COMBINATORIAL DESIGN OF VIRTUAL SIALIC ACID ANALOGUES AGAINST INFLUENZA A HEMAGGLUTININ USING STRUCTURE AND FRAGMENT BASED APPROACHES

By

Mohammed Noor Al-deen Mahmoud Al-qattan

Thesis submitted in fulfillment of the requirements for the degree of Master of Science

February 2009

To all who believe in Allah

ACKNOWLEDGEMENTS

In the name of Allah, the Most Beneficent, the Most Merciful,

"و العصر (۞) إن الإنسان لفي خسر (۞) إلا الذين آمنوا وعملوا الصالحات وتواصوا بالحق وتواصوا بالصبر (۞)"

Means “By (the Token of) time (through the Ages), (۞) Verily Man is in loss, (۞) Except such as have Faith, and do righteous deeds, and (join together) in the mutual teaching of Truth, and of Patience and Constancy. (۞)” Holly Quran, Al-Asr (1-3).

All thanks to ALLAH Who help me to finish this project after about one year of misguidance. Allah created the ideas and gave the assistance. And this is the message I want to deliver to all students that come after me and read this thesis: “Dear friend or brother (if you are Muslim), you should think about one important matter before proceeding in life, do you know why you are on this planet?, and for what purpose?. Be sure that getting a scientific degree is not more important than understanding your entire life. After few hours, days, or years certainly you will die. Then there will be no time for thinking but collecting results of what you have done in your present life. Really, try to spend some time for thinking rather than doing harness jobs. If you could understand this critical matter in your life, certainly you will be able to understand the other non critical subsistences. And then, you will recognize that everything in your life even your acts are provided by the One Who has created you so you can easily ask for guidance and support. But, you cannot think properly unless you purify your heart from the darkness of sins, as prophet Mohammad (pbuh) said (Surely there is in the body a small piece of flesh; if it is good, the whole body is good, and if it is corrupted, the whole body is corrupted, and that is surely the heart)”

All thanks and obligedness to Allah who gave me a faithful wife that supports and advices me. She tolerated the living away from her parents for two years in order to accompany me during my study. Great thanks to Allah who gave me a lovely daughter (Zubaidah) and a good son (Musa) here in Malaysia.

Thanks to my father who supported me financially and by counsel, and may Allah's mercy and forgiveness be upon my mother. Thanks to my father's wife who encouraged me to travel for study. Best wishes to my brothers Mohanned, may Allah give him a good wife and to my brother Zakaria, may Allah facilitate the study for him and give him a good wife too. Best regards to my faithful friend, Ali Abdul Hakim who kept in touch during my stay in Malaysia.

Finally, thanks and appreciations to Universiti Sains Malaysia, Centre For Drug Research which supports my project by a research fellowship award. This award helped me to do more and concentrate better on my project. And special thanks to Dr. Mohammad Nizam bin Mordi who gave me friendship beside supervision.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Å	Angstrom
ΔG	Binding energy
ΔG_0	Adjustable parameter for binding energy
ΔG_{ele}	Energy term describes electrostatic interaction
ΔG_{vdW}	Energy term describes van der Waals interaction
ΔG_{hbond}	Energy term describes hydrogen bonding interaction
$\Delta G_{conform}$	Energy term describes the cost of deviation from ideal bond length, angle, and torsion
ΔG_{rot}	Energy term describes the entropic effect of fixing the rotable bonds.
ΔG_{sol}	Energy term describes the desolvation effect
$\sum$	Summation
ΔG_{hb}	Energy coefficient for hydrogen bonding energy term
ΔG_{ionic}	Energy coefficient for ionic interaction energy term
ΔG_{lipo}	Energy coefficient for lipophilic interaction energy term
A_{lipo}	Lipophilic contact surface
$f(\Delta R, \Delta\alpha)$	A scaling function that penalizes the deviations from the ideal geometry
N_{rot}	Number of rotable bonds
U_{ij} (r_{ij})	Energy of interaction between atoms i and j as function of distance separating them (r_{ij})
A_{ij}, B_{ij}	Coefficients for van der Waals interaction energy
r_{ij}	The distance between the ligand atoms, i, and protein atoms, j
$E(t)$	A directional weight based on the angle, t, between the ligand atom and the protein atom.
C_{ij}, D_{ij}	Coefficients for hydrogen binding energy
ε	Depth of the energy well
r_{eqm}	The equilibrium distance between two atoms where the interatomic interaction energy equals to the depth of the energy well.
E_{hbond}	The estimated average energy of hydrogen bonding of water with a polar atom
q_i, q_j	Partial atomic point charges for ligand atoms, i, and protein atoms, j.
$\varepsilon(r_{ij})$	Distance-depended dielectric constant
S_i	Salvation parameter for ligand atoms, i.
V_j	fragmental volume for protein atoms, j.
σ	Gaussian distance constant.
ΔG_p	Internal ligand energy
δ	The distance pairs of equivalent atoms from docked and crystallographic ligand conformations
ψ	no definitive value was available and only the lower limit of K_d or IC50 was considered
$\psi\ \psi$	no inhibition was observed at the highest concentration tested
D_{ij}	The distance between atoms i and j which is less than 1.5 Å

LIST OF ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
A, Ala	Alanine
ACD	Available Chemical Directory
AnA	Anchor atom
AnAs	Anchor atoms
APS	Accumulated penalized score
CADB08	Commercially Available Data Base version 0.8
cLogp	Calculated logarithm of octanol/water partition coefficient
CMC	Comprehensive Medicinal Chemistry
CPK	Corey-Pauling-Koltun
CS	Crude score
CSD	Cambridge Structural Database
D. Asp	Aspartic acid
DLG	Docking Log file
DPF	Docking parameter file
DTP	Developmental Therapeutics Program
EFEB	Estimated free energy of binding
F	Phenylalanine
FDE	Final docked energy
FF	First principle force-field
Fu	Fucose
Gas const.	Gas constant
G, Gly	Glycine
GA	Genetic algorithm
Gal	Galactose
GalNAc	Galactose amine
Glc	Glucose
GlcNAc	Glucose amine
Q, Gln	Glutamine
E, Glu	Glutamic acid
GPF	Grid parameter file
H, His	Histidine
HA	Hemagglutinin
HA0	The precursor of HA1 and HA2
HA1	First subunit of HA monomer
HA2	Second subunit of HA monomer
HBA	Hydrogen-bond acceptor
HBD	Hydrogen-bond donor
HvA	Heavy atom
HvAs	Heavy atoms
I	Isoleucine
IC50	Concentration produces 50% inhibition
InterMIE	Intermolecular interaction energy
IntraMIE	Intramolecular interaction energy
K, Lys	Lysine
kcal	Kilo calorie

K_d	Dissociation constant
K_i	Inhibitory constant
L, Leu	Leucine
LGA	Lamarckian genetic algorithm
LogSW	Logarithm of Intrinsic water solubility
M1	Internal matrix protein
M2	Trans-membranal protein
MC	Monte Carlo
MD	Molecular dynamics
MDL	Molecular design limited
MM/FF	Molecular mechanics/force-field
MW	Molecular weight
MW/10	Molecular weight divided by 10
N, Asp	Asparagine
NA	Neuraminidase
NCI	National Cancer Institute database
NMR	Nuclear magnetic resonance
OFEB	Observed free energy of binding
P	Proline
PDB	Protein data bank
PS	Penalized score
QM/MM	Quantum mechanics/ molecular mechanics
r	Correlation coefficient
r^2	Correlation coefficient
R, Arg	Arginine
RA	Relative affinity
RB	Number of rotable bonds
RMSD	Root mean square of deviation
RMSEP	Root mean square error of prediction
RO5	Rule of five
RTI	Record Type Indicator
S, Ser	Serine
SA	Sialic acid
T, Thr	Threonine
TCL	Tool command language
TDOF	Torsional degree of freedom
Temp	Temperature in Kelvin
tPSA	Topological polar surface area
Y, Tyr	Tyrosine
V	Valine
VDW	van der Waals
W, Trp	Tryptophan
WOMBAT	WOrld of Molecular BioAcTivity

LIST OF APPENDICES

REKABENTUK KOMBINATORIAL BAGI ANALOG MAYA ASID SIALIK TERHADAP HEMAGGLUTININ DARIPADA INFLUENZA A DENGAN MENGGUNAKAN PENDEKATAN STRUKTUR DAN FRAGMEN

Abstrak

Perencatan virus influenza A untuk menghindarkan daripada morbiditi dan mortaliti merupakan perkara utama yang diambil kira semasa epidemik dan sangat penting semasa pandemik. Terdapat dua jenis glikoprotein permukaan yang membentuk permukaan utama penentu antigenik virus influenza A iaitu hemagglutinin (HA) dan neuraminidase (NA). HA bertanggungjawab untuk perlekatan viral pada sel yang dijangkiti melalui pengikatan dengan moieti asid sialik (SA) permukaan. NA pula bertanggungjawab untuk menghidrolisis ikatan glikosidik yang menghubungkan SA dengan membran sel dan menyebabkan pemisahan viral. Kaedah rekabentuk dadah berasaskan struktur telah berjaya digunakan dalam mereka bentuk secara klinikal perencat NA iaitu Zenamivir dan Oseltamivir yang menghalang pemutusan virus progeni. Walau bagaimanapun, tiada perencat dengan berat molekul rendah yang telah direka berkesan untuk bertindak terhadap HA dan seterusnya mencegah pengikatan viral pada sel perumah.

Dalam kajian ini, kaedah pemodelan molekul telah digunakan untuk mereka bentuk pangkalan data analog maya SA dengan penggantian tunggal sama ada pada kedudukan C2, C5 atau C6 rangka SA. Molekul fragmen yang terdapat secara komersial telah digunakan sebagai calon pengganti. Dengan menggunakan pendekatan pendokkan molekul, fragmen molekul telah didok pada kantung ikatan HA di tapak pengikatan kristalografi C2, C5 dan C6 kumpulan berfungsi SA natural dan analognya yang lain. Kemudian, fragmen yang telah diorientasi digabungkan

secara automatik pada rangka SA dengan/atau tanpa bantuan molekul penghubung menggunakan algoritma empirikal yang dibangunkan secara dalaman. Oleh itu, tiga pangkalan data analog SA dengan penggantian fragmen tunggal pada C2, C5 dan C6 telah berjaya dihasilkan. Ketiga-tiga pangkalan data kemudiannya didokkan pada keseluruhan tapak ikatan SA di kantung pengikat HA menggunakan kaedah pendokkan yang disahkan untuk menentukan ketepatan konformasi dan afiniti ikatan. Keputusan pendokkan menunjukkan afiniti analog yang dihasilkan adalah lebih tinggi (mencecah 30,000 kali ganda) daripada SA natural. Tenaga ikatan yang lebih baik menunjukkan tenaga ikatan daripada fragmen dan rangka SA kristal boleh digabungkan ke dalam analog yang dihasilkan.

Dengan menggunakan terbitan C5 dan C6 analog SA yang menunjukkan afiniti tinggi dan penyimpangan dari kedudukan rangka kristal SA yang kecil, satu pangkalan data analog kombinatorial SA telah dihasilkan dengan mengesktrak kumpulan penukarganti C5 dan C6 dan kemudiannya menggabungkan mereka secara sistematik ke dalam rangka tunggal molekul SA. Peraturan Lima Lipinski telah diaplikasikan untuk membentuk hanya analog dengan pembolehdapatan oral. Keputusan pendokkan menunjukkan afiniti analog kombinatorial terhadap kantung HA adalah lebih tinggi berbanding analog penggantian tunggal dan affinitinya melebihi 100,000 kali ganda daripada SA natural memandangkan kebanyakkan analog SA yang direka bentuk boleh mengikat SA pada tapak ikatan HA dengan afiniti yang lebih tinggi berbanding SA asal, mereka mempunyai potensi untuk merencat virus influenza A daripada terikat kepada membran sel perumah dan seterusnya bertindak sebagai agen anti-flu.

COMBINATORIAL DESIGN OF VIRTUAL SIALIC ACID ANALOGUES AGAINST INFLUENZA A HEMAGGLUTININ USING STRUCTURE AND FRAGMENT BASED APPROACHES

Abstract

Inhibition of influenza A virus to avoid morbidity and mortality is of main concern during epidemics and of major concern during pandemics. Two types of surface glycoprotein form the main surface antigenic determinants of influenza A virus i.e. hemagglutinin (HA) and neuraminidase (NA). HA is responsible for viral attachment to the infected cell through surface-bound sialic acid (SA) moieties, while NA is responsible for hydrolysing the glycosidic bond that connects SA with the cell membrane resulting in viral detachment. Structure-based drug design approach has been successfully used in designing the clinically available NA inhibitors Zenamivir and Oseltamivir which restrict the progeny virus detachment. However, there is no effective low molecular weight inhibitor that has been developed to target HA and prevent the initial viral attachment to the host cell.

In this study molecular modeling techniques were used to design databases of virtual SA analogues by a single substitution at either of C2, C5 or C6 positions of SA scaffold. A commercially available molecular fragment was used for the substitution candidate. By using molecular docking approach, the molecular fragments were docked against the HA binding pocket at the crystallographic binding sites of C2-, C5- and C6-natural functional groups of SA and its analogues. Then, the oriented fragments were connected automatically to the SA scaffold with or without the incorporation of molecular linkers using in-house developed empirical algorithms. Thus, three databases of SA analogues with single substituted fragments

at positions C2, C5 or C6 were successfully generated. The three databases were then docked against the whole SA binding site using a validated docking tool to estimate the accurate binding conformations and affinities. Our docking results showed that the affinities of the generated analogues were higher (up to 30,000 fold) than the natural SA. The improvement in binding energies indicates that the favourable binding energies of the oriented fragments and the crystal SA scaffolds were additively merged within the generated analogues.

Using the C5-derived and C6-derived SA analogues that showed higher affinities with little deviations from the crystal SA scaffold's position, a database of combinatorial SA analogues was generated by extracting the C5- and C6-designed substitutions and combining them systematically on a single SA scaffold molecule. The Lipinski's rule of five was applied to construct only the oral bioavailable analogues. The docking results showed that the affinities of combinatorial analogues were higher than the analogues of single substitution and exceed 100,000 fold the affinity of natural SA. As many of the designed SA analogues could bind the SA binding site of HA with higher affinity than the natural SA, they have the potential to inhibit influenza A virus from attachment to host cell membrane and consequently act as anti-flu agents.

CHAPTER ONE

INTRODUCTION

1.1 Problem statement

Influenza A virus is an enveloped negative strand RNA virus belongs to the *Orthomyxoviridae* family and responsible for the annual influenza epidemics and recurrent pandemics. There are many subtypes of influenza A classified by antigenicity of their corresponding surface glycoproteins i.e. hemagglutinin (HA) and neuraminidase (NA). Currently, 16 serotypes of HA and 9 serotypes of NA are available (Fouchier *et al.*, 2005; Baker *et al.*, 1987). HA glycoprotein is responsible for sticking the virus to the host cell before being engulfed by endocytosis and this attachment is mediated by surface-bound sialic acid (SA) moieties of the cell membrane (Skehel & Wiley, 2000), while NA is responsible for releasing the progeny viruses from the infected cell by hydrolysing O-glycosidic bond between the terminal SA which is bound to HA and the penultimate sugar moiety that connect SA to host cell membrane (Air & Laver, 1989). As functions of HA and NA oppose each other, a balanced effect is required for effective viral infection (Wagner *et al.*, 2002). HA and NA are highly vulnerable to mutagenic changes by shift and/or drift in response to the pressure of host's immune system (Lewis, 2006), that caused vaccination against influenza A virus is ineffective and pandemics are recurrent (Kilbourne, 1975).

SA (or 5-amino-3,5-dideoxy D-glycero-D-galacto nonurosonic acid) is a member of the natural SAs family and is the natural ligand for both of HA and NA (Varki & Varki, 2007; Schauer & Kamerling, 1997; Furuhata, 2004). SA is connected to the penultimate galactose moiety of the host cell membrane by two different modes. In the first mode, SA C2 is connected by α-O-glycosidic linkage to C3 of galactose (α(2,3)). This linkage predominates in the avian intestine. While in the second mode, the connection of SA to C6 of galactose gives the α(2,6) linkage which predominates in human respiratory tract. The binding between SA and HA is a simple bimolecular association. SA binds to the conserved amino acids mainly by bristling hydrogen bonds. No chemical reactions took place and no apparent conformational changes occurred at the binding site upon binding SA.

Pyranose ring forms the scaffold of SA molecule, to which different functional groups are connected through carbon atoms C2, C4, C5, and C6. Changes in SA functional groups may confer changes in affinity toward HA as well. Several studies have been conducted to monitor the effect of modifying natural SA functional groups on the affinity toward HA to produce monovalent inhibitors, or incorporate several molecules of SA analogues of low affinity on large molecular weight carrier to produce polyvalent inhibitors (Matrosovich & Klenk, 2000).

The HA of influenza A H3N2 (X-31) virus is of H3 serotype which is preferentially binds human type SA receptor (SA-α2,6-galactose) (Rogers & Paulson, 1983). There is a theory of periodical recirculation of H3N2 in human population (Masurel & Marine, 1973). H3 bound to various SA analogues was exclusively studied by X-ray crystallography (Weis *et al.*, 1988; Sauter *et al.*, 1992a; Ha *et al.*,

2003), NMR (Sauter *et al.*, 1989, Sauter *et al.*, 1992a; Machytka *et al.*, 1993), alongside with inhibitory assays (Pritchett, 1987; Pritchett *et al.*, 1987; Kelm *et al.*, 1992; Toogood *et al.*, 1991). However, no effective monovalent agents have been developed yet which target the HA and inhibit viral attachment, even though HA has been long identified as a probable target for inhibitor design (Pritchett *et al.*, 1987; Weis *et al.*, 1988; Sauter *et al.*, 1989; Sauter *et al.*, 1992a; Machytka *et al.*, 1993; Lentz, 1990).

Molecular modelling techniques are developing fast, with some branches are matured enough and effectively participated in designing and screening drug candidates. With respect to HA, AutoDock3.05 has been used to reproduce crystal conformation of SA within HA primary binding site with estimated binding energy close to the observed value (Morris *et al.*, 1998). The scope of this research is to design virtual databases of SA analogues of higher affinities toward the HA of influenza A H3N2 by substituting the natural SA functional groups with commercially available molecular fragments.

1.2 Influenza virus

Influenza viruses belong to the *Orthomyxoviridae* family of RNA viruses which include influenza A, influenza B, influenza C, Thogoto (sometimes referred as influenza D), and Isa viruses (Cox *et al.*, 2000). Influenza A can infect human, mammals, and birds (Webster *et al.*, 1992)., while influenza B infects human and seals (Osterhaus *et al.*, 2000) and influenza C infects human and pigs (Yuanji *et al.*, 1983).

Influenza A is a major cause of morbidity and mortality in humans during annual epidemics and the recurrent pandemics. Influenza A mainly infects the epithelium of the upper and lower respiratory tracts and typically results in an abrupt onset of illness that usually includes high fever, coryza, cough, headache, prostration myalgia besides upper respiratory tract congestion and inflammation. These symptoms persist for 7 to 10 days while weakness and fatigue may extend for weeks. Pneumonia is a frequent manifestation of more severe infection. Influenza A infection is also a relatively common aetiology of laryngotracheitis (croup) in children and bronchiolitis. Myocarditis, encephalitis, and other extra-respiratory tract diseases are rarely occurred during the course of influenza infection. Viral infection is also associated with an increased incidence of subsequent otitis media, and influenza A pneumonia which may be complicated by subsequent infection with bacterial pathogens such as *Staphylococcus aureus*. Finally, influenza A infection is an important trigger of reactive airway disease in those with pre-existing asthma, and it may also promote allergic sensitization to environmental proteins (Lewis, 2006; Taubenberger & Morens, 2008).

Influenza A virus is an enveloped negative single-stranded RNA virus. It is a spherical to rod particle of 120 nm in diameter. Eight RNA segments are present inside the viral capsid and encode several proteins which include; HA and NA glycoproteins which are expressed on the viral surface, M2 protein which forms an ion channel that cross the viral lipid bilayer, and M1 protein which forms the internal matrix of the virus and used to encapsulate the genetic material while budding from the host cell membrane in the production of progeny viruses (Figure 1.1) (Bourmakina & García-Sastre, 2005; Lewis, 2006).

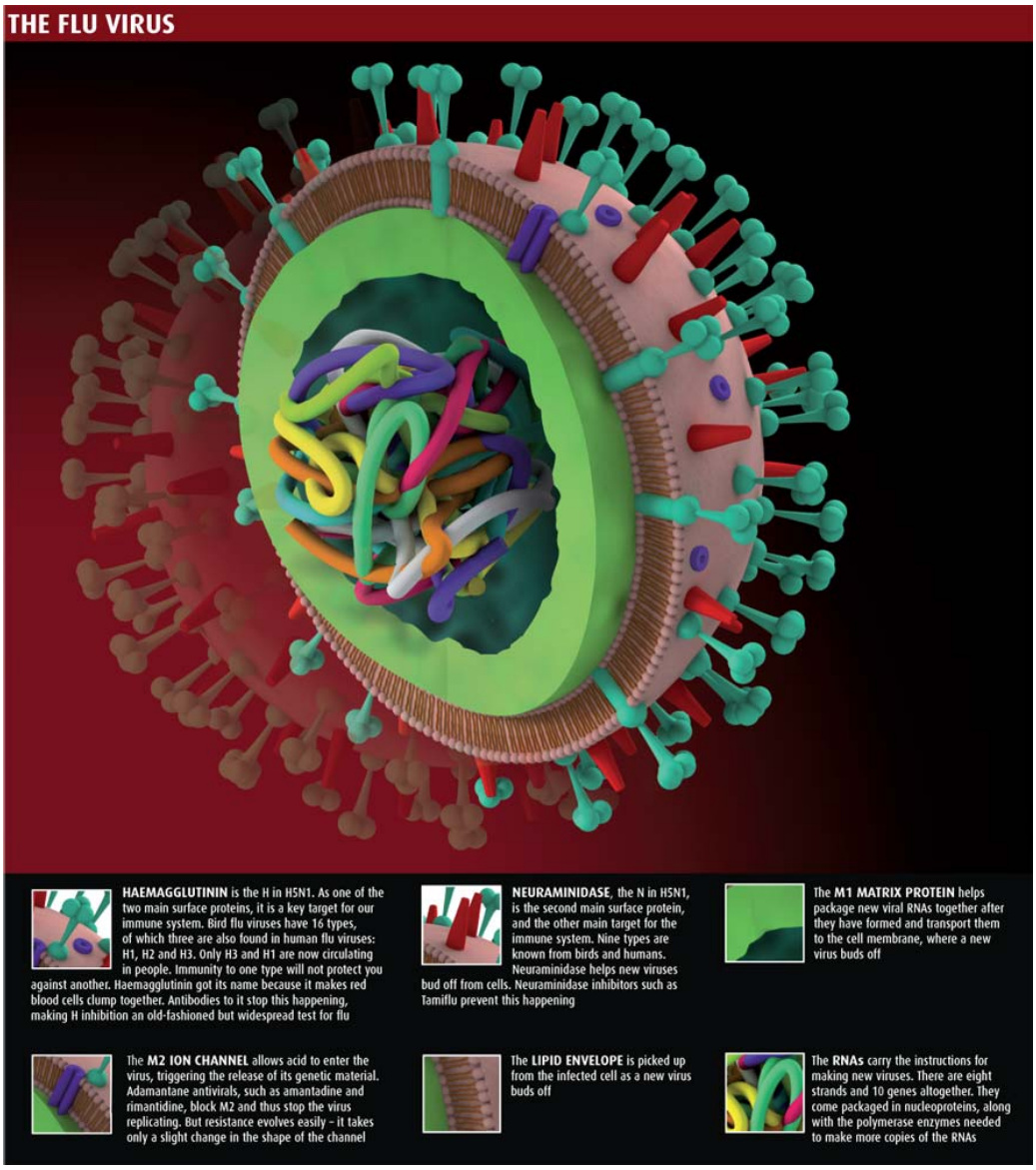


Figure 1.1: Influenza A virus with the external glycoproteins (HA and NA), transmembranal proteins (M2), internal protein matrix (M1) and RNA segments.

1.2.1 Hemagglutinin (HA)

HA is a major antigenic determinant of influenza A virus, it is kind of lectins which are sugar-binding proteins (Lis & Sharon, 1998). Some protozoa, bacteria and viruses use SA-recognizing lectins to attach themselves to the cells of the host organism to initiate infection (Table 1.1). The influenza virus initiates infection by attachment to the host cell membrane followed by endocytosis and fusion with endosomal membranes. This attachment is mediated by interaction of terminal cell-surface SA with viral surface glycoproteins (HA in influenza type A and B or hemagglutininesterase in influenza type C) (Paulson, 1985; Herrler *et al.*, 1995).

Each influenza virus contains about 500-1000 HA homotrimer (Ruigrok, *et al.*, 1984 cited in Glick *et al.*, 1991). HA monomer is synthesized as a single polypeptide (HA0) that is cleaved by host protease into two subunits (HA1 and HA2), in which HA1 is more variable antigenically compared to the HA2. Changes in this glycoprotein are responsible for uncontrolled recurrence of influenza epidemics (Webster & Laver 1975). Each HA monomer has two SA binding sites. The primary SA binding site is located at HA1 and responsible for viral sticking to the host cell, while the secondary binding site is located at the interface between HA1 and HA2 (Figure 1.2) (Weis *et al.*, 1988; Sauter *et al.*, 1992a; Sauter *et al.*, 1992b). The two binding sites are formed of well conserved amino acid residues through all subtypes of influenza A and strains of H3 serotype (Table 1.2) (Ward & Dopheide, 1981; Nobusawa *et al.*, 1991).

During influenza A infection cycle, HA is first attached to the terminal SA residues spreaded on host cell surface. Subsequently, the virus is engulfed by the cell to form an endosome. The acidic environment of the endosomal compartment drives the necessary HA conformational changes to fuse the viral and the endosomal membranes which results in the intracellular release of the virion content (Skehel & Wiley, 2000) (Figure 1.3).

Table 1.1: SA recognizing lectins in protozoa, bacteria and viruses (Varki *et al.*, 2008).

Protozoa Parasite lectins: Merozoite erythrocyte-binding antigens (EBAs) (*Plasmodium falciparum*)
Bacteria Bacterial adhesins: S-adhesin (*Escherichia coli* K99), SabA and SabB (*Helicobacter pylori*) Bacterial toxins: Cholera toxin (*Vibrio cholerae*), tetanus toxin (*Clostridium tetani*), botulinum toxin (*Clostridium botulinum*), pertussis toxin (*Bordetella pertussis*) Mycoplasma lectins: *Mycoplasma pneumoniae* hemagglutinin
Viruses Hemagglutinins: Influenza A and B viruses, primate polyomaviruses, rotaviruses Hemagglutinin neuraminidases: Newcastle disease virus, Sendai virus, fowl plague virus Hemagglutinin esterases: Influenza C viruses, human and bovine coronaviruses

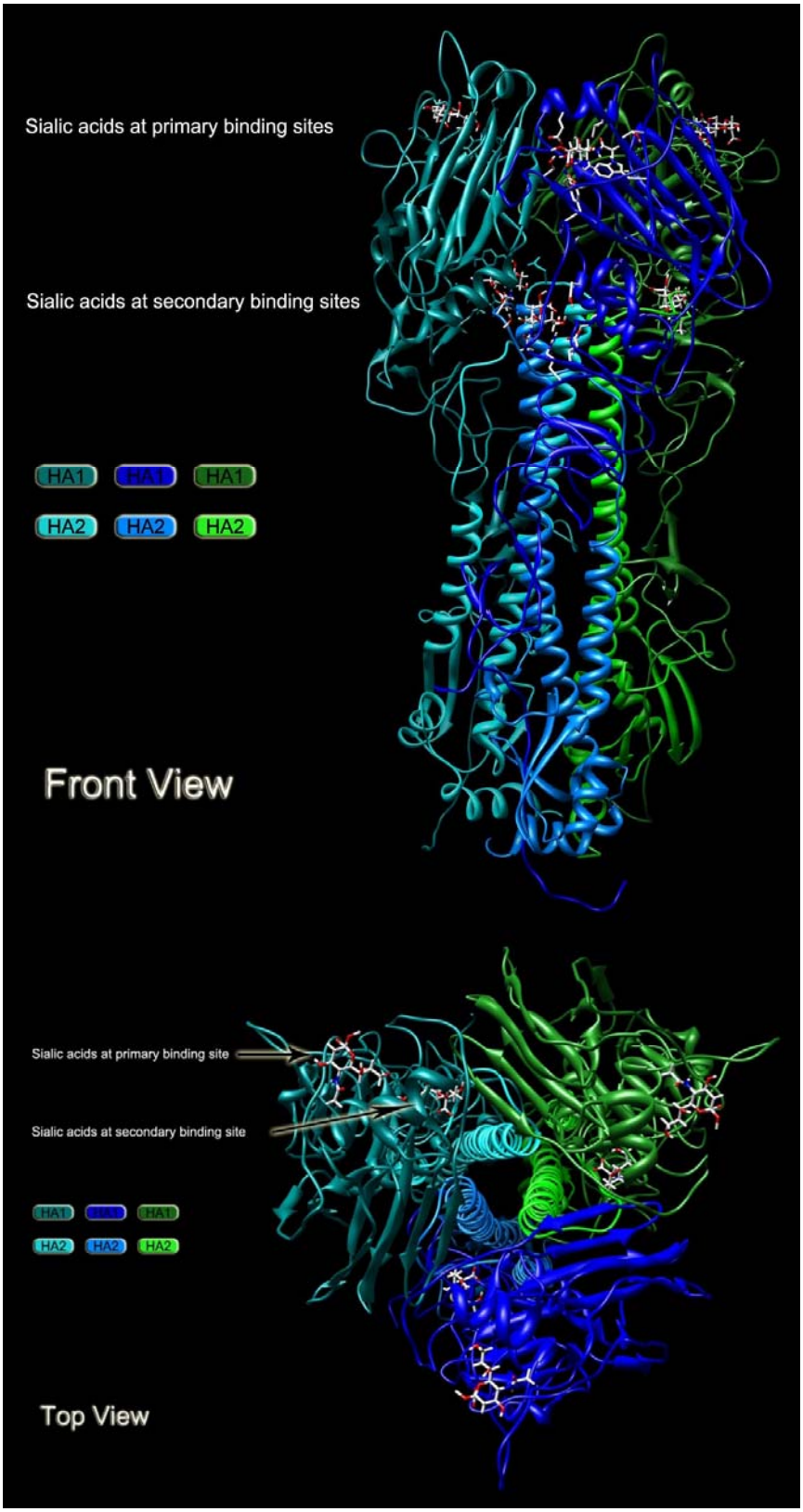


Figure 1.2: Influenza A HA homotrimer with the primary and secondary SA binding sites.

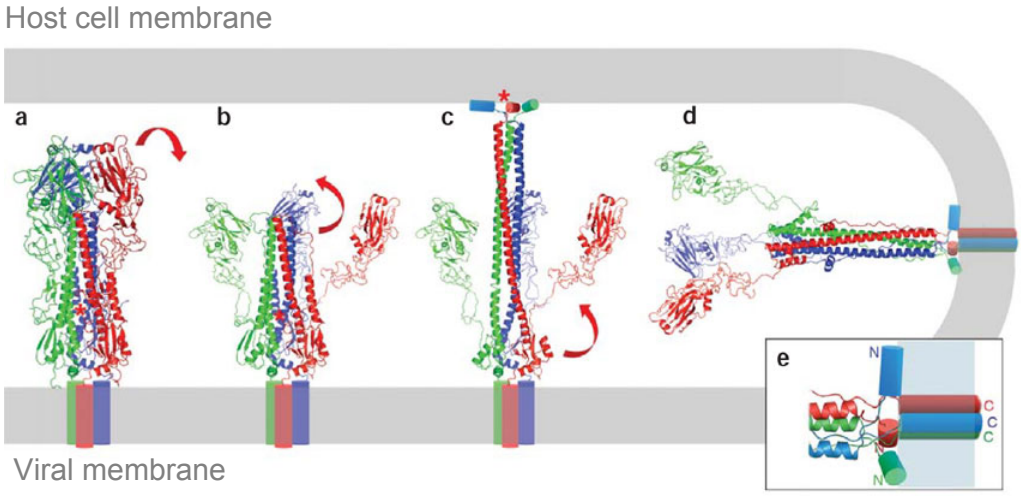


Figure 1.3: The process of fusion between viral and endosomal membranes mediated by viral HA. a) Influenza A HA exposed on the viral surface (bottom) and pointed toward the host cell membrane (top). b) HA1 subunits displaced aside from locations over HA2, c) The loops between shorter and longer helices within each HA2 subunits are extended. Red asterisk represents the exposed fusion peptides. d) Collapse of the extended intermediate loops to generate the post-fusion conformation. e) Magnified fusion point showing the N and C terminal for each of the three HA2 subunits (Harrison, 2008).

Table 1.2: The amino acids of the primary and secondary SA binding sites.

Serotype	Strain		Primary SA binding site														Secondary SA binding site						
			HA1																		HA2		
		98	134	135	136	137	153	155	183	186	190	194	219	226	227	228	89	105	106	269	69	71	72
H1		Y	G	V	T	A	W	T	H	S	D	I	E	Q	A	G	E	Y	E	I	E	N	K
H2		Y	G	G	S	R	W	T	H	I	E	L	T	Q	G	G	E	Y	E	K	E	G	N
H3	HK/1/68 (X-31)	Y	G	G	S	N	W	T	H	S	E	L	S	L	S	S	E	Y	A	R	E	S	E
	NT/60/68	Y	G	G	S	N	W	T	H	S	E	L	S	L	S	S	E	Y	A	R	E	S	E
	Aichi/2/68 (X-31)	Y	G	G	S	N	W	T	H	S	E	L	S	L	S	S	E	Y	A	R	E	S	E
	Mem/1/71	Y	G	G	S	N	W	T	H	S	E	L	S	L	S	S	-	Y	A	R	E	S	E
	Mem/102/72	Y	G	G	S	N	W	Y	H	S	E	L	S	L	S	S	E	Y	A	R	E	S	E
	Vic/3/75	Y	G	G	S	S	W	Y	H	S	E	L	S	L	S	S	E	Y	A	R	E	S	E
H4		Y	G	K	S	G	W	V	H	S	E	L	S	Q	S	G	E	Y	Q	N	E	E	Q
H5		Y	G	V	S	S	W	I	H	N	E	L	T	Q	S	G	E	Y	E	R	E	N	N
H6		Y	G	V	T	R	W	I	H	P	E	L	A	Q	R	G	E	V	E	F	E	S	N
H7		Y	G	T	T	S	W	L	H	G	E	L	T	Q	S	G	E	E	E	S	E	T	N
H8		Y	G	T	S	K	W	T	H	P	E	L	P	Q	Q	G	E	L	E	Q	E	S	E
H9		Y	G	T	S	R	W	T	H	P	E	L	P	Q	Q	G	E	L	E	K	E	N	E
H10		Y	G	T	T	K	W	V	H	S	E	L	A	Q	S	G	E	E	E	S	E	S	E
H11		Y	G	V	T	A	W	I	H	A	E	L	T	Q	A	G	E	E	E	R	E	S	E
H12		Y	G	T	S	K	W	T	H	P	E	L	P	Q	Q	G	E	Q	E	K	E	S	E
H13		Y	G	T	T	S	W	I	H	V	E	L	V	Q	R	S	E	N	G	Q	E	N	Q

1.2.2 Neuraminidase (NA)

NA is the second major glycoprotein distributed on the influenza A viral surface. NA is arranged in tetramers and there are about 100-200 copies in each virus (Laver, 1973 cited in Glick *et al.*, 1991; Schulze, 1973 cited in Glick *et al.*, 1991). Each tetramer is composed of trans-membranal part, thin stalk, and globular head which has the ability to hydrolyse the O-glycosidic linkage that connects SA with the penultimate sugar moiety of the host cell membrane (Seto & Rott, 1966). During the course of influenza infection, NA could also function as scaventure to destroy the epithelial cells to facilitate the viral infection (Air & Laver, 1989). At the end of viral replication cycle, NA facilitates the release of the progeny viruses and prevents HA-mediated viral aggregation (Palese *et al.*, 1974). As the function of HA and NA oppose each other, a balanced effect is required for effective viral infection (Wagner *et al.*, 2002).

In each virus, NA exhibits specificity for terminal SA linkages similar to its relevant HA. Thus NA derived from avian influenza A virus can hydrolyze the α(2,3) glycosidic linkage between SA and penultimate galactose molecule, while NA derived from human influenza A virus can hydrolyse both of α(2,3) and α(2,6) glycosidic linkages (see Section 1.4.3).

1.2.3 Influenza A virus subtypes

There are many subtypes of Influenza A virus which can be classified according to the antigenicity of their corresponding HA and NA for example H1N1, H2N2, H3N2, and H5N1. Up to now there are about 16 serotypes of HA (Ron *et al.*, 2005) and 9 serotypes of NA (Baker *et al.*, 1987). These numbers are vulnerable to increase as a consequence of antigenic shift and drift (Lewis, 2006). All the available HA serotypes besides their host ranges are listed in Table 1.3.

Table 1.3: The available HA serotypes and their host ranges.

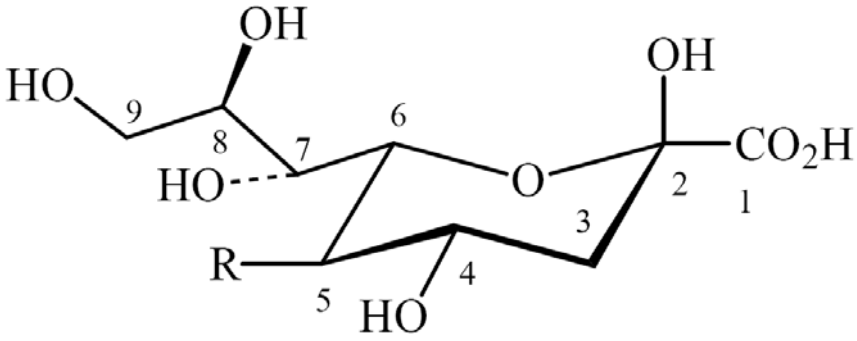
HA serotype	Host range			
	Avian	Human	Equine	Swine
H1	X	X		X
H2	X	X		X
H3	X	X	X	X
H4	X			
H5	X			
H6	X			
H7	X		X	
H8	X			
H9	X			X
H10	X			
H11	X			
H12	X			
H13	X			
H14	X			
H15	X			
H16	X			

1.3 Sialic Acids (SAs)

SAs is a name given to a group of more than 50 different analogues of the parent compound neuraminic acid (Neu) (5-amino-3,5-dideoxy-D-glycero-D-galacto-non-2-ulopyranosonic acid) (Schauer, 2004). They are electronegatively charged acidic monosaccharides which participate in the structural diversity of complex carbohydrates that constitute the major part of proteins, cell membranes lipids and secreted macromolecules (Varki *et al.*, 2008).

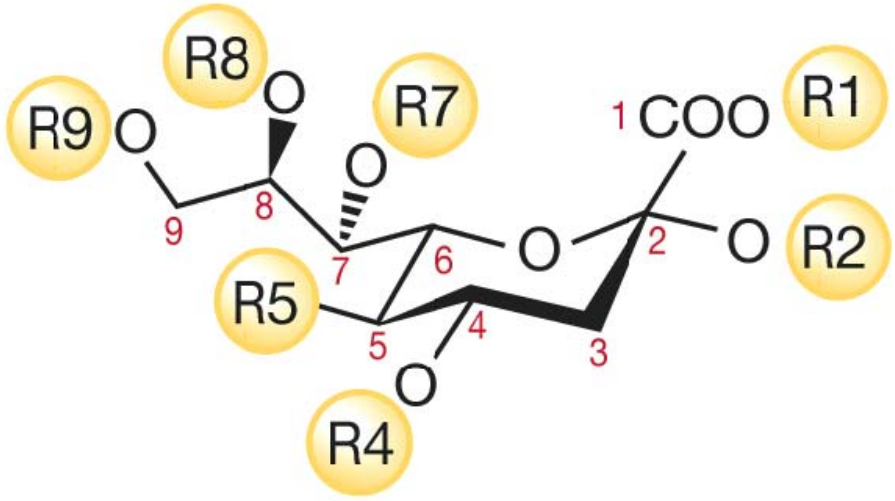
1.3.1 Chemistry of SAs

SAs are naturally occurring deoxy nononic acids of acetylated, sulphated, methylated, and lactylated derivatives comprising a large diverse family of compounds. Such diversity characterizes SA among other sugars (Angata & Varki, 2002). Certain rules are followed in the nomenclature of SA derivatives (Blix *et al.*, 1957; Reuter & Schauer, 1988). Neu5Ac, Neu5Gc, KDN, and Neu are the four main SA molecules (Figure 1.4) from which other analogues are derived by carrying one or more additional substitutions at the hydroxyl groups on C-4, C-7, C-8, and C-9 (Figure 1.5). For more information please see Schauer (1982) or Schauer and Kamerling (1997).



Neu5Ac: R=NHAc
Neu5Gc: R=$NHCOCH_2OH$
KDN: R=OH
Neu: R=H

Figure 1.4: Different substitutions on C5 give the four main SAs molecules.



R1 = H (on dissociation at physiological pH, gives the negative charge of SA); can form lactones with hydroxyl groups on the same molecule or on other glycans; can form lactams with a free amino group at C-5; or tauryl group.
R2 = H; alpha linkage to Gal(3/4/6), GalNAc(6), GlcNAc(4/6), SA (8/9), or 5-O-Neu5Gc; oxygen linked to C-7 in 2,7-anhydro molecule; anomeric hydroxyl eliminated in Neu2en5Ac (double bond to C-3).
R4 = -H; -acetyl; anhydro to C-8; Fu; Gal.
R5 = Amino; N-acetyl; N-glycolyl; hydroxyl; N-acetimidoyl; N-glycolyl-O-acetyl; N-glycolyl-O-methyl; N-glycolyl-O-2-Neu5Gc.
R7 = -H; -acetyl; anhydro to C-2; substituted by amino and N-acetyl in Leg.
R8 = -H, -acetyl, anhydro to C-4, -methyl, -sulfate, SA, Glc.
R9 = -H, -acetyl, -lactyl, -phosphate, -sulphate, SA, OH substituted by H in Leg.

Figure 1.5: Diversity in the SAs. The nine-carbon backbone common to all known SA is shown. The possible variations at the carbon positions are indicated. Glc stands for Glucose, Gal; Galactose, GlcNAc; Glucose amine, GalNAc; Galactose amine, Fu; Fucose, and Leg for legionaminic acid (Varki *et al.*, 2008).

1.3.2 Biological roles of SAs

The cell-surface of both eukaryotic and prokaryotic organisms contains glycocongugates which aid in cellular communications and adhesions. SAs have dual roles in the human body by masking the recognition sites against autoimmune response and in being the binding sites for various pathogens (Kelm & Schauer, 1997).

1. SAs positioning at the outer surface of the cell membrane shields the cell from infective organisms and autoimmune attack. Removing SAs regarded as mechanism for infectivity of various pathogens such as *Vibrio cholerae* (Taylor, 1996). In addition, removing surface SAs predispose the cell to the humoral immunity, a mechanism used for removing thrombocytes from the circulation (Kluge *et al.*, 1992). Therefore, the malignant cells are protected from being attacked by the immune system by over expressing surface SA (Schauer, 2004).
2. SAs on host cells act as anchors for adhesion and subsequent infection by various pathogens. *Plasmodium falciparum* use SAs on the erythrocyte surface as receptors for invasion (DeLuca *et al.*, 1996). Accordingly, studies have been conducted to examine the ability of SAs in blocking the binding site on merozoits which could suppress the infectivity of malaria (Vanderberg *et al.*, 1985). SAs associated with the host cell contribute to invasion by *Trypanosoma cruzi* (Schenkman *et al.*, 1993). Influenza virus A and B are attached to the host cell surface through SA moieties using HA glycoprotein while the *paramyxoviruses* (parainfluenza viruses) have HA-NA

glycoproteins system that enables the virus to attach the host cell through SA followed by membranes fusion by fusion glycoprotein (Colman *et al.*, 1993).

Due to their electronegativity, SAs also participate in transformation of positively charged pharmaceuticals and in mutual repulsion between erythrocytes in the blood stream (Kelm & Schauer, 1997; Schauer & Kamerling, 1997).

1.3.3 SA is the natural ligand of influenza A HA

The C5-N-acetyl substituted neuraminic acid (Neu5Ac) is the natural ligand for HA in all subtypes of influenza A. The molecular structure of SA is composed of central pyranose ring from which different functional groups are protruded. The functional groups include C2-axial carboxylate, C2-equatorial hydroxyl, C4-equatorial hydroxyl, C5-equatorial N-acetyl (acetamido) group, and C6-equatorial glycerol (Figure 1.6). Methyl-α-Neu5Ac is the simplest SA analogue that has been studied crystallographically in complex with HA.

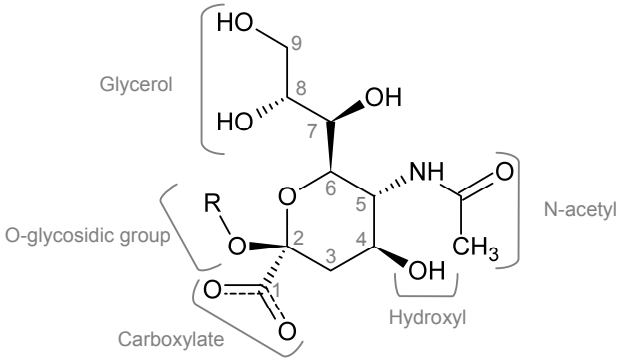


Figure 1.6: The molecular structures of natural SA (R=H) and methyl-α-Neu5Ac ($R=CH_3$).

1.4 HA-SA molecular interaction

1.4.1 The binding site of SA at HA1 (primary SA binding site).

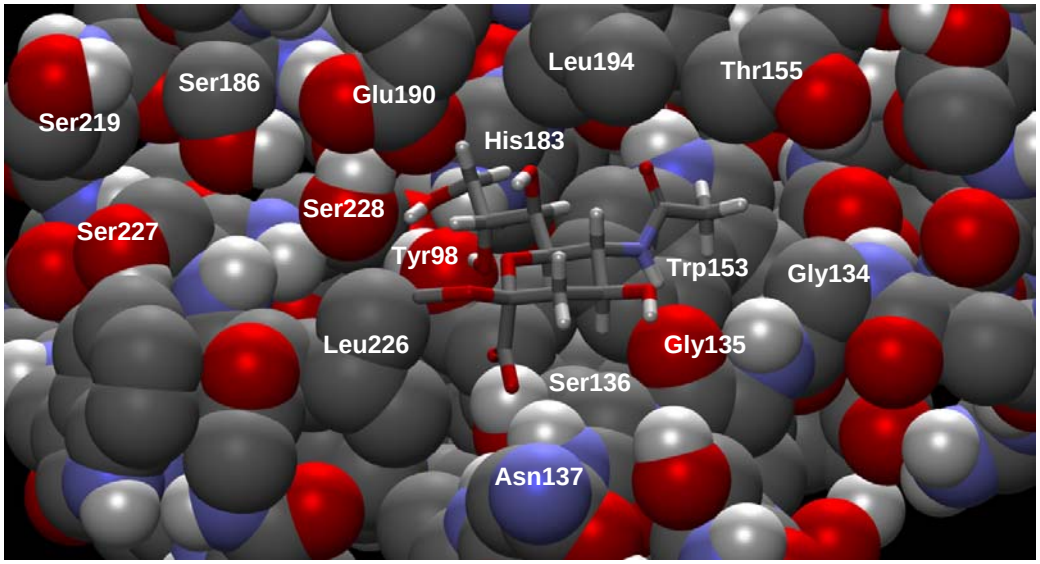
The primary SA binding sites are located at HA1 subunits of HA and formed from well conserved amino acid residues. The phenolic hydroxyl of Tyr98 and the aromatic ring of Trp153 form the bottom of the binding site. There are three polypeptide loops forming the boundaries of this binding site; i.e. loop 130 (includes Gly135, Ser136, Asn137, and Ala138), loop 220 (includes Arg224, Gly225, Leu226, Ser227, and Ser228), and α-helix 190 which form the rear of the site from which the side chains of Glu190 and Leu194 are projected down toward the binding site (Figure 1.7)

1.4.2 The interaction between SA and HA1 binding site.

The interactions between SA and HA1 binding site follow simple bimolecular interaction (Sauter *et al.*, 1989). At the binding site, one face of the SA's pyranose ring faces the bottom of the site while the other face is exposed to the solution. The axial carboxylate, acetamido nitrogen, and two of glycerol hydroxyls are interacted by hydrogen bonds with conserved amino acid residues (Figure 1.7).

C2-carboxylate forms the most stable and important interactions with HA1 binding site where one of the carboyxlate oxygens accepts hydrogen bond from the side-chain of Ser136 while the other oxygen accepts hydrogen bond from Asn137 main chain amide. The C4-hydroxyl group projects outside the binding site, and

appears not to participate in binding due to its equatorial epimerization. With respect to C5-acetamido nitrogen, a hydrogen bond is donated to the main chain carbonyl of Gly135, while the terminal methyl group is in van der Waals contact with six-membered ring of Trp153. The C7-hydroxyl group and the C5-acetamido carbonyl form intra-molecular hydrogen bond and both are in van der Waals contacts with Leu194. The C8-hydroxyl group forms hydrogen bond with the side chain of Tyr98 while C9-hydroxyl group establishes hydrogen bonds with the side chains of Tyr98, His183, Glu190, and Ser228 (Weis *et al.*, 1988; Sauter *et al.*, 1989; Sauter *et al.*, 1992a).



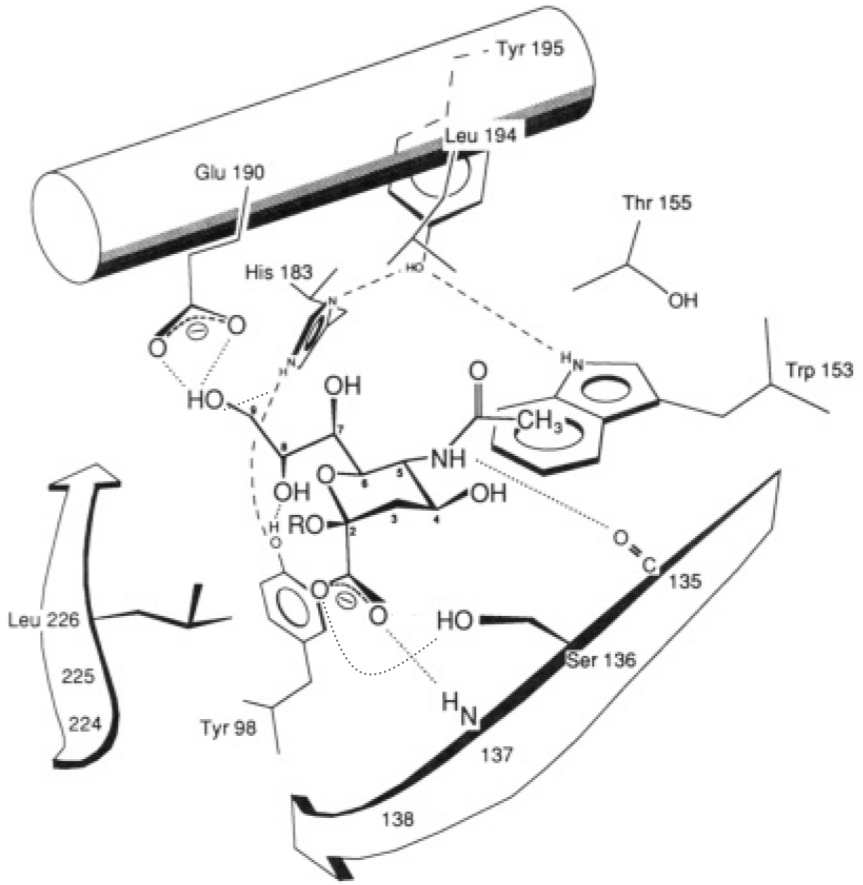


Figure 1.7: Natural SA binding site at HA1 of influenza A virus H3N2 (X-31). Dotted lines indicate hydrogen bonds between SA and HA, while dashed lines show potential hydrogen bonds within the protein (Sauter *et al.*, 1989).

1.4.3 Specificity of the interaction between HA and SA.

1.4.3.1 Specificity of HA1 toward SA follows the type of SA-O-glycosidic linkage.

The type of glycosidic linkage that connects SA moiety to the penultimate galactose residue at terminal cell-membranal carbohydrates determines the affinity whether toward human or avian influenza HA, and vice versa (Rogers & Paulson, 1983). Accordingly, Neu5Ac-α(2,3)-Gal is recognized by avian viruses while Neu5Ac-α(2,6)-Gal is recognized by human viruses (Figure 1.8).

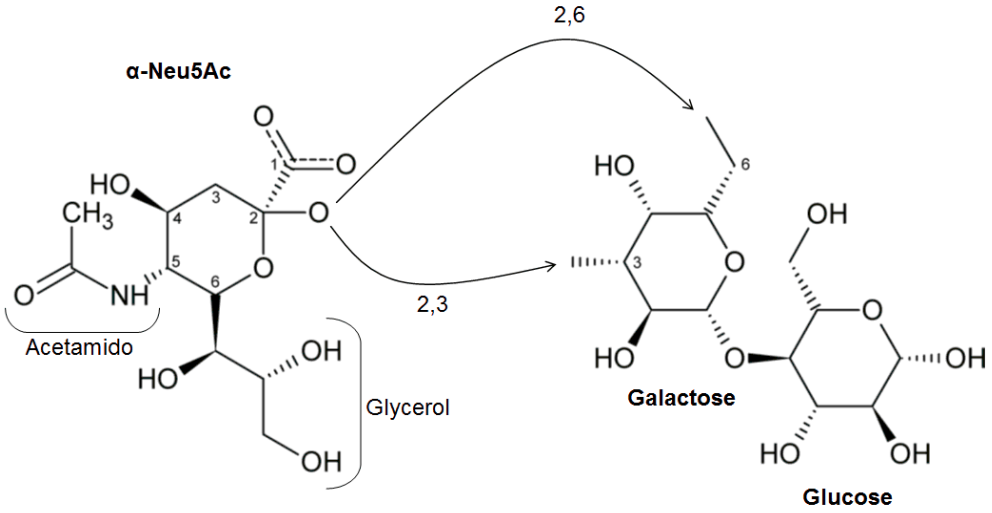


Figure 1.8: Different HAs have different specificities in recognizing these linkages. The HA derived from wild type A/Hong Kong/68, X-31 H3N2 influenza A virus with Leu226 (Human virus) has high affinity to Neu5Ac-α(2,6)-Gal while the Leu226Gln mutant (Avian virus) favours Neu5Ac-α(2,3)-Gal linkage.

The HA specificity governs the viral host range. The humans viruses recognize α(2,6) linkage, and those from avians and equines recognize α(2,3) linkages, while swine viruses recognize both types of linkages (Rogers & D'Souza, 1989; Connor *et al.*, 1994; Matrosovich *et al.*, 1997; Gambaryan *et al.*, 1997; Ito & Kawaoka, 2000). Accordingly, the viral inter-species transferences are limited. Humans pandemics occur when the HA specificity of virus from other species change specificity from α(2,3) to α(2,6) for which humans have no immunity. The worldwide pandemics in 1918, 1957, and 1968 (WHO, 1980) are caused by H1, H2, and H3 viruses, respectively, when the HA specificity have been changed (Ha *et al.*, 2001; Rogers & D'Souza, 1989; Matrosovich *et al.*, 1997). Accordingly, the reason why H5 avian influenza outbreak in 1997 failed to develop pandemic was due to the inappropriate α(2,3) specificity.

H5N1 virus could infect cells in the human's lower respiratory tract where α(2,3) terminal SA dominates, this limitation is responsible for inefficient human to human transmission in 1997 (Shinya *et al.*, 2006). Therefore, for the virus to be disseminated it must infect the upper respiratory tracts where it can be shed out by sneezing and coughing. To infect the cells of upper respiratory tracts, the virus should have the ability to bind α(2,6) terminal SAs (Baum & Paulson, 1990; Couceiro *et al.*, 1993). Because both of α(2,3) and α(2,6) linkages are present in the human respiratory tracts (Shinya *et al.*, 2006; Couceiro *et al.*, 1993), while α(2,3) linkages are present in the avian intestine (Naeve *et al.*, 1984), the influenza A infection in human, equines, and swines is respiratory, while it is enteric in avians. Interestingly, pigs acquire both types of linkages in their respiratory tract thus can be infected by both of avian and human viruses (Hinshaw *et al.*, 1981; (Kida *et al.*,

1994). Accordingly, pigs serve as "mixing –vessels" providing the molecular basis of developing human-avian influenza A virus reassortants, similar to those responsible for 1957 and 1968 pandemics (Ito *et al.*, 1998) (Figure 1.9).

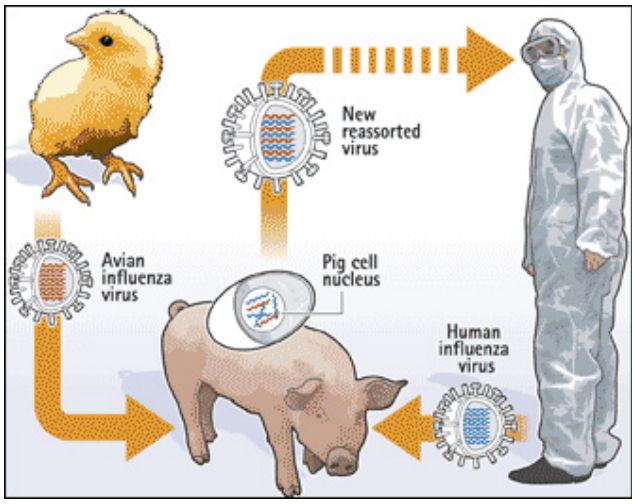


Figure 1.9: Avian influenza A virus is reassorted in pigs to become human infective (Alberto Cuadra, 200?)

In Europe during 1979 H1N1 avian and H3N2 human viruses co-circulated in pigs and eventually reassorted to this host (Castrucci *et al.*, 1993) generating a new type of viruses with HA and NA harbor both avian and human genotypic characteristics capable of infecting humans (Claas *et al.*, 1994). Viruses isolated between 1979 and 1984 were capable of recognizing α(2,3) and α(2,6) linkages, while those isolated after 1985 can recognize only α(2,6) (i.e. human specific). Although humans and nonhuman primates can be experimentally infected with avian viruses, the limited viral replication in these hosts has led to the conclusion that avian influenza viruses are not directly transmitted to humans in nature (Horimoto & Kawaoka, 2001).

1.4.3.2 Determinants of HA1 specificity

Sequence analysis of HA isolated from various avian and human strains revealed that some amino acid substitutions at locations close and far away from the HA1 binding site can affect HA specificity toward the type of SA linkage. Similar to H2 serotypes, H3 serotypes viruses undergo Gln226Leu and Gly228Ser mutations converting the avian specific HA to human specific one (preferring α(2,6) linkages) (Rogers & Paulson, 1983; Rogers *et al.*, 1983; Sauter *et al.*, 1989; Pritchett *et al.*, 1987; Nobusawa *et al.*, 1991; Naeve *et al.*, 1994;; Connor *et al.*, 1994; Vines *et al.*, 1998). These mutations make the SA binding site at HA1 little opened and allow the residues within it to rearrange (Ha *et al.*, 2003). Narrower binding site is observed for H5 serotype of 1997 H5N1 avian virus which has Gln226 and Gly228 and has high affinity for α(2,3) linkages and low affinity for α(2,6) linkages (Figure 1.10). The low affinity to α(2,6) linkages was responsible for H5N1 infectivity to humans in 1997 (Ha *et al.*, 2001). In H1 serotype the Gln226 and Gly228 were maintained while Glu190Asp mutation converts the virus from avian to human specific (Matrosovich *et al.*, 2000).