# Molecular modelling and lead design of substituted zanamivir derivatives as potent anti-influenza drugs

*A Major Project dissertation submitted*

*in partial fulfilment of the requirement for the degree of*

**Master of Technology**

**In**

**Bioinformatics**

*Submitted by*

**Dhwani Dholakia**

**(2K12/BIO/07)**
**Delhi Technological University, Delhi, India**

*Under the supervision of*

Dr. Asmita Das

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road,
Delhi-110042, INDIA

# DECLARATION

I, Dhwani Dholakia hereby declare that the report entitled "**Molecular modelling and lead design of substituted zanamivir derivatives as potent anti-influenza drugs**" submitted in partial fulfilment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is a record of original and independent research work done by me under the supervision and guidance of Dr. Asmita Das, Assistant Professor, Delhi Technological University (DTU), the thesis has not formed the basis of the award of any Degree/Diploma/Associateship/Fellowship or other similar title to any candidate of any university/institution.

Date:                                                                                    Signature of candidate

## ACKNOWLEDGEMENT

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

GQSAR – Group-based Quantitative structure activity structure relationship.

QSAR- Quantitative activity structure relationship.

HA- Hemagglutinin.

NA-Neuraminidase.

RNA- Ribonucleic acid.

DNA- Deoxyribonucleic acid.

NP- Nucleoprotein.

NEP-Nuclear export Protein.

mRNA- Messenger RNA.

vRNP- Viral Ribonucleoprotein.

HIS- Histidine.

TRP- Tryptophan.

cRNA- Complementary RNA.

PA- Polymerase Acidic protein.

PB- Polymerase Basic protein.

ARDS -Acute respiratory distress-like syndrome.

2-D- Two dimension.

3-D- Three dimension.

IC- Inhibitory constant.

XP-Extra precision.

OPLS- Optimized Potentials for Liquid Simulations.

MD-Molecular dynamics.

RMSD- Root mean square deviation.

SASA -Solvent accessible surface area.

# Molecular modelling and lead design of substituted zanamivir derivatives as potent anti-influenza drugs

Dhwani Dholakia

Delhi Technological University, Delhi, India

# 1. ABSTRACT

Influenza virus spreads infection via two main surface glycoproteins namely Hemagglutinin (HA) and Neuraminidase (NA). NA cleaves the sialic acid receptors eventually releasing newly formed virus particles. These released viral particles then invades new cells. Inhibition of NA could limit the infection to one round of replication of virus which is not enough to cause the disease. In this study, novel series of acylguanidine zanamivir derivatives were used to develop a Group-based QSAR (GQSAR) model targeting NA in different strains of influenza virus *viz.*H1N1 and H3N2. Unlike other traditional QSAR approaches, GQSAR gives the flexibility to study fragments within a molecule and its contribution in the inhibitory effect of the compound. A correlation analysis was carried out comparing the statistics of the measured IC50 values with the predicted ones. The descriptors were interpreted graphically. A combinatorial library was developed, activities of the compound were predicted using these GQSAR models and docking study was performed on the top scoring compounds. Docking study revealed the binding orientation of these inhibitors at the active site of NA (150-loop). The top compound (AMA) was selected for carrying out molecular dynamics simulations for 15ns which provided insights into the time dependent dynamics of the designed leads. AMA possessed a docking score of -8.26 Kcal/mol with H1N1 strain and -7.00 Kcal/mol with H3N2 strain. Ligand-bound complexes of both H1N1 and H3N2 were observed to be stable for 11 ns and 7 ns respectively. In addition, absorption, distribution, metabolism and excretion (ADME) descriptors were also determined to evaluate their pharmacokinetic properties. These studies provide valuable insight for designing more potent and selective inhibitors for the treatment of Influenza.

# 2. INTRODUCTION

Type A influenza virus, member of Orthomyxoviridae family (Nelson *et al*., 2007) is one of the most lethal and virulent strains of influenza virus which has been responsible for worldwide havoc including seasonal epidemics and major pandemic breakthroughs(Neumann *et al.,* 2009). Influenza could have originated via two major mechanisms: direct transmission from birds to humans as in 1918"Spanish Influenza" virus (H1N1) or through transfer of genetic material from avian influenza virus to human influenza virus as happened in the case of 1968 "Hong Kong" influenza virus (H3N2) (Carrat *et al.,* 2007). It is a highly contagious virus and causes severe respiratory associated problems. Complications include post influenza encephalitis, secondary bacterial pneumonia and change in cardiac electrocardiogram.

Subtypes of the type A virus has been classified based on the serological activity of the glycoprotein antigens hemagglutinin (HA) and neuraminidase (NA).Sixteen serotypes of HA have been found to circulate in mammalian and avian hosts. HA is a surface envelop protein of influenza virus and performs crucial viral functions like host recognition and membrane fusion (Skehel *et al*., 2000). HA often recognizes sialic acid receptors found in the human upper respiratory tract which act as initial key step of viral infection (Wiley *et al*., 1987). The second glycoprotein NA is a sialidase which destroys HA present on the surface of the virus allowing release of the infected viral progeny from infected cell thus preventing their self-aggregation(Gong *et al*., 2007). Thus, inhibiting NA prevents second round of replication of influenza virus therefore culminating further influenza infection. Sequence analysis of nine subtypes of NA separates them into two major phylogenetic groups. Group 1 consists of N1, N4, N5 and N8 while group 2 consists of N2, N3, N6, N7 and N9.

Active site of NA contains 150-loop which includes residues from 147 to152 and is present in two forms. First is an open conformation which adopts 150-loop formation and the other is a closed conformation in which active site lacks 150-loop conformation (Amaro *et al*., 2011). Analysis of X-ray crystal structure (Russell *et al*., 2006) shows an open conformation for NA in group 1 and a closed conformation for those in group 2. However molecular dynamics simulation suggested the presence of 150-loop not only in group 1 but also in group 2 (Amaro *et al*., 2011). These findings provide deep insight into the design and synthesis of new NA inhibitors targeting the 150-loop cavity. Based on these structures FDA approved drugs(Gubareva *et al*., 2000) like Oseltamivir(Tamiflu), Zanamivir(Relenza) and Peramivir are commercially available to treat infected patients. Oseltamivir, an oral prodrug (administered as phosphate) is hydrolysed hepatically to its active form carboxylate while Zanamivir, is administered via nasal inhalation because of high polar compounds. To alleviate the potential consequence of suboptimal bioavailability and clumsy inhalational devices an intravenous peramivir antiviral drug was used. However, this type of treatment has limited control as the developed vaccines frequently become ineffective because of mutation in influenza viral antigen

taking place at a rapid rate in forms of antigenic shift or drift (Zambon, M.C. 1999). Thus a search for new influenza drug with broad spectrum activity is the need of hour.

Considerable amount of work has been done to target the 150-loop cavity through modification of the existing inhibitors by attaching various additional groups with appropriate size, shape and hydrophobicity. *In silico* methods provide substantial contribution to drug design and development of lead compounds in limited time and resources. Quantitative Structure Activity Relationship (QSAR) is a method of ligand-based drug designing that establishes relationships between structure and inhibitory activity of inhibitors. Group-based QSAR (GQSAR) gives flexibility to traditional QSAR methods by calculating descriptors for the fragment of a molecule rather than calculating descriptors for whole molecule. Various 2-D descriptors calculated for fragments prepared by applying specific chemical rules include molecular weight, element counts, topological index etc are calculated for fragments at various substitution sites. Unlike the traditional QSAR methods, this method can be applied to both congeneric as well as non-congeneric series of compounds. Internal validation and Cross validation validate QSAR models. Using combinatorial library approach with QSAR is helpful in designing novel H1N1 and H3N2 inhibitors.

# 3. REVIEW OF LITERATURE

Subtypes of influenza virus including A, B, C are very similar in their structure. The viral particle have an average diameter size of 100 nm and is usually spherical in shape, though filamentous form also occur and is seen in C subtypes. However, differences in shapes the viral particles have similar composition. The core particle contains RNA genome and other viral proteins that are required for packaging. The complete genome of influenza consists of 13,588 nucleotide and 8 single stranded negative sense ribonucleic acid (RNA) segments. Each RNA consist of one or two genes which codes protein. Genome of virus contains 11 genes on total eight pieces of RNA. These protein includes NA, HA, nucleoprotein (NP), NS1, NS2 Nuclear export protein (NEP), M1, M2 (Matrix) and three polymerase PA, PB1 and PB2. The two glycoproteins HA and NA together comprises maximum glycoprotein in 4:1 ratio. Nucleotides at the 5' and 3' end of the genome are complementary to each other. Individual segments are encapsulated in nucleocapsid and is surrounded by outer envelope.

## 3.1 Viral proteins

### 3.1.1 Hemagglutinin Protein (HA)

The protein was named based on its ability to agglutinate erythrocytes. There are 16 different subtypes (H1-H16) described in wild water fowl. Among all subtypes only H1, H2 and H3 have been found in human's seasonal influenza epidemics and pandemics. The protein forms a rod shaped trimeric complex. HA0 precursor is made up of two subunits HA1 and HA2.Both the subunits remains attach via disulphide bond. HA1 forms globular domain and attaches itself to the sialic acid of the glycosylated protein or lipids. The virus is internalized via pH dependent endocytosis. The acidified medium of endosomes provokes a conformational change in the H1-H2 structure exposing HA2 subunit thus promoting viral fusion with endosomal region. Besides role of receptor recognition and fusion HA is a major antigen which is recognized by adaptive immune system of host. The selection pressure created by neutralizing antibodies can help escape mutants mostly found in H1 domain.

### 3.1.2 Neuraminidase protein (NA)

NA protein was named after its ability to catalyze the cleavage of neuraminic acid that also known as sialic acid. It forms a homo-tetramer complex that has globular domain on the top short stem. The removal of sialic acid from cell surface protein by NA allows efficient release of mature, newly synthesized particles from host cell membrane. This is crucial for virus to propagate making NA a suitable drug target. Besides HA protein, NA also generates an immune response that results in non-neutralizing antibodies. Even though receptor-binding of the virus is not affected, viral spread can be restricted to a certain degree that may allow the host to control the viral infection.

### 3.1.3 NS proteins- NS1 and NEP/NS2

The NS segments code for two proteins, NS1 and NEP/NS2. NS1 is a translated product of fill length NS mRNA while NEP/NS2 is a splice variant of NS mRNA. Both the proteins share first nine amino acids from their N-Terminus. The probability of export of unspliced mRNA is very high therefore its chances of being spliced is minimized (Alonso-Caplen *et al*., 1992). Thus the amount of spliced viral transcript is 10% of the spliced transcript (Lamb *et al*., 1980). This further confirms that NS1 protein is present during the early stage of virus replication to thwart antiviral response. This mechanism avoids early NEP/NS2 protein expression and thus prevents premature export of vRNP.

- **NEP/NS2 nuclear export protein**

The NEP/NS2 protein acts as an adaptor protein for nuclear receptor namely Crm1 and viral M1 protein. It directs the export of newly synthesized RNP that are complex with M1. The NEP/NS2 with M1 blocks the nuclear localization signal (NLS) in M1 which may prevent the re-entry of RNP into nucleus.

### 3.1.4 M proteins (M1 and M2)

The M segment codes for viral proteins M1 and M2. The first eight amino acids of both the proteins are identical. M1 protein is translated from unprocessed mRNA while the M2 protein is translated from splice variant.

- **M1 protein**

M1 protein is abundant in an infected cell along with NP protein. M1 associates with lipid membrane. Although M1 lacks transmembrane domain it interacts with cytosolic domain of transmembrane proteins HA, NA and M2.In the infected cell M1 is necessary for export of RNP out of the nucleus. During the export M1 protein interacts with both vRNA and NP protein (Martin *et al*., 1991; Elton *et al*., 2001; Watanabe, *et al*., 2001). The M1/RNP complex binds to target via M1 to NEP/NS2 protein and then eventually to Crm1. Crm1 is export receptor and thus directs the export of M1/RNPNEP out of the nucleus.
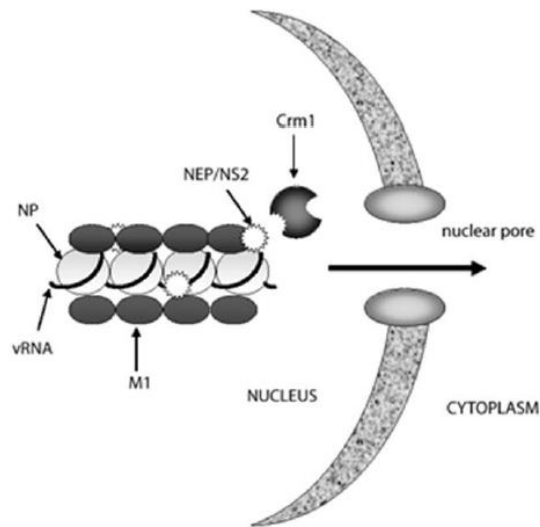
Figure 1:- Illustration of influenza viral RNP nuclear export. The nuclear export of newly synthesized viral RNP complexes is mediated by the viral NEP/NS2. It interacts with the cellular export factor Crm1 and with the viral RNPs via the M1 protein. The interaction with the cellular export machinery facilitates the transport of the viral genome through the nuclear pore into the cytoplasm.

- **M2- ion channel protein**

The M2 forms an ion channel protein that remains rooted in the lipid bilayer of the virion. The ion channel is activated at low pH (Pinto *et al*., 1992) and has some early and late role in viral life cycle (Hay *et al*., 1985). During early infection it pumps protons from acidified endosome into virion. The low pH weakens the binding of M1 to RNP allowing release into cytosol after fusion with host cell membrane. The late role includes regulation of pH in Trans Golgi network by pumping protons out. This thus prevents premature conformational change in HA1/HA2 protein which is further transported to budding site. HIS 32 acts as pH sensor (Wang *et al*., 1995) and TRP 41 blocks C-terminal of the closed channel . The side chains of both HIS and TRP forms a narrow channel which rarely allow anything (Tang *et al*., 2002) to pass through.

The protonation of His37 plays a major role in channel opening. The positively charged rings and their electrostatic repulsion destabilizes the helix-helix packaging eventually widening and opening the pore. Once the channel is opened the pH gradient drives the proton from region of low concentration to high concentration. The M2 is also considered as one of the target besides NA protein. Derivatives of compounds such as adamantane were fund to inhibit influenza viral replication (Davies *et al*., 1964). Though it is a question of debate that weather drug binds on the inside or outside of the M2 tetramer to block the ion channel activity (Schnell *et al*., 2008), drug binding stabilizes the tightly packed closed channel which makes it difficult to initiate conductance. Accumulation of certain mutations like L26F, S31N destabilize the closed conformations allowing opening of the ion channel by activation of HIS 37. Increasing rate of resistance led to impediment for treatment of influenza.

### 3.1.5 PB1-F2 protein

The accessory protein PB1-F2 is expressed from alternative ORF in PB1 segment of influenza virus (Chen *et al*., 2001). Studies have shown this factor to be pathogenic and pandemic in influenza strain of 1918 and in virulent strain of H5N1 Hong Kong virus (Conenello *et al*., 2007). The factor may exert its virulence function by modulating immune response and enhancing secondary bacterial infection, causing morbidity in influenza patients (McAuley *et al*., 2007). It has also been shown that the protein factor localizes to mitochondrial membranes and induces apoptosis by changing the membrane potential. It forms pore in mitochondrial membrane thereby releasing cytochrome C.

### 3.1.6 The Polymerase complex PA, PB1 and PB2

The viral dependent RNA polymerase complex consist of three subunits PA, PB1 and PB2 proteins. Both the transcription and replication takes place inside the nucleus and all three proteins have nuclear localization signals (Nath *et al*., 1990; Martin *et al*., 1991). PA and PB1 enters nucleus as dimer (Fodor *et al*., 2004) and then forms complex with PB2 which is imported independently. Once the viral RNA is transported to nucleus they are reverse transcribed into mRNA which are then capped and polyadenylated. The replication is two-step process. The negative-sense vRNA is copied into positive sense strand known as cRNA (1st step of replication). This sense strand serves as template to replicate more vRNA (2nd round of replication). The viral polymerase catalyzes three different reactions: vRNA- directed mRNA synthesis, vRNA directed cRNA synthesis and cRNA directed vRNA synthesis.

- **PA- polymerase acidic protein**

This polymerase protein is required for transcription and replication(Fodor *et al*., 2004) and has shown to have endonuclease activity. The N terminal domain has cap dependent endonuclease activity. The viral polymerase has shown to have highly conserved activity. The polymerase was shown to have been associate with host cell RNA polymerase II. It specifically degrades phosphorylated subunit of RNA polymerase II which correlates with the shutdown of host cell transcription and onset of viral replication and transcription.

- **PB1 –polymerase basic protein 1**

The polymerase protein has conserved motifs with RNA-dependent RNA polymerase activity (Biswas *et al*., 1994). It catalyzes the sequential addition of nucleotides to RNA transcripts. PB1 imitates replication and transcription by attaching itself to both the ends of cRNA or vRNA.

- **PB2 –polymerase basic protein2**

This protein binds to the caps of host pre-mRNA (Fechter *et al*., 2003).. It deliver these RNA to PA subunit which has endonuclease activity and cleaves 9-17 nucleotides downstream of cap. This process is called "cap snatching".

### 3.1.7 Nucleoprotein (NP)

NP forms a major component of PNP. It binds to RNP and forms helical rod like structures (Klumpp *et al*., 1997; Noda *et al*., 2006). It is one of the most abundant viral protein and performs various viral functions like viral RNA replication, trafficking of viral genome, viral genome packaging and viral assembly.

### 3.2 Viral Replication

Viral replication is multi step process (Colacino *et al*., 1999). Four major steps are:-

1)    Attachment and penetration of virus into host
2)    Viral genome transcription
3)    Replication of the viral genome
4)    Release of viral particle

1)    **Attachment and penetration of virus into host**- Virus attach itself to host cell by attaching to HA glycoprotein, which is considered as primary step for infection (Colacino *et al*., 1999. HA forms spikes on the lipid membrane of virus. HA is synthesized as inactive precursor which is then cleaved by trypsin to give active proteins (Steinhauer, D.A. 1999). The linkage between sialic acid and carbohydrates of glycoproteins play vital role. Human viruses recognize α 2-6 linkage while equines and avian recognize α 2-3 linkages (Bertram *et al*., 2010). Swine recognize booth types of linkages, thus suggesting dangerous pathogenicity of swine as a mixing vessel of both human and avian influenza virus. Infection is initiated when HA1 binds to its receptor sialic acid and then penetrate to host by cell membrane fusion. The virus enters into host cell via receptor mediated endocytosis. The low PH of host induces conformational changes in virus exposing the HA2 fusion peptide. The acidic environment not only aids in fusion but also opens M2 ion channel. M2 ion channel is a proton selective ion channel. Opening of ion channel acidifies the viral core releasing RNP which is free to enter host cell's cytoplasm.

2)    **Transcription of Viral genome-** Transcription and replication of virus occurs inside the nucleus of the host. As the viral RNP's enter host, the viral endonuclease cleaves host mRNA about 13-15 bases from the 5' capped end.  The small cleaved part is used as a primer by the virus to synthesize complementary mRNA. These mRNA further undergo translation leading to formation of polypeptide which are subsequently cleaved to form viral proteins.

3)    **Replication of the viral genome-** Genome replication does not require a primer instead the viral RNA dependent RNA polymerase initiates the replication.   During replication complementary RNA (cRNA) is produced which is complementary to the RNA genome. This cRNA act as template for production of viral genomic segments with the help of polymerase protein. The newly synthesized RNA associates with nucleoprotein (NA) to form ribonucleoprotein (RNP) (Neumann *et al*., 2004).

4) **Release of viral particle**: - Newly formed RNP particles are transported to cytoplasm of host cell where it associates with other protein to form virion. On maturity the virus buds off through outer cell membrane and the NA proteins then cleaves sialic acid form glycoproteins, thus releasing virus particle's (Neumann *et al*., 2004).

### 3.3 Viral variation

A high rate of mutation ($> 7.3 * 10^{-5}$ base per cycle of replication) occurs during viral replication. Such high rate of error is due to lack of proofreading by viral RNA polymerase. These mutations in viral genes cause changes in viral proteins which results in antigenic drift. Some of the substitutions of amino acids can reduce the binding affinity of antibodies. This results in increased virulence of the new strain within nonspecific immune population (Carrat *et al*., 2007). In poultry, it was seen that LPAI virus circulating in wild birds for several months underwent antigenic drift and became HPAI virus (Ito *et al*., 2001). Another variation antigenic shift also was responsible for various pandemic. Antigenic shift occurs when a host is simultaneously infected by two or more strains of influenza virus (Cox *et al*., 2000). It is abrupt change in influenza A virus that results in new HA or NA. Shift result in a new combination of HA and NA that has emerged new in population and is completely different from the subtype that normal human possess. Such shift occurred in 2009 when H1N1 virus emerged with a new combination of genes that infect and spread quickly among people. When shift occurred most people had very little protection against newly emerged virus. Usually influenza viruses change by antigenic drift but undergoes antigenic shift very rare. Results in it has been seen that pigs are susceptible to both human and avian type influenza virus (Li *et al*., 2004). Thus swine is considered as mixing vessel for genetic resortment

### 3.4 Influenza transmission

Influenza transmission mainly occurs through respiratory droplets from the affected patients via coughing, sneezing or talking. These droplets come in direct contact with individual or is transmitted through inanimate object. Two major types of droplets can be produced by a sick individual: aerosol and large droplets. Aerosol particles are smaller than 5 micrometers and are generally formed in the lower respiratory tract while large droplets are usually large in size and are greater than 10 micrometers makes large droplets. They are found in upper respiratory tract especially nasopharyngeal region.

Aerosol nuclei can remain suspended in air for prolonged periods and can be disseminated by air currents or can be inhaled by the host. When the droplets are inhaled they remain deposited in the lower respiratory tract because of the small size. Aerosol transmission can be of concern in a small room with improper ventilation like in closed space or in airplanes. Small particles tend to be more dangerous as they can settle deeper into lungs. Hence proper ventilation and air handling becomes a major concern to prevent spread of droplet nuclei. In contrast to small droplets, large droplets transmission occurs when the droplets nuclei are spread

3 to 6 feet through sneezing or coughing or through contact with affected human directly or indirectly with secretions settled on surfaces. This type of secretions do not stay suspended in air or affected by special type of handling of air. It has also been a topic of debate that large droplets are major mode of influenza transmission

Influenza virus can survive on variety of surfaces at $28^{o}$ C and 35-49% humidity. Virus can survive 24-48 hours on non-porous surface like steel, plastic and 8-12 hours on porous surfaces like cloth or tissue. However life of particle is also influenced by temperature, moisture and characteristic of virus. Various interventions have been proposed to reduce the transmission which includes improved hygiene (proper hand washing and alcohol sanitizer), wearing surgical mask and proper air handling prevents the spread of infection (Roberts *et al*., 2000).

### 3.5 Influenza manifestation

Infection begins when the virus particles (virion) enters the host cell and begins to replicate. Initially the host has no detectable symptoms and does not release virus into environment this period called latent period. After 1-3 days of latent period the infection begins and the person starts showing symptoms like fever, sore throat. The time from initial infection till the appearance of symptom is called incubation period. In addition people with seasonal influenza can become infectious without symptoms or very negligible symptoms. These 'healthy carriers comprises major infected population. This latent symptoms makes it difficult to distinguish healthy from sick and avoid infection when well.

### 3.6 Influenza Past Pandemics

Pandemics in the pre-virology era are difficult to recognize but are usually characterized by excess mortality that overlapped with global illness that clinically and epidemiologically resembles influenza. The first pandemic spread occurred in 1580 which spread from Asia to Africa then Europe and America. Pandemics occurred continually before the isolation of first influenza virus. The major outbreaks include 1729-33, 1781-82, 1830-33, 1889-90 and 1918-19.

First pandemic which occurred in 1918 was one of the most destructive outbreak which killed about 100 million people, which is more than total of both world war. Studies shows that virus decreased the average life of people by 10 years. The virus that had originated in the Kansas (US) came to be known as Spanish flu, since Spain was first to admit the impact of outbreak. Eventually the flu spread across the world even remote areas like pacific and artic islands. This pandemic caused havoc worldwide which forced schools, house, church, theaters to close to stop infection. The pandemic begun in military camp in the spring session which was followed was two more waves of fall and winter. The average fatality rate was 2.5% but was seen high in pregnant women ranging from 23-71%. It is anticipated that excess death caused by 1918-19 may be a result of severe acute respiratory distress-like syndrome (ARDS) or secondary bacterial pneumonia. It appears that viral infection banquets the respiratory tract causing severe damage

which is often followed by secondary bacterial invasion. In addition to that investigators have postulated that a healthy immune response in population induced a "cytokine storm" that caused ARDS and may have contributed to high mortality rate. Spanish flu had had both economic and health effects for those born during pandemic. All these factors differentiate Spanish flu from other pandemics.

Another pandemic outbreak occurred in 1957 called Asian flu pandemic. This flu was originated in East Asia that subsequently spread worldwide. This outbreak was caused by influenza virus A known as H2N2 or Asian flu virus. It was found that the virus was a mixed strain originating from strains of human and avian influenza. In 1960 they underwent minor genetic modifications via antigenic drift which produced periodic epidemics. The virus first spread throughout china and its surrounding regions. By midsummer it reached United States. The first wave attack in US was amongst school age children while the second wave affected elderly population. During this time there was a limited supply of vaccine and 10-35% of the population were affected. The overall mortality rate was around 2 million worldwide.

After 10 years of second pandemic the H2N2 virus disappeared which was replaced by a new influenza subtype H3N2 via antigenic shift. This resulted in third pandemic which took place in 1968 in Hong Kong. After detection in Hong Kong, the virus spread worldwide and had a cumulative death of about 1 million people. Many theories explained the lower fatality rate that neuraminidase was similar to that of Asian flu and thus prior infection during early pandemics would have provided some immunity. Another reason may be that antibiotics and medical care were available for the affected people.

### 3.7 Antiviral drugs for influenza

Treatment recommended for influenza primarily includes of antipyretics, intake of fluids unless fever and other symptoms lessens. Certain antiviral drugs have been found effective in lessoning viral infections. They play a vital role in management of influenza reactions. They directly target influenza viruses and are effective when administered during early course of infection. Two classes of antiviral agents are currently FDA approved: Neuraminidase inhibitors (Itzstein *et al*., 2009) and M2 ion channel blockers. The M2 blockers include rimantadine and amantadine (Hay *et al*., 1985) are effective against A subtype viruses but not against B subtype. Efficacy of these drugs has been greatly reduced by rapid emergence of mutational resistance among subtype A virus including H1N1 and H3N2. The recently emerged 2009 virus have also shown resistance against adamantine. Two currently approved NA drugs includes zanamivir and oseltamivir against both type A and type B viruses including novel 2009 pandemic virus. Research on NA inhibitor peramivir (though not licensed) is still on clinical trials using intramuscular or intravenous formulation. Parallel trends were observed worldwide during 2008-2009 season with many countries reporting 100% oseltamivir resistance on H1N1 viruses. Hence it is critical to boost investigation on susceptibility of, mainly for those having N1 enzyme (eg
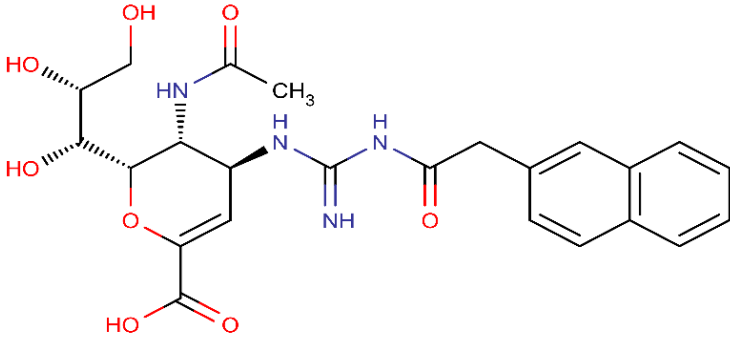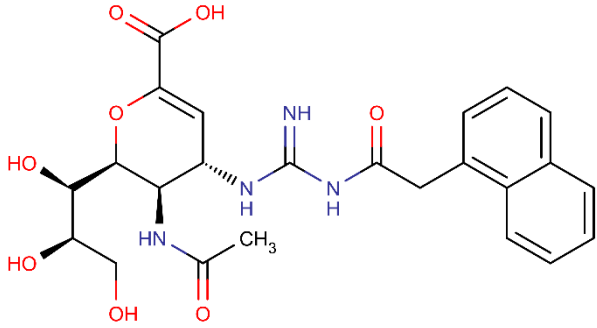
H1N1 and H3N2) as NAI are the only and most effective treatment for both pandemic and seasonal influenza. Comparatively zanamivir has been often less prescribed due to certain current limitations. Thus resistance of zanamivir has been a rare case. Such few cases includes Zanamivir resistance in type B patients with R152K mutations and a novel Q136K mutations. Some cases of invitro zanamivir resistance have also been found including E119G, D, A and R292K.Currently detection of NAI resistant virus is conducted using either chemiluminescent or NA inhibition assays where IC50 values of the test are compared with those of sensitive control cases. Even though the elevated IC50 values are not enough for defining NAI resistance in virus, a combination of known molecular markers of resistant must be used.
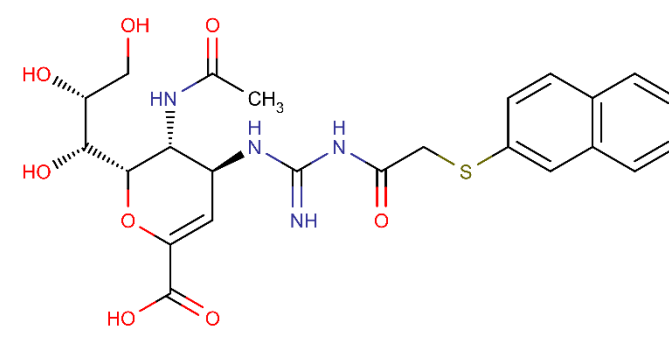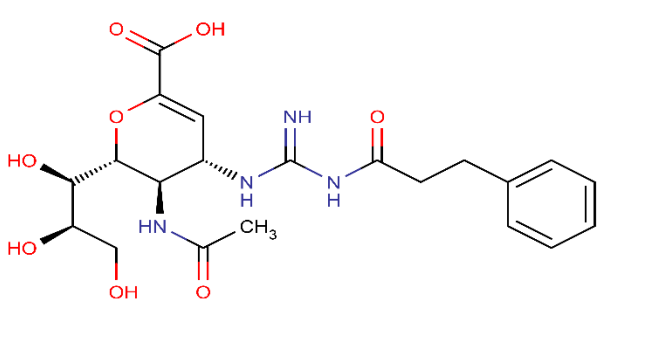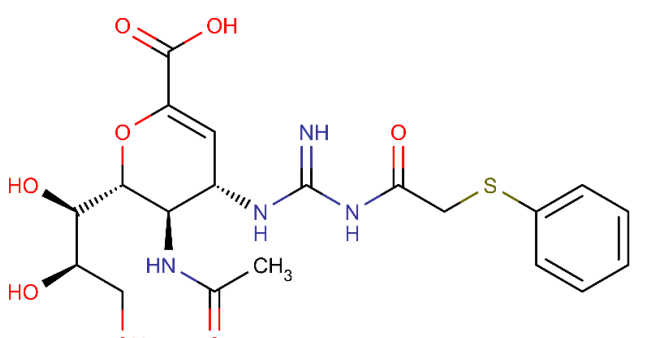
# 4. MATERIALS AND METHODS

### 4.1 Preparation and optimization of data set

Marvin sketch (ChemAxon Ltd., http://www.chemaxon.com/marvin) was used to draw experimentally reported 24 acylguanidine zanamivir derivatives. The compounds were drawn in 2-D format and then converted to 3-D using VlifeEngine module of VLifeMDS (Singla *et al.*, 2010).The prepared compounds were minimized using force field batch minimization platform of VlifeEngine ver 4.3 provided by Vlife Sciences, Pune on Intel® Xeon(R).

Table 1: Structures and anti-influenza activity of acylguanidine zanamivir derivatives.

| Serial no | Compound | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|---|---|---|---|---|---|
| f |  | 54.9 | -1.740 | 58.3 | -1.766 |
| g |  | 98.3 | -1.93 | 84.4 | -1.926 |

| Serial no | Compound | | | | | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|---|---|---|---|---|---|---|---|---|---|
| j | | | | | | 20.1 | -1.303 | 25.5 | -1.407 |
| l | | | | | | 203 | -2.307 | 234 | -2.369 |
| m | | | | | | 189 | -2.276 | 211 | -2.324 |

| Serial no | Compound | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|---|---|---|---|---|---|
| q |  | 332 | -2.521 | 472 | -2.674 |
| r |  | 176 | -2.246 | 209 | -2.320 |
| s |  | 1200 | -3.079 | 1400 | -3.146 |

| Serial no | Compound | | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|---|---|---|---|---|---|---|
| v |  | | 421 | -2.624 | 339 | -2.530 |
| w |  | | 180 | -2.255 | 173 | -2.238 |
| x |  | | 397 | -2.599 | 472 | -2.674 |

| Serial no | Compound | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|---|---|---|---|---|---|
| aa |  | 775 | -2.889 | 868 | -2.939 |
| ab |  | 541 | -2.733 | 615 | -2.789 |
| ac |  | 203 | -2.307 | 196 | -2.292 |

| Serial no | Compound | Ic50 values for H1N1 | Pic50 values for H1N1 | Ic50 values for H3N2 | Pic50 values for H3N2 |
|-----------|----------|----------------------|-----------------------|----------------------|-----------------------|
| ae |  | 34644 | -4.540 | 38303 | -4.583 |
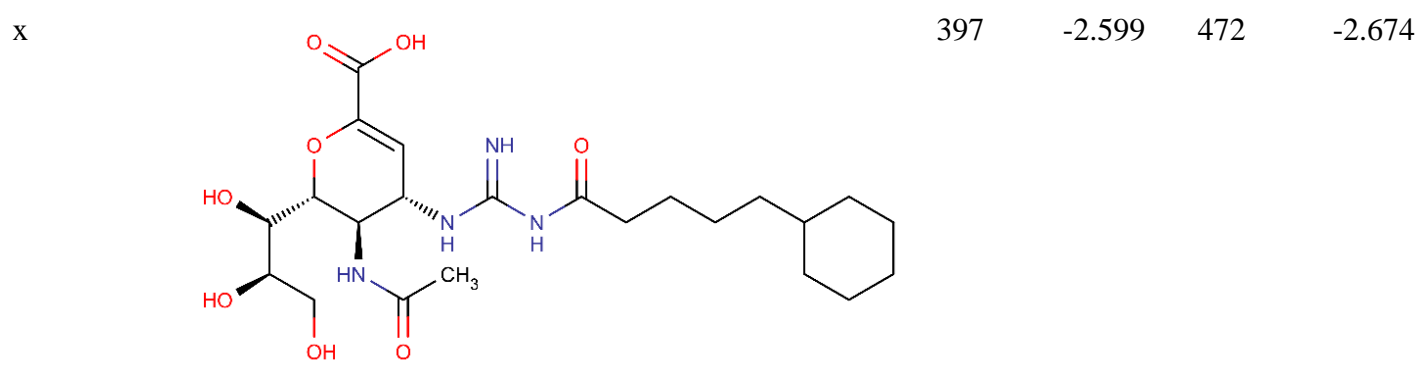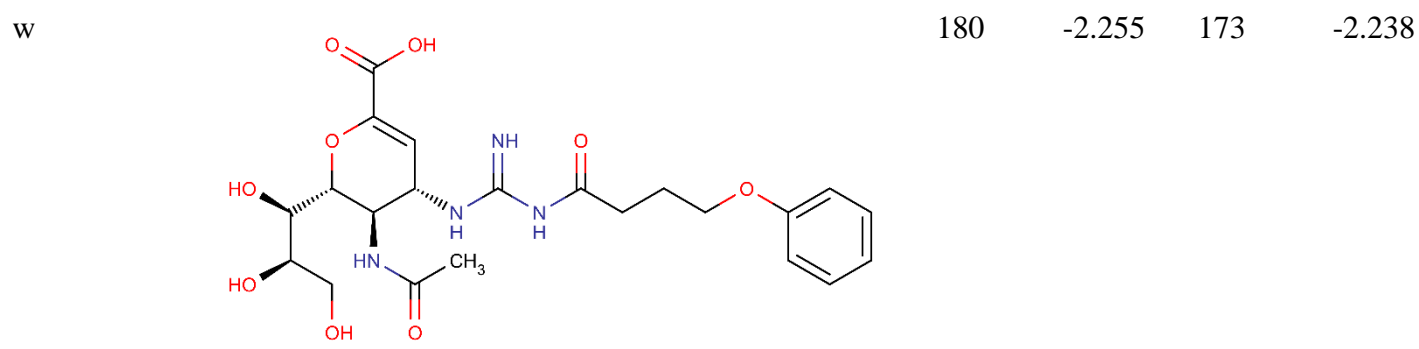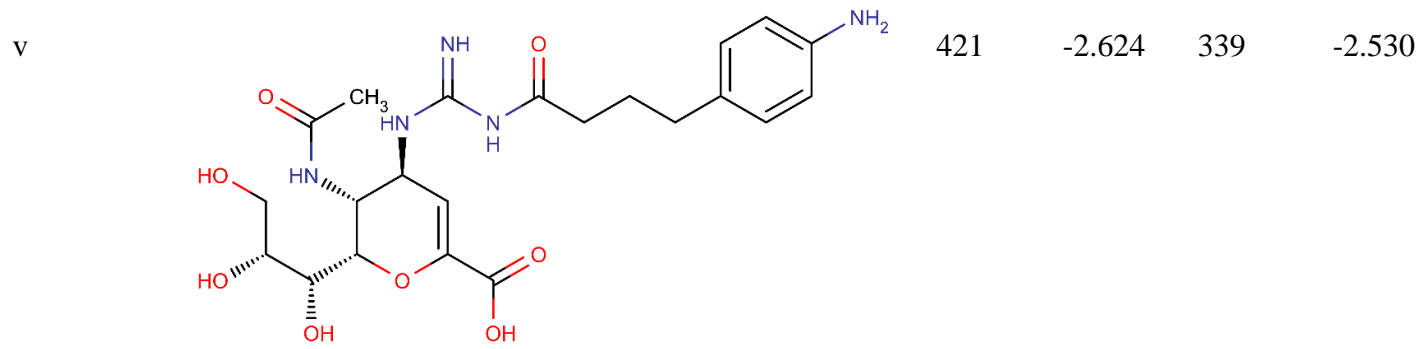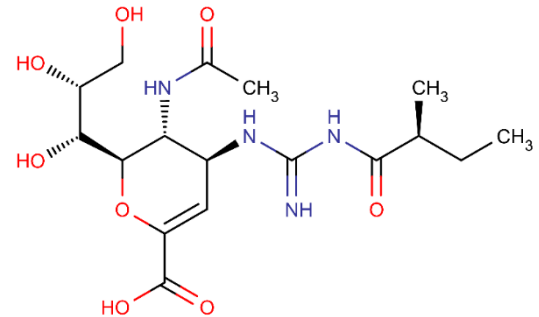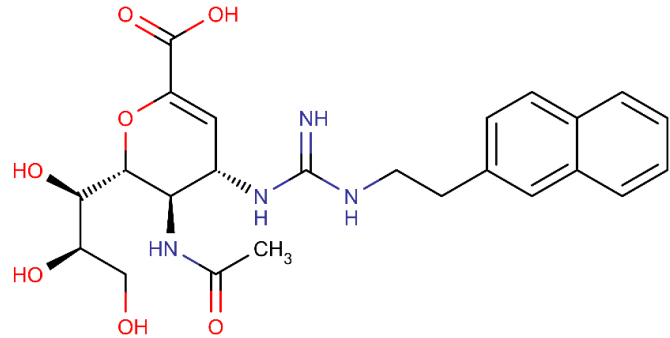
## 4.2 Calculation of Descriptors for GQSAR model development

In this GQSAR study various descriptors correlating the inhibitory activity of molecules were identified. Descriptors are numerical values of the chemical information that is encoded by a molecule which is calculated by certain logical and mathematical operations. A predefined nomenclature is adapted for descriptors in which the particular substitution site is added as a prefix to the calculated descriptor. Since the template molecule selected for this study had only one substitution site R1 descriptors were named as R1_SdOE_index, R1_chiV3 and R1_kappa2 etc. GQSAR model was built using the GQSAR module of VlifeMDS. The common scaffold, representative of all the structures was used as a template for the GQSAR study. Using Modify module of VLifeMDS, template (Figure 2) was created by replacing dummy atoms at R1 on the common moiety i.e. template.
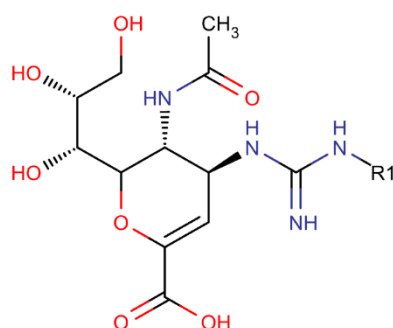


Figure 2:- Representation of common template for acylguanidine zanamivir derived compounds.R1 represents the substitution site of congeneric series.

Optimized set of compounds and template molecule were then imported for template based GQSAR model building. Experimentally reported IC50 values (half maximal inhibitory concentration) were converted to pIC50 scale (-log IC50) to narrow down the range (Table 1).). Thus, a higher value of pIC50 exhibits a more potent compound. These values were then manually incorporated in VLifeMDS. Physicochemical 2-D descriptors were calculated for functional group at substitution site (R1). Total of 101 descriptors out of 343 descriptors were further used for QSAR analysis while rest were removed owing to invariability.

## 4.3 Development of GQSAR model using multiple regression method

For developing a robust and efficient model, the data set of compound were divided into training and test set. For GQSAR against NA of H1N1, 16 molecules were grouped into training set while8 molecules namely f, l, n, o, q, t, y and Ae were grouped in test set. For the second NA target of H3N2, 16 molecules were chosen for training set and 8 molecules namely ac, ae, j, m, q, r, w, y were selected for test set. After manual selection, the unicolumn statistics for both the training and test sets were calculated which provides validation of the chosen training and test sets. Stepwise-forward method was used as variable selection. The next step involved, building of a GQSAR model using multiple regression analysis which predicts the activity using the

selected descriptors. Regression analysis is process which estimates the relationship between a dependent variable and one or more independent variable. For this model Column containing the activity values (pIC50) was selected as dependent variable while rests other were selected as independent variables. In general multiple regression can be explained in the terms of following equation:

$$Y = \alpha + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Where Y is the independent variable, $\alpha$ is the intercept, $\beta_n$ is the slope for $n^{th}$ independent variable X.

### 4.4 Validation and Evaluation of the developed Model

This step was done to test both the stability and predictive ability of the developed GQSAR models. Various statistical parameters (Golbraikh *et al*., 2000) like k (number of variables), n(number of compounds), $r^2$(Squared degree correlation), $q^2$(cross validated correlation coefficient), Pred_$r^2$(for external test set),Z score (Randomization test), F-Test, best_ran_$q^2$ (Highest value of $q^2$ in randomization test), best_ran_$r^2$ (highest value of $r^2$ in randomized test) and standard error were calculated to test goodness of fit of the developed model. For a model to be robust, values should be above threshold i.e. $r^2$ >0.6, $q^2$>0.6, and Pred_$r^2$>0.5 (Afantitis, *et al.,* 2009; Golbraikh *et al*., 2000; Golbraikh *et al*., 2002). F-Test calculates the ratio of the variance observed by the model and variance due to the error in regression. Higher values of F-Test and lower values of standard error of Pred_$r^2$se, $r^2$_se and $q^2$_se indicate a statistically reliable model.

#### 4.4.1 Internal Validation

Internal validation was carried out using leave one out ($q^2$ LOO) method. For calculating $q^2$, a molecule in the training set was eliminated in each iteration and then the activity of the eliminated molecule was predicted based on the remaining molecules thus, accounting for prediction of biological value of the removed molecule. $q^2$ was calculated using the equation:

$$q^2 = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - y_{mean})^2}$$

where $y_i$ and $\widehat{y}_i$ represents actual and predicted activity respectively of the $i^{th}$ molecule in the training set and $y_{mean}$ is the average activity of total molecules in the training set.

#### 4.4.2 External Validation

External validation was performed to predict the activity of each molecule in the test set using models generated from training set. The pred_r2 values is calculated using the equation:

$$Pred\_r^2 = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - y_{mean})^2}$$

where $y_i$ and $\hat{y}_i$ represents actual and predicted activity respectively of the $i^{th}$ molecule in the test set and $y_{mean}$ is the average activity of total molecules in the test set.

### 4.4.3 Randomization Test

To compute the statistical significance of a GQSAR model of actual data, One Tail hypothesis testing was used (Zheng *et al*., 2000).The robustness of models generated for training set was examined by comparing the models derived from random data sets. Random sets were generated by randomly shuffling the molecules and their activities in the training set. Statistical parameter Z score was used to calculate the significance of the models obtained by comparing individual scores with the mean score of the entire data set. Z score was calculated using the following equation:

$$Z\ score = \frac{(h - \mu)}{\sigma}$$

Where h represents $q^2$ values for actual data set, μ represents the average $q^2$ and σ represents the standard deviation for various iterations used to build models from different random data sets.

## 4.5 Development of combinatorial library

Combinatorial library was generated using the Leadgrow module of VLifeMDS by substituting various chemical groups at the substitution site R1 site. The library generated consisted of 189 molecules. Prediction of activity and descriptor for each of the substituted site was calculated using the developed GQSAR model via generic prediction module.

## 4.6 Protein and ligand preparation for docking studies

The protein crystal structure of both H1N1 (PDB ID: 3BEQ) and H3N2 (PDB ID: 4GZ0) were retrieved from protein databank. Since the structures obtained were homomer complex structures, only the monomer chain was selected and rest including water and non-bonded atoms were removed using Accelyrs Viewer lite 5.0 (VLifeMDS). The combinatorial library compounds with good predicted activity were selected and prepared using Ligprep. Schrodinger offers a comprehensive protein preparation wizard which ensures chemical correctness and also optimizes the protein. In protein preparation wizard, hydrogen ions were added, disulphide bonds were created, selinomethionine was converted to methionine, terminal residues were capped and bond lengths were optimized.

### 4.6.1 Receptor grid generation

A Glide scoring grid around the receptor was generated using receptor grid generation platform of Schrodinger's Glide modules (Friesner *et al*., 2004; Halgren *et al*., 2004).This utility of Glide defines receptor structure, determines and mark active site position. All the parameters were kept default and a grid of size $20 \times 20 \times 20$ Å with inner box size of $10 \times 10 \times 10$ Å was generated.

### 4.6.2 Docking and scoring

The prepared combinatorial library compounds were docked against NA of H1N1 and H3N2 using extra precision GlideXP platform. The selected poses were further minimized on pre-computed OPLS-2005 electrostatic and van der Waals grid for receptor. Ultimately lowest energy poses were subjected to Monte Carlo minimization and rescored using Glide Score function. The complexes with least XP score (highest magnitude) were selected for molecular dynamics simulations.

### 4.7 Molecular dynamics simulations

Docked complex of protein and ligand were prepared in protein preparation wizard of maestro. Desmond software was then used to study the molecular dynamics of ligand inside the active site of NA for both H1N1 and H3N2 using the Optimized Potentials for Liquid Simulations 2005 (OLPS) force field (Jorgensen *et al*., 2005). Structures were uploaded in Desmond for further process of molecular dynamics simulations. Proteins were solvated in explicit water, employing periodic boundary condition in a cubic box using TIP4P water model(Jorgensen *et al*., 1983). Distance between edge of the box and protein complex was set greater than 10 Å to avoid direct interaction with its own image. Steepest descent method was used for energy minimization of the prepared systems using up to maximum 5000 steps until gradient threshold (5Kcal/mol/Å) was achieved followed by Loe-memory broyden-fletcher-Goldfarb-Shanno quasi-Newtoninan minimizer (L-BFGS) until a threshold of 1kal/mol/Å was attained. The system were equilibrated using default parameters given in Desmond. MD simulations were performed with periodic boundary conditions in NPT ensemble. A constant temperature of 300 K for desired period of time and a constant pressure of 1 atm for a time step of 2fs. Long distance electrostatic interactions were calculated simulation smooth particle mesh method whereas Cutoff method was selected to define the short range electrostatic interactions. A cutoff of 9 Å radius (default), was used. The docked complexes were then simulated for 15ns using above parameters. Frames of trajectory were recorded for each 10 ns time step. The root mean square deviations (RMSD) for the docked complexes were calculated for the entire simulations trajectory with reference to their respective frames.

### 4.8 ADME Prediction

Absorption, distribution, metabolism and elimination (ADME) of the selected compounds were predicted *in silico* using QikProp module of Schrödinger suite (Jorgensen *et al*., 2002). Ligands were initially prepared using LigPrep. It predicts physically significant descriptors and relevant pharmaceutical properties. In addition to the molecular descriptors, QikProp also provides their range values by comparing an individual molecule property with those known 95% drug.

# 5. RESULT

**5.1 Separation of data into training and test set**

A QSAR model was developed for acylguanidine zanamivir derivatives (Lin *et al.*, 2013) considering the activity and various physiochemical descriptors for both H1N1 and H3N2. 70% of total compounds were trained in training set and the rest were selected as test set. A Unicolumn statistics (Table 2 and Table 3) was generated to check the correctness of the selection criteria. Biological activities of all the compounds in training and test sets was found to follow the pattern of unicolumn statistics in which maximum of test should be less than maximum of training set and minimum of test should be greater than minimum of training set (Jain *et al.*, 2011 )

Table 2:- Unicolumn statistics for training and test sets for *influenza* H1N1 Neuraminidase inhibitory activity.

| Data set | Average | Max. | Min. | Std dev | Sum |
|----------|---------|------|------|---------|-----|
| Training | -2.4963 | -1.3032 | -4.5955 | 0.6975 | -39.9406 |
| Test | -2.5855 | -1.7396 | -4.5396 | 0.8352 | -20.6838 |

Table 3:- Unicolumn statistics for training and test sets for *influenza* H3N2 Neuraminidase inhibitory activity.

| Data set | Average | Max. | Min. | Std dev | Sum |
|----------|---------|------|------|---------|-----|
| Training | -2.5530 | -1.7657 | -4.4713 | 0.6407 | -40.8485 |
| Test | -2.5821 | -1.4065 | -4.5832 | 0.9057 | -20.6564 |

**5.2 Analysis of GQSAR models developed against H1N1 and H3N2**

A robust GQSAR model was developed which explained correlation between the physiochemical parameters and contribution of each substitution site. Several models were developed and the best model with significant values based on statistical parameters was chosen.

### 5.2.1 H1N1 Model

The chosen model for H1N1 exhibited significant statistical values of $r^2 = 0.95$, $q^2 = 0.90$, Pred_$r^2 = 0.95$, F-Test=92.99 while standard errors were observed to be $r^2\_se = 0.15$, $q^2\_se = 0.23$, Pred_$r^2$se = 0.18. Low standard error values indicated absolute quality of the model.

Three descriptors namely R1-SdOEindex, R1-6ChainCount and R1-SssSE-index were selected by the model for all the compounds. The model had good internal and external prediction. The model can be given by the equation:

$$pIC50 = (23.61 * R1 - SdOEindex) + (47.12 * R1 - 6ChainCount) - (39.90 * R1 - (SssSEindex) - 5.26.$$

With n= 16, degree of freedom= 12, ZScore $R^2$ =3.35, ZScore $Q^2$= 0.69, "n" represents total number of compounds in the training set. The derived model shows a good correlation between aforementioned descriptors and biological activity as the derived coefficient of correlation is 0.95 with minimum standard error of 0.15. The model incorporates various descriptors as shown in Table 4.

Table 4:- Physicochemical descriptors with predicted activity values for training and test set for first model.

| Column | R1-SdOE-index | R1-SaaaCE-index | R1-SdsCHcount | R1-chiV4 | Prediction |
|--------|---------------|-----------------|---------------|----------|------------|
| 1186 | 17.51 | 0 | 0 | 0 | -0.823 |
| 1185 | 17.200 | 0 | 0 | 0 | -0.894 |
| 1184 | 16.25 | 0 | 0 | 0 | -1.112 |

R1-SdOEindex which is an electro-topological descriptor indices gives information about number of –OH groups connected with one double bond. The positive contribution of 58.02% (Figure 3) indicates that presence of –OH group increases the inhibitory activity of the NA inhibitors. R1-6ChainCount is one of the most influential descriptors which signify the total number of six-membered rings in a compound. Thus a positive contribution of 28.93% indicates presence of aromatic compounds like phenyl that could improve the inhibitory potency of compounds targeting NA. R1-SssSEindex, another statistically significant descriptor, shows the importance of electronic environment of sulfur atom bonded with two single non hydrogen atoms in the molecule. A negative contribution value of 13.04% suggests decrease in E-state contribution of sulfur atom in either aromatic ring or free sulfur attached to two non-hydrogen atom in a molecule that could improve the inhibitory activity. Thus, it can be deduced that the model is reliable and predictive, which can also be seen in the line graph of observed vs. predicted activity and also the radar plots of observed and predicted activity for both training and test set.
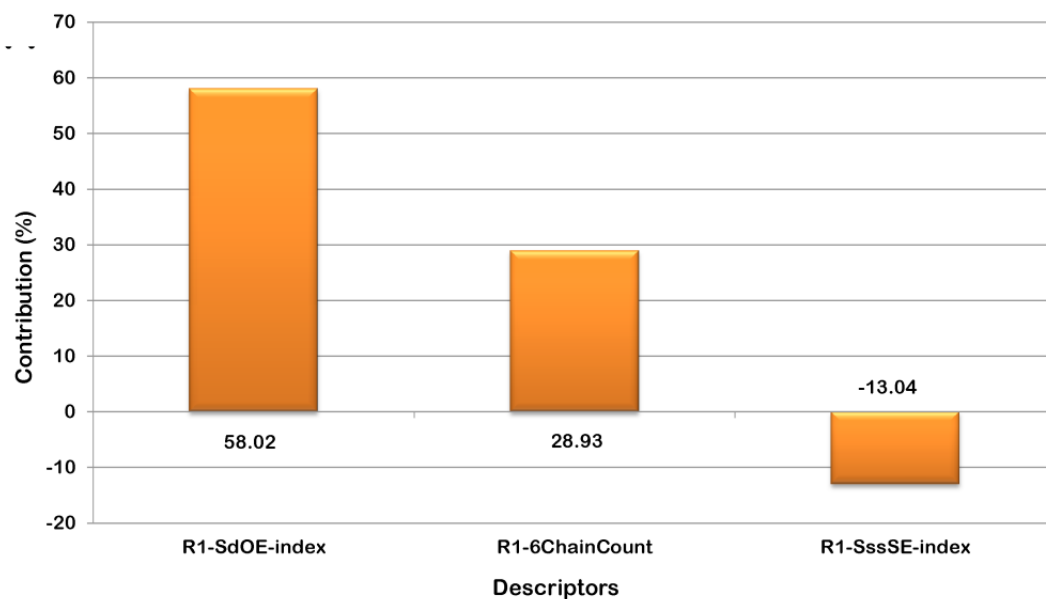
Figure 3:- Descriptor contribution plot of H1N1 GQSAR model in which R1-SdOEindex and R1-6ChainCount have a positive contribution of 58.02% and 28.93% while R1-SssSEindex has a negative contribution value of 13.04%
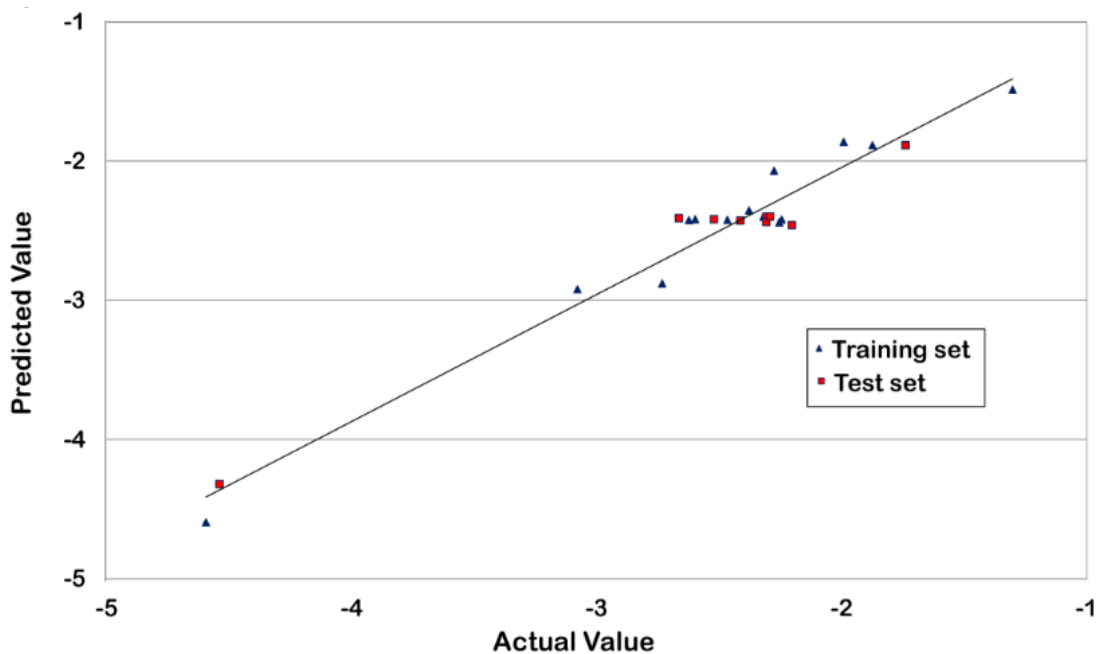


Figure 4: Graph showing the observed vs. predicted activity for training and test set
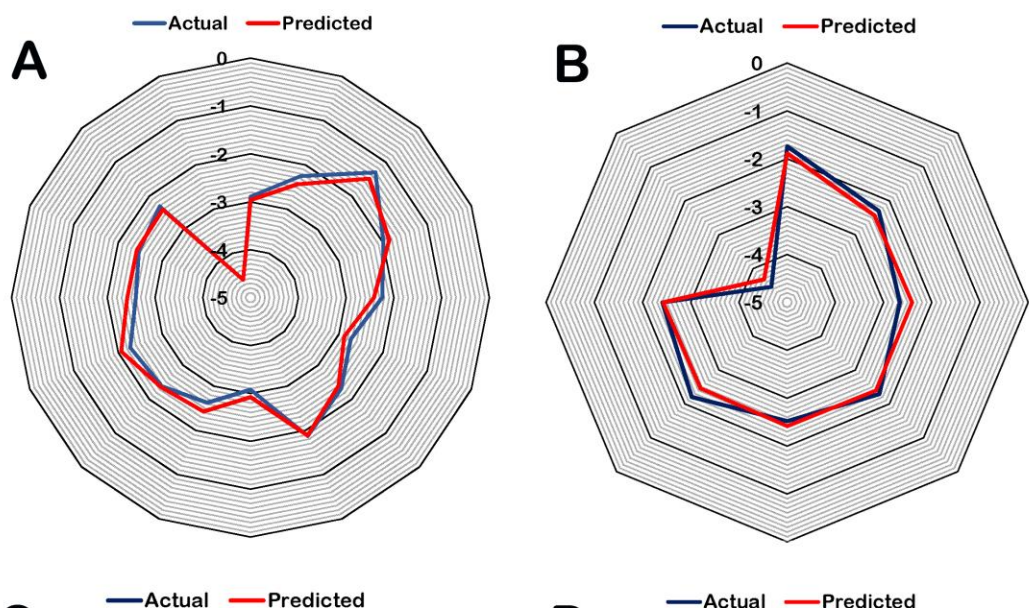
Figure 5:- Radar plots showing observed and predicted values of (a) training set and (b) test set for H1N1.

### 5.2.2 H3N2 Model

The model developed against H3N2 showed values of $r^2 = 0.95$, $q^2 = 0.93$, Pred_$r^2 = 0.87$ and F-test = 61.02 while the standard errors had values: $r^2$_se = 0.15, $q^2$_se = 0.19, Pred_$r^2$_se = 0.32. A line graph of observed vs. predicted activity is shown in Figure 6.
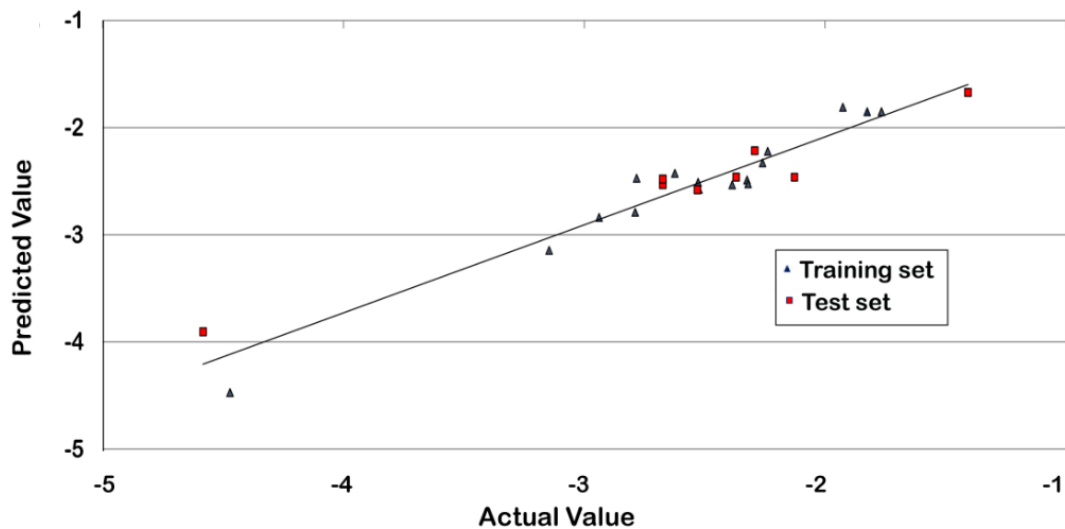


Figure 6:- Graph of observed vs. predicted activity for training and test set of H3N2

Low standard error and high values of internal and external prediction indicate robustness of the model. Thus, it can be inferred that the model is reliable and predictive, which can also be seen in the radar plots of the observed and predicted activity for both training and test set. Four descriptors were selected for model namely R1-SdOEindex, R1-SaaaCEindex, R1-SdsCHcount, R1-chiV4.The developed model had a good internal as well as external prediction. The model can be explained via equation:

$$pIC50 = (22.90 * R1 - SdOEindex) + (20.31 * R1 - SaaaCE\_index) - (25.88 * R1 - SdsCHcount) + (26.58 * R1 - chiV4) - 4.83$$

with n= 16, degree of freedom= 11, ZScore R^2 = 5.94, ZScore Q^2= 0.71, "n" represents total number of compounds in the training set.

The equation obtained above contains three physicochemical descriptors as shown in Table 5.

Table 5:- Physicochemical descriptors with predicted activity values for training and test set for second model.

| Column | R1-SdOE-index | R1-6ChainCount | R1-SssSE-index | Prediction |
|--------|---------------|----------------|----------------|------------|
| 1186   | 17.51         | 0              | 0              | -1.1278    |
| 1185   | 17.20         | 0              | 0              | -1.2019    |
| 1189   | 13.03         | 2              | 0              | -1.2442    |

R1-SdOEindex gives information about number of –OH groups connected with one double bond. The positive contribution of 54.91% (Figure 7) indicates that presence of –OH group increases the inhibitory activity of the NA inhibitors.R1-SaaaCEindex which is an electro topological descriptor which indicates the number of carbon atoms that are connected with three aromatic bonds. A positive contribution indicates that increase in SaaaCE properties would enhance the inhibitory effect of lead compound while R1-SdsCHcount descriptor highlights total number of –CH groups connected with one double and one single bond. Negative contribution of 15.18% indicates that increase in length of -CH atoms at the substitution site of NA inhibitors could be detrimental to the inhibitory activity. R1-chiV4 is a steric property descriptor that helps in discrTiminating molecules according to size, degree of branching, shape and overall flexibility. A positive contribution of 11.27% this descriptor indicates that increasing the steric properties at R1 will account for increased inhibitory activity.
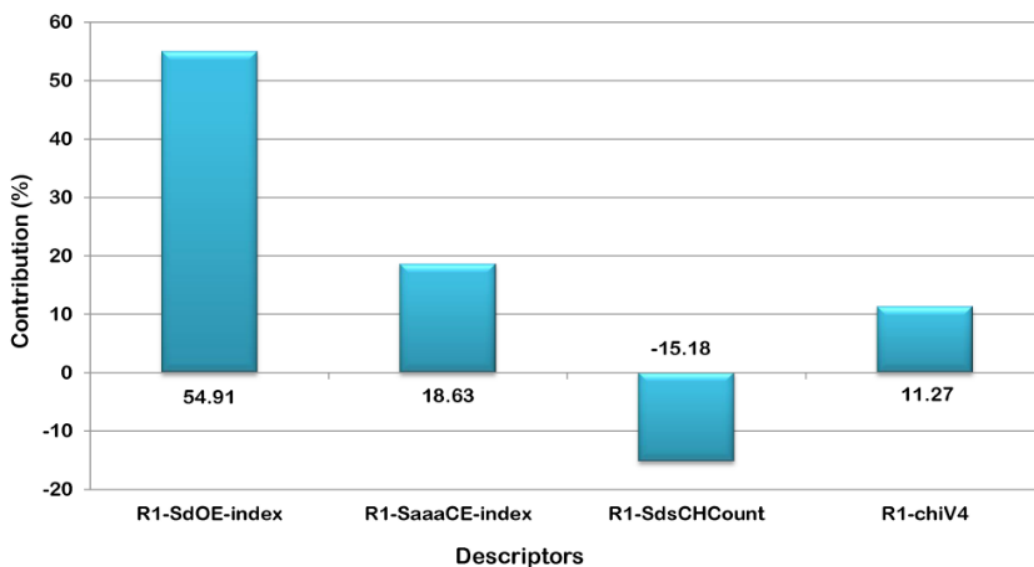
Figure 7: Descriptor contribution plot of H3N2 GQSAR model in which R1-SdOEindex, R1-SaaaCEinde, R1-chiV4 have a positive contribution of 54.91%, 18.63%, 11.27% while R1-SdsCHcount have a negative contribution of 15.18% .
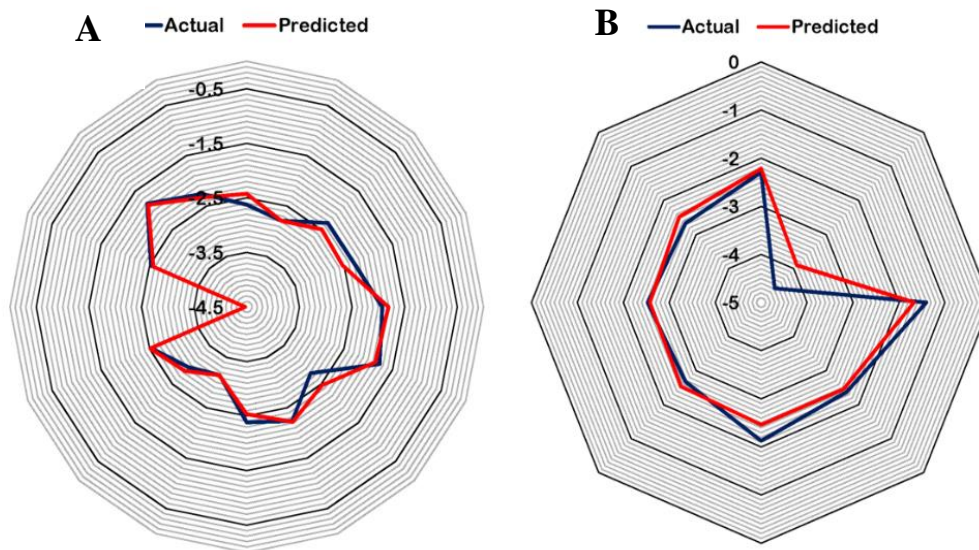


Figure 8:- Radar plots showing observed and predicted values of (a) training set and (b) test set for H3N2

## 5.3 Combinatorial Library analysis and selection of lead compound

Combinatorial library was generated after analyzing the above two models and their activities were predicted. Various substituting groups like alkanes, atoms, aromatic rings, ketone, ester etc. were added. The developed library contained 189 molecules. Molecules having activity values more than that reported in congeneric series were selected and the compound having highest predicted activity was chosen as lead compounds.It was seen that lead compound (Figure 11) was substituted with sulphite group (–s=ooo) at R1 position and had good predicted activity value of for H1N1 and H3N2. Docking studies were performed on lead compound and further molecular dynamics was also performed to check its stability in aqueous environment. The IUPAC name of the compound is (2R,3R,4S)-3-acetamido-4-{[(sulfoamino)methanimidoyl]amino}-2-[(1R,2R)-1,2,3-trihydroxypropyl]-3,4-dihydro-2H-pyran-6-carboxylic acid (AMA)
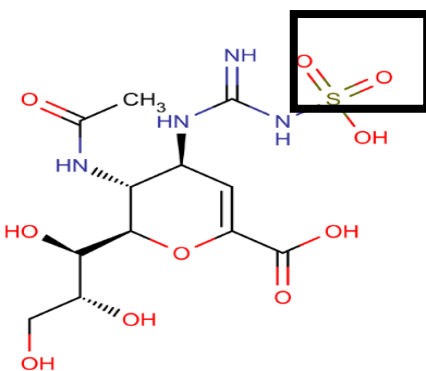


Figure 9:- Designed novel lead compound AMA with sulphite group at the substitution site.

## 5.4 Docking and Molecular Dynamics simulations

Docking study of the top scored compound was performed using Glide to study the interaction with crystal structure of H1N1 and H3N2. Docking of lead compound with H1N1 was performed and the binding energy of the compound was found to be -8.26 Kcal/mol. Weak bonding interactions like hydrophobic and hydrogen bonds are vital parameters that stabilize interactions between ligand and protein. AMA formed various hydrogen bonds with protein via residues namely Arg152, Arg156, Trp178, Glu277, Asn294, Arg371, Arg292 (Figure 10(a)). It also made hydrophobic interactions with nonpolar protein residues viz. Glu119, Asp151, Ser 179, Arg224, Glu227 Ser246, Glu276, Asn347 and Tyr406 (Figure 10(b)). Post-MD simulations AMA was found to form hydrogen bonds with residues Arg156, Asn294, Glu227, Arg371, Tyr406 (Figure 10(c)) and hydrophobic interactions were made with Glu119, Asp 151, Agr152, Trp178 and Ser179 (Figure 10(d)).
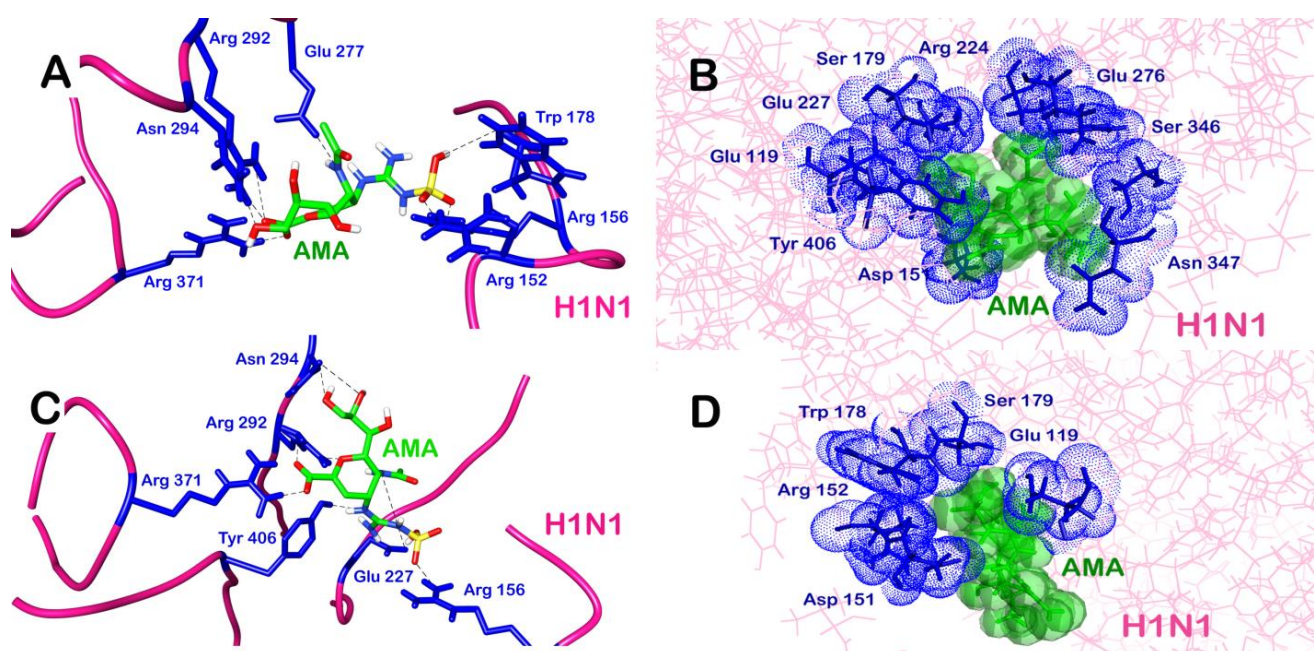


Figure 10: Molecular interactions of H1N1 Neuraminidase (pink) with AMA (green) depicting (a) hydrogen bond before MD simulations and (b) hydrophobic interactions before MD simulations. (c) Hydrogen bond after MD simulations and (d) hydrophobic interactions after MD simulations.

The same lead compound when docked against H3N2 showed different bonding patterns and binding energy. The compound when docked had a binding energy of -7.00 Kcal/mol. It made hydrogen bonds with Arg118, Glu119, Arg371, Asp151 and Arg292. Hydrophobic interactions were also made via residues Val 149, Tyr 406, Arg430, and Lys431. Different hydrogen bonding and hydrophobic interactions were observed post-MD simulations. AMA formed hydrogen bonds with protein residues Lys431 and Glu432 while hydrophobic interactions with Val149, Arg292, Arg371, Arg403 and Arg430.
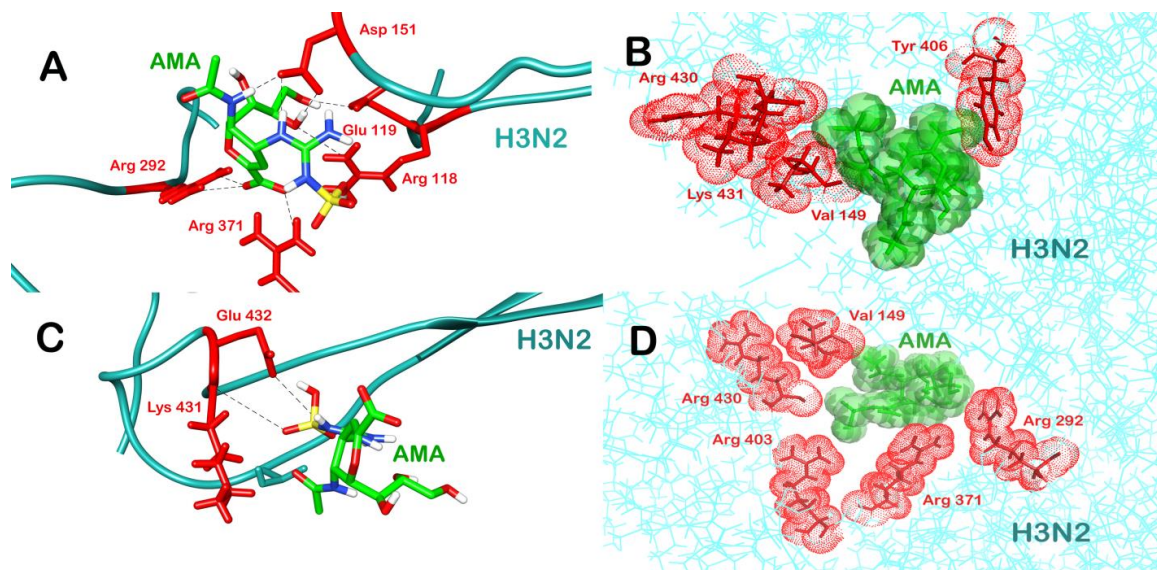


Figure 11: Molecular interactions of H1N1 Neuraminidase (pink) with AMA (green) depicting (a) hydrogen bond before MD simulations and (b) hydrophobic interactions before MD simulations. (c) Hydrogen bond after MD simulations and (d) hydrophobic interactions after MD simulation.

Molecular dynamics study was performed on this lead compound and RMSD was recorded for first 15 nanoseconds to study fluctuations and conformational changes in protein which gives a measure of the stability of protein in vivo. For both H1N1 and H3N2, the chosen lead compound was found to be stable after 4 and 8 ns respectively. This implied that protein underwent significant structural changes during initial stages and was stable for later stage during simulation.
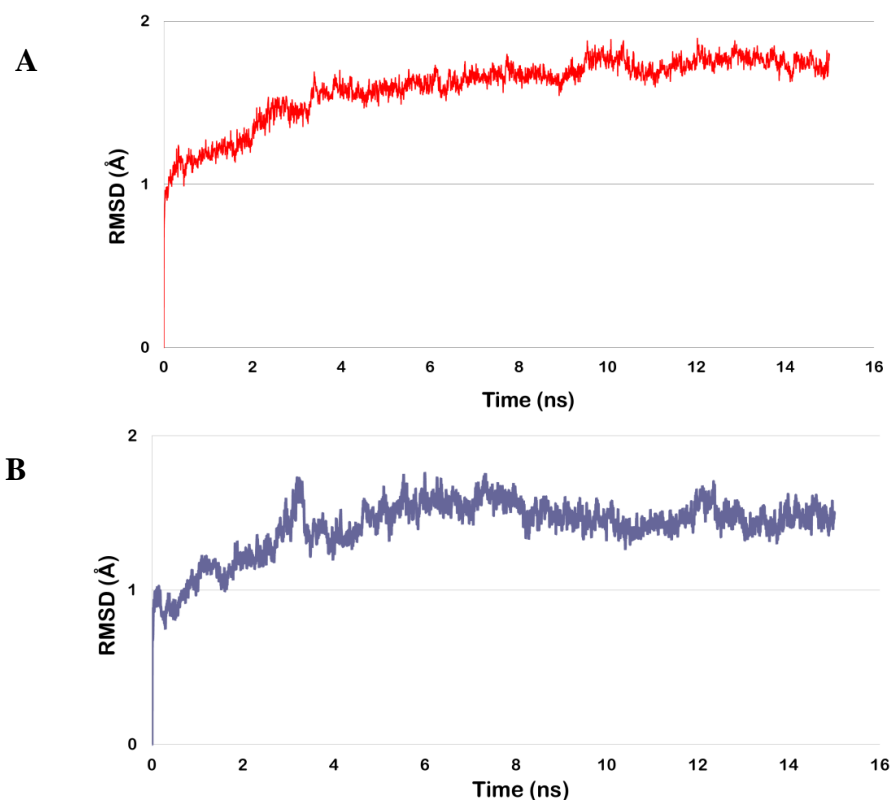


Figure 12:- RMSD plot of molecular dynamics simulations of lead compound against NA of (a) H1N1 (b) H3N2

## 5.5 ADME Value Prediction

ADME properties were predicted using QikProp program (Schrodinger Inc). Lipinski rule of five was used to calculate the drug like properties. It was found that AMA, highest scoring molecule followed three rule of Lipinski filter. Various descriptors were evaluated for ADMET properties. The range values for each descriptors were given based on the known values of 95% of drugs. Molecular weight of AMA was found to be 412.371(ideal molecular weight 130.0 - 725.0). Descriptors considered for drug permeability includes molecular volume of solute, hydrogen bond acceptor and liophilicity. Molecular volume of the compound was found to be 1107.43 (range value 500.00-2000.00) while hydrogen accepter was found to be 12.850 (range value 2.0 – 20.0). The latter parameter estimates average number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Rotatable bond count is

one of the widely used descriptor that inversely correlates with oral bioavailability. Rotatable bonds of this compound had value of 12.00 (Range 0.00-15.00). Various computational parameters were also calculated to analyze the solubility from its 2-D structure. Solvent accessible surface area (SASA) is one influential parameter, it defines the surface area of biomolecule that can be accessed by solvent. It is usually performed using 14 Å radius which generates various components *viz.* total solute accessible area (SASA) whose value was found to be 618.80 (Range values 300.00-1000.00), solute hydrophobic SASA (FOSA) with value 154.12 (Range value 0.00-750.00), solute Carbon Pi SASA (PISA) whose value was found to be 25.71 (range value 0.00-450.00) and solute weekly polar SASA (WPSA), which includes surface area for all sulphur, halogens and phosphorous atom. Its value was found to be 2.00 (Range value 0.00-175.00). Distribution of lead compound in solution is calculated using the ionization potential parameter which affects the availability of the compound for further physical, chemical or biological reactions. The calculated descriptor Solute Ionization Potential (eV) was found to be 9.672 (Range value 7.90 -10.50). Various other electrochemical descriptors like Solute Globularity (Sphere = 1) = 0.837 ( 0.75 / 0.95, Solute Electron Affinity (eV) = 1.262 ( -0.9 / 1.7) were also calculated. This lead compound was found to be similar to various compounds like Voglibose 68.40, Valganciclovir 68.04, Aminopterin 65.66, Lisinopril 64.63 and Methotrexate 64.39. All these above parameters suggest that AMA can be a potential drug molecule and a good lead candidate.

# 6. DISCUSSION

In this study we developed a novel GQSAR model based on congeneric series of acylguanidine zanamivir derivatives (Lin *et al.,* 2013). The QSAR analysis helped to derive highly applicable models that permitted the design of novel and reactive molecules. Preliminary structure–activity relationship (SAR) studies indicated that the Zanamivir moiety is required for inhibition of NA activity. As Neuraminidase (NA) plays a key role in the release of virion particles inhibition of NA has emerged as one of the most promising strategies for the treatment of Influenza. Zanamivir derivatives were therefore built and screened virtually using the docking method. The main purpose of our study was to develop a robust GQSAR model to identify relation between structure and biological activity of the set of zanamivir derivatives as a function of fragments done at substitution site. Developed model predicted the relationship between anti-influenza activity and electro-chemical properties of the derivatives with high efficiency. Same set of congeneric series were counter screened against NA of both H1N1 and H3N2.

**QSAR study**

Three-dimensional structures were built for 24 compounds and optimized in a V Life working environment. Molecular dynamics simulations were carried out for each molecule and molecular descriptors were determined followed by QSAR linear regression study. The correlation model showed a linear plot for both measured IC50 values and values predicted for the training set of molecules. The correlation analysis showed an RMSD of $r^2$ value of 0.95 for both model. This linearity for the model indicates the reliability of QSAR to compare and predict the activity of a test set molecules. Molecular modeling has often been proven to be a powerful tool for rationalizing ligand-target interactions and for making this information available to virtual screening techniques. Molecular modeling studies were performed using neuraminidase of H1N1 and H3N2, since they represent the pharmacological target for the development of new drugs for the treatment of influenza. The binding site of NA was identified from literature (Li *et al*., 2013).

This model was used to generate a combinatorial library of molecules. Library had 189 compounds. The activities of these compounds were predicted using the generated QSAR model. The docking studies performed validated the strong binding affinity of NA with these inhibitors. The molecular dynamics studies confirmed that the protein-ligand complex was stable in vivo. All compounds showed proper drug-like and ADMET properties.

**Molecular Descriptors**

Two models were generated targeting NA of both H1N1 and H3N2 influenza strains. The developed model generated various descriptors. R1-SdOEindex and R1-6ChainCount have a positive contribution of 58.02% and 28.93% and R1-SssSEindex has a negative contribution value of 13.04% for H1N1 model while R1-SdOEindex, R1-SaaaCEinde, R1-chiV4 have a positive contribution of 54.91%, 18.63%, 11.27% and R1-SdsCHcount have a negative contribution of 15.18%. A positive contribution suggests increase in contribution of that

descriptor could be beneficial for inhibitory activity while a negative contribution indicates that those descriptors are detrimental for inhibitory activity.

## Molecular Docking

Top three compounds were selected from the combinatorial libraries above a threshold predicted value of -1.3 and -1.4 nM respectively for H1N1 and H3N2.These compounds were further taken for docking studies. The studies were performed using Glide module of Maestro, a Schrodinger project. AMA was docked against the active site of NA and a satisfactory docking score of $-8.26$ and $-7.00$ Kcal/mol was observed for H1N1and H3N2 respectively.

## Molecular Dynamics

Molecular dynamics study was performed to study the stability of ligand protein complex in vivo. AMA was found to form hydrogen bonds with residues Arg156, Asn294, Glu227, Arg371, Tyr406 and hydrophobic interactions were made with Glu119, Asp 151, Agr152, Trp178 and Ser179 with NA of H1N1 while AMA formed hydrogen bonds with Lys431 and Glu432 and hydrophobic interactions with Val149, Arg292, Arg371, Arg403 and Arg430 with NA of H3N2.The RMSD shows the stability of the complex, AMA was stable from 6 to 15 ns.

## ADME predictions

ADME properties are important conditions and major parts of pharmacokinetics. Viable drugs should have perfect ADME properties for it to be approved as a drug in clinical tests. The ADME predictions of compounds show satisfactory results. Generally, the degree to which any drug binds to plasma protein influences not only the drug action but also its disposition and efficacy. Usually, the drug that is unbound to plasma proteins will be available for diffusion or transport across cell membranes and thereby finally interact with the target. Herein with respect to ADME, the percent of drug bound with plasma proteins was predicted and the compounds were predicted to bind strongly. It therefore seems that by combining all these evaluation procedures, large numbers of new compounds can be screened as possible inhibitors of NA. Since the present results indicate that AMA should be excellent candidates for inhibition of NA, they can be used as a starting point for developing even more potent analogues for the treatment of the Influenza disease**.**

# 7. CONCLUSION

A GQSAR model was generated for a congeneric series of zanamivir derivatives having inhibitory activity against NA. The model was generated using statistical method of multiple regression. The statistical measures r2, q2, F-test and standard error for the training set and the pred_r2 for the test set fulfilled the conditions for a model to be considered robust and predictive. The developed model was used to predict the activity values for a large set of compounds generated by substituting different fragments on the scaffold. We also developed one novel inhibitor (AMA) using the combinatorial library approach which displayed substantial binding affinity for NA in both H1N1 and H3N2 pandemic influenza strains. AMA had a better affinity for NA in comparison to the most potent compound of the congeneric series with pIC50 of -1.3 and -1.4 respectively for H1N1 and H3N2 as observed from the interaction pattern between these compounds and NA. Since AMA action is restricted only to virus, the inhibitors would be non-toxic to humans. Complex structure of ligand and inhibitor was found to be energetically stable post MD Simulations. Thus this provides evidence that the novel compound could serve as potent anti-influenza drugs with improved binding properties and low IC50 values than traditional drug. The present study provides substantial evidence for considering these compounds as prospective lead inhibitor against Influenza having enhanced inhibitory Neuraminidase activity which not only provides graphical results but also has the ability to forecast the activity or potency of compounds being considered for inhibition of target protein. As QSAR approach already plays an important role in lead structure optimization, it is anticipated that it will soon become essential for handling large amount of data generated using combinatorial chemistry.

# 8. REFERENCES

- Afantitis, A., Melagraki, G., Sarimveis, H., Igglessi-Markopoulou, O., Kollias, G. (2009). A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-aryl-[1,4] diazepane ureas. European Journal of Medicinal Chemistry. **44**(2), 877-884.

- Alonso-Caplen, F. V., Nemeroff, M. E., Qiu, Y., Krug, R. M. (1992). Nucleocytoplasmic transport: the influenza virus NS1 protein regulates the transport of spliced NS2 mRNA and its precursor NS1 mRNA. Genes Dev. **6**(2), 255-267.

- Amaro, R. E., Swift, R. V., Votapka, L., Li, W. W., Walker, R. C., & Bush, R. M. (2011). Mechanism of 150-cavity formation in influenza neuraminidase. Nat Commun. **2**, 388.

- Bertram, S., Glowacka, I., Steffen, I., Kuhl, A., & Pohlmann, S. (2010). Novel insights into proteolytic cleavage of influenza virus hemagglutinin. Rev Med Viro. **20**(5), 298-310.

- Biswas, S. K., & Nayak, D. P. (1994). Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J Virol. **68**(3), 1819-1826.

- Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. Vaccine. **25**(39–40), 6852-6862.

- Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Yewdell, J. W. (2001). A novel influenza A virus mitochondrial protein that induces cell death. Nat Med. **7**(12), 1306-1312.

- Colacino, J. M., Staschke, K. A., Laver, W. G. (1999). Approaches and strategies for the treatment of influenza virus infections. Antivir Chem Chemother.**10**(4), 155-85.

- Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T., Palese, P. (2007). A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. PLoS Pathog. **3**(10), 1414-1421.

- Cox, N. J., & Subbarao, K. (2000). Global epidemiology of influenza: past and present. Annu Rev Med. **51**, 407-421.

- Davies, W. L., Grunert, R. R., Haff, R. F., McGahen, J. W., Neumayer, E. M., Paulshock, M., Hoffmann, C. E. (1964). Antiviral Activity of 1-Adamantanamine (Amantadine). Science. **144**(3620), 862-863.

- Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., Digard, P. (2001). Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. J Virol. **75**(1), 408-419.

- Fechter, P., Mingay, L., Sharps, J., Chambers, A., Fodor, E., & Brownlee, G. G. (2003). Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. J Biol Chem. **278**(22), 20381-20388.

- Fodor, E., & Smith, M. (2004). The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. J Virol. **78**(17), 9144-9153.

- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T.,Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. **47**(7), 1739-1749.

- Golbraikh, A., & Tropsha, A. (2000). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. Molecular Diversity. **5**(4), 231-243.

- Golbraikh, A., & Tropsha, A. (2002). Beware of q2! Journal of Molecular Graphics and Modelling. **20**(4), 269-276.

- Gong, J., Xu, W., Zhang, J. (2007). Structure and functions of influenza virus neuraminidase. Curr Med Chem. **14**(1), 113-122.

- Gubareva, L. V., Kaiser, L.,Hayden, F. G. (2000). Influenza virus neuraminidase inhibitors. The Lancet. **355**(9206), 827-835.

- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T.,Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. **47**(7), 1750-1759.

- Hay, A. J., Wolstenholme, A. J., Skehel, J. J., Smith, M. H. (1985). The molecular basis of the specific anti-influenza action of amantadine. EMBO J. **4**(11), 3021-3024.

- Ito, T., Goto, H., Yamamoto, E., Tanaka, H., Takeuchi, M., Kuwayama, M.,Otsuki, K. (2001). Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. J Virol. **75**(9), 4439-4443.

- Itzstein, M., & Thomson, R. (2009). Anti-Influenza Drugs: The Development of Sialidase Inhibitors. In H.-G. Kräusslich & R. Bartenschlager (Eds.). Antiviral Strategies. **189**(111-154).

- Jain, S., Vishwakarma, S., & Nayak, P. (2011). 3D QSAR analysis on pyrrolidine derivatives as DPP IV inhibitors. International Journal of Research in Pharmaceutical and Biomedical Sciences. **2**(3), 1021-1032.

- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. **79**(2), 926-935.

- Jorgensen, W. L., Duffy, E. M. (2002). Prediction of drug solubility from structure. Advanced Drug Delivery Reviews. **54**(3), 355-366.

- Jorgensen, W. L.,Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proceedings of the National Academy of Sciences of the United States of America. **102**(19), 6665-6670.

- Klumpp, K., Ruigrok, R. W., Baudin, F. (1997). Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. EMBO J. **16**(6), 1248-1257.

- Lamb, R. A., Choppin, P. W., Chanock, R. M., Lai, C. J. (1980). Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. Proc Natl Acad Sci U S A. **77**(4), 1857-1861.

- Li, K. S., Guan, Y., Wang, J., Smith, G. J., Xu, K. M., Duan, L., Peiris, J. S. (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature. **430**(6996), 209-213.

- Lin, C. H., Chang, T. C., Das, A., Fang, M. Y., Hung, H. C., Hsu, K. C., Lin, C. C. (2013). Synthesis of acylguanidine zanamivir derivatives as neuraminidase inhibitors and the evaluation of their bio-activities. Org Biomol Chem. **11**(24), 3943-3948.

- Martin, K., & Helenius, A. (1991). Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell. **67**(1), 117-130.

- McAuley, J. L., Hornung, F., Boyd, K. L., Smith, A. M., McKeon, R., Bennink, J., McCullers, J. A. (2007). Expression of the 1918 influenza A virus PB1-F2 enhances the pathogenesis of viral and secondary bacterial pneumonia. Cell Host Microbe. **2**(4), 240-249.

- Nath, S. T., Nayak, D. P. (1990). Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). Mol Cell Biol. **10**(8), 4139-4145.

- Nelson, M. I., & Holmes, E. C. (2007). The evolution of epidemic influenza. Nat Rev Genet. **8**(3), 196-205.

- Neumann, G., Brownlee, G. G., Fodor, E., & Kawaoka, Y. (2004). Orthomyxovirus Replication, Transcription, and Polyadenylation. In Y. Kawaoka (Ed.), Biology of Negative Strand RNA Viruses: The Power of Reverse Genetics. **283** (121-143)

- Neumann, G., Noda, T., & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature. **459**(7249), 931-939.

- Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., & Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. Nature. **439**(7075), 490-492.

- Pinto, L. H., Holsinger, L. J., Lamb, R. A. (1992). Influenza virus M2 protein has ion channel activity. Cell. **69**(3), 517-528.

- Roberts, L., Smith, W., Jorm, L., Patel, M., Douglas, R. M., McGilchrist, C. (2000). Effect of infection control measures on the frequency of upper respiratory infection in child care: a randomized, controlled trial. Pediatrics. **105**( 1), 738-742.

- Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., Skehel, J. J. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. Nature, **443**(7107), 45-49.

- Schnell, J. R., & Chou, J. J. (2008). Structure and mechanism of the M2 proton channel of influenza A virus. Nature. **451**(7178), 591-595.

- Singla, R. K., & Bhat G, V. (2010). QSAR model for predicting the fungicidal action of 1,2,4-triazole derivatives against Candida albicans. Journal of Enzyme Inhibition and Medicinal Chemistry. **25**(5), 696-701.

- Skehel, J. J., & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem. **69**, 531-569.

- Steinhauer, D. A. (1999). Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology. **258**(1), 1-20.

- Tang, Y., Zaitseva, F., Lamb, R. A., Pinto, L. H. (2002). The Gate of the Influenza Virus M2 Proton Channel Is Formed by a Single Tryptophan Residue. Journal of Biological Chemistry. **277**(42), 39880-39886.

- Wang, C., Lamb, R. A., Pinto, L. H. (1995) Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. Biophysical Journal. **69**(4), 1363-1371.

- Watanabe, K., Takizawa, N., Katoh, M., Hoshida, K., Kobayashi, N., Nagata, K. (2001). Inhibition of nuclear export of ribonucleoprotein complexes of influenza virus by leptomycin B. Virus Res. **77**(1), 31-42.

- Wiley, D. C., & Skehel, J. J. (1987). The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem. **56**, 365-394.

- Zambon, M. C. (1999). Epidemiology and pathogenesis of influenza. J Antimicrob Chemother. **44** Suppl B, 3-9.

- Zheng, W., & Tropsha, A. (2000). Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci. **40**(1), 185-194.

# 9. APPENDIX

**RMSD values of stable AMA-NA (H1N1) complex for 15ns**

| Time (ns) | RMSD of backbone atom |
|---|---|
| 0.0 | 0.00000000000000726 |
| 0.0192 | 0.873614107 |
| 0.0624 | 0.939648145 |
| 0.1056 | 0.987624134 |
| 0.1488 | 1.088898909 |
| 0.192 | 1.062972725 |
| 0.2352 | 1.139049394 |
| 0.2784 | 1.073387003 |
| 0.3216 | 1.161321768 |
| 0.3648 | 1.160302111 |
| 0.408 | 1.13145811 |
| 0.4512 | 0.988320641 |
| 0.4944 | 1.054241289 |
| 0.5376 | 1.166677523 |
| 0.5808 | 1.161502649 |
| 0.624 | 1.129360901 |
| 0.6672 | 1.115780151 |
| 0.7104 | 1.141330804 |
| 0.7536 | 1.10729932 |
| 0.7968 | 1.148664346 |
| 0.84 | 1.123882363 |
| 0.8832 | 1.216437868 |
| 0.9264 | 1.194160332 |
| 0.9696 | 1.204930048 |
| 1.0128 | 1.20488492 |
| 1.056 | 1.177890087 |
| 1.0992 | 1.222714368 |
| 1.1424 | 1.220246445 |
| 1.1856 | 1.225403048 |
| 1.2288 | 1.242393832 |
| 1.272 | 1.226678445 |
| 1.3152 | 1.215879851 |
| 1.3584 | 1.242109621 |
| 1.4016 | 1.202084902 |
| 1.4448 | 1.236674878 |
| 1.488 | 1.222798527 |
| 1.5312 | 1.268530065 |

| | |
|---|---|
| 1.5744 | 1.186265911 |
| 1.6176 | 1.218090473 |
| 1.6608 | 1.21079541 |
| 1.704 | 1.256650406 |
| 1.7472 | 1.204502258 |
| 1.7904 | 1.28290694 |
| 1.8336 | 1.272993386 |
| 1.8768 | 1.311398334 |
| 1.92 | 1.227716318 |
| 1.9632 | 1.271077982 |
| 2.0064 | 1.357378367 |
| 2.0496 | 1.310698306 |
| 2.0928 | 1.469732345 |
| 2.136 | 1.433396463 |
| 2.1792 | 1.334105187 |
| 2.2224 | 1.387334446 |
| 2.2656 | 1.427438177 |
| 2.3088 | 1.463410547 |
| 2.352 | 1.525612589 |
| 2.3952 | 1.397912415 |
| 2.4384 | 1.463960482 |
| 2.4816 | 1.42591618 |
| 2.5248 | 1.508151067 |
| 2.568 | 1.489614248 |
| 2.6112 | 1.475339989 |
| 2.6544 | 1.436702419 |
| 2.6976 | 1.530930345 |
| 2.7408 | 1.460757428 |
| 2.784 | 1.463105753 |
| 2.8272 | 1.3721594 |
| 2.8704 | 1.467310859 |
| 2.9136 | 1.408926708 |
| 2.9568 | 1.430862494 |
| 3 | 1.439162963 |
| 3.0432 | 1.405207705 |
| 3.0864 | 1.497661611 |
| 3.1296 | 1.446802638 |
| 3.1728 | 1.491934378 |
| 3.216 | 1.422598617 |
| 3.2592 | 1.42016069 |
| 3.3024 | 1.497293207 |

| | |
|---|---|
| 3.3456 | 1.52282989 |
| 3.3888 | 1.688350131 |
| 3.432 | 1.568986209 |
| 3.4752 | 1.544973642 |
| 3.5184 | 1.543768107 |
| 3.5616 | 1.582940566 |
| 3.6048 | 1.575978734 |
| 3.648 | 1.554438558 |
| 3.6912 | 1.521870485 |
| 3.7344 | 1.519812462 |
| 3.7776 | 1.524184166 |
| 3.8208 | 1.603749016 |
| 3.864 | 1.602475012 |
| 3.9072 | 1.616610446 |
| 3.9504 | 1.597446093 |
| 3.9936 | 1.578707779 |
| 4.0368 | 1.637973161 |
| 4.08 | 1.617569575 |
| 4.1232 | 1.529271025 |
| 4.1664 | 1.594615344 |
| 4.2096 | 1.601329056 |
| 4.2528 | 1.577590296 |
| 4.296 | 1.52184709 |
| 4.3392 | 1.565656647 |
| 4.3824 | 1.565182776 |
| 4.4256 | 1.562785 |
| 4.4688 | 1.510987425 |
| 4.512 | 1.466469012 |
| 4.5552 | 1.504673439 |
| 4.5984 | 1.551636248 |
| 4.6416 | 1.5694253 |
| 4.6848 | 1.569393166 |
| 4.728 | 1.598975276 |
| 4.7712 | 1.620211712 |
| 4.8144 | 1.608315493 |
| 4.8576 | 1.578966491 |
| 4.9008 | 1.608446258 |
| 4.944 | 1.540851377 |
| 4.9872 | 1.604861128 |
| 5.0304 | 1.61584513 |
| 5.0736 | 1.530788195 |

| | |
|---|---|
| 5.1168 | 1.619660911 |
| 5.16 | 1.62492368 |
| 5.2032 | 1.592163885 |
| 5.2464 | 1.577631252 |
| 5.2896 | 1.598783306 |
| 5.3328 | 1.563546872 |
| 5.376 | 1.549540684 |
| 5.4192 | 1.609912399 |
| 5.4624 | 1.584276137 |
| 5.5056 | 1.598458562 |
| 5.5488 | 1.61045727 |
| 5.592 | 1.632073401 |
| 5.6352 | 1.62294154 |
| 5.6784 | 1.619982987 |
| 5.7216 | 1.632702599 |
| 5.7648 | 1.636290082 |
| 5.808 | 1.625263254 |
| 5.8512 | 1.564275796 |
| 5.8944 | 1.636185959 |
| 5.9376 | 1.603817067 |
| 5.9808 | 1.594144464 |
| 6.024 | 1.662703401 |
| 6.0672 | 1.608444894 |
| 6.1104 | 1.657902504 |
| 6.1536 | 1.685531956 |
| 6.1968 | 1.603200074 |
| 6.24 | 1.531291487 |
| 6.2832 | 1.566567468 |
| 6.3264 | 1.570746419 |
| 6.3696 | 1.585061769 |
| 6.4128 | 1.630968349 |
| 6.456 | 1.664270133 |
| 6.4992 | 1.579620138 |
| 6.5424 | 1.689575833 |
| 6.5856 | 1.57540478 |
| 6.6288 | 1.642641038 |
| 6.672 | 1.615077598 |
| 6.7152 | 1.61109466 |
| 6.7584 | 1.62287358 |
| 6.8016 | 1.650537091 |
| 6.8448 | 1.607112803 |

| | |
|---|---|
| 6.888 | 1.700402975 |
| 6.9312 | 1.621440294 |
| 6.9744 | 1.643775053 |
| 7.0176 | 1.617225237 |
| 7.0608 | 1.638515494 |
| 7.104 | 1.679237398 |
| 7.1472 | 1.661511791 |
| 7.1904 | 1.65991906 |
| 7.2336 | 1.646957854 |
| 7.2768 | 1.6509076 |
| 7.32 | 1.688834586 |
| 7.3632 | 1.676415016 |
| 7.4064 | 1.745003914 |
| 7.4496 | 1.663276809 |
| 7.4928 | 1.65714912 |
| 7.536 | 1.650595101 |
| 7.5792 | 1.680505474 |
| 7.6224 | 1.643772461 |
| 7.6656 | 1.663359371 |
| 7.7088 | 1.726491798 |
| 7.752 | 1.767296754 |
| 7.7952 | 1.652379681 |
| 7.8384 | 1.65474065 |
| 7.8816 | 1.718208182 |
| 7.9248 | 1.647365931 |
| 7.968 | 1.687234013 |
| 8.0112 | 1.72710503 |
| 8.0544 | 1.67727134 |
| 8.0976 | 1.644181872 |
| 8.1408 | 1.633501032 |
| 8.184 | 1.624477228 |
| 8.2272 | 1.666218925 |
| 8.2704 | 1.646079737 |
| 8.3136 | 1.706524473 |
| 8.3568 | 1.682767617 |
| 8.4 | 1.615649069 |
| 8.4432 | 1.705557724 |
| 8.4864 | 1.677570981 |
| 8.5296 | 1.631817648 |
| 8.5728 | 1.716378299 |
| 8.616 | 1.668549991 |

| | |
|---|---|
| 8.6592 | 1.678579517 |
| 8.7024 | 1.663968036 |
| 8.7456 | 1.679662026 |
| 8.7888 | 1.623664216 |
| 8.832 | 1.628565528 |
| 8.8752 | 1.612046063 |
| 8.9184 | 1.589387413 |
| 8.9616 | 1.586630292 |
| 9.0048 | 1.654388855 |
| 9.048 | 1.643483811 |
| 9.0912 | 1.632928127 |
| 9.1344 | 1.701211352 |
| 9.1776 | 1.765393262 |
| 9.2208 | 1.666621002 |
| 9.264 | 1.626157309 |
| 9.3072 | 1.655953107 |
| 9.3504 | 1.675763262 |
| 9.3936 | 1.725491695 |
| 9.4368 | 1.676260716 |
| 9.48 | 1.752372359 |
| 9.5232 | 1.78752399 |
| 9.5664 | 1.761855016 |
| 9.6096 | 1.815240355 |
| 9.6528 | 1.830058235 |
| 9.696 | 1.784583154 |
| 9.7392 | 1.7446123 |
| 9.7824 | 1.753147137 |
| 9.8256 | 1.766066306 |
| 9.8688 | 1.838084632 |
| 9.912 | 1.761560809 |
| 9.9552 | 1.722152278 |
| 9.9984 | 1.814218292 |
| 10.0416 | 1.798865824 |
| 10.0848 | 1.759099437 |
| 10.128 | 1.762080454 |
| 10.1712 | 1.754511643 |
| 10.2144 | 1.77718331 |
| 10.2576 | 1.776549144 |
| 10.3008 | 1.764974718 |
| 10.344 | 1.85680791 |
| 10.3872 | 1.738416394 |

| | |
|---|---|
| 10.4304 | 1.717116256 |
| 10.4736 | 1.677373911 |
| 10.5168 | 1.673885995 |
| 10.56 | 1.690654365 |
| 10.6032 | 1.716054139 |
| 10.6464 | 1.711331425 |
| 10.6896 | 1.673669811 |
| 10.7328 | 1.677698765 |
| 10.776 | 1.742993637 |
| 10.8192 | 1.670041059 |
| 10.8624 | 1.720181037 |
| 10.9056 | 1.690589308 |
| 10.9488 | 1.681789739 |
| 10.992 | 1.678451653 |
| 11.0352 | 1.680042275 |
| 11.0784 | 1.640942949 |
| 11.1216 | 1.693291742 |
| 11.1648 | 1.694720055 |
| 11.208 | 1.715913731 |
| 11.2512 | 1.710281059 |
| 11.2944 | 1.673256658 |
| 11.3376 | 1.725762382 |
| 11.3808 | 1.706146041 |
| 11.424 | 1.681898635 |
| 11.4672 | 1.812623514 |
| 11.5104 | 1.706285184 |
| 11.5536 | 1.762840281 |
| 11.5968 | 1.735703043 |
| 11.64 | 1.708541391 |
| 11.6832 | 1.66351682 |
| 11.7264 | 1.754779242 |
| 11.7696 | 1.695894127 |
| 11.8128 | 1.712509086 |
| 11.856 | 1.735788547 |
| 11.8992 | 1.72374845 |
| 11.9424 | 1.764961952 |
| 11.9856 | 1.704513858 |
| 12.0288 | 1.828607507 |
| 12.072 | 1.788744625 |
| 12.1152 | 1.810194894 |
| 12.1584 | 1.745465636 |

| | |
|---|---|
| 12.2016 | 1.710875267 |
| 12.2448 | 1.746609515 |
| 12.288 | 1.820276782 |
| 12.3312 | 1.79199895 |
| 12.3744 | 1.729051727 |
| 12.4176 | 1.733241269 |
| 12.4608 | 1.74699221 |
| 12.504 | 1.799734344 |
| 12.5472 | 1.788307937 |
| 12.5904 | 1.765126832 |
| 12.6336 | 1.758125042 |
| 12.6768 | 1.749707752 |
| 12.72 | 1.706622028 |
| 12.7632 | 1.741796375 |
| 12.8064 | 1.802716422 |
| 12.8496 | 1.851671848 |
| 12.8928 | 1.796916116 |
| 12.936 | 1.811252003 |
| 12.9792 | 1.785407686 |
| 13.0224 | 1.781649946 |
| 13.0656 | 1.803334724 |
| 13.1088 | 1.745512237 |
| 13.152 | 1.74181128 |
| 13.1952 | 1.822397098 |
| 13.2384 | 1.738178887 |
| 13.2816 | 1.803590024 |
| 13.3248 | 1.749927488 |
| 13.368 | 1.785422018 |
| 13.4112 | 1.754627479 |
| 13.4544 | 1.804634637 |
| 13.4976 | 1.77986497 |
| 13.5408 | 1.730055395 |
| 13.584 | 1.79800814 |
| 13.6272 | 1.739432137 |
| 13.6704 | 1.829467984 |
| 13.7136 | 1.762962434 |
| 13.7568 | 1.847845042 |
| 13.8 | 1.753771475 |
| 13.8432 | 1.72791281 |
| 13.8864 | 1.761007507 |
| 13.9296 | 1.785672892 |

| | |
|---|---|
| 13.9728 | 1.799647959 |
| 14.016 | 1.707984001 |
| 14.0592 | 1.697976756 |
| 14.1024 | 1.724073064 |
| 14.1456 | 1.652639842 |
| 14.1888 | 1.676695565 |
| 14.232 | 1.789525363 |
| 14.2752 | 1.750978319 |
| 14.3184 | 1.766499347 |
| 14.3616 | 1.767503133 |
| 14.4048 | 1.743555488 |
| 14.448 | 1.714632558 |
| 14.4912 | 1.75169264 |
| 14.5344 | 1.73702212 |
| 14.5776 | 1.810526328 |
| 14.6208 | 1.78546778 |
| 14.664 | 1.718767963 |
| 14.7072 | 1.727155123 |
| 14.7504 | 1.689796529 |
| 14.7936 | 1.761587226 |
| 14.8368 | 1.67595428 |
| 14.88 | 1.665071687 |
| 14.9232 | 1.711378191 |
| 14.9664 | 1.680956835 |
| 15.00 | 1.796979571 |

**RMSD values of stable AMA-NA (H3N2) complex for 15ns**

| Time(ns | RMSD of the backbone atom |
|---|---|
| 0.0 | 8.61440514654E-15 |
| 0.0192 | 0.709656 |
| 0.0624 | 0.992212 |
| 0.1056 | 0.969377 |
| 0.1488 | 1.026919 |
| 0.192 | 0.895113 |
| 0.2352 | 0.842084 |
| 0.2784 | 0.810291 |
| 0.3216 | 0.89773 |
| 0.3648 | 0.897431 |
| 0.408 | 0.841359 |
| 0.4512 | 0.869518 |
| 0.4944 | 0.877201 |
| 0.5376 | 0.89892 |
| 0.5808 | 0.904215 |
| 0.624 | 0.924042 |
| 0.6672 | 0.94409 |
| 0.7104 | 1.004793 |
| 0.7536 | 1.039002 |
| 0.7968 | 0.998231 |
| 0.84 | 1.059601 |
| 0.8832 | 1.131619 |
| 0.9264 | 1.050013 |
| 0.9696 | 1.049567 |
| 1.0128 | 1.041959 |
| 1.056 | 1.105437 |
| 1.0992 | 1.176281 |
| 1.1424 | 1.197385 |
| 1.1856 | 1.156682 |
| 1.2288 | 1.124695 |
| 1.272 | 1.136837 |
| 1.3152 | 1.136765 |
| 1.3584 | 1.180581 |
| 1.4016 | 1.173193 |
| 1.4448 | 1.152715 |
| 1.488 | 1.149063 |
| 1.5312 | 1.049723 |
| 1.5744 | 1.017701 |
| 1.6176 | 1.060565 |

| | |
|---|---|
| 1.6608 | 1.07788 |
| 1.704 | 1.144423 |
| 1.7472 | 1.151148 |
| 1.7904 | 1.189304 |
| 1.8336 | 1.128764 |
| 1.8768 | 1.263031 |
| 1.92 | 1.219648 |
| 1.9632 | 1.140071 |
| 2.0064 | 1.205144 |
| 2.0496 | 1.227513 |
| 2.0928 | 1.18538 |
| 2.136 | 1.295881 |
| 2.1792 | 1.156754 |
| 2.2224 | 1.217749 |
| 2.2656 | 1.172683 |
| 2.3088 | 1.357905 |
| 2.352 | 1.148766 |
| 2.3952 | 1.183689 |
| 2.4384 | 1.194965 |
| 2.4816 | 1.271098 |
| 2.5248 | 1.228593 |
| 2.568 | 1.230427 |
| 2.6112 | 1.248566 |
| 2.6544 | 1.277901 |
| 2.6976 | 1.304069 |
| 2.7408 | 1.329599 |
| 2.784 | 1.418971 |
| 2.8272 | 1.413848 |
| 2.8704 | 1.362266 |
| 2.9136 | 1.435936 |
| 2.9568 | 1.376291 |
| 3 | 1.455136 |
| 3.0432 | 1.326991 |
| 3.0864 | 1.517811 |
| 3.1296 | 1.544572 |
| 3.1728 | 1.646338 |
| 3.216 | 1.655141 |
| 3.2592 | 1.708748 |
| 3.3024 | 1.669382 |
| 3.3456 | 1.423574 |
| 3.3888 | 1.30205 |

| | |
|---|---|
| 3.432 | 1.376569 |
| 3.4752 | 1.230665 |
| 3.5184 | 1.276794 |
| 3.5616 | 1.455747 |
| 3.6048 | 1.35073 |
| 3.648 | 1.387281 |
| 3.6912 | 1.417196 |
| 3.7344 | 1.408031 |
| 3.7776 | 1.414364 |
| 3.8208 | 1.446937 |
| 3.864 | 1.316805 |
| 3.9072 | 1.288187 |
| 3.9504 | 1.256303 |
| 3.9936 | 1.3114 |
| 4.0368 | 1.338943 |
| 4.08 | 1.401929 |
| 4.1232 | 1.383385 |
| 4.1664 | 1.365978 |
| 4.2096 | 1.285024 |
| 4.2528 | 1.293707 |
| 4.296 | 1.340858 |
| 4.3392 | 1.439261 |
| 4.3824 | 1.336972 |
| 4.4256 | 1.310834 |
| 4.4688 | 1.384986 |
| 4.512 | 1.410181 |
| 4.5552 | 1.389562 |
| 4.5984 | 1.409602 |
| 4.6416 | 1.576264 |
| 4.6848 | 1.519271 |
| 4.728 | 1.453228 |
| 4.7712 | 1.502391 |
| 4.8144 | 1.50561 |
| 4.8576 | 1.53957 |
| 4.9008 | 1.462255 |
| 4.944 | 1.491809 |
| 4.9872 | 1.504337 |
| 5.0304 | 1.495184 |
| 5.0736 | 1.550255 |
| 5.1168 | 1.524927 |
| 5.16 | 1.567794 |

| | |
|--------|----------|
| 5.2032 | 1.396599 |
| 5.2464 | 1.531743 |
| 5.2896 | 1.52675 |
| 5.3328 | 1.506704 |
| 5.376 | 1.520256 |
| 5.4192 | 1.582136 |
| 5.4624 | 1.480867 |
| 5.5056 | 1.52657 |
| 5.5488 | 1.689535 |
| 5.592 | 1.537041 |
| 5.6352 | 1.581382 |
| 5.6784 | 1.545786 |
| 5.7216 | 1.538439 |
| 5.7648 | 1.539347 |
| 5.808 | 1.564916 |
| 5.8512 | 1.491635 |
| 5.8944 | 1.626867 |
| 5.9376 | 1.62051 |
| 5.9808 | 1.614126 |
| 6.024 | 1.548311 |
| 6.0672 | 1.539581 |
| 6.1104 | 1.552923 |
| 6.1536 | 1.595418 |
| 6.1968 | 1.545294 |
| 6.24 | 1.60833 |
| 6.2832 | 1.527135 |
| 6.3264 | 1.509198 |
| 6.3696 | 1.592705 |
| 6.4128 | 1.549387 |
| 6.456 | 1.582589 |
| 6.4992 | 1.525612 |
| 6.5424 | 1.507 |
| 6.5856 | 1.564213 |
| 6.6288 | 1.589757 |
| 6.672 | 1.637061 |
| 6.7152 | 1.586871 |
| 6.7584 | 1.511799 |
| 6.8016 | 1.560414 |
| 6.8448 | 1.502504 |
| 6.888 | 1.553591 |
| 6.9312 | 1.582912 |

| | |
|---|---|
| 6.9744 | 1.576327 |
| 7.0176 | 1.482554 |
| 7.0608 | 1.423675 |
| 7.104 | 1.609515 |
| 7.1472 | 1.53811 |
| 7.1904 | 1.622969 |
| 7.2336 | 1.587604 |
| 7.2768 | 1.708199 |
| 7.32 | 1.756216 |
| 7.3632 | 1.69992 |
| 7.4064 | 1.594302 |
| 7.4496 | 1.666819 |
| 7.4928 | 1.588702 |
| 7.536 | 1.596091 |
| 7.5792 | 1.562426 |
| 7.6224 | 1.675217 |
| 7.6656 | 1.579361 |
| 7.7088 | 1.600347 |
| 7.752 | 1.622772 |
| 7.7952 | 1.553817 |
| 7.8384 | 1.616175 |
| 7.8816 | 1.502378 |
| 7.9248 | 1.596826 |
| 7.968 | 1.654451 |
| 8.0112 | 1.562359 |
| 8.0544 | 1.505045 |
| 8.0976 | 1.542434 |
| 8.1408 | 1.546087 |
| 8.184 | 1.464263 |
| 8.2272 | 1.383284 |
| 8.2704 | 1.36801 |
| 8.3136 | 1.513042 |
| 8.3568 | 1.503767 |
| 8.4 | 1.497261 |
| 8.4432 | 1.513332 |
| 8.4864 | 1.455984 |
| 8.5296 | 1.440599 |
| 8.5728 | 1.590284 |
| 8.616 | 1.548446 |
| 8.6592 | 1.606799 |
| 8.7024 | 1.560618 |

| | |
|---|---|
| 8.7456 | 1.499093 |
| 8.7888 | 1.425986 |
| 8.832 | 1.440639 |
| 8.8752 | 1.362536 |
| 8.9184 | 1.427548 |
| 8.9616 | 1.449762 |
| 9.0048 | 1.384203 |
| 9.048 | 1.491938 |
| 9.0912 | 1.522893 |
| 9.1344 | 1.516129 |
| 9.1776 | 1.432614 |
| 9.2208 | 1.502176 |
| 9.264 | 1.475579 |
| 9.3072 | 1.462198 |
| 9.3504 | 1.410816 |
| 9.3936 | 1.516475 |
| 9.4368 | 1.49907 |
| 9.48 | 1.328766 |
| 9.5232 | 1.418762 |
| 9.5664 | 1.454273 |
| 9.6096 | 1.490314 |
| 9.6528 | 1.370008 |
| 9.696 | 1.450797 |
| 9.7392 | 1.386921 |
| 9.7824 | 1.498342 |
| 9.8256 | 1.431223 |
| 9.8688 | 1.421854 |
| 9.912 | 1.44311 |
| 9.9552 | 1.426387 |
| 9.9984 | 1.479705 |
| 10.0416 | 1.49778 |
| 10.0848 | 1.572563 |
| 10.128 | 1.438808 |
| 10.1712 | 1.350815 |
| 10.2144 | 1.41537 |
| 10.2576 | 1.355348 |
| 10.3008 | 1.430472 |
| 10.344 | 1.448543 |
| 10.3872 | 1.324944 |
| 10.4304 | 1.375273 |
| 10.4736 | 1.401504 |

| | |
|---------|----------|
| 10.5168 | 1.348931 |
| 10.56 | 1.334269 |
| 10.6032 | 1.373769 |
| 10.6464 | 1.463865 |
| 10.6896 | 1.338558 |
| 10.7328 | 1.363383 |
| 10.776 | 1.459674 |
| 10.8192 | 1.46591 |
| 10.8624 | 1.410954 |
| 10.9056 | 1.379859 |
| 10.9488 | 1.459426 |
| 10.992 | 1.431149 |
| 11.0352 | 1.385874 |
| 11.0784 | 1.459647 |
| 11.1216 | 1.463197 |
| 11.1648 | 1.457838 |
| 11.208 | 1.47768 |
| 11.2512 | 1.40039 |
| 11.2944 | 1.368511 |
| 11.3376 | 1.425544 |
| 11.3808 | 1.512207 |
| 11.424 | 1.354555 |
| 11.4672 | 1.396302 |
| 11.5104 | 1.475458 |
| 11.5536 | 1.345749 |
| 11.5968 | 1.481249 |
| 11.64 | 1.499606 |
| 11.6832 | 1.431692 |
| 11.7264 | 1.430522 |
| 11.7696 | 1.503571 |
| 11.8128 | 1.477345 |
| 11.856 | 1.465281 |
| 11.8992 | 1.446661 |
| 11.9424 | 1.613205 |
| 11.9856 | 1.569338 |
| 12.0288 | 1.554128 |
| 12.072 | 1.648947 |
| 12.1152 | 1.587505 |
| 12.1584 | 1.494279 |
| 12.2016 | 1.543083 |
| 12.2448 | 1.559115 |

| | |
|---|---|
| 12.288 | 1.554486 |
| 12.3312 | 1.592849 |
| 12.3744 | 1.454823 |
| 12.4176 | 1.470954 |
| 12.4608 | 1.430739 |
| 12.504 | 1.341805 |
| 12.5472 | 1.475249 |
| 12.5904 | 1.44086 |
| 12.6336 | 1.536867 |
| 12.6768 | 1.481571 |
| 12.72 | 1.503716 |
| 12.7632 | 1.394038 |
| 12.8064 | 1.508602 |
| 12.8496 | 1.531258 |
| 12.8928 | 1.507515 |
| 12.936 | 1.480227 |
| 12.9792 | 1.507468 |
| 13.0224 | 1.367251 |
| 13.0656 | 1.430068 |
| 13.1088 | 1.370019 |
| 13.152 | 1.448946 |
| 13.1952 | 1.344809 |
| 13.2384 | 1.400805 |
| 13.2816 | 1.474761 |
| 13.3248 | 1.374178 |
| 13.368 | 1.376538 |
| 13.4112 | 1.48932 |
| 13.4544 | 1.484599 |
| 13.4976 | 1.418915 |
| 13.5408 | 1.419193 |
| 13.584 | 1.449135 |
| 13.6272 | 1.397648 |
| 13.6704 | 1.465132 |
| 13.7136 | 1.412281 |
| 13.7568 | 1.384695 |
| 13.8 | 1.523745 |
| 13.8432 | 1.443807 |
| 13.8864 | 1.433838 |
| 13.9296 | 1.540675 |
| 13.9728 | 1.565991 |
| 14.016 | 1.489911 |

| | |
|---|---|
| 14.0592 | 1.445833 |
| 14.1024 | 1.421674 |
| 14.1456 | 1.592264 |
| 14.1888 | 1.5356 |
| 14.232 | 1.509411 |
| 14.2752 | 1.478861 |
| 14.3184 | 1.467728 |
| 14.3616 | 1.456309 |
| 14.4048 | 1.481135 |
| 14.448 | 1.457657 |
| 14.4912 | 1.359265 |
| 14.5344 | 1.496221 |
| 14.5776 | 1.558125 |
| 14.6208 | 1.487824 |
| 14.664 | 1.453931 |
| 14.7072 | 1.459721 |
| 14.7504 | 1.535703 |
| 14.7936 | 1.512492 |
| 14.8368 | 1.479148 |
| 14.88 | 1.377779 |
| 14.9232 | 1.516169 |
| 14.9664 | 1.392568 |
| 15.00 | 1.491859 |