COMPUTER-AIDED DRUG DESIGN OF POTENTIAL NEURAMINIDASE INHIBITORS FROM PLANT EXTRACTS

by

MUCHTARIDI

Thesis submitted in fulfillment of the requirements for degree of Doctor of Philosophy

September 2013

ACKNOWLEDGEMENT

In the name of Allah SWT, the Most Gracious and the Most Merciful, I offer my humble thank gratitude You for giving me the strength to complete this thesis.

First and foremost, my sincere gratitude and appreciation goes to my supervisors Professor Habibah A. Wahab and Dr. Aisyah Saad Abdul Rahim for their constructive ideas, criticism, guidance and patience throughout the long duration of preparing this thesis. Especially to Professor Habibah A Wahab, she guide me everything like brother. I learned a lot about “what is true researcher” from her. I could not say in words how great they have guided me to a lot of wisdom that I got. Hopefully I can implement this lesson to the end of my life. I think they have successfully guided me through some stressful times and were always willing to sharpen my understanding of this thesis and other academic activities.

I am also greatly indebted to Associate Professor Nornisah Mohamed, Professor Anas Subarnas and Professor Unang Supratman, for their valuable insights, comments, ideas and suggestions in the course of improving and completing this thesis.

I wish to thank Professor Munavvar Zubaid Abdul Sattar and Professor Syeid Azhar Sulaiman (Dean and Former Dean of School of Pharmaceutical Sciences, USM) and other Faculty members as well as the support staff of USM School of Pharmaceutical Sciences for their encouragement and kind assistance rendered to me throughout my study.

I wish to remember my beloved late mother, Rusmiaty bt Anda Hadi, you are one of the reasons I've been here. Thank you for your pray, I dedicate this PhD to you.

I also wish to thank to my friend of PhDS lab members; Dr. Yam Way Kit, Dr. Choi Sy Bing, Dr. Belal Najar Omar, Muhammad Yusuf, Nur Saira Khairul Ikram, Maywan Hariono, and Nci. Sufyan, for their assistance in completing my Thesis. I give special appreciation to M. Yusuf who kept encouraging me to finish this PhD. Of course, I will always remember my faithful close friend Ida Musfiroh to help me in any circumstances. I wish to thanks to Khairul Ikram and Family who had raised me like a son and brother.

A note of thanks also goes to all my colleagues in Universitas Padjadjaran, Dean of Faculty of Pharmacy and all of staff, especially member of Department of Pharmacochemistry.

Finally and most importantly, I would like to extend my gratitude and affection to my beloved Mia Kusmiatin (my wife) who be patient to serve me in everything in lovely, my daughter Widya Siva Gramitha, and my son Javier Hafiz Gibran who always pray that I'm back home. Thank you for providing me the support with your patience, love, encouragement, and inspiration that has greatly facilitated the completion of this challenging effort.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

1D-NMR	One dimension of Nuclear Magnetic Resonance
2D-NMR	Two dimension of Nuclear Magnetic Resonance
3D	Three-dimension
ADT	Autodock Tools
APT	Attached Proton Test
AUC	Area Under Curve
AIV	Avian Influenza Virus
Ala	Alanine
Arg	Arginine
Asp	Asparagine
CADD	Computer-Aided Drug Design
CD_3OD	Deuterated methanol
$CDCl_3$	Chloroform
C. perpringens	*Clostridium perfringens*
DEPT	Distortionless Enhancement by Polarization Transfer
DCM	Dichloromethane
DMSO	Dimethyl sulfoxide
DS	Discovery Studio
ESIMS	Electrospray ionization mass spectrometry
EtOAc	Ethyl acetate
FEB	Free Energy Binding
GA	Genetic Algorithm
GAF	*Garcinia atroviridis* fruits
GAL	*Garcinia atroviridis* leaves
GCL	*Garcinia celebica leaves*
Glu	Glutamic acid
GML	*Garcinia mangostana* leaves
GMP	*Garcinia mangostana* pericarps
GS4071	Oseltamvir carboxylate
GS4104	Oseltamvir esther
HA	Hemaglutinin
HBA	Pharmacophore Feature of Hydrogen Bond Acceptor
HBD	Pharmacophore Feature of Hydrogen Bond Donor

HRESIMS	High resolution electrospray ionisation mass spectrometry
HY	Pharmacophore Feature of Hydrophobic
IC_{50}	The concentration of the inhibitor that is required for 50 % inhibition of the enzyme
Ile	Isoleusine
IR	Infra Red
Ki	Inhibition constant : the dissociation constant of the Enzyme Inhibitor complex
LBDD	Ligand-Based Drug Design
LGA	Lamarrckian Genetic Algorithm
MeOH	Methanol
MLR	Multiple Linear regression
MUNANA	2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid
NA	Neuraminidase
NADI	Natural Drug Discovery
NCI	National Cancer Institute
Neu5Ac2en or DANA	5-N-acetyl-2-deoxy-2,3-didehydro-neuraminic acid
NI	Pharmacophore Feature of Negative Ionizable
OTV	Oseltamivir
PDB	Protein Data Bank
PI	Pharmacophore Feature of Positive Ionizable
PLC	Preparative Column Chramotography
QSAR	Quantitative Structure Activity Relationship
RMSD	Root Mean Square Deviation
ROC	Receiver operating characteristic
SBDD	Structure-Based Drug Design
Ser	Serine
SIV	Swine Influenza Virus
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultra Violet

LIST OF SYMBOLS

Σ	Sum
W	Weight of the hypothesis function
Disp	the distance between the center of a particular pharmacophoric sphere
Fit	Fit value
Å	Angstrom
MaxOmitFeat	Maximum Omitted Features
ΔG	The free energy
ΔG_{vdw}	The free energy of van der waals interaction
ΔG_{HBOND}	The free energy of hydrogen bond interaction
ΔG_{elec}	The free energy of electrostatic interaction
ΔGconform	The free energy of conformation
ΔG_{tor}	The free energy of torsion
ΔGsol	The free energy of solvation

REKABENTUK DRUG BERBANTU KOMPUTER UNTUK PERENCAT NEURAMINIDASE YANG BERPOTENSI DARI EKSTRAK TUMBUHAN

ABSTRAK

Neuraminidase (NA) dari virus influenza bertanggungjawab dalam proliferasi dan jangkitan virus, menyebabkan beberapa usaha untuk menemui dan mengoptimumkan perencat baru neuraminidase digiatkan. Tujuan kajian ini adalah untuk menemukan perencat baru NA melalui penggabungan kaedah berbantu komputer (model farmakofor dan pendokkan molekular) dan kaedah pemencilan berpandu bioassai dari sumber produk semulajadi.

Dua model Hypogen telah digunakan sebagai model sievers dalam pengskrinan maya (ROC AUC =0.91 and 0.80; r_{70}^2 = 0.92 and 0.90). Model ini telah disepadukan dengan pendokan molekular untuk menyaring 3000 sebatian dari pangkalan data NADI.

Keputusan penyaringan virtual telah menunjukkan bahawa sebatian zanton dari *G. mangostana* (manggis) adalah hit utama diperolehi, menjadikan pokok ini terpilih untuk selanjutnya dilakukan pemencilan sebatian.

Tujuh sebatian diperolehi daripada manggis memberikan perencatan lebih dari 80% melalui kaedah pemencilan sebatian berpandukan bioassai. Potensi kesan perencatan neuraminidase (NA) (*C. perfringens*-NA dan H1N1-NA) dalam kajian ini adalah seperti berikut: garsinona D> γ-mangostin> α-mangostin> garsinona C> 3-isomangostin> gartanin> 8-deoksigartanin. Tujuh sebatian ini menunjukkan pendokkan berkesan terhadap kawasan pengikat aktif NA.

Selepas itu, *G. atroviridis* dan *G. celebica* telah dipilih bagi mengukuhkan hipotesis

bahawa *Garcinia* sp memberikan aktiviti rencatan neuraminidase. Asid garsinia dan naringenin yang diperolehi dari buah dan daun *G. atroviridis* juga memberikan perencatan neuraminidase tetapi kurang aktif berbanding *G. mangostana*.

Friedelina, katekin and dua terbitan lanastona (metil-3α, 23-dihidroksi-17,14-friedolanstan-8,14,24-trien-26-oat dan 24E-3a,9,23-trihidroksi-17,14-friedolanostan-14,24-dien-26-oat) telah diperolehi dari daun *G. celebica* melalui kaedah pemencilan berpandukan bioassai. Katekin dan dua terbitan lanastona menunjukkan perencatan yang kurang aktif, sedangkan friedelina tidak aktif.

Hanya lima sebatian (garsinona C, garsinona D, asid garcinia, naringenin, dan katekin) menepati ciri-ciri "*drug-like*' melalui kaedah peraturan lima. Sebatian yang paling aktif didapati adalah garsinona D (IC_{50} = 6.7 μM)

COMPUTER-AIDED DRUG DESIGN OF POTENTIAL NEURAMINIDASE INHIBITORS FROM PLANT EXTRACTS

ABSTRACT

Neuraminidase (NA) of influenza virus is responsible for the proliferation and infections of the virus progeny, prompting several efforts to discover and optimize new neuraminidase inhibitors. The main aim of this study is to discover new potential neuraminidase inhibitor from the natural product source using computer-aided (pharmacophore modelling-molecular docking) drug design method combined with bioassay-guided isolation. Two Hypogen models were selected as screening sievers (ROC AUC =0.91 and 0.80; r_{70}^2= 0.92 and 0.90). These models were integrated with molecular docking to screen 3000 compounds from NADI database.

Virtual screening results showed that xanthone derivatives from *G. mangostana* (mangosteen) were the top hits, thus provide rationale to select this plant for further isolation. Seven compounds obtained from mangosteen showed inhibition more than 80 % by bioassay-guided isolation. The potency of neuraminidase (NA) (*C. perfringens*-NA and H1N1-NA) inhibitory effect in this study is as follows: garcinone D > γ-mangostin > α-mangostin > garcinone C > 3-isomangostin > gartanin > 8-deoxygartanin. These seven compounds were favourably docked to the binding site of NA.

Subsequently, *G. atroviridis* and *G. celebica* were chosen to further prove the hypothesis that *Garcinia* sp gave neuraminidase inhibition activity. Garcinia acid and naringenin were obtained from *G. atroviridis* fruits and leaves and showed potent against neuraminidase but less active compared to compounds from *G.*

mangostana.

Friedeline, catechin and two lanastone derivatives (methyl-3α, 23-dihydroxy-17,14-friedolanstan-8,14,24-trien-26-oat and 24E-3a,9,23-trihydroxy-17,14-friedolanostan-14,24-dien-26-oate) were obtained from *G. celebica* leaves by bioassay-guided isolation method. Catechin and two lanastone derivatives were less active while friedeline was inactive.

Only five compounds (garcinone C, garcinone D, garcinia acid, naringenin, and catechin) fullfill as drug-like according to “rule of five”. The most active compound of garcinone D showed the most potent activity against NA with IC_{50}=6.70 µM.

CHAPTER ONE

INTRODUCTION

1.1. Statement of the Problem

World Health Organisation (WHO) reported that influenza virus affected nearly 10-20% of the world population, causing millions of people to be hospitalized and resulted in a quarter to a half a million deaths per year (De Filette et al., 2005). Influenza virus can be subdivided into three types A, B and C, but type C virus is less pathogenic than A and B viruses. Avian and swine influenza are variants of type A virus; with avian influenza represented the most pathogenic virus (Varghese, 1999, Webster et al., 1997). Indonesia appeared to have the highest fatality rate caused by avian flu in the world, with 160 deaths from 192 infected cases as of February 2013 (WHO, 2013). Cases of avian flu in China, Indonesia and several other Asian countries in 2005 have led to awareness on the pandemic threat and prompted efforts to control this disease. In June 2009, the swine flu outbreak (SIV) has alerted WHO to classify it as a pandemic (level 6) (WHO, 2009) indicating widespread community transmission. The pandemic was declared to be over in August 2010 and Malaysia being also affected by this infection has reported 15,520 cases and 92 deaths up to August 2010 (Bernama, 2010).

In general, influenza A virus consists of a membrane-enveloped, segmented, negative-strand RNA with hemagglutinin (HA) and neuraminidase (NA) located on the surface. HA and NA play a major role in viral replication. HA is responsible for the attachment on the surface of the cell receptors that are connected with terminal sialic acid (Colman, 1994), whereas NA is responsible for the proliferation and

infections of the virus. In contrast to HA, NA has highly conserved amino acid residues in its protein sequence. For this reason, the design of anti-influenza drug is more attractive to the inhibition of NA. Existing NA inhibitors designed using structure based drug design (SBDD) method such as oseltamivir (OTV) and zanamivir (ZANA) are potent in treating influenza in human but suffer problem due to resistance. Whereas peramivir (Sidwell and Smee, 2002) showed significantly weak inhibition towards virus due to physiological effect (Bantia et al., 2006). Therefore, the discovery of new active compounds that are potent against NA and not susceptible to its resistances has become an important goal of drug discovery in anti-influenza drug and this has also been the goal of this project.

To date, new NA inhibitors have been developed from synthetic chemicals, whereas the use of bioactive compounds from natural product as starting materials is still relevant. Synthesis of oseltamivir for example, utilise shikimic acid, which cannot be obtained economically by synthesis but can be efficiently isolated from Chinese star anise (Anderson, 2008). This project thus aims to leverage both Malaysia and Indonesia's rich natural resources to identify natural products that can be used as drug components or the starting materials for drugs and drug products, especially in the case of anti influenza drug.

1.2. Swine Influenza Virus (SIV)

In 1918, Spanish flu was first documented as a global pandemic that killed 50 million people around the world with an estimated 550.000 people died in the US. Spanish flu is an epizootic disease which has many clinical and pathologic similarities to human influenza that appeared among pigs in the north-central, U.S. In 1930, Smith

first reported that swine flu was caused by a virus. SIV belongs to the viral family of Orthomyxoviridae of type A with H1N1 subtype. They are RNA viruses with a segmented genome that comprise eight negative-sense, single-stranded RNA segments (Webster et al., 1992). Two others pandemics appeared in 1957 and 1968 are known within the history of the 20th century (Webster et al., 1997). The viruses that appeared in 1957 (Asian, H2N2) and 1968 (Hong Kong, H3N2) pandemics were the products of gene reassortment between human and avian viruses, when avian hemagglutinin and basic polymerase 1 (PB1) genes were brought to the human population (Desenclos, 1996). There is evidence that the swine-origin influenza (H1N1) virus is more pathogenic than seasonal influenza A virus (Desenclos, 1996, Calore et al., 2011).

H1N1 influenza A virus was first detected in California USA in April 2009 and subsequently referred to as swine-origin influenza A virus (Reid and Taubenberger, 2003b, Fanning et al., 2002). The virus was found in swine where H3 and N2, (HA and NA, respectively) segments circulating concurrently in humans to replace the avian-like H1N1 by reassortment (Castrucci et al., 1993, Claas et al., 1998, Reid and Taubenberger, 2003a). This virus is not epidemic in pigs but is easily transmitted between humans as found in Mexico (Fanning et al., 2002, Michaelis et al., 2009). The arrangement of the genome into segments allows reassortment resulting a diverse strains (Reid and Taubenberger, 2003b). The latest H1N1 is a descendent of 1918 pandemic strain which obtained new H2, N2, and PB1 genes through reassortment with an avian influenza A virus (Kawaoka et al., 1989).

There are three types of SIV (H1N1, H2N2 and H3N2) found in pig and human (Scholtissek, 1994). The cells of the swine respiratory tract contain

sialyloligosaccharides receptor possessing both α-2,3-galactose and α-2,6-galactose of N-acetylneuraminic acid. α-2,3-galactose-N-acetylneuraminic acid is the favoured receptor for avian influenza viruses, whereas α-2,6-galactose-N-acetylneuraminic is the preferred receptor for mammalian influenza viruses (Ma et al., 2008, Scholtissek, 1994). This provides supportive evidence for the "mixing vessel" theory. In this evolutionary process, swine might play an important role as a "mixing vessel" (Ma et al., 2008), because swine are susceptible to infection with both avian and human influenza viruses. Genetic reassortment between human and avian influenza viruses occurs when these viruses co-infect an individual pig (Scholtissek, 1994, Scholtissek, 1990).

1.3. Avian Influenza Virus (AIV)

H5N1 avian influenza virus (AIV) is pathogenic for poultry and human (Liu et al., 2005). Highly pathogenic avian influenza (HPAI) began distributing in Hong Kong in 1997 (Subbarao et al., 1998). It circulated rapidly in Asia, Europe, and Africa since 2003 and claimed at least 53 human lives (Asmara, 2006). In February 2007, 83 people reported to have been infected by avian influenza that led to the deaths of 63 people. WHO reported that the highest number of cases came from Egypt (29 cases), followed by Indonesia (9), Vietnam (7), China (2) and Cambodia (1) in 2010 (WHO, 2010b). Previously, these countries have also reported human cases of AIV in 2009 (WHO, 2010a). In addition to the poultry, avian influenza virus could also infect mammalian thus experts feared the possible emergence of a new subtype of influenza virus that is able to transmit from human to human now that H5N1 virus is proven to have been able to spread from birds to human.

To date only viruses of H5 and H7 subtypes (and sometimes H9 subtype) have been proven to cause HPAI. Amino acid composition of HA, NA, NS and PB2 proteins contribute to the antigenic properties, virulence, and host specificity of the viruses (Alexander, 2007). AIV has the ability to perform reassortment and genetic mutation allowing the virus to change antigenic properties, pathogenicity, and host specificity. One of the factors that play a role in AIV infection is the match between the viruses with receptors on the surface of host cells. For the occurrence of infection, the AIV binds to the cell surface glycoprotein and glycolipid containing sialyl-galactosyl moieties [Neu5Ac(α2-3)Gal] or [Neu5Ac(α2-6)Gal)]. AIV isolated from chickens binds with [Neu5Ac (α2-3) Gal] while the human virus isolated has specificity against with [Neu5Ac (α2-6)-Gal] (Connor et al., 1994, Matrosovich et al., 2000, Thompson et al., 2006, Matrosovich et al., 1992), thus AIV from chickens could not easily infect to human (Matrosovich et al., 2004).

Research on the biomolecular aspect showed that in addition to sialic acid receptors α-2.6, which is the majority, human also has a network tracheobronchial ciliated cells containing receptors with functional α-2,6 bond-sialic acid and α-2,3-sialic acid in low proportion. α-2,3 sialic acid receptors can be found on pneumosit in the lower respiratory tract (Thompson et al., 2006). The existence of this receptor is probably the main entry of avian viruses to human.

1.4. The Structure of Influenza Virus

Influenza virus belongs to the Orthomyxoviridae family. It is subdivided into A, B, and C based on the antigenic differences of nucleoprotein (NP) and matrix (M1) protein. Avian virus is the type A virus (Figure 1.1) and it causes epidemic and

pandemic in human and animals (Oxford, 2000). The virus contains three main antigens on its surface, hemaglutinin (HA, neuramindase (NA), and M protein. Both HA and NA contained in total around 500 bumps covering envelope glycoprotein lipid bilayer derived from the host. HA is a trimeric glycoprotein consists of three subunits encoded in lipid membrane (Wagner et al., 2002). HA acts as the first access to the host cells on the surface that conjugated with glycoconjugate. Glycoconjugate contains ketosidically-α-N-terminal connected to acetylneuraminic acid residue (Couceiro et al., 1993). These glycoproteins contribute to the internalization through merging the virus envelope with the host cell and is assisted by neuraminidase enzyme (NA).

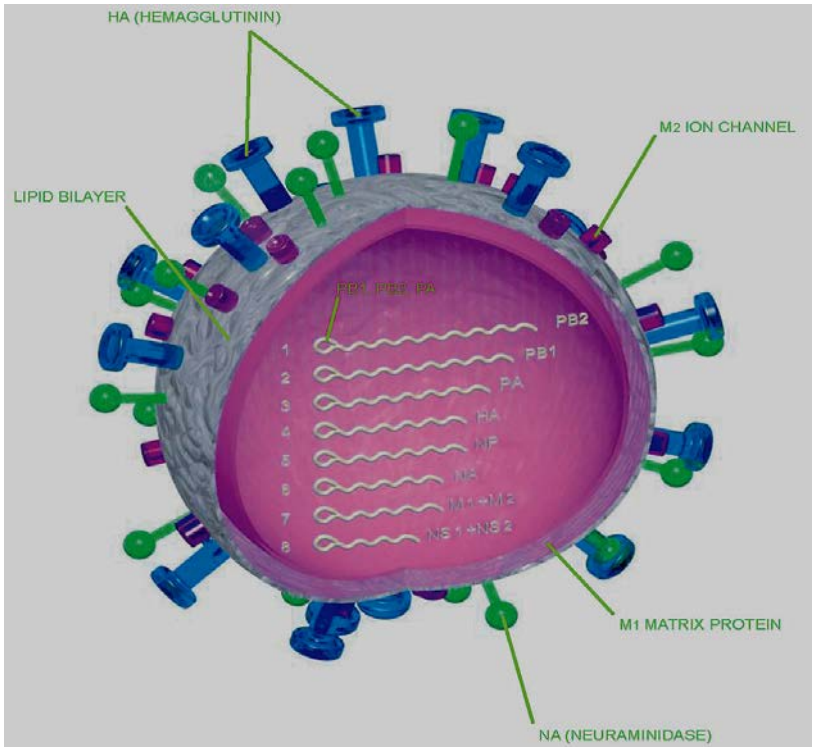


Figure 1.1. Structure of influenza A virus. The virus genome is made of 8 strands of RNA. The virus contains three main antigens on its outside: hemagglutinin, neuraminidase, and the M2 ion channel (Romanova, 2006).

Neuraminidase is an enzyme that prevents aggregation of the virions within the host cell and facilitates the cell-to-cell spread by cleavage of glycosidic linkages to sialic acid. The protein M1 controls the assembly of virion proteins and is situated in the envelope lipid bilayer. The M2 protein which is an ion channel bridges the inside and outside of the virus. In the viral envelope is a transcription complex required for viral replication. It consists of RNA nucleoprotein (NP) and, polymerase PA, PB1, and PB2 (Bardiya and Bae, 2005, Webster et al., 1992).

NA is a tetrameric glycoprotein which composed of four identical subunits (Colman and Ward, 1985) and acts as a glycohydrolase that eradicates α-ketosidically at N-terminal which is connected by acetylneuraminic acid residues from glycoconjugates (Colman and Ward, 1985). Serologic and genetic analysis showed that virus avian Influenza consists of 16 HA and 9 NA (WHO, 1980, Lee et al., 2005, Zhang et al., 2009). N1 is currently the most common pathogenic NA in circulation. A detailed structure of NA is explained below.

1.5. Neuraminidase (NA)

Neuraminidase (NA) is responsible in cleaving sialic acid in terminal receptors, releasing new viruses from infected cells. NA is found particularly in diverse virus families and bacteria, as well as in protozoa, some invertebrates and mammalian (Schwerdtfeger and Melzig, 2010, Sander-Wewer et al., 1982). They have differences in binding affinity or substrate preference, but they have conserved domains and structural similarities (Schwerdtfeger and Melzig, 2010). NA plays a vital role in influenza virus replication, and has a conserved active site residues thus inhibition of NA can delay the release of virus progeny from infected cells. This will

reduce the virus population and will give time for the immunity of the host cell in the body to eliminate the virus (Garman and Laver, 2004). NA hydrolyze α-2,3-sialic acid from sugar (galactose), and is also involved in hydrolysis at α-2,6-sialic acid-galactocyl, but this is less efficient especially if it is weak aglycon (sialic acid).

There are nine subtypes of neuraminidase from influenza A viruses (N1-N9) (Liu et al., 1995). Type A influenza neuraminidases form two genetically distinct groups: Group 1 consists of subtypes N1, N4, N5 and N8 while group-2 consists of N2, N3, N6, N7 and N9. Group-1 has a 150-loop cavity adjacent to the active site that serves as a gateway for the ligand to interact with NA (Russell et al., 2006). The cavity is suitable for the active site in the development of new anti-influenza drugs (Rudrawar et al., 2010).

1.6. The Active Site of Neuraminidase

The active site of NA has highly conserved active residues which are very specific to the sialic acid as the natural ligand. NA active site contains 18 residues (6 basic, 7 acidic, 3 polar, and 2 hydrophobic) (Colman, 1994). Based on the chemical bonding and interaction, NA active site can be divided into sub-pockets (Figure 1.2). Subsite 1 (S1) consists of triarginyl cluster (Arg118, Arg292, and Arg371), which has a pocket of positive charge; thus it will interact with the carboxylic groups of the ligand (Itzstein et al., 1996). S1 is called basic pocket which is important for designing lead compound for NA inhibitors (Taylor, 2009). S2 subsite is negatively charged and is composed of Glu 119 and Glu 227, and it interacts with the amine group on the acetamido of sialic acid. S3 consists of Trp178 and Ileu222 and has hydrophobic properties. The two residues are adjacent to Arg152 that binds to the

water molecules. S4 consists of Ala246 and Arg224, which are adjacent to the Ileu222 pocket and it is unoccupied by the functional groups of sialic acid (Stoll et al., 2003). The pocket accommodates a methyl group from SA and Neu5Ac2en (DANA) (Taylor, 2009). S4 is a new target for the development of new NA inhibitors. S5 has a unique pocket with mixed polarity environment depending on the incoming ligand. This site consists of carboxylate of Glu276 (trans-conformation) and methyl of Ala 246. During enzymatic reaction, Glu276 and Glu277 form hydrogen bonds with Tyr406 to stabilize the oxocarbonium ion with sialic acid. Glu276 interacts with O8-O9 in glycerol (SA) (Taylor and Itzstein, 1994). In addition to these amino acid residues, Asp151 has also an important role but not defined on the S1-S5 sites. This carboxylic residue does not make direct contact with DANA, but is believed to play an important role in catalysis by polarizing the bond α-2,3-sialic acid-glycosidic. Asp115 with Glu119 and Glu227 are also involved in sialic acid hydrolysis through the involvement of water molecules (Taylor and Itzstein, 1994).

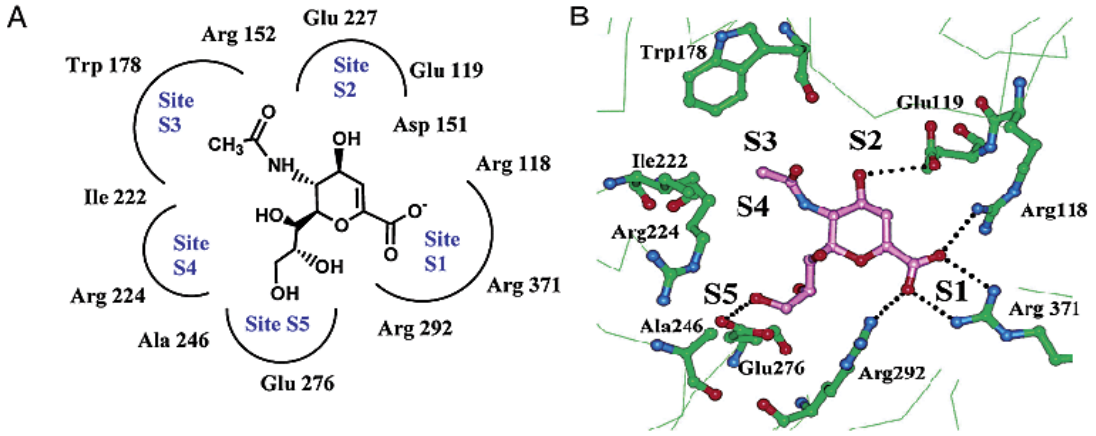


Figure 1.2. The interaction of DANA inside a neuraminidase active site in (a) 2D and (b) 3D representative (taken from PDB ID : 1NNB) (from Stoll et al., 2003).

The availability of three-dimensional of structure of NA has played an important role in the discovery of new inhibitor of neuraminidase. Varghese et al. (1983) determined the first three-dimensional of NA (N2 subtype structure) using X-ray crystallography at 2.9 Å resolution (Varghese et al., 1983).They also solved the first complex structure of NA-sialic acid in 1992 (PDB id; 2bat) (Varghese et al., 1992a).

To date, there have been many crystal structures of neuraminidase available in Protein Data Bank (PDB). Of the 109 neuraminidase proteins (155 from viruses, 52 from bacteria, and 39 from eukaryota) stored in PDB, only 141 protein neuraminidase were derived from influenza virus. They represent three subtypes of neuraminidase and which12 neuraminidase belonging to N1 subtype were found as shown in Table 1.1.

Table 1.1. List of some influenza A of neuraminidase structures of influenza A virus in Protein Data Bank (www.rscb.org).

PDB code	sub-type	Strain	Ligand	Resolution [Å]	Reference
3T16	N1	B/BEIJING/1/87	Apo	2.5	(Russell et al., 2006)
2HTY	N1	B/BEIJING/1/87	Apo	2.5	(Russell et al., 2006)
2HUO	N1	B/BEIJING/1/87	OTV	2.95	ibid
2HU4	N1	B/LEE/40	OTV	2.5	ibid
3BEQ	N1	A/BrevigMission/1/1918 H1N1	Apo	1.64	(Xu et al., 2008)
3CYE	N1	H1N1 1918	Apo	1.65	ibid
3B7E	N1	A/BrevigMission/1/1918	OTV	1.45	ibid(Xu et al., 2008)
3CKZ	N1	H5N1	ZANA	1.9	(Collins et al., 2008)
3CL0	N1	H5N1	OTV	2.2	ibid
3CL2	N1	H5N1	OTV	2.54	ibid
3NSS	N1	A/California/04/2009/H1N1	Apo	1.9	(Li et al., 2010)
4B7J	N1	H1N1 2009: Resistance I223R	OTV	2.42	(van der Vries et al., 2012)
1ING	N2	A/TOKYO/3/67	BANA5[1]	2.4	(Singh et al., 1995)
1INH	N2	A/TOKYO/3/67	BANA6[2]	2.4	ibid
1INW	N2	A/TOKYO/3/67	AXP[3]	2.4	ibid
1INX	N2	A/TOKYO/3/67	EQP[4]	2.4	ibid
1IVC	N2	A/TOKYO/3/67	BANA2[5]	2.4	(Jedrzejas et al., 1995)
1IVD	N2	A/TOKYO/3/67	BANA1[6]	1.8	ibid
1IVE	N2	A/TOKYO/3/67	BANA3[7]	2.4	ibid
1IVF	N2	A/TOKYO/3/67	DANA	2.4	ibid
2BAT	N2	A/TOKYO/3/67	Sialic	2	(Varghese et al., 1992b)
1NNB	N9	A/TERN/AUSTRALIA/G70C/75	DANA	2.8	(Bossart-Whitaker et al., 1993)
1NNC	N9	A/TERN/AUSTRALIA/G70C/75	ZANA	1.8	ibid
1INY	N9	A/TERN/AUSTRALIA/G70C/75	EQP	2.4	(White et al., 1995)
1MWE	N9	A/TERN/AUSTRALIA/G70C/75	Sialic	1.7	(Varghese et al., 1997)
2QWI	N9	A/TERN/AUSTRALIA/G70C/75	G20[8]	2	(Varghese et al., 1998)
2QWJ	N9	A/TERN/AUSTRALIA/G70C/75	G28[9]	2	ibid
2QWK	N9	A/TERN/AUSTRALIA/G70C/7	G39[10]	1.8	ibid
1BJI	N9	A/TERN/AUSTRALIA/G70C/75	DPC[11]	2	(Taylor et al., 1998)
1F8D	N9	A/TERN/AUSTRALIA/G70C/75	9-amino-DANA	1.4	(Smith et al., 2001)
1F8E	N9	A/TERN/AUSTRALIA/G70C/75	4,9-amino-DANA	1.4	ibid

[1]BANA5; 4-(acetylamino)-3-[(hydroxyacetyl)amino]benzoic acid
[2]BANA6; 4-(acetylamino)-3-[(aminoacetyl)amino]benzoic acid
[3]AXP; 4-acetamido-2,4-didexoy-d-glycero-β-d- galacto-octopyranosylphosphonic acid (an axial phosphonate)
[4]EQP; 4-acetamido-2,4-dideoxy-d-glycero-α-d-galacto-1-octopyranosyl)phosphonic acid
[5]BANA2; 4-(acetylamino)-5-amino-3-hydroxybenzoic acid
[6]BANA1; 4-(acetylamino)-5-amino-3-hydroxybenzoic acid
[7]BANA3; 4-(acetylamino)-3-amino benzoic acid
[8]G20; 4-acetyl-4-guanidino-6-methyl(propyl)carboxamide- 4,5-dihydro-2h-pyran-2-carboxylic acid
[9]G28; 5-n-acetyl-4-amino-6-diethylcarboxamide-4,5-dihydro-2h-pyran-2-carboxylic acid
[10]G39; (3R,4R,5S)-4-(acetylamino)-5-amino-3-(pentan- 3-yloxy)cyclohex-1-ene-1-carboxylic acid
[11]DPC; 5-acetylamino-4-amino-6-(phenethyl-propyl-carbamoyl)-5,6-dihydro-4h-pyran-2-carboxylic acid

1.7. Computer-Aided Drug Design (CADD)

The use of computational techniques has been shown to further increase the efficiency of drug discovery and development (Zhang, 2011, Marshall, 1987). Computer-aided molecular design (CAMD) or *in silico* or computer-aided drug design (CADD) is being applied to expedite and assist hit-to-lead selection, hit identification, optimize the absorption, distribution, metabolism, excretion (ADME) and profile toxicity (Kapetanovic, 2008).

CADD could be devided into; (1) ligand based design, (2) structure based design, and (3) *de novo* design. There are various methods of ligand-based drug design (LBDD) can be applied, if protein structures are unknown, such as this method Quantitative Structure Activity Relationship (QSAR) and pharmacophore modelling (Kapetanovic, 2008, Zhang, 2011). The knowledge of ligand such as pharmacological effect and bioactivity is important in LBDD. A ligand set with known activity can generate model by computational approach such as QSAR, pharmacophore modelling, and database mining (Zhang, 2011). The retrieval of all 3D structures from a database considered as being similar to a given target structure is comparable with 2D similarity searching. However, 3D similarity searching reduce the problem of conformational flexibility (Terfloth and Gasteiger, 2003). The successful stories of LBDD approaches in facilitating drug discovery have been reported by Kubinyi (Kubinyi, 2006, Kubinyi, 1993).

In SBDD, structural knowledge obtained from ligand–protein complexes (X-ray crystallography or NMR data) can primarily facilitate the design of focused structure-based libraries by optimizing ligand–receptor complementary interactions, in an effort to increase potency and specificity (Orry et al., 2006, Hubbard, 2010).

The applications of SBDD include the discovery of potent and selective HIV protease inhibitors (Rubin et al., 1978, Wlodawer and Vondrasek, 1998), thrombin inhibitors (Wagner et al., 1998, Bohm et al., 1999), and neuraminidase inhibitors (Itzstein et al., 1993). Recent example of this method, peramivir (BCX-1812) which is developed using structure based method utilise the crystallography structure of a highly conserved NA active site and its substrate interactions (Young et al., 2001, Babu et al., 2000).

On the other hand, De *novo* method is practically used when the ligand is unknown, with a known target. *De novo* ligand design will be able to test many structures in a short period of time and arrange them into a ranked list based on an accurate prediction of binding free energies since the latter reflects actual binding propensities (DeWitte and Shakhnovich, 1996, Bohm, 1998, Hartenfeller and Schneider, 2010).

One of CADD tools, which is the most popular in the last 10 years is virtual screening (Lengauer et al., 2004). Both LBDD and SBDD approaches are powerful technologies, which can be applied to virtual screening (VS) for lead identification and optimization (Zhang, 2011).

1.8. Neuraminidase as a Target for Drug Discovery using SBDD technique

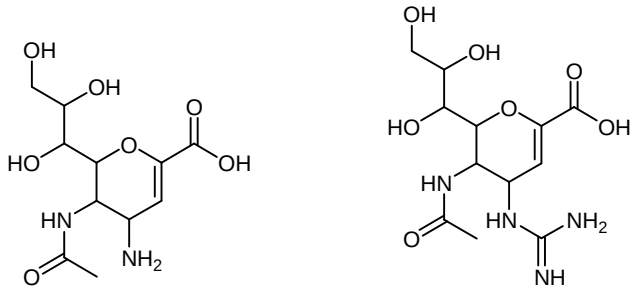
As mentioned in Section 1.7, SBDD is a powerful technique in the process of discovery and development of drug. SBDD approach is responsible for evaluating the complementarities and predicting the possibility of binding modes and affinities between ligands and their macromolecular receptors (Zhang, 2011). The availability of X-ray crystal structures of the influenza virus NA with and without a ligand such as α-Neu5Ac and Neu5Ac2en (Colman et al., 1983, Varghese et al., 1983, Varghese

et al., 1992a) provides the key in designing NA inhibitors. Edmond and co-workers (Edmond et al., 1966) together with Meindle's group (Meindl and Tuppy, 1969) started a random screening method in drug discovery. This method is based on guided activity that focused on trial and error, but the method did not work as the compounds easily produced drug resistant. Goodford (Goodford, 1985) calculated the interaction energy between ligand and the target by using computational method and found that interaction of both ligand and target can be predicted by using predictive software programs such as GRID (Goodford, 1985). The software program GRID forms the collection of computing resourch that performs task from multiple sites. The GRID programs has been used by von Itztein *et al.* (Itzstein et al., 1996) to design NA inhibitors using SBDD approach. Energy interaction between sialic acid (SA) and NA in the complex crystal structure becomes the basis for drug design a NA inhibitors (Wade, 1997).

It is possible to design highly potent NA inhibitors with SBDD as has been reported in the discovery of zanamivir (ZANA) (Taylor and Itzstein, 1994). Using GRID (Goodford, 1984, Goodford, 1985), the active site of NA is explored for the ability to accommodate a variety of groups such as carboxylates, amine, methyl and phosphate functional groups to get a potent and effective inhibiting NA (Itzstein et al., 1996). Several compounds have been successfully modified and optimized based on charge and shape of the character of active sites through SBDD methods, such as ZANA (Itzstein et al., 1996) and OTV (Kim et al., 1997). Based on the results of computational chemistry, von Itztein et al. (Itzstein et al., 1996) replaced the hydroxyl group at C-4 from the Neu5Ac2en with amine base groups into 4-amino-4-deoxy-Neu5Ac2en (Figure 1.3(a)) and further replaced with a guanidino group

(ZANA) (Figure 1.3(b)). Based on these data, C-4 group on the guanidino of ZANA successfully interact with carboxylic groups on the site active residues (Glu119 and Glu227) which leads to better inhibition of NA of Neu5Ac2en. The importance of NA in the history of the pathogenesis of influenza virus infection and the properties of the active side residue which is highly conserved lead to a concrete reason to design of small molecule, which is selective and effective towards NA.

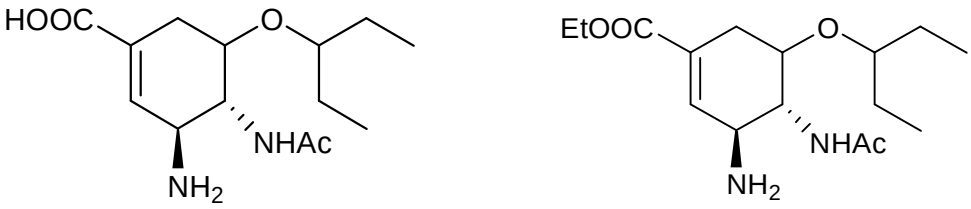
The glycerol moieties of ZANA interact with the active site of NA in a similar way to DANA. The *in silico* results showed that the replacement of glycerol with a more hydrophobic group makes the ligand more stable in solid form (oral administration), whereas ZANA is stable only in the form of solution (intravenous). In addition, QSAR studies explained that the replacement of glycerol with considerations chain length, branches, and stereochemistry of alkyl groups also improve the inhibition against NA. This is the basis in designing OTV (GS4071) (Figure 1.4(a)). Kim *et al.* performed optimization by the replacement of glycerol with 3-pentyl ether but maintaining acetamido and amino groups in the GS 4071(Kim et al., 1997). In this discovery, GS4104 (Figure 1.4(b)) has been developed for the purpose of drug formulations, which is the ester derivative of OTV (Lew et al., 2000, Lew et al., 1999, Li et al., 1998).



(A) (B)

Figure 1.3. The structures of (a) 4-amino-4-deoxy-Neu5Ac2en and (b) zanamivir.

Based on the success the discovery of ZANA (Itzstein et al., 1996) and OTV (Lew et al., 2000), SBDD has played an important role in the discovery of NA inhibitors. However, the constant threat of pandemic avian influenza (Monto, 2005) and the emergence of strains resistant to OTV (Tamiflu) makes the development of new effective NA inhibitor interesting for drug development research.



(A) (B)

Figure 1.4. (A) oseltamivir carboxylate (GS 4071) (B) oseltamivir (GS 4104).

1.9. Pharmacophore Modelling

IUPAC defines pharamcophore as the ensemble of steric and electronic features, which are required to ensure the optimal molecular interactions with a specific biological target structure and to trigger its biological response (IUPAC, 2007).

As QSAR theory that was developed by Hansch and Fujita showed that the sum of the steric, electronic, and hydrophobic effects of substituent's in a compound determines its biological activity, pharmacophoric patterns are also influenced by streometry, steric (atoms are accessible to the receptor), and electrostatic (Kubinyi, 1993). Technique of building pharmacophore is based on generating conformational search then mapping common groups in terms of a particular or specific atom type (e.g. hydrogen bond donor or acceptor), functional groups or some other similar properties. After that, the molecules are aligned and superimposed at the specified

points in a defined mode and then pharmacophore is created by joining the sites in common and calculating distances by averaging sites of superposition in a least-squares fit calculation. Spatial relationships between the pharmacophoric groups and relationships can be expressed as distances, angles, and torsions. The 3D (distances, angles, and torsions) pharmacophore features may describe the properties of the geometric points as to the type of interaction with the receptor (Langer et al., 2007).

Typically, pharmacophore features include hydrogen bond donor, hydrogen bond acceptor, aromatic, hydrophobic, positive ionizable, and negative ionizable. In comparing a ligand and a pharmacophore, the quality of the mapping were indicated by the fit value, a higher fit value represents a better fit.

1.9.1. Pharmocophore Software Tools

Gund et al. (Gund, 2000) developed the first pharmacophore software derived by using 3D database, but at that time (around 1974) only a few 3D databases available. The availability of 3D databases significantly increased the benefits to reach electronic pharmacophore models that could predict bioactivity. DISCO was the first pharmacophore software which incorporated 3D-searching in 1990 (Martin, 1995). GASP (Genetic Algorithm Superposition Program) was created for pharmacophore identification by Jones et al. (Jones et al., 1995). CATALYST was recently one of the most popular pharmacophore modelling. CATALYST algorithms utilized to generating pharmacophore hypothesis include HipHop, HipHoprefine, Hypogen, and Hyporefine (Poptodorov et al., 2005, Clement and Mehl, 2000).

HipHop algorithm is able to identify 3D-conformation of chemical features common to a set of molecules (Barnum et al., 1996). There are three steps to generate feature-

based 3D-pharamcophore aligment: (a) Generation of conformational model for each molecules in the training set is generated, (b) Evaluating of each conformer is evaluated for the presence of chemical feature, (c) Determination of 3D-configuration of chemical features to input molecule is determined (Clement and Mehl, 2000). Beside HipHop module, CATALYST also includes Hypogen pharmacophore generation algorithm that is a ligand-based QSAR tool using pharmacophoric overlap to predict activity (Li et al., 2000).

1.9.2. Basic Principle of Hypogen Algorithm

HipHop as mentioned above is a qualitative model that could not predict activity. However, it is also the aim of CATALYST pharmacophore modelling to generate quantitative model based on a set of compounds with biological activities to predict the activity (Chaudhaery et al., 2010) and it is called Hypogen. Hypogen correlates fit value with activity, thus the model generated can estimate the activity. Information of ligands of activities such as IC_{50} or Ki is used to construct hypotheses model that best correlates the activities between the estimated and that determined in experiment of assay (Li et al., 2000).

There are three phases in Hypogen: constructive, subtractive, and optimization phase. The first phase (constructive) is very similar to HipHop algorithm processing and uses a pruned exhaustive search of all the conformations (Poptodorov et al., 2005, Kurogi and Guner, 2001).

Hypogen identifies common configurations of features among discrete conformations of a set of molecules (Barnum et al., 1996). The training set is divided

into three categories; most active, moderate active and inactive. The most active ligands can be calculated by the equation in below:

MA * UncMA — (A/ UncA) > 0.0…….Equation 1.1

Where MA is the activity of the most active compound, **UncMA** is the uncertainty in the measured activity and **A** is the activity of the active compound. Unc is the uncertainty of the compounds and the value of an ‘‘Uncertainty’’ is 3 by default (Kurogi and Guner, 2001, Guner et al., 2004). Compounds are considered inactive when their activity is located at 3.5 logarithmic units below that of the most active compound.

Constructive phase limits to the maximum of eight (Li et al., 2000). HypoGen creates all acceptable pharmacophores containing up to five features among the two most active compounds. The two most active in the list an investigated to identify systematically by overlaying all their conformation, and only hypothesis that fit a minimum subset of features of the rest a active compounds are maintained (Poptodorov et al., 2005).

Subtractive phase will process programme to examine the hypothesis that were created in the constructive phase and deleted pharmacophores from the data structure that are not likely to be useful or hypotheses that fit inactive training compounds (Taha et al., 2010, Poptodorov et al., 2005, Li et al., 2000).

A particular training compound is considered as being inactive if the activity of the compound was 3.5 fold less than that of the most active compounds as following this equation:

log (A) — log (MA) > 3.5…….. Equation 1.2

where A and MA are the activity of the compound in the query and the activity of the most active compound, respectively.

In optimization phase, simulated annealing algorithm is employed to apply the predictive power of the hypotheses. Improving the score can be attempted by applying small perturbation in the algorithm to the pharmacophores created (Li et al., 2000). In the last step of optimization phase, quantitative extension of Occam's razor selects the simplest models to estimate the activity correctly (Kurogi and Guner, 2001, Li et al., 2000).

1.9.3. Cost function in Hypogen

Hypogen generates the best hypothesis by using Cost function. Cost function in Hypogen is used as the first step for validation process. Cost function is used to evaluate the power of activity data of a compound as an input parameter and dataset. CATALYST will generates two cost hypotheses (blank without activity data and dataset with activity) to show the significance of the results (Poptodorov et al., 2005). Fixed cost and null cost as theoretical costs are calculated in Hypogen as output for correlation data between biological data reference and prediction (Sutter et al., 2000). The value of weight cost is derived from gaussian form and the value increases as the feature deviates from an ideal (2.0). Root mean squares (*rms*) differences between the estimated and measured experiment activity for the data set molecules is calculated as error component value. The configuration component is called as constant cost. This value depends on the complexity of the hypothesis space being optimized and equal to the entropy of the hypothesis space. The configuration should not be greater than 17.0 in standard HypoGen mode.

1.10. Molecular Docking Simulation

Molecular docking is a key tool that combines computational and structural molecular biology to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure (Morris and Lim-Wilby, 2008). The general schematic of molecular docking is illustrated in Figure 1.5 (Morris and Lim-Wilby, 2008). The macromolecule as target and ligand must be first chosen based on the disease type and then the structure selected has to be prepared in correspondence with the needs of the docking method being employed. Finally, the results must be analyzed based on ranking of scoring function. Binding area consists of its position (x-, y-, and z-translations), orientation (Euler angles, axis-angle, or a quaternion), and, if the ligand is flexible, its conformation is defined by the torsion angles for each rotatable bond (Morris and Lim-Wilby, 2008).

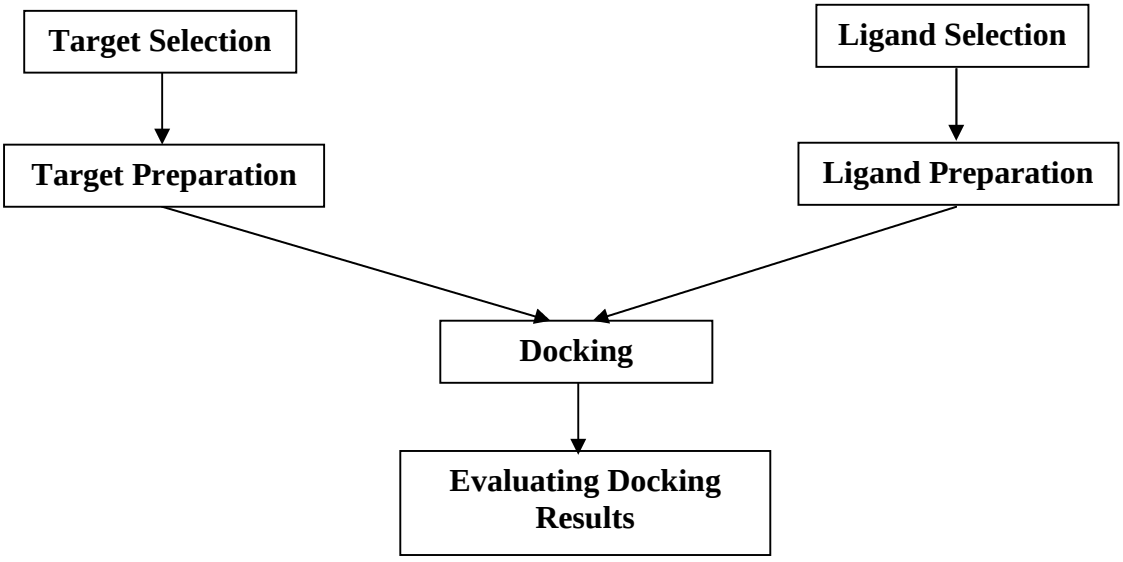


Figure 1.5. Schematic chart of a protocol for docking simulation (Morris and Lim-Wilby, 2008).

There are two terms to understand how well this method performs. The terms are matching and scoring, and this will explain how to determine successful docking between ligand and macromolecule (protein) (Kroemer, 2003, Rarey et al., 1999). Matching is a computational method to find optimum suitability between ligand and protein based on sterical factor evaluation and its electrostatic contribution which play a role in receptor-ligand interaction (Kroemer, 2003, Halperin et al., 2002), while scoring function is used to distinguish correct or wrong solution and sort the correct solution (Halperin et al., 2002).

It is a common concept for docking and fragments placement algorithms in computer-aided drug design (Rarey et al., 1999). The chemical matching algorithm generally ranks known inhibitors better than does matching based on shape (Shoichet and Kuntz, 1993). Matching methods is introduced with different treatment of ligand, conformational exploration, interaction ligand-protein representation, and prediction of binding affinity.

Algorithms approach was developed by considering ligand as flexible molecule (Moustakas et al., 2006, Rarey et al., 1999, Shoichet and Kuntz, 1993). Autodock algorithm used in this study is based on the concepts of natural selection, genetic algorithms, operated by creating a set of proposed solutions (‘population’) to a problem of interest, evaluating them (‘fitness pressure’), and then using the best solutions to develop a new set of proposed solutions (‘breeding’) (Douguet et al., 2000), (Pegg et al., 2001, Zabinsky, 2009).

In Autodock, local search technique is embedded with GA which transforms genotypic representations into phenotypic and is called Lamarckian Genetic Algorithm (LGA) that enhanced performance relative to simulated annealing and GA

alone. Three new search methods in Autodock 3 include a genetic algorithm; local search method; and adaptive global-local search method based on Lamarckian genetic algorithm.

Autodock uses semiempirical free energy force field to evaluate conformation during docking simulation. The force field is parameterized using ligand-protein complexes with known inhibition constant (Ki) (Morris et al., 2009).

Figure 1.6 describes how Autodock calculate the binding free energy. Ligand and protein initially present as unbound conformation states. In the first step, intramolecular energy is calculated from unbound conformation state of ligand that binds to protein. In the second step, force field evaluates ligand to protein intermolecular energy (in bound conformation) (Morris et al., 2009).

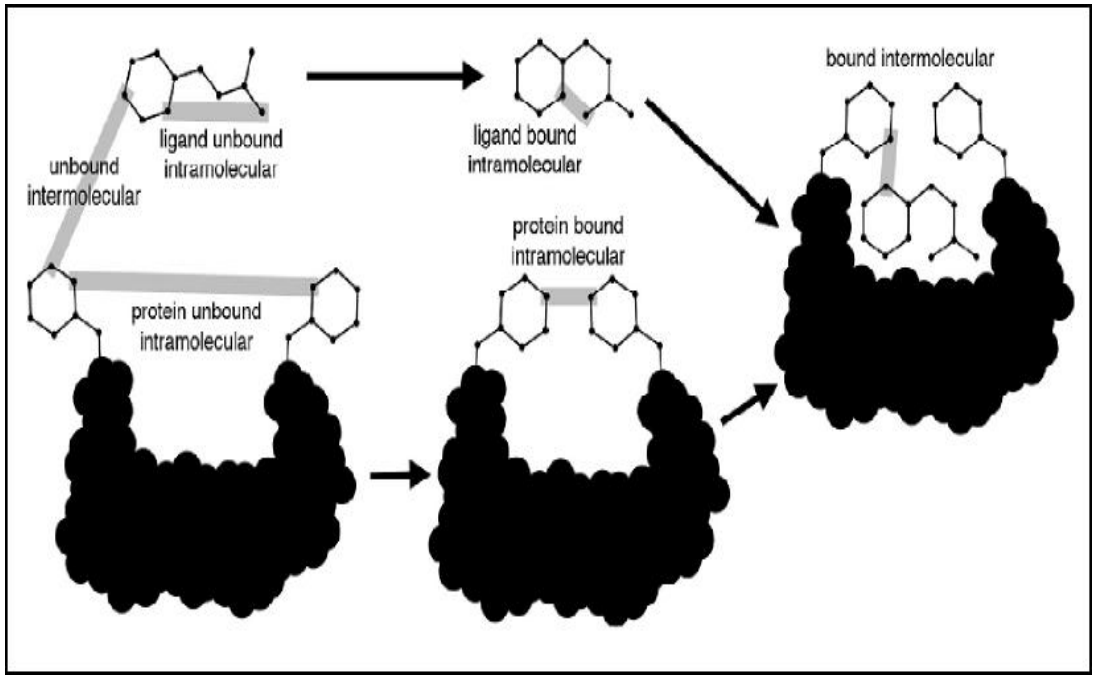


Figure 1.6. Illustration of calculation step of energy in autodock (Morris, *et al.*, 2009).

In Autodock, the equation to calculate free energy of binding also calculates entropic terms to the molecular mechanics, Equations 1.3 (Morris et al., 1998, Budin et al., 2001).

$\Delta G = \Delta G_{vdw} + \Delta G_{HBOND} + \Delta G_{elec} + \Delta G\text{conform} + \Delta G_{tor} + \Delta G\text{sol}$..... Equation 1.3

The above four terms are the typical molecular mechanics for hydrogen bonding, dispersion or repulsion, electrostatics, and deviations from covalent geometry, respectively; ΔGtor are the restriction of internal rotors, global rotation and translation; and ΔGsol is desolvation upon binding and the solvent hydrophobic effect wherein the entropy changes the interfaces between solute-solvent.

Autodock 3 software consists of three programs; AutoTors, AutoGrid, and AutoDock (Goodsell and Olson, 1990). AutoTors assists the input of ligand coordinates thus the ligand has flexibility with rotatable torsion angles and partial charge. AutoGrid facilitates to pre-calculate a three-dimensional grid of interaction energy based on macromolecular coordinates. It constructs three-dimensional grid surrounding the coordinates for the protein target. Finally, AutoDock performs docking simulation between ligand into macromolecules (Goodsell et al., 1996). In this step, the ligand explores six spatial degrees of freedom of rotation-translation and an arbitrary number of torsional degrees of freedom (Morris et al., 2008).

1.11. Ligand-Based Drug Design: Discovery of Neuraminidase Inhibitors

In recent time, sialic acid or Neu5Ac2en derivatives have been synthesized and evaluated for their influenza virus sialidase inhibitory potential. There at least 268 compounds of Neu5Ac2en derivatives have been synthesized worldwide (www.bindingDB.org) (Chen et al., 2002, Liu et al., 2007). In addition, molecular