2023

On the Origins of Life — Modelling the Initial Stages of Complex Coacervate Droplet Formation

Yixuan Wu

Western University, London, ON

Follow this and additional works at: https://ir.lib.uwo.ca/wlura

Part of the Biochemistry, Biophysics, and Structural Biology Commons, Computational Chemistry Commons, Condensed Matter Physics Commons, and the Other Life Sciences Commons

Recommended Citation


This Essay is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Western Libraries Undergraduate Research Awards (WLURAs) by an authorized editor of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.
On the Origins of Life —
Modelling the initial stages of complex coacervate droplet formation

Amber Yixuan Wu
*The University of Western Ontario, Department of Medical Biophysics, Honours Specialization in Medical Biophysics and Biochemistry*

Supervisor:
Dr. Styliani Consta
*The University of Western Ontario, Department of Chemistry*

Graduate Student Mentor:
Keith Christian Afas
*The University of Western Ontario, School of Biomedical Engineering*
Abstract

Introduction: Coacervate droplets are considered a plausible model for protocells due to their spontaneous formation and ability to compartmentalize macromolecules essential for life, such as nucleic acid and peptides. Although experimental studies have observed and synthesized coacervates under different laboratory conditions, little is known about how these charged polyelectrolytes affect surrounding water on an atomic level. Previous computational studies used coarse grained molecular dynamics to simulate large systems of charged homopolymers, but the resolution of such models can become restrictive. Here we present atomistic molecular dynamics simulations of the initial stage of interactions between water and oppositely charged homopolymers and copolymers. In addition, since coacervation is known to be a salt-dependent process, it is expected that the ion distributions would be different compared to the surrounding bulk solution. Such partitioning observed on an atomic level would serve as a model for the initial stages of complex coacervate droplet formation.

Methods: Molecular visualization software PyMOL (Version 2.5.4) was used to construct homopolymers and copolymers. Two strands of lysine and two strands of glutamic acid homopolymers each of 50 residues in length were placed in square arrays with different configurations. Four strands of glycine copolymers of the same length were constructed with two containing lysine residues in position 1, 15, 30, 45 and two containing glutamic acid residues located at the same sites. Visual Molecular Dynamics (VMD, Version 1.9.4a57-x86_64-Rev12) were then used to solvate and ionized the protein structures with periodic boundary conditions. The trajectory of each atom was calculated using Nanoscale Molecular Dynamics (NAMD, Version 2.14) on the Narval cluster, through Advanced Research Computing platform provided by Digital Research Alliance of Canada. Using an extended Lagrangian method implicit to NAMD, the number density distributions of both protein and water molecules with regards to their positions from the protein centre of volume was analyzed using MATLAB (Version R2022b). Two-tailed t-Tests were performed to assess the statistical significance of any changes in number density.

Results: Charged copolymers formed a globular, mostly intrinsically disordered protein, while oppositely charged homopolymers formed cylindrical helix quadruplex, or “coiled coil” structures regardless of their initial configuration. Three-dimensional number density models for spherical and cylindrical protein geometries were constructed and displayed using MATLAB. Data analyses showed that sparsely charged copolymers had no significant effect on the number density of the surrounding bulk solution, whereas statistically significant changes in water and total number density was detected in systems containing homopolymers. The dissociation of a copolymer subunit was observed in the presence of 0.15M of NaCl.

Discussion: The significant increase in number density within, and around the protein may represent the interface of coacervate liquid-liquid phase separation. The presence of salt caused a local dissociation of the charged copolymer due to screening effects. This is one of the very first studies to explore the interface of coacervate droplets and surrounding bulk solution on an atomistic level. One limitation of atomistic modelling is its computational expensiveness, so only small-scale systems can be examined with high resolution. Longer simulations are needed to obtain an average conformation of the structures and confirm our results. Future studies should be conducted using a larger system with more complex mixture of protein peptides and nucleic acids, and simulated under more precise prebiotic conditions.
Keywords: coacervate droplets, soft condensed matter, polyelectrolytes, molecular dynamics, origins of life, computational chemistry

Acknowledgements

First, I would like to acknowledge my supervisor, Dr. Styliani Consta, for her patient mentoring and support throughout this project; I would also like to thank other members of the Consta Research Group: Matt Rossini, Nick Bainbridge, Mingyeong Jung, and Neel Vajaria for their constructive criticisms during group discussions, generous sharing of resources, and creating a positive and mentally stimulating research environment.

This project would not be possible without the generous fundings from NSERC and Digital Research Alliance of Canada.

I would like to express my sincere gratitude towards one of my dear friends, Keith Christian Afas, for his invaluable insight into membrane modelling and guidance with constructing the models used in this project.

I would also like to thank my parents, Melody and Tong, and my grandparents for their unconditional love and encouragement. And finally, my two parrotlets, Pineapple and Smokey, for being my emotional support through stressful times.
1 INTRODUCTION .....................................................................................................................1
    1.1 Coacervate droplets and the Origins of Life.................................................................1
    1.2 Motivations ...................................................................................................................2
    1.3 Objectives and hypotheses ............................................................................................3

2 METHODS .................................................................................................................................4
    2.1 Constructing the initial structures using PyMOL..........................................................4
    2.2 Solvate and Ionize systems using VMD .................................................................5
    2.3 Calculating atomic trajectories using NAMD ..........................................................6
    2.4 Trajectory Analysis using MATLAB ............................................................................7
    2.5 Statistical Analysis ........................................................................................................9

3 RESULTS .................................................................................................................................10
    3.1 Analyses of the water structure in the presence of copolymer aggregate.................10
    3.2 Analyses of the water structure in the presence of homopolymer aggregate..........13

4 DISCUSSION ...........................................................................................................................16
    4.1 Copolymer aggregate .................................................................................................16
    4.2 Homopolymer aggregate ..........................................................................................19

5 CONCLUSIONS ......................................................................................................................20

6 REFERENCES ..........................................................................................................................22

7 APPENDIX ..............................................................................................................................24
List of Figures

Fig. 1. Example of a Coarse-Grain model

Fig. 2. Initial copolymer conformations input into NAMD

Fig. 3. Diagonally placed Lys\textsubscript{50}, Glu\textsubscript{50} ionized by 0.15M NaCl

Fig. 4. Unionized and ionized copolymer initial input structures solvated using VMD

Fig. 5. Spherical and cylindrical models used for number density analyses

Fig. 6. Number density analysis of unionized copolymer

Fig. 7. Number density analysis of copolymer ionized with 0.15M NaCl

Fig. 8. Analysis of the main structure number density of copolymer ionized with 0.15M NaCl

Fig. 9. Analysis of the Gly-Glu strand number density of copolymer ionized with 0.15M NaCl

Fig. 10. Helical “coiled coil” structure of Lys\textsubscript{50} and Glu\textsubscript{50} homopolymers in cylindrical model

Fig. 11. Number density analyses of each ring in the 2D cylindrical plot
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origins</td>
<td>Origins of Life</td>
</tr>
<tr>
<td>LLPS</td>
<td>Liquid-liquid phase separation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>CG</td>
<td>Coarse-grained</td>
</tr>
<tr>
<td>CC</td>
<td>Coiled coil</td>
</tr>
<tr>
<td>M</td>
<td>Molarity</td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
</tr>
<tr>
<td>Lys$_{50}$</td>
<td>Lysine homopolymers 50 amino acids in length</td>
</tr>
<tr>
<td>Glu$_{50}$</td>
<td>Glutamic acid homopolymers 50 amino acids in length</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
<tr>
<td>NAMD</td>
<td>Nanoscale Molecular Dynamics</td>
</tr>
<tr>
<td>PBCs</td>
<td>Periodic boundary conditions</td>
</tr>
<tr>
<td>CHARMM36</td>
<td>Chemistry at Harvard Macromolecular Mechanics (CHARMM), v36</td>
</tr>
<tr>
<td>TIP3P</td>
<td>Transferable intermolecular potential with 3 points</td>
</tr>
<tr>
<td>NVT</td>
<td>Constant number (N), volume (V), temperature (T)</td>
</tr>
<tr>
<td>NPT</td>
<td>Constant number (N), pressure (P), temperature (T)</td>
</tr>
<tr>
<td>fs</td>
<td>Femtoseconds ($10^{-15}$ seconds)</td>
</tr>
<tr>
<td>ps</td>
<td>Picoseconds ($10^{-12}$ seconds)</td>
</tr>
<tr>
<td>ns</td>
<td>Nanoseconds ($10^{-9}$ seconds)</td>
</tr>
<tr>
<td>COV</td>
<td>Centre of volume</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 Coacervate droplets and the Origins of Life

One of the most fundamental questions in the field of Origins of Life (abbrv. Origins) research is to understand how low concentrations of organic molecules present in the primordial world can come together and interact in the absence of a lipid cell membrane\(^1,2\). Membraneless cellular compartments such as nucleolus\(^3\) and stress granules\(^4\) have been described and studied since the 1830s\(^5\). Liquid-liquid phase separation (LLPS), such as “oil in water”, are thought to be a major mechanism of formation for such compartments\(^6\). LLPS can be extended to systems of charged biological macromolecules in aqueous solutions. Certain concentrations of charged biological macromolecules such as proteins and RNA may coalesce and form soft droplets, termed “coacervates” or “coacervate droplets”\(^7\). These droplets were first described and named by two Dutch chemists, Hendrik G. Bungenberg de Jong and Hugo R. Kruyt in 1929\(^7\). There are two types of coacervates: simple coacervates includes only one type of polymer, and complex coacervates includes two or more types of polymers. For the interest of this paper, we will only be focusing on complex coacervates, as oppositely charged RNA and peptides likely interacted to form the coacervate droplets relevant for the emergence of the earliest form of cells called protocells.

Complex coacervation forms a dense, polyelectrolyte-rich liquid phase that separates from the bulk solution, as the result of a series of complex interactions between charged macromolecules to balance the electrostatics, hydrophobics, van der Waals, and other contributions, leading to an overall stability of the system.

In the 1920s, Russian biochemist Aleksander Oparin\(^8\) and British biologist J.B.S. Haldane\(^9\) independently put forward the concept of primordial soup, which refers to the dilute mixture of organic molecules that could have existed in water bodies as a result of prebiotic photochemistry.
In Oparin’s subsequent book “Origins of Life” published in 1936 (English version in 1938\(^1\)), he described the studies on coacervates by Bungenberg de Jong and proposed that organic molecules in early Earth's oceans could be up concentrated and enclosed by coacervate droplets, resulting in the formation of the first cells that eventually leads to life. The Oparin-Haldane hypothesis remains a prominent reference guide in the Origins field to this day.

1.2 Motivations

Coacervate droplets have been observed and synthesized for over a century, and techniques of experimental coacervate synthesis have been developed and improved over many years. The accelerated chemical reactivity within the coacervate droplets\(^{10}\) has attracted considerable attention in recent decades. Studies have shown that these droplets can up-concentrate functional RNA in their active folded states, fragments upon decay\(^{11}\), and even proliferate under certain conditions\(^{12}\). These characteristics not only made coacervates a plausible model for protocells in the field of prebiotic chemistry\(^{13}\), but also opened a new window into biomedical research of diseases related to membraneless organelles\(^{14}\), small molecule drug delivery using coacervate droplets\(^{15}\), and even material sciences and biosensor development\(^{16}\).

In comparison, much fewer numbers of theoretical and computational studies of the interfacial formation, structure, tension, and dynamics of coacervates have been proposed and published. Most of the molecular dynamics (MD) model used Coarse-Grain (CG) modelling to examine the behaviours of highly charged polyelectrolytes in coacervates\(^{17}\), their phase behaviour and salt partitioning\(^{18}\), and the effect of polymer length and salt concentrations on coacervation\(^{19}\). However, CG modeling eliminated a lot of crucial structural details by combining and treating
multiple atoms as a single entity, which increased the computational efficiency at the cost of lowering the accuracy.

![Image of multiple atoms as a single entity](image.png)

**Figure 1.** Example of a CG model used by Tsanai, M. *et al.* to study coacervate formation 20. This method is useful in studying large systems, however, many structural details are lost due to the low resolution.

1.3 **Objectives and hypotheses**

Here, atomistic modelling was used on two small-scale polyelectrolyte systems made of copolymers and homopolymers, to construct a high-resolution model of the initial stages of coacervate formation, examine the effect of oppositely charged polyelectrolytes on the surrounding water molecules, its dependency on salt concentration, and explore the interface of coacervate LLPS.

We hypothesize that sparsely charged, glycine-rich copolymer strands would coalesce into an intrinsically disordered globular structure, while the highly charged lysine and glutamic acid homopolymers would form a helical, coiled coil (CC) structure 21. The total number density of both structures was expected to be higher than the number density of pure water, and the presence of an isosurface was expected to be detected and visualized, marking the interface of LLPS.
2 METHODS

Atomistic MD simulation were performed at 310K under neutral and 0.15M sodium chloride (NaCl) ionized conditions, using PyMOL, VMD and NAMD. The details of the simulations are described in the following sections.

2.1 Constructing the initial structures using PyMOL

Molecular visualization software PyMOL (Version 2.5.4)\textsuperscript{22} was used to construct the initial structures for the simulation. Since it generates high-resolution, ray-traced images for easy visualization, it was also used to view the final structure of the protein complex.

Glycine, lysine, and glutamic acid were chosen as polyelectrolytes in this project because they all existed in the prebiotic environment, either exogenous (from outer space) or endogenous (synthesized on earth) in origin\textsuperscript{23}. Moreover, lysine and glutamic acids are known to form coacervates from experimental studies\textsuperscript{16,19}.

2.1.1 Composition of copolymers and homopolymers

Four copolymer strands each 50 amino acids in length was constructed, which were mainly composed of glycine, with two strands of lysine and two strands of glutamic acids substitutions at residues 1, 15, 30 and 45, placed in a square array configuration 20 Å apart (Figure 2a).

Similar to copolymers, four strands 50 Å in length composed of two lysine homopolymer (Lys\textsubscript{50}) and two glutamic acid homopolymer (Glu\textsubscript{50}) were constructed and placed in the same configuration (Figure 2b&2c). In both cases, polyelectrolytes of the same charge were placed diagonally from each other, with homogenous N and C terminal orientations.
Figure 2. Initial polyelectrolyte conformations input into NAMD. Glycine, lysine, and glutamic acid are represented in grey, yellow, and cyan, respectively. a) Copolymer. b) Homopolymer. c) Aerial view of the C-terminus, with side chains shown as sticks. Dark blue: nitrogen atom; red: oxygen atom.

2.2 Solvate and Ionize systems using VMD

VMD (Version 1.9.4a57-x86_64-Rev12)\textsuperscript{24} is a molecular modelling program commonly used to visualize large biomolecular systems, prepare files for molecular dynamics simulations, and analyze the output data. As a result, it was used in conjunction with Nanoscale Molecular Modelling (NAMD). In this study, periodic boundary conditions (PBCs) were used in all simulations to approximate bulk solutions, all-atom additive force field CHARMM36 was used to examine protein interactions, and the structures were solvated in TIP3P (transferable intermolecular potential with 3 points) water model to reproduce certain experimental properties. All systems were simulated under unionized conditions, where the effect of ions was also explored in copolymers using 0.15M NaCl (Figure 4).
2.2.1 Solvation with polyelectrolytes

Figure 4. Unionized and ionized copolymer initial input structures solvated using VMD, viewing from two different angles. a) Solvated with no salt ions. b) Solvated with 0.15M of NaCl, where Na$^+$ and Cl$^-$ are shown in purple and green, respectively.

2.2.2 Solvation with pure water

A cubic simulation box made of purely TIP3P water 100Å in length was constructed as a control group using the same CHARMM36 force field, and simulated under identical conditions as the system of interest to obtain the expected pure water number density.

2.3 Calculating atomic trajectories using NAMD

NAMD Version 2.14$^{25}$ was used for atomistic molecular dynamics simulations in this study. The simulations were performed on the Narval cluster through the Advanced Research Computing platform, provided by the Digital Research Alliance of Canada.

All simulations were first equilibrated for 1ns (nanosecond) using isothermal–isobaric ensemble (NPT), then up to 70ns of production runs were performed using canonical (NVT)
ensemble. Both equilibration and production runs were thermalized with Langevin dynamics, which introduced random damping forces to reassign velocities.

The output trajectory (.dcd) files were visualized in VMD and PyMOL. Molecular dynamics (MD) integrates Newton’s 2\textsuperscript{nd} law for each atomic site in the system, expressed as

$$\vec{F} = -\frac{\partial U}{\partial \vec{r}} = m \frac{d^2\vec{r}}{dt^2}$$

where $\vec{F}$ is the force, $\vec{r}$ is position, $t$ is time, $m$ is mass, $U(r)$ is the potential energy of interaction between atomic sites. The velocity Verlet algorithm\textsuperscript{26} is then used to calculate trajectories of each atom at the same timestep

$$x(t + dt) = x(t) + v(t)\ dt + \frac{1}{2}a(t)\ dt^2$$

$$v(t + dt) = v(t) + \frac{a(t) + a(t + dt)}{2}\ dt$$

Where $x$ is the position of the atom, $v$ is the velocity, $a$ is the acceleration, $t$ is the current time point, and $dt$ is the timestep. This equation may have additional coupling terms between the system and a heat bath, or a pressure control. In these simulations, the extended Lagrangian method was used for temperature control\textsuperscript{27,28}. The timestep for calculations was 2 femtoseconds (fs), the resulting atom pressure and energy were saved every 2 picoseconds (ps), and the atomic coordinates were saved every 0.2 ps.

2.4 Trajectory Analysis using MATLAB

The trajectory files obtained from NAMD was analyzed in MATLAB, and the number density of the systems were examined using a spherical and a cylindrical model constructed in Maple\textsuperscript{TM} 29.
2.4.1 Spherical partitioning for the number density analyses of copolymers

For Gly-Lys and Gly-Glu copolymers, the spherical model was fitted over the protein centre of volume (COV) using the original dcd. file output in Cartesian coordinates. The dimensions of each model was determined by size and geometry of the proteins after reaching thermal equilibrium. Under unionized condition, a spherical model with maximum radius of 35 Å was divided into 70 shells, with constant radial partitioning \( \delta r \). Under 0.15M ionized condition, a maximum radius of 30Å was divided into 100 shells with constant \( \delta r \).

2.4.2 Cylindrical partitioning for the number density analyses of homopolymers

To analyze the number density of Lys\(_{50}\) and Glu\(_{50}\) homopolymers using the cylindrical model, the Cartesian coordinates of all atoms must be first converted into polar coordinates

\[
\begin{align*}
 r &= \sqrt{x^2 + y^2} \\
 \theta' &= \arctan \left( \frac{y}{x} \right)
\end{align*}
\]
where $r$ is the radius of the cylinder, $x$ and $y$ are the Cartesian coordinates, $\theta'$ is the angle in radian around the circle.

Note that solving for $\theta'$ only returns the resultant angle from either quadrant I or quadrant IV. To include all four quadrants in the plot, the “atan2” command was used in MATLAB to obtain the final $\theta$. Protein COV was defined as (0,0,0).

Since the protein orientations change with every frame of the trajectory, a rotational matrix was necessary to align the length (height) of the protein to the Z axis

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

$$Y = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}$$

$$Z = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and $X,Y,Z$ were multiplied together to create the new coordinates that aligns the length of the protein to the Z axis of cylindrical model.

The maximum radius was set to be 40 Å, divided into 25 rings; polar angle partitioning ($\theta$) was set to be $2\pi / 75$, and maximum Z partitioning of 40Å, divided into 20 layers. Note that the radial, theta, and Z partitioning in the cylindrical figure shown above are arbitrary.

### 2.5 Statistical Analysis

Two-tailed t-Tests assuming unequal variance were performed to assess the significance of any changes in number density within the system, where a $p$-value less than or equal to 0.05 was deemed to be statistically significant.
3 RESULTS

3.1 Analyses of the water structure in the presence of the copolymer aggregate

3.1.1 Unionized solution

After reaching thermal equilibrium, the copolymer coalesced and folded into an intrinsically disordered protein with a roughly globular geometry, after simulated for 18ns under unionized condition. The distance between protein COV and the edge of the PBCs was measured to determine the maximum radius of interest.

![Image of protein structure and radius measurement](image.png)

**c)**

*Radius from protein COV vs. number density*

[Graph showing radius vs. number density for different conditions]
Figure 6. Number density analysis of the unionized copolymer. a) Gly-Lys and Gly-Glu copolymer folded into an intrinsically disordered, roughly globular structure. b) The spherical model was fitted onto the protein to analyze its density distribution, with the entire protein coloured red for easier visualization. c) Number density within each shell of the spherical model was plotted against the radius from protein COV.

A two-tailed t-Test was then performed to assess the statistical significance of the difference between total number density and expected pure water density, and a $p$-value of 0.69 was obtained.

### 3.1.2 Ionized solution with 0.15M NaCl

0.15M of salt was then added to the system, and the same spherical model with a maximum radius of 30 Å divided into 150 shells was fitted onto the protein, simulated for 16ns. One of the Gly-Glu strands dissociated from the main structure, which contained two Gly-Lys and one Gly-Glu after separation. The Gly-Glu strand occasionally interacted with the main structure but would quickly dissociate again (Figure 6).

![Figure 7](image)

Figure 7. Number density analysis of copolymer ionized with 0.15M NaCl. a) One strand of Gly-Glu copolymer dissociates from the three-stranded main structure. b) The distance from Gly-Glu COV was measured to be approximately 52.6 Å away from the closest edge of the main structure,
and conversely, the distance between COV of the main structure to the edge of Gly-Glu was approximately 67.2 Å.

Due to the large distance between the two components and enough water padding around both, the two structures were analyzed separately. First, we removed the Gly-Glu strand and analyzed the main structure.

**Figure 8.** Analysis of the main structure number density, where “Edge of Protein” was defined as radius with the first continuous zero number density. a) Total number density of protein and water plotted against radius from protein COV. b) Number density of NaCl plotted against radius from protein COV.

A two-tailed t-Test was performed to assess the statistical significance of the difference between total number density and expected pure water density, and a *p*-value of 0.84 was obtained.

The single Gly-Glu strand was then analyzed using the same method:
Figure 9. Analysis of the Gly-Glu strand number density. **a)** Total number density of protein and water plotted against radius from protein COV. **b)** Number density of NaCl plotted against radius from protein COV.

A two-tailed t-Test was performed to assess the statistical significance of the difference between total number density and expected pure water density, and a $p$-value of 0.82 was obtained.

### 3.2 Analyses of the water structure in the presence of the homopolymer aggregate

Lys$_{50}$ and Glu$_{50}$ homopolymers formed a stable helical quadruplex structure. The cylindrical model was fitted over the strands based on their COV and polar coordinates, and the water number density was normalized by the local maximum (excluding water inside the helix), and visualized by a colour gradient from black (minimum) to blue (maximum).
Figure 10. a) Lys$_{50}$ and Glu$_{50}$ homopolymers coalesce and formed a helical quadruplex structure. b) Cylindrical model used for homopolymer density analysis. c) Fitted water density map around the protein.

Recall that protein COV was defined as (0,0,0) in polar coordinates. Since the height of the homopolymer helix was measured to be 80.3 Å, the maximum X,Y,Z value were all set to 40 Å extend to both side of the origin, and the density plot was shown in a 50 Å$^3$ 3D space (Figure 8c).

It is important to note that X,Y,Z should be considered as the absolute radius from COV in units of Å, as mathematically radius cannot be negative values.

The average of all voxels along the Z axis was then calculated, and “compressed” down into a 2-dimensional density map. The number density of both protein, water, and their sum in each radial partition ring was plotted into two histograms.
Figure 11. Number density analyses of each ring in the 2D cylindrical plot. a) Average of normalized number density along Z axis, plotted around protein COV. b) Histogram of ring number from protein COV, plotted against number density of both water and protein. Blue and yellow represents water and protein, respectively. c) Histogram of ring number from protein COV, plotted against total number density of both water and protein.
Two-tailed t-Tests was then performed to assess the significance of the changes in number density. Both the water and total number density changes were significant, where \( p < 0.001 \) (\( p = 1.0748 \times 10^{-14} \) and \( p = 9.7036 \times 10^{-29} \), respectively).

4 DISCUSSION

In this study, we investigated the effect of oppositely charged polyelectrolytes have on the water and total number density within and around the structures, under zero and physiological salinity.

4.1 Copolymer aggregate

In unionized environment, the four strands of copolymer coalesced into an approximately globular intrinsically disordered protein. This was expected, as glycine, lysine and glutamic acid are known to be disorder-promoting amino acids. The abundance of non-polar, uncharged glycine residues and the resultant low sequence complexity also contributed to its final confirmation. This result agreed with our hypothesis. The number density of protein and water within each shells of the spherical model was analyzed (Figure 6), and the result showed that beyond the protein periphery, the total number density within the PBC converged with the expected density of pure water. Within the structure of copolymer, the total number density was observed to fluctuate more and appeared to be slightly higher than the expected value of pure water. However, after performing the two-tailed t-Test, this fluctuation and apparent increase was deemed to be non-statistically significant.

The system was then ionized with 0.15M of NaCl, where one of the Gly-Glu strand was observed to dissociate from the rest of the structure. This is likely due to the screening effects of \( \text{Na}^+ \) and \( \text{Cl}^- \) ions, where they neutralized the sparsely charged systems, making the Gly-Glu strand
difficult to maintain interaction with the main structure. It is likely that this is a short-living, local minimum energy conformation, and the Gly-Glu strand will eventually merge into the main structure.

In the meantime, the two constituents were analyzed separately (Figure 8 & Figure 9).

Figure 8a showed the number density of the main structure within each of the 100 shells within the spherical model. Similar to unionized condition, the total number density converges with the expected pure water density beyond the periphery of the protein, and the fluctuation and apparent increase was deemed to be non-statistically significant. Figure 8b showed the NaCl number density within each shell. The first major peak of Na$^+$ appears within the copolymer structure, at 13.2Å from protein COV. Referring back to Figure 8a, this was the exact radius where the number density of protein and water converged. The second major peak of Na$^+$ appeared where the protein density first approached zero, at 21.9Å.

Similarly, Figure 9a showed the number density of the Gly-Glu strand within each shell of the spherical model. This single strand is much smaller than the 3-stranded main structure, thus its effect on number density was elicit on a smaller scale, closer to the protein COV. The total number density converges with the expected pure water density beyond the periphery of the protein, and the fluctuation was non-statistically significant. Figure 9b showed the NaCl number density in each shell, where the first major peak of Na$^+$ also appeared within the copolymer structure, at 6.0Å. According to Figure 9a, this was also the approximate radius where the number density of protein and water converged. The second major peak of Na$^+$ appeared immediately after the protein periphery, at 19.2Å.

The results showed the location of two critical events that affect salt distribution: first, when the protein density and water density within the copolymer structure converged, and second,
at the periphery of the protein. In figure 8, there were some minute fluctuations in the order of $10^{-4}$ between 21Å and 26.4Å (“Edge of Protein”), likely due to lose ends of the intrinsically disordered protein present within the shell. In this region, some shells showed zero number density, and the subsequent one would have positive non-zero values. This observation was puzzling at first, but after some rationalization, since the main structure was made of three independent strands of copolymer, some gaps might exist between them, and the spherical shells may reside in this gap, resulting in the non-continuous zero number densities.

In Figure 9, the Edge of Protein was shown to be flanked by both a Cl$^-$ peak and an Na$^+$ peak of similar magnitude on either side, but its significance and mechanism need to be further investigated.

It is also worth noting that in Figure 8b, the 3-stranded main structure was expected to have an overall positive charge after the dissociation of the weak, negatively charged Gly-Glu strand. But the results suggest that both structures had a higher affinity to Na$^+$ than Cl$^-$. Since Na$^+$ ions have a smaller size than Cl$^-$ ions, it is speculated that Na$^+$ ions are able to diffuse in and out of the copolymer relatively freely, while the larger Cl$^-$ are more likely to be found around and beyond the periphery of the polyelectrolytes.

Figure 8b and 9b also showed alternating shells of number density distribution of Na$^+$ and Cl$^-$ within the radius of interest, and a near-perfect exponential decay of the number density was observed. The mechanism behind such effects of copolymer on the number density of ions at these two critical sites remains unclear based on a single “snapshot” of its conformation. The results are expected to be much more accurate and reliable once the average of more than 5000 conformations we obtained through NAMD can be analyzed.
4.2 Homopolymer aggregate

After reaching thermal equilibrium, the homopolymer structure composed of purely lysine and glutamic acid formed a helical, or “coiled coil (CC)” structure as shown in Figure 10a, which is consistent with our hypothesis based on previous literature and verified the accuracy and reliability of our NAMD simulations. The number density within and around the CC structure was then analyzed using the cylindrical model and plotted in a 3D space (Figure 10c), where each dot represents the number density within each voxel of the cylindrical model, as shown in Figure 10b. The colour of homopolymer was kept constant, while the colour of water was normalized using the local maximum number density value. It is worth noting that the local maximum number density was use instead of the global maximum because in the hollow centre of CC structure, the water density was remarkably high, and if the colour map was normalized using these values, all other voxels will be displayed as black due to the relative low values of number density. After omitting the water density within the helix, the change in water density can be visualized much more clearly.

For direct analysis, the 3D cylindrical density map was reduced to a 2D plot by taking the average number density along the Z axis (Figure 11a). The number density and distribution within each of the 25 rings was then examined and plotted as a bar graph (Figure 11b). As the number density of protein increased, a statistically significant decrease in water number density was detected \((p<0.001)\). This result was expected, as protein structure occupies certain volume, displaces water and lowering its number density. Beyond the periphery of the protein, the water density returns to the expected values, which corresponds to the fluctuating number density distribution beyond ring 13 of the 2D plot shown in Figure 11a.
However, when the sum of the total number density of both protein and water within each ring was calculated, there seemed to be an increase in the total number density, and both the one-tailed and two-tailed t-Tests proved that this change was also statistically significant ($p < 0.001$).

This increase in both water and total number density within the CC structure agreed with our hypotheses, and showed the effect of charged homopolymer had on the system.

## 5 CONCLUSIONS

Coacervate droplets are plausible models of protocells; however, very few studies have been conducted to understand LLPS on an atomistic level. Here we constructed a high-resolution atomistic model to examine the initial stages of interactions between water and oppositely charged homopolymers and copolymers leading to complex coacervation.

The results showed that sparsely charged copolymers had no significant effect on the number density in the bulk solution; in contrast, statistically significant changes were detected in both the water density and total number density in the presence of highly charged homopolymers. Thus we can conclude that the primary structures of oppositely charged polyelectrolytes directly affects their final configuration and number density distribution within and around the structures.

One of the major limitations of atomistic modelling is that it is very computationally expensive, thus only the initial stages of coacervation can be simulated with such high resolution. In addition, the biophysics of their shape fluctuations, coalescence, and fragmentation are also very difficult to be describe mathematically. In this study, we presented three modelling results based on snapshots of their conformational changes, thus longer simulations are needed to confirm that the system has reached its lowest energy conformation, and an average of all conformations should be taken to increase the accuracy and reliability of our results.
For future studies, a larger system with more polyelectrolytes needs to be simulated under more precise prebiotic conditions, such as pH level, salinity, temperature, and pressure. Nucleic acids should be included into the polyelectrolyte mixture to study their interactions with proteins and effects on coacervation. Moreover, the effect of oppositely charged polyelectrolytes on the number density distributions of Na\(^+\) and Cl\(^-\) ions, and their interactions with both protein and water are of great interest for future studies, as such interactions are not visible using CG modelling. It is also important to relate the results of atomistic modelling to coarse grain modelling, as their strength and weaknesses well complement each other.

Although the scale of our model is too small to show a spherical density isosurface as observed in experimental studies, the significant changes in total water density indicated the initial stages of LLPS and complex coacervate droplet formation, allowing us to glimpse into the early steps of compartmentalization leading to the origins of life.
6 REFERENCES

1 Oparin, A. & Morgulis, S. *The Origin of Life*; (Macmillan, 1938).


9 Haldane, John B. S. "The Origin of Life". (1929).


