# Membrane catalysed formation of nucleotide clusters and its role in the origins of life[†]

Rajlaxmi Saha,[a] Prathyush Poduval,[b,c] Krishnakanth Baratam[b], Jayashree Nagesh,[d*] and Anand Srivastava[b*]

One of the mysteries in studying the molecular "Origin of Life" is the emergence of RNA and RNA-based life forms, where non-enzymatic polymerization of nucleotides is a crucial hypothesis in formation of large RNA chains. The non-enzymatic polymerization can be mediated by various environmental settings such as cycles of hydration and dehydration, temperature variations and proximity to a variety of organizing matrices such as clay, salt, fatty acids, lipid membrane and mineral surface. In this work, we explore the influence of different phases of the lipid membrane towards nucleotide organization and polymerization in a simulated prebiotic setting. We calculate the free energy cost of localizing a mononucleotide, Uridine monophosphate (UMP), in distinct membrane settings and we perform all-atom molecular dynamics (MD) simulations to estimate the role of the monophasic and biphasic membrane in modifying the behavior of UMPs localization and their clustering mechanism. Based on the free-energy and diffusion data from our MD calculations, we develop a lattice based model to explore the thermodynamic limits of the observations made from the MD simulations. The mathematical model substantiates our hypothesis that the lipid layers can act as unique substrates for 'catalyzing' polymerization of mononucleotides due to the inherent spatiotemporal heterogeneity and phase change behavior.

## 1 Introduction

The universe we live in has an estimated age of around 13.7 billion years and dating studies have placed the age of the earth to be around 4.5 billion years[1]. The earliest form of life has been estimated to have evolved around 3.7 billion years ago[2]. One of the most interesting and unsolved scientific question of this century is how the living world evolved from the purely non-living "chemical world" on the prebiotic Earth. The open ended nature of the origin of life (OoL) problem and disparate avenues of research has led to the existence of numerous and eclectic theories in the field[3-6]. So much so that there are conflicts about accepted definition of "life" itself[7]. Nonetheless,

almost all theories acknowledge that cellular life has its origin in common chemistry of the prebiotic chemical world. The general postulate in the field is that simplest possible form of earliest life would be a protocell[8] defined by the presence of information encoding genetic material in a spatially localized compartment[9]. Among the many theories, the "RNA world" theory has gained substantial traction due to the well-known roles of RNA in metabolism, enzymatic reactions and its capability to navigate the conflicting demands of phenotype and genotype[10-12]. In this regard, RNA is thought to be the candidate for the first genetic material on the prebiotic Earth[13-17], which would have enabled the transition from chemistry to biology.

Nucleotides, the building blocks for RNA, 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[18,19]. 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 time[20]. An ongoing investigation in the molecular OoL research has been the understanding of the formation and synthesis of the first polymers of nucleic acids[21-23]. The comprehension of how RNA-like

[a] *Department of Biological Sciences, Indian Institute of Science Education and Research, Kolkata, West Bengal 741246, India*

[b] *Molecular Biophysics Unit, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India*

[c] *Department of Physics, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India*

[d] *Solid State and Structural Chemistry Unit, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India*

∗ Corresponding authors, Email: jayashreen@iisc.ac.in, anand@iisc.ac.in

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

molecules were produced and its emergence in the absence of the catalysts and metabolism is elementary but rapidly evolving[24-33].

The formation of RNA polymer from its constituent nucleotides as well as its stability in an aqueous solution is thermodynamically unfavourable. It is generally believed that for an extended half-life, the prebiotic RNA needed to be sequestered in compartments away from the prebiotic buffer[34,35]. RNA polymer enclosed in an assembly of lipid molecules is hypothesised to be one of the earliest protocell. Not only could the lipids safely encapsulate the nucleotides, it is likely to facilitate their polymerization as well. Nearly a decade ago, the concept of guided polymerisation was introduced[36-38], and it was later discovered that RNA-like polymers could be synthesised nonenzymatically[39,40] in different settings reproducing prebiotic cycles such as fluctuations in hydration[26,34,41-43] and the presence of various organising matrix[26-31,44]. Nucleotides in solution are highly diffusive, but diverse substrates provide different type of confinement environments. For example, phospholipid bilayers (membranes) allow 2D confinement[26,31,32], salt confines them in between salt crystallites[30], or nucleotides can confine in the layered structure of clay[27,28,45-47]. Lipid membranes are of particular interest in this regard considering their role as a substrate for pre-polymer formation as well as in encapsulating the polymer thus formed. A sufficient variety of organic materials were available in the prebiotic settings that allowed lipid-like amphiphiles to self assemble into membranes or membranous vesicles[48]. For cellular life to evolve, a system must have defined boundaries from its environment[49-56]. Fatty-acid based membranes would provide such a plausible boundary structure for the earliest cellular compartmentation. The ordering of mononucleotides, Adenine Monophosphate (AMP) and Uridine Monophosphate (UMP), has been shown in 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid matrix using X-ray diffraction[31,32].

However, in the exploration of lipid layers as the nucleotide organizing matrix, the unique ability of the lipid membrane to undergo phase transition due to changing temperature and hydration has been overlooked so far. In membranes, the "liquid-ordered" ($L_o$) phase, comprising mostly saturated lipids is characterized by high conformational order and sub-diffusive behaviour. "Liquid-disordered" ($L_d$) phase is a highly fluid state where the movement of the lipids are unhindered and can cover the entire surface of the membrane[57,58]. A phase separated membrane can bypass 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. Past experimental studies have shown the varied organisation and ordering of mononucleotides in lipid environments[26,31,32,59] but a molecular-level insight into the mechanism leading to their assembly has been unexplored.

Stochastic-modeling based computational models[42,60] have provided evidence towards lipid membranes as ideal substrate for polymerization of nucleotides. The phase rich behaviour of lipids as a function of composition, temperature and hydration[61] makes it appealing to explore their role in the prebiotic era of formation of protocells. Membrane phases provide a dynamic multi-phasic substrate upon which nucleotides can assemble and polymerize. In this work we use all-atom molecular dynamics along with enhanced sampling technique (Umbrella Sampling) to evaluate the affinity of uridine monophosphate(UMP) nucleotides to $L_o$ and $L_d$ phases of lipid bilayers to evaluate the role of slow and fast diffusing lipid membranes on the assembly of mononucleotides. Further, we generated a lattice model to study aggregation of nucleotides in the presence of these phases. The rest of the article is organised as follows. The article is organized as follows. In Materials and Methods, we discuss the protocol used for all-atom molecular simulations along with post simulation analysis methods followed by detailed explanation of lattice model. We discuss our main results in the Results and Discussion. Finally, we conclude with a summary and future directions in the Conclusions section.
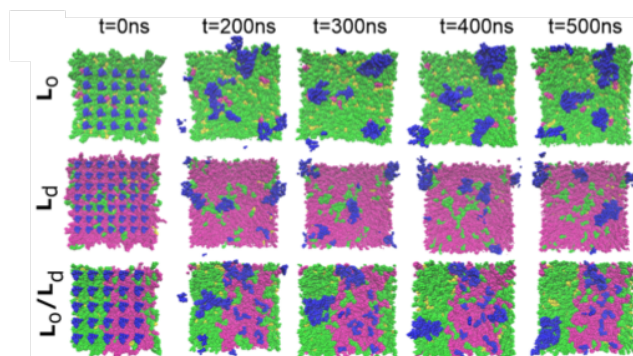


**Fig. 1 Simulation of nucleotides in a PSM/DOPC/Chol lipid environment using AA-MD:** The t = 0 ns shows the nucleotides arranged in arrays on the membranes, and at t = 500ns presents the final snapshot of the system. PSM and DOPC is shown in green and magenta respectively, Chol is represented by yellow, UMP is displayed in blue.

**Table 1** Description of PSM/DOPC/Chol membranes in various phases and their composition

| Phase | Composition | Temp (K) | Lipid Lipid | Area per Lipid |
|---|---|---|---|---|
| $L_o$ | 0.64/0.03/0.33 | 295 | 560 | 40.4 |
| $L_d$ | 0.15/0.82/0.03 | 295 | 574 | 62.7 |
| $L_o/L_d$ | 0.43/0.38/0.19 | 295 | 452 | (N/A) |

## 2 Methods

### 2.1 All-atom molecular dynamics (AA-MD) simulations

Three distinct phases $L_o, L_d$ and $L_o/L_d$ with varying composition of Palmitoylsphingomyelin(PSM), 1-2dioleoyl-sn-glycero-3-phosphocholine(DOPC) and cholesterol(Chol) (Table 1) are considered in this work. An ensemble of UMPs are arranged in arrays on each membrane (Fig. 1 (A)) and have been named as $L_o$-UMP, $L_d$-UMP and $L_o/L_d$-UMP systems (Table 2). All the simulations were performed using GROMACS 5.1.4[62] molecular dy-

namics engine employing CHARMM-36[63] force field for 500 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 integrator[64]. The temperature was held constant with the Nosé-Hoover thermostat[65,66] at 295 K and pressure was maintained by Parrinello-Rahman barostat[67] at 1 bar with relaxation times of 1.0 ps for each case. Standard procedures were used to include the influence of long-range electrostatics[68] and Lennard-Jones (LJ) interaction calculations[69,70]. For van der Waals interactions, cutoff and shift distances of 1.2 nm and 0.9 to 1.2 nm, respectively. Analysis tools included those in GROMACS as well as custom-built scripts.

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

| System | Temp | Molecule or ion | Number of molecules | Simulation time(ns) |
|---|---|---|---|---|
| $L_o$-UMP | 295K | PSM | 362 | 500 |
| | | DOPC | 14 | |
| | | Chol | 184 | |
| | | UMP | 60 | |
| | | Na$^+$ | 120 | |
| $L_d$-UMP | 295K | PSM | 84 | 500 |
| | | DOPC | 460 | |
| | | Chol | 30 | |
| | | UMP | 64 | |
| | | Na$^+$ | 128 | |
| $L_o/L_d$-UMP | 295K | PSM | 194 | 500 |
| | | DOPC | 170 | |
| | | Chol | 88 | |
| | | UMP | 72 | |
| | | Na$^+$ | 144 | |

*Post processing tools.* These include tools for (1) estimation of interaction energy between the nucleotide and the membrane phases, (2) characterization of clusters, and (3) potential mean force (PMF) profiles. We briefly outline these methods:

(1) Interaction energy estimates: To estimate the interaction energies 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[71,72] for the nucleotides is given by

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

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

(2) Characterization of clusters: DBSCAN, or Density-Based Spatial Clustering of Applications with Noise, a clustering algorithm is a density oriented approach to clustering[73] and was implemented using Python programming[74]. It is deterministic, always generates the same clusters when given the same data in the same order. The DBSCAN algorithm uses two parameters:

(a) *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.

(b) *eps* ($\varepsilon$) - 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.

(3) Potential of mean force calculations for UMP-Membrane association using Umbrella Sampling method Umbrella Sampling simulations were performed in three distinctive environments: a ternary mixture constituting of PSM, DOPC, and Chol at various molar ratio. Figure 2A demonstrates the three membrane systems considered and an umbrella sampling simulation was envisioned for each combination of the membrane environment and UMP. For $L_o$ and $L_d$ membranes, the reaction coordinate along the bilayer normal, z, was utilized and the distance between the membrane midplane and the solute molecule was used as the collective variable. In the phase-separated membrane, the reaction coordinate was along the x-axis on the phosphate plane, the distance between the centre of mass on the $L_d$ to the centre of mass of lipid headgroups on the $L_o$ was considered. The individual simulation consisted of at least 50 windows separated every 0.1 nm, where the force constant of the harmonic restraint adjusted to 1000 kcal mol$^{-1}$ nm$^{-2}$.

All the systems were energy minimized and equilibrated and then continued for production runs. The final production runs were carried out in NPT ensemble at a pressure P=1 bar with relaxation times of 1.0 ps for each case using the Parrinello-Rahman barostat[67] and the temperature was kept constant at T=295K using the Nosé-Hoover thermostat[65,66]. Estimation of PMF profiles was carried out by applying the weighted histogram method (WHAM)[75] and the relative errors were estimated by the bootstrapping analysis[76] as enforced in GROMACS[77].

## 2.2 A Lattice model of the membrane to simulate nucleotide clustering behaviour

We developed a spatio-temporal lattice model (Figure 3), which describes the phenomenon of nucleotide cluster formation on a lipid membrane. This phenomenological model was inspired from the theory of lattice models[78]. The primary idea is to equate each nucleotide as a discrete variable that keeps moving around a lattice, and all interactions are specified only between nearest neighbour pairs, and such concepts have been employed in previous works[79–82]. Most noted is to describe polymers[79,80]
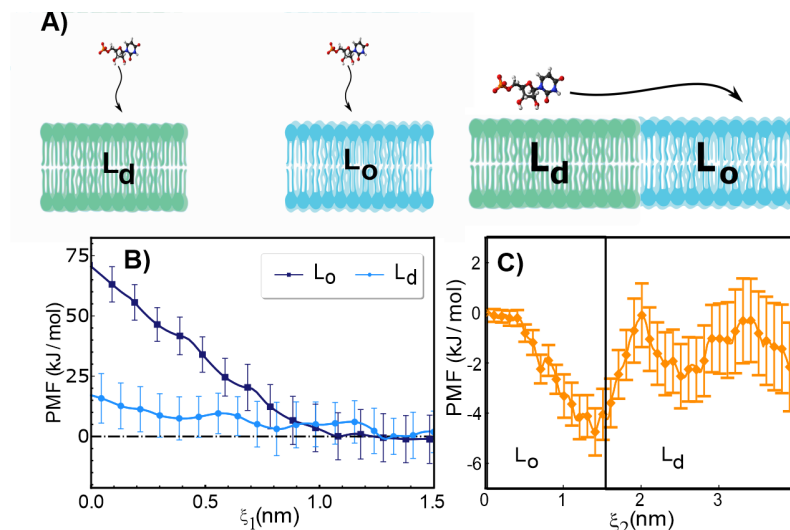
**Fig. 2** (A) L$_o$ & L$_d$ system - The UMP is pulled perpendicular to the plane of the bilayer from the solution to the inside of phosphate plane. L$_o$/L$_d$ system - In the phase-separated system, the UMP is pulled across the phosphate plane of the membrane from the L$_d$ region to the L$_o$ domain. (B & C) **PMF of the UMP in different phases of PSM/DOPC/Chol bilayers at 295 K:** (B) PMF of the L$_o$ and L$_d$ domain of PSM/DOPC/Chol plotted as a function of the distance of UMP from the phosphate head group of the central lipid in the respective phases ($\xi_1$). (C) PMF of UMP in the phase-separated PSM/DOPC/Chol as a function of the distance of UMP from the phosphate head group of the central lipid of the L$_o$ phase ($\xi_2$). 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$_d$.

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 × 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, i.e., either ordered or disordered.

The distribution of the UMP in different phases of the membrane is computationally simulated using a random walk[83-85] on a two dimensional `L × L` lattice with periodic boundary conditions (PBC)[86] 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 sets of parameters; `N` - the number of nucleotides, `L` - the size of the square lattice, $p$ - the hop parameter, $\tau, \tau_1$ - the interaction parameters and `I`$_0$, `I`$_1$, `I`$_2$ - the interface interactions. At each time step, two decisions are made by the nucleotide at a grid point. These are whether the nucleotide hops or not, followed by the choice of location to hop. These decisions are influenced by (a) the probability parameter `p` and (b) the interaction parameters between the nucleotides $\tau$ and $\tau_1$. The probability that a nucleotide decides to stay at its position at a given time

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, and 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. The intuition here is that on a disordered lipid, the nucleotides are more free to move around and thus have a lower probability to stay on a site thus having a lower value of $p$. Similarly because the nucleotides are more restricted on the ordered lipid and prefers to stay at its site and thus will have a higher value of $p$. Note that these are motivated by the diffusion behavior observed in all-atom molecular simulations.

The effect of $\tau$ and $\tau_1$ on the decisions made by a nucleotide at a grid point `(x,y)` are through the 'potential' matrix denoted as `POT[x][y]`. The matrix is a measure of the effect of any nearby nucleotides on the hopping decision of a given nucleotide; it is initialized to 0 at the beginning of the simulation. The matrix then changes as a function of time where each element `POT[x][y]` is updated based on the distribution of nucleotides in the lattice. If there is a nucleotide at the grid point `(x,y)` then the `POT[x][y]` is increased by a value of $\tau$ and the values at `POT[x±1][y]`, `POT[x][y±1]` and `POT[x±1][y±1]` are increased by $\tau_1$ such that $\tau > \tau_1$. This is repeated for all grid points based on whether there is a nucleotide present or not, and all the contributions to a given grid point $(x,y)$ are added. This will result in a distribution of values for the `POT` matrix that will influence the hopping probabilites in the next time step. 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
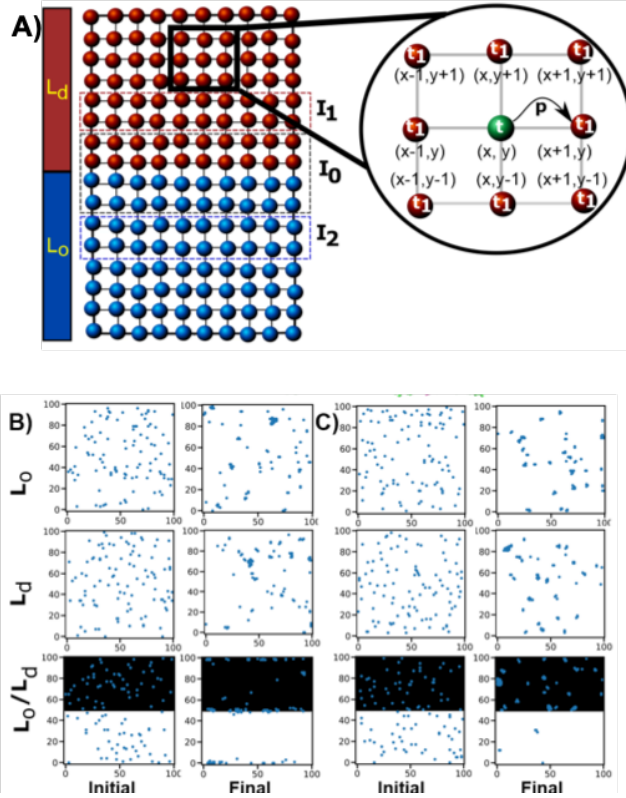
**Fig. 3 (A)Schematic representation of the spatio-temporal model developed in this work:** The phase separated bilayer matrix has been represented using blue beads for $L_o$ and red beads for $L_d$ respectively. The nucleotide has been shown in green shade. The parameters used in the model have been explained; $p$ is the probability of the nucleotide to move in the lipid matrix which differs in the $L_o$ and $L_d$ phase. $\tau$ and $\tau_1$ 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. **(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) and (C) represent simulations of 100 nucleotides on $100 \times 100$ lattice, with the phases $L_d$ and $L_o$ characterized by $p_{dis} = 0.4$ and $p_{ord} = 0.2$ respectively, and the interface 'potential' parameters $I_0 = 65$, $I_1 = 25$, and $I_2 = 5$. (B) shows the steady state results of simulations with $\tau = 100$ and $\tau_1 = 90$, and (C) with $\tau = 700$ and $\tau_1 = 500$.

$(x, y)$ to one of the nearest neighbour $(x', y')$ with probability given as follows:

$$\frac{\text{POT[x'][y']} + 10 - \tau_1}{\sum_{i=\pm 1}\left(\ \text{POT[x+i][y]} + \text{POT[x][y+i]} + 20\ \right) - 4\tau_1} \quad (2)$$

In equation 2, the subtraction of $\tau_1$ is exercised to discard values contributed by the nucleotide itself. To calculate the probabilites, *a biasing term*, in this case 10, was introduced to POT matrix to avoid undefined division by 0. Note that the choice of biasing term does not affect the results of our model if the parameters are scaled appropriately. The motivation behind the biasing term is to ensure non-zero probability of receiving a nucleotide to sites with zero potential. The $\tau$ and $\tau_1$ in the calculations have to be scaled appropriately to the biasing term.

The update procedure can be related to that of a Monte Carlo(MC) simulation[87,88]. In a Monte Carlo simulation, the particle would first choose a spot to hop to, and then make a choice whether to accept or reject the hop. Here, the decision to hop is made first and the spot to hop to is made next. The propagation of the model simulations is summarized as follows: At the initial time point, the nucleotides are randomly distributed on the lattice and POT=0 is updated based on the nucleotide distribution. At the next time step, hops are made based on the $p$ value and the POT matrix is updated based on the new distribution. This is then repeated for the desired length of the simulation.

The interface is modelled at the line $y = \frac{L}{2}$, where $L$ is the length of the lattice, with all points above and below this is described as an ordered and disordered phase respectively of the membrane in the simulation. Due to PBC, the interface was also situated at $y = 0, L$. We introduced three additional parameters $I_0$ for interface, $I_1$, and $I_2$ to model the potential at the closeby interface region of the $L_d$ and $L_o$ phases in a phase-separated membrane respectively. The potential value was increased at and two layers above and below the interface by $I_0$. The two boundaries/interface of the $L_d$ phase have an additional potential of $I_1$ and the two layers interface of the $L_o$ phase are modeled with an increase in potential of $I_2$.

*Parameters for the model* In the model, $p$ is a measure of how "ordered" a lipid 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 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 $I_0$.

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. Table 3 summarises the various simulation setups with different parameter combinations.

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

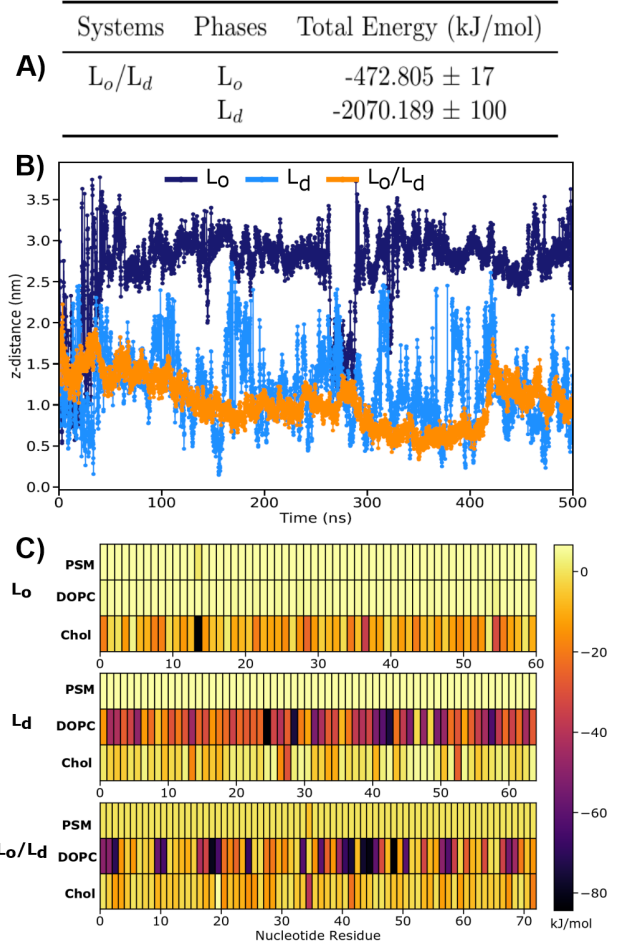| Case | $p_{ord}$ | $p_{dis}$ | $\tau$ | $\tau_1$ | $I_0$ | $I_1$ | $I_2$ |
|---|---|---|---|---|---|---|---|
| 1A | 0.2 | 0.2 | 100 | 90 | - | - | - |
| 1B | 0.2 | 0.2 | 700 | 500 | - | - | - |
| 2A | 0.4 | 0.4 | 100 | 90 | - | - | - |
| 2B | 0.4 | 0.4 | 700 | 500 | - | - | - |
| 3A | 0.4 | 0.2 | 100 | 90 | 65 | 25 | 5 |
| 3B | 0.25 | 0.2 | 100 | 90 | 65 | 25 | 5 |
| 4A | 0.4 | 0.2 | 700 | 500 | 65 | 25 | 5 |
| 4B | 0.25 | 0.2 | 700 | 500 | 65 | 25 | 5 |
| 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 |

## 3 Results and Discussion

### 3.1 AA-MD results

*Preference of nucleotides amongst different lipid membranes.* Lipid membranes constitute one kind of lipid aggregates, which are formed when the lipid head and tail groups are similar in size. In each layer, the hydrophobic tails are oriented towards each other. Recent X-ray diffraction studies examined the ordering of AMP and UMP mononucleotides in the DMPC lipid matrix and found that the nucleotides organized themselves into stacks of about 10 monomers[31,32]. Here we explore the effect of the membrane phases and their chemical composition on the aggregation of the mononucleotides on the membrane preference of each nucleotide.

Figure 2(A) shows the different membrane phases considered in this work. In order to assess the relative preference of the UMP between the $L_d$ and $L_o$ phases, we plot the potential of mean force (PMF) profiles of the approach of UMP towards the isolated phases in Fig. 2(B) and the phase-separated system in Fig. 2(C). In Fig. 2(B), we see that when UMP approaches the $L_o$ or $L_d$ phases (seen from right to left as $\xi_1$ decreases, the mononucleotide sees a higher barrier to an entry into the $L_o$ phase ($\sim 75$ kJ/mol) as compared to $L_d$ phase ($\sim 24$ kJ/mol). This suggests that the UMP will prefer the $L_d$ phase over the $L_o$ phase when we consider a single membrane system. On the other hand, when we take a composite $L_o/L_d$ system, the PMF profile in Fig. 2(C) conveys a preference of the interface (marked by a vertical line). As UMP moves from the $L_d$ to the $L_o$ phase, we can see that the landscape is much more rugged in the $L_d$ side because of the local heterogeneities in the $L_d$ phase. Nevertheless we expect the mononucleotide to find its most stable configuration at the interface. Any excess UMP will then prefer the $L_d$ phase because of the low-lying minima.

To understand the relative preference of the UMPs towards the $L_d$ phase, we explored two things: the interaction energy between the UMPs and the membrane, and the variation in the



| Systems | Phases | Total Energy (kJ/mol) |
|---|---|---|
| $L_o/L_d$ | $L_o$ | $-472.805 \pm 17$ |
| | $L_d$ | $-2070.189 \pm 100$ |

**Fig. 4** (A) Total interaction energy, Coulombic and van der Waals interactions, between UMP and the $L_o$ and $L_d$ domains of the phase separated membrane system (B) **The z-distance calculation between UMP and phosphate plane of the membrane:** UMPs move close 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.

average vertical distance between the UMPs and the phosphate head groups over time. Firstly, we evaluated the interaction energy between UMP and the phases; the interaction energy is composed of Coulombic and van der Waals' contributions (Fig. 4(A)). We see that the $L_d$ phase offers greater stabilization for the entering UMP molecule. In Figure 4(B), we show the evolution of ensemble-averaged vertical distance with time, and we see that the UMPs prefer to interact with the composite system as well as with the isolated $L_d$ phases thereby achieving lower z-distance values. With time, as expected, the UMPs are able to come closer to the lipids. The average z-distance is also therefore an indicator of the increased propensity of the UMPs towards the $L_d$ phase. Further, we also analysed the individual contributions
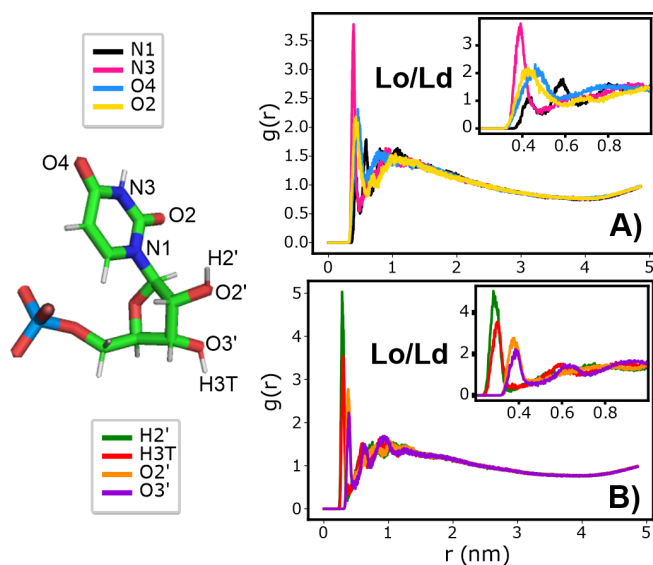
**Fig. 5** Left: Chemical structure of UMP and the nomenclature of their atoms. (A) represents the radial distribution function profiles of the pyrimidine base atoms of UMP around the phosphorous atoms of the membrane lipids. (B) shows plots of the RDF profiles of the ribose sugar atoms of UMP around the phosphorous atoms of the membrane lipids. The N3 atom of the nitrogenous base and H2' atom of the sugar are seen to approach closest to the lipid phosphate head groups.

(PSM, DOPC and Chol) of each component of the membrane towards the interaction energy. Figure 4(C) shows a heat map of such energies for various lipids in the membrane, and we see that the DOPC interacts more with the nucleotides in both $L_d$ and composite $L_o/L_d$ systems. When unsaturated lipids are not present then the cholesterol component of the membrane is seen to interact more with the nucleotides ($L_o$). Together, these observations suggest the differential partitioning of nucleotides in a membrane that can co-exist with highly diffusive ($L_d$) and sub-diffusive ($L_o$) phases. We hypothesize that the nucleotides could have exploited the unique ability of membranes to exist in different phases as an organizing substrate for synthesizing pre-polymers and/or RNA-like polymers.

In order to analyse the propensity of the UMP atoms to interact with the lipid phosphate groups, we plotted the radial distribution functions (Fig. 5) of selected atoms of the pyrimidine nitrogenous base as well as the sugar groups and the phosphorous atom of the lipid phosphate group. The plots reveal a distinct difference between (a) N3 atom and the rest (N1,O2,O4) considered in pyrimidine, and (b) H2' atom and the rest (O2',O3' and H3T). The $g(r)$ peaks are sharper for N3 and H2' atoms, indicating a shorter distance of approach to the phosphate head group (and therefore leading to a stronger interaction) and also a tigher packing of these atoms near the lipid phosphate layer. All of these observations are in line with earlier investigations[89]. Similar results were obtained for the isolated phases and are shown in the supplementary information (Fig. S1).

*Analysis of clusters in AA-MD simulations.* The DBSCAN algorithm revealed the existence of several clusters: eight in $L_o$-UMP, 12 in $L_d$-UMP and 14 in $L_o/L_d$-UMP systems. Such self-assembly

of UMPs in the lipid environment have been observed before and our findings are consistent with these[26,31,32,89]. Additionally, the change of the solvent-accessible surface of UMPs with time (Fig. S2) and their respective solvation energy confirms their clustering behaviour. The solvent accessible surface area decreases over time indicating that the process is driven by hydrophobic effects. This accounts for the corresponding reduction in the solvation energy of the UMPs with time. It establishes that the clusters were formed in the 500 ns of simulation time. The radius of gyration ($R_g$) of the biggest clusters formed were in the range of ~4-5 nm. However, it was found the UMPs in solution form cluster with higher $R_g$, ~6-7 nm (data not shown). We anticipate that the clusters are stabilised by hydrogen bonding or $\pi$-$\pi$ interaction between the constituent mononucleotide. It can be further stabilised by cation-$\pi$ interaction in the proximity of the $Na^+$ counterions as observed for clusters in solution. 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.

## 3.2 Lattice model results

In this section, we describe the observations from implementing a lattice model of clustering behaviour of nucleotides and compare them to the results from the AA-MD simulations. The most important aspect of this exercise is that it provides us insight into the physical processes that arise due to the presence of an interface.

*Effect of the interface on clustering behavior.* Figures 3 (B) and 3(C) show the results of simulating 100 nucleotides on a 100×100 lattice with two sets of inter-nucleotide interactions $\tau$ and $\tau_1$. In the pure phases, any difference of interaction parameters does not make a visible difference to the number and size of the clusters formed. When an interface is introduced between $L_o$ and $L_d$ phases however, an increase in the interaction strength ($\tau$ going from 100 to 700 and $\tau_1$ increasing from 90 to 500) allows for the formation of larger clusters on the whole. Additionally, because of the additional potentials $I_0$, $I_1$ and $I_2$, the nucleotides also tend to aggregate at the interface. In the case of lower interaction, the clusters tend to dissipate as evident from the smaller sized clusters; conversely the higher interaction allows for formation of larger clusters shown in Fig. 1(C) for $L_o/L_d$ phase. Our speculation is that such clusters when they do migrate back into the bulk phases, do so with different probabilites as suggested by the energy landscape in Fig. 2(C). This then allows for formation of larger clusters and bonds between species with repeated migration attempts. The results from the model simulation agree reasonably with the AA-MD results in Fig. 1(A); due to size limitations, the AA-MD simulations could not be run for as long as the lattice model simulations.

*Effect of the phases on clustering behavior.* Figure 6(A) shows the
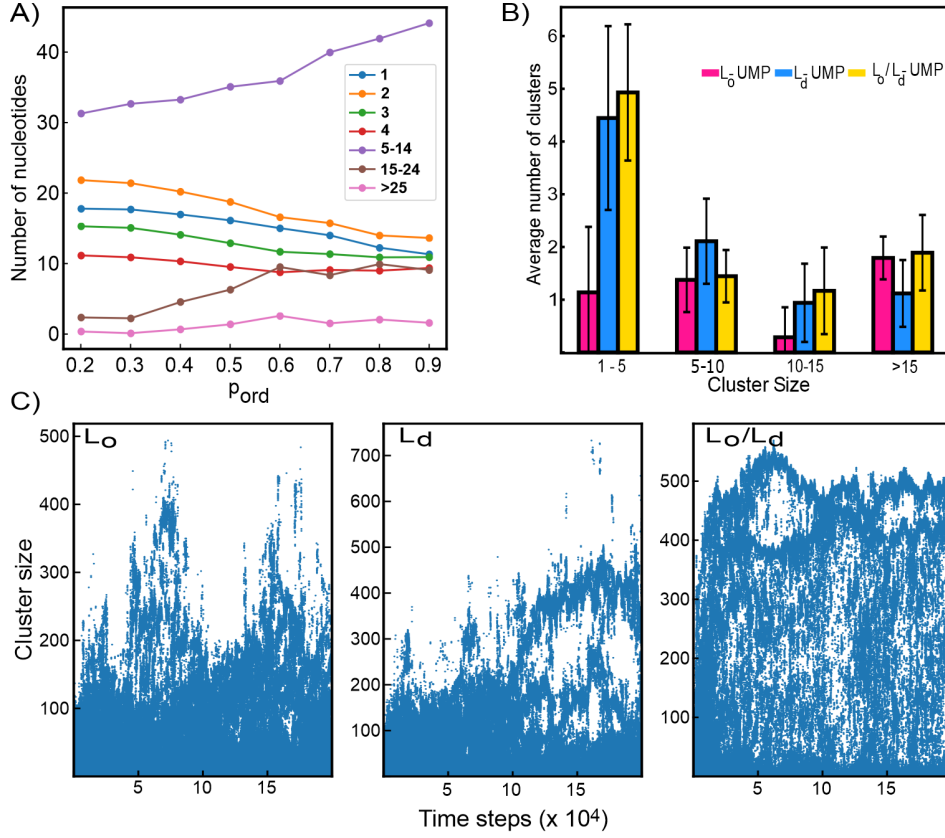
**Fig. 6** (A) Plot of cluster size variation with different amounts of ordering in the phase ($p_{ord}$ parameter), (B) Histogram plot of the average number of clusters in AA-MD simulations for different cluster sizes in the last 100ns, (C)Temporal cluster distribution in different phases of membrane using computer modelling in `100 × 100` matrix with `1000` nucleotides, $L_o$ phase characterized by $p_{ord}$=0.2, $L_d$ phase characterized by $p_{dis}$=0.4, `I`$_0$=65, `I`$_1$=25, `I`$_2$=5, $\tau = 100$ and $\tau_1 = 90$.
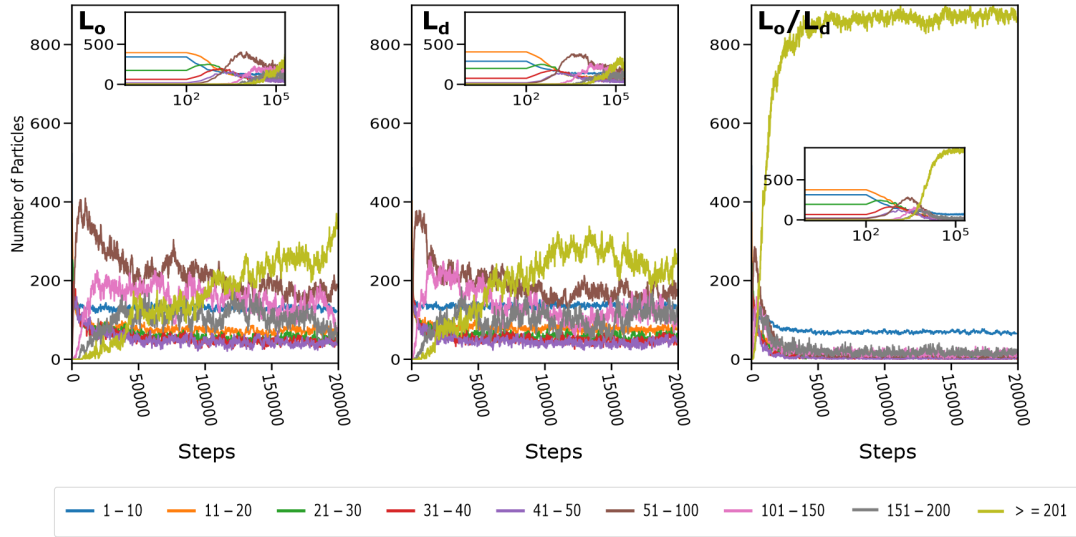


**Fig. 7 Cluster evolution and steady state distribution in `100 × 100` matrix with `1000` nucleotides:** (B) $L_o$ with $p_{ord}$=0.2, $\tau$=100 and $\tau_1$=90, $L_d$ with $p_{dis}$=0.4, $\tau$=100 and $\tau_1$=90, $L_o/L_d$ with $p_{ord}$=0.2 and $p_{dis}$=0.4, $\tau$=100 and $\tau_1$=90, `I`$_0$=65, `I`$_1$=35, and `I`$_2$=5.

variation in the cluster sizes with $L_o/L_d$ membrane modeled with no additional potential for the interface `I`$_0$ =`I`$_1$ =`I`$_2$ = 0 as a function of the $p_{ord}$ and fixed $p_{dis}$ for a system of 100 nucleotides. The

tendency to form larger clusters increases with increase in difference between $p_{ord}$ and $p_{dis}$. We see that if the difference in the diffusive behaviour between the two phases is large enough, then

we can get clustering behaviour without any interface potential.

Next, we analyze a system with 1000 nucleotide and a phase separated region that includes a potential interface (case 3A in Table 3), and a uniform phase system of either ordered or disordered (cases 1A and 2A in Table 3) without including any interfacial potential parameters. (Figs. 6C). The latter corresponds to a pure phase. In either the purely ordered or disordered phase of lipids, the cluster sizes near 500 are not very populated as a function of time and thus these clusters if they form have a very short lifetime. However, the addition of an interface stabilizes these clusters. This can be seen because of the large clusters of size 500 that forms in the presence of the interface. These results suggest that both the interface and existence of different phases play a key role in influencing cluster formation. We have verified that the same conclusions hold for cases 1B,2B and 3B in Table 3.

*Cluster distribution and dynamics in the lattice model.* The distribution of clusters in the lattice model is analysed in two ways: Fig. 6(C) show the time evolution of different sized clusters and Fig. 7 represents a plot of the number of nucleotides belonging to different clusters evolving over time. The plot in Fig. 6(B) shows the average number of different sized clusters formed in the last 100ns in the AA-MD simulations; these reach about a value of 5 for the smaller sized clusters (1-5) and about 2 for the largest sized clusters (>15). The atomistic simulations are limited in the sense that one cannot simulate clusters of several hundreds in a reasonable amount of time with the current resources, which lead us naturally to testing such a time evolution using the lattice model. The plots in Fig. 7 clearly show that as we reach the end of the simulation, the $L_o/L_d$ phase is able to generate a much larger number of very big (> 500 participating nucleotides) clusters compared to the pure phases in the lattice model simulations. The latter plot reveals that a greater number of nucleotides engage in clusters of size > 200 towards the end of the simulations. Both essentially point to the fact that the presence of the phases as well as the inter-phase potential is critical in the formation of large-sized clusters, which in turn can seed the formation of complex assemblies that might have evolved in the course of evolution.

## 4  Conclusions and Outlook

In this work, we have focused on the importance of lipid substrate as a organizing matrix for nucleotide assembly. The work borrows from past work in the field that investigated membranous compartments as substrates for nucleotides' (UMP and/or AMP) non-enzymatic polymerisation[26,31,32]. However, in our work, we have emphasized the role played by the multiphasic nature of lipid layers. 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 compositions[90-94]. 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 that 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 on membranes, which is in accordance[31,32,89] 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. Our MD simulations provide detailed molecular insight for the rearrangement of nucleotides into clusters. Further, guided by inputs from molecular dynamics simulations based interactions and to scale our modeling to large thermodynamic limits, we developed a mathematical model and showed 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, together our work shows how the amphiphilic phospholipid can possibly act 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 as an organising matrix as it could act as a driving factor in the pre-polymer or RNA-like molecule formation, which is an indispensable prelude to the molecular OoL.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## Notes and references

1  G. B. Dalrymple, *The Paleontological Society Papers*, 1999, **5**, 17–22.

2  A. Nutman, V. Bennett, C. Friend and M. Van Kranendonk, *Nature*, 2016, **537**,.

3  K. Michaelian, *Heliyon*, 2017, **3**, e00424.

4  K. Michaelian, *Earth System Dynamics*, 2011, **2**, 37–51.

5  K. Michaelian, *arXiv preprint*, 2009, arXiv:0907.0042.

6  B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy and J. D. Sutherland, *Nature chemistry*, 2015, **7**, 301–307.

7  J. Gayon, C. Malaterre, M. Morange, F. Raulin-Cerceau and S. Tirard, *Origins of Life and Evolution of the Biosphere*, 2010, **40**, 119.

8  I. A. Chen and P. Walde, *Cold Spring Harbor perspectives in biology*, 2010, **2**, a002170.

9  J. Schrum, T. Zhu and J. Szostak, *Cold Spring Harbor perspectives in biology*, 2010, **2**, a002212.

10  A. Rich, *Horizons in biochemistry*, 1962, 103–126.

11  F. H. Crick, *Journal of molecular biology*, 1968, **38**, 367–379.

12  C. R. Woese, *Naturwissenschaften*, 1973, **60**, 447–459.

13  M. Neveu, H.-J. Kim and S. A. Benner, *Astrobiology*, 2013, **13**, 391–403.

14  S. D. Copley, E. Smith and H. J. Morowitz, *Bioorganic chemistry*, 2007, **35**, 430–443.

15  G. F. Joyce, *Nature*, 2002, **418**, 214–221.

16  M. P. Robertson and G. F. Joyce, *Cold Spring Harbor perspectives in biology*, 2012, **4**, a003608.

17  T. R. Cech, *Cold Spring Harbor perspectives in biology*, 2012, **4**, a006742.

18  R. Langridge and P. J. Gomatos, *Science*, 1963, **141**, 1024–1024.

19  J. D. Watson and F. H. Crick, Cold Spring Harbor symposia on quantitative biology, 1953, pp. 123–131.

20  J. Sheng, L. Li, A. E. Engelhart, J. Gan, J. Wang and J. W. Szostak, *Proceedings of the National Academy of Sciences*, 2014, **111**, 3050–3055.

21  A. I. Oparin, *The origin of life*, Dover Publications, 1953.

22  A. I. Oparin *et al.*, *The origin of life on the earth.*, Oliver & Boyd, Edinburgh & London, 1957.

23  C. Scharf, N. Virgo, H. J. Cleaves, M. Aono, N. Aubert-Kato, A. Aydinoglu, A. Barahona, L. M. Barge, S. A. Benner, M. Biehl *et al.*, *A strategy for origins of life research*, 2015.

24  G. Costanzo, R. Saladino, G. Botta, A. Giorgi, A. Scipioni, S. Pino and E. Di Mauro, *ChemBioChem*, 2012, **13**, 999–1008.

25  M. Morasch, C. B. Mast, J. K. Langer, P. Schilcher and D. Braun, *ChemBioChem*, 2014, **15**, 879–883.

26  S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti and D. Deamer, *Origins of Life and Evolution of Biospheres*, 2008, **38**, 57–74.

27  J. P. Ferris, A. R. Hill, R. Liu and L. E. Orgel, *Nature*, 1996, **381**, 59–61.

28  J. P. Ferris, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2006, **361**, 1777–1786.

29  F. Olasagasti Arsuaga and S. Rajamani, *Lipid-Assisted Polymerization of Nucleotides*, MDPI, 2019.

30  L. Da Silva, M.-C. Maurel and D. Deamer, *Journal of molecular evolution*, 2015, **80**, 86–97.

31  S. Himbert, M. Chapman, D. W. Deamer and M. C. Rheinstädter, *Scientific reports*, 2016, **6**, 31285.

32  L. Toppozini, H. Dies, D. W. Deamer and M. C. Rheinstädter, *PloS one*, 2013, **8**, e62810.

33  T. P. Fraccia, G. Zanchetta, V. Rimoldi, N. A. Clark and T. Bellini, *Origins of Life and Evolution of Biospheres*, 2015, **45**, 51–68.

34  D. W. Deamer and C. D. Georgiou, *Astrobiology*, 2015, **15**, 1091–1095.

35  P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell and D. Braun, *Proceedings of the National Academy of Sciences*, 2007, **104**, 9346–9351.

36  D. Deamer, *Chemical Society Reviews*, 2012, **41**, 5375–5379.

37  A. Blumstein, N. Kitagawa and R. Blumstein, *Molecular Crystals and Liquid Crystals*, 1971, **12**, 215–227.

38  E. M. Barrall and J. F. Johnson, *Journal of Macromolecular Science—Reviews in Macromolecular Chemistry*, 1979, **17**, 137–170.

39  R. Lohrmann, P. Bridson and L. Orgel, *Science*, 1980, **208**, 1464–1465.

40  V. DeGuzman, W. Vercoutere, H. Shenasa and D. Deamer, *Journal of molecular evolution*, 2014, **78**, 251–262.

41  M. Hargrave, S. K. Thompson and D. Deamer, *Journal of molecular evolution*, 2018, **86**, 501–510.

42  P. Higgs, *Life*, 2016, **6**, 24.

43  B. Damer and D. Deamer, *Life*, 2015, **5**, 872–887.

44  D. Deamer, S. Singaram, S. Rajamani, V. Kompanichenko and S. Guggenheim, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2006, **361**, 1809–1818.

45  K. Kawamura and J. P. Ferris, *Journal of the American Chemical Society*, 1994, **116**, 7564–7572.

46  K. Kawamura and J. P. Ferris, *Origins of Life and Evolution of the Biosphere*, 1999, **29**, 563–591.

47  G. Ertem and J. P. Ferris, *Origins of Life and Evolution of the Biosphere*, 2000, **30**, 411–422.

48  W. Hargreaves, S. Mulvihill and D. Deamer, *Nature*, 1977, **266**, 78–80.

49  D. S. Tawfik and A. D. Griffiths, *Nature biotechnology*, 1998, **16**, 652–656.

50  P. L. Luisi, P. Walde and T. Oberholzer, *Current opinion in colloid & interface science*, 1999, **4**, 33–39.

51  J. W. Szostak, D. P. Bartel and P. L. Luisi, *Nature*, 2001, **409**, 387–390.

52  A. Lazcano, *Early life on Earth (ed. S. Bengtson)*, 1994, 70–80.

53  A. Lazcano, Nobel Symp, 1994, pp. 60–69.

54  D. W. Deamer, *Early life on Earth*, 1994, 107–123.

55  D. W. Deamer, *Microbiology and Molecular Biology Reviews*, 1997, **61**, 239–261.

56  G. R. Fleischaker, *Origins of Life and Evolution of the Biosphere*, 1990, **20**, 127–137.

57  M.-C. Giocondi, D. Yamamoto, E. Lesniewska, P.-E. Milhiet, T. Ando and C. Le Grimellec, *Biochimica et Biophysica Acta*

(BBA)-Biomembranes, 2010, **1798**, 703–718.

58 H.-M. Wu, Y.-H. Lin, T.-C. Yen and C.-L. Hsieh, *Scientific reports*, 2016, **6**, 20542.

59 D. Deamer, *Journal of molecular evolution*, 2016, **83**, 159–168.

60 M. Wu and P. Higgs, *Biology direct*, 2012, **7**, 42.

61 R. Koynova and B. Tenchov, in *Phase Transitions and Phase Behavior of Lipids*, 2012.

62 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.

63 J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr and R. W. Pastor, *The journal of physical chemistry B*, 2010, **114**, 7830–7843.

64 W. F. Van Gunsteren and H. J. Berendsen, *Molecular Simulation*, 1988, **1**, 173–185.

65 S. Nosé and M. L. Klein, *The Journal of Chemical Physics*, 1983, **78**, 6928–6939.

66 H. C. Andersen, *The Journal of chemical physics*, 1980, **72**, 2384–2393.

67 M. Parrinello and A. Rahman, *Journal of Applied physics*, 1981, **52**, 7182–7190.

68 T. Darden, D. York and L. Pedersen, *The Journal of chemical physics*, 1993, **98**, 10089–10092.

69 M. Allen, *Tildesley. DJ Computer simulation of liquids*, 1987.

70 P. Lagüe, R. W. Pastor and B. R. Brooks, *The Journal of Physical Chemistry B*, 2004, **108**, 363–368.

71 X. Lin, A. A. Gorfe and I. Levental, *Biophysical journal*, 2018, **114**, 1936–1944.

72 A. Centi, A. Dutta, S. H. Parekh and T. Bereau, *Biophysical Journal*, 2020.

73 M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, Kdd, 1996, pp. 226–231.

74 G. Van Rossum *et al.*, USENIX annual technical conference, 2007, p. 36.

75 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *Journal of computational chemistry*, 1992, **13**, 1011–1021.

76 B. Efron, in *Breakthroughs in statistics*, Springer, 1992, pp. 569–593.

77 J. S. Hub, B. L. De Groot and D. Van Der Spoel, *Journal of chemical theory and computation*, 2010, **6**, 3713–3720.

78 E. Ising, *Ph.D. thesis*, Hamburg, 1924.

79 A. I. Pesci and K. F. Freed, *The Journal of Chemical Physics*, 1989, **90**, 2003–2016.

80 P. J. Flory, *Principles of Polymer Chemistry*, Cornell University, Ithaca, 1953, 1953.

81 R. J. Baxter, *Journal of Statistical Physics*, 1978, **19**, 461–478.

82 A. Ghosal, M. Randeria and N. Trivedi, *Physical Review B*, 2001, **65**, 014501.

83 S. Karlin, J. McGregor *et al.*, *Illinois Journal of Mathematics*, 1959, **3**, 66–81.

84 E. W. Montroll and G. H. Weiss, *Journal of Mathematical Physics*, 1965, **6**, 167–181.

85 G. H. Weiss and R. J. Rubin, *Adv. Chem. Phys*, 1983, **52**, 363–505.

86 G. Makov and M. Payne, *Physical Review B*, 1995, **51**, 4014.

87 J. Griesbauer, A. Wixforth, H. M. Seeger and M. F. Schneider, *Method for the Monte Carlo based Simulation of Lipid-Monolayers including Lipid Movement*, 2010.

88 J. Kotze, *Introduction to Monte Carlo methods for an Ising Model of a Ferromagnet*, 2008.

89 S. Sasidharan, S. Pochinda, P. N. Elgaard-Jørgensen, S. Rajamani, H. Khandelia and V. Raghunathan, *Soft matter*, 2019, **15**, 8129–8136.

90 D. Brown and E. London, *Annual review of cell and developmental biology*, 1998, **14**, 111–136.

91 T. Harayama and H. Riezman, *Nature Ecology & Evolution*, 2018, 1–18.

92 G. Van Meer, D. R. Voelker and G. W. Feigenson, *Nature reviews Molecular cell biology*, 2008, **9**, 112–124.

93 K. Jacobson, O. G. Mouritsen and R. G. Anderson, *Nature cell biology*, 2007, **9**, 7–14.

94 K. Simons and E. Ikonen, *nature*, 1997, **387**, 569–572.

95 M. Ponce, R. van Zon, S. Northrup, D. Gruner, J. Chen, F. Ertinaz, A. Fedoseev, L. Groer, F. Mao, B. C. Mundim *et al.*, in *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*, 2019, pp. 1–8.

96 J. C. Candoré, J. Bodnar, V. Detalle and P. Grossel, Journal of Physics: Conference Series, 2010, p. 012068.

# Supporting Information

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

Rajlaxmi Saha,[a] Prathyush Poduval,[b,c], Krishnakanth Baratam[b], Jayashree Nagesh[d] and Anand Srivastava[b]*

[a] Department of Biological Sciences, Indian Institute of Science Education and Research, Kolkata, West Bengal 741246, India

[b] Molecular Biophysics Unit, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India

[c] Department of Physics, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India

[d] Solid State and Structural Chemistry Unit, Indian Institute of Science Bangalore, C. V. Raman Road, Bengaluru, Karnataka 560012, India
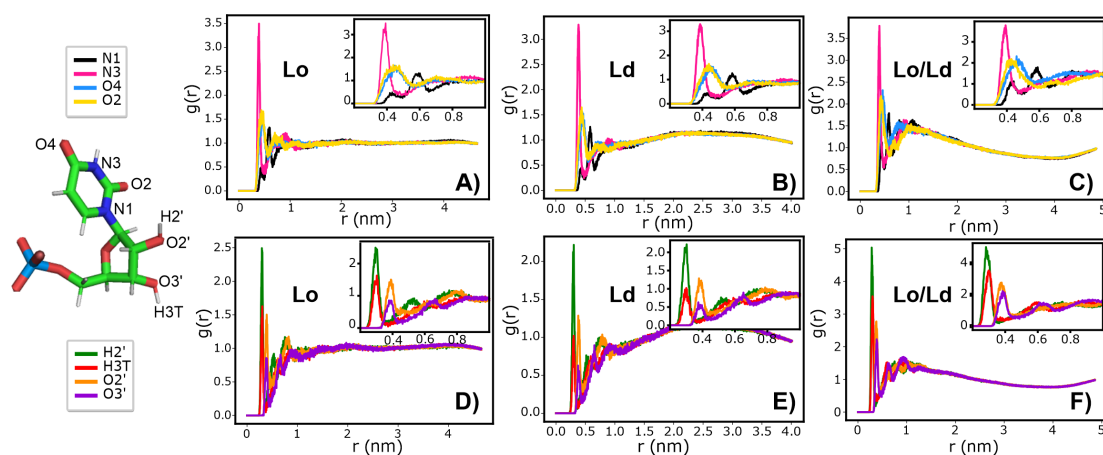
* Corresponding author, Email: anand@iisc.ac.in

Fig. S1: (Left panel) Chemical structure of Uracil Monophosphate and the nomenclature of their atoms according to the PDB (Right panel) RDF of nitrogenous base and ribose sugar atoms of the UMPs around the phosphorous atom of the lipids in membrane. In the nitrogenous base, the N1 of the base shows strongest interaction with the phosphate of the lipids, followed by O4 and O2 atoms. For the ribose sugar atoms, the strongest interaction is between H2' atom of ribose and the phosphate of lipids
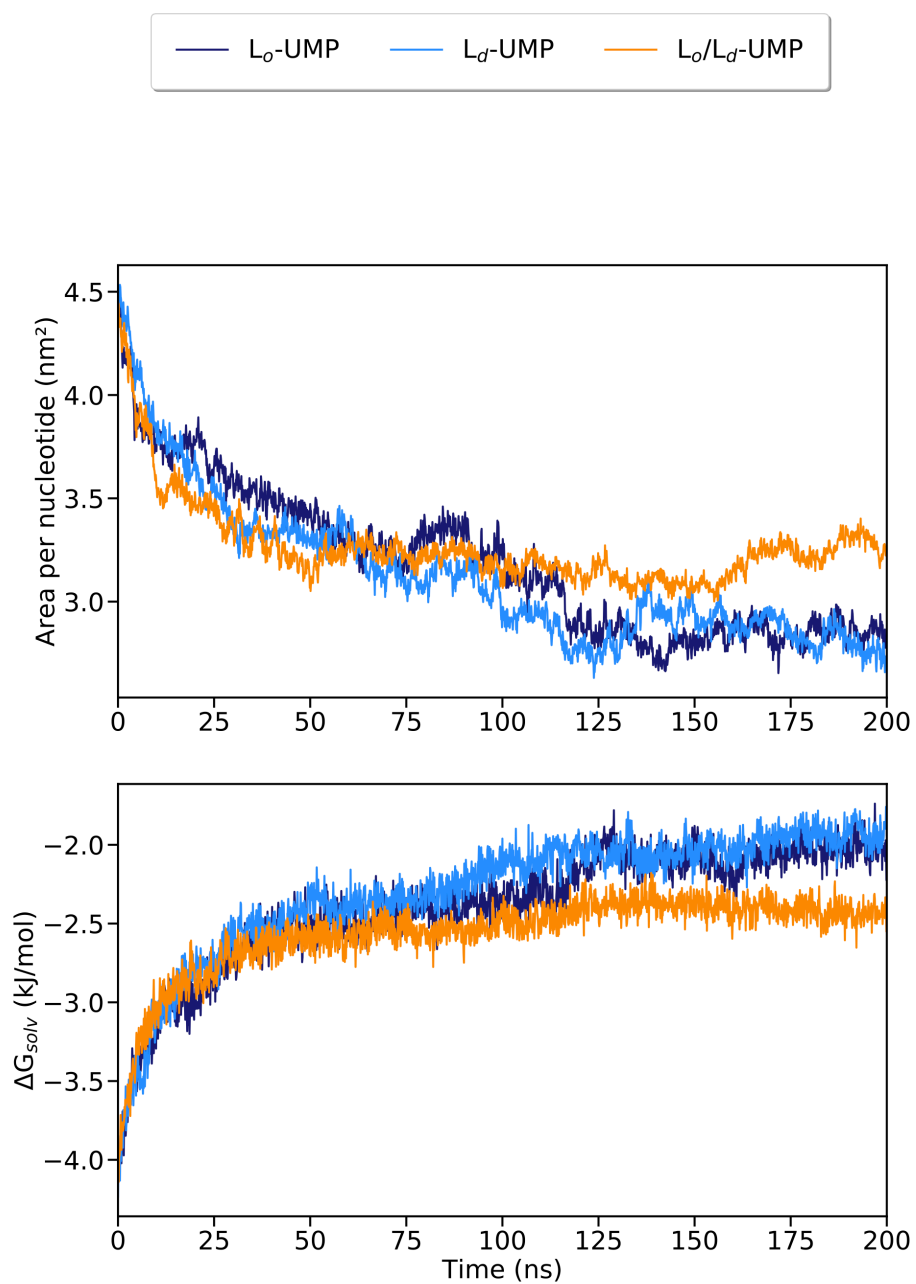
Fig. S2: SASA and the solvation energy of UMPs in the three distinct model membrane systems: (A) The solvent-accessible surface area of UMPs and its change over time in a lipid environment (B) The solvation energy of UMPs