COMPUTATIONAL ANTIVIRAL DRUG DESIGN

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

LISHAN LIU

DR. JASON W RIBBLETT

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2010
ACKNOWLEDGEMENT

I would like to extend a heartfelt thanks to everyone who supported me to accomplish the study in Ball State University and this research thesis. Without your love, encourage and confidence in me, I can not reach this place.

I would like to thank my parents, Xianqiao Liu and Morou Cheng. You are always there for me throughout all my life. No matter I success or fail, you always support me, encourage me to chase my dream and be proud of me.

I would like to thank my advisor, Dr. Jason Ribblett. You are the best mentor and advisor I can imagine of. Thank you for refreshing my view of chemistry and science. And also thank you for giving me the opportunity to complete this research.

I would like to thank my other committee members, James Poole and Scott Pattison. I will never forget that pleasant time in your classes. Thank you for your time and support of this research and thesis work.

I also would like to thank the whole Ball State University Chemistry Department gave me this opportunity to have this wonderful experience. It enlightens me and helps me made up my mind to chase a higher degree in the future. This experience will be a treasure of my whole life.
ABSTRACT

THESIS: Computational Antiviral Drugs Design

STUDENT: Lishan Liu

DEGREE: Master of Science

DEPARTMENT: Chemistry

DATE: July, 2010

PAGES: 74

This study designed and computational docked a group of ligands intended to find potent inhibitors for Neuraminidase 4 which would have strong interactions with 8 conserved amino acids in the active site.

Several trials of ligands were designed based on derivatives of neuraminic acid and evaluated as inhibitors of influenza neuraminidase. Optimized geometries of those ligands were determined using HF/B3LYP/6-311++G** techniques. Binding energies of the ligands bound to the N4 subtype of the neuraminidase protein were determined using AutoDock 4.0. Currently used inhibitors for influenza viruses will also be analyzed in the exactly same way. Comparing the binding information of those candidates and current ligands can provide a useful data about the potential of these species as antiviral drugs.
CONTENTS

Chapter 1: Introduction and Literature Review...........................................................1

1.1 Influenza: Information about the Virus and Current Treatments......................1

1.2 Computational Chemistry: Theory, methods and Uses in Drug Discovery..........5

1.3 References ..............................................................................................................8

Chapter 2: Methods and Procedures...........................................................................10

2.1 Protein Selection ....................................................................................................10

2.2 Ligand Development .............................................................................................11

2.3 Ligand build and optimization: Gaussian Procedure and Theory.......................21

2.4 Protein- Ligand interaction: AutoDock Procedure and Theory..........................23

2.5 References..............................................................................................................28

Chapter 3: Results and Conclusion..............................................................................29

3.1 Blind Docking of Neu5Ac and Interaction Analysis.............................................29

3.2 Ligand Energy Ranking and Interaction Analysis................................................33

3.3 Docking analysis of licensed anti-virus drug: AA4.............................................38

3.4 Docking analysis of AD3......................................................................................41

3.5 Conclusion..............................................................................................................43
APPECDIX: Interaction Diagrams for Potential Neuraminidase 4 Inhibitors ..............45

Figure A.2: Interaction between AA1 and N4......................................................45

Figure A.2: Interaction between AA2 and N4......................................................46

Figure A.3: Interaction between AA3 and N4......................................................47

Figure A.4: Interaction between AA4 and N4......................................................48

Figure A.5: Interaction between AA5 and N4......................................................49

Figure A.6: Interaction between AA6 and N4......................................................50

Figure A.7: Interaction between AB4 and N4......................................................51

Figure A.8: Interaction between AC3 and N4......................................................52

Figure A.9: Interaction between AC4 and N4......................................................53

Figure A.10: Interaction between AD3 and N4....................................................54

Figure A.11: Interaction between AD4 and N4....................................................55

Figure A.12: Interaction between AE4 and N4....................................................56

Figure A.13: Interaction between BA4 and N4....................................................57

Figure A.14: Interaction between BB4 and N4....................................................58

Figure A.15: Interaction between BC4 and N4....................................................59

Figure A.16: Interaction between BD4 and N4....................................................60

Figure A.17: Interaction between BE4 and N4....................................................61

Figure A.18: Interaction between CA4 and N4....................................................62

Figure A.19: Interaction between CB4 and N4....................................................63

Figure A.20: Interaction between CC3 and N4....................................................64

Figure A.21: Interaction between CC4 and N4....................................................65

Figure A.22: Interaction between CD3 and N4....................................................66
Figure A.23: Interaction between CD4 and N4.................................67
Figure A.24: Interaction between CE4 and N4.................................68
Figure A.25: Interaction between DA4 and N4.................................69
Figure A.26: Interaction between DB4 and N4.................................70
Figure A.27: Interaction between DC3 and N4.................................71
Figure A.28: Interaction between DC4 and N4.................................72
Figure A.29: Interaction between DD4 and N4.................................73
Figure A.30: Interaction between DE4 and N4.................................74
LIST OF TABLES

Table 2.1: Ligands built with conserved R₂.................................................................19

Table 2.2: Ligands built with varied R₂..........................................................................20

Table 3.1: Docking results of 20 ligands with conserved R₂ group..............................33

Table 3.2: Docking results of 6 ligands with conserved back-bone and R₁.................34

Table 3.3: Docking results of 5 ligands belong to the selected group.......................35

Table 3.4: Docking energy ranking for all 30 ligands.................................................36
LIST OF FIGURES

Figure 1.1:  Structure of Rimantadine and Amantadine.................................................4
Figure 1.2:  Structure of Oseltamivir and Zanamivir..........................................................5
Figure 2.1:  Visualization of 2HTV, N4 neuraminidase......................................................10
Figure 2.2:  8 conserved amino acid residues in the active site of influenza.........................12
Figure 2.3:  Neuraminidase protein catalysis mechanism....................................................13
Figure 2.4:  Superposition of group-1 neuraminidases and superposition of group-1 with group-2 neuraminidases......................................................................................14
Figure 2.5:  “Worm” molecule surface image of active of Group-1 and Group-2..............15
Figure 2.6:  Other differences between group-1 and group-2 neuraminidase.................17
Figure 2.7:  “backbone structure”..........................................................................................18
Figure 2.8:  Ligands built in GaussView..................................................................................22
Figure 2.9:  Gaussian Calculation Parameters Set up.............................................................23
Figure 2.10: Free energy force field evaluate the binding affinity in two steps..............25
Figure 3.1:  3D X-ray crystal structure and docking image in molecular surface model.........................................................................................................................29
Figure 3.2:  Detailed interaction image between Neu5ac and N4.....................................30
Figure 3.3:  Interaction image generated form 3D X-ray crystal structure..................31
Figure 3.4:  Clustering energy for Neu5ac..............................................................................32
Figure 3.5:  Docking image as spheres....................................................................................32
Figure 3.6:  Docking image of AD4 to N4............................................................................38
Figure 3.7:  Docking image of AD4 to N4 in molecule surface model.............................39
Figure 3.8:  Interaction image of AA4 and N4 in stick and ball model.........................40
Figure 3.9:  Interaction between AD3 and N4 in stick and ball model.........................41
Figure 3.10: Molecule surface space filling image of AD3.............................................42
Figure 3.11: Docking image of AD3 in stick-and-ball model.............................43
Chapter 1: Introduction and Literature Review

1.1 Influenza: Information about the Virus and Current Treatments

Influenza (the flu) is a contagious respiratory illness caused by a group of RNA viruses—influenza viruses. It can cause symptoms such as fever, muscle pains, headache, coughing and at times can lead to pneumonia and death.\(^1\) Influenza virus can be subdivided into three serologically distinct types: A, B and C, each with hundreds of strains.\(^2\)

Influenza virus has been producing epidemics all throughout human history. The 1918 “Spanish flu” pandemic which was caused by influenza A (H1N1) virus, was responsible for the deaths of an estimated 50 million people worldwide.\(^{13}\) There have been several human influenza pandemics during the past few years, which were caused by several extremely aggressive influenza viruses. In 1997, Highly pathogenic avian influenza A viruses or so the called “H5N1” virus began spreading in Hong Kong and then throughout Asia in 2003. In 2009 the World Health Organization (WHO) declared a pandemic of 2009 H1N1 flu (so-called “swine flu”). Those serious outbreaks of Influenza virus did not only kill thousands of lives but also had a worldwide socioeconomic impact; especially in Southeast Asia, which has high population density.\(^{14}\)
However, even with a well-established medical care system, about 36,000 people died of seasonal flu-related causes each year during the 1990s in the United States. Annual averages of 94,735 pneumonia and influenza hospitalizations were associated with influenza virus infections. Annual averages of 226,054 respiratory and circulatory hospitalizations were associated with influenza virus infections.\[^{15}\]

Such repeated and frequent outbreaks of influenza result from highly contagious strains of influenza virus, as well as the virus’ ability to constantly produce new mutant strains which are resistant to current treatments. The threat of influenza pandemics emphasizes the need for therapeutic strategies to combat these pathogens.

There are some current options for prevention and treatment for the Influenza virus. Vaccination is always recommended to all people, especially children and the elderly, or those who have asthma, diabetes, heart disease, or are immuno-compromised.\[^{16}\] Alternatively, there are two classes of anti-virus drugs which target the influenza virus directly: inhibitors of the M2 proton channel and neuraminidase inhibitors.

The type A viruses are generally considered to be the most virulent human pathogens among the three influenza types and cause the most severe symptoms. Because this virus subtype is usually associated with epidemics, these evolving viruses reproduce new mutant strains resistant to treatment constantly and quickly.\[^{3}\]

The viral particles of type A viruses are made of a central core wrapped in a viral envelope containing three main types of glycoproteins.\[^{4}\] The central core contains several pieces of segmented, single-stranded, negative-sense RNA and some proteins that package and protect those RNA.\[^{5}\] On the outside of the viral particle, there are three important components on the surface: Hemagglutinin (HA), neuraminidase (NA) and the M2 proton channel, which play vital roles in the viruses replication.\[^{6}\]
Hemagglutinin (HA) is responsible for binding the virus to the cell that is being infected.\textsuperscript{[7]} There are at least 16 different HA subtypes which are labeled H1 through H16.\textsuperscript{[8]} In order to infect the host cell and replicate itself, influenza viruses need to bind through hemagglutinin onto sialic acid sugars on the surfaces of epithelial cells, such as the cells found in mammalian nose, throat and lungs. Then the virus is imported into the host cell by endocytosis.\textsuperscript{[9]}

The M2 proton channel plays a role adjusting the acidity inside the virus, allowing the viral RNA to infect the host cell.\textsuperscript{[10]} Once the virus is inside the host cell, the hemagglutinin is activated by the acidic conditions of the endosome and fuses the viral envelope with the vacuole’s membrane. The M2 proton channel allows the protons to move into the viral envelope and decrease the pH value inside the core of the virus. This acidity will cause the core to disassemble and release the viral RNA into the host cell.\textsuperscript{[11]}

After RNA and other viral proteins which form the future viruses are replicated and assembled, hemagglutinin and neuraminidase molecules cluster into a bulge in the cell membrane. The virus RNA and viral core proteins leave the host cell and enter this membrane protrusion. At this time, this mature virus adheres to the membrane of host cell through hemagglutinin until neuraminidase cleaves sialic acid residues from the host cell.\textsuperscript{[12]} Neuraminidase is responsible for spreading of the influenza viruses. Nine subtypes of influenza neuraminidase have been found, which are labeled as NA: N1-N9.\textsuperscript{[8]}

In the propagation of influenza, hemagglutinin, neuraminidase and M2 are very important. If one of these proteins can be deactivated, the infection of influenza virus can be controlled.

As mentioned above, the M2 proton channel can adjust the pH value which can
initiate virus RNA release. If the M2 channel is blocked, replication of virus RNA and protein can’t commence. This idea lead to the invention of the first class of anti-virus drugs for influenza virus: the M2 blockers. Currently, there are two licensed M2 blockers: Amantadine and Rimantadine. (See Figure 1.3) If provided in the early period of infection, Amantadine and Rimantadine can reduce and shorten the symptoms of influenza A. However, this kind of drug can not cure the type B influenza due to the absence of protein channel in these kind of viruses. They have also been associated with several Central Nervous System (CNS) side effects, including nervousness, anxiety, agitation, insomnia and difficulty in concentrating. Also, these blockers have given rise to the rapid emergence of drug-resistant viral strains. In the 2008/2009 flu season, the CDC found that 100% of seasonal pandemic flu samples tested shows resistance to Adamantane. The CDC even issued an alert to doctors to prescribe the neuraminidase inhibitors instead of Amantadine and Rimantadine for treatment of the current circulating flu.

![Figure 1.1: structure of Rimantadine (left) and Amantadine (right)](image)

Another class of anti-viral drugs are neuraminidase inhibitors, which can bind to the active site of neuraminidase and make it inactive. Once the inhibitor binds, the
neuraminidase protein can’t cleave the virus from the sialic acid residue host cell, which means virus can not spread to infect other cells. Currently, there are two neuraminidase inhibitors on the market: Oseltamivir (Tamiflu) and Zanamivir (Relenza). (See Figure 1.4)\textsuperscript{[21]} However, different strains of influenza viruses have differing degrees of resistance against these antiviral drugs, and it is impossible to predict what degree of resistance a future pandemic strain might have. \textsuperscript{[22]} According to the CDC, Oseltamivir was not very effective in the 2009 seasonal H1N1 virus due to acquired resistance in 99.6% strains, up from 12% in the 2007-2008 flu season.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{oseltamivir_zanamivir}
\caption{structure of Oseltamivir (left) and Zanamivir (right)}
\end{figure}

The war between mutants of influenza virus and anti-viral drugs is endless. It is indisputable that there is always a large need for a practical drug for the treatment of influenza. The goal of this research is to design a new antiviral drug which is effective against influenza A. This ideal drug should be efficacious but have minimal side effects and less resistance than the current drugs on the market.
1.2 Computational Chemistry: Methods and Uses in Drug Discovery

In this project, computational methods were employed to help in the development of new potent ligands. Computational chemistry or molecular modeling is a branch of chemistry that uses computers to assist in solving chemical problems. One of the first mentions of the term "computational chemistry" can be found in the 1970s, but similar methods had been used for nearly 50 years since the 1920s. It uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules. To date, countless algorithms and computer programs have been developed. With the help of these, molecular structure properties and spectra can be calculated, chemical reactivity and reaction pathways can be modeled, and interactions between molecules can be evaluated.

Usually, the programs used in computational chemistry are based on different quantum-chemical methods that solve the Schrödinger equation based on the molecular Hamiltonian. Methods that do not include any empirical or semi-empirical parameters in their equations are called ab initio methods. They are derived directly from theoretical principles, with no inclusion of experimental data. The simplest type of ab initio electronic structure calculation is the Hartree–Fock (HF) scheme in which individual electron–electron repulsion is not specifically taken into account; only its average effect is included in the calculations.

There are also several Semi-empirical (SE) methods which employ some experimental data and results. Density functional theory (DFT) methods are one of these. In DFT, the total energy is expressed in terms of the electron density rather than the wave function. Because electron density is a function of position, it just has only
three variables (x, y, and z). So DFT methods can be usefully accurate for little computational cost.\cite{23}

Molecular mechanics (MM) is another SE method. In some cases, large molecular systems can be modeled successfully while avoiding quantum mechanical calculations entirely. MM is based on a ball and spring model of molecules which uses a single classical expression in the estimation of the energy of a compound: the harmonic oscillator. These methods can be applied to proteins and other large biological molecules, and allow for the study of the approach and interaction (docking) of potential drug molecules.\cite{24}

With the development of sophisticated computer technology in last decade, computational chemistry has made great improvements and plays an essential role in study of the interactions between molecules, and in drug design. There is no doubt that computational docking provides an effective alternative for chemical and pharmaceutical research. It is more time and cost effective, while also is more friendly to the environment.\cite{24} With the help of computational docking, a large number of potential ligands now can be tested for a specific protein in a relatively fast and easy way, when compared to traditional methods where the actual synthesis of every ligand is performed. Reducing synthesis and testing of potent ligands results in the prevention of a large amount of chemical waste.

As discussed before, the goal of this project is to design a new antiviral drug which is effective against influenza A. The neuraminidase subtype 4 was chosen as the target protein, and computational docking technique was used to study interactions between potent ligands and protein.
1.3 References

[7] Tumpey, Terrence M.; Basler, Christopher F.; Aguilar, Patricia V.; Zeng, Hui; Solorzano, Alicia; Swayne, David E.; Cox, Nancy J.; Katz, Jacqueline M.; Taubenberger, Jeffery K.; Palese, Peter; Science 2005, 310(5745), 77-80.

[24] Finer-Moore, Janet S.; Blaney, Jeff; Stroud, Robert M. *Facing the Wall in Computational Based Approaches to Drug Discovery*. In *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*; Stroud, Robert M.; Finer-Moore, Janet; Royal Society of Chemistry: Cambridge, UK, 2008.
Chapter 2: Methods and Procedures

2.1 Protein Selection

To discover potent ligands to which the type A influenza virus has less resistance, the target protein must be selected first. For this project, the neuraminidase subtype N4 (PDB ID 2HTV) was chosen for study, which can be downloaded from Protein Data Bank. (See Figure 2.1) Its structure was first released on September 9, 2005, and the latest modification was on February 24, 2009. It is a strain of influenza A virus which has a molecular weight of approximately 240 kDa and consists of a mushroom-shaped head of four identical subunits in a square-planar arrangement. Each subunit of the enzyme is made up of two polypeptide chains with 390 amino acids. [1]

Figure 2.1: Visualization of 2HTV, N4 neuraminidase
As mentioned above, there are 9 subtypes of neuraminidase (N1-N9). They can be divided into two distinct groups. Group-1 contains N1, N4, N5 and N8, while group-2 contains N2, N3, N6, N7 and N9. Superposition of the structures of each group shows that neuraminidase subtypes which belong to the same group have virtually identical active sites. However, there are crucial conformational differences between Group-1 and Group-2 neuraminidases. These distinctions are responsible for the resistance of recent pandemic influenza to current anti-viral drugs. Thus, N4 was chosen as one of the group-1 neuraminidase subtypes which has had fewer investigations involving antiviral activity.

2.2 Ligand Development

After the target protein was chosen, a list of potential ligands must be created for docking according to the active site of this specific protein. Each protein active site has a unique size, shape, set of residues and hydrophilicity, which are associated with its function. Those properties can determine the size, shape and complexity of possible ligands.

Neuraminidase is believed to play at least two critical roles in the replication of the virus, including the facilitation of virion progeny release and general mobility of the virus in the respiratory tract. It catalyzes the hydrolysis of terminal sialic acid residues from the newly formed virions and from the host cell receptors. Although the active sites of the neuraminidase have slight differences between each other, X-ray structural studies have indentified eight conserved residues which interact with substrates in all neuraminidases subtypes across influenza A and B viruses. (See Figure 2.2) This provides the opportunity for the development of compounds that
should target all influenza A and B virus strains.\textsuperscript{[4]}

\textbf{Figure 2.2:} 8 conserved amino acid residues in the active site of influenza neuraminidase around the substrate: \(\text{N-Acetylneuraminic acid (Neu5ac)}\)

The enzyme catalysis process has four steps. (See Figure 2.3)\textsuperscript{[5]} The first step involves the distortion of the \(\alpha\)-sialoside from a \(^2\text{C}_5\) chair conformer to a boat conformer, which occurs when the sialoside binds to the sialidase. The second step leads to an oxocarbocation ion intermediate, the sialosyl cation. The third step is the formation of the second transition state and the last step affords \(\alpha\)-Neu5Ac. According to this mechanism,
there are two charged transition-state present in this process which have better affinity than Neu5Ac to this active site. So this structure is being chosen as the template for the ligands. Modifications were made to the template to allow for a maximum number of interactions between new ligands and the eight conserved amino acids to increase the probability that this drug is more effective against all of the influenza viruses.

**Figure 2.3:** Neuraminidase protein catalysis mechanism

Based on this template, two anti-viral drugs-- Relenza (Zanamivir) and
Tamiflu (Oseltamivir) were discovered. However, group-1 viruses show high resistance to these two drugs, which do work well with group-2 viruses.

**Figure 2.4:** (top) Superposition of group-1 neuraminidases and (bottom) superposition of group-1 with group-2 neuraminidases (Russell, Rupert J.; Haire, Lesley F.; Stevens, David J.; Collins, Patrick J.; Yi Pu Ling, Blackburn, G. Michael.; Hay, Alan J.; Garmblin, Steven J.; Sketel, John J. Nature 2006, 443 (7): 45-49.)
A recent report on the structure of H5N1 avian influenza neuraminidase suggests that the loop-150, which forms one corner of the enzyme active site, is able to exist in at least two stable conformations, designated as open and closed (See Figure 2.4). It can be observed that the active sites of group-1 neuraminidase subtypes prefer the same conformation in loop-150: the open conformation. Evidently the closed conformation is energetically preferred in group-2 neuraminidase subtype, both in the absence and
presence of current know inhibitors. On the other hand, group-1 neuraminidases usually show an open conformation when ligand is absent. This causes a larger size of group-1 neuraminidase active sites. (See Figure 2.5) But after binging to a ligand like Oseltamivir, the 150-loop changes its conformation, where Glu199 and Asp151 are now both oriented towards the bound ligand. And the size of the active site is then much the same as it is for group-2 neuraminidases subtypes. This induced fit causes an increase in the energy barrier of the binding of current anti-virus drug.

There are other distinguishing characteristics between the two groups of influenza A viruses (See Figure 2.6). Instead of having an asparagine as residue 347 in group-2 neuraminidase subtypes; group-1 neuraminidase subtypes have a tyrosine. The 270-loop in group-1 subtypes have a tighter turn around residue 273 compared to group-2. Consequently, a conserved tyrosine residue at position 252 forms hydrogen bonds to the main-chain carbonyl at position 273, the peptide amide at 250 and to the His 274. By contrast, in group-2 enzymes, there is a smaller threonine residue at position 252 which does not have this network of hydrogen bonds involved. This hydrogen network effects the position and orientation of Glu276. [6]
Figure 2.6: Other differences between group-1 and group-2 neuraminidase. (Russell, Rupert J.; Haire, Lesley F.; Stevens, David J.; Collins, Patrick J.; Yi Pu Ling, Blackburn, G. Michael.; Hay, Alan J.; Garmblin, Steven J.; Sketel, John J. Nature 2006, 443 (7): 45-49.)

Taking these differences into account, the following "backbone-structure" was made (See Figure 2.7).
To create a stronger affinity to the 150-cavity, a guanidine group was included on C4. Four different cyclic structures were used and combined with five different R₁ groups substituted at the C1 position. The R₂ C6-C9 side chain was varied in six different substituent groups. Combinations of all these modifications give us 120 different ligands. For time efficiency, the R₂ group remained the same, while the cyclic structures and the R₁ groups were changed. 20 ligands were created by this modification and docked with N4 protein first. (See Table 2.1)
Another six ligands were built and docked with N4. This group of ligands has a conserved back-bone and R₁ group combined with 6 different R₂ groups (See Table 2.2). After comparing the docking energy of these 25 ligands, several ligands which would be expected to have better affinity will be selected from the remainder of the possible 120 ligands. Those ligands will be built and docked with the N4 protein.
<table>
<thead>
<tr>
<th>Table 2.2: Ligands built with varied $R_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA1</td>
</tr>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA2</td>
</tr>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA3</td>
</tr>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA4</td>
</tr>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA5</td>
</tr>
<tr>
<td><img src="image" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>AA6</td>
</tr>
</tbody>
</table>
2.3 Ligand build and optimization: Gaussian Procedure and Theory

Before actual computational docking can be attempted, each ligand must be modeled to obtain its lowest energy conformation. Optimization calculations were performed in a variety of ways. Four programs were used in this part: Gauss View\textsuperscript{[7]}, Gaussian 03\textsuperscript{[8]}, WinSCP \textsuperscript{[9]} and PuTTY \textsuperscript{[10]}. With the exception of Gaussian 03, the other three programs run under a Windows platform. Gaussian is run on the College of Sciences and Humanities Cluster (cluster).

Gaussian is a computational chemistry software program initially released in 1970 by John Pople’s group.\textsuperscript{[11]} This electronic structure modeling programs starts from the fundamental laws of quantum mechanics. It can predict the energies, molecular structures, vibration frequencies and properties of molecules and reactions in a wide variety of chemical environments. It can be applied to both stable species and compounds which in some cases are difficult or impossible to observe experimentally, such as short-lived intermediates and transition structures. GaussView acts as a graphical user interface and allows the user to build a molecular structure and see a visual representation of the output file.

For this research, B3LYP/6-311++G** techniques were used to perform geometry optimization calculations for all ligands. After the ligand molecules were built in GaussView, the job type was set to optimization; method was set to ground state using Hartree-Fock/B3LYP with the basis set to 6-311++G**.
The molecular structures for each ligand were built on GaussView (See Figure 2.8). And geometry optimization calculation selections were set up (See Figure 2.9). These calculations were performed on the cluster. Gaussian jobs were parallelized using Linda software. Output geometry files were visualized and converted to .mol2 format for later use in AutoDock.
2.4 Protein-Ligand interaction: AutoDock Procedure and Theory

After the ligands had been built and optimized, AutoDock 4.1 was used for docking simulation. AutoDock runs on a Linux platform, Ubuntu.

AutoDock is a suite of automated docking tools which is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.\textsuperscript{14,15} A fast atom-based computational docking tool to optimize the bound conformation is essential to a docking simulation. There are two broad categories of reported techniques for automated docking: matching methods and docking simulation methods. Matching methods create a model of the active site including steric sites and sites of hydrogen bonding. Then a given inhibitor structure is docked into the model as a rigid body by matching its geometry to that of the active site. Docking simulation methods model the docking of a ligand in a much more detailed way. The ligand starts randomly outside the target protein and translational,
orientational and conformational changes are performed until an ideal site is found. Thanks to the latter technique, flexible docking is possible. AutoDock is one of the most widely used examples of the technique which has been applied with great success in predicting bound conformations of enzyme-inhibitor complexes, peptide-antibody complexes and even protein-protein interactions.\textsuperscript{[14]}

AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein; AutoGrid pre-calculates these grids. In addition to these, it also has a graphical user interface named AutoDockTools or ADT for short, which can help to modify the ligand before docking and to visually analyze dockings after completion. It can be used to set up which bonds will be treated as rotatable in the ligand. And it also allows the user to observe the interaction between the ligand and protein. This program has applications in X-ray crystallography, structure-based drug design, protein-protein docking and chemical mechanism studies.

Docking simulations methods require two basic parts: a search method for exploring the conformational space available to the system and a force field to evaluate the energies of each a conformation.

A Lamarckian genetic algorithm (LGA) is used to find optimal binding conformation of the ligand in this specific research. This algorithm is the hybrid of the genetic algorithm (GA) with an adaptive local search (LS) method. GA mimics the characteristics of natural genetics and biological evolution. In a docking process, the protein and ligand arrangement can be described by a set of values of the ligands’ translations, orientations and conformations with respect to the protein. These are the state variables of the ligand, and each of them corresponds to a gene. The ligand’s state corresponds to the genotype while its atomic coordinates correspond to the
phenotype, which can be converted from the genotype with a developmental mapping function. The fitness of the offspring (new arrangement of ligand-protein) is determined by the total interaction energy of the ligand with the protein. Offspring are created by mating random pairs of individuals using crossover. Some of those offspring undergo random mutation by changing one gene by a random amount. Selection of the offspring is made based on the fitness. Thus offspring with better fitness survive and reproduce to create the next generation; whereas, poorer suited individuals die. The adaptive LS methods perform in phenotypic space instead of in genotypic one. When the LS arrive at a local minimum, an inverse mapping function is used to convert from its phenotype to its corresponding genotype.\[^{15}\]

**Figure 2.10:** free energy force field evaluate the binding affinity in two steps. (Morris, Garrett M.; Goodsell, David S.; Halliday, Robert S.; Huey, Ruth; Hart, William E.; Belew, Richard K.; Olson, Arthur J. Journal of Computational Chemistry **1998**, 19: 1639-1662.)

AutoDock uses a free energy force field as its scoring function. The free energy force field estimates the energies of the process of binding of two molecules using pair-wise terms to evaluate the interaction between the two molecules. Six pair-wise evaluations (\(V\)) and an estimation of the conformational entropy lost upon
binding are involved:

$$\Delta G = (V_{\text{bound}}^{\text{L-L}} - V_{\text{unbound}}^{\text{L-L}}) + (V_{\text{bound}}^{\text{P-P}} - V_{\text{unbound}}^{\text{P-P}}) + (V_{\text{bound}}^{\text{P-L}} - V_{\text{unbound}}^{\text{P-L}} + \Delta S_{\text{conf}})$$

In which, the L refers to the “ligand” and P refers to “protein” in a protein-ligand complex. The first two terms are intramolecular energies for the bound and unbound states of the ligands. The second two terms are intramolecular energies for the bound and unbound states of the protein. When the protein is rigid, the difference between these two terms is equal to zero. And the last two terms represent the change in intermolecular energy between the bound and unbound states. It is assumed that the two molecules are sufficiently distant from one another in the unbound state so the last term is zero. The term for the loss of torsional entropy upon binding is directly related to the number of rotatable bonds in the molecule:

$$\Delta S_{\text{conf}} = W_{\text{conf}} N_{\text{tors}}$$

This force field shows improvement in redocking simulations over the former force field and give a good prediction for the docking energy.\(^{[15]}\)

The first step for docking is ligand preparation. Ligand files in the .mol2 format were opened in AutoDock. Automatically, charges were added and all non-polar hydrogen atoms were merged on the ligand. Then the root atoms need to be detected and all bonds within the ligand were set as rotatable or fixed. After all torsions were determined and set, the file was saved as a .pdbqt file type.

Then the protein file, which was downloaded from Protein Data Bank, was prepared for docking. The .pdb file for the target protein downloaded from this web site then was opened directly with AutoDock. The neuraminidase 4 (PDB code: 2HTV)
protein was opened in AutoDock. After that, the excess water molecules were isolated and deleted from the structure, while all hydrogen atoms were added to the protein structure. These changes were then saved. The rigid and flexible residues of the protein were created as a file _rigid.pbdqt and a file _flex.pbdqt.

After the ligand and protein were ready for docking, a set of grid maps were constructed with the help of the AutoGrid tool. The _rigid.pbdqt file of protein was opened in the screen, and then the ligand was chosen to set the type of mapping for docking. A grid box was used to select which part of the protein would be mapped. To begin, the whole protein was chosen for blind docking to confirm the interaction of N-Acetylatedaminic acid, the product of this catalyst reaction, with N4. The resulting bound structure was compared with the 3D x-ray crystal structure to see if AutoDock shows a good prediction by confirming that the global minimum docking site in this N4 is consistent with the reported active site. Once these were confirmed, only the active site area was chosen for grid map preparation in order to save time.

When the grip map was ready, this docking was submitted to AutoDock. In this final step, both the rigid protein file and the ligand file were selected. The LGA was set up, where the number of every generation, the ratio of mutation and the number of conformations returned were chosen. Then the docking parameters were set and a docking product file was created. After all of these were completed, AutoDock was launched to run. The results of docking conformations produced were returned in the .dpf file which was opened and analyzed in AutoDock.
2.5 References


[3] Finer-Moore, Janet S.; Blaney, Jeff; Stroud, Robert M. Facing the Wall in Computational Based Approaches to Drug Discovery. In Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions; Stroud, Robert M.; Finer-Moore, Janet; Royal Society of Chemistry: Cambridge, UK, 2008.


Chapter 3: Results and Conclusion

3.1 Blind Docking of Neu5Ac and Interaction Analysis

Neu5Ac was docked 25 times with N4 through the whole protein with a resulting average docked energy of –8.21 kcal/mol. A blind docking technique was used for this docking. This method does not limit the docking site to one small area, but instead searches the entire surface of the protein. Blind docking results in conformations of the ligand binding to multiple sites on the protein, which forms a cluster of docking sites. By comparing the docking image of the 3D X-ray crystal structure with the docking result, the best conformation of Neu5Ac was found to be docked in the reported active site every time (See Figure 3.1).

Figure 3.1: 3D X-ray crystal structure (left) and docking result (right) in molecular surface model
In Figure 3.1, both protein and ligand are shown in a “worm” molecule surface model. The left image is the 3D x-ray crystal structure; the gray part indicates the N4 protein and the black part the ligand. The right image is the docking image produced by AutoDock, the gray part indicates the N4 protein and the blue part the ligand. We can see in both images, the ligand fits the protein in an identical site, which is the reported active site.

Figure 3.2: Detailed interaction image between Neu5ac and N4 generated from AutoDock
The detailed interaction also can be compared (See Figure 3.2). In the 3D x-ray crystal structure, hydrogen bonding is represented by the green dotted lines. Hydrogen bonds are formed between a hydroxyl group with glu119 on the enzyme, the C5 carbonyl group with arg152, the C1 carboxyl group with tyr406 and arg371 and the C6-C9 side chain with glu277. Those interactions also are found in the lowest energy structure predicted by AutoDock.

The program was able to determine the most likely site for binding within the protein even though a blind docking procedure was employed for this analysis. Figure 3.3 shows a typical energy conformation diagram of the ten ligand conformations returned by one of the AutoDock runs. In this diagram, the majority of the returned conformations had an energy ranging around \(-8.5\) kcal/mol. And in Figure 3.4 the ligands are represented by little green spheres. Each sphere represents one docking
conformation. It shows the docking position on the surface of the protein in the active site all ten times.

**Figure 3.4:** Clustering energy for Neu5ac

**Figure 3.5:** Docking image as spheres
3.2 Ligand Energy Ranking and Interaction Analysis

Because the active site has been confirmed in AutoDock for this specific project, a grid box focusing on the active site was created and all ligands were “forced” to dock in this site.

25 dockings were performed on each of the 20 ligands which have the same C6~C9 side chain (R₂ group) with the protein structure. The dockings returned conformations along with estimations of the docked energy. Lower docked energies result in better affinity of the ligands to protein. The lowest energy conformations for each of the 25 trials were compared visually to ensure they were consistent. And average docked energies for each ligand were calculated (see Table 3.1).

Table 3.1: Docking results of 20 ligands with conserved R₂ group (Energy in kcal/ mol)

<table>
<thead>
<tr>
<th>R₂</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH</td>
<td>-7.64</td>
<td>-6.30</td>
<td>-6.60</td>
<td>-7.30</td>
</tr>
<tr>
<td>OCH₃</td>
<td>-6.91</td>
<td>-5.69</td>
<td>-6.18</td>
<td>-7.52</td>
</tr>
<tr>
<td>OCH₂CH₃</td>
<td><strong>-7.83</strong></td>
<td>-7.26</td>
<td>-7.38</td>
<td><strong>-7.41</strong></td>
</tr>
<tr>
<td>CH₂OH</td>
<td><strong>-7.64</strong></td>
<td>-7.47</td>
<td><strong>-7.65</strong></td>
<td>-6.80</td>
</tr>
<tr>
<td>CH₂CH₂OH</td>
<td>-7.41</td>
<td>-7.08</td>
<td>-7.12</td>
<td>-6.52</td>
</tr>
</tbody>
</table>

Another group of ligands were then docked with the N4 protein keeping the back-bone and R₁ group conserved while modifying the C6~C9 side chain. The lowest
energy conformations for each of the 25 trials were compared visually to ensure they were consistent. And average docked energies for each ligand were calculated (see Table 3.2).

**Table 3.2**: Docking results of 6 ligands with conserved back-bone and \( R_1 \) (Energy in Kcal/mol)

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Structure</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA1</td>
<td><img src="image" alt="AA1 Structure" /></td>
<td>-7.02</td>
</tr>
<tr>
<td>AA2</td>
<td><img src="image" alt="AA2 Structure" /></td>
<td>-7.21</td>
</tr>
<tr>
<td>AA3</td>
<td><img src="image" alt="AA3 Structure" /></td>
<td>-7.85</td>
</tr>
<tr>
<td>AA4</td>
<td><img src="image" alt="AA4 Structure" /></td>
<td>-7.64</td>
</tr>
<tr>
<td>AA5</td>
<td><img src="image" alt="AA5 Structure" /></td>
<td>-6.89</td>
</tr>
</tbody>
</table>
Five ligands were made by comparing these two sets of docked energies and then submitted to AutoDock for simulation. Average docked energies for each ligand were the calculated (see Table 3.3).

**Table 3.3:** Docking results of 5 ligands belong to the selected group (Energy in kcal/mol)

<table>
<thead>
<tr>
<th>Molecule structure</th>
<th>Binding Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC3</td>
<td>-7.48</td>
</tr>
<tr>
<td>AD3</td>
<td>-7.99</td>
</tr>
<tr>
<td>CC3</td>
<td>-7.24</td>
</tr>
</tbody>
</table>
All 30 ligands were ordered from lowest to highest docked energy and a relative docked energy was found by comparing the docked energy of each ligand to the AA4 ligand. (see Table 3.4)

Table 3.4: Docking energy ranking for all 30 ligands (Energy in kcal/mol)

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Binding Energy (kcal/mol)</th>
<th>Standard deviation</th>
<th>Relative Binding Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD3</td>
<td>-7.99</td>
<td>0.670</td>
<td>-0.35</td>
</tr>
<tr>
<td>AA3</td>
<td>-7.85</td>
<td>0.308</td>
<td>-0.21</td>
</tr>
<tr>
<td>AC4</td>
<td>-7.83</td>
<td>0.426</td>
<td>-0.19</td>
</tr>
<tr>
<td>CD3</td>
<td>-7.65</td>
<td>0.592</td>
<td>-0.01</td>
</tr>
<tr>
<td>CD4</td>
<td>-7.65</td>
<td>0.103</td>
<td>-0.01</td>
</tr>
<tr>
<td>AD4</td>
<td>-7.64</td>
<td>0.266</td>
<td>0</td>
</tr>
<tr>
<td>AA4</td>
<td>-7.64</td>
<td>0.499</td>
<td>0</td>
</tr>
<tr>
<td>DB4</td>
<td>-7.52</td>
<td>0.238</td>
<td>0.12</td>
</tr>
<tr>
<td>AC3</td>
<td>-7.48</td>
<td>0.258</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>BD4</td>
<td>-7.47</td>
<td>0.379</td>
<td>0.17</td>
</tr>
<tr>
<td>DC4</td>
<td>-7.42</td>
<td>0.377</td>
<td>0.22</td>
</tr>
<tr>
<td>AE4</td>
<td>-7.41</td>
<td>0.651</td>
<td>0.23</td>
</tr>
<tr>
<td>CC4</td>
<td>-7.38</td>
<td>0.421</td>
<td>0.26</td>
</tr>
<tr>
<td>DA4</td>
<td>-7.30</td>
<td>0.692</td>
<td>0.34</td>
</tr>
<tr>
<td>DC3</td>
<td>-7.26</td>
<td>0.469</td>
<td>0.38</td>
</tr>
<tr>
<td>BC4</td>
<td>-7.26</td>
<td>0.435</td>
<td>0.38</td>
</tr>
<tr>
<td>CC3</td>
<td>-7.24</td>
<td>0.565</td>
<td>0.4</td>
</tr>
<tr>
<td>AA2</td>
<td>-7.21</td>
<td>0.229</td>
<td>0.43</td>
</tr>
<tr>
<td>CE4</td>
<td>-7.12</td>
<td>0.140</td>
<td>0.52</td>
</tr>
<tr>
<td>BE4</td>
<td>-7.08</td>
<td>0.586</td>
<td>0.56</td>
</tr>
<tr>
<td>AA1</td>
<td>-7.02</td>
<td>0.202</td>
<td>0.62</td>
</tr>
<tr>
<td>AB4</td>
<td>-6.91</td>
<td>0.428</td>
<td>0.65</td>
</tr>
<tr>
<td>AA6</td>
<td>-6.90</td>
<td>0.498</td>
<td>0.64</td>
</tr>
<tr>
<td>AA5</td>
<td>-6.89</td>
<td>0.516</td>
<td>0.65</td>
</tr>
<tr>
<td>DD4</td>
<td>-6.80</td>
<td>0.630</td>
<td>0.74</td>
</tr>
<tr>
<td>CA4</td>
<td>-6.60</td>
<td>0.575</td>
<td>0.94</td>
</tr>
<tr>
<td>DE4</td>
<td>-6.52</td>
<td>0.645</td>
<td>1.02</td>
</tr>
<tr>
<td>BA4</td>
<td>-6.30</td>
<td>0.386</td>
<td>1.22</td>
</tr>
<tr>
<td>CB4</td>
<td>-6.18</td>
<td>0.439</td>
<td>1.34</td>
</tr>
<tr>
<td>BB4</td>
<td>-5.69</td>
<td>0.543</td>
<td>2.3</td>
</tr>
</tbody>
</table>
3.3 Docking analysis of licensed anti-virus drug: AA4

Among these 30 ligands, one is Zanamivir, which was labeled as AA4. The average docked energy of this ligand is -7.64 Kcal/mol which is the fifth most stable inhibitor according to our result. Figure 3.5 displays the docking of the ligand within the N4 protein. The protein is colored by atom type. The red color represents oxygen atoms; the dark blue color represents nitrogen atoms; the grey color represents carbon atoms; and the light blue color represents the polar hydrogen atoms, those bonded to nitrogen or oxygen atoms. The yellow areas indicate the presence of sulfur atoms. The white spot highlights are shown to give the image the appearance of three-dimensionality. The ligand is shown using a green stick model. The ligand is situated within a deep pocket of the protein. A molecular surface space filling model
better displays the ligand filling the pocket (See Figure 3.6).

Figure 3.7: Docking image of AA4 to N4 in molecule surface model

A detailed interaction image is also returned by AutoDock (See Figure 3.7). In this image, hydrogen bonds are formed between R1 group with Arg118, Glu119 while the R2 side chain with Arg371 and Arg292. This conformation gives an unfavorable orientation to the guanidine group and the C5 carbonyl group compared to the substrate Neu5ac. As a result, there are no interactions between any residue with the C5 carbonyl group; and the guanidine group does not interact well with the 150-loop. Also, it fails to interact with Glu276.
Figure 3.8: Interaction image of AA4 and N4 in stick and ball model
3.4 Docking analysis of AD3

Among these 30 ligands, AD3 has the lowest docking energy, which indicates that it has the best affinity to the protein active site. A detailed docking interaction image is shown above (See Figure 3.8).

In this image, hydrogen bonds are presents as green dotted lines. 6 hydrogen bonds are forming: three between the R₁ group and Arg371, Try406 and Arg118, two between the C4 guanidine group and Leu113 and Arg151, one between the R₂ group

Figure 3.9: Interaction between AD3 and N4 in stick and ball model
and Glu276. In addition, interactions between the 150-loop and the C5 carboxyl group and Glu227, R₁ with Glu119 and Arg292 also can be observed. So all eight conserved amino acid all have interactions with AD3.

![Molecule surface space filling image of AD3](image)

**Figure 3.10:** Molecule surface space filling image of AD3

The docking image of this ligand also is obtained in both space-filling and stick-and-ball model. It can be observed that AD3 fits in the active site well (See Figure 3.9 and Figure 3.10).
3.5 Conclusion

A series of conclusions can be drawn from these results. Since AA4 is a licensed anti-viral drug for influenza, it can be argued that 5 ligands on the list which have lower docking energy might be potential targets as anti-virus drugs. And AD3, which returns a lower binding energy then AA4 shows a more favorable interactions with the amino acids in the active site. It shows interactions with all eight conserved residues in the active site, which indicates that it could be a universal inhibitor for all the group-1 neuraminidase subtypes.

There are some areas of further inquiry. One is some possible candidates for this protein may have been overlooked, because a select subset of the 120 ligands was chosen to reduce times. Ligands, such as AA3, AE3, BD3 and DB3, might return good docking results. Another issue is that the protein was treated rigidly in the
docking with all ligands. In future studies, flexible docking can be employed to see if more accurate structure results. Finally, all conclusions are just based on the free energy change in this binding process. Other properties such as transportation characteristics, toxicity, and solubility are still unknown for all ligands in this study.
Figure A.1: Interaction between AA1 and N4
Figure A.2: Interaction between AA2 and N4
Figure A.3: Interaction between AA3 and N4
Figure A.4: Interaction between AA4 and N4
Figure A.5: Interaction between AA5 and N4
Figure A.6: Interaction between AA6 and N4
Figure A.7: Interaction between AB4 and N4
Figure A.8: Interaction between AC3 and N4
Figure A.9: Interaction between AC4 and N4
Figure A.10: Interaction between AD3 and N4
Figure A.11: Interaction between AD4 and N4
Figure A.12: Interaction between AE4 and N4
Figure A.13: Interaction between BA4 and N4
Figure A.14: Interaction between BB4 and N4
Figure A.15: Interaction between BC4 and N4
Figure A.16: Interaction between BD4 and N4
Figure A.17: Interaction between BE4 and N4
Figure A.18: Interaction between CA4 and N4
Figure A.19: Interaction between CB4 and N4
Figure A.20: Interaction between CC3 and N4
Figure A.21: Interaction between CC4 and N4
Figure A.22: Interaction between CD3 and N4
Figure A.23: Interaction between CD4 and N4
Figure A.24: Interaction between CE4 and N4
Figure A.25: Interaction between DA4 and N4
Figure A.26: Interaction between DB4 and N4
Figure A.27: Interaction between DC3 and N4
Figure A.28: Interaction between DC4 and N4
Figure A.29: Interaction between DD4 and N4
Figure A.30: Interaction between DE4 and N4