MOLECULAR DYNAMICS SIMULATION OF TRANSLATIONAL AND ROTATIONAL DIFFUSION OF LIQUID ISOQUINOLINE

NORARIZA BINTI AHMAD

UNIVERSITI SAINS MALAYSIA

2010

MOLECULAR DYNAMICS SIMULATION OF TRANSLATIONAL AND ROTATIONAL DIFFUSION OF LIQUID ISOQUINOLINE

by

NORARIZA BINTI AHMAD

Thesis submitted in fulfillment of the requirements for the degree of Master of Sciences

DECEMBER 2010

ACKNOWLEDGEMENT

Alhamdulillah, 'His' willing has made it possible for me to complete my research successfully.

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Associate Professor Dr Rohana Adnan for her personal guidance which has provided a good basis for my project as well as the thesis. I am deeply grateful to my co-supervisor, Professor Claude Millot from Universitè Henri Poincarè, for his guidance during my first step into molecular dynamic simulation studies. Thank you very much for the constructive comments, co-operations and supports throughout this work.

I owe my loving thanks to my mother, Zainab Long for her continuous support, patience and understanding, who always give me encouragement to finish this work. My special gratitude goes to my brother, Azree and sister, Nurafzan for their endless support. To the memory of my father, Ahmad Che Soh – Al-Fatihah

I wish to thank all my friends, especially Lee-Sin, Zati Zuber and Husna Halim who contribute directly and/or indirectly in my life and this thesis. Thank you very much and may Allah bless all of us.

Special thanks to all the staffs in School of Chemical Sciences for their helps.

Last but not least, to Universiti Sains Malaysia for the financial support under FRGS 203/PKIMIA/671147. The members of Faculté des Science et Techniques, Université Henri Poincaré, Nancy, France are acknowledged for their helps and assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

avacf	angular velocity autocorrelation functions
avccf	angular velocity cross correlation functions
B3LYP	Becke 3-parameter exchange hybrid functional with Lee Yang Parr correlation functional
DSLI	depolarized light scattering intensity
ESP	electrostatic potential
MC	Monte Carlo
MD	Molecular dynamics
MP2	Moller-Plesset second order correction
MSD	Mean square displacement
NMR	Nuclear magnetic resonance
NPT	isobaric-isothermal ensemble, constant number of molecules, pressure, and energy
NVE	microcanonical ensemble, constant number of molecules, volume and energy
NVT	canonical ensemble, constant number of molecules, volume and temperature
OHD-RIKES	optical heterodyne-detected Raman-induced Kerr effect spectroscopy
OKE	optical Kerr effect
OPLS-AA	Optimized paramater for liquid system – all atom
OPLS-CS	Optimized paramater for liquid system – charge separated
PBC	Periodic boundary condition
QLRF	quasilattice random flight

RAM	random access memory
rdf	radial distribution functions
SED	Stokes-Einstein-Debye
vacf	velocity autocorrelation function
vdW	van der Waals
μVT	grand canonical ensemble, constant chemical potential, volume and temperature

LIST OF SYMBOLS

μ_0	magnetic permittivity of vacuum
γ_x	gyromagnetic factor
ε_0	dielectric constant
$\rho_i^q(\mathbf{r})$	Gaussian distribution
κ	width of distribution
ΔH_{vap}	enthalpy of vaporization
$\mathbf{a}_i$	acceleration of atom i
C_l	reorientational correlation functions; l = 1, 2
$D_{rot,a,b}$	rotational diffusion of diffusion tensor; a, $b = x, y, z$
$D_{rot,a'b'}$	rotational diffusion of inertia referential; $a', b' = x', y', z'$
D_{trans}	translational diffusion
E_a	activation energy
E_{angle}	energy associated to bond angle
E_{bond}	energy associated to bond length
E_{bonded}	bonded energy
$E_{dihedral}$	energy associated to dihedral angle
$E_{electrostatic}$	electrostatic energy
E_{LJ}	Lennard-Jones energy
$E_{nonbonded}$	nonbonded energy
E_{pot}	potential energy
E_{total}	total potential energy

E_{vdW}	van der Waals energy
$\mathbf{F}_i$	the force on atom *i*
$J_{l,m}$	spectral density
m	mass
$\mathbf{p}_i$	momentum of atom *i*
q_i	electric charge on atom *i*
q_j	electric charge on atom *i*
R	gas constant
r	vector position
r_{ij}	the interatomic distance between atom *i* and *j*
T	temperature in Kelvin
T_1	longitudinal relaxation time
ε_{ij}	well depth in Lennard-Jones potential of atom *i* and *j*
η	viscosity
σ_{ij}	sum of radii of atom *i* and *j* in Lennard-Jones potential
τ_1	first order reorientational correlation time
τ_2	second order reorientational correlation time

LIST OF PUBLICATIONS

1. N. Ahmad, R. Adnan, J.C. Soetens, C. Millot, *Molecular Dynamics simulation of liquid isoquinoline,* School of Chemical Sciences Research Colloquium 2009, Penang, 2-3 November 2009.

2. N. Ahmad, R. Adnan, J.C. Soetens, C. Millot, *Molecular Dynamics of translational and rotational diffusion of liquid isoquinoline,* USM-UHP Colloquium, Penang, November 2009.

3. N. Ahmad, R. Adnan, J.C. Soetens, C. Millot, *Molecular Dynamics simulation of translational and rotational of liquid isoquinoline,* International Conference on Neutron and X-ray Scattering (ICNX), Kuala Lumpur in July 2009.

SIMULASI DINAMIK MOLEKUL BAGI PEMBAURAN TRANSLASI DAN PUTARAN KE ATAS CECAIR ISOKUINOLIN

ABSTRAK

Simulasi dinamik molekul ke atas cecair isokuinolin dijalankan pada suhu di antara 300 hingga 365 K bagi mengkaji evolusi pembauran translasi dan putaran terhadap perubahan suhu. Interaksi antaramolekul dan intramolekul dimodelkan menggunakan potensi Coulomb dan Lennard-Jones yang diperoleh daripada medan daya Optimized Potential Liquid for Liquid-All Atoms (OPLS-AA). Pengiraan entalpi pengewapan yang diperoleh daripada simulasi ini bernilai 16.01 hingga 14.86 kcal.mol^{-1} daripada 300 hingga 365 K iaitu dalam lingkungan 10.5-11.5 % berbanding kajian eksperimen. Pekali pembauran translasi, D_{trans} diperolehi daripada pengiraan fungsi halaju autokorelasi (vacf) dan kaedah penyesaran kuasa dua min (MSD) manakala pekali pembauran putaran, D_{rots} diperoleh daripada fungsi halaju sudut autokorelasi (avacf). Nilai D_{trans} yang diperoleh daripada vacf iaitu 0.0061 hingga 0.0472 Å^2.ps^{-1} adalah lebih besar berbanding nilai $D_{trans,}$ (MSD) iaitu 0.0066 hingga 0.0500 Å^2.ps^{-1}. Perbezaan yang lebih besar didapati pada suhu lebih rendah. Pekali pembauran putaran pada paksi *x*, *y* dan *z* yang dihitung sebagai berkadaran dengan suhu mempamerkan sifat yang berbeza iaitu nilai $D_{rot,z}/D_{rot,y}$ hanya berubah daripada 2.23 hingga 2.32. Dalam julat 300 hingga 365 K, nisbah $D_{rot,z}/D_{rot,x}$ didapati lebih besar dengan nilai 1.20 hingga 1.43. Masa korelasi orientasi semula, τ_1 dan $\tau_{2,}$ bagi setiap vektor C-H diperoleh daripada penyesuaian fungsi korelasi ke atas fungsi dwieksponen. Pengiraan ke atas masa pengenduran membujur, T_1 bagi ^{13}C diperolehi daripada nilai τ_2 yang memberikan nilai tenaga pengaktifan, E_a bagi

empat vektor CH yang selari dengan paksi inertia dalam satah adalah 5.62 dan 7.32 kcal.mol^{-1} masing-masing pada suhu yang lebih tinggi dan lebih rendah. Manakala E_a bagi tiga vektor CH yang lain masing-masing bernilai 5.69 dan 7.17 kcal.mol^{-1}. Kajian ke atas sifat dinamik molekul bagi gerakan translasi, orientasi semula serta masa pengenduran membujur, T_1 memberikan dua plot Arrhenius yang bersilang pada suhu 320 hingga 325 K. Pemerhatian ini difahamkan sebagai tidak berkaitan dengan peralihan struktur tetapi lebih dipercayai berpunca daripada pergantungan semulajadi terhadap sifat dinamik sistem terhadap suhu.

Kata kunci: Dinamik molekul, isokuinolin, medan daya OPLS, pembauran translasi, pembauran putaran.

MOLECULAR DYNAMICS SIMULATION OF TRANSLATIONAL AND ROTATIONAL DIFFUSION OF LIQUID ISOQUINOLINE

ABSTRACT

Molecular dynamics (MD) simulations of isoquinoline in liquid phase were performed at several temperatures in the range of 300 to 365 K in order to investigate the evolution of translational and rotational diffusions with temperature. The intermolecular interactions are modeled using Coulombic plus Lennard-Jones potentials. The Lennard-Jones parameters are taken from the Optimized Potential of Liquid Simulation – All Atoms (OPLS-AA) force field. The computed vaporization enthalpies vary from 16.01 to 14.86 kcal.mol^{-1} from 300 to 365 K and are within 10.5-11.5 % of the experimental studies. Translational (D_{trans}) and rotational (D_{rot}) coefficients were computed using velocity (vacf) and angular velocity autocorrelation functions (avacf), respectively. The calculated D_{trans} from vacf was found to be larger, 0.0061 to 0.0472 Å^2.ps^{-1}, than the values calculated from mean square displacement (MSD), 0.0066 to 0.0500 Å^2.ps^{-1} with the larger difference observed at lower temperature. The rotational diffusion coefficients in the *x*, *y* and *z* axes computed as a function of temperature exhibited quite different behaviours. $D_{rot,z}/D_{rot,y}$ only changed from 2.23 to 2.32. Between 300 and 365 K a larger ratio was observed for $D_{rot,z}/D_{rot,x}$ and the value varies from 1.20 to 1.43. Reorientational correlation times, τ_1 and τ_2, of each C-H vector were obtained from fitting the correlation function to a biexponential function. The NMR longitudinal relaxation times T_1 of ^{13}C were determined from τ_2, giving the activation energy for higher and lower temperatures equal to 5.62 and 7.32 kcal.mol^{-1} respectively for the four CH

vectors parallel to one inplane-inertia axis. The E_a are 5.69 to 7.17 kcal.mol^{-1}, respectively for the three other CH vectors**.** The investigations of the dynamic properties of translational and reorientational motion as well as T_1 relaxation times yielded two Arrhenius plots which cross at about 320 to 325 K. This observation was understood to be unrelated to the structural transition, but linked with the natural temperature dependence of the dynamical properties of the systems.

Keywords: Molecular dynamics, liquid isoquinoline, OPLS-AA force field, rotational diffusion, translational diffusion.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Computer simulation including Monte Carlo (MC) and molecular dynamics (MD) simulation is a multidisciplinary field built on laws and theories stemming from mathematics, physics and chemistry. Nowadays, MD simulations have been widely used with the aim of understanding the properties of molecules assemblies in term of their structure and the microscopic interactions between them. To date, MD simulations have been applied in many areas of physics, chemistry, engineering and so on to model detailed microscopic dynamical behaviour of various type of systems including electrolytes (Balbuena *et al.*, 1996), polymer (Grest & Kremer, 1986; Bharadwaj *et al.*, 2000; Doi, 2003), biological molecules such as proteins (Elber & Karplus, 1987; Brooks III *et al.*, 2004; Karplus & Kuriyan, 2005), DNA (Cheatham *et al.*, 1995; Cheatham *et al.*, 1998; MacKerell Jr & Banavali, 2000) and polysaccharides (Brisson *et al.*, 1997; Okamoto & Yashima, 1998) as well as membranes (Van der Ploeg & Berendsen, 1983; Chiu *et al.*, 1995), liquid crystals (Egberts & Berendsen, 1988; Cross & Fung, 1994) and zeolites (Demontis *et al.*, 1989; Demontis & Suffritti, 1997). Processes such as melting (Honeycutt & Andersen, 1987; Gay *et al.*, 2002), adsorption and segregation (Glosli & Philpott, 1992; Xia & Landman, 1993; Groß & Scheffler, 1998), formation of molecular complexes (Behr & Lehn, 1976; Winkler *et al.*, 2002), protein denaturation and stability (Ibragimova & Wade, 1998), enzyme reactivity (Daniel *et al.*, 1998; Yon *et al.*, 1998) and membrane permeability (Nikaido, 2003) can be investigated.

By means of computer simulations the study of equilibrium and transport properties of model atomic liquids such as the hard sphere fluid (Alder *et al.*, 1970; Erpenbeck, 1984), the Lennard-Jones liquid (Nijmeijer *et al.*, 1988; Mecke *et al.*, 1997) and mixture models (Schoen & Hoheisel, 1984; Ferrario *et al.*, 1990; Vishnyakov *et al.*, 2001) have been performed. A detailed knowledge of molecular dynamics in liquids is important for the understanding of the microscopic mechanisms of the chemical reaction in liquid phase.

In addition, MD simulation as well as MC simulations can also be employed to investigate the structural properties by analyzing the radial distribution functions or molecular distribution functions of the system (Hansen & McDonald, 2006). Moreover, MD simulations are used to obtain the information on the liquid dynamics by extracting time correlation functions of the atoms and molecules (Hansen & McDonald, 2006). Analysis of the velocity autocorrelation functions and the power spectra obtained by the Fourier transformation of the velocity autocorrelation functions can be used as a tool for the elucidation of the microscopic molecular motions which can be related to the experimental and theoretical works of liquid dynamics. Moreover, by utilizing MD simulation, various cross correlation functions can be obtained between the initial movement of a central molecule and the motions of the surrounding molecules from which, knowledge of collective dynamic properties of the molecular system can be obtained.

Liquids exhibit many fascinating behaviours including structure and association, thermodynamics and molecular motion. Liquids may be studied in bulk phase or at liquid-liquid, liquid-solid and liquid-air interfaces. They have the ability

to flow even in viscous condition (Barton, 1974). Liquids play a major role in our daily life such as in the chemical activity occurring in our body which keeps the system to function properly. Such important features have attracted many researchers from both experimental and theoretical sides to explore and improve the understanding of the structure and dynamics of liquids. The study of the interactions in liquids is complicated by the diversity of supramolecular structures and the strength of the interactions. Therefore, the progress towards the understanding of the behaviour of liquids has been much slower than the understanding of gas and crystal phases. The real advances in the investigation of liquids, regarding liquid state as an intermediate state between solid and gas states have started about seventy years ago (Hansen, 1976).

The most obvious difference between a solid or a liquid and a gas is that a solid or a liquid maintains its shape provided that it is left undisturbed (Trevena, 1975). Unlike solids, liquids and gases must be contained in a vessel and the liquid occupies a given volume separated from the vapour by an interface. In contrast to liquids and solids which posses the property of cohesion (Pryde, 1967), a gas does not exhibit a boundary surface but will always occupy the entire volume available in a closed container. Solids have the ability to keep its shape which can support shear stress if it is not too large whereas fluids cannot support shear stresses no matter how small they are but can flow in response to them (Pryde, 1967). However, the cohesive forces in liquids are not strong enough to prevent translational motions of individual molecules. Liquids can experience a hydrostatic pressure or tension which leads to a situation that at any point in a stationary liquid, the pressure is the same in all directions (Rice & Gray, 1965).

The three states of matter experience phase changes from one to another depending on the temperature and pressure. At sufficiently low temperatures and high pressures all substances exist in solid state whereas as gas at sufficiently high temperatures and low pressures. A stable liquid state is formed at intermediate temperatures and pressures.

1.2 Problem statements

This work is conducted in order to investigate the structural, thermodynamic, and dynamic properties of liquid systems composed of rigid molecule model, isoquinoline. The preliminary studies of this work have been performed on the simulation of liquid benzene and hexafluorobenzene using Smith-Jaffe and modified Williams potentials respectively (Ahmad *et al.*, 2010). It has been observed experimentally by NMR spectroscopy (Jalabert *et al.*, 1991; Jalabert *et al.*, 1993; Robert *et al.*, 1997; Gauthier *et al.*, 2000) that liquid quinoline and liquid isoquinoline have a different temperature evolution for several dynamic properties such as translational and rotational diffusions as well as reorientational correlation times. From the previous experimental studies, isoquinoline presents a perfect Arrhenius behaviour in the liquid range between 300 and 365 K whereas quinoline exhibits two different Arrhenius behaviours between 276 and 320 K, with a change occurring around a transition temperature, T_t = 290 – 293 K (Jalabert *et al.*, 1991; Jalabert *et al.*, 1993; Robert *et al.*, 1997; Gauthier *et al.*, 2000). The experimentalists proposed that the anomaly in the Arrhenius plot was due to a structural transition within the normal liquid phase. This work was done to investigate whether this behaviour correspond to a liquid-liquid structural transition.

1.3 Objectives

The objectives of this work were to investigate the physical properties such as T_1 NMR relaxation times as a function of temperature as well as the translational and rotational diffusion coefficients of liquid isoquinoline using MD simulations. In addition, this work was performed in order to reproduce the experimental trends and compare the results between simulation and experimental works as well as to compare with the simulated liquid quinoline and other related planar molecules such as benzene and pyridine.

1.4 Thesis outline

This thesis is arranged as follows: in Chapter 2 we give a brief introduction on modeling of liquids with general concepts of computer simulations. Chapter 3 covers the literature review on liquids with some examples from experimental and theoretical works. In Chapter 4, we describe the methods and details of MD simulation of liquid isoquinoline. In Chapter 5, we discuss the simulation results in terms of its thermodynamic, dynamical and structural properties. Finally, Chapter 6 contains our conclusions and suggestions for future works.

CHAPTER 2

MODELLING OF LIQUIDS

2.1 Intermolecular potential

Intermolecular potential describes the interaction between molecules that demonstrates the properties of a liquid. A suitable potential function is required to evaluate the interaction between particles. A force field is a combination of parameters and analytical functions which gives complete description of the potential energy of an atomic (molecular) system as a function of its configuration. The general form of a force field consists of bonded and nonbonded interactions which are used to describe the potential energy as given by:

$$E_{total} = E_{bonded} + E_{nonbonded} \tag{2.1}$$

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \tag{2.2}$$

$$E_{nonbonded} = E_{electrostatic} + E_{vdW} \tag{2.3}$$

From equation (2.2), the E_{bonded} term describes the variations of energy associated to the variation of: (i) bond lengths A-B, (E_{bond}), (ii) the bond angles A-B-C, (E_{angle}) and (iii) torsional angles, involving a sequence of 4 atoms A-B-C-D, ($E_{dihedral}$). Other basic forms can be added in E_{bonded}, such as the out of plane bending terms (Leach, 2001). The $E_{nonbonded}$ term in equation (2.3) accounts for the interactions between atoms separated by 3 or more covalent bonds and the intermolecular interactions.

The most important of these interactions are the electrostatic and the van der Waals interactions.

A force field is generally based on a pairwise sum over all the different pairs present in the system. The electrostatic interactions ($E_{electrostatic}$) are usually described by distributions of partial electric charges treated as point charges on atoms (or arbitrary) which interact by Coulomb's law. The interaction between pairs of atoms is usually described by Coulomb potential (Fowler & Buckingham, 1983).

$$E_{electrostatic} = \sum_{I} \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J} \frac{q_i q_j}{r_{ij}} \times \frac{1}{4\pi\varepsilon_0} \qquad (2.4)$$

Where q_i and q_j are the electric charges of atoms i and j in molecule I and J respectively, while r_{ij} is the distance between atoms i and j. In some force fields such as MM2 (Allinger, 1977) and MM3 (Allinger *et al.*, 1989; Lii & Allinger, 1989), point dipoles on atoms or bonds can also be used.

The van der Waals interaction is usually modeled by a Lennard-Jones potential. The Lennard-Jones potential can be described using the equation (2.5):

$$E_{vdW} = \sum_{I} \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad (2.5)$$

where ε_{ij} is the well depth and σ_{ij} is the sum of site radii for the pair (i,j) while r_{ij} is the interatomic distance between atom i and j (Leach, 2001). These parameters can be obtained from fitting experimental data or quantum calculations. The r^{-12} term corresponds to the repulsive energy which arises at short distance. The r^{-6} term dominates at large distance, it is attractive and corresponds to the dispersion energy. It arises from the interactions between fluctuation of dipoles on both atom i and j; this interaction has a quantum origin and no classical counterpart.

The Buckingham potential, known as exp-6 potential (Buckingham *et al.*, 1988) is an alternative to the Lennard-Jones potential to describe van der Waals interactions. The Buckingham potential is written as:

$$E_{vdW} = \sum_{I} \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J} A_{ij} \exp\left(-B_{ij} r_{ij}\right) - \frac{C_{ij}}{r_{ij}^{6}} \qquad (2.6)$$

where, A_{ij} and B_{ij} are the parameters for the repulsive interactions, C_{ij} is the attractive (dispersion) term and r_{ij} is the interatomic distance between atom i and j.

Force fields have been determined for different classes of molecules including small organic molecules (Allinger *et al.*, 1989; Cornell *et al.*, 1995; Jorgensen *et al.*, 1996), biopolymers (Cornell *et al.*, 1995), inorganic molecules (Mayo *et al.*, 1990) and biological systems (Cornell *et al.*, 1995; Karplus & McCammon, 2002). For organic molecules with a variety of functional groups, the MM2, MM3 and MM4 are widely used force fields in molecular mechanics studies.

These force fields were developed by the Allinger group (Allinger, 1977; Allinger *et al.*, 1989; Lii & Allinger, 1989; Allinger *et al.*, 1996). They have been widely used to study the conformation of organic molecules for over 30 years. Another well known force field is Assisted Model Building with Energy Refinement (AMBER) which is used to study biomolecules (Weiner & Kollman, 1981). This force field was developed by Peter Kollman's group of University of California, San Francisco. The parameters used in this force field were chosen to reproduce the experimental structure and conformational energy differences for individual molecule.

Similar to AMBER another force field which is also often used is the all atoms Optimized Potential for Liquid Simulations (OPLS-AA) developed by Jorgensen group (Jorgensen *et al.*, 1984; Jorgensen *et al.*, 1996) at Yale University. OPLS-AA force field was optimized to fit experimental properties of liquids such as heat of vaporization and density. The calculation of potential energy forces is the crucial part in the simulation which determines the quality of MD and MC simulations. Therefore, the parameters for the atomic properties and their interactions in a system have to be defined properly in order to obtain a correct description of the system.

The potential energy functions and parameters are typically derived from experimental observations and quantum calculations. They have to be able to reproduce physical properties measured by experimental techniques including structural data obtained from x-ray crystallography and NMR, the dynamic data obtained from spectroscopy, inelastic neutron scattering and thermodynamic data. The parameters for bonded interactions are obtained from experimental data such as

gas phase geometry from IR spectroscopy as well as ab initio calculations. For the nonbonded interactions, the parameters can be obtained from various sources such as density, dipole moments and enthalpy of vaporization.

2.2 Molecular Mechanics

Molecular mechanics uses the application of classical mechanics to model the geometry and motions of molecules. In this method, only nuclei are taken into calculations. All the other electrons and orbitals are ignored. Using molecular mechanics, a large number of molecules up to thousands of atoms can be computed. Moreover, this method covers vacuum as well as implicit or explicit solvent environment.

In the field of molecular modeling, molecular mechanics is the least expensive (fastest) method (Jensen, 1999; Leach, 2001). Molecular mechanics method is often used as a low level optimization strategy before the optimization with ab-initio method is performed (Cramer & Bickelhaupt, 2002). This is done to ensure that the molecule is in its lowest energy conformation. A good force field is required for the molecule under study in order to obtain a structure which is close to the experimental structure. However, the drawback of this method is the lack of available parameters for many types of compounds and generally the temperature effects are not taken into account (Jensen, 1999). Moreover, since the electrons and orbitals are not explicitly taken into account in molecular mechanics, the study of chemical reactions as well as prediction of the reactivity of molecules cannot be performed.

2.3 Computer simulation methods: General concepts

Computer simulation is known as a branch of physics and chemistry which uses approximations to provide solutions for problems in statistical mechanics, physical chemistry and biophysics. Computer simulation is a test of theories where the results can be compared with real experimental data and also a test of intermolecular and intramolecular force fields. More generally, it is a test of the physical description of the system by a chosen model (classical force field or quantum modeling). Due to the advent of computing power and improvements in programming language, the use of computer simulation has increased in areas ranging from material sciences and chemistry to pharmaceutical science and molecular biology.

A programming language is used in computer simulation to build a program with certain algorithms to sample the configurations and motions of a system at a given thermodynamic conditions for example at given temperature and pressure or a given volume and energy. Moreover, these algorithms are able to mimic the configurations and motions in real system under given conditions. For example, using programming language, a model is set up in a hypothetical box of arbitrary shape (cube, parallelpiped, truncated octahedron). With a limited number of molecules in the box, typically from several hundreds to several hundred thousands the application of periodic boundary conditions (PBC) allows us to mimic the properties of macroscopic system by suppressing largely the surface effects. Programming language is useful in computer simulation in order to deal with the numerical iteration problems such as computing the interaction energies of the system. There are several available programming language that are often used in the

development of molecular system such as FORTRAN (FORmula TRANslation), C, C++, Visual basic and Java.

The application of computer simulation generates detailed insights into the structure, dynamics and the function of (bio) molecular systems (Beveridge & DiCapua, 1989; Orozco & Luque, 2000; Voth, 2006; Swanson *et al.*, 2007) as well as the microscopic interactions between them.

The essence of computer simulation is to complement the experimental work in order to provide a better understanding of the relation between microscopic properties and macroscopic behaviour (van Gunsteren & Berendsen, 1990). Moreover, computer simulation can be used to monitor the quantities that are inaccessible to experiment such as studies at extreme temperature and pressure (Kitchen *et al.*, 1992; Monaco *et al.*, 2003; Yan-Ning *et al.*, 2008) and determination of critical temperature and pressure (Rovere, 1993; Okumura & Yonezawa, 2002) for molecular systems which do not have critical point because they suffer a chemical transformation below the critical conditions. As a result, one may gain new insights into mechanisms and processes of the desired molecular systems. Moreover, testing a molecule using computer simulation is faster, safer and less expensive than synthesizing and characterizing it in a real experiment.

There are two main molecular simulation methods which are used to study the properties of liquids: (i) Monte Carlo (MC) simulation and (ii) Molecular Dynamics (MD) simulation. Both of these methods show similarity in the sense that they are both applied to small finite systems and also use the periodic boundary

conditions (Allen & Tildesley, 1990) in order to overcome the surface effect problem. The difference between the two methods is, Monte Carlo simulations sample only the configurations whereas Molecular Dynamics simulations sample the phase space (particle positions and velocities) of the systems.

2.3.1 Ensembles

Imposing some constraints to a system like constant number of particles N, constant volume V, constant total energy E, and constant temperature, T, leads to restriction to the accessible phase space (Allen & Tildesley, 1990). The set of phase space points (positions and velocities of the particles) accessible within a given constraints is called a statistical ensemble. There are different statistical ensembles introduced for MC and MD simulations. In standard MD method, the number of particles N is fixed as well as the volume V of the box and the total energy is conserved. Thus, the ensemble generated by this kind of simulation is the microcanonical or NVE ensemble. The collection of all phase space points corresponding to a thermodynamic state characterized by a fixed number of particles N, a fixed volume V as well as temperature T is called the canonical or NVT ensemble where the system is coupled with a thermostat (Nosé, 1984). Other frequently used ensemble in molecular simulations is the isobaric-isothermal or NPT ensemble where the number of particles N, pressure P and temperature T are fixed (Evans & Morriss, 1983). An ensemble where the chemical potential, μ, is fixed along with volume V and temperature T, is called grand canonical or the μVT ensemble. As in this ensemble, the number of particles N fluctuates, therefore it is more suitable to use it in MC simulations than in MD simulations.

2.3.2 Periodic Boundary Condition (PBC)

In MC and MD simulations, the N particles (atoms, molecules, or ions) are confined in a container. This situation is not satisfactory for simulation of bulk liquids, due to surface effects. Once a simulation is performed, there is always a large fraction of molecules close to the surface of the box containing these particles. Molecules close to the surface and molecules in the bulk have different properties and behaviours. To overcome the drawback of surface effects, periodic boundary conditions are commonly applied (Allen & Tildesley, 1990; Leach, 2001). Figure 2.1 shows the diagram of periodic boundary conditions. This enables a simulation to be performed using any number of particles, in such a way that the particles experience forces as if they were in bulk liquids.

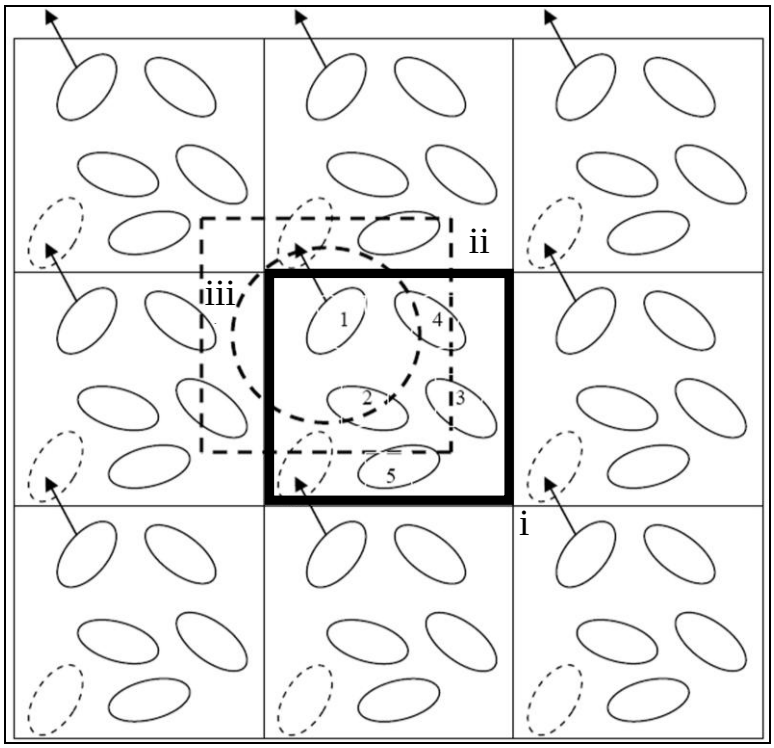

Figure 2.1: Diagram showing periodic boundary conditions in two dimensions with a square box, minimum image convention and cutoff. i) Central box N molecules. ii) Box containing N molecules, minimum image convention centre of position of molecule 1. iii) Same as 2 with a spherical cutoff $R_c = a/2$, a being the edge of square box.

The original simulation box is replicated throughout the space to form an infinite lattice. During the simulation, when a molecule in the original box moves, its periodic images in each of the surrounding boxes move in exactly the same way. Thus, as a molecule leaves the central box, one of its images will enter through the opposite face. The boundary of the central box contains no walls and the system has no surface. The coordinates of the N molecules are measured by the axis system formed by the box. As a result, the number of molecules in the central box as well as in the entire system is conserved. It is not necessary to store the coordinates of all the images in a simulation because they are related by translation to the position of the molecules in the central box.

The crucial part of molecular dynamics simulations is calculating the potential energy of a system and forces acting on each molecule. The calculation of potential energy does not only involve the interactions between molecule 1 and all the other molecules in the simulation box but also the interactions with all the images in the image boxes. In practice, it is impossible to calculate an infinite number of interactions. In order to deal with this kind of problem, the minimum image convention is applied where molecule 1 interacts with N-1 molecules contained in a box centered on molecule 1, belonging to the central box or the first image boxes. If the force is short-ranged, one can only consider the molecules close to molecule 1 while other interactions with far distant molecules can be ignored.

This situation can be significantly improved using another approximation called potential truncation. For a short-range potential, the main contribution to energy and force comes from the nearest particles. Therefore, the potential can be

truncated at certain distance called cutoff distance. When a cutoff distance is employed, the interactions between pairs of atom that are further apart than the cutoff value are set to zero taking into account the closest image. The cutoff distance must not be larger than half of the box length when simulating atomic liquids as it is to keep it consistent with the minimum image convention. Further explanation on PBC can be found in the literature elsewhere (Allen & Tildesley, 1990; Leach, 2001).

2.3.3 Long range interactions

Long range interactions are defined as the interaction with the decay not faster than r^{-d} where d is the dimensionality of the system (Fowler & Buckingham, 1983; Leach, 2001). The examples of long range interactions are the electrostatic charge-charge, charge-dipole, and dipole-dipole interactions which as decays of r^{-1}, r^{-2} and r^{-3}, respectively. The long range interactions have to be considered in order to obtain a qualitative and quantitative description of the dielectric properties of liquids. However, these interactions can be a problem as their range is often larger than half of the box length. Although this problem can be solved by introducing a larger simulation box which would diminish the cutoff artifact however this is not practical because more computing time is required to run such a simulation. Moreover, dielectric properties are also sensitive to the boundary condition. Therefore, appropriate treatments of long range interactions in molecular simulations of condensed phase have been introduced to overcome this problem. There are two most common methods to treat the long range interactions: the Ewald summation technique and the reaction field method. More detailed discussion of both methods can be found in Allen and Tildesley (1990), Leach (2001), etc.

(i) Ewald summation

The Ewald summation is a technique to compute the interaction between charged particles in the cube of edge L and between these particles and all their images in an infinite image boxes. For N charges q_i, the potential energy can be written as

$$U = \frac{1}{2}\sum_{\mathbf{n}}{}' \left(\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \left|\mathbf{r}_{ij} + \mathbf{n}\right|^{-1} \right) \qquad (2.7)$$

where q_i, q_j, are the charges, **n** is the vector with integer components, with **n** = (n_x,L, $n_y L$, $n_z L$) which specifies the box position. The prime indicates that in the central box (**n** = 0) the i = j is omitted.

The Ewald summation introduces the idea of using charge distribution at each point charge location and to replace lattice sum by two rapidly converging sums. The electrostatic interactions are evaluated through this infinite lattice sum. In this method, each point charge is surrounded by a Gaussian charge distribution of opposite sign but equal in magnitude. The Gaussian distribution, $\rho_i^q(\mathbf{r})$ is taken to be (Allen & Tildesley, 1990):

$$\rho_i^q(\mathbf{r}) = \frac{q_i \kappa^3}{\pi^{3/2}} \exp\left(-\kappa^2 \mathbf{r}^2\right) \qquad (2.8)$$

where the arbitrary parameter κ determines the width of the distribution and **r** is the position relative to the centre of the distribution. The charge distribution will partly screen the point charge and the interaction become short ranged. Then, a canceling

charge distribution of the same sign as the original charge is added to obtain the overall potential with the original set of charges. This canceling distribution is summed in reciprocal space. Figure 2.2 shows the charge distribution in Ewald summation technique taken from Allen and Tildesley (1990).

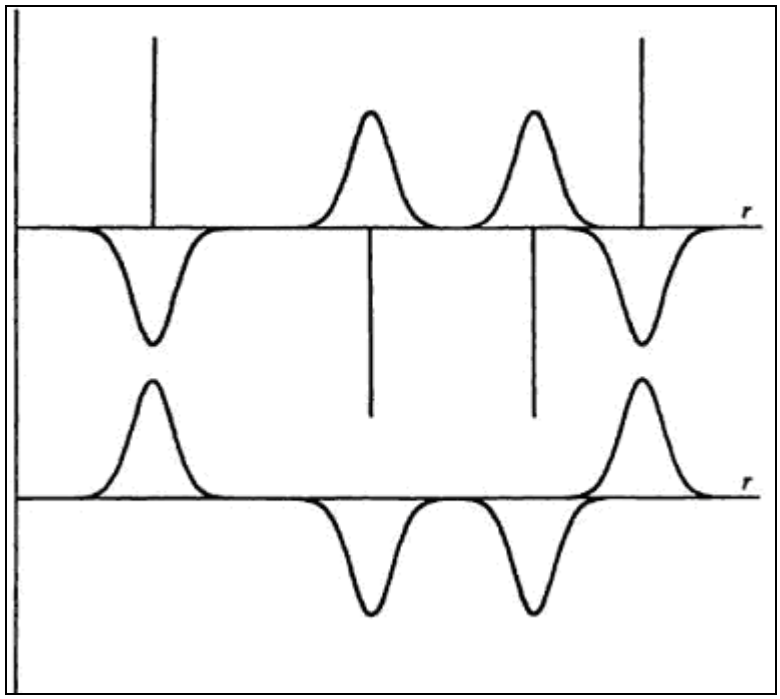

Figure 2.2: Charge distribution in the Ewald summation. (a) Original point charges plus screening distribution. (b) Canceling distribution (Allen & Tildesley, 1990).

The final potential energy of Ewald summation will contain a real space sum plus a reciprocal space sum with the subtraction of the interactions of the canceling distribution centered at $\mathbf{r}_i$ with itself (self-term) plus the surface term (corresponding to the $\mathbf{k} = 0$ term in the reciprocal space summation) giving the final result, of

$$U_{qq}(\varepsilon_s = 1) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left({\sum_{|\mathbf{n}|=0}^{\infty}}' q_i q_j \frac{\operatorname{erfc}\left(\kappa|\mathbf{r}_{ij}+\mathbf{n}|\right)}{|\mathbf{r}_{ij}+\mathbf{n}|} + \left(1/\pi L^3\right)\sum_{\mathbf{k}\neq\mathbf{0}} q_i q_j \left(4\pi^2/k^2\right)\exp\left(-k^2/4\kappa^2\right)\cos\left(\mathbf{k}.\mathbf{r}_{ij}\right)\right) - \left(\kappa/\pi^{1/2}\right)\sum_{i=1}^{N} q_i^2 + \left(2\pi/3L^3\right)\left|\sum_{i=1}^{N} q_i \mathbf{r}_i\right|^2 \qquad (2.9)$$

where *erfc(x)* is the complementary error function with $erfc(x) = \left(2/\pi^{1/2}\right)\times\int_x^{\infty}\exp\left(-t^2\right)dt$ which falls to zero with increasing *x*. The second term is a sum over reciprocal vectors $\mathbf{k} = 2\pi\mathbf{n}/L^2$. Equation (2.9) applies to systems with a boundary condition such that the system is surrounded by a dielectric continuum with a static dielectric constant $\varepsilon_s = 1$ (vacuum). If the dielectric constant is ε_s, a further boundary term must be added (reaction field term):

$$-\frac{2\pi}{3L^3}\left(\frac{2(\varepsilon_s - 1)}{2\varepsilon_s + 1}\right)\left|\sum_i q_i \mathbf{r}_i\right|^2 \qquad (2.10)$$

It is interesting to note that if $\varepsilon_s = \infty$ (conducting boundary condition), this term cancels the last term on the right side of equation (2.9). For molecular systems with $N: 10^3$, this method increases the computer time by a factor ~ 2.5 with respect to the simulation with a spherical cutoff. This technique is usually used to study larger molecular systems such as proteins and DNAs and is more suitable for the simulations of polar molecules and ionic crystals.

(ii) Reaction field method

Reaction field method is another way to deal with the long range intermolecular interactions. In this method, a sphere is constructed around the molecule with a radius equals to the cutoff radius, r_c. When the spherical cutoff is applied, all interactions are truncated at the cutoff radius, r_c. In this method, the interactions beyond r_c are approximated by treating the system as a polarizable continuum of dielectric constant, ε_{RF} (Barker & Watts, 1973). Each molecule of the liquid experiences an electric field arising from the net dipole moment within the cut off sphere (cavity) centered on that molecule. The reaction field acting on molecule *i* due to the surrounding dielectric is given by:

$$E_{i_{\mathrm{RF}}} = 2\left(\frac{\varepsilon_{RF}-1}{2\varepsilon_{RF}+1}\right)\left(\frac{1}{r_c^3}\right)\sum_{j;r_{ij}\le r_c}\boldsymbol{\mu}_j \qquad (2.11)$$

where $\boldsymbol{\mu}_j$ are the dipoles of the neighbouring molecules that are within the cutoff distance (r_c). The contribution to the energy due to the reaction field is given as:

$$U_{\mathrm{RF}} = -\frac{1}{2}\sum_{i=1}^{N} E_{i_{\mathrm{RF}}}\cdot\boldsymbol{\mu}_i \qquad (2.12)$$

This approach assumes that the liquid is structureless beyond r_c so that a continuum approximation can be applied. The reaction field method was first applied by Baker and Watts (1973) in a simulation of water.

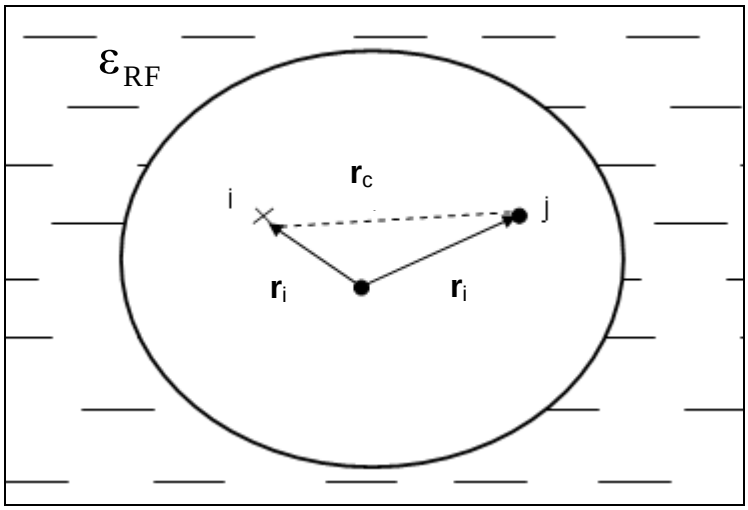

Figure 2.3: Schematic drawing of reaction field between charge *i* and *j*.

In reaction field method, the charge q_i creates a potential V_{ij} at position of charge q_j with

$$V_{ij} = \frac{q_i}{r_{ij}} + q_i \left[\sum_{l=0}^{\infty} \frac{(l+1)(1-\varepsilon_{\mathrm{RF}})}{(\varepsilon_{\mathrm{RF}}(l+1)+l)} \cdot \frac{(r_i r_j)^l}{r_c^{2l+1}} P_l(\cos\ \theta_{ij}) \right] \qquad (2.13)$$

where $\varepsilon_{\mathrm{RF}}$ is the dielectric constant, r_c is the radius surrounded by a dielectric continuum, P_l is the Legendre polynomial and $\cos\ \theta_{ij}$ is the angle between $\mathbf{r}_i$ and $\mathbf{r}_j$. The corresponding energy of N charges is given by:

$$U = \frac{1}{2} \sum_i \sum_j V_{ij} \cdot q_j$$

$$= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ q_i q_j \sum_{l=0}^{\infty} f_l (r_i r_j)^l P_l(\cos\theta_{ij}) \right\} \qquad (2.14)$$

with $f_l = \frac{(l+1)(\varepsilon_{RF}-1)}{\varepsilon_{RF}(l+1)+l} \cdot \frac{1}{r_c^{2l}+1}$

$$U = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left\{\frac{q_i q_j}{r_{ij}} - q_i q_j \sum_{l=0}^{\infty} f_l \left(r_i r_j\right)^l P_l\left(\cos\theta_{ij}\right)\right\} - \frac{1}{2}\sum_{i}^{\infty} q_i^{\,2}\left(\sum_{i}^{\infty} f_l \cdot \left(r_i\right)^{2l}\right) \quad (2.15)$$

On the right side of equation (2.15), the first term is a direct pair term. The second term is the interactions of charges q_i with the reaction potential created by charges $q_i \neq q_j$. The third term is self term, containing the interaction of the charge q_i with their own reaction potential. Through approximate (the discontinuity at r_c introduces substantial artifact), this method has been widely used to study the solvent effects in chemical and biological systems. The main advantage of this method is that it does not increase significantly the computer time requirement with respect to the simulation with a spherical cutoff.

2.4 Monte Carlo simulation

Monte Carlo (MC) simulation is a method introduced by Metropolis *et al.* (1953) where the initial coordinates are chosen arbitrarily for systems of N particles which interact through some known potential. The particles are the randomly displaced one at a time for successive configurations. The individual configuration should posses the rules that appear with probability proportional to the Boltzmann constant, $\exp(-U / k_B T)$ where U is the potential energy of the system, k_B is the Boltzmann's constant and T is the temperature. In that way, the method estimates canonical ensemble averages in which the number of particles (N), total volume (V), and temperature (T) are fixed parameters. MC method also has been adapted in the

framework of the isothermal-isobaric (NPT) (Wood, 1970; Hansen & McDonald, 2006) and grand canonical ensembles (Owicki & Scheraga, 1977; Im *et al.*, 2000). The flowchart of MC simulation is given in Figure 2.4.

In MC method, the thermodynamic properties are calculated using the methods of statistical mechanics (Barker & Henderson, 1976). MC method provides only the information on static properties alone. Moreover, in MC method, the movements of molecules are not associated with real time. The statistical accuracy depends on the length of the chain of configurations and the system size. However, a satisfactory accuracy can be attained with reasonable amounts of computer time (Leach, 2001). A famous simulation package using MC is the BOSS package (Jorgensen, 1999).

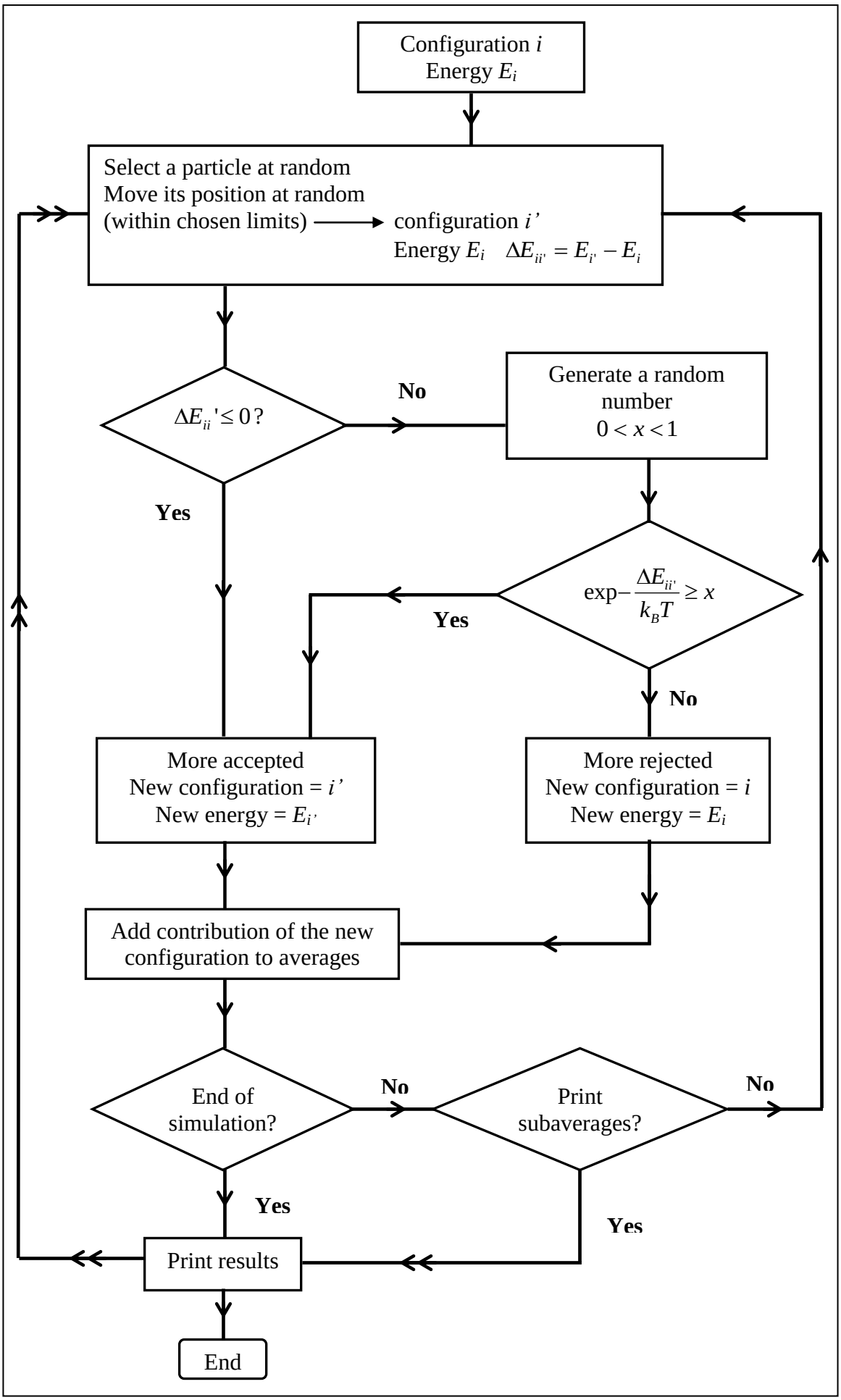

Figure 2.4: Flowchart of an MC simulation (Metropolis *et al.*, 1953).

2.5 Molecular dynamics simulation

In MD simulation, a system of N particles comprising of either atoms or molecules or ions is positioned within a box of fixed volume and a given shape. The initial position and velocities of each particle are assigned. The velocities are chosen with respect to a Boltzmann distribution. Given an initial velocities and positions of each atom, the velocities and position at later time may be calculated numerically by solving the Newton's equations of motions and the process is repeated in a sequence of discrete time steps. The flowchart of MD simulation is presented in Figure 2.5.

Apart from the choice of initial configurations (positions and velocities), an MD simulation is in principle entirely deterministic. During the simulation process, both the total energy and volume are naturally conserved, thus sampling the microcanonical ensemble with constant volume and energy (NVE). Basically, MD simulation relies on three main ideas; (i) the force field used to model the interaction between particles; (ii) the needs of an integrator which propagates the positions and velocities of atoms or molecules from initial to later time; (iii) an ensemble has to be chosen to control the thermodynamic quantities such as pressure (p), temperature (T) and number of particles (N) (van Gunsteren & Berendsen, 1990; van Gunsteren & Mark, 1998; Rapaport, 2004).

In contrast to MC simulation, thermodynamic properties as well as dynamical properties are calculated using time averages. In consequence, MD simulation allows the calculation of the static properties such as pressure, temperature, density as well as the transport properties (Barton, 1974). Moreover, MD simulation provides a