1

Membrane catalysed formation of nucleotide clusters and its role in the origins of life

Rajlaxmi Saha,[†] Prathyush Poduval,[‡] Jayashree Nagesh,[¶] and Anand
Srivastava^{*,§}

*[†]Department of Biological Sciences, Indian Institute of Science Education and Research,
Kolkata*

2

Campus Road, Mohanpur, West Bengal 741246

[‡]Indian Insititute of Science, Bangalore

C.V. Raman Road, Bangalore Karnataka 560 012, India

[¶]Solid State & Structural Chemistry Unit, Indian Institute of Science, Bangalore

C.V. Raman Road, Bangalore Karnataka 560 012, India

[§]Molecular Biophysics Unit, Indian Insititute of Science, Bangalore

C.V. Raman Road, Bangalore Karnataka 560 012, India

E-mail: anand@iisc.ac.in

Abstract

Non-enzymatic polymerisation explained how life could assemble spontaneously on the early Earth. Prior research has discovered that the non-enzymatic polymerisation can be mediated by various environmental settings, most widely accepted are the reactions driven by the cycles of hydration and dehydration and organising matrices such as clay, salt, membranes and mineral surface. Membranes have a unique feature which prepares it as a sensitive organising matrix, as compared to mineral surface and clay, which transforms its phase with small changes in temperature, hydration and composition of lipids. Multi-component membranes can show phase separation where two or more distinct phases can co-exist. To the best of our knowledge, this is for the first time the influence of different phases of the membrane in the nucleotide organisation and polymerisation in a simulated prebiotic setting in this work have been explored. This study encompasses the energetics of inserting a mononucleotide, UMP, in distinct membrane settings which is probed utilising the PMF. The preferential partitioning of UMP is also investigated; which is coupled to the location of the insertion and the composition of the PSM/DOPC/Chol membranes. Combined with umbrella-sampling calculations, AA-MD simulations were performed to estimate the role of the monophasic and diphasic membrane in modifying the behaviour of UMPs and their clustering mechanism. A mathematical model was also developed based on lattice model and random walk and some of the parameters were trained from the AA-MD simulations. The results indicated that the membranes were capable of concentrating nucleotides and organised into anisotropic clusters, and the thermodynamic studies along with mathematical model showed that it had a sharp preference for the L_o/L_d domain interface for clustering.

Keywords: lipids, phases of membranes, domain interface, uridine monophosphate, nucleotide clusters, origin of life, astrobiology

Author Contributions: RS and AS conceived and designed the research; RS

34 and PP performed research; AS provided computational resources; RS, PP, and AS
35 analyzed data; and RS and AS wrote the paper.

Significance/Authors' Summary

The organisation of the mononucleotides and the progress in the formation of clusters in a lipid environment highlighting the importance of its phases has been the emphasis of this work. The fluidity of the lipids and 2D confinement of the membranes enable the nucleotides to form anisotropic clusters, consisting of 2 to > 20 mononucleotides. The UMPs prefer interacting with the L_o/L_d domain interface, followed by the L_d domain of the membranes. The catalytic formation of nucleotide clusters and pre-polymers is an essential inaugural step in the “*origin of life*” since the availability of a catalyst, in this work *phospholipid bilayers*, and the mononucleotides conceive the formation of the RNA-like molecules which initiated the process. Thus our work has highlighted a phenomenon in the scenario which describes the precursor to the RNA-world.

47 Introduction

48 One of the most extensive unsolved scientific problems of this century is understanding
49 the origin of life (OoL) on the prebiotic Earth. There is an inadequacy of a widely ac-
50 cepted definition of life and its phenomenon¹, but there are more obstacles to gain a
51 complete insight regarding the process. It is agreed upon that the OoL have emanated
52 from a long process, involving a series of processes, for the complexity of the cell, could
53 not have arisen suddenly. OoL should have initiated with simple abiotic processes, asso-
54 ciated with energy and matter and is based on concepts such as entropy, complexity and
55 information. The diverse views and disparate avenues of research, including experiments
56 and theory, has led to the existence of considerable and eclectic theories on OoL.

57
58 One of the hypothesis, Thermodynamic Dissipation Theory of the Origin and Evolution
59 of Life²⁻⁴, suggests that the OoL is the microscopic dissipative structuring of organic pig-
60 ments and their proliferation over the entire Earth surface. An alternative chemical paths
61 to life⁵ suggests that biomolecules such as lipids and amino acids arise from homologation
62 hydrogen cyanide and its derivatives, and hence cellular life could have started from com-
63 mon chemistry. On the contrary to this theory, Alex Rich first proposed that RNA, as a
64 biopolymer acting in two capacities, might have sustained both genetics and catalysis at
65 the OoL⁶. The capability of the RNA to do both genetics and catalyse would not necessi-
66 tate multiple biopolymers to spontaneously emerge together^{7,8}. Amongst all theories, the
67 “RNA world” has garnered wide acceptance because of its compelling evidence^{9,10}. It has
68 gained momentum due to the well-known roles of RNA in terrain metabolism, catalytic
69 properties, capability to converge the conflicting demands of phenotype and genotype
70 and lastly the absence of a competitive model where metabolism is placed first¹⁰. In this
71 regard, RNA is thought to be the candidate for the first genetic material in the prebiotic
72 Earth⁹⁻¹³, which would have enabled the transition from chemistry to biology.

73
74 The processes in the modern biology are governed by polymers, hence the origin of
75 the informational molecules must have been an important stepping stone in the OoL

on Earth. One of the challenges in studying abiogenesis is the involvement of multiple biopolymers in basic function of reproduction and metabolism¹⁴. The “RNA first” model avoids not only the chicken-or-egg problem but also the dual improbability that multiple biopolymers coincidentally emerged to interact in an encoder-encoded relationship¹⁰. An ongoing investigation for researchers studying the molecular OoL^{15–17} research has been the understanding of the formation and synthesis of the first polymers of nucleic acids. The comprehension of how RNA-like molecules were produced and its emergence in the absence of the catalysts and metabolism is elementary but rapidly evolving^{18–27}. It was believed that there must be a process which sufficiently concentrates organic compounds^{28,29}.

Nucleic acids form the foundation of life, and nucleotides form their building block. Nucleotides^{30,31} are organic molecules composed of three distinctive chemical subunits, a phosphate group and a nucleoside, which consists of a nitrogenous and a sugar (ribose or deoxyribose) group. The nonenzymatic polymerisation of RNA, a condensation reaction where phosphodiester bonds are formed, is thought to be a crucial step in origin of simple life forms in the prebiotic time³². Nearly a decade ago, the concept of guided polymerisation was introduced^{33–35}, and it was later discovered that RNA-like polymers could be synthesised non-enzymatically^{36,37} in different settings reproducing prebiotic cycles such as fluctuations in hydration^{20,28,38–40} and the presence of various organising matrix^{20–25,41}; conditions which procreate prebiotic environment. Nucleotides in solution are highly diffusive, but diverse substrates provide different type of confinement environments, such as phospholipid bilayers (membranes) allow 2D confinement^{20,25,26}, salt confines them in between salt crystallites²⁴, or nucleotides can confine in the layered structure of clay^{21,22,42–44}.

Lipids are organic molecules which have a hydrophobic, “*water-fearing*”, tail group and a hydrophilic, “*water-loving*”, head group and hence amphiphilic in nature. Due to the hydrophobic effect, where the hydrophilic head interacts with water protecting hydrophobic

105 tails, lipids can form micelles, bilayers, and liposomes in an aqueous environments; this
106 mechanism also depends on the type of lipids such as their charge, chemical composition,
107 cross-sectional area of the head groups and tails. A special type of lipid aggregate, *mem-*
108 *branes*, are formed when the lipid head group and tail groups have similar cross-sectional
109 areas. In a membrane, the hydrophobic tails of each monolayer is orientated towards
110 each other's tails where the tail groups of the upper leaflet and lower leaflet interact and
111 membrane structure is affected by the composition of the head and tail groups of the
112 lipids; it determines their packing density, membrane thick, deuterium order parameter,
113 surface area per lipid (APL) among other things. Lipids antecede the OoL⁴⁵. Enough
114 organic materials were available in the prebiotic settings which allowed amphiphiles to
115 self assemble into membranes or membranous vesicles⁴⁶. For cellular life to evolve, a
116 system must have defined boundaries from its environment⁴⁷⁻⁵⁴. Membranes would pro-
117 vide a plausible boundary structure that would have initiated cellular compartmentation.
118 The lipid membranes envelopes organic molecules and biopolymers; thus protecting the
119 growing genetic material. Therefore, the function of membranes as a container would be
120 indispensable to life. The importance of membranes in the abiogenesis is immense and
121 its role as an organising matrix is has also been studied in the prebiotic condition settings
122 of the Earth^{20,25,26,55}. The ordering of mononucleotides, Adenine Monophosphate (AMP)
123 and Uridine Monophosphate (UMP), has been shown in 1,2-dimyristoyl-sn-glycero-3-
124 phosphocholine (DMPC) lipid matrix using X-ray diffraction^{25,26}.

125
126 The plasma membrane is a self assembly of lipids, sterols and proteins⁵⁶ with the unique
127 ability to undergo phase transitions in different environmental conditions like tempera-
128 ture and hydration. A familiar and more straightforward analogy is the solid, liquid and
129 gaseous phases of simpler systems. In membranes, the "liquid-ordered" (L_o) phase, com-
130 prising mostly saturated lipids, has tight packing of the lipids; the sterol molecules provide
131 the relative rigidity to the arrangement. Ordered phase is characterized by high confor-
132 mational order and translational disorder. This is reflected in the low deuterium order
133 parameters of the lipids, higher membrane thickness and lower area per lipid. "Liquid-

disordered" phase (L_d) is a highly fluid state where the movement of the lipids are unhindered and can cover the entire surface of the membrane. The lipids in L_o have recognizable subdiffusive behaviour when compared to L_d phase^{57,58}. The lack of long-range intermolecular conformation order is connotative of the L_d phase. It is distinguished by its high deuterium order parameter, reduced membrane thickness and correspondingly large APL owing to the irregular packing of the individual lipids. The phase transition is a consequence of the change in entropy in the system; and in response to changed the free energy, the lipids are reorganised. Spatial and dynamic heterogeneities in membrane arise from their differential molecular interaction with diverse constituents^{59,60}. These heterogeneities promote phase separation within membrane constituents which forms distinct domains characterised by their length and timescales. The domains share role in vital biological functions^{61–64} such as cell signalling^{65,66}, membrane⁶⁷ and vesicle trafficking⁶⁸, and protein complex assembly^{65,69}.

Previously studies have shown the various organisation and ordering of mononucleotides has been shown in lipid environments^{20,25,26,55} but their phases and corresponding chemical composition of the membrane has been unexplored. The unique characteristics of membranes are that it can *exist* in many phases, owing to its diverse chemical composition^{59,70–72}, and have different properties such as diffusion^{57,58} and lipid packing. Our research focused on employing a unique concept in regards to understanding the nucleotide polymerisation in the molecular OoL for the first time. The L_d phase is more fluidic, the lipids are more diffusive, than the lipids in L_o phase. A phase separated membrane can compensate the necessity of the cycle of hydration and dehydration for concentrating and organising nucleotides as a possible organising matrix; the fluidic L_d phase could replace the hydration or the solution phase, and the L_o phase can act as the dehydrating phase. Thus membrane bilayer would have aided in the non-enzymatic polymerisation of various biomolecules, including nucleotides, in the prebiotic Earth settings. Hence the importance and relevance of membranes would be indisputable during abiogenesis, so we studied its role and the importance of its phases computationally in prominent prebiotic

processes, nucleotide organisation and non-enzymatic polymerisation of nucleotides. This study focuses on understanding the role of different phases of membrane on formation of pre-polymers or clusters of nucleotides. Towards this, we applied AA-MD simulations and advanced sampling method, called umbrella sampling, for thermodynamics studies. Further, we developed a mathematical model for the nucleotides and the membranes in the systems and their interactions and used data collected from the AA-MD simulations to parametrise the model.

The rest of the article is organised as follows. In Materials and Methods, we discuss our simulation, mathematical model and characterisation techniques. We discuss our main results in the Results and Discussion. Finally, we conclude with a summary and future directions in Conclusion.

Materials and Methods

In this section, we describe the details of the simulation methodology and the mathematical modelling. We begin by summarising the details of AA-MD simulations and the data processing in the first subsection. Next we discuss the advanced sampling method called umbrella sampling used in the work, followed by the mathematical construct we developed to model the nucleotide interactions with lipids and other nucleotides.

All atom MD simulations

All the membranes were borrowed from Edward Lyman’s group at the University of Delaware. They were generated on the Anton supercomputer. The three distinct Palmitoylsphingomyelin (PSM)/ 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC)/ Cholesterol (Chol) bilayer were in different phases, namely L_o , L_d and L_o/L_d (Table 1). The nucleotides, UMP, were arranged in arrays and simulated in a PSM/DOPC/Chol bilayer (Figure 5), which existed in varying composition and hence phases, and the systems have been named accordingly and the details described in Table 2. Here on, the systems will be referred to as mentioned in Table 2.

All the simulations were performed using GROMACS 5.1.4⁷³ molecular dynamics engine employing CHARMM-36⁷⁴ force field for 200 ns. The Molecular Dynamics (MD) simulations were performed in the NPT ensemble at atmospheric pressure (1 bar), with a time step of $\delta t = 2$ fs and using the Leap-Frog integrator⁷⁵. The temperature was held constant with the Nosé-Hoover thermostat^{76,77} at 295 K and pressure was maintained by Parrinello-Rahman barostat⁷⁸ at 1 bar with relaxation times of 1.0 ps for each case. Standard procedures were used to include the influence of long-range electrostatics⁷⁹ and Lennard-Jones (LJ) interaction calculations^{80,81}. For van der Waals interactions, cutoff and shift distances of 1.2 nm and 0.9 to 1.2 nm, respectively.

Post processing

The AA-MD data evaluation was conducted using the extensive library for trajectory analysis tools available in GROMACS⁷³. Other techniques implemented in our study have been discussed in this subsection.

Relative interaction preference

To estimate the interaction polarity of nucleotide with L_o and L_d lipid membranes, we calculated the total interaction energy (Coulombic and van der Waals interactions) between the nucleotide and the different phases of the membranes. The relative interaction preference^{82,83} for the nucleotides will be given by:

$$R = \frac{E_{Lo}}{E_{Lo} + E_{Ld}} \quad (1)$$

where $R=1$ indicates entirely L_o preference and $R=0$ imply complete L_d preference.

DBSCAN algorithm

DBSCAN, or Density-Based Spatial Clustering of Applications with Noise, a clustering algorithm is a density oriented approach to clustering⁸⁴ and was implemented using Python programming⁸⁵. It is deterministic, always generates the same clusters when given the same data in the same order. The DBSCAN algorithm uses two parameters:

1. *min_samples* - The minimum number of points together in a region to be treated as dense. It determines the tolerance of the algorithm towards noise.
2. *eps* (ϵ) - It is the distance measure that is used to find the points in the neighbourhood of any point. It is an important parameter, as it determines the adjacent points of any given point. If the value is chosen to too small, most data will not be clustered, where as, if value is too high the entire data set will be merged into a huge cluster.

A cluster is defined as a set of core samples, close to each other and a set of non-core samples; they are on the fringes of the cluster, that is close to the core samples.

225 The SAXS profile were computed using Debye formula⁸⁶ (Equation 2) implemented in
 226 Fast X-Ray Scattering (FoXS) webserver⁸⁷. The webserver calculates the profiles based
 227 on the given molecular structure with all the inter-atomic distances.

$$I(q) = \sum_{i=1}^N \sum_{j=1}^N f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}} \quad (2)$$

228 where $I(q)$ is the intensity as a function of momentum transfer q , d_{ij} is the distance
 229 between atoms i and j , N is the number of atoms in the molecule and form factors $f_i(q)$
 230 and $f_j(q)$.

231 Umbrella Sampling

232 Umbrella Sampling simulations were performed in three distinctive environments: a
 233 ternary mixture constituting of PSM, DOPC, and Chol at various molar ratio. Fig-
 234 ure 2A demonstrates the three membrane systems considered and an umbrella sampling
 235 simulation was envisioned for each combination of the membrane environment and UMP.
 236 For L_o and L_d membranes, the reaction coordinate along the bilayer normal, z , was uti-
 237 lized and the distance between the membrane midplane and the solute molecule was used
 238 as the collective variable. In the phase-separated membrane, the reaction coordinate was
 239 along the x -axis on the phosphate plane, the distance between the centre of mass on the
 240 L_d to the centre of mass of lipid headgroups on the L_o was considered. The individual
 241 simulation consisted of at least 50 windows separated every 0.1 nm, where the force con-
 242 stant of the harmonic restraint adjusted to $1000 \text{ kcal mol}^{-1} \text{ nm}^{-2}$.

243

244 All the systems were energy minimized and equilibrated and then continued for pro-
 245 duction runs. The final production runs were carried out in NPT ensemble at a pres-
 246 sure $P=1$ bar with relaxation times of 1.0 ps for each case using the Parrinello-Rahman
 247 barostat⁷⁸ and the temperature was kept constant at $T=295\text{K}$ using the Nosé-Hoover
 248 thermostat^{76,77}. Estimation of PMF profiles was carried out by applying the weighted

histogram method (WHAM)⁸⁸ and the relative errors were estimated by the bootstrap-
ping analysis⁸⁹ as enforced in GROMACS⁹⁰.

Mathematical modelling

We developed a spatio-temporal model (Figure 1) which describes the phenomenon of nucleotide cluster forming on a lipid membrane. This phenomenological model was inspired from the theory of lattice models⁹¹. The primary idea is to identify degree of freedom as discrete variables that keep moving around a lattice, and all interactions are specified only between nearest neighbour pairs, and such concepts have been employed in previous works^{92–95}. Most noted is to describe polymers^{92,93} on a first approximation basis where each monomer inhabits a single site, and double occupation of a site is forbidden. A polymer chain will form when there exists a sequence of connected occupied sites, and the interaction is limited between nearest neighbour monomers.

The model we developed for nucleotide aggregation is similar in spirit to the polymer model, with some added intricacies. The fundamental idea is intact; the nucleotide moves around on a $L \times L$ square matrix which represents the lipid bilayer and the interaction effect from one nucleotide extends to the nearest sites. In addition, we added a parameter, p , that determines the rate at which the nucleotide hops around on the lattice and its value depends on the phase of the membrane the lattice site belongs to, either ordered or disordered.

The organisation of the UMP in different phases of the membrane is computationally simulated using random walk^{96–98} on a two dimensional $L \times L$ lattice with periodic boundary conditions (PBC)⁹⁹ and interaction amongst the N nucleotides. The nucleotides occupy individual lattice sites and at each time frame their location is asynchronously updated. The model has five main sets of parameters; N - The number of nucleotide, L - the size of the square lattice, p - the hop parameter, τ, τ_1 - the interaction parameters and I_0, I_1, I_2 - the interface interactions.

At each time frame, two key decisions are made by the nucleotide. First decision includes whether the nucleotide hops or not, followed by the second choice which location to hop. These opportunities are modelled as nucleotide interaction between the nucleotide and the lipid bilayer modelled as follows:

(a) Nucleotides position of the membrane which is decided by probability parameter p .

(b) Interaction between nucleotides determined by parameters, τ and τ_1 .

The probability that a nucleotide decided to stay at its position at a time frame is determined by a Bernoulli random variable with probability $p = 1 - q$. The parameter p depends on the phase of the lipid, a lower value is assigned for the disorder phase where a higher value is allowed for ordered phase. Notations p_{ord} and p_{dis} refer to the probability parameter in L_o and L_d phase of the membrane respectively. Note that these are motivated by kinetic and not energetic considerations.

290

The interactions between the nucleotides are facilitated via two parameters τ , τ_1 and a matrix referred to as $POT[L][L]$ which is initialized to 0 throughout the lattice. At a given time frame, a nucleotide at the position (x, y) undergoes (a) $POT[x][y]$ increased by τ and (b) $POT[x \pm 1][y]$, $POT[x][y \pm 1]$ and $POT[x \pm 1][y \pm 1]$ increased by τ_1 . This can be described as an *attractive forcefield* generated at its own point, and at all the other points on the square that surrounds the nucleotide. The probability τ can be related to the energy difference, E , between the destination site and the the origin site as $\tau \sim e^{-\frac{E}{RT}}$. These are not motivated by energetic considerations but the values of τ is chosen higher than τ_1 such that a nucleotide has a increased chance of residing on a site directly adjacent of the second nucleotide rather than on the next nearest site.

301

At each time frame, the position of the nucleotides are updated asynchronously based on the probability p based on which nucleotides decision to reside. If the nucleotide decides to move, then it hops from its current position (x, y) to one of the nearest neighbour

305 $(\mathbf{x}', \mathbf{y}')$ with probability given as follows:

$$\frac{\text{POT}[\mathbf{x}'] [\mathbf{y}'] + 10 - \tau_1}{\sum_{i=\pm 1} (\text{POT}[\mathbf{x}+\mathbf{i}] [\mathbf{y}] + \text{POT}[\mathbf{x}] [\mathbf{y}+\mathbf{i}] + 20) - 4\tau_1} \quad (3)$$

306 In equation 3, the subtraction of τ_1 is exercised to discard values contributed by the
 307 nucleotide itself. To calculate the probabilities, a *biasing term*, in this case 10, was intro-
 308 duced to POT matrix to avoid undefined division by 0. Note that the choice of biasing
 309 term does not affect the results of our model if the parameters are scaled appropriately.
 310 The motivation behind the biasing term is to ensure non-zero probability of receiving a
 311 nucleotide to sites with zero potential. The τ and τ_1 in the calculations have to be scaled
 312 appropriately to the biasing term.

313

314 This update procedure can be related to that of a Monte Carlo(MC) simulation^{100,101}. In
 315 a Monte Carlo simulation, the particle would first chose a spot to hop to, and then make
 316 a choice whether to accept or reject the hop. Here, the decision to hop is made first and
 317 the spot to hop to is made next.

318

319 The interface is modelled at the line $y = \frac{L}{2}$, with all points above and below this is
 320 described as an ordered and disordered phase respectively of the membrane in the sim-
 321 ulation. Due to PBC, the interface was also situated at $y = 0, L$. We introduced three
 322 additional parameters I_0 for interface, I_1 , and I_2 to model the potential at the closeby
 323 interface region of the L_d and L_o phases in a phase-separated membrane respectively.
 324 The potential value was increased at and two lines above and below the interface by I_0 ,
 325 the further two lines by I_1 in the L_d phase and the subsequent two lines by I_2 in the L_o
 326 phase.

327 Parameters for the mathematical model

328 Due to lack of an exact correspondence between results in the model and MD, we qual-
 329 itatively connect the parameters. In the model, p is a measure of how “ordered” a lipid
 330 is. Higher value of p indicates a more ordered lipid and decreased diffusive ability of

the lipid. We ran two sets of simulations; the first one using parameters $p_{dis} = 0.2$ and $p_{ord} = 0.4$ and the second simulation where parameters are chosen as $p_{dis} = 0.2$ and vary p_{ord} from 0.2 to 0.9 to study the aggregation property while removing the interface potential.

For the interface parameters, we were able to estimate their order of magnitudes by considering the PMF profiles (Figure 2B and 2C). The nucleotide interaction energy with a L_o/L_d membrane, is $E \sim 7$ kJ/mol at domain interface and $E \sim 5$ kJ/mol at the disordered phase. This provides us estimates of $I_0 \sim 65$, $I_1 \sim 25$ and $I_2 \sim 5$ by assuming $I \sim e^{-E/kT}$ and scaling the values by 10 accounting for the biasing term (we have considered a small non-zero value for I_2).

For the interaction parameters, we chose two main sets in comparison to the interface parameters. One set is $(\tau, \tau_1) = (100, 90)$ which is similar in the order of magnitude to the interface potentials (I_0 value) and the second is $(\tau, \tau_1) = (700, 500)$ which is drastically larger than that.

We ran the simulations with $L = 100$ and $N = 100$ for 50,000 steps and $N = 1000$ for 200,000 steps. The result of one simulation is prone to statistical error hence each simulation was run 100 times for $N = 100$ and 20 times for $N = 1000$ respectively. The simulations we performed and their parameters are summarised in Table 3.

Results and Discussion

In this section, we highlight the important results of the study. We commence by showing the differential preference of nucleotides for membrane phases, followed by their predilection for certain lipid constituents of the membranes. Further we present the formation of nucleotide aggregates and its detailed description in the lipid environment, continued by the clustering behaviour of the nucleotide in relevance of the phases of the membrane. Finally, we discuss the assortment of clusters and their dynamics in presence of the monophasic and biphasic membranes.

Nucleotide partitioning in lipid membranes

We used the thermodynamic information embodied in PMFs to evaluate the relative preference of the mononucleotide between different surroundings: the phase-separated membrane and one-phase component domains (L_o and L_d). A prominent energy difference was observed when the UMP approaches the phosphate plane of the L_o and the L_d bilayer (Figure 2B). The UMPs have a significant energy barrier to cross the membrane phosphate plane to insert itself into the L_o membrane; the difference is of the ~ 75 kJ/mol. On the contrary, there is little energetic resistance, ~ 20 kJ/mol, for the mononucleotide when it enters into the phosphate plane (Figure 2B).

According to the PMF profile, the energy minimum is at the domain interface in L_o/L_d membrane with the next lower energy state being the L_d domain (Figure 2C). If UMP fully occupies the domain interface in the phase-separated PSM/DOPC/Chol membrane, then the excess UMP would prefer the L_d because of the lower energy. The domain partitioning of the UMP can be ascertained by (1) its interaction with lipids, which are determined by their chemical compositions and (2) the difference in lipid packing in the domains and steric effects. A possible explanation for the unsymmetric PMF profile might be an inherent polarity of the UMPs with a moiety in the nucleotide prefers the L_d domain and vice versa. Although, PMFs denote the preference of the L_d domain over L_o

in a single phase membrane system, however in a membrane with the two distinct phases it prefers the interface between the two phases followed by the L_d phase.

The trends from the free energy calculations are consistent with the energy computations (Figure 3A). According to the Eq. 1, $R=0.186$ for UMPs in the phase-separated membrane, which suggests a preference for its L_d domain. These results are entirely in accordance with the results from the PMF profiles, hence validating the AA-MD simulations for detailed analyses on the nucleotide dynamics and the membrane - nucleotide interaction. We computed the RDF to analyse the detailed interactions of the UMP atoms and the PSM/DOPC/Chol lipids (Figure 4B and 4C). The results indicate a strong interaction between the phosphate of the lipid headgroups and the ribose sugar and pyrimidine nitrogenous base of the nucleotides. The strongest interaction is between H2' followed by H3T and O2' atoms of the ribose sugar in nucleotide and the phosphate of the lipids (Figure 4C). The phosphate of the lipids also interacts with the nitrogenous base, namely the N1 atom, other weaker interactions with O4 and O2 atoms (Figure 4B); which emphasises on the importance of the OH group in the ribose sugar and its interaction with the membrane. These results establish the interaction between the UMPs and the lipid headgroup area and corroborates with an earlier study¹⁰². Furthermore, the nucleotides have lower interaction energy for the DOPC lipids, which is the predominant lipid in the L_d domain; this substantiates the preferential interactions with specific membrane lipids (Figure 3C). In the L_d and phase-separated membrane, the UMPs show strong interaction with DOPC suggesting that these interactions may enhance the L_d -lipid affinity. The partitioning effect of the UMP drives the difference in the lipid interaction of the nucleotides. Thus it substantiates lipid affinity in addition to phase affinity as shown by the nucleotides.

In addition, we explored the distribution of the UMPs from the membrane in AA-MD simulations to verify the nucleotide's predilection. The z-distance of the nucleotides are estimated by calculating the z-coordinates of the phosphate plane, and then their dif-

ference between the individual UMPs are determined; finally the ensemble average is considered (Figure 3B). The UMPs prefer to interact with the phase-separated bilayer and L_d phase and with time the UMPs move closer to the lipids. In the L_o membrane, these nucleotides drift away from the phosphate plane of the bilayer, indicating their lower preference to associate with the L_o bilayer. The UMPs in the presence of Na^+ counterions engage in interaction with them but do not penetrate the membrane headgroups but still have the competence to interact to the membrane. The bias of the nucleotides for a specific phase of the membrane bilayer is reestablished. The nucleotides could have exploited the unique ability of membranes to exist in various phases as an organising substrate for synthesizing pre-polymers and/or RNA-like polymers.

Accumulation of nucleotide and the role of membrane phases

Self-assembly of UMPs were observed in the lipid environment (Figure 5 and 6), which is supported by earlier work^{20,25,26,102}, and the DBSCAN clustering algorithm reported that the occurrence of 8 clusters in L_o -UMP, 12 clusters in L_d -UMP and 14 clusters in L_o/L_d -UMP membrane systems. The most scattered clusters emerges in the L_o/L_d -UMP system. In addition to that, the change of the solvent-accessible surface of UMPs with time (Figure S2) and their respective solvation energy confirms their clustering behaviour. The SASA decreases which is driven by hydrophobic effects and hence burial of the non-polar moiety of the nucleotides. This accounts for the corresponding reduction in the solvation energy of the UMPs with time. It establishes that the clusters were formed in the 200ns of simulation time.

The flexibility of the dihedrals in the nucleotides explains the UMPs' ability to accommodate Na^+ counterions and align with the other mononucleotide residue (Figure S3). The χ dihedral, which connects the ribose sugar with the nitrogenous base, is constant for most UMPs but the remaining dihedrals show considerable deviation from their mean, indicating their adjustability in the molecule. The dihedral, β , between the ribose sugar and phosphate group indicates the higher movement of the phosphate group, perhaps

to minimise interaction with the negatively charged lipid headgroups. However, the dihedral, α , in the phosphate group is rigid, which demonstrates the mobility within the group is limited. The dihedral angles connecting the ribose sugar with the nitrogenous base or phosphate group show considerable flexibility, indicating intra-molecule adjustability. Despite different lipid environment around the UMPs there is negligible difference in their dihedrals, implicating similar intrinsic movements with itself.

The SAXS profile provides further details about the local arrangement of the UMPs in the lipid environment; it indicates that their clusters are anisotropic with a higher polydispersity (Figure S1 and Table S1). The radius of gyration (R_g) of the biggest clusters formed were in the range of ~ 4 -5 nm. The clusters were, perhaps, stabilised by hydrogen bonding or π - π interaction between the constituent mononucleotide. It can be further stabilised by cation- π interaction in the proximity of the Na^+ counterions. The lipid environment not only has reduced the diffusivity of the nucleotides but also of the Na^+ counterions. The membranes have aided the easy access to finding other UMPs in the otherwise aqueous solution by reducing the search space. Due to AA-MD limitations, we further studied the clustering of nucleotides using bigger patches of membrane and more nucleotide monomers employing the mathematical model.

To verify the importance of the phase separation in the membranes in nucleotide clustering, we used the same values of p for ordered and disordered phases of the membrane (Case 1A, 1B, 2A and 2B), while the parameters determining interfacing effect (I_0 , I_1 , I_2) were not considered (Figure S4). The results indicate that the cluster aggregates formed are considerably lower than the Case 3A, 3B, 4A and 4B where the the different phases had distinct p values; which demonstrates the importance of the two unique phases in clustering of nucleotides (Figure 6D, 6E, 7B and 7C). We also ran simulations within the L_o/L_d membrane where the interface effects were turned off, but the p_{ord} is varied from 0.2 to 0.9 while p_{dis} is kept fixed. There was a increasing trend for occupation of large cluster sizes as we increase the difference between the p_{ord} and p_{dis} (Figure

6D). It suggests that an interface potential is necessary when the difference in diffusive abilities to the two phases is not too extreme. With 100 nucleotides in simulations, the results suggests that without the interface, the large clusters reach a size of utmost 25 nucleotide. However with a interface, we saw improvements in the cluster distributions with aggregates reaching sizes of more than 30 (Figure S5 and S7). Similar pattern was observed in membrane simulated with 1000 nucleotides, they formed and preferred to be in bigger clusters with 200 particles or more. These results suggests that both the effects of interface and different phases play a key role in influencing cluster aggregation.

Distribution and dynamics of nucleotide clusters on membranes

The UMPs organised into various assortment of clusters (Figure 6C) ranging from dimers to clusters with more than 20 monomeric constituents. Smaller clusters consisting of 2-5 UMPs are the most abundant in lipid environments. Clusters of more substantial size were detected; however, not too many. At least one cluster across the various size ranges has been identified. The L_o lipid environment restricted the formation of a cluster comprising of 5-10 monomers whereas in the other systems at least 2 clusters of similar sizes were formed. We conducted further simulations with the mathematical model and found that 20 clusters of size smaller than 5 were formed while the MD simulation results show that they can form as many as 8. Unlike the AA-MD simulations, there is formation of dimer stragglers which reduces its mobility. This can be improved by incorporating continuous diffusion model.

The Figure 7 illustrates the dynamics and the stability of the cluster formation in the distinct lipid systems. The constant association and dissociation of nucleotides in the clusters are evident from their temporal fluctuations. The monophasic L_o and L_d domain have exhausted most of their monomers in the system and appear to have reached equilibrium with the clusters of various sizes. However, in the phase-separated system (Figure 7A) shows monomers remaining in the solutions; implying the UMPs and the clusters did not reach equilibrium, and the clusters can grow further. Due

493 to such shortcomings, we apply our mathematical model to the study the steady state
 494 of the nucleotides in the system. To verify the steady state of the nucleotides in the
 495 simulation we modelled the system as a Markov chain. We assigned each nucleotide
 496 to a state labelled by the size of the cluster it belonged. For example, all the nu-
 497 cleotides belonging to a cluster of size 4 would be assigned to the state 4 and then
 498 and enumerated the transition and stationary matrix of the system. The clusters were
 499 categorised into groups containing 1, 2, 3, 4, 5 – 14, 15 – 24, ≥ 25 mononucleotides and
 500 1 – 10, 11 – 20, 21 – 30, 31 – 40, 41 – 50, 51 – 100, 101 – 150, 151 – 200, ≥ 201 mononu-
 501 cleotides for N=100 and N=1000 nucleotides respectively (Figure 7B and 7C). The tem-
 502 poral data of 100 steps were binned and averaged, the corresponding transition matrix
 503 was calculated and subsequently a stationary matrix.

504

505 The most important aspect of this model is that this provides us an insight for the
 506 physical process that results in the prominent role of an interface which is supported by
 507 mathematical reasoning. The process being describes the trajectory of the nucleotides
 508 in the lipid environment. We appeal to a classical theorem by Polya¹⁰³ which states
 509 that a 2D random walk always returns to the origin. Due to the higher potential of the
 510 interface, once a nucleotide reaches the spot it effectively is trapped at the interface (on
 511 the lines denoted by $y = 0$, $y = \frac{L}{2}$ and $y = L$). In case the nucleotide leaves the interface,
 512 according to the theorem the nucleotide almost always returns back to the interface at a
 513 nearby position which explains why most of the nucleotides are conclusively trapped at
 514 the interface (Figure 5B and 5C). Moreover, the nucleotide prefers the disordered side of
 515 the interface, and hence it would first leave the interface through the disordered side with
 516 a higher probability and then it would return to the interface faster due to the lower p
 517 value of the disordered side.

Conclusion and outlook

Previous studies have investigated membranous compartments as substrates for nucleotides' (UMP and/or AMP) non-enzymatic polymerisation^{20,25,26} and estimated it to be satisfactory. With this work, we explored further into the role of membranes as an organising matrix. Membranes have a unique property which allows it exist it multiple phases in a short range of temperature and hydration; these phases can co-exist at certain settings, environment or compositional^{59,60,70-72}. We explored the significance of such phases and their coexistence in membranes and their contribution to the non-enzymatic polymerisation of the nucleotides. Our results indicate that the domain interface followed by the disordered phase of a L_o/L_d membrane is the prime location for nucleotide cluster aggregation. We also report heterogeneous preference for the lipids in the membrane which promote fluid-like behaviour in the membrane bilayers and the interaction at an atomic level between the ribose sugar and nitrogenous groups of the UMPs with the lipids.

Our work show the formation of nucleotide clusters in multilamellar membranes which is in accordance^{25,26,102} with prior research in OoL. The results reconfirms the understanding that RNA-like molecules could have formed by the self-assembly of the mononucleotides in lipid environment. We provide detailed molecular insights for the rearrangement of nucleotides into clusters. Further, we developed a mathematical which mimics the nucleotide interactions with lipids and nucleotides and from it we infer that the nucleotide and its clusters finally attain a steady state in the lipid environment. The mathematical model corroborates our AA-MD simulation results and provides additional theoretical insight to the problem. The model can be improved into a continuous diffusion model which would be computationally intensive to implement or by including a *bond* between nucleotides.

Thus our work conclusively shows that the amphiphilic phospholipid acts as an ordering agent, and the membrane phases enhances the guided non-enzymatic polymerisation reaction. Therefore, it is important to emphasize on the phase of the membrane when used

547 as an organising matrix as it could act as a driving factor in the pre-polymer or RNA-like
548 molecule formation which is indispensable to the molecular OoL.

549 Acknowledgments

550 The authors are grateful to the SciNet HPC Consortium, ComputeCanada^{104,105} and
551 Supercomputer Education and Research Center, IISc for the computational resources.

Figure 1: **Schematic representation of the spatio-temporal model developed in this work:** The phase separated bilayer matrix has been represented using blue beads for L_o and red beads for L_d respectively. The nucleotide has been shown in green shade. The parameters used in the model have been explained; p is the probability of the nucleotide to move in the lipid matrix which differs in the L_o and L_d phase. τ and τ_1 is the potential at the position of and around the nucleotide - replicating a small attractive forcefield about it. The domain interface potential is defined by I_0 , and the corresponding area in the L_d and L_o is defined by I_1 and I_2 respectively.

Table 1: Detailed description of PSM/DOPC/Chol membranes in various phases and their composition

Phase	Mixture	Composition	Temp	Total Lipid	Area per Lipid
L_o	PSM/DOPC/Chol	0.64/0.03/0.33	295K	560	40.4Å
L_d	PSM/DOPC/Chol	0.15/0.82/0.03	295K	574	62.7Å
L_o/L_d	PSM/DOPC/Chol	0.43/0.38/0.19	295K	452	(N/A)

Table 2: Details of the AA-MD simulations of the UMPs in the presence of PSM/DOPC/-Chol bilayers.

System	Phase	Temp	Molecule or ion	No of molecules	Simulation time(ns)
L_o -UMP	L_o	295K	PSM	362	200
			DOPC	14	
			Chol	184	
			UMP	60	
			Na^+	120	
L_d -UMP	L_d	295K	PSM	84	200
			DOPC	460	
			Chol	30	
			UMP	64	
			Na^+	128	
L_o/L_d -UMP	L_o/L_d	295K	PSM	194	200
			DOPC	170	
			Chol	88	
			UMP	72	
			Na^+	144	

Table 3: Details of the modelling simulations with nucleotides on a membrane with various parameters determining the diverse interactions.

		P_{ord}	P_{dis}	τ	τ_1	I_0	I_1	I_2
Case	1A	0.2	0.2	100	90	-	-	-
	1B	0.2	0.2	700	500	-	-	-
Case	2A	0.4	0.4	100	90	-	-	-
	2B	0.4	0.4	700	500	-	-	-
Case	3A	0.4	0.2	100	90	65	25	5
	3B	0.25	0.2	100	90	65	25	5
Case	4A	0.4	0.2	700	500	65	25	5
	4B	0.25	0.2	700	500	65	25	5
Case	5A	0.2	0.2	100	90	0	0	0
	5B	0.3	0.2	100	90	0	0	0
	5C	0.4	0.2	100	90	0	0	0
	5D	0.5	0.2	100	90	0	0	0
	5E	0.6	0.2	100	90	0	0	0
	5F	0.7	0.2	100	90	0	0	0
	5G	0.8	0.2	100	90	0	0	0
	5H	0.9	0.2	100	90	0	0	0

Figure 2: (A) **Schematics representing the umbrella sampling set-up comprising of UMP and PSM/DOPC/Chol bilayer:** L_o & L_d system - The UMP was pulled perpendicular to the plane of the bilayer, from the solution to the inside of phosphate plane. L_o/L_d system - In the phase-separated system, the UMP was pulled across the phosphate plane of the membrane from the L_d region to the L_o domain. (B & C) **PMF of the nucleotides in the different phases of PSM/DOPC/Chol bilayers at 295K:** (B) PMF of the L_o and L_d domain of PSM/DOPC/Chol. (C) PMF of the phase-separated PSM/DOPC/Chol. The UMPs have a distinct preference for the L_d phase over the L_o in the monophonic membranes. However, in the L_o/L_d the nucleotide prefers the interface between the L_o and L_d .

Figure 3: (A) Total interaction energy, Coulombic and van der Waals interactions, between the nucleotide and the L_o and L_d domains of the phase separated membrane system (B) **The z-distance calculation between the UMPs and phosphate plane of the membrane:** The UMPs move closes to the phase-separated bilayer which indicates its preference for L_o/L_d whereas in the L_d membrane, these nucleotides drift away from the phosphate plane of the bilayer. This reinstates the bias of the nucleotides. (C) **Interaction energy of UMPs and the various lipids of the PSM/DOPC/Chol membrane:** The nucleotides have the lowest interaction energy with the DOPC, which is the predominant lipid in the L_d phase. This reinstates the lipid bias of the nucleotides; the differential interaction between UMP and the lipids drive the partitioning effect.

Figure 4: (A) Chemical structure of Uracil Monophosphate and the nomenclature of their atoms according to the .pdb file (B) **RDF of nitrogenous base atoms of the UMPs around the phosphorous atom of the lipids in membrane:** The N1 of the base shows strongest interaction with the phosphate of the lipids, followed by O4 and O2 atoms. (C) **RDF of ribose sugar atoms of UMP around the phosphate plane of the membrane:** The strongest interaction is between H2' atom of ribose and the phosphate of lipids

Figure 5: (A) **Simulation of nucleotides in a PSM/DOPC/Chol lipid environment using AA-MD:** The $t=0\text{ns}$ shows the nucleotides arranged in arrays on the membranes, and at $t=500\text{ns}$ presents the final snapshot of the system. PSM and DOPC is shown in green and magenta respectively, Chol is represented by yellow, UMP is displayed in blue. (B & C) **Simulation of nucleotides using mathematical model:** In the phase-separated systems, the L_o phase is represented in black where as the L_d phase is represented in white and UMP is shown in blue. (B) 1000 nucleotides on 100×100 matrix: L_d with $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, L_o with $p_{ord}=0.2$, $\tau=100$ and $\tau_1=90$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=65$, $I_1=25$, and $I_2=5$. (C) 100 nucleotides on 100×100 matrix: L_d with $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, L_o with $p_{ord}=0.2$, $\tau=700$ and $\tau_1=500$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, $I_0=65$, $I_1=25$, and $I_2=5$. The initial and the final arrangement of nucleotides and their clusters which reach their reach steady state in the corresponding membranes are shown. We observed that the clusters of nucleotides aggregated at the interface and later migrated to either of the phases in the membrane, especially the L_o phase.

Figure 6: (A & B) **Cluster of UMPs in lipid environment:** UMPs organised as clusters (A) as seen from XY-plane and (B) 3D-perspective. UMPs were observed to self assemble in the lipid environments. (C) **Distribution of various UMP clusters in lipid environments:** Clusters with 2-5 mononucleotides, followed by 5-10, were the predominant in the system. However, clusters of substantial sizes were detected but not too many. (D) **Importance of different phases in cluster formation:** It highlights the nucleotide cluster distribution as a function of p_{ord} on phase separated membranes. There was a distinct increase for occupation of large cluster sizes as we increase the difference between the p_{ord} and p_{dis} . (E & F) **Temporal cluster distribution in different phases of membrane using computer modelling in 100×100 matrix with 1000 nucleotides:** (E) L_o with $p_{ord}=0.2$, $\tau=100$ and $\tau_1=90$, L_d with $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=65$, $I_1=25$, and $I_2=5$. (F) L_o with $p_{ord}=0.2$, $\tau=700$ and $\tau_1=500$, L_d with $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, $I_0=65$, $I_1=25$, and $I_2=5$.

Figure 7: (A) **Trajectory of the UMP clusters in distinct membrane bilayers:** The mononucleotides are exhausted and have attained an equilibrium with the clusters in the L_o and L_d membranes; where as the monomers are not depleted and the clusters are still capable of further growth in the phase separated bilayer. It also establishes the association and dissociation of UMP from the clusters. (B & C) **Cluster evolution and steady state distribution in 100×100 matrix with 1000 nucleotides:** (B) L_o with $p_{ord}=0.2$, $\tau=100$ and $\tau_1=90$, L_d with $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=65$, $I_1=35$, and $I_2=5$. (C) L_o with $p_{ord}=0.2$, $\tau=700$ and $\tau_1=500$, L_d with $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, L_o/L_d with $p_{ord}=0.2$ and $p_{dis}=0.4$, $\tau=700$ and $\tau_1=500$, $I_0=65$, $I_1=35$, and $I_2=5$.

Figure S1: **SAXS profile of UMPs in the lipid environment:** This reestablishes that the clusters formed are anisotropic. Despite the similar profiles the size of clusters are varied across the systems.

Table S1: **Details of the radius of gyration of the most significant clusters in different lipid and aqueous environment:** The biggest clusters formed in former is ~ 4 -5 nm.

Systems	Radius of Gyration ($\text{\AA}$)
L_o -UMP	53.03
L_d -UMP	40.21
L_o/L_d -UMP	46.10

Figure S2: **SASA and the solvation energy of UMPs in the three distinct model membrane systems:** (A) The solvent-accessible surface area of UMPs and its change over time in a lipid environment (B) The solvation energy of UMPs

Figure S3: **Dihedrals of UMPs simulated in the lipid environment:** The flexibility χ and β dihedral of the UMP ensure intrinsic movement of the nitrogenous base and phosphate group within the mononucleotides.

Figure S4: **Cluster evolution and steady state distribution in 100×100 matrix with 1000 nucleotides:** (A,C) L_d with $p=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=0$, $I_1=0$, and $I_2=0$. (C,D) L_o with $p=0.2$, $\tau=700$ and $\tau_1=500$, $I_0=0$, $I_1=0$, and $I_2=0$. (A,B) Cluster Evolution (C,D) Steady State Occupation.

Figure S5: **Temporal cluster distribution in different phases of membrane using computer modelling in 100×100 matrix with 100 nucleotides:** (A) L_d with $p=0.4$, $\tau=100$ and $\tau_1=90$ (B) L_o with $p=0.2$, $\tau=100$ and $\tau_1=90$ (C) L_o/L_d with $p(L_o)=0.2$ and $p(L_o)=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=65$, $I_1=25$, and $I_2=5$.

Figure S6: **Cluster evolution and steady state distribution in 100×100 matrix with 100 nucleotides:** (A,D) L_d with $p=0.4$, $\tau=100$ and $\tau_1=90$ (B,E) L_o with $p=0.2$, $\tau=100$ and $\tau_1=90$ (C,F) L_o/L_d with $p(L_o)=0.2$ and $p(L_o)=0.4$, $\tau=100$ and $\tau_1=90$, $I_0=65$, $I_1=25$, and $I_2=5$. (A,B,C) Cluster Evolution (D,E,F) Steady State Occupation.

Figure S7: **Temporal cluster distribution in different phases of membrane using computer modelling in 100×100 matrix with 100 nucleotides:** (A) L_d with $p=0.4$, $\tau=700$ and $\tau_1=500$ (B) L_o with $p=0.2$, $\tau=700$ and $\tau_1=500$ (C) L_o/L_d with $p(L_o)=0.2$ and $p(L_o)=0.4$, $\tau=700$ and $\tau_1=500$, $I_0=65$, $I_1=25$, and $I_2=5$.

Figure S8: **Cluster evolution and steady state distribution in 100×100 matrix with 100 nucleotides:** (A,D) L_d with $p=0.4$, $\tau=700$ and $\tau_1=500$ (B,E) L_o with $p=0.2$, $\tau=700$ and $\tau_1=500$ (C,F) L_o/L_d with $p(L_o)=0.2$ and $p(L_o)=0.4$, $\tau=700$ and $\tau_1=500$, $I_0=65$, $I_1=25$, and $I_2=5$. (A,B,C) Cluster Evolution (D,E,F) Steady State Occupation.

Umbrella Sampling

An enhanced sampling method that is referred as Umbrella Sampling calculates the free energy and its profile along a reaction coordinate. To obtain a free energy profile in 1-D:

1. Choose a reaction coordinate ξ .

2. **Biased molecular dynamics simulations:** Simulations are performed at different ξ , where the ξ is constrained with harmonic potential.

3. **Calculation of the Potential of Mean Force from Umbrella Sampling**

Data: At each ξ the potential mean force, F_ξ is calculated and it is integrated over ξ to obtain ΔG .

$$\frac{\delta G}{\delta \xi} = -F_{\xi\xi}$$

Transition and Stationary matrix

A Markov model¹⁰⁶ consists of a system which can belong to a set of D states labelled by i . At each time instant, the system changes from state i to state j with probability $\pi[j][i]$ where π is called the transition matrix. At each time instant t , the state can be in the state i with probability $v_t[i]$. v_t is a vector that contains the probability distribution of the state at time t . From v_t , we can find v_{t+1} using $v_{t+1} = \pi v_t$. A state v such that $v = \pi v$ is called as the stationary state and the system is thus in equilibrium.

Bibliography

- (1) Gayon, J.; Malaterre, C.; Morange, M.; Raulin-Cerceau, F.; Tirard, S. Defining life: conference proceedings. *Origins of Life and Evolution of the Biosphere* **2010**, *40*, 119.
- (2) Michaelian, K. Microscopic dissipative structuring and proliferation at the origin of life. *Heliyon* **2017**, *3*, e00424.
- (3) Michaelian, K. Thermodynamic dissipation theory for the origin of life. *Earth System Dynamics* **2011**, *2*, 37–51.
- (4) Michaelian, K. Thermodynamic origin of life. *arXiv preprint arXiv:0907.0042* **2009**,
- (5) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature chemistry* **2015**, *7*, 301–307.
- (6) Rich, A. On the problems of evolution and biochemical information transfer. *Horizons in biochemistry* **1962**, 103–126.
- (7) Crick, F. H. The origin of the genetic code. *Journal of molecular biology* **1968**, *38*, 367–379.
- (8) Woese, C. R. Evolution of the genetic code. *The Science of Nature* **1973**, *60*, 447–459.
- (9) Copley, S. D.; Smith, E.; Morowitz, H. J. The origin of the RNA world: co-evolution of genes and metabolism. *Bioorganic chemistry* **2007**, *35*, 430–443.
- (10) Neveu, M.; Kim, H.-J.; Benner, S. A. The “strong” RNA world hypothesis: Fifty years old. *Astrobiology* **2013**, *13*, 391–403.
- (11) Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, *418*, 214–221.

- (12) Robertson, M. P.; Joyce, G. F. The origins of the RNA world. *Cold Spring Harbor perspectives in biology* **2012**, *4*, a003608.
- (13) Cech, T. R. The RNA worlds in context. *Cold Spring Harbor perspectives in biology* **2012**, *4*, a006742.
- (14) Ward, P.; Kirschvink, J. *A new history of life: the radical new discoveries about the origins and evolution of life on earth*; Bloomsbury Publishing USA, 2015.
- (15) Oparin, A. I. The origin of life. **1953**,
- (16) Oparin, A. I., et al. The origin of life on the earth. *The origin of life on the earth*. **1957**,
- (17) Scharf, C.; Virgo, N.; Cleaves, H. J.; Aono, M.; Aubert-Kato, N.; Aydinoglu, A.; Barahona, A.; Barge, L. M.; Benner, S. A.; Biehl, M., et al. A strategy for origins of life research. 2015.
- (18) Costanzo, G.; Saladino, R.; Botta, G.; Giorgi, A.; Scipioni, A.; Pino, S.; Di Mauro, E. Generation of RNA Molecules by a Base-Catalysed Click-Like Reaction. *ChemBioChem* **2012**, *13*, 999–1008.
- (19) Morasch, M.; Mast, C. B.; Langer, J. K.; Schilcher, P.; Braun, D. Dry Polymerization of 3, 5-Cyclic GMP to Long Strands of RNA. *ChemBioChem* **2014**, *15*, 879–883.
- (20) Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres* **2008**, *38*, 57–74.
- (21) Ferris, J. P.; Hill, A. R.; Liu, R.; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59–61.
- (22) Ferris, J. P. 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* **2006**, *361*, 1777–1786.

- (23) Olasagasti Arsuaga, F.; Rajamani, S. Lipid-Assisted Polymerization of Nucleotides. **2019**,
- (24) Da Silva, L.; Maurel, M.-C.; Deamer, D. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *Journal of molecular evolution* **2015**, *80*, 86–97.
- (25) Himbert, S.; Chapman, M.; Deamer, D. W.; Rheinstädter, M. C. Organization of nucleotides in different environments and the formation of pre-polymers. *Scientific reports* **2016**, *6*, 31285.
- (26) Toppozini, L.; Dies, H.; Deamer, D. W.; Rheinstädter, M. C. Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: Insights into assembly of nucleic acid for primitive life. *PloS one* **2013**, *8*, e62810.
- (27) Fraccia, T. P.; Zanchetta, G.; Rimoldi, V.; Clark, N. A.; Bellini, T. Evidence of liquid crystal-assisted abiotic ligation of nucleic acids. *Origins of Life and Evolution of Biospheres* **2015**, *45*, 51–68.
- (28) Deamer, D. W.; Georgiou, C. D. Hydrothermal conditions and the origin of cellular life. *Astrobiology* **2015**, *15*, 1091–1095.
- (29) Baaske, P.; Weinert, F. M.; Duhr, S.; Lemke, K. H.; Russell, M. J.; Braun, D. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proceedings of the National Academy of Sciences* **2007**, *104*, 9346–9351.
- (30) Langridge, R.; Gomatos, P. J. The structure of RNA. *Science* **1963**, *141*, 1024–1024.
- (31) Watson, J. D.; Crick, F. H. The structure of DNA. Cold Spring Harbor symposia on quantitative biology. 1953; pp 123–131.
- (32) Sheng, J.; Li, L.; Engelhart, A. E.; Gan, J.; Wang, J.; Szostak, J. W. Structural insights into the effects of 2-5 linkages on the RNA duplex. *Proceedings of the National Academy of Sciences* **2014**, *111*, 3050–3055.

- (33) Deamer, D. Liquid crystalline nanostructures: organizing matrices for non-enzymatic nucleic acid polymerization. *Chemical Society Reviews* **2012**, *41*, 5375–5379.
- (34) Blumstein, A.; Kitagawa, N.; Blumstein, R. Polymerization of p-methacryloyloxy benzoic acid within liquid crystalline media. *Molecular Crystals and Liquid Crystals* **1971**, *12*, 215–227.
- (35) Barrall, E. M.; Johnson, J. F. A review of the status of polymerization in thermotropic liquid crystal media and liquid crystalline monomers. *Journal of Macromolecular Science—Reviews in Macromolecular Chemistry* **1979**, *17*, 137–170.
- (36) Lohrmann, R.; Bridson, P.; Orgel, L. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **1980**, *208*, 1464–1465.
- (37) DeGuzman, V.; Vercoutere, W.; Shenasa, H.; Deamer, D. Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *Journal of molecular evolution* **2014**, *78*, 251–262.
- (38) Hargrave, M.; Thompson, S. K.; Deamer, D. Computational models of polymer synthesis driven by dehydration/rehydration cycles: repurination in simulated hydrothermal fields. *Journal of molecular evolution* **2018**, *86*, 501–510.
- (39) Higgs, P. The effect of limited diffusion and wet–dry cycling on reversible polymerization reactions: implications for prebiotic synthesis of nucleic acids. *Life* **2016**, *6*, 24.
- (40) Damer, B.; Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life. *Life* **2015**, *5*, 872–887.
- (41) Deamer, D.; Singaram, S.; Rajamani, S.; Kompanichenko, V.; Guggenheim, S. Self-assembly processes in the prebiotic environment. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2006**, *361*, 1809–1818.

- (42) Kawamura, K.; Ferris, J. P. Kinetic and mechanistic analysis of dinucleotide and oligonucleotide formation from the 5'-phosphorimidazolidine of adenosine on Na⁺-montmorillonite. *Journal of the American Chemical Society* **1994**, *116*, 7564–7572.
- (43) Kawamura, K.; Ferris, J. P. Clay catalysis of oligonucleotide formation: kinetics of the reaction of the 5'-phosphorimidazolidines of nucleotides with the non-basic heterocycles uracil and hypoxanthine. *Origins of Life and Evolution of the Biosphere* **1999**, *29*, 563–591.
- (44) Ertem, G.; Ferris, J. P. Sequence- and regio-selectivity in the montmorillonite-catalyzed synthesis of RNA. *Origins of Life and Evolution of the Biosphere* **2000**, *30*, 411–422.
- (45) Deamer, D.; Lancet, D.; Segre, D.; Ban-Eli, D. The lipid world. *Orig Life Evol Biosph* **2001**, *31*, 119–145.
- (46) Hargreaves, W.; Mulvihill, S.; Deamer, D. Synthesis of phospholipids and membranes in prebiotic conditions. *Nature* **1977**, *266*, 78–80.
- (47) Tawfik, D. S.; Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nature biotechnology* **1998**, *16*, 652–656.
- (48) Luisi, P. L.; Walde, P.; Oberholzer, T. Lipid vesicles as possible intermediates in the origin of life. *Current opinion in colloid & interface science* **1999**, *4*, 33–39.
- (49) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. Synthesizing life. *Nature* **2001**, *409*, 387–390.
- (50) Lazcano, A. The RNA world, its predecessors, and its descendants. *Early life on Earth* (ed. S. Bengtson) **1994**, 70–80.
- (51) Lazcano, A. The transition from nonliving to living. Nobel Symp. 1994; pp 60–69.
- (52) Deamer, D. W. Self-assembly and function of primitive membrane structures. *Early life on Earth* **1994**, 107–123.

- (53) Deamer, D. W. The first living systems: a bioenergetic perspective. *Microbiology and Molecular Biology Reviews* **1997**, *61*, 239–261.
- (54) Fleischaker, G. R. Origins of life: an operational definition. *Origins of Life and Evolution of the Biosphere* **1990**, *20*, 127–137.
- (55) Deamer, D. Membranes and the origin of life: a century of conjecture. *Journal of molecular evolution* **2016**, *83*, 159–168.
- (56) Singer, S. J.; Nicolson, G. L. The fluid mosaic model of the structure of cell membranes. *Science* **1972**, *175*, 720–731.
- (57) Giocondi, M.-C.; Yamamoto, D.; Lesniewska, E.; Milhiet, P.-E.; Ando, T.; Le Grimellec, C. Surface topography of membrane domains. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2010**, *1798*, 703–718.
- (58) Wu, H.-M.; Lin, Y.-H.; Yen, T.-C.; Hsieh, C.-L. Nanoscopic substructures of raft-mimetic liquid-ordered membrane domains revealed by high-speed single-particle tracking. *Scientific reports* **2016**, *6*, 20542.
- (59) Simons, K.; Ikonen, E. Functional rafts in cell membranes. *nature* **1997**, *387*, 569–572.
- (60) Brown, D.; London, E. Functions of lipid rafts in biological membranes. *Annual review of cell and developmental biology* **1998**, *14*, 111–136.
- (61) Long, S. B.; Tao, X.; Campbell, E. B.; MacKinnon, R. Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* **2007**, *450*, 376–382.
- (62) Hansen, S. B.; Tao, X.; MacKinnon, R. Structural basis of PIP 2 activation of the classical inward rectifier K⁺ channel Kir2. 2. *Nature* **2011**, *477*, 495–498.
- (63) McLaughlin, S.; Wang, J.; Gambhir, A.; Murray, D. PIP2 and proteins: interactions, organization, and information flow. *Annual review of biophysics and biomolecular structure* **2002**, *31*, 151–175.

- (64) Hammond, G. R.; Hong, Y. Phosphoinositides and membrane targeting in cell polarity. *Cold Spring Harbor perspectives in biology* **2018**, *10*, a027938.
- (65) Stone, M. B.; Shelby, S. A.; Núñez, M. F.; Wisser, K.; Veatch, S. L. Protein sorting by lipid phase-like domains supports emergent signaling function in B lymphocyte plasma membranes. *Elife* **2017**, *6*, e19891.
- (66) Gaus, K.; Chklovskaya, E.; Fazekas de St. Groth, B.; Jessup, W.; Harder, T. Condensation of the plasma membrane at the site of T lymphocyte activation. *The Journal of cell biology* **2005**, *171*, 121–131.
- (67) van Meer, G.; Stelzer, E.; Wijnaendts-van Resandt, R. W.; Simons, K. Sorting of sphingolipids in epithelial (Madin-Darby canine kidney) cells. *The Journal of cell biology* **1987**, *105*, 1623–1635.
- (68) Salaün, C.; James, D. J.; Chamberlain, L. H. Lipid rafts and the regulation of exocytosis. *Traffic* **2004**, *5*, 255–264.
- (69) Gaus, K.; Le Lay, S.; Balasubramanian, N.; Schwartz, M. A. Integrin-mediated adhesion regulates membrane order. *The Journal of cell biology* **2006**, *174*, 725–734.
- (70) Harayama, T.; Riezman, H. Understanding the diversity of membrane lipid composition. *Nature Ecology & Evolution* **2018**, 1–18.
- (71) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature reviews Molecular cell biology* **2008**, *9*, 112–124.
- (72) Jacobson, K.; Mouritsen, O. G.; Anderson, R. G. Lipid rafts: at a crossroad between cell biology and physics. *Nature cell biology* **2007**, *9*, 7–14.
- (73) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

- (74) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr, A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B* **2010**, *114*, 7830–7843.
- (75) Van Gunsteren, W. F.; Berendsen, H. J. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation* **1988**, *1*, 173–185.
- (76) Nosé, S.; Klein, M. L. A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. *The Journal of Chemical Physics* **1983**, *78*, 6928–6939.
- (77) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **1980**, *72*, 2384–2393.
- (78) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.
- (79) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.
- (80) Allen, M. Tildesley. *DJ Computer simulation of liquids*. 1987.
- (81) Lagüe, P.; Pastor, R. W.; Brooks, B. R. Pressure-based long-range correction for Lennard-Jones interactions in molecular dynamics simulations: application to alkanes and interfaces. *The Journal of Physical Chemistry B* **2004**, *108*, 363–368.
- (82) Lin, X.; Gorfe, A. A.; Levental, I. Protein partitioning into ordered membrane domains: insights from simulations. *Biophysical journal* **2018**, *114*, 1936–1944.
- (83) Centi, A.; Dutta, A.; Parekh, S. H.; Bereau, T. Inserting small molecules across membrane mixtures: Insight from the potential of mean force. *Biophysical Journal* **2020**,

- (84) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd.* 1996; pp 226–231.
- (85) Van Rossum, G., et al. Python Programming Language. USENIX annual technical conference. 2007; p 36.
- (86) Debye, P. Zerstreuung von röntgenstrahlen. *Annalen der Physik* **1915**, *351*, 809–823.
- (87) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic acids research* **2010**, *38*, W540–W544.
- (88) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **1992**, *13*, 1011–1021.
- (89) Efron, B. *Breakthroughs in statistics*; Springer, 1992; pp 569–593.
- (90) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. g-wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of chemical theory and computation* **2010**, *6*, 3713–3720.
- (91) Ising, E. Beitrag zur theorie des ferro-und paramagnetismus. Ph.D. thesis, Hamburg, 1924.
- (92) Lattice models of polymer fluids: Monomers occupying several lattice sites. II. Interaction energies. *The Journal of Chemical Physics*
- (93) Flory, P. J. Principles of Polymer Chemistry.
- (94) Baxter, R. J. Variational approximations for square lattice models in statistical mechanics. *Journal of Statistical Physics* **1978**, *19*, 461–478.
- (95) Ghosal, A.; Randeria, M.; Trivedi, N. Inhomogeneous pairing in highly disordered s-wave superconductors. *Physical Review B* **2001**, *65*, 014501.

- 795 (96) Karlin, S.; McGregor, J., et al. Random walks. *Illinois Journal of Mathematics*
796 **1959**, *3*, 66–81.
- 797 (97) Montroll, E. W.; Weiss, G. H. Random walks on lattices. II. *Journal of Mathemat-*
798 *ical Physics* **1965**, *6*, 167–181.
- 799 (98) Weiss, G. H.; Rubin, R. J. Random walks: theory and selected applications. *Adv.*
800 *Chem. Phys* **1983**, *52*, 363–505.
- 801 (99) Makov, G.; Payne, M. Periodic boundary conditions in ab initio calculations. *Phys-*
802 *ical Review B* **1995**, *51*, 4014.
- 803 (100) Griesbauer, J.; Wixforth, A.; Seeger, H. M.; Schneider, M. F. Method for the Monte
804 Carlo based Simulation of Lipid-Monolayers including Lipid Movement. 2010.
- 805 (101) Kotze, J. Introduction to Monte Carlo methods for an Ising Model of a Ferromag-
806 net. 2008.
- 807 (102) Sasidharan, S.; Pochinda, S.; Elgaard-Jørgensen, P. N.; Rajamani, S.; Khan-
808 delia, H.; Raghunathan, V. Interaction of the mononucleotide UMP with a fluid
809 phospholipid bilayer. *Soft matter* **2019**, *15*, 8129–8136.
- 810 (103) Pólya, G. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Ir-
811 rfahrt im Straßennetz. *Mathematische Annalen* **1921**, *84*, 149–160.
- 812 (104) Ponce, M.; van Zon, R.; Northrup, S.; Gruner, D.; Chen, J.; Ertinaz, F.; Fe-
813 doseev, A.; Groer, L.; Mao, F.; Mundim, B. C., et al. *Proceedings of the Practice*
814 *and Experience in Advanced Research Computing on Rise of the Machines (learn-*
815 *ing)*; 2019; pp 1–8.
- 816 (105) Candoré, J. C.; Bodnar, J.; Detalle, V.; Grossel, P. Non destructive testing in
817 situ, of works of art by stimulated infra-red thermography. *Journal of Physics:*
818 *Conference Series*. 2010; p 012068.
- 819 (106) Gagniuc, P. A. *Markov chains: from theory to implementation and experimenta-*
820 *tion*; John Wiley & Sons, 2017.