TRANSPORT AND THERMODYNAMIC PROPERTIES OF AMOXICILLIN IN WATER

A THESIS SUBMITTED TO THE

CENTRAL DEPARTMENT OF PHYSICS

INSTITUTE OF SCIENCE AND TECHNOLOGY

TRIBHUVAN UNIVERSITY
NEPAL

FOR THE AWARD OF

DOCTOR OF PHILOSOPHY
IN PHYSICS

BY
SHYAM PRAKASH KHANAL

NOVEMBER 2021

TRANSPORT AND THERMODYNAMIC PROPERTIES OF AMOXICILLIN IN WATER

A THESIS SUBMITTED TO THE

CENTRAL DEPARTMENT OF PHYSICS

INSTITUTE OF SCIENCE AND TECHNOLOGY

TRIBHUVAN UNIVERSITY
NEPAL

FOR THE AWARD OF

DOCTOR OF PHILOSOPHY
IN PHYSICS

BY
SHYAM PRAKASH KHANAL

NOVEMBER 2021

DECLARATION

Thesis entitled **"Transport and Thermodynamic Properties of Amoxicillin in Water"** which is being submitted to the Central Department of Physics, Institute of Science and Technology (IOST), Tribhuvan University, Nepal for the award of the degree of Doctor of Philosophy (Ph.D.) is a research work carried out by me under the supervision of Prof. Dr. Narayan Prasad Adhikari (Central Department of Physics, Tribhuvan University) and co-supervision of Dr. Ali Hassanali (The Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy).

This research is original and has not been submitted earlier in part or full in this or any other form to any university or institute, here or elsewhere, for the award of any degree.

Shyam Prakash Khanal

RECOMMENDATION

This is to recommend that **Mr. Shyam Prakash Khanal** has carried out research entitled **"Transport and Thermodynamic Properties of Amoxicillin in Water"** for the award of Doctor of Philosophy (Ph.D.) in **Physics** under our supervision. To our knowledge, this work has not been submitted for any other degree.

He has fulfilled all the requirements laid down by the Institute of Science and Technology (IOST), Tribhuvan University, Kirtipur for the submission of the thesis for the award of Ph.D. degree.

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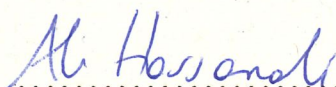
Dr. Narayan Prasad Adhikari
Supervisor
(Professor)
Central Department of Physics
Tribhuvan University
Kirtipur, Kathmandu, Nepal

....

Dr. Ali Hassanali,
Co-Supervisor
(Research Scientist)
The Abdus Salam International Centre for
Theoretical Physics (ICTP), Trieste, Italy

[November 2021]

LETTER OF APPROVAL

[Date: 30/11/2021]

On the recommendation of **Prof. Dr. Narayan Prasad Adhikari** and **Dr. Ali Hassanali**, this Ph.D. thesis submitted by **Mr. Shyam Prakash Khanal**, entitled **"Transport and Thermodynamic Properties of Amoxicillin in Water"** is forwarded by Central Department of Research Committee (CDRC) to the Dean, IOST, T. U..

. . . .

Dr. Om Prakash Niraula
Professor
Head
Central Department of Physics,
Tribhuvan University
Kirtipur, Kathmandu
Nepal

ACKNOWLEDGMENTS

and care. My family members always encourage and support to handle ups and downs in my life. At last, I want to acknowledge to all my relatives, friends and well-wishers for their inspiration in each and every step of my life.

. . . .

Shyam Prakash Khanal

November 2021

ABSTRACT

Out of prominent factors, dissolution rate and solubility of drugs have significant role in oral drug absorption. The solubility of drugs depends upon many factors including diffusion coefficient, solvation free energy in particular solvent environment. The solubility of drugs can be estimated from the calculation of solvation free energy. In this work, molecular dynamics study of amoxicillin, an antibacterial agent, has been carried out to understand its transport and thermodynamic properties in aqueous medium. We used water (SPC/E and TIP3P models) as solvent; and all atom OPLS force field parameters to model the solute.

We have studied the diffusion phenomenon of amoxicillin in water through the estimation of self diffusion coefficients of both solute as well as solvent and their binary diffusion coefficient using Einstein's and Darken's relations respectively. We have examined the effect of temperature on diffusion coefficients of the molecular system under study. For this, we performed simulations at different temperatures; and it has been observed that the coefficients follow Arrhenius's behavior. In addition, the estimated values of diffusion coefficient from simulations carried out under periodic boundary conditions are also affected by the size of the simulations box due to screening effect caused by long range hydrodynamic interactions. So, we have also analyzed the size dependency of diffusion coefficient by performing simulations taking systems of different sizes. It has clearly been noticed that the estimated values of diffusion coefficient from simulation are size dependent; and the estimated values increase with increase in size of simulation box. From this study, the size independent values of diffusion coefficient, i.e., diffusion coefficients estimated with system of infinite size of solvent as well as solution have been reported.

Besides the study of diffusion process, solvation free energy of amoxicillin in aqueous medium has been estimated at 310 K using thermodynamic integration (TI) and free energy perturbation (FEP) based methods: TI, TI-cubic, BAR and MBAR. Only non-bonded Lennard Jones and Coulomb interactions were manipulated to couple solute and solvent molecules. During the estimation of free energy difference between two thermodynamic states, concept of decoupling was used, i.e., initial state indicates that solute and solvent are fully coupled and final state means they are independent. The estimated values of free energy of solvation of amoxicillin in water from different approaches are in good agreement; and solvation energy has positive sign. This ensures that amoxicillin is soluble in water as expected. Also, we have analyzed the individual contribution of van der Waals and Coulomb interactions to solvation free energy; and

it has been found that Coulomb interaction has major contributions to the solvation of amoxicillin in water.

Motivating the influence of solvent on transport and thermodynamic properties of solute, we have studied the diffusion and solvation process of amoxicillin in other solvent i.e., ethanol. Both self and binary diffusion coefficients of amoxicillin in ethanol have been estimated at 298 K temperature; and free energy of solvation of amoxicillin in ethanol has been reported at 310 K temperature. We used aforementioned methods to study transport and thermodynamic properties of amoxicillin in ethanol.

Furthermore, we have reported the solvent accessible surface area (SASA) of solute and number of hydrogen bonds between solute and solvent molecules to get more insight into effect of solvent environments on solvation free energy. The large value of SASA as well as hydrogen bonds in water than in ethanol also support the higher estimated value of solvation free energy of amoxicillin in water in comparison to ethanol.

LIST OF ACRONYMS AND ABBREVIATIONS

AMBER   : Assisted Model Building with Energy Refinement
BAR   : Bennett Acceptance Ratio
CHARMM   : Chemistry at Harvard Macromolecular Mechanics
DNA   : Deoxyribo-Nucleic Acid
EM   : Energy Minimization
ER   : Excluded Region
FEP   : Free Energy Perturbation
FPP   : First Peak Position
FPV   : Fist Peak Value
fs   : Femtosecond
GROMACS   : Groningen Machine for Chemical Simulation
GROMOS   : GROningen Molecular Simulation
IUB   : International Union of Biochemistry
IUPAC   : International Union of Pure and Applied Chemistry
LAMMPS   : Large-scale Atomic/Molecular Massively Parallel Simulator
LINCS   : LINear Constraint Solver
LJ   : Lennard Jones
MD   : Molecular Dynamics
MDP   : Molecular Dynamics Parameters
MSD   : Mean Squared Displacement
NAMD   : Nanoscale Molecular Dynamics
NMR   : Nuclear Magnetic Resonance
NPT   : Constant Number, Pressure and Temperature
NVE   : Constant Number, Volume and Energy
NVT   : Constant Number, Volume and Temperature
PBC   : Periodic Boundary Conditions
PBPs   : Penicillin Binding Proteins
PDB   : Protein Data Bank
PME   : Particle Mesh Ewald
OPLS-AA   : Optimized Potentials for Liquid Simulations-All Atom
RDF   : Radial Distribution Function
RNA   : Ribonucleic Acid

SASA	: Solvent Accessible Surface Area
SPC/E	: Extended Simple Point Charge
SPP	: Second Peak Position
SPV	: Second Peak Value
TI	: Thermodynamic Integration
TIP3P	: Transferable Intermolecular Potential with 3 Points
TIP4P/2005	: Transferable Intermolecular Potential with 4 Points/2005
vdW	: van der Waals
VMD	: Visual Molecular Dynamics
WHAM	: Weighted Histogram Analysis Method

LIST OF SYMBOLS

k_B : Boltzmann constant

k_{ij}^b : Force constant for bond stretching potential

b_{ij}^0 : Equilibrium bond length

k_{ijk}^θ : Force constant for bond angle potential

θ_{ijk}^0 : Equilibrium bond angle

k_ϕ : Force constant for dihedral angle potential

δ : Equilibrium dihedral angle in degree

k_ξ : Force constant for improper dihedral angle potential

ξ_0 : Equilibrium improper dihedral angle in degree

ϵ_0 : Permittivity of free space

ϵ_r : Dielectric constant

ϵ : Strength of potential in L-J potential

σ : Length scale in L-J potential

k_i : Force constant for i^th window in umbrella sampling

LIST OF TABLES

LIST OF FIGURES

xv

TABLE OF CONTENTS

CHAPTER 1

1. INTRODUCTION

1.1 General Consideration

Bacterial infections cause several types of diseases like Pneumonia, Meningitis and so on. There are wide varieties of agents available for the treatment of bacterial infection. A bacteriostatic agent does not kill bacteria but does inhibit their reproductive growth; whereas a bactericidal agent actually kills bacteria. Antibiotic inhibits the growth or even destroys micro organism (Nogrady et al., 2005). Out of many antibiotics, amoxicillin is a semi-synthetic broad-spectrum antibiotic of penicillin family and is widely used in the treatment of bacterial infections (Danelon et al., 2006). When an amino group is added to penicillin, an effective antibiotic is created. This combination of amino group and the penicillin, termed as aminopenicillins for example amoxicillin, can fight against the biotic infection (Todd & Benfield, 1990).

Amoxicillin is in the class of β-lactam antibiotics; in which β-lactam ring is responsible for antibacterial activity. The bacterial cell-wall is strengthened due to cross-linked D-alanyl-alanine portion of peptide in the presence of penicillin binding proteins (PBPs) (Bush & Bradford, 2016). The β-lactam ring inhibits the process of cross-linking in cell wall synthesis of both Gram-negative and Gram-positive bacteria by binding with penicillin binding proteins (PBPs) and thus interfering cell wall synthesis (Rice, 2012; Kapoor et al., 2017). The semi-synthetic penicillins are classified in to three types: acid-resistant alternative to penicillin G, penicillinase resistant penicillins and the extended spectrum penicillin (Tripathi, 2013). Out of these three categories, amoxicillin belongs to the third one.
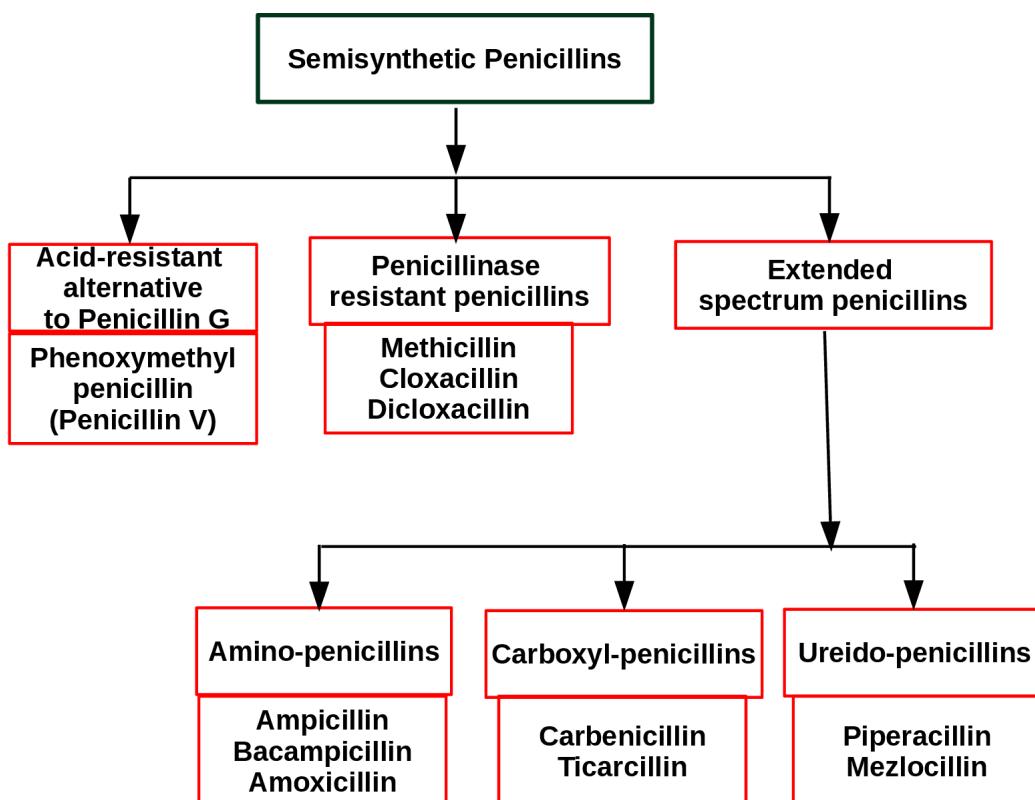
Figure 1: Classification of semisynthetic penicillin.

Orally administered drugs are absorbed at different parts of digestive tract. Absorption begins from mouth with the maximum amount absorbed through the intestinal tissues and finally enters into the portal vein. Oral drug absorption is basically affected by three factors: solubility, dissolution and intestinal permeability. Taking solubility and permeability in account, the drugs are classified into four classes: class I (High solubility and high permeability), class II (Low solubility and high permeability), class III (High solubility and low permeability) and class IV (Low solubility and low permeability). Amoxicillin is a class III type of antibiotic. The permeability of this drugs can be enhanced by co-administration of bioenhancers, the agents which themselves are not therapeutic entities but when combined with an active drug, it accelerate the pharmacological effect of drugs (Barve & Ruparel, 2015).

1.2 Composition of Amoxicillin

Amoxicillin is a drug molecule with molecular formula $C_{16}H_{19}N_3O_5S$ and IUPAC name - (2S,5R,6R)-6-[(2R)-2-4(4-hydroxyphenyl)-acetyl]amino-3,3-dimethyl-7-oxo-4-thia-1-az-abicyclo[3.2.0]heptane-2-carboxylic acid. The molecular weight, melting point and aqueous solubility of amoxicillin are 365.404 g/mol, 194 °C and ≈ 4 mg/mL respectively. It has 5 double bonds, 46 σ-bonds, 5 π-bonds, one carboxyl group and one

hydroxyl group. It has the following mass percentages for different atoms: 52.591 % of carbon (C), 5.241 % of hydrogen (H), 11.499% of nitrogen (N), and 8.752 % of sulphur (S). The molecule has four hydrogen bond donor and seven hydrogen bond acceptor. Amoxicillin is one of the widely used antibiotics.

Amoxicillin is white or almost white powder/crystalline solid having a bit sulphurous odour. Amoxicillin sodium is highly soluble in water, sparingly soluble in anhydrous ethanol, slightly soluble in acetone, while Amoxicillin trihydrate is slightly soluble in water, slightly soluble in ethanol (96 percent), practically insoluble in fatty oil. It dissolves in dilute acids and dilute solutions of alkali hydroxides (Kaur et al., 2011). As the solubility of amoxicillin sodium is very high, it is used for intra-venous infusion. Due to low solubility and high stability, amoxicillin trihydrate is used for oral suspension. Figure shows the 3-D geometry (projected in plane) of an amoxicillin molecule.
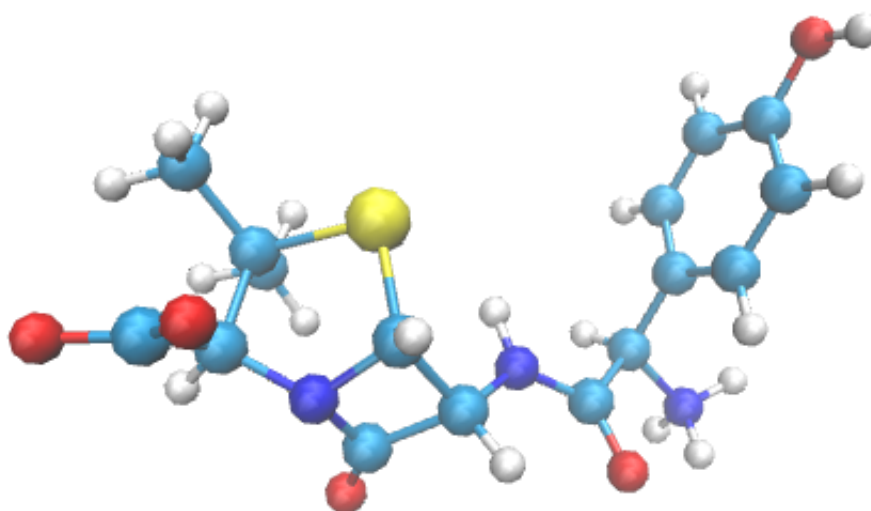
Figure 2: Amoxicillin molecule.

Amoxicillin is stable in presence of gastric acid. It has solubility ≈4 mg/L in water. Several studies were carried out to study about the stability of Amoxicillin after dissolution in tap water. Amoxicillin dissolves fast, has a high solubility and a good stability if pH is buffered at 8 (Boeren et al., 2006).

1.3 Mechanical Action in Bacteria

The cell wall of a bacteria consists of peptidoglycan. Figure 3 is the schematic representation of formation of peptidoglycan (Humphrey et al., 1996). It is a polymer of N-acetylmuramic acid (NAM) alternating with N-acetylglucosamine (NAG). The amino acid polymers bridge the neighbouring strands of peptidoglycan. In each amino acid polymer, five amino acids are linked together, L-Alanine (L-Ala), D-Glutamic acid (D-Glu), L-Lysine (L-Lys) and two D-Alanine (D-Ala) (Vollmer et al., 2008). The

cross-linking occurs from the third position amino acid that is attached to the first strand of peptidoglycan to the carboxyl group of D-Ala at the fourth position of polypeptide in the next strand via directly or 5 glycine residues (Höltje, 1998). The energy for binding of two polypetides of two strands is provided in the expense of cleavage of fifth position D-Ala in both the strands. When this process is repeated, long thick cell wall is formed in the bacterial cell. Thus, the cross-linking provides the stability and rigidity of the cell wall and prevents osmotic pressure that acts from cytoplasm of the bacterial cell. An enzyme named transpeptidase stimulates and mediates the process of the amino acid linkage between two peptedoglycan strands (Scheffers & Pinho, 2005). The β-lactam
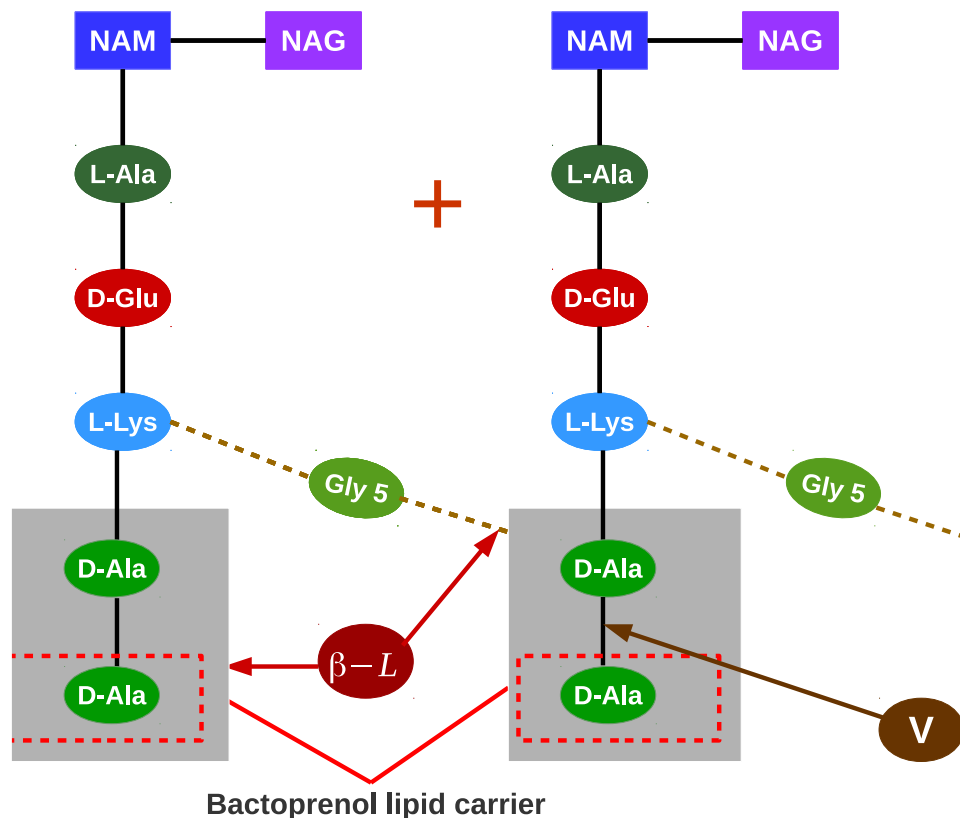
Figure 3: Formation of peptidoglycan with amino acid polymer.

antibiotics prevent the binding of polypeptide side chains by deactivating the action of transpeptidase. The transpeptidase and the corresponding amino acids under the linkage region constitute the penicillin (amoxicillin is a type of penicillin) binding proteins (PBPs). Since the penicillin inhibits the formation of cell wall, the osmotic pressure inside the cell wall tends to swell the cell and then burst out. Finally, the bacteria die out.

1.4 Bacterial Infections and Antibiotics

Bacteria are the small single-celled microorganisms. They are found everywhere on the earth. Most of the bacteria are harmless, but a few of them are infectious. When bacteria enters into the body, they spread throughout due to their fast breading ability. They invade the body via several openings that link the interior part of body to the surroundings. Swelling in infected parts is a common symptom of bacterial infection. When harmful bacteria enter the body, the defensive mechanism of our body kills them. Specifically, white blood cells (WBCs) fight against bacteria. The immune system easily copes and fights with the infection caused by such deteriorative biomolecule (Karthikeyan & Meyer, 2006). Even though immune system fights against such harmful lives, the system fails in some situations. Our body's immune system is affected if the multiplication of bacteria is excessively high. So, external medication is essential. In such conditions, antibiotics are useful (Karthikeyan & Meyer, 2006). Antibiotics are the agents that fight against bacteria. They are used for the prevention and treatment of biotic infections. Before the discovery of antibiotics (before 1920), people died of even a minor bacterial infection.

The discovery of antibiotics is considered as a breakthrough in the field of medicine. Golden era of antibiotics development was between 1940s to 1960s when most the currently used antibiotics were developed. After 1970s new antibiotics developments were faltering so that only few antibiotics are in discovery pipelines. A wide use of natural penicillins resulted in the emergence of penicillinase (beta-lactamase) producing strains among gram positive bacteria and, thus, resulting antibiotic resistance. This directed the research towards development of beta-lactamase-resistent semi-synthetic penicillins (i.e., second generation penicillins) such as oxacillin, methicillin, and dicloxacillin. To overcome the narrow spectrum of second generation penicillins, in 1960s, the third generation and broad spectrum penicillins also known as aminopenicillins were introduced. Amoxicillin and ampicillin are the major antibiotics of this group which have broad spectrum coverage including gram-negative bacteria (as Hemophilus influenza, Escheria coli, Salmonela spp. and Sigella spp.) (Sutherland, 1964). The last generation of penicillins which include carboxypenicillins and ureidopenicillins had further broaden spectrum covering Gram-negative bacteria as Pseudomonas aeruginosa (Lobanovska & Pilla, 2017).

After the invention of antibiotics, life expectancy of people increased, surgeries got softer and life became easier. There are basically two types of antibiotics: bactericidal antibiotic and bacteriostatic antibiotic. The bactericidal antibiotic kills bacteria, whereas the bacteriostatic antibiotic stops the multiplication of bacteria. Antibiotic is not effective against the viral diseases.

1.5 Medical Uses and Side Effects

Amoxicillin is used to treat many types of infections like ear infection, throat infection, pneumonia, skin infection, urinary tract infection etc. It is a common medicine for people of all ages from childhood to old age. It is mostly taken orally, and sometimes from intravenous injections. It is diffused easily into the body tissues and fluids. It is used alone or in combination with clavulanic acid. In some cases, amoxicillin alone is superior to that combination with clavulanic acid, but the combination is more effective and has low side effects in several other cases (Gresser, 2001). The side effects as skin reaction are higher if amoxicillin alone which can be reduced by combining it with clavulanate acid. But the combination has greater unwanted effects on gastrointestinal, hepatic and haematological system. This combination has also higher risk of hepatic problems, diarrhoea, nausea, vomiting, loose stools, and abdominal discomfort (Caron et al., 1991).

Nowadays, medical chemists have a challenge to find or make compounds that have potency with minimum side effects on the host cells. In general, chemists used trial and screening method to design and synthesize the bioactive molecules that is very expensive. But the present need of chemists is the fast access to reliable and accurate information on the synthesized molecules with accurate mathematical model of chemical structures and their interactions with drug receptor sites. In this direction of development, molecular dynamics can be used as a fundamental technique. Molecular dynamics can be used to study internal molecular motion and to find out the structure of medium sized molecules (Mosher, 1992). In medical sciences, the study about interactions of drugs with targets and their dynamics pay great attention. Even though, many experimental and theoretical techniques can be used to understand the structural, transport as well as thermodynamic properties of biological systems, molecular dynamics simulation also has been routinely used to understand such properties in molecular level. Among many tools, computer simulations including Molecular Dynamics (MD) simulations have gained on ever increasing role in addressing key structural, dynamical, thermodynamic and kinetic features at a molecular level (Harmandaris et al., 2007; Malloci et al., 2015). Dynamical events may vary the functional properties of biomolecules. Synthesis of macromolecules is very expensive. Computer simulation can be the alternative way to study about the properties of macromolecules commonly used in the pharmaceutical industry (Ercolessi, 1997). Various experimental studies for the improvement of quality of Amoxicillin have been conducted so far but the researches regarding the measurement of its diffusivity, viscosity, etc. have not been noticed. The present work may play a significant role to understand the transport and thermodynamic properties of amoxicillin in molecular level.

1.6   Rationale of the Study

The majority of biological phenomena happen in aqueous environment. The under-standing about transport, thermodynamics properties like diffusion, viscosity, free en-ergy difference can play a vital role in many disciplines of science including industrial process. Diffusion, a transport property, is the process of transfer of matter from one part to another part of system as a result of random molecular motion (Crank, 1979). In living organisms, diffusion is a crucial phenomenon (Khanal et al., 2019). It is due to the consequence of constant thermal motion of atoms, molecules and particles moving from high concentration to low concentration. The transfer rate is affected by many fac-tors including temperature, viscosity of fluid and the size (mass) of the particles as well system used during simulation. The dynamics of biomolecules in human body happen through diffusion process. The kinetic term; dissolution rate of drugs depends upon the diffusion coefficient, a transport term, and thermodynamic term: solubility (Noyes & Whitney, 1897). Such phenomenon provides idea about inter/intra atomic/molecular interactions (Umecky et al., 2006). However, the experimental measurement of diffusion rate is difficult.

The knowledge about solubility has significant role in many areas including industrial process (Amidon et al., 1995). The aqueous solubility has important role in pharmaceu-tical industries (Yalkowsky & Valvani, 1980). Out of many factors, solubility is a major factor affecting bioavailability of drugs. The solubility of drug in a particular medium also controls the rate of dissolution and quantity that is dissolved in the medium (Noyes & Whitney, 1897; Skyner et al., 2015; Dizaj et al., 2015). There are many approaches to study about the solubility including MD simulations. The solubility of a solute in solvent can be estimated from the knowledge of free energy of solvation (Matos et al., 2017; Bergström & Larsson, 2018). Also, we can estimate the relative solubility of a solute in two different solvents from the estimation of free energy of solvation (Bellucci et al., 2019). The free energy of solvation measures the interactions of solute with solvent environment; and can be estimated from the change in free energy between the two states i. e., the state of fully interacting solute and solvents through different interactions and state of no interactions between them at definite temperature and pressure (Dasari & Mallik, 2020).

The role of different computational methods including molecular dynamics (MD) sim-ulations has been increasing in different stages of drug designing (Mortier et al., 2015; De Vivo et al., 2016). With the availability of high speed computational power, computa-tional methods have routinely been used in the discovery of new drug candidate (Borhani & Shaw, 2012). It becomes a powerful alternative approach which also provides guide-lines for experimental study. Simulation technique can also be used for those extreme

conditions like high pressure and high temperature in which it is difficult to perform experiments. The simulations play a role of bridge between theoretical and experimental results (Allen & Tildesley, 1987). During simulations, the macroscopic properties of interest like transport properties, free energy calculations etc. can be extracted from microscopic properties like masses of atoms, interactions between them and so on the system. Simulations can be taken as a complement to get more insight about results obtained from theoretical and experimental methods (Frenkel & Smit, 2002).

1.7 Objectives of the Study

In human body, the dynamics of biomolecules happen through diffusion process. The physiological transport of digested food materials, hormones and enzymes that occur in living cell is due to diffusion (Bhandari & Adhikari, 2016). The delivery of drug and its efficiency in our body involves the diffusive transport of molecules through membrane. Out of many factors, solubility and dissolution rate of drug are the major factors affecting oral drug absorption. The solubility can be estimated from the knowledge of free energy of solvation. The diffusion rate and solubility affect the dissolution rate. So, the study about the transport and thermodynamic properties of drugs has significant role. The general objective of this work is:

- To study the transport and thermodynamic properties of amoxicillin

And, the specific objectives are:

- To study the diffusion processes and calculate the diffusion coefficient of amoxicillin in water at different temperatures

- To study the effect of system size on diffusion coefficient

- To estimate the solvation free energy of amoxicillin in water

1.8 Organization of the Thesis

The structure of this thesis is organized as follows:

 (i) In chapter 2, we present literatures which provide a background to set up objectives of this work.

 (ii) Chapter 3 describes the theoretical background required to complete this work. In this chapter, we present in brief the Fick's law to explain diffusion phenomenon, Einstein's relation to estimate self diffusion coefficient, Darken's relation to estimate binary diffusion coefficient and effect of system size on diffusion coefficient. We also discuss the free energy perturbation (FEP) and thermodynamic integra-

8

tion methods to estimate the free energy of solvation. This chapter is named as Materials and Methods.

(iii) The main findings of this work are presented and discussed in the chapter 4 entitled 'Results and discussion'. Section 4.1 introduces the background for the whole chapter. We discuss the structural analysis through the calculation of radial distribution function (RDF) between different molecules, the estimation of diffusion coefficients at different temperature, size dependence of diffusion coefficient and estimation of solvation free energy of amoxicillin in different solvents using different approaches in sections 4.2, 4.3, 4.4 and 4.5 respectively.

(iv) We briefly summarize the conclusions with possible extension of the work in chapter 5. Finally, the summary is presented in chapter 6 followed by the references.

CHAPTER 2

2. LITERATURE REVIEW

In this chapter, we briefly discuss some literatures relevant to the present work. Firstly, we review literatures related to antibiotics followed by amoxicillin, a penicillin type of antibiotic. Then we present literatures that are focused on diffusion and thermodynamic phenomena. We also discuss the literatures related to molecular dynamics study.

During last few decades, many research works were performed about physiochemical properties of different antibiotics. The term antibiotics had been originated from French word antibiose which was used as an antonym to symbiosis to describe the antagonistic action between different microorganisms, coined by Paul Vuillemin (Vuillemin, 1889). First man made antibacterial agent was Salvarsan (arsphenamine) which was synthesized by chemist Alfred Bertheim in the laboratories of Paul Ehlrich who had been working in the dyes that specifically stain bacterial cell wall. It was approved in 1910 as a drug and was used for the treatment of syphilis (Ehrlich & Bertheim, 1912).

Discovery of penicillin in 1929 by Alexander Fleming superseded the sulphonamides which were the first truly effective broad spectrum antimicrobials and are still in use (Fleming, 1929). However, he could not produce purified penicillin and it almost took 10 years for its development in purified form. In 1939, Howard Florey with his team in Oxford University, successfully purified penicillin and tested it against streptococcus in mice (Chain et al., 1940). It was major breakthrough that enabled the development of semi-synthetic derivatives which help to bypass the penicillin resistance. Alexander Fleming, Ernst Boris Chain and Howard Walter Florey jointly awarded Nobel prize in 1945 for the discovery of penicillin with its therapeutic value (Raju, 1999).

Amoxicillin was discovered in 1972 which differ from ampicillin structurally by addition of a hydroxyl group in the benzene ring. Addition of the hydroxyl group results in a drug that is more lipid soluble and thus increased bioavailability and duration of action (Sutherland et al., 1972). Although amoxicillin was first synthesized in 1970, it was formulated in tablet in 1981 in United Kingdom (Geddes et al., 2007). Figure 4 shows chart of antibiotic development (Hutchings et al., 2019).
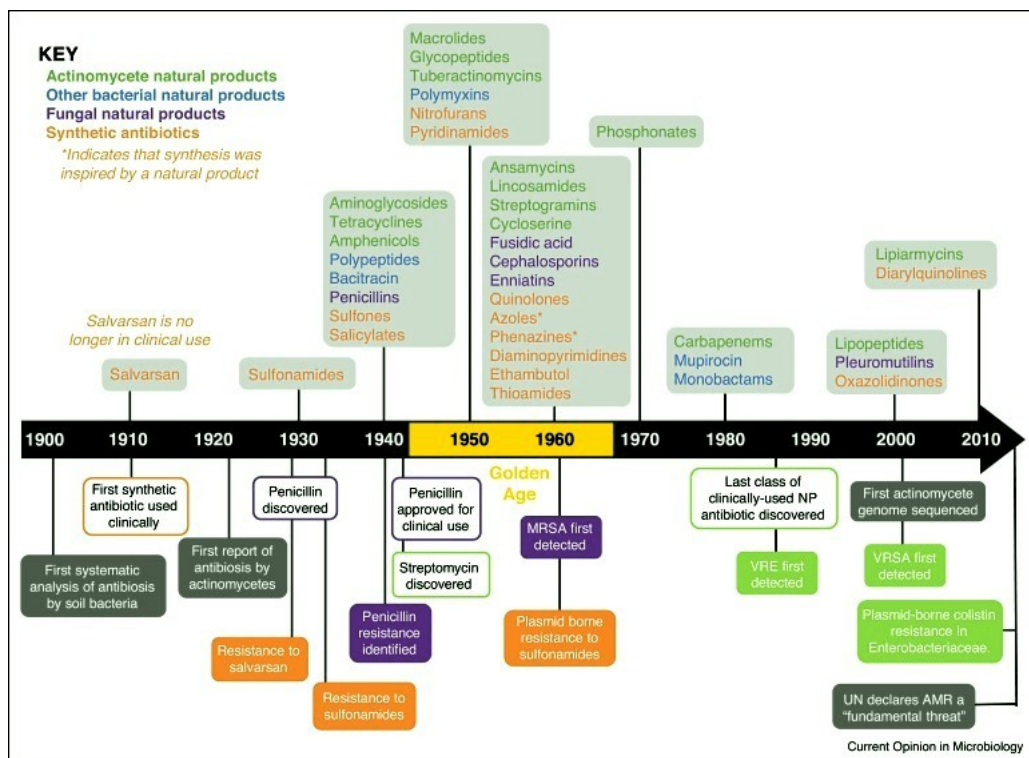
Figure 4: Flow chart of time-line development of antibiotics.

Mainly, the research works were focused on the study of solubility, permeability and diffusivity of antibiotics to enhance the potency of drugs against bacteria. Bodey and Nance studied the antibacterial activity of amoxicillin and compared it to that of ampicillin and cephalothin (Bodey & Nance, 1972). They reported that amoxicillin was more effective over ampicillin on the basis of serum levels produced after oral administration. The comparison of antibacterial activity between amoxicillin and ampicillin were also carried out by Sutherland and co-workers; and they suggested amoxicillin show similar antibacterial activity as ampicillin (Sutherland et al., 1972; Rolinson, 1973). The absorption after oral dose for ampicillin is 30-55 % with 25-45 % recovered unchanged in urine and peak serum concentration of 2-6 mg/L after 500mg dose. Oral absorption of amoxicillin was better with bioavailability of 77-90 % with 50-80 % drug recovered unchanged in urine and peak serum concentration of 7-10 mg/L after 500 mg oral dose (Brogden et al., 1975).

The absorption half life of amoxicillin was reported 0.72 hours by Arancibia et al. (1980). The electrochemical behavior of amoxicillin in aqueous medium at pH 10.5 using ferrocenedicarboxylic acid as mediator in Carbon nanotubes (CNTs) was studied by Fouladgar et al. and they reported that oxidation of amoxicillin takes place at a potential of ≈ 0.5 V (Fouladgar et al., 2011). Amoxicillin distributes well into liver, lungs, prostate, muscle, middle ear effusions, maxillary sinus secretions, bone, gallbladder, bile and into ascetic and synovial fluids, however, poorly penetrate into cerebrospinal

fluid (Huttner et al., 2020).

The crystal structure analysis of amoxicillin was performed by Boles and co-workers (Boles et al., 1978). They reported that amoxicillin has orthorhombic unit cell with $P2_12_12_1$ space group having dimensions a = 15.677 Å, b = 18.785 Å and c = 6.645 Å. It has four molecules per unit cell. They also reported the coordinates, bond length and bond angles between atoms. Bebu and co-workers reported the optimized molecular geometry and molecular electrostatic potential using density functional theory (DFT) with B3LYP functional and 6-31G(d) basis set (Bebu et al., 2011). They also reported the FTIR, FT_Raman and SERS spectrum determined from quantum calculation and using spectroscopies; and also compared the experimental and theoretical spectrum. The crystal structure of amoxicillin was studied by Santos et al. using NMR crystallography (Santos et al., 2013). Similarly, the optimized geometry of amoxicillin and ampicillin as well as their IR spectra were reported by Kariper using quantum calculations (Kariper, 2017). NMR data, molecular electrostatic potential as well as molecular docking study of the molecules were also reported in this article.

The dissociation constant of 7 different β-lactam antibiotics in water as well as mixture of acetonitrile-water were evaluated using spectrophotometric and chromatographic method by Demiralay and co-workers (Demiralay et al., 2012). On the other hand, Crea et al. (2012) reported the solubility of amoxicillin and other two penicillin type antibiotics in water as well as aqueous solution of NaCl salt experimentally using shake-flask method. Similarly, the solubility of amoxicillin at different pH and temperatures in mixtures of ethanol and water was calculated and compared with previously reported values by Felix et al. (2016). They also analyzed the effect of pH and ethanol concentration on the solubility of amoxicillin.

The idea about interaction of drugs with amino acids in molecular level has significant role to understand the transport of drugs through membrane and other drug action. Singla and co-workers studied the interactions of amoxicillin with some amino acids: glycine, L-alanine, L-valine and L-leucine at three different temperatures 305.5 K, 310.5 K and 315.5 K and they reported that the interaction increases with increase in molar mass as well as concentration of amino acids (Singla et al., 2014).

Hancock and co-worker studied the antibiotic uptake into gram-negative bacteria. Antibiotics taken up into gram negative bacteria face two major diffusion barriers, the outer and cytoplasmic membrane. This study is concerned with the outer diffusion barrier. In this study, the uptake in porin-deficient and porin-sufficient mutants compared. It was found that particularly β-lactam pass across the outer membrane through the water filled channel of protein called porins in porins sufficient mutants (Hancock & Bell, 1989). Furthermore, Kalyani Bavre and co-workers focused their study about the improvement

of bioavailability of amoxicillin and they observed that the bioavailability of amoxicillin can be significantly increased by combining with bioenhancers like Piperine and Ginger resis (Barve & Ruparel, 2015).

The effects of the macromolecular solute on the translational mobility of surrounding solvent water, and Na$^+$ and Cl$^-$ ions were studied by Valdimir A. Marker and Co-workers using Molecular Dynamics. In this study, the average diffusion coefficient as a function of distance from the closest solute atom of myoglobin and DNA decamer was determined. It was observed that overall diffusion rate at the interface is lower than in bulk (Makarov et al., 1998). Furthermore, Jean and co-workers focused their review on the problem of bacterial adaptation to reduce influx of antibiotics through porins. The study was concentrated on the bacterial response towards antibiotic stress on altered membrane permeability and the recent molecular approaches for the improvement of the physio-chemical parameters that govern the translocation of antibiotics through porin channel (Pages et al., 2008).

Himmelblau reviewed in detail about the diffusion of dissolved gases (slightly soluble in water) in water (Himmelblau, 1964). The author explained the diffusion theory as well as different experimental methods including diaphragm cells, gas absorption in steady-state laminar flow systems, ringbom apparatus, interferometric techniques to determine diffusion of dissolved gases. Also, he presented the data analysis of diffusion with effect of temperature, pressure and concentration.

Molecular Dynamics

In 1957, Alder and Wainwright published first paper based on molecular dynamics simulations (Alder & Wainwright, 1957; Ercolessi, 1997). They studied phase behavior of hard sphere system from simulation carried out in rectangular box under periodic boundary conditions. Gibson and co-workers used finite difference method to solve equation of motion during simulations (Gibson et al., 1960). Rahman, a pioneer work in 1964, reported the diffusion coefficient of liquid argon estimated from molecular dynamics simulations performed with pairwise Lennard Jone potential (taking parameters: $\epsilon/k_B = 120$ K and $\sigma = 3.4$) (Rahman, 1964). Simulation was carried out in cubic box of size 10.229 σ under periodic boundary conditions; and estimated value of diffusion coefficient of liquid argon at 94.4 K is 2.43×10^{-6} cm^2/s using mean square displacement of particles. In this work, Rahman also explained the pair correlation and auto correlation functions between argon atoms. For hard sphere, the detail about velocity autocorrelation function was explained by Alder and Wainwright (Alder & Wainwright, 1967). The dynamical behavior of pancreatic trypsin protein was studied using molecular dynamics simulation by McCammon et al. (1977). Similarly, Levitt and Warshel

performed simulation to understand the protein folding mechanism (Levitt & Warshel, 1975).

Shivakumar et al. (2010) reported the solvation free energy of 239 different molecules in neutral state in SPC water model using free energy perturbation (FEP) method. In this paper, the authors also compared the solvation free energy of 13 different molecules in different water models: SPC, TIP3P and TIP4P. The solvation free energy of mono-, di- and tri- ethylene in two solvents: water and methane was presented from MD simulations using FEP, BAR and TI methods (Olsen et al., 2016). The simulations were carried at 1 and 80 atm pressure at 298 K and 283 K respectively. The detailed methodology to calculate free energy of molecular solids using MD used presented by Noya et al. (Noya et al., 2008). They also reported the free energy of ice using Einstein's molecule approach and also compared with the previously reported value from Einstein's crystal method. A software to perform molecular simulation and analysis trajectory: GROMACS was developed by Bekker and co-workers (Bekker et al., 1993; Lindahl et al., 2001; Hess et al., 2008). They also explained different algorithms to solve equations of motion. Aragones and co-authors illustrated the calculations of free energy of atomic solids using GROMACS and LAMMPS (Aragones et al., 2012). Similarly, in 2013, the detail of methodology and procedure to estimate the free energy of molecular solids using GROMACS was explained by Aragones and co-workers (Aragones et al., 2013). Authors used Einstein molecule approaches; and applied this methodology for the determination of free energy of solid methanol and water.

In 1935, Kirkwood explained the thermodynamic integration method to evaluate the free energy (Kirkwood, 1935). Similarly, Zwanzig described the perturbation method to estimate thermodynamic properties in 1954 (Zwanzig, 1954). The detailed about the estimation of free energy difference between two states using simulations was explained by Bash and co-workers (Bash et al., 1987). The authors also reported the solvation free energy of amino acids as well as analogues side chain in TIP3P water model at 300 K under 1 atm pressure. Furthermore, theoretical explanation regarding the free energy differences was elaborated by Jarzynski (1997). Klimovich and co-workers explained the analysis tools to estimate the free energy using simulation (Klimovich et al., 2015). They explained the different approaches of free energy calculations based on thermodynamic integration and free energy perturbation based methods as well as reported the idea about convergence of free energy calculations.

Adhikari and co-workers from the Central Department of Physics already carried out MD study to estimate diffusion coefficients of heavy water, oxygen, nitrogen, carbon monoxide, nitric oxide etc. in water (Dahal & Adhikari, 2012; Thapa & Adhikari, 2013; Sharma & Adhikari, 2014; Poudyal & Adhikari, 2014; Pokharel et al., 2016). DNA protein interaction was studied by Adhikari and co-workers from MD simulations (Koirala,

Pokhrel, et al., 2021). They also reported the change in binding free energy during interactions using umbrella sampling method.

Bellucci and co-workers used MD technique to estimate the solubility of paracetamol in ethanol at different temperatures from the estimation of absolute solid free energy using GROMACS software package and compared with experimentally determined values (Bellucci et al., 2019). On the other hand, Dasari and Mallik performed MD study of LASSBio-294, poorly water soluble drug, in water and other ionic solvents to understand the solvation process (Dasari & Mallik, 2020). They reported the solvation free energy as well relative solubility of the drug at 298.15 K temperature; and also analyzed the individual contributions of van der Waals and Coulomb contributions to solvation free energy. The all-atom force field parameters of 40 different antimicrobial compounds including antibiotics and beta-lactamase inhibitors were reported by Malloci et al. (2015). They also studied many properties including number of hydrogen bonds, hydrophobic and hydrophilic surfaces etc. of the compounds using molecular dynamics simulations.

Research Gap

From the review of relevant literatures, it has been observed that many experimental works were carried out to understand the solubility of amoxicillin in different medium. The crystal structure of amoxicillin was already reported. Similarly, optimized geometry and different spectrum of amoxicillin were studied by different groups using DFT calculations. Although the experimental studies have been going on about the calculation of solubility and other thermodynamic parameters, the solvation free energy as well as transport properties of amoxicillin in aqueous medium have not been performed in molecular level. The estimation of free energy difference between two thermodynamic states of drugs pay great attention to understand their binding mechanism with targets; and also solubility of drug in particular solvent can be understood from the solvation free energy of the drug in that solvent. Also, from the knowledge of transport phenomena of drug, we can get more information about the transport of drug to target area. In this context, the study of thermodynamic and transport properties of amoxicillin is relevant to get more insight about solvation process in molecular level.

CHAPTER 3

3. MATERIALS AND METHODS

3.1   General Consideration

In this chapter, we have discussed the methodology used to accomplish the objectives of the present work. The chapter follows the following organization. In section 4.1, we have discussed about radial distribution function (RDF) to analyze structural properties. Similarly, we have addressed about diffusion theory: Fick's law of diffusion, Einstein's equation to measure diffusion coefficient and effect size of system on diffusion coefficient in section 4.2. Section 4.3 deals with theory related to estimate free energy of solvation. Also, details about molecular dynamics simulations have discussed in section 4.4. Finally, the different systems under study are presented in section 4.5.

Computer simulations technique like molecular dynamics has been widely used in large areas of research with availability of high speed computers. Simulations can be used to understand about the structural, dynamics, thermodynamics and so on properties of drugs, macromolecules like protein, DNA, RNA etc. (Allen & Tildesley, 1987; Ercolessi, 1997). Methods of simulations can also be applied in those extreme conditions like high pressure which are not feasible in experiments (Frenkel & Smit, 2002). The experimental study of some phenomenon like diffusion, is difficult. Computer simulations can be taken as complements to get knowledge about the phenomenon like diffusion whose experimental study is difficult. It also provides guidelines for experimental study. In the computer experiment, we first prepare a model of the system under consideration and follow a recipe to calculate the different properties of interest (Adhikari et al., 2004). Although simulations techniques are widely used, the validity of results obtained from the simulations depends on the accuracy of models and validity can be checked by comparing the results with previously reported experimental values. Figure 3.2 shows schematic diagram of connection among experimental study, theoretical study and computer simulations (Allen & Tildesley, 1987).
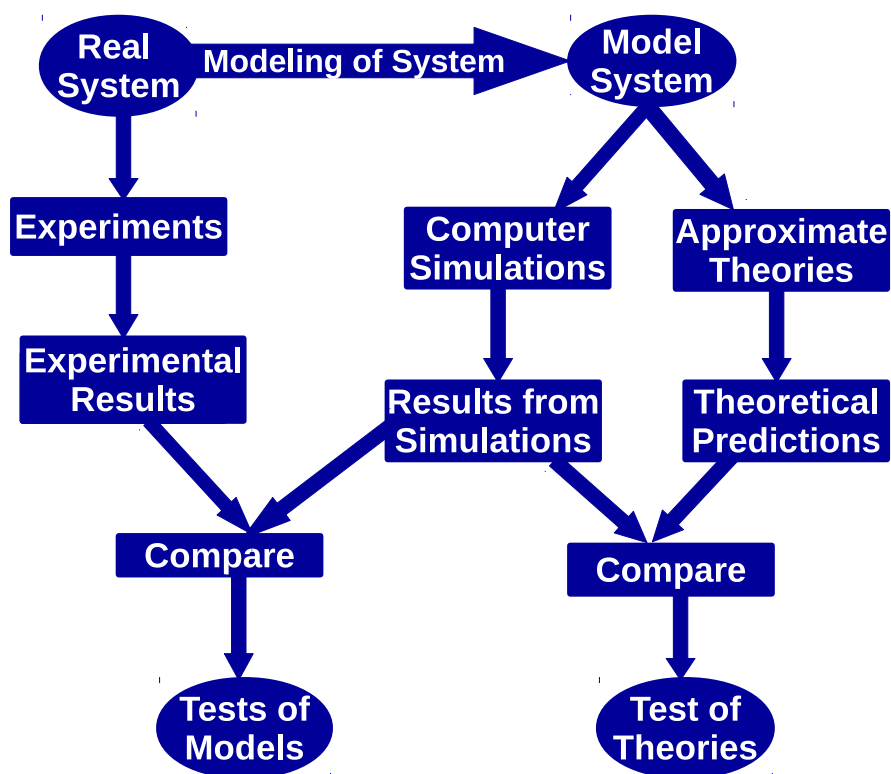
Figure 5: Schematic diagram showing the connection of computer simulations with experiment and theoretical study.

3.2 Radial Distribution Function (RDF)

The radial distribution function (RDF), which gives the probability of finding a pair of atoms located at distance 'r', has been used to analyze the structural properties of system (Allen & Tildesley, 1987). The RDF provides idea about the distribution of molecules around another molecule (McQuarrie, 2000; Hansen & McDonald, 2013). The fourier transform of RDF provides the structure factor. X-ray diffraction technique can be used to study the RDF, liquid exhibits more diffuse pattern of RDF. Alternatively, we can estimate the RDF between molecules from molecular dynamics simulations technique. Figure 6 shows the RDF for liquid argon (Chandler, 1987).

For liquids, the RDF shows an oscillatory in nature (McQuarrie, 2000; Hansen & McDonald, 2013). The oscillation persists upto certain distance from reference molecule and after than the probability distribution function becomes unity, which indicates that no correlation between two molecules. The region from $r = 0$ up to which the RDF becomes zero is known as excluded region (ER).

The radial distribution function $g_{XY}(r)$ between particles of type X and Y is defined in
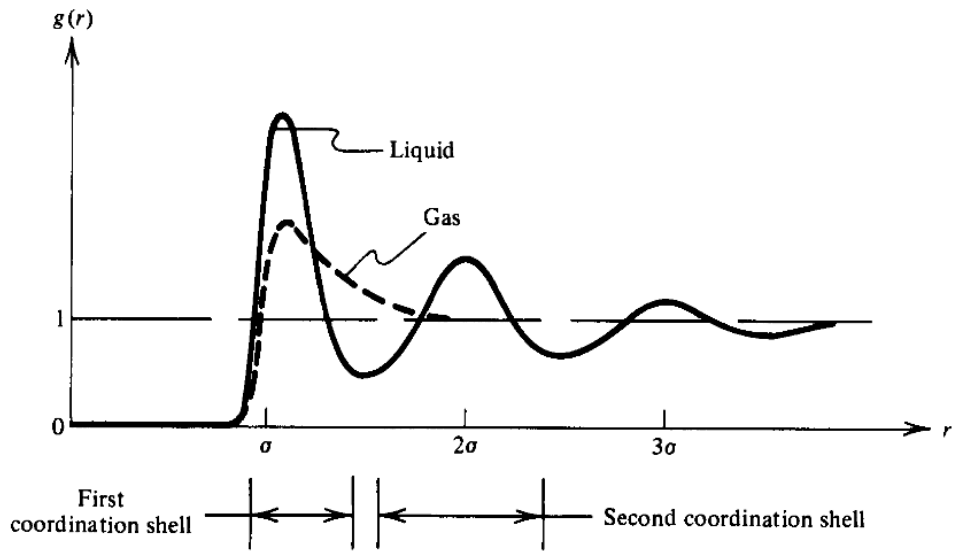
Figure 6: RDF between atoms of liquid argon.

the following way (David et al., 2005):

$$g_{XY}(r) = \frac{\langle \rho_Y(r) \rangle}{\langle \rho_Y \rangle_{\text{local}}} = \frac{1}{\langle \rho_Y \rangle_{\text{local}}} \frac{1}{N_X} \sum_{i \in X}^{N_X} \sum_{j \in X}^{N_Y} \frac{\delta(r_{ij} - r)}{4\pi \, r^2} \qquad (3.1)$$

In this equation, $\langle \rho_Y(r) \rangle$ is the particle density of type Y at a distance r around particles X and $\langle \rho_Y \rangle$ is the particle density of type Y averaged over all the spheres around particles X. Figures show different peaks in RDF.
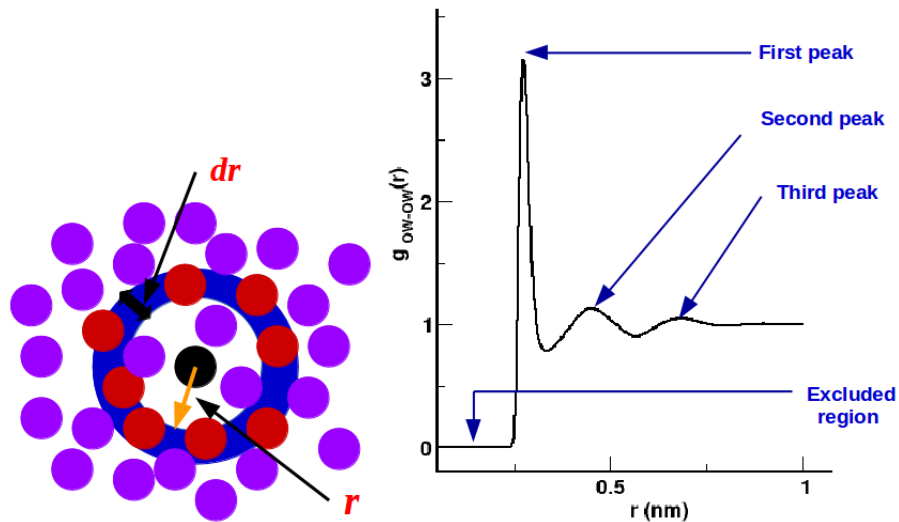
Figure 7: Schematic diagram of RDF (left); and RFD between two oxygen atoms of water taking SPC/E model (right).

18

3.3 Diffusion Theory

Diffusion, a transport phenomenon, is the process of transfer of matter from the region of higher concentration to lower concentration region. Diffusion occurs in a homogeneous medium without chemical concentration gradient is self diffusion and diffusion of the constituent particles occurs in binary mixture is binary diffusion (Frenkel & Smit, 2002). Such phenomenon is measured in terms of diffusion coefficient; the response of a system to a perturbation relates the particle flux to a concentration gradient. The experimental measurement of diffusion coefficient which relates the particle flux to a concentration gradient is very difficult as well as expensive. In this context, molecular dynamics simulation has been considered to be the best alternative way to study diffusion phenomenon (Poudyal & Adhikari, 2014).

In continuous system, the diffusion coefficient is defined by Fick's law. According to Fick's law, the rate of particle flux, the rate of transfer of diffusing substance through unit area of section, is proportional to the concentration gradient measured normal to the surface i.e., (Fick, 1855)

$$\boldsymbol{J} = -D\nabla C(\boldsymbol{r}, t) \tag{3.2}$$

where, $C(\boldsymbol{r}, t)$ is the concentration of diffusing substance which is function of position $\boldsymbol{r}$ and time; and D is proportionality constant known as diffusion coefficient. The significance of negative sign in Equation (3.2) is that the diffusion occurs in the direction opposite to that of increasing concentration. If both terms $\boldsymbol{J}$ and $C(\boldsymbol{r}, t)$ in the equation are expressed in same unit e.g. gram or gram molecules, then the D has dimensions of $(\text{length})^2(\text{time})^{-1}$ (Crank, 1979).

We assume that at $t = 0$, the substance was concentrated at origin of our coordinate system. The time evolution of the concentration profile can be calculated by relating the Fick's law with the continuity equation i.e.,

$$\frac{\partial C(r, t)}{\partial t} + \nabla \cdot \boldsymbol{J}(\boldsymbol{r}, t) = 0 \tag{3.3}$$

Combining equations (3.2) and (3.3), we get

$$\frac{\partial C(r, t)}{\partial t} + \nabla \cdot [-D\nabla C(\boldsymbol{r}, t)] = 0$$

$$\text{or} \quad \frac{\partial C(r, t)}{\partial t} - D\nabla^2 C(\boldsymbol{r}, t) = 0$$

$$\text{or} \quad \frac{\partial C(r, t)}{\partial t} = D\nabla^2 C(\boldsymbol{r}, t) \tag{3.4}$$

Equation (3.4) can be solved by applying the boundary condition $C(\boldsymbol{r}, 0) = \delta(\boldsymbol{r})$, where

$\delta(\mathbf{r})$ is the Kronecker delta. The solution is

$$C(\mathbf{r}, t) = \frac{1}{2(\pi D t)^{\frac{d}{2}}} \exp\left(\frac{-r^2}{4Dt}\right) ; \tag{3.5}$$

where d is the dimensionality of the system. To determine the diffusion coefficient, the time of second moment is defined as

$$\langle r^2(t) \rangle = \int d\mathbf{r} C(\mathbf{r}, t) r^2 ; \tag{3.6}$$

where, $C(\mathbf{r}, t)$ satisfied the normalization condition $\int d\mathbf{r} C(\mathbf{r}, t) = 1$

We can directly obtain an equation for the time evolution of $\langle r^2(t) \rangle$ by multiplying equation (3.4) by r^2 and integrating over all space. This gives

$$\frac{\partial}{\partial t} \int d\mathbf{r}\, r^2\, C(\mathbf{r}, t) = D \int d\mathbf{r}\, r^2 \nabla^2 C(\mathbf{r}, t)$$

$$\text{or} \quad \frac{\partial}{\partial t} \langle r^2(t) \rangle = D \int d\mathbf{r}\, r^2 \nabla^2 C(\mathbf{r}, t)$$

$$\text{or} \quad \frac{\partial}{\partial t} \langle r^2(t) \rangle = D \int d\mathbf{r} \nabla \cdot (r^2 \nabla C(\mathbf{r}, t)) - D \int d\mathbf{r} \nabla r^2 \cdot \nabla C(\mathbf{r}, t)$$

$$\text{or} \quad \frac{\partial}{\partial t} \langle r^2(t) \rangle = 0 - 2D \int d\mathbf{r} \mathbf{r} \cdot \nabla C(\mathbf{r}, t)$$

$$\text{or} \quad \frac{\partial}{\partial t} \langle r^2(t) \rangle = -2D \int d\mathbf{r} \nabla \cdot \mathbf{r} C(\mathbf{r}, t) + 2D \int d\mathbf{r}\, (\nabla \cdot \mathbf{r}) C(\mathbf{r}, t)$$

$$\text{or} \quad \frac{\partial}{\partial t} \langle r^2(t) \rangle = 6D \int d\mathbf{r} C(\mathbf{r}, t)$$

Since $\nabla \cdot \mathbf{r} = 3$ for three dimensional system. Using the normalization condition: $\int d\tau C(\mathbf{r}, t) = 1$, we get

$$\frac{\partial}{\partial t} \langle r^2(t) \rangle = 6D$$

$$\text{or} \quad D = \frac{\partial}{6 \partial t} \langle r^2(t) \rangle \tag{3.7}$$

The Equation (3.7) relates the macroscopic quantity transport coefficient (D) to the microscopic quantity mean squared displacement (MSD) $\langle r^2(t) \rangle$ of diffusing particles using statistical mechanics known as Einstein's relation to measure self diffusion coefficient (Einstein et al., 1905). From Equation (3.7), it is seen that the instantaneous value of diffusion coefficient is given by the slope of curve of MSD of the diffusing particles versus time. For MSD that behaves as a straight line after a period of prolongated time, the Equation (3.7) reduces to

$$D = \lim_{t \to \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t} \tag{3.8}$$

20

Here, $\langle\ldots\rangle$ denotes the ensemble average of parameters inside the angled bracket. Hence, one sixth of the slope of graph plotted between mean squared displacement versus time estimates the value of self diffusion coefficient.

Also, the binary or mutual diffusion coefficient in binary mixture of two constituents X and Y can be calculated using Darken's phenomenological relation (Darken, 1948). According to the relation, if D_X and D_Y are self diffusion coefficients of two individual constituents X and Y with mole fraction N_X and N_Y respectively, the binary diffusion coefficient D_{XY} of the system is:

$$D_{XY} = N_Y D_X + N_X D_Y \tag{3.9}$$

Out of many other factors, temperature is a factor upon which diffusion depends. Density of system decreases with increase in temperature and hence the region for random walk of diffusing particles become more. As a result, particle's velocity increases with increase in temperature and hence rate of transfer of particle i.e., diffusion coefficient increases (Thapa & Adhikari, 2013).

3.3.1 Effect of System Size on Diffusion Coefficient

Nature of hydrodynamics interactions is long range. Long range interaction depends on size of system. Simulations of system with large size is not easy in practice due to computational capacity (Allen & Tildesley, 1987). Simulations with small size of system introduces a problem of surface effect which is illuminated by performing the simulations under periodic boundary conditions (PBC). The dynamical properties estimated under periodic boundary conditions (PBC) also depends on the size of system on account of hydrodynamics interactions which are long range in nature. So, some modification is needed on diffusion coefficient estimated under periodic boundary conditions (PBC) (Jamali et al., 2018; Dünweg & Kremer, 1993; Yeh & Hummer, 2004). From the self diffusion coefficient estimated under PBC (D_{PBC}), the size independent value of self diffusion coefficient (D_0) (i.e., value estimated with infinite system size) can be estimated by (Yeh & Hummer, 2004):

$$D_0 = D_{\mathrm{PBC}} + \frac{2.84\ k_{\mathrm{B}}T}{6\pi\eta L}$$
$$\text{or} \quad D_{\mathrm{PBC}} = D_0 - \frac{2.84\ k_{\mathrm{B}}T}{6\pi\eta L}\ ; \tag{3.10}$$

where, k_{B}, T and η are Boltzmann constant, absolute temperature of system and viscosity coefficient respectively.

Using Equation (3.10), size independent value of diffusion coefficient can be estimated

if viscosity coefficient of solvent η is known. Also, we can estimate the size independent value of diffusion coefficient as well as viscosity coefficient from the graph plotted between D_{PBC} versus $1/L$ if viscosity coefficient is unknown. For this, at first we should estimate values of $_{PBC}$ taking simulations box of different size. After linear fit, the nature of graph between $_{PBC}$ and $1/L$ is as shown in Figure 8. Values of D_0 and η can
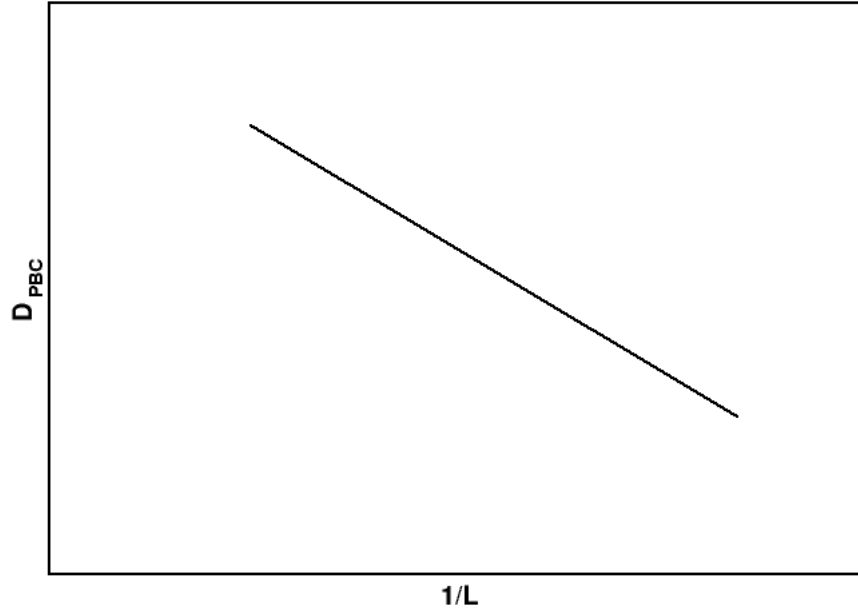
Figure 8: Nature of graph between D_{PBC} and $1/L$.

be evaluated from the intercept and slope respectively of the fitted graph. So,

$$\text{Intercept} = D_0$$
$$\text{Slope} = -\frac{2.84\ k_{\mathrm{B}}T}{6\pi\eta}$$

3.4 Free Energy Calculation

One of the fundamental objective of MD simulations is the estimation of free energy. In this section, we briefly describes the different approaches to estimate free energy of solvation. A thermodynamics quantity equivalent to the capacity of a system to work is known as free energy of the system (i.e., the difference between internal energy of the system and the amount of energy that cannot be used to perform work). Helmholtz free energy $A(N,V,T)$ and Gibb's free energy $G(p,V,T)$ are defined as:

$$A(N,V,T) = U - TS$$
$$G(p,V,T) = A + PV = U - TS + PV = \mu N$$

The difference in free energy during the transformation of a system from initial state A to final B can be calculated using Free Energy Perturbation (FEP) methods (Tuckerman, 2010). Suppose a system with potential energy functions $U_A(r_1, \cdots, r_N)$ and $U_B(r_1, \cdots, r_N)$ for the states A and B respectively. Then, the canonical partition functions of the states A and B are defined as

$$Q_A(N, V, T) = \frac{Z_A(N, V, T)}{N!\,\Lambda^N}$$
$$\text{or}\quad Q_A(N, V, T) = \frac{1}{N!\,\Lambda^N} \times \int e^{-\beta U_A(r_1, \cdots, r_N)}\, dr_1, \cdots, dr_N \tag{3.11}$$

and

$$Q_B(N, V, T) = \frac{Z_B(N, V, T)}{N!\,\Lambda^N}$$
$$\text{or}\quad Q_B(N, V, T) = \frac{1}{N!\,\Lambda^N} \times \int e^{-\beta U_B(r_1, \cdots, r_N)}\, dr_1, \cdots, dr_N \tag{3.12}$$

where, Λ is the de-Broglie wavelength; and $Z_A(N, V, T)$ and $Z_B(N, V, T)$ are the configurational partition functions for the states A and B respectively. If A_A and A_B are the Helmholtz free energies of the states A and B respectively then the difference in free energy between the two states is given by

$$\Delta A_{AB} = A_B - A_A$$
$$\text{or}\quad \Delta A_{AB} = -k_B T \ln Q_B(N, V, T) - (-k_B T \ln Q_A(N, V, T))$$
$$\text{or}\quad \Delta A_{AB} = -k_B T \ln\left[\frac{Q_B}{Q_A}\right]$$
$$\text{or}\quad \Delta A_{AB} = -k_B T \ln\left[\frac{Z_B}{Z_A}\right]$$
$$\text{or}\quad \Delta A_{AB} = -k_B T \ln\left[\frac{\int e^{-\beta U_B}\, d^N r}{\int e^{-\beta U_A}\, d^N r}\right]$$
$$\text{or}\quad \Delta A_{AB} - k_B T \ln\left[\frac{\int e^{-\beta(U_B - U_A)}\, d^N r}{\int d^N r}\right]$$
$$\text{or}\quad \Delta A_{AB} = -k_B T \ln\langle \exp[-\beta(U_B - U_A)]\rangle_A \tag{3.13}$$

The $\langle \cdots \rangle_A$ gives the ensemble average of difference in potential energy function with respect to the state A. Equation (3.13) is the Free Energy Perturbation (FEP) formula (Tuckerman, 2010) to calculate the free energy difference between the two states A and B. If the difference in the potential energy function $(U_B - U_A)$ is high i.e., if there is insufficient overlap of phase space between two states A and B, then the value of the term $\exp[-\beta(U_B - U_A)$ becomes very small; and hence the estimated value may not be converges. To overcome the problem of convergence, one can introduce $M - 2$

unphysical intermediate states with potential energy function $U_\alpha(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N)$, where $\alpha = 1 \cdots M$. After then the net change in free energy of the system in between the two states is

$$\Delta A_{AB} = -k_{\mathrm{B}}T \sum_{\alpha=1}^{M-1} \ln\langle \exp-(\beta\Delta U_{\alpha,\alpha+1})\rangle_\alpha \tag{3.14}$$

Alternatively, the transformation can be handled by introducing a new potential energy function in terms of switching parameter λ as

$$U(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N, \lambda) = f(\lambda)U_A + g(\lambda)\, U_B \tag{3.15}$$

Here, λ is coupling parameter such that $0 \leq \lambda \leq 1$. The value of $\lambda = 0$ represents initial state A, 1 represents final state B and other values lies between 0 and 1 represents non-physical intermediate states. There is no change is estimated values of free energy difference between two states with many non-physical intermediate states as free energy is a state function. Figure shows schematic diagram of representation of many non-physical intermediate states in between initial state A and final state B.
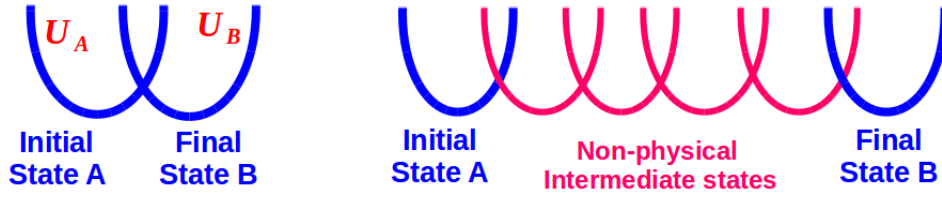
Figure 9: Schematic diagram to illustrate two thermodynamic states.

Then, we can define the partition function and hence the free energy for the potential defined by the equation (3.15) as

$$A(N, V, T, \lambda) = -k_{\mathrm{B}}T\, \ln Q_B(N, V, T, \lambda)$$

Differentiating on both sides with respect to λ, we get

$$\frac{\partial A}{\partial \lambda} = -\frac{k_{\mathrm{B}}T}{Q}\frac{\partial Q}{\partial \lambda} = -\frac{k_{\mathrm{B}}T}{Z}\frac{\partial Z}{\partial \lambda} = -\frac{k_{\mathrm{B}}T}{Z}\frac{\partial}{\partial \lambda}\int e^{-\beta U(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N, \lambda)}\, d^N\boldsymbol{r}$$

$$\text{or} \quad \frac{\partial A}{\partial \lambda} = \langle\frac{\partial U}{\partial \lambda}\rangle_\lambda$$

Then, the difference in free energy during the transformation of the system from state A to state B is evaluated using the equation

$$\Delta A_{AB} = \int_0^1 \frac{\partial A}{\partial \lambda}d\lambda = \int_0^1 \langle\frac{\partial U}{\partial \lambda}\rangle_\lambda\, d\lambda \tag{3.16}$$

Equation (3.16) is known as Thermodynamic Integration (TI) formula. If we consider the functions as $f(\lambda) = 1 - \lambda$ and $g(\lambda) = \lambda$, equation (3.16) becomes

$$\Delta A_{AB} = \int_0^1 \langle U_B - U_A \rangle_\lambda \ d\lambda \qquad (3.17)$$

Now, the free energy of solvation i.e., the free energy difference between the states with and without interaction of solute and solvent molecules is evaluated as

$$\Delta G_{\text{sol}} = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad (3.18)$$

$$\text{or} \quad \Delta G_{\text{sol}} = \int_0^1 \langle U_B - U_A \rangle_\lambda d\lambda \qquad (3.19)$$

Besides FEP and TI methods, another approach used to calculate the free energy difference is Bennett Acceptance Ratio (BAR) method (Bennett, 1976). The assumption behind this is that same microstates are shared by both initial and final states. Here, we briefly describe the theory of BAR method. The canonical configurational partition function is defined as

$$Q(N, V, T) = \int \exp\left[-\beta \, U(\boldsymbol{q}_1, \cdots, \boldsymbol{q}_N)\right] d\boldsymbol{q}_1, \cdots, d\boldsymbol{q}_N \qquad (3.20)$$

Now, the ratio of partition functions Q_0 for initial state A and Q_1 for final state B in terms of weighting function $W(q_1, \ q_2, \ ......q_N)$ is given by

$$\frac{Q_1}{Q_0} = \frac{Q_1 \int W \ \exp\left[-\beta \ (U_1 + U_0)\right] \, dq^N}{Q_0 \int W \ \exp\left[-\beta \ (U_0 \ + \ U_1)\right] \, dq^N}$$

$$\text{or} \quad \frac{Q_1}{Q_0} = \frac{\langle W \ \exp\left[-\beta \ U_1\right] \rangle_0}{\langle W \ \exp\left[-\beta \ U_0\right] \rangle_1} \qquad (3.21)$$

Now, the free energy difference is given by

$$\beta \, \Delta F = \ln\left(\frac{Q_1}{Q_0}\right) = \ln\left(\frac{\langle W \ \exp\left[-\beta \ U_1\right] \rangle_0}{\langle W \ \exp\left[-\beta \ U_0\right] \rangle_1}\right)$$

$$\text{or} \quad \beta \, \Delta F = \ln \langle W \ \exp\left[-\beta \ U_1\right] \rangle_0 - \ln \langle W \ \exp\left[-\beta \ U_0\right] \rangle_1 \qquad (3.22)$$

With suitable chosen of the weight function W, we can estimate the free energy with minimum error and the best choice of weight function is (Bennett, 1976)

$$W = \text{constant} \times \left[\left(\frac{Q_0}{n_0}\right) \exp(-\beta U_1) + \left(\frac{Q_1}{n_1}\right) \exp(-\beta U_0)\right]^{-1} ; \qquad (3.23)$$

where, n_0 and n_1 are the number of independent configurations related with two states.

By combining Equations (3.21) and (3.23), we get

$$\frac{Q_1}{Q_0} = \frac{\left\langle \{1 + \exp\left[\beta\left(U_1 - U_0 + C\right)\right]\}^{-1} \right\rangle_0}{\left\langle \{1 + \exp\left[\beta\left(U_0 - U_1 - C\right)\right]\}^{-1} \right\rangle_1} \exp(\beta C)$$

Here, we defined C as: $\exp(\beta C) = (Q_1 n_0 / Q_0 n_1)$. With Fermi function defined as $f(x) = (1 + \exp(x))^{-1}$, the above equation can be expressed as

$$\frac{Q_1}{Q_0} = \frac{\langle f(U_1 - U_0 + C) \rangle}{\langle f(U_0 - U_1 - C) \rangle} \exp(\beta C) \tag{3.24}$$

3.5 Molecular Dynamics Simulations

In this work, we have performed molecular dynamics (MD) simulations to study the transport and thermodynamic properties of amoxicillin using Groningen Machine for Chemical Simulations (GROMACS) software package (David et al., 2005). Molecular dynamics is a deterministic approach. It is a statistical method in which ensemble average over configurations represents physical quantities. During simulations, macroscopic properties interest can be obtained from microscopic properties through the analysis of trajectory followed in phase space (Gunsteren & Berendsen, 1990; Allen & Tildesley, 1987). With given set of initial position and velocity, the computer calculates their trajectory in a 6N-dimensional phase space, which provides a set of configurations required in statistical mechanics to represent physical quantities (Frenkel & Smit, 2002). The measurement of physical quantities by simulation is obtained as an arithmetic average of the instantaneous values assumed by that quantity during MD run (Rapaport, 2004; Ercolessi, 1997). Such simulations technique plays a crucial role to get more insight about different interactions in atomic/molecular levels. During simulations, experimental environment can be mimicked through the interactions between atoms (Khanal et al., 2019). In classical MD simulations, new configurations can be generated after solving Langevin equation (Lemons & Gythiel, 1997)

$$m_i \frac{\partial^2 r_i(t)}{\partial t^2} = F_i\{r_i(t)\} - \gamma_i \frac{dr_i(t)}{dt} m_i + R_i(t)$$

In the Langevin equation, 'F_i' is the force on i^{th} particle of mass 'm_i' due to interaction of other particles, the second represents the damping term and third terms is random force term.

The classical molecular dynamics simulation is used for the generation of non-equilibrium ensembles and for the analysis of dynamics events like viscosity, diffusion processes, dynamics of defects in crystal etc. The Molecular Dynamics simulations technique can be applied in large area of research like structural and dynamical properties including

biomolecules like protein, amino-acids; to study about phase transition behaviors; to estimate transport coefficients; to calculate free energy of solvation, protein-ligand binding energy and so on (Rapaport, 2004).

In molecular dynamics simulation, first we prepare a sample: we select a model system consisting of N-particles and we solve Newton's equations of motion to equilibrate the system. After then, we perform the actual measurement. The overall process of molecular dynamics simulation can be divided into following four steps (Frenkel & Smit, 2002):

- Modeling the physical system

- Initialization

- Force Calculation

- Integrating Equation of motion

3.5.1 Modeling of the System

The first step of molecular dynamics simulations is the preparation of a model of the system under study. The results obtained from the simulations also depends on modeling of system. To construct model for simulations, we specify the molecules under study in terms of atomic masses, charges, van der Waals radii, well depth of energies, spring constant, force field etc.. In order to describe the different interactions between atoms/molecules, we define functional form of empirical potential to constitute force fields. The forces are evaluated from the negative gradient of the potentials; and the accuracy of simulations also depends upon the selection of force field parameters. The most popular force fields for bimolecular simulations are GROMOS96, OPLS-AA, CHARM and AMBER with reasonable differences in specification and resulting parameters (Oostenbrink et al., 2004; Jorgensen et al., 1996; Brooks et al., 2009; Wang et al., 2004).

In classical force fields, the potential functions are derived from empirically to describe atomic interaction in which atoms are treated as spherically symmetric particles connected through covalent bonds forming molecules. Each atom experiences a force resulting from its pairwise interaction with the rest atoms of the system. The total potential V_{total} which is the sum of bonded and non-bonded interactions given by (Jorgensen et al., 1983; MacKerell Jr et al., 1998; Thapa & Adhikari, 2013).

$$V_{\text{total}} = V_{\text{bonded}} + V_{\text{non-bonded}} \tag{3.25}$$

The bonded interactions includes bond stretching potential (V_{bond}), bond angle poten-

tial (V_{angle}), proper dihedral potential (V_{dihed}) and improper dihedral potential (V_{impr}). Similarly, the non-bonded terms is sum of van der Waals potential (V_{vdW}) and Coulomb potential (V_{Coulomb}). So,

$$V_{\text{bonded}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihed}} + V_{\text{impr}}$$

and $\quad V_{\text{non-bonded}} = V_{\text{vdW}} + V_{\text{Coulomb}}$

Out of the bonded interactions, bond stretching potential (V_{bond}), bond angle potential (V_{angle}) and improper dihedral potential (V_{impr}) are harmonic potentials but the proper dihedral potential (V_{dihed}) (Lindahl et al., 2010).

3.5.1.1 Bonded Interactions

The bonded interactions include the contributions of bond stretching, bond angle and proper as well as improper dihedral potentials (Lindahl et al., 2010). The different bonded potentials are presented in below.

3.5.1.2 Bond Stretching Potential

The bond stretching potential between two bonded atoms x and y is harmonic and defined as

$$V_{\text{bond}}(r_{xy}) = \frac{1}{2}k_{xy}^{\text{bond}}(r_{xy} - b_{xy})^2 \tag{3.26}$$

where k_{xy}^{bond}, r_{xy} and b_{xy} are force constant for bond stretching potential, distance between the two atoms and equilibrium bond length respectively. Figure shows the bond between x and y atoms.
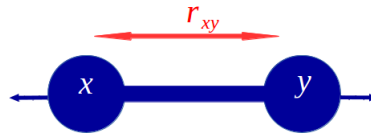
Figure 10: Schematic diagram to represent bond stretching between two atoms separated by a distance r_{xy}.

3.5.1.3 Bond Angle Potential

The bond angle vibration of three atoms $x - y - z$ is also a harmonic potential and defined as

$$V_{\text{angle}}(\theta_{xyz}) = \frac{1}{2}k_{xyz}^{\theta}(\theta_{xyz} - \theta_{xyz}^0)^2 \tag{3.27}$$

where k_{xyz}^{θ}, θ_{xyz} and θ_{xyz}^{0} are the force constant for bond angle potential, bond angle and equilibrium bond angle respectively. Figure shows the bond angle vibration between x, y and z atoms.
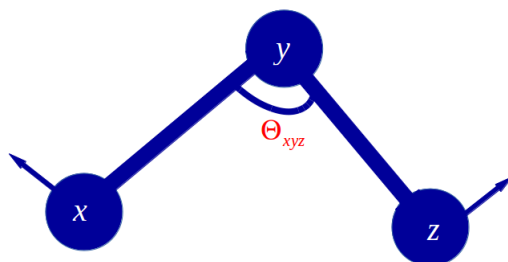
Figure 11: Schematic diagram to represent bond angle vibration between three atoms $x - y - z$.

3.5.1.4 Improper Dihedrals

The improper dihedral angle ξ is defined as the angle between planes (w, x, y) and (x, y, z) as shown in the schematic diagram below.
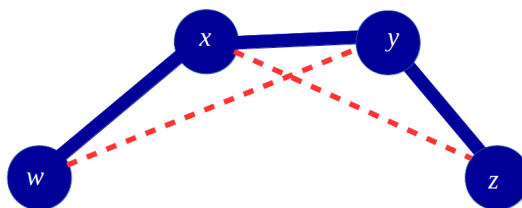
Figure 12: Schematic diagram of improper dihedral.

The simplest improper dihedral potential is harmonic potential and defined as

$$V_{\text{impr}}(\xi_{wxyz}) = \frac{1}{2}k_{wxyz}^{\xi}(\xi_{wxyz} - \xi_0)^2 \tag{3.28}$$

In this equation, k_{wxyz}^{ξ}, ξ_{wxyz} and ξ_0 represent the force constant for improper potential, improper dihedral and equilibrium improper dihedral angles respectively.

3.5.1.5 Proper Dihedrals

According to the IUPAC/IUB convention, the proper dihedral angle is the angle between the planes wxy and xyz as shown in schematic diagram below.
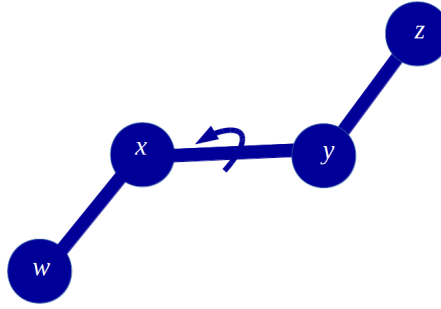
Figure 13: Schematic diagram for proper dihedral.

The potential due to proper dihedral is periodic and defined as

$$V_{\text{dihed}}(\phi_{wxyz}) = k_\phi(1 + \cos(n\phi - \phi_s)) \tag{3.29}$$

In above equation, ϕ represents the dihedral angle.

3.5.1.6 Non-bonded Interactions

The non-bonded interactions are in general sum of two contributions: Electrostatic or Coulomb and van der Waals interactions.

3.5.1.7 Coulomb Potential

The electrostatic interaction between two point charges q_i and q_j as shown in figure is defined in terms of Coulomb potential as (Phillips et al., 2005)
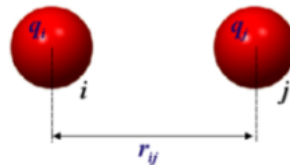
Figure 14: Coulomb interaction.

$$V_{\text{Coulomb}} = \frac{q_i q_j}{4\pi \epsilon_r \epsilon_0 r_{ij}} \tag{3.30}$$

In this Equation (3.30), ϵ_0, ϵ_r and r_{ij} represent the permittivity of free space, dielectric constant or relative permittivity of the medium and distance between the charges respectively. If atoms/molecules are not ionized, they are not charged. But due to unequal distributions of charges in the atoms/molecules, partial charges arises; the partial charges contribute to the electrostatic interaction (Thapa & Adhikari, 2013). The figure shows the variation of Coulomb potential (Poudyal & Adhikari, 2014).
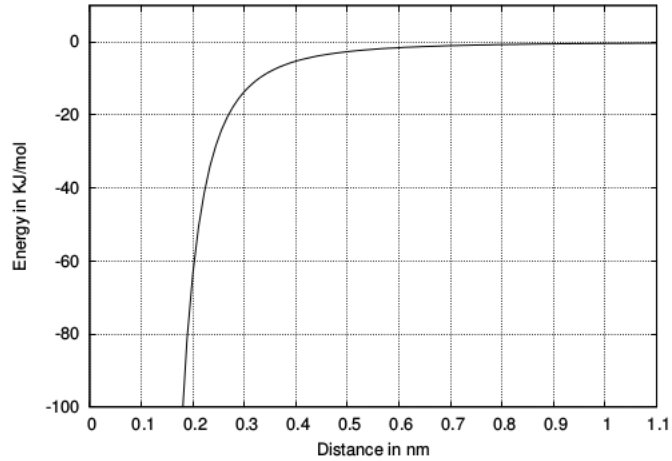
Figure 15: Variation of Coulomb Potential.

3.5.1.8 Lennard-Jones (L-J) Potential

Another non-bonded, pairwise potential suitable for many areas of condensed matter physics is Lennard-Jones (L-J) potential. Such potential includes the strongly repulsive term due to overlapping of electronic cloud at short distance and attractive term caused by dipole-dipole interaction for large distance. The L-J potential between two atoms separated at a distance r is given by (Leach, 2001; Ercolessi, 1997)

$$\phi_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \tag{3.31}$$

In above Equation 3.31, the terms ϵ and σ stand for strength of the interaction and distance at which potential becomes minimum respectively. The variation of L-J potential is shown in figure given below (Thapa & Adhikari, 2013).

31
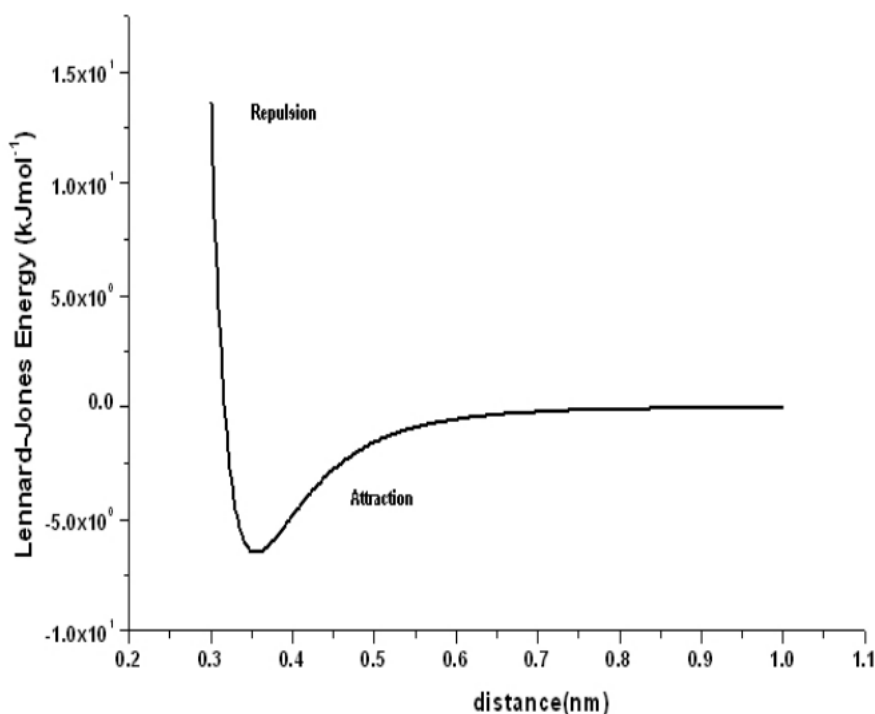
Figure 16: Variation of Lennard - Jones potential with distance between atoms.

3.5.1.9 Potential Box and Periodic Boundary Conditions(PBC)

Construction of simulation box is another important part of modeling of a system. In general simulation box is taken to be cubic which contains all molecules of the system under study. In general, simulations has been carried out taking having small number of molecules i.e., $10 \leq N \leq 10000$. With small system, we encounter a problem that the fraction of molecules lie on the surface to within volume of the simulation box is large known as surface effect (Reed & Flurchick, 1996; Allen & Tildesley, 1987). Surface effect is arised due to the reason that the molecules on the surface of simulation box experience different forces from other molecules lie within volume. Such problem plays important role for relatively small system. The problem of surface effect can be removed by introducing the periodic boundary condition (PBC) (Allen & Tildesley, 1987). When we use the concept of PBC, the box enclosing the molecules is replicated to infinity by rigid translation in all directions as shown in Figure 17. During the simulation if a particle moves in original box, its periodic images in each of neighbouring boxes move in exactly the same way that the number density in central box is conserved.

Although, the PBC virtually solves the surface effect problem, finite size effect still present (Reed & Flurchick, 1996; De Souza & Ornstein, 1997). Also, the number of interacting pairs goes enormously large with PBC. As a result, we must handle infinite number of turns to estimate pairwise potential due to short range non-bonded interaction. The addition of new complexity can be reduced by introducing the concept of the minimum image criterion. According to which, among all possible images of a

32
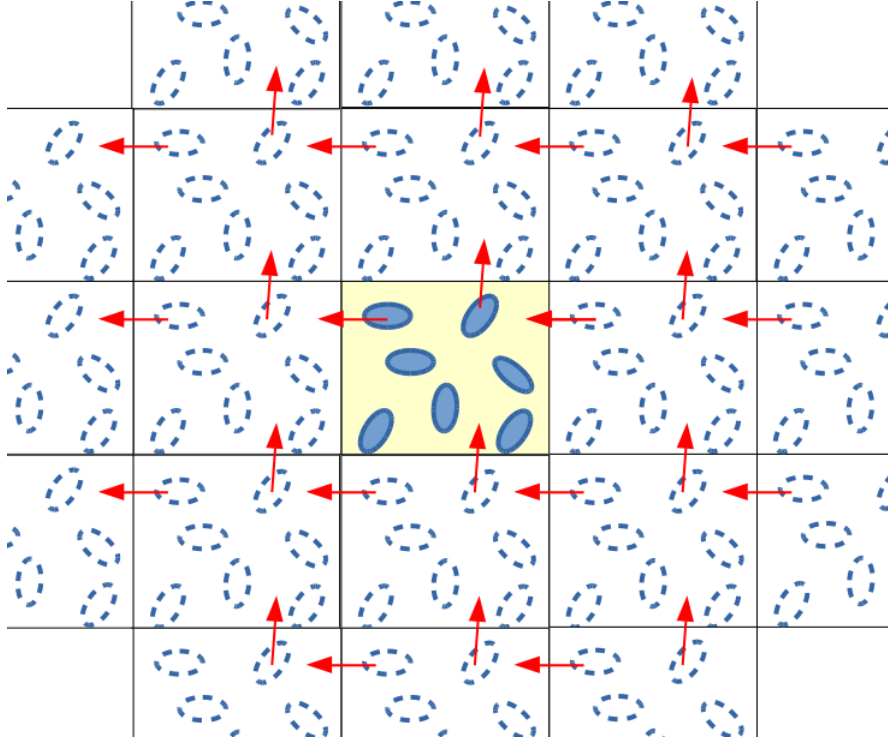
Figure 17: Schematic diagram of periodic boundary condition.

particle, only the closest are selected for interaction (Lindahl et al., 2010). Figure 18 illustrates the schematic diagram of minimum image convention. Here, we present the potential due to interaction of particle A with others obeying minimum image criterion.

3.5.2 Initialization

In order to start simulation, we should assign initial positions and velocities of all molecules in the system. The molecule positions should be chosen compatible with the structure that is being aimed to simulate. In any event, the molecules should not be positioned at positions that result in an appreciable overlap of the atomic or molecular cores. This is achieved by initially placing the molecules on a cubic lattice. The density and initial temperature are chosen such that the simple cubic lattice is mechanically unstable and melts rapidly. At first each molecule is put on its lattice site and for every molecule the value of each velocity component is attributed from a uniform distribution which is not Maxwellian. Then we shift all velocities in such a way that the total momentum is zero. In thermal equilibrium (Frenkel & Smit, 2002)

$$\left\langle \frac{1}{2}mv_\alpha^2 \right\rangle = \frac{k_\mathrm{B}T}{2}$$

or $\quad \langle v_\alpha^2 \rangle = \dfrac{k_\mathrm{B}T}{m}$ \hfill (3.32)
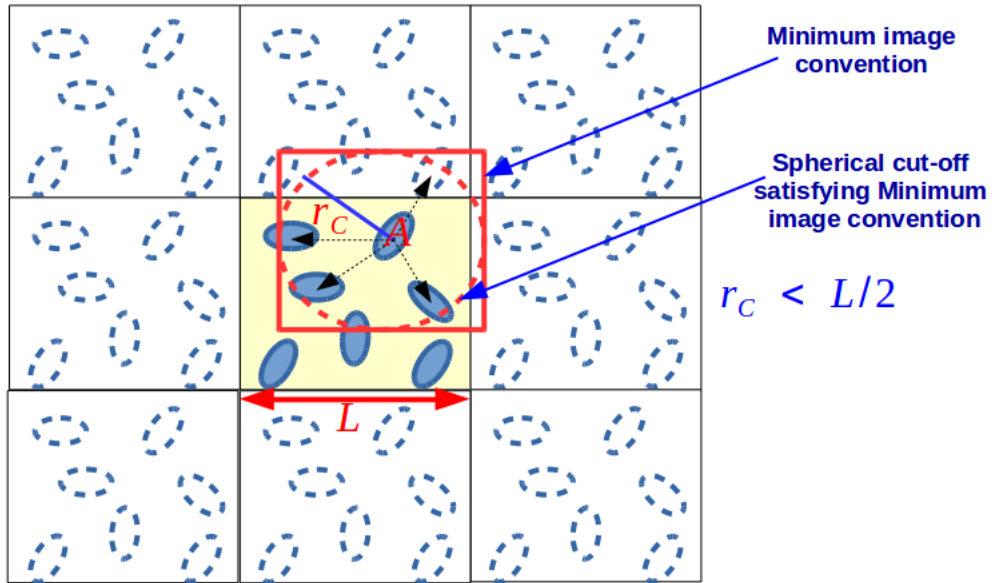
Figure 18: Schematic diagram for use of cutoff and minimum image criterion.

In the equation (3.32), m is the mass of molecule, v_α is the α component of velocity, k_B is the Boltzmann constant and T is the temperature.

3.5.3 Force Calculation

The major challenging and most time consuming part of MD simulations is determination of force acting on each particles (Frenkel & Smit, 2002). In MD simulations, we model our system of interest through pairwise additive interactions: so we take the contribution to the force on i^{th} particle due to all its neighbours. If we account all such interactions, $N(N-1)/2$ pair interactions must be evaluated to get force on i^{th} particle for N particles system. In this case, the required time scale of force calculations is order of N^2. In order to reduce the simulations time, some tricks are implemented during force calculation. For this, we use the concept of cut off distance i. e., we cut off the interaction between two particles if the distance between them exceeds some values known as cut off distance r_c (Rapaport, 2004). The value of cut off for both coulomb and van der Waals interactions must be less than half the size of simulation box. Such technique reduces the time scale to order of N. The use of spherical cut off with minimum image convention is shown in Figure 18.

3.5.3.1 Time Integration Algorithm

In MD study, the desired properties of simulated system have been drawn from trajectory analysis followed by the particles in phase space during simulation. Thus, we first generate the trajectory followed by system after solving the equation motion. The time integration algorithm based on finite difference methods is used to integrate equation of

motion (Allen & Tildesley, 1987; Frenkel & Smit, 2002). By knowing the positions and their time derivatives at any instant of time (t), we can calculate the same quantities at latter time $t + \Delta t$. The finite difference methods are usually based on Taylor expansion truncated at some terms. Due to truncation, some errors are introduced known as truncation errors which are intrinsic to algorithm. Another type of error is introduced due to finite number of digits used in computer arithmetic called round off error. Here we discuss the frequently used algorithms in computer simulations which are:

Verlet Algorithm

It is derived from the Taylor expansion of the coordinate of a particle at $(t + \Delta t)$ and $(t - \Delta t)$ as given below.

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{F}(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\frac{d^3 r}{dt^3} + \theta(\Delta t^4) \tag{3.33}$$

$$\boldsymbol{r}(t - \Delta t) = \boldsymbol{r}(t) - \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{F}(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\frac{d^3 r}{dt^3} + \theta(\Delta t^4) \tag{3.34}$$

Now by adding equations (3.33) and (3.34), we get

$$\boldsymbol{r}(t + \Delta t) + \boldsymbol{r}(t - \Delta t) = 2\boldsymbol{r}(t) + \frac{\boldsymbol{F}(t)}{m}\Delta t^2 + \theta(\Delta t^4)$$

$$\text{or} \quad \boldsymbol{r}(t + \Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t) + \frac{\boldsymbol{F}(t)}{m}\Delta t^2 + \theta(\Delta t^4) \tag{3.35}$$

From Equation (3.35) it is seen that an error of order Δt^4 with time Δt s introduced when we measure position in Molecular Dynamics. From Equation (3.35), it is also seen that Verlet algorithm does not use the velocity to calculate the new position. However, we can calculate the velocity from the knowledge of trajectory. For this, subtract Equation (3.34) from Equation (3.33) as

$$\boldsymbol{r}(t + \Delta t) - \boldsymbol{r}(t - \Delta t) = 2\boldsymbol{v}(t)\Delta t + \theta(\Delta t^3)$$

$$\text{or} \quad \boldsymbol{v}(t) = \frac{\boldsymbol{r}(t + \Delta t) - \boldsymbol{r}(t - \Delta t)}{2\Delta t} - \theta(\Delta t^2) \tag{3.36}$$

The Equation (3.36) is only accurate to Δt^2. However, it is possible to calculate velocity more accurately and hence kinetic energy using other Verlet-like algorithm. Also from Equation (3.36) it is seen that Verlet algorithm is time reversible.

3.5.3.2 Leap-Frog Algorithm

Out of several Verlet equivalent algorithms, one is Leap-frog algorithm (Gunsteren & Berendsen, 1988). 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.

$$v\left(t - \frac{\Delta t}{2}\right) = \frac{r(t) - r(t - \Delta t)}{\Delta t} \tag{3.37}$$

$$\text{and} \quad v\left(t + \frac{\Delta t}{2}\right) = \frac{r(t + \Delta t) - r(t)}{\Delta t} \tag{3.38}$$

From Equations (3.37) and (3.38) , we can obtain the expression for new position from old position and velocity as,

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{3.39}$$

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

$$v\left(t + \frac{\Delta t}{2}\right) = v(t) + \frac{F(t)}{2m}\Delta t \tag{3.40}$$

$$\text{and} \quad v\left(t - \frac{\Delta t}{2}\right) = v(t) - \frac{F(t)}{2m}\Delta t \tag{3.41}$$

Subtracting equation (3.41) from (3.40), we get

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m}\Delta t \tag{3.42}$$

Equation (3.42) represents the expression for new velocity based on old velocity. 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.
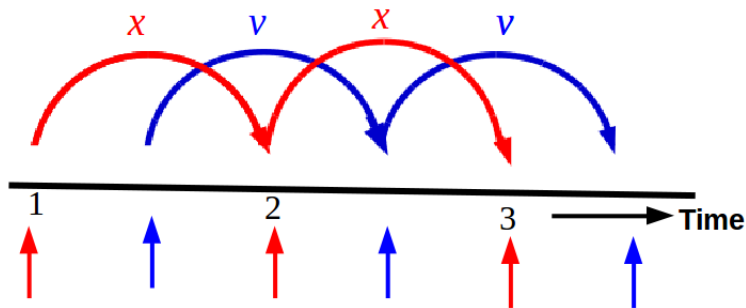
Figure 19: Schematic diagram of the Leap-Frog integration.

3.5.3.3 Constraint Dynamics

In molecular system, the particle also interacts through intra-molecular interactions which are accounted due to change in bond, bond angle etc. The simulation time step depends upon high frequencies vibrations (Leach, 2001). Due to the high frequency bond vibrations, the time step becomes extremely small which introduces complexity for long simulations. Such problem can be handled using the concept of constraint dynamics which enables to constraint some internal degree of freedom of the molecular systems (Allen & Tildesley, 1987). Some popular constraint algorithms are SHAKE, LINCS etc. (Ryckaert et al., 1977; Hess et al., 1997). SHAKE algorithm is an iterative method in which constraint coordinate is set from unconstrained (Lindahl et al., 2010). On the other hand, LINCS algorithm is not iterative method which resets the correct values of bonds in two steps.

3.5.4 Statistical Ensembles in Molecular Dynamics

Molecular dynamical simulations is a statistical approach. The information about different properties are generated at microscopic level from trajectory followed by the system in phase space and relate with macroscopic properties using the concept of statistical mechanics (Allen & Tildesley, 1987; Frenkel & Smit, 2002). For example, the macroscopic property diffusion coefficient is estimated from microscopic properties mean square displacement (MSD) of particles of the system. So, here we briefly discuss the concept of statistical mechanics. Suppose a system defined by thermodynamics parameters: number of particles (N), volume of system (V) and temperature of the system (T). The micro states of the system at any instant is defined by $3N$ position and $3N$ momenta coordinates. So, $6N$-dimensional space is required to specify the micro states of the system known as phase space and the point in the space with coordinates (q_i, p_i) is representative point of the system. The time evolution of the system is governed by equation of motion. If A be the instantaneous value of some observable property, then the macroscopic property that observed experimentally A_{obs} is given by the time average as (Tuckerman, 2010)

$$A_{\text{obs}} = \langle A \rangle_{\text{time}} = \lim_{t_{\text{obs}} \to \infty} \frac{1}{t_{\text{obs}}} \int_0^{t_{\text{obs}}} A(t) \ dt \qquad (3.43)$$

We can only taken the average for long finite time but the integration cannot extend to infinite time; exactly same thing happens in MD simulations. During computer simulations, large number of simulations steps (n_{step}) defined as $n_{\text{step}} = t_{\text{obs}}/\delta t$, where

δt is the time step taken. So, the time average may be defined as

$$A_{\text{obs}} = \langle A \rangle_{\text{time}} = \frac{1}{t_{\text{obs}}} \sum_{1}^{t_{\text{obs}}} A(t) \tag{3.44}$$

In order to get good results from the simulations, the system must covered the sufficient area in phase space. Also, the time average for systems having large number of molecules is not suitable in statistical mechanics. Instead of time average, the concept of ensemble average is used (McQuarrie, 2000). Ensemble is a collection of large number of identical systems having same macro states. Some ensembles which are used in MD simulations are:

1. Microcanonical (NVE) ensemble: Collection of infinite number of isolated systems having equal number of particles (N) in equal volume (V) with equal energy (E) of each system. The micro-canonical partition function is defined as

$$Q_{NVE} = \sum \delta(H(p,q) - E) \tag{3.45}$$

2. Canonical (NVT) ensemble: Collection of infinite number of closed systems having equal number of particles (N) in equal volume (V) at equal temperature (T). The partition function for canonical ensemble is defined as

$$Q_{NVT} = \frac{1}{h^{3N}} \int e^{-\beta H(p,q)} \, dp \, dq \tag{3.46}$$

3. Grand-canonical (μVT) ensemble: Collection of infinite number of open systems having fixed value of chemical potential (μ) in equal volume (V) at equal temperature (T). The partition function for grand-canonical ensemble is defined as

$$Z_{\mu VT} = \sum_{N} z^{N\mu/k_{\text{B}}T} Q_{NVT} \tag{3.47}$$

4. Isothermal-isobaric (NPT) ensemble: Collection of infinite number of closed systems having fixed number of particles (N) with equal pressure (P) and temperature (T). The partition function for isothermal-isobaric ensemble is defined as

$$Z_{NPT} = \sum_{V} e^{(-PV/k_{\text{B}}T)} Q_{NVT} \tag{3.48}$$

The ensemble average of observable properties is related with time average by ergodic hypothesis. According this hypothesis, the time average of any observable say $A(p,q)$ is equal to its ensemble average i. e.,

$$\langle A(p,q) \rangle_{\text{time}} = \langle A(p,q) \rangle_{\text{ensemble}} \tag{3.49}$$

3.5.5 Temperature and Pressure Control

Although MD has been already considered as a tool to study many phenomenon like dynamical properties of liquid, such technique cannot be used to study the transport properties taking NVE ensemble (Berendsen et al., 1984). Also, to get experimental environments as well as study the effect of temperature on many properties of interest, we must performed simulations keeping constant temperature and pressure. The algorithms used to control temperature and pressure are known as thermostat and barostat respectively.

During MD simulations, the instantaneous temperature of a system is evaluated through average kinetic energy as (Leach, 2001)

$$\langle K \rangle = \frac{3}{2} N k_{\text{B}} T$$

where, K and k_{B} are the instantaneous kinetic energy of the system and Boltzmann constant respectively. Then, the instantaneous value temperature $T(t)$ at any instant of time t is given as

$$T(t) = \frac{2K}{3N k_{\text{B}}} \tag{3.50}$$

From the Equation (3.50), it is seen that the temperature can be changed by scaling velocities. If velocities are multiplied by a factor λ, then change in temperature is

$$\Delta T = \frac{1}{2} \sum_i \frac{2 m_i (\lambda v_i)^2}{3N k_{\text{B}}} - \frac{1}{2} \sum_i \frac{2 m_i v_i^2}{3N k_{\text{B}}}$$

$$\text{or} \quad \Delta T = \frac{1}{2} \sum_i \frac{2 m_i v_i^2}{3N k_{\text{B}}} (\lambda^2 - 1)$$

From equation (3.50), we get

$$\Delta T = T(t)(\lambda^2 - 1)$$

$$\text{or} \quad \lambda = \sqrt{\frac{T_{\text{new}}}{T(t)}} \tag{3.51}$$

Thus, we can control the temperature of a system under simulations by multiplying the velocities at each step of time by a factor $\sqrt{T_{\text{req}}/T_{\text{curr}}}$; where T_{curr} is the instantaneous temperature of the system evaluated from kinetic energy and T_{req} target temperature. Such method to control temperature by scaling velocities is velocity rescaling thermostat (Bussi et al., 2007). Furthermore, we can control the temperature of the system through coupling it to external bath at fixed temperature which behaves as the source of thermal energy to the system. Such method is known as Berendsen temperature coupling (Berendsen et al., 1984). The bath either can supply heat to the system or remove from the system such that in each step , the rate of change of temperature is proportional the difference in temperature between the system and bath i. e.,

$$\frac{dT(t)}{dt} = \frac{1}{\tau} \left(T_{\text{bath}} - T(t) \right) \tag{3.52}$$

Such equation indicates that the deviation of temperature decays in exponential ways towards desired temperature. In each successive time step, the temperature changes as

$$\Delta T = \frac{\delta t}{\tau} \left(T_{\text{bath}} - T(t) \right) \tag{3.53}$$

And the scaling factor for velocity is

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{\text{bath}}}{T(t)} - 1 \right) \tag{3.54}$$

Similarly, we can control the pressure of a system by scaling volume of the system. Alternatively, pressure can be controlled by using a concept of pressure bath like temperature bath in case of thermostat. Such concept was introduced by Berendsen (Berendsen et al., 1984). Berendsen barostat also gives first order kinetic decay of pressure to obtain reference temperature. According to the Berendsen barostat, the rate of change of pressure is proportional the difference in pressure between the system and pressure bath i. e.,

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} \left(P_{\text{bath}} - P(t) \right) \tag{3.55}$$

3.5.6 Software Packages

There many software available to perform MD simulations including GROMACS, NAMD, LAMMPS, CHARM with distinct features. The present work has been carried out using GROningen MAchine for Chemical Simulation (GROMACS) package (Berendsen et al., 1995; David et al., 2005). The main features of the package are user-friendly, high compatibility, versatile, open source etc.. We also used other soft-

wares including Visual molecular dynamics (VMD) (Humphrey et al., 1996), Xmgrace (Turner, 2005), NAMD (Phillips et al., 2005) and so on.

3.5.7 Systems Setup

In this work, we have performed MD simulations to study the transport and thermo-dynamics properties of amoxicillin. We have studied the diffusion phenomenon and calculate the diffusion coefficients of amoxicillin in water. We have also analysed the effect of system size and temperature on the estimated values of diffusion coefficient. Furthermore, we have estimate the free energy of solvation of amoxicillin in two solvent environments: water and ethanol. To study such properties, we considered following systems.

3.5.7.1 Systems to Study Structural and Transport Properties:

In order to study the structural and transport properties, a system of binary mixture of 2 amoxicillin and 2160 water molecules (System-I) was set up under periodic boundary conditions in cubic simulations box of size 4.05 nm. Figure 20 represents molecular system of amoxicillin in water.
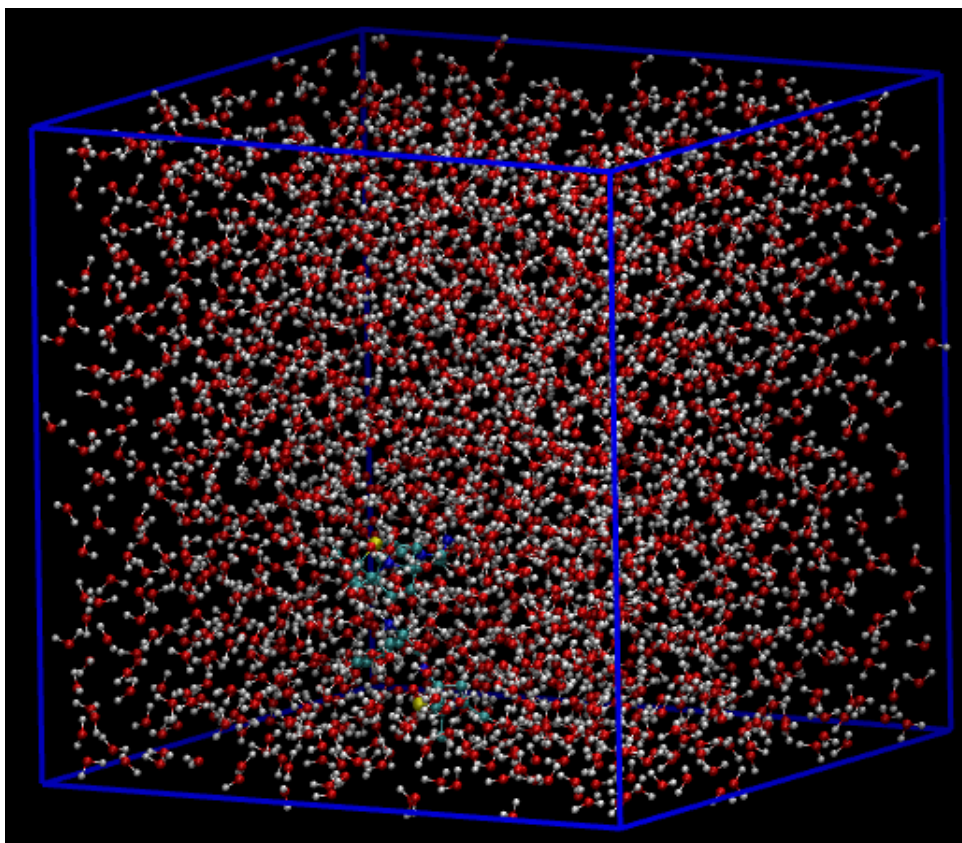
Figure 20: VMD snapshot of the system under study (amoxicillin in water).

Optimized Potentials for Liquid Simulations-all atom (OPLS-AA) (Jorgensen et al., 1996) was used for modeling amoxicillin molecule; and extended simple point charge (SPC/E) (Berendsen et al., 1987) was used. For SPC/E water model, the intra-molecular interactions have been accounted using harmonic potential and the values of parameters for intra-molecular potential i.e., force constant for inter-atomic bond between oxygen and hydrogen atoms (K_{OH}), equilibrium bond distance between oxygen and hydrogen atom (b_{OH}), force constant for bond angle potential (K_{HOH}) and equilibrium bond angle (θ_0) are 3.45×10^5 kJ mol^{-1} nm^{-2}, 0.1 nm, 3.45×10^2 kJ mol^{-1} nm^{-2} and 109.47° respectively (Thapa & Adhikari, 2013).

During the modeling of the system, we considered both non-bonded: Lennard-Jones and Coulomb interactions and the total intermolecular potential due to the non-bonded interaction is given by

$$U(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_m r_{ij}} \tag{3.56}$$

where r_{ij}, q_i and q_j are the distance between the ith and j^{th} atoms, permittivity of the medium, charge of ith and j^{th} atoms respectively; σ and ε are the distance at which LJ interaction becomes zero and depth of the well due to LJ interaction respectively. For

the SPC/E model, the hydrogen and oxygen atom carries partial charges of $+0.4238e$ and $-0.8476e$ respectively, where e is charge of electron. Also, the non-bonded parameters the distance at which LJ interaction is zero ($\sigma_{\text{OW-OW}}$) and well depth ($\varepsilon_{\text{OW-OW}}$) for oxygen atom of SPC/E water model are 0.3165 nm and 78.2 k_{B} respectively. Also, the combination rules to determine the non-bonded LJ parameters between two distinct atoms for OPLS-AA force field parameters are Lindahl et al. (2010).

$$\sigma_{\alpha\beta} = \left(\sigma_{\alpha\alpha} \times \sigma_{\beta\beta}\right)^{1/2} \tag{3.57}$$

$$\varepsilon_{\alpha\beta} = \left(\varepsilon_{\alpha\alpha} \times \varepsilon_{\beta\beta}\right)^{1/2} \tag{3.58}$$

In addition, other two systems in different sizes of simulation box: 2 amoxicillin in 4071 water molecules in cubic box of size 4.98 nm (System-I(a)) and 2 amoxicillin in 6504 water molecules in cubic box of size 5.82 nm (System-I(b)) were set up to study the effect of system size on diffusion coefficient and estimate its size independent value. Moreover, the solvent environment i.e. viscosity of solvent also has significant role on diffusion phenomenon (Skyner et al., 2015). In this context, other system was set up taking ethanol as solvent. The system consists of binary mixture of 2 amoxicillin molecules and 2168 ethanol molecules in cubic simulation box of size 5.96 nm (System-I(c)). OPLS-AA force field was used during modeling of ethanol.

3.5.7.2   Systems to Estimate Free Energy of Solvation:

Besides the study of transport properties, free energy of solvation of amoxicillin was estimated in two different solvent environments: water and ethanol. In order to estimate the free energy of solvation of amoxicillin, simulations were carried out at 310 K under PBC taking following systems: one amoxicillin in 1020 water molecules (System II) and one amoxicillin in 340 ethanol molecules (System III(a)). We used two different water models: SPC/E and TIP3P (Berendsen et al., 1987; Jorgensen et al., 1983). Amoxicillin and ethanol molecules were modeled using OPLS-AA force field parameters.

At the end of this chapter, we have summarized different steps followed during simulations in schematic diagram. Figure represents the flow chart showing different steps of MD simulation using GROMACS software package.
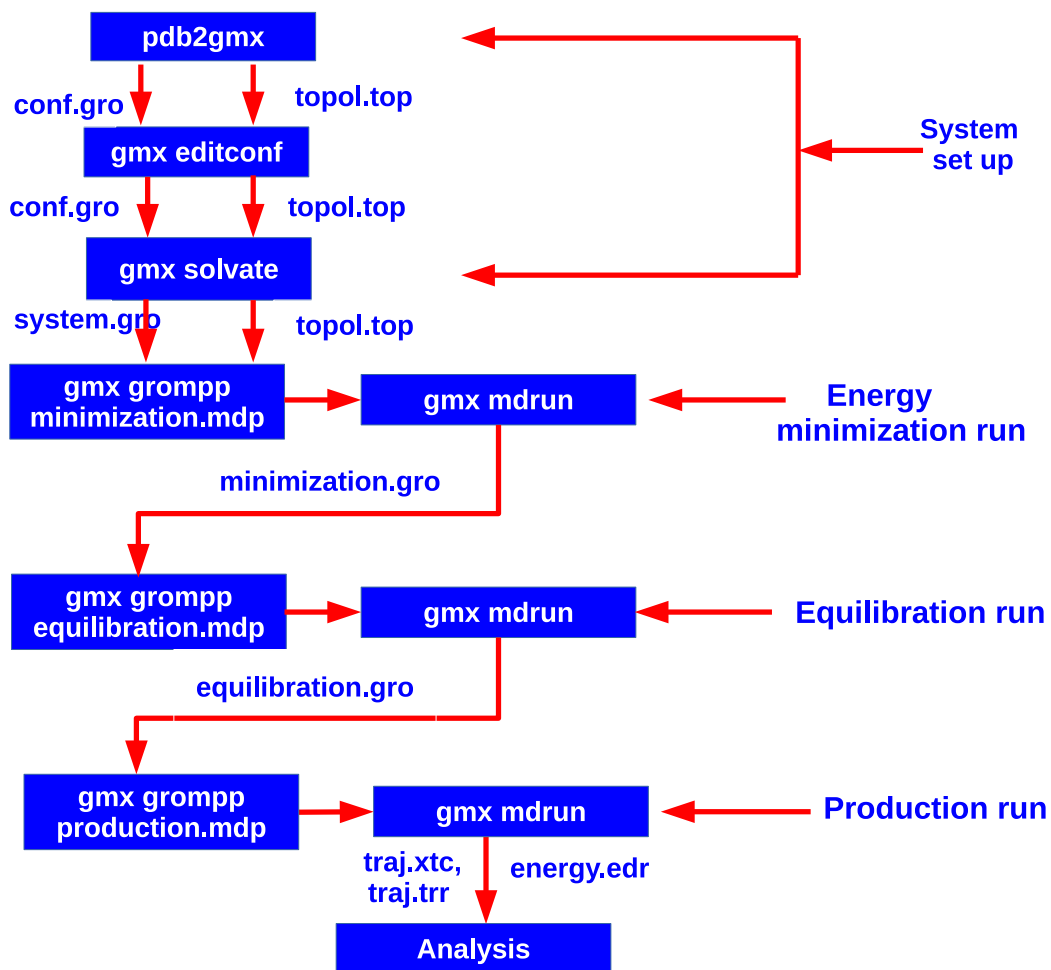
Figure 21: Flow chart showing different steps of molecular dynamics simulation using GROMACS software package.

CHAPTER 4

4. RESULTS AND DISCUSSION

4.1   General Consideration

In this chapter, we have discussed the main findings of the present work. The structural analysis of the System-I has been performed by estimating the RDF between solute-solvent and solvent-solvent at different temperatures in the first section. The study about transport properties has been presented after the structural analysis. In this section, the estimation of self diffusion coefficient of amoxicillin and water is followed by their binary diffusion coefficient; and effect of temperature on diffusion coefficient have been presented. After this, the effect of size of system has been studied; and the size independent value of diffusion coefficient, viscosity coefficient of solvent as well as solution have presented. Finally, we have focused our attention to estimate the free energy of solvation of amoxicillin in two different solvent environments (i. e., water and ethanol) at 310 K temperature using thermodynamic integration (TI) and free energy perturbation (FEP) based method (Lindahl et al., 2010; Bennett, 1976).

At first, we would discuss different steps of MD simulations before analysis of results. After setting up the systems, we first performed the following three runs: Energy minimization (EM), Equilibration and Production for each system at each temperature. Then, the different parameters required to energy minimization, equilibration and production runs with Molecular Dynamics Parameters (MDP) files have been presented for System-I. To remove the van der Waal's bad contact, we first carried out energy minimization run using Steepest-descent method (David et al., 2005). The different parameters used during energy minimization run are presented in the Molecular Dynamics Parameters (MDP) file (see Appendix A.1). The energy minimization run was carried out taking 150 $kJmol^{-1}nm^{-1}$ and 1 nm cut off parameter for both LJ and Coulomb interactions. PME was used to handle long range interaction. After energy minimization run, the system attains the minimum potential energy state (Koirala et al., 2019; Moktan et al., 2012). After energy minimization run, we have plotted a graph of potential energy versus time which is shown in Figure 22.
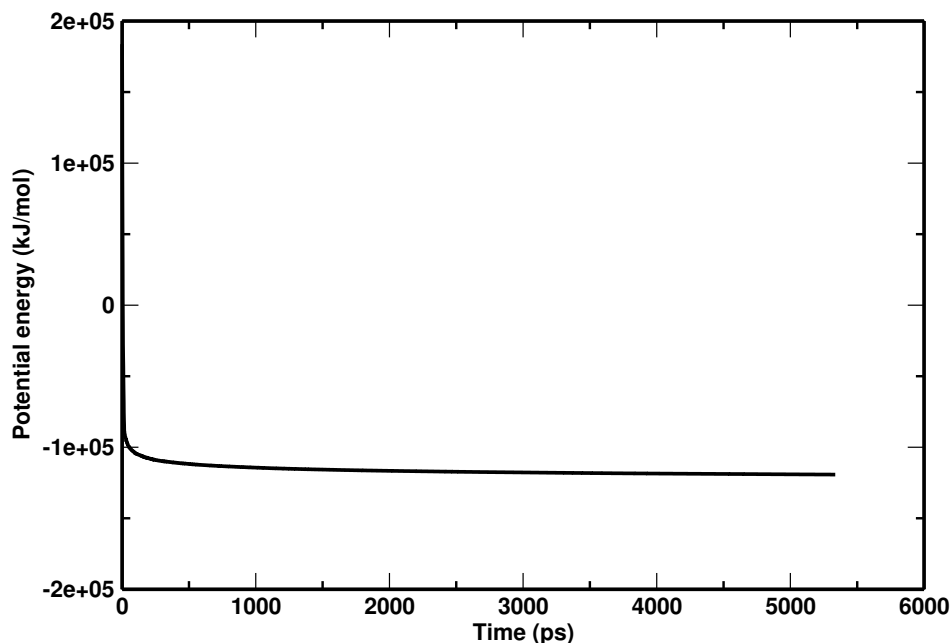
Figure 22: Plot of potential energy of the system after energy minimization.

Many dynamical and thermodynamic properties including diffusion, viscosity, free energy depend upon different parameters like temperature, pressure, density etc. of the system under study (Sharma & Adhikari, 2014). In this regard, each system under study must be in the state of thermodynamic equilibrium; and equilibration run brings the state of thermodynamic equilibrium (Gallo et al., 2009; Koirala, Thapa, et al., 2021). In addition, it is also interesting to study the effect of temperature on transport properties. For this, we performed equilibration run of the System-I at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K under isobaric-isothermal (NPT) ensemble. The MDP parameters used during equilibration run of the system-I at 298 K temperature are presented in MDP files (see Appendix A.2). Each equilibration run was propagated under PBC at 1 atm for each above mentioned five different temperatures. We first assigned initial velocities for each particle using Maxwell-Boltzmann distribution (Andersen, 1980; Frenkel & Smit, 2002). All bonds were subjected to constraint using SHAKE algorithms during equilibration run (Ryckaert et al., 1977). We also considered isothermal compressibility of 4.6×10^{-5} bar^{-1}. For both short range LJ and Coulomb interactions, the cut off parameter of 1 nm was chosen; and long range Coulomb interaction was handled using Particle Mesh Ewald (PME) method (Darden et al., 1993). In order to control the temperature of the system, we used velocity rescaling thermostat with coupling time 0.01 ps; and pressure was controlled by Berendsen barostat with 0.8 ps time of coupling (Bussi et al., 2007; Berendsen et al., 1984). During simulations, the new positions and velocities of each particles are determined after solving equations of motion from old sets. The equations of motion were solved using Leap-frog algorithms (Gunsteren & Berendsen, 1988). At each temperature, we performed

equilibration run for 200 ns taking 0.002 ps time step.

After equilibration run, the temperature and density profiles of the equilibrated system were plotted at each temperature; and also we compared the density of the equilibrated system at each temperature with previously reported experimental values. Figures 23 and 24 show the temperature and density profiles of the system under study at 305 K temperature after equilibration run.
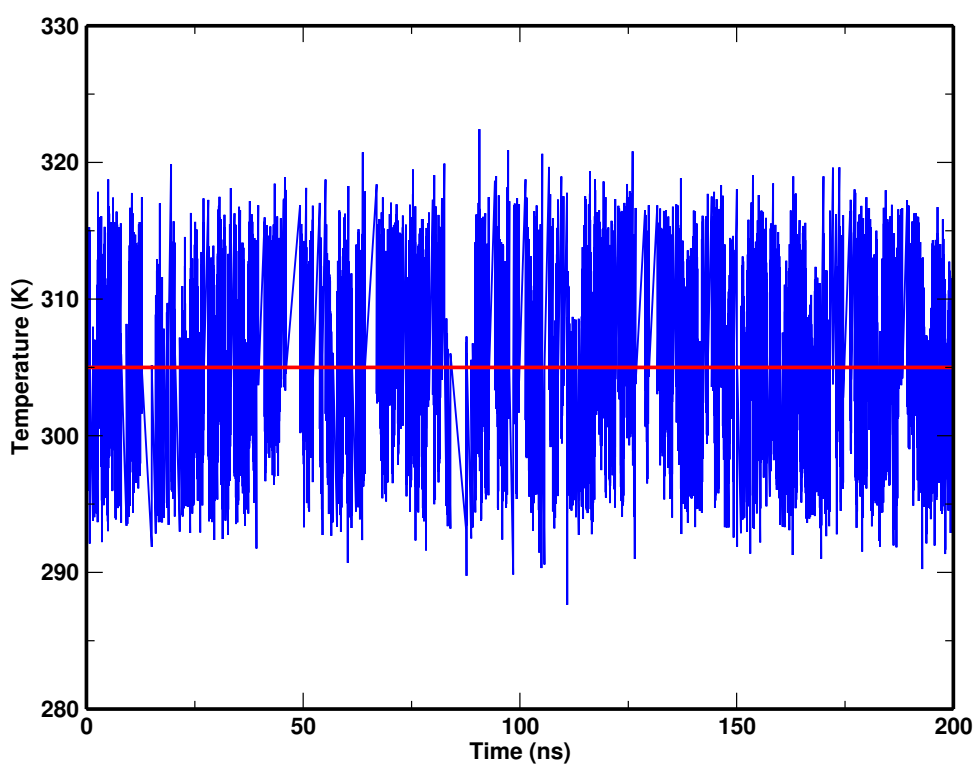
Figure 23: Temperature profile of the system at 305 K temperature after 200 ns equilibration run.
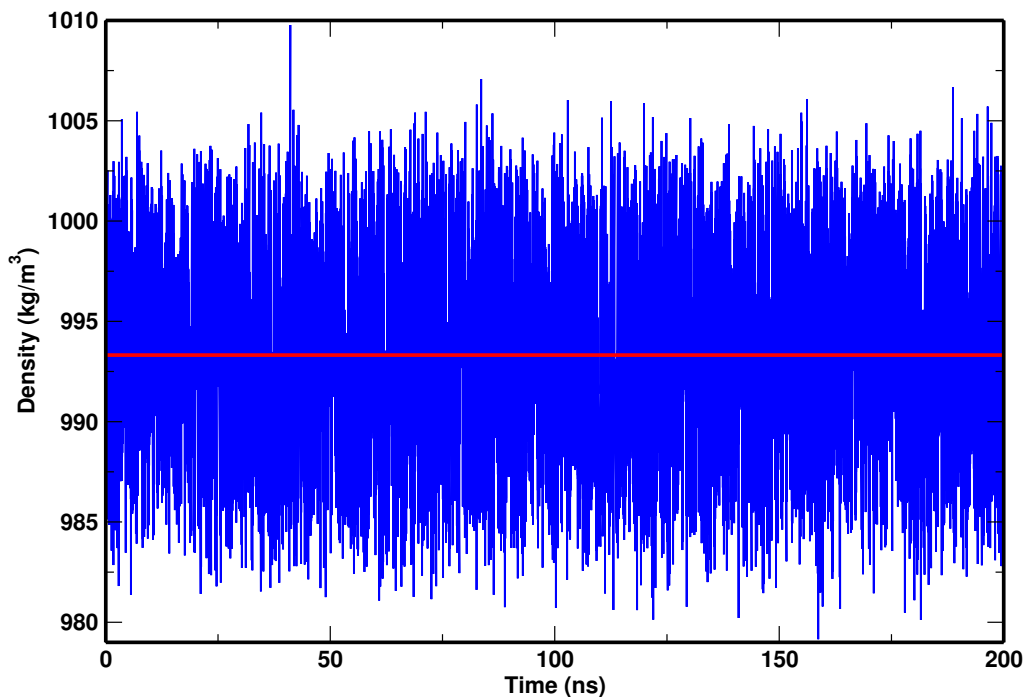
Figure 24: Density profile of the system at 305 K temperature after 200 ns equilibration run.

Table 1 shows the temperature and density profile of the system after 200 ns equilibration run at different temperature with previously reported experimental values of density of water. From the Table, it is seen that the values of simulated density of the system are in closed agreement within 0.5 % with previously reported experimental values of water. This is because our system has large number of water molecules in comparison to 2 amoxicillin molecules.

Table 1: Density and temperature profiles of our system at five different temperatures along with experimental values after 200 ns equilibration run.

S.N	T (K)	T_{sim} (K)	ρ_{system} (kg/m^3)	ρ_{water}(Baysinger, 2015) (kg/m^3)
1	298	298.00 ± 0.01	997.17 ± 0.01	997.05
2	303	303.00 ± 0.01	994.47 ± 0.02	995.65
3	305	305.00 ± 0.03	993.32 ± 0.03	995.03
4	310	310.00 ± 0.01	990.48 ± 0.03	993.33
5	313	314.00 ± 0.01	988.70 ± 0.02	992.22

In addition, we have analyzed the energy profile of the equilibrated system with individual contribution of different interactions. During simulations, we used cut off parameter for both LJ and Coulomb interaction. So, here we present only short range LJ and Coulomb interactions. We plot the contribution of different energies to total energy of the system after 200 ns equilibration run. Although, the total potential energy is the sum of bond angle, dihedral, LJ and Coulomb interactions, the contribution of bonded bond angle and dihedral is small in comparision to other. The contribution of LJ is

48

positive and Coulomb is negative; and the total potential energy is also negative due to large contribution of Coulomb. Furthermore, the kinetic energy also has positive contribution. The total energy is the sum of kinetic energy and total potential energy which has negative value. This indicates that Coulomb has dominant role to stabilize and bound the system under study (Pokharel et al., 2016). Figure 25 shows the energy profile of our equilibrated system at 305 K temperature.
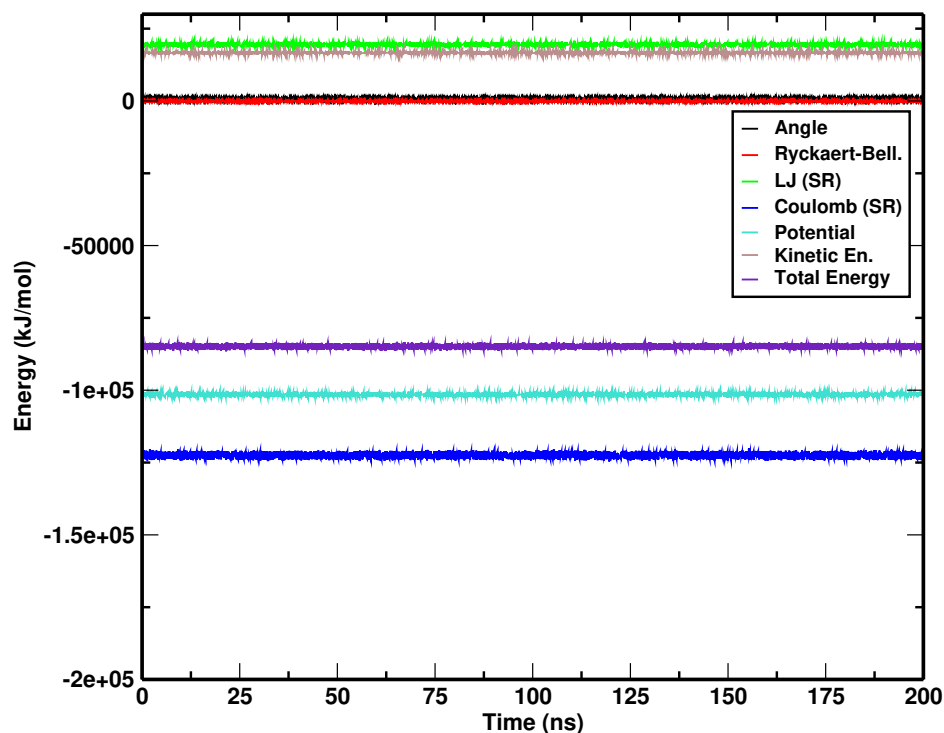
Figure 25: Energy profile of the system at 305 K temperature after 200 ns equilibration run.

In Table 2, we present the contributions of LJ, Coulomb, and kinetic energy to total energy of the system after 200 ns equilibration at different temperatures. From the Table, it is observed that potential due to Coulomb increases with increase in temperature (less negative value with increase in temperature), however, LJ potential decreases with the increase in temperature. The Coulomb has dominating role over LJ interaction, and hence the total potential energy also increases with the increase in temperature. Moreover, the kinetic energy of the system follows increasing trends with temperature. Finally, the total energy, the sum of kinetic and total potential energies of the system, follow the same trend of increasing with temperature.

Table 2: Energy profile of the system at five different temperatures after 200 ns equilibration run.

T (K)	LJ (kJ/ mol)	Coulomb (kJ/mol)	Potential (kJ/mol)	Kinetic (kJ/ mol)	Total (kJ/mol)
298	19867.4 ± 1.0	-123829.0 ± 2.5	-102464.0 ± 1.6	16267.4 ± 0.1	-86196.5 ± 1.5
303	19602.3 ± 0.5	-122897.0 ± 1.2	-101784.0 ± 0.9	16540.2 ± 0.2	-85243.4 ± 1.0
305	19485.1 ± 3.6	-122509.0 ± 7.1	-101516.0 ± 1.2	16649.6 ± 0.1	-84866.6 ± 1.2
310	19236.1 ± 3.5	-121609.0 ± 3.5	-100844.0 ± 0.4	16922.6 ± 0.2	-83921.1± 0.4
313	19087.6 ± 2.7	-121073.0 ± 5.0	-100445.0 ± 1.0	17086.1 ± 0.2	-83359.3 ± 1.1

Finally, the equilibrated system at each temperature was propagated to the production run under isothermal-isochoric (NVT) ensemble. In the beginning of production run, the initial position and velocities of each particle were assigned from the final step of the equilibration run; and Leap-frog algorithm was used to generate trajectory after solving equations of motion. During each production run, pressure coupling was turned off and other parameters were chosen same as in equilibration run. Each production was carried out with 0.002 ps time step for the time of 200 ns. During production run, taking System-I at 298 K temperature, the parameters considered in during production run are presented in MDP file (see Appendix A.3). After the production runs at above mentioned five different temperatures, we have analyzed the trajectory to estimate different properties of our interest including structural and transport.

4.2 Structural Analysis

In this section, we have studied about structural properties of System-I using the concept of RDF (Allen & Tildesley, 1987). For this, at first, the RDF between solvent-solvent and after then solvent-solute of System-I at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K have been plotted from the trajectory obtained from simulations. From the RDF plot, we can get idea about the distribution of molecules around a reference molecule. For liquid, RDF shows an oscillation upto certain distance and becomes unity (Hansen & McDonald, 2013). The region from $r = 0$ up to which the RDF becomes zero is known as excluded region (ER). In order to find the structural properties of the system under study, we have calculated the RDF of oxygen of water and oxygen of water $g_{OW-OW}(r)$, oxygen of water and sulphur of amoxicillin $g_{OW-S}(r)$ and oxygen of water and nitrogen of amoxicillin $g_{OW-N}(r)$ at different temperatures. The van der Waals radius ($r^{1/6} \sigma$) for OW-OW, OW-S and OW-N pairs have also been estimated.

Figure 26 shows the RDF between oxygen atoms of water molecules at temperatures: 298 K, 303 K, 305 K, 310 K and 313 K. From the Figure 26, we have noticed that the graph has oscillatory nature upto three peaks and becomes fairly straight line with value unity as expected. At $r \to 0$, the RDF is zero due to strong repulsive force between molecules contributed by Coulomb interaction along with r^{-12} term of LJ

interaction. There is certain region from r = 0 upto which we observed zero value of RDF i.e., ER region. Three peaks have been noted beyond region of zero probability region. After third peak, the RDF becomes straight with value of unity which means no correlation occurs between molecules in this region (Dahal & Adhikari, 2012). The position of first peak indicates the most favorable position of water molecules from reference water molecule. Furthermore, second and third peak positions indicate other favorable positions. Also, the estimated values of the excluded region (ER), first peak
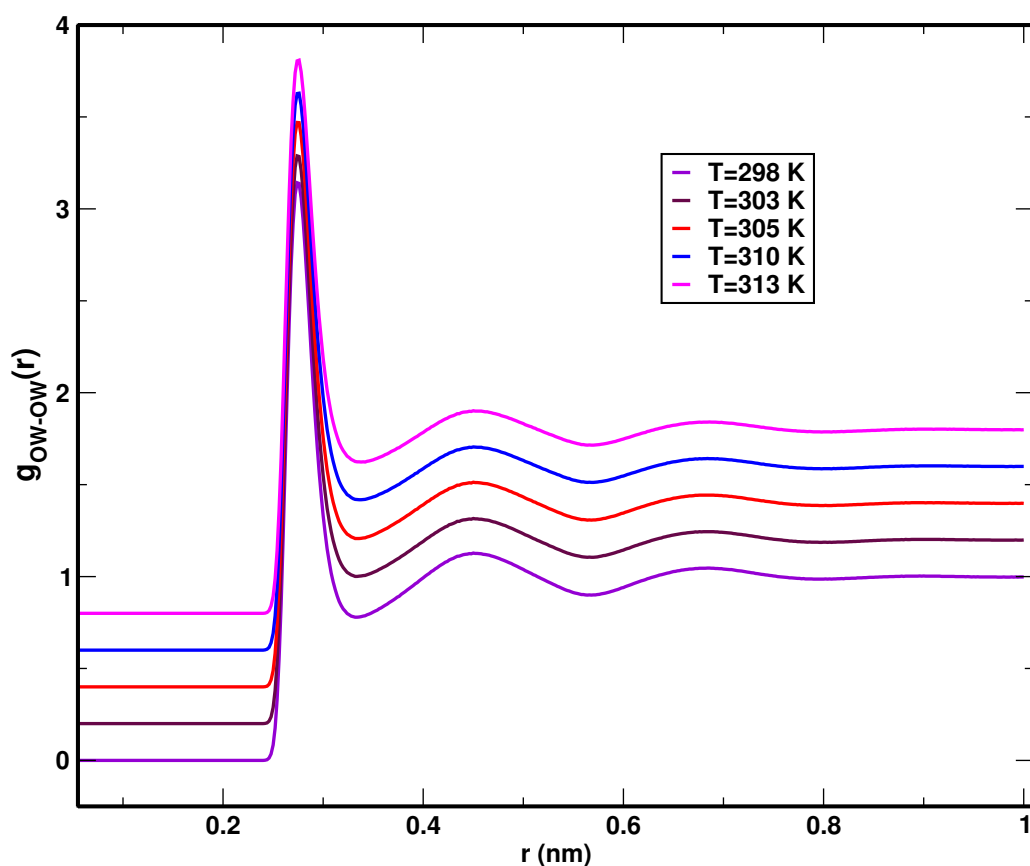
Figure 26: Radial distribution function (RDF) between oxygen atoms of water molecule at five different temperatures.

position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP) and third peak value (TPV) for RDF between oxygen atoms of water molecules (OW-OW) are extended in Table 3.

Table 3: Simulated data for the RDF between the water molecules ($g_{OW\text{-}OW}(r)$) at different temperatures.

T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
298	0.248	0.275	3.135	0.453	1.113	0.683	1.027
303	0.248	0.274	3.063	0.454	1.101	0.686	1.036
305	0.248	0.274	3.058	0.451	1.101	0.686	1.020
310	0.244	0.275	3.024	0.451	1.100	0.687	1.030
313	0.244	0.275	3.004	0.455	1.090	0.685	1.042

Furthermore, the value of (σ_{OW-OW}) and the calculated value of van der Waals radius ($r^{1/6}\sigma$) for OW-OW are 0.316 nm and 0.355 nm respectively (Lindahl et al., 2010). From Table 3, we clearly notice that value of both excluded region (ER) and first peak position (FPP) of RDF between OW-OW are smaller in comparison to corresponding van der Waals radius. This observation signifies that only van der Waals interaction can not contribute to the structure stability i.e., other potentials including Coulomb as well as many body effects along with van der Waals potential contribute to the structural stability (Poudyal & Adhikari, 2014; Khanal, Koirala, et al., 2021).

Similarly, Figure 27 shows the radial distribution function of oxygen atom of water molecule and sulphur atom of amoxicillin molecule at different temperatures. The value of (σ_{OW-S}) is 0.338 nm and the calculated value of van der Walls radius ($r^{1/6}\sigma$) for OW-S is 0.378 nm. The excluded region (ER), first peak position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP) and third peak value (TPV) for OW-S are presented in Table 4.
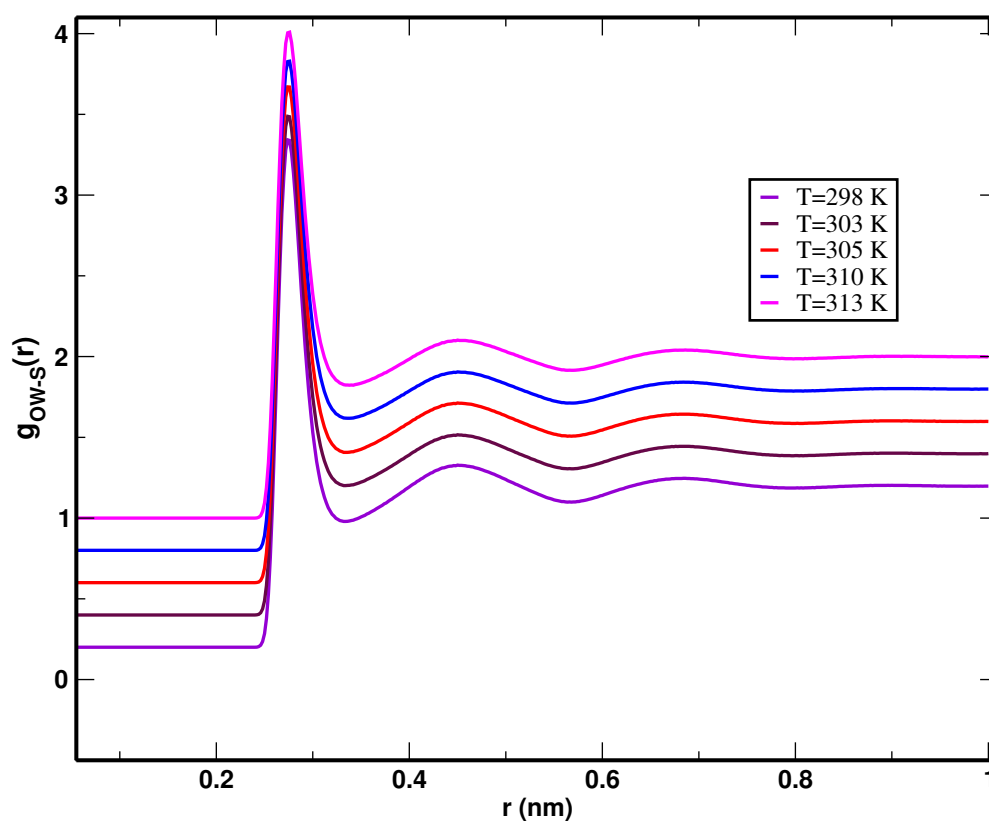
Figure 27: Simulated data for the RDF between the oxygen of water and sulphur of amoxicillin ($g_{OW-S}(r)$) at five different temperatures.

Table 4: Simulated data for the RDF between the oxygen of water and sulphur of amoxicillin ($g_{OW-S}(r)$) at different temperatures.

T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
298	0.241	0.275	3.027	0.451	1.117	0.682	1.038
303	0.241	0.275	3.080	0.452	1.006	0.686	1.040
305	0.241	0.274	3.109	0.452	1.001	0.686	1.037
310	0.241	0.274	3.140	0.452	1.095	0.687	1.042
313	0.245	0.275	3.007	0.455	1.083	0.684	1.032

Figure 28 is the radial distribution function of oxygen atom of water molecule and nitrogen atom of amoxicillin molecule at five different temperatures. The value of (σ_{OW-N}) is 0.321 nm and the calculated value of van der Waals radius ($r^{1/6}\sigma$) for OW-N is 0.360 nm. The excluded region (ER), first peak position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP) and third peak value (TPV) for OW-N are presented in Table 5.
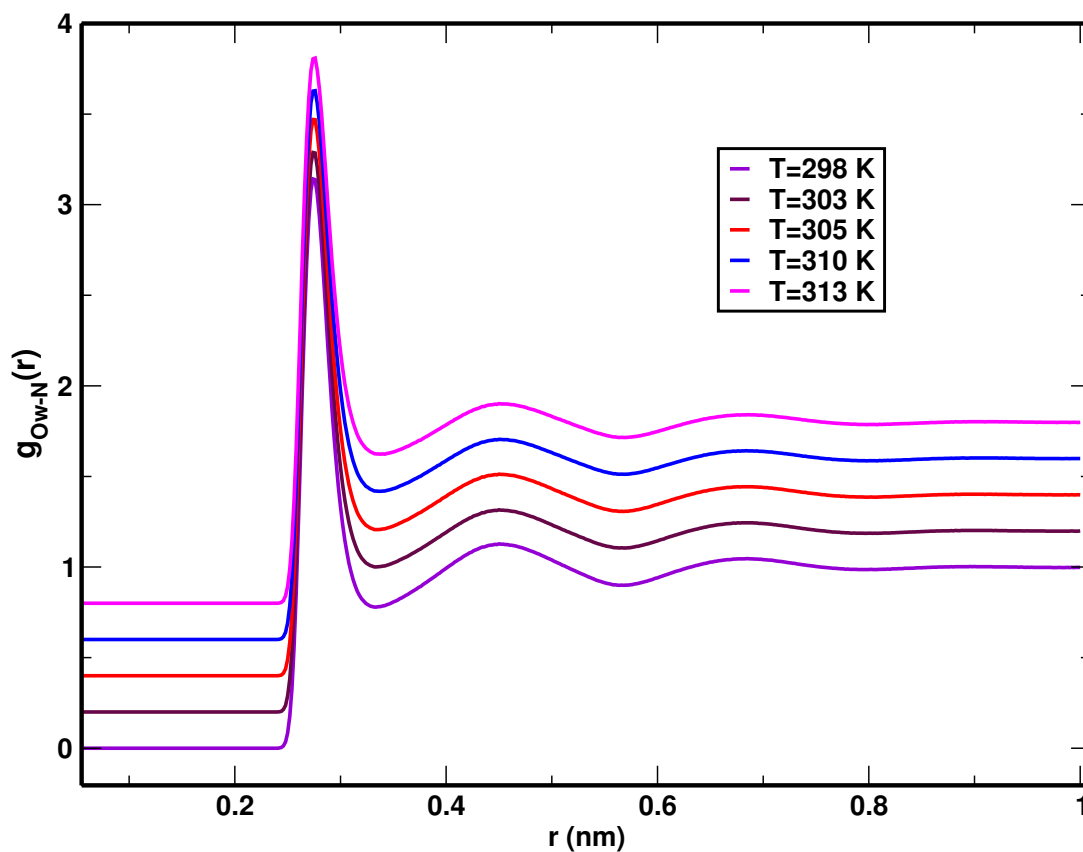
Figure 28: Simulated data for the RDF between the oxygen of water and sulphur of amoxicillin ($g_{OW-S}(r)$) at five different temperatures.

Table 5: Simulated data for the RDF $g_{OW-N}(r)$ analysis between oxygen atom of water molecule and nitrogen atom of amoxicillin molecule at five different temperatures.

T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
298	0.241	0.275	3.139	0.450	1.1226	0.684	1.044
303	0.242	0.274	3.085	0.451	1.110	0.685	1.040
305	0.241	0.274	3.168	0.450	1.011	0.686	1.030
310	0.242	0.274	3.025	0.451	1.090	0.686	1.028
313	0.241	0.275	3.052	0.452	1.079	0.681	1.028

From above mentioned Tables 4 and 5, it is observed that the values of excluded region (ER) and that the first peak position (FPP) are less than corresponding van der Waals radius ($r^{1/6}\sigma$). It indicates that the van der Waals potential as well as other potentials contribute to the structural properties of the system.

4.3 Transport Properties

Besides the study about structural properties, we have studied the transport properties of amoxicillin in water. In this section, we have explained the diffusion phenomenon in terms of diffusion coefficients as well as the effect of temperature and system size on diffusion coefficient. The self diffusion coefficients of solute and solvent of the system under study have been estimated from the trajectory followed by the particles in phase space using Einstein's relation (3.8) and their binary diffusion coefficient using Darken's relation (3.9). In order to estimate self diffusion coefficient using Einstein's relation, we first plotted the MSD versus time graph for both solute and solvent at different temperatures using trajectory from 200 ns production run. The statistics is better in beginning region of trajectory in comparison to ending region (Pokharel et al., 2016). For this region, we truncated the graph between mean square displacement (MSD) versus time and then linearly fitted taking 5 ns for water and 1 ns time for amoxicillin respectively although 200 ns production runs were performed (Sharma & Adhikari, 2014; Bhandari & Adhikari, 2016).

At first, we have plotted MSD versus time graph in logarithmic scale. Figures 29 and 30 show the log-log plots between MSD versus time for water and amoxicillin respectively at 303 K temperature.
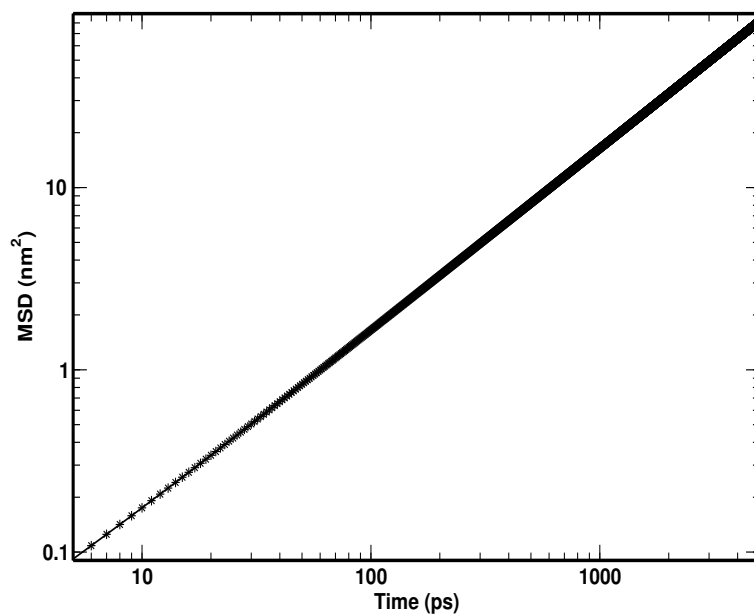
Figure 29: MSD versus time plot in logarithmic scale for water at temperature 303 K.
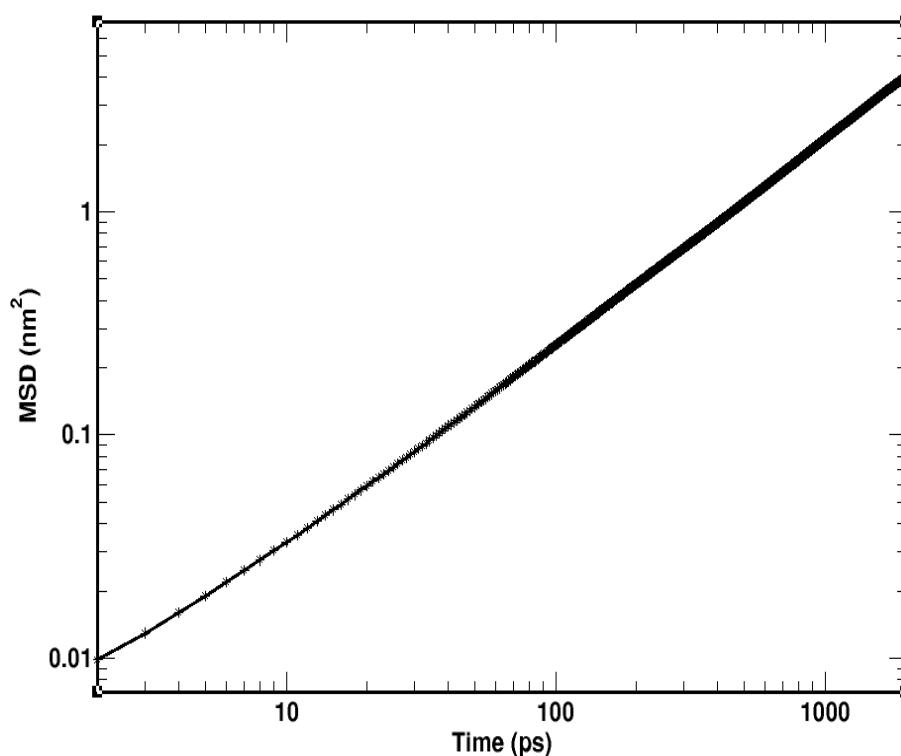
Figure 30: MSD versus time plot in logarithmic scale for amoxicillin at temperature 303 K.

From the plots, it is seen that the graphs are not linear in the beginning which indicates ballistic regime (Caspi et al., 2000). This happens due to ballistic motion of the molecules in the beginning. After ballistic regime, the plots become straight line indicating diffusive regime (Thapa & Adhikari, 2013). The value of self diffusion coefficient is given by the straight portion of the graph. The diffusive regions start beyond around 40 ps and

150 ps time for water and amoxicillin respectively. We have also studied the variation of diffusion coefficient with time. In addition, Figure 31 represents variation of diffusion coefficient of amoxicillin with time at 303 K.
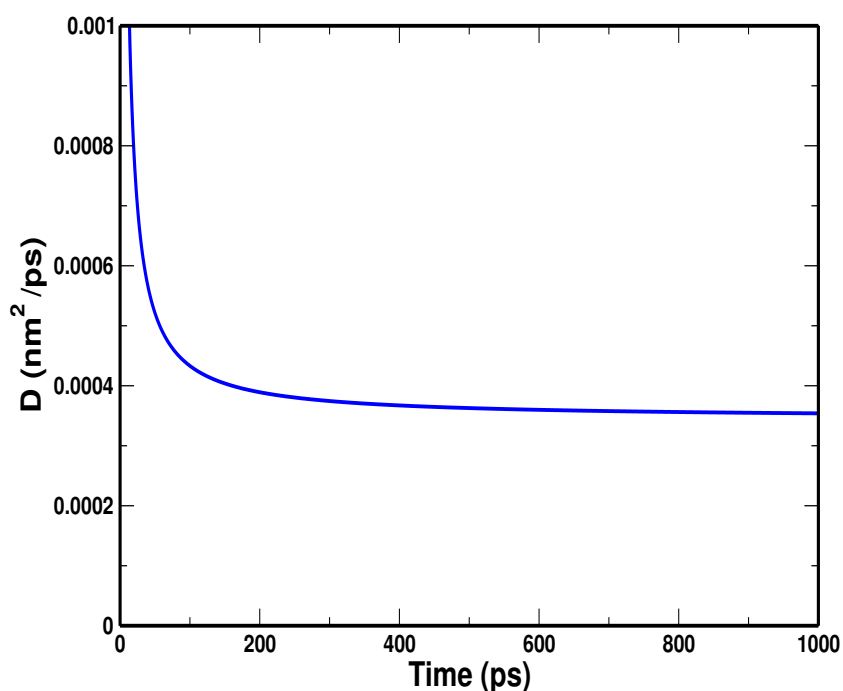
Figure 31: Plot of self diffusion coefficient D versus time of amoxicillin at 303K.

From the Figure, it is clearly seen that the diffusion coefficient is very high at first indicating ballistic regime as also observe in log-log plot; and afterwards, the graph is almost constant with time. The constant value of diffusion coefficient implies that our system reaches at equilibrium. The value of self diffusion coefficient is given by the straight portion of the graph (Pokharel et al., 2016).

From the above MSD versus time plots in logarithmic scale and diffusion coefficient versus time plot, we concluded that the beginning portion of graphs taking time 1 ns and 5 ns are sufficient to estimate the diffusion coefficient although we performed 200 ns production runs.

Now, we have discussed the estimation of self diffusion coefficients of solute and solvent at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K using Einstein's relation. For this, we have first plotted the MSD versus time graphs for both solute and solvent at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K; and the data have been linearly fitted taking 5 ns and 1 ns time for water and amoxicillin respectively. The MSD versus time graphs for water and amoxicillin at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K are presented in the Figures 32 and 33 respectively (Khanal & Adhikari, 2022).
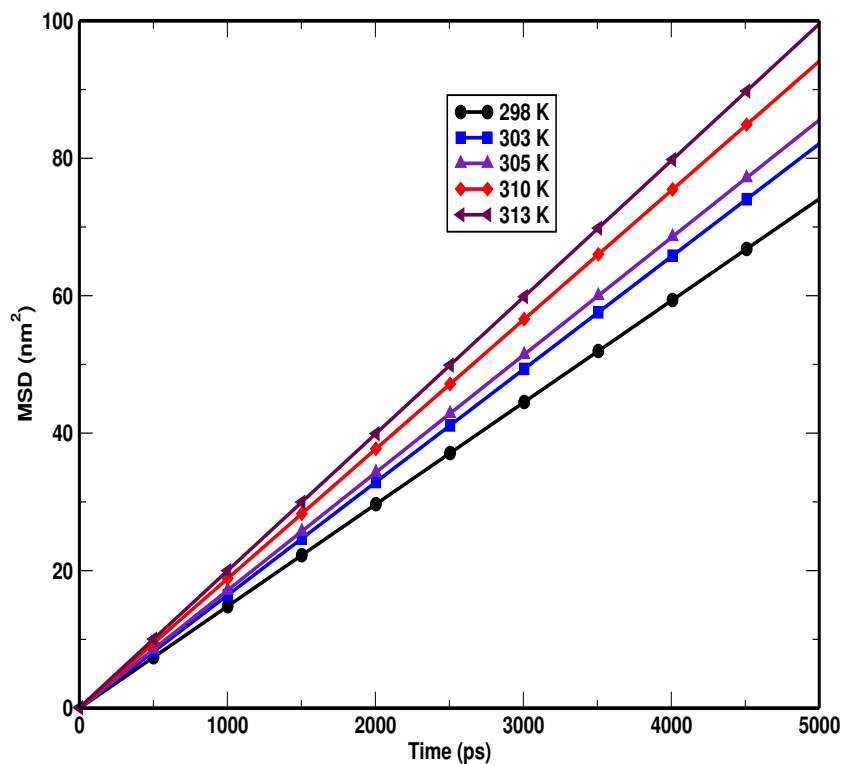
Figure 32: Variation of MSD with time for water at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K.
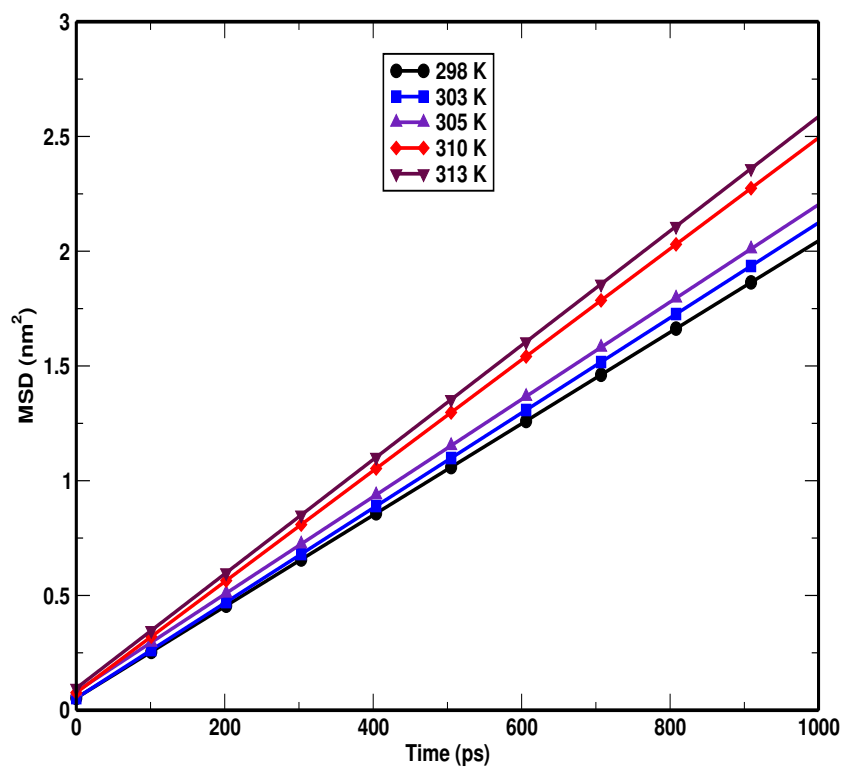
Figure 33: Variation of MSD with time for amoxicillin at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K.

The slope of the Figures 32 and 33 provides the value of self diffusion coefficient. From

both Figures, it has been clearly noticed that the slope of graph increases with increase in temperature. This indicates that the self diffusion coefficient of both solute and solvent goes increasing with temperature. This happens because the kinetic energy and hence velocity of diffusing particles increases with increase in temperature (Mehrer, 2007; Bhandari & Adhikari, 2016). In addition, with the increase in temperature, the density of system decreases; and thus more space is available for random walk of diffusive particles (Sharma & Adhikari, 2014). The estimated values of self diffusion coefficients of solute and solvent at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K are presented in Table 6 (Khanal & Adhikari, 2022).

Table 6: The estimated values of self diffusion coefficients of solute and solvent as well as their mutual diffusion coefficient at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K.

Diffusion Coefficient ($\times 10^{-9}$ m^{2}s^{-1})				
	Self			
Temperature (K)	For amoxicillin	For water		Mutual
	MSD	MSD	Ref.	
298	0.30±0.05	2.49±0.02	2.30 [a]	0.30
303	0.34±0.04	2.74±0.01	2.60 [a]	0.35
305	0.35±0.03	2.85±0.01	2.80 [b]	0.36
310	0.40±0.04	3.14±0.01	–	0.41
313	0.43±0.04	3.32±0.00	3.22 [a]	0.43

Ref. [a]- (Holz et al., 2000) and [b]- (Pokharel et al., 2016)

Table 6 shows that the estimated values of self diffusion coefficient of both amoxicillin and water increases with the increase in temperature. The reason behind increase in the diffusion coefficient with temperature is the increase in kinetic energy and thus velocity of each diffusive particles with increase in temperature (Pokharel et al., 2016). Furthermore, we have compared the estimated values of self diffusion coefficient of water with previously reported experimental values; and observed that the estimated values of self diffusion coefficient of water are in close agreement (within 8.5 %) with previously reported experimental values (Holz et al., 2000). Holz co-worker used H. PFG NMR method for the determination of diffusion coefficient.

In addition, the mutual diffusion coefficient of the System-I has been calculated using Darken's relation (Darken, 1948). The calculated values are presented in Table 6. The values of mutual diffusion coefficient also increase with increase in temperature. In addition, the calculated values of mutual diffusion coefficient are almost equal to the corresponding self diffusion coefficient of amoxicillin. This happens due to low concentration of amoxicillin in the our system.

4.3.1 Temperature Dependence of Diffusion Coefficient

The temperature dependence of diffusion coefficient (shown in Table 6) suggests that diffusion is a phenomenon that is thermally activated (Leach, 2001). It is generally observed that diffusion coefficient follows Arrhenius equation (Mehrer, 2007):

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

In this equation, D_0, E_a, N_A, k_B and T are pre-exponential factor, activation energy, Avogadro number (6.02×10^{23} mol^{-1}), Boltzmann constant (1.38×10^{-23} J/K) and absolute temperature respectively (Mohr et al., 2008). The activation energy means the energy required to confirm diffusion process. The above equation can be written as,

$$\ln\left(\frac{D}{D_0}\right) = -\left(\frac{E_a}{N_A k_B T}\right) \tag{4.1}$$

Equation 4.1 implies that if diffusion coefficient obey Arrhenius behavior, the graph between $\ln(D)$ versus $1/T$ must be straight such that slope of the straight line provides activation energy.

To check the Arrhenius behavior of diffusion coefficient, we have plotted the graph between $\ln(D)$ versus $1/T$ for self diffusion of both solute and solvent. Figures 34 and 35 show the variation of self diffusion coefficient of amoxicillin and water with temperature respectively (Khanal & Adhikari, 2022).
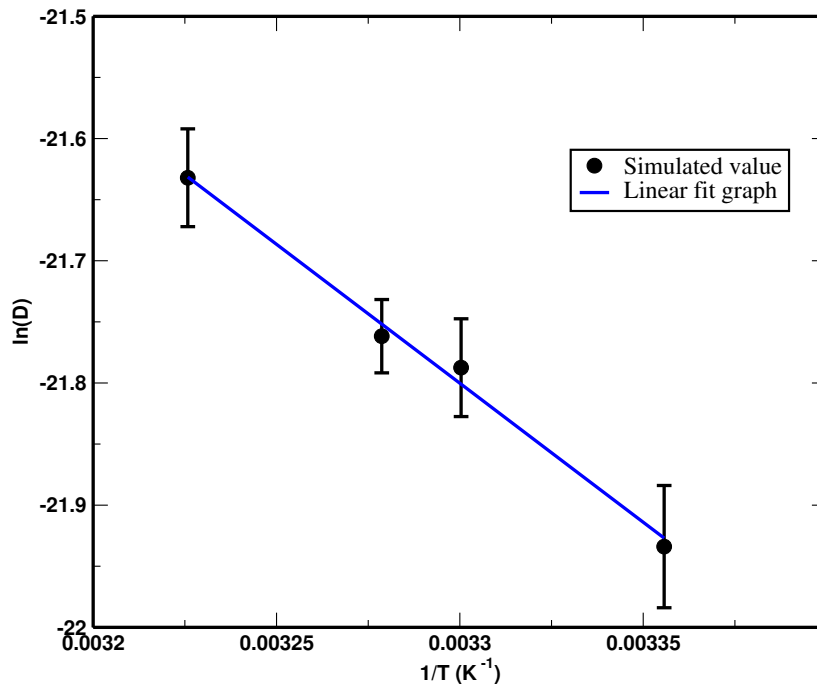
Figure 34: Arrhenius plot (i.e., graph between $\ln(D)$ versus ($1/T$) from the estimated values of self diffusion coefficient of amoxicillin from simulations.

From the Figure 34, it has been observed that the self diffusion coefficient of amoxicillin follows Arrhenius behavior. Also, we can estimate the value of activation energy for diffusion of amoxicillin from the slope of the graph. And the estimated value of activation energy for diffusion of amoxicillin is 0.017 kJ/mol.
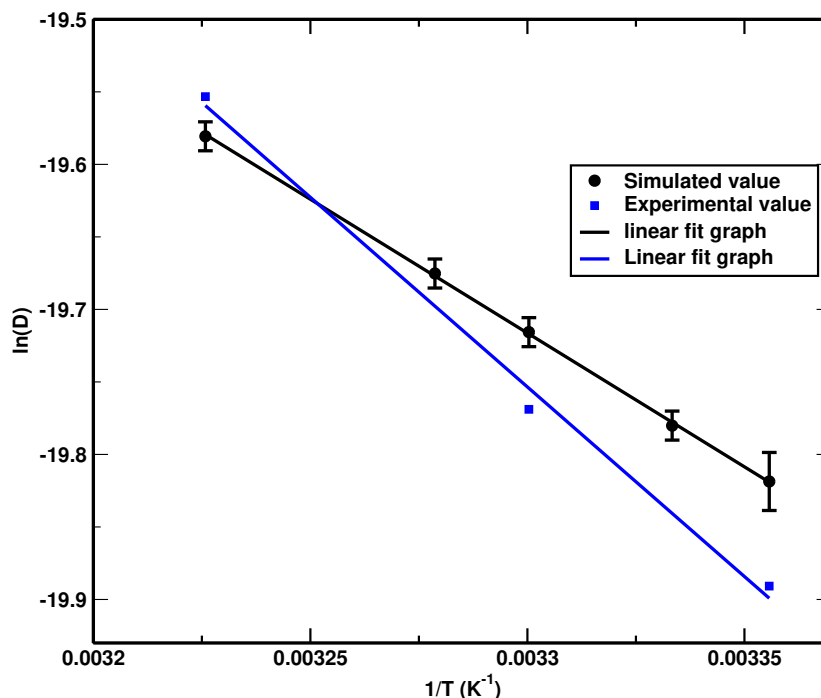
Figure 35: Arrhenius plot (i.e. Graph between ln(D) versus 1/T) from the estimated and experimental values of self diffusion coefficient of water.

From the Figure 35, it has been noticed that the self diffusion coefficient of water also shows Arrhenius behavior. Deviation between simulated and experimental values large at low temperature range in comparison to high temperature range. This may be due to different factors like force field parameters used are suitable in high temperature in comparison to low temperature, accuracy of experimental data etc..

Table 7: Activation energy of water estimated from its simulated and experimental values of self diffusion coefficient.

Activation energy of water (kJmol^{-1})	
Simulated data	Experimental data
15.31	17.43 (Holz et al., 2000)

The activation energy for diffusion of water from simulated values has also been calculated and compared with the value calculated from experimental values of self diffusion coefficient of water. The estimated and experimental values of activation energy are presented in Table 7. From the Table, it is seen that the calculated value of activation energy of water is in close agreement (within 12.2 %) with experimental value (Holz et al., 2000; Easteal et al., 1989).

On the other hand, the diffusion phenomenon also depends upon the solvent environment (Skyner et al., 2015). So, to understand diffusion process in different solvents, we performed simulation taking another System-I(c) in which ethanol is solvent at 298 K under PBC. System-I(c) consists of binary mixture of 2 amoxicillin and 2168 ethanol molecules in cubic box of size 5.96 nm. During the simulation of the System-I(c), same molecular dynamics parameters were used as in System-I except cut off parameter for both non-bonded LJ and Coulomb interactions. Equilibration run was performed after energy minimization followed by production run. We considered 1.2 nm cut off for both LJ and Coulomb interactions during equilibration and production runs. We performed 200 ns simulation for both equilibration and production runs with 0.002 ps time step. After production run, the trajectory has been analyzed to estimate the self diffusion coefficient of amoxicillin and ethanol along with their binary diffusion coefficient. To estimate the self diffusion coefficient of solute and solvent, MSD has been plotted as a function of time for both ethanol and amoxicillin as shown in Figures 36 and 37 respectively.
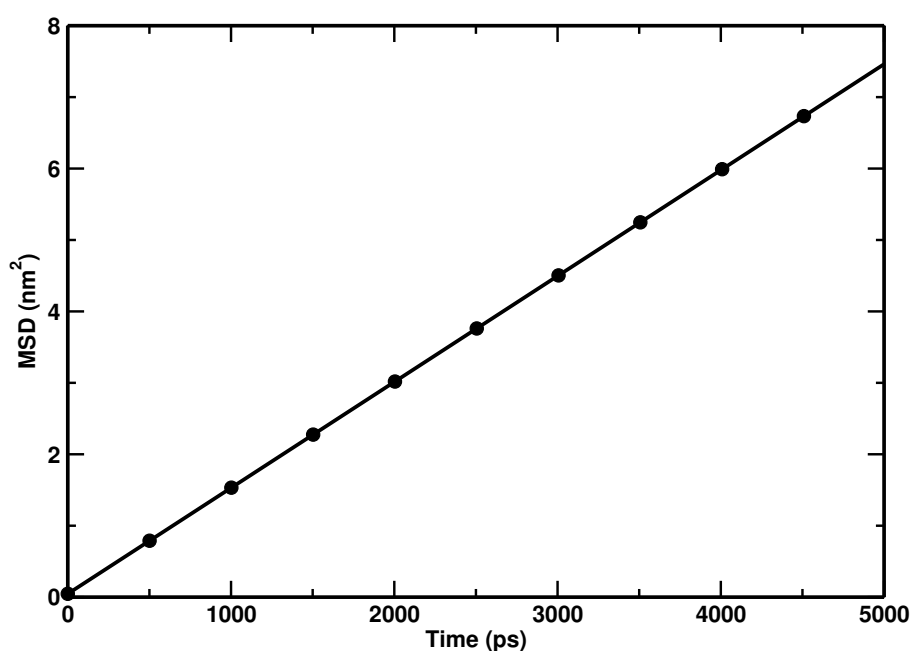
Figure 36: MSD versus time graph for ethanol after 200 ns production run taking System-I(c) at 298 K temperature.
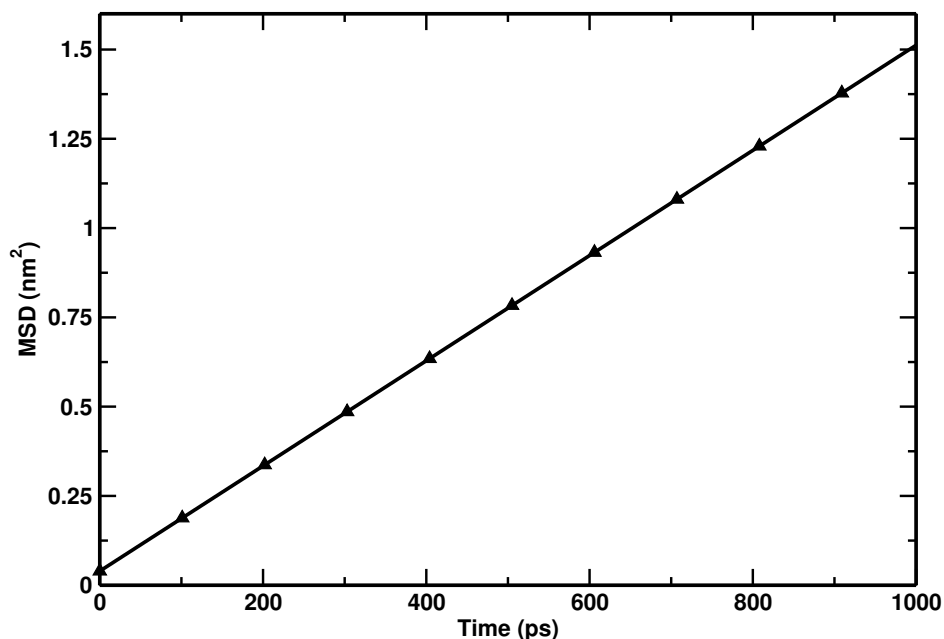
Figure 37: MSD versus time graph for amoxicillin after 200 ns production run taking System-I(c) at 298 K temperature.

Although, we have propagated production run for 200 ns, we have truncated the graphs after 5 ns and 1 ns for ethanol and amoxicillin respectively. This is because the statistic is better in beginning part of trajectory in comparison to ending region. From the slope of the graphs, we determine the self diffusion coefficient of both solute and solvent. The estimated values of self diffusion coefficients of solute and solvent with mutual diffusion coefficient are demonstrated in Table 8. The mutual diffusion coefficient has been estimated using Darken's relation (3.9).

Table 8: The estimated values of self diffusion coefficient of amoxicillin and ethanol as well as their mutual diffusion coefficient at 298 K temperature after 200 ns production taking System-I(c).

Diffusion Coefficient ($\times 10^{-9}$ m^{2}s^{-1})				
Self				Mutual
Amoxicillin	Ethanol			
Estimated	Estimated	Expt.	% Error	
0.26 ± 0.01	1.29 ± 0.02	1.05 (Rathbun & Babb, 1961)	22.8	0.26

We have also compared the estimated value of self diffusion of ethanol with previously reported value. The estimated value is in agreement with previously reported value within 23 %. The accuracy on estimated values from simulations depends upon many factors including force field parameters and modeling of the system. Also, the previously reported value depends upon experimental technique that is used to measure it.

4.3.2 Effect of System Size

The diffusion phenomenon also depends upon size of the simulation box under PBC (Dünweg & Kremer, 1993; Yeh & Hummer, 2004). This happens because of long range nature of hydrodynamics, which causes screening effect under periodic boundary conditions (Koniakhin et al., 2015; Jamali et al., 2018). Infinite size of simulation box is not viable in practice due to computational capacity and finite size box affects on diffusion coefficient (Reed & Flurchick, 1996; De Souza & Ornstein, 1997). Hence, in addition to the effect of temperature on diffusion, we studied the effect of system size in diffusion processes.

To understand the size effect on diffusion coefficient, simulations were carried out by considering another system, glycine as solute and water as solvent. For this system, GROMOS53A6 force field and SPC/E water model were taken (Oostenbrink et al., 2004, 2005; Berendsen et al., 1987). During modeling, we considered united atom model for CH_2. We performed simulations at 298.2 K under PBC by taking three systems of different size with different number of solute and solvent molecules: 3 glycine and 2078 water molecules in cubic box of size 3.98 nm (System-III(a)), 5 glycine and 3464 water molecules in cubic box of size 4.72 nm (System-III(b)) and 7 glycine and 4849 water molecules in cubic box of size 5.28 nm (System-III(c)). The number of solute and solvents were chosen in such a way that the mole fraction remains almost equal.

To remove bad contact between particles in aforementioned systems, we first performed energy minimization run using Steepest-descent method followed by equilibration run (David et al., 2005). Each equilibration run was done under NPT ensemble at 1 atm pressure; and the initial velocities for each particles were allocated using Maxwell-Boltzmann distribution (Andersen, 1980; Frenkel & Smit, 2002). We considered 1 nm cut off parameter for both LJ and Coulomb interactions; and PME method was selected to account long range Coulomb interaction (Darden et al., 1993). All bonds were constraint through LINCS algorithms (Hess et al., 1997). Isothermal compressibility of 4.6×10^{-5} bar^{-1} was considered. During equilibration run for each of above three systems, the temperature was controlled using velocity rescaling thermostat with coupling time 0.01 ps; and to control pressure, Berendsen barostat with 0.8 ps time of coupling was chosen (Bussi et al., 2007; Berendsen et al., 1984). After assigning initial velocities at initial state, the new positions and velocities of each particle at each time step were determined by solving equation of motion using Leap-frog algorithms (Gunsteren & Berendsen, 1988). After thermodynamic equilibrium attained by each system, production run was propagated under NVT-ensemble by taking initial velocities of each particles from the final stage of the equilibration run. All parameters used during production run were same as in equilibration run except barostat which was turned off. Both equilibration as well as production runs for each of the three systems were executed for

200 ns with 0.001 ps time step.

After production run for each system, MSD versus time graphs were plotted for both solute as well as solvent. Although, we performed 200 ns production run, the statistic is better in beginning region. So, only small portion of each graph (5 ns) has been taken and then linearly fitted. From slope of the linearly fitted lines, self diffusion coefficients of both solute and solvent have been estimated using (3.8); and estimated values are extended in Table 9 (Khanal et al., 2019).

Table 9: Estimated values of self diffusion coefficients of solute (i.e. glycine) and solvent (i.e. SPC/E water) from 200 ns production run for three systems of different sizes at 298.2 K temperature.

System	Size (L) (nm)	Diffusion coefficients (D_{PBC})($\times 10^{-9}$m^2 s^{-1})					% Error
		Self			Mutual		
		Glycine	Water				
		Estimated	Estimated	Expt.	Calculated	Expt.	
System-III(a)	3.98	1.13	2.57		1.13		6.60
System-III(b)	4.72	1.16	2.62	2.30 [d]	1.16	1.06 [e]	9.43
System-III(c)	5.28	1.17	2.64		1.17		10.94

Ref: [d]: (Holz et al., 2000) and [e]: (Longsworth, 1953)

From the Table 9, it is clearly understood that the value of diffusion coefficient increases with the increase of size of box (*L*) as expected. The reason behind this observation is that we performed all simulations in finite size box under PBC. In case of finite size box, screening effect due to hydrodynamics interactions of long range play role on estimated value of diffusion coefficient. To overcome this problem, one can perform simulations with large size system which demands high computational capacity. In this regards, Dünweg & Kremer (1993) and Yeh & Hummer (2004) suggested that some modification is needed in the estimated values of diffusion coefficient from simulations under PBC; and the Equation (3.10) gives the necessary correction terms on such diffusion coefficient. From the Equation (3.10), we can also determine the size independent value of diffusion coefficient (D_0) as well as viscosity coefficient (η) if we plot graph between diffusion coefficient estimated from simulation under PBC (D_{PBC}) versus reciprocal of size of simulation box ($1/L$). The intercept and slope of graph between D_{PBC} and $1/L$ offer the values of D_0 and η respectively.

We have plotted graph between D_{PBC} versus ($1/L$) for solvent i.e., water and solution of glycine in water; and Figures 38 and 39 represent the plots for water and solution of glycine in water respectively at 298.2 K.
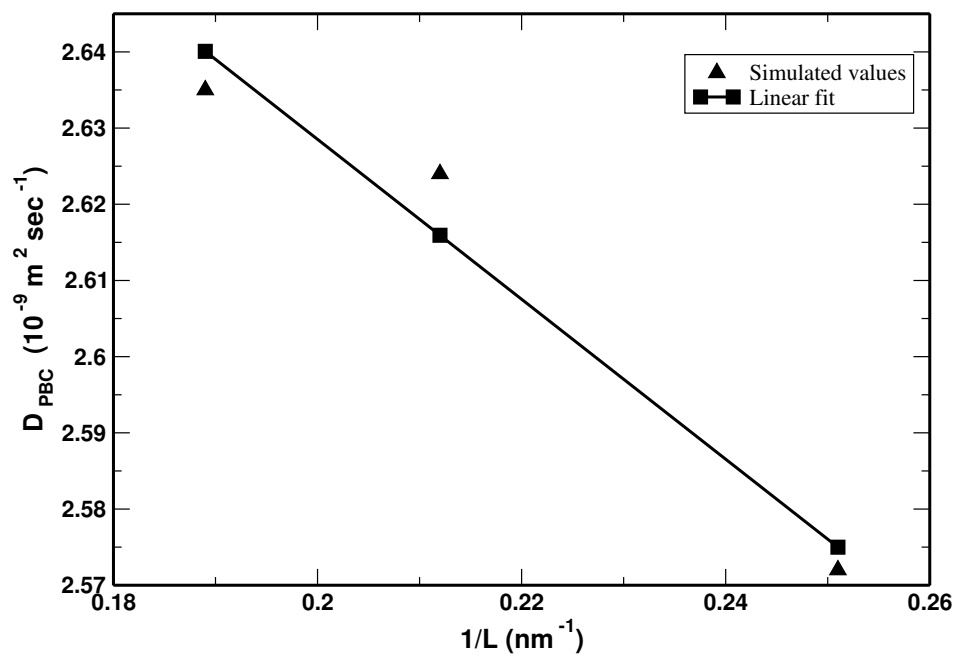
Figure 38: Plot between estimated values of diffusion coefficient (D_{PBC}) of water from simulations under PBC versus reciprocal of size of simulation boxes ($1/L$) at 298.2 K temperature.
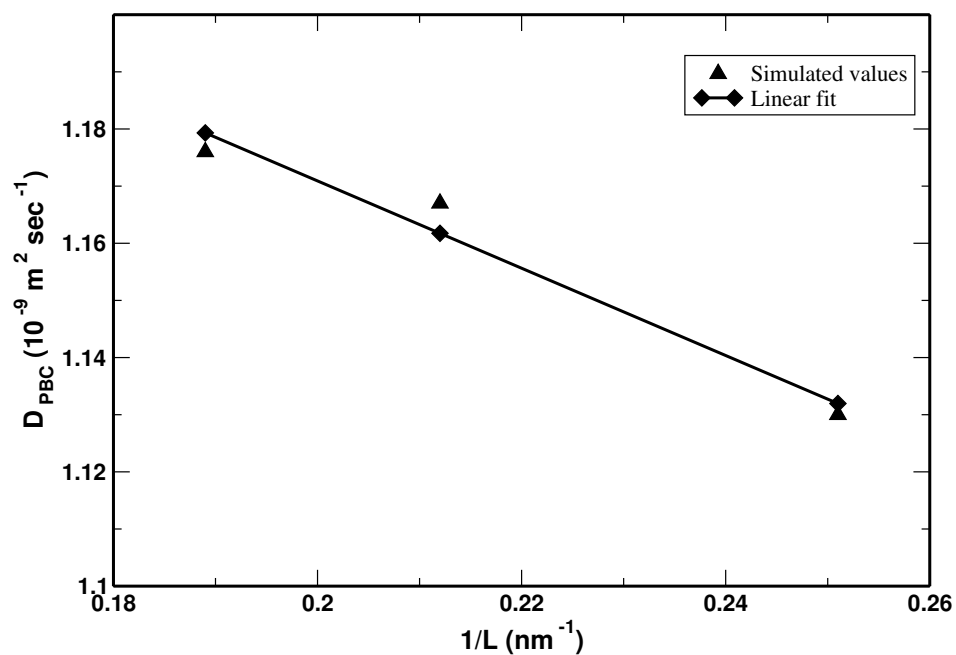
Figure 39: Plot between estimated values of diffusion coefficient (D_{PBC}) of solution (glycine in water) from simulations under PBC versus reciprocal of size of simulation boxes ($1/L$) at 298.2 K temperature.

Table 10: Calculated values of system-size independent values of diffusion coefficient (D_0) of solvent (i.e., water) and solution (i.e., glycine in water) with previously reported values at 298.2 K temperature taking SPC/E water model.

System	D_0 ($\times 10^{-9} m^2 s^{-1}$)			% Error
	Calculated	Ref.	Expt.	
Solvent	2.83	2.97 (Tazi et al., 2012)	2.30 (Holz et al., 2000)	23.04
Solution	1.32	–	1.06 (Longsworth, 1953)	24.53

The estimated values of D_0 for solvent (i.e., SPC/E water model) and solution (i.e., glycine in water) are presented in Table 10 (Khanal et al., 2019). From Tables 9 and 10, it is clearly noticed that the system size independent values of diffusion coefficient (D_0) for solvent as well as solution are higher than the corresponding values of diffusion coefficients estimated from simulations in finite size simulations box under PBC (D_{PBC}) as expected. From this observation, we conclude that diffusion coefficient estimated from simulations under PBC also depends upon size of system.

In addition, we have also compared the calculated values with previously reported value obtained from simulation as well as experimental methods (10). And, it is observed that the calculated value D_0 for solvent is better than previously reported value obtained from simulations, however, the values for both solvent as well as solution are higher than previously reported experimental values. This is because SPC/E model was used during simulations which overestimate the diffusion coefficient although it is better than other models (Tazi et al., 2012). The estimated values are in agreement (within < 25 %) of experimental values.

The viscosity coefficient (η) of both solvent as well as solution have been calculated from the slopes of the Figures 38 and 39; and the calculated values are shown in Table 11 (Khanal et al., 2019). The calculated values of viscosity coefficient of both solvent as well solution are smaller than the corresponding previously reported values. As we used the SPC/E water model during our simulations which underestimate the viscosity coefficient; and transferable intermolecular potential 4 point (TIP4P/2005) water model is better for viscosity coefficient as suggested by Markesteijn et al. (2012).

Table 11: Calculated values of viscosity coefficient η of solvent (i.e., water) and solution (i.e., glycine in water) with previously reported values at 298.2 K temperature taking SPC/E water model.

System	η ($\times 10^{-4}$ Nm^{-2} sec)			% Error
	Calculated	Ref.	Expt.	
Solvent	5.91	6.40 (Tazi et al., 2012)	8.90 (Harris & Woolf, 2004)	33.60
Solution	8.10	–	8.99 (Yan et al., 1999)	9.88

Moreover, the accuracy of results from simulations also depends upon the water models used during system set up (Tazi et al., 2012; Markesteijn et al., 2012). Due to distinct quality of different water models, selection of model must be suitable for particular properties of interest (Tazi et al., 2012). In this regard, we motivated to estimate diffusion coefficient using another four point water model, i.e., transferable intermolecular potential 4 point (TIP4P/2005) (Abascal & Vega, 2005). For this, simulation was performed at 298.2 K taking TIP4P/2005 water model. After simulation, MSD versus time graph was plotted. Figure 40 shows the MSD versus time graph for two water models: SPC/E and TIP4P/2005 water models at 298.2 K temperature.
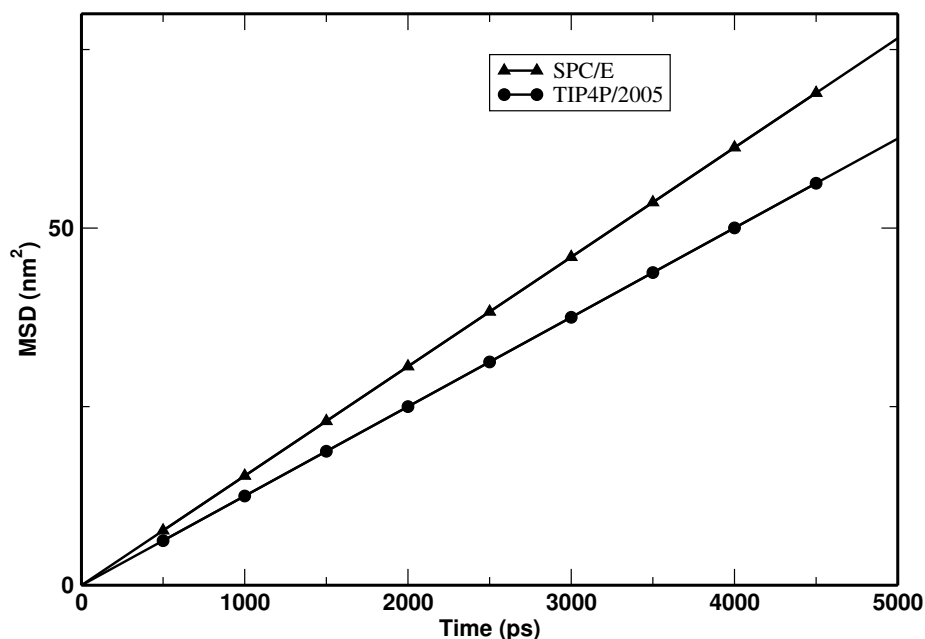
Figure 40: MSD versus time plot for two water models: SPC/E and TIP4P/2005 at temperature 298.2 K.

From the Figure 40, it is distinctly noticed that the estimated value of diffusion coefficient also strongly depends on water model selected during simulation. The simulated values of self diffusion coefficient as well as their previously reported values for different water models are presented in Table 12 (Khanal et al., 2019).

Table 12: Estimated values of self diffusion coefficient of water with previously reported values for two different water models: SPC/E and TIP4P/2005 at 298.2 K temperature.

Water models	Self diffusion coefficients ($\times 10^{-9} m^2 s^{-1}$)			
	Estimated	Ref.	Expt.	% Error
SPC/E	2.55 ± 0.05	2.49 (Shirts & Pande, 2005b)	2.30 (Holz et al., 2000)	10.87
TIP4P/2005	2.08 ± 0.01	2.08 (Abascal & Vega, 2005)		9.56

It is observed from the Table 12 that the estimated value of self diffusion coefficient from the simulation is exactly equal to previously reported values that obtained from simulation for TIP4P/2005 (Abascal & Vega, 2005; Shirts & Pande, 2005b). The estimated value of self diffusion coefficient is smaller for TIP4P/2005 and greater for SPC/E than previously reported experimental value of water. The estimated values are in close agreement with experimental value within 10 % and 11 % for TIP4P/2005 and SPC/E models respectively (Holz et al., 2000). Moreover, the SPC/E water model overestimate the diffusion coefficient but TIP4P/2005 underestimates per expectation and shown in Figure 41 (Shirts & Pande, 2005b; Tazi et al., 2012).
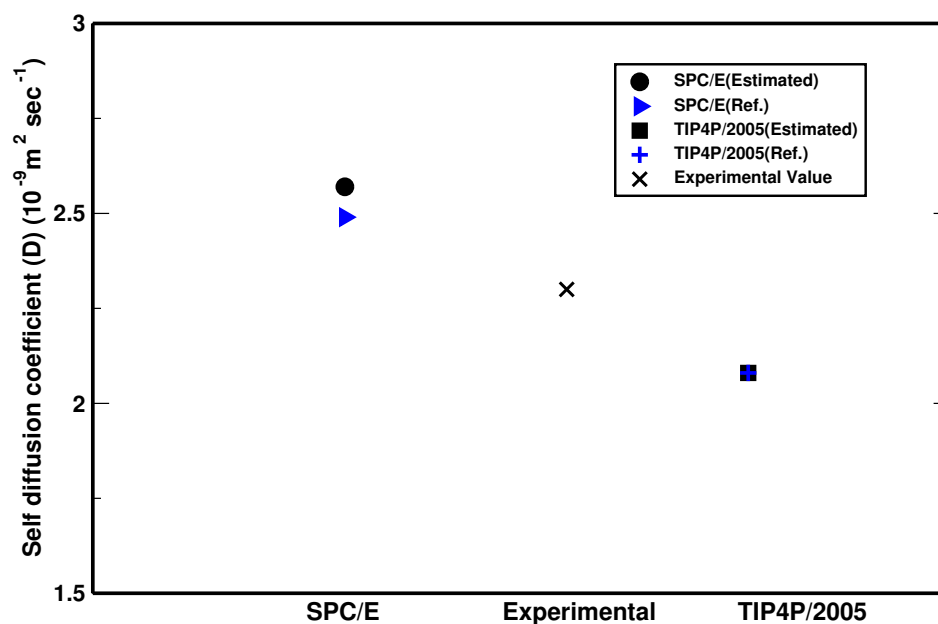
Figure 41: Estimated value of self diffusion coefficients of two different water models: SPC/E and TIP4P/2005 with previously reported simulated as well experimental values at 298.2 K.

Besides the study about effect of system size on diffusion coefficient using a system of binary mixture of glycine and water, we also used the system of amoxicillin in water to understand size effect. In order to understand the effect of system size on diffusion coefficient as well as to estimate size independent value of diffusion coefficient (i.e., diffusion coefficient estimated using simulation box of infinite size), we considered other two systems of different sizes: System-I(a) (2 amoxicillin molecules in 4071 water molecules in cubic box of size 4.98 nm) and System-I(b) (2 amoxicillin molecules in 6504 water molecules in cubic box of size 5.82 nm). Simulations were carried out at 298 K temperature under PBC. During simulations of the above two systems, same parameters in MDP file were used as for System-I. With above two systems, at first, we carried out energy minimization runs. After energy minimization, we performed equilibration run followed by production run for each system. Equilibration and production runs were propagated for 200 ns with time step of 0.002 ps. After production runs, MSD versus time graph were potted for both solute as well as solvent. Figures 42 and 43 are the MSD versus time graph for water and amoxicillin respectively at 298 K temperature taking three systems of different sizes.
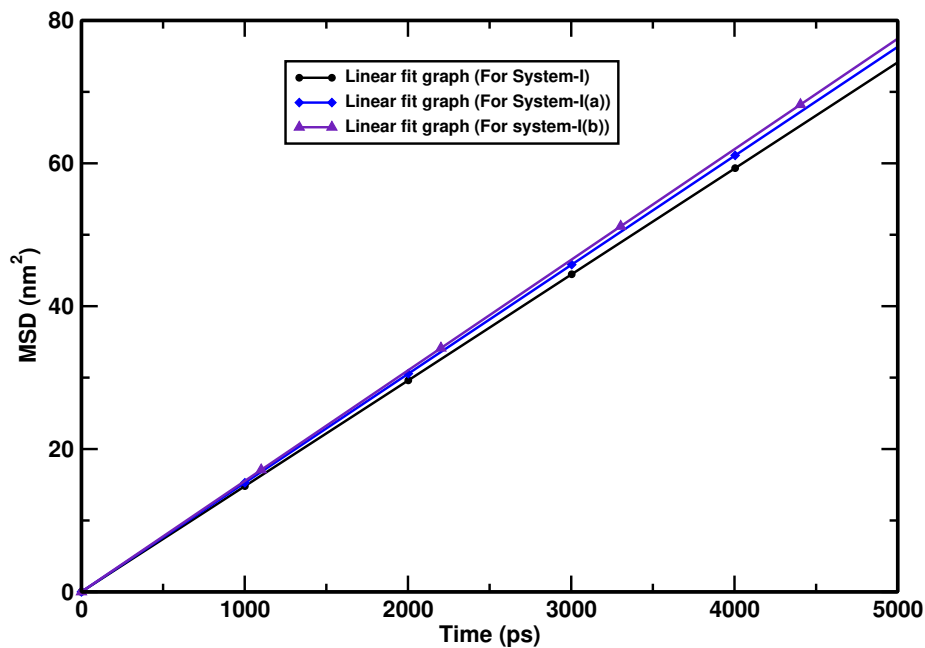
Figure 42: MSD versus time graph for water plotted taking three systems of different size: System-I, System-I(a) and System-I(b) at 298 K temperature.
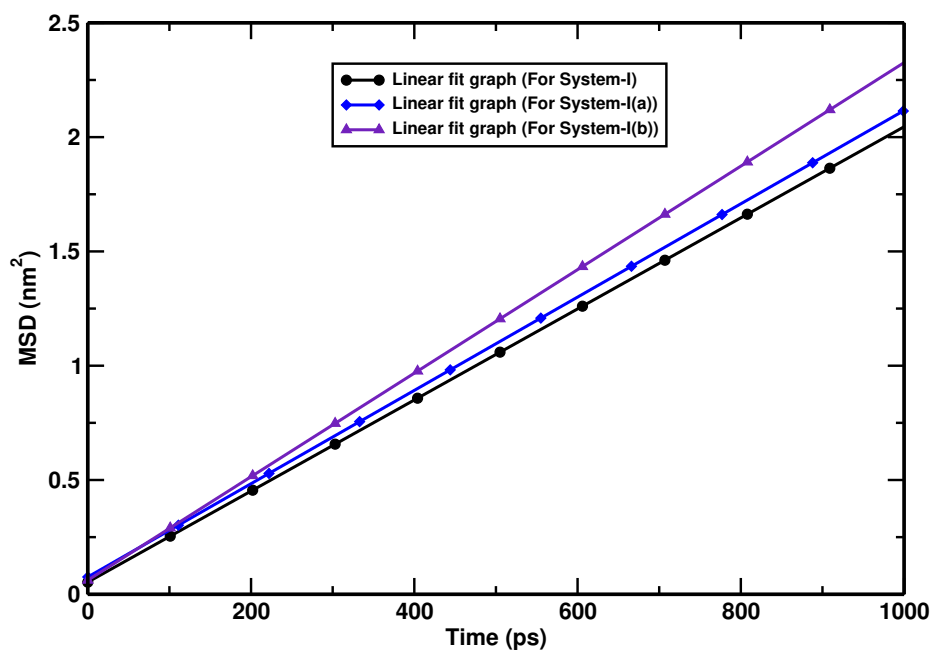
Figure 43: MSD versus time graph for amoxicillin plotted taking three systems of different size: System-I, System-I(a) and System-I(b) at 298 K temperature.

From the Figures 42 and 43, it is observed that the slope the MSD versus time graph increases with the increase in size of system used during our simulations. This clearly indicates that the diffusion coefficient also depends upon the size of system; and it increases with the increase in size of system. We have also estimated the size dependent values of diffusion coefficients using Einstein's relation taking three different size of

simulation boxes; and compared with previously reported experimental values of water (Table 13) (Khanal & Adhikari, 2022).

Table 13: Estimated values of self diffusion coefficients of amoxicillin and water (SPC/E) as well as their mutual diffusion coefficient along with experimental values of self diffusion coefficient of water at 298 K.

System	L (nm)	Diffusion coefficients (D_{PBC})($\times 10^{-9}$ m^2 s^{-1})					Mutual
			Self				
		For amoxicillin	For water				
		Simulated	Simulated	Expt.	% Error		
I	4.05	0.30 ± 0.05	2.49 ± 0.02		7.63		0.30
I(a)	4.98	0.34 ± 0.04	2.54 ± 0.01	2.30 (Holz et al., 2000)	10.43		0.34
I(b)	5.82	0.38 ± 0.02	2.58 ± 0.01		12.17		0.38

From the Table 13, it is noticed that the diffusion coefficient depends on simulation box size and it increases with the increase in size of simulation box (L). This happens due to long range nature of hydrodynamics interactions (Dünweg & Kremer, 1993; Yeh & Hummer, 2004). This indicates that some corrections are necessary to the estimated value of diffusion coefficients from simulations taking finite size box under PBC.

After analyzing the effect of system size on diffusion coefficient, we are motivated to estimate the size independent values of diffusion coefficient of solvent and solution of amoxicillin in water. The size independent value of diffusion coefficient can be estimated using Equation (3.10). To estimate the size independent value of diffusion coefficient (D_0), at first, we have plotted graph between D_{PBC} versus ($1/L$), where D_{PBC} represents the diffusion coefficient estimated from simulation in box of finite size under PBC. From the intercept of graph, we can estimate the size independent value of diffusion coefficient.

For this, we have plotted graphs between D_{PBC} versus ($1/L$) for water and solution of amoxicillin in water at 298 K temperature; and Figures 44 and 45 represent the graph for water and solution of amoxicillin in water respectively. Also, from the Figures, it is clearly observed that the diffusion coefficients estimated from simulations under PBC increase with the increase in size of the system under study. Moreover, the estimated size independent values of diffusion coefficient D_0 of water and solution of amoxicillin in water (i.e., mutual diffusion coefficient) estimated from the intercept of the Figures 44 and 45 are presented in Table 14 (Khanal & Adhikari, 2022).
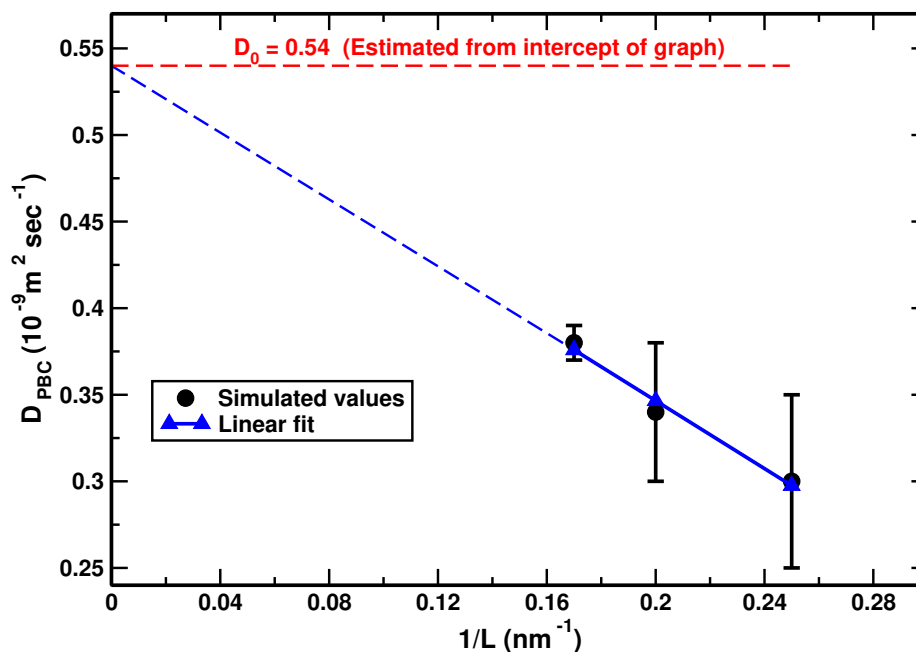
Figure 44: D_{PBC} versus $1/L$ plot for water at 298 K temperature.
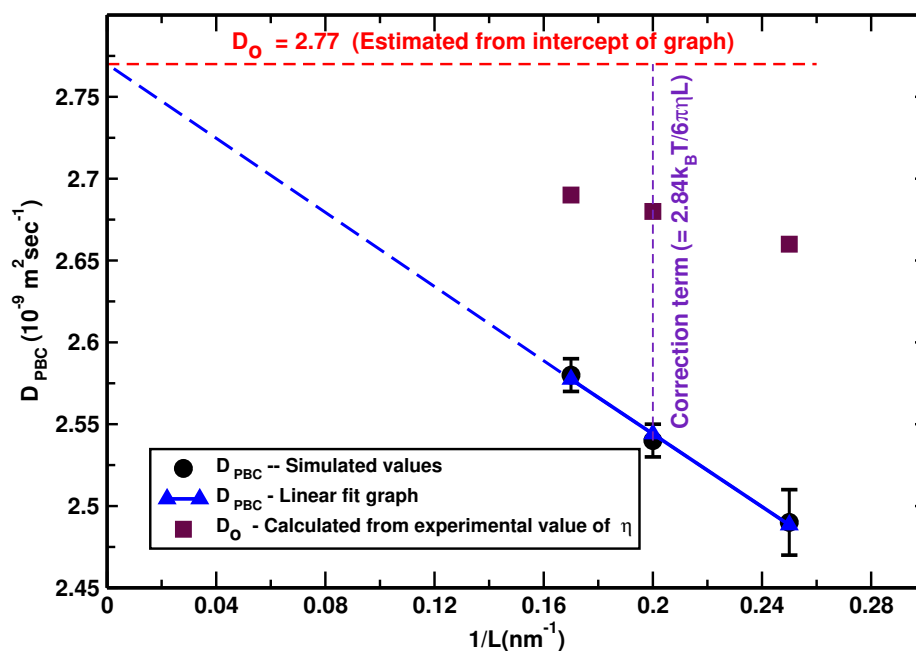
Figure 45: D_{PBC} versus $1/L$ plot for solution of amoxicillin in water at 298 K temperature.

Table 14: System size independent values of diffusion coefficient (D_0) of water and solution of amoxicillin in water at 298 K temperature taking SPC/E water model.

System	D_0 ($\times 10^{-9}$ m^2 s^{-1})			% Error
	Calculated	Ref.	Expt.	
Solvent (water)	2.77	2.97 (Tazi et al., 2012)	2.30 (Holz et al., 2000)	20.4 (Harris & Woolf, 2004)
Solution (amoxicillin in water)	0.54	–	–	–

Table 14 shows that the calculated value of D_0 for water is less than the previously reported value obtained from simulation (Tazi et al., 2012). However, the calculated

value is higher than the experimental value of self diffusion coefficient of water; and the discrepancy between them is 20.4% (Holz et al., 2000). This is because we have taken SPC/E water model as solvent during simulation and such model overestimates diffusion coefficient (Tazi et al., 2012; Markesteijn et al., 2012). Furthermore, the viscosity coefficient (η) of solvent as well as solution can be determined from the slope of D_{PBC} versus ($1/L$) graph (Jamali et al., 2018; Koirala et al., 2020). Here, the viscosity coefficient of both solvent as well as solution of amoxicillin in water has been estimated from the slopes of above two graphs, i.e., Figures 44 and 45. And, the estimated values of (η) for solvent, i.e., water and solution of amoxicillin in water are introduced in Table 15.

Table 15: System size independent value of diffusion coefficient (D_0) and shear viscosity (η) at 298 K temperature taking SPC/E water model.

System	η ($\times 10^{-4}$ Nm^{-2} sec)			% Error
	Calculated	Ref.	Expt.	
Solvent (water)	5.57	6.40 (Tazi et al., 2012)	8.90 (Harris & Woolf, 2004)	37.4
Solution (amoxicillin in water)	6.32	–	–	–

It is found that the calculated value of η for water is smaller than both previously reported experimental as well as simulated values. The value of η is smaller than previously reported values because we used SPC/E model as solvent which underestimates viscosity coefficient (Tazi et al., 2012; Koirala et al., 2020). We further calculate the correction terms on diffusion coefficient due to size effect and the estimated values are present in Table 16 (Khanal & Adhikari, 2022).

Table 16: Estimated values of correction term on diffusion coefficient (D_{PBC}) determined from intercept of graphs 44 and 45; and also compared with corresponding value determined from experimental value of η of water at 298 K temperature taking SPC/E water model.

System	L (nm)	Correction term calculated from($\times 10^{-9}$ m^2 s^{-1})	
		Intercept of graph between D_{PBC} versus $1/L$	$\frac{2.84\ k_{\text{B}}T}{6\pi\eta^{\text{Expt.}}L}$ (Using experimental value of $\eta^{\text{Expt.}}$)
Water	4.05	0.28	0.17
	4.98	0.23	0.14
	5.82	0.19	0.11
Solution	4.05	0.24	–
	4.98	0.20	–
	5.82	0.16	–

From Table 16, we clearly find that the correction term on diffusion coefficient decreases with increase in size of simulation box as expected. Also, the correction terms on diffusion coefficient of water determined from the graphs 44 and 45 are greater than the corresponding values evaluated (from experimental value of η for water) using $\frac{2.84\ k_{\text{B}}T}{6\pi\eta^{\text{Expt.}}L}$. The reason behind this observation is that we used SPC/E water model during the simulations which underestimate viscosity coefficient.

We also analyzed the effect of coupling time on results obtain from simulation. We used Nose-Hoover thermostat instead of velocity-rescaling thermostat taking three different coupling time of 0.5 ps, 1 ps and 2 ps at 298.2 K temperature by considering another System of GABA in water i.e., system consisting of binary mixture of 7 GABA and 2415 SPC/E water molecules. OPLS-AA force field parameters were used to model GABA. For the system of GABA in water, estimated values of diffusion coefficients are reported in Table 17.

Table 17: Estimated values of diffusion coefficients taking system-III at 298.2 K temperature using Nose-Hoover thermostat.

Coupling time	Diffusion coefficients (D_{PBC})(10^{-10} m^2 s^{-1})						
	Self				Binary		
	GABA	Water					
	MSD	MSD	Exp.	% Error	Calculated	Exp.	% Error
0.5 ps	9.32	25.11		9.2	9.37		11.8
1 ps	9.49	25.29	22.99 (Holz et al., 2000)	10.0	9.54	8.38 (Yui et al., 2013)	13.8
2 ps	8.90	25.28		10.0	8.95		6.8

From the Table 17, it has been found that no significant effect on the estimated values of diffusion coefficients when Nose-Hoover thermostat was used instead of velocity-rescaling thermostat with different values of coupling time.

4.4 Free Energy of Solvation

The knowledge of thermodynamic properties, specifically estimation of free energy difference between two thermodynamic states, of drugs, amino acids, proteins and so on in aqueous medium has significant role to understand many process including working mechanism of drugs, protein ligand binding and protein folding (Levy & Onuchic, 2006). The estimation of solvation free energy using different approaches including molecular dynamics play vital role in different discipline of science including pharmaceutical industry (Abel et al., 2017; Khanal et al., 2021b). From the free energy of solvation, we can also estimate the solubility, a major factor for affecting oral drugs absorption (Matos et al., 2017). The estimation of free energy of solvation of a solute is also affected by solvent environment. Similarly, the solvation process also depends upon solvent environments. In this regard, we were inspired to estimate the free energy of solvation of amoxicillin, drug molecule, in two different solvent environments: water and ethanol using molecular dynamics simulation technique. In this section, we have discussed about the estimation of free energy of solvation of amoxicillin in two solvents: water and ethanol at 310 K temperature using TI and FEP based methods. During estimation of solvation free energy of amoxicillin in water, we used two different water models: TIP3P and SPC/E.

The free energy of solvation is the free energy difference between two thermodynamic

states: coupling state (i.e., solute and solvent molecules are coupled through different interactions); and decoupling state (i.e., no interactions between solute and solvent which means solute molecule is in gaseous state). With only two states, there may be a problem of convergence if the difference in potential energy between the states is high. Such a problem of convergence can be handled by introducing many non-physical intermediate states. To introduce intermediate states, we can defined a parameter, i.e., coupling parameter (λ) such that potential energy is also a function of λ and the potential is defined as

$$U(r;\lambda) = (1 - \lambda)\ U_A + \lambda\ U_B \tag{4.2}$$

where, U_A and U_B represent the potential energy of initial state A and final state B respectively. We considered the concept of decoupling to estimate free energy difference which means the solute and solvent which initially in the state of coupling are gradually transformed into the state of decoupling through different intermediate states. In order to couple the solute and solvent molecules, only non-bonded LJ and Coulomb interactions were manipulated without modeling bonded interactions; and the non-bonded interactions were controlled through coupling parameter (λ) (Shivakumar et al., 2010; Khanal et al., 2021b).

To introduce intermediate states, we defined 21 evenly spaced values of coupling parameters for both LJ and Coulomb interactions in case of TIP3P water and ethanol as solvents (Fiorentini et al., 2020). While defining the number of intermediate states, we considered the fact that there must be sufficient overlapping in configurational space between two neighboring states. The 21 different values of coupling states for LJ (λ_{vdW}) and Coulomb ($\lambda_{Coulomb}$) interactions are:

Coupling states: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

λ_{vdW} = 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00,0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90 and 1.00; and

$\lambda_{Coulomb}$ = 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00 and 1.00.

However, we used 0.75 instead of 0.7 in case of SPC/E water model as solvent. Here, the states represented by $\lambda = 0$, 1 and other intermediate values represent the fully interacting solute and solvent molecules, independent and different strength of interaction between solute and solvent molecules respectively. During calculation of solvation free energy with potential defined by Equation (4.2), there is a problem of singularity for small value of r. In order to keep away from such problem introduced by decoupling, LJ

interaction between solute-solvent molecules were modified by soft core potential of the type (Zacharias et al., 1994; Paluch et al., 2011)

$$U_{\text{LJ}}^{\text{SC}}(r_{\text{ij}}, \lambda) = 4\epsilon_{\text{ij}}\lambda^{\text{n}} \left(\frac{1}{\left[\alpha(1-\lambda)^m + \left(r_{\text{ij}}/\sigma_{\text{ij}}\right)^6\right]^2} - \frac{1}{\alpha(1-\lambda)^m + \left(r_{\text{ij}}/\sigma_{\text{ij}}\right)^6} \right) \quad (4.3)$$

where ϵ_{ij} and σ_{ij} are the well depth and length scale parameter for LJ potential respectively; and α is positive constant with typical value of 0.5. Although different values of m and n can be considered, 1 is more effective value of both exponents m and n.

For each 21 different values of coupling parameter, simulations were performed at temperature 310 K and 1 atm pressure under PBC using GROMACS software package. All three systems (System-III(a), System-III(b) and System-III(c)) were treated with energy minimization to overcome bad contact between particles, if any, using Steepest-descent method (David et al., 2005; Pokharel, Khanal, et al., 2019). After attaining the minimum potential energy state with energy minimization, we propagated each system for each value of coupling parameter with equilibration run to attain the state of thermodynamic equilibrium at 310 K temperature under PBC. Each system was equilibrated in two steps: in NVT ensemble at first followed by NPT ensemble. Finally, production runs with each above mentioned equilibrated systems were carried out in NVT ensemble at 310 K under PBC. Time step of 2 fs was chosen for each run. During each run, electrostatic interactions were accounted with PME of order 6; and 1.2 nm cut-off was used for both short range LJ as well as Coulomb interactions (Darden et al., 1993). Temperature was controlled using Berendsen barostat with 0.5 ps of coupling time (Berendsen et al., 1984). Isothermal compressibility 4.5×10^{-5} bar was chosen during each simulation. Also, the value of soft core potential parameter α and σ were set to 0.5 and 0.3 respectively; and 1 was selected for both m and n for soft core potential defined by Equation (20). Bonds were constrained using LINCS algorithms (Hess et al., 1997). Furthermore, Maxwell-Boltzmann distribution was chosen to assign the initial velocities of each particle during equilibration in NVT ensemble. Langevin dynamics was used to solve the equations of motion (Lindahl et al., 2010). We performed 10 ns simulation for each equilibration and productions runs.

After production runs for each value of coupling parameter, we used alchemical-analysis.pl tool to extract free energy difference from *.xvg files (Klimovich et al., 2015; alchemical analysis, 2020). Although different approaches can be used to estimate solvation free energy, TI and FEP methods have been used in our calculations. From the ensemble average of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle$, we can estimate free energy difference in TI approach. During our calculations, two TI based methods: TI-1 and TI-CUBIC have been implemented. TI-1 and TI-3 approaches use different integration methods, i.e.,

the trapezoidal rule (a first-order polynomial) and a cubic spline are used in TI-1 and TI-CUBIC respectively.

As we know in TI based methods, the free energy difference along the integrating path described by coupling parameter is evaluated from the ensemble average $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$. So, we first plot the graph between $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ versus λ taking each solvent. Figures 46, 47 and 48 are graph between $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ versus λ for solvents: TIP3P water, SPC/E water and ethanol respectively (Khanal & Adhikari, 2022).

Figure 46: Graph between $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ versus λ taking TIP3P water model as solvent at 310 K temperature.

Figure 47: Graph between $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda}$ versus λ taking SPC/E water model as solvent at 310 K temperature.
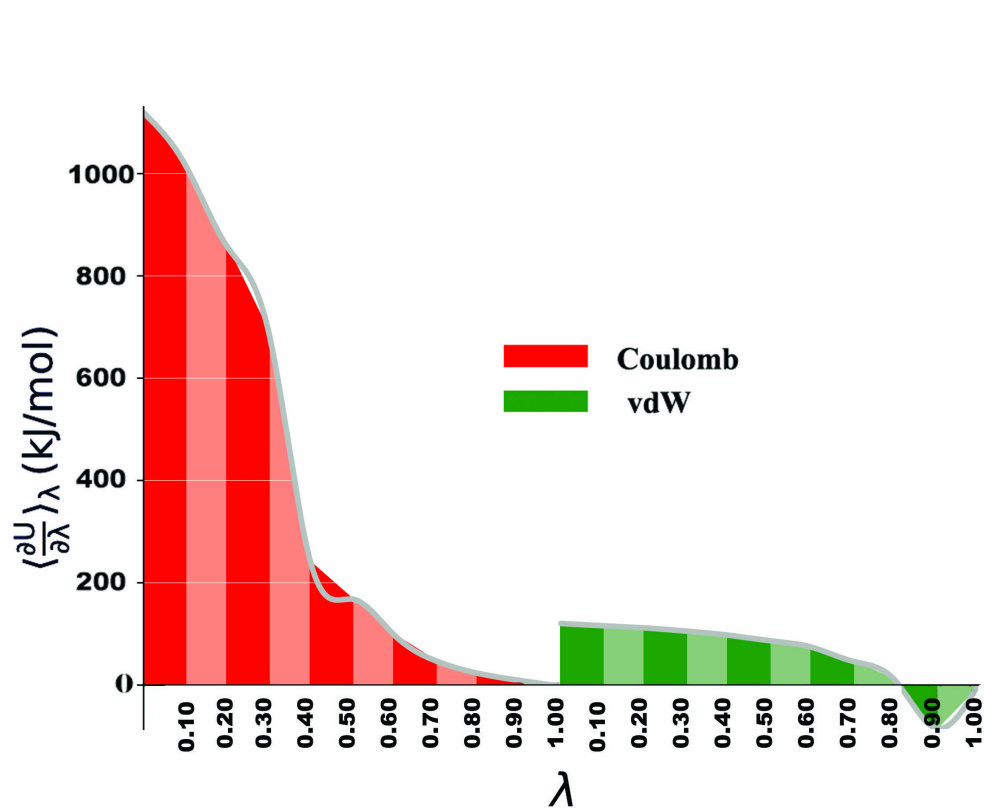
Figure 48: Graph between $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda}$ versus λ taking ethanol as solvent at 310 K temperature.

In the plots, red and green colors indicate the individual contribution of Coulomb and vdW interactions respectively (color online). Also, the estimation of free energy difference using Trapezoidal rule is represented by shaded area and silver colored curve represents the estimation using Cubic spline method.

Similarly, out of different perturbation based methods, BAR and MBAR are used. In the BAR method, the minimum free energy variance is obtained from the sampling in both direction. In BAR method, convergence can be obtained with small phase space overlapping between two thermodynamics states in comparision of other techniques (Shirts & Pande, 2005a; Shirts & Mobley, 2013). The estimated values of free energy differences between two consecutive states at 310 K temperature using TI-1, TI-CUBIC, BAR and MBAR methods are presented in Tables 18, 19 and 20 respectively.

Table 18: Estimated values of free energy difference between two consecutive states in TIP3P water model as solvent at 310 K using different thermodynamic (TI) and free energy perturbation (FEP) based methods: TI, TI-CUBIC, BAR and MBAR.

Coupling States	ΔG_{sol} in kJ/mol with method			
	TI	TI-CUBIC	BAR	MBAR
0 − 1	137.65 ± 0.06	137.66 ± 0.07	137.79 ± 0.07	137.79 ± 0.07
1 − 2	120.65 ± 0.09	121.26 ± 0.11	120.98 ± 0.10	120.98 ± 0.10
2 − 3	99.48 ± 0.15	99.14 ± 0.17	99.76 ± 0.21	99.77 ± 0.21
3 − 4	77.08 ± 0.21	78.45 ± 0.24	76.64 ± 0.25	76.64 ± 0.25
4 − 5	45.70 ± 0.19	45.06 ± 0.22	43.21 ± 0.83	43.21 ± 1.35
5 − 6	21.11 ± 0.12	18.90 ± 0.14	20.97 ± 0.13	21.07 ± 0.12
6 − 7	14.49 ± 0.09	14.98 ± 0.10	14.54 ± 0.08	14.59 ± 0.07
7 − 8	8.88 ± 0.04	8.64 ± 0.05	8.87 ± 0.03	8.93 ± 0.03
8 − 9	4.38 ± 0.02	4.31 ± 0.02	4.30 ± 0.02	4.41 ± 0.01
9 − 10	1.03 ± 0.01	0.97 ± 0.01	1.01 ± 0.01	1.03 ± 0.01
10 − 11	9.02 ± 0.01	9.03 ± 0.01	9.04 ± 0.01	9.03 ± 0.01
11 − 12	8.61 ± 0.01	8.62 ± 0.01	8.61 ± 0.01	8.64 ± 0.01
12 − 13	8.07 ± 0.01	8.10 ± 0.01	8.11 ± 0.01	8.07 ± 0.01
13 − 14	7.11 ± 0.01	7.17 ± 0.01	7.14 ± 0.01	7.12 ± 0.01
14 − 15	5.41 ± 0.02	5.50 ± 0.02	5.45 ± 0.02	5.45 ± 0.02
15 − 16	3.02 ± 0.02	2.94 ± 0.03	3.15 ± 0.02	3.17 ± 0.02
16 − 17	-0.45 ± 0.04	0.33 ± 0.05	-0.13 ± 0.04	-0.13 ± 0.04
17 − 18	-9.61 ± 0.09	-9.80 ± 0.10	-8.97 ± 0.14	-8.98 ± 0.14
18 − 19	-15.52 ± 0.09	-17.13 ± 0.10	-18.61 ± 0.09	-18.63 ± 0.09
19 − 20	-7.19 ± 0.02	-7.67 ± 0.03	-6.98 ± 0.02	-6.97 ± 0.02

Table 19: Estimated values of free energy difference between two consecutive states in SPC/E water model as solvent at 310 K using different thermodynamic (TI) and free energy perturbation (FEP) based methods: TI, TI-CUBIC, BAR and MBAR.

Coupling	ΔG_{sol} in kJ/mol with method			
States	TI	TI-CUBIC	BAR	MBAR
0 − 1	143.80 ± 0.10	143.82 ± 0.10	144.05 ± 0.11	144.05 ± 0.11
1 − 2	125.63 ± 0.10	126.28 ± 0.12	125.84 ± 0.12	125.85 ± 0.12
2 − 3	103.22 ± 0.18	102.71 ± 0.21	103.40 ± 0.25	103.41 ± 0.25
3 − 4	80.74 ± 0.19	82.15 ± 0.22	80.29 ± 0.22	80.29 ± 0.22
4 − 5	49.09 ± 0.19	48.54 ± 0.21	48.76 ± 1.01	48.78 ± 1.76
5 − 6	23.57 ± 0.18	21.29 ± 0.20	23.60 ± 0.16	23.64 ± 0.16
6 − 7	16.32 ± 0.09	16.87 ± 0.10	16.30 ± 0.08	16.38 ± 0.08
7 − 8	9.78 ± 0.04	9.50 ± 0.05	9.72 ± 0.04	9.76 ± 0.04
8 − 9	4.61 ± 0.02	4.50 ± 0.03	4.47 ± 0.02	4.60 ± 0.02
9 − 10	1.02 ± 0.01	0.93 ± 0.02	0.99 ± 0.01	1.02 ± 0.01
10 − 11	9.35 ± 0.01	9.35 ± 0.01	9.34 ± 0.01	9.35 ± 0.01
11 − 12	8.88 ± 0.01	8.90 ± 0.01	8.91 ± 0.01	8.90 ± 0.01
12 − 13	8.25 ± 0.01	8.26 ± 0.01	8.27 ± 0.01	8.32 ± 0.01
13 − 14	7.36 ± 0.02	7.40 ± 0.02	7.45 ± 0.02	7.48 ± 0.01
14 − 15	6.00 ± 0.02	6.07 ± 0.02	6.05 ± 0.02	6.07 ± 0.02
15 − 16	3.77 ± 0.03	3.88 ± 0.04	3.93 ± 0.03	3.93 ± 0.03
16 − 17	-7.47 ± 0.14	-4.05 ± 0.23	-4.60 ± 0.23	-4.62 ± 0.23
17 − 18	-8.74 ± 0.07	-8.94 ± 0.07	-8.87 ± 0.08	-8.89 ± 0.08
18 − 19	-18.88 ± 0.10	-21.25 ± 0.14	-21.37 ± 0.09	-21.29 ± 0.09
19 − 20	-7.60 ± 0.02	-7.45 ± 0.04	-7.28 ± 0.02	-7.27 ± 0.02

Table 20: Estimated values of free energy difference between two consecutive states in ethanol as solvent at 310 K using different thermodynamic (TI) and free energy perturbation (FEP) based methods: TI, TI-CUBIC, BAR and MBAR.

Coupling States	ΔG_{sol} in kJ/mol with method			
	TI	TI-CUBIC	BAR	MBAR
$0-1$	106.79 ± 0.18	107.34 ± 0.22	106.87 ± 0.19	106.85 ± 0.19
$1-2$	93.71 ± 0.21	93.48 ± 0.24	93.84 ± 0.27	93.84 ± 0.27
$2-3$	77.32 ± 0.24	79.31 ± 0.27	77.69 ± 0.29	77.70 ± 0.29
$3-4$	46.44 ± 0.22	45.94 ± 0.26	43.46 ± 1.79	43.47 ± 2.11
$4-5$	20.40 ± 0.18	18.02 ± 0.21	20.47 ± 0.19	20.52 ± 0.18
$5-6$	13.32 ± 0.18	13.88 ± 0.21	13.22 ± 0.19	13.28 ± 0.17
$6-7$	7.28 ± 0.14	6.88 ± 0.17	7.06 ± 0.14	7.22 ± 0.11
$7-8$	3.52 ± 0.07	3.41 ± 0.08	3.44 ± 0.06	3.73 ± 0.04
$8-9$	1.61 ± 0.03	1.56 ± 0.03	1.60 ± 0.03	1.89 ± 0.01
$9-10$	0.33 ± 0.02	0.33 ± 0.02	0.40 ± 0.02	0.42 ± 0.01
$10-11$	11.81 ± 0.02	11.81 ± 0.02	11.84 ± 0.01	11.83 ± 0.01
$11-12$	11.39 ± 0.01	11.40 ± 0.01	11.38 ± 0.01	11.38 ± 0.01
$12-13$	10.87 ± 0.01	10.88 ± 0.01	10.87 ± 0.01	10.88 ± 0.01
$13-14$	10.16 ± 0.01	10.19 ± 0.02	10.25 ± 0.01	10.21 ± 0.01
$14-15$	9.22 ± 0.02	9.22 ± 0.02	9.14 ± 0.02	9.17 ± 0.02
$15-16$	8.05 ± 0.02	8.16 ± 0.03	8.07 ± 0.02	8.00 ± 0.02
$16-17$	6.03 ± 0.04	6.01 ± 0.04	6.16 ± 0.04	6.18 ± 0.04
$17-18$	2.81 ± 0.07	3.38 ± 0.08	3.18 ± 0.07	3.19 ± 0.07
$18-19$	-3.74 ± 0.11	-4.31 ± 0.12	-3.81 ± 0.16	-3.82 ± 0.16
$19-20$	-4.27 ± 0.09	-5.56 ± 0.12	-5.67 ± 0.08	-5.67 ± 0.08

From the Tables 18, 19 and 20, it is seen that the free energy difference between two consecutive coupling states is high at the beginning and decreases with increase in coupling state. This happened because in the beginning, we consider both interactions in full strength and slowly turned off the interactions such that we first turned of Coulomb interaction; after then van der Waals interaction. The total free energy difference between initial and final states gives the solvation free energy. The estimated values of total free energy difference between initial and final states, i.e., free energy of solvation of amoxicillin in two solvents: water (TIP3P and SPC/E models) and ethanol using TI, TI-CUBIC, BAR and MBAR methods are presented in Table 21. During the estimation of free energy of solvation, we also analyzed the individual contributions of van der Waals (vdW) and Coulomb interactions to the total solvation free energy in each solvents: and the individual contributions are also presented in the Table 21 (Khanal & Adhikari, 2022).

Table 21: Estimated values of free energy of solvation of amoxicillin (ΔG_{sol}) in water (TIP3P and SPC/E models) and ethanol at 310 K using TI, TI-CUBIC, BAR and MBAR methods taking individual contribution of vdW and Coulomb along with total contributions due to both the interactions.

Solvent	Model	Method	ΔG_{sol} in kJ/mol with		
			vdW only	Coulomb only	Total
Water	TIP3P	TI	8.47 ± 0.20	530.47 ± 0.53	538.94 ± 0.57
		TI-CUBIC	7.09 ± 0.19	529.38 ± 0.53	536.46 ± 0.57
		BAR	6.82 ± 0.17	528.07 ± 0.91	534.90 ± 0.93
		MBAR	6.77 ± 0.20	528.43 ± 1.43	535.20 ± 1.44
	SPC/E	TI	0.93 ± 0.25	557.78 ± 0.57	558.71 ± 0.62
		TI-CUBIC	2.16 ± 0.26	556.60 ± 0.57	558.76 ± 0.63
		BAR	1.81 ± 0.26	557.42 ± 1.09	559.24 ± 1.12
		MBAR	1.97 ± 0.29	557.76 ± 1.84	559.73 ± 1.86
Ethanol	–	TI	62.35 ± 0.23	370.73 ± 0.74	433.08 ± 0.77
		TI-CUBIC	61.18 ± 0.25	370.15 ± 0.75	431.33 ± 0.79
		BAR	61.40 ± 0.20	368.05 ± 1.87	429.45 ± 1.88
		MBAR	61.35 ± 0.23	368.92 ± 2.22	430.26 ± 2.23

During the estimation of solvation free energy, our initial thermodynamic state represents full interaction between solute and solvent molecules through vdW and Coulomb interactions and final state means there is no interaction between them. So, the positive value of solvation free energy in particular solvent ensures that solvation process occurs in this solvent. From the Table 21, it is seen that the free energy of solvation of amoxicillin in both solvents: water and ethanol has positive values when contributions of both van der Waals and coulomb (vdW-q) interactions has been taken into account. The positive values of free energy of solvation ensures that the solvation of amoxicillin occurs in both solvents. Also, it is observed from the Table 21 that the individual contribution of vdW as well as Coulomb to solvation free energy of amoxicillin in both solvents has positive values which indicates each individual interaction can contribute to the solvation of amoxicillin in water. Although both interactions contribute to the solvation of amoxicillin in both the solvents, it is seen that Coulomb interaction has major contributions for the solvation of amoxicillin in both the solvents: water and ethanol. The estimated values of free energy of solvation using different TI and FEP based methods are in close agreement.

Furthermore, due to unavailability of previously reported value for solvation free energy of amoxicillin, zwitterionic glutamic acid has taken as reference for comparison. The previously reported data of electrostatic contribution for solvation free energy of zwitterionic glutamic acid in aqueous medium is 471.70 kJ/mol and our estimated value of contribution of Coulomb interaction for solvation free energy of amoxicillin in TIP3P water is ≈ 530 kJ/mol (Dixit et al., 1997). Our estimated value of contribution of Coulomb interaction to solvation free energy of amoxicillin in water is in good agreement with previously reported value of electrostatic contribution to solvation free energy

of glutamic acid in zwitterionic form in aqueous medium.

Moreover, we have also estimated the Gibb's free energy of solvation of amoxicillin in water using Born equation, continuum solvation method (Born, 1920);

$$\Delta G = -\frac{N_A Z^2 e^2}{8\pi\epsilon_0 r_0}\left(1 - \frac{1}{\epsilon_r}\right)$$

In this equation, N_A, Ze, ϵ_0, r_0 and ϵ_r represent Avogadro number, charges, permittivity of free space, effective radius of ion and dielectric constant of solvent respectively; and we take $r_0 = 0.5$ nm. Our estimated value using the Born equation is 547.31 kJ/mol at 35 °C which closely agree with our estimated values from simulations.

The free energy estimation from the simulations must be converged. In our free energy calculation, the convergence of calculation has been analyzed by plotting time series plot, i.e., graph between estimated value of solvation free energy as a function of simulation time. Figures 49, 50 and 51 show the time series plots during estimation of solvation free energy of amoxicillin in water (TIP3P and SPC/E models) and ethanol respectively.
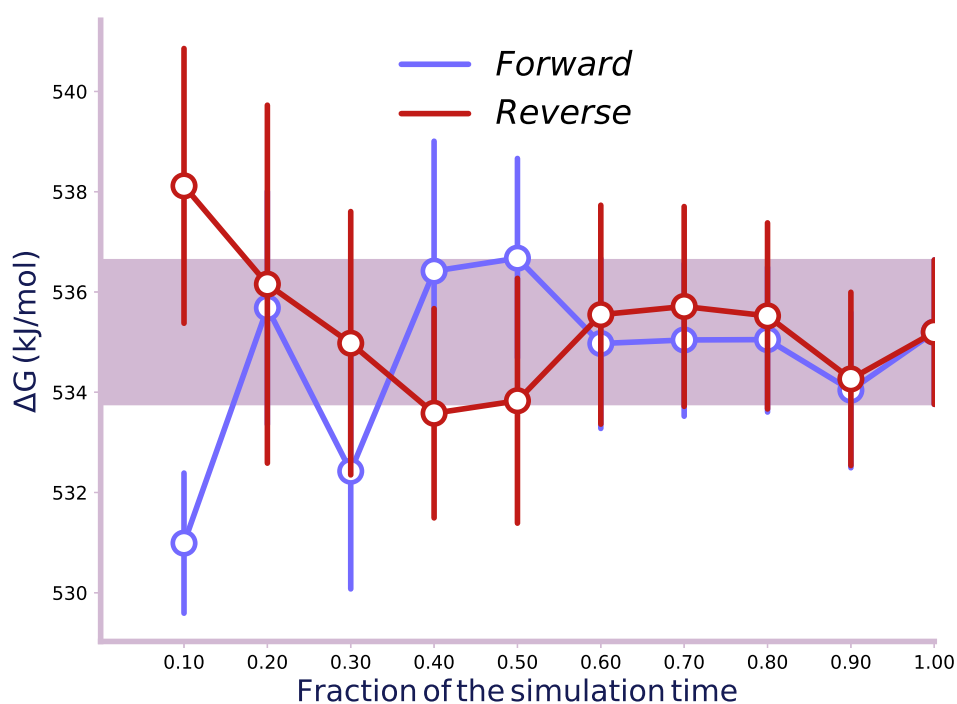
Figure 49: Variation of estimated value of solvation free energy of amoxicillin in TIP3P water as a function of simulation time, i.e., time series plot in both forward and reverse directions.
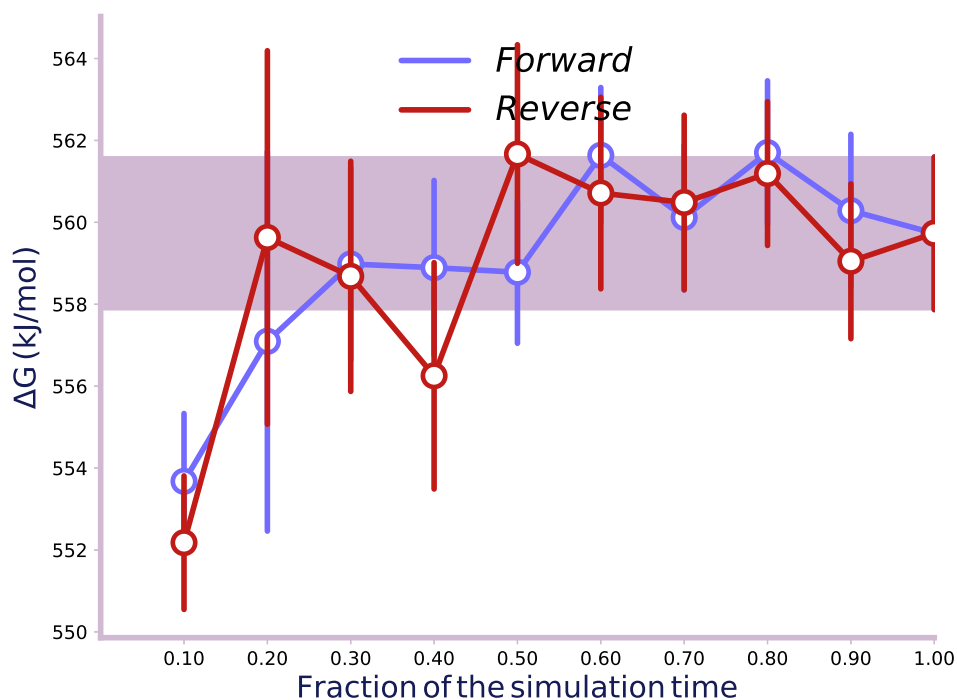
Figure 50: Variation of estimated value of solvation free energy of amoxicillin in SPC/E water as a function of simulation time i.e. time series plot in both forward and reverse directions.
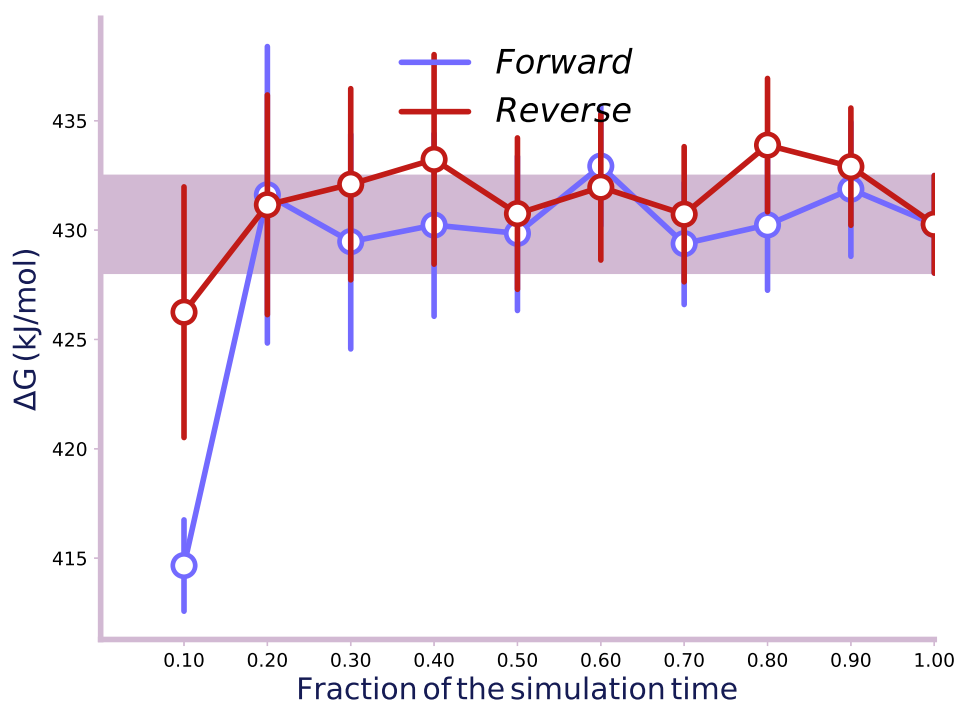
Figure 51: Variation of estimated value of solvation free energy of amoxicillin in ethanol as a function of simulation time i.e. time series plot in both forward and reverse directions.

From the above plots 49, 50 and 51, it is seen that convergence of our free energy calculation occur after 0.4 fraction of simulation time. This observation is in accordance with guidelines for the free energy calculations by Klimovich and co-workers (Klimovich

et al., 2015). From this observation, we conclude that our calculation of free energy follow convergence.

Solvent Accessible Surface Srea (SASA) and Hydrogen Bond Analysis

We have also analyzed the effect of solvent environment on solvation free energy through the concept of Solvent Accessible Surface Area (SASA) of solute and number of hydrogen bonds between solute and solvent molecules. The SASA measures the surface area of solute accessible to interaction with solvents; and higher value of SASA indicates more area is available to solvent for interactions. Figure 52 is the SASA of solute molecules, i.e., amoxicillin in two solvent environments: water (TIP3P and SPC/E models) and ethanol at 310 K for initial ($\lambda = 0$) state (Khanal & Adhikari, 2022).
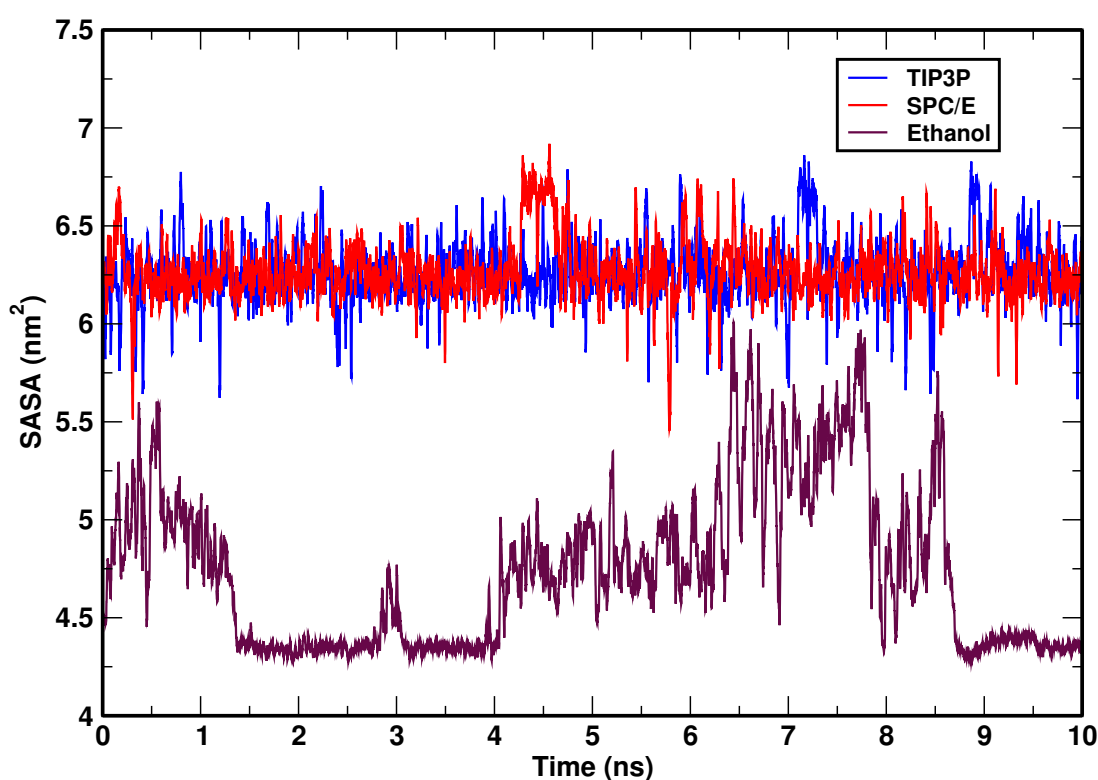
Figure 52: Time evolution of solvent accessible surface area (SASA) of solute molecule at 310 K temperature for initial state represented by $\lambda = 0$ which indicates solute and solvent molecules are fully interacting through vdW and Coulomb interactions in two different solvent environments: water (TIP3P and SPC/E models) and ethanol.

From the Figure 52, it is clearly seen that SASA remains almost constant with simulation time for water as solvent. Also, we observe that SASA of solute to ethanol has lower value in comparison to water by ≈ 1 nm^2. The lower value of SASA in ethanol indicates that the lower affinity of dissolving of amoxicillin in ethanol than in water. Such results also support to the previously reported experimental observation that amoxicillin has smaller solubility in ethanol than in water. This is also in agreement with our observation that lower value of solvation free energy of amoxicillin in ethanol in comparison of water.

Similarly, the hydrogen bond analysis between solute and solvent molecules provides more insight on effect of solvent environment. The hydrogen bonds play a vital role in formation of intra- and inter-molecular structure. Figures 53 and 54 are the time evolution and average number of hydrogen bonds between solute (amoxicillin) and solvents (TIP3P and SPC/E water models) respectively at 310 K temperature for initial state $\lambda = 0$ (Khanal & Adhikari, 2022). During hydrogen bond analysis, we used cut off parameters 0.35 nm and 30° for length and angle respectively. From the plots, we clearly notice that number of hydrogen bonds remains almost constant with time of simulations and also more number of the bonds observe in water than in ethanol. Also, the solute in water has higher number of hydrogen bonds than in ethanol (almost 7). Our calculations of number of hydrogen bonds in both solvents are also in agreement to higher value of solvation free energy in water than in ethanol.
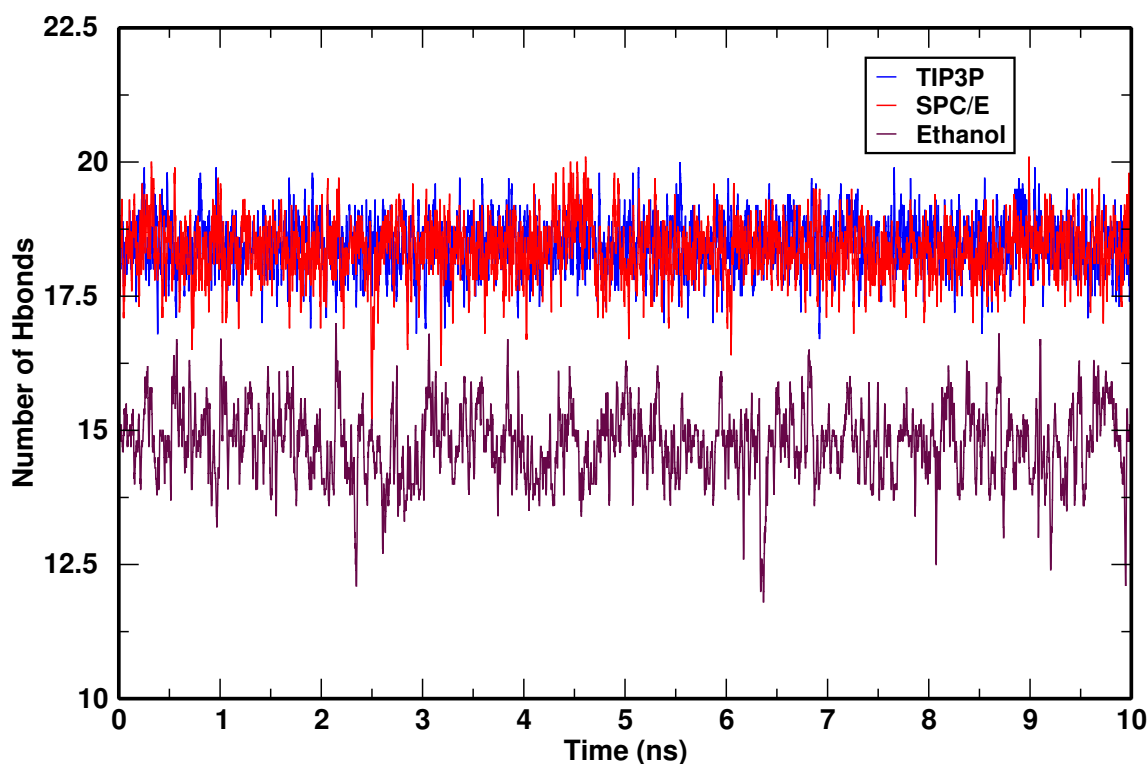
Figure 53: Time evolution of number of hydrogen bonds between solute and solvents (TIP3P water, SPC/E water and ethanol) at 310 K temperature for $\lambda = 0$.
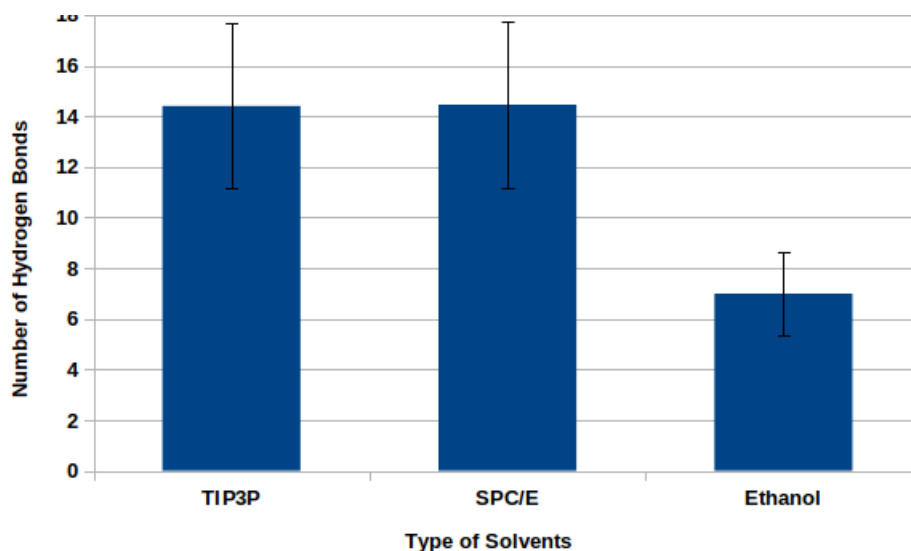
Figure 54: Average number of hydrogen bonds between solute and solvents (TIP3P water, SPC/E water and ethanol) at 310 K temperature for $\lambda = 0$.

From above analysis, it has been found that the values of SASA of solute and number of hydrogen bonds between solute and solvent molecules are large in case water as solvent than ethanol. These findings imply that SASA and number of hydrogen bonds also play supportive role in our observation that higher value of solvation free energy of amoxicillin in water than in ethanol.

CHAPTER 5

5. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

We have performed molecular dynamics study of amoxicillin in water to understand transport and thermodynamic properties like diffusion phenomenon and solvation free energy. Simulations were carried out at 1 atm pressure under periodic boundary conditions using GROMACS software package.

Diffusion is measured in terms of diffusion coefficient. The self diffusion coefficients of solute and solvent as well as their binary diffusion coefficient have been reported at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K from the slope of MSD versus time graph using Einstein's and Darken's relations respectively. The estimated values of self diffusion coefficient of water at different temperatures are in good agreement with experimentally reported values in literature within 8.5 %. It has been observed that both self and binary diffusion coefficients increase with increase in temperature as expected. The effect of temperature on self diffusion coefficients of both solute and solvent has been analyzed by plotting Arrhenius equation; and we observe that the estimated values of self diffusion coefficients follow Arrhenius behavior. The calculated values of activation energy for diffusion of amoxicillin and water are 0.017 kJ/mol and 15.31 kJ/mol respectively. The calculated value of activation energy of water is in agreement within $\approx$12.2 % with previously reported experimental value.

We have also studied the effect of system size on diffusion coefficient; and estimated the size independent value of diffusion coefficient as well as viscosity coefficient of water and solution of amoxicillin in water. In order to study the effect of system size, we extended our simulations taking two different size box at 298 K temperature. We have also reported the correction terms on diffusion coefficient.

Besides transport properties, the solvation free energy of amoxicillin in two solvent environments: water (TIP3P and SPC/E models) and ethanol, has been estimated at 310 K temperature using thermodynamic integration (TI) and free energy perturbation

(FEP) based methods: TI, TI-Cubic, BAR and MBAR. From the estimation of solvation free energy, it has been observed that the amoxicillin is soluble in both solvents. The solvation free energy also depends on solvent environment. The estimated value of solvation free energy of amoxicillin is higher in water than in ethanol as expected. Also, the individual contribution of van der Waals and Coulomb interaction on solvation free energy has been analyzed; and we observe that Coulomb interaction has major contribution on solvation of amoxicillin in both the solvents. We also examined the free energy convergence by plotting time series graphs.

In order to get more insight in molecular level regarding the effect of solvent environment on solvation free energy, the SASA of solute and number of hydrogen bonds between solute and solvent molecules have been reported for initial thermodynamic state. We also compared the SASA and number of hydrogen bonds by taking TIP3P, SPC/E and ethanol as solvents. The reported values of SASA and number of hydrogen bonds are also in agreement to our observation showing higher value of solvation free energy of amoxicillin in water than that in ethanol.

Recommendations

The solubility is one of the fundamental factor for oral drug absorption. The detail idea about solubility of drug in molecular level can play a crucial role in pharmaceutical industry. Solubility of drugs can be estimated from different technique including analytical approaches, activity coefficients models etc. Moreover, solubility can be estimated from the free energy calculation using molecular dynamics approach. The estimation of solubility of drug from molecular dynamics study gives molecular level characteristics during the interaction of targeted molecule with surrounding environment; and also provides guideline for experimental measurement. On the other hand, the knowledge of transport properties provides information about dynamics of drug in our body which reveals the inter molecular interaction. The dissolution rate of drug depends upon many factors including diffusion coefficient, solvation free energy and solubility of the drugs in particular solvent environment.

The solubility of drug can be estimated from the solvation free energy and absolute solid free energy of the drug in crystal. Different methods including Einstein molecule approach can be used to estimate absolute free energy of molecular solid. This work can be extended to estimate the absolute free energy of amoxicillin crystal; and hence the solubility of amoxicillin in different solvents. Viscosity coefficient can be estimated from non-equilibrium dynamics.

The estimation of transport and thermodynamic properties depends upon the solute concentration as well as solvent environments. In this regards, the work can be extended to

estimate the properties by taking different solute concentration and solvent environments can be fruitful. Moreover, other force field parameters can be used to check the validity of results. This study can also be utilized to explore the translocation of drug through the membrane of intestinal wall.

CHAPTER 6

6. SUMMARY

In this work, the transport and thermodynamic properties of amoxicillin, an antibiotic with β-lactam ring responsible for antibacterial activity, have been studied using molecular dynamics simulation technique. As we know that most of biological processes happen in aqueous medium through diffusive transport, the delivery of drugs upto target also happens through diffusive transport. We can get information about the inter/intra molecular interactions from the knowledge of diffusion phenomenon. On the other hand, out of many factors, solubility of drugs is one of the factor affecting oral drug absorption; it can be measured from solvation free energy. Furthermore, the information about diffusion coefficient and solubility is also useful to understand the dissolution rate of drugs.

To understand the transport phenomenon of amoxicillin in aqueous medium, our system which consist of binary mixture of amoxicillin and water was subjected to simulation at 1 atm pressure under periodic boundary conditions at five different temperatures: 298 K, 303 K, 305 K, 310 K and 313 K. We have also studied the effect of temperature on diffusion coefficient; and it has observed that the diffusion coefficient for both solute and solvent follow Arrhenious behavior. Also, to check the size dependency of diffusion coefficient estimated from simulations under periodic boundary condition, simulations have been carried by taking three systems of different size; and it is clearly observed that diffusion coefficient increases with increase in size of simulation box.

Similarly, the solvation free energy of amoxicillin in two different solvents: water (SPC/E and TIP3P models) and ethanol, has been estimated at 310 K temperature using thermodynamic integration and free energy perturbation based methods; and it has been observed that coulomb interaction has major contribution to solvation free energy of amoxicillin in both the solvents.

REFERENCES

Abascal, J. L., & Vega, C. (2005). A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of Chemical Physics*, *123*(23), 234505.

Abel, R., Wang, L., Harder, E. D., Berne, B., & Friesner, R. A. (2017). Advancing drug discovery through enhanced free energy calculations. *Accounts of Chemical Research*, *50*(7), 1625–1632.

Adhikari, N. P., Peng, X., Alizadeh, A., Ganti, S., Nayak, S. K., & Kumar, S. K. (2004). Multiscale modeling of the surfactant mediated synthesis and supramolecular assembly of cobalt nanodots. *Physical Review Letters*, *93*(18), 188301.

alchemical analysis. (2020). https://github.com/mobleylab/alchemical-analysis (accessed: 2020-09-01).

Alder, B. J., & Wainwright, T. (1967). Velocity autocorrelations for hard spheres. *Physical Review Letters*, *18*(23), 988.

Alder, B. J., & Wainwright, T. E. (1957). Phase transition for a hard sphere system. *The Journal of Chemical Physics*, *27*(5), 1208–1209.

Allen, M. P., & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Oxford University Press, New York.

Amidon, G. L., Lennernäs, H., Shah, V. P., & Crison, J. R. (1995). A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharmaceutical Research*, *12*(3), 413–420.

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, *72*(4), 2384–2393.

Aragones, J., Noya, E. G., Valeriani, C., & Vega, C. (2013). Free energy calculations for molecular solids using GROMACS. *The Journal of Chemical Physics*, *139*(3), 034104.

Aragones, J., Valeriani, C., & Vega, C. (2012). Note: Free energy calculations for atomic solids through the Einstein crystal/molecule methodology using GROMACS and LAMMPS. *The Journal of Chemical Physics*, *137*(14), 146101.

Arancibia, A., Guttmann, J., Gonzalez, G., & Gonzalez, C. (1980). Absorption and disposition kinetics of amoxicillin in normal human subjects. *Antimicrobial Agents and Chemotherapy*, *17*(2), 199–202.

Barve, K., & Ruparel, K. (2015). Effect of Bioenhancers on Amoxicillin bioavailability. *ADMET and DMPK*, *3*(1), 45–50.

Bash, P., Singh, U., Langridge, R., & Kollman, P. (1987). Free energy calculations by computer simulation. *Science*, *236*(4801), 564–568.

Baysinger, G. (2015). CRC handbook of chemistry and physics. *National Institute of Standards and Technology*.

Bebu, A., Szabó, L., Leopold, N., Berindean, C., & David, L. (2011). IR, Raman, SERS and DFT study of amoxicillin. *Journal of Molecular Structure*, *993*(1-3), 52–56.

Bekker, H., Berendsen, H. J. C., Dijkstra, E., Achterop, S., Vondrumen, R., van der Spoel, D., ... Renardus, M. (1993). Gromacs-a parallel computer for molecular-dynamics simulations. In *4th International Conference on Computational Physics (PC 92)* (pp. 252–256).

Bellucci, M. A., Gobbo, G., Wijethunga, T. K., Ciccotti, G., & Trout, B. L. (2019). Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments. *The Journal of Chemical Physics*, *150*(9), 094107.

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*(2), 245–268.

Berendsen, H. J. C., Grigera, J., & Straatsma, T. (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry*, *91*(24), 6269–6271.

Berendsen, H. J. C., Postma, J. v., Gunsteren, W. F. V., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690.

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1-3), 43–56.

Bergström, C. A., & Larsson, P. (2018). Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *International Journal of Pharmaceutics*, *540*(1-2), 185–193.

Bhandari, D., & Adhikari, N. (2016). Molecular dynamics study of diffusion of krypton in water at different temperatures. *International Journal of Modern Physics B*, *30*(11), 1650064.

Bodey, G. P., & Nance, J. (1972). Amoxicillin: in vitro and pharmacological studies. *Antimicrobial Agents and Chemotherapy*, *1*(4), 358–362.

Boeren, M., Michiels, D., Verhoeve, P., Aken, K. v., & Van de Keijser, H. (2006). Amoxicillin, stability and solubility. In *12th European Poultry Conference*.

Boles, M., Girven, R., & Gane, P. (1978). The structure of amoxycillin trihydrate and a comparison with the structures of ampicillin. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, *34*(2), 461–466.

Borhani, D. W., & Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of Computer-aided Molecular Design*, *26*(1), 15–26.

Born, M. (1920). Volumen und hydratationswärme der ionen. *Zeitschrift für Physik*, *1*(1), 45–48.

Brogden, R., Speight, T., & Avery, G. (1975). Amoxycillin: A review of its antibacterial and pharmacokinetic properties and therapeutic use. *Drugs*, *9*(2), 88–140.

Brooks, B. R., Brooks III, C. L., Mackerell Jr, A. D., Nilsson, L., Petrella, R. J., Roux, B., . . . others (2009). CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–1614.

Bush, K., & Bradford, P. A. (2016). β-Lactams and β-lactamase inhibitors: an overview. *Cold Spring Harbor Perspectives in Medicine*, *6*(8), a025247.

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, *126*(1), 014101.

Caron, F., Ducrotte, P., Lerebours, E., Colin, R., Humbert, G., & Denis, P. (1991). Effects of amoxicillin-clavulanate combination on the motility of the small intestine in human beings. *Antimicrobial Agents and Chemotherapy*, *35*(6), 1085–1088.

Caspi, A., Granek, R., & Elbaum, M. (2000). Enhanced diffusion in active intracellular transport. *Physical Review Letters*, *85*(26), 5655.

Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J., & Sanders, A. G. (1940). Penicillin as a chemotherapeutic agent. *The Lancet*, *236*(6104), 226–228.

Chandler, D. (1987). *Introduction to Modern Statistical Mechanics* (Vol. 5). Oxford University Press, New York.

Crank, J. (1979). *The Mathematics of Diffusion*. Oxford university press.

Crea, F., Cucinotta, D., De Stefano, C., Milea, D., Sammartano, S., & Vianelli, G. (2012). Modeling solubility, acid–base properties and activity coefficients of amoxicillin, ampicillin and (+) 6-aminopenicillanic acid, in nacl (aq) at different ionic strengths and temperatures. *European Journal of Pharmaceutical Sciences*, *47*(4), 661–677.

Dahal, U., & Adhikari, N. P. (2012). Molecular dynamics study of diffusion of heavy water in normal water at different temperatures. *Journal of Molecular Liquids*, *167*, 34–39.

Danelon, C., Nestorovich, E. M., Winterhalter, M., Ceccarelli, M., & Bezrukov, S. M. (2006). Interaction of zwitterionic penicillins with the OmpF channel facilitates their translocation. *Biophysical Journal*, *90*(5), 1617–1627.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092.

Darken, L. S. (1948). Diffusion, mobility and their interrelation through free energy in binary metallic systems. *Transactions of the Metallurgical Society of AIME*, *175*, 184–201.

Dasari, S., & Mallik, B. S. (2020). Solubility and solvation free energy of a cardiovascular drug, LASSBio-294, in ionic liquids: A computational study. *Journal of Molecular Liquids*, 112449.

David, v. d. S., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718.

Demiralay, E. Ç., Koç, D., Daldal, Y. D., & Çakır, C. (2012). Determination of chromatographic and spectrophotometric dissociation constants of some beta lactam antibiotics. *Journal of Pharmaceutical and Biomedical Analysis*, *71*, 139–143.

De Souza, O. N., & Ornstein, R. L. (1997). Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophysical Journal*, *72*(6), 2395–2397.

De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, *59*(9), 4035–4061.

Dixit, S. B., Bhasin, R., Rajasekaran, E., & Jayaram, B. (1997). Solvation thermodynamics of amino acids assessment of the electrostatic contribution and force-field dependence. *Journal of the Chemical Society, Faraday Transactions*, *93*(6), 1105–1113.

Dizaj, S. M., Vazifehasl, Z., Salatin, S., Adibkia, K., & Javadzadeh, Y. (2015). Nanosizing of drugs: effect on dissolution rate. *Research in Pharmaceutical Sciences*, *10*(2), 95.

Dünweg, B., & Kremer, K. (1993). Molecular dynamics simulation of a polymer chain in solution. *The Journal of Chemical Physics*, *99*(9), 6983–6997.

Easteal, A. J., Price, W. E., & Woolf, L. A. (1989). Diaphragm cell for high-temperature diffusion measurements. Tracer diffusion coefficients for water to 363 K. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, *85*(5), 1091–1097.

Ehrlich, P., & Bertheim, A. (1912). ber das salzsaure 3.3'-diamino-4.4'-dioxy-arsenobenzol und seine nchsten verwandten. *Berichte der Deutschen Chemischen Gesellschaft*, *45*(1), 756–766.

Einstein, A., et al. (1905). On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der Physik*, *17*(549-560), 208.

Ercolessi, F. (1997). A molecular dynamics primer. *Spring College in Computational Physics, ICTP, Trieste*, **19**.

Felix, I., Moreira, L., Chiavone-Filho, O., & Mattedi, S. (2016). Solubility measurements of amoxicillin in mixtures of water and ethanol from 283.15 to 298.15 k. *Fluid Phase Equilibria*, *422*, 78–86.

Fick, A. (1855). Ueber diffusion. *Annalen der Physik*, *170*(1), 59–86.

Fiorentini, R., Kremer, K., & Potestio, R. (2020). Ligand-protein interactions in lysozyme investigated through a dual-resolution model. *Proteins: Structure, Function, and Bioinformatics*, *88*(10), 1351–1360.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology*, *10*(3), 226.

Fouladgar, M., Hadjmohammadi, M. R., Khalilzadeh, M. A., Biparva, P., Teymoori, N., & Beitollah, H. (2011). Voltammetric determination of amoxicillin at the electrochemical sensor ferrocenedicarboxylic acid multi wall carbon nanotubes paste electrode. *International Journal of Electrochemical Science*, *6*, 1355–1366.

Frenkel, D., & Smit, B. (2002). *Understanding Molecular Simulation: From Algorithms to Applications* (Vol. 1). Academic Press, USA.

Gallo, M. T., Grant, B. J., Teodoro, M. L., Melton, J., Cieplak, P., Phillips Jr, G. N., & Stec, B. (2009). Novel procedure for thermal equilibration in molecular dynamics simulation. *Molecular Simulation*, *35*(5), 349–357.

Geddes, A. M., Klugman, K. P., & Rolinson, G. N. (2007). Introduction: historical perspective and development of amoxicillin/clavulanate. *International Journal of Antimicrobial Agents*, *30*, 109–112.

Gibson, J., Goland, A. N., Milgram, M., & Vineyard, G. (1960). Dynamics of radiation damage. *Physical Review*, *120*(4), 1229.

Gresser, U. (2001). Amoxicillin-clavulanic acid therapy may be associated with severe side effects-review of the literature. *European Journal of Medical Research*, *6*(4), 139–149.

Gunsteren, W. F. V., & Berendsen, H. J. C. (1988). A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, *1*(3), 173–185.

Gunsteren, W. F. V., & Berendsen, H. J. C. (1990). Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, *29*(9), 992–1023.

Hancock, R. E., & Bell, A. (1989). Antibiotic uptake into gram-negative bacteria. *Perspectives in Antiinfective Therapy*, 42–53.

Hansen, J.-P., & McDonald, I. R. (2013). *Theory of Simple Liquids: With Applications to Soft Matter*. Academic Press.

Harmandaris, V., Adhikari, N., van der Vegt, N. F., Kremer, K., Mann, B., Voelkel, R., . . . Liew, C. (2007). Ethylbenzene diffusion in polystyrene: United atom atomistic/coarse grained simulations and experiments. *Macromolecules*, *40*(19), 7026–7035.

Harris, K. R., & Woolf, L. A. (2004). Temperature and volume dependence of the viscosity of water and heavy water at low temperatures. *Journal of Chemical & Engineering Data*, *49*(4), 1064–1069.

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. (1997). LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447.

Himmelblau, D. (1964). Diffusion of dissolved gases in liquids. *Chemical Reviews*, *64*(5), 527–550.

Höltje, J.-V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiology and Molecular Biology Reviews*, *62*(1), 181–203.

Holz, M., Heil, S. R., & Sacco, A. (2000). Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. *Physical Chemistry Chemical Physics*, *2*(20), 4740–4742.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38.

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current Opinion in Microbiology*, *51*, 72–80.

Huttner, A., Bielicki, J., Clements, M. N., Frimodt-Møller, N., Muller, A. E., Paccaud, J.-P., & Mouton, J. W. (2020). Oral amoxicillin and amoxicillin–clavulanic acid: properties, indications and usage. *Clinical Microbiology and Infection*, *26*(7), 871–879.

Jamali, S. H., Wolff, L., Becker, T. M., Bardow, A., Vlugt, T. J., & Moultos, O. A. (2018). Finite-size effects of binary mutual diffusion coefficients from molecular dynamics. *Journal of Chemical Theory and Computation*, *14*(5), 2667–2677.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*(14), 2690.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236.

Kapoor, G., Saigal, S., & Elongavan, A. (2017). Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology, Clinical Pharmacology*, *33*(3), 300.

Kariper, S. E. (2017). Spectroscopic and Quantum Chemical Studies on Some β-Lactam Inhibitors. *Turkish Computational and Theoretical Chemistry*, *1*(2), 13–26.

Karthikeyan, K., & Meyer, M. T. (2006). Occurrence of antibiotics in wastewater treatment facilities in Wisconsin, USA. *Science of the Total Environment*, *361*(1-3), 196–207.

Kaur, S. P., Rao, R., & Nanda, S. (2011). Amoxicillin: a broad spectrum antibiotic. *International Journal of Pharmacy and Pharmaceutical Sciences*, *3*(3), 30–7.

Khanal, S. P., & Adhikari, N. P. (2022). Thermodynamic and transport properties of amoxicillin. *Journal of Molecular Liquids*, 118865.

Khanal, S. P., Kandel, Y. P., & Adhikari, N. P. (2019). Transport properties of zwitterion glycine, diglycine, and triglycine in water. *AIP Advances*, *9*(6), 065303-10.

Khanal, S. P., Koirala, R. P., Mishra, E., & Adhikari, N. P. (2021a). Molecular dynamics study of structural properties of γ-aminobutyric acid (gaba). *BIBECHANA*, *18*(1), 67–74.

Khanal, S. P., Poudel, B., Koirala, R. P., & Adhikari, N. P. (2021b). Solvation free energy of protonated lysine: Molecular dynamics study. *Journal of Nepal Physical Society*, *7*(2), 69–75.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, *3*(5), 300–313.

Klimovich, P. V., Shirts, M. R., & Mobley, D. L. (2015). Guidelines for the analysis of free energy calculations. *Journal of Computer-aided Molecular Design*, *29*(5), 397–411.

Koirala, R. P., Bhusal, H. P., Khanal, S. P., & Adhikari, N. P. (2020). Effect of temperature on transport properties of cysteine in water. *AIP Advances*, *10*(2), 025122.

Koirala, R. P., Khanal, S. P., & Adhikari, N. P. (2019). Transport properties of cysteine dimer in water. *Himalayan Physics*, *8*, 11–18.

Koirala, R. P., Pokhrel, R., Baral, P., Tiwari, P. B., Chapagain, P. P., & Adhikari, N. P. (2021). Structural insights into the repair mechanism of AGT for methyl-induced DNA damage. *Biological Chemistry*, *402*(10), 1203–1211.

Koirala, R. P., Thapa, B., Khanal, S. P., Powrel, J., Adhikari, R. P., & Adhikari, N. P. (2021). Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor. *Journal of Physics Communications*, *5*(3), 035010.

Koniakhin, S., Eliseev, I., Terterov, I., Shvidchenko, A., Eidelman, E., & Dubina, M. (2015). Molecular dynamics-based refinement of nanodiamond size measurements obtained with dynamic light scattering. *Microfluidics and Nanofluidics*, *18*(5-6), 1189–1194.

Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*. Pearson education.

Lemons, D. S., & Gythiel, A. (1997). Paul langevin's 1908 paper "on the theory of brownian motion"["sur la théorie du mouvement brownien," cr acad. sci.(paris) 146, 530–533 (1908)]. *American Journal of Physics*, *65*(11), 1079–1081.

Levitt, M., & Warshel, A. (1975). Computer simulation of protein folding. *Nature*, *253*(5494), 694–698.

Levy, Y., & Onuchic, J. N. (2006). Water mediation in protein folding and molecular recognition. *Annual Review of Biophysics and Biomolecular Structure*, *35*, 389–415.

Lindahl, E., Hess, A., Buuren, R. v., Apol, E., Meulenhoff, P., Tieleman, D., ... Berendsen, H. J. C. (2010). Gromacs User Manual, Version 4.5. 6.

Lindahl, E., Hess, B., & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modeling Annual*, *7*(8), 306–317.

Lobanovska, M., & Pilla, G. (2017). Focus: drug development: Penicillin's discovery and antibiotic resistance: lessons for the future? *The Yale Journal of Biology and Medicine*, *90*(1), 135.

Longsworth, L. (1953). Diffusion measurements, at 25, of aqueous solutions of amino acids, peptides and sugars. *Journal of the American Chemical Society*, *75*(22), 5705–5709.

MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., ... others (1998). All-atom empirical potential for molecular modeling

and dynamics studies of proteins. *The Journal of Physical Chemistry B*, *102*(18), 3586–3616.

Makarov, V. A., Feig, M., Andrews, B. K., & Pettitt, B. M. (1998). Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophysical Journal*, *75*(1), 150–158.

Malloci, G., Vargiu, A. V., Serra, G., Bosin, A., Ruggerone, P., & Ceccarelli, M. (2015). A database of force-field parameters, dynamics, and properties of antimicrobial compounds. *Molecules*, *20*(8), 13997–14021.

Markesteijn, A., Hartkamp, R., Luding, S., & Westerweel, J. (2012). A comparison of the value of viscosity for several water models using Poiseuille flow in a nano-channel. *The Journal of Chemical Physics*, ***136***(13), 134104.

Matos, G. D. R., Kyu, D. Y., Loeffler, H. H., Chodera, J. D., Shirts, M. R., & Mobley, D. L. (2017). Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *Journal of Chemical & Engineering Data*, *62*(5), 1559–1569.

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

McQuarrie, D. (2000). *Statistical Mechanics*. University Science Books.

Mehrer, H. (2007). *Diffusion in Solids: Fundamentals, Methods, Materials, Diffusion-Controlled Processes* (Vol. 155). Springer Science & Business Media.

Mohr, P. J., Taylor, B. N., & Newell, D. B. (2008). CODATA recommended values of the fundamental physical constants: 2006. *Journal of Physical and Chemical Reference Data*, *80*(3), 633–1284.

Moktan, H., Panday, A., & Adhikari, N. (2012). Molecular dynamics study of diffusion of different inert gases like neon and argon in water at different temperatures. *International Journal of Modern Physics B*, *26*(03), 1250016.

Mortier, J., Rakers, C., Bermudez, M., Murgueitio, M. S., Riniker, S., & Wolber, G. (2015). The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discovery Today*, *20*(6), 686–702.

Mosher, M. (1992). *Organic Chemistry. (Morrison, Robert Thornton; Boyd, Robert Neilson).* ACS Publications.

Nogrady, T., Weaver, D. F., et al. (2005). *Medicinal Chemistry: A Molecular and Biochemical Approach*. Oxford University Press.

Noya, E. G., Conde, M., & Vega, C. (2008). Computing the free energy of molecular solids by the einstein molecule approach: Ices xiii and xiv, hard-dumbbells and a patchy model of proteins. *The Journal of Chemical Physics*, *129*(10), 104704.

Noyes, A. A., & Whitney, W. R. (1897). The rate of solution of solid substances in their own solutions. *Journal of the American Chemical Society*, *19*(12), 930–934.

Olsen, R., Kvamme, B., & Kuznetsova, T. (2016). Free energy of solvation and henry's law solubility constants for mono-, di-and tri-ethylene glycol in water and methane. *Fluid Phase Equilibria*, *418*, 152–159.

Oostenbrink, C., Soares, T. A., Van der Vegt, N. F. A., & Gunsteren, W. F. V. (2005). Validation of the 53A6 GROMOS force field. *European Biophysics Journal*, *34*(4), 273–284.

Oostenbrink, C., Villa, A., Mark, A. E., & Gunsteren, W. F. V. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, *25*(13), 1656–1676.

Pages, J.-M., James, C. E., & Winterhalter, M. (2008). The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. *Nature Reviews Microbiology*, *6*(12), 893–903.

Paluch, A. S., Cryan III, D. D., & Maginn, E. J. (2011). Predicting the solubility of the sparingly soluble solids 1, 2, 4, 5-tetramethylbenzene, phenanthrene, and fluorene in various organic solvents by molecular simulation. *Journal of Chemical & Engineering Data*, *56*(4), 1587–1595.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802.

Pokharel, S., Khanal, S. P., & Adhikari, N. P. (2019). Solvation free energy of light alkanes in polar and amphiphilic environments. *BIBECHANA*, *16*, 92–105.

Pokharel, S., Pantha, N., & Adhikari, N. (2016). Diffusion coefficients of nitric oxide in water: A molecular dynamics study. *International Journal of Modern Physics B*, *30*(27), 1650205-20.

Poudyal, I., & Adhikari, N. P. (2014). Temperature dependence of diffusion coefficient of carbon monoxide in water: A molecular dynamics study. *Journal of Molecular Liquids*, *194*, 77–84.

Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Physical Review*, *136*(2A), A405.

Raju, T. (1999). The Nobel chronicles. 1945: Sir Alexander Fleming (1881-1955); Sir Ernst Boris Chain (1906-79); and Baron Howard Walter Florey (1898-1968). *Lancet (London, England)*, *353*(9156), 936–936.

Rapaport, D. C. (2004). *The Art of Molecular Dynamics Simulation*. Cambridge University Press.

Rathbun, R., & Babb, A. (1961). Self-diffusion in liquids. iii. temperature dependence in pure liquids. *The Journal of Physical Chemistry*, *65*(6), 1072–1074.

Reed, M. S., & Flurchick, K. M. (1996). Investigation of artifacts due to periodic boundary conditions. *Computer Physics Communications*, *95*(1), 39–46.

Rice, L. B. (2012). Mechanisms of resistance and clinical relevance of resistance to β-lactams, glycopeptides, and fluoroquinolones. In *Mayo clinic proceedings* (Vol. 87, pp. 198–208).

Rolinson, G. (1973). Laboratory evaluation of amoxycillin. *Chemotherapy*, *18*(Suppl. 1), 1–10.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327–341.

Santos, S. M., Rocha, J., & Mafra, L. (2013). NMR crystallography: toward chemical shift-driven crystal structure determination of the β-Lactam antibiotic amoxicillin trihydrate. *Crystal Growth & Design*, *13*(6), 2390–2395.

Scheffers, D.-J., & Pinho, M. G. (2005). Bacterial cell wall synthesis: new insights from localization studies. *Microbiology and Molecular Biology Reviews*, *69*(4), 585–607.

Sharma, K., & Adhikari, N. P. (2014). Temperature dependence of diffusion coefficient of nitrogen gas in water: A molecular dynamics study. *International Journal of Modern Physics B*, *28*(14), 1450084.

Shirts, M. R., & Mobley, D. L. (2013). An introduction to best practices in free energy calculations. In *Biomolecular Simulations* (pp. 271–311). Springer.

Shirts, M. R., & Pande, V. S. (2005a). Comparison of efficiency and bias of free energies computed by exponential averaging, the bennett acceptance ratio, and thermodynamic integration. *The Journal of Chemical Physics*, *122*(14), 144107.

Shirts, M. R., & Pande, V. S. (2005b). Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *The Journal of Chemical Physics*, *122*(13), 134508.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., & Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of Chemical Theory and Computation*, *6*(5), 1509–1519.

Singla, M., Kumar, H., & Jindal, R. (2014). Solvation behaviour of biologically active compounds in aqueous solutions of antibacterial drug amoxicillin at different temperatures. *The Journal of Chemical Thermodynamics*, *76*, 100–115.

Skyner, R., McDonagh, J., Groom, C., Van Mourik, T., & Mitchell, J. (2015). A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Physical Chemistry Chemical Physics*, *17*(9), 6174–6191.

Sutherland, R. (1964). The nature of the insensitivity of Gram-negative bacteria towards penicillins. *Microbiology*, *34*(1), 85–98.

Sutherland, R., Croydon, E., & Rolinson, G. (1972). Amoxycillin: a new semi-synthetic penicillin. *British Medical Journal*, *3*(5817), 13–16.

Tazi, S., Boţan, A., Salanne, M., Marry, V., Turq, P., & Rotenberg, B. (2012). Diffusion coefficient and shear viscosity of rigid water models. *Journal of Physics: Condensed Matter*, *24*(28), 284117-4.

Thapa, S., & Adhikari, N. (2013). A molecular dynamics study of oxygen gas in water at different temperatures. *International Journal of Modern Physics B*, *27*(08), 1350023.

Todd, P. A., & Benfield, P. (1990). Amoxicillin/clavulanic acid. *Drugs*, *39*(2), 264–307.

Tripathi, K. (2013). *Essentials of Medical Pharmacology*. JP Medical Ltd.

Tuckerman, M. (2010). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press.

Turner, P. (2005). XMGRACE, Version 5.1. 19. *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR*.

Umecky, T., Kuga, T., & Funazukuri, T. (2006). Infinite dilution binary diffusion coefficients of several α-amino acids in water over a temperature range from (293.2 to 333.2) K with the Taylor dispersion technique. *Journal of Chemical & Engineering Data*, *51*(5), 1705–1710.

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, *32*(2), 149–167.

Vuillemin, P. (1889). *Antibiose et symbiose*.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174.

Yalkowsky, S. H., & Valvani, S. C. (1980). Solubility and partitioning I: solubility of nonelectrolytes in water. *Journal of Pharmaceutical Sciences*, *69*(8), 912–922.

Yan, Z., Wang, J., Liu, W., & Lu, J. (1999). Apparent molar volumes and viscosity B-coefficients of some α-amino acids in aqueous solutions from 278.15 to 308.15 K. *Thermochimica Acta*, ***334***(1-2), 17–27.

Yeh, I.-C., & Hummer, G. (2004). System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *The Journal of Physical Chemistry B*, *108*(40), 15873–15879.

Yui, K., Noda, Y., Koido, M., Irie, M., Watanabe, I., Umecky, T., & Funazukuri, T. (2013). Binary Diffusion Coefficients of Aqueous Straight-Chain Amino Acids at Infinitesimal Concentration and Temperatures from (298.2 to 333.2) K. *Journal of Chemical & Engineering Data*, *58*(10), 2848–2853.

Zacharias, M., Straatsma, T., & McCammon, J. (1994). Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *The Journal of Chemical Physics*, *100*(12), 9025–9031.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, *22*(8), 1420–1426.

APPENDIX

A.1   For Energy Minimization Run:

```
cpp                 =    / l i b / c p p
define              =    −DFLEX_SPCE
constraints         =    none
integrator          =    steep   ; steepest descent method
nsteps              =    5000000   ; total number of steps
;
;         Energy minimizing stuff

emtol               =    150       ; force tolerance (KJ mol^−1 nm^−1)
emstep              =    0.001     ; initial step size (nm)
nstcomm             =    1         ; frequency for central of mass motion
        removal
ns_type             =    grid      ; neighbour searching type
rlist               =    1.0       ; cut off distance for the short range
        neighbor list
coulombtype         =    PME
vdwtype             =    cut−off
rcoulomb            =    1.0       ; Coulmob cut off
rvdw                =    1.0       ; vanderwaals cut off
nstxtcout           =    20
Tcoupl              =    no
Pcoupl              =    no
gen_vel             =    no
```

A.2   For Equilibration Run:

```
;PREPROCESSING parameters
cpp                 =    / l i b / c p p
integrator          =    md
dt                  =    0.002
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.
vdwtype              = cut-off
coulombtype          = PME        ; Particle Mesh Ewald for long-range
fourierspacing       = 0.12       ; grid spacing for FFT
pme-order            = 4
ewald-rtol           = 1e-05
rcoulomb             =   1.0
rvdw                 =   1.0
epsilon-r            =   1.0
;BERENDSEN TEMPERATURE COUPLING is on in two groups
Tcoupl               =   v-rescale
tc-grps              =   system
tau_t                =   0.01
ref_t                =   298
;PRESSURE COUPLING is on
Pcoupl               =   berendsen
tau_p                =   0.8
compressibility      =   4.6e-5
ref_p                =   1.0
;SIMULATED ANNEALING parameters are not specified.
;GENERATE VELOCITIES is on at 298K
gen_vel              =   yes        ; generate initially
gen_temp             =   298
gen_seed             =   173529     ;give different values for different
     trials.
ld_seed              =   1993
;BONDS parameters
constraints          = all-bonds
constraint-algorithm = SHAKE
unconstrained-start  = no
pbc                  =   xyz
```

A.3   For Production Run:

```
;PREPROCESSING parameters
tinit                = 0
integrator           = md
dt                   = 0.002
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.
vdwtype                = cut-off
coulombtype            = PME
fourierspacing         = 0.12      ; grid spacing for FFT
pme-order              = 4
ewald-rtol             = 1e-05
rcoulomb               = 1.0
rvdw                   = 1.0
epsilon-r              = 1.0
;BERENDSEN TEMPERATURE COUPLING is on in two groups
Tcoupl                 = v-rescale
tc-grps                = system
tau_t                  = 0.01
ref_t                  = 298

;PRESSURE COUPLING is on
Pcoupl                 = no
;SIMULATED ANNEALING parameters are not specified.
;GENERATE VELOCITIES is on at 298K
gen_vel                = no;        ; generate initially
gen_temp               = 298
gen_seed               = 173529   ;give different values for different
    trials.
ld_seed                = 1993
;BONDS parameters
constraints            = all-bonds
constraint-algorithm   = SHAKE
unconstrained-start    = yes
pbc                    = xyz
;SIMULATED ANNEALING parameters are not specified.
```

B. Articles published in International Journals

Khanal, S. P., & Adhikari, N. P. Thermodynamic and transport properties of amoxicillin. (2022). *Journal of Molecular Liquids*, *354*, 118865.

Khanal, S. P., Kandel, Y. P., & Adhikari, N. P. (2019). Transport properties of zwitterion glycine, diglycine, and triglycine in water. *AIP Advances*, *9*(6), 065303.

Koirala, R. P., Bhusal, H. P., Khanal, S. P., & Adhikari, N. P. (2020). Effect of temperature on transport properties of cysteine in water. *AIP Advances*, *10*(2), 025122.

Koirala, R. P., Thapa, B., Khanal, S. P., Powrel, J., Adhikari, R. P., & Adhikari, N. P. (2021). Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor. *Journal of Physics Communications*, *5*(3), 035010.

C. Articles published in National Journals

Khanal, S. P., Koirala, R. P., Mishra, E., & Adhikari, N. P. (2021). Molecular dynamics study of structural properties of γ-aminobutyric acid (GABA). *BIBECHANA*, *18*(1), 67–74.

Khanal, S. P., Poudel, B., Koirala, R. P., & Adhikari, N. P. (2021). Solvation free energy of protonated lysine: Molecular dynamics study. *Journal of Nepal Physical Society*, *7*(2), 69–75.

Pokharel, S., Khanal, S. P., & Adhikari, N. P. (2019). Solvation free energy of light alkanes in polar and amphiphilic environments. *BIBECHANA*, *16*, 92–105.

Koirala, R. P., Khanal, S. P., & Adhikari, N. P. (2019). Transport properties of cysteine dimer in water. *Himalayan Physics*, *8*, 11–18.

Koirala, R. P., Thapa, B., Khanal, S. P., Adhikari, R. P., & Adhikari, N. P. (2020). Intra-molecular conformational stability in human growth hormone. *Journal of Nepal Physical Society*, *6*(2), 41–49.

D. Participation

- **Thermodynamic and transport properties of amoxicillin.**
  Shyam Prakash Khanal, and Narayan P. Adhikari. APS March meeting conference, March 15-19, 2021. Presenter: Shyam Prakash Khanal

- **Solvation Free Energy of Protonated Lysine in Water.**
  Shyam Prakash Khanal, Bhuwan Poudel, and Narayan P. Adhikari. ANPA conference, July 16-18, 2021. Presenter: Shyam Prakash Khanal

- **Molecular Dynamics Study of Amoxicillin in water.**
  Shyam Prakash Khanal, and Narayan P. Adhikari. ANPA conference, July 17-19, 2020. Presenter: Shyam Prakash Khanal

- Participated in the academic discussion with Prof. Dr. Nisanth N. Nair and his group at Indian Institute of Technology (IIT), Kanpur, India during January 13-19, 2020.

- **Molecular Dynamics Study of Diffusion of Amoxicillin in water.**
  Shyam Prakash Khanal, and Narayan P. Adhikari. 35th Annual General Meeting of Nepal Physical Society on November 24, 2018. Presenter: Shyam Prakash Khanal

- **Molecular Dynamics Study of Diffusion of Amoxicillin in water.**
  Shyam Prakash Khanal, and Narayan P. Adhikari. I^{st} NRN GLOBAL KNOWLEDGE CONVENTION Organized by NRNA on October 12-14, 2018 in Kathmandu Presenter: Shyam Prakash Khanal

- Participated in **21 hours lecturer series on Research Methodology** delivered by **Prof. Dr. Subodh R. Shenoy**, during November 6 November to 22 December 2017 held on Central Department of Physics, Kirtipur.

- Participated in the academic discussion on the Condensed Matter Physics and interacted with the faculty and PhD Scholars at Indian Institute of Technology (IIT), Kharagpur, India during June 5-7, 2017.

Thermodynamic and transport properties of amoxicillin

Check for updates

Shyam Prakash Khanal, Narayan Prasad Adhikari *

Central Department of Physics, Tribhuvan University, Kathmandu, Nepal

ABSTRACT

We have carried out MD simulations of amoxicillin, a drug molecule used as antibacterial agent, in water and ethanol to study thermodynamic properties like solvation free energy and transport phenomena like diffusion, viscosity. The free energy of solvation of amoxicillin in water and ethanol have been estimated at 310 K using thermodynamic integration (TI) and free energy perturbation (FEP) based methods. It has been obtained that the free energy of solvation of amoxicillin in water is higher by $\approx$ 105 kJ/mol in TIP3P model than in ethanol. We have also reported the solvent accessible surface area (SASA) of solute as well as the hydrogen bonds between solute and solvent for different solvents. The analysis of SASA and hydrogen bond provides more insight on effect of solvent environment in solvation process of amoxicillin. We have also estimated the diffusion coefficient and our results show that diffusion coefficients of amoxicillin depends on solvent environment, as expected, and it is higher in water than that of ethanol. Also, applying the idea of system size dependence of diffusion coefficient, the viscosity coefficient of water and solution of amoxicillin in water have also been estimated. Furthermore, the effects of temperature on diffusion has been studied. The estimated values of self-diffusion coefficients of water are in good agreement with previously reported experimental data within $\approx$ 13.5%.

1. Introduction

Amoxicillin is a penicillin-type, semi-synthetic, β-lactam antibiotic widely used to treat bacterial infections [1]. Amoxicillin molecule has a β-lactam ring, responsible for antibacterial activity; one hydroxyl and one carboxyl groups [2]. Fig. 1 represents snapshot of amoxicillin molecule in zwitterionic form. Amoxicillin acts by interrupting bacterial cell-wall formation by covalent binding to essential penicillin-binding proteins (PBPs), enzymes involved in the terminal steps of peptidoglycan cross-linking in both Gram-negative and Gram-positive bacteria [3]. Cross-linking of D-alanyl-alanine by glycine, which strengthens the cell wall, occurs with PBPs; and β-lactam ring bind with the PBPs. As a result, the PBPs are inactivated and unavailable for cell wall synthesis; guide to lysis of bacterium [4,5]. The optimized geometry and spectra of amoxicillin were already studied using density functional theory (DFT) [6]. The kinetic analysis of amoxicillin in aqueous medium was also performed [7].

Solubility, dissolution rate and intestinal permeability are the major factors that affect oral drugs absorption [8]. On the basis of solubility and intestinal permeability, amoxicillin is a class III type antibiotic with high solubility and low permeability. The

understanding about the solubility has significant role in many areas including pharmaceutical industry [9,10]. The dissolution rate and solubility are the essential parameters to understand about solvation; Noyes-Whitney relates them by [11]

$$\frac{dm}{dt} = \frac{DA(C_s - C)}{L}. \tag{1}$$

In this equation, $\frac{dm}{dt}$, D, A, C, C_s and L are the rate of dissolution, diffusion coefficient, solute surface area in contact with solvent, instantaneous solute concentration in the bulk solvent, saturation solubility of drug in solution and thickness of the diffusion layer respectively. The Noyes-Whitney equation also establish a connection among i) dissolution rate: kinetic term, ii) solubility: thermodynamic term and iii) diffusion: transport property [12]. The dissolution rate can be enhanced by increasing effective surface area of drug particles [13].

Many experimental techniques can be used to estimate solubility of drugs, a key factor to understand bioavailability of drugs. The solubility of paracetamol in organic solvent was reported using experimental as well as molecular dynamics (MD) study [9]. It can also be estimated from the knowledge of solvation free energy [14–17]. The solvation free energy, change in free energy due to transfer of a solute molecule from solvent to ideal gas state (or vice versa) at particular temperature and pressure, can be determined from understanding of interactions between solute and solvent

* Corresponding author.
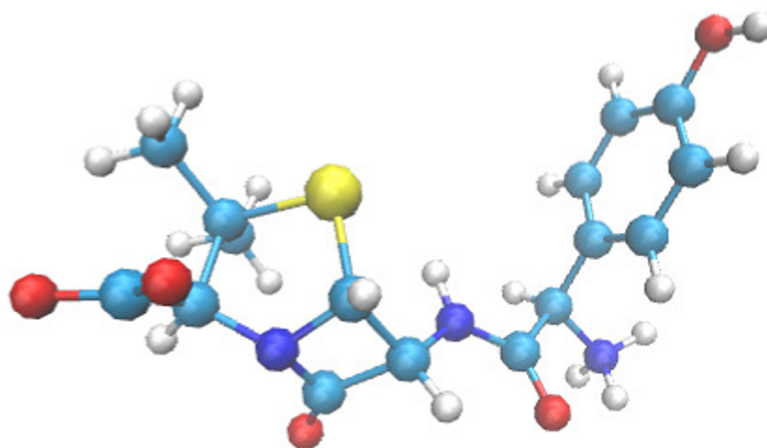*E-mail address:* narayan.adhikari@cdp.tu.edu.np (N.P. Adhikari).

Fig. 1. Snapshot of amoxicillin molecule using VMD.

using molecular dynamics (MD) study [18,19]. The solvation free energy of LASSBio-294 in different medium and estimation of relative solubility of the drug in logarithmic scale was already performed from MD study [18]. The electrostatic contributions to the solvation free energy of zwitterionic glutamic acid was reported as 471.70 kJ/mol using Finite Difference Poisson–Boltzmann method [20]. Similarly, many of biological processes in human body happens through diffusion; and diffusion also depends upon viscosity coefficient of solvent. In this context, the knowledge about transport properties like diffusion coefficient, viscosity etc. of biomolecules including drug has significant role to understand about diffusive transport and hence solvation process. Such transport properties also can be estimated from molecular dynamics study [21,22].

In drugs designing, the estimation of free energy difference between two thermodynamic states from simulations has substantial role [23]. From the estimation of difference in free energy between two thermodynamic states, we can understand about many phenomenon including protein ligand binding mechanism, protein folding mechanism, transport of drug through membrane. There are different methods to calculate the free energy difference between two states [24]. MD can be considered as powerful tool in different steps of drugs discovery. Different techniques like free energy perturbation (FEP) with umbrella sampling, steered MD etc. can be used to understand the drug target binding from thermodynamics and kinetics views [25]. Furthermore, the relative binding free energies (RBFEs), applicable in discovery of drugs, can also be estimated from MD simulation [26].

MD plays a significant role to get detail knowledge about the thermodynamic, kinetic, dynamical as well as structural properties at molecular level [27]. During simulations, the macroscopic properties can be connected with microscopic properties using statistical mechanics [28,29]. Such technique can also be used to understand the transport phenomena like diffusion, viscosity etc. which provide the information about the interaction between solute and solvent molecules [30,31].

From the solvation free energy, diffusion coefficient etc. of drugs, we can get information about the dissolution rate of the drugs as suggested by Noyes-Whitney equation. To the best of our knowledge, molecular dynamics study of amoxicillin has not been carried out to study about the transport properties, free energy calculation etc. The solvation free energy, transport properties also depends upon solvent environments. In this work, we have reported the free energy of solvation of amoxicillin in zwitterionic form in two solvents: water and ethanol using MD simulations. Also, diffusion coefficients of zwitterionic amoxicillin as

well as solvent have been reported; and we have also analyzed the effects of temperature on diffusion coefficients. Furthermore, we have also studied the effects of system size on diffusion coefficient; and also the viscosity coefficient of water and solution of amoxicillin in water have been reported.

2. Methods and methodology

2.1. Theoretical background

Estimation of free energy is one of fundamental objectives of MD simulation. During MD simulation, different approaches including free energy perturbation (FEP) and thermodynamic integration (TI) are used to estimate free energy difference between two thermodynamic state. According to FEP method, the free energy difference during the transformation of a system from initial state A to final state B at a temperature T (ΔF_{AB}) is given by [28]

$$\Delta F_{AB} = -k_B T \ln\langle \exp[-\beta(U_B - U_A)]\rangle_A. \tag{2}$$

In Eq. 2, U_A and U_B represent the potential energy function of states A and B respectively, k_B is Boltzmann constant, $\beta = 1/k_B T$; and $\langle \cdots \rangle_A$ represents the ensemble average of difference in potential energy function between two thermodynamic states A and B with respect to the state A. If the potential energy difference between two states is very high, there is a problem of convergence. In this case, the system may not cover sufficient phase space; however there must be sufficient overlap in phase space for reliable estimation of free energy difference. Such problem of convergence can be resolved by introducing many unphysical intermediate states between the two states. The intermediate states are defined by introducing coupling parameter λ with values $0 \leqslant \lambda \leqslant 1$ such that the potential energy is also a function of the coupling parameter. The 0, 1 and other values of λ represent the initial, final and other intermediate states respectively. The free energy of solvation can be estimated using thermodynamic integration (TI) method as [17]

$$\Delta G_{sol} = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \, d\lambda \tag{3}$$

Out of many factors, the rate of dissolution of drug also depends upon its diffusion coefficient. Diffusion, a transport phenomena, is measured in terms of diffusion coefficients [32]. The self diffusion coefficient is calculated using Einstein's relation. According to Einstein, the macroscopic quantity diffusion coefficient D for 3-dimensional system is calculated from the microscopic quantity mean square displacement (MSD) as [21,22],

$$D = \lim_{t \to \infty} \frac{1}{6} \frac{\langle [r(t)]^2 \rangle}{t} \qquad (4)$$

Due to long range nature of hydrodynamic interactions, the size of system used during simulation also has effect on the dynamical properties calculated under periodic boundary conditions (PBC) [33]. Some correction is necessary on the values of diffusion coefficients estimated under PBC; and the size independent value of self diffusion coefficient (D_0) from size dependent value (D_{PBC}) calculated from simulation in cubic box of size L with correction term can be estimated using the equation [34,35]:

$$D_0 = D_{PBC} + \frac{2.84 k_B T}{6 \pi \eta L} \qquad (5)$$

where k_B, T and η are the Boltzmann constant, absolute temperature of system and viscosity coefficient respectively. Also, from the Eq. 4, the viscosity coefficient of both solvent as well as solution can be estimated after performing MD simulations taking box of different sizes.

2.2. Modeling of the system

The classical MD simulations have been carried out under periodic boundary conditions using GROMACS [36] software package. During the simulations, two water models: SPC/E and TIP3P; and ethanol were used as solvents in order to understand the effects of solvents [37,38]. All atom force field model was used during the simulations. From the OPLS-AA, we investigate amoxicillin and ethanol [38]. Both non-bonded Lennard-Jones (LJ) and Coulomb potentials were taken into account for intermolecular interaction. The total intermolecular potential is defined as

$$U(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{q_i \, q_j}{4 \pi \epsilon_m \, r_{ij}} \qquad (6)$$

where $r_{ij}, \varepsilon, \sigma, \epsilon_m, q_i$ and q_j are the distance between the ith & j^{th} atoms, well depth of the LJ interaction, distance at which LJ interaction is zero, permittivity of the medium, charge of ith and j^{th} atom respectively. The partial charges for hydrogen and oxygen atom of water are $+0.4238e$ and $-0.8476e$ respectively, where e is electronic charge. Also, the non-bonded parameters for water σ_{OW-OW} and ε_{OW-OW} are 0.3165 nm and 78.2 k_B respectively. The combination rules used for the non-bonded LJ interaction parameters between two different atoms α and β are $\sigma_{\alpha\beta} = (\sigma_{\alpha\alpha} \times \sigma_{\beta\beta})^{1/2}$ and $\varepsilon_{\alpha\beta} = (\varepsilon_{\alpha\alpha} \times \varepsilon_{\beta\beta})^{1/2}$ [39].

2.3. Simulation details

2.3.1. Estimation of free energy

To estimate the free energy of solvation, we performed MD simulations taking systems of binary mixture of one amoxicillin in 1020 water molecules (system-I(a)); and one amoxicillin in 340 ethanol molecules (system-I(b)) in a cubic simulation box at 310 K temperature. The simulations were performed at 1 atm pressure under PBC. We used two TIP3P and SPC/E water models to estimate the free energy of solvation of amoxicillin. To remove van der Waals bad contact, we performed energy minimization run using steepest descent method [39]. Before production run, the system should be in the state of thermodynamic equilibrium [40]. For this, we performed equilibration run, for 10 ns with time step of 1 fs at 1 atm pressure, under NVT ensemble at first and then NPT. During each equilibration run, LINCS algorithms and Langevin dynamics [39,41] were used to constrain all bonds and solve the equations of motion respectively. Also, we used Berendsen barostat with coupling time 0.5ps and value of isothermal compressibility

4.5×10^{-5}bar respectively during each equilibration run [42]. Further, Maxwell–Boltzmann distribution was chosen to assign the initial velocities of each particle [39]. Cut off distance of 1.2nm was taken for the non-bonded interactions i.e. Coulomb and LJ; and the Coulomb long range interaction was accounted using Particle Mesh Ewald (PME) method. After the equilibration run, production run was carried for 10 ns with 1 fs time step using Langevin dynamics with same parameters used during equilibration run [39].

During free energy calculations, only non-bonded vdW and Coulomb interactions were manipulated to introduce intermediate states keeping the bonded interactions as it is. In order to transform the system from initial state to final state, we consider 21 different values of evenly spaced coupling parameter λ as: $\lambda_{Coulomb}$ = 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00 and 1.00 for Coulomb interaction; and λ_{vdW} = 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00,0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90 and 1.00 for van der Waals interaction; but in case of SPC/E water, we consider 0.75 instead of 0.7 [43]. Here, state $\lambda = 0$ means the solute and solvent are fully coupled through non-bonded Coulomb and vdW interactions, $\lambda = 1$ means solute and solvent molecules are decoupled and other intermediate values means they are coupled with different strengths.

2.3.2. Transport properties

We considered also a system having binary mixtures of 2 amoxicillin and 2160 (System-II) water molecules in cubic simulation box. The System-II was simulated at six different temperature: 293 K, 298 K, 303 K, 305 K, 310 K and 313 K. Also, the effects of system size on diffusion coefficients estimated under PBC have been analyzed from the simulations at 298 K temperature considering other three systems in cubic box of different sizes: 2 amoxicillin in 4071 water molecules (System-III(a)), 2 amoxicillin in 6504 water molecules (System-III(b)) and 2 amoxicillin in 9523 water molecules (System-III(c)). Diffusion phenomenon also depends upon the solvent environment i.e. viscosity of solvent [12]. In this context, we also performed simulations taking system with amoxicillin as solute and ethanol as solvent. The system consists of binary mixture of 2 amoxicillin molecules and 2168 ethanol molecules in cubic simulations box (System-IV).

We first performed energy minimization run for the systems II, III(a), III(b) and IV using Steepest-descent method [39]; and each of the system was equilibrated at each temperature taking NPT ensemble for 200 ns with time step of 2 fs using Leap-frog algorithms [39]. Maxwell–Boltzmann distribution was used to assign initial velocity for each particle. SHAKE algorithm was used to constraint bonds [44]. Velocity rescaling thermostat with coupling time of 0.01 ps and Berendsen barostat with coupling time 0.8 ps were used to ensure constant temperature and pressure respectively [42,45]. Also, we used cut off parameters of 1 nm for Systems-II & III, and 1.4 nm for System-IV for both non-bonded interactions; and long range Coulomb was accounted using Particle Mesh Ewald (PME) method with fourier spacing of 0.12 nm. We compared the density of each equilibrated system with previously reported experimental data of solvent and found that simulated values were in close agreement. Finally, we performed production run to each system at each temperature for 200 ns with time step of 2 fs under NVT ensemble. Velocity-rescaling thermostat with coupling time of 0.01 ps during the production runs; and initial velocities of each particle were taken from final step of equilibration run.

3. Results and discussion

In this section, we present and discuss the findings of the present work. At first, we present the free energy of solvation of amox-

icillin in water and ethanol; and then we discuss the self and binary diffusion coefficients of the systems: amoxicillin in water and amoxicillin in ethanol as well as effect of temperature and system size on diffusion coefficient. Finally, we present the viscosity coefficient of water and solution of amoxicillin in water.

3.1. Free energy of solvation

The free energy of solvation of amoxicillin has been estimated from trajectory of production run for each value of lambda using Python tool [46,47]. The free energy of solvation of the amoxicillin has been estimated in two different solvent environment: water and ethanol at 310 K temperature. Although different approaches can be used to estimate the free energy of solvation, we used two thermodynamic integration based methods: TI and TI-CUBIC. The only difference in the two methods is that trapezoidal and cubic spline rules are used by TI and TI-CUBIC for numerical integration respectively. Also, two approaches Bennett Acceptance Ratio (BAR) and Multistate Bennett Acceptance Ratio (MBAR) based on FEP method are used [48,49]. As discussed already, it is necessary to introduce many non-physical intermediate states between initial and final states. The intermediate states can be introduced using coupling different values of coupling parameter λ [46]. We consider 21 different states including the initial state of full interaction between solute and solvent through non-bonded vdW and Coulomb interactions and final state with no interaction between them. Since, the free energy of solvation can be estimated from the variation of the ensemble average of change in free energy with respect to λ i.e. $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ as a function of coupling states using Eq. 3. The variation $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ as function of λ of has been plotted for both solvents: water and ethanol. Fig. 2 (a-c) show the variation of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ as function of λ for the solvents water (TIP3P and SPC/E models) and ethanol respectively at 310 K. In the Fig. 2 (a-c), red and green color (online) represent the contribution of Coulomb and vdW interactions respectively to the solvation free energy. From the plots, it is seen that the free energy difference between two consecutive coupling states is high at beginning and decreases with increase in coupling state. This happened because in the beginning, we considered both interactions in full strength. Slowly we turned off the interactions; at first Coulomb interaction and then vdW interaction. The total free energy difference between initial state of coupling of solute to solvent and final state of decoupling between them provides the solvation free energy of the solute. The estimated values of free energy of solvation of amoxicillin in water and ethanol at 310 K using four different methods: TI, TI-CUBIC, BAR and MBAR are presented in the Table 1.

From the Table 1, it is seen that the free energy of solvation of amoxcillin in both solvents: water and ethanol has high negative values when contributions of both vdW and Coulomb interactions have been taken into account. The negative value of free energy of

solvation ensures that the solvation of amoxicillin occurs in both the solvents; and the reason for high value is that we used zwitterion form. Also, the individual contribution of vdW and Coulomb interactions have been analyzed; and it is observed from the Table 1 that the free energy of solvation of amoxicillin in both solvents has small negative value when only vdW interaction is considered. This means vdW interaction has small contribution to the solvation of amoxicillin in both solvents. In addition, it is seen that Coulomb interaction has major contributions for the solvation of amoxicillin in both the solvents: water and ethanol. The estimated values of free energy of solvation using different methods are in close agreement within 1.2%. We have also compared the estimated value of free energy of solvation of amoxicillin in water by taking the contributions of oly Coulomb interaction with previously reported value of electrostatic contributions on solvation free energy of glutamic acid in zwitterionic form. Our estimated values of solvation free of amoxicillin are $\approx$ – 530 kJ/mol and $\approx$ – 557 kJ/mol in TIP3P and SPC/E water models respectively; and the previuosly reported value for glutamic acid is – 471.70 kJ/mol using Finite Difference Poisson–Boltzmann (FDPB) method [20]. For TIP3P water model, our estimated value is in close agreement within 12.3%. Furthermore, the estimated value of free energy of solvation of amoxicillin in water is greater than in ethanol. This may happen because the solubility of amoxicillin in water has larger value than in ethanol i. e. the experimental value of solubility of amoxiciilin in water and ethanol are 4 mg/mL and 3.4 mg/mL respectively [50].

During simulations, the free energy difference between two thermodynamic states has been estimated from trajectory analysis; and hence the trajectory must be sampled at equilibrium to converge the estimated value. In order to check the convergence, we have plotted the free energy difference as a function with time of simulation for both forward and time-reversed data; and Fig. 3 (a-c) are those plots for TIP3P model, SPC/E model and ethanol respectively.

Since, it has been expected that during free energy calculation, the first 40% of simulated data for all lambda are in the non-equilibrium [47]. From the time series plots, it is seen that our calculations also converges after 0.4 fraction of simulation time as expected.

3.1.1. Hydrogen bonds and Solvent Accessible Surface Area (SASA)

We have analyzed the hydrogen bonds between solute and solvent molecules using trajectory after 10 ns production for initial i.e., coupling state represented by $\lambda = 0$. The hydrogen bond analysis also provides information about the effect of solvent environment. During hydrogen bond analysis, we used cut off parameters 0.35 nm and 30° for length and angle respectively. Fig. 4 show the plot of time evolution of number of hydrogen bonds between solute (i. e. amoxicillin) and solvents: water (TIP3P model
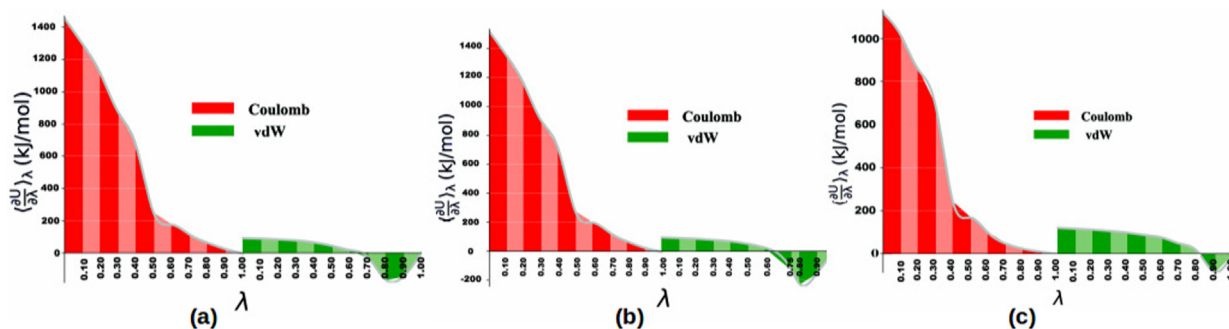
Fig. 2. Variation of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ as a function of λ at 310 K temperature taking water (TIP3P model (a), SPC/E model (b)) and ethanol (c) as solvents.

Table 1

Estimated values of free energy of solvation of amoxicillin (ΔG_{sol}) in water (TIP3P and SPC/E models) and ethanol at 310 K using TI, TI-CUBIC, BAR and MBAR methods taking individual contribution of vdW and Coulomb along with total contributions due to both the interactions.

Solvent	Model	Method	ΔG_{sol} in kJ/mol with		
			vdW only	Coulomb only	Total
Water	TIP3P	TI	- 8.47 ± 0.20	−530.47 ± 0.53	- 538.94 ± 0.57
		TI-CUBIC	- 7.09 ± 0.19	- 529.38 ± 0.53	- 536.46 ± 0.57
		BAR	- 6.82 ± 0.17	- 528.07 ± 0.91	- 534.90 ± 0.93
		MBAR	- 6.77 ± 0.20	- 528.43 ± 1.43	- 535.20 ± 1.44
	SPC/E	TI	- 0.93 ± 0.25	- 557.78 ± 0.57	- 558.71 ± 0.62
		TI-CUBIC	- 2.16 ± 0.26	- 556.60 ± 0.57	- 558.76 ± 0.63
		BAR	- 1.81 ± 0.26	- 557.42 ± 1.09	- 559.24 ± 1.12
		MBAR	- 1.97 ± 0.29	- 557.76 ± 1.84	- 559.73 ± 1.86
Ethanol		TI	- 62.35 ± 0.23	- 370.73 ± 0.74	- 433.08 ± 0.77
	–	TI-CUBIC	- 61.18 ± 0.25	- 370.15 ± 0.75	- 431.33 ± 0.79
		BAR	- 61.40 ± 0.20	- 368.05 ± 1.87	- 429.45 ± 1.88
		MBAR	- 61.35 ± 0.23	- 368.92 ± 2.22	- 430.26 ± 2.23

Fig. 3. Estimated value of free energy differences of amoxicillin as a function of simulation time for both forward and reverse directions at 310 K temperature taking water (TIP3P model (a) and SPC/E model (b)) and ethanol (c) as solvents.
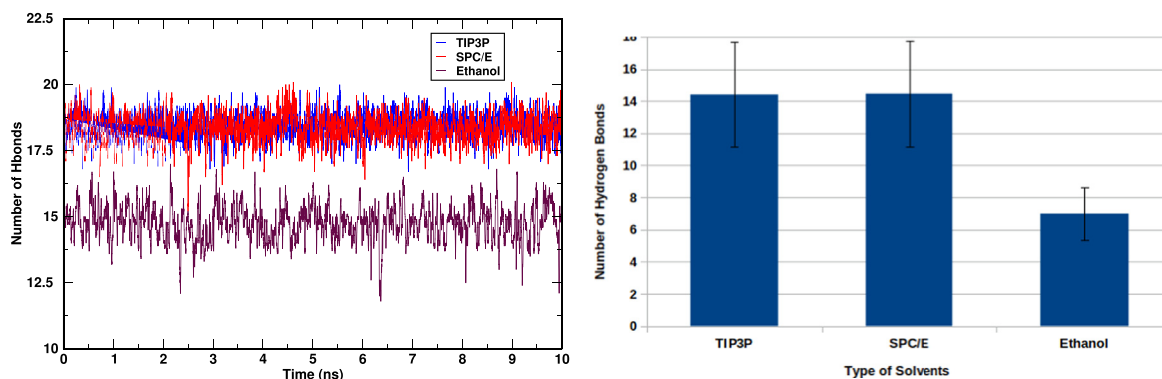
Fig. 4. Comparison of time evolution of number of hydrogen bonds between solute and different solvents (left) and average number of hydrogen bonds (right) at 310 K temperature for coupling state represented by λ=0.

and SPC/E model) and ethanol as a function of simulation time for coupling state represented by λ=0 (left) and average number of hydrogen bonds between solute and solvents (right) at 310 K temperature. From the above plots, we clearly observe higher number of hydrogen bonds with water as solvent in comparison to ethanol as solvent throughout the 10 ns simulations. The average number of hydrogen bonds in water are higher than ≈ 6.6 in ethanol; and this grater number of hydrogen bonds in water than in ethanol is in agreement with our observation with large value of solvation free energy in water than in ethanol.

Solvent Accessible Surface Area (SASA) evaluates how the solute is accessible to dissolve in a particular solvent; higher the SASA gives the greater affinity of solute to dissolve in the solvent. We have also estimated SASA in both solvent environments: water and ethanol from the trajectory after 10 ns production run. Fig. 5

represents the time evolution of SASA in two different solvents water (TIP3P and SPC/E model) and ethanol for coupling state between solute and solvent represented by λ=0. From the Fig. 5, it is seen that SASA has smaller value with ethanol as solvent in comparison to water as solvent. This information of SASA in two different solvents also supports to our estimation of solvation free energy i.e. the solvation free energy of amoxicillin in water is higher in comparison of ethanol as shown in Table 1.

3.2. Transport properties

3.2.1. Self-diffusion coefficients

Together with the estimation of free energy of solvation of amoxicillin, the transport properties of amoxicillin in two solvents water and ethanol have also been studied. In order to estimate the
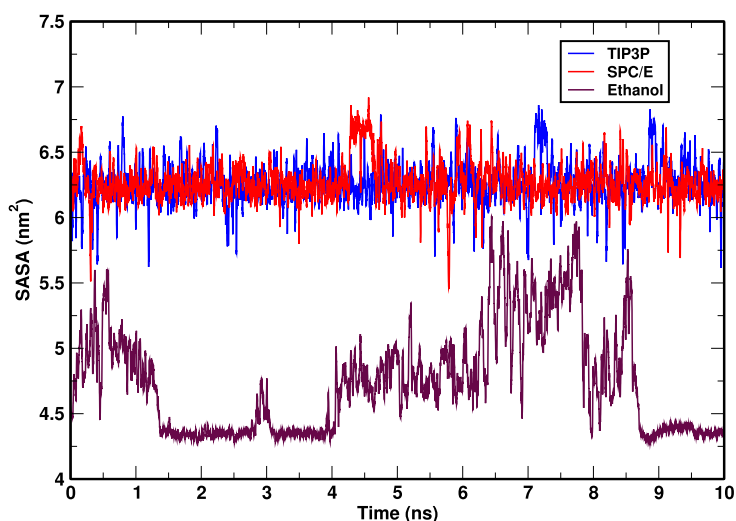
Fig. 5. Comparison of time evolution of Solvent Accessible Surface Area (SASA) of solute molecule at 310 K temperature for coupling state λ=0 in two different solvent environments: water (TIP3P and SPC/E model) and ethanol.

self diffusion coefficient of solute and solvent of the system using Einstein's relation (4), at first the mean square displacement (MSD) as a function of time in logarithmic scale has been plotted. The graphs are not linear near origin due to ballistic motion of the molecules and after some time, the graph becomes straight indicating the uniform motion of the molecules. The value of self-diffusion coefficient is given by the straight portion of the graph. We have plotted graph between mean square displacement (MSD) versus time and the data have been linearly fitted taking 1 ns time for amoxicillin and 5 ns for water and ethanol at different temperature [51]. Figs. 6 and 7 show the MSD versus time graph for amoxicillin and water respectively at different temperature. The estimated values of self diffusion coefficient of amoxicillin and water at different temperature along with their binary diffusion coefficients are presented in Table 2. The binary diffusion coefficients are calculated using Darken's relation [52].

From Table 2, it is found that the values of self diffusion coefficient of amoxicillin and water increase with the increase in tem-

perature. It is because the velocity increases and density decreases with the increase in temperature [51]. The estimated values of self diffusion coefficient of water are in close to previously reported values. From the table, it has been also observed that the values of binary diffusion coefficient are almost equal to corresponding values of self diffusion coefficient of amoxicillin due to low concentration of amoxicillin in the system. In addition, the self diffusion coefficients of amoxicillin and ethanol as well as their binary diffusion coefficient have been estimated at 298 K temperature. The estimated values of diffusion coefficients with previously reported experimental value are presented in the Table 3. From the Table 3, it is seen that the estimated value of self diffusion coefficient of ethanol is in agreement with previously reported value within 23%.

We have also studied the temperature dependence of diffusion coefficient of solute and solvent. Figs. 8 and 9 show the variation of diffusion coefficient of amoxicillin and water with temperature respectively. Further, the activation energy of amoxicillin and
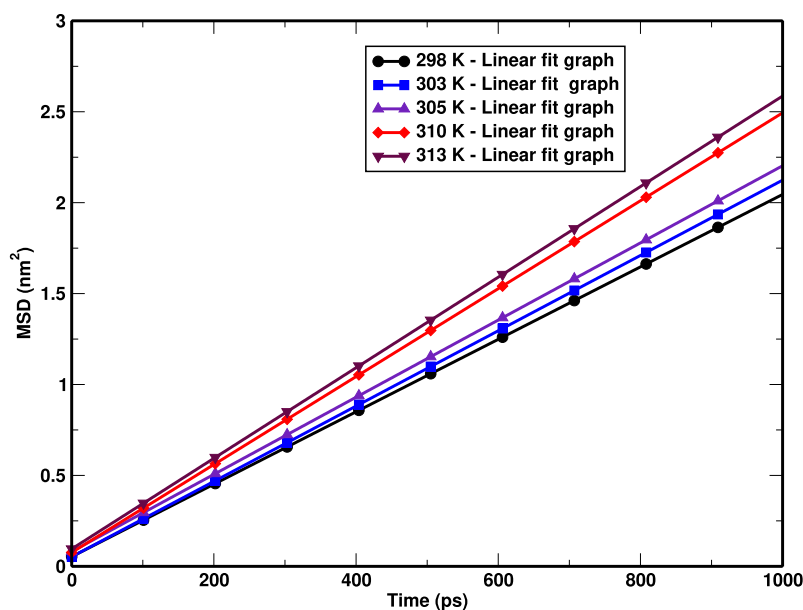
Fig. 6. MSD versus time plot for amoxicillin at different temperature.
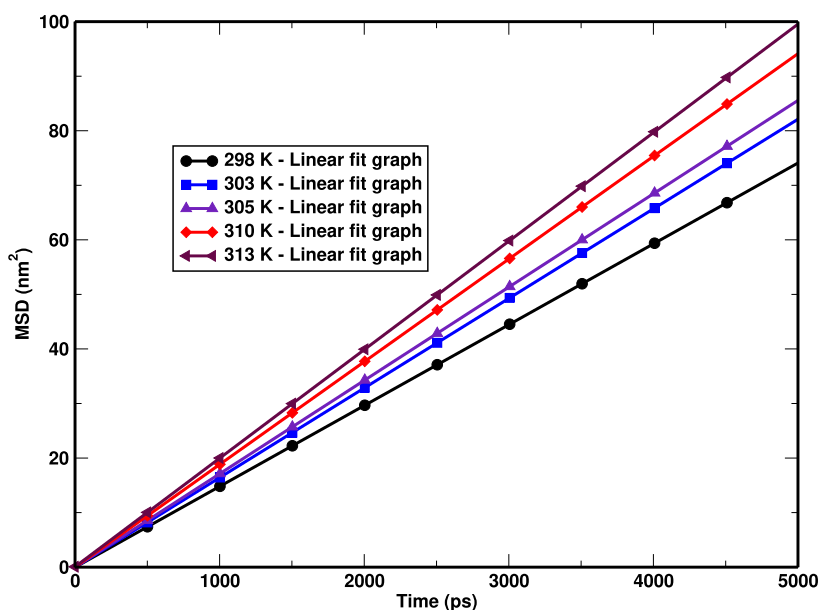
Fig. 7. MSD versus time plot for water at different temperature.

Table 2
The simulated values of self diffusion coefficients as well as mutual diffusion coefficient at different temperature taking the System-II.

Temperature (K)	Diffusion Coefficient ($\times 10^{-9}$ m^{2}s^{-1})			Mutual
	Self			
	For amoxicillin	For water		
	MSD	MSD	Ref.	
293	0.26 ± 0.03	2.20 ± 0.00	2.02 [a]	0.26
298	0.30 ± 0.05	2.49 ± 0.02	2.30 [a]	0.30
303	0.34 ± 0.04	2.74 ± 0.01	2.60 [a]	0.35
305	0.35 ± 0.03	2.85 ± 0.01	2.80 [b]	0.36
310	0.40 ± 0.04	3.14 ± 0.01	–	0.41
313	0.43 ± 0.04	3.32 ± 0.00	3.22 [a]	0.43

Ref. [a]- [53] and [b]- [51].

Table 3
The simulated values of self diffusion coefficient of solute and solvent as well as mutual diffusion coefficient at 298 K temperature taking System-IV.

Diffusion Coefficient ($\times 10^{-9}$ m^{2}s^{-1})				Mutual
Self				
For amoxicillin	For ethanol			
MSD	MSD	Ref.	% Error	
0.26 ± 0.01	1.29 ± 0.01	1.05 [54]	22.8	0.26

water have calculated using Arrhenius formula [32]. The estimated value of activation energy for amoxicillin is 0.017 kJ/mol. Also, the estimated and experimental values of activation energy of water are 15.31 kJ/mol and 17.43 kJ/mol [55] respectively. The estimated value is in agreement with experimental value within 12.2%.

3.2.2. Effects of system-size

Under periodic boundary conditions, the size of system taken during simulations also has effect on the estimated value of diffusion coefficient [34,33,35]. The effect of system size has been studied from the simulations taking other three systems (i.e. III(a), III(b) & III(c)) of different size under PBC. To estimate the diffusion coefficient at 298 K taking systems of different size, we first plot the MSD versus time graphs for amoxicillin and water as shown in Figs. 10 and 11 respectively. The size dependent values of self dif-

fusion coefficient of amoxicillin and water; and their binary diffusion coefficient estimated under PBC at 298 K are presented in Table 4. It is seen from the Table 4 that the diffusion coefficient also depends on size of system under consideration and increase with the increase in the size of simulation box, which is due to the effect of long-range hydrodynamics interactions in simulations under PBC. Also, size independent values of diffusion coefficient (D_0) and viscosity coefficient (η) have been estimated from the graph plotted between diffusion coefficient D_{PBC} calculated under PBC and reciprocal of size of simulation box ($1/L$) using Eq. 4. Figs. 12 & 13 represent the graph between D_{PBC} versus ($1/L$) for water and solution of amoxicillin in water at 298 K temperature with the size independent value for water calculated using experimental value of viscosity coefficient $D_0^{Expt.}$. We can estimate D_0 and η from the intercept and slope of the plot respectively. The estimated val-
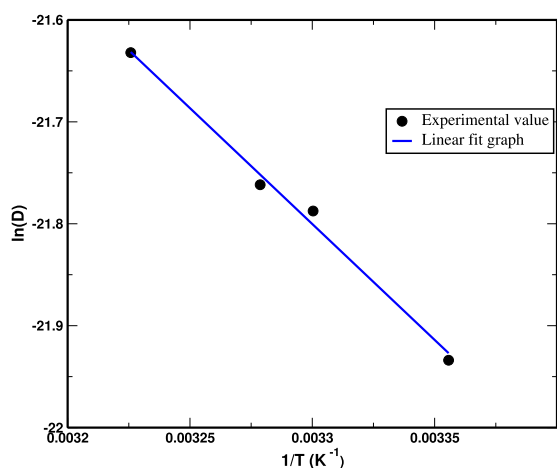
Fig. 8. Arrhenius plot of the simulated values of amoxicillin.
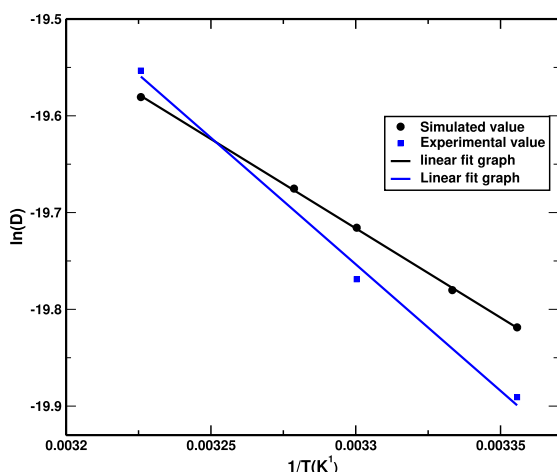
Fig. 9. Arrhenius plot of the simulated and experimental values of water.

ues are presented in Table 5. From the Table 5, it has been observed that the estimated value of D_0 for solvent is higher and smaller than the previously reported experimental and simulated values respectively. The estimated value is in agreement within 20.4% with the experimental value. Also, it has been seen that the calculated value of η for solvent is smaller than experimental value. This is because we used SPC/E model of water which overestimates diffusion coefficient and underestimates the viscosity coefficient [56,58]. The correction on diffusion coefficient estimated under PBC can be calculated in two ways: from intercept of the graphs 12 & 13 and using experimental value of viscosity coefficient.

We have also estimate th correction terms on diffusion coefficient of water and solution and the estimated values of correction term on diffusion coefficient estimated under PBC are presented in Table 6. From the Table 6, we clearly observe the effect of system size on diffusion coefficient estimated under PBC taking finite size simulation box. It is seen from the Table that the correction term decreases with increase in size of simulation box. In case of water, the correction terms estimated from experimental value of viscosity coefficient are smaller in comparison of corresponding value calculated from intercept of the graph. This is because we used SPC/E water model during our simulations and SPC/E model underestimate viscosity coefficient; and as a result, the correction terms becomes larger.

4. Conclusions and concluding remarks

In this work, molecular dynamics study of amoxicillin in two solvents: water and ethanol has been carried out at different temperatures under periodic boundary conditions. During the simulations, OPLS-AA forcefield, and TIP3P and SPC/E water model were used. The free energy of solvation of amoxicillin in water and ethanol have been estimated at 310 K temperature using thermodynamic integration (TI) and Free Energy Perturbation (FEP) based methods. It has been observed that solvation free energy of amoxicillin in water has higher value in comparison to ethanol; and Coulomb has major contribution for solvation of amoxicillin in both solvents. We also compare the electrostatic contribution to solvation free energy of amoxicillin in water with previously reported value of Coulomb contribution to solvation free energy of zwitterionic glutamic acid
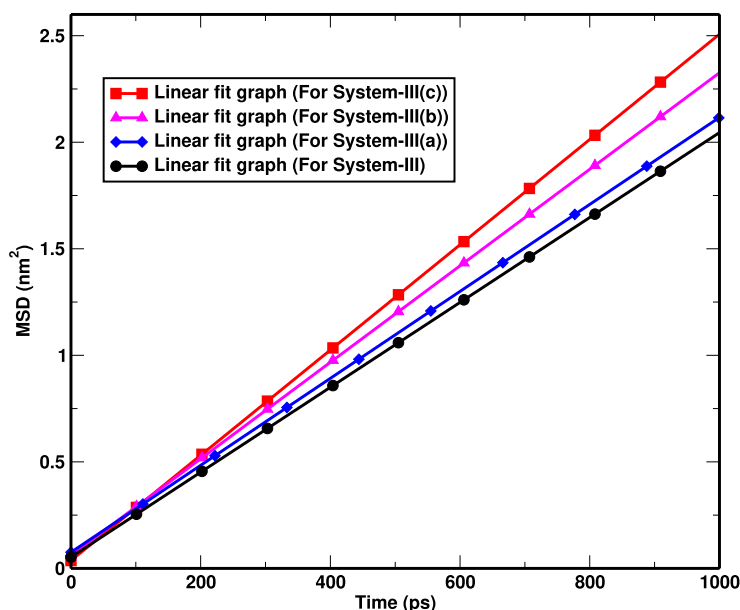
Fig. 10. MSD versus time plot for amoxicillin at 298 K temperature taking four systems of different size.
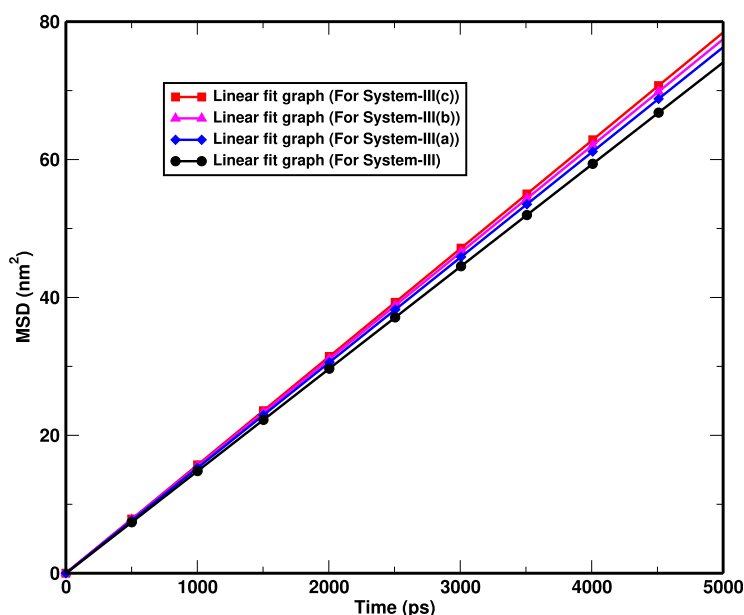
Fig. 11. MSD versus time plot for water at 298 K temperature taking four systems of different size.

Table 4
Simulated values of diffusion coefficients for different size systems at 298 K temperature.

System	L (nm)	Diffusion coefficients (D_{PBC})($10^{-9}m^2s^{-1}$)				
		Self				Binary
		For amoxicillin	For water			
		MSD	MSD	Expt.	% Error	
II	4.05	0.30 ± 0.05	2.49 ± 0.02	2.30 [53]	7.63	0.30
III(a)	4.98	0.34 ± 0.04	2.54 ± 0.01		10.43	0.34
III(b)	5.82	0.38 ± 0.01	2.58 ± 0.01		12.17	0.38
III(c)	6.60	0.41 ± 0.05	2.61 ± 0.01		13.48	0.41
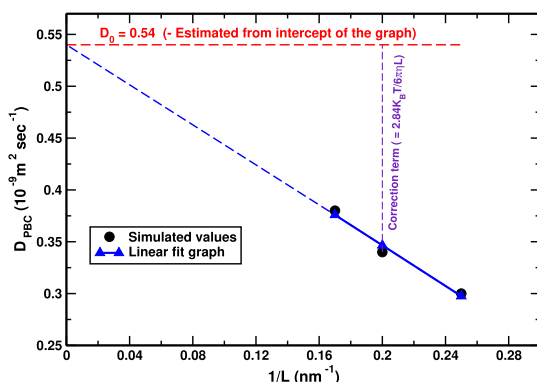
Fig. 12. D_{PBC} versus $(1/L)$ plot for water.
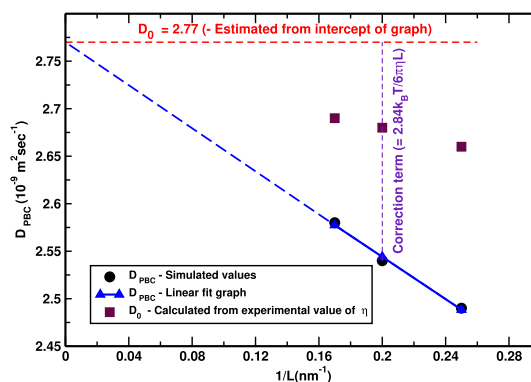
Fig. 13. D_{PBC} versus $(1/L)$ plot for solution of amoxicillin in water.

in aqueous medium; our estimated value of solvation free energy of amoxicillin in water is in good agreement with the solvation free energy of glutamic acid. Number of hydrogen bonds between solute and solvent molecules as well as SASA of solute for initial coupling have been reported at 310 K temperature taking water (i.e. TIP3P and SPC/E models) and ethanol as solvents. The analysis of number of hydrogen bond and SASA also support to higher value of solvation free energy of amoxicillin in water than in ethanol. Also, the self diffusion coefficients of amoxicillin and water as well as their binary diffusion coefficient have been reported at different temperatures under PBC using Einstein's and Darken's relations respectively.

The estimated values of binary diffusion coefficients are almost equal to respective self diffusion coefficient of solute due to infinitely dilute solution. In addition, the effect of system size on diffusion coefficient estimated from simulation under periodic boundary conditions has been studied; and size independent values of diffusion coefficients have been calculated at 298 K temperature. The viscosity coefficient of water as well as solution of amoxicillin in water have been estimated at 298 K temperature. Furthermore, the self diffusion coefficients of amoxicillin and ethanol, and their binary diffusion coefficient have been reported at 298 K temperature using aforementioned methods.

Table 5
System-size independent value of diffusion coefficient (D_0) and shear viscosity (η) of solvent (i. e. water) and solution (i. e. amoxicillin in water) at 298 K temperature taking SPC/E water model.

System		Calculated	Ref.	Expt.	% Error
Water	D_0 $(10^{-9} m^2 s^{-1})$	2.77	2.97 [56]	2.30 [53]	20.4
	η $(10^{-4} Nm^{-2} sec)$	5.57	6.40 [56]	8.90 [57]	37.4
Solution	D_0 $(10^{-9} m^2 s^{-1})$	0.54	–	–	–
	η $(10^{-4} Nm^{-2} sec)$	6.32	–	–	–

Table 6
Correction term on diffusion coefficient (D_{PBC}) estimated from intercept of D versus 1/L graph along with the values calculated from experimental value of η of water at 298 K temperature taking SPC/E water model.

System	L (nm)	Correction term calculated from ($\times 10^{-9} m^2 s^{-1}$)	
		Intercept of graph between D_{PBC} versus 1/L	$\frac{2.84 k_B T}{6\pi \eta^{Expt.} L}$ (Using experimental value of $\eta^{Expt.}$)
Water	4.05	0.28	0.17
	4.98	0.23	0.14
	5.82	0.19	0.11
Solution	4.05	0.24	–
	4.98	0.20	–
	5.82	0.16	–

Data availability

All the data that are used to produce the Figures/Tables in this paper are available from the corresponding author in case reproduction is needed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] C. Danelon et al., Interaction of zwitterionic penicillins with the OmpF channel facilitates their translocation, Biophys. J. 90 (5) (2006) 1617–1627.

[2] S.P. Kaur, R. Rao, S. Nanda, Amoxicillin: a broad spectrum antibiotic, Int. J. Pharm. Pharm. Sci. 3 (3) (2011) 30–37.

[3] K. Bush, P.A. Bradford, β-Lactams and β-lactamase inhibitors: an overview, Cold Spring Harb. Perspect. Med. 6, 8 (2016) a025247.

[4] G. Kapoor, S. Saigal, A. Elongavan, Action and resistance mechanisms of antibiotics: A guide for clinicians, J. Anaesthesiol. Clin. Pharmacol. 33 (3) (2017) 300.

[5] L.B. Rice, Mechanisms of resistance and clinical relevance of resistance to β-lactams, glycopeptides, and fluoroquinolones, Mayo Clin. Proc. 87 (2) (2012) 198–208.

[6] A. Bebu et al., IR, Raman, SERS and DFT study of amoxicillin, J. Mol. Struct. 993 (1–3) (2011) 52–56.

[7] S. Basha et al., On the adsorption/photodegradation of amoxicillin in aqueous solutions by an integrated photocatalytic adsorbent (IPCA): experimental studies and kinetics analysis, J. Mol. Struct. 10 (6) (2011) 1014–1022.

[8] K. Barve, K. Ruparel, Effect of Bioenhancers on Amoxicillin bioavailability, ADMET DMPK 3 (1) (2015) 45–50.

[9] M.A. Bellucci et al., Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments, J. Chem. Phys. 150 (9) (2019) 094107.

[10] G.L. Amidon et al., A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability, Pharm. Res. 12 (3) (1995) 413–420.

[11] A.A. Noyes, W.R. Whitney, The rate of solution of solid substances in their own solutions, J. Am. Chem. Soc. 19 (12) (1897) 930–934.

[12] R.E. Skyner et al., A review of methods for the calculation of solution free energies and the modelling of systems in solution, Phys. Chem. Chem. Phys. 17 (9) (2015) 6174–6191.

[13] S.M. Dizaj et al., Nanosizing of drugs: effect on dissolution rate, Res. Pharm. Sci. 10 (2) (2015) 95.

[14] A.S. Paluch, D.D. Cryan III, E.J. Maginn, Predicting the solubility of the sparingly soluble solids 1, 2, 4, 5-tetramethylbenzene, phenanthrene, and fluorene in various organic solvents by molecular simulation, J. Chem. Eng. Data 56 (4) (2011) 1587–1595.

[15] C.A.S. Bergström, P. Larsson, Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting, Int. J. Pharm. 540 (1–2) (2018) 185–193.

[16] L. Li, T. Totton, D. Frenkel, Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials, J. Chem. Phys. 149 (5) (2018) 054102.

[17] G.A.D.R. Matos et al., Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database, J. Chem. Eng. Data. 62 (5) (2017) 1559–1569.

[18] S. Dasari, B.S. Mallik, Solubility and solvation free energy of a cardiovascular drug, LASSBio-294, in ionic liquids: A computational study, J. Mol. Liq. (2020) 112449.

[19] M.A. Cuendet, M.E. Tuckerman, Alchemical free energy differences in flexible molecules from thermodynamic integration or free energy perturbation combined with driven adiabatic dynamics, J. Chem. Theory Comput. 8 (10) (2012) 3504–3512.

[20] S.B. Dixit, R. Bhasin, E. Rajasekaran, B. Jayaram, Solvation thermodynamics of amino acids assessment of the electrostatic contribution and force-field dependence, J. Chem. Soc. Faraday Trans. 93 (6) (1997) 1105–1113.

[21] D. Frenkel, B. Smit, Understanding Molecular simulation: from algorithms to applications, Academic Press, USA, 2002.

[22] M.P. Allen, D. Tildesley, Computer Simulation of Liquids New York, Oxford University Press, USA, 1991.

[23] M.R. Shirts et al., Best practices in free energy calculations for drug design, J. Chem. Theory Comput. (2012) 425–467.

[24] N. Hansen, W.F. Van Gunsteren, Practical aspects of free-energy calculations: a review, J. Chem. Theory Comput. 10 (7) (2014) 2632–2647.

[25] M. De Vivo et al., Role of molecular dynamics and related methods in drug discovery, J. Med. Chem. 59 (9) (2016) 4035–4061.

[26] Z. Cournia, B. Allen, W. Sherman, Relative binding free energy calculations in drug discovery: recent advances and practical considerations, J. Chem. Inf. Model. 57 (12) (2017) 2911–2937.

[27] G. Malloci et al., A database of force-field parameters, dynamics, and properties of antimicrobial compounds, Molecules 20 (8) (2015) 13997–14021.

[28] M. Tuckerman, Statistical mechanics: theory and molecular simulation, Oxford University Press, 2010.

[29] D.W. Borhani, D.E. Shaw, The future of molecular dynamics simulations in drug discovery, J. Comput. Aid. Mol. Des. 26 (1) (2012) 15–26.

[30] S.P. Khanal, Y.P. Kandel, N.P. Adhikari, Transport properties of zwitterion glycine, diglycine, and triglycine in water, AIP Adv. 90 (2019) 065303.

[31] R.P. Koirala et al., Effect of temperature on transport properties of cysteine in water, AIP Adv. 10 2 (2020) 025122.

[32] J. Crank, The Mathematics of Diffusion, second ed., Oxford University Press, Oxford, 1975.

[33] I.C. Yeh, G. Hummer, System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions, J. Phys. Chem. B 108 (2004) 15873–15879.

[34] S.H. Jamali et al., Finite-size effects of binary mutual diffusion coefficients from molecular dynamics, J. Chem. Theory Comput. 14 (5) (2018) 2667–2677.

[35] B. Dünweg, K. Kremer, Molecular dynamics simulation of a polymer chain in solution, J. Chem. Phys. 99 (9) (1993) 6983–6997.

[36] V.D.S. David et al., GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (16) (2005) 1701–1718.

[37] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (24) (1987) 6269–6271.

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

[39] M.J. Abraham, et al. GROMACS User Manual version 5.1.1, 2015.

[40] R.P. Koirala et al., Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor, J. Phys. Commun. 5 3 (2021) 035010.

[41] B. Hess et al., LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (12) (1997) 1463–1472.

[42] H.J.C. Berendsen et al., Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (8) (1984) 3684–3690.

[43] R. Fiorentini, K. Kremer, R. Potestio, Ligand-protein interactions in lysozyme investigated through a dual-resolution modelProteins: Struct, Funct. Genet. 88 (10) (2020) 1351–1360.

[44] J. Ryckaert, G. Ciccotti, H. Berendsen, JC Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (3) (1977) 327–341.

[45] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescalingJ, Chem. Phys. 126 (1) (2007) 014101.

[46] alchemical-analysis. https://github.com/MobleyLab/alchemical-analysis, (Accessed: 2020-09-01).

[47] D. Mobley, P. Pavel Klimovich, M. Shirts, Guidelines for the analysis of free energy calculations, J. Comput. Aided Mol. Des. 29 (2015) 5.

[48] J.G. Kirwood, Statistical mechanics of fluid mixtures, J. Chem. Phys. 3 (5) (1935) 300–313.

[49] C.H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, J. Comput. Phys. 22 (2) (1976) 245–268.

[50] P.K. Bhattacharyya, W.M. Cort, Analytical Profiles of Drug Substances, vol. 7, Elsevier, 1978, pp. 19–41.

[51] S. Pokharel, N. Pantha, N.P. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, Int. J. Mod. Phys. B 30 (2016) 1650205–1650220.

[52] L.S. Darken, Diffusion, mobility and their interrelation through free energy in binary metallic systems, Trans. AIME 175 (1948) 184–201.

[53] M. Holz, S.R. Heil, A. Sacco, Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements, Phys. Chem. Chem. Phys. 2 (20) (2000) 4740–4742.

[54] R.E. Rathbun, A.L. Babb, Self-diffusion in liquids. iii. temperature dependence in pure liquids1, J. Phys. Chem. 65 (6) (1961) 1072–1074.

[55] A.J. Easteal, W.E. Price, L.A. Woolf, Diaphragm cell for high-temperature diffusion measurements. Tracer diffusion coefficients for water to 363 K.J, Chem. Soc. Faraday Trans. 85 (5) (1989) 1091–1097.

[56] S. Tazi, A. Boțan, M. Salanne, V. Marry, P. Turq, B. Benjamin, J. Phys.: Condens. Matter 24 (28) (2012) 284117.

[57] K.R. Harris, L.A. Woolf, J. Chem. Eng. Data 49 (4) (2004) 1064–1069.

[58] A.P. Markesteijn, R. Hartkamp, S. Luding, J. Westerweel, J. Chem. Phys. 136 (2012) 134104.

Transport properties of zwitterion glycine, diglycine, and triglycine in water

View Online   Export Citation   CrossMark

Shyam P. Khanal,[a] Yadav Prasad Kandel,[a] and Narayan P. Adhikari[a] (iD)

AFFILIATIONS

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu 44600, Nepal

[a]shyamkhanal1989@gmail.com, kandelypg@gmail.com, npadhikari@tucdp.edu.np

ABSTRACT

Diffusion, transport of mass in response to concentration and thermal energy gradient, is an important transport property, vital in material science and life science. In the present work, we have studied about the diffusion of zwitterion glycine, zwitterion diglycine and zwitterion triglycine in SPC/E model of water using classical molecular dynamics. Self and binary diffusion coefficients of aqueous solution of these molecules have been estimated using Einstein's method. Our results agree with experimental data reported in literatures. Temperature dependency of diffusion of glycine in water has been explored using estimated values of self and binary diffusion coefficients at four different temperature. Effects of peptide bond formation in diffusion has been studied using peptide chain composed of up to three monomers of glycine. The system-size dependence of diffusion coefficient has been studied and the shear viscosity of solvent and system has been calculated. Also, the diffusion coefficient of zwitterion glycine in other water model TIP4P/2005 has been estimated. The structure of the system has been analyzed using radial distribution function of different atoms.

I. INTRODUCTION

Amino acids, the fundamental building blocks of proteins, are the organic substances which contain both amine (-NH$_2$) and acidic (-COOH) functional groups. Out of 300 naturally found amino acids, only 20 serve as building blocks of protein.[1] Glycine (gly), a major inhibitory neurotransmitter in the spinal cord and brain stem, is the simplest of all the amino acids and plays significant role in biological systems.[2] Glycylglycine or diglycine (dgl) and glycylglycylglycine or triglycine (tgl) are peptides of glycine having two and three monomer units in chain respectively.[3] The measurement of mass transfer rates of amino acids has significant role in science. Diffusion provides information about the inter and intra atomic/molecular interactions. It also helps in the understanding of the dynamics of amino acids, other bio-molecules, protein and protein folding.[4]

Study of diffusion of biomolecules in aqueous and other medium has attracted a significant number of researchers, both in experiment as well as in simulation. At 298.2 K temperature, the binary diffusion coefficient of aqueous glycine, at infinite dilution, has been reported to be 10.55, 10.64, 10.59 and 10.62 by Longsworth,[5] Lyons and Thomas,[6] Woolf et al.[7] and Ma et al.[8] respectively; all in the unit of 10^{-10} m^2s^{-1}. Similarly, Umecky et al.[4] have studied the temperature dependency of binary diffusion coefficient of glycine and reported the diffusion coefficient to be 9.36 x 10^{-10} m^2s^{-1} at 293.2 K and increase in it by 12.83 x 10^{-10} m^2s^{-1} corresponding to 40 K increase in temperature. Yan et al.[9] have measured densities and viscosities of aqueous solutions of some α-amino acids as a function of temperature and concentration. Also, Dünweg and Kremer[10] have performed MD simulations of a polymer chain in solvent and explained the influence of finite system-size on dynamical properties, and Yeh et al.[11] have studied dependence of diffusion coefficients on system-size and calculate the coefficient viscosity. Furthermore, Changwei et al.[12] have shown that the binary diffusion coefficient of aqueous glycine depends upon concentration and changes from 10.4011 to 9.4258 x 10^{-10} m^2s^{-1} with the change in concentration from 0.1057 to 0.9045 $mol.L^{-1}$. Campo[13] have carried out molecular dynamics (MD) study of hydration and structure of glycine in water. Many other works have been done to study about different properties of glycine, diglycine, and triglycine. Furthermore, Ventura et al.[14] have developed a coarse grained model which explains the interactions between bio-molecules, and

between antibody and substrates. The knowledge about the interactions between the bio-molecules play a key role in different areas of science including biotechnology. Similarly, Tazi et al.[15] have calculated the system-size independent values of diffusion coefficient and shear viscosity of three water models: SPC/E, TIP4P/2005 and Dang-Chang.

Molecular dynamics simulations has become a powerful tool, which can provide a guideline for experimental study. During the simulations, the experimental environment could be mimicked by modeling different interactions between the atoms and molecules. Nowadays, it is routinely used to understand the structure and dynamics of bio-molecules.[16] Inspite of having many experimental works been done to study about the diffusion of glycine in aqueous medium, it is equally important to study about the diffusion phenomenon of the system by using molecular dynamics simulations. To the best of our knowledge, no realistic molecular dynamics simulations has been carried out to study about transport properties of this system.

In the present work, self and binary diffusion coefficient of aqueous solution of zwitterion glycine has been estimated at different temperatures from molecular dynamics simulation, as per mole fraction matching to the experimental values reported by Umecky et al.,[4] and Longsworth.[5] Diffusion coefficients of aqueous solution of zwitterion diglycine and zwitterion triglycine have also been estimated. The structure of systems and its effect on diffusion have been discussed. Furthermore, as the diffusion coefficient depends upon size of system due to screening effect of hydrodynamics flow under periodic boundary conditions (PBC),[10,11] the system-size dependence of diffusion coefficient has also been studied. And, the shear viscosity of solvent as well as solution has been calculated. All the simulations have been carried out using SPC/E water model. Likewise, self diffusion coefficients of zwitterion glycine and water have been calculated using other water model TIP4P/2005,[17] for the dynamical properties significantly depends upon water models taken during simulations.[15] All the simulations are carried out using GROMACS software package.

In this paper, we have discussed theoretical background of the present work in section II, computational details in section III, results in section IV, and conclusions in section V.

II. THEORY

Diffusion is the dynamic property by which particles are transported from one region to other due to their random motion. It takes place in an account of concentration inhomogeneity and thermal gradient. Diffusion is an essential function in living organisms and has many applications in modern material science and technology.

Diffusion of particle in homogeneous medium with no chemical concentration gradient is called the self diffusion coefficient. Rate of the self-diffusion is measured in terms of self-diffusion coefficient. Under the assumption of isotropy of medium, if $r(t) - r(0)$ is the change of position of diffusing particle in time t, then the macroscopic transport property i.e. self diffusion coefficient can be related to microscopic property, mean squared displacement of material, by using Einstein's relation, as:[18]

$$D = \lim_{t \to \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6\,t} \qquad (1)$$

Here, $\langle \ldots \rangle$ represents the ensemble average of quantity inside the angled bracket, which is the square of displacement. Thus, the self diffusion coefficient of any species is one sixth of slope of mean squared displacement plotted as a function of time.

Binary or mutual diffusion in binary mixture is the diffusion of constituent species in that mixture and the corresponding diffusion coefficient is called binary or mutual diffusion coefficient. If self diffusion coefficients of two individual species, A and B are D_A and D_B, with mole fraction N_A and N_B respectively, the binary diffusion coefficient D_{AB} of these species, according to Darken's phenomenological relation, is:[19]

$$D_{AB} = N_B D_A + N_A D_B \qquad (2)$$

The simulation is carried out under periodic boundary conditions. Under periodic boundary conditions, the dynamical properties of system also depends upon size of simulation box due to effect of long range hydrodynamics interaction. So, it is necessary to account the system-size dependence of diffusion coefficient. The system-size independent value of diffusion coefficient (i.e. the diffusion coefficient of particles in an infinite system size) can be estimated by:[10,11]

$$D_0 = D_{PBC} + \frac{2.84\,k_B T}{6\pi\eta L} \qquad (3)$$

where D_0 is the system-size independent value of diffusion coefficient, D_{PBC} the simulated value of diffusion coefficient in cubic box of size L under periodic boundary condition (PBC), k_B the Boltzmann constant, T the absolute temperature of system and η the shear viscosity of solvent.

$$or,\ D_{PBC} = D_0 - \frac{2.84\,k_B T}{6\pi\eta L} \qquad (4)$$

Thus, if values of D_{PBC} are known for systems with different sizes, D_0 can be estimated from the intercept of the graph and the η from its slope plotted between D_{PBC} versus $1/L$.

III. COMPUTATIONAL DETAILS

In the present work, zwitterion form of glycine, diglycine and triglycine were taken and modeled in GROMOS53A6 force field platform.[20,21] Specific bonds, and bond angles were taken in g96 format to configure the topology of the molecules. Different proper dihedrals were defined to prevent rotation around a bond and improper dihedrals were defined to confine four atoms in plane or tetrahedral configuration. We took CH_2 as united atom in which position of united atom is the position of the heaviest carbon(C) atom. Fig. 1 shows the models of zwitterion glycine, diglycine and triglycine molecules respectively. To estimate the diffusion coefficient of zwitterion glycine in water and to check its dependency in temperature, two zwitterion glycine molecules were dissolved in 11,112 water molecules in system-I. It was simulated at four different temperature: 293.2 K, 303.2 K, 313.2 K, and 333.2 K. In order to estimate variation of diffusion coefficient with increase in monomer units in peptide chain, two zwitterion glycine, two zwitterion diglycine, and two zwitterion triglycine molecules were separately dissolved in 1,385 water molecules in system-II, system-III, and system-IV respectively; and diffusion coefficients were estimated at temperature 298.2 K.[22] SPC/E model of water[23] was taken
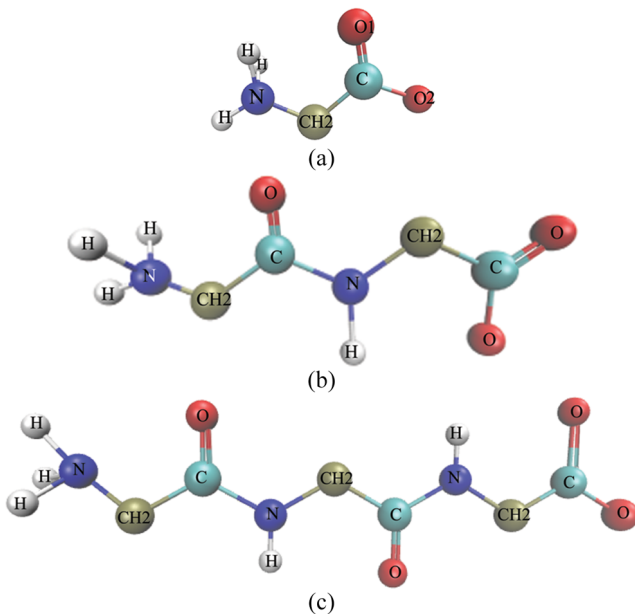
FIG. 1. Zwitterion glycine (a), diglycine (b) and triglycine (c) with CH2 as united atom centered at position of atom C.

as solvent in system-I, II, III and IV. Furthermore, to analyse the system-size dependence of diffusion coefficient and to calculate shear viscosity, we performed simulations with SPC/E water model taking cubic box of different sizes (i.e. systems with different number of solute and solvent molecules). In addition, in system-V, two zwitterion glycine molecules were dissolved in 1385 water molecules with TIP4P/2005[17] water model at 298.2 K to check the dependence of diffusion coefficients on water models used during simulations. All the simulations were carried out at pressure of one atmosphere and under periodic boundary conditions using cubic simulation box. The number of molecules were chosen to match mole fraction of experimentally reported data.[4,5]

During the simulation, if the simulation box contains overlapped particles or particle with bad contact, molecular dynamics may explode and never bring system to equilibrium.[18] To avoid bad contacts and bring the system in minimum potential state, energy minimization was carried out for each of the systems using Steepest-descent method.[24]

Dynamical properties like diffusion, thermal conduction, etc. depend strongly upon the temperature, and pressure of system.[25] To bring system under study at a condition that best mimics experimental environment, the systems ware equilibrated for 200 ns with time step of 0.002 ps using NPT (Isothermal-isobaric) ensemble. LINCS and Leap-frog algorithms[24] were used to constrain all bonds and to integrate equation of motion respectively. The initial velocities were generated using Maxwell-Boltzmann distribution. The velocity-rescaling thermostat and Berendsen barostat[24] with a coupling time of 0.01 ps and 0.8 ps respectively were used to maintain constant temperature and pressure. Also, the value of isothermal compressibility was taken to be 4.5×10^{-5} bar and cut off distance for both of the non bonded interactions- coulomb and LJ was 1.0 nm.

Particle Mesh Ewald (PME) was used for long-range coulomb interaction. Density of the equilibrated system was tallied with experimentally reported values to ensure proper equilibration and to be sure that the force field parameters are suitable for the system under consideration. In no case, density of system after equilibration differed from experimental values by 1.5 percent.

After equilibration of each system, production run was carried out to estimate equilibrium properties of the system taking NVT ensemble. During production run, velocity-rescaling thermostat with coupling time of 0.01 ps was used and initial velocities were taken from final step of equilibration run. Each production run was carried out for 100 ns with time step of 0.002 ps. Structural property of system after production run was explored using radial distribution function (RDF).

IV. RESULTS AND DISCUSSION

We have calculated the self diffusion coefficients of glycine and water, and their binary diffusion coefficient at four different temperature using system-I. Diffusion coefficients of aqueous glycine, diglycine, and triglycine have been calculated at temperature 298.2 K. These diffusion coefficients have been calculated by Einstein method,[18] where self diffusion coefficient of any species in three dimensional isotropic medium is one sixth of slope of mean squared displacement (MSD) versus time plot as in equation (1). Structure of systems have been studied using the radial distribution function (RDF), which gives preferred position of one particle around other particles.

A. Mean squared displacement (MSD)

Mean squared displacement (MSD) as a function of time is used to calculate self diffusion coefficient in Einstein's method. As statistics is better at beginning than ending region, certain portion has been taken and linear fit of MSD of different molecules at different temperature are plotted as a function of time for 5 ns. Fig. 2 is MSD plot of glycine at four different temperature. The MSD is becoming steeper with increase in temperature. This shows that rate of diffusion is increasing with the increase in temperature. Fig. 3 is MSD plot of water at four different temperature and shows
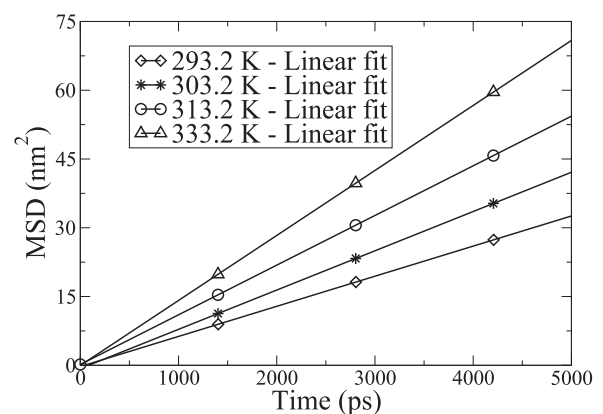
FIG. 2. MSD plot of glycine at four different temperature.
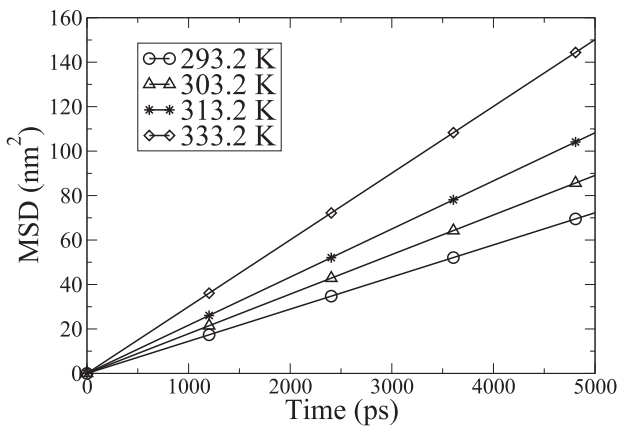
FIG. 3. MSD plot of water at four different temperature.
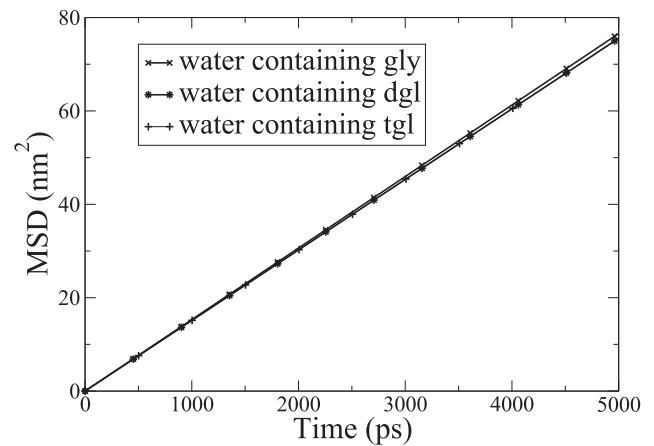
FIG. 5. MSD plot of water with different solute at temperature 298.2 K.
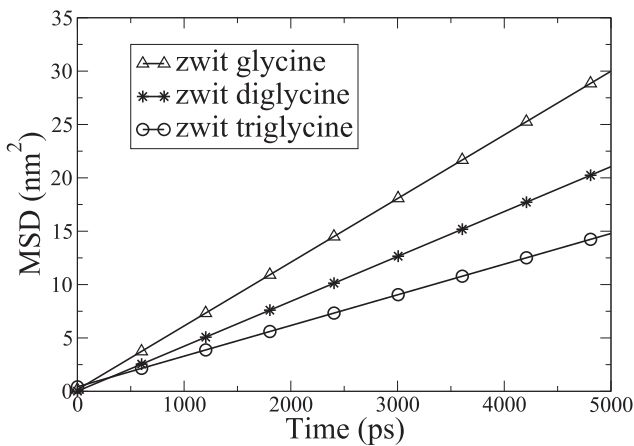
FIG. 4. MSD plot of glycine, diglycine and triglycine at temperature 298.2 K.

similar behaviour of MSD plot of glycine in Fig. 2. MSD plot of glycine, diglycine and triglycine at temperature 298.2 K is shown in Fig. 4. The MSD line is getting less steep with increase in molecule size indicating smaller rate of diffusion of larger molecule than that

of smaller molecule at same temperature. MSD plot of water at 298.2 K containing three different solutes is shown in Fig. 5. The MSD curve is steeper when lighter solute molecules are present. It indicates that water diffuses faster at given temperature when smaller solute molecules are present. Figs. 4 and 5 show the effect of chain length of solute on MSD and diffusion.

B. Self diffusion coefficient

Table I presents the estimated values of self diffusion coefficient of different molecules at different temperature. The self diffusion coefficient of zwitterion glycine and water is increasing with increase in temperature. This happens because random velocity increases with the increase in temperature.

Self diffusion coefficients of zwitterion glycine, zwitterion diglycine and zwitterion triglycine each dissolved separately in 1,385 water molecules at temperature 298.2 K and pressure 1 bar is presented in the same table. The estimated self diffusion coefficient of these molecules have decreased with the increase in monomer unit in chain. It could be attributed to size of the molecules. At given temperature, molecule of larger molecular weight attains smaller velocity than lighter molecule. Therefore, in the same environment

TABLE I. Estimated values of self diffusion coefficient of different molecules at one atm pressure and different temperature.

Temperature (K)	Solute	D_{self}^{solute} $(10^{-9}m^2.s^{-1})$	Water $(10^{-9}m^2.s^{-1})$	
			D_{self}^{est}	D_{self}^{exp}
293.2	Glycine	1.09	2.48	2.02[26]
298.2	Glycine	0.99	2.55	2.30[26]
	Diglycine	0.70	2.52	2.30[26]
	Triglycine	0.48	2.51	2.30[26]
303.2	Glycine	1.43	2.97	2.60[26]
313.2	Glycine	1.81	3.61	3.22[26]
333.2	Glycine	2.36	5.00	4.72[27]
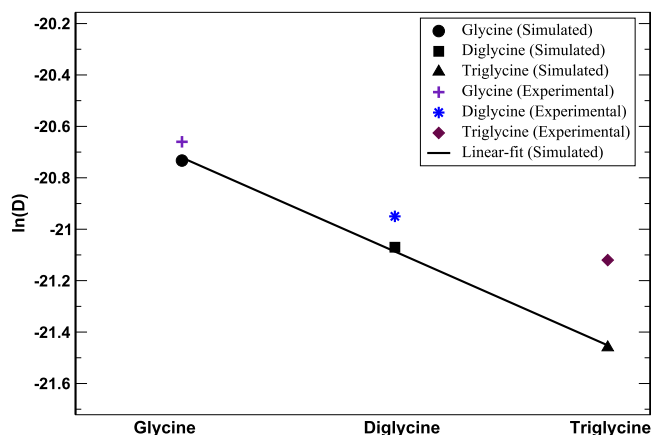
FIG. 6. Binary diffusion coefficient of zwitterion glycine, diglycine and triglycine at 298.2 K temperature.

and physical conditions, large molecules diffuse slowly than small molecules. As the chain length and molecular weight of molecule increases from zwitterion glycine, zwitterion diglycine to zwitterion triglycine, the rate of diffusion decreases. The variation of self diffusion coefficient with size of glycine is shown in Fig. 6. From the figure, it is noticed that self diffusion coefficient decreases linearly with increase in monomer unit in chain.

Also from Table I, we observe that at temperature 298.2 K, self diffusion coefficient of glycine appears less than that at 293.2 K. It could be because of higher concentration of glycine at 298.2 K compared to that at 293.2 K. The value of self diffusion coefficient of water has increased with increase in temperature. Estimated values are slightly greater than experimental values reported in literatures. The reason is SPC/E water model was used during the simulations which overestimate the diffusion coefficient.[15] The errors has been reported to be less than 1% and 0.2% respectively by Holz[26] and Eastel et al.[27] respectively. Further, the self diffusion coefficient could have been affected by presence of solute. The self diffusion coefficient of water at 298.2 K is slightly smaller in the presence of larger molecules because of increased hindrance in motion of water molecules with increased solute size.

C. Binary diffusion coefficient

Binary diffusion coefficient has been calculated for zwitterion glycine-water mixture at four different temperature, and mixture of zwitterion glycine - water, zwitterion diglycine - water and zwitterion triglycine - water at temperature 298.2 K using Darken's relation (2). Binary diffusion coefficients of different pairs of molecules at different temperature are presented in Table II and are compared with experimental values reported in literature.[4,5] D_{binary}^{est} is binary diffusion coefficient estimated in present work and D_{binary}^{exp} is experimental value of binary diffusion coefficient reported in literature. With the increase in temperature, thermal agitation of molecules increases, which boosts the diffusion. This means diffusion coefficient should be greater at higher temperature.

From Table II, it is noticed that the binary diffusion coefficient of glycine in water is increasing with increase in temperature,

TABLE II. Binary diffusion coefficient of different pairs of molecules at different temperatures.

Solvent	Solute	Temperature (K)	Diffusion coeff. $(10^{-9}m^2.s^{-1})$		% Error
			D_{binary}^{est}	D_{binary}^{exp}	
Water	Glycine	293.2	1.10	0.94[4]	17.02
	Glycine	298.2	0.99	1.06[5]	6.60
	Diglycine	298.2	0.70	0.79[5]	11.39
	Triglycine	298.2	0.48	0.67[5]	28.36
	Glycine	303.2	1.43	1.22[4]	17.21
	Glycine	313.2	1.81	1.50[4]	20.67
	Glycine	333.2	2.36	2.22[4]	6.31

like self diffusion coefficient. Binary diffusion coefficient of glycine, diglycine and triglycine in water is also following the decreasing nature of self diffusion coefficient with increase in molecular weight. Mass of molecules increases from glycine to diglycine and then to triglycine; and it is seen that the value of binary diffusion coefficient decreases sharply with increase in number of monomers in peptide chain as in Fig. 6. The error have been reported to be less than 2% Umecky et al.[4] The standard error in the estimation of self as well as binary diffusion coefficients are very small and insignificant compared to the estimated values, where symmetric round off errors in each estimation of diffusion coefficient is of the order 10^{-11}. The value of D_{binary}^{ext} for triglycine has deviated from experimentally reported value by about 28 percent. It is possible that the united atom modeling and corresponding force field parameters that we used might not be adequate for large molecule like triglycine. Moreover, we observe that the effects of united atom model on calculation of diffusion coefficients increases as the number of atom in chain increases.

D. Effect of temperature on diffusion

Diffusion is transport of mass due to concentration and thermal gradient. It strongly depends upon temperature. We have checked the temperature dependency of diffusion coefficient of glycine, water and their binary mixture using Arrhenius formula:[28,29]

$$D = D_e \exp\left(-\frac{E_a}{N_A k_B T}\right) \qquad (5)$$

where D_e is frequency factor (i.e. pre-exponential factor), E_a the diffusion activation energy, T the absolute temperature, N_A Avogadro number, and k_B is the Boltzmann constant. The activation energy of diffusion process corresponds to the slope of Arrhenius plot, which is plot between $\ln(D)$ and reciprocal of absolute temperature.

Fig. 7 shows the Arrhenius diagram of self diffusion coefficient of zwitterion glycine, self diffusion coefficient of water, and their binary diffusion coefficient. From the Figure, it is observed that diffusion coefficients increase with the increase in temperature. In the plot, points obtained from simulations have aligned around straight line. Hence, we can say that diffusion coefficient of these molecules at different temperature follows Arrhenius behaviour. As the self diffusion coefficient of glycine and binary diffusion
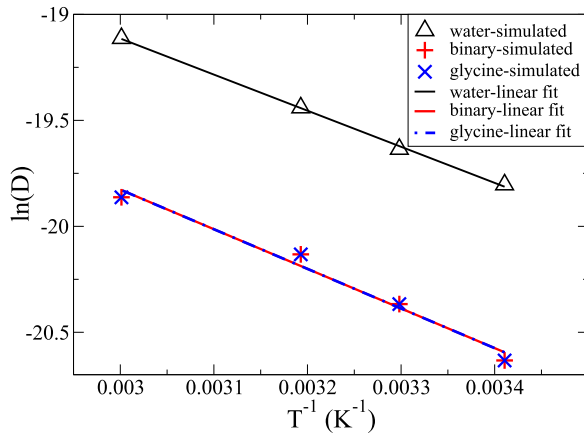
FIG. 7. Arrhenius diagram of diffusion coefficient of water, zwitterion glycine, and their binary mixture.

TABLE III. Activation energy of diffusion.

System	Activation energy (E_a) (kJ.mol^{-1})		
	Simulated	Experimental	% Error
Water	14.14	17.41[16]	18.78
Binary mixture-simulated	15.52	18.47	15.97

coefficient of aqueous glycine are nearly equal, Arrhenius plots have overlapped. This is because of very small mole fraction of glycine. The activation energies are presented in Table III. The experimental values of activation energy of the binary mixture is obtained by fitting the experimental data of diffusion coefficient. From the table, it is seen that the simulated values of activation energy for both glycine and binary mixture agree with previously reported values within 19%.

E. Effect of system-size

The diffusion coefficient also depends on system-size due to screening effect of hydrodynamics under periodic boundary

conditions.[10,11,30] After the calculation of system-size dependent values of diffusion coefficients under periodic boundary conditions, we can calculate the system-size independent value of diffusion coefficient and also shear viscosity. So, we have studied the dependence of diffusion coefficient on size of system from simulations of the systems: (a) 3 glycine molecules and 2078 water molecules in cubic simulation box of size 3.98 nm [System-a], (b) 5 glycine molecules and 3464 water molecules in cubic simulation box of size 4.72 nm [System-b] and (c) 7 glycine molecules and 4849 water molecules in cubic simulation box of size 5.28 nm [System-c]. The simulations have been carried out under periodic boundary at 298.2 K temperature with SPC/E water model using same thermostat and barostat as in the system of 1385 water molecules and 2 glycine molecules. The calculated values of binary diffusion coefficients for the systems along with simulated values of self diffusion coefficients are presented in Table IV.

From the Table IV, it is noticed that the diffusion coefficients increase with the increase in the size of simulation box, which may be due to the effect of long-range hydrodynamics interactions in simulations under periodic boundary conditions.

The system-size independent value of diffusion coefficient D_0 is estimated from the graph between simulated values of diffusion coefficients D_{PBC} with periodic boundary conditions versus reciprocal of size of simulation box (1/L).

Figs. 8 and 9 show the graph between D_{PBC} versus (1/L) for water taking SPC/E model and solution of glycine and water respectively. Then, we can estimate the system-size independent values of diffusion coefficient (D_0) from intercept and also values of coefficient of viscosity (η) from slopes of the graphs. The estimated values of D_0 and η are presented in the Table V. From the Table, it is seen that the calculated values of both D_0 and η for water are smaller than the previously reported simulated values. Also, from the Table, we observe that the calculated values of D_0 are greater than previously reported experimental values but values of η are smaller than the experimental values for both water as well as solution.[32] This may be due to the fact that SPC/E water model overestimate diffusion coefficient but underestimate viscosity and TIP4P/2005 water model is better for viscosity measurement.[15,31]

Furthermore, the calculation of dynamical properties like diffusion coefficient, viscosity etc., from simulation depends not only on simulation parameters like cut-off radius, temperature etc, but also on water models taken in simulations. As the different models of water taken in simulations have distinct features on the basis of their

TABLE IV. Simulated values of diffusion coefficients for different size systems at 298.2 K temperature.

System	L (nm)	Diffusion coefficients (D_{PBC}) ($10^{-9}m^2\,s^{-1}$)					
		Self			Binary		
		For glycine	For water				
		MSD	MSD	Expe.	Calculated	Expe.	% Error
System-a	3.98	1.13	2.57		1.13		6.60
System-b	4.72	1.16	2.62	2.30[26]	1.16	1.06[5]	9.43
System-c	5.28	1.17	2.64		1.17		10.94
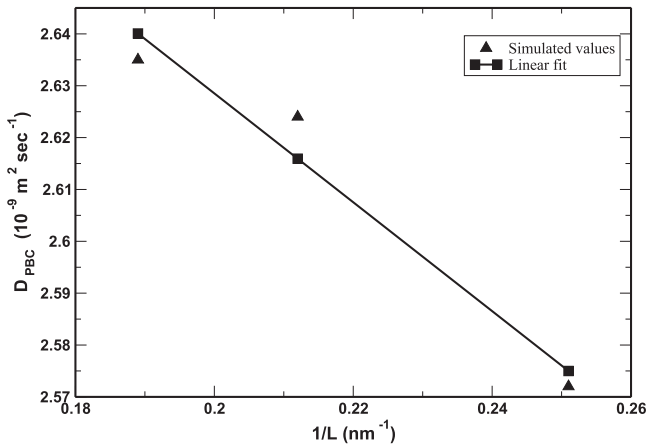
FIG. 8. D_{PBC} versus $(1/L)$ plot for water.
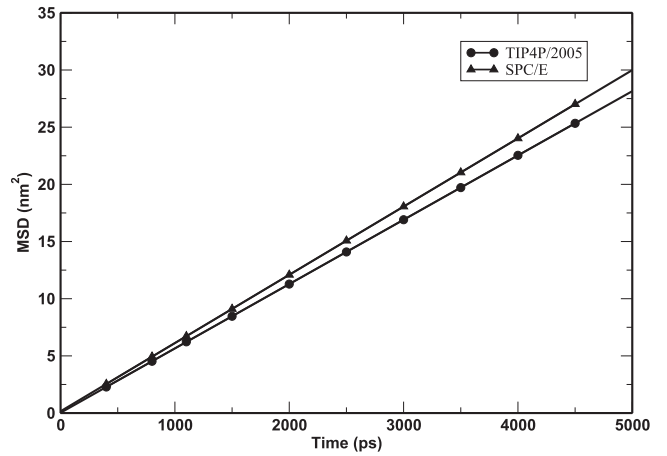
FIG. 10. MSD plot of glycine using different water models at temperature 298.2 K.
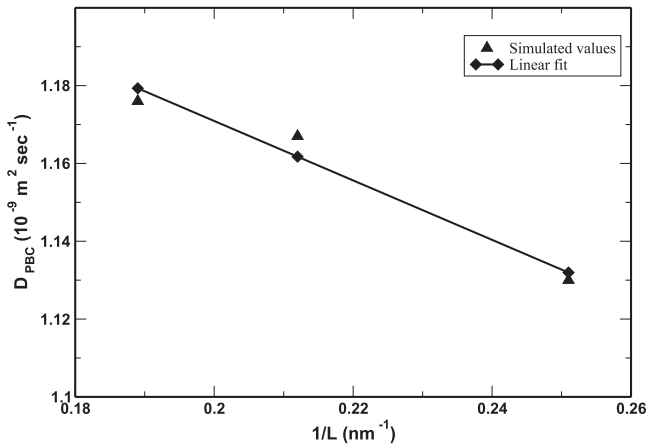
FIG. 9. D_{PBC} versus $(1/L)$ plot for system of glycine in water.
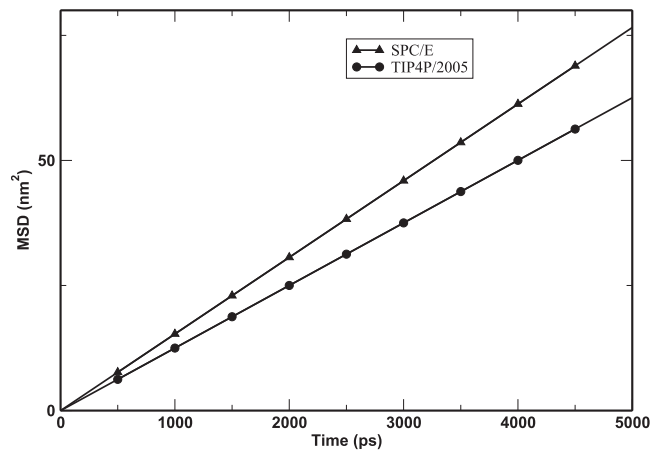
FIG. 11. MSD plot of water for different water models at temperature 298.2 K.

number of interaction sites, flexibility or rigidity etc., they matter the performance of particular properties of the system under consideration.[15] Thus, we have also performed the simulations of glycine in water at temperature 298.2 K using TIP4P/2005 water model for system-V. Figs. 10 and 11 are MSD plot for glycine and water respectively at temperature 298.2 K using SPC/E and TIP4P/2005 water

models. From the Figs. 10 and 11, it is seen that the diffusion of glycine as well as water also depends upon the water models used in simulations.

The simulated values of self diffusion coefficient as well as their previously reported values for different water models are presented in Table VI and are plotted in Fig. 12. From the table, it is noticed

TABLE V. System-size independent value of diffusion coefficient (D_0) and shear viscosity (η) at 298.2 K temperature taking SPC/E water model.

System	D_0 ($10^{-9} m^2 s^{-1}$)				η ($10^{-4} Nm^{-2} sec$)			
	Calculated	Ref.	Expe.	% Error	Calculated	Ref.	Expe.	% Error
Water (SPC/E)	2.83	2.97[15]	2.30[26]	23.04	5.91	6.40[15]	8.90[32]	33.60
Solution (Glycine and water)	1.32	–	1.06[5]	24.53	8.10	–	8.99[9]	9.88

TABLE VI. Simulated values of self diffusion coefficient of glycine and water at temperatures 298.2 K using different water models.

| Water models | Self diffusion coefficients ($10^{-9} m^2 s^{-1}$) | | | | |
| | For glycine | For water | | | |
	MSD	MSD	Ref.	Expe.	% Error
SPC/E	0.99	2.55	2.49[33]	2.30[26]	10.87
TIP4P/2005	0.94	2.08	2.08[17]		9.56

that the SPC/E water models over estimate the value of diffusion coefficient but TIP4P/2005 underestimate.[15,33] Also, it is observed that the estimated value of self diffusion coefficient for water agrees with experimental value[26] for TIP4P/2005 water model within 10% and within 11% for SPC/E model. It is also noticed that the both SPC/E and TIP4P/2005 water models are better to estimate the diffusion coefficients.

F. Radial distribution function (RDF)

The radial distribution function (RDF), which gives the probability of finding a pair of atoms located at distance 'r', has been used to analyse the structural properties of system. The RDF that provides idea about the distribution of molecules around another molecule as reference shows an oscillation upto certain distance and becomes unity for liquids. It gives the preferred position of one particle with respect to other particle. The region from $r = 0$ up to which the RDF becomes zero is known as excluded region. For isotropic system, it is only the function of distance between particles.[18,34]

Fig. 13 is the RDF of oxygen atom of water (OW) in reference to oxygen atom of other water molecule (OW). Value of r in this plot gives the preferred distance of oxygen atom in a water molecule around other water molecules and value of g(r) gives relative probability of finding OW. Only three peaks are observable and beyond that value of g(r) is unity which means there is no correlation between oxygen in water molecules beyond the position of third peak. van der Waals radius of oxygen atom is

FIG. 13. Radial distribution function of oxygen of water (OW) in reference of oxygen of water (OW).

$2^{1/6}\sigma = 0.355$ nm.[24] In the plot, first peak positions are 0.274 nm, 0.275 nm, 0.276 nm, and 0.276 nm and corresponding peak values are 3.151, 3.053, 2.982, and 2.835 at temperature 293.2 K, 303.2 K, 313.2 K, and 333.3 K respectively. It is observed that the values of excluded region (ER) and that the first peak position (FPP) are less than van der Waals radius which indicates that the van der Waals potential as well as other potentials contribute to the structural properties of the system. Also, the hydrogen and oxygen atom in SPC/E model of water has partial positive and negative charges which introduces Coulomb interaction. Thus, the Coulomb interaction is responsible for first peak position being smaller than the van der Waals radius. Although, Lennard-Jones (LJ) plus Coulomb potentials have major contribution to structural properties, many body effects also contributes.[16]

Fig. 14 is the Radial distribution function of oxygen (OW) in water molecule in reference of oxygen (O1) in zwitterion glycine.

FIG. 12. Self-diffusion coefficients of different water models at 298.2 K.
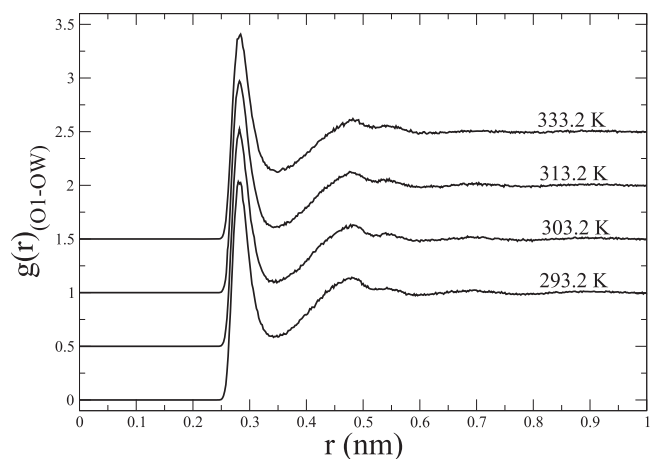
FIG. 14. Radial distribution function of oxygen (OW) in water molecule in reference of oxygen (O1) in zwitterion glycine.
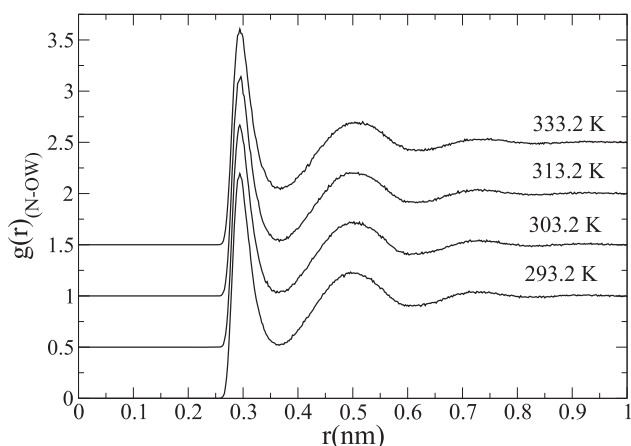
FIG. 15. Radial distribution function of oxygen of water (OW) in reference of nitrogen (N) in zwitterion glycine.

Fig. 15 is the RDF of oxygen (OW) in water molecule in reference of oxygen (O1) in glycine. It gives average distribution of OW and hence the water molecules around O1 in zwitterion glycine. The first peak positions of OW around O1 at temperature 293.2 K, 303.2 K, 313.2 K, and 333.2 K are 0.280 nm, 0.282 nm, 0.282 nm, and 0.284 nm while the corresponding peak values are 2.034, 2.024, 1.971, and 1.907 respectively. Fig. 15 is the RDF of OW in reference of nitrogen (N) in glycine. It gives the distribution of water molecules around $NH3^+$ terminal of zwitterion glycine. The first peak positions of OW around N at temperature 293.2 K, 303.2 K, 313.2 K and 333.2 K are 0.294 nm, 0.294 nm, 0.296 nm, and 0.298 nm while the corresponding peak values are 2.191, 2.161, 2.128, and 2.062 respectively.

In all three RDF plots, with increase in temperature, the first peak positions have moved farther while heights of first peak have decreased and their width have increased. These phenomenon indicate the increased random motion with temperature. Further, wide RDF at higher temperature means more space in between the molecules. This allows the molecules to move more freely, that results the increase in diffusion coefficient. Therefore, the increase in width of RDF means increase in diffusion.

V. CONCLUSIONS AND CONCLUDING REMARKS

We carried out realistic classical molecular dynamics simulation of glycine, diglycine, and triglycine in SPC/E water using GROMACS software package, where concentration of solutes were same as previously reported in experiments. The solutions used in simulation were very dilute, and hence calculated values of binary diffusion coefficients were nearly equal to self diffusion coefficients of molecule of extremely small mole fraction. The effect of increase in chain length from zwitterion glycine to triglycine on diffusion coefficient was studied and it was observed that diffusion coefficients decrease with increase in chain length. Also, the effect of temperature on diffusion coefficient of glycine, SPC/E water and their binary mixture was studied and we found that diffusion coefficients follow Arrhenius behaviour. The estimated values were in good agreement with previously reported values. Furthermore, we performed

the simulation of glycine in SPC/E water taking simulation box of different sizes under periodic boundary conditions to study the size dependence of diffusion coefficient, and hence the system-size independent values of diffusion coefficient as well as shear viscosity of both solvent and solution were estimated. We also performed simulations of glycine in water for other TIP4P/2005 water model and it was noticed that TIP4P/2005 water model also suitable for diffusion calculation. In addition, the equilibrium structural properties of the system was studied using radial distribution function (RDF).

In future, we intend to extend our work to study the transport properties of polymer of glycine in other water models, and transport properties of other amino acids.

REFERENCES

[1] G. Wu, Amino Acids **37**(1), 1–17 (2009).
[2] R. Y. Gundersen, P. Vaagenes, T. Breivik, F. Fonnum, and P. K. Opstad, Acta Anaesthesiologica Scandinavica **49**(8), 1108–1116 (2005).
[3] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 6th ed. (Worth Publishers, New York, 2000).
[4] T. Umecky, T. Kuga, and T. Funazukur, J. Chemical Eng. Data **51**, 1705–1710 (2006).
[5] L. G. Longsworth, JACS **75**, 5705–5709 (1953).
[6] M. S. Lyons and J. V. Thomas, Am. Chem. Soc. **72**, 4506–4511 (1950).
[7] L. A. Woolf, D. G. Miller, and L. J. Gosting, J. Am. Chem. Soc. **84**, 317–331 (1962).
[8] Y. Ma, C. Zhu, P. Ma, and K. T. Yu, J. Chem. Eng. Data **50**, 1192–1196 (2005).
[9] Z. Yan, J. Wang, W. Liu, and J. Lu, Thermochimica Acta **334**, 17–27 (1999).
[10] B. Dünweg and K. Kremer, J. Phys. Chem. B **99**, 6983–6997 (1993).
[11] I. C. Yeh and G. Hummer, J. Phys. Chem. B **108**, 15873–15879 (2004).
[12] Z. Changwei, L. Jiding, M. Peisbeng, and X. Shuqian, Chinese J. Chem. Eng. **15**, 285–290 (2005).
[13] M. G. Campo, J. Chem. Phys. **125**, 114511–114518 (2006).
[14] B. D. Ventura, A. Ambrosio, A. Fierro, R. Funari, F. Gesuele, P. Maddalena, D. Mayer, M. P. Ciamarra, R. Velotta, and C. Altucci, ACS Appl. Mater. & Interfaces **8**(33), 21762 (2016).
[15] S. Tazi, A. Boţan, M. Salanne, V. Marry, P. Turq, and B. Benjamin, J. Phys.: Condens. Matter **24**(28), 284117 (2012).
[16] S. Pokharel, N. Pantha, and N. P. Adhikari, International Journal of Modern Physics B **30**, 1650205 (2016).
[17] J. L. F. Abascal and C. Vega, The Journal of Chemical Physics **123**, 234505 (2005).
[18] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford University Press, USA, 1989).
[19] L. S. Darken, Trans. AMIE **175**, 184–201 (1948).
[20] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. V. Gunsteren, J. Comput. Chem. **25**, 1656–1676 (2004).
[21] C. Oostenbrink, T. A. Soares, N. F. A. Van der Vegt, and W. F. V. Gunsteren, Eur. Biophys. J. **34**, 273–284 (2005).
[22] J. Danielsson, J. Jarvet, P. Damberg, and A. Gräslun, Mag. Reso. Chem. **40**, S89 (2002).
[23] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269–6271 (1987).

[24]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).

[25]S. K. Thapa and N. P. Adhikari, Int. J. Mod. Phys. B **27**, 1350023–1350040 (2013).

[26]M. Holz, S. Manfred, and A. Sacco, Phys. Chem. Chem. Phys. **2**, 4740–4742 (2000).

[27]A. J. Eastel, W. E. Price, and L. A. Woolf, J. Chem. Soc., Faraday Trans. I **85**(5), 1091–1097 (1989).

[28]J. Crank, *The Mathematics of Diffusion*, 2nd ed. (Oxford University Press, Oxford, 1975).

[29]H. Mehrer, *Springer series in solid state science 155, diffusion in solids, fundamentals, methods, materials, diffusion-controlled processes* (Springer, Berlin, 2007).

[30]S. V. Koniakhin, I. E. Eliseev, I. N. Terterov, A. V. Shvidchenko, E. D. Eidelman, and M. V. Dubina, Microfluidics and Nanofluidics **18**(5-6), 1189–1194 (2015).

[31] A. P. Markesteijn, R. Hartkamp, S. Luding, and J. Westerweel, The Journal of Chemical Physics **136**, 134104 (2012).

[32]K. R. Harris and L. A. Woolf, J. Chem. Eng. Data **49**(4), 1064–1069 (2004).

[33]M. R. Shirts and V. S. Pande, The Journal of Chemical Physics **122**, 134508 (2005).

[34]D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).

Effect of temperature on transport properties of cysteine in water SCI F

Rajendra Prasad Koirala,[a] Hem Prasad Bhusal,[b] Shyam P. Khanal,[c] and Narayan Prasad Adhikari[d] (iD)

AFFILIATIONS

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

[a]rpkoirala@tucdp.edu.np
[b]hemprasadbhusal@gmail.com
[c]shyamkhanal1989@gmail.com
[d]Author to whom correspondence should be addressed: npadhikari@tucdp.edu.np

ABSTRACT

Molecular dynamics simulations have been performed to study the transport properties of the dilute solution of cysteine in water at different temperatures. Structural analysis of the system has been carried out using radial distribution functions between different atoms of the solvent and solute. The self-diffusion coefficients of the solute and solvent are estimated from the slope of the mean square displacement vs the time plot using Einstein's equation and their binary diffusion coefficients from Darken's relation. The temperature dependency of diffusion is demonstrated via Arrhenius plots. We have further extended our work to study the effects of the system size on diffusion and to calculate the viscosity coefficients of both the solvent and solution. The calculated values are in close agreement with the previously reported results available in literature.

INTRODUCTION

Cysteine, a non-essential amino acid, contains a sulfur atom in its side chain (CH_2–S–H). The covalent link between the cysteine molecules is of two types: ordinary peptide bond (CO–NH) and disulfide bond (R–S–S–R'). Although methionine, an amino acid, also contains a sulfur atom in its side chain, it does not form a disulfide bond, neither by itself nor with any other molecules. Thus, cysteine possesses identical behavior in the formation of a covalent link during polypeptide chain synthesis. The disulfide bond plays a significant role in protein folding, stability, and functional variation.[1,2]

Cysteine is a white crystalline solid having a molar mass of 121.15 g/mol and a melting point of 513 K. Its solubility in water is 16 g per 100 ml at 288 K. However, cysteine exhibits a hydrophobic nature, due to which it generally resides in the interior of proteins. Cysteine is essential for the synthesis of highly anti-oxidative glutathione, which is important in the detoxification and protection of various tissues and organs in the body. Furthermore, cysteine contributes to the absorption of nutrients from the intestinal wall and in the metabolism of lipids. It enhances fertility and strengthens the immune system, thus aiding prevention of dementia, multiple sclerosis, and Parkinson's disease.[3] It is also recognized as an anti-aging amino acid. All these functions place cysteine in a special position that cannot be substituted by any other amino acid.[4]

The cysteine molecule as a residue in the protein chain plays a crucial role in the DNA–protein interaction. In the DNA methylation repair mechanism, cysteine works in the suicidal reaction in methyl-transfer from methyl-DNA to cysteine itself.[5] Cysteine not only plays a role in biomolecular interactions, but also acts as the catalytic agent in the electroreduction process of metals such as bismuth and gold in the appropriate solvents.[6] Therefore, the study of the mechanical properties of cysteine is necessary in biological and material sciences.

The term transport phenomenon means the process by which the mass, linear momentum, angular momentum, energy, and charge are transferred from one part of the system to another due to non-uniformity or inhomogeneity of the system. Diffusion, an important transport property, is the phenomenon in which mass is transferred as a result of random molecular motion. Various experimental techniques such as the peak-height method,[7] nuclear magnetic resonance (NMR), and molecular dynamics (MD)

simulations[8] have been performed to study the diffusion phenomenon of amino acids in water. These studies were mainly concerned about the effect of concentration, polarity, and temperature on the diffusion of amino acids.

Based on the consideration of chemical parallel of its sulfhydryl (R–S–H) and hydroxyl (O–H) groups in the side chain of other polar amino acids, the cysteine molecule appears due to the hydrophilic nature; however, free cysteine molecules are found in the hydrophobic region of proteins.[9,10] Since the cysteine side chain is hydrophobic in nature, it tends to enhance the diffusion in water. However, the sulfur atom at the cysteine side chain tends to reduce the diffusion, as it has a relatively higher atomic mass than basic elements in organic compounds, such as carbon, nitrogen, and oxygen.[11–13] The study of the effect of the hydrophobic interaction for relatively heavy molecule on the diffusion coefficient would be exciting. Importantly, the thiol –SH site of the cysteine residue in antibodies is functionally active in the interaction with metals such as gold and in bio-sensing. The cysteine molecule, after breakage of the disulfide bond in the peptide chain, plays a very important role in functionalizing the gold surface and immobilizing the antibody. Although gold is an inert metal, it can be made chemically active by interacting with a peptide sequence, basically interacting with sulfur available in the thiol group –SH of the cysteine residue.[14] Moreover, as the cysteine in combination with tryptophan can act as the strong link to bind refractory bio-receptors, it has wide applications in bio-sensing.[15]

Since amino acids are the building blocks of protein molecules, they have many similarities with more complex biomolecules such as antibiotics. Antibiotics are widely used in medicines and nutrients.[16] Therefore, the measurement of diffusivity of amino acids is important in designing the drugs. Moreover, its efficiency of movement in solution can be quantitatively measured by determining the coefficient of viscosity in aqueous solutions.

Thus, a comprehensive study of the diffusion process and viscous property of amino acid molecule in water is essential to understand life processes and the physical mechanism of inorganic compounds. Many researchers have already studied the mechanical properties of some amino acids.[6,17] To the best of our knowledge, the diffusion coefficient and coefficient of viscosity of the cysteine molecule in water using MD simulation has not yet been studied. Therefore, we intend to study these properties of cysteine.

DIFFUSION

Diffusion is a dynamic property of matter in which its particles are transported from the higher concentration region to the lower concentration region. It occurs due to the concentration inhomogeneity and thermal agitation of particles.[18] Diffusion plays many important roles in non-living substances as well as in living organisms. The diffusion in a homogeneous system having no chemical concentration gradient is called self-diffusion, and the corresponding diffusion coefficient is termed self-diffusion coefficient.[19] Einstein's equation is used to calculate the self-diffusion coefficients, which relates the diffusion coefficient with the mean square displacement (MSD) of the particles,[20,21]

$$D = \lim_{t \to \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t}. \tag{1}$$

In Eq. (1), $r(t) - r(0)$ is the displacement of the particle from the reference point during the course of time t, $[r(t) - r(0)]^2$ is the square of displacement, and $\langle \cdots \rangle$ represents the ensemble average, and hence, $\langle [r(t) - r(0)]^2 \rangle$ gives the MSD of the particle.

Binary diffusion is the diffusion of particles in the mixture of two different substances. It is the quantitatively measured diffusion coefficient using Darken's relation[22]

$$D_{12} = N_2 D_1 + N_1 D_2. \tag{2}$$

In Eq. (2), D_{12} is the binary diffusion coefficient, D_1 and D_2 are the self-diffusion coefficients of substances 1 and 2, respectively, and N_1 and N_2 are the corresponding mole fractions.

In order to estimate the diffusion coefficients, simulations are carried out using periodic boundary conditions (PBC). Under PBC, the diffusion strongly depends on the size of the simulation box as suggested by Yeh and Hummer[23] due to the long range nature of hydrodynamics interaction. The effect of the system size on the diffusion coefficient (D_{PBC}) under periodic boundary conditions is accounted for by[23–25]

$$D_0 = D_{PBC} + \frac{2.84 \, k_B T}{6\pi\eta L},$$

where D_0 is the system-size independent value of the diffusion coefficient, D_{PBC} is the simulated value of the diffusion coefficient in the cubic box of size L under periodic boundary conditions (PBC), k_B is the Boltzmann constant, T is the absolute temperature of the system, and η is the shear viscosity of the solvent,

$$or, \; D_{PBC} = D_0 - \frac{2.84 \, k_B T}{6\pi\eta L}. \tag{3}$$

From the intercept and slope of Eq. (3), the values of D_0 and η are estimated.

COMPUTATIONAL DETAILS

Molecular dynamics simulations were performed for the system of 3 cysteine and 1039 water molecules in a cubic box of size 3.17 nm at five different temperatures; 288 K, 293 K, 303 K, 313 K, and 323 K. The Extended Simple Point Charge (SPC/E) water model and Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force field parameters were used in the simulations. All the bonded and non-bonded interaction parameters are assigned in the OPLS-AA force field by default, and the parameters for SPC/E[26] water model are included in the file *spce.itp* inherent to GROMACS 5.1.1.[27] In addition, the same atom possesses different partial charges based on the group of attachment. The Coulomb interaction occurs due to the partial charge existing in the atoms/molecules. Likewise, the van der Waal's interaction occurs as a result of the induced dipole interaction.

The coordinates assigned for the molecules in the .pdb file may not be the equilibrium structures, rather they are initial guess from the electron probability density map produced by x-ray diffraction (XRD) or nuclear magnetic resonance (NMR). In addition, molecules may have been under steric hindrance, which may produce unnecessary strain in the system. In order to remove the effects and to bring the system at the minimum potential energy state, the energy minimization process was carried out using the Steepest-descent algorithm.[27]

After energy minimization, the system is ready to study the dynamic properties. However, the dynamical properties, such as diffusion and viscosity, usually vary with the parameters such as temperature, pressure, and density.[28] Therefore, before starting the production run, these aforementioned parameters should be kept constant during simulation, and the system under study is to be brought in the state of thermodynamic equilibrium, which is known as equilibration of the system. To bring the system to the state of thermodynamic equilibrium, we performed the equilibration run for each system in NPT ensemble. The velocity rescaling thermostat with a temperature coupling time of 0.01 ps and the Berendsen barostat with a coupling time of 0.8 ps were used during each simulation.[27] The isothermal compressibility of 4.6×10^{-5} bar^{-1} was taken. The Linear Constraint Solver (LINCS) constraint algorithm was applied to convert all bonds into constraints during the equilibration run.[27] The cutoff distance of 1 nm was taken for both Coulomb and Lennard-Jones (LJ) interactions, and the long range Coulomb interaction was accounted for using the PME (Particle-mesh Ewald) method with a Fourier spacing of 0.12 nm. To start up the equilibration run, the velocity of the molecules should be assigned. Maxwell–Boltzmann distribution was used to generate the initial velocities for the particles in the system. To obtain new positions and velocities of the particles after each time step, the leapfrog algorithm[27] was chosen. Each equilibration run was performed for 50 ns with a 1 fs time step.

Consequently, the production run of each system was performed to calculate the transport properties of the system in the NVT ensemble for 50 ns with a time step of 1 fs taking velocity-rescaling thermostat with a coupling time of 0.01 ps. Furthermore, it is not required to generate the initial velocities in the NVT run as the simulation continues with the velocities generated in the equilibration run.

RESULTS AND DISCUSSION

In this section, we present structural analysis and transport properties of the system at different temperatures.

STRUCTURE OF THE SYSTEM

The radial distribution functions (RDF) between the pair of atoms are used to analyze the structural properties of the system. For this, we have plotted the RDF between oxygen atoms of water molecules [$g_{OW-OW}(r)$] and carboxyl oxygen of cysteine and oxygen of water [$g_{OC-OW}(r)$] at five different temperatures.

Figure 1 shows the RDF between oxygen atoms of water molecules at the following temperatures: 288 K, 293 K, 303 K, 313 K, and 323 K.

The values of the excluded region (ER), first peak position (FPP), first peak value (FPV), second peak position (SPP), second peak value (SPV), third peak position (TPP), and third peak value (TPV) of RDF $g_{OW-OW}(r)$ are presented in Table I.

In Fig. 1, there are three distinct peaks. The first peak, which is located at the separation of about 0.27 nm from centered atom's position, is the highest and sharpest. This implies that, at this position, the maximum number of oxygen atoms is clustered from the reference oxygen atom. In other words, the probability of finding oxygen atoms at the first peak position is the highest. This is the
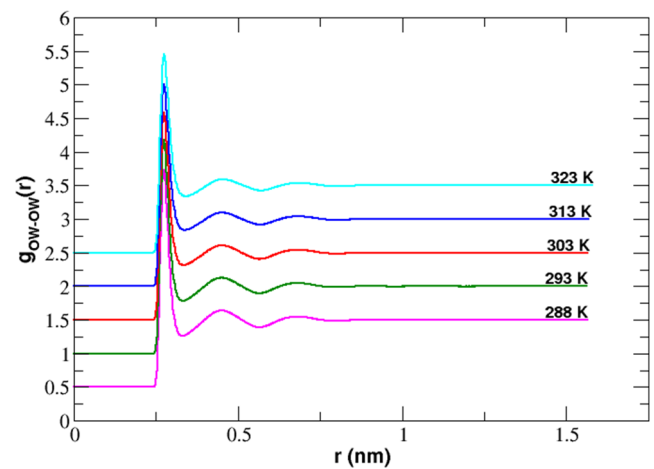
FIG. 1. RDF plot of oxygen–oxygen of water molecules, $g_{OW-OW}(r)$, at different temperatures.

most preferable position or minimum energy position from the centered atom. The value of the Lennard-Jones parameter σ of oxygen in water is 0.3166 nm and the van der Waal's radius is $(2^{\frac{1}{6}}\sigma) \approx 0.36$ nm. However, the FPP in our system is 0.27 nm less than 0.36 nm. This reveals the fact that there is not only the LJ interaction between oxygen atoms of water, but also other interactions such as Coulomb and bonded interactions are present.

The second and third peaks are relatively shorter and wider, which are located approximately at positions 0.45 nm and 0.68 nm, respectively. The excluded region, in which RDF is zero, has extended up to 0.24 nm from the center of the reference oxygen atom. Any other oxygen atom cannot exist within the excluded region due to strong repulsive forces, namely, the r^{-12} term of the LJ interaction and repulsive Coulomb interactions.[29] We have also studied the effect of temperature on RDF. With the increase in temperature, the peak positions are shifted to right, heights of the peaks are decreased, and widths are increased (see Table I). This reflects that our system has become less organized with the increase in temperature. The increase in thermal agitation of atoms in the system with rising temperature accounts for this fact. Furthermore, beyond the third peak graph is the straight line possessing the unit value on

TABLE I. Details of RDF of oxygen–oxygen atoms of water molecules at different temperatures.

			RDF analysis of OW–OW atoms				
T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV	TPP (nm)	TPV
288	0.240	0.274	3.230	0.450	1.140	0.682	1.049
293	0.240	0.274	3.175	0.450	1.129	0.680	1.048
303	0.240	0.274	3.077	0.450	1.110	0.686	1.043
313	0.240	0.276	3.000	0.450	1.096	0.690	1.041
323	0.240	0.276	2.945	0.450	1.091	0.686	1.037
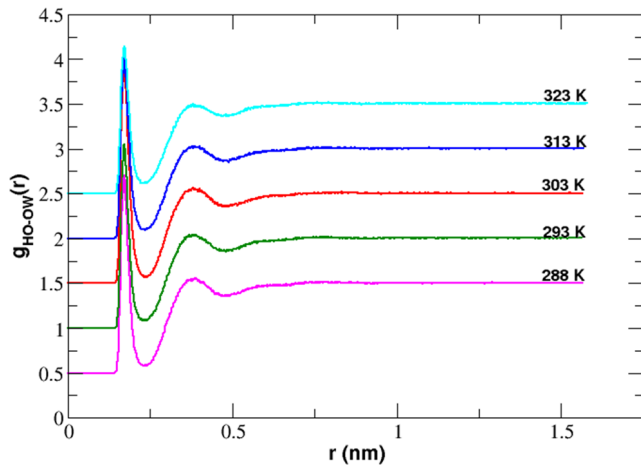
FIG. 2. RDF plot of carbonyl oxygen of cysteine and oxygen of water, $g_{OC-OW}(r)$, at different temperatures.

FIG. 3. MSD vs time plot of cysteine at different temperatures.

average. This indicates that there is no pair correlation of the oxygen atoms.

The RDF $g_{OC-OW}(r)$ gives the insight into how the carbonyl oxygen atoms of cysteine are organized around the oxygen atom of water. Figure 2 represents the RDF $g_{OC-OW}(r)$ at the above-mentioned temperatures. From the figure, it is clearly seen that the RDF has two noticeable peaks. The values of ER, FPP, FPV, SPP, and SPV are tabulated in Table II.

The first peak, which is located at the separation of about 0.33 nm from the position of the reference oxygen atom of water, is the highest and sharpest. This implies that, at this position, maximum number of carbonyl oxygen atoms of cysteine clustered from the reference oxygen atom. Therefore, this is the most preferred position of carbonyl oxygen atoms to cluster around the oxygen atom of water. The second peak is relatively shorter and wider, which is located approximately at the position of 0.58 nm. The excluded region extends up to 0.24 nm from the center of the reference oxygen atom. It is not possible to find any carbonyl oxygen within the excluded region due to strong repulsive forces. Beyond the second peak, there is no pair correlation of carbonyl oxygen atoms and the reference oxygen atom of water.

DIFFUSION COEFFICIENTS

The self-diffusion coefficients of cysteine and water have been calculated for five different temperatures by using the corresponding MSD curves. We have determined the self-diffusion coefficient from the slope of the MSD plot according to Einstein's equation (1).

We have plotted the MSD curves for 3 ns for all temperatures, even though the production run was done for 50 ns as statistics is better at the beginning than the ending region of the plot. Figures 3 and 4 show the MSD vs time plot for cysteine and water at 288 K, 293 K, 303 K, 313 K, and 323 K temperatures. The study has shown that as the temperature increases, the slope of the MSD curves also increases, which in turn increases the self-diffusion coefficient. The estimated values of self-diffusion coefficients of cysteine and water and their binary diffusion coefficients are presented in Table III with previously reported experimental values. Table III demonstrates that the simulated values of the

TABLE II. Details of RDF of carbonyl oxygen of cysteine and oxygen of water at different temperatures.

	RDF analysis of OC–OW atoms				
T (K)	ER (nm)	FPP (nm)	FPV	SPP (nm)	SPV
288	0.240	0.336	1.125	0.584	1.018
293	0.242	0.338	1.094	0.590	1.020
303	0.242	0.336	1.185	0.578	1.031
313	0.242	0.336	1.092	0.580	1.026
323	0.242	0.336	0.985	0.586	1.003

FIG. 4. MSD vs time plot of water at different temperatures.

TABLE III. Estimated values of self and binary diffusion coefficients at different temperatures.

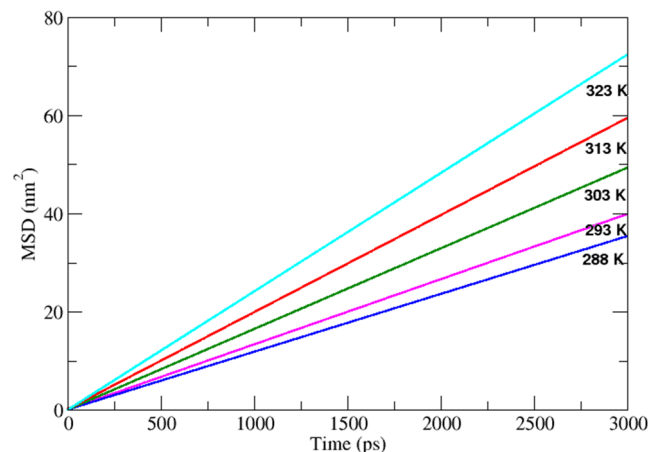|		Diffusion coefficients (D_{PBC}) (10^{-10} m^2 s^{-1})						
		Self				Binary		
		For cysteine	For water					
S.No.	Temp. (K)	MSD	MSD	Expt.[30]	Error (%)	Cal.	Expt.[7]	Error (%)
1	288	8.17 ± 1.10	19.63 ± 0.01	17.66	11.16	8.20	7.90	3.80
2	293	9.01 ± 0.27	22.12 ± 0.14	20.25	9.23	9.05	. . .	. . .
3	303	10.81 ± 0.41	27.38 ± 0.03	25.97	5.43	10.86	. . .	. . .
4	313	13.78 ± 2.79	32.95 ± 0.20	32.22	2.26	13.84	. . .	. . .
5	323	16.29 ± 0.51	40.19 ± 0.05	39.83	0.90	16.36	. . .	. . .

self-diffusion coefficients are in agreement with that of experimental values within 12% error. The error in the experimental values of self-diffusion coefficient of water, as reported by Holz,[30] is less than 1%.

The self-diffusion coefficients of cysteine and water at particular temperatures obtained in the previous sections are now used for the calculation of binary diffusion coefficients by using Darken's relation (2). We have simulated three cysteine molecules and 1039 water molecules, 1042 molecules in total. Thus, the mole fraction of cysteine is 0.003 and that of water is 0.997. The calculated values of binary diffusion coefficients and the corresponding experimental values are shown in Table III. The calculated value of the binary diffusion coefficient agrees within 4%[7] with the experimental value at 288 K. In addition, the calculated values of diffusion coefficients increase with the increase in temperature because the thermal energy of molecules increases with the increase in temperature but the density of the system decreases, which in turn increases the available space for diffusion. Thus, the molecular movement in the system is enhanced, and hence, the diffusion coefficient increases at higher temperatures.

Temperature dependency of diffusion

As observed in Table III, the diffusion phenomenon is strongly dependent on temperature. The temperature dependent behavior of diffusion is given by the Arrhenius equation[29]

$$D = D_0 \, e^{\frac{-E_a}{N_A k_B T}}. \tag{4}$$

In Eq. (4), D is the diffusion coefficient, D_0 represents the pre-exponential factor, E_a is the activation energy for diffusion, N_A is the Avogadro's number whose value is 6.022×10^{23} mol^{-1}, k_B is the Boltzmann's constant whose value is 1.38×10^{-23} J K^{-1}, and T is the absolute temperature. On taking natural logarithm in Eq. (4), we get

$$\ln D = \ln D_0 - \frac{E_a}{N_A k_B T}. \tag{5}$$

The activation energy E_a for diffusion can be obtained from the slope of $\ln D$ vs $\frac{1}{T}$ plot (Arrhenius plot) as

$$E_a = -N_A \, k_B \frac{\partial \ln D}{\partial (1/T)}. \tag{6}$$

The intercept when extrapolated to the $1/T \to 0$ in the Arrhenius plot gives the pre-exponential factor.

Figure 5 depicts the Arrhenius plot of the simulated values of self-diffusion of cysteine. The activation energy for self-diffusion of cysteine calculated using the slope of the linear fit of simulated values is found to be 15.49 kJ mol^{-1}.

Figure 6 portrays the Arrhenius plot of both the simulated and experimental values of self-diffusion of water. The activation energies for self-diffusion of water calculated using the corresponding slope of the linear fit of simulated values and experimental values are found to be 15.67 kJ mol^{-1} and 17.88 kJ mol^{-1}, respectively.

Figure 7 displays the Arrhenius plot of simulated values of binary diffusion of cysteine in water. The activation energy for binary diffusion of cysteine in water calculated using the slope of the linear fit of simulated values is found to be 15.50 kJ mol^{-1}.
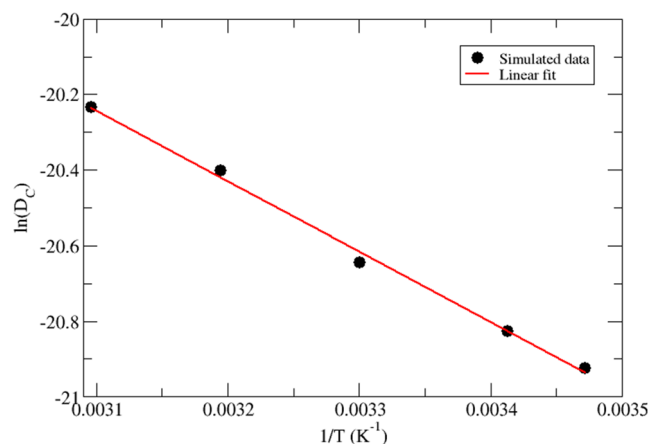
FIG. 5. Arrhenius diagram for self-diffusion coefficients of cysteine.
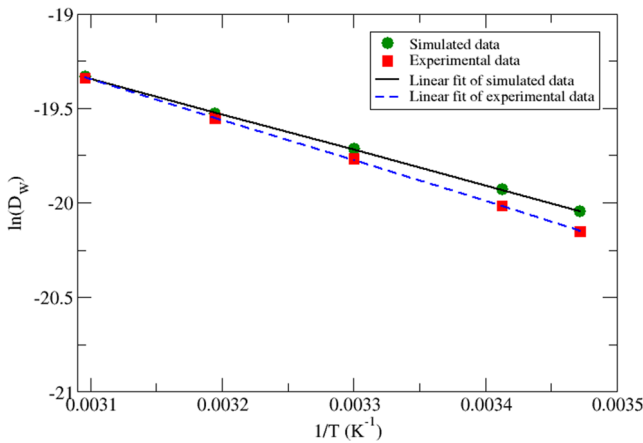
FIG. 6. Arrhenius diagram for self-diffusion coefficients of water
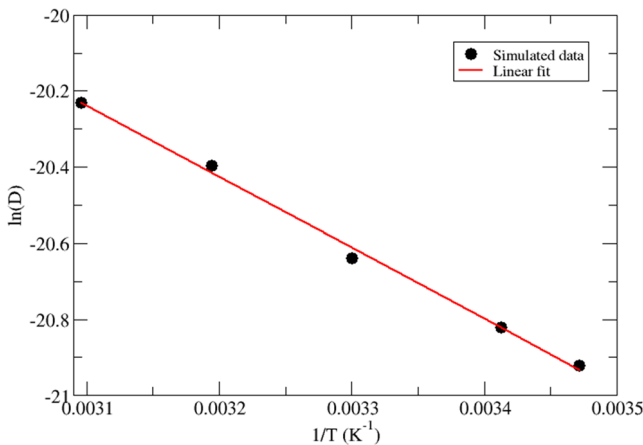
FIG. 7. Arrhenius diagram for binary diffusion coefficients of the binary mixture of cysteine and water.

Figures 5–7 demonstrate the temperature dependency of diffusion. From these plots, it is seen that the diffusion coefficients are found to increase with temperature. We have calculated the activation energies for diffusion of cysteine, water, and their binary mixture by using the slopes of the respective Arrhenius plots, which are tabulated in Table IV.

From Table IV, it is observed that the activation energies for self-diffusion of cysteine and for the binary diffusion of cysteine

TABLE IV. Activation energies for diffusion.

System	Activation energy (E_a) (kJ mol^{-1})		
	Simulated	Experimental	Error (%)
Cysteine	15.49		
Water	15.67	17.88[30]	12.36
Binary mixture	15.50		

TABLE V. Estimated values of diffusion coefficients for different sizes (L) of systems at 288 K.

	Diffusion coefficients (D_{PBC}) (10^{-10} m^2 s^{-1})					
	Self			Binary		
	For cysteine	For water				
L (nm)	MSD	MSD	Expt.	Calculated	Expt.	Error (%)
2.76	8.08	19.36	17.66[30]	8.11	7.90[7]	2.7
3.17	8.17	19.63		8.20		3.8
3.75	8.47	20.21		8.50		7.6

in water are almost same. This implies that the concentration of cysteine in the system is infinitesimal. Furthermore, the activation energy calculated for simulated and experimental values of self-diffusion of water is in agreement with the error of 13%.

Effect of system size on diffusion

Moreover, the diffusion coefficient under periodic boundary conditions (PBC) also depends on the size of the system.[23] In the above calculation, the diffusion coefficient has been calculated at different temperatures under periodic boundary conditions. Now, the simulation was extended to find how the diffusion coefficients vary by changing the size of the box. For this, other two systems were set up: (i) 2 cysteine in 693 water molecules in the box of size 2.76 nm and (ii) 5 cysteine in 1732 water molecules in the box of size 3.75 nm. The estimated values of diffusion coefficients under periodic boundary conditions with simulation boxes of different sizes are tabulated in Table V.

In addition, the values of the viscosity coefficient of water and the solution of cysteine in water were determined at 288 K. They were calculated by plotting D_{PBC} vs $1/L$ in accordance with Eq. (3) as in Figs. 8 and 9. The estimated values of the viscosity coefficient of water at targeted temperature are in agreement with the experimental value within 38%. This error might be the nature of the water
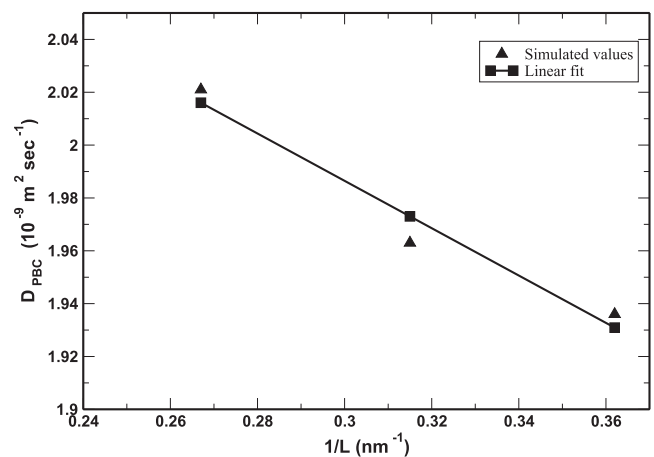
FIG. 8. D_0 vs ($1/L$) plot for water at 288 K.
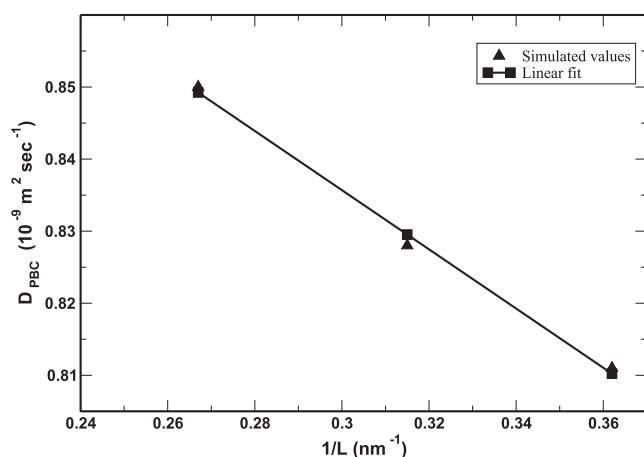
FIG. 9. D_0 vs $(1/L)$ plot for solution of cysteine in water at 288 K.

TABLE VI. System-size independent value of the diffusion coefficient (D_0) and viscosity coefficient (η) for water and solution at 288 K.

System	D_0 (10^{-10} m^2 s^{-1})			η (10^{-4} Nm^{-2} s)		
	Cal.	Expt.	Error (%)	Cal.	Expt.	Error (%)
Water	22.39	17.66[30]	26.8	7.10	11.37	37.6
Solution	9.59	7.90[7]	21.4	14.57		

model SPC/E used for solvation, which underestimates the value of viscosity.[31] The viscosity coefficient of solution is found to be greater than that of water as presented in Table VI.

CONCLUSIONS

In this work, we have performed the molecular dynamics study of transport properties of cysteine in SPC/E water molecules at 288 K, 293 K, 303 K, 313 K, and 323 K temperatures using the GROMACS 5.1.1 software package. We used OPLS-AA force field parameters throughout the simulation. The structures of the solute and solvent of the system are studied via radial distribution functions between atoms. The analysis of the RDF plots at different temperatures reveals that the system becomes less organized with the increase in temperature. Moreover, Lennard-Jones and Coulomb interactions, including many body effects, contribute to the structural properties of the system.

The self-diffusion coefficients of both cysteine and water are determined by using the Einstein's equation. In addition, the diffusion of the binary mixture of cysteine and water is calculated by using Darken's relation. The simulated values obtained are compared with the corresponding experimental values. The simulated values of self-diffusion coefficients of water show excellent agreements with experimental values, especially at higher temperatures and with a small deviation (~11%) at low temperatures (288 K). Likewise, the simulated values of the binary diffusion coefficient of cysteine in water were compared with the available experimental values at 288 K. This comparison shows very little deviation of about 8%.

Furthermore, the estimated values of diffusion coefficients increase with the increase in temperature, which follows the Arrhenius plots. We also estimate the activation energies of diffusion. We have compared the calculated value of activation energy for self-diffusion of water with the experimental value, which is in agreement with the error of 13%. In addition, we have studied the effect of the system size on diffusion, and the viscosity coefficients of both water and solution are estimated at 288 K.

As the further extension of this work, we have a plan to study the transport properties of polycysteine in aqueous medium.

REFERENCES

[1] G. Bulaj, Biotechnol. Adv. **23**, 87–92 (2005).
[2] W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga, Biochemistry **39**, 4207–4216 (2000).
[3] R. M. Canet-Avilés, M. A. Wilson, D. W. Wilson, R. Ahmad, M. Rili, C. McLendon, S. Bandyopadhyay, M. J. Baptista, D. Ringe, G. A. Petsko, and M. R. Cookson, Proc. Natl. Acad. Sci. U. S. A. **101**, 9103–9108 (2004).
[4] S. I. Rizvi and R. Jha, Expert Opin. Drug Discovery **6**, 89–102 (2011).
[5] D. S. Daniels, T. T. Woo, K. X. Luu, D. M. Noll, N. D. Clarke, A. E. Pegg, and J. A. Tainer, Nat. Struct. Mol. Biol. **11**, 714 (2004).
[6] A. Nosal-Wiercińska, J. Electroanal. Chem. **681**, 103–108 (2012).
[7] W. Jin and H. Chen, Chromatographia **52**, 17 (2000).
[8] A. S. Virk, T. S. Gardner, S. A. Willis, A. M. Torres, and W. S. Price, Front. Phys. **3**, 1 (2015).
[9] P. Heitmann, Eur. J. Biochem. **3**, 346–350 (1968).
[10] N. Nagano, M. Ota, and K. Nishikawa, FEBS Lett. **458**, 69–71 (1999).
[11] E. B. Liu and M. Hilpert, Comput. Geosci. **15**, 379–384 (2011).
[12] D. Sandrin, D. Wagner, C. E. Sitta, R. Thoma, S. Felekyan, H. E. Hermes, C. Janiak, N. de Sousa Amadeu, R. Kühnemuth, H. Löwen et al., Phys. Chem. Chem. Phys. **18**, 12860–12876 (2016).
[13] Y. Ma, C. Zhu, P. Ma, and K. T. Yu, J. Chem. Eng. Data **50**, 1192–1196 (2005).
[14] B. Della Ventura, A. Ambrosio, A. Fierro, R. Funari, F. Gesuele, P. Maddalena, D. Mayer, M. Pica Ciamarra, R. Velotta, and C. Altucci, ACS Appl. Mater. Interfaces **8**, 21762–21769 (2016).
[15] B. Della Ventura, L. Schiavo, C. Altucci, R. Esposito, and R. Velotta, Biomed. Opt. Express **2**, 3223–3231 (2011).
[16] M. Ibba, C. Stathopoulos, and D. Söll, Curr. Biol. **11**, R563–R565 (2001).
[17] L. G. Korothchkina, E. M. Ciszak, and M. S. Patel, Arch. Biochem. Biophys. **429**, 171–179 (2004).
[18] J. Crank, The Mathematics of Diffusion, 2nd ed. (Oxford University Press, Ely House, London, WI, 1975).
[19] H. Hirakawa, Y. kamei, and O. Yasumichi, Bull. Chem. Soc. Jpn. **46**, 2659 (1973).
[20] D. Frenkel and B. Smith, Understanding Molecular Simulation from Algorithms to Applications, 2nd ed. (Academic Press, USA, 2002).
[21] M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids (Oxford University Press, USA, 1989).
[22] L. S. Darken, Trans. AIME **175**, 184 (1948).

[23] I. C. Yeh and G. Hummer, J. Phys. Chem. B **108**, 15873–15879 (2004).

[24] B. Dünweg and K. Kremer, J. Phys. Chem. B **99**, 6983–6997 (1993).

[25] S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. H. Vlugt, and O. A. Moultos, J. Chem. Theory Comput. **14**, 2667–2677 (2018).

[26] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269–6271 (1987).

[27] D. Spoel et al., GROMACS User Manual, version 5.1.1, 2016.

[28] S. P. Khanal, Y. P. Kandel, and N. P. Adhikari, AIP Adv. **90**, 065303 (2019).

[29] I. Poudyal and N. P. Adhikari, J. Mol. Liq. **198**, 77 (2014).

[30] M. Holz, S. R. Heil, and A. Sacco, Phys. Chem. Chem. Phys. **2**, 4740 (2000).

[31] A. P. Markesteijn, R. Hartkamp, S. Luding, and J. Westerweel, J. Chem. Phys. **136**, 134104 (2012).

Journal of Physics Communications

Binding of SARS-CoV-2/SARS-CoV spike protein with human ACE2 receptor

Rajendra P Koirala[1]**, Bidhya Thapa**[2]**, Shyam P Khanal**[1]**, Jhulan Powrel**[1]**, Rajendra P Adhikari**[3]** and Narayan P Adhikari**[1]

[1] Central Department of Physics, Tribhuvan University, Kathmandu, Nepal
[2] Padma Kanya Multiple Campus, Tribhuvan University, Kathmandu, Nepal
[3] Department of Physics , Kathmandu University, Dhulikhel, Nepal

E-mail: narayan.adhikari@cdp.tu.edu.np

CrossMark

Abstract

SARS-CoV-2 virus is the serious health concern throughout the world. A comprehensive investigation of binding of SARS-CoV-2 active site with host receptor protein hACE2 is important in designing effective drugs. In the present work, the major amino acid binding partners between the virus CTD and host receptor have been studied and are compared with SARS-CoV RBD binding with hACE2. Our investigation show that some unique hydrogen bond pairs which were not reported in previous work. Along with hydrogen bonding, salt-bridges, hydrophobic interactions and contributions of electrostatic and van der Waals contacts play significant role in binding mechanism. The binding affinity of SARS-CoV-2 CTD/hACE2 is greater than SARS-CoV RBD/hACE2. This outcome is also verified from the free energy estimation by using umbrella sampling.

1. Introduction

Corona virus disease (COVID-19) pandemic, caused by severe acute respiratory syndrome (SARS)-like corona virus (SARS-CoV-2), is a serious health concern for the global community [1–4]. Although the origin of the virus is still unclear, it has been spread all over the world threatening the human civilizations after its initial outbreak from China in 2019. Till date, more than 52 millions infected population has been reported globally and more than 12 hundred thousands people have lost their lives [5]. There is no approved drug or vaccine against the COVID-19, even though several antiviral drugs have been proposed and are also in clinical trials [6]. Understanding more about interactions of this virus with human body is highly demanding at this pandamic time to design drugs. SARS-CoV and other viruses had also threatened the human society at different periods; however the influence of SARS-CoV-2 is significantly higher than other viruses throughout the globe.

SARS-CoV-2 has more than 70 percent of structural similarity with SARS-CoV; most of the residues at binding interface are similar [7, 8]. SARS-CoV-2 similar with SARS-CoV and other corona viruses utilize same human angiotensin converting enzyme 2 (hACE2) receptor to enter into human cell. This entry process is mediated by the spike(s) glycoprotein which are embedded in the capsid of SARS-CoV-2 [9]. The spike protein is subdivided into two receptor binding entities S1 and S2. S1 is responsible in the detection of receptor, whereas S2 plays important role in membrane fusion. Similar to SARS-CoV, C-Terminal Domain(CTD) of S1 subunit of spike protein in SARS-CoV-2 acts as receptor binding domain (RBD) [10, 11]. Even though both SARS-CoV and SARS-CoV-2 have same binding domain, the binding affinity of SARS-CoV-2 is different from that of SARS-CoV [12, 13].

Immediately after the COVID-19 outbreak, several researches have been carried out to identify the nature and location of binding of SARS-CoV-2 CTD with hACE2 using static crystal structure [14, 15]. Although, these researches have attempted to detect the entry process and binding mechanism of SARS-CoV-2 with hACE2, the breakthrough on drug designing is still challenging. Several works are in the way of hopeful future, exploration of detail binding mechanism is still being essential. Moreover, the detail dynamics of SARS-CoV-2 and hACE2 can be very important to design the drug. When we were independently working on the binding mechanism of SARS-CoV-2 with host receptor

hACE2, in the mean time similar type of research works have been published [12]. However our technique as well as some results are different from previous works.

We focus on the estimation of free energy difference of virus protein and hACE2 complex. The free energy calculation provides in-depth insight on the binding mechanism between the protein molecules [16]. There are several experimental techniques of measuring binding free energy such as isothermal titration calorimetry (ITC) [17], fluorescence resonance energy transfer (FRET) [18], nuclear magnetic resonance (NMR) [19], surface plasmon resonance (SPR) [20] and many others. The computational approach can be the best complement for large scale investigations [21–23]. Out of many computational approach, umbrella sampling is one of the widely used method for the estimation of free energy in large molecular systems [24, 25]. It improves the sampling system by designing and implementing the biasing potentials as a function of reaction coordinates [26, 27]. If an energy barrier exists in between two regions of configuration states, there may be poor sampling, despite the long simulation run being carried out. The applied biasing potential bridges such configuration states and makes it easier in searching local or global minima, which can be considered as the structurally favorable state in the molecular complex [28]. Besides these techniques, free energy can be calculated directly from steered molecular dynamics (SMD) [29, 30].

In the present work, we have carried out molecular dynamics (MD) simulations for the comprehensive study of binding mechanism of SARS-CoV-2 CTD with hACE2 and also compared with SARS-CoV-RBD/hACE2. In addition, umbrella sampling method has been executed to estimate the binding free energy of SARS-CoV-2 CTD/hACE2. Required windows for the umbrella sampling have been taken from steered molecular dynamics (SMD) [31] simulations. In SMD, SARS-CoV-2 CTD has been translated taking the hACE2 as the reference molecule. The quantitative estimation of binding affinity between the targeted molecules facilitates in silico-drug designing. We have also performed comparative study of various interactions such as hydrogen bonding, salt bridges, hydrophobic, electrostatic and van der Waals interactions at the binding interface of SARS-CoV-2 and SARS-CoV with hACE2. Furthermore, the contact surface area of these complexes have been estimated and compared to investigate the stability.

2. Methods

2.1. System set up
Two molecular structures, PDB IDs 6LZG and 2AJF, were taken for the molecular dynamics simulations. The PDB ID 6LZG contains the complex of SARS-CoV-2 CTD and hACE2 receptor protein (i.e., SARS-CoV-2 CTD/hACE2 complex) and that of PDB ID 2AJF contains the complex of SARS-CoV RBD and hACE2 receptor protein (i.e., SARS-CoV RBD/hACE2 complex) [32]. CHARMM-GUI [33] was used to create the pdb and psf files of these complexes. Then, both the complex structures were solvated using TIP3P [34] water and electrically neutralized by adding NaCl. We have added the NaCl in the system with concentration 0.15 M by using CHARMM-GUI. In SARS-CoV-2 CTD/hACE2 complex system 220 Na^+ ions and 197 Cl^- ions were added to neutralize the system. Similarly in SARS-CoV RBD/hACE2 complex system 214 Na^+ and 188 Cl^- ions were added so that the system became neutral. A cubical box of dimensions $144 \times 144 \times 144$ $\mathring{A}^3$ was prepared for NPT simulation of the complex SARS-CoV-2 CTD/hACE2 and another cubical box of dimensions $131 \times 131 \times 131$ $\mathring{A}^3$ was prepared for NPT simulation of the complex SARS-CoV RBD/hACE2. Further, two equal sized orthorhombic simulation boxes were prepared in order to estimate the free energy differences of above complexes by changing the dimensions to $250 \times 90 \times 90$ $\mathring{A}^3$ and electrically neutralized by adding NaCl with concentration 0.15 M.

2.2. Molecular dynamics simulation
All atom molecular dynamics (MD) simulations were performed using NAMD [31] simulation package. The CHARMM36m [35] force field was used for each simulations. The Particle Mesh Ewald (PME) [36] was used for the long-range interactions with a 12.0 $\mathring{A}$ non-bonded cut-off. The energy minimization was performed for 10,000 steps, using the conjugate gradient and line search algorithm [37, 38]. The system was then equilibrated at 310 K for 10 ns with harmonically restrained heavy atoms taking 1 fs time step. Finally, the production run was propagated for 250 ns simulation under NPT conditions by using Langevin dynamics with a damping constant of 1 ps^{-1} taking time step of 2 fs.

2.3. Molecular dynamics and umbrella sampling
To perform the umbrella sampling, sample windows were chosen from steered molecular dynamics (SMD) trajectories. During SMD, CTD/RBD of SARS-CoV-2 CTD/SARS-CoV RBD were pulled correspondingly towards the negative x-direction with constant velocity pulling method of velocity 0.00005 $\mathring{A}$/fs. In this process, the alpha carbons of hACE2 protein were taken as the fixed atoms and alpha carbons in CTD/RBD part of spike protein of the systems were taken as the dummy atoms. CTD/RBD of spikes were pulled from their center of mass (COM) along the negative x-direction with constant velocity $(\vec{v} = d\vec{x}/dt)$ in water and ions environment. Then the SMD atom experiences the force $\vec{F}(t) = k(\vec{v}\,t - \vec{\Delta}x)$, providing the external potential energy,

$$U(x, t) = \frac{1}{2}k(\vec{v}\ t - \Delta\vec{x}.\hat{n})^2 \qquad (1)$$

where, k (=5 kcal mol^{-1} Å^{-2}) is the spring constant and gives the stiffness of the applied harmonic restraining force, and $\Delta\vec{x}\ (t) = \vec{x}\ (t) - \vec{x}_0$, is the displacement of SMD molecules from initial position $\vec{x}_0$ to instantaneous position $\vec{x}\ (t)$ and $\hat{n}$ is the unit vector along the direction of pulling.

Out of many other techniques of free energy calculations [39], umbrella sampling was performed to investigate the free energy difference during the translation of SARS-CoV-2 CTD from hACE2 protein for system SARS-CoV-2 CTD/hACE2; and identical condition is applied for system SARS-CoV RBD/hACE2. SMD trajectories were used to select the appropriate windows. Identifying the information on the termination of molecular interactions from SMD, we estimated the number of umbrella windows in both the systems. Ten windows were prepared in SARS-CoV-2 CTD/hACE2 and six windows were prepared for SARS-CoV RBD/hACE2 complexes. Every successive window was taken from the SMD trajectories during the translation of 1 Å along the negative x-direction. The window size ensures the sufficient overlapping of successive windows to cover the entire reaction coordinate space. The reaction coordinate was chosen as the distance between the center of mass (COM) of hACE2 and CTD/RBD spike along the negative x-axis. To make the necessary overlapping reaction coordinates, a bias potential of the ith window V$_i$(x) was used to force the system to fluctuate in coordinate space, which is given by,

$$V_i(x) = \frac{1}{2}k(x_i - x_0)^2 \qquad (2)$$

where x_0 is the harmonic constraint defining a center of window i (i=1 to 10 for for SARS-CoV-2 and 1 to 6 for SARS-CoV), and force constant k is the window width. We used the force constant of 1.5 kcal mol^{-1} Å^{-2}.

2.4. Data analysis

Visual Molecular Dynamics (VMD) [40] and Pymol [41] were used to visualize as well as generate images of the complex structures . VMD analysis tools were used to identify and analyze non-bonded interactions by using the simulation trajectories. The NAMD energy plugin, available in VMD, was used to calculate the non-bonded interaction energy contributions. Pycontact [42] software package was used to analyzed the hydrophobic interactions and salt bridges between the targeted protein residues in CTD/RBD of spike protein and ACE2. Weighted Histogram Analysis Method (WHAM) program [43] was used to estimate the free energy from umbrella sampling simulation. The free energy calculation of large molecular system is generally computationally demanding. This method minimizes the statistical errors as well as increases the efficiency of computational simulation. Moreover, it has the advantage of multiple overlapping of probability distributions for obtaining better estimation of free energy calculations.

3. Results

3.1. Conformational variation in complexes

To examine the conformational variation during the dynamics, we have estimated the root mean square deviation (RMSD) of each molecule on SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes. Besides RMSD, contact surface area between the molecules within the each complex has also been calculated for both complexes.

We have calculated the RMSD for all atoms of proteins backbone without taking hydrogen atoms. The structure from first step of the simulation was taken as the reference to calculate the RMSD. The RMSD of hACE2 and spike CTD/RBD have been compared separately to evaluate the structural integrity of the molecules. Figures 1(a) shows the RMSD of hACE2 molecule in SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 complexes and figure 1(b) represents the RMSD of spike CTD/RBD. From the figure, it is observed that RMSD of hACE2 of SARS-CoV-2 CTD/hACE2 is smaller than that of SARS-CoV RBD/hACE2. RMSD of both the systems are stable with the values below 3.0 Å and 4.5 Å for CoV-2 and CoV respectively. SARS-CoV-2 CTD has the RMSD of 2.5 Å, whereas the RMSD of SARS-CoV RBD is 4.0 Å. This shows the large rearrangements of structure in SARS-CoV RBD, while SARS-CoV-2 CTD structure remains relatively stable.

To get more insight into stability, we also analyzed the contact surface area between the spike protein CTD/RBD and hACE2 receptor using MD trajectories. Contact area is the surface buried at the interface between two proteins which contributes to bind and stabilize the protein-protein complexes. Larger contact surface indicates greater stability of the structure [44]. The estimated values of contact surface area are presented in table 1. From the table 1, it has been observed that SARS-CoV-2 CTD/hACE2 has larger contact area than SARS-CoV RBD/hACE2 by 77.02 ± 2.46 Å^2. The contact surface area for SARS-CoV-2 CTD is more in comparison to SARS-CoV RBD indicating the greater binding affinity of SARS-CoV-2 with receptor.
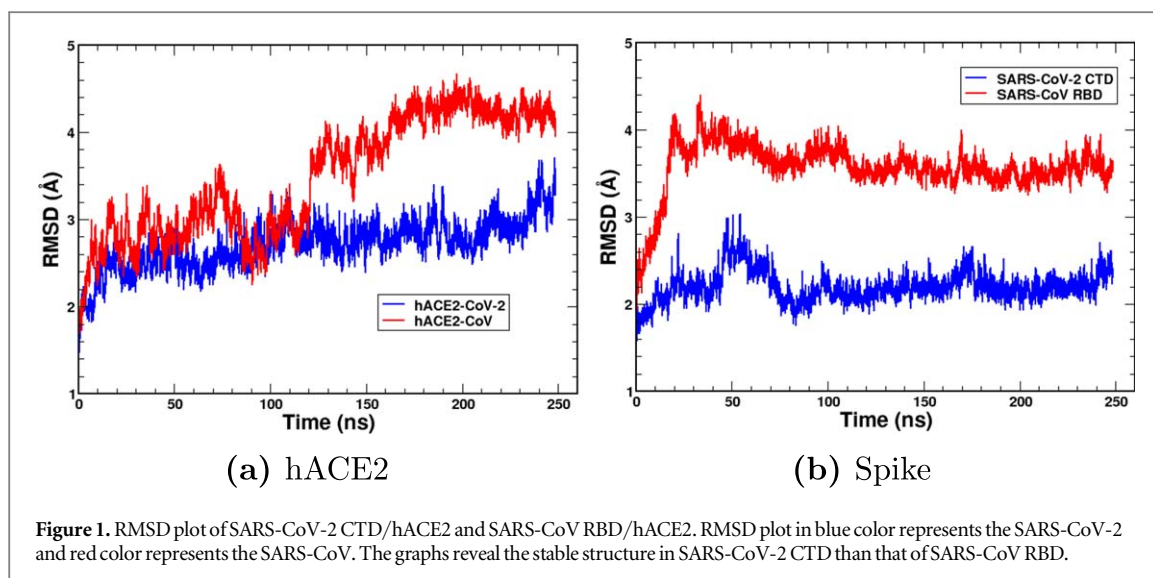
Figure 1. RMSD plot of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2. RMSD plot in blue color represents the SARS-CoV-2 and red color represents the SARS-CoV. The graphs reveal the stable structure in SARS-CoV-2 CTD than that of SARS-CoV RBD.

Table 1. Calculation of contact surface area; A_1, A_2 & A_3 are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV-2 CTD and their complexes; and A_1', A_2' & A_3' are the average solvent accessible surface area (SASA) of hACE2, SARS-CoV RBD and their complexes.

Complex		SASA (Å^2) for				Net contact area (A)[44]
SARS-CoV-2 CTD/hACE2	hACE2 (A_1)	SARS-CoV-2 CTD (A_2)	complex (A_3)	$A_1 + A_2 - A_3$	$(A_1 + A_2 - A_3)/2$
	29 100.36	11227.97	38 549.02	1779.31	889.66 ± 1.07
SARS-CoV RBD/hACE2	hACE2 (A_1')	SARS-CoV RBD (A_2')	complex (A_3')	$A_1' + A_2' - A_3'$	Net contact area (A')
	29 092.62	10859.54	38326.88	1625.27	812.64 ± 2.22

3.2. Non-bonded interactions

Furthermore, we studied in details the hydrogen bonds, salt-bridges, hydrophobic, electrostatic and vdw interactions between the residues residing at the interface between SARS-CoV-2 CTD/hACE2 and obtained results are compared with SARS-CoV RBD/hACE2.

3.2.1. Hydrogen bonds

At the interface region, hydrogen bonds play pivotal role in binding the molecules to form a stable complex. Wang *et al* (2020) and Lan *et al* (2020) have studied the atomic interactions at the interface of static crystal structure of SARS-CoV-2 CTD/hACE2 complex [14, 15], whereas we have investigated the hydrogen bonds at the interface of two complexes by analyzing the MD simulations trajectories. The cut-off distance for hydrogen bond was taken to be 3.5 Å [14]. We monitored the time evolution of number of hydrogen bonds formed at the interface between SARS-CoV-2 CTD/hACE2 and also compared with that of SARS-CoV RBD/hACE2 as shown in figure 2 (also see supplementary table S 1). Hydrogen bonds were found to be consistently existing in both complexes. Total hydrogen bonds formed during the simulations were seen to be more in case of SARS-CoV; however the strength and life time of potential hydrogen bonds were found to be greater in case of SARS-CoV-2.

Many research works have been published by analyzing the hydrogen bonds pair partners between the molecules in the complexes. Even though our investigations regarding the hydrogen bonds in the complexes are in consistent with those published papers, some pair partners are not consistent with these previously published outcomes. Ali *et al* reported three unique hydrogen bonds in SARS-CoV-2 CTD/hACE2 complex, which were not detected in SARS-CoV RBD/hACE2 complex [12]. We found consistent result in GLU35-TYR449 pair, however the result is not consistent with other two pair partners: TYR449-ASP38 and GLN498-LYS353. We have clearly observed ASP38-TYR436 and LYS353-GLY488 pairs in SARS-CoV-RBD/hACE2. Furthermore, a strong hydrogen bond has been detected between GLN498 with GLN42. In static structures, no hydrogen bond was formed by SER19 of hACE2 with ASP463 residue of SARS-CoV RBD [14, 15]. Our investigation shows two potential hydrogen bonds formed between main-main & main-side chains of SER19 of hACE2 with side chain of ASP463 of SARS-CoV RBD and similar type of bond has been detected between SER19 and ALA475 in SARS-CoV-2 CTD/hACE2 complex (see figure 4(a) and supplementary figures S1 and S2).

In the present work, interactions of molecules in each complex has been observed considering three main regions where the interfacial residues of hACE2 take part actively in binding with the spike CTD/RBD section of

Figure 2. Comparison of time evolution of number of hydrogen bonds at the interfaces of SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2.

Figure 3. The arrangement of major interaction residues at the interface of SARS-CoV RBD/hACE2 complex. Three subdivided parts of the interface region are circled in 3(a). Figure 3(a) represents the molecular structure of the complex in the beginning of the simulations and that of at the end of the simulation is represented by figure 3(b). The contact pairs in color contrast are SER19-ASP463 (red), ASP30-TYR442 (green), LYS31-ASN479 (blue), ASP355-THR486 (magenta), and LYS353-GLY488 (violet).

the virus molecule as shown in supplementary figures S1(a) and S2(a). The hydrogen bonds formed at three key regions of interface between SARS-CoV RBD/hACE2 in the begining of the simulation run are shown in figures (see supplementary figures: S1(a), S1(b) & S1(c)). Some residue pairs whose hydrogen bond (Hbond) occupancy percentage greater than 20% is shown in figure 3 and detail of Hbond occupancy percentage in three key regions are shown in figures 4(a), (b) and (c). We have observed some differences in the atomic interactions at the interface of both virus proteins and hACE2 than that of static crystal structure. The two different approaches might be the reason of variation in the number of interactions.

The rearrangement of aminoacids at the binding vicinity have been detected in both the complexes. Different color contrast have been used to indentify the pair partners as shown in figures 3 and 5. The binding affinity of the complex was observed increasing due to the allignments of pair partners at the interface.

At the binding interface of SARS-CoV-2 CTD and hACE2 receptor, three key regions where most of the polar contacts are actively participated to form the hydrogen bonds. The interactions at the interface in which Hbond occupancy greater than 20% are shown in figure 5, and also the details of Hbond occupancy in these regions are shown in figures 6(a), (b) and (c). In region 1, SER19, GLN24 and TYR83 of hACE2 form hydrogen bonds with ALA475 and ASN487 of SARS-CoV-2 CTD as in figures (see supplementary figures: S2(a), S3(a), S4(a), S5(a) & S6(a)). In region 2, there are interactions between the residues LYS417, TYR453 and GLN493 of SARS-CoV-2 CTD forming hydrogen bonds with ASP30, LYS31 and GLU35 of hACE2 (see supplementary figures: S2(b), S3(b), S4(b), S5(b) & S6(b)). Similarly, in region 3, there is extensive network of hydrogen bonds between SARS-CoV-2 CTD residues TYR449, GLN498, THR500, ASN501, GLY502 and TYR505 with the hACE2 residues GLU37, ASP38, TYR41, GLN42, LYS353 and ASP355 (see supplementary figures: S2(c), S3(c), S4(c), S5(c) & S6(c)). Because of dynamical nature of our system, there is continuous formation, breaking and

Figure 4. Hbond occupancy percentage details of major interacting residues in SARS-CoV RBD/hACE2 complex. The double bars for SER19-ASP463 in bar diagram shows the significant bonding of these residues in side chain—main chain and main chain—main chain interactions. 4(a), 4(b) and 4(c) are the Hbond occupancy percentage for three key regions respectively.

Figure 5. Major interacting residues at the interface of SARS-CoV-2 CTD/hACE2 complex. Three subdivided regions are circled in which 4(a) represents the molecular complex at the beginning stage of the simulation run and 4(b) represents the structure rearrangement of residues of molecular complex at the end of simulation run. The pairs in color contrast are TYR83-ASN487 (red), ASP30-LYS417 (green), GLU35-CLN493 (violet), LYS353-GLY502 (lime), TYR41-THR500 (orange) and ASP355-THR500 (magenta).

Figure 6. Details of Hbond occupancy percentage of major interacting residues in SARS-CoV-2 CTD/hACE2 complex. 6(a), 6(b) and 6(c) are the Hbond occupancy percentage for three key regions respectively.

reformation of hydrogen bonds during the simulations. The variation of hydrogen bonds formed during the five representative frames of 250 ns simulation of SARS-CoV-2 has been shown in supplementary figures S2–S6.

3.2.2. Salt-bridges
In addition to extensive network of interfacial hydrogen bonds, another important contributions to protein-protein binding comes from salt-bridge interactions. MD trajectory analysis has shown three salt-bridges, having different bond length and strength, formed at the interface of SARS-CoV-2 CTD/hACE2. The salt-bridge formed between the residue LYS417 of SARS-CoV-2 CTD with ASP30 of hACE2 is found to be the strongest one among them owing to its short bond length. The remaining residues GLU484 and LYS458 of
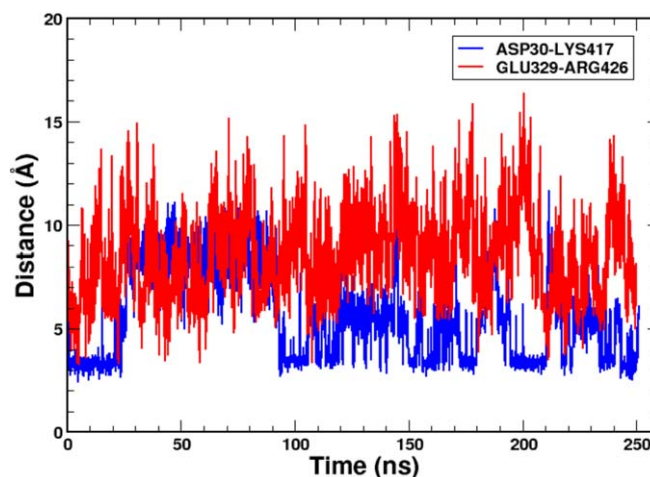
Figure 7. Comparison of time evolution of salt-bridge bond length at the interface of SARS-CoV RBD/hACE2 and SARS-CoV-2 CTD/hACE2. The residue pair ASP30-LYS417 corresponds to the SARS-CoV-2 CTD/hACE2 system and GLU329-ARG426 corresponds to the SARS-CoV RBD/hACE2 system.

SARS-CoV-2 CTD have formed salt-bridges with LYS31 and GLU23 of hACE2 respectively. In contrast, we found only one salt-bridge formed between ARG426 of SARS-CoV RBD and GLU329 of hACE2 which is weaker than that of SARS-CoV-2 because of larger bond length as in figure 7.

3.2.3. Hydrophobic interactions

3.2.4. Electrostatic and van der Waals (vdw) interactions

The electrostatic and van der Waals (vdw) interactions in two complexes SARS-CoV RBD/hACE2 and SARS-CoV-2 CTD/hACE2 have been studied. Supplementary figure S7 depicts the comparative analysis of energy due to electrostatic and vdw interactions as a function of time. In the beginning of simulations, the electrostatic contributions of SARS-CoV-2 CTD/hACE2 was distinctly higher than SARS-CoV RBD/hACE2, however most of the simulation time, the contributions were almost equal. In addition, the potential energy contributed by vdw interactions were consistently almost equal for both the systems throughout the simulations. It reveals that electrostatic and vdw interactions are almost equally contributed in binding both the complexes.

3.3. Free energy

To investigate the energetic difference in binding of hACE2 with SARS-CoV-2 CTD and SARS-CoV RBD, the free energy differences have been estimated using umbrella sampling technique. Umbrella windows were taken from the trajectories of SMD simulations. The interactions between the molecules in SARS-CoV-2 CTD/hACE2 were found terminated after traversing 9 Å distance away from the original position. To incorporate all interacting pathways, ten windows with 1 Å distance separation were taken for every successive window. On the other hand, the interactions between the molecules in SAR-CoV RBD/hACE2 were found ceased after traversing 5 Å distance from the original position. Therefore, six windows were prepared separating 1 Å distance away for every successive window. To ensure the overlapping of consequent data sets in umbrella sampling, we have plotted the distributions of data obtained from the simulations. We found sufficient overlapping of data sets. The graphs for distribution versus COM distance have been included in the supplementary figures S8 and S9. Figure 8 shows the change in free energy during the translation of virus CTD/RBD from active pocket of hACE2 for both complexes. The center of mass (COM) distance as a reaction coordinate allows us to track the free energy changes for SARS-CoV-2 CTD in complex with hACE2 and SAR-CoV RBD in complex with hACE2. Free energy has also been used to compare the differences in the binding affinity for the two complexes. The SARS-CoV-2 CTD in complex with hACE2 is found to have the greater binding free energy of ~1.91 kcal/mol compared to the SAR-CoV RBD in complex with hACE2. This, as well as the nature of the free-energy curve, provides an insight on binding mechanisms of the complexes.

4. Discussion

COVID-19 pandemic has seriously threatened public health throughout the globe. Since there is no approved drug till date to combat against the SARS-CoV-2 virus, more comprehensive study is essential through the
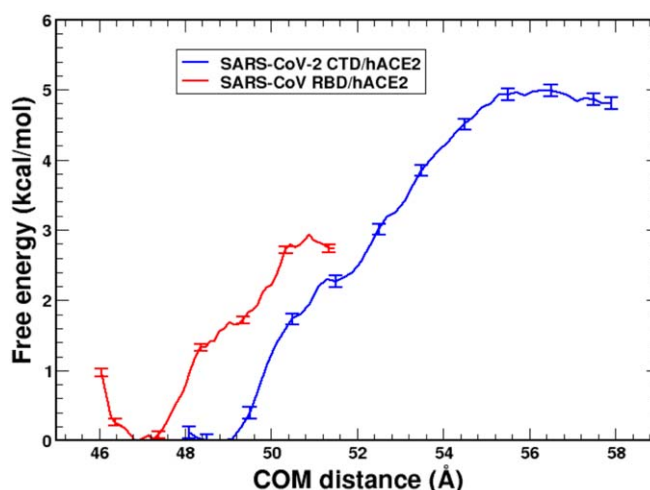
Figure 8. Free energy curve for SARS-CoV-2 CTD/hACE2 and SARS-CoV RBD/hACE2 during the translation of spikes CTD/RBD from hACE2.

various aspects at molecular level. The fundamental necessity is to understand the entry mechanism of the virus into the human cell, which is really helpful to discover the drug against the virus. To deal the entry mechanisms and dynamical characteristics of the virus cell in complex with hACE2 receptor, we used various computational techniques. C-Terminal Domain (CTD) of S1 subunit of spike protein, being the active interacting region, has been taken into consideration in SARS-CoV-2. We performed the comparative analysis of the key residues and atomic interactions responsible for the binding of the SARS-CoV-2 CTD and SARS-CoV RBD with human ACE2 receptor.

Estimation of structural variation during the simulation is the foremost judgement of molecular stability in molecular dynamics study. RMSD is the measure of stability of molecular structure in the cellular environment. Well equilibrated system with consistent RMSD ensures us to proceed for the further study of binding affinity and energy variations of the molecular complexes. Moreover, contact surface area between the molecules identifies the binding strength of the complex. Therefore, we have obtained the contact surface area of both complexes calculating the solvent accessible surface area (SASA). SASA has been determined from time evolution data generated from the 250 ns NPT run. Then, average value of contact area for both the systems have been presented in table 1 and are interpreted graphically in figure 2. Larger contact surface area in SARS-CoV-2 CTD/hACE2 complex depicts the stronger binding of this complex than that of SARS-CoV RBD/hACE2 [45].

Our results show considerable similarity in the binding sites, interfacial residues and important atomic interactions in both viral protein receptor binding domain (i.e., SARS-CoV-2 CTD and SARS-CoV RBD). However, there are some variations in loop between two structures in the binding region and some residues at the binding sites are different. This facilitates more and stronger atomic contacts between SARS-CoV-2 CTD and hACE2 interface and thereby enhancing its binding affinity. Polar residues residing at the interface form an extensive network of hydrogen bonds and salt-bridge interactions [46–49]. Our study reveals that interfacial hydrogen bonds, salt-bridges and hydrophobic interactions play an important role in the binding of SARS-CoV-2 CTD to host cell receptor. Furthermore, comparative analysis of the binding mechanism of two viral proteins with hACE2 show that binding affinity of SARS-CoV-2 is greater than that of SARS-CoV. Notably, more residues are engaged in the binding of SARS-CoV-2 CTD with hACE2. We find the greater number of potential hydrogen bonds formed in the case of SARS-CoV-2 CTD which contributes to higher binding affinity. More and stronger salt-bridges formed in case of SARS-CoV-2 CTD establish stronger binding to the receptor than SARS-CoV RBD. Additionally, we observe hydrophobic interactions are stronger in case of SARS-CoV-2 which also contribute to enhanced binding.

The contributions of electrostatic and vdw contacts are significant to form a stable protein-protein complex [50, 51]. The potential energy in binding the virus CTD/RBD and host receptor are compared in both the systems. Though, initially the electrostatic energy is observed relatively larger in SARS-CoV-2 CTD/hACE2 than that of SARS-CoV RBD/hACE2, the dynamical results show almost equal contributions in both the complexes. This shows that the contributions of hydrogen bonds, salt bridges and hydrophobic interactions are responsible to provide the greater binding strength in SARS-CoV-2 CTD/hACE2.

The binding mechanisms of the complexes are further analyzed to estimate the free energy differences from umbrella sampling method. SMD trajectories are taken for the appropriate samples that ensure the sufficient

overlapping on windows [52]. In SMD, the virus CTD/RBD are pulled upto that distance, beyond which no interactions persists. We find the interactions of molecules in complex SARS-CoV RBD/hACE2 have been terminated after the displacement of RBD by 5 Å from host receptor, whereas the interactions sustain upto 9 Å displacement from the initial position in SARS-CoV-2 CTD/hACE2. Comparisons of free energy of two complexes have provided the insight of bonding affinity between the virus CTD/RBD and hACE2 molecules. The greater free energy difference between SARS-CoV-2 CTD in complex with hACE2 depicts the stronger binding strength than the complex of SARS-CoV RBD and hACE2. As the further investigation, we plan to calculate the solvation free energy of SARS-CoV-2 and SARS-CoV molecule in the aqueous environment.

Acknowledgments

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

References

[1] Yan R *et al* 2020 Structural basis for the recognition of the 2019-nCoV by human ACE2 *Science* **367** 1444–8
[2] Guan W *et al* 2020 Clinical characteristics of coronavirus disease 2019 in China *N. Engl. J. Med.* **382** 1708–20
[3] Zhu N *et al* 2020 A novel coronavirus from patients with pneumonia in China, 2019 *N. Engl. J. Med.* **382** 727–33
[4] Wells C R *et al* 2020 Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak *Proc. Natl. Acad. Sci.* **117** 7504–9
[5] https://covid19.who.int/table.
[6] Folegatti P M *et al* 2020 Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial *The Lancet* **396** 887–97
[7] Shang J *et al* 2020 Clinical characteristics of coronavirus disease 2019 in China *Nature* **581** 221–4
[8] Hoffmann M *et al* 2020 SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor *Cell* **181** 271–80
[9] Lu G, Wang Q and Gao G F 2015 Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond *Trends Microbiol.* **23** 468–78
[10] Lu G *et al* 2013 Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26 *Nature* **500** 227–31
[11] Li F *et al* 2005 Structure of SARS coronavirus spike receptor-binding domain complexed with receptor *Science* **309** 1864–8
[12] Ali A and Vijayan A 2020 Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms *Sci. Rep.* **581** 1–12
[13] Spinello A Saltalamacchia A and Magistrato A 2020 Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atoms Simulations *J. Phys. Chem. Lett.* **11** 4785–90
[14] Wang Q *et al* 2020 Structural and functional basis of SARS-CoV-2 entry by using human ACE2 *Cell* **181** 894–904
[15] Lan J *et al* 2020 Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor *Nature* **581** 215–20
[16] Peng Y *et al* 2018 Predicting protein-DNA binding free energy change upon missense mutations using modified MM/PBSA approach: SAMPDI webserver *Bioinformatics* **34** 779–86
[17] Velázquez-Campoy A *et al* 2004 Isothermal titration calorimetry *Curr. Protoc. Cell Biol.* **23** 17–8
[18] Hillisch A, Lorenz M and Diekmann S 2001 Recent advances in FRET: distance determination in protein-DNA complexes *Curr. Opin. Struct. Biol.* **11** 201–7
[19] Campagne S, Gervais V and Milon A 2011 Nuclear magnetic resonance analysis of protein-DNA interactions *J. R. Soc. Interface.* **8** 1065–78
[20] Teh H *et al* 2007 Characterization of protein- DNA interactions using surface plasmon resonance spectroscopy with various assay schemes *Biochemistry* **46** 2127–35
[21] Donald J E, Chen W W and Shakhnovich E I 2007 Energetics of protein-DNA interactions *Nucleic Acids Res.* **35** 1039–74
[22] Jones S *et al* 1999 Protein-DNA interactions: a structural analysis *J. Mol. Biol.* **287** 877–96
[23] Zhang C *et al* 2005 A knowledge-based energy function for protein- ligand, protein- protein, and protein- DNA complexes *J. Med. Chem.* **48** 2325–35
[24] Torrie G M and Valleau J P 1977 Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling *J. Comput. Phys.* **23** 187–99
[25] Kästner J 2011 Umbrella sampling *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1** 932–42
[26] Zheng L, Chen M and Yan W 2008 Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems *Proc. Natl. Acad. Sci.* **105** 20227–32
[27] Yang M *et al* 2014 Combine umbrella sampling with integrated tempering method for efficient and accurate calculation of free energy changes of complex energy surface *J. Chem. Phys.* **141** 07B618_1

[28] Yang Y *et al* 2016 Efficient sampling over rough energy landscapes with high barriers: A combination of metadynamics with integrated tempering sampling *J. Chem. Phys.* **144** 094105

[29] Sun Z and Zhang J Z 2021 Thermodynamic insights of base flipping in TNA duplex: force fields, salt concentrations, and free-energy simulation methods *CCS Chemistry* **3** 1026–39

[30] Sun Z *et al* 2019 Sulfur-substitution-induced base flipping in the DNA duplex *Phys. Chem. Chem. Phys.* **21** 14923–40

[31] Phillips J *et al* 2005 Scalable molecular dynamics with NAMD *J. Comput. Chem.* **26** 1781–802

[32] Berman H *et al* 2000 The protein data bank *Nucleic Acids Res.* **28** 235–42

[33] Lee J *et al* 2016 CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field *J. Chem. Theory Comput.* **12** 405–13

[34] Jorgensen W *et al* 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926–35

[35] Huang J *et al* 2017 CHARMM36m: an improved force field for folded and intrinsically disordered proteins *Nat. Methods* **14** 71–71

[36] Harvey M J, Giupponi G and Fabritiis G De 2009 ACEMD: accelerating biomolecular dynamics in the microsecond time scale *Nat. Methods* **5** 1632–9

[37] Khanal S P, Kandel Y P and Adhikari N P 2019 Transport properties of zwitterion glycine, diglycine, and triglycine in water *AIP Adv.* **9** 065303–10

[38] Koirala R P, Bhusal H P, Khanal S P and Adhikari N P 2020 Effect of temperature on transport properties of cysteine in water *AIP Adv.* **10** 025122

[39] Wang X and Sun Z 2019 Determination of Base-Flipping Free-Energy Landscapes from Nonequilibrium Stratification *J. Chem. Inf. Model* **59** 2980–94

[40] Humphrey W *et al* 1996 VMD: visual molecular dynamics *J. Mol. Graph.* **14** 33–8

[41] Schrödinger L L C 2015 The PyMOL Molecular Graphics System *Version 1.8.*

[42] Scheurer M *et al* 2018 PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations *Biophys. J.* **114** 577–83

[43] Kumar S *et al* 1992 The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method *J. Comput. Chem.* **13** 1011–21

[44] Zou X *et al* 2012 Recognition of methylated DNA through methyl-CpG binding domain proteins *Nucleic Acids Res.* **40** 2747–58

[45] Ma B *et al* 2003 Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces *Proc. Natl. Acad. Sci.* **100** 5772–7

[46] Teague S J 2003 Implications of protein flexibility for drug discovery *Nat. Rev. Drug Discov.* **2** 527–41

[47] Chen J, Sawyer N and Rayan L 2013 Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area *Protein Sci.* **22** 510–5

[48] Jones S and Thornton J M 1997 Analysis of protein-protein interaction sites using surface patches *J. Mol. Biol.* **272** 121–32

[49] Xu D, Tsai C and Nussinov R 1997 Hydrogen bonds and salt bridges across protein-protein interfaces *Protein Eng.* **10** 999–1012

[50] Jones S *et al* 2003 Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins *Nucleic Acids Res.* **31** 7189–98

[51] DiStasio R A Jr., Gobre V V and Tkatchenko A 2014 Many-body van der Waals interactions in molecules and condensed matter *J. Phys. Condens. Matter* **26** 213202

[52] Isralewitz B *et al* 2001 Steered molecular dynamics investigations of protein function *J. Mol. Graph. Model.* **19** 13–25

BIBECHANA

Molecular dynamics study of structural properties of γ-aminobutyric acid (GABA)

Shyam P. Khanal, Rajendra Prasad Koirala, Esha Mishra, Narayan P. Adhikari[*]
Central Department of Physics, Tribhuvan University, Kirtipur, Nepal
[*]Email: narayan.adhikari@cdp.tu.edu.np

ABSTRACT

The study of structural conformation of Gamma-aminobutyric acid (GABA) exhibits its biological and chemical activities. The GABA molecule is responsible in neurotransmission from one neuron to another neuron and activates the ion channels to pass the chlorine and sodium ions in nerve cells. Its conformation in solid state and gas state are extremely different and it also shows five different conformations in aqueous solution. The study of its structure in such environment can reveal its activity in cellular environment. We have performed the classical molecular dynamics study of this system of GABA in aqueous medium to deal its structure. Radial distribution function (RDF) has been used to study the structural properties of the system.

1. Introduction

Gamma-aminobutyric acid (GABA), a chief inhibitory neurotransmitter, plays major role in reducing the neuronal excitability throughout the central nervous system [1-3]. It is mostly found in nervous system of highly developed brain of mammals. It is, in fact, a chemical messenger that transmits the signals across chemical synapses from one neuron to another neuron, gland cells or to the muscles [4]. It is also used to treat high blood pressure, stress and anxiety; and to stimulate the secretion of natural growth hormone of body. The disorder of GABA in the body may cause the neurologic and psychiatric conditions. In addition, GABA is also detected in the other part from central nervous system like intestines, kidneys, uterus, ovaries, lungs etc. So, it has several functions in body mechanisms [5].

GABA is synthesized from anion of glutamic acid, called glutamate, via the enzyme glutamate decarboxylase with pyridoxal phosphate as a cofactor [6]. Its chemical formula is $C_4H_9NO_2$ and molar mass 103.120 g/mol. It is a white microcrystalline powder with density 1.11 g/mL. It is soluble in water with solubility value 130 g/100 mL. The temperature difference between solid state

and the gas state is quite narrow, with melting point 203.7°C and boiling point 247.9°C [7,8].

It contains a primary amine group and a carboxylic acid functional group, due to which it is categorized in amino acid group; however, the amino group (-NH$_2$) does not link to the alpha carbon as the ordinary amino acid contains. That is why, the GABA is not incorporated into the protein molecule. It is mostly found in zwitterion form [9, 10]. The interesting characteristic of GABA is its nature of conformation with its surroundings. It is mostly found highly folded configuration in gas phase and extended form in solid phase. Its character is surprising in solvent. It has high solvent effect with five different conformation [11]. The biochemical functions of biomolecules are greatly influenced by their structural conformation. Since the GABA can be at different conformations in various phases and also in the surrounding conditions, its structural study makes the great sense to understand the neurotransmission in nervous system, the spinal cord and hyperpolarization condition in ion exchange process in nerve cells [12]. Several researches have been carried out to deal the structural conformations of GABA in aqueous solution; however its structural variations in the living body are still unclear. Solvation of GABA molecule in water resembles real body.

Ashby *et al*. studied about different interactions with binding site that involved in GABA binding in molecular level; and Zafar and Jabeen studied the structure and function of GABA transporters (GATs) using computational method [13, 14]. Also, many experimental studies about transport properties of GABA have been already performed. Umecky *et al.* and Yui *et al.* measured the binary diffusion coefficient of GABA in infinitesimal aqueous medium at different temperature using Taylor dispersion method [15,16]. Also, viscosity of aqueous solution of GABA was estimated by Romero and Beltron [17].

Molecular dynamics (MD) study can be considered as an alternative technique to study about many properties including structural analysis [18]. MD also provides guideline for experimental study. To our best knowledge, the structural properties of GABA in water have not been studied using molecular dynamics. We expect that this study will help to learn the structural conformation of the molecule in water environment.

In this paper, we have discussed methods and methodology in section 2. The results of the work are presented and discussed in section 3; and finally conclusions and concluding remarks are presented in section 4.
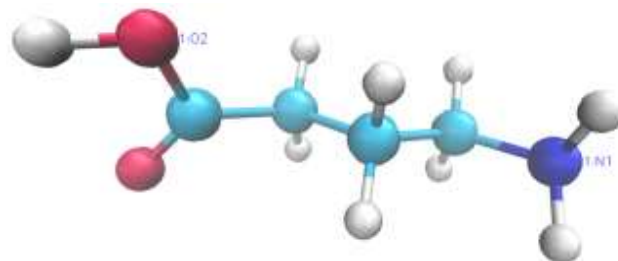
Fig. 1: Snapshot of GABA molecule.

2. Methods and Methodology

Modeling of system

We performed classical molecular dynamics simulations of the system of aqueous solution of GABA. GROMACS 5.1.1 software package was used for the simulations [19]. The OPLS-AA (optimized potentials for liquid simulations – all atom) force field for modeling the GABA molecule and three point SPC/E [20] water model were used during the simulations. In the classical molecular dynamics simulations, we solve the Newton's equation of motion [21]. To account the intra-molecular interactions, the bonded interaction i.e. bond stretching, bond angle and dihedral potentials are considered. And, the non-bonded Coulomb and Lennard-Jones (LJ) interactions contribute to the inter-molecular interactions. The Coulomb interaction arises due to partial charges of the

atoms in GABA and water molecules. For SPC/E water model, the partial charges of hydrogen and oxygen atoms are +0.4238e and -0.8476e respectively where e is elementary charge; and Lennard-Jones (LJ) parameters are 0.316 nm and 78.2k_B. Now, the inter-molecular interaction is caused due to non-bonded interactions which can be expressed as

$$V(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_m r_{ij}}$$

where r_{ij} is the distance between i^{th} & j^{th} atoms of charges q_i & q_j respectively, ε & σ are LJ parameters and ϵ_m is permittivity of medium between the charge.

Computational details

We performed the simulations of the system of 3 γ-aminobutyric acid (GABA) as solute and 1035 water molecules as solvent at 1 atm pressure at five different temperature: 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2. The simulations were carried out in cubic simulation box under periodic boundary conditions (PBC). At first, to remove van der Waals bad contact and to obtain the minimum potential energy state, energy minimization of the system was carried out using Steepest-descent method taking 50 kJ/mole-nm force tolerance [22, 23].

Many properties of the system under study depend upon the parameters like temperature, pressure etc. So, the system must be in thermodynamics equilibrium. For this, the system was equilibrated at each temperature for 200 ns time taking time step of 0.002 ps using isothermal-isobaric (NPT) ensemble. During equilibration run, velocity rescaling thermostat with 0.01 ps coupling time was used to control temperature and Beresdsen barostat with coupling time of 0.8 ps was used to keep constant pressure [24]. LINCS algorithm and Maxwell-Boltzmann distribution were taken to constraint all bonds and to assign initial velocities for each particle respectively [24]. Also, Particle Mesh Ewald (PME) method was chosen to account the long range Coulomb interaction; and cut off parameters of 1 nm was taken for both short range Coulomb & Lennard-Jones (LJ) interactions. In order to solve the equations of motion, Leap-frog algorithm was used [24]. Figure 2 represents the temperature and density profiles of the system after equilibration run at temperature of 303.2 K. Also, the simulated values of temperature and density are presented in the table 1.
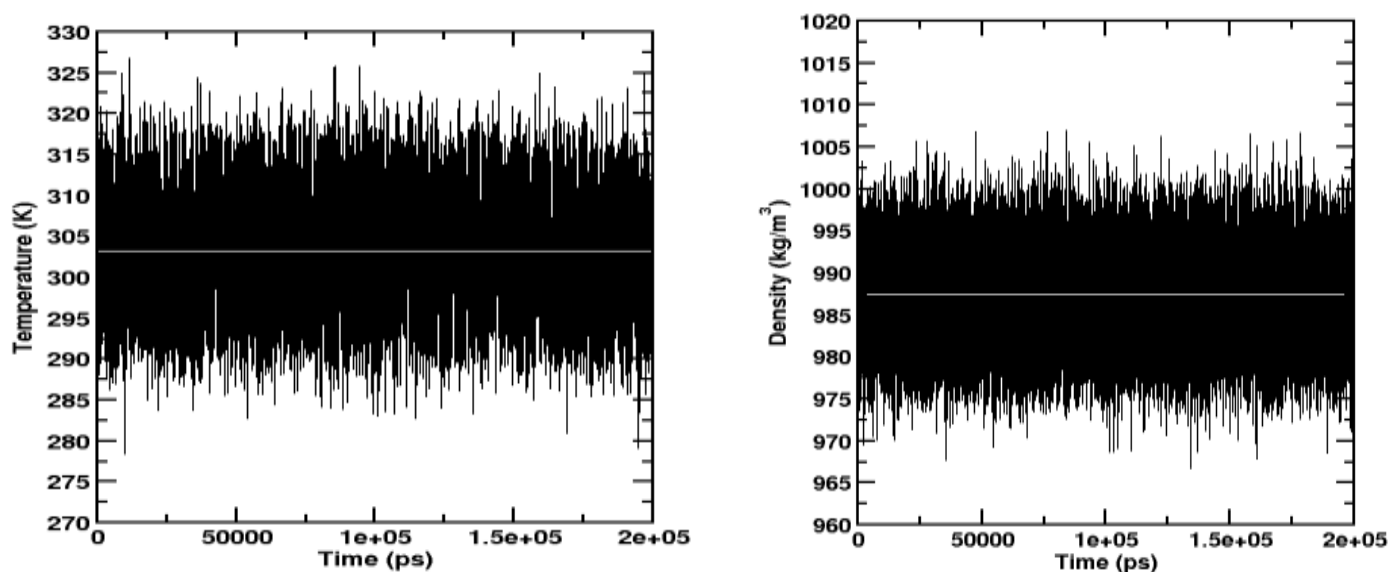

Fig. 2: Temperature (left) and density (right) profiles of the system after equilibration run at 303.2 K temperature.

69

Table 1: Simulated values of temperature and density at five different coupling temperature.

Coupling Temperature (K)	Simulated Temperature (K)	Simulated Density (kg/m³)	Experimental Density (kg/m³)[25]
298.2	298.20±0.01	990.07±0.02	997.03
303.2	303.19±0.00	987.49±0.02	995.63
313.2	313.19±0.04	981.46±0.03	992.19
323.2	323.19±0.01	974.45±0.02	988.19
333.2	333.19±0.02	967.19±0.02	983.17

From the table , it is seen that the simulated values of densities at different coupling temperature agree within 2% with previously reported experimental values.

After the equilibrating the system, the production run was done at each temperature using canonical (NVT) ensemble. During the production run, velocity rescaling thermostat with coupling time of 0.01 ps was used to control temperature. The velocities of final step of equilibration run were taken as initial velocities for production run. Each production was performed for 100 ns with time step 0.002 ps.

3. Results and Discussion

In this section, we present the RDF between different atoms of GABA and Water molecules.

Radial Distribution Function (RDF)

Structural properties of the system has been studied by using radial distribution function (RDF). RDF, that provides the idea of distribution of molecules around another molecule which is taken as reference, gives the probability of finding a pair of atoms located at distance `r' [26]. For liquid, RDF shows an oscillation up to certain distance and becomes unity which means that there is no correlation between molecules after the distance [27, 28]. The GABA molecule contains the functional groups amide ($-NH_2$) and carboxyl (-COOH). Thus, to find the structural properties of the system under study, we have calculated the RDF of oxygen of water & oxygen of water ($g_{OW-OW}(r)$), oxygen of water & nitrogen of amide ($-NH_2$) group of GABA ($g_{OW-N1}(r)$), and oxygen of water & oxygen of carboxylic group (-COOH) of GABA ($g_{OW-O2}(r)$). Figures 3, 4 and 5 represent the $g_{OW-OW}(r)$, $g_{OW-N1}(r)$ and $g_{OW-O2}(r)$ at five different temperature 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2 K respectively.

In the RDF plots, there is a region from the reference atom up to which the value of RDF is zero. In this region, the probability of finding another atom is zero. Such region is known as excluded region (ER). Beyond the zero probability region, some peaks are observed and the value of RDF becomes unity beyond certain distance from reference atom. The unity value of RDF indicates that no correlation between atoms at that region i.e. correlation between atoms takes place up to certain distance from reference position [23]. Three peaks are observed between excluded and unity regions. The peaks indicate the favorable position of the atoms/molecules from reference. The first peak means the most favorable position of the atoms from reference. The values of excluded region (ER), first peak position (FPP), first peak value (FPV)), second peak position (SPP), second peak value (SPV)), third peak position (TPP) and third peak value (FPV) for $g_{OW-OW}(r)$, $g_{OW-N1}(r)$ and $g_{OW-O2}(r)$ are presented in Tables 2, 3, and 4 respectively.
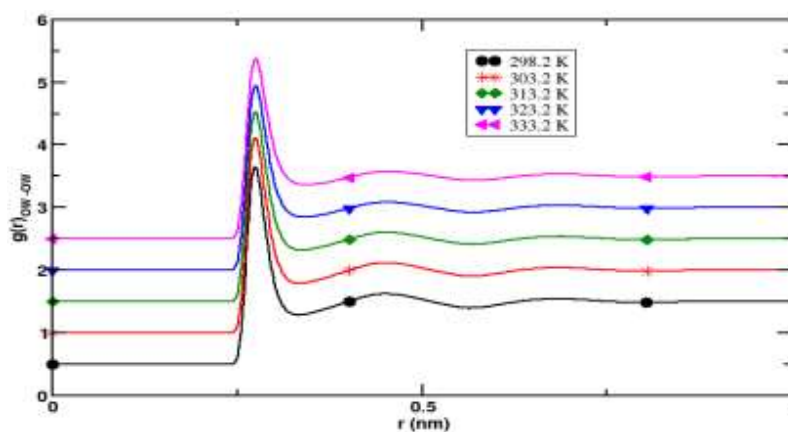
Fig. 3: RDF between oxygen atoms of water (g $_{OW-OW}$(r)) at different temperature.
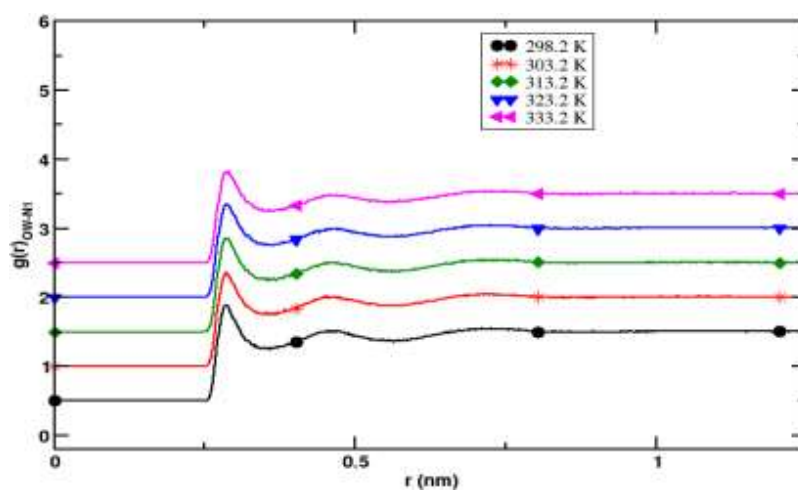

Fig. 4: RDF between nitrogen of GABA and oxygen of water (g $_{OW-N1}$(r)) at different temperature.
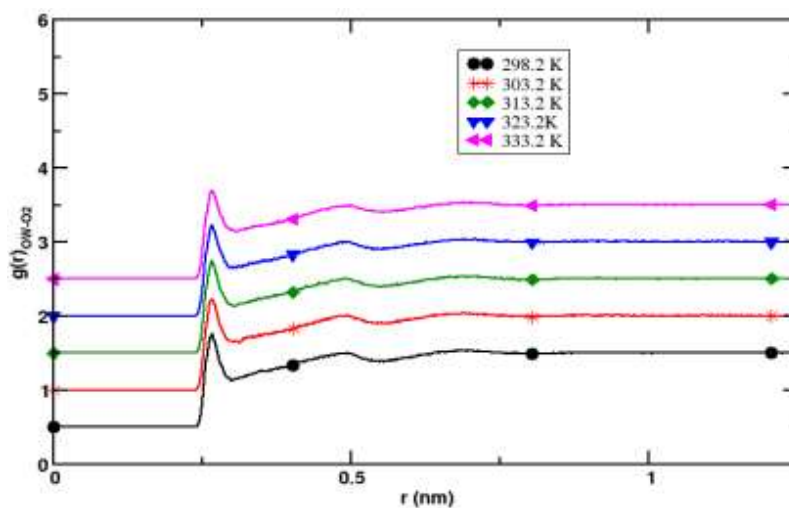

Fig. 5: RDF between oxygen of GABA and oxygen of water (g $_{OW-O2}$(r)) at different temperature.

Table 2: Simulated data for Radial distribution function (RDF) between water molecules (g $_{OW-OW}$(r)) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.240	0.274	3.141	0.450	1.126	0.686	1.046
303.2	0.240	0.276	3.104	0.450	1.120	0.690	1.040
313.2	0.240	0.276	3.018	0.450	1.103	0.688	1.037
323.2	0.240	0.276	2.936	0.450	1.087	0.694	1.036
333.2	0.240	0.276	2.874	0.450	1.075	0.686	1.034

Table 3: Simulated data for Radial distribution function (RDF) between nitrogen (N1) of GABA and water molecules (g $_{OW-N1}$(r)) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.248	0.284	1.379	0.466	1.016	0.726	1.047
303.2	0.248	0.286	1.352	0.452	1.017	0.712	1.047
313.2	0.246	0.284	1.359	0.470	1.004	0.730	1.048
323.2	0.248	0.286	1.344	0.468	0.997	0.724	1.045
333.2	0.248	0.290	1.321	0.458	0.985	0.726	1.040

Table 4: Simulated data for Radial distribution function (RDF) between oxygen (O2) of GABA and water molecules (g $_{OW-O2}$(r)) at different temperature.

Temperature(K)	ER(nm)	FPP(nm)	FPV	SPP(nm)	SPV	TPP(nm)	TPV
298.2	0.234	0.266	1.258	0.494	1.003	0.700	1.039
303.2	0.236	0.266	1.231	0.494	1.008	0.700	1.039
313.2	0.234	0.266	1.242	0.490	1.006	0.684	1.034
323.2	0.234	0.266	1.228	0.494	0.999	0.690	1.037
333.2	0.234	0.266	1.188	0.488	0.991	0.684	1.031

The values of Lennard-Jones parameter (σ) for OW-OW, OW-N1 and OW-O2 are 0.316 nm, 0.323 nm and 0.308 nm respectively. And, the calculated values of van der Waals radius ($2^{1/6}$σ) for OW-OW, OW-N1 and OW-O2 are 0.355 nm, 0.363 nm and 0.346 nm respectively. From the Tables 2-4, it is clearly observed that the values of first peak position (FPP) are smaller than the respective van der Waals radius, which indicates that other potentials along with Lennard-Jones (LJ) also contribute for stability of the system. On the other hand, Coulomb interactions arise in the system due to partial charges of hydrogen and oxygen of

SPC/E water model, and of N1 from -NH2 & O2 from -COOH group of GABA molecule. Thus, both Coulomb as well as LJ including many body effects are responsible for structural properties of the system [19, 20]. From the Table, it is also observed that the values of excluded region (ER) are less than corresponding values of the van der Waals radius.

4. Conclusions and concluding remarks

We performed classical molecular dynamics of a system of 3 γ-aminobutyric acid (GABA) as solute and 1035 water as solvent at five different

temperature: 298.2 K, 303.2 K, 313.2 K, 323.2 K and 333.2 K using GROMACS 5.1.1 package. During the simulations, SPC/E water model and OPLS-AA force field parameters were used. Radial distribution function (RDF) was taken to analyze the structural properties of the system. For this, we plotted the RDF between different atoms at five different temperature i.e. (i) oxygen atoms of water molecules ($g_{OW-OW}(r)$), (ii) nitrogen of –NH2 group of GABA and oxygen of water ($g_{OW-N1}(r)$) and (iii) oxygen of –COOH group of GABA and oxygen of water ($g_{OW-O2}(r)$). From all the plots, we observe that both excluded region (ER) as well as first peak position (FPP) are smaller than the corresponding van der Waals radius. This indicates that along with Lennard Jones (LJ), Coulomb potential which arises due to the partial charges of different atoms as well as many body effects also contribute to the structural properties of the system.

In near future, we intend to study about the free energy calculation of the system.

Acknowledgements

References

[1] H. Lodish et al. Neurotransmitters, synapses, and impulse transmission, In Molecular Cell Biology, 4th edition. WH Freeman, (2000).

[2] L, Sivilotti & A. Nistri, = GABA receptor mechanisms in the central nervous system, Progress in Neurobiology 36(1) (1991) 35-92.
https://doi.org/10.1016/0301-0082(91)90036-Z

[3] K. Łątka, J. Jończyk, and M. Bajda, Structure modeling of γ-aminobutyric acid transporters– Molecular basics of ligand selectivity, Inter. J. Biol. Macromol (2020).
https://doi.org/10.1016/j.ijbiomac.2020.04.263

[4] T. A. Rocheleau, J. C. Steichen & A. E. Chalmers, A point mutation in a Drosophila GABA receptor confers insecticide resistance, Nature 363(6428) (1993) 449-451.
https://doi.org/10.1038/363449a0

[5] D. Ma et al., Structure and mechanism of a glutamate–GABA antiporter, Nature 483 (7391) (2012) 632-636.
https://doi.org/10.1038/nature10917

[6] P. Rorsmanet et al., Glucose-inhibition of glucagon secretion involves activation of GABA A-receptor chloride channels, Nature 341(6239) (1989) 233-236.
https://doi.org/10.1038/341233a0

[7] A. J. Dobson & R. E. Gerkin, γ-Aminobutyric acid: a novel tetragonal phase, Acta Crystallographica Section C: Crystal Structure Communications 52(12) (1996) 3075-3078.
https://doi.org/10.1107/S0108270196010001

[8] O. A. Petroff, Book review: GABA and glutamate in the human brain, The Neuroscientist 8(6) (2002) 562-573.
https://doi.org/10.1177/1073858402238515

[9] A. Schousboe & H. S. Waagepetersen, GABA: homeostatic and pharmacological aspects, Progress in Brain Research 160 (2007) 9-19.
https://doi.org/10.1016/S0079-6123(06)60002-2

[10] N. Ottosson, M. Pastorczak, S. T. van der Post, & H. J. Bakker, Conformation of the neurotransmitter γ-aminobutyric acid in liquid water, Physical Chemistry Chemical Physics 16(22) (2014) 10433-10437.
https://doi.org/10.1039/C4CP00671B

[11] G. Deniau, et al., Synthesis, Conformation and Biological Evaluation of the Enantiomers of 3-Fluoro-γ-Aminobutyric Acid ((R)-and (S)-3F-GABA): An Analogue of the Neurotransmitter GABA, Chem Bio Chem 8(18) (2007) 2265-2274.
https://doi.org/10.1002/cbic.200700371

[12] C. Lee, & A. J. de Silva, Interaction of Neuromuscular Blocking Effects of Neomycin and Polymyxin B, Anesthesiology: The Journal of the American Society of Anesthesiologists 50(3) (1979) 218-220.
https://doi.org/10.1097/00000542-197903000-0001

[13] J. A. Ashby, et al., GABA binding to an insect GABA receptor: a molecular dynamics and mutagenesis study, Biophysical Journal 103(2012) 2071-2081.
https://doi.org/10.1016/j.bpj.2012.10.016

[14] S. Zafar, and I.t Jabeen, Structure, function, and modulation of γ-aminobutyric acid transporter 1 (GAT1) in neurological disorders: a pharmacoinformatic prospective, Frontiers in Chemistry 6 (2018) 397.
https://doi.org/10.3389/fchem.2018.00397

[15] T. Umecky, K. Ehara, S. Omori, T. Kuga, K. Yui, and T. Funazukur, Binary diffusion coefficients of aqueous phenylalanine, tyrosine isomers, and aminobutyric acids at infinitesimal concentration and temperatures from (293.2 to 333.2) K, Journal of Chemical & Engineering Data 58 (2013) 1909-1917.
https://doi.org/10.1021/je3012698

[16] K. Yui, Y. Noda, M. Koido, M. Irie, I. Watanabe, T. Umecky and T. Funazukur, Binary Diffusion Coefficients of Aqueous Straight-Chain Amino Acids at Infinitesimal Concentration and Temperatures from (298.2 to 333.2) K, Journal of Chemical & Engineering Data 58 (2013) 2848-2853.
https://doi.org/10.1021/je301370s

[17] C. M. Romero and A. Beltrán, Effect of temperature and concentration on the viscosity of aqueous solutions of 3-aminopropanoic acid, 4-aminobutanoic…, Revista Colombiana de Química 41 (2012) 123-131.

[18] M. P. Allen and D. Tildesley, Computer Simulation of Liquids New York, USA: Oxford University Press, 1991.

[19] D. van der Spoel, et al., GROMACS: fast, flexible, and free, Inc. J Comput Chem 26 (2005) 1701-1718.
https://doi.org/10.1002/jcc.20291

[20] H. J. C. Berendsen J. R. Grigera, and T. P. Straatsma, The missing term in effective pair potentials, Journal of Physical Chemistry 91 (1987) 6269-6271.

[21] M. E.Tuckerman, Mark, Statistical mechanics: theory and molecular simulation, Oxford university press, 2010.

[22] S. P. Khanal, Y. P. Kandel, and N. P. Adhikari, Transport properties of zwitterion glycine, diglycine, and triglycine in water, AIP Advances 9 (2019) 065303.
https://doi.org/10.1063/1.5099069

[23] R. P. Koirala, et al, Effect of temperature on transport properties of cysteine in water, AIP Advances 10 (2020) 025122.
https://doi.org/10.1063/1.5132777

[24] M. J. Abraham, et al., GROMACS User Manual Version 5.1. 1, GROMACS Development Team (2016).

[25] www.mrbigler.com/misc/Pvap-H2O.PDF

[26] D. A. McQuarrie, Statistical Mechanics, Harper and Row, New York 1976.

[27] S. Pokharel, N. Pantha, and N. P. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, International Journal of Modern Physics B 30 (2016) 1650205.
https://doi.org/10.1142/S0217979216502052

[28] K. Sharma and N. P. Adhikari, Temperature dependence of diffusion coefficient of nitrogen gas in water: A molecular dynamics study, International Journal of Modern Physics B 28 (2014) 1450084.
https://doi.org/10.1142/S0217979214500842

Solvation Free Energy of Protonated Lysine: Molecular Dynamics Study

S. P. Khanal, B. Poudel, R. P. Koirala and N. P. Adhikari[*]
Central Department of Physics, Tribhuvan University, Kathmandu, Nepal
[*]Corresponding Email: narayan.adhikari@cdp.tu.edu.np.

ABSTRACT

In the present work, we have used an alchemical approach for calculating solvation free energy of protonated lysine in water from molecular dynamics simulations. These approaches use a non-physical pathway between two end states in order to compute free energy difference from the set of simulations. The solute is modeled using bonded and non-bonded interactions described by OPLS-AA potential, while four different water models: TIP3P, SPC, SPC/E and TIP4P are used. The free energy of solvation of protonated lysine in water has been estimated using thermodynamic integration, free energy perturbation, and Bennett acceptance ratio methods at 310 K temperature. The contributions to the free energy due to van der Waals and electrostatics parameters are also separately computed. The estimated values of free energy of solvation using different methods are in well agreement with previously reported experimental value within 14 %.

Key words: Molecular dynamics simulations, Solvation free energy, Free energy perturbation and Thermodynamic integration.

1. INTRODUCTION

Solvation free energy plays a central role in protein folding, protein function and molecular recognition [1, 2]. Theoretical and computer simulation inspection on the thermodynamic properties of amino acids and the role of free energy in particular, in this context, become very important in a broad range of fields from chemistry, biology, and pharmaceuticals. Such studies can pave the way for identification of pharmacological targets as well as in the drug discovery [3]. Since most of the biological processes happen in an aqueous solution, this is why solvation free energy is equivalently referred to as hydration free energy that originates from the interactions between the solute and solvent i.e., water. Additionally, free energy calculations often provide an efficient route to estimate kinetic and dynamic characteristics of bio-chemical and physical processes, such as solubility, reaction rates, partition coefficients, associations, dissociations, and binding constants [4].

Number of experimental studies has been carried out to determine the solvation properties of amino acids in past decades [5–8]. Because of the very different physico-chemical properties among the naturally occurring amino acids, the experimental techniques may not be free from intrinsic errors. Therefore, it is highly desirable to complement experimental studies with theoretical approaches using molecular dynamics simulations with explicit solvent molecules [9]. Several researchers performed the molecular dynamics simulations [10–12] and even monte carlo simulations [13] to determine the solvation free energies of amino acids in different solvent environments using variety of available force fields. These calculations show that it is possible to reproduce the experimental solvation free energies by modeling different interactions between the solute and solvents. Motivated from aforementioned studies, we have chosen the protonated lysine system for the calculation of solvation free energy in aqueous solution.

Lysine, an essential amino acid used in biosynthesis of proteins, is harvested from external food stuffs. Also, it is required for growth, tissue repair and improves immune system [14]. Going

through the literature for our case, the experimental value of the solvation free energy of protonated form of lysine in water is known to be 277.80 kJmol^{-1} [15] and value 246.22 kJmol^{-1} have been predicted with use of molecular dynamics simulations [16]. To our best knowledge, there are no simulations have been done in order to estimate the solvation properties of lysine in it's protonated form using Molecular Dynamics (MD) simulations with implementation of variety of Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) based methods in aqueous solution. In this work, we have carried out MD simulations to estimate free energy of solvation of the protonated lysine in aqueous medium. The simulations have performed taking four different water models: TIP3P, SPC, SPC/E and TIP4P. Then, the solvation free energy has been estimated using TI and FEP methods. The obtained results are compared with available previously reported data. A standard comparison is made between each method with potential sources of errors. The solvation free energies due to change in van der Waals parameters and electrostatic parameters are also computed individually.
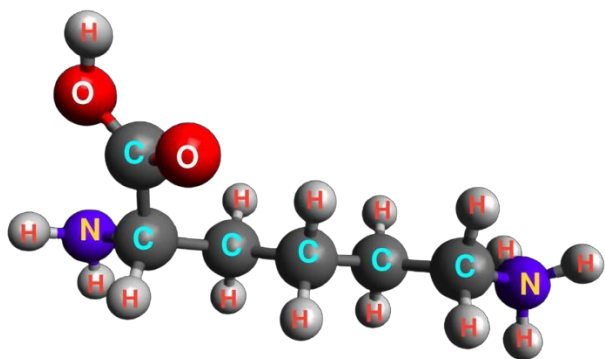

Fig. 1: Protonated lysine molecule.

2. THEORY

The free energy difference between two states say A and B of a system can be calculated using thermodynamic integration (TI) method as [17]:

$$\Delta F_{AB} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \text{ ... (1)}$$

Here, the parameter λ is used to define intermediate states between initial and final states. For this the potential energy is defined such that it is also a function of the coupling parameter λ. The equation (1) can be evaluated for the ensemble average at a number of discrete λ-points by performing series of simulations for each chosen λ-point. We then use numerical integration to determine the integral.

When the free energy difference between two states of a system is small, another approach based on perturbation, called free energy perturbation (FEP) method, can be used [18]. According to this method, the free energy difference, is defined by [19].

$$\Delta F_{AB} = k_B T ln \left\langle e^{-\beta[U(\lambda_B) - U(\lambda_A)]} \right\rangle_{\lambda_A} \text{ ... (2)}$$

Here, $\beta = (k_B T)^{-1}$, where k_B is the Boltzmann constant and T is absolute temperature. In this method, to obtain convergence, significant overlap of the low energy regions of the two ensembles is required.

The asymmetric biased arises in equation (2) due to the configuration being sampled either via λ_A or λ_B can be removed by so called Bennett Acceptance Ratio (BAR) method [20]. The BAR method requires sampling and energy evaluation of the system configurations from both states to estimate the free energy difference [21].

$$\sum_{i=1}^{n_i} \frac{1}{1+exp\left[ln\left(\frac{n_i}{n_j}\right)+\beta\left(\Delta U_{ij}-\Delta A\right)\right]} = \sum_{j=1}^{n_j} \frac{1}{1+exp\left[ln\left(\frac{n_j}{n_i}\right)+\beta\left(\Delta U_{ij}-\Delta A\right)\right]} \text{ ...(3)}$$

The expression (3) minimizes the free energy variance and makes BAR more efficient [22]. In order to increase the overlap between each pair of end states, free energy differences are usually calculated by introducing number of intermediate states in addition to the two end states. A multistate extension of BAR, called the Multistate Bennetts Acceptance Ratio (MBAR), has been proposed [23]. In this approach, a series of weighting functions are derived to minimize the uncertainties in free energy differences between all states considered simultaneously. For the case when only two states are considered MBAR reduces to BAR. Among all the methods discussed here, MBAR has the lowest variance, and is apparently the most decisive estimator of the solvation free energy calculations [23, 24].

3. COMPUTATIONAL DETAILS AND METHODOLOGY

In this work, lysine in protonated form was used as solute and four different models of water: transferable intermolecular potential with 3 points (TIP3P), simple point charge (SPC), extended simple point charge (SPC/E) and transferable intermolecular potential with 4 points (TIP4P) were used as solvent [25–27]. The geometry of solute was mimicked by using Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force field parameters [28]. To prepare the simulation cell, a cubical box of size 3 nm was taken. The solute was placed at the center of this box and filled with 875 water molecules at the experimental density. Simulation was performed under periodic boundary conditions (PBC) at 310 K using GROMACS (5.1.2) package [29, 30].

The approach to find the solvation free energy is by turning off the interactions; bonded and non-bonded, between the solute and the solvent molecules. For our case, we only manipulate non-bonded interactions. In determining the solvation free energy, we can use the fact that free energy is a state function; it is independent of the path taken for the transformation going from state A to state B. For this, we define the potential of the system as a function of two parameters λ_{vdW} for van der Waals potential, and λ_{ele} for electrostatic potential. Thus, the complete description of our system for state A ($\lambda = 0$), where there was interaction (couple) between the solute and solvent, and for state B ($\lambda = 1$), where there was no interaction (decouple) between the solute and solvent. This was done in our case by scaling the system by taking 21 windows for values of λ between 0 and 1. The 21 different values of coupling parameter for Coulombic and Lennard-Jones interactions were defined as: λ_C = 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00 and λ_{LJ} = 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.75, 0.80, 0.90, 1.00 respectively.

At first, the energy minimization was carried out using steepest-descent method [31] with tolerance force set to 10 kJ mol^{-1} nm^{-1}. To generate the initial velocities Maxwell-Boltzmann distribution was used. The Brownian dynamics friction coefficient was set to zero with random speed -1. The leap-frog stochastic dynamics integrator [32] was used to integrate the equations of motion. The temperature was kept constant at 310 K by using the Langevin

thermostat [33]. The pressure was maintained constant by coupling to a reference pressure of 1 bar using the Parrinello-Rahman barostat [34] with compressibility 4.5 × 10^{-5} bar^{-1}. For the simulations, the coupling time for both thermostat and barostat was set to1 ps. The time step used in the simulations was 2 fs throughout. A 5 ns simulation was performed to equilibrate each of the systems before the start of the actual free energy calculations.

In this study, a neighbor list of 1.2 nm updated every twenty steps was used for the short range interactions. Particle Mesh Ewald (PME) was used to evaluate the Coulombic interactions, with a real space cutoff of 1.2 nm and a PME order of 6 [35]. The Fourier spacing was chosen to be as close to 0.12 nm as possible given the box size and the need for integer numbers of grid points. The distance for van der Waals cutoff was set to 1 nm. All the bond angles were constrained using the LINCS algorithm [36]. For the calculation of free energies, we gave the final production run of 5 ns with time step 2 fs. During the production run, the system was coupled only using temperature. In the TI, in order to remove the singularities in the potentials, we used λ dependence of potentials; the soft-core potential [37].

4. RESULTS AND DISCUSSION

The solvation free energy of protonated lysine in water has been estimated at 310 K using TI and FEP methods [38]. In TI method, the free energy change for a path composed of 'm' states is computed as a weighted sum of the ensemble averages of the derivative of potential energy function with respect to coupling parameter λ. There are different approaches available for numerical integration of TI. But in our calculation, we have implemented TI-1 and TI-3 which use the trapezoidal rule (a first-order polynomial) and a cubic spline respectively. The nature of these TI methods depends on the nature of the curve being integrated and hence, it depends on underlying data and the shape of the $\partial U/\partial \lambda$ the alchemical path chosen. Perturbation based methods include a broad range of techniques. The direction dependent transformation of FEP originates from under-sampling the tail regions of the potential distributions, which results biased free energy [39]. We have used BAR method that uses samples of potential energy in both direction to obtain the minimum free energy variance. In BAR, the free energy change between two adjacent states is computed to yield the minimum variance and gives

data for single pair of states, while another class of BAR is MBAR that finds the best estimation of free energy changes between all states simultaneously by optimizing the matrix of the ΔA variance.

The BAR method provides a maximum likelihood estimation of the free energy that is given by the samples from the two states. The BAR requires significantly less phase space overlap between these states in order to converge results as compared to other methods [40, 41]. Note, however, that BAR requires sampling and energy evaluation of the system configurations from both states to estimate the free energy difference. As phase space overlap affects the reliability of the estimate, free energy differences are most often calculated by simulating several intermediate states in addition to the two end states, in order to increase the overlap between each pair of states. A multistage extension of BAR method called as MBAR has been devised [23]. In this approach, a series of weighting functions are derived to minimize the uncertainties in free energy differences between all states considered simultaneously. MBAR reduces to BAR when only two states are considered.

To estimate the solvation free energy of our system, we have first plotted the $\partial U/\partial \lambda$ as a function of coupling parameter λ.

different water models: TIP3P, SPC, SPC/E and TIP4P calculated from TI, TI-CUBIC, BAR and MBAR with previously reported experimental value are presented in the Table (1).
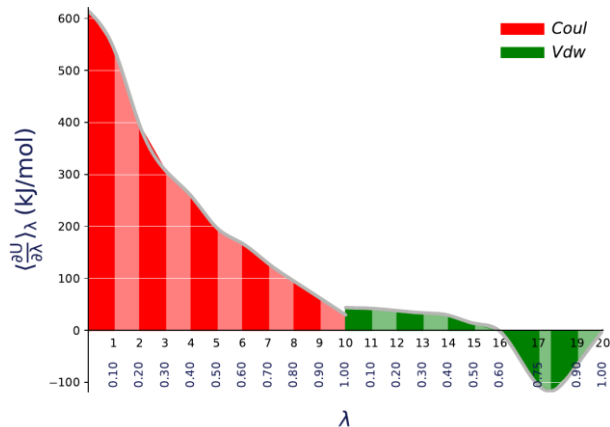
Fig. 3: Variation of $\partial U/\partial \lambda$ as a function of λ taking SPC/E water model as solvent.
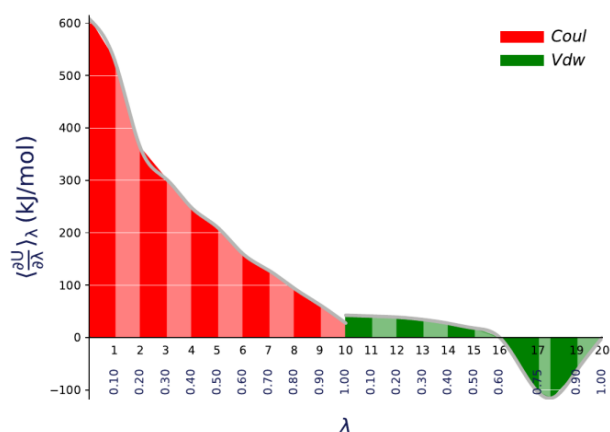
Fig. 4: Variation of $\partial U/\partial \lambda$ as a function of λ taking SPC water model as solvent.

Fig. 2: Variation of $\partial U/\partial \lambda$ as a function of λ taking TIP3P water model as solvent.

Figures 2, 3, 4 and 5 represent the variation of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle$ as function of coupling parameter λ taking TIP3P, SPC/E, SPC and TIP4P water models as solvent respectively. The estimated values of solvation free energy of protonated lysine in four

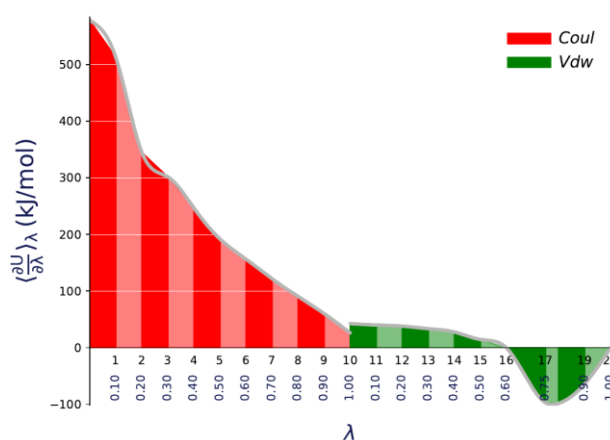**Fig. 5:** Variation of $\partial U/\partial \lambda$ as a function of λ taking TIP4P water model as solvent.

Table 1: Estimated values of solvation free energies for protonated lysine in different water models: TIP3P, SPC/E, SPC and TIP4P at 310 K temperature using TI, TI-CUBIC, BAR and MBAR methods with previously reported experimental value.

Water models	Interactions	TI (kj-mol^{-1})	TI-CUBIC (kj-mol^{-1})	BAR (kj-mol^{-1})	MBAR (kj-mol^{-1})	Expt.[15] (kj-mol^{-1})
TIP3P	Coulomb	247.36±0.35	247.23±0.35	247.37±0.22	247.71±0.24	
	vdW	-5.45±0.18	-5.61±0.20	-4.53±0.16	-4.42±0.18	
	Total	241.92±0.40	241.62±0.40	242.84±0.27	243.29±0.38	
SPC/E	Coulomb	247.22±0.	247.07±0.40	247.52±0.26	247.51±0.26	
	vdW	-8.44±0.24	-8.58±0.27	-6.91±0.22	-6.96±0.2.40	277.80
	Total	238.78±0.47	238.49±0.48	240.60±0.34	240.55±0.36	
SPC	Coulomb	241.12±0.61	241.02±0.63	238.88±0.35	238.82±0.41	
	vdW	-7.42±0.19	-7.56±0.22	-6.14±0.19	-6.09±0.21	
	Total	233.70±0.64	233.46±0.67	232.73±0.40	232.74±0.46	
TIP4P	Coulomb	232.70±0.43	232.76±0.42	238.33±0.27	240.90±0.17	
	vdW	-6.08±0.22	-6.61±0.25	-5.08±0.20	-4.99±0.22	
	Total	226.62±0.49	226.14±0.49	233.25±0.33	235.91±0.27	

Since the process we conducted in this work was the decoupling of protonated lysine in water, so the positive sign appears in each values. The reverse process is also possible. For our simplicity we just take the absolute value. The free energy change from λ = 0 to λ = 1 is simply the sum of the free energy changes of each pair of neighboring λ simulations. For all four water models: TIP3P, SPC, SPC/E & TIP4P, we have also estimated the individual contributions of electrostatic and vdW interactions to the solvation free energy. From the Table 1, it is seen that the contribution of vdW to solvation free energy is negative but of Coulombic interaction is positive. Also, in the current alchemical transformation protocol, the vdW component of the solvation free energy has found to be far less than that of the electrostatic component. From this observation, we have concluded that the solvation free energy of protonated lysine in water is solely due to the electrostatic component. Also, it is seen that the estimated value of solvation free energy is closer to the experimental value for TIP3P water model than that of other models.

For the TIP3P model, the estimated value of solvation free energy using MBAR method is in close agreement to the previously reported value rather than the values estimated using other methods. The estimated value using MBAR is in 12% agreements with the previously reported value. Also, other methods show a maximum difference of > 2 kJ/mol from the value estimated using MBAR method. We also observed from the Table 1 that the solvation free energy in each water model has almost same values using different methods indicated that λ spacing was sufficient for our sampling distributions which were found to be overlapped properly.

Furthermore, the high value of solvation free energy suggests that the lysine in its protonated form is highly soluble in water. This hydrophilic nature of lysine can be explained as follows: Lysine is a simple basic amino acid. In spite of a long and potentially hydrophobic chain, it has a basic NH_2 at the end of side chain. In its basic deprotonated form, lysine is neutral and hydrophilic; however, if found in physiological pH, lysine will pick up an H$^+$ from solution to form an NH_3^+ salt. Salts are charged and therefore definitely hydrophilic.

5. CONCLUSIONS AND CONCLUDING REMARKS

In this work, we performed molecular dynamics (MD) simulation to estimate the solvation free energy of protonated lysine in aqueous medium. OPLS-AA force field parameters and four different water models: TIP3P, SPC, SPC/E and TIP4P were used during the simulations. We have used TI, TI-cubic, BAR and MBAR methods to estimate the salvation free energy. It has been observed that the estimated values of free energy of solvation of protonated lysine with TIP3P water as solvent using different methods have closer values with experimental value in comparison to other models. The estimated values of free energy of solvation of protonated lysine in TIP3P water model at 310 K temperature are 241.92, 241.62, 242.84 and 243.29 in kJ/mol from TI, TI-cubic, BAR and MBAR methods respectively. Obtained numerical values of free energies demonstrated that all these methods are able to reproduce experimental free energies of solvation in water solvent. We have also analyzed the contribution of van der Waals (vdW) and electrostatic interactions to estimate the free energy of solvation; and it has been observed that the electrostatic interaction has major contributions to the solvation free energy. The estimated values of free energy of salvation in TIP3P water model using different methods are in agreement with previously reported experimental value within 14%. To extend this work in near future, we plan to study the solvation free energy of lysine in other solvent environment. Further, we plan to study the solvation free energy of lysine peptides and observe the effect on solubility with increase in chain length of solute.

REFERENCES

[1] Prigogine, I.; and S. A. Rice, *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics,* **148,** John Wiley & Sons (2009).

[2] Levy, Y.; and Onuchic, J. N. Water mediation in protein folding and molecular recognition, *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 389-415 (2006).

[3] Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; and Friesner, R. A. Advancing drug discovery through enhanced free energy calculations, *Accounts of chemical research*, **50**(7): 1625-1632 (2017).

[4] Straatsma, T. P.; Berendsen, H. J. C.; and Postma, J. P. M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water, *The Journal of chemical physics*, **85**(11): 6720-6727 (1986).

[5] Wolfenden, R.; Andersson, L.; Cullis, P. M.; and Southgate, C. C. B. Affinities of amino acid side chains for solvent water, *Biochemistry*, **20**(4): 849-855 (1981).

[6] Cabani, S.; Gianni, P.; Mollica, V.; and Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution, *Journal of Solution Chemistry*, **10**(8): 563-595 (1981).

[7] Radzicka, A.; and Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution, *Biochemistry*, **27**(5): 1664-1670 (1988).

[8] Wimley, W. C.; Creamer, T. P.; and White, S. H. Solvation energies of amino acid side chains and backbone in a family of host− guest pentapeptides, *Biochemistry*, **35**(16): 5109-5124 (1996).

[9] Shirts, M. R.; Pitera, J. W.; Swope, W. C.; and Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins, *The Journal of chemical physics*, **119**(11): 5740-5761 (2003).

[10] Villa, A.; and Mark, A. E. Calculation of the free energy of solvation for neutral analogs of amino acid side chains, *Journal of computational chemistry*, **23**(5): 548-553 (2002).

[11] Gu, W.; Rahi, S. J.; and Helms, V. Solvation free energies and transfer free energies for amino acids from hydrophobic solution to water solution from a very simple residue model, *The Journal of Physical Chemistry B*, **108**(18): 5806-5814 (2004).

[12] Shirts, M. R.; and Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models, *The Journal of chemical physics*, **122**(13): 134508 (2005).

[13] Chang, J.; Lenhoff, A. M. and Sandler, S. I. Solvation free energy of amino acids and side-chain analogues. *The Journal of Physical Chemistry B*, **111**(8): 2098-2106 (2007).

[14] "Lysine," www.drugbank.ca/drugs/DB00123, accessed: 2021-01-25.

[15] Linstrom, P. J.; and Mallard, W. G. The NIST Chemistry WebBook: A chemical data resource on the internet, *Journal of Chemical & Engineering Data*, **46**(5): 1059-1063(1998).

[16] Bash, P. A.; Singh, U. C.; Langridge, R.; and Kollman, P. A. Free energy calculations by computer simulation, *Science*, **236**(4801): 564-568 (1987).

[17] Kirkwood, J. G. Statistical mechanics of fluid mixtures, *The Journal of chemical physics*, **3**(5): 300-313 (1935).

[18] Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases, *The Journal of Chemical Physics*, **22**(8): 1420-1426 (1954).

[19] Tuckerman, M. *Statistical mechanics: theory and molecular simulation* (Oxford university press, 2010).

[20] Pohorille, A.; Jarzynski, C.; and Chipot, C. Good practices in free-energy calculations, *The Journal of Physical Chemistry B*, **114**(32): 10235-10253 (2010).

[21] Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data, *Journal of Computational Physics*, **22**(2): 245-268 (1976).

[22] Wu, D.; and Kofke, D. A. Asymmetric bias in free-energy perturbation measurements using two Hamiltonian-based models, *Physical Review E*, **70**(6): 066702 (2004).

[23] Shirts, M. R.; and Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states, *The Journal of chemical physics*, **129**(12): 124105 (2008).

[24] Paliwal, H.; and Shirts, M. R. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods, *Journal of chemical theory and computation*, **7**(12): 4115-4134 (2011).

[25] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; and Hermans, A. J. *Intermolecular forces*, 331–342 (1981).

[26] Berendsen, H. J. C.; Grigera, J. R.; and Straatsma, T. P. The missing term in effective pair potentials, *Journal of Physical Chemistry*, **91**(24): 6269-6271 (1987).

[27] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, **79**(2): 926-935 (1983).

[28] Jorgensen, W. L.; Maxwell, D. S.; and Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *Journal of the American Chemical Society*, **118**(45): 11225-11236 (1996).

[29] Berendsen, H. J.; van der Spoel, D.; and van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation, *Computer physics communications*, **91**(1-3): 43-56 (1995).

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

[31] Khanal; S. P., Kandel; Y. P.; and Adhikari, N. P. Transport properties of zwitterion glycine, diglycine, and triglycine in water, *AIP Advances*, **9**(6): 065303 (2019).

[32] Van Gunsteren, W. F.; and Berendsen, H. J. A leap-frog algorithm for stochastic dynamics,. *Molecular Simulation*, **1**(3): 173-185 (1988).

[33] Hoover, W. G.; Ladd, A. J.; and Moran, B. High-strain-rate plastic flow studied via nonequilibrium molecular dynamics, *Physical Review Letters*, **48**(26): 1818 (1982).

[34] Parrinello, M.; and Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied physics*, **52**(12): 7182-7190 (1981).

[35] Darden, T.; York, D.; and Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems, *The Journal of chemical physics*, **98**(12): 10089-10092 (1993).

[36] Hess, B.; Bekker, H.; Berendsen, H. J.; and Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations, *Journal of computational chemistry*, **18**(12): 1463-1472 (1997).

[37] Steinbrecher, T.; Mobley, D. L., and Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations, *The Journal of chemical physics*, **127**(21): 214108 (2007).

[38] "alchemical-analysis,"accessed:2020-09-01. https://github.com/MobleyLab/alchemical-analysis.

[39] Lu, N.; Singh, J. K.; and Kofke, D. A. Appropriate methods to combine forward and reverse free-energy perturbation averages, *The Journal of Chemical Physics*, **118**(7): 2977-2984 (2003).

[40] Shirts, M. R.; and Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration, *The Journal of chemical physics*, **122**(14): 144107 (2005).

[41] Shirts, M. R.; and Mobley, D. L. An introduction to best practices in free energy calculations, *Biomolecular Simulations*, 271-311 (2013).

BIBECHANA

Solvation free energy of light alkanes in polar and amphiphilic environments

Sunil Pokharel, Shyam Prakash Khanal, N. P. Adhikari*

Central Department of Physics, Tribhuvan University, Kirtipur, Nepal
*Email: npadhikari@gmail.com

Abstract

Computer simulations of molecular models are powerful technique that have improved the understanding of many biochemical phenomena. The method is frequently applied to study the motions of biological macromolecules such as protein and nucleic acids, which can be useful for interpreting the results of certain biophysical experiments. In this work, we have estimated the solvation free energy for light alkane (methane, ethane, propane and n-butane) dissolved in water and methanol respectively over a broad range of temperatures, from 275 K to 375 K, using molecular dynamics simulations. The alkane (methane, ethane, propane and n-butane), and methanol molecules are described by the OPLS-AA (Optimized Potentials for Liquid Simulations-All Atom) potential, while water is modeled by TIP3P (Transferable Intermolecular Potential with 3-Points) model. We have used the free energy perturbation method (Bennett Acceptance Ratio (BAR) method) for the calculation of free energy of solvation. The estimated values of solvation free energy of alkane in the corresponding solvents agree well with the available experimental data.

Keywords: Alkane; Free Energy; Molecular dynamics; BAR.

1. Introduction

Solubility, lack of solubility and other solvation properties of atoms, molecules and ions in aqueous solutions play a crucial role in biological processes and industrial applications. The free energy of solvation (specifically, hydration) is one of the most important properties in the study of solvent effects which determines solubilities, partition coefficients, association, dissociation, and binding constants, phase equilibria, and for transition states, reaction rates in many bio-chemical and physical processes [1]. For instance, protein folding occurs spontaneously because of a favorable change in the interactions between the protein and the surrounding water molecules. The folded proteins, caused by combined

effects of solvent and hydrogen bonding, are stable by 5-10 kcal/mol . The solvation is driven by minimizing the number side-chains, exposed to water and hydrophobic in nature, covered by the folded protein at its centre [2]. The host-guest complexes, whose binding constant depends on solvent polarity, caused by hydrophobic pores present in host molecules also act as driving force for solvation [3]. Furthermore, the interactions can be used in biological system to delivery the hydophobic drugs. Hydration affects electronic and vibrational properties of biomolecule [4, 5].

Computer simulation and modeling is performed in order to comprehend the properties of assemblies of molecules in terms of their structure and the microscopic interactions between them. In a classical framework, the main computational methods are molecular dynamics (MD) and Monte Carlo (MC) simulations. Molecular Dynamics (MD) simulation is a powerful approach for predicting and knowing the structure, function, dynamics, and interactions of atoms and molecules starting from a simple to complex systems in biophysical phenomenon of materials sciences study. These techniques are complement to conventional experiments, enabling us to learn something new, something that cannot be found out in other ways [6-11].

Alkanes are saturated hydrocarbons that consist only of the elements carbon (C) and hydrogen (H), where each of these atoms are linked together exclusively by single bonds. Alkanes belong to a homologous series of organic compounds in which the members differ by a constant molecular mass of 14 that is CH_2 [12]. First four members (lighter alkanes) of alkane series are methane, ethane, propane, and butane with molecular formula CH_4, C_2H_6, C_3H_8, and C_4H_{10} respectively. The most important sources of alkanes are natural gas and crude oil. Alkanes are non-polar solvents as only C and H atoms are present. Methane, ethane, propane and n-butane are hydrophobic molecules, and as such its solubility in water is rather low, and they tend to aggregate when solvated in water, but freely soluble in non-polar solvent like ether and benzene. This behavior is more clearly exhibited by longer n-alkane chains, which may be considered as polymers of methane. The complicated organic compounds that once made up living plants or animals have transformed into a mixture of alkanes [12-16]. Furthermore, the first four members of alkanes are also neutral analogs of amino acid side chain. Amino acid side chain analogs represent a natural test case for biomolecular interaction [17, 18]. Transport properties such as diffusion, viscosity and thermodynamic properties like free energy of hydration/solvation of hydrocarbons (alkanes) in aqueous environment is a basic consideration in many processes like processing of natural gases and petroleum, understanding the tertiary structure of proteins, as well as the important role it plays as a driving force in a number of processes occurring within living cells [19-21]. The free energy of hydration of light alkanes (a polar molecule) for a few temperatures and pressures have been repeatedly measured by computer simulations but the experimental values are rare. On the other hand, there is no literature data for the free energy of solvation of light alkanes in methanol (an amphiphilic molecule). So, we are motivated to calculate the free energy of hydration/solvation of light alkanes in water and ethanol molecules. Our findings from the numerical simulation can also be used as a crude reference for any further studies of hydrophobicity and solubility of organic and inorganic substances in different solvent environments.

The outline of the paper is as follows: In Sec. II, we discuss the theoretical background of the free energy of solvation and method of calculation. Computational details of our work are stated in Sec. III. Results of the work are presented in Secs. IV. Our conclusions are collected in Sec. V.

2. Theoretical Background

A thermodynamic quantity equivalent to the capacity of a system to do work, the difference between internal energy of the system and the amount of energy that cannot be used to perform work, is known as free energy of the system. Mathematically, the Helmholtz free energy $A(N,\ V,\ T)$ and the Gibbs free energy $G(N,\ P,\ T)$ are defined as:

$$A = U - TS \tag{1}$$

$$G = A + PV = U - TS + PV = \mu N \tag{2}$$

where, $U,\ S$ and μ are internal energy, entropy and chemical potential of the system respectively. The Helmholtz free energy $A(N, V, T)$ describes a closed, isochoric, isothermal assembly, so it is a function of temperature (T), volume (V), and number of molecules (N). The Gibbs free energy $G(N,\ P,\ T)$ describes a closed isobaric, isothermal assembly, so it is a function of temperature (T), pressure(P), and number of molecules (N) [22, 23]. The chemical potential is given by

$$\mu = \left(\frac{\partial A}{\partial N}\right)_{T,V} = \left(\frac{\partial G}{\partial N}\right)_{T,P} \tag{3}$$

Two ensembles are particularly useful for the calculations of free energy: the canonical (NVT) ensemble and the isobaric-isothermal (NPT) ensemble. The Helmholtz free energy and Gibbs free energy are determined by the corresponding partition functions Q defined by [22, 23]:

$$A(N,\ V,\ T) = k_B\,T\ln Q_N\,(V,\ T) = \frac{1}{N!h^{3N}} \times \int e^{-\beta H_0(p,q)}\,dp\,dq \tag{4}$$

$$\text{and} \quad G(N,\ P,\ T) = k_B\,T\ln Q_N\,(P,\ T) = \frac{1}{N!h^{3N}} \times \int e^{-\beta\,P\,V}\,Q_N(V,T)\,dV \tag{5}$$

where $\beta = 1/k_B\,T$, H_0 is the unperturbed Hamiltonian of the system and V is the volume. In simulation, the direct measurement of the free energy is not possible but difference in free energy can be calculated. For any other system differing in the Hamiltonian by a perturbation of the potential energy(U), $H_p = H_0 + U$, the difference in Helmholtz free energy is,

$$\Delta A = A_p - A_0 = -k_B\,T\ln\left(\frac{Q_P}{Q_0}\right) \tag{6}$$

Free energy calculations have a number of practical applications, of which some of the more common ones include free energies of solvation/hydration and free energy of binding for a small molecule to some larger receptor biomolecule (usually a protein). Equilibrium free energy methodologies share the common strategy of generating equilibrium ensembles of configurations at multiple values of the scaling parameter λ. The commonly used methods are thermodynamic integration [24], adaptive integration [25], multistage free energy perturbation [26], and multistage equilibrium Bennett analysis [27]. In GROMACS, different approaches, including "slow-growth" can be used to calculate the free energy differences. In the "slow-growth" method, the Hamiltonian of system, that changes slowly to remain in equilibrium from one system A to other system B, is modified by making the Hamiltonian as function of coupling parameter λ as $H(p, q\,;\lambda)$ such that $H(p, q\,;0) = H^A(p, q)$ and $H(p, q\,;1) = H^B(p, q)$. The Helmholtz free energy $A(N, V, T)$ in terms of partition function Q is defined as [22, 23, 36]:

$$A(\lambda) = -k_BT\ln Q_T(V, T) = \frac{1}{N!h^{3N}} \times \int e^{-\beta H_0(p,\ q;\lambda)}\,dp\,dq \tag{7}$$

where $\beta = 1/k_BT$, $H(p, q;\lambda)$ is the Hamiltonian of the system.

In order to calculate free energy difference, we define potential energy function which linearly depends on coupling parameters λ i.e. $U(\lambda)$ such that, for $\lambda = 0$, $U_A(\lambda = 0)$ represents the potential energy of system A and $U_B(\lambda = 1)$ represents the potential energy of system B, then

$$U(\lambda) = (1 - \lambda)\,U_B + \lambda U_A = U_A + \lambda(\,U_B - U_A) \tag{8}$$

Also, the derivative of Helmoltz free energy Eq. (7) with respect to coupling parameter λ and using Eq. (8), we get the relation as:

$$\frac{dA}{d\lambda} = \frac{\int \left(\frac{dH}{d\lambda}\right) e^{-\beta\,H(p,q;\lambda)}\,dp\,dq}{\int e^{-\beta\,H(p,q;\lambda)}\,dp\,dq} = \left\langle\frac{\partial H}{\partial\lambda}\right\rangle_{NVT;\lambda} = \left\langle\frac{\partial U(\lambda)}{\partial\lambda}\right\rangle_{NVT;\lambda} \tag{9}$$

The $\langle...\rangle$ brackets represent the ensemble average. After integrating the Eq.(9) from $\lambda = 0$ to $\lambda = 1$ using thermodynamic integration (TI) method, we can evaluate the free energy difference between the system A and B as

$$\Delta A = A^B(V,T) - A^A(V,T) = \int_0^1 \left\langle\frac{\partial H}{\partial\lambda}\right\rangle_{NVT;\lambda}\,d\lambda = \int_0^1 \left\langle\frac{\partial U(\lambda)}{\partial\lambda}\right\rangle_{NVT;\lambda}\,d\lambda \tag{10}$$

According to the Bennett Acceptance Ratio (BAR) [27] method, the ratio of partition function Q_0 for $\lambda=0$ and Q1 for $\lambda=1$ is given by:

$$\frac{Q_0}{Q_1} = \frac{Q_0 \int w(r^N)\,e^{-\beta(U_0+U_1)}\,dr^N}{Q_1 \int w(r^N)\,e^{-\beta(U_0+U_1)}\,dr^N} = \frac{\langle w\,e^{-\beta U_0}\rangle_1}{\langle w\,e^{-\beta U_1}\rangle_0} \tag{11}$$

where 'w' is an arbitrary weight function. In terms of 'w', the Helmoltz free energy difference is given by

$$\beta\,\Delta A = \ln\langle w\,e^{-\beta U_0}\rangle_1 - \ln\langle w\,e^{-\beta U_1}\rangle_0 \tag{12}$$

It is also possible to use Bennetts method to combine the information normally used for forward and reverse free energy perturbations. In this approach, we compute the free energy difference between successive λ values δA according to the relation

$$\left\langle\left[1 + e^{\beta\left(U_{\lambda_{i+1}}(x_i) - U_{\lambda_i}(x_i) - \delta A_i\right)}\right]^{-1}\right\rangle_{\lambda_i} = \left\langle\left[1 + e^{\beta\left(U_{\lambda_{i+1}}(x_{i+1}) - U_{\lambda_i}(x_{i+1}) + \delta A_i\right)}\right]^{-1}\right\rangle_{\lambda_{i+1}} \tag{13}$$

Then the sum of these δA i is the total Helmholtz free energy difference [27],

$$\Delta A = \sum_{i=0}^{n-1} \delta A_i \tag{14}$$

3. Computational Model

A. Molecular Models

The TIP3P (Transferable Intermolecular Potential with 3-Points) model [28] is used in all the simulation for water as a solvent. The OPLSS-AA (Optimized Potentials for Liquid Simulations-All Atom) potential model [29] is used for alkanes (methane, ethane, propane, n-butane) and methanol. The all atom model of the studied alkane system is shown in figure (1). The system under study consists of 1 alkane (methane, ethane, propane, n-butane) molecule and 596 water, and 1 alkane (methane, ethane, propane, n-butane) molecule and 354 methanol, a separate system separately. In classical force fields like OPLS-AA, 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 interac- tions with the rest of the system. The total potential energy U tot includes contributions from both bonded and non-bonded interactions [36]. The bonded interactions are bond stretching (2-body), bond angle (3- body) and dihedral angle (4-body) interactions. A special type of dihedral interaction (called improper dihedrals) is used to force atoms to remain in a plane or to prevent transition to a configuration of opposite chirality (a mirror image). The non-bonded interactions are represented by the Lennard-Jones potential and Coulomb potential. Therefore, the total potential energy function of a system can be written as [36]:

$$U_{tot} = U_{bonded} + U_{non-bonded} = U_{bond} + U_{angle} + U_{dihedral} + U_{LJ} + U_{Coul} \tag{15}$$
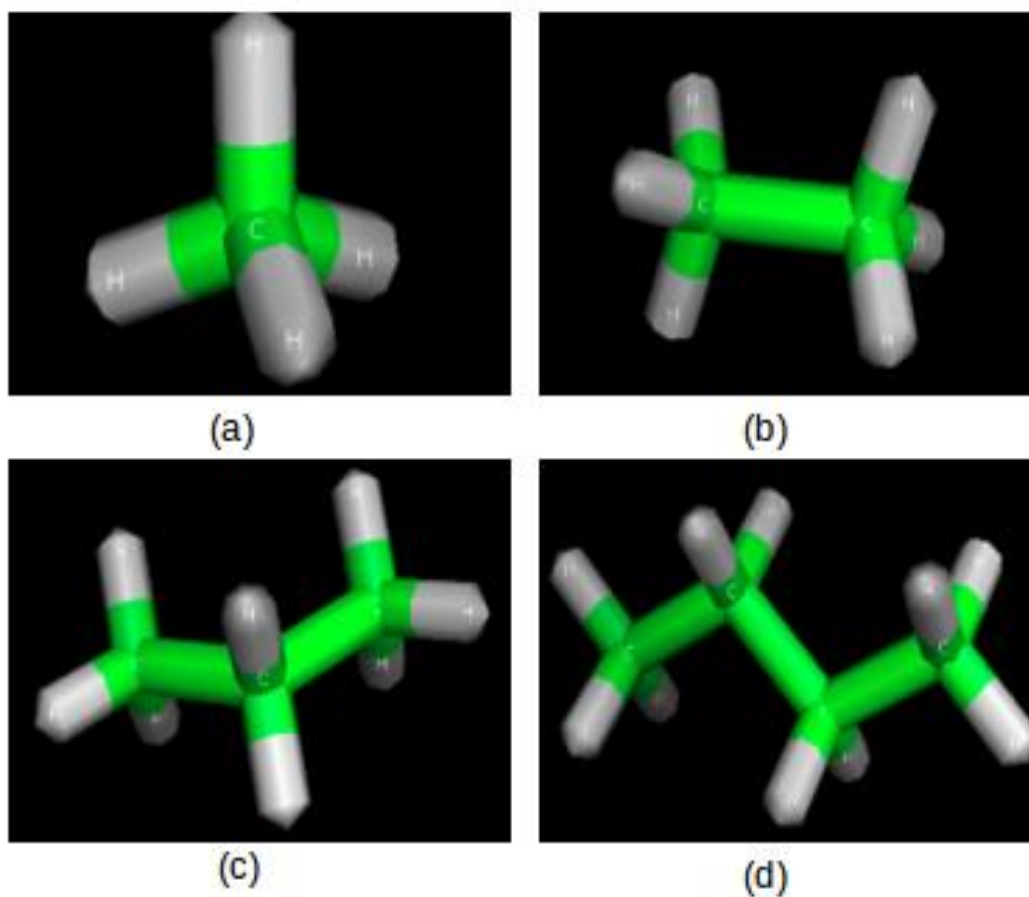
Fig. 1: Pymol snapshots of light alkanes (a) Methane, (b) Ethane, (c) Propane and (d) n-Butane model used in this work.

The bond stretching between two covalently bonded atoms i and j is represented by harmonic potential [36]

$$U_{bond}(r_{ij}) = \frac{1}{2}k_{ij}^{b}(r_{ij} - b_{ij})^{2} \tag{16}$$

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

The bond angle vibration between a triplet of atoms $i - j - k$ is also represented by a harmonic potential on the angle Θ_{ijk} [36]

$$U_{angle}(\Theta_{ijk}) = \frac{1}{2}k_{ijk}^{\Theta}(\Theta_{ijk} - \Theta_{ijk}^{0})^{2} \tag{17}$$

where k_{ijk}^{Θ} is the force constant and Θ_{ijk}^{0} is the equilibrium bond angle.

The proper dihedral angle is defined as the angle between the ijk and jkl planes. In this study, we have used the following dihedral potential (Ryckaert-Bellmans potential) [36] for alkanes:

$$U_{RB} = c_0 + c_1(1 + \cos\varphi) + c_2(1 - \cos2\varphi) + c_3(1 + \cos3\varphi) \tag{18}$$

where φ is the dihedral angle and c_0, c_1, c_2, c_3 are constants. The bonded parameters for water and alkanes are given in the table (I).

Table I: Force-field (bonded) parameters for TIP3P wa- ter and OPLS-AA alkanes and methanol. The units of equilibrium bond length (b) and equilibrium bond angle (Θ_0) are nanometer (nm) and degrees (o) respectively. Similarly, the units of k^b, k^Θ and c_i (c_0, c_1, c_2, c_3) are kJmol−1nm−2, kJmol^{-1}rad^{-2} and kJmol^{-1} respectively.

TIP3P	k^b_{OH}	502415.0	b_{OH}	0.09572
water	k^θ_{HOH}	628.02	θ^0_{HOH}	104.52
	k^b_{CH}	284512.0	b_{CH}	0.1090
OPLS-AA	k^b_{CC}	224262.4	b_{CC}	0.1529
Alkanes	k^θ_{HCH}	276.144	θ^0_{HCH}	109.47
	k^θ_{HCC}	313.800	θ^0_{HCC}	109.47
	k^θ_{CCC}	488.273	θ^0_{CCC}	109.47
Dihedral Potential (Alkanes)				
H-C-C-H	C_0	0.62760	C_0	1.88280
	C_2	0.00000	C_2	-2.51040
H-C-C-C	C_0	0.62760	C_0	1.88280
	C_2	0.00000	C_2	-2.51040
	C_0	2.92880	C_0	-1.46440
C-C-C-C	C_2	0.20920	C_2	-1.67360
Dihedral Potential (Methanol)				
H-C-OH-HO	C_0	0.94140	C_0	2.82420
	C_2	0.00000	C_2	-3.76560
H-C-C-C	C_0	0.62760	C_0	1.88280
	C_2	0.00000	C_2	-2.51040
C-C-C-C	C_0	2.92880	C_0	-1.46440
	C_2	0.20920	C_2	-1.67360

The non-bonded inter-atomic interaction is the sum of Lennard-Jones interaction (ULJ) and Coulomb interaction (UCoul), that can be written as:

$$U_{\alpha\beta}(r_{ij}) = 4\epsilon_{\alpha\beta}[(\frac{\sigma_{\alpha\beta}}{r_{ij}})^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^6] + \frac{q_{i\alpha}\ q_{j\beta}}{4\pi\epsilon_0 r_{ij}} \tag{19}$$

where r_{ij} is the Cartesian distance between the two atoms i and j; α and β indicate the type of the atoms. The non-bonded parameters for alkanes and water is given in the table (II).

Table II: Force-field (non-bonded) parameters for TIP3P water and OPLS-AA alkanes and methanol.

	Atoms	σ(nm)	ε(kJ/mol)	Charge (q)
TIP3P	OW	0.315061	0.636386	-0.834 e
Water	HW	0.000000	0.000000	+0.417 e
	C(CH$_4$)	0.35000	0.276144	0.000 e
OPLS-AA	C(CH$_3$)	0.35000	0.276144	0.000 e
Alkanes	C(CH$_2$)	0.35000	0.276144	0.000 e
	H	0.25000	0.125520	0.000 e
	C(CH$_3$)	0.35000	0.276144	+ 0.145 e
OPLS-AA	C(CH$_2$)	0.25000	0.125520	+ 0.040 e
Methanol	OH	0.312000	0.711280	- 0.683 e
	HO	0.00000	0.000000	+ 0.418 e

Here OW and HW represent the oxygen and hydrogen atoms of the water molecules respectively and $C(CH_4)$, $C(CH_3)$ and $C(CH_2)$ are the methane, methyl and methylene carbon atoms of the alkane molecules respectively. Similarly, OH and HO are oxygen and hydrogen atoms of hydroxyl group of methanol respectively. The parameters for the non-bonded Lennard-Jones interaction between two different atoms for OPLS-AA force field are written as [36]:

$$\sigma_{\alpha\beta} = (\sigma_{\alpha\alpha} \times \sigma_{\beta\beta})^{\frac{1}{2}} \qquad (20)$$

$$\epsilon_{\alpha\beta} = (\epsilon_{\alpha\alpha} \times \epsilon_{\beta\beta})^{\frac{1}{2}} \qquad (21)$$

B. Simulation Procedure

Molecular dynamics simulation was carried out in a cubic box with periodic boundary conditions [7] using GROMACS 5.1.1 [37, 38, 47]. The distance to the edge of the box from the solute (alkane) is an important parameter for defining the size of the box. Since we are using periodic boundary conditions, we must satisfy the minimum image convention. That is alkane (solute) should never see its periodic image, otherwise the forces calculated will be spurious. The size of the box defined here is sufficient for just about any cutoff scheme commonly used in simulations. Our system consists of, 596 water and 1 alkane molecule, 354 ethanol and 1 alkane molecule a separate system. After defining a system in a simulation box, energy minimization is carried out for each values of λ from 0 to 1, for 21 different values to avoid unphysical van der Waals contact caused by the atoms that are too close [36]. Energy minimization brings the system to equilibrium configuration, removes all the kinetic energy from the system, reduces thermal noise in structure and brings the system to one of the local minimum. Steepest descent algorithm followed by L-BFGS (limited- memory- Broyden-Fletcher-Goldfarb-Shanno quasi- Newtonian-mimimizer) [30, 31] algorithm has been used for energy minimization [36]. This combination (steepest descent and L-BFGS) yields a thoroughly-minimized structure suitable for starting equilibration and subsequent data collection. The energy (potential) of the system after energy minimization is shown in figure (2).
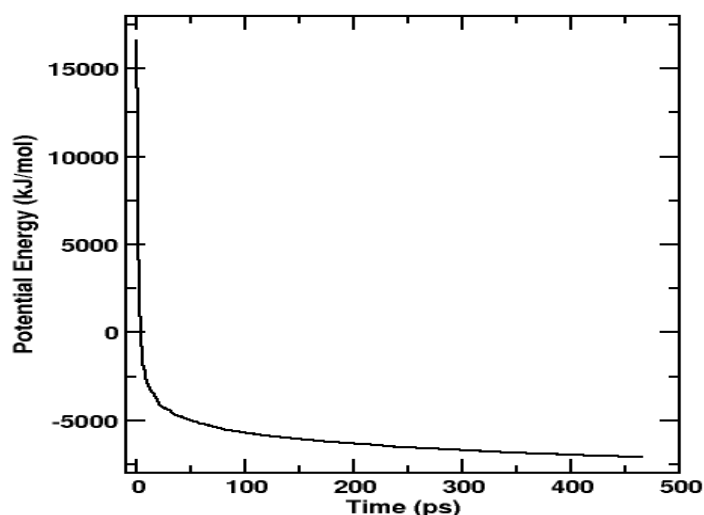

Fig. 2: Plot of potential energy as a function of time after energy minimization for methane-methanol system at $\lambda=0$

After energy minimization, NVT equilibration of 5×10^5 steps for 1 ns and isobaric-isothermal (NPT) equilibration of 2.5×10^6 of 5 ns was carried out at different temperature, from 275K to 375 K and a pressure of 105 Nm^{-2} by using velocity-rescaling thermostat [32] and Berendsen barostat [33] at a coupling time τ_t = 1.0 ps and τ_p = 0.5 ps respectively. Integration of the equations of motion were performed using an accurate leap-frog stochastic dynamics (SD) algorithm [34], and all bonds were constrained using LINC algorithm [35]. During equilibration, the long range Coulomb interac- tion is handled via the PME (Particle Mesh Ewald) algorithm [39, 40] with fourier spacing 0.12 with a real space cutoff of 1.2 nm and a PME order of 6. We monitored the temperature, pressure, density, and energy of each studied system to bring it in thermodynamic equilibrium. After equilibration run we performed the production run to calculate the equilibrium properties of the system, that is free energy of solvation for each values of λ by fixing the number of particles, volume and temperature i.e. NVT ensemble. We use velocity-rescale thermostat for this case. We don't couple the system to a fixed pressure and use the structure obtained after equilibration run by which we fix the volume of the system. The production run was carried out for 1 ns with the time step of 2fs.

4. Results and Discussion

Using the computational details described above, free energy of solvation calculations were carried out in a cubic box in infinite dilution for two separate systems: (i) 596 TIP3P water molecules and 1 OPLS-AA alkane (methane, ethane, propane and n-butane) molecule (ii) 354 OPLS-AA methanol and 1 OPLS-AA alkane (methane, ethane, propane and n-butane) molecule. The free energy of solvation of alkanes in different solvent environments, water and methanol were estimated by Eq.(14). The extent to which Hamiltonian of the system has been perturbed is measured by the free energy change of transforming a system from state A (λ = 0) to state B (λ = 1), ΔA, as a function of a coupling parameter, λ. For decoupling van der Waals interactions, we used an equidistant λ spacing of 21 different λ's from 0 to 1. Thus, the free energy change from λ = 0 to λ = 1 is simply the sum of the free energy changes of each pair of neighboring λ simulations. The free energy changes of each pair of neighboring λ and the cumulative free energy change which is negative of free energy of solvation of butane in water is shown in figure (3).
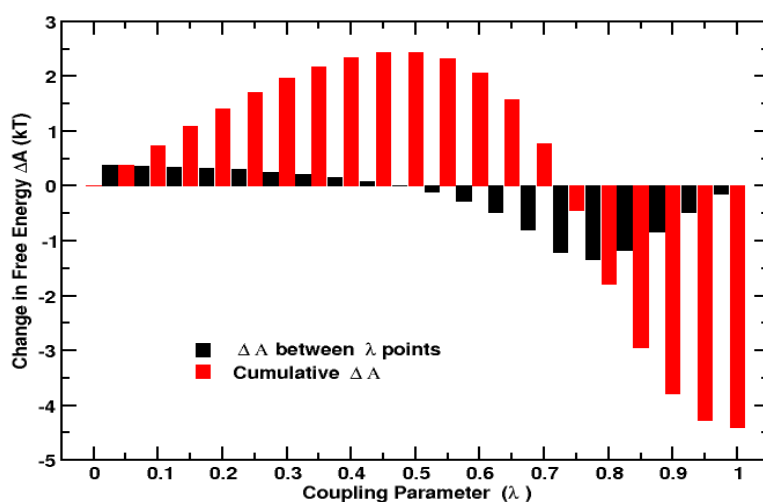

Fig. 3: Difference in free energy (ΔA) and cumulative free energy difference (ΔA) for different λ points for simulation of butane in water at T = 300 K.

The calculated values of solvation free energy of alkane (methane, ethane, propane, n-butane) in water and methanol along with the references (if available) at different temperatures are presented in table (III). There is good agreement between experimental values at 298 K and calculated values at 300 K of the solvation free energy of alkanes in water within 15 % error, but at higher temperatures there is no available experimental results. And there is no experimental available result for solvation free energy of alkanes in methanol.

Table III: Calculated and available experimental solvation free energies (kJmol^{-1}) at different temperature of light alkanes in water and methanol.

| Molecule	T(K)	Water		Methanol	
		Calculated	Expt.[41] (298.15K)	Calculted	Ref.
Methane	275	8.33±0.05	--	2.94±0.02	--
	300	9.08±0.12	8.08	3.05±0.10	--
	325	10.01±0.14	--	3.36±0.01	--
	350	10.32±0.09	--	3.58±0.03	--
	375	10.63±0.07	--	3.79±0.05	--
Ethane	275	8.22±0.14	--	0.20±0.15	--
	300	9.02±0.19	7.41	0.70±0.12	--
	325	11.12±0.07	--	0.90±0.10	--
	350	11.20±0.10	--	0.99±0.10	--
	375	11.52±0.10	--	1.01±0.07	--
Propane	275	8.77±0.14	--	-2.73±0.21	--
	300	9.61±0.21	8.28	-1.28±0.30	--
	325	11.27±0.08	--	-0.93±0.16	--
	350	11.80±0.06	--	--	--
	375	12.70±0.20	--	--	--
n-Butane	275	9.10±0.13	--	-4.92±0.19	--
	300	10.99±0.18	9.03	-3.77±0.18	--
	325	11.82±0.11	--	-3.60±0.19	--
	350	12.95±0.20	--	-2.75±0.16	--
	375	13.93±0.11	--	--	--
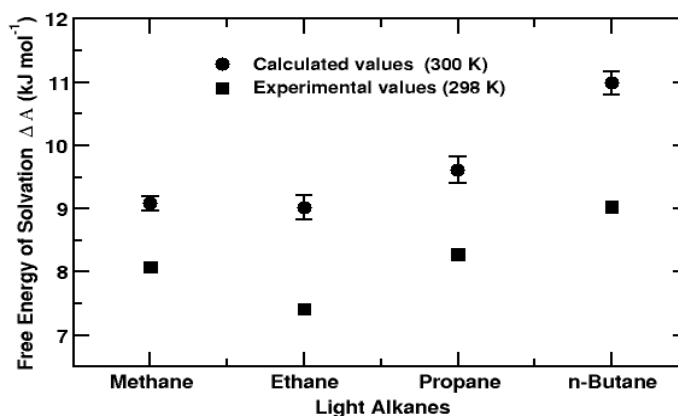
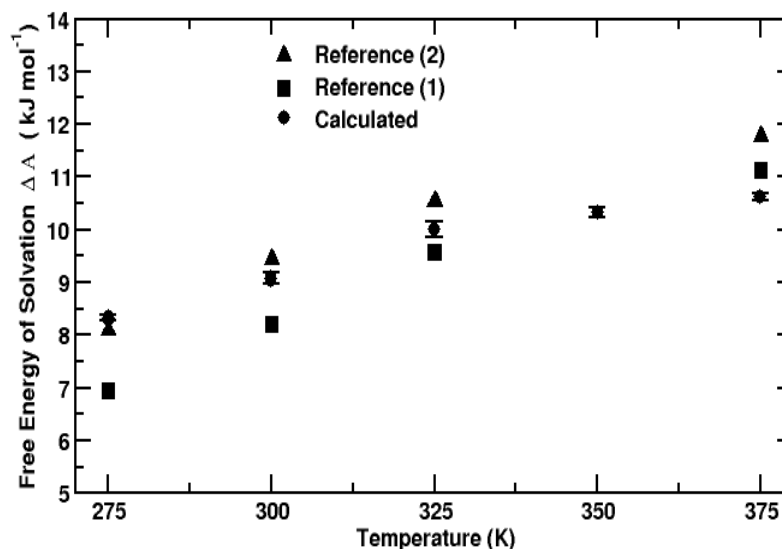**Fig. 4:** Solvation free energy of alkanes in TIP3P water at T=300 K.

Fig. 5: Comparision of calculated and literature values (Reference (1) [42, 43], Reference (2) [44] of solvation free energy of methane in water at different temperature.

The comparison between the calculated values with error bars of solvation free energy of alkane in water at 300 K and the corresponding experimental values at 298 K is shown in figure (4). Similarly, the comparison of calculated and literature values of free energy of solvation of methane in water at different temperature in shown in figure (5) and tabulated in table (IV). The free energy of solvation of alkanes (methane, ethane, propane and n-butane) in methanol as a fuction of temperature is plotted in figure (6).

Table IV: Calculated and reference (experimental and literature) values of solvation free energies (kJ/mol) at different temperature of methane in water.

Molecule	T(K)	Calculated	Ref.(1)[42, 43]	Ref.(2) [44]
	275	8.33±0.05	6.95	8.09
	300	9.08±0.12	8.21	9.45
Methane	325	10.01±0.14	9.58	10.54
	350	10.32±0.09	--	--
	375	10.63±0.07	11.13	11.78

The solvation process is considered to consist of two steps, (i) the formation of a repulsive cavity of appropriate size, and (ii) the introduction of the soa lute into this cavity. The positive values of calculated free energies of solvation (ΔA) in TIP3P water for all of the alkanes (methane, ethane, propane and n-butane) indicate their low solubilities in water that means alkanes are hydrophobic in nature. The simulations also show, in accordance with experiment, that ΔA decreases from methane to ethane, but then increases with increasing carbon number for longer up to butane at 300 K. This shows that the methyl group of alkane molecules have a preferential tendency to be dissolved in the vicinity of water molecules and that this tendency decreases with chain length.
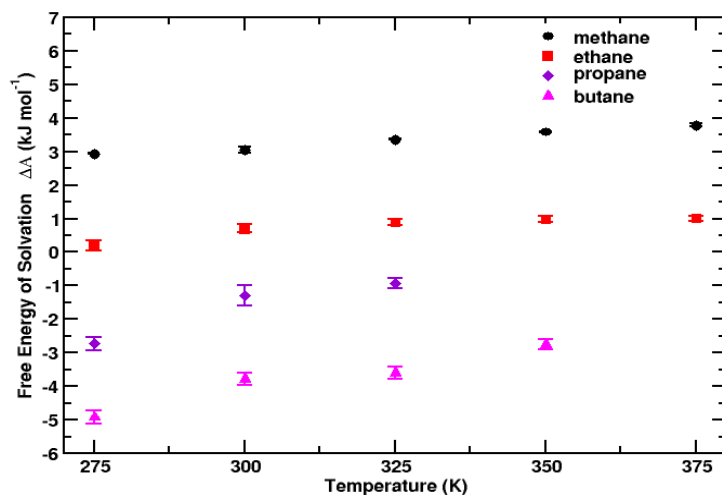
Fig. 6: Solvation free energy of alkane in methanol at different temperature from T = 275 K to T = 375K.

The increase in free energy of solvation with increase in temperature describes that the formation of repulsive cavity in water is perturbed by the thermal agitation of the water molecules. That means when temperature is incresed we dont have new interactions that are strong enough to introduce some important change in enthalpy and change in Helmoltz free energy A is mostly connected with entropy or reordering of hydrogen bonds. By placing alkane in water we are perturbing the hydrogen bond network so that water molecules need to reorganize themselves around the solvent in a particular way that makes possible for average number of bonds for one molecule to remain constant. Since hydrogen bonds are directional this leads to a smaller configurational space for water molecules and change in entropy will be negative 45, 46]. Again , we have calculated free energy of solvation of alkanes in methanol at different temperature. From table (III) and figure (6), free energy of solvation ΔA in methanol becomes more negative as the alkane chain increases. The positive values of solvation free energy of methane and ethane shows that they are insoluble in methanol but the negative values of it for propane and butane indicates they are soluble in the methanol. Methanol is amphiphilic organic substance. There is a competition between hybrophobic group (methyl-CH_3) and a hydrophilic group (hydroxyl-OH). Amphiphilic nature makes methanol interesting solvent because alkanes should show greater solubility in it than in water. For methane and ethane the hydrophobic groups dominates over the hydrophilic group, on the other hand for propane and butane the reverse situation occurs.

4. Conclusion and Concluding Remarks

In this work, we have computed free energy of solvation (Helmholtz free energy A) of alkane (methane, ethane, propane, n-butane) molecules in different solvent environments - water as a polar and methanol as an amphiphilic solvent, for various temperatures 275 K, 300 K, 325 K, 350 K, 375 K, using molecular dynamics simulation technique. The Transferable Intermolecular Potential with 3-Points (TIP3P) model of water and the Optimized Potential for Liquid Simulations- All Atom (OPLS-AA) model of alkane and methanol were used. Here alkane (methane, ethane, propane and n-butane) acts as a solute and water and methanol act as a polar and an amphiphilic solvent respectively. The free energy of solvation or hydration

of alkane in water is positive, and this values increases with increase in change length of alkanes. The calculated values of the free energy are in agreement with the available literature values. The free energy of solvation of alkanes in methanol shows different trends, it is negative for propane and butane but positive for methane and ethane. This shows that methane and ethane are insoluble whereas propane and butane are soluble in methanol.

This study reports the first complete description of the solution thermodynamics calculating solvation free energy of alkanes in water and methanol by computer simulations, using Bennet Acceptance Ratio (BAR) method. This study could be the basis for understanding the biomolecular interactions and calculation of free energy of binding for a small molecule to some larger receptor biomolecule (usually a protein). In the near future, we also plan to study the solvation/hydration of a larger series of solutes and calculation free energy of binding between a ligand and a receptor in different aqueous environment.

Acknowledgements

References

[1]   T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water, J. Chem. Phys. 85 (1986) 6720.

[2]   C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala, Forces contributing to the conformational stability of proteins, FASEB Journal. 10 (1996) 7583.

[3]   J. W. Steed, and J. L. Atwood, Supramolecular Chemistry, 2nd ed. Wiley, (2013).

[4]   Alireza Mashaghi et al., Hydration strongly affects the molecular and electronic structure of membrane phospholipids, J. Chem. Phys. 136 (2012) 114709.

[5]   M Bonn, H.J. Bakker, Y. Tong, E. H. Backus, No ice-like water at aqueous biological interfaces, Biointerphases. 7 (2012) 20.

[6]   D. Frenkel, B. Smit, Understanding Molecular Simulation From Algorithms to Applications, Academic Press, U. S. A, (2002).

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

[8]   H. J. C. Berendsen, Simulating the Physical World , Cambridge University Press, (2007).

[9]   A. Satoh, Introduction to Practice of Molecular Simulation, Elsevier, (2011).

[10]   D. C. Rapport, The Art of Molecular Dynamics Simulation, Cambridge University Press, Second Edition, (2004).

[11]   O. M. Becker, A. D. MacKerell, Jr. B. Roux, M. Watanabe, Computational Biochemistry and Biophysics, Marcel Dekker, Inc. (2001).

[12]   R. T. Morrison, R. N. Boyd, Organic Chemistry, Pearson, 7 th Edition, (2011).

[13]   J. McMurray, Organic Chemistry, Seventh Edition, Physical Sciences: David Harris, (2008).

[14]   W. H. Brown and T. Poon, Introduction to Organic Chemistry, 5 th Edition, John Wiley and Sons, Inc. (2014)

[15]   J. D. Roberts, M. C. Caserio, Basic Principles of Organic Chemistry, W.A. Benjamin Inc, 2nd Edition, (1977).

[16]   T. A. Weber, Simulation of n-butane using a skeletal alkane model, J. Chem. Phys. 69 (1978) 2347.

[17]   M. R. Shirts et al., Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins, J. Chem. Phys. 119 (2003) 5740.

[18] M. R. Shirts and V. S. Pande, Solvation free energies of amino acid side chain analogs for common molecular mechanics water models, .J. Chem. Phys. 122, 134508 (2005).

[19] S. K. Murphy, and Park, J. W. Park and Cruz, F. A. Cruz and V. M. Dong, Rh-catalyzed C--C bond cleavage by transfer hydroformylation, Science, 347 (2015) 6217.

[20] P. L. Privalov and S. J. Gill, Stability of protein structure and hydrophobic interaction, Adv. Protein Chem. 39 (1988) 191.

[21] J. Kyte, Structure in Protein Chemistry, Second Edition, Taylor & Francis Group (2007).

[22] R. K. Pathria and P. D. Beale, Statistical Mechanics, Academic Press, 3 rd Edition, (2011).

[23] K. Huang, Statistical Mechanics, 2nd Edition, John Wiley (1987).

[24] J. G. Kirkwood, Statistical mechanics of fluid mixtures, J. Chem. Phys. 3 (1935) 300.

[25] M. Fasnacht, R. H. Swendsen, and J. M. Rosenberg, Adaptive integration method for Monte Carlo simulations, Phys. Rev. E, 69 (2004) 056704

[26] R. W. Zwanzig, High-temperature equation of state by a perturbation method, I. Nonpolar gases. J. Chem. Phys. 22 (1954) 1420.

[27] C. H. Bennett, Efficient estimation of free energy differences from Monte Carlo data, J. Comput. Phys. 22 (1976) 245

[28] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R.W. Impey, M. L. Klein,   Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926935.

[29] G. A. Kaminski, R.A. Friesner, J.Tirado-Rives and W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. B 105 (2001) 6474.

[30] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing 16 (1995) 1190.

[31] C. Zhu, R. H. Byrd, P. Lu, J. Nocedal,   Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans, Math. Softw. 23 (1997) 550

[32] G. Bussi, D. Donadio, M. Parrinello,   Canonical sampling through velocity rescaling, J. Chem. Phys. 126,(2007) 014101.

[33] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 36843690

[34] W. F. van Gunsteren, H. J. C. Berendsen, A leap-frog algorithm for stochastic dynamics, Mol. Sim. 1 (1988) 173.

[35] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: a linear constraint solver for molecular simulations. J. Comp. Chem. 18, 1463-1472, (1997).

[36] 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 5.1.1, (2016).

[37] D. van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, GROMACS: fast, flexible, and free, J. Comp. Chem. 26 (2005) 1701.

[38] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comp. 4 (2008) 435.

[39] T. Darden, D. York, L. Pedersen,   Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089.

[40] U. Essmann et al., A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577.

[41] A. Ben-Naim, and Y. Marcus, Solvation thermodynamics of nonionic solutes, J. Chem. Phys. Physics 81 (1984) 2016.

[42] D. Paschek, Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models, J. Chem. Phys. 120 (2004) 6674.

[43] R. F. Prini and R. Crovetto, J. Phys. Chem. Ref. Data, 18 (1998) 1231.

[44] H. Docherty, A. Galindo, C. Vega and E. Sanz, A potential model for methane in water describing correctly the solubility of the gas and the properties of the methane hydrate, J. Chem. Phys. 125 (2006) 074510.

[45] Jianzhong Wu and John M. Prausnitz, Pairwise-additive hydrophobic effect for alkanes in water, PNAS 105 (2008) 9515.

[46] Michael H. Abraham and Enrico Matteoli, The temperature variation of the hydrophobic effect, J. Chem. Soc., Faraday Trans. 1, 84 (1988) 1985-2000.

[47] S. Pokharel, N. Pantha and N. P. Adhikari, Diffusion coefficients of nitric oxide in water: A molecular dynamics study, International Journal of Modern Physics B 30 (2016) 1650205

Transport properties of cysteine dimer in water

Research Article

Rajendra Pd. Koirala*, **Shyam P. Khanal, Narayan P. Adhikari**

Central Department of Physics, Tribhuvan University, Kirtipur, Kathmandu, Nepal

Abstract: Disulphide bond in cysteine residues plays vital role in structural stability and fuctional variation of protein molecules. Study of cysteine dimer linking with disulphide bond reveals the nature of stability of tertiary and quarternary structure in polypeptide chain. In order to study the transport properties of cysteine dimer, the molecular dynamics (MD) simulations have been performed at different temperature. The self diffusion coefficients of both cysteine dimer and TIP3P water model have been estimated at four different temperature from the slope of mean square displacement (MSD) versus time plot using Einstein's relation and their binary diffusion coefficients from Darken's relation.

Keywords: Molecular dynamics • Diffusion coefficient • Disulphide bond • Dimer

1. Introduction

Cysteine is a non-essential amino acid. It contains a thiol group -SH in its side chain. It is a white crystalline solid having physical properties: molar mass 121.15 gram per mole, melting point 513 K and solubility in water is 16 gram per 100 mL at 288 K [1, 2]. Cysteine plays vital role in the absorption of nutrients in the inner wall of small intestine. Its defensive mechanism against some diseases is essential for human body. Cysteine residue in a protein molecule strengthens the immune system and defends against the dementia, parkinson and multiple sclerosis [3, 4].

Two cysteine molecules bond together via two different methods: peptide bond (CO-NH) and disulphide bond $(R - S - S - R')$ [1], where R & R' are the side chains of two cysteine residues as shown in Fig. 1. Peptide bond is ordinary bond to form a polypeptide chain, whereas the disulphide bond is formed between two side chains derived from two thiol groups. Disulphide bond in two cysteine residues are very important components to form tertiary structure of proteins, besides to some weak interactions like hydrogen bonding, hydrophobic interactions salt bridges and weakly polar interaction. Although methionine, another amino acid, contains sulphur atom in its side chain, it does not form disulphide bond. Hence, the study the characteristics of disulphide bond is important. [1, 5].

---

* *Corresponding Author: rpkoirala@tucdp.edu.np*

Due to the oxidation of the sulfhydryl group of cysteine residues in a protein molecule, disulfide bond is formed. In such process, the thiol part -SH of cysteine molecules, contributes in disulfide bond formation, is deprotonated and covalent bond is formed between them [5–7]. Two cysteine molecules also covalently linked through peptide bond. Peptide bond is formed after the elimination of a water molecule during the interaction of carboxyl (C-terminus) and amine (N-terminus) regions of two amino acids. This bond has partial double bond character, i.e., stronger than single bond and weaker than double bond. It prevents the rotation of residues in protein molecules [1, 2]. This work is basically focus on disulphide bonding in cysteine dimer.
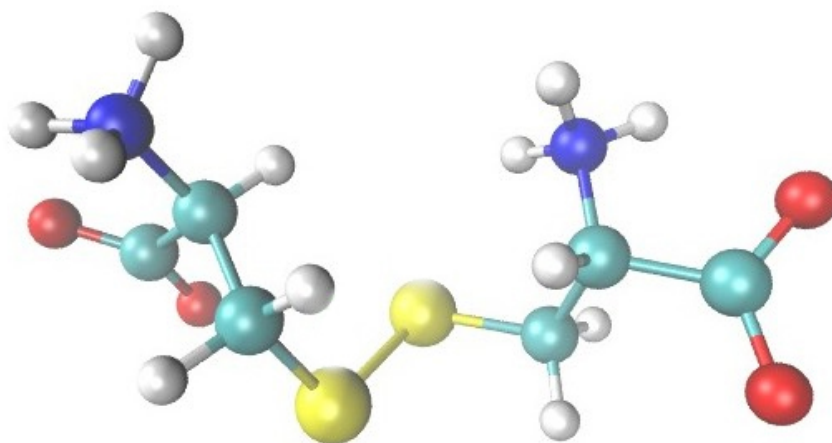

Figure 1.   Cysteine dimer.

Diffusion is an important physical phenomenon to understand the rate of transport of mass from the region of higher concentration to lower concentration [8–11]. Many bio-molecules like insulin are stabilized due to the formation of disulphide bond. The detailed understanding of effect of disulphide bond in any molecule and its effect on diffusion phenomenon in aqueous environment plays an important role in different areas including drugs designing. With our best knowledge, the diffusion coefficient of cysteine dimer, covalently bonding with disulphide bond, in water using classical molecular dynamics technique has not been studied yet, which motivate us to study the diffusion phenomenon of cysteine dimer.

2.   Methodology

Diffusion

Diffusion is a transport property of matter due to which particles flows from the region of higher concentration to lower concentration due to the concentration inhomogeneity [12]. The knowledge about diffusion phenomenon of biomolecules plays an important role to understand many phenomena in living organisms, like transport of biomolecules at different part of body through body fluids. The diffusion in a homogeneous system in the absence of chemical concentration gradient is called self diffusion and is measured in terms of self diffu-

sion coefficient [13] which is estimated from slope of mean square displacement(MSD) versus time graph using Einstein's equation as [14, 15];

$$D = \lim_{t \to \infty} \frac{\langle [r(t) - r(0)]^2 \rangle}{6t}. \tag{1}$$

In equation (1), $r(t) - r(0)$ is the displacement of particle from reference point during the course of time t, $[r(t) - r(0)]^2$ is the square of displacement and $\langle ... \rangle$ represents the ensemble average and hence $\langle [r(t) - r(0)]^2 \rangle$ gives MSD of particle.

Further, the binary diffusion, the diffusion of particles in the mixture of two different species, is measured in terms of binary diffusion coefficient using Darken's relation as [16];

$$D_{12} = N_2 D_1 + N_1 D_2 \tag{2}$$

In equation (2), D_{12} is the binary diffusion coefficient, D_1 and D_2 are the self-diffusion coefficients of substances 1 and 2 respectively, and N_1 and N_2 are the corresponding mole fractions.

Computational Details

In this work, we performed classical molecular dynamics (MD) simulations for the system of 1 cysteine dimer and 1984 water molecules in cubic simulation box of size about 3.90 nm at four different temperature; 288 K, 298 K, 303 K and 308 K using GROMACS 5.1.2 software package [17]. The cysteine dimer was extracted from the insulin molecule (pdb entry: 3i40) taking seventh residue of chain A and also seventh residue of chain B. The input files for MD simulation were generated from CHARMM-GUI online software program [18]. The dimer was solvated in a periodic boundary condition (pbc) box with TIP3P water sample. All the bonded and non-bonded parameters are used, assigned by CHARMM36 force field [19].

MD simulation was begun from energy minimization run using Steepest-descent algorithm [17]. This run removes steric clashes in which the undesirable coordinates of atoms of side chain or backbone that may occupy wrong position in the same coordinate space and bring the system in the condition of minimum potential energy state. Since the transport properties like diffusion depends on the thermodynamic parameters like temperature, pressure etc. of the system [20], the system should be stabilized for thermodynamic parameters. Then, the equilibration run was carried out to stabilize the temperature, pressure, density etc. The equilibrated bio-molecular system saves the computation efforts by removing external unwanted forces [21, 22]. To bring the system in thermodynamics equilibrium, equilibration run of 100 ns with time step of 1 fs was carried out in NPT ensemble for the system at each temperature. During the equilibration run; LINCS algorithms, Berendsen barostat with coupling time of 0.8 ps and Velocity rescaling thermostat with coupling time of 0.01 ps were used to constraint all the bonds, maintain constant pressure and maintain constant temperature respectively [17]. Also, 1 nm cut-off distance was taken for Lennard-Jones and Coulomb interactions; and PME (Particle-mesh Ewald) method was used to handle long range coulomb interaction. Furthermore, Maxwell-Boltzmann distribution and was used to

assign initial velocities of each particle; and the new positions and velocities of the particles was calculated using leapfrog algorithm [17] after each time step respectively.

After equilibration of the system, production run was performed in NVT ensemble for 100 ns with time step of 1 fs for each temperature. During the equilibration run, velocity-rescaling thermostat with coupling time of 0.01 ps was used and the initial velocities of each particle was taken from the final step of equilibration run.

3.   Results and Discussion

Diffusion coefficient

In this section, we present the self diffusion coefficients of water and cysteine dimer as well as their binary diffusion coefficient at different temperature. The self diffusion coefficient of both solute as well as solvent are estimated from their MSD versus time graph. Although, all the production run were carried out for 100 ns, we have plotted the MSD graph for 3 ns and 5 ns for cysteine dimer and water respectively due to the region that statistics is better in beginning region of the graph. Figs. 2 and 3 show the msd versus time graph for cysteine dimer and water respectively at four different temperatures: 288 K, 298K, 303 K and 308 K. From the Figures, it is clearly seen that the slope of the graph increases with temperature for both cysteine dimer as well as water. This indicates that self diffusion increases with temperature.
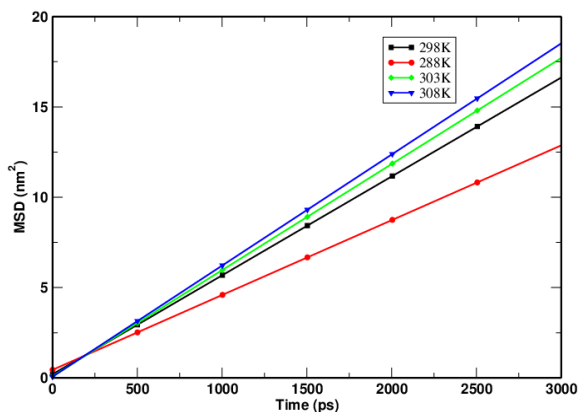

Figure 2. MSD versus time plot of cysteine dimer at four different temperature.
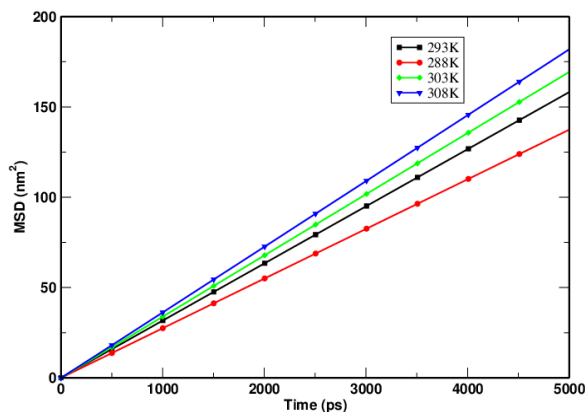
Figure 3. MSD versus time plot of water at four different temperature.

Also, the estimated values of the self diffusion coefficients of cysteine dimer and water estimated from the slope of the MSD versus time graph using Einstein relation; and their binary diffusion coefficients using Darken's relation are presented in Table 1.

From the Table 1, it is seen that both self and binary diffusion coefficients increases with increase in temperature. This is due to the increase in random velocity of the particles and decrease in density of the system with increase in temperature. As a result, the available space for diffusion increases with temperature. The simulated value of self diffusion coefficient at 298 K is in closely agreement with previously reported simulated value

[23]. Also, from the Table, we observe that the estimated values of self diffusion coefficient of water at different temperature are greater than the previously reported experimental values. The reason for this observation is that we used TIP3P water model during the simulations which overestimate the value of diffusion coefficient. Furthermore, we observed that the binary diffusion coefficient of the system equals to corresponding value of self diffusion coefficient of cysteine dimer, which is due to infinite dilute concentration of cysteine dimer in the solution.

Table 1. Estimated values of self and binary diffusion coefficients at different temperature.

		Diffusion coefficients $(D_{PBC})(10^{-9}m^2\ s^{-1})$				
SN	Temp.	Self				Binary
		For cysteine dimer	For Water			
	(K)	MSD	MSD	Reference[23]	Experiment [24]	Calculated
1.	288	0.69	4.58		1.77	0.69
2.	298	0.91	5.27	5.4	2.30	0.91
3.	303	0.98	5.65		2.60	0.98
4.	308	1.02	6.06		2.90	1.02

Temperature Dependency of Diffusion

The Table 1 shows the temperature dependent behavior of diffusion coefficient. In order to check whether the behavior is Arrihenius or not, we have plotted graphs between $ln(D)$ versus $(1/T)$. Fig. 4 shows the temperature dependent behavior of self diffusion of water which follows the Arrhenius equation [25]:

$$\ln D = \ln D_0 - \frac{E_a}{N_A k_B T} \qquad (3)$$

In equation (3), D is the diffusion coefficient, D_0 represents pre-exponential factor, E_a is the activation energy for diffusion, N_A is Avogardo's number whose value is 6.022×10^{23} mol^{-1}, k_B is the Boltzmann's constant whose value is 1.38×10^{-23} JK^{-1} and T is the absolute temperature. The intercept when extrapolated to the $1/T \to 0$ in the Arrhenius plot gives the pre-exponential factor.
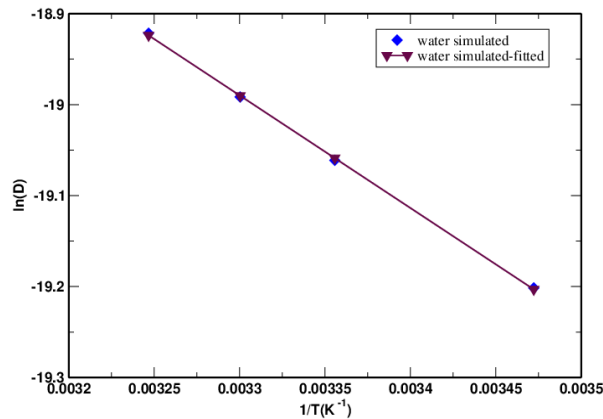

Figure 4. Arrhenius diagram for simulated values of self diffusion coefficient of water.

4. Conclusions

In this work, molecular dynamics simulation has been carried out to study the diffusion properties of cysteine dimer with disulphide bond in infinite dilute solution of water. This study is performed solvating single cysteine dimer molecule in 1984 water molecules at temperature 288 K, 298 K, 303 K, 308 K using GROMACS 5.1.2 software package. CHARMM36 force field parameters and TIP3P water model are used throughout the simulations. Einstein′s equation is used to determine the self diffusion coefficient of cysteine dimer and the water. The estimated values of self diffusion coefficients are found to be higher than the previously reported experimental values. This is due to the fact that TIP3P water model overestimate the diffusion coefficient. Also, the diffusion of binary mixture of cysteine dimer and water was calculated from Darken′s relation. The temperature dependence of diffusion coefficient has been tested from Arrihenous plot.

In near future, we are intended to study the transport properties of poly-cysteine with peptide bond in aqueous medium.

5. Acknowledgements

References

[1] Bulaj G. Formation of disulfide bonds in proteins and peptides. Biotechnology advances. 2005;23(1):87–92.

[2] Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. Disulfide bonds and protein folding. Biochemistry. 2000;39(15):4207–4216.

[3] Canet-Avilés RM, Wilson MA, Miller DW, Ahmad R, McLendon C, Bandyopadhyay S, et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proceedings of the National Academy of Sciences. 2004;101(24):9103–9108.

[4] Rizvi SI, Jha R. Strategies for the discovery of anti-aging compounds. Expert opinion on drug discovery. 2011;6(1):89–102.

[5] Privalov P, Khechinashvili N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. Journal of molecular biology. 1974;86(3):665–684.

[6] Burley S, Petsko G. Weakly polar interactions in proteins. In: Advances in protein chemistry. vol. 39. Elsevier; 1988. p. 125–189.

[7] Sticke DF, Presta LG, Dill KA, Rose GD. Hydrogen bonding in globular proteins. Journal of molecular biology. 1992;226(4):1143–1159.

[8] Thapa S, Adhikari N. A molecular dynamics study of oxygen gas in water at different temperatures. International Journal of Modern Physics B. 2013;27(08):1350023.

[9] Bhandari D, Adhikari N. Molecular dynamics study of diffusion of krypton in water at different temperatures. International Journal of Modern Physics B. 2016;30(11):1650064.

[10] Ghimire S, Adhikari NP. Study of structural and transport properties of argon, krypton, and their binary mixtures at different temperatures. Journal of molecular modeling. 2017;23(3):94–107.

[11] Poudyal I, Adhikari NP. Temperature dependence of diffusion coefficient of carbon monoxide in water: A molecular dynamics study. Journal of Molecular Liquids. 2014;194:77–84.

[12] Crank J. The mathematics of diffusion. Oxford University Press; 1975.

[13] Hirakawa H, Kamei Y, Sugisaki M, Oishi Y. Relationship between Self-Diffusion and Interdiffusion in Gaseous Systems. Bulletin of the Chemical Society of Japan. 1973;46(9):2659–2662.

[14] Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. vol. 1. Elsevier; 2002.

[15] Allen M, Tildesley D. Computer Simulations of Liquids, Oxford: Clarendeon Press. Oxford University Press; 1989.

[16] Darken LS. Diffusion, mobility and their interrelation through free energy in binary metallic systems. Trans Aime. 1948;175:184–201.

[17] Abraham M, van der Spoel D, Lindahl E, Hess B, team atGd. GROMACS User Manual Version 5.1. 2. GROMACS Development Team. 2016;.

[18] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. Journal of chemical theory and computation. 2016;12(1):405–413.

[19] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of computational chemistry. 2010;31(4):671–690.

[20] Khanal SP, Kandel YP, Adhikari NP. Transport properties of zwitterion glycine, diglycine, and triglycine in water. AIP Advances. 2019;9(6):065303.

[21] Walton EB, VanVliet KJ. Equilibration of experimentally determined protein structures for molecular dynamics simulation. Physical Review E. 2006;74(6):061901.

[22] Gallo MT, Grant BJ, Teodoro ML, Melton J, Cieplak P, Phillips Jr GN, et al. Novel procedure for thermal equilibration in molecular dynamics simulation. Molecular simulation. 2009;35(5):349–357.

[23] Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. The Journal of Physical Chemistry A. 2001;105(43):9954–9960.

[24] Holz M, Heil SR, Sacco A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. Physical Chemistry Chemical Physics.

2000;2(20):4740–4742.

[25] Mehrer H. Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes. vol. 155. Springer Science & Business Media; 2007.

Intra-molecular Conformational Stability in Human Growth Hormone

R. P. Koirala[1], B. Thapa[2], S. P. Khanal[1], R. P. Adhikari[3] and N. P. Adhikari[1,*]

[1]Central Department of Physics, Tribhuvan University, Kathmandu, Nepal
[2]Padma Kanya Multiple Campus, Tribhuvan University, Kathmandu, Nepal
[3]Department of Natural Sciences, Kathmandu University, Dhulikhel, Nepal
*Corresponding Email: narayan.adhikari@cdp.tu.edu.np

Abstract

Human growth hormone (hGH) is synthesized, stored and secreted by somatotropic cells within the lateral wings of the anterior lobe of pituitary glands; and is transported to other organs of human body. Study of intra-molecular structure and its binding mechanisms within the molecule gives more insight of structural stability of the molecule and is also essential in drug designing. In this article, we have investigated the various bonded and non-bonded interactions that contribute for the conformation of entire structure of the hGH molecule using molecular dynamics (MD) simulation. The MD outcomes show that the molecule is hydrophobic in nature. In its conformation, several types of interactions exist, such as disulphide bridges (bonded) and nonbonded: hydrogen bond, hydrophobic, aromatic-aromatic, ionic, aromatic-sulphur, cation-pi.

Keywords: Intra-binding, Growth hormone, Hydrophobic, Aromatic.

1. INTRODUCTION

Personal height is the highly concerned matter for everyone. It is also a measure of health conditions [1]. Human growth hormone (hGH) is essential for the proper development of height as well as every part of body. It is a single chain peptide hormone that stimulates growth, reproduction and regeneration of cell in human [2]. Although growth hormone is found in other animals, its effect is significant only in human and old world monkey. Many researches have revealed that the variation of amino acid sequence of hGH is significant in different species [3]. Even though the sequence is different, its role in other animals is also similar to that in human body. This hormone contains 191 amino acid residues with 22,124 daltons molecular weight. It is synthesized, stored and secreted by somatotropic cells within the lateral wings of the anterior lobe of pituitary glands [4]. The hormone after secretion from the gland mixes into the blood stream and is transported into body cells [5]. The structure includes four helices necessary for the functional interaction with the growth hormone (GH) receptor [6]. Human growth hormone is essential in physical development. Main problem of human growth in childhood is the short stature, i.e., insufficient growth in accordance with age; and delayed sexual maturity in adulthood [7].

Human growth hormone, also known as somatotropin, provides important contribution in human development. It increases the concentration of glucose and free fatty acids [1]. It is legally prescribed as a drug to treat the children's growth disorders and adult growth hormone deficiency. Moreover, it is also used in raising livestock more efficiently in industrial agriculture.

Besides its functions of somatic and bone growth as well as increase in the size and mass of organs and tissues, hGH also influences in the functioning of proteins, carbohydrates and lipids metabolisms [8]. These mechanisms caused by hGH are basically due to its ability to bind with specific target cell receptor. hGH surface specifically favors for the inter-molecular bindings at the interfacial region of receptor. Inter-molecular binding occurs in the expense of intra-molecular binding, which ultimately lowers the conformation stability of the molecule [9, 10]. There are several non-bonded

interactions to form a macromolecule a stable structure. Hydrogen bonding play a pivotal role in the formation of secondary structure of a protein. Likewise, disulphide bond is a covalent bonding to give the shape in tertiary structure. Similarly, aromatic-aromatic interactions between two aromatic rings separated by distance 4.5 Å to 7 Å provide the conformation stability. This interaction is energetically favorable and basically applies to form the tertiary and quaternary structure [11, 12]. Similarly, electrostatic and van der Waals interactions are pervasive in intra- and inter-molecular bindings [13]. Likewise, hydrophobic interaction has influences in conformation changes of protein molecule in aqueous environment [14].

In this work, we have used molecular dynamics simulation to study the intra-molecular interactions and the contributions of hydrogen bonding and several other interactions such as hydrophobic, ionic, aromatic-aromatic, aromatic-sulphur and cation-pi that are responsible to provide the stability of hGH molecule. Moreover, we have compared the intra-molecular interactions of amino acid residues in static structure with the structure obtained from molecular dynamics simulations. To our best knowledge, the comprehensive study on molecular stability of this hormone has not been studied yet via MD run. We believe that this work aids the in-depth knowledge about the binding of amino acid residues in order to form the stable structure of the hormone in cellular environment.

2. MATERIALS AND METHODS

To perform molecular dynamics simulation, one needs initial structure of molecule (pdb file) and force field parameters. In this section, we describe system setup of the hGH molecule, force field parameters used in the present work and simulation details.

System setup: The molecular structure of human growth hormone was taken from protein data bank with PDBID 1HGU.pdb [15]. In original file, two amino acids were missing at the N-terminus and C-terminus positions, which were filled with software program CHARMM-GUI [16]. This software program was also used to generate new protein structure file (psf) and pdb files. To mimic the cellular environment, the hormone molecule was solvated in TIP3P water in a cubical box of dimensions $84 \times 84 \times 84$ Å^3. As TIP3P water model resembles real water closely, we chose this model in this work. Furthermore, the molecular system had originally 4 excess negative electronic charges,

so the system was electrically neutralized by adding 4 Na+ ions.

Force fields: Force fields contain topology and parameter files. Topology files are used to generate psf files and parameter files contain the information about parameters of potential energy functions. In the present work, we used CHARMM36m [17] force field.

Molecular Dynamics Simulation: All-atom molecular dynamics (MD) simulations were carried out by using NAnoscale Molecular Dynamics (NAMD) [18] simulation package. The Particle Mesh Edward (PME) was used to treat the long-range interactions with a 12.0 Å non-bonded cut off. The energy minimization was performed for 10,000 steps, using the conjugate gradient algorithm. Energy minimization run removes the unwanted hindrances between the atoms in the system. Since this run is performed in 0 K temperature, the system chooses the local minimum energy state [19, 20]. After energy minimization, the system was equilibrated at 310 K under the isothermal-isochoric conditions for 10 ns with 1 fs time step [21,22]. Then, the production run was propagated for 100 ns under NVT simulation run taking time step 2 fs by using Langevin dynamics with a damping constant of 1 ps^{-1}.

The NAMD energy plugin package available in Visual Molecular Dynamics (VMD) [23] was used to estimate structural stability of the hGH molecule in aqueous environment. Protein interaction calculator (PIC) [24] has been used to analyze the intra-molecular interactions. The study on intra-molecular interactions depicts the detailed insights on folded state of the targeted molecule.

3. RESULTS AND DISCUSSION

This research work has been carried out by molecular dynamics simulation to study the amino acid arrangements, their folding mechanisms, the contributions of hydrogen bonding, and various other non-bonded interactions to form such a structure of human growth hormone in aqueous environment. The solvent accessible surface area and the energy profiles of bonded and non-bonded interactions have also been investigated to know the contact area and the stability of hGH in the aqueous environments.

Structural stability of human growth hormone: The pre-requisite of examining the structural stability of any biomolecular system is the estimation of root mean square deviation (RMSD). RMSD measures the structural stability of entire molecule in the given environment. We have taken

the hGH molecule in aqueous environment in order to resemble the cellular condition. RMSD of the molecule has been calculated from VMD analysis tool "RMSD Trajectory Tool". The mathematical relation to estimate the RMSD is,

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i(t) - r_i^{ref}\right)^2}$$

Where, $r_i(t)$ and r_i^{ref} represent the current and reference coordinates of ith atom respectively and N is total number of atoms.
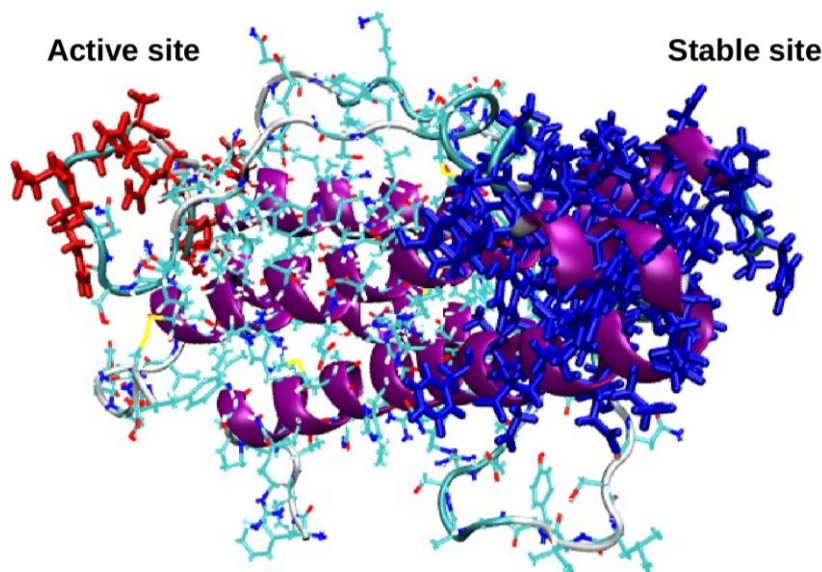
In this process, initial coordinates are taken as the reference and average deviation of molecule from the reference coordinates has been calculated in every frame of simulation. During the 100 ns NVT production run, the RMSD of the hGU was found fairly stable after 4 ns simulation (100 frames) time as shown in Fig. 1(i).
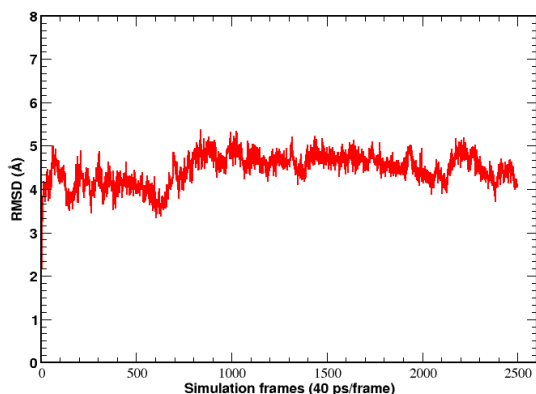
(i)
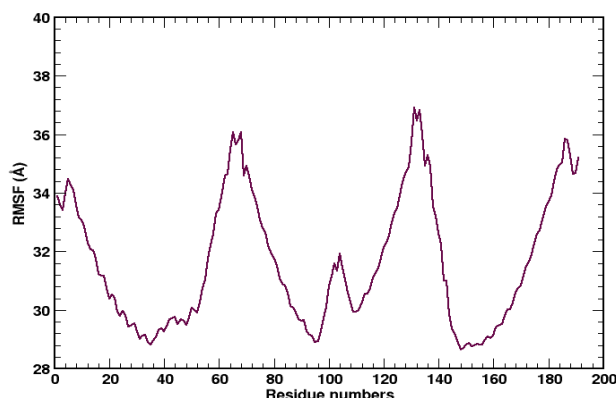
(ii)

(iii)

Fig. 1: Structural characterization of hGH in aqueous environment (i) RMSD (ii) RMSF and (iii) hGH molecule.

In addition, we have also determined the fluctuation of alpha carbons of hGH residues during the simulations through root mean square fluctuation (RMSF) as shown in Fig. 1(ii), which gives the flexibility of residues within the molecular system. Greater value of RMSF of any alpha carbon implies the greater flexibility of corresponding amino acid in the protein chain and vice versa. The mathematical formula to find the RMSF is,

$$RMSF = \sqrt{\left\langle \left(r_i(t) - \langle r_i \rangle\right)^2 \right\rangle}$$

where, $r_i(t)$ and $\langle r_i \rangle$ are the current position and average position respectively.

During the 100 ns simulation run, it has been observed that the regions of amino acid residues 64-68 and 130-136 are most fluctuating and the regions of 23-49, 88-98 and 144-164 are the most stable. The most fluctuating and stable regions are shown in Fig. 1 (iii).

Intra-binding potential energy: There are several interactions existing within a protein molecule to provide it a stable conformation. We have estimated the contributions of bonded and non-bonded potential energy in hGH molecule. Energy

profile for bonded interactions have been determined from 100 ns simulation. From the energy profile, it has been observed that improper dihedral has lowest contributions; and the contributions of harmonic and dihedral angle show almost equal and the largest value as shown in Fig. 2. During the conformation of protein molecules, bonded energy provides the direct covalent link among the atoms, whereas the non-bonded energy contributes to fold the molecule forming the stable structure [25-27]. On analyzing the non-bonded energy, the contributions of electrostatic interaction has found to be significantly higher potential energy than that of van der Waals.
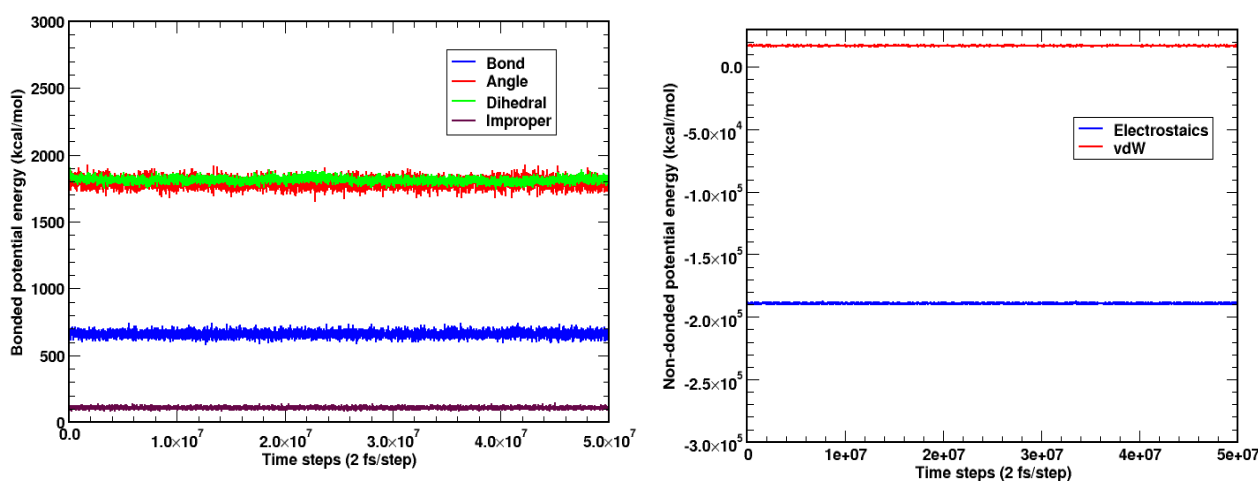

Fig. 2: Energy profiles for hGH (i) bonded (bond, angle, dihedral and improper) interaction (ii) non-bonded (electrostatics and van der Waals) interactions.

Intra-molecular hydrogen bonds: We have also investigated the hydrogen bonding within the molecule to form the hGH molecule. Hydrogen bonding is essential to form secondary structure of protein. The intra-molecular hydrogen bonding provides the important information for drug designing [28, 29]. The number of hydrogen bonds can determine the conformational stability in the molecule. Its role in intermolecular interactions is also very important, which enhances the cooperativity among the molecules in protein-protein, protein-ligand and protein-nucleic acid systems. We have found total 1294 hydrogen bonding in entire 100 ns simulation, and average of 65 hydrogen bonds persisted in each frame of simulation. The number of hydrogen bonds within the cutoff distance of 3.5 Å has been presented in Fig. 3.
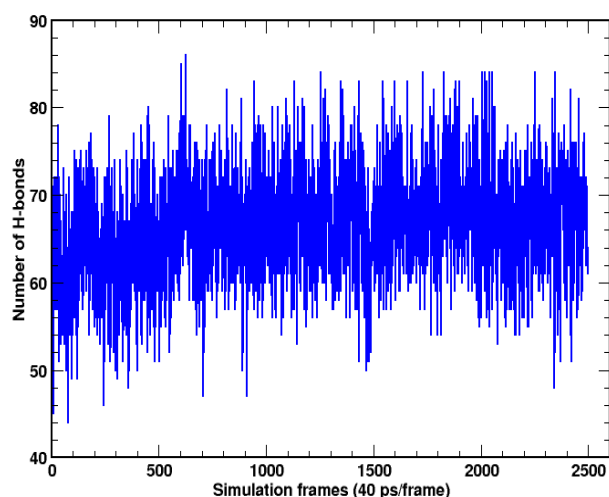

Fig. 3: Time variation of number of hydrogen bonding for intra-molecular binding in hGH.

Many intra-molecular hydrogen bonds were observed in both static and dynamic conditions of the hGH structure. We have shown the hydrogen bonding patterns for three interacting pairs. These three pairs are the representative pairs for main

chain – main chain, side chain – side chain and main chain – side chain as shown in Fig. 4. There are several such type of hydrogen bonds were observed to form the stable structure of the hormone.
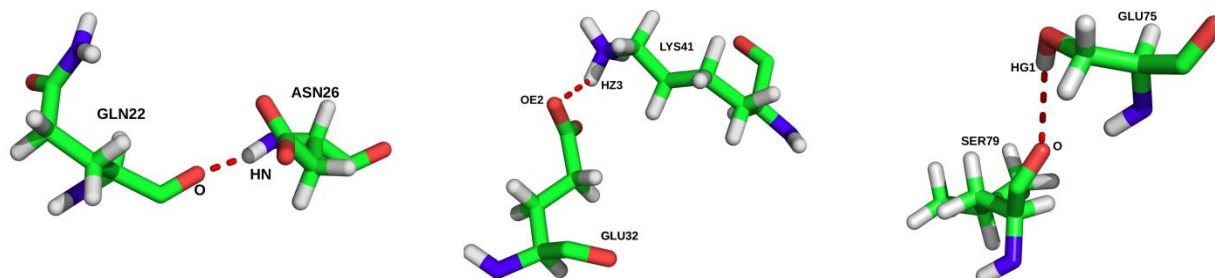

Fig. 4: Intra-molecular hydrogen bonds: main chain – main chain (left), side chain – side chain (middle) and main chain – side chain (right).

Surface accessible surface area (SASA): SASA measures the surface area of a molecule that contacts with the solvent molecules. We have taken water as the solvent to study how the amino acid residues residing on the surface of hGH behave in the aqueous environment. Fig. 5 shows the SASA plot with respect to simulation frame. During the 100 ns simulation, the overall value of SASA has been observed decreasing. The decreasing nature of graph shows the internal rearrangement of amino acid residues that resides on the surface. This shows that the molecule is hydrophobic in nature so that the surface residues tend to aggregate together to minimize the surface area exposed to water.
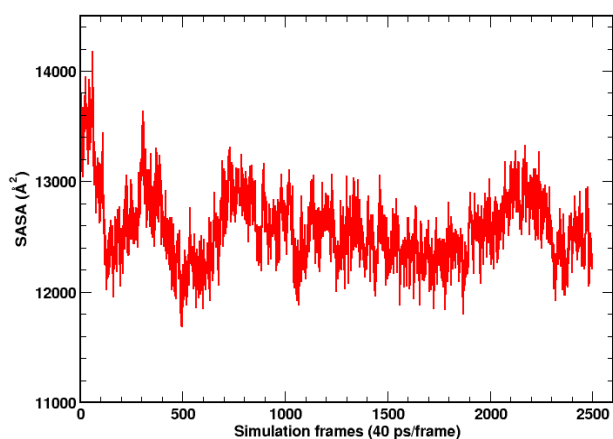

Fig. 5: Solvent accessible surface area (SASA) of hGH in water environment.

Intra-molecular Interactions: We have computed the various interactions present within the human

growth hormone such as hydrophobic interactions, disulphide bridges, ionic interactions, aromatic-aromatic interactions, aromatic-sulphur interactions and cation-pi interactions using protein interaction calculator (PIC) web server. We have compared these intra-molecular interactions in static structure with the most stable structure obtained from molecular dynamics simulations. The most stable structure was taken from the MD trajectories corresponding to the minimum SASA.

There are two disulphide bridges formed between two sulphur atoms of cysteine residues CYS182-CYS189 and CYS53-CYS165 within 2.2 Å in both structures. The distance between two cysteine residues in CYS182-CYS189 and CYS53-CYS165 is 2.02 Å in static structure, whereas 2.06 Å and 2.03 Å respectively in the structure from dynamics.

The hydrophobic interactions are of particular importance for the structural stability of hGH. There are 116 interactions between the hydrophobic residues such as PHE, LEU, ALA, MET, PRO, VAL, ILE, TYR and TRP present in the static structure. We have observed the decrease in the solvent accessible surface area (SASA) of protein during the simulation in aqueous environment that suggests the hydrophobic nature of the molecule. This is corroborated further by the greater number of hydrophobic interactions in the dynamics, which are 143 as compared to 116 in static. The cut off distance of hydrophobic interaction was taken to be 5 Å.

In addition, the ionic interactions between the charged residues lying within the cut off distance of 6 Å present in both structures of hGH protein are depicted in tables 1 and 2. We have observed 29

45

ionic interactions in structure from dynamics, whereas only 11 in static. Salt bridges are the important interactions in forming the tertiary structure of protein molecules. They are the ionic interactions between charged atoms in residues lying within hydrogen bond distance (3.5 Å). In the structure obtained from MD simulation 7 salt-bridges has been observed whereas only 2 salt-bridges are obtained in static structure.
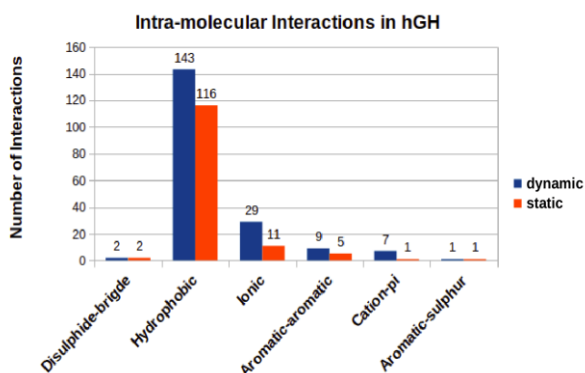

Fig. 6: Bar diagram showing different intra-molecular interactions in hGH molecule.

The bar diagram shown in Fig. 6 summarizes the various intra-molecular interactions present in hGH molecule in both structures i.e., static and dynamic conditions. The comparison diagram shows that hydrophobic interactions have the maximum binding sites and aromatic-sulphur has minimum binding sites.

**Table 1: Ionic Interactions within
6 Å (in static structure)**

Position	Residue	Position	Residue
16	ARG	107	ASP
16	ARG	116	ASP
18	HIS	174	GLU
21	HIS	174	GLU
32	GLU	41	LYS
94	ARG	109	ASP
115	LYS	118	GLU
145	LYS	169	ASP
154	ASP	158	LYS
167	ARG	171	ASP
171	ASP	172	LYS

**Table 2: Ionic Interactions within
6 Å (structure in dynamic)**

Position	Residue	Position	Residue
16	ARG	116	ASP
19	ARG	107	ASP
19	ARG	26	ASP
26	ASP	29	GLU
29	GLU	30	GLU
29	GLU	32	GLU
32	GLU	33	GLU
32	GLU	41	LYS
33	GLU	41	LYS
38	LYS	39	GLU
64	ARG	65	GLU
94	ARG	109	ASP
94	ARG	112	ASP
109	ASP	112	ASP
112	ASP	115	LYS
115	LYS	116	ASP
115	LYS	118	GLU
116	ASP	119	GLU
118	GLU	119	GLU
129	GLU	130	ASP
129	GLU	134	ARG
130	ASP	134	ARG
153	ASP	154	ASP
168	LYS	169	ASP
168	LYS	171	ASP
169	ASP	172	LYS
171	ASP	172	LYS
171	ASP	174	GLU
174	GLU	178	ARG

The aromatic-aromatic interactions also play important role in the structural stability of proteins. These interactions are significant within 4.5 to 7 Å range between aromatic residues. In human growth hormone protein, we have detected 9 and 5 aromatic-aromatic interactions for the dynamic and static structures respectively. These interactions are presented in tables 3 and 4. Moreover, there is an aromatic-sulphur interaction within 5.3 Å present between PHE166 and MET170 residues in both

structures. In static structure, the distance between the centroid and sulphur was observed 4.66 Å and the angle was 72.7°, whereas in dynamic structure the distance between the centroid and sulphur is 5.01 Å and the angle is 44.6°.

The cation-pi interactions occur between the cations of the side chains of arginine or lysine residues with the polarizable pi electron of the aromatic ring. These interactions are distance dependent interactions. Here, we have taken the cut off distance of 6 Å. In static structure, there is only one cation-pi interaction present between the residues LYS41 and TYR164 within the distance of 5.8 Å and at an angle of 119.7°. However, in case of structure obtained from dynamics, we observed 7 cation-pi interactions between the ARG and LYS residues with the aromatic residues PHE and TRP as shown in table 5.

Table 3: Aromatic-Aromatic Interactions within 4.5 and 7 Å (in static structure).

Position	Residue	Position	Residue	D(Centroid-Centroid)	Dihedral Angle
25	PHE	28	TYR	6.93	122.2
31	PHE	35	TYR	4.73	132.8
54	PHE	143	TYR	6.05	25.35
86	TRP	166	PHE	4.93	50.11
160	TYR	164	TYR	6.24	95.19

Table 4: Aromatic-Aromatic Interactions within 4.5 and 7 Å (structure in dynamic).

Position	Residue	Position	Residue	D(Centroid-Centroid)	Dihedral Angle
25	PHE	28	TYR	6.14	144.03
28	TYR	160	TYR	4.74	8.09
28	TYR	164	TYR	6.07	31.76
31	PHE	35	TYR	5	118.96
54	PHE	143	TYR	6.74	50.35
86	TRP	166	PHE	5.63	51.6
86	TRP	97	PHE	6.4	22.9
97	PHE	146	PHE	5.18	123.48
97	PHE	166	PHE	4.79	37.53

Table 5: Cation-Pi Interactions within 6 Å (structure in dynamic)

Position	Residue	Position	Residue	D (Cation-Pi)	Angle
1	PHE	16	ARG	5.15	108.14
1	PHE	8	ARG	5.32	119.02
25	PHE	167	ARG	3.9	13.26
42	TYR	41	LYS	4.54	141.61
111	TYR	94	ARG	3.98	156.72
139	PHE	77	ARG	3.93	149.17
191	PHE	64	ARG	4.45	47.09

4. CONCLUSIONS

We have carried out the molecular dynamics (MD) simulations in order to identify the intra-molecular contacts and their contributions in forming the stable structure of human growth hormone. The MD run was propagated for 100 ns simulation

under NVT condition at body temperature 310 K. We have studied bonded and non-bonded interactions which contribute in the formation of stable structure in aqueous environment. Many non-bonded contacts are compared in static and dynamic conditions.

The decreasing nature of solvent accessible surface area (SASA) depicts the hydrophobic nature of hGH which makes the structure more stable in water. We have compared the various intra-molecular interactions namely disulphide bonds, hydrophobic, ionic, aromatic-aromatic, aromatic-sulphur, cation-pi interactions between the structures from dynamics and statics. In the structure taken from the simulation at the condition of minimum SASA, we have found hydrophobic, ionic, cation-pi, aromatic-aromatic interactions significantly higher than that of static structure. The hydrogen bonding is observed strongly contributing in intramolecular binding. In addition, the estimation of bonded and non-bonded energy profiles show that electrostatic interactions have higher contributions in non-bonded condition. Similarly, the harmonic angle and dihedral angle have almost equal contribution in the conformation of the molecule. Bond energy contribution has also significant role but smaller than that of angle and dihedral.

ACKNOWLEDGEMENTS

REFERENCES

[1] Fayter, D.; Nixon, J.; Hartley, S.; Rithalia, A.; Butler, G.; Rudolf, M.; Glasziou, P.; Bland, M.; Stirk, L.; and Westwood, M. Effectiveness and cost-effectiveness of height-screening programmes during the primary school years: a systematic review, *Archives of disease in childhood,* **93:** 278-284 (2008).

[2] Greenwood, F.; and Landon, J. Growth hormone secretion in response to stress in man, *Nature,* **210:** 540-541 (1966).

[3] Aloj, S.; and Edelhoch, H. The molecular properties of human growth hormone, *Journal of Biological Chemistry,* **247:** 1146-1152 (1972).

[4] Kohler, M.; Püschel, K.; Sakharov, D.; Tonevitskiy, A.; Schänzer, W.; and Thevis, M. Detection of recombinant growth hormone in human plasma by a 2-D PAGE method, *Electrophoresis,* **29:** 4495-4502 (2008).

[5] Li, C. H. Human growth hormone: 1974–1981, *Molecular and cellular biochemistry,* **46:** 31-41 (1982).

[6] Yi, S.; Bernat, B.; Pál, G.; Kossiakoff, A.; and Li, W.-H. Functional promiscuity of squirrel monkey growth hormone receptor toward both primate and nonprimate growth hormones, *Molecular biology and evolution,* **19:** 1083-1092 (2002).

[7] Koirala, R. P.; Pradhan, S.; and Aryal, S. K. Ultrasonic Measurement of Kidney Length in Nepalese People. *Journal of Nepal Physical Society,* **4**(1): 49-53 (2017).

[8] Wells, J. A.; Cunningram, B. C.; Fuh, G.; Lowman, H. B.; Ultsch, M.; Devos, A. M.; Bass, S. H.; Mulkerrin, M. G. The molecular basis for growth hormone–receptor interactions, In *Recent Progress in Hormone Research*, 253-275, Elsevier (1993).

[9] Meiering, E. M.; Serrano, L.; and Fersht, A. R. Effect of active site residues in barnase on activity and stability, *Journal of molecular biology,* **225:** 585-589 (1992).

[10] Schulga, A. A.; Makarov, A. A.; Levichkin, I. V.; Belousova, Y. V.; Lobachov, V. M.; Protasevich, I. I.; Pace, C. N.; and Kirpichnikov, M. P. Increased stability of human growth hormone with reduced lactogenic potency, *FEBS letters,* **528**: 257-260 (2002).

[11] Burley, S.; and Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization, *Science,* **229:** 23-28 (1985).

[12] Bhattacharyya, R.; Samanta, U.; and Chakrabarti, P. Aromatic–aromatic interactions in and around α-helices, *Protein engineering,* **15**: 91-100 (2002).

[13] Dahiyat, B. I.; and Mayo, S. L. Protein design automation, *Protein Science,* **5**: 895-903 (1996).

[14] Kellis, J. T.; Nyberg, K.; and Fersht, A. R. Contribution of hydrophobic interactions to protein stability, *Nature,* **333**: 784-786 (1988).

[15] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; and Bourne, P. E. The protein data bank, *Nucleic acids research,* **28**: 235-242 (2000).

[16] Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L. 3rd, MacKerell, A. D.; Jr., Klauda, J. B.; and Im, W. CHARMM-GUI Input Generator for NAMD,

GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, *J Chem Theory Comput,* **12**: 405-413 (2016).

[17] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; and MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat Methods,* **14**: 71-73 (2017).

[18] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; and Schulten, K. Scalable molecular dynamics with NAMD, *J Comput Chem,* **26**: 1781-1802 (2005).

[19] Koirala, R. P.; Bhusal, H. P.; Khanal, S. P.; and Adhikari, N. P. Effect of temperature on transport properties of cysteine in water, *AIP Advances,* **10**: 025122 (2020).

[20] Ramachandran, S.; Kota, P.; Ding, F.; and Dokholyan, N. V. Automated minimization of steric clashes in protein structures, *Proteins,* 79: 261-270 (2011).

[21] Pantha, N.; Chauhan, B.; Sharma, P.; and Adhikari, N. P. Tuning Structural and Electronic Properties of Phosphorene with Vacancies. *Journal of Nepal Physical Society*, **6**(1): 7-15 (2020).

[22] Walton, E. B.; and Vanvliet, K. J. Equilibration of experimentally determined protein structures for molecular dynamics simulation, *Phys Rev E Stat Nonlin Soft Matter Phys,* 74: 061901 (2006).

[23] Humphrey, W.; Dalke, A.; and Schulten, K. VMD: visual molecular dynamics, *J Mol Graph,* 14: 33-38, 27-38 (1996).

[24] Tina, K.; Bhadra, R.; and Srinivasan, N. PIC: protein interactions calculator, *Nucleic acids research,* 35: W473-W476 (2007).

[25] Neopane, S.; and Pantha, N. First-Principles Study of van der Waals Interactions between Halogen Molecules (Cl2 and I2). *Journal of Nepal Physical Society*, **5**(1): 19-23 (2019).

[26] Thakuria, R.; Sarma, B.; and Nangia, A. 7.03 Hydrogen Bonding in Molecular Crystals, *Comprehensive Supramolecular Chemistry,* **II**: 25-48 (2017).

[27] DiStasio Jr, R. A.; Gobre, V. V.; and Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter, *Journal of Physics: Condensed Matter,* **26**: 213202 (2014).

[28] Yunta, M. It is important to compute intramolecular hydrogen bonding in drug design, *American Journal of Modeling and Optimization,* **5**: 24-57 (2017).

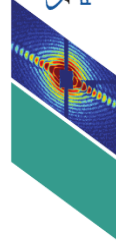[29] Hubbard, R. E.; and Haider, M. K. Hydrogen bonds in proteins: role and strength, *eLS* (2010).

Certificate of Attendance

American Physical Society
March Meeting 2021
March 15–19, 2021 | Online

THIS IS TO CERTIFY THAT

Shyam Prakash Khanal

attended the American Physical Society's March Meeting.
Pre-meeting tutorials and short courses were held on March 13 and 14.

APS physics

MARCH
MEETING 2021

Don Wise, Senior Meetings Registrar
March 19, 2021

NON-RESIDENT NEPALI ASSOCIATION (NRNA)

CERTIFICATE
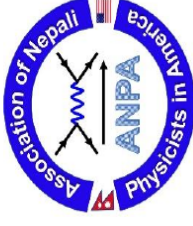OF APPRECIATION

presents this

To MR SHYAM PRAKASH KHANAL

for presenting a paper at the 1ST NRN GLOBAL KNOWLEDGE CONVENTION

organized by NRNA in partnership with the GOVERNMENT OF NEPAL

on 12 - 14 October 2018 in Kathmandu, Nepal

Hem Raj Sharma, PhD
Advisor, NRNA & Chair
1st NRN Global Knowledge Convention

Kumar Panta
Vice-President, NRNA & Chair
1st NRN Global Knowledge Convention

Bhaban Bhatta
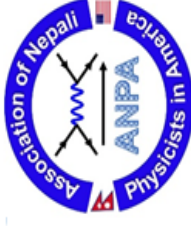*President, NRNA*

Certificate of Appreciation

THIS CERTIFICATE IS PRESENTED TO

MR. SHYAM P. KHANAL

Tribhuvan University, Nepal

FOR YOUR PRESENTATION DURING THE ANPA CONFERENCE 2020

Shree Bhattarai

DR. SHREE K. BHATTARAI
PRESIDENT, ANPA

DR. CHET R. BHATT
CHAIR, ANPA CONFERENCE 2020

Certificate of Appreciation

THIS CERTIFICATE IS PRESENTED TO

MR. SHYAM PRAKASH KHANAL

Central Department of Physics, Nepal

FOR YOUR PRESENTATION DURING THE ANPA CONFERENCE 2021

DR. JAGAN DEVKOTA
PRESIDENT, ANPA

DR. NABIN MALAKAR
CONVENER, ANPA CONFERENCE 2021

Lecture Series on

Research Methodology

6 November – 22 December 2017

Central Department of Physics
Tribhuvan University, Kirtipur, Nepal

Participation Certificate

Shyam Prakash Khanal

Central Department of physics T.U.

participated in **21 hours lecture series** on **Research Methodology** delivered by **Prof. Dr. Subodh R. Shenoy,** TIFR, India during 6 November to 22 December 2017.

Prof. Dr. Subodh R. Shenoy
Guest Speaker
Tata Institute of Fundamental Research, India

Prof. Dr. Binil Aryal
Head
CDP, TU, Kirtipur