MODELLING AND SYNTHESES OF VANILLIN DERIVATIVES TARGETING INFLUENZA VIRUS NEURAMINIDASE

by

MOHD RAZIP ASARUDDIN

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

November 2016

ACKNOWLEDGEMENT

This Thesis Project work has been carried out to meet the academic requirements of Universiti Sains Malaysia (USM) for the completion of Doctor of Philosophy in Pharmaceutic Technology. I would like to put on record, my appreciation and gratitude to all who have rendered their support and input. Without them, it would not have been possible for me to shape this study. I have received immense guidance from my supervisor, Professor Dr Habibah A Wahab. I would therefore like to convey my sincere gratitude to her. I would also like to thank Associate Professor Dr Nornisah Mohammed as my co-supervisor. I would like to convey my appreciation to Ms Khusaira Ikram, Mr Kevin Lim and all USM laboratory staff who gave me support and technical assistant in my chemical syntheses and biological evaluation works. I owe my deepest gratitude to University of Malaysia Sarawak for providing me all the financial support during my study period. I will always remain indebted to the Government of Malaysia for these great opportunities they provided to me in shaping my academic career. Finally I would like to thank my parents and all my family members for bearing my absence for a few years. I whole heartedly thank them all for sending me abundant love, encouragement and support all the way from home from their hearts. I dedicate all my success to each one of them.

TABLES OF CONTENTS **Page**

CHAPTER 2

METHODOLOGY

CHAPTER 4

DISCUSSIONS AND CONCLUSION

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

CNS	Central nervous system
DANA	2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid
FANA	2-deoxy-2,3-dehydro-*N*-trifluoroacetylneuraminic acid
3-D	3-dimension
HA	Hemagglutinin
HPAI	Highly pathogenic avian influenza
HMQC	Heteronuclear multiple quantum coherence
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HY	Hydrophobic
HYPO	Hypothesis
Vanillin	4-hydroxy-3-methoxybenzaldehyde
LPAI	Low pathogenic avian influenza
M	Matrix
4MU	4-Methyumbelliferone
MUNANA	2'-(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid
NIAID	National Institute of Allergy and Infectious Diseases
NADI	Nature-based Drug Discovery Intelligent
NI	Negative ionisable
NA	Neuraminidase
NP	Nucleocapsid
NEP	Nuclear export protein
NMR	Nuclear Magnetic Resonance
Ppm	Part permillion

PA	Polymerase A
PB1	Polymerase B1
PB2	Polymerase B2
RNA	Ribonucleic acid
SA	Sialic acid
SBDD	Structure-based drug design
TMS	Tetramethyl silane
UV	ultra violet
vRNA	viral RNA
vRNP	viral ribonucleoprotein complex
VS	virtual screening
WHO	World Health Organization

PEMODELAN DAN SINTESIS TERBITAN VANILIN YANG MENSASARKAN VIRUS INFLUENZA NEURAMINIDASE

ABSTRAK

Influenza tetap menjadi ancaman global yang serius. Ketika ini perencat neuraminidase (NA) iaitu zanamivir dan oseltamivir menjadi rintang kepada virus influenza. Oleh itu, perencat NA yang mempunyai perancah baru dan berkesan mesti dihasilkan. Melalui kajian ini, pendekatan pemodelan farmakofor digunakan untuk mencari perencat NA yang baru. Dua perisian komputer iaitu Discovery Studio 2.5 dan LigandScout 3.01 telah digunakan. Melalui perisian tersebut, pendekatan farmakofor ligan dan struktur telah dijana. Pendekatan farmakofor ligan dan struktur akan melengkapi antara satu sama lain. Pendekatan pemodelan farmakofor ligan adalah berdasarkan kepada kumpulan berfungsi sama ciri. Pendekatan ini menggunakan set latihan iaitu suatu set sebatian kimia atau ubat-ubatan yang telah terkenal digunakan sebagai perencat NA. Satu lagi pendekatan yang digunakan adalah pemodelan farmakofor struktur, LigandScout 3.01, melalui pendekatan ini interaksi antara perencat NA di dalam tapak aktif dapat digambarkan. Dalam kajian ini, 2HU0 (dari pdb) digunakan sebagai protin sasaran yang digabungkan bersama oseltamivir karboksilat. Sebatian yang menunjukkan skor penjajaran yang lebih besar atau hampir menyamai oseltamivir karboksilat akan berpotensi meningkatkan aktiviti perencatan NA. Hasilnya, kedua-dua pendekatan pemodelan farmakofor ini boleh meramalkan aktiviti perencat NA yang berpotensi. Sebatian yang berjaya melepasi langkah ini akan dinilai seterusnya untuk penilaian biologikal melalui asei MUNANA (aktiviti perencatan NA). Dari tiga ribu sebatian yang diimbas, lebih kurang empat puluh tujuh sebatian yang berpotensi sebagai perencat NA telah dikenalpasti. Beberapa sebatian daripada empat puluh tujuh sebatian tersebut seperti

betalain menunjukkan nilai kemuatan yang sangat baik menandakan ia berpotensi sebagai perencat NA tetapi sebatian-sebatian ini didapati tidak stabil (terurai) semasa proses pengekstrakan dan pemisahan. Sebatian yang mempunyai nilai kemuatan yang baik yang mempunyai kumpulan berfungsi hidroksil telah dikenalpasti dan dipilih. Sebatian yang mempunyai kumpulan hidroksil akan memeta samada melalui penerima ikatan hydrogen (HBA) atau penderma ikatan hydrogen (HBD). Ini akan membantu meningkatkan aktiviti perencatan NA oleh sebatian tersebut. Oleh yang demikian, sebatian yang mempunyai satu kumpulan hidroksil dan metoksi dijangka akan meningkatkan keupayaan pemetaan farmakofor yang melibatkan kedua-dua ciri HBA dan HBD. Oleh sebab itu sebataian benzena yang mempunyai satu kumpulan hidroksil dan metoksi dipilih. Perancah gelang benzena memberikan kelebihan seperti ketidakiralan, sintesis yang lebih ekonomik dan meningkatkan lipofilisiti berbanding dengan gelang dihidropiran. Penyaringan selanjutnya menemukan tiga puluh tiga sebatian perancah gelang benzena yang mempunyai kumpulan hidroksil dan metoksi. Perancah ini adalah sama dengan sebatian vanilin. Oleh yang demikian, pelbagai terbitan vanilin dengan kumpulan azomethin, oksazolin, pirazolin, benzena, sikloheksana, hidrazida dan asetat diperkenalkan melalui pendekatan separa sintetik bertujuan untuk meningkatkan aktiviti perencatan NA. Terbitan vanilin menunjukkan mod pemetaan yang baik sepertimana pemetaan farmakofor pada set latihan. Penjajaran dasar struktur menunjukkan perancah vanilin menunjang dalam tapak aktif NA seperti penjajaran pada oseltamivir karboksilat. Lima terbitan vanillin telah disintesis. Sintesis terbitan vanillin akan membuktikan analisis pemodelan farmakofor mengenalpasti potensinya sebagai perencat NA. Terbitan vanilin tersebut di asei menggunakan asei MUNANA. Keputusan asei menunjukkan bahawa terbitan vanilin menunjukkan aktiviti perencatan NA yang berkesan tetapi lebih lemah

berbanding dengan oseltamivir karboksilat iaitu satu-satunya agen perencat NA yang berada di pasaran. Daripada keputusan asei MUNANA, ternyata bahawa sebatian **20** menunjukkan perencat NA yang terkuat dengan nilai IC_{50} 0.18 µM berbanding terbitan vanillin yang lain-lain diikuti sebatian **18** yang menunjukkan IC_{50} 0.36 µM lebih kuat daripada DANA, **17**, **19** dan **21** dengan IC_{50} 0.44, 1.61, 1.22 dan 0.76 µM. Daripada pemerhatian ini, adalah disimpulkan bahawa kumpulan karboksilat memainkan peranan yang penting bagi meningkatkan aktiviti perencatan NA seperti yang dilihat melalui asei MUNANA.

MODELLING AND SYNTHESES OF VANILLIN DERIVATIVES TARGETING INFLUENZA VIRUS NEURAMINIDASE

ABSTRACT

Influenza remains a serious global threat. To date, zanamivir and oseltamivir used as NA inhibitors were reported to be resistant to influenza virus. Therefore, new and effective NA inhibitors must be discovered. In this study, pharmacophore modelling approach was used to search for NA inhibitors with new scaffold. Two computer softwares such as Discovery Studio 2.5 and LigandScout 3.1 was utilized. Through the softwares, the pharmacophore modelling approaches was generated. Ligand-based and structure-based pharmacophore modelling approaches will complement each other. Ligand-based pharmacophore modelling approach is based on common features functional groups. This approach utilize training set, a set of established compounds used as NA inhibitors. The training set was selected from chemically diverse NA inhibitors. The other approach is structure-based pharmacophore modelling. Through this approach, the interactions between NA inhibitors in active site can be observed. In this study, 2HU0 (from pdb) was used as protein target incorporated oseltamivir carboxylate. Compound which shows greater alignment score towards oseltamivir carboxylate will potentially enhance NA inhibition activity. As a results, both pharmacophore modelling approaches, potentially NA inhibitors could be predicted. Compounds which pass through this step was further to be evaluated for biological evaluation ie MUNANA assay (NA inhibition activity). From the three thousand compounds screened, about fourty seven compounds potentially as NA inhibitors was identified. Some compounds among the fourty seven compounds such as betalain showed very good fit value indicating potential NA inhibitors but these compounds was found to be unstable

(decomposed) upon extraction and isolation process. Compounds with good fit values bearing a benzene ring scaffold with a hydroxyl group were selected. Compounds bearing a hydroxyl group will map either by a hydrogen bond acceptor (HBA) or a hydrogen bond donor (HBD). This will proposed the enhancement of the compounds NA inhibition activity. Thus, it is proposed to search compounds which bearing a hydroxyl and a methoxy group which could enhance the pharmacophore mapping capacity involving both, HBA and HBD features. Therefore, benzene ring bearing a hydroxyl and a methoxy group are selected. The benzene ring scaffold gave the advantages of non-chirality, more economical synthesis and increased lipophilicity compared with the dihydropyran ring. Further screening found thirty three compounds with benzene ring scaffold bearing hydroxyl and methoxy group. The scaffold is similar to vanillin compound. Therefore, various vanillin derivatives with azomethine, oxazoline, pyrazoline, benzene, cyclohexane, hydrazide and acetate groups were introduced through semi-synthetic approach in order to enhance NA inhibition activity. The vanillin derivatives showed a good mapping mode corresponded to the training set pharmacophore mapping. The structure-based alignment showed that the vanillin scaffold anchored in the NA active site similar to oseltamivir carboxylate alignment. Five vanillin derivatives were synthesized. The syntheses of vanillin derivatives will prove the pharmacophore modelling analysis to identify its potential as NA inhibitors. The vanillin derivatives was assayed using MUNANA assay. The assay results shows that vanillin derivatives exerted significant NA inhibition activity but weaker compare to oseltamivir carboxylate which is an NA inhibitor found in the market. From the MUNANA assay results, it was revealed that compound **20** showed the strongest NA inhibition activity with IC_{50} value of 0.18 µM among other vanillin derivatives followed by compound **18**

which showed IC_{50} 0.36 µM greater than DANA, **17**, **19** and **21** with IC_{50} 0.44, 1.61, 1.22 and 0.76 µM, respectively. From this observation, it was concluded that carboxylate group plays an important role in enhancing NA inhibition activity as observed in MUNANA assay.

CHAPTER 1

INTRODUCTION

1.1 Influenza

The influenza virus is a causative agent which causes high morbidity and mortality. The word “Influenza” comes from the Italian word “influence” used since the middle age to describe the disease as it was originally believed that the disease came from the “influence” of the stars [1]. Influenza is a respiratory viral disease infecting the whole respiratory tract including the nose, sinuses, throat, lungs as well as the middle ear [2]. The influenza virus spread through respiratory secretions during coughing and sneezing with direct or indirect contact. After virus replication and the spread of infection throughout the upper and lower respiratory airways, the virus will assemble in the nasopharyngeal airway for five to ten days. The incubation period is about one to four days before obvious symptoms such as flu-like symptoms occurred. Multi-system complications will affect lungs, heart, brain, liver, kidney and muscle and consequently death due to either primary viral infection or secondary bacterial pneumonia [3].

Approximately 20% of children and 5% of adults have symptomatic influenza infection worldwide [4]. Complications from influenza illness such as pneumonia, cardiopulmonary or other chronic diseases lead to more than 200,000 hospitalizations. About 30,000 to 50,000 deaths per year were reported in the United States of America [5]. The risk for complications is higher among the elderly above 65 years old, young children, and those with certain underlying medical conditions. In fact, about 63% of all hospitalizations occur among persons aged more than 65

years old, of which 5-10% were fatal [6]. Hospitalization rates among children aged below 24 months are comparable to the rates reported among the elderly [7].

1.1.1 Influenza virus

Influenza infections are caused by negative strand RNA viruses of the orthomyxoviridae family of influenza virus type A, B and C. The three different types of influenza, A, B and C can be distinguished by antigenic differences between their nucleocapsid (NP) and matrix (M) proteins. Influenza virus type A is further classified according to their surface glycoproteins hemagglutinin (HA) and NA [8]. Influenza A viruses are subdivided into subtypes, and the current nomenclature system includes the host of origin, geographic location of first isolation, strain number, and year of isolation. For example: A/Swine/Iowa/15/30 (H1N1); where, A refers to Influenza A, Swine - the host of origin, Iowa - the geographic location of the first isolates, 15 - strain number and 30 is year of isolation, i.e. 1930. By convention, if the host of origin is human, then it is not included in the nomenclature such as A/Puerto Rico/8/34 (H1N1) [9].

There are sixteen known HA subtypes (H1-H16) and nine known NA subtypes (N1-N9) which give rise to many possible different influenza A subtypes. The subtypes of influenza A, such as H1N1 (1918 pandemic), H2N2 (1957 pandemic) and H3N2 (1968 pandemic) viruses have caused illnesses to people globally in the twentieth century [10]. Three prominent subtypes; H5, H7 and H9 are known to infect both birds and humans. Influenza A viruses are also classified into strains such as low and highly pathogenic on the basis of genetic features and the severity of the illness they can cause. Low pathogenic avian influenza (LPAI) are viruses which usually

associated with mild disease whereas highly pathogenic avian influenza (HPAI) viruses are associated with severe illness and high mortality in poultry [11]. All the three influenza viruses, A, B and C infect humans. However, they have different epidemiologic characteristics. Influenza A viruses caused the flu pandemic which could infect birds, mammals and humans. Wild birds are known to be the natural host of these viruses. Influenza B virus are known to infect humans and seals while influenza C virus are known to infect only humans and pigs [12].

1.1.2 Structure of influenza virus

Influenza virus shows pleomorphic morphology with an average diameter of 120 nm. **Figure 1.1** shows the virus surface glycoprotein membrane spikes which consist of about 500 HA and NA per virus, emerging from the lipid bilayer envelope. For influenza virus A and B, nucleocapsid proteins (NP) of matrix protein M1 contain eight viral RNA (vRNA) segments of different sizes. Each of them is associated with a polymerase complex comprised of polymerase B2 (PB2), polymerase B1 (PB1), and polymerase A (PA) proteins [12]. In viral polymerase complex, PB1 has the polymerase and endonuclease activities [13], PB2 binds to the methylated cap of host cell mRNA [10] and PA is essential for vRNA synthesis as well as for proteolytic activity [12].

The association of these four proteins namely NP, PB1, PB2, and PA as a whole represents the genetic transcription and is known as the viral ribonucleoprotein complex (vRNP) [14]. It is surrounded by the matrix protein (M1) which forms a layer underneath the lipid cell-derived envelope, providing rigidity to it and plays a key role in virus progeny assembly and budding process [15]. The M1 protein also

takes part in transporting the vRNP complex from the nucleus in combination with the Non-structural protein, NS2 or nuclear export protein (NEP), which is also present within the virus [16].

The eight gene segments of influenza A and B viral genome encode for ten proteins, nine of which are incorporated in the progeny virions. Six segments of the genome, PB2, PB1, PA, HA, NP and NA encode for one viral protein. Two open reading frames (ORFs) in segments 7 and 8 encode for M1 and M2 proteins and NS1 and NS2 proteins, respectively [17].

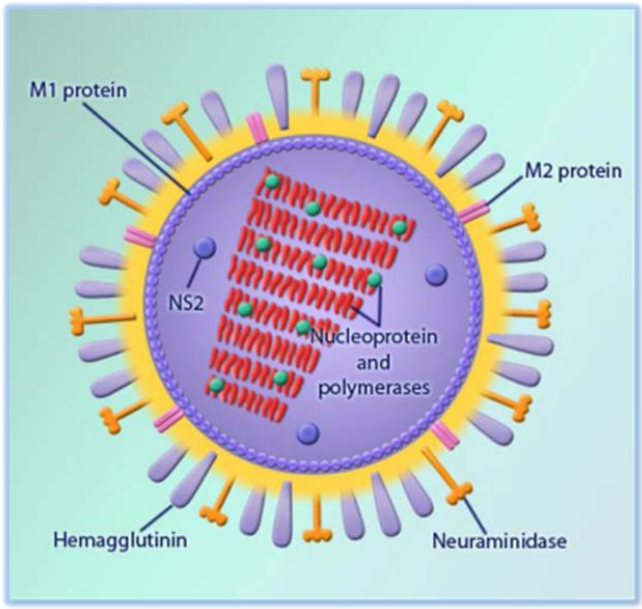


Figure 1.1 The cross section of influenza virus showing the structure of NA which plays a key role in the viral infections, adapted from Sidorenko & Reichl [12].

1.1.3 Life Cycle of the influenza virus

Figure 1.2 illustrates the life cycle of influenza virus and the involvement of NA and HA in the host cell involving multi-steps process. In the first stage, the virus will

fused the host cell through attachment of HA to Sialic acid (SA) on the cells surface of the host respiratory tract system. The host cell engulfs the virus and endocytosis process takes place. As the virus invades the host cell, acidic conditions takes place in the endosome resulting in HA fusing with the host cell vacuole's membrane. The ion channel (M2) opens and allows protons to rush in through the viral envelope and acidify the virus core. This situation promotes the core to dissemble and release the viral RNA and core proteins.

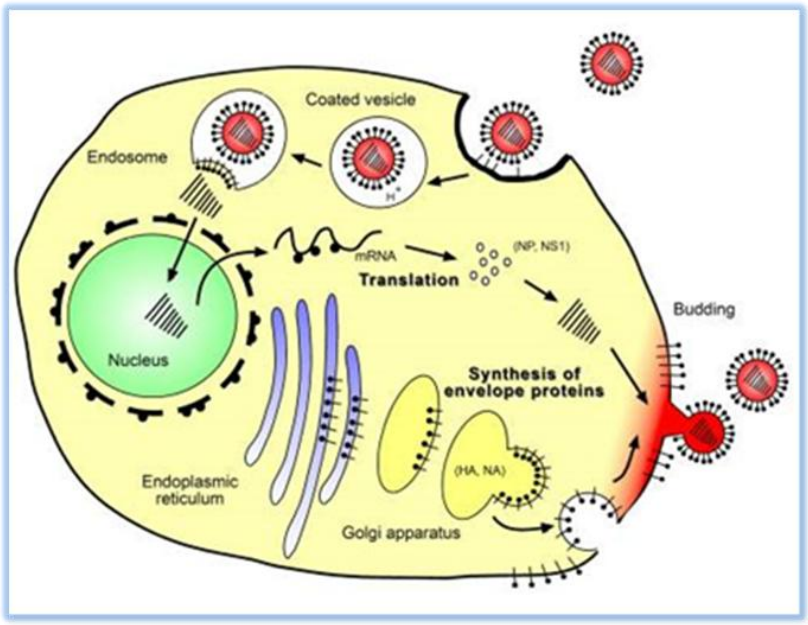


Figure 1.2 Replication cycle of influenza A virus showing binding and entry of the virus, fusion with endosomal membrane and release of viral RNA, replication within the nucleus, synthesis of structural and envelope proteins, budding and release of virus progeny capable of infecting neighbouring epithelial cells, adapted from Cox & Kawaoka [18].

Next, the core proteins and vRNA form a complex which was transported into the cell nucleus where the RNA-dependent RNA polymerase initiates the transcription process. The vRNA is then transferred into the cytoplasm or translated or remained

in the nucleus. The Golgi apparatus then secretes the newly synthesized viral proteins or transports it back into the nucleus to bind with vRNA and finally forms new viral genome materials and assembles them into a viral progeny.

HA and NA found on the host cell's surface glycoproteins were clustered into a bulge in the cell membrane. After leaving the nucleus, the vRNA and viral core proteins enter the cell membrane of the new viral progeny. The viral progeny becomes mature virus and emerges from the cell in a spherical shape from the host membrane, bearing HA and NA. In the last stage, the virus fused to the cell through HA and the mature virus is released once NA cleaves SA which is attached to HA [19].

1.2 Pandemic influenza

An influenza pandemic spreads the deadly disease rapidly around the world and transmits new influenza viral strain from animal species such as pigs, chicken and ducks to human as well as from human to human. Every year, many people are facing the risk due to the spreading of new virus strain. The first recorded influenza pandemic was an outbreak in 1580, which began in Asia and spread to Africa, and then to Europe along two corridors from Asia Minor and North-West Africa [20]. In the same year, the whole of Europe was infected from North to South over a six month period, and the infection subsequently spread to America. The illness rates were high which caused approximately 8,000 deaths reportedly from Rome and some Spanish cities [21].

The pandemics of 1830-1833 and 1918-1920 caused severe fatality cases [22]. The pandemic began in 1830 in China, from where it spread southwards to reach the Philippines, India and Indonesia, and across Russia into Europe. The contagion spread into North America to cause outbreaks in 1831-1832, recurred in Europe at the same time and recurred again in Europe in 1832-1833 [23]. The influenza pandemic was known to occur about three times in a century. The most deadly outbreak was the 1918 flu pandemic of type A influenza H1N1. The death toll was estimated to be approximately within 20 to 100 million people. The H2N2 1957 Asian (replacement of H1N1 subtype) and the H3N2 1968 Hong Kong (replacement of H2N2) pandemics were also associated with high death tolls [24]. The pandemics timeline is shown in **Figure 1.3**.

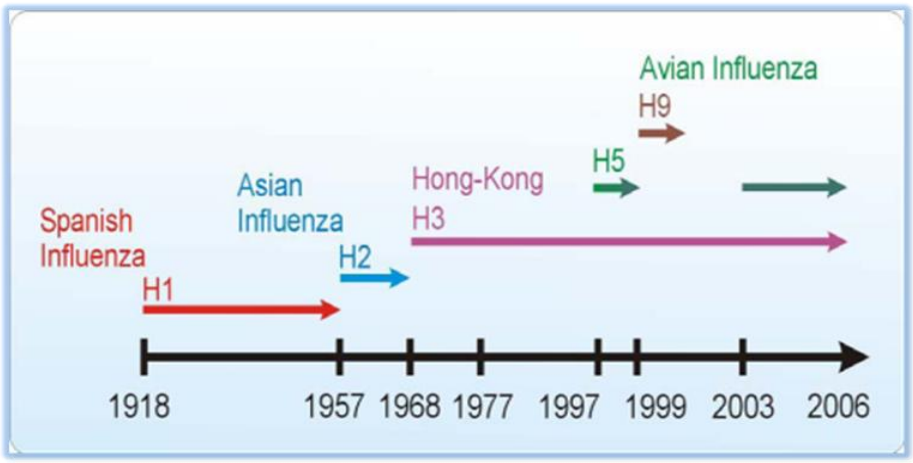


Figure 1.3 Sequence of antigenic shifts in the twentieth century, adapted from David et al., [25].

Figure 1.3 shows the influenza pandemics since 1918 until the emergence of HPAI A/H5N1 in 2006 [26]. The origin of the H1N1 1918 pandemic strain was associated with avian species, while H2N2 1957 and H3N2 1968 pandemics were due to combination of viruses with HA, NA and PB1 genes from avian influenza A viruses

[27]. This great pandemic death toll was due to an extremely high infection rate of more than 50% and the extreme severity of the symptoms caused by cytokine storms [28]. **Figure 1.4** illustrates the involvement of animal species in the pandemic found in avian and human species caused by the combination of HA and NA subtypes.

The influenza virus that caused the 1918-1919 pandemic was observed to make a consistency pattern that persists up to date according to scientists from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, USA [25]. The viruses continue to circulate for nine decades and any antigenic shift with genetics reassortment can result in novel and highly pathogenic strains of human influenza. New influenza viruses will constantly emerge by mutation. This new variant then replaces the older strains as it rapidly spreads to the human population [29].

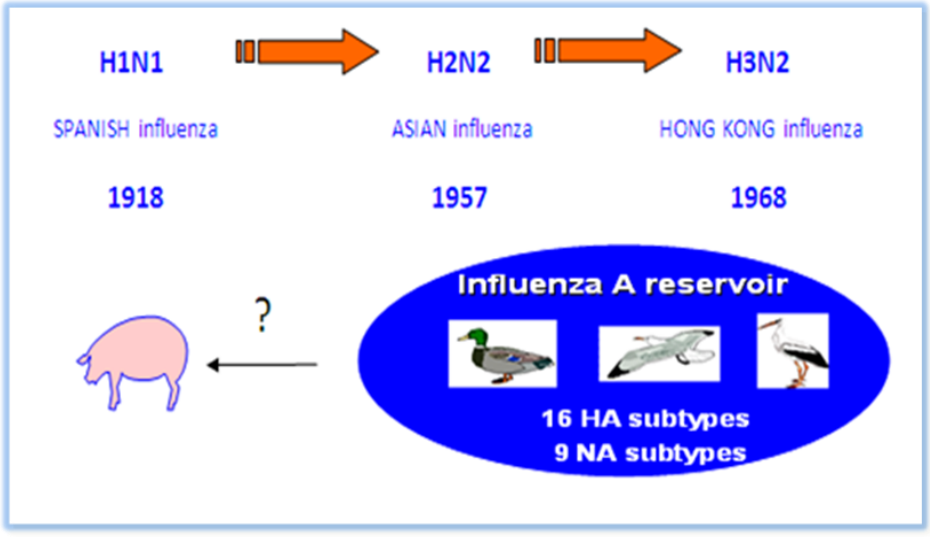


Figure 1.4 Circulation of Influenza A viruses in humans. Of the 144 total combinatorial possibilities, only three HAs and two NAs, in only 3 combinations (H1N1, H2N2, and H3N2), have ever been found in truly human-adapted viruses, adapted from David et al., [25].

As the two of the eight genes i.e. HA and NA promote sixteen and nine subtypes respectively, thus, there are about 144 possible HA-NA combinations. However, only three combinations, H1N1, H2N2 and H3N2 have ever been found in influenza viruses which can infect humans. Other combinations, such as avian influenza H5N1, occasionally infect human as they are bird viruses and not human viruses. The viral genes may assist the virus survival due to the new strength and the ability to infect new host. This situation was likely to initiate the 1918 pandemic. Human immune system was known to promote antibodies as defence factor against influenza virus HA and NA. Through this mechanism, the virus reacts by changing its character and resists the antibodies which recognize it and persists in environment until now [30].

1.3 Anti-Influenza Drug Target

The role of influenza NA in the lifecycle of the virus and the conserved nature of their active site has led the enzyme to be considered as an excellent target for antiviral drug design [31]. The discovery of NA inhibitors programme have been initiated due to the essential role of NA in the virus replication and its highly conserved active site [32]. The crystal structure of influenza virus NA complexed with both *N*-acetylneuraminic acid [33] and 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid (Neu5Ac2en or DANA) [34] proved to be useful in structure based rational drug design [35].

Numerous high resolution crystal structures of influenza NA and its complex with SA and other small molecule inhibitors have been determined [36]. Analysis of these structures showed some common structural features among all influenza NA: (a) NA active site contains well formed, large, and relatively rigid binding pockets; (b) the

key residues make direct contacts with the bound inhibitors are highly conserved and interact in a similar fashion; (c) NA active site contains a large number of polar or charged residues, suggesting that electrostatic interactions might play a critical role for any successful inhibitors [37]. The first clue as to the function of influenza virus NA came from experiments by Seto and Rott which showed that NA activity was associated with the release of virus from infected host cells [38].

1.4 Influenza Neuraminidase (NA)

NA is a glycoprotein that can be found in many life-forms such as viruses, bacteria, protozoa, mycoplasma and fungus including in higher animals, birds and mammals [39]. In influenza A and B viruses, NA adopts a tetramer structure that spikes out from the surface of the virus core [40]. The NA's structure is a box-like head on top of a stalk. The stalk length is considered to be one of the important factors that contribute to the virulence of H5N1 and H1N1 strains. The NA enables the virus to be released from the host cell to cause infection in human. The enzyme cleaves SA from the host cell receptor glycoproteins to promote influenza virus replication [41].

Influenza A virus NA can be classified into two groups; Group 1 and 2 (NA1 and NA 2). The NA 1 group consists of subtypes N1, N4, N5 and N8 while the other group includes N2, N3, N6, N7 and N9 [42]. Subtypes N1 and N2 are known to cause epidemics in human and subtype N3 or N7 are found to cause fatality to human [43]. All these nine subtypes are found in avian hosts and only N1 and N2 subtypes are found in influenza virus strains which infect human. The most prominent structural feature of group NA1 is the formation of a pocket located at the 150-loop

[44]. Contrary to other structurally determined N1 NAs, one of the 2009 H1N1 viruses, lacks the 150-pocket [45].

NA and HA both recognize carbohydrate structures and bind to terminal SA [46] units on the surface of the host cell. Binding of HA to its receptor initiates viral entry into the host cell. NA cleaves the α-(2,3) and α-(2,6) glycosidic linkage of terminal SA and reduces the number of receptor binding sites for HA on the host cells and the viral progeny. This will allow the mature virus to detach from the host cell during release and prevent self-aggregation mediated by HA [47]. Therefore, NA plays a crucial role for virus spread and infection. Inhibition of this enzyme therefore, would provide a potential mechanism to halt the infection.

1.4.1 The structure of Influenza NA

The NA structure composes of four identical subunits forming a tetramer structure of about 240 kDa. The tetramer has a mushroom-like shape, where the head is composed of four catalytic domains and the stalk is formed by the extended N-terminal sequences of the four subunits. The stalk amino sequence can vary in length and glycosylation for different subtypes [48]. In NA crystal structure, the subunits assemble around a four-fold axis. The typical β-sheet dominated six-blade propeller fold of each subunit was first observed in 1983 by Varghese et al. for N2 and is conserved for other subtypes (N9, N4, N6, N8, N1; Influenza B NA). In the centre of each subunit, conserved residues form the catalytic active site [49]. The primary NA structure of various species shows some 30% difference in sequence identity [50]. For instance, bacterial species NA has a conserved 'ASP-box' region which occurs 3 to 5 times in sequence which was not seen in the virus NA. There are some common

features within the catalytic domain observed from the NA X-ray crystal structure from various viruses, bacteria and protozoa. The catalytic site was located at the centre of the anomer structure. The anomer catalytic domain consists of six four-stranded antiparallel β-sheets folded into a propeller shape arrangement [51]. A tetramer crystal structure of NA, (PDB: 09N1) shown in **Figure 1.5** as a ribbon representation.

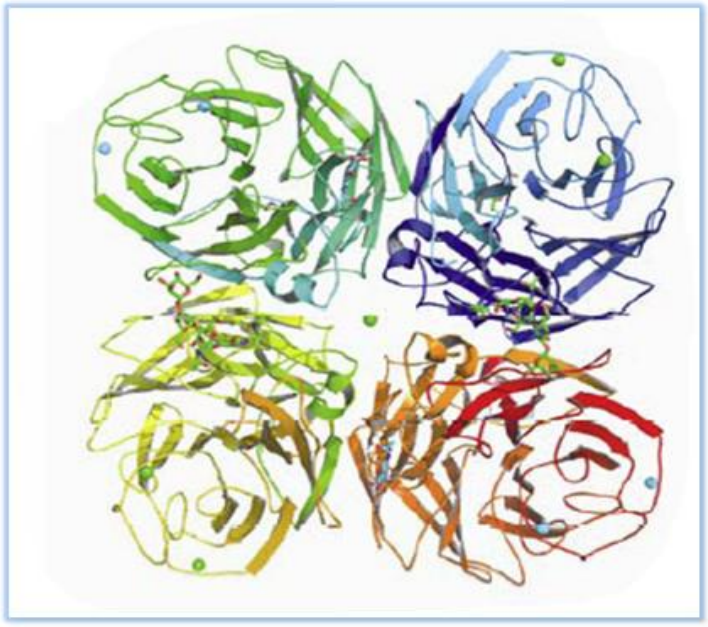


Figure 1.5 A ribbon representation of crystal structure of NA tetramer, adapted from Li et al., [45].

The tertiary structure (i.e. the anomer that is made up of a single polypeptide chain) is linked together by the disulfide bonds between the thiol groups of two peptidyl-cysteine residues in the active site. The active site is located inside the individual anomer and they contain conserved amino acids which will be discussed in detail in the following sections.

1.4.2 The active site of influenza NA and binding mode

Varghese et al successfully characterized the structure of N2 at a resolution of 2.9 Å in 1983 [52]. Four identical polypeptides of N2 subtype head was separated using the pronase enzyme. The tetramer protein molecules were crystallized and elucidated using X-ray diffraction. They identified each anomer of six four-stranded anti-parallel β-sheet arranged around a pseudo six-fold axis. The active site is known to be highly conserved as shown in **Figure 1.6**.

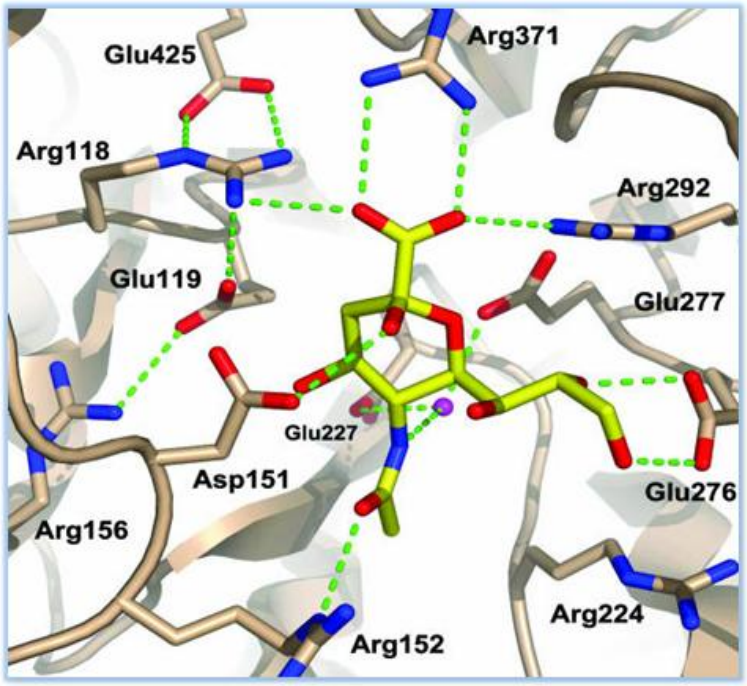


Figure 1.6 Structure of the complex between H2N2 influenza virus NA and SA (PDB code: 2bat). Hydrogen bonds are represented with green dotted lines, adapted from Kim et al., [53].

The residues that made up the active site include ARG371, ARG118, GLU119, ASP151, ARG152, ARG156, ASP198, ARG224, GLU227, ASP243, HIS274, GLU276, GLU277, ARG292, ASP330, LYS350 and GLU425. These amino acid residues were divided into three groups [42] and summarized in **Table 1.1**.

Table 1.1 The three groups of conserved amino acids of Influenza A NA active site.

Group	Amino Acid
Acidic	GLU119, ASP151, ASP198, GLU227, ASP243, GLU276, GLU277, ASP330, GLU245
Basic	ARG118, ARG152, ARG224, HIS274, ARG292, LYS350
Hydrophobic	TYR121, TRP178, LEU134

A number of high-resolution crystal structures of influenza NA and its complex with various small molecule inhibitors have been determined and are available from Brookhaven Protein Data Bank (PDB). Analysis of these structures revealed some common structural features. The NA active site contains some well-formed, large and relatively rigid pockets. All residues making direct contact with the substrate are strictly conserved and interact in a similar fashion with both substrate and inhibitor molecules and the active site contains an unusually large number of polar residues, suggesting that electrostatic interactions plays a key role for any inhibition activity. This is reasonable since NAs carbohydrate substrates are polar in nature and the electrostatic interactions promotes stability in the binding to NA inhibitors.

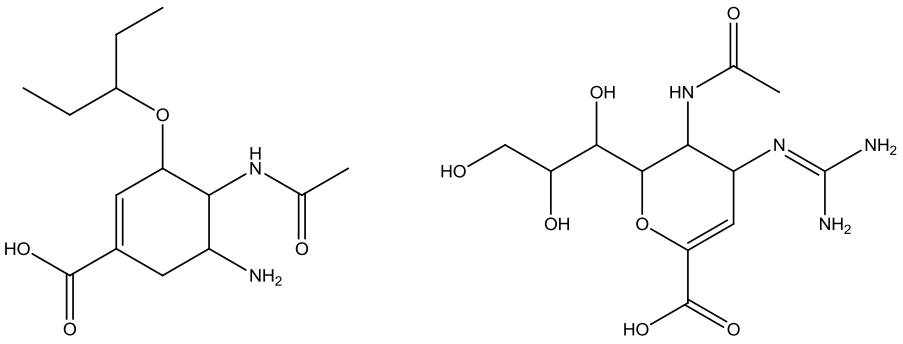
The negatively charged carboxylate group of SA makes strong charge-charge interactions with positively charged side chains of the ARG triad (ARG118, 292 and 371) of NA and the *N*-acetyl group of SA, opposite to the carboxylate group, making both polar and hydrophobic interactions with NA [36]. These two interactions help to anchor the scaffold of SA. The NA active site can be further divided into three major binding pockets. Pocket 1 is formed by GLU276, GLU277, ARG292 and ASN294.

This pocket interacts with the glycerol moiety of SA which proved to be the key to the binding of potent cyclohexene-based NA inhibitors [53]. There is a well-formed hydrophobic cavity of pocket 2 that is not utilized by SA for binding. This pocket is surrounded on one side by ALA246, on the other side by ILE222 and an ARG224 side-chain forms the bottom.

All the three residues are highly conserved and could provide hydrophobic interactions with potential inhibitors. Residues such as GLU119, ASP151, ARG152, TRP178, SER179, ILE222 and GLU227 of NA form Pocket 3. They interacted with both the C4 hydroxyl and *N*-acetyl groups of SA. Pocket 3 is very spacious and deep inside upon inhibitors binding in it which is not fully utilized by SA. Particularly, there are negatively charged residues such as GLU119, GLU227 and ASP151 surrounded C4 hydroxyl group which could be explored for possible charge-charge interactions.

1.4.3 NA inhibitors

Currently, two FDA approved antiviral drugs such as oseltamivir and zanamivir (**Figure 1.7**) act by inhibiting influenza virus NA thereby blocking the release of virions progeny, consequently reducing viral infectivity [54].



(a) (b)

Figure 1.7 Structure of oseltamivir (a) and zanamivir (b)

1.4.3.1 The discovery of zanamivir and oseltamivir

In late 1960s, 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid or Neu5Ac2en (DANA) was shown to have weak NA inhibitory activity. Little progress was made until after a complex crystal structures of NA with SA and DANA were obtained in early 1990's and the anti-influenza research was frequently highlighted [55]. The scheme of zanamivir and oseltamivir discovery shown in **Figure 1.8**, led to the discovery of oseltamivir carboxylate, the derivatives and analogue of oseltamivir ethyl ester.

Replacing the C4-OH group of transition state compound of sialic acid with a basic amino (NH_2) group have strengthen the binding by introducing additional electrostatic interactions with the negatively charged residues in NA active site [53]. The isopentyl side chain of oseltamivir carboxylate makes a hydrophobic interactions with NA active site [56]. The isopentyl group also binds to the highly polar active site [54]. However, experimental data show that oseltamivir carboxylate not only

interacts by using its hydrophobic isopentyl moiety but the interaction is absolutely essential to its inhibitory potency.

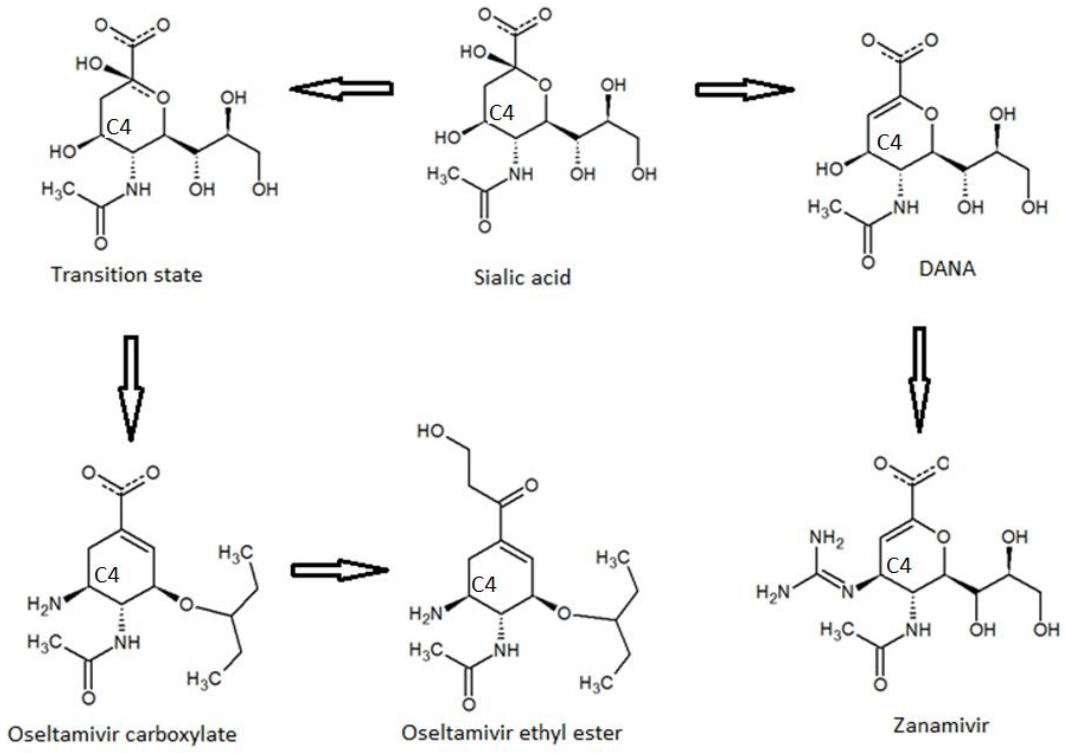


Figure 1.8 Structural modification of SA towards the development of oseltamivir and zanamivir, adapted from Willard et al., [57].

The nature of the interaction by 3-pentyl moiety is purely hydrophobic [58]. Oseltamivir carboxylate has similar inhibitory potency as zanamivir without having to incorporate permanently positively charged guanidino moiety due to the new and novel interactions between NA and oseltamivir carboxylate in NA active site. Introducing the polar carboxylate with an ethyl ester, the prodrug molecule, oseltamivir phosphate has a more balanced polar and hydrophobic property and is readily absorbed through oral administration [57].

Oseltamivir ethyl ester is the prodrug that requires ester hydrolysis to be converted to its active form, oseltamivir carboxylate [40]. The development of oseltamivir resulted from the combination of rational drug design and availability of high resolution X-ray crystal structures of SA and its analogues bound to influenza A and B NA [53]. After the identification of oseltamivir carboxylate as a potent NA inhibitor, modifications of the molecule was done in order to increase its oral bioavailability. Oseltamivir phosphate (Tamiflu, Roche) was licensed by the FDA in October 1999 for the treatment and prophylaxis of influenza infection [53].

The sialic acid which replaces C4-OH to guanidino group of DANA increases its basicity which was resulted in the formation of zanamivir as one of the most potent inhibitors. The highly polar nature of zanamivir makes the membrane permeability difficult to be delivered orally. Therefore, zanamivir is administered through nasal inhalation as dry powder [53]. These functional groups confers zanamivir a tight affinity for the active site of the viral enzyme NA [59].

Although DANA was shown to have weak NA inhibitory activity previously and little progress was made until after a complex crystal structures of NA with SA and DANA were obtained but in early 1990's, the anti-influenza research was frequently highlighted [55]. The discovery of zanamivir in 1989 was the outcome of Computer Assisted Drug Design (CADD) based on the crystal structure of the influenza NA active site resulting in the first potent, selective NA inhibitor [60]. Zanamivir has been commercially known as "Relenza" (GlaxoSmithKline) since 1999 when it was approved by the FDA for marketing in the United States followed by its approval in other seventy countries [61].

The crystal structure of oseltamivir carboxylate complexed with NA shown in **Figure 1.9** indicated that cyclohexene series of inhibitors adopted similar binding mode as SA, i.e., the carboxylate and *N*-acetyl moieties of the inhibitors interact with NA in a similar fashion to SA. The binding interactions in the NA active site anchored the compound scaffold. The crystal structure also revealed that the isopentyl side chain of oseltamivir carboxylate made hydrophobic interactions with pocket 2 [62]. Interestingly, the isopentyl group also binds to the highly polar pocket 1. Pocket 1 is typically considered a polar pocket because the presence of charged or polar residues and the polar interactions with a bound inhibitor's glycerol moiety are essential toward inhibitory activity [63].

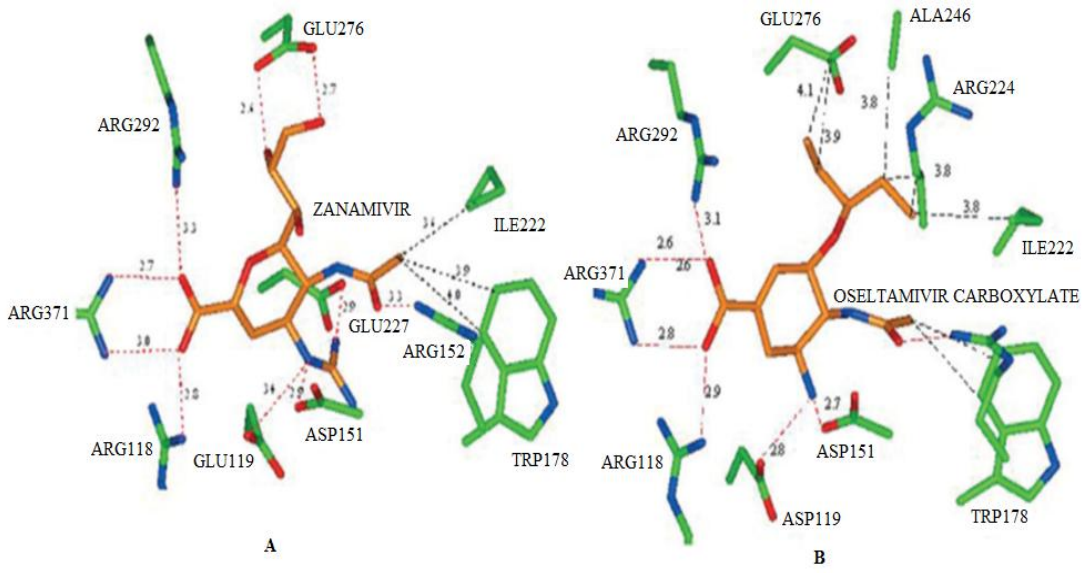


Figure 1.9 Complex structures of NA-zanamivir (A) and NA-oseltamivir carboxylate (B). Red dashes indicate polar interaction and black dashes indicate hydrophobic contacts, adapted from Chun et al., [64].

However, experimental data show that oseltamivir carboxylate not only interacts directly with this pocket using its hydrophobic isopentyl moiety but the interaction is

absolutely essential to its inhibitory potency. The nature of this interaction is purely hydrophobic [65]. Oseltamivir carboxylate has similar inhibitory potency as zanamivir without having to incorporate permanently positively charged guanidino moiety due to the new and novel interactions between NA and oseltamivir carboxylate in pockets 1 and 2. Introducing the polar carboxylate with an ethyl ester, the prodrug molecule oseltamivir ethyl ester has a more balanced polar and hydrophobic property and is readily absorbed through oral administration [57].

1.5 Binding of oseltamivir and zanamivir to H1N1 NA active site

The predicted binding constant for oseltamivir is at least 10 times higher than that for SA in H1N1 NA active site. **Figure 1.10** showed the NA active site of H1N1.

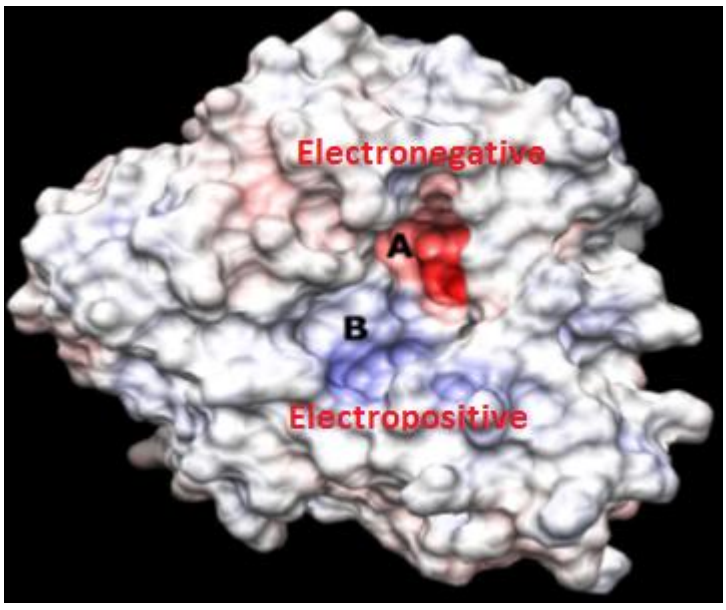


Figure 1.10 A model of NA active site using AutoDock showing the two regions of the H1N1 NA binding site. The red region (A) is the electronegative region, where the amine-rich functional groups of ligands bind, and the large blue region (B) a large electropositive zone, adapted from Lin et al., [66].

The hydrogen bond-rich interaction of SA has not resulted in higher binding affinity. The higher affinity of oseltamivir is probably related to interaction between the amine-rich terminus of the ligand and glutamic acid in the binding site. In their report, zanamivir had a binding affinity similar to that of sialic acid, with docking energy of -12.84 kcal/mol and -12.47 kcal/mol, respectively [67]. Zanamivir also has seven possible hydrogen bonds at the binding site. The significantly higher binding affinity of oseltamivir (-13.94 kcal/mol) as compared with zanamivir, even though both have a similar number of hydrogen bonds indicated that there may be a structural parameter that impacts binding that was not determined in previous studies [68].

1.6 Exploitation of the 150 Cavity in NA active site

All of the nine known NA subtypes from influenza A virus can be divided into two groups on the basis of phylogenetic analysis: Group 1, consisting of N1, N4, N5, and N8 subtypes, and Group 2, consisting of N2, N3, N6, N7, and N9 subtypes [69]. X-ray crystallography of several NAs revealed that the active sites of Groups 1 and 2 differ markedly: In Group 1, a loop of amino acids, consisting of residues 147−152 (150 loop), adopts an open conformation, whereas in Group 2 subtypes, this loop is closed [70]. The 150 cavity of NA active site was shown in **Figure 1.11**.

The open loop conformation, a cavity near the active site (150 cavity) becomes accessible to other ligands of the N1, N4, N5, and N8 subtypes. The discovery of the 150 cavity has led to the development of several inhibitors designed to exploit contacts in this region and increase specificity [71]. However, the recent crystal structure of a partially open 150 loop in a complex of N2 with oseltamivir and

several *in silico* studies have indicated that movement of the 150 loop may not be restricted to group 1 NAs [72].

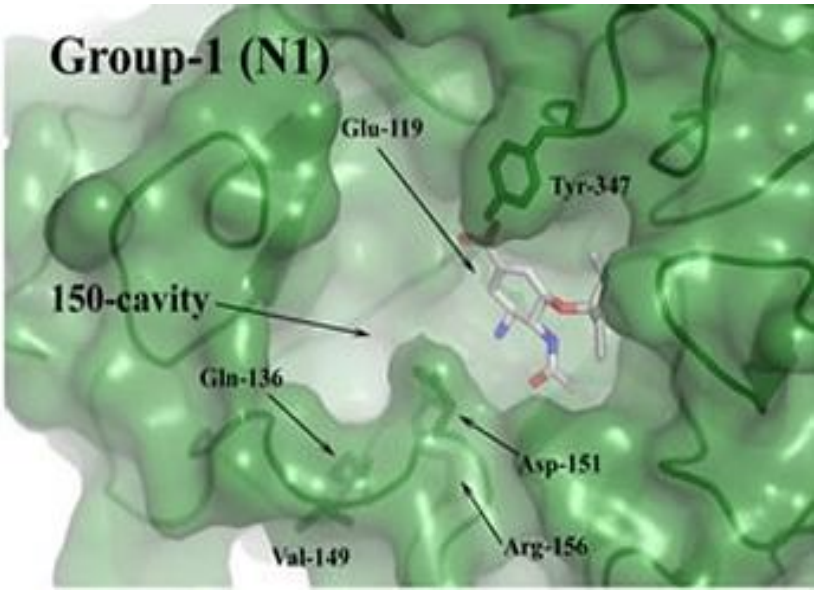


Figure 1.11 The Group 1 NA active site showing 150 cavity adopting an open conformation, adapted from Thompson et al., [69].

1.7 Inhibitor design based on a benzoic acid scaffold

The high potency and oral activity of oseltamivir promoted alternative scaffolds to replace SA. By introducing benzene ring scaffolds, the number of chiral centers could be minimized, potentially simplifying the chemical synthesis [73]. The challenge is to configure substituents on the ring to optimize the interactions with the active site. The residues interacts with SA are conserved in all influenza NAs, but some free SA showed some differences in size and shape between strains and sometimes differ in potency between N1, N2 and B NAs [74] (**Figure 1.12**). In addition, NA inhibitors with a hydrophobic side chain often show differences in potency between the two structural groups of type A NA [75].

Several benzoic acid derivatives have been reported to bind to influenza B NA. The X-ray crystal structures showed that benzoic acid derivatives bind to the active site in the same orientation as SA [76]. The derivatives with common carboxylate substitution at the C2 position, maintains the native interaction with ARG292, ARG118, and ARG371; and the *N*-acetyl substitution at the C5 pocket maintains the interaction with ARG152. The benzene ring scaffold (hydrophobic) fits into the glycerol binding site (C6 pocket) of the SA [77].

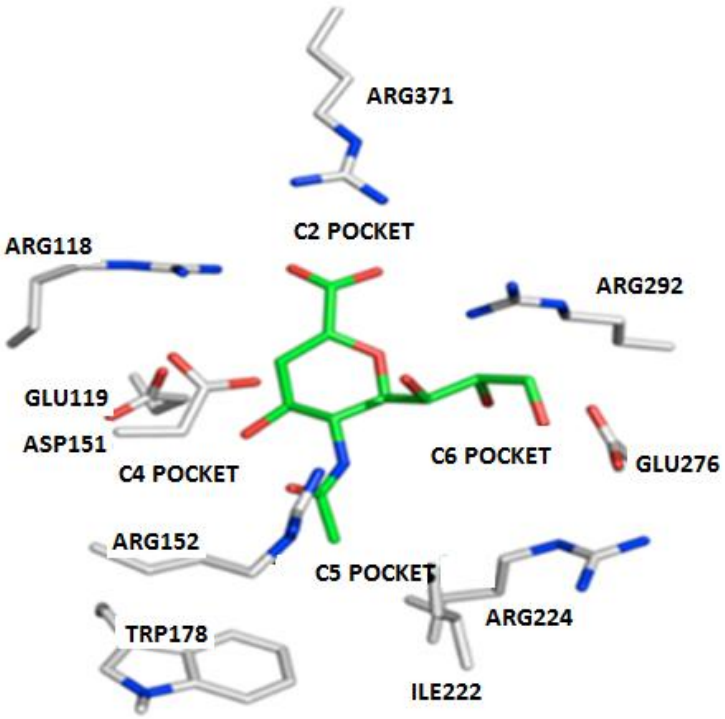


Figure 1.12 The NA pockets that surround the transition state of DANA. In benzoic acid series, pocket 6 is adjacent to C4 of the Benzene ring (PDB ID: 1NNB). Figure made with PDB ID: 1NNB, adapted from Finley et al., [77].

One good example of benzoic acid derivatives which have been studied as NA inhibitor is Inhibitor 1 (**Figure 1.13**).

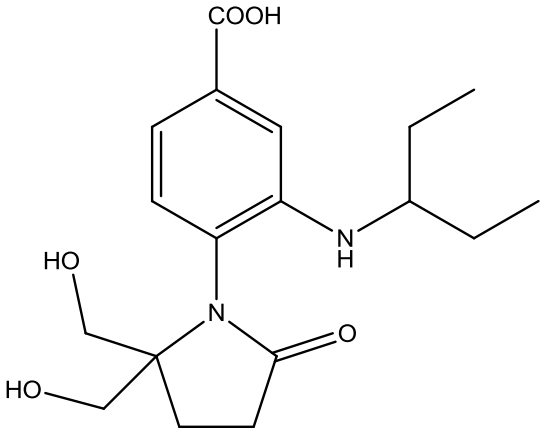


Figure 1.13 Inhibitor 1, adapted from Atigadda et al. [73].

This compound lacks the C4 pocket substituent but showed NA activity as observed with other compounds with a similar hydrophobic substituent on the benzene ring [73]. Inhibitor 1 exhibited moderately potent activity from low to mild (nM) activity against NAs of the N2 and N9 subtypes of influenza A virus, with less NA inhibitory activity against influenza B, as observed for other compounds with a similar hydrophobic substituent on the benzene ring [73].

However, attempts to design substituents on the benzene ring that occupy both the negatively charged C4 pocket and the glycerol binding site (C6 pocket) of SA have been difficult because the two sub sites are off set from the plane of the benzene ring [78]. The results of molecular modeling studies suggested that a different arrangement of the benzene ring may be needed, and one way to alter the arrangement may be to change the size of the hydrophobic substituent, such that an increase or a decrease in steric crowding reorients the benzene ring in the binding site [78].