IN-SILICO INVESTIGATION OF GEOLOGICAL AND BIOLOGICAL MATERIALS BY MOLECULAR DYNAMICS SIMULATIONS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

H M Nasrullah Faisal

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Program:
Materials and Nanotechnology

November 2021

Fargo, North Dakota

North Dakota State University
Graduate School

Title

IN-SILICO INVESTIGATION OF GEOLOGICAL AND BIOLOGICAL MATERIALS BY MOLECULAR DYNAMICS SIMULATIONS

By

H M Nasrullah Faisal

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dinesh R. Katti
Chair

Kalpana S. Katti

Achintya N. Bezbaruah

Bernhardt Saini-Eidukat

Approved:

11/20/2021
Date

Erik K. Hobbie
Department Chair

ABSTRACT

Molecular dynamics (MD) simulation is a computational technique that predicts the time-dependent behavior of a molecular system utilizing molecular mechanics. MD simulations are extensively employed in the scientific arena to investigate a wide range of material systems at the nanoscale (atoms and molecules), including organics, inorganics, polymer, composites, biomacromolecules, etc. This work investigates the properties of a range of geological (Green River oil shale and swelling clays) and biological materials (coronaviral proteins) at the molecular level using MD simulations. Oil shale, a sedimentary rock containing organic crude oil precursor named kerogen trapped in an inorganic mineral matrix, has long been considered an alternative source of petroleum. Molecular dynamics simulation of Green River oil shale Type I kerogen has been performed in the proximity of predominantly present calcite and quartz minerals to identify their binding interactions with trapped kerogen from the mineral matrix for efficient crude oil production. Sodium-montmorillonite (Na-MMT), a member of the smectite group, is one of the swelling clay minerals components that find various geo-environmental and industrial applications due to its high swelling capacity. Steered molecular dynamics (SMD) simulations have been performed to determine the nanomechanical properties of both dry and hydrated Na-MMT clay tactoid. Besides the geological materials, MD and SMD simulations have also been used to computationally inspect the coronaviral protein-ACE2 protein interactions to elucidate the potential reasons why COVID-19 results in significantly more infections and deaths compared to other coronaviruses. The coronaviral attachment to host cell through spike-ACE2 interactions and coronaviral replication mechanism through tri non-structural protein (nsp12-nsp7-nsp8) interactions have been simulated to understand the differences between SARS-CoV and SARS-CoV-2 (COVID-19). The major findings obtained from coronaviral protein interactions may point

towards the underlying reasons behind the severity of COVID-19. Moreover, the potency of different phytochemicals has been examined for breast cancer treatment. Compounds commonly found in Rhodiola, and Oregano plants extracts have been targeted against a series of breast cancer proteins utilizing molecular docking to determine the most potent phytochemical for breast cancer treatment.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Dinesh Katti and co-advisor, Dr. Kalpana Katti, for their advice, instructions, and supervision throughout my graduate study. Their constant patience, allowance, and versatile mindset have made this work possible.

I would like to acknowledge NDSU Center for Computationally Assisted Science and Technology (CCAST) for providing the computational resources required for performing all the studies included here. I would like to acknowledge Mountain Plains Consortium (MPC), National Science Foundation (NSF), Department of Transportation (DOT), ND EPSCoR, ND agricultural products utilization commission, and NDSU Civil, Construction and Environmental Engineering Department for supporting my graduate study through assistantships.

I am highly obliged to my supervisory committee members: Dr. Achintya N. Bezbaruah and Dr. Bernhardt Saini-Eidukat. I am greatly indebted to Dr. Eidukat for teaching me the basics of mineralogy and helping in the projects. I would like to thank Dr. Wenjie Xia for his instructions for the clay coarse-grained modeling study. I also would like to thank Dr. Kalidas Shetty and Dr. Dipayan Sarkar for providing insight into the breast cancer docking study.

I am grateful to my past and present group members: Krishna Kundu, Haneesh Jasuja, Sharad Jaswandkar, Preetham Ravi, Hanmant Gaikwad, and Sibanwita Mohanty, for their assistance in the lab.

Finally, I would like to acknowledge the patience, support, and inspiration I received from all my family members.

DEDICATION

This dissertation is dedicated to all my family members.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MD .....Molecular Dynamics.

SMD.....Steered Molecular Dynamics.

CHARMM .....Chemistry at Harvard Macromolecular Mechanics

3D.....Three Dimensional.

2D.....Two Dimensional.

Na-MMT/Na-MT.....Sodium Montmorillonite.

XRD .....X-Ray Diffraction.

FTIR.....Fourier Transform Infrared.

SEM .....Scanning Electron Microscopy.

VDW .....Van Der Waals.

SARS.....Severe Acute Respiratory Syndrome

CoV .....Coronavirus

ARDS.....Acute Respiratory Distress Syndrome

Nsp .....Non-structural Protein

RdRp .....RNA Dependent RNA Polymerase

RMP .....Remdesivir Monophosphate

1. INTRODUCTION

1.1. Molecular Dynamics Simulation

Molecular Dynamics (MD) simulation is a computational technique to investigate materials at the molecular level. It models the energetics and conformation of a molecular system using molecular mechanics. In molecular mechanics, all the atoms are treated as classical particles, and bonds are considered springs.[1] The forces acting on each atom have been computed using force-field energy equations. In particular, the force field describes the potential energy of the system. The energy equation contains both bonded (covalent) and non-bonded (non-covalent) terms. The bonded terms are bond stretching, angle bending, dihedral, torsional, and impropers. The non-bonded terms are electrostatics and Van der Waals (VDW) energies. After the computation of acting forces, Newton's second law of motion is utilized to compute accelerations and velocities based on which the new atom positions are determined. [2] MD simulations can effectively model the molecular interactions. The most important interactions that need to be investigated in material systems are primarily non-bonded interactions. Non-bonded interactions, the summation of electrostatic and VDW energies, provide information on how different materials or different segments within a material interact with each other. The electrostatic interactions are computed by applying the coulomb's law on the partial charges of interacting atoms. The VDW interactions are weak electrostatic interactions occurring among the neutral atoms due to their electronic polarization. These VDW interactions can be attractive or repulsive based on the interatomic distances. Lennard-Jones (LJ) potential is one of the most commonly used formulas for calculating VDW interactions. The LJ potential formula considers the distance at which interatomic potential is the lowest and the magnitude of this lowest energy. MD simulation is also termed atomistic modeling as it usually models the behavior of atoms.

Molecular dynamics simulations were first developed in the 1950s to investigate the behavior of hard spheres.[3] However, the first molecular dynamics simulation of a realistic system was performed in 1974.[4] Molecular dynamics simulations were initially performed to investigate molecular protein systems. The first MD simulation of protein was performed in 1977.[5] Nowadays, MD simulations have been extensively used for modeling the molecular properties of different types of materials. The continuous improvement in computer facilities has paved the way for performing MD simulations of systems containing millions of atoms.

Many MD simulation packages and force fields have been developed by the researchers. Among them, NAMD (Nanoscale Molecular Dynamics) has been extensively used to model complex molecular systems, including viruses, and for conducting steered molecular dynamics simulations to evaluate mechanical responses at the molecular scale. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana–Champaign.[6] Performing MD simulation requires a molecular model and and various inputs depending on the nature of simulations being performed. The molecular model is generally built using Materials Studio™ or similar software . The required information of the molecular system includes atomic coordinates, partial charge, and mass of atoms, bonding information (bonds, angles, dihedrals, torsion/improper) of the system. The energy calculation of the system requires the declaration of the force field to be used. The chosen force field in this study is CHARMm (Chemistry at HARvard Macromolecular Mechanics) for performing MD simulations.[7] The CHARMm parameters for the bonded atoms are bond stretching, angle bending, dihedral, and improper/torsion constants, as well as the equilibrium values of these bonds, angles, and dihedrals. The CHARMm parameters for the non-bonded atoms are partial charges and LJ parameters to

compute electrostatic and VDW energies. The input files for NAMD are prepared using Visual Molecular Dynamics (VMD) 1.9.3.

Performing MD simulations require significant computational resources, since material systems that are being investigated consist of hundreds of thousands of atoms. The Center for Computationally Assisted Science and Technology (CCAST) at North Dakota State University offers Linux-based high-performance computing clusters in order to meet the necessities of different computational researches throughout the campus. The high-performance computing (HPC) system has more than 180 nodes equivalent to 8,000 CPU cores, including big-memory nodes and general-purpose GPUs. The nodes consist of Intel Xeon 2670v2 "Ivy Bridge" processors.

Steered molecular dynamics (SMD) is a molecular dynamics technique to investigate the mechanical properties of a molecular system under external loading conditions.[6, 8] In this method, constant velocity or constant force is applied on a number of atoms while some other atoms are kept fixed. The force-displacement behavior of a molecular system can be obtained utilizing this method. Steered molecular dynamics is used to examine the unfolding behavior of proteins as well as mechanical properties of different types of materials.[9]

Molecular docking is the static version of molecular dynamics simulation primarily used for molecular structural biology. It is the most popular method of structure-based drug design and is immensely used for the virtual screening of compounds against a specific protein target.[10] In virtual screening, molecular docking is employed to identify the binding orientation of drug candidates against protein targets to predict the drug's affinity and activity.[11] AutoDock Vina, an open-source molecular docking program, is used for performing molecular docking.[12] It was designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps

Research Institute. AutoDock Vina employs **the Royden-Fletcher-Goldfarb-Shanno (BFGS) method for local optimization,** an efficient quasi-Newton method.

1.2. Green River Oil Shale

The continuous advancement of civilization depends upon the seamless availability of energy sources as energy consumption is required in all sectors of life. According to the US energy information administration (EIA), the United States of America consumed energy equivalent to 92.94 quadrillion of British thermal units (BTU) in the year 2020**.** All these energies come from five different sources: petroleum (35%), natural gas (34%), coal (10%), renewable (12%), and nuclear (9%). Renewable energy sources consist of geothermal, solar, hydroelectric, wind, biomass, biofuel, and wood. Petroleum, the single largest energy source, provides 90% of the total transportation fuel and 33% of total industrial energy. As of December 31st, 2019; the proven reserve of crude oil in the United States was 44.2 billion barrels (BBL), while the total petroleum consumption throughout the same year (2019) was 7.49 BBL. The ratio of total petroleum reserve to yearly consumption yields a value close to six (~5.9). This small number suggests a potential shortage of petroleum in the near future given no new petroleum sources are discovered. Therefore, researchers are actively looking for newer and alternative sources of petroleum.

Oil shale is one of the most commonly found alternative sources of petroleum due to its hydrocarbon content named kerogen.[13] Oil shales have been found in 27 countries around the globe, including Estonia, Jordan, Morocco, China, the USA, etc. Among them, China and Estonia have started extracting crude oil from oil shale.[14] The largest deposit of oil shale in the world is the Green River formation in Utah, Wyoming, and Colorado.[15] This 54 million-year old Eocene geologic formation is estimated to be a reserve of 800 billion barrels (BBL) of crude oil, three times higher than the proven reserve of Saudi Arabia.[16] This amount can serve the US oil demand

for approximately 110 years at a consumption rate equal to 7.49 BBL. Oil shale is a dense, laminated sedimentary rock that is an organic-inorganic hetero-composite in composition.[17] In terms of composition, Green River oil shale contains 86.2% mineral matter, 2.76% bitumen, and 11.04% kerogen by weight.[18] Kerogen is one of the most carbonaceous materials on earth, with the significance of being the precursor to crude oil. This feature has provided the oil shale its primary economic value. The large networks of carbon, oxygen, hydrogen, nitrogen, and sulfur build up the kerogen macromolecule that produces shale oil upon pyrolysis.[17, 19] By taking into account the hydrogen to carbon ratio and oxygen to carbon ratio, kerogen can be categorized into Type I, Type II, and Type III kerogen.[20] Green River oil shale contains Type I kerogen, which is trapped inside the mineral matrix.

The minerals that compose the major portion of Green River oil shale primarily belong to carbonate, silicate, and sulfide minerals.[18] The carbonate minerals are Calcite, dolomite, nahcolite, trona, and dawsonite with calcite and dolomite being the dominant ones (43.1%). Montmorillonite (12.9%), quartz (8.6%), analcite (4.3%), and orthoclase (16.4%) are the silicate minerals that are present in Green River oil shale. The only sulfide mineral is pyrite found in a tiny quantity (~0.86%).[18] In essence, the Green River oil shale contains a similar amount of carbonate and silicate minerals.

The most important problem regarding shale oil extraction from oil shale is identify the effects of mineral matrix on the trapped kerogen during the pyrolysis procedure. The qualitative and quantitative distribution of hydrocarbons obtained during the oil shale pyrolysis process is affected by the presence of a mineral matrix.[21] It remains unclear how the organic kerogen interacts with the inorganic mineral matrix during the pyrolysis process as the researchers varied in their findings. A number of TG analyses reported that the whole mineral matrix impedes the kerogen

decomposition during pyrolysis.[22] Other researchers believed that individual mineral or mineral classes have specific impacts on kerogen pyrolysis. It was reported that the elimination of carbonate minerals decreases the amount of extracted hydrocarbons from oil shale.[23] On the contrary, some other researchers found that the amount of extracted organics increased when silicate minerals were drawn out.[24] Another TG study observed that montmorillonite clay facilitates the decomposition of oil shale.[25] It has also been hypothesized that only the majority mineral class affects the pyrolysis process.[26] If the carbonates make up the majority portion of oil shale, they stimulate the pyrolysis process. If the silicates constitute the greater part of oil shale, they suppress the pyrolysis process. As the carbonates and silicates are present in Green River oil almost equal amounts, it is difficult to understand the effect of individual mineral or mineral class effects.

Our group's previous experimental studies provided different information, including nanoscale morphology, nanomechanical properties, and molecular interactions behavior of the Green River oil shale. Scanning electron microscope (SEM) of oil shale showed that kerogen does not exist even in the 100s of nanometer (nm) scale, which infers that kerogen exists in the tens of nm sized pockets.[27] Nanoindentation studies showed that kerogen has an elastic modulus in the range of 5~11 GPa.[27] Fourier Transform Infrared spectroscopy showed that minerals have significant non-covalent interactions with kerogen. [14] Based on the findings of these studies, it is evident that kerogen significantly interacts with the mineral matrix at the nanometer scale. These nanoscale interactions can play a key role in the pyrolysis process of Green River oil shale. These nanoscale molecular interactions can be effectively investigated utilizing molecular dynamics simulations.

Katti research group has been investigating the kerogen-mineral interactions of Green River oil shale utilizing molecular dynamics simulations. The MD simulations of Na-montmorillonite clay with three kerogen moieties showed that pyridine and heptylamine moieties significantly interact with clay while quinaldine does not interact at all.[28] The MD simulations of the 12-unit kerogen model with montmorillonite clay show that larger fragments of kerogen interact more with clay mineral, and the interactions are mostly Van der Waals in nature. Though the carbonates comprise half of the mineral matter, no literature is available regarding the interactions between kerogen and carbonates. Moreover, no literature has beenon how Type I kerogen interacts with quartz, one of the most commonly found silicate minerals on earth. To address the mentioned knowledge gaps, this study conducts molecular dynamics simulations of Green River oil shale Type I kerogen with calcite and quartz to investigate their interactions behavior.

1.3. Swelling Clays

Expansive soil is characterized by the presence of a large proportion of swelling clay minerals. Swelling clay swells due to hydration and applies swelling pressure. Swelling pressure causes damages to civil infrastructures i.e. buildings, bridges, dams etc.[29] Swelling clays are commonly found all over the world, including USA and Canada. In the USA alone, swelling clays cause damage to civil infrastructures, resulting in an average annual damage cost of $20 billion. The swelling clays also find beneficial applications due to their swelling behavior. Swelling clays are commonly used as landfill liners to prevent the leaching of toxic fluids due to their high surface area and low permeability.[30] Swelling clays are also utilized in polymer-clay nanocomposites, pharmaceutical, and cosmetic industry, and modifying asphalt.

Na-montmorillonite (Na-MMT/Na-Mt), a member of the smectite group, is one of the most commonly found swelling clay minerals. Na-MMT is a 2:1 phyllosilicate mineral with two tetrahedral sheets that sandwich one octahedral sheet.[31] Therefore, Na-MMT is sometimes termed as a t-o-t clay mineral. Na-MMT clay mineral consists of interlayers between the clay sheets. The tetrahedral sheets are made of silica tetrahedrons while the octahedral sheet consists of alumina octahedrons. Isomorphous substitution occurs in alumina octahedral sheet where aluminum ions are replaced by magnesium or ferrous ions. This isomorphous substitution results in charge deficiency in MMT clay sheets.[32] The charge deficiency is stabilized by different cations that occupy the interlayer. When sodium (Na) cations balance the negative charge of clay sheets, the clay is termed as Na-montmorillonite (Na-MMT).

The hydration response of Na-MMT is one of the most investigated topics in clay mineralogy due to its widespread significance and applications. The hydration behavior of Na-MMT clay has long been investigated utilizing both experimental and computational approaches in our research group. Scanning electron microscopy (SEM) of hydrated MMT clays showed that increasing hydration reduces the particle size at the micron level.[33] X-ray diffraction (XRD) studies of Na-MMT clays exhibited that increasing hydration increases the interlayer spacing in nm level.[34] The findings of XRD studies have been modeled using MD simulations.[35] Discrete element modeling (DEM) of Na-MMT clay hydration exhibits particle breakdown.[36] All these observations indicate that Na-MMT clay hydration brings changes at both nanoscale and microscale. In order to make a connection between the nanoscale change and microscale change due to hydration, the hierarchical structure of Na-MMT needs to be investigated.

Na-MMT clay consists of a hierarchical structure. A significant number of experimental studies proved this important fact about Na-MMT clays. The hierarchical structure of Na-MMT

clay consists of four different stages: layer, tactoid, single aggregate, multiple aggregates.[37 38] Clay tactoid is termed as the basic clay particle as it houses the interlayer spacings. The molecular interactions and mechanical properties of clay tactoid can be significantly changed due to hydration. However, no literature is available regarding the molecular structure of tactoid and its mechanical properties. This study investigates the properties of tactoid molecular structure both in dry and hydrated conditions using molecular dynamics simulations. The mechanical properties of dry and hydrated tactoid are also investigated by employing steered molecular dynamics (SMD) simulations. The investigated mechanical properties are compression, tension, and shear. Modeling tactoid structure would help to understand clay hierarchical behavior. The unit cell of Na-MMT clay mineral contains 40 atoms with the dimensions of 5.28 Å × 9.14 Å × 6.56 Å, and the environmentally found clay minerals are sized on the order of hundreds of nm.[39]

1.4. Coronaviral Proteins

The beginning of 2020 will be remembered in history because of the emergence of COVID-19 caused by SARS-CoV-2 coronavirus. This ongoing global pandemic has not only caused severe health concerns but also affected every sector of our life as we know it. COVID-19 has made people around the world conform to a "New Normal' life. However, the emergence of coronavirus is not a new incident in human history. Coronaviruses were first discovered in 1956 and got this name due to 'crown-like appearance under an electron microscope.[40] Among the discovered 200 different coronaviruses, only 7 infect people i.e. HCoV-229E, HCoV-NL63, HCoV-OC43, HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2. [41 42] Among these seven coronaviruses, only SARS-CoV, MERS-CoV, and SARS-CoV-2 can pose life-threatening risks. The SARS (Severe Acute Respiratory Syndrome), caused by SARS-CoV, emerged from China's Guangdong province in 2002-03 and spread to 26 countries. SARS-CoV infected 8000 people while leading to the death

of about 774 people.[43] The MERS (Middle East Respiratory Syndrome), caused by MERS-CoV, was first reported in Jeddah of Saudi Arabia in 2012, and spread to 27 countries. MERS-CoV infected 2494 people while leading to the death of 858 people. [44] The first incidence of COVID-19, caused by SARS-CoV-2, was reported in December 2019 in the city of Wuhan, China. Then SARS-CoV-2 spread to 213 countries around the globe.[45] [46] According to the World Health Organization (WHO), COVID-19 has infected more than 248 million people while killing more than 5 million people as of the first week of November 2021. From these data, it's evident that the havoc caused by SARS-CoV-2 is much higher than SARS-CoV and MARS-CoV.

SARS-CoV and SARS-CoV-2 have a number of common features in terms of their origin and lifecycle. Both are bat-originated positive-stranded beta coronaviruses.[47] They also go through the identical procedure during their first two stages of host cell infection i.e. attachment and replication. Both coronaviruses perform the host cell attachment by binding their spike protein to human Angiotensin Converting Enzyme 2 (ACE2) receptor.[48] [49] The spike glycoprotein is a structural coronaviral protein, and it performs the host cell binding as well as membrane fusion.[50] Both SARS-CoV and SARS-CoV-2 perform their replication mechanism through the polymerase activity of RNA dependent RNA polymerase (RdRp) inside the host cell.[51] RdRp is the largest domain of coronaviral non-structural protein 12 (nsp12) and it performs the polymerase activity with the aid of two other non-structural proteins nsp7 and nsp8.[52] [53] Therefore, nsp12-nsp7-nsp8 is considered as the primary replication hub of SARS-CoV and SARS-CoV-2 coronaviruses. Remdesivir, a broad-spectrum antiviral agent against RNA viruses, has been considered for the treatment of coronaviral infection.[54] This drug is targeted against the RdRp domain of nsp12.[55]

SARS and COVID-19 diseases cause a similar response in infected humans as the viral attachment is performed by spike-ACE2 interactions, and replication is performed by nsp12-nsp7-

nsp8 complex in both cases. However, the discrepancy in the number of infections and deaths between SARS and COVID-19 indicates otherwise. The number of ARDS (Acute Respiratory Distress Syndrome) incidents is much higher for COVID-19 compared to SARS. Besides, SARS-CoV-2 has been found to have higher transmissibility than SARS-CoV.[56] Therefore, it might be possible that there exist some differences in spike-ACE2 interactions and nsp12-nsp7-nsp8 interactions between SARS-CoV and SARS-CoV-2.

Molecular dynamics simulations are employed to investigate SARS-CoV and SARS-CoV-2 coronaviral protein-protein interactions (PPI) during attachment and replication. To investigate the spike-ACE2 attachment interactions, both MD and SMD simulations are utilized. MD simulations examine the non-bonded interactions between coronaviral spike and host ACE2 proteins. SMD simulations are used to inspect the ACE2 mechanical behavior in the presence of spike protein. The interactions within nsp12-nsp7-nsp8 replication complex are investigated using MD for both coronaviruses. In the case of SARS-CoV-2, the impact of Remdesivir Monophosphate (RMP) on the RdRp domain of nsp12 is explored using MD to understand the inhibitory effect of Remdesivir on coronaviral replication.

1.5. Breast Cancer Drug Discovery

Breast cancer is one of the most common cancers among women. According to the American Cancer Society, the estimated number of new breast cancer cases in 2021 is higher than any other cancer type. The estimated number of breast cancer-related deaths among women in 2021 is 43600. From 2013-2017, the average annual breast cancer incident rate among women was 126 per 100000. From 2014-2018, the average annual death rate due to Breast cancer is 20.1 per 100000 women. These numbers suggest the severity of breast cancer among women.

Breast cancer treatment measures can be categorized into local and systemic classes.[57] Local treatment is performed through surgery or radiation or by both to remove, destroy and control the growth of breast cancer cells. Systemic treatments employ drugs orally/intravenously to treat breast cancer by killing/controlling the cancer cells throughout the body. Chemotherapy, Hormone therapy, Targeted therapy, and Immunotherapy are the common types of systemic treatments. Chemotherapy (chemo) uses anti-cancer drugs at different stages of cancer that may be given orally or intravenously to kill cancer cells. Types of breast cancer that require estrogen or progesterone hormone for the growth of cancer cells (in other words, breast cancer cells containing these hormone receptors) are treated by Hormone therapeutics that prevent the attachment of hormones to receptors. Targeted therapy halts cancer cells' proliferation by targeting specific substances of cancer cells (generally proteins) while keeping the normal cells mostly unharmed (unlike chemotherapy). Immunotherapy fights cancer by stimulating a person's immune system against cancer by providing lab-made immune-like components inside the body.

Conventional synthetic chemo-preventive agents kill the cancer cells at the cost of cytotoxic effects to normal cells. The effectiveness of chemo-preventive agents can be decreased due to the mutation-induced drug resistance against them. Phytochemicals have long been attracting researchers as anti-cancer agents possessing low toxicity. Phytochemicals can reverse multi-drug resistance (MDR) by affecting the expression of ABC (ATP-binding cassette) transporters.[58] Therefore, phytochemicals are attracting researchers as potential anti-cancer agents.

The potency of phytochemicals as anti-cancerous agents can be explored using Molecular Docking. Molecular Docking is primarily employed to predict the binding orientation of drug candidates against protein targets to predict the drug's affinity and activity.[11] This study utilizes molecular docking to perform the virtual screening of two different plant extracts for treating breast

cancer. The two chosen plants are Rhodiola (***Rhodiola Crenulata*****),** and Oregano ***(Origanum Vulgare*****).** For the docking study, 28 compounds of Rhodiola, and 22 compounds of Oregano have been selected. The selected compounds primarily belong to polyphenols, flavonoids, flavonols, terpenoids, carotenoids etc. These compounds are targeted against a number of breast cancer-related proteins as chemoprevention of breast cancer takes place in various ways. By considering the pathways related to cellular death induction, inhibiting cell proliferation, modifying signal transduction, and blocking hormones to receptors, a total of 33 breast cancer proteins have been selected. AutoDock Vina is used for performing the docking of phytochemicals against breast cancer proteins. The phytochemicals showing higher affinity will be considered as the potential anti-cancer agents.

1.6. Objectives of this Study

The primary objectives of this study are

- To model the interactions between Green River oil shale Type I kerogen and calcite (CaCO3) mineral using molecular dynamics simulations. Moreover, to investigate interactions within kerogen fragments in the presence of calcite mineral.
- To investigate the interactions between Green River oil shale Type I kerogen and quartz (SiO_2) mineral using molecular dynamics simulations. Moreover, to investigate interactions within kerogen fragments in the presence of quartz mineral.
- To inspect the interactions within Na-montmorillonite clay tactoid model using molecular dynamics simulations. The mechanical response of clay tactoid under compression, tension, and shear are also to be investigated utilizing steered molecular dynamics.

- To examine the interactions within hydrated Na-montmorillonite clay tactoid model using molecular dynamics simulations. The mechanical response of hydrated clay tactoid under compression is also to be investigated utilizing steered molecular dynamics.
- To determine the interactions of human ACE2 receptor with SARS-CoV and SARS-CoV-2 spike protein receptor binding domain utilizing molecular dynamics simulations. Moreover, to investigate the mechanical response of ACE2 in the presence of both coronavirus spike proteins utilizing steered molecular dynamics.
- To determine the interactions within coronaviral non-structural protein complex (nsp12-nsp7-nsp8) for both SARS-CoV and SARS-CoV-2 utilizing molecular dynamics simulations. The impact of Remdesivir monophosphate (RMP) on the RNA dependent RNA polymerase domain of SARS-CoV-2 nsp12 to be investigated.
- To perform the virtual screening of Rhodiola, and Oregano compounds against a series of breast cancer proteins to investigate their potency as anti-cancer agents.

1.7. Original Significance

This study utilizes different in-silico approaches to address scientific problems regarding four different material systems. The conducted in-silico approaches are molecular dynamics simulation, steered molecular dynamics, coarse-grained molecular dynamics, and molecular docking. The investigated material systems are oil shale, swelling clays, coronaviral proteins, and plant-derived phytochemicals. The computational approaches are employed to investigate kerogen-mineral interactions in Green River oil shale, Na-montmorillonite clay tactoid structure and its mechanical properties in the dry and hydrated state, coarse-grained modeling of Na-MMT

clay, coronaviral protein-protein interactions, and virtual screening of phytochemicals for breast cancer drug discovery.

Green River oil shale is a reserve of 800 billion barrels of crude oil in solid rock form named kerogen trapped inside a matrix of minerals. Extracting kerogen from the mineral matrix is the first step towards crude oil production. Kerogen can be extracted by interrupting the kerogen-mineral interactions, and efficient extraction depends on identifying kerogen-mineral interactions. This study uses MD simulations to identify the interactions of kerogen with two significant minerals of Green River oil shale i.e., calcite and quartz. These identified kerogen-calcite and kerogen-quartz interactions can be blocked for the efficient isolation of kerogen.

Na-montmorillonite (Na-MMT) is one of the most commonly found swelling clay minerals in the USA. By applying swelling pressure and damaging infrastructures upon hydration, Na-MMT costs annually approximately 20 billion USD only in the US per year. This study investigates their hierarchical structure and mechanical properties both in dry and hydrated conditions at the molecular level utilizing MD and SMD simulations. The findings will help to better understand their hydration behavior at the hierarchical level.

COVID-19 (caused by SARS-CoV-2), the latest addition to coronavirus-related diseases, share much common features with previously emerged SARS disease (caused by SARS-CoV). Though their contributing coronaviruses SARS-CoV-2 and SARS-CoV follow identical attachment and replication mechanisms, the number of infections and deaths due to COVID-19 is much higher. This study performs MD and SMD simulations of spike-ACE2 protein complexes and nsp12-nsp7-nsp8 protein complexes of both coronaviruses to investigate their attachment and replication behavior at the molecular level. The findings indicate how SARS-CoV-2 behaves

differently than SARS-CoV and which features might contribute to the higher deadliness and contagiousness of SARS-CoV-2.

Breast cancer is one of the most predominant life-threatening diseases among women. Phytochemicals are being considered as potential anti-cancer drugs as they are free from cytotoxic effects possessed by traditional chemo-synthetic drugs. This study investigates the potency of Rhodiola, and Oregano extract compounds as breast cancer drug candidates through molecular docking against breast cancer-related proteins. The compounds with high-affinity values can be investigated for breast cancer drug discovery.

1.8. Dissertation Organization

This dissertation is organized into following manner:

1. Chapter 1: Introduction
2. Chapter 2: This chapter investigates kerogen-calcite interactions utilizing MD simulations.
3. Chapter 3: This chapter investigates kerogen-quartz interactions utilizing MD simulations.
4. Chapter 4: This chapter investigates Na-montmorillonite tactoid structure and its mechanical properties utilizing MD and SMD simulations.
5. Chapter 5: This chapter investigates hydrated Na-montmorillonite tactoid structure and its mechanical properties utilizing MD and SMD simulations.
6. Chapter 6: This chapter examines coronaviral spike-ACE2 attachment interactions and ACE2 mechanical behavior for both SARS-CoV and SARS-CoV-2 using MD and SMD simulations.

7. Chapter 7: This chapter examines coronaviral nsp12-nsp7-nsp8 replication interactions and impact of Remdesivir Monophosphate on SARS-CoV-2 utilizing MD simulations.
8. Chapter 8: This chapter investigates the potency of plant-derived phytochemicals against breast cancer target proteins using Molecular Docking.
9. Chapter 9: This chapter presents summary and major conclusions presented in this dissertation.
10. Chapter 10: This chapter discusses the future directions for the research in this field.

1.9. References

(1) Young, D. Computational chemistry: a practical guide for applying techniques to real world problems; John Wiley & Sons, 2004.

(2) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC* **2015**, *8*, 37.

(3) Buehler, M. J. Introduction to Mechanics of Materials. **2007**.

(4) Stillinger, F. H.; Rahman, A. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics* **1974**, *60* (4), 1545-1557.

(5) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267* (5612), 585-590.

(6) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289.

(7) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(8) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered molecular dynamics. In *Computational molecular dynamics: challenges, methods, ideas*, Springer, 1999; pp 39-65.

(9) Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics & Modelling* **2001**, *19* (1), 13-25. DOI: 10.1016/s1093-3263(00)00133-9.

(10) Morris, G. M.; Lim-Wilby, M. Molecular docking. In *Molecular modeling of proteins*, Springer, 2008; pp 365-382.

(11) Chaudhary, K. K.; Mishra, N. A review on molecular docking: novel tool for drug discovery. *databases* **2016**, *3* (4), 1029.

(12) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, *31* (2), 455-461.

(13) Deng, S.; Wang, Z.; Gu, Q.; Meng, F.; Li, J.; Wang, H. Extracting hydrocarbons from Huadian oil shale by sub-critical water. *Fuel Processing Technology* **2011**, *92* (5), 1062-1067. DOI: 10.1016/j.fuproc.2011.01.001.

(14) Alstadt, K. N.; Katti, D. R.; Katti, K. S. An in situ FTIR step-scan photoacoustic investigation of kerogen and minerals in oil shale. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2012**, *89*, 105-113. DOI: 10.1016/j.saa.2011.10.078.

(15) Guo, H.; Lin, J.; Yang, Y.; Liu, Y. Effect of minerals on the self-heating retorting of oil shale: Self-heating effect and shale-oil production. *Fuel* **2014**, *118*, 186-193. DOI: 10.1016/j.fuel.2013.10.058.

(16) Taskforce on Strategic Unconventional Fuels. 2007. America's Strategic Unconventional Fuels. Washington (DC): US Government.

(17) Karabakan, A.; Yurum, Y. Effect of the mineral matrix in the reactions of oil shales: 1. Pyrolysis reactions of Turkish Goynuk and US Green River oil shales. *Fuel* **1998**, *77* (12), 1303-1309. DOI: 10.1016/s0016-2361(98)00045-3.

(18) Cane, R. F. In Oil Shale·, Yen, T. F., Chilingarian, GV, Eds. Elsevier: Amsterdam: 1976.

(19) Yan, J.; Jiang, X.; Han, X.; Liu, J. A TG-FTIR investigation to the catalytic effect of mineral matrix in oil shale on the pyrolysis and combustion of kerogen. *Fuel* **2013**, *104*, 307-317. DOI: 10.1016/j.fuel.2012.10.024.

(20) Razvigorova, M.; Budinova, T.; Tsyntsarski, B.; Petrova, B.; Ekinci, E.; Atakul, H. The composition of acids in bitumen and in products from saponification of kerogen: Investigation of their role as connecting kerogen and mineral matrix. *International Journal of Coal Geology* **2008**, *76* (3), 243-249. DOI: 10.1016/j.coal.2008.07.011.

(21) Eglinton, T. I.; Rowland, S. J.; Curtis, C. D.; Douglas, A. G. Kerogen mineral reactions at raised temperatures in the presence of water. *Organic Geochemistry* **1986**, *10* (4-6), 1041-1052. DOI: 10.1016/s0146-6380(86)80043-2.

(22) Aboulkas, A.; El Harfi, K. Study of the kinetics and mechanisms of thermal decomposition of moroccan tarfaya oil shale and its kerogen. *Oil Shale* **2008**, *25* (4), 426-443. DOI: 10.3176/oil.2008.4.04.

(23) Ballice, L. Effect of demineralization on yield and composition of the volatile products evolved from temperature-programmed pyrolysis of Beypazari (Turkey) Oil Shale. *Fuel Processing Technology* **2005**, *86* (6), 673-690. DOI: 10.1016/j.fuproc.2004.07.003.

(24) Sert, M.; Ballice, L.; Yuksel, M.; Saglam, M. Effect of mineral matter on product yield and composition at isothermal pyrolysis of turkish oil shales. *Oil Shale* **2009**, *26* (4), 463-474. DOI: 10.3176/oil.2009.4.03.

(25) Borrego, A. G.; Prado, J. G.; Fuente, E.; Guillen, M. D.; Blanco, C. G. Pyrolytic behaviour of Spanish oil shales and their kerogens. *Journal of Analytical and Applied Pyrolysis* **2000**, *56* (1), 1-21. DOI: 10.1016/s0165-2370(99)00092-3.

(26) Pan, L.; Dai, F.; Huang, P.; Liu, S.; Li, G. Study of the effect of mineral matters on the thermal decomposition of Jimsar oil shale using TG-MS. *Thermochimica Acta* **2016**, *627*, 31-38. DOI: 10.1016/j.tca.2016.01.013.

(27) Alstadt, K. N.; Katti, K. S.; Katti, D. R. Nanoscale Morphology of Kerogen and In Situ Nanomechanical Properties of Green River Oil Shale. *Journal of Nanomechanics and Micromechanics* **2016**, *6* (1). DOI: 10.1061/(asce)nm.2153-5477.0000103.

(28) Katti, D. R.; Upadhyay, H. B.; Katti, K. S. Molecular interactions of kerogen moieties with Na-montmorillonite: An experimental and modeling study. *Fuel* **2014**, *130*, 34-45. DOI: 10.1016/j.fuel.2014.04.009.

(29) Grim, R. E. Clay Mineralogy McGraw-Hill. *New York* **1968**, 206. Buzzi, O.; Fityus, S.; Sloan, S. W. Use of expanding polyurethane resin to remediate expansive soil foundations. *Canadian Geotechnical Journal* **2010**, *47* (6), 623-634. DOI: 10.1139/t09-132.

(30) Katti, D. R.; Patwary, Z. R.; Katti, K. S. Modelling clay-fluid interactions in montmorillonite clays. *Environmental Geotechnics* **2017**, *4* (5), 322-338. DOI: 10.1680/jenge.14.00027.

(31) Grim, R. E. Modern concepts of clay materials. *Journal of Geology* **1942**, *50* (3), 225-275. DOI: 10.1086/625050.

(32) Hendricks, S. B.; Jefferson, M. E. Structures of kaolin and talc-pyrophyllite hydrates and their bearing on water sorption of the clays. *American Mineralogist* **1938**, *23* (12), 863-875.

(33) Katti, D.; Shanmugasundaram, V. Influence of swelling on the microstructure of expansive clays. *Canadian Geotechnical Journal* **2001**, *38* (1), 175-182. DOI: 10.1139/cgj-38-1-175.

(34) Amarasinghe, P. M.; Katti, K. S.; Katti, D. R. Molecular Hydraulic Properties of Montmorillonite: A Polarized Fourier Transform Infrared Spectroscopic Study. *Applied Spectroscopy* **2008**, *62* (12), 1303-1313.

(35) Pradhan, S. M.; Katti, K. S.; Katti, D. R. Evolution of Molecular Interactions in the Interlayer of Na-Montmorillonite Swelling Clay with Increasing Hydration. *International Journal of Geomechanics* **2015**, *15* (5). DOI: 10.1061/(asce)gm.1943-5622.0000412.

(36) Katti, D. R.; Matar, M. I.; Katti, K. S.; Amarasinghe, P. M. Multiscale modeling of swelling clays: A computational and experimental approach. *Ksce Journal of Civil Engineering* **2009**, *13* (4), 243-255. DOI: 10.1007/s12205-009-0243-0.

(37) Perdrial, J. N.; Warr, L. N. Hydration Behavior of Mx80 Bentonite in a Confined-Volume System: Implications for Backfill Design. *Clays and Clay Minerals* **2011**, *59* (6), 640-653. DOI: 10.1346/ccmn.2011.0590609.

(38) Jullien, M.; Raynal, J.; Kohler, E.; Bildstein, O. Physicochemical reactivity in clay-rich materials: Tools for safety assessment. *Oil & Gas Science and Technology-Revue D Ifp Energies Nouvelles* **2005**, *60* (1), 107-120. DOI: 10.2516/ogst:2005007.

(39) Cadene, A.; Durand-Vidal, S.; Turq, P.; Brendle, J. Study of individual Na-montmorillonite particles size, morphology, and apparent charge. *Journal of Colloid and Interface Science* **2005**, *285* (2), 719-730. DOI: 10.1016/j.jcis.2004.12.016.

(40) Tyrrell, D. A.; Bynoe, M. L. Cultivation of Viruses From a High Proportion of Patients With Colds. *Lancet* **1966**, *1* (7428), 76-&.

(41) Tong, S.; Conrardy, C.; Ruone, S.; Kuzmin, I. V.; Guo, X.; Tao, Y.; Niezgoda, M.; Haynes, L.; Agwanda, B.; Breiman, R. F.; et al. Detection of Novel SARS-like and Other Coronaviruses in Bats from Kenya. *Emerging Infectious Diseases* **2009**, *15* (3), 482-485. DOI: 10.3201/eid1503.081013.

(42) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**. DOI: 10.1038/s41586-020-2012-7.

(43) Chan-Yeung, M.; Xu, R. H. SARS: epidemiology. Respirology 8 Suppl S9–14. 2003.

(44) Lee, J.; Chowell, G.; Jung, E. A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: A retrospective analysis on control interventions and superspreading events. *Journal of Theoretical Biology* **2016**, *408*, 118-126. DOI: 10.1016/j.jtbi.2016.08.009.

(45) Tan, W.; Zhao, X.; Ma, X.; Wang, W.; Niu, P.; Xu, W.; Gao, G. F.; Wu, G. Z. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019−2020. *China CDC Weekly* **2020**, *2* (4), 61-62.

(46) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **2020**, *382* (8), 727-733. DOI: 10.1056/NEJMoa2001017.

(47) Gonzaalez, J. M.; Gomez-Puertas, P.; Cavanagh, D.; Gorbalenya, A. E.; Enjuanes, L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Archives of Virology* **2003**, *148* (11), 2207-2235. DOI: 10.1007/s00705-003-0162-1.

(48) Li, W. H.; Moore, M. J.; Vasilieva, N.; Sui, J. H.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greenough, T. C.; et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426* (6965), 450-454. DOI: 10.1038/nature02145.

(49) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, 1-6.

(50) Beniac, D. R.; Andonov, A.; Grudeski, E.; Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology* **2006**, *13* (8), 751-752. DOI: 10.1038/nsmb1123.

(51) Ahn, D.-G.; Choi, J.-K.; Taylor, D. R.; Oh, J.-W. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Archives of virology* **2012**, *157* (11), 2095-2104.

(52) Velthuis, A.; Arnold, J. J.; Cameron, C. E.; van den Worm, S. H. E.; Snijder, E. J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Research* **2010**, *38* (1), 203-214. DOI: 10.1093/nar/gkp904.

(53) Subissi, L.; Posthuma, C. C.; Collet, A.; Zevenhoven-Dobbe, J. C.; Gorbalenya, A. E.; Decroly, E.; Snijder, E. J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111* (37), E3900-E3909. DOI: 10.1073/pnas.1323705111.

(54) Siegel, D.; Hui, H. C.; Doerffler, E.; Clarke, M. O.; Chun, K.; Zhang, L.; Neville, S.; Carra, E.; Lew, W.; Ross, B. Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo [2, 1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. ACS Publications: 2017.

(55) Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**.

(56) Hui, K. P. Y.; Cheung, M.-C.; Perera, R. A. P. M.; Ng, K.-C.; Bui, C. H. T.; Ho, J. C. W.; Ng, M. M. T.; Kuok, D. I. T.; Shih, K. C.; Tsao, S.-W. Tropism, replication competence, and innate immune responses of the coronavirus SARS-CoV-2 in human respiratory tract and conjunctiva: an analysis in ex-vivo and in-vitro cultures. *The Lancet Respiratory Medicine* **2020**.

(57) Israel, B. e. B.; Tilghman, S. L.; Parker-Lemieux, K.; Payton-Stewart, F. Phytochemicals: Current strategies for treating breast cancer. *Oncology letters* **2018**, *15* (5), 7471-7478.

(58) Tinoush, B.; Shirdel, I.; Wink, M. Phytochemicals: potential lead molecules for MDR reversal. *Frontiers in Pharmacology* **2020**, *11*, 832.

2. MODELING THE BEHAVIOR OF ORGANIC KEROGEN IN THE PROXIMITY OF CALCITE MINERAL BY MOLECULAR DYNAMICS SIMULATIONS[1]

2.1. Introduction

Petroleum has been considered as the principal source of energy for about a century. According to the U.S. Energy Information Administration (EIA), the United States of America consumed energy equivalent to 97.7 quadrillions of British thermal units (BTU) in the year 2017 [1]. The petroleum alone provided 37.1% of the total consumed energy, which is the highest among all the energy sources (petroleum, natural gas, coal, nuclear, renewable). The U.S. EIA has also stated that the proven reserves of crude oil in the United States were 39.2 billion barrels as of December 2017 while the total oil consumption the same year was 7.28 billion barrels (BBL) [2]. These numbers infer that it is crucial to find newer sources of petroleum. Oil shale has long been attracting the attention of researchers as an alternative source of petroleum [3]. Different countries around the globe, including Estonia, Jordan, Morocco, and China, are utilizing oil shale for the production of petroleum [4]. Green River formation, an Eocene geologic formation over 54 million years old, is appraised as the largest storage of shale oil in the world, a reserve of 800 billion barrels equivalent of crude oil [5, 6]. Assuming the petroleum consumption rate same as of 2017 (7.28 BBL/year), the Green River oil shale alone can serve the US oil demand for 110 years.

Oil shale can be described as a fine-grained sedimentary rock containing kerogen and minerals. Kerogen, an organic macromolecule composed of carbon, oxygen, hydrogen, nitrogen, and sulfur, produces shale oil upon pyrolysis [7, 8]. However, in Green River oil shale, the kerogen is locked in 10s of nanometer-sized pores within the minerals [9]. This minerals consists of ten major

[1] The contents of this chapter have been published in Faisal, HM Nasrullah, Kalpana S. Katti, and Dinesh R. Katti. "Modeling the Behavior of Organic Kerogen in the Proximity of Calcite Mineral by Molecular Dynamics Simulations." *Energy & Fuels* 34.3 (2020): 2849-2860. As the lead author of this journal article, the copyright permission has been obtained from American Chemical Society (ACS).

minerals from carbonate, silicate, and sulfide mineral groups [10]. Calcite, dolomite, nahcolite, trona, and dawsonite minerals fall within the carbonate group while montmorillonite, quartz, analcite, and orthoclase belong to the silicate group. Pyrite is the only sulfide mineral present in the oil shale. Among all these minerals, carbonate minerals constitute almost two-fifths (43.1% by weight) of Green River oil shale [10].

The coexistence of carbonate minerals with kerogen in oil shale has been a topic of great interest among the researchers. It is still unclear whether the carbonate minerals have an impact on the pyrolysis of kerogen. Several previous studies concluded that the thermal fracking of kerogen is hindered by carbonate minerals present in oil shale [11, 12]. On the other hand, some experimental results exhibited that carbonates act as catalysts in kerogen pyrolysis as their elimination from oil shale reduces the hydrocarbon output [13, 14]. Another study from different viewpoints stated that the critical factor is the relative quantity of carbonate minerals compared to other minerals. If the carbonate minerals constitute a smaller portion of total mineral content, they do not show any influence on kerogen pyrolysis. However, if the oil shale contains a higher amount of carbonates, they catalyze the pyrolysis process [15]. In Green River oil shale, the carbonate minerals share almost the same quantity (by weight) as other minerals. As a result, the contribution of carbonate minerals on Green River oil shale pyrolysis remains ambiguous. Modeling the interactions between carbonates and kerogen at the molecular level may guide us towards the understanding of carbonate minerals' effect upon kerogen and also provide guidance for efficient extraction of the energy source.

The Katti research group has studied the organic-mineral interactions in various material systems through experiments and computer simulations. Our previous studies indicate that non-bonded interactions control the mechanical properties of organic heteropolymers that reside next

to inorganic minerals in nanometer dimensions. Biological systems such as human bone [16, 17], and seashells [18] exhibit this phenomenon. The same behavior can also be found in hydroxyapatite nanocomposites [19, 20], and polymer clay nanocomposites [21-23]. FTIR photoacoustic experiments of *in situ* Green River oil shale showed changes to kerogen spectra due to kerogen-mineral non-bonded interactions [4]. Molecular dynamics simulation was employed to determine the nonbonded interactions of clay mineral with three moieties of kerogen, i.e., pyridine, quinaldine, and heptylamine [24]. We also constructed a complete 3D molecular model of kerogen to find its interaction with silicate clay minerals [25, 26].

Calcite ($CaCO_3$) is one of the most common geological minerals. It is a stable form of calcium carbonate primarily found in sedimentary formations [27]. It influences different environmental processes like global carbon cycling [28], contaminant transport [29], local pH and alkalinity of terrestrial environments [30], etc. Calcite is a major carbonate (CO_3^{2-}) mineral present in the Green River oil shale. As per our knowledge, there has been no conclusive study regarding the nanoscale interactions between Type I kerogen and calcite mineral of Green River oil shale.

Molecular dynamics (MD) is a computational simulation technique that attempts to mimic the motion and position of atoms and molecules using an energy function, also called as the force field and Newton's laws. A force field containing a set of functions and associated constants is used to define the energy expression for molecular mechanics calculation. This molecular mechanics then predicts the molecular shape and interactions [31]. Quantum Mechanical (*ab-initio*) approaches have been successfully implemented to predict the interfacial behavior of calcium carbonate mineral in biological system [32]. However, present computational resources cannot support the *ab-initio* calculation of large macromolecular (more than 1000 atoms) systems. The aim of our current study is to assess the molecular interactions between kerogen and calcite mineral

of Green River oil shale at Normal Temperature (300K) and Pressure (1.01325 bar) conditions (NTP) by using molecular dynamics simulations. This research seeks to investigate the influence on kerogen due to the presence of calcite mineral and the interactions between the mineral and kerogen in a laboratory environment.

2.2. Methodology

2.2.1. Model Construction

Calcite ($CaCO_3$) has a Trigonal crystal system. It is also considered to possess the Hexagonal crystal system as Trigonal and Hexagonal crystal systems are mathematically equivalent. It falls into a group of anhydrous carbonates called calcite. The other carbonates of this group are Magnesite ($MgCO_3$), Siderite ($FeCO_3$), Rhodochrosite ($MnCO_3$), and Smithsonite ($ZnCO_3$). The triangular carbonate group (CO_3^{2-}) is the basic building unit of calcite mineral [33]. The carbonate group is hypothesized to have a resonant structure, and it does not share oxygen with other carbonate groups. Calcite ($CaCO_3$) can be viewed as a modified NaCl structure whose Na ions are replaced by Ca ions, and Cl ions are replaced by CO_3^{2-} groups. The carbonate groups and Ca ions lie in alternating planes in such a way that the carbonate groups are arranged at right angles to three-fold axis while the Ca ions are positioned at the coordination of six with the oxygen atoms. The space group of calcite is $R\bar{3}C$ (No. 167). The unit cell dimensions of calcite are a=b=4.99 Å, c=17.06 Å, and $\alpha=\beta=90°$, $\Upsilon=120°$ [34]. These lattice parameters and atomic positional parameters have been utilized to build the calcite unit cell **(Figure 1a)**. We have used Materials Studio 7.0 to build this crystal structure. To validate our constructed model, we performed 'X-ray Powder Diffraction' of this calcite unit cell model **(Figure 1b)**. The 'Reflex' module of Materials Studio 7.0 was employed for performing powder diffraction. The output diffraction peaks match

perfectly with experimental diffraction peaks of calcite sample found in nature [35]. MATCH! 3.7.0 was used to match diffraction peaks.

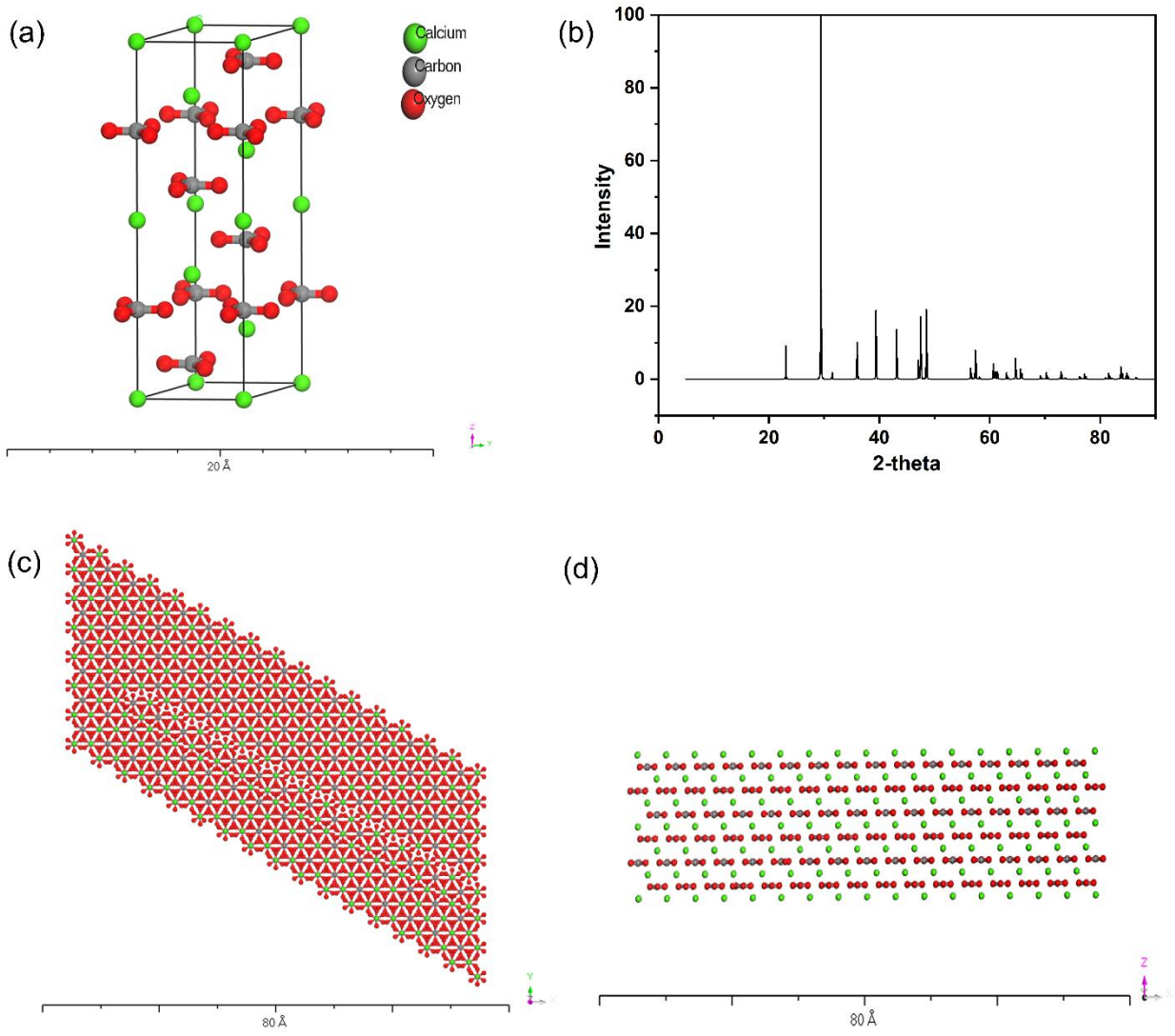

Figure 2.1. (a) Molecular model of a unit cell of calcite ($CaCO_3$), (b) XRD pattern of constructed calcite unit cell model, (c) expanded molecular model (16×7) of calcite along Z-axis and (d) perpendicular to Z-axis. All the molecular representations were generated by Materials Studio 7.0.

We built an expanded model of calcite using Materials Studio 7.0 to observe interactions with the 3D kerogen model obtained from our previous study [25]. The expanded calcite model is 16×7×1-unit cells, with 16 unit cells in X-direction and 7 unit cells in Y-direction **(Figure 1c)**. The resultant dimensions of the calcite expanded model are 79.83 Å × 34.92 Å × 17.06 Å. It constitutes

a total of 3736 atoms. The calcium ions and carbonate groups sit in alternating planes throughout the structure **(Figure 1d).**

The complete 3D kerogen model comprising of 12 kerogen units has been obtained from our previous study [25]. This model has been validated with experimental data and a model for Green River Kerogen available in the literature[36]. The size of the model is 71.92 Å × 88.98 Å × 63.89 Å **(Fig. 2a, 2b)**. Siskin's 2D kerogen model was used as the foundation of the 12-unit 3D kerogen model [36, 37]. Seven hydrocarbon fragments make up each kerogen unit. Hence the 12-unit kerogen model accommodates 12 segments of each fragment. These kerogen fragments are: fragment 1 ($C_{18}H_{38}$), fragment 2 ($C_{18}H_{30}$), fragment 3 ($C_{20}H_{42}$), fragment 4 ($C_{45}H_{60}O$), fragment 5 ($C_{75}H_{117}N_4O_5$), fragment 6 ($C_{102}H_{167}NOS_2$), and fragment 7 ($C_{367}H_{547}N_{10}O_{10}S_2$). Five independent ammonium (NH_4) groups, associated with seven hydrocarbon fragments in each kerogen unit, will be considered as fragment 8. The top kerogen chain (layer) contains six kerogen units, while the bottom kerogen chain (layer) contains the remaining six kerogen units situated at the bottom along Z-axis.

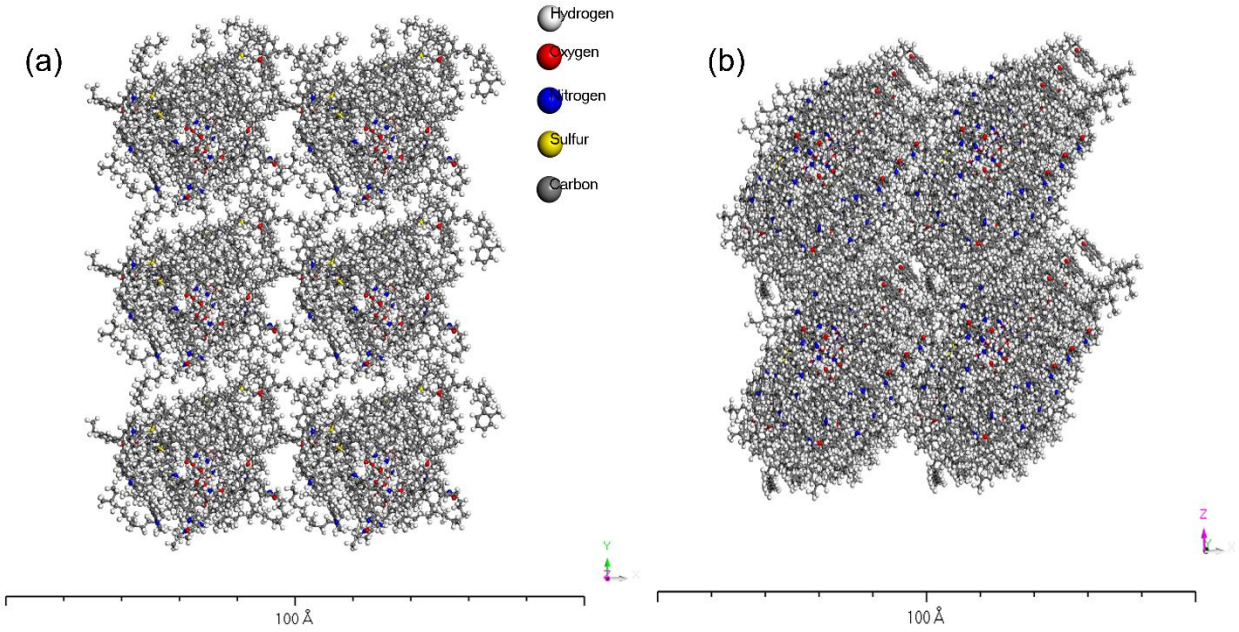

Figure 2.2. Molecular models of 12-unit 3D kerogen (a) along Z-axis, and (b) perpendicular to the Z-axis. All the representations were generated by Materials Studio 7.0.

Each kerogen layer (top and bottom) contains 30 ammonium groups. We merged the structures of expanded calcite model and 12-unit kerogen model using the "Merge Structures" plugin of Visual Molecular Dynamics (VMD) 1.9.3 **(Figure 3a, 3b)**. The expanded calcite model was placed on top of the kerogen model along the Z-axis. FTIR photoacoustic experiments were conducted on Green River oil shale samples by the authors in the range 4000-400 cm^{-1} in a previous study[4]. The resultant FTIR bands confirmed the presence of organic and minerals in the oil shale. The FTIR spectra exhibited the presence of O-H stretching around 3660 cm^{-1} that is attributed to the lattice hydroxyl group of montmorillonite clays. This lattice OH is an integral part of clay structure. Broad bands (unlike the sharp structural OH) in this region are suggestive of adsorbed water. These spectra do not indicate the nature of this water or the amount of water. In a previous study on another organic (TRAP- a protein in tooth enamel [38]) with mineral hydroxyapatite we found that amount of water impacted interactions between TRAP and mineral.

Hence, in our molecular modeling study, we have not considered the presence of any explicit water.

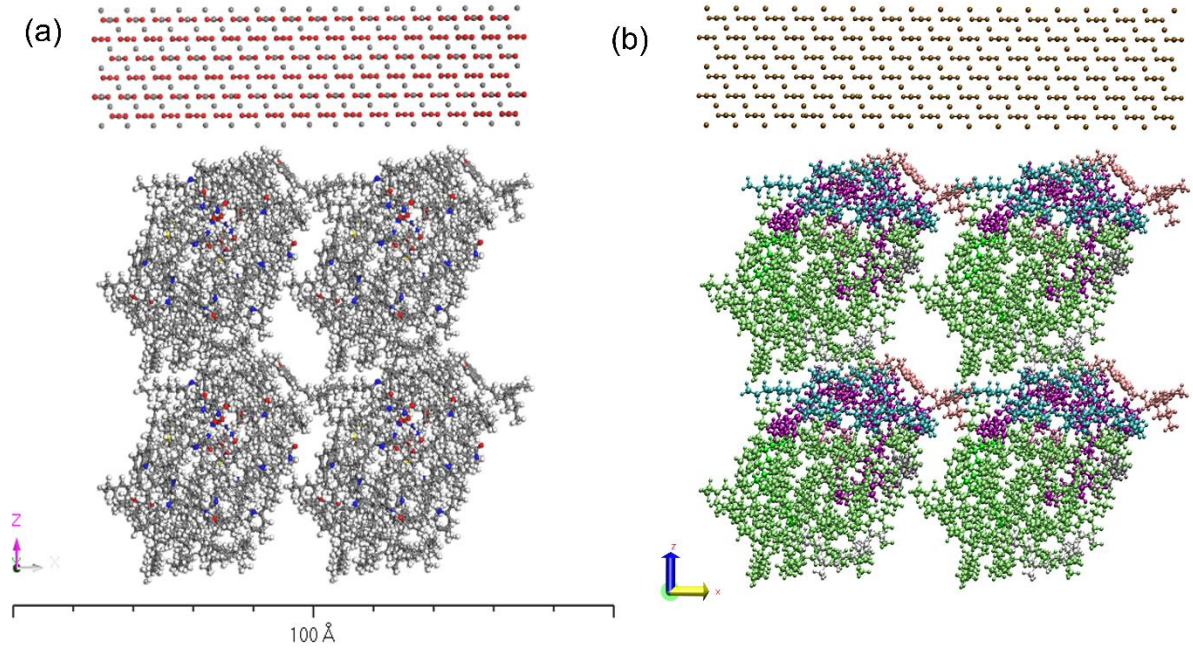

Figure 2.3. Initial molecular model of merged, expanded calcite, and 12-unit kerogen (a) general view generated by Materials Studio 7.0, and (b) fragment-based view generated by VMD 1.9.3. Different colors represent different fragments of kerogen.

2.2.2. Model Parameterization

Several research groups performed computational studies of calcite mineral using different models and force fields. Some researchers derived force field parameters of calcite using core-shell model [39, 40], while others utilized rigid ion model [41-43]. To parameterize our calcite model, we are using the parameters based on rigid ion model [43]. The rigid ion model parameters are transformed to the force field parameters of the desired CHARMm force field. CHARMm (Chemistry at HARvard Macromolecular Mechanics) is used for molecular dynamics simulations [44]. The CHARMm force field can be expressed as:

$$E = E_{bond} + E_{angle} + E_{improper} + E_{VDW} + E_{electrostatic} \quad (2.1)$$

$$E = k_r(r - r_0)^2 + k_\theta(\theta - \theta_0)^2 + k_\psi \quad (\psi - \psi_0)^2 + \epsilon\left[\left(\frac{Rmin}{r}\right)^{12} - 2\left(\frac{Rmin}{r}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}} \qquad (2.2)$$

The first three terms of the right side of the CHARMm equation represent bond-stretching, angle-bending, and improper potential, respectively. The fourth and fifth terms represent Van der Waals (Lennard-Jones) and electrostatic potential, respectively.

From the rigid ion parameters listed in **Table 1**, it is evident that angle, improper, and electrostatic potentials are similar to the CHARMm force field to some extent. On the other hand, Morse potential and Buckinghum potentials are quite different from CHARMm compatible bond stretching potential and Lennard-Jones potential, respectively. In this study, CHARMm force field parameters of calcite mineral have been derived from the rigid ion model parameters. The following sections describe the procedure used for evaluating the CHARMm force field parameters.

Table 2.1. Force field parameters of calcite based on rigid ion model [43]

Potential	Equation	Atom Type	Parameters
Bond stretching (Morse)	$D_0\{e^{-\alpha(r-r_0)}-1\}^2$	C-O	D_0= 6140.92 kJ/mol α=2.5228 Å^{-1} r_0=1.294 Å
Angle bending	$0.5k_b(\theta-\theta_0)^2$	O-C-O	k_b = 173.63 kJ/mol.rad^2 θ_0= 120°
Improper	$k_t(1+\cos\varphi)$	O-C-O-O	k_t = 513.8 kJ/mol φ =180°
Short-range potential (Buckinghum)	$De^{-\beta r}-Cr^{-6}$	Ca-O	D =852875.84 kJ/mol β =4.19939 Å^{-1} C =0 (kJ/mol) Å^6
		O-O	D =3474566.97 kJ/mol β =5.06175 Å^{-1} C =2107.62 (kJ/mol) Å^6
Electrostatic	$\frac{e^2Z_iZ_j}{4\pi\varepsilon_0 r_{ij}}$	Ca	Z_{Ca}= +2.0000
		C	Z_C= +1.3435
		O	Z_O= -1.1450

2.2.2.1. Bond stretching potential

The bond stretching potential of the C-O covalent bond has been obtained from Morse potential. In Morse potential D_0 describes the bond dissociation energy, α is a constant depending upon vibrational force and r_0 is the equilibrium C-O bond length. A relationship exists between the bond stretching frequency (ω) and Morse potential constants (α, D_0) by the equation is given below [45]

$$\alpha = \omega\sqrt{\frac{\mu}{2D_0}} \tag{2.3}$$

Here μ is the reduced or effective mass of the bonded atoms. If m_i and m_j are the masses of bonded atoms,

$$\mu = \frac{m_i m_j}{m_i + m_j} \tag{2.4}$$

The relation between bond stretching frequency (ω) and harmonic spring constant(k) is given by the relation below

$$\omega = \sqrt{\frac{k}{\mu}} \tag{2.5}$$

From the above equations, we get

$$k = 2\alpha^2 D_0 \tag{2.6}$$

As the CHARMm bond stretching constant (k_r) is half of the harmonic spring constant(k), we found $k_r = \frac{1}{2}k = \alpha^2 D_0$. The values of α and D_0 specified in **Table 1** gave the value of k_r as 39084 $kJ/mol\text{Å}^2$. We compared the values of bond stretching potentials using Morse and CHARMm and found that their behavior almost matched up to 1.5% of stretching **(Figure 4)**.

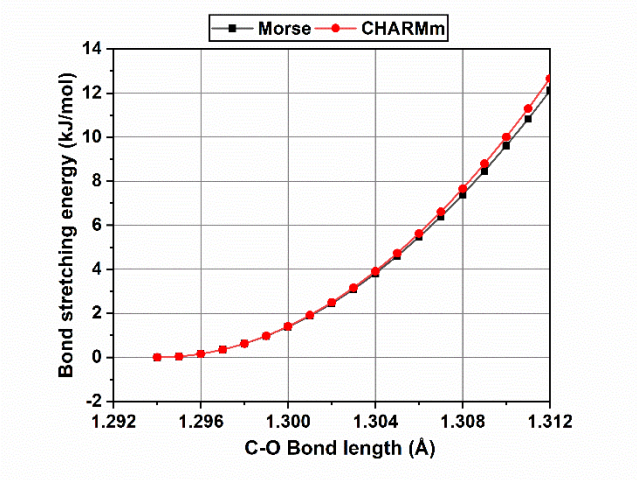

Figure 2.4**.** Comparison of Morse vs. derived CHARMm parameters for C-O bond stretching energy.

2.2.2.2. Angle bending potential

CHARMm angle bending constant (k_θ) of O-C-O has been calculated as half of k_b which equals 86.815 $kJ/molrad^2$. The equilibrium angle (θ_0) of O-C-O is 120°.

2.2.2.3. Improper/Inversion potential

The bond inversion energy arises from the out of a plane distortion of carbonate (CO_3^{2-}) anion. This potential is considered when three atoms (O) are covalently bonded to one central (C) atom. The improper energy constant (k_ψ) has been computed as 256.9 $kJ/molrad^2$ which is the half of four-body constant k_t. As this potential is used to maintain planarity of the carbonate ion, the value of ψ_0 is chosen as 0°.

2.2.2.4. Non-bonded potentials

Non-bonded short-range Van der Waals (VDW) potentials are considered for Ca-O pair and O-O pair. Lennard-Jones (LJ) potential is utilized to describe VDW energy in the CHARMm force field.

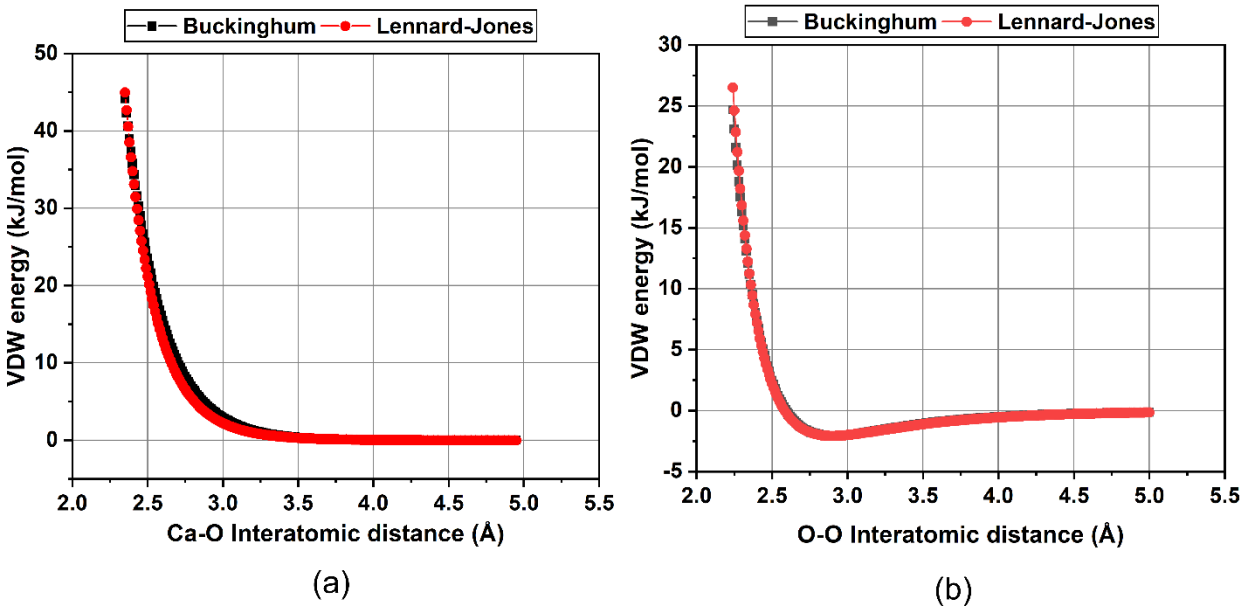

Figure 2.5. Comparison of Buckinghum vs. derived Lennard-Jones VDW potentials for (a) Ca-O pair, and (b) O-O pair.

Table 2.2. CHARMm force field parameters of calcite derived from rigid ion model parameters

Potential	Equation	Atom Type	Parameters
Bond stretching	$k_r(r-r_0)^2$	C-O	$k_r =$ 39084 kJ/molÅ^2 $r_0 = 1.294$ Å
Angle bending	$k_\theta(\theta-\theta_0)^2$	O-C-O	$k_\theta =$ 86.815 kJ/mol.rad^2 θ_0= 120°
Improper	$k_\psi \quad (\psi-\psi_0)^2$	O-C-O-O	$k_\psi = 256.9$ kJ/mol.rad^2 ψ_0 =0°
Van der Waals	$\epsilon\left[\left(\frac{Rmin}{r}\right)^{12} - 2\left(\frac{Rmin}{r}\right)^{6}\right]$	Ca-O	$\epsilon = 6.088 \times 10^{-3}$ kJ/mol $Rmin = 4.946$ Å
		O-O	$\epsilon = -2.078$ kJ/mol $Rmin = 2.90$ Å
Electrostatic	$\frac{q_i q_j}{r_{ij}}$	Ca	q_{Ca}= +2.0000
		C O	q_C= +1.3435 q_O =-1.1450

The LJ parameters ($\epsilon, Rmin$) were derived from Buckinghum potential parameters (D, β, C) by curve fitting method where ϵ was the minimum energy and $Rmin$ was the distance where it occurred **(Figure 5a, 5b)**. Partial electrostatic charges were kept the same for this study. **Table 2** is the tabular representation of all the derived CHARMm force field parameters of calcite. CHARMm force field parameters of 12-unit kerogen were taken from our earlier study [25].

2.2.3. Simulation Details

NAMD 2.12, a parallel molecular dynamics code, was used to simulate the kerogen-calcite model. NAMD was developed by the Theoretical and Computational Biophysics Group in the

Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign [46]. At the beginning of the simulation, the model was minimized at a vacuum condition (0 K and 0 bar) utilizing conjugate gradient method [47]. After performing the minimization, the temperature of the kerogen-calcite model was raised to room temperature (300 K) and pressure (1.0125 bar) condition. The increase to room temperature from 0 K was done in three steps (100 K/step). Next, the pressure was raised from 0 bar to 1.0125 bar in four steps (0.25 bar/ step). The model was simulated at room temperature and pressure for 3 ns during which calcite was harmonically constrained, but kerogen was unconstrained. The timestep used for simulation was 0.5 fs. Center for Computationally Assisted Science and Technology (CCAST), a supercomputing facility at North Dakota State University, was employed to perform all the simulations. Each of the simulations was carried out using one node and 20 Intel Xeon 2670v2 "Ivy Bridge" processors.

2.3. Results

To validate our derived CHARMm force field parameters for calcite mineral, we calculated the radial pair distribution function (RPDF) of our equilibrated calcite model at three ns and measured the lattice parameters of the unit cell.

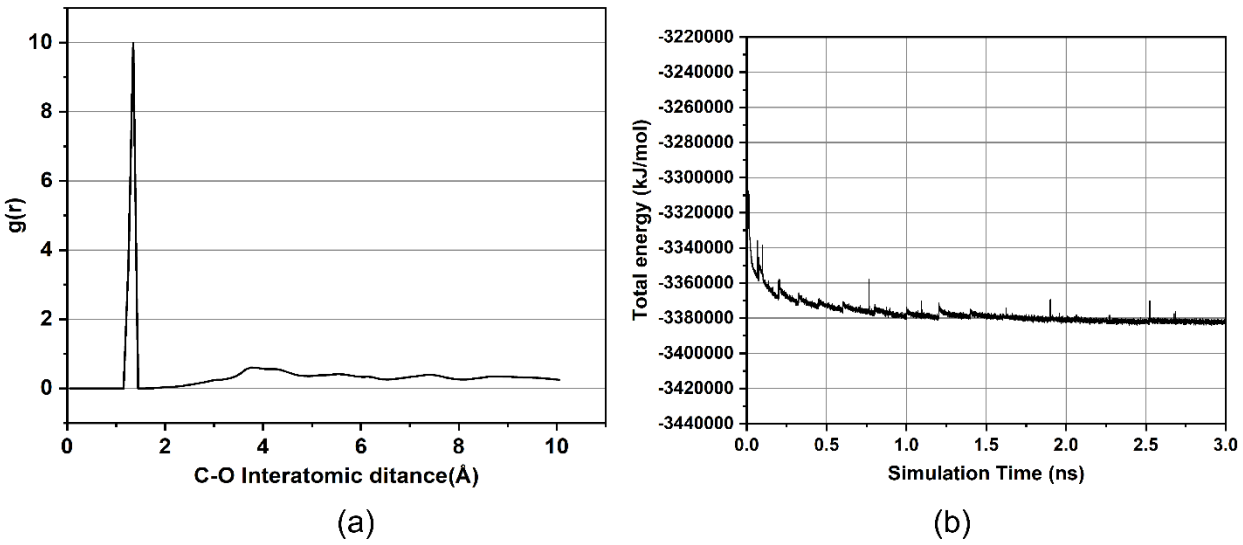

Figure 2.6. (a) Radial Pair Distribution Function (RPDF) plot of calcite model at 3ns, (b) Total energy vs. time plot for kerogen-calcite MD simulation up to 3ns.

The RPDF of the C-O pair of carbonate anion exhibited its peak at 1.30 Å **(Figure 6a),** while unit cell parameters of calcite remained the same. The graph showing total-energy vs. time **(Figure 6b)** indicates that the kerogen-calcite model reached the equilibrium condition within three ns of simulation period at NTP (300K and 1 bar) condition. At the beginning of the simulation, the total-energy significantly dropped within 0.5 ns. Then it decreased very slowly until 2.25 ns and then entered the equilibrium region.

When the simulation between 12-unit kerogen and calcite mineral started at room temperature and pressure, all the hydrocarbon fragments from the top chain (layer) of kerogen except fragment 3 moved towards calcite. Fragments 4 and 5 were pulled to the bottom (XY) and side surfaces (XZ, YZ) of the calcite model. Kerogen fragments 6 and 7 from the top chain (layer) were accumulated under the bottom surface (XY) of calcite mineral. Most of the ammonium ions (fragment 8) belonging to the top chain were slowly moving away from calcite. The bottom chain (layer) of kerogen started to move away from calcite mineral. After one ns of simulation, the kerogen bottom chain (layer) was completely segregated from the top chain (layer) and migrated

away from the carbonate mineral. All the previously attracted fragments from the top chain moved further closer to the calcite surface. Four segments of ‘fragment 3’ from the top chain moved to the bottom chain, while one segment of ‘fragment 4’ from the bottom chain moved towards the top chain. Nine NH_4 ions from the top chain (layer) remained in the proximity of calcite while other ions migrated into the bottom chain. Seven ammonium ions from the bottom chain moved away from the rest of the kerogen fragments. As the simulation was continued for two ns, the position of the interacting fragments from the top chain of kerogen remained almost the same. The non-interacting fragments from the top chain and the entire bottom chain were moved away from calcite. All the interacting and non-interacting fragments of kerogen maintained the same position from 2 ns to 3 ns **(Figure 7a,7b)**.

Table 2.3. Non-bonded interaction energies of kerogen fragments with calcite after three ns of simulation

Kerogen fragment	**Non-bonded energy (kJ/mol)**
1	78
2	49
3	0
4	-884
5	-15587
6	-15484
7	-26425
8	15

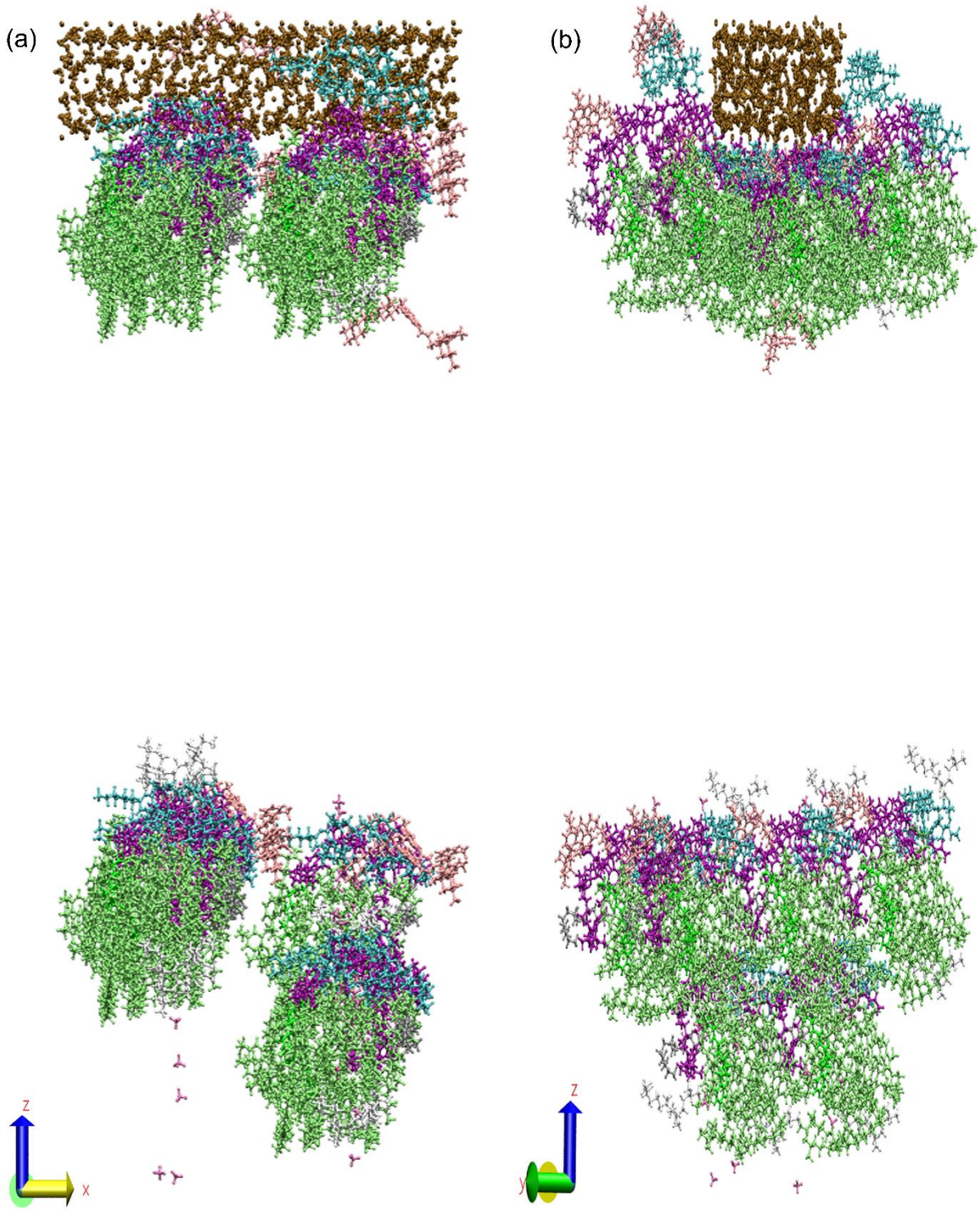

Figure 2.7. Conformation of the kerogen-calcite model at NTP condition after 3ns of simulation (a, b) from two different angles. Kerogen fragments 1, 2, 3, 4, 5, 6, 7, and 8 are represented by silver, green, white, pink, cyan, purple, lime, and mauve colored segments, respectively. Calcite is represented by the ochre-colored segment. All the representations were generated by VMD 1.9.3.

Table 3 indicates that all the fragments of kerogen except fragment 3 interact to some extent with calcite mineral. Fragments 1 and 2 have little interaction with calcite. Their positive interaction energies (78 kJ/mol for fragment 1 and 49 KJ/mol for fragment 2) represent their repulsive behavior towards calcite. Fragment 4 expresses quite attractive (-884 kJ/mol) non-bonded interaction with calcite mineral. Fragments 5, 6, and 7 of kerogen have dominant interactions with calcite mineral. Fragments 5 and 6 have a similar magnitude of attractive interaction (-15587 kJ/mol for fragment 5 and -15484 kJ/mol for fragment 6) with calcite. The interaction energy between calcite mineral and kerogen fragment 7 is the highest (-26425 kJ/mol) among all the organic-mineral interactions. The attractive interaction between fragment seven and calcite is almost 70% greater than the individual interactions of fragments 5 and 6 with calcite. All these interactions continue to be similar throughout the simulation period. Fragment 8 (NH_4 ions) has a smaller interaction with calcite.

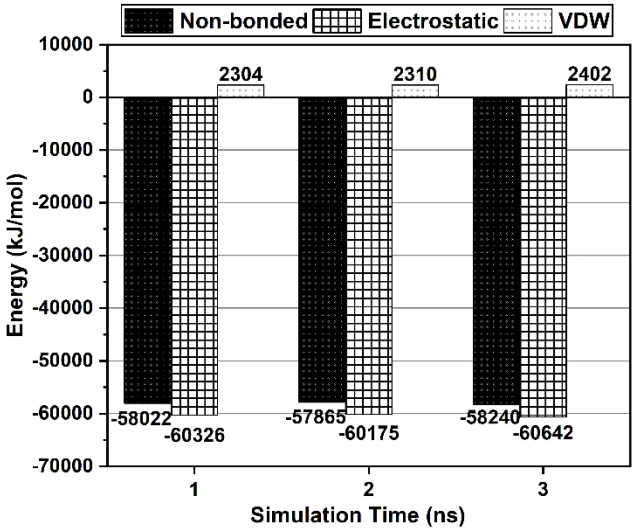

Figure 2.8. A Bar graph showing the total non-bonded energy between 12-unit kerogen and calcite with time.

The bar graph **(Figure 8)** denotes that the total non-bonded interaction between 12-unit kerogen and calcite mineral is mostly electrostatic. It shows that the non-bonded energy between

kerogen and calcite was -58022 kJ/mol at one ns of simulation. Then it decreased to -57865 kJ/mol at two ns and then again increased to -58240 kJ/mol at three ns. The electrostatic interaction is attractive, while the VDW interaction is repulsive. Both electrostatic and VDW energies maintain their respective magnitudes for the simulation period.

We also investigated the inter-fragment non-bonded interactions inside 12-unit kerogen without and with the presence of calcite mineral at three ns of simulation. An energy table is used to represent all the inter-fragment non-bonded interaction energies inside 12-unit kerogen **(Figure 9a,9b).** The inclusion of calcite mineral reduces the attractive interactions between kerogen fragments 1 and 2 (from -76 kJ/mol to -3 kJ/mol), 1 and 3 (from -58 kJ/mol to -10 kJ/mol), 1 and 4 (from -239 kJ/mol to -16 kJ/mol), 1 and 5 (from -115 kJ/mol to -45 kJ/mol) and 1 and 7 (from -825 kJ/mol to -375 kJ/mol). However, the attractive energy increased between fragments 1 and 6 (from -414 kJ/mol to -700 kJ/mol) due to the presence of calcite mineral.

7	-57129						
6	-19795	3870					
5	-13222	2129	364				
4	118	-962	-1073	-665			
3	29	-1203	-82	-289	-86		
2	43	-1458	-84	-96	-85	-38	
1	53	-825	-414	-115	-239	-58	-76
Fragment Number	8	7	6	5	4	3	2

(a)

7	-42888						
6	-16242	5235					
5	-11327	4093	2397				
4	112	-278	-566	-163			
3	52	-529	-32	-134	-12		
2	78	-1163	-34	-15	-12	-5	
1	64	-375	-700	-45	-16	-10	-3
Fragment Number	8	7	6	5	4	3	2

(b)

Figure 2.9. Non-bonded energy (kJ/mol) table of kerogen fragments at 3ns (a) without, and (b) with the presence of calcite. It provides the total non-bonded energy acting between any two fragments (total of eight fragments, including ammonium ions) inside kerogen. For finding the non-bonded energy between the ith and the jth fragment of kerogen (i<j), select i from the vertical column and select j from the horizontal row. Their intersection cell will give the non-bonded energy acting between them.

The calcite mineral also lessened the attractive interactions of fragment 2 with other fragments. The interaction energies declined from -38 kJ/mol to -5 kJ/mol (fragments 2 and 3), -85 kJ/mol to -12 kJ/mol (fragments 2 and 4), -96 kJ/mol to -15 kJ/mol (fragments 2 and 5), -84 kJ/mol to -34 kJ/mol (fragments 2 and 6) and -1458 kJ/mol to -1163 kJ/mol (fragments 2 and 7). The interaction energies of fragment 3 reduced from -86 kJ/mol to -12 kJ/mol (with fragment 4), -289 kJ/mol to -134 kJ/mol (with fragment 5), -82 kJ/mol to -32 kJ/mol (with fragment 6) and -1203 kJ/mol to -529 kJ/mol (with fragment 7) due to addition of calcite mineral. Non-bonded energies subsided from -665 kJ/mol to -163 kJ/mol between fragments 4 and 5, from -1073 kJ/mol to -566 kJ/mol between 4 and 6 and from -962 kJ/mol to -278 kJ/mol between 4 and 7.

Significant changes were seen to take place in interaction energies among certain fragments (5, 6, 7, and 8) due to the presence of calcite mineral. The repulsive interactions of fragment 8 with

fragments 1, 2, and 3 increase from 53 kJ/mol to 64 kJ/mol, 43 kJ/mol to 78 kJ/mol, and 29 kJ/mol to 52 kJ/mol respectively. On the contrary, attractive interactions of fragment 8 with fragments 5, 6, and 7 decreases from -13222 kJ/mol to -11327 kJ/mol, -19795 kJ/mol to -16242 kJ/mol and -57129 kJ/mol to -42888 kJ/mol respectively. The repulsive interactions increased from 364 kJ/mol to 2397 kJ/mol and 2129 kJ/mol to 4093 kJ/mol between fragments 5 and 6 and fragments 5 and 7 respectively. The repulsive energy between fragments 6 and 7 changed from 3870 kJ/mol to 5235 kJ/mol.

The energy between the individual chain (layer) of kerogen and calcite mineral was also calculated. We found that the bottom chain (layer) of kerogen did not interact at all with calcite throughout the simulation. The energy between the top chain (layer) of kerogen and calcite mineral was found to be -58240 kJ/mol at three ns, the same as total non-bonded interaction between calcite mineral and 12-unit kerogen at the same time.

2.4. Discussion

The previously described results from the molecular dynamics simulations of 12-unit kerogen and calcite mineral highlight interesting phenomena. The total non-bonded energy between kerogen and calcite is predominantly electrostatic and attractive. Only the top chain (layer) of kerogen interacts with calcite mineral at NTP condition. All the fragments from the top chain (including NH_4 ions) except fragment 3 interact with the calcite mineral. Fragments 1 and 2 are repulsed by a small magnitude, while fragments 4, 5, 6, and 7 are attracted by calcite. The repulsed fragments (1 and 2) and non-interacting fragment (3) do not contain any olefinic hydrocarbons. Conversely, all the attracted fragments (4, 5, 6, and 7) comprise olefinic hydrocarbons. Among the attracted fragments, the highest interaction takes place between fragment seven and calcite, as fragment 7 is the largest hydrocarbon molecule containing a higher

number of olefins. Most NH_4 ions from the top chain (layer) of kerogen move away from calcite as only 9 of them stay in proximity of mineral. Four segments of 'fragment 3' switch their position from top chain to bottom chain. From the beginning of the simulation, the bottom chain (layer) of kerogen significantly moves away from the calcite mineral. A segment of 'fragment 4' leaves the bottom chain and migrates towards the top chain. About 25% of all ammonium ions belonging to the bottom chain move independently away from its hydrocarbon.

The energy table of kerogen is a complete representation of inter-fragment non-bonded interaction behavior inside 12-unit kerogen. The energy table of equilibrated 12-unit kerogen (without calcite) shows that fragments 1, 2, 3, and 4 have attractive non-bonded interactions with other fragments except fragment 8 (independent ammonium ions). However, fragments 5, 6, and 7 interact with each other in a repulsive manner while each of them interacts with fragment eight attractively. Fragments 5, 6, 7, and 8 participate in the most dominating interactions inside kerogen. The energy table of the equilibrated kerogen-calcite model implies that inter-fragment interaction behavior of 12-unit kerogen is greatly influenced by calcite mineral. The inclusion of calcite changes the magnitude of interaction rather than altering the pattern of interaction.

The repulsive interactions of fragment 8 with fragments 1, 2, and 3 increase while attractive interactions of fragment 8 with fragments 5, 6, and 7 decreases. The migration of seven ammonium ions, leaving the kerogen away from calcite mineral, indicates the resultant decrement of the attractive interaction of fragment eight towards the rest of the kerogen. The repulsive interactions of fragment 5 with fragments 6 and 7 increase significantly. The repulsive energy acting between fragments 6 and 7 also increases.

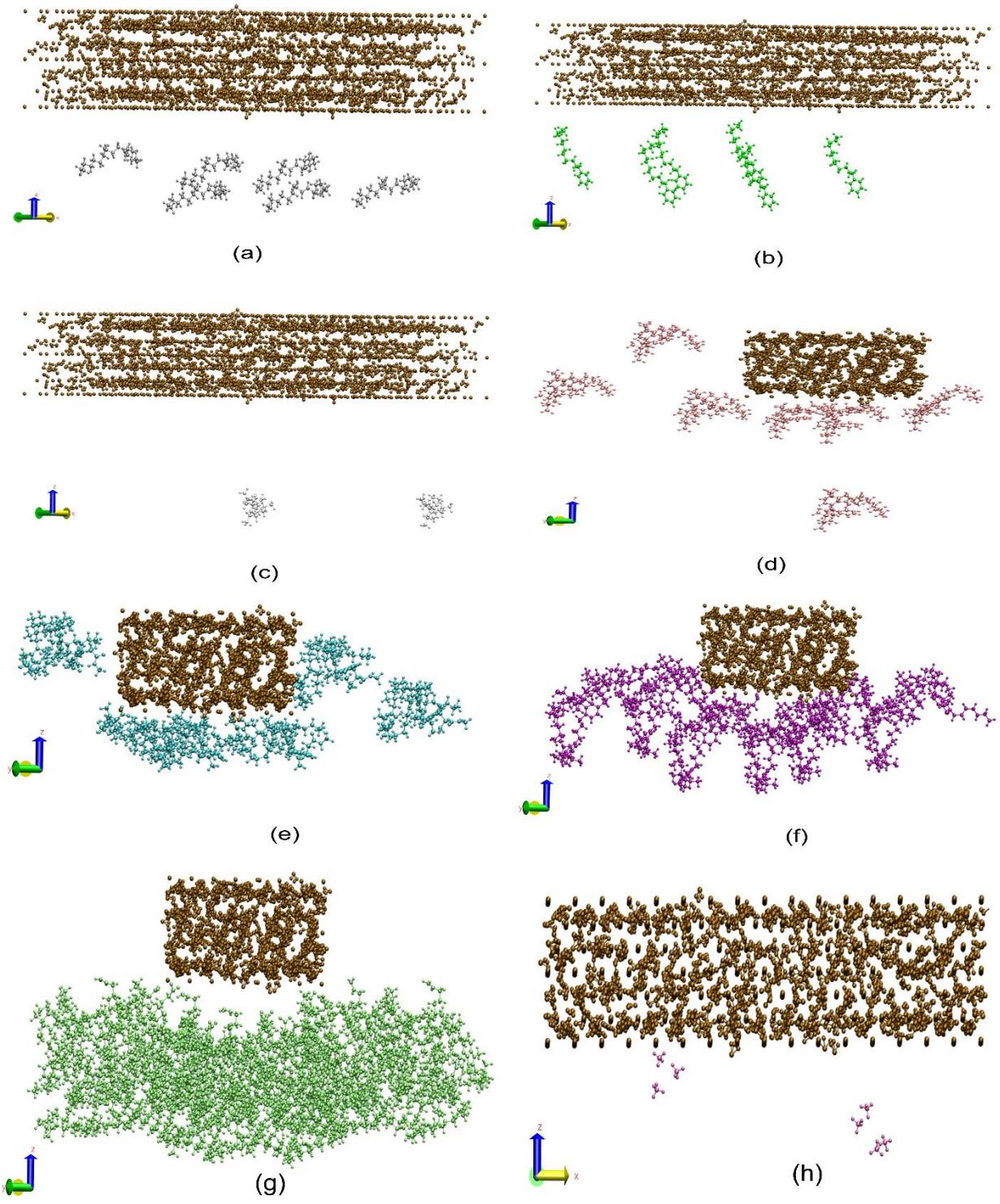

Figure 2.10. Conformation of kerogen (a) fragment 1, (b) fragment 2, (c) fragment 3, (d) fragment 4, (e) fragment 5, (f) fragment 6, (g) fragment 7, and (h) fragment 8 (NH_4 ions) in proximity of calcite mineral after 3 ns of simulation at NTP condition. All the representations were generated by VMD 1.9.3.

The final conformation of kerogen fragments from the top chain (layer) around the calcite mineral is also observed **(Figure 10)**. Fragments 1 and 2 do not change their positions much with respect to their initial conformation **(Figure 10a, 10b)**. Only two segments of fragment 3 remain in proximity of calcite as others move to the bottom chain **(Figure 10c)**. Four segments of fragment 4 attach to the bottom surface (XY) of calcite, while the other two segments stay parallel to the side (XZ, YZ) surfaces of mineral. A segment of fragment 4 from the bottom chain moves to the proximity of mineral **(Figure 10d)**. Three segments of fragment 5 adhere to the bottom calcite surface, while a single segment of fragment 5 adheres to the side mineral surface. The remaining two segments of fragment 5 reside parallel to calcite **(Figure 10e)**. Six segments of fragment 6 move close to the bottom (XY) surface of calcite, while four of them stick to it **(Figure 10f)**. All the segments of fragment 7 are gathered close to calcite, while two of them attach to the calcite bottom face **(Figure 10g)**. The independent NH_4 ions (fragment 8) predominantly move away from calcite, while only nine of them stay in its proximity **(Figure 10h)**. We have also performed another set of simulations where the calcite mineral is positioned at the bottom of kerogen to investigate the variation in kerogen-calcite interactions due to the variation in their orientations. Details of these simulations are presented in the supplemental document. We observe that the magnitude of total interactions change while the relative magnitude of interactions attributed to each kerogen fragment remain almost the same. The pattern and order of non-bonded interactions (attractive/repulsive) between individual kerogen fragments and calcite mineral remain unaltered.

The higher magnitudes of non-bonded interactions between calcite mineral and kerogen fragments can be attributed to the interactions between Ca^{2+} ions of calcite and different functional groups (-COOH, -OH) of kerogen fragments. The interactions of alkaline earth metals of calcite (Ca^{2+}) with different oxygen-containing functional groups of kerogen can lead to the formation of

Ca^{2+}-O complexes or clusters. These clusters are generally hypothesized as the active sites on the oil shale surface as they affect the reactivity of oil shales during oil shale pyrolysis [7, 48, 49].

2.5. Conclusions

In the present study, molecular dynamics simulations have been carried out to find the interactions between 12-unit kerogen and calcite mineral of Green River oil shale. We derived the potentials for the CHARMm force field for calcite from available potentials to parameterize the calcite model. We used a unit cell of calcite to build an expanded model comparable to the size of the 3D kerogen model. Then we merged kerogen and calcite structures and simulated them at room temperature and pressure to analyze their interactions. Some major findings are outlined below:

1. Calcite mineral causes the separation of kerogen chains (layers) where only the top chain (layer) of kerogen interacts with mineral while the bottom chain moves away from the mineral.
2. Apart from fragment 3 ($C_{20}H_{42}$), that had no interaction with calcite, all the other kerogen fragments have non-bonded interactions with calcite. Fragment 8 (independent ammonium ions) also has very little interaction with the mineral.
3. The olefinic hydrocarbons influence the interaction of kerogen fragments with calcite. Fragments that lack olefinic hydrocarbons (fragments 1 and 2) are repelled by calcite. Conversely, the fragments that contain olefinic hydrocarbons are attracted by the carbonate mineral.
4. Attractive electrostatic interactions are the major contributors to the non-bonded energy between calcite mineral and 12-unit kerogen.
5. The proximity of calcite mineral brings changes in the inter-fragment non-bonded interactions inside kerogen. Although the nature of the interactions

(attractive/repulsive) among the fragments are not changed, the magnitude of interactions are significantly modified (increased/decreased).

6. The proximity of calcite mineral significantly reduces the attractive interaction of fragment 8 (ammonium groups) with the rest of the kerogen resulting in the migration of ammonium ions away from the mineral.
7. The most significant kerogen-calcite interaction takes place between fragment 7 ($C_{367}H_{547}N_{10}O_{10}S_2$) of kerogen and calcite because this fragment contains the largest portion of olefinic hydrocarbons of kerogen.
8. This work provides an insight into the binding mechanisms between calcite and kerogen and the role of calcite mineral proximity on the intra-kerogen interactions.

2.6. Acknowledgments

The authors acknowledge the support from Mountain-Plains Consortium (MPC) Grant #: DTRT13-G-UTC38. The computational resources provided by North Dakota State University Computationally Assisted Science and Technology (CCAST) (NSF# 1229316) is much appreciated.

2.7. References

(1) Koebrich, S.; Chen, E. I.; Bowen, T.; Forrester, S.; Tian, T. *2017 Renewable Energy Data Book: Including Data and Trends for Energy Storage and Electric Vehicles*; National Renewable Energy Lab.(NREL), Golden, CO (United States): 2019.

(2) Zborowski, M.; Whitfield, S., E&P Notes (January 2019). *Journal of Petroleum Technology* **2019,** *71* (01), 18-24.

(3) Deng, S.; Wang, Z.; Gu, Q.; Meng, F.; Li, J.; Wang, H., Extracting hydrocarbons from Huadian oil shale by sub-critical water. *Fuel Processing Technology* **2011,** *92* (5), 1062-1067.

(4) Alstadt, K. N.; Katti, D. R.; Katti, K. S., An in situ FTIR step-scan photoacoustic investigation of kerogen and minerals in oil shale. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2012,** *89*, 105-113.

(5) Guo, H.; Lin, J.; Yang, Y.; Liu, Y., Effect of minerals on the self-heating retorting of oil shale: Self-heating effect and shale-oil production. *Fuel* **2014,** *118*, 186-193.

(6) Taskforce on Strategic Unconventional Fuels. 2007. America's Strategic Unconventional Fuels. Washington (DC): US Government.

(7) Karabakan, A.; Yurum, Y., Effect of the mineral matrix in the reactions of oil shales: 1. Pyrolysis reactions of Turkish Goynuk and US Green River oil shales. *Fuel* **1998,** *77* (12), 1303-1309.

(8) Yan, J.; Jiang, X.; Han, X.; Liu, J., A TG-FTIR investigation to the catalytic effect of mineral matrix in oil shale on the pyrolysis and combustion of kerogen. *Fuel* **2013,** *104*, 307-317.

(9) Alstadt, K. N.; Katti, K. S.; Katti, D. R., Nanoscale Morphology of Kerogen and In Situ Nanomechanical Properties of Green River Oil Shale. *Journal of Nanomechanics and Micromechanics* **2016,** *6* (1).

(10) Cane, R. F., In Oil Shale·, Yen, T. F., Chilingarian, GV, Eds. Elsevier: Amsterdam: 1976.

(11) Aboulkas, A.; El Harfi, K., Study of the kinetics and mechanisms of thermal decomposition of moroccan tarfaya oil shale and its kerogen. *Oil Shale* **2008,** *25* (4), 426-443.

(12) Aboulkas, A.; El Harfi, K., Effects of acid treatments on Moroccan Tarfaya oil shale and pyrolysis of oil shale and their kerogen. *Journal of Fuel Chemistry and Technology* **2009,** *37* (6), 659-667.

(13) Sert, M.; Ballice, L.; Yuksel, M.; Saglam, M., Effect of mineral matter on product yield and composition at isothermal pyrolysis of turkish oil shales. *Oil Shale* **2009,** *26* (4), 463-474.

(14) Ballice, L., Effect of demineralization on yield and composition of the volatile products evolved from temperature-programmed pyrolysis of Beypazari (Turkey) Oil Shale. *Fuel Processing Technology* **2005,** *86* (6), 673-690.

(15) Pan, L.; Dai, F.; Huang, P.; Liu, S.; Li, G., Study of the effect of mineral matters on the thermal decomposition of Jimsar oil shale using TG-MS. *Thermochimica Acta* **2016,** *627*, 31-38.

(16) Pradhan, S. M.; Katti, K. S.; Katti, D. R., Multiscale Model of Collagen Fibril in Bone: Elastic Response. *Journal of Engineering Mechanics* **2014,** *140* (3), 454-461.

(17) Katti, D. R.; Pradhan, S. M.; Katti, K. S., Directional dependence of hydroxyapatite-collagen interactions on mechanics of collagen. *Journal of Biomechanics* **2010,** *43* (9), 1723-1730.

(18) Ghosh, P.; Katti, D. R.; Katti, K. S., Mineral proximity influences mechanical response of proteins in biological mineral-protein hybrid systems. *Biomacromolecules* **2007,** *8* (3), 851-856.

(19) Bhowmik, R.; Katti, K. S.; Katti, D. R., Molecular interactions of degradable and non-degradable polymers with hydroxyapatite influence mechanics of polymer-hydroxyapatite

nanocomposite biomaterials. *International Journal of Nanotechnology* **2009,** *6* (5-6), 511-529.

(20) Bhowmik, R.; Katti, K. S.; Katti, D. R., Influence of mineral on the load deformation behavior of polymer in hydroxyapatite-polyacrylic acid nanocomposite biomaterials: A steered molecular dynamics study. *Journal of Nanoscience and Nanotechnology* **2008,** *8* (4), 2075-2084.

(21) Sikdar, D.; Katti, K. S.; Katti, D. R., Molecular interactions alter clay and polymer structure in polymer clay nanocomposites. *Journal of Nanoscience and Nanotechnology* **2008,** *8* (4), 1638-1657.

(22) Sikdar, D.; Pradhan, S. M.; Katti, D. R.; Katti, K. S.; Mohanty, B., Altered phase model for polymer clay nanocomposites. *Langmuir* **2008,** *24* (10), 5599-5607.

(23) Katti, K. S.; Sikdar, D.; Katti, D. R.; Ghosh, P.; Verma, D., Molecular interactions in intercalated organically modified clay and clay-polycaprolactam nanocomposites: Experiments and modeling. *Polymer* **2006,** *47* (1), 403-414.

(24) Katti, D. R.; Upadhyay, H. B.; Katti, K. S., Molecular interactions of kerogen moieties with Na-montmorillonite: An experimental and modeling study. *Fuel* **2014,** *130*, 34-45.

(25) Katti, D. R.; Thapa, K. B.; Katti, K. S., Modeling molecular interactions of sodium montmorillonite clay with 3D kerogen models. *Fuel* **2017,** *199*, 641-652.

(26) Katti, D. R.; Katti, K. S.; Thapa, K.; Faisal, N., Modeling the Nanoscale Kerogen Inclusions in Green River Oil Shale. In *Poromechanics VI*, pp 1968-1975.

(27) Titiloye, J. O.; Parker, S. C.; Mann, S., Atomistic simulation of calcite surfaces and the influence of growth additives on their morphology. *Journal of Crystal Growth* **1993,** *131* (3-4), 533-545.

(28) Walker, J. C. G., Biogeochemistry - an analysis of global change - schlesinger, WH. *Science* **1991,** *253* (5020), 686-687.

(29) Reeder, R. J.; Nugent, M.; Tait, C. D.; Morris, D. E.; Heald, S. M.; Beck, K. M.; Hess, W. P.; Lanzirotti, A., Coprecipitation of uranium(VI) with calcite: XAFS, micro-XAS, and luminescence characterization. *Geochimica Et Cosmochimica Acta* **2001,** *65* (20), 3491-3503.

(30) Stumm, W.; Morgan, J. J., *Aquatic chemistry: chemical equilibria and rates in natural waters*. John Wiley & Sons: 2012; Vol. 126.

(31) Young, D., Computational chemistry: a practical guide for applying techniques to real world problems. John Wiley & Sons: 2004.

(32) Poudel, L.; Tamerler, C.; Misra, A.; Ching, W. Y., Atomic-Scale Quantification of Interfacial Binding between Peptides and Inorganic Crystals: The Case of Calcium Carbonate Binding Peptide on Aragonite. *Journal of Physical Chemistry C* **2017,** *121* (51), 28354-28363.

(33) Klein, C.; Dutrow, B.; Dana, J. D., The 23rd edition of the manual of mineral science:(after James D. Dana). 2007.

(34) Sass, R. L.; Vidale, R.; Donohue, J., Interatomic distances and thermal anisotropy in sodium nitrate and calcite. *Acta Crystallographica* **1957,** *10* (9), 567-570.

(35) Wyckoff, R. W. G., The crystal structures of some carbonates of the calcite group. *American Journal of Science* **1920,** *50* (299), 317-360.

(36) Orendt, A. M.; Pimienta, I. S. O.; Badu, S. R.; Solum, M. S.; Pugmire, R. J.; Facelli, J. C.; Locke, D. R.; Chapman, K. W.; Chupas, P. J.; Winans, R. E., Three-Dimensional

Structure of the Siskin Green River Oil Shale Kerogen Model: A Comparison between Calculated and Observed Properties. *Energy & Fuels* **2013,** *27* (2), 702-710.

(37) Siskin, M.; Scouten, C. G.; Rose, K. D.; Aczel, T.; Colgrove, S. G.; Pabst, R. E., Detailed structural characterization of the organic material in Rundle Ramsay Crossing and Green River oil shales. In *Composition, geochemistry and conversion of oil shales*, Springer: 1995; pp 143-158.

(38) Sharma, A.; Snead, M. L.; Katti, K. S.; Katti, D. R., Mechanics of amelogenin TRAP protein in the proximity of hydroxyapatite mineral is altered by interfacial water. *Chemical Physics* **2019,** *522*, 104-111.

(39) Fisler, D. K.; Gale, J. D.; Cygan, R. T., A shell model for the simulation of rhombohedral carbonate minerals and their point defects. *American Mineralogist* **2000,** *85* (1), 217-224.

(40) Archer, T. D.; Birse, S. E. A.; Dove, M. T.; Redfern, S. A. T.; Gale, J. D.; Cygan, R. T., An interatomic potential model for carbonates allowing for polarization effects. *Physics and Chemistry of Minerals* **2003,** *30* (7), 416-424.

(41) Pavese, A.; Catti, M.; Price, G. D.; Jackson, R. A., Interatomic potentials for caco3 polymorphs (calcite and aragonite), fitted to elastic and vibrational data. *Physics and Chemistry of Minerals* **1992,** *19* (2), 80-87.

(42) Catti, M.; Pavese, A.; Price, G. D., Thermodynamic properties of caco3 calcite and aragonite - a quasi-harmonic calculation. *Physics and Chemistry of Minerals* **1993,** *19* (7), 472-479.

(43) Duckworth, O. W.; Cygan, R. T.; Martin, S. T., Linear free energy relationships between dissolution rates and molecular modeling energies of rhombohedral carbonates. *Langmuir* **2004,** *20* (7), 2938-2946.

(44) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010,** *31* (4), 671-690.

(45) Leach, A., Empirical force field models: molecular mechanics. *A. Leach, Molecular Modelling: principles and applications* **2001**, 165-252.

(46) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** *26* (16), 1781-1802.

(47) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D., Iterative minimization techniques for abinitio total-energy calculations - molecular-dynamics and conjugate gradients. *Reviews of Modern Physics* **1992,** *64* (4), 1045-1097.

(48) Joseph, J. T.; Forrai, T. R., Effect of Exchangeable Cations on Liquefaction of Low Rank Coals. *Fuel* **1992,** *71* (1), 75-80.

(49) Siskin, M.; Aczel, T., Pyrolysis studies on the structure of ethers and phenols in coal. *Fuel* **1983,** *62* (11), 1321-1326.

3. AN INSIGHT INTO QUARTZ MINERAL INTERACTIONS WITH KEROGEN IN GREEN RIVER OIL SHALE[2]

3.1. Introduction

Modern civilization heavily depends on the consumption of energy. The search for 'energy sources' has been one of the most major concerns of humanity for the last two centuries. Crude oil, a critical source of energy, has been in great demand globally. The limited availability of crude oil resources has led the researchers to find newer sources of petroleum. Oil shale is being considered as a potential alternative to traditional petroleum resources (crude oil, coal, gas) [1]. Oil shales have been discovered in 27 countries, and some of them have begun extracting shale oil [2]. Green River formation in Utah, Wyoming, and Colorado is the largest deposit of oil shale in the world [3]. This deposit is estimated to be a resource of 1.5 trillion barrels of shale oil, including a reserve of 800 billion barrels [4].

Oil shale is a dense, fine-grained, laminated sedimentary rock that contains both organic and inorganic materials, and the organic material is considered to be trapped in the inorganic matrix [5]. The organic material is mainly kerogen with small amounts of bitumen, while the inorganic components are minerals. The minerals present are typically carbonates, silicates, and pyrites [6]. Kerogen is the material of interest in oil shale as the thermal cracking of kerogen yields shale oil. Kerogen is a high molecular weight insoluble organic heteropolymer of carbon, oxygen, hydrogen, nitrogen, and sulfur. It has been classified into Type I, Type II, and Type III depending upon its hydrogen to carbon ratio and oxygen to carbon ratio [7]. Green River oil shale contains 11.04 %

[2] The contents of this chapter have been published in Faisal, HM Nasrullah, Kalpana S. Katti, and Dinesh R. Katti. "An insight into quartz mineral interactions with kerogen in Green River oil shale." *International Journal of Coal Geology* 238 (2021): 103729. As the lead author of this journal article, the copyright permission has been obtained from Elsevier.

Type I kerogen, 2.76% bitumen, and 86.2% mineral by weight [8]. Both carbonate and silicate minerals constitute almost half of the mineral matter, while pyrite exists in small amounts.

Early research proved that the presence of the mineral matrix with kerogen affects the qualitative and quantitative distribution of hydrocarbons in the pyrolysis process [9]. However, researchers have different opinions regarding kerogen-mineral interactions during pyrolysis. Some TG analyses revealed that all minerals inhibit the decomposition of kerogen during pyrolysis [10]. On the other hand, other TG analyses showed that montmorillonite clay minerals offer catalytic effect on the decomposition of oil shale [11]. Several experiments demonstrated that the removal of carbonate minerals decreased the hydrocarbon yield while the removal of silicate minerals increased the hydrocarbon yield [12]. Some researchers also hypothesized that the effect of mineral matrix on kerogen primarily depends upon the oil shale mineral composition [13]. If the carbonates make up the greater portion of minerals, they have a catalytic effect on oil shale pyrolysis. However, if the silicates are predominantly present in oil shale, they exhibit an inhibitory effect on kerogen pyrolysis. Nevertheless, as the Green River oil shale contains carbonates and silicates in equal percentage (by weight), their impact on pyrolysis is still unclear.

Our research group has carried out extensive investigations on kerogen mineral interactions in the Green River oil shale. We determined that kerogen exists in pockets of the order of 10s of nm in Green River oil shale, and kerogen-mineral interactions are primarily nonbonded [2, 14]. Our previous studies have shown that nanoscale coexistence of minerals and organics affects the behavior of organic macromolecules in seashells [15], polymer clay nanocomposites [16], human bone [17], and polymer hydroxyapatite nanocomposites [18] mainly because of nonbonded interactions. Solid rock oil shale can be converted into liquid shale oil by pyrolysis procedure [19] . Different physical and chemical methods are utilized to isolate the organic kerogen from the inorganic

mineral matters [20]. Critical point drying, gravity flotation, subcritical water extraction and inorganic acid leaching are the most commonly employed techniques [21]. The effective application of all these procedures is potentially hindered by the lack of detailed understanding of the interactions between the organic kerogen and inorganic minerals [22]. Moreover, the quality and yield of crude oil is also dependent upon the organic-inorganic interactions of oil shale [11, 13]. Therefore, it is essential to investigate the interactions between the organic kerogen and inorganic minerals to extract the trapped kerogen from the mineral matrix. Different experimental studies investigated the organic-inorganic interactions within the oil shale. Supercritical ethanolysis (SE) and stepwise acid treatment of oil shale revealed that carbonate minerals primarily interact with carboxylic acids while silicate minerals form Si-O ether bonds [22]. Liquid chromatography fractionation and Fourier transform infrared spectroscopy of oil shale also showed the similar results [23]. However, these experimental techniques can investigate the interactions for a category of minerals (calcites/silicates etc.) instead of determining the interactions of individual mineral (calcite, dolomite, clay, quartz etc.) with organic kerogen. In-silico investigation of kerogen-mineral interactions through molecular dynamics simulations can be an efficient alternative solution towards this problem. Our research group has been working to devise a computational testbed to model the interactions of organic kerogen with the mineral matrix. As the organic kerogen is trapped inside the mineral matrix in Green River oil shale, the ultimate goal is to model the kerogen in the proximity of all the present minerals through molecular dynamics simulations. Before modeling the kerogen with all minerals, it would be beneficial to model kerogen with each of the individual mineral to identify the interactions that take place between kerogen and only that mineral. Inspecting the interactions of kerogen with individual mineral will help us to identify the major interacting minerals within the oil shale. This virtual screening of interacting minerals of oil

shale will help to model kerogen-mineral matrix interactions in a more efficient manner by considering only the significantly interacting minerals. Different research groups have undertaken the computational approaches to investigate the kerogen-mineral interactions [24, 25]. The previous studies from our group determined the interactions of kerogen with montmorillonite and calcite mineral [26, 27]. For the development of robust testbeds for kerogen interactions with all major Green River Oil Shale minerals, in this study we investigate the molecular interactions of kerogen with quartz mineral using molecular dynamics simulations.

Molecular dynamics (MD) simulation predicts the time-dependent behavior of a molecular system. The energy of the system is computed using molecular mechanics calculations. At first, we performed MD simulations of three kerogen moieties (pyridine, quinaldine, and heptylamine) with Na-montmorillonite clay and verified the findings of MD simulations by performing XRD and FTIR experiments [28]. We then constructed a complete 3D kerogen model (12-unit kerogen) and simulated it with Na-montmorillonite clay, which revealed significant nonbonded interactions between clay and kerogen [26, 29]. Molecular interactions between organic kerogen and carbonate mineral (calcite) have been investigated in terms of energy and conformations [30].

Quartz (SiO_2) is the most commonly found silicate mineral on earth. It makes up about one-fifth of the earth's exposed crust [31]. Apart from Na-montmorillonite clay, quartz is another major silicate mineral of Green River oil shale. It makes up about 8.6% of the Green River oil shale [8] and can have significant interactions with kerogen. Quantum chemical calculations have been used to explore the interactions between kerogen (type II) and silica in gas shale [25]. A model of CS1000 (a disordered porous 3D carbon structure lacking oxygen) [32] was used as type II kerogen to reconstruct interface with α-cristobalite. However, Green River oil shale contains type I kerogen, unlike the gas shale. In this study, we present the results of molecular dynamics

simulations of quartz mineral with type I kerogen to understand their interactions in the Green River oil shale. This study primarily seeks to develop a computational test-bed for investigating the molecular interactions between the organic constituent (Type I kerogen) of Green River oil shale and inorganic quartz mineral in a laboratory environment i.e. Normal Temperature and Pressure conditions.

3.2. Methodology

3.2.1. Model Construction

Quartz is a trigonal crystal system. It belongs to the tectosilicate group that exhibits a complex 3-dimensional framework. The silica (SiO_4) tetrahedron is the principal building block of the 3D frame. These silica tetrahedrons build large framework structure by sharing corner oxygen atoms. Due to this behavior, the corner oxygen atoms are also termed as bridging oxygen atoms. The three-fold screw axis (3_2) is observed in the quartz framework. The quartz structure consists of holes or cavities, usually termed as stuffing sites. Different alkaline earth elements or groups can occupy these sites. In nature, quartz can occur in four different polymorphs (α, β, tridymite, and cristobalite) depending upon its surrounding environment. The α-quartz predominantly occurs at normal temperature and pressure while three other polymorphs usually occur at high temperature and pressure conditions. As we are mainly focused on determining the interactions between quartz mineral and kerogen at NTP condition (300K and 1 bar), we built the crystal structure of α-quartz. The space group of α-quartz is $P3_221$ (No. 154). The unit cell dimensions of alpha-quartz are a=b=4.9134 Å, c=5.4052 Å, and α=β=90º, ϒ=120º [33]. Using this information and fractional coordinates of atoms, the unit cell of α-quartz is constructed using Materials Studio 7.0 **(Fig. 1a, 1b)**. We then performed 'X-ray Powder Diffraction' on our constructed quartz unit cell model using the 'Reflex' module of Materials Studio 7.0. The resultant

diffraction peaks **(Fig. 2a)** are in exact accordance with the diffraction peaks of the quartz sample found in rocks [34]. The matching of the diffraction pattern is performed by MATCH! 3.7.0 Software. This exercise validates our constructed quartz unit cell model. For determining the interactions with the 12-unit 3D kerogen model, we have built a quartz supercell by repeating the quartz unit cell. The dimensions of the quartz supercell are 14.54 Å × 83.51 Å × 10.80 Å. This model contains 1197 atoms in total. In this supercell, the silica tetrahedrons are arranged in such a way to exhibit a three-fold screw axis **(Fig.1c, 1d**) and stuffing sites (**Fig. 2b**). The quartz supercell was also constructed by using Materials Studio 7.0.

The 3D kerogen model of Green River oil shale used in this study has been developed, homogenized through stepwise annealing and validated in our prior work [26]. The construction of the 3D kerogen model was based on Siskin's 2D kerogen model [35, 36]. Siskin generated the type I kerogen model using nondestructive chemical derivatization and characterization via NMR and mass spectroscopy. This structure was later validated against the NMR, XPS (X-ray photoemission spectroscopy), and sulfur XANES (X-ray absorption near edge structure) data [37]. The chemical formula of the kerogen unit is $C_{645}H_{1017}N_{19}O_{17}S_4$. For the convenience of structure building and parametrization the kerogen unit was considered to be made up (agglomerate) of seven different hydrocarbon fragments and four independent ammonium (NH_4) groups associated with the fragments. In the present study, these seven hydrocarbon fragments will be termed as fragment 1 through fragment 7. The independent ammonium ions are considered as fragment 8. The model building procedure was described in details in one of the earlier studies from our group [26]. At first, all the kerogen fragments were built using Materials Studio 7.0. Then all the fragments were merged into unit kerogen model using the VMD 1.9.3. This unit kerogen model was parameterized using the CHARMm General Force Field (CGenFF) which assigns the atom types and partial

charges in a rule-based automated manner. In order to mimic the tens of nanometer sized kerogen pockets, 12 units of kerogen were merged using the VMD. The geologic pressure condition was modeled by applying the in-situ stresses on the 12-unit kerogen model. A series of molecular dynamics simulations were performed on the kerogen model to obtain the equilibrated conformation that will be used for determining the interactions with minerals. At the beginning, minimization was performed on the model at 0 K and 0 bar conditions to remove the artifacts from the model. Then simulated annealing was performed by gradually increasing and decreasing the temperature of the 12-unit kerogen model in a stepwise process for achieving an optimized conformation. The temperature was increased from 0 K to 500 K and then decreased to 300 K while the pressure was increased from 0 bar to 1.01325 bar. At last, the model was equilibrated at NTP condition (300 K and 1.01325 bar) for 5 ns. Radial pair distribution function (RPDF) were calculated for the equilibrated 12-unit kerogen model to validate against the previously reported NMR data [36]. The different stages of model building and validation are shown in consecutive figures of the Supplementary material (Supplementary figures 4-9). Each kerogen unit contains four independent ammonium (NH_4) groups associated with the seven hydrocarbon fragments. In the present study, these seven hydrocarbon fragments will be termed as fragment 1 through fragment 7. Therefore the 12-unit kerogen model consists of 12 segments of each fragment. For the convenience of energy analysis, we will reference all the independent ammonium groups as fragment 8.

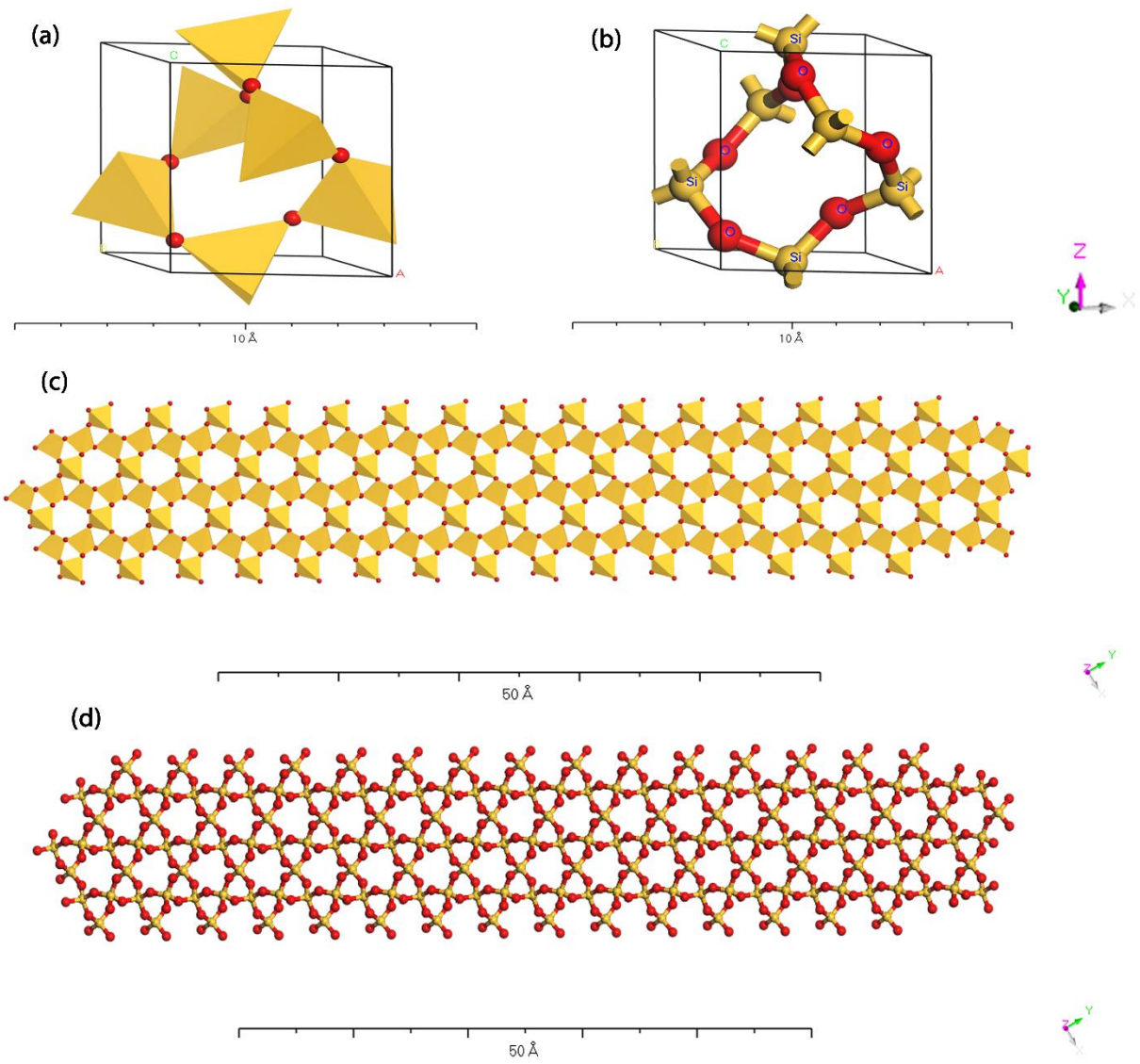

Figure 3.1. Molecular structure of a unit cell of α-quartz (a) polyhedron view, (b) ball and stick view with a label on atoms, and quartz supercell (c) polyhedron view, (d) ball and stick view. All the representations were generated by Materials Studio 7.0.

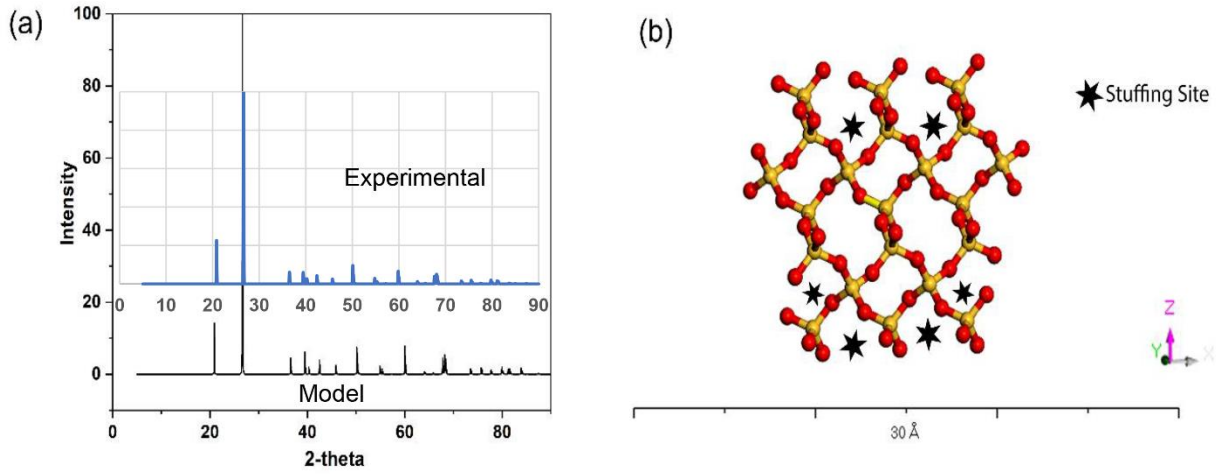

Figure 3.2. (a) The XRD spectrum of constructed α-quartz unit cell model with the experimental spectrum (reproduced from Ikuta et al., 2007) on it , (b) Stuffing sites (cavities) of quartz supercell.

Table 1 shows the chemical composition of kerogen fragments. The detailed description of kerogen fragments is provided in the previous study [26]. The 12 kerogen units are identified as top and bottom chains (layers) along the Z-axis. Thus, each chain has six kerogen units, i.e., 30 NH_4 groups in total. The dimensions of the 12-unit kerogen model are 71.92 Å × 88.98 Å × 63.89 Å **(Fig.3a, 3b)**.

Table 3.1. Chemical composition of kerogen fragments [26]

Kerogen fragment	**Chemical formula**
1	$C_{18}H_{38}$
2	$C_{18}H_{30}$
3	$C_{20}H_{42}$
4	$C_{45}H_{60}O$
5	$C_{75}H_{117}N_4O_5$
6	$C_{102}H_{167}NOS_2$
7	$C_{367}H_{547}N_{10}O_{10}S_2$
8	NH_4

The 12-unit kerogen and quartz supercell structures were merged using the "Merge Structures" plugin of Visual Molecular Dynamics (VMD) 1.9.3.

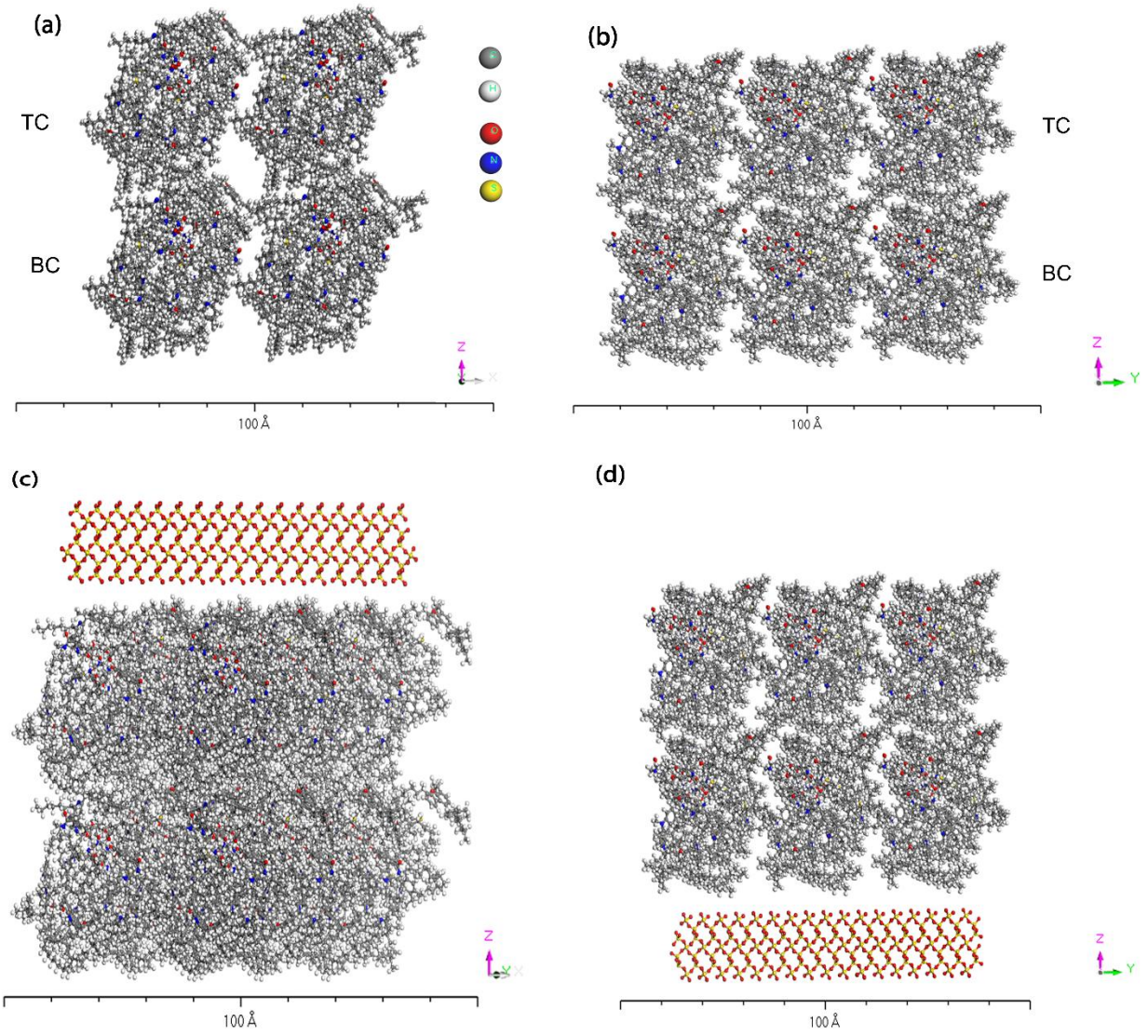

Figure 3.3. Molecular models of (a, b) three-dimensional 12-unit kerogen, (c) quartz supercell on top of kerogen, and (d) quartz supercell at the bottom of kerogen. TC and BC represent the top and bottom chain (layer) of kerogen along the Z-axis, respectively. All the representations were generated by Materials Studio 7.0.

The merging of structures was performed in two different ways to build two different models. In the first model, the quartz supercell is placed on top of the 12-unit kerogen **(Fig. 3c),** while in the second model, quartz supercell is placed at the bottom of kerogen along the Z-axis

(Fig. 3d). The two models allow for evaluating the interactions between the kerogen and quartz for two potential quartz orientations with respect to kerogen. We have not incorporated silanol (SiOH) groups in the model as the previous FTIR study of Green River oil shale did not show the presence of explicit water [2]. A cutoff distance of 16 Å has been employed for computing local non-bonded interactions. In order to be consistent with NAMD periodic boundary algorithms, the periodic boundary conditions have been made sufficiently large so that the quartz supercell can not see the organic kerogen from both sides. Therefore, all the resultant non-bonded interactions originated inside the periodic box rather than across the periodic boxes.

3.2.2. Model Parameterization

The quartz model is parameterized using the CHARMm (Chemistry at HARvard Macromolecular Mechanics) force field 38. It contains a set of functions and associated constants to define energy expression. CHARMm uses five energy terms, including one electrostatic term. The following equation describes the CHARMm force field:

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{VDW} + E_{electrostatic} \tag{3.1}$$

$$E = k_r(r - r_0)^2 + k_\theta(\theta - \theta_0)^2 + k_\varphi \quad [1 + \cos(n\varphi + \delta)] + \epsilon\left[\left(\frac{Rmin}{r}\right)^{12} - 2\left(\frac{Rmin}{r}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}} \tag{3.2}$$

The first three terms of the right side of equations represent bonded energy, while the last two terms represent nonbonded energy. Here $k_r,\ k_\theta\ and\ k\varphi$ express bond stretching, angle bending, and dihedral force constant, respectively. The equilibrium bond length, angle, dihedral and phase shift are symbolized by $r_0, \theta_0, \varphi\ \ and\ \delta$ accordingly. Lennard-Jones VDW parameters are represented by ϵ and $Rmin$. The partial electrostatic atomic charge is represented by q, and the interatomic distance is expressed by r. The CHARMm force field parameters of α-quartz have

been obtained from the literature where the authors proposed a common force field for all silica minerals (α-quartz, α-cristobalite, and amorphous silica) [39]. They determined the structural, mechanical, and vibrational properties of these minerals utilizing the force field and cross-validated them with experimental results. The CHARMm force field parameters of 3D kerogen that were found in our previous study [26] have been utilized.

3.2.3. Simulation Details

Molecular dynamics simulations of both models (quartz supercell on top of kerogen and quartz supercell at the bottom of kerogen along Z-axis) were performed using NAMD 2.12 40. NAMD was developed by the Theoretical and Computational Biophysics Group at the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. Both models were minimized at zero Kelvin temperature (0 K) and 0 bar pressure using the conjugate gradient method 41. Next, both the models were brought to room temperature (300 K) and pressure (1.0125 bar) in seven steps. First, the temperature was raised from 0 K to 300 K in three steps, and then the pressure was increased from 0 bar to 1.0125 bar in four steps. Both models were run at room temperature and pressure (isobaric-isothermal) for 4 ns with a time step of 0.5 fs (8,000,000 steps) as the models reached equilibrium condition within 4 ns. In both cases, the quartz supercell was constrained harmonically to maintain the crystallinity while kerogen was unconstrained. Harmonic constraints compel the atoms to remain in a reference position by applying harmonic force constants. Any movement from the reference position is penalized which restricts the large atomic motion of the constrained atoms. All the simulations were conducted at North Dakota State University Center for Computationally Assisted Science and Technology (CCAST). Each simulation utilized one node, dual socket Intel Xeon 2670v2 "Ivy Bridge" 2.5GHz with 64GB DDR3 RAM at 1866MHz and 20 processors.

3.3. Results and Discussion

3.3.1. Quartz Supercell on Top of Kerogen Model

The total energy of MD simulation is the summation of potential (bonded & nonbonded) and kinetic energy. **Fig. 4a** illustrates that the total interaction energy between the quartz supercell and 12-unit kerogen has reached the equilibrium zone in less than 4 ns of the simulation period. Snapshots of the model were taken at 1 ns intervals to represent the conformations of quartz supercell on top of 12-unit kerogen (**Fig. 5).**

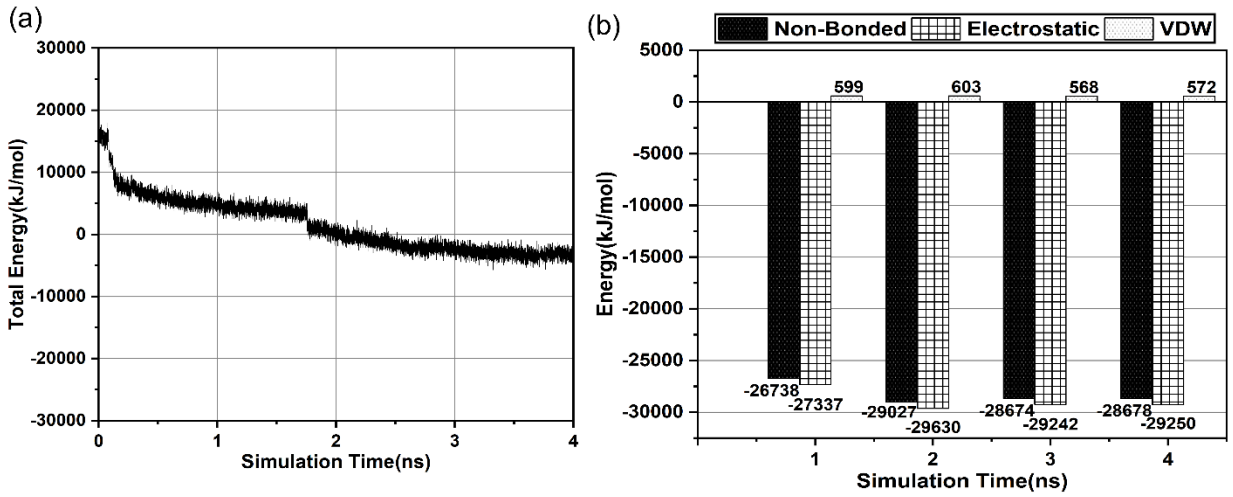

Figure 3.4. (a) Total energy vs. time plot, and (b) bar graph showing nonbonded energy acting between quartz supercell and 12-unit kerogen with time for 'quartz supercell on top of kerogen' model.

Immediately after the simulations were initiated at room temperature and pressure, 13 NH_4 groups among 30 from the top chain (layer) of kerogen moved to the quartz supercell. These NH_4 groups occupied the stuffing sites (holes) of quartz structure **(Fig. 6a)**. All the fragments of kerogen except fragments 4 and 5 slowly moved away from the quartz supercell. At the end of 1 ns, four segments of fragment 4 and three segments of fragment 5 were in close proximity of the supercell **(Fig. 5a)**. All of the interacting fragments belonged to the top chain (layer) of kerogen. The sudden decrease in total energy around 1.7 ns (**Fig. 4a**) indicated the separation of bottom kerogen layer

from the top kerogen layer. After the simulation period of 2 ns, a total of 14 NH_4 groups from the top chain moved to the supercell while other NH_4 groups moved towards the bottom chain of kerogen. Four segments of fragment 4 were still in the close proximity of the quartz. But instead of three, only one segment of fragment 5 had interaction with supercell, which implies that most of the fragments of kerogen were further moving away from quartz **(Fig. 5b and Supplementary Fig. 15)**. At the end of 3 ns, all the fragments of kerogen except the ones close to the quartz supercell finally moved beyond the cutoff distance **(Fig. 5c)**. The top chain of kerogen merged into the bottom chain at that point.

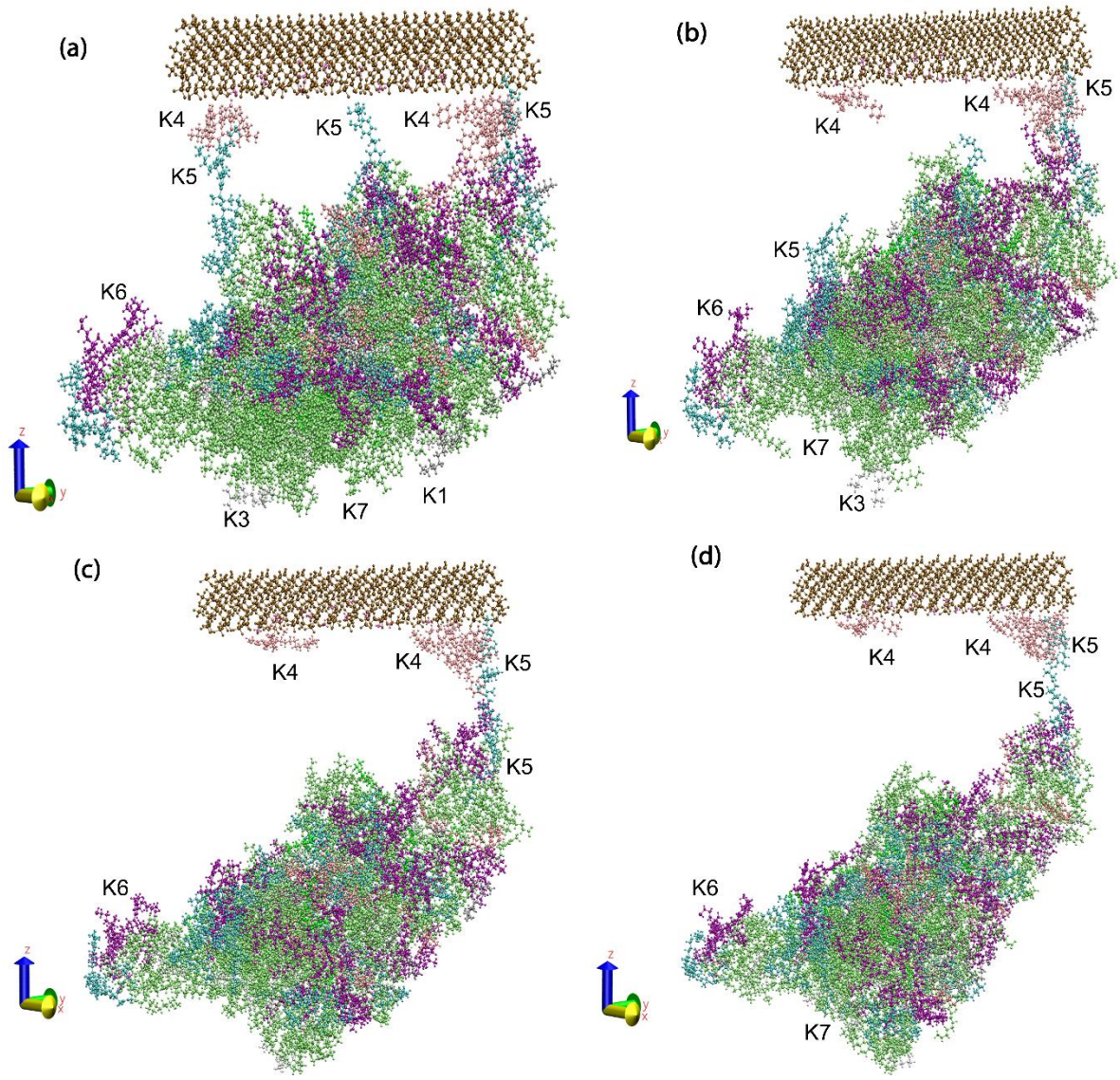

Figure 3.5. The conformation of 'quartz supercell on top of kerogen' model at (a) 1 ns, (b) 2 ns, (c) 3 ns, and (d) 4 ns. All the representations were generated by VMD 1.9.3. Different colored segments represent different fragments of kerogen. Fragment 4 is pink colored, and fragment 5 is cyan colored. Kerogen fragments 1, 2, 3, 4, 5, 6, and 7 are represented by K1, K2, K3, K4, K5, K6, and K7, respectively.

This action increased the interactions between the kerogen chains (layers). As the kerogen fragments did not change their conformations from 3 ns to 4 ns, the simulation is considered to reach the equilibrium zone at 4 ns **(Fig. 5d)**.

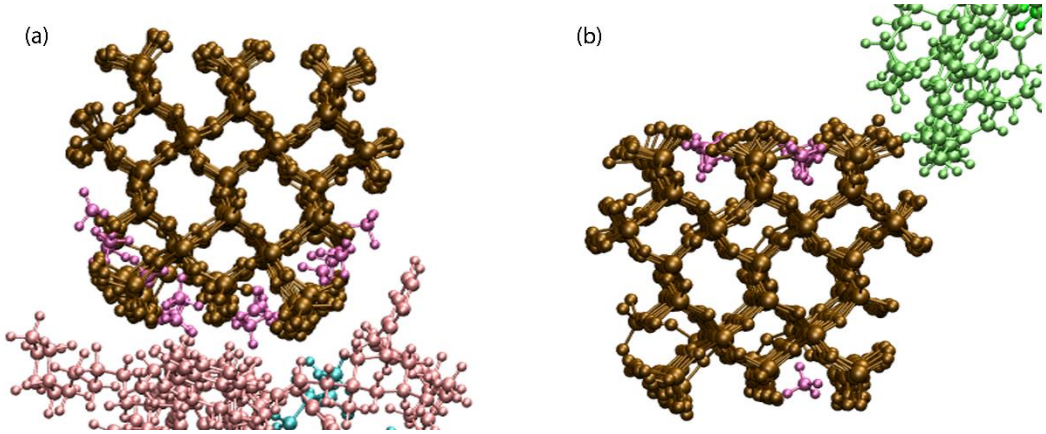

Figure 3.6. A magnified view of stuffing sites (cavities) of quartz supercell providing accommodation to NH_4 ions in (a) quartz supercell on top of kerogen model and (b) quartz supercell at the bottom of kerogen model. The five-atomic mauve-colored segments are ammonium (NH_4) ions.

Table 3.2. Nonbonded interaction energies (kJ/mol) of quartz supercell with kerogen fragments with time in 'quartz supercell on top of kerogen' model

Kerogen fragments	**Total nonbonded energy (kJ/mol)**			
	1 ns	2 ns	3 ns	4 ns
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	-602	-742	-756	-750
5	-858	-592	-450	-442
6	-2	-18	0	0
7	0	-10	0	0
8	-25276	-27665	-27468	-27486

The nonbonded interaction energy is the summation of electrostatic and Van der Waals (VDW) energy. From **Table 2**, it is evident that only the fragments 4, 5, and 8 (NH_4 groups) of 12-unit kerogen interacted with the quartz supercell. Fragments 6 and 7 had very little interaction with the quartz. Approximately 95% of the total attractive interactions are a result of the interaction between NH_4 groups and quartz.

The bar graph (**Fig. 4b**) shows that the total nonbonded energy between quartz supercell and 12-unit kerogen increased from -26738 kJ/mol at 1 ns to -29027 kJ/mol at 2 ns. The negative energy values indicate attractive interactions. Then it reached the equilibrium at 4 ns with the energy slightly decreasing to -28678 kJ/mol. The plot also exhibits that attractive nonbonded interactions are mainly electrostatic. The nonbonded interactions occur only between the top chain (layer) of kerogen and quartz. The nonbonded energies between the top and bottom chain (layer) of kerogen were calculated both in the absence and presence of quartz supercell. In the absence of quartz, the nonbonded energy between the top and bottom chain (layer) was -2656 kJ/mol. However, as the quartz supercell was placed on the top of kerogen, the nonbonded energy between kerogen chains (layers) increased to -4862 kJ/mol at 1 ns, -5135 kJ/mol at 2 ns, -5372 kJ/mol at 3 ns and -5356 kJ/mol at 4 ns of simulation. The merging of both chains and migration of some NH_4 groups from the top chain (layer) to bottom chain (layer) can be attributed to this increase of nonbonded energy.

3.3.2. Quartz Supercell at the Bottom of the Kerogen Model

Figure 7a illustrates that the agglomeration of bonded, nonbonded, and kinetic energies of this model attains the equilibrium condition around 4 ns of simulation. Snapshots were taken at 1ns increments to visualize the changes at various stages of the simulation (**Fig. 8**).

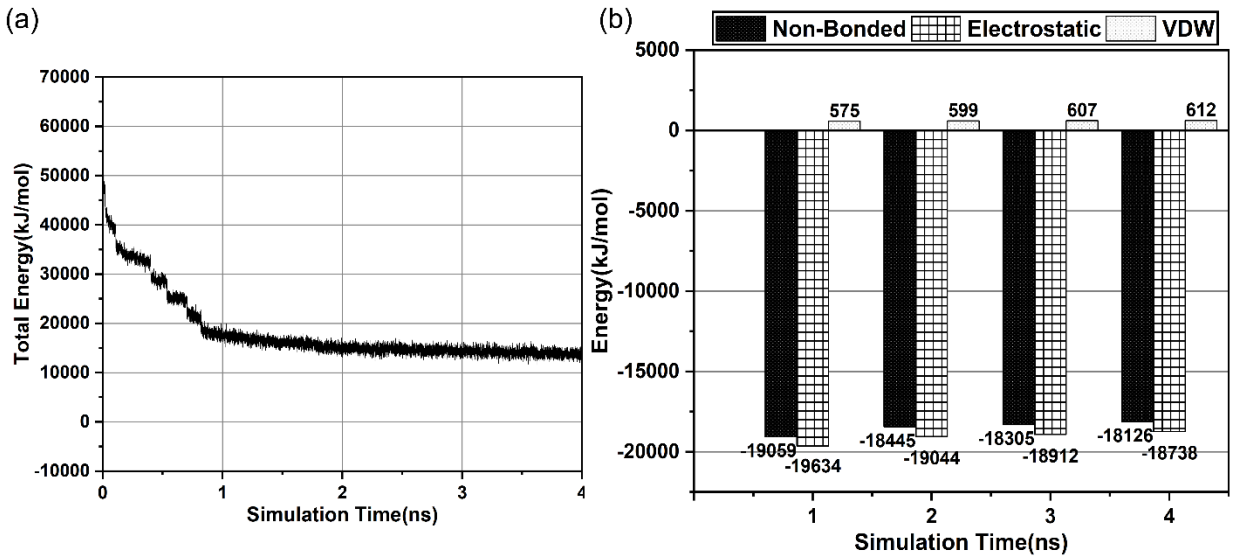

Figure 3.7. (a) Total energy vs. time plot, and (b) bar graph showing nonbonded energy acting between quartz supercell and 12-unit kerogen with time for 'quartz supercell at the bottom of kerogen' model.

The beginning of the simulation at room temperature and pressure (300K and 1.01325 bar) was marked by the rapid migration of 5 NH_4 groups from the bottom chain (layer) of kerogen to quartz supercell. All these NH_4 groups filled some of the stuffing sites of the supercell **(Fig. 6b)**. Fragments 3 and 7 from the bottom chain of kerogen moved slightly towards the quartz. At the end of 1 ns, 5 more NH_4 groups from the bottom chain moved to the stuffing sites of quartz supercell. Fragments 2, 3 and 7 were interacting with the quartz supercell at this point **(Fig. 8a)**

All these interacting fragments belonged to the bottom chain (layer) of kerogen. After simulation of 2 ns, all the fragments except fragment 7 moved away from the quartz supercell. Three segments of fragment 7 from the bottom chain (layer) were interacting with the supercell **(Fig. 8b)**. As the simulation approached 3 ns, the interaction behavior remained similar to the one observed at 2 ns **(Fig. 8c)**. At the end of 4 ns of simulation, only two segments of fragment 7 had interaction with the quartz supercell (Supplementary Fig. 17). A segment of fragment 5 from the bottom chain mostly detached from the rest of kerogen **(Fig. 8d)**.

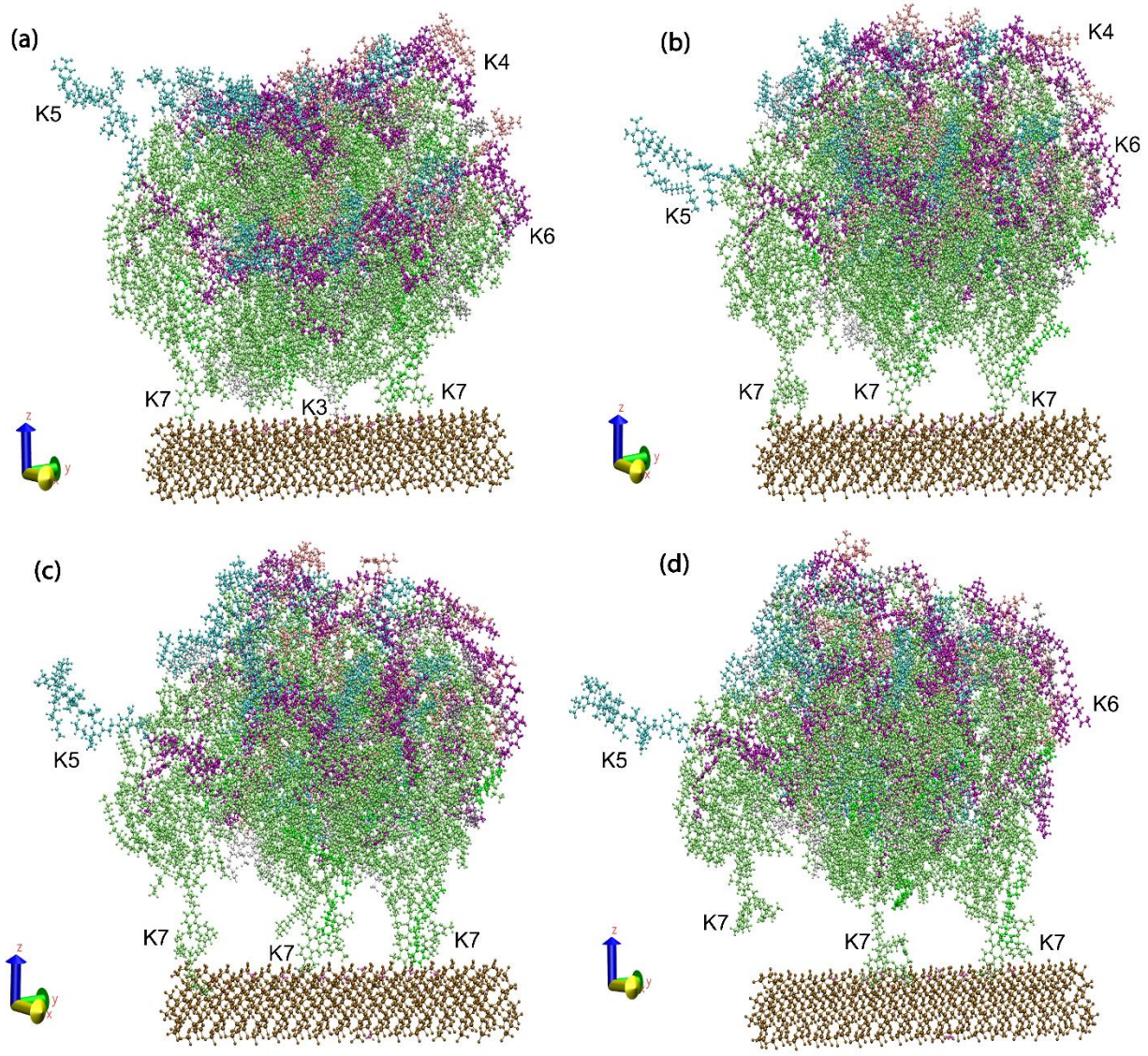

Figure 3.8. The conformation of ‘quartz supercell at the bottom of kerogen’ model at (a) 1 ns, (b) 2 ns, (c) 3 ns, and (d) 4 ns. All the representations were generated by VMD 1.9.3. Different colored segments represent different fragments of kerogen. Fragments 3 and 7 are white and lime-colored, respectively. The ‘mostly detached’ fragment 5 is a cyan colored leftmost segment at 4 ns. Kerogen fragments 1, 2, 3, 4, 5, 6, and 7 are represented by K1, K2, K3, K4, K5, K6, and K7 respectively.

Table 3.3 implies that in this model, only fragments 7 and 8 had significant nonbonded interactions with quartz supercell. Fragments 2 and 3 had little interaction with supercell, which further diminished at 4 ns equilibrium condition.

Table 3.3. Nonbonded interaction energies (kJ/mol) of quartz supercell with kerogen fragments over time in 'quartz supercell at the bottom of kerogen' model

Kerogen fragments	**Total nonbonded interaction energy (kJ/mol)**			
	1 ns	2 ns	3 ns	4 ns
1	-1	0	0	0
2	-73	-48	-45	-11
3	-66	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	-1542	-1026	-910	-767
8	-17377	-17371	-17350	-17348

Compared to the first model, NH_4 groups had 10,138 kJ/mol less nonbonded attractive interactions with quartz in this model. This phenomenon can arise from the difference in the number of NH_4 groups attached to the quartz supercell in the two models. In the first model, 14 NH_4 groups were attached to the supercell, while in the second model, 10 NH_4 groups were attached to the supercell.

The total nonbonded energy between quartz supercell and 12-unit kerogen decreased gradually with simulation time and reached equilibrium at 4 ns **(Fig. 7b)**. The magnitude of nonbonded energy was -19059 kJ/mol at 1 ns, -18445 kJ/mol at 2 ns, -18305 kJ/mol at 3 ns and -18126 kJ/mol at 4 ns. The movement of kerogen fragments away from quartz supercell may be attributed to the observed decrease in the energy. Nonbonded interactions are mainly electrostatic similar to the first model. All the interactions took place between the bottom chain (layer) of kerogen and quartz. The nonbonded energy between the top and bottom chains (layers) of kerogen is also computed for this model. The values are found to be -3100 kJ/mol at 1 ns, which slightly

changed to -3175 kJ/mol at 4 ns. This observation implies that the presence of quartz supercell at the bottom of kerogen did not result in a large change in the inter-chain interaction of kerogen, as observed in the first model (quartz supercell on top of kerogen). In this model, instead of merging into each other, both chains (layers) remained side by side.

We have also investigated the inter-fragmental nonbonded energies inside kerogen for all three conditions (without quartz, quartz on top, and quartz at the bottom). **Table 4** indicates that the presence of quartz caused changes in the interaction energies among the fragments of kerogen.

Table 3.4. Nonbonded interaction energies between kerogen fragments in three different conditions

Interacting kerogen fragments	**Total nonbonded energy (kJ/mol)**		
	Kerogen only	Quartz on top of kerogen	Quartz at the bottom of kerogen
5-6	364	1548	673
5-7	2129	2493	2393
5-8	-13222	-13628	-11972
6-7	3870	3630	3986
6-8	-19795	-18314	-17879
7-8	-57129	-42703	-52200

The attractive interactions of fragment 8 towards the fragments 6 and 7 were reduced when quartz was placed in the proximity of kerogen. The nonbonded energy between fragments 7 and 8 decreased from -57129 kJ/mol (without quartz) to -42703 kJ/mol (quartz on top) and -52200 kJ/mol (quartz at the bottom). Nonbonded energy between fragments 6 and 8 decreased from -19795 kJ/mol (without quartz) to -18314 kJ/mol (quartz on top) and -17879 kJ/mol (quartz at the bottom). Repulsive interactions between fragments 5 and 6 increased from 364 kJ/mol (without quartz) to 1548 kJ/mol (quartz on top) and 673 kJ/mol (quartz at the bottom). The interactions

between fragments 5 and 7 became more repulsive with the presence of quartz (2129 kJ/mol in the absence of quartz, 2493 kJ/mol for quartz on top and 2393 kJ/mol for quartz at the bottom condition).

From the above analysis, it is evident that only the adjacent chain of kerogen interacted with quartz mineral in both orientationsbecause the kerogen chains farther away from the quartz surface will have very small or infinitesimal interaction energy values with quartz. Among the eight fragments of kerogen, fragments 4, 5, 7, and 8 (NH_4 group) interacted significantly with quartz supercell. Fragments 4, 5, and 7 contain aromatic hydrocarbon rings (Supplementary Fig. 19). During the interaction, these hydrocarbon rings were attracted to the quartz. The hydrogen atoms of hydrocarbons were directed toward the oxygen atoms of quartz. The rapid movement of NH_4 groups to the quartz supercell is likely caused by their low molecular weight and higher electrostatic interactions with the supercell. Fragments 5, 7, and 8 contain nitrogen atoms in their structure. Fragment 5 contains 4 nitrogen atoms while fragment 7 contains 10 nitrogen atoms. These nitrogen atoms of fragment 5 and 7 are either aliphatic or aromatic in nature. The non-interacting fragment 6 contains only 1 nitrogen atom which is aliphatic. Therefore, it can be hypothesized that aromatic nitrogen content contributes for the interactions of fragment 5 and 7 with quartz mineral. It proved the hypothesis that one of the major binding interactions in the kerogen-mineral matrix occurs between nitrogen-containing organics and silicate minerals [42]. Previous modeling study also exhibited that kerogen moieties with nitrogen-content have significant nonbonded interactions with phyllosilicate clay minerals [28]. This study identifies the dominant interactions between nitrogen-containing fragments of bulk kerogen model and quartz mineral qualitatively as well as quantitatively. Our observations of MD simulations differ with the observations found by quantum chemical (QM) calculations [25]. These deviations may arise from

the differences in the types of shales and kerogens used. On top of that, we did not consider the presence of water on the silica surface, unlike the QM calculations. It is observed that the different initial model can result in different interaction energy magnitudes for the kerogen-quartz system, however, the relative contributions made by the fragments remains similar. (Supplementary Fig, 10, 11 and supplementary Table 1).

In summary, this is the very first studiey that models the molecular level interactions of macromolecular type-I kerogen found in Green River Oil Shale with quartz mineral. This study reveals which kerogen fragments primarily interact with quartz mineral, the reasons behind interactions and the relative strength of their interactions. This study also shows how the ammonium ions occupy the quartz cavities indicating their strong affinities toward silicate minerals. These models could serve as *in silico* testbeds to screen new methodologies for efficient extraction of kerogen from the oil shale mineral matrix.

3.4. Conclusions

In the current study, we have performed molecular dynamics (MD) simulations on two different arrangements of quartz supercell and 12-unit kerogen model of Green River oil shale. We primarily studied the nonbonded interactions between different fragments of kerogen and α-quartz. We have also explained the conformations of both models at room temperature and pressure. Our findings can be summarized as:

1. The interaction between quartz supercell and 12-unit kerogen is initial-orientation specific. The initial arrangement dictates the amount of nonbonded energy between the mineral and organic constituents.
2. Quartz-kerogen interplay is fragment specific. Only the fragments (4, 5, and 7) with a higher number of aromatic hydrocarbons and Nitrogen elements significantly

interact with quartz. Only kerogen fragments with aromatic hydrocarbons and nitrogen have significant interactions with quartz.

3. The most significant interaction occurs between fragment 8 (NH_4 groups) and quartz. For both models, as soon as the simulation begins, the NH_4 groups from adjacent kerogen chain (layer) migrate into the stuffing sites of the quartz supercell.
4. Nonbonded interactions between kerogen and quartz mineral are predominantly electrostatic.
5. Inter-fragment interaction energies inside kerogen vary depending upon the presence and location of quartz mineral.
6. The presence of quartz mineral significantly affects the inter-chain (inter-layer) energy within the kerogen. This behavior also depends upon the location of the quartz supercell with respect to the kerogen.

3.5. Acknowledgments

The authors would like to acknowledge the financial support from the US Department of Transportation Mountain-Plains Consortium (MPC) Grant #: DTRT13-G-UTC38. We also recognize the computational facilities provided by North Dakota State University Computationally Assisted Science and Technology (CCAST) and NSF MRI grant #1229316 and # #2019077.

3.6. References

(1) Deng, S.; Wang, Z.; Gu, Q.; Meng, F.; Li, J.; Wang, H. Extracting hydrocarbons from Huadian oil shale by sub-critical water. *Fuel Processing Technology* **2011**, *92* (5), 1062-1067. DOI: 10.1016/j.fuproc.2011.01.001.

(2) Alstadt, K. N.; Katti, D. R.; Katti, K. S. An in situ FTIR step-scan photoacoustic investigation of kerogen and minerals in oil shale. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2012**, *89*, 105-113. DOI: 10.1016/j.saa.2011.10.078.

(3) Guo, H.; Lin, J.; Yang, Y.; Liu, Y. Effect of minerals on the self-heating retorting of oil shale: Self-heating effect and shale-oil production. *Fuel* **2014**, *118*, 186-193. DOI: 10.1016/j.fuel.2013.10.058.

(4) Taskforce on Strategic Unconventional Fuels. 2007. America's Strategic Unconventional Fuels. Washington (DC): US Government.

(5) Karabakan, A.; Yurum, Y. Effect of the mineral matrix in the reactions of oil shales: 1. Pyrolysis reactions of Turkish Goynuk and US Green River oil shales. *Fuel* **1998**, *77* (12), 1303-1309. DOI: 10.1016/s0016-2361(98)00045-3.

(6) Yan, J.; Jiang, X.; Han, X.; Liu, J. A TG-FTIR investigation to the catalytic effect of mineral matrix in oil shale on the pyrolysis and combustion of kerogen. *Fuel* **2013**, *104*, 307-317. DOI: 10.1016/j.fuel.2012.10.024.

(7) Razvigorova, M.; Budinova, T.; Tsyntsarski, B.; Petrova, B.; Ekinci, E.; Atakul, H. The composition of acids in bitumen and in products from saponification of kerogen: Investigation of their role as connecting kerogen and mineral matrix. *International Journal of Coal Geology* **2008**, *76* (3), 243-249. DOI: 10.1016/j.coal.2008.07.011.

(8) Cane, R. F. In Oil Shale·, Yen, T. F., Chilingarian, GV, Eds. Elsevier: Amsterdam: 1976.

(9) Eglinton, T. I.; Rowland, S. J.; Curtis, C. D.; Douglas, A. G. Kerogen mineral reactions at raised temperatures in the presence of water. *Organic Geochemistry* **1986**, *10* (4-6), 1041-1052. DOI: 10.1016/s0146-6380(86)80043-2.

(10) Aboulkas, A.; El Harfi, K. Study of the kinetics and mechanisms of thermal decomposition of moroccan tarfaya oil shale and its kerogen. *Oil Shale* **2008**, *25* (4), 426-443. DOI: 10.3176/oil.2008.4.04.

(11) Borrego, A. G.; Prado, J. G.; Fuente, E.; Guillen, M. D.; Blanco, C. G. Pyrolytic behaviour of Spanish oil shales and their kerogens. *Journal of Analytical and Applied Pyrolysis* **2000**, *56* (1), 1-21. DOI: 10.1016/s0165-2370(99)00092-3.

(12) Sert, M.; Ballice, L.; Yuksel, M.; Saglam, M. Effect of mineral matter on product yield and composition at isothermal pyrolysis of turkish oil shales. *Oil Shale* **2009**, *26* (4), 463-474. DOI: 10.3176/oil.2009.4.03.

(13) Pan, L.; Dai, F.; Huang, P.; Liu, S.; Li, G. Study of the effect of mineral matters on the thermal decomposition of Jimsar oil shale using TG-MS. *Thermochimica Acta* **2016**, *627*, 31-38. DOI: 10.1016/j.tca.2016.01.013.

(14) Alstadt, K. N.; Katti, K. S.; Katti, D. R. Nanoscale Morphology of Kerogen and In Situ Nanomechanical Properties of Green River Oil Shale. *Journal of Nanomechanics and Micromechanics* **2016**, *6* (1). DOI: 10.1061/(asce)nm.2153-5477.0000103.

(15) Ghosh, P.; Katti, D. R.; Katti, K. S. Mineral proximity influences mechanical response of proteins in biological mineral-protein hybrid systems. *Biomacromolecules* **2007**, *8* (3), 851-856. DOI: 10.1021/bm060942h.

(16) Sikdar, D.; Katti, K. S.; Katti, D. R. Molecular interactions alter clay and polymer structure in polymer clay nanocomposites. *Journal of Nanoscience and Nanotechnology* **2008**, *8* (4), 1638-1657. DOI: 10.1166/jnn.2008.032.

(17) Bhowmik, R.; Katti, K. S.; Katti, D. R. Mechanisms of Load-Deformation Behavior of Molecular Collagen in Hydroxyapatite-Tropocollagen Molecular System: Steered

Molecular Dynamics Study. *Journal of Engineering Mechanics* **2009**, *135* (5), 413-421. DOI: 10.1061/(asce)0733-9399(2009)135:5(413).

(18) Bhowmik, R.; Katti, K. S.; Katti, D. R. Molecular interactions of degradable and non-degradable polymers with hydroxyapatite influence mechanics of polymer-hydroxyapatite nanocomposite biomaterials. *International Journal of Nanotechnology* **2009**, *6* (5-6), 511-529. DOI: 10.1504/ijnt.2009.024642.

(19) Ballice, L. Stepwise chemical demineralization of Göynük (Turkey) oil shale and pyrolysis of demineralization products. *Industrial & engineering chemistry research* **2006**, *45* (3), 906-912.

(20) Vandenbroucke, M.; Largeau, C. Kerogen origin, evolution and structure. *Organic Geochemistry* **2007**, *38* (5), 719-833.

(21) Forsman, J. P.; Hunt, J. M. Insoluble organic matter (kerogen) in sedimentary rocks. *Geochimica et Cosmochimica Acta* **1958**, *15* (3), 170-182. Saxby, J. D. Isolation of kerogen in sediments by chemical methods. *Chemical Geology* **1970**, *6*, 173-184.

(22) Liu, Q.; Hou, Y.; Wu, W.; Wang, Y.; Wang, Q.; Ren, S. New Insights into the Occurrence and Interaction of Inorganic Minerals and Organic Matter in Huadian Oil Shale. *Energy & Fuels* **2019**, *33* (2), 859-867.

(23) Chang, Z.; Chu, M.; Zhang, C.; Bai, S.; Lin, H.; Ma, L. Compositional and structural variations of bitumen and its interactions with mineral matters during Huadian oil shale pyrolysis. *Korean Journal of Chemical Engineering* **2017**, *34* (12), 3111-3118.

(24) Zhang, Z.; Liu, H.; Wang, J. Energetics of interfacial interactions of hydrocarbon fluids with kerogen and calcite using molecular modeling. *Energy & Fuels* **2020**, *34* (4), 4251-4259.

(25) Hantal, G.; Brochard, L.; Cordeiro, M.; Ulm, F. J.; Pellenq, R. J. M. Surface Chemistry and Atomic-Scale Reconstruction of Kerogen-Silica Composites. *Journal of Physical Chemistry C* **2014**, *118* (5), 2429-2438. DOI: 10.1021/jp406329n.

(26) Katti, D. R.; Thapa, K. B.; Katti, K. S. Modeling molecular interactions of sodium montmorillonite clay with 3D kerogen models. *Fuel* **2017**, *199*, 641-652. DOI: 10.1016/j.fuel.2017.03.021.

(27) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. An insight into quartz mineral interactions with kerogen in Green River oil shale. North Dakota State University: 2019.

(28) Katti, D. R.; Upadhyay, H. B.; Katti, K. S. Molecular interactions of kerogen moieties with Na-montmorillonite: An experimental and modeling study. *Fuel* **2014**, *130*, 34-45. DOI: 10.1016/j.fuel.2014.04.009.

(29) Katti, D. R.; Katti, K. S.; Thapa, K.; Faisal, N. Modeling the Nanoscale Kerogen Inclusions in Green River Oil Shale. In *Poromechanics VI*, pp 1968-1975.

(30) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Modeling the behavior of organic kerogen in the proximity of calcite mineral by molecular dynamics simulations. *Energy & Fuels* **2020**.

(31) Skelton, A. A.; Fenter, P.; Kubicki, J. D.; Wesolowski, D. J.; Cummings, P. T. Simulations of the Quartz(10(1)over-bar1)/Water Interface: A Comparison of Classical Force Fields, Ab Initio Molecular Dynamics, and X-ray Reflectivity Experiments. *Journal of Physical Chemistry C* **2011**, *115* (5), 2076-2088. DOI: 10.1021/jp109446d.

(32) Pikunic, J.; Clinard, C.; Cohaut, N.; Gubbins, K. E.; Guet, J. M.; Pellenq, R. J. M.; Rannou, I.; Rouzaud, J. N. Structural Modeling of porous carbons: Constrained reverse Monte Carlo method. *Langmuir* **2003**, *19* (20), 8565-8582. DOI: 10.1021/la034595y.

(33) Le Page, Y.; Donnay, G. Refinement of the crystal structure of low-quartz. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **1976**, *32* (8), 2456-2459.

(34) Ikuta, D.; Kawame, N.; Banno, S.; Hirajima, T.; Ito, K.; Rakovan, J. F.; Downs, R. T.; Tamada, O. First in situ X-ray identification of coesite and retrograde quartz on a glass thin section of an ultrahigh-pressure metamorphic rock and their crystal structure details. *American Mineralogist* **2007**, *92* (1), 57-63. DOI: 10.2138/am.2007.2228.

(35) Siskin, M.; Scouten, C. G.; Rose, K. D.; Aczel, T.; Colgrove, S. G.; Pabst, R. E. Detailed structural characterization of the organic material in Rundle Ramsay Crossing and Green River oil shales. In *Composition, geochemistry and conversion of oil shales*, Springer, 1995; pp 143-158.

(36) Orendt, A. M.; Pimienta, I. S. O.; Badu, S. R.; Solum, M. S.; Pugmire, R. J.; Facelli, J. C.; Locke, D. R.; Chapman, K. W.; Chupas, P. J.; Winans, R. E. Three-Dimensional Structure of the Siskin Green River Oil Shale Kerogen Model: A Comparison between Calculated and Observed Properties. *Energy & Fuels* **2013**, *27* (2), 702-710. DOI: 10.1021/ef3017046.

(37) Kelemen, S. R.; Siskin, M. Organic matter models of oil shale revisited. *Preprints-American Chemical Society. Division of Petroleum Chemistry* **2004**, *49* (1), 73-76.

(38) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(39) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials* **2014**, *26* (8), 2647-2658. DOI: 10.1021/cm500365c.

(40) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289.

(41) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative minimization techniques for abinitio total-energy calculations - molecular-dynamics and conjugate gradients. *Reviews of Modern Physics* **1992**, *64* (4), 1045-1097. DOI: 10.1103/RevModPhys.64.1045.

(42) Siskin, M.; Brons, G.; Payack, J. F. Disruption of kerogen mineral interactions in oil shales. *Energy & Fuels* **1987**, *1* (3), 248-252. DOI: 10.1021/ef00003a004.

4. MOLECULAR MECHANICS OF THE SWELLING CLAY TACTOID UNDER COMPRESSION, TENSION AND SHEAR[3]

4.1. Introduction

Evaluation and prediction of reactive soil response is a topic of great interest among the researchers. These soils undergo significant changes in volume and microstructure under certain environmental conditions. There are three major types of reactive soils; expansive, collapsible, and chemically reactive. Expansive soils are found extensively in the USA and also in other parts of the world. In the case of expansive soils, the dramatic volume change, degradation of shear strength upon swelling, and the development of considerable swelling pressures when constrained, cause enormous damage to the infrastructure, including distress to buildings, foundations, highways, dams, etc. Smectite clays such as montmorillonite constitute a significant portion of expansive soils. Montmorillonite swells significantly upon interaction with water. Under free swell conditions, montmorillonite can expand as much as 15 times its original volume. However, if the swelling is constrained, it exerts a large amount of swelling pressure [1]. As a swelling clay, montmorillonite also has beneficial applications in geoenvironmental engineering, polymer clay nanocomposites, pharmaceutical, cosmetics, and the petroleum industry.

Montmorillonite is a member of the smectite group. Clay minerals are defined as fine-grained, earthy, and natural materials that exhibit plasticity with the addition of water [2]. Smectite clay belongs to the aluminum phyllosilicate group, with a varying combination of tetrahedral silica sheet (t) and octahedral alumina sheet (o). Sodium montmorillonite (Na-Mt) is a 2:1 (t-o-t) phyllosilicate mineral with an alumina octahedral sheet sandwiched between two tetrahedral silica

[3] The contents of this study have been published in Faisal, HM Nasrullah, Kalpana S. Katti, and Dinesh R. Katti. "Molecular mechanics of the swelling clay tactoid under compression, tension and shear." *Applied Clay Science* 200 (2020): 105908. As the lead author of this journal article, the copyright permission has been obtained from Elsevier.

sheets [3]. Isomorphous substitution of aluminum ions from the octahedral sheet causes the development of negative charge, which is balanced by different interlayer cations (for example, Na, Mg, K, Ca, Cs).

The swelling of Na-montmorillonite is characterized by structural changes both at the nano and micro scale. Clay interlayer space increased with an increasing amount of hydration in the nanometer dimension [4]. The breakdown of Na-Mt particle occurred during swelling [5]. Discrete element modeling (DEM) study on swelling clays simulated particle breakdown [6]. To bridge the swelling behavior between the nano and micro scale, modeling of the clay structure at intermediate levels is essential. According to earlier studies, Na-montmorillonite possesses a hierarchical structure. Investigation of clay porosities using scanning electron microscopy (SEM), small-angle X-ray scattering (**SAXS**), X-ray powder diffraction (**XRD**), and mercury intrusion porosimetry (**MIP**) highlighted three sizes of porosities, interlayer porosity, inter-particle porosity and inter-aggregate porosity [7]. Analysis of the difference between hydration of clay-rich soils and hydration of porous granular soils led to the understanding of the multiscale structure of clays [8]. This multiscale structure consists of the microscopic scale (t-o-t layer), mesoscopic scale (stacks of t-o-t layers), and macroscopic scale (a random combination of stacks). The hierarchical structure of smectite (layer, particle, single aggregate, multiple aggregates) was also described for defining its various hydration sites as a function of scale [9]. Thermoporometric study of mesoscopic swelling of Na-Mt clay mineral, also referred to its stratified architecture [10]. Based on these observations, we can state that the hierarchical structure of Na-Mt clay mineral is composed of four levels.

- Unit clay mineral layer: T-O-T sheet structure with interlayer sodium cations.
- Tactoid: Several clay mineral layers (about 10) stacked one over other in the Z direction to form a clay tactoid.

- Aggregate: Several clay tactoids are gathered in different orientations to form a clay aggregate.
- Multiple aggregate/ Assembly of aggregates: Some clay aggregates are combined to form a multiple-aggregate.

The unit clay mineral layers (t-o-t) of Na-Mt clay mineral form a stack along the vertical (Z) direction. This stack is termed as clay tactoid. Clay tactoid structure is often viewed as the characteristic feature of consolidated natural bentonite bed where significant overloading yields the alignment of clay mineral layers into stacks. Previous X-ray diffraction studies showed the formation of Na-Mt clay tactoids containing ten clay mineral layers stacked to each other in the Z-axis [11]. However, the number of clay mineral layers in a tactoid varies in aqueous solutions with varying ion concentration. Increased ionic concentration (salinity) results in tactoids having higher number of clay mineral layers. Na-montmorillonite tactoid was seen to have 5 to 20 clay mineral layers in solution depending upon the NaCl concentration with an average of 10 clay mineral layers [12]. Clay tactoid structure is also affected by the relative humidity (RH). Nitrogen adsorption study of montmorillonite showed that the hydration at RH of 100% caused the increment of both total clay mineral surface area and micropore (interlayer) surface area i.e. decreased number of clay mineral layers in tactoid [13]. Clay tactoid also behaves as the primary clay particle, which forms aggregates upon organization in different orientations. In the present study, tactoid will be considered to be made up of ten clay mineral layers.

Researchers evaluated different properties of clay minerals through experimental and computational approaches. Influence of temperature, pressure, humidity etc. on clay minerals were investigated to better understand the impact of these environmental factors on the mechanical properties of clays [14]. Molecular dynamics (MD) simulations have long been utilized by the

researchers to investigate various geological materials as well as clay minerals, to predict the time-dependent behavior of molecular systems [15-17]. MD simulations were also used to model the hydration behavior of swelling clay minerals [17, 18]. Nanoscale bending of clay mineral layers and their relationship with cation exchange capacity (CEC) and terminal shear stress were investigated by different researchers [19, 20]. Steered molecular dynamics (SMD) technique was utilized to evaluate the mechanical response of Na-Mt clay mineral [21-24]. The approach behind performing steered molecular dynamics (SMD) is that initially, a subset of atoms is fixed. In contrast, another subset of atoms is pulled/pushed with predetermined constant-velocity or constant-force, and the response of the molecule is recorded. Interactions of various fluid molecules with clay minerals were also studied using molecular dynamics (MD) to evaluate the effect of fluid polarity on swelling [25]. The swelling of Na-Mt clay mineral interlayer spaces inside Na-Mt clay tactoids contributed to the swelling of the tactoid and the clay [9]. Clays are subject to a variety of loading paths such as compression, tension, and shear during construction and operation of civil infrastructure components such as structures, excavations, tunnels, etc. Since clay tactoids are fundamental building blocks of clay aggregates, modeling Na-Mt clay tactoid subject to various loading paths will provide quantitative parameters and an insight into the response of the clay tactoid under various loading conditions. The present study seeks to evaluate the mechanical properties of Na-Mt clay tactoid and investigate molecular-scale mechanisms during loading using both MD and SMD.

4.2. Methodology

4.2.1. Modeling Approach

The authors have performed various experimental and modeling studies on Na-montmorillonite (SWy-2) obtained from clay mineral repository at the University of Missouri, Columbia [5, 23, 24]. Although the structural formula of this Na-montmorillonite is,

$(Ca_{0.12}Na_{0.32}K_{0.05})[Al_{3.01}(III)_{0.41}Mn_{0.01}Mg_{0.54}Ti_{0.02}][Si_{7.98}Al_{0.02}]O_{20}(OH)_4$ [26],

a simplified version $NaSi_{16}(Al_6FeMg)O_{20}(OH)_4$ was utilized for model construction. Previously proposed coordinates [27], partial charges [16], and CHARMm force field parameters [22, 23] were used. Isomorphous substitution occurred in the octahedral sheet (o) as one out of every four Al^{3+} ions was replaced by Mg^{2+} or Fe^{3+} ions. The development of negative charge in the clay mineral layer due to isomorphic substitution was neutralized by introducing interlayer Na cation.

The present study utilized the Na-Mt 6×3 clay model as used in several previous studies [25, 28]. The unit cell dimensions of the clay mineral layer were 5.28 Å×9.14 Å×6.56 Å. Hence, the initial overall dimensions of single 6×3 Na-Mt clay mineral layer (t-o-t) were 31.68 Å×27.44 Å×6.56 Å (**Fig. 1**). The initial interlayer space was 3.44 Å, which resulted in the initial d-value of 10 Å. Each 6×3 clay mineral layer contained 720 atoms. As isomorphic substitution generated a charge of -0.5 e in each unit cell, nine interlayer Na cations were introduced in the interlayer space to balance the charge of the 6×3 clay model.

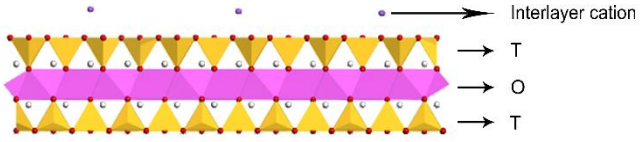

Figure 4.1. Molecular model of 6×3 Na-montmorillonite clay mineral layer with interlayer cations that is replicated to build the tactoid model.

For building the model of Na-Mt clay tactoid, ten 6×3 clay mineral layers were stacked on top of each other in the Z direction. These ten layers were identified as B, C, D, E, F, G, H, I, J, and K layer from bottom to top along the Z-axis, respectively (**Fig. 2**). This model contained nine interlayer spaces. The initial vertical (Z) dimension of Na-Mt clay tactoid was 96.56 Å, the summation of the thickness of 10 clay mineral layers (10×6.56 Å) and nine interlayer spaces (9×3.44 Å). The overall dimensions of Na-Mt clay tactoid were 31.68 Å×27.44 Å × 96.56 Å. Each interlayer space contained nine Na cations. Thus, the full Na-Mt clay tactoid model had 7281 atoms (7200 clay atoms and 81 interlayer Na cations). The distance between the top and bottom clay mineral layers was termed as the 'aggregated d-value' as it was the sum of all the d-values within the Na-Mt tactoid. The initial magnitude of the 'aggregated d-value' was 90 Å. Materials Studio 7.0 was used to build the Na-Mt clay tactoid model.

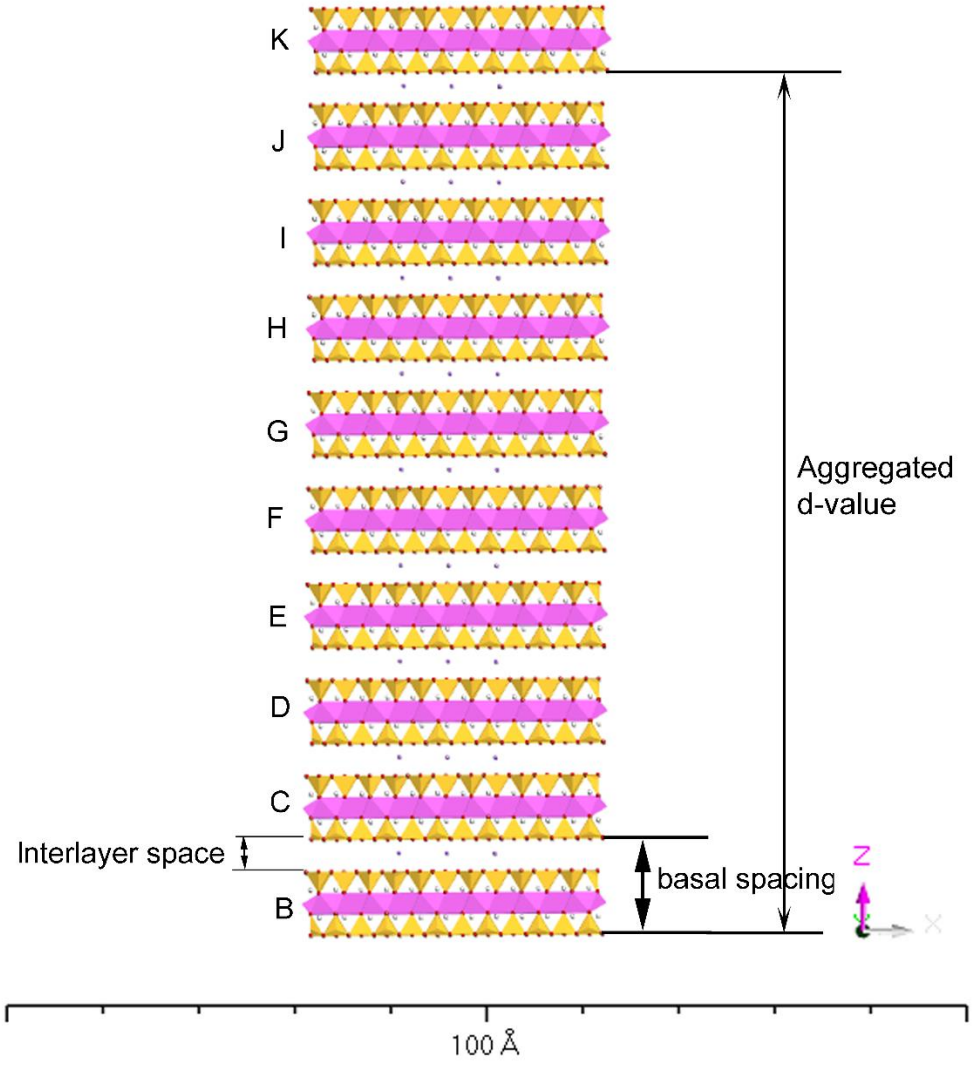

Figure 4.2. Molecular model of Na-montmorillonite tactoid containing ten clay mineral layers. The bottom clay mineral layer was termed as the 'B" layer, and the top clay mineral layer was termed as the 'K' layer.

Na-Mt clay tactoid was parameterized using CHARMm (Chemistry at HARvard Macromolecular Mechanics) force field [29]. A set of functions and associated constants were utilized to describe the energy expression of a molecular system using the CHARMm force field. Consistent force field (CFF) parameters of Na-Mt, developed in an earlier study by MP2 ab intio calculations [16] were transformed to the CHARMm force field [23] and used in this study. A handful of empirical force fields were utilized by researchers to model various properties of different clay

minerals. ClayFF, CementFF, ReaxFF, URF, InterfaceFF etc. are the most commonly used force fields for modeling clay behavior. To facilitate the comparison of these force fields in a single platform, a couple of online force field databases were introduced in a user-friendly manner [30, 31]. However, comparing the performance of force fields is not a straightforward task as it was influenced not only by the accurate reproduction of experimental observables by fitted parameters but also by their computational efficiencies. ClayFF and InterfaceFF were considered the simplest (easily implementable) and most reliable (in terms of partial charges) force fields respectively while ReaxFF was treated as the most unique due to its ability of modeling bond breaking/formation. This particular behavior of ReaxFF limited its capability of modeling the mechanical response [30]. The present study investigated the mechanical response of tactoid which in turn depends on its non-bonded interactions. Therefore, InterfaceFF would be the most suitable option for modeling clay non-bonded interactions. The InterfaceFF force field parameterization was validated by PCFF potential energy expression which employed semiempirical charge equilibration (instead of ab-initio calculations like CFF) to obtain the partial atomic charges of montmorillonite [32]. This semiempirical method was based upon the extended Born model that accounted for both the covalent and ionic bonding through the exploitation of atomization and ionization energies respectively [33]. The partial charges used by the present clay tactoid modeling study (unaltered CFF charges) fell within the range of maximum possible deviations of PCFF semiempirical procedure ($\pm$0.3 e). Therefore, this model efficiently depicted the correct non-bonded interactions within the tactoid like InterfaceFF. The current clay CHARMm parameters were extensively validated with experimental results and other modeling results in the literature and by previous studies [4, 21, 22, 34].

4.2.2. Simulation Details

The MD and SMD simulations of Na-Mt clay tactoid were performed using NAMD 2.12 [35]. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. All the simulations were run on the Center for Computationally Assisted Science and Technology (CCAST), a parallel computing facility at North Dakota State University. The simulations utilized one node and twenty Intel Xeon 2.5 GHz processors with 15GB DDR3 RAM at 1866 MHz.

The Na-Mt tactoid model was first minimized under vacuum conditions (0 K temperature and 0 kPa pressure) using a conjugate gradient method [36]. Next, the model was brought to normal temperature (300 K) and pressure (101.325 kPa) conditions using a stepwise approach. The normal temperature (300 K) was raised from 0 K in three steps (100 K/step). Next, the normal pressure condition (101.325 kPa) was achieved from 0 kPa in four steps (25 kPa/step). The Na-Mt tactoid model simulation was conducted at NTP condition (isobaric-isothermal) for 2 ns to attain the equilibrium condition. As the time step used for MD simulation is 0.5 fs, the model was run at NTP for 4,000,000 steps. All of the MD simulations of Na-Mt tactoid were performed using the periodic boundary conditions (PBC) to implement the Particle Mesh Ewald (PME) electrostatic interactions.

The equilibrated Na-Mt tactoid model was then utilized to evaluate its mechanical response. The tactoid model was subjected to three loading paths, compression, tension, and shear to evaluate the mechanical properties under these loading conditions. The compressive behavior of tactoid was investigated by applying increasing magnitudes of compressive forces on the top clay mineral layer oxygen atoms (layer K) and by keeping the bottom clay mineral layer (layer B)

fixed. The forces applied on each oxygen atom of top clay mineral layer (K) were 0-100 pN in increments of 25 pN followed by increments of 50 pN to 300 pN and up to 2500 pN in increments of 100 pN. The stresses equivalent to these forces were calculated to be ranging from 0 to 37 GPa in 31 increments. Constant force SMD was used to apply these forces on Na-Mt clay tactoid. The d-values of tactoid were measured after these constant-force SMD simulations reached equilibrium conditions for a given load. The d-values were computed by tracking the coordinates of the bottom oxygen atoms of each clay mineral layer. Constant-velocity compression on Na-Mt tactoid was also performed using steered molecular dynamics (SMD) to determine the characteristic spring constant (*k*) and pulling velocity (*v*) of the tactoid. These values were determined by utilizing the constant-force compression response. A number of constant-velocity SMD (compression) simulations were performed on clay tactoid utilizing the combinations of various magnitudes of spring constants (*k= 3, 5, 7, and 9 kcal/mol/Å*^{2}) and velocities (*v=0.5, 1, 1.5, 2, and 2.5 Å/ps*). The specific combination of spring constant (*k*) and velocity (*v*) that provided the identical compression behavior to the constant-force compression was taken as the characteristic spring constant and velocity of tactoid. The values were found to be *k=9 kcal/mol/Å*^{2} and *v=2 Å/ps*, respectively, and further utilized to investigate the tensile and shear properties of tactoid.

In order to investigate the tensile properties of Na-Mt tactoid (stress required to pull off the various number of clay mineral layers from tactoid), tensile (pulling) forces were applied on the top 1 (K), 2 (J, K) and 3 (I, J, K) clay mineral layers by keeping the bottom layer (B) fixed. Constant-velocity SMD (*k=9 kcal/mol/Å*^{2}, *v=2 Å/ps*) was utilized to apply tensile forces. The shearing behavior of Na-Mt tactoid was also studied using constant-velocity SMD (*k=9 kcal/mol/Å*^{2}, *v=2 Å/ps*). We applied force on the surface of the top clay mineral layer (K) along the X-axis

while keeping the bottom clay mineral layer (B) fixed, followed by measuring the corresponding shear deformation.

4.3. Results

4.3.1. Equilibrated Na-Mt Tactoid

The equilibrated Na-Mt tactoid model after 2 ns of simulation at NTP condition (300 K and 101.325 kPa) is shown in **Fig. 3**.

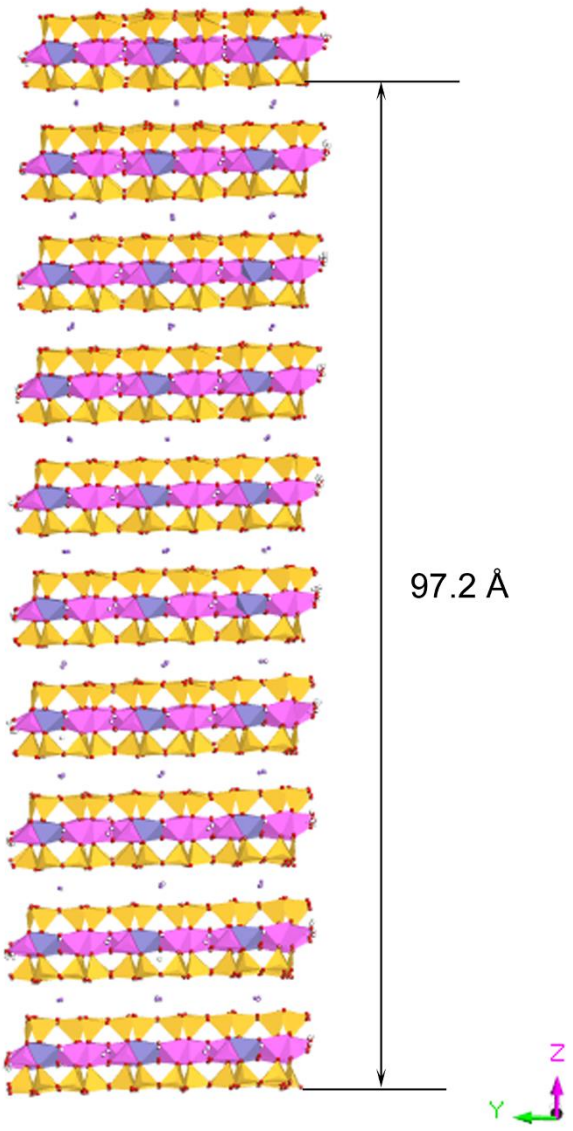

Figure 4.3. NTP equilibrated model of Na-Mt tactoid.

The final height of equilibrated Na-Mt tactoid was 104.04 Å, as opposed to the initial height of 96.56 Å. The final dimensions of Na-Mt tactoid were 33.66 Å×29.08 Å×104.04 Å. The increase in tactoid height resulted from the increase in the interlayer space. The initial and equilibrated aggregated d-value was 90 Å and 97.2 Å, respectively (**Fig. 3**). The change in the thickness of the clay mineral layer was thus minimal (from 6.56 Å to 6.65 Å).

Table 4.1. Calculated d-values between clay mineral layers inside Na-Mt tactoid

Clay mineral layers pair	**Equilibrated d-value (Å)**
B-C	10.79
C-D	10.79
D-E	10.80
E-F	10.80
F-G	10.80
G-H	10.79
H-I	10.80
I-J	10.80
J-K	10.79

Table 1 shows the d-values between each pair of clay mineral layers at equilibrium. As seen, the d-values were quite similar throughout the tactoid, and the layers were parallel to one another. The mean d-value of tactoid was 10.795 Å (≈ 10.80 Å). This was consistent with the previously reported magnitude of d-values from experimental and modeling studies [4, 21, 37]. The average interlayer space was observed to be 4.15 Å. Application of XRD techniques on montmorillonite performed by other researchers provided basal spacing values ranging from 9.6 Å to 12.4 Å [38].

Further, the non-bonded interaction energies for clay mineral layers inside the Na-Mt tactoid were calculated. Non-bonded interaction energies consist of electrostatic and Van der

Waals (VDW) interaction energies. Positive and negative values of the energies referred to repulsive and attractive interaction energies, respectively. Non-bonded energies between each pair of clay mineral layers are shown in **Table 2**.

Table 4.2. Non-bonded energies (kJ/mol) between clay mineral layers in Na-Mt tactoid

Clay mineral layers pair	**Electrostatic (kJ/mol)**	**VDW (kJ/mol)**	**Non-bonded (kJ/mol)**
B-C	751	-2597	-1846
C-D	750	-2596	-1846
D-E	752	-2597	-1845
E-F	746	-2599	-1853
F-G	745	-2599	-1854
G-H	745	-2600	-1855
H-I	744	-2601	-1857
I-J	743	-2602	-1859
J-K	761	-2603	-1842

As seen, the clay mineral layers in the tactoid had relatively uniform attractive pairwise interactions between them with an average value of -1850 kJ/mol. The electrostatic interactions between the clay mineral layers were repulsive, while the Van der Waals interactions were attractive. The average value of Van der Waals energy (-2599 kJ/mol) was found to be 3.5 times greater than average electrostatic energy (749 kJ/mol). Non-bonded energy was also computed for each clay mineral layer to identify its interactions with interlayer Na cations in Na-Mt tactoid.

Table 4.3. Na-clay mineral layer interaction energy in Na-Mt tactoid

Clay-mineral layer	Electrostatic (kJ/mol)	VDW (kJ/mol)	Non-bonded (kJ/mol)
B	-3881	89	-3792
C	-6397	43	-6354
D	-6679	55	-6624
E	-6694	56	-6638
F	-6693	55	-6638
G	-6694	54	-6640
H	-6689	54	-6635
I	-6665	50	-6615
J	-6341	34	-6307
K	-3585	92	-3493

Table 3 indicates that all the clay mineral layers except the top (K) and bottom (B) layers had a similar magnitude of interaction energy with interlayer cations. The interaction energies of interlayer Na cations with top and bottom clay mineral layers were almost one-half (3642 kJ/mol) compared to the interactions with the remaining clay mineral layers (6556 kJ/mol). This difference in the interaction energy occurred because the top and bottom clay mineral layers contained Na cations just on one side of the clay mineral layer, while the other clay mineral layers had cations on both sides. The attractive electrostatic energies were much higher than the repulsive Van der Waals energies among the Na ion and clay mineral layer interaction energies.

The binding energy of a single clay mineral layer to its tactoid is described as the total of Van der Waals and electrostatic energies acting between the clay mineral layer and the rest of the tactoid (other clay mineral layers and interlayer cations). Binding energy reflects how strongly a clay mineral layer is bonded to other constituents in the tactoid.

Table 4.4. Binding energy of each clay mineral layer to tactoid

Clay-mineral layer	Electrostatic (kJ/mol)	VDW (kJ/mol)	Non-bonded (kJ/mol)
B	-3140	-2511	- 5651
C	-5200	-5145	-10345
D	-5200	-5145	-10345
E	-5219	-5147	-10366
F	-5227	-5150	-10377
G	-5230	-5152	-10382
H	-5227	-5154	-10381
I	-5208	-5159	-10367
J	-4851	-5174	-10025
K	-2842	-2515	-5357

The calculated binding energies of the top (K) and bottom (B) clay mineral layers were about 53% of the binding energies of remaining individual clay mineral layers **(Table 4)**. This difference was attributed to the position of the clay mineral layer in the tactoid (top/bottom), the number of interlayer cations interacting with the clay mineral layer, and adjacent clay mineral layer interactions. Electrostatic and Van der Waals energies contributed to about the same order of magnitude to the binding energies of clay mineral layers and both imparted attractive interactions. The total attractive interaction energy between interlayer cations and all the clay mineral layers was -59738 kJ/mol. These results also highlighted the role of other clay mineral layers and Na ions in other interlayers on the binding energy of a clay mineral layer within the tactoid. The binding energy of a single clay mineral layer was higher than its pairwise interactions with adjacent clay mineral layers and cations due to the influence of other clay mineral layers and Na ions in the tactoid.

4.3.2. Compressive Loading on Tactoid

The response of Na-Mt tactoid under an increasing amount of compression was evaluated in the present study. Significant changes occurred in the interlayer space of tactoid due to the application of the compressive stresses. Changes in clay mineral layer thicknesses were minimal. The relationship between increasing compressive stress and aggregated d-value and the resultant stress-strain curves were shown in **Fig 4.**

With the increasing amount of stress, the aggregated d-value of Na-Mt tactoid decreased (**Fig. 4a**). This reduction was the result of the decrease in the interlayer spaces between each pair of clay mineral layers inside the tactoid. The reduction in interlayer space solely contributed to the lessening of d-values between clay mineral layers, since the decrease in clay mineral layer thickness was insignificant. The equilibrated aggregated d-value of Na-Mt tactoid under the compression of 0 GPa was 97.2 Å. As the applied stress increased from 0 GPa to 29.6 GPa, the aggregated d-value gradually decreased from 97.2 Å to 89.10 Å. The Na-Mt tactoid model was deformed at the most by about 8.1 Å under compression because, at this magnitude of compressive stress (29.6 GPa), the interlayer distance approached the Van der Waals radius distance between the clay mineral layers. The stress-strain behavior of Na-Mt tactoid due to compression loading was also determined (**Fig. 4b**).

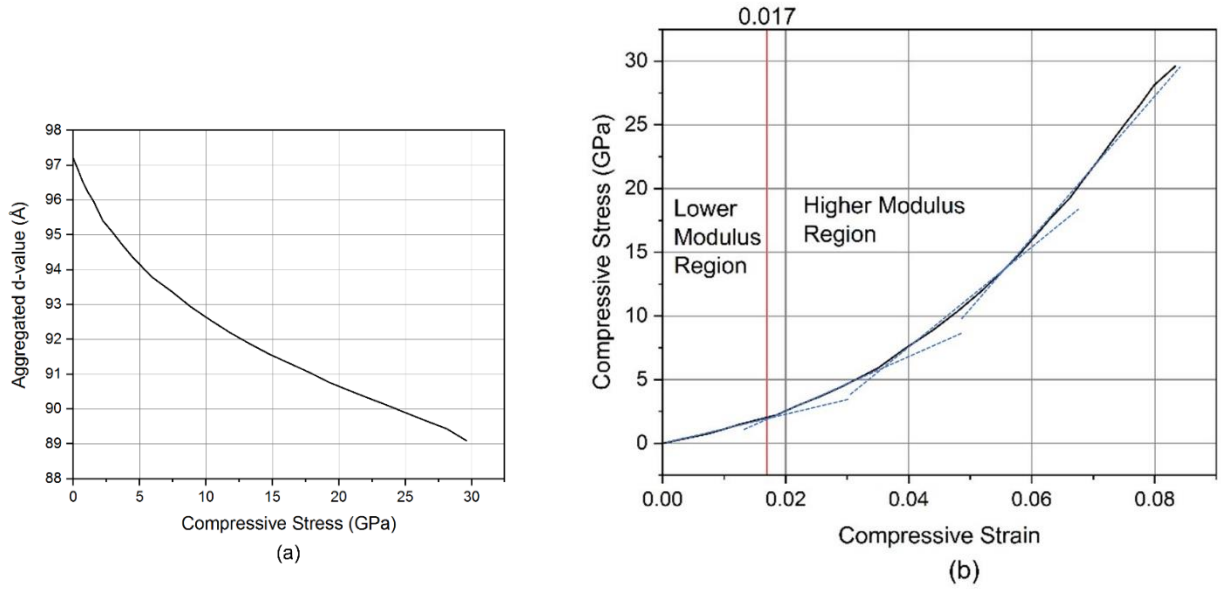

Figure 4.4. Plots showing (a) Aggregated d-value vs. compressive stress and (b) Compressive stress vs. strain of equilibrated Na-Mt tactoid. The vertical line at a strain value of 0.017 specifies the boundary of two different compressive strain behaviors (linear and nonlinear) inside tactoid.

The amount of deformation in aggregated d-value was divided by the initial height of equilibrated tactoid to compute the strain values. This stress-strain plot was primarily composed of two different regions i.e. lower modulus (linear) region and higher modulus (non-linear) region. Within the initial linear region from 0 to 2.22 GPa, the Na-Mt tactoid exhibited a compression modulus of 125 GPa (**Fig. 4b**). The apparent non-linear higher modulus region (2.22-29.6 GPa) actually consists of three different linear zones. The compression moduli of Na-Mt tactoid was calculated as 237.5 GPa, 411 GPa and 643 GPa for the stress regions of 2.22-5.92 GPa, 5.92-14.8 GPa and 14.8-29.6 GPa respectively, as shown in **Table 5**. One of the previous studies from the author's group reported the elastic modulus of clay mineral layer as 696.55 GPa through constant-force SMD simulations [39]. Previously proposed bulk amorphization strength of montmorillonite (60 GPa) validated the compression strength of tactoid as of 29.6 GPa [20]. Moreover, modulus

mapping experiment of clay determined the elastic modulus as of 352.9 GPa, which was the same order of magnitude of derived modulus of tactoid (643 GPa) [39].

Table 4.5. Elastic moduli of a tactoid from compression simulations at various magnitudes of stress

Stress Region (GPa)	**Modulus (GPa)**
0-2.22	125
2.22-5.92	237.5
5.92-14.8	411
14.8-29.6	643

A snapshot of the Na-Mt tactoid model at 29.6 GPa of stress, having an aggregated d-value of 89.10 Å, is shown in **Fig. 5**.

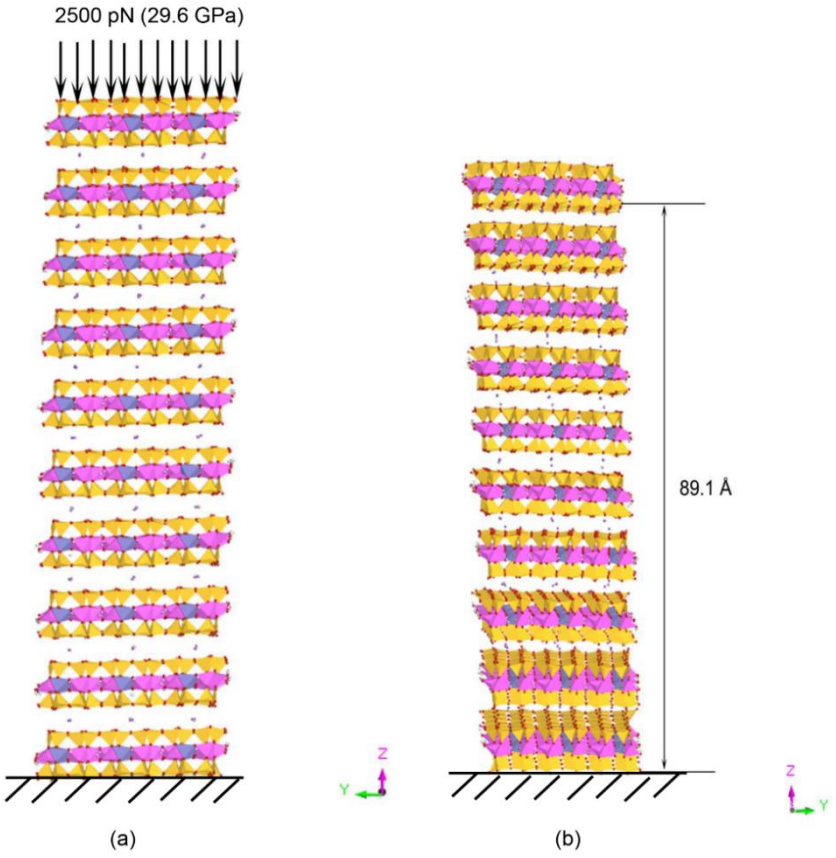

Figure 4.5. Na-Mt tactoid under the compression of 29.6 GPa (a) initial condition, and (b) final condition.

Table 4.6 shows the d-values between each pair of clay mineral layers inside Na-Mt tactoid under 29.6 GPa of compression.

Table 4.6. Calculated d-values between clay mineral layers inside Na-Mt tactoid under 29.6 GPa compressive stress

Clay mineral layer pair	**Equilibrated d-value (Å) under 29.6 GPa**	**Interlayer deformation (Å)**
B-C	10.02	0.77
C-D	9.95	0.84
D-E	9.94	0.86
E-F	9.91	0.89
F-G	9.89	0.91
G-H	9.87	0.92
H-I	9.87	0.93
I-J	9.86	0.94
J-K	9.8	0.95

The top clay mineral layer (H, I, J, and K) interlayers were deformed more, shown by lower d-values while bottom clay mineral layers (B, C, D, and E) were deformed relatively less, demonstrated by higher d-values. Hence the top portion of tactoid was more deformed than the bottom portion of tactoid under compression. The extent of interlayer deformation decreased successively from top to bottom along the tactoid. The minimum interlayer deformation (0.77 Å) occurred between the bottom clay mineral layers (B and C), while the maximum interlayer deformation (0.95 Å) took place between the top clay mineral layers (J and K).

4.3.3. Tensile Loading on Tactoid

The top clay mineral layers of equilibrated Na-Mt tactoid were pulled in the vertical direction to observe the tensile deformation of tactoid. The relationship between the pulling stress and displacement of top clay mineral layers is shown in **Fig 6.** In the case of pulling of only the topmost clay (K) layer, the layer was linearly displaced with increasing stress up to peak stress of

2.87 GPa, which occurred at a displacement of 8.15 Å. As the clay mineral layer moved beyond this distance, the pulling stress required to displace the top clay mineral layer started to reduce, indicating the detachment initiation of the top clay mineral layer from the rest of the tactoid. The interactions between the top clay mineral layer and the rest of the tactoid diminished linearly with the further displacement of the top layer (K).

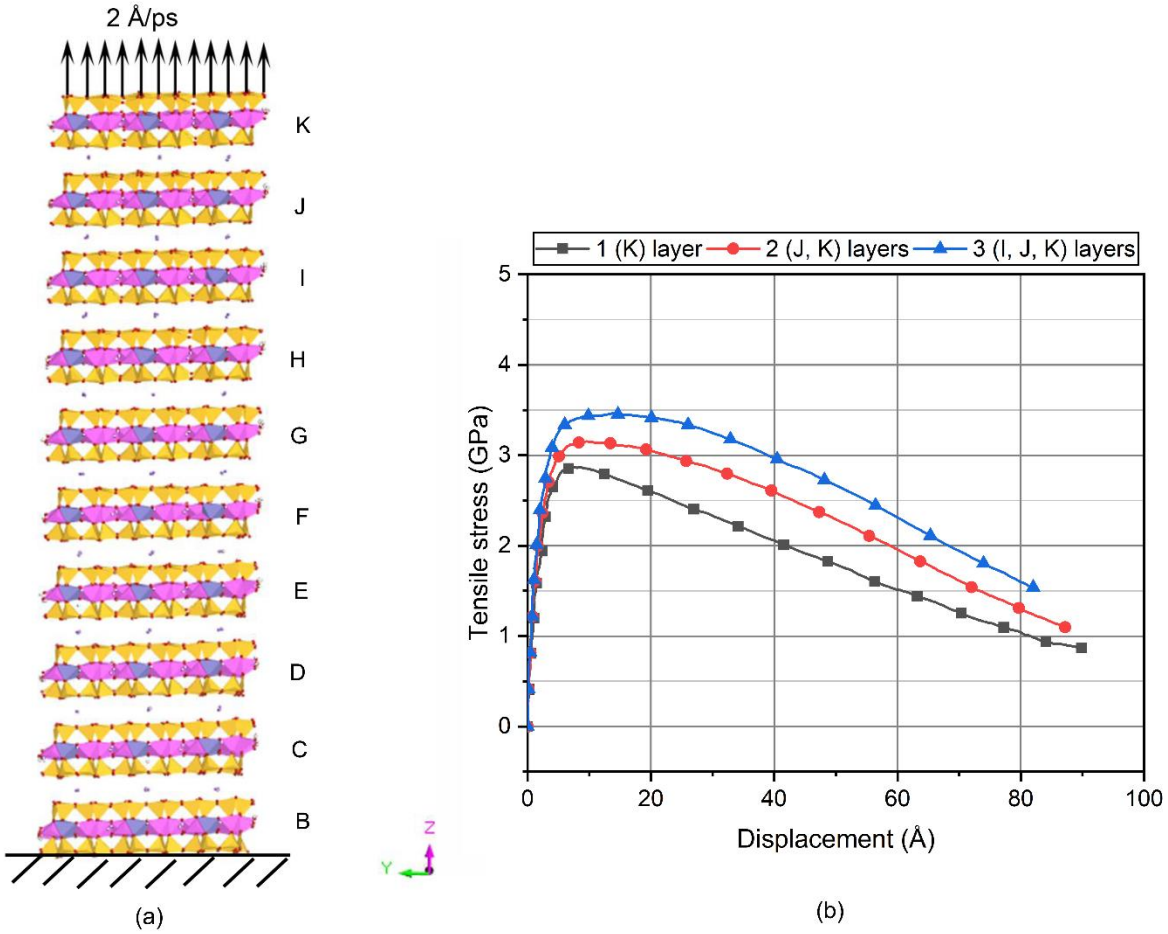

Figure 4.6. (a) Vertical tensile pulling of top clay mineral layer, and (b) tensile stress vs. displacement plot for pulling top 1 (K), 2 (J, K), and 3 (I, J, and K) clay mineral layers.

With the pulling of top two (J, K) and three (I, J, K) clay mineral layers, the peak stress occurred at a distance of about 8.1 Å, a magnitude similar to the case of pulling only the top clay mineral layer (**Fig. 6b**). However, with the increased number of clay mineral layers being pulled,

the magnitude of the maximum pulling stress increased (3.15 GPa for top two clay mineral layers and 3.45 GPa for top three clay mineral layers). The modulus values for tensile pulling of clay mineral layers from the tactoid were 133 GPa for one clay mineral layer (K), 170 GPa for pulling two clay mineral layers (J, K), and 187 GPa for pulling top three clay mineral layers (I, J, K). Beyond a displacement of 20 Å, we observed rigid body motion of the pulled clay mineral layers.

4.3.4. Shear Loading on Tactoid

The shearing of Na-Mt tactoid was performed by pulling the topmost clay mineral layer (K) along its horizontal direction while the bottom layer (B) was fixed (**Fig. 7**). Upon the application of a constant horizontal pulling velocity of 2 Å/ps, we observed a compound response from the clay tactoid.

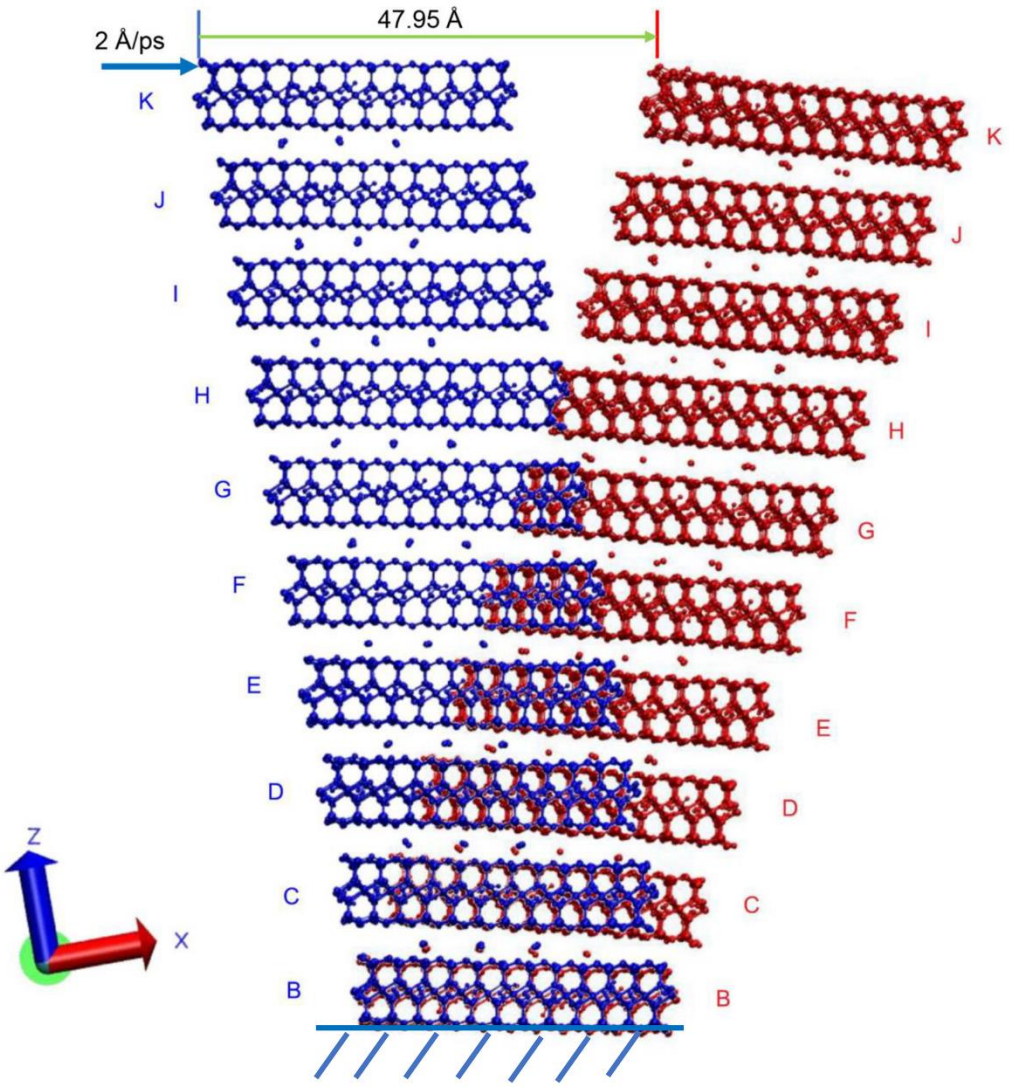

Figure 4.7. Shearing of individual Na-Mt tactoid. The initial (left) and final (right) model of tactoid due to shearing is superimposed to indicate the extent of shearing.

During the application of shear stress on Na-Mt tactoid, the clay mineral layers slided relative to each other except the fixed bottom layer (layer B). This gradual sliding of clay mineral layers deformed the top portion of tactoid the most and bottom portion least, i.e., the distance traveled by the top and bottom clay mineral layers were the largest and smallest, respectively. At the end of the shearing, the top clay mineral layer (K) was horizontally displaced by a total of

47.95 Å to its initial position (Fig. 7). The shear stress-displacement plot of top clay mineral layer K is presented in Fig. 8.

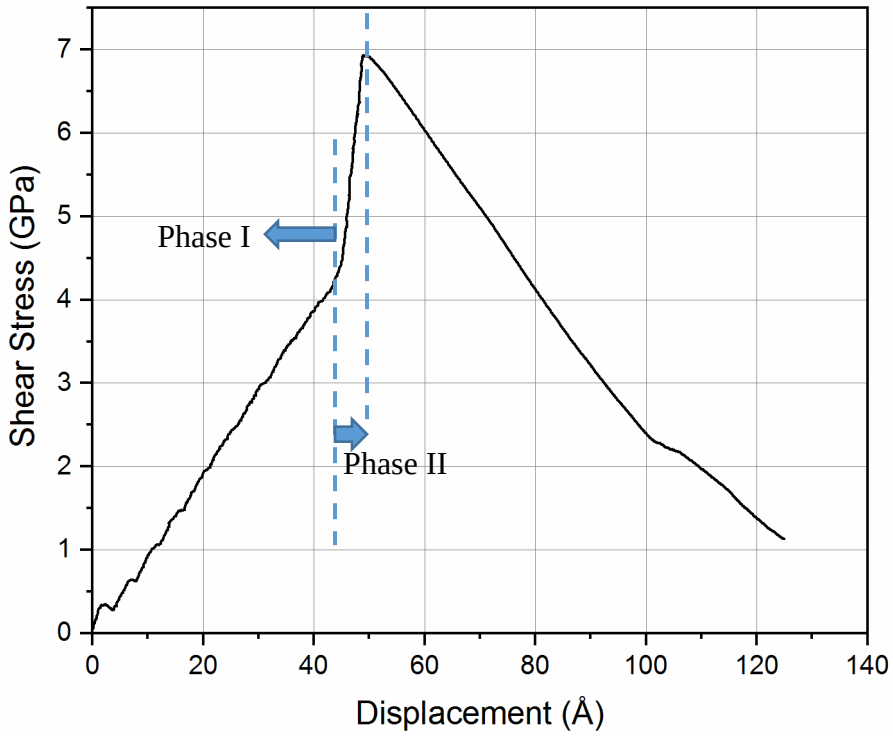

Figure 4.8. Shear stress vs. displacement plot for the top clay (K) layer upon horizontal shearing of tactoid.

This plot displays two distinct phases of tactoid response (sliding of clay mineral layers) due to applied shear stress. The displacement of top clay mineral layer K from 0 Å to 44.8 Å reflected the first phase of the response. During this phase, the stress increased from 0 GPa to 4.2 GPa. The average horizontal relative displacement between each of the clay mineral layers at the end of this phase was about 4.98 Å. After the "first shearing" phase, as the shearing continued, very small (0.30 Å), relative displacements (second phase) between the clay mineral layers were observed. In the shear stress-displacement plot (**Fig. 8**), the second phase was reflected by the steep response of curve where the shear stress sharply increased from 4.2 GPa to 6.87 GPa over the displacement of 3.15 Å of the top clay mineral layer K. Upon further application of constant

horizontal velocity to the top clay mineral layer K, the magnitude of shear stress dropped almost linearly characterizing the complete separation of the top clay mineral layer from the tactoid. During the entire shearing of tactoid, the second from bottom clay mineral layer (C) moved by 5.26 Å in the horizontal direction while the topmost layer (K) moved by 47.95 Å. The total displacement of the top clay mineral layer (layer K) was transformed into the shear strain of tactoid considering the equilibrium height (104.04 Å) of the clay tactoid. Based on the observed response, two different shear moduli for the tactoid, representing the two phases of observed shear deformations, were computed as 10 GPa and 69.56 GPa, respectively. The first phase of shearing (0-4.2 GPa) was the initial elastic shearing of tactoid as the withdrawal of stress brought back the initial tactoid conformation. The second phase of shearing (4.2-6.87 GPa) indicated the transition from elastic to plastic deformation, where the interlayer cations that moved into the tetragonal cavities resisted the deformation through strong binding interactions with clay mineral layers inside the tactoid. During the first phase of shearing, the clay mineral layers smoothly slided past each other as the interlayer cations moved along the direction of shear. This translation of cations allowed the gradual sliding of clay mineral layers from top (K) to bottom (C). The cations in the bottom interlayer space (between B and C layers) got locked up in the tetrahedral cavities after certain amount of translation which in turn initiated the locking of the sliding phenomena of clay mineral layers. Eventually, the clay mineral layers gradually got locked up from bottom (C) to top layer (K) along the tactoid. This specific locking incidence specified the steep shearing response of tactoid. However, the continuous pulling of top clay mineral layer K eventually overcame its binding with the rest of the tactoid and thereby caused its separation.

4.4. Discussion

Inside a Na-Mt tactoid, attractive short-range (Van der Waals) interactions were predominant between clay mineral layers. The attractive interactions between interlayer sodium cations and clay mineral layers are dictated by long-range (electrostatic) interactions. The clay mineral layers located at the two extremes of tactoid (top and bottom) were more loosely bonded as observed by their lesser interactions with interlayer cations. These interactions can be compared with experimental data by computing energy per surface area for both the pairwise interactions between the clay mineral layers (**Table 2**) and their total binding interactions with rest of the tactoid (**Table 4**). The computed values of energy per surface area, 311 mJ/m^2 for pairwise interactions and 928 mJ/m^2 for total binding interactions for a clay mineral layer within the tactoid are found. These values are of the same order of magnitude as the previously reported values i.e. 424 and 205 mJ/m^2 [40]. The maximum compression of Na-Mt tactoid was observed to be dependent upon the Van der Waals radii of tetrahedral oxygen atoms of clay mineral layers. During the compression of tactoid, interlayer Na cations occupied the ditrigonal cavities of clay mineral layers. The Van der Waals radii of tetrahedral oxygen atoms were found to be 1.6 Å. Thus, the effective interlayer space was computed as 0.95 Å by subtracting the VDW radius (1.6 Å) of tetrahedral oxygen atoms from both sides of the interlayer space (4.15 Å) (**Fig. 9**). The compression of Na-Mt tactoid was the result of the reduction in interlayer space. The maximum possible decrease in interlayer space was equal to the magnitude of effective interlayer space (0.95 Å) (**Table 5)**.

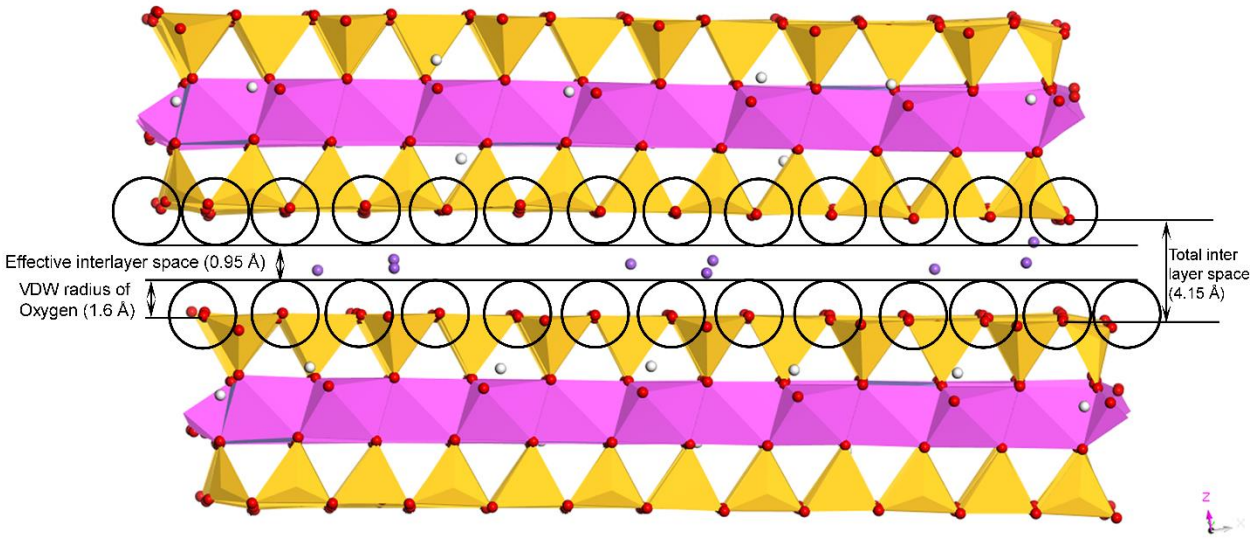

Figure 4.9. Effective interlayer space of Na-Mt clay structure.

In the current study, multiple values of compression moduli were found during the clay tactoid compression. The lower modulus region (0-2.22 GPa) characterized the initial compression of the interlayer space. On the other hand, the higher modulus region (>2.22-29.6 GPa) reflected the full compression of effective interlayer space and the beginning of the compression of the clay mineral layers. The aggregated d-value of Na-Mt tactoid was reduced by 8.1 Å under maximum compression (29.6 GPa). The tensile pulling of a different number of top clay mineral layers caused the permanent separation of them from tactoid beyond 8.15 Å displacement and almost nullified Na-clay mineral layer attractive interactions. In the case of tensile pulling of Na-Mt tactoid, the magnitude of required pulling stress to completely detach the clay mineral layer gradually increased with the number of pulled clay mineral layers. Pulling off one clay mineral layer (K) required 2.87 GPa while pulling off three clay mineral layers (I, J, K) required 3.45 GPa as three clay mineral layers had higher non-bonded interactions with the rest of the tactoid compared to just one clay mineral layer with rest of tactoid. Horizontal shearing of the tactoid consisted of two distinct phases. In the first phase, each clay mineral layer slided to the same extent 4.98 Å and got locked. The locking was consecutive from the top layer to the bottom layer. The locking occurred

when the sodium ions migrated to tetragonal cavities of the clay mineral layers. When the bottom layer was locked, it represented the end of phase I. With the continued increase in shear stress, a very small shear deformation of 0.3 Å was observed. This phase was represented by a much steeper slope than Phase I. The two phases exhibited moduli of 10 GPa and 69.56 GPa, respectively. Further, phase II ended when the top clay mineral layer was separated from the tactoid resulting in the negative slope of the plot, as shown in Figure 8.

4.5. Conclusions

In the present study, for the first time, a molecular model of Na-Mt tactoid is developed and subjected to mechanical loading. We have used molecular dynamics (MD) simulations to evaluate the interaction energies within the equilibrated tactoid and steered molecular dynamics (SMD) simulations to investigate the mechanical response of Na-Mt tactoid under compression, tension, and shear loadings. The key findings are summarized below:

1. In equilibrated Na-Mt tactoid, pairwise non-bonded interactions between the clay mineral layers are identical and attractive in nature. The attractive Van der Waals interactions dominate the total energy as compared to the repulsive electrostatic interactions between the clay mineral layers. The non-bonded interaction energies between the interlayer cations and clay mineral layers at the top and bottom of the tactoid are almost one-half of the interaction energies between cations and other clay mineral layers.
2. During compression, the maximum deformation of the interlayer space is significantly less than the d-value of the interlayer and is dictated by the effective interlayer space.

3. Pulling off an increasing number of clay mineral layers from tactoid requires larger stress.
4. The application of the shear stress to the top of the tactoid results in a relatively uniform displacement of clay mineral layers, with respect to the adjacent layers.
5. Compressive, tensile, and shear moduli of the clay tactoid are computed, and the values depend upon the magnitude of applied stress.

4.6. Acknowledgments

The authors would like to acknowledge the financial support from the US Department of Transportation Mountain-Plains Consortium (MPC) Grant #: DTRT13-G-UTC38. We also recognize the computational facilities provided by North Dakota State University Computationally Assisted Science and Technology (CCAST), and the support of NSF MRI grant #1229316 and NSF OIA NDACES-1946202.

4.7. References

(1) Grim, R. E. Clay Mineralogy McGraw-Hill. *New York* **1968**, 206.

(2) Guggenheim, S.; Martin, R. T. Reply to the comment by DM Moore on "Definition of clay and clay mineral: Joint report of the AIPEA nomenclature and CMS nomenclature committees". *Clays and Clay Minerals* **1996**, *44* (5), 713-715.

(3) Grim, R. E. Modern concepts of clay materials. *Journal of Geology* **1942**, *50* (3), 225-275. DOI: 10.1086/625050.

(4) Pradhan, S. M.; Katti, K. S.; Katti, D. R. Evolution of Molecular Interactions in the Interlayer of Na-Montmorillonite Swelling Clay with Increasing Hydration. *International Journal of Geomechanics* **2015**, *15* (5). DOI: 10.1061/(asce)gm.1943-5622.0000412.

(5) Katti, D.; Shanmugasundaram, V. Influence of swelling on the microstructure of expansive clays. *Can. Geotech. J.* **2001**, *38* (1), 175-182. DOI: 10.1139/cgj-38-1-175.

(6) Katti, D. R.; Matar, M. I.; Katti, K. S.; Amarasinghe, P. M. Multiscale modeling of swelling clays: A computational and experimental approach. *Ksce Journal of Civil Engineering* **2009**, *13* (4), 243-255. DOI: 10.1007/s12205-009-0243-0.

(7) Massat, L.; Cuisinier, O.; Bihannic, I.; Claret, F.; Pelletier, M.; Masrouri, F.; Gaboreau, S. Swelling pressure development and inter-aggregate porosity evolution upon hydration of a compacted swelling clay. *Applied Clay Science* **2016**, *124*, 197-210. DOI: 10.1016/j.clay.2016.01.002.

(8) Jullien, M.; Raynal, J.; Kohler, E.; Bildstein, O. Physicochemical reactivity in clay-rich materials: Tools for safety assessment. *Oil & Gas Science and Technology-Revue D Ifp Energies Nouvelles* **2005**, *60* (1), 107-120. DOI: 10.2516/ogst:2005007.

(9) Perdrial, J. N.; Warr, L. N. Hydration behavior of MX80 bentonite in a confined-volume system: implications for backfill design. *Clays and Clay Minerals* **2011**, *59* (6), 640-653. DOI: 10.1346/ccmn.2011.0590609.

(10) Salles, F.; Beurroies, I.; Bildstein, O.; Jullien, M.; Raynal, J.; Denoyel, R.; Van Damme, H. A calorimetric study of mesoscopic swelling and hydration sequence in solid Na-montmorillonite. *Applied Clay Science* **2008**, *39* (3-4), 186-201. DOI: 10.1016/j.clay.2007.06.001.

(11) Jonas, E. C.; Oliver, R. M. Size and shape of montmorillonite crystallites. *Clays Clay Miner* **1967**, *15*, 27-33.

(12) Pusch, R.; Karnland, O.; Hökmark, H. *GMM-a general microstructural model for qualitative and quantitative studies of smectite clays*; Swedish Nuclear Fuel and Waste Management Co., 1990.

(13) Neaman, A.; Pelletier, M.; Villieras, F. The effects of exchanged cation, compression, heating and hydration on textural properties of bulk bentonite and its corresponding purified montmorillonite. *Applied Clay Science* **2003**, *22* (4), 153-168. DOI: 10.1016/s0169-1317(02)00146-1.

(14) Lloret, A.; Villar, M. V.; Sanchez, M.; Gens, A.; Pintado, X.; Alonso, E. E. Mechanical behaviour of heavily compacted bentonite under high suction changes. *Geotechnique* **2003**, *53* (1), 27-40. DOI: 10.1680/geot.53.1.27.37258.

(15) Chang, F. R. C.; Skipper, N. T.; Sposito, G. Computer-simulation of interlayer molecular-structure in sodium montmorillonite hydrates. *Langmuir* **1995**, *11* (7), 2734-2741. DOI: 10.1021/la00007a064.

(16) Teppen, B. J.; Rasmussen, K.; Bertsch, P. M.; Miller, D. M.; Schafer, L. Molecular dynamics modeling of clay minerals .1. Gibbsite, kaolinite, pyrophyllite, and beidellite. *Journal of Physical Chemistry B* **1997**, *101* (9), 1579-1587. DOI: 10.1021/jp961577z.

(17) Thapa, K. B.; Katti, K. S.; Katti, D. R. Compression of Na-Montmorillonite Swelling Clay Interlayer Is Influenced by Fluid Polarity: A Steered Molecular Dynamics Study. *Langmuir : the ACS journal of surfaces and colloids* **2020**. DOI: 10.1021/acs.langmuir.0c01412.

(18) Rahromostaqim, M.; Sahimi, M. Molecular Dynamics Simulation of Hydration and Swelling of Mixed-Layer Clays. *Journal of Physical Chemistry C* **2018**, *122* (26), 14631-14639. DOI: 10.1021/acs.jpcc.8b03693.

(19) Fu, Y. T.; Zartman, G. D.; Yoonessi, M.; Drummy, L. F.; Heinz, H. Bending of Layered Silicates on the Nanometer Scale: Mechanism, Stored Energy, and Curvature Limits. *Journal of Physical Chemistry C* **2011**, *115* (45), 22292-22300. DOI: 10.1021/jp208383f.

(20) Zartman, G. D.; Liu, H.; Akdim, B.; Pachter, R.; Heinz, H. Nanoscale Tensile, Shear, and Failure Properties of Layered Silicates as a Function of Cation Density and Stress. *Journal of Physical Chemistry C* **2010**, *114* (4), 1763-1772. DOI: 10.1021/jp907012w.

(21) Katti, D. R.; Ghosh, P.; Schmidt, S.; Katti, K. S. Mechanical properties of the sodium montmorillonite interlayer intercalated with amino acids. *Biomacromolecules* **2005**, *6* (6), 3276-3282. DOI: 10.1021/bm0503219.

(22) Katti, D. R.; Schmidt, S. R.; Ghosh, P.; Katti, K. S. Modeling the response of pyrophyllite interlayer to applied stress using steered molecular dynamics. *Clays and Clay Minerals* **2005**, *53* (2), 171-178. DOI: 10.1346/ccmn.2005.0530207.

(23) Schmidt, S. R.; Katti, D. R.; Ghosh, P.; Katti, K. S. Evolution of mechanical response of sodium montmorillonite interlayer with increasing hydration by molecular dynamics. *Langmuir* **2005**, *21* (17), 8069-8076. DOI: 10.1021/la050615f.

(24) Katti, D. R.; Schmidt, S. R.; Ghosh, P.; Katti, K. S. Molecular modeling of the mechanical behavior and interactions in dry and slightly hydrated sodium montmorillonite interlayer. *Can. Geotech. J.* **2007**, *44* (4), 425-435. DOI: 10.1139/t06-127.

(25) Katti, D. R.; Patwary, Z. R.; Katti, K. S. Modelling clay-fluid interactions in montmorillonite clays. *Environmental Geotechnics* **2017**, *4* (5), 322-338. DOI: 10.1680/jenge.14.00027. Katti, D. R.; Thapa, K. B.; Katti, K. S.

(26) Van Olphen, H.; Fripiat, J.-J. *Data handbook for clay materials and other non-metal lic minerals: providing those involved in clay research and indus trial application with sets of*

authoritative data describing the physical and chemical properties and mineralogical composition of the available reference materials; 1979.

(27) Skipper, N. T.; Chang, F. R. C.; Sposito, G. Monte-carlo simulation of interlayer molecular-structure in swelling clay-minerals .1. methodology. *Clays and Clay Minerals* **1995**, *43* (3), 285-293. DOI: 10.1346/ccmn.1995.0430303.

(28) Katti, D. R.; Thapa, K. B.; Katti, K. S. Modeling molecular interactions of sodium montmorillonite clay with 3D kerogen models. *Fuel* **2017**, *199*, 641-652. DOI: 10.1016/j.fuel.2017.03.021.

(29) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(30) Mishra, R. K.; Mohamed, A. K.; Geissbuhler, D.; Manzano, H.; Jamil, T.; Shahsavari, R.; Kalinichev, A. G.; Galmarini, S.; Tao, L.; Heinz, H.; et al. cemff: A force field database for cementitious materials including validations, applications and opportunities. *Cement and Concrete Research* **2017**, *102*, 68-89. DOI: 10.1016/j.cemconres.2017.09.003.

(31) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials* **2014**, *26* (8), 2647-2658. DOI: 10.1021/cm500365c.

(32) Heinz, H.; Koerner, H.; Anderson, K. L.; Vaia, R. A.; Farmer, B. L. Force field for mica-type silicates and dynamics of octadecylammonium chains grafted to montmorillonite. *Chemistry of Materials* **2005**, *17* (23), 5658-5669. DOI: 10.1021/cm0509328.

(33) Heinz, H.; Suter, U. W. Atomic charges for classical simulations of polar systems. *Journal of Physical Chemistry B* **2004**, *108* (47), 18341-18352. DOI: 10.1021/jp048142t.

(34) Katti, K. S.; Sikdar, D.; Katti, D. R.; Ghosh, P.; Verma, D. Molecular interactions in intercalated organically modified clay and clay-polycaprolactam nanocomposites: Experiments and modeling. *Polymer* **2006**, *47* (1), 403-414. DOI: 10.1016/j.polymer.2005.11.055.

(35) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289.

(36) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative minimization techniques for abinitio total-energy calculations - molecular-dynamics and conjugate gradients. *Reviews of Modern Physics* **1992**, *64* (4), 1045-1097. DOI: 10.1103/RevModPhys.64.1045.

(37) Amarasinghe, P. M.; Katti, K. S.; Katti, D. R. Molecular Hydraulic Properties of Montmorillonite: A Polarized Fourier Transform Infrared Spectroscopic Study. *Applied Spectroscopy* **2008**, *62* (12), 1303-1313.

(38) Lin, J. J.; Cheng, I. J.; Wang, R. C.; Lee, R. J. Tailoring basal spacings of montmorillonite by poly(oxyalkylene)diamine intercalation. *Macromolecules* **2001**, *34* (26), 8832-8834. DOI: 10.1021/ma011169f.

(39) Sikdar, D.; Pradhan, S. M.; Katti, D. R.; Katti, K. S.; Mohanty, B. Altered phase model for polymer clay nanocomposites. *Langmuir* **2008**, *24* (10), 5599-5607. DOI: 10.1021/la800583h.

(40) Douillard, J. M.; Salles, F.; Devautour-Vinot, S.; Manteghetti, A.; Henry, M. Study of the surface energy of montmorillonite using PACHA formalism. *Journal of Colloid and Interface Science* **2007**, *306* (1), 175-182. DOI: 10.1016/j.jcis.2006.09.008.

5. AN IN-SILICO INVESTIGATION OF HYDRATED NA-MONTMORILLONITE TACTOID[4]

5.1. Introduction

Swelling clay mineral, an integral component of expansive soil, has long been investigated due to its unique swelling behavior upon hydration. When hydrated, swelling clay applies swelling pressure to civil infrastructures resulting in the damage of buildings, dams, bridges etc.[1] Swelling clays are found all over the United States and cause an average annual damage cost of 9 billion USD.[2] Swelling clays are extensively used as landfill liners due to their high specific surface area, and low permeability.[3] Besides the geotechnical prospect, swelling clays find applications in polymer clay nanocomposites, pharmaceutical, biomedical, and cosmetic industries.[4, 5] [5] The swelling behavior of clays is heavily exploited in all these applications.

Na-montmorillonite (Na-MMT) is one of the most commonly found swelling clay minerals which belong to the smectite group. Na-MMT is a 2:1 phyllosilicate mineral made up of two silica tetrahedral sheets and one alumina octahedral sheet.[6] [7] The clay sheets get negatively charged due to the isomorphous substitution of ions in the octahedral layer compensated by interlayer cations.[8] The clay interlayer properties vary significantly based on the interlayer cations. Na-MMT clay consists of a hierarchical structure at the molecular level. [9] A series of experimental studies exploring clay porosities, clay-rich soil hydration, and hydration sites confirmed the multiscale structure of clays.[10] [11] [12] [13, 14] [15] The hierarchical clay structure contains clay unit layer (t-o-t sheet), tactoid (particle containing several clay layers stacked in vertical direction), aggregate (several tactoids gathered in random orientation), and multi-aggregate (several aggregates gathered). This multi-scale structure houses three layers of porosities i.e., interlayer porosity (within tactoid), inter-

[4] The contents of this chapter are going to be submitted for publication.

tactoid porosity (within aggregate), and inter-aggregate porosity (within multi-aggregate). During clay hydration, clay interlayers contain 86% water while other porosities contain 14% water. [13] Therefore, investigating interlayer hydration behavior carries great important to better explain clay hydration at the molecular level.

Na-MMT clay tactoid is the basic clay particle that contains interlayer porosities. Ten identical clay layers are vertically aligned to form a tactoid. The clay layers of tactoid interact with each other and with interlayer cations. [9] The nanomechanical properties of tactoid depend significantly on interlayer properties. Both experimental and modeling studies showed that clay interlayer spacing increases with increasing hydration. [16] [17] The impact of increasing interlayer hydration on tactoid interactions and mechanical behavior needs to be investigated. Molecular dynamics simulations can be utilized in this regard.

Molecular dynamics (MD) simulation is a computational technique that utilizes molecular mechanics to determine the conformation and dynamics of a molecular system. The atoms are treated as classical spheres, while the bonds are treated as springs. A force field is used to compute the energy of the system. [18] Steered molecular dynamics (SMD) can be employed to investigate the mechanical response of a molecular system. [19] Both MD and SMD simulations have been employed to investigate different materials systems, including oil shale, proteins, coronaviruses, polymer clay nanocomposites etc.[20] [21] [22] [23] [24] [25] [26] Different properties of clay minerals at various environmental conditions have been extensively explored using MD simulations. [27] [28] [29] [30] [31] [32] [33] [34] [35] To examine the hydration mechanism of clay minerals, MD simulations have also been utilized. [36] [37] [38] However, these studies primarily focused on the clay bi-layer systems without taking into consideration the clay hierarchical structure (tactoid, aggregate etc.). In the previous study from Katti research group, the authors inspected the dry Na-MMT clay tactoid behavior

utilizing both MD and SMD simulations.[9] As a continuation of the earlier study, the present study addresses the clay tactoid behavior and nanomechanical properties in hydrated condition. Na-MMT clay properties are determined for three different amounts of interlayer hydration i.e., 10%, 20% and 30% (by weight). This study probes how the clay tactoid d-spacing, interlayer interactions, and compression behavior changes with increasing hydration.

5.2. Methodology

The molecular model of Na-MMT tactoid has been obtained from the previous study. 9 The unit clay layer of Na-MMT was built based on Swy-2 clay with the simplified structural formula of NaSi16(Al6FeMg)O20(OH)4.35 The utilized clay model is 6 × 3 unit cells long in the XY direction with the unit cell dimensions of 5.28 Å × 9.14 Å × 6.56 Å. As the isomorphous substitution of octahedral cations produces a charge of -0.5e per unit cell, nine interlayer Na cations are placed in the interlayer. A tactoid model is built by vertically placing 10 clay layers on top of each other. The 10 clay layers are termed as B, C, D, E, F, G, H, I, J, and K layer from bottom to top respectively. The distance between the bottom and top clay layer is termed as the 'aggregated d-value'. This tactoid contains 9 interlayers and each of them houses 9 interlayer cations. Then water molecules are placed in each interlayer for 10%, 20%, and 30% interlayer hydration (by weight). The total number of water molecules corresponding to 10%, 20%, and 30% (by weight) hydration is 72, 144, and 216, respectively. The water molecules are represented by TIP3P (Transferable Intermolecular Potential 3 Point) model. With increasing amounts of hydration, the initial dimension of tactoid changes only in the Z-direction. The initial size of dry tactoid is 31.68 Å × 27.44 Å × 96.56 Å. The initial size of hydrated tactoids remains the same in XY directions. The vertical height of 10%, 20%, and 30% hydrated tactoids are 114.56 Å, 132.56 Å, and 168.56 Å. The initial magnitude of aggregated d-value for dry, 10%, 20%, and 30% hydrated

tactoid are 90 Å, 108 Å, 126 Å, and 162 Å, respectively. Materials Studio 7.0 was used to construct the clay model, and VMD 1.9.3 was used to add the TIP3P water molecules. CHARMm force field is utilized to parameterize the clay tactoid model, and the parameters are taken from the literature. 39 40 41 These clay parameters are consistent with other force field parameters of clay. 9 42 Previous studies from our group have also validated the clay parameters with respect to the experimental results. 43 41 25 The current study performs the MD and SMD simulations of hydrated tactoids with 3 different amounts of interlayer hydration (10%, 20%, and 30% by weight). The MD simulations are performed to minimize and equilibrate the hydrated tactoid models. The minimization was performed at the vacuum condition, while the equilibration was performed at normal temperature (300 K) and pressure (1.01325 bar) conditions. The equilibration was performed in steps where the temperature was first raised from 0 K to 300 K, and then the pressure was increased from 0 bar to 1.01325 bar. All the models are equilibrated at NTP (normal temperature and pressure) conditions for 2 ns with a timestep of 0.5 fs. The SMD simulations were performed on equilibrated hydrated tactoids to explore the compression behavior of clay tactoids with increasing interlayer hydration. Constant velocity SMD was employed where constant velocity was applied on top clay layer while keeping the bottom clay layer fixed. The constant velocity parameters (k = 9 kcal/mol/Å2 and v = 2 Å/ps) have been obtained from the previous study. 9 NAMD simulation package was used to perform all the MD and SMD simulations. 44 NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. All the simulations were run on Center for Computationally Assisted Science and Technology (CCAST), a parallel computing facility at North Dakota State University.

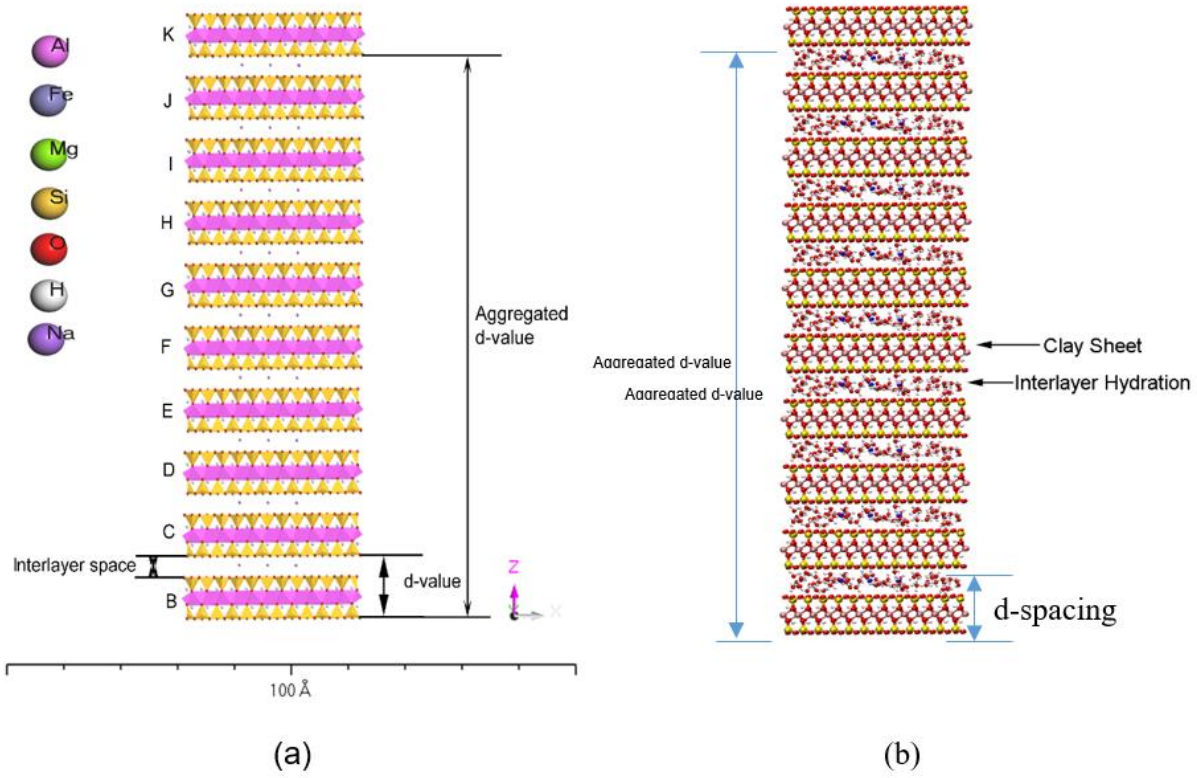

Figure 5.1. (a) Dry Na-MMT tactoid, and (b) hydrated Na-MMT tactoid.

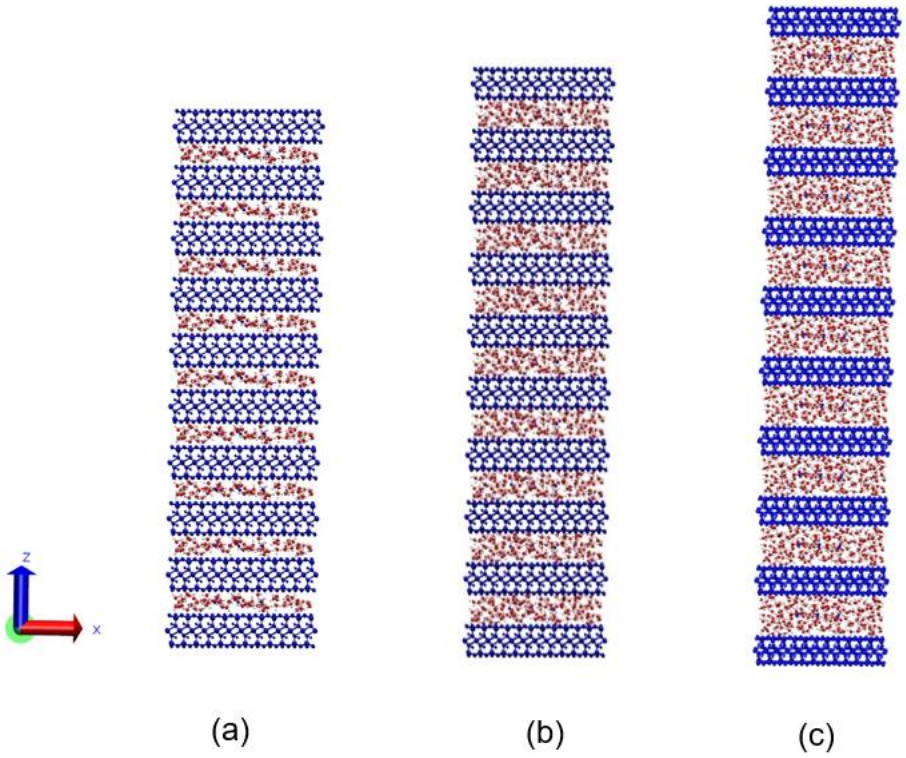

Figure 5.2. Na-MMT clay tactoid with (a) 10%, (b) 20%, and (c) 30% interlayer hydration.

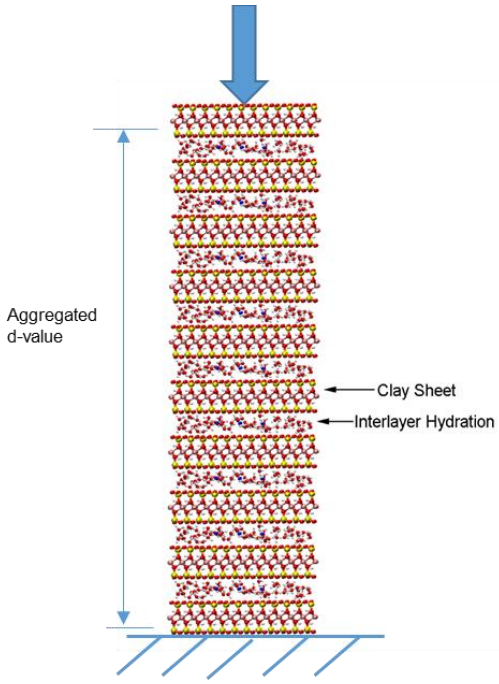

Figure 5.3. Schematic diagram of compression on hydrated tactoid.

5.3. Results and Discussion

The equilibrated d-spacing values and interaction energies among different entities within 10%, 20%, and 30% interlayer hydrated tactoids are given below:

Table 5.1. Equilibrated d-spacing for 10% hydrated tactoid

Clay layers	**d-spacing (Å)**
BC	13.09
CD	13.12
DE	13.12
EF	13.09
FG	13.14
GH	13.12
HI	13.09
IJ	13.12
JK	13.13
Average	13.11

Table 5.2. Interaction energies with equilibrated 10% hydrated tactoid

Clay components	**Electrostatic (kcal/mol)**	**Van der Waals (VDW) (kcal/mol)**	**Interaction energies (kcal/mol)**
Tactoid-water	-10788.1	-1017.3	-11805.4
Na-Water	-10657.8	599.9	-10057.9
Clay sheets-water	-130.3	-1617.2	-1747.5
Na-clay sheets	-10330.5	-155.1	-10485.6

Table 5.3. Equilibrated d-spacing for 20% hydrated tactoid

Clay layers	**d-spacing (Å)**
BC	14.37
CD	14.39
DE	14.65
EF	14.45
FG	14.43
GH	14.36
HI	14.37
IJ	14.52
JK	14.50
Average	14.45

Table 5.4. Interaction energies with equilibrated 20% hydrated tactoid

Clay components	**Electrostatic (kcal/mol)**	**Van der Waals (VDW) (kcal/mol)**	**Interaction energies (kcal/mol)**
Tactoid-water	-16181.8	-2279.4	-18461.2
Na-Water	-14638.2	652.5	-13985.7
Clay sheets-water	-1543.6	-2931.9	-4475.5
Na-clay sheets	-9487.9	-130.6	-9618.5

Table 5.5. Equilibrated d-spacing for 30% hydrated tactoid

Clay layers	**d-spacing (Å)**
BC	16.02
CD	16.03
DE	16.05
EF	16.02
FG	16.06
GH	15.99
HI	16.05
IJ	16.12
JK	16.05
Average	16.04

Table 5.6. Interaction energies with equilibrated 30% hydrated tactoid

Clay components	**Electrostatic (kcal/mol)**	**Van der Waals (VDW) (kcal/mol)**	**Interaction energies (kcal/mol)**
Tactoid-water	-18247.0	-2742.9	-20989.9
Na-Water	-16353.1	650.8	-15702.3
Clay sheets-water	-1893.9	-3393.7	-5287.6
Na-clay sheets	-8591.7	-112.4	-8704.1

Therefore, the aggregated d-value of 10%, 20%, and 30% hydrated tactoids are 118 Å, 130 Å, and 144.4 Å, respectively. Increasing hydration increases the average d-value within tactoid, which in turn results in increasing aggregated d-value. The interactions magnitude within clay tactoid changes with increasing hydration without changing its pattern. The total tactoid-water interaction energies increase with the increasing amount of hydration. The majority of the tactoid-

water interaction energies occur through Na-water interactions. Na-water interaction is mostly electrostatic in nature.

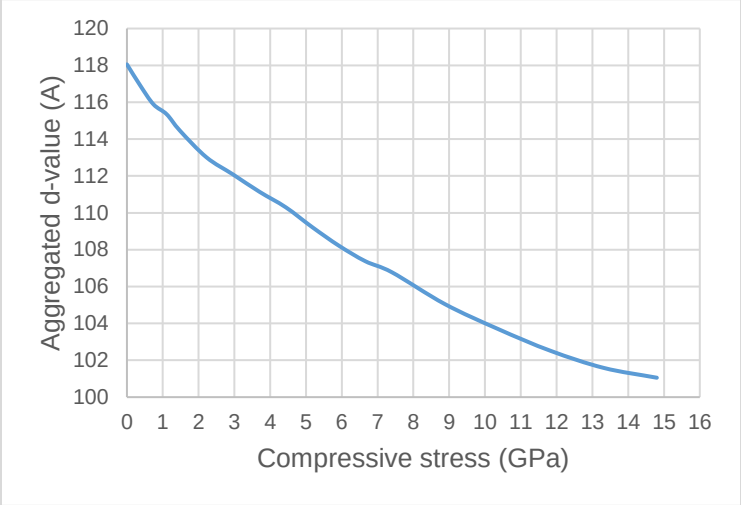

Figure 5.4. Compression mechanism of 10% hydrated tactoid.

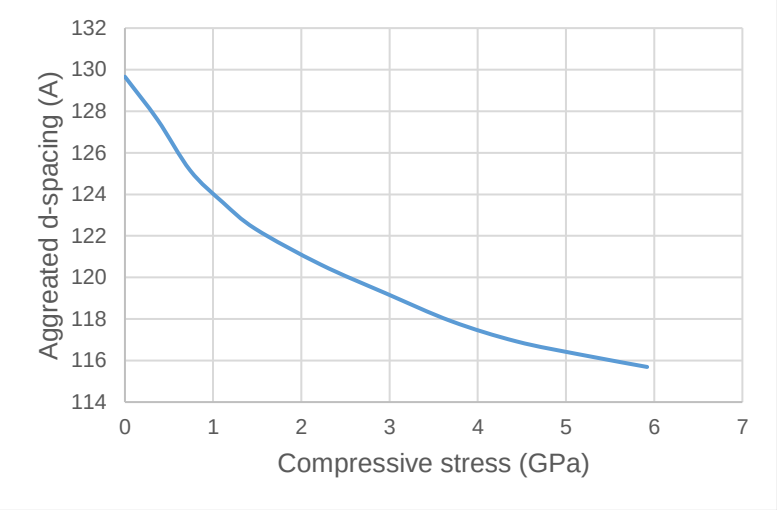

Figure 5.5. Compression mechanism of 20% hydrated tactoid.

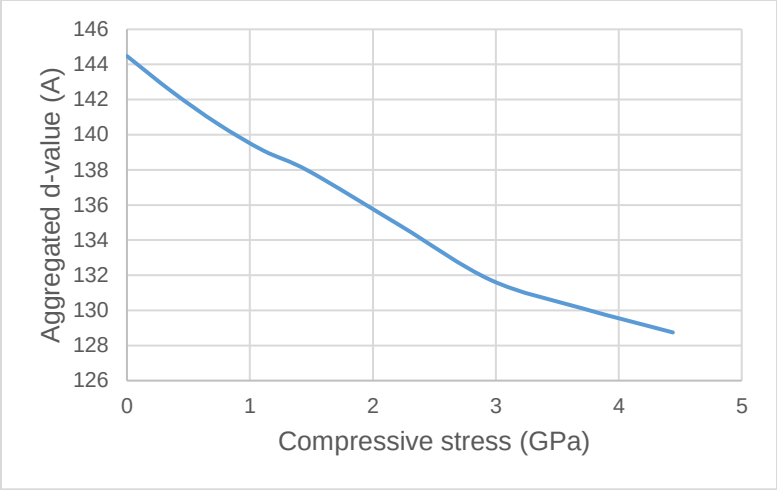

Figure 5.6. Compression mechanism of 30% hydrated tactoid.

Constant-velocity SMD simulations were performed on hydrated tactoids to determine their compression behavior. The constant-velocity SMD applies force on the top clay layer resulting in the reduction of aggregated d-value. With increasing hydration, the compressive strength of tactoid decreases as the compressive strength for 10%, 20%, and 30% hydrated tactoid are 14.8 GPa, 5.9 GPa, and 4.5 GPa. At the maximum compression, the interlayer cations and water molecules start overlapping their VDW radius. In all hydrated cases, the compression strength is much lower than dry tactoid compressive strength, i.e., 29.6 GPa.

5.4. Conclusions

This study performs MD and SMD simulations on three different amounts of hydrated Na-MMT clay tactoid to observe their interactions and nanomechanical behavior. The major findings are:

1. Increasing hydration results in increasing average d-value and aggregated d-value.
2. With increasing hydration, tactoid-water interactions get increased while Na-clay interactions get decreased.
3. Tactoid-water interactions are primarily electrostatic.

4. Increasing hydration reduces the compressive strength of the tactoid.

5.5. References

(1) Buzzi, O.; Fityus, S.; Sloan, S. W. Use of expanding polyurethane resin to remediate expansive soil foundations. *Canadian Geotechnical Journal* **2010**, *47* (6), 623-634. DOI: 10.1139/t09-132.

(2) Pérez, J. C.; Ruge, J. C.; Rojas, J. P.; Orjuela, J. A.; Cáceres, J. R. Experimental assessment of swelling potential in montmorillonite soils. In *Journal of Physics: Conference Series*, 2021; IOP Publishing: Vol. 1938, p 012006.

(3) Thapa, K. B.; Katti, K. S.; Katti, D. R. Compression of Na-Montmorillonite Swelling Clay Interlayer Is Influenced by Fluid Polarity: A Steered Molecular Dynamics Study. *Langmuir : the ACS journal of surfaces and colloids* **2020**. DOI: 10.1021/acs.langmuir.0c01412.

(4) Ambre, A.; Katti, K. S.; Katti, D. R. In situ mineralized hydroxyapatite on amino acid modified nanoclays as novel bone biomaterials. *Materials Science and Engineering: C* **2011**, *31* (5), 1017-1029.

(5) Katti, D. R.; Patwary, Z. R.; Katti, K. S. Modelling clay-fluid interactions in montmorillonite clays. *Environmental Geotechnics* **2017**, *4* (5), 322-338. DOI: 10.1680/jenge.14.00027.

(6) Grim, R. E. Modern concepts of clay materials. *Journal of Geology* **1942**, *50* (3), 225-275.

(7) Grim, R. E. Clay Mineralogy McGraw-Hill. *New York* **1968**, 206.

(8) Hendricks, S. B.; Jefferson, M. E. Structures of kaolin and talc-pyrophyllite hydrates and their bearing on water sorption of the clays. *American Mineralogist* **1938**, *23* (12), 863-875.

(9) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Molecular mechanics of the swelling clay tactoid under compression, tension and shear. *Applied Clay Science* **2021**, *200*, 105908.

(10) Massat, L.; Cuisinier, O.; Bihannic, I.; Claret, F.; Pelletier, M.; Masrouri, F.; Gaboreau, S. Swelling pressure development and inter-aggregate porosity evolution upon hydration of a compacted swelling clay. *Applied Clay Science* **2016**, *124*, 197-210. DOI: 10.1016/j.clay.2016.01.002.

(11) Likos, W. J.; Wayllace, A. Porosity evolution of free and confined bentonites during interlayer hydration. *Clays and Clay Minerals* **2010**, *58* (3), 399-414. DOI: 10.1346/ccmn.2010.0580310.

(12) Jullien, M.; Raynal, J.; Kohler, E.; Bildstein, O. Physicochemical reactivity in clay-rich materials: Tools for safety assessment. *Oil & Gas Science and Technology-Revue D Ifp Energies Nouvelles* **2005**, *60* (1), 107-120. DOI: 10.2516/ogst:2005007.

(13) Perdrial, J. N.; Warr, L. N. Hydration behavior of mx80 bentonite in a confined-volume system: implications for backfill design. *Clays and Clay Minerals* **2011**, *59* (6), 640-653. DOI: 10.1346/ccmn.2011.0590609.

(14) Salles, F.; Douillard, J.-M.; Denoyel, R.; Bildstein, O.; Jullien, M.; Beurroies, I.; Van Damme, H. Hydration sequence of swelling clays: Evolutions of specific surface area and hydration energy. *Journal of Colloid and Interface Science* **2009**, *333* (2), 510-522. DOI: 10.1016/j.jcis.2009.02.018.

(15) Salles, F.; Beurroies, I.; Bildstein, O.; Jullien, M.; Raynal, J.; Denoyel, R.; Van Damme, H. A calorimetric study of mesoscopic swelling and hydration sequence in solid Na-montmorillonite. *Applied Clay Science* **2008**, *39* (3-4), 186-201. DOI: 10.1016/j.clay.2007.06.001.

(16) Amarasinghe, P. M.; Katti, K. S.; Katti, D. R. Molecular Hydraulic Properties of Montmorillonite: A Polarized Fourier Transform Infrared Spectroscopic Study. *Applied Spectroscopy* **2008**, *62* (12), 1303-1313.

(17) Pradhan, S. M.; Katti, K. S.; Katti, D. R. Evolution of Molecular Interactions in the Interlayer of Na-Montmorillonite Swelling Clay with Increasing Hydration. *International Journal of Geomechanics* **2015**, *15* (5). DOI: 10.1061/(asce)gm.1943-5622.0000412.

(18) Young, D. *Computational chemistry: a practical guide for applying techniques to real world problems*; John Wiley & Sons, 2004.

(19) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered molecular dynamics. In *Computational molecular dynamics: challenges, methods, ideas*, Springer, 1999; pp 39-65.

(20) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. An insight into quartz mineral interactions with kerogen in Green River oil shale. *International Journal of Coal Geology* **2021**, 103729.

(21) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Modeling the Behavior of Organic Kerogen in the Proximity of Calcite Mineral by Molecular Dynamics Simulations. *Energy & Fuels* **2020**, *34* (3), 2849-2860.

(22) Jaswandkar, S. V.; Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Dissociation Mechanisms of G-actin Subunits Govern Deformation Response of Actin Filament. *Biomacromolecules* **2021**, *22* (2), 907-917.

(23) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Differences in Interactions Within Viral Replication Complexes of SARS-CoV-2 (COVID-19) and SARS-CoV Coronaviruses Control RNA Replication Ability. *Jom* **2021**, *73* (6), 1684-1695. DOI: 10.1007/s11837-021-04662-6.

(24) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Binding of SARS-COV-2 (COVID-19) and SARS-COV to human ACE2: Identifying binding sites and consequences on ACE2 stiffness. *Chemical Physics* **2021**, *551*, 111353.

(25) Katti, K. S.; Sikdar, D.; Katti, D. R.; Ghosh, P.; Verma, D. Molecular interactions in intercalated organically modified clay and clay-polycaprolactam nanocomposites: Experiments and modeling. *Polymer* **2006**, *47* (1), 403-414. DOI: 10.1016/j.polymer.2005.11.055.

(26) Katti, D. R.; Katti, K. S.; Thapa, K.; Faisal, N. Modeling the Nanoscale Kerogen Inclusions in Green River Oil Shale. In *Poromechanics VI*, pp 1968-1975.

(27) Lloret, A.; Villar, M. V.; Sanchez, M.; Gens, A.; Pintado, X.; Alonso, E. E. Mechanical behaviour of heavily compacted bentonite under high suction changes. *Geotechnique* **2003**, *53* (1), 27-40. DOI: 10.1680/geot.53.1.27.37258.

(28) Pellet, F. L.; Keshavarz, M.; Boulon, M. Influence of humidity conditions on shear strength of clay rock discontinuities. *Engineering Geology* **2013**, *157*, 33-38. DOI: 10.1016/j.enggeo.2013.02.002.

(29) Villar, M. V.; Lloret, A. Influence of temperature on the hydro-mechanical behaviour of a compacted bentonite. *Applied Clay Science* **2004**, *26* (1-4), 337-350. DOI: 10.1016/j.clay.2003.12.026.

(30) Chang, F. R. C.; Skipper, N. T.; Sposito, G. Computer-simulation of interlayer molecular-structure in sodium montmorillonite hydrates. *Langmuir* **1995**, *11* (7), 2734-2741. DOI: 10.1021/la00007a064.

(31) Karaborni, S.; Smit, B.; Heidug, W.; Urai, J.; vanOort, E. The swelling of clays: Molecular simulations of the hydration of montmorillonite. *Science* **1996**, *271* (5252), 1102-1104. DOI: 10.1126/science.271.5252.1102.

(32) Shroll, R. M.; Smith, D. E. Molecular dynamics simulations in the grand canonical ensemble: Application to clay mineral swelling. *Journal of Chemical Physics* **1999**, *111* (19), 9025-9033. DOI: 10.1063/1.480245.

(33) Smith, D. E.; Wang, Y.; Whitley, H. D. Molecular simulations of hydration and swelling in clay minerals. *Fluid Phase Equilibria* **2004**, *222*, 189-194. DOI: 10.1016/j.fluids.2004.06.023.

(34) Skipper, N. T.; Chang, F. R. C.; Sposito, G. Monte-carlo simulation of interlayer molecular-structure in swelling clay-minerals .1. Methodology. *Clays and Clay Minerals* **1995**, *43* (3), 285-293. DOI: 10.1346/ccmn.1995.0430303.

(35) Skipper, N. T.; Sposito, G.; Chang, F. R. C. Monte-carlo simulation of interlayer molecular-structure in swelling clay-minerals .2. Monolayer hydrates. *Clays and Clay Minerals* **1995**, *43* (3), 294-303. DOI: 10.1346/ccmn.1995.0430304.

(36) Rahromostaqim, M.; Sahimi, M. Molecular Dynamics Simulation of Hydration and Swelling of Mixed-Layer Clays. *Journal of Physical Chemistry C* **2018**, *122* (26), 14631-14639. DOI: 10.1021/acs.jpcc.8b03693.

(37) Sun, L.; Tanskanen, J. T.; Hirvi, J. T.; Kasa, S.; Schatz, T.; Pakkanen, T. A. Molecular dynamics study of montmorillonite crystalline swelling: Roles of interlayer cation species and water content. *Chemical Physics* **2015**, *455*, 23-31. DOI: 10.1016/j.chemphys.2015.04.005.

(38) Tao, L.; Tian, X.-F.; Yu, Z.; Tao, G. Swelling of K+, Na+ and Ca2+-montmorillonites and hydration of interlayer cations: a molecular dynamics simulation. *Chinese Physics B* **2010**, *19* (10). DOI: 10.1088/1674-1056/19/10/109101.

(39) Teppen, B. J.; Rasmussen, K.; Bertsch, P. M.; Miller, D. M.; Schafer, L. Molecular dynamics modeling of clay minerals .1. Gibbsite, kaolinite, pyrophyllite, and beidellite. *Journal of Physical Chemistry B* **1997**, *101* (9), 1579-1587. DOI: 10.1021/jp961577z.

(40) Schmidt, S. R.; Katti, D. R.; Ghosh, P.; Katti, K. S. Evolution of mechanical response of sodium montmorillonite interlayer with increasing hydration by molecular dynamics. *Langmuir* **2005**, *21* (17), 8069-8076. DOI: 10.1021/la050615f.

(41) Katti, D. R.; Schmidt, S. R.; Ghosh, P.; Katti, K. S. Modeling the response of pyrophyllite interlayer to applied stress using steered molecular dynamics. *Clays and Clay Minerals* **2005**, *53* (2), 171-178. DOI: 10.1346/ccmn.2005.0530207.

(42) Heinz, H.; Koerner, H.; Anderson, K. L.; Vaia, R. A.; Farmer, B. L. Force field for mica-type silicates and dynamics of octadecylammonium chains grafted to montmorillonite. *Chemistry of Materials* **2005**, *17* (23), 5658-5669. DOI: 10.1021/cm0509328.

(43) Katti, D. R.; Ghosh, P.; Schmidt, S.; Katti, K. S. Mechanical properties of the sodium montmorillonite interlayer intercalated with amino acids. *Biomacromolecules* **2005**, *6* (6), 3276-3282. DOI: 10.1021/bm0503219.

(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289.

6. BINDING OF SARS-COV-2 (COVID-19) AND SARS-COV TO HUMAN ACE2: IDENTIFYING BINDING SITES AND CONSEQUENCES ON ACE2 STIFFNESS[5]

6.1. Introduction

Coronaviruses have been posing mild to serious health concerns for the public since their discovery in 1965[1]. These large positive-stranded RNA viruses were named due to their crown-like appearance observed using electron microscopy[2]. About 200 different coronaviruses have been discovered to date, that infects different creatures, including bats, birds, cattle, dogs, pigs, rodents, monkeys, humans, etc. [3]. The seven coronaviruses found among humans are HCoV-229E, HCoV-NL63, HCoV-OC43, HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2 [4, 5]. Among the mentioned coronaviruses, the first four are commonly found and generally cause symptoms of common cold, while the uncommonly found latter three can be much more deadly, causing severe pneumonia. In 2002-03, the SARS-CoV (Severe Acute Respiratory Syndrome) infected around 8000 people around the globe and caused 774 deaths[6]. Since its first emergence in 2012, the MERS-CoV (Middle East Respiratory Syndrome) has infected 2494 people with a fatality of 858 in 27 countries[7]. The SARS-CoV-2, also known as COVID-19, the recently emerged pandemic, was first reported in Wuhan city, China, in late December 2019[8]. According to the World Health Organization (WHO), the number of globally confirmed cases of COVID-19 is 209,201,939 with 4,390,467 fatalities in 216 countries (as of August 20, 2021) [9].

The SARS-CoV and SARS-CoV-2 are closely related coronaviruses that are classified as beta-coronaviruses, and both have originated in bats[5, 10]. The SARS-CoV-2 genome exhibits an 80% identity match with the SARS-CoV genome[11]. Another remarkable similarity between them

[5] The contents of this chapter have been published in Faisal, HM Nasrullah, Kalpana S. Katti, and Dinesh R. Katti. "Binding of SARS-COV-2 (COVID-19) and SARS-COV to human ACE2: Identifying binding sites and consequences on ACE2 stiffness." *Chemical Physics* (2021): 111353. As the lead author of this journal article, the copyright permission has been obtained from Elsevier.

is their host cell entry mechanism. Both coronaviruses utilize spike glycoproteins (S) to enter host cells by binding with cell surface Angiotensin-Converting Enzyme 2 (ACE2) receptors though their spike (S) genes share only 75% sequence similarity [5, 11, 12]. Spike glycoprotein (S), one of the four structural proteins of coronaviruses, is a class I virus membrane fusion protein[13]. The large ectodomain of spike protein comprises of receptor-binding domain S1 and membrane fusion domain S2[14]. Both the N-terminal domain (NTD) and the C-terminal domain (CTD) of the S1 subunit are attributed to viral host receptor attachment of different coronaviruses[15]. The S2 subunit, the most conserved region of the spike protein, carries the fusion peptide (FP) along with two heptad repeats (HR1 and HR2) for performing viral and host membrane fusion[13, 16].

The ACE2, an essential carboxypeptidase of the renin-angiotensin system (RAS), plays a crucial role in maintaining cardiovascular homeostasis[17]. This Type I membrane protein is primarily expressed in the heart, kidneys, intestine, and lungs [18, 19]. As a homolog of ACE, it negatively regulates the RAS system by cleaving AngI into Ang1-9 and AngII into Ang1-7 [18, 20]. Inside a healthy human lung, alveolar epithelial Type II cells are characterized by abundant expression of ACE2[21]. Downregulation of ACE2 in these cells causes severe lung injury that may be associated with acute respiratory distress syndrome (ARDS) occurring from alveolar collapse due to increased surface tension[22-24]. Both SARS-CoV and SARS-CoV-2 infection have been shown to cause ARDS in severely ill patients[25]. The introduction of host cell infection by these coronaviruses is marked by the molecular interaction of the spike glycoprotein (S) receptor-binding domain (RBD) with ACE2 cell receptor[26]. This interaction ultimately leads to the invasion of the host cell by the virus replicating machinery.

The *in vivo* folding behavior of proteins contributes to their effective functioning[27], and variation in temperature and pH impacts the folded conformation[28]. Cellular motion-induced

mechanical stretching in the extracellular matrix, muscle, and cell receptors also result in protein unfolding [29]. As downregulation of ACE2 cell receptors with cyclic stretching of human lung epithelial cells may be associated with ARDS in case of coronavirus infections (both SARS-CoV and SARS-CoV-2), the molecular interactions and unfolding pathway of ACE2 with and without the presence of spike receptor-binding domain can highlight the deviation of ACE2 behavior due to viral infections[22, 30]. This change in behavior can be modeled through pairwise non-bonded interactions and mechanical response to external forces. Molecular dynamics (MD) simulation is a computational technique that predicts the time-dependent behavior of a molecular system in terms of energy (bonded and non-bonded) and conformation. MD simulations have been employed to investigate different material systems i.e., oil shale [31], swelling clays [32], and proteins [33]. The interactions within coronaviral RNA dependent RNA polymerase (RdRp) have also been analyzed using MD simulations [34].Steered molecular dynamics (SMD) is an *in silico* mechanobiological methodology for investigating the mechanical response of proteins during unfolding as well as the unbinding procedure of ligands from them[35]. In the current study, we report molecular dynamics simulations and steered molecular dynamics simulations of human ACE2 in the proximity of both SARS-CoV and SARS-CoV-2 spike (S) protein receptor-binding domain (RBD) to determine their pairwise non-bonded interactions and effect of these interactions on the mechanical response of ACE2 respectively. We also utilize SMD to explore the binding forces of coronavirus spike RBDs to ACE2. Since ACE2 is the primary cellular receptor for the SARS-CoV and the SARS-CoV-2, any changes in the mechanisms of attachment of ACE2 with SARS-CoV and SARS-CoV-2 spike (S) protein receptor-binding domains (RBD) is relevant to the understanding of the host cell invasion and for developing interventions to prevent attachment.

6.2. Methodology

6.2.1. The SARS-CoV ACE2 and SARS-CoV-2 ACE2 Interaction Model Construction

The initial three-dimensional structures of SARS-CoV spike RBD with ACE2 and SARS-CoV-2 spike RBD with ACE2 have been obtained from RCSB Protein Data Bank. The SARS-CoV model has been developed using X-ray diffraction data [36] while the model for SARS-CoV-2 was constructed using cryo-Electron Microscopy data [37]. Both of these models were experimentally validated before submitting to Protein Data Bank. The corresponding PDB ID of SARS-CoV spike RBD-ACE2 complex and SARS-CoV-2 spike RBD-ACE2 complex are 2AJF and 6M17, respectively [36, 37]. The models were chosen due to their availability and similarity i.e., both models utilized ACE2 homodimer. The SARS-CoV complex model (2AJF) contains two spike-RBD chains (E and F) bound with two ACE2 protein chains (A and B) (**Fig. 1**).

The SARS-CoV-2 model comprises of two spike-RBD chains (E and F) with ACE2 dimer (chain B and D) along with the neutral amino acid transporter B^0AT1 (chain A and C). As the primary objective of this study is to investigate the interactions between coronavirus spike RBD and ACE2, we have removed the amino acid transporter B^0AT1 from model 6M17 (**Fig. 2**).

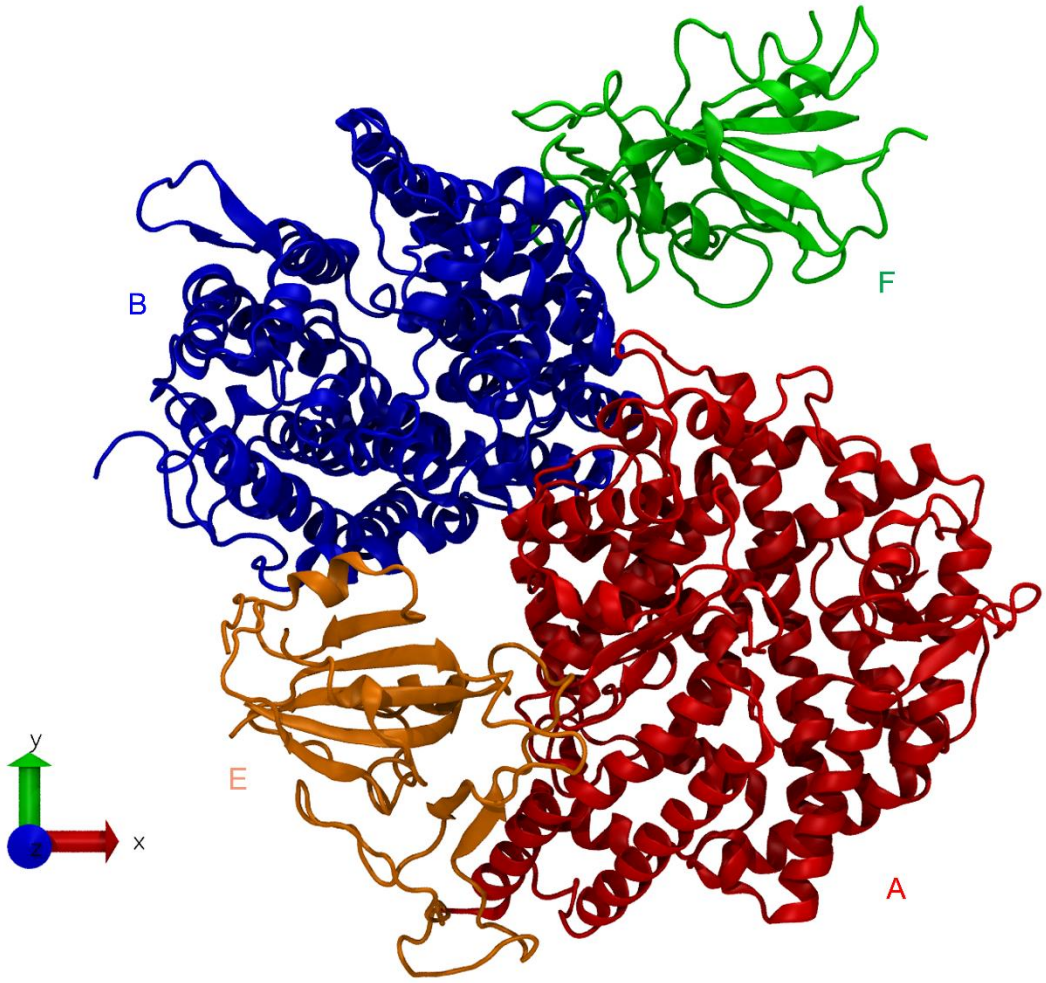

Figure 6.1. Equilibrated structure of human ACE2 with SARS-CoV spike RBD where two chains of SARS-CoV spike RBD (chain E and F) attach to the peptidase domains (PD) of ACE2 homodimer (chain A and B), i.e., chain E attaches on the PD of chain A, and chain F attaches on the PD of chain B. Chain A, B, E and F are represented by red, blue, orange and green colored segments respectively. The initial model has been obtained from RCSB Protein Data Bank (PDB ID: 2AJF) and solvated in a water box. Water molecules are removed from display for the visual clarity of molecular structures.

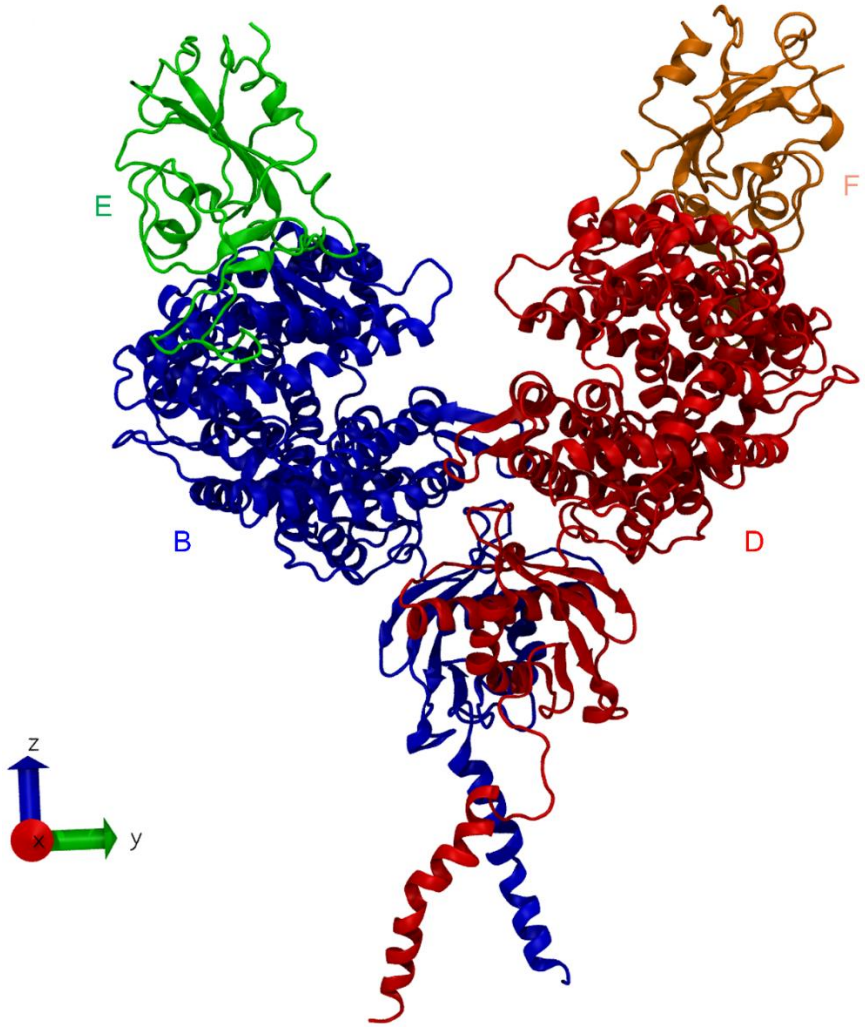

Figure 6.2. Equilibrated structure of full-length human ACE2 (both peptidase and collectrin domains) with SARS-CoV-2 spike RBD where two chains of SARS-CoV-2 spike RBD (chain E and F) attach to the N-terminal peptidase domains (PD) of ACE2 homodimer (chain B and D), i.e., chain E attaches on the PD of chain B, and chain F attaches on the PD of chain D. Chain B, D, E and F are represented by blue, red, green and orange colored segments respectively. The initial model has been obtained from RCSB Protein Data Bank (PDB ID: 6M17).

Further, the AutoPSF plugin of Visual Molecular Dynamics (VMD 1.9.3) has been used for adding H-atoms and assigning partial charges to both models. These protein models were then solvated in a water box of 5 Å thickness. The final dimensions of solvated SARS-CoV spike RBD-ACE2 complex are 118.75 Å × 118.78 Å × 145.36 Å comprising a total of 192,502 atoms. The solvated model of full-length ACE2 with SARS-CoV-2 spike RBD has the dimensions of 106.06 Å × 146.66 Å × 200.60 Å with 295,174 atoms.

6.2.2. Molecular Dynamics and Steered Molecular Dynamics Simulations of SARS-CoV and SARS-CoV-2 RBD Interaction with ACE2

Molecular dynamics (MD) and steered molecular dynamics (SMD) simulations were performed using NAMD 2.12, a parallel molecular dynamics code [38]. NAMD was developed by the Theoretical and Computational Biophysics Group at the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. All the parameters were obtained from CHARMm (Chemistry at HARvard Macromolecular Mechanics) force field [39]. It consists of functions and constants to define energy expression. CHARMm uses both bonded and non-bonded interaction terms. In this study, we utilize non-bonded interactions. At first, both models are minimized at 0 K temperature and 0 bar pressure using conjugate-gradient method [40]. Further, both models are brought to 310 K temperature and 1.01325 bar pressure to mimic the human physiological condition. The models are run for five ns using a timestep of 0.5 fs until they reach the equilibrium condition. Thermodynamic and conformational equilibration of structures are characterized by total energy, and root mean squared deviation (RMSD), respectively. These equilibrated models are further utilized for non-bonded energy calculations and steered molecular dynamics (SMD) simulations. Constant velocity SMD of ACE2 is done to assess its mechanical behavior by pulling its one terminal while keeping the other terminal fixed. In case of SARS-CoV model, the N-terminal was pulled and the C-terminal was kept fixed with and without the presence of spike RBD (**Fig. 3a**). In the SARS-CoV-2 model, the boundary atom of peptidase domain (residue id 615) remained fixed while pulling its N-terminal (**Fig. 4a)**. A constant velocity pulling of the spike RBD is utilized to probe its pull-off force from ACE2. The C-terminal of spike RBD was pulled with a velocity of 0.01 Å/fs by keeping the distant terminal (C-terminal for SARS-CoV and peptidase domain terminal for SARS-CoV-2) of ACE2 fixed (**Fig. 3b and Fig. 4b**). All the

SMD simulations employ a spring constant of 7 kcal/mol/ $Å^2$ and a pulling velocity of 0.01 Å/fs. All the simulations were performed at the Center for Computationally Assisted Science and Technology (CCAST), a parallel computing facility at North Dakota State University. Each simulation utilized one node, dual-socket Intel Xeon 2670v2 "Ivy Bridge" 2.5GHz with 64GB DDR3 RAM at 1866MHz and 50 processors.

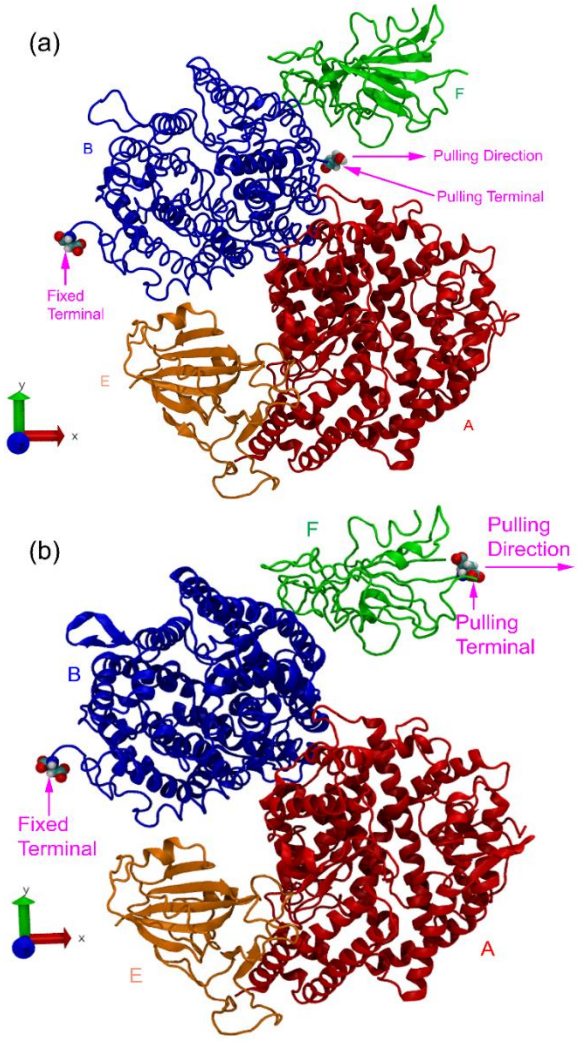

Figure 6.3. **(a)** The constant-velocity pulling of ACE2 in the presence of SARS-CoV spike RBD. The structures of ACE2 homodimer (chain A and B) with SARS-CoV spike RBD (chain E and F) are used. The N-terminal of ACE2 chain B is pulled along the X-axis with a constant velocity by keeping its C-terminal fixed both in the presence and absence of spike RBD chain F. **(b)** The constant-velocity pulling of spike RBD chain F for unbinding from ACE2 chain B performed by pulling the C-terminal of spike RBD chain F while keeping the C-terminal of ACE2 chain B fixed.

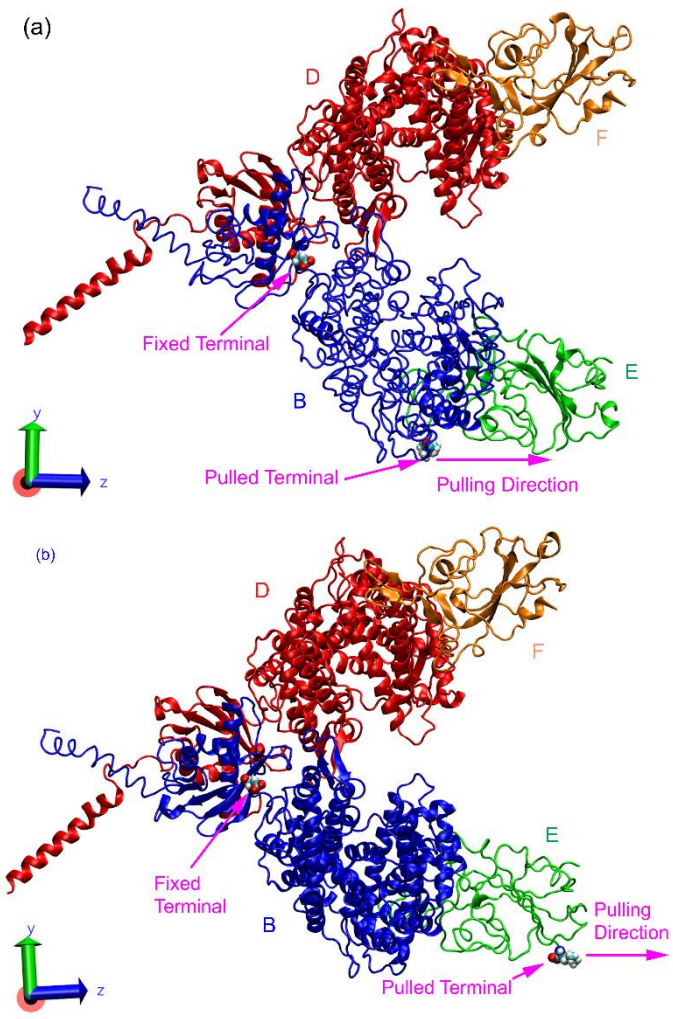

Figure 6.4. **(a)** The constant-velocity pulling of ACE2 in the presence of SARS-CoV-2 spike RBD. The equilibrated structures of ACE2 homodimer (chain B and D) with SARS-CoV-2 spike RBD (chain E and F) have been employed to explore this behavior. The N-terminal of ACE2 chain B was pulled along the Z-axis with a constant velocity by keeping it's peptidase domain terminal fixed both in the presence and absence of spike RBD chain E. **(b)** The constant-velocity pulling of spike RBD chain E towards its unbinding from ACE2 chain B. It has been performed by pulling the C-terminal of SARS-CoV-2 spike RBD chain E while keeping the peptidase domain terminal of ACE2 chain B fixed.

6.3. Results

6.3.1. Interaction Energies of the SARS-CoV Spike RBD with Human ACE2 Complex

The full-length ectodomain structure of human ACE2 is characterized by the claw-like N-terminal peptidase domain and C-terminal collectrin domain [41]. In this case, the N-terminal

peptidase domain of ACE2 serves as the cellular receptor of concave surfaced SARS-CoV spike receptor-binding domain (RBD) [36]. The SARS-CoV spike RBD is 174 residues long, with the terminal residues being cysteine (CYS) and glutamic acid (GLU) [36]. Each unit of ACE2 homodimer consists of 597 residues with serine (SER) and aspartic acid (ASP) as the terminal residues. The spike RBD attaches to each ACE2 protomer resulting in the complex formation with two spike protein chains (E and F) with ACE2 homodimer (A and B), as shown in **Fig. 1**. Further, chain E attaches to peptidase domain (PD) of chain A, while chain F attaches similarly to the chain B. We define the total non-bonded interaction energies between two molecular entities as the sum of electrostatic and van der Waals (VDW) interaction energies between the entities.

The negative values of non-bonded interaction energies constitute attractive interactions, and the positive values are repulsive interactions. In order to determine the interactions between human ACE2 and SARS-CoV spike RBD, non-bonded interaction energies have been computed between chain B of ACE2 and chain F of spike RBD (**Fig. 5a** and **Supplementary Table 1**).

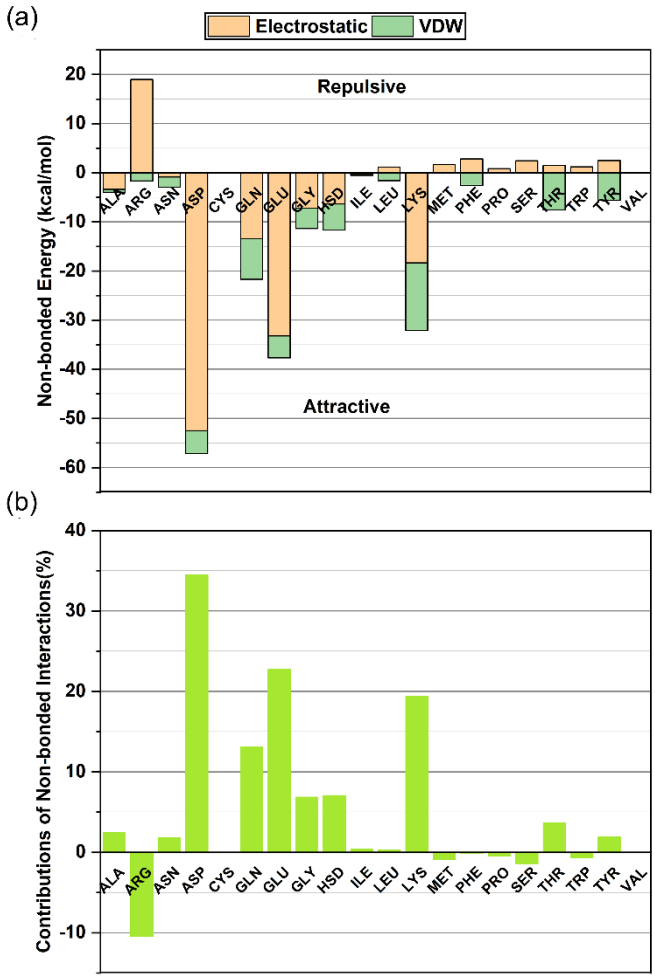

Figure 6.5. **(a)** The non-bonded interaction energies of ACE2 residues with SARS-CoV spike RBD. Negative and positive magnitudes of interaction energies represent the attractive and repulsive interactions, respectively. Here, non-bonded interactions are measured between chain B of ACE2 and chain F of SARS-CoV spike RBD. **(b)** The relative percentage contributions of non-bonded interactions by ACE2 residues with SARS-CoV spike RBD. Four residues of ACE2 (ASP>GLU>LYS>GLN) have significant interactions (attractive) with SARS-CoV spike RBD as they contribute about 89% of the total interactions altogether. ARG is the only residue of ACE2 that interacts significantly but repulsively with SARS-CoV spike RBD.

The total non-bonded interactions between them are -166 kcal/mol, with electrostatic and VDW interactions of -103 kcal/mol and -63 kcal/mol, respectively. Among the 20 different types of residues (amino acids) of ACE2, four different residues interact significantly with spike RBD. Aspartic acid (-57 kcal/mol), glutamic acid (-38 kcal/mol), lysine (-32 kcal/mol) and glutamine (-

22 kcal/mol) of ACE2 contribute approximately 89% of the attractive non-bonded interactions with SARS-CoV spike RBD (**Fig.5(b)**). Other significantly interacting residues of human ACE2 are histidine (-12 kcal/mol), glycine (-11 kcal/mol), and arginine (+17 kcal/mol). The non-bonded interactions of specific residues of ACE2 with SARS-CoV spike RBD are provided in **Supplementary Table 5.** These specific residues have been chosen based on the suggestion of structure resolving study [36] and they been shown to contribute 29.2% (-48.5 kcal/mol) of the total interaction energy.

In terms of the secondary structure, turn and coil components of ACE2 contribute to the majority (62%) of the non-bonded interactions (-103 kcal/mol). The rest of the interactions originate primarily from the alpha-helices of ACE2 (Supplementary Table 2). In terms of the tertiary structure, more than 99% of the interactions arise from the polar residues of ACE2. Also, the non-bonded interactions between the two chains of ACE2 (A and B) are calculated as -285 kcal/mol in the absence of SARS-CoV spike RBD. This interaction energy is reduced to -107 kcal/mol in the ACE2-spike RBD complex. In both cases, the interactions are predominantly electrostatic.

6.3.2. The Binding Force of the SARS-CoV Spike RBD with Human ACE2 Complex

The constant velocity steered molecular dynamics (SMD)[38, 42] method has been used to investigate the mechanical response of ACE2 through its modeling of unfolding to external loading. As shown in **Figure 6a**, Chain B of ACE2 is stretched at constant velocity (0.01 Å/fs) both in the absence and presence of SARS-CoV spike RBD chain F. The molecular stretching is performed by pulling the N-terminal of chain B while keeping the C-terminal fixed. The force-displacement plot (**Fig. 6b)** represents the stretching pathway of chain B, where the peaks represent the unfolding of alpha helices, small turns, and coils (breaking of H-bonds). In the absence of

SARS-CoV spike RBD, the unfolding of the α1 helix of ACE2 occurs at a force of 6624 pN with a corresponding displacement of 101 Å. The highest peak force, 8,408 pN at a displacement of 172 Å, represents the unfolding of α2 helix. The linear stretching of the protein chain characterizes the post-peak downhill region of this peak where no unfolding of coil or helix occurs. In the presence of spike RBD chain F, the breaking of α1 and α2 helices occur at 7,425 pN and 10,021 pN at displacements of 104 Å and 184 Å, respectively. The pulling of ACE2 causes the translation of spike RBD in the same direction.

The binding force of SARS-CoV spike RBD and the human ACE2 has also been explored utilizing the constant velocity SMD simulations. The C-terminal of SARS-CoV spike RBD chain F is pulled away from ACE2 chain B to determine the amount of force required to separate the spike RBD from the spike RBD-ACE2 complex (**Fig. 6(c)**). The detachment of the spike RBD from ACE2 is seen as two important unlatching events at about 20 Å and 121 Å displacement, which are indicated by two sharp peaks in the force-displacement curve. The first peak is observed at 8,030 pN force at 20 Å (**Fig. 6(d)**). The downward slope is characterized by the linear stretching of RBD chain F with no unfolding of coils or helices. Beyond the minimum at 61 Å displacement, the RBD starts to move away from ACE2 due to pulling. The second peak in the plot at a displacement of 121 Å occurs at a force of 10,454 pN and results in complete detachment of the spike RBD from the ACE2 complex. Beyond the second peak, all the smaller peaks and valleys in the plot are related only to the response of spike RBD chain F to pulling.

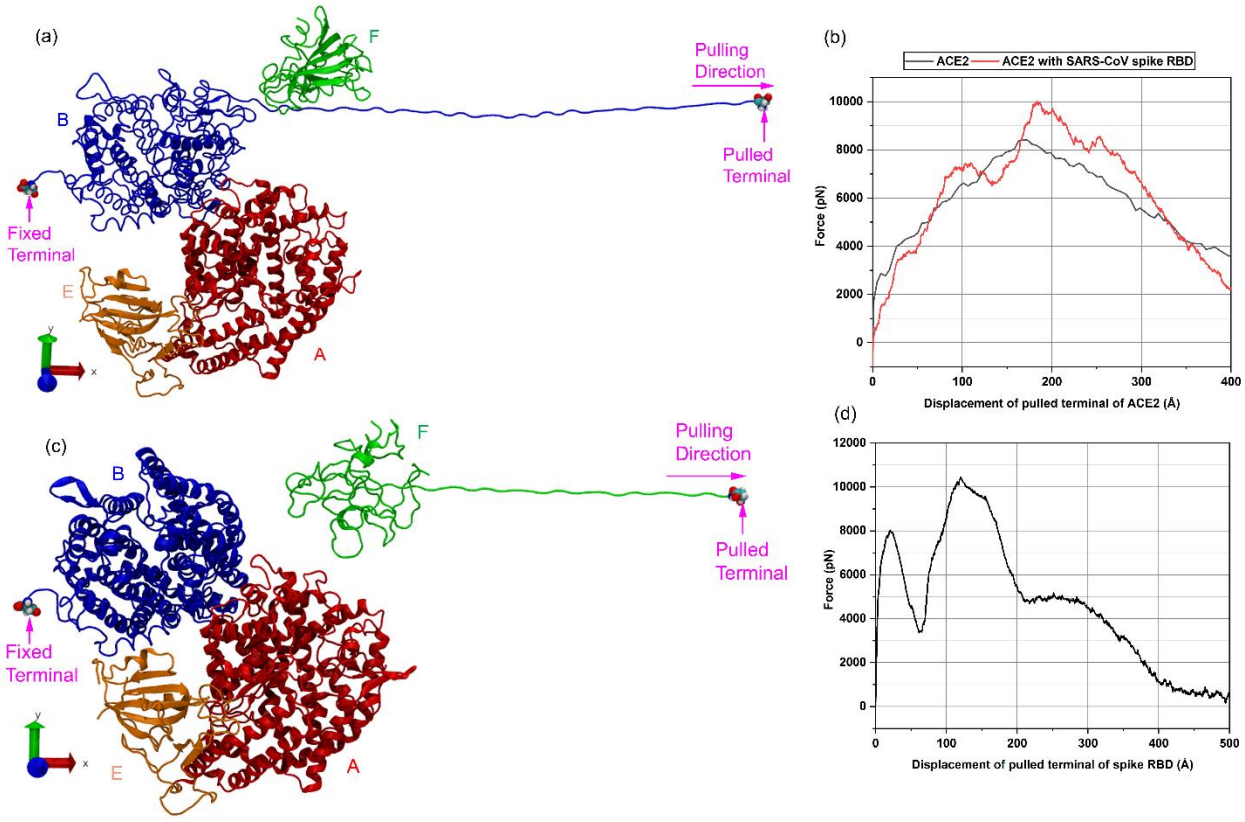

Figure 6.6. (a) Mechanical response of ACE2 chain B with the attcachment of SARS-CoV spike RBD chain F due to constant-velocity pulling. (b) The force-displacement plot of ACE2 with and without bound spike RBD. The plot for bound spike has been corrected for the rigid motion of spike RBD by subtracting the amount of force accountable for this motion. The peaks in force-displacement plots of ACE2 characterize the unwinding of helices (H-bond breaking). The presence of spike RBD increases the required force to unwind the helices/coils, thus increasing the stiffness of ACE2. (c) Pulling off SARS-CoV spike RBD chain F from ACE2 peptidase domain of chain B. (d) Force-displacement plot of pulling of spike RBD chain F for unbinding from ACE2 chain B. The first peak specifies the partial separation of spike RBD (chain F) while its complete separation from ACE2 (chain B) is marked by the second peak. Beyond this point, all the minor peaks and other features in the plot represent the linear stretching of spike RBD.

6.3.3. Interaction Energies of the SARS-CoV-2 (Covid-19) Spike RBD with Human ACE2

The spike (S) protein of SARS-CoV-2 attaches to the peptidase domain (PD) of human ACE237. Due to the homodimerization of ACE2, two spike (S) protein receptor-binding domains (RBDs) attach to ACE2 dimer, where each PD binds with one RBD. Chain B and D of ACE2 are attached to spike RBD chain E and F, respectively (Fig.2). The SARS-CoV-2 spike RBD consists of 183 residues with cysteine (CYS) at the N-terminal and leucine (LEU) at the C-terminal. The

ACE2 protomer is a full-length model having 748 residues, which N-terminal is isoleucine (ILE), and C-terminal is arginine (ARG).37 The peptidase domain consists of about 80% of the total residues of ACE2 (residues 21 to 615). The non-bonded interactions between every residue of chain B of ACE2 and chain E of SARS-CoV-2 spike RBD have been computed to probe the interactions within the SARS-CoV-2 spike RBD-ACE2 complex and are shown in **Fig. 7(a)** and **Supplementary Table 3**. The total non-bonded interactions between them (chain B of ACE2 and SARS-CoV-2 spike RBD chain E) are observed to be -356 kcal/mol with electrostatic and VDW interactions of -289 kcal/mol and -67 kcal/mol respectively. Among the 20 different residues of ACE2, two residues contribute to almost 86% of the total non-bonded attractive interactions between ACE2 and SARS-CoV-2 RBD (**Fig. 7(b)**). The non-bonded interactions of specific residues of ACE2 with SARS-CoV-2 spike RBD are provided in **Supplementary Table 5.** These specific residues have been chosen based on the suggestion of structure resolving study [37] and they been shown to contribute 43.2% (-153.8 kcal/mol) of the total interactions.

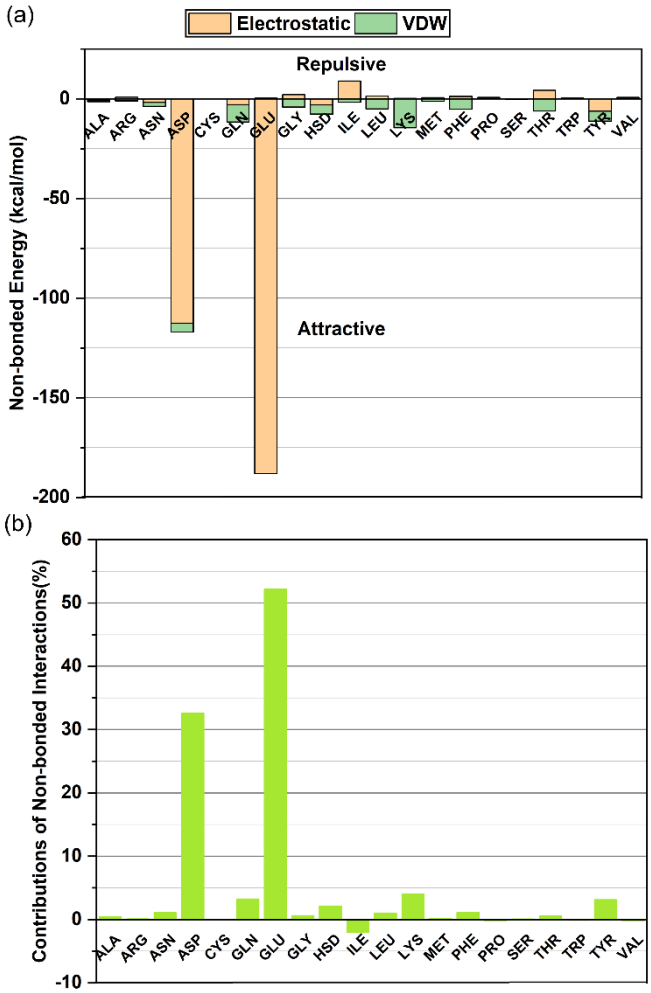

Figure 6.7. **(a)** The non-bonded interactions of ACE2 residues with SARS-CoV-2 spike RBD. Negative and positive magnitudes of interactions represent attractive and repulsive interactions, respectively. Here, non-bonded interactions are measured between chain B of ACE2 and chain E of SARS-CoV-2 spike RBD. **(b)** The relative percentage contributions of non-bonded interactions of ACE2 residues with SARS-CoV-2 spike RBD. Electrostatic interactions yield more than 80% of the non-bonded interactions. Two residues of ACE2 (GLU>ASP) have significant attractive interaction with SARS-CoV-2 spike RBD as they contribute about 86% of the total interaction energy.

The glutamic acid (GLU) alone contributes to more than 52% of the interaction energy (-188 kcal/mol), and Aspartic acid (ASP) contributes to about 33% of the interaction energy (-117 kcal/mol) with SARS-CoV-2 spike RBD. Lysine (LYS), glutamine (GLN), and tyrosine (TYR) are other significantly interacting residues of ACE2. Based on protein secondary structure, the

helices of ACE2 contribute to 94% of the attractive interaction energy with SARS-CoV-2 spike RBD while beta-strands and turns produce the remainder (Supplementary Table 4). Polar residues of ACE2 participate in more than 99% of these interactions. Also, the total non-bonded interactions within the ACE2 dimer (between chain B and D) are observed to be -648 kcal/mol and -808 kcal/mol in the absence and presence of SARS-CoV-2 spike RBD respectively. In both cases, the electrostatic interactions produce more than 80% of the total interactions.

6.3.4. Binding Force of the SARS-CoV-2 (COVID-19) Spike RBD with Human ACE2 Complex

The mechanical response of ACE2 chain B is determined utilizing constant-velocity SMD simulations both in the absence and the presence of SARS- CoV-2 spike RBD chain E as shown in **Fig. 8a**. This is performed by pulling the N-terminal of human ACE2 chain B while keeping its peptidase domain terminal (residue 615) fixed. The SARS-CoV spike RBD-ACE2 model contains only the peptidase domain of ACE2. In order to be consistent with the SARS-CoV model, we have fixed the peptidase domain terminal of full length ACE2 of SARS-CoV-2 spike RBD-ACE2 model. The first peak in the force-displacement plot of ACE2 without spike RBD indicates the beginning of α1 helix unfolding (breaking of H-bonds) (**Fig. 8(b)**). The peak of 3,803 pN at a displacement of 112 Å marks the complete unfolding of α1 helix. The unfolding of α2 helix is accomplished at 4,286 pN force at a displacement of 251 Å. The subsequent peak of 5,456 pN is the result of the unfolding of α3 helix at the displacement of 539 Å, while downward slopes represent the linear stretching of protein without the breaking of coils or helices. The presence of spike RBD in the proximity of ACE2 causes an increase in the peak helix-unfolding forces by a significant amount. In the SARS-CoV-2 spike RBD-ACE2 complex, the unfolding forces of the α1, α2, and α3 helices are computed as 5,346 pN, 4,370 pN, and 6,492 pN at displacements of 102

Å, 254 Å, and 530 Å respectively. The spike RBD chain E also moves in the pulling direction along with the peptidase domain of ACE2.

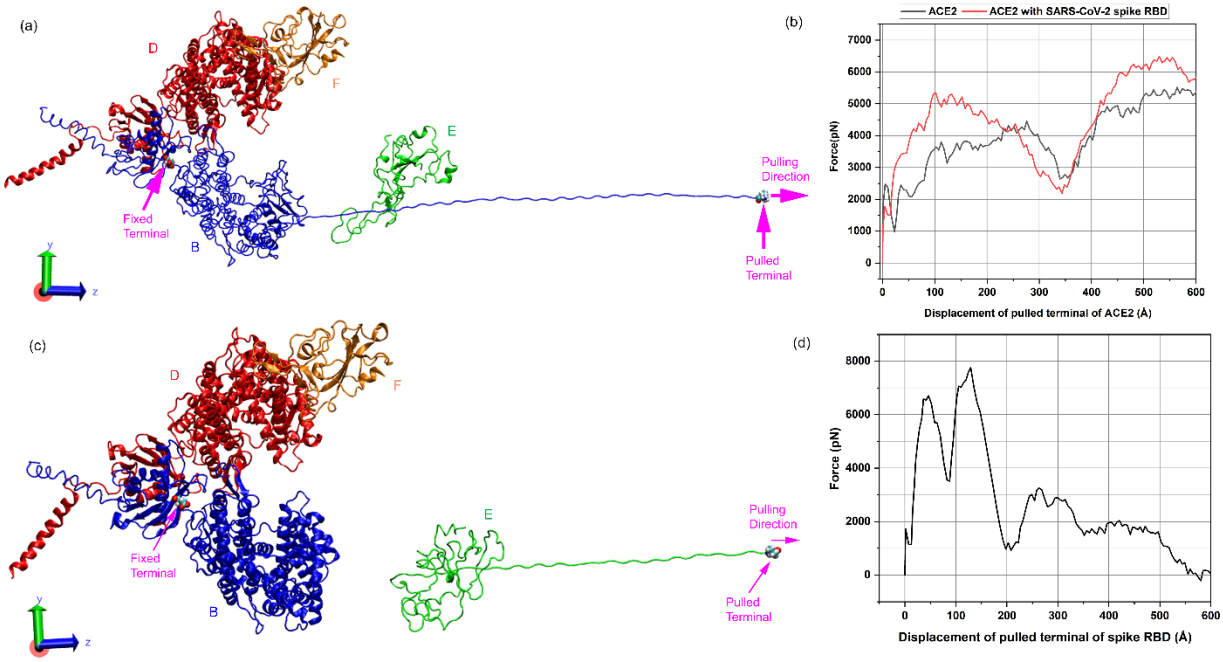

Figure 6.8. **(a)** Mechanical response of ACE2 chain B with the attachment of SARS-CoV-2 spike RBD chain E. **(b)** The force-displacement plot of ACE2 with and without bound spike RBD. The plot for bound spike has been corrected for the rigid motion of RBD by subtracting the amount of force accountable for this motion. The peaks in force-displacement plots of ACE2 characterize the unwinding of helices (H-bond breaking). The presence of spike RBD increases the required force to unwind the helices/coils, thus increasing the stiffness of ACE2. **(c)** Pulling off SARS-CoV-2 spike RBD chain E from ACE2 peptidase domain of chain B. **(d)** Force-displacement plot of pulling of spike RBD chain E towards its unbinding from ACE2 chain B. The first peak specifies the partial separation of spike RBD (chain E) while its complete separation from ACE2 (chain B) is marked by the second peak. Beyond this point, all the minor peaks and other features in the plot represent the linear stretching of spike RBD.

The C-terminal of SARS-CoV-2 spike RBD chain E is pulled away from ACE2 chain B to evaluate the binding force of spike RBD within the SARS-CoV-2 spike RBD-ACE2 complex (**Fig. 8c**). The detachment of the spike RBD from ACE2 is characterized by two unlatching events at about 46.2 Å and 129 Å, which are indicated by two sharp peaks in the force-displacement curve. The first peak has a maximum force of 6713.18 pN at 46.2 Å displacement (**Fig. 8d**). The highest peak of 7759.95 pN at a displacement of 129 Å signifies the complete separation of spike RBD

chain E from the SARS-CoV-2 spike RBD-ACE2 complex. The remainder of the plot is the mechanical response of spike RBD chain E alone.

6.4. Discussion

The simulations indicate that electrostatic interactions dominate the non-bonded interactions between coronavirus spike RBD and ACE2 for both the SARS-CoV-2 and SARS-CoV. One of the major differences between the two coronaviruses spike-RBD interaction with ACE2 is that in the case of SARS-CoV-2, the majority of attractive interaction energies are primarily mediated by just two residues of ACE2; GLU and ASP, whereas, for SARS-CoV, the attractive interaction energies are spread out over four residues GLU, ASP, LYS and GLN (**Fig. 5a & 5b and Fig. 7a & 7b**). The two residue versus multi-residue interaction of the SARS-CoV-2 and SARS-CoV spike RBDs with the human ACE2 is likely to cause significant differences in the attachment of the two coronaviruses with human ACE2. These residues are generally classified as charged or polar residues attributed to salt-bridge and H-bond formations. Further, more than 99% of the spike-ACE2 interactions arise from polar residues of ACE2. The spike-ACE2 interactions for SARS-CoV-2 (-356 kcal/mol) are more than twice the interactions of SARS-CoV spike-ACE2 (-166 kcal/mol). Structural observations in a recent report also suggest higher interactions of SAR-CoV-2 RBD with ACE2 as compared to SARS-CoV RBD[37]. The higher interactions of SARS-CoV-2 are also reflected by the secondary structure of ACE2. In the SARS-CoV-2 spike RBD-ACE2 complex, the majority of the interaction energy arises from the ACE2 helices. On the other hand, in the SARS-CoV spike RBD-ACE2 complex, more than half of the interaction energy is caused by ACE2 turns. Another potential source of differences in interactions arises from the fact that the interacting ACE2 surface areas are 6059 $Å^2$ and 5595 $Å^2$ with SARS-CoV-2 spike RBD and SARS-CoV spike RBD respectively. A higher interacting surface area of

ACE2 results in higher interactions with SARS-CoV-2 spike RBD. A lower dissociation constant (K_D) is suggested from recently reported surface plasmon resonance experiments for the SARS-CoV-2 spike RBD binding of ACE2 (15 nM)[43] as compared to that also reported from surface plasmon experiments for SARS-CoV (185.1 nM)[44]. A lower dissociation constant represents higher binding affinity, i.e., SARS-CoV-2 spike has 10-15 fold higher binding affinity with ACE2 than SARS-CoV as per the above mentioned published studies[43-45]. The simulations presented here indicate a higher attractive non-bonded interaction energy for the SARS CoV-2 spike RBD with ACE (-356 kcal/mole) as compared to that for SARS-CoV RBD (-166 kcal/mole) which is in agreement with the surface plasmon experiments. The surface plasmon experiments are conducted by binding immobilized RBD on a sensor chip followed by injecting varying concentrations of ACE2[43, 44], and studying the association and dissociation of ACE2 to the RBD. Measuring binding energy of the molecular entities that form RBD and ACE2 complex provides a different but complementary view in the attachment of the ACE2 to the coronavirus RBD. The investigation of nine specific residues of ACE2 in terms of their interactions with coronaviral spike RBDs reveal some interesting phenomena. These nine residues cumulatively exhibited almost 3 times higher non-bonded interactions with SARS-CoV-2 spike RBD (-153.3 kcal/mol) relative to SARS-CoV spike RBD (-48.5 kcal/mol). Among these residues, D30 has been found to be most intercating (-97.5 kcal/mol) with SARS-CoV-2 while K353 interacted (-35.3 kcal/mol) most with SARS-CoV. Certain residues (K31, M82, R357) were found to be oppositely interacting (attraction/repulsion) between SARS-CoV and SARS-CoV-2 spike RBD.

The mechanical response of ACE2 is obtained from the force-displacement plot of the complex using constant velocity pulling (**Fig. 6b and Fig. 8b**). The sharp peaks in the plot result from the unfolding of helices or coils in ACE2. The ACE2 force-displacement curve shifts upward,

i.e., increased force is needed to cause the same displacement when the coronaviruses are attached to the ACE2 for both the coronaviruses. This force increment is the result of spike RBD binding interactions with the ACE2 peptidase domain (PD) that results in changes to the ACE2 unfolding behavior making the response stiffer. The force needed to pull the RBD by itself is found to be around 1000 pN, and this magnitude of the force is subtracted from the net force displacement of the RBD-ACE2 complex for both the coronaviruses. The plots shown in **Fig. 6b** and **Fig. 8b** are corrected for the force needed to pull just the RBD. From these observations, it is inferred that the presence of spike RBD makes ACE2 stiffer for both coronaviruses resulting in higher helix/coil unfolding forces. It is to be noted that the ACE2 model utilized with SARS-CoV spike RBD contains only the peptidase domain, while the ACE2 model utilized with SARS-CoV-2 spike RBD contains both peptidase and the collectrin domain [36, 37]. In both cases, the majority of the ACE2 models (spike-RBD ACE2 interacting regions) used are identical, with differences in the back end of the model. The collectrin domain at the back end of ACE2 in the SARS-CoV-2 model is far from spike RBD and does not affect the interactions with the spike RBD. However, in order to compare the force-displacement response of the ACE2 for the two coronavirus attachments, the peptidase domain terminal of full length ACE2 was fixed in the SMD simulations of SARS-CoV-2 spike RBD-ACE2 model. This particular action nullifies the impact of backend portion (collectrin domain) of ACE2 on the external loading response. The comparison of the mechanical response of ACE2 is done by computing the relative changes in ACE2 stiffness due to spike RBD attachment for both coronaviruses. The force displacement response of the ACE2 peptidase domain differs between the two models due to the differences in pulling direction and peptidase domain conformation with spike RBD. Therefore, the direct comparison of ACE2 stiffness is not done between SARS-CoV and SARS-CoV-2 spike attachement. Instead, relative change in

stiffness of ACE2 as a result of RBD attachment are calculated. The stiffness was calculated from the force-displacement curves by finding the slope of the line connecting first peak of the force-displacement curve to the origin from each of the plots in **Figures 6b** and **8b.** The relative change in stiffness was calculated as the ratio of change in stiffness due to spike attachment of ACE2 stiffnesses with spike RBD and the initial stiffness of ACE2 stiffness without spike RBD. The ACE2 exhibits a 54% increase in stiffness (from 33.95 pN/ Å to 52.41 pN/ Å) with the attachment of SARS-CoV-2 RBD, while only a 9% increase in stiffness (from 65.58 pN/ Å to 71.39 pN/ Å) of ACE2 is observed with the attachment of SARS-CoV RBD. The downregulation of ACE2 cell receptors is known to be responsible for ARDS [22, 23]. Dramatic differences are seen in the outcome of the ARDS severity in the SARS-CoV and SARS-CoV-2 infected patients. Here we observe a vivid difference in the change in stiffness of ACE2 (54% versus 9%) between SARS-CoV-2 and SARS-CoV attachment to ACE2. Change in stiffness suggest changes to the unfolding mechanisms of proteins. These altered unfolding characteristics or changes in structure are highly related to biological functionality. Thus, the observed significant differences in stiffness of ACE2 on the coronavirus attachments could point towards differences in biological response of ACE2 and therefore the vast differences in severity of ARDS in the two coronaviruses; resulting in 774 deaths due to SARS-CoV versus over 4,390,467deaths (as of August 20 , 2021) worldwide due to the SARS-CoV-2.

The binding force of spike RBD towards the ACE2 cell receptor is evaluated from the force-displacement behavior obtained by pulling the spike RBD at a constant velocity. We observe two peaks in the force-displacement plots for both coronaviruses resulting from two specific unlatching events that lead to the detachment of the spike RBD from ACE2. Both the spike RBDs from the two coronaviruses exhibit a two-step unlatching, leading to detachment, as shown in

Supplementary Figures 1 and **2**. The SARS-CoV-2 spike RBD shows unlatching events at 6713.18 pN and 7759.95 pN as compared to 8,030 pN and 10,454 pN, for the SARS-CoV spike RBD. Thus unlatching of the SARS-CoV-2 from ACE2 occurs at a lower force than the SARS-CoV. The mechanical pull-off of the coronavirus RBD from ACE2 enabled using SMD is a different phenomenon than measurement of dissociation constant using surface plasmon resonance[43, 44]. On pulling the spike RBD from the attachment to ACE2, conformational changes begin in the RBD, which progressively reduce interaction energies at the ACE2-spike RBD interface. These changes are influenced by factors like ACE2-RBD interface area, two-residue interactions of SARS-CoV-2 spike RBD as opposed to multi-residue interaction of SARS-CoV spike RBD with ACE2, RBD structure, unfolding rate of RBD, ACE2 conformation etc. As seen in Figures 6d and 8d, the pulled terminal of SARS-CoV-2 spike RBD needed to be displaced by a larger distance (580 Å) than SARS-CoV (500 Å) for the force magnitude during pulling to become zero, indicating the influence of factors described above on the deformation. It appears that these factors likely cause faster reduction of non-bonded interactions in spike-ACE2 complex during the pulling of SARS-CoV-2 RBD compared to SARS-CoV RBD, resulting in lower unlatching force of SARS-CoV-2 spike RBD.

Since SMD simulations presented here mimic single protein pulling experiments with AFM, it is suggested that future AFM experiments that evaluate binding mechanisms of these complexes consider the effect of the unfolding of spike RBD. SMD simulations are thus a useful methodology to observe an accurate sequence of events in the pulling away of the spike RBD from ACE2. Understanding the important biological consequence of the formation of the ACE2-coronavirus spike RBD complex is aided by this additional viewpoint of stiffening of ACE2 on

attachment to coronavirus, and the pulling force of the spike RBD from the ACE2-coronavirus spike RBD complex.

6.5. Conclusions

This study utilizes computational techniques to explain the initial host cell response due to coronavirus infections. It has been performed by capturing the molecular interactions and changes in the mechanical response of coronavirus cellular receptor angiotensin-converting enzyme 2 (ACE2) in the presence of SARS-CoV and SARS-CoV-2 spike (S) protein receptor-binding domain (RBD). Molecular dynamics (MD) simulation has been employed to determine the non-bonded interactions, while steered molecular dynamics (SMD) was used to describe the mechanical response of ACE2. The binding force of coronavirus spike (S) RBDs from the ACE2 has also been investigated by using SMD. The major findings of our study are summarized below:

- Of the attractive non-bonded interactions of SARS-CoV-2 RBD with ACE2, 86% result from just two ACE2 residues; GLU and ASP. On the other hand, 89% of the SARS-CoV spike RBD interaction energy is spread over four ACE2 residues, including LYS and GLN, besides GLU and ASP. These observations suggest potential sites of intervention to inhibit attachment of spike RBD to ACE2.
- The non-bonded interaction energies between SARS-CoV-2 spike RBD and ACE2 are more than twice the interaction energies between SARS-CoV spike RBD and ACE2.
- The pull-off force of the spike RBD from the ACE2 is higher in magnitude for the SARS-CoV. On pulling the spike RBD from the attachment of ACE2, continuous conformational changes begin in the RBD, which progressively reduce interaction energies at the ACE2-spike RBD interface and hence influence total pull-off force.

- The attachment of spike RBD with ACE2 results in the stiffening of ACE2 for both SARS-CoV-2 and SARS-CoV. The relative change in stiffness due to the attachment of spike RBD is higher for SARS-CoV-2 (54%) compared to the SARS-CoV (9%). The significantly larger relative stiffness of ACE2 on the SARS-CoV-2 attachment as compared to the SARS-CoV attachment points towards differences in the biological response of ACE2.

Since the host entry modes of the two coronaviruses compared here are similar, it is interesting to note the differences in the mechanisms of interactions, two ACE2 residues for SARS-CoV-2 versus multiple attachment residues for SARS-CoV. The stronger non-bonded interaction energies between SARS-CoV-2 and ACE2 result in a much stiffer ACE2 on attachment to the coronavirus spike RBD than for the SARS-CoV. Overall, the evaluation of these mechanisms of attachment and the resulting binding forces are critical to the development of therapies beyond vaccines that prevent the attachment and subsequent entry into host cells. Further mechanobiological studies that relate mechanical changes to the severity of the ARDS would provide a definitive answer to this important health concern.

6.6. Acknowledgments

The authors gratefully acknowledge the use of computational resources at the North Dakota State University Center for Computationally Assisted Science and Technology (NSF MRI grants #1229316, #2019077 and NSF OIA NDACES-1946202). Discussions on the physiological consequences and biological ramifications with Preeya Katti, Human Biology Department, University of Southern California are acknowledged.

6.7. References

(1) Tyrrell, D. A.; Bynoe, M. L. Cultivation of Viruses From a High Proportion of Patients With Colds. *Lancet* **1966**, *1* (7428), 76-&.

(2) Holmes, K. V.; Lai, M. M. C. Coronaviridae: The viruses and their replication. *Fundamental virology, Third edition* **1996**, 541-559.

(3) Tong, S.; Conrardy, C.; Ruone, S.; Kuzmin, I. V.; Guo, X.; Tao, Y.; Niezgoda, M.; Haynes, L.; Agwanda, B.; Breiman, R. F.; et al. Detection of Novel SARS-like and Other Coronaviruses in Bats from Kenya. *Emerging Infectious Diseases* **2009**, *15* (3), 482-485. DOI: 10.3201/eid1503.081013.

(4) Fehr, A. R.; Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. *Coronaviruses: Methods and Protocols* **2015**, *1282*, 1-23. DOI: 10.1007/978-1-4939-2438-7_1.

(5) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**. DOI: 10.1038/s41586-020-2012-7.

(6) Chan-Yeung, M.; Xu, R. H. SARS: Epidemiology. *Respirology* **2003**, *8*, S9-S14. DOI: 10.1046/j.1440-1843.2003.00518.x.

(7) Lee, J.; Chowell, G.; Jung, E. A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: A retrospective analysis on control interventions and superspreading events. *Journal of Theoretical Biology* **2016**, *408*, 118-126. DOI: 10.1016/j.jtbi.2016.08.009.

(8) Tan, W.; Zhao, X.; Ma, X.; Wang, W.; Niu, P.; Xu, W.; Gao, G. F.; Wu, G. Z. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019−2020. *China CDC Weekly* **2020**, *2* (4), 61-62.

(9) Organization, W. H. Coronavirus disease 2019 (COVID-19): situation report, 124. **2020**.

(10) Hu, B.; Zeng, L.-P.; Yang, X.-L.; Ge, X.-Y.; Zhang, W.; Li, B.; Xie, J.-Z.; Shen, X.-R.; Zhang, Y.-Z.; Wang, N.; et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *Plos Pathogens* **2017**, *13* (11). DOI: 10.1371/journal.ppat.1006698.

(11) Lu, R. J.; Zhao, X.; Li, J.; Niu, P. H.; Yang, B.; Wu, H. L.; Wang, W. L.; Song, H.; Huang, B. Y.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **2020**, *395* (10224), 565-574. DOI: 10.1016/s0140-6736(20)30251-8.

(12) Li, W. H.; Moore, M. J.; Vasilieva, N.; Sui, J. H.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greenough, T. C.; et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426* (6965), 450-454. DOI: 10.1038/nature02145.

(13) Bosch, B. J.; van der Zee, R.; de Haan, C. A. M.; Rottier, P. J. M. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *Journal of Virology* **2003**, *77* (16), 8801-8811. DOI: 10.1128/jvi.77.16.8801-8811.2003.

(14) Kirchdoerfer, R. N.; Cottrell, C. A.; Wang, N.; Pallesen, J.; Yassine, H. M.; Turner, H. L.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Pre-fusion structure of a human coronavirus spike protein. *Nature* **2016**, *531* (7592), 118-121. DOI: 10.1038/nature17200.

(15) Liu, C.; Tang, J.; Ma, Y.; Liang, X.; Yang, Y.; Peng, G.; Qi, Q.; Jiang, S.; Li, J.; Du, L.; et al. Receptor Usage and Cell Entry of Porcine Epidemic Diarrhea Coronavirus. *Journal of Virology* **2015**, *89* (11), 6121-6125. DOI: 10.1128/jvi.00430-15.

(16) Gallagher, T. M.; Buchmeier, M. J. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **2001**, *279* (2), 371-374. DOI: 10.1006/viro.2000.0757.

(17) Wang, W.; McKinnie, S. M. K.; Farhan, M.; Paul, M.; McDonald, T.; McLean, B.; Llorens-Cortes, C.; Hazra, S.; Murray, A. G.; Vederas, J. C.; et al. Angiotensin-Converting Enzyme 2 Metabolizes and Partially Inactivates Pyr-Apelin-13 and Apelin-17 Physiological Effects in the Cardiovascular System. *Hypertension* **2016**, *68* (2), 365-+. DOI: 10.1161/hypertensionaha.115.06892.

(18) Donoghue, M. Hsieh F Baronas E, Godbout K, Gosselin M, Stagliano N, et al. A novel angiotensin converting enzyme related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* **2000**, *87*, E1-E9.

(19) Hamming, I.; Timens, W.; Bulthuis, M. L. C.; Lely, A. T.; Navis, G. J.; van Goor, H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *Journal of Pathology* **2004**, *203* (2), 631-637. DOI: 10.1002/path.1570.

(20) Tipnis, S. R.; Hooper, N. M.; Hyde, R.; Karran, E.; Christie, G.; Turner, A. J. A human homolog of angiotensin-converting enzyme - Cloning and functional expression as a captopril-insensitive carboxypeptidase. *Journal of Biological Chemistry* **2000**, *275* (43), 33238-33243. DOI: 10.1074/jbc.M002615200.

(21) Zhao, Y.; Zhao, Z.; Wang, Y.; Zhou, Y.; Ma, Y.; Zuo, W. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *BioRxiv* **2020**.

(22) Imai, Y.; Kuba, K.; Rao, S.; Huan, Y.; Guo, F.; Guan, B.; Yang, P.; Sarao, R.; Wada, T.; Leong-Poi, H.; et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **2005**, *436* (7047), 112-116. DOI: 10.1038/nature03712.

(23) Dobbs, L. G. Pulmonary surfactant. *Annual review of medicine* **1989**, *40* (1), 431-446.

(24) Ni, W.; Yang, X.; Yang, D.; Bao, J.; Li, R.; Xiao, Y.; Hou, C.; Wang, H.; Liu, J.; Yang, D. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Critical Care* **2020**, *24* (1), 1-10.

(25) Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S. X.; Urbani, C.; Comer, J. A.; Lim, W.; et al. A novel coronavirus associated with severe acute respiratory syndrome. *New England Journal of Medicine* **2003**, *348* (20), 1953-1966. DOI: 10.1056/NEJMoa030781.

(26) Hofmann, H.; Pohlmann, S. Cellular entry of the SARS coronavirus. *Trends in Microbiology* **2004**, *12* (10), 466-472. DOI: 10.1016/j.tim.2004.08.008.

(27) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* **1997**, *48*, 545-600. DOI: 10.1146/annurev.physchem.48.1.545. Pain, R. H. *Mechanisms of protein folding*; IRL Press at Oxford University Press, 1994.

(28) Brooks Iii, C. L. Simulations of protein folding and unfolding. *Current opinion in structural biology* **1998**, *8* (2), 222-226. Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282* (5389), 740-744. DOI: 10.1126/science.282.5389.740.

(29) Kellermayer, M. S. Z.; Smith, S. B.; Granzier, H. L.; Bustamante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **1997**, *276*

(5315), 1112-1116. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997**, *276* (5315), 1109-1112. DOI: 10.1126/science.276.5315.1109.

(30) Zhang, H. B.; Baker, A. Recombinant human ACE2: acing out angiotensin II in ARDS therapy. *Critical Care* **2017**, *21*. DOI: 10.1186/s13054-017-1882-z.

(31) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. An insight into quartz mineral interactions with kerogen in Green River oil shale. *International Journal of Coal Geology* **2021**, 103729.

(32) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Molecular mechanics of the swelling clay tactoid under compression, tension and shear. *Applied Clay Science* **2021**, *200*, 105908.

(33) Jaswandkar, S. V.; Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Dissociation Mechanisms of G-actin Subunits Govern Deformation Response of Actin Filament. *Biomacromolecules* **2021**, *22* (2), 907-917.

(34) Faisal, H. M. N.; Katti, K. S.; Katti, D. R. Differences in Interactions Within Viral Replication Complexes of SARS-CoV-2 (COVID-19) and SARS-CoV Coronaviruses Control RNA Replication Ability. *Jom* **2021**, *73* (6), 1684-1695. DOI: 10.1007/s11837-021-04662-6.

(35) Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics & Modelling* **2001**, *19* (1), 13-25. DOI: 10.1016/s1093-3263(00)00133-9.

(36) Li, F.; Li, W. H.; Farzan, M.; Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **2005**, *309* (5742), 1864-1868. DOI: 10.1126/science.1116480.

(37) Yan, R. H.; Zhang, Y. Y.; Li, Y. N.; Xia, L.; Guo, Y. Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367* (6485), 1444-+. DOI: 10.1126/science.abb2762.

(38) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802, Review. DOI: 10.1002/jcc.20289.

(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(40) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative minimization techniques for abinitio total-energy calculations - molecular-dynamics and conjugate gradients. *Reviews of Modern Physics* **1992**, *64* (4), 1045-1097. DOI: 10.1103/RevModPhys.64.1045.

(41) Towler, P.; Staker, B.; Prasad, S. G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N. A.; et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *Journal of Biological Chemistry* **2004**, *279* (17), 17996-18007. DOI: 10.1074/jbc.M311191200.

(42) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*, Berlin, Heidelberg, 1999//, 1999; Deuflhard, P., Hermans, J.,

Leimkuhler, B., Mark, A. E., Reich, S., Skeel, R. D., Eds.; Springer Berlin Heidelberg: pp 39-65.

(43) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260-+. DOI: 10.1126/science.abb2507.

(44) Kirchdoerfer, R. N.; Wang, N.; Pallesen, J.; Wrapp, D.; Turner, H. L.; Cottrell, C. A.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports* **2018**, *8*. DOI: 10.1038/s41598-018-34171-7.

(45) Geng, Y.-J.; Wei, Z.-Y.; Qian, H.-Y.; Huang, J.; Lodato, R.; Castriotta, R. J. Pathophysiological characteristics and therapeutic approaches for pulmonary injury and cardiovascular complications of coronavirus disease 2019. *Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology* **2020**, *47*, 107228-107228. DOI: 10.1016/j.carpath.2020.107228.

7. DIFFERENCES IN INTERACTIONS WITHIN VIRAL REPLICATION COMPLEXES OF SARS-COV-2 (COVID-19) AND SARS-COV CORONAVIRUSES CONTROL RNA REPLICATION ABILITY[6]

7.1. Introduction

From the beginning of the 21st century, the world has encountered several coronavirus infection outbreaks. During the last twenty years, three different coronavirus outbreaks resulting from zoonotic spillover (transmission from animal to humans) have caused considerable health-related problems, primarily by affecting the human respiratory system and leading to fatal outcome in some cases. The SARS (Severe Acute Respiratory Syndrome), also named novel coronavirus SARS-CoV, emerged from China's Guangdong province in 2002-03 and spread to 26 countries, infecting 8000 people, leading to the death of about 800 people. [1] Another outbreak of zoonotic spillover coronavirus was marked by the emergence of MERS-CoV (Middle East Respiratory Syndrome) in Saudi Arabia in September 2012. The total number of reported cases of MERS-CoV was 2494, with 858 fatalities across 27 countries.[2] Coronavirus Disease 2019 (COVID-19), caused by the novel coronavirus SARS-CoV-2, is the latest and most impactful addition to the list of zoonotic outbreaks. The first incidence of novel coronavirus SARS-CoV-2 was reported in December 2019 in Wuhan, China.[3] Since then, this virus has infected people across 213 countries on all the continents except Antarctica. According to the most recent statistics (as of January 18th, 2021) provided by the World Health Organization (WHO) COVID-19 dashboard, the total number of COVID-19 cases and deaths are 93,611,355 and 2,022,405, respectively.[4] Although SARS-CoV and SARS-COV-2 coronaviruses have similarities in their origin and general structures and both

[6] The contents of this chapter have been published in Faisal, HM Nasrullah, Kalpana S. Katti, and Dinesh R. Katti. "Differences in Interactions Within Viral Replication Complexes of SARS-CoV-2 (COVID-19) and SARS-CoV Coronaviruses Control RNA Replication Ability." *JOM* (2021): 1-12. As the lead author of this journal article, the copyright permission has been obtained from Springer.

target the respiratory system, there are dramatic differences in the infection rates and patient outcomes resulting from infections due to the two coronaviruses. The fatalities because of the SARS-COV-2 virus are 2,022,405 (as of January 18th, 2021) and counting, as against the reported fatalities due to the SARS-COV of about 800.

Coronaviruses are enveloped, large, positive-stranded RNA viruses under the order of *Nidovirales*.[5] This order contains four different virus families, i.e., *Arteriviridae*, *Coronaviridae*, *Mesoniviridae*, and *Roniviridae*. *Coronavirinae* and *Torovirinae* are the constituent sub-families of the *Coronaviridae* family.[6] Alpha, beta, gamma, and delta coronaviruses are the four component genera of *Coronavirinae*.[7] Coronaviruses accommodate the largest multicistronic RNA genome (27-32 kb) containing two open reading frames (ORF1a and ORF1b) at its 5' side.[8] Translation of ORF1a yields the replicative polyprotein 1a (pp1a), while the ribosomal frameshifting of ORF1a and ORF1b make polyprotein 1b (pp1b).[9] Co-translational proteolytic cleavage of polyproteins (pp1a and pp1ab) result in the formation of around 16 non-structural proteins (nsp1-16) that act as the viral replicating machinery.[10] Structural proteins (spike, envelope, membrane, and nucleocapsid) of coronaviruses are predominantly encoded by the 3' side of the RNA genome. [11]

The mechanism of host cell infection by viruses can be divided into five different stages. The very first stage is attachment, when a virus attaches to a host cell through a receptor protein. Then it enters into the host cell either by fusion or endocytosis. Following this, the viral proteins and genome are replicated and assembled inside the cell to make new viruses, followed by their release through lysis or exocytosis. Among these stages, viral replication is considered a critical stage for viral infection of tissues and organs inside the body. RNA-dependent RNA polymerase (RdRp), one of the most conserved elements of the coronaviral life cycle, has been considered the primary tool for the transcription and replication of the virus genome. [12] RdRp encompasses about

two-thirds of the non-structural protein 12 (nsp12). [13] The length and location of nsp12 within the replicase polyprotein affects different features of coronaviruses. [14] Two other non-structural proteins (nsp7 and nsp8) perform as the auxiliary components (cofactors) of the multi-subunit coronaviral replication complex led by nsp12.[15] The participation of cofactors greatly enhances the polymerase activity of RdRp containing nsp12. [16, 17] Due to its RNA replication feature, RdRp has become the focus of pharmaceutical research as a potential drug target for coronavirus infections. [18] For this reason, the use of Remdesivir prodrug, as a suitable ligand of RdRp, has been of interest for SARS-CoV-2 treatment. [17, 19]

SARS-CoV and SARS-CoV-2 diseases share many similarities as both the contributing coronaviruses are bat-originated beta-coronaviruses. [20] Both of them follow the similar host cell attachment mechanism as their spike (S) protein binds to the cellular receptor Angiotensin-Converting Enzyme 2 (ACE2). [21] The nsp12-nsp7-nsp8 complex is utilized as the principal RNA replication/transcription toolbox by SARS-CoV and SARS-CoV-2, where nsp12 consists of RdRp.[16, 17] Though these two coronaviruses pursue almost identical viral lifecycle, the severity of SARS-CoV-2 is being observed as enormously higher than SARS according to the number of reported cases and deaths. Moreover, these two coronaviruses have shown a marked difference in their replication efficiency, and it is still unclear what factors play decisive roles in this aspect.[22] As the nsp12-nsp7-nsp8 complex is considered the hub of viral replication for both coronaviruses, the interactions within this viral protein complex need to be explored.

Remdesivir (GS-5734), an adenosine C–nucleoside prodrug, is considered a broad-spectrum antiviral agent against different RNA viruses. [23] In 2017, Gilead Sciences developed the drug as a potential medication for Ebola viruses.[24] This monophosphoramidate prodrug is effectively metabolized into active nucleoside triphosphate (GS- 443902) when administered into

human cell lines.[25] This active ATP (adenosine triphosphate) analog causes the RNA chain termination by confusing viral RdRp and incorporated into the RNA chain.[25, 26] Remdesivir has attracted researchers' attention again because a recent in-vitro study showed the inhibition of SARS-CoV-2, which utilized it as the inhibitor.[19] Currently, Remdesivir drug has been given emergency approval by the FDA in the United States to treat SARS-CoV-2 infected and hospitalized patients, as Remdesivir treated animals have shown lower lung damage than control animals[27]. As Remdesivir is primarily targeted for the catalytic active site of nsp12 RdRp[28], binding of this prodrug can bring changes in the interactions within the viral complex. Hence, investigating the interactions within the protein subunits of the viral replication complex in the presence of Remdesivir will help to explore its impact on the complex.

Molecular dynamics (MD) simulation is an *in silico* means of inspecting the structure and function of biomolecules.[29] This is a computational technique that predicts the position and motion of atoms and molecules in a time-dependent manner utilizing Newton's second law. MD simulation computes the molecular interactions through molecular mechanics that employs a force field containing energy constants.[30] This study seeks to investigate the nonbonded (electrostatic and Van der Waals) molecular interactions among different subunits and domains inside the viral replication complex (nsp12-nsp7-nsp8) of SARS-CoV and SARS-CoV-2. This study also inspects these interactions when a Remdesivir prodrug (Remdesivir Monophosphate/GS-441524 monophosphate) is in the proximity of RdRp domain of SARS-CoV-2 coronavirus.

7.2. Methodology

The initial three-dimensional models of SARS-CoV and SARS-CoV-2 viral replication complexes (nsp12-nsp7-nsp8 protein complex) have been obtained from RCSB Protein Data Bank (PDB ID 6NUR[16] and 6M71[17] respectively). Each of the models contains one nsp12 polypeptide

chain, one nsp7 chain, and two nsp8 chains. These analogous models are primarily chosen to compare the nonbonded interactions among the protein subunits of corresponding coronaviral replication complexes. The addition of Hydrogen atoms and the assignment of partial charges have been performed by the AutoPSF plugin of Visual Molecular Dynamics (VMD 1.9.4). Next, each of the models was solvated in a water box of 5 Å thickness on all sides around it to approximate the cellular environment. The dimensions of the solvated SARS-CoV and SARS-CoV-2 viral replication complexes are 87.15 Å× 91.74 Å× 126.52 Å and 83.81 Å× 95.22 Å× 117.82 Å, respectively (**Fig. 1a, 1b**). These solvated protein complex models were then simulated using NAMD 2.12.[31] The Theoretical and Computational Biophysics Group developed NAMD at the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. Initially, both models were minimized for achieving stable conformation using the conjugate gradient method.[32] The solvated protein complexes were then brought to 310 K temperature and 1.01325 bar pressure to be consistent with the human physiological condition. The equilibrium condition was achieved at the same temperature and pressure by simulating both models for 5 ns with a timestep of 0.5 fs. Total energy denotes the thermodynamic equilibrium, while Root-Mean-Square Deviation (RMSD) specifies the conformational equilibrium. These equilibrated models were used to compute all the interactions among the subunits and domains within coronaviral replication complexes. CHARMm (Chemistry at HARvard Macromolecular Mechanics) force field was employed for computing all the interaction energies.[33]

To probe the contribution of Remdesivir prodrug on the interactions within the SARS-CoV-2 viral replication complex, a different set of structures have been used. The apo (ligand-free) form (PDB ID 7BV1[28]) and ligand-bound form (PDB ID 7BV2[28]) of SARS-CoV-2 viral complexes are obtained from Protein Data Bank. However, these structures are not analogous as

the apo form (7BV1) contains two chains of nsp8 while ligand-bound form (7BV2) consists of only one chain of nsp8. As the overall structure of apo viral complex (7BV1) is similar to the ligand bound structure (7BV2)[28], we will only utilize the later model both in the ligand-bound and ligand-removed conditions. The protein interactions within the ligand-bound replication complex are compared with those of the ligand-removed complex. Both protein complexes (ligand-bound and ligand-removed) are assigned Hydrogens and partial charges, solvated in 5 Å thick water boxes, minimized, and later equilibrated for 5 ns at the physiological conditions (310 K and 1.01325 bar) utilizing NAMD and CHARMm force field. The initial dimensions of the solvated ligand-bound and ligand-removed complexes are essentially the same (87.75 Å× 92.64 Å× 106.08 Å) as the ligand stays in the cavity of nsp12 (**Fig. 1c, 1d**). All the solvated models consist of TIP3P water molecules.

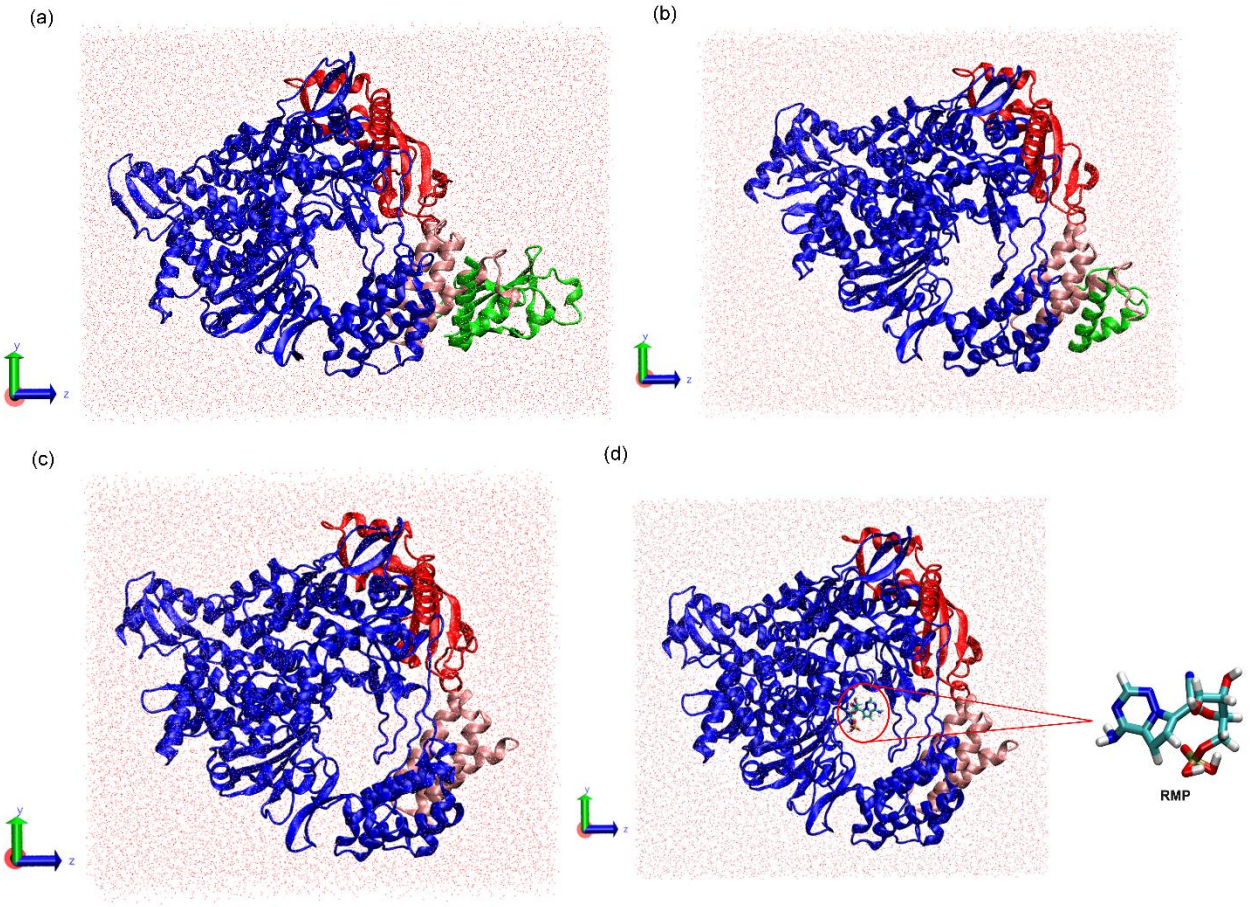

Figure 7.1. Initial solvated models of viral replication complexes (nsp12-nsp7-nsp8) of (a) SARS-CoV (6NUR); (b) SARS-CoV-2 (6M71); (c) SARS-CoV-2 without Remdesivir Monophosphate (RMP) (7BV2) and (d) SARS-CoV-2 with RMP (7BV2) as well as highlighted RMP. In all figures, the blue, red, and pink colored chains represent the full-length nsp12, nsp8, and nsp7, respectively. The green-colored chain represents the second monomer of nsp8 for the first two models (6NUR and 6M71).

Although an ideal ligand should be an active metabolite of Remdesivir (Remdesivir triphosphate/GS-443902), Remdesivir monophosphate (RMP)/ GS-441524 monophosphate has been chosen as the ligand in the present study as it is the only available ligand model experimentally docked into SARS-CoV-2 viral replication complex[28] and deposited into Protein Data Bank as PDB ID 7BV2. Therefore, we will be performing MD simulations on one model of 7BV2 without RMP (Non-RMP model) and another model of 7BV2 with RMP (RMP bound

model). RMP ($C_{12}H_{14}N_5O_7P$) is a small organic molecule with a molecular weight of 371.24 amu. The parameterization of RMP is performed by CHARMm General Force Field (CGenFF) program and the Force Field Toolkit (ffTK) plugin of the Visual Molecular Dynamics (VMD). Initially, the atom types are determined by CGenFF program[33], which is rule-based and programmable. Next, utilizing the ffTK[34] plugin of VMD 1.9.4. and CGenFF topology and parameters, all the atoms are assigned Lennard-Jones parameters according to their atom types. Using the same methodology, the non-polar Hydrogen atoms are assigned a partial charge of +0.09 while all the other atoms are initially assigned a 0.00 charge. Following this step, geometry optimization (Quantum Mechanical calculations) of RMP is performed using the Gausssian09 software package to obtain the lowest energy conformation. This optimized geometry is utilized to derive partial atomic charges of the remaining atoms by reproducing the explicit interactions with TIP3P water molecules. These Quantum Mechanical (QM) water interaction energies of RMP are used to obtain the optimized CHARMm partial charges. The bond stretching and angle bending potentials are derived by performing QM Hessian calculations which employ second derivative of potential energy to model the potential energy surfaces representing the bond and angle distortions. The dihedral parameters of RMP are obtained by QM scanning of the torsional potential energy surfaces. The water interaction energy calculations utilize the Hartree-Fock (HF) theory, while all the bonded parameter potential energy calculations utilize the second-order Møller–Plesset perturbation theory (MP2). All the QM calculation inputs are generated by ffTK plugin, and the simulations are conducted using Gaussian09 software.

We have used the Center for Computationally Assisted Science and Technology (CCAST), a supercomputing cluster at North Dakota State University, to perform all the MD and QM

simulations. Each of the simulations was run on one node and twenty Intel Xeon 2.5 GHz processors with 20 GB DDR3 RAM at 1866 MHz.

7.3. Results

7.3.1. SARS-CoV Replication Complex

The equilibrium structure of the SARS-CoV viral replication complex consists of one nsp12 chain (A), one nsp7 chain (C), and two nsp8 chains (B, D) (**Fig. 2**).[16] The polypeptide chain of nsp12 (A) is 793 amino acids long as obtained from the Protein Data Bank (6NUR). The length of the nsp7 chain (C) and np8 chain (B) is 70 and 115 residues, respectively. The second subunit of nsp8 (chain D) is 109 residues long. The structure of nsp12 can be divided into three different domains, i.e., nucleotidyltransferase (NiRAN), interface, and RdRp domain. The nidovirus RdRp-associated NiRAN domain is positioned at the N-terminal of nsp12 (residues 117-250) while the RNA-dependent RNA polymerase (RdRp) domain lies at its C-terminal (residues 399-920). The interface domain (residues 251-398) is the connecting domain between the NiRAN and RdRp domain. RdRp domain performs the primary polymerase activity of SARS-CoV coronavirus with the assistance of cofactors.[16] The nonbonded interactions (electrostatic and Van der Waals) have been computed within this SARS-CoV viral replication complex to better understand the relationship among different nsps and domains.

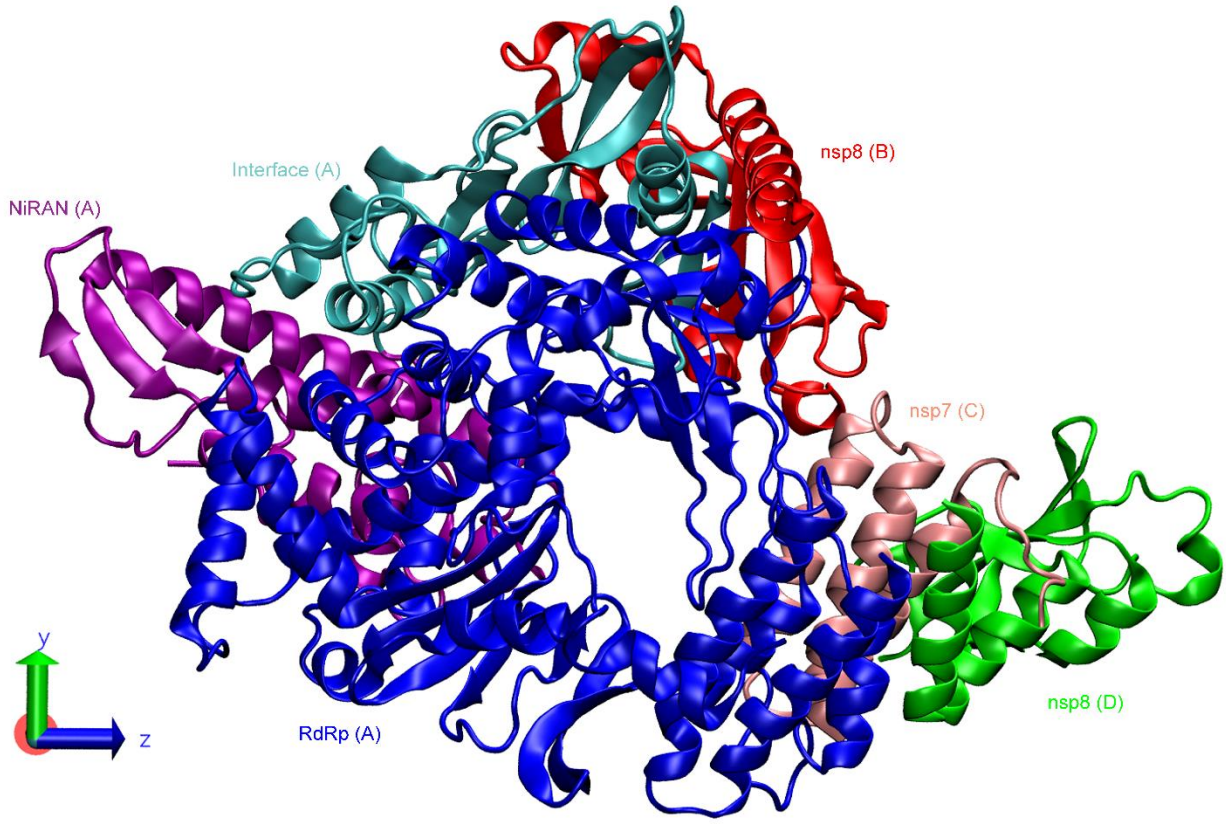

Figure 7.2. Equilibrated structure of SARS-CoV (6NUR) viral replication complex (nsp12-nsp7-nsp8). The purple, cyan, and blue colored segments represent the NiRAN, Interface, and RNA-dependent RNA polymerase (RdRp) domains of nsp12 (chain A). The red and green chains (chain B and D, respectively) represent two monomers of nsp8. The nsp7 is represented by a pink-colored chain (chain C). For the convenience of visualization, the water molecules are not shown.

The total nonbonded interactions between nsp12 and its cofactors (nsp7 and nsp8) are -812.8 kcal/mol where electrostatic, and Van der Waals (VDW/VdW) interactions are -509.1 kcal/mol and -303.7 kcal/mol, respectively (**Supplementary Table S-I**). Three different residues (GLU, ASP, LEU) of nsp12 contribute to more than half of these interaction energies (53.1%) by interacting with the cofactors. The amount of non-bonded interaction energies contributed by GLU, ASP and LEU of nsp12 are -209.9 kcal/mol (25.8%), -122.3 kcal/mol (15.1%) and -105.2 kcal/mol (12.9%) respectively. PHE, ASN, ARG, PRO, and TYR are other significantly interacting residues, i.e., each of them contributed to more than 4% of the total interactions (**Fig.**

3a, 3d). In terms of the nsp12 secondary structure, the sheets interact with the cofactors most (-266.6 kcal/mol) while coils interact the least (-145 kcal/mol) (**Fig. 3b, 3e**). About two-thirds of all the interaction energies (-530.4 kcal/mol) result from the polar residues of nsp12, while the rest (-282.4 kcal/mol) are caused by the hydrophobic residues (**Fig. 3c, 3f**).

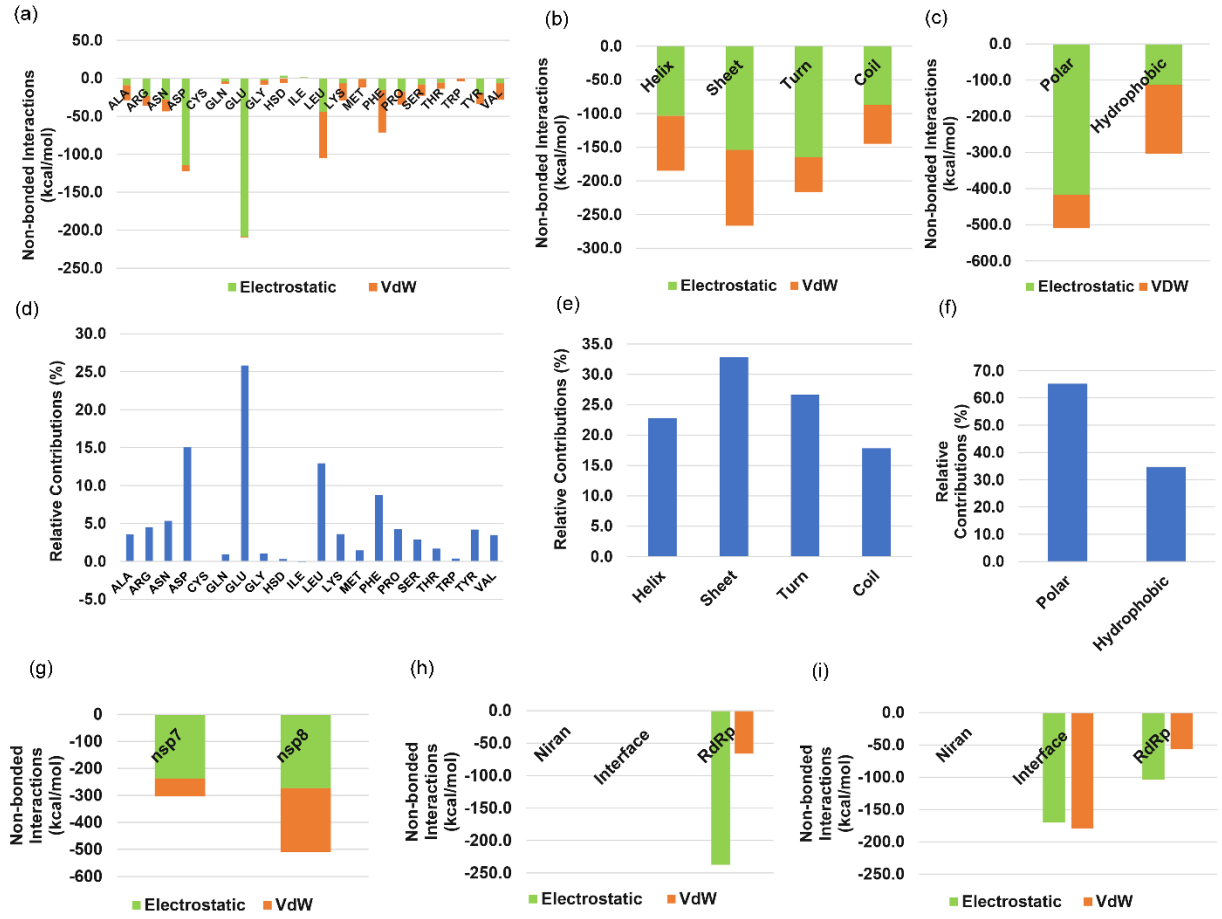

Figure 7.3. Non-bonded interaction (electrostatic and Van der Waals) energies of nsp12 with the co-factors (nsp7 and nsp8) in terms of its (a) primary (b) secondary and (c) tertiary motifs within the equilibrated SARS-CoV viral replication complex (6NUR); Relative contribution (proportion) of interactions made by each (d) primary; (e) secondary and (f) tertiary motif of nsp12; (g) Total non-bonded interactions of nsp12 with nsp7 and nsp8; (h) domain-wise interactions of nsp12 with nsp7; (i) domain-wise interactions of nsp12 with nsp8.

The total non-bonded interactions of nsp12 with its co-factors (-812.8 kcal/mol) are divided into nsp12-nsp7 interactions (-303.5 kcal/mol) and nsp12-nsp8 interactions (-509.3 kcal/mol) (**Fig. 3g**). The total interaction energies between nsp7 and nsp8 are found to be -436.9 kcal/mol. The

total interaction energies of nsp12 with cofactors are also analyzed in terms of the three domains of nsp12. The NiRAN domain does not interact with the cofactors (**Fig. 3h, 3i**). The interface domain of nsp12 contributes to 43% (-349.5 kcal/mol) while RdRp domain contributes to 57% (-463.3 kcal/mol) of total nsp12-cofactor interaction energies. The interface domain of nsp12 does not interact with nsp7 (**Fig. 3h)**. The interaction energies of the RdRp domain of nsp12 with nsp7 and nsp8 have been calculated as -303 kcal/mol and -160.3 kcal/mol, respectively.

7.3.2. SARS-CoV-2 Replication Complex

The replication complex of SARS-CoV-2 has a similar structure to SARS-CoV. The equilibrated SARS-CoV-2 replication complex contains one chain of nsp12 (A), one chain of nsp7 (C), and two chains of nsp8 (B&D).[17] The nsp12 chain (A) consists of 851 residues, while the nsp7 chain (C) contains 70 residues. The chain B of nsp8 has 113 residues. The single monomer of nsp8 (chain D) comprises only 43 residues. NiRAN (residues 60-249), interface (residues 250-365), and RdRp (residues 366-920) domains create the nsp12 structure (**Fig. 4**). The interface domain serves as a connecting domain between NiRAN and RdRp. Like SARS-CoV, the RdRp domain of SARS-CoV-2 is primarily responsible for viral replication in the presence of cofactors.[17]

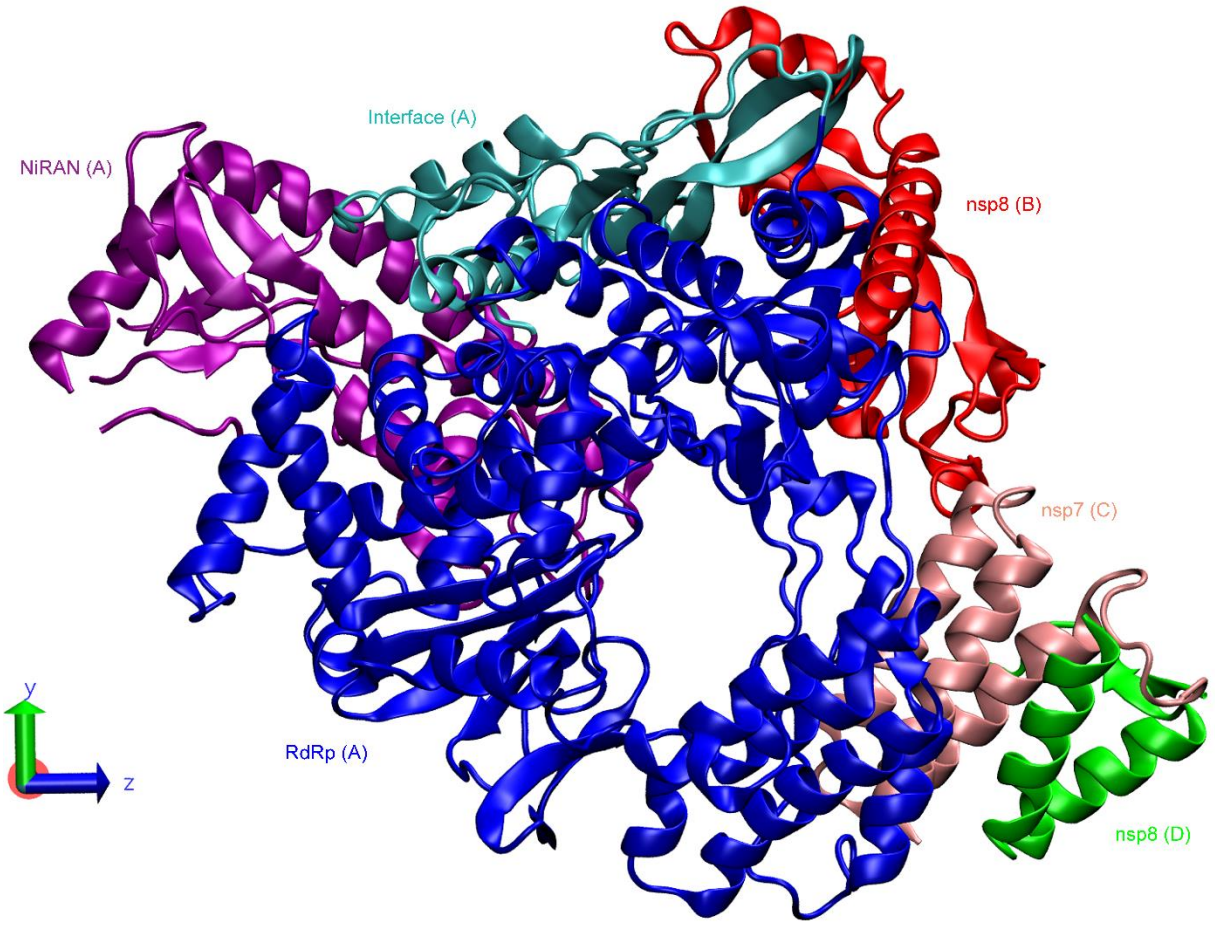

Figure 7.4. Equilibrated structure of SARS-CoV-2 (6M71) viral replication complex (nsp12-nsp7-nsp8). The purple, cyan, and blue colored segments represent the NiRAN, Interface, and RNA-dependent RNA polymerase (RdRp) domains of nsp12 (chain A). The red and green chains (chain B and D, respectively) represent two monomers of nsp8. The nsp7 is represented by a pink-colored chain (chain C). For the convenience of visualization, the water molecules are not shown.

The nsp12 monomer interacts with cofactors (nsp7 and nsp8) with nonbonded energy of -672 kcal/mol (**Supplementary Table S-IV**). Electrostatic interactions contributed 57% (-384 kcal/mol) of the interaction energies while VDW furnished the rest (-288 kcal/mol). ASP (-195.5 kcal/mol), LEU (-105.3 kcal/mol) and GLU (-66.8 kcal/mol) residues of nsp12 together result in about 54.7% of the interaction energies. ALA, ASN, LYS, PHE, PRO, and VAL residues of nsp12 individually provide at least 4% of the interaction energies (**Fig. 5a, 5d**). The helices of nsp12 interact with cofactors with an energy of -223.2 kcal/mol, which is 33.2% of the total interactions.

The sheets of nsp12 interact in almost the same order as of helices (-211.7 kcal/mol) followed by coils (-138.2 kcal/mol) (**Fig. 5b, 5e**). Polar residues of nsp12 yield approximately 59% (-396.4 kcal/mol) of the interaction energies (**Fig. 5c, 5f**).

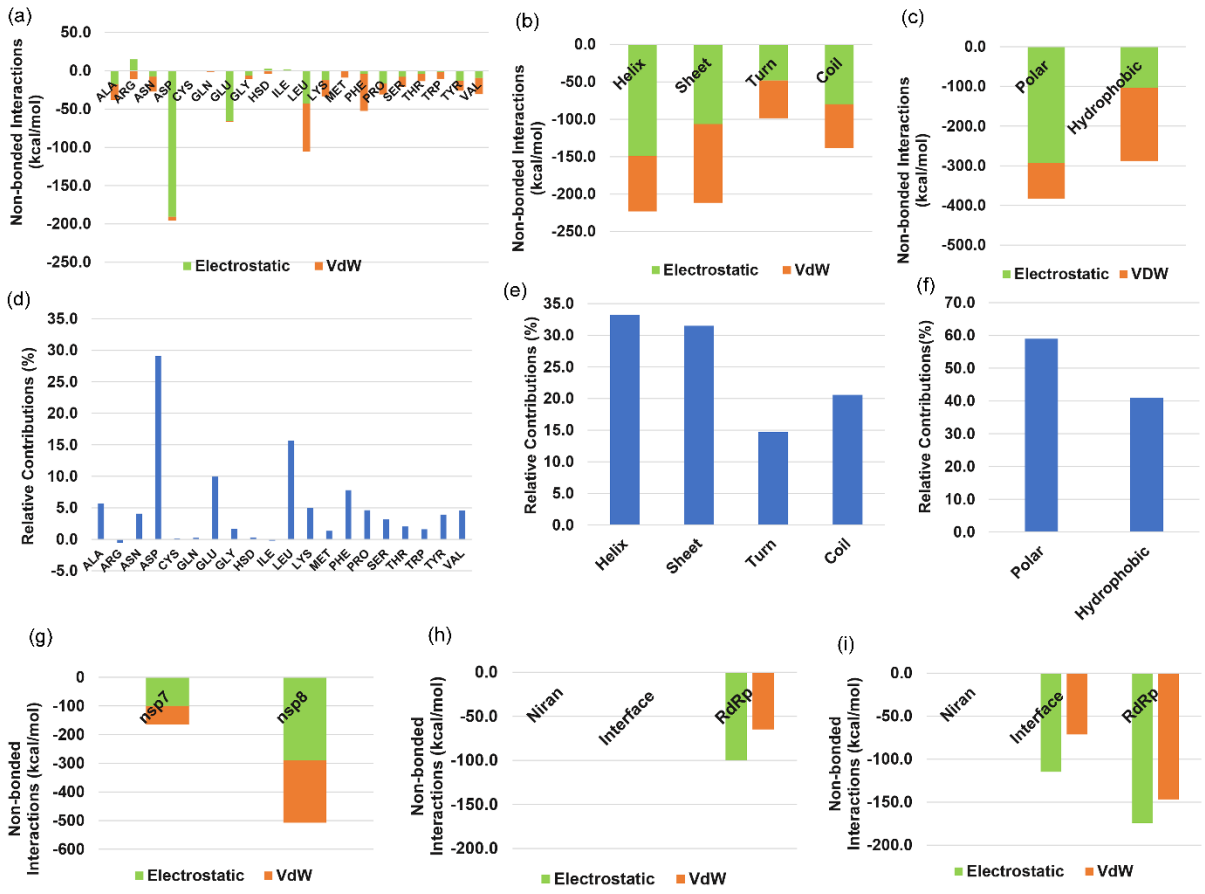

Figure 7.5. Nonbonded interaction energies (electrostatic and Van der Waals) of nsp12 with the cofactors (nsp7 and nsp8) in terms of its (a) primary; (b) secondary and (c) tertiary motifs within the equilibrated SARS-CoV-2 viral replication complex (6M71). Relative contribution (proportion) of interactions made by each (d) primary; (e) secondary and (f) tertiary motif of nsp12; (g) Total non-bonded interactions of nsp12 with nsp7 and nsp8; (h) domain-wise interactions of nsp12 with nsp7; (i) domain-wise interactions of nsp12 with nsp8.

The magnitude of nonbonded interactions between nsp12-nsp7 (A-C) and nsp12-nsp8 (A-BD) have been computed as -164.7 kcal/mol and -507.3 kcal/mol, respectively (**Fig. 5g**). The nonbonded interactions between nsp7 and nsp8 are computed as -476.9 kcal/mol. The NiRAN domain interacts with neither of the cofactors, while the interface domain interacts with only nsp8

(-186.1 kcal/mol) (**Fig. 5h, 5i**). The RdRp domain of nsp12 yields 72.3% (-485.9 kcal/mol) of its total interactions with the cofactors. The interactions of RdRp with nsp7 and nsp8 have been calculated as -164.7 kcal/mol and -321.2 kcal/mol, respectively.

7.3.3. Non-RMP and RMP-Bound SARS-CoV-2 Replication Complex

This viral replication complex of SARS-CoV-2 consists of one nsp12 chain (A), one nsp8 chain (B), and one nsp7 chain (C) (7BV2). The nsp12 chain is composed of NIRAN (residues 115-250), interface (residues 251-365), and RdRp (residues 366-920) domains.[28] Remdesivir monophosphate (RMP), the potent antiviral prodrug, is removed from the 'Non-RMP' model of replication complex while kept in the native conformation in the 'RMP-bound' model of replication complex. The CHARMm compatible parameters of RMP, obtained through rigorous QM calculations and multidimensional optimization, are provided in the **Supplementary Materials (refer to online supplementary material, Supplementary Table S-XV to S-XIX).**

In the equilibrated 'Non-RMP' model (**Fig 6a**), the total interaction energy of nsp12 with its cofactors (nsp7 and nsp8) is found to be -929 kcal/mol. Almost 80% of the interaction energies (-742.2 kcal/mol) are contributed by nsp12-nsp8 interactions while nsp12-nsp7 generates the rest (-186.8 kcal/mol) (**Fig. 6c**). The NiRAN domain of nsp12 has no interactions with cofactors, while the interface domain only interacts with nsp8 (-317 kcal/mol) (**Fig. 6d**). The RdRp domain is responsible for all the interactions of nsp12 with nsp7 (-186.8 kcal/mol), while this domain provides a relatively higher portion (58%) of interactions (-425.2 kcal/mol) between nsp12 and nsp8 compared to the interface domain (42%). The interactions between nsp7 and nsp8 are computed as -87.8 kcal/mol.

In the equilibrated 'RMP-bound' model (**Fig. 6b**), the antiviral ligand RMP stays in the proximity of the catalytic active site of nsp12 RdRp. In this model, the interactions of nsp12-nsp8

and nsp12-nsp7 are determined as -566.7 kcal/mol and -298.6 kcal/mol, respectively (**Fig. 6e**), leading to a total nsp12-cofactor interaction of -865.3 kcal/mol. The nsp12-nsp7 interactions are fully contributed by RdRp while nsp12-nsp8 interactions are contributed by both interface (-188.8 kcal/mol) and RdRp (-377.9 kcal/mol) domains (**Fig. 6f**). The nsp7-nsp8 interactions are determined as -85.5 kcal/mol. The antiviral prodrug RMP is only observed to have interactions with nsp12 RdRp. The total nonbonded interaction energy between the RdRp domain and RMP is determined as -206.9 kcal/mol, 95% of which result from electrostatic interactions.

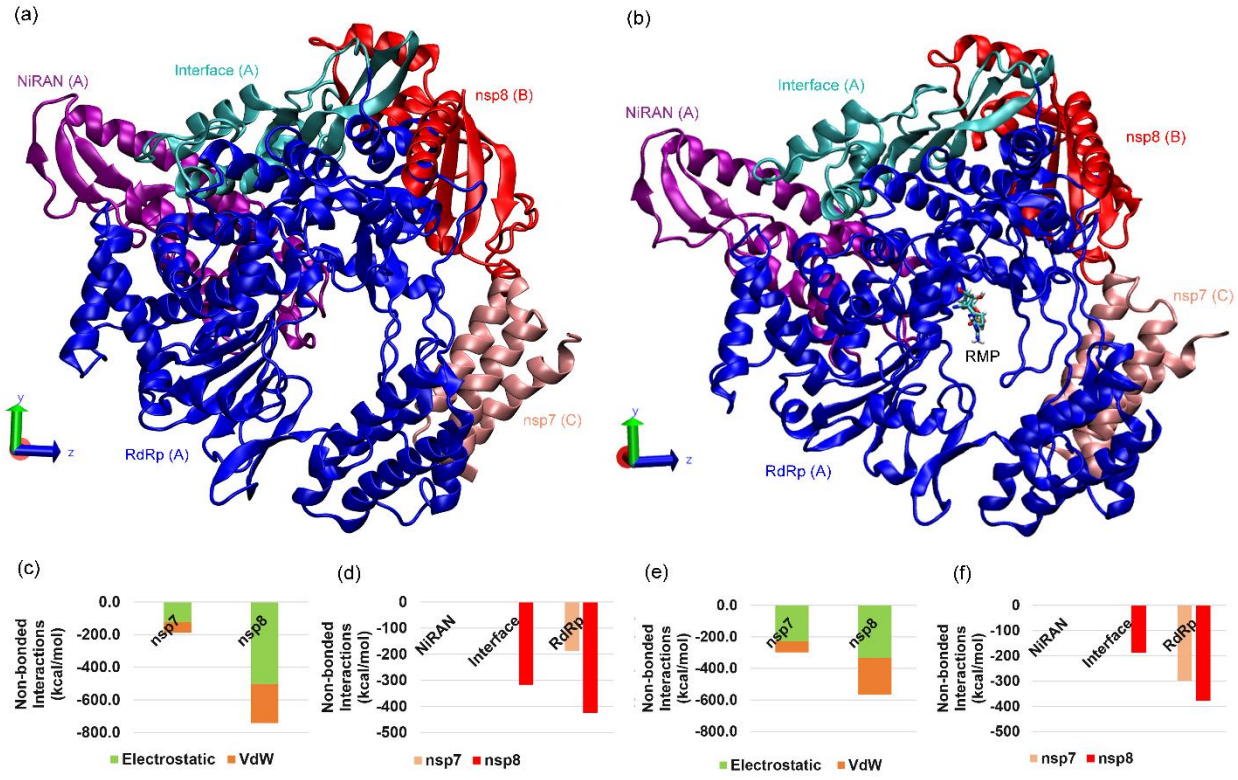

Figure 7.6. Equilibrated structures of (a) Non-RMP and (b) RMP-bound SARS-CoV-2 (7BV2) viral replication complexes (nsp12-nsp7-nsp8). The purple, cyan, and blue colored segments represent the NiRAN, Interface, and RNA-dependent RNA polymerase (RdRp) domains of nsp12 (chain A). The red and pink chains (chain B and C) represent nsp8 and nsp7, respectively. The small licorice structure is identified as RMP in the 'RMP-bound' model. For the convenience of visualization, the water molecules are not shown; (c) Total nonbonded interactions of nsp12 with nsp7 and nsp8 in the 'Non-RMP' model; (d) Domain-wise nonbonded interactions of nsp12 with nsp7 and nsp8 in the 'Non-RMP' model; (e) Total nonbonded interactions of nsp12 with nsp7 and nsp8 in the 'RMP-bound' model; (f) Domain-wise nonbonded interactions of nsp12 with nsp7 and nsp8 in the 'RMP-bound' model.

7.4. Discussion

Protein-protein interactions are at the core of most physiological processes in the biological realm [35]. Viral protein-protein interactions dictate different host cell infection stages, i.e., replication and assembly of virions and their resistance against the host immune system [36]. These protein-protein interactions are divided into covalent (bonded) and non-covalent (nonbonded) interactions at the molecular level. Non-covalent (nonbonded) interactions play critical roles in genome replication, protein folding, signal transduction, enzyme detection, and many other physicochemical processes [37]. Therefore, investigating the nonbonded viral protein-protein interactions in coronaviruses will help better perceive the biological activities (i.e., replication) carried out by the viral replication protein complexes.

SARS-CoV and SARS-CoV-2 coronaviruses comprise the analogous structure of the viral replication complex (nsp12-nsp7-nsp8). In both cases of coronaviruses, a peptide chain of nsp12 (A) performs as the principal replication subunit while one chain of nsp7 (C) and two chains of nsp8 (B, D) act as the replication cofactors.[16, 17] The structural conformation of cofactors can be described as a hetero-dimer of nsp7-nsp8 (C-B) along with an nsp8 monomer (D) for both coronaviral replication complexes. Significant attractive nonbonded (electrostatic and VDW) interactions are observed among the subunits of the replication complex. The nonbonded interactions of SARS-CoV nsp12 with its cofactors (-812.8 kcal/mol) are higher than the interactions of SARS-CoV-2 nsp12 with its cofactors (-672 kcal/mol). Electrostatic interactions yield a greater portion of the nonbonded interactions for both SARS-CoV (62.6%) and SARS-CoV-2 (57.1%). The peptide chain of nsp12 accomplishes more than half of its interactions with cofactors through three different residues, i.e., ASP, GLU, and LEU. GLU is the highest interacting (-209.9 kcal/mol) residue of SARS-CoV nsp12, and ASP is the highest interacting residue (-195.5

kcal/mol) of SARS-CoV-2 nsp12. CYS has been found as the least interacting residue of nsp12 for both coronaviruses. Regarding the secondary structure of nsp12, the highest portion of its interactions is undertaken by beta sheets (-266.6 kcal/mol) and helices (-223.2 kcal/mol) for SARS-CoV and SARS-CoV-2, respectively. The relative contribution by nsp12 sheets remained identical (~32%) for both coronaviruses. Polar residues of SARS-CoV nsp12 are attributed to relatively higher interactions (65.3%) than those of SARS-CoV-2 nsp12 (59%).

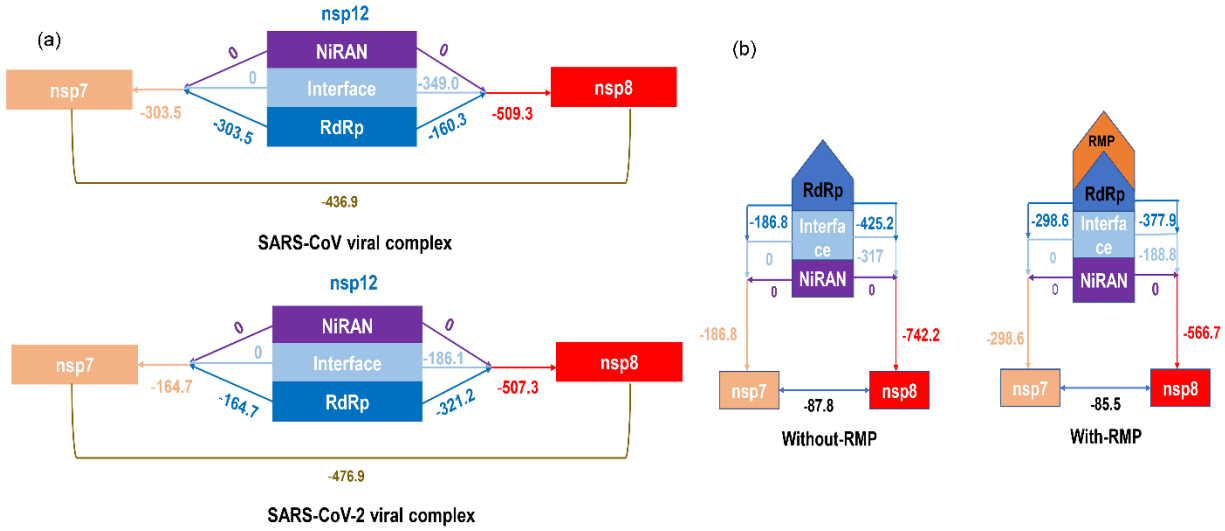

Figure 7.7. Comparison of domain-wise interaction energies of nsp12 with cofactors (nsp7 and nsp8) between equilibrated (a) SARS-CoV (6NUR) and SARS-CoV-2 (6M71) viral replication complexes and (b) Non-RMP (without RMP) and RMP-bound (with RMP) SARS-CoV-2 (7BV2) viral replication complexes. All energies have the unit of kcal/mol.

The interactions of nsp12 with cofactors are analyzed to have a better understanding of the viral replication complexes. The interactions between nsp12 (A) and nsp7 (C) are significantly higher for SARS-CoV (-303.5 kcal/mol) compared to SARS-CoV-2 (-164.7 kcal/mol) (**Fig. 7a**). The SARS-CoV nsp12 interacts with two nsp8 chains (B, D) almost as equally as SARS-CoV-2, i.e., the nsp12-nsp8 interactions are -509.3 kcal/mol and -507.3 kcal/mol for SARS-CoV and SARS-CoV-2 respectively. There are both similarities and differences between the coronaviruses in terms of their domain-wise nsp12 interactions with cofactors. The NiRAN domain does not

interact with cofactors in the replication complex of both coronaviruses. The majority of coronavirus nsp12 interactions (with cofactors) are performed by its RdRp domain. The RdRp domain of nsp12 carries out the primary polymerase activity in the presence of cofactors. The RdRp domain of SARS-CoV-2 exhibits similar interactions (-485.9 kcal/mol) with the two cofactors compared to the SARS-CoV RdRp (-463.3 kcal/mol). However, the proportion of RdRp domain interactions with cofactors towards the total nsp12 interactions is significantly higher for SARS-CoV-2 (72.3%) than SARS-CoV (57%). The interactions of SARS-CoV-2 RdRp with nsp7 co-factor (-164.7 kcal/mol) are 45.6% lower than that of SARS-CoV (-303.5 kcal/mol). On the contrary, the interactions of SARS-CoV-2 RdRp with nsp8 cofactor (-321.2 kcal/mol) are approximately 100% higher than that of SARS-CoV (-160.3 kcal/mol). Thus, significant differences are observed in the nsp12 RdRp domain interactions with the cofactors nsp7 and nsp8, between SARS-COV and SARS-COV-2.

The replication and transcription procedure of the coronavirus RNA genome is guided by the replication complex interactions (nsp12-nsp7-nsp8).[15, 16] Therefore, the severity of coronavirus infections can result from a higher replication/transcription rate inside the host cells. SARS-CoV and SARS-CoV-2 coronaviruses have shown different replication trends in different human cells. SARS-CoV exhibits higher replication in intestinal Caco2 cells, while SARS-CoV-2 demonstrates higher virus production in pulmonary Calu3 cells.[38] SARS-CoV-2 has also been reported to replicate more extensively in bronchial epithelium compared to SARS-CoV. This particular observation has been considered as one of the reasons for the higher transmissibility of SARS-CoV-2.[39] Although it is known that the interactions between the non-structural protein complex (nsp12-nsp7-nsp8) are key to the replication of viral RNA in cells, there is little known about how the RdRp domain interactions specifically with nsp7 or nsp8 contribute to replication in different

types of cells. It might be possible that higher RdRp-nsp7 interactions can contribute to the higher replication of SARS-CoV RNA in specific organs or cell lines. On the other hand, higher replication of SARS-CoV-2 RNA in certain parts of the body can be related to higher RdRp-nsp8 interactions. The results from this work suggest investigating the replication rates and their relationship to RdRp domain interactions with nsp7 and nsp8 for different cell types through carefully crafted experiments. It is important to note that the differences in amino acid compositions of nsp12 between SARS-CoV and SARS-CoV-2 do not contribute to their interactions as they share about 96% sequence identity, and the different residues mostly lie in distant regions from the cofactors (**supplementary Fig. S-2**). The variations in the non-structural proteins' interactions may be attributed to their conformation and relative orientations within the viral replication complex.

Comparison between the 'Non-RMP' and 'RMP-bound' SARS-CoV-2 viral replication complexes (7BV2) indicates the impact of antiviral prodrug RMP on interaction energies inside the RdRp complex. The presence of RMP reduces the interaction energy between nsp12 and nsp8 from -742.2 kcal/mol to -566.7 kcal/mol (**Fig. 7b**). This reduction in interaction energy can also be seen within the interacting individual domains of nsp12. The RdRp domain of nsp12 interacts less (-377.9 kcal/mol) with nsp8 in the RMP-bound model compared to the Non-RMP model (-425.2 kcal/mol). The study results comparing SARS-CoV and SARS-CoV-2 presented earlier have indicated that lower RdRp domain-nsp8 interactions in SARS-CoV versus SARS-CoV-2 may have contributed to the lower replication of RNA in SARS-CoV compared to SARS-CoV-2. Based on this co-relationship, the reduction of RdRp-nsp8 interactions in the presence of RMP can reduce coronaviral replication in case of SARS-CoV-2 infection. The presence of RMP also decreases the interactions between nsp8 and interface domain significantly while increasing the interactions

between RdRp and nsp7. In a recent study, Remdesivir has been shown to inhibit SARS-CoV-2 replication in human lung cells and the primary epithelial airways.[40] The primary inhibition mechanism of active Remdesivir metabolite (Remdesivir Triphosphate/RTP) is to terminate RNA chain synthesis by interfering with the nsp12 polymerase(RdRp) activity.[41] The presence of RMP ligand in the catalytic active site of RdRp modifies its interactions with cofactors, which may lead to the disruption of its regular polymerase activity.

In summary, SARS-CoV nsp12 has higher non-bonded interaction energies (-812 kcal/mol) with co-factors (nsp7 and nsp8) compared to SARS-CoV-2 (-672 kcal/mol). The coronaviral nsp12 consists of three different domains: nucleotidyltransferase (NiRAN), interface, and RNA-dependent RNA polymerase (RdRp) domain. Among these three domains, only the RdRp domain is attributed for viral replication [23,24]. Therefore, only the interactions between the RdRp domain of nsp12 and co-factors need to be considered for viral RNA replication. SARS-CoV RdRp has lower interactions (-463.3 kcal/mol) with co-factors compared to SARS-CoV-2 (-485.9 kcal/mol). This finding is consistent with. higher infectious behavior of SARS-CoV-2. Moreover, by analyzing the pairwise interactions of RdRp with individual co-factor nsp8, it has been observed that SARS-CoV-2 RdRp-nsp8 interaction (-321.2 kcal/mol) is almost twice of that of SARS-CoV (-160.3 kcal/mol). This specific interaction (RdRp-nsp8) might play a key role in higher replication of SARS-CoV-2 in bronchial epithelium which in turn may lead to higher transmissibility of SARS-CoV-2 [55]. The presence of RMP within SARS-CoV-2 viral replication complex has been observed to reduce the interactions between RdRp and nsp8. This reduction in interactions can contribute to the reduction of viral RNA replication activity.

7.5. Conclusions

The invasion of new host cells in coronavirus infections depends on the viral replication complex's virus reproduction process. This coronaviral replication complex comprises three non-structural proteins (nsp12-nsp7-nsp8) that replicate and transcribe new RNA genomes. For the first time, we have compared differences in interactions of viral replication complexes in SARS-CoV and SARS-CoV-2 using molecular dynamics simulations. As Remdesivir prodrug is primarily targeted for the catalytic active site of nsp12 RdRp. Remdesivir prodrug (RMP) treated animals have shown lower lung damage compared to control animals. Here we have also evaluated the impact of Remdesivir prodrug (RMP) on the SARS-CoV-2 viral complex to find the differences between the interactions within the nsps with and without RMP. The significant findings of this study are listed below:

I. In both viral replication complexes, the peptide chain of nsp12 contributes to more than half of its interactions with cofactors (nsp7 and nsp8) through ASP, GLU, and LEU residues.

II. Sheets are the most interacting secondary motifs of SARS-CoV nsp12, while in SARS-CoV-2, helices contribute to the highest interactions.

III. The interaction energies between the nsps in the viral replication complex are significantly different, potentially a clue for the differences seen in the RNA replication rates exhibited by the two coronaviruses. The RdRp-nsp7 interaction energies of SARS-CoV are twice those of SARS-CoV-2. On the other hand, the RdRp-nsp8 interaction energies of SARS-CoV are half of those of SARS-CoV-2.

IV. Remdesivir Monophosphate (RMP) reduces the RdRp-nsp8 interactions of the SARS-CoV-2 viral complex. Reduced RdRp-nsp8 interactions may contribute to

reduced RNA replication rates in SARS-CoV-2, consequently resulting in reduced infections and mortality.

V. This work points towards the potential for using the relative differences in nonbonded interactions between nsps as an indicator of viral RNA replication ability in coronavirus infections. Well-designed experiments based on the simulation results would provide further credence to this assertion.

7.6. Acknowledgments

NSF grants #1229316 and #2019077 are acknowledged for the support of the computational resources at NDSU CCAST. The authors also acknowledge the partial support of the National Science Foundation NSF EPSCoR Track-1 Cooperative Agreement OIA #1946202 and the NDSU Grand Challenge funded Center for Engineered Cancer Testbeds.

7.7. References

(1) Chan-Yeung, M.; Xu, R. H. SARS: epidemiology. Respirology 8 Suppl S9–14. 2003.

(2) Lee, J.; Chowell, G.; Jung, E. A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: A retrospective analysis on control interventions and superspreading events. *Journal of Theoretical Biology* **2016**, *408*, 118-126. DOI: 10.1016/j.jtbi.2016.08.009.

(3) Tan, W.; Zhao, X.; Ma, X.; Wang, W.; Niu, P.; Xu, W.; Gao, G. F.; Wu, G. Z. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019− 2020. *China CDC Weekly* **2020**, *2* (4), 61-62.

(4) (WHO), W. H. O. Weekly Operational Update on COVID-19. 2021.

(5) Lai, M. M. C.; Holmes, K. V. Fields Virology, edited by DM Knipe & PM Howley. Philadelphia: Lippincott Williams & Wilkins: 2001.

(6) Draker, R.; Roper, R. L.; Petric, M.; Tellier, R. The complete sequence of the bovine torovirus genome. *Virus Research* **2006**, *115* (1), 56-68. DOI: 10.1016/j.virusres.2005.07.005.

(7) Woo, P. C. Y.; Lau, S. K. P.; Lam, C. S. F.; Lau, C. C. Y.; Tsang, A. K. L.; Lau, J. H. N.; Bai, R.; Teng, J. L. L.; Tsang, C. C. C.; Wang, M. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of virology* **2012**, *86* (7), 3995-4008.

(8) Nga, P. T.; del Carmen Parquet, M.; Lauber, C.; Parida, M.; Nabeshima, T.; Yu, F.; Thuy, N. T.; Inoue, S.; Ito, T.; Okamoto, K. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog* **2011**, *7* (9), e1002215.

(9) Masters, P. S. The molecular biology of coronaviruses. *Advances in virus research* **2006**, *66*, 193-292. Firth, A. E.; Brierley, I. Non-canonical translation in RNA viruses. *The Journal of general virology* **2012**, *93* (Pt 7), 1385.

(10) Snijder, E. J.; Bredenbeek, P. J.; Dobbe, J. C.; Thiel, V.; Ziebuhr, J.; Poon, L. L. M.; Guan, Y.; Rozanov, M.; Spaan, W. J. M.; Gorbalenya, A. E. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of molecular biology* **2003**, *331* (5), 991-1004.

(11) King, A. M. Q.; Lefkowitz, E.; Adams, M. J.; Carstens, E. B. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*; Elsevier, 2011.

(12) Ahn, D.-G.; Choi, J.-K.; Taylor, D. R.; Oh, J.-W. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Archives of virology* **2012**, *157* (11), 2095-2104.

(13) Velthuis, A.; Arnold, J. J.; Cameron, C. E.; van den Worm, S. H. E.; Snijder, E. J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Research* **2010**, *38* (1), 203-214. DOI: 10.1093/nar/gkp904.

(14) Snijder, E. J.; Decroly, E.; Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Advances in Virus Research, Vol 96: Coronaviruses* **2016**, *96*, 59-126. DOI: 10.1016/bs.aivir.2016.08.008.

(15) Subissi, L.; Posthuma, C. C.; Collet, A.; Zevenhoven-Dobbe, J. C.; Gorbalenya, A. E.; Decroly, E.; Snijder, E. J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111* (37), E3900-E3909. DOI: 10.1073/pnas.1323705111.

(16) Kirchdoerfer, R. N.; Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature Communications* **2019**, *10*. DOI: 10.1038/s41467-019-10280-3.

(17) Gao, Y.; Yan, L. M.; Huang, Y. C.; Liu, F. J.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q. Q.; Ming, Z. H.; Zhang, L. Q.; et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020**, *368* (6492), 779-+. DOI: 10.1126/science.abb7498.

(18) Guan, W.-j.; Ni, Z.-y.; Hu, Y.; Liang, W.-h.; Ou, C.-q.; He, J.-x.; Liu, L.; Shan, H.; Lei, C.-l.; Hui, D. S. C. Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine* **2020**, *382* (18), 1708-1720.

(19) Wang, M.; Cao, R.; Zhang, L. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019 nCoV) in vitro [published online February 4, 2020]. *Cell Res. doi 10*.

(20) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**. DOI: 10.1038/s41586-020-2012-7.

(21) Li, W. H.; Moore, M. J.; Vasilieva, N.; Sui, J. H.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greenough, T. C.; et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426* (6965), 450-454. DOI: 10.1038/nature02145.

(22) V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* **2020**, 1-16.

(23) Siegel, D.; Hui, H. C.; Doerffler, E.; Clarke, M. O.; Chun, K.; Zhang, L.; Neville, S.; Carra, E.; Lew, W.; Ross, B. Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo [2, 1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. ACS Publications: 2017.

(24) Al-Tawfiq, J. A.; Al-Homoud, A. H.; Memish, Z. A. Remdesivir as a possible therapeutic option for the COVID-19. *Travel medicine and infectious disease* **2020**.

(25) Warren, T. K.; Jordan, R.; Lo, M. K.; Ray, A. S.; Mackman, R. L.; Soloveva, V.; Siegel, D.; Perron, M.; Bannister, R.; Hui, H. C. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **2016**, *531* (7594), 381-385.

(26) Tchesnokov, E. P.; Feng, J. Y.; Porter, D. P.; Götte, M. Mechanism of inhibition of Ebola virus RNA-dependent RNA polymerase by remdesivir. *Viruses* **2019**, *11* (4), 326.

(27) Williamson, B. N.; Feldmann, F.; Schwarz, B.; Meade-White, K.; Porter, D. P.; Schulz, J.; van Doremalen, N.; Leighton, I.; Yinda, C. K.; Pérez-Pérez, L.; et al. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nature* **2020**, *585* (7824), 273-276. DOI: 10.1038/s41586-020-2423-5 From NLM.

(28) Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **2020**.

(29) Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **1999**, *28* (1), 155-179.

(30) Young, D. *Computational chemistry: a practical guide for applying techniques to real world problems*; John Wiley & Sons, 2004.

(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289.

(32) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative minimization techniques for abinitio total-energy calculations - molecular-dynamics and conjugate gradients. *Reviews of Modern Physics* **1992**, *64* (4), 1045-1097. DOI: 10.1103/RevModPhys.64.1045.

(33) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive

Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.

(34) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *Journal of computational chemistry* **2013**, *34* (32), 2757-2770.

(35) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-level characterization of protein-protein association. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116* (10), 4244-4249. DOI: 10.1073/pnas.1815431116.

(36) Pan, J. A.; Peng, X. X.; Gao, Y. J.; Li, Z. L.; Lu, X. L.; Chen, Y. Z.; Ishaq, M.; Liu, D.; DeDiego, M. L.; Enjuanes, L.; et al. Genome-Wide Analysis of Protein-Protein Interactions and Involvement of Viral Proteins in SARS-CoV Replication. *Plos One* **2008**, *3* (10). DOI: 10.1371/journal.pone.0003299.

(37) Riley, K. E.; Hobza, P. Noncovalent interactions in biochemistry. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1* (1), 3-17. DOI: 10.1002/wcms.8.

(38) Chu, H.; Chan, J. F.-W.; Yuen, T. T.-T.; Shuai, H.; Yuan, S.; Wang, Y.; Hu, B.; Yip, C. C.-Y.; Tsang, J. O.-L.; Huang, X. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *The Lancet Microbe* **2020**.

(39) Hui, K. P. Y.; Cheung, M.-C.; Perera, R. A. P. M.; Ng, K.-C.; Bui, C. H. T.; Ho, J. C. W.; Ng, M. M. T.; Kuok, D. I. T.; Shih, K. C.; Tsao, S.-W. Tropism, replication competence, and innate immune responses of the coronavirus SARS-CoV-2 in human respiratory tract

and conjunctiva: an analysis in ex-vivo and in-vitro cultures. *The Lancet Respiratory Medicine* **2020**.

(40) Pruijssers, A. J.; George, A. S.; Schäfer, A.; Leist, S. R.; Gralinksi, L. E.; Dinnon Iii, K. H.; Yount, B. L.; Agostini, M. L.; Stevens, L. J.; Chappell, J. D. Remdesivir inhibits SARS-CoV-2 in human lung cells and chimeric SARS-CoV expressing the SARS-CoV-2 RNA polymerase in mice. *Cell reports* **2020**, *32* (3), 107940.

(41) Malin, J. J.; Suárez, I.; Priesner, V.; Fätkenheuer, G.; Rybniker, J. Remdesivir against COVID-19 and other viral diseases. *Clinical microbiology reviews* **2020**, *34* (1).

8. VIRTUAL SCREENING OF RHODIOLA AND OREGANO COMPOUNDS FOR BREAST CANCER DRUG DISCOVERY UTILIZING MOLECULAR DOCKING[7]

8.1. Introduction

Breast cancer is the second most common cancer, followed by skin cancer among women in the USA. Breast cancer is also the second major cause of death among US women, followed by lung cancer [1]. The average chance of a US woman developing breast cancer is 13%, while her chance of falling to death by it is 2.6% [2]. American Cancer Society estimates that about 281,550 US women will be diagnosed with invasive breast cancer while around 43600 will die in 2021. Breast cancer occurs when cells in the breast start growing in an uncontrolled manner [3]. Breast cancer can be divided into four subgroups based on the presence or absence of hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) proteins, i.e., (1) Luminal A: HR+/HER2-, (2) Luminal B: HR+/HER2+, (3) HER2-enriched: HR-/HER2+, and (4) Triple-negative breast cancer: HR-/HER2- [4]. Hormone receptor proteins include estrogen receptor (ER) and progesterone-receptor (PR) proteins, and breast cancer cells can contain either, both, or neither of them. The majority of breast cancer (~70%) tumors express at least one hormone receptor protein while 15% of breast cancers express none [5].

The treatment procedure of breast cancer largely depends on its molecular profile and metastatic condition [6]. Metastasis, an event that accounts for the majority of breast cancer-related deaths, is defined as the spread of breast cancer cells to other parts of the body, i.e., lung, bone, liver, etc. [7]. Breast Cancer treatment practices are divided into two major categories: local and systemic [6, 8]. Local treatment employs surgery or radiation (cryotherapy, laser therapy, topical therapy) or both to remove, destroy and control the growth of breast cancer cells. In non-metastatic

[7] The contents of this chapter are going to be submitted for publication.

breast cancer, local treatment is used to remove the tumor and inhibit recurrence. On the contrary, in metastatic breast cancer, local treatment is primarily used for palliation [9]. Systemic treatments utilize drugs orally/intravenously for killing/controlling the cancer cells throughout the body. Systemic approach is used as a pre-operative or post-operative treatment in both metastatic and non-metastatic cancer [10]. The most common types of systemic treatments are chemotherapy, hormone therapy, targeted therapy, and immunotherapy. Chemotherapy kills cancer cells by employing anti-cancer drugs that damage the genes inside the nucleus of cells [11]. Hormone therapeutics, utilized in hormone receptor-positive breast cancer, inhibit the growth of breast cancer cells by preventing the attachment of hormones to receptors [12]. Targeted therapeutics target specific proteins in the breast cancer cells to cease their proliferation [13]. Immunotherapy stimulates a person's immune system through lab-made immune components to fight cancer [14].

Systemic treatments (chemotherapy, hormone therapy, targeted therapy, immunotherapy) generally employ chemo-synthetic drugs that are cytotoxic and cytostatic. These anticancer drugs target rapidly proliferating cells resulting in their non-specific toxicity to certain normal cells (bone marrow, digestive tract, and hair follicle) that limits their efficacy [15]. These anticancer drugs need to be employed at a maximum tolerated dose (MTD) to have the best possible therapeutic effects i.e., widespread distribution and rapid excretion of the drug from the body. However, the employment of drugs at MTD can cause various side effects including cardiotoxicity, neurotoxicity, immunosuppression etc. [16]. Another limitation of the traditional anticancer agents is multidrug resistance (MDR), an event where the cancer cells can survive the effect of a range of drugs resulting from the reduced drug absorption of cells [17]. Phytochemicals can be an effective alternative solution to the problems mentioned above.

In a broader sense, phytochemicals refer to any plant-based functional foods or isolated plant compounds or compounds present in fruits, vegetables, herbs, and spices. Therefore, a phytochemical can be defined as a non-nutrient bioactive compound present in a plant-based diet [18]. Phytochemicals have been found to possess anticancer activity with low toxicity to normal cells [19, 20]. Dietary phytochemicals have also been observed to reverse the multi-drug resistance [21]. Phytochemicals can be considered anticancer agents due to their diversified contributions, including anti-proliferative, anti-metastatic, anti-angiogenic, apoptotic activities, along with induction of cell cycle arrest [22]. Phytochemicals primarily bind to different protein targets to perform these anticancer activities [19, 23]. The potency of different phytochemicals as anticancer agents can be determined by investigating their binding mechanism to protein targets utilizing molecular docking.

Molecular docking is an in-silico–chemico biological approach that optimizes the protein-ligand (drug candidate) conformation to compute their minimized binding free energy [24]. Molecular Docking is primarily employed to predict the binding orientation of drug candidates against protein targets to predict the drug candidates' affinity and activity. It investigates the protein-ligand interactions along with their strength [25]. Molecular docking has been extensively utilized for breast cancer drug discovery [26, 27]. Phytochemicals have also been investigated for their potential anti breast cancer activity by employing docking [27, 28].

Rhodiola (*Rhodiola Crenulata**)*** and Oregano (*Origanum Vulgare**)*** are two different plants commonly grown in North Dakota as its weather and geological conditions are conducive for their growth. Extracts or specific phytochemicals derived from different Rhodiola species have long been explored for their potency against breast cancer [29]. Different phytochemicals and extracts derived from Oregano species have also been examined for their potency against breast cancer [30].

However, the previous studies were primarily focused on investigating the efficacy of plant extracts or specific phytochemicals on breast cancer cell lines while providing little or no information on which specific protein targets are bound to these compounds. The primary objective of the current study is to investigate the potency of all phytochemicals present in the *Rhodiola Crenulata, and Origanum Vulgare* extracts against breast cancer proteins utilizing molecular docking. This molecular docking study is the virtual screening of Rhodiola and Oregano phytochemicals against a series of proteins related to different mechanisms of breast cancer.

8.2. Methodology

To perform the molecular docking of Rhodiola and Oregano phytochemicals against breast cancer proteins, a total of 33 breast cancer proteins have been selected. This selection has been performed based on several previous review studies on breast cancer [31]. These proteins have various roles in different stages and mechanisms of breast cancer, including hormone attachment, tumor progression, tumor recurrence, apoptosis, autophagy, metastasis, stemness etc. However, some proteins can perform dual or multi-directional roles in breast cancer progression [32]. **Table 8.1** represents the commonly observed contributions/roles conducted by the selected proteins in breast cancer.

Table 8.1. Roles of proteins in breast cancer

Number of targets	Breast cancer target protein	Contribution/Relation to breast cancer/Implication in breast cancer
1	ER-alpha	Estrogen receptors fuel the cancer growth when estrogen hormones attach to them. [33]
2	PR	Progesterone receptors fuel the cancer growth when progesterone hormones attach to them.[34]
3	EGFR	The overexpression of EGFR in Triple negative breast cancer (TNBC) and inflammatory breast cancer (IBC) is considered to be linked with large tumor size, poor differentiation, and poor clinical outcomes.[35]
4	mTOR	Activation of mTOR results in tumor progression as well as reduced autophagy. [36]
5	HER2	The overexpression of HER2 is associated with more aggressive version of cancer marked by higher recurrence rate, and shortened survival [37]
6	HSP90	The overexpression of HSP90 is hypothesized to contribute toward the resistance of breast cancer cells against various stress stimuli (drug effects)[38]
7	Mutant p53 P53 DBD Wild p53	Reduced p53 activity reduces interferes the apoptosis activity and expedites the cancer progression.[39]
8	Murine COX-2	The expression of COX-2 is connected to angiogenesis and lymph node metastasis in human breast cancer[40]
9	VEGFR-2	VEGFR-2 is associated with the development of tumor angiogenesis.[41]
10	HIF-1α	HIF-1α plays in the development and progression of angiogenesis. [42]
11	PI3K	PI3K contributes to breast cancer tumor cell growth and proliferation as a part of PI3K/Akt/mTOR pathway.[43]
12 13	Akt1 Akt2	Activated Akt results in cancer cell survival [44]

Table 8.1. Roles of proteins in breast cancer (continued)

Number of targets	Breast cancer target protein	Contribution/Relation to breast cancer/Implication in breast cancer
14	JAK2	The activity of JAK2 is associated with the stemness (tumor recurrence and drug resistance) of triple negative breast cancer cells.[45]
15	STAT3	Overexpressed and overactivated STAT3 contributes to the metastasis and chemoresistance of breast cancer cells.[46]
16	MMP2	Highly expressed MMP2 is associated with lymph node metastasis and tumor staging.[47]
17	MMP9	Highly expressed MMP9 is associated with lymph node metastasis and tumor staging.[47]
18	P38-alpha (MAPK 14) P38-alpha	P38-alpha is important for tumor progression.[48]
19	JNK1(MAPK 8)	In HER2/neu-positive breast cancer, JNK1 stimulates the cell survival.[49]
20	HSP70	Overexpressed HSP70 is correlated with increased cell proliferation and lymph node metastasis in human breast cancer.[50]
21	ERK 2 (MAPK 1)	Overexpressed ERK2 result in poorer prognosis in Triple negative breast cancer.[51] It also contributes to epithelial-mesenchymal transition (EMT).[52]
22	Bcl-2	Higher level of Bcl-2 is associated with poorer clinical outcome in breast cancer.[53]
23	Caspase-3	Caspase-3 is associated with the apoptotic rates in breast cancer.[54]
24	Caspase-8	The activation of Caspase-8 might restrain the metastatic ability of cancer cells.[55]
25	Caspase-9	Inhibition of Caspase-9 promotes apoptosis of cancer cells.[56]
26	FAK1	Reduced FAK1 expression disrupts cancer cell proliferation.[57]
27	Bax	Bax promotes apoptosis in breast cancer.[58]
28	NF-kB	NF-kB is associated with increased proliferation and reduced apoptosis of human breast cancer cells.[59]

Table 8.1. Roles of proteins in breast cancer (continued)

Number of targets	Breast cancer target protein	Contribution/Relation to breast cancer/Implication in breast cancer
29	TNF-alpha	Highly expressed TNF-alpha promotes the stemness of breast cancer cells.[60]
30	Cyclin-D1	Overexpressed cyclin-D1 may result in breast cancer treatment failure.[61]
31	CDK4	CDK4 plays key role in the proliferation of breast cancer cells.[62]
32	ODC	Downregulation of ODC reduced the proliferation and migratory potential of breast cancer cells.[63]
33	TGF-beta	TGF-beta promotes tumor progression in later stages of breast cancer.[64]

The selected ligands/compounds for molecular docking are phytochemicals derived from Rhodiola and Oregano extracts. The Rhodiola extract refers to the root extract of *Rhodiola Crenulata,* while the Oregano extract refers to the aerial part extract (leaf and flower) of *Oreganum Vulgare.* A total of 28 Rhodiola compounds have been selected from the literature [65]. In the case of Oregano, 22 compounds have been picked from previous studies [66]. The present study performs the molecular docking of a total of 50 (28 Rhodiola, 22 Oregano) compounds against 33 breast cancer proteins. These phytochemicals belong to different categories in their structure i.e., polyphenols, flavonoids, flavonols, terpenoids, carotenoids etc. **Table 2** and **Table 3** lists the Rhodiola and Oregano compounds to be docked, respectively.

Table 8.2. List of Rhodiola Compounds

No.	Compound Name	Chemical formula
1	Gallic acid	$C_7H_6O_5$
2	3-O-methylgallic acid	$C_8H_8O_5$
3	4-(β-D-Glucopyranosyloxy)-3,5-dimethoxybenzoic acid	$C_{15}H_{20}O_{10}$
4	protocatechuic acid	$C_7H_6O_4$
5	vanillic acid	$C_8H_8O_4$
6	vanillic acid 4-O-β-D-glucopyranoside	$C_{14}H_{18}O_9$
7	Tyrosol	$C_8H_{10}O_2$
8	Salidroside	$C_{14}H_{20}O_7$
9	4-Hydroxybenzoic acid	$C_7H_6O_3$
10	4-(beta-D-Glucosyloxy) benzoic acid	$C_{13}H_{16}O_8$
11	Rhodiocyanoside A	$C_{11}H_{17}NO_6$
12	Sarmentosin	$C_{11}H_{17}NO_7$
13	Epigallocatechin gallate	$C_{22}H_{18}O_{11}$
14	(7R*,8R*)-3-Methoxy-3',4,7,9,9'-pentahydroxy-8,4'-oxyneolignan 4-xyloside	$C_{24}H_{32}O_{11}$
15	Isolariciresinol 4'-O-beta-D-glucoside	$C_{26}H_{34}O_{11}$
16	Dehydrodiconiferyl alcohol 4-O-beta-D-glucopyranoside	$C_{26}H_{32}O_{11}$
17	Picein	$C_{14}H_{18}O_7$
18	Icariside D2	$C_{14}H_{20}O_7$
19	Creoside I	$C_{14}H_{24}O_7$
20	Kenposide A	$C_{21}H_{36}O_{10}$
21	Rhodioloside E	$C_{21}H_{38}O_{11}$
22	Rhodiooctanoside	$C_{19}H_{36}O_{10}$
23	coniferoside	$C_{16}H_{22}O_8$
24	dihydroconiferin	$C_{16}H_{24}O_8$
25	triandrin	$C_{15}H_{20}O_7$
26	Vimalin	$C_{16}H_{22}O_7$
27	Pollenitin	$C_{16}H_{12}O_7$
28	Clemastanin A	$C_{25}H_{32}O_{11}$

Table 8.3. List of Oregano Compounds

No.	Compound Name	Chemical formula
1	Carvacrol	$C_{10}H_{14}O$
2	Thymol	$C_{10}H_{14}O$
3	creosol	$C_8H_{10}O_2$
4	Phytol	$C_{20}H_{40}O$
5	P-Cymene	$C_{10}H_{14}$
6	gamma-Terpinene	$C_{10}H_{16}$
7	1-Octacosanol	$C_{28}H_{58}O$
8	luteolin 7-O-glucoside	$C_{21}H_{20}O_{11}$
9	Rosmarinic acid	$C_{18}H_{16}O_8$
10	Luteolin-7-o-glucuronide	$C_{21}H_{18}O_{12}$
11	Apigenin-7-o-glucuronide	$C_{21}H_{18}O_{11}$
12	Linalyl acetate	$C_{12}H_{20}O_2$
13	cis-Sabinene hydrate	$C_{10}H_{18}$
14	4-HYDROXY-4-METHYL-2-PENTANONE	$C_6H_{12}O_2$
15	caffeic acid	$C_9H_8O_4$
16	trans-Sabinene hydrate	$C_{10}H_{18}O$
17	Quercetin 3-O-rutinoside	$C_{27}H_{30}O_{16}$
18	n-Heptanoic acid	$C_7H_{14}O_2$
19	Nitro-L-arginine	$C_6H_{13}N_5O_4$
20	Eriodictyol	$C_{15}H_{12}O_6$
21	Taxifolin	$C_{15}H_{12}O_7$
22	Dihydrokaempferol	$C_{15}H_{12}O_6$

The molecular docking of selected compounds against breast cancer target proteins has been performed using AutoDock Vina 1.1.2.[67] The structures of the selected compounds have been

obtained in .sdf format from PubChem database [68]. The compound sdf files were converted to 3D pdb format using Open Babel [69]. Then AutoDockTools (ADT) 1.5.6 was used to prepare the compound pdbqt files from pdb files by merging the non-polar hydrogen, adding the Gasteiger charges and defining the rotatable bonds. These prepared compound pdbqt files were used as ligand input for AutoDock Vina. On the contrary, the 3D protein structures were collected from RCSB Protein Data Bank. The protein structures were cleaned by removing co-crystallized ligands, water molecules, ions etc. just to keep a single chain of protein using Discovery Studio 2020 [70]. The AutoDockTools was used to prepare the cleaned single-chain protein structure pdbqt file by employing the Gasteiger charges, adding polar hydrogens, and merging non-polar hydrogens. The protein pdbqt files prepared by AutoDockTools were used as protein inputs for AutoDock Vina. The possible active site/ligand binding site on the protein was defined based on the location of the co-crystallized ligand available in literature or by using UniProtKB database[71]. The search grid was specified for each protein with a spacing of 1 Å. In order to have more consistent docking results, the exhaustiveness parameter has been set to 24 [72]. At last, the molecular docking has been performed utilizing AutoDock Vina 1.1.2 in the Linux platform of Center for Computationally Assisted Science and Technology (CCAST), a parallel computing facility at North Dakota State University. The proteins were treated as rigid targets during the docking while the ligands (phytochemicals) were treated as semi-flexible based on their rotatable bonds.

To check the accuracy of docking performed by AutoDock Vina, we have performed the undocking and redocking of co-crystallized ligands in several protein structures. These experimentally (X-ray crystallography, cryo-electron microscopy) derived protein-ligand complexes have already been deposited in RCSB protein data bank. The selected protein-ligand models are ER-4 Hydroxytamoxifen (PDB ID: 3ERT [73]), PR-Ulipristal acetate (PDB ID: 4OAR

[74]), and mTOR-X6K (PDB ID: 4JT6 [75]). These co-crystallized ligands have been undocked and then redocked again by using AutoDock Vina. If the conformation of redocked ligand is within 2 Å RMSD (root-mean-square deviation) of the experimentally docked ligand, then the docking will be considered accurate [67]. The RMSD values were calculated using the PyMOL software [76].

The results of molecular docking include protein-ligand binding poses along with their binding affinity in kcal/mol. The higher binding affinity of protein-ligand complex refers to the higher potency of the ligand as a drug compound. It is essential to consider a threshold value of binding affinity to identify the potent ligand molecules inside different protein cavities. We have performed the molecular docking of five commercially available breast cancer drug compounds against their specific protein targets utilizing AutoDock Vina to establish a threshold binding affinity value. The chosen commercial drug compounds were Alpelisib, Lapatinib, Everolimus, Raloxifene, and Tamoxifen targeted against PI3K, HER2, mTOR, and ER proteins [77]. The binding affinity of these protein-ligand complexes will be utilized to establish the threshold binding affinity. If a ligand binds to the protein with the same binding affinity values as the threshold value or higher than it, then the ligand will be considered as a potent compound.

8.3. Results and Discussion

The RMSD values of redocked co-crystallized ligands have been calculated to check the docking accuracy of AutoDock Vina. The RMSD values of redocked 4 Hydroxytamoxifen, Ulipristal acetate, and X6K within the ER, PR, and mTOR cavities have been computed as 1.67 Å, 0.20 Å, and 1.04 Å, respectively. As all the RMSD values fall under 2 Å, the docking performed by AutoDock Vina can be considered accurate.

AutoDock Vina has been employed to dock the commercial breast cancer drug compounds against specific protein targets, and their binding affinities have been obtained. These binding

affinities are provided in **Table 4**. The negative value of binding affinity refers to attractive interactions.

Table 8.4. Binding affinities of commercial breast cancer drug compounds (ligands) against protein targets

Protein-Commercial Drug	**Binding affinity (kcal/mol)**
PI3K- Alpelisib	-8.3
HER2- Lapatinib	-10.7
mTOR- Everolimus	-8.0
ER- Raloxifene	-10.3
ER- Tamoxifen	-9.7
Average	-9.4

The average binding affinity of commercial drug compounds was computed as -9.4 kcal/mol. However, to reduce the chance of ignoring a potential phytochemical for breast cancer treatment, we will be investigating all the Rhodiola and Oregano compounds with a binding affinity of -9 kcal/mol or higher.

The binding affinity values of all Rhodiola and Oregano compounds against 33 breast cancer proteins are depicted through bubble charts (**Fig. 1, Fig. 2**).

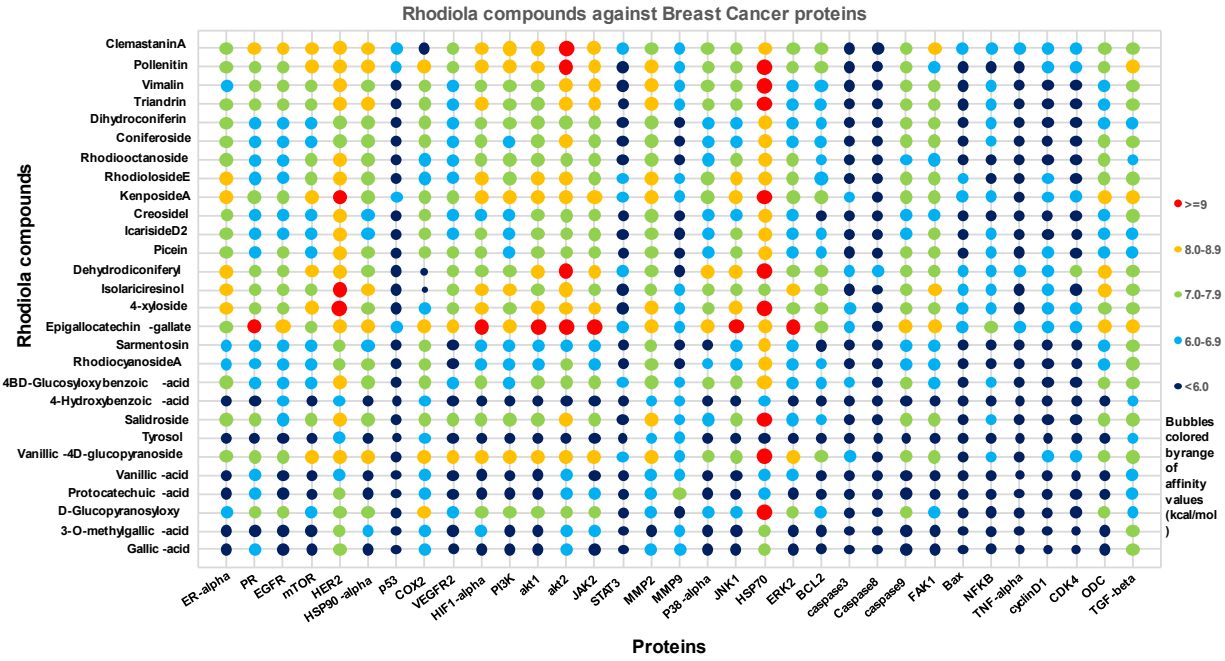

Figure 8.1. Bubble chart showing the binding affinities of Rhodiola compounds against breast cancer proteins in a grouped manner.

Figure 1 shows that certain compounds of Rhodiola show higher binding affinity toward the breast cancer proteins. Among them, epigallocatechin gallate (EGCG) has been found to be the most potent compound as it shows higher binding affinity with a range of proteins. Other Rhodiola compounds like vanillic-acid-4D-glucopyranoside, 4-xyloside, and KenposideA have also been observed to bind to different proteins with moderate to high affinity. Almost all the Rhodiola compounds bind to HSP70 with high affinity. Some Rhodiola compounds bind to AKT (akt1/akt2) proteins with varying affinity.

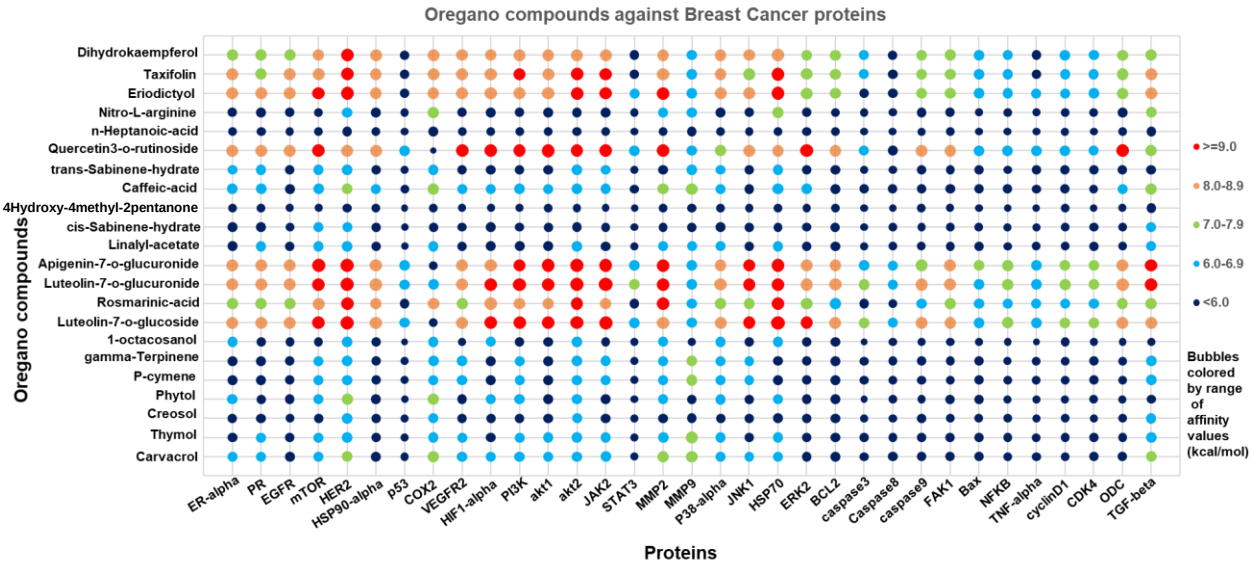

Figure 8.2. Bubble chart showing the binding affinities of Oregano compounds against breast cancer proteins.

Figure 2 shows that Luteolin 7-o-glucoside, Rosmarinic acid, Luteolin-7-o-glucuronide, Apigenin-7-o- glucuronide, and Quercetin 3-o-rutinoside of Oregano exhibit higher binding affinity to a range of proteins, including mTOR, HER2, PI3K, AKT (akt1/akt2), JAK2, MMP2, and HSP70. Other potent compounds of Oregano are found to be Eriodictyol, Taxifolin, and Dihydrokaempferol, which bind to several proteins.

The protein-ligand complexes with binding affinity values of -9 kcal/mol or higher have been listed in **Table 5** and **Table 6**. An extensive literature search was conducted for these protein-ligand complexes to obtain any experimental findings that may define phytochemical's impact on the bound protein in terms of breast cancer treatment.

Table 8.5. Rhodiola compounds against breast cancer proteins with high binding affinity

No	Protein (serialID_name_PDBID-Chain)	Rhodiola Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
1	13_akt2_2jdr_chainA	Epigallocatechin-gallate (EGCG)	13	-10.1	EGCG reduced AKT expression/activation in breast cancer cells [78]
2	20_HSP70_3atu_chainA	vanillic-acid-4-O-Î²-D-glucopyranoside	6	-10.1	No literature found. However, derivatives of vanillic acid has been found to inhibit HSP70 in colorectal cancer cells[79]
3	5_HER2_3pp0_chainA	4-xyloside	14	-10.0	Derivatives of xyloside inhibits HER2 expression in breast cancer cells[80]
4	12_akt1_4ekl_chainA	Epigallocatechin-gallate	13	-10.0	EGCG reduced AKT expression/activation in breast cancer cells [78]
5	20_HSP70_3atu_chainA	Dehydrodiconiferyl	16	-9.9	No literature found.
6	20_HSP70_3atu_chainA	4-xyloside	14	-9.8	No literature found.
7	20_HSP70_3atu_chainA	Kenposide A	20	-9.5	No literature found.
8	13_akt2_2jdr_chainA	ClemastaninA	28	-9.4	No literature found.
9	20_HSP70_3atu_chainA	Pollenitin	27	-9.4	No literature found.
10	20_HSP70_3atu_chainA	Triandrin	25	-9.3	No literature found. However, In neuroglia cells Triandrin might deregulate HSP70.[81]
11	20_HSP70_3atu_chainA	Vimalin	26	-9.3	No literature found.
12	10_HIF1Alpha_3kcx_chainA	Epigallocatechin-gallate (EGCG)	13	-9.2	EGCG inhibited the activation of HIF-1α in breast cancer.[82]
13	5_HER2_3pp0_chainA	Isolariciresinol	15	-9.1	No literature found.

Table 8.5. Rhodiola compounds against breast cancer proteins with high binding affinity (continued)

No	Protein (serialID_name_PDBID-Chain)	Rhodiola Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
14	5_HER2_3pp0_chainA	Kenposide A	20	-9.1	No literature found.
15	13_akt2_2jdr_chainA	Dehydrodiconiferyl alcohol (DHCA)	16	-9.1	No literature found. However, in osteoporosis DHCA decreased the expression level of akt.[83]
16	13_akt2_2jdr_chainA	Pollenitin	27	-9.1	No literature found.
17	14_JAK2_3krr_chainA	Epigallocatechin-gallate	13	-9.1	No literature found. However, EGCG inhibits the activation of JAK2 in human oral cancer [84] and Ewing family tumor cells.[85]
18	20_HSP70_3atu_chainA	Î²DGlucopyranosyloxy	3	-9.1	No literature found.
19	21_ERK2_4n0s_chainA	Epigallocatechin-gallate	13	-9.1	EGCG downregulated the activation of ERK in breast cancer cells.[86]
20	2_PR_4oar_chainA	Epigallocatechin-gallate	13	-9.0	EGCG results in increased steady-state PR mRNA expression.[87]
21	19_JNK1_2h96_chainA	Epigallocatechin-gallate	13	-9.0	EGCG can potentiate the JNK in breast cancer.[88]
22	20_HSP70_3atu_chainA	Salidroside	8	-9.0	No literature found. However, salidroside downregulates HSP70 expression in gastric cancer cells.[89]
23	1_ER_alpha_3ert_chainA	RhodiolosideE	21	-8.9 (~9.0)	No literature found.

Table 8.5. Rhodiola compounds against breast cancer proteins with high binding affinity (continued)

No	Protein (serialID_name_PDBID-Chain)	Rhodiola Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
24	5_HER2_3pp0_chainA	vanillic-acid-4-O-Î²-D-glucopyranoside	6	-8.9 (~9.0)	No literature found. However, phenolic mixture containing vanillic acid inhibits the growth and oxidative stress of HER2 positive breast cancer cells.[90]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
1	13_akt2_2jdr_chainA	Luteolin-7-o-glucuronide	10	-10.7	Luteolin decreased akt levels both in HR positive and Triple negative breast cancer cells.[91]
2	14_JAK2_3krr_chainA	Luteolin-7-o-glucuronide	10	-10.7	Luteolin inhibits the activation of JAK2.[92]
3	20_HSP70_3atu_chainA	luteolin-7-O-glucoside	8	-10.5	No literature found.
4	5_HER2_3pp0_chainA	Luteolin-7-o-glucuronide	10	-10.4	Luteolin decreased HER2 expression in breast cancer cells.[93]
5	14_JAK2_3krr_chainA	luteolin-7-O-glucoside	8	-10.4	Luteolin inhibits the activation of JAK2.[92]
6	20_HSP70_3atu_chainA	Apigenin-7-o-glucuronide	11	-10.3	No literature found. However, apigenin inhibits HSP70 in lung cancer cells.[94]
7	4_mTOR_4jt6_chainA	Apigenin-7-o-glucuronide	11	-10.1	No literature found. However, apigenin decreased mTOR expression in colon cancer cells.[95]
8	5_HER2_3pp0_chainA	luteolin-7-O-glucoside	8	-10.1	Luteolin decreased HER2 expression in breast cancer cells.[93]
9	12_akt1_4ekl_chainA	Quercetin-3-O-rutinoside	17	-10.1	Quercetin inactivates akt in breast cancer cells to induce autophagy.[96]
10	13_akt2_2jdr_chainA	Apigenin-7-o-glucuronide	11	-10.1	Apigenin decreases the akt activation in breast cancer cells.[97]
11	14_JAK2_3krr_chainA	Apigenin-7-o-glucuronide	11	-10.0	Apigenin blocked the activation of JAK2 in breast cancer cells.[98]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
12	12_akt1_4ekl_chainA	Apigenin-7-o-glucuronide	11	-9.9	Apigenin decreases the akt activation in breast cancer cells.[97]
13	13_akt2_2jdr_chainA	luteolin-7-O-glucoside	8	-9.9	Luteolin decreased akt levels both in HR positive and Triple negative breast cancer cells.[91]
14	20_HSP70_3atu_chainA	Rosmarinic-acid	9	-9.9	No literature found.
15	20_HSP70_3atu_chainA	Luteolin-7-o-glucuronide	10	-9.9	No literature found.
16	4_mTOR_4jt6_chainA	Luteolin-7-o-glucuronide	10	-9.8	Luteolin reduces the mTOR levels in breast cancer cells.[99]
17	5_HER2_3pp0_chainA	Apigenin-7-o-glucuronide	11	-9.8	Apigenin reduces the HER2 protein expression in breast cancer cells.[100]
18	20_HSP70_3atu_chainA	Eriodictyol	20	-9.8	No literature found.
19	20_HSP70_3atu_chainA	Taxifolin	21	-9.8	No direct literature found. However, taxifolin suppressed the hsp70 gene expression in human promyeloid leukemia cells.[101]
20	5_HER2_3pp0_chainA	Eriodictyol	20	-9.7	Eriodictyol inhibits the HER2 protein expression in breast cancer cells.[102]
21	11_PI3K_1e8z_chainA	Luteolin-7-o-glucuronide	10	-9.7	Luteolin can modulate the PI3K activity in breast cancer cells.[103]
22	21_ERK2_4n0s_chainA	Quercetin3-O-rutinoside	17	-9.7	Quercetin suppresses ERK activation in breast cancer cells.[104]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
23	4_mTOR_4jt6_chainA	luteolin-7-O-glucoside	8	-9.6	Luteolin reduces the mTOR levels in breast cancer cells.[99]
24	5_HER2_3pp0_chainA	Rosmarinic-acid	9	-9.6	No literature found.
25	9_VEGFR2_1y6b_chainA	Quercetin3-O-rutinoside	17	-9.6	Quercetin reduces VEGFR2 protein expression in breast cancer.[105]
26	10_HIF1Alpha_3kcx_chainA	Luteolin-7-o-glucuronide	10	-9.6	Luteolin reduces HIF-1 activity by downregulating HIF-1α transcriptional activity.[106]
27	12_akt1_4ekl_chainA	Luteolin-7-o-glucuronide	10	-9.6	Luteolin decreased akt levels both in HR positive and Triple negative breast cancer cells.[91]
28	16_MMP2_3ayu_chainA	Quercetin3-O-rutinoside	17	-9.6	Quercetin downregulated MMP2 protein expression in breast cancer cells.[96]
29	12_akt1_4ekl_chainA	luteolin-7-O-glucoside	8	-9.5	Luteolin decreased akt levels both in HR positive and Triple negative breast cancer cells.[91]
30	13_akt2_2jdr_chainA	Eriodictyol	20	-9.5	No literature found. However, eriodictyol downregulated the akt protein expression in lung cancer cells.[107]
31	32_ODC_1d7k_chainA	Quercetin3-O-rutinoside	17	-9.5	No literature found. However, quercetin decreases the ODC activity in colon cancer cells.[108]
32	5_HER2_3pp0_chainA	Taxifolin	21	-9.4	No literature found.

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
33	11_PI3K_1e8z_chainA	luteolin-7-O-glucoside	8	-9.4	Luteolin can modulate the PI3K activity in breast cancer cells.[103]
34	14_JAK2_3krr_chainA	Quercetin3-O-rutinoside	17	-9.4	No literature found. However, quercetin reduced the JAK2 activation in Hepatocellular carcinoma.[109]
35	16_MMP2_3ayu_chainA	Apigenin-7-o-glucuronide	11	-9.4	Apigenin inhibited MMP2 in triple negative breast cancer cells.[110]
36	19_JNK1_2h96_chainA	Luteolin-7-o-glucuronide	10	-9.4	Luteolin suppresses JNK in breast cancer.[111]
37	5_HER2_3pp0_chainA	Dihydrokaempferol	22	-9.3	No literature found.
38	10_HIF1Alpha_3kcx_chainA	luteolin-7-O-glucoside	8	-9.3	Luteolin reduces HIF-1 activity by downregulating HIF-1α transcriptional activity.[106]
39	10_HIF1Alpha_3kcx_chainA	Quercetin3-O-rutinoside	17	-9.3	Quercetin suppresses the HIF-1α in breast cancer cells.[112]
40	13_akt2_2jdr_chainA	Taxifolin	21	-9.3	No literature found. However, taxifolin reduces the akt expression in osteosarcoma cells.[113]
41	16_MMP2_3ayu_chainA	Luteolin-7-o-glucuronide	10	-9.3	Luteolin decreased the MMP2 expression in breast cancer cells.[114]
42	16_MMP2_3ayu_chainA	Eriodictyol	20	-9.3	No literature found. However, eriodictyol inhibits the expression of MMP2 in glioblastoma cells.[115]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
43	11_PI3K_1e8z_chainA	Quercetin3-O-rutinoside	17	-9.2	Quercetin suppresses the levels of PI3K in breast cancer cells.[116]
44	13_akt2_2jdr_chainA	Quercetin3-O-rutinoside	17	-9.2	Quercetin inactivates akt in breast cancer cells to induce autophagy.[96]
45	19_JNK1_2h96_chainA	luteolin-7-O-glucoside	8	-9.2	Luteolin suppresses JNK in breast cancer.[111]
46	4_mTOR_4jt6_chainA	Quercetin3-O-rutinoside	17	-9.1	Quercetin decreases mTOR expression in breast cancer cells.[117]
47	11_PI3K_1e8z_chainA	Apigenin-7-o-glucuronide	11	-9.1	Apigenin suppresses PI3K activity.[118]
48	11_PI3K_1e8z_chainA	Taxifolin	21	-9.1	No literature found. However, taxifolin inhibits PI3K activity in glioma cell lines.[119]
49	16_MMP2_3ayu_chainA	Rosmarinic-acid	9	-9.1	Rosmarinic acid downregulates MMP2 expression in MDA-MB-231 cells.[120]
50	19_JNK1_2h96_chainA	Apigenin-7-o-glucuronide	11	-9.1	No literature found. However, apigenin activates JNK1 resulting in autophagy of solid Ehrlich carcinoma.[121]
51	21_ERK2_4n0s_chainA	luteolin-7-O-glucoside	8	-9.1	Luteolin increases ERK activity towards breast cancer cell death.[122]
52	33_TGFbeta_5e8s_chainA	Luteolin-7-o-glucuronide	10	-9.1	No literature found. However, luteolin inhibits the TGF-beta activity towards epithelial to mesenchymal transition.[123]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
53	33_TGFbeta_5e8s_chainA	Apigenin-7-o-glucuronide	11	-9.1	No literature found. However, apigenin inhibits TGF-beta activity towards VEGF expression in prostate cancer.[124]
54	4_mTOR_4jt6_chainA	Eriodictyol	20	-9.0	No literature found. However, eriodictyol downregulates mTOR expression in lung cancer.[107]
55	13_akt2_2jdr_chainA	Rosmarinic-acid	9	-9.0	No literature found. However, Rosmarinic-acid inhibits the activation of akt in hepatocellular carcinoma.[125]
56	14_JAK2_3krr_chainA	Eriodictyol	20	-9.0	No literature found.
57	14_JAK2_3krr_chainA	Taxifolin	21	-9.0	No literature found. However, taxifolin inhibits the expression of JAK2 in murine psoriasis.[126]
58	6_HSP90Alpha_3tuh_chainA	Luteolin-7-o-glucuronide	10	-8.9 (~9.0)	Luteolin inhibits HSP90 activity in different cancer cells.[127]
59	10_HIF1Alpha_3kcx_chainA	Eriodictyol	20	-8.9 (~9.0)	No literature found.
60	16_MMP2_3ayu_chainA	luteolin-7-O-glucoside	8	-8.9 (~9.0)	Luteolin decreased the MMP2 expression in breast cancer cells.[114]
61	20_HSP70_3atu_chainA	Dihydrokaempferol	22	-8.9 (~9.0)	No literature found.
62	21_ERK2_4n0s_chainA	Luteolin-7-o-glucuronide	10	-8.9 (~9.0)	Luteolin increases ERK activity towards breast cancer cell death.[122]

Table 8.6. Oregano compounds against breast cancer proteins with high binding affinity (continued)

No.	Protein (serialID_name_PDBID-Chain)	Oregano Compounds	Ligand No.	Affinity (kcal/mol)	Experimental Literature
63	25_caspase9_2ar9_chainA	Luteolin-7-o-glucuronide	10	-8.9 (~9.0)	Luteolin enhances caspase9 activity towards apoptosis in breast cancer cells.[128]
64	32_ODC_1d7k_chainA	Luteolin-7-o-glucuronide	10	-8.9 (~9.0)	No literature found.

8.4. Conclusions

The present study performs the molecular docking of Rhodiola and Oregano Phytochemicals against a series of breast cancer proteins. It has been found that among the Rhodiola compounds, EGCG exhibits the most potent behavior as it binds with a number of proteins. The effect of EGCG on these proteins for breast cancer treatment can also be confirmed from different experimental studies. The Oregano compounds that show higher potency are different derivatives of Luteolin, Apigenin, Quercetin, and Rosmarinic acid. Previous experimental studies refer to some of the protein-ligand interactions phenomena for breast or other types of cancer. Planned experimental studies can be performed to investigate the protein-ligand complexes (no previous experimental literature) with higher binding affinity to check the ligand potency on that specific protein toward the breast cancer treatment.

8.5. References

(1) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36

cancers in 185 countries. *CA: a cancer journal for clinicians* **2018**, *68* (6), 394-424. Society, A. C. *ABOUT BREAST CANCER*; 2021.

(2) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. Cancer statistics, 2021. *CA: a cancer journal for clinicians* **2021**, *71* (1), 7-33.

(3) Henry, N. L.; Shah, P. D.; Haider, I.; Freer, P. E.; Jagsi, R.; Sabel, M. S. Chapter 88: cancer of the breast. *Abeloff's Clinical Oncology, 6th edn. Philadelphia: Elsevier* **2020**.

(4) Dai, X.; Xiang, L.; Li, T.; Bai, Z. Cancer hallmarks, biomarkers and breast cancer molecular subtypes. *Journal of Cancer* **2016**, *7* (10), 1281.

(5) Elgene Lim, M. D.; Metzger-Filho, O.; Winer, E. P. The natural history of hormone receptor-positive breast cancer. *Oncology* **2012**, *26* (8), 688.

(6) Waks, A. G.; Winer, E. P. Breast cancer treatment: a review. *Jama* **2019**, *321* (3), 288-300.

(7) Scully, O. J.; Bay, B.-H.; Yip, G.; Yu, Y. Breast cancer metastasis. *Cancer genomics & proteomics* **2012**, *9* (5), 311-320.

(8) Israel, B. e. B.; Tilghman, S. L.; Parker-Lemieux, K.; Payton-Stewart, F. Phytochemicals: Current strategies for treating breast cancer. *Oncology letters* **2018**, *15* (5), 7471-7478.

(9) Lester, J. Local treatment of breast cancer. In *Seminars in oncology nursing*, 2015; Elsevier: Vol. 31, pp 122-133.

(10) Kaufmann, M.; von Minckwitz, G.; Rody, A. Preoperative (neoadjuvant) systemic treatment of breast cancer. *The Breast* **2005**, *14* (6), 576-581.

(11) Iwamoto, T. Clinical application of drug delivery systems in cancer chemotherapy: review of the efficacy and side effects of approved drugs. *Biological and pharmaceutical bulletin* **2013**, *36* (5), 715-718.

(12) Bush, T. L.; Whiteman, M.; Flaws, J. A. Hormone replacement therapy and breast cancer: a qualitative review. *Obstetrics & Gynecology* **2001**, *98* (3), 498-508.

(13) Diaby, V.; Tawk, R.; Sanogo, V.; Xiao, H.; Montero, A. J. A review of systematic reviews of the cost-effectiveness of hormone therapy, chemotherapy, and targeted therapy for breast cancer. *Breast cancer research and treatment* **2015**, *151* (1), 27-40.

(14) Adams, S.; Gatti-Mays, M. E.; Kalinsky, K.; Korde, L. A.; Sharon, E.; Amiri-Kordestani, L.; Bear, H.; McArthur, H. L.; Frank, E.; Perlmutter, J. Current landscape of immunotherapy in breast cancer: a review. *JAMA oncology* **2019**, *5* (8), 1205-1214.

(15) Singh, S.; Sharma, B.; Kanwar, S. S.; Kumar, A. Lead phytochemicals for anticancer drug development. *Frontiers in plant science* **2016**, *7*, 1667. Chari, R. V. J. Targeted cancer therapy: conferring specificity to cytotoxic drugs. *Accounts of chemical research* **2008**, *41* (1), 98-107.

(16) Frei, E.; Elias, A.; Wheeler, C.; Richardson, P.; Hryniuk, W. The relationship between high-dose treatment and combination chemotherapy: the concept of summation dose intensity. *Clinical cancer research* **1998**, *4* (9), 2027-2037.

(17) Zahreddine, H.; Borden, K. Mechanisms and insights into drug resistance in cancer. *Frontiers in pharmacology* **2013**, *4*, 28.

(18) Kapinova, A.; Kubatka, P.; Golubnitschaja, O.; Kello, M.; Zubor, P.; Solar, P.; Pec, M. Dietary phytochemicals in breast cancer research: anticancer effects and potential utility for effective chemoprevention. *Environmental health and preventive medicine* **2018**, *23* (1), 1-18.

(19) Surh, Y.-J. Cancer chemoprevention with dietary phytochemicals. *Nature Reviews Cancer* **2003**, *3* (10), 768-780.

(20) Liu, R. H. Potential synergy of phytochemicals in cancer prevention: mechanism of action. *The Journal of nutrition* **2004**, *134* (12), 3479S-3485S.

(21) Christensen, J. G.; LeBlanc, G. A. Reversal of multidrug resistance in vivo by dietary administration of the phytochemical indole-3-carbinol. *Cancer research* **1996**, *56* (3), 574-581.

(22) Younas, M.; Hano, C.; Giglioli-Guivarc'h, N.; Abbasi, B. H. Mechanistic evaluation of phytochemicals in breast cancer remedy: current understanding and future perspectives. *RSC advances* **2018**, *8* (52), 29714-29744.

(23) Lee, K. W.; Bode, A. M.; Dong, Z. Molecular targets of phytochemicals for cancer prevention. *Nature Reviews Cancer* **2011**, *11* (3), 211-218.

(24) Chaudhary, K. K.; Mishra, N. A review on molecular docking: novel tool for drug discovery. *databases* **2016**, *3* (4), 1029. Morris, G. M.; Lim-Wilby, M. Molecular docking. In *Molecular modeling of proteins*, Springer, 2008; pp 365-382.

(25) Kirkpatrick, P. Gliding to success. *Nature Reviews Drug Discovery* **2004**, *3* (4), 299-299.

(26) Yadav, D. K.; Kumar, S.; Saloni, H. S.; Kim, M.-h.; Sharma, P.; Misra, S.; Khan, F. Molecular docking, QSAR and ADMET studies of withanolide analogs against breast cancer. *Drug design, development and therapy* **2017**, *11*, 1859.

(27) Ahmed, B.; Ashfaq, U. A.; ul Qamar, M. T.; Ahmad, M. Anti-cancer potential of phytochemicals against breast cancer: Molecular docking and simulation approach. ||| *Bangladesh Journal of Pharmacology* **2014**, *9* (4), 545-550.

(28) Acharya, R.; Chacko, S.; Bose, P.; Lapenna, A.; Pattanayak, S. P. Structure based multitargeted molecular docking analysis of selected furanocoumarins against breast cancer. *Scientific reports* **2019**, *9* (1), 1-13.

(29) Tu, Y.; Roberts, L.; Shetty, K.; Schneider, S. S. Rhodiola crenulata induces death and inhibits growth of breast cancer cell lines. *Journal of medicinal food* **2008**, *11* (3), 413-423.

(30) Kubatka, P.; Kello, M.; Kajo, K.; Kruzliak, P.; Výbohová, D.; Mojžiš, J.; Adamkov, M.; Fialová, S.; Veizerová, L.; Zulli, A. Oregano demonstrates distinct tumour-suppressive effects in the breast carcinoma model. *European journal of nutrition* **2017**, *56* (3), 1303-1316.

(31) Schlotter, C. M.; Vogt, U.; Allgayer, H.; Brandt, B. Molecular targeted therapies for breast cancer treatment. *Breast cancer research* **2008**, *10* (4), 1-12.

(32) Platet, N.; Cathiard, A. M.; Gleizes, M.; Garcia, M. Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. *Critical reviews in oncology/hematology* **2004**, *51* (1), 55-67.

(33) Clarke, R.; Liu, M. C.; Bouker, K. B.; Gu, Z.; Lee, R. Y.; Zhu, Y.; Skaar, T. C.; Gomez, B.; O'Brien, K.; Wang, Y. Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. *Oncogene* **2003**, *22* (47), 7316-7339.

(34) Daniel, A. R.; Hagan, C. R.; Lange, C. A. Progesterone receptor action: defining a role in breast cancer. *Expert review of endocrinology & metabolism* **2011**, *6* (3), 359-369.

(35) Richard, J.; Sainsbury, C.; Needham, G.; Farndon, J.; Malcolm, A.; Harris, A. Epidermal-growth-factor receptor status as predictor of early recurrence of and death from breast cancer. *The Lancet* **1987**, *329* (8547), 1398-1402.

(36) Hare, S. H.; Harvey, A. J. mTOR function and therapeutic targeting in breast cancer. *American Journal of Cancer Research* **2017**, *7* (3), 383.

(37) Meric-Bernstam, F.; Hung, M.-C. Advances in targeting human epidermal growth factor receptor-2 signaling for cancer therapy. *Clinical Cancer Research* **2006**, *12* (21), 6326-6330.

(38) Zagouri, F.; Bournakis, E.; Koutsoukos, K.; Papadimitriou, C. A. Heat shock protein 90 (hsp90) expression and breast cancer. *Pharmaceuticals* **2012**, *5* (9), 1008-1020.

(39) Lacroix, M.; Toillon, R.-A.; Leclercq, G. p53 and breast cancer, an update. *Endocrine-related cancer* **2006**, *13* (2), 293-325.

(40) Costa, C.; Soares, R.; Reis-Filho, J. S.; Leitao, D.; Amendoeira, I.; Schmitt, F. C. Cyclo-oxygenase 2 expression is associated with angiogenesis and lymph node metastasis in human breast cancer. *Journal of clinical pathology* **2002**, *55* (6), 429-434.

(41) Yan, J.-D.; Liu, Y.; Zhang, Z.-Y.; Liu, G.-Y.; Xu, J.-H.; Liu, L.-Y.; Hu, Y.-M. Expression and prognostic significance of VEGFR-2 in breast cancer. *Pathology-Research and Practice* **2015**, *211* (7), 539-543.

(42) Peiró, C. H. F.; Encina, J. A.; Perez, M. M.; Aquino, G. S. A.; Veiga, G. L.; Fonseca, F.; Alves, B. C. A. The role of hypoxia-induced factor 1α in breast cancer. *Journal of Cancer Metastasis and Treatment* **2019**, *5*.

(43) Paplomata, E.; O'Regan, R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Therapeutic advances in medical oncology* **2014**, *6* (4), 154-166.

(44) Song, M.; Bode, A. M.; Dong, Z.; Lee, M.-H. AKT as a therapeutic target for cancer. *Cancer research* **2019**, *79* (6), 1019-1031.

(45) Balko, J. M.; Schwarz, L. J.; Luo, N.; Estrada, M. V.; Giltnane, J. M.; Dávila-González, D.; Wang, K.; Sánchez, V.; Dean, P. T.; Combs, S. E. Triple-negative breast cancers with

amplification of JAK2 at the 9p24 locus demonstrate JAK2-specific dependence. *Science translational medicine* **2016**, *8* (334), 334ra353-334ra353.

(46) Ma, J.-h.; Qin, L.; Li, X. Role of STAT3 signaling pathway in breast cancer. *Cell Communication and Signaling* **2020**, *18* (1), 1-13.

(47) Li, H.; Qiu, Z.; Li, F.; Wang, C. The relationship between MMP-2 and MMP-9 expression levels with breast cancer incidence and prognosis. *Oncology letters* **2017**, *14* (5), 5865-5870.

(48) Cánovas, B.; Igea, A.; Sartori, A. A.; Gomis, R. R.; Paull, T. T.; Isoda, M.; Pérez-Montoyo, H.; Serra, V.; González-Suárez, E.; Stracker, T. H. Targeting p38α increases DNA damage, chromosome instability, and the anti-tumoral response to taxanes in breast cancer cells. *Cancer cell* **2018**, *33* (6), 1094-1110. e1098.

(49) Han, J. S.; Crowe, D. L. Jun amino-terminal kinase 1 activation promotes cell survival in ErbB2-positive breast cancer. *Anticancer research* **2010**, *30* (9), 3407-3412.

(50) Yue, L. U.; Xiang, J.-Y.; Sun, P.; Yao, Y.-S.; Sun, Z.-N.; Liu, X.-P.; Wang, H.-B.; Shen, Z.; Yao, R.-Y. Relationship between HSP70 and ERBB2 expression in breast cancer cell lines regarding drug resistance. *Anticancer research* **2016**, *36* (3), 1243-1249.

(51) Lehmann, B. D.; Bauer, J. A.; Chen, X.; Sanders, M. E.; Chakravarthy, A. B.; Shyr, Y.; Pietenpol, J. A. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation* **2011**, *121* (7), 2750-2767.

(52) Shin, S.; Dimitri, C. A.; Yoon, S.-O.; Dowdle, W.; Blenis, J. ERK2 but not ERK1 induces epithelial-to-mesenchymal transformation via DEF motif-dependent signaling events. *Molecular cell* **2010**, *38* (1), 114-127.

(53) Choi, J. E.; Kang, S. H.; Lee, S. J.; Bae, Y. K. Prognostic significance of Bcl-2 expression in non-basal triple-negative breast cancer patients treated with anthracycline-based chemotherapy. *Tumor Biology* **2014**, *35* (12), 12255-12263.

(54) O'Donovan, N.; Crown, J.; Stunell, H.; Hill, A. D. K.; McDermott, E.; O'Higgins, N.; Duffy, M. J. Caspase 3 in breast cancer. *Clinical Cancer Research* **2003**, *9* (2), 738-742.

(55) De Blasio, A.; Di Fiore, R.; Morreale, M.; Carlisi, D.; Drago-Ferrante, R.; Montalbano, M.; Scerri, C.; Tesoriere, G.; Vento, R. Unusual roles of caspase-8 in triple-negative breast cancer cell line MDA-MB-231. *International journal of oncology* **2016**, *48* (6), 2339-2348.

(56) Avrutsky, M. I.; Troy, C. M. Caspase-9: A Multimodal Therapeutic Target With Diverse Cellular Expression in Human Disease. *Frontiers in Pharmacology* **2021**, *12*, 1728.

(57) Tiede, S.; Meyer-Schaller, N.; Kalathur, R. K. R.; Ivanek, R.; Fagiani, E.; Schmassmann, P.; Stillhard, P.; Häfliger, S.; Kraut, N.; Schweifer, N. The FAK inhibitor BI 853520 exerts anti-tumor effects in breast cancer. *Oncogenesis* **2018**, *7* (9), 1-19.

(58) Martinez-Arribas, F.; Martín-Garabato, E.; Zapardiel, I.; Sánchez, J.; Lucas, A. R.; Tejerina, A.; Schneider, J. Bax expression in untreated breast cancer: an immunocytometric study of 255 cases. *Anticancer research* **2008**, *28* (5A), 2595-2598.

(59) Balkwill, F.; Mantovani, A. Inflammation and cancer: back to Virchow? *The lancet* **2001**, *357* (9255), 539-545.

(60) Liu, W.; Lu, X.; Shi, P.; Yang, G.; Zhou, Z.; Li, W.; Mao, X.; Jiang, D.; Chen, C. TNF-α increases breast cancer stem-like cells through up-regulating TAZ expression via the non-canonical NF-κB pathway. *Scientific reports* **2020**, *10* (1), 1-11.

(61) Roy, P. G.; Thompson, A. M. Cyclin D1 and breast cancer. *The Breast* **2006**, *15* (6), 718-727.

(62) Pernas, S.; Tolaney, S. M.; Winer, E. P.; Goel, S. CDK4/6 inhibition in breast cancer: current practice and future directions. *Therapeutic advances in medical oncology* **2018**, *10*, 1758835918786451.

(63) Deng, W.; Jiang, X.; Mei, Y.; Sun, J.; Ma, R.; Liu, X.; Sun, H.; Tian, H.; Sun, X. Role of ornithine decarboxylase in breast cancer. *Acta biochimica et biophysica Sinica* **2008**, *40* (3), 235-243.

(64) Oft, M.; Peli, J.; Rudaz, C.; Schwarz, H.; Beug, H.; Reichmann, E. TGF-beta1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. *Genes & development* **1996**, *10* (19), 2462-2477.

(65) Yang, Y.-n.; Liu, Z.-z.; Feng, Z.-m.; Jiang, J.-s.; Zhang, P.-c. Lignans from the root of Rhodiola crenulata. *Journal of agricultural and food chemistry* **2012**, *60* (4), 964-972.

(66) Veenstra, J. P.; Johnson, J. J. Oregano (Origanum vulgare) extract for food preservation and improvement in gastrointestinal health. *International journal of nutrition* **2019**, *3* (4), 43.

(67) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, *31* (2), 455-461.

(68) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B. PubChem in 2021: new data content and improved web interfaces. *Nucleic acids research* **2021**, *49* (D1), D1388-D1395.

(69) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of cheminformatics* **2011**, *3* (1), 1-14.

(70) Studio, D. Discovery Studio. *Accelrys [2.1]* **2008**.

(71) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2021**, *49* (D1), D480-D489.

(72) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols* **2016**, *11* (5), 905-919.

(73) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95* (7), 927-937.

(74) Petit-Topin, I.; Fay, M.; Resche-Rigon, M.; Ulmann, A.; Gainer, E.; Rafestin-Oblin, M. E.; Fagart, J. Molecular determinants of the recognition of ulipristal acetate by oxo-steroid receptors. *The Journal of steroid biochemistry and molecular biology* **2014**, *144*, 427-435.

(75) Yang, H.; Rudge, D. G.; Koos, J. D.; Vaidialingam, B.; Yang, H. J.; Pavletich, N. P. mTOR kinase structure, mechanism and regulation. *Nature* **2013**, *497* (7448), 217-223.

(76) DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* **2002**, *40* (1), 82-92.

(77) Markham, A. Alpelisib: first global approval. *Drugs* **2019**, *79* (11), 1249-1253. Taskar, K. S.; Rudraraju, V.; Mittapalli, R. K.; Samala, R.; Thorsheim, H. R.; Lockman, J.; Gril, B.; Hua, E.; Palmieri, D.; Polli, J. W. Lapatinib distribution in HER2 overexpressing experimental brain metastases of breast cancer. *Pharmaceutical research* **2012**, *29* (3), 770-781.

(78) Thangapazham, R. L.; Passi, N.; Maheshwari, R. K. Green tea polyphenol and epigallocatechin gallate induce apoptosis and inhibit invasion in human breast cancer cells. *Cancer biology & therapy* **2007**, *6* (12), 1938-1943.

(79) Mai, C. W.; Kang, Y. B.; Nadarajah, V. D.; Hamzah, A. S.; Pichika, M. R. Drug-like dietary vanilloids induce anticancer activity through proliferation inhibition and regulation of bcl-related apoptotic proteins. *Phytotherapy Research* **2018**, *32* (6), 1108-1118.

(80) Einbond, L. S.; Wen-Cai, Y.; He, K.; Wu, H.-a.; Cruz, E.; Roller, M.; Kronenberg, F. Growth inhibitory activity of extracts and compounds from Cimicifuga species on human breast cancer cells. *Phytomedicine* **2008**, *15* (6-7), 504-511.

(81) Panossian, A.; Hamm, R.; Wikman, G.; Efferth, T. Mechanism of action of Rhodiola, salidroside, tyrosol and triandrin in isolated neuroglial cells: An interactive pathway analysis of the downstream effects using RNA microarray data. *Phytomedicine* **2014**, *21* (11), 1325-1348.

(82) Gu, J.-W.; Makey, K. L.; Tucker, K. B.; Chinchar, E.; Mao, X.; Pei, I.; Thomas, E. Y.; Miele, L. EGCG, a major green tea catechin suppresses breast tumor angiogenesis and growth via inhibiting the activation of HIF-1α and NFκB, and VEGF expression. *Vascular cell* **2013**, *5* (1), 1-10.

(83) Lee, W.; Ko, K. R.; Kim, H.-k.; Lee, D. S.; Nam, I.-J.; Lim, S.; Kim, S. Dehydrodiconiferyl alcohol inhibits osteoclast differentiation and ovariectomy-induced bone loss through acting as an estrogen receptor agonist. *Journal of natural products* **2018**, *81* (6), 1343-1356.

(84) Cheng, C.-W.; Shieh, P.-C.; Lin, Y.-C.; Chen, Y.-J.; Lin, Y.-H.; Kuo, D.-H.; Liu, J.-Y.; Kao, J.-Y.; Kao, M.-C.; Way, T.-D. Indoleamine 2, 3-dioxygenase, an immunomodulatory protein, is suppressed by (−)-epigallocatechin-3-gallate via blocking of γ-interferon-induced JAK-PKC-δ-STAT1 signaling in human oral cancer cells. *Journal of agricultural and food chemistry* **2010**, *58* (2), 887-894.

(85) Kang, H.-G.; Jenabi, J. M.; Liu, X. F.; Reynolds, C. P.; Triche, T. J.; Sorensen, P. H. B. Inhibition of the insulin-like growth factor I receptor by epigallocatechin gallate blocks proliferation and induces the death of Ewing tumor cells. *Molecular cancer therapeutics* **2010**, *9* (5), 1396-1407.

(86) Sen, T.; Dutta, A.; Chatterjee, A. Epigallocatechin-3-gallate (EGCG) downregulates gelatinase-B (MMP-9) by involvement of FAK/ERK/NFκB and AP-1 in the human breast cancer cell line MDA-MB-231. *Anti-cancer drugs* **2010**, *21* (6), 632-644.

(87) Manjegowda, M. C.; Deb, G.; Limaye, A. M. Epigallocatechin gallate induces the steady state mRNA levels of pS2 and PR genes in MCF-7 breast cancer cells. **2014**.

(88) Luo, T.; Wang, J.; Yin, Y.; Hua, H.; Jing, J.; Sun, X.; Li, M.; Zhang, Y.; Jiang, Y. (-)-Epigallocatechin gallate sensitizes breast cancer cells to paclitaxel in a murine model of breast carcinoma. *Breast Cancer Research* **2010**, *12* (1), 1-10.

(89) Qi, Z.; Tang, T.; Sheng, L.; Ma, Y.; Liu, Y.; Yan, L.; Qi, S.; Ling, L.; Zhang, Y. Salidroside inhibits the proliferation and migration of gastric cancer cells via suppression of Src-associated signaling pathway activation and heat shock protein 70 expression. *Molecular medicine reports* **2018**, *18* (1), 147-156.

(90) Lage, N. N.; Layosa, M. A. A.; Arbizu, S.; Chew, B. P.; Pedrosa, M. L.; Mertens-Talcott, S.; Talcott, S.; Noratto, G. D. Dark sweet cherry (Prunus avium) phenolics enriched in anthocyanins exhibit enhanced activity against the most aggressive breast cancer subtypes without toxicity to normal breast cells. *Journal of Functional Foods* **2020**, *64*, 103710.

(91) Jeon, Y.-W.; Suh, Y. J. Synergistic apoptotic effect of celecoxib and luteolin on breast cancer cells. *Oncology reports* **2013**, *29* (2), 819-825.

(92) Parker-Athill, E.; Luo, D.; Bailey, A.; Giunta, B.; Tian, J.; Shytle, R. D.; Murphy, T.; Legradi, G.; Tan, J. Flavonoids, a prenatal prophylaxis via targeting JAK2/STAT3 signaling to oppose IL-6/MIA associated autism. *Journal of neuroimmunology* **2009**, *217* (1-2), 20-27.

(93) Chiang, C.-T.; Way, T.-D.; Lin, J.-K. Sensitizing HER2-overexpressing cancer cells to luteolin-induced apoptosis through suppressing p21WAF1/CIP1 expression with rapamycin. *Molecular cancer therapeutics* **2007**, *6* (7), 2127-2138.

(94) Voss, O. H.; Arango, D.; Tossey, J. C.; Calero, M. A. V.; Doseff, A. I. Splicing reprogramming of TRAIL/DISC-components sensitizes lung cancer cells to TRAIL-mediated apoptosis. *Cell death & disease* **2021**, *12* (4), 1-12.

(95) Tong, X.; C Pelling, J. Targeting the PI3K/Akt/mTOR axis by apigenin for cancer prevention. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* **2013**, *13* (7), 971-978.

(96) Jia, L.; Huang, S.; Yin, X.; Zan, Y.; Guo, Y.; Han, L. Quercetin suppresses the mobility of breast cancer by suppressing glycolysis through Akt-mTOR pathway mediated autophagy induction. *Life sciences* **2018**, *208*, 123-130.

(97) Harrison, M. E.; Coombs, M. R. P.; Delaney, L. M.; Hoskin, D. W. Exposure of breast cancer cells to a subcytotoxic dose of apigenin causes growth inhibition, oxidative stress, and hypophosphorylation of Akt. *Experimental and molecular pathology* **2014**, *97* (2), 211-217.

(98) Seo, H.-S.; Ku, J. M.; Choi, H.-S.; Woo, J.-K.; Jang, B.-H.; Shin, Y. C.; Ko, S.-G. Induction of caspase-dependent apoptosis by apigenin by inhibiting STAT3 signaling in HER2-

overexpressing MDA-MB-453 breast cancer cells. *Anticancer research* **2014**, *34* (6), 2869-2882.

(99) Wu, H.-T.; Lin, J.; Liu, Y.-E.; Chen, H.-F.; Hsu, K.-W.; Lin, S.-H.; Peng, K.-Y.; Lin, K.-J.; Hsieh, C.-C.; Chen, D.-R. Luteolin suppresses androgen receptor-positive triple-negative breast cancer cell proliferation and metastasis by epigenetic regulation of MMP9 expression via the AKT/mTOR signaling pathway. *Phytomedicine* **2021**, *81*, 153437.

(100) Way, T.-D.; Kao, M.-C.; Lin, J.-K. Apigenin induces apoptosis through proteasomal degradation of HER2/neu in HER2/neu-overexpressing breast cancer cells via the phosphatidylinositol 3-kinase/Akt-dependent pathway. *Journal of Biological Chemistry* **2004**, *279* (6), 4479-4489.

(101) Rusak, G.; Gutzeit, H. O.; Ludwig-Müller, J. Effects of structurally related flavonoids on hsp gene expression in human promyeloid leukaemia cells. *Food Technology and Biotechnology* **2002**, *40* (4), 267-273.

(102) Way, T.-D.; Kao, M.-C.; Lin, J.-K. Degradation of HER2/neu by apigenin induces apoptosis through cytochrome c release and caspase-3 activation in HER2/neu-overexpressing breast cancer cells. *Febs Letters* **2005**, *579* (1), 145-152.

(103) Cook, M. T. Mechanism of metastasis suppression by luteolin in breast cancer. *Breast Cancer: Targets and Therapy* **2018**, *10*, 89.

(104) Lin, C.-W.; Hou, W.-C.; Shen, S.-C.; Juan, S.-H.; Ko, C.-H.; Wang, L.-M.; Chen, Y.-C. Quercetin inhibition of tumor invasion via suppressing PKCδ/ERK/AP-1-dependent matrix metalloproteinase-9 activation in breast carcinoma cells. *Carcinogenesis* **2008**, *29* (9), 1807-1815.

(105) Balakrishnan, S.; Bhat, F. A.; Raja Singh, P.; Mukherjee, S.; Elumalai, P.; Das, S.; Patra, C. R.; Arunakaran, J. Gold nanoparticle–conjugated quercetin inhibits epithelial–mesenchymal transition, angiogenesis and invasiveness via EGFR/VEGFR-2-mediated pathway in breast cancer. *Cell proliferation* **2016**, *49* (6), 678-697.

(106) Monti, E.; Marras, E.; Prini, P.; Gariboldi, M. B. Luteolin impairs hypoxia adaptation and progression in human breast and colon cancer cells. *European Journal of Pharmacology* **2020**, *881*, 173210.

(107) Zhang, Y.; Zhang, R.; Ni, H. Eriodictyol exerts potent anticancer activity against A549 human lung cancer cell line by inducing mitochondrial-mediated apoptosis, G2/M cell cycle arrest and inhibition of m-TOR/PI3K/Akt signalling pathway. *Archives of medical science: AMS* **2020**, *16* (2), 446.

(108) Linsalata, M.; Orlando, A.; Messa, C.; Refolo, M. G.; Russo, F. Quercetin inhibits human DLD-1 colon cancer cell growth and polyamine biosynthesis. *Anticancer research* **2010**, *30* (9), 3501-3507.

(109) Wu, L.; Li, J.; Liu, T.; Li, S.; Feng, J.; Yu, Q.; Zhang, J.; Chen, J.; Zhou, Y.; Ji, J. Quercetin shows anti-tumor effect in hepatocellular carcinoma LM3 cells by abrogating JAK2/STAT3 signaling pathway. *Cancer medicine* **2019**, *8* (10), 4806-4820.

(110) Huang, C.; Wei, Y.-X.; Shen, M.-C.; Tu, Y.-H.; Wang, C.-C.; Huang, H.-C. Chrysin, abundant in morinda citrifolia fruit water–etoac extracts, combined with apigenin synergistically induced apoptosis and inhibited migration in human breast and liver cancer cells. *Journal of agricultural and food chemistry* **2016**, *64* (21), 4235-4245.

(111) Ahmed, S.; Khan, H.; Fratantonio, D.; Hasan, M. M.; Sharifi, S.; Fathi, N.; Ullah, H.; Rastrelli, L. Apoptosis induced by luteolin in breast cancer: Mechanistic and therapeutic perspectives. *Phytomedicine* **2019**, *59*, 152883.

(112) Lee, D. H.; Lee, Y. J. Quercetin suppresses hypoxia-induced accumulation of hypoxia-inducible factor-1α (HIF-1α) through inhibiting protein synthesis. *Journal of cellular biochemistry* **2008**, *105* (2), 546-553.

(113) Chen, X.; Gu, N.; Xue, C.; Li, B. R. Plant flavonoid taxifolin inhibits the growth, migration and invasion of human osteosarcoma cells. *Molecular Medicine Reports* **2018**, *17* (2), 3239-3245.

(114) Jiang, Y.; Xie, K.-P.; Huo, H.-N.; Wang, L.-M.; Zou, W.; Xie, M.-J. Inhibitory effect of luteolin on the angiogenesis of chick chorioallantoic membrane and invasion of breast cancer cells via downregulation of AEG-1 and MMP-2. *Sheng li xue bao:[Acta physiologica Sinica]* **2013**, *65* (5), 513-518.

(115) Lv, F.; Du, Q.; Li, L.; Xi, X.; Liu, Q.; Li, W.; Liu, S. Eriodictyol inhibits glioblastoma migration and invasion by reversing EMT via downregulation of the P38 MAPK/GSK-3β/ZEB1 pathway. *European Journal of Pharmacology* **2021**, *900*, 174069.

(116) Balakrishnan, S.; Mukherjee, S.; Das, S.; Bhat, F. A.; Raja Singh, P.; Patra, C. R.; Arunakaran, J. Gold nanoparticles–conjugated quercetin induces apoptosis via inhibition of EGFR/PI3K/Akt–mediated pathway in breast cancer cell lines (MCF-7 and MDA-MB-231). *Cell biochemistry and function* **2017**, *35* (4), 217-231.

(117) Li, X.; Zhou, N.; Wang, J.; Liu, Z.; Wang, X.; Zhang, Q.; Liu, Q.; Gao, L.; Wang, R. Quercetin suppresses breast cancer stem cells (CD44+/CD24−) by inhibiting the PI3K/Akt/mTOR-signaling pathway. *Life sciences* **2018**, *196*, 56-62.

(118) Tang, Q.; Gonzales, M.; Inoue, H.; Bowden, G. T. Roles of Akt and glycogen synthase kinase 3β in the ultraviolet B induction of cyclooxygenase-2 transcription in human keratinocytes. *Cancer research* **2001**, *61* (11), 4329-4332.

(119) Yao, W.; Gong, H.; Mei, H.; Shi, L.; Yu, J.; Hu, Y. Taxifolin Targets PI3K and mTOR and Inhibits Glioblastoma Multiforme. *Journal of Oncology* **2021**, *2021*.

(120) Bo-xue, R. E. N.; Huan-huan, R. E. N.; Xiao-yu, C.; De-fang, L. I.; Qiu-sheng, Z. Effects of Rosmarinic Acid on Cell Proliferation, Apoptosis and Migration in Breast Cancer MDA-MB-231 Cells. *Natural Product Research and Development* **2018**, *30* (8), 1300.

(121) Gaballah, H. H.; Gaber, R. A.; Mohamed, D. A. Apigenin potentiates the antitumor activity of 5-FU on solid Ehrlich carcinoma: Crosstalk between apoptotic and JNK-mediated autophagic cell death platforms. *Toxicology and applied pharmacology* **2017**, *316*, 27-35.

(122) Kim, M. J.; Woo, J. S.; Kwon, C. H.; Kim, J. H.; Kim, Y. K.; Kim, K. H. Luteolin induces apoptotic cell death through AIF nuclear translocation mediated by activation of ERK and p38 in human breast cancer cell lines. *Cell biology international* **2012**, *36* (4), 339-344.

(123) Chen, K.-C.; Chen, C.-Y.; Lin, C.-J.; Yang, T.-Y.; Chen, T.-H.; Wu, L.-C.; Wu, C.-C. Luteolin attenuates TGF-β1-induced epithelial–mesenchymal transition of lung cancer cells by interfering in the PI3K/Akt–NF-κB–Snail pathway. *Life sciences* **2013**, *93* (24), 924-933.

(124) Mirzoeva, S.; Franzen, C. A.; Pelling, J. C. Apigenin inhibits TGF-β-induced VEGF expression in human prostate carcinoma cells via a Smad2/3-and Src-dependent mechanism. *Molecular carcinogenesis* **2014**, *53* (8), 598-609.

(125) Wang, L.; Yang, H.; Wang, C.; Shi, X.; Li, K. Rosmarinic acid inhibits proliferation and invasion of hepatocellular carcinoma cells SMMC 7721 via PI3K/AKT/mTOR signal pathway. *Biomedicine & Pharmacotherapy* **2019**, *120*, 109443.

(126) Yuan, X.; Li, N.; Zhang, M.; Lu, C.; Du, Z.; Zhu, W.; Wu, D. Taxifolin attenuates IMQ-induced murine psoriasis-like dermatitis by regulating T helper cell responses via Notch1 and JAK2/STAT3 signal pathways. *Biomedicine & Pharmacotherapy* **2020**, *123*, 109747.

(127) Fu, J.; Chen, D.; Zhao, B.; Zhao, Z.; Zhou, J.; Xu, Y.; Xin, Y.; Liu, C.; Luo, L.; Yin, Z. Luteolin induces carcinoma cell apoptosis through binding Hsp90 to suppress constitutive activation of STAT3. *PloS one* **2012**, *7* (11), e49194.

(128) Park, S.-H.; Ham, S.; Kwon, T. H.; Kim, M. S.; Lee, D. H.; Kang, J.-W.; Oh, S.-R.; Yoon, D.-Y. Luteolin induces cell cycle arrest and apoptosis through extrinsic and intrinsic signaling pathways in MCF-7 breast cancer cells. *Journal of Environmental Pathology, Toxicology and Oncology* **2014**, *33* (3).

9. SUMMARY AND CONCLUSIONS

The major findings obtained from all the studies presented in this dissertation are summarized below:

Modeling the Behavior of Organic Kerogen in the Proximity of Calcite Mineral by Molecular Dynamics Simulations: Interaction with calcite mineral causes the separation of kerogen layers where only the adjacent kerogen layer interacts with the mineral. The relative magnitude of interactions provided by kerogen fragments remain identical for all orientations of kerogen and calcite. Kerogen-calcite interactions are primarily electrostatic and kerogen fragments containing olefinic hydrocarbons carry out the major portion of interactions. The ammonium ions of kerogen do not interact significantly with calcite mineral. Presence of calcite mineral brings changes (increase/decrease) in inter-fragment interactions within kerogen without altering their interaction pattern (attractive/repulsive) i.e., attractive interactions remain attractive and repulsive interactions remain repulsive. Through this study, we have been able to determine the interactions between kerogen and calcite mineral for the very first time. Blocking the determined interactions would help to isolate kerogen from carbonate minerals.

An insight into quartz mineral interactions with kerogen in Green River oil shale: The extent and pattern of kerogen-quartz interaction depends on the initial orientation of kerogen and quartz. Kerogen-quartz interaction is primarily electrostatic and kerogen fragments containing nitrogen or aromatics impart the total interactions. The ammonium ions of kerogen occupy the cavities of quartz supercell while having the largest amount of interactions. The presence of quartz mineral significantly changes the inter-layer interactions within the kerogen. Through this study we have been able to identify the interactions between Type I kerogen macromolecule of Green

River oil shale and quartz mineral. Blocking the determined interactions would help to isolate kerogen from silicate minerals.

Molecular mechanics of the swelling clay tactoid under compression, tension and shear: The average d-spacing of Na-MMT tactoid is 10.8 Å. The clay sheets within Na-MMT tactoid attract each other with equal amount of energies. In terms of binding energy, the boundary clay sheets are loosely bonded compared to other sheets as they have interlayer cations just on one (the inner) side. The maximum permissible compressive stress on clay tactoid is 29.6 GPa which causes the topmost interlayer to be compressed equal to its effective interlayer spacing (0.95 Å). The required tensile stress increases as the number of clay sheets to be pulled from tactoid increases. The shearing of tactoid causes the clay layers slide past each other sequentially until the interlayer cations lock this motion. This is one of the earliest studies that provides structural and mechanical information of clay hierarchical structure.

An in-silico investigation of hydrated Na-montmorillonite tactoid: Hydration increases the average d-spacing of Na-MMT tactoid compared to dry one and increasing hydration even increases the average d-spacing. The average d-spacing of Na-MMT tactoid due to 10%, 20%, and 30% (wt.) interlayer hydration results in the average d-spacing of 13.11 Å, 14.45 Å, and 16.04 Å respectively. Increasing hydration reduces the interactions between clay sheets and clay sheets with cations. The compression behavior of hydrated tactoid shows that increasing hydration gradually reduces the compression strength of tactoid. The compression strength of 10%, 20%, and 30% hydrated is computed as 14.8 GPa, 5.9 GPa, and 4.5 GPa. This is one of the earliest studies that indicate how the Na-MMT hierarchical structure, and its mechanical properties change upon hydration.

Binding of SARS-COV-2 (COVID-19) and SARS-COV to human ACE2: Identifying binding sites and consequences on ACE2 stiffness: The spike protein of SARS-CoV-2 has higher (almost twice) interactions with human ACE2 compared to SARS-CoV spike. Regarding the primary structure, two residues (GLU, ASP) of ACE2 interact with SARS-CoV-2 spike while four residues (GLU, ASP, LYS, GLN) of ACE2 interact with SARS-CoV spike receptor binding domain (RBD). Considering the secondary structure, the helices and turns of ACE2 carry out the majority interactions with spike proteins of SARS-CoV-2 and SARS-CoV respectively. The presence of spike stiffens the ACE2 helices which is six times higher for SARS-CoV-2 than SARS-CoV. The higher interactions of spike-ACE2 and higher increment in ACE2 stiffness may explain the downregulation of ACE2 in epithelial cells and the higher occurrences of ARDS caused by SARS-CoV-2.

Differences in Interactions Within Viral Replication Complexes of SARS-CoV-2 (COVID-19) and SARS-CoV Coronaviruses Control RNA Replication Ability: In the SARS-CoV non-structural tri-protein complex (nsp12-nsp7-nsp8), the RdRp domain of nsp12 has higher interactions with nsp7 than nsp8. On the contrary, the RdRp domain of SARS-CoV-2 has higher interactions with nsp8 than nsp7. Higher RdRp-nsp7 interactions may contribute to the higher replication of SARS-CoV in the intestinal CaCo2 cells. Higher RdRp-nsp8 interactions might result in the higher replication of SARS-CoV-2 in bronchial epithelium cells leading to the higher transmissibility of SARS-CoV-2. The interactions of SARS-CoV-2 RdRp with co-factors (nsp7 and nsp8) change when Remdesivir Monophosphate binds to the RdRp. This change in interactions might block the regular replication behavior of SARS-CoV-2.

Virtual Screening of Rhodiola and Oregano Compounds for Breast Cancer Drug Discovery Utilizing Molecular Docking: Different compounds from Rhodiola, and Oregano have

shown higher binding affinity to breast cancer proteins. Epigallocatechin-gallate (EGCG) is the most potent compound of Rhodiola which shows higher binding affinity with Akt1, Akt2, HIF-1α, ERK2, PR and JNK1. Regarding the Oregano extract, different derivatives of Luteolin, Apigenin, and Quercetin show higher binding affinity with Akt1, Akt2, JAK2, HER2, HSP70, mTOR etc. A significant amount of literature has been found that are consistent with docking results i.e., the effect of specific phytochemical on specific protein has been found to be effective for breast cancer treatment. The study provides virtual screening results of different phytochemicals for breast cancer treatment.

10. FUTURE WORK

The future directions of the studies presented in this dissertation are summarized below:

Oil shale: This study determines the interactions of kerogen with two significantly present minerals in Green River oil shale i.e., calcite ($CaCO_3$) and quartz (SiO_2). Calcite belongs to carbonate, and quartz belongs to the silicate mineral class. The most abundant carbonate mineral present in the Green River oil shale is Dolomite [$CaMg(CO_3)_2$]. Therefore, determining the interactions of kerogen with dolomite will help to better understand kerogen-carbonate interactions in Green River oil shale. Pyrite (FeS_2) is the only sulfide mineral present in the Green River oil shale. The kerogen-pyrite interactions might indicate the effect of sulfide minerals on the Green River oil shale. Molecular dynamics (MD) simulations can be utilized to model kerogen-dolomite and kerogen-pyrite interactions.

Na-MMT clay: The next hierarchical clay structure beyond tactoid is clay aggregate. A number of clay tactoids are randomly oriented to form a clay aggregate. The behavior of clay sheets and tactoids within aggregates needs to be investigated to explain the hydration that takes place outside the interlayer spaces. The interaction behavior and mechanical properties of clay aggregate both in dry and hydrated conditions can be investigated utilizing MD and SMD simulations.

Coronaviral proteins: Viral protein-protein interactions (PPI) contribute to the different phenomena of the viral life cycle and can explain their infection mechanism. This study identifies some discrepancies between SARS-CoV and SARS-CoV-2 in terms of PPI (spike-ACE2 and nsp12-nsp7-nsp8) and mechanobiological (ACE2 stiffness) behavior. These differences are hypothesized to be the reasons behind the higher contagiousness and deadliness of SARS-CoV-2. These spike-receptor interactions, receptor stiffness, and replication protein complex interactions

can be used as a tool for predicting the infection behavior of other previous and potential future coronaviruses. MD and SMD simulations can be employed to investigate the previously emerged MERS, HCoV-OC43 coronaviruses and the newly emerging mutations like SARS-CoV-2 delta.

Phytochemicals for cancer drug discovery: The phytochemicals utilized in this study for targeting breast cancer proteins can also be investigated for prostate cancer drug discovery. A number of breast cancer proteins also contribute to prostate cancer. Moreover, a certain number of proteins can additionally be targeted for prostate cancer drug discovery. Molecular docking of phytochemicals against prostate cancer proteins can identify potent drug candidates. Besides, experimental studies can be performed regarding breast cancer drug discovery. From molecular docking, it is already known which compound binds to which protein and which functional groups are interacting. A protein-ligand complex with high binding affinity determined by molecular docking can be prepared and undergone through FTIR spectroscopy. If the experimental band shifts match with identified functional groups from docking, then the results from molecular docking can be validated.