View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository@USM

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

	Page
DEDICATION	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	v
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF SYMBOLS	xxiii
LIST OF ABBREVIATIONS	xxiv
LIST OF APPENDICES	xxvi
ABSTRAK	xxvii
ABSTRACT	xxix
CHAPTER ONE – INTRODUCTION	
1.1 Problem statement	1
1.2 Influenza virus	4
1.2.1 Hemagglutinin	7
1.2.2 Neuraminidase	12
1.2.3 Influenza A virus subtypes	13
1.3 Sialic acids (SAs)	14
1.3.1 Chemistry of SAs	14
1.3.2 Biological roles of SAs	16
1.3.3 SA is the natural ligand of influenza A HA	17
1.4 HA-SA molecular interaction	18
1.4.1 The binding of SA at HA1 (primary SA binding site)	18
1.4.2 The interaction between SA and HA1 binding site	18
1.4.3 Specificity of the interaction between HA and SA	21
1.4.3.1 Specificity of HA1 toward SA follows the type of SA-O-glycosidic linkage	21
1.4.3.2 Determinants of HA1 specificity	24
1.4.4 Affinity of HA1 toward SA	26

1.5 Anti-influenza A agents 26
1.5.1 NA inhibitors 28
1.5.2 HA inhibitors 30
1.5.2.1 Natural HA inhibitors 30
1.5.2.2 Synthetic HA inhibitors 33
1.6 Influenza A pandemics of the 20th century 36
1.7 Molecular modeling 38
1.7.1 Molecular docking 39
1.7.1.1 Searching algorithms (ligand sampling) 39
1.7.1.2 Scoring functions 44
1.7.1.3 AutoDock3.05 scoring function 49
1.7.2 Molecular descriptors 53
1.7.3 Virtual screening 53
1.7.3.1 Classes of virtual screening 54
1.7.3.2 Databases of molecular structures used in structure-based virtual screening 55
1.7.3.2.1 Databases available for use 55
1.7.3.2.2 Databases generated on demand 57
1.7.3.3 Limitations of structure-based virtual screening 58
1.7.4 Structure-based drug design 59
1.7.5 Combinatorial drug design 63
1.8 Objectives of the study 64

CHAPTER TWO – MATERIALS AND METHODS
2.1 Overview 66
2.2 Molecular preparations 69
2.2.1 Preparation of receptor molecule 69
2.2.2 Preparations of ligand molecules 72
2.3 Stage 1: Validation tests 72
2.3.1 Substage 1: Reproducing the crystal conformation of methyl-α-Neu5Ac at HA1 binding pocket by molecular docking 72
2.3.2 Substage 2: Reproducing experimental affinities for different SA analogues toward influenza A HA by molecular docking 73

2.4 Stage 2: Site-directed fragment-based generation of three databases of single-substituted SA analogues 74

2.4.1 Molecular fragments preparation and orientation at HA1 binding pocket 77

2.4.1.1 Fragments database obtainance and processing 77

2.4.1.1.1 Fragments database disassembly and purification 77

2.4.1.1.2 Fragments conformational optimization 78

2.4.1.1.3 Fragments atomic nomenclature correction, normalization of coordinates, and partial charge assignment followed by PDBQ files preparations 79

2.4.1.2 Fragments docking (orientation) at sites within HA1 binding pocket 79

2.4.2 Substitution of the natural SA functional groups by the oriented fragments to generate three databases of single-substituted SA analogues 83

2.4.3 Preparation and docking of the single-substituted SA analogues databases against HA1 binding pocket 87

2.4.3.1 Conformational optimization of the generated SA analogues, followed by atomic nomenclature correction, partial charge assignment, and PDBQ files preparations 87

2.4.3.2 Docking the three databases of single-substituted SA analogues. 87

2.5 Stage 3: Combinatorial generation of orally bioavailable SA analogues database 88

2.5.1 Filtration of C5- and C6-substituted SA analogues databases by binding energy and deviation of the pyranose ring 91

2.5.2 Extraction of C5- and C6-substituents 91

2.5.3 Possible combinatorial SA analogues that follow the Lipinski's RO5 91

2.5.4 Construction of the best possible combinations 92

2.5.5 Conformational optimization of the generated combinatorial SA analogues followed by LogP filtration, atomic nomenclature correction, partial charge assignment and PDBQ files preparation 92
2.5.6 Docking of the combinatorial SA analogues against HA1 binding pocket 93

CHAPTER THREE – RESULTS AND DISCUSSION
3.1 Validation tests 96
3.1.1 Reproducing the crystal conformation of methyl-α-Neu5Ac at HA1 binding pocket by molecular docking 96
3.1.2 Reproducing experimental affinities of different SA analogues toward influenza A HA by molecular docking 102
3.1.2.1 Correlating EFEB to experimental RA that has been determined by virus hemadsorption inhibition or hemagglutination inhibition assays 104
3.1.2.1.1 Docking SA analogues of RAs measured by viral hemadsorption inhibition assays 104
3.1.2.1.2 Docking SA analogues of RAs measured by hemagglutination inhibition assays 119
3.1.2.2 Correlating EFEB to experimental K_d that was determined by NMR study 126
3.2 Site-directed fragment-based generation of three databases of single-substituted SA analogues 136
3.2.1 The substitution sites on natural SA scaffold 137
3.2.2 Docking (orientation) of molecular fragments within and around HA1 binding pocket 138
3.2.3 Substitution of the natural SA functional groups by the docked (oriented) fragments to generate three databases of single-substituted SA analogues 151
3.2.3.1 Linker molecules 153
3.2.3.2 Fragments attachment process 155

3.2.3.2.1 Step 1: Determination of Anchor Atoms (AnAs) 155
3.2.3.2.2 Step 2: Specifying the best fragment hydrogen atoms that could be substituted by covalent bonds to AnAs and measuring the corresponding inter-atomic distances 156
3.2.3.2.3 Step 3: Converting the measured distances into deviation values in order to calculate the crude score for each AnA 158
3.2.3.2.4 Step 4: Addition of bad contact penalty term to the crude score 162
3.2.3.2.5 Step 5: Hammering the unreasonable growth of molecular linkers 164
3.2.3.2.6 Step 6: Structural refinement 165
3.2.3.2.6.1 Preventing unusual binding modes 165
3.2.3.2.6.2 Evaluating the preservation of C4-hydroxyl of SA scaffold 165
3.2.3.2.6.3 Evaluating the re-inclusion of SA C8-hydroxyl 167
3.2.3.2.6.4 Changing the atomic type of AnA when necessary 168
3.2.3.3 Features of the three designed databases of single-substituted SA analogues 170
3.2.4 Docking of the three generated databases of single-substituted SA analogues against natural SA binding pocket of HA1 176
3.2.4.1 Properties of the docked databases of SA analogues 176
3.2.4.1.1 Binding energies of SA analogues against HA1 binding pocket as calculated by AutoDock3.05 (EFEB) 177
3.2.4.1.2 Stability of designed SA analogues 181

against the natural SA binding site of HA1 during docking simulation
3.2.4.1.3 Preservation of fragments docked (oriented) conformations after attachment to SA scaffolds 186
3.2.4.1.4 Correlation between changing fragments oriented conformations and changing SA scaffolds crystal position during analogues docking 190
3.2.4.1.5 Correlation between activities of the oriented fragments and activities of the generated SA analogues 193
3.2.4.1.6 Effect of linker molecules on activities of the designed SA analogues 196
3.2.4.2 Most actives of the designed single-substituted SA analogues 201
3.2.4.2.1 The most active C2-derived SA analogue 201
3.2.4.2.2 The most active C5-derived SA analogue 207
3.2.4.2.3 The most active C6-derived SA analogue 213
3.3 The generation of orally bioavailable combinatorial (double-substituted) SA analogues database 218
3.3.1 Features of combinatorially designed SA analogues database 220
3.3.2 Docking of the combinatorially designed SA analogues database 221
3.3.3 The most actives of combinatorial SA analogues 225
3.3.3.1 The most active combinatorial SA analogue 225
3.3.3.2 The second most active combinatorial SA analogue 229
3.3.3.3 The third most active combinatorial SA analogue 232

CHAPTER FOUR – CONCLUSION
4.1 Conclusions 237
4.2 Recommendation for future work 240
4.3 Concluding remarks 242

REFERENCES 243

PUBLICATIONS AND AWARDS 259

APPENDICES

I	GPF file for the validation tests	260
II	DPF file for the validation tests	262
III	TCL programming script 1 for salts separation	264
IV	Unix Bash-shell programming script 1 for salts and halogen removal	265
V	TCL programming script 2 for conformational optimization	266
VI	Unix Bash-shell programming script 2 for atomic nomenclature correction, normalization of atomic coordinates, Gasteiger partial charge assignment and PDBQ file preparations	267
VII	DPF for the fragments docking (orientation).	268
VIII	Unix Bash-shell programming script 3 for batch DPF preparation	270
IX	Unix Bash-shell programming script 4 for extracting the atomic coordinates of best fragment's docked conformation in form of .mol file from the .dlg file.	271
X	Unix Bash-shell programming script 5 for fragment attachment	273
XI	DPF for the mediumly slow docking	286
XII	Unix Bash-shell programming script 6 and 7 for recording the analogues docking results and performing fragments plus linkers (substituents) atomic coordinates extraction	288
XIII	Unix Bash-shell programming script 8 for combinatorial search	299
XIV	Unix Bash-shell programming script 9 for combinatorial analogues construction	303

LIST OF TABLES

		Page
1.1	SA recognizing lectins in protozoa, bacteria and viruses (Varki *et al.*, 2008).	8
1.2	The amino acids of the primary and secondary SA binding sites.	11
1.3	The available HA serotypes and their host ranges.	13
1.4	List of the most common docking software, their applications, and types of their searching and scoring functions.	48
2.1	Physicochemical properties of 4541 molecular fragments (from ChemBridge fragments library database) with no bonded halogens and salts.	77
2.2	Linker molecules that were used to attach the oriented fragments to the corresponding SA scaffold's atoms. During attachment process, the chemical group within rectangular brackets will substitute the NH group whenever it is necessary, while the group within crescent brackets will be added to the CH_2 group whenever it is feasible.	85
3.1	Computational values of estimated free energy of binding (EFEB) and inhibitory constant (K_i) calculated by AutoDock3.05, and the RMSD of pyranose ring from the pyranose of crystal methyl-α-Neu5Ac. The experimental values of the relative affinity (RA) have been determined by virus hemadsorption inhibition assay. RAs for analogues 1, 2, 3, 4, 17, 20, 26, 27, 28, and 31 were retrieved from Pritchett, T.J. (1987 cited in Kelm *et al.*, 1992) while RAs for all other analogues including 1 and 2 were retrieved from Kelm *et al.* (1992).	105
3.2	Temperature-factor values for all the atoms forming the natural SA binding site at HA1 binding pocket (PDB ID = 1HGH).	118
3.3	Computational values of estimated free energy of binding (EFEB) and inhibitory constant (K_i) calculated by AutoDock3.05, and the RMSD of pyranose ring from the pyranose of crystal methyl-α-Neu5Ac. The experimental values of relative affinity (RA) have been determined by hemagglutination inhibition experiment (Toogood *et al.*, 1991).	119

3.4 Computational values of estimated free energy of binding (EFEB) and inhibitory constant (K_i) calculated by AutoDock3.05, and the RMSD of pyranose ring from the pyranose of crystal methyl-α-Neu5Ac. The experimental values comprise the dissociation constant (K_d) determined by NMR study and the corresponding Observed Free Energy of Binding (OFEB) value. K_d values for analogues 1, 5, and 6 were obtained from Sauter *et al.*(1989), analogues 2, 3, 4, 7, 8, 14, 15, and 16 were obtained from Sauter *et al.* (1992a) and analogues 9, 10, 11, 12, and 13 were obtained from Machytka *et al.* (1993). 127

3.5 A comparison between the real bond lengths (Atkins & Paula, 2006) and the typical bond lengths considered in the program's calculations. 158

3.6 The equilibrium inter-atomic van der Waals distances between oxygen atom and other heavy atoms present in the fragments database. 166

3.7 The chemical structure, physicochemical properties, and docking results for the next nine C2-derived SA analogues obtained from slow docking experiments. 205

3.8 The chemical structure, physicochemical properties, and docking results for the next nine C5-derived SA analogues obtained from slow docking experiments. 211

3.9 The chemical structure, physicochemical properties, and docking results for the next nine C6-derived SA analogues obtained from slow docking experiments. 216

3.10 The docking results and molecular properties for the 500 combinatorially designed SA analogues. 220

3.11 The chemical structure, physicochemical properties, and docking results for the next seven combinatorial SA analogues obtained from slow docking experiments. 235

LIST OF FIGURES

		Page
1.1	Influenza A virus with the external glycoproteins (HA and NA), trans-membranal proteins (M2), internal protein matrix (M1) and RNA segments.	6
1.2	Influenza A HA homotrimer with the primary and secondary SA binding sites.	9
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).	10
1.4	Different substitutions on C5 give the four main SAs molecules.	15
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).	15
1.6	The molecular structures of natural SA (R=H) and methyl-α-Neu5Ac ($R=CH_3$).	17
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).	20
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.	21
1.9	Avian influenza A virus is reassorted in pigs to become human infective (Alberto Cuadra, 200?)	23

1.10	The HA1 binding site of swine H9 and human H3 that prefer SA of α(2,6) linkage are wider than avian H5 which prefer SA of α(2,3) linkage (Ha *et al.*, 2001).	25
1.11	Inhibition of the influenza A virus replication cycle by the six classes of anti-viral agents (De Clercq, 2006).	27
1.12	A) the NA cleaves off SA from the cell receptor for influenza A virus. B) NA inhibitors sequester the progeny virus from detachment (Moscona, 2005)	28
1.13	Clinically available neuraminidase inhibitors.	29
1.14	Electron micrographs of cells infected with influenza A virus. A) Normal assembly and budding of virus in absence of NA inhibitor, B) Lateral aggregation and formation of large bundles by virus in presence of NA inhibitor (Gubareva *et al.*, 2000).	29
1.15	Antibody mediated immunity against influenza A HA (Subbarao & Joseph, 2007).	30
1.16	Scheme of influenza virus attachment to A) host cell via cooperative interaction of several HA spikes with the complementary array of sialyloligosaccharide moieties of cell-surface glycoproteins and gangliosides, B) soluble macromolecules containing SA prevent viral-cell association by blocking the HA receptor-binding sites and by creating sterical obstacles (Matrosovich & Klenk 2003).	32
1.17	Variants of a multivalent presentation of SA in synthetic inhibitors of influenza virus HA. a) Bivalent sialosides, b) tetrahedrons sialosides, c) dendrimers sialosides, d) liposomic sialosides, and e) polymeric sialosides. (Matrosovich & Klenk 2003).	35
1.18	Areas reporting confirmed occurrence of H5N1 avian influenza in poultry and wild birds since 2003 (www.who.int)	37
1.19	This figure illustrates genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search. The space of the genotypes is represented by the lower horizontal line, and the space of the phenotypes is represented by the upper horizontal line. Genotypes are mapped to phenotypes by a developmental mapping function. The fitness function is $f(x)$. The result of applying the genotypic mutation operator to the parent's genotype is shown on the right-hand side of the diagram, and has the corresponding phenotype shown. Local search is shown on the left-hand side. It is normally	43

performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. (Morris *et al.*, 1998).

2.1 Work flow stages. 68

2.2 The hydrogen bond network formed between conserved amino acids at HA1 binding site (upper part). Histidine amino acid in the two possible tautomerisation states (lower part). 71

2.3 Generation of SA analogues databases by site-directed fragments attachment to SA scaffold (Stage 2). 75

2.4 Part of the crystallographic structure of influenza A HA (PDB ID= 1HGH) shows the SA (sticks representation) at the binding pocket of HA1 with C2-functional group points to site A, C5-functional group extends to site B, and C6-functional group extends to site C. 81

2.5 The natural SA binding site at HA1. Amino acids that form the fragments docking site are represented in Corey-Pauling-Koltun (CPK) model and enclosed by the grid boxes at sites A, B, and C (related to C2-, C5-, and C6-functional groups, respectively). 82

2.6 The three SA scaffold molecules of a, b, and c were used to generate the databases of C2-derived, C5-derived, and C6-derived SA analogues, respectively. 86

2.7 Generation of orally bioavailable combinatorial SA analogues using the best substitutions at C5 and C6 (Stage 3). 90

3.1 Methyl-α-Neu5Ac at HA1 binding site. The protein main chains are represented as solid ribbons and amino acid residues within the binding site are represented as narrow sticks. The crystal and docked SA conformations are represented as bold sticks with atoms coloured by types: carbon, grey (except in docked conformation it is yellow), oxygen, red, nitrogen, blue, and hydrogen, white. Non-polar hydrogen atoms are hidden for image clarity. The hydrogen bonds are represented as red dashed lines for the crystal conformation while it is green for the docked conformation. 99

3.2 Regression analysis shows the correlation between value of the computationally estimated free energy of binding (EFEB) and minus logarithmic value of the experimentally determined relative affinity (-Log(RA)) for all SA analogues listed in Table 3.1. Y stands for affinity. 111

3.3	2D-scatter plots show the correlation between value of computationally estimated free energy of binding (EFEB) and minus logarithmic value of the experimentally determined relative affinity (-log (RA)) for SA analogues of a) modified glycerol side chain, b) modified N-acetyl group, and c) modified C2-substituents. The analogues are listed in Table 3.1.	112
3.4	The docked conformations of SA analogues with modified glycerol side-chain. The structures are colored according to analogues reference numbers listed in Table 3.1 as follow: 1 (yellow), 2 (light gray), 3 (light orange), 4 (dark orange), 5 (light purple), 6 (dark purple), 7 (light cyan), 8 (dark cyan), 9 (pink), 10 (light green), 11 (dark green), 12 (light blue), 13 (dark blue), 14 (dark gray), 15 (black), 16 (brown), and 17 (thin line).	115
3.5	The docked conformations of SA analogues with modified N-acetyl group. The structures are colored according to analogues reference numbers listed in Table 3.1 as follow: 1 (yellow), 2 (light gray), 18 (orange), 19 (brown), 20 (thin line), 21 (light purple), 22 (dark purple), 23 (green), and 24 (blue).	116
3.6	The docked conformations of SA analogues with modified C2-substituents. The structures are colored according to analogues reference numbers listed in Table 3.1 as follow: 1 (yellow), 2 (light gray), 5 (light purple), 25 (thin line colored cyan), 26 (thin line colored blue), 27 (brown), 28 (dark purple), 29 (light green), 30 (dark green), and 31 (dark gray).	117
3.7	Regression analysis shows the correlation between value of the computationally estimated free energy of binding (EFEB) and minus logarithmic value of the experimentally determined relative affinity (-Log (RA)) for all SA analogues listed in Table 3.3. Y stands for affinity.	122
3.8	The docked conformations of SA analogues with C2-substitutions. Different aglycone moieties were attached to C2 through poly-methyl linkers of different lengths. The C2-substituents extend to either of two main nearby grooves at natural HA1 binding site (Grooves A and B). The structures are colored according to analogues reference numbers listed in table 3.3 as follow : 1 (yellow), 2 (light orange), 3 (dark orange), 4 (light purple), 5 (gray), 6 (dark purple), 7 (light cyan), 8 (dark cyan), 9 (light violet), 10 (dark violet), 11 (light green), 12 (dark green), 13 (light blue), 14 (dark blue), and 15 (brown).	124

3.9	2D-scatter plot of EFEB calculated by AutoDock3.05 scoring function against number of heavy atoms and polar hydrogens for all SA analogues listed in Table 3.3.	126
3.10	Regression analysis shows the correlation between value of the computationally estimated free energy of binding (EFEB) and values of the experimentally Observed Free Energy of Binding (OFEB) for all SA analogues listed in Table 3.4.	131
3.11	The docked conformations SA analogues of different substitutions. The structures are colored according to analogues reference numbers listed in Table 3.4 as follow : 1 (yellow), 2 (light orange), 3 (dark orange), 4 (purple), 5 (light gray), 6 (light cyan), 7 (dark cyan), 8 (light pink), 9 (dark pink), 10 (light green), 11 (dark green), 12 (dark green), 13 (light blue), 14 (dark blue), 15 (brown), 16 (black), and 17 (dark gray).	132
3.12	The molecular structures of natural SA (R=H) and the simplest SA analogue, methyl-α-Neu5Ac, (R=CH_3).	137
3.13	The frequency distribution of EFEB values (Kcal/Mol) obtained from docking ChemBridge fragment library database against a) Site A, b) Site B, and c) Site C of HA1.	142
3.14	Fluctuation in values of Intra-Molecular Interaction Energy (IntraMIE) (Kcal/Mol), Molecular Weight (MW/10) and number of Torsional Degrees of Freedom (TDOF) in response to the decrease in value of Estimated Free Energy of Binding (EFEB) (Kcal/Mol). The values were obtained from docking Chembridge fragments library database against a) Site A, b) Site B, and c) Site C of HA1.	143
3.15	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for fragment no. 1122 at site A of HA1 binding pocket.	146
3.16	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for fragment no. 2734 at site B of HA1 binding pocket.	147
3.17	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for fragment no. 1122 at site C of HA1 binding pocket.	148
3.18	The natural SA binding site of HA1 is occupied with three different conformations of one type of fragment molecules, a) fragment no. 1122, and b) fragment no. 2734. All of the three docked conformations (green CPK models) occupy part of crystal binding site of methyl-α-Neu5Ac (stick model).	150

3.19	An example of systematic search that was performed to specify which fragment's hydrogen atoms are nearest to a) AnA1, b) AnA2, c) AnA3, and d) AnA4. The carbon atoms of the fragment molecule are coloured green. The distances between AnAs and nearest hydrogens are denoted by red arrows while the distances between AnAs and the fragment heavy atoms, that carry the nearest hydrogens, are denoted by blue arrows.	157
3.20	The deviations from typical distances are shown between AnAs and both of fragment's Heavy Atom (HvA) (green arrows) and fragment's hydrogen atom (H) (pink arrows). The radiuses of circled orbits around fragment HvA represent the approximate fixed inter-atomic distances between HvA and HvA (outer bold circle of 1.5 Å) and between HvA and H (inner thin circle of 1 Å). The gray hemisphere with radius of 0.5 Å around fragment H represents the sites where AnA could be presented to substitute the hydrogen, with increasing opportunity toward the darker region. With respect to AnA1 and AnA2, the deviations from the typical AnA-HvA distance are equal for both atoms. However the deviation from the typical AnA-H distance is smaller for AnA2 compared to AnA1. Therefore, AnA2 has a more suitable location to substitute H than AnA1.	160
3.21	Attachment of an oriented fragment to the SA scaffold C6 using a) the deviation value of fragment's heavy atom (HvA) alone, or b) combined with the deviation value of fragment's nearest hydrogen atom. Fragment's carbon atoms are coloured green and only the hydrogen atoms of the best HvA are shown. The distance between AnA and the nearest fragment's hydrogen atom is represented as red arrow while the distance between AnA and the best fragment's HvA that carry the nearest hydrogen is represented as blue arrow. The bond to be constructed is coloured yellow.	161
3.22	Attachment of an oriented fragment to the SA scaffold C6 using a) the Penalized Score (PS), and b) the Accumulated Penalized Score (APS).	163
3.23	Attachment of an oriented fragment to the SA scaffold C2 using the Accumulated Penalized Score (APS) with a) no AnAs chain hammering, and b) with AnAs chain hammering.	164
3.24	Two different cases of fragments attachment to C5 of SA scaffold. In the first case a clash occurred between C4-oxygen and one of the fragments heavy atoms which results in dropping of the C4-hydroxyl in the generated SA analogue. While in the second case no clashes was encountered and the C4-hydroxyl is preserved. Fragment's carbon atoms are	167

coloured green and Corey-Pauling-Koltun (CPK) molecular models are used to represent the atomic van der Waals radii.

3.25 Attachment of an oriented fragment to the SA scaffold C5 when AnA2 is a) nitrogen, or b) carbon. In the first case, AnA3 (which is carbon atom) was used for fragment's connection in spite of having poor accumulated penalized score (APS) compared to AnA2. This was to avoid the construction of N-N bond. However in the second case the atomic type of AnA2 was changed to carbon and C-N bond was constructed. 169

3.26 The fluctuation in number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), calculated logarithm of octanol-water partition coefficient (cLogP), and number of torsional degrees of freedom (TDOF) in response to the molecular weight (MW/10) for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 172

3.27 Molecular linkers frequencies for connecting oriented fragments to SA scaffold's a) C2, b) C5, and c) C6. 175

3.28 The distribution frequencies for the values of EFEB (Kcal/Mol) obtained from docking a) C2-derived, b) C5-derived, and c) C6-derived SA analogues against HA1 binding pocket. 179

3.29 2D-scatter plots show the correlation between the number of Torsional Degrees of Freedom (TDOF) and value of EFEB (Kcal/Mol) obtained from docking of a) C2-derived, b) C5-derived, and c) C6-derived SA analogues against HA1 binding pocket. 180

3.30 The frequency distribution of RMSD values of pyranose ring (Å) from the crystal methyl-α-Neu5Ac for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 184

3.31 2D-scatter plots show the correlation between RMSD value of pyranose ring (Å) and the value of EFEB (Kcal/Mol) for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 185

3.32 The frequency distribution of RMSD value between fragment's oriented and attached conformations in a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 188

3.33 2D-scatter plots show the correlation between values of RMSD of attached fragments (Å) and values of EFEB of the generated analogues (Kcal/Mol) for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 189

3.34 The fluctuation in RMSD value of attached fragment in response to the increase in RMSD value of pyranose ring of the analogue for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 192

3.35 2D-scatter plots show the correlation between values of EFEB (kcal/mol) for the oriented fragment and the correspondingly generated analogue for all a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 195

3.36 2D-scatter plots show the correlation between the size of linker molecule incorporated and the mean RMSD value for the attached fragments in all a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 199

3.37 2D-scatter plots show the correlation between the size of linker molecule incorporated and the mean RMSD value of pyranose ring for a) C2-derived, b) C5-derived, and c) C6-derived SA analogues. 200

3.38 The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C2-2734 at HA1 binding pocket. 203

3.39 The superposition of the crystal conformation of methyl-α-Neu5Ac (yellow), the docked conformation of SA-C2-2734 (gray), and the docked conformation of fragment no. 2734 at site A which has been used to generate the analogue (green). 204

3.40 The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C5-2734 at HA1 binding pocket. 209

3.41 The superposition of the crystal conformation of methyl-α-Neu5Ac (yellow), the docked conformation of SA-C5-2734 (gray), and the docked conformation of fragment no. 2734 at site B which has been used to generate the analogue (green). 210

3.42 The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C6-1122 at HA1 binding pocket 214

3.43 The superposition of the crystal conformation of methyl-α-Neu5Ac (yellow), the docked conformation of SA-C6-1122 (gray), and the docked conformation of the fragment no. 1122 at site C which has been used to generate the analogue (green). 215

3.44	a) Accumulated bars show the sum value of EFEB (Kcal/Mol) of parents for all the 4382 combinatorial opportunities. b) Comparison between the combinatorially designed SA analogue and the C5-derived and C6-derived parents concerning EFEB (kcal/mol), c) 2D-scatter plot show the correlation between the sum of EFEB values of the C5-derived and C6-derived parents (kcal/mol) and the EFEB value of the combinatorial analogue (kcal/mol).	223
3.45	a) comparison between the RMSD value of the pyranose ring between the C5-derived and C6-derived parent analogues and the combinatorial analogue, b) the fluctuation in RMSD values of the C5- and C6-attached fragments of the combinatorial analogues in response to the decrease in value of EFEB (kcal/mol).	224
3.46	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C5-62-C6-521 at HA1 binding pocket.	227
3.47	Superposition of the docked conformations for the SA-C5-62-C6-521 (Purple), the C5-derived and C6-derived parents (Gray), and for their corresponding oriented fragments at site B and C, respectively (Green).	228
3.48	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C5-63-C6-1044 at HA1 binding pocket.	230
3.49	Superposition of the docked conformations for SA-C5-63-C6-1044 (Purple), the C5-derived and C6-derived parents (Gray), and for their corresponding oriented fragments at site B and C, respectively. (Green).	231
3.50	The chemical structure, physicochemical properties, docking results, docked conformation and non-bonded interactions for SA-C5-543-C6-62 at HA1 binding pocket.	233
3.51	Superposition of the docked conformations for SA-C5-543-C6-62 (Purple), the C5-derived and C6-derived parents (Gray), and for their corresponding oriented fragments at site B and C, respectively (Green).	234

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

		Page
I	GPF file for the validation tests	260
II	DPF file for the validation tests	262
III	TCL programming script 1 for salts separation	264
IV	Unix Bash-shell programming script 1 for salts and halogen removal	265
V	TCL programming script 2 for conformational optimization	266
VI	Unix Bash-shell programming script 2 for atomic nomenclature correction, normalization of atomic coordinates, Gasteiger partial charge assignment and PDBQ file preparations	267
VII	DPF for the fragments docking (orientation)	268
VIII	Unix Bash-shell programming script 3 for batch DPF preparation	270
IX	Unix Bash-shell programming script 4 for extracting the atomic coordinates of best fragment's docked conformation in form of .mol file from the .dlg file.	271
X	Unix Bash-shell programming script 5 for fragment attachment	273
XI	DPF for the mediumly slow docking	286
XII	Unix Bash-shell programming script 6 and 7 for recording the analogues docking results and performing fragments plus linkers (substituents) atomic coordinates extraction	288
XIII	Unix Bash-shell programming script 8 for combinatorial search	299
XIV	Unix Bash-shell programming script 9 for combinatorial analogues construction	303

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).

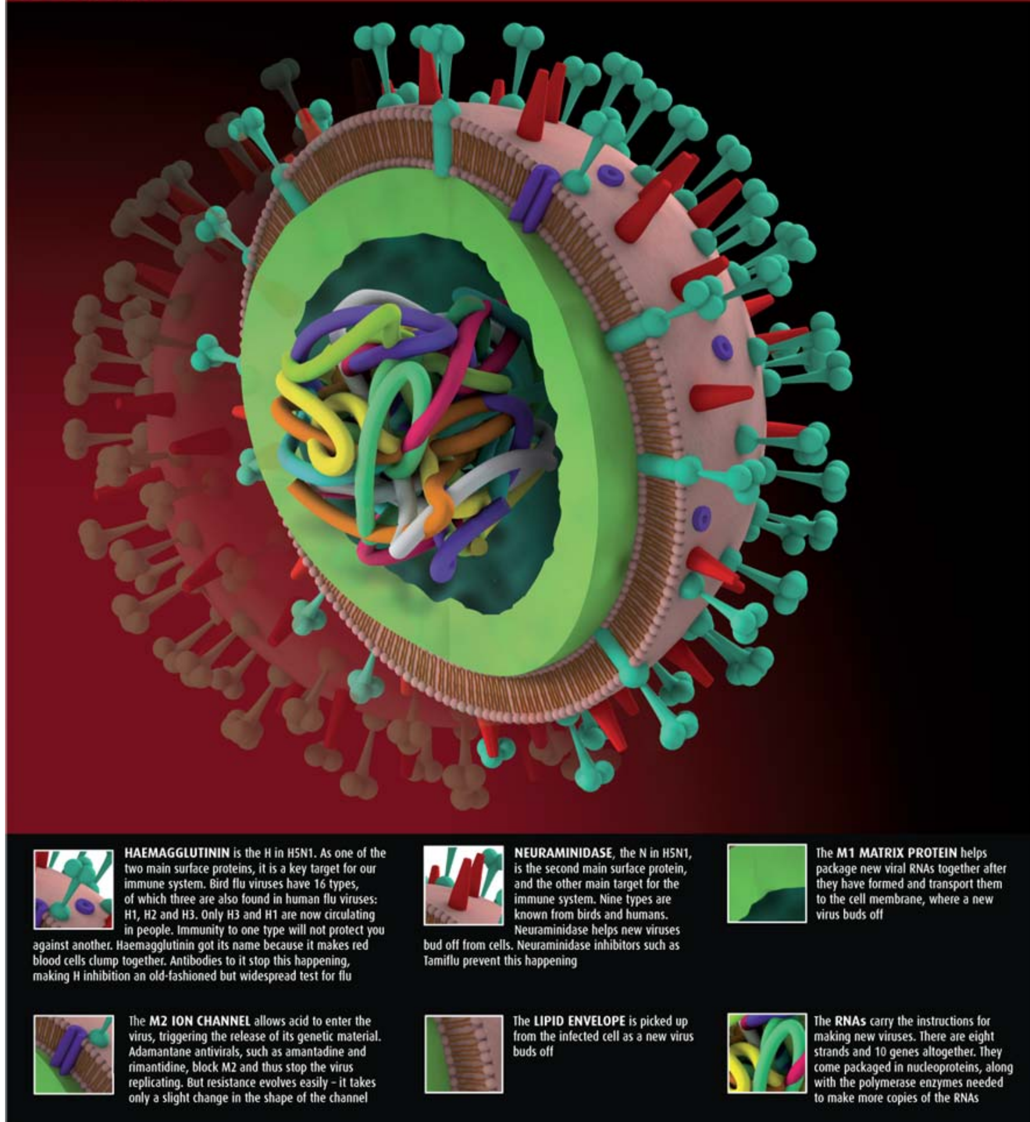
Molecular modelling techniques are developing fast, with some branches are matured enough and effectively participated in designing and screening drug candidates. With respect to HA, AutoDock3.05 has been used to reproduce crystal conformation of SA within HA primary binding site with estimated binding 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).



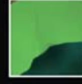


Figure 1.1: Influenza A virus with the external glycoproteins (HA and NA), trans-membranal proteins (M2), internal protein matrix (M1) and RNA segments.

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

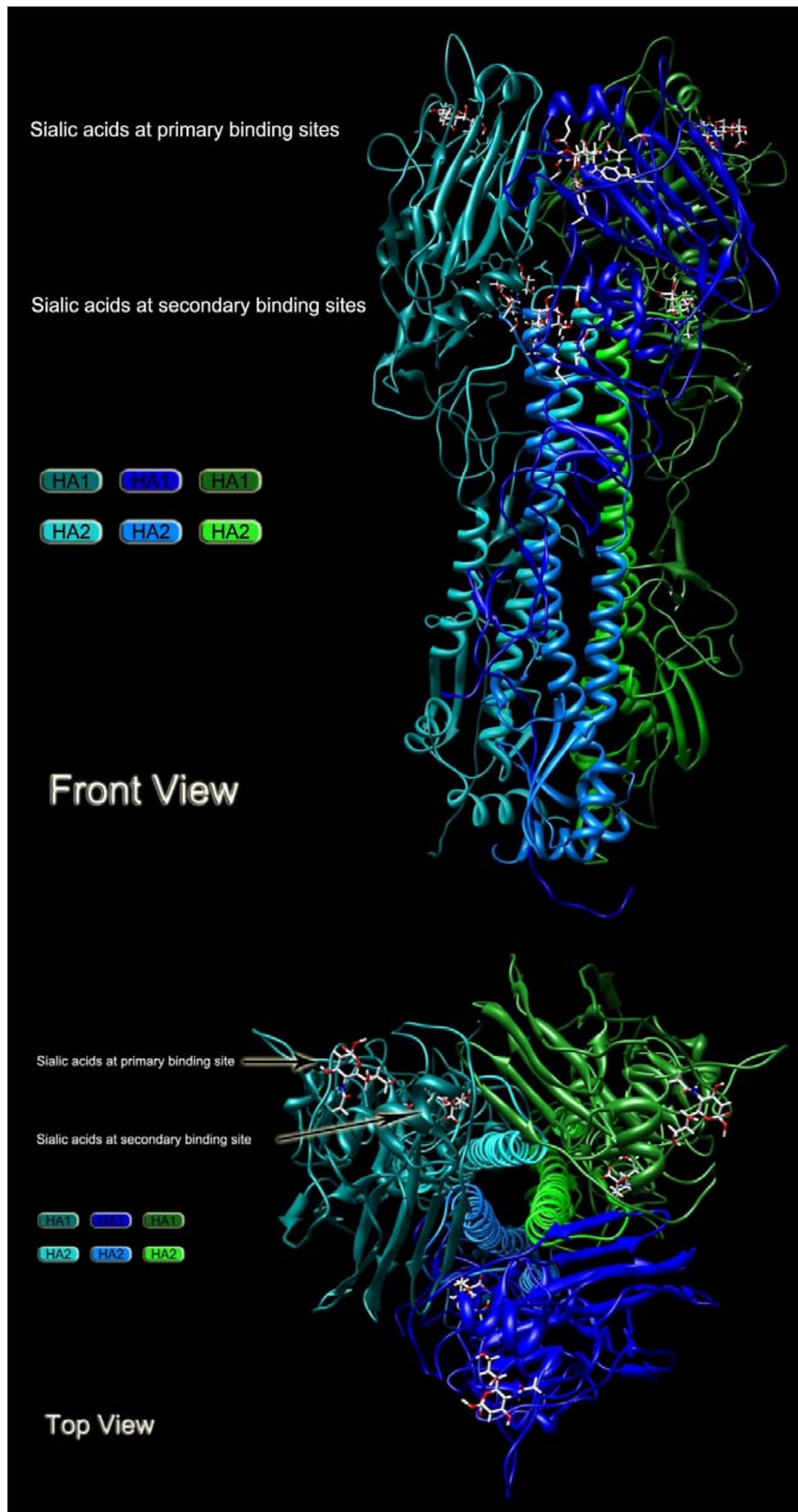


Figure 1.2: Influenza A HA homotrimer with the primary and secondary SA binding sites.

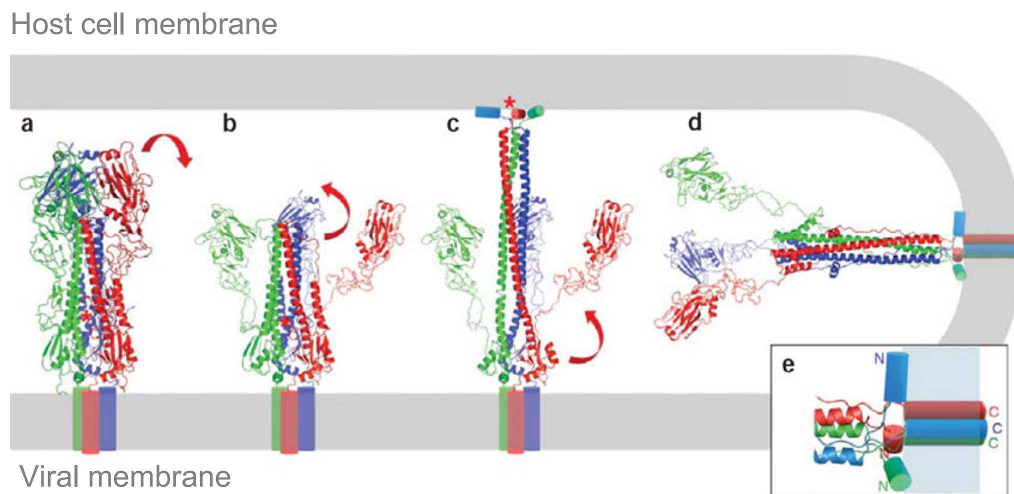


Figure 1.3: The process of fusion between viral and endosomal membranes mediated by viral HA. a) Influenza A HA exposed on the viral surface (bottom) and pointed toward the host cell membrane (top). b) HA1 subunits displaced aside from locations over HA2, c) The loops between shorter and longer helices 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.

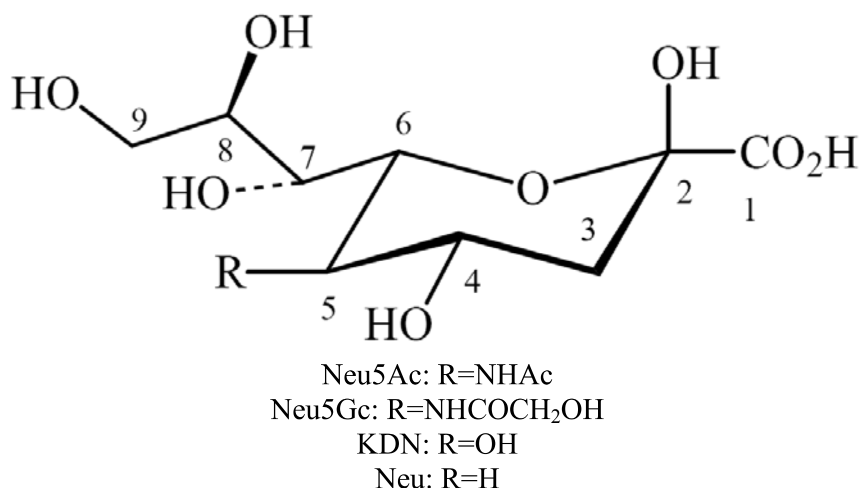
HA serotype	Host range			
	Avian	Human	Equine	Swine
H1	X	X		X
H2	X	X		X
H3	X	X	X	X
H4	X			
H5	X			
H6	X			
H7	X		X	
H8	X			
H9	X			X
H10	X			
H11	X			
H12	X			
H13	X			
H14	X			
H15	X			
H16	X			

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).

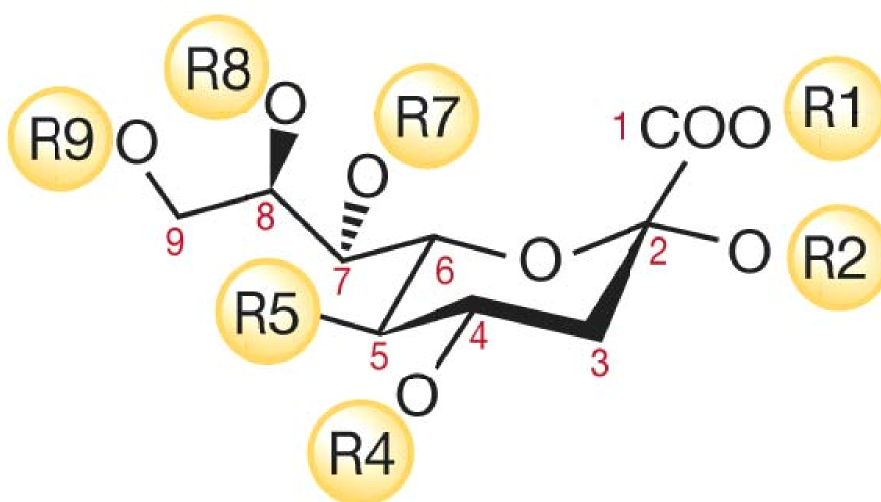
1.3.1 Chemistry of SAs

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.

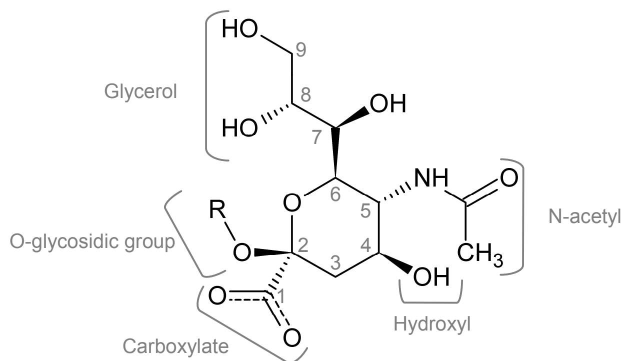


Figure 1.6: The molecular structures of natural SA (R=H) and methyl-α-Neu5Ac (R=CH_3).

1.4 HA-SA molecular interaction

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

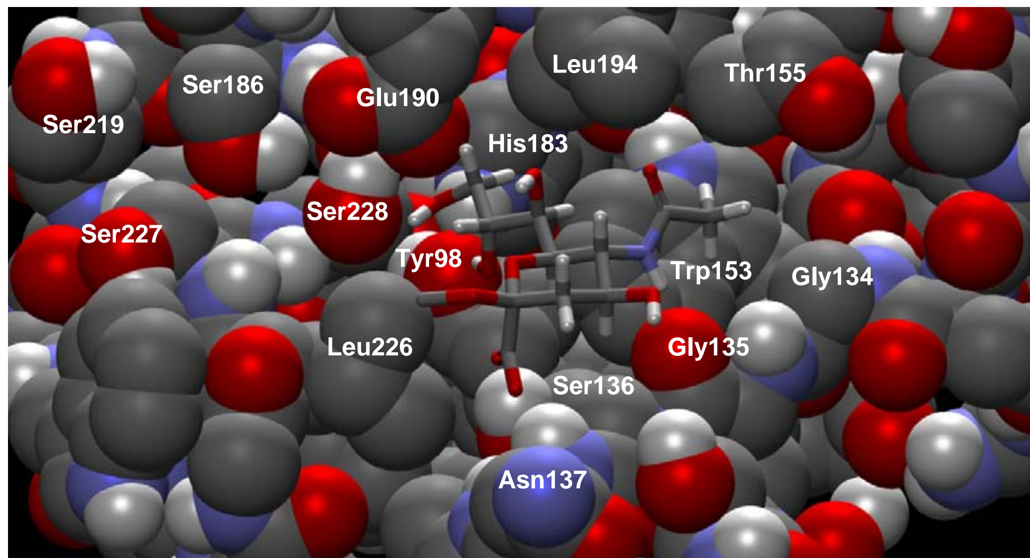
The primary SA binding sites are located at HA1 subunits of HA and formed from well conserved amino acid residues. The phenolic hydroxyl of Tyr98 and the aromatic ring of Trp153 form the bottom of the binding site. There are three polypeptide loops forming the boundaries of this binding site; i.e. loop 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).



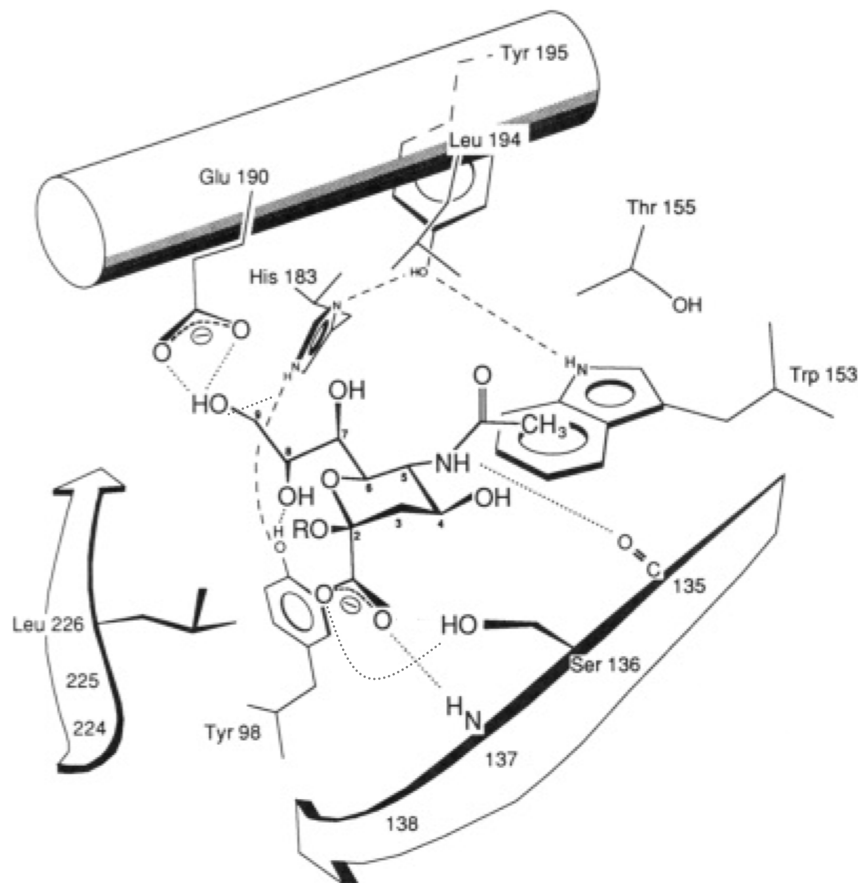


Figure 1.7: Natural SA binding site at HA1 of influenza A virus H3N2 (X-31). Dotted lines indicate hydrogen bonds between SA and HA, while dashed lines show potential hydrogen bonds within the protein (Sauter *et al.*, 1989).

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).

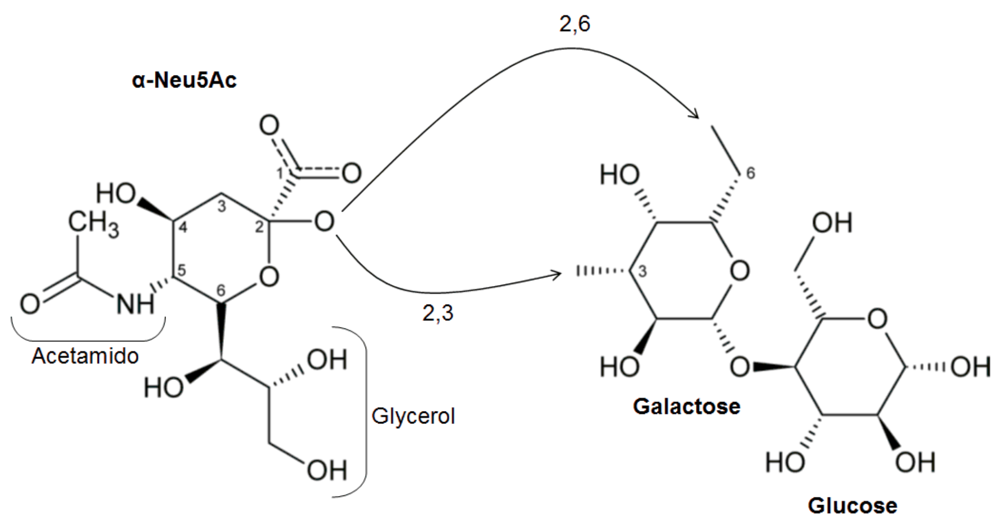


Figure 1.8: Different HAs have different specificities in recognizing these linkages. The HA derived from wild type A/Hong Kong/68, X-31 H3N2 influenza A virus with Leu226 (Human virus) has high affinity to Neu5Ac-α(2,6)-Gal while the Leu226Gln mutant (Avian virus) favours Neu5Ac-α(2,3)-Gal linkage.

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).

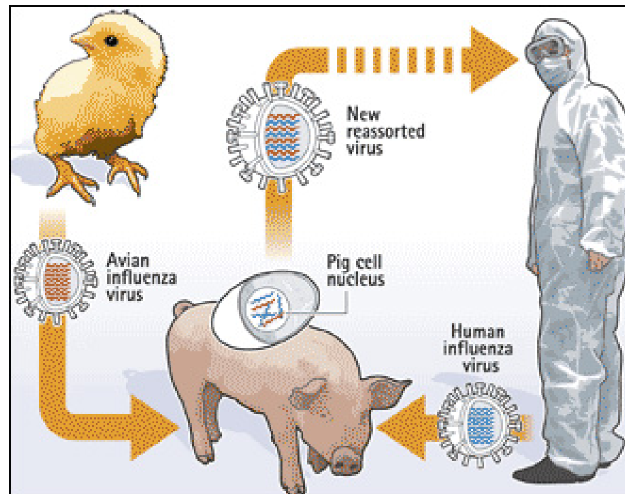


Figure 1.9: Avian influenza A virus is reassorted in pigs to become human infective (Alberto Cuadra, 200?)

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).