# Protein-Surface and Protein-Ligand Interactions: Insights from Atomistic Simulations

by

**Shivam Tiwari**

*A thesis submitted*

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Supervisor

**Dr. K. Anki Reddy**

Department of Chemical Engineering

Indian Institute of Technology Guwahati

Guwahati 781039, India

March, 2024

# Protein-Surface and Protein-Ligand Interactions: Insights from Atomistic Simulations

by

**Shivam Tiwari**

## Abstract

Proteins are one of the most abundant and essential class of organic molecules present in the living systems. Their role in the sustenance and growth of life is unique. However, proteins can turn toxic and unfavorable if they are displaced from their native form or present in an undesirable environment.

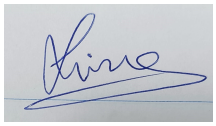
In this thesis, we perform a computational investigation of protein's interaction with various surfaces and chemical entities in varying environments. Various roles of protein are explored in this thesis, ranging from, as a foulant in a desalination membrane, as a cell invader to toxic $\beta$ plaques causing cell damage. Specifically, in the first part (first two chapters) of the thesis, the protein plays a role of a membrane foulant. We investigated the reason for differences found in fouling of reverse osmosis (RO and a forward osmosis (FO) membrane with lysozyme protein. We explained the role of hydration repulsion and electrostatic interactions in the differences found in FO and RO fouling.

Furthermore, we investigated the effect of ionic concentration on membrane fouling with graphene oxide and polyamide as model membrane and bovine serum albumin protein as a model foulant. The study revealed the role of functional groups present on the membrane's surface in deciding the membrane's response to the change in ionic concentration and hence the interaction with protein. Next, we studied the interaction of SARS-CoV2's spike protein (which plays a crucial role in cell entry of the CoV2 virus) with the montmorillonite surface, a clay mineral. The emergence of corona virus has sparked an interest in the search of nanomaterials which can interact strongly with the spike protein and induce structural remodeling

and subsequently defunctionalize it. We report substantial damage to the secondary structure of spike protein in the presence of MMT surface. Also, residues and glycans of spike protein which have been found to play key role in the cell entry were found to be interacting strongly with the MMT surface. In the last two chapters, protein is in the form of toxic $\beta$ sheets, responsible for a range of age-related dementia. Mechanism of binding and disaggregation of A$\beta$ fibrils with a novel peptidomimetic compound was investigated. The compound showed a strong binding on A$\beta$ fibril via hydrophobic interactions and significant destruction of the fibril's $\beta$ sheet content. Finally, the interaction between A$\beta$ fibrils and lipid bilayer was studied using a prodrug peptide $\beta$-Aspartyl. The study showed that $\beta$-Aspartyl inhibits the interaction between A$\beta$ fibrils and lipid bilayer. The study also showed the role of $\pi - \pi$ stacking and cation-$\pi$ interactions in binding $\beta$-Aspartyl on fibrils.

# Declaration

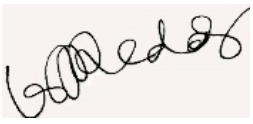
I hereby certify that the work compiled in this thesis is the outcome of the research work, performed by myself, else stated, under the guidance of Dr K. Anki Reddy. Any part of this work has not been submitted for the award of any degree, diploma, associate-fellowship, fellowship or it's equivalent to any university or institution.



Shivam Tiwari,
Registration No: 156107039
Department of Chemical Engineering
IIT Guwahati,
Guwahati-781039, Assam, India.
DATE: 12/06/2024

# CERTIFICATE

It is certified that the work contained in the thesis entitled "Protein-Surface and Protein-Ligand Interactions: Insights from Atomistic Simulations" by Mr Shivam Tiwari, a student of the Department of Chemical Engineering, IIT Guwahati was carried out under my supervision and has not been submitted elsewhere for the award of any degree.



Dr K. Anki Reddy,
Registration No: 156107039
Department of Chemical Engineering
IIT Guwahati,
Guwahati-781039, Assam, India.
DATE: 12/06/2024

# Acknowledgements

I am indebted to my Ph.D. supervisor Dr. K Anki Reddy, for his valuable suggestions, support, and mentorship. I would also like to acknowledge my doctoral committee members: Prof. Ashok Kumar Dasmahapatra, Dr. Amit Kumar, and Dr. Priyadarshi Satpati. Their valuable suggestions and critical comments shaped the presented thesis work. I would also like to thank Prof. Bhubaneshwar Mandal from the Department of Chemistry for sharing his knowledge and time. Besides, I would also like to thank all the Chemical Engineering office, Academic office, Student Affair, and Finance staff members. They were always available whenever I required.

I would also like to thank IITG and the Government of India for providing HPC facility (PARAM-ISHAN) for conducting the simulations, which would have taken a long time otherwise. I want to extend my sincere gratitude to the Ministry of Human Resource Development for providing financial assistance, making the Ph.D. journey much more manageable.

I want to extend my gratitude to all my fellow friends: Dr. Mahendra Baingne, Mr. Bitang Kwrung Tripura, Dr. Avadh Kishore Kumar, Dr. Sunil Kumar, Dr. Piyal Mondal, Mr. Abhishek Kumar, Dr. Vamsi Krishna Reddy, Mr. Kranthi Kumar, Mr. Raja Sekhar Reddy, Mr. Nagendra Prasad, Mr. Adupa Vasistha, for their contributions towards my Ph.D. work directly or indirectly.

Lastly, I would like to thank my family for the unconditional love and support they showered on me. I like to thank my parents, Mr. Shailendra Tiwari and Mrs. Archana Tiwari, my grandparents, Mr. Ramesh Chandra Sharma and Late Mrs. Kamla Sharma, my sister Shivani Tiwari, my wife Nupur Tiwari, and my daughter Darshini Tiwari.

# Contents

# Chapter 1

# Introduction

## 1.1 Overview

Protein derives its root from 'proteios,' meaning "primary." Indeed, proteins are of primary importance in the existence of life on the planet. Miller, through his pioneering work[1] suggested that life could have originated from amino acids, which are the basic units of protein. He demonstrated his hypothesis by synthesizing many amino acids in primitive earth conditions. Hence, it would not be an exaggeration if the protein were called synonyms to life.

Proteins are linear chains of amino acids bonded via amide bonds. There are 20 natural amino acids, and all proteins are made out of these 20 amino acids. All amino acids are structurally identical, i.e., all have a central carbon atom, which is known as alpha carbon ($C_\alpha$). This alpha carbon is bonded to an amino ($NH_2$) group, a carboxyl (COOH) group, a hydrogen atom, and a side chain. All parts except the side chain are identical in all amino acids; hence, the side chain differentiates one amino acid from the other. Also, it is the side chain that defines the chemical characteristic of the amino acid. Figure 1.1 shows the chemical structure of all the basic amino acids and the classification based on the chemical nature. Based on the chemical nature of the side chain, its position in the structure of protein is decided; consequently, the interactions between various amino acids with differing chemical qualities impart the protein its peculiar three-dimensional (3d) structure. The protein's structure can be classified mainly into primary, secondary, and tertiary.

The primary structure refers to the linear amino acid sequence bonded by the peptide bond between the amino group of one amino acid to the carboxyl group of another amino acid.

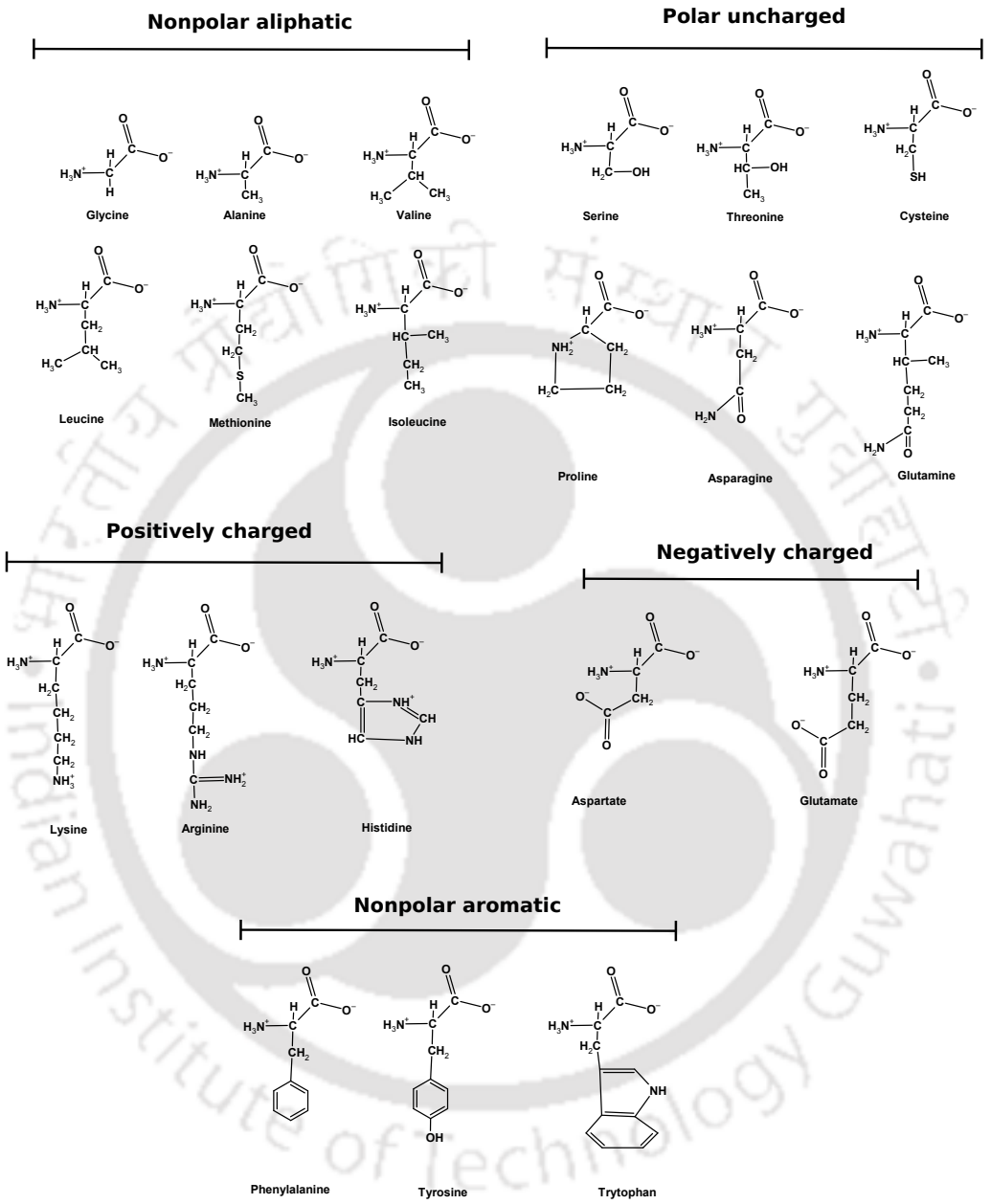



Figure 1.1: Chemical structures of basic amino acids and their classification based on the chemical nature exhibited by their side chains.

The secondary structure is highly ordered local stuctures, stabilized by hydrogen bonding interaction between amino acids. There are broadly two kinds of secondary structures: $\alpha$-helix and $\beta$-sheets; these were proposed first time by Linus Pauling et

al. [2]. Figure 5.5 shows 3d structure of bacteriorhodopsin[3](PDB id: 1FBB). The tertiary structure is the 3d structure that protein gets finally folded into. The interaction between the hydrophobic residues and the water around the protein drives the local secondary structure to get folded up into a 3d globular structure(Figure 5.5).

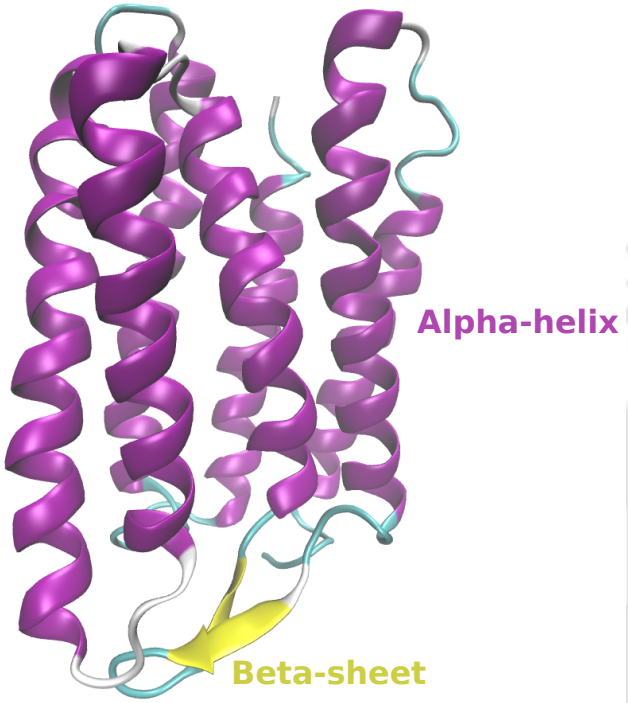



Figure 1.2: 3d structure of bacteriorhodopsin obtained from RCSB protein data bank with pdb id: 1FBB. The purple and yellow color represents alpha-helix and beta-sheet of the protein respectively. The white and cyan color represents coil and turn respectively.

## 1.2 Protein "A Friend turned Foe"

Although proteins are essential biomolecules for sustenance and growth for all living beings, including humans, and hence is our best friend. However, in certain conditions and environments, this friend of ours can also turn against us and become our enemy. Specifically, if proteins are in an environment where they are not desirable

and lose their native form (structure), they can cause harm. Hence, the underlying theme of this thesis is to understand "how proteins in certain situations become undesirable and toxic" at a molecular level via computer simulations.

## 1.3 Protein as a Foulant

Simply put, a membrane is a barrier, which selectively allows some chemical species to pass through it, but stops everything else. Applications of membrane technology are manifold. However, one of the most important uses of membrane technology is in water purification. Reverse osmosis (RO) is one the most widely used membrane water purification technology (MBWT), in which hydraulic pressure is used to push the feed water across a semipermeable membrane. However, forward osmosis (FO) has received increased attention in the past decade as an emerging MBWT[4, 5], mainly because of its excellent energy efficiency. In FO, a draw solution with high solute concentration creates a concentration gradient along a semipermeable membrane and the feed side, which is at the other side of the membrane. However, in both FO and RO and in any MBWT, the heart of the operation is the membrane itself.

However, MBWTs are subject to performance-limiting severe phenomena "membrane fouling." Membrane fouling is undesirable deposition and agglomeration of chemical species and other impurities contained in the feed water on the membrane's surface. The type of fouling layer formed on the membrane's surface can be of variety, e.g., cake formation, pore blocking, organic adsorption, and biofouling. Fouling is one of the major caveats to the membrane performance and can lead to flux decline, increased energy consumption, and even irreversible damage to the membrane[6, 7]. Foulants can be classified mainly into three categories based on their intrinsic chemical characteristic: inorganic, organic, and biofoulants. Inorganic foulants mainly consist of silica and silicate minerals which cause scale formation on the membrane's surface. Organic foulants are one the most ubiquitous category of foulants found in almost all kinds of feed water. They mainly comprise of oils, proteins, other biomacromolecules, and natural organic matter; they get adsorbed on the membrane's surface via hydrophobic interactions, sometimes even irreversibly.

Biofoulants mainly include microorganisms.

Protein falls under the category of organic foulants. Proteins are one of the most important foulants, in a sense that due to their varied range of characterstics imparted by different kinds of amino acid residues they are made of, they are capable of interacting with many surfaces via interactions such as hydrophobic, hydrogen bonding, electrostatic. This makes protein fouling extremely challenging to overcome, and hence protein fouling can act as a stringent test for evaluating antifouling properties of any potential membrane material. Owing to the complex nature of protein fouling and its importance in developing fouling resistant membranes, it has attracted many studies[8–16]. Ang et al.[8] investigated fouling of RO membranes with bovine serum albumin (BSA), a model protein and alginate molecule. They found that fouling is higher when both BSA and alginate are present in the feed than present individually. Also, they observed an increase in BSA fouling with an increase in the ionic concentration of the feed solution. The difference between the nature of fouling and fouling layer structure in FO and RO is an important and long debated topic in membrane research, Lee et al.[10] compared fouling in FO and RO with BSA, alginate, and humic acid as model foulants. They found that flux decline in FO is higher due to the increased cake enhanced osmotic pressure which further drops net osmotic pressure, but they found the fouling layer in FO to be much more easily reversible than RO. Effect of ionic concentration on membrane fouling, is a crucial aspect of the fouling mechanism since membrane can be exposed to a variety of range of salinity depending on the source of the feed water. Miao et al.[13] carried out experiments with BSA as a foulant and polyvinylidene fluoride (PVDF) as a membrane over a range of ionic strengths. They observed an increase in fouling with the increase in ion concentration within a range of 0 to 1mM. However, when the ion concentration is raised to the range of 10-100 mM, a significant decrease in fouling is observed. This was attributed to the increased hydration repulsion between BSA-BSA and BSA-PVDF. Jahan et al.[16] performed atomistic simulation with lysozyme protein and cross-linked polyamide (PA) membrane to study the biofouling in an RO system. They found the surface roughness to play a key role in the membrane's fouling.

## 1.4 Protein as a cause of disease

The phrase 'Form is Function' perfectly defines the mechanism of protein's functioning since its function defines its 3d structure or vice versa. However, deviation from its native structure can cause dysfunction of the process and lead to disorders. There are many diseases attributed to misfolded proteins; the most common ones are dementia diseases caused by the self-assembly of proteins such as Alzheimer's (AD), Parkinson's (PD), type II diabetes (T2D), amyotrophic lateral sclerosis (ALS)[17]. The main characteristic of the diseases mentioned above is fibrillar plaques deposition, made of mainly self-assemblies of misfolded proteins. The key proteins involved in the disease mentioned above are already identified, such as A$\beta$, and tau proteins are considered responsible for AD[18, 19], $\alpha$-synuclein for PD[20] and the superoxide dismutase (SOD1), TAR DNA binding protein 43 (TPD-43) for ALS. Most of these diseases still do not have a proper treatment, largely because a clear understanding of their occurrence is still lacking; hence, it attracted many studies to attempt to understand the underlying mechanism, employing different approaches. Miller et al.[21] using MD simulations, probed the effect of $Zn^{2+}$ on A$\beta$ aggregation. They found the underlying mechanism behind the enhancement of A$\beta$ aggregation rate due to $Zn^{2+}$ binding. Young et al.[22] studied the formation of oligomers of IAPP using ion mobility spectrometry and mass spectrometry. They also tested different polyphenolic compounds as an aggregation inhibitor and were effective but with a different mode of action. Mompeán et al.[23] used a range of experimental and simulation tools to decipher the mechanism behind the TDP-43 aggregates. They observed a novel morphology of TDP-43 aggregates having a cross $\beta$ spine and a tight side packing. Influence of Cu(II) binding on A$\beta_{42}$ dimer and monomer's dynamics was studied by Huy et al.[24]. They observed that upon Cu(II) binding, the aggregation tendency of A$\beta$ decreases, attributed to the increase in flexibility of the salt bridge Asp23-Lys28.

## 1.5 Objectives

Based on the literature survey we proposed following objectives:

- To deciher the underlying reason for differences in fouling in FO and RO.
- To investigate the effect of ionic environment on membrane fouling at molecular level.
- To study the interactions between SARS-CoV2's spike glycoprotein and montmorillonite.
- To study the disaggregation mechanism of A$\beta$ fibrils with a novel peptidomimetic compound.
- To study the interactions between A$\beta$ fibrils and lipid bilayer in the presence of $\beta$-Aspartyl.

## 1.6 Thesis Organization

Chapter 2 briefly discusses the simulation technique (molecular dynamics) used in this thesis and its basic formalism.

Chapter 3 attempts to address the long-standing question of "Difference in fouling in FO and RO." Graphene oxide is used as a model membrane for both FO and RO systems, and lysozyme protein is used as a model foulant.

Chapter 4 deals with yet another important and much-debated question in membrane research: "Effect of ionic environment on membrane fouling. Graphene oxide and polyamide membranes are used as model membranes, while bovine serum albumin protein is used as a model foulant.

Chapter 5 investigate the interaction between SARS-CoV2's spike protein and a clay mineral surface, "montmorillonite."

Chapter 6 examines the capability of a novel peptidomimetic compound to disaggregate the well-organized fibril assembly of A$\beta_{42}$ and also their bindin mechanism to the fibrils.

Chapter 7 tests the performance of a drug peptide "$\beta$-Aspartyl" in screening the interactions between A$\beta_{42}$ fibrils and lipid bilayer. Additionally, the binding of the ligand on fibril was analyzed.

Chapter 8 contains the conclusions and some ideas for future works.

# Bibliography

[1] Stanley L Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, 1953.

[2] Linus Pauling, Robert B Corey, and Herman R Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4):205–211, 1951.

[3] Sriram Subramaniam and Richard Henderson. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature*, 406(6796):653–657, 2000.

[4] Tzahi Y Cath, Amy E Childress, and Menachem Elimelech. Forward osmosis: principles, applications, and recent developments. *Journal of membrane science*, 281(1-2):70–87, 2006.

[5] Devin L Shaffer, Jay R Werber, Humberto Jaramillo, Shihong Lin, and Menachem Elimelech. Forward osmosis: where are we now? *Desalination*, 356: 271–284, 2015.

[6] Mark A Shannon, Paul W Bohn, Menachem Elimelech, John G Georgiadis, Benito J Marinas, and Anne M Mayes. Science and technology for water purification in the coming decades. *Nanoscience and technology: a collection of reviews from nature Journals*, pages 337–346, 2010.

[7] Pierre Le-Clech, Vicki Chen, and Tony AG Fane. Fouling in membrane bioreactors used in wastewater treatment. *Journal of membrane science*, 284(1-2): 17–53, 2006.

[8] Wui Seng Ang and Menachem Elimelech. Protein (bsa) fouling of reverse osmosis membranes: implications for wastewater reclamation. *Journal of Membrane Science*, 296(1-2):83–92, 2007.

[9] Amanda N Quay, Tiezheng Tong, Sara M Hashmi, Yu Zhou, Song Zhao, and Menachem Elimelech. Combined organic fouling and inorganic scaling in reverse osmosis: role of protein–silica interactions. *Environmental science & technology*, 52(16):9145–9153, 2018.

[10] Sangyoup Lee, Chanhee Boo, Menachem Elimelech, and Seungkwan Hong. Comparison of fouling behavior in forward osmosis (fo) and reverse osmosis

(ro). *Journal of Membrane Science*, 365(1):34–39, 2010. ISSN 0376-7388. doi: https://doi.org/10.1016/j.memsci.2010.08.036.

[11] Alberto Tiraferri, Yan Kang, Emmanuel P Giannelis, and Menachem Elimelech. Superhydrophilic thin-film composite forward osmosis membranes for organic fouling control: fouling behavior and antifouling mechanisms. *Environmental science & technology*, 46(20):11135–11144, 2012.

[12] Yi-Ning Wang and Chuyang Y Tang. Protein fouling of nanofiltration, reverse osmosis, and ultrafiltration membranes—the role of hydrodynamic conditions, solution chemistry, and membrane properties. *Journal of Membrane Science*, 376(1-2):275–282, 2011.

[13] Rui Miao, Lei Wang, Na Mi, Zhe Gao, Tingting Liu, Yongtao Lv, Xudong Wang, Xiaorong Meng, and Yongzhe Yang. Enhancement and mitigation mechanisms of protein fouling of ultrafiltration membranes under different ionic strengths. *Environmental science & technology*, 49(11):6574–6580, 2015.

[14] Abdul Rajjak Shaikh, Hamed Karkhanechi, Tomohisa Yoshioka, Hideto Matsuyama, Hiromitsu Takaba, and Da-Ming Wang. Adsorption of bovine serum albumin on poly (vinylidene fluoride) surfaces in the presence of ions: a molecular dynamics simulation. *The Journal of Physical Chemistry B*, 122(6):1919–1928, 2018.

[15] Wesley Beckner, Yi He, and Jim Pfaendtner. Chain flexibility in self-assembled monolayers affects protein adsorption and surface hydration: a molecular dynamics study. *The Journal of Physical Chemistry B*, 120(40):10423–10432, 2016.

[16] Md Symon Jahan Sajib, Ying Wei, Ankit Mishra, Lin Zhang, Ken-Ichi Nomura, Rajiv K Kalia, Priya Vashishta, Aiichiro Nakano, Sohail Murad, and Tao Wei. Atomistic simulations of biofouling and molecular transfer of a cross-linked aromatic polyamide membrane for desalination. *Langmuir*, 36(26):7658–7668, 2020.

[17] Merrill D Benson, Joel N Buxbaum, David S Eisenberg, Giampaolo Merlini, Maria JM Saraiva, Yoshiki Sekijima, Jean D Sipe, and Per Westermark. Amyloid nomenclature 2018: recommendations by the international society of amyloidosis (isa) nomenclature committee. *Amyloid*, 25(4):215–219, 2018.

[18] George G Glenner and Caine W Wong. Alzheimer's disease and down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochemical and biophysical research communications*, 122(3):1131–1135, 1984.

[19] John Hardy and Dennis J Selkoe. The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics. *science*, 297(5580):353–356, 2002.

[20] Matthew S Goldberg and Peter T Lansbury Jr. Is there a cause-and-effect relationship between $\alpha$-synuclein fibrillization and parkinson's disease? *Nature cell biology*, 2(7):E115–E119, 2000.

[21] Yifat Miller, Buyong Ma, and Ruth Nussinov. Zinc ions promote alzheimer a$\beta$ aggregation via population shift of polymorphic states. *Proceedings of the National Academy of Sciences*, 107(21):9490–9495, 2010.

[22] Lydia M Young, Ping Cao, Daniel P Raleigh, Alison E Ashcroft, and Sheena E Radford. Ion mobility spectrometry–mass spectrometry defines the oligomeric intermediates in amylin amyloid formation and the mode of action of inhibitors. *Journal of the American Chemical Society*, 136(2):660–670, 2014.

[23] Miguel Mompeán, Rubén Hervás, Yunyao Xu, Timothy H Tran, Corrado Guarnaccia, Emanuele Buratti, Francisco Baralle, Liang Tong, Mariano Carrión-Vázquez, Ann E McDermott, et al. Structural evidence of amyloid fibril formation in the putative aggregation domain of tdp-43. *The journal of physical chemistry letters*, 6(13):2608–2615, 2015.

[24] Pham Dinh Quoc Huy, Quan Van Vuong, Giovanni La Penna, Peter Faller, and Mai Suan Li. Impact of cu (ii) binding on structures and dynamics of a$\beta$42 monomer and dimer: Molecular dynamics study. *ACS chemical neuroscience*, 7(10):1348–1363, 2016.

# Chapter 2

# Simulation Methodology

This thesis aims to find the microscopic mechanism behind various physical phenomena identified in the objectives. The detailed atomistic characterization of a system is beyond the realm of any experimental technique currently available. However, such microscopic mechanistic details for a plethora of chemical, physical and biological systems can be obtained using molecular simulations. One of the most important molecular simulation techniques is molecular dynamics(MD), which has been remarkably successful in providing atomistic insights into a wide variety of phenomena. The present thesis employs MD simulation to investigate all the systems defined in the objectives. This chapter will discuss the key theoretical and practical aspects of MD simulation.

## 2.1 Introdction to Molecular Dynamics

The fundamental underlying principle of MD simulation is simple, i.e., to solve eqn. 2.1 for all the particles in the system, which is essentially Newton's second law of motion. The eqn. 2.1 is a second-order differential equation that, on solving, gives velocities and positions of all particles at any instant of time using an appropriate force model for force ($F_i$) acting upon each particle in the system. The force model, expressed as a negative gradient of potential(eqn 5.8), is at the heart of MD simulation since the quality of results from MD is as good as the model, commonly called

a force field.

$$a_i = \frac{d^2\vec{x_i}}{dt^2} = \frac{\vec{F_i}}{m_i} \tag{2.1}$$

$$\vec{F}(\vec{r}) = -\nabla U(\vec{r}) \tag{2.2}$$

However, The eqn 2.1 becomes practically impossible to solve analytically for a many-body system having nonlinear terms in the potential expression. Hence, MD simulation employs numerical methods to solve eqn 2.1 for a many-body system. These numerical techniques are discussed in detail later in this chapter.

One of the earliest implementations of MD simulations to study a physical system was performed by Alder and Wainwright[1, 2]. They showed the existence of solid-liquid phase transition using a hard sphere model. Later, Rahman[3] and Verlet[4] independently carried out MD simulation with a Lennard-Jones(LJ) potential for a system of argon atoms. They calculated pair-correlations and other equilibrium properties for the system, which agreed well with the experimental values. Another breakthrough came from a series of studies by Berne and coworkers[5–8], where they investigated molecular relaxation behavior in diatomic liquids using MD simulations. The next major milestone came in the form of the first MD simulations of liquid water by Stillinger and Rahman[9–11] Soon after these studies, Karplus and coworkers carried out the first MD simulation of proteins[12, 13].

In the forthcoming sections, we will discuss essential components of MD calculations and the techniques used to perform them. However, due to the abundant variety of these techniques, we will discuss only those techniques implemented in the MD code NAMD[14], since we performed all the simulations in this thesis using NAMD.

## 2.2 Energy Minimization

It is essential to have a reasonable initial structure of the system under consideration to perform an MD run on it. Usually, the structures obtained from experiments and or constructed using some software, for e.g., protein's structure obtained from the protein data bank(PDB), are subject to local stresses due to nonbonded overlaps

and distorted bond lengths and angles. Hence, it is a standard practice to refine the initial positions of the atoms in the system by performing an energy minimization run using an appropriate technique.

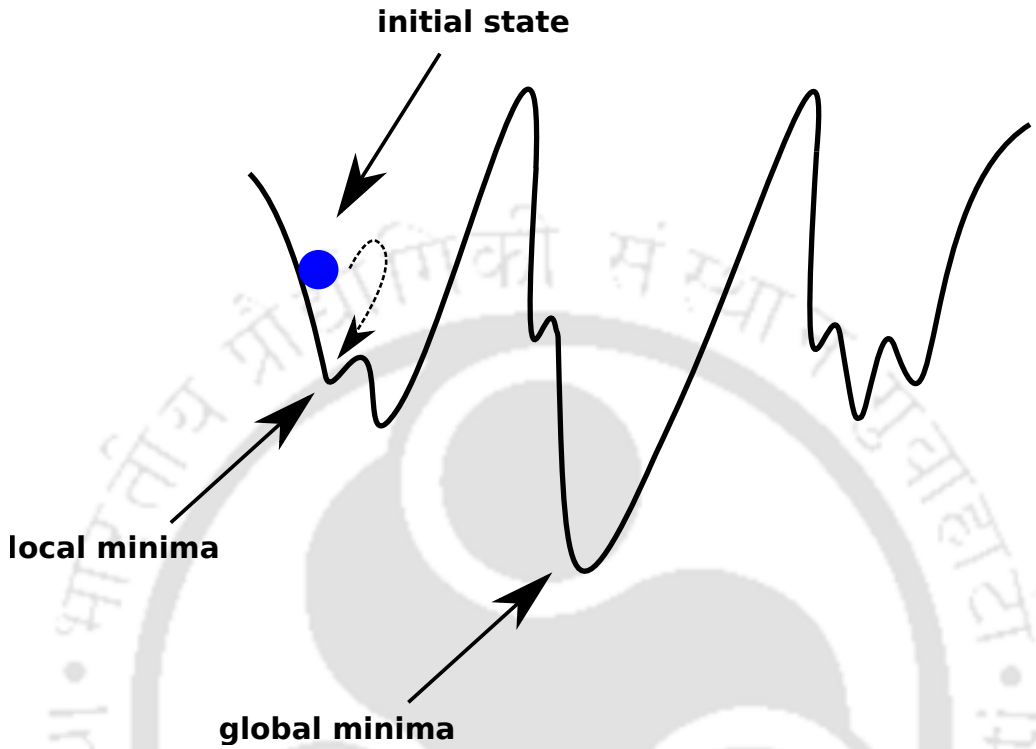



Figure 2.1: A schematic representation of a one dimensional potential energy surface.

Figure 2.1 shows a schematic representation of the potential energy surface of a system. The blue ball and its position on the surface represent the system's initial state in terms of its potential energy. Various trenches along the energy surface are the minimas. The trench nearest to the initial state is called local minima, and the deepest well is called global minima. The goal of energy minimization is to bring the system downhill to the local minima on the potential energy surface. Hence, fundamentally energy minimization is an optimization problem where a function (force-field expression) is given, and it depends upon many independent variables (cartesian coordinates of each atom). The problem to be solved is finding the set of coordinates that provide the minimum value of the potential energy function.

Minimization algorithms can be broadly classified into derivative and non-derivative methods. Consider a Taylor series expansion of the potential energy function $U(x)$

around a minimum point $X_0$.

$$U(X) = U(X_0) + (X - X_0)U'(X_0) + \frac{1}{2}(X - X_0)^2U''(X0) \tag{2.3}$$

Where $U'$ and $U''$ are the first and second derivatives of the function, the X represents a vector with 3N components, where 3N is the total number of cartesian coordinates with N number of atoms. The non-derivative algorithms use only the value of the function at any point during the minimization, and they are also called zero-order algorithms. The first and second-order algorithms employ the value of $U'$ and $U''$, respectively. The accuracy and the computational cost of these algorithms also increase in the same order; that is, the second-order algorithms are the most accurate and, computationally, the most expensive. Hence, the choice of an algorithm must be based on the level of sophistication required, the size of the system, and the amount of computational resources at one's disposal. Therefore, second-order algorithms like the Newton-Raphson method are usually used for quantum mechanical calculations, where the required accuracy level is high, and the number of atoms is less. However, most MD simulation codes employ first-order algorithms such as the steepest descent and conjugate gradient method. The zero order methods such as simplex method are rather crude and are usually used as a precursor to the more sophisticated methods.

The first-order algorithms are the most frequently used in MD simulations since they strike a fair balance between accuracy and efficiency. All the "order 1" methods use the slope of the energy curve(obtained from the first derivative) to drive the ball(blue ball in Figure 5.8) downhill towards the local minima iteratively. In general, all the first-order algorithms inch towards minima with the following equation:

$$\vec{r}_k = \vec{r}_{k-1} + \lambda_k \vec{S}_k \tag{2.4}$$

where $\vec{r}_k$ is the new position at step $k$, $\vec{r}_{k-1}$ represents the position at previous step $k-1$, $\lambda_k$ and $\vec{S}_k$ are step size and direction of the step respectively. Here we discuss two of the most commonly used algorithms in MD codes: the steepest descent and the conjugate gradient method.

### 2.2.1 Steepest Descent

The steepest descent(SD) uses simple logic to perform the energy minimization, that is, to move the system in the direction of the net force since that is the direction downhill towards the local minima. As shown in the eqn 5.8, the force is the negative gradient of the potential; hence we can write the direction term $\vec{S}_k$ in the eqn 2.4 as:

$$\vec{S}_k = -g_k = -\nabla U(\vec{r}) \tag{2.5}$$

after obtaining the direction $\vec{S}_k$, another important choice is the step length $\lambda_k$. In most of SD's implementations, the first value of $\lambda_k$ is an arbitrary value, which is then adjusted during minimization at every iteration based on the energy obtained at that step. If the energy obtained at any step is less than the previous step, then the value of $\lambda_k$ is increased by a preset factor. The increment is repeated for every step as long as there is a reduction in the energy. However, an increase in the energy at the next step indicates that the algorithm has crossed over minima. The energy is then reduced by using a smaller $\lambda_k$ to bring back the system closer to the minima. In this way, the algorithm eventually reaches the minima by oscillating around it and correcting itself at every step.

### 2.2.2 Conjugate Gradient

Although the SD method works well if the system is far from the minima, it becomes inefficient due to its oscillatory behavior and quadratic shape of the energy surface near the minima. Thus SD is good at bringing the system near to minima; however, it rarely converges to the exact point of minimum. The conjugate gradient(CG) method uses a unique approach, where the first step of the CG method is the same as the SD method; that is, the direction at the first step of CG is given by:

$$\vec{S}_1 = -g_1 \tag{2.6}$$

However, for the subsequent steps, the direction is given by the expression:

$$\vec{S}_k = -g_k + b_k \vec{S}_{k-1} \tag{2.7}$$

where $\vec{S}_{k-1}$ is the direction at the previous step and $b_k$ is the weight factor given by the expression:

$$b_k = |g_k|^2/|g_{k-1}|^2 \tag{2.8}$$

Where $g_k$ and $g_{k-1}$ are the current and previous steps' gradients, respectively; hence, the direction in the CG method is calculated as a weighted average of the current step's gradient and the previous step's direction. The CG method performs better than the SD method, even for non-quadratic surfaces, due to its ability to use gradient information from previous steps to evaluate the direction for future steps.

### 2.2.3 Convergence Criteria

Although numerical schemes for energy minimization are good at reducing stress in the structures and obtaining a reasonable starting structure for performing MD simulation, they are not exact as an analytical method. Hence, a numerical minimization algorithm will keep moving the system closer toward the minima forever until it is stopped using appropriate convergence criteria. The criteria could be as simple as the energy difference between two successive steps; if it goes below a certain threshold value and stays there for many steps, the minimization could be considered to be converged. Another simple approach could be to monitor changes in the coordinates of atoms. The minimization could be stopped when no significant changes in the configuration are observed for many steps. However, a more sophisticated criterion is the calculation of root-mean-square gradient(GRMS). The GRMS is calculated by taking the square root of the ratio between the addition of squares of the gradients and the number of coordinates:

$$GRMS = \sqrt{\frac{\sum g^2}{3N}} \tag{2.9}$$

## 2.3 Integration of Equations of Motion

As discussed above, the analytical solution of equations of motion(EOM) is impractical due to the complexity of the potential function and also of solving a many-body problem. Hence, clever algorithms need to be employed to perform the integration.

Many algorithms are available for performing the integration, and we will discuss the two most widely used in the MD codes, namely Verlet and velocity Verlet algorithm.

### 2.3.1 Verlet Algorithm

The idea behind the Verlet algorithm(VA)[4] is quite simple; it uses Taylor series expansion to generate new positions of the atoms. The Taylor expansion for a position $r_i(t + \Delta t)$ can be written as:

$$r_i(t + \Delta t) \approx r_i(t) + \Delta t \dot{r}_i + \frac{1}{2}\Delta t^2 \ddot{r}_i(t) \tag{2.10}$$

The terms higher than second order in the above Taylor expansion are neglected. Also, as is known, the first derivative of the position is the velocity($\dot{r}_i(t) = v_i(t)$), and the second is acceleration($\ddot{r}_i(t) = a_i(t)$), also using Newton's second law we have $a_i(t) = F_i(t)/m_i$. Hence, we can rewrite the eqn 2.10 as:

$$r_i(t + \Delta t) \approx r_i(t) + \Delta t v_i(t) + \frac{\Delta t^2}{2m_i} F_i(t) \tag{2.11}$$

A similar expression could be written for $r_i(t - \Delta t)$ as:

$$r_i(t - \Delta t) \approx r_i(t) - \Delta t v_i(t) + \frac{\Delta t^2}{2m_i} F_i(t) \tag{2.12}$$

Adding eqns 2.11 and 2.12 and rearranging we get

$$r_i(t + \Delta t) \approx 2r_i(t) - r_i(t - \Delta t) + \frac{\Delta t^2}{2m_i} F_i(t) \tag{2.13}$$

Eqn 2.13 and 2.12 are the most crucial parts of VA. The VA's execution can be elaborated in the following steps:

1. A set of energy-minimized coordinates and initial velocities is provided to the integrator at $t = 0$.

2. The eqn 2.11 is used to kick-start the integrator and obtain the first set of new coordinates $r_1(\Delta t), ....., r_N(\Delta t)$.

3. The eqn 2.13 can now be used to obtain future positions and generate an arbitrary length of the trajectory.

4. The velocities can be computed using the centered difference formula on positions as follows:

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} \tag{2.14}$$

Like any algorithm or a numerical technique, VA is also not perfect and has its share of advantages and drawbacks, which could be enlisted as follows:

Advantages:

- One of the key feature of VA is that it is simple yet accurate and is easy to implement. The memory requirements of VA are also modest.
- The VA uses the force evaluation very optimally(once at each integration step), which makes the integrator quite efficient since the force calculation is computationally the most demanding step in any MD simulation.
- Another critical characteristic of VA is that it is time-reversible, a property of EOM that a good integrator should satisfy.

Drawbacks:

- The velocity evaluation in VA is not straightforward; instead, it is inaccurate and inefficient as it involves another set of iterations with eqn 2.14.
- The VA needs an auxiliary eqn (eqn 2.12) to obtain the first step and then move on with eqn 2.13; hence it is not a self-starting algorithm.

### 2.3.2 Velocity Verlet Algorithm

Although the VA is an elegant approach to integrating the EOM, it suffers a serious drawback. It does not evolve velocities simultaneously with positions, and velocities are required for kinetic energy calculations. To overcome the shortcomings of VA, Swope et al. developed the velocity Verlet algorithm(VVA)[15]. The main idea behind the VVA is the realization that the EOM can evolve backward in time. Considering second-order Taylor expansion for position $r_i(t + \Delta t)$ we have:

$$r_i(t + \Delta t) \approx r_i(t) + \Delta t v_i(t) + \frac{\Delta t^2}{2m_i} F_i(t) \tag{2.15}$$

Now if we want to evolve the system backward in time that is from $t + \Delta t)$ to $t$ the Taylor expansion could be written as:

$$r_i(t) \approx r_i(t + \Delta t) - \Delta t v_i(t + \Delta t) + \frac{\Delta t^2}{2m_i} F_i(t + \Delta t) \tag{2.16}$$

substituting the value of $r_i(t + \Delta t)$ from eqn 2.15 in eqn 2.16 and rearranging yields

$$v_i(t + \Delta t) = v_i(t) + \frac{\Delta t}{2m_i}[F_i(t) + F_i(t + \Delta t)] \tag{2.17}$$

Eqn 2.17 is the equation used to perform velocity evolution in VVA and eqn 2.15 for position evolution. Hence, VVA can perform both velocity and position evolution simultaneously. The VVA can be executed in the following steps:

1. A set of energy-minimized coordinates and initial velocities is provided to the integrator at $t = 0$.

2. New set of coordinates are calculated at $t + \Delta t$ using eqn 2.15.

3. Forces are computed at $t + \Delta t$ using the coordinates obtained in the second step.

4. Finally the velocities are calculated at $t + \Delta t$ using the eqn 2.17

The followings are the advantages and drawbacks of VVA: Advantages:

- The most crucial aspect of VVA is its accurate and efficient evaluation of velocities and positions simultaneously.

- It is an algorithm that is straightforward to implement. A code for VVA will be a simple transcription of eqn 2.15 and 2.17 into the particular language's syntax.

- The VVA is time-reversible, which contributes to its numerical stability.

The only serious drawback of VVA is that it is computationally a little more expensive than VA. However, the accuracy and the ease of velocity evaluation provided by VVA outweigh this little expense.

## 2.4 Ensemble

The literal meaning of the word ensemble is "a band or group of musicians." Like in a music band, each member contributes, and the symphony arises from the average of these contributions. Similarly, a macroscopic observable is the resultant of the average of a microscopic property over the possible number of microstates. The idea of the ensemble was first proposed by J.W. Gibbs in his excellent treatise "Elementary Principles in Statistical Mechanics"[16]. Formally, an ensemble could be described as a collection of systems differing in microscopic details(position and momenta) but with the same macroscopic constraints(e.g., number of particles, pressure or total energy, or temperature). The concept of the ensemble could be summarized mathematically with a heuristic expression as follows:

$$M = \frac{1}{\mathcal{Z}} \sum_{i=1}^{\mathcal{Z}} m(X_i) \tag{2.18}$$

Where M is a macroscopic observable, $\mathcal{Z}$ is the number of members in the ensemble, $m(X)$ represents a microscopic phase space function, and $X$ denotes a coordinate on the phase space. In a nutshell, the idea of an ensemble facilitates us to link the macroscopic and microscopic worlds. For a detailed discussion, the reader is referred to excellent texts by Tuckerman[17], Bagchi[18], and others on the subject. Next, we will discuss different types of ensembles. Based on the different macroscopic constraints imposed, there can be many types of ensembles; however, we will discuss the three most common and practically useful ensembles, namely the microcanonical, canonical, and isothermal-isobaric.

### 2.4.1 Microcanonical(NVE) ensemble

The microcanonical(NVE) is the simplest and the most fundamental of all the ensembles. It is characterized by the constant number of particles(N), constant volume(V), and constant total energy(E). One of the most crucial aspects of the NVE ensemble is that it perfectly satisfies the fundamental statistical mechanical postulate of "equal a prior probability." Since *equal a prior probablity* postulate propound that all microstates in a given ensemble are equally probable, this statement inevitably leads to the fact that each microstate possess the same total energy. This feature of

the NVE ensemble enables it to generate detailed dynamics of the given system since all possible microstates can be visited provided enough time. The thermodynamic state function unique to the NVE ensemble is entropy(S), and all other thermodynamic quantities in an NVE ensemble can be obtained from it by the following relations:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V}, \qquad \frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{N,E}, \qquad \frac{\mu}{T} = -\left(\frac{\partial S}{\partial N}\right)_{V,E} \tag{2.19}$$

Where T is the temperature, P is pressure, and $\mu$ is the chemical potential for the given system. The relation between macroscopic observables and a microscopic property for an NVE ensemble is given by the famous "Boltzmann relation":

$$S = k_B ln\Omega \tag{2.20}$$

Where S is the entropy of the system, $k_B$ is known as Boltzmann's constant and is equal to $1.3806 \times 10^{-23} J.K^{-1}$ and $\Omega$ is known as the partition function for NVE ensemble and it defines the number of microstates available to the system. Using the eqn 2.20 and eqns 2.19, all thermodynamic properties can be written in terms of $\Omega$. However, the main limitation of the NVE ensemble is that it does not reflect the systems commonly encountered in most of the experiments and other realistic environments. Since, in most of the experiments, the system is allowed to interact with its surroundings, which makes NVE impractical for most cases.

### 2.4.2 Canonical(NVT) Ensemble

The thermodynamic constraints used for the canonical(NVT) ensemble are the constant number of particles(N), constant volume(V), and constant temperature(T). A clear advantage NVT has over NVE is that NVT mimics the commonly encountered experimental setups more closely. The conditions for an NVT ensemble are achieved theoretically by assuming that the system under consideration is in thermal contact with an external infinite heat bath and the system's energy fluctuates in such a way that its temperature remains constant. The principle thermodynamic state function for the NVT ensemble is the Helmholtz energy(A). The thermodynamic quantities for the NVT ensemble are defined in terms of A as follows:

$$S = -\left(\frac{\partial A}{\partial T}\right)_{N,V}, \qquad \mu = \left(\frac{\partial A}{\partial N}\right)_{V,T}, \qquad P = -\left(\frac{\partial A}{\partial V}\right)_{N,T} \tag{2.21}$$

The connection between macroscopic thermodynamic quantities and microscopic property is given by the relation:

$$A = -k_B T ln Q \tag{2.22}$$

Where Q is known as the canonical partition function and $k_B$ is the Boltzmann's constant.

Although the NVT ensemble exhibits an advantage over NVE of being closer to the realistic systems. However, a straightforward integration of EOM will produce an NVE ensemble. Hence, we require methods to control the temperature in an MD simulation in order to generate an NVT ensemble. These methods are known as thermostats, and we will discuss a few of these methods in the upcoming section.

### Temperature Control or Thermostats

***Velocity Rescaling***: The easiest way to control the temperature is to rescale the velocities at every time step. The relation between velocities and temperature is given by the well-known result,

$$\langle \sum_i \frac{1}{2} m_i {v_i}^2 \rangle = \frac{1}{2} k_B T \tag{2.23}$$

Then velocities can be rescaled by using an appropriate factor given by:

$$\lambda = \sqrt{\frac{T_{desired}}{T(t)}} \tag{2.24}$$

Although convenient in its approach, the method causes significant perturbation of the trajectory. Moreover, the rescaled velocities may not lie within the Boltzmann distribtion; hence, the method does not guarantee a canonical distribution.

***Andersen Thermostat***: The method suggested by Andersen[19] to control the temperature in an MD simulation is based on simple velocity assignment from a Maxwell-Boltzmann distribution of velocities. According to this method, the coupling between the system and the heat bath is represented by stochastic dynamics of the collisions between the system's particles and the heat bath. These collisions occur periodically on randomly selected particles. The particle which undergoes collision is assigned a new velocity from a Maxwell-Boltzmann distribution of velocities corresponding to the target temperature. If the frequency of these stochastic

collisions is given by $\nu$ and t is the time between two successive collisions, then the probability for a collision is by the Poisson process,

$$P(\nu, t) = \nu e^{-\nu t} \tag{2.25}$$

Although the Andersen method successfully generates the canonical ensemble and is also easy to implement, the dynamics generated by it is unphysical. Hence, it must be avoided whenever the goal of the study is to calculate the dynamical properties of the system, e.g., diffusivity, viscosity, thermal conductivity, etc. However, static properties such as pressure, density, etc., remain unaffected.

***Nosé–Hoover Thermostat***: Andersen showed that a stochastic approach could generate a canonical ensemble. However, Nosé–Hoover[20, 21] devised a clever method to generate canonical phase space distribution with a deterministic approach. In this method, the EOM is tweaked by adding additional terms to attain temperature control; such methods are called extended system methods. The main idea behind the Nosé–Hoover thermostat is that an additional term is introduced in the EOM such that it can monitor any deviation in the instantaneous kinetic energy corresponding to the desired temperature and scale the velocities accordingly. The modified Lagrangian of EOM for Nosé–Hoover method is given by:

$$\mathcal{L}_{N-H} = \sum_{i=1}^{N} \frac{1}{2} m_i (s\dot{r}_i)^2 - U(r^N) + \frac{Q\dot{s}^2}{2} - g k_B T ln s \tag{2.26}$$

Where $s$ is the position of the imaginary heat bath,$\dot{s}$ is the conjugate velocity of the heat bath, and $Q$ is the effective mass of the heat bath. The Nosé–Hoover thermostat generates accurate canonical distribution. Additionally, the dynamics generated within it are more realistic than the Andersen thermostat. However, the convergence of Nosé–Hoover thermostat is a bit slower due to the presence of second-order terms. Hence, it must be avoided for the systems that are far away from the equilibrium.

### 2.4.3 Isobaric-Isothermal(NPT) Ensemble

Most of the condensed-phase experiments are conducted under constant temperature and pressure. Moreover, most of the standard thermodynamic data are reported

under constant temperature and pressure conditions. This makes the isothermal-isobaric(NPT) the most useful of all the ensembles, since it is closest to the physical conditions at which experiments are performed. The thermodynamic control variables of NPT ensemble are constant number of particles(N), constant pressure(P) and constant temperature(T). The basic thermodynamic quantities of NPT ensemble are defined using a new state function, the Gibbs free energy(G). The relations are written as:

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{P,T}, \qquad \langle V \rangle = \left(\frac{\partial G}{\partial P}\right)_{N,T}, \qquad S = -\left(\frac{\partial G}{\partial T}\right)_{N,P} \tag{2.27}$$

The link between macroscopic and microscopic worlds for an NPT ensemble is given by the relation:

$$G = -k_B T ln\Delta \tag{2.28}$$

Where all the variables have the same meaning defined previously, and $\Delta$ is the partition function for an NPT ensemble. The concept of free energy obtained in the formulation of NVT and NPT is quite useful since the physical interpretation of it is the work required to change the system's state from state 1 to state 2. Hence free energy is an indicator of the thermodynamic feasibility of a process.

### Pressure Control or Barostats

In the previous section, we discussed methods for controlling the temperature to generate an NVT ensemble. However, in order to perform an MD simulation with an NPT ensemble, it is crucial to control pressure along with the temperature. Here we discuss a couple of methods to maintain pressure in an MD simulation.

***Berendsen Barostat***: In the Berendsen method[22], the system's pressure is controlled by scaling the dimension of the simulation box and coordinates of the particles. The rate of change of pressure is given by the relation:

$$\frac{dP}{dt} = \frac{(P_0 - P)}{\tau_p} \tag{2.29}$$

Where $\tau_p$ is the coupling constant, $P_0$ is the desired pressure, and $P$ is the instantaneous pressure. The scaling factor is given by :

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_p}(P_0 - P) \tag{2.30}$$

Here $\mu$ is the scaling tensor used to scale the box size and coordinates of the particles to maintain the system's pressure at the desired value($P_0$), $\beta$ is the isothermal compressibility of the system and $\Delta t$ is the time step. Berendsen's method is an example of a weak coupling barostat. Hence, it is more appropriate to use it in situations where pressure deviation is significant, mainly in the initial stages of the simulation (e.g., equilibration).

***Anderson Barostat***: The Anderson pressure coupling method[19] uses an imaginary piston to control the pressure. The piston and the original system form an extended system, and the EOM for such a system is given by:

$$\dot{r} = \frac{p_i}{m_i} + \frac{1}{3}\left(\frac{\dot{V}}{V}\right) r_i \tag{2.31}$$

$$\dot{p} = F_i + \frac{1}{3}\left(\frac{\dot{V}}{V}\right) p_i \tag{2.32}$$

$$\ddot{V} = \frac{P - P_0}{M} \tag{2.33}$$

Here $P$ is the instantaneous pressure, $P_0$ is the desired pressure, V is the volume, M is the mass of the imaginary piston, and $r_i, p_i, m_i$ and $F_i$ the position, momentum, mass, and force for the ith particle respectively. The EOM of Anderson barostat(2.31-2.33) samples an isobaric-isoenthalpic ensemble. However, combining an appropriate thermostat with Andersen's barostat will generate the NPT ensemble. The choice of mass for the imaginary piston is crucial here since it will decide the decay of volume fluctuations in the ensemble.

## 2.5 Constraints

Atomistic simulations have a varying degree of freedom(DF) of motions such as bond vibration, angle bending, and torsional motion. The time step for the integration of EOM is usually governed by DF with the fastest motion or the highest frequency present within the system. However, accounting for these higher-frequency motions can slow down the simulation considerably. Moreover, these high-frequency motions are not crucial in general in the overall configurational phase space exploration,

particularly for bigger systems such as biomolecular systems. Hence, it is customary in MD simulations to fix or constrain these DF, which allows the use of a bigger time step and eventually speeds up the simulations.

Implementation of constraints requires modifying the EOM to incorporate the effect of constraints. This modification is done by using the method of Lagrange's undetermined multipliers. The resulting EOM can be expressed as:

$$m_i\ddot{r} = F_i + \sum_{k=1}^{N_c} \lambda_k \nabla_i \sigma_k \tag{2.34}$$

The last term in the eqn. 2.34 represents the force experienced by the atom due to constraints. The $\lambda_k$ is a set of Lagrange's multiplier to enforce the constraints, and $\sigma_k$ are holonomic(depends only on coordinates of atoms involved) constraints. Although it is possible to obtain an exact solution for Lagrange's multiplier for a few atoms, it becomes increasingly impractical as the system size increases. Additionally, as EOM in an MD simulation is solved using one of the numerical schemes discussed in a previous section, incorporating an exact analytical expression for $\lambda_k$ will accumulate errors in the calculations. Hence, to avoid these problems, numerical algorithms such as SHAKE[23] and RATTLE[24] to calculate the multipliers during the simulation run. These algorithms evaluate the multipliers iteratively for each constraint until it converges to a very small tolerance. Although constraints are usually employed to fix the bond vibrations involving hydrogen atoms, they can also be used to fix angle bending in certain molecules by constraining the distance between the two end atoms involved in the angle formation.

## 2.6 Periodic Boundary Condition

A bulk or macroscopic system is characterized as a system containing at least an Avogadro number($10^{23}$) of particles(atoms or molecules). However, even the fastest supercomputer couldn't perform an explicit MD simulation with such an astronomical number of particles. Another serious issue is that in a macroscopic system, the effect of boundaries(wall effect) on the bulk properties is negligible; however, wall effects can't be neglected in a much smaller MD system. Hence, to link the microscopic states generated with the MD simulation and its corresponding macroscopic

system and to reproduce the bulk properties, it is required to treat the boundaries of the MD system cleverly. Periodic boundary conditions(PBC) are one of the most commonly used boundary conditions in MD simulations. In this method, the simulation box is replicated infinitely in all three dimensions, creating a periodic lattice. The motion of particles in these imaginary boxes mimics that of the corresponding particles in the original box. A schematic representation of a two-dimensional PBC is shown in Figure 2.2.

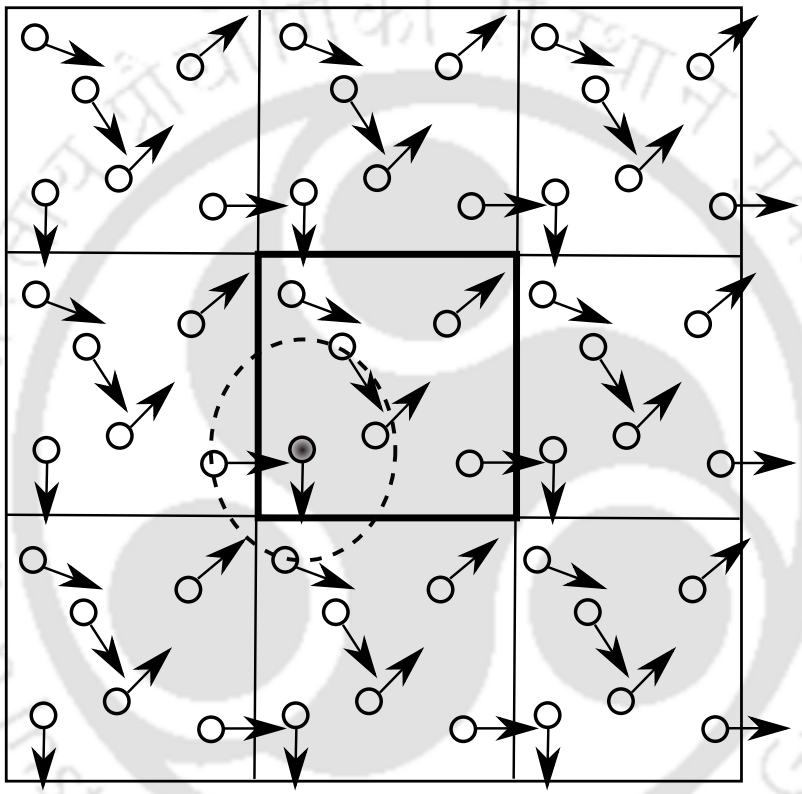


Figure 2.2: A schematic representation of two dimensional periodic boundary condition. The highlighted box in the center is the original simulation box and all other boxes are images of it.

Any particle that leaves one side of the simulation box will re-enter the system from the opposite side, as shown in Figure 2.2. As a result, the system is virtually infinite, and boundary effects are minimized. The PBC allows for the accurate reproduction of bulk properties of the macroscopic system, as the simulation is no longer confined to a small volume but instead simulates an infinitely extended system.

### 2.6.1 Minimum Image Convention and Cutoff

The implementation of PBC resolved the issue of wall effects and facilitated mimicking a bulk system without increasing the size of the actual system. However, since PBC created an infinite system, such a setup poses a challenge for evaluating non-bonded interaction across the periodic boundaries. To overcome this problem, *minimum image convention*(MIC) is used, which essentially governs the non-bonded interactions across the periodic boundaries. The MIC states that, in order to calculate the interactions between two particles across a periodic boundary, the closest periodic image of the other particle should be considered instead of the original one. This is because the interactions between particles decay rapidly with distance, so the contribution of the particles from beyond the cutoff distance is negligible. By considering the closest periodic image, the simulation accurately captures the interaction between the two particles without requiring an excessively large simulation box.

Evaluation of non-bonded interactions is the most time-consuming part of an MD simulation. Although the MIC is a clever trick to compute non-bonded interactions across the boundaries, the calculation of all pairwise interactions, even within the simulation cell, is computationally expensive and, in most cases, unnecessary. In practice, the interactions between particles decay rapidly with distance and become negligible beyond a certain distance. Hence, by introducing a *cutoff* distance, the simulation only calculates interactions between particles that are close enough to each other to have a significant effect and ignores interactions between distant particles that would have a negligible effect. Figure 2.2 shows the particle under consideration in solid color, and the dashed circle represents the cutoff distance for the particle. The choice of cutoff distance is a balance between accuracy and efficiency. A smaller cutoff distance will result in a more accurate simulation, but at the cost of increased computational expense. A larger cutoff distance will result in a less precise simulation with reduced computational cost. The appropriate cutoff distance depends on the simulated system and the accuracy required for the results.

## 2.7 Forcefield

The force model or forcefield is the heart of MD simulation because the results obtained from the simulation are as accurate as the forcefield used. A forcefield is a mathematical expression for the potential energy function of the system in terms of the various modes of interaction among the atoms in the system. These modes of interaction are itself function of coordinates and orientation of the atoms. A simple functional form of the forcefield can be written as:

$$\begin{aligned} U(r_1, ...., r_N) = & \sum_{bonds} \frac{1}{2} K_{bond}(r - r_0)^2 + \sum_{angles} K_{angle}(\theta - \theta_0)^2 \\ & + \sum_{torsion} A_n[1 + cos(C_n\phi + \delta_n)] \\ & + \sum_{i,j \in nb} \left\{ \left[ 4\epsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\} \end{aligned} \tag{2.35}$$

The eqn. 2.35 is a basic forcefield expression commonly used for biomolecular systems. Each term in the eqn 2.35 contains some parameters which are specific to the model used to represent that term, for e.g. $K_{bonds}$(force constant) and $r_0$(equilibrium bond length) for the first term. The evaluation of these parameters is done using quantum mechanical calculations and is called as parametrization of a forcefield. Some popular forcefields frequently used for biological macromolecules(proteins, DNA, RNA etc.) are AMBER[25], GROMOS[26] and CHARMM[27, 28]. Each of these forcefields target to reproduce a specific set of experimental properties, and hence are developed and optimized accordingly. Also, these forcefields differ in the parametrization methodologies and treatment of a particular term. However, the basic skeleton of each of these forcefields is same as eqn. 2.35.

The contribution of each term to the total potential energy in the eqn. 2.35 can be divided into two fundamental modes of interation: *Bonded* and *Non-bonded*.

*Bonded*: The first three terms in the eqn. 2.35 represents the bonded part of the total potential energy. The first term is the energy for all covalently bonded pairs of the atoms. The bond between the two atoms is modeled as a harmonic spring with atoms attached to the ends of the spring. The parameters for the first term are the force constant $K_{bond}$ and the equilibrium bond length $r_0$. The second term is the angle bending energy of all neighboring atoms which have one common covalently

bonded partner. The angle bending contribution is also treated as harmonic spring as for the first term, however, the parameter here are the equilibrium bond angle $\theta_0$ and the force constant for angle bending $K_{angle}$. The third term is the energy due to the torsional motion between four consecutively bonded atoms. The torsional contribution is expressed in the form of a cosine series due to the oscillatory nature of the torsions, containing multiple minima separated by energy barriers of various sizes, representing different possible dihedral conformations. The $A_n$ in the torsional term is called as barrier height and is a measure of barrier to the bond rotation. The $C_n$ is the multiplicity and its value represents the number of minimas as the bond is rotated through 360° and the $\delta_n$ is the phase factor which represents the torsional angle at which the value of torsion potential is minimum.

*Non-bonded*: The last term in the curly bracket is the potential energy contribution due to non-bonded(NB) interactions. The NB term is made up of two contributions, the first is van der Waals(vdW) interaction between two atoms and is modeled using Lennard-Jones(LJ) potential(the term insided the square bracket of the eqn. 2.35). The $\sigma_i$ and $\sigma_j$ are the radius of the two atoms and $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and the atoms are separated by a distance $r_{ij}$ with a potential well depth of $\epsilon_{ij}$. The second part of the NB term is the electrostatic interaction between two atoms having charges $q_i$ and $q_j$ and are separated by a distance $r_{ij}$. The electrostatic contribution is evaluated by Coulomb's law.

## 2.8 Ewald Summation

In the previous section, we discussed the idea of a forcefield in an MD simulation through a simplified mathematical expression for the potential energy. Among the two types of contribution(bonded and NB) to the potential energy, NB is computationally the most expensive one to evaluate in an MD simulation. The vdW interactions of the NB are relatively short-range and hence can be dealt with by an appropriate choice of cutoff as discussed in the section 2.6.1. However, the evaluation of electrostatic interactions is trickier due to the long-range nature of the Coulombic interactions. The relationship between vdW and Coulombic interactions

with the length scale is evident from the eqn. 2.36.

$$U_{NB}(r_1, ...., r_N) = \sum_{i,j \in nb} \left\{ \left[ 4\epsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\} \tag{2.36}$$

To tackle the long-range nature of the Columbic interactions, Ewald introduced a technique now known as Ewald summation[29]. The basic idea of the Ewald summation is that a function that is long-range in real space is short-range in the reciprocal or Fourier space. Hence, if the Columbic term could be transformed into a Fourier representation, it could be evaluated efficiently. We write the reciprocal term in the Columbic contribution of eqn to achieve this. 2.36 as :

$$\frac{1}{r} = \frac{erf(\alpha r)}{r} + \frac{erfc(\alpha r)}{r} \tag{2.37}$$

where the $erf(\alpha r)$ is the error function and $erfc(\alpha r)$ is its complement, both are defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt\, e^{-t^2} \tag{2.38}$$

$$erfc(x) = 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty dt\, e^{-t^2} \tag{2.39}$$

It can be noted that, based on the definitions given above, the first term of the eqn. 2.37 is long-range while the second is short-range. Hence, the short-range term can be evaluated along with the vdW term in the real space, while the long-range term can be computed in the Fourier space. Thus using eqn. 2.37 the Coulombic term can be written as a sum of two terms as follows:

$$U_{Coloumbic} = \underbrace{\frac{q_i q_j erf(\alpha r_{ij})}{r_{ij}}}_{\text{long}} + \underbrace{\frac{q_i q_j erfc(\alpha r_{ij})}{r_{ij}}}_{\text{short}} \tag{2.40}$$

As explained, the short-range term can be evaluated along with the vdW term. The $\alpha$ in the above equations is used to tune in an appropriate cutoff distance for the short-range Coulombic interactions. The long-range part is computed using the Poisson summation rule leading to a Fourier series expansion of the error function.

The final equation obtained is as follows:

$$U_{long} = \frac{1}{V} \sum \frac{4\pi}{|g|^2} exp\left(\frac{|g|^2}{4\alpha^2}\right) \left|\sum q_i exp(ig.r_i)\right|^2 - \frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \tag{2.41}$$

The eqn. 2.41 is called Ewald sum. The $V$ and $g$ are the volume of the simulation cell and the reciprocal space vectors, respectively. Although eqn. 2.41 tackled the calculation of long-range Coulombic interactions ingeniously; direct implementation of it is still computationally expensive, particularly for large systems. Essman et al. introduced the smooth particle mesh Ewald (SPME) method[30] for efficient evaluation of Ewald sum. In SPME, the space is divided into grids, and the charges are uniformly distributed across its nodes; this is achieved using an interpolation scheme with a B-spline function.

## 2.9 Water Models

Water plays a crucial role in many chemical and biological processes. For example, water mediates the folding of protein from peptide bonded linear amino chains to a functional three dimensional structure. Therefore, owing to its importance in driving and/or mediating various phenomenas, it is essential to incorporate water in MD simulations. There are numerous water models available and each one is developed to target a particular set of propeties of the water. Hence, no single water model can reproduce all the properties of water, and one must choose a model based on the target of the investigation under consideration. Further, these models also vary at the level of complexity and sophistication, hence, choice of a model must be made by keeping in the mind the required accuracy and the availability of computational resources.

Among the various water models available three point, and four point models are most commonly used. We will discuss briefly some of the popular water models in each of these categories.

**Three Point Models**

Three point water models are one of the earliest and simplest models. In this model the water molecule is represented by three point charge for hydrogen and oxygen

atoms. Despite their simplicity, three-point water models can provide reasonable approximations of certain properties and behaviors of real water, making them computationally efficient for a wide range of applications. We will discuss two widely used three point models namely TIP3P and SPC/E.

**TIP3P**

Transferable Intermolecular Potential with Three Points (TIP3P) is one of the earliest and most widely used three point model. It was proposed by Jorgensen et al.[31]. In the TIP3P model, each water molecule is represented by three point charges: a negative charge at the oxygen atom (-1.0 e) and two positive charges at the hydrogen atoms (+0.52 e). The atomic charges and Lennard-Jones parameters are fitted to reproduce experimental properties of water, such as the density and heat of vaporization. The O–H bond length and H–O–H bond angle used in TIP3P model are 0.9572 Åand 104.52° respectively.

Pros:

- TIP3P's simplicity reduces the number of interaction terms and makes simulations computationally faster compared to more complex models.
- TIP3P can reasonably reproduce properties like density, radial distribution functions, and self-diffusion coefficients of water.

Cons:

- TIP3P tends to overestimate the hydrogen bond length and underestimate the O–H–O bond angle, resulting in a less accurate representation of the water structure.
- The inaccuracies in the water structure and dynamics make TIP3P less suitable for studies where accurate hydrogen bonding is crucial, such as certain biomolecular systems or phase transitions.

**SPC/E**

SPC/E model[32] is very similar to TIP3P, each water molecule is represented as three point charges: one negative charge located at the oxygen atom (-0.8476 e) and two positive charges located at the hydrogen atoms (+0.4238 e). The O–H

bond length and H–O–H bond angle used in SPC/E model are 1 Åand 109.47° respectively. The target properties of SPC/E are the density and the vaporization enthalpy at room temperature. However, the key difference between TIP3P and SPC/E is the use of a polarization energy correction in case of SPC/E to correctly reproduce the vapourization enthalpy of real water.

Pros:

- One of the major improvements of the SPC/E model over its predecessor is its ability to better reproduce the structural properties of water, such as the O-H-O bond angle and radial distribution functions. The model's parameters are optimized to match experimental data more closely.
- SPC/E is designed to facilitate the formation of hydrogen bonds due to its fractional charges and realistic charge distribution. This makes it more suitable for studies involving hydrogen bond-dependent processes.
- While SPC/E introduces additional complexity compared to three-point models like TIP3P, it remains relatively computationally efficient due to its simplified charge distribution and lack of polarizability.

Cons:

- Despite its improvements, the SPC/E model still makes simplifications that can lead to deviations from real water's behavior. For example, the model does not account for polarization effects, which can play a significant role in some systems.

**Four point water models**

Four point water models improve upon the limitations of three-point models by incorporating a Lennard-Jones site for the lone pair electrons on the oxygen atom. This inclusion allows for better representation of water's structure and dynamics, but it also increases the model's complexity.

**TIP4P**

Being a four point model TIP4P is represented by three point charges and one Lennard-Jones site. However, the key feature of this model is that the site carrying

the negative charge is not located at the oxygen atom but on the H–O–H bisector at a distance of 0.15 Å, effectively accounting for the oxygen atom's lone pair electrons. The TIP4P model was proposed by Jorgensen et al.[31].

Pros:

- The addition of the Lennard-Jones site for the oxygen atom's lone pair electrons allows the TIP4P model to more accurately represent hydrogen bonding interactions compared to TIP3P.
- Like TIP3P, TIP4P is designed to be transferable and can reproduce a variety of water properties across different systems.

Cons:

- The addition of the Lennard-Jones site makes TIP4P slightly more complex than TIP3P, which may result in slightly longer computation times.
- While TIP4P improves upon TIP3P's accuracy, it requires a more intricate parameterization process to ensure that the new interaction site appropriately represents the oxygen atom's lone pair electrons.

## 2.10 Softwares used and Implementation

All MD simulations in this thesis were performed using NAMD[14] package. VMD[33] and PACKMOL[34] were used to generate the initial configurations of the atomistic models. VMD was used to visualize the simulation trajectories and render the snapshots. The postprocessing of snapshots was done using Inkscape graphics software. All the analysis were performed using in-house developed codes written in Tcl, python, and bash. The plots of all the analysis were prepared using Gnuplot and Matplotlib library of python. The simulations were performed with high performance computing (HPC) facility, and the runs were submitted to HPC using SLURM scripts.

Finally we present a general scheme for an MD simulation in Figure2.3, first the coordinates of the structure to be simulated is obtained then that structure is minimized using a suitable minimization algorithm, solvent molecules are added

if required and the initial velocities are assigned using boltzmann distribution and then the equilibration and production run are conducted with the desired ensemble.

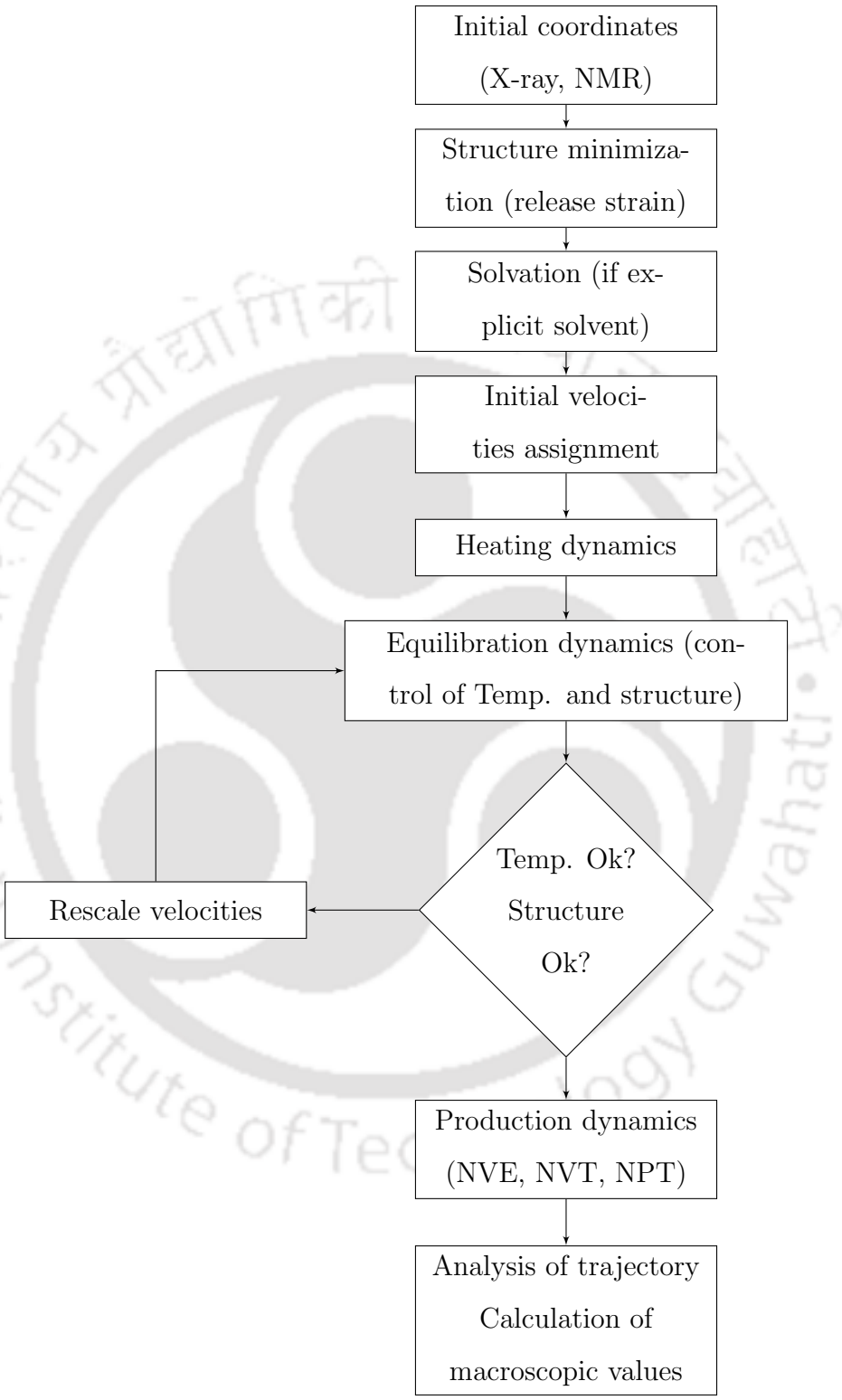



Figure 2.3: A general work flow of operations in general in an MD simulation.

# Bibliography

[1] Berni Julian Alder and Thomas Everett Wainwright. Phase transition for a hard sphere system. *The Journal of chemical physics*, 27(5):1208–1209, 1957.

[2] Berni J Alder and Thomas Everett Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[3] Aneesur Rahman. Correlations in the motion of atoms in liquid argon. *Physical review*, 136(2A):A405, 1964.

[4] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[5] GD Harp and Bruce J Berne. Linear-and angular-momentum autocorrelation functions in diatomic liquids. *The Journal of chemical physics*, 49(3):1249–1254, 1968.

[6] BJ Berne, P Pechukas, and GD Harp. Molecular reorientation in liquids and gases. *The Journal of Chemical Physics*, 49(7):3125–3129, 1968.

[7] GD Harp and BJ Berne. Time-correlation functions, memory functions, and molecular dynamics. *Physical Review A*, 2(3):975, 1970.

[8] BRUCE J BERNE. Time-dependent properties of condensed media. *Physical chemistry, an advanced treatise*, 8:539–716, 1971.

[9] Aneesur Rahman and Frank H Stillinger. Molecular dynamics study of liquid water. *The Journal of Chemical Physics*, 55(7):3336–3359, 1971.

[10] Frank H Stillinger and Aneesur Rahman. Molecular dynamics study of temperature effects on water structure and kinetics. *The Journal of chemical physics*, 57(3):1281–1292, 1972.

[11] Frank H Stillinger and Aneesur Rahman. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, 60(4):1545–1557, 1974.

[12] J Andrew McCammon, Bruce R Gelin, Martin Karplus, and PETER G WOLYNES. The hinge-bending mode in lysozyme. *Nature*, 262(5566):325–326, 1976.

[13] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. Dynamics of folded proteins. *nature*, 267(5612):585–590, 1977.

[14] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[15] William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of chemical physics*, 76(1):637–649, 1982.

[16] Josiah Willard Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner's sons, 1902.

[17] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2010.

[18] Biman Bagchi. *Statistical mechanics for chemistry and materials science*. CRC Press, 2018.

[19] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics*, 72(4):2384–2393, 1980.

[20] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

[21] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.

[22] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[23] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[24] Hans C Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of computational Physics*, 52(1):24–34, 1983.

[25] David A Pearlman, David A Case, James W Caldwell, Wilson S Ross, Thomas E

Cheatham III, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.

[26] Wilfred F van Gunsteren, SR Billeter, AA Eising, PH Hünenberger, PKHC Krüger, AE Mark, WRP Scott, and IG Tironi. Biomolecular simulation: the gromos96 manual and user guide. *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*, 86:1–1044, 1996.

[27] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[28] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

[29] Paul P Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der physik*, 369(3):253–287, 1921.

[30] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[31] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[32] Herman JC Berendsen, J Raul Grigera, and Tjerk P Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[33] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[34] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario

Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.

# Chapter 3

# What Governs Fouling in FO and RO?

## 3.1 Introduction

Since the advent of membrane-based water purification techniques, a plethora of research and development has revolutionized the large scale production of potable water in terms of performance. Currently, Reverse Osmosis(RO) is the commercially most viable membrane technology available to produce the highest quality of purified water. In the RO system, the water with impurities is passed through a semipermeable membrane with the application of hydraulic pressure with a magnitude higher than the osmotic pressure of feed solution.

In the past decade, Forward osmosis(FO) has emerged as a promising water purification technology. Ever since then, FO is finding applications in various fields such as food processing, desalination, industrial water reclamation. [4][8] [28] FO system exploits phenomena of natural osmosis as its driving force. An FO system has two sections, namely, draw solute (DS), which is a solution with very high osmotic pressure and feed solution (FS) separated by a semipermeable membrane. The net movement of the solvent molecules from the feed side to draw side is achieved through the difference in the osmotic pressure between feed solution and DS [4]. A crucial factor for why is FO drawing huge attention over RO is that FO can be operated at much lower energy input if a concentrated DS with high regeneratibility

is available.

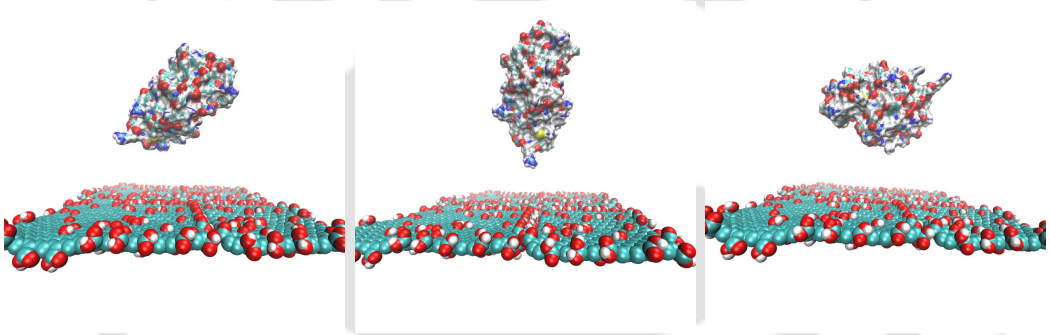
Almost all membrane separation techniques suffer from a performance hampering phenomena called membrane fouling.[46] [35] [2] [25] Membrane fouling is accumulation and adsorption of foulant molecules on the membrane surface, causing a loss in performance of membrane in terms of water flux and salt rejection. Apart from the advantage of low energy consumption that FO has over RO, fouling performance is also an important parameter need to be considered to make a fair comparison between both the technologies(FO and RO). The available literature on FO and RO have broadly two opposing views over the issue of fouling behavior of FO and RO. The conventional view is that RO suffers much severe fouling because of hydraulic compaction of foulants over the membrane surface, which makes fouling irreversible in RO. While in FO, lack of hydraulic pressure results in loosely compacted foulant layer and hence is easily reversible.[30] [3] [26] Lee et al.[26] compared fouling behaviors for FO and RO for a cellulose-based membrane, they found that although there is higher decline in water flux in FO, it can be completely recovered while in RO the fouling is almost irreversible. Mi and Elimelech [30] also demonstrated the reversible nature of fouling for FO and hence very high water flux recovery (almost 100%) after cleaning while for RO it was about 70%.

Apart from the above studies, there are few studies which are in contrast with the claim that RO is more fouling prone than FO.[24][21][42][38] In a recent experimental study[38] Siddiqui et al. compared fouling for FO and RO with cellulose triacetate(CTA) membrane and demonstrated a higher fouling propensity for FO. They quantified the membrane fouling with foulant layer resistance and thickness. Also, they showed that quantifying fouling based solely on water flux could be error prone as it overlooks the effect of concentration polarization, which is inherent in FO. They also tested the effect of hydraulic pressure on fouling for RO and found no effect on the compaction of the foulant layer and water flux. The claim the above study made is that although fouling is more severe in FO, but the nature of fouling in FO is reversible and hence is more fouling resilient, and the mechanism given in support of the above claim are internal concentration polarization(ICP) self-compensation effect, cake enhanced concentration polarization(CECP) and reverse solute diffusion.

Despite quite a number of experimental studies on FO and RO fouling, there is a need of exploring fouling at a molecular scale, because at that scale we can understand the initial interaction of foulant and membrane explicitly, which can suggest a possible mechanism of fouling. Since such resolutions are not accessible through experiments, molecular dynamics (MD) simulations with its high temporal and spatial resolution can play a pivotal role in exploring membrane fouling. In the past few years MD simulations have become an integral part of membrane research. There have been huge number of simulation studies exploring performance of various membranes[5][6][16][7][27][11] but there are relatively less number of simulation studies investigating membrane fouling. Xiang et al.[44][45] performed steered MD (SMD) to investigate the effects of ions($Na^+$,$Ca^{2+}$) on alginate fouling of polyamide(PA) membranes. They demonstrated the formation of ionic bridges due to ions between alginate and membranes carboxylate(COOH) groups and hence explained aggravation of fouling in presence of ions. Hughes and Gale[17] tested three foulants (glucose, phenol and oxygen) in a simulation with a RO module of a PA membrane. They calculated free energies and diffusion coefficents for all three foulants. In a recent study Shaikh et al.[36] performed an MD simulation of bovine serum albumin(BSA) a protein as a model foulant with Poly(vinylidene fluoride) (PVDF) membrane.They also investigated the interaction between protein and PVDF membrane and effect of salt concentration on the interaction. The study demonstrated that the excess ions repels the protein away from membrane due to both membrane and protein surrounded by negatively charged chlorine ions. So it is evident that MD simulation is an important tool which can help to have a much closer and better visualization of fouling mechanism.

In the current investigation, we aim to understand the fouling mechanism for FO and RO at the atomistic level through MD simulations. We used layered graphene oxide (GO) as the membrane for both FO and RO processes. To characterize the impact of foulant–membrane and foulant–foulant interactions on fouling behavior, we used two molecules of lysozyme as a model foulant, which was used in many experimental studies[13][43][48]. Furthermore, to analyze the effect of protein orientation relative to the membrane surface, we simulated three systems with different protein conformation, namely, orientation I (O1), II (O2), and III (O3) as shown

in Fig 3.1. To unravel the nature of interactions between protein–protein and protein–membrane, we analyzed the interaction energy and distance evolution between them. We also illustrated the influence of solvent and ions on fouling and how it gave rise to different fouling layer structures in FO and RO through the analysis of the radial distribution function (RDF). Finally, we detected the key interacting residues in lysozyme and the nature of its interaction with the functional groups present on GO. Overall we proposed a probable mechanism for the different fouling behavior in FO and RO, detected the changes concerning the energetics of the system, and the key interacting residues to gain a detailed insight into the interaction between lysozyme and GO in FO and RO.



(a) (b) (c)

Figure 3.1: Three different initial orientation of protein with respect to GO surface, namely (a) O1 (b) O2 (c) O3

## 3.2 Methodology

### 3.2.1 Layered Graphene Oxide Membrane

For the construction of the layered GO membrane, first individual GO nanosheets are constructed using using Visual Molecular Dynamics (VMD) [18] and Avogadro [14]. Based on the previous experimental and simulation studies [41][37][11] [12] the composition of the GO nanosheets is considered as $C_{10}O_1(OH)_1(COOH)_{0.5}$. The hydroxyl (-OH) and epoxy (-O-) functional groups are located on the basal plane

of the GO nanosheets and the carboxyl (-COOH) functional groups are located on the edges. Although practically GO contains negatively charged groups, which are usually deprotonated carboxyl groups, the charge and protonation state of a GO sheet in a solution are governed by the pH of that solution[37]. Since in the present study pH effects are not important we assume the carboxyl groups to remain protonated. A representative image of the GO nanosheet is shown in Figure 3.2. The sizes of the GO nanosheets considered in the present study are $48 \times 48$ Å$^2$, $40 \times 48$ Å$^2$, $32 \times 48$ Å$^2$ and $22 \times 48$ Å$^2$.

These GO nanosheets are then arranged in space to construct the layered GO membrane as shown in Figure 4.1a. The parameters **D**, **W** and **H** shown in Figure 4.1a are termed as pore width, pore offset distance and interlayer distance respectively. In the present study the value of **D**, **W** and **H** are 7.0 Å, 18.0 Å and 10.0 Å respectively [33] [11] [12] [40] [19]. The membrane is then hydrated in a water box of size $143 \times 94 \times 34$ Å$^3$ which contain 9400 water molecules (Figure 4.1b). The water molecules are modeled with TIP3P water model. The water molecules within 2 Å distance of GO nanosheets are removed. This hydrated membrane is then energy minimized and subsequently equilibrated for 5 $ns$ at a temperature of 300 $K$ and 1 $atm$.

### 3.2.2 Foulant solution

The model foulant considered in the present study is lysozyme. The initial structure of lysozyme is obtained (downloaded) from the crystal structure 1AKI from protein data bank (https://www.rcsb.org/) (Figure 3.4). The net charge of lysozyme is $+8e$ which corresponds to its experimental condition at pH 7.0. In the lysozyme structure; histidine (His), arginine (Arg), and lysine (Lys) were protonated, whereas glutamate (Glu) and aspartate (Asp) were deprotonated.

The foulant solution is prepared (modeled) by placing two lysozyme proteins in NaCl solution box of size $143 \times 94 \times 68$ Å$^3$ as shown in Figure 3.5. This NaCl solution contains 27800 water molecules, 297 $Na^+$ ions and 297 $Cl^-$ ions which corresponds to a 0.56 $M$ NaCl solution which has a osmotic pressure of 27 $atm$ equivalent to sea water osmotic pressure. Since each protein has a net charge of $+8e$, additional

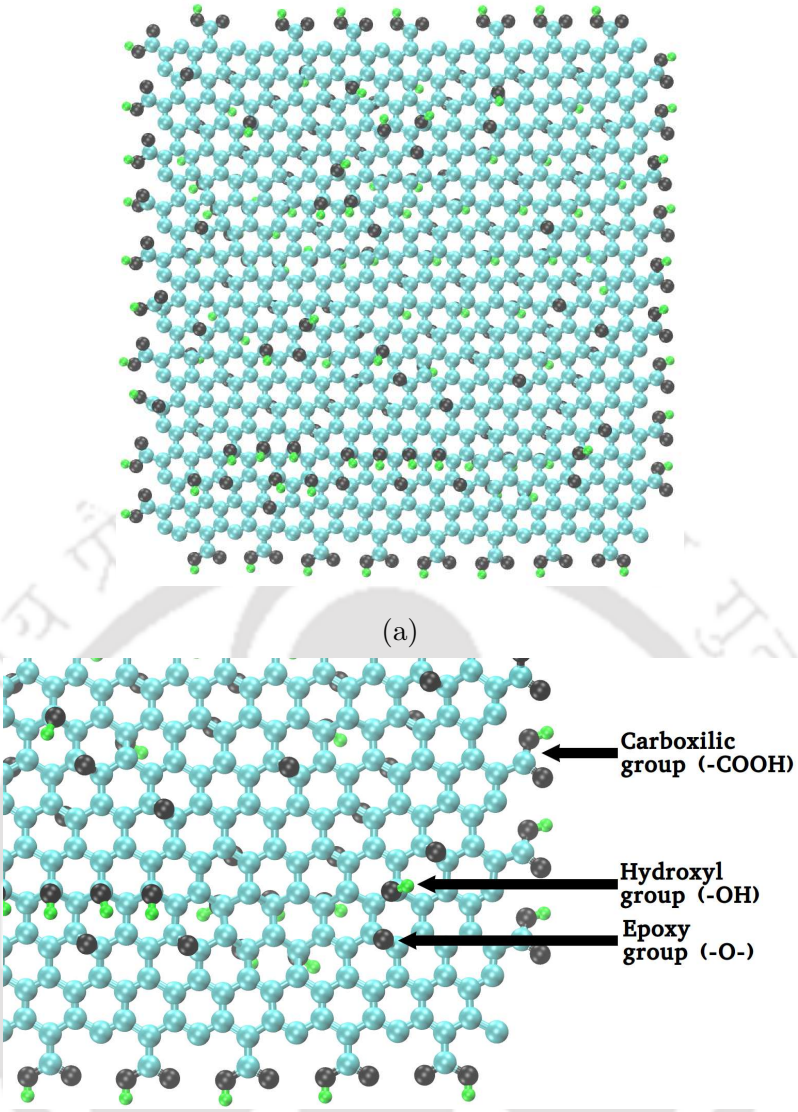
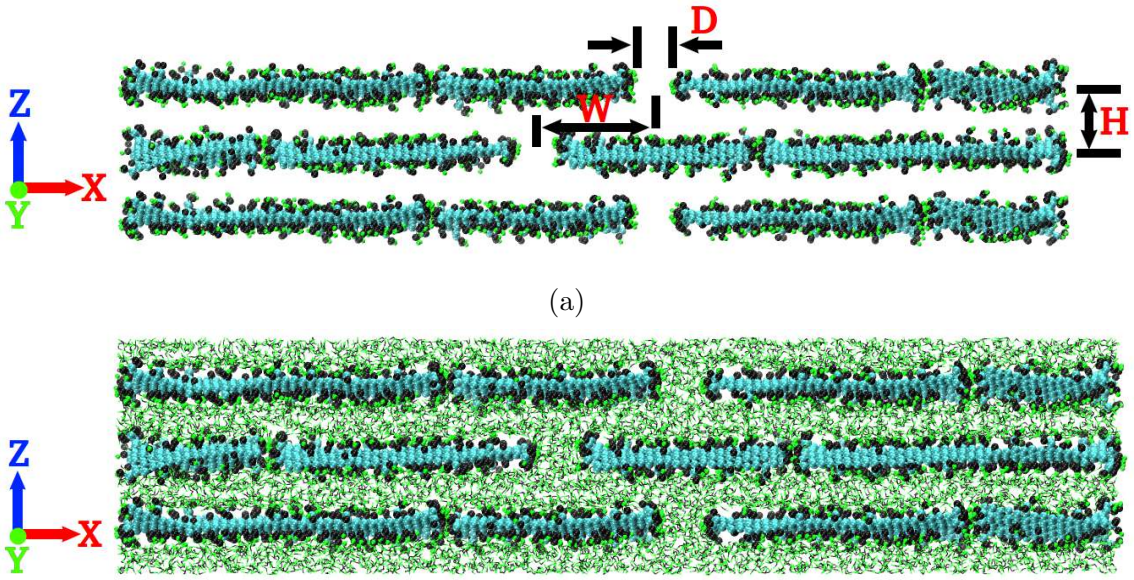



Figure 3.2: (a) GO nanosheet. (b) The distribution of functional groups on GO nanosheet. The green color is for hydrogen atoms, the black color is for oxygen atoms and the cyan color is for carbon atoms.

16 $Cl^-$ ions are added to get a charge neutral system. The initial distance between the two lysozymes is 15 Å along $X$ direction (Figure 3.5). This foulant solution is then energy minimized to remove any energetically unfavorable contacts within the system followed by an equilibration run of 100 $ps$ at a temperature of 300 $K$ and 1 $atm$ pressure. During this equilibration run the protein atoms were fixed to their




(a)




(b)

Figure 3.3: (a) Layered GO membrane. (b) Hydrated layered GO membrane. The green color is for hydrogen atoms, the black color is for oxygen atoms and the cyan color is for carbon atoms.

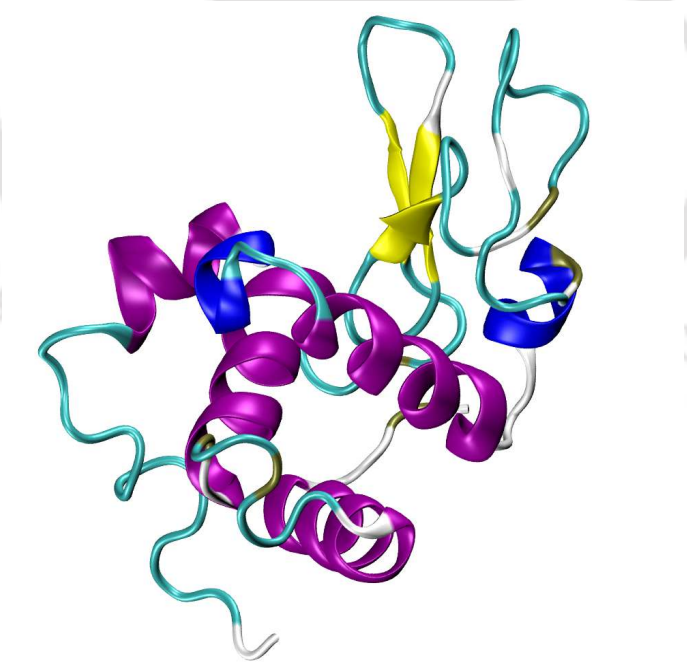


Figure 3.4: Structure of lysozyme (https://www.rcsb.org/).

initial positions. Subsequently, the protein atoms were relaxed and the system is equilibrated for another 3 $ns$ at a temperature of 300 $K$ ans 1 $atm$ pressure.

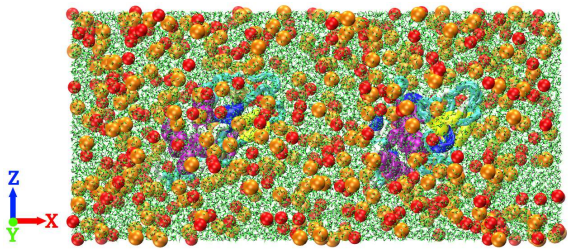



Figure 3.5: Foulant solution. The green color is for hydrogen atoms, the black color is for oxygen atoms, red color is for $Na^+$ ions and orange color is for $Cl^-$ ions.

### 3.2.3 Draw solution

One of the important component of a forward osmosis (FO) process is the draw solution. In the present study a solution of 3 $M$ $MgCl_2$ and 0.05 $M$ $Al_2(SO_4)_3$ is considered as the draw solution (Figure 3.6) which has a osmotic pressure of 228 $atm$. It contains 30000 water molecules, 1783 $Mg^{2+}$ ions, 3566 $Cl^-$ ions, 60 $Al^{3+}$ ions and 90 $SO_4^{2-}$ ions.

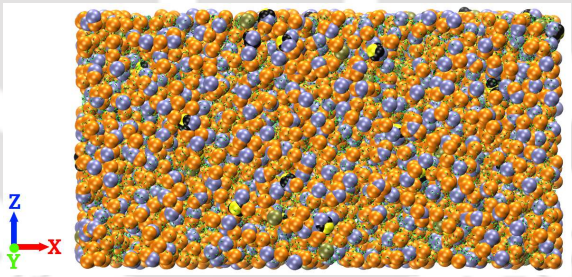



Figure 3.6: Draw solution. The green color is for hydrogen atoms, the black color is for oxygen atoms, red color is for $Na^+$ ions, orange color is for $Cl^-$ ions, iceblue color is for $Mg^{2+}$ ions, tan color is for $Al^{3+}$ ions and yellow color is for sulfur atoms.

### 3.2.4 Simulation system and Methodology

After the construction (modeling) of the hydrated layered GO membrane, foulant solution and the draw solution the simulation system for forward osmosis (FO) and

reverse osmosis (RO) are prepared.

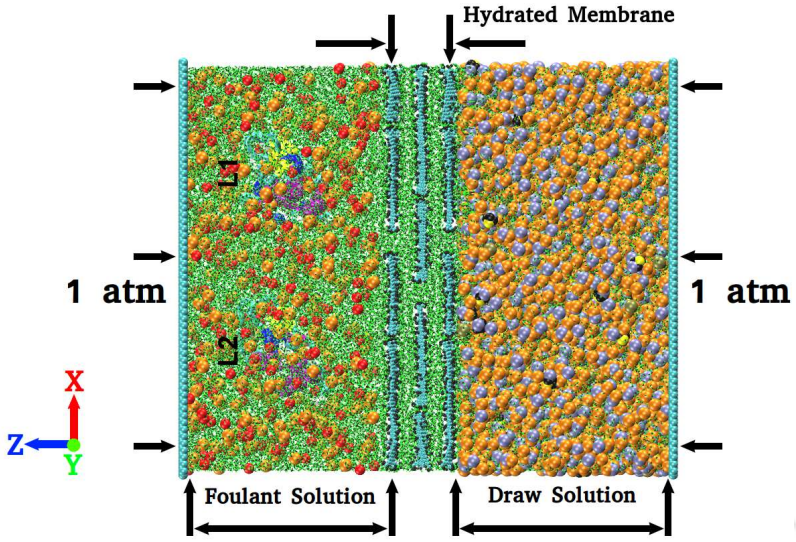



Figure 3.7: Simulation setup for FO process. The green color is for hydrogen atoms, black color is for oxygen atoms, cyan color is for carbon atoms, red color is for $Na^+$ ions, orange color is for $Cl^-$ ions, iceblue color is for $Mg^{2+}$ ions, tan color is for $Al^{3+}$ ions and yellow color is for sulfur atoms.

The simulation system for FO process is constructed by placing the hydrated layered GO membrane between the foulant solution and the draw solution as shown in Figure 3.7. The DS has a very high salt concentration(3M) in comparison to the FS(0.56M), as mentioned in the previous sections 3.2.3 and 5.1 respectively. The high difference in the concentration between FS and DS creates an osmotic pressure and facilitates a natural movement of water molecules from FS to DS. To avoid the lateral movement of the membrane, the carbon atoms of the basal plane of all the GO sheets were constrained in the XY direction. At the edges (along $Z$ direction) of the FS and DS a graphene sheet is placed which will serve two purpose (1) act as a piston to apply pressure. (2) prevent the mixing of FS and DS due to periodic boundary condition. In both the pistons a pressure of 1 $atm$ is applied in opposite directions (along the permeating or $Z$ direction) to mimic the external atmospheric condition in an FO system, during production simulation run as depicted in Figure 3.7. The external pressure applied on the graphene sheets is employed using TclForces module in NAMD[34].

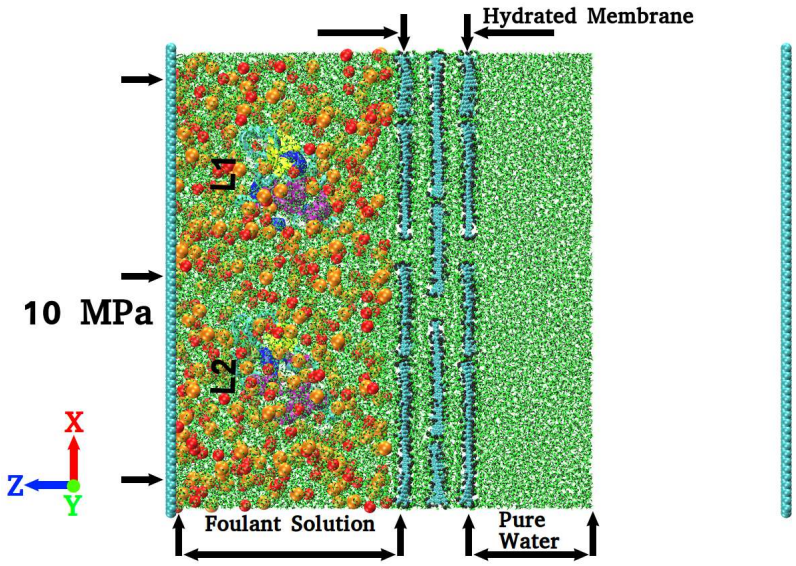



Figure 3.8: Simulation setup for RO process. The green color is for hydrogen atoms, black color is for oxygen atoms, cyan color is for carbon atoms, red color is for $Na^+$ ions, orange color is for $Cl^-$ ions.

For the RO process, the simulation system is constructed by placing the hydrated membrane between the foulant solution and pure water, as shown in Figure 3.8. A single sheet of graphene is placed just outside (along $Z$ direction) the foulant solution, which will act as a piston. Another graphene sheet is placed below the pure water box at a distance of 30 Å (along $Z$ direction), which will be fixed to its initial position during the simulation. The transmembrane pressure on the FS is generated by applying an external force(using the TclForces facility in NAMD) on each atom of the graphene piston along the $Z$ direction. Consequently, the graphene piston pushes the FS toward the membrane. Consistent with the practical RO process, a pressure of 10 $MPa$ is applied on the foulant solution through the graphene piston along the $-Z$ direction (Figure 3.8). In contrast, most of the previous atomistic simulation on the RO process is carried out with very high transmembrane pressure (up to 1000 of $MPa$).

After the construction of the simulation systems (both FO and RO) the systems are first energy minimized and then equilibrated for 500 $ps$ at a temperature of 300 $K$ and 1 $atm$ pressure. Then the production runs are performed for 20 $ns$ at a temperature of 300 $K$. For both FO and RO a vacuum of 50 Å is applied along the

$+Z$ and $-Z$ direction to eliminate the interaction between the simulation box and its periodic image along the $Z$ direction.

The simulations in the present study are performed with NAMD [34] package. All the atomic interactions are computed using OPPLS-AA force field parameters [22]. The bond length of the water molecules (TIP3P water model) are constrained using SETTLE algorithm [32]. The non-bonded interactions are composed of van der Waals interactions and electrostatic interactions. The van der Waals interactions are computed using Lennard-Jones potential with 12.0 Å cut off distance. Particle mesh Ewald (PME) method [9] is used for long range electrostatic interactions. A time step of 1 $fs$ is used for all the simulations reported in the present study. For controlling the temperature Langevin dynamics is used with a damping factor of 5 $ps^{-1}$. Pressure is controlled by modified Nosé-Hoover method [29] [10] with 0.3 $ps$ barostat oscillation time and damping factor.

## 3.3 Results and discussion

MD simulations of explicitly solvated systems were carried out to assess the interaction between lysozyme and the GO membrane and between two lysozyme molecules, and also to quantify the effect of solvent and ions on those interactions. Also, to explore the impact of protein orientation on the fouling mechanism, three different initial setups were generated by rotating the protein relative to the membrane surface as depicted in Fig 3.1.

For a qualitative assessment of the protein–membrane interaction, we first inspected the closest distance between protein and the surface of the layered GO membrane. As shown in Figure 3.7 and 3.8 we refer the two lysozyme proteins as L1 and L2. Figure 3.9a shows the time evolution of distance between protein and the membrane surface in RO and FO. The distance indicated here is the closest separation between the protein and GO at any instant. From Figure 3.9a, we observe that L2 in RO consistently moved towards the membrane surface from the very starting of the simulation and at around 8 ns L2 directly interacts with the surface, which can be inferred as an adsorption event. Also in case of FO, L2 diffuses towards the membrane surface initially but again went back to the bulk. Even L1 in RO, made

an unsuccessful attempt to reach the surface. Distance for L1 in case of FO was initially fluctuating around the initial position, but as the simulation progresses, it started moving away from the membrane surface. However, in the case of the other two orientations, an adsorption event is not observed as can be seen from Figure (3.9b and 3.9c), yet the nature of distance evolution is somewhat similar to O1, in the sense that in both the configuration there is a decrease in the distance for RO. Only in the case of L2–O2 for RO, there is diffusion away from the membrane surface, which could have resulted from an interplay of protein–protein interaction and orientation, since as can be observed the distance analysis for L2-RO in O2 is closely correlated with the corresponding protein–protein distance (Fig 3.10). Overall, it was found that in almost all three cases lysozyme molecules shows a tendency to move towards the membrane surface in RO, while in FO the molecules usually remain in bulk. However, the analysis also emphasizes that evolution of distance may not be sufficient to characterize the fouling behavior and inspection of interactions among the two lysozyme molecules can give a clearer picture. Hence we delve into protein–protein interactions in the following discussions.

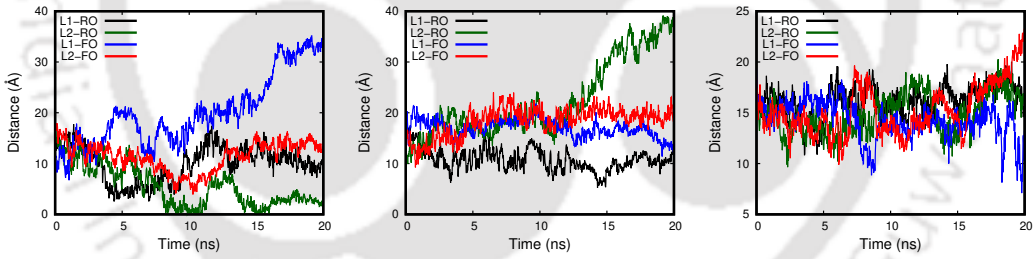
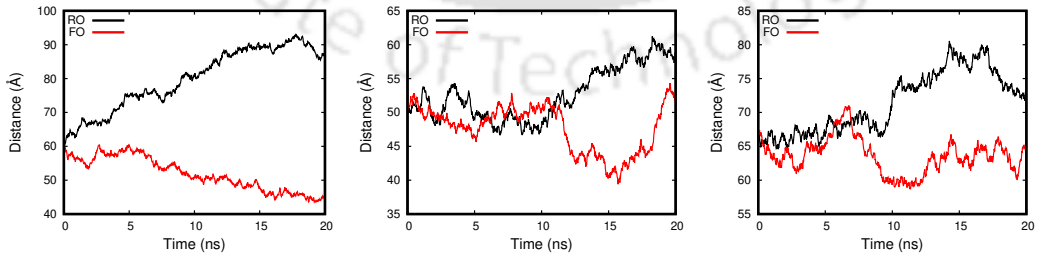



Figure 3.9: Time evolution of closest distance between membrane surface and protein for orientation (a) O1 (b) O2 (c) O3

It is a general consensus in membrane fouling research that foulant-foulant interactions are significant to understand the fouling mechanism and to characterize antifouling performance of a membrane [1]. So to gain insight into the kind of protein-protein interaction, we computed the time evolution of distance between two lysozyme monomers. The distance here is calculated as the separation between the centers of the two proteins at any instant($d_{pp}$). The initial distance between

the centers was 60 Å. From Figure 3.10a, it is observed that for RO,the changes are much steeper and are increasing, while for FO, there is a decrease, but the descent is not as sharp. The distance at the end of the simulation in case of RO is at around 85 Å and that in case of FO it is about 45 Å.

Moreover, we found almost similar trends for $d_{pp}$ evolution for O2 and O3. As shown in Fig 3.10b and 3.10c, the $d_{pp}$ for RO and FO is increasing and decreasing respectively, although in contrast to O1, these differences in O2 and O3 evolved much later in the simulation. The implication of the above result could be that in FO, proteins are more prone to agglomeration since they are coming close to each other, while in the case of RO proteins prefer to stay away from each other. The above result can also be interpreted as a microscopic measure of the structure of the fouling layer in FO and RO. Since proteins tend to remain closer to each other in FO, it implies a thicker fouling layer in FO. On the other hand, in RO, the proteins are going away, so the foulant layer is thinner than FO. The above result corroborates well with the findings of many experimental studies [26, 38, 47] which reports a higher foulant density in FO than RO. The reason for this observation could be that since in RO there is a hydraulic pressure being applied, it pushes water molecules and ions and they diffuse in between the two proteins because of which they move away from each other, while in FO there is not much change in the distance between the proteins, since there is no hydraulic pressure in FO. The effect of solvent and ions on protein–protein interactions is discussed in detail in the analysis of radial distribution function ahead.




(a) (b) (c)

Figure 3.10: Time evolution of closest distance between the two lysozyme protein molecules for orientation (a) O1 (b) O2 (c) O3

We analyzed the orientation evolution of protein as the change in the angle between protein and the membrane surface as a function of time. The orientation is quantified as the angle ($\theta$) between the vector formed by joining $C_\alpha$ atoms of residues Arg 128 and Arg 5 and the plane of membrane's surface. Figure 3.11a shows the changes in the angle for proteins(L1,L2) for FO and RO during the simulation. Interestingly we can see a very close correlation between protein-surface distance and rotation of protein, as shown in the Figure 3.11a, $\theta$ for L2-RO continuously approaches 0 (parallel orientation relative to the surface) after around 9 ns, and simultaneously the distance between L2 and the membrane also reduces (Fig 3.9a). Although L2 in FO also approaches a parallel orientation, it is not able to hold, which again correlates well with distance data, since as L2 comes near to the surface $\theta$ also decreases and gets converged for some time, but as L2 diffuses back to the bulk again, $\theta$ also increases simultaneously. Similarly, Figure 3.11b shows that in the case of O2 also, there is a strong correlation between distance and orientation evolutions.

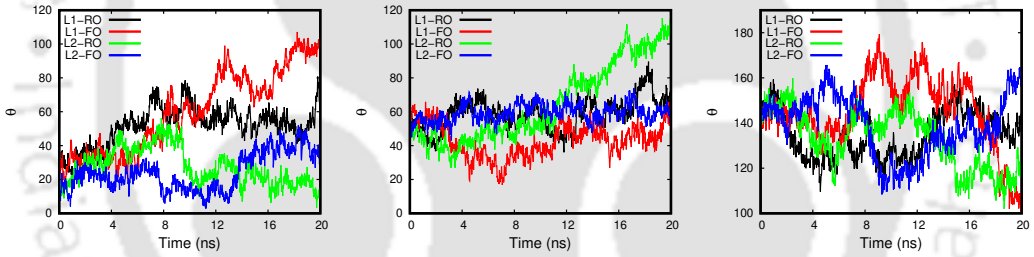



Figure 3.11: Time evolution of orientation of protein with respect to membrane surface for (a) O1 (b) O2 (c) O3

But in the case of O3, as shown in Figure 3.11c, the evolution of orientation is highly chaotic, a probable reason could be that since in O3 the maximum area of protein is exposed to solvent permeation among all the three systems, it gives rise to fluctuations in orientation.

To get a more detailed insight on the interactions between proteins and the layered GO membrane surface, we evaluate time evolution of protein-surface ($E_{ps}$) and protein-protein ($E_{pp}$) interaction energy. As shown in Figure 3.12a the coulombic interaction between protein-surface are repulsive and on an average remains un-

changed for both the proteins. However, only in the case of L2-RO, we observe van der Waals interactions during the last 5 ns of the simulation, since only L2 diffuses quite close to the surface. For all other cases, no van der Waals interaction is observed. Although there is a repulsive interaction between protein and surface, still L2 diffuses towards the membrane surface, the reason could be that since the charges on lysozyme and GO are not uniformly distributed, the electrostatic interactions are dominated by hydraulic pressure which drives the protein towards the membrane surface. Although L1 also diffuses near to the membrane as observed in Figure 3.9a, it failed to find a favorable interaction site; hence, we don't observe a strong interaction between L1 and GO. The mechanism of L2–GO interaction is further discussed along with key interacting residues ahead. As shown in rest of the plots in Figure 3.14, the interaction energies for O2 and O3 have a similar trend to that of O1.

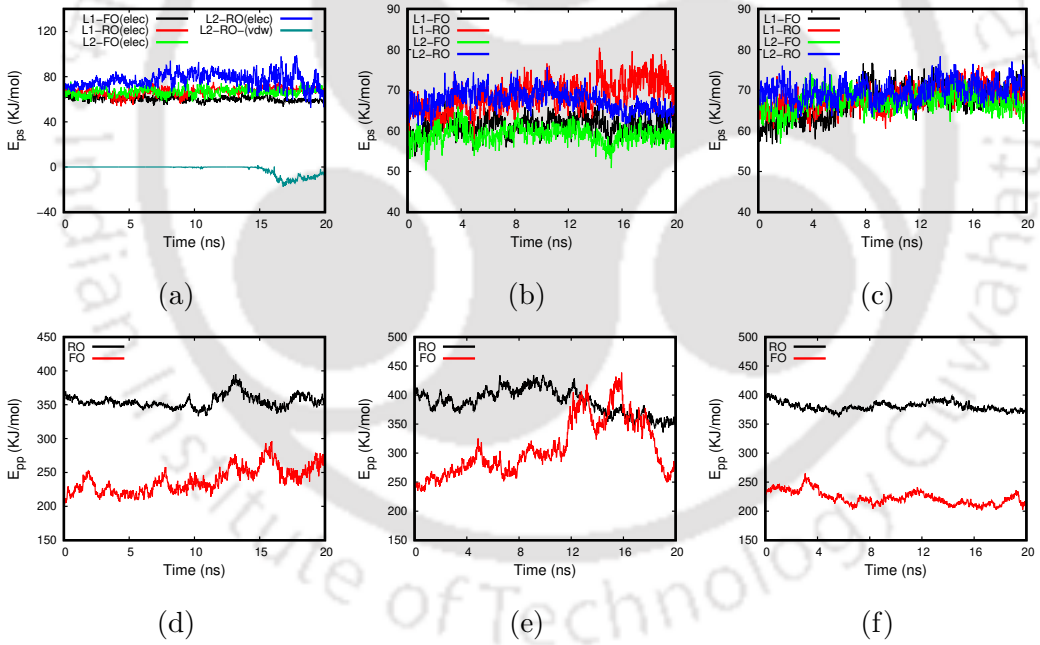



Figure 3.12: Time evolution of interaction energy between membrane and protein for (a) O1 (b) O2 (c) O3. Time evolution of interaction energy between proteins for (c) O1 (d) O2 (e) O3.

Figure 3.12d shows temporal evolution of the interaction energy between L1 and L2. here is no van der Waals interaction observed between the two proteins,

and they are interacting only through Coulombic interaction. The nature of the interaction between proteins is repulsive and is stronger in the case of RO. Another important observation is that the interaction between protein molecules is much stronger than the protein-surface interaction, which emphasizes the crucial role of foulant-foulant interactions in determining a mechanism for membrane fouling. Also, there is a decrease in interaction energy between the proteins for RO because there is an increase in the distance between them as can be seen in the distance evolution (Figure 3.10a). Although there is a repulsive Coulombic interaction between L1 and L2 for FO, according to distance analysis the distance between the proteins is decreasing, which suggests that the dynamics of proteins may not be understood solely based on interaction energy analysis. In further discussions, we show that this is indeed true, with the interaction between protein-solvent and protein-ions playing a crucial role in interpreting the dynamics of proteins.

To demonstrate the important role of solvent and ions in the dynamics of proteins and to elucidate their effect on the fouling layer structure in FO and RO, we investigated the radial distribution function (g(r)) for protein–solvent and protein–ions. We computed g(r) for protein–solvent with non-backbone polar atoms (nitrogen, oxygen) of protein and oxygen of water and similarly for protein–ions, with ions (chloride, sodium) present in the solution. Here, we reported g(r) only for L1, since g(r) for L2 (not shown) is similar to L1 for all the cases. Interestingly, we found that in all the cases (Figure 3.13a-c) g(r) for RO has much higher intensity peaks than FO. Figure 3.13a shows g(r) between protein and water; as can be observed, RO exhibits peaks with much higher intensity than FO, which demonstrates that there is a very high density of water molecules around the protein in RO than FO. Also, we observe that, although nitrogen is having a less density of water around it than oxygen, it has a second distinctive peak at around 5 Å, which indicates that nitrogen has a second hydration shell with water. Interestingly g(r) for ions shows similar trends as that of water. Figure 3.13b and c shows g(r) for oxygen (protein)–ions and nitrogen (protein)–ions respectively; again RO exhibits a much higher density of ions for both nitrogen and oxygen than FO. Specifically, nitrogen shows a very high density of chloride ions in the case of RO, while with sodium ions, nitrogen has almost negligible interaction in both RO and FO.

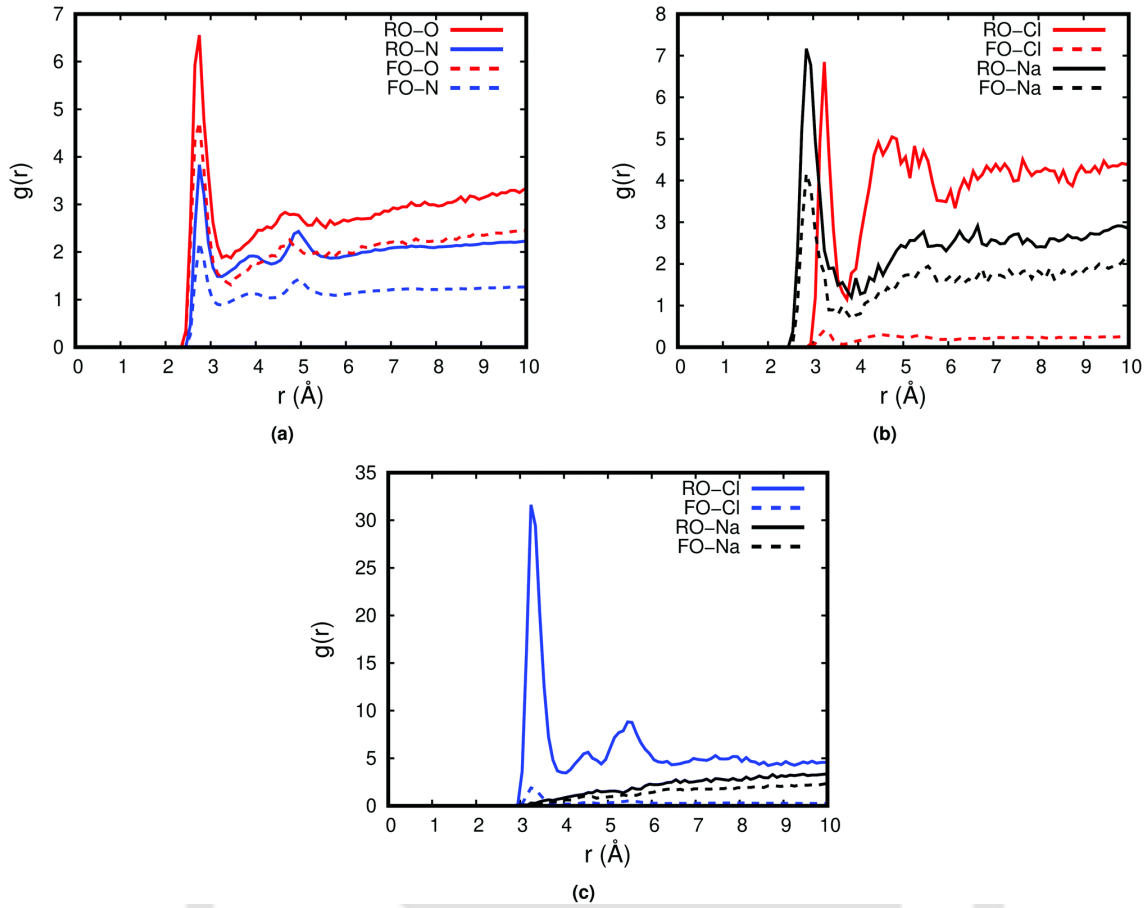



Figure 3.13: RDF between (a) oxygen (L1)–oxygen (water) and nitrogen (L1)–oxygen (water) shown in red and blue color respectively, (b) oxygen (L1)–chloride (ions) and oxygen (L1)–sodium (ions) shown in red and black respectively, and (c) nitrogen (L1)–chloride (ions) and nitrogen (L1)–sodium (ions) shown in blue and black color respectively in RO (solid lines) and FO (dashed lines).

The g(r) analysis reveals the presence of hydration repulsion force acting between protein–protein and protein–surface. Hydration repulsion is the force acting between well solvated and ionized surfaces; it is a fundamental force, common in biological systems and solution chemistry([20], [31], [15]. As discussed in protein–protein distance analysis, due to the presence of hydraulic pressure in RO, more number of water molecules and ions diffuse towards the proteins, which in turn creates a high density region of water molecules and ions around the proteins, due to which they experience much stronger repulsion, and hence, this explains, why proteins move away from each other in RO. Similarly, the proteins experience hydration repulsion

from the membrane surface also, but in RO, the hydraulic pressure compensates for it, while in FO there is no compressing force to balance this repulsion. Furthermore, as observed, there is also a very high density of ions around the proteins, particularly chloride ions, because of which there is a screening of electrostatic interaction between proteins and the membrane surface, and also the density of chloride ions around both proteins adds to the repulsion between them. Overall, g(r) analysis indicates that fundamental forces like hydration repulsion and charge screening can lead to the distinction in the fouling mechanism for FO and RO.

In order to know which residues are directly interacting with the membrane surface, the distance of each residue at different time intervals is monitored. The distance is calculated as the separation between the $C_\alpha$ atom of a residue and the closest atom of GO. Figure 3.14a shows the distance of residues for L1 in RO at different times of simulation. As can be noticed, there are not much changes for L1 in RO, but for FO, an increase in the distance between residues of L1 and GO can be observed, particularly in both terminal regions (C terminal and N terminal). For L2 in FO (Figure 3.14d), the distances remain almost the same throughout the trajectory, except for some minor changes in the terminal regions. Adsorption takes place only for L2 in RO; as can be observed in Figure 3.14b, the residues directly involved in the interaction with GO (Arg 5, Arg 114, Arg 125, and Arg 128) are very close to the surface at the end of the simulation. As can be observed, all the key interacting residues are basic in nature (Figure 3.15).

Figure 3.16 shows the visual representation of a probable mechanism of protein adsorption. The protein initially diffused towards the region where the edges of two GO sheets are close enough since the presence of the carboxyl group at the GO sheet's edges creates a high oxygen density region. The residues Arg 128 and Arg 5 acted as an anchor and pulled the whole protein (L2) towards the GO surface. The anchoring residues (Arg 5 and Arg 128) belong to the N terminal (Arg 5) and C terminal (Arg 128) regions which were also found to be the regions responsible for adsorption in a study by Kubiak et al.[15] The above explanation also implies that as the protein diffuses with its positively charged residues facing towards the oxygen-rich (negatively charged) region of the GO surface, this triggers an electrostatic interaction between these residues and functional groups of the GO

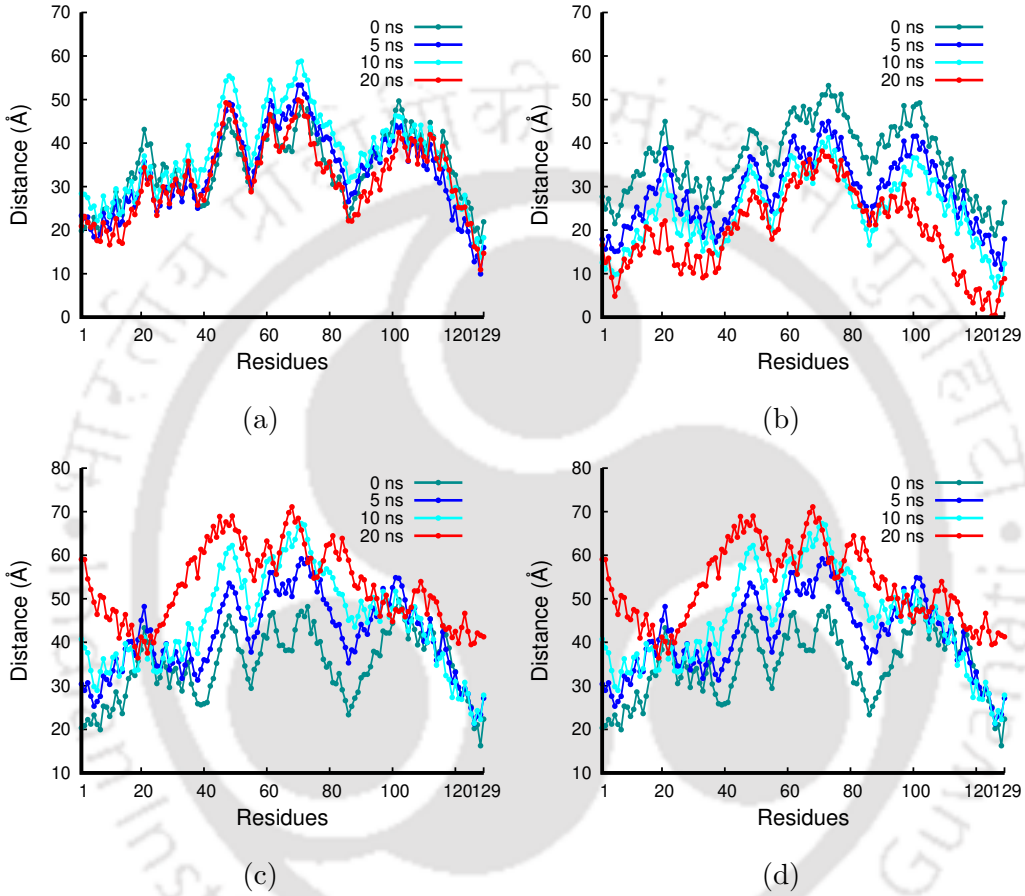
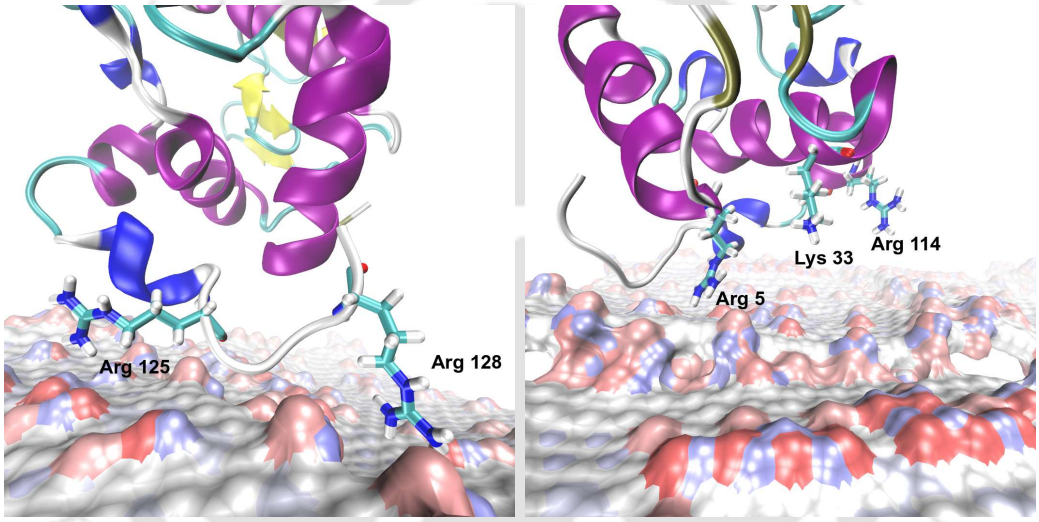



Figure 3.14: Closest distance between the alpha carbon of protein's residues and the membrane surface at 0 ns (cyan), 5 ns (blue), 10 ns (light blue) and 20 ns (red) for L1 (a) and L2 (b) in RO and L1 (c) and L2 (d) in FO.

surface. The electrostatic attraction between interacting residues and functional groups steers the whole protein towards the GO surface. Interestingly, although the electrostatic interaction energy shown in Figure 3.12a shows a positive (repulsive) energy, still the diffusion of protein can be observed, the reason could be that, since the charge density on both lysozyme and GO is not uniform, there is a possibility that positively charged residues of lysozyme can interact with GO at a region where negative charge density is high. So the interaction between lysozyme and GO must be electrostatically driven, which was also found in an experimental study by Smith et al.[39]. Moreover, Kubiak et al.[23] in their MD study showed how a negatively charged protein can get adsorbed onto a negatively charged surface.




(a) (b)

Figure 3.15: Final snapshot of the key interacting residues shown in licorice representation and the membrane surface shown in surf representation.

## 3.4 Conclusion

We have performed MD simulations to study foulant-membrane interactions and to gain molecular insight into the mechanism of fouling in FO and RO. We demonstrated that fouling is a result of a complex interplay between hydraulic pressure, protein-solvent interactions, and protein–membrane interactions. Our results suggested that interfacial water on the membrane's surface and hydration shell of the

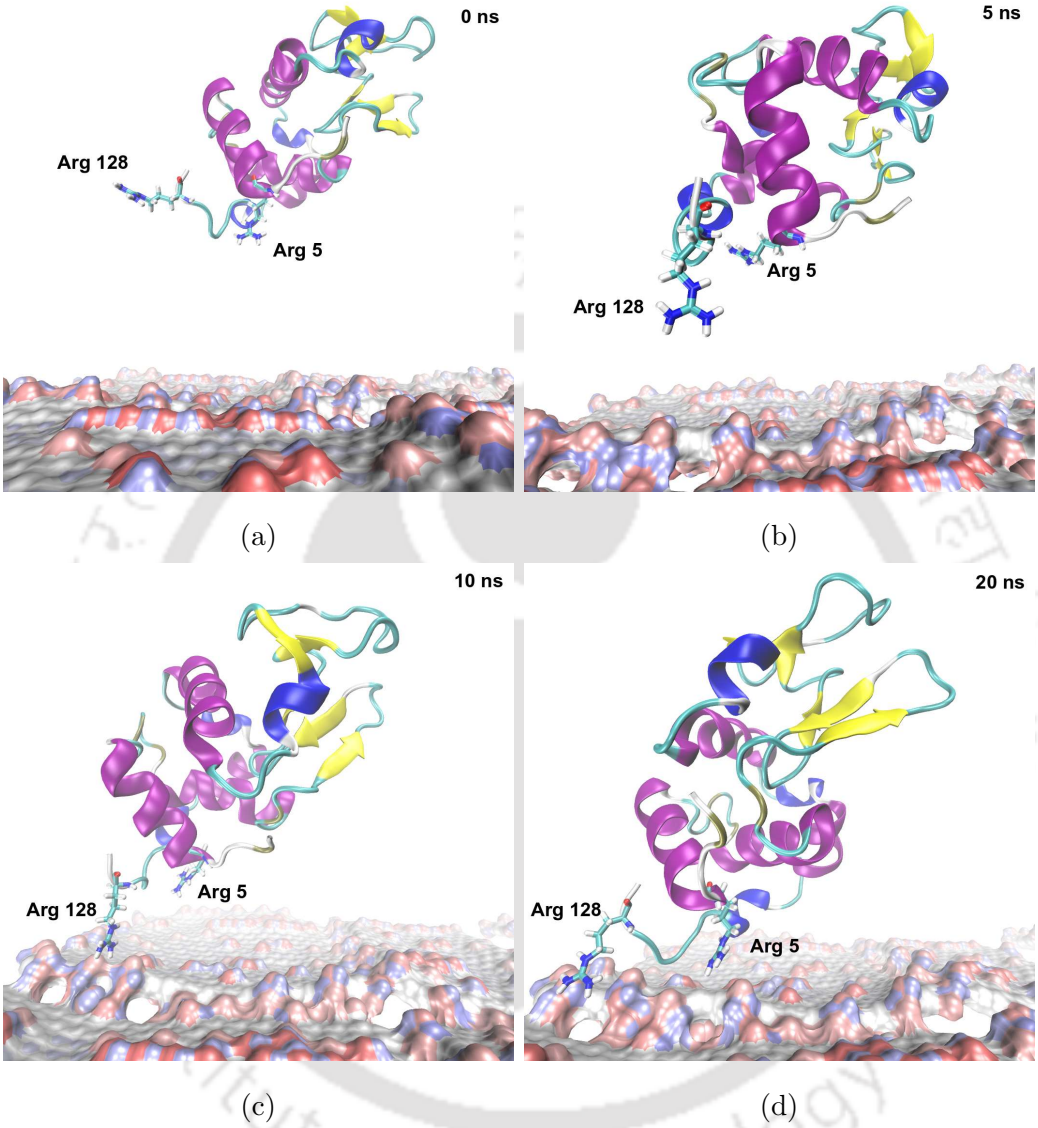



Figure 3.16: Snapshots of L2 shown at different simulation times and the key interacting residues Arg 128 and Arg 5.The membrane layer is represented in surface representation, the key interacting residues are shown in licorice representation and the rest of the protein is shown in New Cartoon representation.

protein plays a significant role in determining the fouling mechanism. In particular, hydraulic pressure in RO leads to a much higher density of water molecules and ions around proteins and on the membrane surface than in FO, which in turn leads to much higher hydration repulsion force in the case of RO. Due to the increased hydration repulsion in RO, proteins move away from each other, which implies a thinner foulant density in RO. On the other hand in FO, due to the lesser ion and water density than RO, the hydration repulsion is also much less between the two proteins. Hence, the distance between the two proteins does not change much, which implies the agglomeration of proteins in FO. Also, due to the lack of hydraulic pressure in the case of FO, proteins are not convected towards the membrane surface as strongly as in the case of RO, which makes the fouling layer loose in FO. Moreover, we also found O1 to be the most favorable orientation in terms of protein-membrane interaction, as only in O1, we observed an adsorption event between GO and lysozyme within the simulated time. A probable mechanism of the adsorption could be that the protein is transported near to the membrane surface due to hydraulic pressure, and protein diffuses towards the oxygen-rich region on the surface, which triggers Coulombic interactions between basic residues of protein facing the surface and functional groups containing oxygen of GO. The key interacting residues which lead to the adsorption of lysozyme are Arg 128, Arg 5, Arg 125, Arg 114 (arginine) and Lys 33 (lysine). The present work provides key molecular insights into the structure of the fouling layer for FO and RO and can help in the development of high-performance antifouling membranes.

# Bibliography

[1] Wui Seng Ang and Menachem Elimelech. Protein (bsa) fouling of reverse osmosis membranes: implications for wastewater reclamation. *Journal of Membrane Science*, 296(1-2):83–92, 2007.

[2] Langming Bai, Fangshu Qu, Heng Liang, Jun Ma, Haiqing Chang, Meilian Wang, and Guibai Li. Membrane fouling during ultrafiltration (uf) of surface water: Effects of sludge discharge interval (sdi). *Desalination*, 319:18–24, 2013.

[3] Chanhee Boo, Menachem Elimelech, and Seungkwan Hong. Fouling control

in a forward osmosis process integrating seawater desalination and wastewater reclamation. *Journal of Membrane Science*, 444:148–156, 2013.

[4] Tzahi Y Cath, Amy E Childress, and Menachem Elimelech. Forward osmosis: principles, applications, and recent developments. *Journal of membrane science*, 281(1-2):70–87, 2006.

[5] David Cohen-Tanugi and Jeffrey C Grossman. Water permeability of nanoporous graphene at realistic pressures for reverse osmosis desalination. *The Journal of chemical physics*, 141(7):074704, 2014.

[6] David Cohen-Tanugi, Li-Chiang Lin, and Jeffrey C Grossman. Multilayer nanoporous graphene membranes for water desalination. *Nano letters*, 16(2):1027–1033, 2016.

[7] Minxia Ding, Anthony Szymczyk, Florent Goujon, Armand Soldera, and Aziz Ghoufi. Structure and dynamics of water confined in a polyamide reverse-osmosis membrane: A molecular-simulation study. *Journal of membrane science*, 458:236–244, 2014.

[8] Magda I Dova, Konstantinos B Petrotos, and Harris N Lazarides. On the direct osmotic concentration of liquid foods. part i: Impact of process parameters on process performance. *Journal of Food Engineering*, 78(2):422–430, 2007.

[9] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, , and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

[10] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.*, 103(11):4613–4621, 1995.

[11] A. Gogoi, T. J. Konch, K. Raidongia, and K. A. Reddy. Water and salt dynamics in multilayer graphene oxide (GO) membrane: Role of lateral sheet dimensions. *J. Membr. Sci.*, 563:785–793, 2018.

[12] A. Gogoi, K. A. Reddy, and P. Mondal. Multilayer graphene oxide membrane in forward osmosis: Molecular insights. *ACS Appl. Nano Mater.*, 1(9):4450–4460, 2018.

[13] Yangshuo Gu, Yi-Ning Wang, Jing Wei, and Chuyang Y Tang. Organic fouling of thin-film composite polyamide and cellulose triacetate forward osmosis

membranes by oppositely charged macromolecules. *Water research*, 47(5):1867–1874, 2013.

[14] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4:17:1–17, 2012.

[15] Yi He, Yung Chang, Jason C Hower, Jie Zheng, Shengfu Chen, and Shaoyi Jiang. Origin of repulsive force and structure/dynamics of interfacial water in oeg–protein interactions: a molecular simulation study. *Physical Chemistry Chemical Physics*, 10(36):5539–5544, 2008.

[16] Zak E Hughes and Julian D Gale. A computational investigation of the properties of a reverse osmosis membrane. *Journal of Materials Chemistry*, 20(36):7788–7799, 2010.

[17] Zak E Hughes and Julian D Gale. Molecular dynamics simulations of the interactions of potential foulant molecules and a reverse osmosis membrane. *Journal of Materials Chemistry*, 22(1):175–184, 2012.

[18] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[19] W. S. Hung, Q. F. An, M. De Guzman, H. Y. Lin, S. H. Huang, W. R. Liu, C. C. Hu, K. R. Lee, and J. Y. Lai. Pressure-assisted self-assembly technique for fabricating composite membranes consisting of highly ordered selective laminate layers of amphiphilic graphene oxide. *Carbon*, 68:670–677, 2014.

[20] Jacob Israelachvili and Håkan Wennerström. Role of hydration and water structure in biological and colloidal interactions. *Nature*, 379(6562):219–225, 1996.

[21] Yongsun Jang, Hyeongrak Cho, Yonghyun Shin, Yongjun Choi, Sangho Lee, and Jaewuk Koo. Comparison of fouling propensity and physical cleaning effect in forward osmosis, reverse osmosis, and membrane distillation. *Desalination and Water Treatment*, 57(51):24532–24541, 2016.

[22] W. L. Jorgensen, D. S. Maxwell, and J. T. Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[23] Karina Kubiak-Ossowska, Barbara Jachimska, and Paul A Mulheran. How negatively charged proteins adsorb to negatively charged surfaces: A molecular

dynamics study of bsa adsorption on silica. *The Journal of Physical Chemistry B*, 120(40):10463–10468, 2016.

[24] Winson CL Lay, Tzyy Haur Chong, Chuyang Y Tang, Anthony G Fane, Jinsong Zhang, and Yu Liu. Fouling propensity of forward osmosis: investigation of the slower flux decline phenomenon. *Water science and technology*, 61(4):927–936, 2010.

[25] Pierre Le-Clech, Vicki Chen, and Tony AG Fane. Fouling in membrane bioreactors used in wastewater treatment. *Journal of membrane science*, 284(1-2):17–53, 2006.

[26] Sangyoup Lee, Chanhee Boo, Menachem Elimelech, and Seungkwan Hong. Comparison of fouling behavior in forward osmosis (fo) and reverse osmosis (ro). *Journal of Membrane Science*, 365(1-2):34–39, 2010.

[27] Bo Liu, Adrian Wing-Keung Law, and Kun Zhou. Strained single-layer c2n membrane for efficient seawater desalination via forward osmosis: A molecular dynamics study. *Journal of Membrane Science*, 550:554–562, 2018.

[28] Kerusha Lutchmiah, ARD Verliefde, Kees Roest, Luuk C Rietveld, and ER Cornelissen. Forward osmosis for application in wastewater treatment: a review. *Water research*, 58:179–197, 2014.

[29] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5):4177–4189, 1994.

[30] Baoxia Mi and Menachem Elimelech. Organic fouling of forward osmosis membranes: fouling reversibility and cleaning without chemical reagents. *Journal of membrane science*, 348(1-2):337–345, 2010.

[31] Rui Miao, Lei Wang, Na Mi, Zhe Gao, Tingting Liu, Yongtao Lv, Xudong Wang, Xiaorong Meng, and Yongzhe Yang. Enhancement and mitigation mechanisms of protein fouling of ultrafiltration membranes under different ionic strengths. *Environmental science & technology*, 49(11):6574–6580, 2015.

[32] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.

[33] R. R. Nair, H. A. Wu, P. N. Jayaram, I. V. Grigorieva, and A. K. Geim. Unimpeded permeation of water through helium-leak-tight graphene-based membranes. *Science*, 335(6067):442–444, 2012.

[34] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D.Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with namd. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[35] Hamid Rezaei, Farzin Zokaee Ashtiani, and Amir Fouladitajar. Effects of operating parameters on fouling mechanism and membrane flux in cross-flow microfiltration of whey. *Desalination*, 274(1-3):262–271, 2011.

[36] Abdul Rajjak Shaikh, Hamed Karkhanechi, Tomohisa Yoshioka, Hideto Matsuyama, Hiromitsu Takaba, and Da-Ming Wang. Adsorption of bovine serum albumin on poly (vinylidene fluoride) surfaces in the presence of ions: A molecular dynamics simulation. *The Journal of Physical Chemistry B*, 122(6):1919–1928, 2018.

[37] C. J. Shih, S. Lin, R. Sharma, M. S. Strano, and D. Blankschtein. Understanding the ph-dependent behavior of graphene oxide aqueous solutions: A comparative experimental and molecular dynamics simulation study. *Langmuir*, 28(1):235–241, 2012.

[38] Farrukh Arsalan Siddiqui, Qianhong She, Anthony G Fane, and Robert W Field. Exploring the differences between forward osmosis and reverse osmosis fouling. *Journal of Membrane Science*, 565:241–253, 2018.

[39] Sean C Smith, Farid Ahmed, Krystal M Gutierrez, and Debora Frigi Rodrigues. A comparative study of lysozyme adsorption with graphene, graphene oxide, and single-walled carbon nanotubes: potential environmental applications. *Chemical Engineering Journal*, 240:147–154, 2014.

[40] P. Sun, K. Wang, J. Wei, M. Zhong, D. Wu, and H. Zhu. Effective recovery of acids from iron-based electrolytes using graphene oxide membrane filters. *J. Mater. Chem. A*, 2:7734–7737, 2014.

[41] H. Tang, D. Liu, Y. Zhao, X. Yang, J. Lu, and F. Cui. Molecular dynamics study of the aggregation process of graphene oxide in water. *J. Phys. Chem. C*, 119(47):26712–26718, 2015.

[42] Emily W Tow et al. Quantifying osmotic membrane fouling to enable comparisons across diverse processes. *Journal of Membrane Science*, 511:92–107, 2016.

[43] Jing Wang, Zhi Wang, Jixiao Wang, and Shichang Wang. Improving the water

flux and bio-fouling resistance of reverse osmosis (ro) membrane through surface modification by zwitterionic polymer. *Journal of Membrane Science*, 493:188–199, 2015.

[44] Yuan Xiang, Yaolin Liu, Baoxia Mi, and Yongsheng Leng. Hydrated polyamide membrane and its interaction with alginate: a molecular dynamics study. *Langmuir*, 29(37):11600–11608, 2013.

[45] Yuan Xiang, Yaolin Liu, Baoxia Mi, and Yongsheng Leng. Molecular dynamics simulations of polyamide membrane, calcium alginate gel, and their interactions in aqueous solution. *Langmuir*, 30(30):9098–9106, 2014.

[46] Kang Xiao, Xiaomao Wang, Xia Huang, T David Waite, and Xianghua Wen. Combined effect of membrane and foulant hydrophobicity and surface charge on adsorptive fouling during microfiltration. *Journal of Membrane Science*, 373(1-2):140–151, 2011.

[47] Ming Xie, Jongho Lee, Long D Nghiem, and Menachem Elimelech. Role of pressure in organic fouling in forward osmosis and reverse osmosis. *Journal of Membrane Science*, 493:748–754, 2015.

[48] Jun Xu, Zhi Wang, Jixiao Wang, and Shichang Wang. Positively charged aromatic polyamide reverse osmosis membrane with high anti-fouling property prepared by polyethylenimine grafting. *Desalination*, 365:398–406, 2015.

# Chapter 4

# Effect of Ionic Envrionment on Fouling

## 4.1 Introduction

The development of thin-film composite (TFC) membranes has revolutionized the production of potable water. Especially, the invention of polyamide (PA) TFC[1] has advanced the seawater desalination significantly. Its excellent water permeation and salt rejection[2] has made PA a huge commercial success. A PA membrane is created by cross-linking trimesoyl chloride (TMC) and m-phenylenediamine (MPD) via interfacial polymerization. However, PA membranes suffer severely by the performance limiting phenomena of membrane fouling[3, 4]. The high fouling propensity of PA membranes is mainly attributed to the high roughness of the PA surface[5].

In the past few years, graphene oxide (GO) has emerged as a promising candidate as a next-generation membrane material for water desalination[6, 7]. Apart from offering amazing water permeability[8, 9], high salt rejection[9, 10], and strong hydrophilicity[11], GO also exhibits excellent antifouling properties[8]. Several studies have focused on testing and comparing GO's antifouling capabilities with conventional PA membranes, and GO has always outperformed the conventional membrane in terms of antifouling performance[12–14].

An important aspect of fouling studies, which remains largely unresolved and debated, is the effect of ions on membrane fouling. Since the feed water that is to

be treated can have a varied range of ionic concentration depending on the source of the feed water, it is important to have an understanding of the effect of ion concentration on membrane fouling. Although, there are number of studies on the subject[15–20], yet a general consensus is lacking in the literature. Some studies claim that membrane fouling increases with an increase in the number of ions and attributes the charge screening and compression of the electric double layer for this[15–17]. Conversely, Chan et al. found that with an increase in ionic strength, the fouling decreases. The reason given for this observation is the increased solubility of protein with an increase in salt concentration due to which the protein becomes less prone to aggregation[21]. Miao et al. investigated protein fouling with bovine serum albumin (BSA) on polyvinylidene fluoride (PVDF) at different ionic strengths[18]. They found, at much higher ionic strength (10 and 100 mM), fouling decreases significantly, mainly because of increased hydration repulsion forces. Smith et al. studied the adsorption of lysozyme on graphene, GO, and single-walled carbon nanotube (SWNT)[20]. They investigated the effect of ionic concentration on protein adsorption and found the effect of ions different for all three nanomaterials. In particular, GO exhibited a sharp decrease in protein adsorption, which was mainly due to charge screening. Conversely, for SWNT, a sharp increase in protein adsorption was observed, for which cation-$\pi$ interactions was suggested as a possible mechanism, and for graphene, no significant change was observed.

Besides numerous experimental studies on membrane fouling available in the literature, there are fewer computational studies on the topic[19, 22–24]. Molecular Dynamics (MD) simulations, with its high resolution and the ability to access atomic-level detail, can provide critical insights into the mechanism governing membrane fouling. Xiang et al. studied the binding of alginate with a PA membrane using steered MD[22]. They explained how calcium ions play a role of an ionic bridge between alginate and the PA, and hence aggravate the fouling. Cruz et al. studied alginate fouling with a PA and multiwalled carbon nanotube PA (CNT-PA) composite membrane using experiments and MD simulations[24]. They found that alginate can foul a PA membrane through forming an ionic bridge with PA via calcium ions or by getting unfolded and spreading over the membrane surface. Recently Shaikh et al. performed MD simulations with PVDF membrane and BSA

molecule as a foulant[19]. They studied the effect of ionic strength on the protein adsorption. They found that an increase in ionic strength delayed the protein adsorption, mainly due to the adsorption of chloride ions on the membrane surface and increasing its electro-negativity. It is evident that MD simulations can be an indispensible tool in investigating the nature of membrane fouling and can provide subtle details and information about various factors affecting fouling and membrane surface. Hence MD simulations can be useful to provide more clarity over the effect of ions on membrane fouling.

In the present study, all-atom simulations were performed to shed light on the effect of ionic strength on protein-surface interaction(PSI) with GO and PA used as membranes. Proteins are one of the main constituents in the category of organic foulants. Proteins are complex molecules made up of different combinations of 20 different amino acids bonded together, each amino acid differ in chemical nature based on the type of side chain it contains. The different chemical nature of amino acids can be broadly classified as charged polar groups, hydrophobic residues, and aromatic groups. This combination of amino acids with varied nature give rise to a very complex chemistry, which makes protein different from other organic molecules. BSA, a globular protein, which is one of the most widely used organic foulants in membrane fouling studies[12, 16, 18, 19], was used as the model foulant in this study. One of the main factors affecting the PSIs is the environment's pH because pH can change the protonation states of the titratable residues present in the protein and the protonation states of functional groups present on PA and GO surface. But, since in the present study, the system considered represents a seawater feed solution, for which the average pH is around 7.5 to 8.3[25]. Hence, we can assume a constant protonation state for all the titratable protein residues and functional groups of GO and PA. To mimic the aqueous seawater environment, we created systems with seawater ion concentration for each membrane. In total, four simulation systems were generated, two (system with seawater ion concentration and without excess ion system) for each membrane. PSI, structural changes in the protein, distribution of ions, and the effect of ions on them were evaluated and compared for both the membranes.

## 4.2 Membranes model & system preparation

### 4.2.1 Polyamide membrane

An atomistic model of cross-linked PA membrane was built by mimicking the interfacial polymerization process. A heuristic method was implemented, which involves cross-linking the monomers TMC and MPD in a vacuum based on a distance criterion. Subsequently, the prepared membrane was solvated to mimic its swelling in water [26] [27] [28]. The structure of MPD and hydrated TMC (TMO) monomer is shown in Figure 4.1a.

Initially 2000 m-phenylenediamine (MPD) and 2000 TMO monomers were packed inside a box of size $125 \times 95 \times 46$ Å$^3$ using packmol [29]. This mixture was then energy minimized and subsequently equilibrated for 2 $ns$ at constant temperature and pressure (340 $K$ temperature and 1 $atm$ pressure). After equilibrating the mixture, the polymerization runs were performed at a constant temperature of 340 $K$ [27]. The polymerization runs were performed for 2000 steps, where each step was simulated for 1000 $fs$ with a time step of 1 $fs$. During this polymerization process, linkages between the monomers take place based on the user-defined cut-off distance. When the distance between C atoms of -COOH group (TMO monomer) and N atoms of the -$NH_2$ group (MPD monomers) is less than the cut-off distance, then a linkage between the monomers are formed. In this present work, for the first 200 steps, the cut-off distance is 3 Å. The cut-off distance is gradually increased with an increment of 0.4 Å at every 200 steps up to 4.6 Å. Then for the steps in the range $1001 - 1500$, the cut-off distance is 4.8 Å. Finally, in the range of $1501 - 2000$, the cut-off distance is 5 Å.

After the polymerization process, the unreacted monomers (monomers which did not form any linkage) were removed from the mixture. The mixture is then energy minimized and subsequently hydrated in an equilibrated water box of size $125 \times 95 \times 137$ Å$^3$. This hydrated mixture (or membrane) is then energy minimized and subsequently equilibrated for 2 $ns$ at a constant temperature of 298.15 $K$ and 1 $atm$ pressure. The equilibrated hydrated membrane is then subjected to a heating-annealing process between 368.15 $K$ to 298.15 $K$ for four cycles. After annealing,

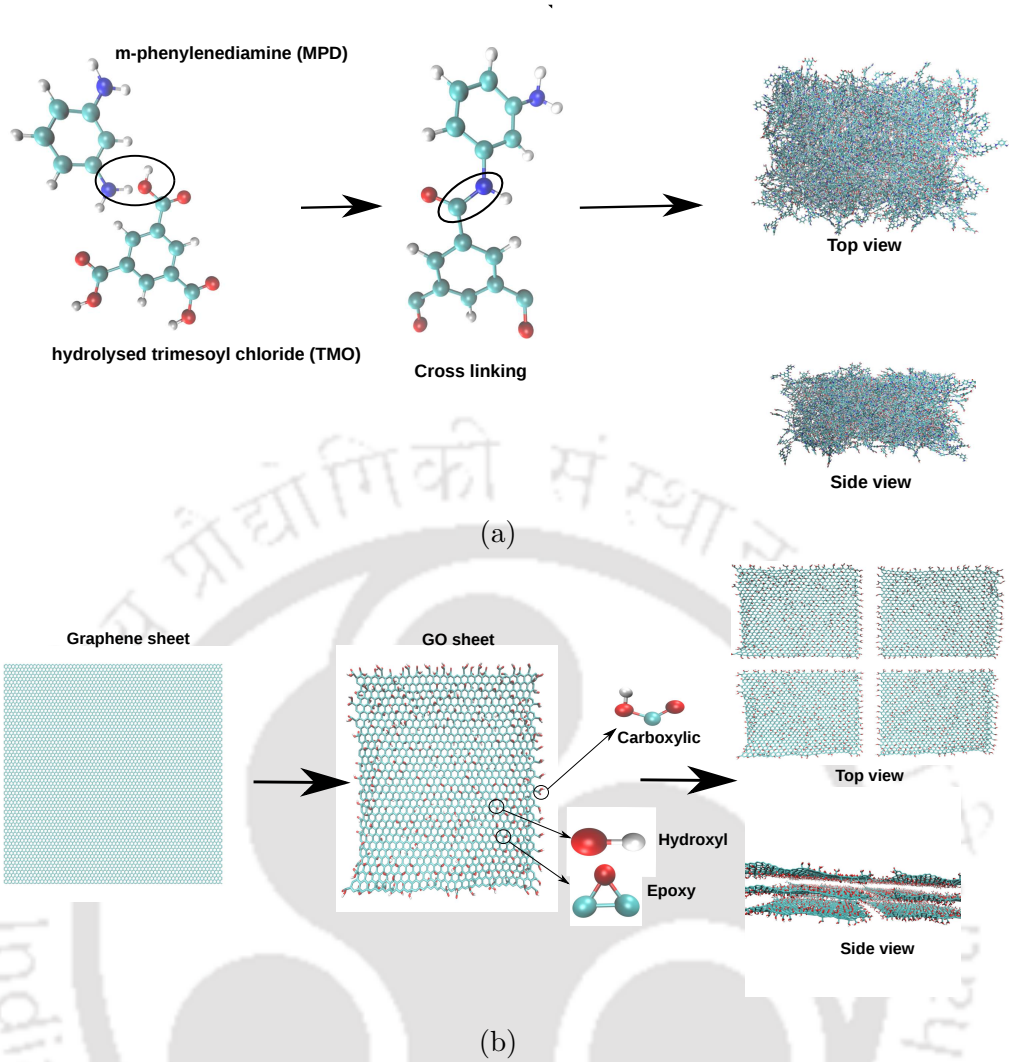



Figure 4.1: A representative depiction of the steps adopted in the construction of (a) PA membrane and (b) GO membrane. The colors cyan, white, red, blue represents carbon, hydrogen, oxygen and nitrogen atoms respectively.

the hydrated membrane is equilibrated for 5 $ns$ at a constant temperature of 298.15 $K$ and 1 $atm$ pressure. The size of this final hydrated membrane is $125 \times 95 \times 128$ $Å^3$.

### 4.2.2 GO membrane

A nanoporous GO membrane was constructed by stacking GO sheets in a configuration, as shown in Figure 4.1b. The figure 4.1b shows the steps followed to create

the GO membrane. A graphene sheet is first generated using the Nanotube builder plugin available in Visual Molecular Dynamics(VMD)[30], further various functional groups (carboxyl, hydroxyl, and epoxy) were added to the graphene sheet with Avogadro [31] to create a single GO sheet. The GO sheet was modeled based on the Lerf-klinowiski model [32], according to which the hydroxyl (-OH) and epoxy (-O-) were bonded randomly to the carbon atoms present on the basal plane of the graphene sheet, while the carboxy (-COOH) group was attached randomly to the carbon atoms present at the edge of the graphene sheet. The partial charges and other parameters for GO were taken directly from the work of Shih et al.[33]. The single GO sheet prepared has a dimension of $70 \times 70$ Å$^2$. The GO sheets were then stacked to form a membrane, as shown in Figure 4.1b. The prepared membrane is then solvated in a water box of dimension $146 \times 146 \times 57$ Å$^3$ and water molecules within 2 Å of GO sheets were removed. The hydrated membrane is then subjected to energy minimization, followed by an equilibration run for 5 ns at 300 K and 1 atm.

### 4.2.3 BSA solution

In the present work, BSA was used as a model foulant to study the membrane fouling. BSA is a protein with 583 amino acid residues and has been used as a model organic foulant in many experimental and computational studies based on membrane fouling[5, 16, 19, 34]. The initial structure of a single BSA molecule was extracted from Protein Data Bank (PDB ID: 4F5S)[35]. All the titratable residues in the protein were allotted appropriate protonation states at neutral pH (7.4), and accordingly, the missing hydrogen atoms were added, which resulted in a total protein charge of -17e. For two membranes (GO and PA), protein is solvated in two separate water boxes with dimensions $146 \times 146 \times 100$ Å$^3$ and $125 \times 95 \times 106$ Å$^3$ for GO and PA respectively. Further, to investigate the effect of ion concentration, for each of the solvated systems, two separate systems were generated such that one of the systems has the ion($Na^+$ and $Cl^-$) concentration equal to the seawater concentration(O.56 M). In contrast, the other system has just enough counterions to neutralize the total charge of the system. All systems were then energy minimized

for 2000 steps followed by an equilibration run at 300 K in an NPT ensemble for 200 ps with protein atoms fixed. Finally, the protein atoms were set free for a final equilibration run for 6 ns in an NPT ensemble at 300 K and 1 atm.

### 4.2.4 Simulation systems and details

After preparing hydrated membranes and protein solutions individually, they were merged to construct the simulation system. In total, there are four simulation systems, i.e., for both GO and PA, there were two systems each, such that one system is without excess ions and another with seawater concentration of ions. Hereafter, the systems without excess ions are referred to as "GN" and "PN" and systems with seawater ion concentration as "GS" and "PS" for GO and PA, respectively. The merged systems were equilibrated for 1 ns at 300 K and 1 atm. The protein was kept initially at the closest distance of 15 Å from the membrane for all systems. A graphene sheet is placed below the membrane, and a vacuum in Z direction was kept on both sides of the system to avoid any interaction between the simulation box and the periodic image. All the GO sheet's basal plane carbon atoms were constrained in x and y direction using harmonic constraints to avoid the lateral movement of the GO membrane and keep it stable. The prepared systems (Figure 4.2) were then submitted for a production run of 200 ns.

All equilibration simulations were performed in the NPT ensemble, while production simulations were carried out in the NVT ensemble using NAMD [36] package. Atomic interactions for GO systems were computed using OPLS-AA force field parameters [37], while the CHARMM force field [38] was used for PA systems. The water molecules were modeled using the TIP3P water model [39], and their bond length was constrained using SETTLE algorithm[40]. Velocity Verlet algorithm [36] was used to integrate the equations of motion with a time step of 1 fs. The van der Waals interactions were computed using Lennard-Jones potential with a cut off of 12 Å and a switching distance of 10 Å. The long-range electrostatic interactions were treated using the particle mesh Ewald (PME) [41].

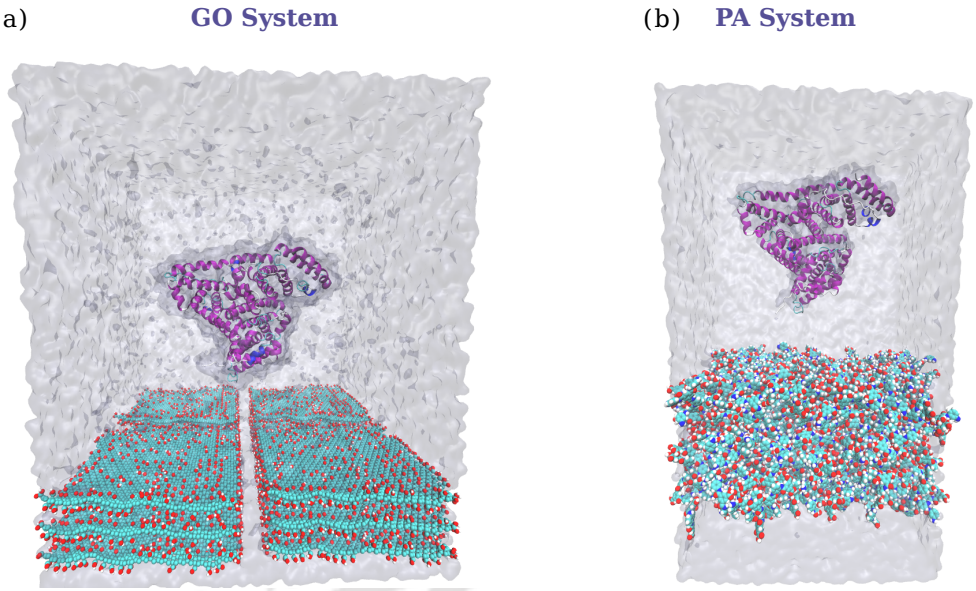



Figure 4.2: Simulation systems for (a) GO and (b) PA. The water is shown as Quicksurf, BSA in New Cartoon and membranes and graphene sheet in VdW representations. The simulation systems shown above are generic setup and ions are not shown for clarity.

## 4.3 Results and discussion

### 4.3.1 Interaction between BSA and Membrane Surface

To evaluate the behavior of BSA in the presence of a membrane surface, we computed the closest distance between BSA and the membrane's surface for all the systems. Figure 4.3a and b show the evolution of the closest distance between BSA and the membrane surface for GO and PA systems, respectively. In the GN system, the distance between the GO surface and protein (Figure 4.3a) increases continuously, indicating protein diffusion away from the surface towards the bulk. The diffusion of BSA away from the surface indicates that BSA prefers to stay in the bulk, avoiding the GO surface. In the case of the GS system, we observe that the BSA diffuses away from the GO surface; however, the distance plot (Figure 4.3a) shows that the diffusion of BSA in the GS system is slower than in the GN system. The distance evolution for PA systems (Figure 4.3b) shows that BSA adsorbs strongly on the PA surface in the PN system. The distance between BSA and PA surface in the PN

system initially decreases quickly till 100 ns, and it diffuses slightly back towards the bulk for the next 50 ns. However, the distance again quickly decreases until 200 ns and remains at a distance of 5 $\AA$ from the surface up to the end of the simulation. Conversely, the distance

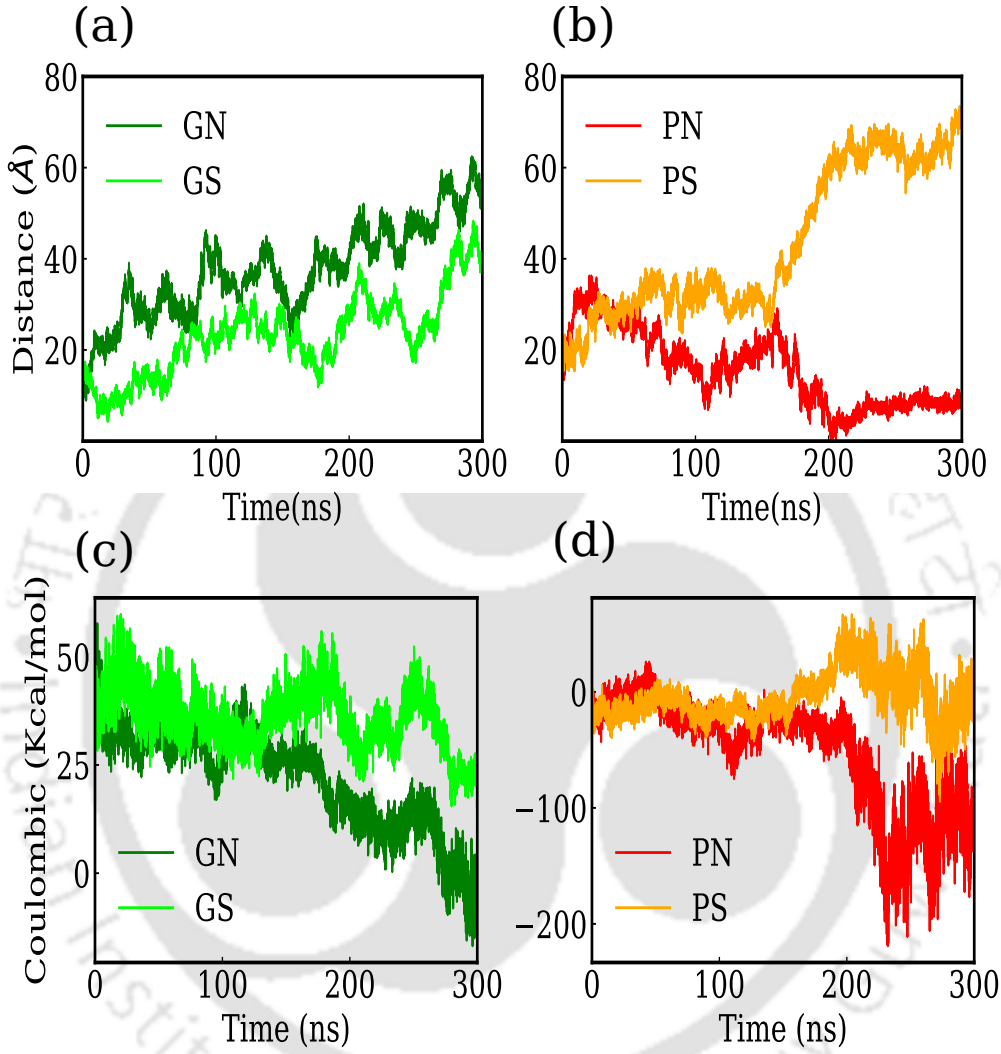



Figure 4.3: Time evolution of closest distance between BSA and (a) GO systems and (b) PA systems. Time evolution of interaction (Coulombic) energy between BSA and (c) GO systems and (d) PA systems.

evolution for PS (Figure 4.3b) shows that a high concentration of ions makes the BSA diffuse away from the PA surface towards the bulk. The distance between BSA and PA surface in the PS system does not change much initially and remains between 20 $\AA$ to 30 $\AA$ up to around 170 ns. However, after 170 ns, there is a drastic increase in the distance as it increases to 65 $\AA$ till 200 ns and remains nearly constant

until the end of the simulation.

To gain further insights into PSI and decipher the nature of the interaction between BSA and the surface in all the systems, we analyzed the time evolution of the Coulumbic energy between BSA and the surface for all the systems. The interaction energy between BSA and the surface in all the systems is dominantly Coulombic in nature, and no vdW interactions were observed except in the PN system since only in the case of the PN system the BSA diffuses close enough to PA to be able to interact via vdW interactions. In this section we report only the Coulombic interaction energy between BSA and all the membrane surfaces and the vdW interaction energy between BSA and PA in PN system will be reported in an upcoming section where we will discuss adsorption of BSA on PA in PN system in detail. We can also observe from the interaction energy plots (Figure 4.3c and 4.3d) that the time evolution of Coulombic energy for all the systems is closely correlated with the corresponding distance analysis. Figure 4.3c shows time evolution of Coulombic energy between BSA and the membrane surface in GO systems. The GS and GN system shows positive interaction energy indicating a repulsive interaction between BSA and GO. We can observe that as the distance between BSA and GO in GN system increases, the interaction energy between them quickly decreases to zero. Similarly, the interaction energy for GS system also decreases as time progresses, however, the decline is slower than in GN, this is again correlated with the corresponding trend in distance evolution of GS. Figure 4.3d shows the time evolution of interaction energy between BSA and the membrane in PA systems. The interaction energy trends for PA systems also show correlation with the corresponding distance evolution. The BSA and PA in PN system initially shows weaker interaction as observed in the interaction energy plot for PN system the energy fluctuate around zero. However, at around 180 ns the Coulombic energy for PN system starts decreasing, indicating a favourable interaction. The interaction energy for PN system converges at around -150 Kcal/mol, indicating an adsorption event, and the lowest energy observed for PN system was around -220 Kcal/mol. In the PS system also initially we observe a weak interaction between BSA and PA, shown by a very small interaction energy. As the simulation progresses the interaction energy for the PS system approaches a convergence to zero, this again correlates with the diffusion of BSA away from the

membrane shown in the distance analysis.

Overall, the distance and interaction energy analysis show that the presence of ions strongly affects the PSI in both PS and GS systems. Specifically, the presence of excess ions screens the interaction between the membrane surface and the BSA. Also, the PSI between BSA and GO are strongly repulsive in nature, while in case of PA, interactions are favourable for the adsorption of BSA.

### 4.3.2 Structural and Conformational Changes in the Protein

It has been demonstrated that the proteins undergo structural and conformational changes in the presence of surfaces [19, 42]. Also, it is a well-known fact that the presence of ions impacts the structural stability of the protein [43, 44]. In this section, we investigated the time evolution of root-mean-square deviation (RMSD), root-mean-squre fluctuations (RMSF), and orientation of protein relative to the membrane surface to observe the evolution of structure and conformation of the protein.

The time evolution of RMSD reflects the changes in the protein structure due to the presence of the membrane surface. Figure 4.4a shows the time evolution of RMSD of backbone $C_\alpha$ atoms of the BSA in GO systems. For GS system the RMSD increases initially and converges to a value of around 3 $\mathring{A}$ and remain constant till the end of the simulation. The RMSD of GS system suggests an overall stable BSA structure. The RMSD of BSA in GN system initially increases upto around 3 $\mathring{A}$ and remains constant upto 100 ns. The RMSD increases to 4.5 $\mathring{A}$ after 100 ns and remains constant till the end of the simulation. The RMSD of GN system shows perturbation of BSA structure in the presence of GO. The RMSD analysis of GO system shows that the presence of excess ions conserve the structure of protein by screening the interaction between BSA and GO surface. The RMSD in both PA systems (Figure 4.4b) increases initially and stabilizes quickly at around 2.5 $\mathring{A}$. The RMSD evolution for PA systems shows that BSA shows resilience to any changes in its structure. This could be probably due to the weaker interaction of BSA with PA, as observed in the interaction energy analysis. Interestingly, eventhough we

observed adsorption of BSA on PA surface in the PN system, the RMSD of PN reflects that the structure of BSA is unpurturbed.

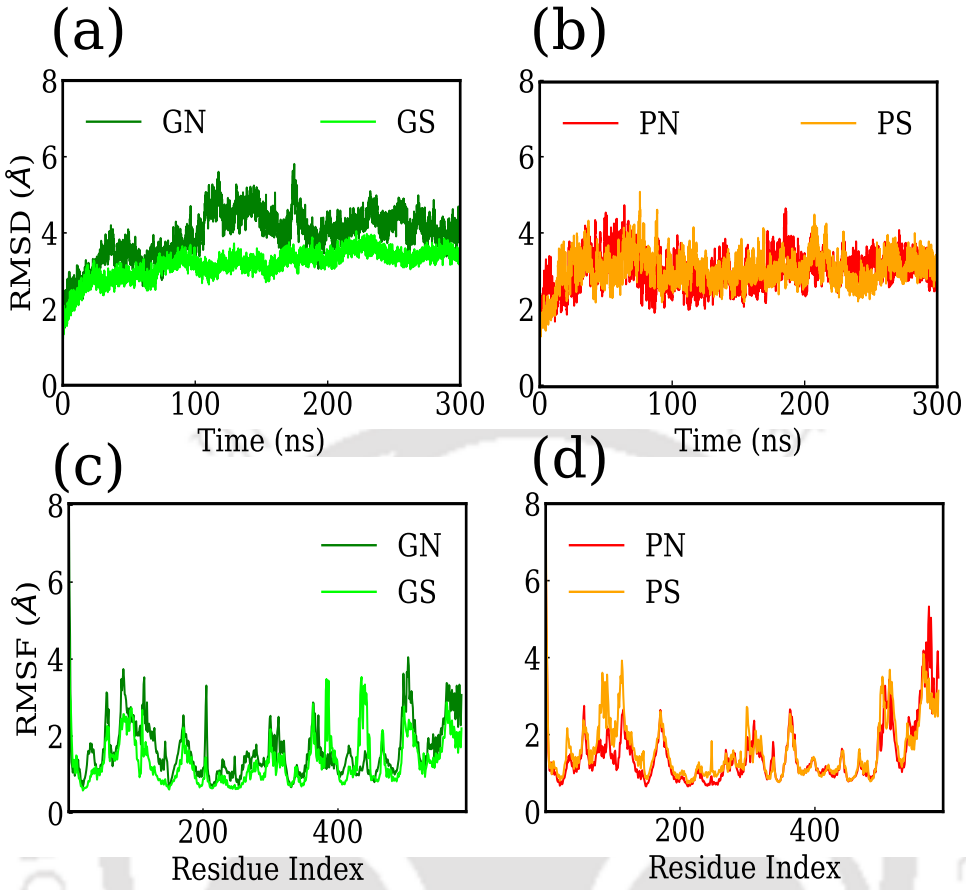



Figure 4.4: Time evolution of RMSD of C$\alpha$ atoms of BSA in (a) GO systems and (b) PA systems. RMSF of C$\alpha$ atoms of BSA in (c) GO systems and (d) PA systems.

To further decipher the local structural changes in protein, we calculated the RMSF for each residue of BSA. As we observe the RMSF analysis for GO systems (Figure 4.4c) shows that most of the residues in PN system shows higher RMSF values in comparison to their counterparts in PS system. The highest difference in the RMSF was found in the N-terminal region of the BSA. The RMSF values of most of the residues in PN and PS systems (Figure 4.4d) overlaps, as was also observed in time evolution of RMSD for PA systems. However, the residues in the N-terminal region in PS system shows higher RMSF than in PN system. Also, the end residues of C-terminal region in PN system shows higher RMSF than PS system and also much higher in comparison to those residues in GO systems, indicating pronounced flexibility of end residues in C-terminal of PN system. The higher RMSF of C-

terminal region in PN system must be due to the participation of residues from this region in the interaction with PA surface. Overall, the RMSD analysis shows that their are no striking changes in the global structure of the BSA, however, the RMSF analysis shows high RMSF for C-terminal region of BSA, which is probably due to its participation in the adsorption process.

Another critical parameter to characterize PSI is the orientation of the protein with respect to the surface. The orientation of the protein is defined as the angle made by the protein's dipole moment vector with the normal of the surface (Figure 4.5a). Figure 4.5 shows the time evolution of $\theta$ for all the systems. The $\theta$ for the GN system initially decreases to around 10° till 50 ns. However, from 50 ns to 150 ns, the $\theta$ increases rapidly to a value of around 100$^circ$, and afterward, it converges and remains between 100° and 115°. The $\theta$ evolution for the GN system shows that as BSA diffuses towards bulk, it continuously rotates itself in order to find an optimum conformation to interact with the GO surface. A visual analysis (Figure 4.6) of the orientation of BSA relative to the surface shows that the BSA orients itself to present its subdomain IIIB towards the GO surface. Kubiak et al.[**?** ] also reported the reorientation of BSA to offer its subdomain IIIB for adsorption on a negatively charged surface (silica).

The $\theta$ for the GS system does not show drastic changes during the simulation time; up to 200 ns, the $\theta$ remains between 50° and 30°. After 200 ns, the $\theta$ reduces to 25° and stays at it till the end of the simulation. The $\theta$ evolution for the GS system shows that the BSA does not change its orientation, this must be because of the screening of PSI due to the presence of high concentration of ions between GO and BSA. Figure 4.6 shows the BSA's orientation with subdomain IIA facing GO in the GS system. The $\theta$ in the PN system remains stable initially around 50° up to 150 ns. However, as the BSA approaches the PA surface, the $\theta$ decreases rapidly to 0° and then, after 220 ns, stabilizes to 25°. The $\theta$ evolution for the PN system shows that as the BSA diffuses towards the PA surface, the rotation of BSA gets restricted, and it finally adsorbs with its subdomain IB, IIA, and IIB in direct contact with the PA (Figure 4.6). The $\theta$ for the PS system does not change rapidly initially and remains stable.

at around 25°. After 175 ns, as the BSA diffuses away from the PA surface, the

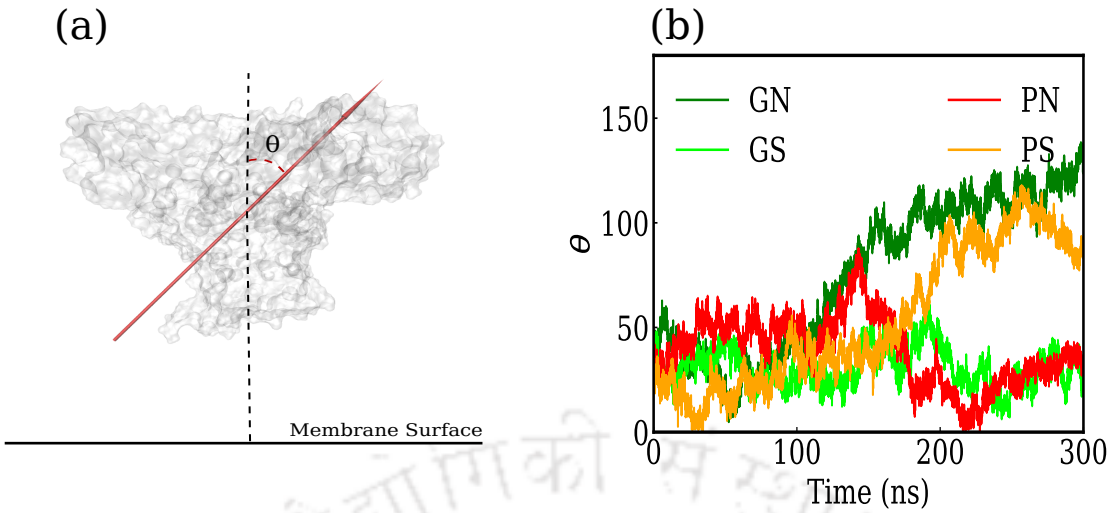



Figure 4.5: (a) A representative definition of the angle ($\theta$) between the dipole moment vector (red arrow) of BSA (shown in Quicksurf and ghost representation) and the normal (dotted vertical line) of the membrane's surface (horizontal solid line). (b) Time evolution of the $\theta$ for all the systems.

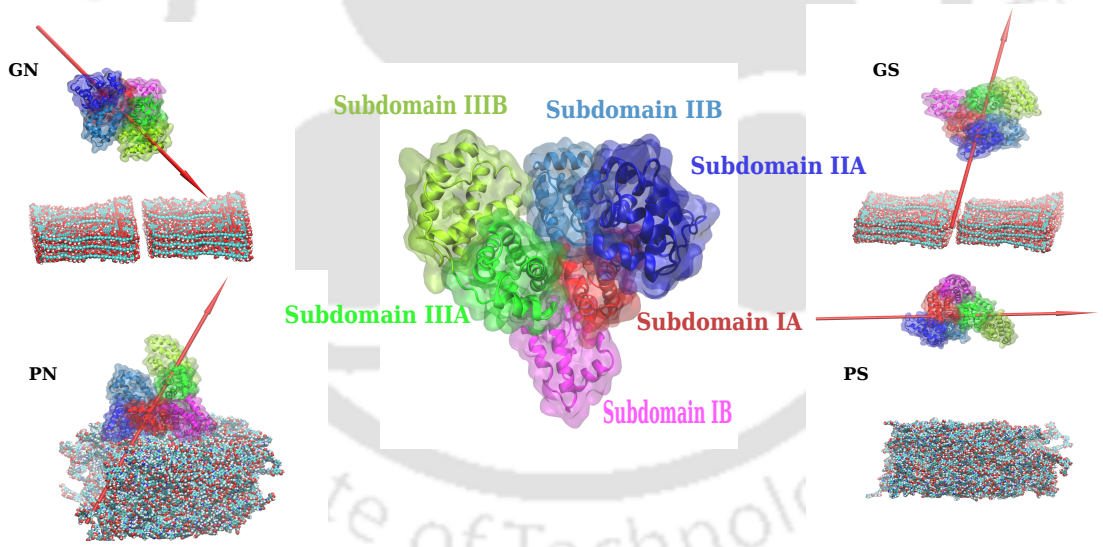



Figure 4.6: BSA represented in New Cartoon drawing with its various domains and subdomains (center) shown in different colors. BSA orientation relative to surface in different systems shown with the dipole moment vector at the end of the simulation.

$\theta$ increases rapidly up to 110°, indicating a drastic change in the orientation of BSA. The end simulation snapshot for the orientation of BSA in the PS system (Figure 4.6) shows a parallel orientation between BSA's dipole moment vector and

PA surface; however, we did not observe a convergence of $\theta$ for PS system within the simulated time. We observe the differences in the conformational evolution of BSA in PS and GS systems, indicating that high salt concentration has a varied effect on the PSI based on the membrane surface under consideration. We speculate that the differences observed are due to the different hydration layer structures at the surface of the two membranes, which arise from the surface chemistry and morphology of the membrane. The PA surface has a higher density of hydrophilic residues, which can interact readily with water molecules and ions, creating a tighter hydration layer and hence offering a strong hydration repulsion. However, in the case of GO, a higher density of hydrophilic residues is present only at the edges of the GO sheet in the form of the carboxyl group, and the central region of the GO sheet is predominantly hydrophobic which will lead to a weaker hydration layer and consequently a weaker hydration repulsion.

### 4.3.3 Adsorption of BSA on PA

The time evolution of closest distance and interaction energy (Figure 4.3) between BSA and membrane surfaces indicates adsorption of BSA on PA surface in the PN system. To analyze the adsorption event in detail and decipher the nature of interaction between BSA and PA, we computed the contact numbers and interaction energies between residues of BSA and the PA surface for the PN system.

#### Key interacting residues

Figure 4.7a shows the number of contacts between the residues of BSA and PA surface. Histidine(HSE 3) shows the strongest interaction as indicated by the highest contact number for HSE in the contact number plot (Figure 4.7a). Figure 4.7b shows the van der Waals (vdW) interaction energies between residues of subdomain IA (SD IA) of BSA and PA surface. The vdW energies are highest for HSE 3 and LYS 4, which are in agreement with the contact number analysis. Although, contact number plot shows many residues adjacent to HSE 3 having considerable contact number, however similar participation is not observed by them in vdW interaction energies except LYS 4, this indicates that they only show close contacts due to a

strong interaction of HSE 3 with PA surface. The high contact number and vdW interaction energy of histidine with the PA surface suggest a possible presence of $\pi$-$\pi$ interaction between the benzene ring of HSE 3 and the TMO or MPD of PA.

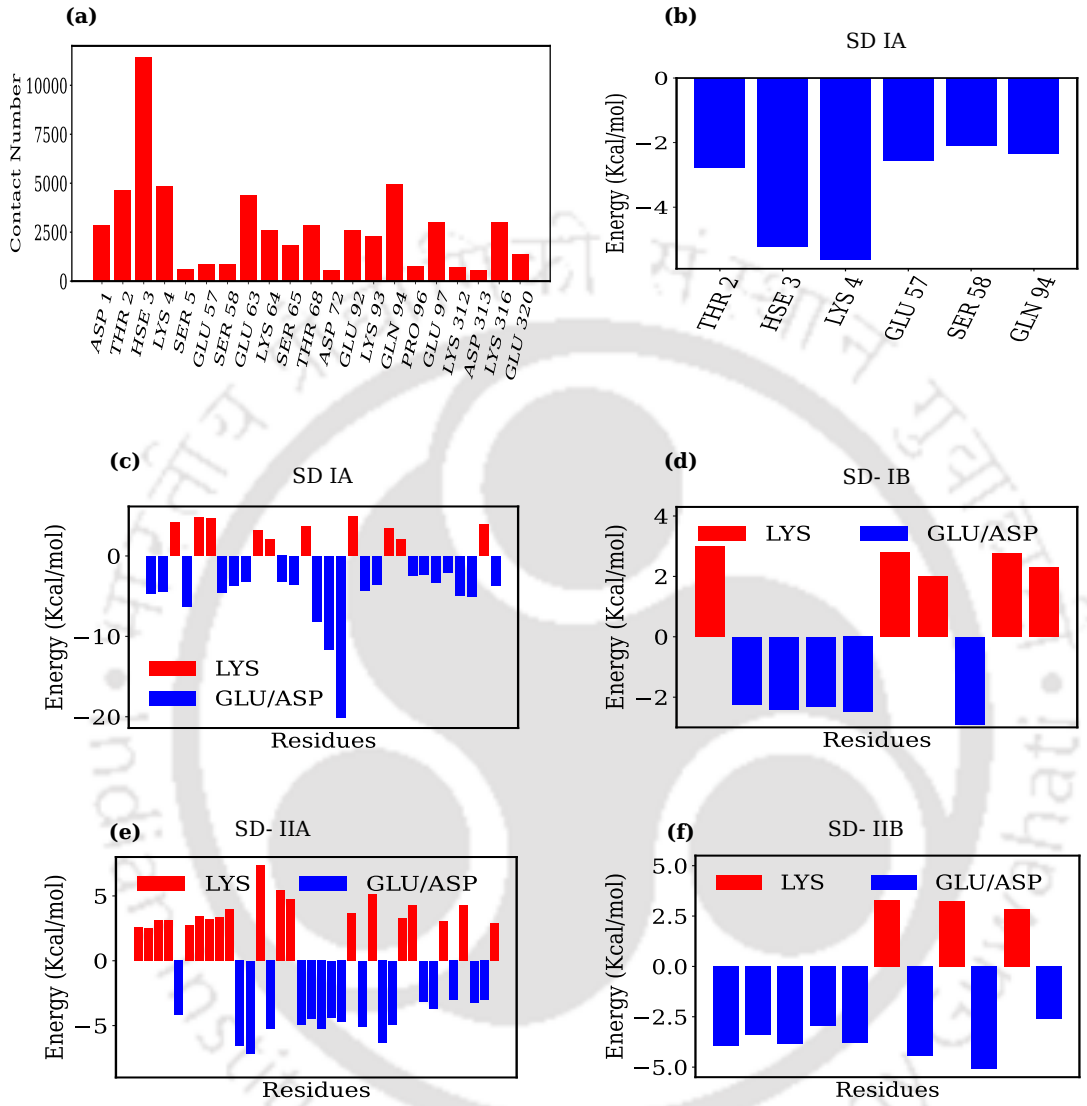



Figure 4.7: (a) Number of contacts between residues of BSA and PA surface. (b) Van der Waals interaction energies of subdomain IA of BSA with the PA surface. Coulombic interaction energies between PA surface and residues in (c) subdomain IA, (d) subdomain IB, (e) subdomain IIA and (f) subdomain IIB of BSA.

Also high vdW interaction energy for positively charged lysine (LYS 4) indicates towards a cation-$\pi$ interaction with the benzene rings of MPD and TMO. Except for SD IA residues of other domains didn't show any significant vdW interaction

energy with the PA surface. We also computed Coulombic interaction energies between residues of various domains of BSA and the PA surface. Almost, all the residues showing repulsive interaction with PA were found to be lysine, while the one having attractive are either apartic acid or glutamic acid Figure 4.7c shows the residuewise interaction energies for SD-IA. As can be observed from the Figure 4.7c SD-IA have higher number of residues which have attractive interaction with PA then the residues having repulsive interaction. Specifically, ASP 56 (-8.098 $\pm$ 5.09), GLU 57 (-11.61 $\pm$ 10.57), and GLU 63 (-20.15 $\pm$ 11.73) shows very strong attraction towards the PA surface. The SD IB(Figure 4.7d) shows relatively weaker Coulombic interaction in comparison to the SD-IA.The SD-IB exhibits almost equal number of residues having repulsive and attractive interaction with PA and with almost equal strength. The SD-IIA (Figure 4.7e) shows strong interaction with the PA surface. Although, relatively there are more number of residues having repulsive interation in SD-IIA, the attractive residues have higher strength of interaction in comparison to the repulsive one. Finally, compared to SD-IIA, SD-IIB (Figure 4.7f) have less number of residues which have a significant interaction with PA surface. However, the number of residues and also the strength of interaction is higher for the attractive interaction for SD-IIB, indicating a net attractive interaction with the PA surface.

Overall, the residuewise interaction analysis shows that the BSA has a weaker vdW interaction with the PA surface and the adsorption process is mainly driven by electrostatic interactions. The Coulombic interaction energies also show that the PA surface can interact with both positively (LYS) and negatively(GLU/ASP) charged residues of BSA, however, a stronger net attractive interaction was found with the negatively charged residues. The SD-IA and SD-IIA among the subdomains were found to be interacting strongly with the PA surface. The contact analysis and vdW energies also indicated towards a probable $\pi - \pi$ and cation-$\pi$ interaction.

## 4.4 Conclusions

We studied the effects of ionic concentration on the interaction between GO/PA and BSA protein via atomistic simulations. To investigate the effect of the ionic

environment on PSI for both the membrane surfaces, systems with excess ions (seawater ion concentration) and without excess ions were created for each membrane. We observed that the presence of high concentration of ions strongly effect the protein-membrane interactions in case of both GO and PA. Specifically, GO shows a repulsive interaction while the PA shows attractive interaction with the BSA in the absence of ions and the time evolution analysis of distance and interaction energy between membrane and protein shows that the presence of ions strongly screens these interactions. Time evolution of RMSD for BSA showed that the BSA gained stability in the presence of ions for GO system, showing a strong screening effect on the BSA structure. However, RMSD for PA systems shows that the BSA structure remains more or less uneffected by the presence of ions. We also observed an adsorption event between BSA and PA surface in PN system. We computed contribution of inidividual residues of BSA in the interaction energy and number of contacts between BSA and PA surface. The contact and interaction energy analysis shows that the BSA adsorption on PA is mostly electrostatically driven and vdW interactions were found to be weak.

Our work provides crucial insights into the fouling behavior of two important membrane material (GO and PA) and the effect of ionic concentration on it. Also, the study reveals the important role played by the functional groups present on the membrane's surface in mediating the surface-ion interaction, which consequently effects the protein-surface interactions. The study may lead to new surface functionalization strategies for better antifouling performance of desalination membranes.

# Bibliography

[1] JE Cadotte, RJ Petersen, RE Larson, and EE Erickson. A new thin-film composite seawater reverse osmosis membrane. *Desalination*, 32:25–31, 1980.

[2] Menachem Elimelech and William A Phillip. The future of seawater desalination: energy, technology, and the environment. *science*, 333(6043):712–717, 2011.

[3] Baoxia Mi and Menachem Elimelech. Organic fouling of forward osmosis mem-

branes: fouling reversibility and cleaning without chemical reagents. *Journal of membrane science*, 348(1-2):337–345, 2010.

[4] Yangshuo Gu, Yi-Ning Wang, Jing Wei, and Chuyang Y Tang. Organic fouling of thin-film composite polyamide and cellulose triacetate forward osmosis membranes by oppositely charged macromolecules. *Water research*, 47(5):1867–1874, 2013.

[5] Yoshihiro Takizawa, Shigeki Inukai, Takumi Araki, Rodolfo Cruz-Silva, Noriko Uemura, Aaron Morelos-Gomez, Josue Ortiz-Medina, Syogo Tejima, Kenji Takeuchi, Takeyuki Kawaguchi, et al. Antiorganic fouling and low-protein adhesion on reverse-osmosis membranes made of carbon nanotubes and polyamide nanocomposite. *ACS applied materials & interfaces*, 9(37):32192–32201, 2017.

[6] RK Joshi, S Alwarappan, M Yoshimura, V Sahajwalla, and Yuta Nishina. Graphene oxide: the new membrane material. *Applied Materials Today*, 1(1): 1–12, 2015.

[7] RR Nair, HA Wu, PN Jayaram, IV Grigorieva, and AK Geim. Unimpeded permeation of water through helium-leak–tight graphene-based membranes. *Science*, 335(6067):442–444, 2012.

[8] Meng Hu and Baoxia Mi. Enabling graphene oxide nanosheets as water separation membranes. *Environmental science & technology*, 47(8):3715–3723, 2013.

[9] Aaron Morelos-Gomez, Rodolfo Cruz-Silva, Hiroyuki Muramatsu, Josue Ortiz-Medina, Takumi Araki, Tomoyuki Fukuyo, Syogo Tejima, Kenji Takeuchi, Takuya Hayashi, Mauricio Terrones, et al. Effective nacl and dye rejection of hybrid graphene oxide/graphene layered membranes. *Nature nanotechnology*, 12(11):1083, 2017.

[10] Hanaa M Hegab and Linda Zou. Graphene oxide-assisted membranes: fabrication and potential applications in desalination and water purification. *Journal of Membrane Science*, 484:95–106, 2015.

[11] Jiguo Zhang, Zhiwei Xu, Wei Mai, Chunying Min, Baoming Zhou, Mingjing Shan, Yinglin Li, Caiyun Yang, Zhen Wang, and Xiaoming Qian. Improved hydrophilicity, permeability, antifouling and mechanical performance of pvdf composite ultrafiltration membranes tailored by oxidized low-dimensional carbon nanomaterials. *Journal of Materials Chemistry A*, 1(9):3101–3111, 2013.

[12] Meng Hu, Sunxiang Zheng, and Baoxia Mi. Organic fouling of graphene oxide membranes and its implications for membrane fouling control in engineered osmosis. *Environmental science & technology*, 50(2):685–693, 2016.

[13] Wansuk Choi, Jungkyu Choi, Joona Bang, and Jung-Hyun Lee. Layer-by-layer assembly of graphene oxide nanosheets on polyamide membranes for durable reverse-osmosis applications. *ACS applied materials & interfaces*, 5(23):12510–12519, 2013.

[14] Francois Perreault, Marissa E Tousley, and Menachem Elimelech. Thin-film composite polyamide membranes functionalized with biocidal graphene oxide nanosheets. *Environmental Science & Technology Letters*, 1(1):71–76, 2014.

[15] Yi-Ning Wang and Chuyang Y Tang. Protein fouling of nanofiltration, reverse osmosis, and ultrafiltration membranes—the role of hydrodynamic conditions, solution chemistry, and membrane properties. *Journal of Membrane Science*, 376(1-2):275–282, 2011.

[16] Wui Seng Ang and Menachem Elimelech. Protein (bsa) fouling of reverse osmosis membranes: implications for wastewater reclamation. *Journal of Membrane Science*, 296(1-2):83–92, 2007.

[17] Huajuan Mo, Kwee Guan Tay, and How Yong Ng. Fouling of reverse osmosis membrane by protein (bsa): effects of ph, calcium, magnesium, ionic strength and temperature. *Journal of Membrane Science*, 315(1-2):28–35, 2008.

[18] Rui Miao, Lei Wang, Na Mi, Zhe Gao, Tingting Liu, Yongtao Lv, Xudong Wang, Xiaorong Meng, and Yongzhe Yang. Enhancement and mitigation mechanisms of protein fouling of ultrafiltration membranes under different ionic strengths. *Environmental science & technology*, 49(11):6574–6580, 2015.

[19] Abdul Rajjak Shaikh, Hamed Karkhanechi, Tomohisa Yoshioka, Hideto Matsuyama, Hiromitsu Takaba, and Da-Ming Wang. Adsorption of bovine serum albumin on poly (vinylidene fluoride) surfaces in the presence of ions: A molecular dynamics simulation. *The Journal of Physical Chemistry B*, 122(6):1919–1928, 2018.

[20] Sean C Smith, Farid Ahmed, Krystal M Gutierrez, and Debora Frigi Rodrigues. A comparative study of lysozyme adsorption with graphene, graphene

oxide, and single-walled carbon nanotubes: potential environmental applications. *Chemical Engineering Journal*, 240:147–154, 2014.

[21] R Chan and V Chen. The effects of electrolyte concentration and ph on protein aggregation and deposition: critical flux and constant flux membrane filtration. *Journal of Membrane Science*, 185(2):177–192, 2001.

[22] Yuan Xiang, Yaolin Liu, Baoxia Mi, and Yongsheng Leng. Hydrated polyamide membrane and its interaction with alginate: a molecular dynamics study. *Langmuir*, 29(37):11600–11608, 2013.

[23] Yuan Xiang, Yaolin Liu, Baoxia Mi, and Yongsheng Leng. Molecular dynamics simulations of polyamide membrane, calcium alginate gel, and their interactions in aqueous solution. *Langmuir*, 30(30):9098–9106, 2014.

[24] Rodolfo Cruz-Silva, Yoshihiro Takizawa, Auppatham Nakaruk, Michio Katouda, Ayaka Yamanaka, Josue Ortiz-Medina, Aaron Morelos-Gomez, Syogo Tejima, Michiko Obata, Kenji Takeuchi, et al. New insights in the natural organic matter fouling mechanism of polyamide and nanocomposite multiwalled carbon nanotubes-polyamide membranes. *Environmental science & technology*, 53(11):6255–6263, 2019.

[25] GM Marion, Frank J Millero, MF Camões, P Spitzer, R Feistel, and C-TA Chen. ph of seawater. *Marine Chemistry*, 126(1-4):89–96, 2011.

[26] T. Wei, L. Zhang, H. Zhao, H. Ma, Md S. J. Sajib, H. Jiang, and S. Murad. Aromatic polyamide reverse-osmosis membrane: An atomistic molecular dynamics simulation. *J. Phys. Chem. B*, 120(39):10311–10318, 2016. doi: 10.1021/acs.jpcb.6b06560.

[27] Y. Luo, E. Harder, R. S. Faibish, and B. Roux. Computer simulations of water flux and salt permeability of the reverse osmosis FT-30 aromatic polyamide membrane. *J. Membr. Sci.*, 384(1-2):1–9, 2011. doi: 10.1016/j.memsci.2011.08.057.

[28] E. Harder, D. E. Walters, Y. D. Bodnar, R. S. Faibish, and B. Roux. Molecular dynamics study of a polymeric reverse osmosis membrane. *J. Phys. Chem. B*, 113(30):10177–10182, 2009. doi: 10.1021/jp902715f.

[29] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. Packmol: A

package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.*, 30(13):2157–2164, 2009. doi: 10.1002/jcc.21224.

[30] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996. doi: 10.1016/0263-7855(96)00018-5.

[31] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1): 17, 2012.

[32] Heyong He, Jacek Klinowski, Michael Forster, and Anton Lerf. A new structural model for graphite oxide. *Chemical physics letters*, 287(1-2):53–56, 1998.

[33] Chih-Jen Shih, Shangchao Lin, Richa Sharma, Michael S Strano, and Daniel Blankschtein. Understanding the ph-dependent behavior of graphene oxide aqueous solutions: a comparative experimental and molecular dynamics simulation study. *Langmuir*, 28(1):235–241, 2012.

[34] Qingye Lu, Jun Huang, Omar Maan, Yang Liu, and Hongbo Zeng. Probing molecular interaction mechanisms of organic fouling on polyamide membrane using a surface forces apparatus: implication for wastewater treatment. *Science of The Total Environment*, 622:644–654, 2018.

[35] Anna Bujacz. Structures of bovine, equine and leporine serum albumin. *Acta Crystallographica Section D: Biological Crystallography*, 68(10):1278–1289, 2012.

[36] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[37] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118 (45):11225–11236, 1996.

[38] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and

dynamics studies of proteins. *The journal of physical chemistry B*, 102(18): 3586–3616, 1998.

[39] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[40] Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry*, 13(8):952–962, 1992.

[41] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[42] Christian Mucksch and Herbert M Urbassek. Molecular dynamics simulation of free and forced bsa adsorption on a hydrophobic graphite surface. *Langmuir*, 27(21):12938–12943, 2011.

[43] Alvaro H Crevenna, Nikolaus Naredi-Rainer, Don C Lamb, Roland Wedlich-Söldner, and Joachim Dzubiella. Effects of hofmeister ions on the $\alpha$-helical structure of proteins. *Biophysical journal*, 102(4):907–915, 2012.

[44] Brian N Dominy, Dieter Perl, Franz X Schmid, and Charles L Brooks III. The effects of ionic strength on protein stability: the cold shock protein family. *Journal of molecular biology*, 319(2):541–554, 2002.

[45] Raul Araya-Secchi, Tomas Perez-Acle, Seung-gu Kang, Tien Huynh, Alejandro Bernardin, Yerko Escalona, Jose-Antonio Garate, Agustin D Martínez, Isaac E García, Juan C Sáez, et al. Characterization of a novel water pocket inside the human cx26 hemichannel structure. *Biophysical journal*, 107(3):599–612, 2014.

[46] Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler John Edward Reddy, Manuel N Melo, Sean L Seyler, Jan Domanski, David L Dotson, Sébastien Buchoux, Ian M Kenney, et al. Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2019.

[47] Naveen Michaud-Agrawal, Elizabeth J Denning, Thomas B Woolf, and Oliver Beckstein. Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, 32(10):2319–2327, 2011.

# Chapter 5

# Structural and Dynamic Insights into SARS-CoV-2 Spike Protein-Montmorillonite Interactions

## 5.1 Introduction

The year 2020 was one of a remarkable era in human history, although not a memorable one. The coronavirus disease 2019 (COVID-19) pandemic posed an unprecedented global challenge for public health and the economy. The pandemic, since its emergence, has claimed millions of lives and pushed billions into economic and social jeopardy. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is responsible for the spread of COVID-19, is highly contagious in comparison to its close relative SARS-CoV[1], which was linked to an epidemic in the year 2002[2].

The SARS-CoV-2's spike (S) glycoprotein plays a crucial role in the viral invasion of the host cell. Human angiotensin-converting enzyme 2 (hACE2) is an enzyme present in the cells of the lungs, heart, and kidney, which helps regulate the blood pressure[3, 4]. The hACE2 also acts as a host receptor for the SARS-CoV-2 virus, and the interaction between S glycoprotein and hACE2 is found to be the mechanism for the cell entry[5–7] of the virus. The receptor binding domain (RBD) in the

"up" conformation[7] among many of the functional domains of S protein is the one that binds to the hACE2. The critical role of S protein-hACE interactions in the viral hijacking makes it the primary target for many therapeutic and vaccine approaches[5, 8, 9].

Nanoparticles (NPs) are revolutionizing almost all aspects of our lives, including health care[10]. NPs show great potential in the development of therapeutic materials and targeted drug delivery. The efficacy of NPs as an efficient drug delivery medium or any other biomedical application can be attributed to their high surface-to-volume ratio, which facilitates adsorption of biomolecules and a variety of other chemical species. Hence, owing to the excellent physicochemical properties offered by the NPs, they have attracted the attention of the biomedical research community for as long as they have been discovered. In fact, NP-biomolecular interactions are studied extensively, both experimentally[11–13] and computationally[14–16]. Zhang et al.[17] used a nanosized graphene oxide sheet as a carrier of anticancer drugs for targeted delivery to MCF-7 cells and human breast cancer cells. They found that the nanocarriers delivered the drug to the target with high specificity and reduced toxicity. Nano clays, particularly montmorillonite are also a class of NPs which are extensively studied for many potential applications such as drug delivery,[18–20] MMT-organic interaction[21] and interactions with virus[22] and protein.[23–25] MMT surface belongs to the smectite class of minerals, a 2:1 phyllosilicate; an octahedral alumina layer is sandwiched between two tetrahedral silica layers. The MMT surface also undergoes isomorphic substitution of $Al^{3+}$ atoms present in the central alumina layer, with a lower charge species such as $Mg^{2+}$, which creates a net negative charge on the surface. Hence, the MMT surface can attract the positively charged species in the environment. Hence, in its natural state the MMT's chemical formula is $(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_10)(OH)_2.nH_2O$. Block et al.[22] studied the interaction between $\phi 6$ virus and MMT clay via transmission electron microscopy. They noted severe disassembly of virus and loss of envelope, which was attributed to the strong electrostatic and van der Waals forces between MMT and virus. Anderson et al. investigated the interaction between model protein Gb1 and different clay surfaces, including MMT, via MD simulations. They observed marginal influence of the presence of mineral surfaces on the protein structure, except for birnessite

surface, which showed significant perturbation of protein's coformation[23]. Köhler et al.[14] performed molecular dynamics (MD) simulations for studying fibrinogen (Fg)-(mica, graphite) interactions. They observed a weaker Fg-mica interaction than Fg-graphite. The Fg-mica interaction was driven by electrostatic effects between Fg and mica surface and the solvated counterions present on the mica surface. Kubaik et al.[15] noted in the MD simulations that the adsorption of lysozyme on silica was mainly driven by electrostatics and supported by weaker hydrophobic forces. Simulations also showed minor variation in lysozyme's conformation due to adsorption, while experiments reveal the effects of pH on the adsorption. Lecot et al. built atomistic models of different silane monolayers functionalized on silica($SiO_2$) surfaces and extensively studied their surface properties.[26] Additionally, they studied adsorption of ACE2(present as ACE2-RBD complex)[27] and streptavidin(present as streptavidin-biotin complex) onto these silane monolayer surfaces via MD simulations. Lecot et al. reported that adsorption of ACE2 on silane monolayers leads to a substantial increase in the binding energy between ACE2 and RBD, but the conformation of ACE2 is not affected much; hence, the bioactivity of ACE2 remains conserved[27]. NPs are also contributing to tackling the COVID-19 pandemic in every sense possible. Silver NPs have been hypothesized to be used against SARS CoV-2[28], as it is found to have antiviral activity in a previous study[29], and hence can help in blocking the viral entry to the host cell. Computational investigation of the interaction between S protein and model surfaces (gold and silica)[30] have shown the exposure of RBD and have some interesting findings.

In this study, we performed explicit molecular dynamics (MD) simulations to investigate the interaction between the montmorillonite (MMT) clay model surface and SARS-CoV's S protein. Specifically, the study aims to unravel the nature of S protein-MMT surface interaction and its effect on S protein's structure. Also, the effect of two different counterions ($Na^+$, $Ca^{2+}$) on protein-surface interactions was investigated.

## 5.2 Methods

In total, three simulation systems were prepared : (1) S protein without any surface (control), (2) S protein-Na-MMT (3) S protein-Ca-MMT. All simulations were performed with NAMD[31] simulation package. The equilibration runs were performed in the NPT ensemble, while the production runs were executed in the NVT ensemble. The production simulations for Na-MMT and Ca-MMT were run for 120 ns, while the control run was simulated for 60 ns. Since, only time-averaged properties were evaluated for control run, which were found to be well converged within the simulated time scale. The montmorillonite surface and S protein were modeled using CLAYFF[32] and CHARMM36[33] force-field, while TIP3P[34] model was used to model water molecule. Periodic boundary conditions were applied in all directions for all the systems. The simulation box length is kept such that the minimum distance between protein and the box's edge in all directions is more than 10 $\mathring{A}$. A time step of 1 fs was used to integrate the equations of motion. van der Waals interactions were cutoff at 12 $\mathring{A}$ with a switching distance of 10 $\mathring{A}$. The long-range electrostatic interactions were evaluated using particle mesh ewald[35] (PME) summation. The bonds involving hydrogen atoms were constrained with SHAKE[36] algorithm.

### 5.2.1 Protein-Surface Setup

The fully glycosylated structure of S trimer obtained from CHARMM-GUI's[37] COVID-19 archive was used as the starting structure for the simulations. The S trimer structure prepared in CHARMM-GUI is a head-only model with 1146 residues and was modeled using the Cryo-EM structure[38] (PDB ID: 6VXX). The titratable residues of protein were assigned correct protonation states assuming a pH of 7.0. The protein was then solvated in a water box of size 185 $\mathring{A}$ × 91.5 $\mathring{A}$ × 199 $\mathring{A}$ with SOLVATE plugin in visual molecular dynamics[39] (VMD). The glycosylated S protein trimer contains a charge of -15e; hence an equal number of counterions ($Na^+$) were added to the solvation box, which totaled the system to 670461 atoms. The solvated and ionized system is then subjected to the minimization of 10,000 steps with the conjugate gradient method. The step was followed by a run of 200 ps, in which the temperature is increased slowly in steps from 0 K to 310 K in an

NVT ensemble. The backbone atoms of protein were frozen, while the side chain and glycan atoms were harmonically constrained with force constant of 0.5 Kcal/mol during the above two steps. In the next step, all constraints were released, further the system was minimized for 5000 steps and was thermalized incrementally from 0 K to 310 K in an NPT ensemble for 5 ns. The system is then submitted for the production run in an NVT ensemble as a control system.

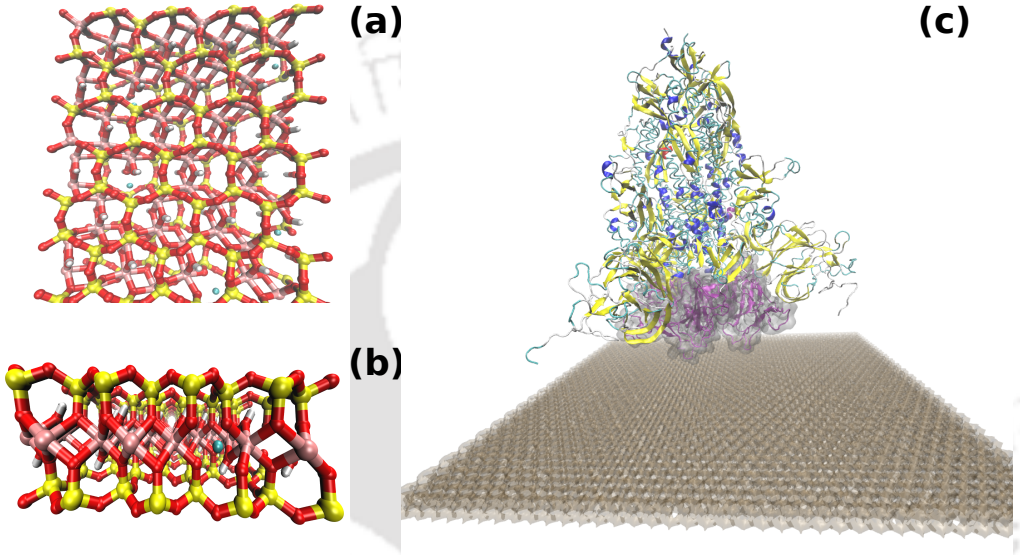



Figure 5.1: (a) Top view and (b) side view of atomistic model of MMT surface represented in licorice representation. Yellow, red, pink and cyan colors represent silicon, oxygen, aluminium and magnesium respectively.(c) Atomic model of simulation system with S protein and MMT surface in new cartoon and quick surf representation, respectively, while RBD of S protein is highlighted in Surf drawing.

The atomistic model of montmorillonite (MMT) surface was designed with Atomistic Topology Operations in MATLAB's atom[40] package. A unit cell with a basic formula as $Al_2Si_4O_{10}(OH)_2$ is replicated in X and Y directions. Then about two-thirds of the central aluminium atoms were replaced with $Mg^{2+}$ atoms in the obtained crystal lattice to mimic the isomorphic substitution of an MMT clay surface (Figure 5.1a and 5.1b). The constructed MMT surface has an XY dimension of 196 $\mathring{A} \times 196\ \mathring{A}$ and a charge of -628e.

To create S protein-MMT systems, the coordinates of equilibrated S protein from the control equilibration run (described in the first part of this section) were taken

as the initial structure of S protein. The protein was then solvated in a water box of size 196 $\mathring{A}$ × 196 $\mathring{A}$ × 188 $\mathring{A}$. The system was then ionized with counterions, taking the protein and MMT charges into account. In this step, two systems were created, namely Ca-MMT and Na-MMT, with $Ca^{2+}$ and $Na^{+}$ as counterion species, respectively. The total number of atoms in Ca-MMT and Na-MMT were 736518 and 735876, respectively. The boxes were then placed on top of the MMT surface (Figure 5.1). The protein is positioned so that the RBD is facing the MMT surface (Figure 5.1), and the minimum protein to surface distance is 8$\mathring{A}$. Figure 5.2 shows the electric dipole moment vector for RBDs of all three chains in both(Na-MMT, Ca-MMT) the systems at the end of the simulation. The electric dipole moments are in agreement with a similar work dealing with the adsorption of a negatively charged protein on a negatively charged surface.[41] The merged system is then minimized for 2500 steps, followed by an NVT run for 200 ps with protein and surface atoms fixed. Further, the system is subjected to an unconstrained minimization of 2500 steps and stepwise temperature increment in an NPT ensemble for 2 ns. Further, the two systems (Ca-MMT, Na-MMT) were submitted for the production runs.

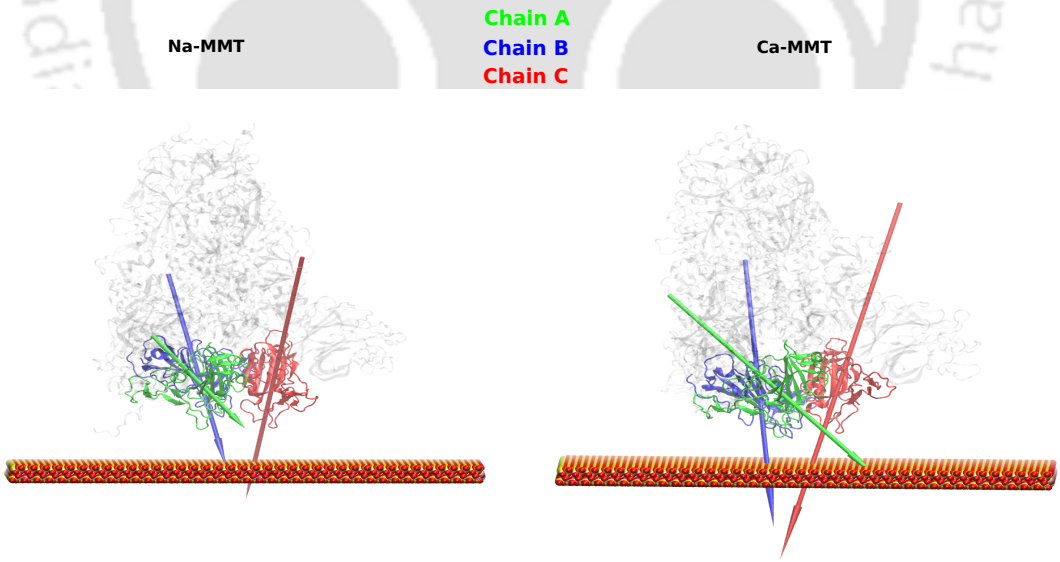



Figure 5.2: Representation of dipole moment of RBD of chain A, chain B and chain C in Na-MMT and Ca-MMT. The dipole moment vectors are color coded green, blue and red for chain A, chain B and chain C respectively. The rest part of the S protein is shown in ghost representation Δ.

### 5.2.2 Analysis

All the analyses was performed using tools present in VMD and NAMD and in-house Tcl scripts. Root-mean-square fluctuation(RMSF) for residues of S protein was calculated to quantify the protein's conformational changes using the "measure rmsf" command in the VMD. The RMSF was calculated, using the initial structure as the reference and taking into account all the frames of the trajectory. The interaction energy between S protein and MMT surface is calculated to understand the nature of interactions using the "pair-interaction" utility in NAMD. Secondary structure calculation on S protein was performed using STRIDE algorithm[42] in VMD. Contact analysis between MMT-ions, MMT-residues, and MMT glycans was performed using the "measure contact" utility in VMD. The distance cut-off used for calculating the contacts was 5 $\AA$ for all the cases. In the case of MMT-ions, only those ions were considered where the ion comes in contact with MMT for at least 40 % frames of the whole trajectory. Electrostatic potential maps were calculated using PMEpot[43] module in VMD.

## 5.3 Results and discussion

### 5.3.1 Effect on Residue Flexibility

Root mean-square fluctuation (RMSF) for each chain of the S trimer (chain A, B, and C) were calculated for all the three systems (control, Ca-MMT, Na-MMT). Figure 5.3 shows the distribution plots for RMSF values of amino acid residues of all the three chains A, B, and C. Figure 5.4 shows the deviation of RMSF the deviation of RMSF values of residues in MMT systems from the values in the control system. The deviation is represented by $\Delta\text{RMSF} = (\text{RMSF})_{control} - (\text{RMSF})_{(Na-MMT/Ca-MMT)}$. An offset of 5 and -5 was added to $\Delta$RMSF values of Ca-MMT and Na-MMT, respectively. The offset is added to separate the curves of Ca-MMT and Na-MMT and bring more clarity to the visualization. Hence, the $\Delta$RMSF values above the baseline indicate a "decrease" in the RMSF value of the residue in comparison to the control system and vice versa for the value below the baseline. In general, it can be observed that due to the presence of the surface, the flexibility of the

residues has reduced, indicated by the decrease in RMSF values of many residues. As can be observed from the Figure 5.3a, control shows highest number of residues in higher RMSF value range, among all the three systems . The distribution for chain A (Figure 5.3a) shows that the curve for control system is broader and spread across higher RMSF values, while the distributions for Na-MMT and Ca-MMT are narrower and are shifted towards smaller values of RMSF. Also, the control distribution has a peak at around 1.5 $\mathring{A}$ while the other two systems have a peak at lower RMSF values. This implies there are some residues that have an increase in RMSF value due to the presence of the surface. This can also be observed from Figure 5.4a, the N terminal residue and residues in the C terminal domain have high RMSF values for the control system. Chain B seems to be the most affected by the presence of MMT surface, as the distribution (Figure 5.3b) shows a drastic increase in the number of residues with smaller RMSF values for Na-MMT and Ca-MMT. The Ca-MMT system has the highest peak for the lowest RMSF value at around 1.2 $\mathring{A}$. The control's distribution for chain B is again broader and lacks any significant peak compared to the other two systems. Notably, the residues from 250 to 600 have got affected the most (Figure 5.4b), which is the RBD region (331-528) of chain B. Chain C shows moderate changes in the presence of MMT surface(Figure 5.3c), only a few residues (Pro 251, Gly 252, Asp 253, Arg 683, Ala 684 and Arg 682) showed significant changes (Figure 5.4c).

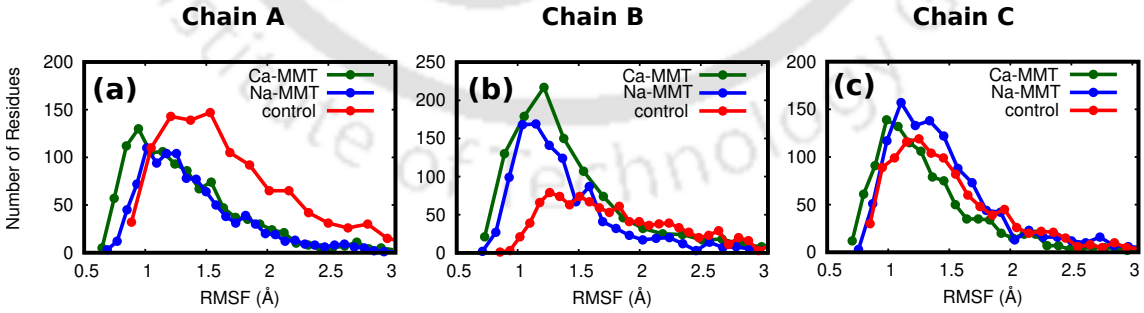



Figure 5.3: Distribution of RMSF values for (a) chain A (b) chain B and (c) chain C in control, Na-MMT and Ca-MMT systems.

Overall, the RMSF analysis shows that the MMT surface considerably affects the flexibility of all the three chains of the S protein, particularly chain B. It indicates

a strong interaction between the MMT surface and the S protein. Moreover, the RBD region of chain B indicated the strongest interaction with surface among all the chains, which is reflected by reduction in RMSF values of most residues in the presence of MMT surface (both Ca-MMT and Na-MMT). The Ca-MMT system has a stronger effect on residue flexibility than Na-MMT. Further, we characterize the nature of interaction and effects on other structural features of the protein.

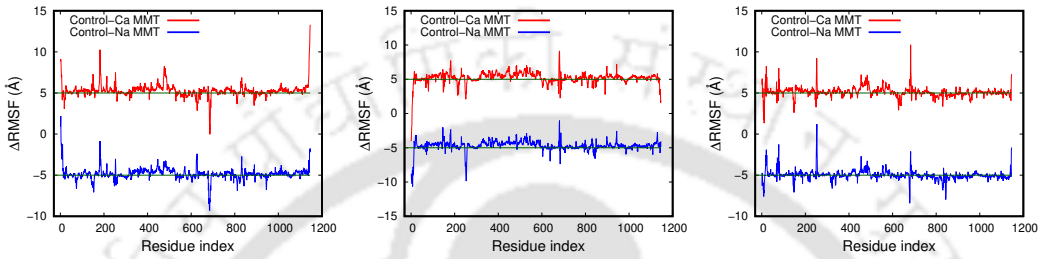



Figure 5.4: Residue-wise ΔRMSF in Ca-MMT and Na-MMT for (a) Chain A (b) Chain B and (c) Chain C. ΔRMSF of a residue represents the difference in its RMSF value in the MMT system(Na-MMT or Ca-MMT) and the control system.

### 5.3.2 Influence on Secondary structure of S Protein

Protein's structure is intimately connected to its function. Hence, changes in secondary structure of a protein can give critical insights about its interaction with the surrounding. Therefore, we evaluated the changes in the secondary structure of S protein due to its interaction with the MMT surface. Figure 5.5 shows the average probability of secondary structure contents (coil, turns, beta-sheet, and helix) for all the three systems. It can be observed that there is a heavy loss of helical content in the presence of the MMT surface as the average probability for helix reduced from 21 % in the control system to 6.4% and 6.3% in Ca-MMT and Na-MMT system respectively.

Conversely, the average probability for turns increased in Na-MMT and Ca-MMT to 39.7% for both the systems from 23.7% in the control system. The turns must have

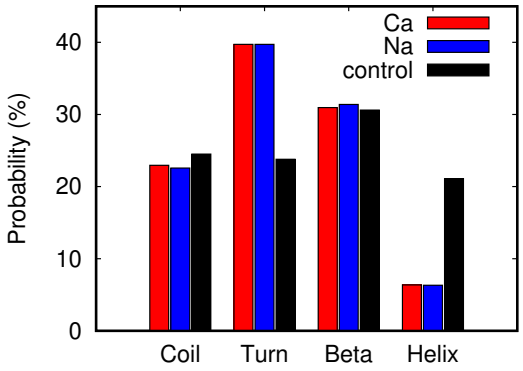
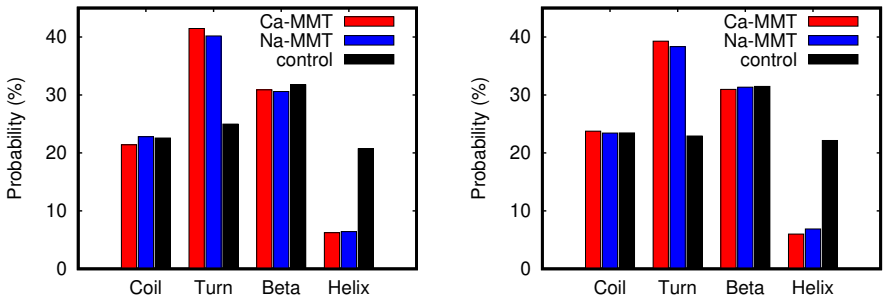



Figure 5.5: Secondary structure probablity of S protein in control, Na-MMT and Ca-MMT systems.

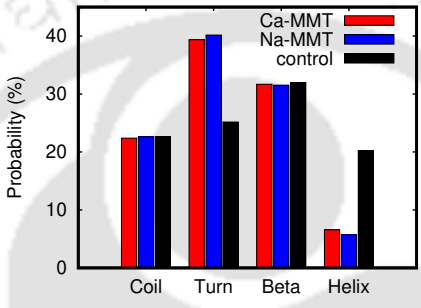
increased at the cost of helical content, i.e., the residues which lost the helical content must be the ones that got converted to turns. All other structure contents (coils and beta sheets) showed minor variations. As shown in the Figure 5.6, changes in the secondary structure of all the three chains were also similar to the changes observed in overall protein structure. Another important point to be noticed here is that the changes in secondary structure that occurred are irrespective of the counterions, i.e., changes are almost identical in both Na-MMT and Ca-MMT. The loss of helical content of the S protein must be due to strong, attractive interaction with the MMT surface. A similar observation was also made by Kubiak and Mulheran[44], where they found loss of helical content of lysozyme protein in the presence of charged Mica surface.

### 5.3.3 Nature of Protein-Surface Interaction and Influence of Counterions

To understand the nature of the interaction between the spike protein and MMT surface, time evolution of interaction energy is calculated between all three chains and MMT surface for both Na-MMT and Ca-MMT (Figure 5.7). The interaction analysis revealed electrostatics to be the dominant mode of interaction between S protein and MMT surface, which is in line with previous studies[23, 44]. However, as found in the subsequent section on "contact analysis," the protein can interact with the surface via hydrophobic contacts. As can be observed from Figure 5.7a




(a) (b)




(c)

Figure 5.6: Secondary structure probablity comparison between control, Na-MMT and Ca-MMT for (a) chain A (b) chain B and (c) chain C.

and 5.7b, all chains except chain A in the Ca-MMT system have negative interaction energy indicating attraction with the MMT surface. The average interaction energy in Ca-MMT for chains A, B, and C was 773, -1483, and -157 KJ/mol, respectively, and for Na-MMT, it is -269, -1047, and -241 KJ/mol for A, B, and C, respectively. Although the average interaction energies show attractive interaction between chains and surface, only chain B shows consistently strong attraction and is more pronounced for the Ca-MMT system. Other chains show relatively weaker interaction with the surface. The highest interaction energy between chain B and surface explains the maximum structural deviation for chain B in RMSF analysis. Since chain B is the most flexible among all three chains, which lets it interact with the MMT surface more readily, we observe the strongest interaction between chain B and the MMT surface.

MMT surface's high negative charge (-628e) is balanced by the presence of posi-

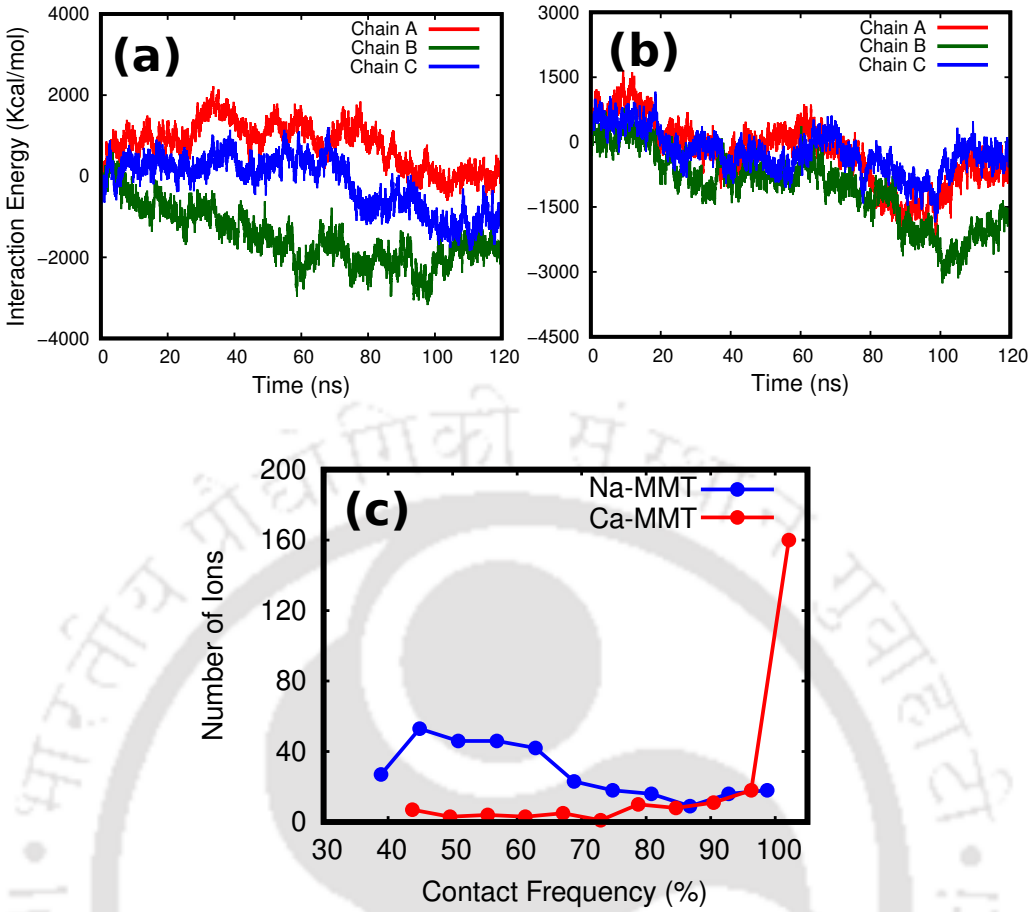



Figure 5.7: Time evolution of interation energy between chains of S protein and MMT surface in (a) Ca-MMT and (b) Na-MMT system. (c) Distribution of contact frequency of ions ($Na^+$, $Ca^{2+}$) on MMT surface in Na-MMT and Ca-MMT systems.

tively charged ions ($Na^+$, $Ca^{2+}$). The presence of ionic species is known to influence protein-surface interaction[23, 41, 45]. To quantify the extent of ion-surface interaction, we evaluated the contact frequency of ions with the surface. Figure 5.7c shows the number of ions and the corresponding percentage of contact frequency with the MMT surface for Na-MMT and Ca-MMT systems. It can be observed from Figure 5.7c that MMT surface in the Ca-MMT system has a very strong interaction with $Ca^{2+}$ ions, as indicated by a very high number of $Ca^{2+}$ ions with 100 % contact frequency. $Na^+$ ions in the Na-MMT system show weaker interaction with MMT surface, as can be observed from significantly less number of $Na^+$ ions in the high contact frequency range. However, there are a substantial number of $Na^+$ ions in the 40 to 60 % range of contact frequency, but total number of $Na^+$ ions in contact

with MMT surface are still very less in comparison to the number of $Ca^{2+}$ ions. The binding of $Ca^{2+}$ ions firmly to the MMT surface than $Na^{+}$ ions was also found in a previous study[23].

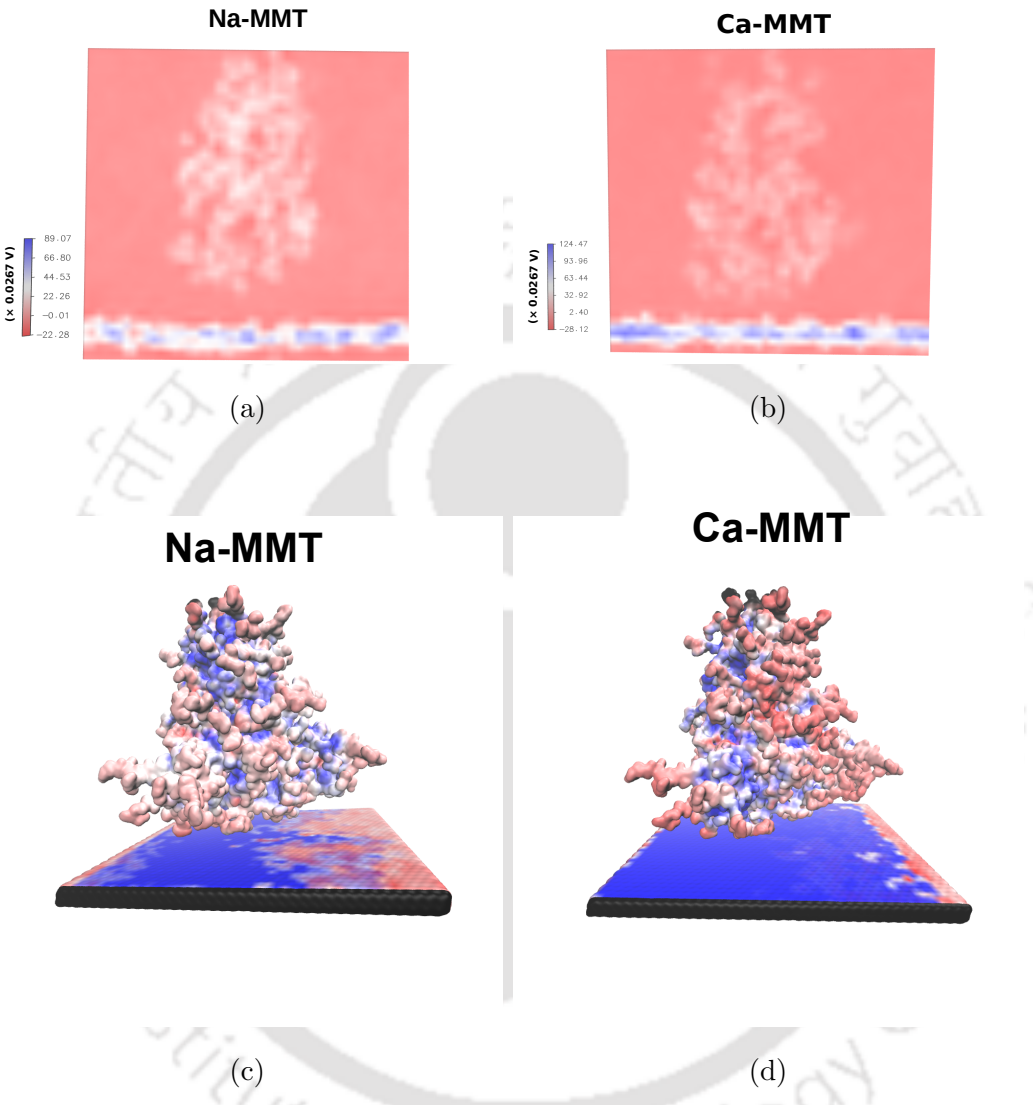



Figure 5.8: Electrostatic potential maps for (a) Na-MMT and (b) Ca-MMT systems. The color bar shows the range of colors based on potential around that region. S protein and MMT surface colored based on electrostatic potential for (c) Na-MMT and (d) Ca-MMT systems. Red, blue and white color represents negative, positive and neutral potential respectively.

Figure 5.8 shows electrostatic maps of systems. It can be observed that the Ca-MMT system (Figure 5.8b) shows higher positive potential density at the MMT surface, indicated by the blue color, than in the case of the Na-MMT system (Figure

5.8a). However, it can be observed that the region around protein in potential maps is white for both the systems and the color is much denser in the case of Na-MMT than in the Ca-MMT system. The denser white color in Na-MMT could be because the $Na^+$ ions readily diffuse from the surface to the bulk and interact with negatively charged polar residues of the protein, creating a net neutral environment around the protein. However, divalent $Ca^{2+}$ ions hold on firmly to the MMT surface and are not readily available for protein to directly interact with; therefore, a lighter white color is observed in the potential map of the Ca-MMT system. This is in agreement with Anderson et al.'s study[23], where they proposed the surrounding of Gb1 protein by $Na^+$ ions, screening the protein-surface interactions and consequently preserving the protein's conformation in the Na-MMT system. While for the Ca-MMT system, they found that the Gb1 protein's conformation was disturbed since $ca^{2+}$ ions were tightly bound to the MMT surface, and the protein is more exposed than in Na-MMT.

To examine the effect of ions on S protein's conformation, we compared the root-mean-square deviation(RMSD)(Figure 5.9) of all three chains of S protein in Na-MMT and Ca-MMT systems. However, as seen in Figure 5.9, we do not find any significant difference in the time evolution of the RMSD between the Na-MMT and Ca-MMT. Although the RMSD for chain A(Figure 5.9a) showed slight differences, they are not remarkable enough. To understand which chain is most affected due to the presence of the MMT surface, we compared the RMSD evolution of each chain within a system for Na-MMT and Ca-MMT (Figure 5.9d and 5.9e). We can observe from Figure 5.9d and 5.9e that chain B has the highest RMSD, and the RMSD curve for chain B is unstable compared to other chains in both systems. The RMSD analysis indicates that chain B conformation is the most affected in the presence of MMT surface. Moreover, structural representation colored with electrostatic potential shows the distribution of potential for the systems (Figure 5.8c and 5.8d). It can be observed that the surface in the Ca-MMT system (5.8d) shows more part of it covered with blue color than Na-MMT system (5.8c), which indicate positive potential and must be because of stronger adsorption of $Ca^{2+}$ ions than $Na^+$ ions on the MMT surface. Also, a difference can be observed in the distribution of potential on protein's structure between both systems. The colors

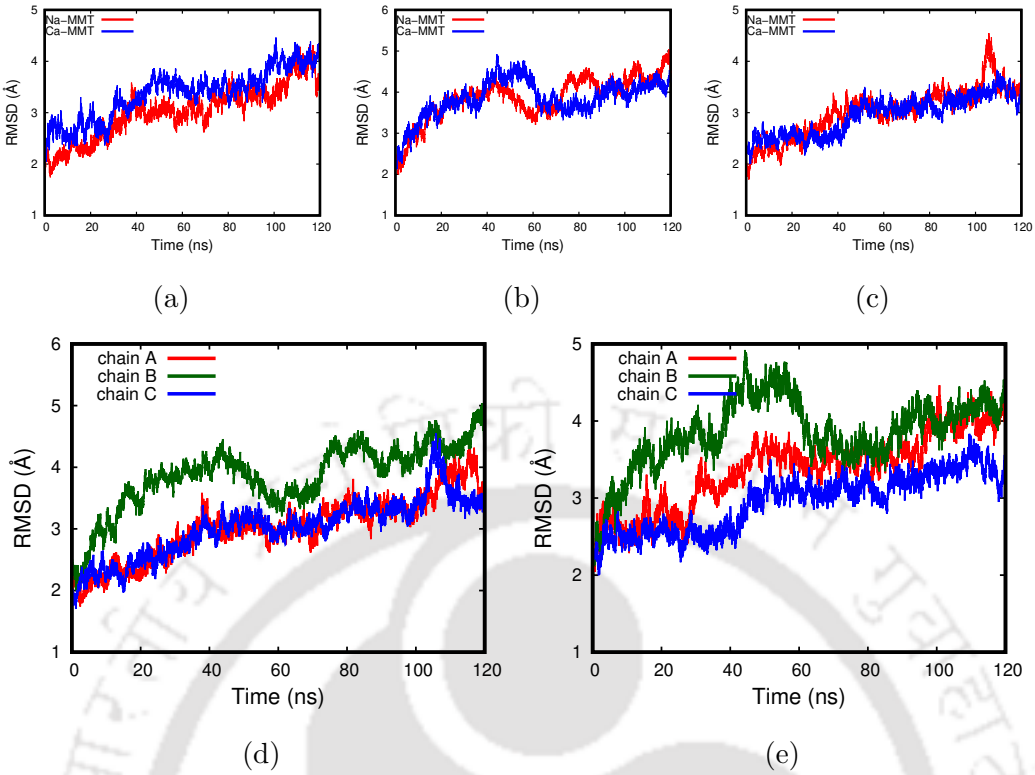



Figure 5.9: Comparison of time evolution of RMSD between Na-MMT and Ca-MMT for (a) chain A (b) chain B and (c) chain C and between all three chains within a system for (d) Na-MMT (e) Ca-MMT.

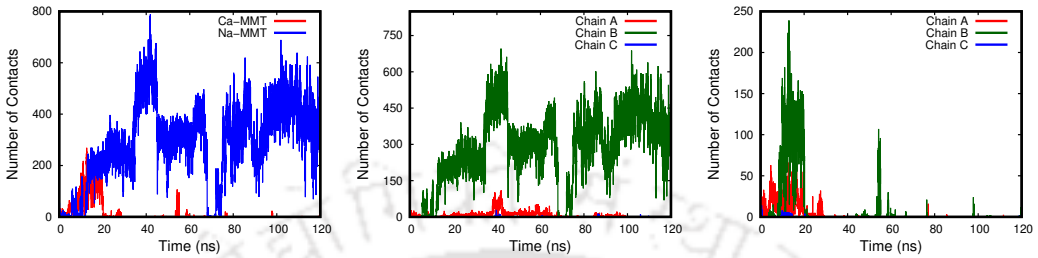
of some regions of protein's structure in the Ca-MMT system (5.8d) are bright red, while the same region in the Na-MMT system (5.8c) are paler in color. This is again due to $Na^+$ encompassing the protein, while in the Ca-MMT system, most of $Ca^{2+}$ ions are strongly adsorbed on the MMT surface.

### 5.3.4 Interacting Residues and Glycans

Contact analysis is performed to identify the critical residues and glycans of S protein closely involved in the interaction with the MMT surface. We searched for all the pairs of atoms, with one atom from S and the other from MMT, which are within $5\mathring{A}$ of each other.

Figure5.10 shows the time evolution of the number of contacts between protein atoms and MMT surface atoms. It can be observed from Figure 5.10a that protein in the Na-MMT system shows much higher and sustained contact with MMT surface

than in Ca-MMT. Figure 5.10b and 5.10c shows contact number for each protein chains in Na-MMT and Ca-MMT respectively. It can be noted that the maximum contribution for protein-MMT surface contact is from chain B in the case of Na-MMT.




(a) (b) (c)

Figure 5.10: (a) Time evolution of number of contacts between all protein atoms and MMT surface atoms for Na-MMT and Ca-MMT system. Comparison of contact number evolution with time for all protein chains for (b) Na-MMT (c) Ca-MMT.

At the same time, in the case of Ca-MMT also, chain B shows the maximum number of contacts, but chain A also shows a substantial contribution. We calculated the residue-wise number of contacts to know the type of residues that play a dominant role in the close interaction between S protein and MMT surface and consequently the nature of interactions. In our simulations, we found residues from the N terminal region and some residues from RBD of S protein in close contact with the MMT surface. Figure 5.11a and 5.11b shows the number of contacts of interacting residues for chain A and chain B, respectively. It can be noted from Figure 5.11a that for chain A, VAL445, which belongs to the RBD region of S protein, has the highest number of contacts in both the systems (Na-MMT, Ca-MMT). The GLY446 is another residue that shows contact with the MMT surface in chain A for both systems. LYS444 residue of chain A in the Na-MMT system also shows a significant number of contacts. The interacting residues of chain B show much higher contacting frequency (almost ten times) (Figure 5.11b) than Chain A.

Moreover, the number of interacting residues in chain B is more than in chain A. In the case of Na-MMT, all interacting residues of chain B belong to the N

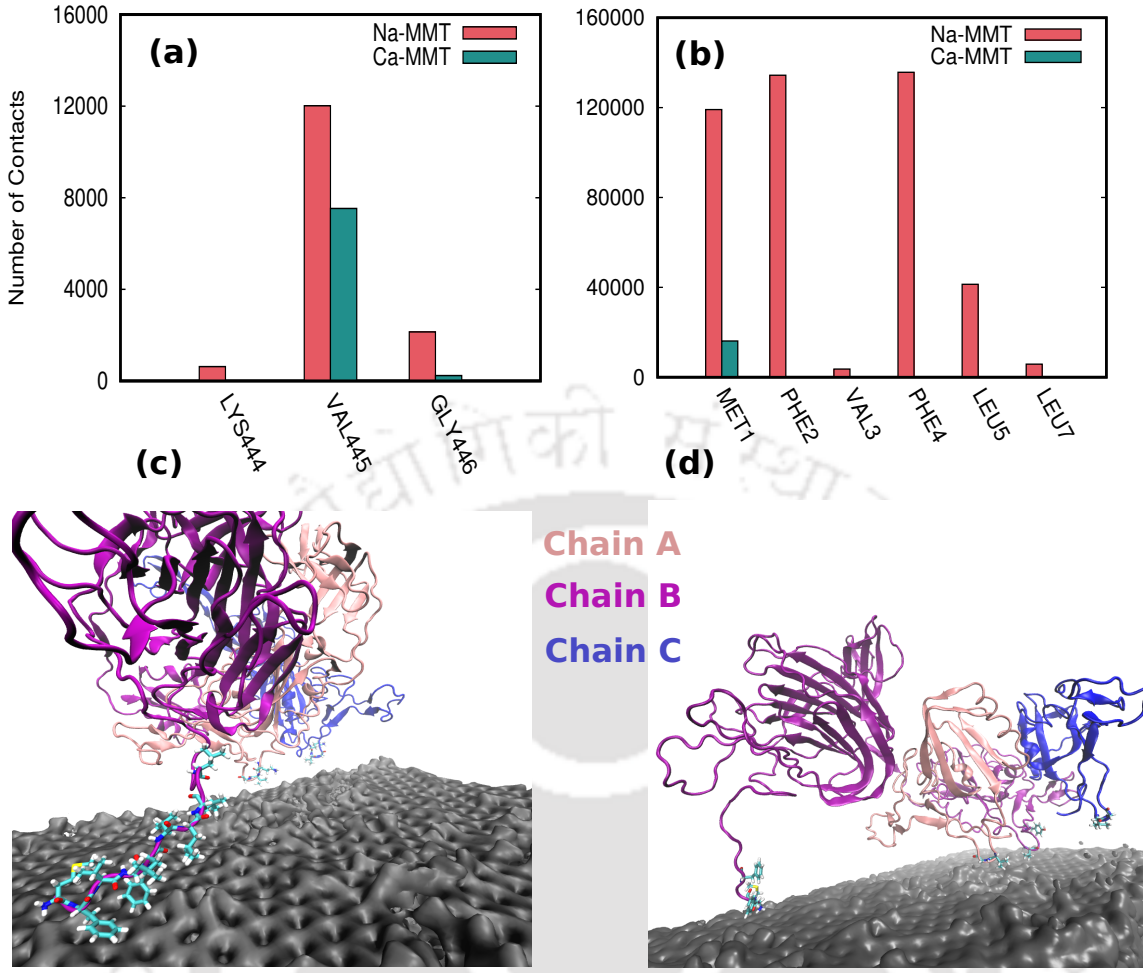


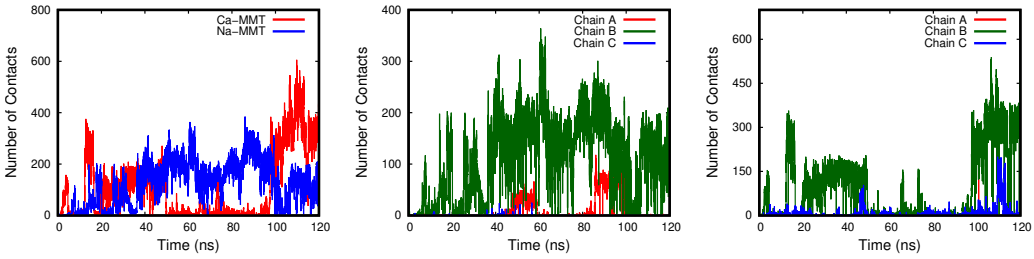

Figure 5.11: Number of contacts between key interacting residues and MMT surface in Na-MMT and Ca-MMT system for (a) chain A and (b) chain B. Representative snapshot of interacting residues and MMT surface in (c) Na-MMT and (d) Ca-MMT system. The chains and interacting residues are shown in new cartoon and licorice and MMT surface is shown in quick surf representation.

terminal. Residues with the highest number of contacts in chain B for the Na-MMT system are MET1, PHE2, PHE4, and LEU5, while VAL3 and LEU7 also show a significant number of contacts. However, chain B in Ca-MMT has only MET1 showing noticeable contact; however, VAL 445 also shows some degree of contact towards the surface(Table 5.2b). Figure 5.11c and 5.11d show the representative snapshot of Na-MMT and Ca-MMT system's trajectory. Table 5.2 shows the list of all interacting residues for Na-MMT(Table 5.2a) and Ca-MMT(Table 5.2b). It can be observed from the contact analysis of S protein and MMT surface that most of the

interacting residues are hydrophobic, such as methionine, valine, phenylalanine, and leucine. It implies the presence of hydrophobic interaction between these residues and the MMT surface. Moreover, the presence of phenylalanine, which contains an aromatic side chain, indicates a possibility of cation-$\pi$ interaction with ions($Na^+$, $Ca^{2+}$) adsorbed on top of the MMT surface. Also, S protein residues show much higher contact in the Na-MMT system than Ca-MMT system. The reason must be less interference of surface cations in the case of Na-MMT since, as discussed in the earlier section, the $Ca^{2+}$ ions are strongly adsorbed and have broader coverage of MMT surface than $Na^+$ in the Na-MMT system. Hence, residues in the Na-MMT system have more surface available for interaction.

The S protein is extensively glycosylated, and the glycans are known to play a crucial role in viral cell entry[46–48]. Since S protein is densely covered with glycans, it can play an important role in the interaction between S protein and the MMT surface. Hence, we calculated the number of contacts and identified glycans with high contact frequency with the MMT surface. Figure 5.12a shows the time evolution of the total number of contacts between glycans and MMT surface for Ca-MMT and Na-MMT. The time evolution of the number of contacts for glycans shows similar behavior to that of protein chains, as the glycans in Na-MMT show higher and more sustained contacts than in Ca-MMT. We also examined the chain-wise time evolution of contacts of glycans. It can be observed that glycans attached to chain B have the maximum number of contacts in both Na-MMT (Figure 5.12b) and Ca-MMT (Figure 5.12c), as observed similarly in protein residue-MMT contacts.

Next, to identify the key interacting glycans, we computed the number of contacts for each glycan. Figure 5.13a and 5.13b shows the number of contacts of the key interacting glycans for chain B and chain C, respectively. We found N17 from N terminal domain (NTD), N165, and N343 from RBD in the case of chain B to be highly interactive with MMT surface (Figure 5.13a) for both the systems. However, N17 in Na-MMT shows much higher contacts than in Ca-MMT, while N165 exhibits more contacts in Na-MMT than in Ca-MMT. Similarly, N149 and N165 of chain C show high contacting affinity for the Ca-MMT system, while chain C in Na-MMT also have these two glycans coming in contact with the MMT surface, but the frequency is much lesser (Table 5.1a).




(a) (b) (c)

Figure 5.12: (a) Time evolution of number of contacts between all glycan atoms and MMT surface atoms for Na-MMT and Ca-MMT system. Comparison of contact number evolution with time for glycan atoms of each chain for (b) Na-MMT (c) Ca-MMT.

However, chain A has only one glycan having contact with the surface per system (Table 5.1), with only N165 in Na-MMT showing high contact frequency (Table 5.1a). In particular, N165 is a key interacting glycan across all the chains in both systems. N165 was also found to play an important role in the interaction of S protein with ACE2 and antibodies in previous studies[48, 49]. Also, N343, which is found to be a key interacting glycan, was found to play a critical role in the binding of various antibodies to S protein[49]. Figure 5.13c and 5.13d show representative snapshots of the interacting glycans along with the MMT surface for the Na-MMT and Ca-MMT system, respectively.

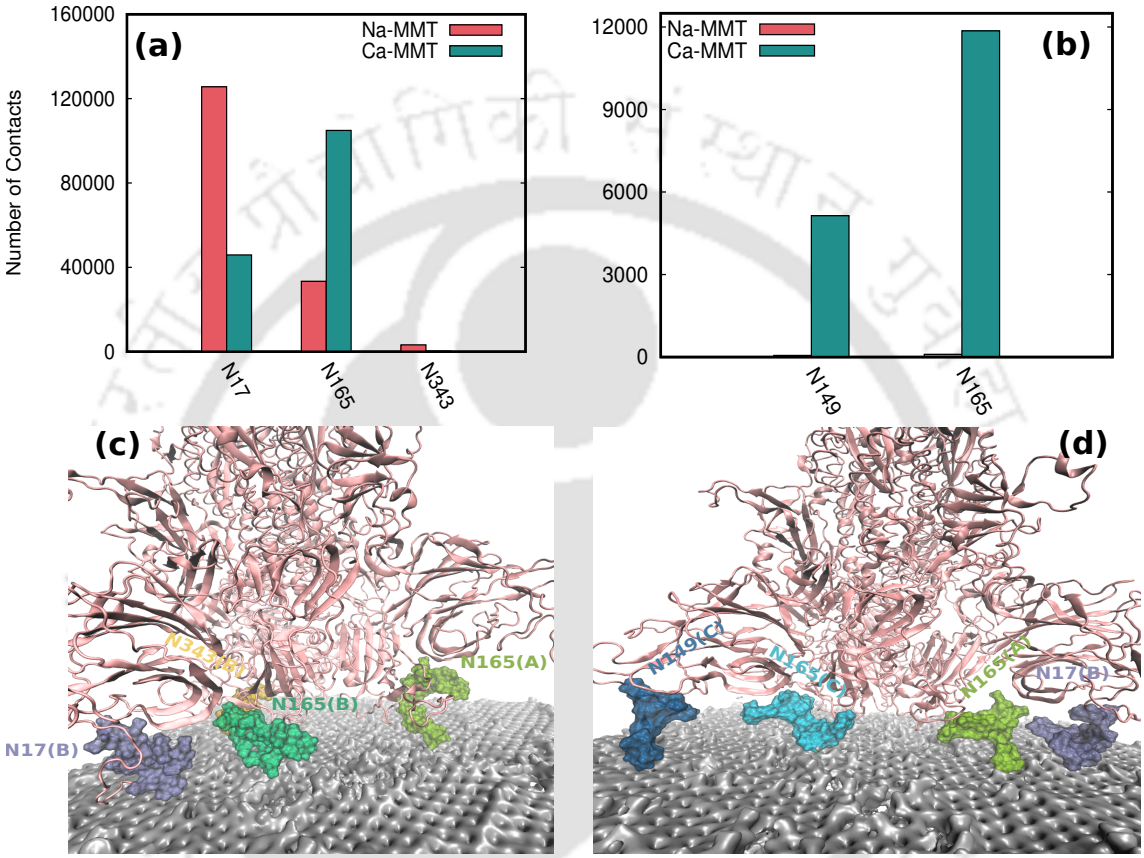



Figure 5.13: Number of contacts between key interacting glycans and MMT surface in Na-MMT and Ca-MMT system for (a) chain B and (b) chain C. Representative snapshot of interacting glycans and MMT surface in (c) Na-MMT and (d) Ca-MMT system. The S protein, interacting glycans and MMT surface are shown in new cartoon, licorice and quick surf representation.

| Glycan | Contacts | Chain |
|---|---|---|
| N165 | 19160 | A |
| N17 | 125601 | B |
| N149 | 15 | B |
| N165 | 33389 | B |
| N343 | 3254 | B |
| N17 | 60 | C |
| N165 | 100 | C |

(a)

| Glycan | Contacts | Chain |
|---|---|---|
| N165 | 207 | A |
| N17 | 4587 | B |
| N165 | 104904 | B |
| N343 | 318 | B |
| N149 | 5140 | C |
| N165 | 11860 | C |

(b)

Table 5.1: Glycans closely interacting with MMT surface, number of contacts they have with MMT surface throughout the trajectory and the chain of S protein to which they belong to for (a) Na-MMT (b) Ca-MMT.

| Residue | Contacts | Chain |
|---|---|---|
| LYS444 | 625 | A |
| VAL445 | 12015 | A |
| GLY446 | 2141 | A |
| GLY447 | 14 | A |
| TYR449 | 13 | A |
| GLN498 | 17 | A |
| THR500 | 19 | A |
| MET1 | 119140 | B |
| PHE2 | 134420 | B |
| VAL3 | 3605 | B |
| PHE4 | 135708 | B |
| LEU5 | 41350 | B |
| VAL6 | 1323 | B |
| LEU7 | 5808 | B |
| LEU8 | 344 | B |
| LEU10 | 27 | B |
| VAL445 | 13 | B |
| VAL445 | 152 | C |
| GLY446 | 57 | C |
| THR500 | 12 | C |

(a)

| Residue | Contacts | Chain |
|---|---|---|
| VAL445 | 7530 | A |
| GLY446 | 234 | A |
| MET1 | 16103 | B |
| PHE2 | 26 | B |
| PHE4 | 29 | B |
| VAL445 | 491 | B |
| GLY446 | 13 | B |
| TYR449 | 16 | B |
| VAL445 | 178 | C |
| GLY446 | 44 | C |

(b)

Table 5.2: Residues closely interacting with MMT surface, number of contacts they have with MMT surface throughout the trajectory and the chain of S protein to which they belong to for (a) Na-MMT (b) Ca-MMT.

## 5.4 Conclusions

In this work, via MD simulations, we investigated fully glycosylated S protein-MMT surface interaction and the effect of two different counterions, $Na^+$ and $Ca^{2+}$, on them. The structural analysis revealed a reduction in flexibility of many residues and heavy loss of helical content in the S protein in the presence of the MMT surface. Chain B of all three chains of S protein was found to have the strongest interaction with the MMT surface. The nature of the interaction between S protein and MMT surface was found to be predominantly electrostatic. The interaction energy analysis revealed stronger electrostatic interaction in the case of the Ca-MMT system. However, we also found many neutral hydrophobic residues, such as valine, methionine, leucine, and phenylalanine, closely interacting with the MMT surface, indicating the presence of hydrophobic interaction between S protein and MMT surface. These hydrophobic interactions were found to be more pronounced in the Na-MMT system. Also, The $Ca^{2+}$ ions showed stronger adsorption than $Na^+$ on the MMT surface, which impacts the mechanism of protein-surface interaction in Na-MMT and Ca-MMT. Additionally, we found many N-glycans closely interacting with the MMT surface. Particularly, N165 and N343 among the interacting residues were found to have a prominent role in S protein-antibody and S protein-ACE2 interactions in recent studies. Overall, our work provides detailed atomistic information about the nature of S protein-MMT interaction, which can lead to the design of better therapeutic agents in the future.

# Bibliography

[1] Coronaviridae Study Group of the International et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nature microbiology*, 5(4):536, 2020.

[2] Christian Drosten, Stephan Günther, Wolfgang Preiser, Sylvie Van Der Werf, Hans-Reinhard Brodt, Stephan Becker, Holger Rabenau, Marcus Panning, Larissa Kolesnikova, Ron AM Fouchier, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England journal of medicine*, 348(20):1967–1976, 2003.

[3] Mary Donoghue, Frank Hsieh, Elizabeth Baronas, Kevin Godbout, Michael Gosselin, Nancy Stagliano, Michael Donovan, Betty Woolf, Keith Robison, Raju Jeyaseelan, et al. A novel angiotensin-converting enzyme–related carboxypeptidase (ace2) converts angiotensin i to angiotensin 1-9. *Circulation research*, 87(5):e1–e9, 2000.

[4] Inge Hamming, Wim Timens, MLC Bulthuis, AT Lely, GJ van Navis, and Harry van Goor. Tissue distribution of ace2 protein, the functional receptor for sars coronavirus. a first step in understanding sars pathogenesis. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 203(2):631–637, 2004.

[5] Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, et al. Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *cell*, 181(2):271–280, 2020.

[6] Wenhui Li, Michael J Moore, Natalya Vasilieva, Jianhua Sui, Swee Kee Wong, Michael A Berne, Mohan Somasundaran, John L Sullivan, Katherine Luzuriaga, Thomas C Greenough, et al. Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus. *Nature*, 426(6965):450–454, 2003.

[7] Daniel Wrapp, Nianshuang Wang, Kizzmekia S Corbett, Jory A Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S Graham, and Jason S McLel-

lan. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science*, 367(6483):1260–1263, 2020.

[8] Changhai Lei, Kewen Qian, Tian Li, Sheng Zhang, Wenyan Fu, Min Ding, and Shi Hu. Neutralization of sars-cov-2 spike pseudotyped virus by recombinant ace2-ig. *Nature communications*, 11(1):1–5, 2020.

[9] Vanessa Monteil, Hyesoo Kwon, Patricia Prado, Astrid Hagelkrüys, Reiner A Wimmer, Martin Stahl, Alexandra Leopoldi, Elena Garreta, Carmen Hurtado Del Pozo, Felipe Prosper, et al. Inhibition of sars-cov-2 infections in engineered human tissues using clinical-grade soluble human ace2. *Cell*, 181(4):905–913, 2020.

[10] Ibrahim Khan, Khalid Saeed, and Idrees Khan. Nanoparticles: Properties, applications and toxicities. *Arabian journal of chemistry*, 12(7):908–931, 2019.

[11] Tommy Cedervall, Iseult Lynch, Stina Lindman, Tord Berggård, Eva Thulin, Hanna Nilsson, Kenneth A Dawson, and Sara Linse. Understanding the nanoparticle–protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. *Proceedings of the National Academy of Sciences*, 104(7):2050–2055, 2007.

[12] Susanne Schöttler, Greta Becker, Svenja Winzen, Tobias Steinbach, Kristin Mohr, Katharina Landfester, Volker Mailänder, and Frederik R Wurm. Protein adsorption is required for stealth effect of poly (ethylene glycol)-and poly (phosphoester)-coated nanocarriers. *Nature nanotechnology*, 11(4):372–377, 2016.

[13] Julie Goossens, Hla Sein, Shaohong Lu, Magdalena Radwanska, Serge Muyldermans, Yann G-J Sterckx, and Stefan Magez. Functionalization of gold nanoparticles with nanobodies through physical adsorption. *Analytical Methods*, 9(23): 3430–3440, 2017.

[14] Stephan Köhler, Friederike Schmid, and Giovanni Settanni. Molecular dynamics simulations of the initial adsorption stages of fibrinogen on mica and graphite surfaces. *Langmuir*, 31(48):13180–13190, 2015.

[15] Karina Kubiak-Ossowska, Monika Cwieka, Agnieszka Kaczynska, Barbara Jachimska, and Paul A Mulheran. Lysozyme adsorption at a silica surface using

simulation and experiment: effects of ph on protein layer structure. *Physical Chemistry Chemical Physics*, 17(37):24070–24077, 2015.

[16] Christian Mücksch and Herbert M Urbassek. Accelerated molecular dynamics study of the effects of surface hydrophilicity on protein adsorption. *Langmuir*, 32(36):9156–9162, 2016.

[17] Liming Zhang, Jingguang Xia, Qinghuan Zhao, Liwei Liu, and Zhijun Zhang. Functional graphene oxide as a nanocarrier for controlled loading and targeted delivery of mixed anticancer drugs. *small*, 6(4):537–544, 2010.

[18] Dongzhi Hou, Ruyi Gui, Sheng Hu, Yi Huang, Zuyong Feng, Qineng Ping, et al. Preparation and characterization of novel drug-inserted-montmorillonite chitosan carriers for ocular drug delivery. *Advances in nanoparticles*, 4(03):70, 2015.

[19] Sarvaiya Jayrajsinh, Gauri Shankar, Yadvendra K Agrawal, and Lateef Bakre. Montmorillonite nanoclay as a multifaceted drug-delivery carrier: A review. *Journal of Drug Delivery Science and Technology*, 39:200–209, 2017.

[20] Tingting Li, Lele Zhao, Ziliang Zheng, Min Zhang, Yidan Sun, Qingping Tian, and Shuqiu Zhang. Design and preparation acid-activated montmorillonite sustained-release drug delivery system for dexibuprofen in vitro and in vivo evaluations. *Applied Clay Science*, 163:178–185, 2018.

[21] Jennifer AR Willemsen, Satish CB Myneni, and Ian C Bourg. Molecular dynamics simulations of the adsorption of phthalate esters on smectite clay surfaces. *The Journal of Physical Chemistry C*, 123(22):13624–13636, 2019.

[22] Karin A Block, Adrianna Trusiak, Al Katz, Paul Gottlieb, Alexandra Alimova, Hui Wei, Jorge Morales, William J Rice, and Jeffrey C Steiner. Disassembly of the cystovirus $\phi$6 envelope by montmorillonite clay. *MicrobiologyOpen*, 3(1): 42–51, 2014.

[23] Amity Andersen, Patrick N Reardon, Stephany S Chacon, Nikolla P Qafoku, Nancy M Washton, and Markus Kleber. Protein–mineral interactions: molecular dynamics simulations capture importance of variations in mineral surface composition and structure. *Langmuir*, 32(24):6194–6209, 2016.

[24] Pu Chen and Lina Zhang. Interaction and properties of highly exfoliated soy

protein/montmorillonite nanocomposites. *Biomacromolecules*, 7(6):1700–1706, 2006.

[25] Ali Assifaoui, Lucie Huault, Cyrielle Maissiat, Chloé Roullier-Gall, Philippe Jeandet, Jérôme Hirschinger, Jésus Raya, Maguy Jaber, Jean-François Lambert, Philippe Cayot, et al. Structural studies of adsorbed protein (betalactoglobulin) on natural clay (montmorillonite). *RSC advances*, 4(105):61096–61103, 2014.

[26] Solène Lecot, Antonin Lavigne, Zihua Yang, Thomas Géhin, Claude Botella, Vincent Jousseaume, Yann Chevolot, Magali Phaner-Goutorbe, and Christelle Yeromonahos. Arrangement of monofunctional silane molecules on silica surfaces: Influence of alkyl chain length, head-group charge, and surface coverage, from molecular dynamics simulations, x-ray photoelectron spectroscopy, and fourier transform infrared spectroscopy. *The Journal of Physical Chemistry C*, 124(37):20125–20134, 2020.

[27] Solène Lecot, Yann Chevolot, Magali Phaner-Goutorbe, and Christelle Yeromonahos. Curious binding energy increase between the receptor-binding domain of the sars-cov-2 spike protein and angiotensin-converting enzyme 2 adsorbed on a silane monolayer from molecular dynamics simulations. *The Journal of Physical Chemistry B*, 125(39):11078–11090, 2021.

[28] Subhasish Sarkar. Silver nanoparticles with bronchodilators through nebulisation to treat covid 19 patients. *Journal of Current Medical Research and Opinion*, 3(04):449–450, 2020.

[29] Rania S El-Mohamady, TA Ghattas, MF Zawrah, and YGM Abd El-Hafeiz. Inhibitory effect of silver nanoparticles on bovine herpesvirus-1. *International journal of veterinary science and medicine*, 6(2):296–300, 2018.

[30] Linda Cerofolini, Marco Fragai, Claudio Luchinat, and Enrico Ravera. Orientation of immobilized antigens on common surfaces by a simple computational model: Exposition of sars-cov-2 spike protein rbd epitopes. *Biophysical chemistry*, 265:106441, 2020.

[31] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant

Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[32] Randall T Cygan, Jian-Jie Liang, and Andrey G Kalinichev. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *The Journal of Physical Chemistry B*, 108(4):1255–1266, 2004.

[33] Jing Huang and Alexander D MacKerell Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of computational chemistry*, 34(25):2135–2145, 2013.

[34] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[35] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n· log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[36] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[37] Jumin Lee, Xi Cheng, Jason M Swails, Min Sun Yeom, Peter K Eastman, Justin A Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, et al. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of chemical theory and computation*, 12(1):405–413, 2016.

[38] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, 181(2):281–292, 2020.

[39] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[40] Michael Holmboe. Atom: A matlab package for manipulation of molecular systems. *Clays and clay minerals*, 67(5):419–426, 2019.

[41] Karina Kubiak-Ossowska, Barbara Jachimska, and Paul A Mulheran. How negatively charged proteins adsorb to negatively charged surfaces: a molecular

dynamics study of bsa adsorption on silica. *The Journal of Physical Chemistry B*, 120(40):10463–10468, 2016.

[42] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4): 566–579, 1995.

[43] Aleksij Aksimentiev and Klaus Schulten. Imaging $\alpha$-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophysical journal*, 88(6):3745–3761, 2005.

[44] Karina Kubiak and Paul A Mulheran. Molecular dynamics simulations of hen egg white lysozyme adsorption at a charged solid surface. *The Journal of Physical Chemistry B*, 113(36):12189–12200, 2009.

[45] Solène Lecot, Antonin Lavigne, Zihua Yang, Yann Chevolot, Magali Phaner-Goutorbe, and Christelle Yeromonahos. Effects of the chemical and structural properties of silane monolayers on the organization of water molecules and ions at interfaces, from molecular dynamics simulations. *Langmuir*, 37(18):5563–5572, 2021.

[46] Yasunori Watanabe, Thomas A Bowden, Ian A Wilson, and Max Crispin. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1863(10):1480–1497, 2019.

[47] Rahul Raman, Kannan Tharakaraman, V Sasisekharan, and Ram Sasisekharan. Glycan–protein interactions in viral pathogenesis. *Current opinion in structural biology*, 40:153–162, 2016.

[48] Lorenzo Casalino, Zied Gaieb, Jory A Goldsmith, Christy K Hjorth, Abigail C Dommer, Aoife M Harbison, Carl A Fogarty, Emilia P Barros, Bryn C Taylor, Jason S McLellan, et al. Beyond shielding: the roles of glycans in the sars-cov-2 spike protein. *ACS central science*, 6(10):1722–1734, 2020.

[49] Yiwei Cao, Yeol Kyo Choi, Martin Frank, Hyeonuk Woo, Sang-Jun Park, Min Sun Yeom, Chaok Seok, and Wonpil Im. Dynamic interactions of fully glycosylated sars-cov-2 spike protein with various antibodies. *Journal of chemical theory and computation*, 17(10):6559–6569, 2021.

# Chapter 6

# Role of Binding Site Specificity in the Disaggregation of A$\beta_{42}$ Fibrils via Synthetic Paratope

## 6.1 Introduction

Despite decades of extensive research, Alzheimer's disease (AD) continues to pose a significant challenge in the field of neurodegenerative disorders. The exact mechanism and cause of AD remain elusive. However, one of the most common pathological hallmarks of AD is the presence of amyloid beta (A$\beta$) plaques and neurofibrillary tangles of Tau proteins[1] in the brains of AD patients. Therefore, misfolding and aggregation of A$\beta$ peptide are widely considered as one of the most accepted hypotheses for the cause of AD[1–5].

A$\beta$ peptide is produced by the cleavage of amyloid precursor protein(APP) by $\beta$ secretase and $\gamma$ secretase[2]. The splitting of APP by $\gamma$ secretase produces A$\beta$ peptides of varying lengths such as A$\beta$1-36 to A$\beta$1-43[6], A$\beta$4-42 and A$\beta$5-42[7], A$\beta$1-26 and A$\beta$1-30[8]. However, the fibrils of species A$\beta$1-40 and A$\beta$1-42 are the major constituents of the senile plaques found in AD patients' brains[2]. Among the two, A$\beta$1-42 is less abundant but is higher in toxicity and aggregation propensity than A$\beta$1-40[9]. The formation of mature A$\beta$ fibrils occurs through a complex multistep self-assembly process through a nucleation-condensation and polymerization mecha-

nism, which involves multiple intermediate metastable species such as oligomers and protofibrils, which are also toxic in nature[10–13]. A$\beta$ fibril also exhibits structural polymorphism demonstrated by various studies reporting models having U-shaped, S-shaped, and LS-shaped topology[14–17].

The finding "A$\beta$ fibrillation is the primary cause of AD" revolutionized the field of AD research and led all the drug development and therapeutical approaches toward finding a way to inhibit and reverse the A$\beta$ fibrillation process. These approaches involve experimental investigation of many potential inhibitors such as nanoparticles[18–20], peptides containing amyloidogenic core region(KLVFF) of A$\beta$ and other short peptides[21–23], small molecules[24–26] and antibodies[27, 28]. Furthermore, these experimental studies also inspired many simulation studies[29–33] to deepen our understanding of the interaction of the drug and A$\beta$ and the mechanism of inhibition/disaggregation of A$\beta$ fibrils at a molecular level. Lemkul et al. performed an MD simulation to study the molecular interactions involved in the destabilization of A$\beta$ fibrils by a flavonoid called morin. They found that morin can block the attachment of a new peptide by binding at the end of the preformed A$\beta$ fibril. Also, morin diffuses to the core of the A$\beta$ fibril, consequently disrupting crucial hydrophobic interactions[29]. Viet et al. used MD simulations to study the inhibition of A$\beta$ oligomerization by two breaker peptides, KLVFF and LPFFD. Although both breaker peptides showed inhibitory effects, LPFFD exhibited stronger interference with aggregation and higher binding affinity to A$\beta$16–22, attributed to favorable hydrophobic interactions[30]. Agrawal et al. employed MD simulations to study the disruption of U-shaped A$\beta$40 trimer by 12-crown-4 ether. The study revealed that the 12-crown-4 ether enters the hydrophobic core region and causes the loss of $\beta$ sheet content by interacting strongly with key hydrophobic residues. Also, it destabilizes the Asp23-Lys28 salt bridge by interacting with Lys28[31]. Zhang et al. used explicit MD simulations to study the interaction between graphene nanosheet and preformed A$\beta$ fibrils. The graphene sheet is found to be interacting strongly with A$\beta$ fibrils leading to structural deformation, particularly for residues having outer side chains. The van der Waals forces were recognized to be the dominant driving force for A$\beta$- graphene binding, while solvent was found to mediate the interaction between them[32]. Zhan et al. conducted explicit atomistic simulations to

understand the mechanism behind the disruption of A$\beta$42 protofibril by green tea extracts epigallocatechin-3-gallate(EGCG) and epigallocatechin(ECG). The work concluded that the EGCG exhibited higher disruptive capacity than EGC, owing to the presence of the gallic acid ester group in EGCG. The study sheds light on various atomistic interactions through which EGCG and EGC interact with A$\beta$42 protofibril such as $\pi$-$\pi$, cation-$\pi$, hydrophobic and hydrogen bonding interaction[33].

Recently, Paul et al.[34] designed and synthesized a flexible, hairpin-like synthetic paratope(SP) as a potential drug candidate for inhibiting the aggreagation of A$\beta$ and disaggregating the preformed fibrils. A paratope is a part of the antibody which recognizes and binds to the epitope region of the antigen. The SP's design was inspired by a peptide fragment(LVFFA) of A$\beta$. The study also tested the efficacy of the newly developed drug(SP) against the inhibition of A$\beta$ aggregation and disaggregation of preformed fibrils and the results were quite promising. Herein, we performed explicit MD simulation to investigate the disassembly mechanism of A$\beta$ fibrils via interaction with SP at a molecular level. In our simulations we observed significant disruption of A$\beta$ fibrils in the presence of SP. Specifically, we analyzed the $\beta$ sheet content of A$\beta$ fibrils and found to have a marked reduction in it due to SP binding. Moreover, the salt bridge between K28 side-chain and terminal residue A42's $COO^-$ group, a crucial stabilizing fchapter04orce for the structural integrity of A$\beta$ fibril, was found to have disrupted in SP's presence. We were able to identify locations of various binding sites of SP on A$\beta$ fibril and the participating residues of A$\beta$ and SP. In particular, hydrophobic and aromatic residues were found to be playing the central role in the binding of SP on A$\beta$ fibril. $\pi - \pi$ interactions were found to be the dominant mode of interaction between A$\beta$ fibril and SP.

## 6.2 Methods

**Control System Design.** The initial structure of the A$\beta_{42}$ protofibril was obtained from the protein data bank entry: 5OQV, which is derived using cryo-EM[16]. The chosen model contains nine chains of LS-shaped conformation of full-length A$\beta_{1-42}$ monomers. The nine chains are two multimers: tetramer and pentamer(Figure 7.1a). The nonamer(nine chains) protofibril is solvated with water molecules in a cubic box

with an edge length of 9.4 nm. The box length is designed such that the closest distance between the protofibril and box boundary is at least 1.2 nm. The solvated system is further neutralized by adding twenty-seven sodium($Na^+$) ions. The final system contains 78120 atoms in total.

**Ligand and Ligand System Design.** The structure of the ligand(SP)(Figure 7.1d) was obtained from the work by Paul et al.[34]. As shown in Figure 7.1d, the structure of SP has two copies of short peptide fragment(LVFFA) from the amyloidogenic core region of A$\beta$, connected with a flexible turn region made with adipoyl, polyethylene glycol, and succinyl, to create a $\beta$ hairpin-like structure. For a detailed account of the design and synthesis of SP, the reader is referred to the work by Paul et al.[34]. The atomistic model of SP for our MD simulations was constructed, optimized, and energy minimized using Avogadro[35]. The CGenFF program[36] was used to generate CHARMM-compatible forcefield parameters and partial charges for the SP molecule. These parameters were further refined using the forcefield toolkit[37] in the visual molecular dynamics(VMD) program[38]. Subsequently, a merged system is constructed with SP molecules placed randomly around A$\beta_{42}$ nonamer at a distance of 1.2 nm from the nonamer and from each other(Figure 7.1b). The number of SP molecules in the ligand system was kept in a ratio of 1:1 to the number of A$\beta_{42}$ chains. The merged system is solvated with water molecules in a rectangular box of dimension $9.2 \times 10.5 \times 12.7$ $nm^3$. The box dimensions were so designed that any atom of SP or A$\beta_{42}$ protofibril is at least 1.2 nm away from the simulation box's boundary. Further, 27 sodium ions were added to the system to maintain charge neutrality. The final ligand system contains 114876 atoms.

**Simulation Details.** All simulations were performed using the NAMD 2.14 MD simulation package[39]. The Charmm36 force field[40] was used for the A$\beta$ fibril and TIP3P water model[41]. All systems were energy minimized using 5000 steepest descent steps. The systems were then equilibrated for 100 ps using the canonical (NVT) ensemble, followed by a further 100 ps equilibration simulation with the isobaric-isothermic (NPT) ensemble. The production runs for all the systems were performed in the NPT ensemble. All the covalent bonds involving hydrogen atoms in water molecules were constrained using SETTLE algorithm[42], and those except water molecules were constrained with SHAKE algorithm[43]. We used a time step of

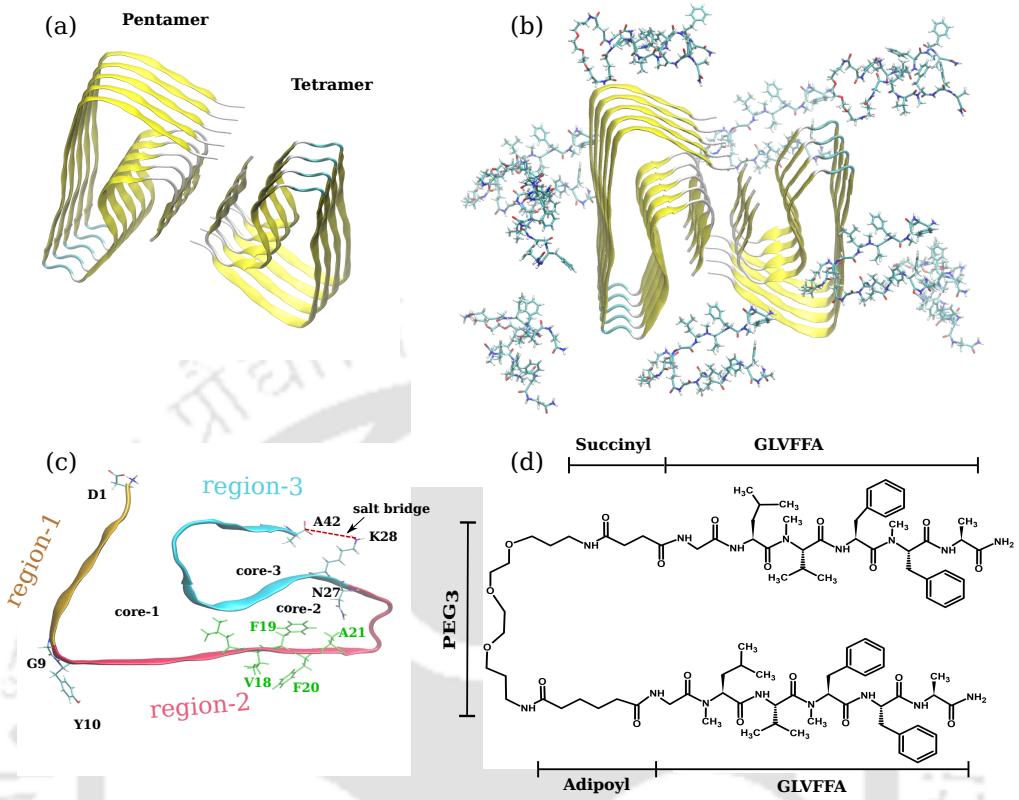



Figure 6.1: (a) LS-shaped A$\beta_{42}$ nonamer(pentamer + tetramer) protofibril obtained from cryo-EM resolved structure(PDB ID: 5OQV). (b) The initial configurtation of A$\beta_{42}$ + SP system with SP molecules randomly placed around A$\beta_{42}$ nonamer. (c) Structural features of LS-shaped A$\beta_{42}$ protofibril shown on a single chain. The protofibril structure is divided into three regions region-1(D1-G9), region-2(Y10-N27) and region-3(K28-A42) encapsulating the three hydrophobic cores. The residues shown in the green color(LVFFA) belongs to the amyloidogenic region of the A$\beta$, these residues are also the core ingredients of the ligand(SP) used in the study. The salt bridge between K28 and A42 shown by red dotted line plays a crucial role in the stability of the A$\beta_{42}$ protofibril. (d) Molecular structure of the ligand, a hairpin like synthetic paratope where two strands of the amyloidogenic region(LVFFA) of A$\beta$ are joined by a flexible chain.

2 fs for integrating the equations of motion. Particle mesh Ewald (PME) method[44] was used to calculate long-range electrostatic interactions with a real space cutoff

of 1.2 nm. The van der Waals (vdW) interactions were calculated using a cutoff of 1.2 nm. All MD simulations were performed at a temperature of 310 K, and the temperature control was achieved using a Langevin thermostat[45]. A constant pressure of 1.01325 bar was maintained in an NPT ensemble by using Nosé-Hoover Langevin barostat[46, 47].

**Analysis Methods.** All the analysis was performed using MDTraj[48], VMD[38], NAMD[39], and our in-house developed Tcl, python, and bash codes. The RMSD of C$\alpha$ atoms of A$\beta_{42}$ protofibril was computed with MDTraj. The initial structure of A$\beta_{42}$ protofibril was taken as a reference, and the RMSD for each chain in the tetramer/pentamer was calculated and averaged over all the chains in the tetramer/pentamer. The secondary structure calculations were performed with the STRIDE algorithm[49] in VMD. The K28-A42 salt bridge plays a critical role in the structural stability of A$\beta_{42}$ protofibril. The K28-A42 salt bridge is said to exist if the distance between positively charged N$\zeta$ of K28 and negatively charged C$\gamma$-carboxylate of A42 is within 0.4 nm. Residue-wise interaction energies were extracted using the *pari-interaction* facility in NAMD. All the interaction energies reported here are vdW energies unless otherwise stated. In the analysis of contacts between the residues of tetramer/pentamer(contact maps in Figure 6.5), the contacting frequency is computed between two residues separated by two peptide bonds. The contacting frequency was calculated for residue pairs in a single chain and then averaged over all the chains in the tetramer/pentamer. A contact is said to exist if the distance between two non-hydrogen atoms is within 0.5 nm. All the visualizations were performed with VMD[38].

## 6.3 Results and discussion

We have performed six independent 600 ns MD simulations for two sets of systems: 1) A$\beta_{42}$ (2) A$\beta_{42}$ + SP. Figure 7.1a and 7.1b shows the atomistic models of isolated A$\beta_{42}$ and A$\beta_{42}$ + SP systems. The A$\beta_{42}$ protofibril model used in this study is a full-length (1-42) LS-shaped A$\beta$ protofibril, which is a nonamer(containing nine chains) in the form of two separate protofibrils, one with five chains(pentamer) and another with four chains(tetramer) as shown in the Figure 7.1a. Figure 7.1c shows the

key structural features of the LS-shaped A$\beta_{42}$ protofibril model used in the present study. The overall structure of the protofibril is stabilized by three hydrophobic cores (Figure 7.1c), and the residues forming the stabilizing hydrophobic contacts inside these cores are as follows: (1) core-1: A2, F4, L34, and V36 (2) core-2: L17, E19, and I31 (3) core-3: A30, I32, M35, and V40. The salt bridge between the positively charged $NH_3^+$ group of K28's side chain and negatively charged $COO^-$(C-terminus) of A42 is also an important interaction for the A$\beta_{42}$ protofibril's structural stability. Figure 7.1d shows the chemical structure of the SP molecule. The most crucial part of the SP's structure is the sequence "LVFFA," chosen from the central hydrophobic core(CHC) region of A$\beta_{42}$ protofibril(Figure 7.1c). The reader is referred to the Methods section for a detailed discussion about the ligand design and system preparation.

**SP Disrupts the Structure of A$\beta_{42}$ Protofibril, with Severe Structural Destruction of Tetramer.** Root-mean-squared-deviation(RMSD) is a frequently used measure for assessing the structural stability of proteins, and any perturbation to the protein's structure will reflect in the RMSD. We calculated the time evolution of RMSD for $C_\alpha$ atoms of A$\beta_{42}$ protofibril for tetramer and pentamer in isolated A$\beta_{42}$ system and A$\beta_{42}$ + SP system with respect to their corresponding initial structure. From now on, we will refer to the isolated A$\beta_{42}$ system as control and A$\beta_{42}$ + SP system as ligand unless otherwise stated. Figure 6.2a, and 6.2b compares control and ligand' $C_\alpha$ RMSD of A$\beta_{42}$ for tetramer and pentamer, respectively. The RMSD analysis shows that the SP causes profound disruption of A$\beta_{42}$ protofibril's structure in the case of tetramer(Figure 6.2a). The RMSD for the ligand system in tetramer shows a sharp increase during the first 50 ns and then converges and fluctuates around 0.2 nm up to 300 ns; during this interval(50 ns - 300 ns), the RMSD does not show any drastic changes and remain in between 0.2-0.25 nm. However, after 300 ns, we observe an acute increase in the RMSD, reaching up to 0.5 nm, indicating a severe disruption of the tetramer's structure in the presence of SP. Contrarily, the tetramer's RMSD in the control system increases quickly during the beginning of the trajectory and then converges at around 0.12 nm. The RMSD suggests that the structure of A$\beta_{42}$ tetramer in the control system remains stable throughout the trajectory, except for a slight increase at around 200 ns and 370 ns, the RMSD stays

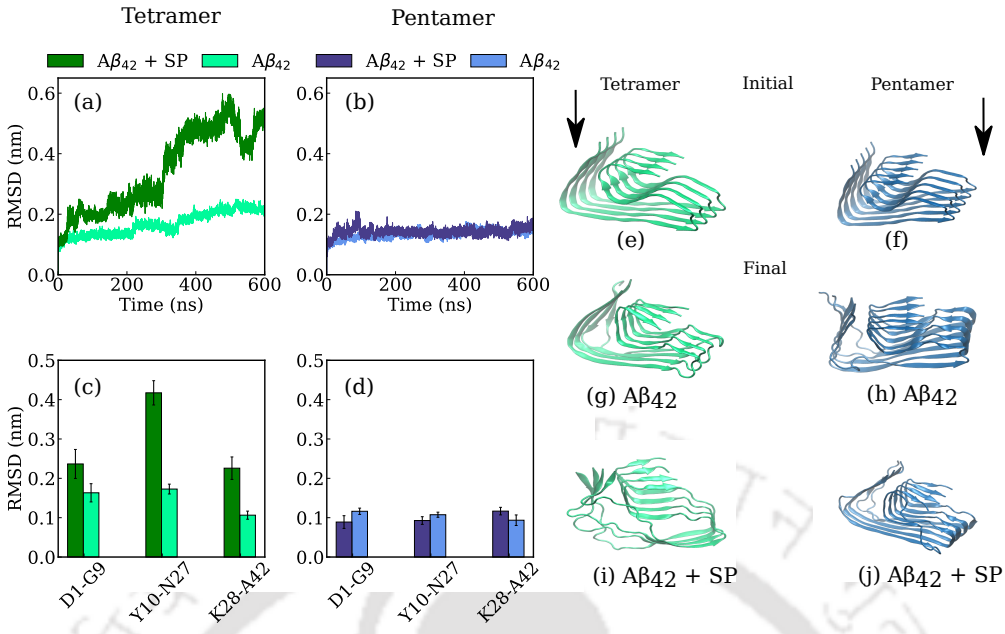



Figure 6.2: Time evolution of C$\alpha$ RMSD of (a) Tetramer in control(light green) and ligand(dark green) system (b) Pentamer in control(light blue) and ligand(dark blue) system. C$\alpha$ RMSD values of different regions of A$\beta_{42}$ protofibril averaged over last 200 ns for the (c)Tetramer in control(light green) and ligand(dark green) system (d) Pentamer in control(light blue) and ligand(dark blue) system. Initial structure of A$\beta_{42}$ protofibril of (e) Tetramer and (f) Pentamer. Structures of A$\beta_{42}$ protofibril at the end of the simulation time (g) Tetramer (h) Pentamer in the control system and (i) Tetramer (j) Pentamer in the ligand system.

within 0.12-0.19 nm. In the case of pentamer, the RMSD(Figure 6.2b), for both the control and ligand system, stabilizes at around 0.15 nm after an initial increase. The RMSD analysis for the pentamer reveals that the presence of SP does not have a marked influence on the structure since the RMSD for the ligand system shows no noticeable deviation from that of the control system. Furthermore, a comparison between the RMSD of the tetramer and pentamer in the control system shows that the pentamer's RMSD stabilizes at 0.15 nm and that of the tetramer at 0.19 nm, implying that the pentamer possesses higher inherent stability than the tetramer.

To further illuminate the effect of SP on the structural stability of A$\beta_4$2 protofibril and to explore its most affected region due to the presence of SP, we calculated the average RMSD of different regions in the protofibril's structure for pentamer

and tetramer. The three selected regions are shown in Figure 7.1c. The selected regions are region-1: D1-G9, region-2: Y10-N27, and region-3: K28-A42. Figure 6.2c shows the average RMSD for tetramer in the control and ligand system. The RMSD suggests instability in all three regions of tetramer in the presence of SP, indicated by higher RMSD in the ligand system. Interestingly, region-2(Y10-N27) in the tetramer shows the highest RMSD(0.42 nm). Region-2 contains the CHC(LVFFA) of A$\beta_4$2, which is also the key ingredient in the SP's design and is known to be a self-recognition unit. Hence, the high RMSD of region-2 indicates that the SP can recognize the CHC region and bind around it. However, in other sections, we will discuss the specific binding site of SP on A$\beta_4$2. The RMSD for D1-G9 and K28-A42 regions for control/ligand was found to be 0.18/0.22 and 0.21/0.10, respectively. RMSD analysis for tetramer shows that SP significantly destabilizes all three regions of the protofibril. Contrarily, pentamer's RMSD(Figure 7.1d) shows that SP's presence has minimal effect on the protofibril's structure. The RMSD(control/ligand) for region D1-G9(0.11/0.09) and Y10-N27(0.11/0.09) shows a slight decrement in the presence of SP, suggesting that SP is stabilizing the protofibril in the case of pentamer. Conversely, the K28-A42 region in the pentamer shows a little increase in RMSD(0.09/0.11) in the ligand system, indicating a slight perturbation of the structure due to SP. The RMSD analysis reveals that the presence of SP strongly affects the tetramer's structure, particularly the CHC-containing region-2 is severely affected. However, pentamer do not show any significant structural changes in the presence of SP. A visual analysis of the initial and final conformation of pentamer and tetramer protofibril(Figure 7.1e-j) also agrees with the results obtained by the RMSD analysis.

**SP Destroys Tetramer's $\beta$ Sheet Structure.** $\beta$ sheets are the characteristic structural features of amyloid beta fibrils. Hence, the disruption of ordered $\beta$ sheets is a crucial indicator of the efficacy of a potential drug aiming to disaggregate the A$\beta$ fibrils. In order to evaluate the impact of SP binding on the $\beta$ sheet structure of A$\beta_{42}$ protofibril, we calculated per-residue $\beta$ sheet probability for tetramer and pentamer in control and ligand systems. Interestingly, the $\beta$-sheet probability analysis of tetramer(Figure 6.3) shows a substantial loss of $\beta$-sheet content, in the presence of SP, around the region-2 of the protofibril. Specifically, residues between Y10 and

F19(Y10, E11, V12, H13, H14 and F19) shows drastic reduction in the $\beta$ sheet probablity. These observations for region-2 reveal crucial information about the SP binding and the protofibril disaggregation. The residues observing the highest reduction in the $\beta$-sheet probability are all sequential(Y10-H14), implying that the disaggregation process primarily occurs around these residues; this can be verified visually from Figure 6.2e and 6.2i. It can be observed that the region around core-1 has been disrupted severely, and the $\beta$-sheets have turned to random coils for all the chains in the tetramer. Moreover, the heavy loss of $\beta$-sheet content in sequential residues suggests that the SP preferentially attacks region-2 and probably enjoys the strongest binding around these residues(precise details of SP binding will be discussed in another section). However, it can also be observed from Figure 6.3a that there is an increase in $\beta$-sheet probability for some residues in the tetramer, such as D1, E3, E22, D23, and G29. In contrast, the $\beta$-sheet probability for pentamer(Figure 6.3b) shows an increase in $\beta$-sheet content for most of the residues in the presence of SP. The residues showing a significant increase in $\beta$-sheet probability are E22, D23, and V24, and residues between S8-H14. Intriguingly, S8-H14 is the same region that witnessed a drastic reduction in the $\beta$-sheet content for tetramer in the presence of SP. Another interesting observation here is that the $\beta$-sheet around the C-terminal region(A30-V40) remains stable and relatively unaffected by the SP's presence in case of both tetramer and pentamer. Overall, the $\beta$-sheet probability reveals interesting details about residues and the regions of A$\beta_4$2 protofibril, involved in the disaggregation process and about the probable region of SP binding.

**SP Dismantle the Salt Bridge Network in Tetramer and Pentamer, But Tetramer Shows Higher Degree of Disruption.** Salt bridges plays a prominent role in stabilizing the A$\beta$ protofibril's structure. The presence of a salt bridge as an essential structural feature has been found across all the different morphological species of A$\beta$ protofibril reported in the literature[16, 50–52]. The salt bridge in A$\beta_{42}$ protofibril(used in the present study), forms between of lysine(K28) and of alanine(A42). These salt bridges contributes to the stability of core-3 by stabilizing the C-terminus(Figure 7.1c and Figure 6.4g) of all the chains in the protofibril. To quantify the effect of SP binding on the stability of the salt bridges, we calculated the probablity density function(PDF) of the distance between the center of mass

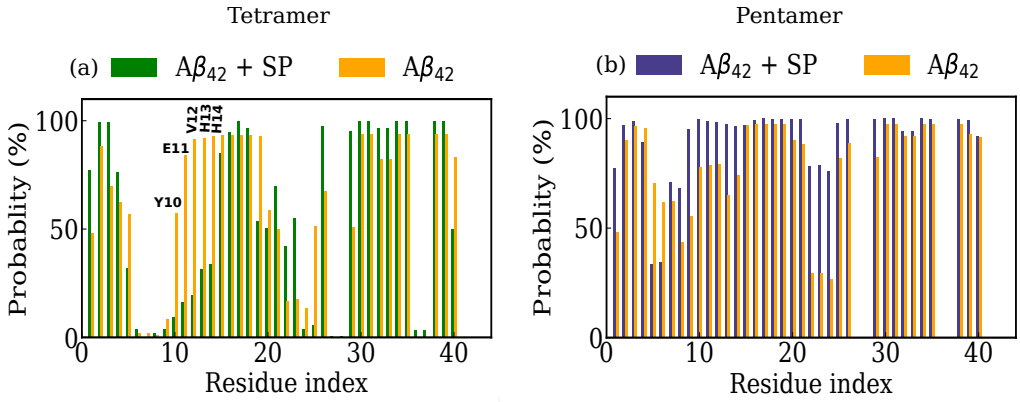



Figure 6.3: Residue-wise $\beta$ sheet probability of A$\beta_{42}$ protofibril for (a) Tetramer and (b) Pentamer. The $\beta$ sheet probability calculations were performed on the last 200 ns of the trajectory.

of $NH_3^+$ group of K28 and $COO^-$ group of A42. Figure 6.4a and b shows PDF of intra-chain salt bridge distance for tetramer and pentamer respectively. The tetramer PDF in control system(Figure 6.4a) shows a sharp and high peak at around 0.45 nm with an average probability value of 0.85. The presence of a narrow and sharp peak in PDF indicate a stable network of salt bridges for tetramer in conrol system. On the other hand, PDF curve for the ligand system is spread across the higher salt bridge distances and also the area under the curve is much wider with a smaller peak(probability value = 0.2) at around 1.4 nm. These observations suggests that the intra-chain salt bridge network of tetramer is critically destabilized in the presence of SP. In case of pentamer(Figure 6.4b), we observed a stable salt bridge network for the isolated A$\beta_{42}$ protofibril, indicated by two sharp peaks, one at 0.36 nm and a higher peak at 0.4 nm. A comparison between PDFs of tetramer and pentamer in the control system reveals that the intra-chain salt bridge network in pentamer is more stable than tetramer, since the dominant peak for pentamer's PDF appears at a smaller distance(0.4 nm) than of tetramer(0.45 nm). The PDF of pentamer in ligand system shows a single distinctive peak at around 0.8 nm with a probability value of 0.6, also the area under ther curve of ligand's PDF is broader than control's. These findings suggests that the SP effectively disrupts the intra-chain salt bridge network in the pentamer, however, the disruption is not as severe as in tetramer. The pentamer's PDF also explains why average RMSD for

region-3 shows an increase while region-1 and region-2 shows a decrease with the SP binding(Figure 6.2d). Since PDF shows that SP binding disrupts salt bridge network in pentamer and both the residues(K28, A42) participating in the formation of the salt bridge lies in region-3, hence, this disturbance get reflected in the RMSD.

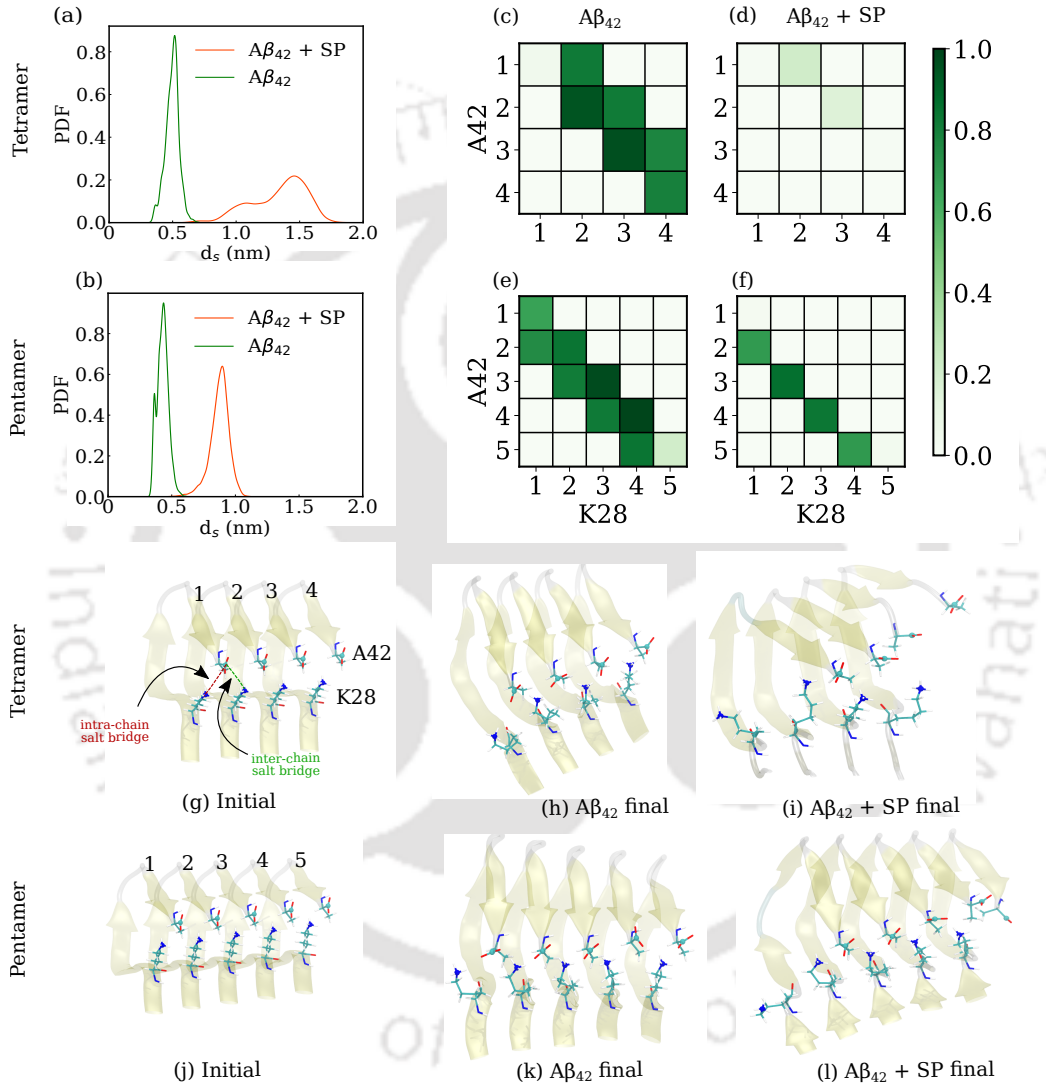



Figure 6.4: Effect of SP binding on inter-chain/intra-chain K28-A42 salt bridge. Probability distribution function of the intra-chain distance between K28 and A42 for (a) Tetramer and (b) Pentamer. Contact probability maps for all the possible inter-chain and intra-chain K28-A42 salt bridges in tetramer: (c) control system (d) ligand system and pentamer: (e) control system (f) ligand system. Snapshots of salt bridges at the start and end of simulation for (g-i) tetramer and (j-l) pentamer.

We also investigated the effect of SP binding on the stability of inter-chain salt bridges between the adjacent chains of A$\beta_{42}$ protofibril. We calculated the probability contact maps of intra-chain and inter-chain salt bridges of tetamer and pentamer in control and ligand systems(Figure 6.4$\{c-f\}$). As shown in the Figure 6.4c, the contact map for tetramer in the control system is well populated indicating a stable network of intra and inter salt bridges. Although, the contact map suggests a weaker intra-chain salt bridge for chain-1 indicated by a very light color(almost blank) in the grid (1,1), it is comepnsated by a strong inter-chain salt bridge between A42 of chain-1 and K28 of chain-2 suggested by the strong green color of the grid (1,2). In the contact map for tetramer in the ligand system(Figure 6.4d) except for grid (1,2) and (2,3) showing a very light color, all grids are blank reflecting a complete disruption of both intra and inter salt bridge network of the tetramer in the presence of SP. In case of pentamer, a comparison between the contact map for control(Figure 6.4e) and ligand(Figure 6.4f) system reveals that the SP binding disrupts the salt bridge network in pentamer, however, only intra-chain bridges are effected and the inter-chain salt bridges remain intact. A visual inspection of the initial and final states of the trajectories of control and ligand systems(Figure 6.4g-l) corroborates well with the PDF and contact map analysis. The tetramer's final state(Figure 6.4 shows a completely disordered salt bridge network, while the pentamer's a slight disruption of the salt bridge network.

**SP Destabilizes the Hydrophobic contacts in Tetramer and Pentamer, Critically in Case of Tetramer.** The three hydrophobic cores(Figure 7.1c) are hallmark of LS-shaped A$\beta_{42}$ protofibril. These cores are stabilized by the contacts formed between the hydrophobic residues present in these cores. The residues involved in hydrophobic interactions in the three cores are core-1: (A2, F4, L34, V36), core-2: (L17, E19, I31), and core-3: (A30, I32, M35, V40). To examine the perturbation caused to these contacts due the presence of SP, we computed pairwise contacting frequency of these residues for each system(control and ligand). Figure 6.5 shows the contact maps for all the systems. As shown in the Figure 6.5a the tetramer in the control system have strong contacts in all three cores indicated by the high contacting frequency observed for the residues in these cores. Alternatively, the contact map for tetramer in the ligand system(Figure 6.5b) shows weaker contact-

ing freqency for the key residue pairs in all three hydrophobic cores. This indicates that SP has caused disruption of key hydrophobic contacts in the tetramer of A$\beta_{42}$ protofibril. In particular, the contacting frequency for the residue pairs F4-L34, F4-V36 in core-1, L17-I31, E19-I31 in core-2, and A30-V40, I32-V40 in core-3 has greatly decreased. Core-3 observes the most drastic disruption of the contacts among all the cores of the tetramer, all the dark colors(implying high contact fequency) in the grids of core-3 box (red box) of the control system(Figure 6.5a) have almost vanished in the ligand system(Figure 6.5b). Moreover, a comparison of contacts in core-3 between control(Figure 6.5a) and ligand(Figure 6.5) system also reveal the critical disruption of intra-chain salt bridge network(K28-A42) in tetramer, which is in agreement with the results obtained in the analysis of salt bridges in the previous section. However, the hydrophobic contacts A2-V36 and M35-V40 of the core-1 and core-3 of tetramer respectively, are unaffected by the presence of SP and remains intact. The contact analysis of tetramer reveals that SP has caused significant disruption of key hydrophobic contacts in all the three cores of the protofibril. Figure 6.5c shows the contact map of pentamer in the control system. The contact map reveals weaker interactions between the hydrophobic residues of core-1 for the pentamer of isolated protofibril. The only residue pair showing substantial contacting frequency in the core-1 is A2-V36. Although, the contact map for pentamer in the control system revealed weaker hydrophobic interactions in core-1, the presence of SP further weakend these interactions, as can be seen in Figure 6.5b. The core-2 of the pentamer shows stable contacts indicated by high contacting frequency of residue pairs inside the core-2 box(green box)(Figure 6.5c). Although, there is a decrease in contacting frequency for many residue pair in core-2 in the presence of SP as can be seen in the Figure 6.5d, the key hydrophobic contacts L17-I31 and E19-I31 remains unaffected. Finally, the bold contacting frequencies of the residues in core-3 of the pentamer(red box in Figure 6.5c) suggests a stable core region in the control system. The presence of SP although influenced some residue pairs in core-3(red box in Figure 6.5d), however, the critical hydrophobic contacts A30-V40, I32-V40 and M35-V40 remain unaltered. Moreover, the contact maps also shows that the intra-chain salt bridge(K28-A42) of the pentamer, which is also a crucial stabilizing interaction of core-3,stay almost undisturbed by the SP binding. This

observation for pentamer's intra-chain salt bridges is in corroboration with the salt bridge analysis in the previous section.

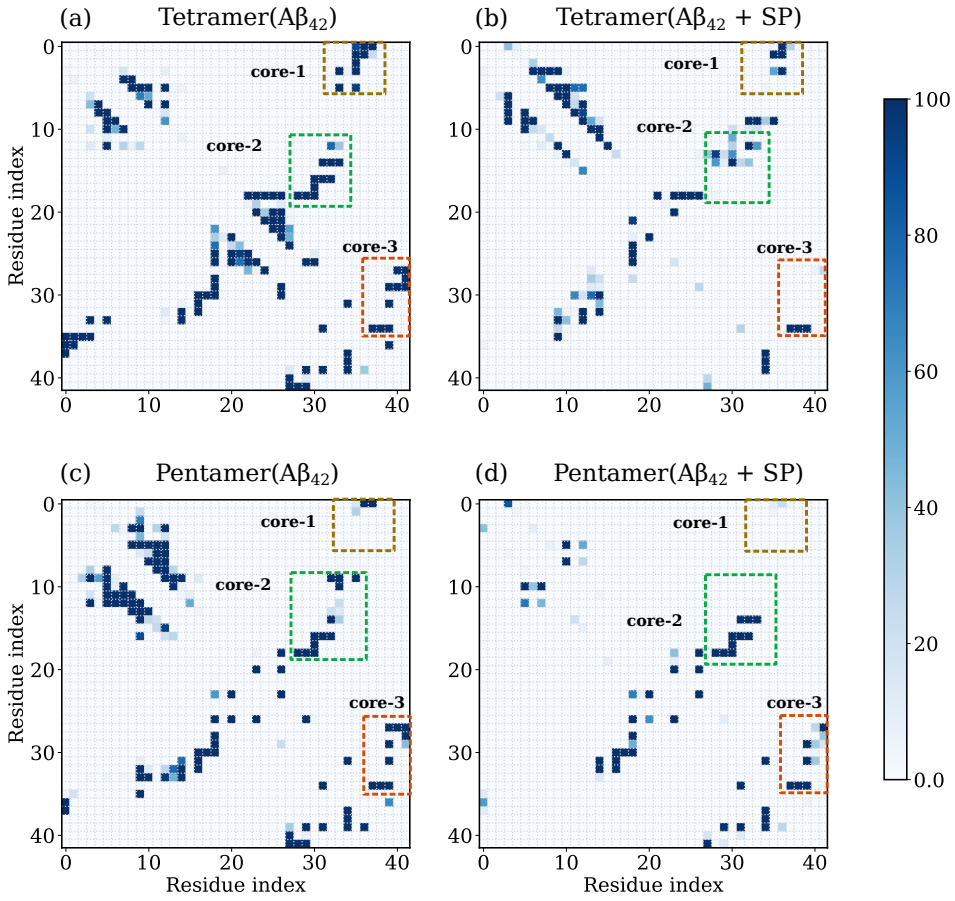



Figure 6.5: Residue-residue contacting frequency maps for tetramer in (a) control system (b) ligand system and pentamer in (c) control system and (d) ligand system. Residue contact pairs for core-1, core-2 and core-3 of A$\beta_{42}$ protofibril are highlighted in brown, green and red dashed boxes respectively. The contacting frequency maps were computed with last 200 ns of the trajectory.

**Binding Site Analysis and Governing Interactions.** So far, we discussed the effect of SP on the structural integrity of A$\beta_{42}$ protofibril, and the results demonstrated that the fibrillar structure is severely disrupted due to SP's presence. How-

ever, it is also noteworthy that the pentamer's fibril structure is relatively stable and unaffected by the SP's presence, while the tetramer suffers a drastic loss of $\beta$ sheet content and disruption of salt bridges, this suggests that the SP is interacting differently with pentamer and tetramer. Hence, to identify the specific binding sites of SP on tetramer and pentamer and decipher the nature of these interactions, we calculated residue-wise average interaction energies(Figure 6.6a and b) and contact numbers(Figure 6.6c and d) between A$\beta_{42}$ protofibril residues and SP molecules.

We can note from Figure 6.6a that the tetramer's residues showing high binding affinity with SP are F20, V18, K16, and I31 in the order of decreasing interaction energy. The presence of phenylalanine, valine, and isoleucine in the above list of the key interacting residues reveals the prominent role of hydrophobic and aromatic residues in the SP binding. Interestingly, the residues K16, V18, and F20 belong to the amyloidogenic core region(KLVFFA), and the subunit LVFFA from this region is the primary ingredient in the design of SP. The high binding affinities of these residues indicate that the SP can recognize the subunit in A$\beta_{42}$ protofibril and bind strongly to it. Figure 6.6c shows the number of contacts each residue of tetramer has with SP. The contact number analysis tells a similar story as the interaction energies for tetramer. The residues with high contact numbers are the same with high interaction energies and in the same order. Additionally, the contact analysis highlights the role of side-chain interaction in the binding of SP. As can be observed from Figure 6.6c, all the key interacting residues have higher side chain contacts, indicating the presence of aromatic and hydrophobic interactions with SP. For pentamer, the key interacting residues involved in SP binding, as revealed by interaction energy(Figure 6.6b and contact(Figure 6.6d) analysis are: F20, Y10, L34, D23, G25, G9 and F4 in the order of decreasing interaction energy and contact number. Again, the residues in pentamer that show high affinity towards SP are primarily aromatic(F20, Y10, and F4) and hydrophobic(L34). Also, the contact analysis for pentamer(Figure 6.6d) shows that the key interacting residues exhibit higher side-chain contacts.

Although the number of key interacting residues in the case of pentamer is higher than tetramer, a comparison of interaction energies(Figure 6.6 a and d) and number of contacts(Figure 6.6 c and d) between tetramer and pentamer shows that tetramer

exhibits higher binding affinity towards SP, particularly F20 of tetramer displays interaction energy and contact number that are at least twice than any residue in pentamer. Moreover, as explained above, the tetramer residues with high binding affinity to SP belong to the amyloidogenic core region, while these residues in pentamer show relatively weaker interaction with SP. Furthermore, the markedly strong interaction of F20 with SP suggests weakening of the $\pi - \pi$ interaction between the benzene rings of F20 residue of adjacent chains of the tetramer. And as these interactions play a central role in the formation of A$\beta$ fibrils[**? ?** ], any perturbation to these interactions will cause modification of the fibril's structure. Thus, all the above observations explain the pronounced destabilization of fibrillar structure in the case of tetramer while not so much for pentamer. Figure 6.6e shows the relative position and orientation of the key interacting residues in the LS-shaped A$\beta_{42}$ protofibril. We can observe from Figure 6.6e that residues with high binding affinity to SP have the side chains oriented outside the surface of the protofibril. The preferential binding of SP to the residues with side chains oriented outwards can be attributed to the complex hairpin-like chemical structure of the SP(Figure 6.6h). Due to the structural restrictions, it would be difficult for SP to interact with the residues with side chains facing inward; therefore, it preferentially binds to the outward-facing residues. However, Figure 6.6e shows few inward-facing residues(F4, I31, and L34) also, showing favorable interactions with SP. These residues must belong to the corner chains of the protofibril(tetramer/pentamer), where they are easily accessible to SP for interactions. Still, these residues exhibit relatively weaker interactions with SP than the outward-facing residues, as observed from the interaction energy and contact number plots.

Residue-wise interaction energy and contact analysis provided information about the residues of A$\beta_{42}$ protofibril participating in the binding with SP. However, to gain a deeper understanding of SP-A$\beta$ binding, we computed contacting frequency maps(Figure 6.6f and g) residue pairs between A$\beta_{42}$ and SP. Figure 6.6h shows different residues of SP in different colors and their corresponding name and number. Figure 6.6f shows a contact map between the residues of SP and tetramer. The residue K16 of tetramer shows a high contacting frequency with SUC9(succinyl), alanine(A10), and phenylalanine(F11) residues of SP. A strong binding between

K16(positively charged side-chain) of A$\beta$ and F11(aromatic side-chain) of SP suggest the presence of cation-$\pi$ interaction. L17(leucine) shows a high contact frequency with only F11 of SP, indicating a hydrophobic interaction between them. V18(valine) of tetramer shows pronounced contacts with F11, ADP7(adipoyl), F2, and F3 of SP. All these residues are hydrophobic/aromatic, implying a strong hydrophobic interaction with V18. F19 of tetramer exhibits high contacting frequency with F2, and a medium one with SUC9 and F11 suggests the presence of $\pi - \pi$ and hydrophobic interactions. As evident from interaction energy and contact analysis for tetramer, F20 shows the highest number of high-frequency contact pairs with SP residues. The residues of SP in contact with F20 are F2, F3, V4, ADP7, F11, V13, and F12. As can be observed, many aromatic and hydrophobic residues show close interaction with F20, pointing towards the presence of $\pi - \pi$ and hydrophobic interactions between F20 and SP. The residue I31 exhibits aromatic and hydrophobic contacts with F2, SUC9, and A10. Finally, I32 and G33 show marked contacts with SUC9 and A10. Figure 6.6j shows a snapshot depicting $\pi - \pi$ interaction between F20 of tetramer and F2 of SP.

A superficial overview of the contact map for pentamer and SP(Figure 6.6g) shows that pentamer has more binding sites than tetramer, evident from interaction energy and contact number analysis. F4 of pentamer shows strong interaction with F2, F3, and F11, implying $\pi - \pi$ interaction with each of these residues with F4. Residues D7(aspartate) and S8(serine) show strong interaction with PEG8(polyethylene glycol) of SP and weak interaction with A1(alanine). These contacts are mainly with the main chain of the pentamer's residues, as observed from the contact number analysis(Figure 6.6d), and hence are primarily hydrophobic. Similarly, G9 of pentamer also shows strong main-chain contacts with PEG8 and V13(valine) of SP. Further, Y10(tyrosine) of pentamer exhibits pronounced contacts with F11, F12, V13, and PEG8, indicating strong hydrophobic and $\pi-\pi$ interaction; Y10 also shows a mild interaction with ADP7. For the V18 residue of pentamer, we observe two contacts with A1 and ADP7 with medium contact frequency and a weak one with PEG8. The F20 of pentamer establishes strong interactions with A1, V4, and ADP7 of SP while interactions of medium strength with F3 and PEG8. The F20 has strong hydrophobic but a weaker aromatic interaction with SP. Residues

D23, V24, and G25 have high contact frequency with F11 of SP, and D23 also shows firm contact with V13. Finally, L34 of pentamer shows strong hydrophobic contacts with F2 and considerable contacts with F3.

Overall, the residue-wise interaction energy and contact number analysis revealed crucial residues in tetramer/pentamer participating in binding with SP. Although pentamer showed higher binding residues, tetramer's residues exhibited stronger interaction, particularly in the amyloidogenic core region. The contact maps between the residues of SP and tetramer/pentamer identified the residue pairs in close and sustained contact, consequently revealing the nature of the interaction between SP and A$\beta_{42}$ protofibril. We found a range of interactions at play; however, the $\pi - \pi$ and hydrophobic interaction were the most dominant mode of interaction.

## 6.4 Conclusion

In this study, we investigated the effects and molecular mechanism of a hairpin-like synthetic paratope binding on a preformed LS-shaped A$\beta_{42}$ protofibril using MD simulations. Our simulations show that SP can destabilize the overall structural integrity of A$\beta_{42}$ protofibril. The RMSD analysis revealed that the tetramer of the A$\beta_{42}$ protofibril is severely disrupted, while pentamer remains almost unaffected by the presence of SP. The region Y10-N27 in the tetramer suffered the maximum structural distortion, while D1-G9 and K28-A42 also displayed significant deviation from their native structures in the presence of SP. The SP binding caused a considerable loss of $\beta$ sheet content in tetramer, particularly for the segment Y10-H14, while pentamer showed no serious changes in its $\beta$ sheet due to the SP binding. The K28-A42 salt bridge is a crucial factor in the A$\beta_{42}$ protofibril's structural stability, and our simulations showed that SP causes severe damage to intra-chain and inter-chain salt bridges in the case of the tetramer. The contacts in the three hydrophobic cores of A$\beta_{42}$ protofibril are strongly affected by SP in case of tetramer, especially the key hydrophobic contacts in these cores such as core-1:[(F4-L34), (F4-V36)], core-2:[(L17-I31), (E19-I31)], and core-3:[(A30-V40), (I32-V40)] are heavily disrupted due to SP binding.

The binding site analysis revealed interesting details about the differences in the

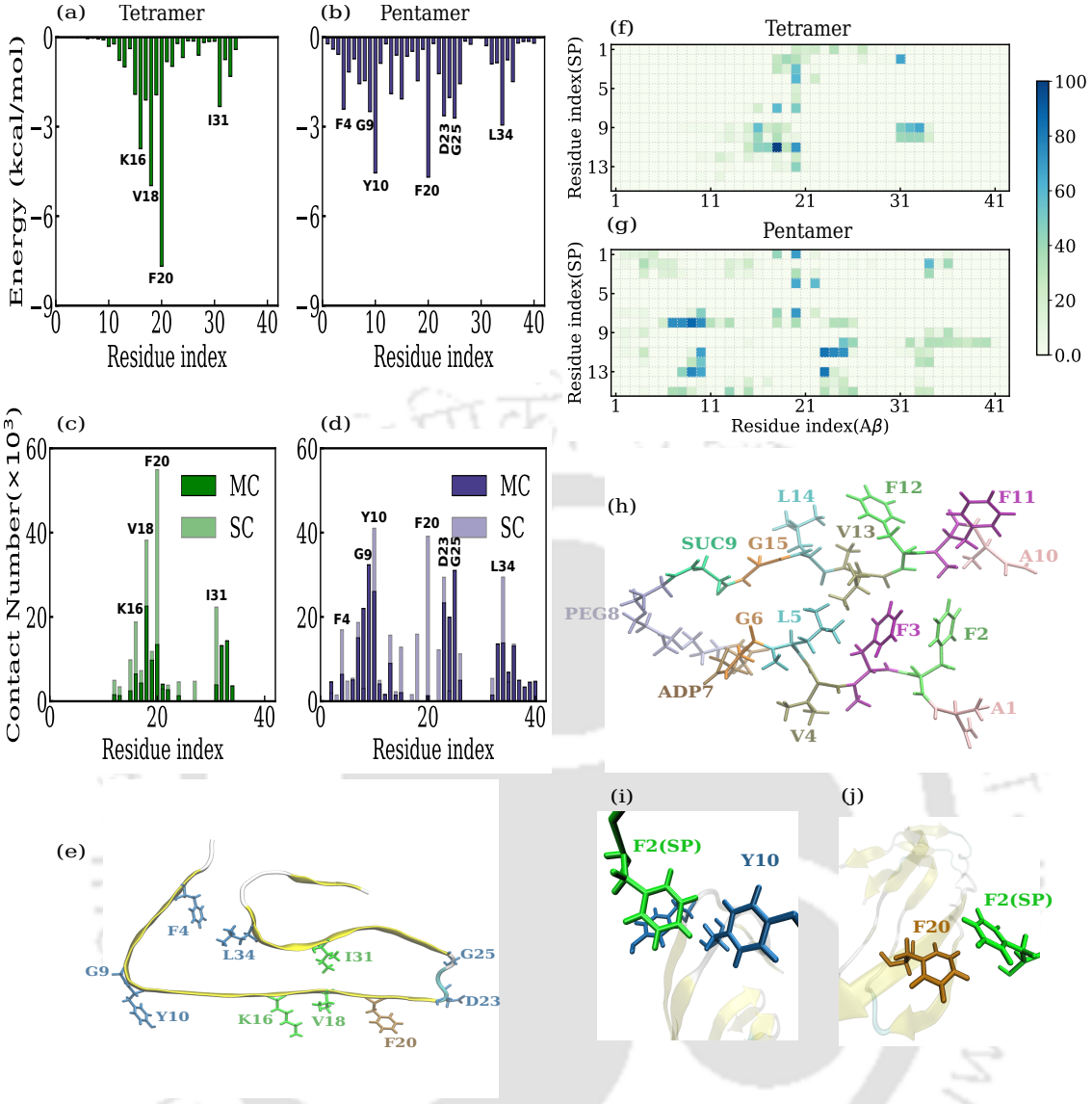



Figure 6.6: Interaction energy between each residue of (a) tetramer and (b) pentamer and SP. Number of contacts between main-chain(MC) and side-chain(SC) of each residue of (c) tetramer and (d) pentamer and SP. Visual representation of the key interacting residues of tetramer(green), pentamer(blue), and residues common to tetramer and pentamer(brown). Contacting frequency maps between residue pair of SP and (f) tetramer and (g) pentamer. (h) Licroice representation of the structure of SP. The residues are shown in different colors and are annotated with residue name and number. Snapshot of $\pi - \pi$ interaction between one of the phenylalanine(F2) of SP with (i) Y10 of pentamer and (j) F20 of tetramer.

binding of SP with tetramer and pentamer and the binding mechanism of SP with A$\beta_{42}$ protofibril. Interestingly, for tetramer, the SP was able to self-recognize the LVFFA region on the protofibril, which is the key component of the SP's design, and the SP binds strongly to this region with F20, V18, and K16 as the participating residues. The key binding residues for pentamer are F20, Y10, D23, L34, and F4. Although pentamer shows more binding sites than tetramer, the pentamer's residues have much weaker binding with SP, which explains why we do not observe any significant disaggregation of fibrils in the case of pentamer. The SP preferentially binds to the residues with side chains oriented outward from the LS-shaped A$\beta_{42}$ protofibril surface. The dominant modes of interaction identified between A$\beta_{42}$ protofibril and SP are $\pi - \pi$ and hydrophobic interaction. Our study provides interesting atomistic details about SP binding onto A$\beta_{42}$ protofibril. It also sheds light on the disaggregation mechanism of the fibrils due to SP binding. The results of our study can prove to be useful in developing therapeutics for AD.

# Bibliography

[1] John Hardy and Dennis J Selkoe. The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics. *science*, 297(5580):353–356, 2002.

[2] Dennis J Selkoe. Alzheimer's disease: genes, proteins, and therapy. *Physiological reviews*, 2001.

[3] Henry W Querfurth and Frank M LaFerla. Alzheimer's disease. *New England Journal of Medicine*, 362(4):329–344, 2010.

[4] Lars M Ittner and Jürgen Götz. Amyloid-$\beta$ and tau—a toxic pas de deux in alzheimer's disease. *Nature Reviews Neuroscience*, 12(2):67–72, 2011.

[5] Dennis J Selkoe and John Hardy. The amyloid hypothesis of alzheimer's disease at 25 years. *EMBO molecular medicine*, 8(6):595–608, 2016.

[6] Brenda D Moore, Paramita Chakrabarty, Yona Levites, Tom L Kukar, Ann-Marie Baine, Tina Moroni, Thomas B Ladd, Pritam Das, Dennis W Dickson, and Todd E Golde. Overlapping profiles of a$\beta$ peptides in the alzheimer's disease and pathological aging brains. *Alzheimer's research & therapy*, 4(3): 1–15, 2012.

[7] Alan Rembach, Noel G Faux, Andrew D Watt, Kelly K Pertile, Rebecca L Rumble, Brett O Trounson, Christopher J Fowler, Blaine R Roberts, Keyla A Perez, Qiao-Xin Li, et al. Changes in plasma amyloid beta in a longitudinal study of aging and alzheimer's disease. *Alzheimer's & Dementia*, 10(1):53–61, 2014.

[8] Ji Won Um, Adam C Kaufman, Mikhail Kostylev, Jacqueline K Heiss, Massimiliano Stagi, Hideyuki Takahashi, Meghan E Kerrisk, Alexander Vortmeyer, Thomas Wisniewski, Anthony J Koleske, et al. Metabotropic glutamate receptor 5 is a coreceptor for alzheimer a$\beta$ oligomer bound to cellular prion protein. *Neuron*, 79(5):887–902, 2013.

[9] Karie N Dahlgren, Arlene M Manelli, W Blaine Stine, Lorinda K Baker, Grant A Krafft, and Mary Jo LaDu. Oligomeric and fibrillar species of amyloid-$\beta$ peptides differentially affect neuronal viability. *Journal of Biological Chemistry*, 277(35):32046–32053, 2002.

[10] Massimo Stefani and Christopher M Dobson. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of molecular medicine*, 81:678–699, 2003.

[11] Jeremy D Schmit, Kingshuk Ghosh, and Ken Dill. What drives amyloid molecules to assemble into oligomers and fibrils? *Biophysical journal*, 100 (2):450–458, 2011.

[12] Samuel IA Cohen, Sara Linse, Leila M Luheshi, Erik Hellstrand, Duncan A White, Luke Rajah, Daniel E Otzen, Michele Vendruscolo, Christopher M Dobson, and Tuomas PJ Knowles. Proliferation of amyloid-$\beta$42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences*, 110(24):9758–9763, 2013.

[13] Thomas CT Michaels, AnĐela Šarić, Johnny Habchi, Sean Chia, Georg Meisl, Michele Vendruscolo, Christopher M Dobson, and Tuomas PJ Knowles. Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annual review of physical chemistry*, 69:273–298, 2018.

[14] Jun-Xia Lu, Wei Qiang, Wai-Ming Yau, Charles D Schwieters, Stephen C Meredith, and Robert Tycko. Molecular structure of $\beta$-amyloid fibrils in alzheimer's disease brain tissue. *Cell*, 154(6):1257–1268, 2013.

[15] Anant K Paravastu, Richard D Leapman, Wai-Ming Yau, and Robert Tycko. Molecular structural basis for polymorphism in alzheimer's $\beta$-amyloid fibrils. *Proceedings of the National Academy of Sciences*, 105(47):18349–18354, 2008.

[16] Lothar Gremer, Daniel Schölzel, Carla Schenk, Elke Reinartz, Jörg Labahn, Raimond BG Ravelli, Markus Tusche, Carmen Lopez-Iglesias, Wolfgang Hoyer, Henrike Heise, et al. Fibril structure of amyloid-$\beta$ (1–42) by cryo–electron microscopy. *Science*, 358(6359):116–119, 2017.

[17] Marius Kollmer, William Close, Leonie Funk, Jay Rasmussen, Aref Bsoul, Angelika Schierhorn, Matthias Schmidt, Christina J Sigurdson, Mathias Jucker, and Marcus Fändrich. Cryo-em structure and polymorphism of a$\beta$ amyloid fibrils purified from alzheimer's brain tissue. *Nature communications*, 10(1):4760, 2019.

[18] Hadas Skaat, Ravit Chen, Igor Grinberg, and Shlomo Margel. Engineered poly-

mer nanoparticles containing hydrophobic dipeptide for inhibition of amyloid-$\beta$ fibrillation. *Biomacromolecules*, 13(9):2662–2670, 2012.

[19] Min Zhang, Xiaobo Mao, Yue Yu, Chen-Xuan Wang, Yan-Lian Yang, and Chen Wang. Nanomaterials for reducing amyloid cytotoxicity. *Advanced materials*, 25(28):3780–3801, 2013.

[20] Guanbin Gao, Mingxi Zhang, Dejun Gong, Rui Chen, Xuejiao Hu, and Taolei Sun. The size-effect of gold nanoparticles and nanoclusters in the inhibition of amyloid-$\beta$ fibrillation. *Nanoscale*, 9(12):4107–4113, 2017.

[21] Tadamasa Arai, Daisuke Sasaki, Takushi Araya, Takeshi Sato, Youhei Sohma, and Motomu Kanai. A cyclic klvff-derived peptide aggregation inhibitor induces the formation of less-toxic off-pathway amyloid-$\beta$ oligomers. *ChemBioChem*, 15(17):2577–2583, 2014.

[22] Gene Hopping, Jackson Kellock, Ravi Pratap Barnwal, Peter Law, James Bryers, Gabriele Varani, Byron Caughey, and Valerie Daggett. Designed $\alpha$-sheet peptides inhibit amyloid formation by targeting toxic oligomers. *Elife*, 3:e01681, 2014.

[23] Tadamasa Arai, Takushi Araya, Daisuke Sasaki, Atsuhiko Taniguchi, Takeshi Sato, Youhei Sohma, and Motomu Kanai. Rational design and identification of a non-peptidic aggregation inhibitor of amyloid-$\beta$ based on a pharmacophore motif obtained from cyclo [-lys-leu-val-phe-phe-]. *Angewandte Chemie International Edition*, 53(31):8236–8239, 2014.

[24] Jan Bieschke, Martin Herbst, Thomas Wiglenda, Ralf P Friedrich, Annett Boeddrich, Franziska Schiele, Daniela Kleckers, Juan Miguel Lopez del Amo, Björn A Grüning, Qinwen Wang, et al. Small-molecule conversion of toxic oligomers to nontoxic $\beta$-sheet–rich amyloid fibrils. *Nature chemical biology*, 8 (1):93–101, 2012.

[25] Janet C Saunders, Lydia M Young, Rachel A Mahood, Matthew P Jackson, Charlotte H Revill, Richard J Foster, D Alastair Smith, Alison E Ashcroft, David J Brockwell, and Sheena E Radford. An in vivo platform for identifying inhibitors of protein aggregation. *Nature chemical biology*, 12(2):94–101, 2016.

[26] Darryll MA Oliver and P Hemachandra Reddy. Small molecules as therapeutic

drugs for alzheimer's disease. *Molecular and Cellular Neuroscience*, 96:47–62, 2019.

[27] Ali Reza A Ladiwala, Moumita Bhattacharya, Joseph M Perchiacca, Ping Cao, Daniel P Raleigh, Andisheh Abedini, Ann Marie Schmidt, Jobin Varkey, Ralf Langen, and Peter M Tessier. Rational design of potent domain antibody inhibitors of amyloid fibril assembly. *Proceedings of the National Academy of Sciences*, 109(49):19965–19970, 2012.

[28] Fan Liao, Aimin Li, Monica Xiong, Nga Bien-Ly, Hong Jiang, Yin Zhang, Mary Beth Finn, Rosa Hoyle, Jennifer Keyser, Katheryn B Lefton, et al. Targeting of nonlipidated, aggregated apoe with antibodies inhibits amyloid accumulation. *The Journal of clinical investigation*, 128(5):2144–2155, 2018.

[29] Justin A Lemkul and David R Bevan. Destabilizing alzheimer's a$\beta$42 protofibrils with morin: mechanistic insights from molecular dynamics simulations. *Biochemistry*, 49(18):3935–3946, 2010.

[30] Man Hoang Viet, Son Tung Ngo, Nguyen Sy Lam, and Mai Suan Li. Inhibition of aggregation of amyloid peptides by beta-sheet breaker peptides and their binding affinity. *The Journal of Physical Chemistry B*, 115(22):7433–7446, 2011.

[31] Nikhil Agrawal and Adam A Skelton. 12-crown-4 ether disrupts the patient brain-derived amyloid-$\beta$-fibril trimer: Insight from all-atom molecular dynamics simulations. *ACS chemical neuroscience*, 7(10):1433–1441, 2016.

[32] Nan Zhang, Xiaoling Hu, Ping Guan, Kaiyang Zeng, and Yuan Cheng. Adsorption mechanism of amyloid fibrils to graphene nanosheets and their structural destruction. *The Journal of Physical Chemistry C*, 123(1):897–906, 2018.

[33] Chendi Zhan, Yujie Chen, Yiming Tang, and Guanghong Wei. Green tea extracts egcg and egc display distinct mechanisms in disrupting a$\beta$42 protofibril. *ACS chemical neuroscience*, 11(12):1841–1851, 2020.

[34] Ashim Paul, Sourav Kumar, Sujan Kalita, Sourav Kalita, Dibakar Sarkar, Anirban Bhunia, Anupam Bandyopadhyay, Amal Chandra Mondal, and Bhubaneswar Mandal. An explicitly designed paratope of amyloid-$\beta$ prevents neuronal apoptosis in vitro and hippocampal damage in rat brain. *Chemical Science*, 12(8):2853–2862, 2021.

[35] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva

Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1): 1–17, 2012.

[36] Kenno Vanommeslaeghe and Alexander D MacKerell Jr. Automation of the charmm general force field (cgenff) i: bond perception and atom typing. *Journal of chemical information and modeling*, 52(12):3144–3154, 2012.

[37] Christopher G Mayne, Jan Saam, Klaus Schulten, Emad Tajkhorshid, and James C Gumbart. Rapid parameterization of small molecules using the force field toolkit. *Journal of computational chemistry*, 34(32):2757–2770, 2013.

[38] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[39] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[40] Jing Huang and Alexander D MacKerell Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of computational chemistry*, 34(25):2135–2145, 2013.

[41] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[42] Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry*, 13(8):952–962, 1992.

[43] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[44] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n· log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[45] P Turq, F Lantelme, HL Friedman, and F Lantelme. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys*, 66:3045, 1977.

[46] Glenn J Martyna, Douglas J Tobias, and Michael L Klein. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*, 101(5):4177–4189, 1994.

[47] Scott E Feller, Yuhong Zhang, Richard W Pastor, and Bernard R Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of chemical physics*, 103(11):4613–4621, 1995.

[48] Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical journal*, 109(8): 1528–1532, 2015.

[49] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4): 566–579, 1995.

[50] Aneta T Petkova, Yoshitaka Ishii, John J Balbach, Oleg N Antzutkin, Richard D Leapman, Frank Delaglio, and Robert Tycko. A structural model for alzheimer's $\beta$-amyloid fibrils based on experimental constraints from solid state nmr. *Proceedings of the National Academy of Sciences*, 99(26):16742–16747, 2002.

[51] John J Balbach, Aneta T Petkova, Nathan A Oyler, Oleg N Antzutkin, David J Gordon, Stephen C Meredith, and Robert Tycko. Supramolecular structure in full-length alzheimer's $\beta$-amyloid fibrils: evidence for a parallel $\beta$-sheet organization from solid-state nuclear magnetic resonance. *Biophysical journal*, 83(2): 1205–1216, 2002.

[52] Shuang Hou, Ruo-Xu Gu, and Dong-Qing Wei. Inhibition of $\beta$-amyloid channels with a drug candidate wgx-50 revealed by molecular dynamics simulations. *Journal of chemical information and modeling*, 57(11):2811–2821, 2017.

# Chapter 7

# A$\beta_{42}$ - Lipid Bilayer interactions in the presence of $\beta$-Aspartyl

## 7.1 Introduction

Alzheimer's disease (AD) is one of the most common form of dementia. Its one of the key characterstics is deposition of fibril like structures, which mainly contains amyloid beta (A$\beta$) peptides. These fibrillar structures are proposed to form a gated channel on the lipid membrane and further cause disruption of the membrane[1, 2].

The source of cytotoxicity to the lipid membranes remains debated till date. Some studies[3, 4] proposed that the intermediate oligomers are the toxic species. In other studies[5, 6] the well stabilized fibrillar structures were found to be the source of toxicity. Milanesi et al[7] used cryoelectron tomography to demonstrate the damaging effect of fibrils on membrane by 3D visualization of the process. They observed a strong interaction between liposomes and distortion of the membrane. Chang et al [8] employed single-molecule microscopy to track the formation and diffusion of A$\beta$ oligomer on to a lipid bilayer at physiological concentration of the oligomers. They found the monomers to tightly bound to the membrane, but was highly mobile while higher dimensional oligomers largely remains immobilized. Chang et al.[9] performed MD simumaltions with three types of A$\beta$25-35 barrels on a lipid bilayer to understand its distortion and perturbation effect on the bilayer. They found significant damage to the membrane by these barrels, in some cases membrane got

damage with water leakage.

In the present study we performed explicit MD simulations to study the interactions of well organized A$\beta_{42}$ protofibril with a POPC lipid bilayer. We also examine the performance of a drug molecule $\beta$-Aspartyl in screening the interactions between the protofibril and the lipid bilayer and therefore avoiding the membrane damage by fibrils.

## 7.2 Methods

Extensive MD simulations were performed to study the interactions between A$\beta$ fibril, ligand and lipid membrane. In total two systems were prepared: (1) system with lipid bilayer, A$\beta_{42}$ protofibril and ligand, (2) system with lipid bilayer, A$\beta_{42}$ protofibril (without ligand). Starting structure of A$\beta_{42}$ protofibril was obtained from protein data bank with pdb id:5OQV[10]. Missing hydrogens were added to the protofibril via psfgen plugin in VMD[11]. The initial atomic structure of ligand $\beta$-Aspartyl (Figure 7.1a) was constructed and optimized using AVOGADRO[12] chemical editor package. The forcefield parameters and partial charges were generated by the CGENFF server[13]. The parameters obtained with CGENFF were further optimized with FFTK plugin in VMD[11]. Phosphatidylcholine (POPC) lipid bilayer membrane was created with MEMBRANE plugin in VMD. Finally lipid bilayer, A$\beta_{42}$ protofibril and ligands were merged into a single system(Figure 7.1c). The protofibril is placed over the lipid membrane at a closest distance of 12 $\mathring{A}$. The ligands were randomly placed around the protofibril at a distance of 10 $\mathring{A}$ in the ligand system. Next, we solvated the whole system with water molecules and ionized the system with sufficient counterions to neutralize the whole system. Figure 7.1 b and c shows the merged lipid-A$\beta_{42}$ system without ligand and with ligand respectively. The prepared system were then submitted for a minimization run of 5000 steps with conjugate gradient method. Furthermore, the minimized systems were submitted for the equilibration run in an NPT ensemble for 40 ns. Finally, the equilibrated systems were submitted for production runs in NVT ensemble with a simulation time of 230 ns each.

All simulations were performed with NAMD 2.11 package[14]. CHARMM36[15]

and TIP3P[16] force-fields were used for modeling the interactions for A$\beta_{42}$ protofibril and water molecules respectively. A time step of 1 fs was used for the integration of equation of motions. Vdw interactions were cutoff at 12 $\mathring{A}$ with switching distance of 10 $\mathring{A}$. The long range electrostatic interactions were evaluated using particle mesh ewald[17] (PME) summation, beyond 12 $\mathring{A}$ distance. The bonds involving hydrogen atoms were constrained with SHAKE[18] algorithm.

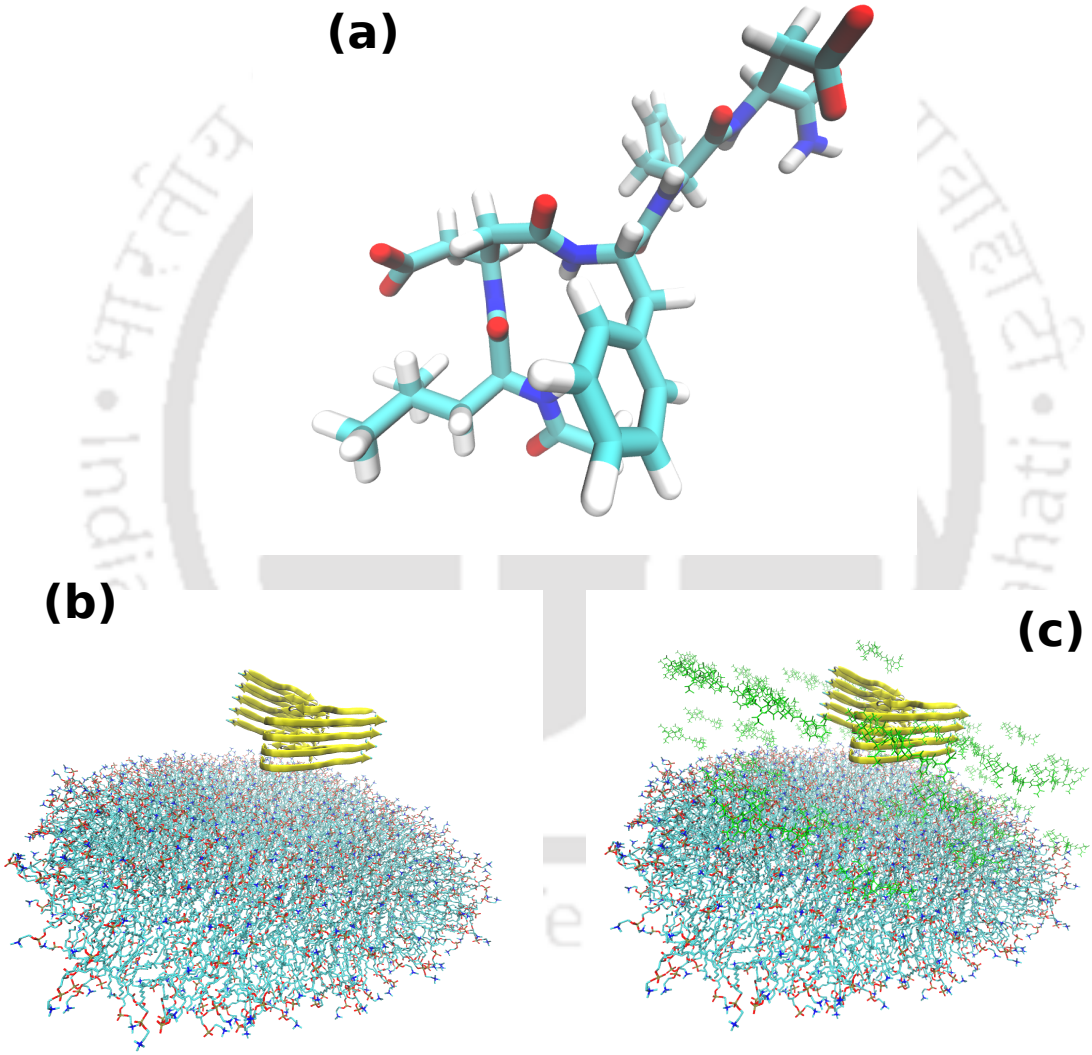
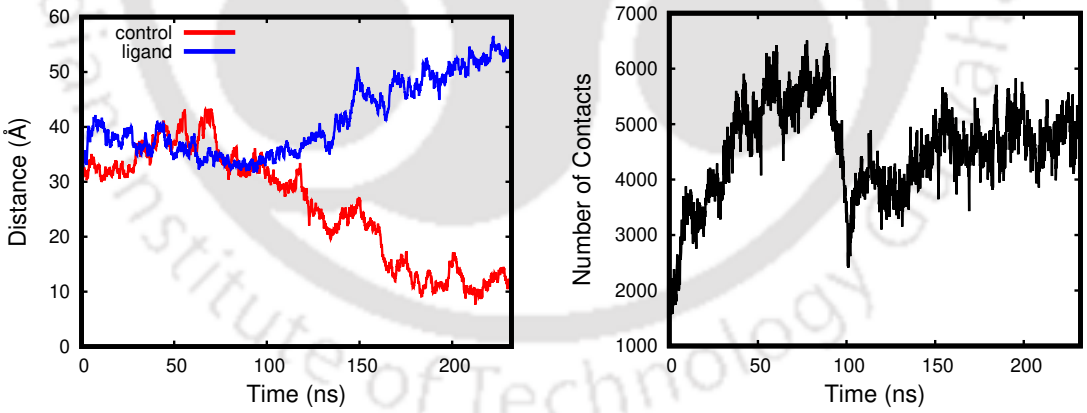



Figure 7.1: Atomic models of (a) ligand $\beta$-Aspartyl and the colors red, blue, cyan, white represents oxygen, nitrogen, carbon and hydrogen. A representation of the merged system (b) control system (without ligand) (c) system with ligand.

## 7.3 Results and discussion

### 7.3.1 Influence on the interaction between A$\beta_{42}$ and lipid bilayer

To characterize the effect of the ligand's presence on the interaction between A$\beta_{42}$ fibrils and lipid bilayer, time evolution of center of mass (COM) distance between the P atoms on the top surface of the lipid bilayer (surface facing the A$\beta_{42}$ protofibril) and A$\beta_{42}$ protofibril is calculated. As shown in the Figure 7.2a the COM distance for control system after around 80 ns start decreasing continuously, indicating that protofibril after 80 ns moved continuosly towards the lipid membrane, this suggests a strong interaction between A$\beta_{42}$ fibril and lipid bilayer. However, COM distance for ligand system shows exactly opposite behaviour than the control. Initially, the COM distance for ligand system decreases, suggesting the protofibril's movement towards the membrane, but after around 90 ns the protofibril gradually and constantly moves away from the membrane. Evolution of COM distance suggests strong influence of ligand on the interaction between A$\beta_{42}$ and lipid bilayer.




(a) (b)

Figure 7.2: Time evolution of (a) center of mass distance between lipid bilayer and A$\beta_{42}$ protofibril and (b) number of contacts between ligand and A$\beta_{42}$ protofibril.

To further quantify the effect of ligand on A$\beta_{42}$-bilayer interaction and to correlate the effect of ligand with COM distance evolution, time evolution of number of contacts between ligand and the protofibril were calculated. A pair of atoms is said

to be in contact if the distance between them is not more than 5$\mathring{A}$. It can be noticed from the Figure 7.2b that number of contacts intially rises steeply with simulation time, representing the attempts of ligands to bind on protofibril. However, after 100 ns, we observe flattening of the curve, which means number of contacts remains almost same throught out the simulation after 100 ns, this indicate strong binding of ligands on protofibril. Furthermore, it can be observed that the binding of the binding of ligands to A$\beta_{42}$ protofibril (evidenced by flattening of number of contacts curve ) is accompanied by moving of protofibril away from the bilayer (Figure 7.2a), as they are happening in the same range of time (100 - 230 ns). This verifies the role of ligands in screening the interactions between A$\beta_{42}$ and lipid bilayer.

### 7.3.2 A$\beta_{42}$-Ligand interaction and its Impact on Fibril's Structure

To charaterize the nature of interaction we calculated interaction energy between ligand and A$\beta_{42}$ protofibril. Figure 7.3a and 7.3b shows time evolution of vdw and coulombic energy between ligand and A$\beta$ fibril. It can be noted that as simulation progresses the interaction energy (both vdw and coulombic) quickly decreases, indicating an attractive interaction.Furthermore, the vdw energy (Figure 7.3a) get equilibrated after around 100 ns, similarly to the time evolution of number of contacts, hence, again indicating binding of ligand to the protofibril. Coulombic energy also decreases initially, but again jump to 0 and fluctuate around it. However, at 100 ns the coulombic energy starts decreasing and takes a dip of -500 KJ/mol at 150 ns, this must be possibly due to binding of the ligand, since it happened around the same time scale which was suggested by other measures to be a binding event. The interaction energy analysis suggests that vdw interaction are major mode of interaction between ligand and A$\beta_{42}$ fibril, with a significant contribution from coulombic interactions.

To get a clearer picture of the non-covalent interactions between ligands and A$\beta_{42}$ and to know about the residue wise participation in the ligand binding we computed contact probablity for A$\beta_{42}$ residues. As described previously a contact is considered if a ligand's atom is not more than 5 $\mathring{A}$ away from a proteins atom. Only

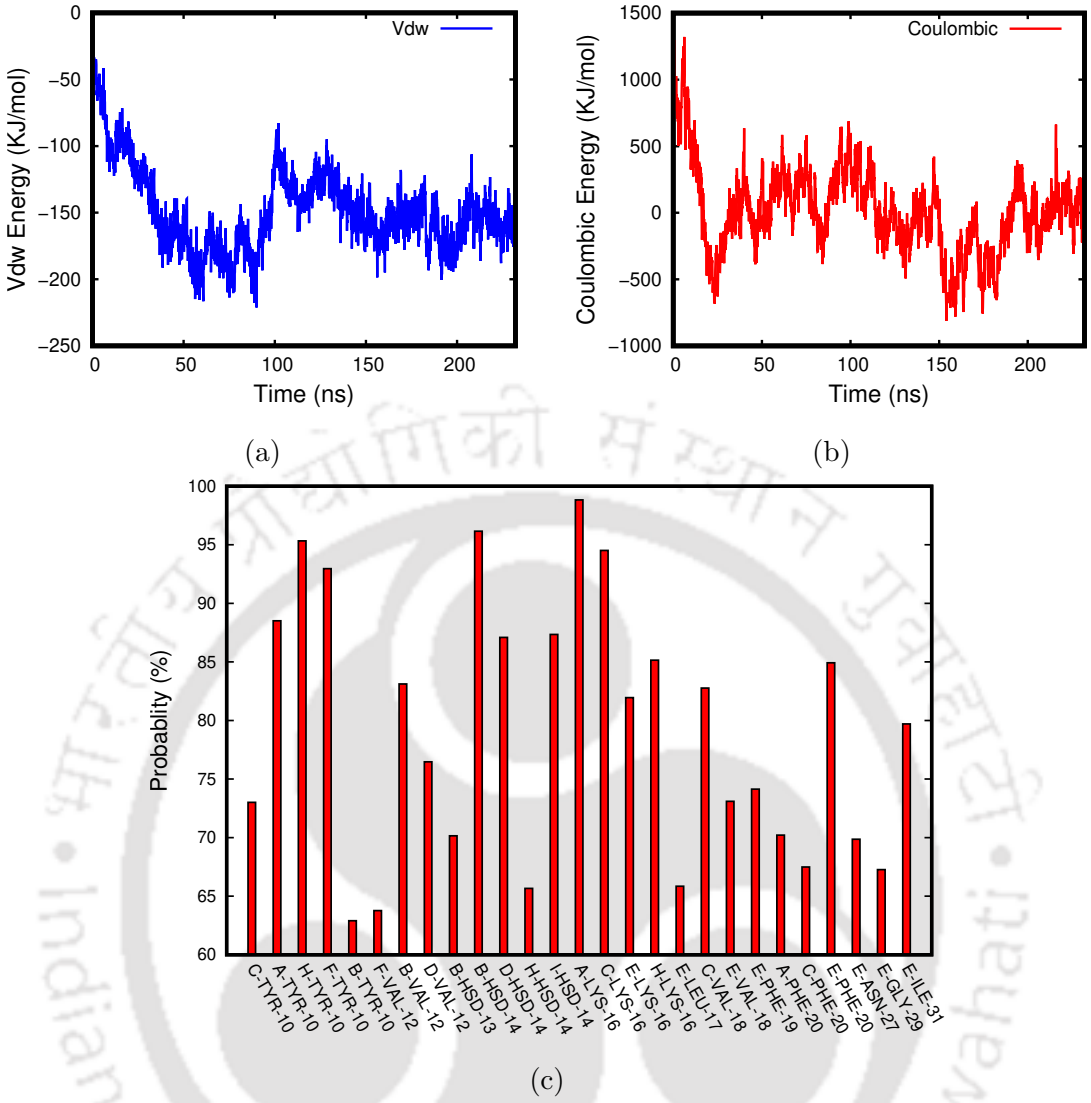



Figure 7.3: Time evolution of (a) Vdw energy (b) Coulombic energy between ligands and A$\beta_{42}$ protofibril. (c) Contact probablity of residues of all chains of A$\beta_{42}$ protofibril with ligands.

residues having more than 60% probablity were considered. Figure 7.3c shows the residues participating in the interaction with ligand and the corresponding contact probablity. The x axis in the Figure 7.3c shows the name of the chain A$\beta_{42}$ fibril to which the particular residue belongs to, residue's name and residue index. It can be noted that there are mainly three types of residues which are in close contact with ligands: hydrophobic (valine, leucine, glycine, isoleucine), aromatic (tyrosine, phenylalanine) and positively charged (lysine, histidine). Hydrophobic residues shows the dominance of hydrophobic interactions and also verify the presence of strong vdw

interaction energy. The aromatic residues in the contact list suggests a possibility of $\pi - \pi$ interactions. Moreover, the positively charged residues found to be in close contact with the ligands, indicate towards cation-$\pi$ interactions with the aromatic ring present on ligand, as these positively charged residues are found to be involved in these interactions.

## 7.4 Conclusion

MD simulations were performed to observe the effect of ligand ($\beta$-aspartyl) on A$\beta_{42}$ protofibril and lipid bilayer interactions. It was found that the ligand is capable of screening the interactions between lipid bilayer and fibrils. The ligands were found to bind strongly to A$\beta_{42}$ protofibril indicated by different measures (interaction energy, number of contacts), and the binding of ligand caused shielding of the inteactions between bilayer and fibrils, indicated by COM distance evolution. The interaction energy analysis showed a strong attractive interaction between ligand and A$\beta_{42}$ protofibril and vdw inteactions were identified as dominant mode of interaction, with a significant contribution from electrostatic interaction. Contact analysis revealed the key residues of A$\beta_{42}$ fibril involved in the interaction with ligand molecules. It also indicated the presence of non covalent interactions between ligand and fibril such as $\pi - \pi$ stacking and cation-$\pi$ interactions based on the class of residues present in the contact list (hydrophobic, aromatic, positively charged). Overall, the present study revealed the nature of interactions between ligand molecules and A$\beta_{42}$ fibril and the key residues of A$\beta_{42}$ involved in the ligand binding. The work also showed that the $\beta$-aspartyl binding to A$\beta_{42}$ can disrupt and shield its interaction with lipid bilayer.

# Bibliography

[1] Nelson Arispe, Harvey B Pollard, and Eduardo Rojas. The ability of amyloid $\beta$-protein [a$\beta$p (1–40)] to form ca2+ channels provides a mechanism for neuronal death in alzheimer's disease. *Annals of the New York Academy of Sciences*, 747 (1):256–266, 1994.

[2] Nelson Arispe, Eduardo Rojas, and Harvey B Pollard. Alzheimer disease amyloid beta protein forms calcium channels in bilayer membranes: blockade by tromethamine and aluminum. *Proceedings of the National Academy of Sciences*, 90(2):567–571, 1993.

[3] Hilal A Lashuel and Peter T Lansbury. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Quarterly reviews of biophysics*, 39(2):167–201, 2006.

[4] Charles G Glabe. Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiology of aging*, 27(4):570–575, 2006.

[5] Laura Pieri, Karine Madiona, Luc Bousset, and Ronald Melki. Fibrillar $\alpha$-synuclein and huntingtin exon 1 assemblies are toxic to the cells. *Biophysical journal*, 102(12):2894–2905, 2012.

[6] Jennifer L Tomic, Anna Pensalfini, Elizabeth Head, and Charles G Glabe. Soluble fibrillar oligomer levels are elevated in alzheimer's disease brain and correlate with cognitive dysfunction. *Neurobiology of disease*, 35(3):352–358, 2009.

[7] Lilia Milanesi, Tania Sheynis, Wei-Feng Xue, Elena V Orlova, Andrew L Hellewell, Raz Jelinek, Eric W Hewitt, Sheena E Radford, and Helen R Saibil. Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proceedings of the National Academy of Sciences*, 109(50):20455–20460, 2012.

[8] Chun-Chieh Chang, Elin Edwald, Sarah Veatch, Duncan G Steel, and Ari Gafni. Interactions of amyloid-$\beta$ peptides on lipid bilayer studied by single molecule imaging and tracking. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1860(9):1616–1624, 2018.

[9] Zhongwen Chang, Yin Luo, Yun Zhang, and Guanghong Wei. Interactions of

a$\beta$25- 35 $\beta$-barrel-like oligomers with anionic lipid bilayer and resulting membrane leakage: an all-atom molecular dynamics study. *The journal of physical chemistry B*, 115(5):1165–1174, 2011.

[10] Lothar Gremer, Daniel Schölzel, Carla Schenk, Elke Reinartz, Jörg Labahn, Raimond BG Ravelli, Markus Tusche, Carmen Lopez-Iglesias, Wolfgang Hoyer, Henrike Heise, et al. Fibril structure of amyloid-$\beta$ (1–42) by cryo–electron microscopy. *Science*, 358(6359):116–119, 2017.

[11] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[12] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1): 1–17, 2012.

[13] Kenno Vanommeslaeghe and Alexander D MacKerell Jr. Automation of the charmm general force field (cgenff) i: bond perception and atom typing. *Journal of chemical information and modeling*, 52(12):3144–3154, 2012.

[14] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[15] Jing Huang and Alexander D MacKerell Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of computational chemistry*, 34(25):2135–2145, 2013.

[16] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[17] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n· log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[18] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints:

molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

# Chapter 8

# Conclusions

In this thesis, protein's behaviour in different environment is studied via molecular dynamics (MD) simulations. Specifically, in the first two objectives the protein play the role of a foulant present near to a desalination membrane, while in the last two objectives the proteins are toxic species present within a living cell and which can cause disorder and disease. The third objective presents the protein as an invader of a foreign cell.

The objectives of the presented dissertation can be summarized as follows:

- Dominance of hydration repulsion force in case of FO make the foulant layer thick and loose, while in case of RO the hydration repulsion between protein and surface is overpowered by the hydraulic pressure being applied.

- We observed that the presence of high concentration of ions strongly effect the protein-membrane interactions in case of both GO and PA. Specifically, GO shows a repulsive interaction while the PA shows attractive interaction with the BSA in the absence of ions and the time evolution analysis of distance and interaction energy between membrane and protein shows that the presence of ions strongly screens these interactions. The structural analysis showed that the BSA gained stability in the presence of ions for GO system, showing a strong screening effect on the BSA structure. However, for PA systems, the BSA structure remains more or less uneffected by the presence of ions. We also observed an adsorption event between BSA and PA surface in PN system. The contact and interaction energy analysis shows that the BSA adsorption

on PA is mostly electrostatically driven and vdW interactions were found to be weak.

- The SARS-CoV-2's spike protein shows strong interaction with MMT surface. The nature of S protein-MMT interaction was found to be predominantly electrostatic. Secondary structure of the protein was also found to be severly affected due to the presence of MMT surface. Specifically, chain B of all three chains of S trimer protein shows the strongest interaction with the MMT surface. Surprisingly, chain B is known to interact with hACE of the host cell in the cell entry mechanism of SARS-CoV-2 virus.

- Our simulations show that SP can destabilize the overall structural integrity of A$\beta_{42}$ protofibril. The structural analysis of the protofibrill revealed that the tetramer of the A$\beta_{42}$ protofibril is severely disrupted, while pentamer remains almost unaffected by the presence of SP. The binding site analysis revealed interesting details about the differences in the binding of SP with tetramer and pentamer and the binding mechanism of SP with A$\beta_{42}$ protofibril. Interestingly, for tetramer, the SP was able to self-recognize the LVFFA region on the protofibril, which is the key component of the SP's design, and the SP binds strongly to this region with F20, V18, and K16 as the participating residues. The key binding residues for pentamer are F20, Y10, D23, L34, and F4. Although pentamer shows more binding sites than tetramer, the pentamer's residues have much weaker binding with SP, which explains why we do not observe any significant disaggregation of fibrils in the case of pentamer. The SP preferentially binds to the residues with side chains oriented outward from the LS-shaped A$\beta_{42}$ protofibril surface. The dominant modes of interaction identified between A$\beta_{42}$ protofibril and SP are $\pi - \pi$ and hydrophobic interaction.

- The binding of ligand $\beta$-Aspartyl to A$\beta_{42}$ protofibril, was found to have a screening effect on the interactions between A$\beta_{42}$ and lipid bilayer. Although, Van der Waals interaction were found to be predominant between ligand and A$\beta_{42}$ protofibril, coulombic interactions were also found to be significant. The key interacting residues in A$\beta_{42}$ with the interaction with ligand were found to

be of following classes: aromatic, hydrophobic and positively charged. The interaction energy along with the key interacting residues empasize the presence of $\pi - \pi$ stacking, cation-$\pi$ and hydrophobic interactions.

## 8.1 Some Ideas for Future Work

Some ideas that sprouted during the reserach of current objectives and could be interesting are as follows:

- Incorporation of silica scalant along with protein as a foulant, can give a more realistic picture of fouling in FO and RO.
- Taking Ph into account via constant-Ph MD simulations can give a detailed idea of the effect of ionic environment on fouling.
- Mediation of clay nanoparticles on the interaction between SARS-CoV-2's spike protein and host cell's hACE2.
- Inhibition of A$\beta$ fibril formation with ligand can be tested with A$\beta$ monomer.
- Interaction of lipid bilayer with preformed oligomers in the presence of $\beta$-Apsartyl.

# Research Output

## Articles Published

- Shivam Tiwari, Abhijit Gogoi, K. Anki Reddy. What governs the nature of fouling in forward osmosis (FO) and reverse osmosis (RO)? A molecular dynamics study. Physical Chemistry Chemical Physics vol. 21 (2019).
- Shivam Tiwari, Abhijit Gogoi, K. Anki Reddy. Effect of an ionic environment on membrane fouling: a molecular dynamics study. Physical Chemistry Chemical Physics vol. 23 (2021).
- Shivam Tiwari, Vasista Adupa, Dhanesh Sing Das, K Anki Reddy, Tadikonda Venkata Bharat. Structural and Dynamic Insights into SARS-CoV-2 Spike-Protein–Montmorillonite Interactions. Langmuir vol. 38 (2022).

## Manuscripts under preparation

- Shivam Tiwari, K Anki Reddy, Bhubaneswar Mandal. The Role of Binding Site Specificity in the Disaggregation of A$\beta$42 Fibrils through a Synthetic Paratope. (Submitted to ACS Chemical Neuroscience. Received comments. Revised version is under preparation.)
- Shivam Tiwari, Vasista Adupa, K Anki Reddy, Bhubaneswar Mandal. How a synthetic peptide "Beta Aspartyl" affects the interaction between amyloid beta aggregate and a lipid bilayer.