

ตำแหน่งยึดจับที่เหมาะสมของน้ำในโพรงของ เอช ไอ วี-1 โปรตีนเอส
โดยวิธีการคำนวณทางเคมีควอนตัม



นางสาวจิตติมา เล่าห์พงศ์ไพศาล

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเคมี ภาควิชาเคมี

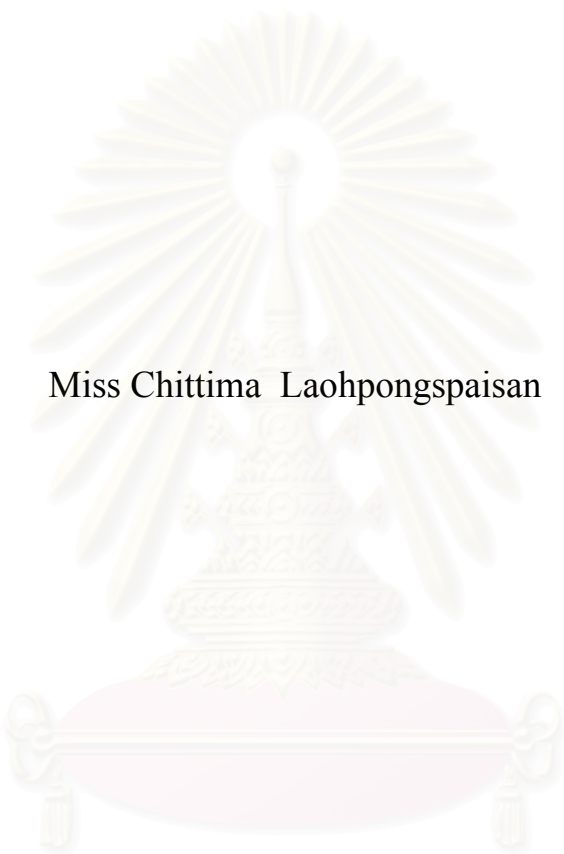
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2545

ISBN 974-17-2684-8

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

PREFERENTIAL BINDING SITES OF WATERS IN THE HIV-1 PROTEASE
POCKET BY QUANTUM CHEMICAL CALCULATIONS



Miss Chittima Laohpongspaisan

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Chemistry

Department of Chemistry

Faculty of Science

Chulalongkorn University

Academic Year 2002

ISBN 974-17-2684-8

Thesis Title Preferential Binding Sites of Waters in the HIV-1 Protease
Pocket by Quantum Chemical Calculations
By Miss Chittima Laohpongspaisan
Field of Study Chemistry
Thesis Advisor Associate Professor Supot Hannongbua, Ph. D.
Thesis Co-advisor Associate Professor Vudhichai Parasuk, Ph. D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial
Fulfillment of the Requirements for the Master's Degree

..... Dean of Faculty of Science
(Associate Professor Wanchai Phothiphichitr, Ph. D.)

THESIS COMMITTEE

..... Chairman
(Associate Professor Sirirat Kokpol, Ph. D.)

..... Thesis Advisor
(Associate Professor Supot Hannongbua, Ph. D.)

..... Thesis Co-advisor
(Associate Professor Vudhichai Parasuk, Ph. D.)

..... Member
(Assistant Professor Surapong Pinitglang, Ph. D.)

..... Member

จิตติมา เลาห์พงศ์ไพศาล: ตำแหน่งยี่ดัดจับที่เหมาะสมของน้ำในโพรงของ เอช ไอ วี-1 โปรตีเอส โดยวิธีการคำนวณทางเคมีควอนตัม (PREFERENTIAL BINDING SITES OF WATERS IN THE HIV-1 PROTEASE POCKET BY QUANTUM CHEMICAL CALCULATIONS) อาจารย์ที่ปรึกษา: รศ. ดร. สุพจน์ หารหนองบัว, อาจารย์ที่ปรึกษา
ร่วม: รศ. ดร. วุฒิชัย พาราสุข, 104 หน้า. ISBN 974-17-2684-8.

งานวิจัยนี้ได้ทำการศึกษาทะกอบของโมเลกุลน้ำในโพรงของ เอช ไอ วี-1 โปรตีเอส พีอาร์ ด้วยระเบียบวิธีการคำนวณทางเคมีควอนตัม โดยศึกษาแรงกระทำระหว่างโมเลกุลน้ำกับคลัสเตอร์ของกรดอะมิโน เปรียบเทียบกันระหว่างระเบียบวิธีทางเคมีคอมพิวเตอร์หลายวิธี และใช้ข้อมูลโครงสร้างของ เอช ไอ วี-1 พีอาร์ และ ของอะตอมออกซิเจนของน้ำได้จากโปรตีน ดาตา แบงค์ งานวิจัยนี้ ทำการศึกษาน้ำ 6 โมเลกุล ที่พบอยู่ในโครงสร้างผลึกทางเอกซ์เรย์ของ เอช ไอ วี-1 พีอาร์ ดังกล่าว โดยกำหนดให้ตำแหน่งอะตอมออกซิเจนคงที่ และปรับการวางตัวของโมเลกุลน้ำให้อยู่ในตำแหน่งที่พลังงานต่ำที่สุด ในส่วนของเอนไซม์นั้นได้ปรับ ขนาดของคลัสเตอร์จนกระทั่ง อันตรกิริยาระหว่างน้ำ กับคลัสเตอร์ของกรดอะมิโนมีค่าคงที่ ซึ่งพบว่า รัศมีของคลัสเตอร์ของกรดอะมิโนที่เหมาะสมมีค่าประมาณ 5.0-5.5 อังสตรอม รอบอะตอมออกซิเจนของน้ำ และมีจำนวนของกรดอะมิโน ประมาณ 4-8 กรดอะมิโน นอกจากนี้ยังพบว่า ระเบียบวิธีกลศาสตร์เชิงโมเลกุล ซึ่งรวมไปถึง ระเบียบวิธีผสมระหว่างกลศาสตร์ควอนตัมและกลศาสตร์เชิงโมเลกุล ไม่สามารถอธิบายระบบที่ศึกษาได้ ขณะที่วิธีทางเคนซิติ ฟังก์ชันนอล มีความเหมาะสมมากที่สุด นั้นสามารถที่จะอธิบาย ระบบดังกล่าวได้ ผลการศึกษานี้ได้วิเคราะห์และเปรียบเทียบบทบาทของน้ำทั้งที่มีส่วนในการเร่งปฏิกิริยา และที่ช่วยพยุงโครงสร้างในรูปของอันตรกิริยาระหว่างน้ำและคลัสเตอร์ของกรดอะมิโน ซึ่งค่าพลังงานที่คำนวณได้ สำหรับน้ำ W301, W607, W566, W608, W406 และ W426 ดังกล่าวมีค่า - 11.49, - 8.92, - 11.35, - 10.05, - 2.88 และ - 8.11 กิโลแคลอรีต่อโมล ตามลำดับ สำหรับน้ำ W406 นั้น ค่าอันตรกิริยา - 2.88 กิโลแคลอรีต่อโมล ชี้ให้เห็นว่าน้ำตัวนี้มีบทบาทในการพยุงโครงสร้างของ เอนไซม์ต่างจากน้ำโมเลกุลอื่น กล่าวคือ แทนที่จะทำหน้าที่ตรึงโครงสร้างโปรตีนด้วยพันธะไฮโดรเจน แต่สันนิษฐานว่าน่าจะมีส่วนในการผลัดกับเรซิดิวส์ของเอนไซม์เพื่อป้องกันไม่ให้โครงสร้างโปรตีนยุบตัวลง

ภาควิชา.....เคมี.....ลายมือชื่อ.....
สาขาวิชา.....เคมี.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2545.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

43722319 : MAJOR CHEMISTRY

KEY WORD: HIV-1 PR / 1HPX / Protease / Proteinase / Water's Role / Quantum
Chemical Calculations / DFT / *ab initio* / Interaction Energy

CHITTIMA LAOHPONGSPAISAN: PREFERENTIAL BINDING SITES OF
WATERS IN THE HIV-1 PROTEASE POCKET BY QUANTUM CHEMICAL
CALCULATIONS. THESIS ADVISOR: ASSOC. PROF. DR. SUPOT
HANNONGBUA, Ph.D., THESIS COADVISOR: ASSOC. PROF. DR.
VUDHICHAI PARASUK, Ph.D., 104 pp. ISBN 974-17-2684-8.

Role of water in the HIV-1 Protease pocket, has been investigated, based on quantum chemical calculations. The interaction energies between amino acid cluster and water molecules were evaluated and compared using various methods. X-ray structure of the HIV-1 PR and of the oxygen atoms of water were taken from Protein Data Bank. The six water molecules in the crystallographic structure of the HIV-1 PR have been taken into consideration. Position of the oxygen atom of water was fixed but its orientation around the oxygen was optimized. The size of amino acid cluster was increased until consistency of the interaction energy was reached. The results show that the optimal radius of the cluster is between 5.0 and 5.5 Å around water molecule accommodating 4-8 amino acid residues. The MM method, and hence the QM/MM fail to describe the investigated system while the DFT/B3LYP is the most appropriate method. Role of water molecules both in hydrolysis mechanism and in stabilizing the enzyme structure have been analyzed in terms of interaction energy and found that the calculated values for W301, W607, W566, W608, W406 and W426 are as follows; - 11.49, - 8.92, - 11.35, - 10.05, - 2.88 and - 8.11 kcal/mol, respectively. For W406, the interaction of - 2.88 kcal/mol suggested that its role in stabilizing enzyme structure is supposed to differ from the other structural water molecules, i. e., instead of holding the enzyme by forming the hydrogen bond, it supposes to repel the residues in order to prevent them from getting collapse.

Department.....Chemistry.....Student's signature.....
Field of Study.....Chemistry..... Advisor's signature.....
Academic year 2002.....Co-advisor's signature.....

ACKNOWLEDGEMENTS

First of all, I would like to give my deep gratitude to my parents for their exceedingly understanding, encouragement and all supports. I would also like to thank my uncle and aunt for taking care of me during the period of my study and away from my family.

Next, thanks all of my brothers and sisters in Christ at the Hope of Bangkok Church. Special thanks to Ms. Naampeung, CU leader (just not position) who one time was my buddy, her encouragement strengthening me to do all works. And above all is our heavenly Father for his blessing.

I would like to express my sincere thanks to Assoc. Prof. Dr. Supot Hannongbua, my advisor and Assoc. Prof. Dr. Vudhichai Parasuk, my co-advisor, for useful guidance, kind helps, long endurance against my innocence, and valuable suggestions both before work started and until the final stages. Their work-hard lives for public will also be typical of my life. I am grateful to Assoc. Prof. Dr. Sirirat Kokpol, Assist. Prof. Dr. Surapong Pinitglang, and Dr. Vannajan Sanghiran Lee for their advices as the thesis committee.

Finally, I would like to thank the Austrian-Thai Center for Computer assisted Chemical Education and Research (ATC) and Computational Chemistry Unit Cell, Chulalongkorn University, Thailand for computer resources and other facilities.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CONTENTS

	Pages
ABSTRACT IN THAI.....	iv
ABSTRACT IN ENGLISH.....	v
ACKNOWLEDGEMENTS.....	vi
CONTENTS.....	vii
LISTS OF FIGURES.....	xi
LISTS OF TABLES.....	xiii
ABBREVIATIONS.....	xv
CHAPTER 1 AN INTRODUCTION TO AIDS.....	1
1.1 Worldwide AIDS Epidemic.....	1
1.2 HIV life Cycle.....	3
1.3 AIDS Healing.....	4
1.4 Role of Water in HIV-1 PR	6
1.5 Overall objective.....	7
CHAPTER 2 HIV-1 PROTEASE ENZYME	8
2.1 HIV Protease: Biology, Biochemistry and Structure.....	8
2.1.1 Structure and Mechanism of HIV Protease.....	8
2.1.2 Structure of HIV Protease Inhibitor Complexes.....	9
2.1.3 Role of HIV Protease in the Viral Life Cycle.....	10
2.2 Design of HIV Protease Inhibitors.....	13
2.3 Drug Resistance.....	14
2.3.1 Resistance Mechanisms.....	15
2.3.2 Active Site Mutants.....	15
2.3.3 Non-Active Site Mutants.....	16
2.3.4 Cleavage Site Mutants and Subsite Specificity.....	16

CONTENTS (Continued)

	Pages
CHAPTER 3 THEORETICAL BACKGROUND.....	18
3.1 An Introduction to Quantum Mechanics.....	18
3.2 Solution of Schrödinger Equation of Molecular Systems.....	20
3.2.1 Schrödinger Equation.....	20
3.2.2 Born-Oppenheimer Approximation.....	22
3.2.3 <i>Ab Initio</i> method.....	23
3.2.4 Molecular Orbital Model.....	25
3.3 Basis Set.....	27
3.3.1 Minimal Basis Sets.....	28
3.3.2 Extended <i>sp</i> Basis Sets.....	28
3.3.3 Polarization Basis Sets.....	29
3.3.4 Basis Sets Incorporating Diffuse Function.....	30
3.4 Molecular mechanics method.....	31
3.5 Semi-empirical method.....	34
3.5.1 MNDO.....	35
3.5.2 AM1.....	35
3.5.2 PM3.....	36
3.6 Density Functional Method.....	37
3.7 QM/MM.....	38
3.5.1 Non-Automated Procedures.....	38
3.5.2 Partitioning of Energy.....	39
3.5.3 Energy Subtractions.....	40
CHAPTER 4 CALCULATION DETAILS.....	42
4.1 HIV-1 Protease Crystal Structure: Completely Initial Structure.....	43

CONTENTS (Continued)

	Pages
4.2 Model Representations.....	45
4.2.1 The Expansion of Model Size: Model I.....	46
4.2.2 The Expansion of Model Size: Model II.....	46
4.3 Computations: Single point calculations.....	47
4.3.1 QM.....	47
4.3.2 QM/MM	47
4.4 Evaluations: Water-Enzyme Interactions.....	47
4.4.1 Interaction with Each Water Molecule.....	47
4.4.2 Interaction with Two Water Molecules.....	47
4.4.3 Interaction with Three Water Molecules within 6 Å.....	48
CHAPTER 5 RESULTS.....	50
5.1 Optimal Calculated Model.....	50
5.2 Optimal Cluster Size and Its Interaction for the Single Water Cluster, H₂O(Res)_{n(r)}	53
5.3 Optimal Cluster Size and Its Interaction for the Double Water Cluster, (H₂O)₂(Res)_{n(r)}	63
5.4 Optimal Cluster Size and Its Interaction for the Triple Water Cluster, (H₂O)₃(Res)_{n(r)}.....	63
CHAPTER 6 DISCUSSION.....	64
6.1 Cluster Sized-Expanding of Amino Acid.....	64
6.2 Water-Enzyme Interactions.....	66
6.2.1 Single Water Cluster.....	66
6.2.2 Double Water Cluster.....	67

CONTENTS (Continued)

	Pages
6.2.3 Triple Water Cluster.....	68
CHAPTER 7 CONCLUSIONS.....	69
7.1 Model Representation.....	69
7.2 The Optimum Method for Water-Amino Acid Cluster.....	69
7.3 Water's Role from this Study.....	70
7.4 Concluding Comments.....	70
REFERENCES.....	72
APPENDICES.....	78
APPENDIX I: Semi-empirical Results.....	79
APPENDIX II: Amino Acid Sequence of HIV-1 PR.....	83
APPENDIX III: Lists of Amino Acids.....	84
APPENDIX IV: Manuscript.....	85
BIOGRAPHY.....	104

List of Figures

Figures	Pages
1.1 HIV life cycle.....	3
1.2 Sketch of the structures and the hydrogen bonds of the two loops containing active sitetriads.....	7
2.1 Chain tracings of HIV PR from crystal structures of (a) unbound and (b) inhibited forms of the enzyme. Inhibitor is drawn as a stick figure.....	8
2.2 Proposed possible active site interactions in HIV PR with a putative substrate or peptide-based inhibitor.....	10
2.3 Genome organization and translational strategy for HIV.	11
2.4 Cleavage site sequences in Gag and Gag-Pol polyproteins recognized by HIV PR	12
4.1 Schematic representation of the calculated procedures.....	42
4.2 X-ray crystal structure of HIV-1 PR (pdb code:1HPX). Six waters were shown in ball and stick model. Catalytic aspartic acids were shown in stick model. Another residues shown in wire-frame model.....	43
4.3 The position of studied six water molecules taken from X-ray crystallographic structure	44
5.1 Schematic representations of the amino acid residue around WAT607 in the H ₂ O(Res) _m cluster (Model I) where (a) – (d) for m = 1-5, respectively ...	51
5.2 Schematic representations of the amino acid residue around WAT607 in the H ₂ O(Res) _{n(r)} cluster (Model II) varying from r = 3.0 – 6.0 Å.....	53
5.3 Schematic representations of the amino acid residue around WAT301 in the H ₂ O(Res) _{n(r)} cluster (Model II) varying from r = 3.0 – 6.5 Å.....	54

List of Figures (Continued)

5.4	Schematic representations of the amino acid residue around WAT406 in the $H_2O(Res)_{n(r)}$ cluster (Model II) varying from $r = 3.0 - 6.0 \text{ \AA}$	56
5.5	Schematic representations of the amino acid residue around WAT566 in the $H_2O(Res)_{n(r)}$ cluster (Model II) varying from $r = 3.0 - 6.0 \text{ \AA}$	57
5.6	Schematic representations of the amino acid residue around WAT426 in the $H_2O(Res)_{n(r)}$ cluster (Model II) varying from $r = 3.0 - 6.0 \text{ \AA}$	59
5.7	Schematic representations of the amino acid residue around WAT607 in the $H_2O(Res)_{n(r)}$ cluster (Model II) varying from $r = 3.0 - 6.0 \text{ \AA}$	60
5.8	Schematic representation of the six water molecules lying in the amino acid environments in the HIV-1 protease pocket	61
5.9	Changes of the water-enzyme interaction energy in the $H_2O(Res)_{n(r)}$ cluster calculated using different methods for each water molecule, where r denotes spherical expansion radius according to Model II.....	62

List of Tables

Tables	Pages
5.1 The water-enzyme interaction energies for lytic water (W607) where cluster size was increased according to the Model I (MM energy (Real) = -1.45 kcal/mol).....	50
5.2 The water-enzyme interaction energies for lytic water (W607) where cluster size was increased according to the Model II (MM energy (Real) = -1.45 kcal/mol).....	52
5.3 The water-enzyme interaction energies of tetrahedrally hydrogen-bond structural water molecule (W301) according to Model II (MM energy (Real) = -0.69 kcal/mol).....	55
5.4 The water-enzyme interaction energies of W406 according to Model II (MM energy (Real) = -1.38 kcal/mol).....	55
5.5 The water-enzyme interaction energies of W566 according to Model II (MM energy (Real) = +0.47 kcal/mol).....	58
5.6 The water-enzyme interaction energies of W426 according to Model II (MM energy (Real) = -2.53 kcal/mol).....	58
5.7 The water-enzyme interaction energies of W608 according to Model II (MM energy (Real) = -1.17 kcal/mol).....	60

List of Tables (Continued)

Tables	Pages
5.8 The water-enzyme interaction energies according the <i>Model II</i> expansion of all investigated water were shown in three different methods: Interaction with each water molecule	61
5.9 The water-enzyme interaction energies for the two closely water molecules, where the cluster size was increased reaching to the Model II.....	63
5.10 The water-enzyme interaction energies for the three closely water molecules, where the cluster size was increased reaching to the Model II.....	63

ABBREVIATIONS

INTRODUCTION

AIDS	Acquired Immunodeficiency Syndrome
HIV	Human Immunodeficiency Virus
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
RT	Reverse Transcriptase
PR	Protease
IN	Integrase
RSV	Rous Sarcoma Virus
PDB	Protein Data Bank

THEORETICAL BACKGROUND

QM	Quantum Mechanics
MM	Molecular Mechanics Methods
QM/MM	Combined Quantum Mechanics and Molecular Mechanics or Quantum Mechanics/ Molecular Mechanics
HF	Hartree-Fock Method
GTO	Gaussian-type orbital
MO	Molecular Orbital Method
LCAO	Linear Combination of Atomic Orbitals Approximation
STO-3G	Slater-type Atomic Orbital in terms of 3 gaussian functions
MNDO	Modified Neglect of Diatomic Overlap
AM1	Austin Model 1
PM3	Parameterization Model 3

DFT	Density Functional Theory
B3LYP	Becke's three parameter and Lee-Yang-Parr for the correlation

SYMBOLS

\hat{H}	Hamiltonian Operator
Ψ	Wave Function
E	Total Energy
∇^2	Laplacian Operator
m_i	mass of particle i
q_i	charge of particle i
r_{ij}	distance between particles i and j
$E^{eff}(R)$	effective electronic energy which depends on relative nuclear coordinates denoted by R
ψ	Molecular orbital
ϕ	one-electron function
ΔH_f	Heat of Formation

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 1

INTRODUCTION

1.1 Worldwide AIDS Epidemic

Since the identification of Acquired Immunodeficiency Syndrome (AIDS) in developed country in the early 1980s, the AIDS epidemic has resulted in a total 11.7 million deaths, including the deaths of 4.0 million women and 2.7 million children. According to the World Health Organization, 5.8 million people were infected with the Human immunodeficiency virus (HIV) in 1997 alone, and 30.6 million people are currently living with HIV infection. HIV was identified as the causative agent for AIDS in 1983. The virus infects CD4⁺ lymphocytes and causes their destruction with a half-life of less than two days (1).

Efforts to control the AIDS epidemic have focused heavily on studies of the biology, biochemistry, and structural biology of HIV and on interactions between viral components and new drug candidates. The reverse transcriptase inhibitor AZT (zidovudine) was first approved by the U.S. Food and Drug Administration (FDA) for treating AIDS in 1987, and seven FDA-approved reverse transcriptase (RT) inhibitors (including nine nucleosides and three non-nucleoside inhibitors) are now commercially available (2). Although these drugs delay the progression of the disease, they do not prevent it, as infection readily leads to drug-resistant mutants. More recently, new class of drugs that target the HIV protease (PR) was introduced, and seven different PR inhibitors are currently on the market (2). These drugs were developed via structure-based rational drug design strategies, in which drug candidates were designed, tested, and modified on the basis of high-resolution three-dimensional structural information obtained for PR and PR-inhibitor complexes. As for the RT inhibitors, the virus is capable of developing resistance to the PR inhibitors. This is likely to be due to the low fidelity of RT, which does not have a proofreading function. Although a given a given cell is believed to be infected only

once by HIV, it has been estimated that at 10^9 new cells are infected per day in HIV-infected patients, and that point mutations occur along the entire length of the genome at the rate of 10^4 to 10^5 times per day (1).

Recently developed “drug cocktails” containing combinations of PR and RT inhibitors can reduce viral loads to undetectable levels, and these low levels can be maintained for periods of two years or more. Although there are grounds for optimism that current drug cocktails may keep the virus at bay for extended periods, it appears unlikely that the current repertoire will lead to a cure. The most serious problem is that the virus apparently can be maintained in reservoirs that are not susceptible to the current drugs. In addition, the current drug regimens are expensive and compliance is difficult, and it is therefore prudent to continue to pursue other viral component or potential drug targets (1).

HIV is a lentivirus and belongs to the family Retroviridae, which are enveloped, positive-sense, single-stranded RNA viruses (3). Retroviruses induce a variety of neoplastic diseases and are widely distributed among vertebrate species. A defining characteristic of these viruses is their use of an RNA-dependent DNA polymerase, or reverse transcriptase (RT), for replication of the viral RNA. Many of the molecular events specific to HIV-1 infection have been characterized including functions common to the retrovirus life cycle (Figure 1.1). Infection by a retrovirus results in the synthesis of one or more double-stranded DNA intermediates by the action of RT and at least one DNA copy of the viral genome is integrated into the host DNA. This proviral DNA serves to direct its own transcription, translation, and assembly of new virions. This ability of retroviruses to incorporate their genomes into that of their host cells endows them with the capability to be stably maintained during the life of the host and even to be transmitted through the germ line. While most germ line “endogenous” retroviruses found in animals and humans are believed to be nonpathogenic, tumor-causing retroviruses are fairly common in animals, and their discovery dates back to 1901.

1.2 HIV life cycle

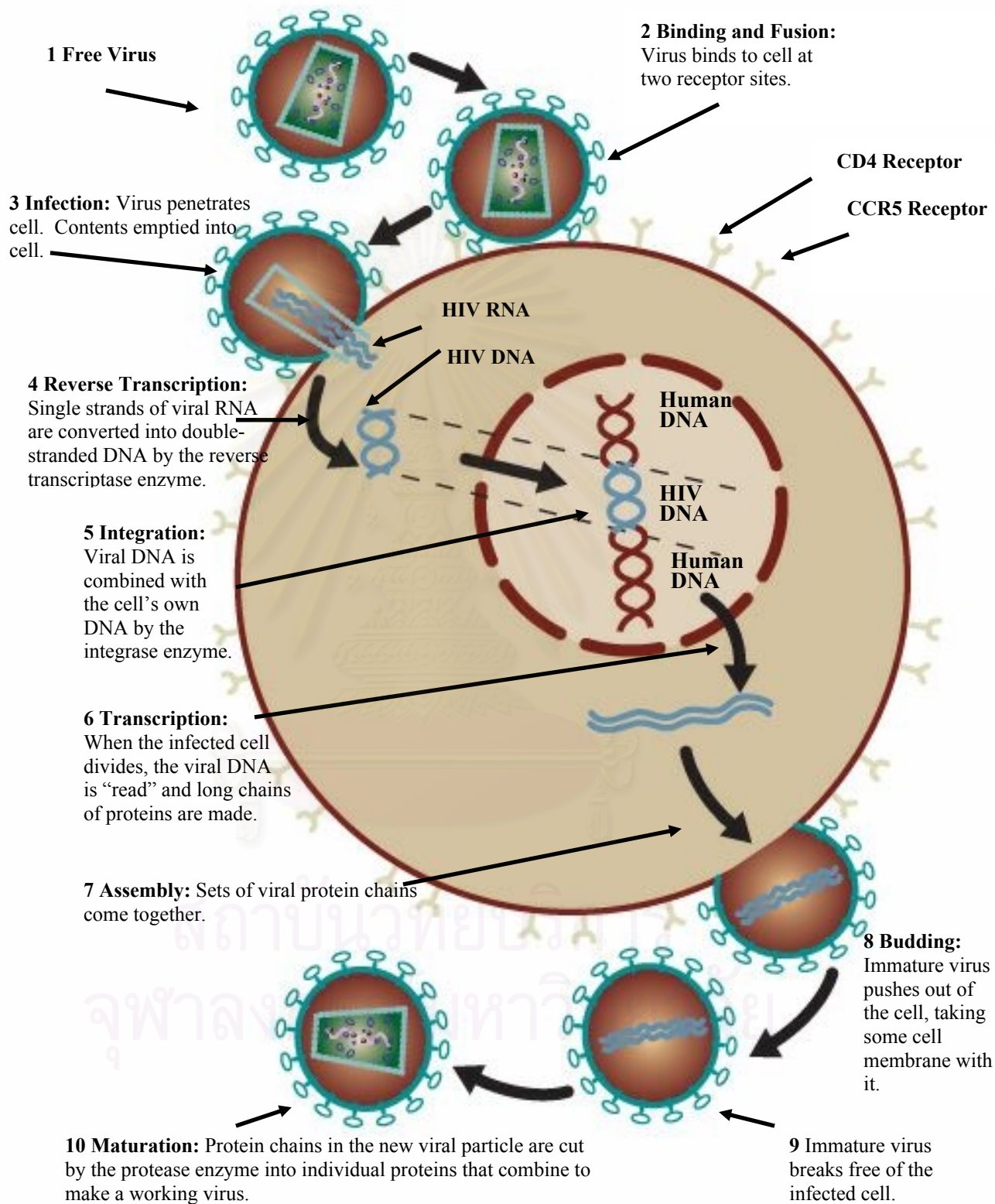


Figure 1.1 HIV Life Cycle (5).

HIV infects a T cell via recognition of the CD4 receptor on the cell surface. Fusion of the viral envelope and cell membranes leads to cytoplasmic invasion by the nucleoprotein core of the virus. Proviral DNA is synthesized using the virion-associated RT and tRNA as a primer. Integration of proviral DNA is mediated by a viral integrase (IN), also present in the infecting virion, and host factors. Transcription of proviral DNA into spliced and unspliced RNAs provides mRNAs for translation of the gag, pol, and env gene products as well as viral RNA for packaging. Assembly of the Gag and Gag-Pol precursor proteins and packaging of the viral RNA occurs at the cell membrane. Extracellular budding of virions results in the acquisition of an envelope which contains the viral env proteins required for subsequent rounds of receptor recognition and fusion. Processing of the Gag and Gag-Pol polyproteins occurs during budding and release and is mediated by a viral PR. The viral enzymes are products of the pol gene. Pol encodes the enzymes PR, RT, RNase H, and IN. As with the Gag gene products, these are produced by cleavage of a larger precursor polyprotein. In this case, the protein is a Gag-Pol polyprotein, Pr160^{gag-pol}, which is made as the result of a site-specific frame-shifting event during translation of the viral mRNA (Gag and Pol are encoded in different reading frames on the HIV-1 genome, and the two open reading frame overlap). Note that is a self-cleaving protein. PR is a homo-dimeric enzyme which cleaves the HIV-1 Gag-Pol polyprotein. Its crystal structure and site-specificity have been defined, allowing the design of a variety of enzyme inhibitors. RT is an RNA-dependent DNA polymerase which synthesizes DNA from RNA template. It also has RNase H activity, which allows it to degrade RNA from RNA:DNA hybrids (this is a necessary step during reverse transcription). The crystal structure of RT has been defined and many specific inhibitors of this enzyme has been derived. One of the important properties of RT is that it lacks proofreading activity. As a result, RT has a high error rate and generates a large number of mutations. IN catalyzes the insertion of linear double-stranded HIV-1 DNA into the host cell chromosome. This is an obligatory step in the HIV-1 life cycle (see Figure 1.1).

1.3 AIDS Healing

The discovery of HIV-1 in 1984 led to a parallel explosion of research on the molecular virology of this infectious agent and to an intensive, global search for a cure for this fatal disease that continues to the present. Efforts to inhibit HIV continue to represent one of the most active areas of antiviral research today, and nearly every aspect of the viral life cycle has become a target for antiviral drug discovery. Modern antiviral strategies are turning more and more to the use of structure- and mechanism-based approaches for the design of safer, more specific, and effective drugs.

The successful introduction of HIV PR inhibitors for AIDS treatment is a testimonial to the importance of structure- and mechanism-based approaches in the discovery and design of more effective, less toxic antiviral therapies. To date there are seven drug-approved PR inhibitors: Saquinavir, Ritonavir, Indinavir, Nelfinavir, Amprenavir, Lopinavir, and Kaletra (2). There are a number of drugs for RT both nucleosides and non-nucleosides, as follows; zidovudine (AZT), didanosine (ddI), zalcitabine (ddC), stavudine (d4T) and lamivudine (3TC). An IN enzyme is a recent target which has also been studied for drug inhibiting HIV infection. These compounds are highly potent and play a central role in the development of the highly active anti-retroviral therapies that comprise the current standard of care and that, for the first time in the brief history of HIV infection, provide dramatic and durable suppression of HIV replication. It is probably safe to assume that much of the attention currently being paid by the pharmaceutical industry to other viral proteases as drug design targets is a coattail effect based on HIV success story (3). It can be argued that much of the current focus on viral proteases as targets for antiviral therapy really stems from the collision of two disparate disciplines—protease biochemistry and virology. Thus, it is worth reflecting on the scientific developments that led up to the identification of HIV PR as an attractive target for drug design.

1.4 Role of water in the HIV-1 PR

It is generally accepted that water molecules play an important role in the binding affinity or specificity of HIV-1 PR inhibitors (10-17). A prominent example is a lytic water molecule which holds by the two catalytic Asp residues through the hydrogen bonding and plays role in inducing the protein hydrolysis (1, 4). Another important water molecule is in flap region, called WAT301, which observes to participate in hydrogen-bonding interactions between the flaps and the inhibitors. This water plays role in controlling movement of the flaps. it is found also in the crystal structure of almost every HIV-1 protease-inhibitor complexes bridged the backbone amide protons of Ile50 and Ile150 to inhibitor (presumably substrate) carbonyls through hydrogen bonding. Among the two water molecules mentioned above, the presence of other structurally important water molecules inside and outside the active pocket of the protease were also reported (14, 16), WAT301, WAT406, WAT426, WAT566, WAT607, WAT608 (see Figures 1.2). These examples clearly demonstrate role of water in the catalytic process. They are, hence, taken into account in a structure-based drug design strategy (Figures 2.1-2.2).

Several experimental techniques (30-38) such as NMR spectroscopy and X-ray crystallography have been used to map out the hydration water molecules both in free form and its complexes with inhibitors. On the other side, the water positions were investigated theoretically within HIV-1 protease complexes using computer simulations (9-11). Most theoretical data regarding water molecules in the free protease and its complexes so far have relies on the use of intermolecular interaction based on molecular mechanics (MM) parameterization (11). As a matter of fact, this kind of potentials is developed to preserve transferability for a wide range of systems and usually leads to the loss of some specific details. For this reason, some doubts arise when the MM method was used to represent the interaction between protease or its complexes and water molecules in which hydrogen bonding could be very important and better represented by quantum chemical calculations. Due to the above reasons, quantum mechanics calculation and combined quantum mechanic and molecular mechanic (QM/MM) have been applied to investigate precise position and orientation of water molecules in the binding pocket of the PR.

CHAPTER 2

HIV-1 PROTEASE ENZYME

2.1 HIV Protease: Biology, Biochemistry, and Structure

2.1.1 Structure and Mechanism of HIV Protease

The HIV PR dimer consists of two identical, non-covalently associated subunits of 99 amino acid residues associated in a twofold (C₂) symmetric fashion (Figure 2.1a).

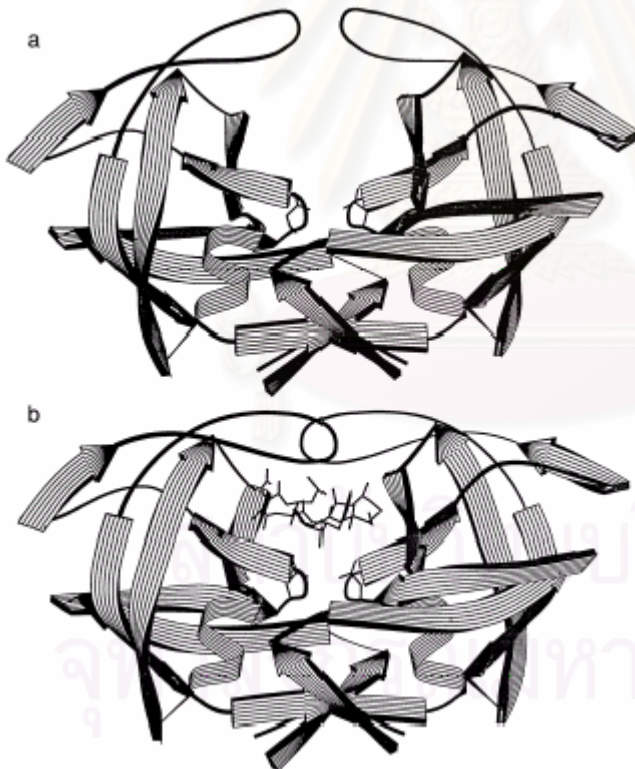


Figure 2.1 Chain tracings of HIV PR from crystal structures of (a) unbound and (b) inhibited forms of the enzyme. Inhibitor is drawn as a stick figure (1, 4).

acid information APPENDIX II).

The dimer is stabilized by a four-stranded antiparallel β -sheet formed by the interlocking N- and C-termini of each subunit. The active site of the enzyme is actually formed at the dimer interface and contains two conserved catalytic aspartic acid residues, one from each monomer (33, 34). The substrate binding cleft is composed of equivalent residues from each subunit and it bound on one side by the active site aspartic acids, Asp25 and Asp25', and on the other by a pair of twofold related, antiparallel β -hairpin structures, or “flaps” (see amino

Comparison of the structure of HIV PR with that of a complex with a peptide-based inhibitor shows that the flap undergoes significant structural changes upon binding and that it makes several direct interactions with inhibitor (Figure 2.1b). Molecular dynamics studies indicate that the flaps are highly flexible and must undergo large localized conformational changes during the binding and release of inhibitors and substrates. Crystal packing forces apparently maintain the flap in a conformation that is unsuitable for substrate binding in the structure of the enzyme. The crystal structures of Rous Sarcoma Virus Protease (RSV) and HIV PR revealed that, despite the apparent lack of sequence homology, aspartic proteases and retroviral proteases display considerable structural homology at the backbone level. Fully one-third of the main chain atoms of RSV PR can be superposed onto the backbone of porcine pepsin to within 1.5 Å root-mean-square deviation. As expected, most of the structural correspondence is in the active site region. However, the overall chain topologies of the two families of enzymes are more similar than a simple superposition analysis reveals and are indicative of a distant but definite relationship to a common, ancestral aspartic protease gene.

2.1.2 Structure of HIV Protease Inhibitor Complexes

To date, several hundred crystal structures have been solved for various HIV protease/inhibitor complexes—a testimony to the importance placed on structural information in the process of inhibitor design (4). Structural comparison of the inhibitor complexes reveals certain common features (see Figure 2.2). The inhibitor and enzyme make a pattern of complementary hydrogen bonds between their backbone atoms. In some instances, these hydrogen bonds are mediated by bridging water molecules. A unique feature found in the structure of HIV PR/inhibitor complexes is the presence of a water molecule that forms bridging hydrogen bonds between the NH atoms of Ile50 and Ile150 in the two flaps and the P2 and P1' backbone carbonyl groups of the inhibitor. Since a similar pattern of hydrogen bonds is believed to be made for both substrates and peptidomimetic inhibitors, specificity is

believed to reside in the pattern of largely nonpolar subsite interactions between inhibitor and enzyme side-chain atoms (6).

Overall, knowledge of the structure and function of HIV PR and its relationship to other aspartic proteases has led to the successful development of a wide variety of potent and chemically diverse inhibitors (39-41).

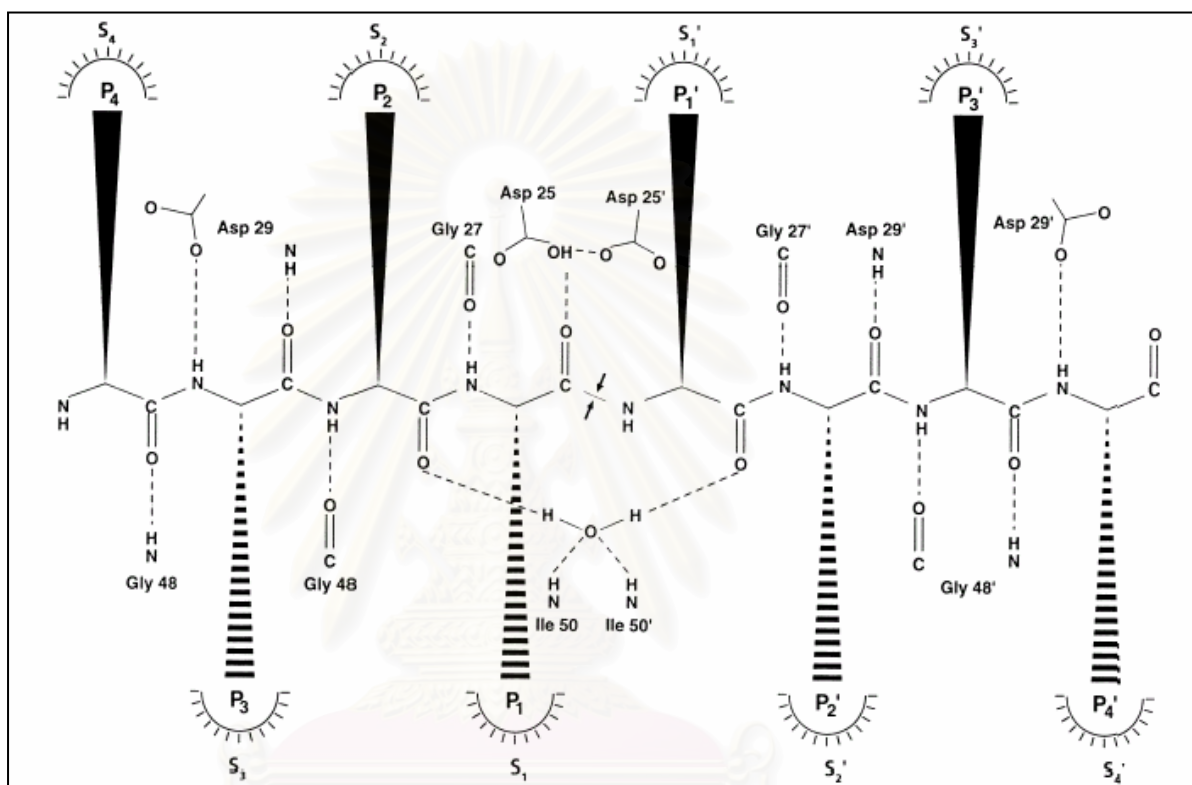


Figure 2.2 Proposed possible active site interactions in HIV PR with a putative substrate or peptide-based inhibitor (4).

2.1.3 Role of HIV PR in the viral life cycle

The HIV genome, like all other retroviral genomes, is a single-stranded, positive-sense RNA molecule that is organized into three major coding elements: the gag, pol, and env genes. The gag and pol gene product are translated from a single unspliced polycistronic mRNA that encodes both genes (Figure 2.3). A stop codon in the unspliced RNA leads to the translation of a 55-kDa Gag polyprotein, Pr55gag, that contains sequences of the structural proteins of the virion-matrix (MA), capsid (CA), and nucleocapsid (NC)-along with the peptides p2, p1, and p6 that are involved in the

assembly and morphogenesis of mature capsids. The *pol* gene encodes the viral enzymes necessary for replication—protease (PR), reverse transcriptase (RT), and integrase (IN). These proteins are also translated as part of a larger polyprotein precursor, Pr160gag-pol, which results from ribosomal frameshift and readthrough during translation of the *gag* gene. The Pr55 precursor protein is believed to play a central role in directing virion assembly and RNA packaging based on studies with other retroviruses which show that enveloped nucleoprotein core particles can form from Gag precursor proteins in the absence of *pol* and *env* gene products.

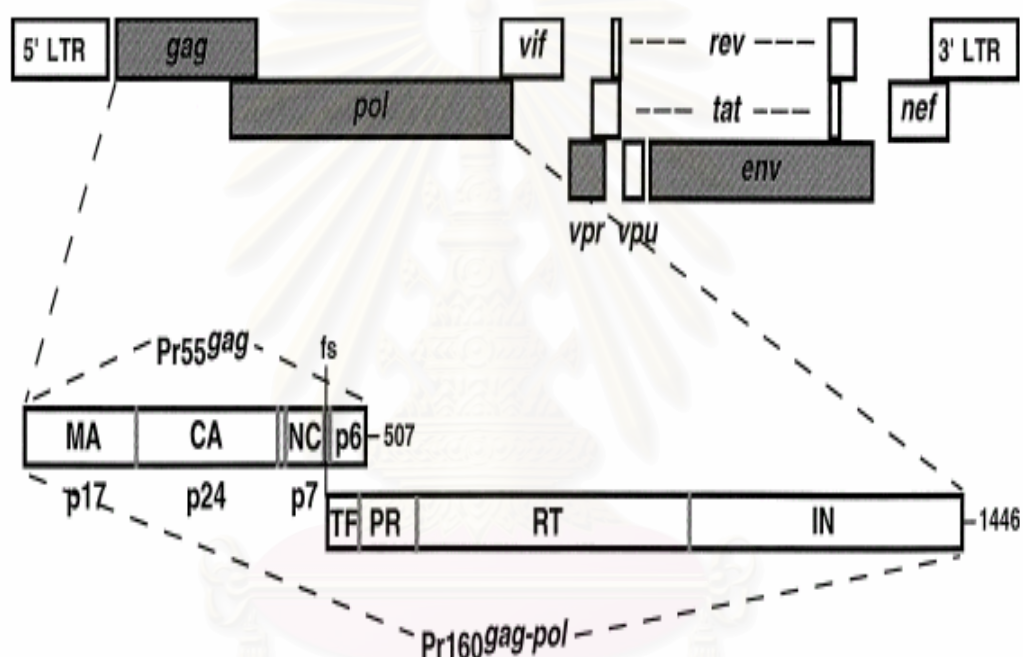


Figure 2.3 Genome organization and translational strategy for HIV. Structural (*gag*, *pol*, and *env*) genes are shaded; regulatory (*tat* and *rev*) and accessory (*vif*, *nef*, *vpr*, and *vpu*) genes are clear. Common to all retroviruses, the *gag* and *pol* gene products are translated on free ribosomes in the cytoplasm from newly synthesized unspliced viral RNA. Translation usually occurs through to a stop codon at the 3' end of the *gag* gene resulting in the structural polyprotein Pr55gag. About 5% of the timeribosome frameshifting during translation of *gag* results in the synthesis of a Gag-Pol fusion protein, Pr160gag-pol. The frameshift site (fs) is located upstream of the Gag p6 protein such that a transframe polypeptide, TF, is incorporated into Gag-Pol in place of p6. The functions of the p6 and TF proteins are unclear. The total number of amino acids contained by each polyprotein is indicated at the end of each molecule (4).

Site	Processing Sites for HIV-1 Protease (P4 P3 P2 P1 P1' P2' P3' P4')
MA/CA	-Ser-Gln-Asn- Tyr/Pro -Ile-Val-Gln-
CA/p2	-Ala-Arg-Val- Leu/Ala -Glu-Ala-Met
P2/NC	-Ala-Thr-Ile- Met/Met -Gln-Arg-Gly-
NC/p1	-Arg-Gln-Ala- Asn/Phe -Leu-Gly-Lys-
P1/p6	-Pro-Gly-Asn- Phe/Leu -Gln-Ser-Arg-
TF/PR	-Ser-Phe-Asn- Phe/Pro -Gln-Ile-Thr-
PR/RT	-Thr-Leu-Asn- Phe/Pro -Ile-Ser-Pro-
RT/IN	-Arg-Lys-Val- Leu/Phe -Leu-Asp-Gly-
RT (internal)	-Alu-Glu-Thr- Phe/Tyr -Val-Asp-Gly-

Figure 2.4 Cleavage site sequences in Gag and Gag-Pol polyproteins recognized by HIV PR. Cleavage occurs between residues in the P1/P1' positions and they are indicated in bold and separated by a slash. The RT (internal) represents a PR-mediated cleavage at the junction of the p51/Rnase H domains which yields the active p66/p51 heterodimer found in isolated virus particles. The nomenclature of Schechter and Berger (1967) is used to designate residue positions in the substrate sequence (4).

Proteolytic processing of Pr55gag and Pr160gag-pol during virus assembly and maturation is performed by the viral PR, which is itself encoded by the pol gene. The env gene product, gp160, is processed into gp120 and gp41 by a cellular protease. The processing products of HIV PR include the gag-encoded structural proteins and peptides-MA, CA, NC, p1, p2, and p6-and the pol enzymes-RT, IN, and PR. All of these products are found in mature infectious viral particles and result from cleavages at unique amino acid sequences that span the N- and C-termini of the mature protein (Figure 2.2). The sequences recognized by HIV PR are diverse, but certain general features emerge. Hydrophobic amino acids are preferred at the P1/P1' residues that

flank the scissile peptide bond, aliphatic and Glu/Gln residues are often found at P2', aromatic residues are almost never found at P3', and small residues are preferred at P2. Several sequences contain an aromatic residue at P1 followed by a Pro at P1'. Although all retroviral proteases appear to be structurally and functionally related, their cleavage site preferences vary widely. Efforts to predict cleavage sites for HIV Pr have met with limited success and the understanding of the basis of PR specificity is incomplete. However, identification of cleavage site sequences quickly led to the successful generation of a variety of synthetic substrates that facilitated the design of rapid and quantitative assays of PR activity

2.2 Design of HIV protease inhibitors

The design of clinically effective HIV PR inhibitors has been a major success story for structure-based design (3, 7, 8). Several HIV PR inhibitors are currently in widespread use for the treatment of patients with AIDS e.g. Saquinavir, Ritonavir, Indinavir, Nelfinavir, Amprenavir, Lopinavir, and Kaletra. These compounds represent a new class of therapeutic agents that complement already-licensed antivirals-AZT, ddI, ddC and d4T- all of which inhibit HIV RT. Although the initial lead compounds have been generated in various ways, the availability of protein crystal structures of these leads with HIV PR facilitated structure-based approaches to the optimization of interactions to increase potency. Once potent inhibitors of the enzyme were generated considerations such as improving antiviral potency, improving bioavailability, and reducing cost could be addressed within the known structural limitations provided by these crystal structures (39-41).

The close structural and functional relationship between the retroviral and cellular aspartic proteases, together with knowledge of the HIV PR cleavage site sequence on Gag and Gag-Pol protein polyproteins, immediately opened the avenue of peptidomimetic substrate-based approaches that had been developed for designing inhibitors of human rennin, an aspartic protease that was a popular target for drug design of antihypertensive agents in the 1980s.

The crystal structure of HIV PR immediately provided as a structural basis for the development of a new generation of inhibitors that did not need to rely on substrate or peptide mimicry. Virtually all of the different approaches to improving potency have incorporated some aspect of structure into their design efforts, and numerous protein crystal structures of the enzyme with inhibitors have been generated. Structure-based design has been used to identify templates for elaboration as well as to optimize the interactions found in the original peptide-based substrates and to replace peptide moieties by nonpeptide group (49, 50).

Two groups have been pursuing a different cyclic nonpeptide template for the development of HIV PR inhibitors. Reviews of these efforts have appeared (51).

Irreversible enzyme inhibitors are often ignored as therapeutic targets because of toxicity and bioavailability concerns. A number of weak irreversible inhibitors of HIV protease have been described and crystal structures of the resulting alkylated enzyme have been generated (52).

Unlike the case for the majority of RT inhibitors, pharmacokinetics, most HIV PR inhibitors have serious pharmacokinetic limitations. Poor oral absorption, serum protein binding, liver enzyme metabolism, and other factors can all but eliminate the antiviral benefits of many potent PR inhibitors. As a result, the failure or drop-out rate of patients on PR inhibitor therapy tends to be relatively high, the compliance of treatment with PR inhibitors is likely to be poor, and the development of resistance to these highly effective drugs is a growing problem (53).

2.3 Drug resistance

Although many promising new anti-HIV drugs have been developed, their effectiveness has been hampered by the emergence of drug-resistant variants (4). Mutant viruses emerge in the presence of antiviral agents whenever the balance of mutant virus replication is favorable, i. e. the mutant provides a selective advantage to the virus in the presence of the drug.

An understanding of the biological and structural mechanisms of resistance to HIV PR inhibitors is necessary in order to predict optimal salvage therapies; indeed,

the full implications of drug resistance for disease outcome have not yet been realized. A great deal of progress has been made at unraveling the structural, biochemical, and virologic mechanisms of resistance to PR (2) and RT inhibitors in light of the three-dimensional atomic structures of these proteins. An important goal of these studies is to gain insights that may lead to new strategies to combat resistance to anti infectious disease agents.

2.3.1 Resistance mechanisms

A variety of resistance mechanisms to HIV PR inhibitors have been proposed based on our understanding of the structural biochemistry of the PR and on the nature of inhibitor binding to then enzyme (4). They can be classified as active sites vs nonactive site mutations according to whether they occur inside or outside the inhibitor binding subsites. Mutations of specificity-determining residues that would directly interfere with inhibitor binding and lead to loss of potency constitute an obvious mechanism for resistance to HIV PR inhibitors. Active site mutations are necessary but often are not sufficient for high-level resistance in the clinical setting. An explanation for this observation comes from biochemical studies that reveal a negative impact of many resistance-conferring active site mutations on the enzyme activity, suggesting that such mutations result in suboptimal virus. A second mechanism of resistance involves non-active site mutations that indirectly alter the active site architecture via long-range structural perturbations.

2.3.2 Active Site Mutants

The first-described resistance mutation for HIV PR was a single substitution of a valine residue by alanine at position 82 (V82A mutation) that was selected using a symmetric diol inhibitor (54). Since then, resistance mutations have been observed in each of the specificity pockets, S3, S2, S1 and, by symmetry, S1', S2', S3'. However, only a subset of all residues that constitute a particular subsite mutates in response to

a particular drug. The structural effects of mutations on drug binding have been modeled using the crystal structures of the appropriate wild-type enzyme/inhibitor complexes and used to rationalize the effects of specificity mutations on drug binding. Most of the subsite mutations, like I84V and V82A, -I or -F, affect hydrophobic and van der Waals interactions and can be considered to be packing mutants, somewhat analogous to hydrophobic mutations in a protein. While numerous crystal structures of wild-type HIV PR/inhibitor complexes have been published, crystal structures of mutant HIV PR/inhibitor complexes have appeared with less frequency in the literature.

2.3.3 Non-active site mutants

While the precise structural mechanism of drug resistance can often be pinpointed for active site mutations that directly affect inhibitor binding, the evaluation of non-active site mutants is more challenging. Some mutations might act in concert with active site mutations by compensating for a functional deficit caused by a latter.

For example, the R8Q (Arginine replaced by Glutamine at the number of position 8) mutation is found almost exclusively in combination with one or more mutations outside the active site region, such as M46I (55). Mutations of Met46 is in the flap of HIV PR and molecular dynamics simulations on flap movement have shown that the M46I mutant exhibits a markedly different dynamical behavior than the wild-type enzyme and presumably exhibits altered enzyme kinetics. A role for Met46 in polypeptide substrate recognition is also possible.

However, the fact that these mutations are only observed in the presence of drug means that they must provide the mutant virus with a competitive growth advantage over wild-type HIV. However, this mutant has not been observed either in vitro or in vivo in the presence or absence of inhibitors.

2.3.4 Cleavage site mutants and subsite specificity

Active site mutations can strongly affect catalytic efficiency of HIV PR (56), but the magnitude of the effect depends on the substrate sequence (57). The addition of one or more non-active site mutations may compensate for a catalytically defective active site mutation. These conclusions are based on enzymology studies with recombinant HIV PR mutants.

Recent longitudinal studies of individual drug treated patients demonstrate an ordered accumulation of mutations in which one or two active site mutations usually occur early and are followed by numerous non-active site mutations. Thus, the clinical evolution of drug resistance to protein inhibitors seems to qualitatively mirror expectations based on the enzymology studies. Since active site mutations may be expected to alter the rate of one or more cleavages that must occur during viral maturation, one may imagine that compensating mutations in the cleavage sites on the Gag or Gag-Pol polyproteins might result in better substrates for particular mutant enzymes. This is an important finding since it confirms the possibility for this drug resistance mechanism to be operative in the clinical setting. Several groups have followed the lead of these investigators and have confirmed the presence of cleavage site mutations in clinical isolates from multidrug-resistant HIV in preliminary reports.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 3

THEORETICAL BACKGROUND

In this chapter, a brief concept of quantum mechanics (QM) starting from elementary to advance technique was overviewed. Molecular mechanics (MM) as well as combined QM/MM methods were also summarized.

3.1 Introduction to Quantum Mechanics

Toward the end of the nineteenth century, many physicists felt that all the principles of physics had been discovered and little remained but to clear up a few minor problems and to improve experimental methods in order to investigate the next decimal place. This attitude was somewhat justified by the great advances in physics that had been made up to that time (48).

The first half of the nineteenth century was an intensely active period for discovery of electric and magnetic effect, best exemplified by the brilliant career of Michael Faraday and the complete unification of many diverse experimental observation by Maxwell. Not only did Maxwell's prediction of the electromagnetic nature of light unify the fields of optics and electricity and magnetism, but its subsequent experimental demonstration by Hertz in 1887 appeared to be a final blow to the corpuscular theory of light. The body of these accomplishments is now considered to be the development of what scientists now call "classical physics." Little was it realized in this justifiably heady era of success that physics was about to enter a period of profound discovery and growth whose effects have filtered not only into the field of chemistry, biology, and engineering but into technology and politics as well.

The twentieth century saw the birth of the theory of relativity and quantum mechanics. The first, due to Einstein alone, completely altered our ideas of space and time and is an extension of classical physics to the region of high velocities and

astronomical distances. Quantum mechanics, on the other hand, was developed over several decades by many people and is an extension of classical physics to subatomic, atomic, and molecular sizes and distances. Relativity theory and quantum theory constitute what is now called modern physics. Although relativity theory has played a profound role in our everyday life through nuclear energy it has not yet played an important role in the field of chemistry. Quantum mechanics, however, in dealing in the atomic and molecular region has played a very important role in chemistry, so much so that an introductory course in quantum mechanics and its applications to chemistry, often called quantum chemistry, is an integral part of any chemistry curriculum.

QM which is the correct mathematical description of the behavior of electrons i.e. has never been found to be wrong, give a mathematical description of the behavior of electrons. The postulates and Theorems of quantum mechanics form the rigorous foundation for the prediction of observable chemical properties from first principles. The expressed somewhat loosely, the fundamental postulates of quantum mechanics assert that microscopic systems are described by wave functions that completely characterize all of the physical properties of the system. In particular, there are quantum mechanical operators corresponding to each physical observable that, when applied to the wave function, allow one to predict the probability of finding the system to exhibit a particular value or range of values for that of observable.

In theory, QM can predict any property of an individual atom or molecule exactly. However, the quantum mechanical equation have never been solved exactly for any chemical system other than the hydrogen atom or one electron systems. Thus, the entire field of computational chemistry is built around approximate solutions. Some of these solutions are very crude and others are expected to be more accurate than any experiment that has been yet designed. There are several implications of this situation. A myriad collection of methods has been developed for approximating the solution for multiple electron systems. This approximations can be very useful, but this requires an amount of sophistication on the part of researcher to know when each approximation is valid and how accurate the results are likely to be.

3.2 Solution of the Schrödinger Equation of Molecular Systems

3.2.1 The Schrödinger Wave Equation

The Schrödinger equation is a fundamental equation of quantum mechanics. The solutions to the Schrödinger equation are called wave functions (46).

As mentioned, the energy as well as properties of a stationary state of a molecule can be obtained by solving of the Schrödinger equation,

$$\hat{H} \Psi = E \Psi . \quad (3.1)$$

Here \hat{H} is the *Hamiltonian* operator, a differential operator representing the total energy. E is the numerical value of the energy of the state, that is, the energy relative to a state in which the constituent particles (nuclei and electrons) are infinitely separated and at rest. Ψ is the wave function. It depends on the Cartesian coordinates of all particles (which may take any value from $-\infty$ to $+\infty$) and also on the spin coordinates (which may take only a finite number of values corresponding to spin angular momentum components in a particular direction). The square of wave function, Ψ^2 (or $|\Psi|^2$ if Ψ is complex), is interpreted as a measure of the probability distribution of the particles within the molecule.

The Hamiltonian operator \hat{H} , in general,

$$\hat{H} = \hat{T} + \hat{V} \quad (3.2)$$

where for a molecule

$$\hat{T} = \hat{T}_n + \hat{T}_e = -\sum_{A=1}^N \frac{1}{2M_A} \nabla_A^2 - \sum \frac{1}{2} \nabla_i^2 \quad (3.3)$$

where ∇_i^2 and ∇_A^2 is the Laplacian operator acting on the particles which are both electrons and nuclei, respectively.

$$\hat{V} = \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} = -\sum_{i=1}^n \sum_{A=1}^N \frac{Z}{r_{iA}} + \sum_{i=1}^n \sum_{j<i}^n \frac{1}{r_{ij}} + \sum_{A=1}^N \sum_{B<A}^N \frac{Z_A Z_B}{R_{AB}} \quad (3.4)$$

From (3.3) to (3.4), the molecular Hamiltonian is,

$$\hat{H} = -\sum \frac{1}{2} \nabla_i^2 - \sum_{A=1}^N \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^n \sum_{A=1}^N \frac{Z}{r_{iA}} + \sum_{i=1}^n \sum_{j<i}^n \frac{1}{r_{ij}} + \sum_{A=1}^N \sum_{B<A}^N \frac{Z_A Z_B}{R_{AB}} \quad (3.5)$$

where A and B refer to nuclei and i and j refer to electrons. The first and second terms in (3.5) is the operator for the kinetic energy of the electrons and nuclei, respectively. The third term is the electron-nuclear attraction where R_{iA} being the distance between electron i and nucleus A . The fourth term is the electron-electron repulsion where r_{ij} being the distance between electron i and j . The last term is the nuclear-nuclear repulsion where R_{AB} being the distance between nuclei A and B with atomic numbers Z_A and Z_B , respectively.

This formulation is the time-independent, nonrelativistic Schrödinger equation. Additional terms can appear in the Hamiltonian when relativity or interaction or fields are taken into account. This also ceases to be appropriate as the velocities of the particles, particularly electrons approach the velocity of light. Although this is a significant effect for the inner-shell electrons of heavy atoms. Certain small magnetic effects, for example, spin-orbit coupling, spin-spin interactions, and so forth, are also omitted in this Hamiltonian; these are usually of minor significance in discussions of chemical energies.

One other restriction has to be imposed on the wave functions. The only solution of Schrödinger equation that are physically acceptable are those with appropriate symmetry under interchange of identical particles. For boson particles, the wave function is unchanged, that is symmetric, under such interchange. For fermion particles, the wave function must be multiplied by -1 , that is, antisymmetric.

Electrons are fermion, so that Ψ must be antisymmetric with respect to interchange of the coordinates of any pair of electrons. This is termed the antisymmetry principle.

3.2.2) Born-Oppenheimer Approximation

The first major step in simplifying the general molecular problem in quantum mechanics is the separation of the nuclear and electronic motions. This is possible because the nuclear masses are much greater than those of the electrons, therefore, nuclei move much more slowly. As a consequence, the electrons in a molecule adjust their distribution to changing nuclear position rapidly. This makes it reasonable approximation to suppose that the electron distribution depends only on the instantaneous positions of the nuclei and not on their velocities. In other words, the quantum mechanical problem of the electron motion in the field of fixed nuclei may first be solved, leading to an effective electronic energy $E^{eff}(R)$ which depends on the relative nuclear coordinates, denoted by R . This effective energy is then used as a potential energy for a subsequent study of the nuclear motion. $E^{eff}(R)$ will depend on all of the relative nuclear coordinates. For a diatomic molecule, only the internuclear distance, R , is required and $E^{eff}(R)$ is the potential curve for the molecule. For a polyatomic system, more relative coordinates are needed, and $E^{eff}(R)$ is termed the potential surface for the molecule. This separation of the general problem into two parts is frequently called the adiabatic or Born-Oppenheimer approximation. It was examined quantitatively by Born and Oppenheimer, who showed that it was valid, provided that the ratio of electron to nuclear mass was sufficiently (46).

Quantitatively, the Born-Oppenheimer approximation may be formulated by writing down the Schrödinger equation for electrons in the field of fixed nuclei,

$$\hat{H}^{elec} \Psi^{elec}(r, R) = E^{eff}(R) \Psi^{elec}(r, R). \quad (3.6)$$

Here, Ψ^{elec} is the electronic wave function which depends on the electronic coordinates, r , as well as on the nuclear coordinates, R . The electronic Hamiltonian, \hat{H}^{elec} , corresponds to motion of electrons only in the field of fixed nuclei and is

$$\hat{H}^{elec} = \hat{T}^{elec} + \hat{V}, \quad (3.7)$$

where \hat{T}^{elec} is the electronic kinetic energy, and \hat{V} is the coulomb potential energy,

$$\hat{H}^{elec} = -\sum_i^{elec} \frac{\nabla_i^2}{2} - \sum_A \sum_i^{elec} \frac{Z_i}{r_{iA}} + \sum_{i<j}^{elec} \sum \frac{1}{r_{ij}} \quad (3.8)$$

The first part of (3.6) corresponds to the kinetic energy of the electrons only. The second term is the attraction of electron to nuclei. The third term is the repulsion between electrons.

The main task of theoretical studies of electronic structure is to solve, at least approximately, the electronic Schrödinger equation (3.6), and hence find the effective nuclear potential function $E^{eff}(R)$. From this point, we omit the superscripts in (3.5); it is assumed that the Hamiltonian, \hat{H} , wave function, Ψ , and energy, E , refer to electronic motion only, each quantity being implicitly a function of the relative nuclear coordinated, R .

3.2.3 *Ab initio* methods

The term *ab initio* is a latin for “form the beginning”. This name is given to computations that are derived directly from theoretical principles with no inclusion of experimental data. This is an approximate mathematical calculation. The approximation which is made, are usually mathematical approximations such as using a simpler functional form for a function or finding an approximate solution to a differential equation.

Hartree Fock (HF) approximation, is the most common type of *ab initio* calculation. The primary approximation is the central field approximation, which means that the Coulombic electron-electron repulsion is taken into account by integrating the repulsion term. This gives an average effect of the repulsion, but not explicit repulsion interaction. This is a variational calculation, meaning that the

approximate energies calculated are all equal to or greater than exact energy. One of the advantages of this method is that one electron equation is solved to yield a single-electron wave function, called an orbital, and the energy, called an orbital energy. The second approximation in HF calculations is due to the fact that the wave function must be described by some mathematical function, which is known exactly for only a few one-electron systems. The functions used most often are linear combinations of Gaussian-type orbital $\exp(-\alpha r^2)$, abbreviated GTO. The wave function is formed from linear combination of atomic orbitals or, stated more correctly, from linear combinations of basis functions. Gaussian functions are multiplied by an angular function in order to give the orbital the symmetry of s, p, d, and so on. A constant angular term yields s symmetry. Angular terms of xy , xz , yz , x^2-y^2 , $4z^2-2x^2-2y^2$ yield d symmetry. This pattern can be continued for the other orbitals. These orbitals are then combined into a determinant. This is done to satisfy two requirements of quantum mechanics. One is that the electrons must be indistinguishable. The second requirement is that the wave function for fermions must be antisymmetric with respect to interchanging of 2 particles. Thus, if electron 1 and electron 2 are switched, the sign of the total wave function must change and only the sign can change. This is satisfied by a determinant because switching two electrons is equivalent to interchanging two columns of the determinant, which changes its sign. The functions put into the determinant do not need to be individual GTO functions, called Gaussian primitives. They can be a weighted sum of basis functions on the same atom are often used to make the calculation run faster. Sum of basis functions on different atoms are used to give the orbital a particular symmetry. For example, a water molecule with C_{2v} symmetry will have orbitals that transform as A_1 , A_2 , B_1 , B_2 , which are irreducible representations of the point group. The resulting orbitals that use functions from multiple atoms are called molecular orbitals. This is done to make the calculation run much faster. Any overlap integral over orbitals of different symmetry does not need to be computed because it is zero by symmetry.

One of limitation of HF calculations is that they do not include electron correlation. This means that HF take into account the average affect of electron repulsion, but not explicit electron-electron interaction. Within HF theory the probability of finding an electron at some location around an atom is determined by the distance from the nucleus but not to the to each other electrons. This is not

physically true, but it is the consequence of the central field approximation, which defines the HF method.

A number of types of calculations begin with a HF calculation and then correct for correlation. Some of these methods are Møller-Plesset perturbation theory (MPn, where n is the order of correction), configuration interaction (CI), and coupled cluster theory (CC). These methods are referred to as correlated calculations.

3.2.4 Molecular Orbital Model

The theoretical models discussed in this thesis are all based on molecular orbital (MO) theory. This approximate treatment of electron of electron distribution and motion assigns individual electrons to one-electron function termed *spin orbitals*. These comprise a product of spatial function, termed *molecular orbitals*, $\psi_1(x, y, z)$, $\psi_2(x, y, z)$, $\psi_3(x, y, z)$, ..., and either α or β *spin components*. The spin orbitals are allowed complete freedom to spread throughout the molecule, their exact form being determined variationally to minimize the total energy. In the simplest version of the theory, a single assignment of electrons to orbitals (sometimes called an electron configuration) is made. These orbitals are then brought together to form a suitable many-electron many electron wave function Ψ which is the simplest MO approximation to the solution of the Schrödinger equation.

In practical calculations, the molecular orbitals ψ_1, ψ_2, \dots are further restricted to be linear combinations of a set of N known one-electron function $\phi_1(x, y, z), \phi_2(x, y, z), \dots, \phi_N(x, y, z)$:

$$\psi_i = \sum_{\mu} c_{\mu i} \phi_{\mu}. \quad (3.9)$$

The function $\phi_1, \phi_2, \dots, \phi_N$ (which are defined in the specification of the model) are known as one-electron basis functions, or simply as basis function, or simply as basis functions. They constitute the basis set. If the basis function are the atomic orbitals for the atoms making up the molecule, Eq. (3.9) is often described as *the linear*

combination of atomic orbitals (LCAO) approximation, and is frequently used in qualitative descriptions of electronic structure.

Given a basis set, the unknown coefficients $c_{\mu i}$ are determined so that the total electronic energy calculated from the many-electron wave function is minimized and according to the variational theorem, is as close as possible to the energy corresponding to exact solution of the Schrödinger equation. This energy and the corresponding wave function represent the best that can be obtained within the Hartree-Fock approximation, that is, the best given the constraint imposed by: (a) the use of limited basis set in the orbital expansion, and (b) the use of single assignment of electrons to orbitals.

Hartree-Fock models are the simplest to use for chemical applications and have been employed in many of the studies carried out to date. To specify the model in full, it is only necessary to define a unique basis set for any nuclear configuration. This is conveniently done by having a standard set of basis functions for each nucleus, centered at nuclear position, which depend only on the corresponding atomic number. Thus, there would be a set of functions for each hydrogen atom and other sets for each carbon and so forth. In the simplest Hartree-Fock models, the number of basis functions on each atom will be as small as possible, that is, only large enough to accommodate all the electrons and still maintain spherical symmetry. As consequence, the molecular orbitals (3.1) will have only limited flexibility. If larger basis sets are used, the number of adjustable coefficients in the variational procedure increases, and an improved description of the molecular orbitals is obtained. Very large basis sets will result in nearly complete flexibility. The limit of such an approach, termed the Hartree-Fock limit, represents the best that can be done with a single electron configuration.

The main deficiency of Hartree-Fock theory is its incomplete description of the *correlation* between motions of the electrons. Even with a large and completely flexible basis set, the full solution of the Schrödinger equation cannot be expressed in terms of a single electron wave function, that is, a unique assignment of single electron wave function, that is, a unique assignment of electrons to orbitals. To correct the deficiency, it is necessary to use wave functions that go beyond the Hartree-Fock level, that is, that represent more than a single electron configuration. If Ψ_0 is the full

Hartree-Fock many electron wave function, the extended approximate form for the more accurate wave function Ψ is

$$\Psi = a_0 \Psi_0 + a_1 \Psi_1 + \dots \quad (3.10)$$

Here Ψ_1, Ψ_2, \dots are wave functions for other configurations, and the linear coefficients a_0, a_1, \dots are to be determined. Inclusion of wave function for all possible alternative electron configurations (within the framework of given basis set) is termed *full configuration interaction*. It represents the best that can be done using that basis set. Practical methods, which may be sequenced in order of increasing sophistication and accuracy, seek either to limit the number of configurations or to approximate the effect which their inclusion has on the total wave function.

3.3 Basis Sets

Basis set is a set of functions used to describe the shape of the orbitals in an atom. Molecular orbitals and entire wave functions are created by taking linear combinations of basis functions and angular functions. The molecular orbitals ψ_i in a Hartree-Fock treatment are expressed as linear combinations of N nuclear-centered basis function ϕ_μ ($\mu = 1, 2, \dots, N$),

$$\psi_i = \sum c_{\mu i} \phi_\mu \quad (3.11)$$

A limiting of Hartree-Fock treatment would involve an infinite set of basis functions ϕ_μ . This is clearly impractical since the computational expense of Hartree-Fock molecular orbital calculations is formally proportional to the fourth power of the total number of basis functions. Therefore, the ultimate choice of basis set size depends on a compromise between accuracy and efficiency.

3.3.1 Minimal Basis sets

The simplest level of *ab initio* molecular orbital theory involves the use of a minimal basis set of nuclear-centered functions. In the strictest sense, such a representation comprises exactly that number of functions required to accommodate all of the electrons of the atom, while maintaining overall spherical symmetry. For example, within framework of a minimal basis set, hydrogen and helium are represented by a single s-type function, lithium and beryllium by a pair of such functions, and the remaining first-row elements (boron to neon) by two s functions and a complete set of three p-type functions (46).

3.3.2 Extended *sp* Basis sets

Minimal basis sets have several inherent inadequacies. Because the number of atomic basis functions is not apportioned according to electron count—for example, the lithium atom, which has only three electrons, is provided with the same number of functions (five) as fluorine with its nine electrons—it follows that minimal basis set descriptions of compound containing elements such as oxygen and fluorine are likely to be poorer than those for molecules comprising elements with fewer electrons. A second problem arises because a minimal basis set using fixed Gaussian exponents is unable to expand and contract in response to differing molecular environments. This is because a minimal basis set contains only a single valence function of each particular symmetry type, for example, $2s, 2p_x$. Thus, in the absence of radial exponent optimization for each atom in a molecule, there is no mechanism for the individual sets of functions to adjust their sizes. Finally, minimal representations lack the ability to describe adequately the non-spherical anisotropic aspects of molecular charge distributions.

In principle, the first two deficiencies may be alleviated simply by allowing for more than a single valence function of each symmetry type in the basis set description. In this way, the number of basis functions for all elements, not just those on the left-hand side of the Periodic Table, would be substantially in excess of the number actually required. Furthermore, the allocation of two or more valence basis functions of each given symmetry type would provide the needed flexibility for

overall radial size to be determined simply by the adjustment of the relative weights of the individual components in the variation procedure.

The third deficiency, the inability of a minimal basis set to describe properly anisotropic molecular environments, may be alleviated in one of two ways. The conceptually simpler way would be to allow each of the x, y and z p components describing the valence region of a main group element to have different radial distribution, that is to employ an anisotropic rather than an isotropic minimal basis set. Similarly, the five d-function components describing the valence manifold of a first- or second-row transition metal could be independently specified.

A basis set formed by doubling all functions of a minimal representation is usually termed a double-zeta basis. An even simpler extension of a minimal basis set is to double only the number of basis functions representing the valence region. While the inner-shell electrons are important with regard to total energy, their effect on molecular bonding is of a little consequence (46).

3.3.3 Polarized Basis sets

One feature common to the basis sets discussed thus far is that they comprise functions constrained to be centered at the nuclear positions. The description of highly polar molecules, and of systems incorporating small strained rings, requires that some allowance be made for the possibility of non-uniform displacement of charge away from the atomic centers. Without such allowance, comparisons of properties between, say, small ring compounds and their acyclic isomers are likely to lead to unreliable results. This situation may be dealt with in one of two ways. Conceptually the simpler approach is to allow for the inclusion in the basis set of functions not associated with any particular center. The first way, the valence basis set is supplemented by gaussian s- and p-type primitives placed along the bonds. Such an approach of allowing for functions to be placed away from the nuclear positions is not, however, without disadvantages. It 's generally not size consistent. That is to say, the number of non-nuclear centered functions, those placed along bonds, does not necessarily increase in direct proportion to molecular size. Also, while the position of non-nuclear centered basis functions are easily and uniquely specified in molecule

characterized by 2-center, 2-electron bonds, their placement is not obvious in molecules involving multi-center bonding.

Another way to allow the small displacements of the of electronic charge away from the nuclear positions is simply to include function of higher angular quantum number (d-type functions on heavy atoms and p-type functions on hydrogen) in the basis set.

A basis set that incorporates function of higher angular momentum number than are needed by the atom in its electronic ground state is called a polarization basis set; it provides for displacement of electronic charge way form the nuclear centers, that is charge polarization (46).

3.3.4 Basis sets incorporating diffuse functions

The basis sets that have been discussed thus far, although developed with no extended bias, are more suitable for molecules in which the electrons are tightly held to the nuclear centers than they are for species with significant electron density far removed form those centers. Calculations involving anions pose special problems. Since the electron affinities of the corresponding neutral molecules are typically quite low, the extra electron in the anion is only weakly bound. Even the largest basis sets considered do not incorporate functions with significant amplitude far distant form their center, and therefore do not provide a completely adequate description of molecules in which a large portion of the valence-electron density is allocated to diffuse lone-pair or to anti-bonding orbitals. For most stable anions, application of these basis sets yields positive energies for the highest occupied molecular orbitals, indicating erroneously that the outermost valence electron are unbound.

One way to overcome problems associated with anion calculations is to include in the basis representation one or more sets of highly diffuse functions. These are then able to describe properly the long-range behavior of molecule orbitals with energies close to the ionization limit. It has been shown that the addition of diffuse functions has dramatic effects on calculated electron affinities, proton affinities, and inversion barriers (46).

3.4 Molecular Mechanics (MM) Methods

The most severe limitation of *ab initio* methods is the limited size of the molecule that can be modeled on even the largest computers. Molecular mechanics method is still possible to model its behavior avoiding quantum mechanics totally (47).

The molecular mechanics energy expression consists of a simple algebraic equation for the energy of a compound. It does not use a wave function or total electron density. The constants in this equation are obtained either from spectroscopic data or *ab initio* calculations. A set of equation with their associated constant is called a force field. The fundamental assumption of molecular method is the transferability of parameters. In other words, the energy penalty associated with a particular molecular motion, say, the stretching of a carbon-carbon single bond, will be the same from one molecule to the next. This gives a very simple calculation that can be applied to very large molecular systems. The performance of this technique is dependent on four factors as follow, the functional form of the energy expression, the data used to parameterize the constants, the technique used to optimize constants from that data, and the ability of the user to apply the technique in a way consistent with its strengths and weaknesses.

In order for the transferability of parameters to be a good description of the molecule, force fields use atom types. This means that a sp^3 carbon will be described by different parameters than a sp^2 carbon, and so on. Usually, atoms in aromatic rings are treated differently from sp^2 atoms. Some force fields even parameterize atoms for specific functional groups. For example, the carbonyl oxygen in the carboxylic acid may be described by different parameters than the carbonyl oxygen in a ketone.

The energy expression consists of the sum of simple classical equations. These equation describe various aspects of the molecule, such as bond stretching, bond bending, torsion, electrostatic interactions, van der Waals forces, and hydrogen bonding. Force fields differ in the number of terms in the energy expression, the complexity of those terms, and the way in which the constants were obtained. Since electrons are not explicitly included, electronic processes cannot be modeled.

Terms in the energy expression that describe a single aspect of the molecular shape, such as bond stretching, angle bending, ring inversion, or torsional motion, are called valence terms. All force fields have at least one valence term and the most have three or more.

For terms in the energy expression that describe how one motion of the molecule affects another are called cross terms. A cross terms commonly used is a stretch-bond term, which describes how equilibrium bond lengths tend to shift as bond angles are changed. Some force fields have no cross terms and may compensate for this by having sophisticated electrostatic functions. The MM4 force field is at the opposite extreme with nine different types of cross terms.

Force fields may or may not include an electrostatic term. The electrostatic term most often used is the Coulombs law term for the energy of attraction or repulsion between charged centers. These charges are usually obtained from non-orbital-based algorithm designed for use with MMs. These charges are meant to be the partial charges on the nuclei. The modeling of molecules with a net charge is described best by using atom types parameterized for describing charged centers. A dielectric constant is sometimes included to model solvation effects.

Bond stretching is most often describing by a harmonic oscillator equation. It is sometimes described by a Morse potential. In rare cases, bond stretching will be described by Lennard-Jones or quartic potential. Cubic equations have been used for describing bond stretching, but suffer from becoming completely repulsive once the bond has been stretched past a certain point.

Bond bending is the most often described by a harmonic equation. Bond rotation is generally described by a cosine expression. Intermolecular forces, such as van der Waals interactions and hydrogen bonding, are often described by Lennard-Jones equation. Some force fields also used a combined stretch-bond term. The choice of equation functional forms is particularly important for computing energies of molecules distorted from the equilibrium geometry, as evidence by the difference between a harmonic potential and a Morse potential.

All the constants in these equations must be obtained from experimental data or an *ab initio* calculation. The database of compounds used to parameterized against a specific class of molecules, such as proteins or nucleotides. Such a force field would be only expected to have any relevance in describing other proteins or nucleotides.

Other force fields are parameterized to give a reasonable description of a wide variety of organic compounds. A few force fields have even been parameterized for all the elements.

Some force fields simplify the complexity of the calculations by omitting most of hydrogen atoms. The parameters describing each backbone atom are then modified to describe the behavior of the atoms with the attached hydrogens. Thus, the calculation uses a CH₂ group rather than an sp³ carbon bonded to two hydrogens. These are called united atom force fields or intrinsic hydrogen methods. This calculation is most often employed to describe very large biomolecules. It's not generally applied if the computer hardware available is capable of using the more accurate explicit hydrogen force fields. Some force fields have atom types for use with both implicit and explicit hydrogens.

The way in which the force field parameters are obtained from this original data is also important. Bond stretching and bending are relatively stiff motions. Thus, they can often be described very well by using equilibrium values obtained from x-ray diffraction results and force constants from vibrational spectroscopy. On the other hand, torsional behavior is sensitive to both the torsional behavior of the isolated bond and the non-bonded interaction between distant sections of the molecule and surrounding molecules. The choice of fitting procedure becomes important because it determines how much of the energy is from each contributing process. A force field can also be parameterized to best predict vibrational motion or intermolecular forces.

The energies computed by MMs are usually conformational energies. This means that the energy computed is meant to be an energy that will reliably predict the difference in energy from one conformation to the next. The effect of strained bond lengths or angles is also included in this energy. This is not the same as the total energies obtained from *ab initio* programs or the heat of formation from a semiempirical program. The actual value of the conformational energy does not necessarily have any physical meaning and is not comparable between different force fields. MMs methods can be modified to compute heats of formation by including a database or computation scheme to yield bond energies that might be added to the conformational energy.

MMs methods are not generally applicable to structures very far from equilibrium, such as transition state calculations that use algebraic expressions to

describe the reaction path and transition structures are usually semiempirical algorithms. These calculations use an energy expression fitted to an *ab initio* potential energy surface for that exact reaction, rather than using the same parameters for every molecule.

3.5 Semiempirical Methods

Semiempirical calculations are set up with the same general structure as a HF calculation in that they have a Hamiltonian and 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. Parameters to estimate the omitted values are obtained by fitting the results to experimental data or *ab initio* calculations. Often, these parameters replace some of the integrals that are excluded (47).

The advantage of semiempirical calculations is that they are much faster than *ab initio* calculation. The disadvantage of semiempirical calculation is that the results can be erratic and fewer properties can be predicted reliably. 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. For example, the carbon atoms in cyclopropane and cubane have considerably different bond angles from those in most other compounds; thus, these molecules may not be predicted well unless they were included in the parameterization. However, semiempirical methods are not as sensitive to the parameterization set as are molecular mechanics calculations.

Semiempirical 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 properties, such as electronic spectra or NMR chemical shifts.

Semiempirical calculations can be used to compute properties other than those in the parameterization set.

Many semiempirical methods compute energies as heats of formation. The researcher should not add zero-point corrections to these energies because the thermodynamic corrections are implicit in the parameterization.

CIS calculations from the semiempirical wave function can be used for computing electronic excited states. This is good technique for modeling compounds that are not described properly by a single-determinant wave function. Semiempirical CI calculations do not generally improve the accuracy of results since they include correlation twice: once by CI and once by parameterization.

Semiempirical 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 semiempirical methods have been devised specifically for the description of inorganic chemistry as well. For our work, AM1 and PM3 used to test the model, has next reviewed.

3.5.1 The modified neglect of diatomic overlap (MNDO)

The MNDO method has been found to give the reasonable result qualitative results for many organic systems. It has been for many organic systems. It has been incorporated into several popular semiempirical programs as well as the MNDO program. Today, it is still used, but the more accurate AM1 and PM3 methods have surpassed it in popularity.

3.5.2 The Austin Model 1 (AM1)

AM1 method is still popular for modeling organic compounds. This method generally predicts the heat of formation (ΔH_f) more accurately than MNDO, although a few exceptions involving Br atoms have been documented. Depending on the nature of the system and information desired, either AM1 or PM3 will often give the most accurate results obtainable for organic molecules with semiempirical methods. There are some known strengths and limitations in the results obtained from these methods.

Activation energies are improved over MNDO. AM1 tends to predict results for aluminum better than PM3. It tends to poorly predict nitrogen pyramidalization. AM1 tends to give O-Si-O bonds that are not bent enough. There are some known limitations to AM1 energies, such as predicting rotational barriers to be one-third the actual barrier and predicting five-membered ring to be too stable. Nitro groups are too positive in energy. The peroxide bond is too short by about 0.17Å. Hydrogen bonds are predicted to have the correct strength, but often the wrong orientation. On average, AM1 predicted energies and geometries better than MNDO, but not as well as PM3. Computed bond enthalpies are consistently low .

3.5.3 Parameterization Model 3 (PM3)

PM3 uses nearly the same equation as AM1 method along with an improved set of parameters. This method is also currently extremely popular for organic systems. It is more accurate than AM1 for hydrogen bond angles, but AM1 is more accurate for hydrogen bond energies. The PM3 and AM1 methods are also more popular than other semiempirical methods due to the availability of algorithms for including solvation effect in these calculations.

These are also some known strengths and limitations of PM3. Overall heats of formation are more accurate than with MNDO or AM1. Hypervalent molecules are also predicted more accurately. PM3 tends to predict that the barrier to rotation around the C-N bond in peptide is too low. Bond between Si and states for germanium compounds. It tends to predict sp³ nitrogen as always being pyramidal. Some spurious minima are predicted. Proton affinities are not accurate. Some polycyclic rings are not flat. The predicted charge on nitrogen is incorrect. Nonbond distances are too short. Hydrogen bonds are too short by about 0.1Å, but the orientation is usually correct. On average, PM3 predicts energies and bond lengths more accurately than AM or MNDO.

3.6 Density functional theory

Density functional theory (DFT) has established as a very popular method in recent years. This is justified based on the pragmatic observation that is less computationally intensive than other methods with similar accuracy. This theory has been developed more recently than other *ab initio* methods. Because of this, there are classes of problems not yet explored with this theory, making it all the more crucial to test the accuracy of the method before applying it to unknown systems (47).

In DFT formulation, the energy of a molecule can be determined from the electron density instead of a wave function. This theory originated with theorem by Hohenberg and Kohn that state this was possible. The original theorem only applied to finding the ground-state electronic energy of a molecule. A practical application of this theory was developed by Kohn and Sham who formulated a method similar in structure to the HF method.

In this formulation, the electron density is expressed as a linear combination of a basis function similar in mathematical form to the HF orbitals. A determination is then formed from these functions, called Kohn-Sham orbitals. It is the electron density from this determination of orbitals that is used to compute the energy. This procedure is necessary because Fermion systems can only have electron densities that arise from an anti-symmetric wave function. There has been some debate over the interpretation of Kohn-Sham orbitals. It is certain that they are not mathematically equivalent to either HF orbitals or natural orbitals from correlated calculations. However, Kohn-Sham orbitals do describe the behavior of electrons in a molecule, just as the other orbitals mentioned do. DFT orbital eigenvalues do not match the energies obtained from photoelectron spectroscopy experiments as well as HF orbital energies do. The questions still being debated are how to assign similarities and how to physically interpret the differences.

A density functional is then used to obtain the energy for the electron density. A functional is a function of a function, in this case, the electron density. The exact density functionals is not known. Therefore, there is a whole list of different functionals that may have advantages or disadvantages. Some of these functionals were developed from fundamental quantum mechanics and some were developed by parameterized function to best reproduce experimental results. Thus, there are in

essence *ab initio* and semiempirical versions of DFT. DFT tends to be classified either as an *ab initio* method or in a class by itself.

The advantage of using electron density is that the integrals for coulomb repulsion need to be done only over the electron density, which is a three dimensional function, thus scaling N^3 . Furthermore, at least some electron correlation can be included in the calculation. This results in faster calculations than HF calculation (which scale as N^4) and computations that are a bit more accurate as well. The better DFT functionals give results with an accuracy similar to that of an MP2 calculation.

Density functionals can be broken down into several classes. The simplest is called the $X\alpha$ methods. This type of calculation includes electron exchange but not correlation. The $X\alpha$ method is similar in accuracy to HF and sometimes better.

The simplest approximation to the complete problem is one based only on the electron density, called a local density approximation (LDA). For high-spin systems, this is called the local spin density approximation (LSDA).

A more complex set of functionals utilizes the electron density and its gradient. These are called gradient-corrected methods. These are also hybrid methods that combine functional forms other methods with pieces of a Hartree-Fock calculation, usually the exchange calculations.

The accuracy of results from DFT calculations can be poor to fairly good, depending on the choice of basis set and density functional. The choice of density functional is made more difficult because creating new functionals is still an active area of research. For our work, the B3LYP hybrid functional (also called Becke3LYP) was chose for applying our system in the part of QM.

3.7 Combined quantum mechanics and molecular mechanics (QM/MM) Methods

Various computational methods have strengths and weaknesses. Quantum mechanics (QM) can compute many properties and model reactions. For molecular mechanics (MM) is able to model very large compounds quickly. It is possible to combine these two methods into one calculation, with models a very large compound using MM and

one crucial section of the molecule with QM. This calculation is designed to give results that have very good speed when only one region needs to be modeled quantum mechanically. It can also be used to model a molecule surrounded by solvent molecules. This type of calculation is called a QM/MM calculation (47).

3.7.1 Nonautomated procedures

The earliest combined calculations were done simply by modeling different parts of the system with different techniques. For example, some crucial part of the system could be modeled by using an *ab initio* geometry-optimized calculation. The complete system could then be modeled using the rest of the molecule.

This procedure yielded a geometry for the whole system, although there is no energy expression that reflects nonbonded interactions between the regions. One use is to compute the conformational strain in ligands around a metal atom, which is important in the determining the possibility of binding. In order to do this, the metal atom is removed from the calculation, leaving just the ligands in the geometry from the complete system. Two energy calculations on these ligands are then performed by with and without geometry optimization. The difference between these two energies is the conformational strain that must be introduced into the ligands in order to form the compound.

Another technique is to use an *ab initio* method to parameterize force field terms specific to a single system. For example, an *ab initio* method can be used to compute the reaction coordinate for a model system. An analytic function can then be fitted to this reaction coordinate. A MM calculation can then be performed, with this analytic function describing the appropriate bonds, and so on.

3.7.2 Partitioning of energy

Quantitative energy values are one of the most useful result from computational techniques. In order to develop a reasonable energy expression when two calculations are combined, it is necessary to know not only the energy of two regions, but also the

energy of interaction between those regions. There have been a number of energy computation schemes proposed. Most of these schemes can be expressed generally as

$$E = E_{QM} + E_{MM} + E_{QM/MM} + E_{pol} + E_{boundary} \quad (3.12)$$

The first two terms are the energies of the individual computations. The $E_{QM/MM}$ term is the energy of interaction between these regions, if we assume that both regions remain fixed. It may include van der Waals terms, electrostatic interaction, or any term in the force field being used. E_{pol} is the effect of the either region changing as a result of the presence of the other region, such as electron density polarization or solvent reorganization. $E_{boundary}$ is a way of representing the effect of the rest of the surroundings, such as the bulk solvent.

3.7.3 Energy subtraction

An alternative formulation of QM/MM is the energy subtraction method. In this method, calculation are done on various regions of the molecule with various levels of theory. Then, the energy are added and subtracted to give suitable corrections. This results in computing an energy for correct number of atoms and bonds analogous to an isodesmic strategy.

Three such methods have been proposed by Morokuma and coworkers. The integrated MO + MM (IMOMM) method combines an orbital-based technique with an MM technique. The integrated MO + MO method (IMOMO) integrates two different orbital-based techniques. The our own n-layered integrated MO and MM method (ONIOM) allows for three or more different technique.

This technique can be used to model a complete system as small model system and the complete system. The complete system would be computed using only the lower level of theory. The model system would be computed with both levels of theory, would then be

$$E = E_{low,complete} + E_{high,model} - E_{low,model} \quad (3.13)$$

Likewise, a three-layer system could be broken down into small, medium, large regions, to be compute with low, medium, and high level of theory (L, M, and H respectively). The energy expression would then be

$$E = E_{H,small} + E_{M,medium} - E_{M,small} + E_{L,large} - E_{L,medium} \quad (3.14)$$

This method has the advantage of not requiring a parameterized expression to describe the interaction of various regions. Any systematic errors in the way that the lower level of theory describe the inner regions will be canceled out. The geometry of one region will affect the geometry of the other because interaction between regions is not a systematic effect. If we assume transferability of parameters, this method avoids any over counting of the nonbonded interaction.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 4

CALCULATION DETAILS

As mentioned in chapter 1, interest of this work is centered on the role of water molecules, in terms of their interactions and orientations, in the HIV-1 protease. Brief process of this work were illustrated in Figure 4.1.

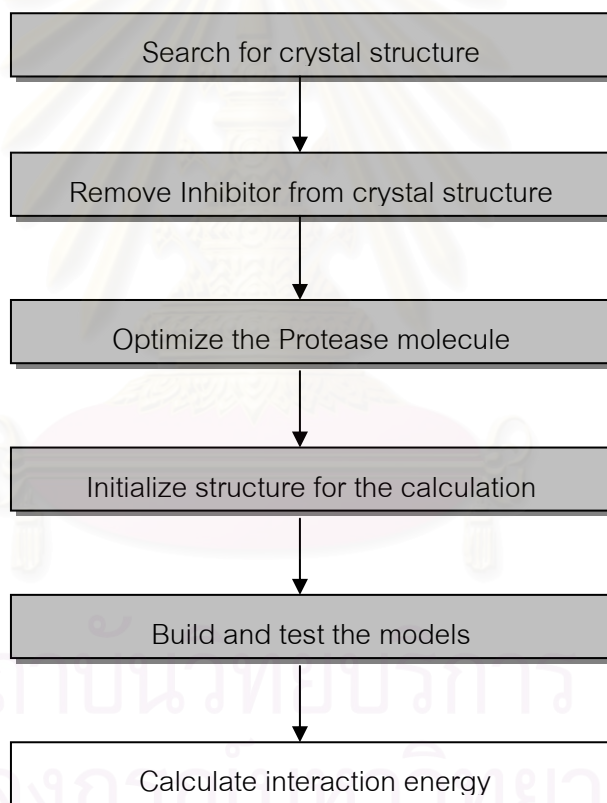


Figure 4.1 Schematic representation of the calculated procedures.

Three dimensional structure of the enzyme-inhibitor complex was taken from Protein Data Bank (PDB), PDB code: 1HPX. Because the complete X-ray structure of free protease is not available. Therefore, the complex structure was used for preparing the overall structure of the free enzyme. This was done as the following. First,

coordinate of the inhibitor was deleted. Then, the hydrogen atom was added and the enzyme structure was optimized using AMBER 5.0 program. The superimposition between free and complex forms the protein shows no significant difference. The obtained structure was used as the protein structure for calculations, where, the atomic coordinates were kept constant throughout. For water, coordinates of oxygen atom of water molecule taken from the PDB of the HIV-1 protease-inhibitor complex were fixed at their original position.

4.1 HIV-1 Protease Crystal Structure: Completely Initial Structure

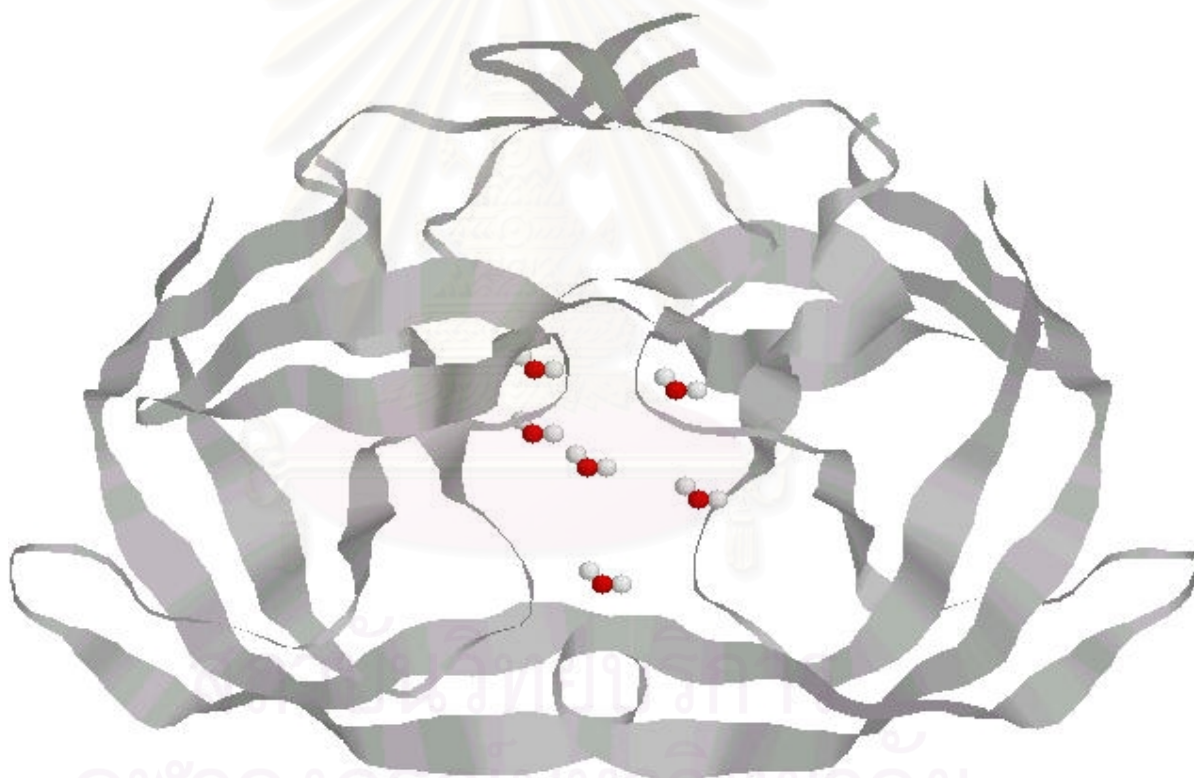


Figure 4.2 X-ray crystal structure of HIV-1 PR (pdb code:1HPX). Six waters were shown in ball and stick model. Catalytic aspartic acids were shown in stick model. Another residues shown in ribbons diagram.

As hydrogen atoms are not available in the X-ray structure. These all hydrogen coordinates were then added to the X-ray crystallographic coordinates not only hydrogen in amino acids but also in the ionization states of side chains for charged amino acid residue as well as at N-terminal amine and C-terminal carboxylic groups.

All added hydrogens were later optimized using AMBER 5.0 program. The getting structure was used as an initial structure.

Definition of Studied Water in HIV-1 Protease

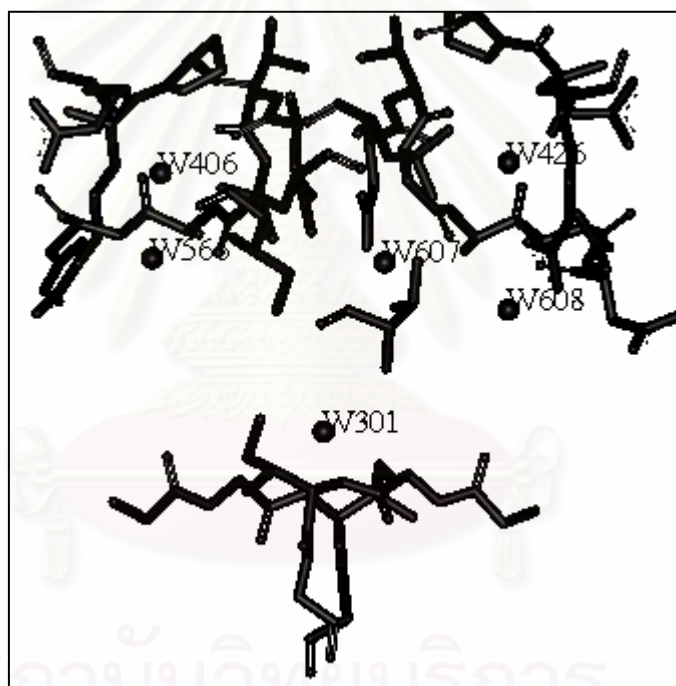


Figure 4.3 The position of studied six water molecules taken from X-ray crystallographic structure.

The positions of water were collected from the PDB. Their positions and names are known as:

Notation	PDB position	Description
WAT301, W301	301	Water at the flap site (tetrahedrally H-bonded water molecule)
WAT406, W406	406	Left side water
WAT426, W426	426	Right side water
WAT566, W566	566	Left side water
WAT607, W607	607	Lytic (catalytic) water molecule
WAT608, W608	608	Right side water

4.2 Model representations

Due to the size of the HIV-1 protease which consists of 198 amino acids (Figure1), it is not possible to take into account the whole structure in quantum chemical calculations, even at a very low level of accuracy. Thus, the QM/MM approach which designed to alleviate the size problem was used in the study.

To find the model of optimal size, the WAT607 and amino acids in its vicinity were used. Models I and II were investigated. In QM part, two methods, HF and DFT (B3LYP) were applied and before and results were compared.

The bigger cluster of amino acid were expanded according to Model I and Model II, (see 4.2.1 and 4.2.2). The interaction energy were calculated and reported as the previous model until the consistency of interaction energy. The model of cluster were expanded, the investigated energy shown in Table 5.1 and Table 5.2.

In this step, each cluster of amino acids from previous step was completely added hydrogen, both of all added hydrogens and water hydrogens was optimized with Hyperchem 6.0 program.

4.2.1 The expansion of Model Size: Model I

Model I: The interaction between water molecule and amino acid residues (Res) in the $H_2O(\text{Res})_m$ clusters, $\Delta E_{(\text{mod } 1)}$ where, $m = 1, 2, 3, \dots$ was calculated according to equation [4.1];

$$\Delta E_{(\text{mod})} = E(\text{cst}) - E(H_2O) - E(\text{Res})_m \quad [4.1]$$

where the energies on the right hand side are total energies obtained from quantum chemical calculations of the cluster; $E(\text{cst})$, water; $E(H_2O)$ and m amino acid; $E(\text{Res})_m$, respectively. Here, the first amino acid ($m = 1$) in the $E(\text{Res})_m$ cluster is the one that lying closest to oxygen atom of the water molecule. For $m = 2, 3, 4, \dots$, the next residue which binds to the previous one in the amino acid sequence was included. Extension was made until either consistency of the interaction energy (from eq. [4.1]) or limitation in computation was reached.

4.2.2 The expansion of Model Size: Model II

Model II: The water-residues interaction, $\Delta E_{(\text{mod } 2)}$ in the $H_2O(\text{Res})_{n(r)}$ cluster was calculated using equation [4.1]. In this model, cluster size was increased as a function of spherical radius, r center at oxygen atom of the water, i.e., n is defined as number of amino acid residues lying within a sphere of radius r , where $r = 3, 3.5, 4, 4.5, \dots \text{ \AA}$, from oxygen atom of the water. The condition is that if each atom of any amino acid is detected within the spherical radius, the whole molecule of that residue will be included into the model. The same criteria as that of *Model I* has been also used to characterize the investigated model, i.e., cluster size was extended, in terms of r_i until consistency of the interaction energy was reached. At this state, the obtained results, is considered as the optimal cluster size or optimal spherical radius, r_{opt} for each water molecule.

4.3 Computation: Single point calculation

Calculation techniques used in this study can be categorized into two types i.e. the full quantum mechanics calculation (QM) and the combined quantum mechanics and molecular mechanics calculation (QM/MM). All calculations were performed using Gaussian 98 program on the PC-workstation the Austrian-Thai Center for Computer Assisted Chemical Education and Research (ATC), Chulalongkorn University.

For water molecule, experimental geometry with the O-H distance of 0.975 Å and H-O-H angle of 104.5° taken from the literature, has been applied and kept constant throughout. As mentioned in 4.1, coordinate of the oxygen atom taken from the PDB structure of the HIV-1 protease was kept fixed. Here, its orientation around oxygen atom was optimized using *ab initio* calculation with 6-31G** basis set.

4.3.1 QM Calculations

The Density Functional Theory with B3LYP functional and the Hartree-Fock method were used as QM methods of choice. Types of basis sets, i.e., 6-31G* for all atoms of amino acids and 6-31G** basis set for water molecule.

4.3.2 QM/MM Calculations

To perform QM/MM calculations, the molecule was partitioned into two levels where two different levels of theory were treated. Two methods of high level, HF and DFT (B3LYP), have been parallel applied to the crucial part or “model system”. The second level, the low level or “real system”, was treated using Molecular Mechanics method (MM). The single point energy of QM/MM was then calculated.

4.4 Evaluations: Water-Enzyme Interactions

The calculations have been performed for the $H_2O(\text{Res})_m$ (Model I) and $H_2O(\text{Res})_{n(r)}$ (Model II) clusters, varying m and r as mention in (4.2). Three types

of method namely, molecular mechanic (MM) with the MM+ forcefield, Hartree-Fock (HF) and density functional theory (DFT) with (B3LYP) method with 6-31G* basis set was applied for all amino acid and 6-31G** for water molecule. MM calculations were performed using Hyperchem 6.0 program.

4.4.1 interaction with each water molecule

The interaction energies of six centered water molecule (W607, W301, W406, W566, W426, and W608) with amino acid residues have been calculated according eq. [4.1].

4.4.2 Interaction with two water molecules

Questions arise concerning reliability of the models proposed in (i) when the two water molecules lie within the R_{opt} of each other, for instance W406 and W566. To clarify this, quantum chemical calculations have been performed for the bigger cluster consisting of the two nearest water molecules and all amino acids which any of their atoms locate within the spherical radius of R_{opt} from oxygen atom of the two water molecules. The interaction energy per mole of water can be computed from half of ΔE_{2w} showing in eq. [4.2],

$$\Delta E_{2w} = E(cst) - 2E(H_2O) - E(Res1, Res2) \quad [4.2]$$

where Res1 and Res2 are amino acid residues lying within R_{opt} for the two water molecules.

4.4.3 Interaction with three water molecules

The same reason as in case of two water molecules, the interaction energy between the three water molecules, W426, W607, and W608, with all amino acids lying in their R_{opt} has been calculated using equation [4.3]

$$\Delta E_{3w} = E(cst) - 3E(H_2O) - E(Res1, Res2, Res3) \quad [4.3]$$

in which number of amino acid residues lying within R_{opt} of the three water molecules, Res1, Res2, Res3, were obtained from (4.3).



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 5

RESULTS

The results have been shown in terms of binding energies. They are divided into two parts to search for the optimal models, to search for the cluster of optimal size for single, double and triple water molecules.

5.1 Optimal calculated model

To seek for the cluster of optimal size, W607 was chosen to be tested system. Binding energies yielded from the MM, HF, DFT and QM/MM methods for models I and II (detailed in 4.2) were summarized in Table 5.1 and 5.2 with the corresponding images shown in Figures 5.1 and 5.2, respectively.

Table 5.1 The water-enzyme interaction energies for lytic water (W607) where cluster size was increased according to the Model I (MM energy (Real) = -1.45 kcal/mol).

Model I: H ₂ O(Res) _m			Δ E (kcal/mol)			Δ E _{QM/MM} (kcal/mol)	
m	Residues	No. of atoms	MM	HF	DFT	HF	DFT
2	D25, D25'	27	-0.28	-3.45	-4.35	-4.62	-5.52
4	D25, T26, D25', T26'	61	-0.33	-3.43	-4.24	-4.55	-5.36
6	D25, T26, G27, D25', T26', G27'	75	-0.81	-3.19	-4.92	-3.83	-5.56
8	D25, T26, G27, A28, D25', T26', G27', A28'	95	-0.91	-4.06	-6.40	-4.60	-6.94
10	D25, T26, G27, A28, D29, D25', T26', G27', A28', D29'	121	-0.93	-4.14	-5.76	-4.66	-6.28

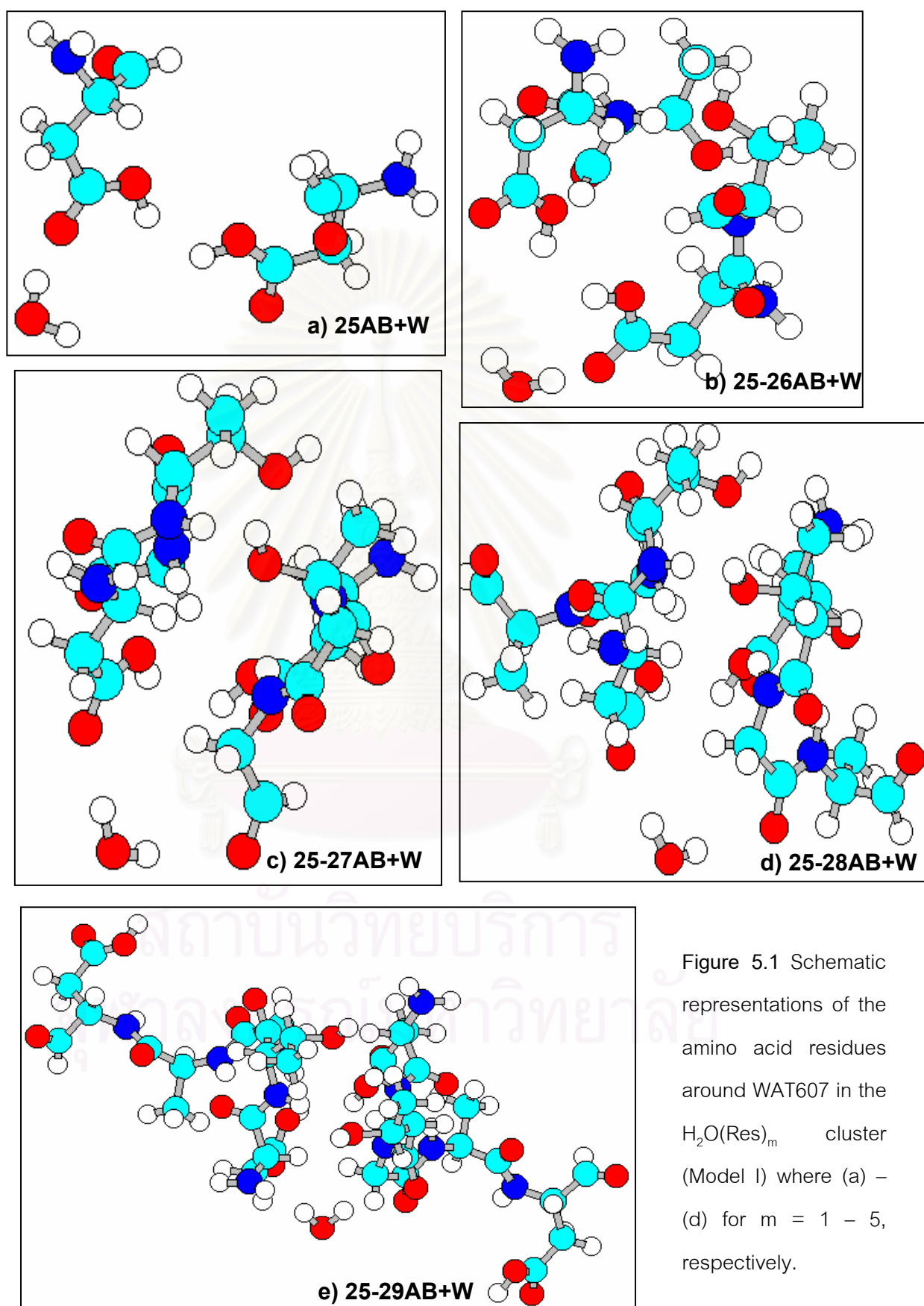
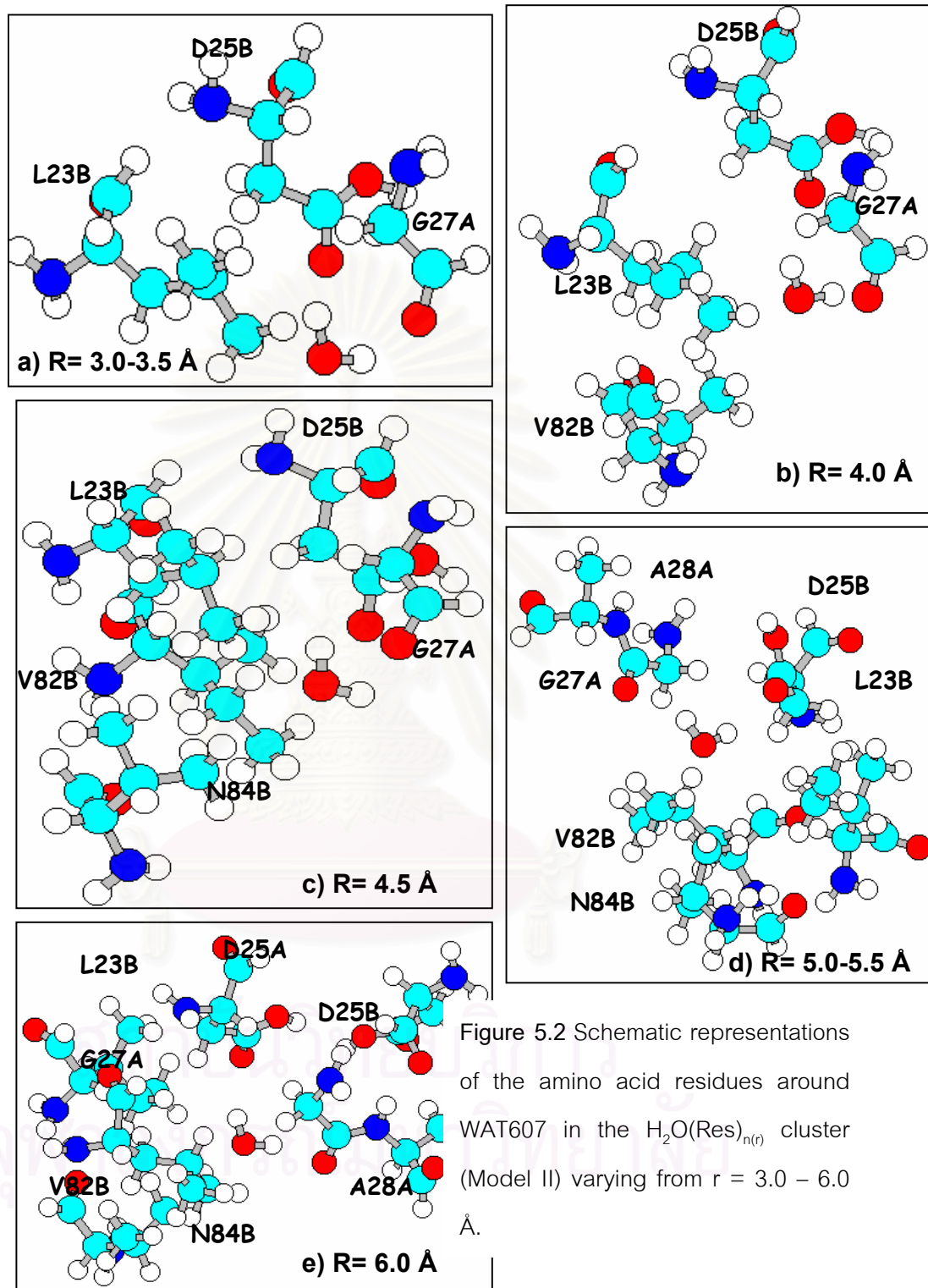


Table 5.2 The water-enzyme interaction energies for lytic water (W607) where cluster size was increased according to the Model II (MM energy (Real) = -1.45 kcal/mol).

Model II: H ₂ O(Res) _{n(r)}				Δ E (kcal/mol)			Δ E _{QM/MM} (kcal/mol)	
<i>r</i> (Å)	<i>N</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.5	3	G27,A23',D25'	48	-0.24	+1.60	-1.21	+0.39	-2.42
4.0	4	G27,A23',D25',V82'	66	-0.59	+1.31	-1.71	+0.45	-2.57
4.5	5	G27,L23',D25',V82',I84'	87	-0.99	+1.06	-2.24	+0.60	-2.70
5.5	6	G27,A28,L23',D25',V82',I84'	97	-1.06	-5.64	-8.92	-6.03	-9.31
6.0	7	D25,G27,A28,L23',D25',V82',I84'	112	-1.11	-5.30	-8.62	-5.64	-8.96

5.2 Optimal cluster size and its interaction for the single water cluster, H₂O(Res)_{n(r)}

With the optimal cluster model, model II, this strategy of increasing cluster size has been applied throughout. All interaction energies of single water cluster, H₂O(Res)_{n(r)}, were reported in Tables 5.3 - 5.7 and, with the corresponding image in Figures 5.3 – 5.7. Schematic representation of the six water molecules were given, again, in Figure 5.8. Energy data was plotted in Figure 5.9. The optimal cluster size extrapolated from the plots and collected from Table 5.3 – 5.7 were summarized in Table 5.8.



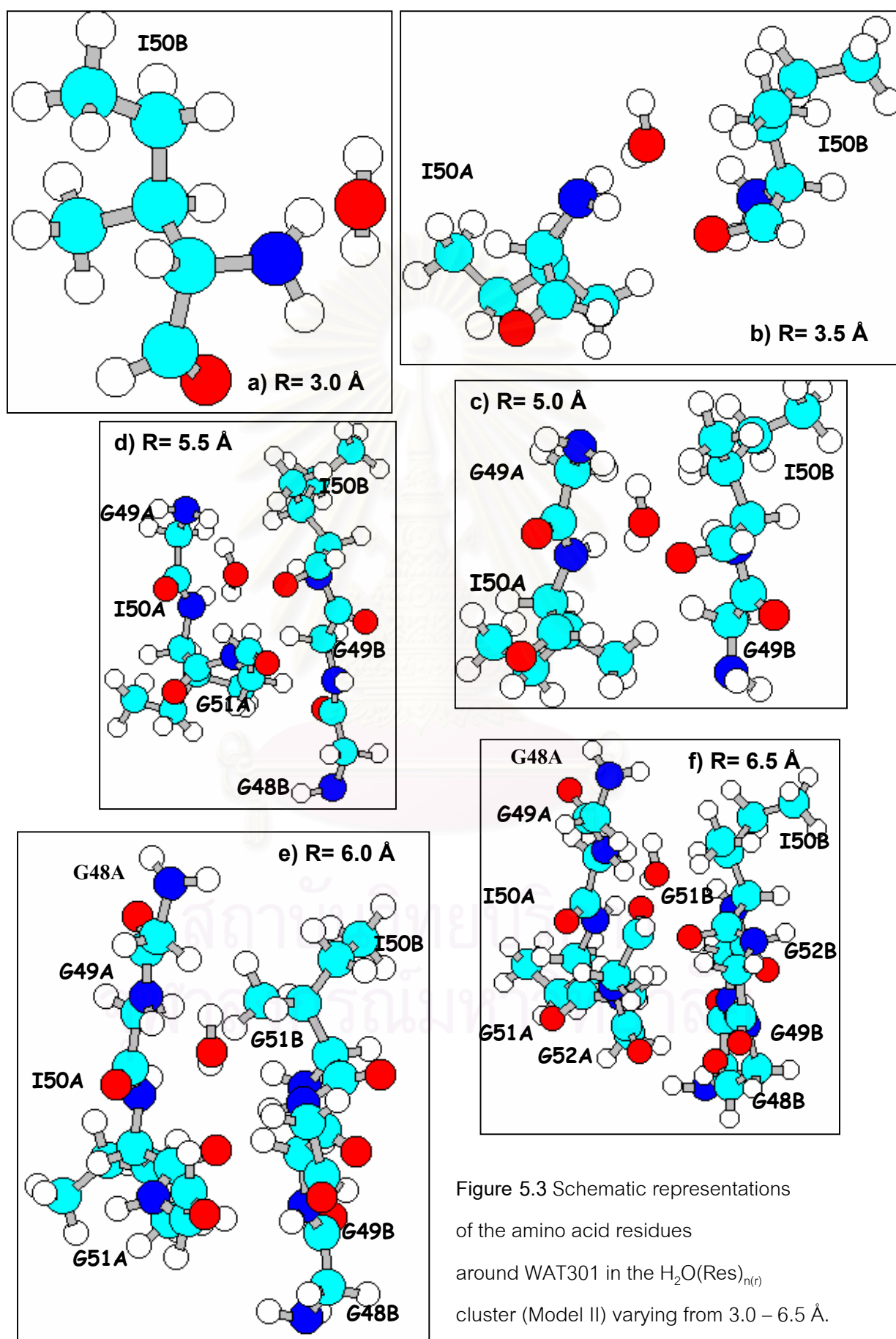


Table 5.3 The water-enzyme interaction energies of tetrahedrally hydrogen-bond structural water molecule (W301) according to Model II (MM energy (Real) = - 069 kcal/mol)

Model II: H ₂ O(Res) _{n(r)}				Δ E (kcal/mol)			Δ E _{QM/MM} (kcal/mol)	
<i>r</i> (Å)	<i>n</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.0	1	I50'	24	+0.50	-2.04	-4.14	-3.23	-5.33
3.5	2	I50,I50'	45	+0.31	-1.62	-2.61	-2.62	-3.61
5.0	4	G49,I50,G49',I50'	59	-0.18	-5.42	-9.36	-5.82	-9.87
5.5	6	G49,I50,G51,G48',G49',I50'	73	-0.29	-7.18	-10.99	-7.58	-11.39
6.0	8	G48,G49,I50,G51,G48',G49',I50',G51'	87	-0.35	-7.75	-11.49	-8.09	-11.83
6.5	10	G48,G49,I50,G51,G52,G48',G49',I50',G51',G52	101	-0.40	-7.89	-11.53	-8.18	-11.82

Table 5.4 The water-enzyme interaction energies of W406 according to the Model II (MM energy (Real) = -1.38 Kcal/mol)

Model II: H ₂ O(Res) _{n(r)}				Δ E (kcal/mol)			Δ E _{QM/MM} (kcal/mol)	
<i>r</i> (Å)	<i>n</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.5	2	T26',R87'	46	+0.89	+1.76	-2.02	-0.51	-4.29
4.0	7	R8,P9,L23,T26',G27',D29',R87'	130	-0.67	+4.65	-2.69	+3.94	-3.40
5.0	8	R8,P9,L23,T26',G27',A28',D29',R87'	138	-0.99	+5.19	-2.88	+4.80	-3.27
5.5	9	Q7,R8,P9,L23,T26',G27',A28',D29',R87'	155	-1.05	+5.08	-2.69	+4.75	-3.02
6.0	11	Q7,R8,P9,L23,L24,D25',T26',G27',A28',D29',R87'	187	-1.13	+4.30	-3.36	+4.05	-3.61

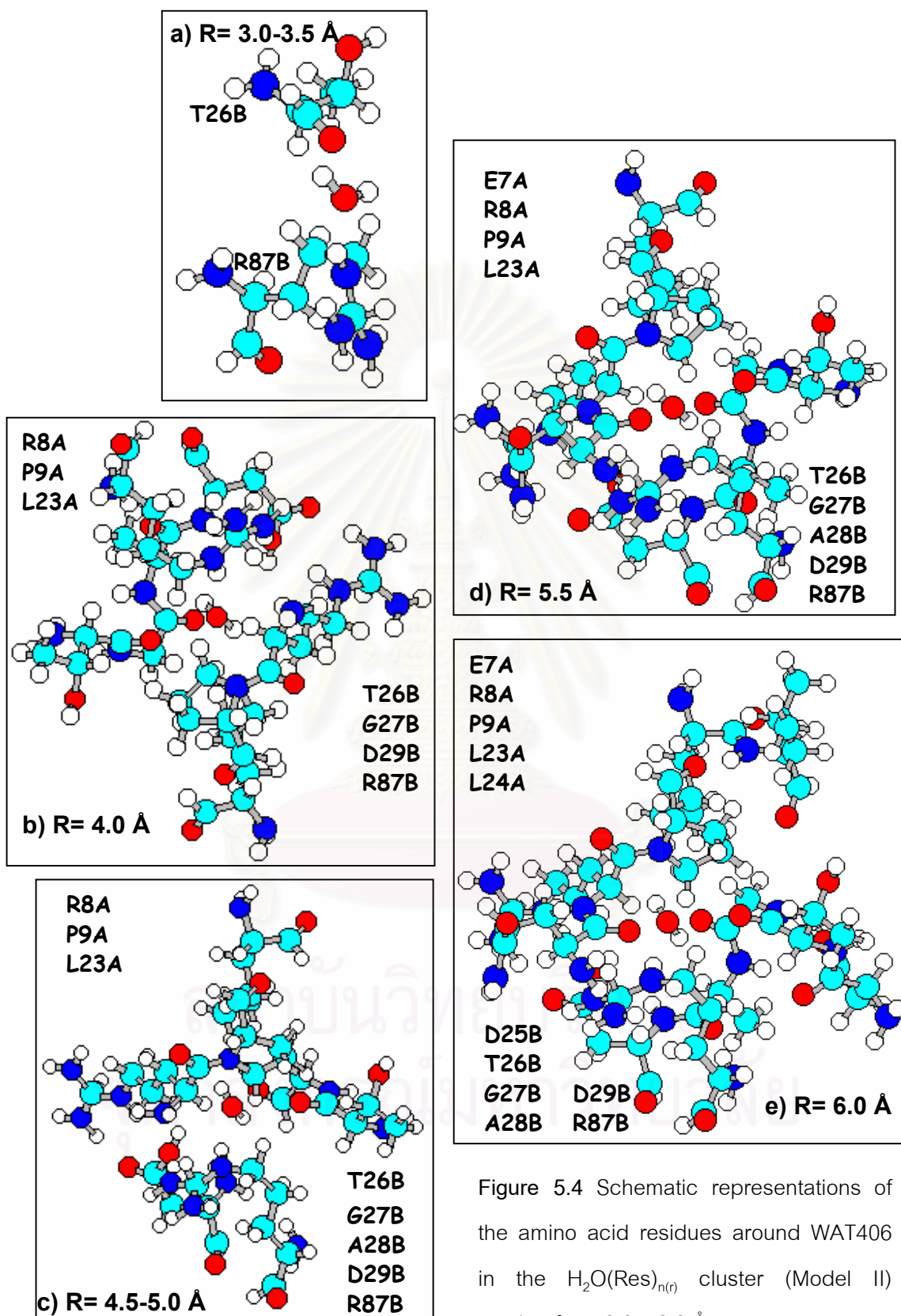


Figure 5.4 Schematic representations of the amino acid residues around WAT406 in the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster (Model II) varying from 3.0 – 6.0 Å.

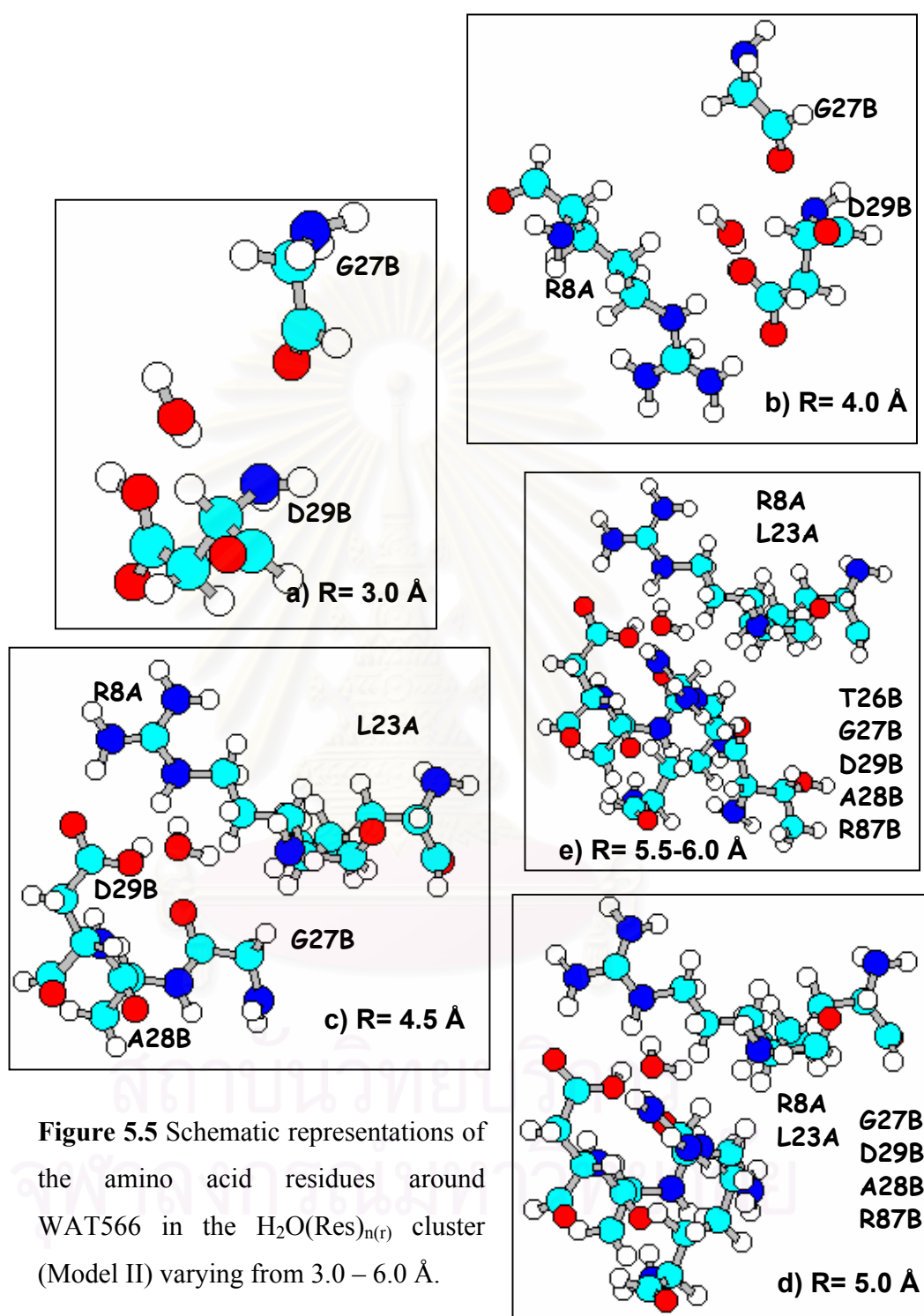


Table 5.5 The water-enzyme interaction energies of W566 according to the Model II
(MM energy (Real) = +0.47 kcal/mol)

Model II: H ₂ O(Res) _{n(r)}				Δ E (kcal/mol)			Δ E QM/MM	
<i>r</i> (Å)	<i>n</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.0	2	G27',D29'	27	+2.82	+1.52	-1.51	-0.83	-3.86
4.0	3	R8,G27',D29'	54	+2.21	-8.80	-12.53	-10.54	-14.27
4.5	5	R8,L23,G27', A28',D29'	83	+1.65	-10.19	-13.87	-11.27	-15.05
5.0	6	R8,A23,G27',A28', D29',R87'	110	+1.57	-11.28	-14.35	-12.38	-15.44
6.0	7	R8,L23,T26',G27',A2 8',D29',R87'	124	+1.51	-11.64	-14.52	-12.68	-15.56

Table 5.6 The water-enzyme interaction energies of W426 according to the Model II
(MM energy (Real) = -2.53 Kcal/mol).

Model II: H ₂ O(Res) _{n(r)}				Δ E (kcal/mol)			Δ E QM/MM (kcal/mol)	
<i>r</i> (Å)	<i>n</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.0	2	T26,D29	34	-0.12	-1.52	-3.73	-3.93	-6.14
3.5	3	T26,G27,D29	41	-0.64	-0.22	-2.91	-2.11	-4.80
4.0	6	T26,G27,A28,D29, R8',P9'	90	-1.25	-3.09	-7.05	-4.37	-8.33
4.5	7	T26,G27,A28,D29, R8',P9',L23'	111	-1.65	-3.27	-7.47	-4.15	-8.35
5.5	8	T26,G27,A28,D29, R87,R8',P9',L23'	138	-1.32	-1.31	-8.11	-2.52	-9.32
6.0	10	D25,T26,G27,A28, D29,R87, Q7',R8',P9',L23'	168	-2.11	-2.62	-8.77	-3.04	-9.19

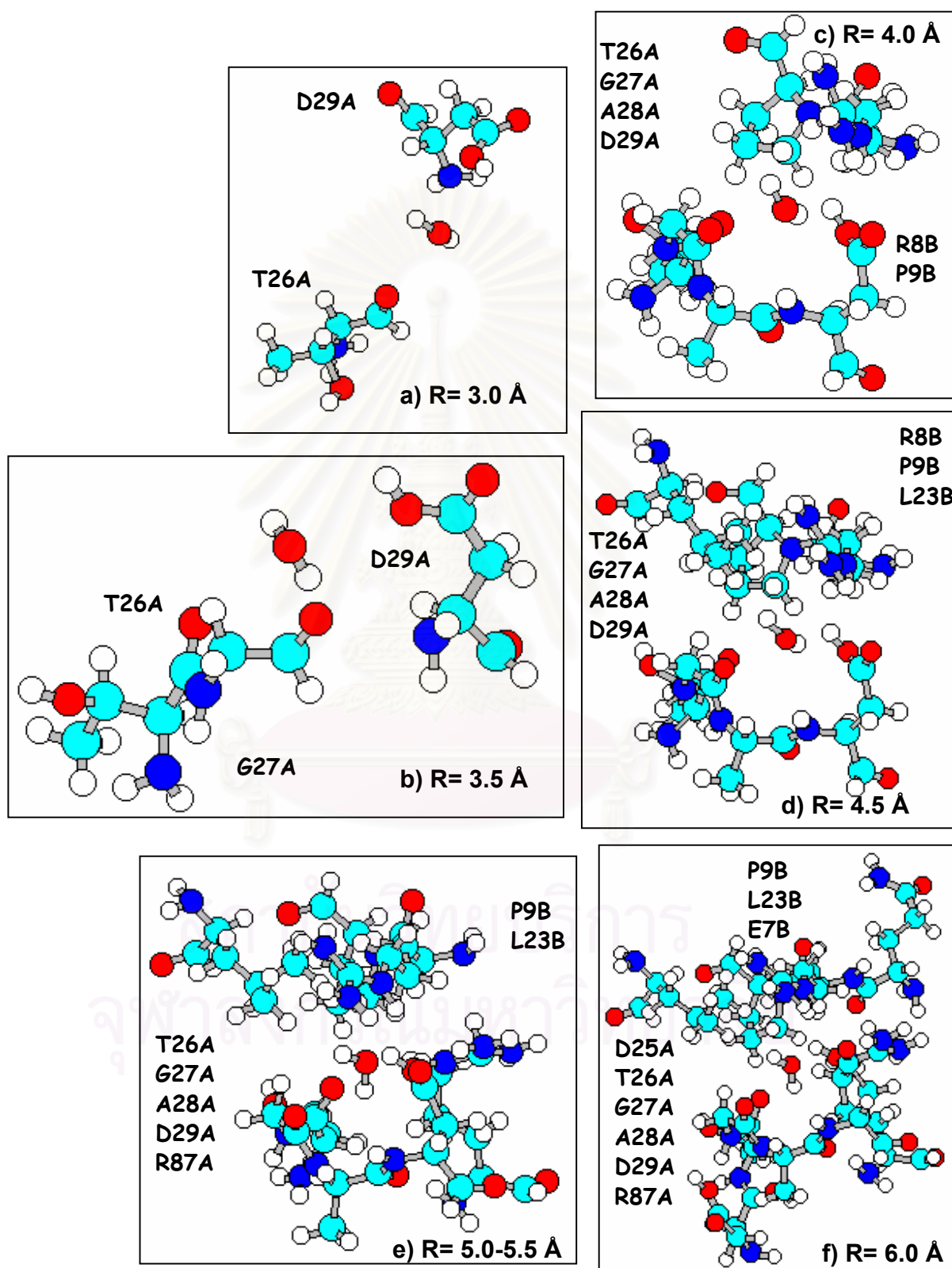


Figure 5.6 Schematic representations of the amino acid residues around WAT566 in the $H_2O(Res)_{n(f)}$ cluster (Model II) varying from 3.0 – 6.0 \AA .

Table 5.7 The water-enzyme interaction energies of W608 according to the Model II(MM energy (Real) = -1.17 Kcal/mol).

Model II: H ₂ O(Res) _{n(r)}				ΔE /(kcal/mol)			$\Delta E_{QM/MM}$ kcal/mol)	
r (Å)	n	Residues	No. of atoms	MM	HF	DFT	HF	DFT
3.5	2	G27,D29	27	-0.83	-4.90	-5.98	-5.24	-6.32
4.5	3	G27,A28,D29	35	-0.58	-8.19	-9.55	-8.78	-10.14
5.0	4	G27,A28,D29, D30	48	-0.71	-8.50	-10.05	-8.96	-10.51
6.0	6	G27,A28,D29, D30,G48,R8'	84	-0.80	-9.37	-10.78	-9.74	-11.15

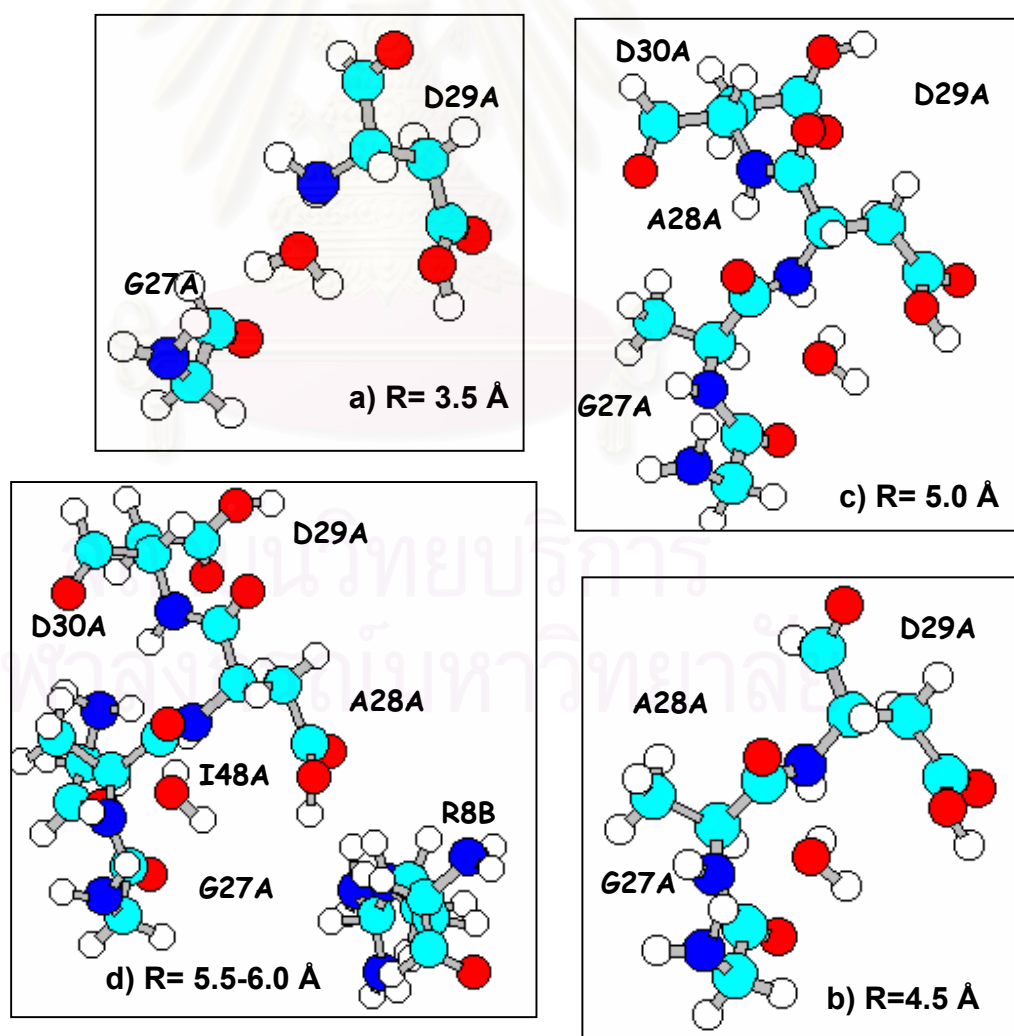


Figure 5.7 Schematic representations of the amino acid residues around WAT608 in the H₂O(Res)_{n(r)} cluster (Model II) varying from 3.0 – 6.0 Å.

Table 5.8. The water-enzyme interaction energies according the *Model II* expansion for all investigated water were shown in three different methods: Interaction with each water molecule.

Model II: H ₂ O(Res) _{n(r)}					ΔE /(kcal/mol)			$\Delta E_{QM/MM}$ (kcal/mol)	
WAT No.	<i>r</i> (Å)	<i>n</i>	Residues	No. of atoms	MM	HF	DFT	HF	DFT
301	6.0	8	G48,G49,I50, G51, G48',G49',I50', G51'	87	-0.36	-7.75	-11.49	-8.09	-11.83
607	5.5	6	G27,A28, L23',D25',V82', I84'	97	-1.06	-4.83	-8.92	-6.03	-9.31
406	5.0	8	R8,P9,L23,T26', G27', A28',D29',R87'	138	-0.99	+5.19	-2.88	+4.80	-3.27
426	5.5	8	T26,G27,A28, D29,R87, R8',P9',L23'	138	-2.02	-1.31	-8.11	-2.52	-9.32
566	5.0	6	R8,L23, G27',A28',D29', R87'	110	+1.57	-11.28	-14.35	-12.38	-15.44
608	5.0	4	G27,A28,D29, D30	48	-0.71	-8.50	-10.05	-8.96	-10.51

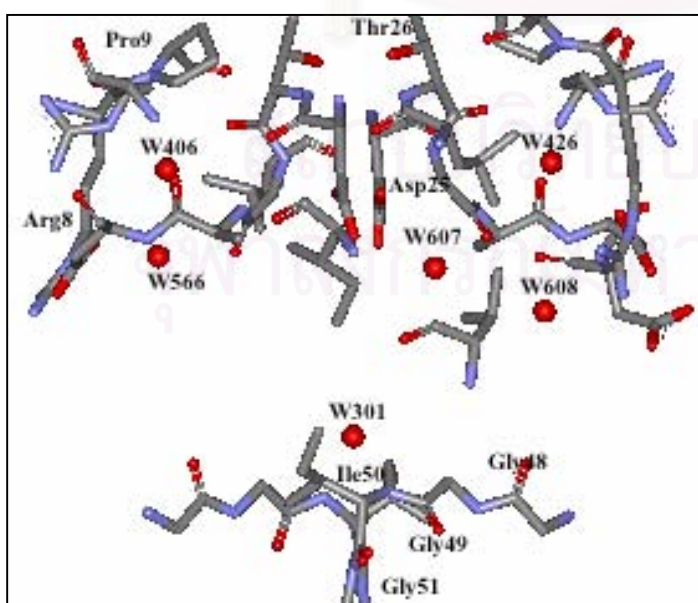


Figure 8. Schematic representation of the six water molecules lying in the amino acid environments in the HIV-1 protease pocket.

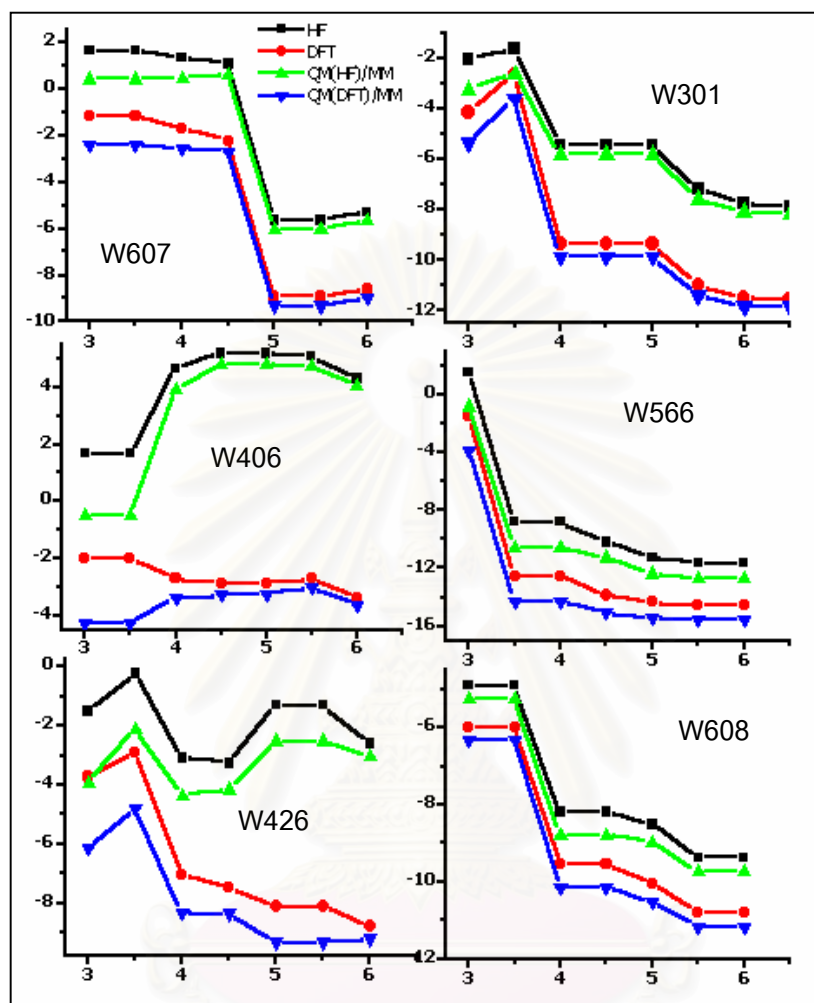


Figure 5.9 Changes of the water-enzyme interaction energy in the $\text{H}_2\text{O}(\text{Res})_n(r)$ cluster calculated using different methods for each water molecule, where r denotes spherical expansion radius according to model II (see text for more details)

5.3 Optimal cluster size and its interaction for the double water cluster, $(\text{H}_2\text{O})_2(\text{Res})_{n(r)}$

To clarify reliability of the single water model, $\text{H}_2\text{O}(\text{Res})_{n(r)}$, the cluster containing of 2 water molecules, $(\text{H}_2\text{O})_2(\text{Res})_{n(r)}$ has been considered. The calculated results are given in Table 5.9.

Table 5.9 The water-enzyme interaction energies for the two closely water molecules, where the cluster size was increased reaching to the model II. (MM energy (Real) = -0.39 and -3.72 Kcal/mol, respectively)

Model II			ΔE (kcal/mol)			$\Delta E_{\text{QM/MM}}$ (kcal/mol)	
WAT no.	Included Residues	No. of atoms	MM	HF	DFT	HF	DFT
406+566	R8,P9,L23,T26',G27',A28',D29',R87'	141	+0.94	-4.19	-17.5	-5.52	-18.38
426+608	T26,G27,A28,D29,D30,R87,R8',P9',L23'	154	-2.89	-11.71	-19.03	-12.54	-19.86

5.4 Optimal cluster size and its interaction for the triple water cluster, $(\text{H}_2\text{O})_3(\text{Res})_{n(r)}$

The same reason as for double-water cluster model, the triple water one, $(\text{H}_2\text{O})_3(\text{Res})_{n(r)}$ has been calculated. The results are in Table 5.10.

Table 5.10 The water-enzyme interaction energies for the three closely water molecules, where the cluster size was increased reaching to the model II. (MM energy (Real) = -5.22 Kcal/mol)

Model II			ΔE (kcal/mol)			$\Delta E_{\text{QM/MM}}$ (kcal/mol)	
WAT no.	WAT no.	WAT no.	MM	HF	DFT	HF	DFT
426+608+607	T26,G27,A28,D29,D30,R87,R8',P9',L23',D25',V82',I84'	211	-4.22	-11.09	-22.38	-12.09	-23.38

CHAPTER 6

DISCUSSION

Aims of this study is to investigate water's role in terms of interaction energies between water and enzyme. It's generally known and have already mentioned before that QM calculation is too expensive or impossible to perform using whole protein molecule. Therefore, QM/MM is alternative way for this propose.

The calculated results have been shown in Tables 5.1 to 5.10. Two possible strategies for expanding the cluster size (details in Topic 4.2) were applied. The interaction energy between water and amino acid cluster were evaluated for single water cluster, $\text{H}_2\text{O}(\text{Res})_{n(r)}$. The results were given in Tables 5.2 - 5.7 and the optimal cluster size for each water was summarized in Table 5.8. The double water and triple water clusters, $(\text{H}_2\text{O})_2(\text{Res})_{n(r)}$ and $(\text{H}_2\text{O})_3(\text{Res})_{n(r)}$ were also investigated and the result shown in Tables 5.9 and 5.10, respectively.

6.1 Cluster sized-Expansion of Model

To seek for the optimal expansion method, WAT607 was used to be tested model. The calculations have been performed for the $\text{H}_2\text{O}(\text{Res})_m$ and $\text{H}_2\text{O}(\text{Res})_{n(r)}$ clusters, varying m and r as already mentioned and the results were given in Tables 5.1 and 5.2. In the $\text{H}_2\text{O}(\text{Res})_m$ cluster where $m = 2$, residues 25 and 25', the water (WAT607) – residues interaction obtained from the HF calculations in Table 5.1 of - 3.45 kcal/mol is slightly higher than that of - 4.35 kcal/mol from the DFT method. This leads, consequently, to the lower ΔE -QM/MM yielded from the DFT than that from the HF calculation. As expected, the MM water-residues interaction for the $\text{H}_2\text{O}(\text{Res})_2$ cluster is much higher than the QM

and QM/MM energies. Increase of the cluster size, according to the linear model, the interaction decrease slowly.

As can be seen in Table 5.2 for Model II, $\text{H}_2\text{O}(\text{Res})_{n(r)}$, the cluster with $r = 3.5 \text{ \AA}$ accumulates 3 amino acids ($n = 3$), residues 27, 23' and 25'. The corresponding MM, HF and DFT interaction energies are - 0.24, + 1.60 and - 1.21 kcal/mol, respectively. Unexpected result is that of the HF calculation in which the interaction is slightly positive. This situation remains unchanged up to $r = 4.5 \text{ \AA}$. Similar to that of Model I (linear expansion), the MM energies for Model II are almost the same when the cluster size increases. The most appropriate trend is the decrease of interaction energy according to the DFT method in which the energy is slightly decrease for $r \leq 4.5 \text{ \AA}$ and remains constant at about - 9 kcal/mol for $r \geq 5.0 \text{ \AA}$. This value is significantly lower than that of the model I (DFT energy for $m = 8 - 10$ in Table 5.1). A clear conclusion is that spherical expansion of the cluster size, model II, is more-reliable than the linear one. Therefore, Model II was, then, used for further investigations. Additional weak point of model I is due to the fact, for examples, that residues 23', 82' and 84' which locate at the distances cluster than 3.0 \AA , 4.0 \AA and 4.5 \AA to the oxygen atom of WAT607, respectively, were not taken into consideration.

Taken into account all the energy data shown in Table 5.1 and 5.2, same other conclusions can be also made. (i) MM method, and hence QM/MM, fails to describe the water-enzyme interactions, at least for the investigated system. (ii) Between the HF and the DFT/B3LYP, the latter is superior. (iii) For WAT607, the optimal $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster size for the DFT calculations is $r = 5.0 \text{ \AA}$, consisting of 6 amino acid residues, 27, 28, 23', 25', 82' and 84' (Table 5.2).

As a function of the cluster size defined by Models I and II (see equation 4.2.1), the correspondence; residue numbers were displayed in Figure 5.1 and 5.2, respectively. According to the Tables 5.1 and 5.2, the interaction energies indicate that model II is the suitable model for the investigation. Therefore, Model II was, then applied for the other water molecules. The cluster size increasing according to Models I and II has been shown in Figure 5.1 (a-e) and Figure 5.2 (a-e) in the stick and ball representation. The energy results of each method have been shown. Variances in both model and method lead to different in computed interaction energy.

The result in Table 5.1 showed that the expansion of Model I which increases the cluster size along amino acid chain is not appropriated because interaction energy does not change when number of amino acid increase. For Model II shown in Table 5.2, the interaction energy decreases as a function of cluster size and is consistent at certain radius, 5.0-5.5 Å. As a results of two Tables, Model II-cluster sized expansion is the appropriate for investigation. And the optimal cluster size of amino acid around WAT607 is 5.5 Å.

Additionally, semi-empirical method have been applied to examine our tested model results. The cluster size $\text{H}_2\text{O}(\text{Res})_{n(r)}$ was increased beyond the QM and QM/MM calculations. The results were in Appendix I. It was found that the semi-empirical methods fail to describe the system.

6.2 Water-Enzyme Interaction energies

6.2.1 Single Water Cluster

With the optimal expansion strategy (Model II), six single waters were investigated by considering the interaction energies when the cluster of amino acid radius was increased. The best model of six single waters were shown in Table 5.8. Four waters; WAT301, WAT607, WAT566 and WAT608, the energy trends and the difference of HF and DFT seem to be normal which is around 3 – 5 kcal/mol. These confirm what reported experimentally that WAT301 plays role in controlling movement of the flaps by participating in the hydrogen bonding between the flaps and the inhibitor. The observed stabilization energies of – 14.35 kcal/mol for WAT566 and of – 10.65 kcal/mol for WAT608 also confirm their role in stabilizing the overall structure of the HIV-1 protease. However, the interactions of some waters is abnormal e.g. WAT406 and WAT426. For the these two structural water molecules, WAT426 and WAT406, the calculated energy of – 8.11 kcal/mol for WAT406 supports its role in holding the enzyme structure. On the other hand, the energy of – 2.88 kcal/mol for WAT426 indicates very clear that it doesn't

form hydrogen bond with the amino acid residues. Therefore, its role in stabilizing enzyme structure is supposed to differ from the other structural water molecules. Instead of holding the enzyme by forming the hydrogen bond, it supposes to repel the residues in order to prevent them from getting collapse.

The interactions are shown to be repulsive or mildly attractive force. Both waters stay in region of side or wing of protease. Moreover, it seems that other waters near by contribute to stabilize the overall structure of protease and for many processing. From this results, the interaction of each single water with amino acid cluster is not enough. The double water interactions, consequently, were calculated with their cluster of amino acid.

The side waters (WAT406 and WAT566 in left side or WAT426 and WAT608 in right side) do not far from each other. And it seems to be affected for the interaction energies. Therefore, the interaction with two waters had been calculated and shown in Table 5.9. As in Table 5.10, there are three waters within 6 Å. These interaction energies were calculated to investigate the interaction of water molecules and amino acid environment.

6.2.2 Double Water Cluster

Table 5.8, water-enzyme interactions for the $(\text{H}_2\text{O})_2(\text{Res})_{n(r)}$ cluster were summarized. The results show that sum of the DFT interaction energies obtained from each single-water cluster models of -17.23 kcal/mol for WAT406+WAT566 and of -18.16 kcal/mol for WAT426+WAT608 are almost the same as those of the double-water cluster models shown in Table 5.9. This indicates that individual cluster is sufficient to represent water-enzyme interaction.

6.2.3 Triple Water Cluster

From Table 5.10, the interaction energies of three closely related water molecules with amino acid within $r = 6 \text{ \AA}$ were calculated. Sum of the DFT individual interaction of WAT 426, WAT608 and WAT607 of -27.08 kcal/mol is significantly lower than that of -22.38 kcal/mol of the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster (Table 5.10). The difference of 4.06 kcal/mol can be arisen from different amino acid residues in the cluster and interaction energy among water themselves.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER 7

CONCLUSION

7.1 Model Representation

To calculate water-HIV-1 PR interaction, it was found that spherical expansion of the cluster size, model II, is more reliable than the linear one, Model I. The optimal cluster size of amino acid around WAT607 is 5.5 Å. Interaction energy has been affected by spherical increasing the cluster size of amino acid.

Additional weak point of model I is due to the fact that it did not include amino acid at the distances closer even than 3.0 Å to the oxygen atom of WAT607.

7.2 The Optimum Method for Calculating the Water Amino Acid Cluster

The MM water-residues interaction is much higher than the QM and QM/MM energies. Unexpected result is that some of the HF calculations for W406 give positive values while the DFT, for the same cluster, yields attractive interaction. The most appropriate trend is the decrease of interaction energy according to the DFT method in which the energy is slightly decrease for $r \leq 4.5$ Å and remains relatively constant for $r \geq 5.0$ Å.

7.3 Water's Role from this Study

The interaction energies of WAT301, WAT607, WAT566, WAT608 are as follows; - 11.49, - 8.92, - 14.35 and - 10.05 kcal/mol, respectively. Role of water molecules both in hydrolysis mechanism and in stabilizing the enzyme structure have been analyzed in terms of interaction energy and found that the calculated values for W301, W607, W566, W608, W406 and W426 are as follows; - 11.49, - 8.92, - 11.35, - 10.05, - 2.88 and - 8.11 kcal/mol, respectively. For W406, the interaction of - 2.88 kcal/mol suggested that its role in stabilizing enzyme structure is supposed to differ from the other structural water molecules, i. e., instead of holding the enzyme by forming the hydrogen bond, it supposes to repel the residues in order to prevent them from getting collapse.

7.4 Concluding Comments

It is the first time that interaction energies between water molecules and the surrounded amino acids have been investigated systematically using high accurated quantum chemical calculations. The obtained results confirm their role either in catalytic process or in stabilizing the enzyme structure. Due to the requirement of a numerous computer time, this prevents the case of the high accurate methods. Therefore, suggestions for further study, in which at that time computer would be much cheaper and the speed would be much faster, to increase reliability of the results are to perform calculations:

- (i) using the bigger basis set and/or the higher corrected method,
- (ii) using the whole six-water cluster, $(\text{H}_2\text{O})_6\text{Res}_n(r)$,

- (iii) by optimizing oxygen coordinates of water molecules, or the best if it is possible is to
- (iv) optimize all positions and orientations of all water molecules and all atoms of residues which are included in the clusters.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

REFERENCES

- 1) Turner, B. G.; Summers, M. F. Structural Biology of HIV., *J. Mol. Biol.* **1999**, 285, 1-32.
- 2) <http://www.niaid.nih.gov/daids/dtpdb/intro.htm>
- 3) Tomasselli, A. G.; Heinrikson, R. L. Targeting the HIV-protease in AIDS therapy: a current clinical perspective., *Biochim. et Biophys. Acta* **2000**, 1477, 189-214.
- 4) Erickson, J. W.; Eissenstat, M. A., Protease of Infected Agent, Academic Press. **1999**.
- 5) <http://www.aidsinfony.org>
- 6) Todd, M. J.; Semo, N.; Freire, E. The Structural Stability of the HIV-1 Protease. *J. Mol. Biol.*, **1998**, 283, 475-488.
- 7) Mitsuya, H.; Yarchoan, R.; Broder, S. Molecular Target for AIDS Therapy *Science*, **1990**, 249, 1533-1544.
- 8) Blundell, T.; Pearl, L. A Second front against AIDS., *Nature*, **1989**, 596-599.
- 9) Resat, H.; Mezei, M. Grand Canonical Monte Carlo Simulation of Water Positions in Crystal Hydrates., *J. Am. Chem. Soc.* **1994**, 116, 7451-7452.
- 10) York, D. M.; Darden, T. A.; Pederson, L. G.; Anderson, M. W. Molecular Dynamics Simulation of HIV-1 Protease in a Crystalline Environment and in Solution., *Biochemistry*, **1993**, 32, 1443-1453.
- 11) Marrone, T. J.; Resat, H.; Hodge, C. N.; Chang, C. -H.; McCammon, J. A. Solvation studies of DMP323 and A76928 bound to HIV protease: Analysis of water sites using grand canonical Monte Carlo simulation *Prot. Sci.* **1998**, 7(3), 573-579.
- 12) Gerstein, M.; Levitt, M. Molecular of Life., *Scientific American*, **1998**, 101-105.
- 13) Otting, G.; Liepinsh, E.; Wuthrich, K. Protein Hydration in Aqueous Solution., *Science*, **1991**, 254, 974-980.

- 14) Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshino, T.; Tsuda, M. Molecular Dynamics Study of HIV-1 Protease-Substrate Complex: Roles of the Water Molecules at the Loop Structures of the Active Site., *J. Am. Chem. Soc.* **2000**, 122, 5613-5622.
- 15) Sussman, F.; Villaverde, M. C.; Davis, A. Solvation Effects are responsible for the reduced inhibitor affinity of some HIV-1 PR mutants. *Prot. Sci.*, **1997**, 6, 1024-1030.
- 16) Wang, Y. -X.; Freedberg, D. I.; Grzesiek, S.; Torchia, D. A.; Wingfield, P. T.; Kaufman, J. D.; Stahl, S. J.; Chang, C-H; Hodge, C. N. Mapping Hydration Water Molecules in the HIV-1 Protease/DMP323 Complex in Solution by NMR Spectroscopy., *Biochemistry*.**1996**, 35, 12694-12704.
- 17) Grzesiek, S.; Bax, A.; Nicolson, L. K.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hodge, C. N.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C. -H. NMR Evidence for the Displacement of a Conserved Interior Water Molecule in HIV Protease by a Non-Peptide Cyclic Urea-Based Inhibitor., *J. Am. Chem. Soc.* **1994**, 116, 1581-1582.
- 18) Schmidtke, G.; Holzhtter, H-G; Bogoy, M; Kairies, N.; Groll, M; Giuli, R. d.; Emch, S.; Groettrup, M. How an Inhibitor of the HIV-I protease Modulates Proteasome Activity., *J. Biol. Chem.* **1999**, 274 (50), 35734-35740.
- 19) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site., *J. Am. Chem. Soc.* **2001**, 123, 8730-8737.
- 20) Okimoto, N.; Hata, M.; Hoshino, T.; Tsuda, M. Protein hydrolysis mechanism of HIV-1 protease: Investigation by the ab Initio MO calculations., *RIKEN Rev.* 29, **2000**, 100-102.
- 21) David, L.; Luo, R.; Head, M. S.; Gilson, M. K. Computational Study of KNI-272, a Potent Inhibitor of HIV-1 Protease: On the Mechanism of Preorganization., *J. Phys. Chem. B* **1999**, 103, 1031-1044.

- 22) Venturini, A.; Lopez-Ortiz, F.; Alvarez, J. M.; Gonzalez, J. Theoretical Proposal of a Catalytic Mechanism for the HIV-1 Protease Involving an Enzyme-Bound Tetrahedral Intermediate. *J. Am. Chem. Soc.* **1998**, 120, 1110-1111.
- 23) Weber, I. T.; Cavanaugh, D. S.; Harrison, R. W. Models of HIV-1 protease with peptides representing its natural substrated. *J. Mol. Struc.* **1998**, 423, 1-12.
- 24) Liu, H.; Plathe, M. F.; Gunsteren, W. F. v. A Combined Quantum/Classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease., *J. Mol. Biol.*, **1996**, 261, 454-469.
- 25) Lee, H.; Darden, T. A.; Pederson, L. G. An ab Initio Quantum Mechanical Model for the catalytic Mechanism of the HIV-1 protease., *J. Am. Chem. Soc.* **1996**, 118, 3946-3950.
- 26) Mavri, J. Irreversible Inhibition of the HIV-1 Protease: A Theoretical Study. *Inter. Quan. Chem.* **1998**, 69, 753-759.
- 27) Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. Hydrolysis Mechanism of the Phenylalanine-Proline Peptide Bond Specific to the HIV-1 Protease: Investigation by the ab Initio Molecular Orbital Method., *J. Am. Chem. Soc.* **1999**, 121, 7349-7354.
- 28) Benedict, W. S.; Gailar, N.; Plyer, E. K. Rotation-Vibration Spectra of Deuterated Water Vapor. *The J. Chem. Phys.* **1956**, 24, 6, 1139-1165.
- 29) Beveridge, A. J.; Heywood, G. C. A Quantum Mechanical Study of the Active Site of Aspartic Proteinases., *Biochemistry*, **1993**, 32, 3325-3333.
- 30) Skalka, A. M. Retroviral Proteases: First Glimpses at the Anatomy of a Processing Machine. *Cell* **1989**, 56, 911-914.
- 31) Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C-H.; Weber, P. C. NMR and X-ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and a Non-Peptide Cyclic Urea-Based Inhibitor., *J. Am. Chem. Soc.* **1994**, 116, 10791-10792.

- 32) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insight into the mechanisms of drug resistance., *Prot. Sci.* **2000**, 9, 1898-1904.
- 33) Navia, A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C-T.; Heimbach, C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1., *Nature* **1989**, 337, 615-620.
- 34) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. Conserved Folding in Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease., *Science* **1989**, 245, 616-621.
- 35) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, 247, 1149-1152.
- 36) Jaskolski, M.; Tommasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5-Å Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-Based Inhibitors., *Biochemistry* **1991**, 30, 1600-1609.
- 37) Miller, M.; Jaskolski, M.; Rao, J. K. M.; Leis, J.; Wlodawer, A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **1989**, 337, 576-579.
- 38) Rapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature* **1989**, 342, 299-302.
- 39) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. -H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science (Washington, D. C.)* **1994**, 263, 380-384.
- 40) DeCamp, D.; Ogden, R.; Kuntz, I.; Craik, C. S. Site-Directed Drug Design *Pro. Eng.* **1996**, 467-505.

- 41) Wang, W.; Kollman, P. A. Computational Study of protein specificity: The molecular basis of HIV-1 protease drug resistance. *PNAS* **2001**, *98*, 14937-14942.
- 42) Freedberg, D. I.; Wang, Y-X.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Kiso, Y.; Torchia, D. A. Flexibility and Function in HIV Protease: Dynamics of the HIV-1 Protease Bound to the Asymmetric Inhibitor Kynostatin 272 (KNI-272). *J. Am. Chem. Soc.* **1998**, *120*, 7916-7923.
- 43) Wang, Y-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. Solution NMR Evidence That the HIV-1 Protease Catalytic Aspartyl Groups Have Different Ionization States in the Complex Formed with the Asymmetric Drug KNI-272. *Biochemistry*, **1996**, *35*, 9945-9950.
- 44) Velazquez-Campoy, A.; Luque, I.; Todd, M. J.; Milutinovich, M.; Kiso, Y.; Freire, E. Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. *Pro. Sci.* **2000**, *9*, 1801-1809.
- 45) Bader, R. F. W. (1994). *Atoms in Molecules: A Quantum Theory*.
- 46) Hehre, W. J.; Radom, L.; Schleyer, P. v.R.; Pople, J. A. (1986). *Ab initio Molecular Orbital Theory*.
- 47) Young, C. D. (2001). *Computational Chemistry*. John Wiley & Sons.
- 48) McQuarrie, D. A. (1984) *Quantum Chemistry*. University Science Books.
- 49) Wlodawer, A.; Erickson, J. W. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **1993**, *62*, 543-585.
- 50) Gait, M. J.; Karn, J. Progress in anti-HIV structure-based drug design. *Trends Biotech.* **1995**, *13*, 430-438.
- 51) Lee, C. S.; Choy, N.; Park, C.; Choi, H.; Son, Y. C.; Kim, S.; Ok, J. H.; Yoon, H.; Kim, S. C. Design, Synthesis, and Characterization of dipeptide isostere containing cis-epoxide for the irreversible inactivation of HIV protease. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 589-594.
- 52) Yu, Z.; Caldera, P.; Mcphee, F.; De Voss, J. J.; Jones, P. R.; Burlingame, A. L.; Kuntz, I. D.; Craik, C. S.; Ortiz de Montellano, P. R. Irreversible inhibition of the HIV-1 protease: Targetting alkylating agents to the catalytic aspartate groups. *J. Am. Chem. Soc.* *118*, 5846-5856.

- 53) Deeks, S. G.; Smith, M.; Holodnity, M.; Kahn, J. O. HIV-1 protease inhibitors: A review for clinicians. *J. Am. Med. Assoc.* **1997**, 277 (2), 145-153.
- 54) Otto, M. J.; Garber, S.; Winslow, D. L.; Reid, C. D.; Aldrich, P.; Jadhav, P. K.; Patterson, C. E.; Hodge, C. N.; Cheng, Y. –S. E. In vitro isolation and identification of human immunodeficiency virus (HIV) variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proc. Natl. Acad. Sci. USA* **90**, 7543-7547.
- 55) Ho, D. D.; Toyoshima, T.; Mo, H.; Kempf, D. J.; Norbeck, D.; Chen, C. –M.; Wideberg, N. E.; Burt, S. K.; Erickson, J. W.; Singh, M. K. Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C₂-symmetric protease inhibitor., *J. Virol.* **68**(3), 2016-2020.
- 56) Gulnik, S. V.; Suvorov, L. I.; Lui, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **1995**, 34(29), 9282-9287.
- 57) Ridky, T. W.; Kikonyogo, A.; Leis, J.; Gulnik, S.; Copeland, T.; Erickson, J.; Wlodawer, A.; Kurinov, I.; Harrison, R. W.; Weber, I. T. Drug-resistant HIV-1 proteases identify enzyme residues important for substrate selection and catalytic rate., *Biochemistry* **1998**, 37(39), 13835-13845.



APPENDICES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

APPENDIX I

The Semiempirical Results

Table I.1 The WAT607-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model II: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E SemiEmpirical (kcal/mol)	
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1	PM3
3.0,3.5	27,23',25'	48	-0.24	+1.60	-1.21	-0.95	+0.26
4.0	27,23',25',82'	66	-0.59	+1.31	-1.71	-1.06	+0.14
4.5	27,23',25',82',84'	87	-0.99	+1.06	-2.24	-1.00	+0.07
5.0,5.5	27,28,23',25',82',84'	97	-1.06	-5.64	-8.92	-3.65	-1.56
6.0	25,27,28,23',25',82',84'	112	-1.11	-5.30	-8.62	-3.44	-1.33
6.5	25,26,27,28,8',23',25',83',82',84'	-	-	-	-	-0.57	+0.50
7.0	25,26,27,28,29,50,8',23',24',25',28',81',83',82',84'	-	-	-	-	-1.40	-0.38

Table I.2 The WAT301-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model II: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E SemiEmpirical (kcal/mol)
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1
3.0	50'	24	+0.50	-2.04	-4.14	-2.56
3.5	50,50'	45	+0.31	-1.62	-2.61	-2.20
4.0,5.0	49,50,49',50'	59	-0.18	-5.42	-9.36	-5.57
5.5	49,50,51,48',49',50'	73	-0.29	-7.18	-10.99	-6.68
6.0	48,49,50,51,48',49',50',51'	87	-0.35	-7.75	-11.49	-6.90
6.5	48,49,50,51,52,48',49',50',51',52'	101	-0.40	-7.89	-11.53	-6.91

Table I.3 The WAT406-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model II: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E _{SemiEmpirical} (kcal/mol)
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1
3.0,3.5	26',87'	46	+0.89	+1.76	-2.02	+0.54
4.0	8,9,23,26',27',29',87',	130	-0.67	+4.65	-2.69	+1.17
4.5,5.0	8,9,23,26',27',28',29',87'	138	-0.99	+5.19	-2.88	+1.54
5.5	7,8,9,23,26',27',28',29',87'	155	-1.05	+5.08	-2.69	+1.47
6.0	7,8,9,23,24,25',26',27',28',29',87'	187	-1.13	+4.30	-3.36	+1.07

Table I.4 The WAT566-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model II: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E _{SemiEmpirical} (kcal/mol)
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1
3.0	27',29'	27	+2.82	+1.52	-1.51	-0.69
3.5,4.0	8,27',29'	54	+2.21	-8.80	-12.53	-3.76
4.5	8,23,27',28',29'	83	+1.65	-10.19	-13.87	-4.73
5.0	8,23,27',28',29',87'	110	+1.57	-11.28	-14.35	-5.39
5.5,6.0	8,23,26',27',28',29',87'	124	+1.51	-11.64	-14.52	-5.52

Table I.5. The WAT426-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model II: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E _{SemiEmpirical} / (kcal/mol)
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1
3.0	26,29	34	-0.12	-1.52	-3.73	-1.42
3.5	26,27,29	41	-0.64	-0.22	-2.91	-0.27
4.0	26,27,28,29,8',9'	90	-1.25	-3.09	-7.05	-2.59
4.5	26,27,28,29,8',9',23'	111	-1.65	-3.27	-7.47	-3.32
5.0,5.5	26,27,28,29,87,8',9',23'	138	-1.32	-1.31	-8.11	-1.97
6.0	25,26,27,28,29,87,7',8',9',23'	168	-2.11	-2.62	-8.77	-2.72

Table I.6. The WAT608-enzyme interaction energies according to Model II, investigated using semiempirical method.

Model 2: H ₂ O(Res) _{n(r)}			Δ E (kcal/mol)			Δ E _{SemiEmpirical} (kcal/mol)
n(r)	Residues	No. of atoms	MM	HF	DFT	AM1
3.0,3.5	27,29	27	-0.83	-4.90	-5.98	-3.51
4.0,4.5	27,28,29	35	-0.58	-8.19	-9.55	-4.79
5.0	27,28,29,30	48	-0.71	-8.50	-10.05	-5.01
5.5,6.0	27,28,29,30,48,8'	84	-0.80	-9.37	-10.78	-5.44

Table I.7 The optimal model of single water-enzyme interaction energies according to Model II by semiempirical method, summarized.

Model II: H ₂ O(Res) _{n(r)}			ΔE (kcal/mol)		ΔE SemiEmpirical (kcal/mol)
n(r)	Residues	No. of atoms	HF	DFT	AM1
301	48,49,50,51,48',49',50',51'	87	-7.75	-11.49	-6.90
607	27,28,23',25',82',84'	97	-4.83	-8.92	-3.65
406	8,9,23,26',27',28',29',87'	138	+5.19	-2.88	+1.54
426	26,27,28,29,87,8',9',23'	138	-1.31	-8.11	-1.97
566	8,23,27',28',29',87'	110	-11.28	-14.35	-5.39
608	27,28,29,30	48	-8.50	-10.05	-5.01

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

APPENDIX II

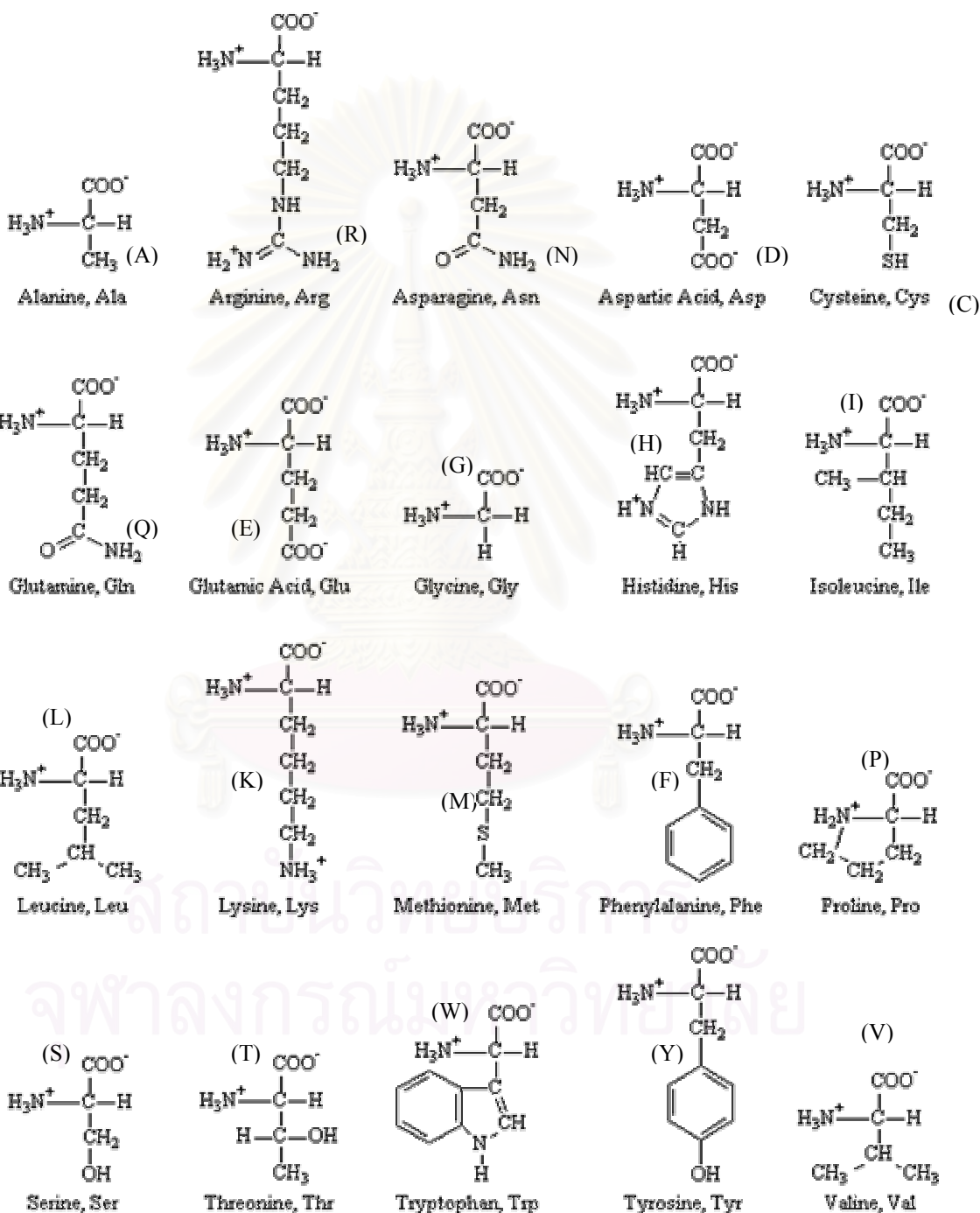
Amino Acid Sequence of HIV-1 Protease from 1HPX.pdb

5								10				
PRO	GLN	ILE	THR	LEU	TRP	GLN	ARG	PRO	LEU	VAL	THR	ILE
LYS	ILE	GLY	GLY	GLN	LEU	LYS	GLU	ALA	LEU	LEU	ASP	THR
GLY	ALA	ASP	ASP	THR	VAL	LEU	GLU	GLU	MET	SER	LEU	PRO
GLY	ARG	TRP	LYS	PRO	LYS	MET	ILE	GLY	GLY	ILE	GLY	GLY
PHE	ILE	LYS	VAL	ARG	GLN	TYR	ASP	GLN	ILE	LEU	ILE	GLU
ILE	CYS	GLY	HIS	LYS	ALA	ILE	GLY	THR	VAL	LEU	VAL	GLY
PRO	THR	PRO	VAL	ASN	ILE	ILE	GLY	ARG	ASN	LEU	LEU	THR
GLN	ILE	GLY	CYS	THR	LEU	ASN	PHE					
PRO	GLN	ILE	THR	LEU	TRP	GLN	ARG	PRO	LEU	VAL	THR	ILE
LYS	ILE	GLY	GLY	GLN	LEU	LYS	GLU	ALA	LEU	LEU	ASP	THR
GLY	ALA	ASP	ASP	THR	VAL	LEU	GLU	GLU	MET	SER	LEU	PRO
GLY	ARG	TRP	LYS	PRO	LYS	MET	ILE	GLY	GLY	ILE	GLY	GLY
PHE	ILE	LYS	VAL	ARG	GLN	TYR	ASP	GLN	ILE	LEU	ILE	GLU
ILE	CYS	GLY	HIS	LYS	ALA	ILE	GLY	THR	VAL	LEU	VAL	GLY
PRO	THR	PRO	VAL	ASN	ILE	ILE	GLY	ARG	ASN	LEU	LEU	THR
GLN	ILE	GLY	CYS	THR	LEU	ASN	PHE					

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

APPENDIX III

Brief information of amino acids



Modified from

<http://www.stark.kent.edu/~cearley/PChem/AAcids/classify.htm>

APPENDIX IV



MANUSCRIPT

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Quantum Chemical Calculations on the Structure and Bonding of Water Molecules in the HIV-1 Protease (PR) Enzyme

Chittima Laohpongspaisan^a, Atchara Wijitkosoom^a, Sulapong Pinichklang^b, Vudhichai Parasuk^a, Supot Hannongbua^{a,*}

^aDepartment of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

^bDepartment of Food Science, Faculty of Science, The University of Thai Chamber of Commerce, Bangkok 10400, Thailand

Abstract: Role of water molecule in the HIV-1 Protease (PR), has been investigated, based on quantum mechanical calculations. The obtained data in terms of binding energies and water binding sites were taken into consideration in the modeling and designing of the potent inhibitors in the HIV-1 protease pocket. The interaction energies between amino acid cluster and water molecules taken from X-ray structure of the HIV-1 PR from Protein Data Bank (PDB) were evaluated and compared among various methods. For water molecule, coordinate of the oxygen atom is taken from the PDB while its X-ray geometry with the O-H bond length of 0.975 Å and H-O-H bond angle of 104.5° is applied and kept constant throughout. Here, position of the oxygen atom of water is fixed but its orientation around the oxygen is optimized. The size of amino acid cluster was increased as a function of spherical radius centered at oxygen water atom until consistency of the interaction energy was reached. The results show that strategy to increase the size of the model takes one important part in the study of water in protein. The six water molecules in the crystallographic structure of the HIV-1 PR, namely W301, WAT406, WAT426, WAT566, WAT607 and WAT608 have been taken into consideration studied. The obtained binding energies for the six water molecules are -11.49, -2.88, -8.11, -14.35, -8.92 and -10.05 Kcal/mol, respectively. The results indicate also that catalytic water, WAT607, plays role not only in hydrolysis mechanism but also in stabilizing the enzyme structure. The other water

molecules were also found to have a substantial role in maintaining the specific conformation of enzyme structure.

1. Introduction

The HIV-1 protease has been the most important target (1,2) for drug development against HIV-1 infection due to its key role in viral maturation (3-5). The HIV-1 protease is a symmetrical homo-dimer (Figure 1) composed of identical subunits of 99 residues each (2). The binding pocket is formed by the loop structure containing the active site triads and flap region (flaps) which are presumably related with the entry and affinity of the substrate to enzyme (8-11). The crystallographic structures of the free enzyme as well as its complexes with several inhibitors have been determined at high resolution (14-22). Several HIV-1 protease inhibitors are already in clinical therapies. It is generally accepted that water molecules play an important role in the binding affinity or specificity of an inhibitor (6-13). A prominent example is a lytic water molecule which is held by the two catalytic Asp25A/25B residues through the hydrogen bonding. Its role is to induce the protein hydrolysis. Another important water molecule is in flap region, called WAT301 and observed to participate in hydrogen-bonding interactions between the flaps and the inhibitors. This water plays role in controlling movement of the flaps (8,9,11). It is also found, in the crystal structure of almost every HIV-1 protease-inhibitor complexes, to bridge the backbone amide protons of Ile50A/50B to inhibitor (presumably substrate) carbonyls through hydrogen bonding (11,12). Among the two water molecules mentioned above, the presence of other structurally important water molecules inside and outside the active pockets of the protease were also reported (8,9,11,13). These examples clearly demonstrate role of water in the catalytic process (23-33), and hence a structure-based drug design strategy (34-36,38-41).

Several experimental techniques such as NMR spectroscopy (14-17) and X-ray crystallography (18-22) have been used to map out the hydration water molecules both in free form and its complexes with inhibitors. On the other side, the water positions were investigated theoretically within HIV-1 protease complexes using computer simulations

(9,13,37). Most theoretical data regarding water orientation and its interaction in the free protease and its complexes so far have relies on the use of molecular mechanics (MM) calculations (9). As a matter of fact, this method usually leads to the loss of some specific details. For example, some doubts arise when the MM method was used to represent the interaction of the system in which hydrogen bonding could be very important and better represented by quantum chemical calculations. Due to the above reasons, quantum mechanics calculation and combined quantum mechanic and molecular mechanic (QM/MM) has been applied to investigate precise position and orientation of water molecules in the binding pocket of the protease.

2. Calculation Details

2.1 Preparation of initial structure.

Since the complete X-ray crystal structure of free HIV-1 protease is not available. Therefore, that of the complex (1HPX in Protein Data Bank code), refined at a resolution of 2.0 Å was used as the starting structure. Here, the inhibitor was eliminated. Then, hydrogen atoms have been added and their positions and orientations have been optimized using AMBER 5.0 program. The obtained structure was used and kept constant throughout. To clarify reliability of this structure, geometry optimization has been applied using the AMBER program. It was found that its backbone geometry is insignificantly changes, RMSD (root mean square displacement) = 0.002 Å, in comparison to the X-ray structure of the complex. For water molecule, experimental geometry with the O-H distance of 0.975 Å and H-O-H angle of 104.5° taken from the literature (32), has been applied and kept constant. Coordinate of the oxygen atom taken from the PDB structure of the HIV-1 protease complex was kept fixed. Here, its orientation around oxygen atom was optimized.

2.2 Model representations

As it is known that enzyme molecule is too big to perform quantum chemical calculations, even with small basis set and the high performance computers. Therefore, the protease enzyme has to be represented by the cluster of optimal size, by varying number of amino acid residues around the investigated water molecules. Here, two types of expansion were proposed.

Model I: linear expansion

Interaction between water molecule and amino acid residues (Res) in the $H_2O(Res)_m$ clusters, $\Delta E_{(mod1)}$ where, $m = 1, 2, 3, \dots$, have been calculated according to equation [2.1];

$$\Delta E_{(mod1)} = E(cst) - E(H_2O) - E(Res)_m \quad [2.1]$$

where terms on the right hand side are total energies obtained from quantum chemical calculations of the cluster, water and m amino acid, respectively. The first amino acid ($m = 1$) in the cluster is that lying closest to oxygen atom of the investigated water molecule. For $m = 2, 3, 4, \dots$, these refer to the next residue which binds to the previous one in the amino acid sequence. Extension has been made until either consistency of the interaction energy (from eq. [2.1]) or limitation in computation was reached.

Model II: spherical expansion

Water-residues interaction, $\Delta E_{(mod2)}$ in the $H_2O(Res)_{n(r)}$ cluster has been calculated using the same equation, equation [2.1]. In this model, cluster size was increased as a function of spherical radius, r , centered at oxygen atom of the investigated water, i. e., n is defined as number of amino acid residues lying within a sphere of radius r , where $r = 3, 3.5, 4, 4.5, \dots \text{ \AA}$, from oxygen atom of the central water. If each atom of any amino acid is detected within the spherical radius, the whole molecule of that residue will be included into the model. Similar criteria as that of model I has been also used to characterize the

investigated model, i. e., cluster size was expanded, in terms of r_i , until consistency of the interaction energy was reached. At this state, the obtained results, is considered as the optimal cluster size or optimal spherical radius, r_{opt} , for each water molecule.

2.3 QM and QM/MM calculations

Calculation techniques used in this study can be categorized into two types, the full QM and the QM/MM. All calculations performed in this work were done using the GAUSSIAN 98 program. Quantum chemical calculations have been performed based on the Hartree-Fock (HF) and the Density Functional Theory (DFT) with B3LYP function. The basis sets 6-31G* and 6-31G** were applied for amino acid cluster and water, respectively.

For the QM/MM calculations the enzyme molecule was partitioned into two parts, both mentioned QM methods, HF and DFT, have been applied for the model system which consists of water molecule and amino acid residues lying in the cluster (Model I and Model II). For the rest of molecules, real system, the MM calculations have been performed.

3. Results and Discussion

3.1 Optimal calculated model

To seek for the optimal expansion method, WAT607 was used as the tested model. The calculations have been performed for the $H_2O(Res)_m$ and $H_2O(Res)_{n(r)}$ clusters, varying m and r as already mentioned. Interaction energies for the two model types were shown in Tables 1 and 2.

In the $H_2O(Res)_2$ cluster (model I, $m = 2$) interaction between water molecule (WAT607) and two residues, Asp25 and Asp25', has been calculated. The HF energy of -3.45 Kcal/mol is slightly higher than that of -4.35 Kcal/mol obtained from the DFT calculations. This leads, consequently, to the higher DFT-QM/MM energy than the HF-QM/MM one. As expected, the MM water-residues interaction for the $H_2O(Res)_2$ cluster is

much higher than the QM and QM/MM energies. Increase of the cluster size, according to the linear model, the interaction decrease slowly.

For model II, $H_2O(Res)_{n(r)}$, the cluster with $r = 3.0 - 3.5 \text{ \AA}$ accumulates 3 amino acid ($n = 3$), residues 27, 23' and 25'. The corresponding MM, HF and DFT interaction energies are -0.24, 1.60 and -1.21 Kcal/mol, respectively. Unexpected result is that of the HF calculation in which the interaction is slightly positive. This situation remains unchanged up to $r = 4.5 \text{ \AA}$. Similar to that of model I (linear expansion), the MM energies for model II are almost the same when the cluster size increases. The most appropriate trend is the decrease of interaction energy according to the DFT method in which the energy is slightly decrease for $r \leq 4.5 \text{ \AA}$ and remains constant at about -9 Kcal/mol for $r \geq 5.0 \text{ \AA}$. This value is significantly lower than the optimal interaction yielded from model I (DFT energy for the cluster $H_2O(Res)_m$ where $m = 8 - 10$ in Table 1). A clear conclusion is that spherical expansion of the cluster size, model II, is more reliable than the linear one. Therefore, model II was, then, used for further investigations. Additional weak point of model I is due to the fact, for examples, that residues 23', 82' and 84' which locate at the distances closer than 3.0 \AA , 4.0 \AA and 4.5 \AA to the oxygen atom of WAT607, respectively, were not taken into consideration.

Taken into account all the energy data shown in Tables 1 and 2, some other conclusions can be also made. (i) MM method, and hence QM/MM, fails to describe the water-enzyme interactions, at least for the investigated system. (ii) Among the HF and the DFT/B3LYP, the second one is superior. (iii) For WAT607, the optimal $H_2O(Res)_{n(r)}$ cluster size for the DFT calculations is $r = 5.0 \text{ \AA}$, consisting of 6 amino acid residues, 27, 28, 23', 25', 82' and 84' (Table 2).

3.2 Optimal cluster size and its interaction in the single water cluster, $H_2O(Res)_{n(r)}$

With the conclusion drawn in section 3.1, the appropriated expansion strategy, model II, has been applied for each water. Six single-water molecules were used as a center for the spherical expansion in the $H_2O(Res)_{n(r)}$ cluster. By varying r centered on

oxygen atom of water, interaction between water and n amino acid residues, $(\text{Res})_n$, has been calculated. Changes of HF, DFT, QM(DFT)/MM and QM(HF)/MM interaction energies as a function of r were plotted for each water in Figure 3.

Failure of the HF method has been, again, exhibited for WAT406 and WAT426 in which the interaction energies are increased when r increases (Figure 3). Although the QM(DFT)/MM energy plots for all water molecules are observed to be the optimal ones, but they are excluded from consideration because the MM energies are almost constant for any cluster sizes (see also section 3.1). Therefore, although the HF, DFT, QM(HF)/MM and QM(DFT)/MM energies, the corresponding cluster sizes (see model II in section 2.2 for more details) as well as their characteristics for each water molecule were extracted from the optimization and summarized in Table 3, but only those of the DFT method were used for future investigations.

It can be seen from Table 3 as well as the plots in Figure 3 that the optimal radius for any water molecule in the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster (where the interaction energy is consistent when r increases) is between 5.0 Å and 5.5 Å. The clusters contain 4-8 amino acid residues. It is interesting to note that the DFT interaction energies for WAT566, WAT301 and WAT301 are superior, in comparison to those of the other water molecules. These confirm what reported experimentally that WAT301 plays role in controlling movement of the flaps (4) by participating in the hydrogen bonding between the flaps and the inhibitor. The observed stabilization energies of -14.35 Kcal/mol for WAT566 and of -10.65 Kcal/mol for WAT608 also confirm their role in stabilizing the overall structure of the HIV-1 protease (3,5).

For the other two structural water molecules, WAT426 and WAT406, the calculated energy of -8.11 Kcal/mol for WAT406 supports its role in holding the enzyme structure. On the other hand, the energy of -2.88 kcal/mol for WAT426 indicates very clear that it doesn't form hydrogen bond with the amino acid residues. Therefore, its role in stabilizing enzyme structure is supposed to differ from the other structural water molecules. Instead of holding the enzyme by forming the hydrogen bond, it suppose to repel the residues in order to prevent them from getting collapse.

For WAT301 which is known to participate in the catalytic process (9,10), its interaction with 8 amino acids is -8.92 Kcal/mol.

3.3 Optimal cluster size and its interaction in the double- and triple-water cluster

To clarify reliability of the single-water cluster model, $H_2O(Res)_{n(r)}$, especially for those lying within the r_{opt} of the other water molecule, calculations have been performed for the double-, $(H_2O)_2(Res)_{n(r)}$, and triple-, $(H_2O)_3(Res)_{n(r)}$, water cluster models. Total interaction energy between 2 (or 3) water molecules with n amino acids lying within $r = r_{opt}$ of the 2 (or 3) water molecules have been calculated according to equation [2.2]

$$\Delta E = E(cst) - xE(H_2O) - E(Res)_n \quad [2.2]$$

where $x = 2$ and 3 for the double- and triple-water cluster models, respectively. The calculated energies were summarized in Table 4.

The results show that sum of the DFT interaction energies obtained from the two single-water cluster models of -17.23 Kcal/mol for WAT406+WAT566 and of -18.16 Kcal/mol for WAT426+WAT608 are almost the same as those of the double-water cluster models shown in Table 4. This indicates that individual cluster is sufficient to represent water-enzyme interaction.

The DFT interaction energies of the three closely related water molecules with amino acid in the triple-water cluster model of -22.38 Kcal/mol (Table 4) are higher than that of -27.08 Kcal/mol yielded from sum of the three individual water. The difference of 4.06 kcal/mol can be assigned to the repulsion among the amino acid residues in the cluster. However, this value is negligible for such big cluster $(H_2O)_3(Res)_{n(r)}$ containing 3 water molecules and 12 amino acid residues.

Acknowledgements.

Computing facilities provided by the Austria-Thai Center for Chemical Education and Research at Chulalongkorn University, Bangkok, Thailand are gratefully acknowledged.

References

1. Tomasselli, A. G.; Heinrikson, R. L. *Biochim. et Biophys. Acta* **2000**, 1477, 189-214.
2. Erickson, J. W.; Eissenstat, M. A., *Protease of Infected Agent*, Academic Press. **1999**.
3. Turner, B. G.; Summers, M. F. *J. Mol. Biol.* **1999**, 285, 1-32.
4. Todd, M. J.; Semo, N.; Freire, E. J. *Mol. Biol.*, **1998**, 283, 475-488.
5. Mitsuya, H.; Yarchoan, R.; Broder, S. *Science*, **1990**, 249, 1533-1544.
6. Gerstein, M.; Levitt, M. *Scientific American*, **1998**, 101-105.
7. Otting, G.; Liepinsh, E.; Wuthrich, K. *Science*, **1991**, 254, 974-980.
8. Okimoto, N.; Tsukui, T.; Kitayama, K.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* **2000**, 122, 5613-5622.
9. Marrone, T. J.; Resat, H.; Hodge, C. N.; Chang, C. -H.; McCammon, J. A. *Prot. Sci.* **1998**, 7(3), 573-579.
10. Sussman, F.; Villaverde, M. C.; Davis, A. *Prot. Sci.*, **1997**, 6, 1024-1030.
11. Wang, Y. -X.; Freedberg, D. I.; Grzesiek, S.; Torchia, D. A.; Wingfield, P. T.; Kaufman, J. D.; Stahl, S. J.; Chang, C-H; Hodge, C. N. *Biochemistry*. **1996**, 35, 12694-12704.
12. Grzesiek, S.; Bax, A.; Nicolson, L. K.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hodge, C. N.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C. -H. *J. Am. Chem. Soc.* **1994**, 116, 1581-1582.
13. York, D. M.; Darden, T. A.; Pederson, L. G.; Anderson, M. W. *Biochemistry*, **1993**, 32, 1443-1453.
14. Skalka, A. M. *Cell* **1989**, 56, 911-914.
15. Yamazaki, T.; Nicholson, L. K.; Torchia, D. A.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Ru, Y.; Jadhav, P. K.; Chang, C-H.; Weber, P. C. *J. Am. Chem. Soc.* **1994**, 116, 10791-10792.
16. Miller, M.; Jaskolski, M.; Rao, J. K. M.; Leis, J.; Wlodawer, A. *Nature* **1989**, 337, 576-579.

17. Rapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. *Nature* **1989**, 342, 299-302.
18. Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. *Prot. Sci.* **2000**, 9, 1898-1904.
19. Navia, A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C-T.; Heimbach, C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. *Nature* **1989**, 337, 615-620.
20. Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. *Science* **1989**, 245, 616-621.
21. Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. *Science* **1989**, 247, 1149-1152.
22. Jaskolski, M.; Tommasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Biochemistry* **1991**, 30, 1600-1609.
23. Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, 123, 8730-8737.
24. Okimoto, N.; Hata, M.; Hoshino, T.; Tsuda, M. *RIKEN Rev.* 29, **2000**, 100-102.
25. David, L.; Luo, R.; Head, M. S.; Gilson, M. K. *J. Phys. Chem. B* **1999**, 103, 1031-1044.
26. Venturini, A.; Lopez-Ortiz, F.; Alvarez, J. M.; Gonzalez, J. *J. Am. Chem. Soc.* **1998**, 120, 1110-1111.
27. Weber, I. T.; Cavanaugh, D. S.; Harrison, R. W. *J. Mol. Struct.* **1998**, 423, 1-12.
28. Liu, H.; Plathe, M. F.; Gunsteren, W. F. v. *J. Mol. Biol.*, **1996**, 261, 454-469.
29. Lee, H.; Darden, T. A.; Pederson, L. G. *J. Am. Chem. Soc.* **1996**, 118, 3946-3950.
30. Mavri, J. *Inter. Quan. Chem.* **1998**, 69, 753-759.
31. Okimoto, N.; Tsukui, T.; Hata, M.; Hoshino, T.; Tsuda, M. *J. Am. Chem. Soc.* **1999**, 121, 7349-7354.
32. Benedict, W. S.; Gailar, N.; Plyer, E. K. *The J. Chem. Phys.* **1956**, 24, 6, 1139-1165.
33. Beveridge, A. J.; Heywood, G. C. *Biochemistry*, **1993**, 32, 3325-3333.

34. am, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. -H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. *Science* (Washington, D. C.) **1994**, 263, 380-384.
35. DeCamp, D.; Ogden, R.; Kuntz, I.; Craik, C. S. *Pro. Eng.* **1996**, 467-505.
36. Wang, W.; Kollman, P. A. *PNAS* **2001**, 98, 14937-14942.
37. Resat, H.; Mezei, M. *J. Am. Chem. Soc.* **1994**, 116, 7451-7452.
38. Freedberg, D. I.; Wang, Y-X.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Kiso, Y.; Torchia, D. A. *J. Am. Chem. Soc.* **1998**, 120, 7916-7923.
39. Wang, Y-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. *Biochemistry*, **1996**, 35, 9945-9950.
40. Velazquez-Campoy, A.; Luque, I.; Todd, M. J.; Milutinovich, M.; Kiso, Y.; Freire, E. *Pro. Sci.* **2000**, 9, 1801-1809.
41. Blumdell, T.; Pearl, L. *Nature*, **1989**, 596-599.

TABLE CAPTIONS

Table 1

Water (WAT607)-enzyme interaction energy in the $\text{H}_2\text{O}(\text{Res})_m$ cluster calculated using different methods, where m denotes number of amino acid residues expanded linearly according to model I (see text for more details).

Table 2

Water (WAT607)-enzyme interaction energy in the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster calculated using different methods, where n denotes number of amino acid residues expanded spherically according to model II (see text for more details).

Table 3

Water-enzyme interaction energy in the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster calculated using different methods for each water molecule, where n is number of amino acid residues and r denotes spherical expansion radius according to model II (see text for more details).

Table 4

Water-enzyme interaction energy in the $(\text{H}_2\text{O})_2(\text{Res})_{n(r)}$ and $(\text{H}_2\text{O})_3(\text{Res})_{n(r)}$ clusters calculated using different methods, where n is number of amino acid residues and r denotes spherical expansion radius according to model II (see text for more details).

FIGURE CAPTIONS

Figure 1

HIV-1 protease in free (a) and complex (b) forms taken from literatures (6, 9).

Figure 2

Schematic representation of the six water molecules lying in the amino acid environments in the HIV-1 protease pocket.

Figure 3

Changes of the water-enzyme interaction energy in the $\text{H}_2\text{O}(\text{Res})_{n(r)}$ cluster calculated using different methods for each water molecule, where r denotes spherical expansion radius according to model II (see text for more details).

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

BIOGRAPHY

Chittima Laohpongspaisan

Born December 7th, 1976

Education

- 1979-1983 Preprimary School
Sangmanee, Bangkok, Thailand
- 1984-1987 Primary Level
Mongkolkul Vithaya, Nakornrathchasi, Thailand
- 1988-1990 Secondary School
Sikhiu Sawadpadung Vithaya, Nakornrathchasi, Thailand
- 1991-1993 High School
Sikhiu Sawadpadung Vithaya, Nakornrathchasi, Thailand
- 1994-1997 Bachelor of Science (Chemistry)
Khonkaen University, Khonkaen, Thailand
- 2000-2002 Master of Science (Chemistry)
Chulalongkorn University, Bangkok, Thailand

Experience

- 1998-1999 Lecturer position Department of Chemistry,
Faculty of Science, Khonkaen University