Molecular Dynamics Simulations of Aragonite

by

© **Parisa Mirzaei**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science.

Department of Physics and Physical Oceanography
Memorial University

December 2024

St. John's, Newfoundland and Labrador, Canada

Abstract

This work presents a molecular dynamics simulation study of aragonite. Aragonite is a naturally occurring crystal form of calcium carbonate found in shells of marine organisms and sedimentary rocks. The vibration of atoms inside the crystal is fundamental to mineral science and solid-state physics studies. Raman and infrared spectroscopy are the two primary techniques for vibrational spectroscopy. These methods provide unique molecular fingerprints for a sample that identifies crystal structures. The vibrational density of states (VDOS) is a key parameter accessible through simulations. A VDOS is a histogram counting the number of the vibrational modes of a given energy and depends on how interactions between atoms are modelled. A force field represents these interactions. There are two recently developed force fields for aragonite. One includes explicit bonding terms between carbon and oxygen atoms within carbonate units (Bond), and the other does not include the bond (No-Bond). These aragonite force fields were developed to work with the OPLS-AA force field, which is commonly used for aqueous and biological systems. In this thesis, our first goal is to reproduce reported results in the literature for the No-Bond model, primarily unit cell parameters, and to test the effect of including the C-O bond on these parameters and on VDOS. As part of this goal, we study the impact of simulation details on the results, including those related to the control of pressure and temperature. Having determined a simulation protocol, we calculate VDOS for the two force field models. The generated VDOS plot for the No-Bond model shows the qualitative agreement with experimental data. Incorporating the spring bond between the carbon and oxygen atoms significantly and detrimentally changes the overall VDOS structure and mode energies.

Acknowledgements

I would like to express my special thanks to my supervisors, Dr. Ivan Saika-Voivod and Dr. Kristin M. Poduska, for providing me with this excellent opportunity to do this project. I gratefully acknowledge the Department of Physics and Physical Oceanography and the School of Graduate Studies at the Memorial University of Newfoundland for providing financial support for this project. Finally, I would like to thank my family for their support.

Statement of contribution

This thesis is the result of my own work, completed under the supervision and guidance of Dr. Ivan Saika-Voivod and Dr. Kristin M. Poduska. I performed all the simulations, collected and analyzed the results and wrote the corresponding sections. My supervisors, Dr. Ivan Saika-Voivod and Dr. Kristin M. Poduska provided me with comments and feedback.

Contents

List of Tables

List of Figures

List of abbreviations

AMCSD	American Mineralogist Crystal Structure Database
BH	Buckingham
BR	Berendsen
CG	Congjugate Gradient
DFT	Density Functional Theory
EM	Energy Minimization
GROMACS	GROningen MAchine for Chemical Simulations
GULP	General Utility Program
IR	Infrared
LJ	Lennard Jones
MD	Molecular Dynamics
MDP	Molecular Dynamics Parameters
NMA	Normal Mode Analysis
NPT	constant Number of particles, Pressure, Temperature
NVT	constant Number of particles, Volume, and Temperature
OPLS-AA	Optimized Potentials for Liquid Simulations (All-Atom)
PR	Parrinello-Rahman
PME	Particle Mesh Ewald
VDOS	Vibrational Density Of States
VMD	Visual Molecular Dynamics

Chapter 1

Introduction

Calcium carbonate ($CaCO_3$) is one of the most abundant materials on Earth, constituting approximately 4% of the Earth's crust [1]. This material has numerous industrial applications, including construction, paper production, pharmaceuticals, and biomedicine [2, 3, 4, 5]. There are three crystalline polymorphs of calcium carbonate: aragonite, calcite, and vaterite. These polymorphs have the same chemical formula but different crystal structures. Aragonite is a mineral that is mainly found in marine sediments. This crystal is the main component of the shells, pearls, coral reefs, and other biominerals [6, 7].

Studying the vibrational properties of crystals has always been an interesting topic in solid-state physics. The vibrational spectroscopic technique offers a non-destructive way to measure the vibrational energy levels related to the oscillations of atoms within the crystal. Two primary types of vibrational spectroscopy are infrared (IR) and Raman spectroscopy. These techniques provide unique crystallographic fingerprints for the sample that identify the crystal structure [8]. One of the key concepts in this area is the vibrational density of states (VDOS). This quantity can be used to study different properties of the materials, including vibrational, thermodynamic, and transport properties [9].

VDOS is typically shown as a histogram of the number of vibrational modes for each frequency. In this way, the peaks in a VDOS histogram correspond to collections of normal modes with approximately the same frequency. VDOS is accessible through simulations. GROMACS (Groningen Machine for Chemical Simulations) is a free and open-source molecular dynamics (MD) software package that can be used to

calculate the VDOS. This software solves Newton's equation of motion for a system that contains hundreds to millions of particles and includes tools to calculate VDOS [10, 11, 12, 13, 14, 15, 16]. A VDOS enumerates the number of vibrational modes of a given energy, and in simulations, it depends on how interactions between atoms are modelled. These interactions are represented by a force field, which is a collection of all the interactions in the system. Two recently developed force fields have been used to study aragonite [17, 18]. In this thesis, these force fields are used to conduct molecular dynamics simulations of aragonite and generate VDOS.

1.1 Atomic Vibration in Crystals

A simple classical mass-spring model can be used to study the vibrations of the atoms in a crystal. Based on this model, the atoms in the crystal are the masses, and the bonds between the atoms are the springs. Considering Hooke's law for a system of two masses, m_1 and m_2, which are separated by r, the potential energy is,

$$V = \frac{1}{2}K(r - r_0)^2, \tag{1.1}$$

where K is the force constant and r_0 is the equilibrium separation. For this system, the frequency of vibration is,

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{\mu}}, \tag{1.2}$$

where μ indicates the reduced mass, which is given by,

$$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2}. \tag{1.3}$$

A lattice can be considered a system of masses and springs in three dimensions. Every two neighbouring masses are connected by a spring. The vibration of this system is not as simple as the system of two masses; however, the fundamentals are the same. In a crystal, atoms vibrate around their equilibrium positions, and the chemical bonds between atoms act like springs with specific force constants. Therefore, the frequencies of the oscillations can be determined using a similar approach to that of a simple mass-spring system. However, we must consider the complexity of a many-body system in three dimensions. For a harmonic solid such as the spring-mass

systems, the vibrational motion of atoms can be decomposed into contributions from independent modes of collective vibrations or normal modes [19].

In a real crystal, the vibrational energy of these modes is quantized. A phonon, which is a quantum of lattice vibrational energy, has an energy that is proportional to the frequency of vibration and that can be probed with spectroscopy [9]. Similar to how a photon is a quantum of electromagnetic energy, a phonon represents the discrete energy associated with atomic vibrations [9]. The normal mode frequencies obtained from classical MD simulations thus provide a link to spectroscopic experiments.

1.2 Normal Mode Analysis

Normal modes of a system are vibrational modes in which all the system's atoms vibrate with the same frequency. To study the normal modes of a system, if the generalized Cartesian coordinates of the system relative to the equilibrium are denoted by q_k, it is expedient to expand the potential energy of the system around equilibrium [11, 19],

$$V(q_k) = V(0) + \sum_k \left(\frac{\partial V}{\partial q_k}\right)_0 + \frac{1}{2}\sum_{j,k} q_j \left(\frac{\partial^2 V}{\partial q_j \partial q_k}\right)_0 q_k. \tag{1.4}$$

For this system, the kinetic energy can be written as,

$$T = \frac{1}{2}\sum_j m_j \left(\frac{dq_j}{dt}\right)^2, \tag{1.5}$$

where m_j is the mass associated with the jth coordinate. Considering the potential at equilibrium to be zero, Eq. 1.4 can be written to leading order, since the first-order derivatives are zero, as,

$$V = \frac{1}{2}\sum_{j,k} q_j V_{jk} q_k. \tag{1.6}$$

In this equation, V_{jk} indicates the second derivative of the potential energy. Using Eqs. 1.5 and 1.6, we can calculate the equation of motion from the Lagrangian,

$$m_j \frac{d^2 q_j}{dt^2} = -\sum_k V_{jk} q_k. \tag{1.7}$$

Introducing the mass-weighted Cartesian coordinates as,

$$x_j = q_j\sqrt{m_j}, \tag{1.8}$$

Eq. 1.7 in this coordinate can be written as,

$$\frac{d^2x_j}{dt^2} = -\sum_k H_{jk}x_k. \tag{1.9}$$

where $H_{jk} = \frac{V_{jk}}{\sqrt{m_j m_k}}$ is called the mass-weighted Hessian matrix. Eq. 1.9 is an equation for harmonic motion. Therefore, a solution for x_j can be written as,

$$x_j(t) = \sum_m A_{jm}\cos(\omega_m t), \tag{1.10}$$

where A_{jm} is the coefficient of the contribution to the solution for $x_j(t)$ from the normal mode of angular frequency ω_m. Substitution of Eq. 1.10 into the Eq. 1.9 results in,

$$\omega_m^2 A_{jm} = \sum_k H_{jk}A_{km}. \tag{1.11}$$

This equation indicates an eigenvalue equation. In matrix form, Eq. 1.11 can be written as,

$$\mathbf{HA} = \lambda\mathbf{A}. \tag{1.12}$$

where λ is a diagonal matrix of eigenvalues with elements $\lambda_m = \omega_m^2$. In this equation, each column of A represents an eigenvector corresponding to a particular normal mode. The eigenvectors indicate the displacements of particles in the system with respect to one another, and the eigenvalues are the squared angular frequencies of the normal modes.

For a system of N atoms, there are $3N$ degrees of freedom. The number of non-zero normal modes for this system is $3N - 6$, since the translations and rotations of the entire system are zero-frequency modes. In a solid like $CaCO_3$, a carbonate unit behaves like a molecule. Therefore, similar to a molecule, two important types of vibrations can be considered for the system: stretching and bending. Stretching vibrations involve changes in the bond length and are divided into two categories: symmetric and asymmetric. Bending vibrational modes involve the angle changes between bonds [20].

1.3 Raman and Infrared Spectroscopy

Raman and IR spectroscopy are the two primary techniques for vibrational spectroscopy. Raman spectroscopy is a method based on the inelastic scattering of light from a sample. This technique was introduced by Raman in 1928 [21]. In this method, when light interacts with a sample, the frequency of the scattered light differs from that of the incident light. The shift in the frequency corresponds to specific vibrational modes.

In IR spectroscopy, infrared radiation passes through the sample, and the resulting spectrum is analyzed using a spectrometer. In this method, the vibrational modes of the sample are excited when infrared light is absorbed. Typically, the spectrum shows the intensity of absorbed light versus wavenumber [8, 22]. In recent years, Raman and IR spectroscopy have attracted more attention in structural studies. Examples of uses of these techniques include the study of structural differences in mineral samples [23, 24], and structural deformation [25].

1.4 Aragonite

Aragonite has an orthorhombic symmetry and belongs to the Pmcn space group. The crystalline axes are perpendicular in this symmetry, but their lengths differ. The cell parameters of aragonite at ambient conditions are $a = 4.960$ Å, $b = 7.7964$ Å, and $c = 5.5738$ Å [26]. Figure 1.1 illustrates the unit cell of aragonite. The unit cell contains four $CaCO_3$ units, represented by blue (carbon atoms), red (oxygen atoms), and green (calcium ions) spheres. This configuration results in a total of 20 atoms per unit cell [27]. The carbonate unit in aragonite, CO_3, is nearly planar. This means that the carbon atom is not quite in the same plane as the oxygen atoms. It is positioned 0.020 Å out of the oxygen atoms plane [28].

1.5 Outline

This thesis presents a molecular dynamics (MD) simulation study of aragonite using GROMACS to calculate the VDOS through normal mode analysis. It is organized

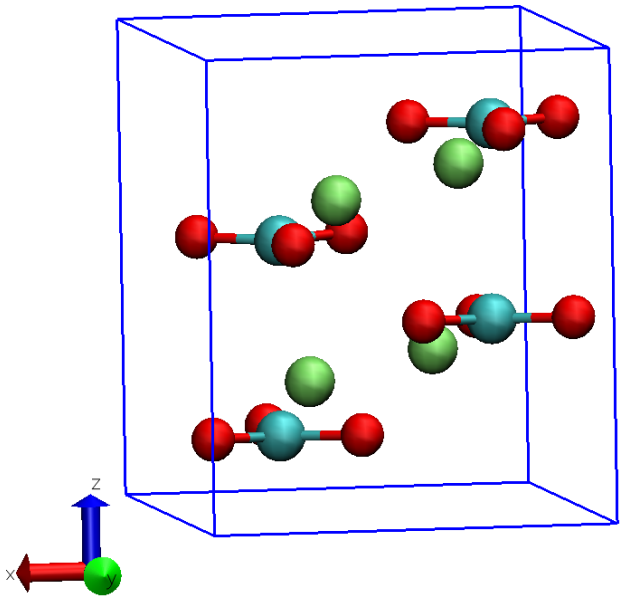


Figure 1.1: Aragonite unit cell, which contains four $CaCO_3$ molecules, totaling 20 atoms. Calcium ions are represented by green spheres, and CO_3 ions are shown with the blue and red spheres.

as follows. In Chapter 2, the methods used to simulate aragonite are explained: The theory of MD simulations, the stages involved in performing MD simulations, and the required input parameters are described. In Chapter 3, two recently developed force fields for aragonite are tested. One force field does not include an explicit harmonic bond term between the carbon and oxygen in carbonate units [17], referred to as the No-Bond model in this thesis, and the other does explicitly include the harmonic bond [18], referred to as the Bond model. In the first part of Chapter 3, MD simulation is used to calculate the unit cell parameters of aragonite for both models and to compare them with those reported in previous work [17]. In the second part of Chapter 3, simulation input parameters are examined to determine a protocol for calculating the VDOS for both the Bond and No-Bond models. In Chapter 4,

VDOS is calculated through normal mode analysis, and the results are compared with the published results of Carteret et al. [29], who used a density functional theory that successfully recapitulated experimental data. Finally, conclusions and outlook for future work are given in Chapter 5.

Chapter 2

Methodology

In this thesis, molecular dynamics (MD) simulations are used to study the vibrational density of states (VDOS) of aragonite through normal mode analysis. A detailed explanation of MD simulations is available in references [30, 31]. This chapter briefly reviews MD simulations, including all necessary GROMACS input files and the stages that MD simulations follow.

2.1 Molecular Dynamics Simulations

Molecular dynamics simulation is a method that models the interaction of particles classically and uses Newton's equations of motion. In a many-body system, if the mass of particle i is m_i, its position is $\mathbf{r}_i$, and the net force on it is $\mathbf{f}_i$, Newton's equations of motion can be written as [31],

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{f}_i. \tag{2.1}$$

We can use computer programs to solve this equation for a many-body system. In this thesis, GROMACS [32] is used to integrate the equations of motion. This software offers different algorithms to integrate the equations of motion, including the leap-frog [33] (denoted as md in the GROMACS .mdp file described below) and velocity Verlet (denoted as md-vv in the .mdp file) [11]. Both algorithms are variations of the Verlet algorithm. Verlet is a simple algorithm that conserves mechanical energy. To study

this algorithm, we can use the following Taylor series expansions [31],

$$x(t+h) = x(t) + h\frac{dx(t)}{dt} + \frac{1}{2!}h^2\frac{d^2x(t)}{dt^2} + \frac{1}{3!}h^3\frac{d^3x(t)}{dt^3} + \dots \tag{2.2}$$

$$x(t-h) = x(t) - h\frac{dx(t)}{dt} + \frac{1}{2!}h^2\frac{d^2x(t)}{dt^2} - \frac{1}{3!}h^3\frac{d^3x(t)}{dt^3} + \dots \tag{2.3}$$

By adding Eqs. 2.2 and 2.3 the second derivative can be written as,

$$\frac{d^2x(t)}{dt^2} = \frac{x(t+h) - 2x(t) + x(t-h)}{h^2} + \mathcal{O}(h^2). \tag{2.4}$$

Substituting Eq. 2.4 in Eq. 2.1 for the x direction results in,

$$\frac{x_i(t+h) - 2x_i(t) + x_i(t-h)}{h^2} = \frac{f_{xi}}{m}. \tag{2.5}$$

Using Eq. 2.5, Newton's equation of motion in the x direction can be written as,

$$x_i(t+h) = 2x_i(t) - x_i(t-h) + \frac{h^2}{m}f_{xi}(t). \tag{2.6}$$

The Verlet algorithm uses Eq. 2.6 to calculate the positions of the particles in the system at any time step. In this algorithm, velocity is not directly used but is needed to calculate kinetic energy. The velocity is calculated through,

$$\mathrm{v}(t) = \frac{1}{2h}\left(x(t+h) - x(t-h)\right). \tag{2.7}$$

Another algorithm GROMACS offers to solve Newton's equations of motion is velocity Verlet. To use this algorithm, considering the definition of velocity as the first-order derivative of position and acceleration as the second-order derivative of position, the following equations are obtained from Eq. 2.2 [31],

$$\mathbf{r}_i(t+h) = \mathbf{r}_i(t) + h\mathbf{v}_i(t) + \frac{h^2}{2m_i}\mathbf{f}_i(t), \tag{2.8}$$

$$\mathbf{v}_i(t+h) = \mathbf{v}_i(t) + \frac{h}{2m_i}(\mathbf{f}_i(t) + \mathbf{f}_i(t+h)). \tag{2.9}$$

The terms of order higher than h^2 are neglected in these calculations, and the term

$\mathbf{f}_i(t+h)$ is added to the force. This term indicates the force at the next time step. Calculating $\mathbf{v}(t+h)$ can not be performed in one step since $\mathbf{f}(t+h)$ depends on $\mathbf{r}(t+h)$. In this thesis, the velocity Verlet integrator is used to solve Newton's equations of motion.

2.2 Force Field

The interactions between atoms in a system can be categorized into bonded and non-bonded terms. The formula that describes all these interactions is called a force field. A typical representation for a force field might take the following form [34],

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{dihedral} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{elec} \frac{q_i q_j}{r_{ij}}. \tag{2.10}$$

The first four terms in Eq. 2.10 represent the bonded interactions, including bond stretching, angle bending, dihedral, and improper dihedral interactions, respectively. In this equation, k_b is the bond force constant, r_0 is the equilibrium bond length, and r indicates the distance between two bonded atoms for a bond stretching term, modelled by Hooke's law with a harmonic potential. In the angle bending potential, k_a is the angle force constant, θ_0 is the equilibrium angle between three bonded atoms, and θ is the variable angle.

When four atoms (A, B, C, D) are connected, dihedral and improper dihedral energies are introduced to calculate the potential energy. For atoms in a linear chain, V_n is the potential barrier height, and the dihedral (or torsional) angle ϕ is defined as the angle between planes (A, B, C) and (B, C, D). In this equation, n is the number of minima or maxima in a period, and δ is the phase angle. For improper dihedral interaction, V_{imp} is the potential energy. To describe this energy, using the carbonate unit as an example, the improper dihedral angle is defined as the angle between the planes formed by $O_1 - C - O_2$ and $O_2 - C - O_3$.

The non-bonded interactions are represented by the last two terms, which include

the Lennard-Jones (LJ) potential and the Coulombic interactions. The LJ interaction models the van der Waals forces between atoms, with σ_{ij} as the distance with zero potential energy and ϵ_{ij} representing the depth of the potential well for the interaction between atom pairs i and j. The LJ potential includes the repulsive (r^{-12}) and attractive (r^{-6}) terms, with r_{ij} as the interatomic distance between atoms i and j. The last term in Eq. 2.10 indicates the electrostatic interactions, where q_i and q_j are the charges of atoms i and j. This term calculates the energy between two charged particles using Coulomb's law [34].

In this thesis, two recently developed force fields are used to study aragonite. The first force field was developed by Xiao et al. [17]. Before this force field was introduced, the Buckingham (BH) potential [35] was used to model the van der Waals interactions. In the Xiao et al. model, the LJ potential replaced the BH potential. This modification allowed for easier integration with existing force fields for proteins and water, making it a practical choice for future studies involving these components. In addition, using the LJ potential results in faster calculations than the BH potential, which includes an exponential term [17]. The following equations represent BH and LJ potentials,

$$V_{BH} = A \exp\left(-\frac{r}{\rho}\right) - \frac{C}{r^6}, \tag{2.11}$$

$$V_{LJ} = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6}. \tag{2.12}$$

In these equations, r indicates the distance between atoms. A, ρ, C, C_6, and C_{12} are BH and LJ potentials parameters. The force field developed by Xiao et al. includes the following terms [17],

$$\begin{aligned} V_{total} &= V_{nonbond} + V_{angle} + V_{dihedral} \\ &= \sum_{i=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \frac{e^2 z_i z_j}{r_{ij}} + \left(\frac{c_{12}}{r_{ij}^{12}} - \frac{c_6}{r_{ij}^{6}} \right) \right\} + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ &+ \sum_{plane} K_\psi [1 + cos(2\psi)]. \end{aligned} \tag{2.13}$$

The first two terms indicate non-bonded electrostatic and LJ potentials. The last two terms show bonded interactions, including angle potential, which includes θ, the O$-$C$-$O angle, and improper dihedral potential with angle Ψ being the angle between

planes $O_1 - C - O_2$ and $O_2 - C - O_3$.

In this model, there are no bond terms between carbon and oxygen atoms within the carbonate unit, which is referred to as the No-Bond model in this thesis. To confirm their model, Xiao et al. compared the bulk aragonite parameters with experimental results. They studied unit-cell structural parameters, axial elastic modulus, and shear modulus. These parameters were studied through calculations using GULP [36] (general utility program designed for molecular dynamics simulations). The second force field used in this thesis was introduced by Cruz-Chu et al. [18] and is referred to as the Bond model in this thesis. This force field is a modified version of the Xiao et al. force field and includes an explicit bond term between carbon and oxygen atoms within carbonate units. This extra bond term is,

$$V_b(r_{\mathrm{CO}}) = 0.5 \times k_b(r_{\mathrm{CO}} - b_{\mathrm{CO}})^2, \tag{2.14}$$

where r_{CO} is the separation between carbon and oxygen in the carbonate unit, and k_b and b_{CO} are constants. The harmonic bond replaces the Coulombic and LJ interactions between the bonded atoms. This force field was used to study the flaw insensitivity of aragonite nacre. Nacre, known as the mother of pearl, is a natural material with outstanding fracture resistance and toughness. These outstanding properties of nacre are related to its composition, including 95% of aragonite and 5% of an organic layer. Cruz-Chu et al. conducted MD simulations using GROMACS and reported that the organic layer increases the mechanical stability of nacre at the nanoscale. Compared to the Xiao et al. model, adding this stretching bond term between carbon and oxygen preserves the structural integrity of aragonite slabs in nacre and makes the model more suitable for mechanical investigations [18].

2.3 GROMACS Input Files

2.3.1 .top File

The topology (.top) file is one of the input files that GROMACS requires to initiate molecular dynamics simulations. This file contains all the essential information about the system, including the types of particles used in the simulated system and the

parameters that describe the bonded and non-bonded interactions between atoms. The topology file also specifies constraints, which indicate bonds that are treated as rigid constraints during the simulation. Another consideration is exclusions, which specify how many bonds apart two atoms must be before they interact with non-bonded interactions. In GROMACS, the topology is formatted as a simple text file that can be created using the `pdb2gmx` command [11] or from scratch by the user.

In this thesis, the topology file is created in two different ways. Firstly, the supporting information from Cruz-Chu et al. [18], which includes all the initial files needed to create a topology, is used. These files contain detailed information about atoms and their interactions, specified within different text files with extensions like .atp (general information about atom types), .itp file (which shows the bond types and angles), and .rtp (residue topology parameter), which defines a single unit in a large molecule [11]. The `pdb2gmx` tool generates the topology file from the initial files. The topology generated in this way is used in most of the simulations in this thesis.

The second way topology files are created in this thesis is "by hand "without using the `pdb2gmx` command. We start with a single carbonate unit and a $CaCO_3$ unit to study how GROMACS calculates bonded and non-bonded energies. Using the information from Cruz-Chu et al. [18], we generate the topology files for the carbonate and $CaCO_3$ units. Different energies for these systems are calculated using the GROMACS `energy` tool. To help check how precisely the potential energy is calculated by GROMACS, we wrote a Python code that calculates the system's energies. Studying the energies of the CO_3 unit, we confirm that when the bond terms between carbon and oxygen are presented, the LJ and Coulomb potentials (the non-bonded energies) calculated by GROMACS are zero. For the $CaCO_3$ unit, the results indicate that the LJ and Coulomb energies are the sum of the energies between calcium and oxygen atoms ($Ca - O$) and the calcium and carbon atoms ($Ca - C$). GROMACS calculates the bond, angle, and improper dihedral energies, as detailed in Section 2.2.

In the No-Bond model, the LJ and Coulomb energies are non-zero. Therefore, the carbon and oxygen atoms are still considered bonded in the sense that the Coulomb and LJ interactions between them tend to keep an optimal separation, at least for small amplitudes of motion. In this model, the bond, angle, and improper dihedral energies are also calculated by GROMACS.

The topology file, which is created by hand, contains directives such as non-bonded

parameters, atoms, bonds, angles, etc. This topology file is referred to as **our.top** in this study and is included in Appendix B. The **our.top** file also has a simpler structure compared to those from Cruz-Chu et al. [18]. The GROMACS files were not provided by Xiao et al. [17]. Comparing the results we calculated from our topology and the results from modifying the files by Cruz-Chu et al. helps confirm the correct implementation by Xiao et al.

2.3.2 .gro File

One of the input files that GROMACS needs to start the simulations is the file with the extension .gro. This file contains molecular structure information, including the names and numbers of atoms, as well as their coordinates and velocities [11]. In this thesis, to create the .gro file, we use the supporting information provided by Cruz-Chu et al. [18]. This information includes a .pdb file, a standard format for storing atomic coordinates. The .pdb file provided by Cruz-Chu et al. consists of the coordinates of the atoms in a unit cell. Using the `pdb2gmx` command, we generate a .gro file from the initial .pdb file.

Additionally, we create a .gro file for the aragonite unit cell using the crystal structure from the American Mineralogist Crystal Structure Database (AMCSD). Using the symmetry of aragonite in the Pmcn space group and its structural information, we create the .gro file for the aragonite unit cell. This file and the program to generate the simulation box are included in Appendix B.

2.3.3 .mdp File

The molecular dynamic parameter (.mdp) file is another input for GROMACS simulations. This file contains different parameters and settings that describe how the simulation should be conducted, including the integrator to be used, the pressure and temperature control parameters, the time step, and the number of steps [11]. Typically, multiple simulation stages are performed to obtain equilibrium data, which are outlined below. A different .mdp file is used for each simulation stage. The parameters in each .mdp file are adjusted to meet the specific requirements of that particular stage.

2.4 Stages in Performing an MD Simulation

Conducting molecular dynamics (MD) simulations involves several key stages. These stages include initialization, energy minimization, and two equilibration stages: NVT and NPT. In the following sections, we will briefly review these stages and explain how they are employed in our study.

2.4.1 Initialization

The first stage in an MD simulation is knowing the particles' initial positions. For a crystal, the initial positions of the particles are accessible in the crystallographic database. After preparing the unit cell of the system, a larger system can be generated by combining multiple unit cells. As explained in Section 1.4, aragonite has an orthorhombic crystal structure with space group Pmcn. The unit cell of aragonite contains four $CaCO_3$ molecules, totalling 20 atoms, as shown in Figure 1.1. The simulation box in this work contains 60 unit cells, five unit cells in the x direction, three in the y direction, and four in the z direction ($5 \times 3 \times 4$), so that the simulation box is almost cubic. To generate this box, supporting information from Cruz-Chu et al. [18], which includes a script-based tool, and the program VMD [37] is used. We also confirm this by writing our own scripts to generate initial atomic coordinates.

2.4.2 Energy Minimization

Before conducting MD simulations, the energy of the initial configuration should be minimized to stabilize the system. Energy minimization prior to conducting MD simulations reduces numerical instability resulting from large forces that may be present in the system after the initial placement of atoms. In a crystal, the experimentally determined atomic positions may not correspond to the minimum potential energy for the force field employed, which can result in large initial forces. Energy minimization results in the forces on all the particles being zero to within a specified tolerance. The energy function of a system that contains all bonded and non-bonded interactions is related to the positions of the atoms. So, the energy changes by changing the positions.

To minimize the system's energy, the atomic positions are iteratively optimized to achieve a configuration with a lower energy state. In a many-body system such as a crystal, if f is the energy function of the system and the coordinates of the atoms are the x_i, to minimize the energy, two following equations should be satisfied [38],

$$\frac{\partial f}{\partial x_i} = 0, \tag{2.15}$$

and

$$\frac{\partial^2 f}{\partial x_i^2} \geq 0. \tag{2.16}$$

Different numerical methods can be used to minimize the energy function of a complex system like a crystal. GROMACS offers three algorithms for energy minimization: steepest descent (denoted as steep in the GROMACS .mdp file), conjugate gradients (denoted as cg in the .mdp file), and the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian minimizer (denoted as lbfgs in the .mdp file) [11]. The energy minimization algorithms find a local energy minimum rather than a global one, as they always proceed downhill on the energy surface. The process of minimization determines a nearby low-energy state (local minimum), but in this state, the configuration of the system is usually not in the absolute lowest energy (global minimum).

In this thesis, the steepest descent algorithm is used to minimize the system's energy in the first minimization stage. We will also repeat energy minimization before normal mode analysis using the conjugate gradient (cg) algorithm. Both techniques iteratively minimize potential energy by adjusting atomic coordinates. The steepest descent method reduces energy by moving in the direction of the steepest decrease, but it is slow. The conjugate gradient algorithm is faster and more efficient.

During the minimization stage, the potential energy of the system decreases with each iteration of the algorithm. Figure 2.1 shows the changes in potential energy over time using the steepest descent algorithm. To stop the minimization algorithm, the parameter *emtol* (maximum force tolerance) is defined in the .mdp file. This parameter specifies the maximum force that should be met before the algorithm stops. The typical value for this parameter, suggested in the GROMACS manual, is 10 kJ $\text{mol}^{-1}\text{nm}^{-1}$ [11]. In this thesis, we fixed *emtol* at 0.01 kJ $\text{mol}^{-1}\text{nm}^{-1}$ during the first

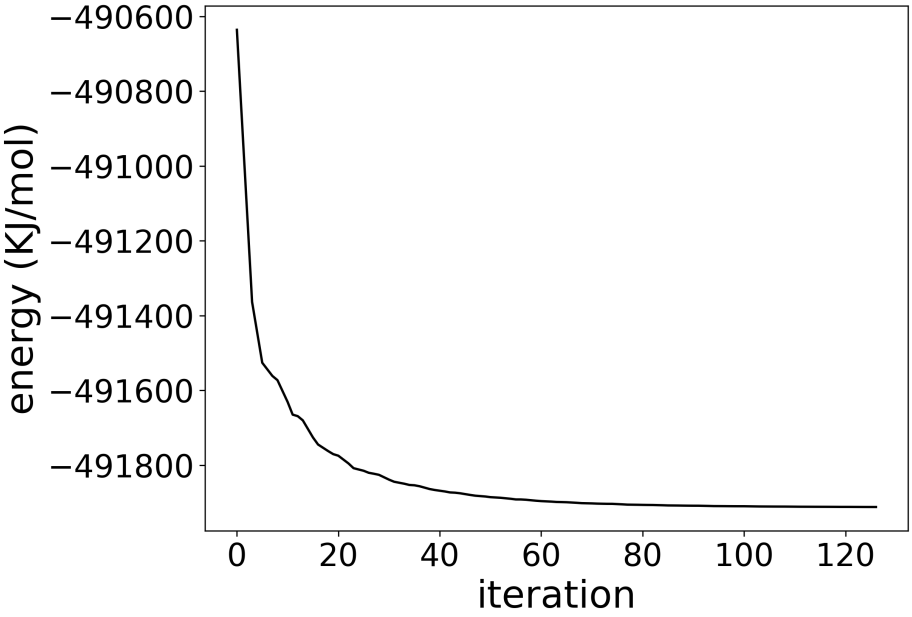


Figure 2.1: The potential energy of the system is minimized using the steepest descent minimization algorithm.

minimization stage and 0.001 kJ $\text{mol}^{-1}\text{nm}^{-1}$ during the minimization stage before normal modes analysis. A tighter tolerance is required for the normal mode analysis since the algorithm assumes that the system is at a local energy minimum, while the initial energy minimization is only used to reduce initial forces.

2.4.3 NVT

After preparing the system at a local potential energy minimum, it is necessary to equilibrate it at the desired temperature and pressure. The equilibration is performed in two separate phases. The first phase is carried out in the NVT ensemble, in which the simulation proceeds at a constant number of particles N, volume V, and temperature T. This stage adjusts the temperature to the desired value and calculates the pressure. In this thesis, the desired temperature is 300 K and is specified in the .mdp file. Different algorithms, known as thermostats, are offered by GROMACS to perform NVT equilibration. Basic thermostats adjust the temperature by rescaling the velocities. The temperature of the system is related to the average kinetic energy through the equipartition theorem [39],

$$\langle K \rangle = \frac{3}{2} N k_B T, \tag{2.17}$$

where k_B is the Boltzmann constant. The kinetic energy is also related to the velocity via,

$$K = \frac{1}{2} \sum_i m_i v_i^2. \tag{2.18}$$

Considering Eqs. 2.17 and 2.18, the velocities of the particles should be modified to keep the temperature of the system constant. This modification can be performed using a rescaling factor λ. This factor for the simplest velocity rescaling algorithm is,

$$\lambda = \sqrt{\frac{T_0}{T(t)}}, \tag{2.19}$$

where T_0 is target temperature, and $T(t)$ is the instantaneous temperature. This thermostat applies sudden velocity adjustments to the system, which can introduce non-physical fluctuations in the system. A gentler variation of velocity rescaling is accomplished by the Berendsen thermostat. Conceptually, this thermostat couples the system to an external bath. The bath is fixed at the desired temperature and acts as a thermal energy source by adding or removing heat from the system. In this algorithm, the scaling factor λ is given by [40],

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau} \left(\frac{T_0}{T\left(t - \frac{\delta t}{2}\right)} - 1 \right)}, \tag{2.20}$$

where δt is the time step and τ is the coupling constant, determining how quickly the thermostat acts. The main advantage of this algorithm is that the τ can be chosen based on the system under study. If $\tau \gg \delta t$, it results in $\lambda = 1$, which means no change in the velocity and no coupling between system and bath. If $\tau = \delta t$, it results in a strong coupling. In this situation, the equilibration happens too fast. Generally speaking, the Berendsen thermostat does not yield the energy fluctuations expected for the NVT ensemble.

Another popular thermostat for adjusting temperature is the Nosé-Hoover thermostat. The idea behind this thermostat is that the system is coupled to an external bath

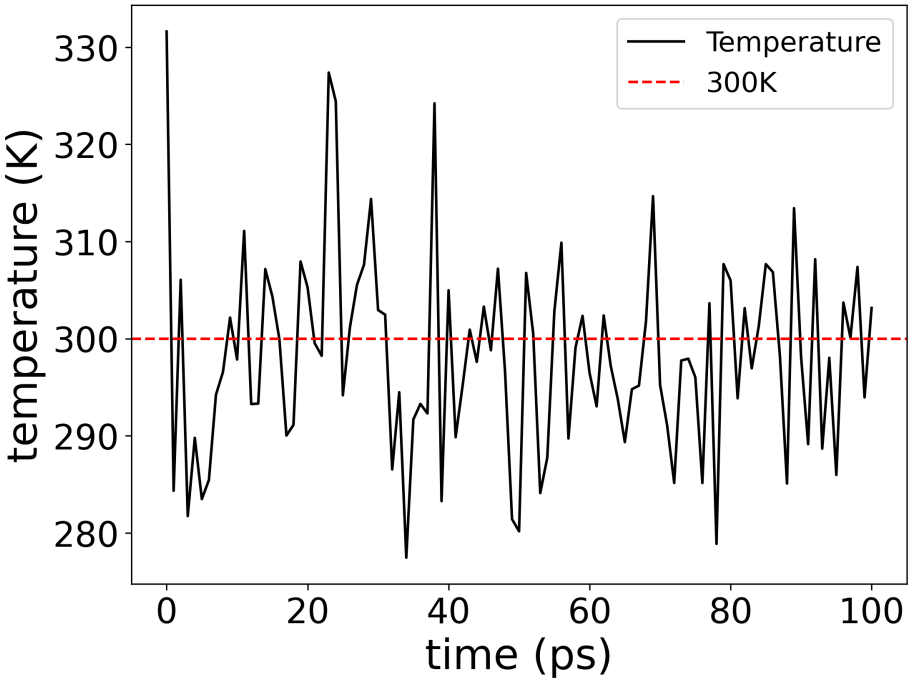


Figure 2.2: Using Nosé-Hoover thermostat, an NVT run is performed for 100 ps to bring the system to the desired temperature of 300 K.

through an additional degree of freedom. This algorithm uses an extended Lagrangian that includes an additional velocity term (technically a friction term) to adjust the temperature by increasing or decreasing the particles' velocities. This thermostat controls the temperature smoothly and is an appropriate option for reproducing the canonical ensemble.

In this thesis, we use the Nosé-Hoover thermostat to adjust the temperature during the NVT stage. Figure 2.2 illustrates the temperature changes over time for a 100 ps run, with a time step of 0.002 ps and a total of 50000 steps. The results indicate that the system's average temperature approaches the desired value of 300 K. At the next phase of equilibration in this thesis, the NPT run, we use a stochastic variant of the velocity rescaling thermostat (v-rescale) that also reproduces canonical fluctuations to maintain the temperature at 300 K.

2.4.4 NPT

The next stage in MD simulations is the NPT equilibration (constant N, P, and T). In this ensemble, the system's pressure is adjusted by changing the system's volume while the temperature is also kept constant. The volume of the simulation box is scaled to achieve the desired pressure. Berendsen (BR) and Parrinello-Rahman (PR) are two common pressure coupling algorithms or barostats provided by GROMACS. The BR barostat is analogous to the BR thermostat. The volume scaling factor is represented by λ [41],

$$\lambda = \left(1 - \frac{k\delta t}{\tau}(P_0 - P(t))\right), \tag{2.21}$$

where P_0 is the desired pressure, $P(t)$ is the instantaneous pressure, k is an estimate of the isothermal compressibility, and τ is the coupling constant. For isotropic scaling, the scaling of the volume is,

$$V_f = V_i \cdot \lambda, \tag{2.22}$$

where the simulations box vectors and all particle coordinates are scaled by the same factor in each of the x, y, and z directions. In this thesis, we use the BR barostat for the first NPT run. This run is 100 ps with the time step 0.002 ps and the number of steps 50000. The desired pressure is set to 1 bar. Figure 2.3 shows the change in pressure over time using the BR barostat. It can be seen that the average pressure approaches the desired value of 1 bar. The BR barostat is not constructed to yield the volume fluctuations expected for the NPT ensemble.

Another barostat that is used to adjust the system's pressure is the Parinello-Rahman barostat. This pressure coupling uses a method similar to the Nosé-Hoover thermostat detailed in Section 2.4.3. In this algorithm, the system is coupled to an external pressure bath. An additional term in the equations of motion is responsible for adjusting the system's pressure. For a non-cubic crystal, isotropic volume scaling will generally result in internal stresses, as quantified by the pressure tensor described below. Both BR and PR allow for non-isotropic scaling of the system in order to achieve a diagonal pressure tensor with equal diagonal terms (isotropic pressure).

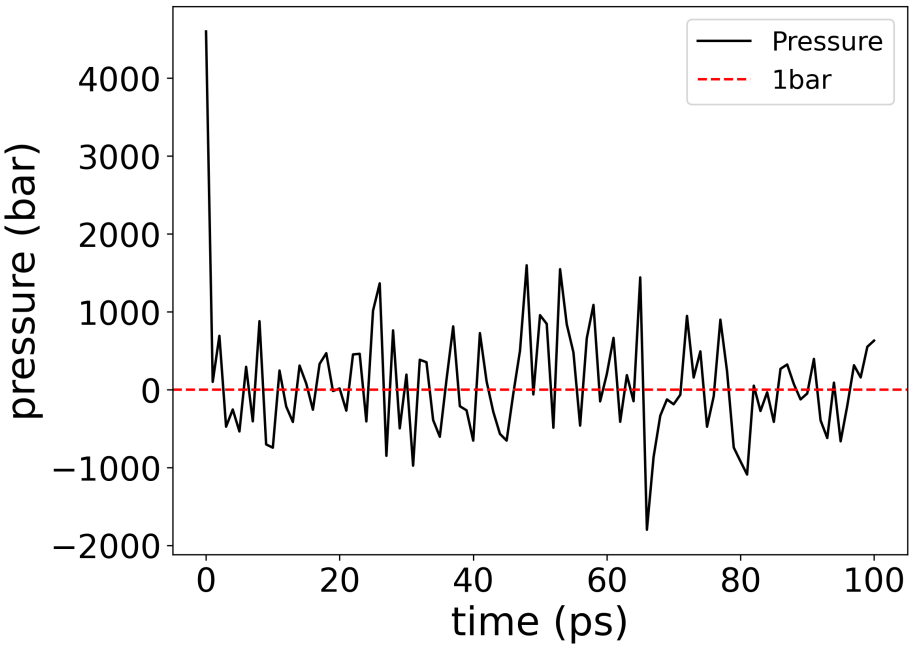


Figure 2.3: Using Berendsen barostat, an NPT run is conducted for 100 ps to adjust the pressure of the system to the desired value of 1 bar.

2.5 Pressure Tensor

One of the quantities we study in this thesis is the pressure tensor. The pressure tensor can be calculated using the virial, which is defined as [11, 41],

$$\boldsymbol{\Xi} = -\frac{1}{2}\sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij}, \tag{2.23}$$

where $\mathbf{r}_{ij}$ is the separation between particles i and j and $\mathbf{F}_{ij}$ is the force between them. The pressure tensor is,

$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{\mathrm{K}} - \boldsymbol{\Xi}), \tag{2.24}$$

where $\mathbf{E}_K$ is the kinetic energy tensor, and V is the volume of the simulation box. Once the system is equilibrated, the pressure tensor elements are reported using the GROMACS `energy` tool. These elements give the pressure exerted by the system in each direction. The tensor is represented by a 3×3 matrix and contains six independent elements: P_{xx}, P_{yy}, P_{zz}, $P_{xy} = P_{yx}$, $P_{xz} = P_{zx}$, and $P_{yz} = P_{zy}$. The

diagonal components P_{xx}, P_{yy}, and P_{zz} indicate the normal pressures in the x, y, and z directions.

Chapter 3

Optimizing the Simulation Protocol

The goal of this chapter is to find an optimized simulation protocol for aragonite before studying the vibrational density of states (VDOS). We examine two recently developed force fields for aragonite, the models introduced by Xiao et al. [17] (No-Bond) and Cruz-Chu et al. [18] (Bond), as explained in Section 2.2. We test different variants of pressure control algorithms, including isotropic, semi-isotropic, and anisotropic, for each model. We also confirm our understanding of the force fields by writing our own GROMACS topology file. Table 3.1 summarizes the simulations carried out in an effort to reproduce results on unit cell volume and shape reported by Xiao et al. [17]. Several other simulation details are then studied to find an optimized set of simulation parameters that can be used for VDOS analysis.

3.1 Models Types

In this thesis, we study two recently developed force field models for aragonite, referred to as the No-Bond and Bond models. The No-Bond model follows the force field developed by Xiao et al. [17]. In this model, there is no stretching bond term between carbon and oxygen in the carbonate units. The Bond model includes explicit harmonic bond terms between carbon and oxygen in the carbonate units [18]. Detailed

System	**Force Field**	**Duration (ps)**	**Compressibility β (bar^{-1})**	**Reference pressure (bar)**
Aniso-no-bond-triclinic	Xiao et al.	100	$\beta_{xx} = 5 \times 10^{-5}$ $\beta_{yy} = 5 \times 10^{-5}$ $\beta_{zz} = 5 \times 10^{-5}$ $\beta_{xy/yx} = 5 \times 10^{-5}$ $\beta_{xz/zx} = 5 \times 10^{-5}$ $\beta_{yz/zy} = 5 \times 10^{-5}$	$p_{xx} = 1$ $p_{yy} = 1$ $p_{zz} = 1$ $p_{xy/yx} = 0$ $p_{xz/zx} = 0$ $p_{yz/zy} = 0$
Aniso-bond-triclinic	Cruz-Chu et al.	100	$\beta_{xx} = 5 \times 10^{-5}$ $\beta_{yy} = 5 \times 10^{-5}$ $\beta_{zz} = 5 \times 10^{-5}$ $\beta_{xy/yx} = 5 \times 10^{-5}$ $\beta_{xz/zx} = 5 \times 10^{-5}$ $\beta_{yz/zy} = 5 \times 10^{-5}$	$p_{xx} = 1$ $p_{yy} = 1$ $p_{zz} = 1$ $p_{xy/yx} = 0$ $p_{xz/zx} = 0$ $p_{yz/zy} = 0$
Aniso-no-bond-orthorhombic	Xiao et al.	100	$\beta_{xx} = 5 \times 10^{-5}$ $\beta_{yy} = 5 \times 10^{-5}$ $\beta_{zz} = 5 \times 10^{-5}$ $\beta_{xy/yx} = \beta_{xz/zx} =$ $= \beta_{yz/zy} = 0$	$p_{xx} = 1$ $p_{yy} = 1$ $p_{zz} = 1$ $p_{xy/yx} = 0$ $p_{xz/zx} = 0$ $p_{yz/zy} = 0$
Aniso-bond-orthorhombic	Cruz-Chu et al.	100	$\beta_{xx} = 5 \times 10^{-5}$ $\beta_{yy} = 5 \times 10^{-5}$ $\beta_{zz} = 5 \times 10^{-5}$ $\beta_{xy/yx} = \beta_{xz/zx} =$ $= \beta_{yz/zy} = 0$	$p_{xx} = 1$ $p_{yy} = 1$ $p_{zz} = 1$ $p_{xy/yx} = 0$ $p_{xz/zx} = 0$ $p_{yz/zy} = 0$
Semiiso-nobond	Xiao et al.	100	$\beta_{x-y} = 5 \times 10^{-5}$ $\beta_z = 5 \times 10^{-5}$	$p_{x-y} = 1$ $p_z = 1$
Semiiso-bond	Cruz-Chu et al.	100	$\beta_{x-y} = 5 \times 10^{-5}$ $\beta_z = 5 \times 10^{-5}$	$p_{x-y} = 1$ $p_z = 1$
Iso-no-bond	Xiao et al.	100	$\beta = 5 \times 10^{-5}$	$p = 1$
Iso-bond	Cruz-Chu et al.	100	$\beta = 5 \times 10^{-5}$	$p = 1$
Iso-no-bond-ourtop	Xiao et al.	100	$\beta = 5 \times 10^{-5}$	p= 1
Iso-bond-ourtop	Cruz-Chu et al.	100	$\beta = 5 \times 10^{-5}$	$p = 1$
Iso-bond-1ns	Cruz-Chu et al.	1000	$\beta = 5 \times 10^{-5}$	$p = 1$
Iso-bond-ourtop-1ns	Cruz-Chu et al.	1000	$\beta = 5 \times 10^{-5}$	$p = 1$

Table 3.1: Simulations carried out to test effects of pressure coupling. Two force fields, No-Bond (Xiao et al. [17]) and Bond (Cruz-Chu et al. [18]), and three pressure algorithms, isotropic, semi-isotropic, and anisotropic, are used to define the simulations. For the anisotropic pressure algorithm, both triclinic and orthorhombic variants are studied. For nine simulations, the topology file as described in [18] is used, and for three simulations, we use a topology file that we generated, labelled as 'ourtop'. The simulation duration for most models is 100 ps, with two models having a duration of 1000 ps. The reference temperature is set to 300 K. The values for compressibility and reference pressure components employed in the pressure coupling are indicated in the table.

descriptions of these two models are provided in Section 2.2. Three pressure algorithm couplings are used for each model: isotropic, semi-isotropic, and anisotropic. The anisotropic pressure coupling is further refined into two sub-cases that reflect crystal structures with distinct symmetries: triclinic and orthorhombic. To study the simulation box (unit cell), we introduce the box vectors. These vectors that define the orthorhombic simulation cell are (b_x, 0, 0), (0, b_y, 0), and (0, 0, b_z). When the triclinic pressure coupling is used, the zero elements can become non-zero. The results that are reported as the box vectors for triclinic are still b_x, b_y, and b_z.

In isotropic pressure coupling, the box vectors are scaled uniformly in the x, y, and z directions. This scaling requires one value for compressibility (β) and one value for reference pressure [11]. In this thesis, these values are set to 5×10^{-5} bar^{-1} and 1 bar, respectively (see the bottom six rows in Table 3.1).

In semi-isotropic coupling, the x and y directions scale uniformly, while the z direction changes independently. This means that the $b_x : b_y$ ratio is fixed, but b_z changes independently. The semi-isotropic coupling uses two values for compressibility and reference pressure: one for the x/y direction and one for the z direction. In this thesis, these values are set to 5×10^{-5} bar^{-1} and 1 bar, respectively. This compressibility value is an appropriate order of magnitude for condensed matter systems.

In anisotropic pressure coupling b_x, b_y, and b_z vary independently. This algorithm requires six values for the compressibility and reference pressure, corresponding to the diagonal and off-diagonal components of the target pressure tensor: xx, yy, zz, xy/yx, xz/zx, and yz/zy. In the orthorhombic version of the algorithm, the off-diagonal components of compressibility and pressure are set to zero. In this case, the off-diagonal elements of the target pressure do not affect the system; they are set to zero only nominally. In this structure, the three axes can have different lengths, but all the angles between them are 90 degrees. To allow for triclinic crystal structure (under hydrostatic pressure), all components of the compressibility should be non-zero (with the same value), and the off-diagonal elements of the target pressure should be set to zero.

Using the Bond and No-Bond models with isotropic, semi-isotropic, and anisotropic pressure coupling, twelve systems are defined. Information about these systems is listed in Table 3.1. We use the supporting information from Cruz-Chu et al. for nine systems to generate the topology file. This information includes the initial files that

GROMACS uses through the `pdb2gmx` command to generate the topology file. For the remaining three systems, specified in Table 3.1, we use "ourtop", the topology file we generated ourselves, as explained in Section 2.3.1. Our topology file is listed in Appendix B.

3.1.1 The Simulated System

All the models listed in Table 3.1 are simulated in a box containing 60 unit cells of aragonite. To construct the simulation box, with all atoms placed in the aragonite structure, we use the supplemental files provided by Cruz-Chu et al. and the program VMD [37]. Xiao et al. (No-Bond model) did not provide the files and scripts to set up the system. Therefore, we use the Cruz-Chu et al. information to create a topology for the No-Bond model. We remove the bond potential from the Cruz-Chu et al. force field and use the `pdb2gmx` to create the topology for the No-Bond model.

To confirm this modification, we provide our own topology based on the No-Bond model, as explained in Section 2.3.1, and use our own scripts to place the atoms. In this way, we can double-check that the modifications required to change the Cruz-Chu et al. topology file so that it yields the Xiao et al. force field were implemented correctly. The results of using our own topology are presented as 'ourtop' in this thesis. To set up the simulation box, the unit cells are placed in a configuration of five in the x direction, three in the y direction, and four in the z direction ($5 \times 3 \times 4$) so that the simulation box is almost cubic. Each aragonite unit cell contains 20 atoms, as shown in Figure 1.1, resulting in 1200 atoms in the simulated system.

3.1.2 Comparing the Box Vectors with Different Models

Xiao et al. [17] used their model to examine several properties of aragonite, including unit cell parameters. They compared their results with those from previous studies and experiments. Xiao et al. reported the cell parameters for the aragonite as $a =$ 4.96 Å, $b = 7.97$ Å, and $c = 5.74$ Å in the x, y, and z directions, respectively. The unit cell parameters are not reported by Cruz-Chu et al. [18]. Based on the simulation box size in this thesis, which contains 60 unit cells ($5 \times 3 \times 4$), the box vectors are calculated to be $5 \times a = 2.48$ nm in the x direction, $3 \times b = 2.39$ nm in the y direction, $4 \times c = 2.29$ nm in the z direction. These values are labelled as the expected values

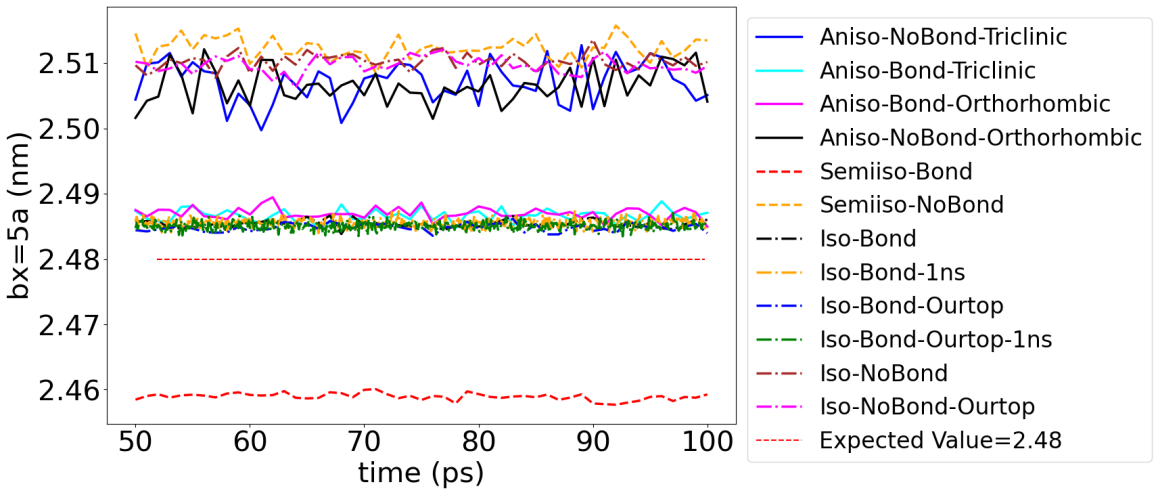


Figure 3.1: Box vectors in the x direction for the simulations introduced in Table 3.1. The No-Bond simulations show higher values, ranging from 2.50 nm to 2.52 nm, while the Bond simulations display smaller box vectors, ranging from 2.48 nm to 2.49 nm. The semi-isotropic Bond model indicates the smallest b_x value, around 2.455 nm. The expected value is based on the results reported by Xiao et al. [17].

in this thesis.

To calculate the box vectors, after the box preparation and energy minimization, two equilibration stages, NVT and NPT, are performed. These two stages aim to bring the system to the desired temperature at 300 K and the desired pressure at 1 bar. The `energy` tool in GROMACS software is a comprehensive resource for analyzing various system properties, such as box vectors, pressure, volume, temperature, and potential energy. After the NPT run, the box vectors are reported using the `energy` tool. The results are then compared with those reported by Xiao et al. [17].

Figure 3.1 shows the results for the box vectors in the x direction for the simulations listed in Table 3.1. The graph displays b_x for the second half of the time series so as to avoid any transients at early times. As shown in Figure 3.1, the results for b_x in the No-Bond simulations range between 2.50 nm and 2.52 nm. When bonds are included, b_x decreases to a range between 2.48 nm and 2.49 nm. The semi-isotropic pressure coupling for the Bond model results in the smallest box vector among all simulations, approximately 2.45 nm. These results indicate that adding bonds between carbon and oxygen atoms leads to a reduction in b_x in all simulations. The results of the isotropic and anisotropic Bond simulations are, oddly, closer to the expected value, which is the value reported by Xiao et al. for the No-Bond model [17]. The

results of the y and z directions are presented in Appendix A, Figures A.1 and A.2.

3.1.3 Comparing Pressure Tensors with Box Vectors

After the NPT equilibration, the pressure tensor elements, as explained in Section 2.5, are reported using the GROMACS `energy` tool. These elements represent, in a sense, the pressure acting on the system in each direction. The tensor is represented by a 3× 3 matrix containing six independent components: P_{xx}, P_{yy}, P_{zz}, $P_{xy} = P_{yx}$, $P_{xz} = P_{zx}$, and $P_{yz} = P_{zy}$. The diagonal components P_{xx}, P_{yy}, and P_{zz} indicate the normal pressures in the x, y, and z directions, respectively.

The left panel in Figure 3.2 shows b_x, and the right panel shows P_{xx} for the various simulations. The results are calculated for the second half of the time series. The error bars represent the standard deviation of the box vectors in each simulation. The red open triangles show the results of the No-Bond runs, and the filled blue circles represent the results of the Bond runs. The dashed black line is the expected value based on the Xiao et al. [17]. In the left plot, adding the bond term results in a reduction in b_x for isotropic and anisotropic couplings. All No-Bond runs show average b_x ranging from 2.50 to 2.52 nm, while the results for Bond runs are between 2.48 to 2.49 nm. This decrease is more significant for the semi-isotropic pressure coupling, where b_x decreases from approximately 2.52 nm (No-Bond) to about 2.46 nm (Bond). The results indicate that adding the bond terms leads to a decrease in the box vectors in all simulations.

The right plot in 3.2 shows the results for P_{xx}. The results indicate that adding the bond term to the force field with an isotropic pressure algorithm results in a significant deviation from 1 bar of pressure (which on the scale of pressures shown is essentially zero), yielding P_{xx} to be approximately -6000 bar. Adding the bond and assuming the same box shape as determined by Xiao et al. results in large stress in the system. For the No-Bond case, the isotropic coupling results confirm that the cell shape determined by Xiao et al. is correct since no stress is apparent, at least in the x direction.

The anisotropic pressure algorithm, by design, allows the simulation box to adjust to stress in the system. For the Bond model, P_{xx} is reduced as expected while b_x remains approximately unchanged despite the fact that the bond lies in the xy plane.

The results show that anisotropic triclinic and anisotropic orthorhombic give the same results. This means that adding the bond preserves the unit cell as orthorhombic as expected.

The model with semi-isotropic pressure coupling behaves differently with the bond term, exhibiting an average pressure in the x direction of approximately 8000 bar. This indicates that the system is under compression, consistent with the lowest b_x value in the left plot. In contrast, the No-Bond models exhibit an average pressure variation close to 1 bar, indicating no significant stress in the system. In this coupling, it is insufficient to allow b_z to vary while preserving the $b_x : b_y$ ratio. Adding the bond, however, changes all three unit cell lengths.

The same graphs are plotted for the y and z directions and presented in Appendix A, Figures A.3 and A.4. Regarding the results we calculated so far, none of the pressure couplings for the No-Bond model reproduced the expected value for b_x. However, the No-Bond simulations confirm the shape of the unit cell, as reported by Xiao et al., showing no stresses even when simulated with the isotropic pressure algorithm. Therefore, for the next part of this thesis, we select the No-Bond model with an isotropic pressure algorithm to study the impact of the simulation details on the box vectors.

3.2 Isotropic No-Bond Model: Detailed Parameters Study

Regarding the discussion in Section 3.1.2, the isotropic No-Bond model is chosen to examine the impact of simulation details on the box vectors. These details were not described in Xiao et al. [17]. We study the effect of different parameters in the .mdp file on the box vectors. The results are compared to Xiao et al. [17]. For each test, only one parameter is changed in the .mdp file during the NPT stage. The following sections detail the parameters and explain the results.

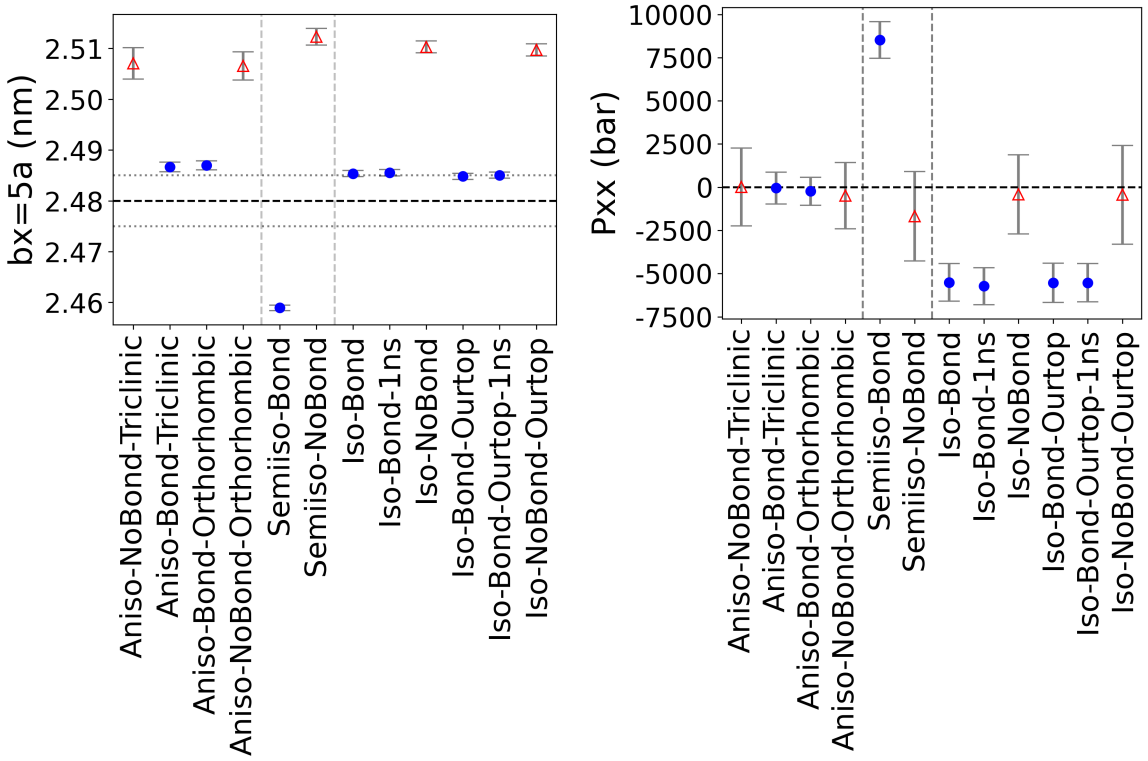


Figure 3.2: The error bars represent the standard deviation of the b_x (left) and pressure P_{xx} (right) for three pressure algorithms and two force fields. Filled blue circles represent models with the bond term in the force field, and the red open triangles show No-Bond models. The graphs indicate that adding bonds to the force field decreases the box size but less stress on the systems with the No-Bond model. The black dashed lines represent the expected values for b_x (reported by Xiao et al. [17]) and pressure (1 bar), and the dotted lines the uncertainty that we infer from Ref. [17]. All of the simulations are conducted at a temperature of 300 K.

3.2.1 Dispersion Correction

For the non-bonded interactions, a cut-off distance is used to reduce the computational cost. The cut-off specifies the distance within which the non-bonded interactions are calculated, and interactions beyond this distance are ignored. Using the cut-off distance introduces errors since the exact result would have an infinite cut-off distance. Dispersion correction (DispCorr) is introduced in the .mdp file to address these errors by applying long-range corrections. GROMACS offers three options for this parameter: EnerPres, Ener, and no correction. To study the impact of this parameter on box vectors, we compare the EnerPres correction with no correction.

The EnerPres option applies correction to both energy and pressure [11]. Figure 3.3 indicates the changes in the b_x due to adding a dispersion correction. The results show that switching from the EnerPres correction to no correction does not significantly change b_x, and the results remain in the same range. The results for the y and z directions are presented in Appendix A, Figure A.5.

3.2.2 Vdw-Modifier

As mentioned in Section 3.2.1, we use a cut-off for the non-bonded interactions (LJ and Coulomb potential). This option introduces discontinuities in the potential energy and force calculations at the cut-off point. To minimize the impact of these discontinuities, GROMACS offers several approaches: potential shift, force switch, potential switch, and none [11]. In the potential shift case, the potential energy function is shifted by a constant so that it goes to zero at the cutoff distance. This should affect neither the pressure calculation nor the box vectors. The force and potential switches are two variations of a technique to smoothly transition both the forces and potentials to zero over a specified range before reaching the cutoff distance. The default setting for this parameter offered by GROMACS is the potential shift.

To study the impact of these options on the box vectors, we change the relevant parameter in the .mdp file during the NPT stage. The results for the box vectors for different vdw-modifier options are shown in Figure 3.3. The force switch and potential shift option show similar results. However, the potential switch shows a small decrease in the box vector. Overall, we can conclude that changing the vdw-modifier parameter does not lead to substantial changes in the box vectors. To calculate the Coulomb interactions, which converge conditionally and very slowly, we use the particle mesh Ewald (PME) technique. In this technique, the calculation of interaction energies between particles is divided into two parts. PME performs a direct sum for the short-range interactions in real space; for the long-range interaction, a Fourier transform is used [11].

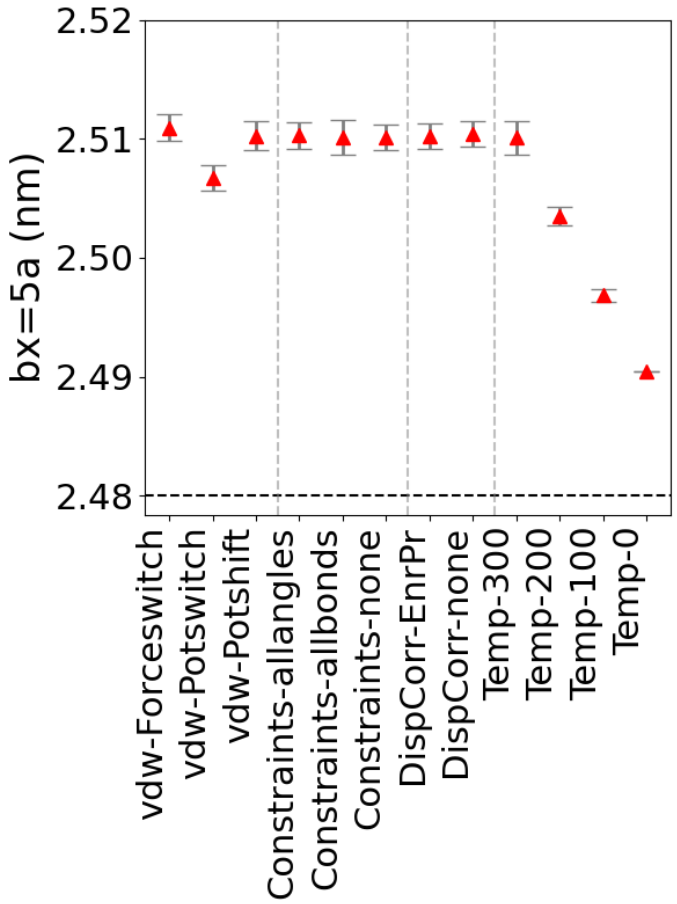


Figure 3.3: The error bars represent the standard deviation of the box vector in the x direction for the isotropic No-Bond model using different input parameters. The results show that the average b_x remains at around 2.510 nm for most of the studied parameters. A significant change in b_x is observed only with a reduction in temperature. Our results indicate that b_x decreases from approximately 2.510 nm at 300 K to 2.487 nm at 0 K. The dashed line indicates the expected value reported by Xiao et al. [17].

3.2.3 Constraints

The constraints in the .mdp file determine how the lengths of certain bonds and the angles between atoms are fixed throughout the simulation. Replacing bonded interactions with constraints eliminates the need to integrate vibrational motion, typically allowing a longer time step to be used. GROMACS provides different options for constraints, including all-bonds (constraints all bonds to be of fixed length), all-angles (all

bonds and angles are converted to constraints), h-bonds (the bonds containing hydrogen atoms are converted to constraints), h-angles (all angles containing the hydrogen atoms and bonds convert to constraints) and none [11]. We study the impact of all-bonds, all-angles, and no constraints on the box vectors. This parameter is changed in the .mdp file during the NPT stage, and the resulting box vectors are calculated. Figure 3.3 shows the b_x variations for the second half of the time series. It can be seen that adding constraints to the system does not affect the box vectors. The all-angles and all-bonds constraints illustrate results similar to those of the no-constraints.

3.2.4 Temperature

During the NVT and NPT stages, the temperature of the system is fixed to the desired value, which is 300 K in this thesis. To study the impact of temperature changes on box vectors, we carry out simulations at 300 K, 200 K, 100 K, and 0 K. The resulting b_x values from the second half of the time series of the NPT run are reported in Figure 3.3. The results indicate that decreasing the temperature leads to a decrease in b_x. This result is consistent with the expectation that lower temperatures result in a smaller volume. At 300 K, b_x is approximately 2.515 nm, which decreases to approximately 2.505 nm at 200 K and further to about 2.495 nm at 100 K. When the temperature is reduced to 0 K, b_x drops to approximately 2.487 nm; this result is the closest to the value reported by Xiao et al. [17] (Expected value = 2.48 nm). However, Xiao et al. did not specify the temperature used in their simulations nor uncertainties in cell parameters, so we strongly suspect that their GULP calculations were carried out at 0 K.

In conclusion, as shown in Figure 3.3, temperature is the only parameter significantly affecting the box vectors. Changes to other parameters in the .mdp file do not significantly change the box vectors. The results for the y and z directions are presented in Appendix 3.3, Figure A.5. We are satisfied that the discrepancy in b_x between our simulations of the No-Bond model and those of Xiao et al. is likely due to temperature.

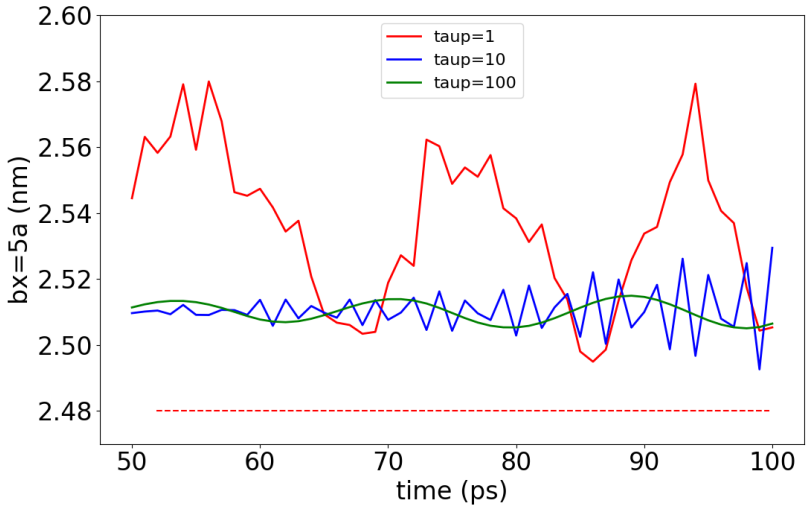


Figure 3.4: Box vector in the x direction (b_x) for the isotropic No-Bond model. The results are plotted for the second half of the time series of the NPT simulation with the Parrinello-Rahman barostat. The resulting box vectors for three tau-p values (1, 10, and 100 ps) are plotted. Increasing tau-p from 1 ps to 10 and 100 ps results in a decrease in the fluctuations in b_x. For tau-p= 10 ps and 100 ps, the fluctuations in b_x are still growing by the end of the simulation.

3.2.5 Tau-p

In MD simulations, tau-p refers to the pressure relaxation time constant within the barostat algorithm; it indicates how long it takes for the algorithm to impose a target pressure on the system. In this section, we study the impact of changing tau-p on the box vectors. We start with the No-Bond model using an isotropic pressure algorithm without constraints with tau-p = 1 ps. During the NPT run, the BR barostat is used to stabilize the system's pressure at 1 bar. After this equilibration, another NPT run is conducted using the final configuration of the first NPT. In the new NPT run, the PR barostat is used to adjust the system's pressure. After the NPT run with the PR barostat, the average box vectors are calculated. The result for the x direction is shown in Figure 3.4.

We calculate the box vectors for tau-p = 1, 10, and 100 ps to study the impact of changing tau-p on the box vectors. Figure 3.4 shows the results for b_x during the second half of the time series. It is observed that increasing the pressure coupling time to 10 ps and 100 ps significantly reduces the fluctuations in b_x. The results for the simulation with tau-p = 10 ps and 100 ps demonstrate an increasing pattern over

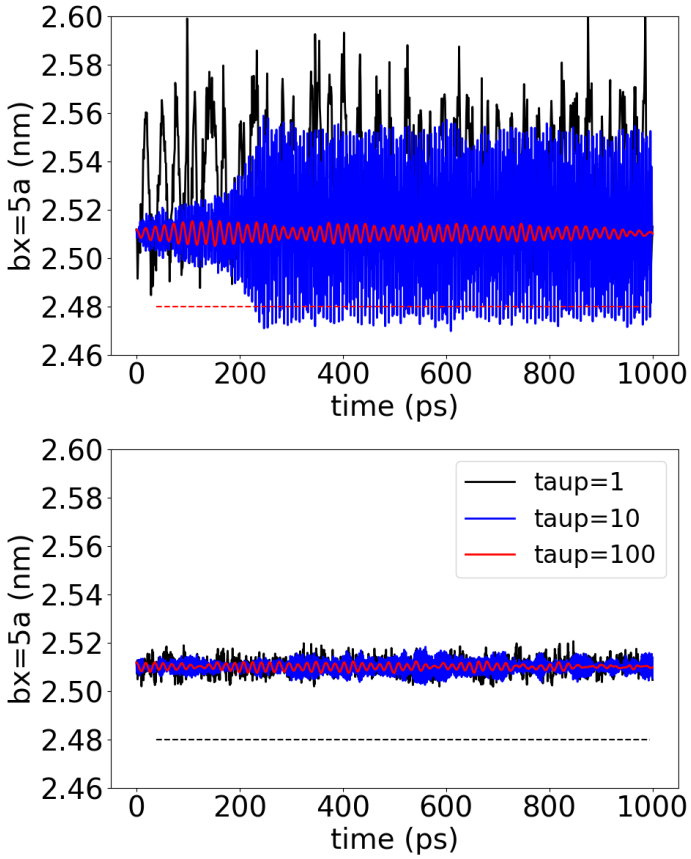


Figure 3.5: Box vectors in the x direction (b_x) for the isotropic No-Bond model over 1000 ps. Equilibration begins with the Berendsen barostat, followed by a plotted NPT run with the Parrinello-Rahman barostat. The resulting box vectors are plotted for tau-p=1, 10, and 100 ps. The top plot corresponds to simulations with parameters dt = 0.002 ps and nstep = 500000, and the bottom plot to parameters dt = 0.001 ps and nstep = 1000000. The results show that decreasing the time step results in smaller fluctuations in b_x.

time. To further analyze this increasing amplitude, we extend the simulation time to 1000 ps (one nanosecond) using two approaches.

In the first approach, the time step is kept at dt = 0.002 ps, with the number of steps at nstep = 500000, extending the simulation time to 1000 ps (1 ns). In the second approach, to extend the simulation time to 1000 ps, the parameters are set to dt = 0.001 ps and nstep = 1000000. We analyze the fluctuations of the box vectors after the second NPT run. Figure 3.5 (top) shows the results for the box vectors using the first approach. The average box size remains around 2.51 nm for tau-p values 10 ps and 100 ps. For tau-p = 1 ps, the average box size is larger, approximately 2.53

nm. The results for tau-p = 10 ps and 100 ps show significant fluctuations.

The results obtained using the second approach are demonstrated in Figure 3.5 (bottom). The average b_x for all three tau-p values is approximately 2.51 nm. A comparison of the two plots reveals that reducing the time step to 0.001 ps leads to significantly smaller fluctuations in the box vectors. The larger time step of 0.002 ps leads to unphysically large fluctuations. Based on these results, we choose the time step of $dt = 0.001$ ps and the tau-p = 10 ps for the optimized simulation protocol.

3.3 Anisotropic Orthorhombic Pressure Coupling

In this section, we study the anisotropic orthorhombic pressure coupling. As explained in Section 3.1, in an anisotropic algorithm, b_x, b_y, and b_z can change independently, which is physically more meaningful for an orthorhombic crystal. After the EM and NVT stages, a 100 ps NPT run using the BR barostat is conducted. We conduct a 1000 ps NPT run using the last configuration of this stage. For the second NPT run, PR and BR barostats are used in two separate simulations. We investigate various parameters related to each barostat and calculate the box vectors. The results are then compared with those reported by Xiao et al. [17].

3.3.1 Parrinello-Rahman Barostat

3.3.1.1 Comparing the System Size Effect

In this section, we study the anisotropic orthorhombic pressure coupling using the PR barostat for two box sizes: 60 unit cells ($5 \times 3 \times 4$) and 480 unit cells ($10 \times 6 \times 8$). After preparing the simulation box using the supplemental files from Cruz-Chu et al. [18], three stages are conducted: EM, NVT, and two NPT runs. During the first NPT, the following parameters are used: barostat = Berendsen, $dt = 0.002$ ps, nstep = 50000, and tau-p = 1 ps, with constraints = all-angles. We conduct a second NPT run for 1000 ps with the PR barostat using the last configuration of the first NPT run. The parameters for this stage are selected based on the results of Section 3.2.5, including $dt = 0.001$ ps, nstep = 1000000, and tau-p = 10 ps. All constraints are removed from the system (constraints = none). b_x, b_y, and b_z are reported using

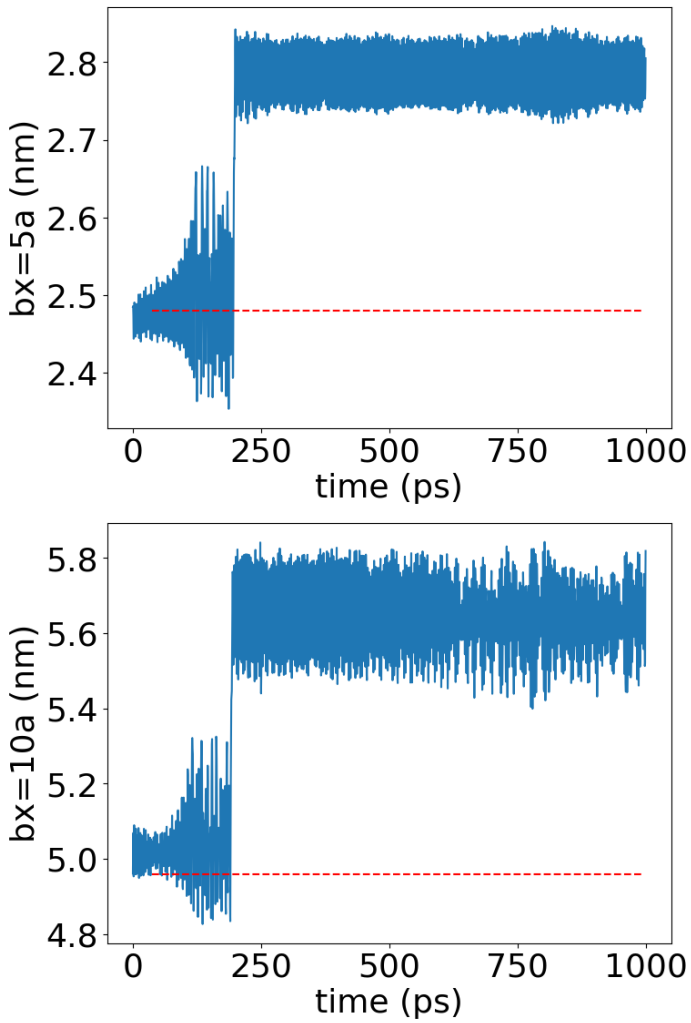
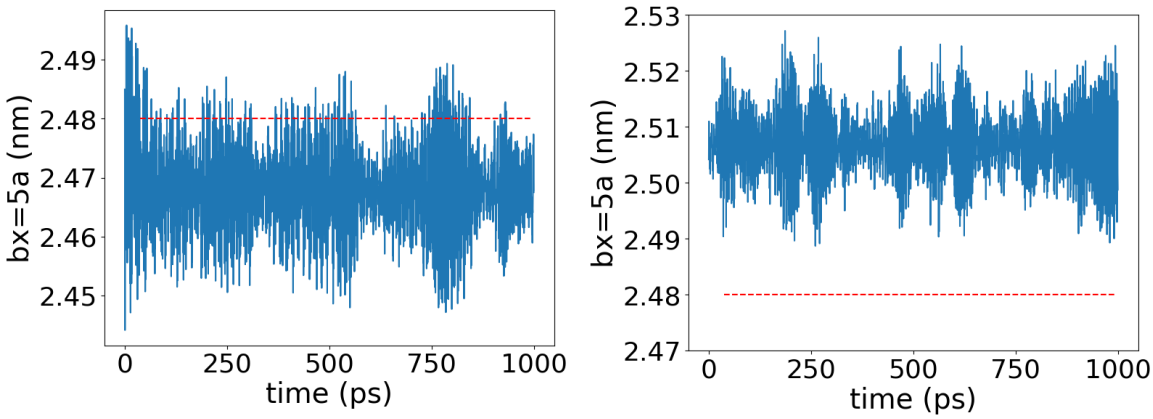


Figure 3.6: The box vector in the x direction for the anisotropic 60 unit cells ($5\times3\times4$) Bond and the 480 unit cells ($10\times6\times8$) No-Bond models. The Berendsen barostat is initially employed to equilibrate the system during an NPT run. Another NPT run is performed with the Parrinello-Rahman barostat using the last configuration of the first NPT run. For this run, parameters are set to tau-p = 10 ps, $dt = 0.001$ ps, and no constraints. The top plot shows the result for the 60 unit cells Bond model. At around 200 ps, a jump in the box vector occurs, indicating a phase transition to another crystal. The bottom plot shows the result for the 480 unit cells in the No-Bond model. The phase transition happens at around 200 ps.

the `energy` tool. Figure 3.6 (top) displays the results for b_x for a 60 unit cell Bond model. From the beginning to around 200 ps, there is an increasing trend in b_x. At around 200 ps, a noticeable jump in b_x occurs, indicating a phase transition. We observe that the crystal structure changes at this time using VMD software.



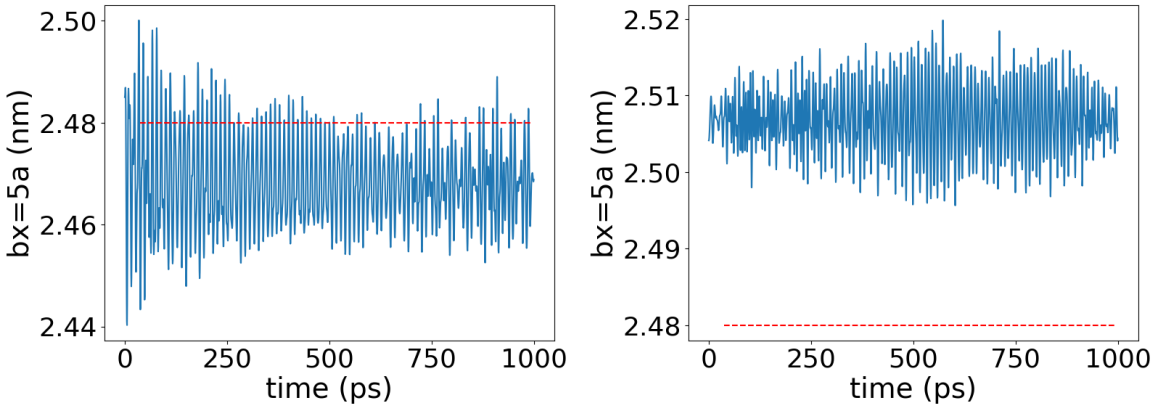
(a) Bond model-taup 1

(b) No-Bond model-taup 1

After observing a phase transition for the 60 unit cells, the number of unit cells is increased in each direction by a factor of two, resulting in a larger box containing 480 unit cells ($10 \times 6 \times 8$). Following a similar procedure with the same input parameters, b_x for this larger box is calculated. A similar phase transition is observed for this larger box. Figure 3.6 (bottom) illustrates the results for b_x of the No-Bond model. Similar to the 60 unit cells, the phase transition occurs at around 200 ps, and the crystal structure changes at this time. Phase transitions are also observed for the y and z directions. These phase transitions are undesirable for the system under study and likely the spurious result of the large box fluctuations that reemerge when the anisotropic pressure coupling is used. Therefore, to stabilize the crystal structure and prevent these phase transitions, we study the simulation parameters that can affect the system's behaviour.

3.3.1.2 Study of Other Input Parameters Affecting the Pressure Algorithm

In this section, we study some GROMACS input parameters and their effect on the phase transition observed in Section 3.3.1.1. The parameters studied are nstlist, nsttcouple, nstpcouple, and tau-t. The nstlist specifies the number of steps for updating the neighbour list, which is a list of the particles whose interactions are considered in the simulation. The number of steps to update pressure and temperature is called nstpcouple and nsttcouple, respectively. The time constant for coupling the temperature is called tau-t [11].



(c) Bond model- taup 10 (d) No-Bond model- taup 10

Figure 3.7: b_x for anisotropic orthorhombic pressure coupling. The Berendsen barostat is initially employed for system equilibration during an NPT stage. Another NPT run, shown, is conducted with the Parrinello-Rahman barostat using the last configuration of the first NPT run. The input parameters for this stage are nstlist = nstpcouple = nsttcouple = 1, tau-t = 1 ps. The box vectors are plotted for two tau-p values: 1 and 10 ps. Panels (a) and (b) display the results for the Bond and No-Bond models with tau-p = 1 ps, and panels (c) and (d) indicate the results for tau-p = 10 ps. Compared to Figure 3.6, no phase transition occurs in the crystal structure. The red lines indicate expected values reported by Xiao et al. [17].

The default values recommended for these parameters can be found in the GROMACS manual. In this thesis, we use GROMACS version 2020.1, with the recommended values of nstlist = nstpcouple = nsttcouple = 10 [11]. For the first NPT stage, we use these recommended values and tau-t = 0.4 ps with the BR algorithm. To study the impact of these parameters on the box vectors, we change the values of these parameters in the second NPT stage, for which we use the PR barostat. During the second NPT run, these parameters are set to nstlist = nstpcouple = nsttcouple = 1, and the tau-t = 1 ps.

The results for b_x for the Bond and No-Bond models are plotted for tau-p = 1 and 10 ps. In Figure 3.7, panels (a) and (b) show the results for tau-p = 1 ps, and panels (c) and (d) illustrate the results for tau-p = 10 ps. The results indicate that with the change in the mentioned parameters, there is no phase transition in the crystal structure. Thus, for our system, the default of updating P, T, and neighbour lists every 10 steps is insufficient. The average box vector for the Bond models is approximately 2.465 nm. For the No-Bond models, the average b_x is around 2.505

nm. Although there is no phase transition in the new results, significant fluctuations are still evident in the box vector. To investigate these fluctuations in the next section, we study the compressibility of the system.

3.3.1.3 Compressibility Consistency

To study the volume fluctuations, we examine the compressibility of the system. Isothermal compressibility, β_T, is a quantity that describes how the volume changes with pressure at a constant temperature. It is defined as,

$$\beta_T = \frac{-1}{V}\left(\frac{\partial V}{\partial P}\right)_T. \tag{3.1}$$

If NPT simulations are carried out at several equally spaced pressures, we can approximate the derivative as,

$$\frac{\partial V}{\partial P} = \frac{V_{i+1} - V_{i-1}}{2\Delta P}. \tag{3.2}$$

The compressibility can also be calculated from the volume fluctuations as [42],

$$\beta_T = \frac{1}{k_B T}\frac{\langle \Delta V^2 \rangle_{NPT}}{\langle V \rangle_{NPT}}, \tag{3.3}$$

where k_B is Boltzmann's constant, $\langle \Delta V^2 \rangle_{NPT}$ is the variance of the volume,

$$\langle \Delta V^2 \rangle = \langle V^2 \rangle - \langle V \rangle^2, \tag{3.4}$$

and $\langle V \rangle_{NPT}$ represents the average volume of the system. In this thesis, we calculate the compressibility using two equations, 3.1 and 3.3, and we refer to the results as the derivative and variance methods, respectively. Another way of approximating the derivative is to fit the $V(P)$ data to a quadratic polynomial,

$$V_{fit} = V_0 + V_1 P + V_2 P^2. \tag{3.5}$$

Then, the system's compressibility is calculated using the derivative,

$$\frac{\partial V_i}{\partial P} = V_1 + 2V_2 P_i. \tag{3.6}$$

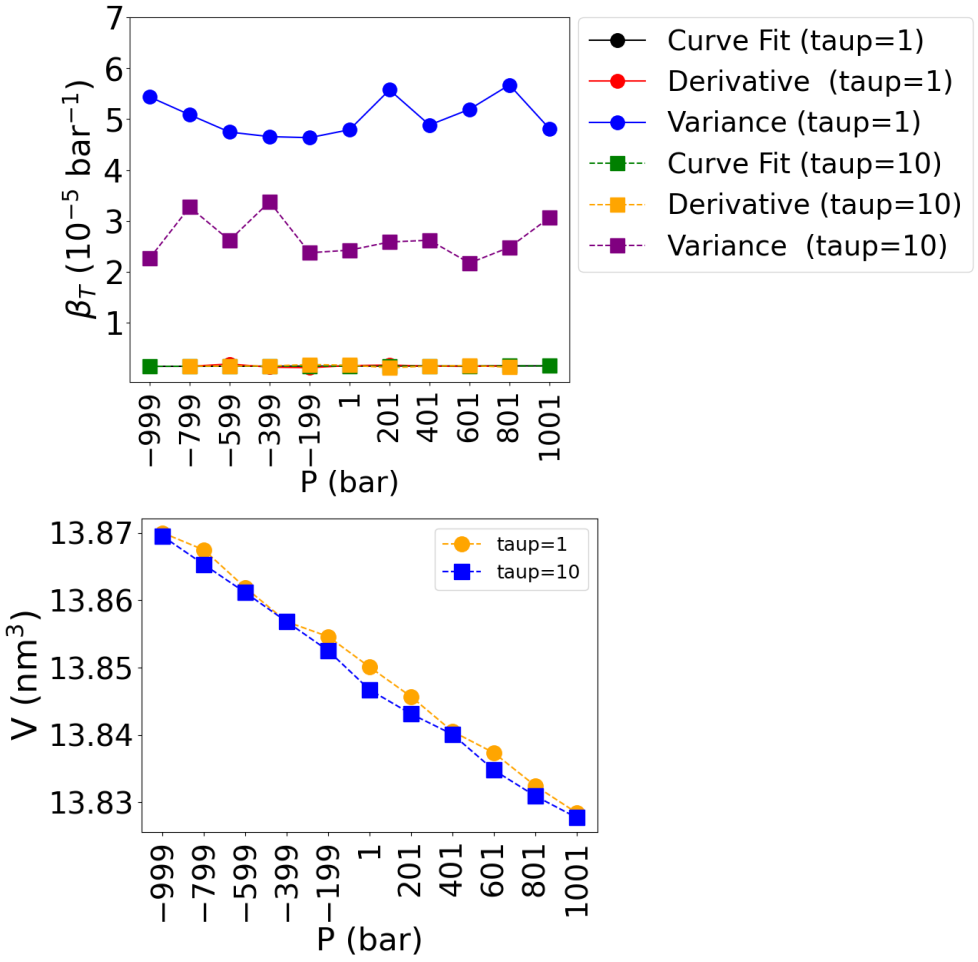


Figure 3.8: The compressibility-pressure (top) and volume-pressure (bottom) relationships for the anisotropic orthorhombic Bond model using the Parrinello-Rahman barostat with parameters nstlist = nstpcouple = nsttcouple = 1, and tau-t = 1 ps. The results are plotted for two pressure time constants: tau-p = 1 and 10 ps. Compressibility is calculated using derivative, variance, and curve-fitting methods. The compressibility calculated from the variance method is an order of magnitude larger than the results from the derivative equation, both for tau-p = 10 ps and tau-p = 1 ps. On this scale, the derivative and curve-fitting methods overlap for both values of tau-p. This indicates unphysically large volume fluctuations in the system. The volume-pressure plot shows a decreasing trend, as expected. The average volume for tau-p = 10 ps is slightly larger than that from tau-p = 1 ps.

We refer to β_T obtained in this way as the "curve-fitting" method. After NPT equilibration with the Berendsen barostat for 100 ps, the second NPT run is performed for the models described in Section 3.3.1.2. A series of eleven distinct NPT simulations for each case are conducted to study the volume variations and to compute

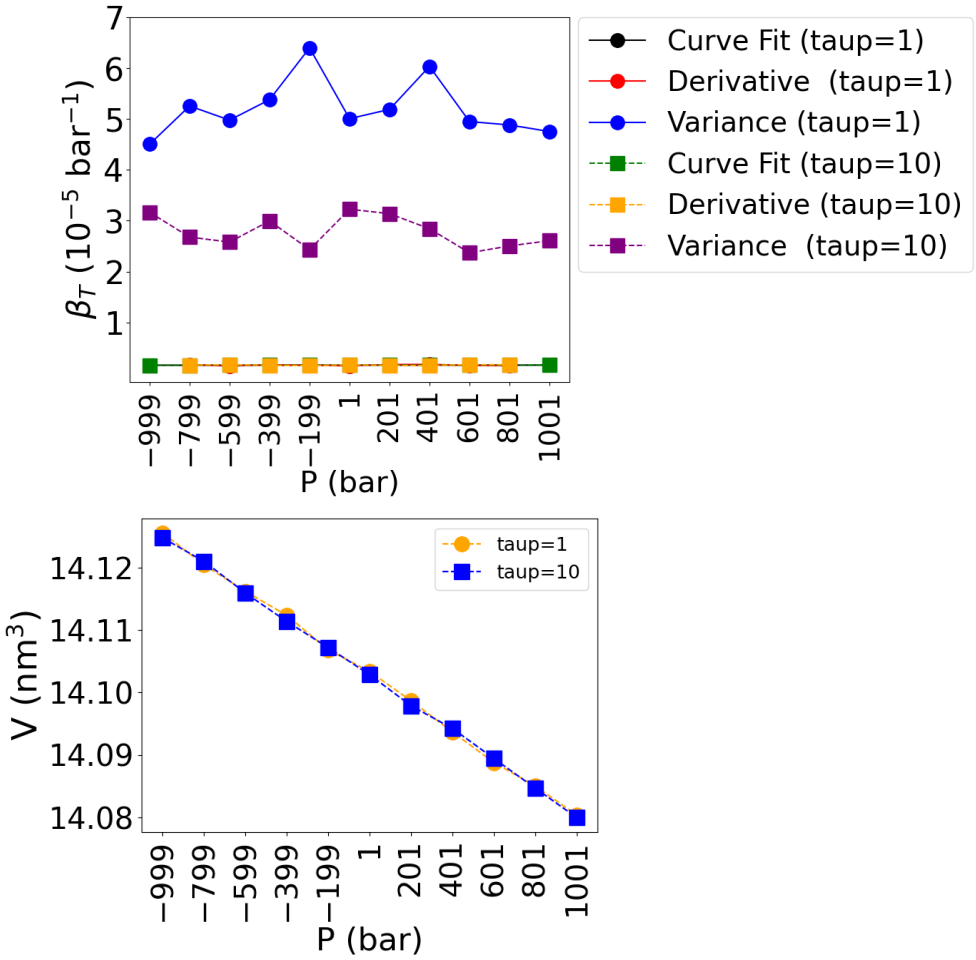


Figure 3.9: The compressibility-pressure (top) and volume-pressure (bottom) plots for the anisotropic orthorhombic No-Bond model using Parrinello-Rahman barostat with parameters nstlist = nstpcouple = nsttcouple = 1, and tau-t = 1 ps. The results are plotted for two pressure time constants: tau-p = 1 and 10 ps. Compressibility is calculated using derivative, variance, and curve-fitting methods. The compressibility calculated from the variance method is larger than the results of the derivative method by an order of magnitude. This indicates unphysically large volume fluctuations in the system. The volume-pressure plot shows a decreasing trend, with the average volume remaining consistent for both tau-p values.

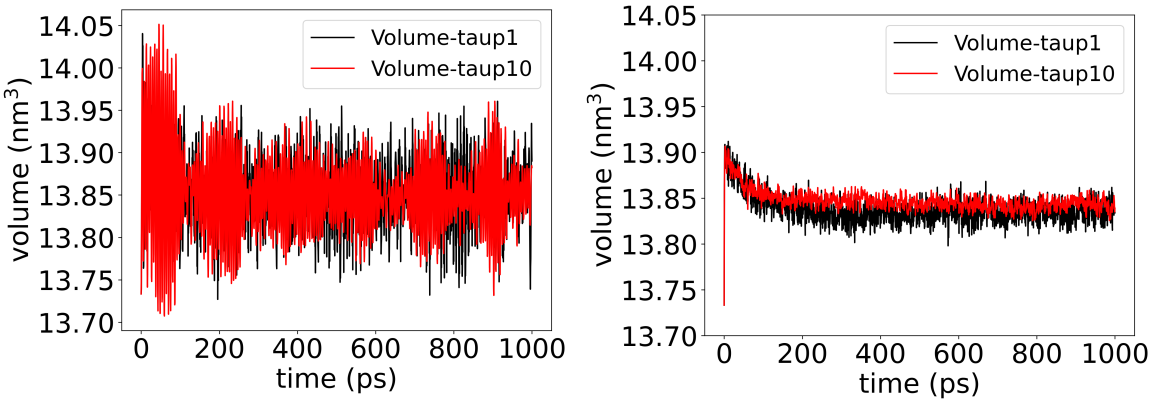
compressibility. These simulations cover a range of target pressures from -999 to 1001 bar, including -999, -799, -599, -399, -199, 1, 201, 401, 601, 801, and 1001 bar. The negative pressures in this range correspond to isotropic tension. The volume time series of each simulation is obtained using the `energy` tool in GROMACS. The system's compressibility is calculated using derivative, variance, and curve-fitting methods for

the second half of the time series. For the Bond and No-Bond models, we calculate compressibility for two tau-p values, 1 and 10 ps.

Figure 3.8 shows the compressibility-pressure (top) and volume-pressure (bottom) results for the Bond model. The compressibility-pressure plots indicate that the compressibility values obtained from the variance method are significantly higher than those from the derivative method. Specifically, for tau-p = 10 ps, the compressibility values from the variance method average around 2.5×10^{-5} bar^{-1}, 10 times larger than the average compressibility of 0.25×10^{-5} bar^{-1} obtained from the derivative method. Similarly, for tau-p = 1 ps, the compressibility values from the variance method average around 5.5×10^{-5} bar^{-1}, 20 times larger than from the derivative method.

Similar results are calculated for the No-Bond model. We calculate the compressibility (top) and volume (bottom) versus pressure for two tau-p values in Figure 3.9, using the same target pressures. The compressibility values from the derivative method average around 0.25×10^{-5} bar^{-1} for both tau-p values. For tau-p = 10 ps, the average compressibility calculated from the variance method is approximately 3×10^{-5} bar^{-1}, 10 times larger than the derivative method. For tau-p = 1 ps, the compressibility values from the variance method average around 5×10^{-5} bar^{-1}, 20 times larger than those obtained from the derivative method. These results indicate the volume fluctuations are too large, and they are not consistent with β_T obtained from the thermodynamic definition of β_T.

The volume-pressure plots for both tau-p values exhibit similar decreasing trends in Bond and No-Bond models. In the Bond model, Figure 3.8, the average volume for tau-p = 10 ps is slightly larger than the tau-p = 1 ps for each target pressure. In the No-Bond model, illustrated in Figure 3.9, the volumes for two tau-p values are mostly the same. The overall average volume in the No-Bond model shows larger magnitudes compared to the Bond model. Both models demonstrate that increasing the pressure results in a decrease in volume, which is expected. However, the study of compressibility demonstrates large fluctuations in volume. Strange oscillations are observed when we visualize the system with VMD using the PR barostat. To further study, we plot the volume over time for the simulated system.



(a) PR -Bond model

(b) BR -Bond model

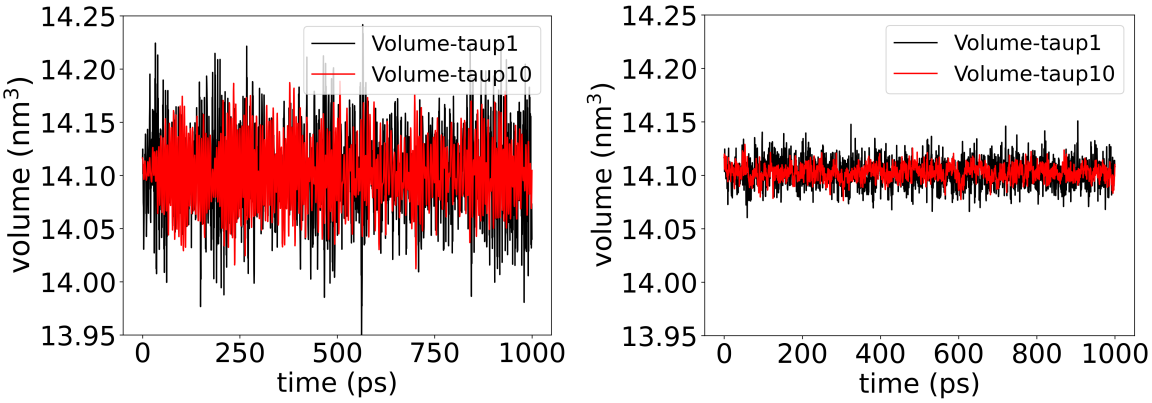
3.3.1.4 Volume Time Series

In this section, we study the volume fluctuations over time for a target pressure of 1 bar to check the unphysical artifacts using the PR barostat. The results are calculated for two tau-p values: 1 ps and 10 ps. In Figure 3.10, panels (a) and (c) display the volume time series results for the Bond and No-Bond models introduced in Section 3.3.1.2. Looking at the two models, we observe similar volume changes over time for both tau-p values. In the Bond model, fluctuations are larger initially but decrease around 100 ps. The average volume for this system is approximately 13.85 nm^3. The No-Bond model shows consistent fluctuations throughout the simulation, with an average volume of approximately 14.10 nm^3. These results are consistent with the volume-pressure plot in Figure 3.8, illustrating similar average volumes at a pressure of 1 bar for both models. These plots show relatively long-period oscillations with the PR barostat, particularly visible in Figure 3.10, which is not desirable.

3.3.2 Berendsen Barostat

3.3.2.1 Volume fluctuations

After observing the large volume fluctuations in the anisotropic orthorhombic model using the PR barostat, we study the fluctuations when using the BR barostat. The parameters for the initial 100 ps NPT run are as follows: $dt = 0.002$ ps, constraints = all-angles, nstep = 500000, tau-p = 1 ps, and barostat = Berendsen. The final configuration from this run is used to conduct a second NPT run for one nanosecond.



(c) PR No-Bond model

(d) BR No-Bond model

Figure 3.10: The volume-time plots for the anisotropic orthorhombic system using Parrinello-Rahman (PR) and Berendsen (BR) barostats. For the simulation with PR barostat, the parameters are set to nstlist = nstpcouple = nsttcouple = 1, and tau-t = 1 ps. The results are plotted for two taup values, 1 and 10 ps, in a 1000 ps NPT run. Panels (a) and (c) show the results for the Bond models, and panels (b) and (d) illustrate the results for the No-Bond models. The average volumes for tau-p = 1 and 10 ps are similar in each plot. In the bond models using both barostats, the volume averages around 13.85 nm^3, while for the No-Bond models, it is approximately 14.10 nm^3. Significant volume fluctuations are evident when the barostat is set to Parrinello-Rahman, compared with the Berendsen barostat.

The parameters for this stage are dt = 0.001 ps, nstep = 1000000, tau-p = 1 ps, constraints = none, and barostat = Berendsen. Figure 3.10 shows the results for V for the Bond model, panel (b), and the No-Bond model, panel (d), with good convergence to average values in the second half of the time series. Comparing with the PR results in Fig 3.10 (a) and Figure 3.10 (c) shows that using the BR barostat significantly reduces the fluctuations in the volume with fluctuations decreasing with increasing tau-p. In the next section, we find the value of tau-p yields consistency between the derivative and variance routes to compressibility.

3.3.2.2 Compressibility Consistency

Following NPT equilibration with the BR barostat for 100 ps, a series of eleven distinct NPT runs are conducted. These simulations cover a range of target pressures from -999 to 1001 bar, including -999, -799, -599, -399, -199, 1, 201, 401, 601, 801, and 1001 bar, to study the volume variations and calculate the compressibility. The

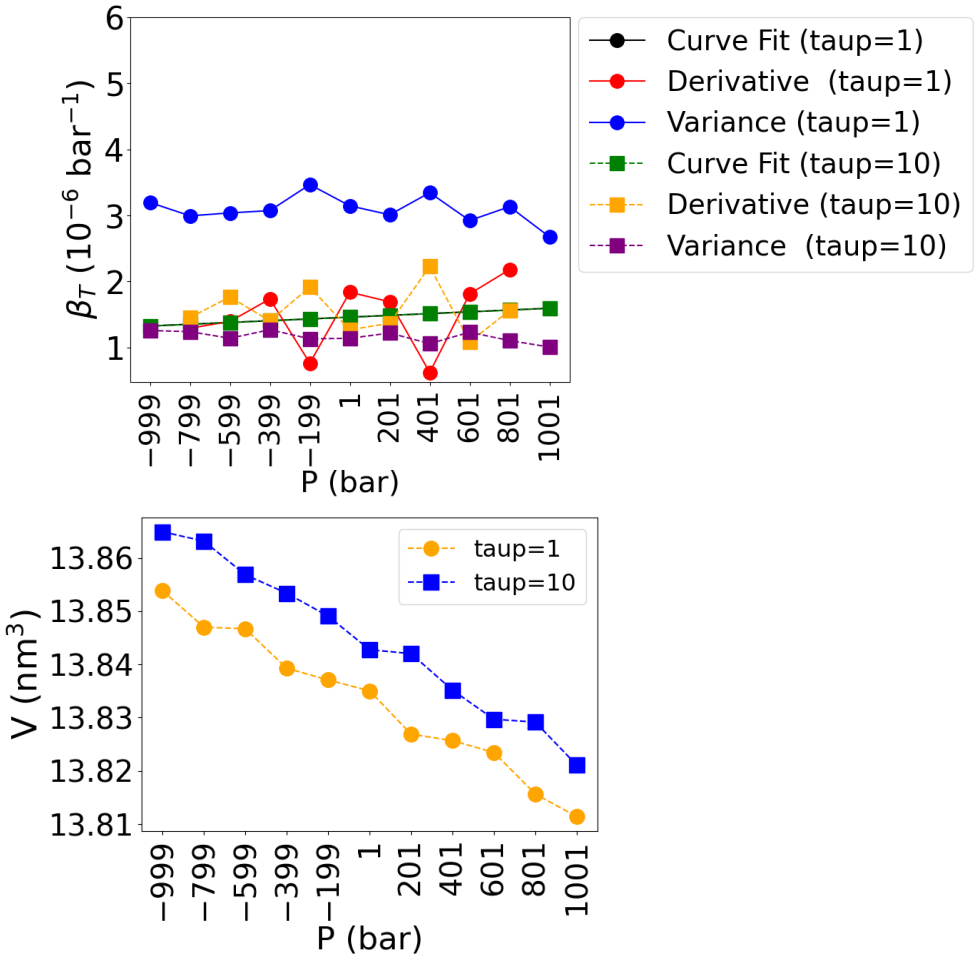


Figure 3.11: Compressibility-pressure (top) and volume-pressure (bottom) results for the anisotropic orthorhombic Bond model using the Berendsen barostat for two pressure time constants: tau-p = 1 and 10 ps. Compressibility is calculated using the variance, derivative, and curve-fitting methods. For tau-p = 10 ps, the results obtained from the three methods are similar. However, using tau-p = 1 ps produces a larger compressibility than with the variance. The volume-pressure plots show a decreasing trend for both tau-p values. A larger average volume is obtained for tau-p = 10 ps.

following parameters are used for these simulations: $dt = 0.001$ ps, nstep = 1000000. For this run, the barostat is set to Berendsen, and no constraints are applied (constraints = none), as detailed in Section 3.3.2. The compressibility of the system is calculated for the second half of the time series using the derivative, variance, and curve-fitting methods detailed in Section 3.3.1.3.

These stages are calculated for two values of tau-p: 1 ps and 10 ps. Figure 3.11 displays the plots of compressibility-pressure (top) and volume-pressure (bottom) for

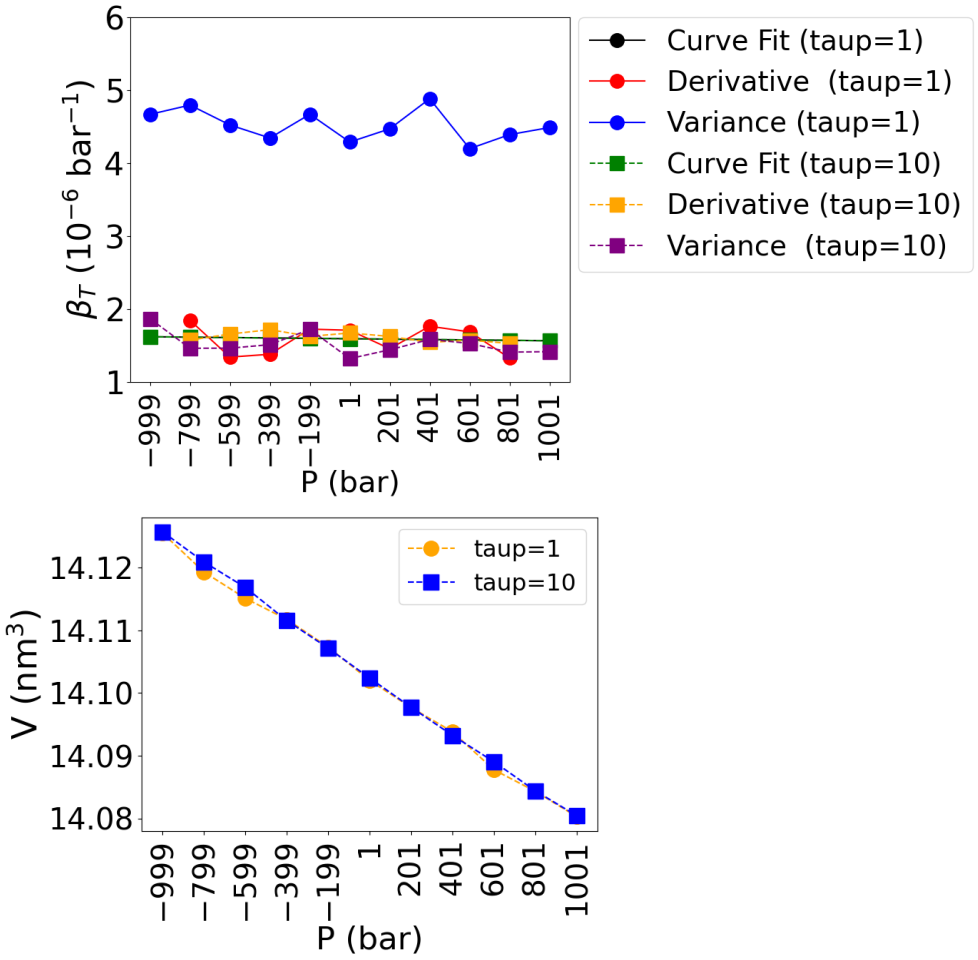


Figure 3.12: Compressibility-pressure (top) and volume-pressure (bottom) results for the anisotropic orthorhombic No-Bond model using the Berendsen barostat for two pressure time constants: tau-p = 1 and 10 ps. Compressibility is calculated using the variance, derivative, and curve-fitting methods. For tau-p = 10 ps, the results obtained from the three methods show similar results. However, using tau-p = 1 ps produces a larger compressibility than with the variance. The average volumes for both tau-p values are mostly consistent. The results demonstrate that as pressure increases, the volume decreases.

the Bond model. In the compressibility-pressure plot, the results of the derivative, variance, and curve-fitting methods exhibit similar values for tau-p = 10 ps, all averaging around 1.5×10^{-6} bar^{-1}. However, for tau-p = 1 ps, the variance method gives a larger compressibility, averaging approximately 3.2×10^{-6} bar^{-1}. Similar results are observed in the No-Bond model presented in Figure 3.12. For the tau-p = 10 ps,

the derivative, variance, and curve-fitting methods result in an average compressibility around 1.5×10^{-6} bar^{-1}. However, for tau-p = 1 ps, the results of the variance method indicate an average compressibility around 4.5×10^{-6} bar^{-1}. Comparing the volume-pressure plots shows that, for the Bond model at tau-p = 10 ps, the average volume is slightly larger than at tau-p = 1 ps, as observed for the PR algorithm. The No-Bond model indicates very similar results for both tau-p values. The overall average volume in the No-Bond model is larger than that in the Bond model.

It is concluded that using the Berendsen barostat with tau-p = 10 ps yields consistency between β_T calculated from fluctuations and that obtained from the derivative of $V(P)$. We then use this barostat and this value of coupling time for simulations that generate configurations for normal mode analysis in the next chapter.

Chapter 4

Normal Mode Analysis

4.1 VDOS Using the Anisotropic Orthorhombic Berendsen Barostat

In this section, the vibrational density of states (VDOS) is calculated through the normal mode analysis. We calculate VDOS for Bond and No-Bond models employing the BR barostat with anisotropic orthorhombic pressure coupling as explained in Section 3.3.2. The vibrational modes of the models are visualized, and the results are compared with those reported by Carteret et al. [29]. They have used a combination of experimental techniques, including IR and Raman spectroscopy and theoretical quantum-mechanical calculations, to study the vibrational spectrum of aragonite. They have studied the normal modes of aragonite, and their results are presented as graphical animations on the Crystal website [43].

Other researchers have used vibrational analysis to study carbonates for various purposes. Zhuravlev and Atuchin [44] studied the normal vibrations of various carbonate compounds using density functional theory (DFT). They compared the IR and Raman spectra of these minerals with experimental results. Their analysis showed that the intensities and wavenumbers of individual modes are consistent between theory and experiment. La Pierre et al. [45] interpreted the Raman spectra of powder and single crystals of calcite and aragonite using DFT. They compared the calculated spectra with experimental measurements, which showed satisfactory agreement. The simulations indicated differences between these two minerals in the powder spectra.

Matas et al. [46] employed VDOS to study the thermodynamic properties of five different carbonates. From the spectroscopic models, the thermodynamic properties were calculated as a function of temperature and pressure. The results were compared with available measurements to study the impact of pressure and temperature on the vibrational modes, thermodynamic behaviour, and phase stability.

4.1.1 Normal Mode Analysis for the No-Bond Model

After completing the second NPT run for the No-Bond model using the Berendsen barostat for 1000 ps (1 ns), as explained in Section 3.3.2, 100 configurations are extracted from the NPT run trajectory at 10 ps intervals. This process aims to average over the different volumes sampled by the system. It is expected that VDOS would be the same for two configurations that have the same box vectors. Before starting the normal mode analysis, another energy minimization is performed so that the system is at a minimum in potential energy, as required for calculating the Hessian matrix. For this EM run, we use the conjugate gradient (cg) integrator, which is generally faster than the steepest descent. During this stage, each configuration is quenched to the minimum potential energy with a tolerance factor, *emtol*, of 0.001 kJ mol^{-1} nm $^{-1}$. There are no bond constraints used in this stage or in subsequent stages.

Normal modes of each configuration can be calculated using GROMACS normal mode analysis. Initially, to calculate the Hessian matrix H (explained in Section 1.2), the nm integrator is used. Normal mode analysis involves diagonalizing the Hessian matrix. Using the `nmeig` tool, eigenvectors and eigenvalues are calculated by diagonalizing H. During normal mode analysis, GROMACS sorts eigenvectors in order of increasing eigenvalues. The eigenvalues obtained from normal mode analysis represent the squared frequencies of the normal modes, $\omega^2 = \lambda$, as explained in Section 1.2. In this equation, ω is the angular frequency and related to frequency ν through $\omega = 2\pi\nu$. For ease of comparison with Raman and IR spectroscopy, GROMACS reports the frequency of the normal modes in terms of the spectroscopic absorption wavenumber $k = \nu/c$ reported in units of cm^{-1} [47, 48], and which is also proportional to the energy.

The `xmgrace` tool is employed to visualize the distribution of normal mode wavenumbers. Figure 4.1 (top) shows the normal mode wavenumber plotted against the normal

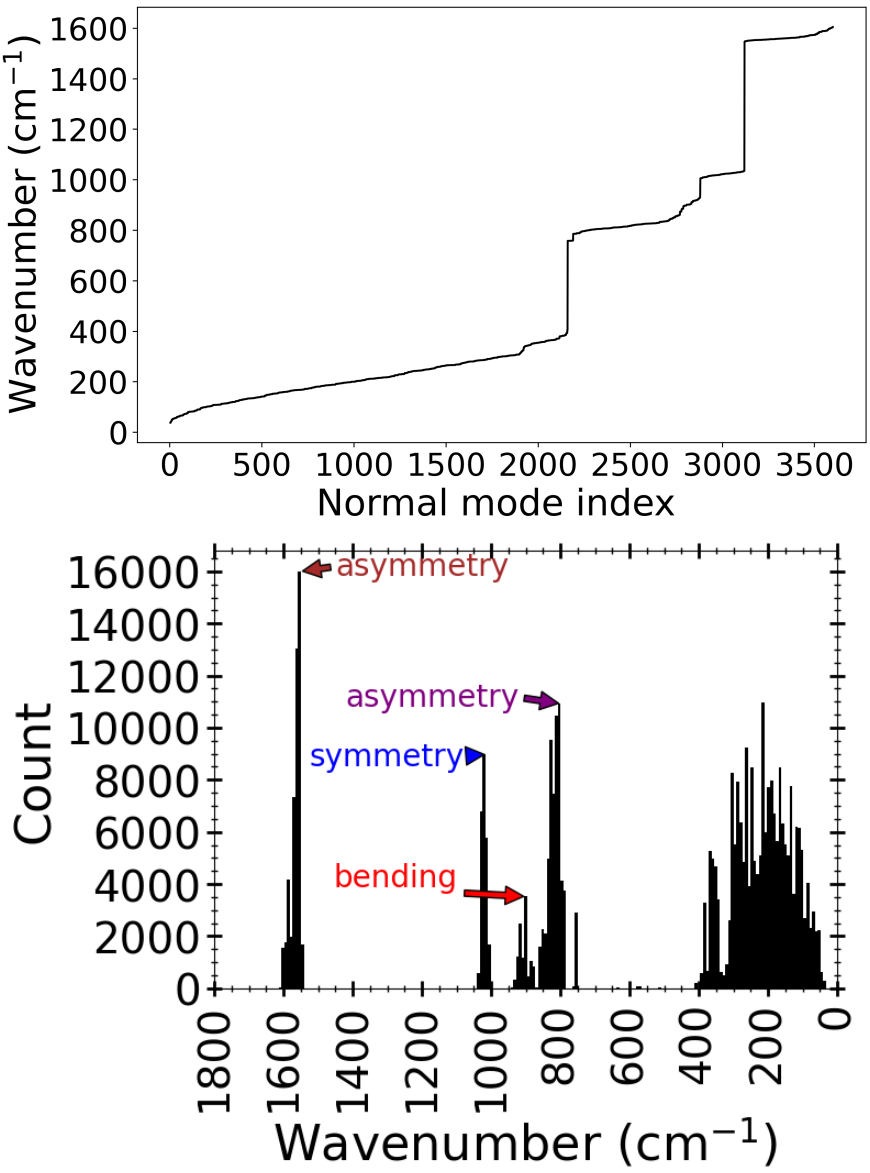
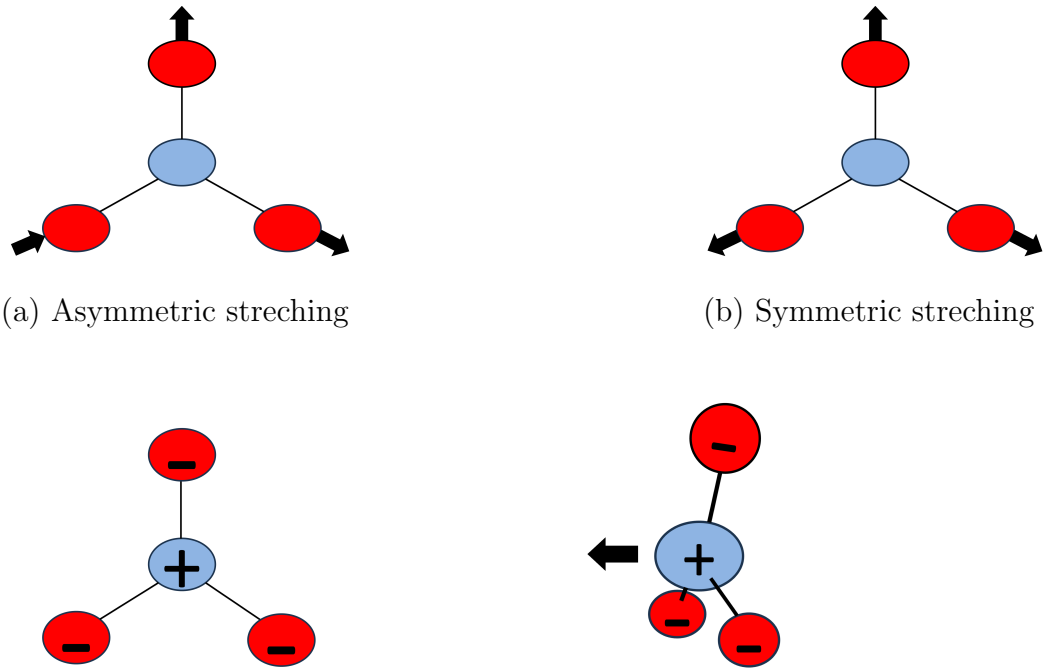


Figure 4.1: Normal mode wavenumber distributions for the anisotropic orthorhombic No-Bond model using the Berendsen barostat. In the top panel, wavenumbers are plotted in increasing order for a single configuration. A histogram of the vibrational density of states (VDOS) was generated from data for 100 sampled configurations (bottom panel). The first and the second prominent peaks are located at 1560 cm^{-1} and 821 cm^{-1}, corresponding to asymmetric stretching modes. The peak at 1017 cm^{-1} corresponds to the symmetric stretching mode, and the peak at 941 cm^{-1} corresponds to a bending vibrational mode.

mode index for one configuration. In this plot, regions with low slopes, such as for indexes between 2300 and 2800, indicate many vibrational modes that are close in



(a) Asymmetric streching

(b) Symmetric streching

(c) Bending

Figure 4.2: Asymmetric (a), symmetric (b), and bending (c) normal modes of a carbonate unit. In the asymmetric stretching mode, the oxygen atoms move towards and away from the carbon atom in an out-of-phase manner. However, in the symmetric stretching mode, all oxygen atoms move towards and away from the carbon atom in phase. In the bending vibration mode, the carbon atom moves perpendicularly to the plane of the oxygen atoms, as shown in two views.

energy. Conversely, regions with high slopes, such as around index 2200, indicate few or no vibrational modes.

To better study the distribution of normal modes, we bin them in a histogram and thus create a VDOS plot, Figure 4.1 (bottom). The `nmtraj` tool in GROMACS is then used to generate trajectories from these eigenvectors, showing how atoms oscillate around their equilibrium positions in different modes. The result is saved in the .gro file, which can be viewed using VMD to create movies of the atomic vibrations. To increase the motion amplitude and enhance the visibility, we can adjust the temperature setting in `nmtraj` to a higher value [11].

For the No-Bond model, in order of decreasing energy, the most prominent peak is located at 1560 cm^{-1}. Visualizing this mode reveals that it resembles an asymmetric stretching vibration mode but with carbon-oxygen-carbon angles changing during the

motion. In this vibrational mode, the oxygen atoms remain stationary, and the carbon atom moves towards and away from the oxygen atoms. The subsequent peak, at 1017 cm^{-1}, represents a symmetric stretching mode of the carbonate unit. In this mode, the carbon atom remains stationary, while the three oxygen atoms move towards and away from the carbon atom in phase, as illustrated in Figure 4.2 (b). The peak at 821 cm^{-1} is an asymmetric stretching mode. In this vibrational mode, the carbon atom remains stationary, and the oxygen atoms move toward and away from the carbon in an out-of-plane manner, as illustrated in Figure 4.2 (a)

The peak at 941 cm^{-1} corresponds to a bending vibrational mode. In this mode, the carbon moves perpendicular to the plane of the oxygen atoms, as shown in Figure 4.2 (c). In the low-energy region between 400 cm^{-1} and 50 cm^{-1}, a collection of weak peaks is seen. These peaks correspond to vibrational modes characterized by low frequencies and, consequently, lower energies. These broad-spectrum peaks represent less distinct and overlapping vibrational modes, which typically provide less information about the structure of the material.

4.1.2 Normal Mode Analysis for the Bond Model

Using the approach explained in Section 4.1.1, we generate VDOS for the Bond model, shown in Figure 4.3. A comparison of the VDOS plots of the Bond and No-Bond models indicates that adding bonds in the carbonate units affects the VDOS significantly. Adding the bond terms results in a gap between the first and second prominent peaks in the VDOS in the Bond model. In this model, the first prominent peak is located at 1625 cm^{-1} and corresponds to an asymmetric stretching vibrational mode. The second prominent peak at 737 cm^{-1} corresponds to the symmetric stretching vibrational mode. However, the left side of this peak at 776 cm^{-1} contains asymmetric stretching modes. We thus see a merging of two peaks that were well separated in the No-Bond model. The next dominant peak, at 589 cm^{-1}, corresponds to a bending vibrational mode. In addition, some weak peaks appear in the Bond model in the range of 600 cm^{-1} to 400 cm^{-1}, which are not present in the No-Bond model. Similar to the No-Bond model, a collection of weak peaks is present in the low-energy region between 400 cm^{-1} to 50 cm^{-1}.

Comparing the results for the No-Bond and Bond models, as shown in Figures 4.1 and 4.3, we observe that VDOS is sensitive to how the carbonate C-O bond is

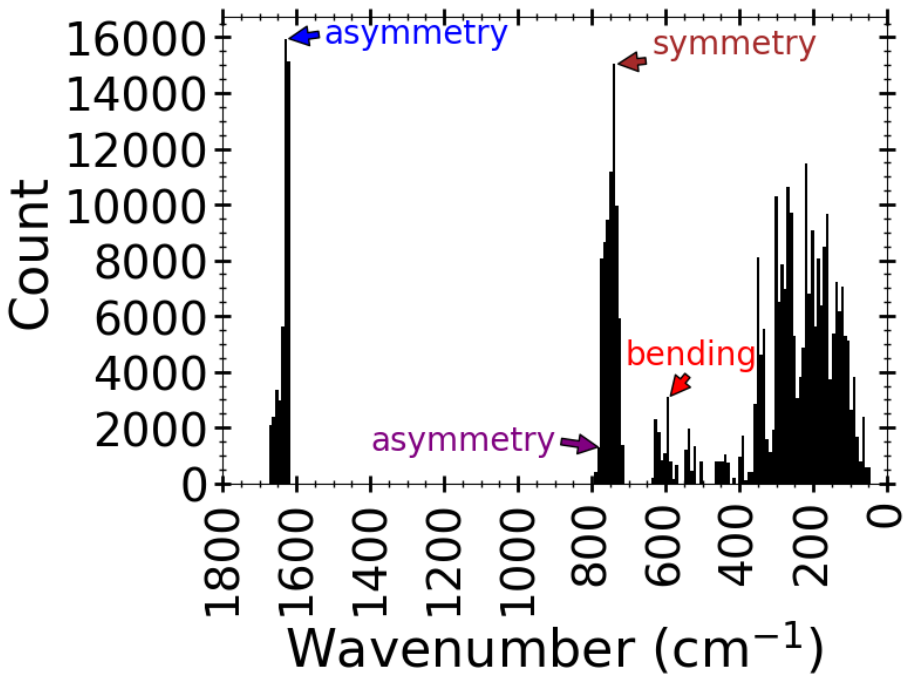


Figure 4.3: The vibrational density of states (VDOS) for the anisotropic orthorhombic Bond model using the Berendsen barostat. The first prominent peak at 1625 cm^{-1} corresponds to an asymmetric stretching mode. The next prominent peak at 737 cm^{-1} indicates a symmetric stretching mode. This peak is merged with a peak at 776 cm^{-1}, which contains asymmetric stretching modes. The peak at 589 cm^{-1} corresponds to a bending mode.

modelled. This sensitivity is evident in the shift in energy observed between the two models. The peak corresponding to the asymmetric mode, located at 1560 cm^{-1} in the No-Bond model, shifts by +65 cm^{-1} to 1625 cm^{-1} in the Bond model. The second asymmetric mode, at 821 cm^{-1} in the No-Bond model, experiences a shift of -45 cm^{-1} to 776 cm^{-1} after adding explicit, harmonic bonds. The symmetric vibrational mode, located at 1017 cm^{-1} in the No-Bond model, shows a significant shift to 737 cm^{-1} (-280 cm^{-1}) with the introduction of bonds. The bending vibrational mode in the No-Bond model, located at 941 cm^{-1}, shifts downward by -352 cm^{-1} to 589 cm^{-1} after adding bonds.

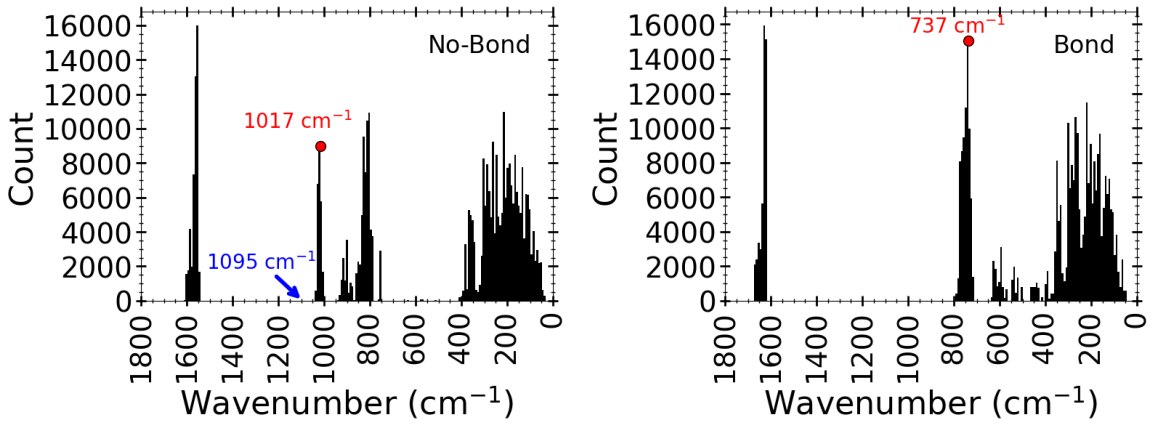


Figure 4.4: The vibrational modes calculated through GROMACS for the No-Bond (left) and Bond (right) models compared with the results of density functional theory (DFT). The peaks are identified with arrows: 1017 cm^{-1} for the No-Bond model (left) and 737 cm^{-1} for the Bond model (right), both indicating symmetric stretching vibrational modes. The results from the DFT, presented on the Crystal website [29], show the same vibrational mode at 1095 cm^{-1}. The No-Bond model shows closer agreement with a small shift of -78 cm^{-1}, while the Bond model shows a significant shift of -358 cm^{-1}.

4.2 Comparison with Density Functional Theory (DFT)

In the last part of this thesis, we compare the results of the normal mode analysis of aragonite, calculated using GROMACS, with those obtained through density functional theory (DFT). We compare our results with the published work by Carteret et al. [29]. After calculating VDOS for the Bond and No-Bond models, we visualize the vibrational modes using the VMD software and compare our results with the graphical animations on the Crystal website.

For the No-Bond model, the peak at 1017 cm^{-1} corresponds to a symmetric stretching vibrational mode, which is compatible with the mode at 1095 cm^{-1} presented on the Crystal website [29], as shown in Figure 4.4 (left). It shows a shift of -78 cm^{-1} in wavenumber. A similar vibrational mode for the Bond model is located at 737 cm^{-1}, which shows a large shift of -358 cm^{-1}, shown in Figure 4.4 (right). Considering these results, we can conclude that adding the bond terms in carbonate units significantly changes the vibrational properties of aragonite.

The VDOS analysis indicates that the inclusion of the bond terms in the force field significantly changes the VDOS, particularly affecting the vibrational modes corresponding to the frequencies between 1100 cm^{-1} to 400cm^{-1}. After adding bonds, some peaks are shifted downward, and some minor peaks show up in the range of 600 cm^{-1} to 400 cm^{-1}. These changes can be attributed to the absence of LJ and Coulomb interactions in the carbonate unit when the bond terms are included in the force field, as detailed in Section 2.3.1. When the harmonic bond is added between carbon and oxygen, the LJ and Coulomb energies in the carbonate unit are zero. The normal mode analysis indicates that the No-Bond model, which includes LJ and Coulomb interactions between C and O atoms in the carbonate unit, provides results that are more consistent with experimental data.

We conclude that when Cruz-Chu et al. [18] added an explicit C-O bond to the No-Bond model, they better recovered the unit cell parameters at ambient temperature of aragonite but worsened the resulting Bond model's VDOS.

Chapter 5

Conclusions and Future Work

This thesis focused on MD simulations of aragonite using GROMACS software. Our goal was to find an optimized simulation protocol to study aragonite's vibrational density of states (VDOS). There are two related recently-developed force fields for aragonite, the first of which excludes an explicit bond term between carbon and oxygen atoms in the carbonate units (No-Bond) [17], and the second of which includes the bond term (Bond) [18].

The first part of this thesis started by reproducing the results of the No-Bond model using three pressure algorithms: isotropic, semi-isotropic, and anisotropic. The anisotropic algorithm includes variants for triclinic and orthorhombic crystal structures. Using the GULP software, Xiao et al. [17] have reported aragonite's primary unit cell parameters for their No-Bond model. We confirmed their results by finding no significant stress in the system when simulated with isotropic pressure coupling. However, we were unable to reproduce unit cell parameters near standard conditions. We thus tested the impact of several MD options on cell vectors, including details pertaining to the truncation of short-range interactions. The results indicate that most of these parameters do not significantly impact the cell vectors. The only parameter that affected the cell vector was temperature. Our study indicated that the results reported by Xiao et al. were obtained at T=0 K, which they did not mention in their paper [17].

We also investigated the unit cell shape for the Bond model, and the results indicated that adding bonds leads to a change in the unit cell shape. Another finding

from the first part of the thesis was that both the Bond and No-Bond models preserve the orthorhombic symmetry of aragonite. Therefore, we next chose anisotropic orthorhombic pressure coupling to find an optimized simulation protocol before studying the VDOS. We tested the Berendsen and Parrinello-Rahman barostats for the anisotropic orthorhombic pressure coupling and the parameters that affect their functioning. In our tests, we looked for consistency in calculating the isothermal compressibility from the equation of state and fluctuations. The results indicated that using the Parrinello-Rahaman barostat induces non-physical volume fluctuations. However, the results using Berendsen baraostat allowed for more consistency with thermodynamics with appropriately chosen parameters. The parameters that we finalized in our study with the Berendesn barostat are tau-p=10 ps, tau-t = 1 ps, nstlist = nsttcouple = nstpcouple =1, dt = 1 fs, and constraints = none. Using the optimized simulation protocol, we determined the unit cell parameters. The calculated results for the Bond model are a = 4.93 Å, b = 8.01 Å, and c = 5.83 Å, and for the No-Bond model a = 5.01 Å, b = 8.31 Å, and c = 5.79 Å at 300 K and 1 bar. The experimental results for the unit cell parameters reported by Xiao et al. [17] are a = 4.96 Å, b = 7.97 Å, and c = 5.74 Å.

In the last part of this thesis, we calculated the VDOS for the Bond and No-Bond models using normal mode analysis. The difference in the VDOS plots for the two models indicated that adding the bond term changes the VDOS of the system. We visualized the vibrational modes for the peaks in the models using the GROMACS software and VMD. The results of the normal mode analysis were compared with the results presented by Carteret et al. [29]. Our comparison showed that our simulations reproduce the symmetric stretch peak but that the positions are quite different between the No-Bond model (1017 cm^{-1}) and the Bond model (737 cm^{-1}), with the former closest to the 1095 cm^{-1} peak position reported by Carteret et al. [29].

For future work, it would be wonderful if the C-rescale barostat could be used to study the unit cell parameters. Unfortunately, this barostat has no anisotropic version implemented in GROMACS yet. This barostat is similar to the Berendsen barostat. The crucial difference is that the C-rescale barostat contains a stochastic term that ensures correct fluctuations in the volume [47, 49]. The expectation is that using this barostat offers more stable volume adjustments and reduces the non-physical oscillations observed with other barostats.

Additionally, VDOS analysis can be performed at various temperatures to investigate the vibrational properties of aragonite. Our findings in this thesis indicate that both increasing and decreasing the temperature lead to changes in the box vectors, which will affect the VDOS of the system, as would pressure changes, defects, chemical impurities and changes in the environment surrounding the aragonite such as when the aragonite is part of a biocomposite. Consequently, the shifts in the VDOS peaks can be examined at various conditions to understand how structural variations influence the vibrational modes at different temperatures and pressures. This will aid in interpreting spectroscopic results in terms of the microstructure of aragonite-based materials.

Bibliography

[1] Fadia, P., Tyagi, S., Bhagat, S., Nair, A., Panchal, P., Dave, H., Dang, S., & Singh, S. (2021). Calcium carbonate nano- and microparticles: Synthesis methods and biological applications. *3 Biotech*, *11*, 457.

[2] Kawashima, S., Hou, P., Corr, D. J., & Shah, S. P. (2013). Modification of cement-based materials with nanoparticles. *Cement and Concrete Composites*, *36*, 8–15.

[3] Barhoum, A., Rahier, H., Abouzeid, R., Rehan, M., Dufour, T., Hill, G., & Dufresne, A. (2014). Effect of cationic and anionic surfactants on application of calcium carbonate nanoparticles in paper coating. *ACS Applied Materials and Interfaces*, *6*, 2734–2744.

[4] Munyemana, J. C., He, H., Fu, C., Wei, W., Tian, J., & Xiao, J. (2018). Trypsin-calcium carbonate hybrid nanospheres based enzyme reactor with good stability and reusability. *New Journal of Chemistry*, *42*, 18388–18394.

[5] Kiranda, H. K., Mahmud, R., Abubakar, D., & Zakaria, Z. A. (2018). Fabrication, characterization and cytotoxicity of spherical-shaped conjugated gold-cockle shell derived calcium carbonate nanoparticles for biomedical applications. *Nanoscale Research Letters*, *13*, 1.

[6] Ian Miller, D. H. (2011). *Encyclopedia of modern coral reefs: Structure, form and process* (1st ed.). Springer.

[7] Meldrum, F. (2003). Calcium carbonate in biomineralization and biomimetic chemistry. *International Materials Reviews*, *48*, 187–224.

[8] Yoshikawa, M. (2023). *Advanced optical spectroscopy techniques for semiconductors* (1st ed.). Springer.

[9] Kittel, C. (2004). *Introduction to solid state physics* (8th ed.). Wiley.

[10] Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, *7*, 306–317.

[11] Lindahl, Abraham, & Spoel, V. D. (2021). Gromacs 2020.5 manual. *Zenondo*.

[12] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1-2*, 19–25.

[13] Berendsen, H., Van Der Spoel, D., & Van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*, 43–56.

[14] Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, Load-Balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, *4*, 435–447.

[15] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, *29*, 845–854.

[16] Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). *Tackling exascale software challenges in molecular dynamics simulations with gromacs*.

[17] Xiao, S., Edwards, S., & Gräter, F. (2011). A new transferable forcefield for simulating the mechanics of $CaCO_3$ crystals. *The Journal of Physical Chemistry C*, *115*, 20067–20075.

[18] Cruz-Chú, E. R., Xiao, S., Patil, S. P., Gkagkas, K., & Gräter, F. (2017). Organic filling mitigates flaw-sensitivity of nanoscale aragonite. *ACS Biomaterials Science and Engineering*, *3*, 260–268.

[19] Goldstein, H. (1980). *Classical mechanics* (2nd ed.). Addison-Wesley.

[20] El-Azazy, M. (2018). Introductory chapter: Infrared spectroscopy - a synopsis of the fundamentals and applications. In M. El-Azazy (Ed.). IntechOpen.

[21] Raman, C. V. (1953). A new radiation. *Proceedings Indian Academy of Sciences*, *37*, 333–341.

[22] Robert M. Silverstein, D. K., Francis X. Webster. (2005). *Solutions manual for spectrometric identification of organic compounds* (7th ed.). Wiley.

[23] Regev, L., Poduska, K. M., Addadi, L., Weiner, S., & Boaretto, E. (2010). Distinguishing between calcites formed by different mechanisms using infrared

spectrometry: archaeological applications. *Journal of Archaeological Science*, *37*, 3022–3029.

[24] Gao, B., & Poduska, K. M. (2023). Tracking amorphous calcium carbonate crystallization products with Far-Infrared spectroscopy. *Minerals*, *13*, 110.

[25] Wang, X., Ye, Y., Wu, X., Smyth, J. R., Yang, Y., Zhang, Z., & Wang, Z. (2018). High-temperature raman and ftir study of aragonite-group carbonates. *Physics and Chemistry of Minerals*, *46*, 51–62.

[26] Dickens, B., & Bowen, J. (1971). Refinement of the crystal structure of the aragonite phase of $CaCO_3$. *Journal of Research of the National Bureau of Standards*, *75A*, 27.

[27] Bragg, W. L. (1924). The structure of aragonite. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*, *105*, 16–39.

[28] De Villiers, J. P. (1971). Crystal structures of aragonite, strontianite, and witherite. *American Mineralogist*, *56*, 758–767.

[29] Carteret, C., De La Pierre, M., Dossot, M., Pascale, F., Erba, A., & Dovesi, R. (2013). The vibrational spectrum of $CaCO_3$ aragonite: A combined experimental and quantum-mechanical investigation. *The Journal of Chemical Physics*, *138*, 014201.

[30] Frenkel, D., & Smit, B. (1996). *Understanding molecular simulation: From algorithms to applications* (2nd ed.).

[31] Satoh, A. (2010). *Introduction to practice of molecular simulation: Molecular dynamics, monte carlo, brownian dynamics, lattice boltzmann and dissipative particle dynamics* (1st ed.). Elsevier.

[32] Hess, B., Kutzer, C., van der Spoel, D., & Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, *4*, 435–447.

[33] Cuendet, M., & van Gunsteren, W. (2007). On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm. *The Journal of Chemical Physics*, *127*, 184102.

[34] Gonzalez, M. (2011). Force fields and molecular dynamics simulations. *École Thématique de la Société Française de la Neutronique*, *12*, 169–200.

[35] Pricopi, D., & Popescu, E. (2016). Phase-space structure of the Buckingham's two-body problem. *Astrophysics and Space Science*, *361*, 1–6.

[36] Gale, J. D. (1997). Gulp: A computer program for the symmetry-adapted simulation of solids. *Journal of the Chemical Society Faraday Transactions*, *93*, 629–637.

[37] Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38.

[38] Mackay, D. H. J., Cross, A. J., & Hagler, A. T. (1989). The role of energy minimization in simulation strategies of biomolecular systems (1st ed.), Springer.

[39] Adcock, S., & McCammon, J. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, *106*, 1589–615.

[40] Basconi, J. E., & Shirts, M. R. (2013). Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, *9*, 2887–2899.

[41] Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*, 3684–3690.

[42] Tuckerman, M. E. (2010). *Statistical mechanics: Theory and molecular simulation* (1st ed.). Oxford University Press.

[43] Carteret, C., De La Pierre, M., Dossot, M., Pascale, F., Erba, A., & Dovesi, R. (2020). *Vibrational spectrum of aragonite*. Retrieved September 16, 2024, from https://www.crystal.unito.it/vibs/aragonite/

[44] Zhuravlev, Y. N., & Atuchin, V. V. (2020). Comprehensive density functional theory studies of vibrational spectra of carbonates. *Nanomaterials*, *10*, 2275.

[45] De La Pierre, M., Demichelis, R., Wehrmeister, U., Jacob, D. E., Raiteri, P., Gale, J. D., & Orlando, R. (2014). Probing the multiple structures of vaterite through combined computational and experimental raman spectroscopy. *The Journal of Physical Chemistry C*, *118*, 27493–27501.

[46] Matas, J., Gillet, P., Ricard, Y., & Martinez, I. (2000). Thermodynamic properties of carbonates at high pressures from vibrational modelling. *European Journal of Mineralogy*, *12*, 703–720.

[47] Abraham, M., Alekseenko, A., Bergh, C., Blau, C., Briand, E., Doijade, M., Fleischmann, S., Gapsys, V., Garg, G., Gorelov, S., Gouaillardet, G., Gray, A., Irrgang, M. E., Jalalypour, F., Jordan, J., Junghans, C., Kanduri, P., Keller, S., Kutzner, C., ... Lindahl, E. (2023). Gromacs 2023 manual. https://doi.org/10.5281/zenodo.7588711

[48] Abraham, M., Alekseenko, A., Basov, V., Bergh, C., Briand, E., Brown, A., Doijade, M., Fiorin, G., Fleischmann, S., Gorelov, S., Gouaillardet, G., Gray, A., Irrgang, M. E., Jalalypour, F., Jordan, J., Kutzner, C., Lemkul, J. A., Lundborg, M., Merz, P., ... Lindahl, E. (2024). *GROMACS 2024.3 source code* (Version 2024.3). https://gitlab.com/gromacs/gromacs

[49] Bernetti, M., & Bussi, G. (2020). Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics*, *153*, 114107.

Appendix A

The Simulation Results

A.1 Box Vectors

Figures A.1 and A.2 display the box vectors in the y and z directions for the models introduced in Table 3.1. The results are plotted for the second half of the time series. Figure A.1 shows that for the Bond models using the isotropic pressure algorithm, the box vector ranges between 2.39 and 2.40 nm, which is smaller than in the No-Bond models, where it ranges between 2.41 and 2.43 nm. For the anisotropic and semi-isotropic algorithms, adding the Bond term results in a box vector smaller than the expected value. For the semi-isotropic algorithm Bond model, the average box vector is around 2.37 nm, and for the anisotropic algorithm Bond model, it ranges between 2.31 and 2.32 nm. However, for the No-Bond models, all the results are larger than the expected value, ranging between 2.41 and 2.44 nm.

In the z direction, as shown in Figure A.2, the anisotropic and semi-isotropic simulations exhibit an inverse behaviour in the No-Bond model compared to the y direction. These pressure algorithms result in the largest box vector, with semi-isotropic values ranging between 2.36 and 2.37 nm and anisotropic values between 2.38 and 2.39 nm. In contrast, the Bond model with the isotropic pressure algorithm shows smaller values, ranging from 2.29 to 2.40 nm. For the No-Bond model, the semi-isotropic and anisotropic pressure algorithms result in a smaller box vector (2.29-2.32 nm) compared to the isotropic algorithm (2.32-2.33 nm).

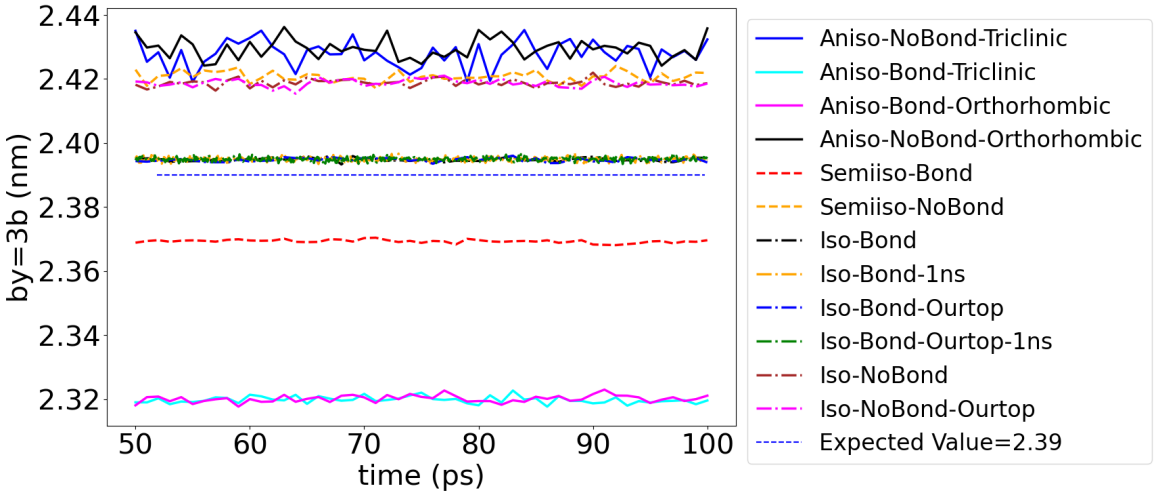


Figure A.1: Comparison of box vector in the y direction (b_y) for the Bond and No-Bond models using different pressure algorithms. The isotropic Bond models indicate a smaller value(2.39–2.40 nm) than the No-Bond models (2.41–2.43 nm). For the semi-isotropic and anisotropic algorithms, the Bond models show box vectors of around 2.37 nm and 2.31–2.32 nm, respectively, while the No-Bond models exhibit larger b_y (2.41–2.44 nm).

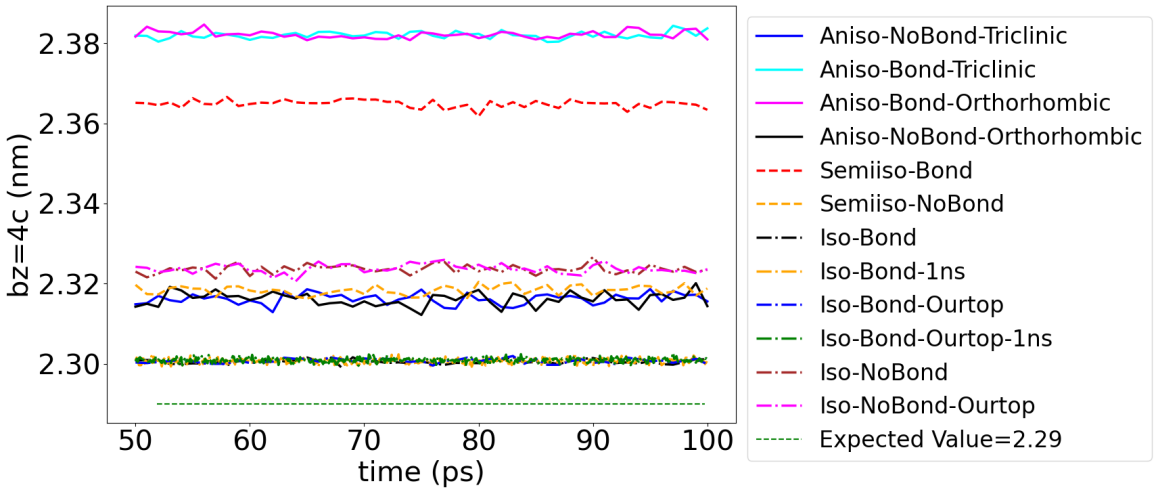


Figure A.2: In the z direction, anisotropic and semi-isotropic simulations of the No-Bond model show the largest box vector (2.36-2.37 nm for semi-isotropic, 2.38-2.39 nm for anisotropic), inversely compared to the y direction. The Bond model with the isotropic pressure algorithm has smaller values (2.29-2.40 nm). For the No-Bond model, the semi-isotropic and anisotropic algorithms yield smaller magnitudes (2.29-2.32 nm) than the isotropic algorithm (2.32-2.33 nm).

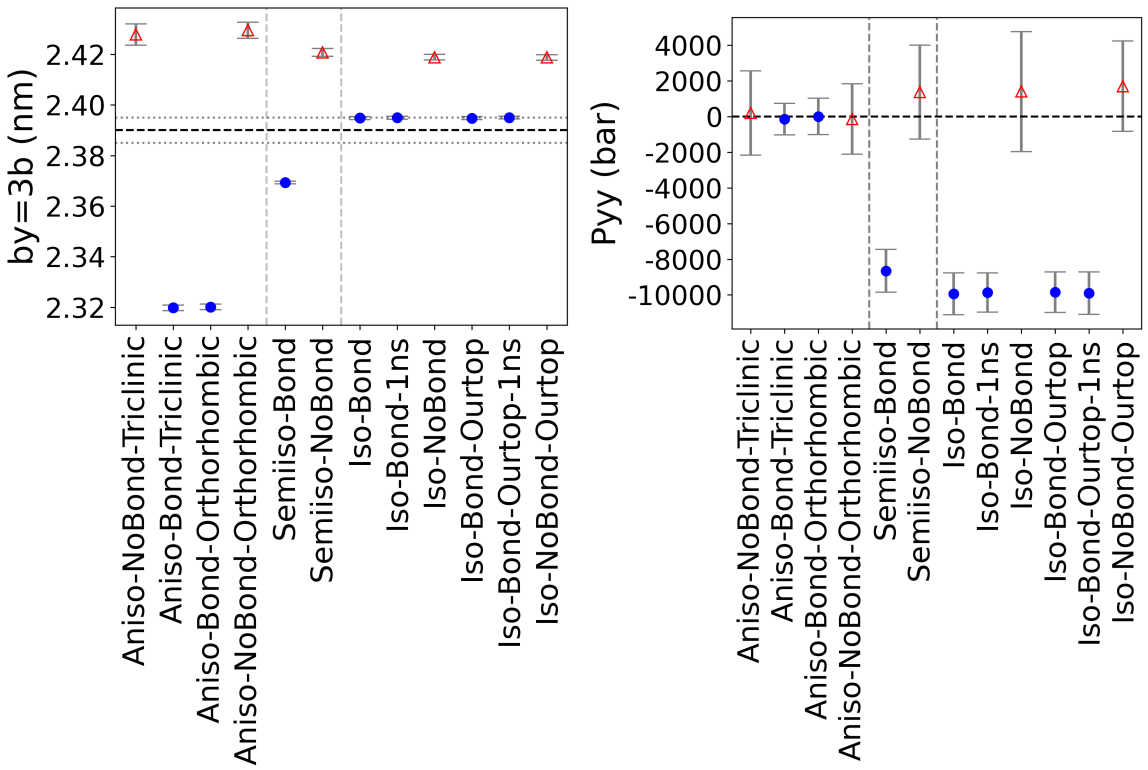


Figure A.3: Box vector in the y direction (left) and pressure tensor element P_{yy} (right) for three pressure algorithms and two force fields are presented. Blue markers represent models with the bond term in the force field, and the red markers show no-bond models.

A.2 Pressure Tensor

Figures A.3 and A.4 show the box vector (left panel) and pressure tensor (right panel) in the y and z directions. In the left graphs, it can be seen that for the No-Bond models, the box vectors are larger than the expected value for all pressure algorithms. However, adding the bond term for the isotropic pressure algorithm results in a smaller box vector. Different behaviours are observed for anisotropic and semi-isotropic pressure algorithms. In the y direction, adding the bond term results in a box vector smaller than the expected value, while in the z direction, the box vector is larger than the expected value.

Looking at pressure plots on the right (P_{yy} and P_{zz}), for both the Bond and No-Bond models, when using an anisotropic pressure algorithm, the average pressure is close to zero, indicating no stress on the system. For the semi-isotropic and isotropic

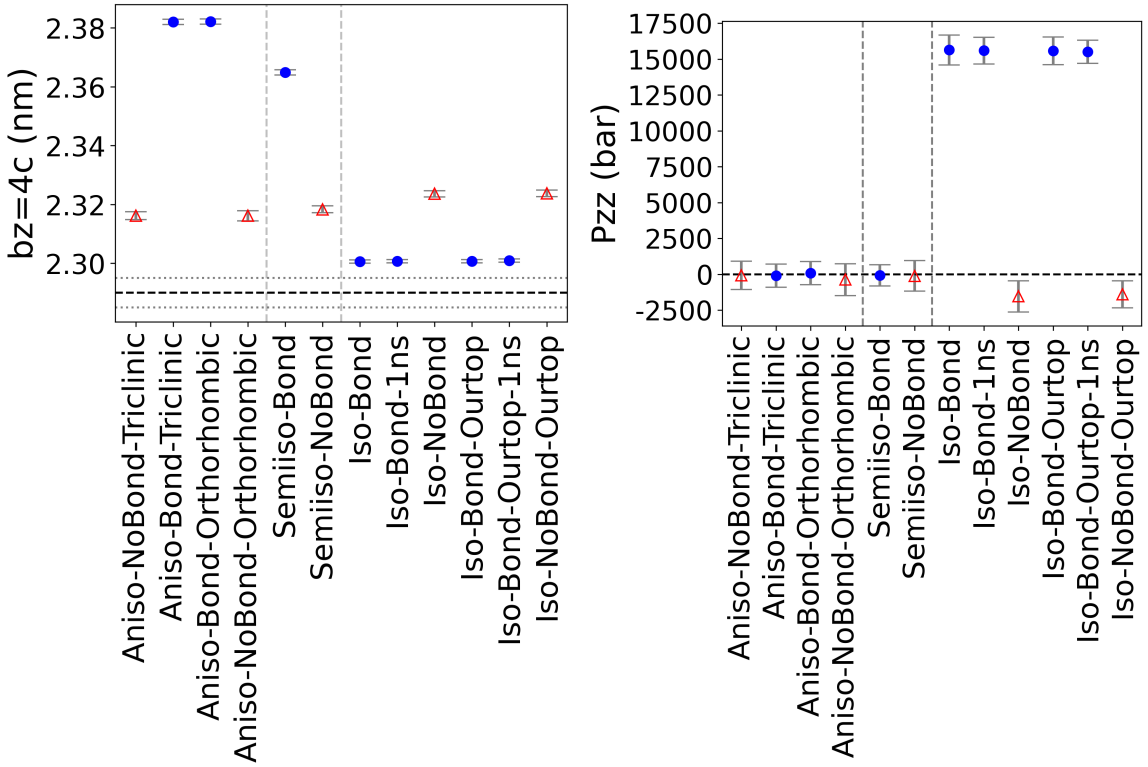


Figure A.4: Box vector in the z direction (left) and pressure tensor element P_{zz} (right) for three pressure algorithms and two force fields are presented. Blue markers represent models with the bond term in the force field, and the red markers show no-bond models.

algorithms, the No-Bond models show pressure values closer to zero. In contrast, adding the bond term results in a pressure deviation from zero, indicating stress on the system.

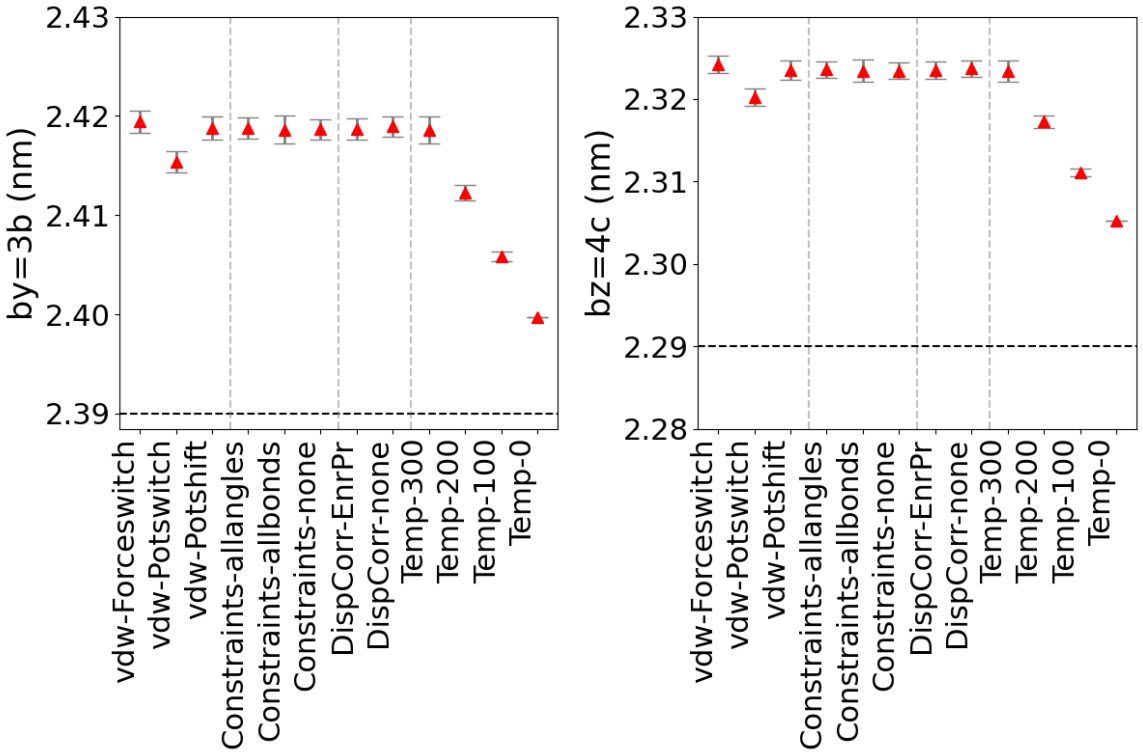


Figure A.5: Box vector in the y and z directions for the isotropic No-Bond model. The results show that the average box vector remains constant for most of the studied parameters. The only significant change observed in box size occurs with a reduction in temperature. The results indicate that the average box vector decreases from approximately 2.420 nm at 300 K to 2.398 nm at 0 K in the y direction and from 2.325 nm at 300 K to 2.305 nm at 0 K in the z direction.

A.3 Isotropic No-Bond Model: Detailed Parameters Study

The input parameters in the .mdp file are studied, and the box vectors are calculated. The results show that temperature is the only parameter significantly affecting the box vectors. Decreasing the temperature from 300 K to 0 K results in a decrease of the box vectors in the y and z directions. Changes to other parameters in the .mdp file do not significantly alter the box vectors, as illustrated in Figure A.5.

Appendix B

Files to Generate the Simulation Box

B.1 .gro File

Using the crystal structure of aragonite from the American Mineralogist Crystal Structure Database and the Pmcn space group, we generate a '.gro' file for aragonite unit cell as follows:

```
ARAGONITE
 20
    1arag  Ca     1   0.124   0.331   0.436
    1arag  C      2   0.124   0.607   0.525
    1arag  O1     3   0.124   0.735   0.519
    1arag  O2     4   0.235   0.543   0.525
    1arag  O3     5   0.483   0.254   0.050
    2arag  Ca     6   0.372   0.466   0.138
    2arag  C      7   0.372   0.190   0.050
    2arag  O1     8   0.372   0.062   0.055
    2arag  O2     9   0.261   0.254   0.050
    2arag  O3    10   0.013   0.543   0.525
    3arag  Ca    11   0.372   0.729   0.425
    3arag  C     12   0.372   0.209   0.337
```

```
  3arag  O1    13   0.372   0.337   0.342
  3arag  O2    14   0.261   0.144   0.337
  3arag  O3    15   0.013   0.653   0.238
  4arag  Ca    16   0.124   0.068   0.149
  4arag  C     17   0.124   0.588   0.238
  4arag  O1    18   0.124   0.460   0.232
  4arag  O2    19   0.235   0.653   0.238
  4arag  O3    20   0.483   0.144   0.337
 0.49500   0.79600   0.57400
```

B.2 Program to Generate the Simulation Box

```
      implicit real*8 (b-h,o-z)
    character*12 a1,a2,a3
    parameter (nmolx=5,nmoly=3,nmolz=4,nmax=nmolx*nmoly*nmolz)
    parameter (nmaxmol=4000)
    dimension rca(nmaxmol,3,4)
    dimension rc(nmaxmol,3,4)
    dimension ro(nmaxmol,3,3,4)
    character*77 filename

    open(unit=1,file="aragonite-OneMol.gro",status="old")


100 continue
    read(1,*,end=999)
    read(1,*)
    do i=1,1
     do j=1,4
      read(1,*)a1,a2,a3,rca(i,1,j),rca(i,2,j),rca(i,3,j)
      read(1,*)a1,a2,a3,rc(i,1,j),rc(i,2,j),rc(i,3,j)
      read(1,*)a1,a2,a3,ro(i,1,1,j),ro(i,2,1,j),ro(i,3,1,j)
```

```
        read(1,*)a1,a2,a3,ro(i,1,2,j),ro(i,2,2,j),ro(i,3,2,j)
        read(1,*)a1,a2,a3,ro(i,1,3,j),ro(i,2,3,j),ro(i,3,3,j)
       enddo
      enddo
      read(1,*)box_a, box_b, box_c
      goto 100

999   close(1)

      ns=0
      i=0
      tconf=0.d0
      facx=0.0d0
      facy=0.0d0
      facz=0.0d0

      write(filename,'("aragonite-",i4.4,"unitcell.gro")')nmax

      open(11,file=filename,status="unknown")

c      write(11,99001)tconf
      write(11,99001)
      write(11,99002)nmax*20

      do imol=1,nmolz
      facy=0.d0
      do jmol=1,nmoly
      facx=0.d0
      do kmol=1,nmolx
      do j=1,4
      i=i+1
      ns=ns+1
      write(11,99004)i,ns,rc(1,1,j)+facx*box_a,rc(1,2,j)+facy*box_b
     x               ,rc(1,3,j)+facz*box_c
```

```
 ns=ns+1
 write(11,99005)i,ns,ro(1,1,1,j)+facx*box_a,ro(1,2,1,j)+facy*box_b
x               ,ro(1,3,1,j)+facz*box_c
 ns=ns+1
 write(11,99006)i,ns,ro(1,1,2,j)+facx*box_a,ro(1,2,2,j)+facy*box_b
x               ,ro(1,3,2,j)+facz*box_c
 ns=ns+1
 write(11,99007)i,ns,ro(1,1,3,j)+facx*box_a,ro(1,2,3,j)+facy*box_b
x               ,ro(1,3,3,j)+facz*box_c
 enddo
  facx=facx+1.d0
 enddo
  facy=facy+1.d0
 enddo
  facz=facz+1.d0
 enddo

 facz=0.d0
 do imol=1,nmolz
 facy=0.d0
 do jmol=1,nmoly
 facx=0.d0
 do kmol=1,nmolx
 do j=1,4
 i=i+1
 ns=ns+1
 write(11,99003)i,ns,rca(1,1,j)+facx*box_a,rca(1,2,j)+facy*box_b
x               ,rca(1,3,j)+facz*box_c
 enddo
  facx=facx+1.d0
 enddo
  facy=facy+1.d0
 enddo
  facz=facz+1.d0
```

```
      enddo


      print*,  facx, facy, facz
      write(11,99008)nmolx*box_a,nmoly*box_b,nmolz*box_c


      close(11)

c99001 FORMAT ('aragonite t=',f13.5)
99001 FORMAT ('aragonite ')
99002 FORMAT (i5)
99003 FORMAT (i5,'calc  CAM',i6,f8.3,f8.3,f8.3)
99004 FORMAT (i5,'carb  CM0',i6,f8.3,f8.3,f8.3)
99005 FORMAT (i5,'carb  OM1',i6,f8.3,f8.3,f8.3)
99006 FORMAT (i5,'carb  OM2',i6,f8.3,f8.3,f8.3)
99007 FORMAT (i5,'carb  OM3',i6,f8.3,f8.3,f8.3)
99008 FORMAT (f10.5,f10.5,f10.5)

c99001 FORMAT ('aragonite t=',f13.5)
c99002 FORMAT (i5)
c99003 FORMAT (i5,'aragonite  Ca',i6,f8.3,f8.3,f8.3)
c99004 FORMAT (i5,'aragonite  C ',i6,f8.3,f8.3,f8.3)
c99005 FORMAT (i5,'aragonite  O1',i6,f8.3,f8.3,f8.3)
c99006 FORMAT (i5,'aragonite  O2',i6,f8.3,f8.3,f8.3)
c99007 FORMAT (i5,'aragonite  O3',i6,f8.3,f8.3,f8.3)
c99008 FORMAT (f10.5,f10.5,f10.5)



      end
```

B.3 .top File

The topology file can be generated by hand without using the `pdb2gmx` command. This topology file is used for the simulation with the Bond model using the isotropic pressure algorithm.

```
[ defaults ]
; nbfunc        comb-rule        gen-pairs        fudgeLJ        fudgeQQ
  1                1                yes                0.5        0.5

[atomtypes]
;name     mass       charge   ptype    C6           C12
CA       40.08000    1.668       A   1.42E-003    2.52E-007
CM       12.01100    0.999       A   4.61E-003    1.44E-005
OM       15.99940    -0.889      A   2.03E-003    1.77E-006

[ nonbond_params ]
; C parameter below incorporates patch
  CA       OM      1       0            9.49E-007
  CA       CA      1       1.42E-003    2.52E-007
  CA       CM      1       2.55E-003    1.89E-006
  CM       CM      1       1.43E-002    4.61E-006
  CM       OM      1       3.08E-004    9.04E-010
  OM       OM      1       5.21E-005    5.94E-007

[moleculetype]
; name nrexcl
carb  3

[atoms]
; nr type resnr residu atom cgnr charge
1     CM   1     carb   CM0  1    +0.999        12.01100
2     OM   1     carb   OM1  1    -0.889        15.99940
3     OM   1     carb   OM2  1    -0.889        15.99940
4     OM   1     carb   OM3  1    -0.889        15.99940
```

```
[ bonds ]
; i   j   func    b0      kb
  1   2    1      0.12940 263173.6
  1   3    1      0.12940 263173.6
  1   4    1      0.12940 263173.6


[ angles ]
; i j k type theta_0 K_theta
2 1 3 1 120 1852
3 1 4 1 120 1852
2 1 4 1 120 1852

[ dihedrals ]
;i j k l type phi_s K_phi multiplicity
2 3 1 4 4 180 28.9 2


[moleculetype]
; name nrexcl
calc  3

[atoms]
; nr type resnr residu atom cgnr charge
1     CA   1     calc   CAM  1    1.668     40.08000

[system]
aragonite melting

[molecules]
carb        240
calc            240
```