โครงสร้างและซอลเวชันของเอชไอวี-1-อาร์ที/สารยับยั้งและฟูริน/เอชเอ-ซับสเตรต โดยการจำลองพลวัตเชิงโมเลกุล

นางสาวพนิตา เดชา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเคมี ภาควิชาเคมี
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

STRUCTURE AND SOLVATION OF HIV-1-RT/INHIBITORS AND FURIN/HA-SUBSTRATE BY MOLECULAR DYNAMICS SIMULATIONS

Miss Panita Decha

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Chemistry
Department of Chemistry
Faculty of Science
Chulalongkorn University
Academic year 2008

Thesis Title	STRUCTURE AND SOLVATION OF HIV-1-RT/INHIBITORS AND FURIN/HA-SUBSTRATE BY MOLECULAR DYNAMICS SIMULATIONS
By	Miss Panita Decha
Field of Study	Chemistry
Advisor	Professor Supot Hannongbua, Dr. rer. nat.
Co-Advisor	Associate Professor Vudhichai Parasuk, Ph. D.
Co-Advisor	Professor Peter Wolschann, Ph. D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

…….. Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr. rer. nat.)

THESIS COMMITTEE

…….. Chairman
(Associate Professor Sirirat Kokpol, Ph. D.)
…….……. Advisor
(Professor Supot Hannongbua, Dr. rer. nat.)
…….. Co-Advisor
(Associate Professor Vudhichai Parasuk, Ph. D.)
…….. Co-Advisor
(Professor Peter Wolschann, Ph. D.)
…….. Examiner
(Assistant Professor Somsak Pianwanit, Ph. D.)
…….. Examiner
(Assistant Professor Rath Pichyangkura, Ph. D.)
…….. External Examiner
(Professor Kritsana Sagarik, Dr. rer. nat.)

พนิตา เดชา: โครงสร้างและซอลเวชันของเอชไอวี-1-อาร์ที/สารยับยั้งและฟูริน/เอชเอ-ซับสเตรตโดยการจำลองพลวัตเชิงโมเลกุล. (STRUCTURE AND SOLVATION OF HIV-1-RT/INHIBITORS AND FURIN/HA-SUBSTRATE BY MOLECULAR DYNAMICS SIMULATIONS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. ดร. สุพจน์ หารหนองบัว, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร. วุฒิชัย พาราสุข, Prof. Dr. Peter Wolschann, 102 หน้า.

ได้ใช้เทคนิคการจำลองพลวัตเชิงโมเลกุล ศึกษาการเปลี่ยนแปลงทางโครงสร้าง พลวัตและซอลเวชันของโปรตีน 2 ระบบ *(i)* โปรตีนที่เกิดสารประกอบเชิงซ้อนกับสารยับยั้งได้แก่ เอชไอวี-1 รีเวอร์สทรานสคริปเทส (อาร์ที) กับกับสารยับยั้งในกลุ่ม NNRTIs สี่ชนิดประกอบด้วย EFV, EMV, ETV และ NVP และ *(ii)* โปรตีนที่เกิดสารประกอบเชิงซ้อนกับซับสเตรตได้แก่ เอนไซม์ฟูรินกับซับสเตรตฮีแมกกลูตินิน (เอชเอ) ในส่วนแรกของงานวิจัย ผลการคำนวณของสารประกอบเชิงซ้อนเอชไอวี-1อาร์ทีกับสารยับยั้งสี่ชนิด พบว่าค่าพลังงานยึดจับอิสระลดลงตามลำดับดังนี้ EFV ~ ETV > EMV > NVP ซึ่งสอดคล้องเป็นอย่างดีกับผลการทดลองในรูปของค่า IC_{50} สำหรับการคำนวณพลังงานอันตรกิริยา พบว่ากรดอะมิโนที่ให้พลังงานอันตรกิริยาสูงกับสารยับยั้งได้แก่ L100, V106, Y181, Y188, F227, W229, P236, K101 และ K103 ซึ่งสอดคล้องกับผลทางคลินิกที่พบว่ามีการกลายพันธุ์บ่อยที่ตำแหน่งกรดอะมิโนเหล่านี้ อันเป็นผลทำให้เกิดการดื้อยา

สำหรับส่วนที่สองของงานวิจัยใช้เทคนิคการจำลองพลวัตเชิงโมเลกุลในการหาสาเหตุของความรุนแรงในการก่อโรคสูงของไข้หวัดนกชนิด H5N1 เนื่องจากการเพิ่มเข้ามาของลำดับกรดอะมิโน –RRRKK– ที่บริเวณที่ถูกตัด (cleavage site) ของเอชเอ เปรียบเทียบกับสายพันธุ์ที่มีความรุนแรงในการก่อโรคต่ำซึ่งไม่มีการเพิ่มของกรดอะมิโนดังกล่าว โดยให้เอชเอทั้งสองยึดจับกับเอนไซม์ฟูริน จากผลการคำนวณพบว่าบริเวณที่ถูกตัดของสายพันธุ์ที่มีความรุนแรงในการก่อโรคสูงชนิด H5 ยึดจับอย่างแข็งแรงกับบริเวณยึดจับของเอนไซม์ฟูริน จึงมีความเหมาะสมที่จะทำให้เกิดกลไกการตัดพันธะเกิดขึ้น ผลที่ได้สอดคล้องเป็นอย่างดีกับผลการทดลองที่พบว่าเมื่อมีการเพิ่มเข้ามาของ –RRRKK– เอชเอถูกตัดด้วยเอนไซม์ฟูรินได้ดียิ่งขึ้น และสามารถตอบคำถามที่ว่าทำไมเอชเอสายพันธุ์ที่มีความรุนแรงในการก่อโรคสูงชนิด H5 ถูกตัดด้วยเอนไซม์ฟูรินได้ดีกว่าเอชเอสายพันธุ์ที่มีความรุนแรงในการก่อโรคต่ำชนิด H3 และ H5

ภาควิชา เคมี ลายมือชื่อนิสิต .....

สาขาวิชา เคมี ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2551 ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม .....

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม .....

4873829923 : MAJOR CHEMISTRY
KEYWORDS : HEMAGGLUTININ/ FURIN / H5N1 / HIV-1 RT / NNRTIs / MOLECULAR DYNAMICS SIMULATION

PANITA DECHA: STRUCTURE AND SOLVATION OF HIV-1-RT/INHIBITORS AND FURIN/HA-SUBSTRATE BY MOLECULAR DYNAMICS SIMULATIONS. THESIS ADVISOR: PROF. SUPOT HANNONGBUA, Dr. rer. nat., THESIS CO-ADVISOR: ASSOC. PROF. VUDHICHAI PARASUK, Ph. D., PROF. PETER WOLSCHANN, Ph. D., 102 pp.

The molecular dynamics (MD) simulations were applied in order to study changes of the structure, dynamic and solvation of protein via complexation with inhibitors and substrates, for the two systems: *(i)* HIV-1 reverse transcriptase (RT) complexed with the four non-nucleoside reverse transcriptase inhibitors (NNRTIs), efavirenz (EFV), emivirine (EMV), etravirine (ETV) and nevirapine (NVP) and *(ii)* Furin complexed with substrates, hemagglutinin subtypes H5 and H3 of the influenza A virus. In the first part, the MD simulations of the HIV-1 RT complexed with one of each of the four NNRTIs provide good evidence which deal with questions related to the mutation and free energy data. The binding free energy was found to decrease in the following order: EFV ~ ETV > EMV > NVP which is in good agreement with the experimentally derived IC_{50} values. On the basis of the calculation of the interaction energies, the L100, V106, Y181, Y188, F227, W229, P236, K101 and K103 amino acid residues show the largest contributions. Moreover, these residues are considered to be the most frequently detected mutated amino acids during treatment by various NNRTIs and, therefore, those most likely to find in the population for resistance. For the second part, origin of high pathogenicity of an emerging avian influenza H5N1 due to the –RRRKK– insertion at the cleavage loop of the hemagglutinin H5, was studied using the MD technique, in comparison with those of the non-inserted H5 and H3 bound to furin active site. The cleavage loop of the highly pathogenic H5 was found to bind strongly to the furin cavity, serving as a conformation suitable for the proteolytic reaction. Experimentally, the –RRRKK– insertion was also found to increase in cleavage of hemagglutinin by furin. The simulated data provide a clear answer to the question of why inserted H5 is better cleaved by furin than the other subtypes, explaining the high pathogenicity of avian influenza H5N1.

Department	Chemistry	Student's signature	
Field of study	Chemistry	Advisor's signature	
Academic year	2008	Co-advisor's signature	
		Co-advisor's signature	

ACKNOWLEDGEMENTS

I would like to begin by expression my deep gratitude to my advisor, Professor Dr. Supot Hannongbua, for his valuable guidance, excellent support and kind encouragement throughout the course of my graduate study. I have learned from him not only in aspect of scientific research but also many great things: positive thinking, good performance and kind attitude. I am deeply appreciated and special thanks to my co-advisor, Professor Dr. Peter wolschann, who kept an eye on my progressive work and always provides me for valuable guidance, considerably helpful comments, and especially for his kind hospitality during my visits in Vienna. I also deeply thanks to Associate Professor Dr. Vudhichai Parasuk for his helpful suggestions, and considerably helpful comments, discussion on various aspects to my research as well as Assistant Professor Dr. Pornthep Sompornpisut.

Furthermore, I would like to thank my committees, Associate Professor Dr. Sirirat Kokphon, Professor Dr. Kritsana Sagarik, Assistant Professor Dr. Somsak Pianwanit and Assistant Professor Dr. Rath Pichyangkura, for the time spent reading my thesis and contributing to my examination.

A great thanks goes to my Bio-Science research team at Computational Chemistry Unit Cell (CCUC) in particular Dr. Ornjira Aruksakunwong, Dr. Thanyada Rungrotmongkol, Dr. Nadtanet Nunthaboot, Dr. Pathumwadee Intharathep, Dr. Chittima Laohpongspaisan, Miss Maturos Malaisree, Mister Nopporn Kaiyawet and Miss Thanyarat Udommaneethanakit as well as CCUC members those I did not mention by name for their valuable assistance, continuing supports in various ways, and advising me at each state of the Ph. D. program.

Full financial support by the Commission on Higher Education (Sandwich Ph. D. program) and supporting from the National Center for Genetic Engineering and Biotechnology (BIOTEC) were gratefully acknowledged. The CCUC, Faculty of Science, Chulalongkorn University and Institute of Theoretical Chemistry, University of Vienna are gratefully acknowledged for, software packages and computing time.

Finally, I would like to express my most profound gratitude to my mother, my brothers and my sister for their love, sincere care, constant support and great encouragement throughout the duration of my educations.

CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Page

Page

LIST OF ABBREVIATION

3TC	=	Lamivudine
AIDS	=	Acquired Immune Deficiency Syndrome
Ala (A)	=	Alanine
Asn (N)	=	Asparagine
Arg (R)	=	Arginine
Asp (D)	=	Aspatic acid
AZT	=	Zidovudine
Cys (C)	=	Cysteine
DC	=	Decomposition Energies
DE	=	Desolvation
ddC	=	Zalcitabine
ddI	=	didanosine
DNA	=	Deoxyribonucleic acid
ELEC	=	Electrostatics
EFV	=	Efavirenz
EMV	=	Emivirine
ETV	=	Etravirine
FDA	=	Food and Drug Administration
FFT	=	Fast Fourier Transform
FR	=	Furin
Gln (Q)	=	Glutamine
Glu (E)	=	Glutamic acid
Gly (G)	=	Glycine
HA	=	Hemagglutinin
His (H)	=	Histidine
HIV-1	=	Human Immunodeficiency Virus Type 1
HIV	=	Human Immunodeficiency Virus
HPAI	=	High-Pathogenicity Avian Influenza
HPH5	=	High-Pathogenicity Avian Influenza subtype H5
IN	=	Integrase
Ile (I)	=	Isoleucine

Leu (L)	=	Leucine
LPAI	=	Low-Pathogenicity Avian Influenza
LPH3	=	Low-Pathogenicity Avian Influenza subtype H3
LPH5	=	Low-Pathogenicity Avian Influenza subtype H5
Lys (K)	=	Lysine
MD	=	Molecular Dynamics
Met (M)	=	Methionine
MM	=	Molecular Mechanics
MM-PBSA	=	Molecular Mechanics Poisson-Boltzmann Surface Area
MS	=	Molecular Surface
NA	=	Nuraminidase
NNRTIs	=	Non-Nucleoside Reverse Transcriptase Inhibitors
NP	=	Nucleoprotein
NRTIs	=	Nucleoside Reverse Transcriptase Inhibitors
NVP	=	Nevirapine
PDB	=	Protein Data Bank
Phe (F)	=	Phenylalanine
PME	=	Particle Mesh Ewald
PR	=	Protease
Pro (P)	=	Proline amino acid
PSC	=	Pair-wise Shape Complementarity
RDF	=	Radial Distribution Function
RMSD	=	Root Mean Square Deviation
RNA	=	Ribonucleic acid
RNaseH	=	Ribonuclease H
RNP	=	Ribonucleoprotein
RT	=	Reverse Transcriptase
SASA	=	Solvent Accessible Surface Area
SCRs	=	Structurally Conserved Regions
Ser (S)	=	Serine
Thr (T)	=	Threonine
Trp (W)	=	Tryptophan
Tyr (Y)	=	Tyrosine

UNAIDS	=	The Joint United Nations Programme on HIV/AIDS
Val (V)	=	Valine
VRs	=	Variable regions
WHO	=	World Heath Organization

CHAPTER I

INTRODUCTION

1.1 Research rationale

Understandings of known enzyme-substrate and enzyme-inhibitor interactions are the key to success to design and discover new inhibitors that fit better to the catalytic site of the targets of both wild and mutant types. This issue becomes the main goal of this work in which the following systems were examined:

- Reverse transcriptase (RT) of Human Immunodeficiency Virus type 1 (HIV-1) complexed with non-nucleoside reverse transcriptase inhibitors (NNRTIs)

- Furin complexed with substrates, hemagglutinin subtypes H5 and H3 of the influenza A virus.

Investigations are based on the molecular dynamics (MD) simulations where structural and dynamic properties as well as solvation of enzymes via complexation with inhibitors and substrates were taken into consideration.

Acquired immune deficiency syndrome (AIDS) is a set of symptoms and infections that lead to the reduction of the effectiveness of immune system. AIDS is caused by the human immunodeficiency virus (HIV). This virus is passed from one person to another through blood-to-blood and sexual contact. The HIV-1 RT is a major target for anti-AIDS drugs. To seek for more potent drugs, information on the structure and interaction between drug and enzymes are needed.

Avian influenza A virus, especially subtype H5N1, becomes one urgent issue since its first infection in Hong Kong in 1997. The virulent H5N1 is highly pathogenic and has been shown to cross the species barrier and infect humans. This disease has caused many human deaths in various countries in Asia and Europe during the past few years. Hemagglutinin (HA) cleavability plays a pivotal role in determining of the virulent of influenza viruses.

Both AIDS/HIV and H5N1 avian influenza are emerging infectious diseases and have recently attracted substantial scientific and popular attention. Significance and backgrounds of the two diseases are given as the following.

1.2 HIV-1 RT inhibitors complexes

1.2.1 History and properties of HIV/AIDS

HIV, a causative agent of AIDS, was first discovered in 1981 and the FDA approves the first HIV antibody test for blood supplies in 1985 [1]. The patient infected with HIV leading to AIDS died continually and the number of people living with HIV/AIDS rapidly increase. AIDS is thus one of the most devastating diseases over twenty years now and deaths due to AIDS are still remarkably rising in the developing countries. By the end of 2007, between 30.6 and 36.1 million people were believed to live with HIV, and it killed an estimated 2.1 million people that year, including 330,000 children; there were 2.5 million new infections (see Table1.1) [2].

Table 1.1 The statistics on the incidence of AIDS/HIV in 2007 reported by UNAIDS [2].

The latest statistics on the incidence of AIDS and HIV in 2007		
	Estimate (millions)	**Range (millions)**
People living with HIV/AIDS	33.2	30.6–36.1
Adults living with HIV/AIDS	30.8	28.2–33.6
Children living with HIV/AIDS	2.5	2.2–2.6
People newly infected with HIV	2.5	1.8–4.1
Adults newly infected with HIV	2.1	1.4–3.6
Children newly infected with HIV	0.42	0.35–0.54
AIDS deaths	2.1	1.9–2.4
Adult AIDS deaths	1.7	1.6–2.1
Child AIDS deaths in 2007	0.33	0.31–0.38

HIV belongs to a class of viruses called retrovirus in which its genetic material is composed of RNA (Ribonucleic Acid). This means that it has to make a DNA (deoxyribonucleic acid) copy of its RNA in order to replicate. Structure of HIV (see Figure 1.1) exists as roughly spherical particles (100-150 nm), their particles surround themselves with a coat of fatty material known as the viral envelope. Projecting from this are around 72 little spikes, which are formed from the proteins gp120 and gp41. Just below the viral envelope is a layer called the matrix. Inside the core are three enzymes required for HIV replication called reverse transcriptase (RT), integrase (IN) and protease (PR).

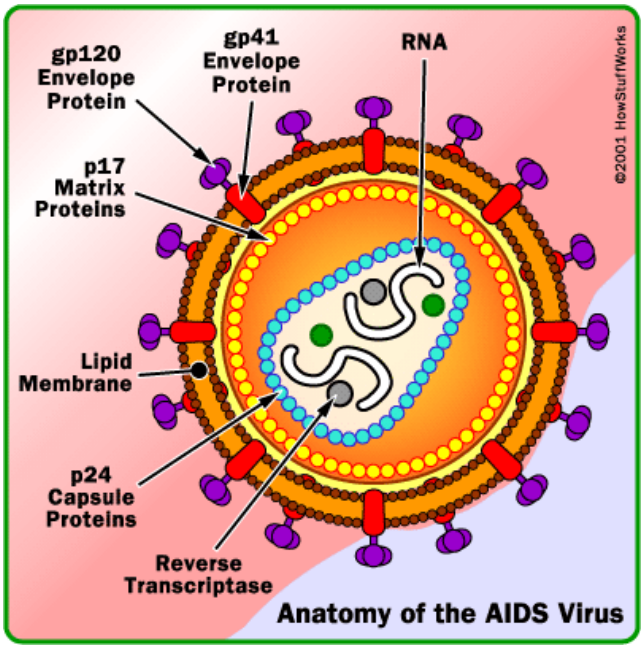


Figure 1.1 Structure and genome organization of the HIV-1 virus [3].

Like all viruses, HIV can replicate only inside cells, commandeering the cell's machinery to reproduce. Only HIV and other retroviruses, however, once inside a cell, use an enzyme to convert their RNA into DNA, which can be incorporated into the host cell's genes. The life cycle of HIV [3] shown as following:

Binding - The HIV attaches to the immune cell when the gp120 protein of the HIV virus binds with the CD4 protein of the T-helper cell. The viral core enters the T-helper cell and the virion's protein membrane fuses with the cell wall.

Reverse transcription - The viral enzyme, RT, copies the virus's RNA into DNA, which is compatible with human genetic material.

Integration - The newly created DNA is carried into the cell's nucleus by the enzyme, IN, and it binds with cell's DNA. HIV DNA is called a provirus.

Transcription - The viral DNA in the nucleus separates and creates messenger RNA (mRNA), using the cell's own enzymes. The mRNA contains the instructions for making new viral proteins.

Translation - The mRNA is carried back out of the nucleus by the cell's enzymes. The virus then uses the cell's natural protein-making mechanisms to make long chains of viral proteins and enzymes.

Assembly - RNA and viral enzymes gather at the edge of the cell. An enzyme, called protease, cuts the polypeptides into viral proteins.

Budding - New HIV virus particles pinch out from the cell membrane and break away with a piece of the cell membrane surrounding them. The newly matured HIV particles are ready to infect another cell and begin the replication process all over again. In this way the virus quickly spreads through the human body. And once a person is infected, they can pass HIV via to others in their bodily fluids.

1.2.2 HIV-1 RT as a molecular target for drug therapy

The HIV-1 RT has been the subject of intensive study due to the importance of its key role in the life cycle of the HIV-1 virus. HIV-1 RT is part of the HIV capsid and has an essential role in the replication of the AIDS virus. This enzyme transcribes the single-stranded RNA genome of the AIDS retrovirus to a double-stranded DNA which can then be integrated into the host genome [4]. RT is a heterodimer that consists of a 66 kDa subunit (p66) and a 51 kDa subunit (p51) as

shown in Figure 1.2. Both subunits contain ‘finger’, ‘palm’, ‘thumb’, named according to their resemblance to a right hand, and connection subdomains, with the ribonuclease H (RNaseH) domain found only in the p66 subunit [5]. The p51 subunit is originated from p66 by proteolytic cleavage of the C-terminal RNaseH domain. Consequently, both subunits have the same amino acid sequence. However, the structural organization of the subdomains is completely different, resulting in an asymmetric dimer with the active site only located in the p66 subunit. RT exhibits both DNA polymerase and RNaseH activities to complete the reverse transcription. The p66 subunit containing the catalytic activity region has a functional polymerase active site and a large cleft that binds the template-primer DNA, a DNA-binding cleft, formed by the p66 fingers, palm and thumb subdomains. Even though only the p66 subunit seems to play an important role in the catalytic function of HIV-1 RT, the p51 subunit is necessary for the enzyme to be fully active conformation of the p66 subunit. For this reason, the p66 subunit would be focus on discussing the catalytic activities of this enzyme. The p66 palm subdomain not only contains the polymerase active site but also the NNRTIs binding pocket (allosteric site) buried within RT near the polymerase active site (see in Figure 1.2).

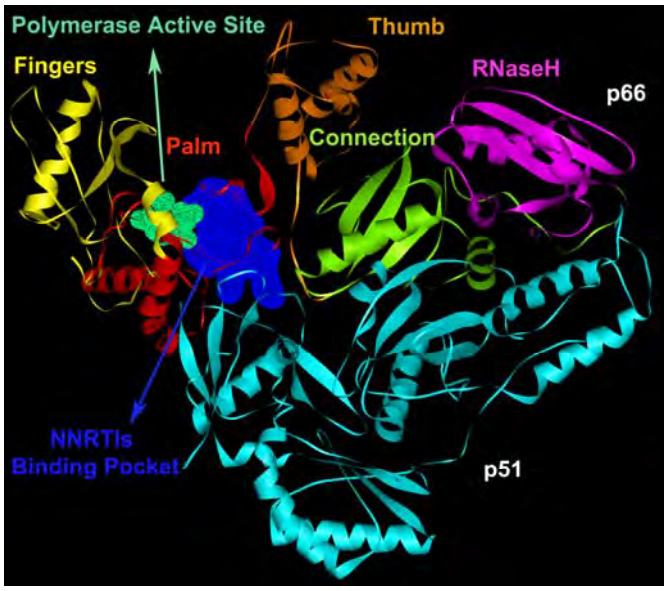


Figure 1.2 Structure of the HIV-1 RT catalytic complex**.**

Crystal structure analysis of HIV-1 RT has revealed important features of the enzyme's structure and function. Structures of unliganded HIV-1 RT, in complexes with various NNRTIs [6-12], and in complex with a DNA and an antibody Fab fragment [13, 14] have been reported. Structural studies of the crystallographic structure comparison provided information on the RT has several extremely flexible regions which have been proposed to be essential for the polymerization process. The binding site, which is a large hydrophobic cavity, is stabilized by hydrophobic interactions and hydrogen bonding. Additionally, the NNRTIs-binding pocket is elastic, and its conformation depends on the size, the specific chemical structure, and the binding mode of individual NNRTIs [15]. This information should be useful in the development of new or improved inhibitors of HIV-1 RT.

1.2.3 HIV-1 RT inhibitors

Currently approved drugs targeting on HIV-1 RT to treat HIV infection can be divided into two classes: nucleoside-analogous RT inhibitors (NRTIs) and nonnucleoside RT inhibitors (NNRTIs). NRTIs, such as AZT (1-[4-azido-5-(hydroxymethyl)oxolan-2-yl]-5-methyl-pyrimidine-2,4-dione), ddC (4-amino-1-[5-(hydroxymethyl)oxolan-2-yl]-pyrimidin-2-one) and ddI (9-[5-(hydroxymethyl)oxolan-2-yl]-3*H*-purin-6-one) [16, 17] are competitive inhibitors of the nucleotide substrate which lack the 3'-OH group at the terminal DNA primer found on normal nucleosides and act as chain terminators when incorporated into viral DNA by HIV-1 RT. However, their application is limited due to the high serious side effects, especially damage to mitochondria [18-20] and the emergence of drug-resistant mutants. Important mutations giving rise to NRTI resistance include M41L, A62V, K65R, D67N, K70R, V75I, F77L, F116Y, Q151M, M184V/L, T215Y/F and K219Q. Previous studies suggest that two main factors are involved in HIV mutation to resist NRTIs: rescue of DNA synthesis by increased repolymerization reaction [21] such as AZT resistance associated with M41L, D67N, K70R, T215Y/F, K219E/Q: and change in drug/substrate molecular recognition [22-24] such as resistance to the 3TC analogues (for example, the M184V mutant has increased steric hindrance, blocking the appropriate binding of 3TC derivatives but still permitting the incorporation of normal nucleosides) and mutations (K65R) giving cross-resistance to ddI, ddC and 3TC. NNRTIs are non-competitive inhibitors which are highly specific

for HIV-1 RT at a common allosteric site approximately 10 Å from the polymerase active site. In addition, they are less likely to cause adverse side effects by disruption of the normal DNA polymerase activity [19, 20]. Interest is, then, focused to the NNRTIs group, aiming to understand and to provide detailed information insight into HIV-1/NNRTIs interaction.

The structures of NNRTIs, e.g., Efavirenz (6-chloro-4-(2-cyclopropylethynyl)-4-(trifluoromethyl)-2,4-dihydro-1*H*-3,1-benzoxazin-2-one), nevirapine (11-cyclopropyl-4-methyl-5,11-dihydro-6*H*- dipyrido[3,2-*b*:2′,3′-*e*][1,4]diazepin-6-one), delavirdine (*N*-[2-([1]carbonyl)-1*H*-indol-5-yl]methanesulfonamide, emivirine (6-benzyl-1-(ethoxymethyl)-5-isopropyluracil) and etravirine (4-[6-Amino-5-bromo-2- [(4-cyanophenyl)amino] pyrimidin-4-yl]oxy- 3,5-dimethylbenzonitrile) [15, 18, 25, 26], are non-competitive inhibitors. It is thought that these act by distorting the HIV-1 RT enzyme in several important regions near the polymerase active site [26]. In addition, they are less likely to cause adverse side effects by disruption of the normal DNA polymerase activity. However, the rapidity of the selection of drug resistant HIV in patients was such that mutations in the virus made NNRTIs unusable in therapy. Among the mutations in RT were described that the K103 mutation affects nevirapine (NVP) and especially efavirenz (EFV) potency by stabilizing the closed form of the binding pocket through the formation of a hydrogen bond between the K103 side-chain amide and the Y188 phenoxyl oxygen, reducing the rate of inhibitor entry [27]. In contrast, the Y181C and Y188L mutations affect the binding of an EFV, emivirine (EMV) and especially NVP by loss of favorable aromatic ring interactions. Mutation of L100 residue causes resistance to NVP and EFV that probably can be described in the term of steric interference [15].

This growing problem of drug resistance makes finding new potent inhibitors a vital. It is known that in spite of intensive experimental investigations, the detailed origins of the binding affinity of inhibitors interacting in NNRTIs active site of HIV-1 RT remain unclear. Also, understanding of which amino acid residues are essential for the interaction with the mentioned inhibitors, leading to sources of drug resistance. Recently, the decomposition free energy, representing interaction between HIV-1 protease inhibitors and each residues of enzyme was found to be used as the criteria for mutation prediction, i.e., direct correlation between experimental drug

resistance and the calculated binding free energies was observed [28]. In addition, the distribution and binding of water molecules in the cavity of the enzyme-inhibitor complex is known to play an important role to the drug binding.

Aims of this part of work are to provide detailed information insight into drug-target interaction, structure, solvation, and thermodynamics properties using molecular dynamics simulations. Thus, attempts were made to seek for reflection between clinical mutation data and the simulated molecular properties in order to understand source of mutation at molecular level of the four HIV-1 RT inhibitors, EFV, NVP, EMV and etravirine (ETV), embedded in the binding site of HIV-1 RT. This fundamental structural information is helpful for the understanding of the enzyme-inhibitor interaction, especially, for the design of more potent inhibitors active against mutant HIV-1 RT.

1.3 Furin – substrates (Hemagglutinin) complexes

1.3.1 History and properties of influenza virus

Influenza A viruses have been isolated from a variety of animals, including humans, pigs, horses, sea mammals, and birds. Influenza pandemics, defined as global outbreaks of the disease due to viruses with new antigenic subtypes, have exacted a high death toll from human populations. The most devastating pandemic, the so-called Spanish influenza of 1918 to 1919, resulted from an H1N1 virus and caused the deaths of at least 20 million people worldwide. Other much less catastrophic pandemics occurred in 1957 (Asian influenza [H2N2 virus]), 1968 (Hong Kong influenza [H3N2 virus]), and 1977 (Russian influenza [H1N1 virus]). Then in 1997, an H5N1 influenza virus was isolated from a 3-year-old boy in Hong Kong [29, 30], who died of extensive influenza pneumonia. The avian H5N1 virus is becoming more deadly in a growing number of bird species and mammals. This flu was unique because it seemed to be carried by chickens and moved directly from chickens to people. Figure 1.3 presents the cumulative number of countries from around the world with human cases of H5N1 since 1997 through December 31, 2008. In addition number of confirmed human cases of avian influenza A H5N1 from 2004-2009 reported by WHO show in Table1.2. With the Hong Kong outbreak, the risk that it will acquire the ability to have efficient and sustained human-to-human transmission

is present as long as opportunities for human infections occur. These opportunities will persist as long as the virus continues to circulate in domestic birds, perhaps for years to come.

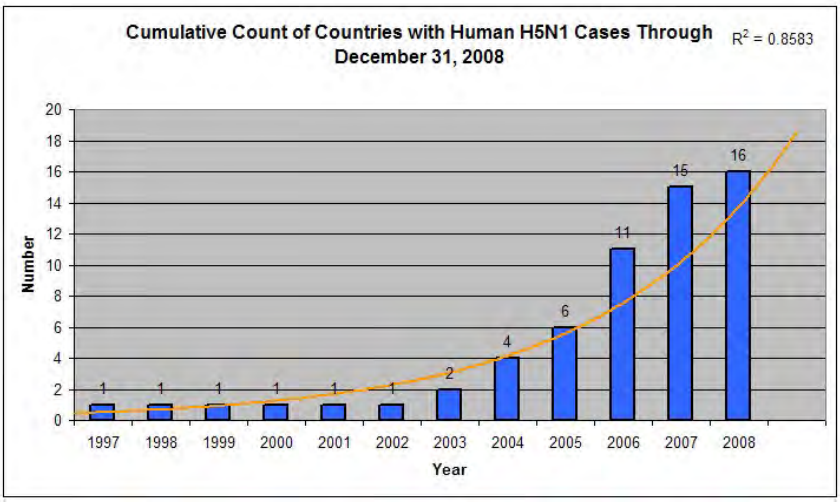


Figure 1.3 Cumulative count of countries with human cases of H5N1 since 1997 through December, 2008 [31].

Influenza viruses, which belong to the *Orthomyxoviridae* family, are classified as A, B, and C based on antigenic differences in their nucleoprotein (NP) and matrix (M1) protein. All avian influenza viruses are classified as type A, it genomes consist of eight separate (Figure 1.4). The current understanding of the influenza virus genes and their functions provides in Table 1.3. Influenza type A subtypes is based on the antigenicity of two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). Currently, 16 HA and 9 NA subtypes have been identified among influenza A viruses [29]. The name of virus strains can be classified by type of virus, geographic origin, strain, and subtype (see Figure 1.5).

Table 1.2 Cumulative number of confirmed human cases of avian influenza A/(H5N1) from 2004-2009 reported by WHO [32].

Country	2004		2005		2006		2007		2008		2009		Total	
	cases	death	cases	death	cases	death	cases	death	cases	death	cases	death	cases	death
Azerbaijan	0	0	0	0	8	5	0	0	0	0	0	0	8	5
Bangladesh	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Cambodia	0	0	4	4	2	2	1	1	1	0	0	0	8	7
China	0	0	8	5	13	8	5	3	4	4	7	4	37	24
Djibouti	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Egypt	0	0	0	0	18	10	25	9	8	4	9	0	60	23
Indonesia	0	0	20	13	55	45	42	37	24	20	0	0	141	115
Iraq	0	0	0	0	3	2	0	0	0	0	0	0	3	2
Lao	0	0	0	0	0	0	2	2	0	0	0	0	2	2
Myanmar	0	0	0	0	0	0	1	0	0	0	0	0	1	0
Nigeria	0	0	0	0	0	0	1	1	0	0	0	0	1	1
Pakistan	0	0	0	0	0	0	3	1	0	0	0	0	3	1
Thailand	17	12	5	2	3	3	0	0	0	0	0	0	25	17
Turkey	0	0	0	0	12	4	0	0	0	0	0	0	12	4
Viet Nam	29	20	61	19	0	0	8	5	6	5	2	2	106	51
Total	46	32	98	43	115	79	88	59	44	33	18	6	409	252

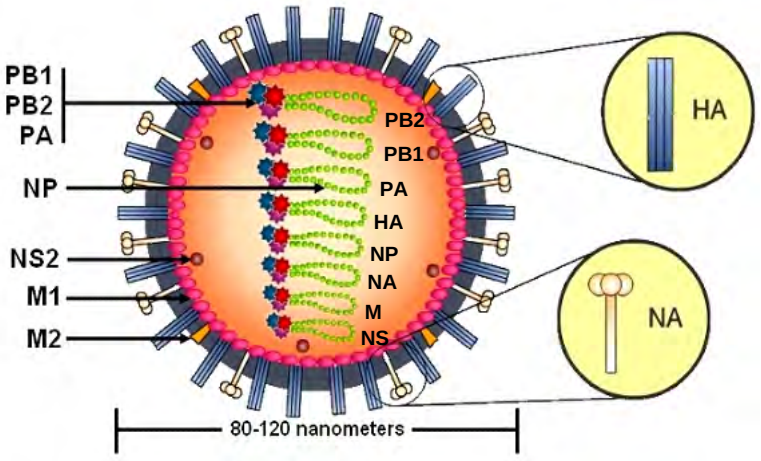


Figure 1.4 Structure and genome organization of influenza A virus.

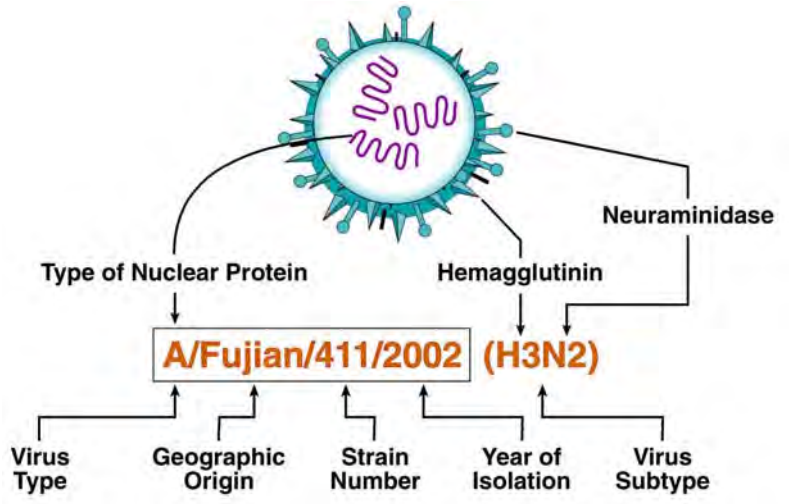


Figure 1.5 Diagram of influenza nomenclature.

The sequence of steps of influenza‘s replication cycle involves;

(1) HA binding to sialic acid-containing receptors on the membrane surface, the virus enters the cell by receptor mediated endocytosis.

(2) A low pH in the endosome induces a conformational change in HA, resulting in membrane fusion between the viral envelope and the endosomal membrane. Within the endosome, the M2 proton channel exposes the viral core to low pH, resulting in dissociation of M1 from ribonucleoprotein (RNP) and leading to a release of RNP to the cytoplasm.

(3) RNP is then transported to the nucleus, most probably by nuclear localization signals in proteins composed of the RNP complex (PB1, PB2, and NP).

(4) New viral proteins are translated from transcribed messenger RNA.

(5) New viral RNA is encased in the capsid protein, and together with new matrix protein is then transported to sites at the cell surface where envelope HA and NA components have been incorporated into the cell membrane.

(6) Progeny virions are formed and released by the NA protein removes SA to liberate newly synthesized viruses from infected cells.

The avian influenza A virus (H5N1) is highly pathogenic, which the HA glycoprotein plays a pivotal role. Proteolytic activation of the HA is essential for viral infectivity and for spread of the avian influenza virus through the host's body [33, 34]. Understanding source of high pathogenicity of influenza A virus subtype H5N1 is the goal of this study.

Table 1.3 The influenza virus genes and their functions [30]**.**

Genes of influenza A and their presumed functions	
PB2 gene	Codes for an RNA polymerase involved in cap binding (sealing end of molecule); part of transcriptase, which is an enzyme that converts DNA into types of RNA
PB1 gene	Codes for an RNA polymerase involved in elongation of the molecule; part of transcriptase
PA gene	Codes for an RNA polymerase that may serve as a protease; part of transcriptase
HA gene	Codes for hemagglutinin; involved with virus attach to host cell and membrane fusion
NP gene	Codes for the nucleoproteins; Types A, B, and C have different nucleoproteins; part of transcriptase complex
NA gene	Codes for neuraminidase; involved with release of virus from the host cell
M1 gene M2 gene	Matrix protein; different sections of M2 gene the genetic code of the gene are read to produce the two proteins that open channels in the cell membrane and allow charged atoms/molecules to pass through
NS1 gene NS2 gene	Codes for two different nonstructural proteins whose function is still unknown; as above, different sections of the code are used for each

1.3.2 Hemaglutinin and pathogenicity

As already mentioned, the HA glycoprotein plays a pivotal role in the pathogenicity of the avian viruses. The HA is initially synthesized as an inactive precursor (HA0, Figure 1.6) which is then activated through a controlled proteolytic cleavage by protease into HA1 and HA2 subunits. HA1 mediates virus binding to host-cell receptors while HA2 promotes the release of the viral RNA complexed with polymerase through membrane fusion [29, 33-37]. Without proteolysis, the fusion peptide cannot occur and therefore the virus is essentially non-infectious. The x-ray crystal structure of HA0 has been solved and shows that the cleavage site forms a loop that extends from the surface of the molecule, and it is the composition and structure of the cleavage loop region that dictate the range of proteases that can potentially activate infectivity. Amino acid sequence comparisons between naturally occurring high-pathogenicity avian influenza (HPAI) viruses and low-pathogenicity avian influenza (LPAI) viruses have shown that HA with restricted cleavability. LPAI usually have a single arginine (R) whereas those with high cleavability (HPAI) have the insertion of the –RRRKK– (where K is Lysine) residues situated upstream of the HA cleavage site, which have potential to cause devastating pandemics in the future (Table 1.4) [29, 34, 36, 38, 39].

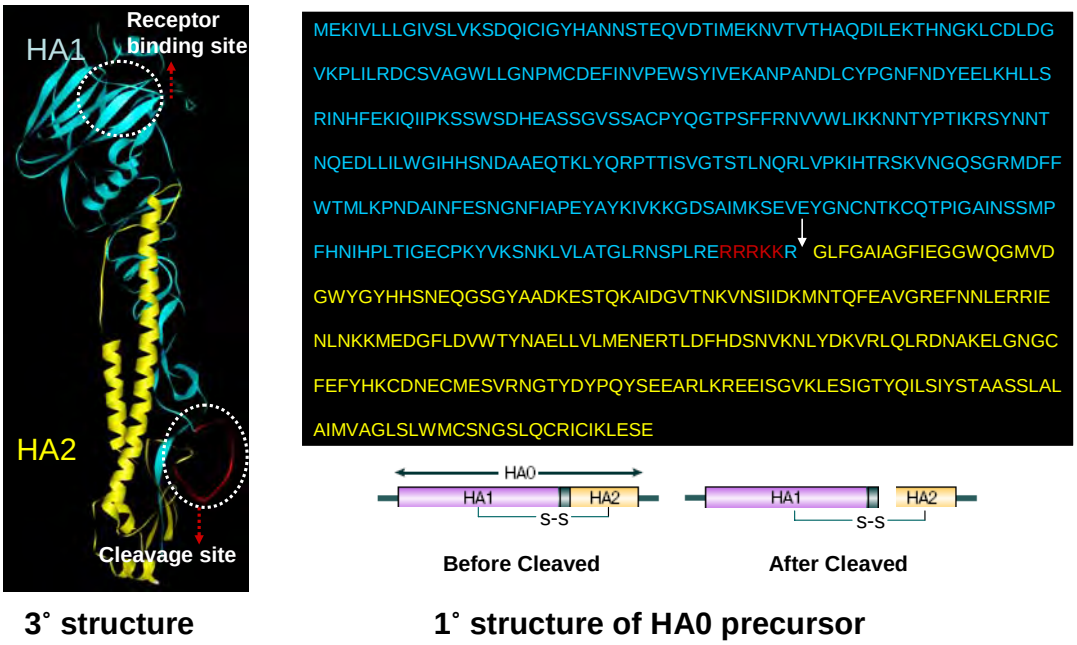
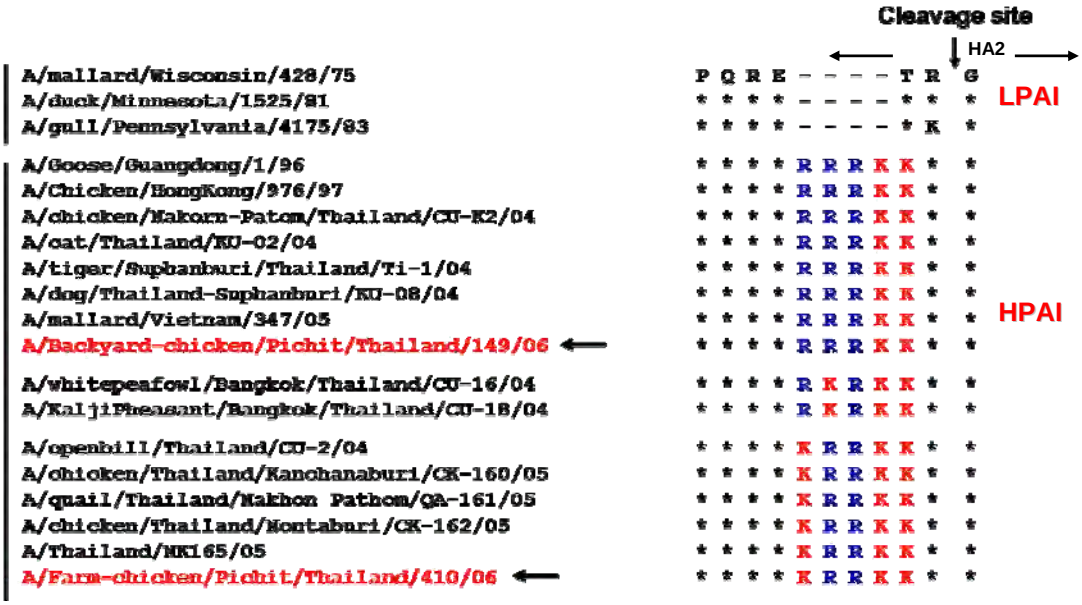


Figure 1.6 The 3D structure and primary sequences of precursor H5N1 Hemagglutinin (HA0) whereas, residues at cleavage site in circle) [37].

Table 1.4 The comparison of amino acid sequences at HA's cleavage site between HPAI viruses and LPAI viruses.



1.3.3 Furin protease

The HA cleavability relies on the presence of basic amino acids and relates directly to pathogenicityof the influenza virus. The HAs of LPAI are usually cleaved in only a limited number of organs, so that the viruses cause only localized infections in the respiratory or intestinal tract. In contrast, the HAs of HPAI viruses, which are cleaved by furin, ubiquitous proteases, are present in a broad range of different host cells [39-41].

Furin, as seen in Figure 1.7, is a mammalian subtilisin/Kex2p-like endoprotease that has a ubiquitous tissue distribution and cycles between the trans-Golgi network, the cell surface, and the endosomes. Furin appears to be a highly specific enzyme, cleaving pro-protein precursors at specific consensus sequence –RXK/RR–, usually to produce biologically active products [40, 42-47]. Based on the X-ray structure of mouse furin complexed with the dec-RVKR-cmk inhibitor, the

furin's catalytic ability is considered to be originated from the catalytic triad residues (S368, H194 and D153) and the oxyanion hole (formed by the carboxamide nitrogens of N295 and backbone nitrogen of S368 to carbonyl oxygen of the inhibitor/substrate's centered arginine) at the active site [43]. The rational here is that the D153-H194-S368 catalytic triad of furin [43] is a structural analogue of the conserved D102-H57-S195 residues found in serine proteases [48]. Thus, the cleavage mechanism of HPAI subtype H5 (HPH5) virus by furin is proposed as the most likely reaction mechanism of serine proteases, as shown in Figure 1.8. The catalytic H194 is envisaged to act as a general base to accept a proton from S368. The deprotonated S368 then becomes a nucleophile which directly attacks the carbonyl carbon of the S1-Arg, the reacting residue on the HPH5's cleavage site, with the formation of the tetrahedral intermediate. A pair of electrons from the double bond of the carbonyl oxygen moves to the oxygen, then this electrons move back to break the nitrogen-carbon peptide bond of HA substrate.

Figure 1.7 Crystal structure of furin/dec-RVKR-cmk inhibitor complex [43].

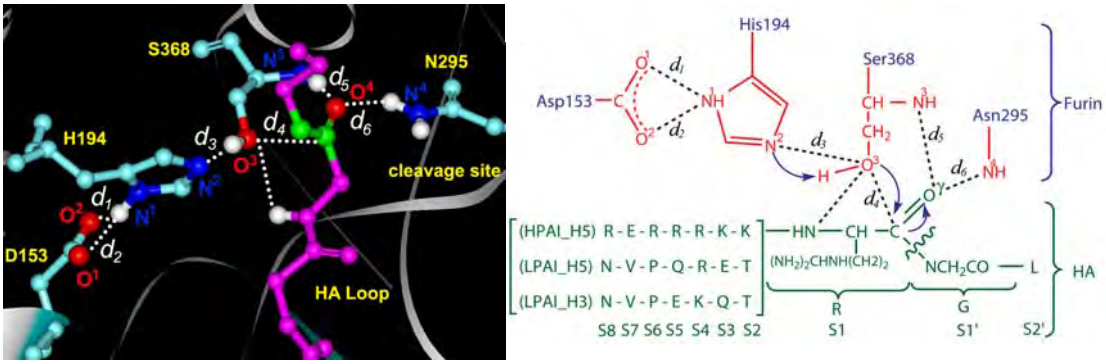


Figure 1.8 Proposed cleavage mechanism of HA by furin and definitions of d_1-d_6.

The objective of this part of work is to seek for an answer on the HPAI of H5. MD simulations were carried out for both high and low pathogenic forms of H5. For comparison, simulations were also extended to LPAI H3.

1.4 Research objectives

In the present study, MD simulations were applied for the following systems**.**

1.4.1 HIV-1 RT/inhibitors complexes, aimed:

- To understand the molecular basis of each of the four HIV-1 RT/NNRTI interactions.
- To calculate the binding free energies and water accessibilities in the binding site of each of the four HIV-1 RT/NNRTI complexes.

1.4.2 Furin/hemagglutinin-substrates, aimed:

- To understand the molecular basis of the inserted hemagglutinin strains, HPAI, and non-inserted strains, LPAI, complexed with furin.
- To calculate the intra- and intermolecular interactions and geometries of the substrate-furin complex, potentially involved in the cleavage mechanism.

- To compare the substrate specificity and recognition of furin at the cleavage site of the HPAI virus subtype H5 and LPAI subtypes H5 and H3.

CHAPTER II

THEORY

2.1 Homology modeling

One method that can be applied to generate reasonable models of protein structures is homology modeling [49]. This procedure develops a three-dimensional model from a protein sequence based on the structures of homologous proteins. The steps to creating a homology model are as follows:

2.1.1 Identifying homologues

First, selecting the sequence of a protein is needed to model the three dimensional structure (the unknown). Apply sequence search methods to identify proteins with unknowns ,then, proteins are homologous with unknown and use the three-dimensional structures of these proteins to develop a model of the structure of unknown.

2.1.2 Aligning sequences

A critical step in the development of a homology model is the alignment of the unknown sequence with the homologues. Factors to be considered when performing an alignment are (1) which algorithm to use for sequence alignment, (2) which scoring method to apply, and (3) whether and how to assign gap penalties.

Sequence alignments generally are based on the dynamic programming algorithm of Needleman and Wunsch [50]. Current methods include FASTA, Smith-Waterman, and BLASTP, with the last method differing from the first two in not allowing gaps.

2.1.3 Identification of structurally conserved and structurally variable regions

After the known structures are aligned, they are examined to identify the structurally conserved regions (SCRs) from which an average structure, or framework, can be constructed for these regions of the proteins. Variable regions (VRs), in which each of the known structures may differ in conformation, also must

be identified. The VRs usually lie on the surface of the proteins and form the loops where the main chain turns.

2.1.4 Generating coordinates for the unknown structure

When generating coordinates for the unknown structure, one needs to model main chain atoms and side chain atoms, both in SCRs and VRs. For the SCRs, it is straightforward to generate the coordinates of the main chain atoms of the unknown structure from those of the known structure(s). Side chain coordinates are copied if the residue type in the unknown is identical or very similar to that in the known homologues. For other side chain coordinates one can apply a side chain rotamer library in a systematic approach to explore possible side chain conformations. It may be desirable to weight the contribution of each homologue in each SCR based on the extent of similarity with the unknown. In the event that some coordinates in the unknown are undefined in the SCRs, regularization can be used to build and relax both main chain and side chain atoms in those regions. Note that this procedure should be used only if the region of undefined atoms is one or two residues in length.

For the VRs, a variety of approaches may be applied in assigning coordinates to the unknown. Recall that these regions will correspond most often to the loops on the surface of the protein. If a loop in one of the known structures is a good model for that of the unknown, then the main chain coordinates of that known structure can be copied. Side chain coordinates of residues that are similar in length and character also may be copied.

2.1.5 Evaluation and refinement of the structure

For a homology model from any source, it is important to demonstrate that the structural features of the model are reasonable in terms of what is know about protein structures in general. That is, researchers have analyzed three-dimensional structures of proteins from which basic principles of protein structure and folding have been developed. Several programs are available to assist in this analysis of correctness of a homology model. The most popular analysis of correctness is main chain conformations in acceptable regions of the Ramachandran map by using program such as PROCHECK [51] , WHAT IF [52] and Ramachandran plot 3.2 [53].

2.2 Molecular docking

The biological activities of many proteins depend on the specific recognition of proteins. Although many high-resolution (X-ray) structures of protein complexes are available in the protein data bank (PDB), a vast number of protein complex structures are not yet determined. Thus, computational determination of complex structures by docking studies will continue to be an invaluable tool, not only for the insight they yield on the thermodynamics of molecular recognition, but also for their potential utility in understanding protein interaction.

2.2.1 ZDOCK Program

ZDOCK is a rigid body protein–protein-docking algorithm. It fully explores three translational and three rotational degrees of freedom. The rotational space is searched explicitly, and the translational space is searched by using Fast Fourier Transform (FFT) algorithm [54]. Program ZDOCK2.3 had the most powerful scoring function by combining pair-wise shape complementarity (PSC) with desolvation (DE) and electrostatics (ELEC), PSC+DE+ELEC [55]. ZDOCK is freely available to academic researchers.

The basic procedure of ZDOCK can be summarized in Figure 2.1 and following steps:

(i) Start with processed the three-dimensional (3D) structures of the receptor and the ligand by removing or blocking residues that we postulated not to be part of the binding site.

(ii) Then ran ZDOCK on the processed structure files and generated 1000 predictions.

(iii) Next clustered the structures.

(iv) We manually inspected all the clusters with a prediction ranked < 200 by ZDOCK and selected the top rank 5 or 10 cluster.

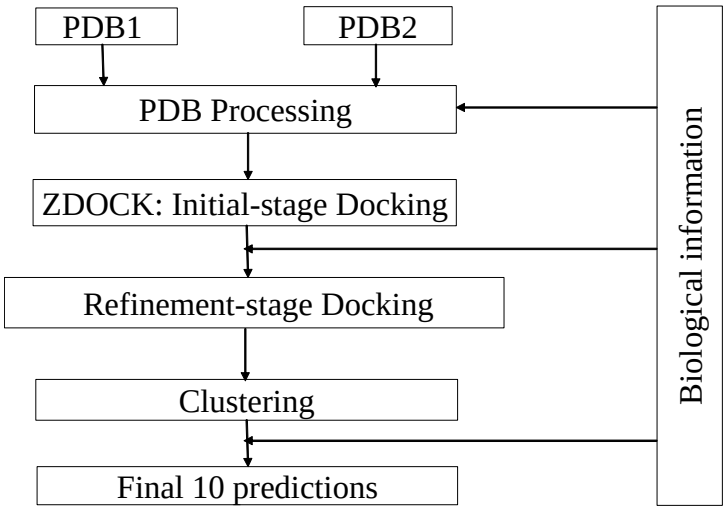


Figure 2.1 Schematic of the basic procedure of ZDOCK program [54]..

2.2.2 Scoring function

The basic search algorithm of ZDOCK has been described previously. The electrostatics energy was computed by using the Coulombic formula, which is expressed as a function of the electrical potential generated by the receptor (*R*, the larger protein) and the partial charges of ligand (*L*, the smaller protein) atoms. The resulting electrostatics energy was multiplied with a scaling factor β and adds it to the PSC and DE scores. In practice, this sum can be directly evaluated by using the FFT search algorithm. Four discrete functions on an N×N×N grid, $R_{PSC+ELEC}$, $L_{PSC+ELEC}$; R_{DE} and L_{DE} are used to describe the shape and desolvation properties of the receptor and ligand, and the PSC + DE + LELC scoring function, $S_{PSC + DE + ELEC}$, is expressed as correlations of these four functions:

$$\mathrm{Re}[R_{PSC+ELEC}] = \mathrm{Re}[L_{PSC+ELEC}]$$

$$= \begin{cases} 3.5 & \text{solvent excluding surface layer of the protein} \\ 3.5^2 & \text{protein core} \\ 0 & \text{open space} \end{cases}$$

$$\mathrm{Im}[R_{PSC+ELEC}]$$

$$= \begin{cases} \beta \times (\text{electric potential of all open space receptor atom}) \\ 0 \quad \text{otherwise} \end{cases}$$

$$\mathrm{Im}[L_{PSC+ELEC}] = \begin{cases} -1 \times (\text{atom charge}) \\ 0 \quad \text{otherwise} \end{cases}$$

$$\mathrm{Re}[R_{DE}] = \mathrm{Re}[L_{DE}] = \begin{cases} \text{sum of PSC and atomic contact energy scores of all open space nearby atoms} \\ 0 \quad \text{otherwise} \end{cases}$$

$$\mathrm{Im}[R_{DE}] = \mathrm{Im}[L_{DE}] = \begin{cases} 1 \text{ if this grid point is the nearest grid point of a atom} \\ 0 \quad \text{otherwise} \end{cases}$$

$$S_{PSC+DE+ELEC} = \mathrm{Re}[R_{PSC+ELEC} \bullet L_{PSC+ELEC}] + \frac{1}{2}(\mathrm{Im}[R_{DE} \bullet L_{DE}]) \qquad (2.1)$$

where Re and Im denote the real and imaginary parts of a complex function. If a protein atom has > 1 $Å^2$ solvent-accessible area, calculated by using a water probe radius of 1.4 Å, it is considered a surface atom. Otherwise, it is a core atom. The "solvent excluding surface layer of the protein" is defined by the grid points corresponding to surface atoms. All other grid points corresponding to core atoms are in the "protein core." "Nearby atoms" are atoms within the distance cutoff of a grid point. $\mathrm{Im}[R_{DE} \bullet L_{DE}]$ is divided by 2 because each atom pair has been counted twice.

2.3 Potential energy function

The same for all approaches, theoretical studies to investigate the relationships between structures, function and dynamics at the atomic level are based on several assumptions. As many of the problems, especially in biological systems involve many atoms, quantum mechanics is too expensive to treat these systems. However, the problems become much more adaptable when turning to empirical potential energy functions. One of the most important is that no drastic changes in electronic structure are allowed, for instance, no events like bond making or breaking can be modeled.

Potential energy function looks at the bonds as springs which can be stretched called bond stretching. Compressed, bent at the bond angles, twisted in torsional angles and non-bonded interactions between atoms are also considered. It can be interpreted in terms of a relatively simple of the intra- and inter- molecular forces within the system. This is the general equation of the potential energy function

$$V(R) = V_{bonded}(R) + V_{non-bonded}(R) \tag{2.2}$$

2.3.1 Bonded interactions

The $V_{bonded}(R)$ is sum of the following three terms

$$V_{bonded}(R) = E_{bonds} + E_{bond-angles} + E_{torsion-angles} \tag{2.3}$$

Which correspond to three types of atom movement, r_{23}, θ_{234} and ϕ_{1234} shown in Figure 2.2.

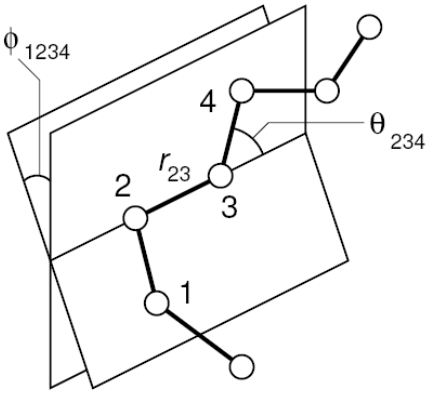


Figure 2.2 Geometry of a simple chain molecule, illustrating the definition of interatomic distance r_{23}, bend angle θ_{234}, and torsion angle ϕ_{1234} [56].

Bonds stretching (E_{bonds}): The interaction between two atoms directly bonded to each other is assumed to be harmonic. Bond stretching energy can be represented as following,

$$E_{bonds} = \sum_{n=1}^{i} k_b (r - r_0)^2 \tag{2.4}$$

Where r is the length of the ith bond (Å)

r_0 is the equilibrium length for the ith bond (Å)

k_b is the bond stretching constant (kcal·mol^{-1} (Å)^2)

Bond angles ($E_{bond\text{-}angles}$): The interaction between three connected atoms is also assumed to be harmonic. Angle bending energy is a function of angular displacement. The bending energy equation is also based on Hooke's law

$$E_{bonds-angles} = \sum_{n=1}^{i} k_\theta (\theta - \theta_0)^2 \tag{2.5}$$

Where θ is the angle between two adjacent bonds (°)

θ_0 is the equilibrium value for the ith angle (°)

θ_k is the angle bending constant (kcal·mol^{-1} (°)^2)

Dihedral (Torsion) angles: There are four atoms connected by three bonds, so that if we look straight down the second bond, the dihedral angle is the angle between the first and third bonds. The energies associated with dihedral angles are treated using a cosine series. The torsion energy is modeled by a simple periodic function.

$$E_{torsion-angles} = \sum_{1,4\ pair} K_\phi \left[1 - Cos\left(n\phi\right)\right] \tag{2.6}$$

Where K is the torsional barrier (kcal·mol^{-1})

n is the periodicity

ϕ is the torsional angle

2.3.2 Non-bonded interactions

The non-bonded interactions are contributed by two functions which are van der Waals interaction energy and the electrostatic interaction energy.

$$V_{non-bonded}(R) = E_{vdw} + E_{electrostatic} \tag{2.7}$$

The van der Waals energy, E_{vdw}, describes the repulsion or attraction between atoms that are not directly bonded. This term can be interpreted as the part of the interaction which is not related to atomic charges. The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The repulsive force comes up at short distances because the electron-electron interaction is strong. The attractive force, also referred to as dispersion force, arises from fluctuations in the charge distribution in the electron clouds. The fluctuation in the electron distribution on one atom or molecule gives rise to an instantaneous dipole which, in turn, induces a dipole in a second atom or molecule giving rise to an attractive.

The Lennard-Jones potential is the most commonly used form. It can be written as

$$E_{LJ}(R) = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] \tag{2.8}$$

with two parameters: r_0, the diameter, and ε, the well depth, r^{-6} represents the attraction interaction and the repulsive part is given by r^{-12}.

Another part of non-bonded interaction is $E_{electrostatic}$ which is due to interaction distribution of the electron. It is created by negative and positive parts of the molecule. The electrostatic interaction between a pair of atoms is represented by Colomb potential; D is the effective dielectric function for the medium and r is the distance between two atoms having charges q_i and q_k.

$$E_{electrostatic} = \sum_{\substack{nonbonded \\ pair}} \frac{q_i q_k}{D r_{ik}} \tag{2.9}$$

Up to date development potential energy functions (or force field) provide a reasonably good compromise between accuracy and computational efficiency. They are often calibrated to experimental results and quantum mechanical calculations of small model compounds. Their facility to reproduce physical properties calculable by experiment is tested; these properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy and thermodynamic data. The most commonly used potential energy functions are AMBER [57], CHARMM [58], GROMACS [59] and OPLS [60] force fields. The continuing development of force fields remains an intense area of research with implications for both fundamental researches as well as applied researches in the pharmaceutical industry.

2.4 Energy minimization

Using the force field that has been assigned to the atoms in the system it is essential to find a stable point or a minimum on the potential energy surface by adjusting the atomic coordinates, in order to begin dynamic derivative provide information that can be very useful in minimization procedure. At the minimum the net force on each atom vanishes, *i.e.* the derivative or gradient $-\nabla V(R) = 0$. There can be more than one minimum for a large molecule. The minima are called *local minima*. The ideal solution of geometry minimization is the *global minimum*. Due to numerical limitations, however, it is impossible to exactly reach the global minimum or even the local minimum. In practice, local minimum refers to a point on the potential energy surface where the applied minimization procedure cannot further reduce the function value. Mostly, the magnitude of the first derivative is a rigorous way to characterize convergence. The minimum has converged when the derivatives are close to zero. The typical tolerance, for example, in AMBER program [61, 62] is in the range of 10^{-5} to 10^{-6} kcal·mol^{-1}·Å^{-1}. To reach the minimum the structure must be successively updated by changing the coordinates (taking a step) and checking for the convergence. Each complete cycle of differentiation and stepping is known as minimization iteration. The efficiency of minimization can be judged by both the number of iterations required to converge and number of function evaluations needed per iteration. Typically, thousands of iterations are required for macromolecules to reach the convergence.

Two first-order minimization methods, which are frequently used in molecular modeling, are steepest descent and conjugate gradient methods. Both techniques use the first derivatives of the potential function. Additionally, the Newton-Raphson method which uses both the first and the second derivatives to locate the minimum, namely the second-order method, is also widely used.

2.4.1 Steepest descents method

The steepest descents method uses the first derivatives to determine the direction to move towards the minimum. This direction is defined by the negative first derivative of the potential energy, $-\nabla V(R)$. However, the derivative (gradient) merely points downhill of potential energy surface but not necessarily direct the minimum. Thus the technique so-called *line search*, which is used to locate the minimum along the gradient direction, is required to decide how far to move along the direction or to determine the step size (see blue line in Figure 2.3).

The method is commonly used in the initial step for relaxing the poorly refined structure either resolved from the crystallography or model building since it reasonably converges at the initial step and requires minimal computing time. However, the progress becomes slow when approaching the minimum. This leads to another minimization method known as conjugate gradient to be used.

2.4.2 Conjugate gradient method

The conjugate gradient method also uses the first derivatives of the potential energy. But instead of using local gradient for going downhill as in the steepest descent method, the conjugate gradient technique defines the new gradient direction for each iteration by using information from previous gradient directions to determine the optimum direction for line search. Using an algorithm that produces a complete basis set of mutually conjugate directions, each successive step continually refines the direction toward the minimum. Therefore, the conjugate gradient method is more efficient and gives smaller number of iterations to reach the convergence, comparing to the steepest descents method. The plot of minimization path of the conjugate gradient method is shown in Figure 2.3 (red line).

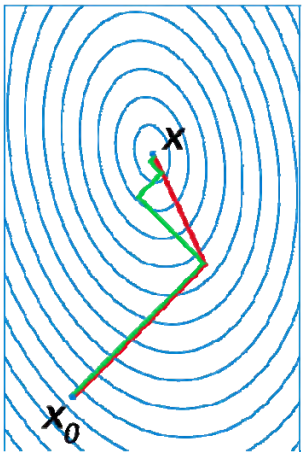


Figure 2.3 A comparison of the convergence of steepest descent with optimal step size (in green) and conjugate gradient (in red) for minimizing the quadratic form associated with a given linear system [63].

Generally, this method converges in approximately *M* steps for a quadratic function, where *M* is the number of degrees of freedom of the function. Note that several terms in the potential energy are quadratic. Nevertheless, the disadvantage is that the line minimizations need to be performed accurately in order to ensure that the conjugate direction is set up correctly and thus time consuming. In addition, the method can be unstable if conformation is so far away from a local minimum.

2.4.3 Newton-Raphson method

The Newton-Raphon method uses the second derivatives providing information about curvature of the function, as well as the first derivatives providing the gradient. In addition use of the gradient to identify a search direction, the curvature of the function is also applied to predict where the function passes through a minimum along that direction, i.e. to predict where along the gradient the function will change direction. This give rise to the Newton-Raphon method converge faster (if the starting geometry is not too far from the minimum) and more accurate with a tolerance up to 10^{-9} kcal·mol^{-1}·Å^{-1} in comparing to that of 10^{-6} kcal·mol^{-1}·Å^{-1} of the conjugate gradient method. However, for molecule of *N* atoms it required not only the vector of *3N* first derivatives but also Hessian matrix of $(3N)^2$ second derivatives,

which must be inverted. The need to calculate the Hessian matrix, the iteration makes this algorithm computationally expensive and large storage requirement.

From the above description, it can be seen that different minimization algorithms have different advantages and disadvantages. To optimize the minimization procedure it is usually best to combine several algorithms in the minimization scheme. For instance, the steepest descents method should be used for the first 10-100 steps of minimization procedure to obtain configuration close to the local minimum, then the conjugate gradient or the Newton-Raphson method (dending on how large the system) is employed to reach the minimum.

2.5 Molecular dynamics (MD) simulation

A molecular dynamics simulation numerically solves Newton's equations of motion, thus allowing structural fluctuations to be observed with respect to time. Dynamic simulation methods are widely used to obtain information on the time evolution of conformations of proteins and other biological macromolecules [64] and also kinetic and thermodynamic information. Simulations can provide fine detail concerning the motions of individual particles as a function of time. They can be utilized to quantify the properties of a system at a precision and on a time scale that is otherwise inaccessible, and simulation is, therefore, a valuable tool in extending our understanding of model systems. Theoretical consideration of a system additionally allows one to investigate the specific contributions to a property through "computational alchemy", that is, modifying the simulation in a way that is nonphysical but nonetheless allows a model's characteristics to be probed. One particular example is the artificial conversion of the energy function from that representing one system to that of another during a simulation. This is an important technique in free-energy calculations. Thus, molecular dynamics simulations, along with a range of complementary computational approaches, have become valuable tools for investigating the basis of protein structure and function.

2.5.1 Basic theory of molecular dynamics

The molecular dynamics simulations method is based on Newton's second law or the equation of motion, $F=ma$, where F is the force exerted on the

particle, *m* is its mass and *a* is its acceleration. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Newton's equation of motion is given by

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{2.10}$$

Where F_i is the force exerted on the particle *i*, m_i is the mass on the particle *i*, and a_i is its acceleration. The force can also be expressed as the gradient of the potential energy.

$$F_i = \frac{dV}{dr_i} \tag{2.11}$$

Combining these two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{2.12}$$

where *V* is the potential energy of the system. Newton's equation of motion can be related the derivative of the potential energy to the changes in position as a function of time.

In case of property calculations of the system,

$$F_i = m_i a_i = \frac{dv}{dt} = m_i \frac{d^2 r_i}{dt^2} \tag{2.13}$$

and the acceleration is constant,

$$a = \frac{dv}{dt} \tag{2.14}$$

Expression for the velocity after integration is obtained,

$$v = at + v_0 \tag{2.15}$$

and since

$$v = \frac{dr}{dt} \tag{2.16}$$

therefore,

$$r = at + r_0 \tag{2.17}$$

Combining this equation for the velocity, it can be obtained the following relation which gives the value of r at time t as a function of the acceleration, a, the initial position, r_0 and the initial velocity, v_0.

The acceralation is given as the derivative of the potential energy with respect to the position, r,

$$a = -\frac{1}{m}\frac{dE}{dr} \tag{2.18}$$

Therefore, to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic, e.g., the position and the velocities at time zero determine the positions and velocities at all other times, t. The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein. It is possible to calculate the new position of each atom in the system at the end of each time step. In order to propagate a trajectory for the system of particles, the equation (2.18) must be integrated for all particles in the system at each time step of simulation. This is not a trivial task because the motions of all atoms in the system are coupled together and thus analytically solving the equations of motion is impossible. Therefore, these equations must be integrated using numerical methods.

2.5.2 Integration algorithms

Numerous numerical algorithms have been developed for integrating the equations of motion. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion as shown in equation (2.19) – (2.21):

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + \frac{1}{24}c(t)\Delta t^4 .... \quad (2.19)$$

$$v(t+\Delta t) = v(t) + a(t)\Delta t + \frac{1}{2}b(t)\Delta t^2 + \frac{1}{6}c(t)\Delta t^3 .... \quad (2.20)$$

$$a(t+\Delta t) = a(t) + b(t)\Delta t + \frac{1}{2}c(t)\Delta t^2 + ... \quad (2.21)$$

where *r* is the position, *v* is the velocity (the first derivative with respect to time), *a* is the acceleration (the second derivative with respect to time), and etc. Mostly MD software packages, including AMBER, use a variety of methods based on the Verlet algorithm for propagation of atomic coordinates, which are designed to allow velocities to be calculated at each step.

The basic idea of Verlet algorithm [65] is to write two third-order Taylor expansions for the positions *r(t)*, one forward and one backward in time. Calling *v* the velocities, *a* the accelerations, and *b* the third derivatives of *r* with respect to *t*, as following:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + ... \quad (2.22)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - .... \quad (2.23)$$

Adding the two expressions gives

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \quad (2.24)$$

The equation (2.24) is the basic form of the Verlet algorithm. Since the integrating Newton's equation, *a(t)* is just the force divided by the mass and the force is in turn a function of the position r(t),

$$a(t) = \left(\frac{1}{m}\right).....v(r\Delta t) \tag{2.25}$$

A problem of the Verlet algorithm is that velocities are not directly generated. While they are not needed for the time evolution, their knowledge is sometimes necessary. Moreover, they are required to compute the kinetic energy *K*, whose evaluation is necessary to test the conservation of the total energy $E = K+V$. This is one of the most important tests to verify that a MD simulation is proceeding correctly. One could compute the velocities from the positions by using equation (2.26).

$$v(t) \quad = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \tag{2.26}$$

2.5.3 Bond constrained

The bond constrained is the effective techniques to be proposed to increase the efficiency of a computer simulation. Improvements in efficiency are often obtained through freezing the fastest modes of vibration by constraining the bonds to hydrogen atoms to fixed lengths using algorithms such as SHAKE [66]. This algorithm (otherwise known as the constrained Verlet method) is a straightforward modification of the Verlet algorithm to impose constraints on the internal coordinates such as bond lengths and bond angles. The length of the time step is restricted by the requirement that Δt is small compared to the period of the highest frequency motions being simulated. For the biomolecular systems of interest, the highest frequency motions are the bond stretching vibrations, yet these vibrations are generally of minimal interest in the study of biomolecular structure and function. Thus, algorithms, such as SHAKE, that constrain the bonds to their equilibrium lengths are useful. In essence, they may be considered as averaging out the highest frequency vibrations.

2.5.4 Periodic boundary condition

MD simulations are used to investigate on macroscopic behavior and to obtain information that is not easily got from experiments. Nevertheless, computer still cannot model more than a few million atoms at one time, despite the rapid advancement of computer power. In order to model a macroscopic system in terms of a finite simulation, duplication the system periodically in all directions is often employed.

The idea is represent in Figure 2.4. Typically, a cubic lattice is used for replication of the central cubic box. The atoms outside the central box are simply images of the atoms simulated in that box. So-called periodic boundary conditions ensure that all simulated atoms are surrounded by neighboring atoms, whether those neighbors are images or not. The so-called minimum image convention guarantees that duplicate interactions between atoms are not included by calculating only one pair wise interaction for each pair of atoms. For atoms *i* and *j*, the interaction is that between the original atom *i* and whichever copy of atom *j*, original or image, is closest to atom *i*.

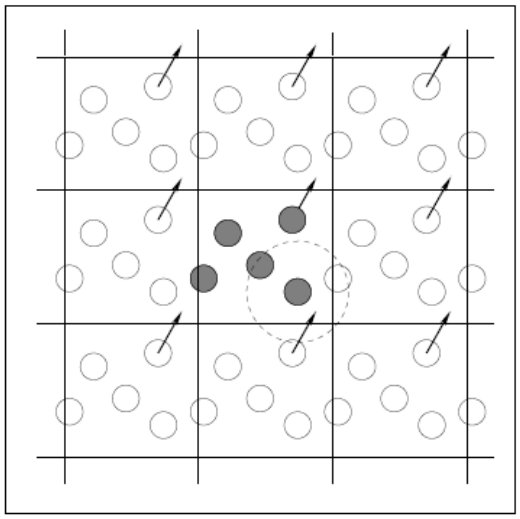


Figure 2.4 Periodic boundary conditions: as a particle moves out of the simulation box, an image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbors are included.

2.5.5 Treatment of long-range interaction

Long-range electrostatic interactions play a dominant role in protein structural stability. In order to improve the treatment of the long-range electrostatic interactions, the Ewald summation method [67] including the particle mesh Ewald (PME) summation method [68], have been implemented in MD packages.

Ewald summation was developed in 1921 to study energetic of ionic crystals. In this method a particle interacts with all the other particles in the simulation box and with all their images in an infinite array of periodic cells. For systems with large periodic boxes and high dielectric solvents, the artifacts observed in simulations with the Ewald summation methods are insubstantial. These methods were demonstrated as being relatively efficient.

PME methods have been developed that approximate the reciprocal space term of the standard Ewald summation by a discrete convolution on an interpolation grid, using the discrete FFT. The advantages of the PME are the method for the treatment of long-range forces in macromolecule systems. The high accuracy can be obtain with relatively little increase in computational cost and efficiently implemented into usual MD algorithms such as AMBER package [62, 68].

2.6 Basic steps in AMBER

MD simulations of proteins, which began about 25 years ago, are by now widely used as tools to investigate structure and dynamics under a variety of conditions; these range from studies of ligand binding and enzyme reaction mechanisms to problems of denaturation and protein refolding to analysis of experimental data and refinement of structures. “AMBER” is the collective name for a suite of programs that allows users to carry out and analyze molecular dynamics simulations, particularly for proteins, nucleic acids and carbohydrates. None of the individual programs carries this name, but the various parts work reasonably well together, providing a powerful framework for many common calculations. The term AMBER sometimes also refers to the empirical force fields. It should be recognized, however, that the code and force fields are separate; several other computer packages

have implemented the AMBER force fields, and other force fields can be used within the Amber programs.

The principal flow of information is three main steps: system preparation, simulation, and trajectory analysis.

2.6.1 Preparatory programs

The main preparation programs are antechamber (which assembles force fields for residues or organic molecules that are not part of the standard libraries) and LEaP (which constructs biopolymers from the component residues, solvates the system, and prepares lists of force field terms and their associated parameters).

2.6.2 Simulation programs

The main MD program is called SANDER. It is the basic energy minimizer and molecular dynamics program. This program relaxes the structure by iteratively moving the atoms down the energy gradient until a sufficiently low average gradient is obtained. The molecular dynamics portion generates configurations of the system by integrating Newtonian equations of motion. MD will sample more configurational space than minimization, and will allow the structure to cross over small potential energy barriers. Configurations may be saved at regular intervals during the simulation for later analysis, and basic free energy calculations using thermodynamic integration may be performed.

2.6.3 Analysis programs

The task of analyzing MD trajectories faces two main programs, ptraj and MM-PBSA. Ptraj is a general purpose utility for analyzing and processing trajectory or coordinate files created from MD simulations (or from various other sources), extractions of coordinates, calculation of bond/angle/dihedral values, atomic positional fluctuations, analysis of hydrogen bonds, analysis of radial distribution function (RDF) etc. MM-PBSA is a script that automates energy analysis of snapshots from a molecular dynamics simulation using ideas generated from continuum solvent models.

2.7 Analysis of MD trajectories

In this part, the trajectories, the production from molecular dynamics simulations, were obtained, and then analyzed in different way. Here, many properties were investigated by module of Amber e.g. root mean square deviation, hydrogen bonding, inter-atomic distances, fluctuations of bond-angle and dihedral angle, RDF in the system, binding free energies and so on.

2.7.1 The root mean square deviations (RMSD)

The RMSD is the measure of the average distances of certain atoms in molecule with respect to a reference structure. This value that measure for equilibration and protein flexibility can be calculated by least-square fitting the structure to the reference.

$$RMSD(t_1, t_2) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i \left| r_i(t_1) - r_i(t_2) \right|^2} \qquad (2.27)$$

where $M = \sum_{i=1}^{N} m_i$ and $r_i(t)$ is the position of atom i at time t.

2.7.2 The hydrogen bonding

Hydrogen bonding is determined between all possible H-bond donor atoms and acceptor atoms based on the following criteria: (i) proton donor-acceptor distance, r, $\leq$ 3.5 Å, and (ii) donor-H-acceptor bond angle,α, $\geq$ 120° (see Figure 2.5).

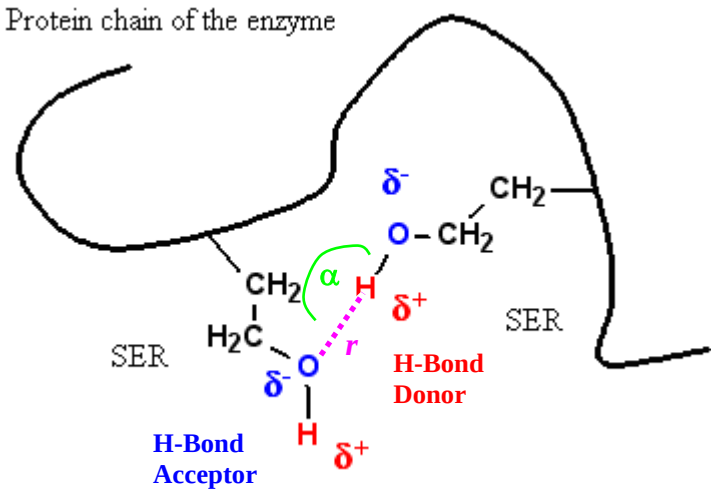


Figure 2.5 H-bond geometry between two molecules**.**

2.7.3 Radial distribution function (RDF)

The RDF, *g(r)*, or the pair correlation function, is a measure to determine the correlation between particles within a system. Specifically, it is a measure of, on average, the probability of finding a particle at a distance of *r* away from a given reference particle, relative to that for an ideal gas. The general algorithm involves determining how many particles are within a distance of *r* and *r*+*dr* away from a particle. Basic Schematic of the RDF shows in Figure 2.6, where the red particle is the reference particle, and blue particles are those which are within the circular shell, dotted in red .

The RDF is usually determined by calculating the distance between all particle pairs and binning them into a histogram. The histogram is then normalized with respect to an ideal gas, where particle histograms are completely uncorrelated. For three dimensions, this normalization is the number density of the system multiplied by the volume of the spherical shell, which mathematically can be expressed as $N_{i.g.}(r) = 4\pi r^2 \rho dr$, where ρ is the number density.

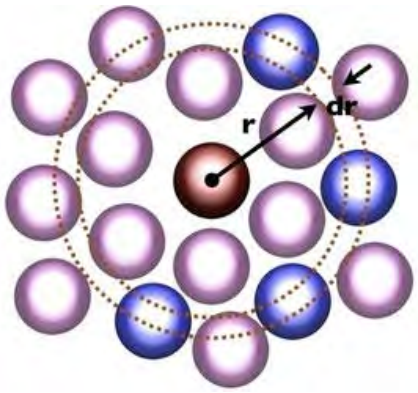


Figure 2.6 Basic Schematic of the RDF [69].

2.7.4 The binding free energy ($\Delta G_{binding}$) calculations

In general, the free energy of the inhibitor binding, $\Delta G_{binding}$, can be obtained from the difference between the free energy of the receptor-ligand complex (G_{cpx}) and the unbound receptor (G_{rec}) and ligand (G_{lig}) as follows:

$$\Delta G_{binding} = G_{cpx} - \left(G_{rec} + G_{lig}\right) \tag{2.28}$$

The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) approach calculates $\Delta G_{binding}$ on the basis of a thermodynamic cycle. Therefore, Eq.2.28 can be approximated as model is method to estimate a binding free-energy change. The MM-PBSA method partitions the free energy into molecular mechanical (MM) energies, continuum solvation energies, and solute entropy terms as follows

$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{PBSA} - T\Delta S_{MM} \tag{2.29}$$

where $\Delta G_{binding}$ is the average standard free energy of the molecule of interest, which can be the ligand, the receptor, or their complex. ΔG_{PBSA} is the molecular solvation free energy. The solute's entropy term may be estimated using a number of methods. The average molecular mechanical energy, ΔE_{MM}, is typically defined as

$$E_{MM} = E_{bond} + E_{angle} + E_{torsion} + E_{vdw} + E_{elec} \quad (2.30)$$

where E_{bond}, E_{angle}, E_{torsion}, E_{vdw}, and E_{elec} are the bond, angle, torsion, van der Waals, and electrostatics terms of molecular energy, respectively.

The molecular solvation free energy can be further decomposed to

$$\Delta G_{PBSA} = \Delta G_{PB} + \Delta G_{SA} \quad (2.31)$$

where ΔG_{PB} was calculated using a continuum solvent model with a Poisson-Boltzmann solution for calculating the electrostatic free energy of solvation. The ΔG_{SA} is estimated from the solvent accessible surface area (SASA) for calculating the nonpolar free energy of solvation.

2.7.4.1 Poisson-Boltzmann (PB) model

Calculations of electrostatic potentials required for evaluating the ionization and interaction energies are based on the PB model of electrostatic interactions. The foundation of the PB model constitutes the fundamental equation of electrostatics, differential Poisson equation that describes the electrostatic potential $\Phi(r)$ in a medium with a dielectric scalar field $\varepsilon(r)$ and with a charge density $\rho(r)$.

$$\nabla \cdot \varepsilon(r) \nabla \Phi(r) = -4\pi\rho(r) \quad (2.32)$$

When one deals with a macromolecule (in this case the protein) immersed in an aqueous medium with mobile ions, the charge density $\rho(r)$ can be separated into two components:

$$\rho(r) = \rho^{f}_{\text{int}} + \rho^{m}_{ext} \quad (2.33)$$

where ρ_{int}^{f} describes the interior charge distribution of the fixed (^{f}) positions of all the charges in the protein and ρ_{ext}^{m} is a mobile (^{m}) exterior charge density modeled by a Boltzmann distribution. Equation 2.32 takes the form

$$\nabla \cdot \varepsilon(r)\nabla\Phi(r) = -4\pi\left[\rho_{\text{int}}^{f}(r) + \lambda(r)\sum_{i} q_i n_i \exp\left(-q_i\Phi(r)/kT\right)\right] \tag{2.34}$$

where n_i is the bulk number density of type-i ions, q_i is the ionic charge, k is the Boltzmann constant, and λ equals 1 for ion-accessible regions and 0 elsewhere.

Linearizing the exponential terms in the mobile charge distribution and assuming equal number of positively and negatively charged ions one obtains

$$\nabla \cdot \varepsilon(r)\nabla\Phi(r) = -4\pi\left[\rho_{\text{int}}^{f}(r) + \lambda(r)\sum_{i} q_i n_i q_i\Phi(r)/kT\right] \tag{2.35}$$

Next, the modified Debye-Hueckel parameter is introduced

$$\bar{k}^{2} = \frac{4\pi\sum_{i} n_i q_i^{2}}{kT} = \frac{8\pi e^{2} N_A I}{1000kT} \tag{2.36}$$

where N_A is the Avogadro constant, e the charge of an electron, and I the ionic strength. The ionic strength is represented as

$$I = \frac{1}{2}\sum_{i} c_i z_i^{2} \tag{2.37}$$

where c_i and z_i are the molar concentrations and the valencies of the mobile ions, respectively.

Introducing $\bar{k}^2$ in equation 2.36, one obtains the linearized Poisson-Boltzman equation where the protein is represented as a low dielectric medium containing fixed charges and the solvent is represented as a medium of

dielectric constant of 80 which contains mobile ions that screen the fixed charges according to the Debye-Hueckel model:

$$\nabla \cdot \varepsilon(r)\nabla\Phi(r) = -4\pi\rho_{\text{int}}(r) + \lambda(r)\bar{k}^2\phi(r) \tag{2.38}$$

The charges are usually located at nucleic positions determined by x-ray crystallographic methods or high-resolution Nuclear Magnetic Resonance spectroscopy.

2.7.4.2 Solvent accessible surface area

Since proteins inside our cells are in an aqueous environment, considering a protein's interactions with solvent molecules, particularly water, is very important for appropriately modeling them. Recall that one of the phenomena that determine the structure of a protein is the hydrophobic effect: some amino acid residues are stabilized by the presence of water. Therefore, quantitative modeling of the strength of interaction with solvent often involves computing the SASA. Computing SASA can be done by regarding each solvent molecule as a sphere of set radius. When this sphere rolls about the molecule, its center delineates the SASA. One can think of the SASA of a molecule as the result of growing each atom sphere by the radius of the solvent sphere. Instead, by taking what is swept out by the front of the solvent sphere, the molecular surface (MS) model of the molecule was obtained (see in Figure 2.7). Alternatively, the MS can be obtained by removing a layer of solvent radius depth from the SASA model.

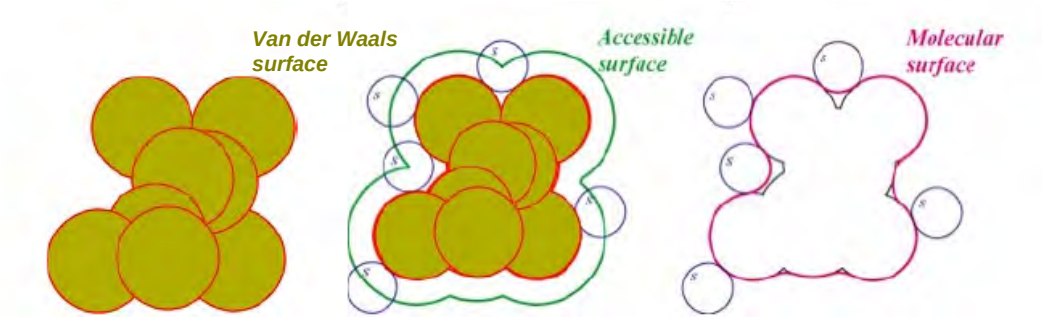


Figure 2.7 Schematic of the SASA [70].

CHAPTER III

MD SIMULATION OF HIV-1/NNRTI COMPLEXES

3.1 Introduction

Since HIV-1, a causative agent of AIDS, was first discovered in 1981 [1], the HIV-1 RT has been the subject of intensive studies. HIV-1 RT is part of the HIV capsid and has an essential role in the replication of the AIDS virus. This enzyme transcribes the single-stranded RNA of the AIDS retrovirus to a double-stranded DNA which can then be integrated into the host genome. Due to its key role in the HIV life cycle, RT is an important target for antiviral agents in the treatment of AIDS. Anti-HIV-1 RT drugs can be divided into two main classes: NRTIs and NNRTIs. NRTIs are competitive to the nucleotide substrates and act as chain terminators when incorporated into viral DNA by HIV-1 RT. However, they instigate serious side effects, especially damages to mitochondria [18-20]. NNRTIs are non-competitive inhibitors which are highly specific for HIV-1 RT at a common allosteric site approximately 10 Å apart from the polymerase active site. In addition, they are less likely to cause adverse side effects by disruption of the normal DNA polymerase activity [19, 20]. Consequentially, interest is currently focused upon the NNRTIs group, aiming to understand and to provide a detailed insight into the HIV-1/NNRTIs interactions.

Residues that have been reported to play an important role in binding to the NNRTI consists approximately of 15 amino acid residues from the p66 subunit, L100, K101, K103, V106, T107, V108, V179, Y181, Y188, V189, G190, W227, W229, L234 and Y318 plus the E138 residue from p51. It appears that these binding pocket residues are flexible to some extent depending on molecular shape and size, the specific chemical structure, and the binding mode of the individual NNRTIs [7, 15, 71]. The NNRTIs considered in this study are EFV, EMV, ETV and NVP. Their structure in the corresponding binding pockets of HIV-1 RT, their chemical composition and numbering of selected atoms are given in Figure 3.1. NVP is classified as the member of first generation NNRTIs. EMV has been found to inhibit HIV strains that have developed resistance against NVP, albeit at a higher concentration than that required to inhibit the wild-type virus, but it failed in clinical

trials [72]. EFV and ETV are classified as second generation inhibitors. EFV is the most prescribed NNRTI use for patients in the first-line antiretroviral therapy because of its potent activity on wild-type HIV-1 [73]. ETV shows potent efficacy and retains

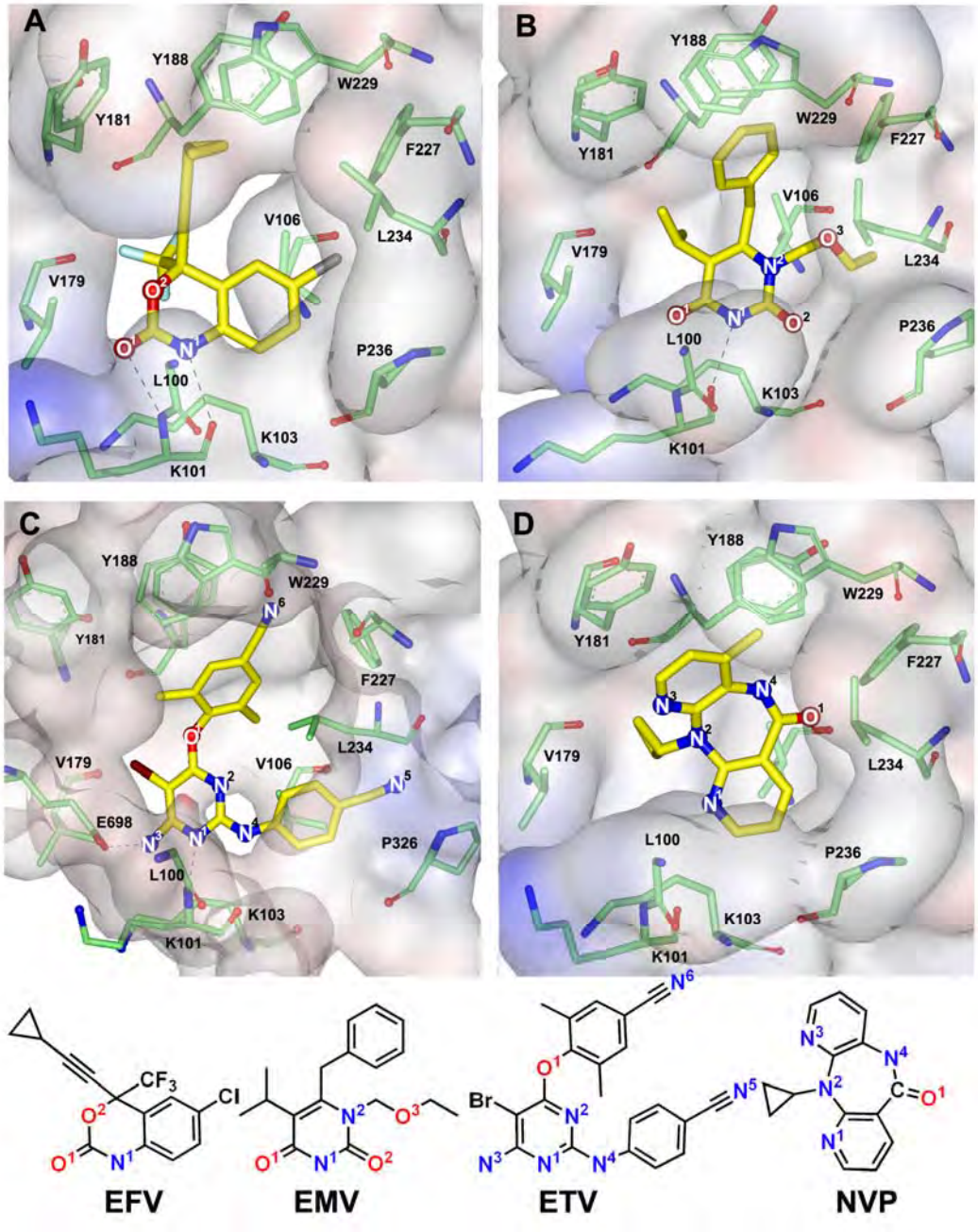


Figure 3.1 Geometries of EFV (A), EMV (B), ETV (C) and NVP (D) corresponding the binding pockets of HIV-1 RT. Chemical structures of NNRTIs and definition of atoms are also given.

potency in patients infected with NNRTI-resistant HIV-1 variants [15, 74]. Experimentally derived IC_{50} values of the wild-type HIV-1 RT treated with NNRTIs were reported in the following order: EFV > ETV > EMV > NVP [15, 75-77]. However, more NNRTIs are needed to overcome drug resistant mutations in the enzyme. Evidently, there exists no final treatment solution against AIDS. NNRTI-resistant mutations cluster between position 98 - 108, 179 - 190, and 225 - 238 in the p66 subunit. K103N is probably the predominant mutation observed in patients receiving NNRTIs. Y188L has also been linked to high-level cross-resistance, mainly induced by NVP. Also ten other residues with selected mutations, L100I, K101P, V106M/A, V108I, Y181C/I/V, G190A/S/C/E/Q, P225H, M230L, P236L and Y318F, have been observed [78].

This work aims to investigate amino acid residues that are responsible for the HIV-1 RT-the NNRTIs interactions by means of molecular dynamics and the free energy calculation methods. Here, we constructed the four model systems consisting of the HIV-1 RT with 1000 residue-long complexed with the four NNRTIs, EFV, EMV, ETV and NVP. The size of the systems for the MD simulations was considerably large compared to what have been studied using the same protein and the same approach [9, 72, 76, 79, 80]. The calculation was focused upon examination of the structure, binding free energy and water accessibilities in the binding site of each of the four HIV-1 RT/NNRTI complexes. In addition, the distribution and binding of water molecules in the cavity of the enzyme-inhibitor complex, which is known to play an important role in the drug binding, was investigated and is discussed with reference to those found in the crystal structures. The information will be helpful for the rational design of new anti-HIV drugs with improved resistance profiles for anti-HIV therapy, as well as prediction of new resistance mutations.

3.2 Computational methods

3.2.1 The initial structures of HIV-1/NNRTI complexes

The initial crystallographic structures of HIV-1 reverse transcriptase complexed with EFV, EMV, ETV and NVP were obtained from the PDB (RCSB PDB, http://www.rcsb.org), entry codes 1FK9 [8], 1RT1 [81], 1SV5 [15] and 1VRT [8], respectively. From the structure of mutation K103N (1SV5), the wild type system

was reconstructed by replacing glutamine by lysine using the LEaP module in the AMBER9 software package [62]. The missing residues of 1FK9, 1RT1 and 1VRT were reconstructed with the help of an other X-ray structure (PDB entry code 1SV5) and subsequent low temperature annealing and energy minimization procedures of the added residues whilst constraining the known X-ray residue positions. All missing hydrogen atoms of the protein were added using the LEaP module in the AMBER9 software package. The ionization states of amino acids with electrically charged side chains were assigned using the PROPKA program [82] (see in Table 3.1)

Table 3.1 The pKa values of selected amino acids of four crystallographic structures of HIV-1 RT obtained from the PROPKA program [82].

Residue/No.	**pKa**			
	1FK9	**1RT1**	**1SV5**	**1VRT**
ASP 76A	2.97	3.23	3.85	3.35
ASP 86A	2.70	4.00	2.20	3.24
ASP 110A	2.31	8.32	3.93	3.46
ASP 113A	3.87	3.91	1.30	3.28
ASP 121A	3.43	2.75	3.22	3.56
ASP 123A	3.80	3.80	3.45	3.29
ASP 177A	3.06	3.21	3.87	3.59
ASP 185A	5.06	4.84	5.06	4.26
ASP 186A	7.69	2.70	4.36	5.65
ASP 192A	3.81	3.63	3.57	3.42
ASP 218A	3.04	3.22	5.89	3.04
GLU 79A	3.94	3.98	4.00	3.58
GLU 122A	4.78	3.97	3.84	5.35
GLU 138A	4.32	4.01	4.50	4.50
GLU 169A	4.57	3.84	4.64	4.50
GLU 194A	2.88	4.49	3.76	3.52
GLU 203A	3.51	4.00	4.78	4.09
GLU 204A	4.57	4.30	3.91	3.91
GLU 224A	3.98	4.64	4.50	4.71
GLU 233A	4.57	4.08	4.57	4.11
GLU 248A	4.12	3.71	4.50	4.78
GLU 291A	4.01	3.89	4.64	3.79
GLU 89A	4.92	5.37	4.83	4.82

Residue/No.	pKa			
	1FK9	1RT1	1SV5	1VRT
GLU 297A	4.50	4.50	4.50	4.50
GLU 298A	4.38	3.39	4.64	4.58
GLU 138B	3.70	6.42	5.54	3.09
HIS 96A	0.45	0.35	0.12	1.54
HIS 198A	3.09	3.70	3.39	3.65
HIS 208A	7.36	6.43	3.93	5.32
HIS 221A	6.43	6.50	6.29	6.66
HIS 235A	6.43	6.29	6.15	7.03
CYS 38A	10.05	10.02	7.63	8.69
TYR 56A	12.82	12.24	12.41	12.22
TYR 115A	12.20	12.58	13.25	11.00
TYR 127A	10.83	10.46	10.90	10.69
TYR 144A	12.94	13.68	12.34	13.24
TYR 146A	11.87	11.59	11.43	12.13
TYR 181A	12.10	11.54	11.41	11.26
TYR 183A	10.14	10.00	10.00	10.14
TYR 188A	13.66	14.18	14.62	12.42
TYR 232A	11.18	11.74	11.09	11.00
TYR 271A	9.38	9.20	9.62	9.48
TYR 318A	11.77	12.30	11.16	11.17
TYR 319A	11.98	11.05	11.29	11.78
LYS 73A	7.78	7.88	9.87	9.07
LYS 82A	10.29	10.29	10.50	10.29
LYS 101A	10.93	10.89	12.29	11.02
LYS 104A	10.50	10.50	10.50	10.01
LYS 126A	10.22	10.15	10.08	10.01
LYS 154A	9.32	9.49	9.39	9.48
LYS 166A	9.77	9.79	10.18	10.37
LYS 173A	10.50	9.80	9.06	10.50
LYS 201A	10.22	10.36	10.50	10.08
LYS 219A	10.29	10.50	10.22	10.50
LYS 223A	10.36	10.29	10.15	10.36
LYS 238A	10.50	9.73	10.15	10.50
LYS 249A	9.59	10.22	10.50	9.87
LYS 259A	9.80	10.50	10.29	10.22
LYS 263A	10.43	10.50	10.22	10.43
LYS 275A	10.43	10.36	10.36	10.29
LYS 281A	10.22	10.43	10.22	10.43
ARG 72A	11.94	12.01	12.22	12.15
ARG 78A	12.29	12.29	12.29	12.29
LYS 220A	10.43	10.50	10.50	10.36

Residue/No.	pKa			
	1FK9	1RT1	1SV5	1VRT
ARG 83A	11.73	11.59	11.73	11.73
ARG 125A	11.29	10.87	11.06	11.52
ARG 143A	11.41	11.30	11.31	11.11
ARG 199A	12.50	12.22	12.50	12.36
ARG 206A	11.31	11.80	11.94	11.59
ARG 211A	11.80	12.15	12.50	12.50
ARG 277A	12.29	12.43	11.49	12.36
ARG 284A	12.22	12.08	12.08	12.15
ARG 307A	11.31	11.24	11.59	11.45

3.2.2 Force field parameters for the inhibitors

Starting structures and force field parameters for the four inhibitors were obtained as follows. Ligand geometries were optimized using Gaussian03 [83] with the HF/6-31G* method. Then single-point calculations were carried out to compute the electrostatic potentials around each compound using the same basis set and level of theory. The electrostatic potential were calculated by RESP [84]. Partial charge generation and assignment of the force field [57] were performed using the Antechamber suite [85].

3.2.3 Molecular dynamics simulations

Energy minimization and MD simulations were performed using the SANDER module of AMBER9 software package. An all-atom representation of the system was used employing the *ff03* [86, 87] force field to assign parameters for the standard amino acids. To incorporate the solvent and counter ions under consideration, each system was solvated using the TIP3P water model [88] and shown in Figure3.2. Neutralization was performed by the counter ions using the LEaP module. The total numbers of atoms are 142662, 141659, 141220 and 143341 for the RT/EFV, RT/EMV, RT/ETV and RT/NVP systems, respectively. The periodic boundary condition with the NPT ensemble was applied. The simulation steps consisted of thermalization, equilibration and production phases. Initially, the temperature of the system was gradually increased from 0 K to 298 K during the first

50 ps. Then, the system was maintained at 298 K until the MD simulations reach 1.5 ns. Finally, the production phases hold up from 1.5 ns to 3.0 ns. A Berendsen coupling time of 0.2 ps was used to maintain the temperature and standard pressure of the system [89]. The SHAKE algorithm [66] was applied to constrain all bonds involving hydrogens, and the simulation time step of 2 fs was used. All MD simulations were run with a 10 Å residue-based cutoff for non-bonded interactions, and the PME method was used for an adequate treatment of long-range electrostatic interactions [68]. The convergence of energies, temperature, pressure and RMSD was monitored to verify the stability of the systems. After reaching an almost stable RMSD value, the production phase was selected from 1.5 ns to 3.0 ns for the four systems.

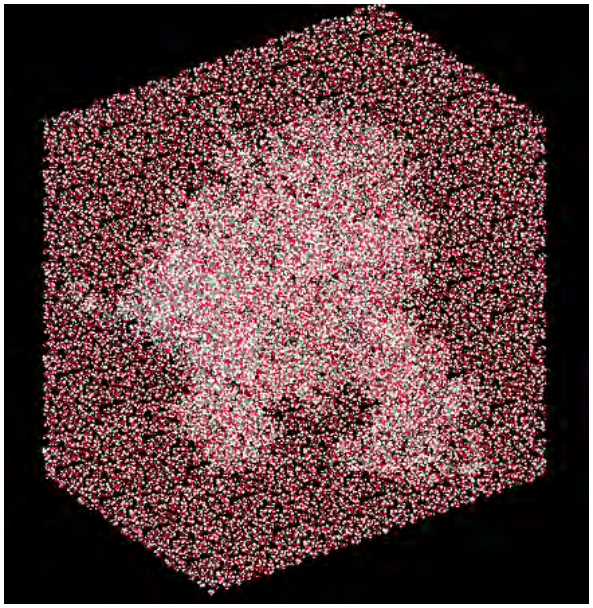


Figure 3.2 The complex of HIV-1 RT/inhibitor in solvation box.

3.2.4 Free energy calculation

The binding free energies ($\Delta G_{binding}$) of the four systems were estimated by the MM-PBSA approach [9, 90]. In this method, the $\Delta G_{binding}$ value can be calculated according to: *(i)* the molecular mechanics interaction energy of protein-inhibitor complex in vacuum, composed of the electrostatic interaction and the van der Waals interaction *(ii)* the contribution of solute entropy to binding and *(iii)* the

solvation free energy, composed of the polar and the nonpolar contribution. In our calculations, the solvation energy was calculated using the dielectric constant of 1.0 for the protein interior, and 80.0 for the water molecule. An extensive determination of the entropic effect from normal mode analysis requires high computational demands, and thus, is costly. We could assume that the entropic differences among the four systems, RT/EFV, RT/EMV, RT/ETV and RT/NVP, should be very small, because they are very similar. Moreover, a number of previous studies [91-93] have shown that the entropic change is minor in many protein complex systems. Therefore, the contribution of the entropy *(TΔS)* was not included in the present study. One hundred MD snapshots from the data collection period in MD trajectory were extracted during the calculation, every 75 ps from 1500 to 3000 ps. The electrostatic solvent energies were calculated using the pbsa program [94] while the nonpolar contribution of the solvent was determined using the molsurf program [95].

3.3 Results and discussion

3.3.1 Overall enzyme-inhibitor Structure

To monitor the stability of the four simulated systems, RT-EFV, RT-EMV, RT-ETV and RT-NVP, RMSD of the structures obtained during the 3ns MD simulations relative to the initial structure (heavy atoms only) in the enzyme-inhibitor complex and the inhibitors were evaluated and plotted in Figure 3.3. It can be seen from the plots that RMSD of complexes were larger than that of inhibitors. RMSD of the three complexes were stabilized at ~ 3.5 Å after 1.5 ns, whereas RMSD of the inhibitors stabilized at ~ 1.0 Å. Conformation of EFV and ETV does not significantly change when compared with the initial x-ray structure while EMV shows small fluctuation NVP changes to another conformation at 1.5 ns and returns back to initial conformation at 2.0 ns. Therefore, on analyze to the RMSD plot, the production phase was selected from 1.5 ns to 3.0 ns for the three systems.

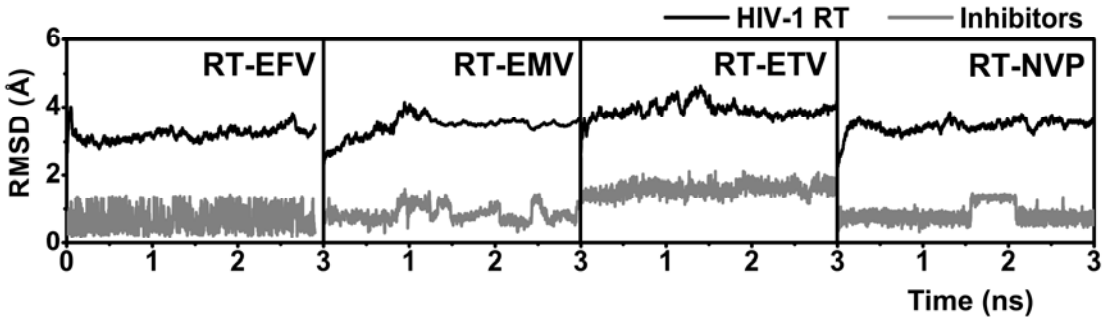


Figure 3.3 RMSDs relative to the initial structure of the enzyme-inhibitor complexes (black) RT-EFV, RT-EMV, RT-ETV and RT-NVP and their corresponding inhibitors, EFV, EMV, ETV and NVP (grey).

3.3.2 Hydrogen bonds between inhibitors and the enzyme

To analyze the ligand-enzyme interaction, the occurrences of hydrogen bonds between the inhibitors and the binding pocket residues with percentage of occupations were determined based on the following criteria: *(i)* proton donor-acceptor distance ≤ 3.5 Å, and *(ii)* donor-H-acceptor bond angle ≥ 120°. The analysis was carried out on the trajectories after equilibration. The results are given in Table 3.2.

As can be seen from the table, the hydrogen bond patterns for EFV and EMV are slightly different. For EFV, two strong and one weak hydrogen bond, with the occupation of 100%, 73% and 13%, were detected with K101 only, no interaction with K103 could be found. For EMV, the ligand binds strongly to the binding pocket via K101 with 100% hydrogen bond occupation. Two weaker hydrogen bonds with 10% and 4% occupation only were also observed via K103.

In the case of ETV, one hydrogen bond is formed between ETV and the backbone carbonyl of K101 but with only 5% occupation. An additional hydrogen bond is observed between the nitrogen of the inhibitor and the carboxylate group of E698 (E138 in p51 subunit) with 24% occupation. Such an interaction with E698 is unusual for HIV-1 RT/NNRTIs complexes, and could be found only in some special cases [96].

For the NVP, no hydrogen bond formation within the binding pocket of RT was found. This observation is in agreement with the X-ray structure of the HIV-1 RT complexed with NVP [81], and suggestions of Kohlstaedt *et al*. [6] and

Das *et al*.[15], who stated that NVP generally contains aromatic rings and forms π-π interactions with aromatic amino acid residues (Y181, Y188, F227, W229, and Y318) in the HIV-1 RT binding pocket without any hydrogen bonds.

Taking into account all the hydrogen bond data given above, EFV and EMV bind much more tightly into the catalytic site of the HIV-1 RT than ETV while for NVP no hydrogen bond interaction was found. These hydrogen bond patterns seem to correspond with the binding affinities, and subsequently, with the IC_{50} values of the drugs. However, the HIV-1 RT binding pocket of NNRTIs is a hydrophobic cavity and, consequently, other types of interactions need to be considered too.

Table 3.2 Percentage of occupations for hydrogen bonds between amino acid residues in the binding pocket and the inhibitors in the four simulated systems, RT-EFV, RT-EMV, RT-ETV and RT-NVP (see definition of NNRTIs'atoms in Figure3.1).

NNRTI	Type		%Occupation
EFV	N1-H…O=C	K101	100
	N1…H-N	K101	13
	O1…H-N	K101	73
EMV	N1-H…O=C	K101	100
	O1…H-NZ	K103	10
	O1…H-NZ	K103	4
ETV	N1-H…O=C	K101	5
	N3-H…OE1	E698	24
NVP	-		-

3.3.3 Which RT residues are important for binding?

Interaction energies between inhibitors and individual amino acid residues can be calculated using the decomposition energy module of Amber 9, in order to identify residues which are important for the binding affinities. The plot of the decomposition energies (DC) of those amino acids of RT, which are located within 5 Å of the binding pocket, is shown in Figure 3.4. Their interaction energies

vary from -4 to -1 kcal·mol^{-1} for all systems. It is interesting to note that L100, K101, K103, V106, Y181, Y188, F227, W229 and L234 show the largest contributions to the HIV-1 RT/NNRTIs interaction energies. These observations support clinical data [78, 97], which found mutations to occur on these residues mainly.

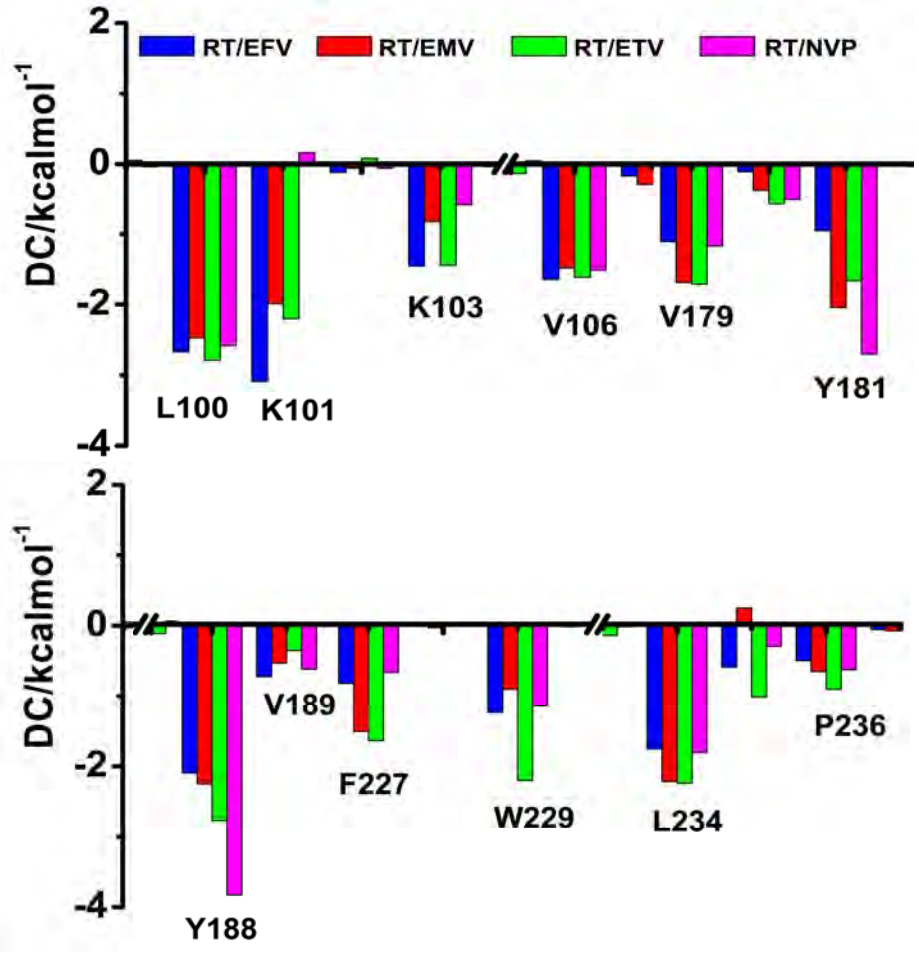


Figure 3.4 Per residue interaction energies of the HIV-1 RT to EFV, EMV, ETV and NVP with those residues which contribute most to the inhibitor-surrounding.

As can be seen in Figure 3.4, the interaction energies of RT concerning residues L100, V106, V179, V189, L234 and P236 with the four inhibitors are not significantly different. The amino acids of HIV-1 RT with the highest interaction energies for EFV are K101 and K103. In the RT/NVP system, the interaction energies with the Y181 and Y188 are the highest compared with the other three inhibitors. The

calculated interaction energies of K101, K103, Y181 and Y188 are comparable to those found experimentally in which the K103N mutation affects highly resistance to EFV by reducing the rate of inhibitor entry [27] and mutations of Y181C and Y188L amino acid residues cause highly resistance to NVP by loss of favorable aromatic ring interactions [15]. Therefore, the data lead us to conclude that the importance of these residues on the interaction energies is a primary source of mutation of NNRTIs.

3.3.4 HIV-1 RT/inhibitors binding energy

The total binding free energies ($\Delta G_{binding}$) averaged over the trajectories between 1.5 and 3.0 ns were calculated for RT/EFV, RT/EMV, RT/ETV and RT/NVP complexes using MM-PBSA. The free energies and decomposed free energies are shown in Table 3.3.

The van der Waals interactions, part of the ΔE_{solute}, appear to be the major contribution to the -HIV-1 RT/inhibitors binding (~80%) for all complexes in accordance with the fact that the binding pocket of the HIV-1 RT is considerably hydrophobic. Among the four complexes, the order of ΔE_{vdw} interactions in the gas phase (solute) are RT/ETV ~ RT/EMV > RT/EFV ~ RT/NVP, while those for ΔE_{ele} interactions are RT/ETV > RT/EFV > RT/EMV > RT/NVP. However, the binding free energies ($\Delta G_{binding}$) of RT with EFV and ETV are in about the same magnitudes but larger than those of other two inhibitors, EMV and NVP. The solvation free energy term (ΔG_{sol}) is the most important term for total binding free energy. In cases of EMV and ETV the gas phase interaction energies are somewhat larger than those of EFV and NVP, which is consequential to some extent by the overall free solvent energies of these molecule (including binding pocket and complexes). This is in agreement with the solvation data (section 3.4.5) where more water molecules were found in the neighborhood of ETV and EMV than that of EFV. The absolute values of the predicted binding free energies are consistent with experimental IC_{50} as shown in Table 3.3, however, the values depend on the method used [15, 75-77].

Table 3.3 Calculated binding free energy and its components (kcal·mol^{-1}) as well as the experimental IC_{50} (in μM) of the NNRTIs complex with RT.

	RT-EFV	**RT-EMV**	**RT-ETV**	**RT-NVP**
ΔE_{ele}	-10.98 ± 2.59	-7.65 ± 2.06	-17.54 ± 5.66	-5.71 ± 1.92
ΔE_{vdw}	-42.35 ± 1.80	-49.74 ± 2.70	-50.32 ± 2.88	-42.14 ± 1.98
ΔE_{solute}	-53.33 ± 2.54	-57.39 ± 3.25	-67.85 ± 5.02	-47.85 ± 3.23
$\Delta G^{nonpolar}$	-3.60 ± 0.14	-4.11 ± 0.21	-6.47 ± 0.20	-3.59 ± 0.25
ΔG^{ele}	25.07 ± 2.14	32.66 ± 3.03	43.25 ± 3.37	25.46 ± 2.59
ΔG_{sol}	21.47 ± 2.12	28.55 ± 3.05	36.78 ± 3.33	21.87 ± 2.61
$\Delta G_{binding}$	-31.86 ± 2.37	-28.83 ± 2.81	-31.07 ± 4.45	-25.99 ± 3.72
IC_{50}	0.001 (15) 0.001 (72) 0.004 (74)	0.004 (73)	0.002 (15) 0.0014 (72) 0.029 (74)	0.085 (15) 0.76 (72) 0.039 (74) 0.034 (73)

3.4.5 Water molecules in the cavity of HIV-1 RT

The distribution of water molecules in the cavity of enzyme-inhibitor complex is known to play an essential role in the drug binding. To analyze this information, RDFs, $g_{xy}(r)$—the probability of finding the particle of type y in a spherical radii r, around the particle of type x—were evaluated. Here, the RDFs for donor and acceptor atoms of the four inhibitors (EFV: N^1, O^1-O^2, EMV: N^1-N^2, O^1-O^3, ETV: N^1-N^5, O^1 and NVP: N^1-N^4, O^1 (see labels in Figure 3.1), to the oxygen atom of the water molecule were calculated. The results as well as the running coordination numbers integrated up to the first minimum (marked by an arrow) of the corresponding RDF and chemical structures of each inhibitors together with definition of atoms were summarized in Figure 3.5 and Figure 3.6.

To estimate the total number of water molecules in the pocket of the HIV-1 RT, any oxygen atom of water molecules lying within the spherical radius of the first minimum of the donor atoms of NNRTIs were counted. Four inhibitors were increasingly solvated in order of ETV > EMV > NVP ~ EFV (Figure 3.5), considered

in terms of both coordination number and distances to solvent molecules. To visualize the above-mentioned hydration, the snapshots which accumulate water molecules lying under the first peak ~ 4 Å of the RDFs were displayed in Figure 3.6.

For EFV, the RDF shows a sharp and narrow first peak at ~ 2.9 Å with the corresponding coordination number (CN) of 1 indicating that O^1 is solvated by one water molecule. In terms of hydrogen bonding (O-O distance 3.0 Å), this accounts for 100% occupation in good agreement with the X-ray structure (PDB: 1FK9). The solvation of N^1 atom can be accessed by 0.5 water molecules with the distance of ~ 4.5 Å (Figure 3.6) suggesting that this is the same water found at the O^1 site which swinging back and forth between the nearby binding site. While the small broad peak which is pronounced at 4.5 Å with the CNs of 0.5 water molecule was found for O^2. This indicates a weak hydrogen bond, around 30% occupation, between O^2 of EFV and the water molecule.

For the RDFs of EMV, two water molecules were found in the binding pocket in agreement with the X-ray structure (PDB: 1RT1). One can observe the 1^{st} sharp peak of O^1 and broad peak of N^1 which is pronounced at 3 Å and 4 Å with the CNs of 1 and 0.5 water molecules, respectively. For O^2, N^2 and O^3, the first peaks were observed at 3.0, 4.5 and 4.5 Å with CNs of 1, 0.5 and 0.5, respectively. Owing to detailed analysis of the simulated trajectories, the various coordination numbers are resulted from one water molecule which located in the vicinity of the side chain of the inhibitor.

For ETV, solvation of N^3 can be accessed by 2.5 water molecules (the configuration is shown in Figure 3.6) at distance ~3.0 Å (Figure3.5). However, the minimum of the first sharp peak above zero demonstrates solvent exchange between the first water and another water with N^3. A small broad peak which is pronounced at 4.0 Å with the CNs of 0.5 results from the same water molecule found at N^3. A small sharp peak pronounced at 3.0 Å with the CNs of 0.5 water molecule was found for N^5 which indicates a strong hydrogen bond between N^5 of ETV and the water molecule. However, this hydrogen bond is not seen during the whole simulation time of the MD simulation. No water molecule was detected around the O^1, N^2, N^4 and N^6 atoms.

For NPV, RDFs of N^1 and N^3 donor atom show the first sharp peaks at 3.0 Å, with CN of 1 which indicates that N^1 and N^3 are solvated by one water

molecule (this is the same as those found in X-ray; 1VRT). The first broad peak of N^2 RDF observed at 4.0 Å with CN of 1 indicates that N^2 is solvated by the same water molecule at as N^3. In contrast, N^4 and O^1 were observed to be free from solvation.

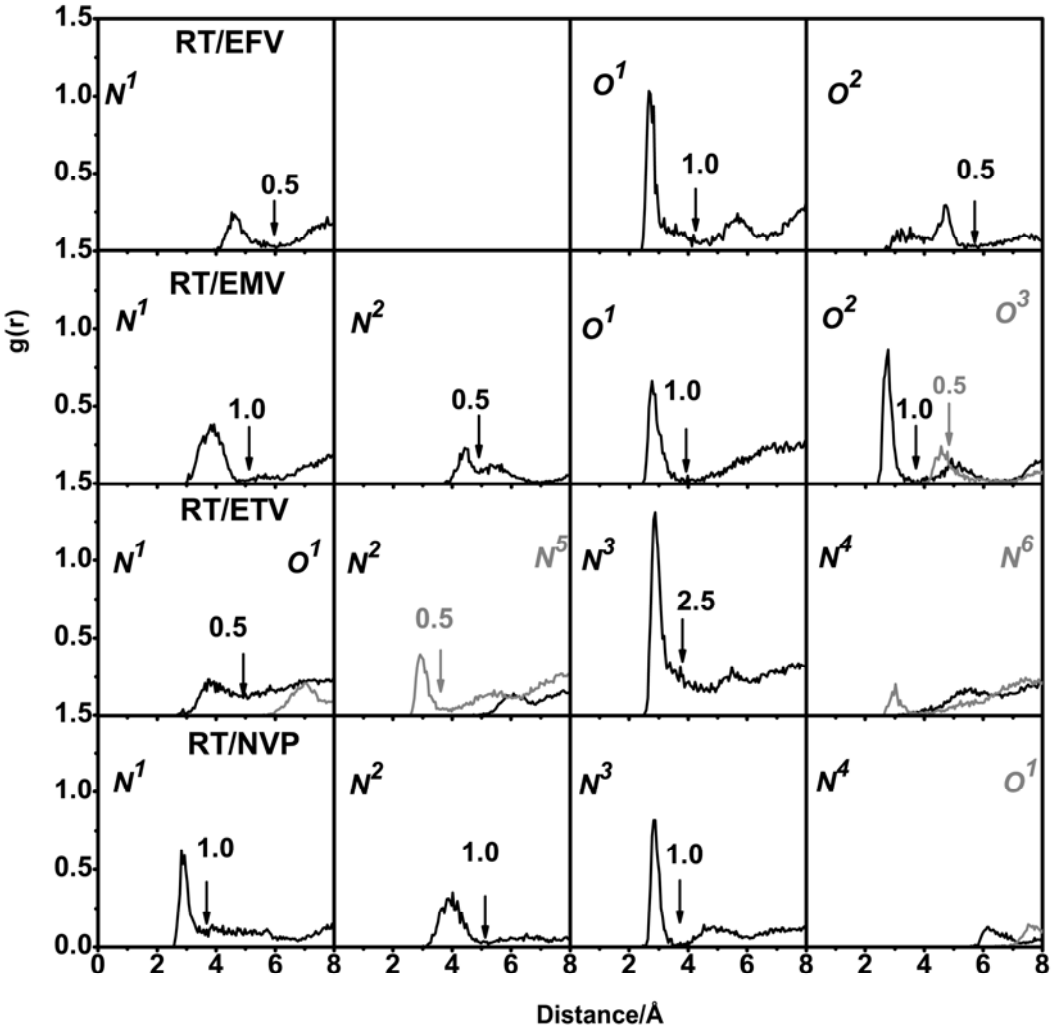


Figure 3.5 RDFs, *g(r)*, centered on the inhibitor atoms to the oxygen atoms of modeled water of the four complexes, RT/EFV, RT/EMV, RT/ETV and RT/NVP, in which chemical structures of each inhibitors together with definition of atoms and running integration number up to the first minimum (marked as arrow) were also given.

The RDFs calculation is in good agreement with the X-ray structure of EFV, EMV and NVP, whereas the RDF of ETV could not be compared because its X-ray structure does not include water coordinates. Taking into account all the data and the discussion given above, the EFV atoms can be much less accessed by water

molecules than the other three inhibitors, considered in terms of both coordination number and distance to solvent molecules. This implies that there is less space available between the EFV atoms and the enzyme residues. This conclusion was strongly supported by the hydrogen bond formation (Table 3.2) and enzyme/inhibitor binding energies (Table 3.3), *i.e*., among the four inhibitors, the main contribution of EFV was exhibited in terms of number as well as percentage occupation of hydrogen bond shown in Table 3.3.

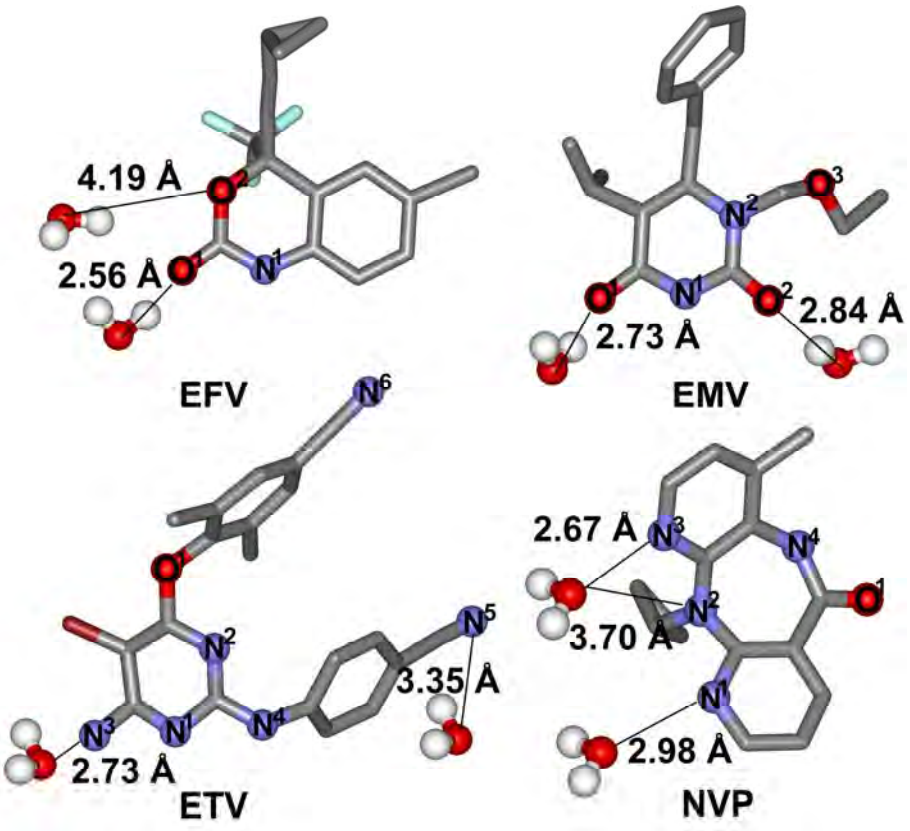


Figure 3.6 Snapshot of the MD simulation showing the surrounding water molecules of the first hydration shell of the four NNRTIs in the HIV-1 RT binding pocket and the distance between selected inhibitor atoms and oxygen atom of water are also given.

CHAPTER IV

MD SIMULATION OF HPH5–FR, LPH5–FR and LPH3–FR COMPLEXES

4.1 Introduction

Proteolytic activation of the HA is essential for viral infectivity and for spread of the avian influenza virus through the host's body [29, 34, 35]. This process is determined by a cleavage reaction at HA's cleavage site, a conserved arginine, by host proteases [33, 36, 37]. Insertion of the –RRRKK– residues into the LPAI viruses cleavage site is known to potentially activate infectivity of viruses, *i.e.,* the LPAI viruses, which then become HPAI viruses, allowing highly virulent strains to be cleaved by furin, an ubiquitously expressed protease. Understanding this fact, why furin cleaves the inserted HA strains better than non-inserted strains, is the goal of this study. The HA is initially synthesized as an inactive precursor (HA0) which is then activated through a controlled proteolytic cleavage by protease into HA1 and HA2 subunits. HA1 mediates virus binding to host-cell receptors while HA2 promotes the release of the viral RNA complexed with polymerase through membrane fusion [29, 33-37]. Without proteolysis, the fusion peptide cannot occur and therefore the virus is essentially non-infectious.

The HA cleavage site relies on the presence of basic amino acids and relates directly to influenza virus pathogenicity, *i.e*., the LPAI virsuses, possess a single arginine at the cleavage site, while the HPAI viruses, the polybasic insertion upstream of the cleavage site of the H5 subtype, which have potential to cause devastating pandemics in the future [34, 38]. In addition, in vitro cleavage of a series of peptide substrates demonstrated that the insertion of the –RRRKK– residues at cleavage site leads to an increase in cleavage by furin protease [35, 39, 42].

Furin (FR), a subtilisin-like serine endoprotease, appears to be a highly specific enzyme, cleaving pro-protein precursors at specific consensus sequence –RXK/RR–, usually to produce biologically active products [40, 43, 98-101]. Based on the X-ray structure of mouse furin complexed with the dec-RVKR-cmk inhibitor, the

furin's catalytic ability is considered to be originated from the catalytic triad residues (S368, H194 and D153) and the oxyanion hole (formed by the carboxamide nitrogens of N295 and backbone nitrogen of S368 to carbonyl oxygen of the inhibitor/substrate's centered arginine) at the active site [43]. The proposed cleavage mechanism is shown in Figure 4.1 in chapter 1[48, 102].

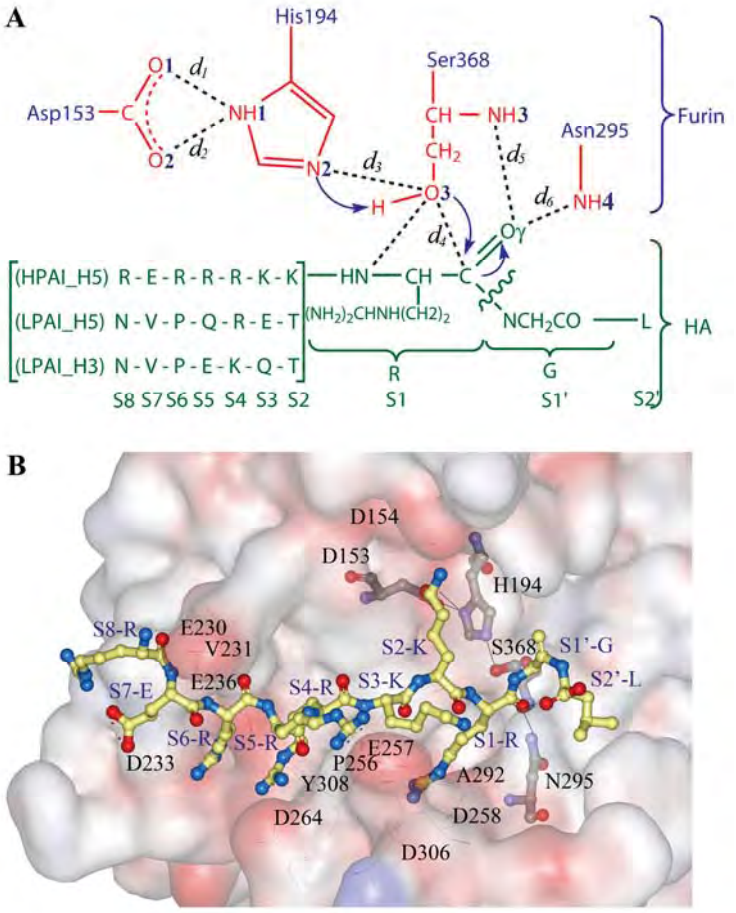


Figure 4.1 (A) Proposed cleavage mechanism of HA by furin and definitions of d_1-d_6. **(B)** Loop of HPAI H5 (ball and stick model) in the electrostatic surface of furin. Blue and red represent positively and negatively charged amino acid residues, respectively.

This work aims to seek for source of high pathogenicity in avain influenza A virus subtype H5 in comparison with LPAI subtypes H5 and H3. Therefore, three MD simulations were carried out for the three complexes, HPH5–FR, LPH5–FR and LPH3–FR. Investigation was focused to intra- and intermolecular interactions and

geometries of the substrate-furin complex potentially involved to the cleavage mechanism.

4.2 Computational methods

4.2.1 The initial structures of individual protein and substrates

The crystal structure of furin with bound dec-RVKR-cmk inhibitor [43] used as receptor model for MD simulations was obtained from PDB, code: 1P8J. This study covers the loop of HA substrate features of subtypes H5 (both HPAI and LPAI) and H3 (known as LPAI). The sequence of the HPAI subtype H5 isolated during the 2006 influenza outbreaks in Thailand was taken from Genbank LOCUS ID ABK13784 (A/chicken/Thailand/PC-168/2006(H5N1)) [103]. While sequence of the LPAI subtype H5 was obtained from PDB entry code: 1JSM [104].

The initial model for the HPAI_H5 loop (RERRRKKRGL) was built up using the sequence alignments and the atomic coordinates of the X-ray structure (residues 322-331: NVPEKQTQGL) of the HA0 of H3 (PDB entry code: 1HA0) [41] and dec-RVKR-cmk inhibitor of furin [43] as a template, thus, it has 50% sequence identity with its templates. With this module, the H3 loop above was mutated at residue 329 from glutamine (Q) to arginine (R). In homology modeling performed by using the homology module in the Studio Discovery program [105]. The final sequence alignments show in Figure 4.2.

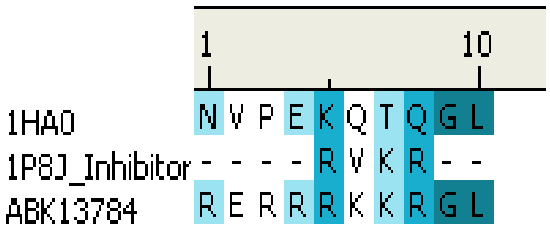


Figure 4.2 Sequence alignment of HPH5 (A/chicken/Thailand/PC-168/2006) ID BK13784 with the dec-RVKR-cmk inhibitor of furin (PDB code 1P8J) and the HA0 of H3 (PDB code 1HA0).

The Needleman-Wunsch algorithm [106] was used for pairwise alignment to identify the correspondence region. The coordinates of the SCRs—in which amino acid sequences in the template and the model are identical—of the template were copied to be those of the model. In addition to the similar SCRs, only the backbone coordinates were replicated and the sidechain atoms were added in a manner that preserves the model protein's residue type. The structure of HPAI_H5 loop modeled according to the multiple templates is displayed in Figure 4.3. Then, the geometry and stereochemistry of the loop were validated by using Ramachandran plot 2.0 program [53, 107]. Ten residues of cleavage loop are located in the fully allowed (90%) and the additionally allowed (10%) regions of the plot (see in Figure 4.4) indicating that the initial structure of HPAI_H5 loop is reasonable.

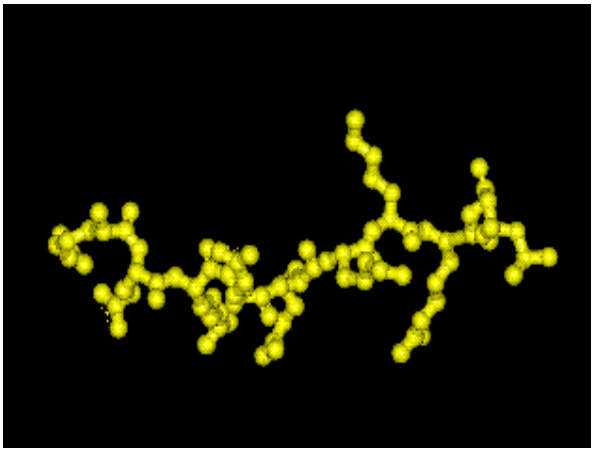


Figure 4.3 3D structure of HPAI_H5 loop modeled according to the multiple templates.

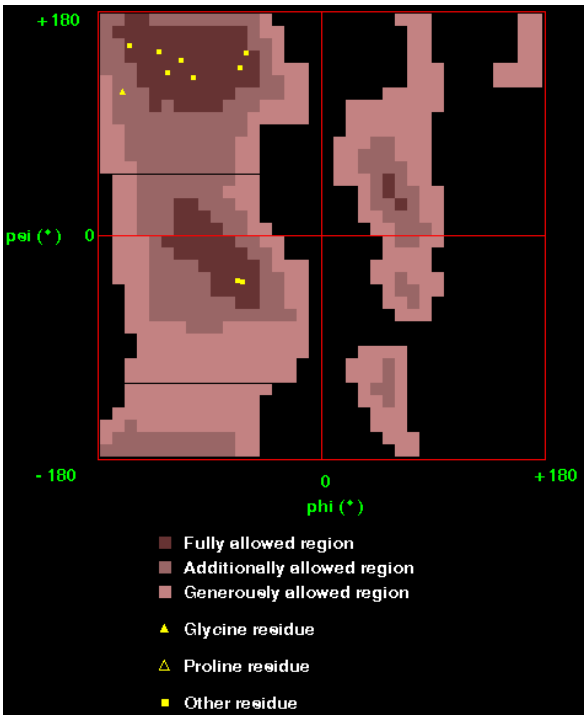


Figure 4.4 A Ramachandran plot for the modeled HPAI_H5 loop**.**

For LPAI subtype H5, the initial structure of the cleavage loop (NVPQRETRGL) [104] was constructed using the backbone atoms of the HPH5 loop built previously and then the side chains was generated by the LEaP module of AMBER 7 [61]. With this module, the H3 loop above was mutated at residue 329 from Q to R.

4.2.2 The initial structure of enzyme-substrate complexes

The HA's cleavage loop complexed with furin was generated by the following procedures (see Figure 4.1): *(i)* All HA heavy atoms of S1-S4 were superimposed with the crystal structure of the dec-RVKR-cmk inhibitor while the HA backbone atoms of S5-S8 and S1'-S2' were superimposed with the HA0 loop of H3. The coordinates of S1-S4 residues were changed according to the -RVKR- coordinates of inhibitor leading to a newly formed conformation of the HA loop.; *(ii)* Then, this loop was manually placed into the furin active site by superimposition between the -R_KR- residues of the HA loop and the dec-RVKR-cmk with a creation

of the furin-substrate complex (Figure 4.5).; *(iii)* The complex was minimized by keeping the S1-S4 residues and the furin fixed.; and *(iv)* Finally, three steps of the restrained MD simulations at 298 K were carried out for relaxing the modeled systems with the restrain factors of 10, 5 and 2.5 kcal·mol^{-1}-Å^{2} for 200 ps, 200 ps and 200 ps, respectively. Therefore, the conformation of each cleavage loop was adapted (from the initial model) to fit better with the furin cavity. The last snapshot obtained from restrained MD procedure was used as the starting structure of the substrate-enzyme complex for the next MD simulations with all atoms allowed to move freely for 2 ns.

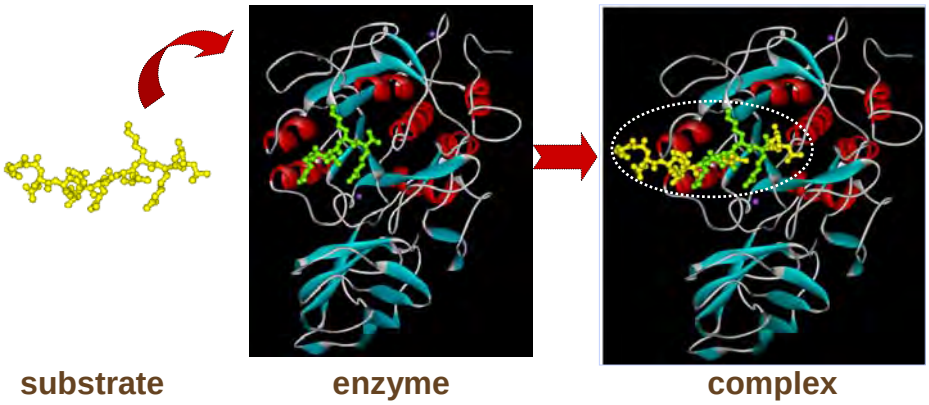


Figure 4.5 The step of initial structure of enzyme-substrate complexes was generated by superimposition.

4.2.3 Molecular dynamics simulations

Three MD simulations for the HA's cleavage loop complexed with furin, HPH5–FR, LPH5–FR and LPH3–FR, were performed where their initial structures were generated as described above. The simulated systems were neutralized by 6, 11 and 11 Na^{+} ions and solvated by TIP3P water molecules leading to total atoms of 54500, 56663 and 56667 for the HPH5–FR, LPH5–FR and LPH3–FR complexes, respectively (Figure 4.6). The dimensions of the obtained simulation boxes for the three systems are 86 Å x 90 Å x 88 Å. The PBC with the NPT ensemble was employed. Energy minimizations and MD simulations were carried out using the SANDER module of AMBER 7 [61] with the Cornell force field [57]. A Berendsen

coupling time of 0.2 ps was used to maintain the temperature and standard pressure of the system [89]. The SHAKE algorithm [66] was applied to constrain all bonds involving hydrogens. The simulation time step of 2 fs was used. All MD simulations were run with a 12 Å residue-based cutoff for nonbonded interactions and the particle mesh Ewald method was used for an adequate treatment of long-range electrostatic interactions [68]. The MD simulations were carried out for 2.0 ns where the production phase starts from 0.75 ns to 2.0 ns. The convergences of energies, temperature, pressure, and global RMSD (root mean square displacement) were used to verify the stability of the systems. The MD trajectories were collected every 0.2 ps. Analysis of all MD trajectories, *i.e.*, RMSD, distances, hydrogen bonds, etc., were carried out using the Ptraj, CARNAL and DC modules of AMBER 7.

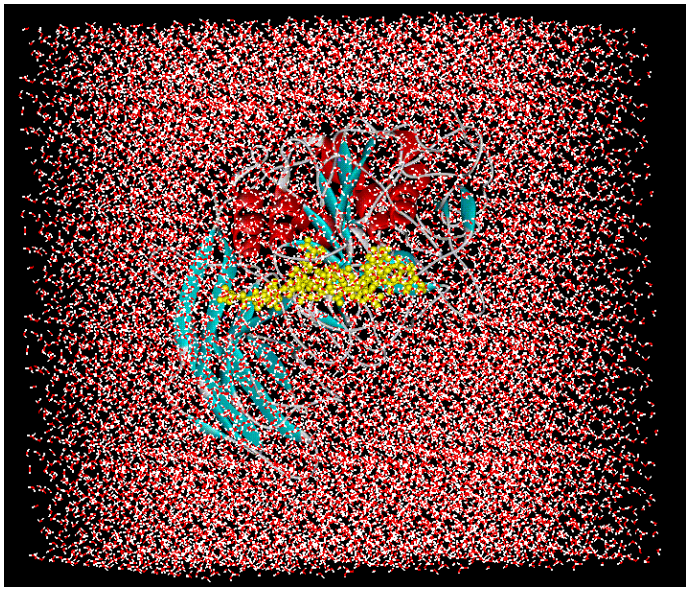


Figure 4.6 The complex of HPH5–FR in solvation box.

4.3 Results and discussion

4.3.1 Overall enzyme-substrate structure

To monitor the stability of the three simulated systems, HPH5–FR, LPH5–FR and LPH3–FR, RMSD of the structures obtained during the 2ns MD simulations relative to the initial structure of all heavy atoms in the substrate-enzyme

complex and the 10 residues of HA substrate were evaluated and plotted in Figure 4.7. It can be seen from the plot of RMSD plots versus simulation time that all three systems were found to reach equilibrium at 0.75 ns.

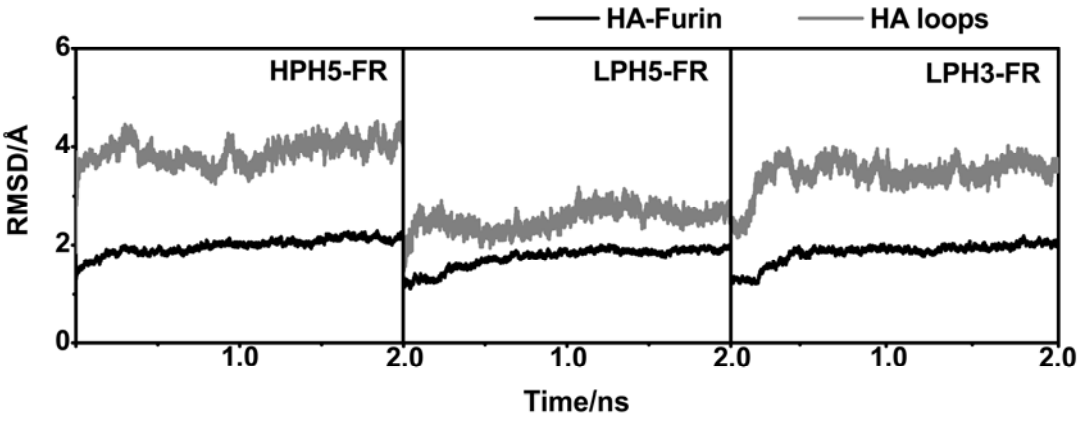


Figure 4.7 RMSDs relative to the initial structure for all heavy atoms of the substrate-furin complexes (black) and the HA loops (grey) for the three systems studied: HPH5–FR, LPH5–FR and LPH3–FR.

4.3.2 Enzyme-substrate conformation

The main results refer to the schematic representation shown in Figure 1 where the conformation structures of the complexes are described in terms of d_1-d_6 (Figure 4.8) in which the reaction coordinates involved in the acylation in three regions, the catalytic triad of furin, nucleophilic attack, and oxyanion hole, were described by the d_1-d_3, d_4 and d_5-d_6, respectively (see Figure 4.1). Accordingly, percentage H-bond occupations were evaluated using the CARNAL module in AMBER7 based on the following criteria: *(i)* proton donor-acceptor distance ≤ 3.5 Å, and *(ii)* donor-H-acceptor bond angle ≥ 120°. The results were summarized in Table 4. 1.

At the catalytic triad (D153, H194 and S368) region of the HPH5–FR, the simulated distances of $d_1 = d_2 = d_3 = 2.85$ Å (Figure 4.8) are comparable to those found experimentally for the inhibitor-furin complex of 2.53 Å, 3.10 Å, and 2.95 Å, respectively, [43]. This indicates that the three catalytic residues, D153, H194, and S368, are in the configuration suitable for initiating the nucleophilic reaction. The

situation is different for the bound LPAI, H5 and H3, hemagglutinin loops, where only d_3 of both systems and d_1 of LPH5–FR fall within the range whereby the reaction can take place.

The nucleophilic attack is determined by the distance between the O^3-hydroxyl oxygen of S368 and the carbonyl carbon of the S1-Arg, d_4 in Figure 4.1. The d_4 bond-making distances were found to be at 3.05 Å, 3.45 Å, and 3.75 Å for the HPH5–FR, LPH5–FR, and LPH3–FR systems, respectively (Figure 4.8). In addition to a short d_4, the detected sharp and narrow peak was found only in the HPH5–FR complex, signifying the rigidity of the complex which thus serves as a more appropriate configuration for the nucleophilic attack. In contrast, for the other two complexes the d_4 peak shows a broad distribution and takes place at significantly longer distances.

At the oxyanion hole region of the furin active site (d_5 and d_6 in Figure 4.1), the hole is generally formed by the backbone nitrogen of S368 and the carboxamide nitrogen of N295, which specifically binds with the carbonyl group of the S1 reacting residue [43]. In Figure 4.8, the d_5 distance of 2.95 Å for the HPH5–FR (solid line) is slightly shorter than that of 3.15 Å for the LPH5–FR (dashed line) with the presence of a strong hydrogen bond to S368, 100% and 86% occupations (d_5 in Table 4.1), respectively. The percentage occupation of 15% for the LPH3–FR confirms a very weak interaction and highly flexible of the complex. This conclusion was also confirmed by a broad distribution of d_5 distance varying from 3.00 Å to 5.85 Å (Figure 4.8). In terms of d_6 distance, the distribution plot for the HPH5–FR complex shows two apparent peaks centered at 2.85 Å and 5.45 Å where the sharper first peak indicated a preferential interaction. The fluctuation of d_6 distance between the carbonyl oxygen of S1-Arg and carboxamide nitrogen of N295 was back and forth (Figure 4.9) suggesting the medium to low H-bond interaction between these two atoms. When N295 lost the interaction with S1-Arg, its carboxamide nitrogen was solvated by water molecules (see H-bond numbers 3-5 in Table 4.2). For LPAI systems, the d_6 distance is detected at 3.05 Å in H3 while this distance is significantly longer in H5. This means that interaction between S1 residue and Asn295 was almost lost in the LPH5–FR system. The observed intermolecular distances agree very well with the percentage pair of hydrogen bond through the d_5 and d_6 in which those values (d_5 = 100% and d_6 = 34%) for the HPH5–FR are higher than those (d_5 = 86% and d_6 =

6%) for the LPH5–FR. On the contrary, bonding via these two distances almost disappear for the LPH3–FR complex. Note that the experimental d_5 and d_6 distances for the inhibitor-furin complex are 3.34 Å and 2.78 Å, respectively [43].

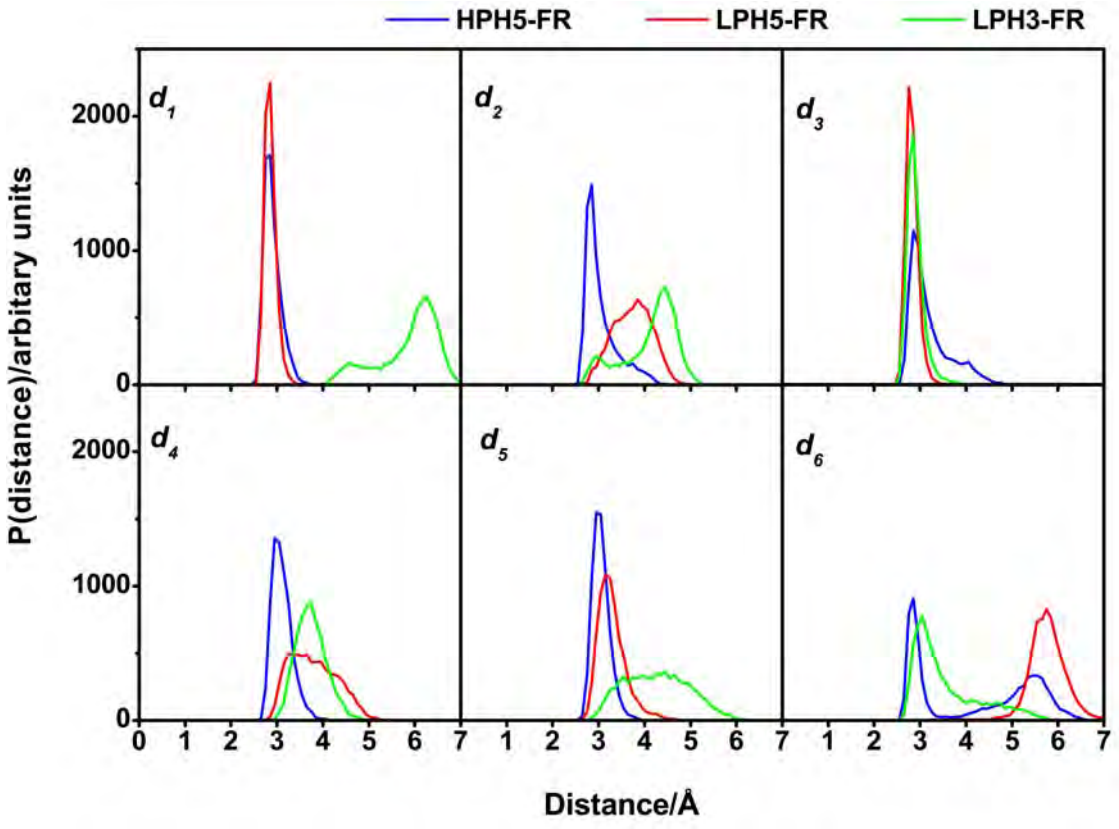


Figure 4.8 Distributions of the d_1-d_6 distances defined in Figure 4.1 for the three simulated systems, sampling from 0.75 to 2.0 ns in MD simulations.

Table 4.1 Percentage of hydrogen bond according to the d_1-d_3 and d_5-d_6 distances defined in Figure 4.1 for the three simulated systems: HPH5–FR, LPH5–FR and LPH3–FR.

Distance	H-bond	HPH5–FR	LPH5–FR	LPH3–FR
d_1	D153_O^1…HN^1_H194	100	100	0
d_2	D153_O^2…HN^1_H194	96	75	32
d_3	H194_N^2…HO^3_S368	81	90	98
d_5	S1-R_ O^γ … HN^3_ S368	100	86	15
d_6	S1-R_ O^γ … HN^4_N295	34	6	34
	S368_O^3…HN_S1-R	71	10	0

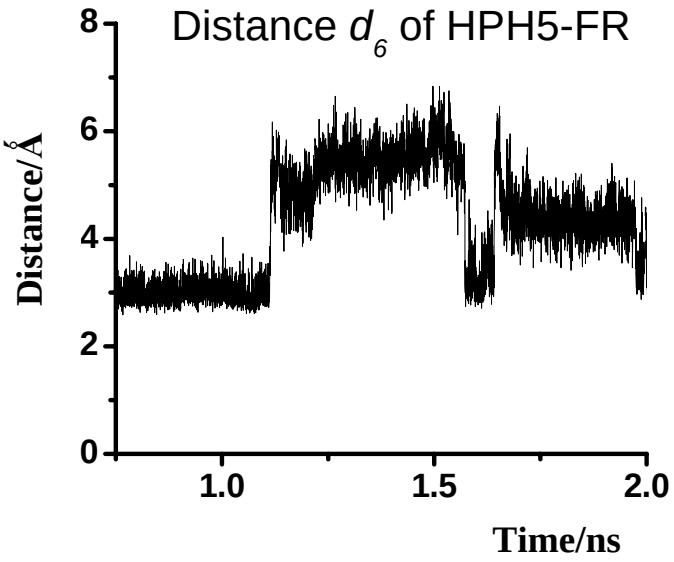


Figure 4.9 Distance d_6 of HPH5–FR plot versus simulation time, sampling from 0.75 to 2.0 ns in MD simulations.

Table 4.2 Percentage of hydrogen bond of N295 for the HPH5–FR system.

	% H-bond
1) N295_HD21…O_S1	34%
2) N295_HD22…O_S2’	27%
3) N295_HD21…O_WAT5094	10%
4) N295_HD21…O_WAT8550	10%
5) N295_HD21…O_WAT6334	11%
6) N295_H…OD1_D258	25%
7) N295_H…OD2_D258	75%
8) N295_H…OE1_E331	59%
9) N295_O…H1_WAT5689	16%
10) N295_O…H2_WAT5689	78%

4.3.3 Hydrogen bonds between enzyme-substrate

To assess to molecular interaction at the cleavage loop of hemagglutinin, percentage and number of hydrogen bond between each HA residue

and the active site residues of furin, were evaluated and plotted in Figure 4.10. The criteria used are the same as those mentioned earlier. Here, the experimentally detected hydrogen bonds for the inhibitor-furin complex [43] were also given for comparison, the residue with a box around the amino acid label. Interest was focused on S1-Arg of HA and furin's flat groove – a cavity formed by residues P256, D258, A292 and D306. The later was found to play an important role in facilitating the cleavage reaction by forming hydrogen bonds to the guanidinium side chain (see Figure 4.10) of P1-Arg of the inhibitor [43] (Si denotes a residue of the substrate while Pi represents one of the inhibitor, where i = 1, 2, 3, …) [108]. Strong hydrogen bonds between the four flat groove residues and S1-Arg were obviously detected for the HPH5–FR complex (S1 in Figure 4.10). The cleavage reaction in the HPH5–FR system was also promoted by the other hydrogen bonds from S1-Arg to S368 (2 bonds, one of which is d_5 in Figure 4.1) and S253 (1 bond) of furin with almost 100 % occupation. These observations are strongly supported by the experimentally detected hydrogen bonds shown in the same Figure. This is in contrast with LPH5–FR in which no hydrogen bond with the flat groove was detected and only a single bond with the catalytic S386 (d_5 in Figure 4.1) was formed. The LPH3–FR loop shows much less interaction with the furin residues, forming no hydrogen with either the flat groove or the catalytic S368 residue.

Some comments could be made concerning the two hydrogen bonds formed between the reacting S1-Arg and the catalytic S386 of HPH5–FR (Figure 4.10). Besides that defined by d_5, another bond is S368_O^3…NH_S1-Arg (71 % occupation). This hydrogen bond which has never been found in the oxyanion hole region of any other HA subtype or in the inhibitor-furin complex [43], could be a reason for the shortening and rigidity (narrow and sharp peaks shown in Figure 4.8) of the intermolecular d_4 distance, which could directly facilitate the nucleophilic reaction in HPH5–FR.

Considering role of the –RRRKK– insertion (Figure 4.1), more hydrogen bonds and a higher percentage occupation between the S2-S6 residues of HA and the surrounding residues of furin (Figure 4.10) were found for HPH5–FR in comparison with the two LPAI systems. This means that the –RRRKK– insertion can directly help to hold the substrate in place.

Taking into account all the hydrogen bond data given above, the hemagglutinin loop of HPH5–FR is observed to bind much more tightly into the catalytic site of furin than the LPH5–FR and LPH3–FR systems.

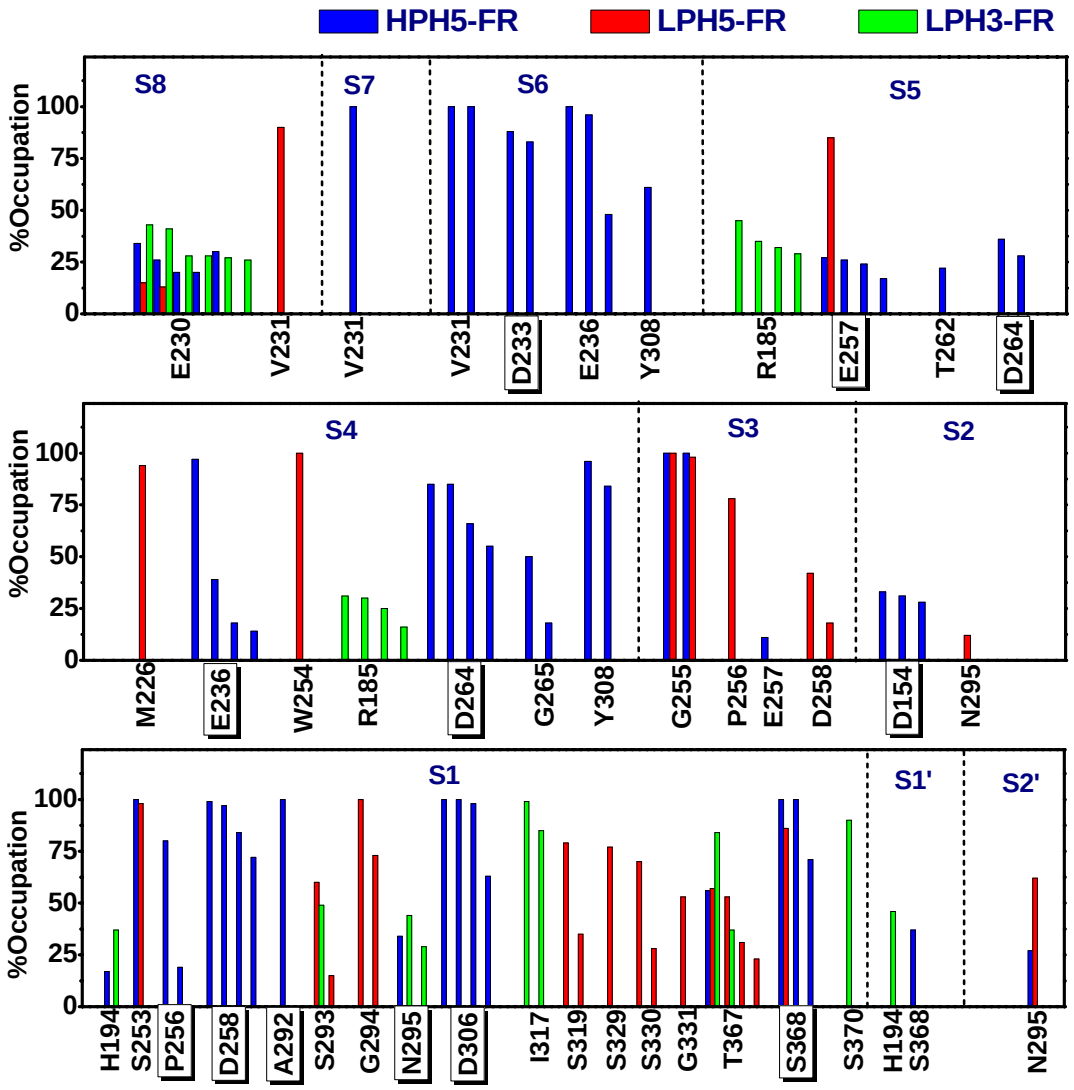


Figure 4.10 Percent occupations of hydrogen bonds between furin and the ten HA residues (defined in Figure 4.1) where the residues with a box around the label represent experimentally detected bonds for the inhibitor-furin complex.

4.3.4 Per residue enzyme-substrate interactions

To seek for the fundamental basis of substrate-furin interactions, the interaction energy between individual furin residue and the three HA loops and vice

versa were calculated using the DC module of AMBER 7. Plots of the DC of the HA residues as well as the selected furin residues that located within 5 Å from the ten residues of HA loop were shown in Figure 4.11. The overall DCs of furin residues were observed in the following order: HPH5–FR >> LPH5–FR ~ LPH3–FR (Figure 4.11B). For the HA loops, the DCs for the residues of HPH5–FR are significantly lower than for the corresponding residues of LPH5–FR and LPH3–FR (Figure 4.11A). This is especially true for the S1, S4 and S6 residues. The calculated results are consistent with experimental results for the inhibitor-furin complex, which stated that P1, P4 and P6 residues of the inhibitor were observed to interact strongly with furin [47, 98]. In addition, these two P4 and P6 residues are required to dramatically increase the cleavage reaction [109]. The DCs for the furin and HA residues agree very well with the hydrogen bond data shown (Figure 4.10) and discussed above.

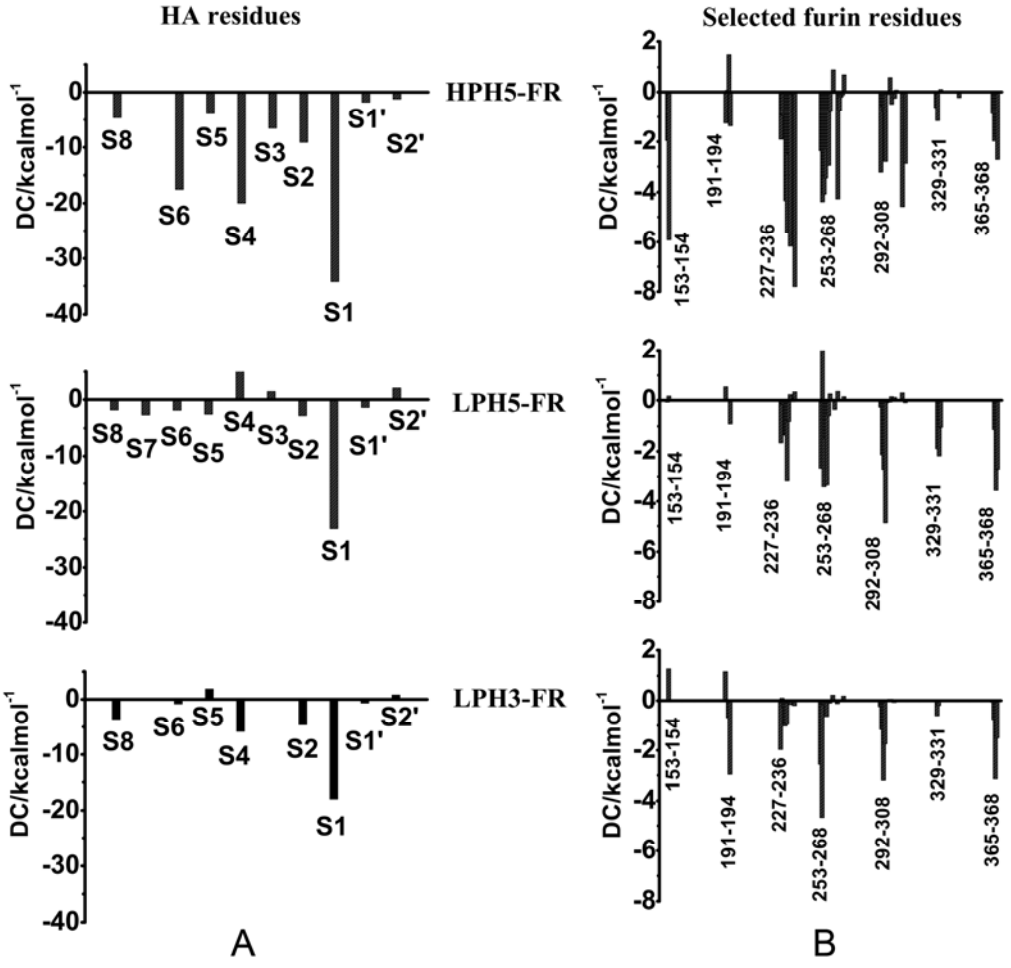


Figure 4.11 Decomposition energies (DC) of **(A)** ten HA residues (defined in Figure 4.1) and **(B)** selected furin residues within 5 Å of those ten HA residues.

4.3.5 Water accessibilities in catalytic mechanism of furin

Hydrophilicity which determined by ligand's solution in the cavity region of enzyme-substrate complex is known to play essential role in the catalytic mechanism of biological systems. To seek for such information, the plot of the RDFs, $g_{xy}(r)$ –the probability of finding the particle of type *y* in a spherical radii, *r*, around the particle of type *x* was evaluated. Here, the RDFs from the acceptor atoms of furin pocket (O^1, O^2:Asp153; N^1, N^2:His194; N^4:Asn295 and N^3, O^3:Ser368) and the reactive HA residue (O^γ:S1-Arg), see label in Figure 4.1, to oxygen atom of water was calculated. The results for the three simulated systems were shown in Figure 4.12 together with the corresponding running integration numbers up to the first minimum of the RDF plot.

The RDFs for all oxygen atoms, O^1, O^2, O^3 and O^γ, of the LPH3–FR complex show sharp and narrow first peaks at ~2.9 Å with the integration number (coordination number, CN) of 1.5, 1, 1 and 2 water molecules, respectively (see Figure 4.12). The CN denotes the number of water molecules which locate in the first hydration shell around the central atom. This is in contrast for the LPH5–FR complex where only O^2 can be accessed by water molecule whereas no water was detected around the four oxygen atoms of HPH5–FR.

The RDFs centered on N^1 and N^4 atoms of the three systems are almost similar, the first broad peaks at ~3.8 Å and sharp peaks at ~ 3 Å, respectively. Note that, N^3 atom of only LPH3–FR was solvated by water molecule. Interest is focused to N^2 atom of His194 where the catalytic water was observed and proposed to play role in the reaction mechanism of serine protease (14). In good agreement with our simulated results, one water (CN=1) was detected to preferentially coordinate (first peak centered at ~2.9 Å) to N^2 atom of only for the HPH5–FR complex. This is not the case for the other two low-pathogenic systems, LPH3–FR and LPH5–FR, where less water, 0.5 and 0.2 were, respectively, was found (Figure 4.12, green and red lines). In addition, distance to the LPH3-FR water molecule of ~3.45 Å is significantly longer than that of the HPH5–FR complex.

Taking into account the solvation data shown above, number of overall water molecules around the selected central atoms are in the following order: LPH3–

FR > LPH5–FR > HPH5–FR. The obtained data indicates that more water molecules can take place in the catalytic pocket of furin in the complex whose substrate does not fit well with the furin active site, giving a subsequent result of more cavities or less binding between substrate and enzyme. This suggestion agrees very well with the conformational structures of the complexes (Figure 4.8) and the hydrogen bond data (Figure 4.10) shown and discussed previously.

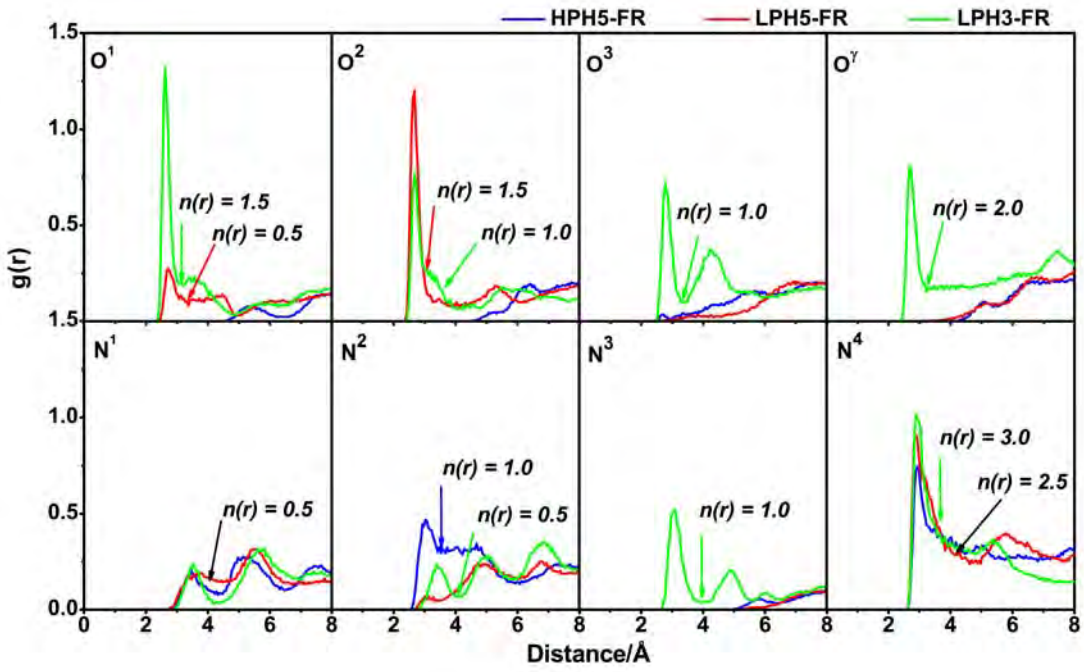


Figure 4.12 Radial distribution functions, g(r), from acceptor atoms of furin catalytic residues and HA loops (see Figure 4.1 for definition) to oxygen atoms of water molecules for the three simulated systems, HPH5–FR, LPH5–FR and LPH3–FR complexes.

CHAPTER V

CONCLUSION

5.1 HIV-1 RT/ inhibitors complexes

MD simulations of the HIV-1 RT complexed with EFV, EMV, ETV and NVP provide information on hydrogen bonding, important residues of the HIV-1RT/inhibitor interaction energies, binding free energies and hydration structures of the complexes. The simulation results report good evidence which deal with questions related to the basic mutation and free energy data. Concerning the interaction energies, L100, K101, K103, V106, Y181, Y188, F227, W229 and L234 show the largest contributions to the enzyme/inhibitor interaction energies. The obtained results support clinical data which reveal that these residues are the most frequent amino acids when mutation against NNRTIs takes place. The data lead us to conclude that the residues which are of high importance for the interaction energy are primary sources for mutation of NNRTIs. In addition, the binding affinities calculated by MM-PBSA, $\Delta G_{binding}$ values were observed in the following order: EFV ~ ETV > EMV > NVP in which this order was decreasingly solvated by water molecules. This implies that EFV atoms can be much less accessed by water molecules than the other three inhibitors. The $\Delta G_{binding}$ values agree well with the experimental IC_{50} values. In addition, the distribution of water molecules that is comparable to those found in the crystal structures and is in good agreement with enzyme/inhibitor binding energies in which the EFV atoms can be much less accessed by water molecules than the other three inhibitors.

5.2 Furin/HA-substrate complexes

The technique of MD simulations has been applied to seek for detailed information focused to intra- and intermolecular interactions and geometries of the three substrate-furin complexes, HPH5–FR, LPH5–FR and LPH3–FR. From the obtained results, the –RRRKK– insertion in the HPH5–FR in particular two arginines at S4 and S6 positions helps directly to hold the HA's cleavage loop in place by forming many strong hydrogen bonds between residues of HA and furin. This is well supported by the corresponding strong-decomposition interaction energies and leads

consequently to accumulate less water molecules accessible into the furin cavity. Therefore, an active conformation of the HPH5–FR complex suitable for the acylation reaction by furin was formed. In conclusion, the results yielded from this study provide a clear answer to the question why HPAI hemagglutinin H5 is better cleavaged by furin than the other two HA subtypes.

REFERENCES

[1] Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, C., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzinouset, C., Rozenbaum, W., Montagnier, L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-870.

[2] Joint United Nations Programme on HIV/AIDS World Health Organization (WHO). AIDS epidemic update: World epidemic update 2007 UNAIDS [online]. Available from: http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf

[3] Roshan, P. Health Issues : AIDS[online]. Available from: http://www.roshanpakistan.com/web-magzine/Health/aids.php [2009, February 28].

[4] Wondrak, E. M., Loewer, J., Kurth, R. 1986. Functional purification and enzymic characterization of the RNA dependent DNA polymerase of human immunodeficiency virus. J. Gen. Virol. 67: 2791-2797.

[5] Hostomsky, Z., Hostomska, Z., Fu, T. B., Taylor, J. 1992. Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. J. Virol. 66: 3179-3182.

[6] Kohlstaedt, L. A., Wang, J., Friedman ,J. M., Rice, P. A., Steitz, T. A. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256: 1783-1790.

[7] Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., Stammers, D. 1995. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nat. Struct. Biol. 2: 293-302.

[8] Ren, J., Milton, J., Weaver, K. L., Short, S. A., Stuart, D. I., Stammers D. K., 2002. Structural basis for the resilience of efavirenz (DMP-266) to drug

resistance mutations in HIV-1 reverse transcriptase. Structure Fold. Des. 8: 1089-1094.

[9] Wang, J., Morin, P., Wang, W., Kollman, P. A. 2001. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. 123: 5221-5230.

[10] Pungpo, P., Saparpakorn, P., Wolschann, P., Hannongbua, S. 2006. Computer-aided molecular design of highly potent HIV-1 RT inhibitors: 3D QSAR and molecular docking studies of efavirenz derivatives. SAR QSAR Environ. Res. 17: 353-370.

[11] Saparpakorn, P., Hannongbua, S., Rognan, D. 2006. Design of nevirapine derivatives insensitive to the K103N and Y181CHIV-1 reverse transcriptase mutants. SAR QSAR Environ. Res. 17: 183-194.

[12] Lawtrakul, L., Hannongbua, S. 2007. Molecular dynamics study on the unbinding of HBY 097 in the K103N mutant RT. Monatshefte für Chemie 138: 1029-1034.

[13] Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc. Natl. Acad. Sci. USA. 90: 6320-6324.

[14] Rungrotmongkol, T., Mulholland, A. J., Hannongbua, S. 2007. Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. J. Mol. Graph. Model. 26: 1-13.

[15] Das, K., Clark, A. D., Jr., Lewi, P. J., Heeres, J., De Jonge, M. R., Koymans, L. M. H., Vinkers, H. M., Daeyaert, F., Ludovici, D. W., Kukla, M. J., De Corte, B., Kavash ,R. W., Ho, C. Y., Ye, H., Lichtenstein, M. A., Andries, K., Pauwels, R., De Béthune, M. P., Boyer, P. L., Clark, P., Hughes, S. H.,

Janssen, P. A. J., Arnold, E. 2004. Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. J. Med. Chem. 47: 2550-2560.

[16] Chong, Y., Borroto-Esoda, K., Phillip, A. F., Raymond, F. S., Chung, K. C. 2002. Molecular mechanism of DApd/DXG against zidovudine- and lamivudine- drug resistant mutants: a molecular modelling approach. Antivir. Chem. Chemother. 13: 115-128.

[17] Kedar, P. S., Abbotts, J., Kovacs, T., Lesiak, K., Torrence, P., Wilson, S. H. 1990. Mechanism of HIV reverse transcriptase: enzyme-primer interaction as revealed through studies of a dNTP analogue, 3'-azido-dTTP. Biochemistry 29: 3603-3611.

[18] Hannongbua, S., Nivesanond, K., Lawtrakul, L., Pungpo, P., Wolschann, P. 2001. 3D-Quantitative structure-activity relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on ab initio calculations. J. Chem. Inf. Model. 41: 848-855.

[19] De Clercq, E. 2004. Antiviral drugs in current clinical use. J. Clin. Virol. 30: 115-133.

[20] Sarafianos, S. G., Das, K., Hughes, S. H., Arnold, E. 2004. Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases. Curr. Opin. in Struct. Biol. 14: 716-730.

[21] Meyer P. R., Matsuura S. E., So A. G., Scott W. A. 1998. Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. Proc. Natl. Acad. Sci. USA. 95: 13471-13476.

[22] Klarmann, G. J., Smith, R. A., Schinazi, R. F., North, T. W., Preston, B. D. 2000. Site-specific incorporation of nucleoside analogs by HIV-1 reverse transcriptase and the template grip mutant P157S. Template interactions

influence substrate recognition at the polymerase active site. J. Biol. Chem. 275: 359-366.

[23] Gao, H. Q., Boyer, P. L., Sarafianos, S. G., Arnold, E., Hughes, S. H. 2000. The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase. J. Mol. Biol. 300: 403-418.

[24] Sluis-Cremer, N., Arion, D., Kaushik, N., Lim, H., Parniak, M. A. 2000. Mutational analysis of Lys65 of HIV-1 reverse transcriptase. Biochem. J. 348: 77-82.

[25] Lawtrakul, L., Beyer, A., Hannongbua, S., Wolschann, P. 2004. Quantitative structural rearrangement of HIV-1 reverse transcriptase on binding to non-nucleoside inhibitors. Monatsh. Chemie. 135: 1033-1046.

[26] Saen-Oon, S., Hannongbua, S., Wolschann, P. 2003. Structural flexibility of non-nucleoside HIV-1 reverse transcriptase inhibitor: 9-Cl TIBO as explained by potential energy surface and (13)C and (1)H NMR calculations, based on ab initio and density functional study. J. Chem. Inf. Model. 43: 1412-1422.

[27] Hsiou, Y., Ding, J., Das, K., Clark, A. D., Boyer, P. L., Lewi, P., Janssen, P. A. J., Kleim, J. P., Rösner, M., Hughes, S. H., Arnold, E. 2001. The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. J. Mol. Biol. 309: 437-445.

[28] Wittayanarakul, K., Hannongbua, S., Feig, M. 2007. On the correlation between drug-resistant pattern of HIV-1 protease inhibitors and binding free energy and structural changes. BMC Systems Biology P58: 1-2.

[29] Horimoto, T., Kawaoka, Y. 2001. Pandemic threat posed by avian influenza A viruses. Clin. Microbiol. Rev. 14: 129-149.

[30] Emmeluth, D. 2003. Deadly diseases and epidemics: Influenza. New York: Chelsea House.

[31] Laidback, A. l. Cumulative count of countries with human cases of H5N1 [online]. Available from: http://www.flutrackers.com/forum/showthread. php ?t= 16352 [2009, February 25].

[32] World Health Organization (WHO). Confirmed Human Cases of Avian Influenza A(H5N1) [online]. Available from: http://www.who.int/csr/disease/avian influenza/country/en/. [2009, February 28].

[33] Cross, K. J., Burleigh, L. M., Steinhauer, D. 2001. Mechanisms of cell entry by influenza virus. Expert. Rev. Mol. Med. 6: 1-18.

[34] Horimoto, T., KAwaoka, Y. 2005. Influenza: lessons from past pandemics, warnings from current incidents. Nat. Rev. Microbiol. 3: 591-600.

[35] Horimoto, T., Kawaoka, Y. 1995. The hemagglutinin cleavability of a virulent avian influenza virus by subtilisin-like endoproteases is influenced by the amino acid immediately downstream of the cleavage site. Virology 210: 466-470.

[36] Steinhauer, D. A. 1999. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology 258: 1-20.

[37] Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., Wilson, I. A 2006. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science 312: 404-410.

[38] Neumann, G., Kawaoka, Y. 2006. Host range restriction and pathogenicity in the context of influenza pandemic. Emerging Infect. Dis. 12: 881-886.

[39] Basak, A., Zhong, M., Munzer, J. S., Tien, M. C., Seidah, N. G. 2001. Implication of the proprotein convertases furin, PC5 and PC7 in the cleavage of surface glycoproteins of Hong Kong, Ebola and respiratory syncytial viruses : a comparative analysis with fluorogenic peptides. Biochem. J. 353: 537-545.

[40] Bergeron, F., Ledue, R., Day, R. 2000. Subtilase-like pro-protein convertases: from molecular specificity to therapeutic applications. J. Mol. Endocrinol. 24: 1-22.

[41] Chen, J., Lee, K. L., Steinhauer, D. A., Stevens, J., Skehel, J. J., Wiley, D. C. 1998. Structure of the hemagglutini precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell 95: 409-417.

[42] Stieneke-Gröber, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H. D., Garten, W. 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. EMBO J. 11: 2407-2414.

[43] Henrich, S., Cameron, A., Bourenkov, G. P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W., Than, M. E. 2003. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. Nature Struct. Biol. 10: 520-526.

[44] Kacprzak, M. M., Peinado, J. R., Than, M. E., Appel, J., Henrich, S., Lipkind, G., Houghten, R. A., Bode, W., Lindberg, I. 2004. Inhibition of furin by polyarginine-containing peptides. J. Biol. Chem. 279: 36788-36794.

[45] Oliva, R., Falcigno, L., D'Auria, G., Dettin ,M., Scarinci, C., Pasquato, A., Bello, C. D., Paolillo, L. 2003. Structural investigation of the HIV-1 envelope glycoprotein gp 160 cleavage site, 2 : relevance of an N-terminal helix. ChemBioChem 4: 727-733.

[46] Oliva, R., Leone, M., Falcigno, L., D'Auria, G., Dettin, M., Scarinci, C., Bello C. D., Paolillo, L. 2002. Structural investigation of the HIV-1 envelope glycoprotein gp 160 cleavage site. Chem. Eur. J. 8: 1467-1473.

[47] Rockwell, N. C., Krysan, D. J., Komiyama, T., Fuller, R. S. 2002. Precursor processing by kex2/furin proteases. Chem. Rev. 102: 4525-4548.

[48] Hedstrom, L. 2002. Serine protease mechanism and speecificity. Chem. Rev. 102: 4501-4523.

[49] Bevan, D. R., Homology modeling[online]. Available from http://www.biochem.vt.edu/modeling/homology.html [2009, February 14].

[50] Needleman S. B., Wunsch C. D. 1970. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. J. Mol. Biol. 48: 442-453.

[51] Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26: 283-291.

[52] Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. Model. 8: 52-56.

[53] Gopalakrishnan, K., Sowmiya, G., Sheik, S. S., Sekar, K. 2007. RP (2.0): Ramachandran plot on the web. Protein Peptide Letters 14: 669-671.

[54] Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., Vakser, I.A. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. USA. 89: 2195-2199.

[55] Chen, R., Li, L., Weng, Z. 2003. ZDOCK: an initial-stage protein docking algorithm. Protein 52: 80-87.

[56] Allen, M. P., Attig, N., Binder, K., Grubmuller, H., Kremer, K. 2004. Introduction to molecular dynamics simulation. NIC Series, John von Neumann Institute for Computing 1-28.

[57] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. 1995. A second generation forcefield for the simulation of proteins, nucleic-acids, and organic-molecules. J. Am. Chem. Soc. 117: 5179-5197.

[58] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. 1983. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4: 187-217.

[59] Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F., Postma, J. P. M. 1984. A consistent empirical potential for water-protein interactions. Biopolymers 23: 1513-1518.

[60] Jorgensen, W. L., Maxwell, D. S. 1996. TiradoRives J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. 118: 11225-11236.

[61] Case, D. A., Pearlman, J. W., Caldwell, T. E., Wang, W., Ross, C. L., Simmerling, T. A., Darden, K. M., Merz, R. V., Stanton, A. L., Cheng, J. J., Vincent, M., Crowley, V., Tsui, H., Gohlke, R. J., Radmer, Y., Duan, J., Pitera, I., Massova, G. L., Seibel, U. C., Singh, P. K., Kollman, P. A. 2002. AMBER 7. San Francisco: University of California.

[62] Case, D. A., Darden, T. A., Cheatham, T. E., et. al. 2006. AMBER 9. San Francisco: University of California.

[63] Drexel University. Introduction to numerical methods[online]. Available from: http://www.math.drexel.edu/~pg/520/images/CG.png [2009, February 19].

[64] Adcock, S. A., McCammon, J. A. 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem. Rev. 106: 1589-1615.

[65] Verlet, L. 1967. Computer experiments on classical fluids. *i.* thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159: 98-103.

[66] Ryckaert, J. P., Cicotti, G., Berendsen, H. J. C. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23: 327-341.

[67] Pollock, E. L., GLlosli, J. 1996. Comment on P^3M, Fmm nad the Ewald method for large periodic Coulombic systems. Comp. Phys. Comm. 95: 93.

[68] York, D. M., Darden, T. A., Pedersen, L. G. 1993. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a

comparison of the Ewald and truncated list methods. J. Chem. Phys. 99: 8345-8348.

[69] Frenkel, D., Smit, B. 2002. Understanding molecular simulation : from algorithms to applications. San Diego

[70] Kavraki, L. E. 2007. Molecular shapes and surfaces. Creative commons attribution 1-9.

[71] Zhou, Z., Madrid, M., Evanseck, J. D., Madura, J. D. 2005. Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. J. Am. Chem. Soc. 127: 17253-17260.

[72] Rizzo, R. C., Udier-Blagovic´, M., Wang, D. P., Watkins, E. K., Kroeger, Smith M. B., Smith, R. H., Jr., Tirado-Rives, J., Jorgensen, W. L. 2002. Prediction of activity for nonnucleoside inhibitors with HIV-1 reverse transcriptase based on monte carlo simulations. J. Med. Chem. 45: 2970-2987.

[73] Maggiolo, F. 2007. Efavirenz. Expert Opin. Pharmacother. 8: 1137-1145.

[74] Kehr, H. A., Olin, J. L., Love, B. L. 2008. Etravirine - a non-nucleoside reverse transcriptase inhibitor for the treatment of resistant HIV-1 infection. Formulary 43: 105-114.

[75] Mao, C., Sudbeck, E. A., Venkatachalam ,T. K., Uckun, F. M. 2000. Structure-based drug design of non-nucleoside inhibitors for wild-type and drug-resistant HIV reverse transcriptase. Biochem. Pharmacol. 60 : 1251-1265.

[76] Rodríguez-Barrios, F., Balzarini, J., Gago, F. 2005. The molecular basis of resilience to the effect of the Lys103Asn mutation in non-nucleoside HIV-1 reverse transcriptase inhibitors studied by targeted molecular dynamics simulations. J. Am. Chem. Soc. 127: 7570-7578.

[77] Vingerhoets, J., Azijn, H., Fransen, E., De Baere, I., Smeulders, L., Jochmans, D., Andries, K., Pauwels, R., De Be´thune, M. P. 2005. TMC125 displays a high genetic barrier to the development of resistance: evidence from in vitro selection experiments. J. Virol. 79: 12773-12782.

[78] Martinez-Picado, J., Martinez, M. A. 2008. HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. Virus Res. 134: 104-123.

[79] Shen, L., Shen, J., Luo, X., Cheng, F., Xu, Y., Chen, K., Arnold, E., Ding, J., Jiang, H. 2003. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. Biophys. J. 84: 3547-3563.

[80] Weinzinger, P., Hannongbua, S., Wolschann, P. 2005. Molecular mechanics PBSA ligand binding energy and interaction of Efavirenz derivatives with HIV-1 reverse transcriptase. J. Enzym. Inhib. Med. Chem. 20: 129-134.

[81] Hopkins, A. L., Ren, J., Esnouf, R. M., Willcox, B. E., Jones, E. Y., Ross, C., Miyasaka ,T., Walker, R. T., Tanaka, H., Stammers, D. K., Stuart, D. I. 1996. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. J. Med. Chem. 39: 1589-1600.

[82] Li, H., Robertson, A. D., Jensen, J. H. 2005. Very fast empirical prediction and interpretation of protein pKa values. Proteins 61: 704-721.

[83] Frisch M. J., Trucks G. W., Schlegel H. B., et. al. 2004. Gaussian 03, Revision C.02. Gaussian, Inc.: Wallingford CT.

[84] Cornell, W. D., Cieplak, P., Bayly, C. I., Kollmann, P. A. 1993. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. J. Am. Chem. Soc. 115: 9620-9631.

[85] Wang, J. M., Wang, W., Kollman, P. A. 2001. Antechamber: an accessory software package for molecular mechanical calculations. J. Am. Chem. Soc. 222: U403-U403.

[86] Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24: 1999-2012.

[87] Lee, M. C., Duan, Y. 2004. Distinguish protein decoys by using a scoring function based on a new amber force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins 55: 620-634.

[88] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926-935

[89] Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren ,W. F., DiNola, A., Haak, J. R. 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81: 3684-3690.

[90] Gilson, M. K., Sharp, K. A., Honig, B. H. 1987. Calculating the electrostatic potential of molecules in solution: method and error assessment. J. Comput. Chem. 9: 327-335.

[91] Villa, J., Strajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., Warshel, A. 2000. How important are entropic contributions to enzyme catalysis? Proc. Natl. Acad. Sci. USA. 97: 11899-11904.

[92] Kuhn, B., Kollman, P. A. 2000. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J. Med. Chem. 43: 3786-3791.

[93] Wang, J., Kang, X., Kuntz, I. D., Kollman, P. A. 2005. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. J. Med. Chem. 48: 2432-2444.

[94] Luo, R., David, L., Gilson, M.K. 2002. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. J. Comput. Chem. 23: 1244-1253

[95] Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J. Am. Chem. Soc. 120: 9401-9409.

[96] Udier-Blagovic´, M., Tirado-Rives, J., Jorgensen, W. L. 2003. Validation of a model for the complex of HIV-1 reverse transcriptase with nonnucleoside inhibitor TMC125. J. Am. Chem. Soc. 125: 6016-6017.

[97] Ren, J., Chamberlain, P. P., Stamp, A., Short, S. A., Weaver, K. L., Romines, K. R., Hazen, R., Freeman, A., Ferris, R. G., Andrews, C. W., Boone, L., Chan, J. H., Stammers, D. K. 2008. Structural basis for the improved drug resistance profile of new generation benzophenone non-nucleoside HIV-1 reverse transcriptase inhibitors. J. Med. Chem. 51: 5000-5008.

[98] Holyoak, T., Kettner, C. A., Petsko, G. A., Fuller, R. S., Ringe, D. 2004. Structural basis for differences in substrate selectivity in kex2 and furin protein convertases. Biochemistry 43: 2412-2421.

[99] Thomas, G. 2002. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Molecular Cell Biology 3: 753-766.

[100] Molloy, S. S., Bresnahan, P. A., Leppla, S. H., Klimpel, K. R., Thomas, G. 1992. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. J. Biol. Chem. 267: 16396-16402.

[101] Hosako, M., Nagahama, M., Kim, W. S., Watanabe, T., Hatsuzawa, K., Ikemizu, J., Murakami, K., Nakayama, K. 1991. Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. J. Biol.Chem. 266: 12127-12130.

[102] Ishida, T., Kato, S. 2003. Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. J. Am. Chem. Soc. 125: 12035-12048.

[103] Chutinimitkul, S., Songserm, T., Amonsin, A., Payungporn, S., Suwannakarn, K., Damrongwatanapokin, S., Chaisingh, A., Nuansrichay, B., Chieochansin, T., Theamboonlers, A., Poovorawan, Y. 2007. New strain of influenza A virus (H5N1), Thailand. Emerging Infect. Dis. 13: 506-507.

[104] Ha, Y., Stevens, D. J., Skehel, J. J., Wiley, D. C. 2002. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. EMBO J. 21: 865-875.

[105] Accelrys Software Inc. 2007. Discover Studio Program. San Diego: Accelrys Software Inc.

[106] Durbin, R., Eddy, S., Krogh, A., Mitchison, G. 2004. Needleman-Wunsch Algorithm. Cambridge University.

[107] Sheik, S. S., Sundararajan, P., Hussain, A. S. Z., Sekar, K. 2002. RP: Ramachandran plot on the web. Bionformatics 18: 1548-1549.

[108] Schechter, I., Berger, A. 1968. On the active site of proteases. 3. mapping the active site of papain; specific peptide inhibitors of papain. Biochem. Biophys. Res. Commun. 32: 898-902.

[109] Nakayama, K. 1997. Furin : a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. Biochem. J. 327: 625-635.

APPENDICES

APPENDIX A1: PUBLISHED ARTICLE

128 Biophysical Journal Volume 95 July 2008 128–134

Source of High Pathogenicity of an Avian Influenza Virus H5N1: Why H5 Is Better Cleaved by Furin

Panita Decha,* Thanyada Rungrotmongkol,* Pathumwadee Intharathep,* Maturos Malaisree,* Ornjira Aruksakunwong,† Chittima Laohpongspaisan,* Vudhichai Parasuk,* Pornthep Sompornpisut,* Somsak Pianwanit,* Sirirat Kokpol,* and Supot Hannongbua*

*Department of Chemistry, Faculty of Science, Chulalongkorn University, Patumwan, Bangkok 10330, Thailand; and †Department of Chemistry, Faculty of Science, Rangsit University, Pathumtani 12000, Thailand

ABSTRACT The origin of the high pathogenicity of an emerging avian influenza H5N1 due to the –RRRKK– insertion at the cleavage loop of the hemagglutinin H5, was studied using the molecular dynamics technique, in comparison with those of the noninserted H5 and H3 bound to the furin (FR) active site. The cleavage loop of the highly pathogenic H5 was found to bind strongly to the FR cavity, serving as a conformation suitable for the proteolytic reaction. With this configuration, the appropriate interatomic distances were found for all three reaction centers of the enzyme-substrate complex: the arrangement of the catalytic triad, attachment of the catalytic Ser^{360} to the reactive S1-Arg, and formation of the oxyanion hole. Experimentally, the –RRRKK– insertion was also found to increase in cleavage of hemagglutinin by FR. The simulated data provide a clear answer to the question of why inserted H5 is better cleaved by FR than the other subtypes, explaining the high pathogenicity of avian influenza H5N1.

INTRODUCTION

Proteolytic activation of the hemagglutinin (HA) is essential for viral infectivity and for spread of the avian influenza virus through the host's body (1–3). This process is determined by a cleavage reaction at the HA cleavage site, a conserved arginine, by host proteases (4–6). Insertion of the –RRRKK– residues into the low pathogenic avian influenza (LPAI) cleavage site is known to potentially activate infectivity of viruses, i.e., the LPAI viruses, which then become high pathogenic avian influenza (HPAI) viruses, allowing highly virulent strains to be cleaved by furin (FR), an ubiquitously expressed protease. The goal of this study is to understand why FR cleaves the inserted hemagglutinin strains better than the noninserted strains.

The HA is synthesized initially as an inactive precursor (HA0) that is then activated through a controlled proteolytic cleavage by protease into HA1 and HA2 subunits. HA1 mediates virus binding to host-cell receptors whereas HA2 promotes the release of the viral RNA complexed with polymerase through membrane fusion (1–6). Without proteolysis, the fusion peptide cannot occur and therefore the virus is essentially noninfectious.

The HA cleavage site relies on the presence of basic amino acids and relates directly to influenza virus pathogenicity, i.e., the LPAI viruses possess a single arginine at the cleavage site, while the HPAI viruses containing the polybasic insertion upstream of the cleavage site of the H5 subtype have potential to cause devastating pandemics in the future (3,7). In addition, in vitro cleavage of a series of peptide substrates showed that the insertion of the –RRRKK– residues at cleavage site led to an increase in cleavage by FR protease (2,8,9).

Submitted December 25, 2007, and accepted for publication March 5, 2008.

Address reprint requests to Supot Hannongbua, E-mail: supot.h@chula.ac.th.

Editor: Ron Elber.

0006-3495/08/07/128/07 $2.00

FR, a subtilisin-like serine endoprotease, seems to be a highly specific enzyme, cleaving pro-protein precursors at specific consensus sequence –RXK/RR–, usually to produce biologically active products (10–15). Based on the x-ray structure of mouse FR complexed with the dec-RVKR-cmk inhibitor, the catalytic ability of FR is considered to have originated from the catalytic triad residues (Ser^{368}, His^{194}, and Asp^{153}) and the oxyanion hole (formed by the carboxamide nitrogens of Asn^{295} and backbone nitrogen of Ser^{368} to carbonyl oxygen of the inhibitor/substrate's centered arginine) at the active site (12). The proposed cleavage mechanism is shown in Fig. 1 (16,17).

This work seeks the source of high pathogenicity in avian influenza A virus subtype H5 in comparison with LPAI subtypes H5 and H3. Three molecular dynamics simulations were carried out for the three complexes, HPH5–FR, LPH5–FR, and LPH3–FR. The investigation was focused to intra- and intermolecular interactions and geometries of the substrate-furin complex potentially involved to the cleavage mechanism.

METHODS

Initial structure of individual protein and substrates

The crystal structure of FR with bound dec-RVKR-cmk inhibitor (12) used as receptor model for MD simulations was obtained from Protein Data Bank (PDB), code: 1P8J. This study covers the loop of HA substrate features of subtypes H5 (both HPAI and LPAI) and H3 (known as LPAI). The sequence of the HPAI subtype H5 isolated during the 2006 influenza outbreaks in Thailand was taken from Genbank LOCUS ID ABK13784 (A/chicken/Thailand/PC-168/2006(H5N1)) (18). The sequence of the LPAI subtype H5 was obtained from Protein Data Bank (PDB entry code: 1JSM) (19).

The initial model for the HPAI_H5 loop (RERRRKKRGL) was built up using the sequence alignments and the atomic coordinates of the x-ray

doi: 10.1529/biophysj.107.127456

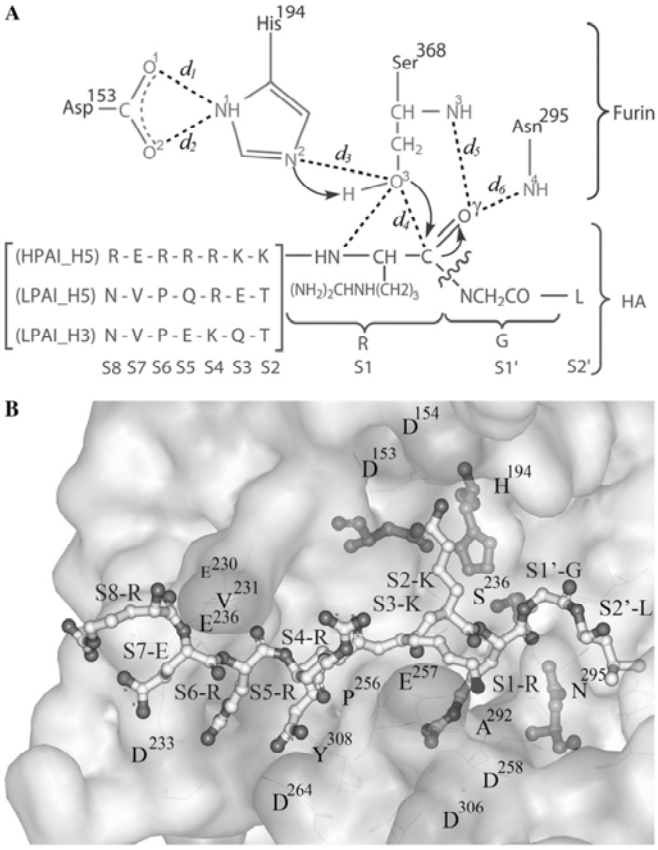


FIGURE 1 (*A*) Proposed cleavage mechanism of HA by furin and definitions of d_1–d_6. (*B*) Loop of HPAI H5 (ball and stick model) in the electrostatic surface of furin.

structure (residues 322–331: NVPEKQTQGL) of the HA0 of H3 (PDB entry code: 1HA0) (20) and dec-RVKR-cmk inhibitor of FR (12) as a template, by using the homology module in the Insight II program (Accelrys, San Diego, CA) (21). The Needleman-Wunsch algorithm (22) was used for pairwise alignment to identify the correspondence region. The coordinates of the structurally conserved regions (SCRs) of the template —in which amino acid sequences in the template and the model are identical—were copied to be those of the model. In addition to the similar SCRs, only the backbone coordinates were replicated and the side chain atoms were added in a manner that preserves the residue type of the model protein.

For LPAI subtype H5, the initial structure of the cleavage loop (NVPQRETRGL) (19) was constructed using the backbone atoms of the HPH5 loop built previously. The side chains were generated by the LEaP module of AMBER 7 (23). With this module, the H3 loop above was mutated at residue 329 from glutamine (Q) to arginine (R).

Initial structure of enzyme-substrate complexes

The cleavage loop of HA complexed with FR was generated by the following procedures (Fig. 1): i), all HA heavy atoms of S1–S4 were superimposed with the crystal structure of the dec-RVKR-cmk inhibitor whereas the HA backbone atoms of S5–S8 and S1′–S2′ were superimposed with the HA0 loop of H3. The coordinates of S1–S4 residues were changed according to the –RVKR– coordinates of inhibitor leading to a newly formed conformation of the HA loop; ii), this loop was then placed manually into the FR active site by superimposition between the –R_KR– residues of the HA loop and the dec-RVKR-cmk with a creation of the furin-substrate complex; iii), the complex was minimized by keeping the S1–S4 residues and the FR fixed; and iv), finally, three steps of the restrained MD simulations at 298 K were carried out for relaxing the modeled systems with the restrain factors of 10, 5, and 2.5 kcal $mol^{-1}\cdot Å^2$ for 200 ps, 200 ps, and 200 ps, respectively. Therefore, the conformation of each cleavage loop was adapted (from the initial model) to fit better with the FR cavity. The last snapshot obtained from restrained MD procedure was used as the starting structure of the substrate-enzyme complex for the next MD simulations with all atoms allowed to move freely for 2 ns.

Molecular dynamics simulations

Three MD simulations for the HA cleavage loop complexed with furin, HPH5–FR, LPH5–FR, and LPH3–FR, were carried out where their initial structures were generated as described above. The simulated systems were neutralized by 6, 11, and 11 Na^+ ions and solvated by TIP3P water molecules leading to total atoms of 54,500, 56,663, and 56,667 for the HPH5–FR, LPH5–FR, and LPH3–FR complexes, respectively. The dimensions of the obtained simulation boxes for the three systems are 86 Å × 90 Å × 88 Å. The periodic boundary condition with the NPT ensemble was used. Energy minimization and MD simulations were carried out using the SANDER module of AMBER 7 (23) with the Cornell force field (24). A Berendsen coupling time of 0.2 ps was used to maintain the temperature and standard

pressure of the system (25). The SHAKE algorithm (26) was applied to constrain all bonds involving hydrogens. The simulation time step of 2 fs was used. All MD simulations were run with a 12 Å residue-based cutoff for nonbonded interactions and the particle mesh Ewald method was used for an adequate treatment of long-range electrostatic interactions (27). The MD simulations were carried out for 2.0 ns where the production phase starts from 0.75 ns to 2.0 ns. The convergences of energies, temperature, pressure, and global root mean-square displacement (RMSD) were used to verify the stability of the systems. The MD trajectories were collected every 0.2 ps. Analysis of all MD trajectories, i.e., RMSD, distances, hydrogen bonds, etc., were carried out using the Ptraj, CARNAL, and MMGB/SA modules of AMBER 7.

RESULTS AND DISCUSSION

Overall enzyme-substrate structure

To monitor the stability of the three simulated systems, HPH5–FR, LPH5–FR, and LPH3–FR, RMSD of the structures obtained during the 2 ns MD simulations relative to the initial structure of all heavy atoms in the substrate-enzyme complex and the 10 residues of HA substrate were evaluated and plotted in Fig. 2. It can be seen from the plot of RMSD plots versus simulation time that all three systems were found to reach equilibrium at 0.75 ns.

Enzyme-substrate conformation

The main results refer to the schematic representation shown in Fig. 1 where the conformation structures of the complexes are described in terms of d_1–d_6 (Fig. 3) in which the reaction coordinates involved in the acylation in three regions, the catalytic triad of FR, nucleophilic attack, and oxyanion hole, were described by the d_1–d_3, d_4, and d_5–d_6, respectively (Fig. 1). Accordingly, percentage H-bond occupations were evaluated using the CARNAL module in AMBER7 based on the following criteria: i), proton donor-acceptor distance $\leq$3.5 Å; and ii), donor-H-acceptor bond angle $\geq$120°. The results were summarized in Table 1.

At the catalytic triad (Asp^{153}, His^{194}, and Ser^{368}) region of the HPH5–FR, the simulated distances of $d_1 = d_2 = d_3 =$ 2.85 Å (Fig. 3) are comparable to those found experimentally for the inhibitor-furin complex of 2.53 Å, 3.10 Å, and 2.95 Å, respectively (12). This indicates that the three catalytic residues, Asp^{153}, His^{194}, and Ser^{368}, are in the configuration suitable for initiating the nucleophilic reaction. The situation is different for the bound LPAI, H5, and H3, hemagglutinin loops, where only d_3 of both systems and d_1 of LPH5–FR fall within the range whereby the reaction can take place.

The nucleophilic attack is determined by the distance between the O^{γ}-hydroxyl oxygen of Ser^{368} and the carbonyl carbon of the S1-Arg, d_4 in Fig. 1. The d_4 bond-making distances were found to be at 3.05 Å, 3.45 Å, and 3.75 Å for the HPH5–FR, LPH5–FR, and LPH3–FR systems, respectively (Fig. 3). In addition to a short d_4, the detected sharp and narrow peak was found only in the HPH5–FR complex, signifying the rigidity of the complex that thus serves as a more appropriate configuration for the nucleophilic attack. In contrast, for the other two complexes the d_4 peak shows a broad distribution and takes place at significantly longer distances.

At the oxyanion hole region of the FR active site (d_5 and d_6 in Fig. 1), the hole is generally formed by the backbone nitrogen of Ser^{368} and the carboxamide nitrogen of Asn^{295}, which specifically binds with the carbonyl group of the S1 reacting residue (12). In Fig. 3, the d_5 distance of 2.95 Å for the HPH5–FR (*solid line*) is slightly shorter than that of 3.15 Å for the LPH5–FR (*dashed line*) with the presence of a strong hydrogen bond to Ser^{368}, 100% and 86% occupations (d_5 in Table 1), respectively. The percentage occupation of 15% for the LPH3–FR confirms a very weak interaction and highly flexible of the complex. This conclusion was also confirmed by a broad distribution of d_5 distance varying from 3.00 Å to 5.85 Å (Fig. 3). In terms of d_6 distance, the distribution plot for the HPH5–FR complex shows two apparent peaks centered at 2.85 Å and 5.45 Å where the sharper first peak indicated a preferential interaction. For LPAI systems, the d_6 distance is detected at 3.05 Å in H3 whereas this distance is significantly longer in H5. This means that interaction between S1 residue and Asn^{295} was almost lost in the LPH5–FR system. The observed intermolecular distances agree very well with the percentage pair of hydrogen bond through the d_5 and d_6 in which those values ($d_5 = 100\%$ and $d_6 = 34\%$) for the HPH5–FR are higher than those ($d_5 = 86\%$ and $d_6 = 6\%$) for the LPH5–FR. On the contrary, bonding via these two distances almost disappear for the LPH3–FR complex. Note that the experimental d_5 and d_6 distances for the inhibitor-furin complex are 3.34 Å and 2.78 Å, respectively (12).

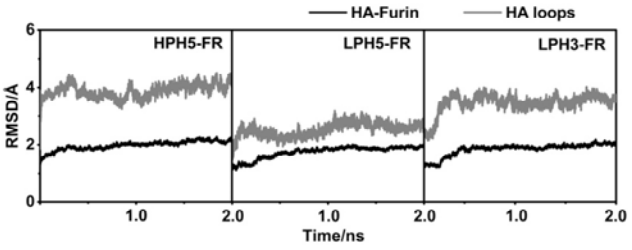


FIGURE 2 RMSDs relative to the initial structure for all heavy atoms of the substrate-furin complexes (*black*) and the HA loops (*gray*) for the three systems studied: HPH5–FR, LPH5–FR, and LPH3–FR.

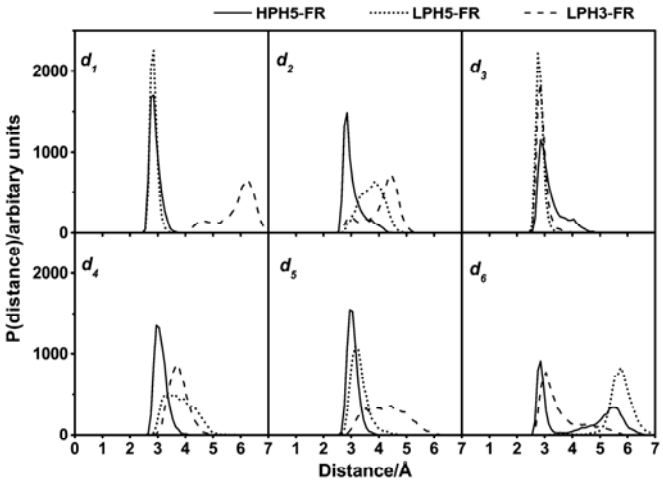


FIGURE 3 Distributions of the d_1–d_6 distances defined in Fig. 1 for the three simulated systems, sampling from 0.75 to 2.0 ns in MD simulations.

H-bonds of the hemagglutinin loop

To assess to molecular interaction at the cleavage loop of hemagglutinin, percentage, and number of hydrogen bond between each HA residue and the active site residues of FR, were evaluated and plotted in Fig. 4. The criteria used are the same as those mentioned earlier. Here, the experimentally detected hydrogen bonds for the inhibitor-furin complex (12) were also given for comparison, the residue with a box around the amino acid label. Interest was focused on S1-Arg of HA and FR's flat groove—a cavity formed by residues Pro^{256}, Asp^{258}, Ala^{292}, and Asp^{306}. The latter was found to play an important role in facilitating the cleavage reaction by forming hydrogen bonds to the guanidinium side chain (Fig. 1) of P1-Arg of the inhibitor (12) (Si denotes a residue of the substrate whereas Pi represents one of the inhibitors, where $i = 1, 2, 3, \ldots$) (28). Strong hydrogen bonds between the four flat groove residues and S1-Arg were highly detected for the HPH5–FR complex (S1 in Fig. 4). The cleavage reaction in the HPH5–FR system was also promoted by the other hydrogen bonds from S1-Arg to Ser^{368} (two bonds, one of which is d_5 in Fig. 1) and Ser^{253} (one bond) of FR with almost 100% occupation. These observations are strongly supported by the experimentally detected hydrogen bonds also shown in Fig. 1. This is in contrast with LPH5–FR in which no hydrogen bond with the flat groove was detected and only a single bond with the catalytic Ser^{386} (d_5 in Fig. 1) was formed. The LPH3–FR loop shows much less interaction with the FR residues, forming no hydrogen with either the flat groove or the catalytic Ser^{368} residue.

Some comments could be made concerning the two hydrogen bonds formed between the reacting S1-Arg and the catalytic Ser^{386} of HPH5–FR (Fig. 4). Besides the bond defined by d_5, another bond is $S^{368}\_O^3\ldots NH\_S1\text{-}Arg$ (71% occupation). This hydrogen bond that has never been found in the oxyanion hole region of any other HA subtype or in the inhibitor-furin complex (12), could be a reason for the shortening and rigidity (narrow and sharp peaks shown in Fig. 3) of the intermolecular d_4 distance, which could directly facilitate the nucleophilic reaction in HPH5–FR.

Considering the role of the –RRRKK– insertion (Fig. 1), more hydrogen bonds and a higher percentage occupation between the S2-S6 residues of HA and the surrounding residues of FR (Fig. 4) were found for HPH5–FR in comparison with the two LPAI systems. This means that the –RRRKK– insertion can directly help to hold the substrate in place.

Taking into account all the hydrogen bond data given above, the hemagglutinin loop of HPH5–FR is observed to bind much more tightly into the catalytic site of FR than the LPH5–FR and LPH3–FR systems.

TABLE 1 Percentage of hydrogen bond according to the d_1–d_3 and d_5–d_6 distances defined in Fig. 1 for the three simulated systems

Distance	H-bond	HPH5–FR	LPH5–FR	LPH3–FR
d_1	$D^{153}\_O^1\ldots HN^1\_H^{194}$	100	100	0
d_2	$D^{153}\_O^2\ldots HN^1\_H^{194}$	96	75	32
d_3	$H^{194}\_N^2\ldots HO^3\_S^{368}$	81	90	98
d_5	$S1\text{-}R\_\ O^\gamma\ldots HN^3\_\ S^{368}$	100	86	15
d_6	$S1\text{-}R\_\ O^\gamma\ldots HN^4\_N^{295}$	34	6	34
	$S^{368}\_O^3\ldots HN\_S1\text{-}R$	71	10	0

Per residue enzyme-substrate interactions

To seek the fundamental basis of substrate-furin interactions, the interaction energy between individual FR residue and the three HA loops and vice versa were calculated using the decomposition energy module of AMBER 7. Plots of the decomposition energies (DC) of the HA residues, as well as the selected FR residues that are located within 5 Å from the 10 residues of HA loop, were shown in Fig. 5. The overall

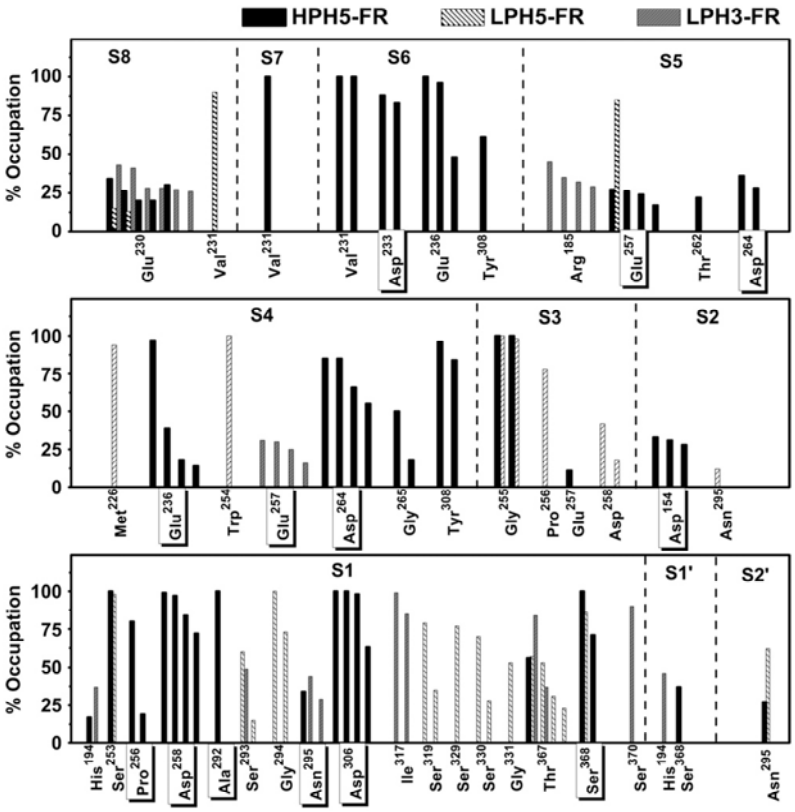


FIGURE 4 Percent occupations of hydrogen bonds between furin and the ten HA residues (defined in Fig. 1) where the residues with a box around the label represent experimentally detected bonds for the inhibitor-furin complex.

DCs of FR residues were observed in the following order: HPH5–FR ≫ LPH5–FR ~ LPH3–FR (Fig. 5 *B*). For the HA loops, the DCs for the residues of HPH5–FR are significantly lower than for the corresponding residues of LPH5–FR and LPH3–FR (Fig. 5 *A*). This is especially true for the S1, S4, and S6 residues. The calculated results are consistent with experimental results for the inhibitor-furin complex, which stated that P1, P4, and P6 residues of the inhibitor were observed to interact strongly with FR (11,29). In addition, these two P4 and P6 residues are required to dramatically increase the cleavage reaction (30). The DCs for the FR and HA residues agree very well with the hydrogen bond data shown (Fig. 4) and discussed above.

Solvation structure

Hydrophilicity that is determined by the ligand solution in the cavity region of enzyme-substrate complex is known to play an essential role in the catalytic mechanism of biological systems. To seek this information, the plot of the radial distribution functions (RDFs, $g_{xy}(r)$)—the probability of finding the particle of type *y* in a spherical radii, *r*, around the particle of type *x*—was evaluated. Here, the RDFs from the acceptor atoms of FR pocket (O^1, O^2:Asp^{153}; N^1, N^2:His^{194}; N^4:Asn^{295}, and N^3, O^3:Ser^{368}) and the reactive HA residue (O^γ:S1-Arg) (see label in Fig. 1), to oxygen atom of water was calculated. The results for the three simulated systems are shown in Fig. 6 together with the corresponding running integration numbers up to the first minimum of the RDF plot.

The RDFs for all oxygen atoms, O^1, O^2, O^3, and O^γ, of the LPH3–FR complex show sharp and narrow first peaks at ~2.9 Å with the integration number (coordination number, CN) of 1.5, 1, 1 and 2 water molecules, respectively (Fig. 6). The CN denotes the number of water molecules that locate in the first hydration shell around the central atom. This is in contrast for the LPH5–FR complex where only O^2 can be accessed by water molecule whereas no water was detected around the four oxygen atoms of HPH5–FR.

The RDFs centered on N^1 and N^4 atoms of the three systems are almost similar, the first broad peaks at ~3.8 Å and sharp peaks at ~3 Å, respectively. Note that, N^3 atom of only LPH3–FR was solvated by water molecule. Interest is focused

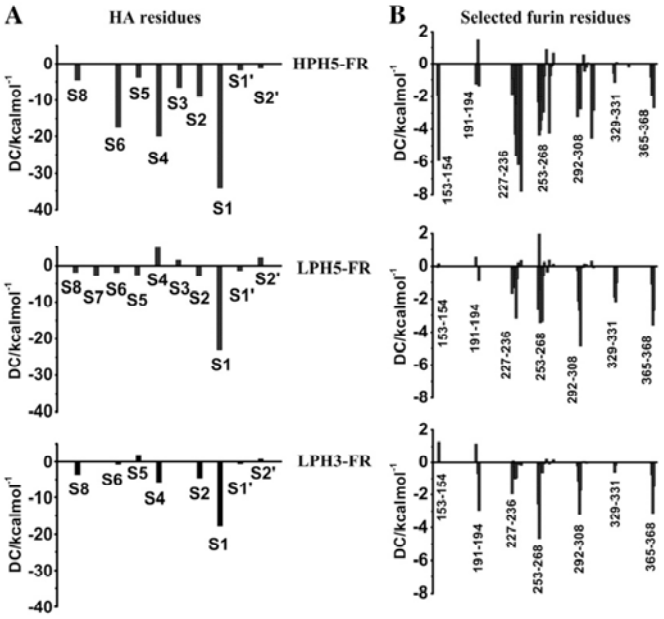


FIGURE 5 Decomposition energies (DC) of (*A*) 10 HA residues (defined in Fig. 1), and (*B*) selected furin residues within 5 Å of those 10 HA residues.

to N^2 atom of His^{194} where the catalytic water was observed and proposed to play role in the reaction mechanism of serine protease (14). In good agreement with our simulated results, one water (CN = 1) was detected to preferentially coordinate (first peak centered at ~2.9 Å) to N^2 atom of only for the HPH5–FR complex. This is not the case for the other two low-pathogenic systems, LPH3–FR and LPH5–FR, where less water, 0.5 and 0.2 were, respectively, was found (Fig. 6, *dashed* and *dotted lines*). In addition, distance to the LPH3-FR water molecule of ~3.45 Å is significantly longer than that of the HPH5–FR complex.

Taking into account the solvation data shown above, the number of overall water molecules around the selected central atoms are in the following order: LPH3–FR > LPH5–FR > HPH5–FR. The obtained data indicates that more water molecules can take place in the catalytic pocket of FR in the complex whose substrate does not fit well with the FR active site, giving a subsequent result of more cavities or less binding between substrate and enzyme. This suggestion agrees very well with the conformational structures of the complexes (Fig. 3) and the hydrogen bond data (Fig. 4) shown and discussed previously.

CONCLUSIONS

The technique of molecular dynamics simulations has been applied to look for detailed information focused on the intra- and

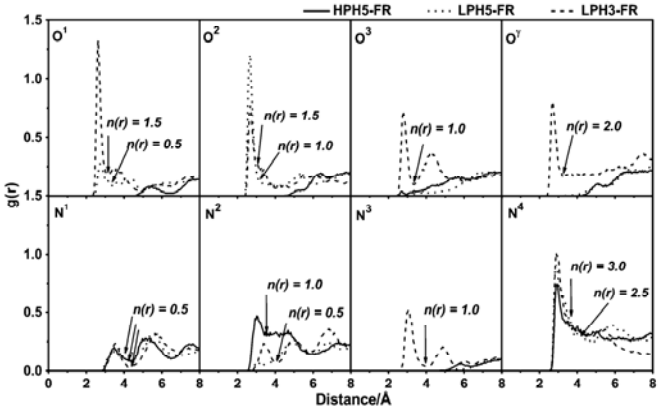


FIGURE 6 Radial distribution functions, $g(r)$, from acceptor atoms of furin catalytic residues and HA loops (see Fig. 1 for definition) to oxygen atoms of water molecules for the three simulated systems, HPH5–FR, LPH5–FR, and LPH3–FR complexes.

intermolecular interactions and geometries of the three substrate-furin complexes, HPH5–FR, LPH5–FR, and LPH3–FR. The results obtained from this study indicate that the –RRRKK– insertion in the HPH5–FR (two arginines at S4 and S6 positions in particular) helps to directly hold the HA cleavage loop in place by forming many strong hydrogen bonds between the residues of HA and FR. This data is well supported by the corresponding strong-decomposition interaction energies and leads consequently to the accumulation of less water molecules accessible to the FR cavity. An active conformation of the HPH5–FR complex suitable for the acylation reaction by FR was formed. In conclusion, the results from this study provide a clear answer to the question of why HPAI hemagglutinin H5 is better cleaved by FR than the other two HA subtypes.

P.D. thanks the Sandwich Ph.D. Program and T. R. thanks the Post Doctoral Program from the Commission on Higher Education. The Computer Center for Advanced Research and the Computational Chemistry Unit Cell, Faculty of Science, Chulalongkorn University provided research facilities, software packages, and computing times.

This work was supported by the Thailand Research Fund.

REFERENCES

1. Horimoto, T., and Y. Kawaoka. 2001. Pandemic threat posed by avian Influenza A viruses. *Clin. Microbiol. Rev.* 14:129–149.
2. Horimoto, T., and Y. Kawaoka. 1995. The hemagglutinin cleavability of a virulent avian influenza virus by subtilisin-like endoproteases is influenced by the amino acid immediately downstream of the cleavage site. *Virology.* 210:466–470.
3. Horimoto, T., and Y. Kawaoka. 2005. Influenza: lessons from past pandemics, warnings from current incidents. *Nat. Rev. Microbiol.* 3:591–600.
4. Steinhauer, D. A. 1999. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology.* 258:1–20.
5. Cross, K. J., L. M. Burleigh, and D. Steinhauer. 2001. Mechanisms of cell entry by influenza virus. *Expert Rev. Mol. Med.* 6:1–18.
6. Stevens, J., O. Blixt, T. M. Tumpey, J. K. Taubenberger, J. C. Paulson, and I. A. Wilson. 2006. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science.* 312:404–410.
7. Neumann, G., and Y. Kawaoka. 2006. Host range restriction and pathogenicity in the context of influenza pandemic. *Emerg. Infect. Dis.* 12:881–886.
8. Basak, A., M. Zhong, J. S. Munzer, M. C. Tien, and N. G. Seidah. 2001. Implication of the proprotein convertases furin, PC5 and PC7 in the cleavage of surface glycoproteins of Hong Kong, Ebola and respiratory syncytial viruses: a comparative analysis with fluorogenic peptides. *Biochem. J.* 353:537–545.
9. Stieneke-Gröber, A., M. Vey, H. Angliker, E. Shaw, G. Thomas, C. Roberts, H. D. Klenk, and W. Garten. 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J.* 11:2407–2414.
10. Bergeron, F., R. Ledue, and R. Day. 2000. Subtilase-like pro-protein convertases: from molecular specificity to therapeutic applications. *J. Mol. Endocrinol.* 24:1–22.
11. Holyoak, T., C. A. Kettner, G. A. Petsko, R. S. Fuller, and D. Ringe. 2004. Structural basis for differences in substrate selectivity in kex2 and furin protein convertases. *Biochemistry.* 43:2412–2421.
12. Henrich, S., A. Cameron, G. P. Bourenkov, R. Kiefersauer, R. Huber, I. Lindberg, W. Bode, and M. E. Than. 2003. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat. Struct. Biol.* 10:520–526.
13. Thomas, G. 2002. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Mol. Cell. Biol.* 3:753–766.
14. Molloy, S. S., P. A. Bresnahan, S. H. Leppla, K. R. Klimpel, and G. Thomas. 1992. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J. Biol. Chem.* 267:16396–16402.
15. Hosako, M., M. Nagahama, W.-S. Kim, T. Watanabe, K. Hatsuzawa, J. Ikemizu, K. Murakami, and K. Nakayama. 1991. Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. *J. Biol. Chem.* 266:12127–12130.
16. Ishida, T., and S. Kato. 2003. Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. *J. Am. Chem. Soc.* 125:12035–12048.
17. Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chem. Rev.* 102:4501–4523.
18. Chutinimitkul, S., T. Songserm, A. Amonsin, S. Payungporn, K. Suwannakarn, S. Damrongwatanapokin, A. Chaisingh, B. Nuansrichay, T. Chieochansin, A. Theamboonlers, and Y. Poovorawan. 2007. New strain of influenza A virus (H5N1), Thailand. *Emerg. Infect. Dis.* 13:506–507.
19. Ha, Y., D. J. Stevens, J. J. Skehel, and D. C. Wiley. 2002. H5 avian and H9 swine influenza virus hemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* 21:865–875.
20. Chen, J., K. L. Lee, D. A. Steinhauer, J. Stevens, J. J. Skehel, and D. C. Wiley. 1998. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell.* 95:409–417.
21. Accelrys Software Inc. Insight II Environment. 2006. Accelrys Software, San Diego.
22. Durbin, R., S. Eddy, A. Krogh, and G. Mitchison. 2004. Needleman-Wunsch Algorithm. Cambridge University:1–3.
23. Case, D. A., J. W. Pearlman, and T. E. Caldwell. J. C. III, W. S.Wang, C. L. Ross, T. A. Simmerling, K. M. Darden, R. V. Merz, A. L. Stanton, J. J. Cheng, M. Vincent, V. Crowley, H. Tsui, R. J. Gohlke, Y. Radmer, J. Duan, I. Pitera, G. L. Massova, U. C. Seibel, P. K. Singh, and P. A. Kollman. 2002. AMBER 7. University of California, San Francisco, CA.
24. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, J. W. C. T. Fox, and P. A. Kollman. 1995. A second generation forcefield for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* 117:5179–5197.
25. Berendsen, H. J. C., J. P. M. Postma, W. F. V. Gunsteren, and A. DiNola. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
26. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1997. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.
27. York, D. M., T. A. Darden, and L. G. Pedersen. 1993. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J. Chem. Phys.* 99:8345–8348.
28. Schechter, I., and A. Berger. 1968. On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* 32:898–902.
29. Rockwell, N. C., D. J. Krysan, T. Komiyama, and R. S. Fuller. 2002. Precursor processing by kex2/furin proteases. *Chem. Rev.* 102:4525–4548.
30. Nakayama, K. 1997. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem. J.* 327:625–635.

APPENDIX A2: SUBMITTED ARTICLE

Theoretical Studies on the Molecular Basis of HIV-1RT/NNRTIs Interactions

Panita Decha[†], *Thanyarat Udommaneethanakit*[†], *Pathumwadee Intharathep*[†], *Pornthep Sompornpisut*[†], *Supot Hannongbua*[†], *Peter Wolschann*[‡] *and Vudhichai Parasuk*[*, †]

[†]Department of Chemistry, Faculty of Science, Chulalongkorn University, Phayathai Road, Patumwan, Bangkok 10330, Thailand. [‡]Institute of Theoretical Chemistry, University of Vienna, Waehringer Strasse 17, Vienna 1090, Austria.

*Corresponding author: phone: +66 22 187602; fax: +66 22 187603; e-mail: vudhichai.p@chula.ac.th

ABSTRACT Molecular dynamics simulations (MD) of the 117 kDa HIV-1 RT complexed with one of each of the four NNRTIs, EFV, EMV, ETV and NVP, were performed to examine the structures, binding free energies and the importance of water molecules in the binding site in relation to the experimentally derived IC_{50} values of the drugs. In terms of hydrogen bonding, EFV and EMV bind much more tightly into the catalytic site of HIV-1 RT than the ETV system, whilst for the least effective NVP drug no such a hydrogen bond was found. The binding free energy, calculated using MM-PBSA, was found to decrease in the following order: EFV ~ ETV > EMV > NVP. The decrease in stability of HIV-1 RT/NNRTI complexes is in good agreement with the experimentally derived IC_{50} values. On the basis of the calculation of the interaction energies, the binding affinity of the protein-inhibitor complexes was found to be essentially associated with the cluster of seven hydrophobic residues, L100, V106, Y181, Y188, F227, W229 and P236, and two basic residues, K101 and K103. Moreover, these residues are considered to be the most frequently detected mutated amino acids during treatment by various NNRTIs and, therefore, those most likely to have been selected in the population for resistance. In addition, the distribution and binding of water molecules were investigated and discussed, referring to those found in the crystal structures.

Decha, P., T. Udommaneethanakit, P. Intharathep, P. Sompornpisut, S. Hannongbua, P. Wolschann, and V. Parasuk. 2009. Theoretical studies on the molecular basis of HIV-1RT/NNRTIs interactions J. Enzym. Inhib. Med. Chem.

APPENDIX A3: PREPARATION

Determinants of the Inserted Sequences of High Pathogenicity H5N1 Hemagglutinin Specificity with Furin

*Panita Decha[1], Thanyada Rungrotmongkol[1], Maturos Malaisree[1], Nadtanet Nunthaboot[2], Pathumwadee Intharathep[1], Nopporn Kaiyawet[1], Vudhichai Parasuk[1], Pornthep Sompornpisut[1], Peter Wolschann[3] and Supot Hannongbua[1]**

[1]Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. [2]Department of Chemistry, Faculty of Science, Mahasarakhamn Universit, Mahasarakhamn 44150, Thailand. [3]Institute of Theoretical Chemistry, University of Vienna, Vienna A-1090, Austria.

[*]Corresponding author: phone: (66)-22-187602; fax: (66)-22-187603; E-mail: supot.h@chula.ac.th

Abstract Molecular dynamics simulations have been applied on furin complexed with the four different inserted sequences of high pathogenicity H5N1 hemagglutinin at cleavage site that identified during the 2004-2006 influenza outbreaks in Thailand: FR-H5Sq1, FR-H5Sq2, FR-H5Sq3 and FR-H5Sq4, aiming to compare the substrate specificity and recognition of furin. The simulated results of hydrogen bonds, decomposition interaction energies, binding energies and water accessibilities suggested that the cleavage loops of the FR-H5Sq1, FR-H5Sq2 and FR-H5Sq3 were found to bind strongly to the furin binding pocket and fit well within the cavity compared to that of FR-H5Sq4. Therefore, the deletion of an amino acid at the cleavage site of HPAI-H5 has caused a decrease in substrate specificity and recognition of furin.

Decha, P., T. Rungrotmongkol, M. Malaisree, N. Nunthaboot, P. Intharathep, N. Kaiyawet, V. Parasuk, P. Sompornpisut, P. Wolschann, and S. Hannongbua. 2009. Determinants of the inserted sequences of high pathogenicity H5N1 hemagglutinin specificity with furin. Preparation.

APPENDIX B: PRESENTATIONS AND WORKSHOPS

Oral presentations

1. **Decha, P**., et al. 2009. Determinants of substrate specificity of furin on the inserted sequences of high pathogenic H5N1 hemagglutinin: a MD study. ANSCSE 13. Kasetsart University, Bangkok, Thailand. 25-27, 2009.
2. **Decha, P**., et al. 2008. Understanding of Pathogenicity of an Avian Influenza Virus H5N1: Why H5 is Better Cleaved by Furin. CHE-USDC Congress I. Pattaya, Thailand. 5-7 September 2008.
3. **Decha, P**., Malaisree, M., Aruksakunwong, O., and Hannongbua, S. 2006. Effects of Mutations on the Susceptibility of Oseltamivir to Influenza A Virus N1 Neuraminidase as Studied by Molecular Dynamics Simulations. The 32nd Congress on Science and Technology of Thailand. Bangkok, Thailand. 10-12 October 2006.
4. Chaodamrongsakul, J., **Decha, P.** and Vaosoongnern, V. 2003. Solid Polymer Electrolytes for an Application as Novel Energy Storage Materials. The 10th Tri-University International Joint Seminar & Symposium, Mie University, Japan. 17-21 October 2003.

Poster presentations

1. **Decha, P**., Wolschann, P. and Hannongbua, S. 2009. Better Cleaved of Hemagglutinin Subtype H5 of the High Pathogenic Avian Influenza Virus by Furin. Naresuan University, Phitsanulok, Thailand. 14-16 January, 2009.
2. **Decha, P.,** Parasuk, V., Hannongbua, S. and Wolschann, P. 2008. Computational analysis of HIV-1 reverse transcriptase-inhibitors as a molecular basis for drug development. The International Conference on Drug Design and Discovery for Developing Countries. Trieste, Italy. 3-5 July 2008.
3. **Decha, P.,** Wolschann, P. and Hannongbua, S. 2008. Understanding of High Pathogenicity of the Avian Influenza Virus H5N1: Why H5 is Better Cleaved by Furin. Computational Biophysics to Systems Biology. Juelich, Germany. 19-21 May 2008.

4. **Decha, P.,** et al. 2007. Structure and Solvation of Furin-Hemagglutnin Complex by Molecular Dynamics Simulations. The 33rd Congress on Science and Technology of Thailand. Nakhon Si Thammarat, Thailand. 18-20 October 2007.
5. **Decha, P.** and Vaosoongnern, V. 2002. Effect of Adding Poly(propylene oxide) on Structures and Properties of Solid Polymer Electrolytes: Poly(ethylene oxide)/Salt System. The 28th Congress on Science and Technology of Thailand. Bangkok, Thailand. 24-26 October 2002.
6. **Decha, P.** and Vaosoongnern, V. 2002. Conformational Properties of Poly(propylene oxide). The 3rd National Symposium on Graduated Research. Suranaree University of Technology, Thailand. 18-19 July 2002.

Workshops and Research

1. The Commission on Higher Education scholarship under the program "Strategic Scholarships for Frontier Research Network". Structure and Solvation of HIV-1-RT/inhibitors and Furin/HA-substrate by Molecular Dynamics Simulations. University of Veinna, Austria. 13 September 2007 – 3 September 2008.
2. Training Course on "Molecular Design and Computer-assisted Combinatorial Chemistry". The International Centre for Science and High Technology-the United Nations Industrial Development Organization (ICS-UNIDO). Trieste, Italy. 30 June -2 July 2008.
3. Summer School Drug Design. Department of Medicinal Chemistry University of Vienna. Vienna, Austria. 16-21 September 2007.
4. The 3rd Thai Summer School of Computational Chemistry. Columbus in Bangkok - A Multireference Methods Workshop. Burapha University, Bang Saen , Thailand. April 2-5, 2006

VITAE

NAME: Miss Panita Decha

BORN: 27 April 1976 in Nakhon Si Thammarat, Thailand

EDUCATION:

1995-1998	B. Sc. in Chemistry, Ramkhamhaeng University, Bangkok, Thailand
2000-2002	M. Sc. in Chemistry, Suranaree University of Technology, Nakhonratchasima, Thailand
2005-2009	Ph. D. in Chemistry, Chulalongkorn University, Bangkok, Thailand

PUBLICATIONS:

1. **Decha, P.,** et al., **2008.** Source of high pathogenicity of an avian influenza virus H5N1: why H5 is better cleaved by furin. Biophys. J. 95:128-134.
2. Rungrotmongkol, T., **Decha, P.,** et al., **2008.** Comment on "Cleavage mechanism of the H5N1 hemagglutinin by trypsin and furin". Amino Acids. 35:511-512
3. Malaisree, M., T. Rungrotmongkol, **Decha, P.,** et al., **2008.** Understanding of known drug-target interactions in the catalytic pocket of neuraminidase subtype N1. Proteins. 71:1908-1918.
4. Aruksakunwong, O., Malaisree, M., **Decha, P.,** et al., **2007.** On the Lower susceptibility of oseltamivir to influenza neuraminidase subtype N1 than to in N2 and N9. Biophys. J. 92:798-807.
5. Intharathep, P., Laohpongspaisan, C., Rungrotmongkol, T., Loisruangsin, A., Malaisree, M., **Decha, P.,** et al., **2008.** How amantadine and rimantadine inhibit proton transport in the M2 protein channel. J. Mol. Graphics Modell. In Press.
6. Rungrotmongkol, T., **Decha, P.,** et al., **2008.** Combined QM/MM mechanistic study of the acylation process in furin complexed with the H5N1 avian influenza virus hemagglutinin's cleavage site. Protein. In Press.
7. Laohpongspaisan, C., Rungrotmongkol, T., Intharathep, P., Malaisree, M., **Decha, P.,** et al., **2008.** Why amantadine loses its function in influenza M2 mutants: MD simulations. J. Chem. Inf. Model. In Press.
8. Malaisree, M., Rungrotmongkol, T., Nunthaboot, N., Aruksakunwong, O., Intharathep, P., **Decha, P.,** et al., **2008.** Source of oseltamivir resistance in avian influenza H5N1 virus with the H274Y mutation. Amino Acids. In Press.
9. **Decha, P.**, et al., **2009**. Theoretical studies on the molecular basis of HIV-1RT/NNRTIs interactions J. Enzym. Inhib. Med. Chem. Submitted.