Exploration of the Chemical and Biological Processes through Molecular Dynamics Simulation and Blocklocalized Wavefunction Method

Yuchun Lin
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations

Part of the Chemistry Commons

Recommended Citation
Lin, Yuchun, "Exploration of the Chemical and Biological Processes through Molecular Dynamics Simulation and Blocklocalized Wavefunction Method" (2008). Dissertations. 788.
http://scholarworks.wmich.edu/dissertations/788

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
EXPLORATION OF THE CHEMICAL AND BIOLOGICAL PROCESSES THROUGH MOLECULAR DYNAMICS SIMULATION AND BLOCK-LOCALIZED WAVEFUNCTION METHOD

by

Yuchun Lin

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Chemistry

Western Michigan University
Kalamazoo, Michigan
June 2008
ACKNOWLEDGMENTS

There are so many people I would like to thank. First of all, I would greatly like to thank Dr. Yirong Mo for having me work in his lab. His advice, support, and patience throughout the course of my studies are genuinely appreciated. He has given me the wonderful opportunity to work on several exciting projects, and has truly made graduate school a wonderful experience. I will truly miss working in his lab, but I know his guidance during the past few years have prepared me for a good future.

I would also like to thank Dr. John Miller, Dr. Donald Schreiber, Dr. Susan Stapleton, and Dr. Brian Tripp for being on my thesis committee and taking time to review my work. I appreciate their guidance and help towards the completion of my dissertation.

I was very fortunate to be in such a friendly and collaborative department. The outstanding teachers throughout my educational career made me different. I would also like to thank Pam McCartney and Annie Dobbs for their nice help all the time when I am in Department of Chemistry.

I would like to thank past and present members of Dr. Mo’s research group. I thank Minghong Liu for her useful discussion on the research and nice friendship. I thank Lingchun Song, who has been a great colleague and mentor. Kazuhito Nakashima, Theresa Obias, Mingli Xiang, and Xin Zhang have been a pleasure to work with. I also would like to thank my friends with whom I have shared lots of great memories in WMU and learned a lot from them. Special thanks to Jie Wang, Chun Wang, Son Duc Tran, Raymond Sung, and Donald Toppen, who have always been there for me when in need.
Acknowledgments—continued

The Keck Foundation, the National Institute of Health, and Dissertation Completion Fellowship from Western Michigan University are acknowledged for financial support of this work.

Last but definitely not least, I would like to thank my parents, sisters and brother in China for their constant encouragements, considerations and supports. To my wonderful wife, Wen Guo, whom I love very much and whose love and support have help me achieve a lot: the life we are sharing is my greatest happiness. This dissertation is dedicated to you, my love. Without you, nothing would be possible.

Yuchun Lin
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... ii

LIST OF TABLES .................................................................................................................. viii

LIST OF FIGURES ............................................................................................................... ix

LIST OF SCHEMES .............................................................................................................. xiv

CHAPTER

I. OVERVIEW ...................................................................................................................... 1

II. METHODOLOGIES ON BIOLOGICAL SYSTEM ......................................................... 8

2.1 Molecular Dynamics ...................................................................................................... 9

   2.1.1 Newtonian Equation of Motion ............................................................................ 10

   2.1.2 Force Fields ........................................................................................................... 13

   2.1.3 Non-bonded Calculations .................................................................................... 16

   2.1.4 Boundary Treatments ......................................................................................... 20

   2.1.5 Solvent Models .................................................................................................... 23

   2.1.6 Temperature and Pressure Regulations ............................................................... 24

   2.1.7 Free Energy Calculation ...................................................................................... 26

   2.1.8 Limitations of MD Methods ............................................................................... 28

2.2 Quantum Mechanics Methods .................................................................................... 29

   2.2.1 Ab initio Methods ................................................................................................ 31

   2.2.2 Semi-empirical Methods ..................................................................................... 32

   2.2.3 Density Functional Theory Methods ................................................................... 33

   2.2.4 Limitations of QM Methods ............................................................................... 35
Table of Contents—continued

CHAPTER

2.3 Hybrid QM/MM Method .................................................. 35

2.3.1 Basic Idea of QM/MM Method .................................. 35

2.3.2 Implementation of QM/MM Method .......................... 37

III. MOLECULAR DYNAMICS SIMULATIONS ON THE
ESCHERICHIA COLI AMMONIA/AMMONIUM TRANSPORTER
PROTEIN AMTB ...................................................... 42

3.1 Introduction and Background ....................................... 42

3.1.1 Specific Role of Ammonium and the Ammonium
Transporter .......................................................... 42

3.1.2 Research on Amt Structure ................................... 47

3.1.3 Computational Studies on AmtB ............................... 54

3.1.4 Mechanistic Questions to Be Addressed .................... 55

3.2 Theoretical Background and Computational Details ......... 58

3.2.1 Computational Model Setup .................................. 58

3.2.2 Coordinates for Energy Profiles .............................. 67

3.2.3 Computational Details ........................................ 71

3.3 Recruitment of NH\textsubscript{4}\textsuperscript{+} to the Binding Vestibule .......... 74

3.4 Gating Mechanism .................................................... 81

3.5 Functional Role of Asp160 .......................................... 83

3.6 QM/MM Simulation on the Deprotonation of Ammonium .... 94

3.7 Continuous Conduction of Neutral NH\textsubscript{3} in the Channel .... 108

3.8 Selectivity of the Ammonia-Conducting Channel toward
Water Molecules ...................................................... 113
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9 Summary and Future Plan</td>
<td>114</td>
</tr>
<tr>
<td>IV. MOLECULAR DYNAMICS SIMULATION ON THE PHOSPHOTRIESTERASE (PTE): DOCKING PROCESS</td>
<td>118</td>
</tr>
<tr>
<td>4.1 Background of Phosphotriesterase (PTE)</td>
<td>118</td>
</tr>
<tr>
<td>4.1.1 Overview of Phosphotriesterase (PTE)</td>
<td>118</td>
</tr>
<tr>
<td>4.1.2 Experimental Studies on Phosphotriesterase (PTE)</td>
<td>120</td>
</tr>
<tr>
<td>4.1.3 Computational Studies on Phosphotriesterase (PTE)</td>
<td>124</td>
</tr>
<tr>
<td>4.1.4 Our Current Objectives</td>
<td>125</td>
</tr>
<tr>
<td>4.2 Computational Model and Computational Detail</td>
<td>127</td>
</tr>
<tr>
<td>4.2.1 QM/MM Model Setup</td>
<td>127</td>
</tr>
<tr>
<td>4.2.2 Computational Detail</td>
<td>128</td>
</tr>
<tr>
<td>4.3 Results and Discussion</td>
<td>130</td>
</tr>
<tr>
<td>4.4 Summary and Future Plan</td>
<td>135</td>
</tr>
</tbody>
</table>

| V. DEVELOPMENT AND APPLICATIONS OF BLOCK-LOCALIZED WAVEFUNCTION (BLW) METHOD | 137  |
| 5.1 Introduction                                                         | 137  |
| 5.2 Methodology                                                          | 139  |
| 5.2.1 Ab initio Valence Bond (VB) Theory                                 | 139  |
| 5.2.2 Block-Localized Wavefunction (BLW) Method                          | 141  |
| 5.2.3 BLW Method at the Density Functional Theory (DFT) Level            | 145  |
| 5.2.4 Energy Decomposition Analysis Based on BLW (BLW-ED)               | 147  |
CHAPTER

5.3 Applications of BLW Method......................................................... 149

5.3.1 π-Cation Interactions between Cations and Benzene........... 150

5.3.2 Charge Transfer in the Solvation of MeₙNH₄ⁿ⁺ (n = 0, 1, 2, and 3) ................................................................. 156

5.3.3 Theoretical Study of the Interchain Conductivity in Poly(p-phenylene) ............................................................... 162

5.4 Summary....................................................................................... 170

APPENDIX

A. Acronym Glossary ........................................................................ 172

BIBLIOGRAPHY .................................................................................. 176
LIST OF TABLES

5.1. BLW-DFT energy decomposition analyses on the π-cation interactions with the 6-311G(d,p) basis set (kcal/mol) ................................................................. 152

5.2. BLW-DFT energy decomposition analyses on the interaction between MeₙNH₄ₙ⁺ and water with the 6-311+G(d,p) basis set (kcal/mol) ......................... 158

5.3. Individual polarization energies (kcal/mol) .......................................................... 159

5.4. Amount of electrons transferred from water to MeₙNH₄ₙ⁺ based on various population analysis schemes ................................................................. 161

5.5. Optimal bond lengths in benzene, benzene cationic radical and benzene anionic radical ................................................................................................. 165

5.6. Coupling energy $V_{AB}$, charge transfer barrier $\Delta E_a$, and reorganization energy $\lambda$ (in eV) at several distances when the donor and acceptor groups are weakly coupled ........................................................................ 170
LIST OF FIGURES

2.1. Scheme of cutoff for non-bonded calculation.............................................. 18
2.2. Illustration of periodic boundary conditions............................................. 21
2.3. Illustration of the combined QM/MM method in protein system............... 36
2.4. Comparison of QM/MM partitioning schemes across covalent bonds......... 41

3.1. Sketch map for the membrane system of Ammonia/Ammonium channel.............................................................. 45
3.2. Topology model for Amt proteins............................................................. 47
3.3. The ammonia-conducting portion of AmtB (1U7G structure) after removal of portions of helices M8, M9, and M10.......................... 50
3.4. A ribbon representation of AmtB viewed from the extracellular surface demonstrates the trimeric structure of AmtB............................ 53
3.5. AmtB membrane protein model (unitary cell) for simulations from side view, and the inset is the top view of the AmtB trimer........... 59
3.7. Final AmtB membrane protein model (unit cell) for simulations (Model A), where the white arrow highlights the ammonia channel and the ribbon structure is shown in yellow.................................................. 63
3.8. The orientations of Asp160 in AmtB (in cyan) and Ala160 in the D160A mutant (in magenta) after superimposing these two protein structures........................................................................................................ 65
3.9. Snapshot showing the surrounding for the ammonium ion in the extracellular binding site ($Rc1 = 15.8 \text{ Å}$)......................................................... 66
3.10. QM part model for deprotonation process calculation............................ 67
List of Figures—continued

3.11. The ammonia-conducting channel in AmtB with one ammonium (Am) outside the channel and three ammonia molecules (Am2, Am3, and Am4) inside the channel in the crystallographic structure [33] .......................... 68

3.12. A snapshot for the initial configuration of an ammonium ion outside the AmtB membrane protein, where the ion is fully solvated (Rc1 = -20.2 Å) .......................... 74

3.13. PMF profiles for the passage of NH3/NH4+ across the AmtB membrane protein ............................................................................................................................. 75

3.14. Snapshots illustrating the surroundings for the ammonium ion in the first binding region: (a) just before reaching the first binding site (Rc1 = -17.8 Å); (b) in the periplasmic recruitment (Rc1 = -16.4 Å). Unit of distances is Angstrom ............................................................................................................................. 77

3.15. Snapshots illustrating the surroundings for the ammonium ion in the second binding region: (a) in the first part of the second binding region (Rc1 = 11.8 Å); (b) in the second part of the second binding region (Rc1 = -9.7 Å). Unit of distances is Angstrom ............................................................................................................................. 79

3.16. The dynamic opening of the gates by (a) the ammonium ion (superimposed snapshots at Rc1) -20.2, -19.3, -17.8, -16.8, -16.3, -14.5, -13.2, -12.0, -10.9, -9.6, -8.1, and -6.7 Å in color from black to magenta); (b) the ammonia (superimposed snapshots at Rc1) -15.6, -14.1, -13.0, -11.3, -9.9, -8.7, -8.0, -7.4, -6.5, and -5.7 Å in color from black to purple) ............................................................................................................................. 82

3.17. The superimposition of two snapshots, where NH4+ is in the bulk solvent (Rc1 = -20.2 Å, in purple) and in the first binding region (Rc1 = -15.8 Å, in multi-colors). ............................................................................................................................. 86

3.18. Potential of mean force profiles for the translocation of ammonium across the AmtB (in red) and the D160A mutant (in blue) proteins ......................................................... 87

3.19. Snapshot showing the surrounding for the ammonium ion in the extracellular binding site (Rc1 = -15.7 Å) in D160A mutant ................................................................. 88

3.20. Snapshots illustrating the surroundings for the ammonium ion (a) in the second binding region of the native AmtB (Rc1 = -11.3 Å) and (b) in a similar position in the D160A mutant (Rc1 = -10.8 Å) ............................................................................................................................. 90
List of Figures—continued

3.21. QM calculations on the binding energy difference due to the carboxyl group .......................................................... 91

3.22. QM calculations on the impact of the negatively charged carboxyl group of Asp160 on the hydrogen bonding between NH$_4^+$ and the backbone carbonyl group of Ala162 .......................................................... 92

3.23. PMF of proton transfer from water molecule W2 to carboxylate group of Asp160 .................................................................................................................. 97

3.24. QM part at $Rc' = 0.48$ Å in the simulation for proton transfer from H$_2$O to carboxylate group of Asp160 .......................................................... 98

3.25. PMF of proton transfer from NH$_4^+$ to the nearest water molecule W1 ...... 98

3.26. QM part structures of proton transfer process ........................................... 100

3.27. Snapshot of AmtB in proton transfer process. It is the initial structure where $Rc2 = -0.80$ Å .......................................................... 101

3.28. Snapshot of AmtB in proton transfer process. It is the structure around the transition state, where $Rc2 = 0.12$ Å .......................................................... 102

3.29. Snapshot of AmtB in proton transfer process. It is the structure around the transition state, where $Rc2 = -0.32$ Å .......................................................... 103

3.30. Snapshot of AmtB in proton transfer process. It is the product structure where $Rc2 = 0.70$ Å .......................................................... 104

3.31. Trajectory of three hydrogen bonds (O1···H1, O2···H2 and O3···H5) within proton transfer process .......................................................... 107

3.32. Snapshots before (in purple) and after (in multicolor) the deprotonation of ammonium ion at the binding vestibule ($Rc1 = -15.8$ Å). The transferred proton is highlighted with a circle in orange .......................................................... 109

3.33. Snapshot of the passage of ammonia alone through the hydrophobic channel of AmtB ($Rc1 = -9$ Å) .......................................................... 111

3.34. Snapshots of the passage of ammonia alone (a) or coupled with a water molecule (b) through the hydrophobic channel of AmtB ($Rc1 = -8.1$ Å) ...... 112
List of Figures—continued

3.35. Potential of mean force (PMF) profile for the passage of a water molecule across the AmtB membrane protein ........................................... 114

3.36. Illustration of the ammonium/ammonia transport mechanism in AmtB. .................................. 117

4.1. A ribbon representation of PTE enzyme (PDB ID:1DPM)........................................... 122

4.2. (a) Crystal structure of PTE with bound substrate analog diethyl 4-methylbenzylphophonate (PDB ID: 1DPM); (b) The reactive site of PTE. The red circle between two zinc ions (black circles) represents a water molecule or hydroxide ion .................................................. 123

4.3. Final model for PTE in a sphere model. ................................................................. 128

4.4. Snapshot of PTE active site and substrate in a state conformation............................. 131

4.5. PMF energy profile for docking process in PTE........................................... 132

4.6. Snapshot of the docking process in PTE. The distance of (P)O···Zn2 is 2.62 Å................................................................. 133

4.7. Snapshot of the docking process in PTE, The distance of (P)O···Zn2 is 1.92 Å................................................................. 134

5.1. Geometries for the π-cation complexes: (a) M⁺(C₆H₆) (M = Li, Na, K); (b) NH₄⁺(C₆H₆); (c) N(CH₃)₄⁺(C₆H₆) .................................................................. 152

5.2. Electron density difference (EDD) maps: (a1) and (b1) show the polarization effect in the Li⁺C₆H₆ and NH₄⁺C₆H₆ complexes (isodensity 3×10⁻³ and 2×10⁻³ a.u., respectively); (a2) and (b2) show the charge transfer effect in Li⁺C₆H₆ and NH₄⁺C₆H₆ (isodensity 1×10⁻³ a.u.) .................. 155

5.3. Electron density difference (EDD) maps for the NH₄⁺(H₂O)₄ cluster model showing: (a1) the polarization effect (isodensity 3×10⁻³); (a2) the charge transfer effect (isodensity 2×10⁻¹ a.u.)........................................... 160

5.4. Correlation between the charge transfer stabilization energy and the amount of NPA charge transferred from water molecules to the cations ....... 161

5.5. Binding energy potentials between a neutral benzene and (a) an anionic benzene and (b) a cationic benzene radical .................................................. 166
List of Figures—continued

5.6. Correlation between the electronic coupling energy ($V_{AB}$ in eV) and the electron-transfer distance between a neutral benzene and (a) an anionic benzene and (b) a cationic benzene radical ............................................................... 167

5.7. Potential energy curves for the charge transfer reactions between benzene and its ionic form at the distances of 5, 6 and 7 Å ........................................... 169
LIST OF SCHEMES

3.1. Proton transfer pathway from ammonium to carboxyl group of Asp160. The hydrogen network is highlighted by the dashed red line ........................................ 95

4.1. PTE hydrolysis reaction PTE by catalyzing the cleavage of P-O and P-F bonds ....................................................................................................................... 119

4.2. Postulated reaction mechanism for the PTE catalyzed hydrolysis of paraoxon ........................................................................................................... 126

5.1. Energy decomposition scheme in our BLW-ED approach ................................................................................................................................. 149

5.2. Two parallel benzene rings used in our studies .................................................. 162
CHAPTER I

OVERVIEW

At the heart of chemistry are atoms and molecules - they are the basis set with which chemical events are expressed. As a chemist, the ability to accurately understand the behavior of atoms and molecules is of fundamental importance for exploring the world and studying the systems from small molecules to biological systems. To obtain this ability, chemists have devoted their whole lifetimes to develop and improve experimental utilities. Nowadays, breath-taking developments in novel biophysical and biochemical techniques have made it possible to monitor[1] and manipulate[2, 3] the behavior of biological systems. Remarkably, researchers are able to study the conformational dynamics of subtle system such as a single biomolecule [4-7]. Considering the complexity of life processes, there is no doubt that theoretical and computational studies will play an important role in terms of interpreting and unifying experimental observations as well as providing guidance to further experimental studies[8]. Computational studies on biological systems have drawn much attention with the advance of computational techniques. Although X-ray crystallographic structures can provide static pictures about the three-dimensional (spatial) arrangement of atoms, it cannot provide the transient conformations in the biological processes. But accumulating evidence shows that dynamics play a critical role in the functions of proteins[9, 10]. Among the breath-taking developments of biological research utilities, time-resolved X-ray crystallography has become a powerful tool to elucidate structural information at different stages along the catalytic reaction coordinates that provide valuable insight into biological processes [11].
However, this research area is considerably challenging due to the difficulty of trapping a reactive intermediate and obtaining quality crystals, as well as uncertainties due to the nature of effects that arise from the crystallization conditions. For these reasons, there is considerable interest in the development of theoretical methods that can aid in the refinement and interpretation of existing experimental data, and provide structural insight into systems where such data are not yet available. Molecular dynamics (MD) simulation, along with experimental structural data, provides a plausible venue for the characterization of biological molecular dynamics in solution and refinement of key mechanistic details. Molecular modeling methods are now routinely used to investigate the structure, dynamics and thermodynamics of inorganic, biological, and polymeric systems. MD simulations of proteins at the atomic level can reveal the dynamic behavior of proteins in biological processes and provide insights into the physical mechanisms and the physical principles underlying these processes [12-19]. The types of biological activities that have been investigated using molecular modeling include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes. [20-22]

During my Ph.D. study under the guidance of Professor Yirong Mo in the Department of Chemistry, Western Michigan University, I carried out several research projects which concern molecular modeling of biochemical systems. This dissertation work tries to explore the biological processes via computational chemistry methods, such as molecular dynamics simulation, block-localized wave-function method which has been developed by Professor Yirong Mo’s research group. Following a chapter of methodologies, there are three different parts within this dissertation: (1) Molecular dynamics simulation on the *Escherichia coli* Ammonia/Ammonium transporter
protein AmtB (Chapter III); (2) Molecular dynamics simulation on the Phosphotriesterase (PTE): Docking Process (Chapter IV); and (3) Development and applications of Block-Localized Wave-function (BLW) method (Chapter V).

Chapter III: The Amt (ammonium transporter)[23], Mep (methylammonium permease) [24-26] and Rh (Rhesus)[27-29] membrane proteins form a super family which is ubiquitous in all domains of life and responsible for the transmembrane transport of ammonium involved in the fundamental nitrogen metabolism [30]. Therefore, the transport of ammonium across cellular membranes is of high biological relevance. For instance, in bacteria and plants, ammonium is an important nutrient that is taken from surroundings such as air (convert from nitrogen) to provide a primary source of nitrogen for amino acid synthesis. For humans, ammonium is toxic at high concentration and is produced primarily in renal cells from glutamine where it plays a critical role in the maintenance of our body’s acid-base balance (pH regulation) through the regulation of renal ammonia excretion [31, 32]. The uptake and secretion of ammonium is mediated by a family of ubiquitous membrane proteins. So far, over 200 genes belonging to this Amt/Mep/Rh family have been discovered. Although there has been a controversy over the ammonium transport mechanism [33-36], it had often been thought that the species transported through the membrane proteins was ammonium, and the transport channels were therefore ion channels. However, recent high resolution X-ray crystallographic structures of Escherichia coli AmtB with and without ammonia or methylammonia clearly demonstrated that it is the neutral NH$_3$ rather than the positively charged NH$_4^+$ that crosses the hydrophobic channel [33-36]. As a consequence, there is a need to elucidate the transport mechanism, deprotonation mechanism, and substrate selectivity for the AmtB channel [37-41]. In this dissertation, computational techniques were used to explore these
unsolved issues, in order to provide new insights into the structure and functions of the AmtB channel. Our simulation results showed that the entrance of NH$_4^+$ into the periplasmic recruitment vestibule requires only 3.1 kcal/mol of energy. This is consistent with the X-ray crystallographic structure, where one NH$_4^+$ is captured in the binding vestibule [33]. In this vestibule, NH$_4^+$ loses one water of hydration, but the loss is compensated by a hydrogen bond, first with the backbone carbonyl oxygen of Phe161 then with the hydroxyl group of Ser219, as well as the stabilizing π-cation interactions with the aromatic rings of Trp148 and Phe107 in the AmtB protein. In the end of this recruitment vestibule, the phenyl ring of Phe107 dynamically switches to an open state. This is coordinated with a slight rotation and shifting of the indole ring of Trp148, which eventually creates a slot for the initially buried carboxylate group of Asp160 to become exposed to the bulk solvent. A hydrogen bond wire between NH$_4^+$ and the carboxylate group of Asp160 via two water molecules was observed. Thus, Asp160 is most likely the proton acceptor from NH$_4^+$. This explains the high conservation of Asp160 in Amt proteins and why mutants Asp160 to alanine (D160A) would completely quench the activity of AmtB [40-43]. Furthermore, convincing evidence was provided to improve the understanding of the deprotonation process of NH$_4^+$ and the specific role of Asp160. In particular, to evaluate the functional role of Asp160 versus Ala160, the potential of mean force (PMF) along the reaction coordinate for both the wild-type enzyme (AmtB) and the D160A mutant were derived. And the comparison of the transport capability between the wild-type AmtB and the D160A variant was discussed. The results based on microscopic molecular dynamics simulations support the hypothesis that Asp160 plays an essential role in the function of AmtB. This special residue is the proton accepter or the driving force within the deprotonation process of NH$_4^+$. To elucidate the deprotonation mechanism,
a combined Quantum Mechanics/Molecular Mechanics (QM/MM) study of the proton transfer process was also presented in this dissertation. Once NH$_4^+$ deprotonates, the phenyl ring of Phe215 rotates to open, and the subsequent passage of NH$_3$ through the channel is straightforward. And finally, a complete story of ammonium transporter AmtB is provided.

Chapter IV: Among the various groups of pesticides, organophosphates are one of the most widely used types, accounting for about 36% of the total world market, with more than 100 compounds commonly used. Their high effectiveness has led to worldwide use; this has caused serious environmental problems. Toxic pesticide waste and chemical stockpiles pose a serious potential threat to both the environment and human health [44]. It is of great interest and significant importance to develop effective and economical methods for detoxification and removal of organophosphates from the environment. Biodegradation provides a safe and efficient way to clean residual organophosphorus pesticides from the environment, and counteract chemical nerve agents. Much attention has been given to Phosphotriesterase (PTE) as a top candidate for engineering to detoxify organophosphate nerve agents due to its two advantages: the highly catalytic properties of its wild-type as well as its applicability under various field conditions [45-52]. However, PTE does not hydrolyze all organophosphate substrates at high rates. Thus, understanding the structural origin of the enzyme activity is pivotal to engineer PTE to increase its enzymatic diversity and activity. Engineered PTE can be utilized as a decontaminant to protect populations at risk for exposure to organophosphate pesticides or chemical warfare agents, as an agricultural tool to protect beneficial predators of undesirable insects, or as a way to clean and preserve the environment. While engineering of PTE is a long term goal, in this dissertation,
our initial focus was simulation of the substrate docking process to elucidate the catalytic mechanism of PTE. This was accomplished by performing QM/MM dynamics simulations. By using umbrella sampling dynamics simulations, one Zn2 cation (second Zn cation in the active site) binding site was identified where the distance of \( O@\text{phosphate} \cdots \text{Zn2} \) is 1.92 Å. An energy barrier of 15.50 kcal/mol is observed for this docking process. At the binding site, the arrangement of the ligands of Zn2 cation changes from an approximately trigonal pyramidal arrangement to an octahedral arrangement. After the binding site, a sharp increase of energy is observed. Thus, the binding site is a good stable point for the second step simulation on the catalytic mechanism of PTE. The results will be instructive for the subsequent computer protein design (bioengineering on the PTE).

Chapter V: For biomolecular systems such as enzymes that are solvated by large numbers of explicit solvent molecules, molecular mechanics (MM) methods have enjoyed remarkable successes. For the MM methods, various force fields have been proposed and used to model the dynamic properties of biomolecules such as protein and DNA [53-55]. Currently, most available force fields assume static charge distribution for atoms. This kind of treatment may lead to the underestimation of intermolecular interaction energy for weakly bound dimers or molecular complexes, especially for the modeling of solvation of ions. Polarizable force fields can significantly improve the computational results [56-58], by introducing polarizable dipoles centered on atomic sites, or allowing charges to fluctuate based on the electronegativity equalization method in response to external electric fields [57, 58]. However there are few theoretical or experimental data to guide the parameterization of polarizable force fields. In addition, present force fields exclude the electron transfer effect, which may be important for the protein/environment interaction. Based
on the Block-Localized Wave function (BLW) method [59-61], an energy decomposition scheme (BLW-ED) which is stable with the variation of basis sets [62, 63] was proposed by Professor Mo. This method is particularly useful to study the polarization and charge-transfer effects between systems, thus may afford clues for the development and refinement of force fields. Currently, we extended the BLW method to the density functional theory (DFT) level [64]. This BLW-DFT method had been incorporated into the quantum mechanical software GAMESS [65] which can be ported to the molecular dynamics simulation software CHARMM [66]. The BLW method can not only evaluate the Pauling-Wheland resonance energy in conjugated systems [59, 67], but can also be used to explore the nature of intermolecular interactions and to decompose the interaction energy in terms of Heitler-London, polarization, and electron transfer energy terms. The Heitler-London energy term can be further decomposed to electrostatic and Pauli exchange interactions [62, 63, 68-70]. After brief description on the development of BLW method and the extension to DFT level, in this dissertation, a few applications of the BLW-DFT code are presented. These applications concern the nature of \(\pi\)-cation interaction between a few cations and benzene, the charge transfer between solute and solvent with the supermolecular models of a positively charged ammonium and its methyl substitutes methylamines \(\text{Me}_n\text{NH}_4\text{H}_{4-n}^+\) (\(n = 0-3\)) plus a few water molecules surrounding each cation, and the theoretical study of the interchain conductivity in Poly(\(p\)-phenylene) (PPP).
CHAPTER II

METHODOLOGIES ON BIOLOGICAL SYSTEM

Computational methods have become a fundamental research tool for scientists from a wide range fields including biology, physics and chemistry and played a central role to bridge theoretical and experimental studies [12-19]. Generally, for studying biological systems, computational methods can be broadly classified into molecular mechanics (MM) and quantum mechanics (QM) approaches. The common feature of MM techniques is the atomistic level description of the molecular systems, in other words, the lowest level of information is individual atoms (or a small group of atoms). Thus, effects of electron are excluded in MM methods. This is in contrast to QM approaches (also known as electronic structure calculations) where electrons are considered explicitly. The benefit of MM techniques is that they reduce the complexity of the system, allowing many more particles (atoms) to be considered during simulations. On the other hand, QM methods model molecular systems at the more sophisticated level of electrons and nuclei by solving the Schrödinger equation. The chief advantage of the QM approach over MM methods is that it can be used to study processes that involve chemical bond breaking/forming processes. By taking advantage of both the accuracy of QM methods and the efficiency of MM methods, the development and improvement of a combined QM/MM approach is currently a very active area of research [20, 71-92]. This combined approach allows one to model a very large compound using MM methods and only the most crucial section of the molecule with QM methods.

The following in this chapter includes the background, methods, and
procedures used for computations on biological systems. The theoretical background of classical molecular dynamics simulations is described in Section 2.1; the background of QM methods is described in Section 2.2, and descriptions of the QM/MM method are presented in Section 2.3. The discussion, however, mainly focuses on the classical molecular dynamics simulations, which are the principal methods used in this dissertation work.

2.1 Molecular Dynamics

Generally, molecular dynamics (MD) and Monte Carlo (MC) methods are two typical simulation methods within molecular mechanics category [93-96]. MD simulations follow the natural time evolution of a molecular system and can be used to predict structural, dynamic and thermodynamic properties. The central idea of Monte Carlo method, which is a stochastic method, is to sample conformational space using random numbers while ensuring that the probability of occurrence of a particular formation is proportional to its Boltzmann factor. Unlike MD, the MC states sampled are not connected in time, i.e., MC is time-independent simulation. In this dissertation work, only MD simulations, rather than MC approaches, were employed and are presented here.

With the time-dependent feature, molecular dynamics provides a method for examination of molecular motion at the level of atoms and thus is a useful tool in chemistry and physics. Remarkably, MD has been well suited for simulations of biological systems, which are often hard to study experimentally due to the short time and distance scales associated with molecular motion and structures [93-101].

In practice, the general steps for a molecular dynamics simulation of an equilibrium system are as follows:
1. Choose initial positions for the atoms. In the case of biomolecules, there are usually obtained from crystal structures deposited in the RCSB Protein Data Bank or Rutgers Nucleic Acid Database [102].

2. Choose an initial set of atom velocities. These are usually chosen to obey a Boltzmann distribution at the initial temperature.

3. Compute the forces on each atom from the energy expression. This is usually a molecular mechanics force field designed to be used in dynamical simulations.

4. Compute new positions for the atoms a short time later, called the time step. These result from a numerical integration of classical Newtonian equations of motion.

5. Compute new velocities and accelerations for the atoms.

6. Repeat steps 3 through 5. Repeat this iteration long enough for the system to reach equilibrium.

7. Once the system has reached equilibrium, begin saving the atomic coordinates every few iterations. This information is typically saved every 5 to 25 iterations. This list of coordinates over time is called a trajectory.

8. Analyze the trajectories to obtain information about the system.

2.1.1 Newtonian Equation of Motion

In a MD simulation, the motion of the system, i.e., positions, velocities, and accelerations of all the atoms and molecules, is simulated by integrating classical Newtonian equation of motion. Given a system of particles, by calculating the forces on each particle at a particular time (see Subsection 2.1.2 for force calculation), one can determine velocities and positions over time, according to Newtonian equation of motion:

\[ F = m \frac{dv}{dt} = m \frac{d^2x}{dt^2} \]  

(2.1)

where \( F \) is the force acting on a particle, \( m \) is the mass of the particle, \( a \) is its acceleration, \( v \) is the velocity, \( x \) is its position, and \( t \) is time.
In reality, the force on the particle depends on its position relative to the other particles. Therefore, it is often very difficult, if not impossible, to describe the motion analytically. Under such circumstances, the equation of motion (eq. 2.1) is generally integrated numerically by using finite difference methods, which break the integration down into small timesteps ($\Delta t$), which are usually selected based on the highest frequency motions of the system. The total force on each particle in the configuration at a time $t$ is calculated as the vector sum of its interactions with other particles. From the force, the accelerations of the particles can be calculated, and those are then combined with the positions and velocities at time $t$ to calculate the new positions and velocities at time ($t + \Delta t$). The forces on the particles in the new positions are then calculated, which gives new positions and velocities at time ($t + 2\Delta t$), and so on.

It is obvious that a useful molecular dynamics program requires a good algorithm to integrate Newtonian equations of motion. By using finite difference methods, most algorithms assume that the positions, velocities, and accelerations of the particles in a system can be approximated by a Taylor series expansion. Over years, the most widely used method is the so-called Verlet algorithm [103], which is based on two truncated Taylor series expansions, a forward expansion ($t + \Delta t$) and a backward ($t - \Delta t$) expansion:

\begin{align*}
  r(t + \Delta t) &= r(t) + v(t)\Delta t + \frac{1}{2} a(t)(\Delta t)^2 + \frac{1}{6} \frac{d^3 r(t)}{dt^3} \bigg|_{\tau = t} (\Delta t)^3 + O(\Delta t^4) \quad (2.2) \\
  r(t - \Delta t) &= r(t) - v(t)\Delta t + \frac{1}{2} a(t)(\Delta t)^2 - \frac{1}{6} \frac{d^3 r(t)}{dt^3} \bigg|_{\tau = t} (\Delta t)^3 + O(\Delta t^4) \quad (2.3)
\end{align*}

where $v$ and $a$ for the first (velocity) and second (acceleration) time derivatives of the position vector $r$, and $O(\Delta t^p)$ is the term of order $\Delta t^p$ or smaller.

These two expansions can be summed together, which gives an algorithm for propagating the positions:

\begin{align*}
  r(t + \Delta t) &= 2r(t) - r(t - \Delta t) + a(t)(\Delta t)^2 + O(\Delta t^4) \quad (2.4)
\end{align*}
This is the basic form of the Verlet algorithm [103]. Since we are integrating Newtonian equations, \( a(t) \) is just the force divided by the mass, and the force is in turn a function of the positions \( r(t) \). The Verlet algorithm is carried out in two steps: use the position at time \( t \) to calculate the force \( F \) and the acceleration at time \( t \), and then use the position at time \( t \) and the position from the previous step, \( r(t-\Delta t) \), together with the acceleration at time \( t \) to calculate the new position at \( t + \Delta t \), \( r(t + \Delta t) \). These two steps are repeated for every timestep for every particle in the system.

The Verlet scheme propagates the position vector with no reference to the particle velocities. Thus, it is particularly advantageous when the position coordinates of phase space are of more interest than the momentum coordinates, e.g., when one is interested in some property that is independent of momentum. However, often we want to control the simulation temperature. This can be accomplished by scaling the particle velocities so that the temperature remains constant (or changes in some defined manner). To propagate the position and velocity vectors in a coupled fashion, a modification of Verlet’s approach called the leapfrog algorithm [104] has been proposed. In this case, Taylor expansions of the position vector truncated at second order about \( t + \Delta t/2 \) are employed:

\[
\begin{align*}
\mathbf{r}(t + \frac{1}{2} \Delta t + \frac{1}{2} \Delta t) &= \mathbf{r}(t + \frac{1}{2} \Delta t) + \mathbf{v}(t + \frac{1}{2} \Delta t) \frac{1}{2} \Delta t + \frac{1}{2} \mathbf{a}(t + \frac{1}{2} \Delta t) \left( \frac{1}{2} \Delta t \right)^2 \\
\mathbf{r}(t + \frac{1}{2} \Delta t - \frac{1}{2} \Delta t) &= \mathbf{r}(t + \frac{1}{2} \Delta t) - \mathbf{v}(t + \frac{1}{2} \Delta t) \frac{1}{2} \Delta t + \frac{1}{2} \mathbf{a}(t + \frac{1}{2} \Delta t) \left( \frac{1}{2} \Delta t \right)^2
\end{align*}
\]

(2.5) (2.6)

When eq. 2.6 is subtracted from eq. 2.5, we get:

\[
\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{1}{2} \Delta t\right) \Delta t
\]

(2.7)

Similar expansions for \( \mathbf{v} \) give:

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

(2.8)
In the leapfrog method, positions depend on the velocities as computed one-half timestep out of phase, thus, scaling of the velocities can be accomplished to control temperature. Particularly, no force-field calculations actually take place for the fractional time steps. Forces (and thus accelerations) in eq. 2.8 are computed at integral time steps, halftime-step-forward velocities are computed therefrom, and these are then used in eq. 2.7 to update the particle positions.

There are also several other variations on the Verlet algorithm have been developed, such as velocity Verlet method [105] and Beeman's algorithm [106]. In this dissertation, the leapfrog integration is used for MD simulations.

### 2.1.2 Force Fields

In practice, a molecular dynamics simulation requires the definition of a potential function, or a description of the terms by which the particles in the simulation will interact. In chemistry and biology, this is usually referred to as a force field. Force fields are used in a variety of computer simulation techniques such as Monte Carlo simulations and energy minimizations in addition to MD simulations [107]. Most commonly used force fields in chemistry are based on molecular mechanics (MM) and embody a classical treatment of particle-particle interactions.

MM force fields are derived from empirical potential functions representative of the quantum mechanical Born-Oppenheimer potential surface. Generally, a force field can be expressed as a sum of the single atom potentials, pair atom potentials, triplet atom potentials, etc. The standard potential energy function of the MD system is represented by the bonded and non-bonded terms and is the sum of the potential energies for bond stretching, bond bending, dihedral rotations and van der Waals and electrostatic interactions:
\[ V_{total} = V_{bonded} + V_{non-bonded} \]

\[ = V_{bonds} + V_{angles} + V_{dihedrals} + V_{electrostatic} + V_{van
der Waals} \] (2.9)

The bond and angle terms are usually modeled as harmonic oscillators in force fields that do not allow bond breaking. A more realistic description of a covalent bond at higher elongation is provided by the more-expensive Morse potential. The functional form for the rest of the bonded terms is highly variable. Proper dihedral potentials are usually included. Additionally, "improper torsional" terms may be added to enforce the planarity of aromatic rings and other conjugated systems, and "cross-terms" that describe coupling of different internal variables, such as angles and bond lengths may also be added. Some force fields also include explicit terms for hydrogen bonds \[108, 109\].

The non-bonded terms are most computationally expensive because they include many more interactions per atom. A popular choice is to limit interactions to pairwise energies. The van der Waals term is usually computed with a Lennard-Jones 6-12 potential and the electrostatic term with Coulomb’s law, although both can be buffered or scaled by a constant factor to account for electronic polarizability and produce better agreement with experimental observations.

Some of the well established classical MM force fields are listed below.

**AMBER (Assisted Model Building and Energy Refinement) \[110\]:** an all atom force field which was parameterized specifically for proteins and nucleic acids. AMBER uses only five bonding and nonbonding terms along with a sophisticated electrostatic treatment. No cross terms are included.

**CHARMM (Chemistry at HARvard Macromolecular Mechanics) \[66\]:** an all atom force field which was initially designed for the simulations of proteins and nucleic acids. It also includes polarizable force fields using two approaches. One is based on the Drude shell or dispersion oscillator model \[56\]. The other is based on the fluctuating charge (FQ) model, also known as Charge Equilibration (CHEQ) \[111, 112\].
GROMOS (GROningen MOlecular Simulation) [113]: a united atom force field primarily developed for the simulations of protein-lipid and protein carbohydrate interactions.

OPLS (Optimized Potential for Liquid Simulations) [114]: an all-atom force field for condensed phase simulations of proteins and carbohydrates.

Since in this dissertation work, CHARMM all atom force fields [53] (CHARMM22 and CHARMM27) were used in the simulations, only the CHARMM force field is presented in the following discussion.

2.1.2.1 CHARMM Force Field.

The CHARMM force field is widely used for MD simulations. In CHARMM, the bonded terms (bond stretching, bond bending, dihedral rotations) can be expressed as below by a harmonic oscillator treatment:

\[
V_{\text{bonds}} = \sum_{\text{bonds}} k_r (r - r_{eq})^2 \tag{2.10}
\]

\[
V_{\text{angles}} = \sum_{\text{angles}} k_\theta (\theta - \theta_{eq})^2 \tag{2.11}
\]

\[
V_{\text{dihedrals}} = \sum_{\text{dihedrals}} k_\phi [1 + \cos (n\phi - \delta)] \tag{2.12}
\]

This includes harmonic terms for bond stretching and bending with force constants \( k_r \) and \( k_\theta \), a Fourier (or sinusoidal) torsional term with force constant \( k_\phi \), periodicity \( n \) and a phase \( \delta \). The variable \( r \) is the bond length between two atoms and \( r_{eq} \) is its equilibrium value; \( \theta \) is the angle formed by three atoms and \( \theta_{eq} \) is its equilibrium value; \( \phi \) is the dihedral angle formed by four atoms.

The non-bonded interactions include the electrostatic potential \( V_{\text{electrostatic}} \) (with the constant dielectric Coulombic interaction) and the van der Waals interaction \( V_{\text{van der Waals}} \) (with Lennard-Jones 6-12 potential):

\[
V_{\text{electrostatic}} = \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \tag{2.13}
\]
where $q_i$ is the charge of atom $i$, $r_{ij}$ is the distance between atoms $i$ and $j$, $\varepsilon$ is the dielectric constant, and $\sigma_{ij}$ is the collision diameter for atom pair $i$ and $j$ (the separation for which the energy is zero).

2.1.3 Non-bonded Calculations

The most time-consuming part of an MD simulation is the evaluation of non-bonded forces. The number of bond, angle, and dihedral terms in the force field are all proportional to the number of particles in the system (the number of internal coordinates is $3N-6$, which is linear in $N$); however, the number of non-bonded terms increases on the order of $N^2$ (non-bonded forces must be calculated between every pair of atoms in system). On the other hand, in principle, the non-bonded attractive and repulsive terms of the Lennard-Jones 6-12 potential decay quickly with distance; at $2.5 \sigma$ ($\sigma$ is the distance at which the force between the two interacting particles is zero, see eq. 2.14) the Lennard-Jones 6-12 potential has only 1% of its value at $\sigma$. Therefore, van der Waals interactions are considered short-range and can be effectively calculated using a spherical cutoff scheme (see Subsection 2.1.3.1). In contrast to van der Waals interaction, charge-charge interactions do not decay rapidly with distance; the electrostatic energy is $\sim 1/r$ for two point charges. As a result, simple truncation methods are not appropriate for the long-range electrostatic forces.

To account for the long-range contributions to the Coulombic interactions, various algorithms such as particle mesh Ewald (PME) [115-117] can be used (see Subsection 2.1.3.2). [95, 101]
2.1.3.1 Cutoff Scheme and Non-bonded Neighbor List

The most popular way of reducing computational time and effort in the calculation of non-bonded interactions is the use of non-bonded cutoff scheme (Figure 2.1). A cutoff distance is simply a radial distance about a given atom which defines a boundary for the calculation of short-range interactions. In simulations where the Lennard-Jones potential is the only non-bonded interaction, a cutoff of $2.5\sigma$ (see eq. 2.14) gives rise to a relatively small error. However, when the long-range electrostatic interactions are involved, the cutoff should be much greater because the electrostatic interactions decay much slower. Usually the cutoff distance is set to 10-14 Å for biomolecular systems. When periodic boundary conditions (PBC) are used, it is important that the cutoff distance is no more than half the shortest cell dimension (for rectangular cells). Using a larger cutoff may result in including interactions of periodic images with themselves. In the cutoff scheme, atoms that are within the cutoff distance of a particular atom will be included in the pair interaction calculations, and those outside will be ignored. The computational time saving comes from only calculating the short-range interactions for pairs that are within a cutoff distance of each other.

Clearly, a cutoff scheme would introduce a discontinuity in both the potential energy and the force near the cutoff distance (Figure 2.1), and this would result in errors. There are two commonly used algorithms, known as shifting and switching functions [101], that deal with this issue. Both methods set the non-bonded energies at zero beyond some distance value; however, the values for interactions at less than the cutoff distance are treated differently. In shifting function, a constant term is subtracted from the potential at all values within the cutoff distance. As the addition term is constant, it disappears when the potential is differentiated and so does not
affect the force calculation in MD. Therefore, there is a discontinuity in the force with the shifting potential; at the cutoff distance, the force will have a finite value which drops suddenly to zero just beyond the cutoff distance. This can introduce instabilities in a simulation. On the other hand, the switching function is a more stable approach, in which a function multiplies the van der Waals interaction and causes it (and possibly its first and second derivatives) to go smoothly to zero at the cutoff distance. This function must, of course, be equal to 1 at short distance. In this dissertation, we chose the switching function to run the MD simulations.

![Figure 2.1](image.png)

**Figure 2.1** Scheme of cutoff for non-bonded calculation. Black dots represent atoms. Distances are based on atom A. $R_{\text{cutoff}}$ is for cutoff distance, $R_{\text{nonbond list}}$ is for non-bonded atom list range. Distance between yellow and blue circles is for switching or shifting function. Reproduced from ref. 101
Even by using a cutoff scheme in MD simulations, the time to compute the number of non-bonded interactions is not dramatically reduced, because calculating all the distances takes almost as much time as calculating the energy itself. Regarding this issue, one method, called the non-bonded neighbor list method, is introduced. By using a non-bonded neighbor list, all atoms within the cutoff distance, together with all atoms that are slightly further away than the cutoff distance, are stored. Then during the simulation, only these atoms are included in the non-bonded interaction calculations. The non-bonded neighbor list, of course, will need to be updated frequently (generally about every 10~20 timesteps). The success of this approach relies on the slowly changing microscopic environment, which implies that the list of neighbors remains valid over a number of timesteps, even for relatively small cutoff distance. It is important to update the neighbor list at the correct frequency. If the update frequency is too high, the procedure is inefficient; if too low, the energies and forces may be calculated incorrectly due to atoms moving within the non-bonded cutoff region. An algorithm that can automatically update the neighbor list and so circumvent these problems is as follows: an array element is set to zero for each atom whenever the neighbor list is updated. This array is used to store the displacement of each atom or molecule in subsequent steps. When the sum of the maximum displacements of any two atoms exceeds the difference between the non-bonded cutoff distance and the neighbor list distance, then it is time to update the neighbor list again.[101]

2.1.3.2 Ewald Summation

For long-range electrostatic interactions, the Ewald summation technique can be applied. It is specific for calculating the electrostatic energy of a system using
periodic boundary conditions. It separates the potential function into short-range and long-range terms. The short-range term is treated by the truncating method discussed above. The long-range term is Fourier transformed into reciprocal space where it behaves as a short-range interaction, so it can then be easily calculated.

\[ V(r) = V_{sr}(r) + V_{lr}(r) \]  

(2.15)

where \( V_{sr}(r) \) is the short-range term that sums quickly in real space and \( V_{lr}(r) \) is the long-range term that sums quickly in Fourier space.

Nowadays, a popular variant Ewald summation technique, the particle mesh Ewald (PME) method [115-117], is available in most MD programs. The basic improvement of PME summation is to use the Fast Fourier transformation, which requires that the density field is evaluated on a discrete lattice in space, to evaluate the Fourier transfer term of the charge density field in the original Ewald summation. A full description about PME method can be found in the articles by Darden [117], Allen [97], Frenkel [100], or Schlick [95]. This dissertation used the PME for the simulations of the AmtB project.

2.1.4 Boundary Treatments

The modeling of bulk solvent is accomplished through the use of boundary conditions. The correct treatment of boundaries and boundary effects is crucial to simulation methods. There are various approaches to doing this, which can be broadly classified into periodic boundary conditions and non-periodic boundary conditions.

2.1.4.1 Periodic Boundary Conditions (PBC)

PBC allow simulations to be performed using a relatively small number of particles so that the particles in the system experience forces as if they were in bulk
solvent. When using PBC, particles are enclosed in a box (unit cell), and the simulation box is replicated to infinity by imaging the system in all three Cartesian directions, completely filling the space. During the simulation, only the coordinates of the unit cell are included. If an atom of the unit cell leaves the unit cell by crossing the boundary, then an image atom enters to replace it from the opposite side (Figure 2.2). Therefore, the number of particles within the unit cell remains constant.

![Illustration of periodic boundary conditions](image)

**Figure 2.2** Illustration of periodic boundary conditions. D atoms are crossing the boundary. Reproduced from ref. 108

The choice of geometry for the periodic cell is important. The cubic cell is the simplest periodic cell; however, it is appropriate to use a periodic cell that reflects the underlying geometry of the system. In principle, any cell shape can be used, if it can fill all the space by translation operations of the unit cell in three dimensions. Five shapes satisfy this condition: the cube (and its close relation, the parallelepiped), the
hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the ‘elongated’ dodecahedron [101]. Currently, most MD programs can simulate all these five shapes. Of the five possible shapes, the cube and the truncated octahedron have been most widely used. The formulae used to translate a particle back into the central simulation box for these shapes are simpler. Thus, it may be preferable to use one of the more common periodic cells even if there are aesthetic reasons for using an alternative. This is because the expressions for calculating the images may be difficult and inefficient to implement. Therefore, the simulations of AmtB project in this dissertation are all performed with PBC in cubic shape.

2.1.4.2 Non-periodic Boundary Methods

PBC is not always used in computer simulations. It does have disadvantages. It may cause difficulties when simulating heterogeneous systems or systems that are not at equilibrium. The other choice to simulate the boundary effect is non-periodic boundary method. This kind of methods is in contrast to PBC methods. One representative non-periodic boundary method is stochastic boundary condition. In this method, the system is divided into two regions. One region, often called the reaction zone, contains all atoms or groups within a given radius $R_j$ of the site of interest. The atoms in the reaction zone are subjected to the full simulation method. The second region (the reservoir region) contains all atoms outside the reaction zone but within a distance $R_2$ of the active site. The atoms in the reservoir region may be kept fixed in their initial positions, or may be restrained so that they stay within the shell defined by $R_1$ and $R_2$. Alternatively, they may be restrained to their initial positions using a harmonic potential. Any atoms further away from the active site than $R_2$ are discarded or may be kept fixed in their initial positions. In some case, one more region, call
buffer region, may be introduced between the reaction region and reservoir region. The atoms in buffer region are generally subjected to Langvin simulation by imposing a friction force. It is important to be aware that restraining or fixing atoms in this way may prevent natural changes occurring and so lead to artificial behavior. This method is particularly useful when we are only interested in a specific part of the solute, such as the active site of an enzyme. In the simulation of the PTE project in this dissertation, this stochastic boundary condition is chosen.

2.1.5 Solvent Models

To be realistic, biomolecular simulations have to include some sort of description of a solvent environment. Classical water models are used for the simulation of water clusters, liquid water, and aqueous solutions with explicit solvent. A common water model is to put equal positive atomic charges on the hydrogen atoms and a negative charge on the symmetry axis, or equal negative charges in the lone pair regions. Such very simple models, with careful parameterization, do remarkably well in reproducing many properties of liquid water, e.g., bulk density, heat of vaporization, compressibility, heat capacity, etc [118-123]. The most successful models along these lines, which are widely used in modern simulations, are the Transferable Intermolecular Potentials 3 and 4 Point charge water models (TIP3P and TIP4P) [118]. The similarly designed SPC (simple point charge) water model also continues to see modern use [119]. In this dissertation, the typical water model TIP3P (see description below) was chosen to simulate the explicit water solvent.

2.1.5.1 TIP3P Model

TIP3P model is now very popular in the MD simulations because of its
simplicity and computational efficiency. It simply requires parameters for new types of oxygen and hydrogen atoms to be inserted into the CHARMM energy equation. It uses a rigid geometry matching the known geometry of the water molecule and relies only on non-bonded interactions. The electrostatic interaction is modeled using Coulomb's law and the dispersion and repulsion forces using the Lennard-Jones potential. It has three interaction sites, corresponding to the three atoms of the water molecule. Each atom gets assigned a point charge, and the oxygen atom also gets the Lennard-Jones parameters. The potential for the TIP3P model is represented by

\[ V = \sum_i \sum_j \left( k_C q_i q_j \frac{1}{r_{ij}} + \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) \]  

(2.16)

where \( k_C \) is the electrostatic constant, which has a value of 332.1 Å·kcal/mol in the units commonly used in molecular modeling; \( q_i \) are the partial charges relative to the charge of the electron; \( r_{ij} \) is the distance between two atoms or charged sites; and \( A \) and \( B \) are the Lennard-Jones parameters. The charged sites may be on the atoms or on dummy sites (such as lone pairs). In most water models, the Lennard-Jones term applies only to the interaction between the oxygen atoms.

2.1.6 Temperature and Pressure Regulations

Among the different thermodynamic ensembles, NVT (constant volume and temperature; Canonical ensemble) and NPT (constant pressure and temperature; isothermal-isobaric ensemble) are the two most common ensembles in MD simulations.

The temperature of system (T) is related to the time average of the kinetic energy \( \langle K \rangle \):

\[ \langle K \rangle_{NVT} = \frac{3}{2} N k_B T \]  

(2.17)

where \( k_B \) is the Boltzmann constant, and \( N \) is the number of particles in system.
Thus, one way to control the temperature of the system is to scale the velocities [124]. If the temperature at time \( t \) is \( T(t) \), the temperature can be controlled by multiplying the velocities at each time step by the factor \( \lambda = \sqrt{T_{\text{req}}/T_{\text{curr}}} \), where \( T_{\text{curr}} \) is the current temperature as calculated from the kinetic energy and \( T_{\text{req}} \) is the desired temperature.

An alternative way to control the temperature is the Berendsen's "temperature bath" method [125]. The basic idea of this method is to couple the system to an external heat bath that is fixed at the desired temperature. Velocities are scaled at each step, such that the rate of temperature change is proportional to the difference in temperature between the bath and the system:

\[
\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{\text{bath}} - T(t))
\]  

(2.18)

where \( \tau \) is a coupling parameter that determines how tightly the bath and system are coupled together. The scaling factor for velocity is thus:

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

(2.19)

If \( \tau \) is large, then the coupling will be weak. Conversely, if it is small, the coupling will be strong. When \( \tau = \delta t \), the method is equivalent to the simple velocity scaling method.

Pressure is another quantity that often needs to be controlled during MD simulations. The simulation in isothermal-isobaric ensemble maintains constant pressure by changing the volume of the simulation cell. The volume fluctuation is related to isothermal compressibility. The methods used for pressure control are analogous to those for temperature control. The pressure can be maintained by simply scaling the volume, and the "pressure bath" method is just like the "temperature bath" method.
2.1.7 Free Energy Calculation

Calculation of free energy change is important in MD simulation. It is particularly useful to know how the free energy changes as a function of a reaction coordinate, such as the distance between two atoms or the torsion angle of a bond in a molecule. A brief introduction on the concept of the potential of mean force (PMF) is presented herein.

2.1.7.1 Potential of Mean Force (PMF)

PMF is an approach to describe how the free energy changes as a particular coordinate is varied. When the system is in a solvent, the PMF incorporates solvent effects as well as the intrinsic interaction between the two particles. When the same two particles were brought together in the gas phase, the free energy would simply be the pair potential, which has only a single minimum. But the PMF between two particles in liquid oscillates with maximum and minimum. For a given separation \( r \) between the two molecules, the PMF describes an average over all the conformations of the surrounding solvent molecules.

Conceptually, for a system with \( N \) molecules, PMF can be expressed as: \[ \begin{align*}
-\nabla_j \omega^{(n)} &= \frac{\int e^{-\frac{1}{2}k_B T \nabla_j \cdot \nabla_j} (-\nabla_j \nu) d\mathbf{q}_{n+1} \cdots d\mathbf{q}_N}{\int e^{-\frac{1}{2}k_B T \nabla_j \cdot \nabla_j} d\mathbf{q}_{n+1} \cdots d\mathbf{q}_N}
\end{align*} \]

where \( -\nabla_j \omega^{(n)} \) is the average force and therefore \( \omega^{(n)} \) is the so-called Potential of Mean Force, which is the potential that gives the average force over all the configurations of all the \( n+1 \ldots N \) molecules acting on a particle \( j \) at any fixed configuration, keeping fixed a set of molecules \( 1 \ldots n \). A particular example would be \( \omega^{(2)}(r_{12}) \) that describes the interaction between two molecules held a fixed distance \( r \) when the remaining \( N-2 \) molecules are canonically averaged over all configurations.
Practically, to describe the free energy changes as a function of a coordinate of the system in MD simulations, we can simplify the calculation of PMF. The PMF $\omega(\chi)$ along a coordinate $\chi$ is defined from the average distribution function $\langle \rho(\chi) \rangle$

$$\omega(\chi) = \omega(\chi^\star) - k_BT \ln \left[ \frac{\langle \rho(\chi) \rangle}{\langle \rho(\chi^\star) \rangle} \right]$$  \hspace{1cm} (2.21)

where $\omega(\chi^\star)$ and $\langle \rho(\chi^\star) \rangle$ are arbitrary reference values. We usually set the most probable distribution such that it corresponds to a free energy of zero. The chosen coordinated $\chi$ can be any other functionality including geometrical coordinate.

However, for processes with activation barrier higher than $k_BT$ the distribution function $\langle \rho(\chi) \rangle$ cannot be computed by a standard MD simulation. Such a computation would not converge due to under sampling of higher energy configurations (i.e., rare-events). Although PMF of enzymatic reactions can be calculated using free energy perturbation [91, 127, 128] to avoid this problem, the traditional way is to use the so-called umbrella sampling technique [129, 130]. This was the strategy chosen for this dissertation.

2.1.7.2 Umbrella Sampling Technique

As an efficient method in sampling low probability events, the umbrella sampling method is now popularly used to overcome the rare-event problem in PMF approach. In umbrella sampling, a biasing potential $W_b(\chi)$ is added to the potential energy $V(q)$:

$$V'(q) = V(q) + W_b(\chi)$$ \hspace{1cm} (2.22)

Usually, in most MD programs, the biasing potential $W_b(\chi)$ is expressed as a harmonic function:

$$W_b(\chi) = \frac{1}{2} k (\chi - \chi_0)^2$$ \hspace{1cm} (2.23)

Thus, we can force the system to compute an ensemble average over a non-
Boltzmann distribution within a small interval of a prescribed value of $\chi$ through controlling the force constant $k$. In practice, unless the entire range of $\chi$ coordinate is spanned in a single simulation, multiple simulations (windows) are performed with different biasing umbrella potentials $W_b(\chi)_i$ that center the sampling in different overlapping regions or windows of $\chi$.

By applying the biasing potential to the total potential energy, we can recalculate the distribution function in eq. 2.21 as $\langle \rho(\chi) \rangle^b$, $b$ indicate biased. Having this distribution function on hand, we can rewrite the final estimate PMF $\omega(\chi)$ as:

$$\omega(\chi)_i = \omega(\chi^*) - k_B T \ln \left[ \frac{\langle \rho(\chi) \rangle^b}{\langle \rho(\chi^*) \rangle} \right] - W_b(\chi)_i - k_B T \ln (e^{-W_b(\chi)_i/k_B T})_i$$

(2.24)

where the ensemble average is accumulated with $W_b(\chi)$ added to the system.

When running umbrella sampling with multiple windows, there is a question about how to handle the overlapping regions rising up. A strategy, the Weighted Histogram Analysis Method (WHAM) [131], was developed to treat this issue.

2.1.8 Limitations of MD Methods

Although MD simulations play a more and more important role in biological discovery, they do have drawbacks. One of the most important limitations of MD is that it cannot be used to describe phenomena involving bond breaking/forming events, excited states, or isomerizations. Furthermore, the reliability of results from MD simulations strongly depends on the accuracy of the force field. Moreover, the force field itself is an approximation. The assumptions made in the process have a direct impact on the accuracy of the results and the validity of conclusions drawn from them. For example, as we discussed in Subsection 2.1.3, the cutoff approach (or Ewald summation in PBC) is generally used for calculations of long ranged electrostatic interactions. This treatment is inadequate for some systems which have
strongly charged residues. Moreover, polarization is another issue which needs to be
addressed in force fields. Most available force fields assume static charge distribution
for atoms. This kind of treatment may lead to the underestimation of intermolecular
interaction energy for weakly bound dimmers or molecular complexes, especially for
the modeling of solvation of ions [56-58]. But, present polarizable force fields, which
introduce polarizable dipoles centered on atomic sites, or allowing charges to
fluctuate based on the electronegativity equalization method in response to external
electric fields [57, 58], are still far from good enough. For example, present force
fields exclude the electron transfer effect, which may be important for the
protein/environment interaction.

2.2 Quantum Mechanics Methods

Quantum mechanics (QM) methods have been developed for several decades.
By correctly mathematical description of the behavior of electrons and thus of
chemistry, QM is currently necessary to understand the behavior of systems at atomic
length scales and smaller.

The basic equation involved in QM is the Schrödinger equation, from which
the energy of the system as well as the positions and related properties of all particles
in the system can be calculated. The Schrödinger equation is expressed as:

\[ \hat{H} \Psi = E \Psi \] 

(2.25)

where \( \hat{H} \) is the Hamiltonian operator, \( \Psi \) is a wave function, and \( E \) is the energy. In
the language of mathematics, an equation of this form is called an eigen equation. \( \Psi \)
is then called the eigenfunction and \( E \) is an eigenvalue. The operator and
eigenfunction can be a matrix and vector, respectively.

The Hamiltonian operator \( \hat{H} \) is, in general,
\[ H = - \sum_{i}^{\text{particles}} \frac{\nabla_{i}^{2}}{2m_{i}} + \sum_{i<j}^{\text{particles}} \frac{q_{i}q_{j}}{r_{ij}} \]  
(2.26)

where \( \nabla_{i}^{2} \) is the Laplacian operator acting on particle \( i \). Particles are both electrons and nuclei. The symbols \( m_{i} \) and \( q_{i} \) are the mass and charge of particle \( i \), and \( r_{ij} \) is the distance between particles. The first term gives the kinetic energy of the particle within a wave formulation. The second term is the energy due to Coulombic attraction or repulsion of particles. This formulation is the time-independent, non-relativistic Schrödinger equation. Additional terms can appear in the Hamiltonian when relativity or interactions with electromagnetic radiation or fields are taken into account.

In practice, one approximation, the Born-Oppenheimer approximation, is applied for separating the nuclear and electron motions. Accordingly, the Hamiltonian for a molecule with stationary nuclei is

\[ H = - \sum_{i}^{\text{electrons}} \frac{\nabla_{i}^{2}}{2} - \sum_{i}^{\text{nuclei}} \sum_{j}^{\text{electrons}} \frac{Z_{i}}{r_{ij}} + \sum_{i<j}^{\text{electrons}} \sum_{i}^{\text{electrons}} \frac{1}{r_{ij}} \]  
(2.28)

Here, the first term is the kinetic energy of the electrons only. The second term is the attraction of electrons to the nuclei. The third term is the repulsion between electrons. The repulsion between nuclei is added onto the energy at the end of the calculation. The motion of nuclei can be described by considering this entire formulation to be a potential energy surface on which nuclei move.

Three popular methods exist within this approach: \textit{ab initio}, semi-empirical, and density functional methods. These three methods are all molecular orbital (MO) based methods. \textit{Ab initio} methods solve the Schrödinger equation based entirely on the laws of quantum physics. No empirical parameters are used in the computation. Semi-empirical quantum methods solve an approximate form of the Schrödinger equation, which relies on parameters, typically obtained from \textit{ab initio} calculations, to
simplify the computation. Density functional methods greatly simplify computation by calculating molecular energies based on the molecule’s electron density rather than many-body wave functions.

There is another approach to applying Schrödinger equation to chemistry, namely the valence bond (VB) method. Basically the MO methods allow atomic orbitals (AOs) to interact to create the MO of a molecule, and do not focus on individual bonds, as typically shown in conventional structural formulas. The VB method, on the other hand, treats the molecule as a sum (linear combination) of structures each of which corresponds to a structural formula with a certain pairing of electrons. In this chapter, MO methods are introduced only, while the VB methods are discussed in chapter V.

2.2.1 *Ab initio* Methods

The term *ab initio* is Latin for “from the beginning”. The computations in *ab initio* method are derived directly from theoretical principles without experimental data. However, it does have some mathematical approximations, such as using a simpler functional form for a function or finding an approximate solution to a differential equation.

The simplest type of *ab initio* electronic structure calculation is the Hartree-Fock (HF) scheme, in which the instantaneous Coulombic electron-electron repulsion is not specifically taken into account. Only its average effect (mean field) is included in the calculation. This is a variational procedure; therefore the obtained approximate energies, expressed in terms of the system's wave function, are always equal to or greater than the exact energy, and tend to a limiting value called the Hartree-Fock limit as the size of the basis is increased. Many types of calculations begin with a
Hartree-Fock calculation and subsequently correct for electron-electron repulsion, referred to also as electronic correlation, such as post-HF methods. Møller-Plesset perturbation theory (MPn) [132] and coupled cluster theory (CC) [133, 134] are examples of these post-HF methods. In some cases, particularly for bond breaking processes, the Hartree-Fock method is inadequate and this single-determinant reference function is not a good basis for post-HF methods. It is then necessary to start with a wave function that includes more than one determinant such as Multi-Configurational Self-Consistent Field (MCSCF) [135] and methods have been developed that use these multi-determinant references for improvements.

2.2.2 Semi-empirical Methods

Semi-empirical calculations are set up with the same general structure as a HF calculation in that they have a Hamiltonian and a wave function. Within this framework, certain pieces of information are approximated or completely omitted. Usually, the core electrons are not included in the calculation and only a minimal basis set is used. Also, some of the two-electron integrals are omitted. In order to correct for the errors introduced by omitting part of the calculation, the method is parameterized by fitting the results to experimental data or \textit{ab initio} calculations.

The advantage of semi-empirical calculations is that they are much faster than \textit{ab initio} calculations. The disadvantage of semi-empirical calculations is that the results can be erratic and fewer properties can be predicted reliably [101, 108, 109]. If the molecule being computed is similar to molecules in the database used to parameterize the method, then the results may be very good. If the molecule being computed is significantly different from anything in the parameterization set, the answers may be very poor. Semi-empirical methods are parameterized to reproduce
various results. Most often, geometry and energy (usually the heat of formation) are used. Some researchers have extended this by including dipole moments, heats of reaction, and ionization potentials in the parameterization set. A few methods have been parameterized to reproduce a specific property, such as electronic spectra or NMR chemical shifts. Semi-empirical calculations can be used to compute properties other than those in the parameterization set.

Semi-empirical calculations have been very successful in the description of organic chemistry, where there are only a few elements used extensively and the molecules are of moderate size. Some semi-empirical methods have also been devised for the description of inorganic chemistry as well [136]. The most commonly used semi-empirical methods in MD simulations are AM1 [137] and PM3 [138]

2.2.3 Density Functional Theory Methods

The central idea in density functional theory (DFT) is that a relationship exists between the total electronic energy and the overall electron density. Initially, Hohenberg and Kohn [139] stated in 1964 that the ground state electronic energy of a molecule can be expressed exactly as a functional of the electron density of the molecule. The term functional refers to a function of a function, which in this case means that the total energy has a functional dependence on electron density, which in turn is dependent on the coordinates of the electrons of the system. The approximate functionals in DFT methods use the following separation of the total electronic energy.

\[ E = E_T + E_V + E_J + E_{XC} \]  \hspace{1cm} (2.29)

\( E_T \) is the kinetic energy term from electron motion, \( E_V \) is the potential energy term and includes nuclear-electron attraction and nuclear-nuclear repulsion, \( E_J \) is
electron-electron repulsion term (also described as the Coulombic self-interaction of the electron density), and $E_{XC}$ is the exchange-correlation term and includes the contributions due to electron exchange and correlation. All terms of eq. 2.29 with the exception the nuclear-nuclear repulsion are functions of the electron density, $\rho$.

Then in 1965, Kohn and Sham[140] suggested that the sum of the kinetic energy of the electrons and the contributions from electronic interactions should be expressed as a sum of three terms: the kinetic energy (which is defined as the kinetic energy of a system with non-interacting electrons), the electron-electron repulsion energy, and the electron exchange-correlation energy. In practice, Kohn-Sham DFT calculations are performed using an iterative approach, analogous to the SCF approach used in HF calculations. An initial guess is made for $\rho$, which allows for derivation of a set of orbitals that leads to an improved value for $\rho$. The improved $\rho$ value is then used in the second iteration and so on until convergence is reached.

The exchange term in eq. 2.29 represents the effects of electron exchange and correlation on the total energy of the system, which is an advantage over HF, which does not treat effects of electron correlation. Unfortunately, there is no known exact expression for $E_{XC}$ (Hohenberg and Kohn showed that $E_{XC}$ is dependent entirely on $\rho$; however, the theorem does not provide the form of this functional) [139, 140]. A number of approximate expressions for $E_{XC}$ have been developed, which in turn lead to a variety of computational methods.

The key feature of DFT methods is the way in which electron exchange and correlation effects are directly incorporated; correlation effects are truly only considered in more complex, post-HF methods, such as MP2. Because DFT methods include both exchange and correlation effects, higher accuracy is achieved compared to Hartree-Fock, which does not include electron correlation effects. However, there
are still difficulties in using DFT to describe properly certain intermolecular interactions, in particular, van der Waals (dispersion) forces.

2.2.4 Limitations of QM Methods

In contrast to MM methods, QM methods can provide relatively more accurate results. However, the higher level of theory in the quantum mechanical approach comes at high computational cost. At present, most QM methods are typically limited to systems consisting of a few hundred atoms or fewer. This restriction essentially excludes macromolecules of biological interest.

2.3 Hybrid QM/MM Method

2.3.1 Basic Idea of QM/MM Method

Although computer simulations play a more and more important role in discovering the chemical and biological worlds, it is still a challenge to reach a result with both accuracy and efficiency. Generally, one biological molecule such as a protein, is composed of hundreds of amino acids with on the order of thousands of non-hydrogen atoms. In this kind of condensed phase system, the overall system and solvation environment have large influences on the reactivity of the biological molecule. Therefore, it is crucial to consider the whole system during the computer simulations. Sophisticated first-principles methods for simulating reactions and electronic processes lead to high accuracy, but these are limited by their computational cost to small molecules. High level \textit{ab initio} calculations performed without considering the whole protein cannot offer information about the role of the part of the protein that is not modeled. On the other hand, MM methods have enjoyed remarkable successes in modeling biosystems in solution. But difficulties remain in
MM methods for the treatment of chemical reactions that involve the forming/breaking of covalent bonds. A plausible way to take advantage of both the accuracy of QM methods and the efficiency of MM methods is to combine QM and MM methods such that a small part of the whole system is treated explicitly by a QM method, while the remaining majority is approximated by a MM method [71, 73]. The development and improvement of QM/MM approach is currently a very active area of research [20, 71-92]. The strategy is to partition the system into a QM region and a classical MM region. The QM region includes the reacting species and some other moieties that are expected to significantly impact the reaction, while the surrounding MM region provides a realistic reaction field. (Figure 2.3)

![Figure 2.3 Illustration of the combined QM/MM method in protein system.](image)

Figure 2.3 Illustration of the combined QM/MM method in protein system.

A force field approach is sufficient for describing the MM atoms as their thermal motion is largely near their equilibriums. On the other hand, a quantum mechanical description of the potential energy surface for the QM atoms is necessary because of their reactive nature. The QM/MM approach synthesizes the generality of
the QM models and the efficiency of MM force fields, and numerous studies have demonstrated that it is an effective method to characterize enzymatic reactions. During the past decade, combined QM/MM methods represent one of the most significant advances in computational chemistry \cite{20, 71-92}. Molecular dynamics simulations with combined QM/MM methods can provide the details related to the catalytic activities and reaction mechanisms.

### 2.3.2 Implementation of QM/MM Method

According to the definition of QM/MM method, the energy in this technique can be expressed as:

$$H_{\text{total}} = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM/MM}}$$  \hspace{1cm} (2.30)

where $H_{\text{QM}}$ accounts for the full interaction energy of all quantum mechanical particles with one other, $H_{\text{MM}}$ accounts for the full interaction energy of all classical particles with one other, and $H_{\text{QM/MM}}$ accounts for the energy of all interactions between one quantum mechanical particle and one classical particle. The final boundary term can be calculated by:

$$H_{\text{QM/MM}} = -\sum_{e} q_{i} r_{e,i} + \sum_{n} Z_{n} q_{i} r_{n,i} + \tilde{R}_{\text{vdW}}$$  \hspace{1cm} (2.31)

where $i$ is summed over all MM partial charges, $n$ over all QM nuclei, and $e$ over all QM electrons. In eq. 2.31, the first term is a 1-electron interaction between QM electron density and MM partial charges. The second term is the standard Coulomb interaction between QM nuclei and MM charges. The final term is required because electron density (and hence dispersion) is explicitly treated in the QM region, but not in the MM region.

Basically, the accuracy of combined QM/MM methods relies on three factors: (1) the level of the QM computation (to calculate $H_{\text{QM}}$ term), (2) the quality of the
empirical potential functions (to calculate $\hat{H}_{MM}$ term), and (3) the partition of QM and MM parts for the targeted complex system (to calculate $\hat{H}_{QM/MM}$ term).

For the first factor, a key issue in combined QM/MM methods is the level of QM computations. Ideally, high level *ab initio* QM methods are preferred. However due to high computational costs (or time consuming), combined QM/MM calculations usually put the QM portion at semi-empirical level (such as AM1 [137] or PM3 [138]), but density functional theory (DFT) is gradually being adopted in combined QM/MM methods [141, 142]. An attractive method is the self-consistent charge density functional tight binding (SCC-DFTB) method [143, 144].

For the second factor, the quality of the empirical potential functions, various force fields have been proposed and used to model the dynamic properties of biomolecules such as protein and DNA [53-55]. Methods for this term $\hat{H}_{MM}$, as well as the first term $\hat{H}_{QM}$, have been discussed the preceding section in this chapter.

The separation of the QM and MM regions usually involves the breaking of one or more covalent bonds. Thus, the appropriate treatment of the QM-MM connection is centrally important to the accuracy of combined QM/MM methods [77, 85, 141, 145-149]. Generally, treatment of the boundary can be classified into two groups. One is the link atom (LA) approach, where the unsaturated valences of QM fragments are satisfied by hydrogen atoms [73, 81, 82, 85], pseudo-halogen atoms [85, 146], or a pseudo-bond [147]. The second group goes to the methods with applying localized orbitals at the boundary. The approaches belong to this group involved the local self-consistent field (LSCF) formalism [76-79], and the frozen orbital approach within the effective fragment potential method [148] or density functional theory (DFT) method [141]. Another approach in the spirit of the LSCF method is the generalized hybrid orbital (GHO) method [145, 149].
2.3.2.1 Link Atom (LA) Approach

The link atom (LA) approach is the most straightforward strategy to treat QM/MM boundary, which uses a hydrogen atom as a capping atom. In general, the position of the QM/MM boundary is selected so that it will not cut across any particularly polar or polarizable bonds. In this way, the two electrons in a single bond can be separated; one is included in the QM region and the other one goes to MM region. In general, C - C bonds between sp³ carbon atoms are the best choice in biological system. For the free valence of the QM fragment, hydrogen is commonly chosen to be a capping atom because of the similarly electronegativity between H and C. A potentially better choice is a pseudo-halogen having seven valence electrons and an electronegativity similar to that of carbon. The ‘lone pairs’ on such a capping atom will then resemble the electrons from the other bonding orbitals that would reside on the atom if the system were fully QM, which may offer a better representation of the system. [147]

The LA approach has been widely used in all kinds of QM/MM applications, due to its simplicity. However, the use of link atoms has been associated with one drawback that it introduces additional degrees of freedom, which will result in the complication of calculations and leads to instability in MD simulations.

2.3.2.2 Localized Orbital Approach

More robust approaches without adding extra atoms are Localized Orbital approaches. In this formalism, one auxiliary region is introduced between the QM and MM regions. Now the full system is partitioned into three regions, which may be called the QM region, the auxiliary region, and the MM region. The new auxiliary region is characterized by nuclei having their normal nuclear charges, and electron
density expressed in some set of basis functions. Eq. 2.30 may now be rewritten as:

$$H_{\text{total}} = H_{\text{QM}} + H_{\text{aux}} + H_{\text{MM}} + H_{\text{QM/aux}} + H_{\text{QM/MM}} + H_{\text{aux/MM}}$$  \hspace{1cm} (2.32)

Compared to eq. 2.30, there are three new terms, all involving the auxiliary region. However, two of these terms are entirely classical, $H_{\text{aux}}$ and $H_{\text{aux/MM}}$. The first is simply the electrostatic interaction of the localized density and its nuclei with themselves, while the second is the interaction of the localized density and its nuclei with the MM point charges and non-bonded LJ terms between the two regions. As for the $H_{\text{QM/aux}}$ term, it is, in principle, not much different than $H_{\text{QM/MM}}$, except that instead of adding one-electron integrals over atomic partial charges it adds two-electron integrals with the orbitals for one electron being localized.

The first application of localized orbital approach is the localized self-consistent-field (LSCF) method [76]. In this case, the chemical bonds connecting the QM and MM fragments are call frontier bonds, and they are represented by a set of strictly localized bond orbitals (SLBOs), which are determined by calculations for small model compounds. The strictly localized character of these orbitals helps to ensure that they are transferable from the model system to the large molecule.

The generalized hybrid orbital (GHO) approach is an more attractive approach with the spirit of LSCF method [145, 149]. In GHO method, a set of four hybrid orbitals is assigned to each MM atom at the QM/MM boundary. Then, the hybrid orbitals of the boundary atoms are partitioned into one "active" orbital which is directed toward the QM atom, and three auxiliary orbitals which are held frozen with paired-spin-density populations equal to one minus one-third of the partial atomic charge the atom would carry for MM purposes, i.e., there is an attempt to spread out the character of the boundary atom over its frozen orbitals. The hybrid active orbitals, along with the atomic orbital basis functions, are optimized in SCF calculations,
whereas the charge density of the auxiliary orbitals acts as an effective potential for the boundary atoms. Thus, there is again a similarity to the link atom procedure, in that there is a fully optimized MO representing each bond at the QM/MM frontier, but in this case the orbital is surrounded by a much more realistic charge environment from the hybrid atom nucleus and its three frozen auxiliary orbitals.

A comparison among LA, LSCF and GHO method is shown in Figure 2.4. A key difference is that the orbital containing the two electrons in the C - Cₓ bond is frozen in the LSCF method, optimized in the context of an Cₓ - H bond in the LA method, and optimized subject only to the constraint that atom C’s contribution be a particular sp hybrid in the GHO method. In the LA and LSCF methods, the MM partial charge on atom C interacts with some or all of the quantum system; in the GHO method, it is only used to set the population in the frozen orbitals. [145]

![Figure 2.4](image)

**Figure 2.4** Comparison of QM/MM partitioning schemes across covalent bonds. Frozen orbitals are in green for the LSCF and GHO methods. Reproduced from ref. 108
3.1 Introduction and Background

3.1.1 Specific Role of Ammonium and the Ammonium Transporter

Ammonia/Ammonium (NH$_3$/NH$_4^+$) plays an important role in fundamental nitrogen metabolism of all domains of life [23-32]. In many organisms, it is the primary nitrogen source for biological synthesis of amino acids, whereas in animals and humans, renal and hepatic ammonia sequestration and excretion are of fundamental importance for the regulation of the systemic pH value and the functioning of the central nervous system [23-32]. For instance, within many living organisms, especially bacteria and eukaryotic microbes, NH$_4^+$ is the preferred and primary nitrogen source. Generally, most of organisms can acquire a variety of nitrogen sources from the environment. And these kinds of nitrogen sources will be consequently transformed into NH$_3$/NH$_4^+$ before they are utilized in amino acid biosynthetic pathways. The important role of NH$_3$/NH$_4^+$ as a primary nitrogen source also means that a number of microbes show attractions towards NH$_3$/NH$_4^+$. And they will change shapes, activities, and even species when they obtain NH$_3$/NH$_4^+$. Thus, this property might also make it possible for ammonium to be a biosensor to bacteria [30, 31]. Moreover, in plant metabolism, beyond functioning as a major nitrogen source, NH$_3$/NH$_4^+$ serves as a major form for nitrogen retrieval in leaves. Thus, along with the fact that NH$_3$/NH$_4^+$ is generated by photorespiration in the mitochondria and
subsequently assimilated in the chloroplast, significant ammonium fluxes are likely to occur across a number of plant cell membranes. Finally, in animals, ammonium transport systems are responsible for active resorption of ammonium in renal cells. Transmembrane flux of ammonium plays a critical role in the maintenance of the body's acid-base balance through the regulation of renal ammonia excretion [32]. A flux of ammonium from neurons to glial cells is also found in most nervous tissues implicating ammonium transport in brain function. [30, 31]

Although ammonium is important in the biological metabolism, it becomes harmful at high concentration in animals. As a universal phenomenon, sensitivity to ammonium occurs in all domain of life, such as animals, plants, and microorganisms, although toxicity levels depend strongly on the type of organism and can vary largely between closely related species [150, 151]. During the past several decades, researchers tried several explanations for the toxic effect of ammonium, such as the dissipation of proton gradients across membranes, the acidification of the external medium in response to ammonium uptake, or a disequilibrium in the acid/base balance [32]. However, these explanations can at best only partially explain the observed ammonium toxicity syndromes and in plants membrane transport of ammonium, in particular efflux from the most sensitive cellular compartments, appears to be crucial for alleviating ammonium toxicity.[151]

Since flux of $\text{NH}_3/\text{NH}_4^+$ occurs in all domain of life, the questions are: "How does the flux happen? Does it depend on the diffusion of $\text{NH}_3/\text{NH}_4^+$ or is there a transporter system?" Quantitative determinations of the permeability coefficient for $\text{NH}_3$ have been carried out for a variety of membranes from bacterial, plant, and animal cells, as well as for artificial bilayers. The accumulated data suggest that $\text{NH}_3$ is indeed capable of permeating through membranes at a rate similar to that for water.
However, ammonia exists in biological solutions in two forms, NH$_3$ and NH$_4^+$. The relative amounts of each are governed by the buffer reaction

\[ \text{NH}_4^+ \rightleftharpoons \text{NH}_3 + \text{H}^+ \]  

(3.1)

This reaction occurs essentially instantaneously and has a $pK_a$ under biologically relevant conditions of $\sim 9.1 - 9.3$. Accordingly, in most biological fluids of pH 7.4 or less, most ammonia is present as NH$_4^+$. At pH 7.4 only $\sim 1\%$ of total ammonia is present as NH$_3$. In other words, at the ambient pH condition and ionic strength of blood plasma, the NH$_4^+$ form is predominant. Moreover, because most biological fluids exist at a pH substantially below the $pK_a$ of this buffer reaction, small changes in pH result in exponential changes in relative NH$_3$ concentration with almost no change in NH$_4^+$ concentration. Consequently, both the diffusion of NH$_3$ and the specific transport of NH$_4^+$ can potentially occur in biological systems. In principle, ammonium transduction across cell membranes could occur either by direct NH$_4^+$ transport or by a combination of H$^+$ and NH$_3$ transport [31]. Because NH$_3$ is a relatively small, uncharged molecule, it can diffuse across most lipid bilayers. However, its membrane permeability is not infinite, and thus transepithelial NH$_3$ gradients can occur in the presence of high rates of either NH$_4^+$ or H$^+$ transport. On the other hand, NH$_4^+$ is a cation with very limited permeability across lipid bilayers in the absence of specific transport proteins. Passive NH$_4^+$ transport might be driven by a concentration gradient or an electrical potential gradient. At high concentration of NH$_3$/NH$_4^+$, passive membrane permeation of neutral NH$_3$ may be effective enough. At low concentration, however, a specific transport system is needed. This transport system is called ammonia/ammonium transporter (Amt) which exists in the Amt/Mep/Rh family of membrane proteins. (Figure 3.1)
NH₄⁺ at high concentration (> 1 mM)

NH₄⁺ at low concentration (< 1 mM)

Periplasm

Cytoplasm

Figure 3.1 Sketch map for the membrane system of Ammonia/Ammonium channel. At high concentration of NH₃/NH₄⁺, passive membrane permeation of neutral NH₃ may be effective enough. At low concentration, however, a specific transport system is needed. Red circles represent NH₄⁺, blue circles represent NH₃.

There is a further question regarding Amt proteins. As total ammonia comprises a mixture of the two species, NH₃ and NH₄⁺, the ratio of which is dependent on pH, there has been an intense controversy on the identification of the molecular species that are transported by Amt/Mep/Rh proteins. Despite numerous studies, there has been uncertainty and much speculation over whether a neutral ammonia (NH₃) [152-156] or a positively charged ammonium ion (NH₄⁺) [157-162] is more probable as the species transported through the 20 Å long ammonium channel.
of AmtB [33]. In other words, the precise substrate of Amt/Mep/Rh proteins in neutral or ion has been debated. Unfortunately, it is difficult to measure ammonium directly, and its transport behavior can only be inferred from its inhibitory effect on methylamine uptake [157, 158] or the measurement of ion current in oocytes [29, 156, 159]. Initially, uptake studies with $^{14}$C-labelled methylammonium in *C. glutamicum* and *AtAMT1/-*transformed yeast showed that inhibitors of the plasma membrane ATPase and protonophores strongly inhibited methylammonium transport, arguing in favour of NH$_4^+$ being transported [26, 155, 163]. The W148L substitution increases flux of CH$_3$NH$_3^+$ and indicates that the substrate is an ion [164]. Furthermore the $K_m$ of AmtB for methylammonium transport in *C. glutamicum* and in *Escherichia coli* is essentially independent of pH, which again favors CH$_3$NH$_3^+$ transport or CH$_3$NH$_2$/H$^+$ co-transport [155, 163]. The conclusions drawn from these observations have been questioned on the basis that rapid assimilation of methylammonium to methylglutamine by glutamine synthesis, or compartmentation to the vacuole in the case of fungi, might create an internal sink, which subsequently would accelerate inward-directed diffusion of CH$_3$NH$_2$ [153].

It would appear that, the recent high resolution X-ray crystallographic structures of *Escherichia coli* AmtB with and without ammonia or methylammonia and of *Archaeoglobus fulgidus* Amt-1 clarified that the neutral NH$_3$ is indeed the species conducted through the channel [33, 34, 165, 166]. This is further endorsed by *in vivo* data which show that AmtB function is independent of either the membrane potential or the intracellular ATP pool [166]. But we also noted that some plant Amt homologues may facilitate net NH$_4^+$ transport. [162]
3.1.2 Research on Amt Structure

To study the property of Amt, the first task was to elucidate its structure. Based on the research reported during the past several decades, the Amt proteins were expected to be integral membrane proteins. Computer analyses of the first cloned sequences revealed them to encode highly hydrophobic proteins with a molecular mass of 50-55 kDa [23, 25, 163, 167-171]. (See example in Figure 3.2) In 2000, a good model of Amt topology was derived from a combination of empirical analysis and multiple alignments of sequences derived from databases. A detailed empirical analysis of the *Escherichia coli* Amt protein (AmtB) used fusions between AmtB and the reporter proteins alkaline phosphatase (which is active in the periplasm) and β-galactosidase (which is active in the cytoplasm) to map the topology. This analysis concluded that the protein contained 11 TM helices with both the N and C-termini located in the cytoplasm [170] (Figure 3.2).

![Figure 3.2 Topology model for Amt proteins. Model based on empirical analysis of the topology of the *Escherichia coli* AmtB[170]. Adapted from ref 170.](image-url)
Then in 2002, Blakey et al. first purified the *Escherichia coli* AmtB protein and found it as a stable homotrimer [172]. The transport channels are located in the middle of the monomers, which are bound together and function independently. And the trimer is stabilized by the hydrophobic interactions, rather than intersubunit disulfide bonds [172]. This kind of oligomeric structure has also been found in other transporter systems [173-175]. For example, the lactose transporter, LacS, behaves as a monomer-dimer [173]. Two-dimensional crystals revealed an engineered form of lactose permease, LacY as trimeric complexes [174]. Likewise, the TetA tetracycline transporter also has trigonal symmetry in two-dimensional crystals [175]. Within the Amt family, all three of the *S. cerevisiae* Mep proteins also show signals compatible with the existence of homomultimers. It may therefore be the case that the native state of all Amt proteins is oligomeric. Indeed the related Rh proteins have been proposed to associate in tetrameric complexes in the erythrocyte membrane [176].

Recently, two research groups independently solved the high resolution X-ray crystallographic structure of AmtB from *Escherichia coli* [33, 34]. These significant advances have made it possible to examine and elucidate the mechanism of ammonium transport at the level of atomic interaction. Moreover, these X-ray crystallographic structures of AmtB also permit further homology modeling within the large Amt/Mep/Rh family and thus make more informative sequence comparisons than hitherto were possible.

*Escherichia coli* AmtB is an archetypal member of the Amt family and contains 406 amino acids which construct 11 transmembrane α-helices with an N-terminal periplasmic extension and a C-terminal cytoplasmic extension. In the native cytoplasmic membrane, AmtB forms a stable homotrimeric structure [172, 177]. The X-ray crystallographic structures as determined from different crystal forms by the
two groups are largely identical except for small differences in the local structure of two cytoplasmic loops. In both of the crystallographic structures, each monomer of the tightly packed trimer comprises the 11 TM α-helices, though showing considerable variation in length, tilt, and internal bends. Their topological arrangement represents a new fold and consists of two pentahelical bundles in opposite arrangement (antiparallel with respect to the membrane normal) and a C-terminal long helix tilted by about 45°. The two pentahelical substructures, probably the result of gene duplication, are related by an approximate local twofold axis, but this relationship is very poor at the sequence level. [178]

Of significant interest, both of the X-ray crystallographic structures identified a pathway for the substrate transportation. This pathway is an approximately 20 Å long narrow pore (with diameter 1.2 Å), that is located between the two pentahelical domains. (Figure 3.3) From the structural and chemical information of the X-ray crystallographic structures, a clue is provided to answer the question: which species is more preferable for binding to the AmtB membrane protein, NH₄⁺ or NH₃? In both crystallographic structures, a substrate binding vestibule has been identified in the intracellular region of each monomer based on the electron density peak. It is viewed as an ammonium ion-binding site to recruit NH₄⁺. Although the chemical nature of the bound species cannot be definitely assigned from the observed electron density since X-ray crystallography cannot usually identify hydrogen atoms, several evidences support this interpretation. First, it is endorsed by the observation of an elongated density peak (in place of a round one) at this position when ammonium is replaced by methylammonium in the crystal buffer [33]. Second, the binding site has two polar groups, Ser219 Oy and an adjacent water, within hydrogen bonding distance and two aromatic side chains (Phe107, Trp148) at distances favorable for cation-π interactions.
Furthermore, this binding site is not selective to Na\(^+\) or K\(^+\) ions even at rather high concentrations. Indeed, the combination of three hydrogen-bonds and one cation-\(\pi\) interaction in a synthetic small molecule ammonium receptor has yielded micromolar affinity for NH\(_4\)^+ and a selectivity of nearly 1,000 against K\(^+\). The combination of these two types of interactions has also been noted in the biomolecular recognition of substituted ammonium ligands [179]. Accordingly, the authors of both crystallographic structures came to the same hypothesis: this binding site is selective for ammonium ion (NH\(_4\)^+). In the other words, AmtB will prefer to bind ammonium ion (NH\(_4\)^+).

![Figure 3.3](image)

**Figure 3.3** The ammonia-conducting portion of AmtB (1U7G structure) after removal of portions of helices M8, M9, and M10. The positions of two histidines near the three NH\(_3\) sites (blue spheres) are shown in yellow and blue stick representation and were not included in the surface electrostatic calculation. Two narrow hydrophobic regions through the channel lie above and below the NH\(_3\) positions (the zone within a dashed rectangle). The orange sphere represents an ammonium ion. Adapted from ref. 33.
As the NH$_4^+$ ion will be preferred to enter the ion-binding site of AmtB, the subsequent questions are: "Which species will go through the 20 Å pore pathway?" and "Will NH$_4^+$ deprotonate before entering the highly hydrophobic channel?" At the periplasmic entry, the narrow pore is blocked by the phenyl ring side chains of Phe107 and Phe215 that are highly conserved in the whole Amt/Mep/Rh family. It is thought that transient structural fluctuations are sufficient to permit frequent passage of small molecules like NH$_3$ or H$_2$O. It is not known whether these fluctuations limit the maximum rate of conduction. Furthermore, the central part of the narrow pore is lined by highly hydrophobic side chains. The predominantly hydrophobic channel is consistent with the viewpoint that the neutral NH$_3$, rather than the positively charged NH$_4^+$, is conducted by AmtB. It should not be ignored, that within the channel there are also two almost totally highly conserved histidines, His168 and His318, which have side chains arranged such that a hydrogen bond forms between their δ nitrogen atoms. The remarkable conservation of this imidazole pair arrangement has given rise to speculations that it may have a role in the deprotonation of ammonium ions before they cross the central part of the channel as uncharged NH$_3$ molecules [34]. However, their location, deeply buried in the hydrophobic pore, raises questions with respect to the energetics of ammonium ion entry and subsequent proton release from the histidine residue. [178]

As mentioned above, there is a little difference between the crystallographic AmtB structures observed in the different crystal forms on the cytoplasmic side. Most of the residues in the long loop connecting transmembrane domains TM5/TM6 are disordered in the structure reported by Zheng (Protein data bank ID: 1XQE) [34] and the cytoplasmic ends of TM9 and TM10 as well as their connection are also affected by disorder. In the fully ordered state structure by Khademi (Protein data bank ID:
1U7G) [33], TM10 starts with a $3_{10}$ helical turn which generates a narrow constriction at the cytoplasmic exit around Val314 before the channel opens into the cytoplasmic vestibule. In the 1XQE structure the same $3_{10}$ helical turn conformation is observed, but the nine preceding residues are disordered. In the 1XQE structure, residues 309 – 314 at the N-terminus of TM10 assume a different, nonhelical conformation, which widens the cytoplasmic constriction by about 2 Å. Furthermore, it renders itself more polar because the main chain carbonyl group of Cys312 points into the pore in place of the Val314 side chain. It has been speculated that this more open-structural state may correspond to the conducting state of the channel and that the fully ordered but more closed state might be induced by the binding of the regulatory protein GlnK known to inactivate AmtB conduction. [33, 34, 42, 180]

As the fact that there are only small differences between the two crystallographic structures (the C-terminal 20 residues are disordered in all crystal forms suggesting that they are not essential for conduction, and truncations after TM11 have no or little effect on ammonium conduction [154, 180]), for the MD simulations, as well as the quantum mechanical calculations in this dissertation, the 1U7G structure which is the X-ray structure of AmtB at an atomic resolution of 1.35 Å determined by Khademi et al. (Figure 3.4) [33] was chosen to build the initial computational model.

Accordingly, the 1U7G AmtB crystallizes as the physiological threefold symmetric trimer of channel-containing proteins (Figure 3.4). The trimer extends ~65 Å parallel to the threefold axis and is 81 Å in diameter in the plane of the membrane. Eleven transmembrane-spanning α-helices (M1 to M11) form a right-handed helical bundle around each channel. Residues from M1, M6, M7, M8, and M9 of one monomer interact with helices M1, M2, and M3 of the neighboring monomer, with a
Figure 3.4 A ribbon representation of AmtB viewed from the extracellular surface demonstrates the trimeric structure of AmtB. Adapted from ref. 33.

Total interacting surface area of $2716 \text{Å}^2$. As described for other membrane proteins, polar aromatic side chains of residues Tyr62 at the periplasmic (termed extracellular) side and Tyr180, Trp250, and Trp297 at the cytoplasmic side would lie in the membrane-aqueous phase interface. At the extracellular side, the threefold axis is surrounded by just three closely packed copies of the M1 of each monomer, which together the central axis against passage. Toward the cytoplasmic side, the three M1s (tilted $+16^\circ$ to each other) veer away from the threefold axis to leave an open pocket ~10 Å across, formed by three copies of M1 and M6. M1 has a kink ($22^\circ$) in the helix secured by the only cis-proline (Pro26) in AmtB, a residue not conserved in the superfamily. M1 and M6 are not long enough to span the bilayer, consistent with the trimer being the stable physiological quaternary structure. The interfaces between
subunits are almost as hydrophobic as the exterior, suggesting that a monomer could be transiently stable in the membrane upon synthesis, before forming trimers. An isolated square planar arrangement of four water molecules, each hydrogen-bonded to each other (average hydrogen bond length 3.0 Å), makes two hydrogen bonds to carbonyl oxygen of Cys56 and Ala102 in an otherwise hydrophobic cavity in the interface. [33]

3.1.3 Computational Studies on AmtB

The discovery of the high resolution of X-ray crystallographic structure triggered computational studies on AmtB. Molecular dynamics simulations and quantum chemical calculations have been conducted in order to elucidate the mechanism of AmtB as well as understand its biological function [37-41, 181-186]. Based on quantum chemistry calculation for the intermolecular interaction energies between the differentially methylated ammonium and the ammonium channel protein AmtB, Liu et al. [39] analyzed the π-cation interaction, which is arguably the strongest non-covalent interaction and is known to play a key role in numerous biological recognition processes and the regulation of enzymatic activity. Their results attribute the molecular determinants for protein-ligand recognition in ammonium transporter AmtB to the aromatic cage formed by three aromatic residues Phe103, Phe107, and Trp148, as well as Ser219 [39]. Using MD simulation approaches, Nygaard et al. [182] also revealed that a highly conserved Asp residue (Asp160), whose mutation is known to destroy the transport capability of AmtB, plays a key role in NH₄⁺ deprotonation. Their results are in good agreement with our published paper [37] except that they proposed that deprotonation occurs near site Am2, after NH₄⁺ donates a proton to Asp160 via the backbone carbonyl group of Ala162 and the amide
N-H of G163 using an imidic acid mechanism. Based on the binding free energies of different species at the external cavities of AmtB and its D160N mutant, Luzhkov et al. [38] proposed that the structural effects of the Asp160 mutation are less important than the replacement of the negative charge. Through $pK_a$ calculations of $\text{NH}_3/\text{NH}_4^+$ at four positions with electron density peak by solving the Poisson-Boltzmann equation, Ishikita et al. [185, 186] similarly recognized the stabilizing role of Asp160 for $\text{NH}_4^+$ at the periplasmic binding site but surmised that Ser219 plays crucial role and is responsible for the deprotonation of $\text{NH}_4^+$. Though lacking adequate strong evidences, such as energetic analysis, Yang et al. [183] proposed a transduction mechanism for $\text{NH}_4^+$ and claimed that AmtB channel acts like a one-way valve for passage of $\text{NH}_3$ and the entrance gate can open spontaneously. Very recently, Bostick et al. [40, 41] showed that Asp160 is engaged in persistent hydrogen bonds with the protein, and that the negative charge of Asp160 stabilized ammonia/ammonium in its protonated form, shifting its apparent $pK_a$ upward by $\sim$4 units. Simulation results indicated that the importance of Asp160, as evidenced by mutational studies, is more likely due to recruitment of $\text{NH}_4^+$ from the periplasm and stabilizing its binding at site Am1. Bostick et al. concluded that Asp160 is not involved in the deprotonation of $\text{NH}_4^+$ and is a primary structural role. Their results suggested that the most plausible proton donor/acceptor at either of these sites is water. [40, 41]

3.1.4 Mechanistic Questions to Be Addressed

Although several experimental and computational studies have been performed on AmtB protein, many questions regarding the biological function and molecular mechanism of Amt/Mep/Rh proteins remain open.

First, the X-ray crystallographic structure reveals a recruitment vestibule for
which involves Phe103, Phe107, Trp148 and Phe215 right at the entrance of the narrow pore on the periplasmic side [33, 34]. The environment of this binding site is highly conserved within the almost if not all Amt protein family. As introduced above, the presence of this ammonium ion-binding site is important for AmtB to capture positively charged \( \text{NH}_4^+ \) rather than neutral \( \text{NH}_3 \). Moreover, the specific features of the binding site may strongly relate to the selectively against other cations and possibly water [178]. In other words, AmtB protein is selective to ammonium ion. Thus, it is of general interests to obtain the energy cost for the process of recruitment of \( \text{NH}_4^+ \) to the binding vestibule, which is the precondition of the whole ammonia transportation process.

Secondly, the transport rate is estimated to be \( 10^{-10} \) ammonium molecules per second per channel, compared with the diffusion limit \( 10^{8} - 10^{9} \) in an open channel [34]. The X-ray crystallographic structure of AmtB showed that AmtB does act as a channel [33, 34]. The most recent structure of Amt-1 from *Archaeoglobus fulgidus* resembles the AmtB structure and the authors made a similar conclusion [165, 187]. A similar channel concept has also been proposed for human Rh proteins [152, 188]. As the fact that there is a big difference of the energy barrier between a transporter and a channel in membrane system, it is also worth to obtain the energy profile for movement of ammonium/ammonia through the whole AmtB channel.

Thirdly, the crystal structures suggest that the phenyl rings of Phe107 and Phe215 also serve as gates for the channel which is highly hydrophobic and located at the center of each monomer of the homotrimeric AmtB. It is therefore desirable to explore the mechanism of gate-opening, which is a dynamical process and hard to be captured experimentally. Since there is no significant conformational change between AmtB structures with or without ammonium [33, 34], it must be the side chains that
control the gate movement and the subsequent ammonia transport. Furthermore, it is valuable to study whether this gate has selectivity to the small molecule such as water molecule.

As the conducted species has now been confirmed to be the neutral ammonia, the subsequent question is where and how the initial ammonium deprotonates. No direct experimental evidence is available so far. There are two highly conserved residues in the mid-membrane center of the pathway, His168 and His318, whose adjacent imidazole rings are arranged such that a lateral hydrogen bond is formed between their N81 nitrogens. It has been speculated that these two histidines may serve as proton acceptors for ammonium ions [33, 34, 178], and the essential role of these twin histidines for Amt protein activity has just been confirmed by site-directed mutagenesis experiments [189]. However, the subsequent diffusion of protons from these histidines back to the periplasmic phase seems problematic as there is no apparent proton transfer route from the inner channel to the external solution with the gates (Phe107 and Phe213) closed.

While the hydrophobic channel obviously disfavors the physiologically important cations Na\textsuperscript{+} and K\textsuperscript{+}, whose concentrations are much higher than that of ammonium, due to the extremely high energetic cost of desolvation, it is an interesting question whether water can go through the channel, considering that oxygen is only slightly more electronegative than nitrogen and H\textsubscript{2}O and NH\textsubscript{3} are of comparable dipole moments and sizes.

To probe the transport mechanism, MD simulations on the transduction of ammonium/ammonia through the membrane protein AmtB in silico have been performed. In particular, free energy profiles were derived based on the umbrella sampling technique. The MD simulations in this dissertation are intended to address
the following mechanistic questions: (1) the recruitment of ammonium ions from the bulk solution to the periplasmic binding vestibule; (2) the gating mechanism of the channel, where the gates are the phenyl rings of Phe107 and Phe215; (3) the ammonium deprotonation mechanism; (4) the energy profile for the passage of ammonia through the channel; and (5) the selectivity of the channel toward water molecules and whether water can pass the channel, either alone or following ammonia in a water-ammonia pair as there is weak hydrogen bonding interaction between them.

3.2 Theoretical Background and Computational Details

3.2.1 Computational Model Setup

During the entire computational study on AmtB project in this dissertation, three different models were built up. As shown below, model A is built for the molecular dynamics simulation on the transductions of ammonia/ammonium through native AmtB. Model B is built for the comparative simulations in D160A mutant protein. Finally, model C is built for the QM/MM calculation on the deprotonation process of NH$_4^+$. Within model C, a description of QM part is also provided.

Model A for the transduction of NH$_3$/NH$_4^+$ through native AmtB Since AmtB is a membrane protein, membrane model is the first choice to be built up for this project. With the advantages of self-assembly in water and tendency to close on themselves, self-sealing (meaning a hole is unfavorable) and ability to be extensive: up to millimeters, lipid bilayers model are excellent for cell membrane. Thus a complete computational membrane model based on the established procedure detailed by Woolf and Roux [190-192] can be built up to retain these properties. In this membrane model, the whole AmtB homotrimeric structure (1U7G structure which is obtained from the protein data bank) is embedded in a phospholipid bilayer
constructed with DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) to imitate the membrane environment. The complete candidate model with unit cell dimensions of 93\times101\times94 \, \text{Å}^3 (Figure 3.5, a cuboid model in which two water slabs of 30 \, \text{Å} thick are added to both the top and bottom of the phospholipid bilayer) consists of about 87,200 atoms, including the AmtB trimer, 167 DMPC molecules, 16787 water molecules, as well as 6-7 chloride ions to neutralize the whole system.

![Figure 3.5](image)

**Figure 3.5** AmtB membrane protein model (unitary cell) for simulations from side view, and the inset is the top view of the AmtB trimer.

However, since AmtB forms a homotrimeric structure in the native cytoplasmic membrane, and the channels are located in the middle of monomers, which are bound together by hydrophobic forces, these channels must function independently. Thus, to reduce the computational costs, a monomeric form of AmtB
was used to run simulations instead. Subsequent MD simulations confirm the feasibility of this monomeric model as the root mean square (RMS) deviation of the monomer with reference to the crystal structure ranges from 1.45 to 1.60 Å, close to the resolution (1.35 Å) of the crystal structure.

To build the simplified model, similarly, the X-ray structure of AmtB at an atomic resolution of 1.35 Å determined by Khademi et al. was used in this work (Protein Data Bank ID code 1U7G) [33]. The three mutated residues (F68S, S126P, and K255L) and engineered Met residues (S atoms were replaced by Se) in 1U7G were modified back to their native states. Hydrogen positions of the protein are incorporated using the HBUILD facility in CHARMM [66] based on heavy atom positions and standard bond lengths and angles. All histidine residues in the crystal structure were analyzed. The protonated states of histidine residues are determined based on their individual microenvironments; for instance, both His168 and His318 are found to be neutral, but the proton is attached to Nδ1 in the imidazole ring for His168, whereas the proton is on Ne2 for His318. The following CHARMM residue types were used. HSE: H100, H145, H318. HSD: H156, H168, H196, H350. After the AmtB protein’s structure is ready, we followed the Woolf and Roux’s procedure [190-192] to build the final model. The general strategy of Woolf and Roux’s procedure for creating a reasonable starting configuration for the protein phospholipid system consists of randomly selecting lipids from a pre-equilibrated and pre-hydrated set, placing them around the protein, and finally reducing the number of core-core overlaps between heavy atoms through systematic rotations (around the Z-axis) and translations (in the XY plane) of the lipids and protein. To provide the initial XY positions for each lipid, the full molecules are first represented as single effective particles corresponding to the average cross-section area of a single phospholipid. The
packing of the effective lipid particles around the protein is determined from a MD simulation in which the large effective particles are harmonically restrained at a given value of Z and moving in the XY plane with periodic boundary conditions. There are 15 steps during the procedure of membrane model setup:

Step 1: Calculates the cross section of the protein in its initial configuration. Essential indication to set the system dimension and the number of lipids on each layer.

Step 2-3: Distributes van der Waals spheres around the protein (top layer in step2, bottom layer in step3). (Figure 3.6a,b)

Step 4: Replaces the spheres by lipids chosen from a library of 2000 DMPC, and eliminates the bad contacts between lipids.

Step 5: Minimizes the system by slowly increasing the van der Waals size of atoms at each 100 steps of minimization. (Figure 3.6c)

Step 6-7: Makes a water box with the same cross section as the system. And minimizes the energy of the water box.

Step 8: Generates the bulk overlay of water for the system. Move the water box in Z until the optimum number of waters in determined. That sets the Z-translational distance.

Step 9: Minimizes the full system, followed by 5 ps of Langevin dynamics with these constraints: peptide fixed, center of mass of each lipid constrained by harmonic forces, penetration of water prevented by harmonic forces.

Step 10: 25 picosecond (ps) of minimization by dynamics with the same constraints.

Step 11: 25 ps of minimization by dynamics with harmonic forces on the peptide and reducing the constraints on the water and lipid molecules.

Step 12: 25 ps of minimization by dynamics reducing the constraints on the water and lipid molecules.

Step 13: 25 ps of minimization by dynamics reducing the constraint on the peptide.
Step 14: 25 ps of minimization by Langevin dynamics (the only constraints present are there to fixed the reference of the system and to prevent the drift of the peptide out of the central unit).

Step 15: 25 ps of minimization by dynamics with speed scaling method (with the same constraints). (Figure 3.6d)

Figure 3.6 Snapshots of structures (base on AmtB monomer) during membrane model setup through procedure detailed by Woolf and Roux [190-192].
Accordingly, in the simulation model, water slabs of 30 Å thick are added to both the top and bottom of the phospholipid bilayer. The final model (Figure 3.7) for product simulations, in the size of $60 \times 60 \times 92$ Å$^3$, consists of about 35000 atoms, including the AmtB protein, 72 1,2-Dimyristoyl-sn-Glycero-3-phosphocholine (DMPC) molecules, 6414 water molecules, as well 2-3 chloride ions to neutralize the whole system [37]. This model has been used in the simulations of (1) transduction of NH$_4^+$ through the native AmtB (3 chloride); (2) transduction of NH$_3$ through the native AmtB (by replacing NH$_4^+$ with NH$_3$; 2 chloride ions); (3) transduction of H$_2$O through the native AmtB (by replacing NH$_4^+$ with H$_2$O; 2 chloride ions).

![Figure 3.7](image)

**Figure 3.7** Final AmtB membrane protein model (unit cell) for simulations (Model A), where the white arrow highlights the ammonia channel and the ribbon structure is shown in yellow. Atom colors are red for oxygen, blue for nitrogen, green for carbon, and white for hydrogen. White arrow represents the channel location.
Model B for the comparative simulation in D160A mutant protein From the product simulation of AmtB-NH$_4^+$ complex based on the Model A, one snapshot where NH$_4^+$ is fully solvated with four water molecules in the first hydration shell was chosen for mutant D160A case. First, Asp160 residue was mutated to alanine with fixing the rest of the system through InsightII biological homology program [193]. After randomly adding one more chloride ion to neutralize the system, the D160A variant, which complexed with a NH$_4^+$ just outside the extracellular binding vestibule, is ready to be the initial structure. Without changing other setting comparing with Model A, the system was then brought to equilibrium after performing 100 ps simulations. By superimposing the final mutant configuration with the wild-type protein, the RMSD is calculated to be 1.16 Å, which suggests the high similarity of both structures. In this model, the surroundings of the 160th residue are more important. In AmtB, Asp160 is a helix-capping residue for the fifth helix and its negatively charged carboxylate group is inaccessible to the solvent and used to fix the backbone of the helical structure by forming hydrogen bonds with the backbone amide groups of residues Gly163, Gly164 and Thr165 [33, 34, 37]. Its main chain carbonyl group, however, directs to the solvent and forms hydrogen bonds with water molecules. With the removal of the carboxylate group, the main chain carbonyl group of Ala160 in the D160A mutant flips to the position originally taken by the carboxylate group of Asp160 (Figure 3.8) to form hydrogen bonds with the main chain amide groups of Gly163 and Gly164 to sustain the framework of the fifth helix. As a consequence, no remarkable structural difference between the wild AmtB protein and its D160A variant is observed and the mutation of Asp160 neither changes the pore structure nor blocks the entrance of the binding vestibule. Then, this model B is ready for the product simulation on D160A mutant.
Figure 3.8 The orientations of Asp160 in AmtB (in cyan) and Ala160 in the D160A mutant (in magenta) after superimposing these two protein structures.

Model C for QM/MM simulation on the deprotonation process. For the QM/MM simulations on the deprotonation process, the computational studies were initiated from the AmtB-NH$_4^+$ complex (also from the previous Model A) where NH$_4^+$ located at the first binding vestibule ($Rc1$ = -15.8 Å). (Figure 3.9, and see the reaction coordinate setting below in subsection 3.2.2). During the dynamics simulation of ammonium movement through the channel, at this state, one ideal 8 Å hydrogen bond wire between NH$_4^+$ and the carboxylate group CO$_2^-$ of Asp160 through two water molecules was observed (see detail discussion in subsection 3.4). Thus, this snapshot was chosen as the initial model for QM/MM calculation on deprotonation.
process. While other settings are similar with the Model A described above, the QM part (totally 17 atoms) consists of the ammonium ion (NH$_4^+$), two water molecules, and the carboxylate group of Asp160 where the QM/MM boundary, saying C$_\gamma$ atom of Asp160 residue, will be treated with the generalized hybrid orbital (GHO) method [145]. (Figure 3.10) The model C has been used in the QM/MM simulation of deprotonation process of NH$_4^+$. 

![Diagram](diagram.png)

**Figure 3.9** Snapshot showing the surrounding for the ammonium ion in the extracellular binding site ($R_{el} = -15.8$ Å). The QM part atoms are highlighted by ball-stick model. Black dash lines indicate the hydrogen bonds. Unit of distances is Angstrom.
Figure 3.10 QM part model for deprotonation process calculation. The QM part consists of the NH$_4^+$, two water molecules, and the carboxylate group of Asp160 where the QM/MM boundary, saying C$_y$ atom of Asp160 residue, will be treated with GHO method. Black dash lines indicate the hydrogen bonds. Black dash lines indicate the hydrogen bonds. Unit of distances is Angstrom.

3.2.2 Coordinates for Energy Profiles

In terms of reaction coordinate for the free energy simulations, we chose a site around the end of the channel as the reference point. Apart from an NH$_4^+$ (Am) in the recruitment vestibule, Khademi et al. observed three weak electron density peaks within the hydrophobic channel which were interpreted to be partially occupied NH$_3$ molecules (Am2, Am3, and Am4 in Figure 3.11) [33]. Interestingly, these four sites are nearly in line. Accordingly, in both the wild-type AmtB calculation (transductions of all species: NH$_4^+$, NH$_3$ and H$_2$O) and the mutation D160A calculation, the position of Am2, which is in the downside of the channel, was chosen as the origin of the coordinate to describe the movement of NH$_3$/NH$_4^+$ and plot the corresponding energy profiles (refer as Rc1 below). It should be noted, however, that the pathways of NH$_3$/NH$_4^+$ are not constrained to be linear at all in MD simulations.
Figure 3.11 The ammonia-conducting channel in AmtB with one ammonium (Am) outside the channel and three ammonia molecules (Am2, Am3, and Am4) inside the channel in the crystallographic structure. [33]

For QM/MM simulation on deprotonation of ammonium, whereas localized proton transfers have been well studied [20, 84, 194-200], long-range proton transfers in biochemical processes are more challenging to quantify as the exact transfer pathways and rate-limiting factors are difficult to identify due to the involvement of many residues and solvents [201-213]. One prominent difficulty in the theoretical studies of long-range proton transfers is the definition of a reaction coordinate to characterize the progress of long-range proton transfer [214]. Whereas geometric coordinates (such as bond distance changes) are appropriate to study chemical
processes in the gaseous phase, for reactions in the condensed phases there are collective solvent reorganizations which should be considered in the reaction coordinates, particularly in the charge-transfer or proton-transfer reactions where the protein and solvent undergo significant reorganization [215]. A promising solution is to define reaction coordinates for long-range proton transfers by considering the movement of the excess charge center [214, 216, 217]. For current case where only two water molecules are involved, however, a linear combination of donor-proton and acceptor-proton distances might be effective [214]. Thus, a commonly used reaction coordinate which is the antisymmetric stretch involving the donor (D), the transferring proton (H), and the acceptor (A) is chosen. In the present work, nitrogen of ammonium will be the donor, oxygen of the nearest water to ammonium will be the acceptor, (Figure 3.10) and then the reaction coordinate (refer as \( Rc2 \)) was defined as:

\[
Rc2 = r_{N-H_1} - r_{O_1-H_1}
\]  

(3.2)

Alternatively, reaction coordinates can be defined in terms of energy. Warshel proposed the empirical valence bond (EVB) method and defined the energy gap between the reactant state and product state as a reaction coordinate [91, 195, 210, 218-221]. Voth et al. have extended the simple EVB ideas to modeling proton transfer reactions in aqueous systems with multistate EVB configurations [222, 223]. Interestingly, Mo and Gao have developed a MOVB method which combines the BLW method with explicit solvent molecules [224]. The solvent molecules are treated molecular mechanically, and the dynamics of chemical reactions in solution can be studied by performing Monte Carlo simulations. The QM-MOVB/MM method has been applied to three nucleophilic substitution reactions in aqueous solution [224-226]. The proton transfer process can be described as a resonance of two states, or

\[
-CO^- \cdots H_2O \cdots H_2O \cdots NH_4^+ \quad (I) \leftrightarrow -CO_2H \cdots H_2O \cdots H_2O \cdots NH_3 \quad (II)
\]  

(3.3)
Although wavefunctions for the above two states can be well defined with Heitler-London-Slater-Pauling (HLSP) functions in \textit{ab initio} VB methods [227-240], the computations of the Hamiltonian and overlap matrix elements between VB functions remain a challenge. In HLSP functions all orbitals are nonorthogonal and different orbitals are for different spins. Goddard's generalized VB (GVB) method retains the VB form for only one or a few focused bonds (perfect-pairs) but accommodates all rest electrons with orthogonal and doubly occupied MO's [241, 242]. A further combination of the VB and MO methods is to represent bond orbitals with nonorthogonal doubly occupied localized orbitals [243-246]. Recently, Mo \textit{et al.} proposed a block localized wave function (BLW) for each diabatic state [247]. In the BLW approach it is assumed that the total electrons and primitive basis functions are partitioned into \( k \) subgroups, in line with the conventional VB ideas. The \( i \)th subspace consists of \( \{ \chi_{i\mu}, \mu = 1,2, \ldots, m \} \) basis functions and accommodates \( n_i \) electrons. Clearly, for a resonance structure every two electrons form a subspace. However, Mo \textit{et al.} extend the definition of resonance structures and allow a subspace to have any number of electrons. The localized MOs for the \( i \)th subspace \( \{ \phi_j, j = 1,2, \ldots, m \} \) are expanded in terms of \( \{ \chi_{i\mu} \} \) only

\[
\phi_j = \sum_{\mu=1}^{m_i} C_{j\mu} \chi_{i\mu}
\]

Subsequently, the BLW is defined using a Slater determinant and in the case of \( S = 0 \)

\[
\Phi_K^{BLW} = M_K \left( N! \right)^{-1/2} \det \left| \phi_{i1}^2 \phi_{i2}^2 \cdots \phi_{i1}^2 \phi_{i2}^2 \phi_{i1}^2 \phi_{i2}^2 \cdots \phi_{i1}^2 \phi_{i2}^2 \right| \]

Orbitals in the same subspace are subject to the orthogonality constraint, but orbitals belonging to different subspaces are nonorthogonal. The energy of the BLW is determined as:
\[ E^{BLW} = \langle \Phi^{BLW} | H | \Phi^{BLW} \rangle = \sum_{\mu=1}^{m} \sum_{\nu=1}^{m} d_{\mu \nu} h_{\mu \nu} + \sum_{\mu=1}^{m} \sum_{\nu=1}^{m} d_{\mu \nu} F_{\mu \nu} \]  

(3.6)

where \( h_{\mu \nu} \) and \( F_{\mu \nu} \) are elements of the usual one-electron and the Fock matrices, and \( d_{\mu \nu} \) is an element of the density matrix, \( D = C C^T S^{-1} C^T \) (\( S \) is the overlap matrix of the basis functions). The optimization of orbitals in BLW can be accomplished using successive Jacobi rotation or Gianinetti et al.'s algorithm [248, 249]. The latter generates coupled Roothann-like equations and each equation corresponds to a block. Thus, the first derivative of the energy with respect to nuclear coordinates \( \{q_i\} \) directly takes the form in conventional HF theory and the second derivatives can be computed numerically.

After obtaining the energies of states I and II using the BLW method, it is straightforward to adopt a two-state model [250] to derive the overall energy of the system and project the overall energy in terms of the energy gap between states I and II. Currently we are extending the BLW method to the DFT level ported to the GAMESS package [251]. A detail description the development and application of BLW method is shown in Chapter V.

3.2.3 Computational Details

All MD simulations were carried out with the molecular dynamics simulation program package CHARMM [66]. All quantum chemistry calculations were carried out with Gaussian98 suit of quantum calculation program [252]. For both of the wild-type AmtB simulation and mutation simulation on D160A, the all-atom empirical potential energy function CHARMM22 force field for proteins [53] and phospholipids [253] was used. For the water molecules, the TIP3P potential [118] was used. Mulliken charges at the HF/6-31G(d) were employed for nitrogen and
hydrogen atoms in ammonium/ammonia, whose Lennard-Jones (LJ) parameters were kept the same as those for amide groups. The list of non-bonded interactions was truncated at 13 Å, and the van der Waals and electrostatic interactions were smoothly switched off in the range of 11-12 Å. Umbrella sampling technique was adopted to generate the free energy profile or potential of mean force (PMF) along the transduction trajectory. A biasing harmonic potential with a force constant of 15-20 kcal/mol was imposed in simulations which were separated by 0.4 Å, that is, there would be 50 independent simulations for the transport of ammonium/ammonia of 20 Å distance. Overall, the trajectory of the ammonium transport was simulated with 40 windows, and either of the two trajectories of the ammonia transport was projected out with 75 simulations. For each window, the first 50 ps simulation brought the system to an equilibrium state, and the second 50 ps simulation generated dynamics data for further analyses. In the simulations, all of the bonds with hydrogen atoms in AmtB will be constrained with the SHAKE algorithm [254]. The whole system is further extended with periodic boundary condition (PBC) which was accomplished by using the image facility in CHARMM and translating the system along three dimensions. Such a periodic treatment made our simulation system behave like an infinite large system in the whole space. If a molecule moved out of the boundaries in the xyz three directions, its image entered the system from the opposite side. The electrostatic interactions are computed with no truncation using the particle mesh Ewald (PME) algorithm [116]. Dynamics of atoms were propagated using the leapfrog-Verlet integration algorithm [103] with the typical time-step of 1.0 fs with isothermal-isobaric (NPT) ensemble at 300 K and 1 atm. And the trajectories were recorded at every 0.1 ps for future analysis.

QM/MM simulations were performed to study the deprotonation of NH₄⁺.
the current AmtB project, to make the calculations efficient, the semi-empirical (PM3) QM/MM method was used to run the dynamics simulations. In the PM3/CHARMM simulations, the QM part, which is the proton transfer pathway, consists of a linear chain of two water molecules, \( \text{NH}_4^+ \) and the carboxylate group of Asp160. The interface (boundary atom) between QM and MM regions, namely C\( \gamma \) of Asp160 residue, was treated using the generalized hybrid orbital (GHO) method, with standard van der Waals parameters of the CHARMM22 all-atom force field. Thus, the charge of QM part is zero. To preserve the linearity and connectivity of the hydrogen-bonded chain, the heavy atoms were subjected to harmonic constraint with 10 kcal/mol/Å\(^2\) force. Similarly, umbrella sampling technique was adopted to generate the free energy profile or potential of mean force (PMF) along the proton transfer trajectory. In this case, simulations were separated by 0.2 Å, and there would be 10 independent simulations. During the simulations, the bonds with hydrogen atoms (except the QM hydrogen atoms) in AmtB will be constrained with the SHAKE algorithm. PME algorithm is now built to involve the QM part using the new CHARMM version. Other settings such as non-bonded definition, PBC extension, time-step setting for simulation and trajectory recording, temperature and pressure etc., are similar with the MM simulation described above.

Within the study for elucidating the specific role of Asp160 residue, to appreciate the binding energy difference due to the carboxylate group, high level quantum mechanical calculations including electron correlation method at the MP2 level were performed on the reduced models with broken bonds saturated with hydrogen atoms by Gaussian 98 program suite. The basis set was chosen to be 6-31G, expanded to include diffusion and polarization functions (6-31+G(d)). (See below at Subsection 3.6)
To visualize the calculated data and generate the figures in this dissertation, the Gaussview program [255] and the graphic molecular modeling software – InsightII [193] were used.

### 3.3 Recruitment of NH$^+$ to the Binding Vestibule

To fully understand the whole process of the ammonium/ammonia transport across AmtB, MD simulations started with a state where the NH$_4^+$ is completely solvated by aqueous solution and has no direct interactions with the membrane protein. The state is outside the protein and has a distance of around 20.7 Å with the reference point for NH$_4^+$. A few umbrella sampling simulations separated by 0.4 Å were performed around this site. Slight energetic fluctuation (0.7 kcal/mol) is observed, and a minimum is located at a distance of about 20 Å. A snapshot at this site is shown in Figure 3.12, where a solvent shell composed of about four water

![Figure 3.12](image.png)  
**Figure 3.12** A snapshot for the initial configuration of an ammonium ion outside the AmtB membrane protein, where the ion is fully solvated ($R_{c1} = -20.2$ Å).
Figure 3.13 PMF profiles for the passage of NH$_3$/NH$_4^+$ across the AmtB membrane protein. The free energy at the starting configuration is referenced as zero. The energy change along the transduction of NH$_4^+$ is plotted in red. As NH$_4^+$ most probably deprotonates in the end of the periplasmic recruitment vestibule (~16 Å), the free energy profiles for the transduction of NH$_3$ alone (in blue) or coupled with one water (in green) are made to overlap with the curve of NH$_4^+$ at this position. Two different binding sites are also indicated.

Molecules form hydrogen bonds with the NH$_4^+$ and the slight stabilization may result from the indirect interaction with the protein environment (i.e., long-range electrostatic interactions). As the solvation free energy of the NH$_4^+$ in water is around 80 kcal/mol [256, 257], NH$_4^+$ reaches its energy minimum at this state. Starting from this configuration, more umbrella sampling simulations were conducted to slowly move the ion toward and through the channel gates. Figure 3.13 plots the free energy profile along the transport of NH$_4^+$ from bulk solvent to the binding vestibule and into.
the channel (curve in red). Once NH$_4^+$ begins to interact with the protein residues directly, two binding regions for the ion, in the ranges of -17.1 to -15.5 Å and -12.2 to -9.2 Å, were identified. Notably, the first binding region (recruitment vestibule) is consistent with the X-ray crystal structure where an NH$_4^+$ is captured (Am in Figure 3.11, $Rc1 = -16.3$ Å). The first binding region is located in the outside of the channel, and the second one is inside the channel.

Javelle et al. [166] examined the temperature dependence of methylammonium uptake by AmtB from 4 to 37 °C and determined an activation energy ($E_a$) of 1.6 kcal/mol, which confirmed that AmtB functions as a channel rather than a transporter. For comparison, the $E_a$ of transporters usually ranges from 11 to 27 kcal/mol [258]. The energy profile (red curve in Figure 3.13) shows that the first ion-binding site for NH$_4^+$ requires only 3.1 kcal/mol of energy, in good agreement with the experimental measurement. Figure 3.14a presents a snapshot (at $Rc1 = -17.8$ Å) just before NH$_4^+$ climbs into the first binding site. At this snapshot, NH$_4^+$ loses one water molecule in its first hydration shell, but the energetic loss is mostly compensated by the hydrogen bond with the backbone carbonyl oxygen of Ala162 (when NH$_4^+$ floats away a little bit, it can also bind to the backbone carbonyl oxygen of Phe161) as well as the $\pi$-cation interaction with Trp148. As a consequence, the overall solvation energy reduction is trivial. With the further movement of NH$_4^+$ toward the channel, NH$_4^+$ falls into the binding vestibule, which is composed of Phe107, Trp148, Phe215, and Ser219. Ser219 plays its role by forming a strong hydrogen bond with NH$_4^+$, whereas the aromatic rings of both Trp148 and Phe107 bind NH$_4^+$ via $\pi$-cation interactions. As a result, in the range from -17 to -16 Å, the free energy changes negligibly. These results support the viewpoint: it's is easy for ammonium ion to be bound to AmtB membrane protein.
Figure 3.14 Snapshots illustrating the surroundings for the ammonium ion in the first binding region: (a) just before reaching the first binding site ($Rc1 = -17.8 \text{ Å}$); (b) in the periplasmic recruitment ($Rc1 = -16.4 \text{ Å}$). Unit of distances is Angstrom.
Interesting finding in this range is that the first gate of the channel (phenyl ring of Phe107) starts to open dynamically. Figure 3.14b shows a snapshot at $Rc1 = -16.4$ Å, where the phenyl ring side chain of Phe107 has rotated nearly $90^\circ$ to be perpendicular to the aromatic ring of Phe215, which still blocks the channel (more discussion can be found in following Section 3.4). Furthermore, when $\text{NH}_4^+$ enters into the narrow, mostly hydrophobic (with the exception of Ser219) recruitment vestibule, we found that one water molecule can also fall into the vestibule and directs its hydrogen to the aromatic ring of Phe215 (Figure 3.14b). Thus, it is of general interest to examine the possible fate of this water molecule which is still in the hydration shell of $\text{NH}_4^+$ by a hydrogen bond with $2.19$ Å long.

The second binding site ($Rc1 = -12.2$ to $-9.2$ Å) is located in the beginning of the channel, and $\text{NH}_4^+$ has already entered into the channel. One water molecule is observed to follow $\text{NH}_4^+$ into the channel due to the strong hydrogen bonding interaction between them. While Phe107 has already returned to its original position and closed the channel, Phe215 starts to open its gate as its phenyl ring, together with Trp212, forms a binding pocket for $\text{NH}_4^+$ via stabilizing $\pi$-cation interactions. Apart from one hydrogen bond with water, the ion forms hydrogen bonds with surrounding residues. Interestingly, there is a noticeable barrier (1.3 kcal/mol) in the middle of the second binding region, which is thus divided into two areas of similar energies. Two representative snapshots for these two areas are shown in Figure 3.15. In both cases, the ammonium ion interacts with the aromatic ring of Phe215 noticeably. In the first area (Figure 3.15a, where $Rc1 = -11.8$ Å), except for one hydrogen pointing toward the phenyl ring of Phe215, the rest three hydrogen atoms in $\text{NH}_4^+$ form hydrogen bonds with water, the backbone carbonyl oxygen at Ala162, and the carbonyl oxygen at Asn216. In the second area (Figure 3.15b, where $Rc1 = -9.7$ Å), however, there is a
Figure 3.15 Snapshots illustrating the surroundings for the ammonium ion in the second binding region: (a) in the first part of the second binding region ($Rc_1 = 11.8 \text{ Å}$); (b) in the second part of the second binding region ($Rc_1 = -9.7 \text{ Å}$). Unit of distances is Angstrom.
significant change for the binding mode of NH$_4^+$, which now binds His168, Thr273, and Trp212. The small barrier between these two areas thus reflects the shifting of hydrogen bonds.

Compared with the first binding region (Figure 3.14b), there is a considerable energy loss for the system at the second binding region by 5.6 kcal/mol, and the barrier between these two sites is about 6.9 kcal/mol (Figure 3.13). This is somewhat lower than the estimated electrostatic barrier of 10 kcal/mol determined by Zheng et al., who also suggested that NH$_4^+$ may lose its proton at this site to His168 [34]. Energetically, it seems probable, although it is certainly more favorable for NH$_4^+$ to deprotonate at the first binding site as the barrier from the first to the second binding site is about 6.9 kcal/mol. However, one concern for this deprotonation mechanism at the second binding site is how the proton can be shipped out of this channel. Also, there is no water found in the crystal structure, suggesting no appropriate hydrogen bond chain via water molecules to facilitate the transduction of a proton out of the channel. Besides, the protonated His168 is still around NH$_3$, and there will be a strong electrostatic attraction between them. Moving NH$_3$ down the channel thus will be energetically very costly.

After the second binding site, a sharp increase of energy is observed if the ammonium ion keeps moving down the hydrophobic channel, largely due to the high desolvation cost and the lack of effective hydrogen bonds with the channel. This strongly confirms that ammonium ions cannot go through the very hydrophobic channel, and the conducted species thus must be neutral ammonia molecules.
3.4 Gating Mechanism

Crystallographic structures of AmtB reveal two gates, namely, the phenyl rings of Phe107 and Phe215 which are highly conserved in most if not all Amt proteins, in the entrance of the channel (Figure 3.11). In both structures with and without ammonia or methylamine, these two gates are parallel and block the narrow pore [33]. Thus, the passage of ammonium/ammonia through the gates must be a dynamic and fast process. However, our simulations on AmtB-NH$_4^+$ complex show that when NH$_4^+$ is far away from the binding vestibule, no significant movement of the two phenyl rings of Phe107 and Phe215 was observed. Interestingly, with the approaching of NH$_4^+$ to the first gate (Phe107), a rotation (or opening) of the phenyl ring of Phe107 to facilitate the passage of the invader was observed. Figure 3.16a illustrates this process by superimposing 12 snapshots along the movement of NH$_4^+$ in the range from -20.2 to -6.7 Å for $Rc1$. Obviously, the gate opening is not a continuous process, rather, a dynamic process. More specifically, there are only two states for Phe107, namely, "closed" (perpendicular to the channel) or "open" (parallel to the channel). When NH$_4^+$ gets close to Phe107, the latter will choose the "open" state; otherwise, it will stay in the "closed" state. This gate-opening process does not require additional energy (by comparing with the free energy profile) and thus is a very fast process which starts to appear in the first binding vestibule. It is assumed that a favorable orientation for the $\pi$-cation interaction between NH$_4^+$ and the aromatic ring of Phe107 may be the cause. Figure 3.14b also shows that, in the end of the first binding vestibule, the gate is already open.
Figure 3.16 The dynamic opening of the gates by (a) the ammonium ion (superimposed snapshots at $R_{cl}$) -20.2, -19.3, -17.8, -16.8, -16.3, -14.5, -13.2, -12.0, -10.9, -9.6, -8.1, and -6.7 Å in color from black to magenta; (b) the ammonia (superimposed snapshots at $R_{cl}$) -15.6, -14.1, -13.0, -11.3, -9.9, -8.7, -8.0, -7.4, -6.5, and -5.7 Å in color from black to purple.)
The continuous transport of \( \text{NH}_4^+ \) gradually leads to the second gate, namely the phenyl ring of Phe215, open. However, this process accompanies a high energy cost as depicted by the free energy profile in Figure 3.13, although this energy cost may be mainly due to the loss of hydrogen bonds with water molecules which are shut out of the channel due to the narrowness and hydrophobicity of the channel. When \( \text{NH}_4^+ \) falls into the second binding region, the first gate is still open, but once \( \text{NH}_4^+ \) leaves the first gate far enough and binds to His168, the first gate returns to the closed state which might due to the decreasing \( \pi \)-cation interaction between \( \text{NH}_4^+ \) and the aromatic ring of Phe107. Thus, we conclude that \( \text{NH}_4^+ \) is responsible for opening of the first gate of Phe107 but not the second gate of Phe215. Instead, we found a deprotonation mechanism (see Subsection 3.6) for \( \text{NH}_4^+ \) in the first binding vestibule, and the subsequent neutral ammonia will open the second gate dynamically and effectively (Figure 3.16b).

3.5 Functional Role of Asp160

The experimental determinations of Amt structures spurred a wave of interests in computational simulations of the \( \text{NH}_3/\text{NH}_4^+ \) transport in AmtB [37-41, 181-186]. Of particular interests, however, are the deprotonation mechanism for ammonium and the role of the highly conserved aspartate residue, Asp160, in the function of AmtB. There are two highly conserved aspartate residues, Asp160 and Asp310, in the Amt proteins of bacteria, fungi, and plants [42, 43, 259, 260]. Whereas there is no evidence yet to show that they actively participate in the ammonia transporting process, their roles are generally believed to be structural as their acidic side chains form hydrogen bonds with main chain N-H groups (residues 163-165 and 314-316, respectively). These hydrogen bonding interactions can fix the backbone of the helical structure
preceding His168 and His318 [34], although structural modeling of the AmtB protein suggested Asp160 as a potential candidate for an initial ammonium binding site on the periplasmic face of the membrane [170]. Mutation of Asp160 to Ala160, nevertheless, completely disables the transporting capability, while the D160E mutant retains 71% of the activity of the wild type [42]. Furthermore, X-ray structures clearly rule out the possibility for Asp160 as a binding site for the ammonium ion. It has been shown that Asp160 is a helix-capping residue for the fifth helix of AmtB and its carboxylate group forms hydrogen bonds with Thr165, Gly164, and Gly163 at the N-terminal end of the fifth helix. Although it is an ionic residue, Asp160 itself is not accessible to any solvent molecule since the indole ring of Trp148 is interposed between Asp160 and the binding vestibule for NH$_4^+$ and thus shields Asp160 from solvent [33]. One of the major conclusions from the analysis of static crystallographic structures is that the role of Asp160 is primarily structural. Based on the binding free energies of different species at the external cavities of AmtB and its D160N mutant, Luzhkov et al. also proposed that the structural effects of the Asp160 mutation are less important than the replacement of the negative charge [38]. Ishikita and Knapp similarly recognized the stabilizing role of Asp160 for NH$_4^+$ at the periplasmic binding site but surmised that Ser219 is responsible for the deprotonation of NH$_4^+$ [185]. Bostick and Brooks proposed that water is the only plausible candidate for accepting a proton from NH$_4^+$ [40, 41].

Of significant interest, our MD simulations showed that, in the periplasmic binding vestibule, the dynamic flipping of the phenyl ring of Phe107 is accompanied by a slight rotation and shifting of the indole ring of Trp148. Meanwhile, the carboxylate group of Asp160 rotates and subsequently reorients the carbonyls (backbone) of Gly163 and Ala162. All these collective movements accumulate to a
slot for Asp160 to become exposed to the bulk solvent, and one water molecule eventually manages to enter into the slot to bond with the carboxylate group. This water molecule connects to another solvent molecule which is in the first hydration shell of NH$_4^+$. As a consequence, an ideal hydrogen bond wire between NH$_4^+$ and the carboxylate group CO$_2^-$ of Asp160 through two water molecules was observed. Figure 3.17 depicts the changes of the surroundings of the carboxylate group of Asp160 with the approach of an ammonium ion from bulk solvent to the first binding vestibule. Therefore, *Asp160 is most likely the proton acceptor from ammonium ions or the driving force for ammonium deprotonation.* This explains the extreme importance of this residue and why the D160A mutant completely quenches the activity of AmtB. Most recently, Marini *et al.* also stated that “the conserved aspartate residue likely plays a preserved functional role in Mep/Amt/Rh proteins” based on their experimental finding that the transport and sensing functions of this aspartate residue are not dissociated [43].

However, at present, there is no direct experimental evidence to verify these computational findings regarding the specific functional role of Asp160. The lack of research in this aspect is mainly due to the fact that, only until very recently when the X-ray structures were accessible, the ammonia transporter proteins were thought to be ion channels. Due to the static nature of X-ray structures, however, the role of Asp160 is only recognized to be structural. To find the probable proton acceptor, other than Asp160, Zheng *et al.* proposed that His168, bridged with His318 via a H-bond, might facilitate the deprotonation of the ammonium ion in the second binding site [34], but as pointed out in the above, there are enough concerns not to pursue this hypothesis. In addition, our energetic analysis showed that the energy barrier to move NH$_4^+$ from the first binding site to the second binding site is 6.9 kcal/mol. Thus, it comes to the
hypothesis that the deprotonation process occurs in the first binding site, particularly, with Asp160 as the proton acceptor or driving force for ammonium deprotonation as water bulk is the ultimate proton acceptor in biological process.

Figure 3.17 The superimposition of two snapshots, where NH$_4^+$ is in the bulk solvent ($R_e 1 = -20.2$ Å, in purple) and in the first binding region ($R_e 1 = -15.8$ Å, in multi-colors). The hydrogen bond chain between Asp160 and NH$_4^+$ mediated by water molecules is highlighted by the dashed red line.

To deeply clarify the specific role of Asp160, extensive molecular dynamics simulations were performed on the homology structure of the D160A mutant and its ammonium transport capability was compared with that of native AmtB. As mentioned in the computation model part (Section 2.2.1), the computational studies were initiated from the AmtB-NH$_4^+$ complex where NH$_4^+$ is fully solvated with four
water molecules in the first hydration shell (the reaction coordinate $Rc1 = -20.7$ Å).

To evaluate the functional role of Asp160 versus Ala160, we simulated the binding and transduction of an ammonium ion from outside the binding site to the periplasmic binding vestibule and further into the channel. Figure 3.18 compares the free energy (or PMF) profiles along the ammonium transduction trajectory through the native AmtB and its D160A mutant channels, where the energies at the starting configurations were referenced as zero. By means of the free energy perturbation and molecular dynamics simulations, Luzhkov et al. found that the mutation of Asp160 to Asn destabilizes the bound $\text{NH}_4^+$ by 10 kcal/mol [38]. Although there is a significant loss of binding and a barrier, $\text{NH}_4^+$ can remain in the periplasmic binding vestibule ($Rc1 = -17.5 \sim -15.5$ Å) of the D160A mutant stably, as shown in Figure 3.18.

**Figure 3.18** Potential of mean force profiles for the translocation of ammonium across the AmtB (in red) and the D160A mutant (in blue) proteins. The free energies at the starting configurations are referenced as zero.
**Figure 3.19** Snapshot showing the surrounding for the ammonium ion in the extracellular binding site ($R_c l = -15.7$ Å) in D160A mutant. Unit of distance is Angstrom.

Figure 3.19 presents a snapshot at $R_c l = -15.7$ Å showing NH$_4^+$ in the extracellular ion-binding vestibule of the D160A mutant. Within this binding vestibule, NH$_4^+$ loses one water molecule in its first hydration shell, but the energetic loss is largely compensated by the hydrogen bond with the hydroxyl group of Ser219 and the backbone carbonyl oxygen of Phe161 or Ala162 successively with the movement of NH$_4^+$ as well as the π-cation interaction with Trp148 and Phe107. This picture is essentially identical to the case in the native AmtB. Similarly, we also observed the dynamic rotation of the phenyl ring of Phe107 at this stage to allow the
passage of NH$_4^+$. However, significant disparity occurs with the continuing advance of NH$_4^+$ into the channel. Due to the high hydrophobic and narrow nature of the channel, NH$_4^+$ will lose its most if not all hydration water molecules and this desolvation process is extremely endothermic, as evidenced by the high solvation free energy of ammonium ion in water (~80 kcal/mol). Thus, a steep increase of energy is observed for the transport of NH$_4^+$ in the D160A mutant after the extracellular binding site (see the blue curve in Figure 3.18). In contrast, in the native AmtB, an intracellular binding site in the range of -12.2 ~ -9.2 Å is identified. Structural analyses showed that the stabilization in the native AmtB comes from the negatively charged carboxylate group of Asp160 which is only ~8 Å away and forms a hydrogen-bond chain with NH$_4^+$ via the backbones of Ala162 and Gly163 (Figure 3.20a). When NH$_4^+$ moves deeper, the only coupled water can participate in the hydrogen-bond wire and make the ammonium ion remain stable in a ~3 Å long range. With two opposing charges located at the two ends of a hydrogen-bond wire, the hydrogen bond interaction or stabilization will be enhanced remarkably. By analyzing the binding energies, Luzhkov et al. also pointed out that the electrostatic interactions, notably from the proximal carboxylate ion of Asp160, are the largest stabilizing factor for NH$_4^+$ at the periplasmic binding vestibule [38]. With the mutation Asp160 with alanine, the carbonyl group of Ala160 flips and is similarly able to form hydrogen-bond wire with NH$_4^+$ via Ala162 and Gly163 (Figure 3.20b). But the magnitude of stabilization is greatly reduced compared with the AmtB case. Thus, stabilization of NH$_4^+$ at the second binding site is mainly contributed by electrostatic interaction between the carboxylate group of Asp160 and NH$_4^+$.

Further QM calculations to explore the binding energies have been done. To appreciate the binding energy difference due to the carboxylate group, we performed
QM calculations at the MP2/6-31+G(d) level on the reduced models shown in Figure 3.20 with broken bonds saturated with hydrogen atoms. Since our focus is on the interactions between the ammonium ion and its surroundings, NH$_4^+$ is defined as monomer A and the rest system as monomer B. Calculation result data is shown in figure 3.21 and 3.22. In the following Tables, $E(AB)$ refers to the total energy of the model, whiles $E(A:AB)$ and $E(B:AB)$ refer the energies of A and B with the basis sets of AB (to calculate basis set superposition error, BSSE), respectively.

**Figure 3.20** Snapshots illustrating the surroundings for the ammonium ion (a) in the second binding region of the native AmtB ($R_c = -11.3$ Å) and (b) in a similar position in the D160A mutant ($R_c = -10.8$ Å). Unit of distance is Angstrom.
Figure 3.21 QM calculations on the binding energy difference due to the carboxyl group.
Figure 3.22 QM calculations on the impact of the negatively charged carboxyl group of Asp160 on the Hydrogen bond between NH$_4^+$ and the backbone carbonyl group of Ala162.
The results showed that the binding energies are -118.8 and -63.8 kcal/mol after BSSE corrections for models a and b in figure 3.21, respectively. This big difference of the binding energy supports that mutating Asp160 to alanine has a large effect on the NH4+ at this position. Furthermore, we also performed calculations similar to energy decomposition to explore the binding energy between the negatively charge Asp160 and NH4+ by studying three model systems in figure 3.22. First, if we further simplify model a by retaining only Asp160 and NH4+ (model e in figure 3.22), the binding energy or pure electrostatic attraction between them is -39.6 kcal/mol. And we calculated the binding energy between the NH4+ and the backbone carbonyl group of Ala162 (model d in figure 3.22), which is -29.0 kcal/mol. The third model we calculated is that included NH4+, the backbone carbonyl group of Ala162, and the negatively charged Asp160 group (model e in figure 3.22). Similarly, we define NH4+ as monomer A and the rest system as monomer B. The results showed that the binding energies are -77.5 kcal/mol after BSSE corrections. The enhancement of the hydrogen bonding due to the carboxylate group of Asp160 can be estimated by ΔE(e) - ΔE(d) - ΔE(e) = -8.9 kcal/mol. Thus, the higher binding energy for NH4+ in the binding site of AmtB is significantly contributed from the negatively charged carboxylate group of Asp160, in both terms of direct electrostatic interaction and charge-enhanced hydrogen bonding with the backbone carbonyl group of Ala162.

Based on the discussion above both of the simulations of the movement of NH4+ within native AmtB and the mutant D160A, our computational results establish the central role of Asp160 in the uptake and deprotonation of ammonium ions in AmtB. In the periplasmic binding vestibule, we observed an ideal hydrogen bond wire between NH4+ and the carboxylate group CO2- of Asp160 through two water molecules. Along with our energetic analysis which showed that the energy barrier to
move NH$_4^+$ from the first binding site to the second binding site is 6.9 kcal/mol, we expect that the deprotonation process occurs in the first binding site and particularly, *Asp160 is most likely the proton acceptor from ammonium ions or the driving force for ammonium deprotonation*, as generally the final proton acceptor is water in biological systems. As the backbone carbonyl group will flip to fix the fifth helix when Asp160 is mutated to Ala160, the role of Asp160 is primarily functional rather than structural role. Although the negatively charged carboxylate group of Asp160 is buried and about 8 Å away from the transported ammonium ion, there is significant stabilizing electrostatic interaction between them. And the hydrogen bond network along the pathway plays important role in the deprotonation process. If NH$_4^+$ does not deprotonate in the periplasmic binding site and keeps moving into the channel, this favorable electrostatic interaction will be further enhanced by a hydrogen bond chain through Ala162 (the backbone C=O group) and Gly163 (the backbone N-H group). Then a huge increase has been observed in the free energy profile. This result explains the occurrence of the intracellular binding site in AmtB but none in the D160A mutant, as well as the high conservation of Asp160 in the Amt proteins and why the D160A mutant would completely lose the transport capability.

3.6 QM/MM Simulation on the Deprotonation of Ammonium

Based on the discussions above, we proposed that the NH$_4^+$ deprotonation occurs in the periplasmic binding vestibule and the highly conserved Asp160 is the proton acceptor or driving force for the deprotonation of NH$_4^+$. Scheme 3.1 depicts the hydrogen-bonded wire connecting the carboxylate group of Asp160 and ammonium ion with the participation of two water molecules. This is analogous to the 8 Å long-range proton transfer between a zinc-bound water and a histidine residue
through 2-4 intervening water molecules in carbonic anhydrase II (CAII) [214]. To advance our understanding of the deprotonation process of ammonium and the specific role of Asp160, direct evidences are needed.

Scheme 3.1: Proton transfer pathway from ammonium to carboxylate group of Asp160. The hydrogen network is highlighted by the dashed red line.

To fully understand the deprotonation process, the dynamics simulations started with a state where the ammonium arrived at the first binding vestibule (Figure 3.9, where \( R_{c1} = -15.8 \) Å). At this state, there is a hydrogen-bond wire mediated by two water molecules between \( \text{NH}_4^+ \) and the carboxylate group of Asp160. Generally, the deprotonation of \( \text{NH}_4^+ \) and the transfer of the proton to Asp160 could occur by a concerted hopping of protons between water molecules in manner analogous to the Grotthuss mechanism [261], which explains the anomalously high mobility of protons in water and also called proton-hopping-mechanism, or by a stepwise mechanism in which the proton is first transferred from \( \text{NH}_4^+ \) to the adjacent water molecule, then to the next water molecule, and finally to the carboxylate group [181]. Very recently,
with simple sphere model, Cao et al. proposed a mechanism which supports a stepwise mechanism rather than a concerted mechanism [181]. In Cao’s mechanism, the migrational substance is the negative rather than the positive charge. They found the carboxylate group first abstracts a proton from the water molecule close to it, and the resulting hydroxide ion abstracts a proton from another water molecule, which in turn abstracts a proton from NH$_4^+$.

However, our simulation results showed that the migrational substance is the positive rather than the negative charge. Furthermore, we found that this deprotonation process is most likely to choose a Grotthuss mechanism, rather than stepwise mechanism.

Basically, there are two different ways to define the reaction coordinate to study the proton transfer process in Scheme 3.1. The reaction coordinate can be set to be $Rc' = r_{02-H5} - r_{03-H5}$ which is between the carboxylate group of Asp160 and the water molecule W2 (see Figure 3.10 for atom number setting). In this setting, Asp160 is viewed to be the trigger in the deprotonation process. The carboxylate group will first abstract a proton from the water molecule W2 and then the whole deprotonation process finishes. This reaction coordinate setting was chosen in Cao’s work [181]. The reaction coordinate can also be defined to $Rc2 = r_{N-H1} - r_{O1-H1}$ (also see the Section 3.2.2) which is between the NH$_4^+$ and the nearby water molecule W1. In this setting, NH$_4^+$ is now viewed as the trigger and one proton of NH$_4^+$ will be derived to transfer to nearby water molecule W1 first and then the whole deprotonation process finishes. In this dissertation, simulations on both definitions were examined.

For the first reaction coordinate setting $Rc'$, from the initial structure (Figure 3.10 for QM portion), 10 window’s umbrella sampling MD simulations were
performed and the energy profile was plotted. As shown in Figure 3.23, an energy barrier of 21.22 kcal/mol was observed. There is only one transition state shown in the energy curve, which indicates that the deprotonation process is a one-step concerted process. During the simulations, the carboxylate group of Asp160 acts as a trigger of the deprotonation to abstract one proton (H5) from nearby water W2. The QM region in the transition state is shown in Figure 3.24. At this state, the proton H5 locates around the mid-point between O3 and O2. And one hydroxide ion is generated. However, at such high barrier energy, deprotonation process is very hard to happen. Therefore, MD simulations on the other reaction coordinate were performed.

![PMF of proton transfer from H2O to Asp160](image)

*Figure 3.23* PMF of proton transfer from water molecule W2 to carboxylate group of Asp160. The energy barrier for the deprotonation process is 21.22 kcal/mol.
Figure 3.24 QM part at $Rc' = 0.48 \text{ Å}$ in the simulation for proton transfer from H$_2$O to carboxylate group of Asp160.

Figure 3.25 PMF of proton transfer from NH$_4^+$ to the nearest water molecule W1. The energy barrier for the deprotonation process is 7.72 kcal/mol.
As the reaction coordinate was then set to $R_{c2} = r_{N-H_1} - r_{O1-H_1}$, after 10 windows’ umbrella sampling simulations, we plotted the free energy profile for the proton transfer process which was shown in Figure 3.25. Along with the transfer of proton from NH$_4^+$ to nearby water molecule (W1), an activation energy ($E_a$) of 7.72 kcal/mol was observed. This energy barrier is in good agreement with Cao’s calculations. Initially at a state where $R_{c2} = -0.8$ Å, the system reached its energy minimum (Figure 3.27 for the snapshot). At this state, while the ammonium ion has a hydrogen-bond network with three water molecules and Ser219 residue, there is a strong hydrogen-bond between the carboxylate group of Asp160 and the OH group Thr165 (distance is about 1.81Å). The hydrogen of OH group of Thr165 points to carboxylate group of Asp160. (Figure 3.27) During the deprotonation process, due to the driving force of the negatively charged Asp160 residue, the first proton (H1) moved from NH$_4^+$ to the most closed water molecule (W1). After conquering the energy barrier of 7.72 kcal/mol, the system arrived at the transition state (Figure 3.26a, $R_{c2}$ is about -0.15 Å) where a hydroxide ion is generated and one proton is observed to locate around the mid-point between nitrogen (N) and oxygen (O1) atoms. Due to the driving force of carboxylate group of Asp160 and the push from the proton H1 of NH$_4^+$, water molecule W1 loses one proton (H2) to the other water molecule (W2) and becomes a hydroxide ion. Meanwhile, carboxylate group of Asp160 got one proton (H5) from the water molecule W2. Figure 3.26a showed the QM portion at this state as a transition state where $R_{c2} = -0.15$ Å. At this point, we found that one proton located around the mid-point between nitrogen and oxygen. And there exists one hydroxide ion. We noted that the generation of a hydroxide ion seems very unusual. However, around the transition state, we found that this hydroxide ion is stabilized by a hydrogen bond network, namely OH$^-$—NH$_3$, ...
HOH···OH', and OH···NH-Trp148 (Figure 3.28, Figure 3.29). After this point, the proton H1 keeps moving from N1 to O1 and finally, it arrived at the product state (Figure 3.26b, 3.30). For comparison, 22.59 kcal/mol energy is required for the combination of positive-negative charges. The overall proton transfer process has a change in free energy of 14.87 kcal/mol. This supports the hypothesis that it is easy for NH₄⁺ to lose one proton at the first binding vestibule (only 7.72 kcal/mol barrier energy) and the Asp160 is the driving force for this deprotonation process.

**Figure 3.26** QM part structures of proton transfer process. a) Transition state of proton transfer where \( R_{e2} = -0.15 \, \text{Å} \); b) Product state structure where the proton has already been bonded to carboxylate group of Asp160, \( R_{e2} = 0.70 \, \text{Å} \).
Figure 3.27 Snapshot of AmtB in proton transfer process. It is the initial structure where $Rc2 = -0.80 \text{ Å}$. While the ammonium ion has a hydrogen-bond network with three water molecules and Ser219 residue, there is a strong hydrogen-bond between the carboxylate group of Asp160 and the OH group Thr165 (distance is about 1.81Å). The OH group of Thr165 points to carboxylate group of Asp160. Unit of the distance shown in the figure is Å.
Figure 3.28 Snapshot of AmtB in proton transfer process. It is the structure around the transition state, where $Rc2 = 0.12$ Å. At this point, there is a hydroxide ion which is stabilized by the hydrogen bonds, namely, OH$^-$···NH$_3$, HOH···OH$^-$, and OH$^-$···NH-Trp148. Unit of the distance shown in the figure is Å.
Figure 3.29 Snapshot of AmtB in proton transfer process. It is the structure around the transition state, where \( R_{c2} = -0.32 \) Å. At this point, the carboxylate group obtains one proton from the nearest water molecule W2. Meanwhile, water molecule W1 lost one proton to W2 and becomes a hydroxide ion. While the ammonium ion has a hydrogen-bond network with three water molecules and Ser219 residue still, now the direction of OH group of Thr165 change about 90°. And the carboxylate group of Asp160 lost a hydrogen-bond to Thr165 (distance goes to 3.41 Å). Unit of the distance shown in the figure is Å.
Figure 3.30 Snapshot of AmtB in proton transfer process. It is the product structure where $R_{c2} = 0.70$ Å. The proton has transferred to the Asp160 residue. While the ammonium ion still has a hydrogen-bond network with water molecules and Ser219 residue, the OH group of Thr165 points back to carboxylate group of Asp160. And there is a strong hydrogen-bond between the carboxylate group of Asp160 and the OH group Thr165 (distance is about 1.94 Å) again. Unit of the distance shown in the figure is Å.
During the proton transfer process, the hydrogen-bond network between NH$_4^+$ and the residues Trp148, Ala162, and Asp160 is retained, but remarkable changes are observed for Thr165. In the initial conformation Asp-NH$_4^+$, while the ammonium ion has a hydrogen-bond network with three water molecules and Ser219 residue, there is a strong hydrogen-bonding interaction between the carboxylate group (O4 atom) of Asp160 and the backbone OH group of Thr165 (H$_{\gamma 1}$ atom). (Figure 3.27, distance is about 1.81 Å) The OH group of Thr165 points to carboxylate group of Asp160. With the movement of the proton H1, this hydrogen bond disappears when the structure arrived at around the transition state (Figure 3.29). At this point, carboxylate group of Asp160 gets one proton (H5) from the water molecule W2. Asp160 becomes a neutral residue. While the ammonium ion still maintains a hydrogen-bond network with three water molecules and Ser219 residue, now the direction of OH group of Thr165 rotated about 90°. And the carboxylate group of Asp160 lost a hydrogen-bond to Thr165 (distance between the O4 atom and H$_{\gamma 1}$ atom goes to 3.41 Å). Finally at the product state, the proton has transferred to the Asp160 residue. Due to the all neutral groups within the pathway, structure analysis showed that the proton (H5) which is bound to carboxylate oxygen changes the orientation to the backbone oxygen of Ala162 residue, and forms a hydrogen bond between them. Along with this directional change, the OH group of Thr165 points back to carboxylate group of Asp160. And there is again a strong hydrogen-bond between the carboxylate group of Asp160 and the OH group Thr165 (distance is about 1.94 Å). (Figure 3.30) These observations support the calculations above (Subsection 3.5) that the charge-enhanced hydrogen bonding plays an important role in the higher binding energy for NH$_4^+$ in the binding site of AmtB and in the deprotonation process of ammonium ion.

To elucidate the relatively changes on different hydrogen bond network in this
deprotonation process, we plotted the trajectories of three different hydrogen-bond distances (O1···H1, O2···H2, and O3···H5) along the reaction coordinate $Rc_2$ shown in Figure 3.31. Within this trajectory picture, since distance of the hydrogen bond O1···H1 is involved in our reaction coordinate ($Rc_2$) setting, it shows a smooth curve in the trajectory picture. From the trajectory curve, it shows that before the transition state of deprotonation process, all of the three hydrogen bonds are in the distance range from 1.5 ~ 3.3 Å ($Rc_2 = -1.2$ Å ~ -0.5 Å). Although they are in fast dynamics movement, this distance range indicates that they are in strong hydrogen bond networks. With the migration of the proton H1 to water molecule W1, there is no significant change on the hydrogen network until the system arrives at the state where the $Rc_2$ is -0.35 Å. At this state, one dynamic movement in both hydrogen-bonds of O2···H2 and O3···H5 was observed. This dynamic movement on trajectory is due to the concerted movement of the protons we discussed in the pretext. During the whole deprotonation process, hydrogen bond O2···H2 has a similar behavior with hydrogen O3···H5. After the dynamic movement point in the trajectory picture ($Rc_2 = -0.5$ Å), along with changing on the distance of O1···H1 bond, both of O2···H2 and O3···H5 hydrogen bonds keep staying at the similar distance (about 1.0 Å) until the proton H1 bonds to O1 where the proton transfer process finishes (Figure 3.30). This distance indicates that the protons have already bonded to the oxygen.

In brief summary, we can clearly elucidate the detailed deprotonation process of NH$_4^+$ and confirm the functional role of Asp160 residue. According to our discussion in Sections 3.4, 3.5 and 3.6, deprotonation of NH$_4^+$ takes place after NH$_4^+$ enters the first ion-binding vestibule and prior to further translocation down to the hydrophobic pore (the channel). At this state, due to the strong interaction between the negatively charged Asp160 residue and the positive NH$_4^+$, the ammonium ion will
initialize the deprotonation process via an about 8 Å long proton transfer pathway. This pathway connects the carboxylate group of Asp160 and the NH₄⁺ through two water molecules. After slowly move the proton H1 from NH₄⁺ to nearest water molecule W1, a fast and concerted movement process happens. During this fast movement process, a hydroxide ion was generated and was stabilized by the strong hydrogen bond network along the pathway. Then the proton H1 keep moving to O1 to form a bond and the proton transfer process finished. The energy cost for this deprotonation process is only 7.72 kcal/mol. During this deprotonation process, Asp160 does play a functional role as the proton acceptor or driving force for proton transfer process, given that the final proton acceptor is water in biological system.

**Figure 3.31** Trajectory of three hydrogen bonds (O1···H1, O2···H2 and O3···H5) within proton transfer process.
3.7 Continuous Conduction of Neutral NH$_3$ in the Channel

As the deprotonation of ammonium ions most probably occurs in the periplasmic ion-binding vestibule and Asp160 is the proton acceptor mediated by two water molecules via a hydrogen bond wire, we proceeded to determine the free energy profile for the conduction of ammonia through the channel starting from this site. We chose an equilibrium configuration after 100 ps simulation for the ammonium ion and AmtB membrane model at $Rc = -15.8$ Å and removed the proton pointing toward the carboxylate group of Asp160. The subsequent 100 ps simulation for the ammonia and AmtB membrane protein model perturbs the amino acid orientations surrounding the neutral ammonia due to the dramatic reduction of hydrogen bonds of substrate (NH$_3$) with Ser219 and water molecules compared with the ammonium ion. Figure 3.32 superimposes two snapshots of NH$_4^+$ and NH$_3$ in the binding vestibule (at $Rc = -15.8$ Å). Consistent with the very low solvation energy of ammonia (4.3 kcal/mol) which is in contrast to the very high value of ammonium ion (~80 kcal/mol [181, 256]), the water molecules around NH$_3$ tend to move away, apparently due to the greatly weakened hydrogen bonds, and ammonia reorientates to make the nitrogen lone pair point toward the solvent side to form a hydrogen bond with solvent molecules. Notably, the water molecule just above the phenyl ring of Phe215 (also see Figure 3.14b) loses the binding to NH$_3$, but it is still trapped in the vestibule. Due to the hydrophobic nature of the deep site formed by Ala162, Phe215, and Phe107, it is difficult for water to reside stably in this pocket as the favorable water-water interactions can slightly stabilize the system. Thus, with the deprotonation of NH$_4^+$, the trapped water molecule has a tendency to move back the bulk solvent. However, our test simulations starting from this site but with different equilibrium conformations suggest that this trapped water may go through the channel coupled
with the ammonia before being able to get out. As a consequence, we performed complete simulations on two trajectories, one with NH$_3$ alone and the other with the NH$_3$-H$_2$O complex, whose free energy profiles are plotted in Figure 3.16. Note that we do not impose any restraint on the NH$_3$-H$_2$O complex in the second case.

**Figure 3.32** Snapshots before (in purple) and after (in multicolor) the deprotonation of ammonium ion at the binding vestibule ($R_{el} = -15.8$ Å). The transferred proton is highlighted with a circle in orange.
The passage of ammonia through the channel is relatively smooth, and the biggest bump is only 3.6 kcal/mol. When the ammonia travels alone, the first energy minimum is found at around -11.1 Å, due to the favorable hydrogen bond between \( \text{NH}_3 \) and His168. The energy maximum appears at -8.3 Å and may result from the environmental effect (i.e., the electrostatic field imposed by the surrounding protein environment) as we do not observe any immediate interactions except the hydrogen bond with His168. After a long movement with the breaking of the hydrogen bond with His168, \( \text{NH}_3 \) starts to stabilize slightly again at 3.8 Å by forming a hydrogen bond with His318. A further transduction without a noticeable barrier makes the ammonia exit the channel at \( R_{c1} = 8 \) Å, where Phe31 holds the gate. However, we do not observe the rotation of the phenyl ring of Phe31. Instead, the ammonia gets out of the channel around the phenyl ring. Since the passage of ammonia through AmtB is a freely reversible process and a specific base (Asp160) is required for the deprotonation of the ammonium on the periplasmic side, there must also be a specific base for deprotonation on the cytosolic side for transport in the reverse (and an acid for the forward transport). We examined the structure and found that, at the exit \( (R_{c1} = 9 \) Å), the ammonia connects via two water molecules to the carboxylate group of Asp313 and the ammonium group of Lys303, which may involve in the protonation/deprotonation of ammonia/ammonium on the cytosolic side (Figure 3.33).

The trajectory of the \( \text{NH}_3\text{-H}_2\text{O} \) complex is somewhat different due to the involvement of the water. For instance, the first minimum comes early at -13.5 Å due to the \( \text{NH}_3\text{-H}_2\text{O}\text{···His168} \) hydrogen bridge. In contrast to the maximum state for the first trajectory at -8.3 Å, the \( \text{NH}_3 + \text{H}_2\text{O} \) complex finds the minimum state at this site. Figure 3.34 compares the snapshots of the two trajectories at this region. Apart from the different orientations of the conducted species, there are no other apparent
discrepancies. We therefore surmise the difference originates from the environment effect. Both energy profiles reach minima together at 3.8 Å as at this moment the water in the second case has already slipped out of the channel. In the simulations, we do not observe any other external water coming into the channel from either the top or the bottom. This is different from the simulations of the transduction of the ammonium ion, where we observe water molecules coming from the bottom due to the strong electrostatic attraction from the cation.

Figure 3.33 Snapshot of the passage of ammonia alone through the hydrophobic channel of AmtB ($R_{c1} = -9$ Å).
Figure 3.34 Snapshots of the passage of ammonia alone (a) or coupled with a water molecule (b) through the hydrophobic channel of AmtB ($Rc1 = -8.1\,\text{Å}$).
3.8 Selectivity of the Ammonia-Conducting Channel toward Water Molecules

Due to the extremely high energetic cost of desolvation, the hydrophobic ammonia-conducting channel of AmtB obviously disfavors the physiologically important cations Na$^+$ and K$^+$ [158, 262], whose concentrations are much higher than that of ammonium. However, it is intriguing whether they can occupy the binding site as they can also form π-cation interactions with surrounding aromatic rings. Experiments showed that there is no significant inhibition of the Amt/Mep/Rh proteins by Na$^+$ and K$^+$, which is further verified by adiabatic free energy calculations [34]. However, it is still an open question whether water can go through the channel, considering that oxygen is only slightly more electronegative than nitrogen and H$_2$O and NH$_3$ are of comparable dipole moments and sizes. To this end, we computed the free energy profile for the passage of water through the ammonia transport channel. Similar to the calculations of AmtB with ammonia/ammonium, we initiated the MD simulations at $Rc = -20.7$ Å (the initial site for NH$_4^+$). Figure 3.35 shows the free energy profile for the transduction of a water molecule through the AmtB channel. There is no apparent periplasmic binding site for water, in contrast to the case of ammonium. Water requires about 6 kcal/mol to pass the two gates, which rotate to allow the water through. Since the ammonium ion needs only 3.1 kcal/mol to enter the periplasmic binding vestibule, the binding site outside the AmtB channel favors the ammonium ions over water molecules. Once the water molecule enters the channel, however, the energy trajectory is relatively smooth with only low barriers, as the hydrophobic channel can incorporate the hydrogen bond acceptor with not only ammonia but also water using weak interactions with C-H bonds [33, 34]. The two highly conserved residues in the mid-membrane center of the pathway, His168 and His318, can interact with water a little more strongly than with ammonia. We note
that Khademi et al. investigated the possible water conductivity by AmtB by the measurement of osmotic permeability [263] and the vesicle shrinkage and swelling [264]. Both experiments indicated no additional water conductivity in AmtB proteoliposomes. The structure of AmtB with \( \text{NH}_4^+ \) also showed no ordered water between the hydrophobic constricted regions of the channel [33].

![PMF of H\(_2\)O in AmtB](image)

**Figure 3.35** Potential of mean force (PMF) profile for the passage of a water molecule across the AmtB membrane protein.

3.9 Summary and Future Plan

We have performed detailed free energy MD simulations on the transduction of ammonium/ammonia through the 20 Å long ammonia-conducting channel of the *Escherichia coli* AmtB membrane protein. In agreement with the analysis of the X-ray crystallographic structures of AmtB complexed with ammonia or methylamine [33],
our simulations revealed a periplasmic binding vestibule which includes Phe107, Trp148, Phe215, and Ser219. This binding vestibule is specific for ammonium ions which form strong non-covalent \( \pi \)-cation interactions with the aromatic rings of Phe107, Trp148, and Phe215. Herein, Ser219 plays its role by forming a strong hydrogen bond with \( \text{NH}_4^+ \). Simulations on the transduction of water indicate that this binding site is not favorable for water molecules which require an activation energy of 6 kcal/mol to go through, since the activation energy for \( \text{NH}_4^+ \) is only 3.1 kcal/mol.

Of significance, with the entrance of \( \text{NH}_4^+ \) into the periplasmic binding vestibule, there are remarkable movements of the side chains of surrounding amino acid residues, that is, the flipping of the phenyl ring of Phe107, the slight rotation and shifting of the indole ring of Trp148. These collective conformational changes open a slot for the carboxylate group of Asp160 to access to the solvent. Asp160 normally is buried inside the Amt proteins and forms strong hydrogen bonds with the backbones of Thr165, Gly164, and Gly163 at the N-terminal end of the fifth helix. On the basis of the determined X-ray structures, Khademi et al. claimed that the role of the highly conserved Asp160 is structural. Our simulations showed that, with the opening of the first gate (the phenyl ring of Phe107), the carboxylate group \( \text{CO}_2^- \) of Asp160 forms hydrogen bonds with the ammonium ion via two water molecules. Thus, the role of Asp160 is functional rather than structural, and Asp160 is actually the proton acceptor from \( \text{NH}_4^+ \). Further simulation on the mutant D160A also supports the central role of Asp160 in the uptake and deprotonation of ammonium ions in AmtB. As the backbone carbonyl group will flip to fix the fifth helix when Asp160 is mutated to Ala160, the role of Asp160 is primarily functional. Although the negatively charged carboxylate group of Asp160 is buried and about 8 Å away from the transported ammonium ion, there is significant stabilizing electrostatic interaction between them.
If \( \text{NH}_4^+ \) does not deprotonate in the periplasmic binding site and moves into the channel, this favorable electrostatic interaction will be further enhanced by a hydrogen bond chain through Ala162 (the backbone C=O group) and Gly163 (the backbone N-H group). This explains the occurrence of the intracellular binding site in AmtB but none in the D160A mutant, as well as the high conservation of Asp160 in the Amt proteins and why the D160A mutant would completely lose the transport capability. The mechanism of deprotonation of \( \text{NH}_4^+ \) at this site were elucidated by detailed QM/MM simulations. Our simulation results showed that the migrational substance is the positive rather than the negative charge. Furthermore, we found that this deprotonation process is most likely to choose a Grotthuss mechanism (a concerted mechanism), rather than stepwise mechanism. The energy barrier for the deprotonation process is only 7.72 kcal/mol, which also support the conclusion that \( \text{NH}_4^+ \) ion deprotonates at the first bind vestibule.

With the deprotonation of \( \text{NH}_4^+ \) to \( \text{NH}_3 \), the second gate (the phenyl ring of Phe215) instantly opens and the subsequent transduction of ammonia through the channel is fairly smooth with only small bumps of less than 3.6 kcal/mol energy. Thus the AmtB protein is certainly a channel rather that a transporter (an 11~20 kcal/mol energy is required for transporter, generally). Figure 3.36 illustrates the overall ammonium/ammonia transport mechanism in the AmtB membrane protein.

Although the ion-binding site at the extracellular pore entry of the AmtB protein is specific for ammonium ions, the barrier for water to pass this site and enter the channel (6 kcal/mol) should be easy to overcome if there is no competition from ammonium ions. Thus, we conclude that water should be able to pass through the AmtB channel alone, but this statement needs further experimental verifications.
Figure 3.36 Illustration of the ammonium/ammonia transport mechanism in AmtB. This figure is a modification of the original mechanism proposed by Khademi et al. [33]

Future Plan Almost all ammonia transporter proteins contain two conserved histidines, which in AmtB are located within transmembrane helices M5 (His168) and M10 (His318) such that their side chains protrude into the channel [188]. Our simulations showed that these histidines might have a role in the transportation of NH3. Site-directed mutants on these histidines also supported their essential roles for optimum substrate conductance [189]. Thus, it is desirable to perform simulations on the H168A and H318A mutants to explore their specific roles.
CHAPTER IV

MOLECULAR DYNAMICS SIMULATION ON THE PHOSPHOTRIESTERASE (PTE): DOCKING PROCESS

4.1 Background of Phosphotriesterase

4.1.1 Overview of Phosphotriesterase (PTE)

Among the various groups of pesticides, organophosphates are one of the most widely used types, accounting for about 36% of the total world market, with more than 100 compounds being commonly used. Their high effectiveness has led to worldwide use, which has subsequently caused serious environmental problems. Since toxic pesticide waste and chemical stockpiles pose a serious potential threat to both the environment and human health [44], it is of great interests and significant importance to develop effective and economical methods for the detoxification and removal of these organophosphates from the environment and battlefield. Biodegradation provides a safe and efficient way to clean the residual organophosphorus pesticides from the environment, and counteract chemical nerve agents. The bacterial phosphotriesterase (PTE), which is isolated from wild-type, soil-dwelling bacteria *Pseudomonas diminuta*, has been viewed as a top candidate to be engineered to detoxify various organophosphate nerve agents due to its two advantages: the highly catalytic properties of its wild-type as well as its applicability under various field conditions [45-52].

PTE works particularly well for degradation of paraoxon to diethyl phosphate and *p*-nitrophenol (Reaction 1 in Scheme 4.1) by catalyzing hydrolysis of the P-O
bond. PTE can also catalyze the cleavage of P-F (e.g., Reaction 2 in Scheme 4.1) and P-S bonds. Consequently, this class of natural organophosphate-degrading enzymes is of considerable interest for detoxifying organophosphate-based insecticides and chemical warfare agents. [265, 266]

**Reaction 1**

\[
\text{EtO} - \overset{\text{O}}{\text{P}} - \overset{\text{O}}{\text{O}} - \overset{\text{O}}{\text{O}} - \overset{\text{O}}{\text{O}} + \text{H}_2\text{O} \rightarrow \text{EtO} - \overset{\text{O}}{\text{P}} - \overset{\text{OH}}{\text{OH}} + \overset{\text{O}}{\text{O}} - \overset{\text{O}}{\text{O}}
\]

Paraoxon

**Reaction 2**

\[
\text{Me} - \overset{\text{O}}{\text{CH}} - \overset{\text{OH}}{\text{OH}} + \text{H}_2\text{O} \rightarrow \text{Me} - \overset{\text{O}}{\text{CH}} - \overset{\text{OH}}{\text{OH}} + \text{HF}
\]

Sarin

**Scheme 4.1** PTE hydrolysis reaction PTE by catalyzing the cleavage of P-O and P-F bonds.

The remarkable enhancement of the hydrolysis of organophosphates catalyzed by the wild-type PTE can be exemplified with paraoxon as a substrate. Raushel and coworkers determined the kinetic rate constants, \(k_{\text{cat}}\) and \(k_{\text{cat}}/K_m\), for the hydrolysis of paraoxon by PTE are 2100 s\(^{-1}\) and \(4 \times 10^7\) M\(^{-1}\)s\(^{-1}\), respectively [267]. In contrast, the second order rate constant for the chemical hydrolysis of paraoxon by KOH (without PTE) at pH 7.0 is only \(7.5 \times 10^2\) M\(^{-1}\)s\(^{-1}\).[267]

However, PTE does not hydrolyze all organophosphate substrates at high rates. Since organophosphates are not natural chemicals and have existed for only about fifty years, the discovery of the natural enzyme phosphotriesterase from bacteria *Pseudomonas diminuta* is fortuitous in many ways. In other words, nature has not
evolved an enzyme to hydrolyze unnatural organophosphates as the evolution process generally takes millions of years. Thus, understanding the structural origin of the enzyme activity is pivotal to engineer PTE to increase its enzymatic diversity and activity. Engineered PTE can be utilized as a decontaminant to protect populations at risk for exposure to organophosphate pesticides or chemical warfare agents, as an agricultural tool to protect beneficial predators of undesirable insects, or as a way to clean and preserve the environment. As a currently burgeoning field, rational protein design can be employed to enhance the activity and to increase the stability of the targeted proteins by finding the optimal sequence. From the experimental point of view, there are basically two techniques to achieve the above goals. One is directed evolution, and the other is site-directed mutagenesis.

4.1.2 Experimental Studies on Phosphotriesterase (PTE)

Extensive experiments, including directed molecular evolution and site-directed mutagenesis, have been conducted to probe the amino acid residues responsible for the PTE catalysis and find mutated PTEs with higher catalytic activities. Directed molecular evolution starts from a known natural protein, and goes through several rounds of mutagenesis, functional screening, selection and propagation of successful sequences, until a few sequences with improved functions and properties are derived [268]. Directed evolution can be used to rapidly evolve any protein without knowledge of its structure. Cho, et al. searched novel PTE variants by means of directed evolution and after two rounds of DNA shuffling and screening, they successfully isolated several mutants with improved hydrolysis of methyl parathion [269]. In particular, one mutant hydrolyzes methyl parathion 25-fold faster than does the wild type. Although unexpected, novel sequences are often reported
based on directed molecular evolution experiments, a vast space of possible sequences potentially having better functions and other enhanced properties, such as stability, have not been explored.

Compared with the directed evolution, site-directed mutagenesis is more knowledge-based and mainly focused on the residues in the reactive pockets or loops which are most possibly involved in the activation, binding or reaction processes. To increase the diversity of organophosphate hydrolysates via rational protein design, site-directed mutagenesis around the active site of enzymes has been employed to find the best mutants towards various targets. A few groups, particularly the Raushel group at Texas A&M, have been conducting extensive studies on the organophosphate hydrolysis mechanism and engineering PTE with higher rate enhancement and selective substrate specificities [51, 270-276]. By comparing the structures of PTE and acetylcholinesterase (AChE) from *Torpedo california*, Gopal et al. mutated residues in the binding pocket and reactive site to simulate the environment of AChE in PTE with the intention to enhance the activity of PTE towards the chemical warfare VX agent [273].

As X-ray crystallographic studies have revealed that PTE is a dimeric metalloenzyme containing two zinc ions per monomer which are coordinated by four histidines, one aspartate and one carbamylated lysine as well as a bridging hydroxyl group [265, 266, 277] (Figure 4.1 for whole structure), it is intriguing to investigate the roles of these ligands in the enzymatic catalysis. di Sioudi et al. conducted a series of mutagenesis studies where the histidinyl residues at positions 254 and 257 are replaced by arginine (H254R) and leucine (H257L) [272]. The kinetics of hydrolysis of three mutants (two single-site mutants and one double-site mutant) are examined and compared with the native PTE towards paraoxon and a few other
organophosphates. Their results demonstrated the ability to significantly enhance the specificity of PTE for various substrates by site-specific modifications, and it is suggested that changes in metal requirements may affect these improved catalytic characteristics by enhancing structural flexibility and improving access of larger substrates to the active site, while simultaneously decreasing the catalytic efficiency for smaller substrates.

Figure 4.1 A ribbon representation of PTE enzyme (PDB ID:1DPM). It is a dimeric structure in the native form.

Within the active site of X-ray crystallographic structure (Figure 4.2), the two zines, which are separated by 3.31Å, are situated at the C-terminal portion of the TIM (triosephosphate isomerase) barrel motif. Lys169, which is carboxylated, and a water molecule (or hydroxide ion) serve as bridging ligands between two zinc cations. If we do not consider the possible metallic bond between the two zines, the more buried Zn1, referred to α-metal, is surrounded by His55, His57, Asp301 and Lys169, and a
hydroxide ion. In other words, Zn1 is coordinated by five ligands in a trigonal bipyramidal arrangement. If we count the metallic bond between two zinc ions, Zn1 could be viewed as octahedrally coordinated. Interestingly, the hydroxide ion binds to two zinc cations in an equal distance with R(O-Zn) of 1.93 Å in the crystal structure. The second zinc cation, Zn2, or β-metal which is more solvent-exposed, is tetrahedrally ligated by His201, His230, Lys169 and the bridging hydroxide ion, in an approximately trigonal pyramidal arrangement. Thus, the valency of Zn2 is not fully filled and there is a dangling orbital in Zn2 towards the solvent-accessible surface, which remains to serve as the catalytic center in the reactions. It is worthwhile to note that the substitution of the zinc ions by cadmium or manganese ions does not affect the enzymatic activity.

Figure 4.2 (a) Crystal structure of PTE with bound substrate analog diethyl 4-methylbenzylphophonate (PDB ID: 1DPM); (b) The reactive site of PTE. The red circle between two zinc ions (black circles) represents a water molecule or hydroxide ion.
4.1.3 Computational Studies on Phosphotriesterase (PTE)

Molecular dynamics simulations and quantum chemical calculations have been conducted in order to elucidate the structure of PTE as well as understand the reactivity in the active site [278-285]. Koca et al. analyzed the large distance between the phosphoryl oxygen and the nearest zinc Zn2 (3.4Å) in the crystal structure where the substrate is the paraoxon analogue EPB, and performed molecular dynamics simulations [284]. They found that the phosphoryl oxygen becomes strongly coordinated with Zn2 if the substrate is paraoxon or sarin, in accord with the high reactivity of PTE towards these two substrates. Of particular interest to the computational chemists, however, is the nature of the bridging oxygen between two zinc atoms found in the crystal structure. Although the oxygen species was initially assumed to be a water molecule, high-level quantum mechanical studies confirmed the species as a hydroxide anion [280]. Similar Zn1-OH-Zn2 bridges exist in other metalloenzymes such as zinc b-lactamase [286].

Since the enzyme contains several thousands of atoms, a pure quantum mechanical investigation is difficult. A conventional way to deal with an enzymatic reaction is to adopt an approximate cluster model (supermolecular model), which usually includes the active site and the first shell of ligands. In some cases, the bulk protein environment surrounding the active site (cluster model) is represented by a reaction-field or effective potential. Often this type of model can provide an intuitively correct yet overly-simplistic explanation about the enzymatic catalysis. Correct detailed and quantitative insights need a more appropriate treatment of the surrounding residues and solvent molecules. Many researchers have employed the cluster models to explore the detoxification in PTE. Kafai and Krauss determined the structure of the active site in PTE at the DFT level and the optimal cluster model is in
reasonable agreement to the X-ray crystal structure [283]. Krauss used a combined cluster model and effective fragment potentials (EFP) approach to determine the inherent electronic and structural characteristics of PTE [282], and explore the influence of water molecules in the active site [279, 287]. Similarly, Zheng et al. determined the structures of the active sites of Cd\(^{2+}\)-containing PTE [278, 279]. They further performed \textit{ab initio} and density functional theory calculations on active site models to study the reaction pathways and energy profiles of the alkaline hydrolysis of paraoxon, diisopropyl phosphorofluoridate (DFP), sarin, soman and O,O-dimethyl phosphonofluoridate.

In the aspect of phosphate hydrolysis, both pure QM methods and hybrid QM/MM methods have been applied to theoretically study the hydrolysis of various phosphates and phosphate esters in gas phase, aqueous solution [288-294] and protein environment [295-298]. For instance, Li and Cui discussed the ATP hydrolysis in myosin using hybrid QM/MM potentials [298], and Wladkowski used simple cluster models to perform QM calculations to explore the catalytic mechanism of the hydrolysis of phosphate esters by RNase A [297].

4.1.4 Our Current Objectives

As the understanding of the structural origin of the enzyme activity is the most fundamental objective in biological science, the goals of these simulations are to study the structural and energetic influences of the enzymatic environment on the hydrolysis of organophosphate pesticides and nerve agents and to evaluate the roles of residues in the binding and catalytic pocket. While the engineering on the PTE is the long term goal, the first objective of this project is to \textit{probe the catalytic mechanism of PTE}. 
As for the reaction mechanism, a $S_N2$ reaction mechanism has been postulated (Scheme 4.2), as experiment with a chiral organophosphate substrate demonstrated that the reaction proceeds along with an inversion of stereochemical configuration at the phosphorus center [299]. Consequently, the enzymatic reaction was driven by the electrostatic attraction between the solvent-accessible $\text{Zn}^2$ cation and the phosphoryl oxygen. With the gradual formation of the $\text{Zn}2$-$\text{O}$ bond, the hydroxide ion, which initially bridges the two zinc ions, breaks its bond to $\text{Zn}2$ and start a nucleophilic attack on the phosphorus center without forming a phosphorylated enzyme intermediate. During the formation of the bond between phosphorus and hydroxide oxygen, the bond between phosphorus and nitrophenol oxygen breaks simultaneously. At last, paraoxon is detoxified with the release of nitrophenol.

![Scheme 4.2 Postulated reaction mechanism for the PTE catalyzed hydrolysis of paraoxon.](image)

Therefore, the first step will be to dock paraoxon into the active site and get the initial equilibrium conformation. This step is instructive for the second step, which is to derive the free energy profile for the hydrolysis. In this dissertation work, we focused on simulating the docking process of the substrate and elucidating the catalytic mechanism of PTE, which was accomplished by performing QM/MM dynamics simulations.
4.2 Computational Model and Computational Detail

4.2.1 QM/MM Model Setup

The X-ray crystallographic structure of PTE with bound substrate analog diethyl-4-methylbenzylphophonate (PDB code: 1DPM, determined at 2.1 Å resolution) is used as the starting protein structure for computational simulations. The bounded substrate analog diethyl 4-methylbenzylphophonate was used as a template for designing the substrate structure, which was done by simply replacing the methyl group by nitro groups. X-ray crystallographic studies revealed that PTE is a dimeric metalloenzyme and each subunit contains a binuclear zinc center. In our MD simulations, the whole dimer structure was used. During the MD simulation, we will focus on the activity of monomer A, since the second monomer of PTE does not have direct functional connections to the first active site. Hydrogen positions of the protein are incorporated using the HBUILD facility in CHARMM [66] based on heavy atom positions and standard bond lengths and angles. All histidine residues in the crystal structure were analyzed. The protonated states of histidine residues are determined based on their individual microenvironments. The following CHARMM residue types were used. $HSE$: H201, H257; $HSD$: H55, H57, H230; $HSP$: H223, H254.

To reduce computational cost, a sphere model was built up. The final model (Figure 4.3) we used for product simulations consists of about 17000 atoms, including the PTE dimer, 4 zine ions, one substrate in monomer A and 2271 water molecules. Within this model, there are 7 positive charges as each zinc ion bears two positive charges.
4.2.2 Computational Detail

All MD simulations were carried out with the molecular dynamics simulation program package CHARMM [66]. The all-atom empirical potential energy function CHARMM22 force field for proteins [53] was used. For the water molecules, the TIP3P potential [118] was used. To minimize the computational costs, stochastic boundary conditions were imposed for this project. In stochastic boundary condition approach, a system is partitioned into three zones, the reaction zone (<25 Å), the buffer zone (25~30 Å), and the reservoir zone (>30 Å). As the geometry center of the two zinc ions in monomer A was set to be the center, all residues and water molecules that have at least one atom with the reaction zone were treated by Newtonian
dynamics. Atoms in the buffer zone were subjected to Langevin dynamics, in which friction coefficients of 200 ps$^{-1}$ and 62 ps$^{-1}$ for all heavy atoms of protein and the oxygen atoms of water molecules respectively. The atoms in buffer region were further restrained with harmonic force fields to maintain the crystal structure, with the corresponding force constants gradually scaled to zero at the reaction zone boundary. The harmonic force constants in units of kcal mol$^{-1}$ Å$^{-2}$ are 1.22 for backbone oxygen atoms, 1.30 for all other backbone atoms, and 0.73 for all side-chain atoms. All atoms in the reservoir region are fixed throughout the MD simulations. The reservoir region acted like a static force field, providing electrostatic interactions between the atoms in the reservoir region and atoms in the inner region. In addition, to contain the reaction zone, a deformable boundary potential corresponding to a 30 Å solvent sphere was applied to all solvent atoms in the system. The leapfrog integration scheme was used in simulations with a time step of 1 fs, and the temperature was set at 298 K.

Umbrella sampling technique was adopted to generate the free energy profile or potential of mean force (PMF) along the transduction trajectory. A biasing harmonic potential with a force constant of 15-20 kcal/mol was imposed in simulations which were separated by 0.2-0.3 Å. For each window, the first 50 ps simulation brought the system to an equilibrium state, and the second 100 ps simulation generated dynamics data for further analyses. The list of non-bonded interactions was truncated at 14 Å, and the van der Waals and electrostatic interactions were smoothly switched off in the range of 13-14 Å. Bond lengths involving hydrogen atoms (except the QM hydrogen atoms) were constrained with the SHAKE algorithm. Throughout the MD simulations, the atom lists for the frictions and random forces in the buffer regions were updated in every 5 time steps.

A hybrid QM/MM simulation was performed to derive the stable PTE-
paraoxon complex and study the active site dynamics. The QM region includes paraoxon and the six Zn ligands His55, His57, Lys169, His201, His230, Asp301 in the QM part which consists of 54 heavy atoms (or 100 atoms included Hydrogen atoms) totally and carries only one positive charge. The six QM-MM boundary atoms will be treated with the GHO method. As the conventional AM1 and PM3 models cannot reasonably produce the reaction barriers and heats for the hydrolysis of phosphates, mainly due to the omission of $d$ orbitals for phosphorous atom, we used a reparameterized PM3 method based on the DFT calculations of the paraoxon hydrolysis in the gas phase with small models, and used the subsequent PM3-SRP (Specific Reaction Parameters) to study the hydrolysis of paraoxon catalyzed by PTE. The PM3-SRP approach has been successfully used to study the enzymatic catalysis. The CHARMM all-atom force field will be used for the bulk MM region.

4.3 Results and Discussion

As mentioned in the model setup section, the bounded substrate analog diethyl 4-methylbenzylphophonate was used as a template for designing the substrate structure, to successfully obtain the initial equilibrium conformation, our simulations started with the state where the substrate paraoxon is in the similar location with diethyl 4-methylbenzylphophonate. After 100ps MD simulation, a few umbrella sampling simulations separated by 0.3 Å were performed around this site. We observed an energy minimum state where the distance between O@phosphate (refer as (P)O) and Zn2 is about 5.4 Å (Figure 4.4). Similar to the crystal structure, Zn1 is coordinated by five ligands in a trigonal bipyramidal arrangement or octahedral arrangement if we count the metallic bond between two zinc ions. The second zinc cation, Zn2, is tetrahedrally ligated by His201, His230, Lys169 and the bridging
hydroxide ion, in an approximately trigonal pyramidal arrangement. At this state, one water molecule was observed around the active site. The distance between (P)O and the hydrogen of this water molecule is only 1.75 Å. Thus, this water molecule might take part in the hydrolysis reaction. The current QM region setting does not involve this water molecule. Therefore, this information provides one probability that enlarged QM region which includes one water molecule might be good.

Figure 4.4 Snapshot of PTE active site and substrate in a state conformation. The distance of (P)O···Zn2 is 5.43 Å.
From this state, to derive the substrate entering the active site of PTE, with setting the distance of \((\text{P})\text{O} \cdots \text{Zn2}\) as reaction coordinate, we derived the simulations and plotted the energy profile or PMF for this docking process (Figure 4.5). Once the substrate leaves the energy minimum state, an energy barrier of 15.50 kcal/mol was identified. This energy barrier probably comes from the large steric effect between the substrate and the enzyme. Furthermore, during this process, the substrate changes its orientation a little bit to let the \((\text{P})\text{O}\) direct to Zn2 (see Figure 4.6). This also costs energy.

**Figure 4.5** PMF energy profile for docking process in PTE.

Figure 4.6 shows the snapshot of the energy maximum in PMF energy curve. At this state, the distance of \((\text{P})\text{O} \cdots \text{Zn2}\) is 2.62 Å. The ligands of Zn2, which are His201, His230, Lys169 and the bridging hydroxide ion, were observed to have slight
shiftings and slightly distorted the trigonal pyramidal arrangement of Zn2. These moving make the second zinc cation, Zn2, more solvent-exposed and the substrate more easily ligates to Zn2. Of particular note, at this transition state, we did not observe any water molecules around the active site, which might come from the reason that we did not involve one water molecule in QM region.

Figure 4.6 Snapshot of the docking process in PTE. The distance of (P)O···Zn2 is 2.62 Å.
After the transition state, due to electrostatic attraction from the more solvent-exposed Zn2 cation, the substrate falls into active site and ligates to the Zn2 cation (Figure 4.7, where the distance of (P)O···Zn2 is 1.92 Å). Now similar to Zn1 cation, the Zn2 cation is coordinated by five ligands in a trigonal bipyramidal arrangement or octahedral arrangement if we count the metallic bond between two zinc ions. Compared with the transition state, there is an energy loss for the system at this Zn2
cation binding pocket by 11.12 kcal/mol (see Figure 4.5). If we continue to drive the substrate close to Zn2 cation, a sharp increase of energy is observed after the binding site. This is due to the highly steric effect between the substrate and the enzyme. This strongly confirms that the system at the binding site, where the distance of (P)O···Zn2 is 1.92 Å, is a good conformation and good for a starting point for the second step simulation on the catalytic mechanism.

4.4 Summary and Future Plan

We have performed MD simulations on the docking process for paraoxon with PTE enzyme. The simulation started from a stable initial conformation where the distance of (P)O···Zn2 is around 5.4 Å. By using umbrella sampling dynamics simulations, one Zn2 cation binding site was identified where the distance of (P)O···Zn2 is 1.92 Å. Compared with the initial conformation, there is a considerable more energy for the system at the binding site by 4.38 kcal/mol, and the energy barrier between these two sites is about 15.50 kcal/mol. At the binding site, the arrangement of the ligands of Zn2 cation changes from an approximately trigonal pyramidal arrangement to a trigonal bipyramidal arrangement. After the binding site, a sharp increase of energy is observed. Thus, the binding site is a good stable point for the second step simulation on the catalytic mechanism of PTE.

Future Plan With the stable structure in binding site on hand, the subsequent objective will be to derive the free energy profile for the hydrolysis mechanism (Scheme 4.2). The PMF for the overall reaction can be determined by umbrella sampling along the reaction coordinate which can be set as the difference between the breaking bond and the forming bind, i.e., \( R_{\text{react}} = R(\text{P-O(ph)}) - R(\text{P-O(H)}) \). Based on the overall energy profile, we can conveniently locate the transition state and estimate
the reaction barrier. Subsequently, we can explore and elucidate the catalytic mechanism.

The third objective will be the difference between diethyl-4-methylbenzylphophonate and paraoxon catalyzed by PTE. Similar MD simulations will be carried out by the substitution of the nitro group by a methyl group and the subsequent PMF profile will be plotted. We hope to answer why PTE can effectively catalyze the hydrolysis of paraoxon, but not diethyl-4-methylbenzylphophonate. The structural information will provide us clues for further enzyme engineering.
CHAPTER V

DEVELOPMENT AND APPLICATIONS OF BLOCK-LOCALIZED WAVEFUNCTION (BLW) METHOD

5.1 Introduction

Molecular mechanics or force fields, where the potential energy function is generally expressed as the summation of various bonded and non-bonded energy terms, play a central role in computer simulations, and ultimately determine the accuracy of computational results. Although remarkable successes have been achieved in molecular dynamics and Monte Carlo simulations of liquids, solutions, and biopolymers, there are also deficiencies in the current generation of force fields. For instance, most of the available force fields assume static charge distribution for atoms by making use of pairwise potentials and fixed partial charges on atoms [53-55]. This kind of treatment may lead to the underestimation of intermolecular interaction energy for weakly bound dimers or molecular complexes, especially for the modeling of solvation of ions. Thus, improving these force fields by incorporating explicit many-body polarization terms into the potential energy function is needed. In addition, present force fields exclude the electron transfer effect, which may be important for the protein/environment interaction [300]. Since these deficiencies have important implications in force field development, an accurate estimate of the polarization and charge transfer interaction as well as the electrostatic component is warranted. Upon that, it is of interest to decompose the total intermolecular interaction energy into specific energy components. This can lead to deeper understanding of intermolecular interactions, and the quantitative results provided an
avenue to guide development of next-generation force fields. Based on Block-
Localized Wave function (BLW) method [59-61], which was recently developed by
Professor Yirong Mo, an energy decomposition scheme (BLW-ED) which is stable
with the variation of basis sets [62, 63] is proposed. This method is useful to study the
polarization and charge-transfer effects between systems, thus may afford clues for
the development and refinement of force fields.

The block-localized wavefunction (BLW) approach is an *ab initio* valence
bond (VB) method incorporating the efficiency of molecular orbital (MO) theory.
Instead of allowing all MOs to be a combination of all atomic orbitals in MO theory,
this BLW method defines the wave function for a resonance structure or diabatic state
by partitioning the overall electrons and primitive orbitals into several subgroups and
expanding each block-localized molecular orbital in only one subspace. As a
consequence, although block-localized molecular orbitals in the same subspace are
constrained to be orthogonal (a feature of MO theory), orbitals between different
subspaces are generally nonorthogonal (a feature of VB theory). In such a way, the
BLW method preserves the characteristics and advantages of both the VB and MO
theories. Significantly, the BLWs for diabatic states are optimized self-consistently,
and the adiabatic state is a combination of a few (usually two or three) diabatic state
wave functions. Currently, we extended the BLW method to the density functional
theory (DFT) level [64], which includes dynamical electron correlation and some
static correlation [140, 301-303]. This BLW-DFT method had been incorporated into
the quantum mechanical software GAMESS [65] which can be ported to the
molecular dynamics simulation software CHARMM [66].

Based on the BLW approach, an energy decomposition scheme (BLW-ED)
has also been developed. An attractive characteristic of the present BLW-ED method
in comparison with other energy decomposition schemes is the construction of an intermediate diabatic state where charge transfer among interacting monomers is quenched and the wavefunction is self-consistently optimized. This kind of intermediate diabatic state corresponds to a resonance structure (often the most stable) within conventional resonance theory. In such a way, a physical separation of polarization and charge-transfer effects and exploration of the charge-transfer effect on geometry, energy, charge redistribution etc. are feasible. It is also worthwhile to point out that most energy-decomposition schemes are based on the analysis of an adiabatic state wavefunction. In addition, in the BLW-ED approach, the basis set superposition error (BSSE) is fully taken into account.

Thus, in the rest of this chapter, a brief description of the development of BLW method and the extension to DFT level, as well as the theoretical background of BLW-ED scheme, will be presented first. After that are a few applications of the BLW-DFT code, such as the nature of \( \pi \)-cation interaction between a few cations and benzene, the charge transfer between solute and solvent with the supermolecular models of a positively charged ammonium and its methyl substitutes methylamines \( \text{Me}_n\text{NH}_4^{+} \) (\( n = 0 \)-3) plus a few water molecules surrounding each cation, and the theoretical study of the interchain conductivity in Poly\( (p\)-phenylene) (PPP).

5.2 Methodology

5.2.1 \textit{Ab initio} Valence Bond (VB) Theory

In VB theory, a resonance structure is constructed with chemical bonds each of which concerns only two atoms and is thus strictly localized. For a system of \( N=2n+2S \) electrons (\( n \) is the number of electron pairs and \( S \) is the spin quantum number), each resonance structure can be uniquely expressed by a Heilter-London-
Slater-Pauling (HLSP) function as

$$\Phi_K = M_K \hat{A} (\varphi_{1,2} \varphi_{3,4} \cdots \varphi_{2n-1,2n} \varphi_{2n+1} \alpha(2n+1) \cdots \phi_\alpha \alpha(N))$$ \hspace{1cm} (5.1)

where $M_K$ is the normalization constant, $\hat{A}$ is the antisymmetrizer and $\varphi_{2i-1,2i}$ is a bond function corresponding to the bond between orbitals $\phi_{2i-1}$ and $\phi_{2i}$ (or a lone pair if $\phi_{2i-1} = \phi_{2i}$)

$$\varphi_{2i-1,2i} = \hat{A} [\varphi_{2i-1} \varphi_{2i} [\alpha(i) \beta(j) - \beta(i) \alpha(j)]]$$ \hspace{1cm} (5.2)

In eq. 5.2 there are $2S$ singly occupied orbitals from $\phi_{2n+1}$ to $\phi_N$. As each bond function can be expanded into 2 Slater determinants, a HLSP comprises of $2^n$ Slater determinants. The overall many-electron wave function for a system is a linear superposition of VB functions [304]

$$\Psi = \sum_K C_K \Phi_K$$ \hspace{1cm} (5.3)

where the coefficients $\{C_k\}$ are determined by solving the secular equation $HC = ESC$. But the evaluation of the Hamiltonian and overlap matrix elements between VB functions remains a challenge (the so-called "N! problem") for ab initio VB methods due to the non-orthogonality of VB orbitals $\{\phi\}$. For instance, the Hamiltonian matrix element based on determinants is expressed as

$$\langle D | H | D \rangle = \sum_{r,s} f_{rs} D(S^r_s) + \sum_{r<u,s<t} (g_{rs,ut} - g_{rs,tu}) D(S^r_s^{st})$$ \hspace{1cm} (5.4)

where $f_{rs}$ and $g_{rs,ut}$ are one-electron and two-electron integrals respectively, and $D(S^r_s)$ and $D(S^r_s^{st})$ are the first and the second order cofactors of the overlap matrix between the two VB determinants, respectively. Over the years, several groups have developed efficient algorithms to simplify the computations of the Hamiltonian and overlap matrix elements [228-230, 232, 236, 305, 306]. One dramatic way to boost the computational efficiency of VB methods is the replacement of the bond function shown in eq. 5.2 with a doubly occupied MO-like localized orbital.
where \( \{\phi'_i\} \) is usually localized over the two bonding atoms and nonorthogonal with others [243, 244]. As such, the VB wavefunction eq. 5.1 is reduced to a single Slater determinant. Bond functions are particularly suitable for the discussion of intramolecular electron delocalization. For example, Sover et al. examined the barrier potential to internal rotation in ethane with this kind of bond-orbital wavefunctions and concluded that the dominant contribution to the barrier is the repulsion between C-H bond orbitals [244]. This form of wavefunction can also be used to study the effects of conjugation and hyperconjugation by substituting the \( \pi \) MO's in the HF wavefunction with ethylene \( \pi \) MO's derived from calculations of ethene with the same basis set [245, 246]. Apparently, the further introduction of orthogonality and delocalization over the whole system for orbitals \( \{\phi'_i\} \) leads to the much familiar HF wavefunction.

The very successful Generalized Valence Bond (GVB) method can be regarded as the hybrid use of eqs 5.2 and 5.5 in \( \Psi \) [241, 242], where the focused perfect pairs are expressed in VB form (eq 5.2), but the rest electrons are put into orthogonal and doubly occupied MO's in the form of eq. 5.5. The introduction of the strong orthogonality constraint between VB orbitals and MO's significantly reduces the computational demand for GVB calculations. The GVB method is particularly advantageous for the study of excited states and photo-dissociation pathways which cannot be well described with a single-determinant HF wavefunction.

5.2.2 Block-Localized Wavefunction (BLW) Method

A further simplification of the VB wavefunction is the use of group functions instead of bond functions by allowing the doubly occupied bond function (eq. 5.5) to
partially delocalize over a fragment of the system instead of only two bonding atoms. This kind of combination of the VB and MO theories has the remarkable advantage of using the least number of diabatic states to describe an overall chemical reaction process. For instance, in the Marcus-Hush model for a donor-acceptor system [307-310], the electron transfer (ET) process is normalized described by two electron-localized diabatic states, namely one pre-ET and one post-ET states. Since the focus is on the electron transfer from the donor to the acceptor, usually the electron delocalization within the donor or acceptor per se is not our concern and thus it is better to use one concise wavefunction instead of several VB wavefunctions for the donor or acceptor to simplify both the numerical computations and conceptual picture. But in terms of the whole donor-acceptor complex, either donor or acceptor is only a fragment, and both the pre-ET and one post-ET states need to be defined individually following the VB concepts. Putting the above considerations together, recently Mo et al. generalized the idea of localized bond functions and proposed the BLW method [59, 62, 67, 311-313]. In the BLW approach it is assumed that the overall electrons and primitive basis functions in a system are partitioned into several physically-defined subgroups, in line with the conventional VB ideas. The \(i\)th subspace consists of \(\{\chi_{\mu i}, \mu=1,2,\ldots,m_i\}\) basis functions and accommodates \(m_i\) electrons. Clearly, for a resonance structure every two electrons form a subspace. However, we extend the definition of resonance structures to diabatic states and allow a subspace to have any number of electrons. The block-localized MOs for the \(i\)th subspace \(\{\phi_{ij}, j=1,2,\ldots,m_i\}\) are expanded with \(m_i\) basis functions \(\{\chi_{\mu i}\}\)

\[
\phi_{ij} = \sum_{\mu=1}^{m_i} C_{ij\mu} \chi_{\mu i}
\]  

(5.6)

Subsequently, the BLW is defined using a Slater determinant and in the case of \(S=0\)
\[ \Phi_{r}^{BLW} = M_r \left( \mathcal{N} \right)^{-1/2} \det \left| \phi_{11}^2 \phi_{12}^2 \cdots \phi_{1n_r}^2 \phi_{21}^2 \cdots \phi_{m_r}^2 \cdots \phi_{kn_r}^2 \right| \] (5.7)

Orbitals in the same subspace are subject to the orthogonality constraint, but orbitals belonging to different subspaces are nonorthogonal. Thus, the BLW method combines the characteristics of both the MO and VB theories. For the example of an SN2 reaction \( A + BC \rightarrow AB + C \), we can define two BLWs for the reactant and product states as

\[ \psi_{r}^{BLW} = M_r \hat{\Delta} [\Phi(A) \Phi(BC)] \] (5.8a)
\[ \psi_{p}^{BLW} = M_p \hat{\Delta} [\Phi(AB) \Phi(C)] \] (5.8b)

where A and BC form two blocks (\( \Phi \) is a successive product of all occupied block-localized MOs in a block) in the reactant state \( \psi_{r}^{BLW} \) and the product state \( \psi_{p}^{BLW} \) consists of AB and C blocks. By defining the electron density matrix

\[ D = C(C^* SC)^{-1} C^* \] (5.9)

where \( S \) is the overlap matrix of the basis functions. The energy of the BLW can be determined as

\[ E^{BLW} = \langle \Psi^{BLW} | H | \Psi^{BLW} \rangle = \sum_{\mu=1}^{m} \sum_{\nu=1}^{m} d_{\mu\nu} h_{\mu\nu} + \sum_{\mu=1}^{m} \sum_{\nu=1}^{m} d_{\mu\nu} F_{\mu\nu} \] (5.10)

where \( h_{\mu\nu} \) and \( F_{\mu\nu} \) are elements of the usual one-electron and the Fock matrices, and \( d_{\mu\nu} \) is an element of \( D \).

The self-consistent optimization of orbitals in the BLW method is the key to distinguish it from other post-SCF localization methods [314-318] and can be accomplished using successive Jacobi rotation [59] or the algorithm by Gianinetti et al. [248, 249] The latter generates coupled Roothaan-like equations and each equation corresponds to a block. For the example of two blocks \( a \) and \( b \), the coefficient matrix takes the diagonal form

\[ C = \begin{pmatrix} C_a & 0 \\ 0 & C_b \end{pmatrix} \] (5.11)
where $C_a$ and $C_b$ are submatrixes. The overlap matrix $S$ can also be partitioned as

$$S = \begin{pmatrix} S_{aa} & S_{ab} \\ S_{ba} & S_{bb} \end{pmatrix}$$  \hspace{1cm} (5.12)$$

The effective overlap matrix $S'$, and effective Fock matrix $F'$ for block $a$ are defined as

$$S'_a = S_{aa} - S_{ab}D_bS_{ba}$$ \hspace{1cm} (5.13a)$$

$$F'_a = (1_a - S_{ab}D_b)F\left(-1_a^b\right) - D_bS_{ba}$$ \hspace{1cm} (5.13b)$$

The general stationary condition for each block, e.g., for $a$, is

$$\begin{align*}
F'_a C_a &= F'_a C_a L_a \\
C'_a S'_a C_a &= 1_a
\end{align*}$$ \hspace{1cm} (5.14)$$

More details on the Gianinettia et al.'s algorithm can be found in their original literature [248, 249]. Obviously, it is straightforward to extend the above two-block algorithm to cases of any number of blocks, as eq. 5.14 can be solved sequentially for each block and the rest is regarded as one block. Furthermore, the first derivative of the energy with respect to nuclear coordinates $\{q_i\}$ directly takes the form in conventional HF theory [248, 249]

$$\frac{\partial E^{\text{BLW}}}{\partial q_i} = 2 \sum_{\mu\nu} d_{\mu\nu} \frac{\partial h_{\mu\nu}}{\partial q_i} + \sum_{\mu\nu\rho\sigma} \left[2d_{\mu\nu}d_{\rho\sigma} - d_{\mu\rho}d_{\nu\sigma}\right] \frac{\partial (\mu\nu|\rho\sigma)}{\partial q_i} - 2 \sum_{\mu\nu} W_{\mu\nu} \frac{\partial S_{\mu\nu}}{\partial q_i}$$ \hspace{1cm} (5.15)$$

where $W_{\rho\sigma}$ is a Lagrangian variable. With the first derivatives derived analytically, the second derivatives can be computed numerically.

Mo et al. have written an independent BLW code at the HF level with high efficiency, and numerous applications endorse its usefulness. For instances, the BLW method had been used to study the charge transfer in the prototype of donor-acceptor complexes BH$_3$NH$_3$ [63], probe the nature of the ethane rotation barrier [319, 320] and the cation-π interactions in δ-opioid receptor binding [321], propose an energetic measure of aromaticity and antiaromaticity based on the Pauling-Wheland resonance
energies [67], and analyze the charge transfer between solute and solvent with up to 1202 basis functions. [69]

However, since VB theory focuses on individual atoms and atomic orbitals, \textit{ab initio} VB methods and the BLW method may not work well if the basis functions lose atomic characteristics, e.g., when a complete basis on a single center for a molecular system is used. As a matter of fact, this unphysical basis set artifact complicates not only \textit{ab initio} VB methods but also MO-based analyses on atomic properties in molecules, and in reality one can use basis functions optimized for individual atoms. Thus, the BLW method is applicable with regular basis sets and with the currently popular basis sets from 6-31G(d) to 6-311+G(d,p) and cc-pVTZ, previous works showed that the basis set dependence is generally trivial for the BLW method. [59, 62, 68, 322]

5.2.3 BLW Method at the Density Functional Theory (DFT) Level

Due to the low computational costs and incorporation of (at least partial) electron correlation, DFT methods provide a sound basis for the development of computational strategies for studying potential energy surfaces, dynamics, various response functions and spectroscopy, excited states, and many more [323]. Although there are several known deficiencies in DFT, e.g., DFT is less accurate for weak interactions (non-bonded interactions) and \(\pi\)-bonded systems, significant and persistent efforts have been put forth to develop new functionals, particularly the critical exchange and correlation functional [324-326]. For the sake of simplicity, approximate dispersion corrections, e.g., a \(C_6/r^6\) term, can be added to DFT calculations directly. [327-329]

In DFT, the self-consistent Kohn-Sham (KS) procedure is strictly analogous to
the Hartree-Fock-Roothaan SCF procedure, except that the HF exchange potential is replaced by a DFT exchange-correlation (XC) potential. And the orbital equations of DFT have the same forms as those in HF theory except with a different Fock matrix

$$ F^\alpha = H + J + F^{\text{XC}\alpha} \quad \text{(5.16)} $$

where $H$ is the one-electron Hamiltonian matrix and $J$ is the Coulomb matrix. The elements of $\alpha$ exchange-correlation matrix $F^{\text{XC}\alpha}$ can be evaluated by a one-electron integral involving the local electron spin densities (LSD methods), or by an integral involving electron densities and their gradients (GGA methods). Thus, it is fairly straightforward to implement the BLW idea into DFT as long as we keep all the equations (eqs 5.6-5.15) unchanged except that the Fock matrix therein is replaced by a DFT one ($F^{\text{XC}\alpha}$). Recently, we extended the BLW method to the DFT level by adopting the block-localized orbitals in the KS-DFT procedure, and the implementation of the BLW-DFT method consists of the following steps:

(i) Construct the DFT Fock matrix and calculate the DFT energy.

(ii) Construct the effective Fock and overlap matrices for each block.

(iii) Solve the generalized secular equations and subsequently form the new coefficient and density matrix.

(iv) Check the variation of the density matrix. Go back to the first step if convergence is not reached; otherwise, print out the final outcome and compute various properties.

Because achieving a self-consistent field with DFT is usually more difficult than with the normal HF method, Pulay's DIIS technique is used to update the Fock matrix and accelerate the convergence [330]. The fluctuation of density matrix in the process of iteration will be taken as the error vector. Fortunately, the present version of GAMESS has the capability of performing DFT calculations [65]; thus we used GAMESS as a platform to implement the BLW-DFT method.
The formulation of the BLW method for open-shell systems is quite similar to eqs 5.6-5.15 where the doubly occupied orbitals are replaced with singly occupied spin-orbitals. In other words, we replace the Fock matrix with the $\alpha$ and $\beta$ Fock matrices in the restricted or restricted open-shell self-consistent equations. The current version of BLW-DFT works for both closed-shell and open-shell systems.

5.2.4 Energy Decomposition Analysis based on BLW (BLW-ED)

A number of energy decomposition schemes have been proposed and are available in most quantum chemistry software. The primary difference of these methods stems from the implementation of the computational procedure.

In our analysis, we make use of the BLW approach, which allows a convenient separation of contributing energy terms. When we take each block for a monomer, the BLW method can be straightforwardly applied to the analysis of molecular interactions, which not only provides insights into the nature of forces binding molecules together, but also affords clues to the force field development. With the inclusion of the correction for basis set superposition error (BSSE), the energy decomposition scheme (BLW-ED) can be developed to decompose the molecular interaction energy ($\Delta E_{\text{int}}$) into the energy terms such as Heitler-London energy ($\Delta E_{\text{HL}}$), polarization energy ($\Delta E_{\text{pol}}$) and charge transfer energy ($\Delta E_{\text{CT}}$)

$$
\Delta E_{\text{int}} = E(\Psi_{AB}) - E(\Psi_A^0) - E(\Psi_B^0) + \text{BSSE} = \Delta E_{\text{HL}} + \Delta E_{\text{pol}} + \Delta E_{\text{CT}}
$$

(5.17)

where $\Psi_{AB}$, $\Psi_A^0$ and $\Psi_B^0$ are the wave functions for the dimer AB, monomers A and B, respectively.

The derivation of these individual energy terms is based on the construction of the initial block-localized wave-function for the dimer $\Psi_{AB}^{\text{BLW0}}$ as well as its self-consistent form $\Psi_{AB}^{\text{BLW}}$. as
In a BLW as eq. 5.18, the orbitals belonging to either monomer A or B are constrained to be mutually orthogonal, as in conventional MO methods, while those belonging to different monomers are nonorthogonal, as in VB methods. The self-consistent optimization of the block-localized orbitals in eq. 5.18b can be accomplished using successive Jacobi rotation as initially adopted, or using Gianinetti et al.'s algorithm. Of significance, Gianinetti and coworkers demonstrated that the self-consistent-field (SCF) solution of a wavefunction like BLW can be decomposed to coupled Roothaan-like equations and each equation corresponds to a monomer.

With the above definitions of initial and optimal BLWs in eq. 5.18, the energy terms in eq. 5.17 can subsequently be expressed as

\[
\Delta E_{\text{HL}} = E(\psi_{AB}^{\text{BLW}_0}) - E(\psi_A^0) - E(\psi_B^0) \quad (5.19a)
\]

\[
\Delta E_{\text{pol}} = E(\psi_{AB}^{\text{BLW}}) - E(\psi_{AB}^{\text{BLW}_0}) \quad (5.19b)
\]

\[
\Delta E_{\text{CT}} = E(\psi_{AB}^{\text{HF}}) - E(\psi_{AB}^{\text{BLW}}) + \text{BSSE} \quad (5.19c)
\]

The Heitler-London energy (eq. 5.19a) is defined as the energy change by bringing monomers together without disturbing their individual electron densities, while the polarization energy (eq. 5.19b) corresponds to the stabilization of the complex due to the mutual relaxation of individual electron densities. In this polarization step, however, there is no penetration of electrons between two monomers. The extension of electron movements from block-localized orbitals to the whole complex further stabilized the complex and this energy variation is denoted as the charge-transfer energy (eq. 5.19c). In this step, the BSSE is also introduced, thus the correction is completely assigned to the charge-transfer energy term. It should be noted that \( \Delta E_{\text{HL}} \) is a sum of electrostatic and Pauli-exchange repulsion energies. Since
the exchange of electrons is a quantum mechanical effect and classical force field approaches have difficulties to formulate the exchange energy separately, here we simply use $\Delta E_{\text{HL}}$ as the electrostatic energy. (Scheme 5.1)

\[ \Delta E_{\text{ex}} = \langle \Psi_{A}^{0} \mid \hat{H} \mid \Psi_{B}^{0} \rangle - E(\Psi_{A}^{0}) - E(\Psi_{B}^{0}) \]

\[ \Delta E_{\text{ex}} = E(\Psi_{AB}^{\text{BLW0}}) - \langle \Psi_{A}^{0} \mid \hat{H} \mid \Psi_{B}^{0} \rangle - E(\Psi_{A}^{0}) - E(\Psi_{B}^{0}) \]

\[ \Delta E_{\text{HL}} = \Delta E_{\text{ex}} + \Delta E_{\text{ex}} \]

\[ \Delta E_{\text{ex}} = E(\Psi_{AB}^{\text{BLW}}) - E(\Psi_{AB}^{\text{BLW0}}) \]

\[ \Delta E_{\text{ct}} = E(\Psi_{AB}^{\text{HF}}) - E(\Psi_{AB}^{\text{BLW}}) + BSE \]

**Scheme 5.1** Energy decomposition scheme in our BLW-ED approach.

### 5.3 Applications of BLW Method

The BLW method at both the HF and DFT levels has been implemented and ported to the general *ab initio* quantum chemistry package GAMESS software, and the code has the geometry optimization capabilities [311]. The BLW method can not only evaluate the Pauling-Wheland resonance energy in conjugated systems [59, 67], but also explore the nature of intermolecular interactions and decompose the interaction energy in terms of Heitler-London, polarization, and electron transfer.
energy terms, where the Heitler-London energy term can be further decomposed to electrostatic and Pauli exchange interactions [62, 63, 68-70]. In the following we will present a few preliminary applications of the BLW-DFT code to the nature of $\pi$-cation interaction between a few cations and benzene, the charge transfer between solute and solvent with the supermolecular models of a positively charged ammonium and its methyl substitutes methylamines $\text{Me}_n\text{NH}_4^{n+}$ ($n = 0-3$) plus a few water molecules surrounding each cation, and the theoretical study of the interchain conductivity in Poly($p$-phenylene) (PPP).

5.3.1 $\pi$-Cation Interactions between Cations and Benzene

As a pilot test for intermolecular interactions, we studied a kind of extremely strong non-covalent interaction, namely $\pi$-cation interaction which even can compete with full aqueous solvation in binding cations. $\pi$-cation interaction plays a key role in biological recognition [331-336], where cations such as simple $\text{Na}^+$ or complex acetylcholine (ACh) binds aromatic components from the amino acids Phe, Tyr and Trp. The elucidation of the nature of this specific interaction will be especially helpful for the understanding of the mechanisms of enzymatic catalysis and ion channels.

We choose benzene as the $\pi$ aromatic system to interact with cations. Although benzene is a non-polar molecule, it has a quadrupole moment, and Dougherty assumed that the electrostatic interaction between the cation and the quadrupole charge distribution of the aromatic is of prime importance in the $\pi$-cation interactions, while additional terms such as induced dipoles, polarizabilities, dispersion forces and charge transfer should be included to quantitatively model the cation-$\pi$ interactions [336-338]. Kollman and coworkers showed the molecular mechanical model with polarizability can model $\pi$-cation interaction energies better.
than two-body additive models [339]. Using a perturbation approach, Cubero et al. explored the importance of cation→aromatic polarization effects on cation-π interactions and found that the polarization energy is 70% the magnitude of the electrostatic energy at the optimal Na⁺-benzene distance of 2.47 Å [340].

To elucidate the origin of π-cation forces, we investigated the interactions between a few simple cations (M⁺ = Li⁺, Na⁺, K⁺, N⁴⁺ and N(CH₃)₄⁺, as shown in Figure 5.1) and the prototypical aromatic system, benzene, with the energy decomposition scheme based on the BLW method [62, 63, 68-70].

Our calculations on the interactions between cations and benzene are performed with geometries optimized at the MP2/6-311G(d,p) level, and subsequent BLW energy analyses are conducted at the B3LYP/6-311G(d,p) level. Table 5.1 compares various energy contributions (electrostatic, polarization and charge-transfer) to the interaction energies ΔE_{int} at the DFT. For comparison, the MP2 interaction energies are also listed to evaluate the residual electron correlation (dispersion energy) which is left out in the DFT calculations. It should be noted that geometries and binding energies for alkali-metal cation complexes with benzene have been extensively studied by Nicholas et al. at various levels [341], and our optimizations produced similar results. For instance, the distances between the cation and the center of benzene are 1.88, 2.42 and 2.79 Å for Li⁺, Na⁺ and K⁺ at the MP2/6-311G(d,p) level, respectively, and data for N⁴⁺ and N(CH₃)₄⁺ are 2.90 and 4.22 Å, respectively. Nicholas et al. also pointed out that MP2 results are well converged with regard the extent of electron correlation [341].
Figure 5.1 Geometries for the π-cation complexes: (a) M⁺(C₆H₆) (M = Li, Na, K); (b) NH₄⁺(C₆H₆); (c) N(CH₃)₄⁺(C₆H₆).

<table>
<thead>
<tr>
<th>Complex</th>
<th>$\Delta E_{\text{HL}}$</th>
<th>$\Delta E_{\text{pol}}$</th>
<th>$\Delta E_{\text{CT}}$</th>
<th>$\Delta E_{\text{int}}^\text{B3LYP}$</th>
<th>$\Delta E_{\text{int}}^\text{MP2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li⁺(C₆H₆)</td>
<td>-8.3</td>
<td>-21.9</td>
<td>-8.9</td>
<td>-39.1(-40.2)</td>
<td>-35.8(-40.1)</td>
</tr>
<tr>
<td>Na⁺(C₆H₆)</td>
<td>-9.6</td>
<td>-11.6</td>
<td>-3.0</td>
<td>-24.2(-25.4)</td>
<td>-21.8(-25.2)</td>
</tr>
<tr>
<td>K⁺(C₆H₆)</td>
<td>-5.4</td>
<td>-7.8</td>
<td>-3.3</td>
<td>-16.6(-17.4)</td>
<td>-17.5(-20.3)</td>
</tr>
<tr>
<td>NH₄⁺(C₆H₆)</td>
<td>-2.9</td>
<td>-7.9</td>
<td>-5.7</td>
<td>-16.5(-17.1)</td>
<td>-18.0(-19.7)</td>
</tr>
<tr>
<td>N(CH₃)₄⁺(C₆H₆)</td>
<td>-0.8</td>
<td>-2.4</td>
<td>-2.4</td>
<td>-5.6(-6.2)</td>
<td>-9.2(-11.1)</td>
</tr>
</tbody>
</table>

a. Data in parentheses are derived without taking the BSSE effect into account.

As listed in Table 5.1, the Heitler-London energy decreases in the order of Na⁺ > K⁺ > NH₄⁺ > N(CH₃)₄⁺, in inversely proportional to the distance between the cation and benzene. But Li⁺ is an exception as the Heitler-London energy with
benzene is lower than Na+. Further analyses reveal that this abnormality comes from the strong Pauli exchange repulsion due to the short distance between Li+ and benzene. Overall, \( \Delta E_{\text{HL}} \) accounts for only 21% (Li+), 40% (Na+), 33% (K+), 17% (NH4+) and 14% (N(CH3)4+) of the interaction energies. This finding is in accord with the failure of previous force field studies based on a pure electrostatic model [339], although the latter does provide correct qualitative ordering for the interaction with aromatic compounds. [337]

Notably, our energy decomposition analysis highlights the importance of the polarization effect, which almost solely comes from the aromatic benzene. For the present cation-benzene complexes, the polarization energy contributes about 50% to the interaction energies, and decreases in the order of Li+ > Na+ > K+ ≈ NH4+ > N(CH3)4+. Like the electrostatic force, the polarization effect decreases with increasing distance between the cation and benzene. Our calculations support previous arguments that the explicit inclusion of polarization in molecular interaction potential is essential to the modeling of the \( \pi \)-cation interactions [339, 340, 342, 343]. Without the inclusion of the polarization effect, even modified OPLS reproducing the quadrupole moment of benzene leads to the Li+-benzene complex enthalpy only -25.3 kcal/mol [339]. Cubero et al. estimated the polarization stabilization energy -9.9 kcal/mol for the interaction of Na+ with benzene [340], which is in good agreement with our result (-11.6 kcal/mol). To explore the origin of the polarization effect, we evaluated the individual polarization energies of the cation, \( \sigma \) part and \( \pi \) part of benzene, and found that the polarization effect is actually dominated by the hybridization of the \( \sigma \) and \( \pi \) parts of benzene.

The polarization of benzene can be visualized by the electron density difference (EDD) between the BLW for the complex and the sum of individual
monomers. Figure 5.2 shows the polarization of benzene in the electrical field of K$^+$ and NH$_4^+$, where the red means the gain of electron density and the blue refers to the loss of electron density. Apparently, the polarization results from the $\sigma \rightarrow \pi^*$ excitation, and the overall effect is the electron density shift from hydrogen ($\sigma$ orbitals) to carbon ($\pi$ orbitals). Other cations have a similar effect; the field effect decreases in order Li$^+ >$ Na$^+ >$ K$^+ >$ NH$_4^+ >$ N(CH$_3$)$_4^+$, in accord with the polarization energies.

Kollman and coworkers' nonadditive model took the polarization effect into account and got the enthalpies close to both experimental and quantum mechanical data. However, the distance from the cation to benzene center is noticeably underestimated by about 0.2 Å in all cases. Adjusting the three-body potential may result in good distances, but enthalpies will be underestimated. For the example of Li$^+(C_6H_6)$, the enthalpy is -32.1 kcal/mol when the $\pi$-cation distance is 1.9Å. The dilemma mainly lies in the omission of the charge transfer effect in their modeling, as Kollman and coworkers assumed. Our analyses endorsed their assumption, and particularly for Li$^+$, the charge transfer stabilizes the Li$^+(C_6H_6)$ complex by 8.9 kcal/mol, which accounts for 23% of the total interaction energy. For other cations, the charge transfer effect is not as prominent as Li$^+$, but still noticeable, particularly for NH$_4^+$ and N(CH$_3$)$_4^+$. The charge transfer effect can be visualized by the EDD maps between BLW and DFT wave functions, as shown in Figure 5.1 for the cases of Li$^+(C_6H_6)$ and NH$_4^+(C_6H_6)$ complexes. The charge transfer mainly occurs from carbon atoms in benzene to Li$^+$ or the protons in NH$_4^+$ pointing toward benzene.
Figure 5.2 Electron density difference (EDD) maps: (a1) and (b1) show the polarization effect in the \( \text{Li}^+\text{C}_6\text{H}_6 \) and \( \text{NH}_4^+\text{C}_6\text{H}_6 \) complexes (isodensity \( 3 \times 10^{-3} \) and \( 2 \times 10^{-3} \) a.u., respectively); (a2) and (b2) show the charge transfer effect in \( \text{Li}^+\text{C}_6\text{H}_6 \) and \( \text{NH}_4^+\text{C}_6\text{H}_6 \) (isodensity \( 1 \times 10^{-3} \) a.u.).

The comparison between the DFT and MP2 interaction energies in Table 5.1 indicates that the counter-poise method [344] may remarkably overestimate the BSSE correction for MP2 energies.[345] The B3LYP calculations result in a BSSE correction of about 1 kcal/mol, but the correction at the MP2 level is 4.3 kcal/mol for \( \text{Li}^+\text{C}_6\text{H}_6 \) and then decreases with the increasing of the \( \pi \)-cation distance (or the weakening of the \( \pi \)-cation interaction). However, we still can envision that the dispersion energy plays a noticeable role in the interactions at least between \( \text{K}^+ \), \( \text{NH}_4^+ \) or \( \text{N(CH}_3)_4^+ \) and benzene.
5.3.2 Charge Transfer in the Solvation of MeₙNH₄⁺ (n = 0, 1, 2, and 3)

Since most chemical reactions and biological processes occur in solution, the simulation of solvent effects has been one of the most active research fields in computational chemistry and significant progresses have been made in both implicit and explicit solvation models [226, 346-351]. In implicit solvation models, a polarizable solvent is efficiently treated as a continuous homogeneous dielectric [347], but the strong and specific solute-solvent interactions, e.g., hydrogen bonding which is a directional short-range force, are not completely accounted for. In explicit solvation models, solvent molecules are usually defined explicitly at the molecular mechanical (MM) level, while the solute at either the same MM level or more advanced quantum mechanical (QM) level. A critical component in the explicit solvation models is the intermolecular potential function that describes intermolecular interactions in the condensed phase, and ultimately determines the success of computer simulations [53-55]. With the recognition of the importance of the solvent polarization effect in solute-solvent interactions, polarizable force fields where explicit polarization terms are added in the potential energy function have been proposed and developed [56-58, 352-356]. However, there have been controversies over the magnitude of charge transfer between solute and solvent molecules [69, 300, 357-360]. We note that the controversies mostly originate from the various definitions of the charge transfer energy term in numerous energy decomposition schemes [361-367]. The uniqueness of our BLW energy decomposition method lies in the construction of an intermediate diabatic state where charge transfer is deactivated and the corresponding wavefunction is self-consistently optimized. Using such a diabatic state as a reference, both the polarization and charge transfer effects can be distinctly differentiated.
Most recently, we performed combined QM/MM simulations on the solvation of two simple ionic systems, acetate and methylammonium, in a water box, followed by BLW energy decomposition analyses at the HF level on a few randomly selected configurations where the first and second hydration shells of water molecules are included in the QM part. We found that the charge transfer term only makes a small fraction of the total solute-solvent interaction energy [69]. However, we note that the force field used in the simulation is nonpolarizable and as a consequence, the distance between the solute and the first hydration shell may be more or less lengthened as the short-range polarization interaction has been diluted to the long-range electrostatic interaction by adjusting the atomic partial changes in nonpolarizable force fields. For instance, the radial distribution function showed the peak of the average acetate oxygen or methylammonium nitrogen and water oxygen in the first hydration shell at 2.95 Å or 2.85 Å, compared with 2.75 Å and 2.85 Å from Car-Parrinello simulations with plan-wavefunction DFT by Dal Peraro et al. [360] Since the charge transfer is very sensitive to the distance and increases in an exponential pattern, it would be of general interests to derive the solute-solvent configurations at the \textit{ab initio} level.

Here we estimated the charge transfer effect in the solvation of ammonium and its methyl substitutes with supermolecular models Me$_n$NH$_4$$_n$+…(H$_2$O)$_{4-n}$ (n = 0, 1, 2, and 3), where each N-H group forms a hydrogen bond with a water molecule. The BLW-DFT calculations and analyses are conducted at the geometries optimized at the MP2/6-311+G(d,p) level. Results are summarized in Table 5.2, where the MP2 interaction energies are also listed for comparison.
Table 5.2

BLW-DFT energy decomposition analyses on the interaction between Me\textsubscript{n}NH\textsubscript{4-n}\textsuperscript{+} and water with the 6-311+G(d,p) basis set (kcal/mol)

<table>
<thead>
<tr>
<th>Complex</th>
<th>$\Delta E_{H\text{L}}$</th>
<th>$\Delta E_{pol}$</th>
<th>$\Delta E_{\text{CT}}$</th>
<th>$\Delta E_{\text{int}}^{\text{B3LYP}}$</th>
<th>$\Delta E_{\text{int}}^{\text{MP2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH\textsubscript{4}(H\textsubscript{2}O)\textsubscript{4}</td>
<td>-50.8</td>
<td>-11.5</td>
<td>-7.7</td>
<td>-70.0(-74.0)</td>
<td>-67.5(-74.5)</td>
</tr>
<tr>
<td>MeNH\textsubscript{3}(H\textsubscript{2}O)\textsubscript{3}</td>
<td>-34.1</td>
<td>-9.4</td>
<td>-6.6</td>
<td>-50.1(-53.2)</td>
<td>-48.7(-54.4)</td>
</tr>
<tr>
<td>Me\textsubscript{2}NH\textsubscript{2}(H\textsubscript{2}O)\textsubscript{2}</td>
<td>-20.2</td>
<td>-6.8</td>
<td>-5.2</td>
<td>-32.2(-34.2)</td>
<td>-31.6(-35.7)</td>
</tr>
<tr>
<td>Me\textsubscript{3}NH\textsuperscript{+}(H\textsubscript{2}O)</td>
<td>-8.8</td>
<td>-3.7</td>
<td>-3.1</td>
<td>-15.6(-16.6)</td>
<td>-15.5(-17.8)</td>
</tr>
</tbody>
</table>

a. Data in parentheses are derived without taking the BSSE effect into account.

Due to the very small size of solvent molecules in the present models, the total solute-solvent interaction energies are much lower than the true cases. For instance, the QM/MM data for the solvation of MeNH\textsubscript{3}\textsuperscript{+} is -122.0 kcal/mol [63], while the present model gave only -50.1 kcal/mol. We believe the difference mostly comes from the long-range electrostatic interactions, plus a small portion from the solvent polarization. With the reduction of the water molecule number in the models, the solute-solvent interaction energy decreases dramatically from -70.0 kcal/mol in NH\textsubscript{4}(H\textsubscript{2}O)\textsubscript{4} to only -15.6 kcal/mol Me\textsubscript{3}NH\textsuperscript{+}(H\textsubscript{2}O), indicating the inappropriateness of these models to study the solvation of ions. However, we note that our objective here is to evaluate the contribution from the charge transfer effect to the solute-solvent interactions, rather than get accurate solvation energies. The current calculations do confirm that the permanent electrostatic energy dominates the solute-solvent interactions, and polarization effect plays the secondary role. The charge-transfer energy is comparable in magnitude with the polarization effect, nevertheless we expect that the inclusion of more water molecules in models will remarkably...
increase both the Heitler-London and polarization energies, but retain the charge transfer energy at the current level. More extensive studies currently are still under way.

If we focus on individual hydrogen bonds in the four systems, the charge transfer energy for each N-H⋯OH₂ bond is -1.9, -2.2, -2.6 and -3.1 kcal/mol, in good correlation with the hydrogen bond distance Rₕ₋ₒ = 2.860, 2.840, 2.818 and 2.792 Å in the four optimal models. Similar to previous work, we also probed individual polarization contributions from the solute and solvent separately and Table 5.4 lists the polarization energy of the solute by the solvent charge density in the absence of the solute, and the polarization energy of the solvent by the solute permanent (gas phase) charge density. Due to the coupling effect, the sum of individual polarization energies is slightly lower than the total polarization energy listed in Table 5.2. But Table 5.3 demonstrates that the solvent polarization effect is far more significant than the solute polarization effect. Figure 5.3(c1) plots that the solvent polarization shifts the electron density from the O-H σ bond to the oxygen side, and this shifted electron density will be subsequently donated to the protons in ammonium ion as manifested by Figure 5.3(c2).

Table 5.3
Individual polarization energies (kcal/mol)

<table>
<thead>
<tr>
<th>Complex</th>
<th>ΔEₚₒˡ(M⁺)</th>
<th>ΔEₚₒˡ(water)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₄⁺(H₂O)₄</td>
<td>-0.4</td>
<td>-11.1</td>
</tr>
<tr>
<td>MeNH₃⁺(H₂O)₃</td>
<td>-1.3</td>
<td>-7.7</td>
</tr>
<tr>
<td>Me₂NH₂⁺(H₂O)₂</td>
<td>-1.4</td>
<td>-4.9</td>
</tr>
<tr>
<td>Me₃NH⁺(H₂O)</td>
<td>-0.9</td>
<td>-2.4</td>
</tr>
</tbody>
</table>
We can further conduct population analyses on the electron densities derived by both the conventional DFT and BLW-DFT computations and take the differences as the charge transferred between the solute and water. Table 5.4 compiled the amount of charge transferred in the Me$_n$NH$_4$$^+$($\text{H}_2\text{O}$)$_4$ complexes based on three population analysis schemes, namely Mulliken, Löwdin and natural population analysis (NPA) [368, 369]. We can see the latter two derive very similar results, and excellent correlation between the electron transfer energies (Table 5.2) and population analyses can be found as shown in Figure 5.4.
Table 5.4

Amount of electrons transferred from water to $\text{Me}_n\text{NH}_{4+n}^+$ based on various population analysis schemes

<table>
<thead>
<tr>
<th>Complex</th>
<th>Mulliken</th>
<th>Löwdin</th>
<th>NPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{NH}_4^+(\text{H}_2\text{O})_4$</td>
<td>0.041</td>
<td>0.060</td>
<td>0.063</td>
</tr>
<tr>
<td>$\text{MeNH}_3^+(\text{H}_2\text{O})_3$</td>
<td>0.050</td>
<td>0.053</td>
<td>0.052</td>
</tr>
<tr>
<td>$\text{Me}_2\text{NH}_2^+(\text{H}_2\text{O})_2$</td>
<td>0.048</td>
<td>0.043</td>
<td>0.039</td>
</tr>
<tr>
<td>$\text{Me}_3\text{NH}^+(\text{H}_2\text{O})_2$</td>
<td>0.033</td>
<td>0.021</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Figure 5.4 Correlation between the charge transfer stabilization energy and the amount of NPA charge transferred from water molecules to the cations.
### 5.3.3 Theoretical Study of the Interchain Conductivity in Poly(p-phenylene)

π-conjugated polymers have received intensive attentions due to their applications in chemical sensors and molecular electronic and photonic devices [370-376]. One of the major classes of π-conjugated polymers is poly(p-phenylene) (PPP) which is associated with their application as the organic component in LED based devices [377, 378]. Due to the limited length and disorder of polymer strands, charge transfer between polymer strands resulted from interchain interactions plays an essential role in charge transport across macroscopic distances [379-381], although interchain transport (or “inter-soliton” transfer [382]) is expected to be significantly slower than intrachain transport [383]. As the charge carriers are expected to be localized over a single unit (e.g., benzene for PPP), the transport mechanism can be described in terms of polaronic hopping between adjacent chains [384-386]. For the instance of PPP, interchain electron or hole transfer can be modeled with two adjacent benzenes which are either neutral or charged [387, 388]. Scheme 5.2 shows a model with two parallel benzene rings.

![Scheme 5.2](image)

**Scheme 5.2** Two parallel benzene rings used in our studies.

Based on the Marcus-Hush model [307-310], two diabatic states where the charge is localized on either benzene (as shown in Scheme 5.2), can be used to describe the above charge transfer process or self-exchange reaction. With the harmonic approximation, the potential energy for each diabatic state is represented by
a parabolic curve and the adiabatic (ground) state is characterized by a double-well potential and the polaronic hopping between the two minima requires overcoming an activation barrier. The most significant issue in this model is the seam of the crossing between the two energy profiles of diabatic states, where charge transfer occurs in line with the Frank-Condon principle. Due to electronic coupling largely dictated by orbital overlap, the ground state energy will be lowered by $V_{AB}$ compared with the diabatic state energies. The significance of $V_{AB}$ lies in its direct relevance to the interchain transfer rate as [389, 390]

$$k_{ET} = \frac{1}{\hbar} |V_{AB}|^2 \left( \frac{\pi}{\lambda k_B T} \right)^{1/2} \exp \left( -\frac{(\Delta_r G^0 + \lambda)^2}{4\lambda k_B T} \right)$$  \hspace{1cm} (5.20)

where $\lambda$ is the reorganization energy owing to the geometric relaxation accompanying charge transfer and $\Delta_r G^0$ is the overall reaction free energy which is zero in present charge self-exchange reactions.

The major challenge therefore is to compute the electronic coupling term $V_{AB}$ (also called transfer integral [391]) and reorganization energy $\lambda$ which center on the electron transfer theory. Much progress has been made during the past decade in computing the coupling term. Among various approaches, the energy-splitting method based on Koopmans’ theorem provides the simplest way to evaluate $V_{AB}$ as half of the energies of the appropriate occupied and unoccupied orbitals [384, 391-395], e.g., half of the energy difference between the HOMO and HOMO-1 of a system with an extra electron. Another direct method to compute the couplings is based on the energy difference between the ground and the first excited states [396]. Newton and coworkers have developed the generalized Mulliken-Hush (GMH) formalism and showed that $V_{AB}$ can be estimated from the transition dipole moment between the states of interests and the diabatic states can be defined by diagonalizing the dipole moment matrix [397-400]. Other approaches include the fragment charge difference
procedure [401], tuning the self-consistent-field (SCF) energy difference with external perturbation [391], perturbed ground state method [402], the block diagonalization method [398, 403], or methods in terms of partitioning theory [404] or Green functions [405].

However, the above approaches are indirect as they share a common feature that the electronic coupling between the diabatic states stems from chosen adiabatic states. In other words, the diabatic states themselves are unknown and thus not directly accessible. An accurate calculation of $V_{AB}$ requires the unambiguous definition of diabatic states with the transported electron localized. Although in general this remains a challenge in molecular orbital (MO) theory due to the delocalization nature of MOs, the valence bond (VB) theory is established on the philosophy that any molecular state is a superposition of certain charge-localized resonating (diabatic) states and each diabatic state can be well described by a Heitler-London-Slater-Pauling (HLSP) function [406]. Unfortunately, high computational costs resulting from the non-orthogonality of orbitals severely limit the wide applications of ab initio VB methods. Thus, with the ability to combine the characters and advantages of both the MO and VB theories [59, 64, 311, 407], the BLW method provides a good choice to study the two-state model.

For the two states in Scheme 5.2, their wave functions can be defined as

$$
\psi_A = \hat{A} \Phi_1(C_6H_5^-)\Phi_2(C_6H_6)
$$

(5.21a)

$$
\psi_B = \hat{A} \Phi_1(C_6H_6)\Phi_2(C_6H_6^+)
$$

(5.21b)

where $\Phi$ denotes a successive product of occupied orbitals which are expanded in only one benzene orbital space. Once we derive the wave functions and energy profiles for the two electron-localized diabatic states, we can easily derive the electronic coupling energy $V_{AB}$ by resolving the following $2 \times 2$ secular equation.
where $H$ and $S$ are the Hamiltonian and overlap matrix, respectively, and $E$ is the energy eigenvalue and has two roots $E_1$ and $E_2$. At the crossing point of energy curves of diabatic states when their energies are identical ($H_{AA}=H_{BB}$), the electronic coupling strength $V_{AB}$ is defined as the difference between the diabatic state energy and the lower root of eq. 5.22

$$V_{AB} = H_{AA} - E_1 = \frac{H_{AA} S_{AB} - H_{AB}}{1 + S_{AB}}$$

(5.23)

Note that $V_{AB}=H_{AB}$ when the overlap between the two diabatic states is neglected.

We employed the above approach to study the self-charge transfer reactions between an ionic benzene radical and a neutral benzene. We first optimized the geometries of monomers at the restricted open HF (ROHF) level, where the neutral benzene adopts a $D_{6h}$ symmetry and ionic radicals are imposed a $D_{2h}$ symmetry. Table 5.5 listed the optimal bond lengths.

**Table 5.5**

Optimal bond lengths in benzene, benzene cationic radical and benzene anionic radical

<table>
<thead>
<tr>
<th>Bond length (Å)</th>
<th>$C_6H_6$ ($D_{6h}$)</th>
<th>$C_6H_6^+$ ($D_{2h}$)</th>
<th>$C_6H_6^-$ ($D_{2h}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(C-C)</td>
<td>1.388</td>
<td>1.379/1.446</td>
<td>1.370/1.428</td>
</tr>
<tr>
<td>R(C-H)</td>
<td>1.076</td>
<td>1.074/1.073</td>
<td>1.081/1.077</td>
</tr>
</tbody>
</table>
As the charge transfer reactions accompany the rearrangement of nuclear conformations and the energy change is related to the reorganization energy $\lambda$, the reaction coordinate $Q_i$ is defined as a linear combination of reactant coordinates $Q_A$ and product coordinates $Q_B$.

$$Q_i = (1-\alpha)Q_A + \alpha Q_B$$

(5.24)

We first study the charge transfer reactions along the variation of the donor-acceptor distance with the transition state geometry ($\alpha = 0.5$). Figure 5.5 shows the binding energy profiles. The minima derived by the two-state model are located at $R = 3.5$ Å and 4.1 Å for the cationic and anionic complexes, respectively. For comparison, the distance between layers in graphite is 3.37 Å [408]. Apparently, the attractive interaction in the cationic complex (48 kJ/mol) is much stronger than in the anionic complex (10 kJ/mol).

![Figure 5.5](image)

**Figure 5.5** Binding energy potentials between a neutral benzene and (a) an anionic benzene and (b) a cationic benzene radical.
It is predicted that $V_{AB}$ decays exponentially with respect to the distance between the donor and acceptor [399, 409]

$$V_{AB} \propto e^{-\beta R_{AB}/2}$$

(5.25)

Figure 5.6 plots the correlation between the natural logarithm of $V_{AB}$ and R. In both the anionic and cationic complexes, we observed a perfect linear relationship with correlation factors close to 1. The exponential decay constants ($\beta$) are 1.32 and 2.76 Å$^{-1}$, respectively, indicating higher electrical conductivity in the electron transfer reaction than in the hole transfer reaction. For comparison, electron transfer across linear alkyl chains polyethylene decay at rates of 1.0 and 0.2 Å$^{-1}$, respectively [250]. Thus, interchain charge transport is significantly slower than intrachain transport [383]. Experimentally, the exponential decay constants for electron tunneling through a frozen toluene glass and a frozen 2-methyl-tetrahydrofuran glass are measured to be 1.23 and 1.62 Å$^{-1}$, respectively [410]. Our computation of the interchain electron transfer in PPP is in good agreement with the experimental data in frozen glasses.

Figure 5.6 Correlation between the electronic coupling energy ($V_{AB}$ in eV) and the electron-transfer distance between a neutral benzene and (a) an anionic benzene and (b) a cationic benzene radical.
The BLW method can also quantitatively compute the reorganization energies. By incorporating the BLW method with Monte Carlo simulation method, the important solvent reorganization energy can be computed [226]. Here, however, we focus on the inner-sphere (solute) reorganization energy. We studied the charge transfer processes with the reaction coordinate $Q_i$ (eq. 5.24) by gradually changing $\alpha$ from 0 to 1 at the donor-acceptor distances of 5, 6 and 7 Å. Figure 5.7 plots the diabatic and adiabatic energy curves. While the diabatic energy curves are essentially independent of the donor-acceptor distance and similar in both reactions, the adiabatic energy curves vary significantly due to the different coupling magnitude at different distances. At $R = 5$ Å, the coupling in the anionic complex is so strong that the adiabatic energy curve exhibits only one minimum at the midpoint, suggesting that the system is conductive. With the increasing separation of the donor and acceptor, the activation energy ($\Delta E_a$) required for moving the charge increase and the adiabatic energy curve shows two minima.

In the diabatic limit and using the parabolic approximation for the energy surfaces, Marcus showed that [307, 308]

$$\Delta G^* = \Delta E_a + V_{AB}^* \approx \lambda / 4$$ (5.26)

Table 5.6 listed the coupling energy $V_{AB}$, charge transfer barrier $\Delta E_a$, and reorganization energy $\lambda$ in the charge self-exchange reactions between benzene ionic radicals and neutral benzene at the separations of 6–10 Å, where the donor and acceptor monomers are weakly coupled (except for the anion at $R = 6$ Å where the coupling is strong, as shown in Figure 5.7). While the coupling energy is highly geometry-dependent and falls off exponentially, the charge transfer reaction barriers converge to around 0.09 eV. For both cationic and anionic complexes, the inner-sphere reorganization energies are similar and have little geometric dependency.
Significantly, we found that the Marcus hypothesis stands and eq. 5.26 holds with negligible errors.

**Figure 5.7** Potential energy curves for the charge transfer reactions between benzene and its ionic form at the distances of 5, 6 and 7 Å. Blue and red curves denote the diabatic energy potentials, whereas the black curve refers to the adiabatic energy potential.
Table 5.6

Coupling energy $V_{AB}$, charge transfer barrier $\Delta E_a$, and reorganization energy $\lambda$ (in eV) at several distances when the donor and acceptor groups are weakly coupled

<table>
<thead>
<tr>
<th>R (Å)</th>
<th>Anion</th>
<th>Cation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V_{AB}$</td>
<td>$\Delta E_a$</td>
</tr>
<tr>
<td>6.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>7.0</td>
<td>0.053</td>
<td>0.046</td>
</tr>
<tr>
<td>8.0</td>
<td>0.028</td>
<td>0.066</td>
</tr>
<tr>
<td>9.0</td>
<td>0.013</td>
<td>0.079</td>
</tr>
<tr>
<td>10.0</td>
<td>5.0x10^{-3}</td>
<td>0.087</td>
</tr>
</tbody>
</table>

5.4 Summary

Due to its intuitive concepts and bonding pictures, VB theory has been attractive to chemists. Unfortunately, due to the high computational demands, *ab initio* VB methods severely lag behind MO-based methods although significant renewed endeavors and progresses have been observed lately [228-230, 232, 236, 238, 239, 305, 306, 411]. Instead, various empirical and semi-empirical VB approaches have been proposed and extensively applied to the elucidation of the correlations between molecular structures and properties, and the studies of chemical reactions in solution and enzyme.

The proposed BLW method takes the advantages of both the MO and VB theories and is an *ab initio* VB-like method with the high efficiency of the HF and DFT methods. Since the BLW method is based on a single Slater determinant, its extension to the DFT level can effectively take electron correlation into account.
Although the BLW method is not general but restricted to specific cases, these cases are sufficiently numerous and interesting to make the method highly useful.

Since diabatic states can be defined by BLWs at the *ab initio* level individually, the important off diagonal matrix element $V_{AB}$ can be subsequently computed and its dependency on environment can also be examined. As an adiabatic state is a combination of two or more diabatic states, BLW-based two-state (or multi-state) approaches can be developed. These kind of two-state approaches can study not only the chemical reactions as done by EVB, but also the electron transfer processes, and thus establish the qualitative Marcus-Hush model at the quantitative level. The combination of the BLW method with MD simulation codes can further allow the combined QM(BLW)/MM approach to study the solvent reorganization effect which is critical in electron transfer theory. [224-226] the BLW method can effectively define the wave function for a diabatic state and the subsequent two-state model can derive quantitatively, such as the coupling energy and reorganization energy that are essential in the charge transfer theory.
Appendix A

Acronym Glossary
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AChE</td>
<td>Acetylcholinesterase</td>
</tr>
<tr>
<td>AM1</td>
<td>Austin model 1</td>
</tr>
<tr>
<td>AMBER</td>
<td>Assisted Model Building with Energy Refinement</td>
</tr>
<tr>
<td>Amt</td>
<td>Ammonium transporter</td>
</tr>
<tr>
<td>AO</td>
<td>Atomic Orbital</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke 3 term, Lee Yang, Parr (Density Functional theory method)</td>
</tr>
<tr>
<td>BSSE</td>
<td>Basis Set Superposition Error</td>
</tr>
<tr>
<td>BLW</td>
<td>Block-Localized Wavefunction</td>
</tr>
<tr>
<td>BLW-ED</td>
<td>Block-Localized Wavefunction Energy Decomposition</td>
</tr>
<tr>
<td>CAII</td>
<td>Carbonic Anhydrase II</td>
</tr>
<tr>
<td>CC</td>
<td>Coupled Cluster theory</td>
</tr>
<tr>
<td>CHARMM</td>
<td>Chemistry at HARvard Macromolecular Mechanics</td>
</tr>
<tr>
<td>CHEQ</td>
<td>Charge Equilibration</td>
</tr>
<tr>
<td>CI</td>
<td>Configuration Interaction</td>
</tr>
<tr>
<td>DFP</td>
<td>Diisopropyl phosphorofluoridate</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>DIIS</td>
<td>Direct Inversion of the Iterative Subspace</td>
</tr>
<tr>
<td>DMPC</td>
<td>1,2-dimyristoyl-sn-glycero-3-phosphocholine</td>
</tr>
<tr>
<td>ECP</td>
<td>Effective Core Potential</td>
</tr>
<tr>
<td>EDD</td>
<td>Electron Density Difference</td>
</tr>
<tr>
<td>EFP</td>
<td>Effective fragment potential</td>
</tr>
<tr>
<td>EPB</td>
<td>diethyl 4-methylbenzylphonate</td>
</tr>
<tr>
<td>ESP</td>
<td>ElectroStatic Potential</td>
</tr>
<tr>
<td>EVB</td>
<td>Empirical Valence Bond</td>
</tr>
<tr>
<td>FQ</td>
<td>Fluctuating Charge</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>GHO</td>
<td>Generalized Hybrid Orbital</td>
</tr>
<tr>
<td>GMH</td>
<td>Generalized Mulliken-Hush</td>
</tr>
<tr>
<td>GVB</td>
<td>Generalized Valence Bond</td>
</tr>
<tr>
<td>GROMOS</td>
<td>GROnigen MOlecular Simulation</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HLSP</td>
<td>Heilter-London-Slater-Pauling function</td>
</tr>
<tr>
<td>KS</td>
<td>Kohn-Sham</td>
</tr>
<tr>
<td>LA</td>
<td>Link Atom</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones function</td>
</tr>
<tr>
<td>LSCF</td>
<td>Local self-consistent field</td>
</tr>
<tr>
<td>MCSCF</td>
<td>Multi-Configurational Self Consistent Field</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>Mep</td>
<td>Methylammonium permease</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular Mechanics</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MO</td>
<td>Molecular Orbitals</td>
</tr>
<tr>
<td>MOVB</td>
<td>Molecular Orbital and Valence Bond</td>
</tr>
<tr>
<td>MP2</td>
<td>Müller-Plesset 2nd-order</td>
</tr>
<tr>
<td>ET</td>
<td>Electron Transfer</td>
</tr>
<tr>
<td>NBO</td>
<td>Natural Bond Order</td>
</tr>
<tr>
<td>NPA</td>
<td>Natural population analysis</td>
</tr>
<tr>
<td>NPT</td>
<td>isothermal-isobaric ensemble</td>
</tr>
<tr>
<td>OPLS</td>
<td>Optimized Potentials for Liquid Simulation</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Condition</td>
</tr>
<tr>
<td>PES</td>
<td>Potential Energy Surface</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PM3</td>
<td>Parameterization Method 3</td>
</tr>
<tr>
<td>PME</td>
<td>Particle Mesh Ewald</td>
</tr>
<tr>
<td>PMF</td>
<td>Potential of Mean Force</td>
</tr>
<tr>
<td>PPP</td>
<td>Poly(p-phenylene)</td>
</tr>
<tr>
<td>PTE</td>
<td>Phosphotriesterase</td>
</tr>
<tr>
<td>QM/MM</td>
<td>Quantum Mechanics / Molecular Mechanics</td>
</tr>
<tr>
<td>RC</td>
<td>Reaction Coordinate</td>
</tr>
<tr>
<td>Rh</td>
<td>Rhesus</td>
</tr>
<tr>
<td>RHF</td>
<td>Restricted Hartree-Fock</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviation</td>
</tr>
<tr>
<td>ROHF</td>
<td>Restricted Open-shell Hartree-Fock</td>
</tr>
<tr>
<td>SCF</td>
<td>Self-Consistent Field</td>
</tr>
<tr>
<td>SCRF</td>
<td>Self-Consistent Reaction Field</td>
</tr>
<tr>
<td>SLBO</td>
<td>Strictly Localized Bond Orbital</td>
</tr>
<tr>
<td>SPC</td>
<td>Simple Point Charge</td>
</tr>
<tr>
<td>SRP</td>
<td>Specific reaction (or range) parameters</td>
</tr>
<tr>
<td>STO</td>
<td>Slater Type Orbital</td>
</tr>
<tr>
<td>TIP3P</td>
<td>Transferable Intermolecular Potentials 3 Point charge water model</td>
</tr>
<tr>
<td>TIP4P</td>
<td>Transferable Intermolecular Potentials 4 Point charge water model</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane</td>
</tr>
<tr>
<td>VX</td>
<td>O-ethyl S-(2-diisopropyl aminoethyl) methylphosphonothioate</td>
</tr>
<tr>
<td>WHAM</td>
<td>Weighted histogram analysis method</td>
</tr>
<tr>
<td>XC</td>
<td>Exchange-Correlation functional</td>
</tr>
</tbody>
</table>


247. Y. Mo and J. Gao, *Block-localized wave function (BLW)*. 2000, University of Minnesota: Minneapolis, M.N.


260. M. Boeckstaens, B. Andre, and A.M. Marini, "The yeast ammonium transport protein Mep2 and its positive regulator, the Npr1 kinase, play an important role in normal and pseudohyphal growth on various nitrogen media through retrieval of excreted ammonium". Molecular Microbiology, 2007. 64(2).534-546.


297. B.D. Wladkowski, P. Ostazeski, S. Chenoweth, S.J. Broadwater, and M. Krauss, "Hydrolysis of cyclic phosphates by ribonuclease A: A computational study


350. C. Amovilli, V. Barone, R. Cammi, E. Cances, M. Cossi, B. Mennucci, C.S. Pomelli, and J. Tomasi, "Recent advances in the description of solvent effects with the polarizable continuum model". *AdV. Quantum Chem.*, 1998. 32.227.


