MOLECULAR DYNAMICS STUDY OF DIFFUSION OF CARBON MONOXIDE IN WATER AT DIFFERENT TEMPERATURES

A DISSERTATION

SUBMITTED FOR PARTIAL FULFILLMENT OF THE REQUIREMENT FOR

MASTERS' DEGREE OF SCIENCE IN PHYSICS

By

ISHWOR POUDYAL

Central Department of Physics

Institute of Science and Technology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

November, 2013

ACKNOWLEDGEMENTS

I would like to acknowledge the guidance and support I have received from my supervisor Dr. Narayan Prasad Adhikari. His encouragement and assistance at all times have been of immense value.

I am especially grateful to Prof. Dr. Lok Narayan Jha (Head, Central Department of Physics) for his valuable co-operation throughout this work. I would like to express my sincere thanks to all my respected faculty members of the Central Department of Physics.

I am thankful to my seniors Mr. Sunil Kumar Thapa, Mr. Anil Adhikari and Mr. Saran Lamichhane for their help and friendly advice.

I am happy to have a good bunch of friends with whom I enjoy every academic and non-academic discussions which adds a morale-boost in my work. I am pleased with Mr. Keshav Sharma with whom I have enjoyed the fruitful discussions regarding the subject matter. I remember my friends Mr. Hemanta Bhattarai, Mr. Saroj Dhakal and Mr. Saroj Dangi who unconditionally helped me in learning software LaTeX, solve the computer trouble I encountered and motivate me in each and every problems I faced.

I acknowledge the partial support from The Abdus Salam International Centre for Theoretical Physics (ICTP) through office of external activities within Net-56. Further I would like to acknowledge University Grants Commission (UGC), Nepal for its partial financial support.

I would like to dedicate this thesis to my parents and family members, whose immense support in each and every step of my life, help to reinforce my morale.

RECOMMENDATION

It is certified that Mr. Ishwor Poudyal has carried out the dissertation work entitled "**Molecular dynamics study of diffusion of carbon monoxide in water at different temperatures**" under my supervision and guidance.

I recommend the dissertation in the partial fulfillment for the requirement of Masters' Degree of Science in Physics.

(Supervisor)
(Dr. Narayan Prasad Adhikari)
Central Department of Physics
Tribhuvan University, Kirtipur
Kathmandu, Nepal

Date: .....

EVALUATION

We certify that we have read this dissertation and in our opinion, it is good in the scope and quality as dissertation in partial fulfillment for the requirement of Masters' Degree of Science in Physics.

Evaluation Committee

(Dr. Narayan Prasad Adhikari)
(Supervisor)

(Prof. Dr. Lok Narayan Jha)
(Head)
Central Department of Physics
Tribhuvan University, Kirtipur
Kathmandu, Nepal

(External Examiner)

(Internal Examiner)

Date : .....

Abstract

In the present work, molecular dynamics simulations of a mixture of carbon monoxide (CO) gas in SPC/E water (H_2O) with CO as a solute and water as a solvent have been performed to understand the self-diffusion coefficient of the components i.e. CO and water along with the mutual diffusion coefficients and structural properties of the system studying the radial distribution function (RDF) of the components present in the system at different temperatures 293K, 298K, 303K, 313K, 323K, 333K, 345K, 360K. The mole fraction of CO in the system is 0.018 and that of water is 0.982. The solvent-solvent, solute-solute and solute-solvent radial distribution functions have been estimated to study the structural properties of the system. The self-diffusion coefficient of CO is calculated using mean square displacement (MSD) and velocity auto correlation function (VACF) and that of water is estimated using MSD method only. The self-diffusion coefficient of CO agrees within around 20% of the experimental results and that of water agrees within around 10% of the experimental results. The mutual diffusion coefficient of solute-solvent is calculated using Darken's relation. The temperature dependence of self-diffusion coefficients of CO and water and mutual diffusion coefficients of CO in water all follow the Arrhenius behavior from where we estimate the activation energy of diffusion process. Thus estimated activation energies also agree to experimental values.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 General Considerations

The chemical and biochemical processes that are essential to living systems take place in aqueous environment. Water, the most abundant liquid on the Earth, participates in the ecological cycle and acts as the essential element of life. Carbon monoxide (CO), on the other hand is toxic to humans and animals when encountered in higher concentration [1]. The study of these molecules is of great importance because of their impact in living beings.

1.2 Carbon Monoxide

Carbon monoxide, normally present in the atmosphere at a concentration less than 0.001%, is an odorless, colorless, and tasteless gas. It is readily recognized for its toxic effects when encountered in higher concentrations though is also produced in normal animal metabolism in low quantities [2].

Worldwide, the largest source of carbon monoxide is natural in origin, due to photochemical reactions in the troposphere that generates about 5×10^{12} kilograms per year. Other natural sources of carbon monoxide include volcanoes, forest fires, and other forms of combustion [1]. Besides the natural production, some production is also due to vehicular and industrial emission.

Carbon monoxide (CO) is a very weak direct greenhouse gas, but has important indirect effects on global warming. CO reacts with hydroxyl (OH) radicals in

the atmosphere reducing their abundance. As OH radicals help to reduce the lifetimes of strong green house gases like methane, carbon dioxide, nitrous oxide, CO indirectly increases the global warming potential of the gases [3].
The physical parameters for CO is mentioned in table (1.1) [4].

Table 1.1: Some Physical Parameters of CO.

Physical Quantity	**Unit**	**CO**
Molar mass	gm/mol	28.010
Melting point (1 atm)	°C	-205.02
Boiling point (1 atm)	°C	-191.50
Density	gm/cc	0.7909
Critical Temperature	°C	-140.29
Critical Pressure	MPa	3.494
Critical Molar Volume	cm^3/mol	93

1.3 Diffusion and its Importance

Molecular diffusion occurs, as different species of a mixture move under the influence of concentration gradients [5]. It plays a key role in a variety of atmospheric and biospheric sciences. Diffusion is fundamental for transport of matter and for ionic conduction in disordered materials [6, 7]. The kinetics of many microstructural changes that occur during preparation, processing and heat treatment of materials include diffusion, typical examples are nucleation of new phases, diffusive phase transformation precipitation and dissolution of a second phase, homogenization of alloys, recrystallization and thermal oxidation [5]. The application of diffusion is on doping during fabrication of microelectronic devices, the operation of solid electrolytes for batteries and fuel cells, surface hardening of steel through carburisation or nitridation etc. [8].

The first attempt to measure self-diffusion (the most basic diffusion process) was done by physico-chemist George Karl Von Hevesy who studied self-diffusion in liquid [9] and in solid lead [10] by using a natural radioisotope ^{210}Pb and ^{212}Pb of lead. From the pioneering work of Alder and Wainwright [11, 12] the simulation of diffusion coefficient has been an area of continuous research. The

equations of Fick, the statistical interpretation of diffusion coefficient by Einstein and Smoluchowski and the Boltzmann-Matano method for concentration dependent diffusion coefficients opened the way for new experimental techniques [8].

Now, there is a magnificent increase in the application of computer modeling and simulation method in different areas of research [13, 14, 15] and to study the diffusion process [6]. The study of the diffusion process have been done using molecular dynamics simulation technique [16, 17, 18]. Rahman (1971) considered 216 rigid water molecules to examine the structural and kinetic behaviour of liquid water at 34.3 ^{o}C [19]. Malenkov et al. studied the physical, structural and dynamical properties of liquid water and ice [20, 21]. Malenkov considered a huge system containing 3456 water molecules to study the correlation coefficient [21].

Due to the availability of the open source packages like GROMACS [22] and the different experimental techniques Ringbom apparatus [23], Diaphragm cell, Infinite Couple, Taylor Dispersion, Nuclear Magnetic resonance [24] the study of diffusion process is familiar in large scientific family. Despite the experimental study of diffusion process of CO [25], to the best of our knowledge, there has been no molecular dynamics study on the diffusion of CO in water.

1.4 Scope of the Present Work

Carbon monoxide (CO) is a pollutant commonly recognized for its toxicological attributes. One of the major causes of deaths in fires has been smoke toxicity. CO, a highly toxic combustion product, is one of the major component of smoke derived from fires, a primary cause of the lethality for smoke in fires. It is rapidly absorbed through the lungs and binds mostly to haemoglobin (Hb) and mildly to intracellular cytochrome oxidase. The affinity of CO to Hb molecule is 200-240 times more than the affinity of Oxygen to Hb. As a result the carboxyhaemoglobin level increases which limits the oxygen carrying capacity of the Hb resulting cellular anoxia which leads to systematic and cutaneous manifestations in pregnant woman [2].

CO is the raw material in the polycarbonate, polyurethane and oxy-alcohol manufacture and is also used in the manufacture of metal carbonyls. It is used in calibration gas mixtures for petrochemical industry, environmental emission monitoring, industrial hygiene monitors and trace impurity analyzers [26].

A test of the diffusing capacity of the lungs for carbon monoxide ($D_L CO$) is one of the most clinically valuable tests of lung function. $D_L CO$ measures the ability of the lungs to transfer gas from inhaled air to the red blood cells in pulmonary capillaries. It has been shown to be a sensitive indicator of gas exchange, being abnormal in patients with interstitial lung disease, pulmonary vascular lung disease and emphysema [27]. So, the study of diffusion of CO has great value in biological and environmental science. Hence, we intend to study the molecular dynamics simulation study of carbon monoxide in water.

The experimental measurement of diffusion coefficients is reputed to be very difficult and very expensive. Molecular dynamics simulation is considered to be the best alternative for study of system dynamics and structure, which is economic and free from experimental hazards. It can also play an important role of guide to the experimental studies [16]. So, we have chosen the molecular dynamics technique to study the diffusion properties of CO in water. We have arranged our work in 5 chapters.

In chapter 2, there is a discussion of the theoretical basis of diffusion. This chapter also introduces the Fick's law, continuity equation and using this we have derived the Einstein's relation. A brief introduction of Green Kubos formula has also been given. The binary diffusion coefficient has been described along with the way i.e. empirical Darken's relation to calculate it. A general introduction of radial distribution function is kept along with brief introduction of molecular dynamics technique with its limitation. The description of GROMACS (**GR**oningen **MA**chine for **C**hemical **S**imulation) is provided.

In chapter 3, the details of simulation have been presented along with the description of the input files and parameters. This chapter along with chapter 4 includes the technical part of the work, which will provide the interested ones, the way for research using this software.

Chapter 4 consists the Results and Discussions part. The structural analysis of the system (carbon monoxide in water) is done using radial distribution function. The self-diffusion coefficient of water and carbon monoxide is calculated using mean square displacement (MSD) method and also compared the self-diffusion coefficient of carbon monoxide using MSD with velocity autocorrelation function (VACF). The binary diffusion coefficient is calculated using Darken's relation. The temperature variation of diffusion coefficient is also studied.

In chapter 5, we provide the conclusion of this research work and opens the way for more future works.

Chapter 2

Theory

2.1 Diffusion Theory

2.1.1 Introduction to Diffusion

When a fluid F_1 is carefully introduced to a recipient which contains a fluid F_2 which is miscible with F_1, after a certain duration we notice that movements related to mixing have ceased and the hydrostatic equilibrium is attained; however, an inhomogeneity can persist in the concentration (this can be observed by differences in color for example). This indicates that the phenomenon is an extremely slow function of time. This migration of the components with respect to one another constitutes a phenomenon known as material *diffusion* which is characterized by the fact all of the components of a mixture do not have exactly the same speed, and these differences lead to variations in the composition of the mixture [28].

2.1.2 Definition

Diffusion is the process by which matter is transported from one part of a system to another as a result of random molecular motions. The diffusion coefficient is a response property of the system to a concentration inhomogeneity. Diffusion also occurs when a system is not in equilibrium and random molecular motion tends to bring everything toward uniformity.

If a droplet of ink is placed without stirring at the bottom of a bottle filled

with water, the color will slowly spread through the bottle. At first, it will be concentrated near the bottom. After a few days, it will penetrate upward a few centimeters. After several days, the solution will be colored homogeneously. The process responsible for the movement of the colored material is diffusion. Diffusion is caused by the Brownian motion of atoms or molecules that leads to complete mixing. In gases, diffusion progresses at a rate of centimeters per second; in liquids, its rate is typically fractions of millimeters per second; in solids, diffusion is a fairly slow process and the rate of diffusion decreases strongly with decreasing temperature: near the melting temperature of a metal a typical rate is about one micrometer per second; near half of the melting temperature it is only of the order of nanometers per second [5].

2.1.3 Self-Diffusion

The diffusion in a homogeneous system where no chemical concentration gradient exists is known as self-diffusion. The corresponding diffusion coefficient is called the self-diffusion coefficient. In other words, it is a simple form of diffusion where the molecules of the diffusing species are identical to the other molecules in every aspect except for a label that does not affect interaction of the labelled molecules with the others [29, 30]. The self-diffusivity describes the random motion of a molecule in the absence of any gradients that would cause a mass flux. This term is used for the diffusion not only in a pure system of single component but also in a homogeneous multicomponent system. The self-diffusion coefficient is the most simple to interpret among all the other diffusion coefficients [29].

There are two common ways to obtain a self-diffusion coefficient. The first is from molecule positions and the second is from velocities [30, 31]. Theoretically, both methods yield the same result. Obtaining the self-diffusivity from the velocities involve integrating the velocity auto-correlation function, an example from what is called Green-Kubo relations. The description of both these techniques is given in the subsequent section.

2.1.4 Binary-Diffusion

The diffusion of two different species in a binary mixture is called binary diffusion and the corresponding diffusion coefficient is called binary diffusion coefficient [29, 31]. This process is also termed as mutual diffusion, inter-diffusion, chemical diffusion, or transport diffusion. The phenomenological expression for binary diffusion coefficient is given by famous Darkens relation [32].

$$D_{AB} = N_B\, D_A + N_A\, D_B \tag{2.1}$$

where D_{AB} is the mutual diffusion coefficient and D_A and D_B are the self-diffusion coefficients of two species A and B respectively. For a binary system under infinite dilution, the binary diffusion coefficient becomes equal to the self-diffusion coefficient of one of the components [33]. From the applications point of view, mutual diffusion is more important than self-diffusion and plays a major role in many physical and chemical processes. It is characterized by the cooperative motion of the molecules and therefore its theoretical description is more complex than that of the self-diffusion [30].

2.2 Mathematical Interpretation

When mass, energy or momentum is transferred through a system, the transport is described to first order by a phenomenological relation of the form

$$\text{Flux} = -\text{coefficient} \times \text{gradient}$$

The flux measures the transfer per unit area in unit time, the gradient provides the driving force for the flux and coefficient characterizes the resistance to flow. The above equation includes Newton's law of viscosity, Ficks's law of diffusion, Fourier's law of heat conduction and Ohm's law of electrical conduction.

2.2.1 Einstein's relation

In an isotropic medium, physical and chemical properties are independent of direction, whereas in anisotropic media properties depend on the direction considered. Diffusion is isotropic in gases and in most liquids whereas anisotropy in non-cubic crystals and in some quasicrystals [5, 34]. In isotropic materials the

diffusivity is a scalar quantity. Anisotropy affects the directional relationship between the vectors of the diffusion flux and of the concentration gradient [5]. We discuss about the isotropic diffusion.

The mathematical theory of diffusion in isotropic substances is based on the hypothesis that the rate of transfer of diffusing substance through unit area of a section is proportional to the concentration gradient measured normal to the section [34] i.e.

$$\mathbf{J} = -D\nabla C(r,t) \tag{2.2}$$

where **J** is the flux of particles (diffusion flux) and C(r,t) their number density (concentration). The negative sign indicates opposite directions of diffusion flux and concentration gradient. The factor of proportionality, D, is denoted as the diffusion coefficient or as the diffusivity of the species considered. The diffusion coefficient can be self-diffusion or mutual diffusion coefficient which is described below.

Units : The diffusion flux is expressed in number of particles (or moles) traversing a unit area per unit time and the concentration in number of particles per unit volume. Thus, the diffusivity D has the dimension of length^2/time and bears the units [cm^2 s^{-1}] or [m^2 s^{-1}] [5].

The equation (2.2) is also known as the *Fick's first law* and is illustrated in figure (2.1) [5].

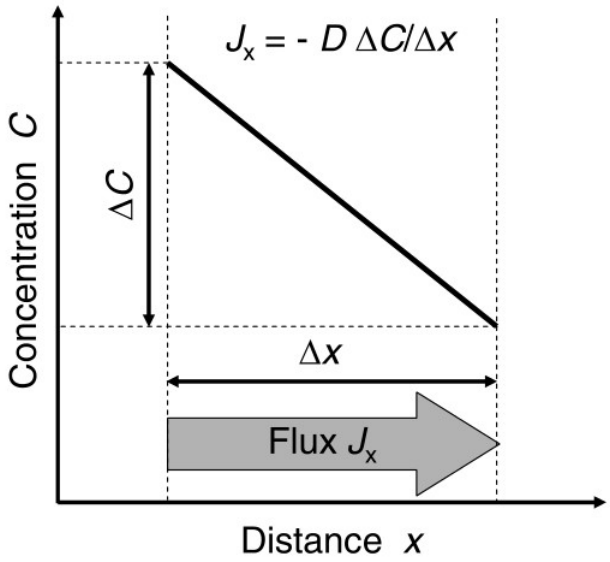


Figure 2.1: Illustration of Fick's first law.

Figure (2.1) explains the Fick's first law. Usually, in diffusion processes the number of diffusing particles is conserved, i.e. the diffusing species neither undergoes reactions nor exchanges with internal sources or sinks. For a diffusing species which obeys a conservation law an equation of continuity can be formulated.

$$\frac{\partial C(r,t)}{\partial t} + \nabla.\mathbf{J}(r,t) = 0 \tag{2.3}$$

The equation (2.3) is the *continuity equation*. Fick's first law (2.2) and the equation of continuity (2.3) can be combined to give an equation which is called *Fick's second law* or sometimes also the *diffusion equation* [15]. Combining equation (2.2) and (2.3) we get

$$\frac{\partial C(r,t)}{\partial t} = \nabla.(D\nabla C) \tag{2.4}$$

From a mathematical view point Fick's second law is a second-order partial differential equation. To determine the diffusion coefficient of the substance it is necessary to determine the concentration profile of the diffusing substance. One of the several ways to determine the concentration profile is under the assumption that at time t=0, the substance was concentrated at the origin of our coordinate frame, i.e. C(r, 0) = $\delta(r)$.

From equation (2.4) we get

$$\frac{\partial C(r,t)}{\partial t} - D\nabla^2 C(r,t) = 0 \tag{2.5}$$

Equation (2.5) can be solved using the boundary condition

$$C(r,0) = \delta(r) \tag{2.6}$$

where $\delta(r)$ is the Dirac delta function. This will yield [33]

$$C(r,t) = \frac{1}{(4\pi Dt)^{\frac{d}{2}}} \exp\left(\frac{-r^2}{4Dt}\right) \tag{2.7}$$

where d is the dimensionality of the system. For a three-dimensional system, as in our case, its value is 3.

To determine the diffusion coefficient the time dependence of second moment of C(r,t) is needed, which is defined as

$$\left\langle r^2(t)\right\rangle \equiv \int d\tau\ C(r,t)\ r^2 \tag{2.8}$$

where C(r,t) satisfies the normalization condition

$$\int d\tau\ C(r,t) = 1 \tag{2.9}$$

The second moment is also defined as mean-squared displacement (MSD) of the diffusing particles. The time evolution of $\langle r^2(t)\rangle$ can be obtained by multiplying equation (2.4) by r^2 and integrating over all space, which yields

$$\int d\tau\ r^2\ \frac{\partial C(r,t)}{\partial t} = \int d\tau\ r^2\ D\nabla.(\nabla C(r,t)) \tag{2.10}$$

Changing the order of derivative and integration, we get

$$\frac{\partial}{\partial t}\int d\tau\ r^2\ C(r,t) = D\int d\tau\ r^2\nabla.(\nabla C(r,t)) \tag{2.11}$$

From equation (2.8), equation (2.11) can be written as

$$\frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle = D\int d\tau\ r^2\ \nabla.(\nabla C(r,t)) \tag{2.12}$$

Since

$$\nabla.(\phi\ \nabla\psi) = \nabla\phi.\nabla\psi + \phi\ \nabla^2\psi \tag{2.13}$$

Equation (2.12) can be written as

$$\frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle = D\int d\tau\nabla.(r^2\ \nabla C(r,t)) - D\ \int d\tau\nabla r^2.\nabla C(r,t) \tag{2.14}$$

Using the Gauss divergence theorem,

$$\int_v d\tau(\nabla.\mathbf{A}) = \oint_s d\mathbf{s}.\mathbf{A} \tag{2.15}$$

in the first part and taking derivative in the second part, we get

$$\frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle = D\oint d\mathbf{s}.(r^2\ \nabla C(r,t)) - 2D\int d\tau\ \mathbf{r}.\nabla C(r,t) \tag{2.16}$$

The surface integral is zero over the surface enclosing the entire space under consideration. Thus, the first term vanishes. Using equation (2.13) in the second integral it is seen that

$$\frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle = -2D\int d\tau\ \nabla.(\mathbf{r}\ C(r,t)) + 2D\int d\tau\ (\nabla.\mathbf{r})\ C(r,t) \tag{2.17}$$

Again the first integral vanishes by divergence theorem and $\nabla.\mathbf{r} = 3$ (for three dimension). Using these, equation (2.17) becomes

$$\frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle = 6D\ \int d\tau\ C(r,t) \tag{2.18}$$

Now, using the normalization condition (2.9), an expression for diffusion coefficient can be obtained as

$$D = \frac{1}{6}\ \frac{\partial}{\partial t}\left\langle r^2(t)\right\rangle \tag{2.19}$$

Equation (2.19) is the famous Einstein's relation which relates the diffusion coefficient, a macroscopic transport property of the system, with the mean-squared displacement, a microscopic quantity of the diffusing particles of the system. As in equation (2.19) the instantaneous value of the diffusion coefficient can be obtained from the slope of the MSD curve with time. For an MSD that behaves as a straight line after a prolonged period of time, equation (2.19) reduces to

$$D = \lim_{t\to\infty}\frac{\left\langle \mathbf{r}^2(t)\right\rangle}{6t} \tag{2.20}$$

2.2.2 Green-Kubo's Formalism

The Green-Kubo's formalism relates macroscopic properties (e.g. the diffusion coefficient, in particular a response property of the system) to microscopic properties (fluctuations of the equilibrium distribution). The diffusion coefficient is a response property of the system to a concentration inhomogeneity. The velocity auto-correlation function is an equilibrium property of the system, because it describes correlations between velocities at different times along an equilibrium trajectory [33].

We have seen from the Einsteins formula for diffusion coefficient equation (2.19) we need to calculate the MSD to get the diffusion coefficient of the particle in the system. If we need to calculate the average value of r^2 called mean square displacement, we need to do the average of $\triangle r^2$ over all the particles at a certain time. $\triangle r^2$ is really how far all these particles have moved.

$$\left\langle \triangle r(t)^2 \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \triangle r_i(t)^2 \tag{2.21}$$

From equation (2.19) we have seen the average of the MSD is equal to two times the diffusion coefficient times the dimensionality of the system. For 3-D,

$$\frac{\partial}{\partial t} \left\langle r^2(t) \right\rangle = 6 \times D \tag{2.22}$$

Let us introduce the position as the integral of velocity,

$$\triangle \mathbf{x}(t) = \int_0^t dt' \ \mathbf{v}_x(t') \tag{2.23}$$

Now, from equation (2.22) the expression average of the mean square displacement is,

$$\left\langle \triangle x(t)^2 \right\rangle = \left\langle \left(\int_0^t dt' \ \mathbf{v}_x(t') \right)^2 \right\rangle \tag{2.24}$$

$$= \int_0^t dt' \int_0^t dt'' \left\langle \mathbf{v}_x(t').\mathbf{v}_x(t'') \right\rangle \tag{2.25}$$

The bracket $\langle\rangle$ denotes the ensemble average on all the particles.

$$= 2 \int_0^t dt' \int_0^{t'} dt'' \left\langle \mathbf{v}_x(t').\mathbf{v}_x(t'') \right\rangle \tag{2.26}$$

We have written the average mean square displacement as an integral of the average product between the velocity at a certain instant and the velocity at different instant. This is where our connection with the equilibrium properties is starting to emerge. This is called velocity autocorrelation function. From equation (2.22) we have seen the Einstein's relation is written as

$$6D = \lim_{t\to\infty} \frac{\partial \langle r^2(t)\rangle}{\partial t} \tag{2.27}$$

When we take the derivative with respect to t one integral cancels out.

$$6D = \lim_{t'\to\infty} 2 \int_0^{t'} dt'' \langle \mathbf{v}_x(t').\mathbf{v}_x(t'')\rangle \tag{2.28}$$

We can use translational invariance in time. We want to look the average value of the product of the velocity at a certain instant t' times the velocity at another instant t'', the average product is not going to be different if we translate it in time, so we can refer it to an arbitrary origin. We shift t'' to 0.

$$6D = \lim_{t'\to\infty} 2 \int_0^{t'} dt'' \langle \mathbf{v}_x(t'-t'').\mathbf{v}_x(0)\rangle \tag{2.29}$$

Hence our final expression for velocity-velocity autocorrelation function can be written as

$$D = \frac{1}{3} \int_0^{\infty} d\tau \langle \mathbf{v}_x(\tau).\mathbf{v}_x(0)\rangle \tag{2.30}$$

2.3 Radial Distribution Function

Radial distribution (pair correlation) functions (RDF) are the primary linkage between macroscopic thermodynamic properties and intermolecular interactions of fluids and fluid mixtures [35]. RDFs are useful way to describe the structure of a system particularly of liquids. RDF also known as pair distribution function, $g(r)$, gives the probability of finding an atom (or molecule, if simulating a molecular fluid) a distance r from another atom (or molecule) compared to the ideal gas distribution, $g(r)$ is thus dimensionless. Higher RDF (e.g. the triplet radial distribution function) can also be defined but are rarely calculated and so references to the radial distribution function are usually taken to mean the

pairwise version and also known as pair distribution function (PDF) [36].

The structure of a liquid the spatial arrangements of its molecules relative to each other is determined mostly by the sudden strong repulsion between molecules when they come close together and that once the structure is thus determined the effects of the much weaker and slowly varying attractive forces can be adequately treated by a simple approximation [37].

The strong repulsion between molecules can often be idealized as those between hard spheres. If so, then the structure of a real liquid (of spherical or nearly spherical molecules) should be nearly the same as that of a hypothetical fluid of non attracting hard spheres of a diameter equal to the real molecules σ (shown in figure (2.2)) at the same number density [37].

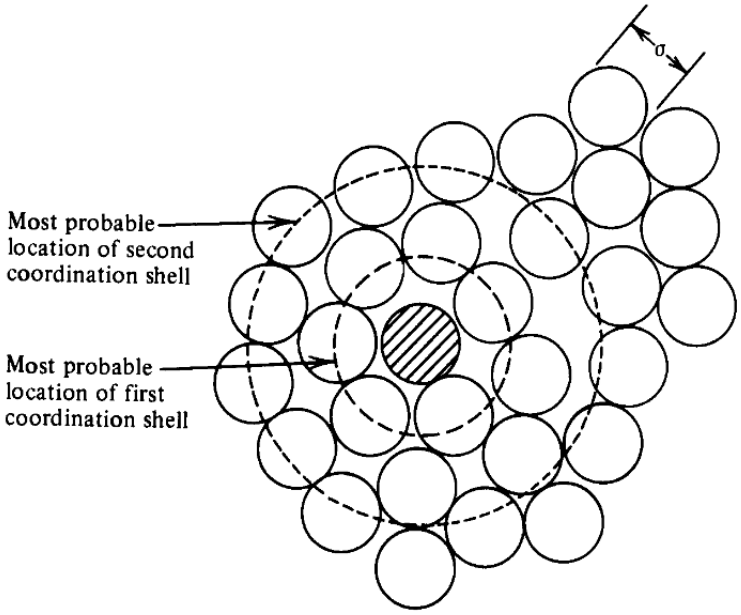


Figure 2.2: Simple liquid Structure.

Figure (2.2) shows the most probable location of first and second coordination shell with respect to a fixed molecule. The definition of $g(r)$ implies that on average the number of particles lying within the range r to $r + dr$ from a reference particle is $4\pi r^2 \rho g(r) dr$ and the peaks in $g(r)$ represents shells of the neighbors around the reference particle. Integration of $4\pi r^2 \rho g(r)$ up to the position of the first minimum therefore provides an estimate of the nearest-neighbor coordina-

tion number [35].

$$\int_0^\infty \rho\, g(r)\, 4\pi\, r^2 dr = N - 1 \approx N \tag{2.31}$$

This probability is not normalized to unity. The function $g(r)$ can also be thought of as the factor that multiplies the bulk density ρ to give a local density $\rho(r) = \rho g(r)$ about some fixed particles.

If the energy of interaction among molecules is additive, so that the total is just the sum of the interactions of all pairs, then this total interaction energy and so ultimately all of the liquids thermodynamic properties including its equation states is given in terms of $g(r)$ [37].

$$E_{\text{intermol}} = \frac{1}{2}\, N\, \rho \int d\tau\, g(r)\, \phi(r) \tag{2.32}$$

We suppose, for simplicity, that the molecules are spherical or nearly so, so that the interaction between them is describable by a spherically symmetric function $\phi(r)$. N is the total number of molecules and ρ is the overall mean density. If we write the above equation in polar form we get,

$$E_{\text{intermol}} = 2\, \pi\, N\, \rho \int_0^\infty dr\, g(r)\, \phi(r)\, r^2 \tag{2.33}$$

In a crystal, the RDF has an infinite number of sharp peaks whose separation and heights are characteristics of the lattice structure. The RDF of a liquid is intermediate between solid and gas with small number of peaks between short distances, superimposed on steady decay to a constant value at longer distances. For short distances, $g(r)$ is zero due to strong repulsive forces. RDF can be measured by X-ray diffraction and can be compared to the RDF calculated from the computer simulation method. By using RDF we can calculate the thermodynamic properties, if pairwise additivity is of the forces is assumed [36]. From the above discussions we can say that the average density $\rho(r)$ at a distance r of a given particle is defined to be $\rho\, g(r)$, i.e.

$$\rho(r) = \rho\, g(r)$$

where ρ is the average density in the fluid and $\rho(r)$ is a conditional density; it is the density at $\mathbf{r}$, given a particle is present in the origin.

The radial distribution function or pair correlation function $g_{AB}(r)$ between particles of type A and B is defined in the following way,

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} \tag{2.34}$$

which can be written as

$$= \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4 \pi r^2} \tag{2.35}$$

where $\langle \rho_B(r) \rangle$ is the particle density of type B at a distance r around particles A and $\langle \rho_B \rangle_{local}$ is the particle density of type B averaged over all the spheres around particles A [22].

2.4 Molecular Dynamics

2.4.1 An Introduction

The molecular dynamics methods were originally devised in the $1950s$, but they only began to receive widespread attention in the mid $1970s$ when the digital computers became powerful and affordable. Today, lots of researchers are interested because of its diverse research areas [38].

Molecular dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms is followed by integrating their equations of motion [39]. In MD, we follow the laws of classical mechanics, and most notably Newton's law.

$$m_i \frac{\partial^2 \mathbf{r_i}}{\partial t^2} = -\nabla_i U(r) = \mathbf{F_i} \quad \text{where} \quad \text{i = 1, 2, . . . , N} \tag{2.36}$$

The right hand side is negative gradient of the potential which is equal to force $\mathbf{F_i}$ acting on the atom and m_i is mass of the atom.

Molecular dynamics is a statistical mechanics method and is a deterministic technique; given an initial set of positions and velocities, the subsequent time evolution is in principle completely determined. In more pictorial terms, atoms will move into the computer, bumping into each other, wandering around (if the

system is fluid), oscillating in waves in concert with their neighbours, perhaps evaporating away from the system if there is a free surface, and so on, in a way pretty similar to what atoms in a real substance would do [39].

The essence of molecular dynamics is simply stated: numerically solve the N-body problem of classical mechanics. Since, the time of Newton, the N-body problem has been viewed as important, but the reasons for its importance have evolved. At the present time, its importance stems from attempts to relate collective dynamics to single-particle dynamics, attempts motivated by the hope that the puzzling behavior of large collections of particles can be explained by examining the motions of individual particle, e.g. how does flow of fluid around an object produce a turbulent wake? How do atoms on a protein molecule move together so the protein folds in life-supporting ways? How do individual molecules combine to form new molecules? Such questions suggest that molecular dynamics can enlighten diverse research areas [38].

Molecular dynamics is basically used for the generation of non-equilibrium ensembles and for the analysis of dynamics events. Molecular dynamics simulation in general calculates the dynamics or non-equilibrium properties of a molecular system such as the viscosity of a liquid, diffusion coefficient, the dynamics of phase changes, reaction kinetics or the dynamics of defects in crystals [38].

The few areas of current interest where MD could bring important contribution are *liquids* (transport phenomena such as viscosity and heat flow can be studied using MD), *Defects* (defects in crystals - crucial for their mechanical properties and therefore of technological interest can be studied using MD), *Fracture* (simulation provide the insight of fracture process that occur in different ways and with different speeds depending of several parameters), *Surfaces* (simulation plays a big role in understanding phenomena such as surface reconstructions, surface melting, faceting, surface diffusion, roughening, etc.), *Friction* (investigations of adhesion and friction between two solids, propelled by the development of the atomic force microscope (AFM)), *Clusters* (metal clusters are extremely important from technological point of view, due to their role as a catalyst in important chemical reactions), *Biomolecules* (MD allows to study the dynamics of large macromolecules including biological systems such as proteins, nucleic acids

(DNA, RNA), membranes) and in *Electronic properties and dynamics* [39].

2.4.2 History

The key papers appeared in the 50's and in the 60's which can be regarded as milestones in molecular dynamics are mentioned below [39].

- The first paper reporting in a molecular dynamics simulation was written by Alder and Wainwright in 1957. The purpose of the paper was to investigate the phase diagram of a hard sphere system, and in particular the solid and liquid regions. In a hard sphere system, particles interact via instantaneous collisions and travel as free particles between collisions. The calculations were performed on a UNIVAC and on an IBM-74.

- The article *Dynamics of radiation damage* by J. B Gibson, A. N Goland, M. Milgram and G. H. Vineyard from Brookhaven National Laboratory appeared in 1960 is probably the first example of a molecular dynamics calculations with a continuous potential based on a finite difference time integration method. The paper deals with the creation of defects induced by radiation damage (a theme appropriate to cold war days) and was performed on an IBM-704.

- Aneesur Rahman at Argonne National Laboratory has been a well known pioneer of MD. In his famous 1964 paper correlations in the motion of atoms in liquid argon he study number of properties of liquid argon.

- Loup Verlet calculated in 1967 the phase diagram of argon using the Lennard-Jones potential and computed correlation functions to test theories of the liquid state.

The overall process of molecular dynamics simulation, where we try to replicate the real physical system in computer readable format so that every phenomenon that occurs at outside world is exactly followed by computer simulation, can be divided into following four major steps [40].

- *Modelling a system*

- *Initialization*

- *Force Calculation*

- *Integrating equation of motion*

2.5 Modelling a System

The main ingredient of a simulation is a model for the physical system. It is the heart of simulation. The small error during this process leads to the large fluctuation in final result which ultimately failures the model. In molecular dynamics simulations we are concerned with the force field, potential, atomic mass, charge, van der Waals constant etc. We have to be careful while preparing topology file which lists the constant attributes of each atom. In addition there are dynamic attributes of atoms their positions, velocities and forces; they do not strictly belong to the molecular topology, and are stored in the coordinate file or trajectory file.

The main art of modelling is the appropriate choice of force field. There are number of force fields like GROMOS, OPLS/AA, AMBER, CHARMM, MARTINI [22]. They show considerable differences in the strategy of parameter specification and the resulting parameters. Nevertheless, the functional forms of these force fields are very similar [40].

In classical force fields the potential functions are derived empirically to describe the atomic interactions. The atoms are treated as spherically symmetric particles and are considered to be connected through covalent bonds to form molecules. Each and every atom experiences a force resulting from its pairwise additive interaction with the rest of the system. The total potential energy U_{total} includes contributions from both bonded interactions and non-bonded interactions. Bonded interactions are based on a fixed list of atoms. They are not exclusively pair interactions, but include 3 and 4-body interactions as well. There are bond stretching (2-body), bond angle (3-body), and dihedral angle (4-body) interactions. A special type of dihedral interaction (called improper dihedral) is used to force atoms to remain in a plane or to prevent transition to a configuration of opposite chirality (a mirror image). Thus, the bonded interaction includes terms for bond stretching U_{bond} , bond-angle bending U_{angle} , bond dihedral U_{dihed} , and out-of-plane distortions alias improper U_{impr} while the non-bonded interactions are represented by the Van der Waals potential U_{vdw} and the

Coulomb potential $U_{coulomb}$. Now the total potential energy of the system is sum of the potential energies due to all above mentioned interactions [36]. Therefore, the total potential energy function of a system can be stated as

$$U_{total} = U_{bond} + U_{angle} + U_{dihed} + U_{impr} + U_{vdw} + U_{coulomb} \tag{2.37}$$

In classical MD studies, the bond stretching, bond angle vibration, and improper dihedral potential energy contributions are approximated in terms of harmonic potentials. However, the potential function for proper bond dihedral is characterized by periodic potentials. Every potential energy function has been defined as below [40].

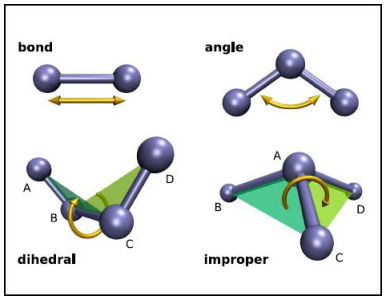


Figure 2.3: Different types of bonding between atoms and atomic planes in a molecule.

2.5.1 Bond Stretching

The bond stretching between two covalently bonded atoms i and j is represented by harmonic potential [22].

$$U_b(r_{ij}) = \frac{1}{2}\, k_{ij}^b\, (r_{ij} - b_{ij})^2 \tag{2.38}$$

and the force is given by

$$\mathbf{F}(\mathbf{r}_{\mathrm{ij}}) = k_{ij}^b\, (r_{ij} - b_{ij})\, \frac{\mathbf{r}_{\mathrm{ij}}}{r_{ij}} \tag{2.39}$$

where b_{ij} is the bond length between two atoms and k_{ij}^b is the force constant.

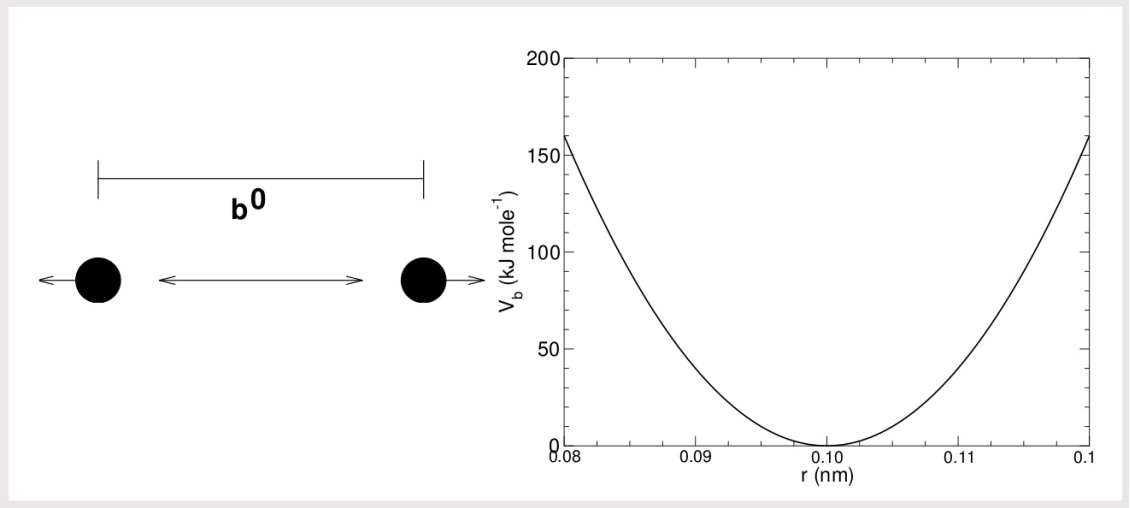


Figure 2.4: Principle of bond stretching and the bond stretching potential for a typical system.

Figure (2.4) represents the bond stretching (left) and the bond stretching potential (right) [22].

2.5.2 Harmonic angle Potential

The bond-angle vibration between a triplet of atoms $i - j - k$ is also represented by a harmonic potential on the angle Θ_{ijk} [22].

$$U_a(\Theta_{ijk}) = \frac{1}{2} K^{\Theta}_{ijk} \left(\Theta_{ijk} - \Theta^0_{ijk}\right)^2 \tag{2.40}$$

As the bond-angle vibration is represented by a harmonic potential, the form is the same as the bond stretching (Figure 2.4). K^{Θ}_{ijk} is the force constant and Θ^0_{ijk} is the equilibrium bond angle.

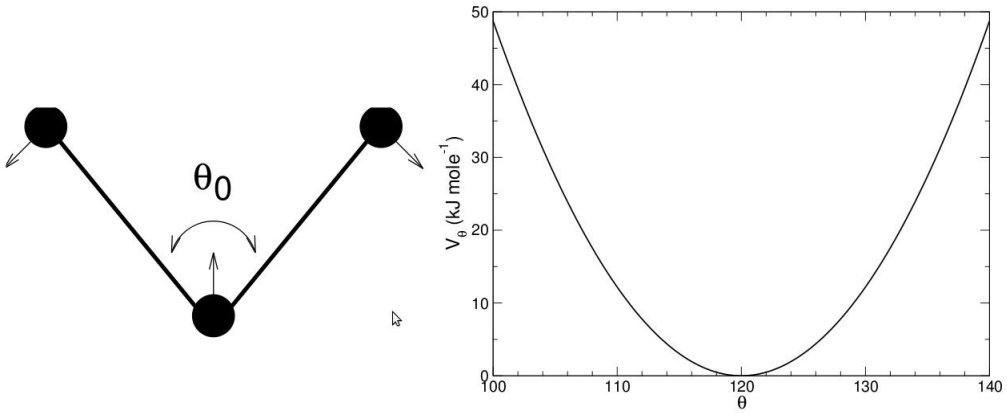


Figure 2.5: Principle of angle vibration and bond angle potential.

The principle of angle vibration is shown in left of figure (2.5) and bond angle potential is shown in right. The numbering i j k is in sequence of covalently bonded atoms. Atom j is in the middle, atoms i and k are at the ends (see figure (2.5)).

2.5.3 Proper Dihedrals

The dihedral is defined by the angle between the two planes determined by atoms $A-B-C$ and $B-C-D$. It measures torsion around bonds. The periodic potential is defined as

$$U_{dihed} = \frac{k_i^{\Phi}}{2}(1 + cos(n\ \Phi_i + \delta_i)) \tag{2.41}$$

Φ_i is dihedral angle, k_i^{Φ} is the barrier height, and δ_i is a reference angle at which the potential is maximum. The periodicity Φ_i counts the number of minima for a full rotation of 360^o. The potential is represented by figure (2.6).

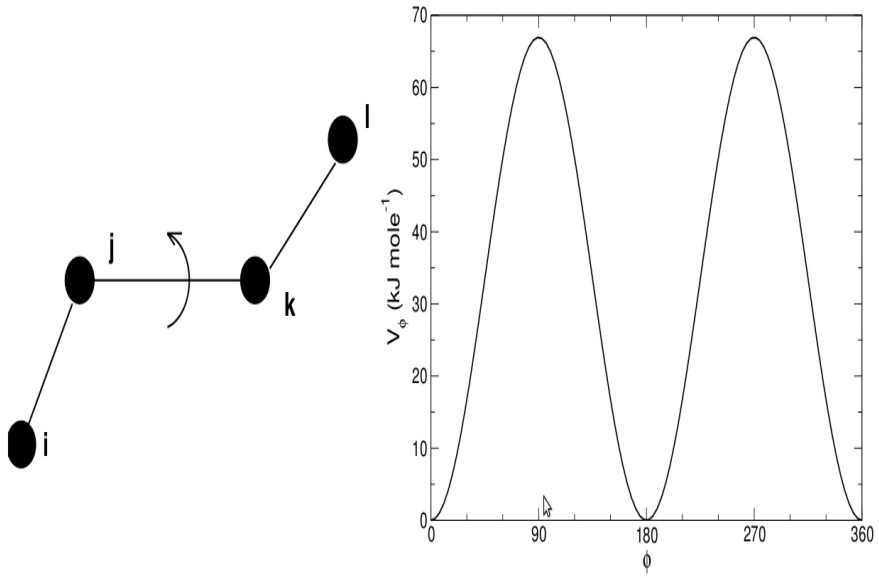


Figure 2.6: Principle of proper dihedral and dihedral angle potential.

2.5.4 Improper Dihedrals

Improper dihedrals are meant to keep planar groups planar, or to prevent molecules from flipping over to their mirror image. The equation describing the improper dihedral potential is as

$$U_{impr} = \frac{1}{2}\, k^{\omega}\, (\omega_{ijkl} - \omega_0)^2 \tag{2.42}$$

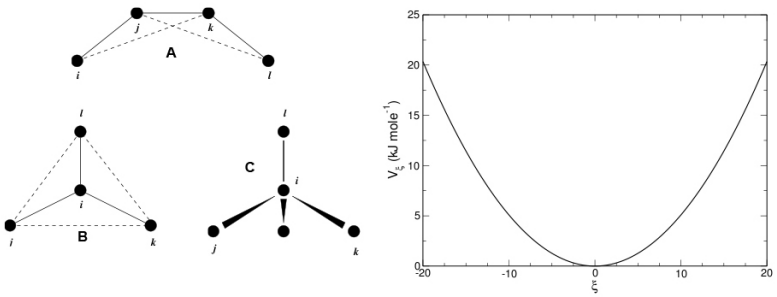


Figure 2.7: Principle of improper dihedral angles.

From figure (2.7) Out of plane bending for rings (A), substituents of rings (B), out of tetrahedral (C). The improper dihedral angle is defined as the angle between planes (i, j, k) and (j, k, l) in all cases. Harmonic improper dihedral potential for improper dihedral for a typical system (right) is shown in figure (2.7).

2.5.5 Lennard - Jone's Interaction

The Lennard-Jone's (L-J) is a non-bonded interaction. The L-J $12-6$ potential is given by the expression.

$$\phi_{LJ}(r) = 4\,\epsilon\,\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.43}$$

for the interaction potential between a pair of atoms. Here ϵ and σ are the Lennard-Jone's constant. The total potential of a system containing many atoms is then given by

$$V(r_1, \ldots, r_N) = \sum_i \sum_{j \geq i} \phi(|\mathbf{r_i} - \mathbf{r_j}|) \tag{2.44}$$

This potential has an attractive tail at large r, it reaches a minimum around 1.122σ and is strongly repulsive at shorter distance, passing through 0 at $r = \sigma$ and increasing steeply as r is decreased further.

The term $1/r^{12}$, dominating at short distance, models the repulsion between atoms when they are brought very close to each other. Its physical origin is related to the Pauli's principle: when the electronic clouds surrounding the atoms start to overlap, the energy of the system increases abruptly. The exponent 12 was chosen exclusively on a practical basis.

The term $1/r^6$, dominating at large distance, constitute the attractive part. This is the term which gives cohesion to the system. A $1/r^6$ attraction is originated by van der Waals dispersion forces, originated by dipole-dipole interactions in turn due to fluctuating dipoles.

These are rather weak interactions, which however dominate the bonding characteristics of the closed-shell systems, that is, rare gases such as Ar or Kr. The parameters ϵ and σ are chosen to fit the physical properties of the material [39].

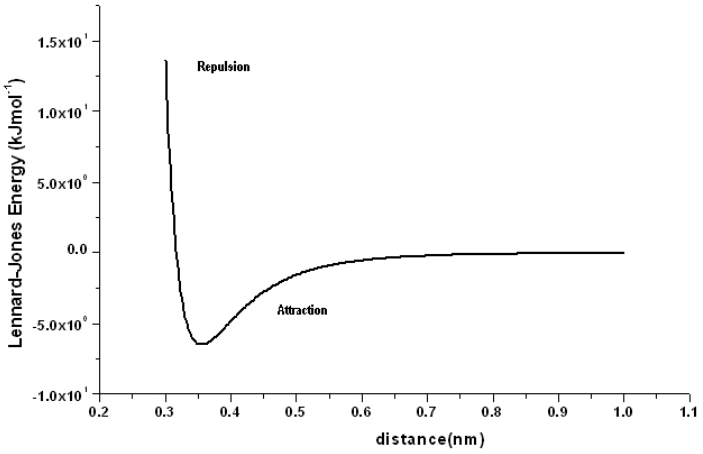


Figure 2.8: Lennard-Jones Potential.

2.5.6 Coulomb Interaction

Another non-bonded interaction that may occur between atoms of same or different molecules is the electrostatic interaction because of presence of some partial charge. The interaction can be defined by Coulomb potential approximated by interaction between two point charge q_i and q_j given below as:

$$U_{coulomb} = \frac{q_i q_j}{4\ \pi\ \epsilon\ \epsilon_0\ r_{ij}} \tag{2.45}$$

ϵ denotes the dielectric constant, ϵ_0 is the permittivity of free space and r_{ij} is the distance between the charges. Since atoms and molecules are not charged, unless they are ions, the atomic charges in this expression are an artificial construct. However, in molecules, the atoms share their valence electrons and the electron density may be shifted due to different electronegativity of the atoms. These different charge densities can be mapped into partial point charges which are represented by q_i and q_j in the expression above [40]. The potential is represented by figure below.

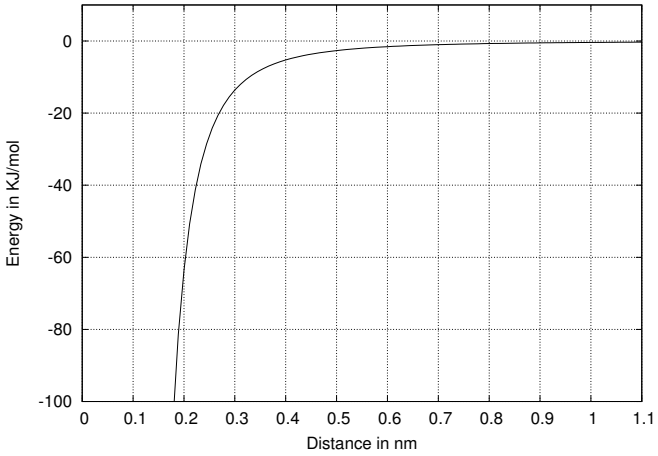


Figure 2.9: Coulomb Potential.

2.6 Simulation Box and Periodic Boundary Conditions

Molecular Dynamics is typically applied to systems containing several hundreds or a few thousand atoms. Such small systems are dominated by surface effects-interactions of the atoms with the container wall [38]. The classical way to minimize edge effects in a finite system is to apply periodic boundary conditions (PBC) [22]. The main purpose of the PBC is to eliminate the surface effects. When using the PBC, particles are enclosed in a box, and we can imagine that this box is replicated to infinity by rigid translation in all the three Cartesian directions, completely filling the space. These replica are called the image cells. In other words, if one of our particle is located at a position $\mathbf{r}$ in the box, we assume that this particle really represents an infinite set of particles located at

$$r + l\mathbf{a} + m\mathbf{b} + n\mathbf{c} \quad (l, m, n = -\infty \ldots \infty)$$

where l, m, n are integer numbers and $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are vectors corresponding to the edges of the box. All these image particles move together, and in fact only one of them is represented in the computer program [39].

In the course of the simulation, as a molecule moves in the original box, its periodic image in each of the neighboring boxes moves in exactly the same way. Thus, as a molecule leaves the central box, one of its images will enter through

the opposite face. There are no walls at the boundary of the central box, and no surface molecules. It is not necessary to store the coordinates of all the images in a simulation but just the molecules in the central box [31]. We follow the minimum image criterion to account the interaction of the particles i.e. Among all the possible images of a particle j, select the closest, to interact with particle i and throw away all the others. In fact only the closest is a candidate to interact and all the others certainly do not. One must always make sure that the box size is at least $2R_c$ (R_c is the cut off distance) along all the directions where PBCs are in effect [39], i.e. R $\leq \frac{1}{2}$ min($|a|,\ |b|,\ |c|$).

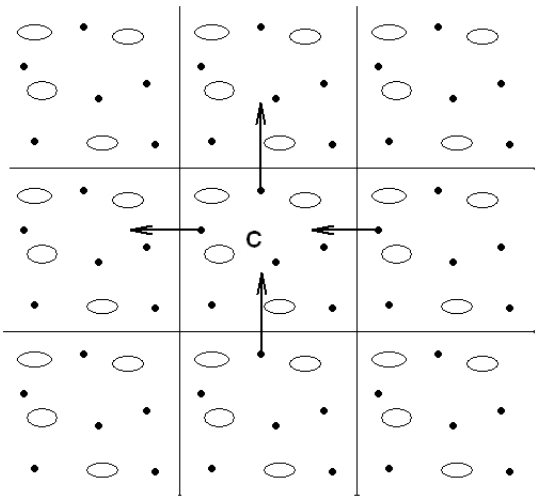


Figure 2.10: Schematic representation of periodic boundary condition.

2.7 Initialization

To start a simulation we need to define the MD box and the set of positions and velocities to assign initially to the particles. If we are starting from a scratch we need to create the set of initial positions and velocities. Positions are usually defined on a lattice, assuming a certain crystal structure. This structure is typically the most stable at T=0 with the given potential. Initial velocities are assigned from a Maxwellian distribution while doing so, the system will have a small total linear momentum, corresponding to a translational motion of the whole system. Since this is inconvenient to have, it is common practice to subtract this component from the velocity of each particle in order to operate in a zero total momentum condition. Another possibility is to take the initial position

and velocities to be the final position and velocities of a previous MD run. This is most commonly used in actual production [39].

Each initial state will not of course correspond to an equilibrium condition however, once the run is started equilibrium is usually reached within a certain time. Initial randomization is usually the only place where chance enters in MD simulation, the subsequent time evolution is completely deterministic [39].

To measure an observable quantity in a MD simulation, we must first of all be able to express this observable as a function of the positions and momenta of the particles in the system [33]. Measuring quantities in MD usually means performing time averages of physical properties over the system trajectory. Physical properties are usually a function of the particle coordinates and velocities.
$A(t) = f(r_1(t), \ldots, r_N(t), v_1(t), \ldots, v_N(t))$ and

$$\langle A \rangle = \frac{1}{N_T} \sum_{t=1}^{N_T} A(t)$$

where N_T is the total number of steps [41]

2.8 Force Calculation

The most time-consuming part of all molecular dynamics simulation is the calculation of the force acting on every particle. In MD forces are derived from a potential energy function V, which depend on the particle coordinates $\mathbf{F_i} = -\nabla V(r_1, \ldots, r_N)$. The problem of modelling a material can therefore be concentrated on finding a potential function $V(r_1, \ldots, r_N)$ for that material.

If we consider a model system with pairwise additive interaction we have to consider the contribution to the force on particle i due to all its neighbors. If we consider only the interaction between a particle and the nearest image of another particle, this implies that, for a system of N particles, we must evaluate $N \times (N-1)/2$ pair distances. This implies that if we use no tricks, the time needed for the evaluation of the forces scales as $O(N^2)$ (Order of N^2). There exists efficient techniques to speed up the evaluation of both short range and

long range force in such a way that the computing time scales as O(N), rather than $O(N^2)$. We first compute the current distance in x, y, z directions between each pair of particles i and j. We then use the cut off at a distance r_c so chosen that to be less than half the length of the periodic box. In that case we can always limit the evaluation of the intermolecular interactions between i and j to the interaction between i and the nearest periodic image of j thus using the periodic boundary conditions [33]. The minimum image criterion is shown in figure below.

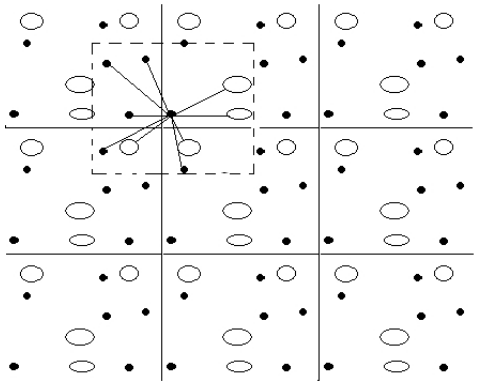


Figure 2.11: Use of cut-off and minimum image criterion for pbc.

Whenever, the interaction beyond r_c is significant, long range correction should be done. This correction is usually based on the concept in which the field beyond r_c is constant (mean field approximation). In the case of Lennard-Jones interaction the long-range correction term is given as [40]

$$U_{LR} = 2\pi N\rho \int_{r_c}^{\infty} dr\ r^2 U_{L-J} \tag{2.46}$$

i.e.

$$U_{LR} = \frac{8}{3}\left[\frac{1}{3}\left(\frac{\sigma_{ij}}{r_c}\right)^9 - \left(\frac{\sigma_{ij}}{r_c}\right)^3\right] \tag{2.47}$$

Another famous long-range correction method is called Ewald summation in which all interactions between the target particle and the replica unit cells surrounding the central unit cell are integrated.

2.9 Integration of Equation of Motion

This step involves calculation of phase space coordinates in order to generate trajectory of the molecules in the system. Both the position coordinates and

velocity components are calculated with the help of known the initial position coordinates and the velocity components. This calculation is done after all forces between the particles have been calculated. Several algorithms have been developed to perform this task. A short list of desirable qualities for a successful simulation algorithm might be as follows [31].

a. It should be fast and require little memory.

b. It should permit the use of long time step δt.

c. It should duplicate the classical trajectory as closely as possible.

d. It should satisfy the known conservation laws for energy and momentum and be time-reversible.

e. It should be simple in form and easy to program.

It is rather difficult to manage all the mentioned points as, the lesser the time step, greater is the precision of classical trajectory and more will be the computational time. The accuracy and stability of the simulation algorithm are measured by its local and global truncation errors. We now mention some specific examples of algorithm that are common in use.

2.9.1 Verlet Algorithm

This is the most widely used method of integrating the equations of motion. This method is based on positions $\mathbf{r}(t)$, accelerations $\mathbf{a}(t)$ and the positions $\mathbf{r}(t-\delta t)$ from the previous step.

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{\mathbf{F}(t)}{2m}\delta t^2 + \frac{\delta t^3}{3!}\frac{d^3r}{dt^3} + O(\delta t^4) \tag{2.48}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{\mathbf{F}(t)}{2m}\delta t^2 - \frac{\delta t^3}{3!}\frac{d^3r}{dt^3} + O(\delta t^4) \tag{2.49}$$

Now adding equations (2.48) and (2.49) we get

$$\mathbf{r}(t+\delta t) + \mathbf{r}(t-\delta t) = 2\mathbf{r}(t) + \frac{\mathbf{F}(t)}{m}\delta t^2 + O(\delta t^4) \tag{2.50}$$

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \frac{\mathbf{F}(t)}{m}\delta t^2 + O(\delta t^4) \tag{2.51}$$

From equation (2.51) it is seen that the estimate of the new position contains an error of order δt^4. The velocities are not needed to compute the trajectories but

they are useful for estimating the kinetic energy (and hence the total energy). To obtain the velocities subtract equation (2.49) from (2.48)

$$\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t) = 2\mathbf{v}(t)\delta t + O(\delta t^3) \tag{2.52}$$

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t} - O(\delta t^2) \tag{2.53}$$

This expression for velocity is accurate to order of δt^2. Now after new positions have been computed the positions at time $t - \delta t$ can be discarded. The current position become the old position and new position become the current position. It is seen that more accurate estimates of $\mathbf{v}(t)$ can be made, if more variables are stored, but this adds the inconvenience seen in the equation (2.53) that $\mathbf{v}(t)$ can only be computed once $r(t + \delta t)$ is known [31].

2.9.2 Leap-Frog Algorithm

There are several algorithms which are equivalent to the Verlet scheme. The simplest among these is the so-called leap-frog algorithm [33]. This algorithm evaluates velocities at half-integer time steps and uses these velocities to compute the new positions. To derive leap-frog algorithm from Verlet algorithm, we start by defining the velocity at half-integer time steps as follows.

$$\mathbf{v}\left(t - \frac{\delta t}{2}\right) \equiv \frac{\mathbf{r}(t) - \mathbf{r}(t-\delta t)}{\delta t} \tag{2.54}$$

and

$$\mathbf{v}\left(t + \frac{\delta t}{2}\right) \equiv \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t)}{\delta t} \tag{2.55}$$

From the latter equation we can immediately obtain the expression for the new positions from old positions and velocities.

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\delta t}{2}\right)\delta t \tag{2.56}$$

Also using Verlet algorithm, the expression for the velocity based on the old velocity can be obtained. Using Taylor series expansion on velocity about t we get,

$$\mathbf{v}\left(t + \frac{\delta t}{2}\right) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{2m}\delta t \tag{2.57}$$

Again,

$$\mathbf{v}\left(t - \frac{\delta t}{2}\right) = \mathbf{v}(t) - \frac{\mathbf{F}(t)}{2m}\delta t \tag{2.58}$$

Subtracting the above two equations we get,

$$\mathbf{v}\left(t + \frac{\delta t}{2}\right) = \mathbf{v}\left(t - \frac{\delta t}{2}\right) + \frac{\mathbf{F}(t)}{m}\delta t \tag{2.59}$$

As the leap-frog algorithm is derived from the Verlet algorithm, it gives rise to identical trajectories. However, the velocities are not defined at the same time as the positions. As a consequence, kinetic and potential energy are also not defined at the same time, and hence total energy cannot be computed directly in the leap-frog algorithm. Figure below gives a schematic representation of the algorithm.

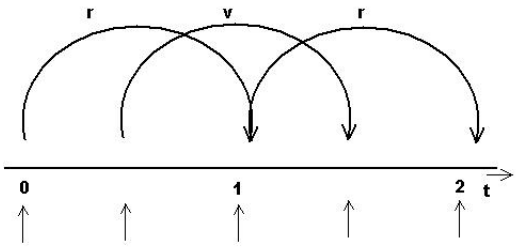


Figure 2.12: Schematic representation of leap-frog algorithm.

The stored quantities during this process are the current positions $\mathbf{r}(t)$ and accelerations $\mathbf{a}(t)$ together with the mid-step velocities $\mathbf{v}\left(t - \frac{1}{2}\delta t\right)$. From equation(2.59) it is seen that velocity leap over the coordinates to give the next mid step values $\mathbf{v}\left(t + \frac{1}{2}\delta t\right)$ hence the name leap-frog algorithm. During this step, the current velocities may be calculated as

$$\mathbf{v}(t) = \frac{1}{2}\left(\mathbf{v}\left(t + \frac{1}{2}\delta t\right) + \mathbf{v}\left(t - \frac{1}{2}\delta t\right)\right) \tag{2.60}$$

2.10 Constraint Dynamics

In polyatomic systems, it becomes necessary to consider not only the stretching of interatomic bonds, but also bending motions, which change the angle between bonds and twisting motions, which alter torsional angles. These effects must be

treated properly in MD, within the classical approximation. It would be quite unrealistic to assume total rigidity of such a molecule, although bond lengths can be thought of as fixed. A special technique has been developed to handle the dynamics of a molecular system in which certain arbitrarily selected degrees of freedom (such as bond lengths) are constrained, while other remains free to evolve under the influence of intermolecular and intramolecular forces. Its great appeal is that it reduces even a complex polyatomic liquid simulation to the level of difficulty of an atomic calculation plus a constraint package based on molecular geometry [31].

2.10.1 SHAKE Algorithm

SHAKE Algorithm is the algorithm for the constraint dynamics simulation. Polyatomic systems may be considered to have internal constraints such as interatomic bonds, bond angles, proper dihedrals, improper dihedrals, etc,. That is, the equilibrium bond lengths, bond angles should remain unchanged throughout the simulation. For such systems, the standard leap-frog method of calculation of new positions and velocities then may not be sufficient. Therefore, a set of independent generalized coordinates must be constructed to obey the constraint-free equations of motion. In these equations the constraints appear implicitly. SHAKE is a method that deals with leap-frog algorithm with internal constraints. In constraint dynamics and hence shake algorithm is used to solve the equations of motion simultaneously satisfying the imposed constraints.

SHAKE uses Cartesian coordinates for each of the atoms to describe the configuration of molecules with internal constraints. It is a procedure that approaches internal constraints by going through the constraints one by one, cyclically, adjusting the coordinates to satisfy each constraints in turn. The procedure is repeated until all constraints are satisfied to within a specified tolerance level [36].

The SHAKE algorithm changes a set of unconstrained coordinates $\mathbf{r}'$ to a set of coordinates $\mathbf{r}''$ that fulfill a list of distance constraints, using a set $\mathbf{r}$ reference as

$$SHAKE(\mathbf{r}' \rightarrow \mathbf{r}''; \mathbf{r}) \tag{2.61}$$

This action is consistent with solving a set of Lagrange multipliers in the con-

strained equations of motion. SHAKE needs a relative tolerance, it will continue until all constraints are satisfied within that relative tolerance.

Let us assume the equations of motion must fulfill K holonomic constraints, expressed as

$$\sigma_k(r_1, \ldots, r_N) = 0; \quad where \quad k = 1, \ldots, K \tag{2.62}$$

For example $(r_1 - r_2)^2 - b^2 = 0$ Then the forces are defined as

$$-\frac{\partial}{\partial r_i}\left(V + \sum_{k=1}^{K} \lambda_k \, \sigma_k\right) \tag{2.63}$$

where λ_k are Lagrange multiplier which must be solved to fulfill the constraint equations. The second part of this sum determines the constraint forces $\mathbf{G}_i$, defined by

$$\mathbf{G}_i = -\sum_{k=1}^{K} \lambda_k \frac{\partial \sigma_k}{\partial r_i} \tag{2.64}$$

The displacement due to the constraint forces in leap-frog or Verlet algorithm is equal to $(\mathbf{G}_i/m_i)(\delta t)^2$. The solutions of a set of coupled equations of the second degree is required in solving the Lagrange multipliers and hence the displacements. These are solved iteratively by SHAKE within a specified tolerance level [22].

2.11 Temperature Calculation and Control

The temperature T is directly related to the kinetic energy by the well-known equipartition formula, assigning an average kinetic energy $k_B T/2$ per degree of freedom.

$$K = \frac{3}{2} N \, k_B \, T \tag{2.65}$$

An estimate of the temperature is therefore directly obtained from the average kinetic energy K. For practical purposes, it is also common practice to define an instantaneous temperature T(t), proportional to the instantaneous kinetic energy K(t), by the relation analogous to the one above.

$$T(t) = \frac{2}{(3N - N_c)k_B} \sum_{i}^{N} \frac{1}{2} m_i \, v_i^2(t) \tag{2.66}$$

where N and N_c represents the number of particles and number of constraints on the system respectively. m_i represents the mass of the ith particle and k_B is the Boltzmaan's constant. The average temperature of the system is given by the ensemble average or time average of the instantaneous temperature as

$$T = \langle T(t) \rangle = \left\langle \frac{2}{(3N - N_c)k_B} \sum_{i}^{N} \frac{1}{2} m_i \, v_i^2(t) \right\rangle \tag{2.67}$$

A constant temperature simulation is required if we wish to determine how the behaviour of the system changes with the temperature such as the unfolding of a protein or glass formation [15].

The initial velocities are generated randomly according to the Maxwell-Boltzmann distribution at the desired temperature, but the velocity distribution does not remain constant as the simulation continues due to various factors. To maintain the desired temperature, the computed velocities need to be adjusted. The temperature-control mechanism should not only maintain the temperature to the right target but also produce the correct statistical ensembles. Different methods are developed for temperature control.

2.11.1 Direct Velocity Rescaling

The temperature of the system is related to the time average of kinetic energy as in equation (2.66). An obvious way to alter the temperature of the system is thus to scale the velocities. If the temperature at time t is $T(t)$ and if the velocities are multiplied by a factor λ, then the associated temperature change can be calculated as

$$\begin{aligned} \Delta T &= \frac{1}{2} \sum_{i=1}^{N} \frac{2}{3} \frac{m_i (\lambda v_i)^2}{N \, k_B} - \frac{1}{2} \sum_{i=1}^{N} \frac{2}{3} \frac{m_i \, v_i^2}{N \, k_B} \\ \Delta T &= (\lambda^2 - 1) T(t) \\ \lambda &= \sqrt{\frac{T_{new}}{T(t)}} \end{aligned} \tag{2.68}$$

The simplest way to control the temperature is thus to multiply the velocities at each time step by the factor $\lambda = \sqrt{T_{req}/T_{curr}}$, where T_{curr} is the current temperature as calculated from the kinetic energy and T_{req} is the desired temperature [15].

2.11.2 Berendsen Scheme

Berendsen method is refined approach to velocity rescaling [40]. In this scheme the system is coupled to an external heat bath, fixed at desired temperature, to maintain the temperature. The bath acts as a source of thermal energy, supplying or removing heat from the system as appropriate. The velocities are scaled at each step, such that the rate of change of temperature is proportional to the difference in temperature between the bath and the system.

$$\frac{dT}{dt} = \frac{1}{\tau}(T_{bath} - T(t)) \tag{2.69}$$

τ is a coupling parameter whose magnitude determines how tightly the bath and the system are coupled together. This method gives an exponential decay of the system towards the desired temperature. The change in temperature between successive time steps is

$$\Delta T = \frac{\delta t}{T}(T_{bath} - T(t)) \tag{2.70}$$

The scaling factor for the velocities is thus

$$\lambda^2 = 1 + \frac{\delta t}{T}\left(\frac{T_{bath}}{T(t)} - 1\right) \tag{2.71}$$

If τ is large, then the coupling will be weak. If τ is small, the coupling will be strong and when the coupling parameter equals the time step ($\tau = \delta t$) then the algorithm is equivalent to the simple velocity scaling method. The advantage of this approach is that it does permit the system to fluctuate about the desired temperature [15].

2.11.3 Nose-Hoover Temperature Coupling

The Berendsen weak-coupling algorithm is extremely efficient for relaxing a system to the target temperature, but once the system has reached equilibrium it might be more important to probe a correct canonical ensemble. This is unfortunately not the case for the weak-coupling scheme [22].

To enable canonical ensemble simulations, the extended-ensemble approach was first proposed by Nose and later modified by Hoover. The system Hamiltonian is extended by introducing a thermal reservoir and a friction term in the equations

of motion. The friction force is proportional to the product of each particles velocity and a friction parameter ξ. This friction parameter (or heat bath variable) is a fully dynamic quantity with its own equation of motion; the time derivative is calculated from the difference between the current kinetic energy and the reference temperature [36].

In Hoover's formulation, the particles equations of motion are replaced by

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \xi \frac{dr_i}{dt} \tag{2.72}$$

where the equation of motion for the heat bath parameter ξ is

$$\frac{d\xi}{dt} = \frac{1}{Q}(T_{cur} - T_{tar}) \tag{2.73}$$

The target temperature is denoted T_{tar} , while T_{cur} is the current instantaneous temperature of the system. The strength of the coupling is determined by the constant Q (usually called the mass parameter of the reservoir) in combination with the reference temperature. To maintain the coupling strength, Q would have to be changed in proportion to the change in reference(target) temperature. For this reason, it is preferable to use τ_c which is the period of the oscillations of kinetic energy between the system and the reservoir. It is directly related to Q and T_{tar} via

$$Q = \frac{\tau_c^2 T_{tar}}{4\pi^2} \tag{2.74}$$

A much more intuitive way of selecting the Nose-Hoover coupling strength similar to the Berendsen method is provided by this relation. Moreover, τ_c is independent of system size and reference temperature [22].

2.12 Pressure Calculation and Control

In the same spirit as the temperature coupling, the system can also be coupled to a pressure bath. GROMACS supports both the Berendsen algorithm that scales coordinates and box vectors every step, the extended ensemble Parrinello-Rahman approach,and for the velocity Verlet variants, the Martyna-Tuckerman-Tobias-klein (MTTK) implementation of pressure control [22].

Pressure is a tensor which can be defined by means of a 3×3 square matrix as

$$\begin{bmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{bmatrix}$$

Each element of the pressure tensor is the force acting on the surface of an infinitesimal cubic volume that has edges parallel to the x, y, and z axes. The first subscript denotes the normal direction to the plane on which the force acts, and the second one denotes the direction of that force.

2.12.1 Berendsen Scheme

The Berendsen algorithm rescales the coordinates and box vectors every step with a matrix μ, which has the effect of a first-order kinetic relaxation of the pressure towards a given reference pressure P_{tar}. The effect of this method is that deviation of the current pressure P_{cur} of the system from the target pressure P_{tar} is slowly corrected in accordance with the relation

$$\frac{d\mathbf{P}_{cur}}{dt} = \frac{\mathbf{P}_{tar} - \mathbf{P}_{cur}}{\tau_p} \tag{2.75}$$

This means that a pressure deviation decays exponentially with a time constant τ_p.

The scaling matrix μ is given by

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p}\, \beta_{ij}[(P_{tar})_{ij} - (P_{cur})_{ij}(t)] \tag{2.76}$$

where β is the isothermal compressibility of the system. In most cases, will be a diagonal matrix, with equal elements on the diagonal, the value of which is generally not known. It suffices to take a rough estimate because the value of β only influences the non-critical time constant of the pressure relaxation without affecting the average pressure itself [22].

2.13 Water Models

Water is the subject of numerous studies due to its biological, chemical significance and its universal presence. It is widely known that water (H_2O) consists

of two hydrogen atoms (H) and one oxygen atom (O) held together by electron pairing, covalent bonds. This leaves the oxygen with a further four unpaired electrons which are arranged to two lone-pair bonds and form an almost perfect tetrahedron. The angle between the two hydrogen atoms, as well as the distance between the hydrogen atoms and the oxygen atom, depend on the isotopes used and the surrounding environment. Naturally occurring water usually consists of a mixture of the different isotopes of its constituents, for example heavy water, whose structure includes two deuterium atoms.

The thermodynamic behavior of water presents important differences compared with those of the other substances, and many of the characteristics of such behavior are often attributed to the existence of hydrogen bonds between water molecules [42]. The water models differ from each other in many aspects, mainly on the way they treat the covalent bonds of the molecule. Water models can be conveniently divided into three types. In the simple interaction site models each water molecule is maintained in a rigid geometry and the interaction between molecules is described using pairwise Coulomb and L-J expressions. Flexible model permits internal changes in conformation of the molecule. Finally, models have been developed that explicitly include the effects of polarization and many body effects. The simple water models use between three and five interaction sites and a rigid water geometry [15]. Each model is good for a particular aspect, some are simple and some are really complicated resulting to more time consuming programming procedure.

The most commonly used water models TIP3P (Transferable Intermolecular Potential 3P), SPC (Simple Point Charge), SPC/E (Extended Simple Point Charge) have three interaction sites for the electrostatic interactions and are similar in nature but the Lennard-Jones (L-J) and Coulombic terms differ and give significant differences in calculated bulk properties for liquid water. In this model the partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge located on the oxygen atom. The van der Waals interaction between two water molecules is computed using a Lennard-Jones function with just a single interaction point per molecule centered on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated.

The four site models such as that of Bernal and Fowler (BF) and TIP4P model shift the negative charge from oxygen atom to a point along the bisector of the HOH angle towards hydrogens. The most commonly used five site model is the ST2 potential of Stillinger and Rahman. Here, charges are placed on the hydrogen atoms and on two lone pair sites on the oxygen. Some simple water models are shown in figure (2.13) [15].

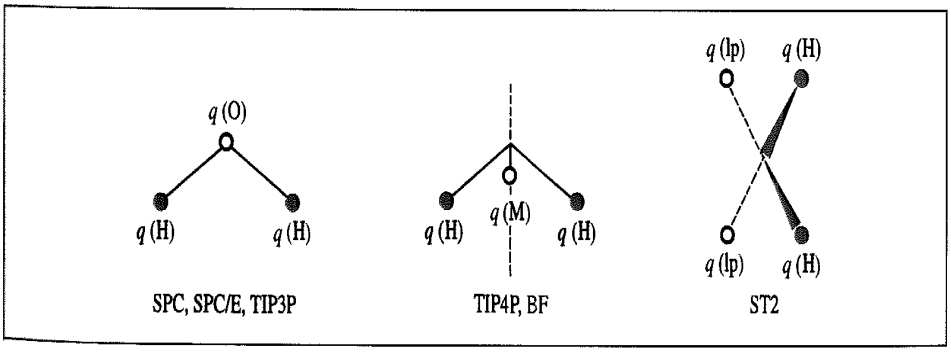


Figure 2.13: Some simple water models.

2.14 Different Software Package for Molecular Dynamics

A few molecular dynamics software packages are available for life science research and simulations. Different softwares have different features and their own merits. These features may not be sufficient to fulfill the need of the simulation problem. Some of them are

- GROMACS (Groningen Machine for Chemical Simulation)
- AMBER (Assisted Model Building and Energy Refinement System)
- CHARMM (Chemistry at Harvard Macromolecular Mechanics)
- LAMMPS (Large-Scale Atomic/Molecular Massively Parallel Simulator)

2.14.1 GROMACS

GROMACS (Groningen Machine for Chemical Simulations) has been developed by Department of Biophysical Chemistry at Groningen University in the Nether-

lands. It is an engine to perform molecular dynamics simulations and energy minimization. It is primarily designed for biochemical molecules like proteins and lipids that have many complicated bonded interactions, but since it is extremely fast at calculating the non-bonded interactions (that usually dominate simulations) it is also used for research on non-biological systems, for example polymers, gases, liquids and so on [40]. The main features of GROMACS are as described below.

1. **User-Friendly:** GROMACS is user-friendly, with topologies, parameter files, and error messages written in clear text format. There is a lot of consistency checking, but no scripting language; all programs use a simple interface with command line options for input and output files. Help files can be accessed through GROMACS directly or via the extensive user manual available for free. When simulations are running, GROMACS can output its progress, giving details of the time and date it expects to finish.

2. **High Compatibility:** GROMACS does not have a force field of its own, but it is compatible with GROMOS, OPLS, AMBER, and ENCAD force fields. Interfaces with popular quantum-chemical packages such as MOPAC, GAMESUK, and GAUSSIAN are provided to perform mixed MM/QM simulations.

3. **Versatility:** The aim of GROMACS is to provide a versatile and efficient MD program with source code, especially directed towards the simulation of biological macromolecules in aqueous and membrane environments. It is able to run on single processors as well as on parallel using standard message passing interface (MPI) communication. It provides not only microcanonical Hamiltonian mechanics, but also stochastic dynamics (SD) including Langevin and Brownian dynamics and energy minimization(EM).

4. **Fastest and Flexible Software:** The highly optimized code makes GROMACS the fastest software for molecular simulations to date. It is faster

in calculating the nonbonded interactions which is dominant in any simulation. Besides, the support of different force fields and the open source (GPL) character make GROMACS very flexible.

5. **Open Source Software:** Unlike AMBER and CHARMM, GROMACS is freely available and can be downloaded from the Internet easily and compiled on any Linux-based operating system. It is available from www.gromacs.org for free under the Groningen University General Public Licence.

These are the main features which make GROMACS so popular and widely used software. Due to these features we have also used this software for our research.

2.15 Limitations of Molecular Dynamics

Molecular Dynamics is a powerful technique but has of course certain limitations. Some of them are described below.

2.15.1 Use of Classical Force

We know that systems at the atomistic level obey quantum laws rather than the classical laws. How can we use Newtons law to move atoms? A simple test of the validity of the classical approximation is based on the de-broglies thermal wavelength defined as

$$\Lambda = \sqrt{\frac{2\ \pi\ \hbar^2}{M\ \mathrm{k}_B\ T}} \tag{2.77}$$

where M is the atomic mass and T is the absolute temperature. The classical approximation is justified if $\Lambda << a$, a is the nearest neighbor separation. Moreover, quantum effects become important in system when T is sufficiently low. Molecular dynamics results should be interpreted with caution in these regions.

2.15.2 Realism of force

In MD, atoms interact with each other. These interactions originate forces which act upon atoms, and atoms move under the action of these instantaneous forces.

As the atoms move, their relative positions change and forces change as well. The essential ingredient containing the physics is therefore constituted by the forces. A simulation is realistic - that is, it mimics the behavior of the real system, only to the extent that interatomic forces are similar to those that real atoms (or, more exactly, the nuclei) would experience when arranged in the same configuration. As the forces are usually obtained as the gradient of a potential energy function depending on the position of the particles. The realism of simulation therefore depends on the ability of the potential chosen to reproduce the behavior of the material under the conditions at which the simulation is run.

2.15.3 Time and Size limitation

Typical MD simulations can be performed on systems containing thousands, or perhaps, millions of atoms, and for simulation times ranging from a few picoseconds to hundred nanoseconds.

A simulation is safe from the point of view of its duration when the simulation time is much larger than the relaxation time of the quantities we are interested in. However, different properties have different relaxation times. In particular, systems tend to become slow and sluggish in the proximity of phase transitions, and it is not uncommon to find cases where the relaxation time of a physical property is orders of magnitude larger than time achievable by simulation. A limited system size can also constitute a problem. In this case one has to compare the size of the MD cell with the correlation lengths of the spatial correlation functions of interest.

However, with time, different simulation softwares like GROMACS have been improved to solve the problems associated with the above limitations. Molecular dynamics, therefore, will be an eminent revolutionary aspect of research and development of molecular science.

Chapter 3

Details Of Simulation

3.1 General Consideration

In this chapter the procedure for the simulation process has been explained in detail. We explain the modeling of carbon monoxide and water, and the sample of the necessary files needed during simulation is provided. A force field contains a large number of parameters even if it is intended for calculations on only a small set of molecules. Parametrization of a force field is not a trivial task. A significant amount of effort is required to create a new force field entirely from scratch, and even the addition of a few parameters to an existing force field in order to model a new class of molecules can be complicated and time-consuming task. We explain the details we follow while adding the parameters in the existing force field inherent in GROMACS.

3.2 Modeling Carbon Monoxide

It consists of one carbon atom and one oxygen atom, connected by a triple bond that consists of two covalent bonds as well as one dative covalent bond. The bond is polar due to difference in electronegativity of the constituent atoms, and there is partial charge in the molecule.

3.2.1 Bonded interaction

In our force field, the covalent bond formed due to overlapping of the electron cloud between carbon and oxygen atoms is approximated by a harmonic potential

of the type [22]

$$V_{12}(r) = \frac{1}{2}K_b(r_{12} - b_0)^2 \tag{3.1}$$

where K_b is the spring constant of the interaction which measures the strength of the bonding and b_0 is the equilibrium distance between two atoms 1 and 2 [22]. Since, there are only two atoms in a carbon monoxide molecule, the contribution from bond angle, dihedral angle, improper dihedral angle don't make any sense. So, we need not consider them in our model.

3.2.2 Non-Bonded Interaction

In our system there are more than one CO molecule. For more than one CO molecule we need to consider the non-bonded Lennard-Jones interaction (L-J interaction) and Coulomb interaction (developed due to polar nature of CO molecule). The L-J interaction is of the following type according to our model [22].

$$U_{ij}(r_{ij}) = 4\ \epsilon_1 \left[\left(\frac{\sigma_1}{r_{ij}} \right)^{12} - \left(\frac{\sigma_1}{r_{ij}} \right)^{6} \right] \tag{3.2}$$

In GROMACS format

$$U_{ij}(r) = \left(\frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^{6}} \right) \tag{3.3}$$

where $C_{12} = 4\epsilon_1\sigma_1^{12}$ and $C_6 = 4\epsilon_1\sigma_1^6$. In equation (3.3), r_{ij} is the distance between two atoms i and j; $2^{1/6}\sigma_1$ is the equilibrium distance between two atoms also called van der Waals radius, and ϵ_1 is the minimum energy depth which measures the strength of the interaction.

Due to different partial charge in atoms of CO molecule there also exists Coulomb potential between two different atoms of different CO molecules expressed as;

$$U_{Coul} = \sum_{\alpha}\sum_{\beta} \frac{q_\alpha(i)\ q_\beta(j)}{4\pi\epsilon_o r_{\alpha\beta}(i,j)} \tag{3.4}$$

The total non-bonded interaction is due to Lennard-Jones and Coulomb contribution given as

$$U_{i,j} = \left(\frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^{6}} \right) + \sum_{\alpha}\sum_{\beta} \frac{q_\alpha(i)\ q_\beta(j)}{4\pi\epsilon_o r_{\alpha\beta}(i,j)} \tag{3.5}$$

The parameters used to parametrize carbon monoxide molecule is presented in table ([43])

Table 3.1: Parameters of carbon monoxide.

$\sigma(A^o)$	$\epsilon\ k_B\ (K)$	bond-length(A^o)	partial charge (δq)
C : 3.49	C: 22.8	1.128	C: 0.0203 e
O : 3.13	O : 63.5		O: -0.0203 e
C-O : 3.31	C-O : 38.05		

The parameters used in the modeling of CO is mentioned in table (3.1) [43]. The value of force constant used is 1.1142×10^6 KJ mol^{-1} nm^{-2} [44]. In GROMACS we use C_6 and C_{12} instead of σ and ϵ, in order to make our calculation easier, so we have to change the value accordingly.

We have defined the forces involved in the model of CO. Now, we need to explain about the files we prepare to include above mention parameters. We need to add an atom and its relevant parameters to an atomic parameter *(.atp)* file, named $ffG43a1.atp$. $ffG43a1$ represents the force field we have used for our system. The extension *.atp* stands for atom type parameter. The type of atoms and relevant parameters entered into the file $ffG43a1.atp$ as

```
Atom    Mass (in amu)
 OB        15.9994       ; Bare Oxygen
 CB        12.0110       ; Bare Carbon
```

where OB and CB represents the oxygen and carbon atom of CO molecule. After this, the information of a molecule is entered in the file name $ffG43a1.rtp$ where *.rtp* stands for *residue type parameter*. GROMACS takes molecules as a residue, and the parameters for CO molecules are entered in the following way.

[*CO*]

[*atoms*]

Name	Type	Charge	Charge group
OB	OB	−0.0203	0
CB	CB	0.0203	0

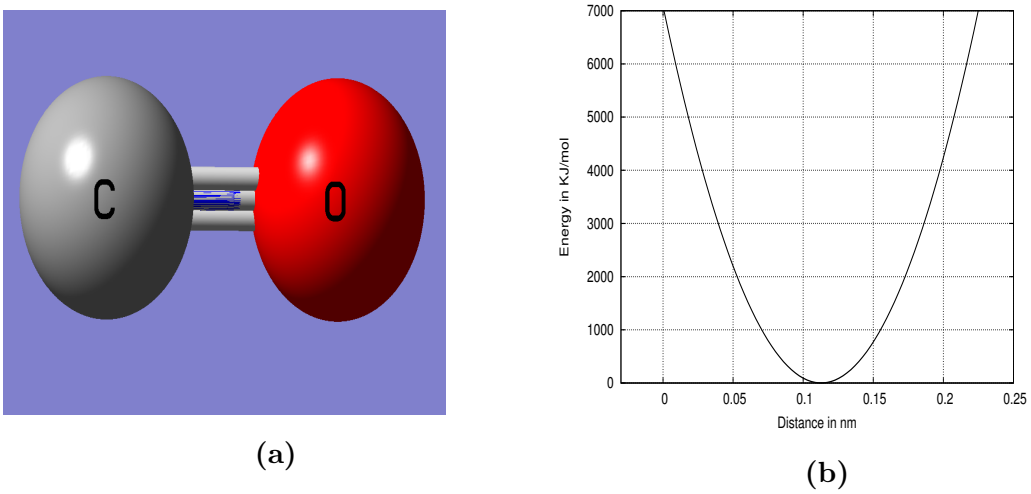
[*bonds*]

Atom 1	Atom 2	Bond information
OB	CB	*gb*_50

where [CO] stands for the residue name for carbon monoxide molecule. The section [atoms] contains the information about the atoms in each molecule (residue). As the partial negative charge in oxygen atom is balanced by the partial positive charge in carbon atom so the atom as a whole is neutral so belongs to charge group 0. The bond information of type *gb*_50 is given in file *ffbonded.itp*. The extension *.itp* is *include topology*. The file *ffbonded.itp* contains information in the form

$$\sharp\text{define} \quad gb\_50 \quad 0.1128 \quad 1.1142 \times 10^6$$

The information in *ffbonded.itp* file are the bond-length 0.1128 nm and the spring constant of the harmonic potential 1.1142×10^6 KJ mol^{-1}nm^{-2}. The structure of the CO molecule and the bond stretching plot of CO molecule is shown in figure (3.1).



(a) **(b)**

Figure 3.1: CO molecule (a) and Bond stretching plot of CO molecule (b).

The left of figure (3.1) represents the model of CO molecule and the right represents the Bond stretching plot of CO molecule which shows the equilibrium bond length of CO molecule is $b_0 = 0.113$ nm.

The non-bonded information (i.e. the Lennard-Jones and Coulomb Potential information) is written in the file name *ffnonbonded.itp* as given below.

[atom types]

name	at.num	mass	charge	ptype	C_6	C_{12}
CB	6	12.0110	0.0203	A	1.3703×10^{-3}	2.4763×10^{-6}
OB	8	15.9994	-0.0203	A	1.9861×10^{-3}	1.8675×10^{-6}

[nonbond-params]

i	j	func	C_6	C_{12}
OB	OB	1	1.9861×10^{-3}	1.8675×10^{-6}

where the name in atom types represents the name of the atom that were entered in *.atp* file. The atoms are particles of type A as understood by GROMACS. In nonbond-params i and j represent Oxygen atoms of different carbon monoxide molecule under consideration. The func. defines the interaction potential. For our force field numeral 1 specifies the Lennard-Jones potential.

In constructing the parameters matrix for the non-bonded L-J-parameters for two different atoms the following combination rule can be used within GROMACS,

(only geometric average) known as type 1 rule in GROMACS.

$$C_{ij}^{6} = (C_{ii}^{6}\, C_{jj}^{6})^{1/2}$$
$$C_{ij}^{12} = (C_{ii}^{12}\, C_{jj}^{12})^{1/2}$$

After defining all these parameters, the bonded and non-bonded parameters, in the GROMACS file system; our next step in the modeling of CO molecule is creation of basic files needed for the simulation. First of all we need the basic structure file. We create this file in the Protein Data Bank (PDB) format

ATOM	1	CB	CO	1	0.000	2.000	3.000	1.000	0.000
ATOM	2	OB	CO	1	1.128	2.000	3.000	1.000	0.000
ATOM	3	CB	CO	2	3.000	1.000	0.000	1.000	0.000
ATOM	4	OB	CO	2	4.128	1.000	0.000	1.000	0.000
ATOM	5	CB	CO	3	-3.000	-2.000	1.000	1.000	0.000
ATOM	6	OB	CO	3	-4.128	-2.000	1.000	1.000	0.000
ATOM	7	CB	CO	4	4.000	0.000	6.000	1.000	0.000
ATOM	8	OB	CO	4	4.000	0.000	7.128	1.000	0.000
ATOM	9	CB	CO	5	1.000	0.000	-3.000	1.000	0.000
ATOM	10	OB	CO	5	1.000	0.000	-4.128	1.000	0.000

The above table shows the pdb (protein data bank) file format for CO molecule. We have five molecules of CO. The keyword ATOM in the first column represents the record which presents atomic coordinates (in Cartesian coordinate format), occupancy, and temperature factor for each atom. The numbers $1, 2, 3, \ldots$ in the second column are atom serial numbers. The third column represents the symbols used to represent atom of CO molecule in the *ffG43a1.atp* file, fourth column represents the molecule name defined in the *ffG43a1.rtp* file and fifth column represents molecule number. The three consecutive columns sixth,seventh and eighth column represents the x, y and z coordinates in Angstrom(A^o). Ninth column is for occupancy and last column represents the temperature factor which is by default zero.

The structure file *CO.pdb* needs to be changed into the structure file readable by GROMACS. For this we use the command **pdb2gmx** inherent in GROMACS which takes *CO.pdb* as input and produce GROMACS structure file *b4edit.gro* and topology file *co.top* as output. The contents of *b4edit.gro* is shown below.

ATOM	1	CB	CO	1	0.000	0.200	0.300
ATOM	2	OB	CO	1	0.113	0.200	0.300
ATOM	3	CB	CO	2	0.300	0.100	0.000
ATOM	4	OB	CO	2	0.413	0.100	0.000
ATOM	5	CB	CO	3	-0.300	-0.200	1.000
ATOM	6	OB	CO	3	-0.413	-0.200	1.000
ATOM	7	CB	CO	4	0.400	0.000	6.000
ATOM	8	OB	CO	4	0.400	0.000	0.713
ATOM	9	CB	CO	5	0.100	0.000	-0.300
ATOM	10	OB	CO	5	0.100	0.000	-0.413
0.82560	0.40000	1.12560					

The gromacs structure file is similar to pdb file except that the coordinates in gromacs structure file is written in nm. The last line in *.gro* file is the information about the size of the box if we use the default value while preparing the box. With this we have completed the parametrization of CO molecule and now we discuss the parametrization of water molecule.

3.3 Modeling Water

There are many force field (pair potentials) which have been applied for the computer simulation of water. Water is such a substance which is used in large scientific community in variety of sectors like biology, chemistry, physics. Parameters that defines the model are structure (geometry); charge(electric); spectroscopic parameters (bond, angle vibrational frequencies); and Lennard-Jones parameters. We use the SPC/E (Extended Single Point Charge) model which is an extension to SPC model developed in 1987. It is a semi-empirical model which consists of a three point charge on each atomic site, two in hydrogen atom and one in oxygen atom [40]. Each hydrogen atom carries a partial charge of +0.410e, and the oxygen atom carries a charge of −0.820e where e is the elementary charge having magnitude 1.602×10^{-19} Coulomb.

3.3.1 Intramolecular Potential Parameters

The intramolecular potential of SPC/E water relies on its geometry. From the perspective of geometry, the SPC/E model is of two types

1. rigid SPC/E model and,

2. flexible SPC/E model

In case of the rigid SPC/E model, the geometrical structure of the water molecule is fixed by fixing the distance between the atoms which means bond length and angle remains fixed. Therefore, the only configurational parameters are positions and velocities of each atomic site. However, the flexible SPC/E model incorporates intramolecular degrees of freedom which implies that the bond distance and bond angle do change in time. The potential represents intramolecular interactions within a certain range of distances and intermolecular interactions elsewhere. In this model the intramolecular potential consists of harmonic bond and angle vibration terms.

$$U_{OH}(r) = \frac{1}{2}\sum K_{OH}(r - b_{OH})^2 \qquad \text{and} \tag{3.6}$$

$$U_{OH}(\Theta) = \frac{1}{2}\sum K_{HOH}\,(\Theta - \Theta_0)^2 \tag{3.7}$$

where K_{OH} is spring constant which measures the strength of interatomic bond between oxygen and hydrogen atom; b_{0H} is the equilibrium bond length between oxygen and hydrogen atoms. Similarly K_{HOH} is spring constant for bond-angle vibration with Θ_0 as the equilibrium HOH bond angle. The parameters that were used for our study are listed in table below [22]. These parameters were presented in the file name *spce.itp* inherent in GROMACS. The curve represented by above two equations is shown in figure (3.2).

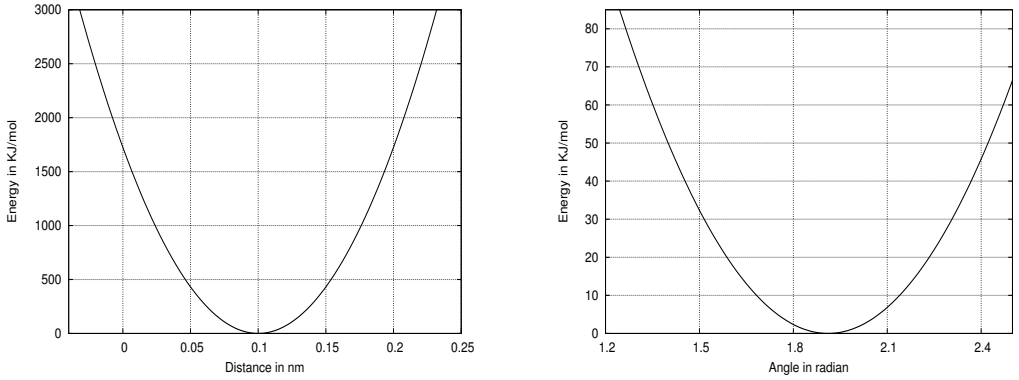


Figure 3.2: Bond Stretching and Angle Stretching potential of water.

Table 3.2: Intramolecular Potential Parameters for Flexible SPC/E Water.

Parameters	Values
K_{OH}	3.45×10^{5} KJ mol^{-1} nm^{-2}
b_{OH}	0.1 nm
K_{HOH}	3.83×10^{2} KJ mol^{-1} rad^{-2}
Θ_o	109.47^o

These parameters were presented in the file name *spce.itp* inherent in GROMACS.

3.3.2 Intermolecular Potential Parameters

So far we have discussed only about the bonded interaction potential and corresponding parameters of SPC/E (Extended Single Point Charge) water. In this section, we discuss about the non-bonded interaction potential and the respective parameters. The intermolecular interaction potential between two SPC/E water molecules consists of two types of interactions namely

1. Lennard-Jones Interaction
2. Coulomb Interaction

The dispersion interaction existing between the two SPC/E water molecules have been approximated by the Lennard-Jones potential and is expressed as:

$$U_{WW}(r_{ij}) = \frac{C_{12}^2}{r_{ij}^{12}} - \frac{C_6^2}{r_{ij}^6} \tag{3.8}$$

where $C_{12}^2 = 4\epsilon_2\sigma_2^{12}$ and $C_6^2 = 4\epsilon_2\sigma_2^6$. For our study, we have taken value of these parameters as follows [36].

Table 3.3: L-J Parameters for SPC/E Water.

Parameters	Values
σ_2	0.3165 nm
ϵ_2	78.1820 k$_B$
C_{12}^2	2.6341×10^{-06} KJ mol^{-1} nm^{12}
C_6^2	2.6173×10^{-03} KJ mol^{-1} nm^{6}

The parameters of L-J interactions were entered in the file *ffnonbonded.itp* as done in carbon monoxide molecule.

[atom types]

name	at.num	mass	charge	ptype	C_6	C_{12}
HW	1	1.0080	0.4100	A	0.000	0.000
OW	8	15.9994	-0.8200	A	2.6173×10^{-03}	2.6341×10^{-06}

[nonbond-params]

i	j	func	C_6	C_{12}
OW	OW	1	2.6173×10^{-03}	2.6341×10^{-06}
OW	HW	1	0.000	0.000

The information about the file is similar to that described for carbon monoxide molecule. A sample graph of L-J energy vs intermolecular distance of water has been shown in figure (3.3).

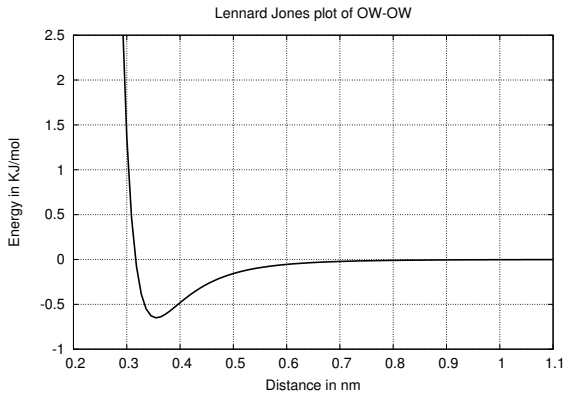


Figure 3.3: Lennard-Jones potential plot for OW-OW.

Due to the presence of partial charge -0.8476e in oxygen and 0.4238e in hydrogen atoms, the non-bonded interaction also includes Coulomb interaction between oxygen and hydrogen atoms of two different water molecules. It is expressed as

$$U_C = \sum_{\alpha=1}^{3}\sum_{\beta=1}^{3} \frac{q_i^{\alpha}\, q_j^{\beta}}{4\pi\epsilon_0 r_{ij}^{\alpha\beta}} \tag{3.9}$$

Our system under study consists of water and carbon monoxide. Till now we have mentioned the non-bonded interaction existing within the same molecules. There

also exists the non-bonded interaction between atoms of different molecules, for that we need to calculate the Lennard-Jones parameter using type first rule as mentioned while modeling carbon monoxide. With this we have completed the modeling of the molecules involved in our system under study.

3.4 Simulation Box and Periodic Boundary Conditions

In this section we will describe, how simulation box is created and how periodic boundary conditions are applied. The shape of the box is presented in figure (3.4). For our study we construct the simulation box of following dimensions.

Table 3.4: Box dimensions.

Box type	Box vectors			Box vector angles			Box volume
	a	**b**	**c**	$<$ **ab**	$<$ **bc**	$<$ **ca**	
Cubic	2.2 nm	2.2 nm	2.2 nm	90^o	90^o	90^o	10.648 nm^3

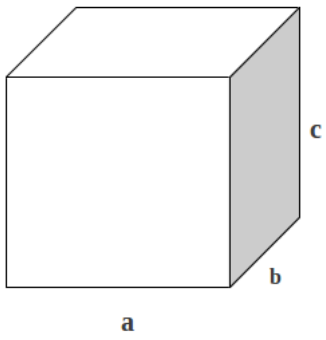


Figure 3.4: Simulation Box.

Figure (3.4) represents the simulation box. $<$ **ab** is the angle between box vectors **a** and **b**. The box is created using command **editconf** inherent in GROMACS with the above specifications. This uses *b4ed.gro* as input file and produces *afed.gro* as output. If we don't specify the box vectors it will take the dimensions we have mentioned in *b4ed.gro*. The contents of *afed.gro* is mentioned below.

```
  10
1CO    OB      1    1.102   1.280   1.260
1CO    CB      2    0.989   1.280   1.260
2CO    OB      3    1.402   1.186   0.960
  ⋮     ⋮      ⋮
2.200  2.200  2.200
```

The information in the file is no different than the one we have described on other files above. It just changes the coordinates of the atoms of the molecules, as it confine the molecules in the box dimension we provide. The last line in the above file gives the dimension of the box. Moreover, **editconf** applies the periodic boundary conditions by default. With these boundary conditions only one, the nearest, image of each particle is considered for short-range non-bonded interaction terms [22].

3.5 Solvation of Carbon monoxide in Water

We describe how we add the desired number of water molecules in the box containing five carbon monoxide molecules to prepare our system under study.

As we know the box prepared in the earlier step already contains the carbon monoxide molecule. We now only have to add the water molecules in the same box. This is done by the command **genbox**, which adds the water molecules in the box. The number of added water molecules is under our control. It takes *spc216.gro*, inherent in gromacs, as a solvent structure file (input file) and *afed.gro* as a solute structure file (input file) and produces *afgen.gro* as a output file. In our work we have added 280 water molecules in the box. Now we need to prepare the topology file specifying both the solute and solvent molecules. This can also be done by editing the topology file, obtained after this step, as our need. The contents of the afgen.gro is shown below.

```
  850
  1CO      OB      1    1.102   1.280   1.260
  1CO      CB      2    0.989   1.128   1.260
 ..........................................
 ..........................................
  6SOL     OW     11    0.569   1.275   1.165
  6SOL     HW1    12    0.476   1.268   1.128
  6SOL     HW2    13    0.580   1.364   1.290
 ..........................................
 ..........................................
 285SOL    HW2   850    2.024   1.583   1.340
  2.200    2.200  2.200
```

where 850 represents the total number of atoms of the system. All the other information is similar to that of *afed.gro* except that it adds the information of atoms of 280 water molecules too. While adding these molecules **genbox** takes care of the condition that the distance between any atom of oxygen molecule and that of the water molecule should be greater than the sum of their van der Waals radii [22]. Each atom associated with the water molecule is assigned respective x, y, and z coordinates in nanometer scale. The topology file (CO.top) is shown below.

```
; Include forcefield parameters
#include "./gromos43a1.ff/forcefield.itp"

[ moleculetype ]
; mol  name           nrexcl
 CO                 1

[ atoms ]
;   nr       type  resnr residue  atom   cgnr     charge      mass  typeB    chargeB
     1         OB      1    CO     OB      1     -0.0203    15.99940   ; qtot -0.203
     2         CB      1    CO     CB      1      0.0203    12.01100   ; qtot  0

[ bonds ]
;  ai    aj funct           c0           c1           c2           c3
    1     2     1   gb_50

; Include water topology
#include "./gromos43a1.ff/spce.itp"

[ system ]
; Name
CO in Water

[ molecules ]
; Compound        #mols
CO                5
SOL               280
```

The term *nrexcl* represents the number of exclusions for the CO molecule. 1 stands for excluding non-bonded interactions between atoms that are no further than 1 bonds away [22]. The term *func* represents the type of potential approximating the covalent bonding between the carbon and oxygen atoms. Numeral 1 stands for the harmonic potential. C0 and C1 represent the bond constant and force constant, respectively which are included in the file *ffbonded.itp*.

3.6 Energy Minimization

With the addition of water in our box, we have finalized our system. However, there may be unphysical van der Waals contacts caused by the atoms that are too close. This results in a configuration of the system which may not be the equilibrium configuration and may be far away from the equilibrium configuration. If a starting configuration is very far from equilibrium, the forces may be excessively large and the MD simulation may fail. In those cases, a robust energy minimization is required. Another reason to perform an energy minimization is

the removal of all kinetic energy from the system: if several snapshots from dynamic simulations must be compared, energy minimization reduces the thermal noise in the structures and potential energies so that they can be compared better [22]. The potential energy function of a (macro)molecular system is a very complex landscape (or hypersurface) in a large number of dimensions. It has one deepest point, the global minimum and a very large number of local minima, where all derivatives of the potential energy function with respect to the coordinates are zero and all second derivatives are non-negative. Energy minimization is the process to bring the system to one of these local minima so that the initial structure for simulation possesses minimum energy and the simulation can run smoothly. There are two different algorithms that can be employed to attain energy minimized state of the system.

a. Steepest Descent Method

b. Conjugate Gradient Method

We have used the steepest descent method for energy minimization in our present work. The details are explained below.

3.6.1 Steepest Descent Method

The steepest descent method moves in the direction parallel to the net force without consideration of the history built up in previous steps, which in our geographical analogy corresponds to walking straight downhill. We define the vector $\mathbf{r}$ as the vector of all 3N coordinates. Initially a maximum displacement h_0 (eg 0.01nm) must be given. First the forces $\mathbf{F}$ and potential energy are calculated, new positions are calculated by

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{max(|\mathbf{F}_n|)} h_n \tag{3.10}$$

where h_n is the maximum displacement and $\mathbf{F}_n$ is the force i.e. negative gradient of the potential V and $|\mathbf{F}_n|$ means the largest of the absolute values of the force components. The forces and energy are again computed for the new positions.

If $(V_{n+1} < V_n)$ then new positions are accepted and $h_{n+1} = 1.2h_n$
If $(V_{n+1} \geq V_n)$ then the new positions are rejected and $h_n = 0.2h_n$.

The algorithm stops when either a user-specified number of force evaluations has been performed or when the maximum of the absolute values of the force (gradient) components is small than a specified value.

This method is simple and sturdy, its convergence can be quite slow, especially in the vicinity of the local minimum. The faster converging conjugate gradient method uses gradient information from the previous steps. In general, steepest descents will bring us close to the nearest local minimum very quickly, while conjugate gradient brings you very close to the local minimum, but performs worse far away from the equilibrium [22].

For the energy minimization the commands **grompp** and **mdrun** are used in succession. The first step is to use command **grompp**. This program requires three input files namely *afgen.gro*, *em.mdp* and *CO.top*. It generates a run input file em.tpr; *afgen.gro* is the output structure file generated by **genbox**; *em.mdp* is a molecular dynamics parameter file that contains the parameters for the simulation for energy minimization and is presented in figure below.

```
define              =  -DFLEX_SPCE ; flexible SPC/E water model used
constraints         =  none ; constraints are not applied
integrator          =  steep ; steepest descent algorithm
nsteps              =  1500 ; number of steps
;
;       Energy minimizing stuff
;
emtol               =  150  ; force tolerence level
emstep              =  0.001 ; initial step size in nm
nstcomm             =  1     ;frequency for center of mass reoval
ns_type             =  grid  ; neighbor search type
rlist               =  1.0   ;  cut off distance for the neighbor list (nm)
rcoulomb            =  1.0  ; Coulomb cut-off distance (nm)
rvdw                =  1.0 ; L-J cut off distance (nm)
nstxtcout           =  20  ; frequency to write coordinates to xtc trajectory
Tcoupl              =  no ; temperature coupling
Pcoupl              =  no ; pressure coupling
gen_vel             =  no ; velocity generation
```

These parameters indicate that GROMACS uses flexible SPC/E water for energy minimization and the system is not coupled with any thermostat or barostat. It creates at least two important output files namely *em.edr* and *em.gro*. The structure file *em.gro* contains the coordinates of the atoms of the last step of energy minimization. The energy file *em.edr* contains different energies due to different

interactions as specified in the force field.

Since the energy minimization changes the structure of the system from chaos to a stable structure hence we see new distribution of molecules within the box which is given in figure (3.5).

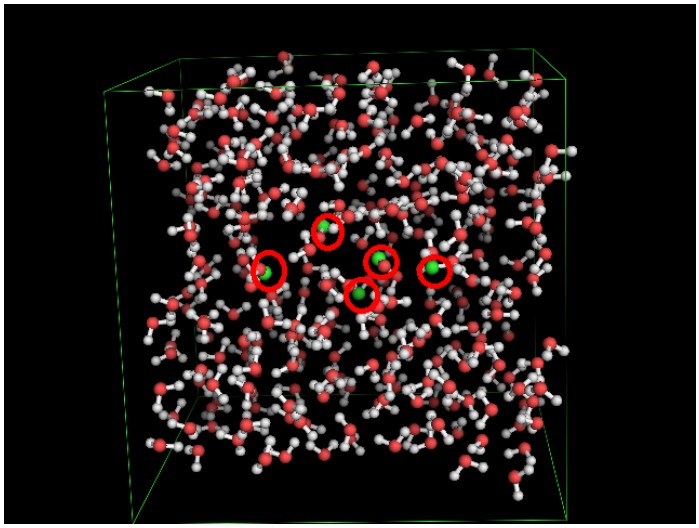


Figure 3.5: Structure after Energy minimization (encircled molecules are carbon monoxide).

Figure (3.5) represents the system structure after energy minimization. The local equilibrium state of the system is associated with its minimum energy after energy minimization. Figure (3.6) shows how the energy of the system changes in time in the course of bringing the system in a local equilibrium state.

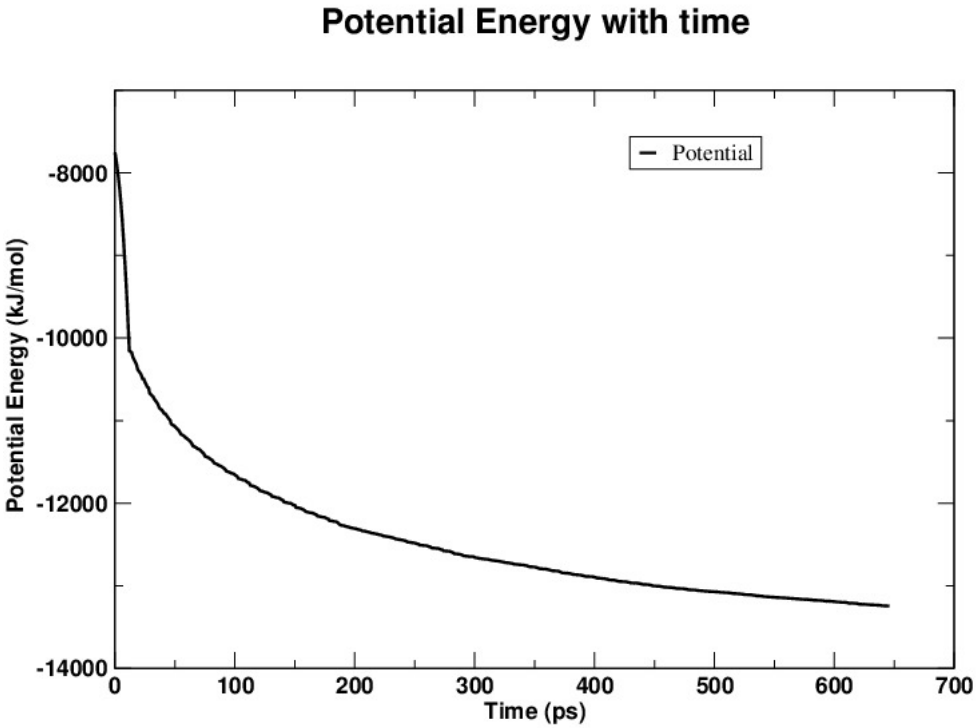


Figure 3.6: Potential energy plot after Energy minimization.

3.7 Equilibration

As soon as we perform energy minimization, we are ready for the simulation of the system for equilibration run and then we perform the production run after which we calculate the various properties of the system. The dynamical variables, however, usually vary with the parameters such as temperature, pressure, density, etc. Therefore, before performing the production simulation, which would reveal the parameters under study, the system must be brought to state of thermodynamic equilibrium. This process is called equilibration. The thermal equilibrium of the system is achieved by coupling the system with a suitable thermostat. This is known as Temperature Coupling. Similarly, a constant pressure subjected to the system is achieved by coupling the system with a suitable barostat. GROMACS understands this process as Pressure Coupling. With this we can say that the system we have taken is subjected to NPT ensemble. Constant temperature is required when we wish to study the property of the system

at different temperatures.

The process to perform equilibration is similar to that of energy-minimization. We use the command **grompp** and **mdrun** in succession for this. Gromacs preprocessor command **grompp** uses *em.gro*, *npt.mdp* and *CO.top* as a input file and produces a binary file *npt.tpr*, *em.gro* is the structure file obtained after energy-minimization and *npt.mdp* is the file containing molecular dynamics parameter which is shown below in figure (3.7) and *CO.top* is the same topology file we used during energy-minimization.The binary file *npt.tpr* is used as a input for process **mdrun** and produces trajectory file *npt.trr*, energy file *npt.edr* and structure file *npt.gro*. The system temperature, pressure, density is obtained from file *npt.edr* using command **g_energy**.

```
;PREPROCESSING parameters
integrator        = md
dt                =.002
nsteps          = 100000000
nstcomm                 = 1

;OUPUT CONTROL parameters.
nstxout         = 500
nstvout         = 500
nstfout         = 500
nstlog          = 500
nstenergy       = 500
nstxtcout       = 500
energygrps      = system
;NEIGHBOUR SEARCHING parameters.
nstlist         = 10
ns_type         = grid
rlist           = 1.0
;ELECTROSTATIC and VdW parameters.
rcoulomb        = 1.0
rvdw            = 1.0
epsilon-r       = 1
;BERENDSEN TEMPERATURE COUPLING is on in two groups
Tcoupl          = berendsen
tc-grps         = system
tau_t           = 0.01
ref_t           = 293
;PRESSURE COUPLING is on
Pcoupl          = berendsen
tau_p           = 0.6
compressibility     = 4.6e-5
ref_p           = 1.0
;GENERATE VELOCITIES is on at 293 K.
gen_vel         = yes;    ; generate initially
gen_temp        = 293
gen_seed        = 173259   ;give different values for different trials.
;BONDS parameters
constraints       = all-bonds
constraint-algorithm = shake
unconstrained-start  = no
pbc             = xyz
```

Figure 3.7: Equilibration parameters at temperature 293K.

Leap-frog algorithm has been used as integrator. A Berendsen thermostat has been used to couple the system with the temperature coupling constant tau/t of 0.01 ps. The reference temperature used in our study are 293K, 298K, 303K, 313K, 323K, 333K, 345K, 360K. The velocities are generated initially using Maxwell-Boltzmann distribution function at the specified temperature of the system. A Berendsen pressure coupling has been applied with the reference pressure of 1 bar and the pressure coupling time constant tau/p has been taken as 0.6 ps. An isothermal compressibility of $4.60 \times 10^{-5}\ bar^{-1}$ has been taken. All the bonds are converted to constraints using SHAKE algorithm and those of SPC/E water were constrained using SETTLE algorithm. The SHAKE algorithm resembles the SETTLE algorithm when all the atoms in a molecules are fixed by fixing the distance between them. We also have fixed the distances between the atoms in carbon monoxide molecules, hence the SHAKE algorithm behaves as SETTLE algorithm. The system has been subjected to equilibration for a time period of 200 ns.

Table (3.5) shows the values of the densities and simulated temperatures at various coupled temperatures.

Table 3.5: Values of simulated temperature (T_{sim}) and density at various coupling temperatures (T_{CO}).

S.N	(T_{co}) K	(T_{sim}) K	$Density_{system}$ (kg/m^3)	$Density_{water}$ (kg/m^3)[45]
1	293	293.097 ± 0.007	1002.010 ± 0.029	998.19
2	298	298.106 ± 0.005	998.949 ± 0.032	997.02
3	303	303.109 ± 0.006	995.696 ± 0.051	995.61
4	313	313.108 ± 0.007	988.868 ± 0.031	992.17
5	323	323.118 ± 0.006	981.432 ± 0.047	987.99
6	333	333.128 ± 0.005	973.478 ± 0.035	983.16
7	345	345.133 ± 0.007	963.254 ± 0.017	
8	360	360.136 ± 0.005	949.663 ± 0.032	

The above table shows the value of simulated temperature (T_{sim}) and density at various coupling temperatures (T_{co}). The value of the system density has been compared with the available data of water density at various temperatures [45]. As the number of carbon-monoxide molecules is less in comparison with water molecules in our system (5-CO and 280-H_2O), the density of the system is

comparable with that of water. The maximum fluctuation of the system density with water density is less than 1% by which we can say, our simulated density is in good agreement with water density. We perform the equilibration run for a time period of 200 ns and the fluctuation of the system densities and temperatures with time is shown in following figures.

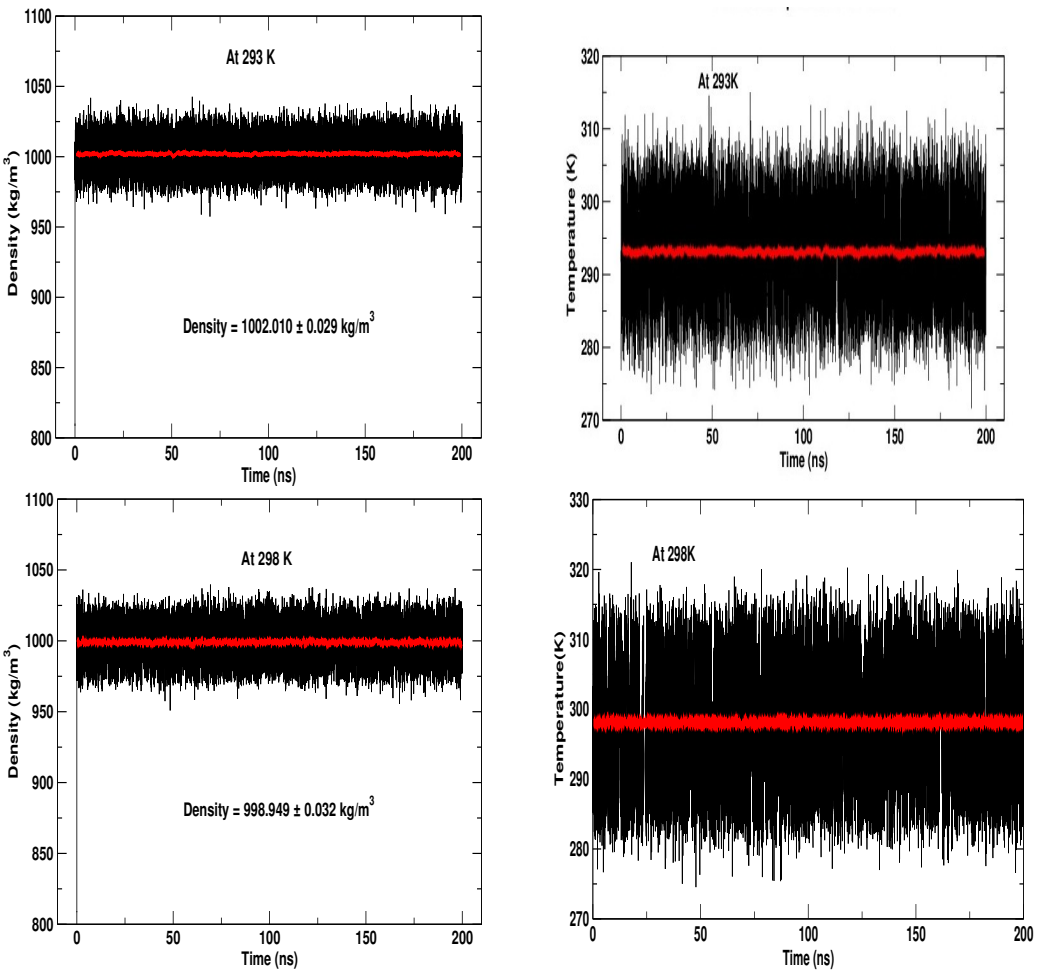


Figure 3.8: Density (left) and simulated temperature (right) vs Time at coupled temperatures 293K, 298K.

Figure (3.8) shows the variation of system temperature and density at temperatures 293K and 298K where it is seen that the system temperature and density attain the equilibrium value at very small time interval. We have simulated the system for time 200 ns to get the desired value of temperature and pressure and it is seen that the system attain the equilibrium state up to that time scale. The nearly horizontal line in figure shows the average value of density and temperatures at desired temperature value. The succeeding figures shows the similar

information for different desired temperatures.

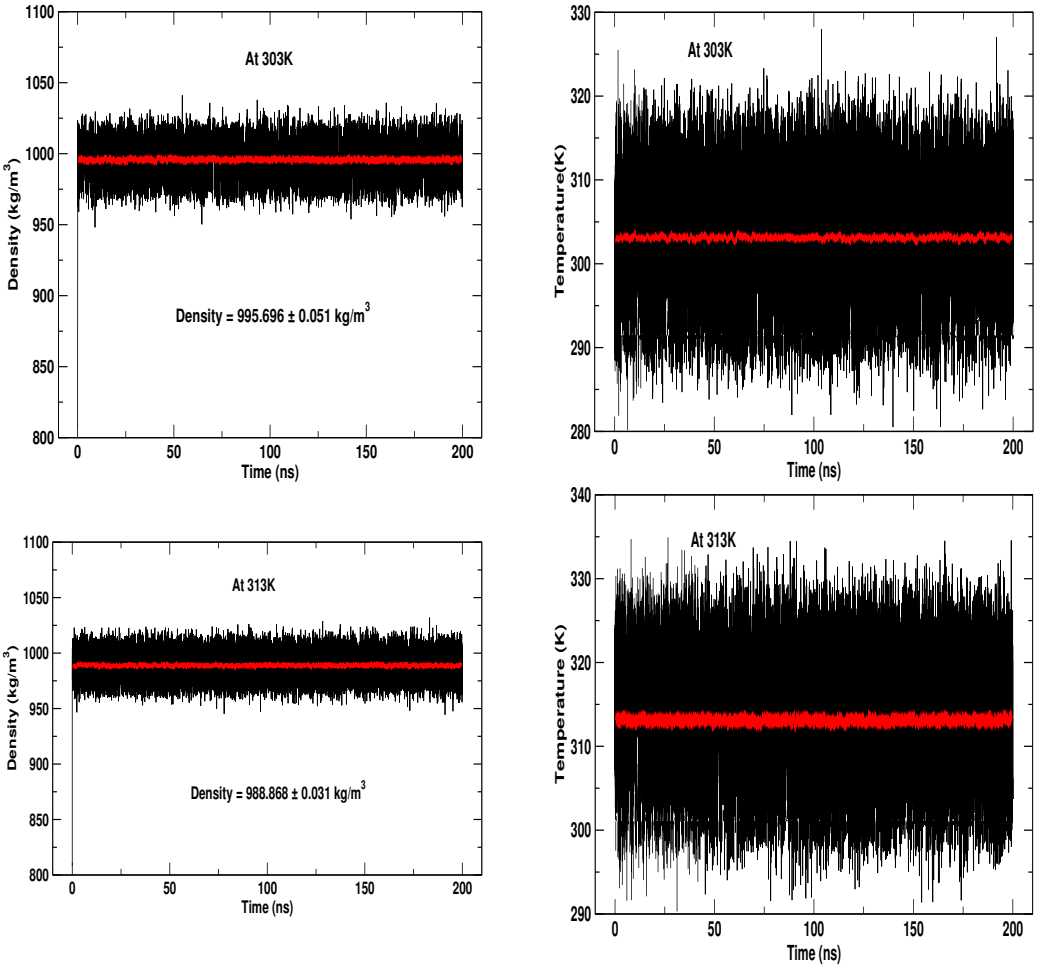


Figure 3.9: Density (left) and simulated temperature (right) vs Time at coupled temperatures 303K, 313K.

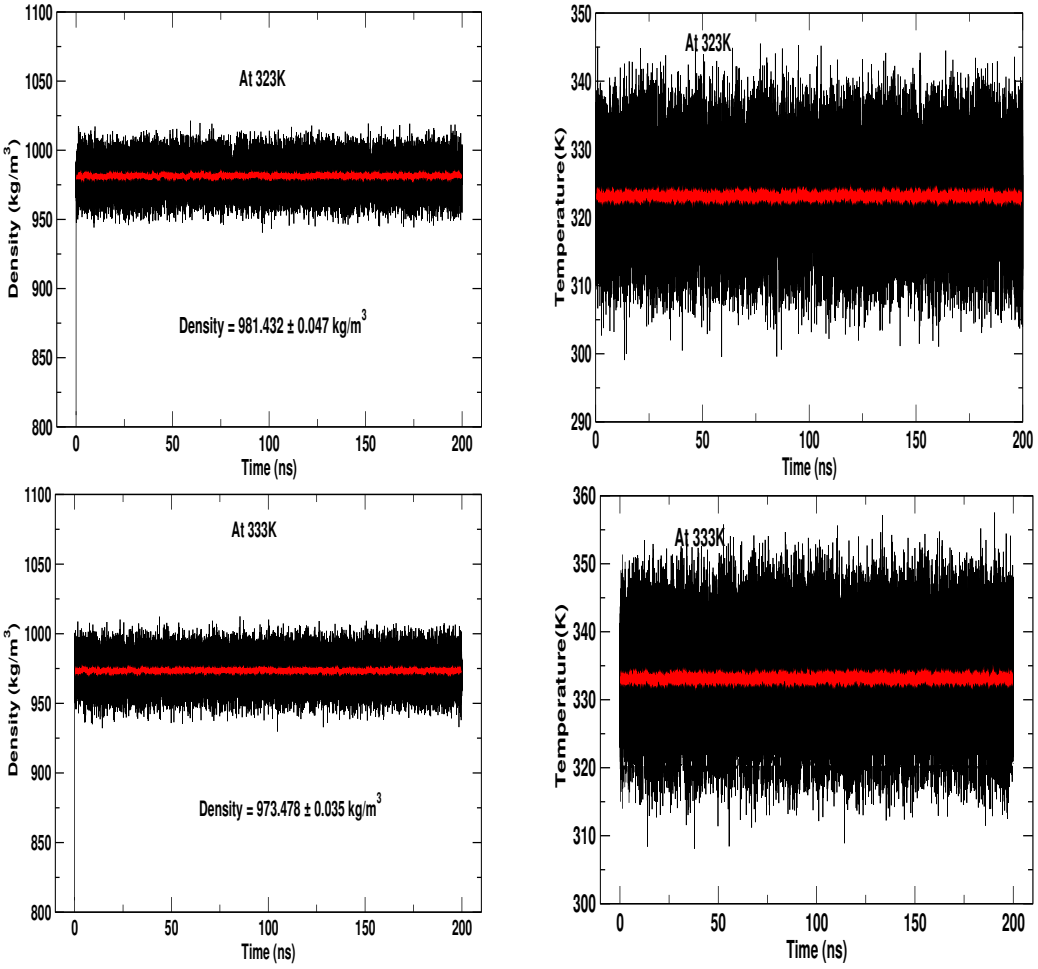


Figure 3.10: Density and simulated temperature vs Time at coupled temperatures 323K, 333K.

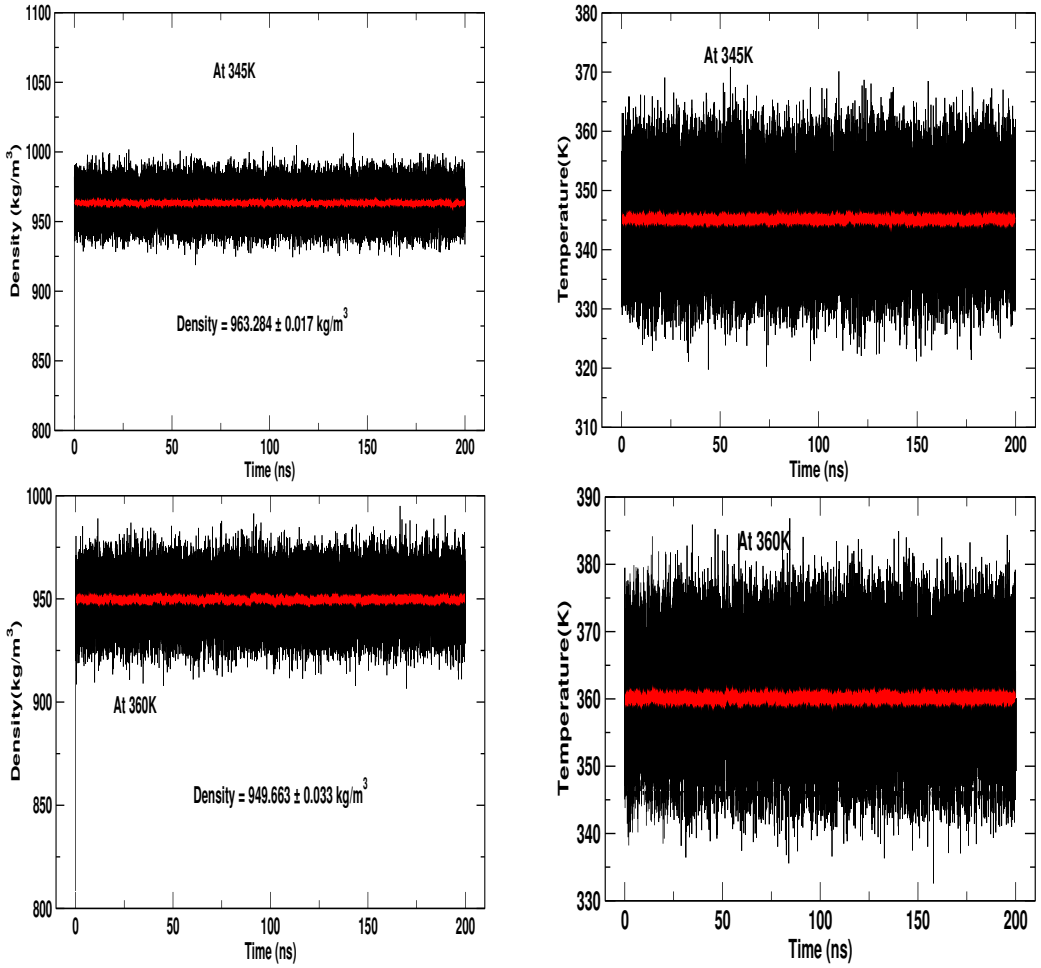


Figure 3.11: Density (left) and simulated temperature (right) vs Time at coupled temperatures 345K, 360K.

3.8 Production

As described in the above section, we have coupled our system to the desired value of the temperature and system pressure of 1 bar. It brings the system to a fixed density (i.e. it changes the box size so as to keep the system in fixed density at given temperature). In this step (i.e. production run) we use the structure file obtained after coupling the system to a fixed temperature and pressure. This means we are having our system volume fixed so this is the case of NVT ensemble. After this run we will be in position to calculate the equilibrium properties of the system (e.g. diffusion).

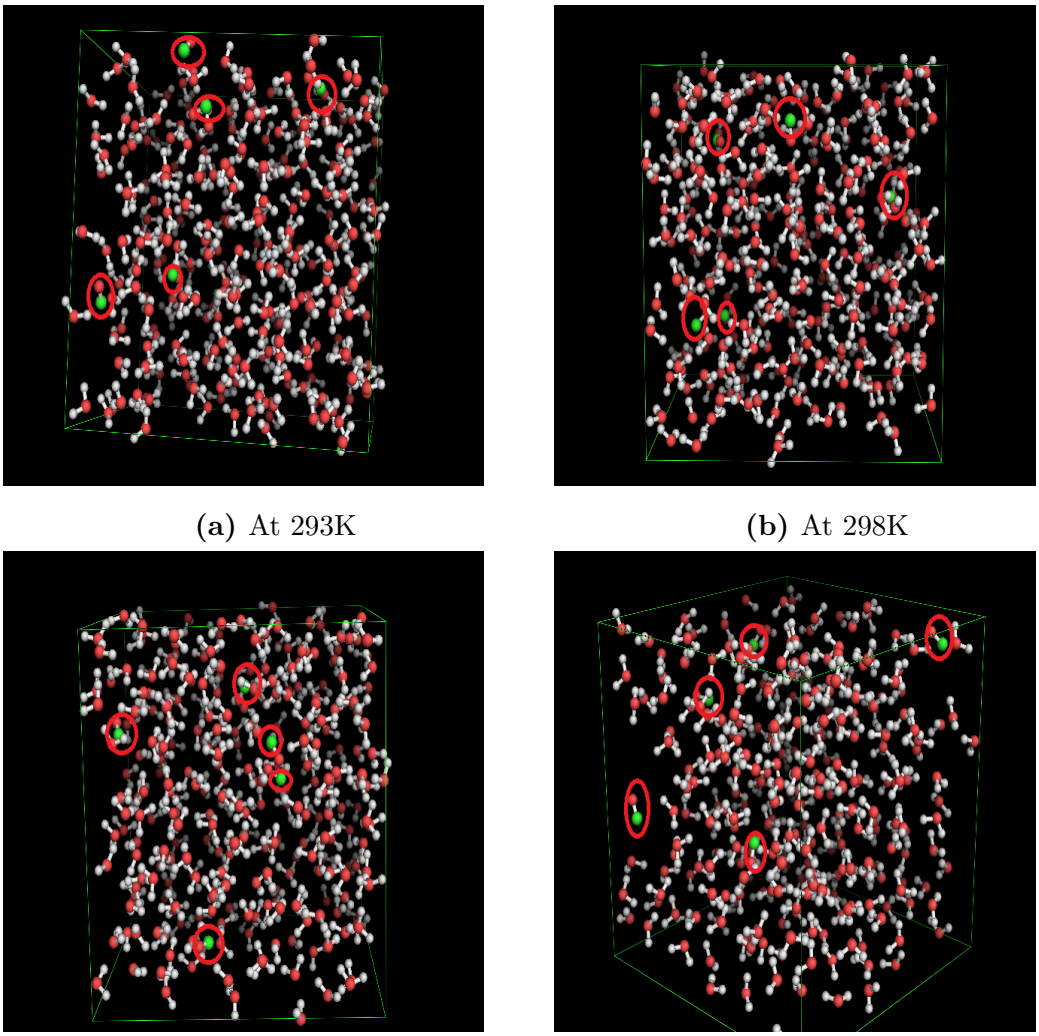
We use the command **grompp** and **mdrun** in succession as in equilibration

run. The command **grompp** uses the file *nvt.mdp*, *npt.gro* and *CO.top* as input file and produce the binary file *em.tpr* as output which is the input file for **mdrun**. *npt.gro* is the structure file which is the output file of equilibration run , *CO.top* is the topology file which is the same as used in equilibration run and *nvt.mdp* is the molecular dynamics parameter file which contains the information about the production run. We have performed the production run for 200 ns. We use no pressure coupling and no generation of velocity, ensuring the NVT ensemble. Moreover, initially the the simulation was unconstrained as there was no generation of velocities.

```
tinit          = 0
integrator      = md
dt               =.002
nsteps         = 100000000
nstcomm              = 1
nstxout         = 500
nstvout         = 500
nstfout         = 500
nstlog          = 500
nstenergy       = 500
nstxtcout       = 500
energygrps      = system
nstlist         = 10
ns_type         = grid
rlist           = 1.0
rcoulomb        = 1.0
rvdw            = 1.0
epsilon-r       = 1
Tcoupl          = berendsen
tc-grps          = system
tau_t           = 0.01
ref_t           = 293
Pcoupl          = no
gen_vel         = no;
constraints     = all-bonds
constraint-algorithm = shake
unconstrained-start  = yes
pbc             = xyz
```

Figure 3.12: Production parameters at temperature 293K.

The system structures after equilibration run at different temperatures are shown in figure below.



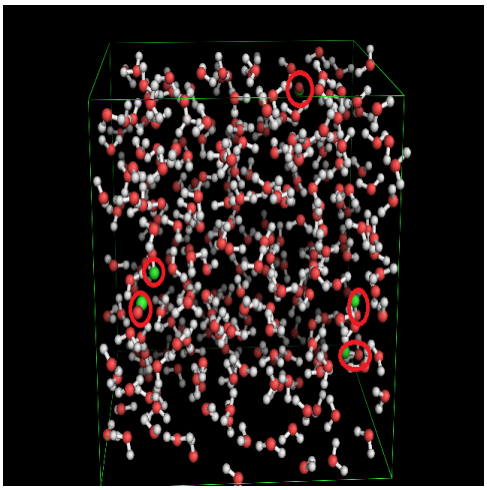
(a) At 293K

(b) At 298K

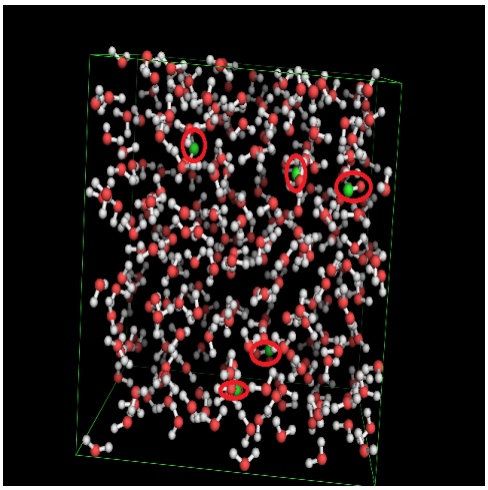
(c) At 303K

(d) At 313K

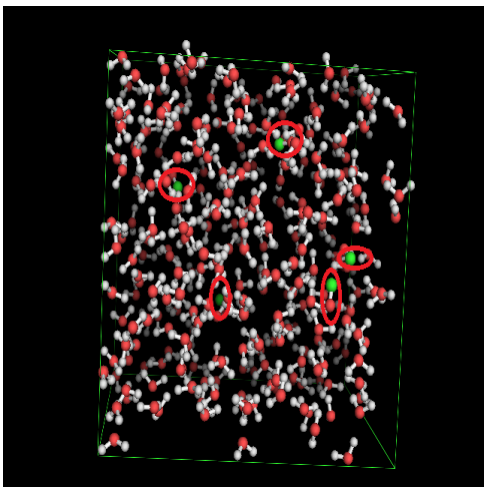
Figure 3.13: Structure after equilibration at temperatures 293K, 298K, 303K, 313K (carbon monoxide molecules are encircled).



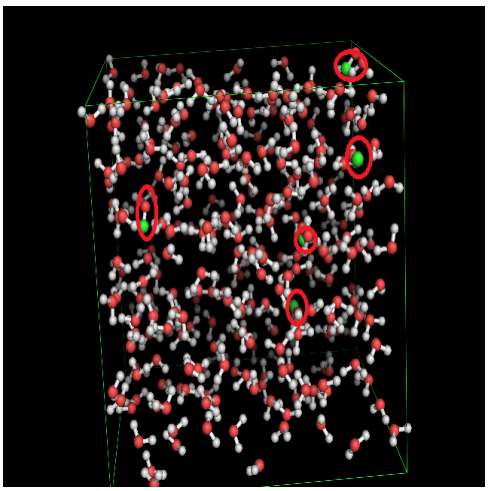
(a) At 323K



(b) At 333K



(c) At 345K



(d) At 360K

Figure 3.14: Structure after equilibration at temperatures 323K, 333K, 345K, 360K (carbon monoxide molecules are encircled).

Chapter 4

Results and Discussion

4.1 General Considerations

In the previous chapter we have presented the details of simulation along with the temperature and density fluctuation of the system at various temperatures. We have also presented the system structure after equilibration run. In this chapter we focus on presentation and discussion of the main findings of the present work. We discuss about the structural analysis of the system by studying radial distribution function (RDF) at various temperatures, energy profile of the system and the transport properties of CO in water i.e. self-diffusion coefficient of CO and water and the mutual diffusion coefficient of CO in water. The study of diffusion coefficient at various temperatures enables us to study the temperature dependence of self-diffusion coefficient.

4.2 Structural Analysis

Each molecule in a liquid is in constant interaction with a large number of its neighbors. When the terminology liquid structure is used, one is referring to quantities like structure factor $s(q)$, radial distribution function $g(r)$. The advantage of molecular simulation is that we can study $g(r)$ or $s(q)$ [7]. Since both give same information we focus on $g(r)$. The central idea in most theories of liquids is the radial distribution function (RDF) [46]. The detail information of $g(r)$ is provided in chapter 2. We have discussed three different radial distribution functions $g_{(OW-OW)}(r)$, $g_{(OC-OW)}(r)$, $g_{(OC-OC)}(r)$. $g_{(OW-OW)}(r)$ represents the RDF of oxygen atoms of water molecule which determine the structure of

the solvent present in the system i.e. SPC/E water. $g_{(OC-OW)}(r)$ represents RDF between the oxygen atom of CO molecule and oxygen atom of water molecule. Similarly $g_{(OC-OC)}(r)$ is the RDF of the oxygen atoms of CO molecule.

The graph of radial distribution function, $g(r)$, as a function of separation distance shows a pattern of peaks and troughs [47]. It tends to unity at large r and vanishes as $r \to 0$ as a consequence of the strongly repulsive forces, that act at small particle separations [35]. It remains 0 up to certain distance which implies that there is zero probability of finding the particle within this range. This distance helps to estimate the excluded volume. Mathematically it is due to r^{-12} term of Lennard-Jones interaction and repulsive Coulomb interaction incorporated in the model of the system. Since particle become effectively hard as $r \to 0$ so $g \to 0$ as it approaches the distance less than the size of the particle and since the influence of the particle at the origin diminishes as r becomes large so $g \to 1$ as $r \to \infty$ [46]. In addition the form of $g(r)$ provides considerable insight into what is meant by the structure of a liquid, at least at the level of pair correlations.

The radial distribution function turns out to be the central importance in the theory of structure of liquids for two reasons. If the total potential energy of the N-body system is pair-wise additive then we can write the thermodynamic functions of the system in terms of $g(r)$. In addition RDF can be determined by diffraction studies on liquids, like X-ray diffraction, Neutron diffraction and it is therefore possible to determine thermodynamic functions of a liquid from diffraction studies. In a solid the molecules are arranged in a regular repeating order, and this leads to a sharp X-ray diffraction pattern from which the order of the molecules in the solid can be obtained. The X-ray diffraction pattern of a liquid is more diffuse, but nevertheless can be used to determine the local short-range order in a fluid. The peaks in the curve represent smeared-out shells of the first nearest neighbors, second nearest neighbors etc, which in a sense are remnants of the ordering found in the solid [46].

The radial distribution function is obtained using command **g_rdf** inherent in GROMACS. The RDF of oxygen atoms in water molecule at temperature $298K$ is shown in figure (4.1).

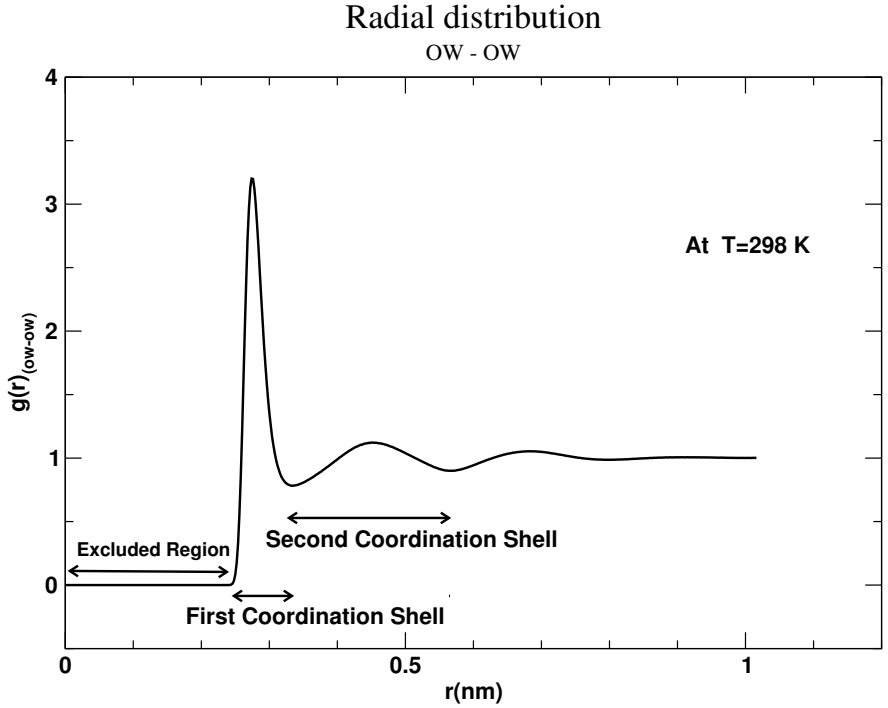


Figure 4.1: RDF between oxygen atoms of water molecule at 298K.

If the system obeys Lennard-Jones $6-12$ interaction then first neighbor shell will be found around r $= 2^{1/6}\sigma$ (Figure (4.1)). The nearest neighbors, which comprise the first coordination shell, tend to exclude the next nearest neighbors from an intermediate region around r $\approx \ 2^{\frac{1}{6}} \times (\frac{3}{2})\sigma$. Thus $g(r)$ will be less than unity in that region, and it will peak above the uncorrelated result near $r = 2 \times 2^{1/6}\ \sigma$ which is the van der Waals radius. In our case there is influence of not only Lennard-Jones interaction but also the Coulomb interaction so the above mentioned statistics doesn't exactly match in the figure (4.1). $g(r)$ for a simple atomic liquid looks like as in figure (4.1). The second peak corresponds to the most probable location for the next nearest neighbors. These neighbors comprise the second coordination shell. This layering manifests the granularity (non-continuum) nature of the liquid. It shows up in an oscillatory form for $g(r)$, which persists until r is larger than the range of correlations (typically a few molecular diameters in a dense liquid). The graph of $g(r)$ shows a finite density of the particles even in unlikely regions like in between first and second peak. This is one of the features that distinguishes a liquid from a crystalline solid. Without it, the possibility of diffusion would be drastically reduced [47]. Figure (4.2) shows the RDF of oxygen atoms of water molecules at different

temperatures.

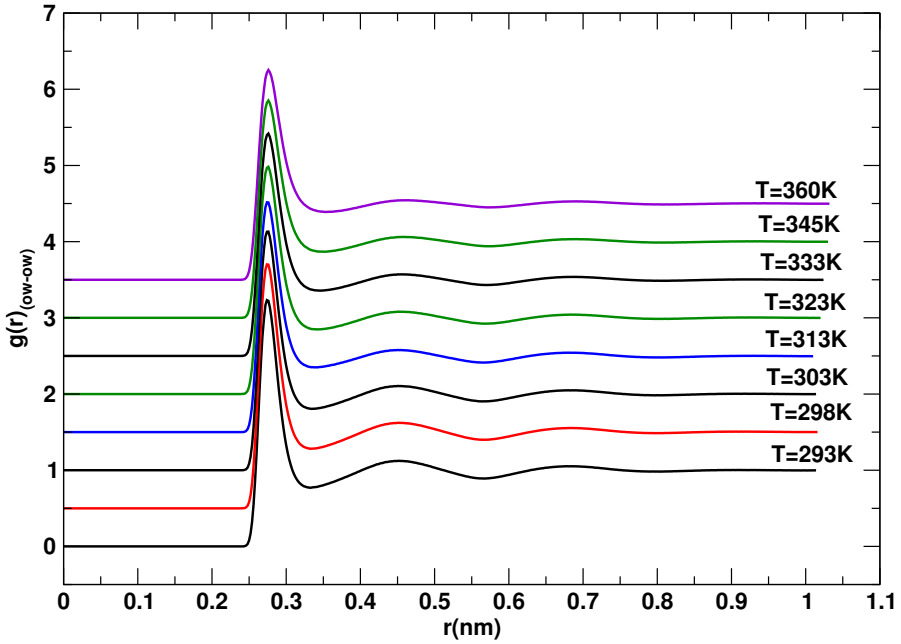


Figure 4.2: RDF of oxygen atoms of water molecules at different temperatures.

From figure (4.2) one can see three different peaks and the value of the peak and position of the peak in detail is provided in table (4.1).

Table 4.1: Simulated data for the RDF analysis between the solvent molecules.

RDF analysis of OW-OW							
T(K)	ER (nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
293	0.232	0.274	3.233	0.450	1.123	0.678	1.053
298	0.232	0.274	3.201	0.450	1.123	0.686	1.054
303	0.234	0.276	3.131	0.450	1.106	0.680	1.049
313	0.234	0.274	3.017	0.450	1.078	0.690	1.042
323	0.232	0.276	2.983	0.450	1.080	0.690	1.042
333	0.234	0.276	2.919	0.454	1.071	0.692	1.037
345	0.234	0.276	2.753	0.462	1.044	0.690	1.029
360	0.232	0.276	2.753	0.462	1.044	0.690	1.029

In table (4.1) the acronyms used are as follows :

ER - Excluded Region

FPP - First Peak Position

FPV - First Peak Value

SPP - Second Peak Position
SPV - Second Peak Value
TPP - Third Peak Position
TPV - Third Peak Value

The value of the Excluded Region gives the value up to which $g_{ow-ow}(r)$ is zero. The value of σ for OW-OW is 0.3165 nm, and the van der Waals radius ($2^{1/6}\sigma$) is 0.3553 nm. From table (4.1) it is seen that the value of the Excluded region is less than the van der Waals radius which is in accordance with the theory that there is no probability of finding the particle below van der Waals radius and the value of van der Waals radius is greater than the first peak position which implies other potential also plays a role in addition to van der Waals potential which is shown in figure (4.3). The FPP of RDF almost remains at the same position ($\pm$ 0.002 nm) and doesn't change with temperature. However, the SPP and TPP shift towards the right with the rise in temperature. Further the height of the peak decreases with the rise in temperature. Moreover, with the rise in temperature the width of the first and second peaks increases by a small amount. These all signify that the solvent becomes less structured as temperature is raised.

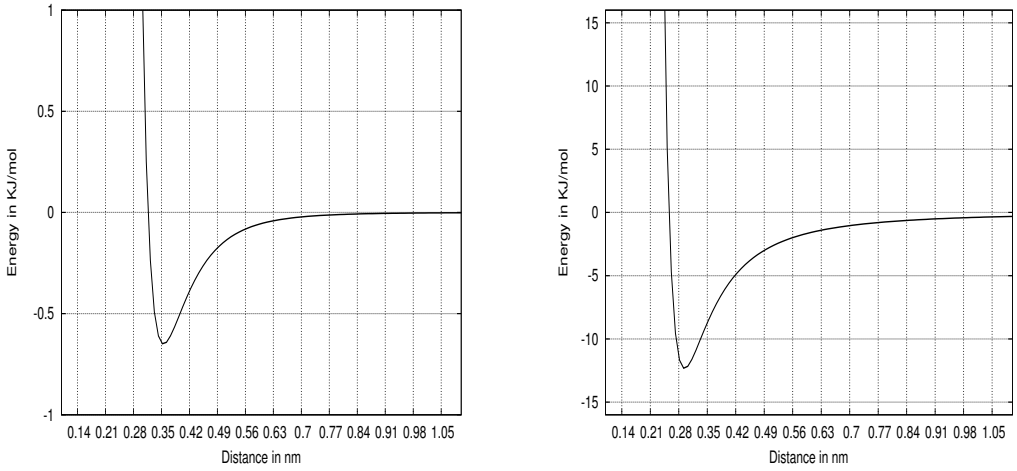


Figure 4.3: Plot of Lennard-Jones potential (left) and Lennard-Jones plus Coulomb potential (right) with distance for two different water molecules.

From figure (4.3) we can say that the maximum probability of finding other oxygen atoms of water molecule with respect to fixed oxygen atom shift left when

we consider both Lennard-Jones and Coulomb potential acted on the particle than only when we consider Lennard-Jones interaction. The FPP in figure (4.2) is nearly at the same position as the minimum of Lennard-Jones plus Coulomb plot in figure (4.3) (right). To be precise FPP is slightly less than the minimum of Lennard-Jones plus Coulomb plot as we have considered only two water molecules and many body effects is actually inevitable in the system. We then study the RDF between oxygen atom of CO molecule and oxygen atom of water molecule which is shown in figure (4.4)

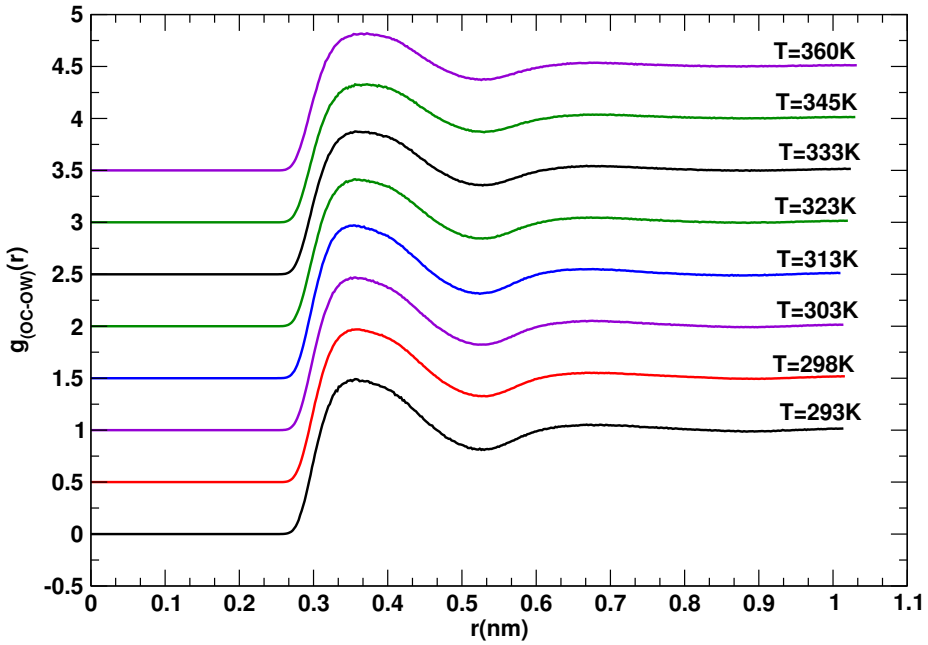


Figure 4.4: RDF of oxygen atom of CO molecule and oxygen atom of water molecule.

From figure (4.4) one can see two different peaks and the value of the peak and position of the peak in detail is provided in table (4.2).

Table 4.2: Simulated data for the RDF analysis between oxygen atom of CO molecule and oxygen atom of water molecule.

RDF analysis of OC-OW					
T(K)	ER (nm)	FPP(nm)	FPV	SPP(nm)	SPV
293	0.252	0.356	1.491	0.668	1.054
298	0.248	0.360	1.470	0.672	1.052
303	0.248	0.356	1.470	0.676	1.053
313	0.250	0.354	1.471	0.676	1.050
323	0.246	0.356	1.416	0.674	1.047
333	0.248	0.356	1.375	0.668	1.044
345	0.248	0.372	1.329	0.700	1.038
360	0.246	0.372	1.319	0.672	1.037

The positions of the first peak and second peak doesn't follow a trend. The value of first peak and second peak decreases slightly with the increase in temperature. This signifies that the correlation between carbon monoxide and water molecules decreases with the rise in temperature which is caused by the increased thermal energy acquired by the molecules. The value of σ_{oc-ow} is 0.315 nm and the van der Waals radius is 0.354 nm. It is seen from table (4.2) that the value of first peak point is greater than van der Waals radius and the excluded region is less than the van der Waals radius of oxygen atoms of two different types of molecule of the system.

Figure (4.5) shows the RDF of oxygen atoms of carbon monoxide molecule.

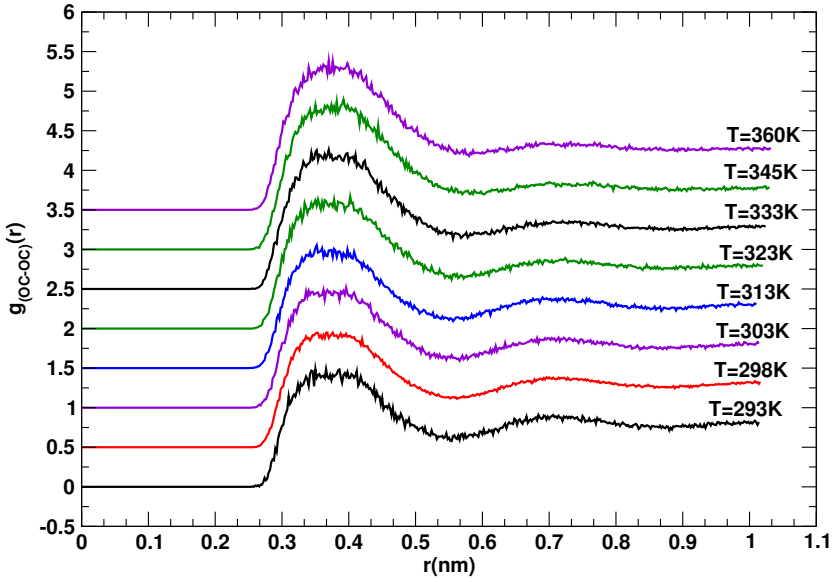


Figure 4.5: RDF between oxygen atoms of CO molecule.

From figure (4.5) one can see two different peaks and the value of the peak and position of the peak in detail is provided in table (4.3).

Table 4.3: Simulated data for the RDF analysis between oxygen atoms of CO molecule.

RDF analysis of OC-OC					
T(K)	ER (nm)	FPP(nm)	FPV	SPP(nm)	SPV
293	0.258	0.366	1.487	0.706	0.908
298	0.256	0.376	1.449	0.696	0.884
303	0.254	0.368	1.488	0.694	0.880
313	0.254	0.352	1.553	0.690	0.894
323	0.256	0.354	1.632	0.718	0.875
333	0.252	0.352	1.752	0.720	0.860
345	0.250	0.382	1.870	0.766	0.856
360	0.250	0.376	1.879	0.706	0.850

The value of σ_{oc-oc} is 0.310 nm and the value of van der Waals radius is 0.351 nm. The excluded region is less than van der Waals radius and first peak position is greater than van der Waals radius. The nature of the rough peak, clearly visible in the fig (4.5), is due to the lesser number of carbon monoxide molecule. This in turn plays a role for undefined trend in the position and values of the peaks, as

temperature increases. Although our simulated system contains only 5 carbon monoxide (CO) molecules and 280 water molecules, the RDF we have calculated informs that these CO molecules do not move independently of each other and are, therefore, correlated to some extent, ascertaining that the system cannot be considered to be perfectly infinitely dilute.

4.3 Energy Profile

In this section we discuss the energy of the system under study. We try to gain some idea how the different energies change with the temperature and time. We have calculated the kinetic energy, potential energy, L-J potential, coulomb potential and total energy. Since we have used the cutoff Lennard-Jones potential and also cutoff coulomb potential the molecules under these potentials will have only short range interaction and thus calculated data for the energy is energy due to short range interactions. The total potential energy is sum of the L-J and coulomb potential energies. Similarly the total energy is the sum of total potential energy and kinetic energy at the desired temperature. The graph of the energies of all temperatures gives the information how the energy profile varies with the temperature. To calculate the energy profile we use the command **g_energy** inherent in GROMACS. The plot of energies for various temperatures is shown in succeeding figures.

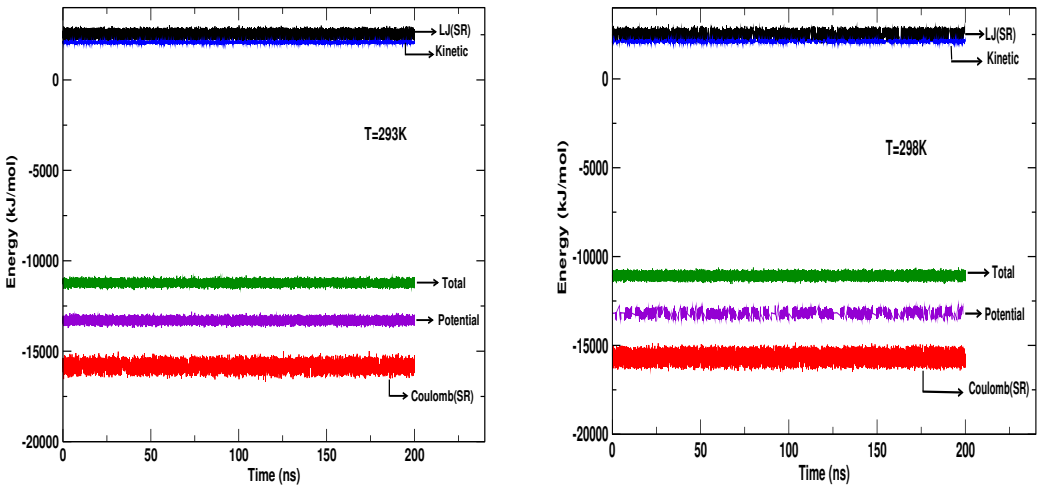


Figure 4.6: Energy profile of the system at temperatures 293K & 298K.

Table 4.4: Energy profile table at temperatures 293K and 298K.

Different energies at temperature 293K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2526.66	0.20	121.938	0.995
Coulomb(SR)	-15823.70	0.51	186.988	-2.853
Potential	-13297.10	0.34	105.632	-1.858
Kinetic	2074.00	0.01	35.681	-0.007
Total	-11223.10	0.35	100.511	-1.866

Different energies at temperature 298K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2482.60	0.26	121.889	-0.649
Coulomb(SR)	-15674.60	0.52	187.706	1.784
Potential	-13192.00	0.28	106.709	1.135
Kinetic	2109.36	0.04	36.208	0.144
Total	-11082.70	0.31	101.474	1.279

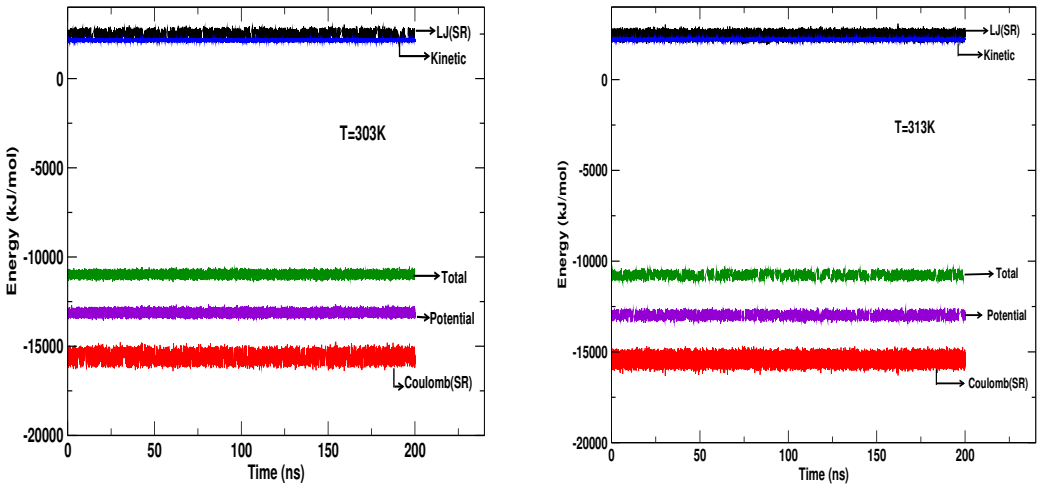


Figure 4.7: Energy profile of the system at temperatures 303K & 313K.

Table 4.5: Energy profile table at temperatures 303K and 313K.

Different energies at temperature 303K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2464.74	0.13	122.571	-0.243
Coulomb(SR)	-15585.10	0.19	187.813	1.062
Potential	-13120.30	0.20	106.795	0.819
Kinetic	2144.83	0.04	36.795	0.084
Total	-10975.50	0.19	101.359	0.903

Different energies at temperature 313K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2433.55	0.91	123.572	0.333
Coulomb(SR)	-15416.50	0.34	187.949	0.054
Potential	-12982.90	0.26	107.138	0.387
Kinetic	2215.68	0.04	37.899	-0.095
Total	-10767.30	0.26	101.417	0.292

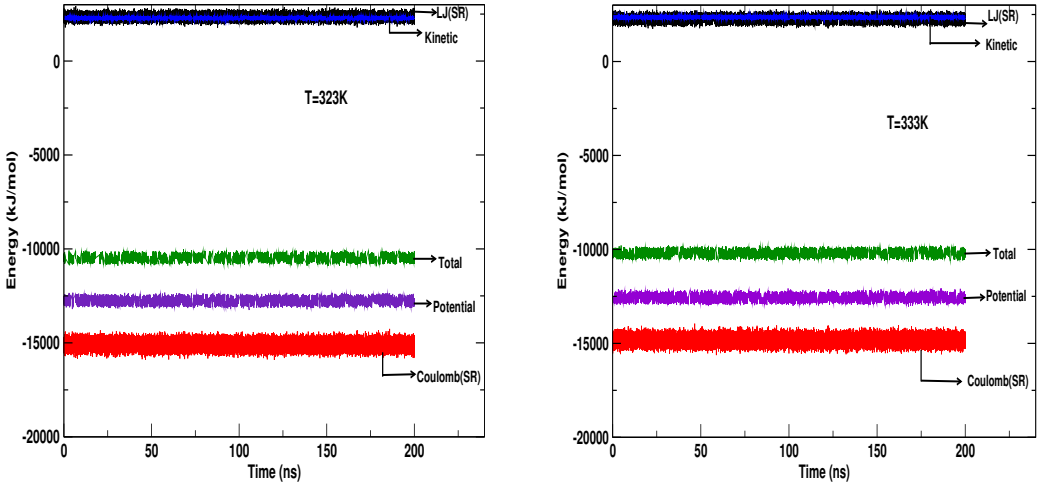


Figure 4.8: Energy profile of the system at temperatures 323K & 333K.

Table 4.6: Energy profile table at temperatures 323K and 333K.

Different energies at temperature 323K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2333.10	0.19	122.908	0.965
Coulomb(SR)	-15099.60	0.34	188.995	-1.664
Potential	-12755.50	0.14	109.582	-0.699
Kinetic	2286.30	0.02	38.869	0.035
Total	-10469.20	0.14	103.667	-0.664

Different energies at temperature 333K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2262.69	0.18	123.149	0.439
Coulomb(SR)	-14825.10	0.26	190.055	-0.767
Potential	-12562.40	0.15	111.267	-0.328
Kinetic	2357.18	0.03	39.888	0.161
Total	-10205.20	0.12	105.178	-0.167

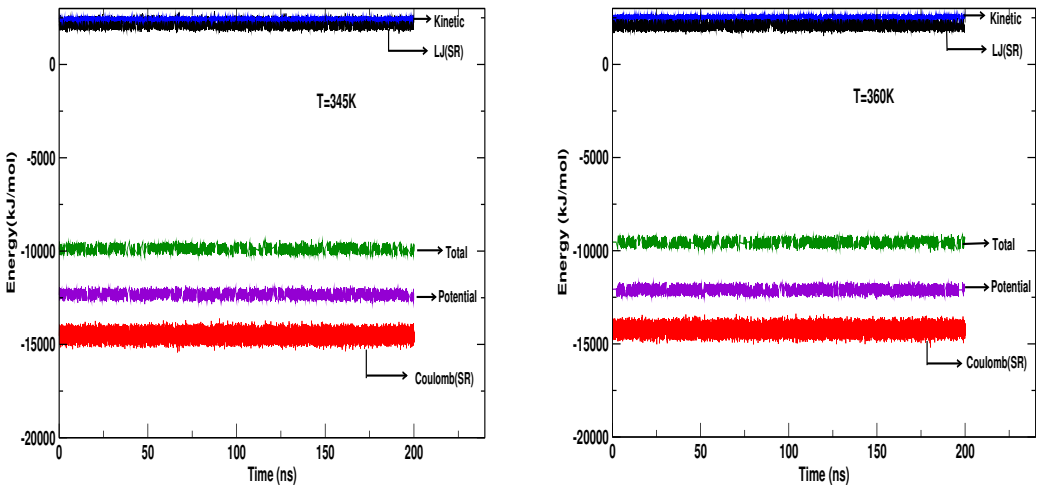


Figure 4.9: Energy profile of the system at temperatures 345K & 360K.

Table 4.7: Energy profile table at temperatures 345K and 360K.

Different energies at temperature 345K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2178.47	0.21	123.081	1.150
Coulomb(SR)	-14825.10	0.26	190.055	-0.767
Potential	-12562.40	0.15	111.267	-0.328
Kinetic	2357.18	0.031	39.888	0.161
Total	-10205.2	0.12	105.178	-0.167

Different energies at temperature 360K in KJ/mol				
Energies	Average	Err.Est	RMSD	Total-Drift
L-J(SR)	2116.33	0.17	124.104	-1.017
Coulomb(SR)	-14220.40	0.30	192.477	1.144
Potential	-12104.10	0.22	114.785	0.127
Kinetic	2548.33	0.026	42.546	-0.180
Total	-9555.73	0.22	108.070	-0.053

The error estimate of the different energies, root mean square deviation (RMSD) values and total-drift in various energies are provided in the tables (4.4 - 4.7). Energy drift is the gradual change in the energy of the system. One of the features of molecular dynamics simulation is that there is an unphysical drift in the energy of the system over long integration periods. The energy drift occurs due to numerical integration artifacts that arises with the use of a finite time step δt. Energy drift can also result from imperfections in evaluating the energy function usually due to simulation parameters that sacrifice accuracy for computational speed e.g. cut off schemes for evaluating electrostatic force, precision of the code [48].

The value of the Lennard-Jones potential (SR) decreases and the coulomb energy (SR) increases with the increase in temperature. The term SR indicates the short range which is quite obvious as we have used the truncated value for coulomb parameter and Lennard-Jones parameter as mentioned in the molecular dynamics parameter file and it encompasses all interactions that are not excluded. The positive value of the L-J potential is in consistent with the radial distribution function of water. The RDF of water (OW-OW) (Figure 4.2) shows that the position of the first peak is less than the value of σ_{ow-ow}, means most of the oxygen atoms lies in a distance less than the Lennard-Jones parameter σ. As we are familiar with the L-J potential versus distance curve that the potential is positive for the value of r less than σ and our system almost consist of water molecules which indicate the energy profile of the system is dominant of water energy, this explains the positive value of L-J potential of the system. The decrease of L-J potential can also be explained from the RDF of water. From table (4.1) it is clearly seen that the first peak value of the RDF decreases with the increase in temperature which means the probability of finding other oxygen atoms of water molecule with respect to the fixed oxygen atom at certain dis-

tance is in decreasing order with the increase in temperature. This means at low temperature large number of atoms take part in van der Waals interaction and increases the L-J potential in comparison to high temperature.

Potential energy is the sum of L-J and Coulomb energy and it increases with the increase in temperature. Kinetic energy also increases with the increase in temperature which is quite obvious as velocity of the particle increase with increasing temperature. The total energy, which is the sum of kinetic and potential energies is negative and increases with the increase in temperature. This characterizes the bound nature of the system and says the system is more stable at low temperature than in high temperature.

4.4 Diffusion Coefficients

We discuss the self-diffusion coefficient of each component present in the system under study. Our system is composed of water (as a solvent) and carbon monoxide gas. We use the values of self-diffusion coefficient of each component to determine the mutual diffusion coefficient using Darken's relation (equation (2.1)). The self-diffusion coefficient of the components are determined separately using Einstein's relation and Green Kubos formula (see chapter 2 for detail). The Green Kubo relation we use to calculate the self-diffusion coefficient is

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}_i(t)\mathbf{v}_i(0)\rangle_{i\in\alpha}\, dt \tag{4.1}$$

α denotes the either type of the component (H_2O or CO) and i is the particle number. The Einstein's relation for self-diffusion coefficient can be written as

$$D_\alpha = \lim_{t\to\infty} \frac{\left\langle (\mathbf{r}_\alpha(t+t_i) - \mathbf{r}_\alpha(t_i))^2 \right\rangle}{6\,t} \tag{4.2}$$

α denotes the either type of the component (H_2O or CO), t_i is any time origin, i is the particle number and $\langle \ldots \rangle$ signifies ensemble average has been taken. The ensemble average is over all atoms of the components α in the simulation and all time origins. The ensemble average at the instant of time t is defined as [49]

$$\left\langle (r_\alpha(t+t_i) - r_\alpha(t_i))^2 \right\rangle \simeq \frac{\Theta(T_{sim} - (t+t_i))\ (\mathbf{r}_\alpha(t+t_i) - \mathbf{r}_\alpha(t_i))^2}{\sum_{i=0}^{N-1} \Theta(T_{sim} - (t+t_i))} \tag{4.3}$$

where $t_i = i\Delta t$ is the starting time of the i^{th} segment of the trajectory obtained after simulating for time T_{sim}, Δt is the restart time, Θ is the Heaviside function.

Δt is made small for the optimal exploitation of data in the trajectory file. At short times, $0 \leq t \leq \Delta t$, the MSD is thus obtained as an average of the displacements observed in the N segments, while obviously less statistics is available at longer time [49].

To calculate the self-diffusion coefficient using Einstein's relation (MSD method) we use the command **g_msd** inherent in GROMACS, which generates the *.xvg* file in the format mean squared displacement as a function of time. Since the statistics is better due to higher averaging at the starting than towards the ending region, hence we take the certain portion which has the best statistics. We then fit linearly using *grace*, the slope divided by 6 gives the value of diffusion coefficient of the desired species under study. In our case we have a simulation time of 200 ns and the best statistics for CO molecule is found within 2 ns which is very small in comparison to simulation time this is due to lesser number of CO molecules. This time extends for (5-10) ns in the case of water, this is again due to the higher number of water molecules, in our case we have taken a time of 5 ns for the linear fit. The time Δt taken for calculation of mean square displacement is 10 ps.

To calculate self-diffusion using velocity auto-correlation function (Green Kubos formula) we have command **g_velacc** inherent in GROMACS. This generates a file in *.xvg* format with $C(\tau) = \langle \mathbf{v}_i(\tau).\mathbf{v}_i(0) \rangle_{i \in \alpha}$ as a function of time. As the information of velocity is considered important in this method, we need to have decent sampling of the velocities since they change rapidly, so the velocities should be written down at least every 10 fs during molecular dynamics. To get the consistent result for diffusion coefficient with MSD calculation we need to integrate $C(\tau)$ up to the time we linear fit the data in MSD calculation. To integrate we use the command **g_analyze** inherent in GROMACS. We calculate the diffusion coefficient using this method only for carbon monoxide and due to high computational cost we are unable to do the same for water molecules.

4.4.1 Diffusion Coefficient of Carbon monoxide

The plots of mean square displacement of CO molecule in various forms are shown in the succeeding figures.

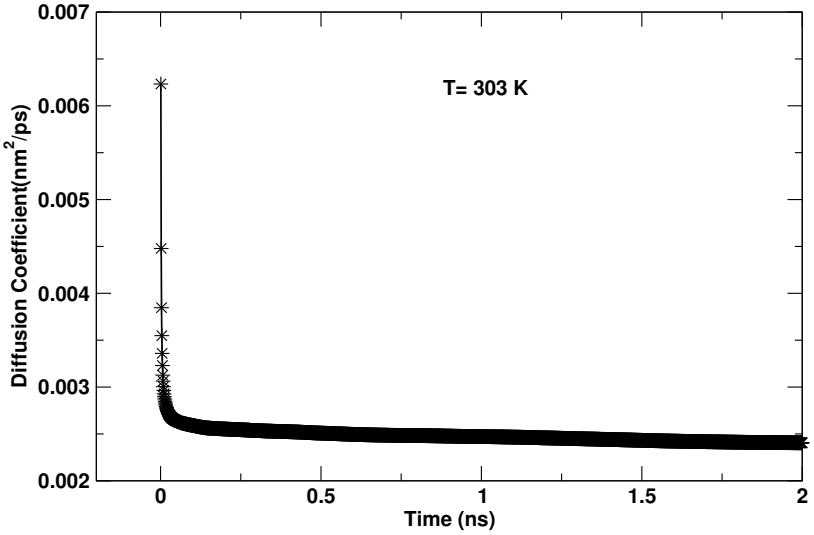


Figure 4.10: Plot of Diffusion coefficient vs time of CO at 303K.

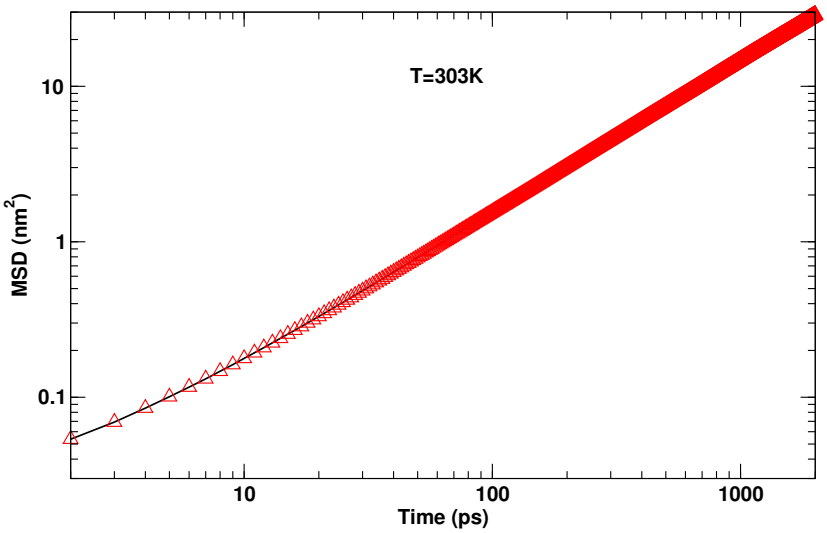


Figure 4.11: Logarithmic plot of MSD with time of CO at 303K.

Figure (4.10) shows the variation of diffusion coefficient with time. At first the diffusion coefficient is high due to ballistic motion and later as time passes it remains constant. This constant portion of the graph gives the diffusion coefficient. The ballistic region is also seen in figure (4.11) which is the log-log plot of mean square displacement with time and is represented by the parabolic region of the graph. In this case the molecules at the start move swiftly in the holes present in the system thereby showing higher value of diffusion coefficient. After certain time the molecules show uniform motion as a result of which we see the linear portion of the plot.

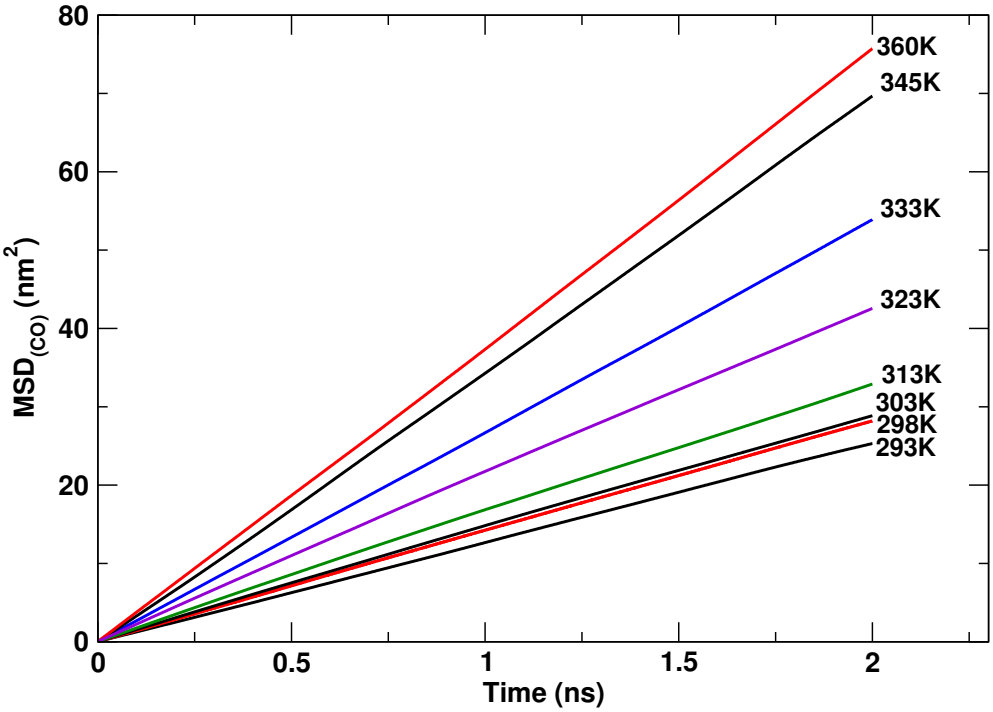


Figure 4.12: Plot of MSD vs time of CO at different temperatures.

Figure (4.12) shows the plot of mean square displacement (MSD) with time for different temperatures. As Einsteins relation mentioned in equation (4.2) has a linear relation of MSD with time so the figure (4.12) represents the appropriate truncated region to calculate self-diffusion coefficient. We first take the linear fit of the above data and then divide the slope by 6 to get the simulated value of self-diffusion coefficient. The comparison of diffusion coefficient with experimental value is shown in table (4.8). Figures (4.13, 4.14) show the variation of velocity autocorrelation function with time.

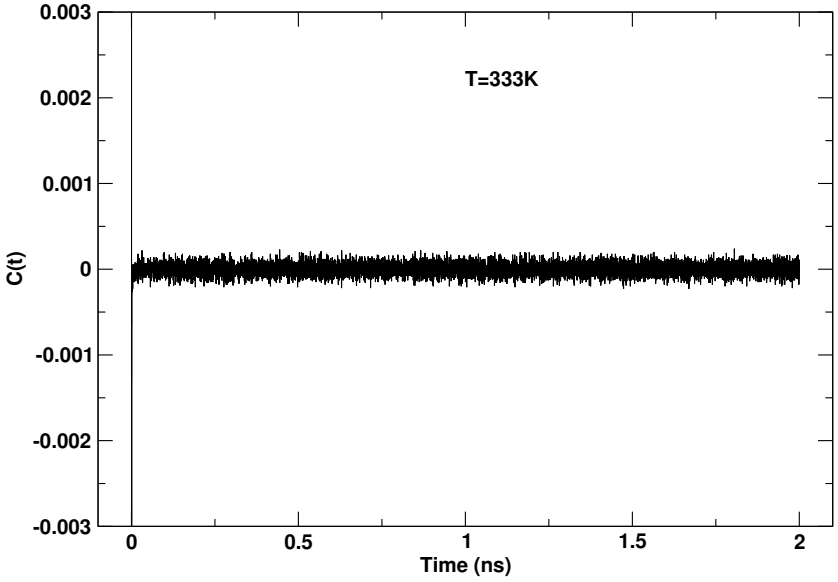


Figure 4.13: Plot of C(t) =($\langle \mathbf{v}_i(t).\mathbf{v}_i(0) \rangle_{i \in \alpha}$) vs time of CO at T=333K.

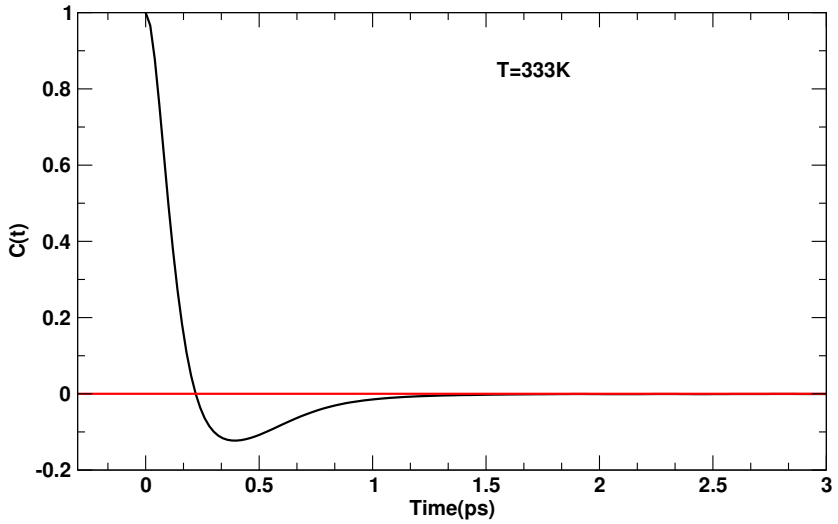


Figure 4.14: Plot of normalized C(t) vs time (truncated to 3 ps) of CO at T=333K.

Figure (4.13) shows the variation of velocity autocorrelation function with time at temperature T=333K. We look at the product of the two velocities at two instances that are apart by time t. If t is very small, the velocity doesn't change very much. Therefore, in the limit of small t the velocity at 0 and the velocity at t are going to be very similar, we can say that there is a lot of correlation which is seen in figure (4.13). As, we move away in time, there is going to be no correlation which is also revealed from figure (4.13). The velocity at 0 and the

velocity at an instant that is very far away in time is not going to be correlated. The nature of the graph for other temperatures is similar so we have taken above graph as a representative graph. We take the correlation of velocity at 0 to other velocities only up to 2 ns in order to match with our MSD calculation where we have truncated the data up to 2 ns. To calculate the self-diffusion coefficient we integrate the above graph and divide the result by 3. Figure (4.14) shows the normalized graph of figure (4.13) and the values are truncated to few ps time to see the variation of the velocity correlation function at small time interval. At very small time t the correlation is 1 as we have normalized the value, and after few ps the value shows there is no correlation. The comparison of the value obtained by using velocity autocorrelation method (VACF), MSD method and experimental value is shown in table (4.8).

Table 4.8: Experimental and Simulated value of Self-Diffusion coefficient of CO molecule.

Self - Diffusion Coefficient of CO ($1 \times 10^{-9} m^2 s^{-1}$)					
Temperature (K)	Experimental [25]	MSD method		VACF method	
		Simulated value	% Error	Simulated value	% Error
293	2.03	2.13	4.93	2.05	0.99
298		2.35		2.31	
303	2.43	2.40	1.23	2.35	3.29
313	3.62	2.77	23.48	2.82	22.10
323	4.11	3.53	14.11	3.48	15.33
333	5.68	4.48	21.13	4.59	19.19
345		5.83		5.92	
360		6.30		6.34	

Table (4.8) shows the experimental and simulated values of the self-diffusion coefficient of CO molecule at different temperatures. The simulated values of the self-diffusion coefficient of CO using MSD method varies within 2% maximum in comparison to VACF method. From this we can infer that the simulated values estimated using two methods (i.e. MSD and VACF) are in good agreement. The self-diffusion coefficient estimated using both the methods varies within 15% to that of experimental value [25] except at temperatures T=313K and T=323K.

The diffusion coefficient increases with the increase of temperature as expected. This is due to the increase in the velocity of the molecules because of the en-

ergy imparted to the molecules in accordance with the increase in temperature. Also, the density of the system decreases with increasing temperature and hence increasing the space available for the carbon monoxide molecules to execute random-walk motion [40]. Because of these, the mean squared displacement increases which by Einsteins relation yields an increased self-diffusion coefficient. The product of the velocity of CO at time 0 and velocity at other instant t increases with the increase in temperature. Then from Green Kubos relation (equation (4.1)) increases the value of self-diffusion coefficient. The dependence of VACF at two temperatures T=293K and T=360K is shown in figure (4.15). The possibility of a diffusion can also be explained by the RDF of OC-OC (Figure 4.5), as there is finite density of the particles in the region between the first peak and the second peak.

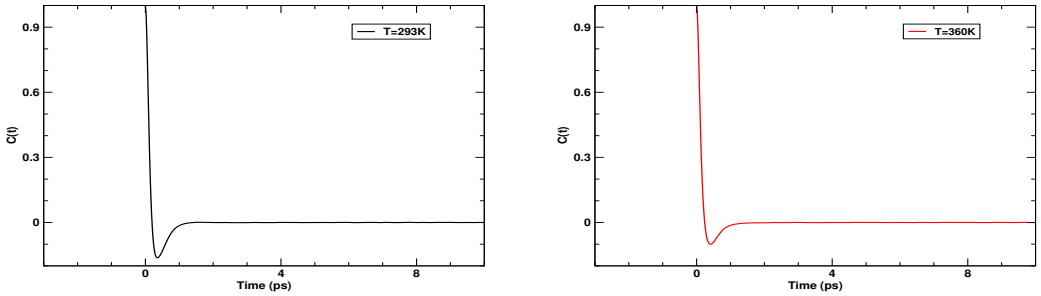


Figure 4.15: Plot of C(t) = ($\langle \mathbf{v}_i(t).\mathbf{v}_i(0) \rangle$) vs time of CO at T=293K (left) and at T=360K (right).

From figure (4.15) it is seen that at higher temperature the negative region of the graph decreases. If we integrate the whole region we get greater value at higher temperature due to lesser contribution from the negative region in comparison to smaller temperature. This ultimately increases the self-diffusion coefficient which is seen from Green Kubos relation (equation (4.1)).

4.4.2 Diffusion-Coefficient of Water

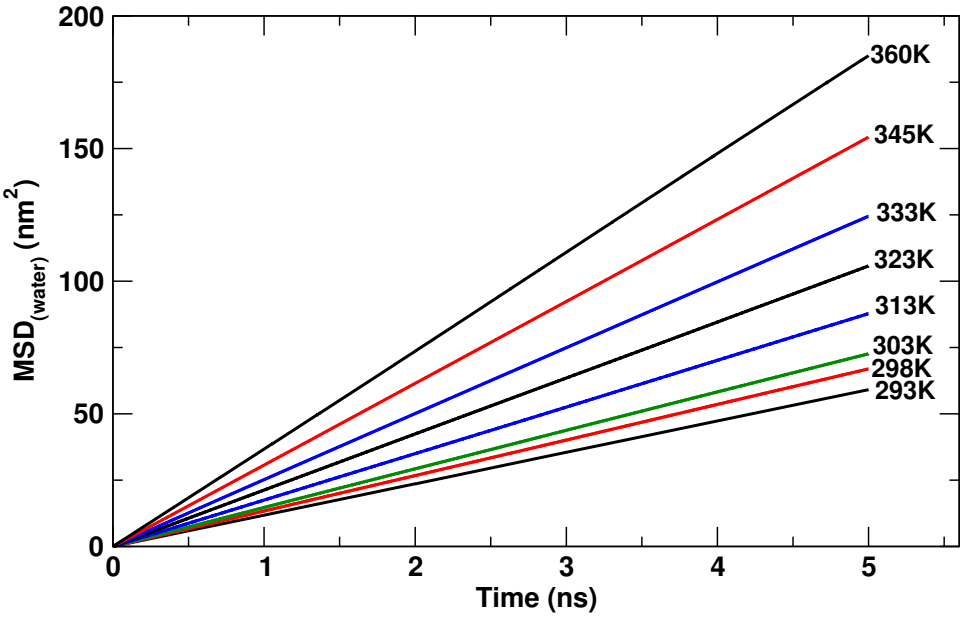


Figure 4.16: Plot of MSD vs time of water at different temperatures.

Figure (4.16) shows the variation of MSD with time of water at different temperatures. We linear fit the the above graph and obtain the self-diffusion coefficient of water at various temperatures by the similar process as described while calculating diffusion coefficient of carbon monoxide. The comparison of the simulated and the experimental value of the diffusion coefficient is shown in table (4.9).

Table 4.9: Experimental and Simulated value of Self-Diffusion coefficient of H_2O molecule.

Self-Diffusion Coefficient of H_2O ($1 \times 10^{-9} m^2 s^{-1}$)			
Temperature (K)	Experimental [50, 51]	MSD method	
		Simulated value	% Error
293	2.022	1.980	2.08
298	2.296	2.232	2.79
303	2.590	2.420	6.56
313	3.240	2.932	9.51
323	3.968	3.520	11.29
333	4.772	4.145	13.14
345		5.141	
360		6.181	

Table (4.9) compares simulated value of self-diffusion coefficient of water with experimental value. It is seen that the experimental and simulated value are in good agreement with the experimental value [50, 51] within about 13% maximum. It is seen that diffusion coefficient increases with the increase in temperature as expected. This is due to the increase in the velocity of the molecules because of the energy imparted to the molecules in accordance with the increase in temperature. Also, the density of the system decreases with increasing temperature and hence increasing the space available for the water molecules to execute random-walk motion [40]. Because of these, the mean squared displacement increases which by Einsteins relation yields an increased self-diffusion coefficient. With this we can say that the water model SPC/E we have chosen, to calculate the macroscopic properties like diffusion coefficient is appropriate model.

4.4.3 Mutual Diffusion Coefficient

Though our system consists of only 5 CO molecules the RDF of OC-OC (Figure 4.5) shows that the CO molecules do not move independently of each other. Therefore the mutual diffusion coefficient of CO in water D_{12} is estimated using Darken's relation (equation 2.1)

$$D_{12} = N_1 \, D_2 + N_2 \, D_1 \tag{4.4}$$

where N_1 is the mole fraction of water molecules, D_1 is the self-diffusion coefficient of water molecules, N_2 is the mole fraction of carbon monoxide molecules and D_2 is the self-diffusion coefficient of carbon monoxide molecules. Since the system consists of 5 CO and 280 H_2O molecules so N_1 is 0.982 and N_2 is 0.018.

Table 4.10: Mutual Diffusion Coefficient of Carbon monoxide in water.

Mutual Diffusion Coefficient of CO in H_2O ($1 \times 10^{-9} m^2 s^{-1}$)			
Temperature (K)	Self-Diffusion Coefficient		Mutual Diffusion Coefficient
	For CO	For H_2O	
293	2.130	1.980	2.127
298	2.350	2.232	2.348
303	2.400	2.420	2.400
313	2.770	2.932	2.773
323	3.530	3.520	3.530
333	4.480	4.145	4.474
345	5.830	5.141	5.818
360	6.300	6.181	6.298

We can infer from table (4.10) that the mutual diffusion coefficient is nearer to the diffusion coefficient of carbon monoxide. For the low solute concentrations studied in this work, the contribution to the mutual diffusion coefficient is nearly equal to the self-diffusion coefficient of the solute in the mixture. If the system is at infinite dilute condition then the mutual diffusion coefficient is equal to self-diffusion coefficient. This is not the case because there is a correlation of the carbon monoxide molecule in the system which is also shown by RDF of OW-OW $g_{ow-ow}(r)$.

4.5 Dependence of Diffusion on Temperature

Diffusion generally depends strongly on temperature, being low at low temperatures but appreciable at high temperatures. The temperature dependence of diffusion coefficients is frequently, but by no means always, found to obey the Arrhenius formula [5].

$$D = D_0 \exp\left(\frac{-E_a}{N_A\, k_B\, T}\right) \tag{4.5}$$

Equation (4.5) can be written as

$$\ln \mathrm{D} = \ln \mathrm{D}_0 - \frac{\mathrm{E}_a}{\mathrm{N}_A \, \mathrm{k}_B \, \mathrm{T}} \tag{4.6}$$

where D_0 denotes the pre-exponential factor, also called frequency factor, E_a is the activation energy for diffusion, T is the absolute temperature, N_A is Avogadros constant whose value is 6.023×10^{23} and k_B is the Boltzmann constant whose value is $1.38 \times 10^{-23}\ JK^{-1}$. Both E_a and D_0 are called the activation parameters of diffusion. The activation energy of diffusion process [5]

$$E_a = -N_A \, k_B \frac{\partial \ln \mathrm{D}}{\partial (1/T)} \tag{4.7}$$

corresponds to negative slope of the Arrhenius diagram. The intercept of the extrapolated Arrhenius line for $T^{-1} \Longrightarrow 0$ yields the pre-exponential factor D_0. In an Arrhenius diagram the logarithm of the diffusivity is plotted versus the reciprocal absolute temperature. To take the logarithm of diffusion coefficient it is made unitless by normalization.

The dependence of self-diffusion coefficients calculated using simulation method with temperature are shown in succeeding graphs. Figure (4.17, 4.18) shows the arrhenius diagram for the diffusion coefficient of CO obtained using MSD and VACF method from where we can see that it obeys the Arrhenius behavior. For water figure (4.19) the data lies in the fitted line from which we can conclude it obeys the Arrhenius relation with perfection.

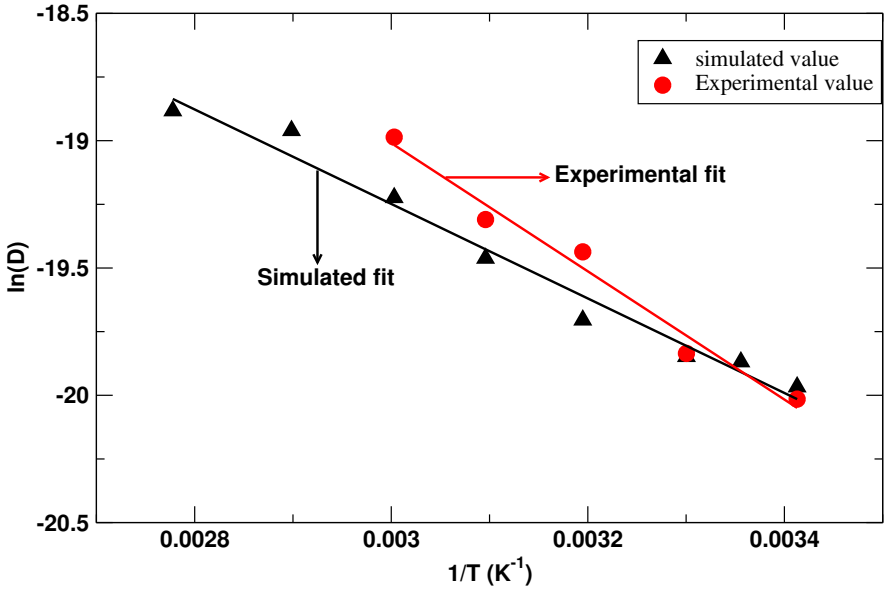


Figure 4.17: Arrhenius diagram* of the simulated (using MSD method) and experimental value of carbon monoxide.

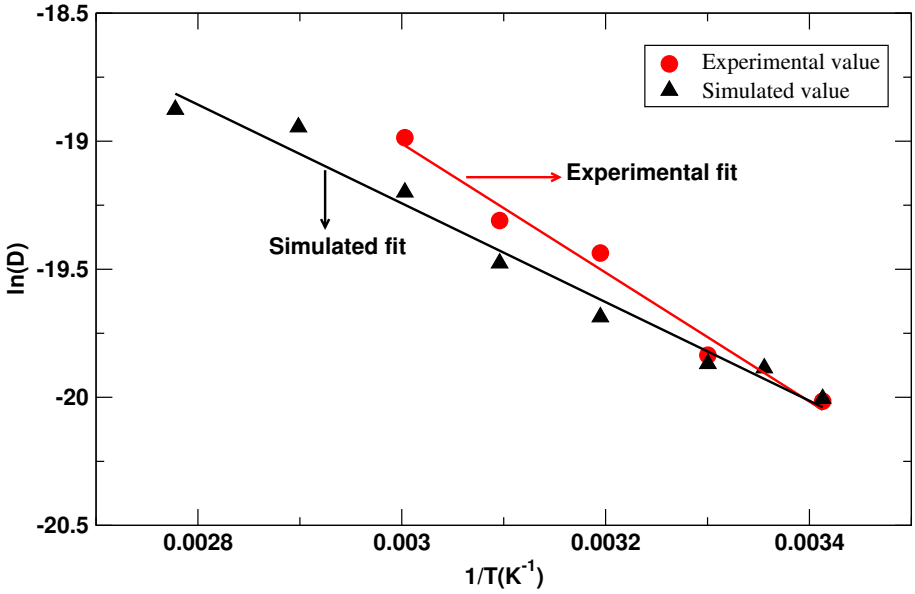


Figure 4.18: Arrhenius diagram* of the simulated (using VACF method) and experimental value of carbon monoxide.

* To obtain lnD, D is made dimensionless via normalization.

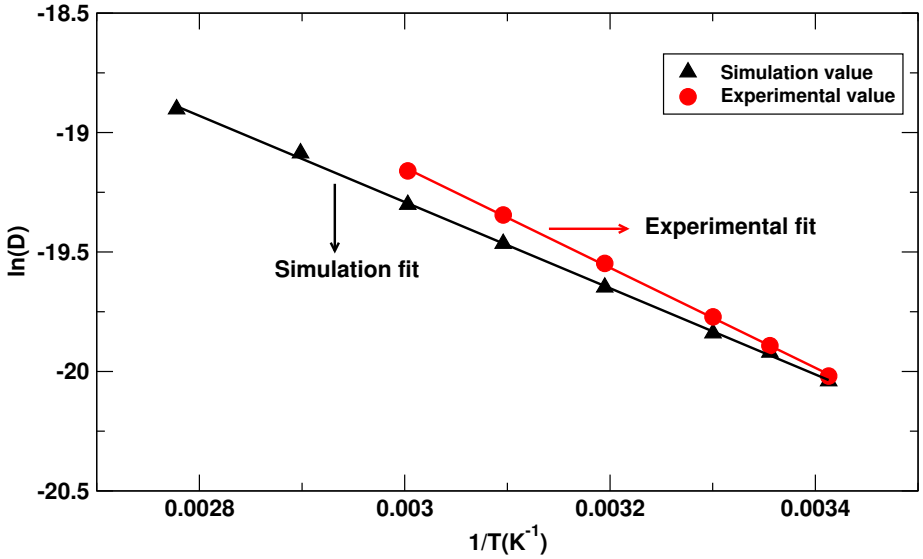


Figure 4.19: Arrhenius diagram* of the simulated and experimental value of water.

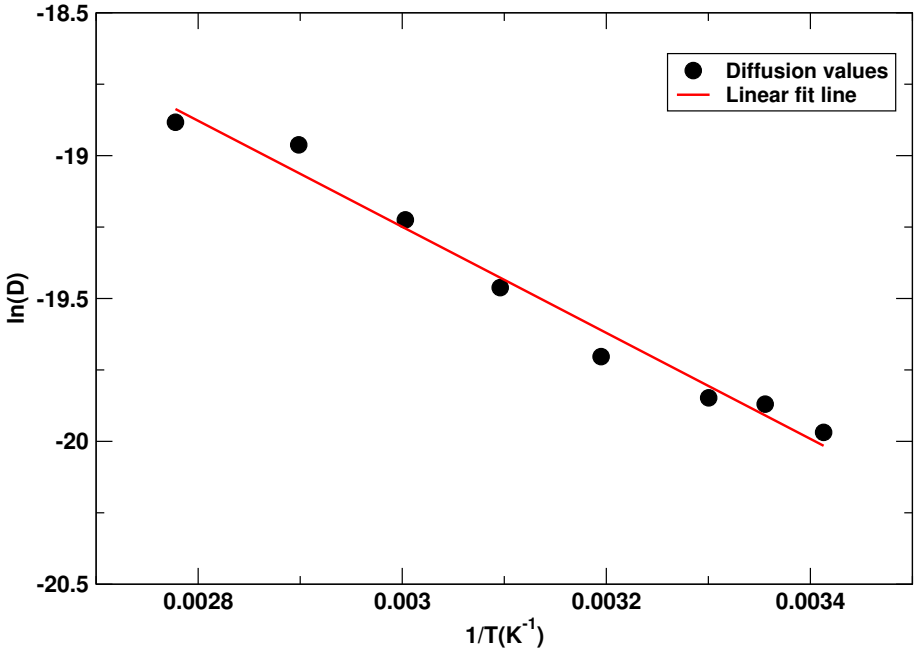


Figure 4.20: Arrhenius diagram* for mutual-diffusion of the system.

From figures (4.17-4.20) we can say that the temperature dependence of self-diffusion coefficient of CO and water and the mutual diffusion coefficient of CO in water follows the arrhenius behavior. The activation energy of diffusion is shown in table (4.11).

* To obtain lnD, D is made dimensionless via normalization.

Table 4.11: Table of Activation Energy.

Activation Energy in KJ mol^{-1}				
CO			Water	
MSD	VACF	Experimental	MSD	Experimental
15.43	16.02	20.93	15.01	17.41

The experimental value of activation energy in table (4.11) is the value obtained from the Arrhenius plot of the experimental data. We have plotted the Arrhenius diagram for both the experimental and simulated value, and table (4.11) shows the details of the plot in data. The activation energy of CO obtained using MSD method deviates by 4% to that obtained from VACF. It is seen that water and carbon monoxide activation energy deviates by 13% and 26% respectively in comparison to the experimental activation energy obtained from the Arrhenius plot of experimental data.

Chapter 5

Conclusions and Concluding Remarks

In this work, the system containing 280 water (H_2O) molecules and 5 carbon monoxide (CO) molecules were taken and the molecular dynamics study for various temperatures 293K, 298K, 303K, 313K, 323K, 333K, 345K, 360K were done. The Extended Simple Point Charge (SPC/E) water model was used. CO acts as a solute and water as a solvent. For simulation procedure GROMACS 4.6.1 software was used. The input parameters were taken so as to be consistent with the experimental values as much as possible. Our working procedure is divided into three important parts, energy minimization, equilibration and production run. Steepest descent method was used for energy minimization process. To equilibrate, the system was simulated for 200 ns in NPT ensemble using Berendsen thermostat and barostat. The system was equilibrated to the desired temperature and a pressure of 1 atm (1.01×10^{5} Pa). The system was then operated for production run using NVT ensemble for 200 ns with the time step of 2 fs. The energy profile of the system were studied to know the equilibrium nature of the system. For temperature 313K and 323K, the structure obtained after production run was again simulated in NVT ensemble to be ensured with the equilibrium nature of the system by which the total energy obtained vary within the error bar of the previous value.

The equilibrium structural properties of both the components (CO and H_2O) were studied calculating corresponding radial distribution function (RDF) namely $g_{ow-ow}(r)$ RDF of oxygen atoms of water molecules, $g_{oc-ow}(r)$ RDF of oxygen

atom of CO and oxygen atom of H_2O, $g_{oc-oc}(r)$ RDF of oxygen atoms of CO molecules. The RDF of $g_{oc-oc}(r)$ showed some convolutions and depression at short distance due to less number of CO molecules.

The main goal of our work was to study diffusion phenomenon of the mixture of water and CO. The self-diffusion coefficients of water was estimated using Einstein's method (MSD) and that of CO was estimated using Einstein's and velocity autocorrelation method (VACF) separately. The simulated values of the self-diffusion coefficient of CO using MSD method varies within 2% maximum in comparison to VACF method. The self-diffusion coefficient estimated using both the methods varies within 15% to that of experimental value [25] except at temperatures T=313K and T=323K, and that of water is deviated within 13% of the available experimental data [50, 51], which highlights the goodness of SPC/E water model. The binary or mutual diffusion coefficient of the system was calculated using Darken's relation. The Arrhenius diagram (plot of natural logarithm of diffusion coefficient versus inverse of temperature) was plotted for self-diffusion coefficient of CO & H_2O separately and it showed temperature dependence of self-diffusion coefficient of both is of Arrhenius type.

It can, therefore, be concluded that classical molecular dynamics simulation technique can be used as a reliable method to study the equilibrium structure and dynamic properties of fluid mixture. The calculated values of the diffusion coefficient may be used as a crude reference for any further fluid studies.

In future, we wish to use the long range van der Waals and Coulomb interaction to study the diffusion coefficient, and use supercritical water model for study. We want to use GROMACS to study the molecular dynamics of biomolecules and free energy calculations for molecular solids. We want to study the diffusion coefficients by varying concentration of the component over wide range of temperature and also study the dynamics of Methane, Ammonia, Krypton.

Bibliography

[1] M. Thompson, *Carbon Monoxide-Molecule of the month*, Winchester College, U.K, 2009.

[2] Z. Palai, H. Saricaoglu, A. Acar, Journal Of the European Academy Of Dermatology and Venereology, **9**, (1997), 152.

[3] *www.ghgonline.org*, 6$^{\text{th}}$ August 2013.

[4] D. R. Lide (Editior-in-Chief), *CRC Handbook of Chemistry and Physics*; CRC Press LLC; 2005.

[5] H. Mehrer, *Diffusion in Solids*, Springer Series in Solid State Science, Berlin, 155.

[6] V. A. Hermandaris, N. P. Adhikari, N. F. A. van der Vegt, K. Kremer, B. A. Mann, R. Voelkel, H. Weiss, C. C. Liew, Macro-molecules, **40**, (2007), 7026.

[7] N. P. Adhikari, X. Peng, A. Alizadeh, S. Ganti, S. Nayak, S. K. Kumar, Physical Review Letters, **93**, (2004), 188301.

[8] H. Mehrer, N. A. Stolwijik, *Heroes and Highlights in the History of Diffusion*; Open-Access Journal for the Basic Principles of diffusion.

[9] J. Groh, G. VonHevesy, Ann-Physik, **63**, (1920), 85.

[10] J. Groh, G. VonHevesy, Ann-Physik, **65**, (1921), 216.

[11] B. J. ALder and T. E. Wainwright, Physical Review Letters, **18**, (1967), 988.

[12] B. J. ALder and T. E. Wainwright, Physical Review A, **1**, (1970), 18.

[13] N. P. Adhikari, R. Auhl, E. Straube, Macromolecular Theory and Simulations, **11**, (2002), 315.

[14] N. P. Adhikari, E. Straube, Macromolecular Theory and Simulations, **12**, (2003), 499.

[15] A. R. Leach, *Molecular Modelling, Principles & Applications*, 2nd Edition, Pearson Education Limited, 2001.

[16] S. K. Thapa, N. P. Adhikari, International Journal of Modern Physics B, **27**, (2013), 1350023.

[17] U. Dahal, N. P. Adhikari, Journal of Molecular Liquids, **167**, (2012), 34.

[18] H. B. Moktan, A. Panday, N. P. Adhikari, International Journal of Modern Physics B, **26**, (2012), 1250016.

[19] A. Rahman and F. H. Stillincer, Journal of Physical Chemistry, **55**, (1971), 3336.

[20] G. Malenkov, Journal of Physics, Condensed matter, **21**, (2009), 283101.

[21] G. Malenkov, Y. Naberukhin, and V. Voloshin, Structural Chemistry, **22**, (2011), 459.

[22] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, *Gromacs User Manual version 4.5.6*, (2010).

[23] R. E. Smith, E. T. Friess, M. F. Moracles, Journal of Physical Chemistry, **59**, (1955), 382.

[24] E. L. Cussler, *Diffusion-Mass Transfer in Fluid Systems*, 2nd Edition, Cambridge University Press, U.K.

[25] D. L. Wise and G. Houghton, Chemical Engineering Science, **23**, (1968), 1213.

[26] Air Liquid Gas Encyclopedia *encyclopedia.airliquid.com.*

[27] D. C. Johnson, Respiratory Medicine, **94**, (2000), 28.

[28] J. L. Puebe, *Fundamentals of Fluid Mechanics and Transport Phenomena*, John Wiley & Sons, Inc. 2009.

[29] H. Hirakawa, Y. Kamei, M. Sugisaki, Y. Oishi, Bulletin of the Chemical Society of Japan, **46**, (1973), 2659.

[30] D. Keffer, *The Working Mans Guide to Obtain Self Diffusion Coefficients from Molecular Dynamics Simulations*, Department of Chemical Engineering, University of TN, Knoxville.

[31] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, U.S.A, 1989.

[32] L. S. Darken, AIME, **175**, (1948), 184.

[33] D. Frenkel, B. Smit, *Understanding Molecular Simulation From Algorithms to Applications*, Academic Press, USA, 2002.

[34] J. Crank, *The Mathematics of Diffusion*, Clarendon Press, 2nd Edition, 1975.

[35] J. P. Hansen, I. R. McDonald, *Theory of Simple Liquids*, 3rd Edition, Elsevier, USA, 2006.

[36] U. R. Dahal, *Molecular Dynamics study of diffusion of heavy water in ordinary water at different temperatures*, Msc Dissertation Central Department of Physics, Tribhuvan University, 2010.

[37] B. Widom, *Statistical Mechanics: A Concise introduction for chemists*, Cambridge University Press.

[38] J. M. Haile, *Molecular Dynamics Simulation Elementary Methods*, John Wiley & Sons, Inc. 1992.

[39] F. Ercolessi, *A Molecular Dynamics Primer*, International School for Advanced Studies (SISSA-ISAS), Trieste, Italy.

[40] S. K. Thapa, *Molecular Dynamics study of diffusion of oxygen in water at different temperatures*, Msc Dissertation Central Department of Physics, Tribhuvan University, 2010.

[41] www.fisica uniudit/ ercolessi/md.

[42] M. G. Campo, Papers in Physics, **2**, (2010), 20001.

[43] P. Kowalczyk, *Molecular insight into the high selectivity of double-walled Carbon nanotubes*; Journal of Physical Chemistry Chemical Physics.

[44] C. N. Banwell and E. M. McCash, *Fundamentals of Molecular Spectroscopy*; Tata McGraw Hill; 4^{th} edition.

[45] www.mrbigler.com/misc/Pvap-H_2O.PDF.

[46] D. A. McQuarrie, *Statistical Mechanics*, University Science Books, USA, 2000.

[47] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[48] http://gromacs.5086.x6.nabble.com/, *Gromacs User Forum.*

[49] V. Ballenegger, S. Picaud, C. Toubin, Chemical Physics Letters, **432**, (2006), 78.

[50] A. J. Easteal, W. E. Price and L. A. Woolf, J. Chem. Soc. Faraday Trans. **1**, (1989), 85, 1091.

[51] R. Mills, Journal of Physical Chemistry, **77**, (1973), 685.