

การจดจำของเฮแมกกลูทีนินในไวรัสไข้หวัดใหญ่ที่พบในสัตว์ปีกต่อไซแอลิก-กาแล็กโทสของ  
มนุษย์โดยใช้เทคนิคอินซูลิโค



นาย นพพร ไกยเวช

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

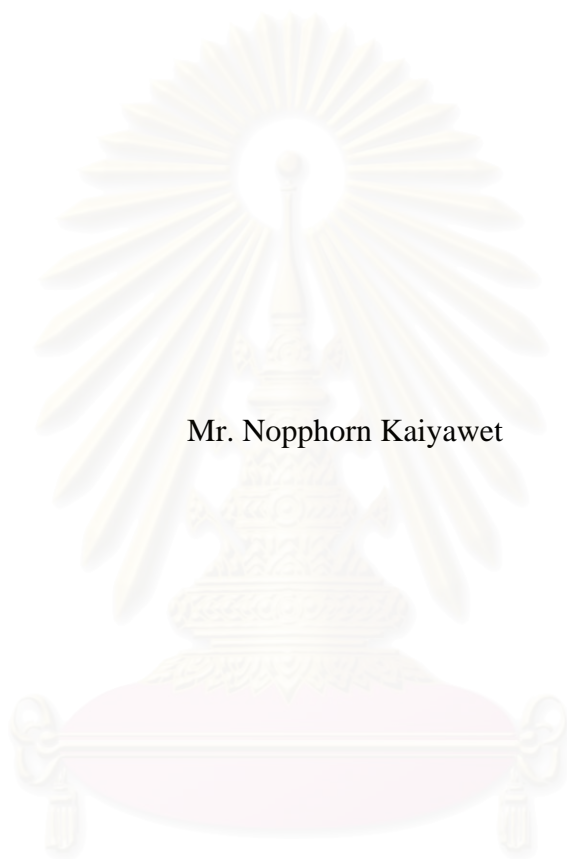
สาขาวิชาเคมี ภาควิชาเคมี

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2551

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

RECOGNITION OF HEMAGGLUTININ IN AVIAN INFLUENZA VIRUS TO  
HUMAN SIALIC-GALACTOSE USING *IN SILICO* TECHNIQUE



Mr. Nopphorn Kaiyawet

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

Department of Chemistry

Faculty of Science

Chulalongkorn University

Academic Year 2008

Copyright of Chulalongkorn University



นพพร ไกยเวช : การจดจำของเฮแมกกลูทีนินในไวรัสไข้หวัดใหญ่ที่พบในสัตว์ปีกต่อ  
 ไซแอลิก-กาแล็กโทสของมนุษย์โดยใช้เทคนิคอินซิลิโค. (RECOGNITION OF  
 HEMAGGLUTININ IN AVIAN INFLUENZA VIRUS TO HUMAN  
 SIALIC-GALACTOSE USING IN SILICO TECHNIQUE) อ.ที่ปรึกษา  
 วิทยานิพนธ์หลัก : ศ. ดร. สุพจน์ หารหนองบัว, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร.  
 พรเทพ สมพรพิสุทธิ์, 60 หน้า.

โปรตีนเฮแมกกลูทีนินของไวรัสไข้หวัดใหญ่มีบทบาทสำคัญต่อการยึดจับกับตัวรับที่อยู่  
 บนผิวเซลล์ การติดต่อกันของเชื้อไวรัสไข้หวัดใหญ่สายพันธุ์ H5N1 จากสัตว์ปีกมาสู่มนุษย์นั้น พบ  
 ได้น้อยเนื่องจากการยึดจับระหว่าง โปรตีนเฮแมกกลูทีนิน H5 กับ ตัวรับชนิดไซแอลิก- $\alpha$ 2,6-กา  
 แล็กโทส ของมนุษย์ยังไม่แข็งแรง พบว่าการกลายพันธุ์ของโปรตีนเฮแมกกลูทีนินที่ตำแหน่ง 129  
 และ 222 ช่วยเพิ่มประสิทธิภาพการยึดจับที่มีต่อตัวรับชนิดไซแอลิก- $\alpha$ 2,6-กาแล็กโทส มากขึ้น  
 การเปรียบเทียบชนิดของกรดอะมิโนภายใน 5 อังสตรอมจากบริเวณยึดจับของโปรตีนเฮแมกกลู  
 ทีนินชนิด H3 และ H5 พบความแตกต่างที่สำคัญในรูปของชนิดกรดอะมิโนและโครงสร้าง  
 งานวิจัยนี้ใช้เทคนิคการออกแบบเชิงโมเลกุลและการจำลองพลวัตเชิงโมเลกุล ในการศึกษาการยึด  
 จับของกรดอะมิโนที่มีผลต่อตัวรับชนิดไซแอลิก- $\alpha$ 2,6-กาแล็กโทส การศึกษานี้ใช้การจำลอง  
 พลวัตเชิงโมเลกุลสำหรับโปรตีนชนิดโวลด์ไทป์และชนิดกลายพันธุ์ ได้แก่ S129N, N182S,  
 E186D, Q222V and G224S ได้คำนวณและเปรียบเทียบค่าพลังงานการยึดจับระหว่างตัวรับ  
 ชนิดไซแอลิก- $\alpha$ 2,6-กาแล็กโทส และ โปรตีน H5 ทั้งหมด ผลการคำนวณแสดงว่า S129N และ  
 Q222V ให้พลังงานเสรีของการยึดจับกับตัวรับมากกว่าโปรตีนชนิดโวลด์ไทป์ การค้นพบนี้น่าจะ  
 เป็นประโยชน์ต่อการทำนายถึงตำแหน่งของกรดอะมิโนที่มีความสำคัญต่อการยึดจับซึ่งอาจ  
 เกี่ยวข้องกับการติดต่อจากสัตว์ปีกสู่คนได้อย่างมีประสิทธิภาพ

ภาควิชา.....เคมี.....ลายมือชื่อนิสิต นพพร ไกยเวช.....  
 สาขาวิชา.....เคมี.....ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....  
 ปีการศึกษา.....2551.....ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม พรเทพ สมพรพิสุทธิ์.....

## 5072313723 : MAJOR CHEMISTRY

KEYWORDS : HEMAGGLUTININ / BIRD FLU / RECOGNITION / BINDING  
FREE ENERGY / MOLECULAR DYNAMICS SIMULATIONS

NOPPHORN KAIYAWET: RECOGNITION OF HEMAGGLUTININ IN AVIAN INFLUENZA VIRUS TO HUMAN SIALIC-GALACTOSE USING IN SILICO TECHNIQUE. THESIS ADVISOR: PROF. SUPOT HANNONGBUA, Dr. rer. nat., THESIS CO-ADVISOR: ASST. PROF. PORNTHEP SOMPORNPIST, Ph. D., 60 pp.

Influenza virus hemagglutinin is an essential protein for the binding to host cell receptor. Rare cases of human infection with avian influenza virus subtype H5N1 is thought to be associated with poor binding of avian viral H5 to human sialic acid- $\alpha$ 2,6-galactose receptor (SA $\alpha$ 2,6Gal). It was found that mutations at positions 129 and 222 enhanced specificity of H5N1 viruses to recognize SA $\alpha$ 2,6Gal. Comparisons between residues within 5 Å from the binding pocket of the H3 and those of the H5 have revealed substantial differences in terms of amino acids and structures. In this study, molecular modeling and molecular dynamics simulation techniques have been employed to investigate the binding and recognition roles of these amino acids to SA $\alpha$ 2,6Gal. Molecular dynamics simulations have been carried out for a number of single mutants including S129N, N182S, E186D, Q222V and G224S. Binding energy of the complexes between the SA $\alpha$ 2,6Gal receptor and all the H5 proteins were computed and compared. The results show that S129N and Q222V exhibit a greater binding free energy to the SA $\alpha$ 2,6Gal receptor compared to the wild-type. This finding would be useful for prediction of the critical amino acid residues which are possibly associated with the efficient transmission of H5N1 subtype from avian to human.

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

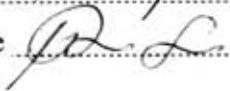
Department : Chemistry

Field of Study : Chemistry

Academic Year : 2008

Student's Signature 

Advisor's Signature 

Co-Advisor's Signature 

## ACKNOWLEDGMENTS

This thesis was completely finished with the excellent helps from my thesis advisor, Prof. Dr. Supot Hannongbua, who always gives me his guiding, advising and encouraging for all my graduate study. This also to my co-advisor Assist. Prof. Dr. Pornthep Sompornpisut

I would like to special thanks to Assoc. Prof. Dr. Sirirat Kokpol, Assist. Prof. Dr. Somsak Pianwanit and Dr. Nadthanet Nunthabut who act as the thesis committee.

Furthermore, I would like to thank Miss Thanyada Rungrotmongkol and Miss Mathuros Malaisree for proved my thesis intensively.

Finally, I would like to thanks Computational Chemistry Unit Cell (CCUC) Department of Chemistry, Faculty of Science, Chulalongkorn University for computer sources and all laboratory members for suggestion.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

# CONTENTS

	Page
<b>ABSTRACT IN THAI</b> .....	iv
<b>ABSTRACT IN ENGLISH</b> .....	v
<b>ACKNOWLEDGMENTS</b> .....	vi
<b>CONTENTS</b> .....	vii
<b>LIST OF TABLE</b> .....	ix
<b>LIST OF FIGURES</b> .....	x
<b>CHAPTER 1 INTRODUCTION</b> .....	1
1.1. Avian influenza virus.....	1
1.2. History of influenza virus .....	2
1.2.1. 1918 Spanish Flu .....	3
1.2.2. 1957 Asian Flu.....	4
1.2.3. 1976 Swine Flu Scare.....	4
1.2.4. 1977 Russian Flu Scare.....	5
1.2.5. 1968 Hong Kong Flu.....	5
1.2.6. 1997 Avian Flu Scare.....	6
1.3. Structure of the influenza virus.....	7
1.4. Specificity and substrate recognition by hemagglutinin.....	11
1.5. The viral genome changing.....	13
1.6. Sequence alignment.....	14
1.7. Representative residue within the catalytic pockets of H5 and H3.....	15
1.8. Scope of this work.....	17
<b>CHAPTER 2 THEORY</b> .....	20
2.1 Molecular mechanics.....	20
2.1.1 Bond stretching .....	21
2.1.2 Bond bending.....	23
2.1.3 Torsion angle.....	23

2.1.4	Out of plane.....	24
2.1.5	Electrostatic energy.....	25
2.1.6	Lennard-Jones energy.....	26
2.2	Molecular dynamics .....	27
2.2.1	Algorithms for Time Dependence.....	28
2.2.2	Periodic boundary condition.....	31
2.2.3	Cutoff methods for the calculation of non-bonding interactions	32
2.2.4	Water models .....	33
2.2.5	Implicit solvation.....	35
2.2.6	Molecular mechanics Poisson-Boltzmann solvent accessibility surface area.....	36
<b>CHAPTER 3 COMPUTATIONAL DETAILS.....</b>		<b>38</b>
3.1	Preparation of the initial structure.....	38
3.2	Molecular dynamics simulations .....	38
3.3	The binding free energy calculations.....	40
<b>CHAPTER 4 RESULTS AND DISCUSSIONS.....</b>		<b>41</b>
4.1	Torsion angle of SA $\alpha$ 2,6Gal pentasaccharide.....	41
4.2	Enzyme-substrate interactions .....	43
4.3	Binding free energy results.....	52
<b>CHAPTER 5 CONCLUSION .....</b>		<b>54</b>
<b>REFERENCES.....</b>		<b>56</b>
<b>VITAE.....</b>		<b>60</b>



## LIST OF TABLES

	Page
<b>Table 1.1</b> Statistical sequence alignment results of hemagglutinin H3 subtype in human host and H5 subtype in avian host.....	15
<b>Table 1.2</b> Representative residues within the catalytic pockets of H5 and H3 (see text for details).....	17
<b>Table 3.1</b> Detailed characteristics of the simulated systems.....	39
<b>Table 4.1</b> Calculated binding free energy and its components for wild-type and mutants of hemagglutinin-SA $\alpha$ 2,6Gal complexes.....	52



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## LIST OF FIGURES

		Page
<b>Figure 1.1</b>	Pandemics of influenza virus subtype H1N1 or “Spanish flu” in 1918.....	3
<b>Figure 1.2</b>	Structural model of an influenza virus particle.....	7
<b>Figure 1.3</b>	Crystal structure of trimer hemagglutinin protein.....	8
<b>Figure 1.4</b>	Crystal structure of tetramer neuraminidase protein.....	9
<b>Figure 1.5</b>	The picture summarizes viral replication process. First, the virus must enter the host cell. Next, the virus will shed its protein coat and begin replication. While the genetic material is replicating, the virus will also produce a new glycoprotein coat. Finally, the new copies of the virus are assembled and released from the cell.	10
<b>Figure 1.6</b>	(a) SA $\alpha$ 2,6Gal and SA $\alpha$ 2,3Gal receptor of human and avian. (b) binding pocket of hemagglutinin protein.....	11
<b>Figure 1.7</b>	Preferential of influenza subtypes in vary hosts.....	12
<b>Figure 1.8</b>	Viral genome changing via genetic shift mechanism.....	14
<b>Figure 1.9</b>	Amino acid residues within a sphere of 5 Å around the hemagglutinin H5 binding pocket.....	16
<b>Figure 2.1</b>	Bond distance between two atoms.....	21
<b>Figure 2.2</b>	The harmonic bond energy term.....	22
<b>Figure 2.3</b>	The Morse function bond energy term.....	22
<b>Figure 2.4</b>	Bond angle between three atoms.....	23
<b>Figure 2.5</b>	Torsion angle or dihedral angle between four atoms.....	23
<b>Figure 2.6</b>	The Lennard-Jones potential energy curve for an atomic pair.....	26

	Page
<b>Figure 2.7</b> The potential energy function for a bond. The initial bond length at 2 angstroms is too long. (a) Molecular mechanics finds the lowest energy state of the molecule. (b) Molecular dynamics find the time dependent motion of the molecule. The vibration continues forever.....	28
<b>Figure 2.8</b> An example of the PBC approximation in two dimensions in which the central, square, shaded box of side L is replicated in both dimensions.....	31
<b>Figure 2.9</b> Example schematic cut-off of Lennard-Jones potential.....	32
<b>Figure 2.10</b> Water molecule models generated by 3, 4 and 5 site of point charge.....	34
<b>Figure 3.1</b> Hemagglutinin protein in the simulation box which solvated by water molecules.....	35
<b>Figure 4.1</b> Definition of torsional angles, T1-T5 lying on the bond linker between two connected unit of SA $\alpha$ 2,6Gal pentasaccharide.....	40
<b>Figure 4.2</b> Distribution of torsional angles, T1-T5, of SA $\alpha$ 2,6Gal pentasaccharide (defined in Fig. 4.1) for the six simulated systems: wild-type, E186D, G224S, N182S, Q222V and S129V..	43
<b>Figure 4.3</b> SA $\alpha$ 2,6Gal pentasaccharide bound to the binding pocket of hemagglutinin.....	44
<b>Figure 4.4</b> Important residues at the hemagglutinin binding pocket directly interacted with the sialic acid of SA $\alpha$ 2,6Gal pentasaccharide where the selected atoms of sialic acid were labeled.....	44
<b>Figure 4.5</b> Percent of hydrogen bond occurrence with amino acid residues in the binding pocket of hemagglutinin protein for all systems: (a) wild-type, (b) S129N, (c) N182S, (d) E186D, (e) Q222V and (f) G224S.....	48

- Figure 4.6** Distance labeled of donor-acceptor of heavyatom between amino acid residues in the pocket of hemagglutinin and sialic acid by a simple model: (a)TYR91 OH-O6, (b)TYR91 OH-O7, (c)VAL131 N-O8, (d)VAL131 O-N1,(e)SER132 OG-O1, (f)SER132 OG-O2, (g)SER133 OG-O1, (h)SER133 N-O1, (i)HIE179 NE2-O7, (j)ASN182 N-O7, (k)GLU186 OE1-O7, (l)GLU OE2-O7, (m)GLN222 NE2-O6, (n) SER224 OG-O7 and ASN129 O-N1..... 49
- Figure 4.7** Distances distribution between donor-acceptor of heavy atoms of amino acid residues in the pocket of hemagglutinin and sialic acid for all systems: (a)TYR91 OH-O6, (b)TYR91 OH-O7, (c)VAL131 N-O8, (d)VAL131 O-N1, (e)SER132 OG-O1, (f)SER132 OG-O2, (g)SER133 OG-O1, (h)SER133 N-O1, (i)HIE179 NE2-O7, (j)ASN182 N-O7, (k)GLU186 OE1-O7, (l)GLU186 OE2-O7, (m)GLN222 NE2-O6, (n) SER224 OG-O7 and ASN129 O-N1..... 51

# CHAPTER I

## INTRODUCTION

Avian influenza hemagglutinin (HA) subtype H5 binds preferentially to the SA $\alpha$ 2,3Gal avian cell surface receptor, whilst the SA $\alpha$ 2,6Gal human receptor is well recognized by the human HA subtype H3. Specificity and substrate recognition by different HA subtypes provide a considerable species barrier between avian and human. To understand of how to overcome this transmission barrier, all binding residues of H5 positioned within a sphere of 5 Å of the sialic acid were singly mutated and modeled to be as the same residue as found in the H3 sequence. It is based on an assumption “an increase in the SA $\alpha$ 2,6Gal binding affinity to the mutated H5 may raise a possibility to enable the avian virus to become infective for the human host”.

### 1.1 Avian influenza virus

The Avian influenza virus is an Influenza Type A of the Orthomyxoviridae family. These viruses most commonly infect poultry as well as many types of wild birds. Some of them are also known to infect a variety of mammals, including humans. There are three types of influenza viruses<sup>[1]</sup>. Type A contain any subtypes and has been the major culprit in causing epidemics and pandemics in the last 100 years. Type B has been responsible for some regional level epidemics. Type C seldom creates major problems and is found only in humans. Neither Type B nor Type C has any known subtypes. Differences in the three types of viruses are caused by differences in the envelope proteins, the viral genetic information the virus contains, and the matrix protein<sup>[1,2]</sup>.

Avian Influenza flu infected during the last decade is caused by the bird flu virus also known as the H5N1 virus<sup>[1]</sup>. Scientists believe that all flu viruses are derived from wild birds. The wild birds carry the viruses inside their intestines and rarely get sick from them<sup>[2]</sup>. The avian bird flu virus is very contagious among the

wild birds but is generally not lethal. However, once it infects domestic birds like chickens, ducks and turkeys it is lethal. Outbreaks of bird flu in domestic poultry have been recorded in the following countries: Asia - Cambodia, China, Indonesia, Kazakhstan, Malaysia, Mongolia, Russia, Thailand, and Vietnam, and also in Turkey, Ukraine and Romania<sup>[3]</sup>.

Bird flu virus is spread through the bird's saliva, nasal secretions and feces<sup>[4,5]</sup>. The virus is also transmitted through contaminated surfaces or materials. Once infected, domestic birds either exhibit the "low pathogenic" form which is a milder form of the flu or the "high pathogenic" form which rapidly spreads through flocks causing death<sup>[1]</sup>.

## **1.2 History of influenza virus**

The symptoms of human influenza may be first described by Hippocrates roughly 2,400 years ago<sup>[6]</sup>. Since then, the virus has caused numerous pandemics. Historical data on influenza are difficult to interpret, because the symptoms can be similar to those of other diseases, such as diphtheria, pneumonic plague, typhoid fever, dengue, or typhus. The first convincing record of an influenza pandemic was of an outbreak in 1580, which began in Russia and spread to Europe via Africa<sup>[7]</sup>. In Rome, over 8,000 people were killed, and several Spanish cities were almost wiped out.

There were three influenza pandemics during the 20<sup>th</sup> century: the 1918 Spanish Flu, the 1957 Asian Flu, and the 1968 Hong Kong Flu (informally, pandemics are named after their presumed sites of origin). Influenza virus subtypes H1N1, H2N2, and H3N2 are the antigenic subtypes that, respectively, caused these influenza pandemics. Each successive pandemic virus strain had hemagglutinin (HA, which may be shortened to H) antigens that differed from the HA antigens of previous seasonal flu virus strains and previous pandemic flu virus strains. True human influenza pandemics are classified by changes in the HA subtypes that arise from genetic

reassortment with animal influenza A viruses (genetic shift). Historically, influenza pandemics have differed from each other in etiology, epidemiology, and disease severity.

### 1.2.1 1918 Spanish Flu

In 1918, the Spanish Flu caused an estimated 50 million deaths worldwide (Fig. 1.1). Genetic material recovered from the preserved (frozen) bodies of people who had died from that flu have enabled scientists to identify the causative influenza virus subtype as having been H1N1<sup>[7]</sup>.

This virus was exceedingly virulent, which meant that those who were infected became very sick, many dying the same day as the first symptoms presented. Of those who did not die in the first few days, a high proportion succumbed later to flu-related complications, such as pneumonia. High infection rates and mortalities were especially common among otherwise healthy adults aged 20-50. High-risk groups, such as the elderly and young children, also had high infection rates and mortalities. No subsequent influenza pandemic has been caused by a virus as virulent as the 1918 influenza A virus.



**Figure 1.1** Pandemics of influenza virus subtype H1N1 or “Spanish flu” in 1918<sup>[9]</sup>.

### **1.2.2 1957 Asian Flu**

The Asian Flu pandemic occurred about 40 years after the Spanish Flu pandemic. It caused an estimated one to two million deaths worldwide. By this time, science and technology had advanced significantly. Therefore, scientists were able to rapidly identify the H2N2 flu virus subtype<sup>[7]</sup>. These HA and neuraminidase (NA) antigens were completely different from the antigens of the 1918 influenza virus. Science and technology advancements also enabled scientists to start developing an appropriate vaccine in May 1957, with a limited vaccine supply becoming available by August 1957.

During this pandemic, attack rates (greater than 50%) were highest among school children (aged 5-19), who spread the virus to their classmates. Those infected children carried the virus back to their families. Infection rates were also high among young adults and pregnant women. The elderly had the highest death rates. The worst seemed to be over by December 1957, but then a second wave of illness occurred among the elderly in February 1958. Prolonged illnesses led to reoccurring outbreaks and more deaths.

### **1.2.3 1976 Swine Flu Scare**

When a novel virus was first identified at Fort Dix, it was labeled the "killer flu." Experts were extremely concerned because the virus was thought to be related to the Spanish flu virus of 1918. The concern that a major pandemic could sweep across the world led to a mass vaccination campaign in the United States. In fact, the virus (later named "swine flu") never moved outside the Fort Dix area. Research on the virus later showed that if it had spread, it would probably have been much less deadly than the Spanish flu.



#### **1.2.4 1977 Russian Flu Scare**

In May 1977, influenza A/H1N1 viruses isolated in northern China, spread rapidly, and caused epidemic disease in children and young adults worldwide. The 1977 virus was similar to other A/H1N1 viruses that had circulated prior to 1957<sup>[7]</sup>. Because of the timing of the appearance of these viruses, persons born before 1957 were likely to have been exposed to A/H1N1 viruses and to have developed immunity against A/H1N1 viruses. Therefore, when the A/H1N1 reappeared in 1977, many people over the age of 23 had some protection against the virus and it was primarily younger people who became ill from A/H1N1 infections. By January 1978, the virus had spread around the world. Because illness occurred primarily in children, this event was not considered a true pandemic. Vaccine containing this virus was not produced in time for the 1977-78 seasons, but the virus was included in the 1978-79 vaccine.

#### **1.2.5 1968 Hong Kong Flu**

This 1968 Hong Kong influenza pandemic of 1968 caused fewer deaths than the previous two pandemics. The virus subtype that caused this pandemic was somewhat analogous to the 1957 influenza virus because it had the same NA antigen, N2, but a different HA antigen<sup>[7]</sup> (H3N2).

The first case of Hong Kong flu was detected in September 1968, just 11 years after the last flu pandemic. The outbreak developed rather slowly, becoming widespread in early December, and with the mortality rate peaking in December and January 1969. Similar to the previous pandemic, schoolchildren suffered the highest attack rate. By the time the pandemic waned it had caused about 700,000 deaths worldwide.

### 1.2.6 1997 Avian Flu Scare

In May 1997, the Government Virus Unit in Hong Kong isolated an influenza A virus from a three-year-old child who was admitted to the hospital with a fever and respiratory symptoms. This child later died of acute respiratory distress and Reye's syndrome. Subsequently, the National Influenza Center in the Netherlands identified the virus that caused the child's death as an influenza A H5N1 subtype<sup>[7]</sup>.

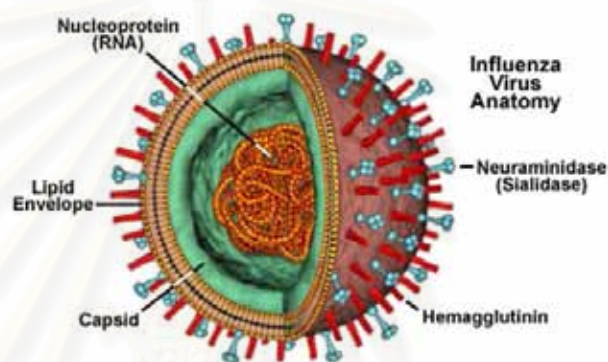
Six months after the virus's identification, 17 human H5N1 infections were documented in hospitalized patients in Hong Kong over a seven-week period. These patients ranged in age from one year to 60 years; seven cases were male and ten were female. Six of the 18 patients died and, with the exception of the first case, symptoms were found to be less severe in children than in adults.

Molecular analysis determined that the influenza A viruses were avian in origin and that genetic reassortment had not occurred. During this time, the same virus was isolated from asymptomatic ducks and geese in local live-bird markets. Further studies revealed that humans became infected as a result of direct human-bird contacts and not person-to-person contact. However, because the influenza A virus strain did not need to pass through an intermediate host such as the pig, but could move directly from bird to human, it represented a novel outbreak.

Similar to the H5N1 virus spread, there have been a few human H9N2 infections, often in association with H5N1 infections. All H9N2 viral isolates are of avian origin. Previously, scientists believed that due to receptor specificity, avian influenza strains could not invade humans. However, since we now know that humans can become infected with H5N1 and H9N2 avian influenza viruses, albeit rarely, this notion has been disproved.

### 1.3 Structure of the influenza virus

The structure of the influenza virus is somewhat variable, but the virion particles are usually spherical or ovoid in shape and 80 to 120 nanometers in diameter<sup>[10]</sup>. Sometimes filamentous forms of the virus occur as well, and are more common among some influenza strains than others. The influenza virion is an enveloped virus that derives its lipid bilayer from the plasma membrane of a host cell.



**Figure 1.2** Structural model of an influenza virus particle<sup>10</sup>.

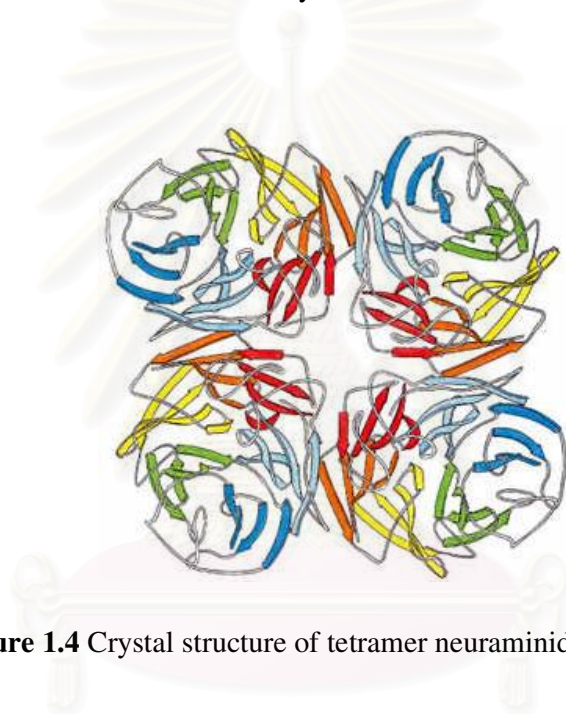
Two different varieties of glycoprotein spike are embedded in the envelope (Fig. 1.2). Approximately 80 percent of the spikes are hemagglutinin, a trimeric protein that functions in the attachment of the virus to a host cell (Fig. 1.3). The remaining 20 percent or so of the glycoprotein spikes consist of neuraminidase, which is thought to be predominantly involved in facilitating the release of newly produced virus particles from the host cell<sup>[1]</sup>. On the inner side of the envelope that surrounds an influenza virion is an antigenic matrix protein lining. Within the envelope is the influenza genome, which is organized into eight pieces of single-stranded RNA (A and B forms only; influenza C has 7 RNA segments)<sup>[1]</sup>. The RNA is packaged with nucleoprotein into a helical ribonucleoprotein form, with three polymerase peptides for each RNA segment<sup>[1]</sup>.

The hemagglutinin molecule is the major antigenic molecule of the virus. This means that the host's immune system usually identifies this molecule on the influenza virus before all others<sup>[11,12]</sup>. Therefore, the hemagglutinin molecule has the greatest selective pressure and thus the greatest diversification among the influenza components. Despite having the most diversity among the influenza components, all of the type A influenza hemagglutinin molecules perform the same function. The hemagglutinin glycoprotein has the important role of mediating the attachment to the target cell, and also the entry of the virus into the target cell<sup>[11-13]</sup>. The hemagglutinin glycoprotein performs these events by binding to a specific molecule on the surface of the target cells. Only the cells that express this molecule will be targeted by the virus for infection. Soon after attaching to the cell, the glycoprotein is induced to undergo a large change in shape, which exposes a part of the glycoprotein that was previously buried within the molecule. The newly exposed part of the hemagglutinin is called the fusion peptide and gets injected into the membrane of the target cell. After several of the virus hemagglutinin molecules insert fusion peptides into the target cell membrane, the virus lipid coat and the cell fuse to form one continuous membrane.



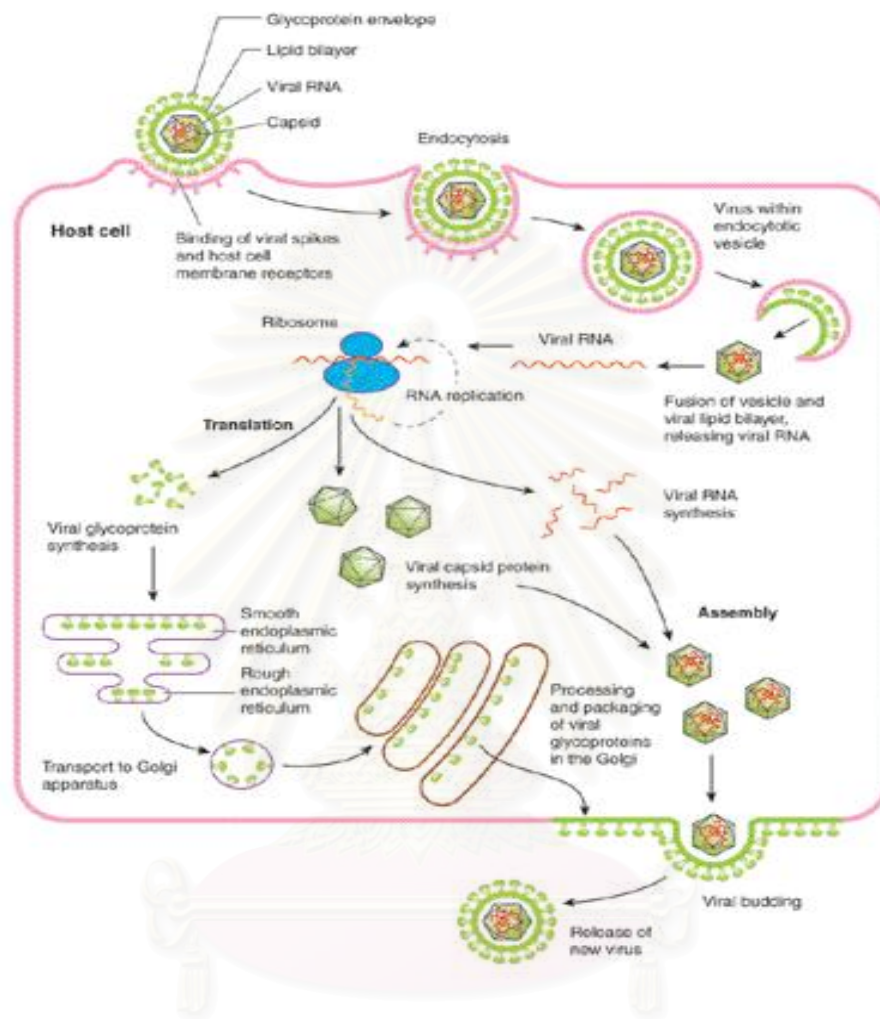
**Figure 1.3** Crystal structure of trimer hemagglutinin protein.<sup>[14]</sup>

The neuraminidase glycoproteins function in what may appear to be the opposite manner (Fig 1.4). Neuraminidase is an enzyme that removes the molecules that the hemagglutinin binds to from the surface of the cell<sup>[15]</sup>. Although this function may seem to be deleterious to the function of the hemagglutinin, the neuraminidase function is important in that it allows new viruses to float away from the cell after they have been assembled. Without the neuraminidase glycoprotein, the viruses would remain attached to the cell in which they were assembled.



**Figure 1.4** Crystal structure of tetramer neuraminidase protein.<sup>[22]</sup>

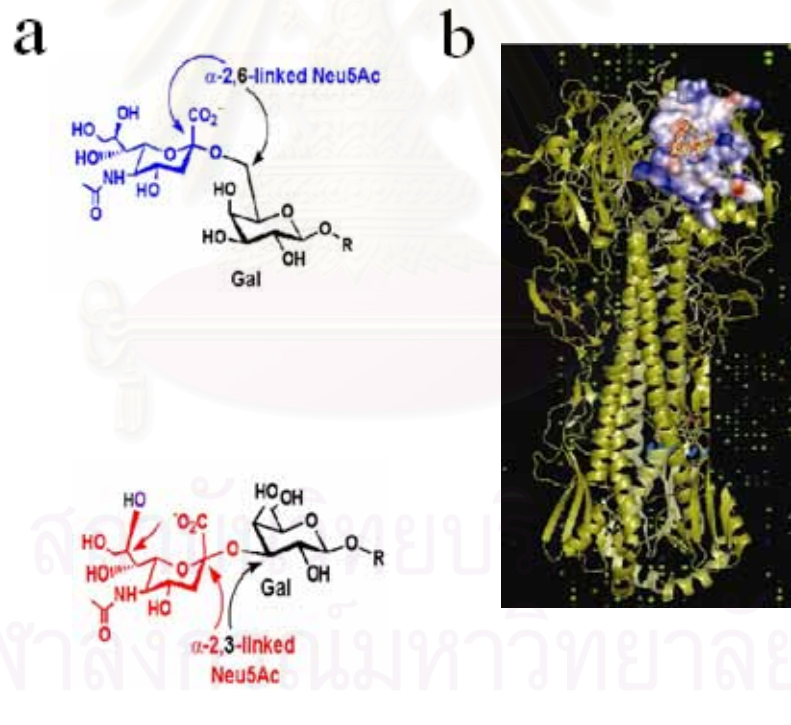
สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



**Figure 1.5** The picture summarizes viral replication process. First, the virus must enter the host cell. Next, the virus will shed its protein coat and begin replication. While the genetic material is replicating, the virus will also produce a new glycoprotein coat. Finally, the new copies of the virus are assembled and released from the cell<sup>[17]</sup>.

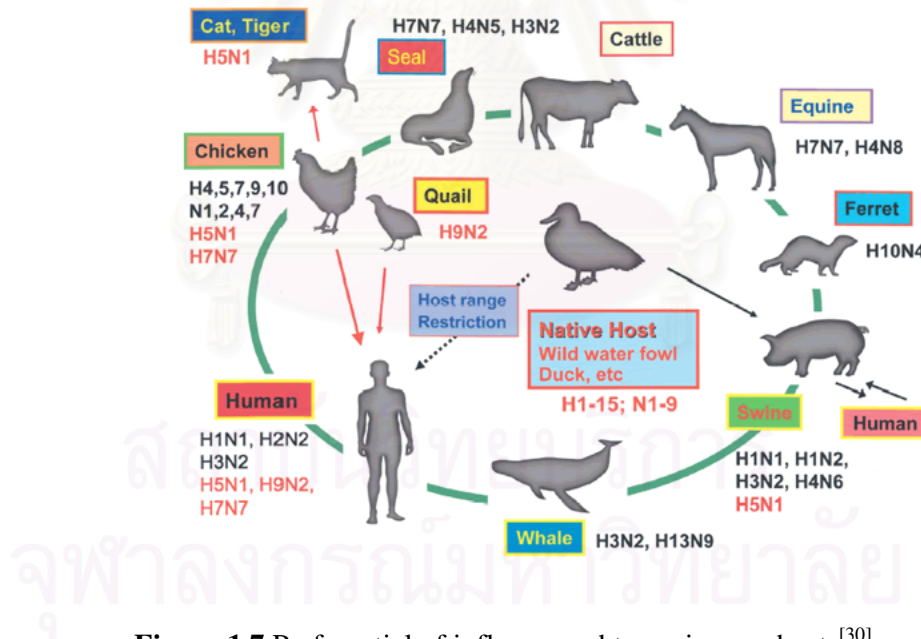
#### 1.4 Specificity and substrate recognition by hemagglutinin

Emergence of new influenza strains in the human population occurs via transmission from other animal species, especially poultry (Fig.1.7). Typically, human and avian influenza viruses are different and are not infectious for both species. However, pigs can become infected with both types of viruses, and act as a “mixing vessel” to produce recombinant viruses capable of transmission to humans<sup>[18]</sup>. Occasionally, direct avian-human transmission can occur, often with enhanced pathogenicity, as demonstrated by the emergence of the recent H5N1 avian influenza in many countries throughout South-East Asia; human-to-human spread of H5N1 avian influenza has not been conclusively documented<sup>[19-22]</sup>.



**Figure 1.6** (a) SA $\alpha$ 2,6Gal and SA $\alpha$ 2,3Gal receptor of human and avian<sup>[23]</sup>. (b) binding pocket of hemagglutinin protein<sup>[24]</sup>.

Interactions between influenza virus hemagglutinin and sialic acid determine the tissue and species tropism of the virus. Human influenza viruses bind to host cell surface targets that contain N-acetylneuraminic acid. Binding to sialic acids occurs via a shallow depression near the tip of the hemagglutinin glycoprotein. Viruses that infect humans bind preferentially to terminal sialic acids containing  $\alpha$ 2,6 linkage<sup>[25]</sup>, whereas others (such as the bird influenza viruses) favor binding to  $\alpha$ 2,3-linked sialic acid<sup>[25-29]</sup>, a receptor binding specificity that correlates with a specific amino acid at the position 226 and 228 of hemagglutinin<sup>[29]</sup> (Fig. 1.6). Pigs can become infected with both avian and human strains because both  $\alpha$ 2,3 and  $\alpha$ 2,6 linked sialic acids occur in the trachea of swine. The cell surface receptors for influenza viruses are widely considered to be sialic acid linked to either glycoprotein or glycolipid.



**Figure 1.7** Preferential of influenza subtypes in vary hosts<sup>[30]</sup>.

Upon completing its cellular replication cycle, the final release of influenza virus from an infected cell surface relies on the action of the viral neuraminidase, which acts to remove sialic acid (the viral receptor) from the surface of the host cells.



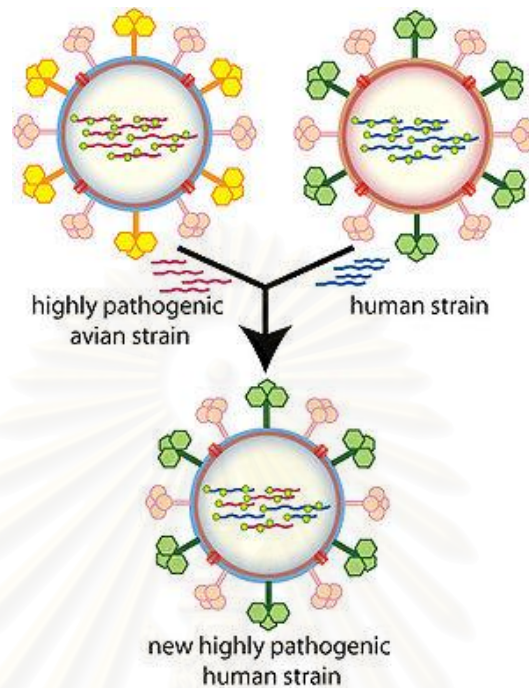
Without this step, the newly forming virus particles would immediately rebind to their receptor and not be efficiently released into the extracellular space, remaining attached to the cell in large clumps. Thus, the establishment of a productive influenza virus infection is dependent on both neuraminidase and hemagglutinin and a delicate balance between the functions of the two glycoproteins.

### **1.5 The viral genome changing**

The influenza virus is constantly changing through mutations, reassortment, and recombination, for example. Different subtypes of Influenza A are found in the environment each winter. Therefore, a new flu vaccine must be produced each year. There are two conditions that are frequently mentioned as the major reasons for the instability of the influenza virus.

First, small changes in the genetic sequence of the HA or NA genes lead to a change in the amino acid sequence of the HA or NA proteins. This happens because the order or sequence of amino acids, which make up a protein. These changes often occur because the influenza virus is an RNA virus. RNA viruses are constantly making “spelling errors” in their genetic sequences when they are being copied. The RNA copying process is flawed. This leads to new genetic sequences. The new genetic sequences lead to new amino acids being put into place in the creation of a protein. This usually leads to a new or altered protein. This series of continual changes is known as genetic drift.

The second mechanism changing is genetic shift. When two strains of virus infect a cell at the same time, the genetic information may not only be copied incorrectly but may also be reassorted or recombined in the new ways. This also could lead to strains of virus that could cause major epidemics or pandemics because the population has no protection against these new strains.



**Figure 1.8** genetic shift mechanism of viral genome changing<sup>[31]</sup>.

### 1.6 Sequence alignment

The amino acid sequences of all H3 which isolated from human host and all H5 which isolated from avian host were collected from National Center for Biotechnology Information (NCBI) database<sup>[32]</sup>. The Needleman-Wunsch algorithm<sup>[33]</sup> was applied to pairwise alignment of all H3 and H5 sequences using Geneious program. The results of sequences alignment were shown in the Table 1.1 which indicated that amino acid residues in all 2442 H3 and 1501 H5 strains provide very similar property.

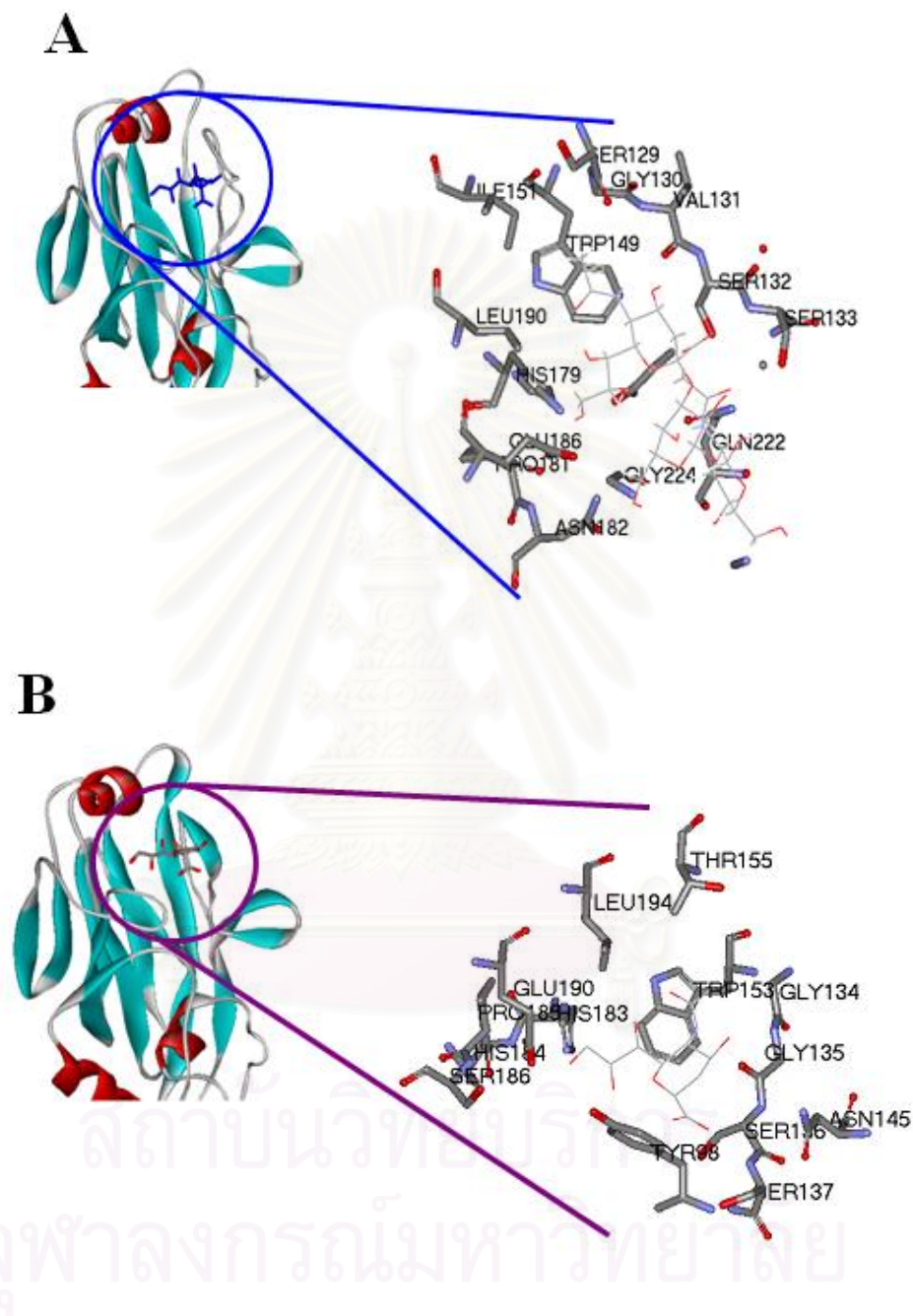
**Table 1.1** Statistical sequence alignment results of hemagglutinin H3 subtype in human host and H5 subtype in avian host.

subtypes	Number of sequences	Percent of	
		identity	similarity
H3-human host	2442	62.2	87.4
H5-avian host	1501	51.6	84.9

### 1.7 Residues within the catalytic pockets of H5 and H3

To seek for the residues within the catalytic pocket of H3-human and H5-avian strains, the following procedures were applied: (i) First set of the residues lying within the sphere of 5 Å around the sialic acid of H5-avian, taken from the crystal structure of H5 (PDB entry code: 1JSN<sup>[14]</sup>) was selected. The resulted 15 residues were found and shown in Fig1.9. (ii) Sequence alignments were performed among 2442 H3-human and 1501 H5-avian strains, only focused to the 15 residue numbers. (iii) Residues with the most probability, among 15 residue numbers, were chosen and summarized in Table 1.2. They were, then, used as the representative within the catalytic pockets of the 2442 H3-human and 1501 H5-human strains.

As a result, among the 15 residues of the two strains, there are 6 different residues (marked as bold in Table 1.2). Here, the 5 H5-avian residues were, then, singly mutated to be that of H3-human subtype and used for further study with the molecular dynamics (MD) simulation framework. In conclusion, the simulated systems of single mutation are S129N, N182S, E186D, Q222V and G224S.



**Figure 1.9** Amino acid residues within a sphere of 5 Å around the hemagglutinin H5 (A) and H3 (B) binding pocket.

**Table 1.2** Representative residues within the catalytic pockets of H5 and H3 (see text for details).

H5 avian		H3 human	
PRO	81	PRO	107
TYR	91	TYR	117
<b>SER</b>	129	<b>ASN</b>	155
GLY	130	GLY	156
SER	132	SER	158
SER	133	SER	159
TRP	149	TRP	175
<b>ILE</b>	151	<b>HIS</b>	177
HIS	179	HIS	205
<b>GLU</b>	186	<b>ASP</b>	212
<b>GLN</b>	222	<b>VAL</b>	248
TRP	149	TRP	175
<b>ASN</b>	182	<b>SER</b>	208
LEU	190	LEU	216
<b>GLY</b>	224	<b>SER</b>	250

### 1.8 Scope of this work

The main goal of this research attempted to evaluate the binding efficiency between hemagglutinin protein (both wild-type and mutated) with SA $\alpha$ 2,6Gal pentasaccharide which indicated in term of binding free energy by means of molecular modeling techniques and the implicit solvent free energy calculations. The

x-ray structure of the H5 wild-type and a set of homology models of H5 mutants complexed with the oligosaccharide SA $\alpha$ 2,6Gal substrate were subjected to molecular dynamics (MD) simulations of fully solvated systems using AMBER force field.

*In silico* site-directed mutations of H5 were performed by substituting the residues within 5 Å of the H5 binding pocket with those aligned residues having different amino acid type of H3. The residues of H5 that are subjected to mutate include S129N, N182S, E186D, Q222V and G224S. Free energy of substrate binding to H5 wild-type and mutants was computed using MD trajectory and molecular mechanic and Poisson-Boltzmann solvent accessible surface area approaches.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

	1	10	20	30	40	50	60
H5_avian	-----	MEKIVLLLAIVSLVKS	DQ----	ICIGYHANNSTE	QVDTIMEK	NTVTVTHA	QDILE
H3_human	MKTIIALS	YILCLVFAQ	KLPGNDN	STATLCLGH	HAVPNGT	IVKTI	TNDQIEVTNATELVQ
H5_avian	KTHNGK	LCDLNGVK	PLILRDCS	VAGWLLGN	PMCEFTN	VPEWSYI	VEKANPANDLCYPGD
H3_human	SSSTGG	ICD-SPHQ	ILDGENCT	LIDALLGDP	QCDGFQN	-KKWDLF	VERSK-AYSNCYPYD
H5_avian	FNDYEEL	KHLLSR	INHFEKI	QIIPKSS	WSNHDAS	SGVSSAC	PHYGRSSFFRN
H3_human	VPDYAS	LRSLIAS	SGTLEFNN	--ESFN	WTG-VTQ	NGTSSAC	KRRSNSSFFSRLNWLTHLK
H5_avian	STYPTIK	RSYNN	TNQEDLL	VLWGIH	HPNDAAE	QTKLYQ	NPTTYISVGTSTLNQRLVPEIA
H3_human	FKYPAL	NVTMPN	NEEFDK	LYIWGV	HHPGTY	NDQISL	YAQASGRITVSTKRSQQTVIPNIG
H5_avian	TRPKVNG	QSGRME	FFWTIL	KPNDAI	NFESNG	NFIAPEY	AYKIVKKGDSAIMKSELEYGNC
H3_human	SRPRIR	DIPSR	ISYWTI	VKPGDI	LLINST	GNLIAP	RGYFKI-RSGKSSIMRSDAPIGKC
H5_avian	NTKCQT	PMGAIN	SSMPFH	NIHPLT	IGECPK	YVKS	NRLVLTATGLRNTPQRERRRKKRGLFG
H3_human	NSECIT	PNGSIP	NDKPFQ	NVNRIT	YGACPK	YVKQNT	LKLATGMRNVPEKQT----RGIFG
H5_avian	AVGREFN	NLERRI	ESLNKK	MEDGFL	DVWWTY	NAELLV	LMENERTLDFHDSNVKNLYDKVRL
H3_human	QIEKEF	SEVEGR	IQDLEK	YVEDTK	IDLWSY	NAELLV	ALENQHTIDLTDSEMNKLFERTKK

**Figure 1.10** Sequence alignment between hemagglutinin H5 avian and H3 human.

## CHAPTER II

### THEORY

#### 2.1 Molecular mechanics

A molecule can possess different kinds of energy such as bond and thermal energy. Molecular mechanics calculates the potential energy of a molecule that means the energy due to the geometry or conformation of a molecule. Energy is minimized in nature, and the conformation of a molecule that is favored is the lowest energy conformation. Knowledge of the conformation of a molecule is important because the structure of a molecule often has a great effect on its reactivity. The effect of structure on reactivity is important for large molecules like proteins. Studies of the conformation of proteins are difficult and therefore interesting, because their size makes many different conformations possible.

Molecular mechanics assumes the potential energy of a molecule to arise from a few, specific interactions within a molecule. These interactions include the stretching or compressing of bonds beyond their equilibrium lengths and angles, torsional effects of twisting about single bonds, the Van der Waals attractions or repulsions of atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to polar bonds.

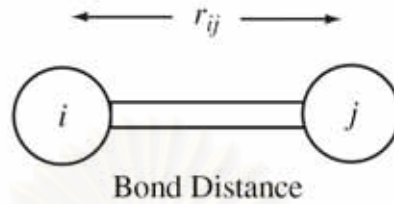
#### **Bonded terms**

To quantify the contribution of each, these interactions can be modeled by a potential function that gives the energy of the interaction as a function of distance, angle, or charge. The total potential energy of a molecule can be written as a sum of the energies of bonded interactions:

$$E_{\text{cov}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} \quad (2.1)$$



### 2.1.1 Bond stretching

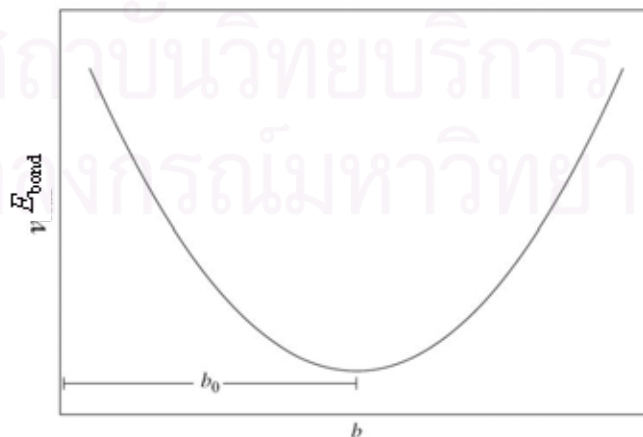


**Figure 2.1** Bond distance defined between two atoms.

$E_{\text{bond}}$  is the energy function for stretching a bond between two atom types A and B. The bond energy is often taken to have a harmonic form:

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 \quad (2.2)$$

Where  $k_b$  is the force constant for the bond,  $b$  is the actual bond length in the structure between the two atoms defining the bond and  $b_0$  is the equilibrium distance for the bond. The sum runs over all the bonds that have been defined in the system. Because the energy is in harmonic form (Fig 2.2) it means that the energy of the bond will increase steadily without limit as it is distorted from its equilibrium value,  $b_0$ .

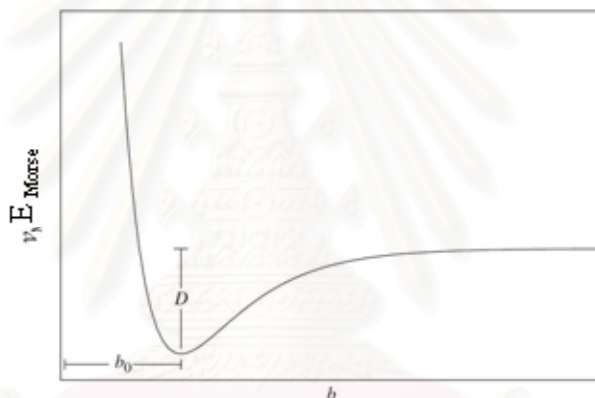


**Figure 2.2** The harmonic bond energy term.

Harmonic terms are sufficient for many studies, but sometimes it is important to have a form for the bond energy that permits dissociation. An example would be if a reaction were being studied. One form of represent the bond behavior is the *Morse* potential which is showed (Fig. 2.3). The Morse energy,  $E_{\text{Morse}}$ , is given by

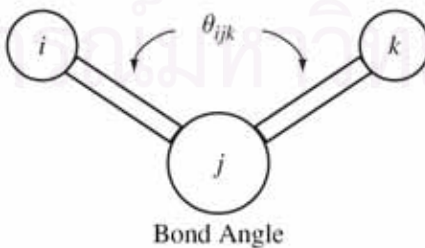
$$E_{\text{Morse}} = \sum_{\text{bonds}} D \{ \exp[-a(b - b_0)] - 1 \}^2 - D \quad (2.3)$$

Where the two new parameters are  $D$ , which is the dissociation energy of the bond, and  $a$ , which determines the width of the potential well.



**Figure 2.3** The Morse function bond energy term.

### 2.1.2 Bond bending



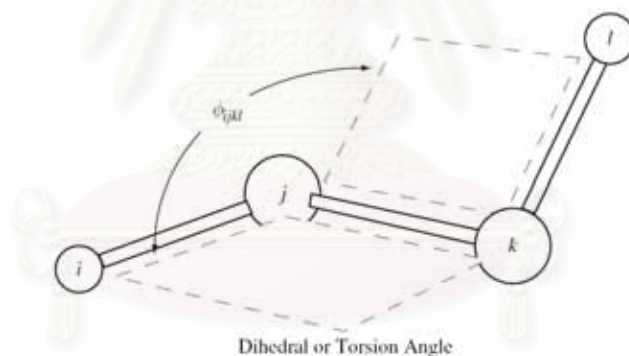
**Figure 2.4** Bond angle defined between three atoms.

The angle energy term is designed to imitate how the energy of a bond angle changes when it is distorted away from its equilibrium position. Like the bond energy term it too is often taken to be harmonic:

$$E_{\text{angle}} = \sum_{\text{angle}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \quad (2.4)$$

The extra parameters are similar to those of the bond energy  $k_{\theta}$  is the force constant for the angle and  $\theta_0$  is its equilibrium value.

### 2.1.3 Torsion angle



**Figure 2.5** Torsion angle or dihedral angle defined between four atoms.

The third type of bonding term is the term that describes how the energy of a molecule changes as it undergoes a rotation about one of its bonds, i.e. the dihedral or torsion energy for the system. In contrast to the bond and angle terms a harmonic form for the dihedral energy is not usually appropriate. This is because, for many dihedral angles in molecules, the whole range of angles from  $0^\circ$  to  $360^\circ$  can be accessible with not too large differences in energy. Such effects can be reproduced with a periodic function that is continuous throughout the complete range of possible angles. The dihedral energy can then be written as

$$E_{\text{dihedral}} = \sum_{\text{dihedral}} \frac{1}{2} V_n (1 + \cos(n\phi - \delta)) \quad (2.5)$$

### 2.1.4 Out of plane

The fourth term in the sum in Equation (2.1) is a more complicated one that describes the energy of out-of-plane motions. It is often necessary for planar groups, such as sp<sup>2</sup> hybridized carbons in carbonyl groups and in aromatic systems, because it is found that use of dihedral terms alone is not sufficient to maintain the planarity of these groups during calculations. A common way to avoid this problem is to define an improper dihedral angle, which differs from the proper dihedral angle in that the atoms which define the dihedral angle, i–j–k–l, are not directly bonded to each other. The calculation of the angle, however, remains exactly the same.

$$E_{\text{improper}} = \sum_{\text{improper}} \frac{1}{2} V_n (1 + \cos(n\omega - \delta)) \quad (2.6)$$

### Non-bonded terms

The bonding energy terms help to define the covalent energy of a molecule. The non-bonding terms describe the interactions between the atoms of different molecules or between atoms that are not directly bonded together in the same molecule. These interactions help to determine the overall conformation of a molecular system.

The non-bonding terms in an empirical force field attempt to reproduce all these types of interaction. Here we shall consider a non-bonding energy consisting of the sum of two terms:

$$E_{\text{nonbond}} = E_{\text{elec}} + E_{\text{vdw}} \quad (2.7)$$

### 2.1.5 Electrostatic energy

The electrostatic energy,  $E_{elec}$ , mimics the energy arising from the electrostatic interactions between two charge distributions. The simplest representation of a charge distribution and the one that is most widely used is one in which a fractional charge is assigned to each atom. This is the total net charge of the atom obtained as the sum of the nuclear charge and the charge in the part of the electron cloud that surrounds it. The electrostatic energy is calculated as

$$E_{elec} = \frac{1}{4\pi\epsilon_0} \sum_{i,j \text{ pair}} \frac{q_i q_j}{r_{ij}} \quad (2.8)$$

Where  $q_i$  and  $q_j$  are the fractional charges on atoms  $i$  and  $j$  and  $r_{ij}$  is the distance between the two particles.

The terms in the pre-factor are  $1/4\pi\epsilon_0$ , which is the standard term when calculating electrostatic interactions in the MKSA (metre, kilogram, second, ampere) system of units, and, which is the dielectric constant that will have the value 1 when the system is in vacuum. The sum in Equation (2.8) runs over all pairs of atoms for which an electrostatic interaction is to be calculated. Note that the fractional charges on the atoms are constants and do not change during a calculation. It is possible to define other representations of the charge distribution. For example, instead of fractional charges at the atoms' centres (i.e. on the nuclei), charges could be assigned off-centre along bonds or higher moments, such as dipoles, could be used. Such representations are not usually favoured because they are more complex and more expensive than the simple point-charge model and they have not been shown to give proportionately better results.

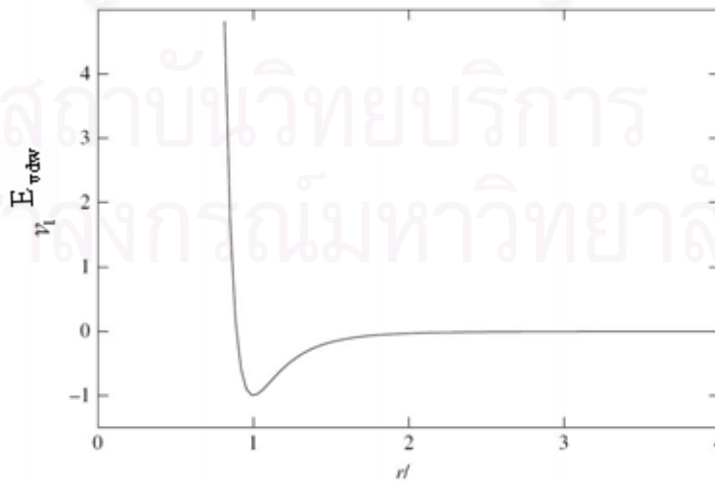
### 2.1.6 Lennard-Jones energy

The second term in Equation (2.7) is the Lennard-Jones energy which mimics the long-range dispersion interactions and the short-range repulsive interactions. It has the form

$$E_{vdw} = \sum_{i,j \text{ pairs}} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (2.9)$$

Where  $A_{ij}$  and  $B_{ij}$  are positive constants whose values depend upon the types of the atoms,  $i$  and  $j$ , and the sum is over all pairs of atoms for which the interaction is to be calculated.

The shape of the Lennard-Jones potential is plotted in Figure 2.6. The repulsive part of the curve is produced by the  $1/r_{ij}^{12}$  term and the attractive part by  $1/r_{ij}^6$ . The inverse sixth power form of the attraction arises naturally from the theory of dispersion interactions. The choice of an inverse twelfth power for the repulsion is less well founded and other forms for the repulsion have been used, including other inverse powers, such as eight and ten, and an exponential form that leads to the so-called Buckingham potential. Most simple force fields seem to use the Lennard-Jones form.



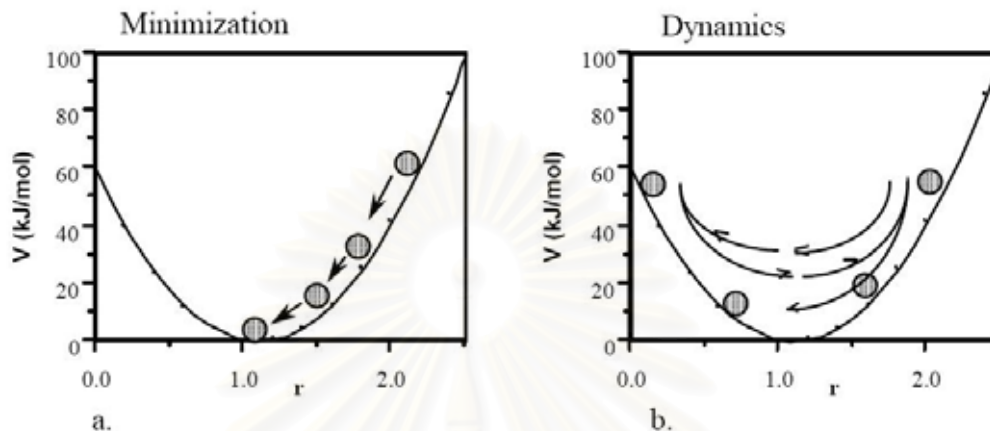
**Figure 2.6** The Lennard-Jones potential energy curve for an atomic pair.

## 2.2 Molecular dynamics

One of the most important developments in macromolecular chemistry is molecular dynamics. Molecular dynamics is the study of the motions of molecules. The time-dependent coordinate of the motion of a molecule is called its trajectory. The trajectory is determined by integrating Newton's equations of motion for the bond stretching, angle bending, and torsions of the molecule. Molecules are always in motion. The motion of molecules is important in essentially all chemical interactions and particular interested in biochemistry. For example, the binding of substrates to enzymes, the binding of antigens to antibodies, the binding of regulatory proteins to DNA, and the mechanisms of enzyme catalysis are enhanced. Different domains of an enzyme can have very different motional freedom.

Most chemistry is done in solution. Molecular dynamics has proved to be an invaluable tool in studies of solvation energy. Solute and solvent interactions are governed by the relative motions of the solute and solvent molecules and the motional-response of the solute to the presence of the solvent.

The difference between molecular mechanics and dynamics can be illustrated with a simple example. If we were to run molecular mechanics, the bond length would decrease until the minimum in the potential energy was reached, Fig. 2.7 (a). Further minimization would not change the bond length. If we were to run molecular dynamics on our stretched bond, the trajectory would decrease the bond length, but the bond length would continue decreasing past the equilibrium length until it was too short. Being too short, the bond length would then begin to increase. Over time the bond length will oscillate about its equilibrium value, never coming to rest, Fig. 2.7 (b). In other words, in mechanics the potential energy is minimized, while the kinetic energy of the molecule is ignored. In a dynamics trajectory, both potential and kinetic energy are studied and the total energy is conserved by the motion.



**Figure 2.7** The potential energy function for a bond. The initial bond length at 2 angstroms is too long. (a) Molecular mechanics finds the lowest energy state of the molecule. (b) Molecular dynamics find the time dependent motion of the molecule. The vibration continues forever.

### 2.2.1 Algorithms for Time Dependence

Once we have calculated the potential and hence the force by differentiation, we have to solve Newton's equation of motion. If  $F_A$  is the force on particle A, whose position vector is  $r_A$  and whose mass is  $m_A$  then

$$\begin{aligned}
 F_A &= m_A \frac{d^2 r_A}{dt^2} \\
 &= m_A a_A
 \end{aligned}
 \tag{2.10}$$

This is a second-order differential equation that I can write equivalently as two first-order differential equations for the particle position  $r_A$  and the velocity  $V_A$



$$v_A = \frac{dr_A}{dt} \quad (2.11)$$

### Leapfrog Algorithm

A simple algorithm for integration of these two equations numerically in small time steps  $\Delta t$  can be found by considering the Taylor expansion for  $\mathbf{v}(t)$ :

$$\begin{aligned} v_A\left(t + \frac{\Delta t}{2}\right) &= v_A(t) + \left(\frac{dv_A}{dt}\right)\left(\frac{\Delta t}{2}\right) + \frac{1}{2}\left(\frac{d^2v_A}{dt^2}\right)\left(\frac{\Delta t}{2}\right)^2 + \dots \\ v_A\left(t - \frac{\Delta t}{2}\right) &= v_A(t) - \left(\frac{dv_A}{dt}\right)\left(\frac{\Delta t}{2}\right) + \frac{1}{2}\left(\frac{d^2v_A}{dt^2}\right)\left(\frac{\Delta t}{2}\right)^2 - \dots \end{aligned} \quad (2.12)$$

Subtracting and rearranging we get:

$$v_A\left(t + \frac{\Delta t}{2}\right) = v_A\left(t - \frac{\Delta t}{2}\right) + a_A(t)\Delta t + \dots \quad (2.13)$$

Then, we could have written  $v_A(t)$  to mean the instantaneous velocity of particle A at time  $t$ . The acceleration  $a_A$  is calculated from the force. Using the same procedure for the Taylor expansion of  $r_A$  at the time point  $t + \Delta t/2$  we get

$$r_A(t + \Delta t) = r_A(t) + v_A\left(t + \frac{\Delta t}{2}\right)\Delta t + \dots \quad (2.14)$$

Equations (2.13) and (2.14) form the so-called leapfrog algorithm, which is reputed to be one of the most accurate and stable techniques for use in molecular dynamics. A suitable time increment  $\Delta t$  for molecular dynamics is a femtosecond ( $10^{-15}$  s). In the leapfrog scheme, the velocities are first calculated at time  $t + \Delta t/2$ . These are used to calculate the positions of the particles at time  $t + \Delta t$  and so on. In

this way the velocities leap over the positions and then the positions leap over the velocities.

### Verlet Algorithm

If instead we start from the Taylor expansion of  $r_{A(t)}$  we have

$$\begin{aligned} r_A(t + \Delta t) &= r_A(t) + \left(\frac{dr_A}{dt}\right)(\Delta t) + \frac{1}{2}\left(\frac{d^2r_A}{dt^2}\right)(\Delta t)^2 + \dots \\ r_A(t - \Delta t) &= r_A(t) - \left(\frac{dr_A}{dt}\right)(\Delta t) + \frac{1}{2}\left(\frac{d^2r_A}{dt^2}\right)(\Delta t)^2 - \dots \end{aligned} \quad (2.15)$$

which gives (assuming that third-order and higher terms are negligible)

$$r_A(t + \Delta t) = 2r_A(t) - r_A(t - \Delta t) + \left(\frac{d^2r_A}{dt^2}\right)(\Delta t)^2 \quad (2.16)$$

This is known as the Verlet algorithm. The acceleration is obtained from the force experienced by atom A at time  $t$ . The velocity does not appear in the expression, but it can be obtained from the finite difference formula

$$v_A(t) = \frac{r_A(t + \Delta t) - r_A(t - \Delta t)}{2\Delta t} \quad (2.17)$$

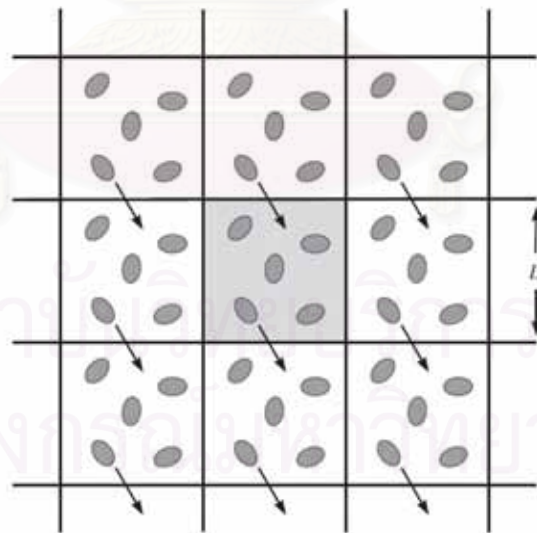
The Verlet algorithm uses positions and accelerations at time  $t$  and the position at time  $t - \Delta t$  to calculate a new position at time  $t + \Delta t$ . All these have to be stored at every iteration. A variant is the velocity Verlet algorithm, which requires only the storage of positions, velocities and accelerations that all correspond to the same time step. It takes the form:

$$r_A(t+\Delta t) = r_A(t) + \left(\frac{dr_A}{dt}\right)(\Delta t) + \frac{1}{2}\left(\frac{d^2r_A}{dt^2}\right)(\Delta t)^2$$

$$v_A(t+\Delta t) = \left(\frac{dr_A}{dt}\right) + \frac{1}{2}\left[\left(\frac{d^2r_A}{dt^2}\right)_t + \left(\frac{d^2r_A}{dt^2}\right)_{t+\Delta t}\right](\Delta t)$$
(2.18)

### 2.2.2 Periodic boundary condition

The classical way to minimize edge effects in a finite system is to apply periodic boundary conditions. The atoms of the system to be simulated are put into a space-filling box, which is surrounded by translated copies of itself (Fig. 2.8). Thus there are no boundaries of the system. The assumption of periodicity immediately makes the simulation of such a system tractable because equivalent atoms in each of the copies behave identically and so do not need to be treated distinctly during a simulation.



**Figure 2.8** An example of the PBC approximation in two dimensions in which the central, square, shaded box of side  $L$  is replicated in both dimensions.

In the PBC approximation an infinite system is constructed by periodically replicating a finite system. The finite system must be of a sufficiently regular shape that it can fill space when it is copied. The most common option in three dimensions is to use finite systems that are cubic. Also common are orthorhombic boxes whose angles are all right angles but whose sides are of different lengths. Other shapes that are possible are triclinic, hexagonal, truncated octahedral and various sorts of dodecahedral boxes. These shapes are geometrically more complicated but they can be required when studying certain types of system.

### 2.2.3 Cutoff methods for the calculation of non-bonding interactions

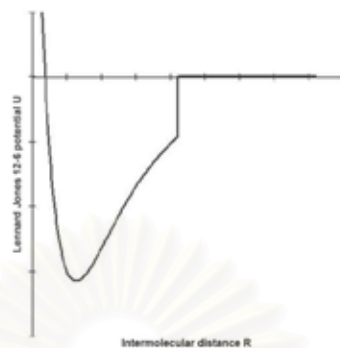
The non-bonding energy for a molecular system with the types of MM force fields that we are using can be written as a sum of electrostatic and Lennard-Jones contributions. The expression for the energy,  $E_{nonbond}$ , is

$$E_{nonbond} = \sum_{i,j \text{ pairs}} \left( \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (2.19)$$

The principal problem with the non-bonding energy is the long-range electrostatic interaction which decays as the reciprocal of the distance between the atoms. The most widely used approximate methods for the evaluation of the non-bonding energy overcome the long-range nature of the electrostatic interaction by modifying its form so that the interactions between atoms are zero after some finite distance. These are the cutoff or truncation methods.

There are several subtleties that have to be addressed when using cutoff schemes. The first is that of how the truncation is to be effected. The easiest way is to use an abrupt truncation and simply ignore all interactions that are beyond the cutoff distance. This is equivalent to multiplying each term in the non-bonding energy expression (Equation (2.18)) by a truncation function,  $S(r)$ , of the form

$$S(r) = \begin{cases} 1 & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (2.20)$$



**Figure 2.9** Example schematic cut-off of Lennard-Jones potential.

### 2.2.4 Water models

In computational chemistry, classical water models are used for the simulation of water clusters, liquid water, and aqueous solutions with explicit solvent. These models use the approximations of molecular mechanics. Many different models have been proposed. They can be classified by the number of points used to define the model (atoms plus dummy sites), whether the structure is rigid or flexible, and whether the model includes polarization effects. An alternative to the explicit water models is to use an implicit solvation model, also known as a continuum model.

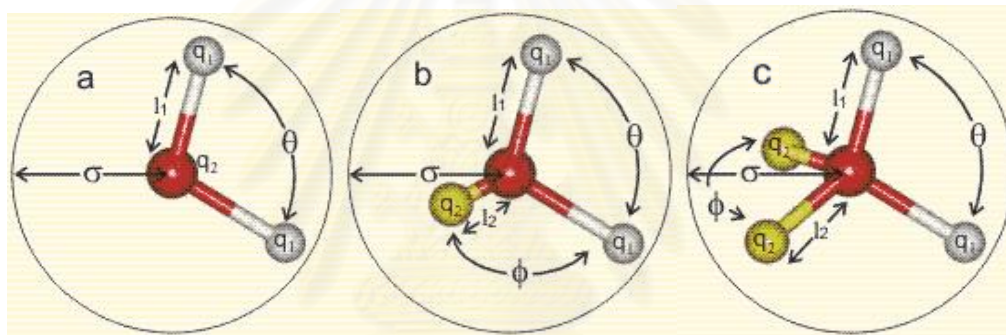
The simplest water models treat the water molecule as rigid and rely only on non-bonded interactions. The electrostatic interaction is modeled using Coulomb's law and the dispersion and repulsion forces using the Lennard-Jones potential. The potential for models such as TIP3P and TIP4P is represented by

$$E_{ab} = \sum_i^{\text{on } a} \sum_j^{\text{on } b} \frac{k_c q_i q_j}{r_{ij}} + \frac{A}{r_{oo}^{12}} - \frac{B}{r_{oo}^6} \quad (2.21)$$

where  $k_c$ , the electrostatic constant, has a value of  $332.1 \text{ \AA} \cdot \text{kcal/mol}$  in the units commonly used in molecular modeling;  $q_i$  are the partial charges relative to the charge

of the electron;  $r_{ij}$  is the distance between two atoms or charged sites; and A and B are the Lennard-Jones parameters.

The charged sites may be on the atoms or on dummy sites (such as lone pairs). In most water models, the Lennard-Jones term applies only to the interaction between the oxygen atoms. The figure below shows the general shape of the 3, 4 and 5 site water models. The exact geometric parameters (the OH distance and the HOH angle) vary depending on the model.



**Figure 2.10** Water molecule models generated by 3, 4 and 5 site of point charge

The simplest models have three interaction sites, corresponding to the three atoms of the water molecule (Fig 2.10(a)). Each atom gets assigned a point charge, and the oxygen atom also gets the Lennard-Jones parameters. The 3-site models are very popular for molecular dynamics simulations because of their simplicity and computational efficiency. Most models use a rigid geometry matching the known geometry of the water molecule.

### 2.2.5 Implicit solvation

Implicit solvation or continuum solvation is a method of representing solvent as a continuous medium instead of individual “explicit” solvent molecules most often used in molecular dynamics simulations and in other applications of molecular mechanics. The method is often applied to estimate free energy of solute-solvent interactions in structural and chemical processes, such as folding or conformational transitions of proteins, DNA, RNA, and polysaccharides, association of biological macromolecules with substrates, or transport of drugs across biological membranes.

The implicit solvation model is justified in liquids, where the potential of mean force can be applied to approximate the averaged behavior of many highly dynamic solvent molecules. However, the interiors of biological membranes or proteins can also be considered as media with specific solvation or dielectric properties. These media are continuous but not necessarily uniform, since their properties can be described by different analytical functions, such as “polarity profiles” of lipid bilayers. There are two basic types of implicit solvent methods: models based on accessible surface areas (ASA) that were historically the first, and more recent continuum electrostatics models, although various modifications and combinations of the different methods are possible.

The accessible surface area (ASA) method is based on experimental linear relations between Gibbs free energy of transfer and the surface area of a solute molecule.<sup>[2]</sup> This method operates directly with free energy of solvation, unlike molecular mechanics or electrostatic methods that include only the enthalpic component of free energy. The continuum representation of solvent also significantly improves the computational speed and reduces errors in statistical averaging that arise from incomplete sampling of solvent conformations, so that the energy landscapes obtained with implicit and explicit solvent are different. Although the implicit solvent model is useful for simulations of biomolecules, this is an approximate method with certain limitations and problems related to parameterization and treatment of ionization effects.

### 2.2.6 Molecular mechanics Poisson-Boltzmann solvent accessibility surface area

There are many computational methods that can be used to calculate the standard and relative free energies of complex formation, but most of them require substantial computational efforts. The MM/PBSA scheme is a post processing method to evaluate the standard free energies of molecules or the binding free energies of molecular complexes in a less computationally demanding manner. The MM/PBSA calculation method combines molecular mechanical (MM) energies and continuum solvation energies as follows:

$$\langle G_{mol} \rangle = \langle E_{MM} \rangle + \langle G_{PBSA} \rangle \quad (2.22)$$

where  $\langle G_{mol} \rangle$  is the average “standard” free energy of the molecule of interest, which can be the substrate, the receptor, or their complex. The average molecular mechanical energy,  $\langle E_{MM} \rangle$ , is typically defined as :

$$\langle E_{MM} \rangle = \langle E_{bond} \rangle + \langle E_{angle} \rangle + \langle E_{torsion} \rangle + \langle E_{vdw} \rangle + \langle E_{elec} \rangle \quad (2.23)$$

where  $E_{bond}$ ,  $E_{angle}$ ,  $E_{torsion}$ ,  $E_{vdw}$ ,  $E_{elec}$  are the bond, angle, torsion, van der Waals, and electrostatics terms of intramolecular energy, respectively. The molecular solvation free energy can be further decomposed to:

$$\langle G_{PBSA} \rangle = \langle G_{PB} \rangle + \gamma \langle A \rangle \quad (2.24)$$

where  $\langle G_{PBSA} \rangle$  denotes the average electrostatic contribution of molecular solvation,  $\gamma$  is the surface tension of water, and  $A$  is the solvent-accessible surface area (SASA). The electrostatic solvation free energy,  $G_{PB}$ , is calculated by:

$$G_{PB} = \sum_{i=1}^N [q_i(\varphi_i^{aq} - \varphi_i^s)] \quad (2.25)$$



where  $N$  is the number of atoms in the molecule,  $q_i$  is the electrostatic charge of atom  $i$ , and  $\phi_i^{aq}$  and  $\phi_i^g$  are the electrostatic potentials of atom  $i$  in the aqueous and gas phase, respectively, which are usually obtained by solving the Poisson–Boltzmann equation:

$$-\nabla \cdot [\varepsilon \varepsilon(r) \nabla \varphi(r)] + \kappa^2(r) \sinh \left[ \frac{e_c \varphi(r)}{k_B T} \right] = \rho_{mol}(r) \quad (2.26)$$

where  $\varepsilon(r)$  is the position-dependent dielectric constant,  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature,  $e_c$  is the electron charge,  $\kappa$  is the inverse Debye–Hückel screening length,

$$\kappa^2(r) = \frac{2Ie_c^2}{k_B T \varepsilon(r)} \quad (2.27)$$

where  $I$  is the ionic strength, and  $\rho_{mol}(r)$  is the charge distribution of the molecule, given by a sum of  $\delta$  functions:

$$\rho_{mol}(r) = \sum_{i=1}^N q_i \delta(r - r_i) \quad (2.28)$$

In principle, the binding free energy of a substrate with its receptor can be calculated as follow:

$$\Delta G_{bind} = \Delta G_{com} - \Delta G_{pro} - \Delta G_{sub} \quad (2.29)$$

where  $\Delta G_{com}$ ,  $\Delta G_{pro}$ , and  $\Delta G_{sub}$  are the standard free energy of the protein, substrate, and their complex, respectively.

## CHAPTER III

### COMPUTATIONAL DETAILS

#### 3.1 Preparation of the initial structure

The crystal structure of wild-type hemagglutinin H5 protein complexed with SA $\alpha$ 2,3Gal and SA $\alpha$ 2,6Gal pentasaccharide were taken from Protein Data Bank (PDB) code 1JSN and 1RVT<sup>[14]</sup>, respectively. Complexation of wild-type hemagglutinin H5 protein and SA $\alpha$ 2,6Gal pentasaccharide was taken by superimposition SA $\alpha$ 2,6Gal pentasaccharide into SA $\alpha$ 2,3Gal at the sialic acid position. After that, moved out the SA $\alpha$ 2,3Gal and used as an initial structure of wild-type complexed. All mutant systems were modeled by single replacement of amino acid residue (as describe above) using LEAP module in AMBER8 software package<sup>[36]</sup> which used an initial structure of wild-type complex as template.

#### 3.2 Molecular dynamics simulations

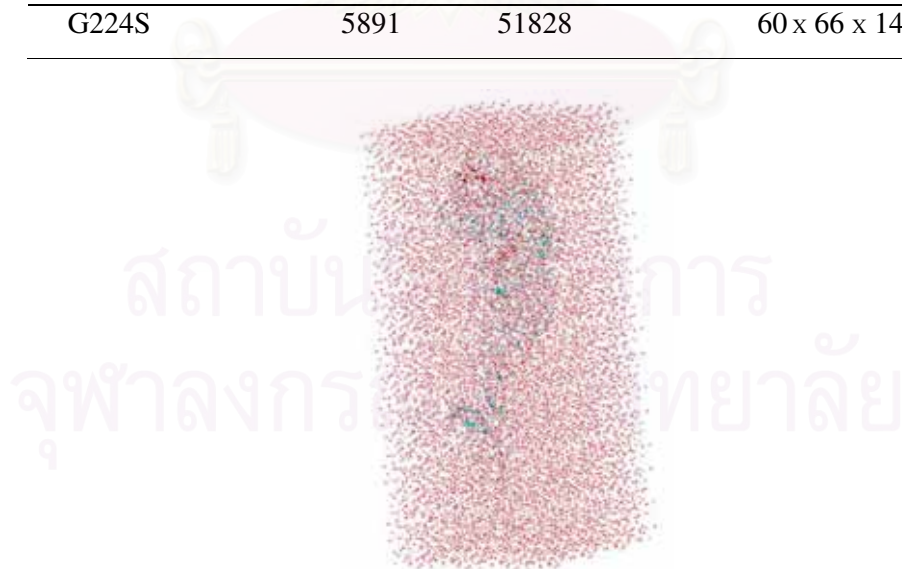
The molecular topology of the systems was generated using LEAP module in AMBER8 program. All hydrogen atoms of the protein were added using the LEAP module in the AMBER8 software package. The Cornell forcefield version ff03<sup>[37]</sup> was used for protein atoms. The force field of the SA $\alpha$ 2,6Gal substrate atoms was taken from RESP atomic charges of the substrate obtained from ANTECHAMBER module. All simulated systems were filled up by solvated TIP3P water molecules. The dimensions of the obtained simulation boxes are approximately 60 x 66 x 143 Å<sup>3</sup>.

The periodic boundary condition with the NPT ensemble was used. Before starting the MD simulations, two steps of energy minimization were performed: (i) only water molecules were minimized 3,000 steps by steepest descent method. (ii) The whole systems energy minimized for 3,000 steps by steepest descent method. Energy minimization and MD simulations were carried out using the SANDER

module in AMBER8. A Berendsen coupling time of 0.2 ps was used to maintain the temperature and standard pressure of the system. The SHAKE algorithm<sup>[35]</sup> was applied to constrain all bonds involving hydrogens. The simulation 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<sup>[36]</sup> was used for an adequate treatment of long-range electrostatic interactions.

**Table 3.1** Detailed characteristics of the simulated systems.

system	Number of		Box dimension (Å <sup>3</sup> )
	water molecules	total atoms	
wild-type	5887	51830	60 x 66 x 143
S129N	5891	51826	60 x 66 x 143
N182S	5886	51833	60 x 66 x 143
E186D	5887	51837	60 x 66 x 143
Q222V	5890	51842	60 x 66 x 143
G224S	5891	51828	60 x 66 x 143



**Figure 3.1** Hemagglutinin protein in the simulation box which solvated by water molecules.

In the pre-equilibration step, multiple stepwise MD simulations were carried out by employing the harmonic position restrained potential to the substrate atoms with the restrain factors of 50, 25, 10, 5, 2.5, 1.2, 0.5 and 0.2 kcal mol<sup>-1</sup> Å<sup>2</sup> for 200, 100, 100, 100, 100, 100, 100 and 100 ps of the simulation, respectively, to gradually relax the modeled systems while maintaining the position of the hemagglutinin protein in the binding pocket.

During an equilibration of the simulation, structure coordinates of the wild-type H5 SA $\alpha$ 2,6Gal complex taken from the MD snapshot were used as a template to construct a set of homology models of H5 mutant complexed with SA $\alpha$ 2,6Gal. The residues of H5 that are subjected to *in silico* site-directed mutagenesis include Ser129, Asn182, Glu186, Gln222 and Gly224 by asparagine, serine, aspartic acid, valine and serine respectively. Thus, a wild-type and total of 5 mutant complexes was subjected to MD with analogous system set up for the simulation. Then, all atoms of the simulated system were subjected to move without any restraints. The production stage of MD run for all the wild-type and mutant systems were performed until achieving 5 ns of the simulation. Structure and dynamics properties of each system were analyzed using the last 3 ns of MD trajectory.

### 3.3 The binding free energy calculations

The binding free energy ( $\Delta G_{bind}$ ) of hemagglutinin wild-type and mutants with SA $\alpha$ 2,6Gal pentasaccharide using MM/PBSA module in AMBER8 program. The enthalpic change in the gas phase upon binding ( $\Delta E_{MM}$ ) was computed by MM module with no cutoff. The electrostatic free energy of solvation was computed using the Poisson-Boltzmann module. The grid spacing was set to 1/3 Å with the boundary conditions of Debye-Hückel potentials. Atomic charges were taken from the Cornell forcefield. The water and protein dielectric constants were set to 80.0 and 1.0, respectively. The solvent-accessible surface area was calculated by SA module.

## CHAPTER IV

### RESULTS AND DISCUSSION

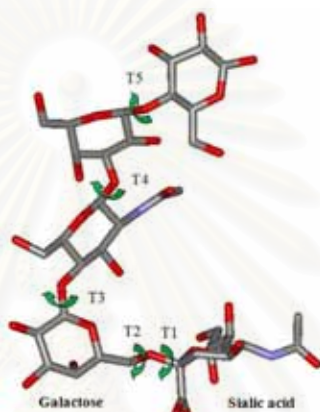
The wild-type and five single mutated (S129N, N182S, E186D, Q222V and G224S) hemagglutinin strains complexed with the SA $\alpha$ 2,6Gal pentasaccharide were treated by molecular dynamics simulations. Only trajectories taken from the production phase of 3-ns simulations were extracted for the extensive analysis comprising of torsional angles of substrate, enzyme-substrate interactions, intermolecular distances, and predictive binding affinity of substrate against hemagglutinins.

#### 4.1 Torsion angle of SA $\alpha$ 2,6Gal pentasaccharide

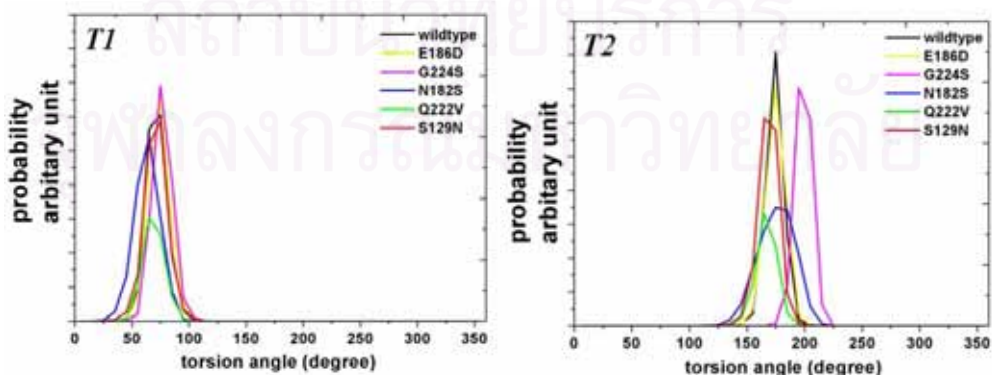
To access the stability of the SA $\alpha$ 2,6Gal pentasaccharide in the HA binding pocket, the torsion angles between the saccharide monomers were monitored. Each angle was defined by a set of the four connected atoms lying on the bond linker between two neighboring monomers as shown in Fig. 4.1. The results were plotted in Fig. 4.2.

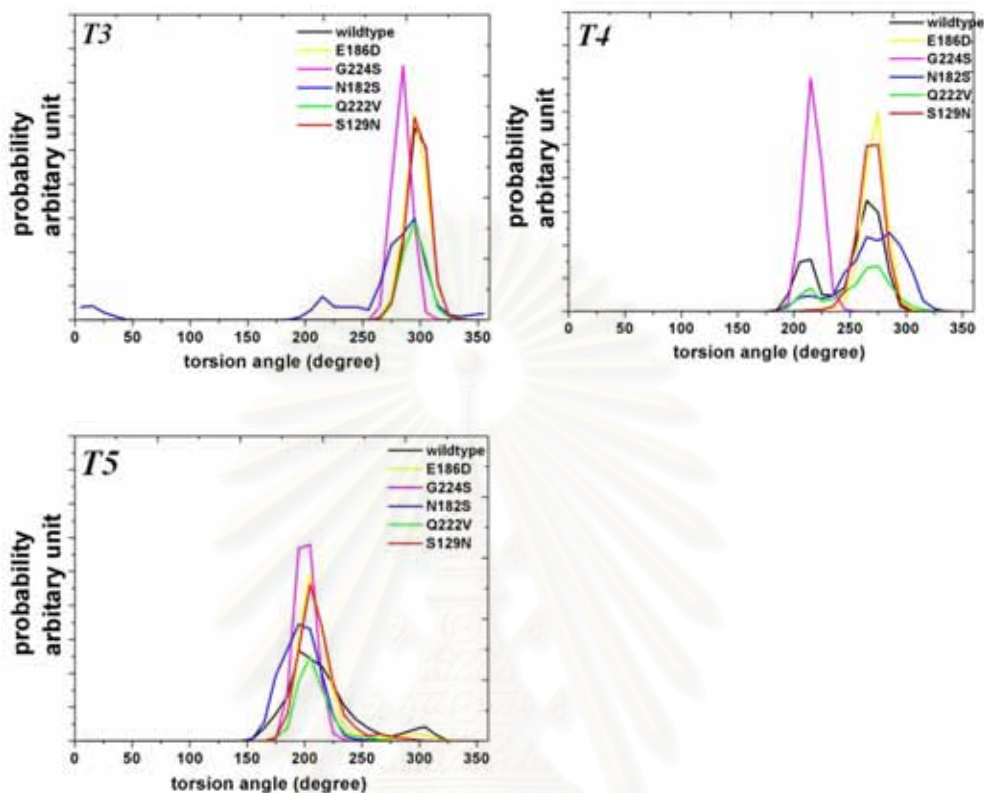
In the wild-type strain, torsion angles linked between sialic acid and galactose, T1 and T2 (see Fig.4.1 for definition), showed a sharp and narrow peak with the maximum at  $\sim 75^\circ$  and  $\sim 170^\circ$ , respectively (Fig.4.2). Similar shape and preferential position of both angles were observed in all mutations. This is except for the T2 in the G224S system which was considerably shifted by  $30^\circ$  to  $210^\circ$  with a result of new conformation of the SA $\alpha$ 2,6Gal structure lying in the HA pocket. Consequently, change of intermolecular interactions to this new conformation was noticeably found in this mutation (discussed in the section of enzyme-substrate interactions).

Not much information can be learnt from the torsion angles, T3-T5, of the last three saccharides of the substrate: N-acetylglucosamine, glucose and terminal glucose, lying outside the binding pocket of enzyme. The ordering of torsion flexibility of  $T5 > T4 \gg T3 > T2 > T1$  is found in all strains suggesting that the two glucoses move rather freely in the solution.



**Figure 4.1** Definition of torsional angles, T1-T5 lying on the bond linker between two connected unit of SA $\alpha$ 2,6Gal pentasaccharide.



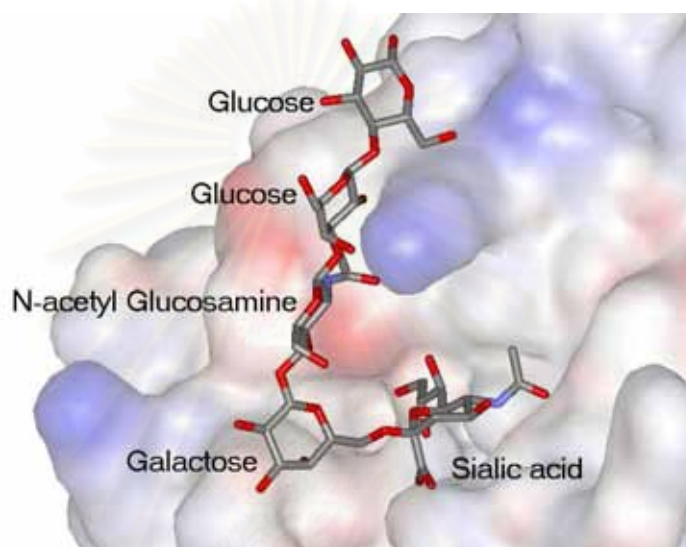


**Figure 4.2** Distribution of torsional angles,  $T1$ - $T5$ , of SA $\alpha$ 2,6Gal pentasaccharide (defined in Fig. 4.1) for the six simulated systems: wild-type, E186D, G224S, N182S, Q222V and S129V.

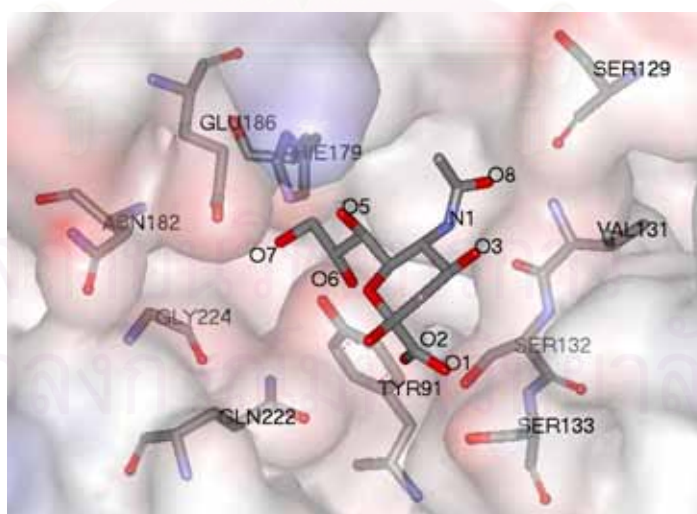
#### 4.2 Enzyme-substrate interactions

The complexed structure of hemagglutinin and SA $\alpha$ 2,6Gal pentasaccharide was shown in Fig. 4.3. Only sialic acid was fitted into the enzyme binding pocket (Fig. 4.3) and thus had direct interactions with the catalytic residues (Fig. 4.4): Tyr91, Ser129, Val131, Ser132, Ser133, Gln222, Gly224, His179, Asn182, and Glu186. To determine the hydrogen bond interactions between enzyme and substrate, the following criteria were used: (i) distance between hydrogen bond donor and acceptor was less than 3.5 Å; and (ii) angle of donor-hydrogen-acceptor was higher than 120°.

The results were depicted in Fig. 4.5. To better understanding, comparison of the detected hydrogen bond interactions between the wild-type and mutated systems were discussed together with the corresponding distances where the definitions and results were shown in Fig. 4.6 and 4.7, respectively.



**Figure 4.3** SA $\alpha$ 2,6Gal pentasaccharide bound to the binding pocket of hemagglutinin.



**Figure 4.4** Important residues at the hemagglutinin binding pocket directly interacted with the sialic acid of SA $\alpha$ 2,6Gal pentasaccharide where the selected atoms of sialic acid were labeled.



Considering the  $\text{-COO}^-$  group of the sialic acid in the wild-type system (Fig. 4.4), the four medium hydrogen bonds (% occupation  $\sim 50$ ) with two serines, Ser132 and Ser133, were detected (Fig. 4.5 a). The O1-oxygen (see Fig. 4.4 for atomic label) was well stabilized by the two serine hydroxyl groups and the backbone nitrogen of Ser133 as follow. The O1-OG and O1-N (Ser133) distances showed a sharp peak with a long tail preferentially positioned at 2.7 Å (Figs. 4.7 g and h). The O1-OG (Ser132) distance was found to exhibit a broad peak ranging from 2.3 Å to 6.0 Å (Fig. 4.7 e). In addition, this hydroxyl group of Ser132 was also interacted with the other oxygen of the  $\text{-COO}^-$  moiety of substrate as shown by a narrow peak with a long tail appeared at 2.7 Å (Fig. 4.7 f). In all single mutations, increase in hydrogen bonds with Ser133 (% occupation  $> 90$  in Figs. 4.5 b-f) was noticeably found indicating the firmly strong electrostatic contribution to the negatively charged  $\text{-COO}^-$  group. These strong hydrogen bonds were confirmed by the detected sharp and narrow peaks at 2.7 Å in Fig. 4.7 g and h, except for two separated peaks of the O1-OG (Ser133) distances for N182S and Q222V (Fig. 4.7 g). The medium hydrogen bonds with Ser132 were maintained in the S129N, N182S and Q222V mutations (% occupation  $\sim 30$ -70 in Figs. 4.5 b, c and e, respectively) whereas those obtained in the other mutations are considerably increased (% occupation  $> 80$  in Figs. 4.5 d and f). This obtained results are in good agreement with the hydrogen bond lengths measured from the hydroxyl group of Ser132 to the two oxygens of the  $\text{-COO}^-$  group of the sialic acid (Figs. 4.7 e and f). Surprisingly, the higher electrostatic contribution from these two serines to the  $\text{-COO}^-$  group via hydrogen bonds maybe not a main reason for an increase in the substrate binding affinity in this particular enzyme.

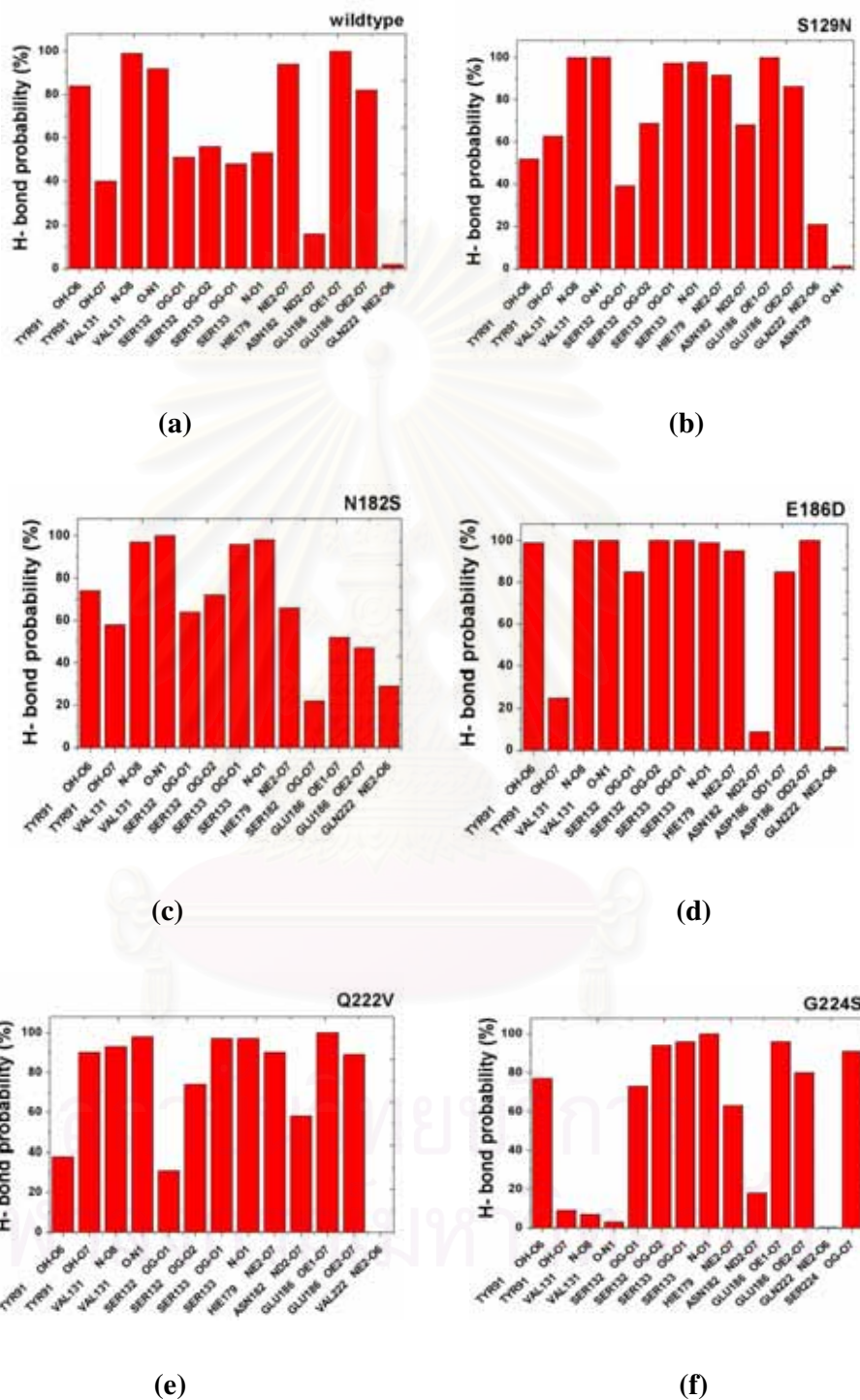
At the  $\text{-NHAc}$  sidechain, the very sharp peak with the maximum at 2.7 Å of the O8-N and N1-O (Val131) distances suggested that there are two strong hydrogen bonds with the backbone heteroatoms of Val131 in both wild-type and mutated hemagglutinins ( $> 90$  % occupation, Fig. 4.5). This is except for the G224S mutation in which these interactions were almost totally disappeared (Fig. 4.5 f), in good agreement with an increase of the two relative distances (O8-N: 7.2 Å and N1-O:  $\sim 5.3$  Å). The extremely long distances are not due to the rotation of the  $\text{-NHAc}$  group of

sialic acid in the direction away from the Val131 backbone (see Fig. 4.4). Instead, it should be raised from the rotation of *T2* by 30° as discussed previously (Fig. 4.2).

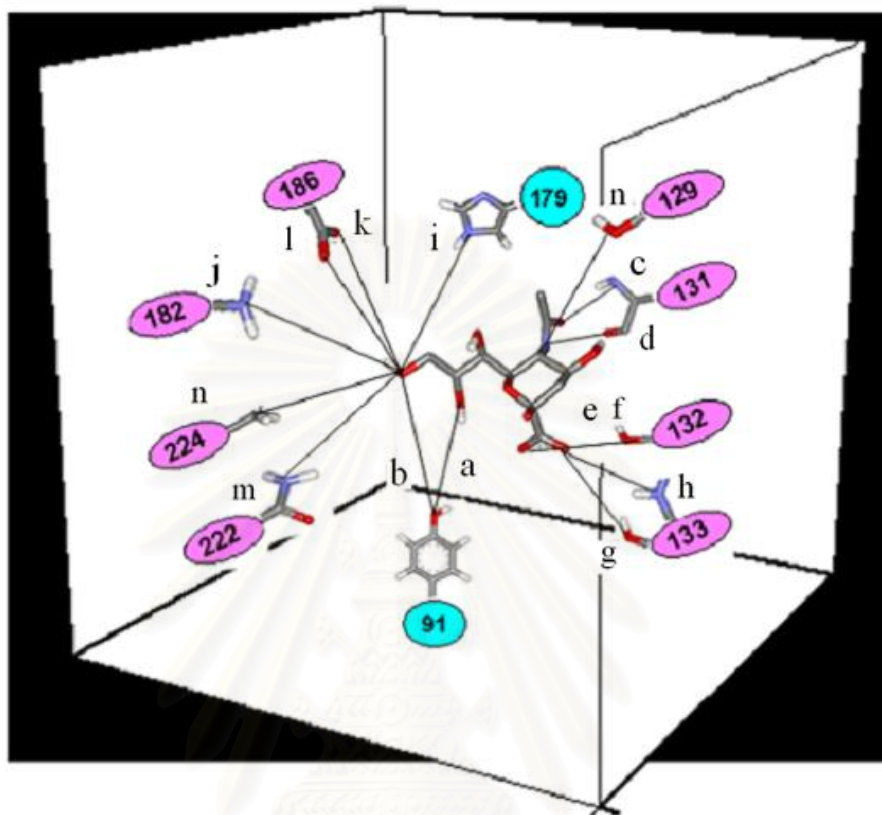
Dramatically detected hydrogen bond variations were observed for the hydrophilic group of sialic acid with three hydroxyl moieties substituted (Fig 4.4). Among the three hydroxyl oxygens, only two of them, namely O6 and O7, had hydrogen bond interactions with the HA binding residues where most of hydrogen bonds were observed at terminal one (O7) in all different strains as shown in Figs. 4.5 a-f.

In wild-type system (Fig. 4.5 a), the O7-oxygen mainly interacted with the carboxylate group of Glu186 and the imidazole ring of His179 with percentage of hydrogen bond occupation more than 80, and slightly formed a weak hydrogen bond with the hydroxyl group of Tyr91 (40 % occupation). This is due to the latter residue gave more hydrogen bond contribution to the O6-oxygen with > 80 %. In all mutated systems, the two hydrogen bonds with Glu186 were firmly detected as found in the wild-type (> 80 %). This is except the N182S mutation in which the smaller mutated residue, Ser182, had created an enlarger space for its neighboring residue Glu186 (Fig. 4.4) and, therefore, the carboxylate sidechain was probably able to move rather freely and lose interactions with the O7-oxygen of substrate (~ 50 % occupation (Fig. 4.5 c) and the lengthening O7-OE1 and O7-OE2 (Glu186) distances by ~ 2 Å (Figs. 4.7 k and l)). Higher flexibility of the Glu186 sidechain may interrupt the attractive interaction between its nearby residue His179 and the O7-oxygen as can be seen by 30 % reduction of the hydrogen bond (Fig. 4.5 c) and the broader peak of the corresponding distance ranging from 2.5 to 6.0 Å (Fig. 4.5 i). Loss of these important hydrogen bonds is a primary source of predicted binding affinity of substrate to the enzyme target (more details in the last section). This differs from the results of the E186D mutation in which all hydrogen bonds detected with the three residues, Tyr91, His179 and Asp186 (the mutated residue), were remained (> 80 % in Fig. 4.5 d and the hydrogen bond distances of ~ 3 Å in Figs. 4.7 i and j) and found to be similar to those detected in the wild-type.

Interestingly, the O7-oxygen in the S129N and Q222V systems had one more hydrogen bond with the amide group of Asn182 (~ 60% occupation (Figs. 4.5 b and e) and the O7-ND2 (Asn182) distance at 3.2 Å (Fig. 4.7 j)). Together with the higher affinity of substrate to the mutated S129N and Q222V strains in the next section, this lead us to conclude somehow that Asn182 could be a key residue in substrate stabilization in particular at the hydrophilic moiety of the sialic acid. On the other hand, it means that either S129N or Q222V mutations has induced more interactions with the human receptor, SA $\alpha$ 2,6Gal pentasaccharide. In the G224S system, a newly formed strong hydrogen bond between the O7-oxygen and the hydroxyl group of the mutated Ser224 (~ 90% in Fig. 4.5 f and the O7-OG (Ser224) distance of 2.7 Å in Fig. 4.7 n) was compensated with loss of interactions at the -NHAc moiety of sialic acid (mentioned above). Good rearrangement of sialic acid into the binding site of the G22S mutation leads to almost equal  $\Delta G_{bind}$  value with that of wild-type (Table 4.1 and more discussion below).

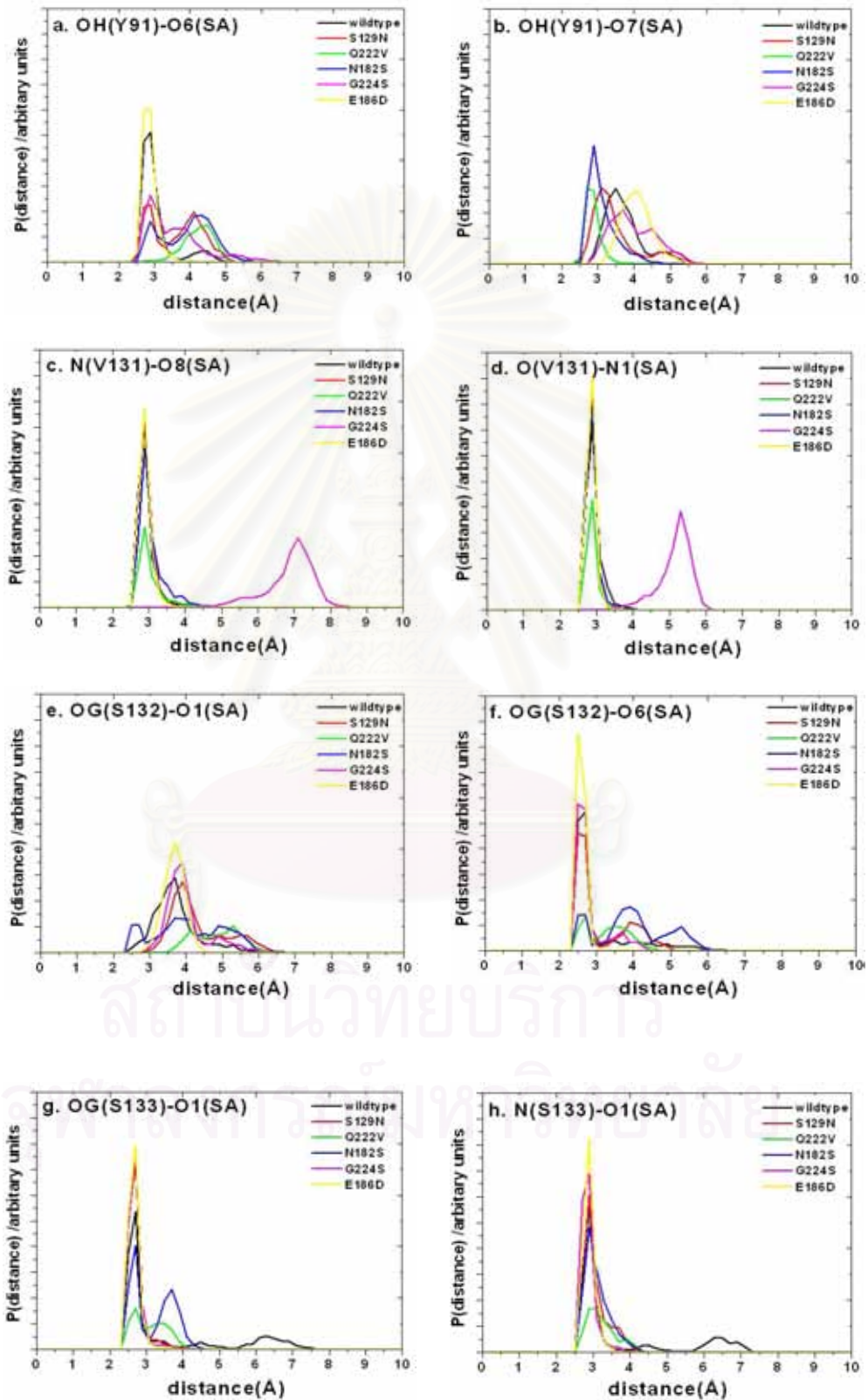


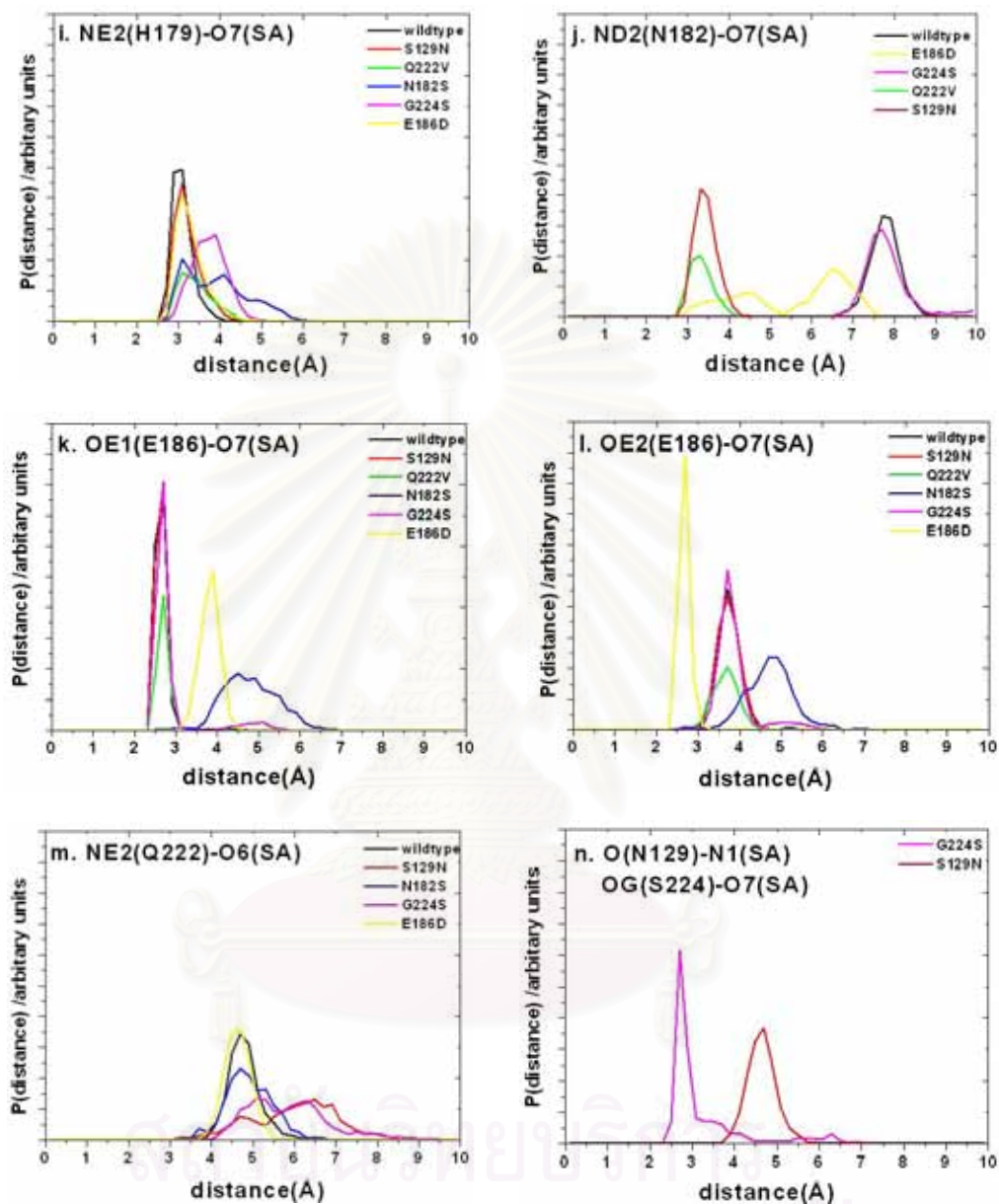
**Figure 4.5** Percent of hydrogen bond occurrence with amino acid residues in the binding pocket of hemagglutinin protein for all systems: (a) wild-type, (b) S129N, (c) N182S, (d) E186D, (e) Q222V and (f) G224S.



**Figure 4.6** Distance labeled of donor-acceptor of heavyatom between amino acid residues in the pocket of hemagglutinin and sialic acid by a simple model: (a)TYR91 OH-O6, (b)TYR91 OH-O7, (c)VAL131 N-O8, (d)VAL131 O-N1,(e)SER132 OG-O1, (f)SER132 OG-O2, (g)SER133 OG-O1, (h)SER133 N-O1, (i)HIE179 NE2-O7, (j)ASN182 N-O7, (k)GLU186 OE1-O7, (l)GLU OE2-O7, (m)GLN222 NE2-O6, (n) SER224 OG-O7 and ASN129 O-N1 .

จุฬาลงกรณ์มหาวิทยาลัย





**Figure 4.7** Distances distribution between donor-acceptor of heavy atoms of amino acid residues in the pocket of hemagglutinin and sialic acid for all systems: (a)TYR91 OH-O6, (b)TYR91 OH-O7, (c)VAL131 N-O8, (d)VAL131 O-N1, (e)SER132 OG-O1, (f)SER132 OG-O2, (g)SER133 OG-O1, (h)SER133 N-O1, (i)HIE179 NE2-O7, (j)ASN182 N-O7, (k)GLU186 OE1-O7, (l)GLU186 OE2-O7, (m)GLN222 NE2-O6, (n) SER224 OG-O7 and ASN129 O-N1 .

### 4.3 Binding free energy results

To estimate the binding affinity between hemagglutinin protein and SA $\alpha$ 2,6Gal pentasaccharide, the free energy was calculated based on the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) methodology. One hundred MD snapshots extracted from the production phase were used as a structural ensemble to evaluate the MM/PBSA binding free energy. The free energies for wild-type and mutant systems were calculated and separated into the electrostatic ( $\Delta E_{ele}$ ), van der Waals ( $\Delta E_{vdw}$ ) and binding energy ( $\Delta E_{MM}$ ) in the gas phase, and binding free energy in condense phase ( $\Delta G_{bind}$ ) components. The results were summarized in Table 4.1.

**Table 4.1** Calculated binding free energy and its components for wild-type and mutants of hemagglutinin-SA $\alpha$ 2,6Gal complexes.

	$\Delta E_{ele}$ [kcal/mol]	$\Delta E_{vdw}$ [kcal/mol]	$\Delta E_{MM}$ [kcal/mol]	$\Delta G_{bind}$ [kcal/mol]
N182S	-143.81 $\pm$ 15.41	-21.15 $\pm$ 4.12	-164.96 $\pm$ 15.65	-5.39 $\pm$ 5.55
E186D	-181.67 $\pm$ 13.84	-26.83 $\pm$ 3.74	-208.50 $\pm$ 13.30	-10.27 $\pm$ 8.64
wild-type	-182.58 $\pm$ 30.67	-25.78 $\pm$ 6.35	-208.36 $\pm$ 32.51	-11.02 $\pm$ 7.16
G224S	-217.85 $\pm$ 21.53	-31.20 $\pm$ 5.09	-249.05 $\pm$ 19.87	-12.56 $\pm$ 6.36
Q222V	-173.18 $\pm$ 14.76	-28.13 $\pm$ 3.26	-201.32 $\pm$ 14.63	-17.60 $\pm$ 6.64
S129N	-186.31 $\pm$ 27.31	-24.84 $\pm$ 4.37	-211.15 $\pm$ 27.61	-18.18 $\pm$ 9.10

An increase of  $\Delta G_{bind}$  in the mutant complexes with respect to the wild-type corresponds to a raise in the substrate recognition between hemagglutinin and SA $\alpha$ 2,6Gal human receptor. The predicted binding free energies for wild-type,



N182S, E186D, G224S, Q222V and S129N are -11.02, -5.39, -10.27, -12.56, -17.60 and -18.18 kcal/mol, respectively, indicating that the binding affinity of substrate to the Q222V and S129N were stronger than that of the native hemagglutinin H5 of avian. This is in good agreement with the enzyme-substrate interaction where more percentage of hydrogen bond occupation with the carboxylate and hydrophilic groups of sialic acid was found in the Q222V and S129N systems than that in the wild-type (Fig. 4.5). In contrast, the weaker binding affinity for N182S mutation has caused by a loss of hydrogen bond interaction between the hydrophilic group of sialic acid and the surrounding residues in particular Glu186 and His179 as mentioned above. However, no significant difference in binding free energy was found for the E186D and G224S systems in comparison with the wild-type system. Taking into account all the enzyme-substrate interactions, the change of single mutated residue from H5 to H3 sequence in Q222V and S129N systems possibly leads to the efficiency transmission of influenza A virus subtype H5 from avian to human.

## CHAPTER V

### CONCLUSIONS

The technique of molecular dynamics simulations was applied to search for detailed information on the intermolecular interactions of SA $\alpha$ 2,6Gal pentasaccharide embedded in the native hemagglutinin subtype H5 and mutated systems, S129N, N182S, E186D, Q222V and G224S. In wild-type strain, the sialic acid showed hydrogen bond interactions with the catalytic residues in three zones: (i) Tyr91, His179 and Glu186 (ii) Val131 (iii) Ser132 and Ser133 interacting with the hydrophilic, -NHAc and -COO<sup>-</sup> groups of sialic acid, respectively. Dramatically detected hydrogen bond variations were observed at three hydroxyls substituted on the hydrophilic group. In all mutated systems, the two hydrogen bonds with Glu186 were firmly detected as found in the wild-type. This is except for the N182S mutation in which the smaller mutated residue had created an enlarger space for Glu186 and lost interactions with the O7-oxygen of the substrate. Loss of the important interaction is found to be a primary source of the reduction in the substrate binding free energy c.a. 5 kcal/mol for N182S mutation relative to the wild-type system. In contrast, the binding of substrate affinity to S129N and Q222V mutation is stronger than what found in the native hemagglutinin by ~ 7 kcal/mol. This is in agreement with the intermolecular interaction where more number and percentage of hydrogen bond occupation at the carboxylate and hydrophilic groups of sialic acid was detected in the two mutations than the other systems, especially the presence of hydrogen bond with Asn182. This leads us to conclude, however, that Asn182 could be a key residue in substrate stabilization. On the other hand, either S129N or Q222V mutation has induced more interactions with the SA $\alpha$ 2,6Gal human receptor and possibly led to the efficient transmission of influenza A virus subtype H5 from avian to human.

### Further work

During the time of writing this thesis, the outbreak of new strain of 2009 influenza A (H1N1) virus initially spread from Mexico has expanded around the globe. World Health Organization confirmed that 4379 cases of infected in human for 29 countries and 49 of them died<sup>3</sup>. The new H1N1 virus contains the combination of gene segments of swine influenza, avian influenza and human influenza viruses. Sequence alignment of hemagglutinin H1 sequences between the new strain which isolated from human in California (A/California/04/2009(H1N1)), and those seasonal influenza viruses from various hosts including human, avian and swine, was performed. It appears that there are some changes of amino acid residues in the binding pocket of the new H1 strain. This might implicate different binding efficiency to the host receptor which may be developed as a research subject of the further study.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## REFERENCES

- [1] Lee, C. W.; and Yehia, M. S. Avian influenza virus: Comparative Immunology. Microbiology and Infectious Diseases 34 (2009): 301-310.
- [2] The Writing Committee of the World Health Organization Consultation on Human Influenza A/H5. Avian Influenza A (H5N1) Infection in Humans. N Engl J Med. 353 (2005): 1374-1385.
- [3] World Health Organization. Avian influenza: Epidemic and Pandemic Alert and Response (EPR) [Online]. (2008). Available from: [http://www.who.int/csr/disease/avian\\_influenza/en/](http://www.who.int/csr/disease/avian_influenza/en/) [2009, June 18 ]
- [4] Hien, T. T.; et al. Avian Influenza A (H5N1) in 10 Patients in Vietnam. N Engl J Med. 350 (2004): 1179-1188.
- [5] Anthony, W. M.; et al. Case Control Study of Risk Factors for Avian Influenza A (H5N1) Disease: Hong Kong 1997. The Journal of Infectious Diseases 180 (1999): 505-508.
- [6] Beigel, J. H. Influenza [Online]. (n.d.). Available from: <http://en.wikipedia.org/wiki/Influenza> [2008, November 3]
- [7] U.S. Department of Health & Human Services. Pandemics and Pandemic Scares in the 20th Century [Online]. (2008). Available from: <http://www.hhs.gov/nvpo/pandemics/flu3.htm>. [2008, December 16]
- [8] Ron, A. M. F.; et al. Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls J. Virol. 79 (2005): 2814-2822.
- [9] National Museum of Health and Medicine, Armed Forces Institute of Pathology, Washington, D.C., United States. Spanish flu hospital [Online]. (2008). Available from:

[http://commons.wikimedia.org/wiki/File:Spanish\\_flu\\_hospital.png26](http://commons.wikimedia.org/wiki/File:Spanish_flu_hospital.png26)

[2008, December 16]

- [10] Michael, W. D. The Influenza (Flu) Virus [Online]. (2008). Available from: <http://micro.magnet.fsu.edu/cells/viruses/influenzavirus.html>. [2008, December 16]
- [11] David, A. S.; Stephen, A. W.; Don, C. W.; and John, J. S. Deacylation of the hemagglutinin of influenza A/Aichi/2/68 has no effect on membrane fusion properties. Virology 184 (1991): 445-448.
- [12] Robert, G. W.; and Rudolf, R. Influenza virus A pathogenicity: The pivotal role of hemagglutinin Cell 50 (1987): 665-666.
- [13] Taisuke, H.; and Yoshihiro, K. Pandemic Threat Posed by Avian Influenza A Viruses Clin. Microbiol. Rev. 14 (2001): 129-149.
- [14] Protein Data Bank. 1JSN [Online]. (2007). Available from: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1JSN>. [2008, June 11]
- [15] Peter, P.; Kiyotake, T.; Masahiro, U.; and Richard, W. C. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology 61 (1974): 397-410.
- [16] Protein Data Bank. 3CL0 [Online]. (2007). Available from: <http://www.rcsb.org/pdb/explore.do?structureId=3CL0>. [2008, June 11]
- [17] Cox, N. J. Influenza Virus Life Cycle [Online]. (2008). Available from: <http://nursingcrib.com/influenza-virus-life-cycle/> [2008, July 6]
- [18] Young, K. C.; et al. Studies of H5N1 Influenza Virus Infection of Pigs by Using Viruses Isolated in Vietnam and Thailand in 2004. J.Virol. 79 (2005): 10821–10825.

- [19] Carolyn, B. B.; et al. Risk of Influenza A (H5N1) Infection among Health Care Workers Exposed to Patients with Influenza A (H5N1), Hong Kong. The Journal of Infectious Diseases 181 (2000): 344-348.
- [20] Debby, van R.; et al. H5N1 Virus Attachment to Lower Respiratory Tract. Science 312 (2006): 399-413.
- [21] Kumnuan, U.; et al. Probable Person-to-Person Transmission of Avian Influenza A (H5N1). N Engl J Med. 352 (2005): 333-340.
- [22] Jacqueline, M. K.; et al. Antibody Response in Individuals Infected with Avian Influenza A (H5N1) Viruses and Detection of Anti-H5 Antibody among Household and Social Contacts. The Journal of Infectious Diseases 180 (1999): 1763-1770.
- [23] James, S.; Ola, B.; James, C. P; and Ian, A. W. Glycam microarray technologies: tools to survey host specificity of influenza viruses. Nature Reviews Microbiology 4 (2006): 857-864.
- [24] Alisa, Z. M. Scientists Develop Flu Virus Early Warning System [Online]. (2008). Available from: <http://www.nigms.nih.gov/News/Results/NIGMS01102006.htm> [2008, November 15]
- [25] John, J. S; and Don, C. W. Receptor binding and membrane fusion in virus entry: The Influenza Hemagglutinin. Annual Review of Biochemistry 69 (2000): 531-569.
- [26] Toshihiro, I.; and Yoshihiro, K. Host-range barrier of influenza A viruses. Veterinary Microbiology 74 (2000): 71-75.
- [27] Gary, N. R.; Thomas, J. P.; Jeri, L. L; and Paulson, J. C. Differential sensitivity of human, avian, and equine influenza a viruses to a glycoprotein inhibitor of infection: Selection of receptor specific variants. Virology 131 (1983): 394-408.

- [28] Gary, N. R.; and James, C. P. Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology 127 (1983): 361-373.
- [29] Robert, J. C.; Yoshihiro, K.; Robert, G. W.; and James, C. P. Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates. Virology 205 (1994): 17-23.
- [30] David, C. Y. Avian Influenza [Online]. (2008). Available from: [http://www.medicalecology.org/diseases/influenza/influenza\\_h](http://www.medicalecology.org/diseases/influenza/influenza_h) [2008, July 16]
- [31] Baron, S. A. Influenza [Online]. (2008). Available from: <http://en.wikipedia.org/wiki/Influenza> [2008, July 16]
- [32] U.S. National Library of Medicine. National Center for Biotechnology Information [Online]. (2008). Available from: <http://www.ncbi.nlm.nih.gov/> [2008, November 4]
- [33] Needleman, S. B.; and Carl, D. W. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. J. Mol. Biol. 48 (1970): 442-453.
- [34] Verlet, L. Computer experiments on classical fluids thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159 (1967): 98-103.
- [35] Ryckaert, J. P.; Ciccotti, G.; and Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23 (1997): 327-341.
- [36] Pollock, E. L.; and Glosli, J. Comment on P3M, Fmm nad the Ewald method for large periodic Coulombic systems. Comp. Phys. Comm. 81 (1996): 93-95.

## VITAE

**NAME :** Nopphorn Kaiyawet

**DATE AND PLACE OF BIRTH:** 14 November 1984 in Bangkok, Thailand

**ELEMENTARY SCHOOL:**

1990-1996 Watbumpennuer

**SECONDARY SCHOOL:**

1996-2002 Setthabutbumpen

**UNIVERSITY EDUCATION:**

2002-2006 B. Sc. In Chemistry, Department of Chemistry, Faculty of Science,  
Chulalongkorn University, Bangkok

2006-2008 M. Sc. In Physical Chemistry, Department of Chemistry, Faculty of  
Science, Chulalongkorn University, Bangkok

จุฬาลงกรณ์มหาวิทยาลัย