Institute of Science Education and Research
(IISER – PUNE)
Dr. Homi Bhabha Road, Pashan, Pune 411 021.

Dr. Sudha Rajamani
Associate Professor
srajamani@iiserpune.ac.in

Certificate

Certified that the work incorporated in the thesis entitled "**Role of co-solutes in nonenzymatic RNA replication on prebiotic Earth**", submitted by Ms. Niraja Vijay Bapat, 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.

Date: October 1st, 2018

Dr. Sudha Rajamani

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.

Niraja Vijay Bapat

Registration number: 20133242

Date: October 1st, 2018

Acknowledgments

Accomplishment of a doctoral degree has been a very enriching journey for me. I have received generous help, support and motivation all the way through, from a lot of wonderful people and the acknowledgements are long due. I take this opportunity to express my gratitude towards them.

First and foremost, my sincere thanks to my PhD advisor, Dr. Sudha Rajamani, for letting me work with her on this very fascinating mystery of origin of life. She has been very helpful and supportive throughout. She is an incredible human being and exceptionally good mentor. She has encouraged me to be an independent researcher from the very beginning. The scientific writing skills and problem solving approach that I have learnt from her will definitely help me in the long run. I would also like to acknowledge my research advisory committee (RAC) members, Dr. B.J. Rao, Dr. Supratim Sengupta and Dr. Thomas Pucadyil, for their timely inputs and suggestions. I sincerely thank our collaborators Harshad Paithankar and Dr. Jeetender Chugh for carrying out the NMR analysis. I am also grateful to Dr. Julien Derr for bringing in his expertise and helping us with the simulation studies. I am thankful to Soumya Bhattacharyya and Dr. Thomas Pucadyil for their help with analysis of the RNA–lipid interactions. I am also thankful to Dr. Yayoi Hongo for the LC-MS analysis of some of our samples. I have learnt the invaluable skill of mass analysis from her, for which I will be always grateful. All these collaborations have truly helped me in shaping my thesis project as well as let me learn a lot of new skills and techniques. I am grateful to Dr. Ramana Athreya for his help with statistical analysis. A particular thanks to Dr. DeLuca from Keck laboratory, Yale, USA, for providing us with the custom-synthesized RNA oligomers.

Next up are all the present and past COoL lab members. It was fun to work in the lively environment that was maintained in the lab. I thank Chaitanya, Manesh, Shikha, Susovan and Anupam for being a great group to work with! A special thanks to Chaitanya for being a wonderful colleague and friend. I appreciate Anupam's efforts of teaching me the basics of organic synthesis. I would also like to mention past members, Gaurav, Yovhan and Raji, for their enthusiasm and help in my work.

I am very thankful to the facilities and financial support provided by the Biology and Chemistry departments at IISER, Pune. I would further like to acknowledge the help provided by Bio-admin and Bio-stores, especially by Kalpesh, Piyush, Shabnam, Mrinalini and Mahesh. I am very grateful to Dr. Ram Krishnamurthy, Dr. H. James Cleaves, Dr. David Deamer and Dr. Tony Jia for all the discussions and their scholarly inputs. Their expert suggestions really helped to fill the intellectual void due to immense lack of the Astrobiology research community in India. I take this opportunity to thank Dr. Nishigandha Naik who inspired me and kindled my passion for science. I am thankful to the financial support received in the form of travel grants from ISSOL, ELSI-EON, DBT, Infosys foundation and the organizing committee of AbGradCon 2018; it gave me the opportunities to attend international conferences and interact with the experts and like-minded colleagues from all over the globe. I acknowledge CSIR, Government of India for my fellowship.

The long stay away from home during my PhD tenure wouldn't have been possible if not for some great friends. I am really grateful to Janhavi Kodilkar, who has always been at my side no matter what. A very special thanks to Manasi Kulkarni and Shubhankar Kulkarni for always being there! I am also thankful to Ravi, Jyoti, Rucha, Amruta, Arunabha, Palak and Mahesh for standing with me through thick and thin. I would also like to mention Tanushree, Kunalika, Sayali, Sampada, Aboli, Ketakee, Libi for their contributions in professional as well as personal life. I would like to thank the entire team of 'NASA spaceward bound India–Ladakh' expedition for once-in-a-life-time experience.

These acknowledgments won't ever be complete without thanking my family. None of this would have been possible without the unconditional support and love from my parents, grandparents and my brother, Ninad. My Aai and Baba have always encouraged me to achieve greater goals; words would dry up describing their immense contribution in shaping my life.

Niraja

Synopsis

Role of co-solutes in nonenzymatic RNA replication on prebiotic Earth

---

Name: Niraja Vijay Bapat

Registration number: 20133242

Name of supervisor: Dr. Sudha Rajamani

Department: Biology, IISER Pune, India

Date of joining: January, 2, 2013

Chapter 1: Introduction

The transition from chemistry to biology, resulting in the origin of life, still remains an intriguing mystery. Based on our understanding of extant biochemical processes, the existence of information carrying molecules is thought to have been indispensible for the formation of the living entity. In this regards, 'RNA' is deemed suitable as the first genetic polymer due to its peculiar structural and functional abilities. The widely accepted 'RNA World' hypothesis presumes that the fundamental roles of catalysis and information transfer were performed by RNA molecules, during the origin and early evolution of life on prebiotic Earth (Gilbert, 1986). This hypothesis is supported by the presence of modern molecular relics such as RNA viruses (in which RNA acts as the genetic polymer), rRNA in the core of the ribosome (in which RNA acts as the catalytic molecule) etc. Given this, chemically driven processes, such as oligomerization of monomers, along with accurate nonenzymatic replication of the encoded information, are thought to have been crucial for the formation and propagation of primitive ribozymes in a putative RNA World.

For the past half century, prebiotic chemists have exploited different means, such as use of metal ions (Orgel, 2004), cyclic nucleotides (Costanzo et al., 2009), ice-water eutectic phases (Kanavarioti et al., 2001), clay minerals (Ferris, 2006) and alternate cycles of dehydration-rehydration (Rajamani et al., 2007), for obtaining nucleic acid oligomers without the aid of enzymes. Once the oligomers are formed, the subsequent step would have been faithful

replication, in order to retain the information and thus the function of these oligo-/polymers. Early studies in this area of research, involved understanding the template-directed oligomerization of the activated (Inoue and Orgel, 1981; Tohidi et al., 1987) and non-activated (Olasagasti et al., 2011) monomers, to get the complementary nucleic acid strands. Nonenzymatic template-directed primer extension has also been widely studied using activated monomers (Orgel, 2004; Stütz et al., 2007). Information has been copied from both homopolymeric (Zhang et al., 2013a) and heteropolymeric templates (Zhang et al., 2013b). Further, the rate of template-copying was found to get enhanced in the presence of downstream helper oligomers (Li et al., 2017). While understanding the fidelity associated with such enzyme-free copying reactions, it was found that addition of incorrect nucleotides during nonenzymatic replication stalls the process (Rajamani et al., 2010). The initial mismatch further leads to a cascade of mismatches (Leu et al., 2013), giving advantage to the faithfully replicating sequences.

However, in all aforementioned studies, a crucial aspect, which is of the inherent complexity of the prebiotic environment, has not been factored. Prebiotic soup would have been replete with different kinds of molecules as a result of a plethora of products and by-products obtained from the relevant syntheses, processes and the exogenous delivery of mixtures of different chemical entities. Hence, the studies carried out in the absence of any background molecules might not be necessarily prebiotically realistic. Therefore, in the present investigation, we set out to understand the role of prebiotically relevant co-solutes on nonenzymatic processes, using template-directed RNA primer extension reactions as our study system.

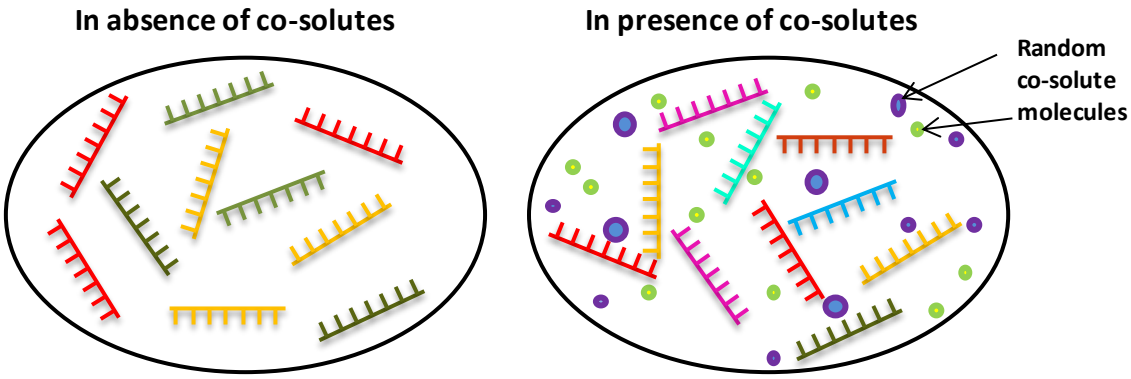
Chapter 2: Effect of co-solutes on template-directed nonenzymatic copying of RNA

In order to increase the complexity of our reaction milieu in a step-wise manner, lipid vesicles and polyethylene glycol (PEG) were used as co-solutes. The presence of co-solutes or background molecules in the heterogeneous prebiotic soup would have resulted in effects similar to that of molecular crowding. Presence of co-solutes and molecular crowding agents is very relevant in extant biology and is also known to affect the kinetics of many protein- and nucleic acid-based biochemical reactions (Ellis, 2001; Nakano et al., 2014). In our study, the lipid

vesicles were chosen as a co-solute due to their relevance in the formation of boundary conditions of the protocell (Szostak et al., 2001). On the other hand, PEG is routinely used as a molecular crowding agent in many biochemical analyses. Its use also served as a proxy for the presence of nonspecific polymers that might have resulted from the polymerization of monomers other than the RNA monomers (Mamajanov et al., 2014; Mungi and Rajamani, 2015). Using imidazole-activated nucleotides, the rate and accuracy of the nonenzymatic template-directed RNA primer extension reactions was systematically studied in the presence of individual and admixture of these co-solutes.

Upon analysis of the rate of monomer addition for all possible 16 combinations (viz. 4 matched additions and 12 mismatched additions), it was observed that the rate of primer extension decreased in presence of individual co-solutes in those reactions, which involved a 'matched' addition of a purine across its cognate template base. This effect was further enhanced in the presence of the admixture of both the co-solutes. The rate of addition of 'G across C' and 'A across U' was found to be lowered by almost 55% in the presence of both the co-solutes as compared to the control conditions. The rates of all the other addition reactions were not notably affected in the presence of either the individual or admixture of co-solutes. The differential effect of the presence of the co-solutes on the rate of purine based matched addition reactions, was further confirmed using statistical analysis.

As a result of the decrease in the rate of two of the four matched addition reactions, the fidelity of our RNA based nonenzymatic copying system decreased in the presence of co-solutes. An overall increase in the misincorporation frequencies against 'C' and 'U' template bases was observed. Specifically, the misincorporation frequency of the addition of 'G across U' was found to have particularly increased in the presence of co-solutes. Furthermore, the overall mutation rate was found to have increased to 26.6 ± 4.4% in the presence of both the co-solutes as opposed to that of 19.8 ± 0.5% in control conditions. When the individual template bases were considered, a prominent increase in mutation rate was observed for the 'C' (3.6 ± 0.3 % in the control condition vs. 9.6 ± 1.6 % in the presence of both the co-solutes) and 'U' (31 ± 4.6 % in the control condition vs. 51.7 ± 3.5 % in the presence of both the co-solutes) template bases.



A schematic depicting the possibility for greater sequence space exploration in the presence of co-solutes. Different colors of strands indicate different sequences.

To analyze the effect of presence of background molecules on the nonenzymatic RNA replication over several cycles, simulation studies were also undertaken in collaboration with Dr. Julien Derr from Paris Diderot University, France. Experimentally generated reaction rates were used to understand *in silico* RNA template replication over multiple cycles. A strong 'GC' bias was observed in the replicated sequences, both in the absence and the presence of the co-solutes. Upon further analysis, higher numbers of new sequences were found to have been generated in the presence of co-solutes. Such effect would have led to a faster exploration of the sequence space in a heterogeneous prebiotic soup. In general, this study highlighted the importance of accounting for prebiotic heterogeneity; it might have directly affected the origin and evolution of functional nucleic acid polymers on prebiotic Earth. A part of this study has been published. Ref: "Bapat, N. V, and Rajamani, S. (2015). Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids. Journal of Molecular Evolution 81, 72–80."

Chapter 3: Characterization of mechanism underlying co-solute based reactions

The presence of lipid vesicles and PEG was found to prominently decrease the rate of addition of 'G across C' and 'A across U' reactions, during nonenzymatic RNA copying (Bapat and Rajamani, 2015). Hence, we wished to further analyze the underlying mechanism for this differential effect observed in these reactions. Presence of co-solutes could, in principle, affect the behavior of monomers or of the RNA primer-template complexes itself. PEG, which was used as one of the co-solutes, is known to affect the diffusion of the reactant molecules by causing crowding effect (Ellis, 2001). Additionally, nucleic acid-lipid interactions are also known

to occur (Kuvichkin, 2002). Therefore, we analyzed the effect of presence of co-solutes on two phenomenon, that of monomer stacking as well as potential RNA–lipid interaction.

Nuclear Magnetic Resonance (NMR) spectroscopy was chosen as the method of analysis for estimating the diffusion and the size of the monomer/monomer clusters, in the absence and presence of PEG. This is mainly because no molecular tagging is required for this method, an advantageous situation for us as tagging could potentially affect the stacking properties of the monomers. The NMR analysis was carried out in collaboration with Harshad Paithankar and Dr. Jeetender Chugh, from the Chemistry department at IISER, Pune, India. The diffusion constants were estimated using DOSY NMR. In the absence of PEG, an increase in the diffusion time was observed for all the nucleotides; the extent of this increase was greater for the purine monomers as opposed to the pyrimidine nucleotides. However, in the presence of PEG, no clear trend could be established, possibly due to the substantial increase in the viscosity of the solution. Therefore, to further estimate the relative size of the molecular clusters, ^{13}C-T_1 relaxation NMR was carried out. These set of experiments reveled comparatively greater stacking tendencies for purine monomers in the presence of PEG. Highly stacked purine nucleotides in PEG would result in lesser availability of these monomers in the reaction milieu to facilitate primer extension. Thus, observations from these NMR experiments explain, at least in part, the reason behind differential decrease in the rate of purine based matched addition reactions in the presence of co-solutes.

We further analyzed RNA–lipid interactions using fluorescence imaging, as the primer used in our studies is Cy3-tagged at its 5'-end. DLPC lipid was used for the analysis, as role of this particular phospholipid as a co-solute had been analyzed in previous experiments (Bapat and Rajamani, 2015). We used silica beads coated with excess lipid reservoir (SUPER templates), and supported lipid bilayer (SLB), for our assays. These were carried out in collaboration with Soumya Bhattacharyya and Dr. Thomas Pucadyil from the Biology department at IISER, Pune, India. However, using these assay systems, no significant interactions were observed between the 20-mer RNA primer and DLPC lipid. This might be due to the very small size of primer RNA, or the absence of effective net charge on DLPC at neutral pH. We envisage that systematic

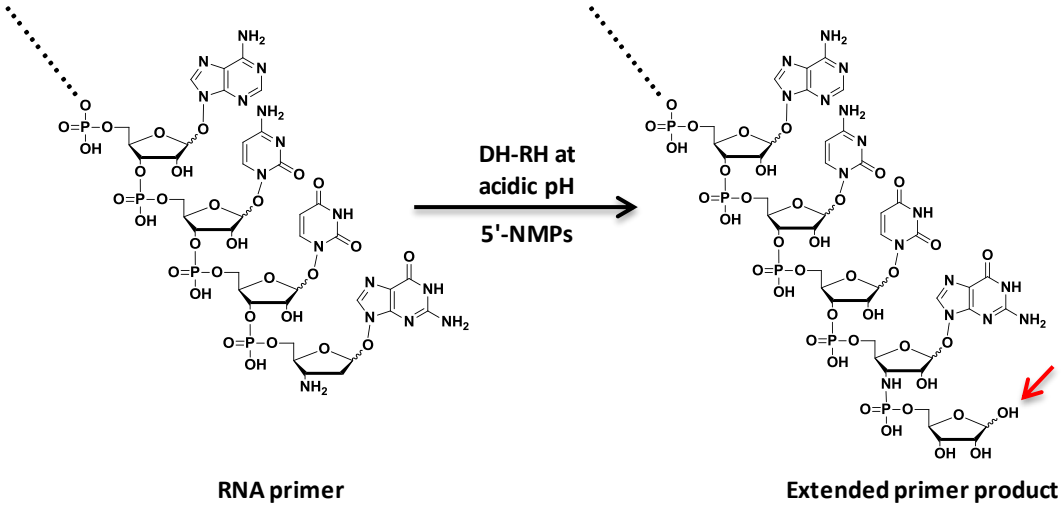
studies involving variation of both, the length of the RNA, and the composition of the lipid bilayer, would shed more light on the role of these factors in RNA-lipid interactions, under our reaction conditions.

Chapter 4: Template-directed replication using nucleoside- 5'-monophosphate monomers under DH-RH conditions

Next, we wanted to analyze the plausibility of template-directed RNA primer extension, using prebiotically realistic 'non-activated' monomers. Studies carried out using activated monomers have substantially advanced our knowledge of nonenzymatic oligomerization and replication processes. Nonetheless, the presence of highly concentrated and stable pools of such monomers on prebiotic Earth remains arguable. Even though the oligomerization of non-activated nucleoside-5'-monophosphate monomers has been demonstrated (Rajamani et al., 2007; Mungi and Rajamani, 2015), enzyme-free information copying using these non-activated monomers remains largely unexplored. Only a single study to date has attempted information copying from ssDNA strand, using 5'-dNMPs under alternate dehydration-rehydration (DH-RH) reaction regimen (Olasagasti et al., 2011). To the best of our knowledge, the current study is the first to report information copying in RNA based system, using non-activated 5'-NMPs.

We carried out template-directed RNA primer extensions using 5'-NMPs in the presence of lipids as co-solutes, under repeated cycles of DH-RH. Alternate cycles of DH-RH, which are also prebiotically relevant, facilitate bond formation between two monomers by providing the requisite energy and by reducing water activity in the reaction environment. The lipids form multilamellar liquid crystalline matrices under the dehydrated conditions, further concentrating the reactants and facilitating the phosphodiester bond formation (Deamer, 2012; Toppozini et al., 2013). Reduced degradation of RNA primer over multiple DH-RH cycles was indeed observed when lipid was included in the reaction mixture. However, upon gel analysis, the extended RNA primer product was found to run in between the level of the staring 20-mer primer and that of an intact 21-mer extended RNA product. Generation of an abasic site on this extended primer product is suspected as the main reason for the aforesaid observation. The loss of base was persistent despite lowering the temperature at which the DH-RH cycles were

carried out, and even after changing the nature of the monovalent ion used in the starting reaction mixture. Furthermore, lowering the amount of the acid in the rehydration solution or variation in rehydrating agent also did not circumvent the loss of base during the extension of the RNA primer using 5'-NMPs. Deglycosylation has been previously reported for oligomerization reaction using 5'-AMPs that was carried out under similar DH-RH conditions (Mungi and Rajamani, 2015).From our observations, we envisage the loss of base being potentially coupled to the chemical addition of the incoming monomer on the pre-existing RNA monomer/oligomer.



Schematic representation of generation of an abasic site (indicated by red arrow) during RNA primer extension under DH-RH, carried out at low pH and high temperature, using 5'-NMPs as monomers

Since the extended primer product contained a plausible abasic site while using 5'-NMPs as monomers; we next investigated the extension of the RNA primer using ribose-5'-monophosphate (5'-rMP), which lacks the nitrogenous base. Long streaks resulting possibly from multiple additions of 5'-rMP to the RNA primer were observed upon gel analysis. Such hybrid oligomers might have played a crucial role of sampling different information carrying moieties during the early stages of emergence of primitive informational polymers (Hud et al., 2013). Significantly, the extension product obtained from the addition of a single 5'-rMP monomer was found to be running at a similar level on a denaturing gel, to the band that was obtained when 5'-AMP was used as monomer. This indirectly confirmed the loss of base during

the addition of 5'-NMP to a pre-existing RNA primer, in our reactions. Furthermore, in preliminary mass analysis, the formation of a covalent bond, between the 5'-rMP monomer and a nucleoside resembling the 3'-terminal monomer of the RNA primer, was confirmed. This was carried out in collaboration with Dr. Yayoi Hongo at the Earth Life Science Institute (ELSI) at Tokyo, Japan. In general this study highlighted the importance of analyzing pertinent nonenzymatic reactions in prebiotically realistic scenarios. Part of this chapter has recently been accepted for publication. Ref: "Bapat, N. V, and Rajamani, S. (2018). Templated Replication (or Lack Thereof) under Prebiotically Pertinent Conditions. Scientific Reports, accepted."

Chapter 5: Summary

The work carried out as part of this thesis emphasizes the importance of the following considerations while discerning prebiotic processes pertinent to our understanding of the origin of life on early Earth: 1) the prebiotic heterogeneity, 2) availability of the reactant moieties and, 3) prebiotically pertinent reaction conditions that might have prevailed. We found that the rate and accuracy of enzyme-free RNA copying decreases in the presence of background molecules. Furthermore, the presence of crowding agent was found to selectively enhance the stacking tendencies of the purine monomers. Additionally, when information copying from ssRNA template was attempted using non-activated 5'-NMPs, a persistent loss of the informational moiety on the extended primer product was seen. These studies thereby importantly demonstrate that the nonenzymatic reactions carried out in buffered conditions, using high concentrations of activated reactants, might not result in a similar set of product/s when studied under prebiotically realistic conditions.

List of publications

Bapat N.V, Outlining 'Abiosignatures': Clues from prebiotic chemistry. *In review*.

Bapat, N. V, and Rajamani, S. (2018). Templated Replication (or Lack Thereof) under Prebiotically Pertinent Conditions. Scientific Reports 8(1), 15032.

Bapat, N. V, and Rajamani, S. (2015). Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids. Journal of Molecular Evolution 81, 72–80.

References

Costanzo, G., Pino, S., Ciciriello, F., and Di Mauro, E. (2009). Generation of Long RNA Chains in Water. The Journal of Biological Chemistry 284, 33206–33216.

Deamer, D. (2012). Liquid Crystalline Nanostructures : Organizing Matrices for Non-enzymatic Nucleic Acid Polymerization. Chemical Society Reviews 41, 5375–5379.

Ellis, R.J. (2001). Macromolecular Crowding : Obvious but Underappreciated. Trends in Biochemical Sciences 26, 597–604.

Ferris, J.P. (2006). Montmorillonite-catalysed Formation of RNA Oligomers: The Possible Role of Catalysis in the Origins of Life. Philosophical Transactions of the Royal Society B: Biological Sciences 361, 1777–1786.

Gilbert, W. (1986). The RNA world. Nature 319, 618.

Hud, N. V, Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The Origin of RNA and "My Grandfather's Axe". Chemistry & Biology 20, 466–474.

Inoue, T., and Orgel, L.E. (1981). Substituent Control of the Poly(C)-Directed Oligomerization of Guanosine 5'-Phosphoroimidazolide. Journal of the American Chemical Society 103, 7666–7667.

Kanavarioti, A., Monnard, P.-A., and Deamer, D. (2001). Eutectic Phase in Ice Facilitates Nonenzymatic Nucleic Acid Synthesis. Astrobiology 1, 271–281.

Kuvichkin, V. V (2002). DNA–Lipid Interactions in vitro and in vivo. Bioelectrochemistry 58, 3–12.

Leu, K., Kervio, E., Obermayer, B., Turk-macleod, R.M., Yuan, C., Luevano, J., Chen, E., Gerland, U., Richert, C., and Chen, I.A. (2013). Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. Journal of the American Chemical Society 135, 354–366.

Li, L., Prywes, N., Tam, C.P., Flaherty, D.K.O., Lelyveld, V.S., Izgu, E.C., Pal, A., and Szostak, J.W. (2017). Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. Journal of the American Chemical Society 139, 1810–1813.

Mamajanov, I., Macdonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Fernandez, 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.

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.

Nakano, S., Miyoshi, D., and Sugimoto, N. (2014). Effects of Molecular Crowding on the Structures , Interactions , and Functions of Nucleic Acids. Chemical Reviews 114, 2733–2758.

Olasagasti, F., Kim, H.J., Pourmand, N., and Deamer, D. (2011). Non-enzymatic Transfer of Sequence Information under Plausible Prebiotic Conditions. Biochimie 93, 556–561.

Orgel, L.E. (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 39, 99–123.

Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2007). Lipid-Assisted Synthesis of RNA-Like Polymers from Mononucleotides. Origins of Life and Evol. of Biospheres 38, 57–34.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., and Chen, I.A. (2010). Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. Journal of the American Chemical Society 132, 1008–1011.

Stütz, A.R., Kervio, E., Deck, C., and Richert, C. (2007). Chemical Primer Extension : Individual Steps of Spontaneous Replication. Chemistry & Biodiversity 4, 784–802.

Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing Life. Nature 409, 387–390.

Tohidi, M., Zielinski, W., Chen, C.B., and Orgel, L.E. (1987). Oligomerization of 3’-Amino-3-Deoxygaunosine-5'-Phosphorimidazolides on a d(CpCpCpCpCp) template. Journal of Molecular Evolution 25, 97–99.

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.

Zhang, S., Zhang, N., Blain, J.C., and Szostak, J. (2013a). Synthesis of N3'-P5'-Linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension. Journal of the American Chemical Society 135, 924–932.

Zhang, S., Blain, J.C., Zielinska, D., Gryaznov, S., and Szostak, J. (2013b). Fast and Accurate Nonenzymatic Copying of an RNA-Like Synthetic Genetic Polymer. Proceedings of the National Academy of Sciences 110, 17732–17737.

Table of Contents

(Hyperlinked)

List of Symbols and Abbreviations

˚C	degree Celsius
^{1}H	Hydrogen-1 (NMR)
2-MeImpC	Guanosine- 5'-phosphoryl-2-methylimidazolide
2-MeImpG	Guanosine- 5'-phosphoryl-2-methylimidazolide
2-MeImpN	5'-phosphoryl-2-methylimidazole-activated RNA monomers
2'-NH_2-ImpddN	2'-amino-2', 3'-dideoxy-5'-phosphorimidazolide monomers
3'-NH_2-ddA	3'-amino 2', 3'-dideoxy adenosine
3'-NP-DNA	3'-5' phosphoramidite linked DNA
5'- AMP	Adenosine-5'-monophosphate
5'-CMP	Cytidine-5'-monophosphate
5'-GMP	Guanosine-5'-monophosphate
5'-NMP	Nucleoside-5'-monophosphate
5'-rMP	Ribose-5'-monophosphate
5'-UMP	Uridine-5'-monophosphate
^{13}C	Carbon-13 (NMR)
α	Level of significance
A	Adenine
A˚	Ångström
ACE	2'-O-[bis[2-(acetyloxy)ethoxy]methyl] (protection group)
AFM	Atomic force microscopy
BSA	Bovine serum albumin
C	Cytidine
cAMP	cyclic Adenosine-5'-monophosphate
cGMP	cyclic Guanosine-5'-monophosphate
CH_3COOH	Acetic acid
CO_2	Carbon dioxide
Cy3	Cyanine 3 (dye)
D	Diffusion coefficient

DH-RH	Dehydration-rehydration
DLPC	1, 2-dilauroyl-sn-glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
D_2O	Deuterated water
DOSY	Diffusion ordered spectroscopy (NMR)
DY547	DyLight^{TM} 547 (dye)
EDTA	Ethylenediaminetetraacetic acid
FRET	Fluorescence resonance energy transfer
FTT synthesis	Fischer–Tropsch type synthesis
G	Guanine
GUVs	Giant unilamellar vesicles
H_2	Hydrogen
H_2SO_4	Sulphuric acid
H_3PO_4	Orthophosphoric acid
HCl	Hydrochloric acid
HCN	Hydrogen cyanide
HCOOH	Formic acid
HNO_3	Nitric acid
HOAt	1-hydroxy-7-azabenzotriazole
HPLC	High performance liquid chromatography
hr	Hour
ImpA	Adenosine-5'-phosphoimidazolide
ImpC	Cytidine- 5'-phosphoimidazolide
ImpG	Guanosine- 5'-phosphoimidazolide
ImpN	Nucleoside- 5'-phosphoimidazolide
ImpU	Uridine- 5'-phosphoimidazolide
ITC	Isothermal titration calorimetry
LC–MS	Liquid chromatography–Mass spectrometry
μ	Mutation rate
μl	Microliter

µM	Micromolar
M	Molar
MALDI–TOF	Matrix-assisted laser desorption ionization–Time-of-flight
min	Minute
mM	Millimolar
ms	Milliseconds
NaCl	Sodium chloride
NaOH	Sodium hydroxide
nm	Nanometer
NMR	Nuclear magnetic resonance spectroscopy
p	Probability value
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PMT	Photomultiplier tubes
POPC	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
P-T complex	Primer-template complex
RNA	Ribonucleic acid
RT	Room temperature
rTCA	Reductive tricarboxylic acid cycle
s^2U	2-thio-U
SELEX	Systematic evolution of ligands by exponential enrichment
SLB	Supported lipid bilayer
SMrT	Supported membrane tubes
ssDNA	Single-stranded deoxyribonucleic acid
ssRNA	Single-stranded ribonucleic acid
SUPER	Supported bilayers with excess membrane reservoir
SUVs	Small unilamellar vesicles
T_1	Spin-lattice relaxation time constant (NMR)
TBDMS	tert-Butyldimethylsilyl (protection group)
TBE buffer	Tris-Borate-EDTA buffer

TEMED	Tetramethylethylenediamine
Tris	Trizma base
U	Uracil
UV	Ultraviolet

List of Figures

(Hyperlinked)

List of Tables

(Hyperlinked)

Chapter 1

Introduction

1.1 The 'Origin of life' conundrum

A very fundamental question that has intrigued scientists and philosophers alike for centuries pertains to deciphering how life might have originated on our planet. Decades of research has gone into hypothesizing and predicting the steps that might have led to the formation of the first living cells. Based on recent discoveries, the tentative time-line indicates that during the 4.5 billion years old history of Earth, the origin of life might have taken place around 3.6 to 4 billion years ago. The earliest available record of life, found in the form of biogenic graphite in Labrador, Canada, dates back to 3.95 billion years ago (Tashiro et al., 2017). Another early evidence for life comes from Nuvvuagittuq belt in northeastern Canada in the form of fossilized microorganisms, which date back to a minimum of 3.77 billion years (Dodd et al., 2017). Biosignatures in the form of stromatolites, preserved gas bubbles etc., have also been uncovered in the ca 3.5 billion year old hot spring deposits in Pilbara craton, Western Australia (Djokic et al., 2017). Stromatolites are the layered deposits caused by the growth of Cyanobacteria, an early unicellular microbial life form. Microorganisms are the simplest unicellular life known to us. However, this microbial life is complicated enough to have arisen on its own from an admixture of chemicals. The real challenge, therefore, lies in understanding the process of transition from chemistry to biology, for which no fossil records might be available.

The first ideas about the abiotic origin of life on Earth came from J.B.S. Haldane and Alexander Oparin. In the 1920s, these two scientists independently proposed that the organic molecules necessary for life could be formed from inorganic materials, in the presence of an external energy source, and in an atmosphere that was reducing. This 'primordial soup' theory (a term coined by Haldane) gained popularity amongst the scientific community and was significantly supported by the famous Urey-Miller experiment, which set the stage for the emergence of the field of prebiotic chemistry. In a Science article from 1953, Stanley Miller and Harold Urey successfully demonstrated the formation of organic molecules from a mixture of water, methane, ammonia and hydrogen, in the presence of electric discharge as the source of external energy (Bada and Lazcano, 2003). They identified five amino acids including a large quantity of Glycine. This piece of work came shortly after another seminal discovery of double

helical nature of DNA by Watson and Crick. Recent re-analysis of the original samples from the Urey-Miller experiment has shown that many more significant organic molecules were formed in the original spark discharge experiment, albeit in smaller quantities. This study, which was carried out using modern day analytical techniques like HPLC and mass analysis, has identified 22 different amino acids and five amines (Johnson et al., 2008). Although the Urey-Miller experiment successfully demonstrated the synthesis of complex organic molecules from simple inorganic components, the correct assembly of these molecules to give rise to living cells remains largely elusive. Almost 65 years later, opinions still remain divided and there are a lot of theories which have been proposed to explain the processes that might have occurred during the origin and early evolution of life on Earth.

This chapter describes a few important theories put forth (the 'Many World' hypotheses), for explaining the origin of life and the reasons behind the widespread acknowledgment of the 'RNA World' hypothesis in the Astrobiology community. In the later sections, the prebiotically pertinent synthesis of nucleotide monomers has been discussed. This is followed by a discussion on their nonenzymatic oligomerization and related replication reactions, which are required for the synthesis of functional RNA molecules, deemed essential for a putative RNA World. In the last segment, the fidelity associated with nonenzymatic replication processes has been elaborated on; this would have been essential to preserve the sequence and thus the function of the RNA molecules.

1.2 Plausible routes to 'Origins of life'

It might not be possible to discern the precise order of events that would have led to the origin of the first cells. However, depending on our knowledge of extant chemical and biological systems, and some understanding about the atmospheric and geochemical conditions that might have been prevalent on the early Earth, there have been multiple hypotheses proposed to explain the probable route for the transition from chemistry to biology. A few of the important and relevant hypotheses are discussed below.

1.2.1 RNA World hypothesis

This is currently the most widely accepted hypothesis in the field. It revolves around ribonucleic acid (RNA) being a crucial molecule and its central role in the origin and early evolution of life on Earth. It was originally proposed almost 50 years ago by Francis Crick, Carl Woose and Leslie Orgel (Robertson and Joyce, 2010). However, it wasn't until the unexpected discovery of ribozymes in the early 1980s by Cech's group that the widespread discussions about the crucial role of RNA on prebiotic Earth were taken seriously. This ultimately gave rise to the term "RNA World", which was coined by Gilbert in his famous Nature article (Gilbert, 1986). This concept mainly states that RNA molecules would have acted, both, as a genetic material (a function performed mostly by DNA in extant biology), and also as a catalyst (a function performed mostly by proteins in extant biology), during the origin and early evolution of life on Earth (Fig 1.1). In other terms, it presumes the existence of catalytic RNAs which were capable of both, storing and propagating genetic information. This idea falls under the 'replication-first' concept of the origin of life, which suggests that the self-replicating genetic material might have preceded the evolution of complex metabolic networks.

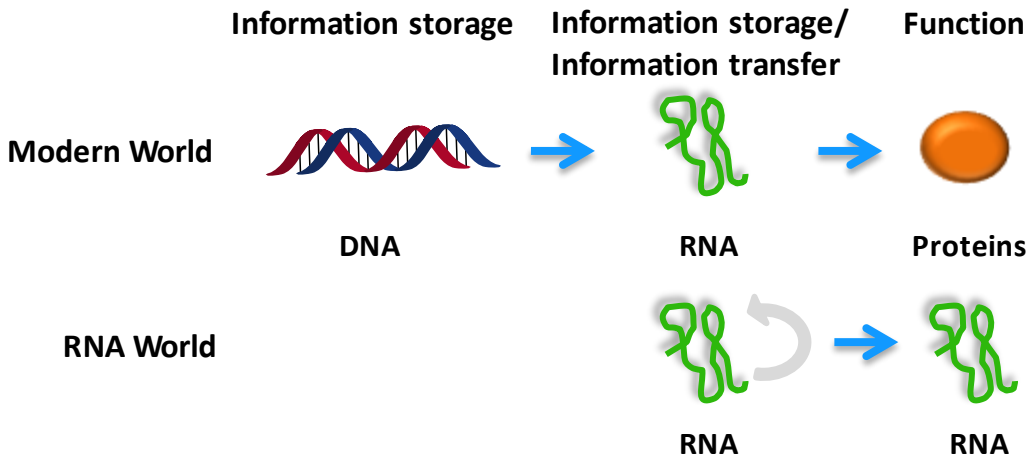


Figure 1.1: Conceptual representation of the 'RNA World' hypothesis. This hypothesis states that during origin and evolution of early life, RNA played an active role in both genetic information transfer and catalysis.

The ability of RNA to base pair with another RNA strand makes it a good candidate for the transfer of genetic material. The peculiar secondary structure forming capability of this single stranded molecule makes it a good candidate for facilitating catalysis as well. Thus, instead of requiring two different kinds of molecules, namely protein and DNA, to carry out two core functions, RNA could have facilitated both, simultaneously. The fact that we have evidences of RNA acting as a genetic material (in RNA viruses), and as a catalyst (core of the ribosome, self-splicing RNA introns etc.), even in modern biology, largely supports the 'RNA World' hypothesis. The use of RNA as the genetic material by certain viruses, like polio virus, influenza virus etc., is well known. Such viruses are also thought to be a vestige of the prebiotic RNA World (Forterre, 2005). On the other hand, the first discovery of catalytic RNAs came independently from the labs of Sidney Altman and Thomas Cech (Kruger et al., 1982; Guerrier-Takada et al., 1983), for which they were awarded the Nobel Prize in Chemistry in 1989. The 'smoking gun evidence', however, is the presence of RNA in the core of the contemporary ribosome's structure. No amino acid side chain is observed within the 18 A˚ active site of the ribosome core, indicating that the peptide-bond formation is catalyzed by RNA (Nissen et al., 2000). Furthermore, functional RNA molecules, like different aptamers and ribozymes, can also be selected for in the laboratory using a methodology called SELEX (Systematic Evolution of Ligands by Exponential Enrichment). Such experimental evidence makes RNA a favorable contender for explaining the origin of life, as it is argued that molecular evolution might have preferred a minimalistic approach in terms of the number of 'functional innovations' required for the formation of first living entities.

1.2.1.1 Pre-RNA world hypothesis

Although RNA is a favorable candidate as the foremost molecule important for life's origin, the structure of RNA molecule itself is very complicated and seems very difficult to have gotten assembled on its own from a mixture of prebiotic entities. *De novo* synthesis of long RNA strands capable of catalysis has proved to be very difficult, and has been termed as the 'Molecular biologists' dream' (Orgel, 2004). Hence, there has been an opinion in the field that considers RNA itself to have resulted from a process of molecular evolution that was predated

by other, possibly related, simpler informational polymers. It has been suggested that other types of nucleic acids, such as glycol nucleic acid (GNA), threose nucleic acid (TNA), peptide nucleic acid (PNA) or alanyl nucleic acid (ANA) (Fig 1.2A) (Robertson and Joyce, 2010), probably formed first, and later got replaced by the more efficient RNA molecules (Hud et al., 2013). The structure of RNA itself might have also evolved from pre-RNA polymers that used other simpler alternatives for sugar, phosphate backbone and the nitrogenous bases (Fig 1.2B, Hud 2013). Recent findings related to nucleoside and nucleotide formation, using non-canonical bases, further strengthens this hypothesis(Chen et al., 2014; Kim and Benner, 2015; Mungi et al., 2016).

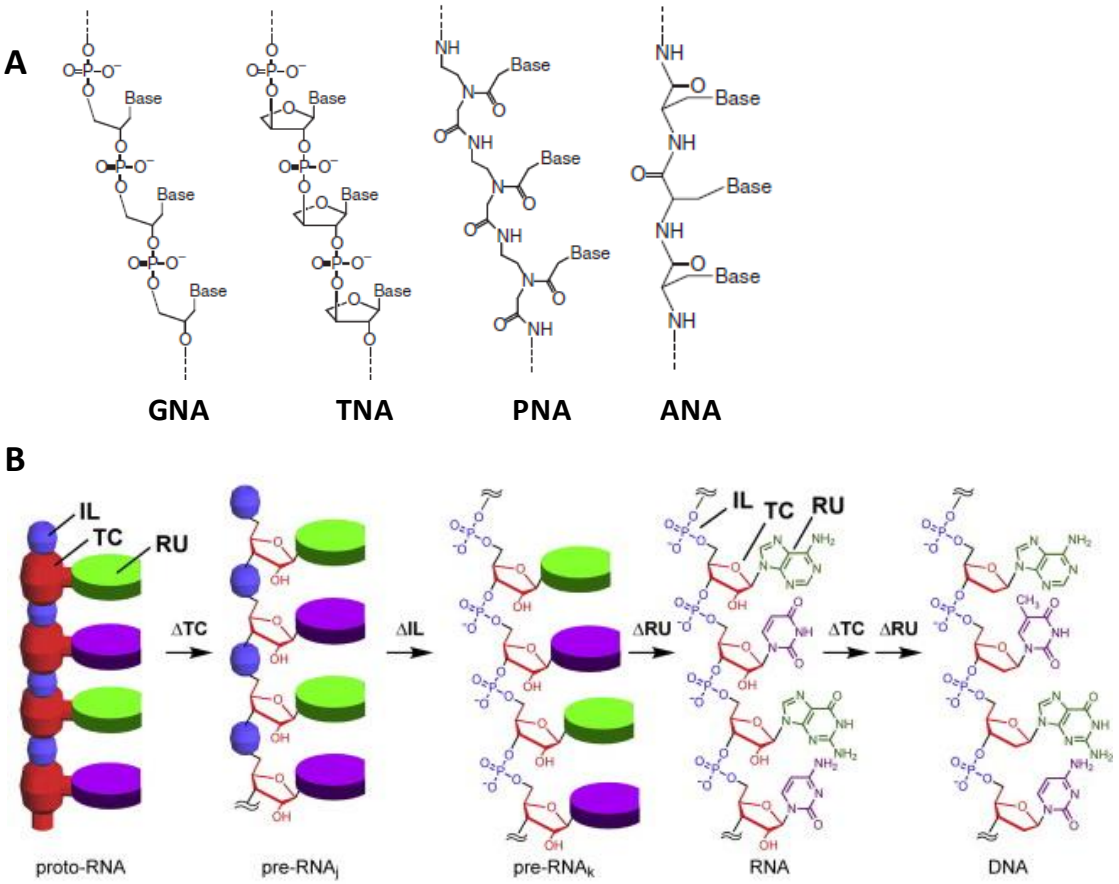


Figure 1.2: RNA as a result of molecular evolution (A) Chemical structures of alternate nucleic acids (adapted from Robertson & Joyce, 2010) (B) Plausible prebiotic evolutionary pathway from proto-RNA to RNA to DNA. IL (ionized linker), TC (trifunctional connecter) and RUs (recognition units) denote the three main parts of the RNA molecule (reproduced form Hud et. al., 2013).

1.2.2 Metabolism-first hypothesis

An alternate concept suggests that self-reproducing and evolving proto-metabolic networks, consisting of simple organic molecules, might have preceded the self-replicating genetic material(s) during the origin of life (Fig 1.3) (Trefil et al., 2009). In other words, this hypothesis states that the earliest stage(s) of life comprised of a self-sustaining chain of chemical reactions potentially associated with mineral surfaces, wherein the genetic information was not an essential part of life (Bada and Lazcano, 2002). The primitive metabolic life might have comprised of autocatalytic reactions based on organic compounds made from simpler constituents like CO_2, CO etc. 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 concept (Trefil et al., 2009; Kitadai et al., 2017). The advantage of this concept is that it does not invoke the need to explain the very complicated synthesis of, both, monomeric and polymeric RNA, which is required for the putative RNA World hypothesis or the replication-first scenario. However, this small molecule approach would also require the presence of relevant niche parameters, like an external energy source, a driver reaction coupled to this energy source, compartments etc., in order to thrive and evolve (Shapiro, 2007).

Different metabolically pertinent reactions, like the formose reaction, reverse TCA cycle, autocatalytic peptide assemblies etc., have been tried and tested to support the notion of self-sustaining reaction networks that might have kick-started the origin of life. However, none of these reactions thus far have been proven to be completely autocatalytic (Orgel, 2008). This remains one of the major critiques of this concept. Therefore, it is generally considered that the autocatalytic-reaction based life could not have evolved in the absence of genetic replication system/s as it would have been difficult for them to maintain heredity and evolution otherwise (Bada and Lazcano, 2002).

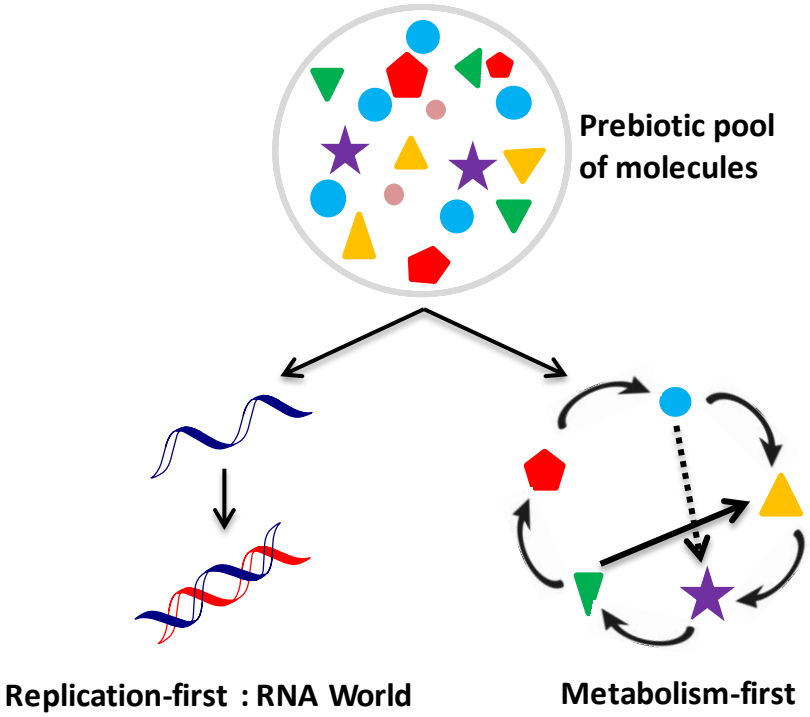


Figure 1.3: Illustration depicting two different views of origins of life that exists in the field. The replication-first scenario presumes formation of genetic material from the prebiotic pool of chemicals for origin and sustenance of life. On the other hand, the metabolism-first scenario conceptualizes formation of self-maintaining proto-metabolic networks for the origin and sustenance of life.

1.2.2.1 Iron-sulfur World

This idea was proposed by Gunter Wächtershäuser in the late 1980s, and is aimed at explaining the metabolism-first hypothesis. The key point to consider in Wächtershäuser's idea is the site of occurrence of these primordial metabolic cycles. He suggested that the proto-metabolic networks occurred on iron sulfide (mainly pyrite) mineral surfaces. Such interaction would be dependent on the anionic binding of the ligands on the positively charged mineral surfaces and thus would eventually select for polyanionic molecules like polyanionic coenzymes, peptides and nucleic acids (Wächtershäuser, 1988). His view of the first life was that of two-dimensional surface chemistry (acellular life) rather than the assembly of organic molecules into a primitive cell. A few of the early experimental support for this idea include formation of acetic acid using (Ni, Fe)S under hydrothermal vent like conditions (Huber and Wächtershäuser, 1997), and the

synthesis of pyruvate using catalytic iron sulfide under high pressure and temperature conditions (Cody et al., 2000). More recent research in this direction involves the synthesis of iron-sulfur clusters using ultraviolet light (Bonfio et al., 2017), and the demonstration of nonenzymatic, metal-catalyzed steps involved in the rTCA cycle (Muchowska et al., 2017).

This idea of 'Iron-sulfur World' was further expanded by Michael Russell, who suggested that life might have originated in iron sulphide bubbles (probotryoids) near alkaline and highly reduced hydrothermal vents (Russell et al., 1994). According to him, proto-metabolism was preceded by the formation of iron sulphide membranes, which might have precipitated on the sea floor between sulphidic, alkaline, and highly reduced hot spring waters, and the acidic, mildly oxidized, iron-bearing Hadean ocean. The hydrothermal vent could have been a source of chemical energy via H_2/CO_2 chemical potential and also would have concentrated the reactants in sufficient quantities in their microporous internal structures (Russell et al., 2007). The CO_2 reducing geochemistry at these vents, thus, might have supported the origin of first autotrophic organisms and the present day vent microorganisms might contain the biochemical pathways that are relics of these early microbial communities (Martin et al., 2008).

1.2.3 RNA-protein World

Another prevalent idea about early life is that of simultaneous presence, and even possible co-evolution, of RNA and proteins/peptides. Presence of RNA at the core of the proteinaceous ribosome structure strengthens this alternate hypothesis about life's origin. Recent studies on phylogenetic analysis of ribosomal RNA and proteins (Harish and Caetano-Anollés, 2012), and the accretion model of ribosome evolution (Petrov et al., 2015), hint at the presence of a putative ribonucleoprotein (RNP) World. The peptidyl-tRNA complex, which plays an important role during extant protein synthesis, is also considered as a relic from this putative RNP World. In its primitive format, the RNA-peptide and/or RNA-amino acid complex/s might have served as the enzyme-coenzyme complex (Guerrier-Takada et al., 1983). A related hypothesis, namely, 'tRNA core hypothesis' considers the proto-tRNA to have played an important role during proto-translation system, thus connecting the RNA World to the RNP World (de Farias et al.,

2016). A parallel concept of a protein world has also been proposed, wherein glycine, alanine, aspartic acid and valine (GADV-protein world hypothesis) are thought to have been the main players that led to the eventual origin of the genetic code (Ikehara, 2005).

1.2.4 Lipid World

All the living cells that we observe today are membrane bound. These membranes protect the cellular contents from the outer environment, while allowing for selective movement of molecules in- and out- of the cell. It is very important to keep the content, both the replicating entity and the metabolic networks, within a boundary, in order to achieve Darwinian evolution. The primary job of primitive encapsulating membranes would have been to shield the genetic material and metabolic networks from the outer environments, and sequestering it from random parasitic entities. The 'Lipid World' hypothesis states that amphiphilic molecules must have played an important role in demarcating this boundary condition in the primitive cells. It gains main support from the fact that amphiphiles can self-assemble and form supramolecular structures, such as micelles and vesicles, resulting in membranes wherein solutes can be encapsulated. It was first proposed by Segre et al. in 1999 (Segre et al., 1999) and was revitalized by work from the Szostak group, who emphasized the importance of encapsulation of genetic material for the formation of protocells (Szostak et al., 2001).

It is generally accepted that the protocellular membrane would have been made of simpler molecules due to the lack of elaborate machinery that is required for the synthesis of complex lipid molecules. Given this, fatty acids, which are the simpler amphiphilic counterparts, are thought to have been more likely a component of protocellular membranes, as against complex phospholipids. It is now well established that fatty acids of chain lengths C8 to C18, with varying degree of unsaturation in their alkyl chain, can self assemble to form vesicles above the critical bilayer concentration, and at pH around the pKa of their head group (Monnard and Deamer, 2003). Prebiotically plausible Fisher-Tropsch Type (FTT) synthesis, resulting in fatty acid formation has been demonstrated (McCollom et al., 1999).Furthermore, membrane-like structures are also found to be formed from non-polar molecules that were extracted from

Murchison, a carbonaceous meteorite (Deamer, 1985). Significantly, the presence of boundary condition, like that of fatty acid membranes, seems essential for, both, replication-first and metabolism-first approaches to the origin of life on prebiotic Earth.

Apart from the major hypotheses mentioned above, other ideas like clay-world hypothesis, PAH (polycyclic aromatic hydrocarbons) world hypothesis etc., have also been proposed. Amongst all these 'many worlds' hypotheses, the RNA World hypothesis remains the most popular amongst several scientists who work in the area of Astrobiology. The presence of probable 'molecular relics' from RNA World, which can be seen in extant biology in the form of RNA viruses, ribosomes etc, has prompted greater scientific enquiry towards understanding the events pertaining to the RNA World. A lot of efforts are ongoing to decipher the probable route from the mixture of prebiotic chemicals to the formation of a functional RNA molecule. A few of these important achievements are summarized in the next sections.

1.3 The RNA World hypothesis

Although the idea that RNA-based life preceded extant DNA/protein-based biochemistry remains speculative, it is largely supported by the fact that experimental models can be constructed to figure out the pathway for the origin and evolution of functional RNA molecules. For the formation of long enough RNA molecules that are also capable of self-replication, e.g. a replicase enzyme, it is important that we find out prebiotically plausible ways of synthesizing the relevant monomers and ways to get them linked together to have polymers of significant length, and so on. Subsequently, to achieve the formation of primitive cells and to study the role of Darwinian evolution on them, faithful replication of informational molecules and the encapsulation of these replicating entities, would have been crucial steps (Fig 1.4). These processes that are thought to have set the stage for the transition from non-life to life on the early Earth are discussed here.

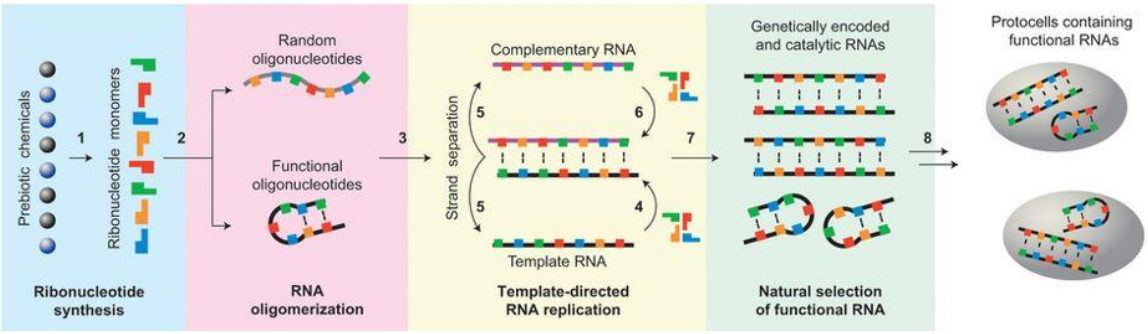


Figure 1.4: The plausible path for the formation of a protocell from a random pool of chemicals in the putative RNA World. The numbers indicate the sequential steps starting from monomer synthesis up to the formation of the primitive cell containing encapsulated RNA as genetic material (reproduced from Hernandez and Piccirilli, 2013).

1.3.1 Abiotic synthesis of monomers

For the formation of reasonable length functional RNA molecules, chemical synthesis of the monomeric units would have been the first crucial step. The ribonucleotides, i.e. the monomers of RNA, consist of three major parts viz. ribose, phosphate and the nucleobase. Hence, historically, scientists have tried to synthesize the RNA monomers by chemically joining these three different parts. Ribose can be synthesized as one of the products of Butlerow's formose synthesis (Orgel, 2004). This synthesis is usually carried out using alkaline solutions of formaldehyde, in the presence of simple mineral catalysts. The starting material first polymerizes to give glycolaldehyde under such conditions. It is later converted into glyceraldehyde and further into different tetrose, pentose and hexose sugars. Additionally, work from Eschenmoser's group has shown the formation of ribose-2, 4-diphosphate as one of the major products when using monophosphates of glycoaldehyde and glyceraldehydes in the formose synthesis (Orgel, 2004). More recently, considerable amount of ribose, along with other pentose sugars, was obtained by irradiating interstellar ice analogs (Meinert et al., 2016).

Phosphate group, which imparts the negative charge to the RNA backbone, might have been available in the mineral form on prebiotic Earth. The list of Haeden phosphate minerals described by Robert Hazen mainly includes the minerals from meteorite and lunar samples, and other phosphate minerals that are formed by metamorphism (Hazen, 2013). Exogenous

delivery by meteorites seems like the most prominent route of phosphorous accumulation on the prebiotic Earth. Four minerals, namely apatite, struvite, luneburgite and schreibersite are considered important as prebiotic phosphorylating agents for organic molecules (Pasek et al., 2017). Among them, luneburgite and schreibersite have been shown to be capable of phosphorylating nucleosides (Gull et al., 2015; Kim et al., 2016). Recently, diamidophosphate (DAP) has been demonstrated as an efficient (amido)phosphorylating agent for not only ribose and contemporary nucleosides, but also for amino acids and lipid precursors in aqueous environment (Gibard et al., 2017).

As far as nucleobase synthesis is considered, the first major research came from Juan Oró and his coworkers. They were able to synthesize adenine in considerable amounts, by refluxing a solution of ammonium cyanide (Oró and Kimball, 1961). Adenine and guanine can also be recovered from the dark solid material resulting from hydrogen cyanide (HCN) polymerization (Orgel, 2004). Adenine can also be obtained from eutectic freezing solutions of HCN (Miyakawa et al., 2002). HCN tetramer is considered as the important intermediate in both of these aforementioned reactions. Furthermore, adenine can also be obtained by directly heating formamide solution (Saladino et al., 2001). On the other hand, pyrimidines like cytosine have been shown to form from cyanoacetylene or its hydrolysis product cyanoacetaldehyde. Uracil can subsequently be formed by the hydrolysis of cytosine (Orgel, 2004). Recent research has also demonstrated the synthesis of all four RNA nucleobases in electric discharge experiments, under reducing atmosphere containing ammonia and carbon monoxide (Ferus et al., 2017). A plethora of nucleobases and nucleobase analogues have also been identified in carbon-rich meteorites (Callahan et al., 2011), suggesting that exogenous delivery might have been an important source of such nucleobases as well during the prebiotic era. Furthermore, nucleobases and nucleoside formation has been observed from liquid formamide in the presence of powdered meteorites and irradiation (Saladino et al., 2015). Thus, synthesis of all three individual parts of the RNA monomer has been demonstrated, albeit under varied reaction conditions.

The attachment of ribose to the nucleobase, however, has proven to be exceedingly difficult so far and is famously known as 'the nucleoside problem'. Direct heating of adenine with β-D-ribose results in only 3% of β-D-adenosine (Fuller et al., 1972; Orgel, 2004). Importantly, no nucleoside formation was observed when cytosine or uracil was used instead of adenine in this reaction. The ribosylation of nucleobase is low yielding, probably due to the presence of many forms of ribose (α/ β, furanose/ pyaranose etc), and many different tautomeric forms of the nucleobase, which are all present in the solution at any given time. This problem can be further enhanced for pyrimidines, due to the lack of a strong nucleophilic centre at the N1 position (Sutherland, 2010). This weak link in the formation of the RNA monomer is still largely unanswered and has prompted the idea of RNA World being predated by other pre-RNA World(s).

John Sutherland's group has taken an alternative approach towards achieving the synthesis of nucleosides. Sutherland and co-workers began with the synthesis of pentose amino-oxazolines (Anastasi et al., 2006), an intermediate that they later used in the successful synthesis of β-ribocytidine-2',3'-cyclic phosphate. They were also able to convert this molecule to β-ribouridine-2',3'-cyclic phosphate by using photochemistry (Powner et al., 2009). The same group also showed formation of a similar intermediate for purine nucleotide synthesis (Powner et al., 2010), however the exact chemical pathway to the actual purine nucleotide using this intermediate still remains elusive. Recent research from Steven Benner's group has used phosphorylated carbohydrates for demonstrating prebiotically plausible synthesis of nucleotides (Kim and Benner, 2017). An alternate prebiotic pathway for purine synthesis using formamidopyrimidines has also been suggested (Becker et al., 2016, 2018).

Although all these above mentioned syntheses have considerably advanced our knowledge in the field of prebiotic chemistry, a few of the daunting problems are yet to be addressed. The yield of a lot of these reactions is very low. Only those reactions, which are usually carefully tinkered to give the expected product(s), seem to yield the required form of the expected canonical nucleotide in considerable amounts. However, it is unclear whether the precise pathways described for these chemical reactions, are actually prebiotically plausible. The next

unanswered question is that of the specific selection of ribose, phosphate and the canonical nucleobases for the construction of the RNA monomers. We still don't have a clear understanding of how and why these three components were selected amongst the vast repertoire of related prebiotically available organic molecules. Therefore, researchers have recently started exploring the chemical space of plausible alternate nucleobases that might have resulted in primitive informational polymers of the pre-RNA World. Recent studies have have demonstrated the addition of non-contemporary bases, like barbituaric acid, melamine, 6-aminouracil etc., to ribose or ribose monophosphate , at both ambient conditions (Cafferty et al., 2016) and at high temperature regimes (Kim and Benner, 2015; Mungi et al., 2016). This has led to a renewed focus on re-understanding the steps that might be involved in the molecular evolutionary pathway, which would have allowed for the transition from the prebiotic starting soup to an RNA World.

1.3.2 Nonenzymatic oligomerization

Given a prebiotic pool of desired monomers in sufficient concentrations, the next step would be to chemically link these monomers to yield RNA oligomers. Nucleotide polymerization is an uphill reaction. Phosphodiester bond formation between two monomers requires removal of a water molecule and hence does not occur spontaneously in bulk water phase. Either pre-activation of the monomers, or a source of external energy, may increase the rate of this otherwise slow reaction. Different strategies have been tried so far in order to improve the yield of nonenzymatic RNA oligomerization reactions, some of which are described below. The pioneering work in this area of research came from Leslie Orgel's group. They started using activated nucleotides to carry out enzyme-free polymerization reactions (please refer to section 1.4.1 for details on activation chemistry). Orgel and coworkers began with studying the polymerization of activated nucleoside 5'-phosphorimidazolides (ImpNs) in solution. They found that such reactions yield a mixture of short, linear and cyclic oligomers, containing both 2'-5' and 3'-5' phosphodiester linkages. The role of metal ions as catalysts in such reactions was also studied, amongst which Pb^{2+} was identified as being one of the best (Lohrmann et al., 1980; Orgel, 2004). The same group also analyzed the polymerization potential of nucleoside cyclic -

2', 3'- monophosphates for the first time. They carried out dry-state polymerization of cyclic 2', 3' AMP (cAMP) in the presence of different catalysts like amines, imidazole etc. Polymers up to 6-mers were observed (Verlander et al., 1973). Polymerization efficiency of cyclic nucleotides was further studied mainly by Ernesto Di Mauro's group. Di Mauro and coworkers have explored the self-polymerization capability of cyclic-3', 5' GMP (cGMP) and cyclic-3', 5' AMP in water. They found that the cyclic-3', 5' GMP forms longer RNA oligomers amongst the two monomers (Costanzo et al., 2009, 2016).

Polymerization of RNA monomers is also achievable in eutectic solutions. Metal-catalyzed polymerization of imidazole-activated nucleotides has been demonstrated in ice-water eutectic phase solutions, at the temperature of around -18.4˚ C (Fig 1.5A) (Kanavarioti et al., 2001; Monnard et al., 2003). Oligomers up to 17-mers have been observed in these polymerization reactions, which were carried out in the absence of any template. Mg^{2+} and Pb^{2+} ions have been shown to act as catalysts in these reactions too. The monomers, along with other solutes such as salts, get concentrated in the brine channels at and below the eutectic point, thus increasing the chances of bond formation between two monomers. Higher yield of oligomers during the eutectic phase polymerization is thought to be promoted due to the preservation of the activated monomers at these lower temperatures, lesser degradation rates of the resultant oligomers and more inter-nucleotide interactions.

Another method that has been explored mainly by James Ferris's group is the clay-catalyzed oligomer formation from monomers. Ferris and coworkers have used montmorillonite, a clay mineral, as a surface catalyst, to concentrate the reacting monomers and facilitate bond formation between them. They first studied the polymerization of 5'- AMP in the presence of carbodiimide as condensing agent (Ferris et al., 1989) and later, more successfully, the polymerization of phosphorimidazolide monomers using Na^{+}-montmorillonite as a catalyst (Ferris and Ertem, 1992). They were able to form oligomers up to 11-mers using ImpA as the monomer. The predominant link between the two monomers was observed to be that of a 3', 5'-phosphodister bond (Ferris and Ertem, 1993). Later, they could achieve polymerization of up to 40-50 mer length molecules using 1-methyladenine-activated monomers and

montmorillonite clay as catalysts, with considerable regioselectivity for the 3', 5'-phosphodister bond (Fig 1.5B) (Huang and Ferris, 2003, 2006). Clay surfaces can facilitate the aforementioned enzyme-free polymerization by concentrating the monomers, and the ions adsorbed on the clay surface can also act as catalysts in these reactions (Ferris, 2006). X-ray diffraction studies suggests the presence of stacked nucleotides on the clay surface as is signified by the peak distance of 3.4 A˚ in the observations (Himbert et al., 2016).

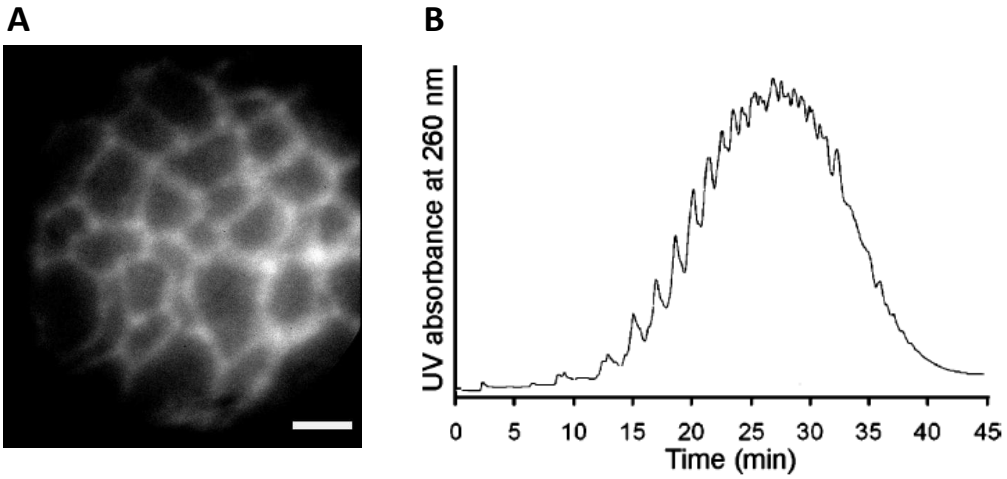


Figure 1.5: Different means of achieving nonenzymatic oligomerization (A) ImpU oligomerization in eutectice ice phase as visualized using acridine orange staining of the oligomers, scale bar =26.7µm (adapted from Kanavarioti et al., 2001) (B) Anion exchange HPLC chromatogram showing montmorillonite clay catalyzed formation of oligomers up to 50-mers from adenine-5'-phosphoro-1-methyladeninium (reproduced from Huang & Ferris, 2006).

Nonenzymatic polymerization of the more prebiotically pertinent non-activated nucleic acid monomers has also been studied, under fluctuating environments such as alternate cycles of dehydration and rehydration (DH-RH). Such cycles are thought to have been prevalent on the prebiotic Earth and might have resulted due to the temperature fluctuations occurring during day-night cycles or seasonal variations. Dehydration would have served as a means to concentrate the starting monomers. This phase can also promote bond formation between the monomers due to the higher temperatures involved in these reactions that promotes

condensation. The first demonstration of polymerization of nucleoside- 5'- monophosphate monomers, under such DH-RH conditions, came from David Deamer's group. The researchers observed formation of 50-75-mer long RNA-like polymers, when the 5'-NMPs were subjected to alternate DH-RH cycles at high temperature, and in the presence of phospholipids (Rajamani et al., 2007). Lipids are known to form liquid-crystalline matrices under dehydrated conditions, facilitating the concentration of the starting reactants (e.g. nucleotides) within the inter-layers of the resultant multilamellar structures (Toppozini et al., 2013; Himbert et al., 2016). Lipids are also thought to protect the growing oligomers from hydrolysis that is prevalent under these harsh reaction conditions. A follow-up study from our laboratory later also showed the polymerization of sugar-phosphate backbones under similar reaction conditions (Mungi and Rajamani, 2015).

1.3.3 Nucleic acid replication

Along with the nonenzymatic polymerization of abiotically synthesized monomers, it would have been necessary to correctly replicate the encoded information so as to maintain the sequence, and thus the function of the resultant RNA polymers. Life can be said to have originated only when the replication of information began to occur (Chen and Nowak, 2012). In this context, a self-replicating 'replicase' like ribozyme is considered to have been of prime importance for the evolution of the RNA World (Robertson and Joyce, 2010).

1.3.3.1 RNA mediated RNA replication

Copying the information from a template can, in principle, be achieved by the ligation of two template-bound RNA molecules. In 1993, Bartel and Szostak for the first time evolved class I RNA ligase ribozyme from a random pool of around 10^{15} RNA oligomers (Bartel and Szostak, 1993). This ligase was able to generate the 3'-5' phosphodiester bond at a rate that was 10^{7} times faster than the uncatalyzed reaction. Gerald Joyce's lab has also extensively worked on ribozymes pertinent to the RNA World. Earlier studies from this group involved the successful formation of 'R3C' ligase ribozyme (Rogers and Joyce, 2001), which was further evolved to

ligate the daughter strands to produce more copies of the ribozyme itself (Paul and Joyce, 2002). This self-replicating R3C ligase was later evolved into a cross-catalytic variant for self-sustained replication of the ribozyme (Kim and Joyce, 2004; Lincoln and Joyce, 2009). Another efficient ligase came in the form of an improved DSL ligase ribozyme, achieved by continuous *in vitro* evolution, and was found to be approximately 10^5 times better catalyst than the original DSL ribozyme (Voytek and Joyce, 2007). Recent advances in this area of research include: a) the formation of a self-replicating ribozyme by the ligation of three shorter RNAs, establishing cooperative networks in an RNA population (Vaidya et al., 2012), b) evolution of ligase ribozymes capable of joining substrate RNAs of opposite chirality (Sczepanski and Joyce, 2014) and c) assembly of active ribozymes from oligonucleotides no longer than 30-mers, using freeze- thaw cycles (Mutschler et al., 2015), to name a few.

The evolutionary potential of ligase ribozymes, however, could be limited as they can only carry out the assembly of comparatively longer stretches of RNA. On the other hand, RNA polymerase ribozymes can copy the information in the template using nucleoside triphosphates, leaving greater room for evolution due to higher possible chances of mutations during copying (Martin et al., 2015). The first ribozyme-catalyzed RNA polymerization was reported almost 22 years ago, wherein an evolved variant of the above-mentioned class I ligase ribozyme (Bartel and Szostak, 1993) was engineered to extend a primer by six nucleotides (Ekland and Bartel, 1996). Another landmark discovery from the same group was the *in vitro* evolution of the 'round 18 (R18) ribozyme' from the aforementioned study, which facilitated templated-primer extension of up to 14 nucleotides (Johnston et al., 2001). This R18 ribozyme was further evolved into a variant named B6.61, for promoting a truly *trans*- primer-template extension of up to 20 nucleotides (Zaher and Unrau, 2007). The efficiency of the R18 ribozyme was found to have increased when the primer-template complex and the ribozyme was co-localized on a micelle by means of attached hydrophobic anchors (Müller and Bartel, 2008). On the other hand, Holliger's group has been trying to increase the efficiency of polymerase ribozymes by evolving them at sub-freezing temperatures (Attwater et al., 2010, 2013). Starting from the R18 sequence, the group has been successful in evolving a more general RNA polymerase ribozyme that was able to synthesize RNAs up to 95 nucleotides long (tC19

ribozyme), and was also able to transcribe information of another active ribozyme (tC19Z ribozyme) (Wochner et al., 2011). Joyce's lab recently has evolved a more efficient polymerase ribozyme capable of copying complex functional RNA sequences (Horning and Joyce, 2016), as well as another ribozyme capable of performing reverse transcription (Samanta and Joyce, 2017).

Although we now have an array of these ligase and polymerase ribozymes, none of them really fulfill the task of total self-replication. Also, most of the above-mentioned ribozymes are sequence specific when it comes to their catalytic activity. Furthermore, these ribozymes are fairly long in length, which is not trivial to achieve simply by chemical polymerization. To evolve functional ribozymes it would have been fundamental to have short RNA sequences, which self-replicated without the help of any enzyme, simultaneously undergoing random ligations and mutations during replication. This process would have been crucial in spanning a considerable sequence space for the evolution of a long enough RNA molecule with catalytic activity. Therefore, it is important to first consider nonenzymatic replication of informational polymers; a vital step prior to the evolution of functional ribozymes. The advent in the research area of nonenzymatic replication of nucleic acids, in the context of an RNA World, is encapsulated in the next section.

1.4 Nonenzymatic replication

During the emergence of life, the template required for information copying could have been obtained using different prebiotically plausible means of nonenzymatic oligomerization as illustrated in section 1.3.2. Depending on the availability of relevant primer/s, this enzyme-free template-directed polymerization could have been achieved in three different ways (Fig 1.6). In the absence of any primer, the information from the template can be copied via assembling the cognate monomers against the templating bases (Fig 1.6A). When another complimentary short oligomer binds to the template and acts as a primer, the information copying can be achieved by extension of this pre-existing primer (Fig 1.6B). Also, when multiple complementary oligomers bind to the same template, the copying can be achieved by inserting the monomers

between two complementary oligomers to obtain the longer stretch of the copied strand (Fig 1.6C). This section describes the abovementioned means of attaining nonenzymatic replication. It also discusses the fidelity of replication associated with such reactions.

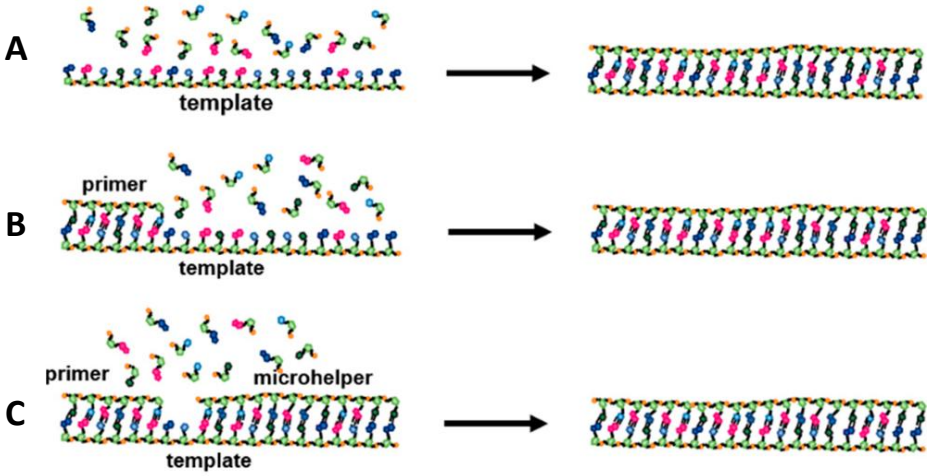


Figure 1.6: Nonenzymatic template-directed information copying by (A) alignment of the monomers against the template (B) primer-extension (C) primer-extension in the presence of downstream helper oligomer (adapted from Kaddour & Sahai, 2014).

1.4.1 Template-directed polymerization

In this approach, the polynucleotide is incubated with the monomers or short oligomers in order to get the complementary strand via ligation driven reaction/s (Fig 1.6A). Along with the studies on nonenzymatic polymerization; pioneering work in this area of nonenzymatic template copying also came from Leslie Orgel and his co-workers. The first experiments pertaining to the transfer of information from an RNA template have used carbodiimide as activating agent. A mixture of oligomers was obtained when either polyuridylic acid or polycytidylic acid were used as templates (Sulston et al., 1968, 1969). These reactions were not as efficient and/or regiospecific. To obtain the product strand yield in higher quantities, and with a specific phosphodiester bond joining the two adjacent nucleotides, researchers have since used 'activated nucleotides' (Fig 1.7). The nucleotides pre-activated with phosphoramidate chemistry on the 5'-phosphate group, mainly phosphorimidazoles, have been widely used to study template-directed copying processes (Orgel, 2004). However, it is to be

noted that the plausible synthesis of such nucleotides prebiotically, and, especially, in sufficient quantities, has not been conclusively demonstrated to date. A very recent study reported the formation of imidazole-activated ribonucleotides by means of a prebiotically viable method (Yi et al., 2018). Nonetheless, another important aspect for consideration is that of the long-term stability of such monomers. Their intactness over a prolonged period remains questionable due to their high intrinsic energy. Despite these drawbacks, the reactions involving activated monomers have shed light on many basic principles and processes involved in chemical polymerization and copying, since these reactions are very difficult to achieve using canonical nucleotides. The knowledge accrued from these studies has given important insights into pertinent set of reactions that might have occurred during molecular evolution on prebiotic Earth.

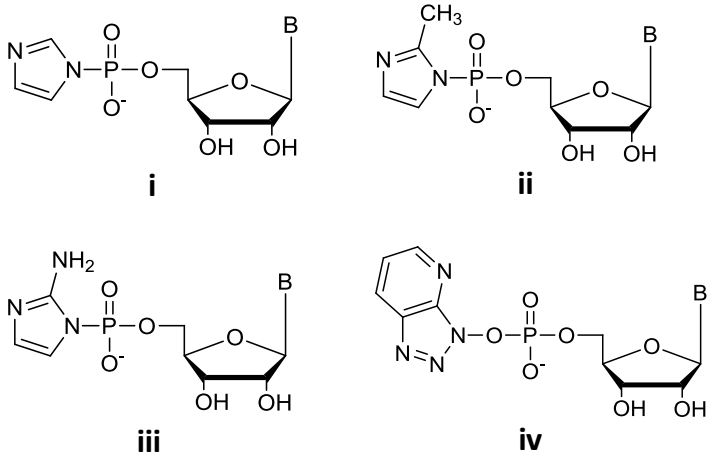


Figure 1.7: Selected leaving groups used in activated monomers (i) imidazole (ii) 2-methylimidazole (iii) 2-aminoimidazole (iv) oxyazabenzotriazole. 'B' represents the nucleobase (B= A/U/G/C).

In the first set of experiments that used imidazole-activated nucleotides to copy the information from a nucleic acid strand, the researchers used guanosine-5'-phosphorimidazolides to synthesize oligo-G strands against the poly-C template, in the presence of metal ions. Different metal ions tested had different specificity towards the formation of 2'-5' vs. 3'-5' phosphodiester bond. For example, use of Pb^{2+} and Mg^{2+} resulted in

higher percentage of the 2'-5' phosphodiester bond in the resultant products (Lohrmann et al., 1980). When the activation group was replaced from imidazole to 2-methylimidazole (i.e. when using 2-MeImpG instead of ImpG), the resultant products predominantly contained the canonical 3'-5' phosphodiester bond (Inoue and Orgel, 1981). The 2-methylimidazole activation chemistry seems superior as longer oligomers were obtained even in absence of any metal ions (other than Mg^{2+}). Interestingly, the activated nucleotides 2-MeImpG and 2-MeImpC were also able to copy a mixed sequence of poly (CG) (Joyce and Orgel, 1986). It was also seen that the templates that adopted an A-type helical structure were better copied using 2-methylimidazole activated monomers. Hence, RNA and other modified nucleic acids, like HNA (Hexitol Nucleic Acid) and ANA (Altritol Nucleic Acid), proved to be better templates than DNA for these reactions (Kozlov and Orgel, 2000). Additionally, to enhance the regiospecificity and efficiency of the reaction, Orgel and his colleagues also modified the 2'- or 3'- hydroxyl terminal of the nucleotide with an amino group, which being more nucleophilic, enhanced the rate of bond formation between two monomers. As far as template-directed oligomer formation was considered, modifying the 2'/3'-hydroxyl terminal with amino group gave similar results as that of replacing the 'Imp' group with '2-MeImp' on the 5'- phosphate end of the nucleotides (Zielinski and Orgel, 1985, 1987). This strategy has been later used widely in analyzing the template-directed primer extension reactions and the results are summarized in the following subsection.

Notably, template copying, using non-activated nucleotides, remains largely unexplored. Thus far, only a single proof-of-concept study has been demonstrated by David Deamer's group, albeit using a DNA-based system. They carried out information transfer from a DNA template, under repeated cycles of dehydration and rehydration, at low pH and high temperature. The product yield was around 0.5%, with an ascertained misincorporation rate of 9.9%, as analyzed by sequencing (Olasagasti et al., 2011). To the best of our knowledge, no RNA-template based information transfer study, using nucleoside-5'-monophosphates, has been reported yet.

1.4.2 Template-directed primer extension

The role of a pre-formed primer, in template copying reactions, was first evaluated using oligomers capable of forming hairpin structures. This provides a pseudo primer-template complex as the oligomer folds upon itself to form the specific secondary structure (Fig 1.8A). Using 2-MeImpNs as monomers in this system, researchers found out that the cognate copying of 'C' template was most efficient while copying of the 'A' template base was the least efficient. More importantly, the presence of two consecutive 'A' residues in the template region, almost always halted the replication (Wu and Orgel, 1992a, 1992b, 1992c).

Other significant contribution to the field of nonenzymatic nucleic acid replication has come from Jack Szostak's group. In a first of its kind study, the group demonstrated template-directed primer extension, using 2'-amino-2', 3'-dideoxy-5'-phosphorimidazolide monomers (2'-NH_2-ImpddN), for both homopolymeric (Mansy et al., 2008) and heteropolymeric (Schrum et al., 2009) templating regions. The rate of the extension was enhanced due to the 2'-amino modification on, both, the incoming nucleotide, and the 3'-terminal nucleotide of the primer (Fig 1.8B). As stated before, the amino group is a better nucleophilic group than the canonical hydroxyl group. Amongst the different templates used, the researchers observed maximum extension efficiency for the RNA and LNA templates, highlighting the importance of A-type helical structure for nonenzymatic copying reactions (Schrum et al., 2009). Use of the aforementioned activation chemistry typically resulted in oligomers linked by 2'-5' phosphodiester bonds, and not the canonical 3'-5' phosphodiester bonds. However, the presence of 2'-5' phosphodiester bonds in the oligomers, has been suggested to be useful for strand separation, thus being advantageous for facilitating repeated cycles of nonenzymatic replication. Furthermore, the presence of up to 25% of 2'-5' phosphodiester bonds, has also been shown to be tolerated by a functional ribozyme, suggesting the possibility of the presence of such mixed linkage backbones during early evolution of functional polymers (Engelhart et al., 2013).

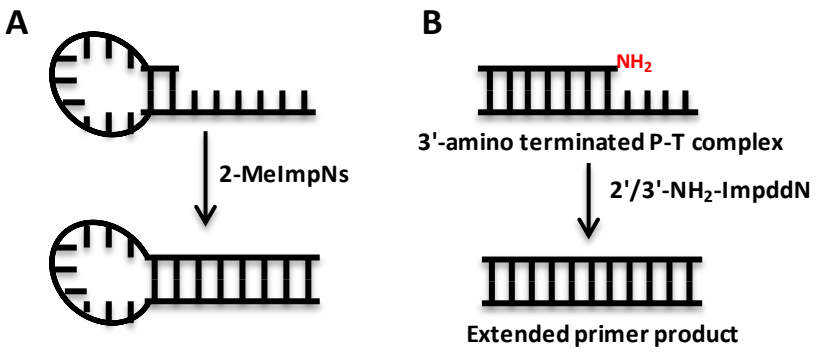


Figure 1.8: Template-directed enzyme-free primer extension using (A) 2-methylimidazole activated monomers and oligomers capable of forming hairpin structure (that provides the templating region) (B) 3'-amino modified primer and 2' or 3'-amino modified imidazole activated monomers.

Richert and colleagues have been contributing to these studies using different activation chemistry, namely 1-hydroxy-7-azabenzotriazole (HOAt) (Fig 1.7). Using this non-phosphoramidate activating group, and 3'-amino terminated primers, the group has been able to achieve faster rates of replication (Stütz et al., 2007). They have also achieved DNA copying in both 3'- and 5'-direction, by alternating the amino and -OAt activation chemistry, on the 5' and 3'end of the nucleotide (Kaiser et al., 2012). On the other hand, Monnard's group has analyzed Pb^{2+}/Mg^{2+} catalyzed template-directed primer-extension, under eutectic ice-water mixture conditions, using imidazole or 2-methylimidazole activated monomers. Although the resultant products contained higher percentage of the non-canonical 2'-5' phosphodiester linkages, enhanced copying of otherwise difficult to copy sequences, viz. 'AA', 'AU' and 'AG', was achieved (Löffler et al., 2013). The synthesis of template-directed 3'-5' linked DNA (3'-NP-DNA) has also been demonstrated, using phosphoramidate activated 3'-amino-2', 3'-dideoxyribonucleotides, on both homopolymeric (Zhang et al., 2013a) and heteropolymeric templates (Zhang et al., 2013b) (Fig 1.8B). The cyclization of 3'-amino activated monomers is a considerable side-reaction in such cases. However, due to the faster rate of template-directed polymerization, the synthesis of the daughter strands was obtained in detectable yields, despite the competing monomer cyclization. The 3'-amino activated monomers also performed better while copying all the four canonical nucleotides, as opposed to the previously studied 2'-amino

activated monomers, which were able to copy ‘G’ and ‘C’ nucleotides to a greater extent than the ‘A’ and ‘U’ nucleotides (Zhang et al., 2013a). A new phosphoramidate-based activation group viz. 2-aminoimidazole (Fig 1.7) was recently used to activate the monomers, and was found to be better than the 2-methylimidazole leaving group in promoting enzyme-free RNA copying. Both, the rate and the yield of primer extension was enhanced due to the higher reactivity of this new activation group (Li et al., 2017).

Further, it is important to understand the replication of membrane encapsulated template-primer complex as it would considerably help in the construction of a viable protocell. A seminal work published in 2008 , reported for the first time the enzyme-free DNA template-directed primer extension inside a model fatty acid vesicle using activated monomers (Mansy et al., 2008). RNA copying, using similar activation chemistry, however, requires considerable amounts of Mg^{2+} ions. Nonetheless, this high a concentration of magnesium is not compatible with fatty acid vesicles, which are known to get destabilized under higher ionic concentrations (Monnard et al., 2002). This poses a remarkable conundrum while integrating the RNA replication system inside the fatty acid vesicle, which is thought to have constituted the initial boundary structures. The ‘Mg^{2+} problem’ was solved by Szostak’s group, at least partially, by using citrate to chelate the Mg^{2+} ions effectively. This allowed for the protection of the fatty acid vesicles while allowing chemical copying of a homopolymeric template to occur inside these vesicles (Adamala and Szostak, 2013). The same group also recently achieved copying of a heteropolymeric RNA template inside fatty acid vesicles using citrate-chelated Mg^{2+} ions and 5'-phosphoro-2-aminoimidazol-activated monomers (O’Flaherty et al., 2018).

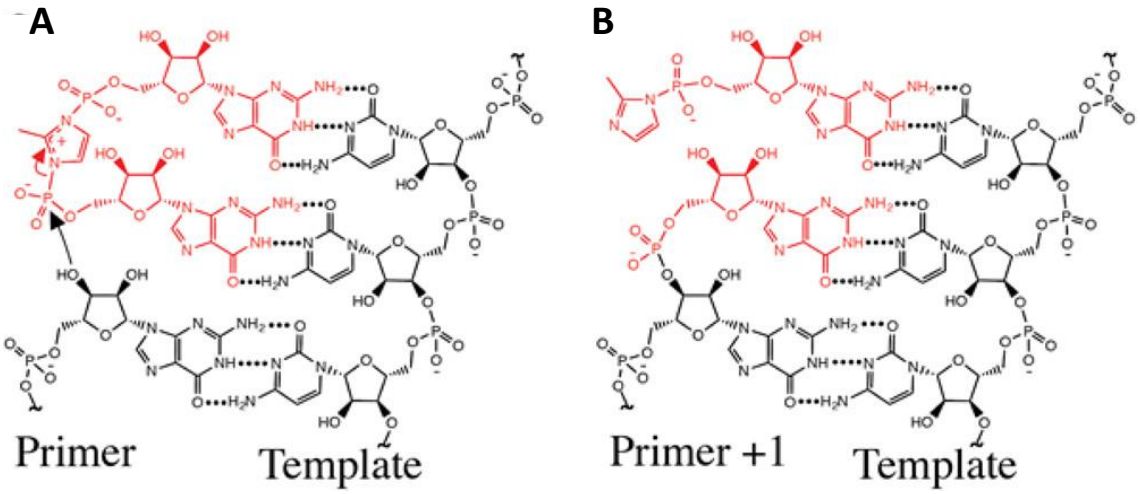


Figure 1.9: Proposed mechanism for imidazolium-bridged dinucleotide intermediate mediated template-directed primer extension. Along with the formation of extended primer product (A), the bridged dinucleotide also results in regeneration of the second activated monomer (B) (reproduced from Walton & Szostak, 2016).

As far as the underlying mechanism for enzyme-free template-directed primer extension goes, for a long time it was thought to include an initial binding of a monomer to its cognate base on the template. This was thought to be followed by the attack of a 3' (or 2') hydroxyl group, present on the extending end of the primer, on the 5'-phosphate of the incoming monomer (S_N2 type reaction), subsequently resulting in primer extension. However, a recent study suggests that two 5'-phosphoryl-2-methylimidazole-activated monomers (2-MeImpNs), react with each other to form an imidazolium-bridged dinucleotide intermediate, prior to binding to the template. The following attack of the 3'-hydroxyl of the primer on this dinucleotide, leads to displacement of an activated nucleotide as the leaving group, and results in primer extension by one nucleotide (Fig 1.9) (Walton and Szostak, 2016). This proposed mechanism was further supported by X-ray crystallographic studies of an RNA template-bound analogous triphosphate-bridged dinucleotide (GpppG). It reveled that the dinucleotide tightly binds to the template via Watson-Crick pairing and forms a highly ordered structure, conducive for the attack of 3'-hydroxyl of the primer (Zhang et al., 2017). In addition, the factors that would result in perturbation of the pre-organized geometry of this reaction centre, such as mismatched base

pairing, were shown to reduce both the rate, and the regiospecificity of the nonenzymatic primer extension (Giurgiu et al., 2017).

1.4.3 Template copying using multiple short oligomers

The early experiments on template-directed ligation of oligomers, involved the formation of T_{12} oligomer from two T_6 oligos, using a dA_{12} template and carbodiimide as the condensing agent (Orgel, 2004). Template-directed ligation of phosphorimidazole-activated short oligonucleotides has been analyzed by Orgel and coworkers, and was found to be inefficient (Ninio and Orgel, 1978). The longer oligonucleotides, activated at 5'-end with triphosphates, showed better ligation competence, resulting in a higher yield of the canonical 3'-5' phosphodiester linked products (Rohatgi et al., 1996). When compared to prior studies of template copying using activated monomers, the oligomer ligation proved to be better, mostly due to the improved Watson-Crick base pairing between the template and long oligomers.

In addition to the use of a primer, downstream 'helper oligonucleotides' have also been used to improve nonenzymatic template-directed primer extension. Richert and co-workers first adopted this approach in order to enhance the rate of template-directed reactions, using -OAt activation chemistry. Using a downstream binding pentadecamer, the researchers were able to copy a purine template base, which is otherwise difficult to achieve (Vogel et al., 2005). The terminal nucleotide of the template is known to replicate at a much slower rate than the one in the internal part of the template. With the help of downstream binding oligomers, the last template base can be pseudo-internalized, thus resulting in enhanced rate of the copying (Stütz et al., 2007). Using this strategy, and immobilized RNA strands, the group also achieved copying of all four templating bases in high yields as analyzed by MALDI-TOF (Deck et al., 2011). Immobilized RNA used in this study, helped in replacing the spent activated monomers, which otherwise compete for the template binding and result in an overall reduction in the rate of the reaction. Recently, Szostak's group has also used downstream helper oligos in their reaction, along with phosphoramidate-based activation chemistry. They have reported use of activated downstream helper trimers to enhance the copying of a heteropolymeric RNA template, with 2-

aminoimidazole activated nucleotides (Li et al., 2017). While probing the thermodynamics using NMR and ITC, these researchers found that binding of 5'-GMP to the template was increased by about two orders of magnitude in the presence of tightly bound downstream oligomers (Tam et al., 2017). Additional structural analysis revealed better pre-organization of the monomer on the primer-template complex in the presence of the helper oligomer, thus favoring the reactive conformation that allowed for the in-line attack of the 3'-hydroxyl group of the primer (Zhang et al., 2018).

1.4.4 Fidelity of the nonenzymatic replication

In order to conserve the function of RNA, it would have been essential to conserve its secondary structure, which is usually sequence dependent. To conserve the sequence, faithful replication of RNA would have played a crucial role. Given the lack of any specific error-correction machinery, nonenzymatic template-directed replication would have been a highly error-prone process. Nevertheless, there is a limit to the number of misincorporations that can be tolerated while propagation of information to subsequent generations. This concept, which is popularly known as the 'error threshold' problem, was originally put forth by Manfred Eigen in 1971. In simple terms, it states that the maximum genome information that can be carried in a sequence is inversely proportional to its mutation rate (Eigen, 1971). This would provide an intrinsic advantage to correctly replicating sequences, over its competitors, in a given pool of replicating sequences. Although such 'superiority' might be limited to short periods of time, especially due to the emergence of novel variants of the sequence, it would at least be advantageous for the initial selection of efficiently replicating sequences, from a pool of not-so-efficient RNA replicators. Such scenarios might have been evolutionarily important in promoting the overall fidelity associated with a self-replicating system. Thus, the maintenance of a high accuracy of replication would have played a crucial role, not only for an elevated propagation of the correctly replicating sequences, but also for increasing the size of the genome during the origin and early evolution of life. It is estimated that in order to conclusively establish an RNA-based living system, a fidelity rate of 99% is required, which can maintain a genome size of around 100 nucleotides (Robertson and Joyce, 2010).

In the experimental set-up, however, it has been excruciatingly difficult to achieve high fidelity rates in the context of nonenzymatic replication of nucleic acid sequences. This is mainly due to the fact that none of the four contemporary DNA/RNA nucleotides, while acting as a template base, get nonenzymatically copied with the requisite rate and accuracy. Early work in this regard from Orgel and his colleagues, uncovered that the incorporation of 'G across C' was the most efficient amongst the lot, while that of 'U across A' was the least efficient. The other two cognate additions occurred with moderate efficiency. In general, it has been very difficult to copy 'AU' rich sequences, and the presence of two consecutive 'A' residues usually was observed to halt the replication process (Orgel, 2004). The wobble pairing of 'G across U' was noted to be the most frequently occurring mismatch, such that a highly 'GC' rich template would not be able to propagate itself faithfully through subsequent cycles of replication. The different activation chemistries (as elaborated above, Fig 1.7) used to enhance the rate of nonenzymatic copying reactions, have not contributed significantly to increasing the accuracy of these reactions. The observed frequency of the incorrect incorporations remained close to 20%, thus limiting the size of the correctly replicating genome, to merely a 5 nucleotides long sequence in length (Hagenbuch et al., 2005). A study has suggested that the transition from RNA to DNA might have helped in increasing the genome size in general, as enzyme-free DNA replication was found to be less error-prone than its RNA counterpart (Leu et al., 2011).

The original 'error threshold' calculations presumed that template-directed polymerization would proceed at a similar rate regardless of any misincorporation that might take place during the process. However this doesn't seem to be the case. A pioneering study carried out by Irene Chen's group demonstrated that the rate of template-directed primer extension in a DNA-based system slowed by up to two orders of magnitude post a mismatched addition (Fig 1.10A). This might help in increasing the genetic information content despite comparatively higher intrinsic mutation rate, as the correctly replicating templates tend to complete their replication faster than the mutated sequences due to post-mismatched stalling (Rajamani et al., 2010). A follow-up study from the same group showed that an initial mismatch during enzyme-free template copying is likely to be followed 54-75% of the time by another mismatched addition. Although, the addition of subsequent cognate nucleotides restored the rate of the reaction, the

researchers found that a cascade of mismatches considerably slowed the rate of the incoming nucleotide addition, thus further favoring the correctly replicating sequences (Fig 1.10B). Simultaneously, simulation experiments revealed that an initial error was likely to be followed by several consecutive mutation events, thus taking larger steps through the sequence landscape. Such events might have helped in rapid exploration of the sequence space, while simultaneously conferring 'superiority' to the correctly replicating template-primer complex/s (Leu et al., 2013). Both these studies underline the importance of a matched primer-template terminus in order to achieve efficient primer extension.

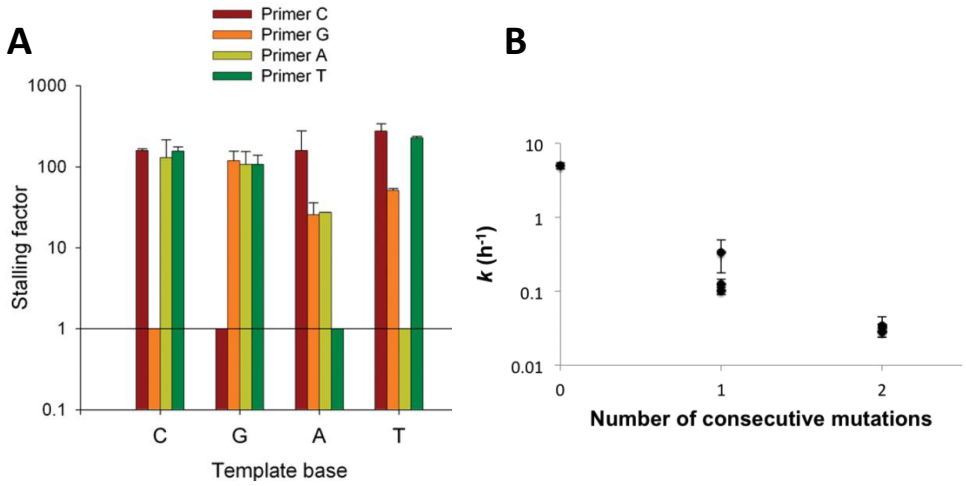


Figure 1.10: Rate of nonenzymatic copying slows down post mismatch/s. (A) Addition of the next matched nucleotide is stalled by up to two orders of magnitude after a mismatch in RNA primer-template system. The X-axis indicates the template base and the primer indicates the 3'-terminal base on the primer (reproduced from Rajamani et al., 2010) (B) Rate of addition of 'G across C' in RNA primer-template system decreases drastically with each consecutive mutation (reproduced from Leu et al., 2013).

Szostak's group, on the other hand, has adopted the strategy of using modified nucleotides in order to increase the fidelity of enzyme-free replication of nucleic acid sequences. About five years ago, the group reported the use of 2-thio modified thymidine nucleotide in order to suppress the G:T and A:C wobble-pairing, during enzyme-free replication in a DNA system. They showed an increase in, both, the rate and fidelity of copying 3'-NP-DNA template, when using 2-thiothymidine instead of a thymidine nucleotide (Zhang et al., 2013b). Similar results were

obtained when using 2-thio-U (s^2U), while copying of RNA template (Heuberger et al., 2015). Structural insights into the RNA heptamer duplex containing s^2U revealed that the presence of this modified nucleotide stabilized U:A base pairing likely via the increased stacking interaction and a more constrained the sugar conformation, without any apparent structural perturbation (Sheng et al., 2014). Furthermore, thermodynamic studies using ITC and UV thermal denaturation revealed that RNA duplexes containing internal s^2U:A base pair, are better stabilized and the stabilization is of an entropic origin (Larsen et al., 2015). In general, the 2-thiolation is now established to better stabilize the C3'-endo conformation, similar to what is seen in structured RNA, thus resulting in higher rate and fidelity of nonenzymatic copying reactions.

1.5 Summary

The emergence of life on Earth remains an intriguing mystery. Many scientific hypotheses have been proposed to delineate the plausible steps towards the formation of first living entities, from a pool of prebiotic chemicals (Krishnamurthy, 2017). Amongst these, the hypothesis of 'RNA World' is most widely accepted due to the presence of possible relics, such as the ribosome core, self-splicing ribozymes, RNA viruses etc., in extant biology. This hypothesis presumes the existence of catalytic RNAs, which were also capable of storing and propagating genetic information (Robertson and Joyce, 2010). The initial formation of such catalytic RNA molecules must have been preceded by chemically driven oligomerization and replication reactions during the origin and early evolution of life. Extensive experimental and *in silico* studies have been undertaken to better understand enzyme-free monomer formation, and their subsequent oligomerization and replication (Kitadai and Maruyama, 2018). Using different strategies like activated molecules, confinement, eutectic phases of water-ice mixtures, helper oligonucleotides, alternate cycles of dehydration and rehydration etc., researchers have been able to obtain nucleic acid polymers chemically (Wachowius et al., 2017). However, the rate and accuracy of these reactions still remains a topic of concern for prebiotic chemists. Another important consideration here is the fact that almost all of these reactions discussed are usually

carried out in prebiotically 'not-so-realistic' scenarios, such as in buffered solutions, starting with high concentrations of pure and activated chemicals. Although, knowledge accrued from such experiments is crucial to our understanding of abiotic formation of life, the reactions themselves might not have proceeded in a similar way in a complex prebiotic milieu. To summarize, despite this large body of existing work, we substantially lack the understanding of progress of pertinent nonenzymatic reactions in prebiotically realistic scenarios.

1.6 Scope of the thesis

Before the advent of a functional 'replicase' like ribozyme, nonenzymatic replication reactions would have played a crucial role in a putative RNA World. Several studies carried out to date have analyzed and tried to improve the rate and accuracy of such reactions (as detailed above). However, almost all these studies have been carried out in a chemically simplistic system, which is mostly not prebiotically realistic. Understanding RNA-based enzyme-free copying, under prebiotically pertinent set-up, will actualize our knowledge related to discerning the transition from chemistry to biology on prebiotic Earth.

Chapter 2: In order to incrementally increase the complexity of the starting reaction mixture, we used lipid molecules and polyethylene glycol as co-solutes. Nonenzymatic template-directed primer extension reactions were used as a proxy to understand the effect of 'random' co-solutes, and molecular crowding agents, on pertinent prebiotic reactions of a putative RNA World. Fidelity and mutation rates were calculated for this system, both, in the absence and presence of co-solutes. Furthermore, computer simulation studies were undertaken in collaboration with Dr. Julien Derr at Paris Diderot University, France. These studies allowed us to analyze these experiments under aforementioned conditions, over multiple replication cycles; a scenario which is otherwise experimentally limited due to the intrinsic issues associated with our study system (e.g. Strand-separation for subsequent replication cycle, RNA hydrolysis over prolonged periods).

Chapter 3: To understand the underlying mechanism of the effect of lipid vesicles, and/or PEG, on enzyme-free RNA copying reactions using activated monomers, we carried out few

biophysical studies. The effect of presence of PEG on the RNA monomers was analyzed using NMR technique. DOSY NMR and T_1 relaxation time measurements were used (undertaken in collaboration with Harshad Paithankar and Dr. Jeetender Chugh, Dept. of Chemistry, IISER, Pune) to estimate the diffusion time and molecular size of the nucleotide clusters at different concentrations, in the absence and presence of PEG 8000. On the other hand, the plausible RNA-lipid interactions were examined using fluorescence imaging using SUPER template and SLB assay systems (carried out in collaboration with Soumya Bhattacharyya and Dr. Thomas Pucadyil, Dept. of Biology, IISER, Pune).

Chapter 4: To understand the prebiotic plausibility of the fundamental reactions discussed in Chapter 2, we also carried out RNA template-based copying, using prebiotically pertinent non-activated nucleotides. To our knowledge, such studies have not been reported to date. Towards this, templated-replication reactions were carried out under repeated cycles of dehydration and rehydration at high temperatures, in the presence of lipids as co-solute molecules. The efficiency of the reaction was analyzed under varying conditions to evaluate the role of parameters like temperature, the nature of the rehydration solution and role of monovalent cations on templated-replication of RNA using nucleoside 5'-monophosphates. Furthermore, we also carried out primer extension reactions using sugar-phosphate monomers, resulting in hybrid polymers. Preliminary mass analysis was carried out in collaboration with Dr. Yayoi Hongo, Earth Life Science Institute (ELSI), Tokyo, Japan. Our results add to the new body of literature that is shining light on the plausibility of having hybrid polymers, which are thought to have been important for the molecular evolution process that eventually would have led to the formation of RNA World from primitive informational polymers.

1.7 References

Adamala, K., and Szostak, J.W. (2013). Nonenzymatic Template-directed RNA Synthesis Inside Model Protocells. Science 342, 1098–1100.

Anastasi, C., Crowe, M.A., Powner, M.W., and Sutherland, J.D. (2006). Direct Assembly of Nucleoside Precursors from Two- and Three-Carbon Units. Angewandte Chemie International Edition 45, 6176–6179.

Attwater, J., Wochner, A., Pinheiro, V., Coulson, A., and Holliger, P. (2010). Ice as a Protocellular Medium for RNA Replication. Nature Communications 1, 76.

Attwater, J., Wochner, A., and Holliger, P. (2013). In-Ice Evolution of RNA Polymerase Ribozyme Activity. Nature Chemistry 5, 1011–1018.

Bada, J.L., and Lazcano, A. (2002). Some Like It Hot, But Not the First Biomolecules. Science 296, 1982 – 1983.

Bada, J.L., and Lazcano, A. (2003). Prebiotic Soup--Revisiting the Miller Experiment. Science 300, 745 – 746.

Bartel, D.P., and Szostak, J.W. (1993). From Prelife to Life: How Chemical Kinetics Become Evolutionary Dynamics. Science 261, 1411 – 1418.

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 – 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. Nature Communications 9, 163.

Bonfio, C., Valer, L., Scintilla, S., Shah, S., Evans, D.J., Jin, L., Szostak, J.W., Sasselov, D.D., Sutherland, J.D., and Mansy, S.S. (2017). UV-light-Driven Prebiotic Synthesis of Iron–Sulfur Clusters. Nature Chemistry 9, 1229–1234.

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. Nature Communications 7, 11328.

Callahan, M.P., Smith, K.E., Cleaves, H.J., Ruzicka, J., Stern, J.C., Glavin, D.P., House, C.H., and Dworkin, J.P. (2011). Carbonaceous Meteorites Contain a Wide Range of Extraterrestrial Nucleobases. Proceedings of the National Academy of Sciences 108, 13995–13998.

Chen, I., and Nowak, M. (2012). From Prelife to Life: How Chemical Kinetics Become Evolutionary Dynamics. Accounts of Chemical Research 45, 2088–2096.

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. Journal of the American Chemical Society 136, 5640–5646.

Cody, G.D., Boctor, N.Z., Filley, T.R., Hazen, R.M., Scott, J.H., Sharma, A., and Yoder, H.S. (2000). Activated Acetic Acid by Carbon Fixation on (Fe,Ni)S Under Primordial Conditions. Science 289, 1337 – 1340.

Costanzo, G., Pino, S., Ciciriello, F., and Mauro, E. Di (2009). Generation of Long RNA Chains in Water . The Journal of Biological Chemistry 284, 33206–33216.

Costanzo, G., Pino, S., Timperio, A.M., Šponer, J.E., Šponer, J., Nováková, O., Šedo, O., Zdráhal, Z., and Di Mauro, E. (2016). Non-Enzymatic Oligomerization of 3', 5' Cyclic AMP. PLOS ONE 11, e0165723.

Deamer, D. (1985). Boundry Structures are Formed by Organic Components of the Murchison Carbonaceous Chondrite. Nature 317, 792–794.

Deck, C., Jauker, M., and Richert, C. (2011). Efficient Enzyme-free Copying of all Four Nucleobases Templated by Immobilized RNA. Nature Chemistry 3, 603–608.

Djokic, T., Van Kranendonk, M.J., Campbell, K.A., Walter, M.R., and Ward, C.R. (2017). Earliest Signs of Life on Land Preserved in ca. 3.5 Ga Hot Spring Deposits. Nature Communications 8, 15263.

Dodd, M.S., Papineau, D., Grenne, T., Slack, J.F., Rittner, M., Pirajno, F., O'Neil, J., and Little, C.T.S. (2017). Evidence for Early Life in Earth's Oldest Hydrothermal Vent Precipitates. Nature 543, 60–64.

Eigen, M. (1971). Sel-forganization of Matter and the Evolution of Biological Macromolecules. Die Naturwissenschaften 58, 465–523.

Ekland, E.H., and Bartel, D.P. (1996). RNA-catalysed RNA Polymerization using Nucleoside Triphosphates. RNA 382, 373–376.

Engelhart, A.E., Powner, M.W., and Szostak, J. (2013). Functional RNAs Exhibit Tolerance for Non-heritable 2'-5' versus 3'-5' backbone heterogeneity. Nature Chemistry 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, 6(2), 15.

Ferris, J.P. (2006). Montmorillonite-catalysed Formation of RNA Oligomers: the Possible Role of Catalysis in the Origins of life. Philosophical Transactions of the Royal Society B: Biological Sciences 361, 1777–1786.

Ferris, J.P., and Ertem, G. (1992). Oligomerization Reactions of Ribonucleotides: The reaction of the 5'-phosphorimidazolide of Nucleosides on Montmorillonite and Other Minerals. Origins of Life and Evolution of the Biosphere 22, 369–381.

Ferris, J.P., and Ertem, G. (1993). Montmorillonite Catalysis of RNA Oligomer Formation in Aqueous Solution. A Model for the Prebiotic Formation of RNA. Journal of the American Chemical Society 115, 12270–12275.

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. Origins of Life and Evolution of the Biosphere 19, 165–178.

Ferus, M., Pietrucci, F., Saitta, A.M., Knížek, A., Kubelík, P., Ivanek, O., Shestivska, V., and Civiš, S. (2017). Formation of Nucleobases in a Miller–Urey Reducing Atmosphere. Proceedings of the National Academy of Sciences 114, 4306–4311.

Forterre, P. (2005). The Two Ages of the RNA World and the Transition to the DNA World: A Story of Viruses and Cells. Biochimie 87, 793–803.

Fuller, W.D., Sanchez, R.A., and Orgel, L.E. (1972). Studies in Prebiotic Synthesis: VI. Synthesis of Purine Nucleosides. Journal of Molecular Biology 67, 25–33.

Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K., and Krishnamurthy, R. (2017). Phosphorylation, Oligomerization and Self-assembly in Water Under Potential Prebiotic Conditions. Nature Chemistry 10, 212–218.

Gilbert, W. (1986). The RNA World. Nature 319, 618.

Giurgiu, C., Li, L., O'Flaherty, D.K., Tam, C.P., and Szostak, J.W. (2017). A Mechanistic Explanation for the Regioselectivity of Nonenzymatic RNA Primer Extension. Journal of the American Chemical Society 139, 16741–16747.

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.

Gull, M., Mojica, M.A., Fernández, F.M., Gaul, D.A., Orlando, T.M., Liotta, C.L., and Pasek, M.A. (2015). Nucleoside Phosphorylation by the Mineral Schreibersite. Scientific Reports 5, 17198.

Hagenbuch, P., Kervio, E., Hochgesand, A., Plutowski, U., and Richert, C. (2005). Chemical Primer Extension: Efficiently Determining Single Nucleotides in DNA. Angewandte Chemie International Edition 44, 6588–6592.

Harish, A., and Caetano-Anollés, G. (2012). Ribosomal History Reveals Origins of Modern Protein Synthesis. PLOS ONE 7, e32776.

Hazen, R.M. (2013). Paleomineralogy of the Hadean Eon : A Preliminary Species List. American Journal of Science 313, 807–843.

Heuberger, B.D., Pal, A., Del Frate, F., Topkar, V. V, and Szostak, J.W. (2015). Replacing Uridine with 2-Thiouridine Enhances the Rate and Fidelity of Nonenzymatic RNA Primer Extension. Journal of the American Chemical Society 137, 2769–2775.

Himbert, S., Chapman, M., Deamer, D.W., and Rheinstädter, M.C. (2016). Organization of Nucleotides in Different Environments and the Formation of Pre-Polymers. Scientific Reports 6, 31285.

Horning, D.P., and Joyce, G.F. (2016). Amplification of RNA by an RNA Polymerase Ribozyme. Proceedings of the National Academy of Sciences 113, 9786 – 9791.

Huang, W., and Ferris, J.P. (2003). Synthesis of 35–40 mers of RNA Oligomers from Unblocked Monomers: A Simple Approach to the RNA World. Chemical Communications 1458–1459.

Huang, W., and Ferris, J.P. (2006). One-Step , Regioselective Synthesis of up to 50-mers of RNA Oligomers by Montmorillonite Catalysis. Journal of the American Chemical Society 128, 8914–8919.

Huber, C., and Wächtershäuser, G. (1997). Activated Acetic Acid by Carbon Fixation on (Fe,Ni)S Under Primordial Conditions. Science 276, 245 – 247.

Hud, N. V, Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The Origin of RNA and “My Grandfather ’s Axe ”. Chemistry & Biology 20, 466–474.

Ikehara, K. (2005). Possible Steps to the Emergence of Life: The [GADV]-Protein World Hypothesis. The Chemical Record 5, 107–118.

Inoue, T., and Orgel, L.E. (1981). Substituent Control of the Poly(C)-Directed Oligomerization of Guanosine 5’-Phosphoroimidazolide. Journal of the American Chemical Society 103, 7666–7667.

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.

Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. (2001). RNA-Catalyzed RNA Polymerization : Accurate and General RNA-Templated Primer Extension. Science 292, 1319–1325.

Joyce, G.F., and Orgel, L.E. (1986). Non-enzymic Template-directed Synthesis on RNA Random Copolymers: Poly(C, G) Templates. Journal of Molecular Biology 188, 433–441.

Kaiser, A., Spies, S., Lommel, T., and Richert, C. (2012). Template-Directed Synthesis in 3'- and 5'-Direction with Reversible Termination. Angewandte. Angewandte Chemie (International Ed. in English) 51, 8299–8303.

Kanavarioti, A., Monnard, P.-A., and Deamer, D. (2001). Eutectic Phase in Ice Facilitates Nonenzymatic Nucleic Acid Synthesis. Astrobiology 1, 271–281.

Kim, D.E., and Joyce, G.F. (2004). Cross-Catalytic Replication of an RNA Ligase Ribozyme. Chemistry & Biology 11, 1505–1512.

Kim, H.J., and Benner, S.A. (2015). Prebiotic Glycosylation of Uracil with Electron-Donating Substituents. Astrobiology 15, 301–306.

Kim, H.-J., and Benner, S.A. (2017). Prebiotic Stereoselective Synthesis of Purine and Noncanonical Pyrimidine Nucleotide from Nucleobases and Phosphorylated Carbohydrates. Proceedings of the National Academy of Sciences 114, 11315 – 11320.

Kim, H.-J., Furukawa, Y., Kakegawa, T., Bita, A., Scorei, R., and Benner, S.A. (2016). Evaporite Borate-Containing Mineral Ensembles Make Phosphate Available and Regiospecifically Phosphorylate Ribonucleosides: Borate as a Multifaceted Problem Solver in Prebiotic Chemistry. Angewandte Chemie International Edition 55, 15816–15820.

Kitadai, N., and Maruyama, S. (2018). Origins of Building Blocks of Life: A Review. Geoscience Frontiers 9, 1117–1153.

Kitadai, N., Kameya, M., and Fujishima, K. (2017). Origin of the Reductive Tricarboxylic Acid (rTCA) Cycle-Type CO2 Fixation: A Perspective. Life 7, 7(4), 39.

Kozlov, I.A., and Orgel, L.E. (2000). Nonenzymatic Template-directed Synthesis of RNA from Monomers. Molecular Biology 34, 781–789.

Krishnamurthy, R. (2017). Giving Rise to Life: Transition from Prebiotic Chemistry to Protobiology. Accounts of Chemical Research 50, 455–459.

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.

Larsen, A.T., Fahrenbach, A.C., Sheng, J., Pian, J., and Szostak, J.W. (2015). Thermodynamic Insights into 2-thiouridine-enhanced RNA Hybridization. Nucleic Acids Research 43, 7675–7687.

Leu, K., Obermayer, B., Rajamani, S., Gerland, U., and Chen, I.A. (2011). The Prebiotic Evolutionary Advantage of Transferring Genetic Information From RNA To DNA. Nucleic Acids Research 39, 8135–8147.

Leu, K., Kervio, E., Obermayer, B., Turk-macleod, R.M., Yuan, C., Luevano, J., Chen, E., Gerland, U., Richert, C., and Chen, I.A. (2013). Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. Journal of the American Chemical Society 135, 354–366.

Li, L., Prywes, N., Tam, C.P., O'Flaherty, D., Lelyveld, V., Izgu, E.C., Pal, A., and Szostak, J. (2017). Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. Journal of the American Chemical Society 139, 1810–1813.

Lincoln, T.A., and Joyce, G.F. (2009). Self-Sustained Replication of an RNA Enzyme. Science 323, 1229–1232.

Löffler, P.M.G., Groen, J., Dörr, M., and Monnard, P.-A. (2013). Sliding Over the Blocks in Enzyme-Free RNA Copying – One-Pot Primer Extension in Ice. PLOS ONE 8, e75617.

Lohrmann, R., Bridson, P., and Orgel, L. (1980). Efficient Metal-Ion Catalyzed Template-Directed Oligonucleotide Synthesis. Science 208, 1464–1465.

Mansy, S., Schrum, J., Krishnamurthy, M., Tobe, S., Treco, D., and Szostak, J. (2008). Template-directed Synthesis of a Genetic Polymer in a Model Protocell. Nature 454, 122–125.

Martin, L.L., Unrau, P.J., and Müller, U.F. (2015). RNA Synthesis by *in vitro* Selected Ribozymes for Recreating an RNA World. Life 5, 5(1), 247–268.

Martin, W., Baross, J., Kelley, D., and Russell, M.J. (2008). Hydrothermal Vents and the Origin of Life. Nature Reviews Microbiology 6, 805–814.

McCollom, T.M., Ritter, G., and Simoneit, B.R.T. (1999). Lipid Synthesis Under Hydrothermal Conditions by Fischer- Tropsch-Type Reactions. Origins of Life and Evolution of the Biosphere 29, 153–166.

Meinert, C., Myrgorodska, I., De Marcellus, P., Buhse, T., Nahon, L., Hoffmann, S. V, D'Hendecourt, L.L.S., and Meierhenrich, U.J. (2016). Ribose and Related Sugars from Ultraviolet Irradiation of Interstellar Ice Analogs. Science 352, 208 – 212.

Miyakawa, S., Cleaves, H.J., and Miller, S.L. (2002). The Cold Origin of Life: B. Implications Based on Pyrimidines and Purines Produced From Frozen Ammonium Cyanide Solutions. Origins of Life and Evolution of the Biosphere 32, 209–218.

Monnard, P., and Deamer, D. (2003). Preparation of Vesicles from Nonphospholipid Amphiphiles. Methods in Enzymology 372, 133–151.

Monnard, P.-A., Apel, C.L., Kanavarioti, A., and Deamer, D.W. (2002). Influence of Ionic Inorganic Solutes on Self-Assembly and Polymerization Processes Related to Early Forms of Life: Implications for a Prebiotic Aqueous Medium. Astrobiology 2, 139–152.

Monnard, P.-A., Kanavarioti, A., and Deamer, D.W. (2003). Eutectic Phase Polymerization of Activated Ribonucleotide Mixtures Yields Quasi-Equimolar Incorporation of Purine and Pyrimidine Nucleobases. Journal of the American Chemical Society 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. Nature Ecology & Evolution 1, 1716–1721.

Müller, U.F., and Bartel, D.P. (2008). Improved Polymerase Ribozyme Efficiency on Hydrophobic Assemblies. RNA 14, 552–562.

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.

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. Physical Chemistry Chemical Physics 18, 20144–20152.

Mutschler, H., Wochner, A., and Holliger, P. (2015). Freeze-thaw Cycles as Drivers of Complex Ribozyme Assembly. Nature Chemistry 7, 502–508.

Ninio, J., and Orgel, L.E. (1978). Heteropolynucleotides as Templates for Non-enzymatic Polymerizations. Journal of Molecular Evolution 12, 91–99.

Nissen, P., Hansen, J., Ban, N., Moore, P., and Steitz, T. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. Science 289, 920–930.

O'Flaherty, D.K., Kamat, N.P., Mirza, F.N., Li, L., Prywes, N., and Szostak, J.W. (2018). Copying of Mixed-Sequence RNA Templates inside Model Protocells. Journal of the American Chemical Society 140, 5171–5178.

Olasagasti, F., Kim, H.J., Pourmand, N., and Deamer, D. (2011). Non-enzymatic Transfer of Sequence Information under Plausible Prebiotic Conditions. Biochimie 93, 556–561.

Orgel, L.E. (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 39, 99–123.

Orgel, L.E. (2008). The Implausibility of Metabolic Cycles on the Prebiotic Earth. PLOS Biology 6, e18.

Oró, J., and Kimball, A.P. (1961). Synthesis of Purines under Possible Primitive Earth Conditions. I. Adenine from hydrogen cyanide. Archives of Biochemistry and Biophysics 94, 217–227.

Pasek, M.A., Gull, M., and Herschy, B. (2017). Phosphorylation on the Early Earth. Chemical Geology 475, 149–170.

Paul, N., and Joyce, G.F. (2002). A self-replicating Ligase Ribozyme. Proceedings of the National Academy of Sciences 99, 12733–12740.

Petrov, A.S., Gulen, B., Norris, A.M., Kovacs, N.A., Bernier, C.R., Lanier, K.A., Fox, G.E., Harvey, S.C., Wartell, R.M., Hud, N. V, et al. (2015). History of the Ribosome and the Origin of Translation. Proceedings of the National Academy of Sciences 112, 15396 – 15401.

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. Journal of the American Chemical Society 132, 16677–16688.

Rajamani, S., Vlassov, A., and Benner, S. (2007). Lipid-assisted Synthesis of RNA-like Polymers from Mononucleotides. Orig Life Evol Biosph.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., and Chen, I.A. (2010). Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. Journal of the American Chemical Society 132, 1008–1011.

Robertson, M.P., and Joyce, G.F. (2010). The Origins of the RNA World. Cold Spring Harbor Perspective in Biology 4, a003608.

Rogers, J., and Joyce, G.F. (2001). The Effect of Cytidine on the Structure and Function of an RNA Ligase Ribozyme. RNA 7, 395–404.

Rohatgi, R., Bartel, D.P., and Szostak, J.W. (1996). Nonenzymatic , Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3'-5' Phosphodiester Bonds. Journal of the American Chemical Society 118, 3340–3344.

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. Journal of Molecular Evolution 39, 231–243.

Russell, M.J., Daniel, R.M., Hall, A.J., and Sherringham, J.A. (2007). On the Origin of Biochemistry at an Alkaline Hydrothermal Vent. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 1887 – 1926.

Saladino, R., Crestini, C., Costanzo, G., Negri, R., and Di Mauro, E. (2001). A Possible Prebiotic Synthesis of Purine, Adenine, Cytosine, and 4(3H)-Pyrimidinone from Formamide: Implications for the Origin of Life. Bioorganic & Medicinal Chemistry 9, 1249–1253.

Saladino, R., Carota, E., Botta, G., Kapralov, M., Timoshenko, G.N., Rozanov, A.Y., Krasavin, E., and Di Mauro, E. (2015). Meteorite-catalyzed Syntheses of Nucleosides and of Other Prebiotic Compounds from Formamide under Proton Irradiation. Proceedings of the National Academy of Sciences 112, E2746–E2755.

Samanta, B., and Joyce, G.F. (2017). A Reverse Transcriptase Ribozyme. eLife 6, e31153.

Schrum, J.P., Ricardo, A., Krishnamurthy, M., Blain, J.C., and Szostak, J.W. (2009). Efficient and Rapid Template-Directed Nucleic Acid Copying Using 2'-Amino-2',3'-dideoxyribonucleoside-5'-Phosphorimidazolide Monomers. Journal of the American Chemical Society 131, 14560–14570.

Sczepanski, J.T., and Joyce, G.F. (2014). A Cross-chiral RNA Polymerase Ribozyme. Nature 515, 440–442.

Segre, D., Ben-Eli, D., Deamer, D.W., and Doron, L. (1999). The Lipid World. Orig Life Evol Biosph 31, 119–145.

Sheng, J., Larsen, A., Heuberger, B.D., Blain, J.C., and Szostak, J.W. (2014). Crystal Structure Studies of RNA Duplexes Containing s^2U:A and s^2U:U Base Pairs. Journal of the American Chemical Society 136, 13916–13924.

Stütz, A.R., Kervio, E., Deck, C., and Richert, C. (2007). Chemical Primer Extension : Individual Steps of Spontaneous Replication. Chemistry & Biodiversity 4, 784–802.

Sulston, J., Lohrmann, R., Orgel, L.E., and Miles, H.T. (1968). Nonenzymatic Synthesis of Oligoadenylates on a Polyuridylic Acid Template. Proceedings of the National Academy of Sciences of the United States of America 59, 726–733.

Sulston, J., Lohrmann, R., Orgel, L.E., Weimann, B.J., and Miles, H.T. (1969). Non-enzymatic Oligonucleotide Synthesis on a Polycytidylate Template. Journal of Molecular Biology 40, 227–234.

Sutherland, J.D. (2010). Ribonucleotides. Cold Spring Harb Perspect Biol a005439.

Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing Life. Nature 409, 387–390.

Tam, C.P., Fahrenbach, A.C., Björkbom, A., Prywes, N., Izgu, E.C., and Szostak, J.W. (2017). Downstream Oligonucleotides Strongly Enhance the Affinity of GMP to RNA Primer–Template Complexes. Journal of the American Chemical Society 139, 571–574.

Tashiro, T., Ishida, A., Hori, M., Igisu, M., Koike, M., Méjean, P., Takahata, N., Sano, Y., and Komiya, T. (2017). Early Trace of Life from 3.95 Ga Sedimentary Rocks in Labrador, Canada. Nature 549, 516–518.

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.

Trefil, J., Morowitz, H.J., and Smith, E. (2009). The Origin of Life: A Case is Made for the Descent of Electrons. American Scientist 97, 206–213.

Vaidya, N., Manapat, M.L., Chen, I.A., Xulvi-Brunet, R., Hayden, E.J., and Lehman, N. (2012). Spontaneous Network Formation among Cooperative RNA Replicators. Nature 491, 72–77.

Verlander, M.S., Lohrmann, R., and Orgel, L.E. (1973). Catalysts for the Self-polymerization of Adenosine Cyclic 2',3'-phosphate. Journal of Molecular Evolution 2, 303–316.

Vogel, S.R., Deck, C., and Richert, C. (2005). Accelerating Chemical Replication Steps of RNA Involving Activated Ribonucleotides and Downstream-Binding Elements. Chemical Communications 4922–4924.

Voytek, S.B., and Joyce, G.F. (2007). Emergence of a Fast-reacting Ribozyme that is Capable of Undergoing Continuous Evolution. Proceedings of the National Academy of Sciences 104, 15288– 15293.

Wachowius, F., Attwater, J., and Holliger, P. (2017). Nucleic acids: Function and Potential for Abiogenesis. Quarterly Reviews of Biophysics 50, E4.

Wächtershäuser, G. (1988). Before Enzymes and Templates: Theory of Surface Metabolism. Microbiological Reviews 52, 452–484.

Walton, T., and Szostak, J.W. (2016). A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension. Journal of the American Chemical Society 138, 11996–12002.

Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011). Ribozyme-catalyzed Transcription of an Active Ribozyme. Science 332, 209–212.

Wu, T., and Orgel, L.E. (1992a). Nonenzymic Template-directed Synthesis on Oligodeoxycytidylate Sequences in Hairpin Oligonucleotides. Journal of the American Chemical Society 114, 317–322.

Wu, T., and Orgel, L.E. (1992b). Nonenzymic Template-directed Synthesis on Hairpin Oligonucleotides. 2. Templates Containing Cytidine and Guanosine Residues. Journal of the American Chemical Society 114, 5496–5501.

Wu, T., and Orgel, L.E. (1992c). Nonenzymatic Template-directed Synthesis on Hairpin Oligonucleotides. 3. Incorporation of Adenosine and Uridine Residues. Journal of the American Chemical Society 114, 7963–7969.

Yi, R., Hongo, Y., and Fahrenbach, A.C. (2018). Synthesis of Imidazole-activated Ribonucleotides using Cyanogen Chloride. Chemical Communications 54, 511–514.

Zaher, H.S., and Unrau, P.J. (2007). Selection of an Improved RNA Polymerase Ribozyme with Superior Extension and Fidelity. RNA 13, 1017–1026.

Zhang, S., Zhang, N., Blain, J.C., and Szostak, J. (2013a). Synthesis of N3’-P5'-linked Phosphoramidate DNA by Nonenzymatic Template-Directed Primer Extension. Journal of the American Chemical Society 135, 924–932.

Zhang, S., Blain, J.C., Zielinska, D., Gryaznov, S., and Szostak, J. (2013b). Fast and Accurate Nonenzymatic Copying of an RNA-Like Synthetic Genetic Polymer. Proceedings of the National Academy of Sciences of the United States of America 110, 17732–17737.

Zhang, W., Tam, C.P., Walton, T., Fahrenbach, A.C., Birrane, G., and Szostak, J.W. (2017). Insight into the Mechanism of Nonenzymatic RNA Primer Extension from the Structure of an RNA-GpppG Complex. Proceedings of the National Academy of Sciences 114, 7659–7664.

Zhang, W., Tam, C.P., Zhou, L., Oh, S.S., Wang, J., and Szostak, J.W. (2018). Structural Rationale for the Enhanced Catalysis of Nonenzymatic RNA Primer Extension by a Downstream Oligonucleotide. Journal of the American Chemical Society 140, 2829–2840.

Zielinski, W.S., and Orgel, L.E. (1985). Oligomerization of Activated derivatives of 3'-amino-3'-Deoxyguanosine on Poly(C) and Poly(dC) Templates. Nucleic Acids Research 13, 2469–2484.

Zielinski, W.S., and Orgel, L.E. (1987). Oligoaminonucleoside Phosphoramidates. Oligomerization of Dimers of 3'-amino-3'-deoxy-nucleotides (GC and CG) in Aqueous Solution. Nucleic Acids Research 15, 1699–1715.

Chapter 2

Effect of co-solutes on template-directed nonenzymatic copying of RNA

2.1 Introduction

The assembly of a replicase-like enzyme in the context of 'RNA World' poses considerable challenge. This is because it implies the construction of a long enough functional RNA molecule from monomers, which needs to also faithfully replicate its own information to maintain heredity. As elaborated in section 1.4 in Chapter1, nonenzymatic polymerization and replication, which can lead to formation of such molecule/s, is an error prone process, which is further compounded by the lack of correction machinery. Initial studies aimed at understanding enzyme-free template-directed primer extensions, have characterized the rate and accuracy of such replication processes (Kervio et al., 2010; Rajamani et al., 2010; Leu et al., 2011, 2013). On a positive note, it appears that faithfully replicating sequences do have an advantage even in the absence of an elaborate enzyme-based error correction system, as the mismatches that occur during enzyme-free replication tend to stall their own course (Rajamani et al.,2010; Leu et al., 2013).

However, all of the above mentioned reactions were carried out using highly pure starting materials and homogeneous, buffered reaction mixtures. To state this in another way, most if not all the nonenzymatic replication reactions studied in the context of 'RNA World' thus far, have been carried out in 'pure systems' devoid of any co-solutes or background molecules. This scenario is not very likely to be prebiotically plausible as the 'prebiotic soup' would have been a mixture of many different molecular species rather than being a homogeneous solution containing only certain kinds of molecules. These different molecular species could result from various prebiotic syntheses pathways, and it is known that the outcome of almost all such reactions is a mixture of products. For e.g., the formose synthesis reaction, which results in a mixture of 4 to 6 carbon containing sugar molecules, never exclusively yields the ribose sugar (Zweckmair et al., 2014). Along with D-ribose, which is a part of the RNA molecule, this reaction also yields a plethora of other sugars such as erythrose, xylose, arabinose, glucose, mannose etc (Fig 2.1A). The Fischer-Tropsch synthesis pathway, which is considered as the prebiotically viable means of synthesis of longer carbon chain compounds from carbon monoxide and hydrogen, is also known to yield a wide range of products. For example, the GC-MS analysis depicted in Fig 2.1B shows the range of products obtained from just heating oxalic acid in a

sealed steel tube at 174 ˚C (McCollom et al., 1999). Similarly, only a fraction of product from Oró's synthesis is the canonical adenine nucleobase (Orgel, 2004). On the other hand, the meteorites that would have fallen on the prebiotic Earth, during the 'Heavy bombardment period' must have also carried a mixture of compounds. All the aforementioned examples clearly indicate that the early Earth would have been replete with different kinds of molecules that would have been present together in the prebiotic soup. Hence, any prebiotically pertinent reaction might not have taken place in a clean buffered condition unless it was completely isolated by some means.

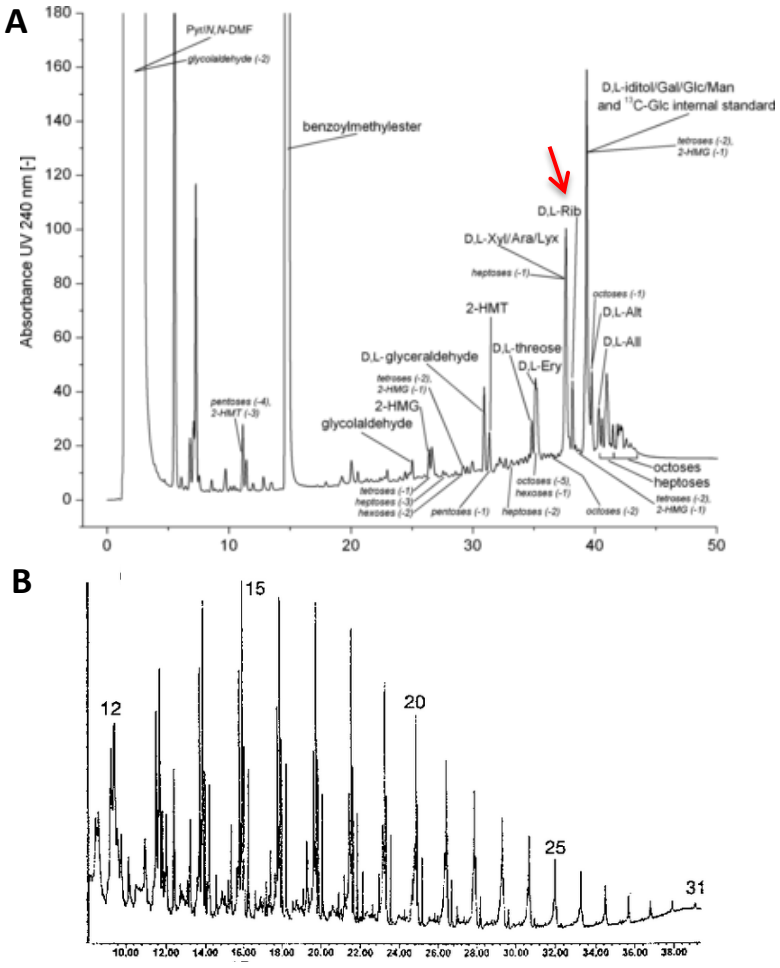


Figure 2.1: Prebiotic syntheses yield a mixture of products (A) LC-UV chromatogram for the products obtained from formose reaction initiated with glycoladehyde. The red arrow indicates the peak for ribose (reproduced from Zweckmair et al., 2013) (B) GC-MS analysis of the total chloroform extract obtained from FTT synthesis obtained from aqueous solutions of oxalic acid (reproduced from McCollam et al., 1999).

The existence of these molecular entities (co-solutes (or) background molecules) would have resulted in a complex prebiotic milieu, which in turn might have given rise to certain inherent properties like molecular crowding. Presence of co-solutes, and the resultant phenomenon of molecular crowding, are known to affect both the kinetics and equilibrium of several biochemical reactions (Ellis, 2001). Furthermore, macromolecular crowding has been shown to affect not only protein folding, stability and the rates of the enzymatic reactions that they catalyze (Minton, 2001; Zhou et al., 2008), but also impact reactions involving nucleic acids (Nakano et al., 2014). A few studies have also shown the effect of molecular crowding on prebiotically relevant molecules like ribozymes. For example, catalytic RNAs are known to get stabilized under crowded conditions (Kilburn et al., 2013; Desai et al., 2014). In addition, it has also been reported that the activity of ribozymes increase in the presence of co-solutes (Nakano et al., 2009; Kilburn et al., 2013; Strulson et al., 2013; Desai et al., 2014). Increase in the rate of self-cleavage (by up to 70-fold) was observed for hammerhead ribozyme in an aqueous-two phase system comprising of polyetheyelne glycol (PEG) and dextran, which is usually used to mimic intracellular compartmentalization and macromolecular crowding (Strulson et al., 2012). Crowding has also been shown to improve the catalytic activity of ribozymes at physiologically relevant ionic strengths (Strulson et al., 2013), which has implications for stability of protocells under such circumstances (Chen et al., 2005; Adamala and Szostak, 2013).

Despite its implicit relevance, the heterogeneity of the prebiotic soup is usually not considered when discerning prebiotic processes like the nonenzymatic formation of catalytic RNA molecules. Hence, characterizing enzyme-free copying reactions in prebiotically relevant scenarios becomes important as they might not advent in a chemically complex environment in the same manner as they would in a chemically simpler set-up. Scenarios where the reaction rates might be affected pronouncedly due to the presence of co-solutes would have had a bearing on the efficiency of successive transfer of genetic information. This is fundamental for the persistence and evolution of catalytic RNAs in an RNA World. Therefore, in this chapter, we have studied the effect of presence of co-solutes on nonenzymatic template-directed RNA primer extension reactions. Poly Ethylene Glycol (PEG) and a double chain surfactant lipid were

used as co-solutes in these reactions. In the presence of these co-solutes, both in isolation and in conjunction, the rate and fidelity of these enzyme-free copying reactions were characterized. Additionally, computer simulation studies were also carried out to analyze the potential effect on multiple consecutive replication cycles.

2.2 Materials and Methods

2.2.1 Chemicals

Analytical grade Trizma base (Tris- hydroxymethyl- aminomethane), sodium chloride, polyethylene glycol (PEG) 8kDa, boric acid and ethylenediaminetetraacetic acid (EDTA) were purchased from Sigma-Aldrich. The double chain surfactant lipid used in the reactions, namely 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC), was purchased from Avanti Polar Lipids Inc., Alabama, USA. Adenosine-5'-phosphoimidazolide (ImpA), guanosine- 5'-phosphoimidazolide (ImpG), cytidine- 5'-phosphoimidazolide (ImpC) and uridine- 5'-phosphoimidazolide (ImpU) were purchased from GLSynthesis Inc. , MA, USA and used without further purification.

A 20-mer length RNA primer and four different 29-mer length RNA templates were used in the reactions, the sequences for which are mentioned below. The templating base is highlighted in bold:

Name	Sequence (5'-> 3')
Primer Amino G	GG GAU UAA UAC GAC UCA CUG-NH_2
Template MisInc_C	AGU GAU CU**C** CAG UGA GUC GUA UUA AUC CC
Template MisInc_G	AGU GAU CU**G** CAG UGA GUC GUA UUA AUC CC
Template MisInc_A	AGU GAU CU**A** CAG UGA GUC GUA UUA AUC CC
Template MisInc_U	AGU GAU CU**U** CAG UGA GUC GUA UUA AUC CC

The RNA primer (Primer Amino G), which was acquired from Keck laboratory, Yale, USA, is terminated with 3'-amino-2', 3'-dideoxynucleotide (Metkinen, Finland). This modification was

done to ensure effective addition of the incoming nucleotide onto the primer, as amino group is known to be a better nucleophile than the canonical hydroxyl group. The primer was labelled with Cy3 fluorescent tag on the 5' end for enabling its detection upon polyacrylamide gel electrophoresis (PAGE) analysis. The acquired primer was first deprotected using triethylamine trihydrofluoride (TEA.3HF) to remove the 2'-TBDMS protection groups. The deprotecetd RNA primer was then gel purified prior to using it in the reactions. The RNA templates were acquired from Thermo Scientific (Dharmacon). The 2'-ACE protection was removed using deprotection buffer (100mM acetic acid adjusted to pH 3.8 using TEMED) following the manufacturer's protocol. The templates were then used without further purification.

2.2.2 Methods

2.2.2.1 Setting-up the enzyme-free primer extension reactions

DLPC vesicles were prepared by first drying the required volume of the 25mg/ml chloroform stock in a clean glass vial. The dried film was then re-suspended in nanopure water. It was followed by vortexing to get the lipid stock of desired concentration. This lipid stock was extruded through a 100 nm membrane using a mini extruder (Avanti Polar lipids Inc., USA), to get small unilamellar vesicles. These uniform-sized vesicles were then used as co-solute in the reactions. For PEG, which was the second co-solute, a 50% w/v working stock was prepared by dissolving required amount of PEG in nanopure water. The RNA stock solutions were made in 10mM Tris buffer (pH 7.0) and their concentrations were estimated by UV spectroscopy (UV-1800 UV/Vis spectrophotometer, Shimadzu Corp., Japan).

In a typical reaction, 0.325µM of primer and 1.3µM of one of the templates were annealed to each other in RNase free water (Sigma), by heating at 95˚C for 5 minutes, followed by cooling at room temperature (RT) for another 5 minutes. Appropriate volumes of 1M Tris (pH 7.0) and 4M NaCl were added to this annealed primer-template (P-T) complex to a final concentration of 100mM and 200mM, respectively. The co-solutes, viz., DLPC and PEG 8000, were added in the reactions to a final concentration of 5mM and 18% w/v, respectively. The primer extension

reactions were initiated by addition of, typically, 10 mM final concentration of Imp A/G/C or 40 mM final concentration of Imp U (Rajamani et al., 2010; Leu et al., 2011).

2.2.2.2 Analysis of primer extension reactions

The time points for the primer extension reactions (including a zero minute 'control' time point, which was collected before the addition of activated monomer) were collected in Tris-Borate-EDTA (TBE) buffer (10.8 gm Trizma base, 5.5 gm boric acid, 4ml of 0.5M EDTA in 1000 ml of distilled water), containing 8 M urea and additional 100 mM EDTA (gel loading buffer). Typically, 1µl of the reaction time point was collected in 19µl of gel loading buffer. Competitor RNA, with a sequence similar to that of Primer Amino G (5' GGG AUU AAU ACG ACU CAC UG), was added in up to 10-20 times excess to the reaction time points. This was done to ensure complete separation of the fluorescently tagged primer from their respective templates. These time point samples were then heated at 95˚C for 5 minutes, cooled at RT for 5 minutes and briefly centrifuged, before loading onto 17cm X 15cm 20% denaturing PAGE gel containing 8M urea (SequaGel- UreaGel system, National diagnostics, USA). 1X TBE buffer was used as the gel running buffer. Bromophenol blue and xylene cyanol were used as tracker dyes to estimate the progress of the of the RNA primer bands during gel electrophoresis. The gels were scanned using Typhoon Trio plus imager (GE Healthcare) at 550 PMT and 200 micron resolution, to detect the presence of fluorescently labeled (excitation laser: 532nm) primer and the extended product. The scans were analyzed using ImageQuant v5.2 software (GE healthcare). The extension of the primer over time was calculated by dividing the intensity of unreacted primer by sum of the intensities of both extended and unreacted primer, for a particular time point. This was plotted on the 'Y' axis and the time (in minutes) was plotted on the 'X' axis. The initial rate of the reactions, wherein the reaction is estimated to follow pseudo-first order kinetics, was calculated by fitting a linear equation to several early data points.

2.2.2.3 Calculation of error rates

The frequency of incorporation ($f_{N,N}$) of a specific nucleotide (ImpN) across a particular template base was calculated by dividing the rate of addition of that nucleotide by sum of the rates of addition of all nucleotides across the same template base in the P-T complex. For example, incorporation frequency of ImpG across 'C' template base ($f_{C,G}$) can be calculated by using the equation ($f_{C,G}= r_{C,G}/(r_{C,G}+ r_{C,A}+ r_{C,U}+ r_{C,C})$). Mutation rate for a given template base (μ_N) was calculated by adding the frequencies of incorrect incorporations across that base. The average mutation rate was calculated by considering all mutation rates of all the template bases in a particular system ($\mu_{avg}= 0.25 *(\mu_A+ \mu_G+ \mu_C+ \mu_U)$). Equal contribution from A, G, C and U in the template was assumed for the calculation of average mutation rates. Mutation rates for genomes composed of equal parts of G+C or A+U were also calculated in a similar manner.

2.2.2.4 Computer simulation for template-directed polymerization

To analyze the effect of co-solutes on template-directed polymerization over multiple replication cycles, we simulated the process using the Gillespie algorithm. 'MisInc_C' sequence was selected as the template for these simulations and the monomers were incorporated in a stepwise fashion in the absence of any primer (similar to the depiction in Fig 1.6A, Chapter 1). No stalling factor or boundary conditions were invoked for these initial simulation trials. The simulation was started with 2000 (N) identical template sequences and the polymerization process was followed for t=500 hrs, with predetermined replication cycle times. The replication cycle time denotes the number of hours for which the *in silico* replication was allowed to continue without strand separation. All the polymers had an equal chance to form a reverse complementary strand by the stochastic addition of monomers, which depended upon experimentally calculated rate constants. If the *in silico* replication of a said sequence was complete before the stipulated cycle period, no more nucleotides were added. After a predetermined replication cycle period was complete, all template strands were separated from their complementary strands. The incomplete daughter strands were discarded and the full complements of the templates were retained (n). This resulted in a new pool of sequences containing a total of N+n strands, out of which N strands were chosen randomly for the next

cycle of replication. The two main observables recorded were (a) the total number of diverse sequences generated and (b) the GC content of the sequences for each replication cycle. This work was carried out in collaboration with Dr. Julien Derr at Paris Diderot University, France.

2.3 Results

2.3.1 Gel analysis of the extended primer product

The progress of the template-directed RNA primer extension reactions, under control and experimental (in presence of co-solutes) conditions, was analyzed using 20% denaturing polyacrylamide gel electrophoresis (containing 8M urea). Tracking dyes (mainly xylene cyanol) were used to standardize the approximate gel running time, so as to get a good separation between the fluorescently tagged primer (20-mer RNA) and the extended primer (21-mer). Adequate separation between the original and extended primer was essential to achieve, as the rate of the reaction was calculated by determining the fraction of the original primer remaining over time, using image analysis software. To achieve this, the standardization of the gel running protocol was carried out till there was a clear one nucleotide length resolution between the bands. Figure 2.2A shows a typical gel image depicting the progress of the RNA primer extension reaction, and the graph therein depicts the fraction of the original primer remaining over time. The apparent first-order rate of addition of all matched as well as mismatched nucleotides, against the four different template bases, were determined for the control conditions and in the presence of co-solutes (a total of 64 different reaction combinations). This was undertaken to characterize reaction rates of such nonenzymatic replicating systems in the presence of background molecules.

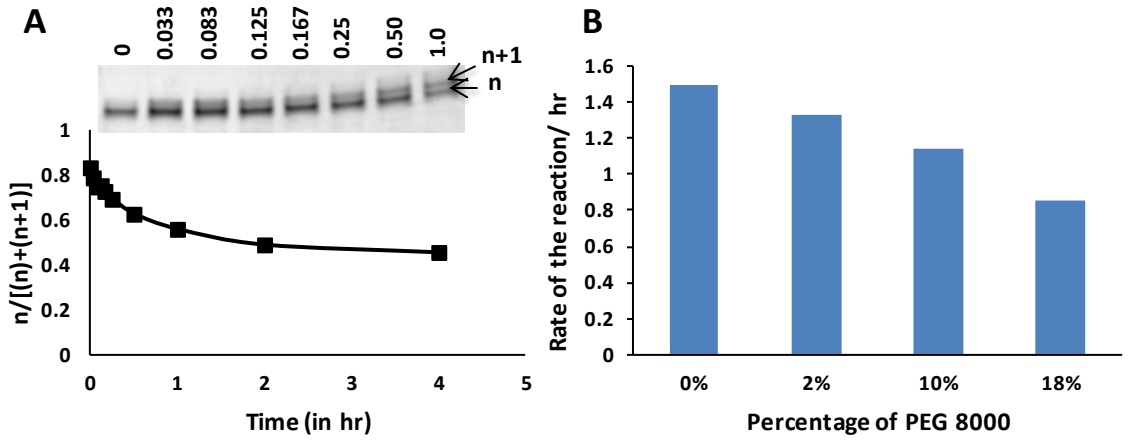


Figure 2.2: (A) Incorporation rate of cognate nucleotide 'Imp A' across the template base U of the 'MisInc_U' template, in the presence of both DLPC vesicles and PEG. The gel image shows extension of the primer over time (initial rate = 0.509/h) (B) Rate of incorporation of 'ImpG' across 'MisInc_C' template, in presence of different concentrations of PEG 8000.

2.3.2 Effect of individual co-solutes on the template-directed primer extension reactions

We have used a double chain surfactant lipid (DLPC) and PEG 8000 as co-solute molecules in our nonenzymatic replication system. To begin with, the effect of presence of the individual co-solute, on nonenzymatic replication, was studied. Firstly, the rate of template-directed addition of a 'correct (cognate)' nucleotide was estimated in individual reactions, i.e. the addition of ImpG/ImpC/ ImpA/ ImpU nucleotides, against C/G/U/ A template base, respectively.. The corresponding control reactions were carried out in buffered solution, without the addition of any background molecules (comparable to the nonenzymatic RNA replication reactions in Leu et. al, 2011). Interestingly, the rate of addition of 'G across C' (addition of ImpG across 'C' template base) was found to have reduced from 1.58 ± 0.17 per hour in the control condition, to 1.09 ± 0.1 per hour in the presence of DLPC ($p < 0.05$, Student's t test), and to 0.79 ± 0.26 per hour in presence of PEG ($p < 0.05$, Student's t test) (Table 2.1). The standard deviation was calculated from three independent reaction replicates. The decrease in the rate of addition of 'G across C' was further found to be dependent on the percentage of PEG polymer used in the starting reaction mixture (Fig 2.2.B). In the case of addition of 'A across U' (i.e. rate of addition

of ImpA across 'U' template base), the rate of the reaction did reduce in the presence of co-solutes when individual replicates were considered. However, the absolute rates varied between the three replicates, giving rise to the large standard deviation in all the three reaction sets (0.92 ± 0.37 per hour in the control condition, 0.73 ± 0.24 per hour in the presence of DLPC and 0.57 ± 0.33 per hour in the presence of PEG) (Table 2.1). Due to this large deviation from the mean value, the decrease in the rate of this cognate addition reaction in the presence of co-solutes was not found to be statistically significant using the Student's t test. In general, the rate of those cognate addition reactions wherein the incoming monomer was a purine nucleotide was found to have reduced in the presence of individual co-solutes (Fig 2.3). When the rate of the other two cognate addition reactions, i.e. ImpU across 'A' template base and ImpC across 'G' template base, was calculated in the absence and presence of co-solutes, no apparent difference was observed (Fig 2.3). Furthermore, the other 12 'mismatched' addition reactions were also carried out under aforementioned conditions. No significant effect on the rate of mismatched additions was observed in the presence of co-solutes, even for those reactions wherein the incoming non-cognate base was a purine nucleotide (Table 2.1).

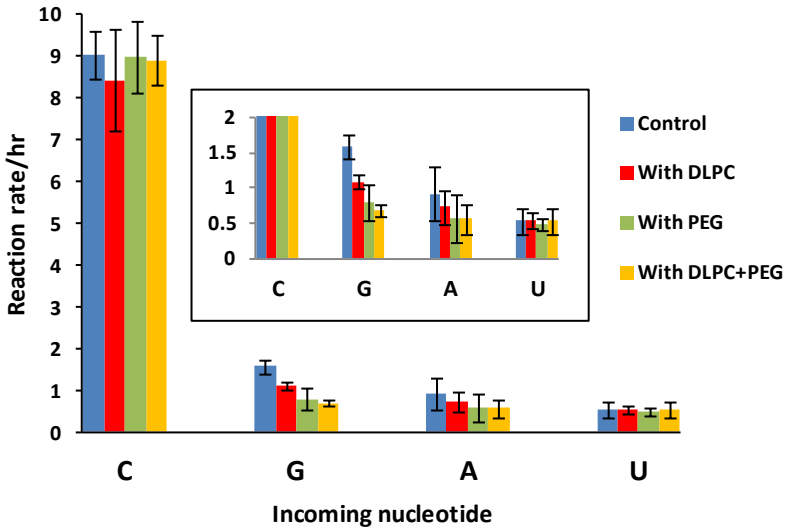


Figure 2.3: Rate of primer extension reactions for addition of a cognate nucleotide across the corresponding template base, in the absence or presence of co-solutes. Error bars indicate standard deviation calculated from experimental triplicates.

Table 2.1: Experimental rates for nonenzymatic primer extension reactions (hr^{-1}) under control reaction conditions and in the presence of either or both of the co-solutes. σ denotes standard deviation from experimental replicates.

Reaction No.			Control		With DLPC		With PEG8000		With DLPC and PEG8000	
	Template	ImpN	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}
1	C	G	1.58	0.17	1.09	0.1	0.79	0.26	0.69	0.08
2	C	U	0.021	0.002	0.022	0.004	0.02	0.002	0.022	0.005
3	C	C	0.022	0.002	0.014	0.008	0.022	0.004	0.028	0.0004
4	C	A	0.017	0.006	0.026	0.01	0.018	0.006	0.027	0.008
5	G	C	9.03	0.55	8.42	1.22	8.97	0.86	8.91	0.6
6	G	U	0.48	0.28	0.40	0.17	0.71	0.16	0.55	0.27
7	G	G	0.16	0.06	0.18	0.04	0.11	0.04	0.18	0.11
8	G	A	0.04	0.02	0.048	0.02	0.048	0.008	0.03	0.003
9	A	U	0.53	0.17	0.54	0.1	0.49	0.08	0.53	0.19
10	A	G	0.17	0.05	0.17	0.01	0.17	0.06	0.2	0.14
11	A	C	0.08	0.001	0.09	0.007	0.11	0.006	0.11	0.007
12	A	A	0.07	0.03	0.056	0.03	0.053	0.007	0.024	0.006
13	U	A	0.92	0.37	0.73	0.24	0.57	0.33	0.43	0.22
14	U	U	0.12	0.07	0.047	0.02	0.058	0.02	0.074	0.03
15	U	C	0.017	0.001	0.018	0.002	0.023	0.002	0.03	0.014
16	U	G	0.38	0.03	0.24	0.09	0.23	0.03	0.49	0.06

Rows indicated in green show the experimental rates of addition of a purine monomer across its cognate template base while those indicated in yellow show the experimental rates of addition of a pyrimidine monomer across its cognate template base. n=3 for all matched addition reactions, n=2 for all mismatched addition reactions.

2.3.3 Effect of admixture of co-solutes on the template-directed primer extension reactions

We then decided to experimentally increase the complexity of the reaction mixture, by adding both of the co-solutes together to the reaction. The combined effect of presence of DLPC and PEG 8000 was studied on all four cognate addition reactions as well as twelve mismatched addition reactions. Since there was a reduction in the rate of cognate purine addition reactions even in the presence of individual co-solutes, a similar effect was hypothesized for these reactions when a combination of both the co-solutes was used as well. As expected, the rate of addition of 'G across C' was clearly observed to have reduced from 1.58 ± 0.17 per hour in the control condition, to 0.69 ± 0.08 per hour in the presence of the admixture of DLPC and PEG 8000 ($p < 0.05$, Student's t test) (Fig 2.3, Table 2.1). For the addition of 'A across U', around 55% decrease in the rate of the reaction was observed in presence of both the co-solutes as compared to the control reaction rates, for each of the three replicates of this reaction. However, the variation in the absolute rates of the individual replicates was again large, giving rise to higher deviation from the mean (0.92 ± 0.37 per hour in the control condition vs. 0.43 ± 0.22 per hour in the presence of DLPC plus PEG 8000) (Fig 2.3, Table 2.1). Nonetheless, for both of these matched purine addition reactions, the combined effect of presence of both the co-solutes was more pronounced than the effect seen when only individual co-solutes were present. No apparent effect was observed on the rate of the remaining fourteen reactions (two cognate pyrimidine-based addition reactions, and twelve non-cognate addition reactions), in the presence of both the co-solutes, as compared to the control reactions (Table 2.1).

2.3.4 Statistical analysis

The difference observed between the rates of the reactions, in the presence of both the co-solutes vs. in the control conditions, was further subjected to methodical statistical analysis to evaluate the significance of these observations. This was accomplished by analyzing the Δ Rate (Δ Rate = rate of control reaction minus the rate of reaction in presence of both the co-solutes) on a two-dimensional plot where the Δ Rate was plotted on the X-axis and the Δ Rate normalized to the control rate was plotted on the Y-axis (Fig 2.4). The Δ Rate was calculated within each replicate for all combinations of incoming monomer–template base (total 16

combinations). To mitigate the differences in the absolute rates between the replicates, the normalized values were plotted on the Y-axis. Thus, on this plot, only those points which are distant from both the axes would point at a significant difference between the rates of the reactions for control vs. that for the experimental condition (in this case the combined effect of both the co-solutes).

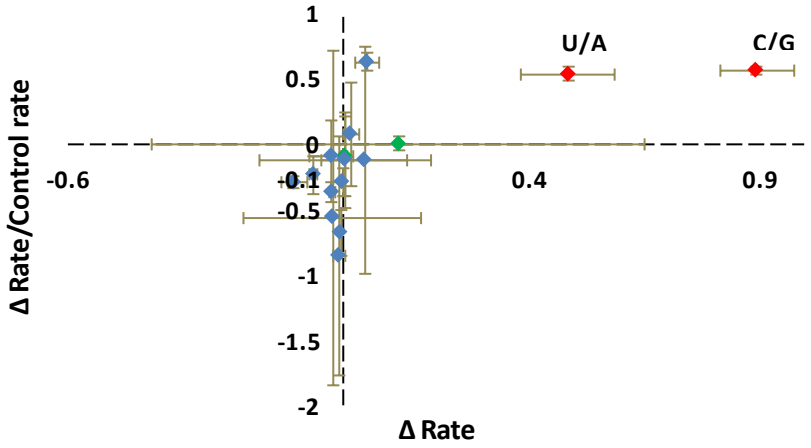


Figure 2.4: The 'Δ Rate normalized by the initial control rate'(Y axis) is plotted against the 'Δ Rate'(X axis) for each reaction. The 'Δ Rate' is significantly different from zero for addition of 'G across C' (point C/G) and addition of 'A across U' (point U/A), indicated as red diamonds. The green diamonds indicate the remaining two matched addition reactions. All mismatched addition reactions are indicated as blue diamonds. Error bars indicate the standard error. n=3 for all cognate addition reactions and n=2 for all mismatched addition reactions.

Δ Rate for only two reactions out of the sixteen studied, namely addition of 'G across C' and 'A across U' (points 'C/G' and 'U/A', respectively, indicated in red), was found to be significantly different from the zero cluster (Fig 2.4). The variation between the replicates can be visualized as the larger error bars along both the axes, which are normalized on the Y-axis. A few reactions e.g. addition of 'C across G', for which the rate of the reaction itself was high, larger error bars were observed along the X-axis. However, the normalized value on the Y-axis is close to zero, suggesting no significant difference between the rates of the reaction in the absence and presence of the co-solutes. On the other hand, for some of the slowest reactions (e.g. addition

of ‘C across U’), spurious error bars were observed on the Y-axis, as the rate of the control reaction (for such combinations) of incoming monomer–template base itself is very close to zero. Therefore, apart from ‘Δ Rate’ for purine cognate addition reactions, all other points are consistent with the zero cluster.

Subsequently, to determine the difference between the base level of Δ Rate and the deviation of the two points (namely, ‘U/A’ and ‘C/G’) from the same, robust iterative statistical analysis was carried out. This analysis, using three times standard deviation (99.5%), yielded 14 points with a mean = -0.005, and standard deviation = 0.054.The point for Δ Rate of addition of ‘G across C’ (‘C/G’) was excluded from this zero cluster at a confidence level of 99.5% ($\alpha = 0.005$). Similarly, the point for Δ Rate of addition of ‘A across U’ (‘U/A’) was excluded from this zero cluster at confidence level of 97.5% ($\alpha = 0.025$). This further confirmed that two out of the 16 possible reaction combinations in our nonenzymatic template-directed RNA primer extension system, were affected in the presence of both the co-solutes in the reaction mixture. Notably, both these reactions were addition of a purine monomer, across its cognate template base.

2.3.5 Estimation of the error rates

Previous experiments have indicated that for a nonenzymatic nucleic acid replication system, the observed misincorporation frequencies are much higher for ‘A’ and ‘U’ as template bases, as opposed to ‘G’ and ‘C’ as template bases (Orgel, 2004; Rajamani et al., 2010). Additionally, G: U wobble pair is, in general, the most commonly observed misincorporation in the nonenzymatic replication system (Orgel, 2004). The estimation of the misincorporation frequencies for our template-directed RNA primer extension system was consistent with these previous results. Amongst the different incorrect incorporations studied, the addition of ‘G across U’ was the most prominent misincorporation. Importantly, the frequency of this mismatched addition was found to have considerably increased in the presence of both the co-solutes (Fig 2.5A vs. Fig 2.5B). This, predominantly stemmed from the decreased rate of addition of the cognate base (ImpA) across ‘U’ template base. On the other hand, the frequency of addition of ‘U across G’ was not affected significantly in the presence of co-solutes. This is

due to the fact that the rate of the cognate addition, that of ImpC across G, was not affected due to the presence of the background molecules. In general, we observed an overall increase in the misincorporation frequencies against 'C' and 'U' template bases, in the presence of co-solutes as opposed to the control condition (Fig 2.5).

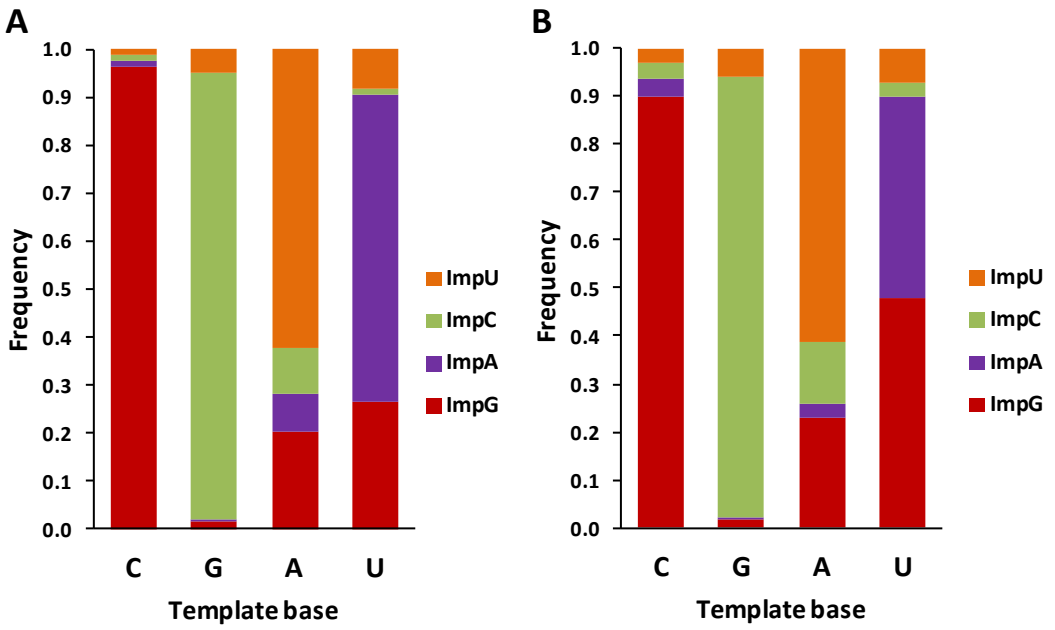


Figure 2.5: Incorporation frequencies for addition of cognate and non-cognate bases (A) In the absence of any co-solutes (B) In the presence of DLPC plus PEG8000 as co-solutes.

Furthermore, the error rates associated with our nonenzymatic replication system, in the absence and presence of the co-solutes, were also calculated. Two independent experimental replicates were considered for this calculation, wherein the rates of all possible sixteen combinations of reactions were experimentally determined, under control and co-solute conditions. For the control condition reactions, the mutation rate was estimated to be 19.8 ± 0.5% for a genome comprised of equal parts of A, U , G and C. In the case of the reactions where both of the co-solutes were present, the mutation rate was found to be 26.6 ± 4.4% (p=0.159, Student's t test) (Table 2.2). In the case of reactions with only the individual co-solutes added, the mutation rate in the presence of only DLPC was found to be 18.9 ± 0.2 %,

and that in the presence of only PEG 8000 was found to be 20.7 ± 0.4 %. Interestingly, the mutation rate for a putative GC-only genome increased from 5.3 ± 0.7 % in the control reactions, to 8.6 ± 0.3 % in the presence of both the co-solutes ($p<0.05$, Student's t test). The mutation rate for a putative AU-only genome was found to have increased from 34.3 ± 0.2 % in the control reactions, to 44.4 ± 9 % in the presence of both the co-solutes. The higher standard deviation for the mutation rate of AU-only genome in the presence of both the co-solutes stems from the inherent low fidelity of copying associated with A and U template bases (Orgel, 2004). When we considered only the individual template bases, the mutation rates for 'G' (6.8 ± 2.1 %) and 'C' (3.6 ± 0.3 %) bases was found to be lesser than 'A' (37.5 ± 4.9 %) and 'U' (31 ± 4.6 %), in the control condition. In the presence of both the co-solutes, the most prominent increase in the mutation rate, of almost 20%, was observed for the 'U' template base (31 ± 4.6 % in the control condition vs. 51.7 ± 3.5 % in the presence of both the co-solutes). The next template base that was most affected was the 'C' template base, for which the mutation rate increased from 3.6 ± 0.3 % in the control condition vs. 9.6 ± 1.6 % in the presence of both the co-solutes. Statistically significant increase was not observed in the mutation rates for 'G' and 'A' template bases in presence of both DLPC and PEG, as opposed to the control conditions. It is to be noted that, the mutation rate has significantly increased for both the pyrimidine template bases as the rate of addition of cognate purine bases against them showed significant reduction in the presence of the co-solutes.

Table 2.2: Experimental mutation rates for nonenzymatic RNA primer extension, in the absence and presence of co-solutes. Standard deviation is calculated from experimental replicates.

	For control reactions	**For reactions in presence of DLPC plus PEG**
Overall Mutation rate	19.8% ± 0.5%	26.6% ± 4.4%
Mutation rate for GC-only	5.3% ± 0.7%	8.6% ± 0.3%
Mutation rate for AU-only	34.3% ± 0.2%	44.4% ± 9%
Mutation rate for 'G' template	6.8% ± 2.1%	7.6% ± 1.1%
Mutation rate for 'C' template	3.6% ± 0.3%	9.6% ± 1.6%
Mutation rate for 'A' template	37.5% ± 4.9%	37.2% ± 17.6%
Mutation rate for 'U' template	31 ± 4.6 %	51.7 ± 3.5 %

2.3.6 Computer simulations for multiple replication cycles

To understand the impact of reduction in the rate of certain nonenzymatic matched addition reactions in the presence of co-solutes, over multiple cycles of replication, stochastic simulations were carried out using the Gillespie algorithm. The simulations were done using the 'MisInc_C' sequence as the template, and the experimentally generated rate constants for addition of all four nucleotides against all the four template bases, both, in the absence and the presence of co-solutes (essentially 64 combinations were used, Table 2.1). The data generated for replication cycle periods of 8hr, 16hr and 24 hr was analyzed. Almost negligible full-length copying of the templates was observed when the replication cycle time was set at 8hr (Fig2.6A). However, as the replication cycle period was increased to 16 hr, higher number of full-length (or intact) replications of the templates was seen, primarily in the control condition (Fig 2.6B, indicated in blue). However, due to the lowered rate of addition of purines against their cognate bases in the presence of co-solutes, longer duration of replication cycle period (viz. 24hr) was required to obtain significant *in silico* replication of the templates, in the presence of the co-solutes (Fig 2.6C, indicated in orange (DLPC), green (PEG) and red (both)).

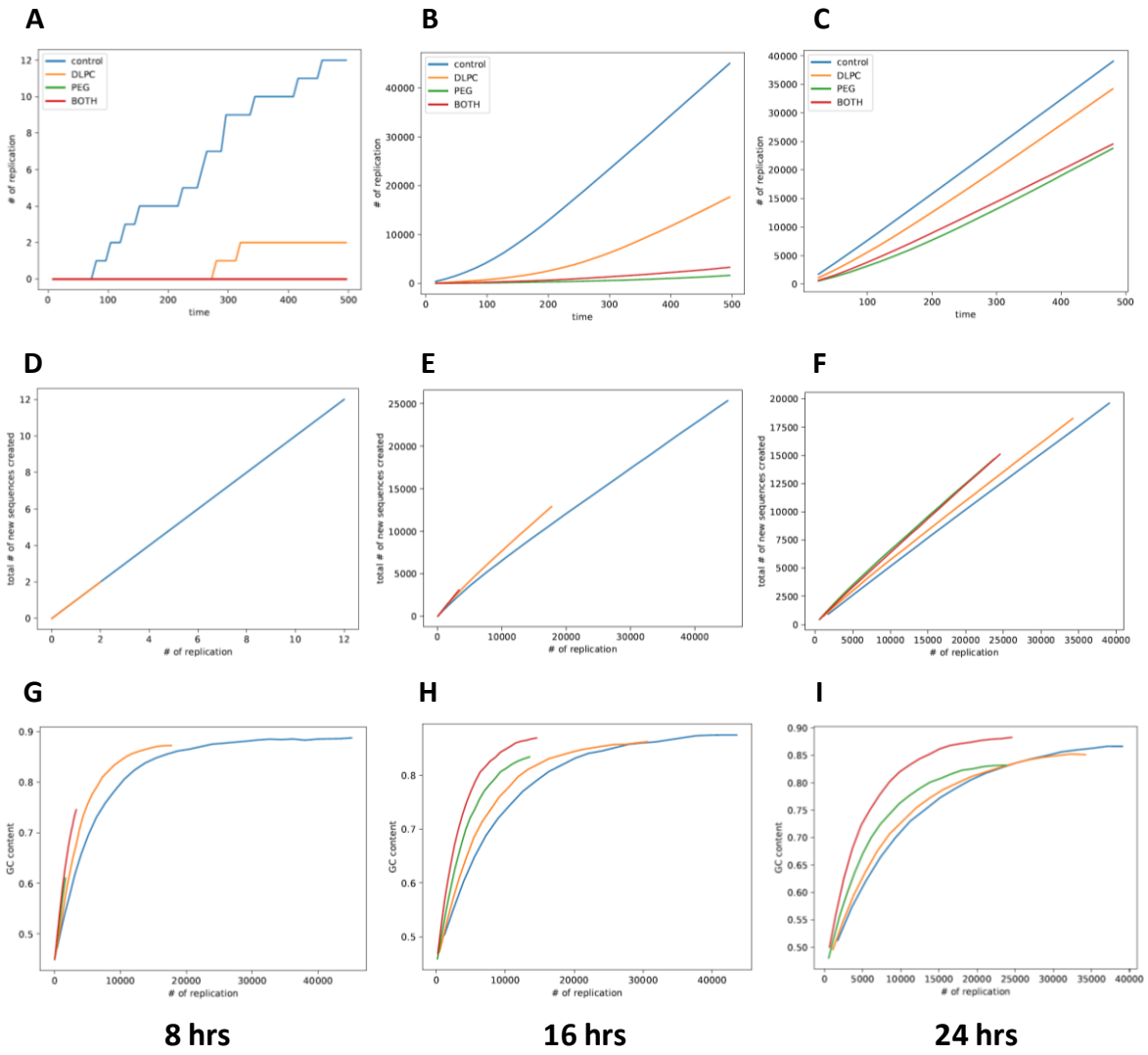


Figure 2.6: Analysis of *in silico* replication of the 'Misinc_C' RNA template, using experimental rates obtained for nonenzymatic copying. The extent of full-length copying of the templates increases with increase in the replication cycle time (A, B, C). Higher numbers of diverse sequences are observed in the presence of co-solutes as the replication cycle time is increased (D, E, F). The replicated sequences show higher GC-bias in the presence of co-solutes (G, H, I). The blue trace indicates *in silico* RNA replication in control conditions. The orange, green and red traces indicate the *in silico* RNA replication in presence of DLPC, PEG and both DLPC plus PEG, respectively. The time (in hrs) is indicative of the replication cycle time.

Furthermore, the total number of new sequences generated during these simulations was characterized. For the 8hr replication cycle period, very low numbers of new sequences were observed, which is concurrent with the lower number of full-length replications that were observed during this short replication duration (Fig 2.6D). Higher number of diverse sequences was observed in the presence of co-solutes over multiple generations of replication, if sufficient time was provided for completion of every replication cycle (Fig 2.6E, F). When the GC content of the resultant sequences was analyzed, GC bias was observed very early on during the replication cycles (Fig 2.6G). This might be due to the higher rate of incorporation of G and C nucleotides as against the A and U nucleotides, in general. The 'GC' enrichment was observed to be taking place at an even faster rate in the presence of co-solutes. In almost all the cases, when enough time was provided for the replication cycle to generate significant number of full-length daughter sequences, a GC bias of close to 85-90% was evident (Fig 2.6 H, I).

2.4 Discussion

Prebiotic milieu would have been replete with a mixture of many different molecular species. This heterogeneity might result from a mixture of products produced during prebiotic syntheses, in addition to the plethora of compounds that were potentially delivered to the Earth by exogenous bodies (e.g. meteorites and asteroids). However the existence of this inherent complexity is usually overlooked in studying pertinent nonenzymatic reactions, while delineating the transition from chemistry to biology. In these particular set of experiments, the effect of presence of DLPC and PEG 8000 as co-solutes, on template-directed RNA primer extension reactions, was characterized. DLPC is a phospholipid, which forms the main fraction of the modern cellular membranes. Phospholipids are the contemporary counterparts of the simpler amphiphiles like fatty acids, which have been implicated in the formation of protocellular membranes. The prebiotic pool is thought to have contained sufficient amounts of amphiphiles, as they could be synthesized using prebiotically viable FTT synthesis (McCollom et al., 1999). Additionally, organic components from certain carbonaceous chondrites have also been shown to form membrane like structures (Segre et al., 1999). Therefore, we wanted to

analyze the effect of this important co-solute on the rate and efficiency of a prebiotically relevant, nonenzymatic copying system. Another important component of the prebiotic soup would have been different kinds of polymers. A few of them would have eventually evolved to form essential functional polymers. However, a lot of non-specific polymers that would have resulted from polymerization of chemical entities like malic acid, lactic acid, ribose-5'-monophosphate etc. (Harshe et al., 2007; Mamajanov et al., 2014; Mungi and Rajamani, 2015) would have also existed in the same milieu, as background molecules. Furthermore, a study has indicated a prebiotically possible mechanism for selective enrichment of longer nucleic acid polymers (Kreysing et al., 2015), which would have further added to the complexity of the prebiotic milieu. PEG was used as a proxy for all such non-specific polymers, which collectively would have had a pronounced crowding effect. PEG is also routinely used for simulating volume exclusion effects in many experimental biochemical reactions. In our reaction set-up, the primer-template complex was pre-annealed before starting the reaction. The template was used in four times excess to ensure efficient primer-template complex formation, which is crucial as the aim of the study was to understand the effect of co-solutes on prebiotically relevant information transfer processes.

2.4.1 Effect of presence of co-solutes on nonenzymatic copying reactions

The rate of two of the fast matched addition reactions decreased in the presence of background molecules in the reaction mixture. This effect was seen in the presence of individual as well as the admixtures of DLPC and PEG 8000. However, background literature survey predicated an increase in the overall rate of information copying in our nonenzymatic RNA replication system that would result from volume exclusion effect/s. For example, the catalytic reaction rate for, both, protein- and nucleic acid-based reactions, is usually known to increase under crowded conditions due to volume exclusion (Ellis, 2001; Nakano et al., 2014). Interestingly, in our experimental system, the rate of only those matched addition reactions was found to have decreased wherein the incoming nucleotide was a purine (ImpG or ImpA). This effect might potentially be due to enhanced stacking interactions of the purine nucleotides in solution, as opposed to pyrimidine nucleotides. This feature is largely attributed to the

double-ring structure of purines vs. the single ring seen in pyrimidines. Such stacking could potentially result in the reduction of the diffusion of the monomers due to the formation of pseudo-oligomers under crowded conditions, like in the case of reactions that were carried out in the presence of co-solutes. This could further lead to the unavailability of monomers in such diffusion-limited, template-directed primer extension reactions, especially when involving a cognate purine addition. Interestingly, reduction in the purine addition across its template base was specially enhanced in the presence of both the co-solutes as opposed to the presence of just the individual co-solute (Table 2.1).

The addition of 'C across G' and that of 'U across A' was not found to have been affected in the presence of co-solutes. This might again be due to better availability of the pyrimidine monomers, even in the presence of co-solutes, due to their lower stacking tendency. Furthermore, the additions of mismatched monomers were also not significantly affected in the presence of DLPC and/or PEG, even though the incoming nucleotide was a purine. General observations indicate that the rate of nonenzymatic mismatched addition reactions is usually very slow, sometimes up-to two-orders of magnitude slower than the matched addition reactions (for e.g. addition of 'C across G' vs. addition of 'A across G', Table 2.1). Due to their inherently slow kinetics, the rate of these reactions might not be affected to a similar extent even if there was potentially a reduction in the mobility of stacked purine monomers, in the presence co-solutes.

2.4.2 Fidelity of nonenzymatic copying reactions in the presence of co-solutes

The nonenzymatic addition of a cognate nucleotide against a template base occurs at faster rate than that of non-complementary nucleotides. Due to the reduction in the rate of two out of these four reactions in the presence of DLPC and/or PEG, a concurrent decrease in the accuracy of replication was likely under similar conditions. In general, an increase in the mutation rates for the pyrimidine template bases (i.e. 'C' and 'U') was observed in the co-solute scenario. This can be explained by observed decrease in the rates of cognate purine additions in the presence of co-solutes vs. in their absence. As mentioned in the 'Results' section, the

addition of ‘G across U’ was the most frequently observed misincorporation, the frequency of which was further enhanced in the presence of DLPC and PEG. To begin with, the addition of ‘A across U’ is an error prone reaction (Orgel, 2004). The rate of this addition reaction was further found to have decreased by almost 55% in the presence of co-solutes. On the other hand, the rate of wobble addition of ‘G across U’ did not get affected in the presence of co-solutes. This resulted in the very pronounced reduction of accuracy of copying of the ‘U’ template base (mutation rate of 31 ± 4.6 % in the control reaction vs. 51.7 ± 3.5 % in the presence of both the co-solutes, $p<0.05$, Student’s t test), as compared to the other three template bases.

Amongst the four matched addition reactions, nonenzymatic addition of ‘G across C’ and vice versa is observed to be faster than the addition of ‘A across U’ and vice versa. Furthermore, due to increase in the error rates for copying of the ‘U’ template base in the presence of co-solutes, the complexity inherent to a prebiotic chemical environment could have adversely affected the copying of ‘AU’ rich genomes to a greater degree than that of ‘GC’ rich genomes. As hypothesized, the mutation rate for a ‘GC’ only genome was affected to a lesser extent (5.3 ± 0.7 % in the control reactions vs. 8.6 ± 0.3 % in the presence of both the co-solutes) as opposed to the ‘AU’-only genome (34.3 ± 0.2 % in the control reactions vs. 44.4 ± 9 % in the presence of both the co-solutes)(Table 2.2). The observed standard deviation for the mutation rate of AU-only genome, in the presence of co-solutes, was due to the greater intrinsic variability of the extension rates of the two reaction replicates for the addition of ‘G across U’ and that of ‘G across A’. Such high mutation rates for the copying of ‘A’ and ‘U’ template bases will bias the genome composition towards a more GC-rich one, which has been suggested earlier for a nonenzymatic replication system (Leu et al., 2011).

2.4.3 Effect of co-solutes over multiple cycles of nonenzymatic replications

Experimental realization of simultaneously occurring addition reactions, over multiple nonenzymatic replication cycles is difficult to achieve. This is mainly due to the limitation of handling a highly complex set of reactions, and side reactions, along with the intrinsically slower reaction rates of nonenzymatic replication, compounded further by problems related to strand

separation. Therefore, stochastic simulations were carried out in order to understand the effect of presence of co-solutes for multiple nonenzymatic replication cycles. Using 'MisInc_C' as the template and experimentally generated reaction rate constants, no significant *in silico* replication was observed for a replication cycle period of 8 hrs. This is due to the overall low rate of nonenzymatic incorporation of monomers across the template sequences, as even the best possible experimentally observed rate for any addition (for incorporation of 'C across G'), in control conditions, was 9.03±0.55 per hr (Table 2.1). Every other cognate addition was 5 to 17 times slower than the aforementioned rate. However, given enough replication cycle time, considerable number of replications are seen in, both, the absence and the presence of co-solutes.

Additionally, to understand the fidelity of this *in silico* replication, the total number of 'new' sequences generated was analyzed over cycles. Due to the lower replication rate, not many sequences were generated when each replication cycle time period was set to 8hrs. However, when sufficient time was provided for every replication cycle (viz. 16 hrs or 24 hrs), higher number of newly generated sequences were observed for the co-solute scenario, as compared to the control condition. This is probably due to the greater mutation rate observed in the presence of co-solutes as opposed to the control condition. Consequently, there is a higher chance of generating a different sequence than the cognate complement, for the templates, in the co-solute scenario and this is reflected in the graphs presented in Fig 2.6E, F. This is also indicative of the possibility that the presence of co-solutes in the prebiotic milieu might have helped in greater exploration of the sequence space. Given enough time, generation of point mutations will occur with greater frequency due to the higher chances of addition of incorrect nucleotides in the presence of co-solutes. This effect could be further enhanced as it has been demonstrated that the addition of one mismatch during enzyme-free copying, is usually followed by a cascade of mutations (Leu et al., 2013). Also, while analyzing the 'GC' content for these replication cycles, a strong GC bias was observed, both, in the absence and the presence of co-solutes. The rate at which this bias was attained, was however, higher in the case of the co-solute scenario. This was mostly due to the reduction in the rate of cognate purine additions, which was further compounded by an increase in the G:U wobble pairing in these

scenarios. This suggests that the intrinsically higher mutation rates associated with nonenzymatic replication systems would in general result in a genome composition enriched for 'GC' content (Leu et al., 2011). This effect could potentially be more pronounced in a complex prebiotic milieu, as suggested by the current study.

2.5 Conclusion

In general, our studies revealed the effect of prebiotic complexity on pertinent nonenzymatic reactions. This effect could arguably be context dependent and could be more pronounced for a certain set of reactions. In our system, even a seemingly 'step-wise' increase in the complexity of the reaction mixture, led to a decrease in the rate of certain nonenzymatic primer extension reactions. This strongly implies that a mixture of co-solutes, similar to what might have comprised a realistic prebiotic soup, could have had a pronounced role to play as far as the rate and accuracy of nonenzymatic replication is concerned. The simulation studies reveal that this effect gets enhanced over multiple generations of nonenzymatic replication. According to the *in silico* replication experiments, more varied sequences are generated in the presence of co-solutes, as opposed to the control conditions. This would lead to rapid exploration of the sequence space in a heterogeneous prebiotic milieu. However, this would also have posed a considerable challenge for information transfer and could have directly affected the process by which a functional molecule evolved under 'realistic' conditions. Therefore, to prevent functional polymers from undergoing such random mutations, it would have been crucial to encapsulate them, as also has been suggested previously (Szostak et al., 2001). Encapsulation has further been shown to play a favorable role in group selection of templates (Fontanari et al., 2006), as well as selection of replicases (Bianconi et al., 2013) in certain *in silico* studies.

2.6 References

Adamala, K., and Szostak, J. (2013). Competition Between Model Protocells Driven by an Encapsulated Catalyst. Nature Chemistry 5, 495–501.

Bianconi, G., Zhao, K., Chen, I.A., and Nowak, M.A. (2013). Selection for Replicases in Protocells. PLoS Computational Biology 9, e1003051.

Chen, I.A., Salehi-ashtiani, K., and Szostak, J.W. (2005). RNA Catalysis in Model Protocell Vesicles. Journal of the American Chemical Society 127, 13213–13219.

Desai, R., Kilburn, D., Lee, H., and Woodson, S.A. (2014). Increased Ribozyme Activity in Crowded Solutions. The Journal of Biological Chemistry 289, 2972–2977.

Ellis, R.J. (2001). Macromolecular Crowding : Obvious but Underappreciated. Trends in Biochemical Sciences 26, 597–604.

Fontanari, J.F., Santos, M., and Szathmáry, E. (2006). Coexistence and Error Propagation in Pre-Biotic Vesicle Models: A Group Selection Approach. Journal of Theoretical Biology 239, 247–256.

Harshe, Y.M., Storti, G., Morbidelli, M., Gelosa, S., and Moscatelli, D. (2007). Polycondensation Kinetics of Lactic Acid. Macromolecular Reaction Engineering 1, 611–621.

Kervio, E., Hochgesand, A., Steiner, U.E., and Richert, C. (2010). Templating Efficiency of Naked DNA. Proceedings of the National Academy of Sciences of the United States of America 107, 12074–12079.

Kilburn, D., Roh, J.H., Behrouzi, R., Briber, R.M., and Woodson, S.A. (2013). Crowders Perturb the Entropy of RNA Energy Landscapes to Favor Folding. Journal of the American Chemical Society 135, 10055–10063.

Kreysing, M., Keil, L., Lanzmich, S., and Braun, D. (2015). Heat Flux Across an Open Pore Enables the Continuous Replication and Selection of Oligonucleotides Towards Increasing Length. Nature Chemistry 7, 203–208.

Leu, K., Obermayer, B., Rajamani, S., Gerland, U., and Chen, I.A. (2011). The Prebiotic Evolutionary Advantage of Transferring Genetic Information from RNA to DNA. Nucleic Acids Research 39, 8135–8147.

Leu, K., Kervio, E., Obermayer, B., Turk-macleod, R.M., Yuan, C., Luevano, J., Chen, E., Gerland, U., Richert, C., and Chen, I.A. (2013). Cascade of Reduced Speed and Accuracy after Errors in

Enzyme-Free Copying of Nucleic Acid Sequences. Journal of the American Chemical Society 135, 354–366.

Mamajanov, I., Macdonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Fernandez, 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.

McCollom, T.M., Ritter, G., and Simoneit, B.R.T. (1999). Lipid Synthesis Under Hydrothermal Conditions by Fischer- Tropsch-Type Reactions. Origins of Life and Evolution of the Biosphere 29, 153–166.

Minton, A.P. (2001). Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. The Journal of Biological Chemistry 276, 10577–10580.

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.

Nakano, S., Karimata, H.T., Kitagawa, Y., and Sugimoto, N. (2009). Facilitation of RNA Enzyme Activity in the Molecular Crowding Media of Cosolutes. Journal of the American Chemical Society 131, 16881–16888.

Nakano, S., Miyoshi, D., and Sugimoto, N. (2014). Effects of Molecular Crowding on the Structures , Interactions and Functions of Nucleic Acids. Chemical Reviews 114, 2733–2758.

Orgel, L.E. (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 39, 99–123.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., and Chen, I.A. (2010). Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. Journal of the American Chemical Society 132, 1008–1011.

Segre, D., Ben-Eli, D., Deamer, D.W., and Doron, L. (1999). The Lipid World. Orig Life Evol Biosph 31, 119–145.

Strulson, C., Molden, R., Keating, C., and Bevilacqua, P. (2012). RNA Catalysis Through Compartmentalization. Nature Chemistry 4, 941–946.

Strulson, C.A., Yennawar, N.H., Rambo, R.P., and Bevilacqua, P.C. (2013). Molecular Crowding Favors Reactivity of a Human Ribozyme Under Physiological Ionic Conditions. Biochemistry 52, 8187–8197.

Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing Life. Nature 409, 387–390.

Zhou, H.-X., Rivas, G., and Minton, A.P. (2008). Macromolecular Crowding and Confinement: Biochemical, Biophysical and Potential Physiological Consequences. Annual Reviews of Biophysics 37, 375–397.

Zweckmair, T., Böhmdorfer, S., Bogolitsyna, A., Rosenau, T., Potthast, A., and Novalin, S. (2014). Accurate Analysis of Formose Reaction Products by LC–UV: An Analytical Challenge. Journal of Chromatographic Science 52, 169–175.

Chapter 3

Characterization of mechanism underlying co-solute based reactions

3.1 Introduction

The prebiotic solution would have been replete with different kinds of molecules as the products of prebiotically relevant syntheses are usually highly diverse. The exogenous delivery of material via meteoritic and other possible sources would have further added to this vast repertoire of molecules. Our effort to better understand the role of prebiotically relevant co-solutes in our reaction regimen resulted in us increasing the complexity of the reaction mixture using relevant co-solutes. This resulted in the decreased rate in two of the sixteen possible combinations tested in the context of nonenzymatic template-directed RNA primer extension reactions. It is peculiar to note that both of these reactions were purine based cognate addition reactions. This effect was evident even when the individual co-solute was used, with a more pronounced effect seen when using the admixture of both the co-solutes (Bapat and Rajamani, 2015). One of the co-solutes which was used, i.e. PEG, is known to affect the rate of contemporary biochemical reactions by facilitating crowding effect (Ellis, 2001). On the other hand, the other co-solute, i.e. the phospholipid, is also known to interact with nucleic acids in certain biological contexts (Kuvichkin, 2002). Nevertheless, it was perplexing as to why only the rate of two specific nonenzymatic reactions was prominently reduced in the presence of co-solutes, under our reaction set-up.

The inside of a typical cell is a crowded environment due to the presence of large amounts of molecules that exclude up to 20-40% of the available space (Ellis, 2001; Miyoshi and Sugimoto, 2008). This affects the way biochemical reactions occur inside a living cell vs. in a test tube. In order to simulate the molecular crowding observed inside the cell, many *in vitro* biochemical studies have extensively used water-soluble polymers like Ficoll, Dextran, PEG etc. These bulky inert polymers limit the volume of solvent available for other solutes in the reaction mixture by occupying significant amount of space. The presence of molecular crowding agents like PEG is known to affect various aspects of a protein, including its shape, folding, the enzymatic activity, and its interactions with other proteins etc. (Kuznetsova et al., 2014; Rivas and Minton, 2016). Their presence in the reaction milieu is known to affect nucleic acid-based reactions as well (Nakano and Sugimoto, 2017). Different sized PEG molecules are commercially available, making it easier to analyze the size-dependent effects of this crowding agent.

One conceivable way of explaining the effect crowding agents have on macromolecule-based reactions is the possible alteration in diffusion, caused by the presence of the bulky co-solute polymers. Crowding agents are known to reduce the diffusion coefficient (D) of both small and large molecules (Ellis, 2001). If the D is reduced, molecules would take much longer time to travel the same distance, thus resulting in reduced chances of encountering neighboring molecules or their interacting partners. Hence, if the reaction is diffusion-limited, its rate would decrease in the presence of co-solute polymers (Ellis, 2001). We envisage that in our template-directed RNA primer extension reactions, the presence of co-solutes is affecting the diffusion of monomers, thus resulting in the observed reduction in the rate of certain purine-based addition reactions. Interestingly, the purine nucleobases are better known for their greater π-stacking ability than the pyrimidine nucleobases (Sigel et al., 2014), which might result in longer stacks (pseudo-oligomers) of purine monomers. Furthermore, the diffusion coefficient, D, has been shown to be dependent on the number of nucleotides present in the ssRNA (Werner, 2011). The presence of PEG, therefore, might affect the stacking of purine pseudo-oligomers to a greater extent than the pyrimidine monomers, ultimately resulting in lesser availability of the purine monomers in a given reaction.

The other possible interaction in our reaction set-up containing co-solutes is that between RNA and lipid vesicles. The interaction between DNA and cationic liposomes has been known for a long time, mainly due to their importance as a gene delivery system (Lasic et al., 1997). Just like DNA, RNA is also known to bind liposomes (Marty et al., 2009). Significantly, Michael Yarus's group has also undertaken studies to discern RNA-lipid interactions in the context of an RNA World. Using the selection-amplification technique, this group found eight distinct RNA sequences of around 95-mer, which were able to bind and modulate the membrane permeability of liposomes comprised of phosphatidylcholine and cholesterol (Khvorova et al., 1999). Later, a heterotrimeric complex was described by the same group that could bind and modulate the permeability of liposomes consisting of only phosphatidylcholine, under physiologically relevant conditions (Vlassov et al., 2001). These interactions were also visualized using AFM and fluorescence microscopy (Janas and Yarus, 2003). This binding of RNA to liposomes showed a positive correlation between the affinity of RNA and the phospholipid

bilayer order (Janas et al., 2006). Additionally, binding of RNA is also known to result in the aggregation of larger vesicles, amongst a mixed population of different sized vesicles, indicating specific interactions between a particular RNA and the lipid in question (Thomas and Luisi, 2005). Similar interactions between DLPC lipid vesicles and RNA, under our reaction set-up, could potentially make the primer-template terminus unavailable for the extension, resulting in the observed decrease in the rate of nonenzymatic copying reactions, in the presence of co-solutes.

In this chapter, we have tried to delineate the underlying mechanism behind the observed effect of presence of PEG and/or DLPC vesicles, on nonenzymatic replication, in our RNA primer-template system. The extent of nucleotide stacking, in the absence and presence of molecular crowding agents, was estimated using Nuclear Magnetic Resonance (NMR) spectroscopy technique and is detailed in the first section of this chapter. Simultaneously, the RNA-lipid interaction was evaluated using fluorescence microscopy and is discussed in the latter half of the chapter.

3.2 Analysis of nucleotide stacking using NMR

3.2.1 Introduction

Historically, diffusion of nucleotides has been studied in the context of its role in macromolecular signaling in biological systems. Understanding the diffusion of second messenger molecules, like cAMP and cGMP, has helped unravel, for e.g., their crucial role in important biological functions like chemotactic response in *Dictyostelium discoideum* cells (Loomis, 2014), and meiosis in mouse ovarian follicles (Shuhaibar et al., 2015), respectively. Different techniques have been employed to understand this process. Early attempts of understanding the diffusion of nucleotides e.g ATP, cAMP etc., in plain solution, made use of simple diffusion cells (Bowen and Martin, 1964; Dworkin and Keller, 1977). Researchers have also used infrared thermal diffusion forced Rayleigh scattering (IR-TDFRS) setup in order to analyze the thermal diffusion of cAMP and cGMP in aqueous solution (Wang et al., 2012). Fluorescence correlation spectroscopy (FCS) is yet another technique that can be used to

understand the mobility of fluorescently tagged nucleotides or the interactions between nucleotides and fluorescent tags in solution (Ranjit and Levitus, 2012). More sophisticated methods have been used for analyzing nucleotide diffusion in biological context. A study published in 2001 used 2'-Cy3 tagged cAMP, in order to follow the diffusion of ligand-receptor complex in *Dictostellium* cells, using real-time epifluroscence microscopy imaging (Ueda et al., 2001). Another study, aimed at understanding second messenger trafficking in mammalian oocytes, used FRET analysis to track cGMP diffusion through gap junctions (Shuhaibar et al., 2015). However, tagging the nucleotides with such bulky fluorescent dyes might in itself affect the diffusion of the nucleotides.

Another efficient way of analyzing nucleotide diffusion without the need of either fluorescence tagging or complex biological machinery is by using Nuclear Magnetic Resonance (NMR) spectroscopy. The pulse sequences required for calculating the diffusion constants using this non-invasive NMR technique were proposed as early as 1965 (Cohen et al., 2005). The early aggregation studies included analysis of self-aggregation of the mononucleotides AMP, CMP, and UMP, in the presence of Mg^{2+}, using the Fourier transform NMR pulsed-gradient spin-echo multicomponent-self-diffusion technique (Stokkeland and Stilbs, 1985). The diffusion data from this study showed no observable aggregation tendency for the pyrimidine nucleotides. An important technical advent in the field of diffusion related NMR came in the early 1990s in the form diffusion ordered spectroscopy (DOSY) (Johnson, 1999). DOSY is a technique which monitors the translational diffusion of the molecules, which is a function of molecular size, shape, solvent temperature and viscosity. In DOSY NMR, pulse field gradients are used to spatially label the molecules in solution. If a molecule moves in a defined diffusion time after the application of the first gradient, then its new position can be monitored by application of the second gradient. The signal is detected as the average intensity over all the molecules in the sample tube. The signal modulation is achieved by varying the strength of the applied field gradient (Stejskal and Tanner, 1965; Johnson, 1999). Fitting of the collected intensities against the applied field strength allows determination of the diffusion constant. These diffusion constants can help to understand the aggregation behavior of nucleic acids, in the absence and presence of co-solutes.

Apart from DOSY, another way of estimating the molecular size using NMR is by tracking a molecule's rotational motion. The rotational motion of molecules can be studied by monitoring the nuclear spin relaxation time. The nuclear spin relaxation processes involve spin-spin relaxation as well as spin-lattice relaxation. The spin-lattice relaxation is the process of reestablishing the equilibrium population of the spin states after application of the radio frequency pulses, which lead to attenuation in these populations. The spin-lattice relaxation time, also known as longitudinal or T_1 relaxation time constant is a temperature dependent quantity. The T_1 relaxation time follows a parabolic function passing through a minimum with respect to molecular size (Fig 3.1). For small molecules, increasing the sample temperature enhances the random thermal fluctuations, thus resulting in an increase in the T_1 relaxation time. However for larger molecules, or in the presence of viscous liquids, increasing temperature reduces the T_1 relaxation time. This property can be used to determine stacking and diffusion of nucleotides, under different experimental conditions. We have used both the aforementioned methods, namely DOSY NMR and estimation of T_1 relaxation time, in order to estimate the extent of nucleotide stacking in our reactions. The work mentioned in this section was carried out in collaboration with Harshad Paithankar and Dr. Jeetender Chugh from the Chemistry department at IISER, Pune, India.

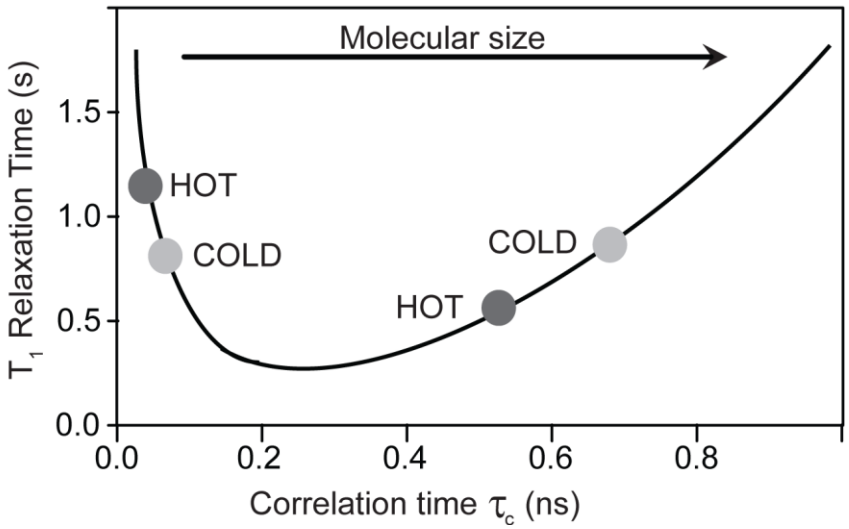


Figure 3.1: T_1 relaxation time as a function of molecular size as affected by temperature (reproduced from Spin dynamics by M H Levitte).

3.2.2 Materials and methods

3.2.2.1 Chemicals

The disodium salts of all four 5'-nucleoside monophosphates (5'-NMPs), viz. adenosine 5'-monophosphate (AMP), guanosine 5'-monophosphate (GMP), uridine 5'-monophosphate (UMP) and cytidine 5'-monophosphate (CMP), were purchased from Sigma Aldrich and used without any further purification. Analytical grade PEG 8000 and D_2O were also purchased from Sigma.

3.2.2.2 Methods

3.2.2.2.1 Sample preparation

The nucleotide stocks were prepared in nanopure water and the concentrations were estimated using UV spectroscopy (UV-1800 UV/Vis spectrophotometer, Shimadzu Corp., Japan). 50% w/v stock solution was prepared for PEG 8000 by dissolving the required amount of powder in nanopure water. Three different concentrations of nucleotides viz. 10mM, 50mM and 100mM were used to record the DOSY NMR data. For T_1 relaxation time measurements, the data was recorded at 10mM, 40mM and 100mM nucleotide concentration. The final concentration of PEG 8000 in the analyzed samples was maintained at 18% in order to keep it comparable to the amount that was used for the experiments mentioned in Chapter 2. 300µl of 1.1 times concentrated samples were first prepared and 10% D_2O was added to these samples for field locking before recording the NMR data.

3.2.2.2.2 NMR data acquisition

All the NMR relaxation experiments were recorded on Acsend^{TM} Bruker 600 MHz NMR spectrometer, furnished with quadruple ($^{1}H/^{13}C/^{15}N/^{31}P$) resonance cryoprobe equipped with X, Y, Z- gradient. DOSY experiments were recorded for the nucleotides in the presence and absence of PEG 8000. For each sample, the diffusion time and the gradient length was optimized to get 5-10% residual signal at 95% gradient strength. Sixteen data points with 8 scans for each point were recorded, with strength of the gradient ranging between 2-95%.

For ^{13}C-T_1 relaxtion time measurement, ^{13}C signals were excited and detected on the attached protons via polarization transfer. ^{13}C data was recorded by providing six inversion recovery delay (5 delay points and a repeat point for error estimation) in the range of 100 to 800 ms, so as to get 70% reduction in the signal intensity. For all the samples (i.e. control samples as well as those containing 18% PEG 8000), the T_1 relaxation data was recorded at two temperatures viz. 10°C (cold) and 25°C (hot).

3.2.2.2.3 Data analysis

The collected data was processed using SimFit algorithm as explained in standard Bruker DOSY data processing manual (Kerssebaum, R., DOSY and Diffusion by NMR, in *User Guide for XWinNMR 3.5*, Version 1.0; Bruker Buospin GmbH: Rheinstetten, Germany, 2002). The intensities extracted by SimFit algorithm were fit using the following two parameter mono-exponential equation in OriginPro 8.5.0 to get the diffusion constant for the solute molecule in the solvent used in the study:

$$I = I_0 \exp\left[-D\gamma^2 g^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right)\right]$$

where, *I* is the observed intensity, I_0 is the reference or un-attenuated intensity, *D* the diffusion coefficient, γ the gyromagnetic magnetic ratio of the observed nucleus, *g* the gradient strength, δ the length of the gradient pulse, and Δ the diffusion time.

^{13}C-T_1 relaxation data was processed and extracted as separate 1D corresponding to each recovery delay time, *t*. Peak picking was done in TopSpin3.2 (Bruker, Inc). The intensity data was then fitted using Mathematica v5.2 script (Spyracopoulos, 2006) to the following mono-exponential decay function, to get the longitudinal relaxation time constant (T_1):

$$I = I_0 \exp(-t/T_1)$$

The error in the T_1 time was obtained as fit error to the above equation from the Mathematica script.

3.2.3 Results

3.2.3.1 DOSY NMR analysis

To begin with, the DOSY data was recorded for the nucleotide samples in the absence of PEG. The diffusion constants (D) obtained for these samples are listed in Table 3.1. Concentration dependent increase in D was observed for all four nucleotides (Table 3.1, Fig 3.2). However, the DOSY data recorded for nucleotide samples in the presence of PEG did not show any clear trend. This could be due to effect of presence of PEG on the diffusion of the nucleotides. Hence, ^{13}C-T_1 relaxation experiment was further used as a tool to compare the relative sizes of the molecules, in the presence and absence of PEG.

Table 3.1: Diffusion constants for different nucleotides in the absence and presence of 18%PEG 8000 using DOSY NMR.

Concentration of Nucleotide (in mM)	Diffusion constant ($1*E^{-10}$) (m^2/s) Without PEG	Diffusion constant ($1*E^{-10}$) (m^2/s) With PEG
5'-AMP		
10	$5.08E^{-10} \pm 0.77E^{-12}$	$1.75E^{-10} \pm 1.18E^{-12}$
50	$4.84E^{-10} \pm 1.31E^{-12}$	$1.62E^{-10} \pm 0.68E^{-12}$
100	$4.51E^{-10} \pm 2.23E^{-12}$	$1.64E^{-10} \pm 0.43E^{-12}$
5'-GMP		
10	$4.98\ E^{-10} \pm 2.18E^{-12}$	$1.93E^{-10} \pm 2.20E^{-12}$
50	$4.57E^{-10} \pm 2.42E^{-12}$	$1.60E^{-10} \pm 0.94E^{-12}$
100	$4.29E^{-10} \pm 0.60E^{-12}$	$1.44E^{-10} \pm 0.45E^{-12}$
5'-CMP		
10	$5.53E^{-10} \pm 1.84E^{-12}$	$1.44E^{-10} \pm 5.05E^{-12}$
50	$5.29E^{-10} \pm 0.62E^{-12}$	$1.98E^{-10} \pm 1.02E^{-12}$
100	$5.06E^{-10} \pm 1.45E^{-12}$	$1.96E^{-10} \pm 0.40E^{-12}$
5'-UMP		
10	$5.57E^{-10} \pm 4.22E^{-12}$	$1.87E^{-10} \pm 5.03E^{-12}$
50	$5.38E^{-10} \pm 1.59E^{-12}$	$1.99E^{-10} \pm 2.15E^{-12}$
100	$5.03E^{-10} \pm 1.29E^{-12}$	$2.02E^{-10} \pm 0.19E^{-12}$

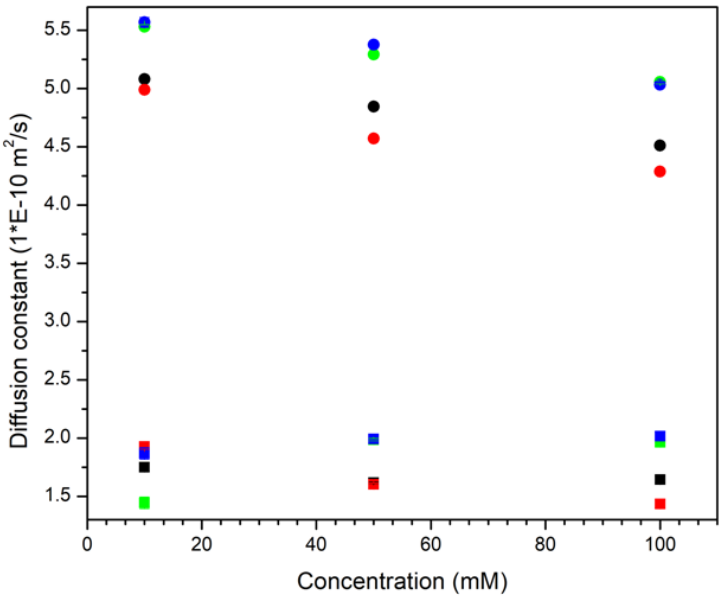


Figure 3.2: DOSY NMR measurements for nucleotide diffusion constants. The diffusion constants were measured at different concentrations, in the absence (indicated by circles), and in the presence of PEG (indicated by squares), for all four nucleotides viz. 5'AMP (indicated in black), 5'GMP (indicated in red), 5'CMP (indicated in green) and 5'UMP (indicated in blue). The X-axis indicates the three different concentrations at which the data was recorded.

3.2.3.2 T_1 relaxation analysis for nucleotides in the absence of PEG

T_1 relaxation data was recorded for all four 5'-NMPs at three different concentrations. The comparison of T_1 relaxation times for all these samples showed an increase with increase in the temperature from 10˚C to 25˚C (Fig 3.3, Table 3.2). This suggested that the nucleotides behaved as small molecules under our analytical conditions. However, as the concentration of nucleotides was increased from 10 to 100mM, a significant decrease (~150ms at 25°C and ~100ms at 10°C) in T_1 relaxation time was observed for purine monomers (5'-AMP and 5'-GMP). However, in the case of pyrimidine monomers (5'-UMP and 5'-CMP), no such significant change was observed. This might stem from the fact that the stacking ability of purine monomers is higher than the pyrimidines, in general (Sigel et al., 2014).

Table 3.2: T_1 relaxation time data for different nucleotides in the absence and presence of 18%PEG 8000.

Concentration of Nucleotide (in mM)	Without PEG		With PEG	
	T_1 (ms) at 10°C	T_1 (ms) at 25°C	T_1 (ms) at 10°C	T_1 (ms) at 25°C
5'-AMP				
10	468.48 ± 11.92	580.62 ± 14.94	271.54 ± 13.63	338.48 ± 10.87
40	361.69 ± 7.63	510.30 ± 8.26	287.24 ± 6.15	338.86 ± 3.57
100	326.10 ± 6.48	431.74 ± 4.39	271.16 ± 6.95	309.71 ± 12.16
5'-GMP				
10	383.65 ± 22.77	469.11 ± 3.26	317.14 ± 15.44	280.72 ± 21.94
40	340.42 ± 4.87	429.74 ± 15.62	244.62 ± 7.48	323.08 ± 4.67
100	297.11 ± 2.90	369.39 ± 2.98	253.91 ± 6.78	315.45 ± 7.27
5'-CMP				
10	391.23 ± 11.70	531.67 ± 5.75	289.84 ± 6.20	342.47 ± 18.37
40	380.12 ± 3.01	520.03 ± 6.81	284.39 ± 1.19	365.06 ± 6.87
100	361.04 ± 2.90	493.06 ± 4.55	271.92 ± 1.97	351.37 ± 4.49
5'-UMP				
10	384.18 ± 5.39	547.65 ± 6.17	272.89 ± 6.57	385.46 ± 15.27
40	378.79 ± 2.81	522.71 ± 4.02	292.04 ± 6.19	369.22 ± 10.55
100	368.43 ± 2.04	506.53 ± 5.27	280.27 ± 1.94	370.59 ± 5.05

3.2.3.3 T_1 relaxation analysis for nucleotides in the presence of PEG

In presence of PEG 8000, the T_1 relaxation time for the nucleotides was observed to be lower at 10°C than what was observed at 25°C (except 10mM GMP in the presence of PEG) (Fig 3.3, Table 3.2). This suggested that the nucleotides behaved as small molecules, for most part, even in the presence of PEG. The recorded T_1 values were in general lower in the presence of PEG, as compared to those obtained in its absence. The size of the nucleotide cluster/pseudo-oligomers, however, was bigger in the presence of PEG, as suggested by lesser difference in the recorded T_1 values at the two different temperatures.

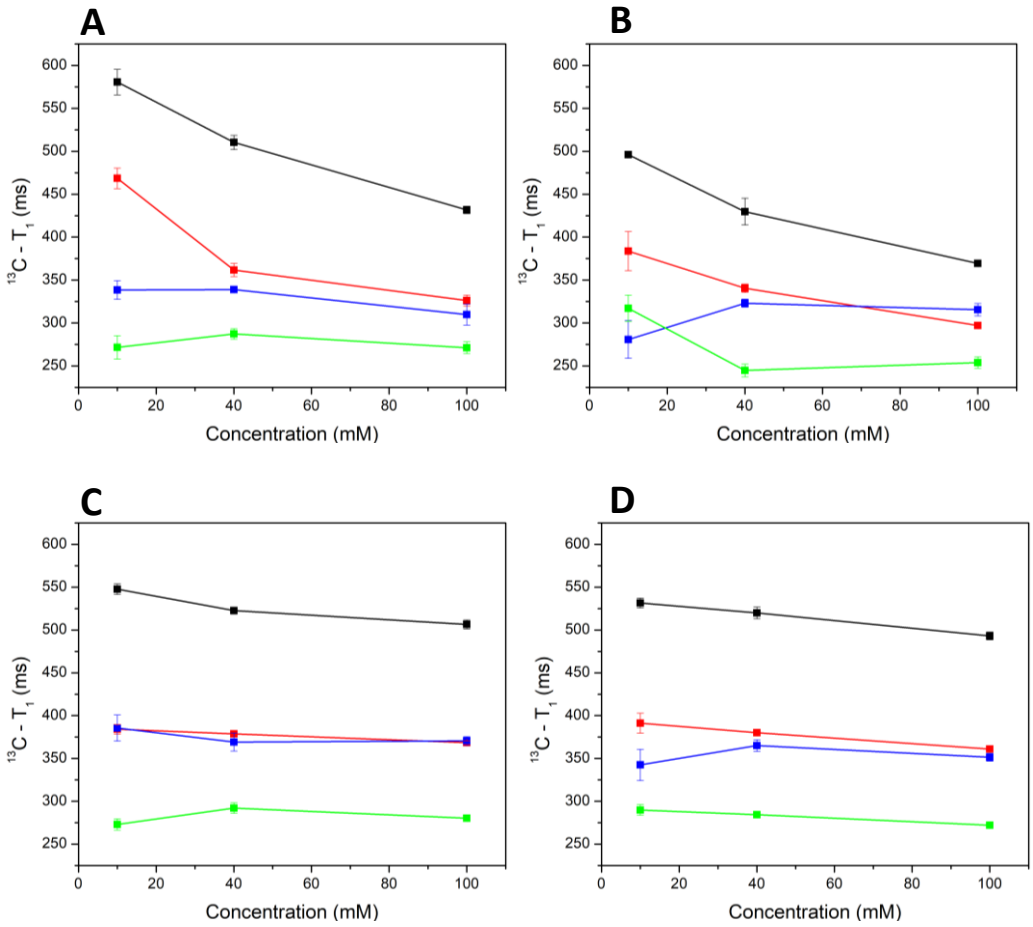


Figure 3.3: T_1 relaxation time data for different concentrations of (A) 5'AMP (B) 5'GMP (C) 5'UMP (D) 5'CMP. The data was recorded for the nucleotides in the absence (black trace: T_1 at 25˚C, red trace: T_1 at 10˚C) and in the presence of PEG (blue trace: T_1 at 25˚C, green trace: T_1 at 10˚C).

3.2.4 Discussion

In our previous experiments, the presence of co-solutes, namely PEG and lipid vesicles, were found to affect the rate and fidelity of template-directed nonenzymatic RNA copying reactions (Bapat and Rajamani, 2015). Given the fact that PEG's role as a molecular crowding agent is widely known, its effect on the diffusion of the reactants was suspected as one of the underlying mechanisms behind the aforementioned observations. NMR analysis was used to probe the effect of presence of PEG on nucleotide stacking, and thereby its role in affecting nucleotide diffusion in our reaction scheme. Along with being a non-invasive method, NMR also

bypasses the need to tag the nucleotides in order to track their movements, which is hugely advantageous. This is because, the attachment of any fluorescent tag would increase the effective nucleotide size as well as would most likely affect their stacking properties, and thereby their diffusion. Two different methods viz. DOSY NMR and T_1 relaxation time were used to estimate the diffusion constant of the nucleotides, and their molecular size.

To begin with, DOSY NMR was chosen as the method for analysis as one can have direct estimation of diffusion constants for molecules under different solution conditions. In the absence of PEG in the solution, the diffusion coefficients showed an increase with the concurrent increase in the concentration, for all the four nucleoside-5'-monophosphates. This suggested formation of higher order aggregates for all the four nucleotides. However the extent of aggregate formation was observed to be higher for purine based monomers than for pyrimidine monomers, as suggested by the extent of increase in the diffusion constants. For example, for 5'-AMP, the diffusion constant at 100mM was recorded to be $4.51E^{-10}$, while it was observed to be $5.03E^{-10}$ for 100mM 5'-UMP. This is most likely due to better stacking efficiency of the purine nucleobases. Due to π-stacking, larger aggregates/pseudo-oligomers can be formed in the case of purines, which would result in the decrease of overall diffusion at higher concentrations. However, this trend was not observed in the presence of PEG. Although the diffusion constants observed in the presence of PEG were higher in general, than those observed in its absence, no concentration dependent increase was observed. This might be due to the overall effect of PEG on the viscosity of the solution as well as the nucleotide stacking, such that no observable differences were detected in the diffusion constants at different concentrations.

Since diffusion constants calculated from DOSY could not provide the differential size estimation for the various nucleotide concentrations studied in the presence of PEG, T_1 relaxation time was subsequently recorded. This was done to get more detailed insights of the molecular size changes that could occur from the formation of pseudo-oligomers in crowded conditions. T_1 relaxation, which is also known as spin-lattice relaxation time, is a temperature-dependent function of the molecule and hence can be used to estimate its size. The diffusion

properties can then be further estimated from the molecular size estimation. When the T_1 data was recorded for three different concentrations of nucleotides in the absence of PEG, an increase in the T_1 value was observed for concurrent increase in the temperature from 10˚C to 25˚C. Therefore, according to the molecular behavior depicted in Figure 3.1 for small vs. large molecules, it was clear that the nucleotides acted as small molecules under this experimental set-up. However, a notable and concentration dependent increase in T_1, with increase in temperature, was observed only for purine (5'-AMP and 5'-GMP) monomers. This suggests a concentration dependent increase in the molecular size for purine monomers as compared to the pyrimidine monomers. This stems from better π- stacking properties exhibited by the purines, which might then lead to the formation of pseudo-oligomers, ultimately resulting in larger molecule size at higher concentrations. This larger molecular size would result in a concurrent decrease in diffusion, such that the stacked purine monomers might be less available for the diffusion-limited reactions, as compared to the pyrimidine monomers. This effect could further get enhanced at higher concentrations of the nucleotides.

In the presence of PEG, the T_1 relaxation time was observed to have decreased for all nucleotides at all concentrations in general, as compared to those samples analyzed in the absence of PEG. This is primarily due to increase in the viscosity of the solution in the presence of bulky PEG polymers. When the data was analyzed closely, it was noted that the overall decrease in the T_1 values at 25˚C, to those observed at 10˚C, was lower for the purine nucleotides as compared to those obtained for pyrimidine nucleotides. For example, the T_1 for 100mM 5'-AMP at 25˚C was found to be around 310ms and that at 10˚C was ~ 270ms (the difference being that of ~40ms). On the other hand, the T_1 relaxation time for 100mM 5'-UMP at 25˚C was found to be 370ms and that at 10˚C was 280ms (i.e. the difference of ~90ms). The reduction in the T_1 value from higher temperature to the lower temperature would be less for a molecule/molecular cluster that is comparatively of larger size (Fig 3.1). Thus, this variation in the difference is suggestive of higher size of molecular clusters/pseudo-oligomers for purine monomers in the presence of PEG, than that for pyrimidine monomers. PEG might facilitate higher degree of stacking for purine monomers, resulting in the sequestration of the monomers into these pseudo-oligomeric structures.

Another peculiar thing that was observed for the relaxation data recorded in the presence of PEG was the T_1 values for 5'-GMP. For 10mM of 5'GMP, the recorded T_1 relaxation time was higher at 10˚C than at 25˚C, in the presence of PEG (Table 3.2). This indicated the presence of large nucleotide aggregates (Fig 3.3). This meant that in the presence of PEG, at lower concentration, 5'-GMP acted as a large molecule, as estimated by the T_1 relaxation data. However at the higher concentrations of 40mM and 100mM of 5'-GMP, the recorded T_1 values in the presence of PEG were again found to be lower at 10˚C (Table 3.2). This indicated that at higher concentrations, the 5'-GMP cluster acted as small molecules as compared to that at 10mM concentration. The molecular crowding induced by PEG is known to promote formation of non-canonical G-quadruplex structures that result from the formation of non-Watson-Crick interactions between guanine (Miyoshi et al., 2006). We therefore suspect that, in the presence of PEG, 5'-GMP might be forming G-quadruplex-like structures at higher concentrations, resulting in the compaction of the nucleotide clusters and thus yielding a T_1 relaxation time trend characteristic of a small molecule. To analyze this phenomenon further, ^{1}H NMR analysis was carried out to confirm the presence of G-quadruplex-like structures at higher concentrations of 5'-GMP in the presence of PEG. The formation of hydrogen bonds in G-quadruplexes results in the stabilization of exchangeable hydrogen, which can be observed as peak at ~10-12 ppm on ^{1}H NMR. The peak corresponding to the presence of G-quadruplex was observed at 40mM 5'GMP concentration and was absent at 10mM 5'GMP concentration (Fig 3.4). This observation supported the T_1 data by confirming the formation of compact G-quadruplex like structures at higher GMP concentration, as compared to the looser stacks present at lower GMP concentrations.

In summary, both the DOSY NMR and T_1 relaxation time data indicates a higher propensity for stacking of the purine monomers as against the pyrimidine monomers. Furthermore, the T_1 data also indicated that this stacking tendency increased in the presence of PEG 8000. Effect of molecular crowding on π-stacking might result in the formation of pseudo-oligomers to a greater extent in the presence of PEG, as opposed to in plain aqueous solution. This would result in the sequestration of higher percentage of purine monomers in these aggregates. These results explain, at least in part, the observed reduction in the rate of purine based

cognate nonenzymatic addition reactions in the presence of PEG. Similar NMR studies will be carried out for analyzing the effect of the other co-solute viz. DLPC vesicles, and also for the admixture of both the co-solutes (viz. PEG 8000 and DLPC vesicles), on the nucleotide stacking behavior. Depending upon our NMR analysis, we suspect that comparable results might be observed for other molecular crowding agents as well, further highlighting the importance of accounting for prebiotic heterogeneity while studying pertinent nonenzymatic copying reactions.

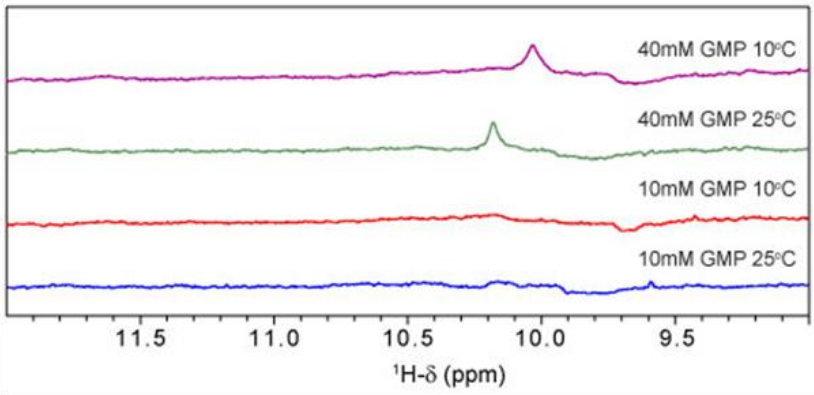


Figure 3.4: ^{1}H NMR profile indicating the presence of G-quadruplex like structure at higher concentration of 5'GMP. A peak near 10 ppm which was observed only for 40mM 5'GMP is characteristic of formation of G-qudrplex structure.

3.3 Analysis of RNA-lipid interaction using imaging

3.3.1 Introduction

Different techniques have been utilized to understand the interactions between lipid vesicles and nucleic acids. Lasic et al. have analyzed DNA-lipid vesicle interactions using structural analysis tools such as small angle X-ray scattering (SAXS) (Lasic et al., 1997). Fourier transform infrared (FTIR), UV-visible, and circular dichroism spectroscopic methods have been used to analyze the lipid binding site, binding constant and tRNA secondary structure in lipid-RNA adducts (Marty et al., 2009). Another technique namely, differential scanning calorimetry (DSC), has been used widely to understand the thermodynamic and kinetic parameters for membrane-nucleic acid interactions (Giatrellis and Nounesis, 2011). Nonetheless, microscopy remains as one of the popular techniques for direct visualization of nucleic acid binding to lipids. Electron

microscopy (EM), including cryo EM, has been used widely to understand nucleic acid-lipid binding (Junquera and Aicart, 2016). On the other hand, Yarus's group has used atomic force and fluorescence microscopy techniques to analyze the binding of *in vitro* selected and evolved RNAs, to liposomes (Janas and Yarus, 2003; Janas et al., 2006). Since the RNA primer used in our studies was tagged with a fluorescent dye at the 5'-end, it was in principle, suitable to use fluorescence imaging based analysis to characterize the RNA-lipid interactions that might be facilitated under our reaction set-up.

The microscopy analysis for RNA-lipid interactions that was carried out by Yarus and his coworkers mainly used giant lipid vesicles (Janas and Yarus, 2003; Janas et al., 2006). However, the co-localization of lipid vesicles and RNA might become difficult to visualize due to drifting of vesicles in solution. To stabilize the lipid component and minimize the drifting while imaging, a new technique was used in our experiment. The lipid vesicles were fused with silica beads in order to form SUPER (supported bilayers with excess membrane reservoir) templates (Pucadyil and Schmid, 2010). These coated beads with excessive membrane reservoir, which mimic vesicular structures, were then used to understand the binding of RNA to the lipid membrane. Another possible way of analyzing lipid-RNA interactions is to study binding of RNA to supported lipid bilayers. A recent study reports the characterization of 'supported membrane tubes (SMrTs)' for their utilization in understanding membrane fission process (Dar et al., 2017). The advantage of working with these systems is that the drifting while image acquisition is minimum as the membrane is supported on a passivated glass coverslip. Therefore, this protocol was used to generate supported DLPC lipid bilayers, which was then used to study the potential binding of RNA to this lipid. The work mentioned in this section was carried out in collaboration with Soumya Bhattacharyya and Dr. Thomas Pucadyil from the Biology department at IISER, Pune, India.

3.3.2 Materials and methods

3.3.2.1 Chemicals

The lipid used in this study, namely 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC), was purchased from Avanti Polar Lipids Inc., Alabama, USA. A 20-mer primer Amino-G was used in this study, the details of which are given in section 2.2.1. The Cy3 and DiO dyes were acquired from Glen research, USA and Thermo Fischer Scientific, USA, respectively. Uniform sized plain silica beads (of 5.2µm size) were purchased from Corpuscular Inc., NY, USA. Other chemicals used for analysis, such as BSA, NaCl, NaOH, PEG etc. were of analytical grade and purchased from Sigma-Aldrich.

3.3.2.2 Methods

3.3.2.2.1 Preparation of small unilamellar vesicles (SUVs)

The glass tube in which the lipid was aliquoted was first cleaned with 1% SDS, followed by methanol and then chloroform. The remnant of chloroform was air dried. To prepare 500µl SUV solution containing 1mM DLPC lipid, 12.5µl of 25mg/ml chloroform stock was then aliquoted into the clean glass tube using a Hamilton syringe. DiO dye was added to a final concentration of 0.01mol%, wherever required. The lipid solution was then dried under a gentle stream of air by swirling the tube. This ensured the formation of an even thin layer of lipid on the inner surface of the glass tube. The tube was subsequently kept under vacuum for 30 minutes to ensure the complete removal of chloroform. 500µl of nanopure water was added to the tube, which was then incubated at 50˚C for 30 minutes to facilitate swelling of the lipid. The tube was then briefly vortexed followed by sonication (Sonics Inc., CT, USA). The sonication of the lipid mixture was carried out for 5 minutes at 2 sec burst and 3 sec wait time, at 30% amplitude. The suspension was then transferred to a low-binding microcentrifuge tube prior to centrifugation at 100,000g for 20 minutes at room temperature. The supernatant containing the SUVs was used for further processing. This suspension can be stored at 4˚C for up to a week.

3.3.2.2.2 SUPER template preparation

The SUPER templates were prepared using a previously reported protocol (Pucadyil and Schmid, 2010). In brief, 70µl of nanopure water was added to a low-binding microcentrifuge tube. 0.2µl of 5M NaCl solution, 20µl of 1 mM SUV suspension and 10µl of plain silica bead stock (5.2µm) were added to the tube, in the mentioned order. The mixture was vortexed briefly before incubation in dark for 30 minutes, at room temperature. 1ml of nanopure water was added after this incubation and the resultant mix was mildly vortexed. It was then subjected to a spin in a swinging bucket rotor, at 90 g for 2 minutes at RT. This wash step was repeated thrice. After the last wash step, 1ml of the supernatant was removed and the particles were re-suspended in remaining 100µl of volume by gently tapping the tube. This suspension was then used to visualize the RNA-lipid interaction.

3.3.2.2.3 Incubation of RNA with SUPER templates and imaging

The SUPER templates and the fluorescently tagged 20-mer RNA were incubated in a LabTek chamber (Nunc, USA) to study the binding of the RNA to the lipid. The 8-well chamber was first treated with 1M NaOH solution for 15 minutes. The well surfaces were then passivated using 1mg/ml solution of Bovine Serum Albumin (BSA). To the passivated chamber, nanopure water, RNA (or dye) stock solution and SUPER template suspension was added, in that order. Typically 10µl of the SUPER template suspension was added to each well and various concentrations of either the RNA, or the fluorescent dye, was added to analyze concentration dependent binding process. The final volume of the incubated samples was maintained at 200µl. The samples were incubated in dark for a stipulated time before imaging using Olympus IX71 inverted microscope through a 100X, 1.4 NA oil-immersion objective. The fluorescent tags were excited using LED light source (Thorlabs). The signals were collected using single band pass filters (Semrock) with excitation/emission band pass of 482 ± 35 nm/536 ± 40 nm for DiO probe, and of 532 ± 35 nm /570 ± 40 nm for the Cy3 tag on an Evolve 512 EMCCD camera (Photometrics). The acquired images were analyzed using ImageJ software (NIH, USA).

3.3.2.2.4 Preparation of supported lipid bilayer (SLB)

The SLB was prepared using a previously reported protocol (Dar et al., 2017). Briefly, a very small quantity of chloroform stock containing 1mM of DLPC lipid, along with 0.01mol% of DiO dye, was spread on a PEGylated coverslip. The chloroform was evaporated completely before assembling the coverslip in a flow cell (Bioptechs, PA, USA). The flow cell was filled with filtered and degassed buffer (150mM KCl, 20mM HEPES, pH 7.4), and the subsequent controlled flow of this buffer (using a peristaltic pump) caused the formation of SMrT templates. The SLB was observed near the source where the chloroform stock was originally spotted.

3.3.2.2.5 Incubation of RNA with SLB and imaging

The SLB, along with the preformed SMrT templates, was first equilibrated in filtered and degassed buffer containing 150mM KCl and 20mM HEPES at pH 7.4. 200µl of primer Amino-G, in buffer, was added and incubated with SLB for 45 minutes in the chamber. Excess of RNA was then rinsed using 1ml of the buffer. The RNA binding was assessed using fluorescence imaging, the details of which are mentioned in the section 'Incubation of RNA with SUPER templates and imaging' (section 3.3.2.2.3).

3.3.3 Results

3.3.3.1 Assay using SUPER templates

The RNA-lipid interaction could be visualized clearly using GUVs and fluorescence microscopy. However, the drifting of the GUVs hindered efficient imaging in our preliminary analysis. Therefore, SUPER templates were chosen to avoid drifting of the lipid vesicles (Fig 3.5A). The RNA binding interactions were assessed using three different concentrations of Cy3 labeled 20-mer RNA viz. 10nM, 40nM and 100nM. A free form of 'Cy3' dye was used as a control at similar concentration levels. After 45 minutes of incubation, we observed a concentration dependent binding for the Cy3 labeled RNA (indicated by the red trace in Fig 3.5B). No such significant trend was observed for the free Cy3 dye as indicated by the black trace in Fig 3.5B. The average fluorescence intensity obtained for RNA binding was greater than that obtained for Cy3 binding,

at any given concentration. However, RNA was also found to be binding to the silica beads, which were not coated with DLPC (indicated by green point in Fig 3.5B). This deemed the SUPER template system unreliable to assess the potential interactions between the Amino-G primer and the DLPC lipid.

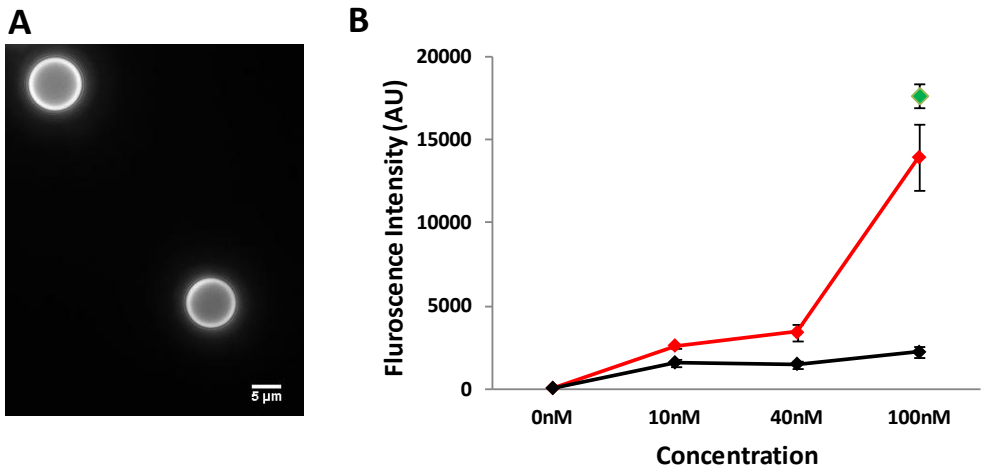


Figure 3.5: (A) Fluorescence microscopy image showing binding of Cy3 labeled RNA to the DLPC coated silica beads (B) Binding of Cy3 alone (black trace) and Cy3 labeled RNA (red trace), to the DLPC coated silica beads at different concentrations. RNA was found to be binding to the uncoated silica beads as well, as indicated by the green point. The error bars represents S.E. (N=50 beads).

3.3.3.2 SLB assay

Since the RNA was found to bind to even those silica beads that were not coated with DLPC lipid (Fig 3.5B), a different assay system was used to analyze the binding of the RNA to the lipid. Herein, a supported lipid bilayer consisting of DLPC, along with the DiO dye, was obtained on a PEGylated coverslip. Formation of the stable bilayer, which ensured minimization of out-of-focus movements in solution, was confirmed using fluorescence microscopy (Fig 3.6, upper panel images). Two different concentrations of Cy3 labeled RNA viz. 100nM and 1µM were used in order to check for RNA binding. The RNA (in buffer solution) was incubated with the SLB for 45 minutes to ensure enough interaction time. However, no RNA binding was observed at both

these concentrations as was evident by lack of any fluorescence signal in the corresponding channel (Fig 3.6, lower panel images).

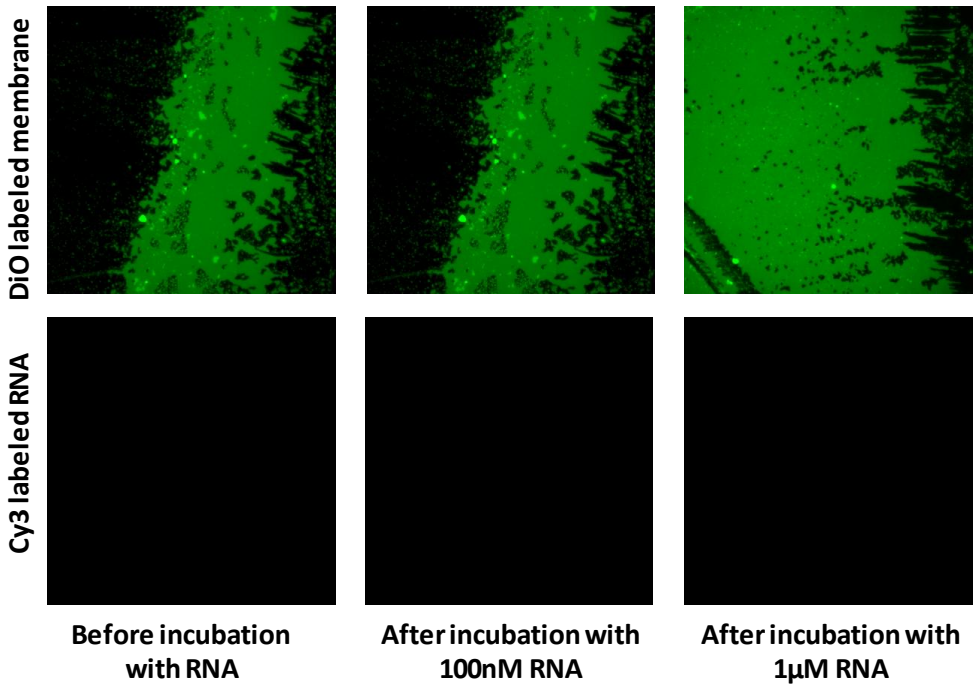


Figure 3.6: SLB assay for RNA-lipid binding analysis. Supported bilayer formation was observed using DiO labeled lipid (upper panel). No fluorescence signal was obtained after incubation of the Cy3 labeled 20-mer RNA with the SLB (lower panel).

3.3.4 Discussion

As detailed in Chapter 2, presence of lipid vesicles in the starting reaction mixture was found to lower the rate of certain enzyme-free template-dependent primer extension reactions (Bapat and Rajamani, 2015b). Previous studies from Michael Yarus's group have demonstrated the interaction between RNA oligomers and lipid vesicles (Janas and Yarus, 2003; Janas et al., 2006). However, the length of the RNAs used for these studies was at least 80 nucleotides long. To check whether stable RNA-lipid interactions are observed even with a smaller length RNA such as the 20-mer RNA primer used in our experiments, binding experiments were carried out. The use of GUVs is prevalent for studying nucleic acid–lipid vesicle interactions. However, drifting of the GUVs could hamper simultaneous imaging of the fluorescently tagged lipids and

nucleic acids. Hence, two newer methods of microscopic analysis were used to assess the potential interactions between the Cy3 tagged 20-mer RNA and DLPC lipid.

To begin with, the binding of RNA to curved membranes was analyzed using silica beads coated with DLPC. The phospholipid DLPC was specifically used as the effect of this particular lipid as a co-solute has been previously analyzed in prebiotic context (Bapat and Rajamani, 2015). To achieve this coating, a previously reported protocol to obtain SUPER templates was followed (Pucadyil and Schmid, 2010). This method exploits charge based interactions for the formation of excess lipid bilayer on the silica beads. With these SUPER templates, concentration dependent binding of the Cy3 tagged RNA primer to the beads, likely coated with the DLPC, was observed (Fig 3.5). To analyze whether this binding was observed due to potential charge-based interactions, or due the partitioning of the Cy3 dye in the lipid phase, parallel 'control' binding assays were carried out with free Cy3 dye. No concentration dependent binding of the Cy3 dye was apparent (Fig 3.5B, as indicated by black trace), hinting at a charge based interaction between negatively charged RNA backbone and the SUPER templates. However, to our surprise, similar interaction was also observed between RNA primer and the silica beads that were not coated with any lipids (Fig 3.5B, as indicated by green point). This indicated inefficiency in the formation of SUPER templates, when using DLPC alone for this process. Since DLPC is neutral at physiological pH, it might not be able to form the excess bilayer on the silica beads by itself. This might also be the reason behind the absence of concentration dependent binding of the Cy3 to the beads. Previous studies have actually indicated the partitioning of the Cy3 dye in the lipid phase (Hughes et al., 2014). Furthermore, it is suspected that the RNA binding that was observed, might be due to the electrostatic interactions between the uncoated/partially coated silica beads and the backbone of the RNA. The issue of the binding of DLPC to the silica beads can be possibly mitigated by using 10-15% of phosphatidylserine-based lipid/s as a dopant, which would enhance the electrostatic interactions between the lipid stock and the silica beads. It is of our interest to standardize the SUPER template assay further for its utility in analyzing the RNA–lipid interactions.

To analyze the binding of small RNA to the supported flat bilayer, SLB assay was used. SLB was obtained near the source of the lipid, when the protocol for the formation of SMrT was followed (Dar et al., 2017). Use of SLB is advantageous as the bilayer is fully supported, thus overcoming any issues related to the drifting of lipid vesicles during imaging. However, no RNA binding to the lipid bilayer was observed using this assay system, even at RNA concentration as high as 1μM of (Fig 3.6). This also indicated that the Cy3 did not partition into the flat DLPC bilayer. Importantly, the absence of RNA binding in this case might also be due to the smaller size of the RNA used in our study as compared to the previous studies, wherein longer RNAs were used to demonstrate RNA-lipid binding as visualized using fluorescence imaging (Janas and Yarus, 2003; Janas et al., 2006).

In summary, no significant binding was observed between the 20-mer RNA and the DLPC lipid, using either SUPER template assay or SLB assay. This could be due to the absence of effective electrostatic interactions between the RNA and the lipid. It is suspected that, in addition to the smaller size of the RNA, the presence of no effective charge on the DLPC lipid at neutral pH, could have contributed towards the observations recorded in aforementioned assay systems. Systematic studies involving variation of both, the length of the RNA, and the composition of the lipid bilayer, would shed further light on the role of these factors during the RNA-lipid binding.

3.4 Conclusion

Understanding the effect of prebiotic heterogeneity on pertinent enzyme-free reactions remains poorly explored to date. Our initial studies regarding increasing the complexity of the reaction milieu of nonenzymatic template-directed primer extension reactions, revealed the effect of co-solutes on certain enzyme-free replication reactions. As detailed in Chapter 2, in the presence of PEG or/and DLPC lipid vesicles, the rate of those matched addition reactions was found to have decreased wherein the incoming nucleotide was a purine monomer (Bapat and Rajamani, 2015). In the present study, we attempted to delineate the underlying mechanism of this observed effect. The presence of co-solutes could potentially affect the

monomers or the RNA primer-template complexes themselves. Using NMR based techniques, the stacking tendency of the purine monomers were found to have increased in the presence of PEG. Increased stacking, primarily caused due to π-stacking interactions, would result in higher chances of sequestration of the monomers into pseudo-oligomers. As PEG was one of the co-solutes used in our reaction mixture, this observation at least in part, explains the decrease in the rate of only the purine based nonenzymatic matched addition reactions. On the other hand, no significant RNA-lipid binding was observed using the SUPER template assay or SLB assay. This also suggests that the decrease in the rate of certain nonenzymatic RNA primer extension reactions in the presence of DLPC as a co-solute, might stem from the effect the lipid vesicles have on the monomers, as opposed to our hypothesis based on RNA-lipid binding. In preliminary experiments undertaken in a collaborator's lab, direct interaction of the nucleic acid monomers has been observed with the lipid bilayer (personal communication with Dr. V. Raghunathan at RRI, Bangalore, India). Further characterization of such interactions under our reaction conditions would be helpful in understanding their fundamental implications for the emergence of early cellular life.

3.5 References

Bapat, N. V, and Rajamani, S. (2015). Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids. Journal of Molecular Evolution 81, 72–80.

Bowen, W.J., and Martin, H.L. (1964). The Diffusion of Adenosine Triphosphate Through Aqueous Solutions. Archives of Biochemistry and Biophysics 107, 30–36.

Cohen, Y., Avram, L., and Frish, L. (2005). Diffusion NMR Spectroscopy in Supramolecular and Combinatorial Chemistry: An Old Parameter—New Insights. Angewandte Chemie International Edition 44, 520–554.

Dar, S., Kamerkar, S.C., and Pucadyil, T.J. (2017). Use of the Supported Membrane Tube Assay System for Real-Time Analysis of Membrane Fission Reactions. Nature Protocols 12, 390–400.

Dworkin, M., and Keller, K.H. (1977). Solubility and Diffusion Coefficient of Adenosine 3' -5' Monophosphate. The Journal of Biological Chemistry 252, 864–865.

Ellis, R.J. (2001). Macromolecular Crowding : Obvious but Underappreciated. Trends in Biochemical Sciences 26, 597–604.

Giatrellis, S., and Nounesis, G. (2011). Nucleic Acid-Lipid Membrane Interactions Studied by DSC. Journal of Pharmacy and Bioallied Sciences 3, 70–76.

Hughes, L.D., Rawle, R.J., and Boxer, S.G. (2014). Choose Your Label Wisely: Water-Soluble Fluorophores Often Interact with Lipid Bilayers. PLoS ONE 9, e87649.

Janas, T., and Yarus, M. (2003). Visualization of Membrane RNAs. RNA 9, 1353–1361.

Janas, T., Janas, T., and Yarus, M. (2006). Specific RNA Binding to Ordered Phospholipid Bilayers. Nucleic Acids Research 34, 2128–2136.

Johnson, C S, J. (1999). Diffusion Ordered Nuclear Magnetic Resonance Spectroscopy : Principles and Applications. Progress in Nuclear Magnetic Resonance Spectroscopy 34, 203–256.

Junquera, E., and Aicart, E. (2016). Recent Progress in Gene Therapy to Deliver Nucleic Acids with Multivalent Cationic Vectors. Advances in Colloid and Interface Science 233, 161–175.

Khvorova, A., Kwak, Y.-G., Tamkun, M., Majerfeld, I., and Yarus, M. (1999). RNAs that Bind and Change the Permeability of Phospholipid Membranes. Proceedings of the National Academy of Sciences 96, 10649–10654.

Kuvichkin, V. V (2002). DNA–lipid Interactions *in vitro* and *in vivo*. Bioelectrochemistry 58, 3–12.

Kuznetsova, M.I., Turoverov, K.K., and Uversky, N.V. (2014). What Macromolecular Crowding Can Do to a Protein. International Journal of Molecular Sciences 15, 12, 23090–23140.

Lasic, D.D., Strey, H., Stuart, M.C.A., Podgornik, R., and Frederik, P.M. (1997). The Structure of DNA−Liposome Complexes. Journal of the American Chemical Society 119, 832–833.

Loomis, W.F. (2014). Cell Signaling During Development of *Dictyostelium*. Developmental Biology 391, 1–16.

Marty, R., N'soukpoé-Kossi, C.N., Charbonneau, D.M., Kreplak, L., and Tajmir-Riahi, H.-A. (2009). Structural Characterization of Cationic Lipid–tRNA Complexes. Nucleic Acids Research 37, 5197–5207.

Miyoshi, D., and Sugimoto, N. (2008). Molecular Crowding Effects on Structure and Stability of DNA. Biochimie 90, 1040–1051.

Miyoshi, D., Karimata, H., and Sugimoto, N. (2006). Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. Journal of the American Chemical Society 128, 7957–7963.

Nakano, S., and Sugimoto, N. (2017). Model Studies of the Effects of Intracellular Crowding on Nucleic Acid Interactions. Molecular BioSystems 13, 32–41.

Pucadyil, T.J., and Schmid, S.L. (2010). Supported Bilayers with Excess Membrane Reservoir: A Template for Reconstituting Membrane Budding and Fission. Biophysical Journal 99, 517–525.

Ranjit, S., and Levitus, M. (2012). Probing the Interaction Between Fluorophores and DNA Nucleotides by Fluorescence Correlation Spectroscopy and Fluorescence Quenching. Photochemistry and Photobiology 88, 782–791.

Rivas, G., and Minton, A.P. (2016). Macromolecular Crowding In Vitro, In Vivo, and In Between. Trends in Biochemical Sciences 41, 970–981.

Shuhaibar, L.C., Egbert, J.R., Norris, R.P., Lampe, P.D., Nikolaev, V.O., Thunemann, M., Wen, L., Feil, R., and Jaffe, L.A. (2015). Intercellular Signaling via Cyclic GMP Diffusion Through Gap Junctions Restarts Meiosis in Mouse Ovarian Follicles. Proceedings of the National Academy of Sciences 112, 5527–5532.

Sigel, A., Operschall, B.P., and Sigel, H. (2014). Comparison of the π-Stacking Properties of Purine versus Pyrimidine Residues. Some Generalizations Regarding Selectivity. Journal of Biological Inorganic Chemistry 19, 691–703.

Chapter 4

Template-directed replication using nucleoside- 5'-monophosphate monomers under DH-RH conditions

4.1 Introduction

Activated nucleotides are routinely used while studying nonenzymatic reactions pertinent to prebiotic chemistry. Several different activation chemistries, including imidazole, 2'-methylimidazole, 1'-methyladenine, 2'-aminoimidazole (Orgel, 2004; Li et al., 2017), oxyazabenzotriazole (Deck et al., 2011) etc., have been explored thus far to successfully demonstrate enzyme-free oligomerization and copying reactions. This is mainly done to achieve the formation of the phosphodiester or phosphoramidite bond/s on laboratory time scale and also to have detectable yields of products. The formation of a phosphodiester bond, being an uphill reaction involving the removal of a water molecule, is otherwise very difficult to achieve in solution. However, a key point to note here is that, presence or existence of significant amounts of most of the aforementioned activated nucleotides, on the prebiotic Earth, is largely arguable. There have hardly been any studies on the demonstration of prebiotically plausible routes for the synthesis of large quantities of activated nucleic acid monomers. One recent study demonstrated a probable route for the synthesis of imidazole activated ribonucleotides. This particular study had used an *in situ* generated cyanogen chloride to form diimidazole amine from imidazole, which in turn activates the 5'-NMPs to their 5'-phosphorimidazolides (Yi et al., 2018). However, the long term stability of such activated monomers still remains questionable. Due to their high intrinsic energy, these monomers tend to lose their activation group quickly while in the solution. The activated monomers can also undergo intramolecular cyclization to yield cyclic monomers (Fig 4.1). Such 'spent monomers' usually compete with the activated monomers and are thought to result in the decrease of the rate of the enzyme-free copying or extension reactions (Kervio et al., 2014).

Nevertheless, possible prebiotic routes for the synthesis of non-activated nucleotides and their precursors have been demonstrated. John Sutherland's group started with the synthesis of precursor pentose amino-oxazolines and finally achieved the synthesis of cyclic pyrimidine nucleotides (Powner et al., 2009). This alternate way of synthesizing non-activated nucleotides has already been elaborated in section 1.3.1 in Chapter 1. Very few studies, however, have demonstrated oligomerization reactions using prebiotically plausible non-activated nucleoside monophosphate monomers (Rajamani et al., 2007; DeGuzman et al., 2014). Only a single study

to date has actually characterized the chemical nature of the products formed in such reactions (Mungi and Rajamani, 2015). Almost all of these polymerization studies have employed alternate cycles of dehydration and rehydration (DH-RH) at elevated temperature, in order to achieve the polymerization of these non-activated moieties. This is because the formation of phosphodiester bond, which is essentially a condensation reaction, is very difficult to achieve at room temperature in bulk water. Polymerization from 5'-NMP is also a slow reaction due to lack of a good leaving group (such as an oxyazabenzotriazole moiety) on the phosphate. Dehydration at higher temperature enhances the loss of water, promoting bond formation by facilitating the condensation reaction. Dehydration also concentrates the monomers, thus increasing the chances of bond formation between them. The subsequent rehydration phase helps in re-distribution of the monomers and oligomers, consequently increasing the overall efficiency of the reaction. Terrestrial geothermal fields and inter-tidal pools are some of the probable prebiotic niches where such alternate DH-RH cycles would have existed. These geological features are thought to have been prevalent on the early Earth. The DH-RH cycles at these niches might have been facilitated by diurnal cycles, seasonal variations etc. Interestingly, similar reaction regimen has also been used to obtain polymers from malic acid (Mamajanov et al., 2014) and lactic acid monomers (Harshe et al., 2007), highlighting the relevance of such niches in promoting prebiotically pertinent reactions.

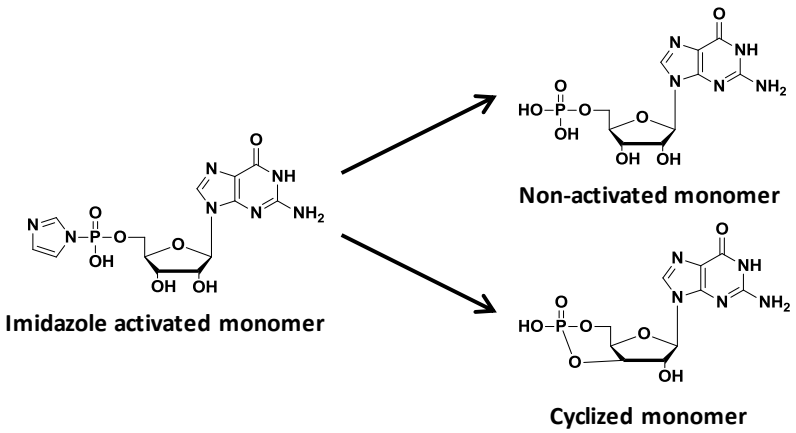


Figure 4.1: Spontaneous degradation of the activated monomers yields either non-activated monomers (by hydrolysis) or cyclic monomers (by internal cyclization).

On the other hand, the nonenzymatic copying reactions involving the use of non-activated nucleotides remains a largely unexplored area of research. In fact, to the best of our knowledge, only a single study thus far has attempted prebiotically relevant information transfer reaction using non-activated nucleotides. In this particular study, which was reported by David Deamer's group, the authors tried to copy the information present in a ssDNA strand, using 2'-deoxyribonucleoside 5'-monophosphates, under alternate cycles of dehydration and rehydration (DH-RH), in the presence of phospholipids (Olasagasti et al., 2011). Lipids are known to form fluid lamellar matrices under anhydrous conditions, facilitating the concentration of the starting reactants (e.g. nucleotides) within the inter-layers of the resultant multilamellar structures (Deamer, 2012) (Fig 4.2A). The reaction yield was estimated to be 0.5%, which was calculated based on the starting concentration of the template (Fig 4.2B). The misincorporation frequency for this DNA based system was observed to be 9.9%. However, no experiments have been reported thus far, for the information transfer from an RNA template using non-activated nucleotides, under prebiotically plausible environmental conditions. Given the importance of RNA molecules in a putative RNA world, it becomes important to gain an insight into this crucial process.

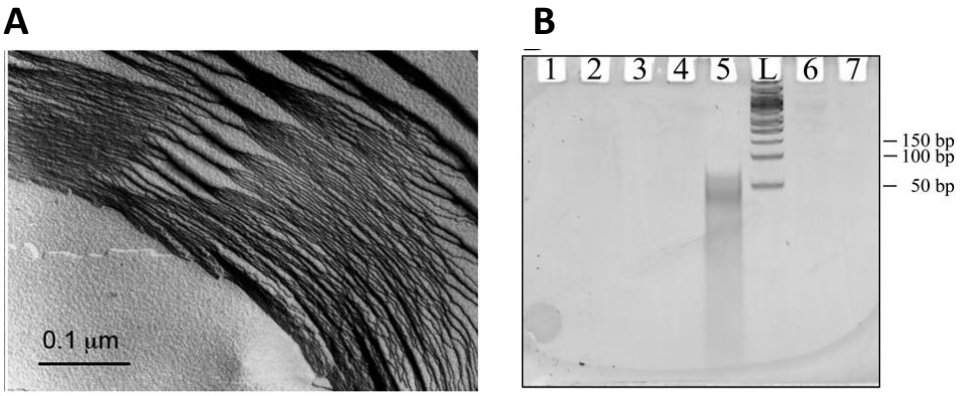


Figure 4.2: (A) Formation of multilamellar structures by phospholipids in dehydrated condition (reproduced from Deamer, 2012) (B) The information from ssDNA could only be copied in the presence of phospholipid in DH-RH reaction regimen at high temperature and low pH, as indicated in the lane 5 (reproduced from Olasagasti et al., 2011).

In this particular chapter, we have studied lipid-assisted template-directed RNA primer extension under prebiotically relevant DH-RH reaction regimen, using relevant non-activated monomers. To obtain the yields of the extended primer in detectable yields, 3'-amino-2', 3'-dideoxynucleotide terminated primer was used. Furthermore, using this primer also makes it easier to compare and analyze our reactions with other nonenzymatic copying studies that have been carried out using the similar primer sequence but with various activated nucleotides (Rajamani et al., 2010; Leu et al., 2013; Bapat and Rajamani, 2015). The effects of different parameters like temperature, rehydration solution etc., on these reactions, have also been systematically characterized. Importantly, the extension of a pre-existing RNA primer, using 5'-ribose monophosphate (5'-rMP), a sugar-phosphate monomer, was also undertaken that resulted in hybrid polymers. Such hybrid polymers have been hypothesized to have played an important role in sampling a variety of bases, during the formation and evolution of primitive information polymers of pre-RNA World(s). Furthermore, preliminary attempts were made to characterize the products obtained from a reaction between the aforementioned RNA monomer and 5'-rMP.

4.2 Materials and Methods

4.2.1 Chemicals

The disodium salts of all four 5'-nucleoside monophosphates (5'-NMPs), viz. adenosine 5'-monophosphate (5'-AMP), guanosine 5'-monophosphate (5'-GMP), uridine 5'-monophosphate (5'-UMP), cytidine 5'-monophosphate (5'-CMP), and ribose 5'-monophosphate (5'-rMP), were purchased from Sigma-Aldrich and used without further purification. 3'-amino 2', 3'-dideoxy adenosine (3'-NH_2-ddA) was acquired from Metkinen, Finland, and used without further purification. Analytical grade Trizma base (Tris- hydroxymethyl- aminomethane), sodium chloride, boric acid, ethylenediaminetetraacetic acid (EDTA), and all the mineral and organic acids used in the reactions were also purchased from Sigma-Aldrich. The double chain surfactant lipid used in the reactions, namely 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC), was purchased from Avanti Polar Lipids Inc., Alabama, USA.

Two 20-mer long RNA primers and four different 29-mer long RNA templates were used in the reactions, the sequences for which are mentioned below. The templating base has been highlighted in bold:

Name	**Sequence (5'-> 3')**
Primer Amino G	GG GAU UAA UAC GAC UCA CUG-NH_2
Primer Hydroxyl G	GG GAU UAA UAC GAC UCA CUG
Template MisInc_C	AGU GAU CU**C** CAG UGA GUC GUA UUA AUC CC
Template MisInc_G	AGU GAU CU**G** CAG UGA GUC GUA UUA AUC CC
Template MisInc_A	AGU GAU CU**A** CAG UGA GUC GUA UUA AUC CC
Template MisInc_U	AGU GAU CU**U** CAG UGA GUC GUA UUA AUC CC

The RNA primer (Primer Amino G), which was acquired from Keck laboratory, Yale, USA, is terminated with a 3'-amino-2', 3'-dideoxynucleotide (procured from Metkinen, Finland). This modification was done to ensure the addition of the incoming nucleotide to the primer, and using amino group facilitates this more efficiently than the canonical hydroxyl group, as the former is a better nucleophile. A control RNA primer (Primer Hydroxyl G), terminating in the canonical 5'-guanosine monophosphate, was also used. The primer Amino G was labeled with 'Cy3', and Primer Hydroxyl G was labeled with 'DY547' fluorescent tags, on the 5' end, for enabling their detection upon polyacrylamide gel electrophoresis (PAGE) analysis. Amino G primer was first deprotected using triethylamine trihydrofluoride (TEA.3HF) to remove the 2'-TBDMS protection groups. The deprotecetd RNA primer was then gel purified prior to using it in the reactions. The Hydroxyl G primer and the RNA templates were acquired from Thermo Scientific (Dharmacon). The 2'-ACE protection was removed using deprotection buffer (100mM acetic acid, adjusted to pH 3.8 using TEMED) and following the manufacturer's protocol. These RNA oligos were then used without further purification.

4.2.2 Methods

4.2.2.1 Setting-up the primer extension reactions

The RNA oligomer stock solutions were made in 10mM Tris buffer (pH 7.0) .The nucleotide monomer stock solutions were made in nanopure water. The concentrations for these reaction components were estimated by UV spectroscopy (UV-1800 UV/Vis spectrophotometer, Shimadzu Corp., Japan). A typical reaction was set up by annealing 0.65µM of the primer and 1.3µM of the template to each other, in RNase free water (Sigma), by heating at 95˚C for 5 minutes and followed by cooling at room temperature (RT). 100mM Tris (pH 7.0) and 200mM NaCl were then added to the annealed primer-template complex. DLPC vesicles were prepared by first drying the required volume of the 25mg/ml chloroform stock in a clean glass vial. The dried film was then resuspended in nanopure water. It was subsequently vortexed to get a lipid stock of the desired concentration. The lipid stock was extruded through a 100 nm membrane using a mini extruder (Avanti Polar lipids Inc., USA), to get small unilamellar vesicles. DLPC vesicles and 5'-NMPs were added to the reaction mixture at a final concentration of 5mM and 10mM, respectively, unless otherwise specified. The reaction mixture (typical final volume of 10µl) was then allowed to dry at elevated temperature, and was held at the same temperature for a total of 30 minutes (including the drying time), unless otherwise specified. The dried mixture was then rehydrated using the rehydration solution. After allowing rehydration for 5 minutes, the procedure was repeated in a similar manner for multiple DH-RH cycles. A sample aliquot was removed for analysis post the rehydration phase, at predetermined time intervals. The sample volumes were adjusted to a standardized amount to compensate for the degradation of RNA primer that occured over multiple DH-RH cycles. Typical volumes withdrawn as aliquots, at different DH-RH cycle intervals, are mentioned in the following table:

DH-RH cycle number	Volume of sample withdrawn
Cycle 0	0.5 µl
Cycle 3	1 µl
Cycle 5	1.5 µl
Cycle 7	3 µl
Cycle 10	4 µl

4.2.2.2 Reaction products analysis

As detailed above, the necessary samples were collected in TBE buffer (10.8 gm Trizma base, 5.5 gm boric acid, 4ml of 0.5M EDTA in 1000 ml of distilled water) containing 8M urea and 100mM EDTA, at the aforementioned time intervals during the course of the DH-RH reaction. A competitor RNA, which did not have a fluorescence tag, and which has a sequence that is exactly similar to that of the tagged primer, was used (5' GG GAU UAA UAC GAC UCA CUG). This was added in at least 10 time excess to the reaction samples, to successfully separate the fluorescent primer in question from the template, for unhindered gel analysis. After the addition of the competitor RNA, the samples were heated at 95˚C for 5 minutes, cooled at RT for 5 minutes and briefly centrifuged, before loading on the gel. The extended primer products were analyzed on 20% denaturing PAGE containing 8M urea as denaturant (SequaGel-UreaGel system, National diagnostics, USA) (Gel size : 17 X 15 cm). 1X TBE buffer was used as the gel running buffer. Bromophenol blue and Xylene cyanol were used as tracker dyes to estimate the progress of the RNA primer bands during gel electrophoresis. The gels were imaged on a Typhoon Trio plus imager (GE Healthcare) at 550PMT and 100 micron resolution setting, using the 532nm excitation laser. This allowed for the efficient detection of the fluorescently labelled primer and the resultant extended products. The gel images were subsequently processed using ImageQuant v 5.2 software (GE healthcare), to minimally adjust the contrast and also for quantifying the band intensities in certain reactions.

4.2.2.3 Mass analysis of reaction products

To understand the nature of products formed by the addition of 5'-rMP to the pre-existing RNA primer, a proxy reaction was set up with just the monomers. The reaction was set up with 10mM each of 3'-NH_2-ddA and 5'-rMP in nanopure water. The pH of the reaction mixture was set to 2 using sulphuric acid. This mixture was subjected to five consecutive cycles of DH-RH at 90˚C. The sample was collected at the end of the DH-RH cycles and the different products were purified using semi-preparative HPLC. Chromatography was carried out using an Agilent 1260 chromatography system (Agilent Technologies, Santa Clara, CA, USA) and DNAPac PA200 column (9 × 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 3 mL/min. The collected samples were lypophilized. The mass analysis for these samples was carried out in collaboration with Dr. Yayoi Hongo from Earth Life Science Institute (ELSI), Tokyo, Japan. For this, Acquity UPLC+ system from Waters, with a CORTCES UPLC C18 column (1.6µm, 2.1 x 50 mm) was used with 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.

4.3 Results

The polymerization of non-activated monomers (5'-NMPs) has been demonstrated under multiple cycles of DH-RH, at low pH and high temperature (Rajamani et al., 2007; Mungi and Rajamani, 2015). It was observed that the reaction rate was further enhanced in the presence of phospholipid molecules. This could have been facilitated by the entrapment of the monomers in the multiple lipid layers that are formed under dehydration conditions (Toppozini et al., 2013; Himbert et al., 2016). We wanted to analyze whether similar reaction regimens would also support the template-directed information transfer involving an RNA template-primer complex and non-activated nucleotides.

4.3.1 RNA primer extension under DH-RH condition

The preliminary reactions were carried out by subjecting the mixture of RNA primer-template complex (0.65µM of primer and 1.3 µM of the template) and 10mM of nucleotide, to alternate DH-RH cycles at 90˚C. 1mM H_2SO_4 was used as rehydration solution to ensure the low pH of the reaction over the cycles. 5mM of 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) lipid was used in the starting reaction mix. DLPC vesicles were extruded prior to adding to the reaction mixture in order to obtain vesicles of uniform size. As mentioned in the methods section, the samples for analysis were collected at regular intervals and were eventually run on a denaturing PAGE gel containing 8M urea. The denaturing gel analysis was employed to avoid occurrence of any extra bands potentially arising from the formation of RNA secondary structures such as primer dimers. Upon gel analysis, a band indicating the extension of the RNA primer appeared

in increasing intensity over multiple DH-RH cycles. However, this new band did not run at the same level as that of an intact 21-mer primer (Fig 4.3, right panel). A 21-mer standard RNA control was obtained by addition of a single nucleoside-5'-phosphoimidazolide monomer to the Amino G primer (the reaction was carried out as detailed in section 2.2.2.1, Chapter 2) and run on one of the lanes in the gel for comparison. The extended primer band from our reactions ran at a level that was in between that of the starting 20-mer RNA primer band, and the aforementioned control 21-mer RNA band (compare 'N+1 lane' and 'Cyc 7 lane' in Fig 4.4A). This observation held true when any one of the four RNA monomers were used in the starting reaction mixture (i.e. for 5'-NMPs, N=A/G/C/U). We suspect the generation of an abasic site is occurring in the extended primer product. This is supported by a previous study in which the abasic sites were generated during polymerization of nucleotide monomers under similar reaction conditions (Mungi and Rajamani, 2015). In a parallel study, we also analyzed template-directed information transfer for Hydroxyl G primer under the same reactions conditions. This primer is not activated at the 3'-terminal and ends in a canonical 5'-GMP. Very low amount of the extended primer product was observed for this primer over multiple DH-RH cycles (Fig 4.3, left panel).

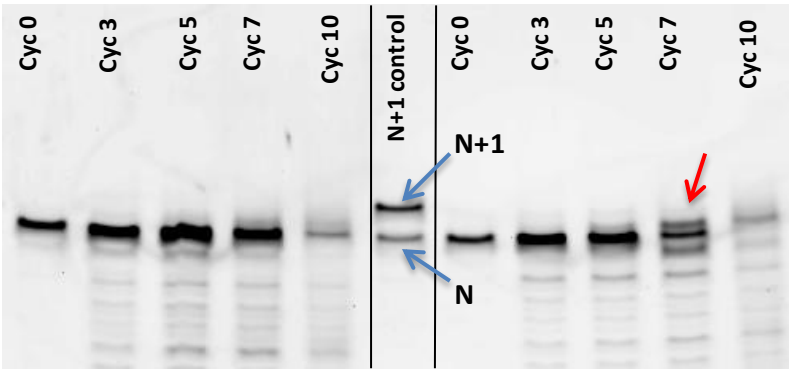


Figure 4.3: Extension of RNA primer ending with a canonical ribonucleotide (left panel) and a 3'-amino-2', 3'-dideoxynucleotide (right panel), over multiple cycles of DH-RH at 90˚C, in the presence of MisInc_U template, 5'-AMP (monomer) and lipid, using 1mM H_2SO_4 as rehydrating agent. The red arrow indicates the extended primer product, which is more apparent from Cyc 7 in the image on the right panel, as against the image on the left panel. 'N' indicates the 20-mer RNA primer, 'N+1' indicates intact extension of the primer by one nucleotide.

4.3.1.1 Role of lipids

To analyze the role of lipids in the above mentioned primer extension reactions, similar reactions in the absence of lipid molecules were also carried out. Higher degradation of the RNA primer was observed in the absence of the lipids (Fig 4.4B). This suggests a protective role stemming from the multilamellar layers that are formed by the lipids during the dehydration stage. Furthermore, another longer chain-length phospholipid, namely 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), was also tested to compare how it fared in comparison to DLPC. Similar results were observed in the POPC experiments too. The degradation of RNA primer, over multiple cycles of DH-RH, was somewhat lower in the presence of POPC. However, the extended primer product still did not run at a similar level as that of an intact 21-mer RNA, again most likely due to the loss of the nitrogeneous base during the extension of the primer. Since the stability of the RNA primer was found to be better in the presence of lipids under our reaction conditions of high temperature and low pH, DLPC was used in the starting reaction mixtures for the rest of the reactions that were carried out under similar conditions. In addition, since the primer Amino G was found to be more efficient in terms of the observed primer extension, all the reactions mentioned henceforth have been carried out using this primer.

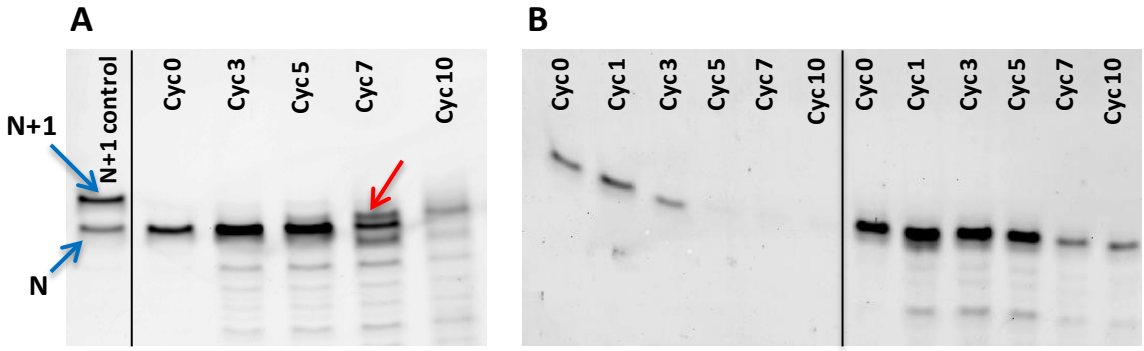


Figure 4.4: (A) Extension of the RNA primer using MisInc_U template and 5'-AMP as the monomer, in the presence of lipid, over repeated cycles of DH-RH. The reactions were carried out at 90°C with 1mM H_2SO_4 as the rehydrating agent. The red arrow indicates the extended product. In the control lane, 'N' indicates the 20-mer RNA primer while 'N+1' indicates extension of the primer by one nucleotide. (B) Stability of the RNA primer over multiple DH-RH cycles, when performed in the absence of lipid (left panel) and in the presence of lipid (right panel) in the starting reaction mixture. The black vertical lines in the above two panels have been used to demarcate two reaction sets that were run on the same gel (Panel b). In Panel a, the black line demarcates a reaction set from the control lane, both of which were analyzed on the same gel.

4.3.2 Effect of temperature and rehydration solution on primer extension

The glycosidic bond, which links the nitrogeneous base to the sugar phosphate moiety in a nucleotide, is known to be susceptible to hydrolysis under low pH and high temperature conditions (Rios et al., 2015). Since the template-directed primer extension reactions, using non-activated monomers under the aforementioned reaction conditions, resulted in an extended primer product with a possible abasic site, DH-RH cycles were performed at lower temperatures to see if this could mitigate the deglycosylation problem. The cycling reaction was carried out at a lower temperature of 80˚C and, subsequently at an even lower temperature of 50˚C. Since the reaction temperature was reduced, the time taken for the dehydration of the reaction mixtures actually increased. Especially for the reactions that were carried out at 50˚C, each dehydration phase lasted for a total of 90 minutes due to slower evaporation rates. Figure 4.5 shows the addition of 5'-AMP across the MisInc_U template, at different temperatures. Primer extension was indeed observed at both of these lower temperatures. However, upon gel

analysis, the extended primer product was still found to have been running at an intermediate level corresponding to a length in between that of a 20-mer and a 21-mer RNA.

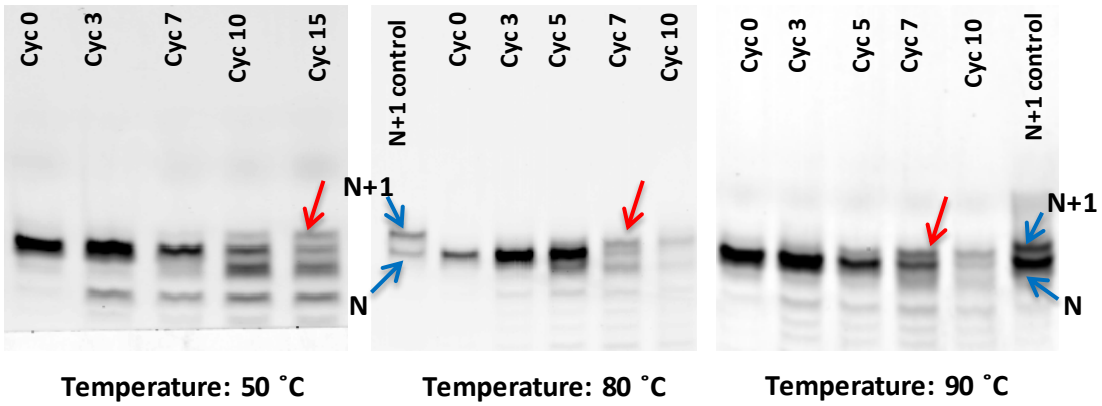


Figure 4.5: Extension of RNA primer in the presence of MisInc_U template, 5'-AMP (as monomer) and lipid, at different reaction temperatures. The red arrow indicates the extended primer product. 'N' indicates the 20-mer RNA primer and 'N+1' indicates intact extension of the primer by one nucleotide.

Next, the effect of different rehydrating agents on the efficiency of RNA primer extension under DH-RH cycles was characterized at 90˚C. As in the previous reactions, sulphuric acid was used as the rehydration solution but the strength of the acid used was varied. 0.05mM to 2mM acid was used to rehydrate the reaction mixtures. As shown in Figure 4.6, at lower strengths of acid no primer extension was observed. The optimum concentration for facilitating extension was found to be 1mM of acid, amongst the experimental variants that were tried. When higher concentrations of H_2SO_4 were used, the RNA broke down rapidly over the course of the DH-RH cycles. This might be due to excess amount of acid that accumulated over the cycles.

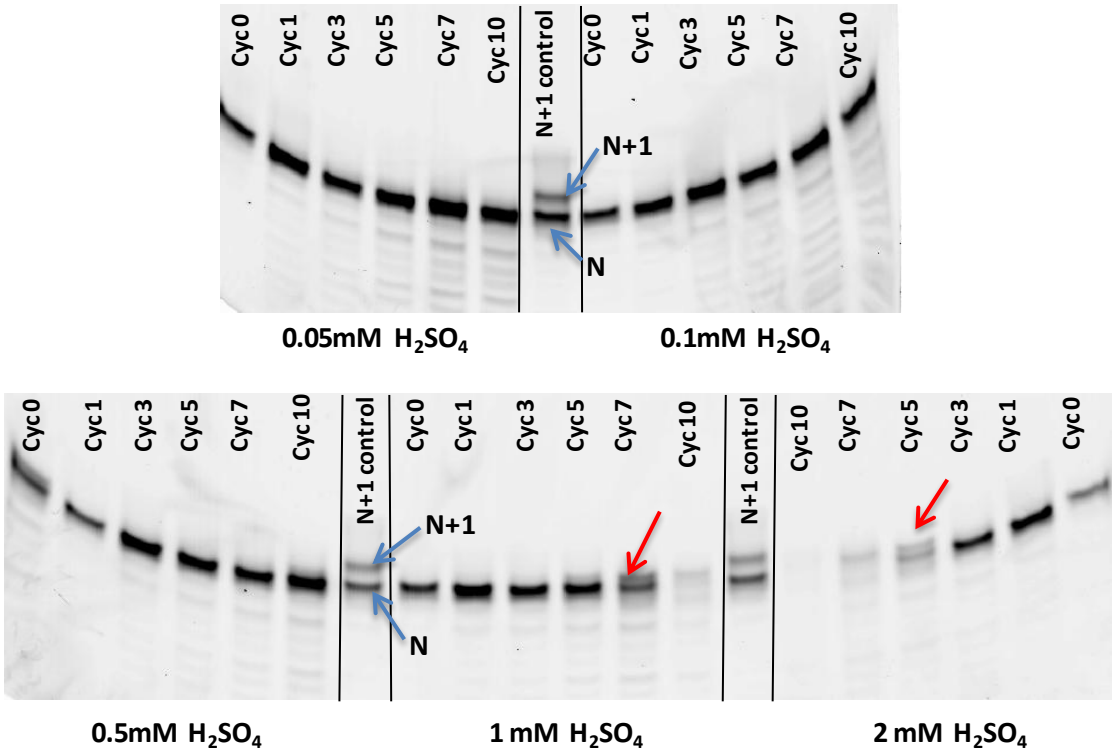


Figure 4.6: Extension of RNA primer over multiple cycles of DH-RH at 90˚C, in the presence of MisInc_U template, 5'-AMP (monomer) and lipid, using different concentrations of sulphuric acid as the rehydrating agent. The red arrow indicates the extended primer product. 'N' indicates the 20-mer RNA primer while 'N+1' indicates intact extension of the primer by one nucleotide. The black vertical lines on the gel images are used to demarcate different reactions and control lanes.

Subsequently, the acid that was used for rehydration was also varied. In total, four different mineral acids and two different organic acids were used as the rehydration solution. The concentration of the mineral acids was kept at 1mM. To compensate for lower acidity of the organic acids in general, 10mM of organic acids was used. Amongst the mineral acids that were used viz. HCl, HNO_3, H_3PO_4 and H_2SO_4, the primer extension was observed when sulphuric acid and phosphoric acid were used as the rehydrating agent (Fig 4.7, upper panel). Primer extension was observed when 10mM HCl was used instead of 1mM HCl. However in the presence of higher amounts of HCl, the RNA broke down very rapidly, thus decreasing the overall efficiency of the reaction. Amongst the organic acids used viz. HCOOH and CH_3COOH,

RNA primer extension was observed when formic acid was used for rehydration of the reaction mixture (Fig 4.7, upper panel). However, the extended primer continued to run slower than an intact N+1 (21-mer) product. No significant primer extension was observed for the other acids when used as rehydration agent (Fig 4.7, lower panel).

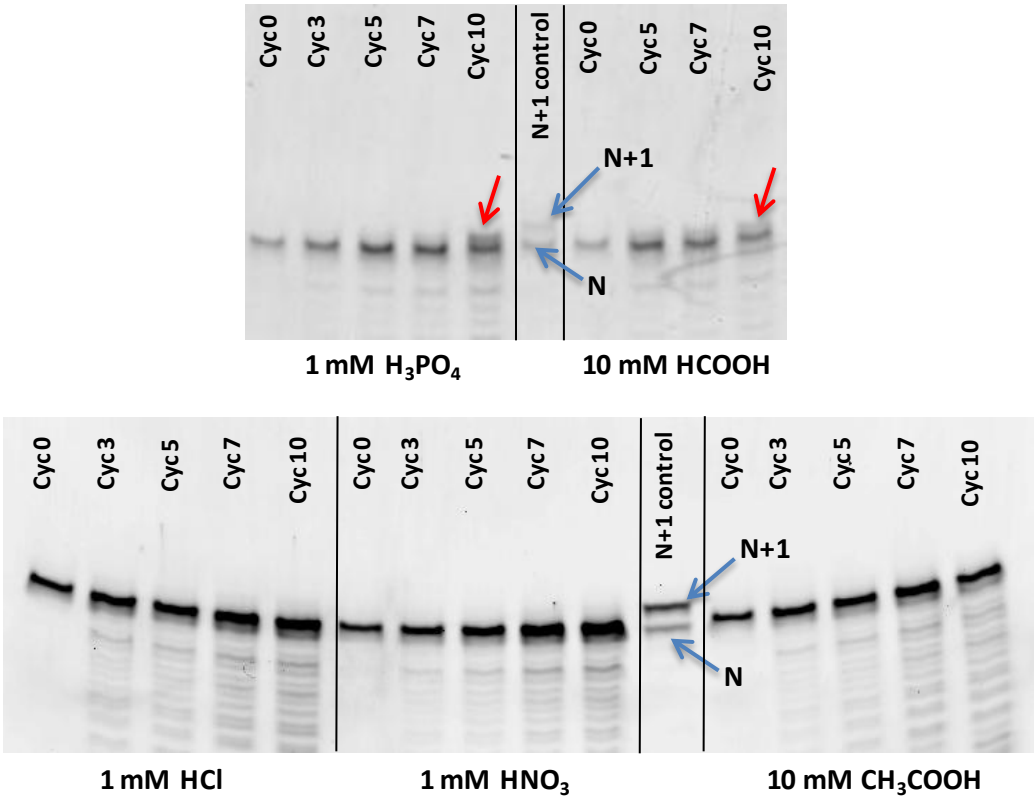


Figure 4.7: Extension of RNA primer in the presence of MisInc_U template, 5'-AMP as monomer and lipid, using different acids as rehydration solution. The red arrow indicates the extended primer product. 'N' indicates the 20-mer RNA primer and 'N+1' indicates intact extension of the primer by one nucleotide.

4.3.2.1 RNA primer extension at lower concentration of H_2SO_4 and monomer

Since the loss of base persisted during template-directed RNA primer extension under DH-RH cycles, despite varying the type and concentration of rehydration solutions, the concentration of the acid in the starting reaction mixture was reduced to see if this might prevent the potential generation of an abasic site. Towards this, sulphuric acid was added in the starting

reaction mixture at a final concentration of 1mM to bring down the pH of the reaction to around 3.5. Nanopure water was subsequently used to rehydrate the dehydrated reaction mixture after every DH-RH cycle. The primer extension was indeed observed even in these scenarios, however the yield of the extended primer product was lower when 1mM H_2SO_4 was used as the rehydrating agent (Table 4.1). Importantly, using this strategy, primer extension was demonstrated using as low as 1mM of nucleotides and lipid in the starting reaction mixture (Fig 4.8). However, upon denaturing gel analysis, the extended primer product continued to run somewhere in between the starting 20-mer primer and the known 21-mer extension product, even at this low effective concentration of the acid in the reaction mixture.

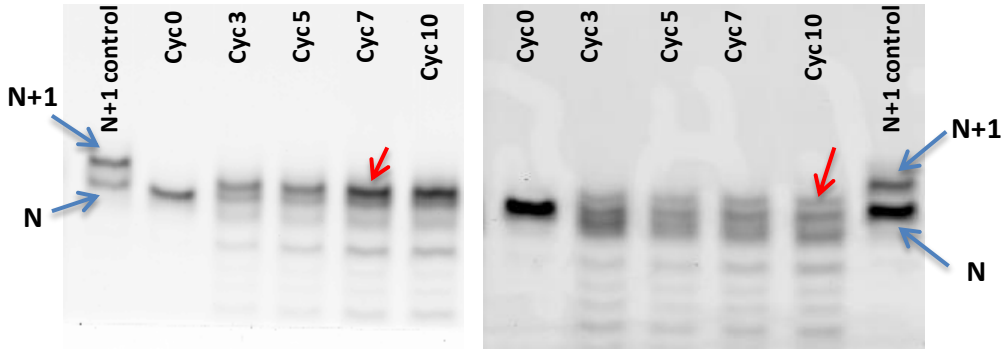


Figure 4.8: Extension of the RNA primer using 1mM 5'-AMP with 5mM lipid (left panel) and 1mM 5'-AMP with 1mM lipid (right panel), in the starting reaction mixture. Red arrows indicate the extended RNA primer product. 'N' indicates the 20-mer RNA primer while 'N+1' indicates intact extension of the primer by one nucleotide.

Table 4.1: Semi-quantitative yields of the extended primer product over multiple DH-RH cycles.

Concentration of 5'-AMP	Rehydrating solution	Yield of extended primer product after 7 DH-RH cycles	Yield of extended primer product after 10 DH-RH cycles
10mM	1mM H_2SO_4	8.5%± 1.6%	2.4%± 0.2%
1mM	Nanopure water	1.8%± 0.07%	1.4%± 0.07%

The reactions were carried out at 90˚C, using MisInc_U template, 5'-AMP as the nucleotide monomer, in the presence of 5mM lipid. The starting monomer concentration and the rehydrating agent were varied in order to vary the amount of acid in the reaction mixture. Yields are calculated with respect to the intact 20-mer primer band intensity (at cycle 0) from two independent experiments, and are represented as mean ± S.D. The reaction mixture with 1mM H_2SO_4 as rehydration agent had 60nmoles of H_2SO_4 at the end of 7th DH-RH cycle and 72nmoles of H_2SO_4 at the end of 10th DH-RH cycle. While, the reaction mixture with nanopure water as rehydration agent contained 10 nmoles of H_2SO_4 throughout all the DH-RH cycles.

4.3.3 Effect of monovalent ions

A study published in 2015 from David Deamer's group demonstrated the oligomerization reactions of non-activated nucleotides in salty environments (Da Silva et al., 2015). The researchers demonstrated the base pairing capacity of the resultant oligomers using ethidium bromide, which is an intercalating dye. They noticed that maximum yield for longer oligomers was obtained when ammonium chloride was used in the starting reaction mixture. Since an extended primer product with a potential abasic site was observed in all of the above-mentioned reactions that we had carried out, the presence of ammonium ions in the reaction mixture was evaluated to test if it might result in an intact extended primer product under DH-RH conditions. To analyze this, NH_4Cl was added to the reaction to a final concentration of 200mM. However, the presence of NH_4^+, either alone, or along with Na^+, did not improve the efficiency of the reaction in terms of the retention of the base on the extended RNA primer. The resultant product band continued to migrate in between that of a known 20-mer and 21-mer RNA on the denaturing gel (Fig 4.9). Interestingly, the use of ammonium ions also resulted in higher amounts of RNA degradation over the course of DH-RH, as opposed to when only

sodium ions were present in the reaction mix (Compare lane 'Cyc 7' in Fig 4.9 vs. lane 'Cyc 7' in Fig 4.4A).

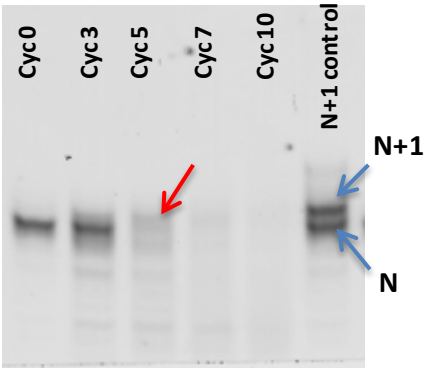


Figure 4.9: Extension of RNA primer, over multiple cycles of DH-RH, in the presence of MisInc_U template, 5'-AMP (monomer), lipid, and 200mM NH4Cl. The red arrow indicates the extended primer product, which is quite faint in this image. As in previous images, 'N' indicates the 20-mer RNA primer while 'N+1' indicates intact extension of the primer by one nucleotide.

4.3.4 Extension of RNA primer using ribose-5'-monophosphate (5'-rMP)

Since, in all of the aforementioned reactions, the product was an extended primer with a potential abasic site, RNA primer extension was carried out using 5'-rMP under the DH-RH set-up. Ribose-5'-monophosphate (5'-rMP) is a monomer containing a phosphate moiety and the ribose sugar, but lacks any nitrogenous base on the 1'-position of the sugar. The reaction was carried out using 5'-rMP instead of 5'-NMP as monomer, under DH-RH cycles at 90˚C, and 1mM H_2SO_4 was used as the rehydration agent. Upon denaturing gel analysis, long streaks of products were observed above the 20-mer primer band (Fig 4.10A, right panel). These extension products were observed starting from the first DH-RH cycle. Such primer extension was also observed in the absence of any template in the starting reaction mixture (Fig 4.10A, left panel). The smudge observed on the top of the primer band in the gel is potentially indicative of multiple additions to the RNA primer. Similar results were observed when 1mM of 5'-rMP was used in the reaction along with 1mM H_2SO_4, and nanopure water was used as the

rehydration solution. Surprisingly, similar primer extension product/s were also observed when the Hydroxyl G primer was instead of Amino G primer (Fig 4.10B).

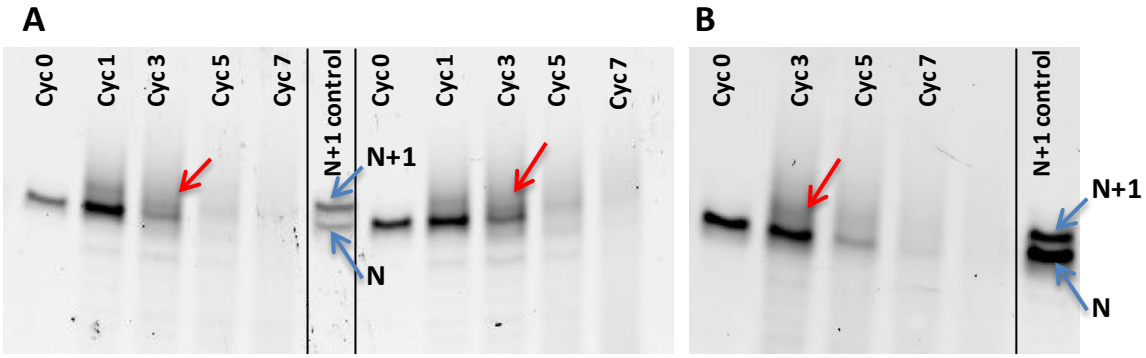


Figure 4.10: (A) Extension of the Amino-G RNA primer using 10mM of 5'-rMP, in the absence (left panel), and the presence (right panel) of the template in the reaction. (B) Extension of the Hydroxyl-G RNA primer using 10mM of 5'-rMP. Reactions were carried out at 90°C, in the presence of 5mM lipid and 1mM H_2SO_4 was used as rehydrating agent. Red arrows indicate extended RNA primer product/s. 'N' indicates the 20-mer RNA primer while 'N+1' indicates intact extension of the primer by one nucleotide.

In a parallel reaction, the 5'-rMP was cycled in the absence of the template-primer complex. No detectable fluorescent signal was observed upon scanning the gels using 532nm laser or even after staining the gel with SyBr gold dye. Significantly, the extension product obtained from the addition of a single 5'-rMP to the RNA primer was found to be running at a similar level to what was obtained when 5'-AMP was used as the monomer (Fig4.11). This gave an indirect confirmation about the loss of base that was occurring during the addition of 5'-NMP to a pre-existing RNA primer, under our experimental DH-RH reaction conditions.

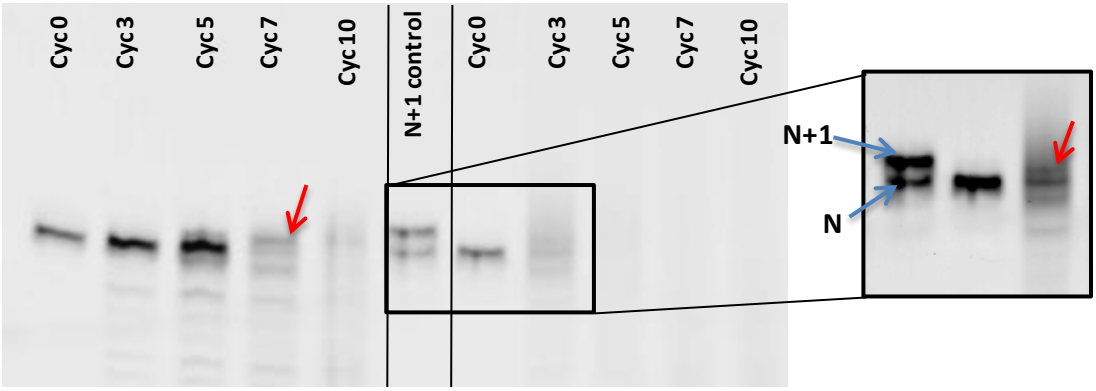


Figure 4.11: Extension of RNA primer over multiple cycles of DH-RH, in the presence of MisInc_U template, lipid, and 5'-AMP as monomer (left panel) or 5'-rMP as monomer (right panel). The inset shows the selected (boxed) part of the gel with the contrast enhanced for clearer visualization. The red arrow indicates the extended primer product. The extended primer products from both the reactions were observed to be running at the same level. The reactions were carried out at 90˚C, using 1mM H_2SO_4 as rehydrating agent. 'N' indicates the 20-mer RNA primer and 'N+1' indicates intact extension of the primer by one nucleotide.

4.3.5 Addition of 5'-rMP to 3'-NH_2-ddA

To further analyze the addition of 5'-rMP to the pre-existing RNA primer, mass analysis of the extended primer products was undertaken. However due to the low yield of the reaction in general, along with the degradation of the RNA template and primer at high temperature, it proved to be very difficult to get clean mass data. Hence, as a proxy, a reaction between a monomer analogous to the 3'-end of the RNA primer viz. 3'-NH_2-ddA, and 5'-rMP was carried out under similar DH-RH reaction conditions at low pH and high temperature (Fig 4.12A). The RNA-like monomer viz. 3'-NH_2-ddA was chosen to be without any phosphate group, to avoid any self-polymerization of this monomer. The products were purified from the unreacted reactants using ion-exchange chromatography. The purified peak (as indicated by red arrow in Fig 4.12B) was further subjected to mass analysis using LC-MS. This purified peak showed mass number corresponding to addition of one 5'-rMP to 3'-NH_2-ddA (Fig 4.12C, Table 4.2). This confirmed, albeit indirectly, the formation of a covalent bond between the incoming 5'-rMP monomer and the pre-existing RNA primer in reactions mentioned in section 4.3.4. Surprisingly,

we also observed a mass number corresponding to that of 5'-AMP (Table 4.2), hinting at the possibility of addition of a free adenine (breakdown product of 3'-NH_2-ddA) to the 5'-rMP, under our reaction conditions.

Table 4.2: Mass numbers observed in MS analysis of reaction between 5'-rMP and 3'-NH_2-ddA carried out at high temperature and low pH, under DH-RH conditions.

Chemical species	**Expected mass**	**Observed mass**
Adenine + H^+	136.0623	136.0635
5'-AMP + H^+	348.07	348.06
Dimer of 5'-rMP and 3'-NH_2ddA + H^+	463.1342	463.1343

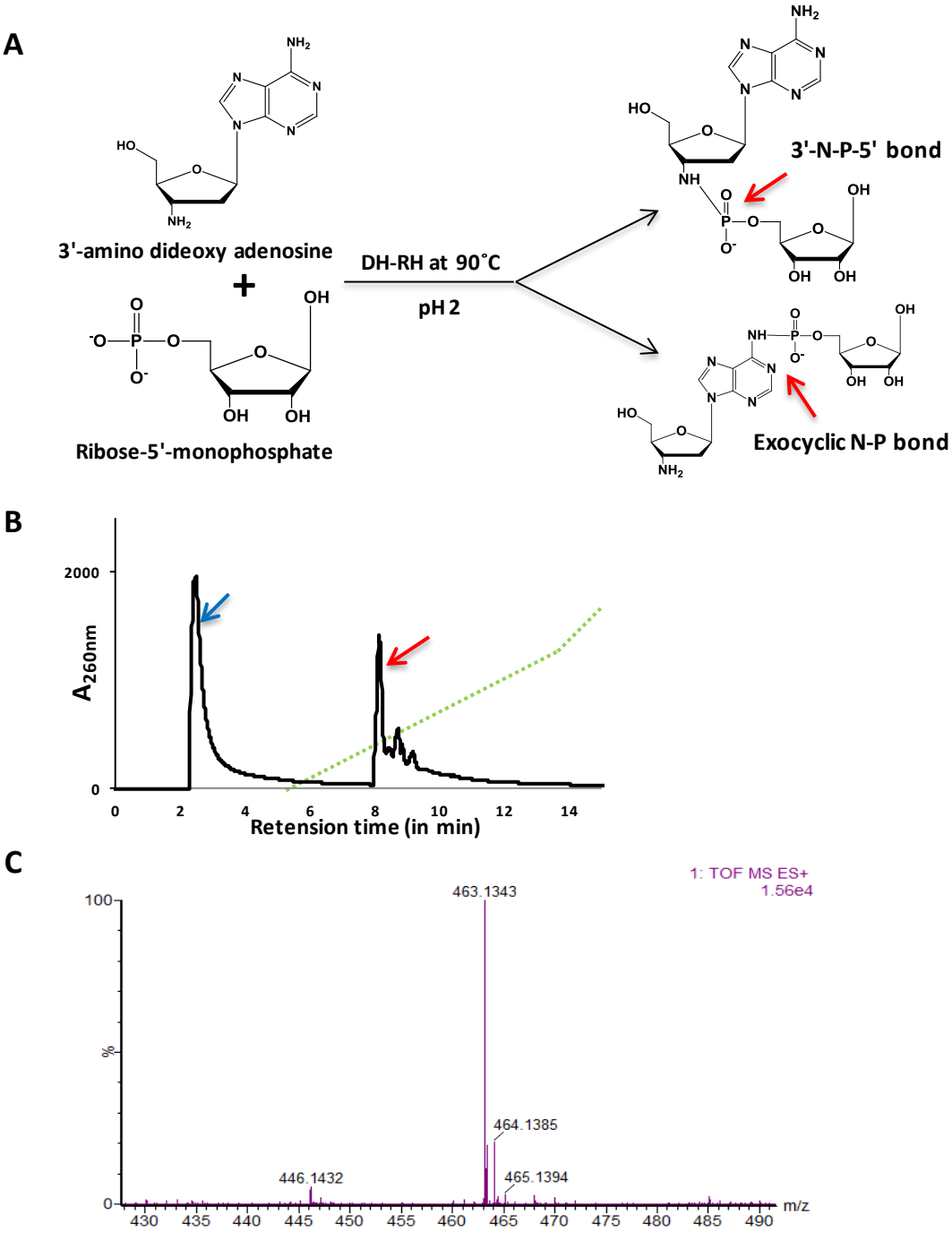


Figure 4.12: (A) Scheme for reaction between 5'-rMP and 3'-NH_2-ddA under DH-RH conditions. The two monomers can get linked between two alternate linkages as depicted (B) HPLC chromatogram of 'Cycle 5' from the reaction showed presence of a dimer product of 5'-rMP and 3'-NH_2-ddA, as indicated by red arrow. The blue arrow indicates the excess 3'-NH_2-ddA and the breakdown product (adenine nucleobase). The green trace indicates the gradient of salt that was used (C) Mass spectrum of the purified peak (indicated by red arrow in B).

4.4 Discussion

Phosphodiester bond formation between two nucleoside monophosphate monomers is a kinetically uphill reaction. Hence, chemically driven oligomerization and template-directed ligation/replication reactions, using non-activated nucleotides, do not occur spontaneously. Therefore, to understand the principles that underlie enzyme-free polymerization and copying reactions, prebiotic chemists have mostly used activated nucleotides. As elaborated in Chapter 1, activation chemistries, including but not limited to imidazole, 2-methylimidazole, oxyazabenzotriazole etc. (Fig 1.7, Chapter 1), have been developed and studied. They have contributed extensively to our current understanding of nonenzymatic reactions, especially in the context of emergence of functional polymers on the prebiotic Earth. However, in all of these reactions, a high concentration of the monomers has been used to achieve the products in detectable yields. The existence of such concentrated pools of activated monomers on the early Earth still remains to be unequivocally determined, as not many studies have conclusively demonstrated their prebiotically plausible syntheses. Furthermore, the stability of such activated monomers for prolonged time periods also remains to be addressed.

On the other hand, few studies to date have been carried out to demonstrate the polymerization of non-activated RNA monomers. To accelerate the process of phosphodiester bond formation in these reactions, the researchers used alternate dehydration-rehydration (DH-RH) cycles at elevated temperature(Rajamani et al., 2007; Mungi and Rajamani, 2015). These conditions are not only prebiotically pertinent but they also provide the requisite temperature and chemical fluxes that facilitate uphill polymerization reactions such as RNA oligomerization. Similar strategy has also been used to achieve oligomer formation from other non-nucleic acid monomers (Harshe et al., 2007; Mamajanov et al., 2014). The dehydration step involved in the DH-RH cycles help in concentrating the monomers, thus enabling phosphodiester bond formation (Deamer, 2012). The elevated temperature also reduces the water activity, thereby facilitating the bond formation by excluding water from the reaction mixture. The subsequent rehydration phase helps in redistribution of, both, the monomers and the resultant oligomers. The alternate DH-RH cycles help in trapping the polymers in a kinetic trap, wherein the rate of polymer formation exceeds the rate of polymer degradation, thus

ultimately resulting in oligomers that form from the monomers present in the starting reaction mixture. Furthermore, these reactions have been observed to work most efficiently at lower pH (Mamajanov et al., 2014; Mungi and Rajamani, 2015). A suggested mechanism, similar to that of acid-catalyzed esterification, hypothesizes the role of protonation of the nucleotide at lower pH, thereby facilitating the formation and accumulation of oligomers over multiple DH-RH cycles (Fig 4.13A)(Olasagasti et al., 2011). Presence of lipids in the DH-RH reactions was found to reduce the breakdown that occurs during these reactions, resulting in an increased yield of the oligomers over cycles (Mungi and Rajamani, 2015). Under dehydrated conditions, the lipids are known to form alternate hydrophilic and hydrophobic layers or liquid crystalline matrices. X-ray diffraction and neutron scattering studies have revealed that the nucleotide monomers get aligned in these alternate hydrophilic layers, increasing the chance for phosphodiester bond formation between the two monomers (Fig 4.13B)(Toppozini et al., 2013; Himbert et al., 2016). Specifically, in the neutron scattering studies, the peak at 3.4 A˚ was evident of such alignment as opposed to the peak observed at 4.6A˚, which signifies randomly distributed nucleotides. Furthermore, the oligomers that form in the dehydration phase can arbitrarily get encapsulated in the lipid vesicles during the subsequent rehydration phase. Such encapsulated nucleic acid polymers are reminiscent of the earliest cells or protocells, which are considered a crucial step during the origin of early life.

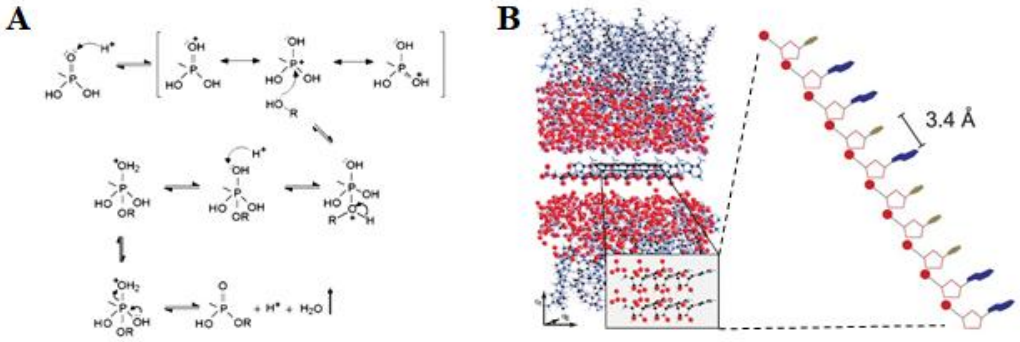


Figure 4.13: (A) Proposed mechanism for phosphodiester bond formation during lipid-assisted nonenzymatic polymerization (reproduced from Olasagasti et al., 2011).
(B) Ordered arrangement of monomers in dehydrated lipid bilayers (adapted from Toppozini et al., 2013, and Himbert et al., 2016).

The reactions discussed above, and other related reactions, would have resulted in nucleic acid polymers. However, faithful replication of these polymers under similar conditions would have been important for them to have acted as catalysts for multiple generations (e.g. ribozymes). To the best of our knowledge, only one relevant study to date has reported information transfer from a ssDNA template, under DH-RH reaction conditions, in the presence of phospholipids at low pH and at elevated temperature of 90˚C (Olasagasti et al., 2011). The authors estimated that the reaction yielded a total of 0.5% of the copied strands, with a copying accuracy of 91.1% as compared to the starting template concentration. However, no study thus far has reported the transfer of the information from a template RNA strand, using relevant RNA monomers, under prebiotically plausible DH-RH conditions. As discussed in Chapter 1, the emergence of DNA is thought to have been preceded by an RNA World due to RNA's capability of carrying out, both, catalytic and information transfer functions (Gilbert, 1986). Therefore, it becomes pertinent to understand RNA template copying under prebiotically relevant DH-RH conditions, specifically using prebiotically relevant non-activated nucleotides.

4.4.1 RNA template copying using non-activated monomers

To begin with, the extension of a pre-existing RNA primer (Amino-G) was analysed using contemporary 5'-NMPs in the starting reaction mixture. The RNA primer-template was pre-annealed like in the previous scenario. The starting template concentration was maintained twice as much as that of the primer to ensure efficient formation of primer-template complexes. The amino group modification on the 3'-end of the Amino-G primer ensures competent attack from this more efficient nucleophilic group (as against the canonical hydroxyl group) on the incoming monomer, yielding an extended primer product in detectable yields. Not surprisingly, when the Hydroxyl G primer, which terminates in a canonical RNA nucleotide, was used instead in the starting reaction mix, very low amounts of the extended primer product was visible upon denaturing-PAGE analysis (Fig 4.3). This is due to the lower ability of the hydroxyl group as compared to the amino group to nucleophilically attack the phosphate group of the incoming monomer. We observed the extension of the primer over multiple cycles of DH-RH, when the reaction was carried out at 90˚C, and when 10mM of the nucleotide and 5mM of

DLPC was used in the starting reaction mixture (Fig 4.4A). 1mM H_2SO_4 was used as the rehydration solution for these reactions, as comparable oligomerization reactions of these non-activated nucleotides have been reported to work optimally at low pH (Mungi and Rajamani, 2015). The lipid in the reaction mixture showed protective effect in terms of limiting the RNA primer degradation that is typically seen over multiple cycles of DH-RH, especially when carried out at low pH and high temperature. We observed lesser RNA primer degradation in the presence of lipids (Fig 4.4B), conceivably due to the seclusion of the RNA in the multilamellar liquid crystalline matrices that are formed by these lipids under dehydrated conditions (Fig 4.2A).

However, the extended primer product was observed to not run at the same level as that of an intact extension product like in a known 21-mer RNA. The intact 21-mer primer control was obtained by the single addition of an activated nucleotide (ImpN), to the Primer Amino G. In this reaction there is an assured formation of a phosphoramidite bond between the penultimate and the new (incoming) nucleotide, resulting in an intact 21-mer extended primer product. This is important to consider as even under our reaction conditions, the bond that would form between the incoming monomer and the 3'- end of the pre-existing RNA primer, would be a phosphoramidite bond (and not a phosphodiester bond) due to the amino modification on the 3'-end of the primer. This way, any differences in the charge on the 21-mer RNA control and the reaction products that might potentially arise due to the disparity in the kind of covalent bond (phosphoramidite vs. phosphodiester) are taken care of. The extended primer product from the reaction was observed to run somewhere in between that of the starting 20-mer RNA primer and the known 21-mer RNA control. A previous study from our lab, which aimed at characterizing the products obtained from oligomerization of 5'-AMP under similar DH-RH conditions, has demonstrated the presence of abasic site/s on the resultant oligomeric products (Mungi and Rajamani, 2015). Our results are the first of its kind to demonstrate that even the template-directed primer extension reactions are probably prone to the generation of abasic site on the extended primer product. The potential base stacking arising from the RNA template sequence, along with the otherwise stabilizing near-neighbour interactions, seem to play no significant role towards the prevention of potential base loss on

the incoming monomer, during the primer extension reactions. The 3'- end of the primer seems to be more susceptible for the nitrogenous base loss. The 5'-end of our RNA primer is tagged with Cy3 for enabling its detection upon gel analysis, and if this end of the primer happened to undergo degradation, it would not be possible to detect the RNA primer (and primer product/s) on the gel. To confirm this, we also subjected the known 21-mer RNA control to repeated cycles of DH-RH under similar reaction conditions. It was observed that the 21-mer primer degraded to a 20-mer oligomer, which corresponded to the starting primer length, over multiple DH-RH cycles (Fig 4.14).

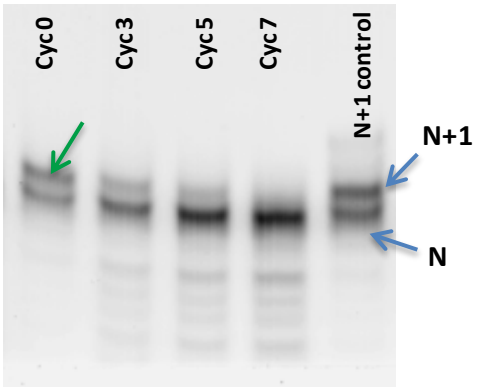


Figure 4.14: Degradation of the intact 21-mer primer (indicated by green arrow) over multiple cycles of DH-RH, in the presence of lipid. The reaction was carried out at 90˚C, using 1mM H_2SO_4 as the rehydrating agent. The 21-mer primer breaks down to the 20-mer primer over time, with no intermediate band observed between the N and N+1 bands. 'N' indicates the 20-mer RNA primer and 'N+1' indicates intact extension of the primer by one nucleotide.

The detection of the 20-mer band clearly indicates that the 5'-end of this primer was still intact. Importantly, no intermediate band was also observed in these reactions, which laid between the bands corresponding to a 20-mer (starting primer) and a 21-mer (extended product) length RNA. This strongly suggests that the loss of the base seen in our 5'-NMP reactions, potentially occurs just before or during the addition of the monomer to the pre-existing primer. It is very unlikely that the 20-mer primer gets extended to 21-mer primer and then the nitrogenous base on only the 3'-terminal nucleotide is selectively lost. Since, even the templating effect is unable

to prevent the generation of the abasic site, it might potentially result from a chemical reaction that takes place during the addition of the incoming nucleotide on to the primer. Collectively, from the above described observations, it is speculated that the extension of the primer and the loss of the base on the incoming monomer could potentially be coupled to each other.

4.4.2 Variation in the reaction parameters

To check whether the loss of base on the extended RNA primer product can be ameliorated, different reaction conditions, like temperature of the reaction, rehydration solution, nature of monovalent cations etc, were systematically varied. To begin with, the reaction temperature was lowered, as the glycosidic bond in the nucleotides is known to be susceptible to breakage at high temperature (Rios et al., 2015). The DH-RH reactions were carried out at temperatures of 80˚C and 50˚C, respectively. The RNA degradation was observed to have reduced at the lower temperature (50˚C) than when the reaction was performed at the higher temperature (80˚C). The extension of the primer at these low temperatures, using 1mM H_2SO_4 as the rehydration solution, was indeed observed. However, the apparent loss of base persisted at temperature even as low as 50˚C (Fig 4.5).

Subsequently, the nature of the rehydration solution was varied in order to assess its effect on the potential loss of base during primer extension, under our reaction conditions. Firstly, the concentration of sulphuric acid that was used as the rehydration solution was altered. Amongst the different variations tested from 0.5mM to up to 2mM of acid, 1mM of H_2SO_4 was found to be the optimum rehydration concentration, in terms of the extended primer product yield, as visualized on the gel. Lower concentrations of the acid probably were insufficient to bring down the pH to the desired level, thereby resulting in no detectable primer extension. On the other hand, repetitive addition of higher concentration of sulphuric acid resulted in excessive degradation of the RNA primer, such that no significant amount of the starting 20-mer primer was available after first few DH-RH cycles. This again underlines the importance of maintaining the optimum pH for achieving the addition of RNA monomer to the pre-existing RNA primer,

under DH-RH conditions. This also indicates the facilitation of an acid-catalysed esterification type reaction during the primer extension, as suggested earlier (Fig 4.13A).

Additionally, instead of sulphuric acid, different mineral and organic acids were also tested as rehydrating agents. Amongst the mineral acids used, primer extension was observed only when either sulphuric acid or phosphoric acid was used to repeatedly rehydrate the reaction mixture. This might be attributed to the fact that other two acids viz. HCl and HNO_3 are thermally labile and the temperature during the dehydration phase for these reactions was maintained at 90˚C. It is also interesting to notice that H_2SO_4 and H_3PO_4 have better dehydration capabilities than HCl and HNO_3. Typically, 1mM of mineral acid and 10mM of the organic acid was used as rehydration agent. The higher concentration of the organic acid was selected to compensate for the lower strength of the organic acids in general. Among the two organic acids evaluated, detectable RNA primer extension was visible only when formic acid was used. This can be attributed to higher acidic strength of formic acid than that of acetic acid, which might enable bringing down the pH of the reaction mixture more effectively (in HCOOH) over the cycles of DH-RH. Significantly, despite varying the nature of the acid used for rehydration, the extended primer product continued to run in between the level of the starting 20-mer primer and that of a 21-mer RNA control, on the denaturing gel. This again points at the possibility of loss of base on the incoming monomer, during RNA primer extension at low pH, in the DH-RH reaction regimen.

Furthermore, the nature of the monovalent ion that was added to the reaction mixture was altered to analyze whether this change might prevent the generation of the potential abasic site on the extended primer product. The role of ammonium ions in this context was specifically checked as a previous study had demonstrated the formation of oligomers from non-activated RNA monomers in salty environment, under DH-RH regimes (Da Silva et al., 2015). Among the different salts used in their study (viz. NH_4Cl, NaCl, LiCl and KCl), ammonium chloride was the most efficient salt that seemed to promote oligomer formation when 1:1 mixture of 5'-AMP and 5'-UMP were cycled under DH-RH reaction conditions at 85˚C. The researchers also claimed to have observed oligomers in their reaction that were capable of base pairing, as was evident

from EtBr staining of the reaction products. However in our reaction set-up, use of NH_4^+ alone, or along with Na^+, did not prevent the apparent loss of base during the primer extension reactions (Fig 4.9).

4.4.3 RNA primer extension using sugar-phosphate monomer

Despite varying multiple parameters that potentially influenced the reaction, the apparent loss of the base on the extended RNA primer product persisted. Therefore, whether the RNA primer could get extended with just the sugar-phosphate unit was analyzed. A previous study from our lab has demonstrated the polymerization of this monomer under similar DH-RH reaction regimen (Mungi and Rajamani, 2015).Towards this end, 5'-rMP, which consists of only the ribose sugar and a phosphate (but lacks any informational moiety), was used as the monomer in the starting reaction mixture. Upon analysing the reaction products on a denaturing gel, long streaks of products were observed above the starting 20-mer RNA primer band. These resultant products were observed at the end of the first DH-RH cycle itself (Fig 4.10A 'Cyc 1' lane). This indicated a higher polymerization efficiency for 5'-rMP than the 5'-ribonucleotides (discussed in detail in result sections 4.3.1 to 4.3.3). The long streaks that resulted from the products formed suggested multiple additions of the sugar-phosphate monomers to the RNA primer. These 5'-rMP are known to form multiple linear and cyclic oligomers under these DH-RH conditions of low pH and high temperature, as was reported in previous work from our lab (Mungi and Rajamani, 2015). It is suspected that in our reaction set-up, the pre-existing RNA primer possibly is getting extended by a mixture of linear and cyclic products, resulting in poor resolution of the product oligomers during gel analysis.

Furthermore, we wanted to confirm that the streaks of the products observed on top of the RNA primer was due to a covalent linkage between 5'-rMP and the 3' terminal of the primer used in the reaction, and not because of oligomerization that resulted from 5'-rMP alone. To address this, 5'- rMP molecules were cycled alone without the primer-template complex and analysed using gel electrophoresis. No fluorescence signals were obtained upon scanning the gel at 532nm wavelength (at which the signals from ‘Cy3’ are usually obtained). No products

were observed upon staining the gel with SyBr gold as well, due to lack of any intercalating bases on 5'-rMP oligomers. This indirectly confirmed that the 5'-rMP monomer and/or oligomers were covalently linked to the pre-existing RNA primer in the reaction mixture, and thus resulted in the poorly resolved smudge of products observed on the gel. Surprisingly similar streaks of products were also observed when using Hydroxyl G primer (Fig 4.10B). This again might be an indicative of better polymerization competence of the sugar-phosphate monomers than contemporary nucleotides.

To further confirm the covalent linkage between the 5'- rMP monomers/polymers and the pre-existing RNA primer, mass analysis was undertaken. However, despite multiple efforts, clean mass data could not be obtained for the extended primer products. This was mainly due to the low yield of the reaction products. This limitation was further compounded by other factors including the degradation of RNA over multiple DH-RH cycles, multiple peaks observed in MALDI-TOF analysis due to the inherent complexity of the unpurified reaction material, and considerable loss of the products during gel purification to obtain the purified products. Therefore, to circumvent this problem, a proxy reaction was set up with a nucleotide analogous to the 3'-end of the RNA primer viz. 3'-NH_2-ddA, and 5'-rMP, under DH-RH reaction conditions. As mentioned above, previous study from our lab has shown generation of oligomers with basic sites when 5'-AMP was subjected to similar reaction conditions. Addition of single or multiple 5'-rMP molecules to the nucleotide monomer would, in principle, result in similar polymers containing abasic site/s. Hence, 3'-NH_2-ddA was chosen to have no phosphate group, so as to avoid its self-polymerization, and generation of possibly similar reactions products as those that might be obtained from addition of 5'-rMP to this nucleotide monomer (i.e. addition of 5'-rMP to 3'-NH_2-ddA-5'-monophosphate). In other words, this strategy ensured that the resultant products were only due to the addition of 5'-rMP to 3'-NH_2-ddA. The reaction product was purified using ion exchange HPLC (indicated by red arrow in Fig 4.12B), for further mass analysis. Since this product eluted out in the salt-gradient on the ion-exchange column, it would have a phosphate group. Additionally, since it was detected using DAD detector (absorbance: 260 nm), it would also have a nucleobase attached to it. Therefore, it was suspected that the purified fraction contained a product with addition of a single 5'-rMP to a 3'-NH_2-ddA molecule.

The mass analysis confirmed this suspicion (Table 4.2), as the exact mass for a dimer of 5'-rMP and 3'-NH_2ddA was obtained. This dimer is suspected to have two prominent covalent linkages as indicated in Fig 4.12A. The nature of the covalent linkage needs further confirmation using NMR analysis. To our surprise, a mass number corresponding to that of 5'-AMP was also obtained (Table 4.2). This hints at the addition of adenine nucleobase (a possible breakdown product from 3'-NH_2-ddA, indicated by blue arrow in Fig 4.12B), to the pre-existing 5'-rMP molecule under the DH-RH conditions. Presence of at least two isomers, namely the canonical 5'-AMP and N6-ribosylated isomer of the AMP, is suspected for this observed mass number. It is of our interest to pursue this reaction further in order to find out the exact chemical structure of the 5'AMP (like) molecule obtained.

Under our reaction regimen, addition of sugar-phosphate monomers to the RNA resulted in the formation of hybrid oligomers. Such hybrid oligomers have the potential to sample multiple nitrogenous bases/informational moieties. They have been hypothesized to have played an important role in exploring a vast chemical space of nitrogenous bases that might have populated a pre-RNA World, ultimately giving rise to modern RNA from pre-RNA informational polymers, in a process facilitated by molecular evolution (Fig 1.2B, Chapter1) (Hud et al., 2013). This transition from pre-RNA to RNA molecules has gained support from the recent experimental demonstrations of addition of non-canonical nucleotides, to pre-existing ribose or ribose monophosphate molecules. A nucleobase analogue, namely 6-aminouracil, was shown to form the corresponding nucleoside under dry heating conditions (Kim and Benner, 2015). 2,4,6-Triaminopyrimidine (TAP) is yet another nucleobase analogue capable of forming TAP-ribose conjugates under prebiotically plausible conditions (Chen et al., 2014). A previous study from our lab has also reported formation of a nucleotide by the incorporation of barbituric acid (BA) to 5'-rMP, under high temperature DH-RH regimens (Mungi et al., 2016).

4.4.4 RNA primer extension using lower monomer concentration

In order to try and retain the informational moiety on the extended primer, we also used lower amounts of acid in the starting reaction mixture. This was done as the glycosidic bond in the

nucleotide monomers is known to be susceptible to hydrolysis at higher temperatures and low pH. However, maintaining the pH in the acidic range seemed important for addition of the monomer to the pre-existing monomer/oligomers (results from Mungi and Rajamani, 2015, and from this study, respectively). Therefore, we added sulphuric acid to the starting reaction mixture to a final concentration of only 1mM and then used nanopure water for rehydration. Since the sodium salt of the monomer has been used in the reactions (the pH for this monomer stock solution is near 8), its amount was reduced to up to 1mM in order to maintain a low enough pH. Surprisingly, primer extension was observed even at this low concentration of monomer. This was true for both the nucleoside-5'-monophosphate and ribose-5'-monophosphate, when they were used as the starting monomer in the respective reaction mixtures. We believe that this result is of importance as the existence of highly concentrated pools of reactants on the prebiotic Earth remains questionable. Additionally, the yields of most prebiotically pertinent reactions are usually really low, which might result in availability of very low amounts of the reactants at any given time. Therefore, it is really crucial to analyze prebiotically pertinent reactions using lower “realistic” concentrations of the starting chemicals, as has been done in this study.

4.5 Conclusion

Certain niches like acidic geothermal pools, present at the edges of volcanic sites, have argued to be the most plausible site for the origin of life on prebiotic Earth. In this study, the feasibility of information transfer process, which is essential for the emergence of functional nucleic acids, was assessed in such environments, under prebiotically pertinent DH-RH reaction regimes. Our studies reveal that, although the RNA primer could get extended using non-activated 5'-NMPs as monomers, the information moiety is lost under these reaction conditions of high temperature and low pH. Presence of lipid in reaction mixture plays a protective role as far as RNA degradation is considered. The loss of base can, however, not be prevented even after changing pertinent reaction parameters such as temperature, rehydration solution and the nature of monovalent cation. This strongly hints at a possible loss of nucleobase, being coupled to the RNA primer extension, in these reaction regimens. In general, this set of experiments

highlight the importance of systematically characterizing prebiotically pertinent reactions in simulated laboratory conditions, to better understand how they might advent in niches that are thought to have supported the origin of life. Particularly, these findings have important implications for discerning relevant mechanisms in such niches that would have eventually enabled the transition from chemistry to biology on the early Earth. Furthermore, the formation of hybrid polymers in our 5'-rMP based reactions, strengthens the likelihood of pre-RNA World hypothesis. Also, the possibility that the same sugar phosphate backbone can potentially sample both contemporary as well as alternate nucleobases, is indicative of the important role such mixed backbones might have played in the transition from a putative pre-RNA World to an RNA World.

4.6 References

Bapat, N. V, and Rajamani, S. (2015). Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids. Journal of Molecular Evolution 81, 72–80.

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. Journal of the American Chemical Society 136, 5640–5646.

Deamer, D. (2012). Liquid Crystalline Nanostructures : Organizing Matrices for Non-enzymatic Nucleic Acid Polymerization. Chem Soc Rev 41, 5375–5379.

Deck, C., Jauker, M., and Richert, C. (2011). Efficient Enzyme-free Copying of all Four Nucleobases Templated by Immobilized RNA. Nature Chemistry 3, 603–608.

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

Gilbert, W. (1986). The RNA world. Nature 319, 618.

Harshe, Y.M., Storti, G., Morbidelli, M., Gelosa, S., and Moscatelli, D. (2007). Polycondensation Kinetics of Lactic Acid. Macromolecular Reaction Engineering 1, 611–621.

Himbert, S., Chapman, M., Deamer, D.W., and Rheinstädter, M.C. (2016). Organization of Nucleotides in Different Environments and the Formation of Pre-Polymers. Scientific Reports 6, 31285.

Hud, N. V, Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The Origin of RNA and "My Grandfather 's Axe ". Chemistry & Biology 20, 466–474.

Kervio, E., Claasen, B., Steiner, U.E., and Richert, C. (2014). The Strength of the Template Effect Attracting Nucleotides to Naked DNA. Nucleic Acids Researchclei 42, 7409–7420.

Kim, H.-J., and Benner, S.A. (2015). Prebiotic Glycosylation of Uracil with Electron-Donating Substituents. Astrobiology 15, 301–306.

Leu, K., Kervio, E., Obermayer, B., Turk-macleod, R.M., Yuan, C., Luevano, J., Chen, E., Gerland, U., Richert, C., and Chen, I.A. (2013). Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. Journal of the American Chemical Society 135, 354–366.

Li, L., Prywes, N., Tam, C.P., Flaherty, D.K.O., Lelyveld, V.S., Izgu, E.C., Pal, A., and Szostak, J.W. (2017). Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. Journal of the American Chemical Society 139, 1810–1813.

Mamajanov, I., Macdonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Fernandez, 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.

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.

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. Physical Chemistry Chemical Physics 18, 20144–20152.

Olasagasti, F., Kim, H.J., Pourmand, N., and Deamer, D. (2011). Non-enzymatic Transfer of Sequence Information under Plausible Prebiotic Conditions. Biochimie 93, 556–561.

Orgel, L.E. (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 39, 99–123.

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., and Benner, S. (2007). Lipid-assisted Synthesis of RNA-like Polymers from Mononucleotides. Orig Life Evol Biosph.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., and Chen, I.A. (2010). Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. Journal of the American Chemical Society 132, 1008–1011.

Rios, A.C., Yu, H.T., and Tor, Y. (2015). Hydrolytic Fitness of N-glycosyl Bonds : Comparing the Deglycosylation Kinetics of Modified , Alternative , and Native Nucleosides. Journal of Physical Organic Chemistry 28, 173–180.

Da Silva, L., Maurel, M., and Deamer, D. (2015). Salt-Promoted Synthesis of RNA-like Molecules in Simulated Hydrothermal Conditions. Journal of Molecular Evolution 80, 86–97.

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.

Yi, R., Hongo, Y., and Fahrenbach, A.C. (2018). Synthesis of Imidazole-activated Ribonucleotides using Cyanogen Chloride. Chemical Communications 54, 511–514.

Chapter 5

Summary

Summary

From time immemorial, the formation of first living entity from a mixture of chemicals has intrigued scientists. Researchers have put forth different theories for drawing out the potential path for this non-trivial transition from 'nonliving' chemistry to 'living' biology. Since the presence of genetic material is deemed indispensable for living beings, most of the 'origin of life' related research has been focused on understanding the formation and evolution of functional genetic polymer(s). 'RNA' has been the most favored candidate in this regard, due to its capability to store information and its potential to act as a catalyst (Robertson and Joyce, 2010). This widely acknowledged 'RNA World' hypothesis gains considerable support from the presence of modern day evidences, such as RNA viruses (wherein RNA acts a genetic material) and self-splicing introns (wherein RNA acts as a catalyst). The 'smoking-gun' evidence, however, is the presence of rRNA in the core of the extant ribosome, which indicates a key role of RNA in the catalysis of peptide-bond formation. Additionally, the availability of experimental models to demonstrate nonenzymatic construction of RNA polymers has further fueled research in understanding RNA-based early life.

Substantial amount of experimental work in the past 50 years has explored enzyme-free formation of nucleic acid monomers, and their subsequent polymerization and replication (Orgel, 2004; Robertson and Joyce, 2010; Wachowius et al., 2017; Kitadai and Maruyama, 2018). The fidelity of these enzyme-free reactions has also been characterized to some extent (Orgel, 2004; Rajamani et al., 2010; Leu et al., 2013). However, during our literature survey, we noticed that almost all of these prebiotically pertinent reactions were carried out without considering the inherent heterogeneity of the prebiotic milieu. The prebiotic soup must have contained different kinds of chemicals as a result of a plethora of products and by-products obtained from relevant synthesis pathways like FTT synthesis, formose reaction etc. The exogenous delivery of mixtures of different chemical entities would have further compounded the complexity of the prebiotic milieu. Hence, it is imperative to analyze relevant prebiotic processes in the presence of background molecules. The parallel in extant biology includes studies wherein the effect of co-solutes and molecular crowding, on various biochemical processes, has been well documented (Nakano et al., 2014; Rivas and Minton, 2016). Therefore,

for the present investigation, the effect of prebiotic heterogeneity on nonenzymatic reactions was systematically analyzed. We selected the 'enzyme-free template-directed RNA primer extension' system as a model prebiotic reaction for this analysis.

To begin with, the complexity of the reaction mixture was increased in a step-wise manner, using DLPC vesicles and PEG as co-solutes. The lipid vesicles were chosen as a co-solute due to their relevance in the formation of protocellular boundary structures (Szostak et al., 2001). On the other hand, PEG is routinely used as a molecular crowding agent in many biochemical analyses. Its use also served as a proxy for the presence of nonspecific polymers that might have resulted from the polymerization of other non-RNA monomers that would have also been present in the milieu (Mamajanov et al., 2014; Mungi and Rajamani, 2015). The rate and accuracy of nonenzymatic template-directed RNA replication was analyzed in the presence of either the individual or the admixture of these co-solutes, using imidazole-activated monomers. We found that the rate of purine based matched addition reactions decreased in the presence of individual co-solutes. This effect was found to get enhanced in reactions where the admixture of both these co-solutes was used. On the other hand, no significant change in the rate of addition was noted for the other two matched additions, along with all the mismatched additions, in the presence of either the individual co-solute or in the admixture of lipid and PEG.

We envisaged that greater π-stacking ability of the purine monomers might underlie these observed effects. In order to analyze this further, we estimated the diffusion constants and the size of the molecular cluster/s of nucleotide monomers using NMR. By estimating DOSY NMR and T_1 relaxation constants for different concentrations of monomers in the absence and presence of the PEG, greater stacking tendency for purine monomers, especially in the presence of PEG, was confirmed (as opposed to what was seen when pyrimidine monomers were used). Elevated stacking of purine nucleotides in the presence of PEG would mean the sequestration of the free monomers into these pseudo-oligomeric entities, and thus lesser availability of these monomers for primer extension in the reaction milieu. Observations from our NMR experiments explain in part, the reason behind the differential decrease seen in the rate of purine based matched additions under our reaction setting.

Due to a decrease in the rate of two out of four of the fast matched addition reactions, the overall fidelity of nonenzymatic RNA replication was also found to have decreased under our experimental conditions. The mutation rate was found to be 19.8 ± 0.5% in control condition, which increased to 26.6 ± 4.4% in the presence of both the co-solutes. The frequency of incorrect addition of 'G across U' considerably increased, resulting in almost 20% rise in the mutation rate for the 'U' template base in the presence of DLPC and PEG (31 ± 4.6 % in the control condition vs. 51.7 ± 3.5 % in the presence of both the co-solutes). We further analyzed the effect of presence of co-solutes on enzyme-free RNA replication for several replication cycles, using *in silico* studies which were parameterized based on experimentally generated reaction rates. The simulation studies revealed strong GC-bias in the daughter sequences, which might stem from the inherent low fidelity of copying 'A' and 'U' template bases (Orgel, 2004). Greater sequence space exploration was observed in the case of RNA template-directed polymerization that occurred in the presence of co-solutes, as was interpreted from the generation of higher number of new sequences under these conditions. Taken together, presence of background molecules was found to directly impact the propagation of genetic information in our prebiotically pertinent experimental conditions.

Going forward, we also wanted to analyze the feasibility of template-directed RNA primer extension, using prebiotically realistic non-activated monomers. These experiments were carried out in the presence of lipids as co-solutes, and under a reaction regimen of alternate DH-RH cycles at low pH and high temperature, to promote the bond formation between the incoming monomer and the RNA primer. Although RNA primer extension was observed under these conditions, the extended primer product contained a plausible abasic site. The loss of base on the extended RNA primer persisted despite changing various parameters including the temperature of the reaction, the nature and amount of acid used for rehydration, and the identity of the monovalent cation used in the reaction mixtures. We hypothesize that the loss of base is potentially coupled to the chemical addition of the incoming monomer on to the pre-existing RNA oligomer.

Since the extended primer product contained a plausible abasic site while using 5'-NMPs as monomers, we also went ahead and investigated the extension of the RNA primer using ribose-5'-monophosphate (5'-rMP; this monomer lacks the nitrogenous base). Long streaks resulting from, possibly, multiple additions of 5'-rMP to the RNA primer were observed upon gel analysis. Such hybrid oligomers of RNA and sugar-phosphate backbones is argued to have played a crucial role in sampling different information carrying moieties, during the early stages of emergence of the informational polymer (Hud et al., 2013). Significantly, the extension product obtained from addition of a single 5'-rMP monomer was also found to be running at a similar level as the band that was obtained when 5'-AMP was used as monomer. This indirectly confirmed the loss of base during the addition of 5'-NMP to a pre-existing RNA primer.

In conclusion, the investigations detailed in this thesis highlight the importance of understanding the progress of prebiotic reactions under realistic prebiotic niche conditions. Our studies reveal that nonenzymatic reactions that are carried out in buffered conditions, using high concentrations of the starting reactants, might not result in a similar set of product/s when subjected to prebiotically realistic scenarios. This is an important consideration as studies like ours, will allow for a better understanding of the mechanisms underlying the transition from chemistry to biology on early Earth. Importantly, it has crucial implications for helping in the reconstruction of the 'origin of life' in a test-tube, using the bottom up approach.

List of publications

Bapat N.V, Outlining 'Abiosignatures': Clues from prebiotic chemistry. *In review*.

Bapat, N. V, and Rajamani, S. (2018). Templated Replication (or Lack Thereof) under Prebiotically Pertinent Conditions. Scientific Reports, 8(1), 15032. doi:10.1038/s41598-018-33157-9

Bapat, N. V, and Rajamani, S. (2015). Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids. Journal of Molecular Evolution 81, 72–80. doi: 10.1007/s00239-015-9700-1

References

Hud, N. V, Cafferty, B.J., Krishnamurthy, R., and Williams, L.D. (2013). The Origin of RNA and "My Grandfather 's Axe ". Chemistry & Biology 20, 466–474.

Kitadai, N., and Maruyama, S. (2018). Origins of Building Blocks of Life: A Review. Geoscience Frontiers 9, 1117–1153.

Leu, K., Kervio, E., Obermayer, B., Turk-macleod, R.M., Yuan, C., Luevano, J., Chen, E., Gerland, U., Richert, C., and Chen, I.A. (2013). Cascade of Reduced Speed and Accuracy after Errors in Enzyme-Free Copying of Nucleic Acid Sequences. Journal of the American Chemical Society 135, 354–366.

Mamajanov, I., Macdonald, P.J., Ying, J., Duncanson, D.M., Dowdy, G.R., Walker, C.A., Engelhart, A.E., Fernandez, 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.

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.

Nakano, S., Miyoshi, D., and Sugimoto, N. (2014). Effects of Molecular Crowding on the Structures , Interactions , and Functions of Nucleic Acids. Chemical Reviews 114, 2733–2758.

Orgel, L.E. (2004). Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 39, 99–123.

Rajamani, S., Ichida, J.K., Antal, T., Treco, D.A., Leu, K., Nowak, M.A., Szostak, J.W., and Chen, I.A. (2010). Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. Journal of the American Chemical Society 132, 1008–1011.

Rivas, G., and Minton, A.P. (2016). Macromolecular Crowding In Vitro, In Vivo, and In Between. Trends in Biochemical Sciences 41, 970–981.

Robertson, M.P., and Joyce, G.F. (2010). The Origins of the RNA World. Cold Spring Harbor Perspective in Biology 4, a003608.

Szostak, J.W., Bartel, D.P., and Luisi, P.L. (2001). Synthesizing Life. Nature 409, 387–390.

Wachowius, F., Attwater, J., and Holliger, P. (2017). Nucleic Acids: Function and Potential for Abiogenesis. Quarterly Reviews of Biophysics 50, E4.

Author's publications

J Mol Evol (2015) 81:72–80
DOI 10.1007/s00239-015-9700-1

ORIGINAL ARTICLE

Effect of Co-solutes on Template-Directed Nonenzymatic Replication of Nucleic Acids

Niraja V. Bapat[1] · **Sudha Rajamani**[1]

Received: 26 May 2015 / Accepted: 23 September 2015 / Published online: 6 October 2015

Abstract The widely acknowledged 'RNA world' theory pertains to how life might have chemically originated on early Earth. It presumes the existence of catalytic RNAs, which were also capable of storing and propagating genetic information. Substantial research has gone into understanding how enzyme-free reactions of nucleic acids might have led to the formation of such catalytic RNA polymers. However, most of these studies involved reactions that were performed in aqueous systems devoid of any "background" molecules. This scenario is not a true representation of the complex chemical environment that might have been prevalent on prebiotic Earth. In the present study, we analyzed the effect of co-solutes ("background" molecules) on the rate and accuracy of template-directed nonenzymatic replication of RNA, in a putative RNA world. Our results suggest that presence of co-solutes in the reaction affects the addition of purine monomers across their cognate template base. Reduction in the rate of these 'fast' cognate addition reactions resulted in an apparent increase in the frequency of mismatches in the presence of co-solutes. However, reactions that involved the addition of a mismatched base were not notably affected. Such a scenario could have led to an accrual of mutations during the propagation of functional sequences on early Earth, unless the relevant sequences were separated from the bulk reaction milieu by some limiting boundary structure (e.g., a membrane). In general, our results suggest that the presence of co-solutes could have affected certain prebiotic reaction rates to a larger extent than others. Even modest changes in nonenzymatic replication reaction rates could have eventually resulted in the accumulation of greater variation in RNA sequences over prolonged time periods. It, therefore, is pertinent to account for the chemical complexity intrinsic to prebiotic environments while studying relevant nonenzymatic reactions.

Keywords Origin of life · Prebiotic complexity · RNA world · Nonenzymatic replication · Co-solutes

Introduction

Last few decades of research has enabled delineating the workings of many contemporary life processes. On the contrary, the abiotic origin of life continues to be a fascinating and challenging mystery. The concept of 'RNA world', a widely accepted hypothesis, pertains to a time in the early history of life's origin when RNA molecules are thought to have acted, both, as a genetic material and a catalyst. This hypothesis assumes that the RNA world was populated by oligomeric RNA molecules that propagated their genetic information by virtue of self-replication (Gilbert 1986). To evolve a functional ribozyme capable of self-propagation, a crucial first step in the process would have involved the generation of RNA molecules and their subsequent replication by chemical means. A significant amount of work has enabled delineating the chemical synthesis of nucleic acid monomers and characterizing plausible routes by which they would have polymerized under different prebiotically relevant conditions (Orgel 2004; Robertson and Joyce 2011). However, for an RNA molecule to act as its own replicase, it would have been

✉ Sudha Rajamani
srajamani@iiserpune.ac.in

[1] Indian Institute of Science Education and Research (IISER), Dr. Homi Bhabha Road, Pashan, Pune, Maharashtra 4110 008, India

crucial that this reasonably long functional molecule replicated its information with high efficiency.

Previous studies have demonstrated that purely chemistry-based, enzyme-free templated primer extension reactions are usually highly prone to mismatched base additions (e.g., G:U pairing) due to the unavailability of an error correction machinery (a function carried out by proteins today). In addition, if the frequency of errors is high, propagation of information over successive generations is hampered as maximum information that can be carried in a replicating genome is known to be inversely proportional to the mutation rate (Eigen 1971). It has previously been shown that the mismatched addition of nucleotides during nonenzymatic replication not only stalls the process (Rajamani et al. 2010), but also leads to a cascade of mismatches after the initial misincorporation (Leu et al. 2013), giving advantage to the faithfully replicating nucleic acids in such scenarios. The reaction kinetics and error rates for such enzyme-free primer extension reactions have been characterized. However, in these cases, the starting reactants used were usually very pure and the reaction mixtures were homogenous (Stütz et al. 2007; Rajamani et al. 2010; Leu et al. 2011). In other words, the systems studied thus far are what can be referred to as "pure systems" devoid of any co-solutes; a scenario that is prebiotically not realistic. It is pertinent to evaluate the aforementioned reaction variables in the presence of co-solutes as their presence is known to affect both the kinetics and the equilibrium of several biochemical reactions (Ellis 2001). Macromolecular crowding, a phenomenon facilitated in reactions with co-solutes, has been shown to not only affect protein folding and its stability, but also the rates of the enzymatic reactions that they catalyze (Minton 2001; Zhou et al. 2008). Such crowded conditions are also known to influence reactions involving nucleic acids (Nakano et al. 2014). For example, catalytic RNAs are known to get stabilized under crowded conditions (Kilburn et al. 2013; Desai et al. 2014). In addition, it has also been reported that the activity of certain ribozymes increases in the presence of co-solutes (Nakano et al. 2009; Kilburn et al. 2013; Strulson et al. 2013; Desai et al. 2014) and in aqueous two-phase systems, another system used to facilitate crowded conditions (Strulson et al. 2012). Interestingly, crowding improves the catalytic activity of ribozymes at physiologically relevant ionic strengths (Strulson et al. 2013), which is advantageous in reactions containing fatty acid protocells that are generally not very stable at higher ionic strengths (Chen et al. 2005; Adamala and Szostak 2013).

The prebiotic soup would have been a composite mixture of many different molecular species rather than being a homogenous solution of only certain kinds of 'interesting' molecules. Therefore, the phenomena of, both, molecular complexity and crowding are pertinent to prebiotic chemical environments. Nonetheless, studies involving the characterization of nonenzymatic replication reactions of nucleic acids in the presence of other chemical entities have not been reported so far. Characterizing enzyme-free copying in prebiotically relevant scenarios is important as they might not advent in a chemically complex environment in the same manner as they would in a chemically simpler scenario. The presence of other molecules could have subtle to pronounced effects on several aspects of nonenzymatic replication, which could potentially influence the overall fidelity of template-directed primer extension reactions. In scenarios wherein the reaction rates might be affected pronouncedly, this could also have a bearing on the efficiency of successive transfer of genetic information, which is imperative for the persistence and evolution of catalytic RNAs in an RNA world.

To the best of our knowledge, this study reports for the first time the effect of the presence of co-solutes on nonenzymatic template-directed primer extension reactions. We used Polyethylene Glycol (PEG) and a double-chain surfactant lipid as co-solutes in our reactions. Compartmentalization of the replicating genetic material is considered to have played a fundamentally important role during the origin and evolution of primitive life on early Earth (Szostak et al. 2001). Double-chain surfactant lipids are the contemporary counterparts of prebiotic amphiphiles, which are thought to have promoted the formation of protocellular membranes. We, therefore, wanted to understand the role of this important co-solute on error rates of nonezymatic replication. PEG, on the other hand, has been extensively used in facilitating volume exclusion effects in several biochemical experiments. In our reactions, PEG is also representative of other random non-specific polymers which might have been present in the prebiotic soup. In the present study, we characterized the effect of the aforementioned co-solutes, both, in isolation and in conjunction, to gain an understanding of the plausible role of co-solutes in prebiotic replication reactions.

Materials and Methods

Materials

RNA primer (Primer Amino G), terminated with 3′-amino-2′, 3′-dideoxynucleotide (Metkinen, Finland), was acquired from Keck laboratory, Yale, USA and used after gel purification. The primer was labeled with Cy3 on the 5′ end for enabling its detection upon polyacrylamide gel electrophoresis (PAGE) analysis. The RNA templates were acquired from Thermo Scientific (Dharmacon) and were used without further purification. The sequences of the primer and templates used are as follows (the template base is indicated in bold):

Primer Amino G: 5′ GG GAU UAA UAC GAC UCA CUG-NH_2
Template MisInc_C: 5′ AGU GAU CU**C** CAG UGA GUC GUA UUA AUC CC
Template MisInc_G: 5′ AGU GAU CU**G** CAG UGA GUC GUA UUA AUC CC
Template MisInc_A: 5′ AGU GAU CU**A** CAG UGA GUC GUA UUA AUC CC
Template MisInc_U: 5′ AGU GAU CU**U** CAG UGA GUC GUA UUA AUC CC

The double-chain surfactant lipid used in all the reactions was 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) and was purchased from Avanti Polar Lipids Inc. Nucleoside 5′-phosphoimidazolides (ImpNs) were purchased from GLSynthesis Inc. All other reagents used were of analytical grade purchased from Sigma-Aldrich.

Methods

Reaction Conditions

0.325 μM of primer and 1.3 μM of template were annealed to each other by heating at 95 °C for 5 min, followed by cooling at room temperature (RT) for another 5 min. Tris (pH 7.0) and NaCl were added to the reaction to a final concentration of 100 mM and 200 mM, respectively. DLPC vesicles were prepared by resuspending the dried film of lipid in nanopure water, followed by vortexing. The vesicles were then extruded through a 100 nm membrane using a mini extruder (Avanti Polar lipids Inc), before adding them to the reaction mix. DLPC was added as a co-solute to the reaction mixture at a final concentration of 5 mM. PEG 8000 was added as a co-solute in the reactions to a final concentration of 18 %. The primer extension reactions were initiated by addition of nucleoside 5′-phosphoimidazolides (ImpNs). Typically, 10 mM final concentration of Imp A/G/C and 40 mM final concentration of Imp U were used in the reactions (Rajamani et al. 2010; Leu et al. 2011).

Analysis of Primer Extension Reactions

The time points for the primer extension reactions were collected in TBE buffer containing 8 M urea and 100 mM EDTA. Competitor RNA, with a sequence similar to that of Primer Amino G (5′ GG GAU UAA UAC GAC UCA CUG), was added in up to 10 times excess to the reaction time points, prior to loading them onto the gel. This was done to ensure complete separation of the fluorescently tagged primer from their respective templates. The time points were then run on 20 % denaturing PAGE. The gels were scanned using Typhoon Trio plus imager (GE Healthcare) to detect the presence of fluorescently labeled primer and the extended product. The scans were analyzed using ImageQuant v 5.2 software. The extension of the primer over time was calculated by dividing the intensity of unreacted primer by sum of the intensities of both reacted (extended) and unreacted primer. The initial rate of the reactions was estimated by fitting a linear equation to several early data points (Rajamani et al. 2010; Leu et al. 2011).

Calculation of Error Rates

The frequency of incorporation of a specific nucleotide (ImpN) across a particular template base was calculated by dividing the rate of addition of that nucleotide by sum of the rates of addition of all nucleotides across the same template base in the primer-template complex. Mutation rate for a given template base (μ_N) was calculated by adding the frequencies of incorrect incorporations across that base. For the calculation of average mutation rates in a system, we have assumed equal parts of A, G, C, and U in the template. The average mutation rate was calculated by taking the average of mutation rates of all the template bases in a particular system ($\mu_{avg} = 0.25 *(\mu_A + \mu_G + \mu_C + \mu_U)$). Mutation rates for genomes composed of equal parts of G+C or A+U were calculated in a similar manner.

Results

To study the effect of co-solutes on enzyme-free copying of RNA templates, we used DLPC and PEG8000 as co-solutes in our reactions. We analyzed the individual effects of these co-solutes, as well as the combined effect of the admixture of these co-solutes, on template-directed primer extension reactions by following the reaction over a course of time. We determined the apparent first-order rate of addition of matched as well as mismatched nucleotides in the aforementioned scenarios, against the different template bases, to characterize potential changes in the reaction rates of such nonenzymatic replicating systems (Fig. 1).

Effect of Individual Co-solutes on Nonenzymatic Replication

We studied the effect of the presence of DLPC or PEG8000 as co-solutes, individually, on nonenzymatic replication reactions by studying the rate of template-directed addition of a single activated nucleotide to the fluorescently labeled primer. These reactions were carried out for the ‘correct’ incorporation of a cognate base against all four template bases i.e., addition of ImpG, ImpC, ImpA, and ImpU across

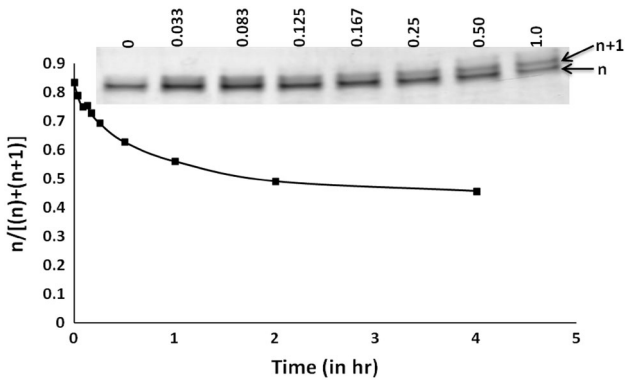


Fig. 1 Incorporation rate of cognate nucleotide 'Imp A' across the template base U of the 'MisInc_U' template, in the presence of both DLPC vesicles and PEG. The gel image shows extension of the primer over time (initial rate = 0.509/h)

C, G, U, and A template bases, respectively. The rates of addition were compared to those obtained for control reactions which were carried out in buffer, without the addition of any co-solutes (comparable to reactions carried out in Leu et al. 2011). Interestingly, the rate of addition of 'G across C' (where C is the template base) was clearly slowed down in the presence of either of the co-solutes. In case of addition of 'A across U' (where U is the template base), the individual co-solutes did seem to slow down the rate of extension in each of the three individual replicate sets analyzed under different reaction conditions (control vs individual co-solute-based reactions). However, the absolute rates varied between the experimental replicates leading to the observed large deviation from the mean values (Table 1). In other words, the rate of primer extension was reduced when the incoming cognate nucleotide was a purine (rates indicated in green in Table 1). Furthermore, we also studied the effect of the presence of the aforementioned co-solutes on 'incorrect' or misincorporation rates (mismatched additions) against all four template bases i.e., a total of twelve mismatched addition reactions. We observed no significant effect on the rate of mismatched additions, even when the incoming non-cognate base was a purine (Table 1).

Effect of Admixture of Co-solutes on Nonenzymatic Replication

As the presence of individual co-solutes was observed to be affecting the rate of some of the primer extension reactions, we decided to study the combined effect of the presence of both DLPC and PEG 8000 on all the matched and mismatched primer extension reactions, against all four template bases. This was done to experimentally increase the complexity in the starting reaction mixture. As suspected, the rates of addition of 'G across C' were clearly affected in these scenarios too (Table 1). For the addition of 'A across U,' we saw an effect similar to aforementioned individual co-solute-based reactions. In each of the three replicate sets, the admixture of co-solutes did reduce the absolute rate of addition of 'A across U' but the error between experiments was again large. Nevertheless, the reduction in the rate observed for each replicate of the aforementioned admixture-based reactions was, on an average, close to 55 % in comparison to that of the corresponding control reaction rate. To analyze whether this change in the rate is significant, we compared the difference in the rate (Δ Rate = rate of control reaction minus the rate of reaction in presence of both the co-solutes) within individual replicates for all combinations of 'incoming nucleotide-template base' (Table 1). The variations in the rate between the replicates for the same reaction can be mitigated to some extent by using the normalized value (Δ Rate/Rate of control reaction). However, this procedure will (spuriously) amplify the Δ Rates and their dispersion when rate of control reaction itself is very small. This issue can be resolved by analyzing the data for difference in the rates of control reaction and that in presence of both the co-solutes on a 2-dimensional plot with 'Δ Rate' along one axis and 'Δ Rate/Rate of control reaction' along the other (Fig. 2). Only those points which are substantially distant from both the axes have a Δ Rate value significantly different from zero.

Figure 2 shows that the 'Δ Rate' for two reactions i.e., 'G across C' and 'A across U' (points 'U/A' and 'C/G', indicated in red) are the only ones distant from both the axes. The large difference in their error bars along the two axes suggests variation from one replicate set to the other, which has been compensated by the normalization. This is also particularly evident for the difference in the rates for addition of 'C across G' (green point along the *x*-axis) which has large error bars along the *x*-axis but the normalized value is at zero. On the other hand, some of the reactions whose rate of control reactions are close to zero (e.g., addition of 'C across U') show the expected (though spurious) large error bars along the *y*-axis. In summary, all the data points, excepting points 'U/A' and 'C/G' (indicated in red), are consistent, with zero difference between rate obtained for control reaction and that in presence of both the co-solutes. Furthermore, we used an iterative robust statistical procedure to determine the base level of the difference in the reaction rates (Δ Rate) and the significance of the deviation of the two values from the same (i.e., the Δ Rate for addition of 'G across C' and 'A across U'). Robust analysis using 3 times the standard deviation (99.5 %) as the cut-off for deviants yielded $N = 14$ points with mean $= -0.005$ and standard deviation $= 0.054$. This analysis indicates that the Δ rates for addition of 'A across U' and that of 'G across C' were both excluded from

Table 1 Experimental rates for nonenzymatic primer extension reactions (h^{-1}) under control reaction conditions and in the presence of either or both of the co-solutes (Color table online)

Reaction No.			Control		With DLPC		With PEG8000		With DLPC and PEG8000	
	Template	ImpN	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}	Rate/hr	σ_{exp}
1	C	G	1.58	0.17	1.09	0.1	0.79	0.26	0.69	0.08
2	C	U	0.021	0.002	0.022	0.004	0.02	0.002	0.022	0.005
3	C	C	0.022	0.002	0.014	0.008	0.022	0.004	0.028	0.0004
4	C	A	0.017	0.006	0.026	0.01	0.018	0.006	0.027	0.008
5	G	C	9.03	0.55	8.42	1.22	8.97	0.86	8.91	0.6
6	G	U	0.48	0.28	0.40	0.17	0.71	0.16	0.55	0.27
7	G	G	0.16	0.06	0.18	0.04	0.11	0.04	0.18	0.11
8	G	A	0.04	0.02	0.048	0.02	0.048	0.008	0.03	0.003
9	A	U	0.53	0.17	0.54	0.1	0.49	0.08	0.53	0.19
10	A	G	0.17	0.05	0.17	0.01	0.17	0.06	0.2	0.14
11	A	C	0.08	0.001	0.09	0.007	0.11	0.006	0.11	0.007
12	A	A	0.07	0.03	0.056	0.03	0.053	0.007	0.024	0.006
13	U	A	0.92	0.37	0.73	0.24	0.57	0.33	0.43	0.22
14	U	U	0.12	0.07	0.047	0.02	0.058	0.02	0.074	0.03
15	U	C	0.017	0.001	0.018	0.002	0.023	0.002	0.03	0.014
16	U	G	0.38	0.03	0.24	0.09	0.23	0.03	0.49	0.06

σ denotes standard deviation from experimental replicates

Rows indicated in green show the experimental rates of addition of a purine monomer across its cognate template base while those indicated in yellow show the experimental rates of addition of a pyrimidine monomer across its cognate template base. n = 3 for all matched addition reactions, n = 2 for all mismatched addition reactions.

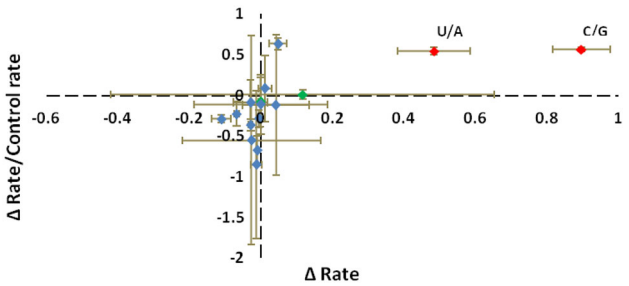


Fig. 2 The 'Δ Rate normalized by the initial control rate'(*Y* axis) is plotted against the 'Δ Rate'(*X* axis) for each reaction. The 'Δ Rate' is significantly different from zero for addition of 'G across C' (point C/G) and addition of 'A across U' (point U/A), both indicated as *red* diamonds. The *green* diamonds indicate the remaining two matched addition reactions. All mismatched addition reactions are indicated as *blue* diamonds. *Error bars* indicate the standard error. $n = 3$ for all cognate addition reactions and $n = 2$ for all mismatched addition reactions (Color figure online)

this "zero" cluster at significance levels of alpha = 0.025 and 0.005, respectively (i.e., at confidence level 97.5 and 99.5 %).

It is to be noted that the combined effect of the presence of more than one chemical entity (or co-solute) was more pronounced than when only one of these was added to the reactions wherein the incoming cognate nucleotide was a purine (indicated in green in Table 1). Additionally, the kinetics of the other two matched reactions i.e., 'U across A' and 'C across G', wherein the incoming cognate nucleotide is a pyrimidine base (indicated in yellow in Table 1), seemed unaffected by the presence of any of the co-solutes, either individually, or when the admixture of co-solutes was used (Fig. 2; indicated by green diamonds). The rates of misincorporation reactions were also not found to have been significantly affected (Fig. 2; indicated by blue diamonds).

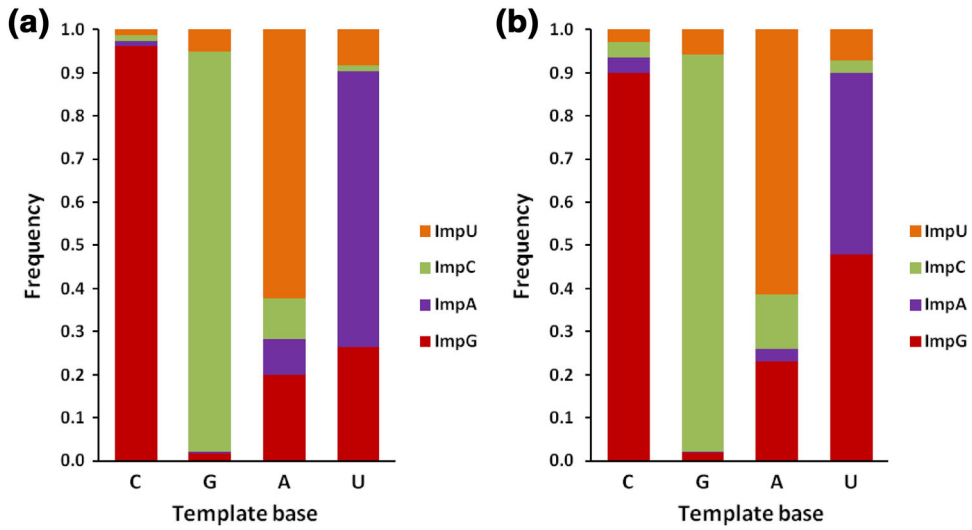


Fig. 3 Incorporation frequencies for addition of cognate and non-cognate bases. **a** In the absence of any co-solutes. **b** In the presence of DLPC and PEG8000 as co-solutes (Color figure online)

Estimation of Experimental Error Rates in the Presence of Co-solutes

Our results indicate that the misincorporation frequencies are higher for A and U than that of G and C template bases, consistent with previous results (Rajamani et al. 2010; Leu et al. 2011). The prominently observed misincorporation was the addition of 'G across U,' the frequency of which considerably increased in the presence of both the co-solutes as compared to the control reaction (Fig. 3). This was a consequence of the decrease in the rate of cognate base addition across 'U' template base in the presence of the co-solutes as compared to the control condition (Reaction 13 as indicated in Table 1). However, misincorporation frequency of 'U across G' was not considerably affected in the presence of the co-solutes apparently due to the faster rate of the corresponding cognate base addition across G (i.e., 'C across G') than of that across U (i.e., 'A across U'). The mutation rates were calculated from two independent reaction sets wherein we calculated the rates for all possible matched and mismatched addition reactions. The overall mutation rate for a genome comprising equal parts of A, U, G, and C was 19.8 ± 0.5 % in our control reactions, which were devoid of any co-solutes. The mutation rate in the presence of only DLPC as co-solute was found to be 18.9 ± 0.2 % and that in the presence of only PEG as co-solute was found to be 20.7 ± 0.4 %. The observed mutation rate increased to 26.6 ± 4.4 % in the reactions, which had both PEG and lipid vesicles as co-solutes. However, this change was not found to be statistically significant ($p = 0.159$, Student's t test). Nonetheless, this effect on mutation rate becomes prominent for copying of GC-genomes (mutation rate of 5.3 ± 0.7 % in control reaction vs. 8.6 ± 0.3 % in the presence of both the co-solutes, $p < 0.05$) as compared to that of AU-genomes (mutation rate of 34.3 ± 0.2 % in control reaction vs. 44.4 ± 9 % in the presence of both the co-solutes, $p > 0.05$). Larger error bars associated with the mutation rates of AU genomes might stem from the inherent low efficiency associated with the copying of A and U template bases (Orgel 2004).

Discussion

Chemical environments on prebiotic Earth, that are thought to have supported the origin of biologically relevant molecules, would have been replete with several different kinds of molecules rather than being a preferentially concentrated solution of only specific types of moieties. In this study, we wanted to discern the effect of the presence of co-solutes on the rate of template-directed nonenzymatic replication reactions. Lipid vesicles and PEG were used as co-solutes in our reactions. Lipids are a crucial part of contemporary cell membranes. Extant/modern lipids are modified versions of simpler amphiphiles like fatty acids, which are suggested to have formed membrane boundary structures around the earliest forms of cells (protocells). Fatty acid molecules are thought to have been formed on prebiotic Earth by Fischer–Tropsch synthesis like reactions (McCollom et al. 1999). Moreover, organic hydrocarbon components of certain carbonaceous meteorites have also been shown to form membrane like structures (Segré et al. 1999). It therefore is conceivable that prebiotic pools could have contained sufficient amphiphilic molecules. For these preliminary studies, we used DLPC, a modern counterpart of prebiotic amphiphiles, to study the effect of amphiphiles on nonenzymatic replication reactions. PEG, on the other hand, is a known molecular crowding agent and has been used extensively in various studies to simulate volume

exclusion effects. The prebiotic soup would have been populated by oligomers that would have resulted from parallel polymerization reactions involving other chemical moieties such as lactic acid, malic acid, and ribose-5'-monophosphate (Harshe et al. 2007; Mamajanov et al. 2014; Mungi and Rajamani 2015). In addition, a very recent study also demonstrated a possible route for selective survival of longer nucleic acid strands in a prebiotically relevant context (Kreysing et al. 2015). The use of PEG in our reactions is to simulate the presence of these other random polymers in order to study their collective volume exclusion effects.

Based on observations from previous background literature, our original hypothesis was that the overall reaction rates of all the co-solute-based reactions might actually increase due to the volume exclusion effect that is facilitated under such conditions. It is known that in crowded conditions, catalytic rates of biomolecular reactions generally tend to increase. This is mostly attributed to concentration-related volume exclusion effects (Minton 2001; Zhou et al. 2008). Interestingly, our results indicated that in the presence of co-solutes only the rates of those fast matched addition reactions were altered wherein the incoming cognate nucleotide was a purine. We speculate this might be driven by the fact that purines can stack better in solution as compared to pyrimidines. Importantly, such a property could affect the availability of purine monomers in reactions wherein the presence of co-solutes facilitate volume exclusion, leading to concentration-related effects. Previous studies have shown that crowded conditions tend to affect the diffusion of biomolecules (Banks and Fradin 2005; Kuznetsova et al. 2015). Moreover, crowding is also known to affect the diffusion of small molecules to a considerable extent (Dauty and Verkman 2004). For this reason, we envisage that in our reactions volume exclusion could be affecting the kinetics of matched reactions involving the addition of a purine and this potentially might stem from reduced diffusion of stacked purine monomers in comparison to that of the free monomers. The purine monomers, therefore, might not be readily available for primer extension, which in turn could be altering the rate of this subset of fast matched reactions of the primer-template complex. This effect was especially enhanced in those reactions where a combination of co-solutes was used to increase the complexity in our reaction mix (rates indicated in green in Table 1). Mismatched reactions, on the contrary, are known to occur at a much slower rate than that of the matched reactions; some of them up to two orders of magnitude slower. Due to the inherent slow kinetics of mismatched reactions, alteration in the diffusional mobility of potentially stacked purine monomers (in the presence of the co-solutes), does not seem to affect their rates to a similar extent in our reactions. Consequently, the rate of mismatched reactions is not obviously affected in the presence of co-solutes even if the incoming nucleotide is a purine.

Although the rate of mismatched reactions was not altered much in the presence of co-solutes, the reduction in the rate of two of the otherwise fast cognate addition reactions resulted in a small decrease in the fidelity of nonenzymatic template-directed primer extension reactions in our co-solute-based scenarios. In general, the presence of co-solutes led to an increase in the mutation rates for those reactions wherein a pyrimidine nucleotide was the template base. As stated earlier, the misincorporation frequency of 'G across U' was more pronounced in the presence of co-solutes (Fig. 3). G: U wobble pairing is known to be one of the most commonly observed mismatches during nonenzymatic reactions (Schrum et al. 2009; Leu et al. 2011). The rate of correct incorporation of 'A across U' was altered in the presence of co-solutes in all experimental replicate sets despite the absolute rates within each replicate set showing variation ($\sim$55 % reduction in rate was observed). Additionally, incorporation of 'G across U' was not affected much, which resulted in an overall reduction of fidelity of copying of the 'U' template base. Therefore, the mutation rate of enzyme-free copying of 'U' template base was observed to have increased in the presence of co-solutes (mutation rate of 31 ± 4.6 % in control reaction vs. 51.7 ± 3.5 % in the presence of both co-solutes, $p < 0.05$), and this was more than in comparison to any other template base.

Addition of 'C across G' and that of 'G across C' were two of the fastest cognate base copying reactions observed in our study, consistent with previous results (Leu et al. 2011, 2013). However, due to the inherent low rate of cognate additions of 'A across U' and vice versa, and an increase in the frequency of G:U wobble pairing in the presence of co-solutes, the complexity inherent in a prebiotic chemical environment would have adversely affected the copying of 'AU' rich genomes to a greater degree than that of 'GC' rich genomes. Interestingly, the standard deviation obtained for fidelity calculation of 'AU' rich genome in the presence of co-solutes was somewhat larger than usual (refer "Estimation of experimental error rates in the presence of co-solutes" section under "Results" section). This is because the two reaction replicates for the addition of 'G across U' and that of 'G across A' reactions had an intrinsically greater variability in their extension rates. We speculate that this might potentially stem from random variability in the stacking interactions of G monomers in the presence of co-solutes. The overall reduction in the fidelity of enzyme-free copying, as we had hypothesized, was observed to be higher for an 'AU' genome than a 'GC' genome in the presence of both the co-solutes. This would tend to bias the genome composition

toward being more GC-rich, as was also suggested in the earlier study (Leu et al. 2011). In general, our results suggest that increasing the complexity of the prebiotic soup could lead to alterations in nonenzymatic reaction rates in a context dependent manner. Importantly, this underlines the need to factor in prebiotic complexity while quantifying reactions that are relevant to understanding the chemical origins of life. Even a step-wise increase in the starting complexity led to changes in the reactions rates of certain nonenzymatic replication reactions, suggesting the possibility that the presence of various co-solutes in the starting mix, like in a realistic prebiotic milieu, could have had more pronounced effects on similar reactions. Consequently, the presence of such chemically complex reaction mixtures would have posed a considerable challenge for the efficient replication of any functional polymer due to the aforementioned and possibly other reasons. Therefore, it would have been crucial to compartmentalize the functionally significant polymers during evolution, as has been previously suggested (Szostak et al. 2001). Interestingly, in certain in silico studies, encapsulation has been shown to be advantageous for group selection of templates (Fontanari et al. 2006) as well as for the selection of replicases (Bianconi et al. 2013).

In conclusion, our study highlights the importance of accounting for the complexity that was prevalent in the chemical environments of prebiotic Earth as it could have significant implications for the origin of functionally important molecules. Importantly, these results also underline the significance of compartmentalization very early on in order to minimize the error rates of nonenzymatic replication during evolution of functionally competent nucleic acids.

Acknowledgments We thank IISER Pune for financial support. NVB receives fellowship from CSIR, Government of India. We are profoundly thankful to Dr. Ramana Athreya for his help with the statistical analysis and grateful to Chaitanya Mungi for useful discussions.

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no conflict of interest.

References

Adamala K, Szostak JW (2013) Competition between model protocells driven by an encapsulated catalyst. Nat Chem 5:495–501. doi:10.1038/nchem.1650

Banks DS, Fradin C (2005) Anomalous diffusion of proteins due to molecular crowding. Biophys J 89:2960–2971. doi:10.1529/biophysj.104.051078

Bianconi G, Zhao K, Chen IA, Nowak MA (2013) Selection for replicases in protocells. PLoS Comput Biol 9(5):e1003051. doi:10.1371/journal.pcbi.1003051

Chen IA, Salehi-ashtiani K, Szostak JW (2005) RNA catalysis in model protocell vesicles. J Am Chem Soc 127:13213–13219. doi:10.1021/ja051784p

Dauty E, Verkman AS (2004) Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: measurement by fluorescence correlation spectroscopy. J Mol Recognit 17:441–447. doi:10.1002/jmr.709

Desai R, Kilburn D, Lee H, Woodson SA (2014) Increased ribozyme activity in crowded solutions. J Biol Chem 289:2972–2977. doi:10.1074/jbc.M113.527861

Eigen M (1971) Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465–523

Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26:597–604. doi:10.1016/S0968-0004(01)01938-7

Fontanari JF, Santos M, Szathmáry E (2006) Coexistence and error propagation in pre-biotic vesicle models: a group selection approach. J Theor Biol 239:247–256. doi:10.1016/j.jtbi.2005.08.039

Gilbert W (1986) The RNA world. Nature 319:618

Harshe YM, Storti G, Morbidelli M, Gelosa S, Moscatelli D (2007) Polycondensation kinetics of lactic acid. Macromol React Eng 1:611–621. doi:10.1002/mren.200700019

Kilburn D, Roh JH, Behrouzi R, Briber RM, Woodson SA (2013) Crowders perturb the entropy of RNA energy landscapes to favor folding. J Am Chem Soc 135:10055–10063. doi:10.1021/ja4030098

Kreysing M, Keil L, Lanzmich S, Braun D (2015) Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. Nat Chem 7:203–208. doi:10.1038/nchem.2155

Kuznetsova IM, Zaslavsky BY, Breydo L, Turoverov KK, Uversky VN (2015) Beyond the excluded volume effects: mechanistic complexity of the crowded milieu. Molecules 20:1377–1409. doi:10.3390/molecules20011377

Leu K, Obermayer B, Rajamani S, Gerland U, Chen IA (2011) The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. Nucleic Acids Res 39:8135–8147. doi:10.1093/nar/gkr525

Leu K, Kervio E, Obermayer B, Turk-Macleod RM, Yuan C et al (2013) Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. J Am Chem Soc 135:354–366. doi:10.1021/ja3095558

Mamajanov I, Macdonald PJ, Ying J, Duncanson DM, Dowdy GR, Walker CA, Engelhart AE 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. doi:10.1021/ma402256d

McCollom TM, Ritter G, Simoneit BRT (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. Orig Life Evol Biosph 29:153–166

Minton AP (2001) Influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J Biol Chem 276:10577–10580. doi:10.1074/jbc.R100005200

Mungi CV, 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. doi:10.3390/life5010065

Nakano S, Karimata HT, Kitagawa Y, Sugimoto N (2009) Facilitation of RNA enzyme activity in the molecular crowding media of cosolutes. J Am Chem Soc 131:16881–16888. doi:10.1021/ja9066628

Nakano S, Miyoshi D, Sugimoto N (2014) Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. Chem Rev 114:2733–2758. doi:10.1021/cr400113m

Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Bio 39:99–123. doi:10.1080/10409230490460765

Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA (2010) Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. J Am Chem Soc 132:1008–1011. doi:10.1021/ja100780p

Robertson MP, Joyce GF (2011) The origins of the RNA world. In: Atkins JF, Gesteland RF, Cech TR (eds) RNA worlds: from life's origins to diversity in gene regulation. CSH Press, New York

Schrum JP, Ricardo A, Krishnamurthy M, Blain JC, Szostak JW (2009) Efficient and rapid template-directed nucleic acid copying using 2′-amino-2′,3′-dideoxyribonucleoside-5′-phosphorimidazolide monomers. J Am Chem Soc 131:14560–14570. doi:10.1021/ja906557v

Segré D, Ben-Eli D, Deamer DW, Lancet D (1999) The lipid world. Orig Life Evol Biosph 31:119–145

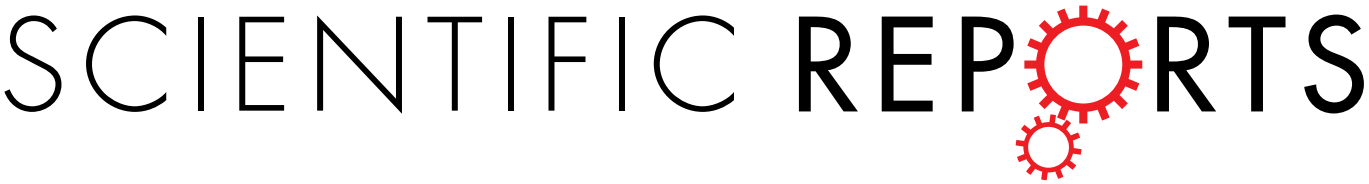
Strulson C, Molden R, Keating C, Bevilacqua PC (2012) RNA catalysis through compartmentalization. Nat Chem 4:941–946. doi:10.1038/nchem.1466

Strulson CA, Yennawar NH, Rambo RP, Bevilacqua PC (2013) Molecular crowding favors reactivity of a human ribozyme under physiological ionic conditions. Biochem 52:8187–8197. doi:10.1021/bi400816s

Stütz AR, Kervio E, Deck C, Richert C (2007) Chemical primer extension: individual steps of spontaneous replication. Chem Biodivers 4:784–802. doi:10.1002/cbdv.200790064

Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390. doi:10.1038/35053176

Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Ann Rev Biophys 37:375–397. doi:10.1146/annurev.biophys.37.032807.125817



OPEN

Templated replication (or lack thereof) under prebiotically pertinent conditions

Received: 11 December 2017
Accepted: 24 September 2018
Published online: 09 October 2018

Niraja V. Bapat & Sudha Rajamani

Accurate replication of encoded information would have been crucial for the formation and propagation of functional ribozymes during the early evolution of life. Studies aimed at understanding prebiotically pertinent nonenzymatic reactions have predominantly used activated nucleotides. However, the existence of concentrated pools of activated monomers on prebiotic Earth is debatable. In this study, we explored the feasibility of nonenzymatic copying reactions using the more prebiotically relevant 5′-nucleoside monophosphates (5′-NMP). These reactions, involving a 20-mer primer, were performed in the presence of amphiphiles, under volcanic geothermal conditions. Interestingly, the extended primer was not comparable to the expected full length 21-mer product. Our results suggest loss of the nitrogenous base in the extended primer. This phenomenon persisted even after lowering the temperature and when different rehydration solutions were used. We envisage that the loss of the informational moiety on the incoming 5′-NMP, might be occurring during addition of this monomer to the pre-existing oligomer. Significantly, when 5′-ribose monophosphate was used, multiple additions to the aforementioned primer were observed that resulted in hybrid polymers. Such hybrid oligomers could have been important for exploring a vast chemical space of plausible alternate nucleobases, thus having important implications for the origin of primitive informational polymers.

A widely accepted hypothesis pertaining to the existence of a putative 'RNA World' presumes that RNA molecules played a central role during the emergence and evolution of early life on Earth. Central to this hypothesis is the notion of an 'RNA replicase' enzyme (a ribozyme) that is capable of total self-replication[1]. However, prior to the advent of ribozymes that could catalyze reactions efficiently, it is thought that enzyme-free propagation of genetic information would have been a fundamentally crucial step. Inspired by this line of thought, studies have been undertaken in the last few decades to characterize nonenzymatic polymerization and information transfer reactions of nucleic acids. Most of these aforementioned studies have involved the use of activated nucleotides, and in some cases chemically modified primers, in order to facilitate the formation of either phosphodiester or phosphoramidite bond and, to also, obtain products in detectable yields. Several different activation chemistries, including imidazole, 2-methylimidazole, 1-methyladenine, 2-aminoimidazole[2,3], oxyazabenzotriazole[4] etc., have been explored thus far to successfully demonstrate enzyme-free oligomerization and copying reactions. Although a recent finding demonstrates a plausible route for the synthesis of imidazole activated ribonucleotides[5], the presence or existence of significant amounts of most of the aforementioned activated nucleotides on the prebiotic Earth, is still arguable. On the other hand, possible prebiotic routes for the syntheses of nucleotides, and of their precursors, have been demonstrated in previous studies[2,6–8]. Very few reports, however, have demonstrated and characterized the products resulting from the oligomerization of non-activated monomers such as nucleoside-5′-monophosphates[9–11]. Importantly, the kinetics and fidelity of enzyme-free template-directed primer extension reactions of oligomers, in a putative RNA World, using non-activated nucleotides, still remains largely unexplored. Therefore, it becomes pertinent and important to study the feasibility of nonenzymatic information transfer and related processes, using non-activated nucleotides, which are prebiotically relevant in comparison to their activated counterparts.

Phosphodiester bond formation between an existing nucleic acid polymer and an incoming 5′-NMP is a slow reaction due to lack of a good leaving group such as an imidazole moiety on the phosphate. Furthermore, polymerization is essentially a condensation reaction involving the loss of a water molecule. Therefore, the rate of this

Indian Institute of Science Education and Research (IISER), Dr. Homi Bhabha Road, Pashan, Pune, 411 008, Maharashtra, India. Correspondence and requests for materials should be addressed to S.R. (email: srajamani@iiserpune.ac.in)

reaction cannot be enhanced in the presence of bulk water. Not surprisingly, alternating cycles of dehydration and rehydration (DH-RH cycles) at elevated temperatures, have been shown to promote the polymerization of nucleic acid monomers[12] as well as alpha hydroxy acids[13] and malic acid monomers[14]. Dehydration at high temperatures not only results in concentration of the reactants, it also enhances the loss of water molecules, facilitating condensation and, thus, promoting bond formation. Such alternate DH-RH cycles would have been facilitated on the prebiotic Earth by diurnal cycles, seasonal variations etc. For e.g., dehydration in relevant niches might have been facilitated by high temperatures, which is followed by rehydration that results due to condensation of water at lower temperatures. It is thought that such alternate DH-RH cycles would have been abundant in terrestrial geothermal fields and in inter-tidal pools, geological features that are thought to have been prevalent on the early Earth. In a related study, researchers studied the copying of information from a single-stranded DNA, using deoxyribonucleotide monomers, in the presence of lipids, under DH-RH conditions[15]. Lipids are known to form liquid-crystalline matrices under dehydrated conditions, facilitating the concentration of the starting reactants (e.g. nucleotides) within the inter-layers of the resultant multilamellar structures[16,17]. This has previously been shown to facilitate the polymerization of 5′-NMPs, under DH-RH conditions and in the presence of phospholipids[9,11]. However, to our knowledge, no study has been reported to date wherein information transfer using RNA, and prebiotically relevant non-activated nucleotides, has been systematically evaluated especially under prebiotically relevant volcanic geothermal conditions. These RNA based studies are crucial as they pertain to a molecule whose relevance to the emergence of early life has been demonstrated in multiple lines of studies.

In this particular study, we aimed to discern phospholipid-assisted template-directed primer extension reactions, under DH-RH regimes and at elevated temperatures. As indicated earlier, this is likely the first study where in template-directed information transfer from RNA has been attempted using non-activated 5′-NMP nucleotides. As in previous comparable studies, we used the 3′-amino-2′, 3′-dideoxynucleotide terminated primer (Amino-G primer), to study pertinent reactions, as it will enable the detection of the extended product, which is known to form in observable yields[18–20]. The effects of various parameters on these reactions have also been systematically characterized. Importantly, we also report the extension of a pre-existing RNA primer using 5′-ribose monophosphate (5′-rMP), a sugar-phosphate monomer, that resulted in hybrid polymers. Such hybrid polymers could have potentially been significant in sampling different bases during the emergence of primitive functional polymers on the early Earth[11].

Results

Previously, few studies reported the polymerization of 5′-NMPs under DH-RH regimens, at low pH and in a range of high temperatures[9,11]. Furthermore, the presence of lipid molecules in the starting reaction mixture was shown to enhance the polymerization process. This is because lipids are known to form multilamellar structures under dehydrated conditions, allowing for the entrapment of the nucleotide monomers, thus facilitating their polymerization. Therefore, we set out to study whether template-directed extension of an existing RNA primer was also possible under similar conditions, using non-activated nucleotides.

Extension of RNA primer under DH-RH regimen. In preliminary reactions, 1, 2-dilauroyl-sn-glycero-3 -phosphocholine (DLPC) lipid vesicles and 5′- NMPs were added to a final concentration of 5 mM and 10 mM respectively, to the pre-annealed RNA primer-template complex. This mixture was subjected to repeated cycles of DH-RH at 90 °C, and 1 mM H_2SO_4 was used as rehydration solution in these reactions to facilitate an acid-catalyzed esterification type reaction that has been demonstrated previously[14,15]. Upon analysis using denaturing gel electrophoresis, a band for the extended primer product was observed over increasing DH-RH cycles. However, this new band was not comparable to the expected 21-mer product (i.e. one nucleotide addition to the original 20-mer primer). In fact, the newly formed band did not run to the same extent as an intact 21-mer band would have, on the 20% denaturing urea gel that was used for analyzing the progress of the reaction (first lane of Fig. 1a, intact 21-mer band indicated by a blue arrow). The band obtained was actually found to be somewhere in between that of a 20-mer and 21-mer RNA (indicated by a red arrow in 'Cyc 10' lane of Fig. 1a). This was true for all of the 5′-NMP-based reactions (N = A/U/G/C), irrespective of which one was used as the starting monomer in a given reaction. We suspect that the extended product in these reactions might have an abasic site, a result that has previously been reported for comparable concatenation reactions that were carried out under similar conditions[11]. Furthermore, we observed higher degradation of the fluorescently-labelled RNA primer over multiple cycles of DH-RH, when similar reactions were carried out in the absence of lipid (Fig. 1b left panel vs right panel). Comparable results were also observed when a longer chain-length phospholipid, namely 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), was used instead of DLPC, in the starting reaction mixture, suggesting a protective role for the lipids under these harsh prebiotically pertinent conditions. Additionally, in complementary studies, we also performed the aforementioned main experiment i.e. attempted templated-replication in the presence of amphiphiles, using a canonical hydroxyl terminated RNA primer (Hydroxyl-G primer). Since hydroxyl group is known to be less nucleophilic than an amino group, we did not see the formation of prominent amount of the extended primer product (Fig. S1).

Effect of temperature and rehydrating solution on primer extension. Loss of the nucleotide base is known to occur under high temperature and low pH conditions as the glycosidic bonds are susceptible to hydrolysis in such conditions[21,22]. Given that the extension of the primer at 90 °C, under DH-RH conditions, resulted in a product with a possible abasic site, the efficiency of the extension reaction was tested at lower temperatures. All the reactions mentioned henceforth were carried out using Amino-G primer, and in the presence of lipids. Extension of the Amino-G primer was observed at lower temperatures, including at 80 °C and at 50 °C (Fig. 2a, middle and left panels respectively). It is to be noted that lowering of the temperature led to an increase in the dehydration time for each cycle. For e.g., the dehydration time at 50 °C increased to up to 90 minutes (from

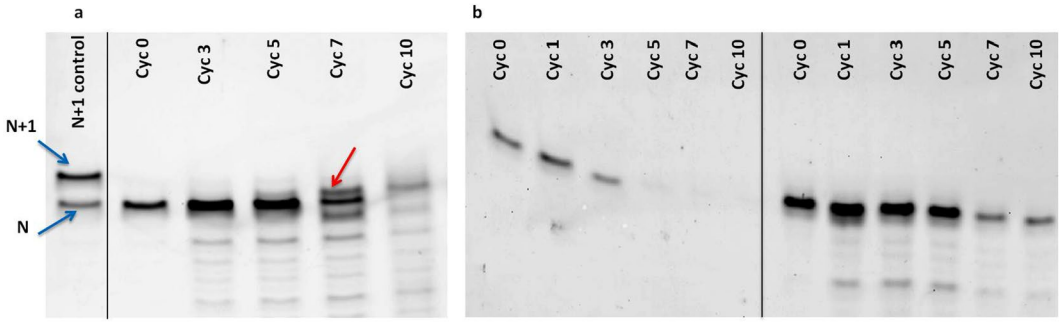


Figure 1. (**a**) Extension of the RNA primer using MisInc_U template and 5′-AMP as the monomer, in the presence of lipid, over repeated cycles of DH-RH. The reactions were carried out at 90 °C with 1 mM H_2SO_4 as the rehydrating agent. The red arrow indicates the extended product. In the control lane, 'N' indicates the 20-mer RNA primer while 'N + 1' indicates extension of the primer by one nucleotide[20]. (**b**) Stability of the RNA primer over multiple DH-RH cycles, when performed in the absence of lipid (left panel) and in the presence of lipid (right panel) in the starting reaction mixture. The black vertical lines in the above two panels have been used to demarcate two reaction sets that were run on the same gel (Panel b). In Panel a, the black line demarcates a reaction set from the control lane, both of which were analysed on the same gel.

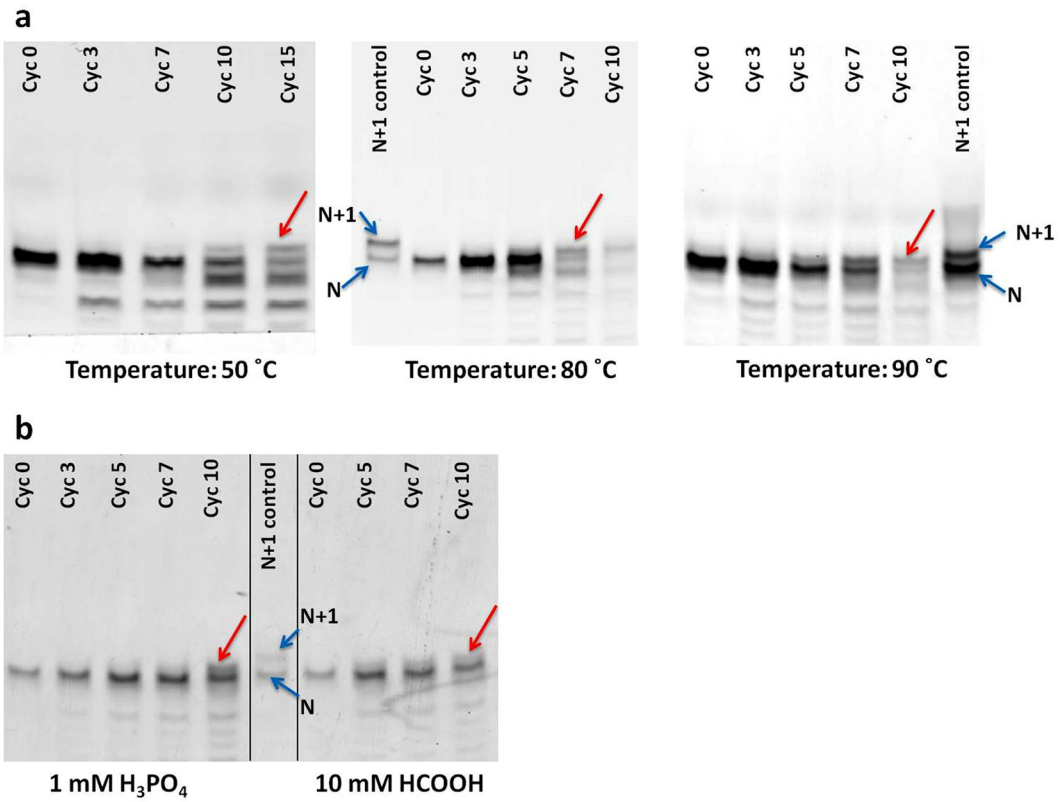


Figure 2. Extension of RNA primer in the presence of MisInc_U template, 5′-AMP as monomer and lipid (**a**) at different reaction temperatures (**b**) using different rehydrating agents. The red arrow indicates the extended primer product. The black vertical lines in the Panel b gel image have been used to demarcate reactions sets run on the same gel. Additionally, the black line has also been used to demarcate reactions from the control lane. 'N' indicates the 20-mer RNA primer and 'N + 1' indicates extension of the primer by one nucleotide.

30 minutes at 90 °C) due to slower evaporation rate. Importantly, the resulting extension product still ran lower than the intact 21-mer primer product that was used as control on the gel.

Furthermore, we also assessed the effect of different rehydrating agents on the potential loss of base during primer extension reactions, under DH-RH cycles at 90 °C. To begin with, different concentrations of H_2SO_4, varying from 0.05 mM to 5 mM, were used, amongst which 1 mM H_2SO_4 was found to be the optimum concentration. Higher concentrations of acid led to huge degradation of RNA during the cycling reactions whereas at lower concentrations of acid, no primer extension was observed (Fig. S2). Subsequently, different types of acids were also

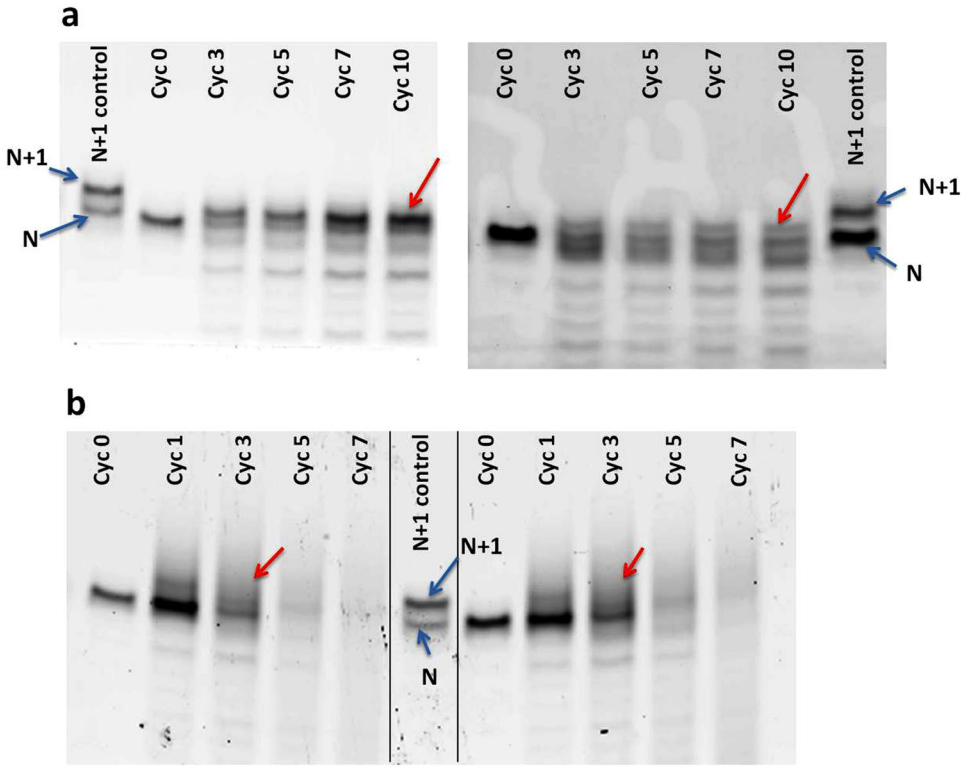


Figure 3. Extension of the RNA primer using (**a**) 1 mM 5′-AMP with 5 mM lipid (left panel) and 1 mM 5′-AMP with 1 mM lipid (right panel), in the starting reaction mixture. The reactions contained 1 mM H_2SO_4 and nanopure water was used as the rehydrating agent. (**b**) 10 mM of 5′-rMP, in the absence (left panel), and the presence (right panel) of the template in the reaction. Reactions were carried out at 90 °C, in the presence of 5 mM lipid and 1 mM H_2SO_4 was used as rehydrating agent. The black vertical lines in the Panel b gel image, demarcates two different reactions run on the same gel. Red arrows indicate extended RNA primer product/s. 'N' indicates the 20-mer RNA primer while 'N + 1' indicates extension of the primer by one nucleotide.

investigated as rehydrating solutions to check if any of them could prevent the possible loss of base during primer extension, under aforementioned conditions. Amongst the different mineral and organic acids that were tested, the extended primer product was detected when phosphoric acid and formic acid were used as rehydrating agents (Figs 2b and S3). However, the extended primer continued to run lower than an intact N + 1 (21-mer) product (bands indicated by red arrows in Fig. 2b).

Consequently, the concentration of the acid in the reaction mixture was kept low to see if this might prevent what seemed like the generation of an abasic site. Towards this attempt, sulphuric acid was added to the starting reaction mixture to a final concentration of 1 mM (to maintain the requisite pH) and nanopure water was used as rehydrating agent for the remainder of the DH-RH cycles. Using this approach, the extended primer product was observed at a starting concentration of 5′-NMP that was as low as 1 mM, albeit at a lower efficiency (Table S1). Furthermore, the extension was also observed at a lower concentration of the lipid (as low as 1 mM) in the starting reaction mixture (indicated by a red arrow in Fig. 3a). However, even at this low concentration of acid in the reaction mixture, the extended primer product still did not run at a level similar to that of a known 21-mer RNA control.

Effect of monovalent cations. A recent study from Da Silva *et al*. reported the formation of oligomers from 5′-NMPs in salty environments, and under cycling conditions at high temperatures[23]. Interestingly, the RNA-like oligomers formed in their reactions were shown to have base-pairing ability, as was evident by the ethidium bromide staining of the resultant products. Amongst the different salts that they had used, ammonium chloride was found to be very efficient in promoting the formation of oligomers. To analyze whether the presence of NH_4^+ ions might result in the formation of an intact extended primer product, NH_4Cl was added, either in isolation or along with NaCl, to the reaction mixtures. The final concentration for each of the salt was maintained at 200 mM. However, the presence of NH_4Cl did not improve the efficiency of our primer extension reactions in terms of resulting in an intact N + 1 extended primer. The resultant product band continued to migrate in between that of a known 20-mer and 21-mer RNA on the denaturing gel (Fig. S4). We also observed higher amounts of RNA degradation during the course of DH-RH reactions in the presence of ammonium ions, as opposed to when only sodium ions were present in the reaction (Compare Fig. S4 vs. Fig. 1a).

RNA primer extension using 5′-ribose monophosphate. 5′- ribose monophosphate (5′-rMP) is a monomer which contains a sugar and a phosphate entity but lacks any information carrying nitrogenous base

(which is otherwise present in 5′-NMPs). Earlier studies indicate that 5′-rMP polymerizes readily under fluctuating DH-RH conditions, to potentially result in linear and cyclic sugar-phosphate backbones[11]. However, the addition of such a backbone onto a pre-existing RNA oligomer has not yet been evaluated. Towards this, primer extension reactions were carried out under DH-RH regimes at 90 °C, using 5′-rMP as monomers as against the 5′-NMPs (used in the earlier reactions). Denaturing gel analysis showed the presence of a long smear of products, even in as early as the end of the first DH-RH cycle. The smudge (as opposed to a clear single extension product band), observed on the top of the RNA primer band, is suggestive of possible multiple additions to the primer. Interestingly, this was seen, both, in the presence and absence of the RNA template (control reaction) in the starting reaction mixture (Fig. 3b right vs left panels). Similar results were also observed when using as low as 1 mM of 5′-rMP in the starting reaction mixture, along-with 1 mM sulphuric acid and nanopure water as rehydration agent (data not shown). In contrast, in a parallel reaction wherein only 5′-rMP was cycled under aforementioned conditions, no detectable fluorescent signal was obtained upon gel analysis. Significantly, the primer extension product from the addition of a single 5′-rMP monomer, and that obtained when using 5′-AMP as monomer (discussed in previous results), were found to be running at the same level upon gel analysis (Fig. S5). This provided an indirect confirmation of the possibility that in our main reactions involving 5′-NMPs, the addition of an abasic nucleotide was resulting in the primer extension band that did not run to the same length as an intact 21-mer product.

Discussion

Chemically driven oligomerization and nonenzymatic template-directed reactions are uphill processes as the formation of phosphodiester bonds between the non-activated nucleotides is not a spontaneous reaction. In order to demonstrate polymer formation from non-activated nucleoside monophosphates, previous studies have reported the use of alternate DH-RH cycles at elevated temperatures, to facilitate the aforementioned processes[9–11,15]. The dehydration phase helps in concentrating the monomers, thus increasing the chances of bond formation between them[12]. Subsequent rehydration helps in the random re-distribution of, both, the monomers as well as the resultant oligomers. This increases the probability of formation of longer oligomers over multiple wet-dry cycles. Therefore, by trapping the polymers in a kinetic trap, wherein the rate of polymer formation exceeds the rate of polymer degradation, alternate DH-RH cycles ultimately help in yielding polymers from the starting monomers. The elevated temperature of the reaction decreases water activity, thus favouring the forward reaction by overcoming the energy barrier for the phosphodiester bond formation. Furthermore, the efficiency of these reactions has been found to be higher at low pH. A suggested mechanism, similar to that of acid-catalyzed esterification, hypothesizes the role of protonation of the nucleotide at lower pH, thus facilitating the formation and accumulation of oligomers over repeated DH-RH cycles[15]. Previous studies have also characterized the use of lipids in such reactions, which reduces the unfavourable back reaction and results in an apparent increase in the yield of oligomers[9]. Phospholipids are known to form multi-lamellar sandwiches of alternating hydrophilic and hydrophobic layers in the dehydrated phase of the reaction. Studies have shown the nucleotide monomers to get confined in the hydrophilic layers, thus increasing their chance of participating in a phosphodiester bond formation[16,17]. Significantly, during the subsequent rehydration phase, some of the oligomers that form in the dehydration phase could get encapsulated in the lipid vesicles, a step that is considered crucial for the emergence of protocells.

Faithful replication of any functional (catalytic) polymer that would have resulted from the aforementioned polymerization reactions under DH-RH scenarios, would have been a crucial step for them to have consistently acted as a catalyst. Only one study has been reported so far, where nonenzymatic replication reactions have been attempted under similar DH-RH reaction regimens. The authors of this study, however, used a mixture of 2′-deoxyribonucleoside 5′-monophspahtes and a single stranded DNA, as template, to look at nonenzymatic replication[15]. We, therefore, set out to test if the DH-RH conditions would also support template-directed primer extensions using RNA monomers, which is directly pertinent to the 'RNA World hypothesis', with imminent implications for the emergence of early life.

To begin with, the extension of a pre-existing RNA primer (Amino-G primer) was analyzed. The 3′- amino group, present at the primer's extending terminus, acts as a better nucleophile than the contemporary hydroxyl group, thus yielding detectable amounts of the extended product (Fig. S1). Extension of the Amino-G primer was observed upon multiple cycles of rehydration and dehydration. The presence of lipids in the reaction mixture seemed to confer some protection under our experimental conditions. This might potentially result from sequestration of the RNA within the hydrophilic layers of the multilamellar structures that are formed during the dehydrated phase[16]. The newly formed product's retention time on the denaturing gel, however, was found to be in between that of the unreacted primer (20-mer) and that of known 21-mer control RNA (Fig. 1a). This is indicative of the product possibly having an abasic site as the charge on the product seems to be more than that of the 20-mer primer from our gel analyses. Previous results have demonstrated the formation of RNA-like, depurinated oligomers from non-activated 5′-AMP, upon multiple DH-RH cycles[11]. Our results suggest that even template-directed primer extension reactions are potentially susceptible to base loss during bond formation. This seems to be more prevalent on the 3′-end of the growing polymer, because if the 5′-end of the RNA primer used in this study happened to be degraded, it cannot be visualized on the gel as the Cy3 fluorescent tag is attached to the 5′-end of the primer. To further confirm the aforementioned, we subjected the 21-mer intact primer to similar reaction conditions. We observed that the 21-mer primer degraded directly to a 20-mer oligomer (corresponding to the starting primer length), over multiple DH-RH cycles (Fig. S6). Additionally, no intermediate product was observed in between the bands corresponding to a 20-mer (starting primer) and 21-mer (extended product) length RNA. This strongly hints at the potential loss of a base on the incoming 5′-NMP ribonucleotide. This possibility could occur just before or during the addition of the incoming nucleotide, onto the pre-existing primer, during the course of our DH-RH reactions. During template-directed primer extension, the template is

thought to stabilize the incoming base via near neighbour interactions. However, templating effect does not seem to be able to prevent the potential loss of base in our reactions, further suggesting that it could result from the chemical reaction that takes place during the actual bond formation between the incoming nucleotide and the primer. Additionally, the temperature at which the DH-RH cycles were carried out was reduced systematically to check for any effect on this phenomenon. However, the apparent loss of the informational entity seemed to persist even at temperatures as low as 50 °C. Therefore, it seems that the low pH of the reaction might possibly be the predominant reason for this effect in our reactions.

We also tried different rehydrating agents other than sulphuric acid to study their effect on template-directed primer extension reactions using 5′-NMPs. In total, four mineral acids (viz. H_2SO_4, HCl, HNO_3, and H_3PO_4) and two organic acids (viz. formic acid and acetic acid) were evaluated. Typically 1 mM of mineral acid was used while up to 10 mM of organic acids had to be used to compensate for their lower intrinsic strength. Among the mineral acids used, we observed the extension of the primer only when either sulphuric acid or phosphoric acid was used as the rehydration agent. This might potentially be due to the thermally labile nature of the other mineral acids that we tried, as the reactions were carried out at 90 °C. Among the organic acids that were used, we observed primer extension only when formic acid was used. This might be attributed to the fact that formic acid is a stronger acid than acetic acid, and hence might be more efficient in bringing down the pH of the reaction. Significantly, the extended product's retention time on the gel continued to not correspond to that of the expected intact N + 1 nucleotide extension. Furthermore, we checked whether addition of ammonium cations to our reaction might facilitate efficient template-directed primer extension as was observed in a previous study[23]. However the presence of NH_4^+, either alone or along with Na^+ ions, did not result in an apparent increase in the efficiency of the extended product formation (Fig. S4). Furthermore, the presence of ammonium ions in the reactions mixture, also lead to greater RNA degradation under our DH-RH reaction regimen.

Finally, to analyze whether the repeated addition of sulphuric acid after every DH-RH cycle, was causing what seemed like an apparent base loss, due to higher concentrations of acid in the reaction, nanopure water was used instead as the rehydrating agent. In these reactions, sulphuric acid was added to the starting mix to reduce the pH to around 3.5 and was kept at a constant 1 mM throughout the course of the DH-RH cycling. Even upon reduction in the amount of acid in the reaction mixture, the extended product continued to run lower than the expected intact N + 1 primer product (Fig. 3a). Interestingly, the extension of the RNA primer continued to occur even while using as low as 1 mM of the monomer. We believe this result is significant, as the existence of highly concentrated pools of nucleotides on the prebiotic Earth is not considered very likely.

Despite varying several parameters of the reaction scheme, including pH, temperature, nature and strength of the rehydration solution, and that of the monovalent cations, the extended primer product had a retention time on the gel that was lower than that of a known 21-mer control. This indicated a strong probability of a base loss over the reaction course and it seemed that the primer was getting extended by the addition of only a sugar-phosphate moiety. Given this possibility, we also analysed whether ribose-5′-monophosphate (5′-rMP) by itself could extend the pre-existing RNA primer. Oligomer formation from 5′-rMP monomers, under alternate DH-RH conditions, has been previously reported[11]. Upon cycling the RNA primer with 5′-rMP, long streaks were observed on the denaturing gel. These resultant products were observed even at the end of the first DH-RH cycle itself (Fig. 3b 'Cyc 1' lane). This seemed to indicate a higher polymerization potential for the rMP monomers in comparison to contemporary ribonucleotides. Also, the presence of long streaks on the denaturing gel indicated possible multiple additions of rMP to the RNA primer. Both linear and cyclic oligomers of rMP have been previously reported to result upon multiple DH-RH cycles[11]. We suspect that the RNA primer is possibly getting extended by a mixture of, both, linear and cyclic rMP molecules, and this, in turn, resulted in poor resolution of these polymers on the denaturing gel. Furthermore, as a control, 5′- rMP molecules were cycled alone (in the absence of the primer and template) and analysed using gel electrophoresis. No streaks were observed upon fluorescence imaging of the gel, and even upon SyBrGold staining (due to lack of bases), confirming the possibility that the rMP oligomers were most likely covalently linked with the RNA primer in the test reactions involving 5′-rMP. This is significant as such hybrid molecules are hypothesized to have been important for exploring a vast chemical space of plausible alternate nucleobases that might have resulted in the formation of primitive information polymers of a pre-RNA World[24]. This hypothesis has gained support from recent studies that have demonstrated the addition of non-contemporary bases to ribose or ribose monophosphate, at both ambient conditions[25] and at high temperature regimes[26,27].

In conclusion, this study highlights the importance of systematically characterizing prebiotically pertinent reactions in simulated laboratory conditions to better understand how they might advent in niches that are thought to have supported the origin of life on prebiotic Earth. These findings particularly have important implications for discerning relevant mechanisms in such niches that would have eventually enabled the transition from chemistry to biology on Earth. Furthermore, the formation of hybrid polymers in our 5′-rMP based reactions, strengthens the likelihood of pre-RNA World hypothesis. Also, the possibility that the same sugar phosphate backbone can potentially sample both contemporary as well as alternate nucleobases, as discussed above, is indicative of the important role such mixed backbones might have played in the transition from a putative pre-RNA World to an RNA World.

Material and Methods

Materials. The disodium salts of all four 5′-NMPs viz. Adenosine 5′-monophosphate (AMP), Guanosine 5′-monophosphate (GMP), Uridine 5′-monophosphate (UMP), Cytidine 5′-monophosphate (CMP), and ribose 5′-monophosphate (5′-rMP) were purchased from Sigma-Aldrich and used without further purification. The RNA primers used in this study are Amino-G (acquired from Keck laboratory, Yale, USA) and Hydroxyl-G (acquired from Thermo Fisher Scientific, USA) primers. Both these primers had fluorescence labels on their 5′-end for facilitating their detection on polyacrylamide gel electrophoresis (PAGE). The Amino-G primer

terminates with a 3′-amino-2′, 3′-dideoxynucleotide (Metkinen, Finland) while the Hydroxyl-G primer terminates in a canonical ribonucleotide. The sequences of the primers and templates are as given below, with the template base indicated in bold:

Primer Amino-G: 5′ GG GAU UAA UAC GAC UCA CUG-NH_2.
Primer Hydroxyl-G: 5′ GG GAU UAA UAC GAC UCA CUG.
Template MisInc_C: 5′ AGU GAU CU**C** CAG UGA GUC GUA UUA AUC CC.
Template MisInc_G: 5′ AGU GAU CU**G** CAG UGA GUC GUA UUA AUC CC.
Template MisInc_A: 5′ AGU GAU CU**A** CAG UGA GUC GUA UUA AUC CC.
Template MisInc_U: 5′ AGU GAU CU**U** CAG UGA GUC GUA UUA AUC CC.

The phospholipid used in this study, namely 1, 2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) was purchased from Avanti Polar Lipids Inc. All other reagents used were of analytical grade purchased from Sigma-Aldrich.

Methods. *Reaction setup.* A typical reaction was set up by annealing 0.65 μM of primer and 1.3 μM of template to each other by heating at 95 °C for 5 minutes, followed by cooling at room temperature (RT). 100 mM Tris (pH 7.0) and 200 mM NaCl were then added to the annealed primer-template complex. A suspension of DLPC vesicles was prepared in nanopure water by hydrating the dried film of lipid, followed by extrusion of the vesicles through a 100 nm membrane using a mini extruder (Avanti Polar lipids Inc). DLPC vesicles and 5′-NMPs were added to the reaction mixture at a final concentration of 5 mM and 10 mM, unless otherwise specified. The reaction mixture was then allowed to dry at elevated temperature. The dried mixture was typically rehydrated after 30 minutes. After allowing the rehydration for 5 minutes, the procedure was repeated for several DH-RH cycles. A sample for analysis was removed in the rehydration phases at regular intervals.

Analysis of reaction products. The samples were collected in TBE buffer containing 8 M urea and 100 mM EDTA. The sample volumes were adjusted to a standardized amount to compensate for the degradation of RNA primer that occurs over multiple cycles of DH-RH. A competitor RNA, without any fluorescence tag and with a sequence that is exactly similar to that of the tagged primer, was used (5′ GG GAU UAA UAC GAC UCA CUG). This was added in at least 10 times excess to the reaction samples, to successfully separate the fluorescent primer in question for unhindered gel analysis. The extended primer products were analyzed on 20% denaturing PAGE and the gels were imaged with a Typhoon Trio plus imager (GE Healthcare) at 550PMT and 100 micron resolution setting, using the 532 nm excitation laser. All the gels were scanned by setting the region of interest (ROI) in the image acquisition software to the lower half of the gel region in the 17 × 15 cm gel, which typically has all bands pertinent to our reaction. This allowed for the efficient detection of the fluorescently labelled primer and the resultant extended products. The gel images were subsequently processed using ImageQuant v5.2 software to minimally adjust the contrast and also for quantification purposes in some cases.

Data Availability Statement

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

1. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).
2. Orgel, L. E. Prebiotic chemistry and the origin of the RNAworld. *Crit. Rev. Biochem. Mol. Bio.* **39**, 99–123 (2004).
3. Li, L. *et al.* Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J. Am. Chem. Soc.* **139**, 1810-1813.
4. Deck, C., Jauker, M. & Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).
5. Yi, R., Hongo, Y. & Fahrenbach, A. Synthesis of imidazole-activated ribonucleotides using cyanogen chloride. *Chem. Comm.* **54**, 511–514 (2018).
6. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible condition. *Nature* **459**, 239–242 (2009).
7. Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* **132**, 16677–16688 (2010).
8. Becker, S. *et al.* A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
9. Rajamani, S. *et al.* Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig. Life Evol. Biosph.* **38**, 57–74 (2008).
10. DeGuzman, V., Vercoutere, W., Shenasa, H. & Deamer, D. Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *J. Mol. Evol.* **78**, 251–262 (2014).
11. Mungi, C. V. & Rajamani, S. Characterization of RNA-like oligomers from lipid-assisted nonenzymatic synthesis: implications for origin of informational molecules on early earth. *Life* **5**, 65–84 (2015).
12. Deamer, D. Liquid-crystalline nanostructures: organizing matrices for nonenzymatic nucleic acid polymerization. *Chem. Soc. Rev.* **41**, 5375–5379 (2012).
13. Harshe, Y. M., Storti, G., Morbidelli, M., Gelosa, S. & Moscatelli, D. Polycondensation kinetics of lactic acid. *Macromol React Eng* **1**, 611–621 (2007).
14. Mamajanov, I. *et al.* Ester formation and hydrolysis during wet—dry cycles: generation of far-fromequilibrium polymers in a model prebiotic reaction. *Macromolecules* **47**, 1334–1343 (2014).
15. Olasagasti, F., Kim, H. J., Pourmand, N. & Deamer, D. W. Non-enzymatic transfer of sequence information under plausible prebiotic conditions. *Biochimie* **93**, 556–561 (2011).
16. Toppozini, L., Dies, H., Deamer, D.& Rheinstädter, M.C. Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: Insights into assembly of nucleic acid for primitive life. *PLoS One* **8**, https://doi.org/10.1371/journal.pone.0062810 (2013).
17. Himbert, S., Chapman, M., Deamer, D. W. & Rheinstädter, M. C. Organization of Nucleotides in Different Environments and the formation of Pre-Polymers. *Sci. Rep.* **6**, 31285, https://doi.org/10.1038/srep31285 (2016).

18. Rajamani, S. *et al*. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.* **132**, 1008–1011 (2010).
19. Leu, K. *et al*. Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. *J. Am. Chem. Soc.* **135**, 354–366 (2013).
20. Bapat, N. V. & Rajamani, S. Effect of co-solutes on template-directed nonenzymatic replication of nucleic acids. *J. Mol. Evol.* **81**, 72–80, https://doi.org/10.1007/s00239-015-9700-1 (2015).
21. Zoltewicz, J. A., Clark, D. F., Sharpless, T. W. & Grahe, G. Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. *J. Am. Chem. Soc.* **92**, 1741–1750 (1970).
22. Rios, A. C., Yu., H. T. & Tor, Y. Hydrolytic fitness of N-glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. *J. Phys. Org. Chem*. https://doi.org/10.1002/poc.3318 (2014).
23. Da Silva, L., Maurel, M. C. & Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *J. Mol. Evol.* **80**, 86–97 (2015).
24. Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, L. D. The origin of RNA and "My Grandfather's Axe". *Chem. Biol.* **20**, 466–474 (2013).
25. Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nat. Commun.* **7**, 11328, https://doi.org/10.1038/ncomms11328 (2016).
26. Mungi, C. V., Singh, S. K., Chugh, J. & Rajamani, S. Synthesis of barbituric acid containing nucleotides and their implications for the origin of primitive informational polymers. *Phys. Chem. Chem. Phys.* **18**, 20144–20152 (2016).
27. Kim, H. J. & Benner, S. A. Prebiotic glycosylation of uracil with electron-donating substituents. *Astrobiology* **15**, 301–306 (2015).

Acknowledgements

The authors are grateful to IISER, Pune, for funding this research and providing the requisite organizational support. N.B. is a recipient of a fellowship from CSIR, Government of India. We wish to extend our heartfelt gratitude to Dr. Ramanarayanan Krishnamurthy (TSRI, CA, USA) and Chaitanya Mungi (IISER, Pune) for critical discussions and their valuable inputs.

Author Contributions

N.B. and S.R. designed the experiments. N.B. performed the experiments. N.B. and S.R. analysed the data, wrote and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-33157-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.