Applications of Molecular Dynamics in Atmospheric and Solution Chemistry

Xin Li

(李昕)

Doctoral Thesis in Theoretical Chemistry and Biology
School of Biotechnology
Royal Institute of Technology
Stockholm, Sweden 2011
To my parents

To my parents
Abstract

This thesis focuses on the applications of molecular dynamics simulation techniques in the fields of solution chemistry and atmospheric chemistry. The work behind the thesis takes account of the fast development of computer hardware, which has made computationally intensive simulations become more and more popular in disciplines like pharmacy, biology and materials science. In molecular dynamics simulations using classical force fields, the atoms are represented by mass points with partial charges and the inter-atomic interactions are modeled by approximate potential functions that produce satisfactory results at an economical computational cost. The three-dimensional trajectory of a many-body system is generated by integrating Newton’s equations of motion, and subsequent statistical analysis on the trajectories provides microscopic insight into the physical properties of the system.

The applications in this thesis of molecular dynamics simulations in solution chemistry comprise four aspects: the $^{113}$Cd nuclear magnetic resonance shielding constant of aqua Cd(II) ions, paramagnetic $^{19}$F nuclear magnetic resonance shift in fluorinated cysteine, solvation free energies and structures of metal ions, and protein adsorption onto TiO$_2$. In the studies of nuclear magnetic resonance parameters, the relativistic effect of the $^{113}$Cd nucleus and the paramagnetic shift of $^{19}$F induced by triplet O$_2$ are well reproduced by a combined molecular dynamics and density functional theory approach. The simulation of the aqua Cd(II) ion is also extended to several other monovalent, divalent and trivalent metal ions, where careful parameterization of the metal ions ensures the reproduction of experimental solvation structures and free energies. Molecular dynamics simulations also provided insight into the mechanism of protein adsorption onto the TiO$_2$ surface by suggesting that the interfacial water molecules play an important role of mediating the adsorption and that the hydroxylated TiO$_2$ surface has a large affinity to the proteins.

The applications of molecular dynamics simulations in atmospheric chemistry are mainly focused on two types of organic components in aerosol droplets: humic-like compounds and amino acids. The humic-like substances, including cis-pinonic acid, pinic acid and pinonaldehyde, are surface-active organic compounds that are able to depress the surface tension of water droplets, as revealed by both experimental measurements and theoretical computations. These compounds either concentrate on the droplet surface or aggregate inside the droplet. Their effects on the surface tension can be modeled by the Langmuir-Szyszkowski equation. The amino acids are not strong surfactants and their influence on the surface tension is much smaller. Simulations show that the zwitterionic forms of serine, glycine and alanine have hydrophilic characteristics, while those of valine, methionine and phenylalanine are hydrophobic. The curvature dependence of the surface tension is also analyzed, and a slight improvement in the Köhler equation is obtained by introducing surface tension corrections for droplets containing glycine and serine.
Through several examples it is shown that molecular dynamics simulations serve as a promising tool in the study of aqueous systems. Both solute-solvent interactions and interfaces can be treated properly by choosing suitable potential functions and parameters. Specifically, molecular dynamics simulations provide a microscopic picture that evolves with time, making it possible to follow the dynamic processes such as protein adsorption or atmospheric droplet formation. Moreover, molecular dynamics simulations treat a large number of molecules and give a statistical description of the system; therefore it is convenient to compare the simulated results with experimentally measured data. The simulations can provide hints for better design of experiments, while experimental data can be fed into the refinement of the simulation model. As an important complementary to experiments, molecular dynamics simulations will continue to play significant roles in the research fields of physics, chemistry, materials science, biology and medicine.
Preface

The work presented in this thesis has been conducted at the Department of Theoretical Chemistry and Biology, School of Biotechnology, Royal Institute of Technology, Stockholm, Sweden and Laboratory for Advanced Materials and Institute of Fine Chemicals, East China University of Science and Technology, Shanghai, P. R. China.

List of papers included in this thesis

**Paper I**  Nuclear Magnetic Shielding of the $^{113}$Cd(II) Ion in Aqua Solution: A Combined Molecular Dynamics/Density Functional Theory Study  
Li, X.; Rinkevicius, Z.; Tu, Y.; Tian, H.; Ågren, H.  

**Paper II**  Paramagnetic Perturbation of the $^{19}$F NMR Chemical Shift in Fluorinated Cysteine by O$_2$: A Theoretical Study  
Li, X.; Rinkevicius, Z.; Tu, Y.; Tian, H.; Ågren, H.  

**Paper III**  Computer Simulations of Aqua Metal Ions for Accurate Reproduction of Hydration Free Energies and Structures  
Li, X.; Tu, Y.; Tian, H.; Ågren, H.  

**Paper IV**  On the Mechanism of Protein Adsorption onto Hydroxylated and Nonhydroxylated TiO$_2$ Surfaces  
Kang, Y.; Li, X.; Tu, Y.; Wang, Q.; Ågren, H.  

**Paper V**  Surface-Active cis-Pinonic Acid in Atmospheric Droplets: A Molecular Dynamics Study  
Li, X.; Hede, T.; Tu, Y.; Leck, C.; Ågren, H.  

**Paper VI**  HULIS in Nanoaerosol Clusters; Investigations of Surface Tension and Aggregate Formation using Molecular Dynamics Simulations  
Hede, T.; Li, X.; Leck, C.; Tu, Y.; Ågren, H.  
Paper VII  Glycine in Aerosol Water Droplets: A Critical Assessment of Köhler Theory by Predicting Surface Tension from Molecular Dynamics Simulations  
**Li, X.;** Hede, T.; Tu, Y.; Leck, C.; Ågren, H.  

Paper VIII  Molecular Dynamics Study on the Surface Tension of Atmospheric Water Droplets Containing Amino Acids  
**Li, X.;** Hede, T.; Tu, Y.; Leck, C.; Ågren, H.  
*In manuscript.*

List of papers not included in this thesis

**Paper I**  Modulation of Iridium(III) Phosphorescence via Photochromic Ligands: A Density Functional Theory Study  
**Li, X.;** Zhang, Q.; Tu, Y.; Ågren, H.; Tian, H.  

**Paper II**  Theoretical Study of Phosphorescence of Iridium Complexes with Fluorine-Substituted Phenylpyridine Ligands  
**Li, X.;** Minaev, B.; Ågren, H.; Tian, H.  

Comments on my contributions to the included papers

I have taken the major responsibility in papers I, II, III, V, VII and VIII, where I am the first author. I as the co-author participated in the discussion and preparation of papers IV and VI.
Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Prof. Hans Ågren, for introducing me to the research field of theoretical chemistry and giving me the opportunity to work on various interesting projects. I have learned much from his profound knowledge and insightful thoughts on scientific research, and I am thankful for all his guidance, support, trust and encouragement, without which this thesis would have been impossible.

I am deeply indebted to my supervisor at East China University of Science and Technology, Prof. He Tian, for his recommendation and the scientific training that I have received from him. I have benefited a lot from his scientific attitude and rigorous style.

I am heartily thankful to my co-supervisor, Dr. Yaoquan Tu, for his careful guidance and considerable help. I am also thankful to Dr. Zilvinas Rinkevicius, Prof. Boris Minaev, Prof. Caroline Leck and Thomas Hede for the nice collaboration and valuable discussions on research projects. Furthermore, I am sincerely grateful to Prof. Yi Luo for his positive advice and helpful suggestions on my academic career. Many thanks to Prof. Faris Gel’mukhanov, Prof. Fahmi Himo, Dr. Paweł Salek, Dr. Ying Fu, Dr. Olav Vahtras, Staffan Hellström and Dr. N. Arul Murugan for their kind help in my laboratory life.

My special gratitude goes to my girlfriend, Yu Kang, for her love, understanding and support. I am also indebted to Dr. Qiong Zhang and Dr. Zhijun Ning, for their help and care in many aspects. I want to thank my colleagues and friends, Dr. Kai Liu, Dr. Jun Jiang, Dr. Bin Gao, Dr. Hui Cao, Dr. Tiantian Han, Dr. Guangde Tu, Dr. Ke Zhao, Xiaofei Li, Keyan Lian, Dr. Katia J. de Almeida, Dr. Hao Ren, Qiang Fu, Dr. Kai Fu, Dr. Xiaohao Zhou, Ying Hou, Hongmei Zhong, Weijie Hua, Fuming Ying, Sai Duan, Xing Chen, Jiayuan Qi, Chunze Yuan, Dr. Shilv Chen, Dr. Rongzhen Liao, Qiu Fang, Yuejie Ai, Dr. Feng Zhang, Guangjun Tian, Xiao Cheng, Bonaman Xin Li, Hongbao Li, Dr. Jicai Liu, Yuping Sun, LiLi Lin, Yong Ma, Xiuneng Song, Quan Miao, Dr. Xifeng Yang, Bogdan Frecus, Ying Wang, Xiangjun Shang, Zhihui Chen, Yongfei Ji, Dr. Junkuo Gao and Lu Sun, who have made my life in Sweden happy and fulfilled. I am so grateful to all of you for frequent academic discussions, leisure activities and all kinds of help in daily life.

The financial support from the Chinese Scholarship Council and the computational resources provided by the Swedish National Infrastructure for Computing are sincerely acknowledged.

Finally, I wish to express my deepest gratitude to my parents for their constant love and support.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMBER</td>
<td>assisted model building with energy refinement</td>
</tr>
<tr>
<td>CCN</td>
<td>cloud condensation nuclei</td>
</tr>
<tr>
<td>CHARMM</td>
<td>chemistry at Harvard macromolecular mechanics</td>
</tr>
<tr>
<td>CPA</td>
<td>cis-pinonic acid</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>GIAO</td>
<td>gauge including atomic orbital</td>
</tr>
<tr>
<td>HSA</td>
<td>human serum albumin</td>
</tr>
<tr>
<td>HULIS</td>
<td>humic-like substances</td>
</tr>
<tr>
<td>LINCS</td>
<td>linear constraint solver</td>
</tr>
<tr>
<td>LRC</td>
<td>long-range correction</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>MP2</td>
<td>second-order Møller-Plesset</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NPT</td>
<td>isothermal-isobaric (ensemble)</td>
</tr>
<tr>
<td>NVE</td>
<td>microcanonical (ensemble)</td>
</tr>
<tr>
<td>NVT</td>
<td>canonical (ensemble)</td>
</tr>
<tr>
<td>OPLS</td>
<td>optimized potentials for liquid simulations</td>
</tr>
<tr>
<td>PAD</td>
<td>pinic acid</td>
</tr>
<tr>
<td>PAL</td>
<td>pinonaldehyde</td>
</tr>
<tr>
<td>PME</td>
<td>particle mesh Ewald</td>
</tr>
<tr>
<td>QM/MM</td>
<td>quantum mechanical/molecular mechanical</td>
</tr>
<tr>
<td>RDF</td>
<td>radial distribution function</td>
</tr>
<tr>
<td>RESP</td>
<td>restrained electrostatic potential</td>
</tr>
</tbody>
</table>
Contents

Chapter 1  Introduction ........................................................................................................1

Chapter 2  Theoretical Background ..................................................................................4

  2.1 Molecular Dynamics Algorithms .............................................................................4
    2.1.1 Equations of Motion and Integrator ..................................................................4
    2.1.2 Periodic Boundary Conditions ..........................................................................6
    2.1.3 Cutoff and Neighbor List ..................................................................................6
    2.1.4 Constraint Algorithms .....................................................................................7
    2.1.5 Thermostats and Barostats ..............................................................................8

  2.2 Inter-atomic Interactions and Force Fields ............................................................10
    2.2.1 Bonded Interaction .......................................................................................11
    2.2.2 Non-Bonded Interaction ...............................................................................12
    2.2.3 Polarization ..................................................................................................14
    2.2.4 Several Force Fields and Water Models .........................................................15

Chapter 3  Applications in Solution Chemistry ............................................................17

  3.1 Cd-113 NMR Shielding Constant of Aqua Cd(II) Ion (Paper I) ............................19
  3.2 Paramagnetic F-19 NMR Shift in Fluorinated Cysteine (Paper II) .......................23
  3.3 Solvation Free Energies and Structures of Metal Ions (Paper III) .......................26
  3.4 Protein Adsorption onto Titanium Dioxide Surface (Paper IV) .........................29
Chapter 4 Applications in Atmospheric Chemistry ..................................................32

4.1 cis-Pinonic Acid in Atmospheric Droplets (Paper V) .............................................34

4.2 Humic-Like Substances in Atmospheric Droplets (Paper VI) .................................37

4.3 Glycine in Atmospheric Droplets (Paper VII) ..........................................................39

4.4 Amino Acids in Atmospheric Droplets (Paper VIII) ...............................................42

Chapter 5 Summary and Outlook ..............................................................................45
Chapter 1  Introduction

*Travelers, there is no path, paths are made by walking.*

— Antonio Machado

The molecular dynamics (MD) technique employs computers to carry out virtual experiments of many-body systems by solving Newton’s equations of motion. By recording and analyzing the varying coordinates, velocities and forces of the particles in the system, the dynamical processes and statistical properties can be tracked and studied in detail. Such computer experiments provide the convenience of performing tests under unusual conditions or investigating non-existing materials, and the development of more accurate and realistic interaction potential functions ensures the reliability and feasibility of the simulations. So far, MD simulations have been widely used in theoretical studies of materials science, biophysics and biochemistry.

The history of MD simulations can be traced back to the middle of the 20th century when computers became available for non-military use. By then, the analytical theories of ideal gas and harmonic crystals had been established; however, the theory for liquids remained a difficult issue [1]. It is known that a many-body system is a chaotic system; a small perturbation would result in a large deviation in the motions of the particles, leading to difficulties in deriving analytical theories for liquids. Computers with the capacity of heavy computations were then utilized to solve the problem numerically. The very first computer simulation of a liquid was carried out by Metropolis et al. [2], using a Monte Carlo method. Later the first MD simulation was reported for a hard-sphere system [3], and then the technique was employed to study real systems like crystalline copper [4] and liquid argon [5]. Since that time, the foundation of MD simulations was laid step by step [6] [7] [8], and the applications were extended to polyatomic molecules such as water [9], hydrocarbon [10], and even proteins [11]. MD simulations originated from the difficult issue of liquid modeling and experienced a fast development with the help of the computers.

The concept of MD simulations can be applied on different scales. From chemical intuition and classical point of view, it is straightforward to use the so-called all-atom model where each atom is represented by a particle with certain mass. For computational efficiency, the inter- and intra-molecular interaction potentials were modeled by simple and approximate functions, such as the Lennard-Jones potential for van der Waals interaction, the point-charge Coulomb potential for electrostatic interaction, the harmonic potential for chemical bonds and angles, and the Fourier series for torsion angles. Fortunately, such approximations have proved useful and efficient, especially for MD simulations of macromolecules. To take a
step towards a higher accuracy in electrostatic treatment, the point-charge model can be improved to explicitly take into account polarization effects, for instance by introducing the Drude particle \cite{12}, an oscillator with partial electric charge to model the polarizability, or by using the fluctuating charge method, where the partial charges on the atoms are no longer fixed. Towards a larger scale of simulation time, system size and configuration space, coarse-grained models can be used, where several atoms are represented by a single bead and the atomic details are dealt with in an averaged manner. Still, all these methods are based on a classical description of the interaction potential. Pioneered by the works of Warshel et al. \cite{13} in 1976 and Car et al. \cite{14} in 1985, more reliable and accurate quantum chemical approaches became available for MD simulations, providing the possibility of investigating electron transitions and chemical reactions which are not accessible in classical models. Today, MD simulations have become a general simulation technique that covers a wide scope of time and scale, ranging from femtoseconds to milliseconds, from atoms to macromolecular aggregates. In this thesis, the applications of MD simulations in the fields of solution chemistry and atmospheric chemistry will be in focus.

The application in solution chemistry has a long history since MD simulations were born for liquid modeling. On one hand, water has become one of the most attractive research topics since it is a most important and interesting liquid solvent and since it constitutes the majority of aqueous solutions. Plenty of water models and potentials have been proposed to simulate various properties of water and to provide microscopic pictures of the structures and dynamics of water molecules. On the other hand, solute molecules are the essential components that make the solution different from pure water. Metal ions, amino acids and proteins are important types of solutes in life science and also in the study of solution chemistry. We here focus on the following topics: the relativistic effect of Cd(II) ion on $^{113}$Cd NMR (nuclear magnetic resonance), the paramagnetic effect of triplet O$_2$ on $^{19}$F NMR, the polarization effects of highly charged metal ions on the structure of surrounding water molecules, and the effects of surface modification of TiO$_2$ on protein adsorption.

The application in atmospheric chemistry is relatively new; the basic concept is to study the atmospheric aerosol droplets from a microscopic point of view for better understanding of the nucleation and growth of these droplets, the properties of which will affect the cloud reflectivity and absorption, and hence the climate. Various types of solutes exist in atmospheric droplets including both inorganic and organic components; the effect of the organic compounds is less well understood than for their inorganic counterparts. Therefore, we employed MD simulations to study atmospheric organic components such as humic-like substances originating from oxidation processes of organic vapor from boreal trees and amino acids from protein decomposition, inserted to the troposphere by bubble bursting processes at the ocean surface. Droplets as inhomogeneous systems with liquid-gas interfaces can be straightforwardly treated by MD simulations with proper interaction potential functions and parameters. Modern computers ensure the simulations of large droplets consisting of
thousands of water molecules, and the surface tension is calculated to facilitate better understanding of the role of the organic components.

The thesis is organized in the following way. First a brief introduction of the theoretical background of MD simulations is given in Chapter 2, followed by discussions on the applications in solution chemistry and atmospheric chemistry in Chapter 3 and Chapter 4, respectively. Finally, the summary of the studies and the outlook of the future are presented in Chapter 5.
Chapter 2  Theoretical Background

The MD simulation technique is essentially based on Newton’s second law, which reveals that the acceleration of a particle is equal to the force acting on it divided by its mass. As time evolves, the positions and velocities of the particles are updated by integrating Newton’s equations of motion over short time steps. The interacting forces between the particles, which determine the accelerations, are modeled either quantum chemically or by the so-called force field which comprises a set of potential functions and corresponding parameters. The latter is much faster and more efficient in simulating large systems.

2.1 Molecular Dynamics Algorithms

A series of algorithms control the computational procedure of MD simulations, including integration, periodic boundary conditions, neighbor searching, constraints, thermostats, barostats, etc. The integrator defines how the equations of motion are solved numerically. Periodic boundary conditions are applied so that a microscopic simulation box can reasonably model an infinite system by replicating the box in three dimensions. The neighbor searching algorithm is used to quickly and efficiently find the pairs of atoms within a predefined cutoff radius. Constraints are employed to maintain certain bonds at constant lengths. Thermostats and barostats are introduced in MD simulations to keep the system under desired temperature and pressure in order to mimic experimental conditions.

2.1.1 Equations of Motion and Integrator

In a system of \( N \) interacting particles with coordinates \( \mathbf{r}_i, i = 1, \ldots, N \), the equation of motion is a second order differential equation according to Newton’s second law

\[
m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i
\]  

(2.1)

where \( m_i \) and \( \mathbf{F}_i \) are the mass of particle \( i \) and the force acting on it, respectively. A direct solution to Equation (2.1) is given by Verlet \[^6\]

\[
\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \mathbf{a}_i(t)\Delta t^2
\]

(2.2)
where \( \mathbf{a}_i(t) = \frac{\mathbf{F}_i(t)}{m_i} \) is the acceleration of particle \( i \) at time \( t \). The Verlet integrator is time-reversible with an error of order \( O(\Delta t^4) \) as implied by the Taylor expansions of \( \mathbf{r}_i(t+\Delta t) \) and \( \mathbf{r}_i(t-\Delta t) \) \[^8\]. The deficiency of the Verlet integrator lies in the mean-value description of the velocity \( \mathbf{v}_i(t) \) (and hence the kinetic energy and temperature) which introduces an error of order \( O(\Delta t^2) \) \[^8\].

\[
\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t} \tag{2.3}
\]

Modifications of the Verlet integrator have been proposed to improve the algorithm. By performing integration over half time step, two successful integrators have been developed after the original Verlet algorithm, namely the leap-frog method \[^{15}\] and the velocity-Verlet algorithm \[^{16}\].

The leap-frog method, as the name implies, produces the position \( \mathbf{r}_i \) at integer time steps and the velocity \( \mathbf{v}_i \) at half integer time steps so that \( \mathbf{r}_i \) and \( \mathbf{v}_i \) jump over each other like leapfrog

\[
\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{1}{2}\Delta t)\Delta t \tag{2.4}
\]

\[
\mathbf{v}_i(t + \frac{1}{2}\Delta t) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \mathbf{a}_i(t)\Delta t \tag{2.5}
\]

The velocity at integer time step is obtained as an average so that the kinetic energy can be evaluated at the same instant as other physical quantities

\[
\mathbf{v}_i(t) = \frac{1}{2}\left( \mathbf{v}_i(t + \frac{1}{2}\Delta t) + \mathbf{v}_i(t - \frac{1}{2}\Delta t) \right) \tag{2.6}
\]

The velocity-Verlet algorithm gives both position \( \mathbf{r}_i \) and velocity \( \mathbf{v}_i \) at integer time steps, yet through the intermediate velocity at half integer time step

\[
\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \tag{2.7}
\]

\[
\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t) + \mathbf{a}_i(t+\Delta t)}{2}\Delta t \tag{2.8}
\]

In practice, the velocities at half integer time steps are first updated, and then the positions, forces and velocities at \( t+\Delta t \) are computed. It can be seen that the velocities are properly handled in the velocity-Verlet algorithm.
The leap-frog method and the velocity-Verlet algorithm, as variants of the Verlet integrator, are very efficient in producing positions and velocities that evolve with time and provide ease and convenience in programming. The leap-frog method is implemented as the main integrator in the GROMACS program package \cite{17} \cite{18} \cite{19}, while the velocity-Verlet algorithm is adopted in the NAMD program package \cite{20}. Both GROMACS and NAMD have been widely used to carry out MD simulations. Besides the Verlet integrator family, some other integration algorithms such as the predictor-corrector algorithm and the Runge-Kutta-Gill method are less frequently used due to their complexity and higher computational cost \cite{8}.

2.1.2 Periodic Boundary Conditions

By integrating Newton’s equations of motion, the particles in the system are able to move through space according to their initial velocities and the interacting forces between them. However, the number of particles that can be treated in MD simulations is limited by the performance and capacity of computer hardware. Today, after dozens of years of development of computers and algorithms, the number of atoms studied in most MD simulations is still within one million, the size being on a too small scale to mimic a macroscopic system.

To bridge the gap between the microscopic system and the macroscopic world, periodic boundary conditions are applied, allowing the simulation unit cell to reasonably model an infinite system by replicating itself over the three dimensions with its periodic images. This treatment works perfectly well for highly ordered crystals but will inevitably introduce artifacts in liquid simulations since liquids are much more disordered than crystals. Such errors can be estimated by performing simulations under various system sizes \cite{21}. Fortunately, the errors are usually very small in large systems containing thousands or more atoms, and fruitful research results have been produced under the periodic approximation.

Needless to say, the periodic boundary conditions require seamless connections between the simulation box and its periodic images. Therefore the simulation box must be in a polyhedron shape that can be stacked in three dimensions. A good choice is the triclinic box with arbitrary shape described by three box vectors \( \mathbf{a}, \mathbf{b}, \) and \( \mathbf{c} \) \cite{21} \cite{22}, its volume being \( V_{\text{box}} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \). The triclinic box type covers the special cases of cubic, rectangular, rhombic dodecahedron, and truncated octahedron boxes \cite{21}.

2.1.3 Cutoff and Neighbor List

A common rule used in periodic boundary conditions is the so-called “minimum image convention”, which requires that each particle only interacts with the nearest image of the other particles in the system except for long-range electrostatic interactions. In this regard, the
cutoff radius for short-range interactions cannot exceed half of the box length, and the number of interacting pairs in a system containing \( N \) particles is on the order of \( O(N^2) \). Therefore the computational cost would be quadrupled when the size of the system is doubled, which is not desired for large-scale simulations. To speed up the simulation while maintaining reasonable accuracy, the cutoff radius for short-range interactions is usually fixed at a constant value (for instance 10–15 Å in all-atom simulations), and then the computational cost is linearly dependent on \( N \).

In a large system where the size of the simulations box is much larger than the cutoff radius, the distances between particles will likely exceed the cutoff radius. In such case it becomes important to quickly and efficiently locate the interacting neighbors within the cutoff radius for each particle, in order to reduce computational costs \[^{23}\]. There’re mainly two types of methods for neighbor searching; the Verlet list method uses a less frequently updated neighbor list with a radius slightly larger than the cutoff radius, while the cell-list method divides the simulation box into many smaller cells so that the neighbors can be found in adjacent cells. For large systems the cell-list method is more efficient.

### 2.1.4 Constraint Algorithms

In a system containing polyatomic molecules, intra-molecular interaction arises. When a force field is used to model the system, the intra-molecular interaction can be decomposed into several bonded potential terms describing bonds, angles, dihedral angles, etc. Among these terms, bonds can be treated as harmonic potential with the minimum at the reference bond lengths; additionally, constraints are often introduced to maintain bonds at constant lengths, mainly for the purpose of increasing the time step of integration. On one hand, the vibrations of the bonds are in higher frequencies than, for instance, the low-frequency conformational changes regarding dihedral angles, and are less interested in MD simulations. Constraining bond vibrations will diminish the high-frequency motions in the system and therefore allow an increase in the integration time step \[^{24}\]. On the other hand, constraints on bonds with hydrogen atoms will prevent the very light hydrogen atoms from moving too far away in a relatively large integration time step.

A typical and conventional constraint algorithm is named SHAKE which processes all constraints one after another \[^{25}\]. The holonomic constraints at time \( t \) take the form

\[
\sigma_k(t) = \left\| r_a(t) - r_b(t) \right\|^2 - d_{ab}^2 = 0, \quad k = 1, \ldots, N_c
\]  

(2.9)

where \( a \) and \( b \) are the two atoms involved in the \( k \)-th constraint, \( d_{ab} \) is the required distance between \( a \) and \( b \), and \( N_c \) is the total number of constraints. Introducing Lagrange multipliers \( \lambda_k \), the constraint forces \( G_i \) are defined as
\[ G_i(t) = -\sum_{k=1}^{N_c} \lambda_k \frac{\partial \sigma_i(t)}{\partial r_i(t)} \]  

(2.10)

The displacements of the atoms are then determined from the combined force of \( F_i \) and \( G_i \). By requiring that \( \sigma_i(t+\Delta t) \) is equal to zero for all the \( N_c \) constraints, a set of non-linear equations can be established. The SHAKE algorithm adopts the Gauss–Seidel method to solve the non-linear equations iteratively until a given tolerance in bond lengths has been satisfied \[25\]. A variant of the SHAKE algorithm named RATTLE \[26\] is developed for use together with the velocity-Verlet integration method. For water molecules an analytical solution of the SHAKE algorithm can be derived for \( N_c = 3 \), namely the SETTLE \[27\] constraint algorithm, which runs much faster than the iterative procedure.

An alternative constraining procedure is the LINCS (linear constraint solver) algorithm which mainly employs matrix manipulations to constrain the bond lengths \[28\]–\[29\]. The LINCS algorithm is non-iterative, and is reported to be three or four times faster than the SHAKE algorithm \[28\]. In the LINCS algorithm, a slightly different set of constraint equations is used

\[ g_k = \| \mathbf{r}_a - \mathbf{r}_b \| - d_{ab} = 0, \quad k = 1, \ldots, N_c \]  

(2.11)

where \( N_c \) is the total number of constraints. The algorithm is carried out in two steps; the first step is to constrain the projection of the new bond on the old bond while the second step is to correct the bond lengthening due to the rotation. The LINCS algorithm adopts the serial expansion in the computation of the inverse of the sparse constraint coupling matrix, and is inherently not suitable for systems with high connectivity, for instance, coupled angle-constraints \[21\]. In molecules where only bond constraints are required, the connectivity is low and the LINCS algorithm is robust and efficient.

### 2.1.5 Thermostats and Barostats

If the equations of motion of an isolated system are integrated exactly according to the leapfrog or the velocity-Verlet algorithm, the energy of the system will be conserved within the accuracy of the algorithm and the computer, and such ensemble is termed as microcanonical or the \( NVE \) ensemble, since the number of particles (\( N \)), the volume of the simulation box (\( V \)) and the energy (\( E \)) are all constant during the simulation. Unfortunately, this is not the case in most experiments where temperature and pressure are present. Therefore the integrator needs modifications to include the influence of external temperature and pressure by introducing thermostats and barostats. Only in this way can the canonical (\( NVT \)) or the isothermal-isobaric (\( NPT \)) ensemble be generated.
A straightforward way of tuning the temperature is the direct scaling of the velocities of the particles so that the temperature of the system is maintained constant. However, this method completely suppresses the fluctuation in the kinetic energy of the system, and the total energy will suffer large oscillations. A more reasonable temperature coupling scheme is the Berendsen thermostat which slowly corrects the temperature \( T \) to the prescribed value \( T_0 \) [30]

\[
\frac{dT}{dt} = \frac{T_0 - T}{\tau_T}
\]

(2.12)

The velocities of the particles in the system is then scaled by a factor \( \lambda \) [30]

\[
\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \left( \frac{T_0}{T} - 1 \right)}
\]

(2.13)

Although the Berendsen thermostat is a weak coupling scheme, it suppresses the fluctuation of the kinetic energy to some extent. The sampling will be problematic and the generated ensemble not exactly canonical [21]. Fortunately, in large systems containing biological macromolecules such errors are very small [21], the Berendsen coupling method is still adopted considering its simplicity and ease in implementation.

The Berendsen pressure coupling algorithm is also a weak coupling scheme similar to the Berendsen temperature coupling [30]

\[
\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}
\]

(2.14)

provided that the system is under isotropic external pressure. Here \( P_0 \) is the prescribed pressure and \( P \) is the isotropic pressure of the system which is given by the volume of the simulation box, the kinetic energy and the virial [30]

\[
P = \frac{1}{3V_{box}} \left( \sum_i m_i v_i^2 + \sum_{i<j} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \right)
\]

(2.15)

where \( \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j \) and \( \mathbf{f}_{ij} \) is the force on particle \( i \) exerted by particle \( j \). The factor for scaling the coordinates and box lengths is

\[
\mu = 1 - \frac{\beta \Delta t}{3\tau_p} (P_0 - P)
\]

(2.16)

Here \( \beta \) is the compressibility which is usually set to the value of water, \( 4.6 \times 10^{-10} \text{ Pa}^{-1} \), since \( \beta \) does not affect the average value of the isotropic pressure \( P \) [21].
While the Berendsen coupling schemes are simple and effective, for small systems more accurate coupling algorithms are necessary to generate the correct ensembles \[^{21}\]. In such cases the extended-ensemble coupling methods are better choices, namely the Nosé-Hoover thermostat \[^{31}\] \[^{32}\] and the Parrinello-Rahman barostat \[^{33}\] \[^{34}\].

The Nosé-Hoover thermostat couples the system to an external reservoir by introducing a variable \(s\) with conjugate momentum \(p_s\) and mass parameter \(Q_s\). Defining the friction parameter \(\zeta = p_s / Q_s\), the equation of motion is modified as \[^{31}\] \[^{32}\]

\[
\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \zeta \frac{d \mathbf{r}_i}{dt} \tag{2.17}
\]

with

\[
\frac{d \zeta}{dt} = \frac{1}{Q_s} \left( \sum m_i v_i^2 - N_{df} k_B T_0 \right) \tag{2.18}
\]

where \(N_{df}\) is the number of degrees of freedom of the system, \(k_B\) is the Boltzmann constant, and \(T_0\) is the prescribed temperature.

The Parrinello-Rahman barostat works in a similar way to the Nosé-Hoover thermostat. Defining the box matrix \(\mathbf{h} = [\mathbf{a}, \mathbf{b}, \mathbf{c}]\) where \(\mathbf{a}, \mathbf{b}\) and \(\mathbf{c}\) are box vectors and the metric tensor \(G = \mathbf{h}^T \mathbf{h}\), the equation of motion of a system under isotropic external pressure becomes \[^{33}\] \[^{34}\]

\[
\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - G^{-1} \frac{d \mathbf{G}}{dt} \frac{d \mathbf{r}_i}{dt} \tag{2.19}
\]

with

\[
\mathbf{W} \frac{d^2 \mathbf{h}}{dt^2} = (P - P_0) \mathbf{\sigma}, \quad \mathbf{\sigma} = V_{box} (\mathbf{h}^T)^{-1} \tag{2.20}
\]

where \(V_{box} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})\) and \(\mathbf{W}\) is the mass parameter matrix that determines the strength of the pressure coupling. The Parrinello-Rahman barostat can be extended to include anisotropic external pressure so that the shape of the simulation box is allowed to change.

### 2.2 Inter-atomic Interactions and Force Fields

A force field uses a combination of predefined potential functions with parameters to approximately describe the potential energy surface of a system. Potential functions are some simple functions used to approximate the interactions between the atoms, and the force acting
Chapter 2  Theoretical Background

on an atom is determined as the negative derivative of potential energy with respect to the position of the atom. Normally the potential energy in a force field is decomposed into two parts, the bonded term and the non-bonded term. The former consists of contributions coming from bonds, angles, dihedral angles and improper dihedral angles, while the latter describes how the van der Waals and electrostatic interactions take effect over space. In some cases where it is necessary to explicitly take into account the polarization effect, the polarization term arises in the expression of the force field. This can be achieved by defining polarizabilities of the molecules or introducing the Drude particle which is attached to the molecule through a harmonic spring.

2.2.1 Bonded Interaction

Bonded interactions arise between atoms linked through one to three bonds. The bond stretching and angle bending are usually approximated by harmonic functions with potential minimum at their reference bond length and angle

\[ E_{b,ij} = \frac{1}{2} k_{b,ij} (r_{ij} - r_{ij,0})^2, \quad E_{a,ijk} = \frac{1}{2} k_{a,ijk} (\theta_{ijk} - \theta_{ijk,0})^2 \]  

(2.21)

Sometimes the interaction between the 1-3 atoms of an angle needs to be taken into account and this is described by the Urey-Bradley potential function which is also a harmonic function

\[ E_{UB,ijk} = \frac{1}{2} k_{UB,ijk} (r_{ijk} - r_{ijk,0})^2 \]  

(2.22)

Clearly the harmonic potential functions are only valid when the bonds and angles are close to their equilibrium configurations. In case bond dissociation occurs, the Morse potential provides an option to describe the anharmonicity

\[ E_{Morse,ij} = D_{ij} [1 - \exp(-\beta_{ij}(r_{ij} - r_{ij,0}))]^2 \]  

(2.23)

where the energy approaches \( D_{ij} \) as the distance \( r_{ij} \) becomes infinity.

Dihedral angles are important factors that determine the conformation of a molecule, and the rotations of dihedral angles have much lower frequencies comparing with the vibrations of bonds and angles. Besides, dihedral angles are periodic in character. A dihedral angle potential term can be properly described by the sum of several cosine functions

\[ E_{d,ijkl} = \sum k_{d,ijkl} [1 + \cos(n \phi_{ijkl} - \phi_{ijkl,0})] \]  

(2.24)

or alternatively by the Fourier series
\[
E_{Fijkl} = \frac{1}{2} [F_0 (1 + \cos \phi) + F_2 (1 - \cos (2\phi)) + F_3 (1 + \cos (3\phi)) + F_4 (1 - \cos (4\phi))] \tag{2.25}
\]

which can be rewritten as the Ryckaert-Bellemans function \cite{21}

\[
E_{RBijkl} = \sum_{n=0}^{5} C_n (\cos (\psi))^n, \quad \psi = \phi - 180^\circ \tag{2.26}
\]

with conversion rules

\[
\begin{align*}
C_0 &= F_2 + (F_0 + F_3) / 2 \\
C_1 &= -(F_0 + 3F_3) / 2 \\
C_2 &= -F_0 + 4F_4 \\
C_3 &= -2F_0 \\
C_4 &= -4F_4 \\
C_5 &= 0
\end{align*}
\tag{2.27}
\]

In case of double bonds or aromatic rings, improper dihedral angles are introduced to maintain the planar structures. The improper dihedral angles can also be used to prevent a chiral center from flipping over to its stereoisomer. Considering the rigidity of the double bonds, aromatic rings and chiral configurations, the improper dihedral angles can be described by harmonic potential functions, similar to bonds and angles

\[
E_{imijkl} = \frac{1}{2} k_{imijkl} (\xi_{ijkl} - \xi_{ijkl0})^2 \tag{2.28}
\]

Caution has to be taken in that the improper dihedral angle is periodic like a regular dihedral angle while the harmonic function is not. If the reference value of an improper dihedral angle \(\xi_{ijkl0}\) is set to 180° or −180°, the potential energy will suffer severe discontinuation as \(\xi_{ijkl}\) oscillates around \(\xi_{ijkl0}\). Fortunately one can always tune the improper dihedral angle by changing the sequence of the four atoms, for instance from \(\xi_{ijkl0}\) to \(\xi_{ijlk0}\), to make sure that the improper dihedral angle is far from ±180°.

2.2.2 Non-Bonded Interaction

In a conventional force field the many-body potential is approximated by the effective pairwise potential, which means that the total non-bonded potential is equal to the sum of the contributions from the interaction of each pair of particles. In this way, the polarization effects in the system are implicitly included in the effective pairwise potential, leading to the limitation that the polarization effect on each molecule is not varying with respect to its
environment, which is not exactly the same case as in realistic systems. Nevertheless, the
effective pairwise potential provides satisfactory description of the systems in most cases, for
instance in aqueous solutions of biological macromolecules [21].

The non-bonded interaction potential between two atoms is described by the sum of the
van der Waals and electrostatic terms, the latter being usually modeled by the Coulomb
potential between point charges, where the dipole, quadrupole and higher-order contributions
are implicitly included in an average manner, since the effective pairwise potentials are
parameterized to reproduce the experimental observations or the results from quantum
chemical calculations. The van der Waals interaction can be conveniently modeled by the
Lennard-Jones potential

\[ E_{LJ,ij} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \]  

(2.29)

and the electrostatic interaction by the Coulomb potential

\[ E_{C,ij} = \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \]  

(2.30)

It should be noted that in a force field the non-bonded interactions between atoms
connected by one or two bonds are excluded, since the constraints and harmonic potential
functions are expected to properly describe the bonds and angles. However, for atoms
connected by three bonds, the dihedral angle potential is usually coupled with a scaled 1-4
non-bonded interaction [21]. Such a treatment enhances the flexibility in the parameterization
of the force field and the transferability of the optimized parameters.

Due to the use of cutoffs in evaluating non-bonded interactions as required by periodic
boundary conditions and computational efficiency, the long-range corrections (LRC) have to
be taken into account for both Lennard-Jones and Coulomb potentials. For the Lennard-Jones
potential, the first term \( (4\epsilon \sigma_{ij}^{12}/r_{ij}^{12}, \text{repulsive term}) \) decays so fast with respect to the distance \( r_{ij} \)
that it becomes totally negligible at the cutoff radius, therefore only the LRC to the second
term \( (-4\epsilon \sigma_{ij}^{6}/r_{ij}^{6}, \text{dispersion term}) \) needs to be evaluated. Assuming that the radial distribution
function or pair correlation function \( g(r) \) is equal to unity beyond the cutoff radius \( r_c \), the
dispersion correction to the potential energy is [8][21]

\[ E_{LJ,LRC} = \frac{1}{2} N \rho \int_{r_c}^{\infty} 4\pi r^2 \left( -\frac{4\epsilon \sigma_{ij}^{6}}{r^6} \right) dr \]

\[ = \frac{8\pi}{3} N \rho \epsilon \sigma_{ij}^{6} r_c^{-3} \]  

(2.31)
where \( \rho = N/V_{\text{box}} \) is the number density of the atoms.

As shown in Equation (2.30), the Coulomb potential is proportional to \( r_{ij}^{-1} \) and therefore long-ranged in character. For the Coulomb potential, an integration similar to Equation (2.31) does not converge at any given cutoff radius \( r_c \), and the evaluation of the long-range electrostatic interactions becomes tricky. The electrostatic interactions between the atoms in the simulation box and their translational periodic images have to be taken into account explicitly to calculate the total Coulomb potential

\[
E_{C,\text{total}} = \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}} \sum_{\mathbf{n}^*} \frac{q_i q_j}{r_{ij,n}}
\]

where \( \mathbf{n} = (n_x, n_y, n_z) \) is the box index vector and the asterisk over \( ij \) means that \( i \neq j \) when \( \mathbf{n} = (0, 0, 0) \). This sum is slowly and conditionally convergent and therefore inconvenient for the computation of the Coulomb potential \cite{21}. Ewald in 1921 proposed the idea of transforming the full Coulomb potential into two quickly-converging terms that can be evaluated in real space and Fourier space, respectively \cite{35} \cite{36}. In this procedure, the first step is to introduce a Gaussian charge distribution with the same magnitude but opposite sign around each point charge so that the resultant Coulomb potential becomes convergent in real space, since the point charges are partially screened by the introduced Gaussians. Then the compensatory Gaussian charge distributions are introduced, whose contributions to the Coulomb potential can be evaluated in the Fourier space. Finally, the self-interactions of the firstly introduce Gaussians are removed from the potential. The Ewald summation therefore provides a practical approach for evaluating long-range electrostatic potentials. Unfortunately its computational cost is on the order of \( O(N^2) \) or \( O(N^{3/2}) \) and the method becomes very inefficient when simulating large systems.

The particle mesh Ewald (PME) method \cite{37} \cite{38} was then developed to improve the performance of the conventional Ewald summation. In PME, the atomic charges are interpolated on grid points which are then transformed by using the fast Fourier transformation algorithm. After such treatment the evaluation of the Fourier part can be speeded up and the computational cost of the PME method becomes on the order \( O(N \cdot \log(N)) \). Therefore, the PME method is much more efficient and practical than the conventional Ewald summation, and has so been widely used in large-scale MD simulations.

### 2.2.3 Polarization

Polarization effects are important in many-body interactions, and conventional force fields usually include the polarization effects implicitly in the effective pairwise potentials. Such a treatment is simple and often produces satisfactory results for biological simulations; however,
the limitations become visible when the polarization effects are pronounced, for instance, in the presence of a highly charged metal ion. In this case the polarization effects have to be taken into account explicitly. For small molecules, a simple solution is to introduce the Drude particle \(^{12}[21]\).

A Drude particle is a massless particle attached to an atom through a harmonic spring. The Drude particle carries some electric charge and is subject to the electric field of the surrounding point charges; it stays at the position where the electrostatic forces arising from the surrounding point charges offset the elastic force of the harmonic spring. Therefore, in a system containing multiple Drude particles, a self-consistent field calculation has to be carried out at each time step to locate the positions of the Drude particles, leading to heavier computational burden than the simulations using a non-polarizable force field. The advantage of the Drude particle is that it keeps the simplicity of the point charge model and that the charge on each site is fixed during simulations \(^{12}\), making it easy to implement in MD programs. However, it should be noted that the Drude particle is difficult to parameterize. So far, it has only been introduced for small molecules.

2.2.4 Several Force Fields and Water Models

Classical all-atom force fields have been very popular in MD simulations of organic and biomolecular systems, including AMBER \(^{39}[40] [41]\), CHARMM \(^{42} [43]\), OPLS \(^{44} [45]\), etc. These force fields adopt similar potential functions, with the exception that the CHARMM force field uses the Urey-Bradley term to account for 1-3 interactions. The AMBER (assisted model building with energy refinement) force field has several variants such as AMBER94 \(^{39}\), AMBER99 \(^{40}\) and AMBER03 \(^{41}\), with its special feature of the use of partial atomic charges from restrained electrostatic potential (RESP) fitting. In the AMBER force field the 1-4 scaling factor is set to 0.5 for the Lennard-Jones potential and 0.8333 for the Coulomb potential. The CHARMM (chemistry at Harvard macromolecular mechanics) force field uses partial atomic charges determined based on the solute-water complexes. The 1-4 scaling factor is not used in the CHARMM force field, in other words, the factor is set to unity. The OPLS (optimized potentials for liquid simulations) force field uses empirically optimized partial atomic charges and electrically neutral subunits. In the OPLS force field, the 1-4 non-bonded scaling factor is 0.5 for both van der Waals and electrostatic interactions. Besides, the general AMBER force field \(^{46}\) and the CHARMM general force field \(^{47}\) have been introduced to model small organic molecules and pharmaceutical compounds.

In particular, water models provide the specially tailored force fields for water molecules. Theoretical investigations and simulations on the structure of water have been very active and plenty of water models have been published in the literature so far. The famous three-site and four-site transferable intermolecular potentials for water, namely TIP3P \(^{48}\) and TIP4P \(^{49}\), are
successful parameterizations of water models using simple potential functions. At the computational cost of an extra site, the TIP4P water model is reported to give better density and quadrupole moment than the TIP3P model. The simple point charge model SPC \(^{[50]}\) and its extended version SPC/E \(^{[51]}\) are also popular due to their simplicity and computational efficiency. The SPC/E model was proposed to improve the diffusion constant of the SPC model. Besides, the SWM4-DP model \(^{[52]}\) is a more recent polarizable water model. It is similar to the TIP4P model in structure, except for the Drude particle attached to the oxygen site through a harmonic spring. The SWM4-DP model is able to reproduce a variety of physical properties of liquid water which are in excellent agreement with experiment.
Chapter 3  Applications in Solution Chemistry

As the most abundant liquid on Earth, water covers 70% of Earth’s surface in its liquid and solid forms. Water participates in the ecological cycle and acts as the essential element of life; the chemical and biochemical processes that are essential to living systems take place in aqueous environment. Besides, many laboratory studies and industrial processes are conducted in solution. The chemistry of solution is therefore very important to both human being and the planet of Earth. Experimental and theoretical studies of the structure of water and the solvated solute molecules have been very abundant in scientific research.

From a theoretical point of view, MD simulations possess the natural advantage of explicitly taking into account the solvent environment, which is less well described in pure quantum chemical calculations. The use of simple potential functions ensures the involvement of a large number of solvent molecules and the sampling over long period of simulation time at an economical computational cost. The structural fluctuation and distortion can be observed at a given temperature, giving rise to the ensemble-averaged properties. Moreover, the free energy of a process can be calculated routinely through free energy algorithms such as free energy perturbation, thermodynamic integration, umbrella sampling, etc.

In MD simulations, several properties can be obtained from the trajectory for better insight into the structural and dynamical behavior of the system. One important structural property is the radial distribution function (RDF) between a pair of particle types of \(A\) and \(B\) \[21\]

\[
g_{AB}(r) = \frac{1}{\rho_B N_A} \sum_{i \in A} \sum_{j \in B} \frac{\delta(r_{ij} - r)}{4\pi r^2}
\]  \hspace{1cm} (3.1)

where \(\rho_B\) is the average number density of particles \(B\) in the sphere with the largest radius (usually half of the box length), and \(N_A\) is the number of particles \(A\). Normally the RDF curve will show distinguished peaks implying the coordination shells and decay to unity at a large \(r\). The calculated RDF can be integrated to obtain the coordination number of each coordination shell. Take a metal ion for example, the position of the first RDF peak and the coordination number in the first coordination shell are the characteristic properties that can be compared with experimental measurements.

Dynamical properties such as the self-diffusion coefficient and the mean residence time can also be calculated from the simulation trajectory. Given a molecule, its self-diffusion coefficient \(D\) can be computed from the mean square displacement, according to the Einstein relation \[53\]
Chapter 3    Applications in Solution Chemistry

\[ 6D = \lim_{t \to \infty} \frac{d}{dt} \langle \| \mathbf{r}(t) - \mathbf{r}(0) \|_{l_2}^2 \rangle \]  

(3.2)

where the angle brackets represents the ensemble average. Provided that the statistics is good enough, the mean square displacement will be linearly dependent on the simulation time, and then the self-diffusion coefficient is readily obtained from the slope. The mean residence time is usually used to describe solvent dynamics around a solute molecule such as a metal ion, and can be calculated by fitting to a function \( n_{\text{ion}}(t) \), according to Impey et al. \[54]\n
\[ n_{\text{ion}}(t) = \left\langle \sum_j P_j(t_0, t_0 + t; t^*) \right\rangle \]

(3.3)

where the summation \( j \) runs over all solvent molecules and the angle brackets denote ensemble average over all \( t_0 \) times. The probability function \( P_j(t_0, t_0 + t; t^*) \) takes the value of unity if the solvent molecule \( j \) is in the vicinity (usually the first solvation shell) of the solute molecule at time \( t_0 \) and \( t_0 + t \) and does not leave the region for any continuous period longer than \( t^* \). In all other cases \( P_j(t_0, t_0 + t; t^*) \) takes the value of zero. The computed \( n_{\text{ion}}(t) \) function decays exponentially, \( n_{\text{ion}}(t) \sim n_{\text{ion}}(0) \exp(-t/\tau_{\text{ion}}) \), and the mean residence time \( \tau_{\text{ion}} \) can be obtained through exponential fitting.

Importantly, free energies differences are accessible in MD simulations. By setting up a thermodynamic cycle, the free energy of a process can be calculated from individual steps. For example, the hydration free energy of a metal ion \( M^{n+} \) can be calculated from the following thermodynamic cycle \[55]\n
\[
\int_{0}^{1} \left[ \frac{\partial H}{\partial \lambda} \right]_{\text{NPT}, \lambda} \, d\lambda
\]

(3.4)

where \( \Delta G_{\text{hyd}} = \Delta G_{\text{cav}} + \Delta G_{\text{elec}} \) and the latter two \( \Delta G \) can be calculated by employing a parameter \( 0 \leq \lambda \leq 1 \) to couple the initial and final states \[21\]

where \( H \) is the Hamiltonian \( H(p, q; \lambda) \). The derivative of \( H \) with respect to \( \lambda \) is computed as an ensemble average and then integrated to give the free energy difference \( \Delta G \).
3.1 Cd-113 NMR Shielding Constant of Aqua Cd(II) Ion (Paper I)

NMR measurements of transition metals are important in biochemical research \[^{[56]}\]. The chemical shift of the transition metal center is sensitive to the chemical environment and can be detected in the NMR experiment, and the coordination details can therefore be determined for better understanding of the structure and function of the metal center. For instance, the Zn(II) ion is very abundant in living organisms and appears in many enzymes. However, NMR measurements of Zn(II) are very difficult due to the low sensitivity of the \(^{67}\)Zn nucleus. As an alternative approach, Cd(II) is usually used to replace Zn(II) ion and a \(^{113}\)Cd NMR measurement is carried out to determine the chemical environment and the coordination structure of the metal center \[^{[57]}\]. In paper I we employed MD and quantum chemical density functional theory (DFT) methods to investigate the \(^{113}\)Cd NMR shielding constant of aqua Cd(II) ions. Two issues were addressed. Firstly, conventional MD simulations using non-polarizable water models do not give reliable Cd-O distance in the coordination structure \[^{[58]}\]. Secondly, conventional quantum chemical calculations do not take into account relativistic effects explicitly \[^{[59]}\].

![Figure 3.1](image)

**Figure 3.1.** (left) Structure of the \([\text{Cd(H}_2\text{O)}_6]^{2+}\) complex. (right) Calculated quantum chemical potential energies (dots) and fitted molecular mechanical potential energy curve (line). Reprinted with permission from the American Chemical Society.

The parameterization of the Cd(II) ion is based on the quantum mechanical second-order Møllner-Plesset (MP2) calculations. First, the potential energies for the \([\text{Cd(H}_2\text{O)}_6]^{2+}\) complex is computed by the MP2 method, and an initial set of the Lennard-Jones parameters for the Cd(II) ion was obtained by fitting the molecular mechanical potential energies to the MP2 results. Subsequent test simulations based on the initial set of Lennard-Jones parameters suggest a coordination number of six for aqua Cd(II) ion. Therefore it is more reasonable to use the \([\text{Cd(H}_2\text{O)}_6]^{2+}\) complex to generate the Lennard-Jones parameters of the Cd(II) ion. As shown in Figure 3.1, the \([\text{Cd(H}_2\text{O)}_6]^{2+}\) complex is constructed to model the first solvation shell of the ion. The use of six water molecules ensures the involvement of the polarization effect of the Cd(II) ion on the surrounding water molecules. By varying the six Cd-O bond lengths simultaneously, the potential energies can be obtained as in Figure 3.1. The molecular
mechanically calculated potential energy curve was then fitted to the MP2 result, giving rise to the Lennard-Jones parameters of the Cd(II) ion. After parameterization, MD simulations were carried out. A Cd(II) ion together with 1024 polarizable SWM4-DP [52] water molecules were put into a cubic unit cell and subject to 250 ps simulations under constant temperature (298 K) and pressure (1 atm). The structure of aqua Cd(II) ion was checked by analyzing the RDF between the cadmium ion and the oxygen atoms in water molecules. The calculated RDF between cadmium and oxygen is shown in Figure 3.2. The Cd-O distance of 2.26 Å and coordination number of six are in good accordance with experimental data [60]. From the last 50 ps simulation trajectory fifty snapshots were extracted for the subsequent quantum chemical calculations of the NMR shielding constants.

![Figure 3.2](image.png)

**Figure 3.2.** Radial distribution function between cadmium ion and oxygen atoms in water molecules. Reprinted with permission from the American Chemical Society.

After MD simulations, quantum chemical calculations were carried out using the gauge including atomic orbital (GIAO) approach [61] for the non-relativistic NMR shielding constant, which is equal to the sum of the diamagnetic and paramagnetic contributions. In quantum chemical calculations it is important to set up a proper computational model, since the whole simulation box far exceeds the available computational power. An ideal choice would be the first solvation shell of Cd(II) ion, since the surrounding water molecules in the first solvation shell are expected to largely influence the electronic structure. To verify the effect of surrounding water molecules on the $^{113}$Cd NMR shielding constant, test calculations were carried out on three structures extracted from the first snapshot. The bare Cd(II) ion, [Cd(H$_2$O)$_6$]$^{2+}$ complex (the first solvation shell), and [Cd(H$_2$O)$_{18}$]$^{2+}$ complex (including most of the second solvation shell) were subject to quantum chemical calculations of the $^{113}$Cd NMR shielding constant, as shown in Figure 3.3. The six water molecules in the first solvation shell give rise to a decrease of more than 1000 ppm in the shielding constant, while twelve more water molecules in the second solvation shell results in a further decrease of only
20 ppm. Therefore, the first solvation shell, which contributes dominantly to the shielding constant, can be used as the computational model to balance computational cost and accuracy.

Figure 3.3. Structures and computed non-relativistic $^{113}$Cd NMR shielding constants of the bare Cd(II) ion, [Cd(H$_2$O)$_6$]$^{2+}$ complex and [Cd(H$_2$O)$_{18}$]$^{2+}$ complex.

One purpose of this study is to explicitly take into account relativistic effects; we therefore adopt the leading-order Breit-Pauli perturbation method $^{[62]}$ $^{[63]}$ to compute the relativistic contribution to the NMR shielding constant. The computed total $^{113}$Cd NMR shielding constant, obtained as the sum of the non-relativistic part and the relativistic corrections, is found to vary with respect to simulation time, as shown in Figure 3.4. During the simulation, the structure of the [Cd(H$_2$O)$_6$]$^{2+}$ complex was distorted by thermodynamic motions of the cadmium ion and water molecules, and the fluctuating Cd-O distances are expected to result in large perturbations on the shielding constant. Therefore we also monitored the time-evolution of the average Cd-O distance in the first coordination shell. It is interesting to find that the two curves have very similar shapes, suggesting a close relationship between the total shielding constant and the average Cd-O distance. Such a relationship can be fitted to a linear function, as shown in Figure 3.4. Similarly, an even closer linear relationship is established between the total NMR shielding constant and the corresponding paramagnetic non-relativistic contribution, as shown in Figure 3.5, indicating the origin of the fluctuation in the total NMR shielding constant.

Figure 3.4. (left) Time-evolution and running average of total shielding constant in comparison with time-evolution of average Cd-O distance. (right) Linear relationship between $\sigma_{\text{total}}$ and average $r_{\text{Cd-O}}$. Reprinted with permission from the American Chemical Society.
Finally the fluctuations of the three contributions to the total shielding constant were analyzed by plotting the instantaneous deviations from their respective average values, as shown in Figure 3.6. The paramagnetic non-relativistic part fluctuates most significantly, while the other two are much more stable. It is interesting to find that the instantaneous deviations of the diamagnetic non-relativistic part and the relativistic part are similar in magnitude while opposite in sign. Moreover, the standard deviations of the diamagnetic non-relativistic, paramagnetic non-relativistic and relativistic parts are determined as 2 ppm, 45 ppm and 3 ppm, respectively. The fluctuation of the relativistic part is so small with respect to the change of surrounding chemical environment that the relativistic contributions in the target compound and the reference compound may cancel each other in the computation of
the \(^{113}\text{Cd}\) NMR chemical shift. Therefore explicit treatment of relativistic effects is not necessary in such cases.

### 3.2 Paramagnetic F-19 NMR Shift in Fluorinated Cysteine (Paper II)

Paramagnetic NMR shift provides a practical approach towards the determination of the structures of trans-membrane proteins, which is an important and difficult issue in biological studies\(^{[64]}\). Taking advantage of the concentration gradient of dioxygen molecules in the membrane, the immersion depth of a properly labeled residue can be deduced from the \(\text{O}_2\)-induced paramagnetic NMR shift\(^{[67]}\). Technically, \(^1\text{H}\), \(^{13}\text{C}\), \(^{15}\text{N}\), and \(^{19}\text{F}\) are all affected by the unpaired electrons in the \(\text{O}_2\) molecule; however, \(^{19}\text{F}\) is the most suitable nucleus for paramagnetic NMR shift measurements, due to its wide chemical shift range of over 900 ppm. Besides, the \(^{19}\text{F}\) nucleus also has advantages such as 100% natural abundance, high sensitivity, and low background of naturally occurring resonances\(^{[68]}\). In proteins, \(^{19}\text{F}\) nuclei can be conveniently introduced through cysteine mutagenesis and subsequent thioalkylation with 3-bromo-1,1,1-trifluoropropanone. In paper II we combined MD and DFT methods to study \(\text{O}_2\)-induced paramagnetic \(^{19}\text{F}\) NMR shifts in fluorinated cysteine. The dynamic behavior of \(\text{O}_2\) was monitored during the MD simulations, and its influence on the \(^{19}\text{F}\) NMR shift was obtained from quantum chemical DFT calculations on discrete snapshots. It is interesting to see i) how the \(\text{O}_2\) molecules reside around \(^{19}\text{F}\) labels and ii) how the structural parameters affect the paramagnetic shift.

![Chemical structure and calculated partial atomic charges of fluorinated cysteine.](image)

The use of proper force fields is crucial for MD simulations. Here the general AMBER force field\(^{[46]}\), which was designed for small organic and pharmaceutical molecules, was adopted to model the fluorinated cysteine. The geometry of the fluorinated cysteine was first optimized at the density functional level of theory, and then the electrostatic potential was calculated to derive the RESP charges\(^{[46]}\) as partial atomic charges. It should be noted that the solvent effect of water was taken into account by the polarizable continuum model\(^{[64]}\), and that the optimized cysteine is in its zwitterionic form. The RESP charges for the fluorinated cysteine molecule are shown in Figure 3.7. For dioxygen, the model by Zasetsky and Svishchev was adopted\(^{[65]}\), which was parameterized to reproduce the experimental
quadrupole moment of the O\textsubscript{2} molecule. The four-site TIP4P water model\textsuperscript{[66]} was used to represent water molecules; this model gives better density and quadrupole moment than the three-site TIP3P model.

To mimic experimental conditions, the simulation box was constructed as a periodic cubic unit cell consisting of one fluorinated cysteine molecule, two O\textsubscript{2} molecules and 850 water molecules, where the concentration of O\textsubscript{2} reflects the oxygen partial pressure of 100 atm according to Henry’s law (\(k_H = 1.3 \times 10^{-3} \text{ mol L}^{-1} \text{ atm}^{-1}\) at 298 K). The simulations were carried out in the \textit{NPT} ensemble, with temperature maintained at 298 K and pressure at 100 atm. After the simulation, the coordinates of the fluorinated cysteine and the dioxygen molecules were extracted from the trajectory. The shortest distance between fluorine atoms in fluorinated cysteine and oxygen atoms in O\textsubscript{2} molecules was monitored, since the F-O distance is expected to dominate the paramagnetic NMR shift. Experience suggests that the O\textsubscript{2}-induced paramagnetic \textsuperscript{19}F NMR shift becomes negligible when the shortest F-O distance is larger than 4.0 Å; therefore this distance is used as a cutoff in extracting the snapshots from the trajectory. Among the 20000 saved frames, there are 1511 frames in which the shortest F-O distance is smaller than 4.0 Å. Moreover, it is found that there is only one adjacent O\textsubscript{2} molecule around the fluorine atoms at one time. Our idea is to compute the average over these 1511 frames and then scale the result by the factor 0.076, which is equal to 1511/20000.

![Figure 3.8. Computed orbital shielding constants in the absence and presence of adjacent O\textsubscript{2} molecule. Reprinted with permission from the American Chemical Society.](image)

To further reduce the computational cost, 75 snapshots were selected from the 1511 frames at regular intervals. In each snapshot, the average orbital NMR shielding constants of the fluorine atoms were computed in the absence and the presence of the adjacent O\textsubscript{2} molecule, using the GIAO method\textsuperscript{[61]}. The computational result is shown in Figure 3.8. The presence of the O\textsubscript{2} molecule results in a small decrease of 0.84 ppm in the orbital shielding constant.
Neglecting the influence of the O$_2$ molecule on the reference compound, the O$_2$-induced orbital shift is equal to $0.84 \times 0.076 = 0.06$ ppm, which is very small in magnitude. The main contribution to the O$_2$-induced $^{19}$F NMR shift comes from the paramagnetic temperature-dependent part of the NMR shielding tensor, which is expressed as the product of the Cartesian components of the electronic $g$-tensor, the hyperfine coupling tensor and the effective spin $S$.\[^{[70]}\] In the very case of fluorinated cysteine and O$_2$ molecules, the expression for the paramagnetic NMR shielding tensor can be significantly simplified by neglecting the spin–orbit contribution to the hyperfine coupling tensor and the deviation between the electronic $g$-tensor and the free electronic factor $g_e$. The paramagnetic shift is then dependent on the Fermi contact part of the isotropic hyperfine coupling constant.

Figure 3.9. Spin density contour in one of the snapshots. Reprinted with permission from the American Chemical Society.

Figure 3.10. (left) Three-dimensional plot showing the dependence of hyperfine coupling constant on the structural parameters. (right) Performance of the fitted function $f(r,\phi) = (c_1\phi^2 + c_2\phi + c_3) / r^5$ in reproducing the contact shift. Reprinted with permission from the American Chemical Society.

Figure 3.9 shows the spin density contour surface in one snapshot of the fluorinated cysteine and the adjacent O$_2$ molecule. It is interesting to find that the spin density distribution, which determines the paramagnetic shift of the $^{19}$F nucleus, is closely related to the structural parameters, the shortest F-O distance $r$ and the corresponding F-O-O angle $\phi$. The F-O distance determines the magnitude of the paramagnetic shift while the F-O-O angle decides
the sign of the shift. An illustrative three-dimensional plot is shown in Figure 3.10, together with the fitted function \( f(r, \phi) = (c_1 \phi^2 + c_2 \phi + c_3) / r^9 \) which reproduces the paramagnetic shift from \( r \) and \( \phi \). Finally the computed paramagnetic shift is plotted with respect to the selected snapshots as shown in Figure 3.11. The shift is averaged and scaled by the factor of 0.076 to give the ensemble average, which is 3.38 ± 0.60 ppm, in agreement with experimental observations for the lipid-bilayer-embedded CF\(_3\)(CF\(_2\))\(_5\)C\(_2\)H\(_4\)-O-maltose (3.5−7 ppm) \(^{71}\) \(^{72}\) and the membrane-embedded diacylglycerol kinase (0− 3 ppm) \(^{73}\). The combined MD and DFT calculation has shown to be useful in treating complex systems dominated by weak interactions where pure quantum chemistry computations cannot be fully put into practice.

![Figure 3.11.](image-url)

**Figure 3.11.** Time-evolution and running average of the O\(_2\)-induced \(^{19}\)F NMR shift in the selected snapshots. Reprinted with permission from the American Chemical Society.

### 3.3 Solvation Free Energies and Structures of Metal Ions (Paper III)

Metal ions can play various roles in nature, and the thermodynamics of the solvation process of the metal ions is an interesting topic from both experimental and theoretical aspects \(^{60}\) \(^{74}\) \(^{75}\). As a powerful numerical approach, MD simulations have been extensively carried out in theoretical studies. However, when using conventional water models the metal-water interaction is not well described since the polarization effects are not explicitly taken into account. The highly charged metal ion is able to significantly polarize the surrounding water molecules and affect the metal-water distance and coordination number. In this regard, combined quantum mechanical/molecular mechanical (QM/MM) simulations usually give more reliable results in the study of ion solvation, yet they are much more computationally expensive. Fortunately, it is possible to alleviate this problem in MD simulations by using a polarizable water model in which the polarization effect is explicitly taken into account, for
instance the SWM4-DP water model which has been proved to be capable of proper description of the structures and thermodynamics of solvated alkali metal ions. In paper III we employ the polarizable SWM4-DP water model to study the solvation structures and thermodynamics of nine metal ions, Al\(^{3+}\), Zn\(^{2+}\), Mg\(^{2+}\), Cd\(^{2+}\), Hg\(^{2+}\), Ca\(^{2+}\), Sn\(^{2+}\), Na\(^+\) and K\(^+\). The aim of this study is to optimize the Lennard-Jones parameters of the ions in order to give consistent hydration structures and free energies with experimental observations, in order to obtain superior accuracy to conventional water models at less computational cost than QM/MM simulations.

The SMW4-DP water model is similar to the TIP4P water model except for the Drude particle which accounts for the polarization effects explicitly. The charge on the Drude particle interacts with the electric field of the surrounding point charges and has an effect on the electric field as well; therefore the positions of the Drude particles are optimized through a self-consistent calculation procedure. In this way, the charge distribution of a water molecule is affected by its environment so that the polarization effect in water is taken into account. For simplicity, the metal ions are represented by point charges with Lennard-Jones parameters. Such a treatment is based on the fact that a closed-shell metal ion is surrounded by a spherical water continuum, where the polarization effect on the metal ion is expected to be negligible on average. A point charge together with Lennard-Jones parameters will suffice in the statistical description of the solvation structure and thermodynamics of the metal ion. Motivated by the successful parameterization in paper I, we again utilize the hexa-coordinated metal-water complex \([\text{M(H}_2\text{O)}_6]^n\) as the model to derive the initial Lennard-Jones parameters for the metal ions. The potential energy curve with respect to the metal-oxygen distance is determined by the quantum mechanical MP2 calculations, and then the molecular mechanical potential energy curve is fitted to the MP2 results to obtain the initial set of Lennard-Jones parameters. This initial set of parameters is then fed into simulations to calculate the hydration structure and free energy, and is further refined by comparing the theoretical results with the experimental data.

Snapshots of the metal-water complexes from MD simulations with the final refined Lennard-Jones parameters are shown in Figure 3.12. Each snapshot illustrates the typical number of water molecules coordinating with the metal ion. The hydration structure is also reflected in the RDFs and the cumulative coordination number curves in Figure 3.13. Al\(^{3+}\) has the largest positive charge and possesses a stable hexa-coordinating structure. Zn\(^{2+}\) and Mg\(^{2+}\) are less positive charged and of larger size compared to Al\(^{3+}\). Both of them are also hexa-coordinated but with longer metal-oxygen distances than that of Al\(^{3+}\). As second-row and third-row transition metal ions, Cd\(^{2+}\), Hg\(^{2+}\) and Sn\(^{2+}\) have even larger sizes and their coordinating numbers exceed six. K\(^+\) is of similar size to Ca\(^{2+}\) and its coordination number is smaller due to its less positive charge. Na\(^+\) has a smaller coordination number than K\(^+\) due to its smaller size. In general both metal-water distances and coordination numbers are well reproduced comparing with experimental data.
Chapter 3 Applications in Solution Chemistry

Figure 3.12. Snapshots of the metal-water complexes from MD simulations: [Al(H$_2$O)$_6$]$^{3+}$, [Zn(H$_2$O)$_6$]$^{2+}$, [Mg(H$_2$O)$_6$]$^{2+}$, [Cd(H$_2$O)$_6$]$^{2+}$, [Hg(H$_2$O)$_2$]$^{2+}$, [Ca(H$_2$O)$_8$]$^{2+}$, [Sn(H$_2$O)$_8$]$^{2+}$, [Na(H$_2$O)$_6$]$^{+}$ and [K(H$_2$O)$_7$]$^{+}$.

Figure 3.13. Metal-oxygen radial distribution functions and coordination numbers for the nine metal ions in aqueous solutions. Reprinted with permission from the American Institute of Physics.

The solvation free energies were computed through the free energy perturbation approach, in which the initial and final states were coupled by a parameter $0 \leq \lambda \leq 1$. By decomposing the solvation process into a thermodynamic cycle described at the beginning of Chapter 3, the solvation free energy can be obtained as the sum of the electrostatic part and the van der Waals contribution. The former was calculated by changing the charged ion M$^{n+}$ to the uncharged M$^0$, and the latter was computed by changing M$^0$ to nothing. The soft-core potential was used in the latter process to avoid numerical instability during the elimination or
creation of the ion. The calculated free energy derivatives, \( \frac{dG_{QQ}}{d\lambda} \) and \( \frac{dG_{LJ}}{d\lambda} \), are shown in Figure 3.14 as functions of \( \lambda \). It can be seen that the net charge of the ion has a significant effect on the \( \frac{dG_{QQ}}{d\lambda} \) curve which contributes greatly to the solvation free energy. The van der Waals contributions are one order-of-magnitude smaller than the electrostatic parts; nevertheless, they are not negligible. Overall, the calculated hydration free energies have been nicely reproduced by the refined Lennard-Jones parameters. Besides, the dynamical properties such as self-diffusion coefficient and mean residence time are also qualitatively reproduced by these derived parameters. Such a fitting-refining procedure is able to give the optimal set of Lennard-Jones parameters and is therefore useful in deriving the physical pictures of the solvated metal ions.

![Figure 3.14](image)

**Figure 3.14.** The computed free energy derivatives \( \frac{dG_{QQ}}{d\lambda} \) and \( \frac{dG_{LJ}}{d\lambda} \) as functions of the coupling parameter \( \lambda \). Reprinted with permission from the American Institute of Physics.

### 3.4 Protein Adsorption onto Titanium Dioxide Surface (Paper IV)

The study of protein adsorption onto implant surfaces is of primary importance in the fields of biomaterials, biomedical devices, and biosensors. The mechanisms of cell adhesion, spreading, and proliferation as well as the regulation of biocompatibility and bioactivity of the implants are significant issues in clinical applications. Specifically, titanium is regarded as a kind of biocompatible material with lightweight and tensile strength, and has been employed as medical implants including dental implants, artificial joints, and blood-contacting devices. This is due to the presence of the titanium dioxide film on the titanium surface, which prevents further corrosion of the metal \(^{77}\) and enhances the biocompatibility and bone-bonding ability by inducing the formation of apatite layers \(^{78}\). Therefore, understanding the protein-TiO\(_2\)-solvent interaction and the dynamic mechanism of the adsorption/desorption process will be of great help in the design of implants with optimal bioactivity. In paper IV we select the subdomain IIIb of human serum albumin (HSA) as a model to investigate the protein adsorption behavior on the non-hydroxylated and hydroxylated rutile TiO\(_2\) (110)
surfaces and to elucidate the mediation of the interfacial water affected by the surface modifications. HSA is the most abundant blood protein involved in the biocompatibility for biomaterials and has been extensively used for investigations of the protein adsorption process by both experiments and simulations [79] [80] [81]. The AMBER03 force field [41] is adopted to describe the HSA_IIIb protein together with the SPC/E water model [51], while the parameters for the rutile TiO\(_2\) substrate are taken from the paper by Předota et al. [82].

Figure 3.15. Snapshots of the adsorbed residues on (left) hydroxylated and (right) non-hydroxylated rutile TiO\(_2\) surfaces. Protein atoms within 7 Å of the surface Ti atoms are shown in the CPK model. Only the interfacial water molecules are shown for clarity; the hydrogen bonding networks among them are displayed as dashed white lines. Reprinted with permission from the American Chemical Society.

Figure 3.15 shows the snapshots taken from the simulation trajectories of the HSA_IIIb protein adsorbing onto hydroxylated and non-hydroxylated rutile TiO\(_2\) surfaces. On both surfaces, the acidic and basic residues show strong binding to the surface, indicating that the anionic carboxylate ion and the cationic ammonium ion are the main adsorbed groups onto the surface, in agreement with experimental measurements [83]. Comparison between the two types of surfaces shows that the TiO\(_2\) surface modified with −OH groups exhibits much greater affinity to HSA_IIIb, supported by the stronger protein-surface electrostatic interaction and more adsorbed residues. This conclusion is also supported by the constant-velocity pulling of the protein from the surface, which shows that the hydroxylated surface has a much stronger affinity to the HSA_IIIb protein.

Further inspection into the adsorption process reveals the adsorption details on both surfaces. On the hydroxylated surface, the side chains of the aspartic acid and threonine residues bind with the water molecules above the −OH groups, while the side chains of the glutamic acid and lysine residues bind directly with the −OH groups on the modified surface, replacing the interfacial water molecules. Differently, the adsorption is weaker on the non-hydroxylated surface; only the carboxylate groups in glutamic acid and aspartic acid and the ammonium group in lysine bind with the interfacial water molecules.

To explore the mechanism of protein adsorption on different interfaces, the water molecules at the interfaces are examined with care. Here the terminal −OH groups on the
hydroxylated surface and the first water layer on the non-hydroxylated surface are denoted as L1, and the water layers above L1 on both surfaces are denoted as L2. Figure 3.16 shows the vertical density distributions of oxygen atoms in L1 and L2 after protein adsorption. It is found that the separation between L1 and L2 peaks is larger for the hydroxylated surface, which can be interpreted by the fact that the highly ordered structure of the modified terminal hydroxyl groups enhances the number of hydrogen bonds inside L1 and reduces the possibility of H-bond formation between L1 and L2. Similar phenomenon has been reported in the paper by Wang et al. In short, it is concluded from MD simulations that protein adsorption onto TiO$_2$ surface is strongly mediated by the interfacial water molecules and can feasibly be manipulated through appropriate surface modification. The redistribution and reorientation of the water molecules at the interface are of great importance in the adsorption process, and the understanding of such behavior will be of help in the design of biocompatible materials, biosensors, and bioactive implants for specific adsorption of proteins.

![Figure 3.16. Vertical density distributions of (left) oxygen atoms in water molecules (named OW) and in terminal hydroxyl groups (named OH) and (right) atoms of the protein after the adsorption, with respect to the distance from the surface Ti atoms. Reprinted with permission from the American Chemical Society.](image_url)
Chapter 4 Applications in Atmospheric Chemistry

Prediction of climate change on Earth is of much importance to humanity and the environment of the planet. However, in global climate model projections, the indirect effect of aerosol particles on cloud reflectivity is associated with large uncertainties \[^85]\). The airborne water-soluble particles, known as cloud condensation nuclei (CCN) \[^86]\), are crucial factors for cloud droplet growth and activation that indirectly affect the optical properties of the clouds. Relevant studies have been carried out on the chemical composition, concentration, activation, surface tension, density, and hydrophilicity of CCNs \[^87]\ \[^88]\ \[^89]\ \[^90]\ \[^91]\).

The Köhler theory is a well-established theory that describes the ability of droplet nucleation and the equilibrium growth of cloud droplets \[^92]\). It is composed of two competing effects, the Kelvin effect and the Raoult effect, resulting in a maximal supersaturation known as the critical supersaturation at a certain radius of the wet particle. In experimental measurements, the fraction of activated particles can be monitored as a function of supersaturation, which is then fitted to a proper curve such as the error function to give the critical supersaturation \[^91]\). Therefore the Köhler equation provides a convenient approach to predict the critical supersaturation which can be compared with experimental observation. The Köhler equation was originally derived for inorganic salts, and has been extended to include organic components. A useful formula of the Köhler theory is expressed as \[^93]\ \[^94]\)

\[
s \equiv \frac{p}{p_0} = a_w \exp \left( \frac{4M_w \sigma}{RT \rho D_p} \right) \tag{4.1}
\]

where \(s\) is the saturation ratio, \(p\) is the vapor pressure of water over a droplet, \(p_0\) is the water vapor pressure over a flat surface of pure water, \(a_w\) is the water activity of the solution, \(M_w\) is the molar mass of water, \(\sigma\) is the surface tension, \(R\) is the gas constant, \(T\) is the temperature, \(\rho\) is the density of the solution, and \(D_p\) is the droplet diameter. The water activity is calculated as

\[
a_w = \frac{n_w}{n_w + \sum_i v_i n_i} \tag{4.2}
\]

where \(n_w\) is the number of moles of water, \(v_i\) and \(n_i\) are the van’t Hoff factor and the number of moles of the \(i\)-th solute molecule, respectively.

The Kelvin term, that is, the exponential term in Equation (4.1), arises due to the curvature of the spherical droplet. The surface tension appearing in the Kelvin term is a factor that has
observable effect on the saturation ratio and consequently the critical supersaturation. Experimental measurements have shown that the presence of atmospheric surfactants, or surface-active organic compounds, is able to decrease the surface tension of water droplets up to 30% \[^95\]. Through simple calculations based on the Köhler equation, atmospheric surfactants have been estimated to affect the optical properties of clouds and hence Earth’s climate \[^95\]. A lower critical supersaturation due to depressed surface tension leads to more possible CCNs for cloud droplet activation, which in turn give rise to more, but smaller cloud droplets, making the cloud whiter with higher reflectivity. To improve the accuracy and availability of the Köhler equation, empirical parameterizations of the surface tension based on experimental measurements have also been introduced \[^89\].

Theoretical evaluation of the surface tension serves as a powerful and useful supplement to experiments, especially when investigating nano-sized systems, by providing insight into the solute-solvent structures and the origin of the change of surface tension. Particularly, in MD simulations the statistical behavior of the molecules is taken into account and the surface tension can be evaluated as an ensemble average. For planar liquid-gas interfaces it is straightforward to compute the surface tension from the virial pressure tensor \[^96\]

\[\sigma_o = \frac{1}{2} L_z \left[ P_{zz} - \frac{1}{2} (P_{xx} + P_{yy}) \right] \tag{4.3}\]

where \(P_{xx}, P_{yy}\) and \(P_{zz}\) are the diagonal components of the pressure tensor and \(L_z\) is the box length along the \(z\)-axis. The long-range correction due to the use of truncated Lennard-Jones potential is given as \[^97\]

\[\sigma_d = 12\pi\varepsilon \sigma^6 (\rho_L - \rho_V)^2 \int_0^\infty \int_0^{\infty} dr dr' (3s^3 - s) r^{-3} \coth \left( \frac{rs}{\xi} \right) \tag{4.4}\]

where \(\varepsilon\) and \(\sigma\) are Lennard-Jones parameters, \(\rho_L\) and \(\rho_V\) are number densities of the liquid phase and the gas phase, respectively, and \(\xi\) is the fitting parameter obtained by fitting the density profile to the hyperbolic tangent function.

The computation of surface tension of spherical liquid-gas interfaces is more complicated than the planar case due to the non-unique definition of the pressure tensor. One solution is to evaluate the surface tension from the Irving-Kirkwood pressure tensor \[^98\], as proposed by Thompson et al. \[^99\]. First the normal component of the Irving-Kirkwood pressure tensor is computed from the number density of the molecules and the inter-molecular forces

\[P_N(r) = P_k(r) + P_U(r) = k_B T \rho(r) + S^{-1} \sum_k f_k \tag{4.5}\]
where $k_B$ is Boltzmann constant, $T$ is temperature, $S$ is the area of the spherical surface of radius $r$, and $f_k$ is the normal component of the force between a pair of molecules acting across the surface $S$. Then the work of formation is obtained by integrating over the radius 

$$W = 2\pi \int_0^{R_\beta} \left[ P_N(r) - P_\beta \right] r^2 \, dr$$

$$= 2\pi \int_0^{R_\beta} P_N(r) r^2 \, dr - \frac{2\pi}{3} P_\beta R_\beta^3$$

(4.6)

where $R_\beta$ is the radius of a sphere in the vapor region and $P_\beta = P_N(R_\beta)$ is the vapor pressure. Finally the effective surface tension is obtained as

$$\sigma = \tilde{\gamma}_e = \frac{3W}{4\pi R_e^2}$$

(4.7)

where $R_e$ is obtained by fitting the density profile to the hyperbolic tangent function or the error function. Here the effective surface tension $\tilde{\gamma}_e$ is adopted as an approximation to the real surface tension. It has been shown that by choosing the equimolar surface as the dividing surface the effective surface tension $\tilde{\gamma}_e$ is equal to $(\gamma_e - R_e d\gamma_e/dR_e)$ and serves as a good approximation to $\gamma_e$ at large $R_e$, especially at the sizes of CCNs which are of interest in atmospheric studies.

4.1 cis-Pinonic Acid in Atmospheric Droplets (Paper V)

Atmospheric surfactants or organic surface-active compounds are important components in CCNs and are able to decrease the surface tension of aerosol droplets. The presence of atmospheric surfactants can result in significant effects on the critical supersaturation and the number concentration of CCNs, and hence the optical properties of the clouds. One of the identified surface-active organic compounds in atmospheric droplets is cis-pinonic acid (CPA), the oxidation product of $\alpha$-pinene, with the ability to depress the surface tension of water even at very low concentrations. In paper V, we studied the behavior of CPA in atmospheric droplets by carrying out MD simulations and focused on the issue how CPA molecules can decrease the surface tension of droplets. The structure of CPA is shown in Figure 4.1. It comprises a hydrophobic four-membered carbon ring, a hydrophilic carboxylic acid group and a less hydrophilic carbonyl group.

The OPLS force field was adopted to describe CPA molecules in atmospheric droplets. This force field is suitable for simulations of small organic molecules and is expected to give accurate solvation free energies. Therefore, the behavior of CPA can be properly simulated by using the OPLS force field. The SPC/E water model was used to
describe the water molecules, since this water model is simple and able to give better surface tension than several other water models\(^{103}\). To simulate a droplet in a periodic unit cell, we first constructed an equilibrated cubic simulation box filled with a desired number of solute and solvent molecules, and then enlarged the box in three dimensions to create the liquid-gas interface. Here the gas phase is represented by vacuum for simplicity. After further equilibration in the NVT ensemble, the initial cubic cluster of molecules spontaneously becomes a spherical cluster under its surface tension, as illustrated in Figure 4.2.

![Figure 4.1. Structure of cis-pinonic acid.](image1)

![Figure 4.2. Construction of a spherical cluster in a periodic unit cell.](image2)

![Figure 4.3. (left) Snapshot of a spherical cluster consisting of 144 CPA molecules and 2000 water molecules. (right) Radial number densities of CPA and water molecules in this cluster. Reprinted with permission from the American Chemical Society.](image3)
A total number of eighteen systems were constructed, consisting of different numbers of CPA and water molecules. Each system was set up according to the above procedure, then energy minimized and subject to a further 3 ns simulation, from which the last 1 ns trajectory was extracted for analysis. The NVT ensemble was employed; the temperature was maintained at 286 K by the Nosé-Hoover thermostat to mimic the atmospheric temperature. During the simulations, the CPA molecules were found to spontaneously concentrate on the droplet surface. A snapshot of the cluster consisting of 144 CPA and 2000 water molecules is shown in Figure 4.3. The radial number densities of water and CPA molecules are also plotted.

The normal component of the Irving-Kirkwood pressure tensor, $P_N(r)$, was calculated in each droplet. Figure 4.4 shows the $P_N(r)$ and $P_N(r)r^2$ curves in the clusters containing 2000 water molecules and different number of CPA molecules. It can be seen that the presence of CPA results in a decrease in the $P_N(r)$ curve, both at the surface ($r \approx 2.2$ nm) and inside the droplet ($r \approx 1.5$ nm). From Equation (4.6) we see that the work of formation is computed from the integration of $P_N(r)r^2$ over $r$, therefore the $P_N(r)r^2$ curves in Figure 4.4 provide a more intuitive picture. For pure water droplet the $P_N(r)r^2$ curve shows a distinguished peak at the surface and a shoulder peak close to the surface, corresponding to the two rightmost peaks in the $P_N(r)$ curve; other $P_N(r)$ peaks at small $r$ become inconspicuous after multiplication with $r^2$. As the number of CPA molecules increases, the height of the $P_N(r)r^2$ curve decreases at both peaks and results in a lower surface tension of the droplet.

![Figure 4.4](image)

Figure 4.4. (left) Normal component of Irving-Kirkwood pressure tensor ($P_N(r)$) in the clusters consisting of 2000 water molecules and different number of CPA molecules. (right) $P_N(r)$ multiplied by $r^2$ in each cluster.

Finally the computed surface tension in droplets with different size were collected and illustrated in Figure 4.5 in terms of the Langmuir-Szyszkowski equation

$$\frac{\Delta \sigma}{\sigma} = - \frac{\Gamma_{\max}}{\sigma} RT \ln(1 + KC) \quad (4.8)$$

where $\Gamma_{\max}$ is the maximal surface excess concentration of the surfactant, $\sigma$ is the surface tension of the pure water droplet, $\Delta \sigma$ is the surfactant-induced change of surface tension, $K$ is
the Langmuir equilibrium constant, and $C$ is the concentration of the surfactant, which is CPA in this case. From Figure 4.5 it can be seen that the computed Langmuir-Szyszkowski curves are in similar shape to the experimental result reported by Shulman et al. [87], and that the decrease in surface tension is dependent on both the size of the droplet and the concentration of the organic compound. As the droplet grows larger, the computed curve is shifted leftward, approaching the experimental data. In short, theoretical MD simulations have shown that the CPA molecules are inhomogeneously distributed in atmospheric droplets and that the surface tension of the droplet varies with respect to both concentration and size.

![Figure 4.5. Surface tension reduction induced by CPA molecules in (red line and dots) the cluster containing 1000 water molecules, (green line and dots) the cluster containing 2000 water molecules, (blue line and dots) the cluster containing 5000 water molecules, and (black line and dots) the experimental data. Reprinted with permission from the American Chemical Society.](image)

4.2 Humic-Like Substances in Atmospheric Droplets (Paper VI)

Observations of airborne aerosol organic mass in both urban and boreal wood areas have identified several surface-active compounds, where humic-like substances (HULIS) are found abundant. One of the oxidation products of $\alpha$-pinene evaporating from trees, CPA, is an organic acid belonging to the group of HULIS. Surprisingly CPA has also been detected in Polar Regions where there is little vegetation [105], indicating long range transport or local/regional marine sources of terpenes. Closely related to CPA are two potentially surface-active compounds: pinic acid (PAD) and pinonaldehyde (PAL). The chemical structures of CPA, PAD, and PAL are shown in Figure 4.6. In paper VI, we use MD techniques to investigate surface tension for clusters consisting of water and surfactants (CPA, PAD and PAL) of atmospheric relevance. In the case of CPA a direct comparison is made with the experimental data by Shulman et al. [87], while for PAD and PAL there is yet no experimental measurement to our knowledge. Previous studies by Zakharov et al. [101] and Chakraborty...
et al. \cite{106} have employed MD simulations to determine surface tension for droplets containing water and fatty acids; the HULIS compounds in this study are expected to make a good model for other surfactants present in the troposphere.

![Chemical structures of CPA, PAD and PAL.](image)

**Figure 4.6.** Chemical structures of CPA, PAD and PAL.

![Radial number densities of water and solute molecules in the droplet containing (left) 162 PAL molecules and (right) 162 CPA molecules. Green lines are the densities amplified ten times to facilitate comparison.](image)

**Figure 4.7.** Radial number densities of water and solute molecules in the droplet containing (left) 162 PAL molecules and (right) 162 CPA molecules. Green lines are the densities amplified ten times to facilitate comparison.

The low solubility of CPA (3.7 g/L \cite{107}) indicates that the CPA molecules may not dissolve in water, but rather aggregate as a result of the hydrophobic effect, driven by the increase of entropy in the system. In paper V it has been shown that the CPA molecules moved to the surface in clusters containing 1000, 2000 and 5000 water molecules. In larger clusters with 10000 water molecules there is enough space or volume for the HULIS molecules to have the chance to aggregate by hiding their hydrophobic parts inside and reducing the number of surrounding water molecules per HULIS molecule. Figure 4.7 shows the radial number density curves of solute molecules in the droplets containing 162 PAL and 162 CPA molecules, respectively. Both droplets contain 10000 water molecules as solvent. The PAL molecules are not centered in the droplet, but rather concentrated at the surface, contributing to the radial number density peak at the radius of around 4.2 nm. For CPA molecules it is found that part of the CPA molecules aggregates inside the droplet while the remaining CPA molecules reside at the surface. This means that the aggregates can possibly be formed even in the inner part of the cluster.
The surface tension values of droplets containing HULIS compounds were also computed and fitted to the Langmuir-Szyszkowski equation \[104\], as shown in Figure 4.8. It can be seen that the Langmuir-Szyszkowski curve for the droplet containing 10000 water molecules shows a better resemblance to the experimental data by Shulman et al. \[87\], in agreement with the trend obtained from simulations of smaller droplets. Comparison among the three HULIS compounds reveals that CPA is the strongest surfactant at low concentrations, followed by PAL and PAD. According to the Köhler equation, a decrease in surface tension will lead to a lower supersaturation for cloud droplet formation and growth, and consequently a greater number of cloud droplets and enhanced cloud reflectivity.

![Figure 4.8. Fitted Langmuir-Szyszkowski curves for (left) CPA-containing droplets with different size and (right) water droplets containing different types of solute molecules.](image)

### 4.3 Glycine in Atmospheric Droplets (Paper VII)

A large proportion of the remote oceanic aerosols have been advocated to originate from bubble bursting processes \[108\] \[109\], during which soluble organic compounds with high molecular weight could rise through the water and inject into the atmosphere \[110\] \[111\]. Proteins and bacterial enzymes have been suspected to be present in the bursting bubbles, supported by the fact that proteinaceous material has been identified in Antarctic cloud water samples \[112\], and that free and combined amino acids are enriched in marine aerosols and precipitation relative to sea water concentrations and mainly in their L-isomers \[113\]. Specifically, glycine as an amino acid has been proved to appear in aerosol particles and behave as an effective CCN material \[91\]. The activation properties of amino acid-water clusters have also been analyzed in laboratory measurements \[91\]. In paper VII, we aimed to study the curvature dependence of the surface tension of pure water droplets and water droplets containing glycine, motivated by the presence of amino acid molecules in atmospheric aerosols. The free glycine molecules in water droplets are in the zwitterionic
forms as shown in Figure 4.9. The OPLS force field \[^{44}\][^{45}\] was adopted to model the zwitterionic glycine molecules.

![A snapshot of a water droplet containing zwitterionic glycine molecules.](image1)

Figure 4.9. A snapshot of a water droplet containing zwitterionic glycine molecules.

First the curvature dependence of the surface tension of pure water droplets was studied by using the so-called semi-flexible SPC/E water model \[^{51}\], in which the two OH bonds were constrained by the LINCS algorithm \[^{28}\][^{29}\] while the HOH angle was kept flexible. This model gives better surface tension than the rigid model and is compatible with the use of a larger time step (2 fs) comparing with the flexible model. A series of water droplets with different size were constructed and simulated to give the radial number densities and normal components \((P_N(r))\) of the Irving-Kirkwood pressure tensors, as shown in Figure 4.10. The radial number densities can be fitted to the hyperbolic tangent function or the error function to give the radius of the equimolar dividing surface \[^{99}\]. The \(P_N(r)\) curves are relatively flat inside the droplet and show peaks at the surface, indicating the presence of the surface tension.

![Curves of radial number densities and normal components of the Irving-Kirkwood tensors for water droplets.](image2)

Figure 4.10. (left) Radial number densities and (right) normal components of the Irving-Kirkwood tensors of pure water droplets.

Further investigations were carried out in water droplets containing glycine molecules. As shown in Figure 4.11, the glycine molecules reside in the bulk phase of the droplet and vanish at the surface, indicating their strong hydrophilicity. Comparing with pure water droplets, the
the $P_N(r)$ curves possess similar peaks at the droplet surface, however, at smaller $r$ the $P_N(r)$ curves were perturbed by the presence of the glycine molecules. The differences shown in the $\Delta P_N(r)$ curves indicate that the glycine molecules have an undulate effect on the normal component of the Irving-Kirkwood pressure tensor.

![Figure 4.11.](image)

Figure 4.11. (left) Radial number densities and (right) normal components of the Irving-Kirkwood pressure tensors in water droplets containing glycine molecules. The $\Delta P_N(r)$ curves denote the difference between $P_N(r)$ of pure water droplets and water droplets containing glycine.

![Figure 4.12.](image)

Figure 4.12. (left) Curvature dependence of surface tension of pure water droplets and water droplets containing glycine molecules. (right) Köhler curve without and with surface tension corrections.

The surface tension of the planar liquid-gas interface was also investigated. By using the semi-flexible SPC/E water model [51] and the OPLS force field [44][45] for glycine, the change of surface tension induced by the presence of glycine was determined as $\frac{d\sigma}{dC} = 0.96 \text{ mJ m}^{-2} \text{ L mol}^{-1}$, in good agreement with experimental data (0.92 [114] and 1.12 [115]). Regarding the planar interface as the limiting case of spherical interface with infinite radius, the curvature dependence of the surface tension was derived by fitting the computed surface tension to a quadratic polynomial function of the inverse of droplet radius, as shown in Figure 4.12. At larger droplet sizes, for example the sizes of CCNs, the surface tension of pure water droplet will be enhanced by the curvature, and the presence of glycine is able to further increase the surface tension. Such behavior of surface tension is taken into account for better prediction of
the critical supersaturation by the Köhler equation. The curvature correction has improved the predicted critical supersaturation from 0.546% to 0.553%, closer to the experimental value reported by Kristensson et al. [91] which is between 0.6% and 0.7%. The correction is not so pronounced, since glycine is not as strong surfactant as cis-pinonic acid and its effect on the surface tension is much smaller. Nevertheless, our approach provides a convenient and efficient way of evaluating the curvature dependence of surface tension of aerosol droplets, which is of specific interest in the studies of the remote marine aerosol with biological origin.

4.4 Amino Acids in Atmospheric Droplets (Paper VIII)

Amino acid molecules as water-soluble organic nitrogen compounds comprise an important fraction of the atmospheric organic species. Several types of amino acids have been identified in both marine and continental atmospheric samples [116] [117] [118] [119] [120], including alanine, arginine, glycine, phenylalanine, proline, serine, valine, etc. In paper VII it has been shown that glycine has a small but non-negligible effect on the surface tension and hence the critical supersaturation of atmospheric water droplets. In paper VIII we extend the studies to the six types of amino acids shown in Figure 4.13.

![Chemical structures of the zwitterionic forms of the L-amino acids: serine (SER), glycine (GLY), alanine (ALA), valine (VAL), methionine (MET) and phenylalanine (PHE).](image)

Figure 4.13. Chemical structures of the zwitterionic forms of the L-amino acids: serine (SER), glycine (GLY), alanine (ALA), valine (VAL), methionine (MET) and phenylalanine (PHE).

For each amino acid, a droplet consisting of 50 amino acid molecules and 5000 water molecules is constructed, simulated and analyzed. The radial number densities are shown in Figure 4.14. According to the radial number density plots, these amino acids can be divided into hydrophilic and hydrophobic classes. Glycine is almost uniformly distributed in the bulk phase of the droplet, while serine has a peak density at the 0.5 < r < 1.0 region. Alanine is slightly dragged towards the surface due to the presence of the methyl group. On the contrary, valine, methionine and phenylalanine are concentrated on the surface due to the strong hydrophobicity of their side chains. From Figure 4.14 it can be seen that the hydrophobicity of these three amino acids follows the order methionine < valine < phenylalanine. Moreover, the molecular orientations are also illustrated by the radial number densities of the backbone and the side chain in serine and valine, respectively. For serine no significant preference of
molecular orientation is found; the densities of the backbone and the side chain leapfrog over each other along the radius axis, implying the thermodynamic motions of molecules inside the droplet. For the hydrophobic valine a clear molecular orientation is observed at the surface; the side chain stays on the surface with the backbone pointing inward.

![Figure 4.14](image1.png)

**Figure 4.14.** (left) Radial number densities and (right) orientations of amino acid molecules in water droplets.

![Figure 4.15](image2.png)

**Figure 4.15.** (left) Normal components of Irving-Kirkwood pressure tensors \( P_N(r) \) in water droplets containing different amino acids and (right) the \( P_N(r) \) curves for water droplets containing serine and valine in comparison with pure water droplets.

The normal components of the Irving-Kirkwood pressure tensors are shown in Figure 4.15. In each \( P_N(r) \) curve the peak at the surface indicates the presence of surface tension. A difference is found at the maximal \( P_N(r) \) value, being less than 50 kJ mol\(^{-1}\) nm\(^{-3}\) for serine, glycine and alanine and close to 55 kJ mol\(^{-1}\) nm\(^{-3}\) for valine, methionine and phenylalanine. Such a difference is expected to result in large deviation in surface tension since the work of formation is computed from the integrand \( P_N(r)r^2 \). The computed surface tension data show that the former three amino acids have small effects on the surface tension of water droplets while the latter three amino acids are able to increase the surface tension significantly. A direct comparison between two representative amino acids, the hydrophilic serine and the hydrophobic valine, is shown in Figure 4.15 together with the \( P_N(r) \) curve for pure water droplet. The \( \Delta P_N(r) \) curves show that serine has an undulate effect on the \( P_N(r) \) curve, similar to glycine as discussed in paper VII. Differently, valine gives rise to a positive peak in the
surface region of the $\Delta P_N(r)$ curve. As a result, the presence of the hydrophobic valine is expected to strengthen the surface tension of the droplet.

![Figure 4.16](image)

**Figure 4.16.** (left) Curvature dependence of surface tension of pure water droplets and water droplets containing serine and glycine. (right) Köhler curve without and with surface tension corrections for water droplets containing serine and glycine.

In the experiments carried out by Kristensson *et al.* [91], the critical supersaturations for glycine and serine have been measured and compared with the prediction from Köhler theory. It is found that the critical supersaturations are underestimated by Köhler theory; therefore we here introduce the surface tension corrections to the Köhler equation. Figure 4.16 shows the Köhler curves for dry glycine and serine particles with diameters of 50 nm. After applying the surface tension corrections to the Köhler equation, the critical supersaturations are slightly increased towards the experimental results [91]. The critical supersaturation for glycine is increased from 0.546% to 0.553%, comparing with the experimental value of ~0.65%. The critical supersaturation for serine is increased from 0.658% to 0.667%, comparing with experimental value of ~0.8%. It is clear that such corrections are relatively small, and that other factors affecting the critical supersaturation cannot be ruled out. As pointed out by Kristensson *et al.* [91], the deviation between the experimental results and the Köhler model possibly arises from several factors including surface tension, non-ideality, and uncertainty in the density of glycine. Therefore, further study is necessary to improve the Köhler model for better prediction of the critical supersaturation.
Chapter 5  Summary and Outlook

In this thesis, the MD simulation technique is employed to study several issues in the fields of solution chemistry and atmospheric chemistry. In all cases water molecules constitute a large part of the systems and play an important role. We find that MD simulations provide a powerful statistical and numerical tool to investigate the physicochemical processes taking place at the atomistic scale in these systems.

The applications in this thesis of MD simulations in solution chemistry involve four aspects. First, the relativistic effect of the Cd(II) ion on the $^{113}$Cd NMR shielding constant is studied through successive MD simulations and quantum chemical calculations. It is revealed that the relativistic contribution is stable with respect to the chemical environment. Second, the paramagnetic $^{19}$F NMR shift induced by triplet O$_2$ is investigated through the same approach. The downfield shift of several ppm is reproduced and the geometrical dependence of the paramagnetic shift is discussed. Third, the hydration free energies and structures of metal ions are calculated using a polarizable water model. A fitting-refining procedure is suggested to optimize the parameters for satisfactory reproduction of static and dynamic properties of the metal ions. Fourth, the adsorption behavior of proteins onto hydroxylated and non-hydroxylated TiO$_2$ surfaces is studied. The interfacial water molecules are found to mediate the adsorption, and surface modification is expected to significantly affect the adsorption process. Through these applications it is shown that MD simulations are versatile and useful in the modeling of liquids and interfaces. Classical all-atom force fields provide a satisfactory description of the inter- and intra-molecular interactions at an economical computational cost, while quantum chemical calculations are able to derive proper parameters for the solvated ions or molecules. The combination of MD simulations and quantum chemical calculations utilizes the advantages of the two techniques and gives insight into both conformational space and electronic structures.

The applications of MD simulations in atmospheric chemistry are mainly focused on the investigation of the physical properties of water droplets containing humic-like substances and amino acids. The droplets remain spherical under the presence of surface tension; the radius of the equimolar dividing surface can be determined by fitting the radial number density curves to the hyperbolic tangent function or the error function. Through the Irving-Kirkwood pressure tensor the work of formation and the surface tension can be computed and analyzed to reveal the roles of the organic components present in the droplets. It is found that the surface-active cis-pinonic acid molecule is able to decrease the surface tension, its effect being dependent on both the size of the droplet and the concentration of the solute. In this regard, the partition of aerosol organic components between the surface and the bulk phases
of the droplet should be taken into account for better understanding of the atmospheric processes of droplet formation and growth. Differently, the presence of glycine can increase the surface tension of large droplets while decrease the surface tension of small droplets. According to the Köhler equation, a perturbation in the surface tension will put an effect on the critical supersaturation of the droplets. For very large atmospheric droplets with diameter of hundreds of nanometers, the critical supersaturation will be enhanced by the curvature dependence of the surface tension and the presence of glycine molecules. Such results are in qualitative agreement with experimental measurements and show the usefulness of MD simulations in atmospheric research.

In summary, MD simulations provide a powerful and versatile technique that allows us to study the statistical and dynamical properties of complex systems at the molecular level. It has been shown in this thesis that the MD technique is capable of providing insight into various chemical and biochemical species ranging from metal ions, amino acids and small organic molecules to macromolecules like proteins. The disadvantage is that, at the current stage, the time scale of MD simulations is usually limited to hundreds of nanoseconds or microseconds, which may not be sufficient to fully sample the rare events or to completely explore the free energy landscape. Therefore the development of faster and smarter algorithms is of great importance in future research of MD simulations, for example, the implicit solvent model and the metadynamics techniques have already shown their power in efficient sampling. From another aspect, the coarse-grained MD simulations provide the opportunity to study extremely large systems or to run simulations on much longer time scale, which is a big step from the atomistic world towards the mesoscopic world. Besides, the combination of quantum mechanics and molecular mechanics provides superior accuracy and capability over classical force fields, making it possible to deal with complex processes such as excitation, coordination, electron transitions and chemical reactions that take place in living systems as well as in the atmosphere. With the fast development of computer hardware and of new emerging algorithms, the computational accuracy and efficiency of MD simulations will be pushed to new frontiers in the future. MD simulations will certainly constitute an important part in both academic and industrial research in the coming era of computer modeling.
References


References


References


References


References


Zhang, Q. and Anastasio, C. (2003). Free and combined amino compounds in atmospheric fine particles (PM2.5) and fog waters from Northern California. *Atmos. Environ.*, Vol. 37, pp. 2247–2258.