AUTODOCK AS A TOOL FOR VIRTUAL SCREENING FOR ANTI-INFLUENZA DRUGS

INTRODUCTION

The influenza virus is an RNA virus that contains two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA) shown in Figure 1, to initiate viral fusion and subsequent budding of new virions from the infected cell (Sears and Wong, 1999). Both of these glycoproteins are presented on the surface of the influenza virus and are essential for virion propagation. Hemagglutinin recognizes target cells via sialic acid binding sites and then promotes viral fusion (Huang *et al.*, 1981). Neuraminidase cleaves terminal sialic acid (Figure 2a) moieties of virus progeny to promote the release and subsequent spreading of new virus particles, as shown in Figure 3. In Figure 3, virus attachment through HA to receptors containing terminal neuraminic acid residues and penetration into host cell, transcription of viral RNA, and translation of viral proteins, replication of viral RNA and assembly of virion, budding, and subsequent release from host cell. Both glycoproteins have been suggested as therapeutic targets to prevent the spread of the influenza virus in the host (Bamford, 1995; Bodian *et al.*, 1993; Hatakeyama, 2005). In the present, more success has been achieved for designing compounds against neuraminidase than hemagglutinin (Babu *et al.*, 2000). Due to neuraminidase role in the life cycle of the influenza virus and its potential to prevent the spread of infection, neuraminidase is an appealing target. Neuraminidase exhibits a high degree of selectivity among structurally similar compounds.

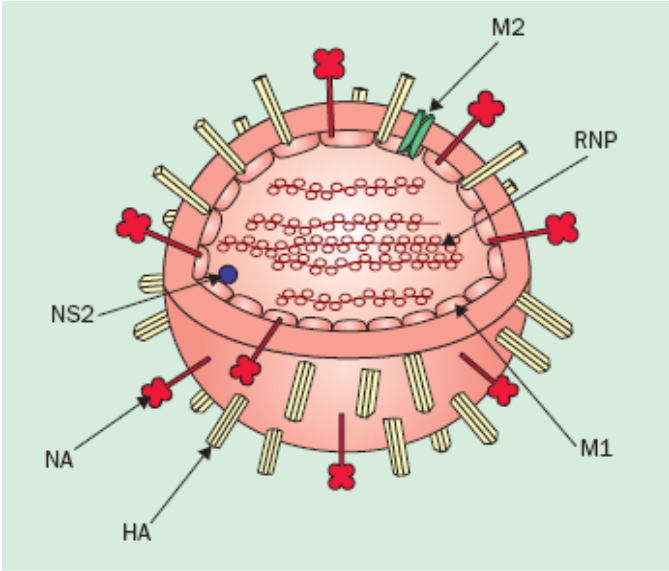


Figure 1 Schematic representation of influenza A virion. Eight ribonucleoprotein segments (RNP) are surrounded by layer of matrix (M1) protein and lipid bilayer taken from host cell at budding. NS2 (NEP) protein is associated with M1. Three viral proteins are incorporated into the lipid bilayer: HA, NA, and M2 protein. HA trimers and NA tetramers form spikes on the surface of the virion. RNP segments contain viral RNA surrounded by nucleoprotein and associated with the polymerase complex.

Source: Gubareva *et al*. (2000)

The first inhibitor of influenza virus neuraminidase, developed by Meindl and Tuppy, (1969) inhibited viral replication but had low potency and specificity. In the 1980s P. M. Colman and his colleagues reported the crystal structure of influenza virus NA complexed with neuraminic acid. Those findings, together with an improved understanding of the mechanism of catalysis, set the stage for the synthesis of neuraminic acid derivatives with enhanced affinity for influenza NA. In 1993, Itzstein and co-workers demonstrated that 4-guanidino-Neu5Ac2en (GG167 or zanamivir, Figure 2b) was a potent and highly specific inhibitor of influenza NA activity that inhibited both *in vitro* and *in vivo* virus replication (Gubareva *et al.*, 2000) . Due to its high polarity, the bioavailability is very poor and oral inhalation or intravenous

injection is necessary. Oseltamivir (Tamiflu, Figure 2c) is a prodrug which has less polarity and better clinical efficacy (Moscona, 2005). The drug resistance does not happen in case of less distributed zanamivir. The virus can rapidly adapt itself when the disease is endemic or the drug is not carefully prescribed. For these reasons, searching for new potent anti-neuraminidase is important for drug discovery.

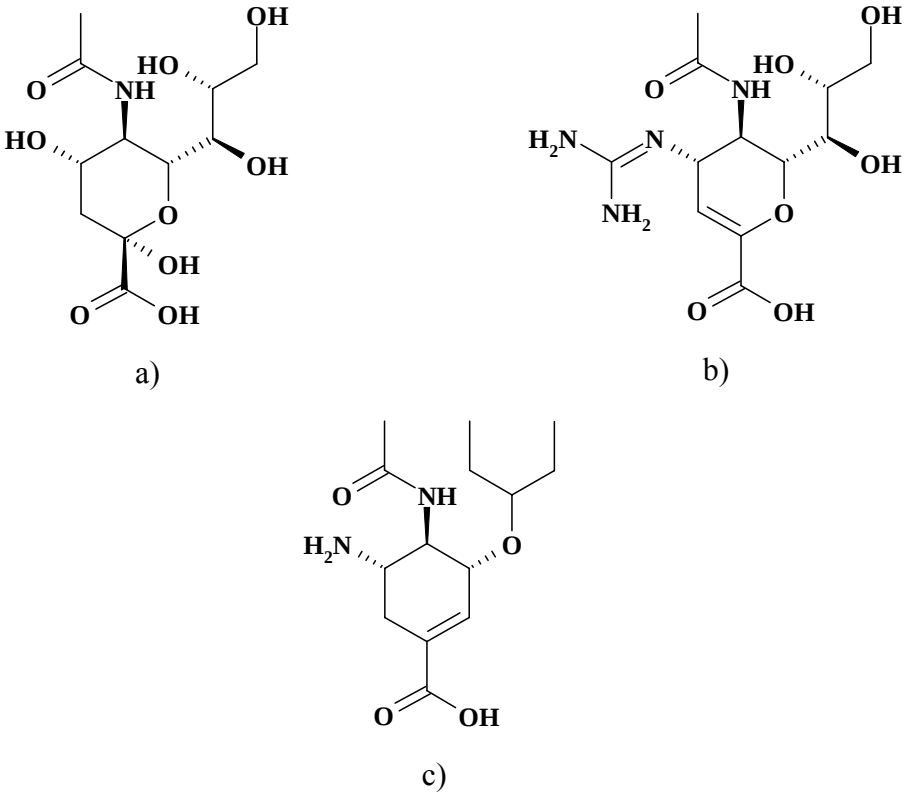


Figure 2 Chemical structures of some influenza neuraminidase inhibitors: a) Sialic acid (N-acetylneuraminic acid, Neu5Ac); b) Zanamivir (4-guanidio-Neu5Ac2en); (c) Oseltamivir (4-N-acetyl-5-amino-3-(1-ethylpropoxy)-1-cyclohexene-1-carboxylic acid methyl-ester prodrug).

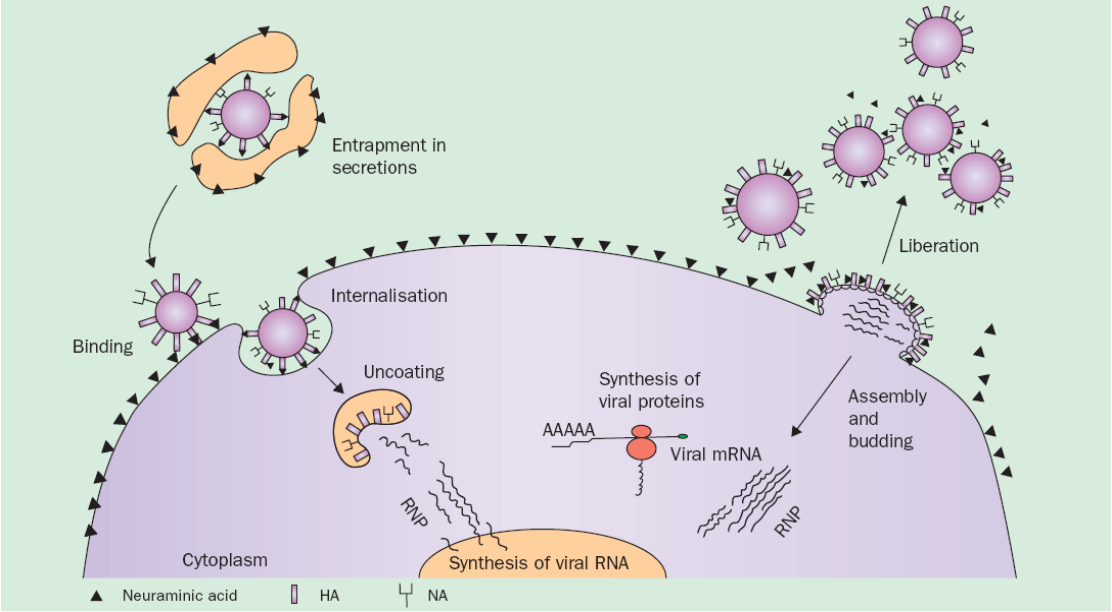


Figure 3 Influenza virus replication cycle.

Source: Gubareva *et al*. (2000)

Since knowledge of 3D-structure of receptor is known, a well-known approach of drug design is called structure-based drug design. This method searched for a ligand whose orientation and conformation achieves the highest degree of complementarities with respect to all details of the receptor's steric constraints and interaction geometries (Bajorath, 2002; Hoffmann *et al*., 1999; Oprea, 2002; Seifert *et al*., 2003; Shoichet, 2004). Structure-based drug design is one of the most powerful method in drug discovery process in recent year because the explosion of structural information has provided hundreds new targets and opportunities for future drug lead discovery (Lyne, 2002; Anderson, 2003). The target's structure reveals the important binding mode of inhibitors indicates the essential aspects determining its binding affinity and generates new ideas about the ways of improving drug efficacy (Klebe, 2000).

Nowadays, several successes using structure-based drug design in discovery of new more potent drugs has been published, shown in literature review section.

Structure-based drug design includes several methods, i.e. 3D database techniques, molecular docking and fragment methods *etc* (Marrone *et al.*, 1997). In 3D database searching, the pharmacophore model was used to search a database of 3D molecular structures and resulted in a molecule. For molecular docking, it is important to the prediction of ligand conformation and orientation within a targeted binding site. Docking compounds from databases to targets of known structure can be utilized to discover or refine new lead compounds which are active against a specific target as a starting point. To avoid testing on too many candidates and test only compounds that meet our criteria, this technique is called "virtual screening".

Virtual screening is an efficient method for generating new leads for drug discovery (Abagyan and Totrov, 2001; Birch *et al.*, 2003). However, virtual screening is hampered by difficulties in obtaining the correct bind-virtual screening. The binding mode is important because it is difficult to believe that correct ranking can be routinely obtained with incorrect binding modes. There are many docking programs available so it is often difficult to judge which is the best program for a given protein target (Taylor *et al.*, 2002). Enzymes conformational changes occur on ligand binding. This may involve only small side chain rotations to maximize interactions with the ligand (Najmanovich *et al.*, 2000), or the change may also be associated with small main chain movements. In extreme cases, large loop movements or even domain shifts are induced on ligand binding. Cases of large induced fit are difficult to account in any docking program (Sherman *et al.*, 2006). A more realistic goal would be a method robust enough to deal with relatively small changes in the active site when an analogous ligand binds. This is especially important in virtual screening of novel compounds where, there is a crystal structure of a different ligand bound to the protein. The sensitivity of results to small structural changes is very important because virtual screening of compound databases against a single protein may often miss hits that require small structural modifications of the protein to bind well. Then, the ideal situation would be a program able to dock a ligand into a protein structure from a different complex easily and with reasonable accuracy. This has been referred to as "cross-docking" (Kramer *et al.*, 1999).

To date, in virtual screening process the small compounds from database are fitted into the target structure using a docking program, such as DOCK (Ewing *et al.*, 2001), FlexX (Rarey *et al.*, 1996), GOLD (Verdonk *et al.*, 2003), AutoDock (Morris *et al.*, 1998), ICM (Chen *et al.*, 2007), GLIDE (Halgren *et al.*, 2004), SLIDE (Zavodszky *et al.*, 2003), LigandFit (Venkatachalam *et al.*, 2003), FRED (McGann *et al.*, 2003) and Surflex (Jain, 2003) as a tool for screening new lead compounds. Among these docking programs for virtual screening, AutoDock is a program that is widely accepted when working on individual molecules. However, this program does not have a good result when used to screen new lead compounds from the large database (Douglas *et al.*, 2004). Due to using this program to screen many compounds on a large database is not successful because the analysis of the several results from this program is not practical. However, there are many program features that are still make it interesting. AutoDock offers no fee for academic license which makes it popular. Moreover, it is one program that always gives good docking results when it is especially used to predict a binding conformation of an inhibitor in protein. It is easy to use. For these reasons, analyzing thousands of docking results at the same time can be overcome. AutoDock could be another powerful virtual screening program.

Therefore, this study has two objectives; the first objective is to investigate the feasibility of using AutoDock program as a tool for virtual screening for search the new lead compound that against the influenza viral coat protein neuraminidase is a response to the continuing threat of influenza epidemics. This information can be used a starting point for discovers of other drugs via the method of virtual screening, and the second objective is to perform a virtual screening using AutoDock program combined with FRED program to screen the many compounds from ChemieBase database for anti-influenza drugs. Therefore, this works have been separated into two parts.

In part I, we were reproduced 24 co-crystals of NA and inhibitors (see Table 1, Materials and Methods) using AutoDock version 3.0.5 in order to observe the calculated docked energy (summation of intermolecular energy term and internal

energy of ligand), binding free energy (summation of intermolecular energy term and torsional free energy term), and the root mean square deviation (rmsd) values. The docking results were compared the binding free energy, docked energy of prediction with the experimental data by using simple linear regression model. The correlation coefficient, r^2, was used to identify the accuracy of calculation model. Moreover, the AutoDock complexes of reproduction were rescored using six different scoring functions in FRED program to choose the best docked conformation and compared with the best rmsd docked conformation. In the next step, cross validation of the 24 co-crystals predicted binding free energy and docked energy. The docking results of cross validation were rescored using various scoring functions in FRED program to ranking the affinity of inhibitor set. From this part, we will obtain the suitable protein template and to choose the suitable procedure for virtual screening step.

In part II, for finding the lead compounds for anti-influenza drugs using the best protein template from the part I, compounds in a database or unknown set will be screened using the chosen procedure.

LITERATURE REVIEW

In search of the new generation of therapeutics against the influenza virus is a response to the continuing threat of influenza epidemics. Recently virtual screening as an efficient method of generating new lead compounds for drug discovery. In this process, molecular docking and estimation of binding affinities are an important goal in structure based drug design applications (Goodsell *et al.*, 1996; Knegtel *et al.*, 1997). Molecular docking is important to the prediction of ligand conformation and orientation within a target binding site. In general, there are two aims of docking studies: accurate structural modeling and correct prediction of activity (Kuntz *et al.*, 1982).

Docking is generally devised as a multi-step process in which each step introduces one or more additional degrees of complexity (Brooijmans and Kuntz, 2003). The process begins with the application of docking algorithms that pose small molecules in the active site, the process of determining whether a given conformation and orientation of a ligand fits the active site. This in itself is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run. Algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. Early scoring functions evaluated compound fits on the basis of calculations of approximate shape and electrostatic complementarities. Relatively simple scoring functions continue to be heavily used, at least during the early stages of docking simulations. Pre-selected conformers are often further evaluated using more complex scoring schemes with more detailed treatment of electrostatic and van der Waals interactions (Gohlke and Klebe, 2002).

In addition to solve problems associated with scoring of compound conformations, other complications exist that makes it challenging to accurately

predict binding conformations and compound activity. These include, among others, limited resolution of crystallographic targets, inherent flexibility, induced fit (Birch *et al.*, 2003) or other conformational changes that occur on binding, and the participation of water molecules in protein-ligand interactions.

For many years ago, many research groups have tried to search new lead compounds in the large database. They used the several different docking programs or scoring functions as a tool for virtual screening (Oda *et al.*, 2006). The ultimate measure of success for the many docking program is their ability to produce significant hit rates while reducing the number of compounds that need to be tested. A number of successful case studies have been published over the years that demonstrate the potential of the approach. Most lead compounds with at least low-micromolar potency were usually found, often without biasing search calculations towards previously identified hits. The results also mirror the general trend that hits in the micromolar range are much more frequently identified than nanomolar hits in these calculations. Major reasons for this explanation are that newly identified active compounds are rarely optimized for potency against a given target and that nanomolar potency is typically obtained after chemical modification in the course of hit-to-lead transition and lead optimization (Greer *et al.*, 1994).

The examples of successful case studies have been reported that is success by structure-based drugs designs are shown in the following.

In 2001, Enyedy *et al.* (2001) described the discovery of novel classes of small-molecule inhibitors targeted at the BH3 binding pocket in Bcl-2. Bcl-2 belongs to a growing family of proteins which regulates programmed cell death (apoptosis). Over expression of Bcl-2 has been observed in 70% of breast cancer, 30-60% of prostate cancer, 80% of B-cell lymphomas, 90% of colorectal adenocarcinomas, and many other forms of cancer. Thereby, Bcl-2 is an attractive new anti-cancer target. The three-dimensional (3D) structure of Bcl-2 has been modeled on the basis of a high-resolution NMR solution structure of template protein. A structure-based computer screening approach has been employed to search the National Cancer

Institute 3D database of 206,876 organic compounds to identify potential Bcl-2 small-molecule inhibitors that bind to the BH3 binding site of Bcl-2. Thirty-five potential inhibitors were tested in this binding assay, and seven of them were found to have a binding affinity (IC_{50} value) from 1.6 to 14.0 μM. The anti-proliferative activity of these seven active compounds has been tested using a human myeloid leukemia cell line, HL-60, which expresses the highest level of Bcl-2 protein among all the cancer cell lines examined.

Next, Kamionka *et al.* (2002) have determined the crystal structure of the insulin-like growth factor-I (IGF-I) in complex with the N-terminal domain of the IGF-binding protein-5 (IGFBP-5). They reported of computer screening results for potential inhibitors of this interaction using the crystal coordinates. From the compounds suggested by *in silico* screens, successful binders were identified by NMR spectroscopic methods. NMR was also used to map their binding sites and calculate their binding affinities. Small molecular weight compounds bind to the IGF-I binding site on the IGFBP-5 with micromolar affinities (IC_{50} = 43 μM) and thus serve as potential starting compounds for the design of more potent inhibitors and therapeutic agents for diseases that are associated with abnormal IGF-I regulation.

Vangrevelinghe *et al.* (2003) used a structure-based approach to screen for alternative human thymidine phosphorylase inhibitors, in particular to identify new scaffolds with submillimolar IC_{50} values suitable for subsequent lead optimization. This strategy involved a hierarchical computational screening of the National Cancer Institute (NCI) database of anticancer compounds using the computer program DOCK to predict protein–ligand binding mode and affinity. For docking 250,000 compounds of the NCI database on to protein thymidine phosphorylase by doing is rigid docking 50,000 orientations per ligand is used and 444 ligands are flexible docked. There are 3 compounds that showed IC_{50} lower than 100 μM.

After that, Peng *et al.* (2003) presented the discovery of novel inhibitors targeted at the catalytic domain of ABL tyrosine kinase by using 3D-database searching techniques. From a database containing 200,000 commercially available

compounds, the top 1000 compounds with the best binding score were selected and subjected to structural diversity and drug likeness analysis, 15 compounds were submitted for biological assay. Eight out of the 15 showed inhibitory activity against K562 cells with IC_{50} value ranging from 10 to 200 mM. Two promising compounds showed inhibition in further ABL tyrosine phosphorylation assay. It is anticipated that those two compounds can serve as lead compounds for further drug design and optimization.

Doman *et al.* (2002) reported that High-throughput screening (HTS) of compound libraries is used to discover novel leads for drug development. When a structure is available for the target, computer-based screening using molecular docking may also be considered. The two techniques have rarely been used together on the same target. The opportunity to do presented itself in a project to discover novel inhibitors for the enzyme protein tyrosine phosphatase-1B (PTP1B). A tyrosine phosphatase has been implicated as a key target for type II diabetes. A corporate library of approximately 400,000 compounds was screened using high-throughput experimental techniques for compounds that inhibited PTP1B. Concurrently, molecular docking was used to screen approximately 235,000 commercially available compounds against the x-ray crystallographic structure of PTP1B, and 365 high-scoring molecules were tested as inhibitors of the enzyme. Of approximately 400,000 molecules tested in the high-throughput experimental assay, 85 (0.021%) inhibited the enzyme with IC_{50} values less than 100 M; the most active had an IC_{50} value of 4.2 M. Of the 365 molecules suggested by molecular docking, 127 (34.8%) inhibited PTP1B with IC_{50} values less than 100 M; the most active of these had an IC_{50} of 1.7 M. Structure-based docking therefore enriched the hit rate by 1,700-fold over random screening. The hits from both the high-throughput and docking screens were dissimilar from phosphotyrosine, the canonical substrate group for PTP1B; the two hit lists were also very different from each other. Surprisingly, the docking hits were judged to be more drug-like than the HTS hits. The diversity of both hit lists and their dissimilarity from each other suggest that docking and HTS may be complementary techniques for lead discovery.

Brenk *et al.* (2003) reported that they used FlexX program for virtual screening for Eubacterial tRNA-guanine transglycosylase (TGT) of the bacterium due to inefficient translation of a virulence protein mRNA. They are described the discovery of a ligand with an unexpected binding mode. On the basis of this binding mode, three slightly deviating pharmacophore hypotheses have been derived. Virtual screening based on this composite pharmacophore model retrieved a set of potential TGT inhibitors belonging to several compound classes. All nine tested inhibitors being representatives of these classes showed activity in the 0.25-256 μM.

The DOCK program has been used to dock the small compound containing around 400,000 structures into protein kinase CK2, of which a dozen were selected for actual testing in a biochemical assay (Enyedy *et al.,* 2001). The compound inhibits the enzymatic activity of CK2 with an IC_{50} value of 80 nM, making it the most potent inhibitor of this enzyme ever reported. Its high potency, associated with high selectivity, provides a valuable tool for the study of the biological function of CK2. Those results suggest that virtual screening of a 3D database by molecular docking is a useful approach for lead finding provided that adapted post docking filtering and reranking procedures are applied to the primary hit list.

Once case study, Rastelli *et al.* (2003) proposed an approach toward molecular docking of the structures contained in the Available Chemicals Directory (ACD) database around 4,000 structures to search for novel inhibitors of PfDHFR. Instead of docking the whole ACD database, specific 3D pharmacophores were used to reduce the number of molecules in the database by excluding a priori molecules lacking essential requisites for the interaction with the enzyme and potentially unable to bind to resistant mutant PfDHFRs. The molecules in the resulting "focused" database were then evaluated with regard to their fit into the PfDHFR active site. Twelve new compounds whose structures are completely unrelated to known anti-folates were identified and found to inhibit, at the micromolar level, the wild-type and resistant mutant PfDHFRs harboring A16V, S108T, A16V + S108T, C59R + S108N + I164L, and N51I + C59R + S108N + I164L mutations. Depending on the functional groups interacting with key active site residues of the enzyme, these inhibitors were classified

as N-hydroxyamidine, hydrazine, urea, and thiourea derivatives. The structures of the complexes of the most active inhibitors, as refined by molecular mechanics and molecular dynamics, provided insight into how these inhibitors bind to the enzyme and suggested prospects for these novel derivatives as potential leads for anti-malarial development.

Perez and Ortiz (2001) compared force field scoring with a combined PMF knowledge-based function. In this study, force-field terms generally performed better than PMF, but steric contributions significantly outweighed electrostatic terms. Correct ligand poses were found in nearly 80% of the cases studied when the force-field function was used, but the success rate dropped to 56% when cross-docking experiments were performed. In an effort to better understand incorrectly calculated effects, linear discriminate analysis was applied; the results indicated that better adjusted molecular volume and hydrogen bonding terms were likely to further improve force-field scoring.

Sotriffer *et al.* (2003) compared DrugScore (knowledge based) with AutoDock (force-field-based) and found that these two scoring methods had very similar abilities in predicting the correct binding modes for 158 complexes. For rigid molecules, 90–100% of cases were predicted correctly. However, as the number of rotatable bonds (and molecular flexibility) increased, the success rate dropped to 44 – 80%, and DrugScore only marginally improved the results over AutoDock. These findings also illustrated the difficulty in finding correct poses for highly flexible ligands.

For the other of case studies, there are many researcher groups were focus to assess the ability of many scoring function. For examples, Charifson *et al.* (1999) showed that most conventional scoring functions were able to place ~50% of active (<100 nM) compounds within the top 1,000 of the score lists but that less active (~1 μM) compounds were much more difficult to identify. Chemscore, the DOCK energy and PLP scores performed well for all three binding sites that were analyzed. In this study, a consensus scoring was applied that combined results from each scoring function and was found to further improve prediction accuracy. In a similar study,

Wang *et al*. (2003) tested on 100 protein-ligand complexes to evaluate their abilities to reproduce experimentally determined structures and binding affinities. They included four scoring functions implemented in the LigFit module in Cerius2 (LigScore, PLP, PMF, and LUDI), four scoring functions implemented in the CScore module in SYBYL (FScore, G-Score, D-Score, and ChemScore), the scoring function implemented in the AutoDock program, and two stand-alone scoring functions (DrugScore and X-Score). In this case study, they used the AutoDock program to generate an ensemble of docked conformations for each ligand molecule. Then, each scoring function is applied to score this conformational ensemble and used to identify the experimentally observed conformation from all of the other decoys. Finally, they found that six scoring functions, *i.e.,* PLP, F-Score, LigScore, DrugScore, LUDI, and X-Score, yield success rates higher than the AutoDock scoring function. The results form this work indicated that AutoDock program is a suitable for generating the structural accuracy and failed in scoring function.

In the next year, Wang and co-worker were used the fourteen popular scoring functions, i.e., X-Score, DrugScore, five scoring functions in the Sybyl software (D-Score, PMF-Score, G-Score, ChemScore, and F-Score), four scoring functions in the Cerius2 software (LigScore, PLP, PMF, and LUDI), two scoring functions in the GOLD program (GoldScore and ChemScore), and HINT, to refined set of 800 diverse protein-ligand complexes with high-resolution crystal structures and experimentally determined K_i or K_d values. This study, they done to assess the ability of these scoring functions to predict binding affinities based on the experimentally determined high-resolution crystal structures of proteins in complex with their ligands. The quantitative correlation between the binding scores produced by each scoring function and the known binding constants of the 800 complexes was computed. X-Score, DrugScore, Sybyl::ChemScore, and Cerius2::PLP provided better correlations than the other scoring functions with standard deviations of 1.8-2.0 log units. These four scoring functions were also found to be robust enough to carry out computation directly on unaltered crystal structures. To examine how well scoring functions predict the binding affinities for ligands bound to the same target protein, the performance of these 14 scoring functions were evaluated on three subsets of protein-ligand

complexes from the test set: HIV-1 protease complexes (82 entries), trypsin complexes (45 entries), and carbonic anhydrase II complexes (40 entries). Although the results for the HIV-1 protease subset are less than desirable, several scoring functions are able to satisfactorily predict the binding affinities for the trypsin and the carbonic anhydrase II subsets with standard deviation as low as 1.0 log unit. Our results demonstrate the strengths as well as the weaknesses of current scoring functions for binding affinity prediction.

For influenza viral, there are many research groups have tried to search and develop a new anti-influenza drug. In 1999, Murray and co-workers described the application of PRO_LEADS to the flexible docking of ligands into crystallographically derived enzyme structures that are assumed to be rigid. PRO_LEADS uses a tabu search methodology to perform the flexible search and an empirically derived estimate of the binding affinity to drive the docking process. To test, the extent to which the assumption of a rigid enzyme compromises the accuracy of the results. All pairs docking experiments are performed for three enzymes (thrombin, thermolysin and influenza virus neuraminidase) based on six or more ligand-enzyme crystal structures for each enzyme. In 76% of the cases, PRO_LEADS can successfully identify the correct ligand conformation as the lowest energy configuration when the enzyme structure is derived from that ligand's crystal structure, but the methodology only docks 49% of the cases successfully when the ligand is docked against enzyme crystal structures derived from other ligands. Small movements in the enzyme structure lead to an under-prediction in the energy of the correct binding mode by up to 14 kJ/mol and in some cases this under-prediction can lead to the native mode not being recognized as the lowest energy solution. This work illustrated that the assumption of a rigid active site can lead to errors in identification of the correct binding mode and the assessment of binding affinity, even for enzymes which show relatively small shift in atomic positions from one ligand to the next. PRO_LEADS can usually dock successfully if there is induced fit in relatively rigid enzymes but there remains the need to develop improved strategies for dealing with enzyme flexibility.

Birch *et al.* (2002) investigated the limits of the rigid protein approximation used by the docking program, GOLD, through cross-docking using protein structures of influenza neuraminidase. Neuraminidase is known to exhibit small but significant induced fit effects on ligand binding. Some neuraminidase crystal structures caused concern due to the bound ligand conformation and GOLD performed poorly on these complexes (Birch *et al.*, 2003). A 'clean' set, which contained unique, unambiguous complexes, they defined. For this set, the lowest energy structure was correctly docked (*i.e.* rmsd < 1.5 Å away from the crystal reference structure) in 84% of proteins, and the most promiscuous protein structure PDB code 1mwe was able to dock all 15 ligands accurately including those that normally required an induced fit movement. This is considerably better than the 70% success rate seen with GOLD against general validation sets. Moreover, inclusion of specific water molecules involved in water-mediated hydrogen bonds did not significantly improve the docking performance for ligands that formed water-mediated contacts but it did prevent docking of ligands that displaced these waters. This data supports the use of a single protein structure for virtual screening with GOLD in some applications involving induced fit effects, although care must be taken to identify the protein structure that performs best against a wide variety of ligands.

McGann *et al.* (2003) compared the robust of FRED and Glide scoring schemes for cases of cyclooxygenase-2, oestrogen receptor,mitogen-activated kinase, gyrase B, thrombin, gelatinase-A and neuraminidase. They reported that the scoring functions in Glide performed overall better than the scoring function in FRED. However, FRED was found to produce accurate results for lipophilic binding sites, especially when hydrophobic effects outweighed electrostatic and hydrogen bonding interactions. Scoring functions therefore respond differently to specific features in binding sites.

In 2005, Maria *et al.* (2005) combined FRED, DOCK, and Surflex for a multistep virtual ligand screening (VLS) procedure to screen the pocket of four different proteins, estrogen receptor (ER), thymidine kinase (TK), coagulation factor VIIa (F7) and neuraminidase (NA). The goal of this study was to evaluate the impact

of chaining “freely available packages to academic users” on docking/scoring accuracy and CPU time consumption. A bank of 65,660 compounds including 49 known actives was generated. This study is successful because docking/scoring parameters are tuned according to the nature of the binding pocket and because a shape based filtering tool is applied prior to flexible docking. They suggested that consensus docking/scoring could be valuable to some drug discovery projects. This protocol could process the entire bank for one receptor in less than a week on one processor, suggesting that VLS experiments could be performed even without large computer resources.

Recently, Park *et al*. (2006) has been reported the first example of assessing the performance of the AutoDock program in virtual database screening in comparison to the other popular docking programs. The automated AutoDock represents an advance in virtual screening of chemical databases with docking simulation. By using the unique 3D potential grids common to all ligands, they have been able to accelerate the overall virtual screening process by more than two times without a significant change in binding free energies of individual ligands. These results exemplify the usefulness of the automated AutoDock as a new promising tool in structure-based virtual screening.

MATERIALS AND METHODS

Methodologies

1. Molecular Docking

Molecular docking is a method to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein (Krovat *et al.*, 2005). A general docking procedure consists of three steps: identification of the binding site, a search algorithm to effectively sample the set of orientation and conformation of ligands, and a scoring function (McConkey *et al.*, 2002). More details of molecular docking method are described in Appendix A. In this study, using AutoDock and FRED program have been applied for performing virtual screening. The details of each docking method are described as following.

1.1 AutoDock

AutoDock (Automated Docking of Flexible Ligands to Receptors) is a docking method combines a positional, orientational and conformational search engine with a grid-based method of energy evaluation. AutoDock 3.0.5 was using a Lamarckian genetic algorithm (LGA) for docking flexible ligands into protein binding sites (Morris *et al.,* 1998). The LGA were combined a genetic algorithm (GA) for global searching and a local search (LS) method to perform energy minimization. The local search method is based on that of Solis and Wets (Gohlke and Klebe, 2002). The LGA switches between "genotypic space" and "phenotypic space." Mutation and crossover occur in genotypic space, while phenotypic space is determined by the energy function to be optimized. Energy minimization (local sampling) is performed after genotypic changes have been made to the population (global sampling) in phenotypic space, which is conceptually similar to Monte Carlo (MC) minimization. The phenotypic changes from the energy minimization are mapped back onto the genes (by changing the ligand coordinates in the chromosome).

AutoDock have implemented a grid-base method for calculation of energy to precalculate grid maps of pairwise atomic interaction energies (Morris *et al.*, 1996). For AutoDock program calculated grid map by using AutoGrid program.

The following figure illustrates the main features of a grid map:

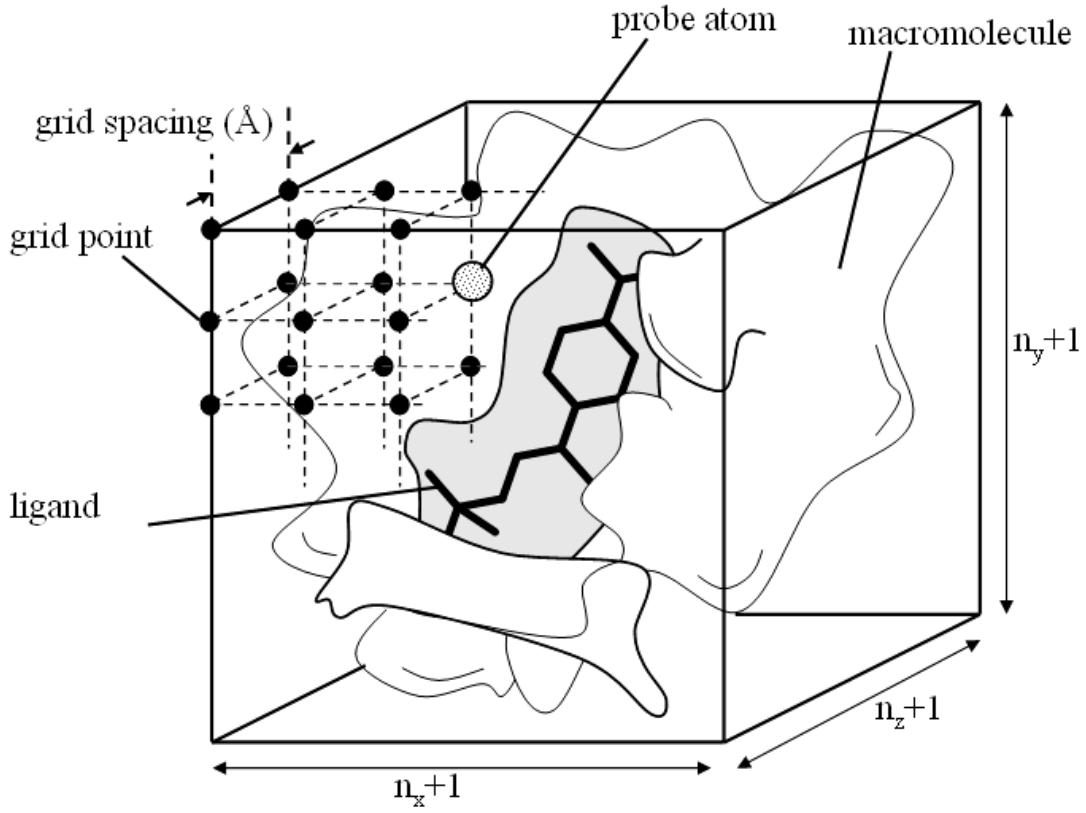


Figure 4 Illustrates the main features of a grid map.

A grid map consists of a three dimensional lattice of regularly spaced points which is called grid point, surrounding (either entirely or partly) and centered on some region of interest of the macromolecule or protein target. Grid point spacing varies from 0.2Å to 1.0Å, although the default is 0.375Å. In the practical, the user must set an even number of grid points in each dimension (n_x, n_y and n_z), because AutoGrid will be add a central point (n_x+1, n_y+1 and n_z+1), and AutoDock requires an odd number of grid points. For each grid point within the grid map stores the potential energy of a 'probe' atom which is summed over all protein atoms within a nonbonded

cutoff radius of 8 Å. The final grid of energies provides a lookup table for the rapid evaluation of interaction energies. Separate grids are calculated for each type of atom in the ligand, including a dispersion/repulsion term and added one electrostatic-potential grid. These grids are read into memory at the start of the docking job, and then sampled by the ligand's atoms type using trilinear interpolation.

The AutoDock force-field parameters are a subset of those of AMBER (Weiner *et al.*, 1984). For a given probe atom, i, and all protein atoms, j, within a nonbonded cutoff distance (R_{cut}). For Van der Waals energies are calculated using a Lennard-Jones 12-6 potential:

$$E_{vdW} = \sum_{i<j, R_{ij}<R_{cut}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) \quad (1)$$

Where R_{ij} is the distance between interacting atoms. The coefficients A_{ij} and B_{ij} are calculated from well depths (ε_{XX}) and equilibrium contact distances ($r_{eqm,XX}$) between the nuclei of two like atoms, X. Combining rules for the Van der Waals radius, r_{eqm}, and the well depth, ε, for two different atoms X and Y, are:

$$r_{eqm,XY} = \frac{1}{2}\left(r_{eqm,XX} + r_{eqm,YY}\right) \quad (2)$$

$$\varepsilon_{XY} = \sqrt{\varepsilon_{XX}\,\varepsilon_{YY}} \quad (3)$$

Where X and Y is different atom types. Assuming that the potential is a minimum at r_{eqm} with a value of -ε, then:

$$A_{ij} = \varepsilon_{XY} r_{eqm,XY}^{12} \quad (4)$$

$$B_{ij} = 2\varepsilon_{XY} r_{eqm,XY}^{6} \quad (5)$$

Hydrogen bonds are treated with a traditional 12-10 potential:

$$E_{H-bond} = \sum_{i<j, R_{ij}<R_{cut}} \left(\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right) \quad (6)$$

Coefficients C_{ij} and D_{ij} may similarly be calculated for hydrogen bonds using the assumption of minimum energy,-ε, at internuclear separation, r_{eqm}, thus:

$$C_{ij} = 5\varepsilon_{XY} r_{eqm,XY}^{12} \quad (7)$$

$$D_{ij} = 6\varepsilon_{XY} r_{eqm,XY}^{10} \quad (8)$$

In addition to the atomic affinity grid maps, AutoDock requires an electrostatic-potential grid map. Partial atomic charges have to be assigned to the macromolecule. The electrostatic grid can be generated by AutoGrid. AutoGrid calculates electrostatic interaction energy grid maps using Coulombic interactions between the macromolecule and a probe of charge e, $+1.60219 \times 10^{-19}$ C; there is no distance cutoff used for electrostatic interactions. The latter uses a sigmoidal distance dependent dielectric function based on the work of Mehler and Solmajer.

$$\varepsilon(r) = A + \frac{B}{1 + ke^{-\lambda Br}} \quad (9)$$

Where: $B = \varepsilon_0 - A$, with $\varepsilon_0 = 78.4$ (the dielectric constant of bulk water at 25 °C), A =-8.5525, k = 7.7839, and $\lambda = 0.003627$ $Å^{-1}$. Electrostatic-potential grids are calculated with a probe carrying a single positive charge. The electrostatic interaction energy of each atom in the ligand is obtained by multiplying the trilinearly interpolated electrostatic potential taken from this grid with the partial charge of the atom.

The docking program AutoDock version 3.0 implemented a semiempirical scoring function which is rewritten by using the thermodynamic cycle of Wesson and Eisenberg. This scoring function includes five terms, as in equation 10:

$$\Delta G = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right)$$
$$+ \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor}$$
$$+ \Delta G_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i \right) \cdot e^{(-r_{ij}^2 / 2\sigma^2)} \qquad (10)$$

The five ΔG terms on the right-hand side are coefficients empirically determined using linear regression analysis from a set of 140 protein-ligand complexes with known binding constants. For the AutoDock 3.0 scoring function, the Van der Waals coefficient (ΔG_{vdW}) will be scaled by 0.1485, the hydrogen bonding coefficient (ΔG_{hbond}) will be scaled by 0.0656 and the electrostatic energy coefficient (ΔG_{elec}) will be scaled by 0.1146. For loss of torsional degrees of freedom upon binding coefficient (ΔG_{tor}) and the ligand desolvation free energy coefficient (ΔG_{tor}) will be scaled by 0.3113 and 0.1711 respectively.

In addition, the first term in the right-hand side of equation 10 is the contribution of Van der Waals force between the ligand and the acceptor to binding free energy using a Lennard-Jones 12-6 potential. The second term is the contribution from hydrogen bond using a Lennard-Jones 12-10 potential, the hydrogen bonding term has directionality owing to the *E(t)* factor which is a function of the angle. The third term is that of electrostatic potential; the fourth is the change of binding free energy aroused by the frozen rotary free energy in ligands; and the last term accounts for desolvation effects. Autodock uses a pairwise, volume-based method to estimate the buriedness of the atom, which is multiplied by the atomic solvation parameter for that atom. This function was evaluated based on precalculated grids for the receptor contributions, and was derived based on 30 protein-ligand complexes.

1.2 FRED

FRED (Fast Rigid Exhaustive Docking) is a protein-ligand docking program. The ligand conformers and protein structure are treated as rigrid during the docking process. FRED's docking strategy is to exhaustively score all possible positions of each ligand in the active site. The exhaustive search is based on rigid rotation and translation of each conformer. Typical docking time for FRED is few second per ligand. The six scoring functions in FRED program are explanted as in following:

1.2.1 Shapegauss

A shape based Gaussian function (McGann *et al.*, 2003). The Gaussian docking function originally suggested by Grant and Pickup is a summation of pairwise interactions between the ligand and protein atoms. Each pairwise interaction, $F_{i,j}$, are calculated from:

$$F_{ij}(d_{ij}) = \left(\frac{32\kappa^3}{9\pi}\right)\left(\frac{\pi R_i^2 R_j^2}{\kappa R_i^2 + \kappa R_j^2}\right)^{3/2} \times \left[1 + \kappa\frac{d_{ij}^2 - D_{ij}^2}{R_i^2 + R_j^2}\right]\exp\left(-\frac{\kappa}{R_i^2 + R_j^2}d_{ij}^2\right) \quad (11)$$

where radii R_i and R_j are the radii of atoms i and j, respectively, $d_{i,j}$ is the distance between the two atom centers, $D_{i,j}$ is sum of the sphere radii ($D_{i,j} = R_i + R_j$), κ is a freely adjustable parameter. The parameter κ controls the distribution of the Gaussian ($\kappa = 1.5$).

1.2.2 PLP

The Piecewise Linear Potential (PLP) is empirical scoring function (Verkhivker *et al.*, 2000). This scoring function is a sum of pairwise linear potentials between ligand and protein heavy atoms with parameters dependent on interaction type. It can be expressed conceptually as

$$E_{total} = E_{H-bond} + E_{repulsion} + E_{contct} \quad (12)$$

Ligand and protein heavy atoms are classified as hydrogen bond donors, acceptors, donor/acceptors, or nonpolar. Each pair of interacting atoms is then assigned one of the three interaction types: hydrogen bonding between donors and acceptors, repulsive donor-donor and acceptor-acceptor contacts, and generic dispersion of other contacts. Both the hydrogen bonding and repulsive terms are modulated by a scaling factor that imparts a crude distance and angular dependence. Small (fluorine and metal ion), medium (carbon, oxygen, and nitrogen), and large (sulfur, phosphorus, chlorine, and bromine) atoms are assigned atomic radii of 1.4, 1.8, and 2.2 Å, respectively. These parameters are derived from interatomic distances observed from a large number of high quality crystal structures.

1.2.3 Chemscore

Chemscore is empirical scoring function is based on the work of Eldridge *et al.* (1997).

$$\begin{aligned}\Delta G_{binding} = {} & \Delta G_0 + \Delta G_{H-bond} \sum_{iI} g_1(\Delta r) g_2(\Delta \alpha) \\ & + \Delta G_{metal} \sum_{aM} f(r_{am}) + \Delta G_{lipo} \sum_{IL} f(r_{IL}) \\ & + \Delta G_{rotor} H_{rot}\end{aligned} \quad (13)$$

The ΔG in the first five terms on the right-hand side are coefficients empirically determined using multiple linear regressions of 82 protein-ligand complexes which known binding constants. The first term accounts for protein-ligand hydrogen bonding. The second term accounts for the coordinate bonding between the ligand and the metal ions residing inside the protein binding pocket. The third term accounts for the hydrophobic effect, which is calculated by summing a distance dependent potential of all the hydrophobic atom pairs formed between the

complexes. The fourth term also counts rotors, but the contribution of each rotor is scaled by a complicated function to reflect the chemical nature of its environment. The first term ΔG_0 is a regression constant.

For the hydrogen bond term, $\sum_{iI} g_1 g_2$ is calculated for all complementary possibilities of hydrogen bonds between ligand atoms, i, and receptor atoms, I. The functions g_1 and g_2 are of the same form as used by Böhm:

$$g_1(\Delta r) = \begin{cases} 1 & \text{if } \Delta r \le 0.25\,\text{Å} \\ 1-(\Delta r - 0.25)/0.4 & \text{if } 0.25\,\text{Å} < \Delta r \le 0.65\,\text{Å} \\ 0 & \text{if } \Delta r > 0.65\,\text{Å} \end{cases} \quad (14)$$

$$g_2(\Delta\alpha) = \begin{cases} 1 & \text{if } \Delta\alpha \le 30^\circ \\ 1-(\Delta\alpha - 30)/50 & \text{if } 30^\circ < \Delta\alpha \le 80^\circ \\ 0 & \text{if } \Delta\alpha > 80^\circ \end{cases} \quad (15)$$

Δr is the deviation of the H---O/N hydrogen bond length from 1.85 Å and $\Delta\alpha$ is the deviation of the hydrogen bond angle N/O-H---O/N from its ideal value of 180°. Water molecules are scored for contact with the receptor and those possessing more than one hydrogen bond to the receptor are treated as if they are part of the receptor when contacts between the ligand and receptor are evaluated.

The metal term, $\sum_{aM} f(r_{aM})$ is calculated for all acceptor and acceptor/donor atoms, a, in the ligand and any metal atoms, M, in the receptor. The function f(r) is a simple contact term and its functional form is illustrated in Figure 5. The parameters R1 and R2 defined by Figure 5 are 2.2 and 2.6 Å, respectively. r_{aM} is the distance between ligand and receptor atoms of the appropriate type.

The lipophilic term, $\sum_{lL} f(r_{lL})$ is calculated for all lipophilic ligand atoms, l, and all lipophilic receptor atoms, L. The parameters R1 and R2

defined by Figure 5 are $r_l^{vw} + r_L^{vw} + 0.5$ and R1 + 3.0, respectively, where r_l^{vw} is the Van der Waals radius of atom l and distances are measured in Å.

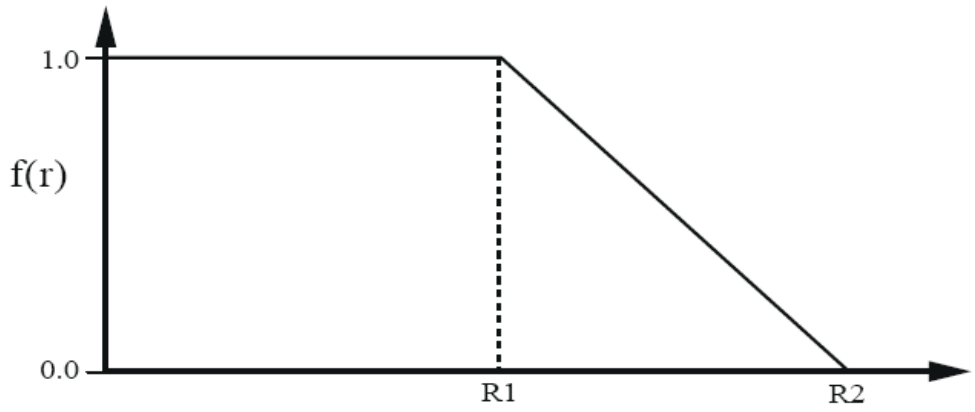


Figure 5 Form of the function f(r) given in equation 13. The parameters R1 and R2 have different values for the different terms in equation 13.

Source: Eldridge *et al.* (1997)

The final term identifies frozen rotatable bonds. The definition of rotatable bonds is any sp^3-sp^3 bond and any sp^2-sp^3 bond, excluding bonds to terminal CH_3, CF_3, NH_2 or NH_3 groups. Bonds are considered frozen if atoms on both sides of the rotatable bond are in contact with the receptor. The following function is used to estimate the flexibility penalty for molecules possessing frozen rotatable bonds:

$$H_{rot} = 1 + (1 - 1/N_{rot}) \sum_r (P_{nl}(r) + P'_{nl}(r))/2 \qquad (16)$$

Where N_{rot} is the number of frozen rotatable bonds, the summation is over frozen rotatable bonds and Pnl(r) and P'nl(r) are the percentages of non-lipophilic heavy atoms on either side of the rotatable bond.

1.2.4 Chemgauss2

This scoring were combined with Chemgauss scoring function (Verdonk *et al.*, 2004) and the Shapegauss scoring function with additional potentials between chemically matched positions around the ligand pose. These chemically complementary positions are generally not only atom positions, but also rather are placed near specific functional groups. For instance acceptors have lone pair positions around them which denote positions where polar hydrogen could be placed to create a hydrogen bonding interaction. Similarly donors have polar hydrogen positions, which denote positions its hydrogen could be in. For simple donors without rotatable bonds these positions correspond to the actual polar hydrogen position, but for rotatable bonds such as hydroxyls there are several positions representing the ring of possible positions for the polar hydrogen. A favorable hydrogen bond score is obtained when a polar hydrogen position on one molecule overlaps a lone pair position on another molecule.

1.2.5 Screenscore

The Screenscore scoring function was derived from PLP and FlexX score (Stahl and Rarey, 2001). The standard scoring function used in FlexX (Rarey *et al.*, 1996) is a modified version of the empirical scoring function by Boehm (Boehm, 1994). It can be written as a sum of five contributions:

$$\Delta G_{bind} = \Delta G_{match} F_{match} + \Delta G_{lipo} F_{lipo} + \Delta G_{ambig} F_{ambig} + \Delta G_{clash} F_{clash} + \Delta G_{rot} F_{rot} \quad (17)$$

Where the ΔG_i are coefficients of functions F_i operating on the protein and ligand coordinates. ΔG_{match} is a sum of scores for directed interactions between receptor and ligand. It consists of individual energy contributions for each hydrogen bond, metal contact, and specific aromatic interaction multiplied by two linear penalty functions for angle and distance deviations from predefined ideal values. The terms F_{lipo} and F_{ambig} provide a measure of hydrophobic contact surface as functions of receptor-ligand atom pairs, F_{lipo} involving only pairs of unpolar atoms

and F_{ambig} involving pairs of one polar and one unpolar atom. Finally, F_{clash} is a penalty function for protein-ligand overlap, and n_{rot} is equal to the number of rotatable bonds in the ligand times a weighting factor.

The final combined PLP-FlexX combination, which is called ScreenScore in the following, has the form:

$$\Delta G_{bind} = \Delta G_{match} + 0.07(F_{lipo} + F_{ambig}) + 0.3F_{PLP} + 1.6n_{rot} \quad (18)$$

1.2.6 Consensus

The poses returned from exhaustive docking are scored by six scoring functions (Shapegauss, PLP, Chemgauss2, Chemscore, Screenscore). For each scoring function a list of the poses is created ordered by rank. Each pose is then assigned a consensus structure score equal to the sum that pose's rank in each list.

2. Mathematical Methods

Due to, AutoDock program calculated the energy of a particular substrate configuration using tri-linear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate. In addition, this work used linear regression analysis to find the best fitting curve between predicted data and experimental data and calculated the correlation coefficient of these data. Then, in this part will be explained about an interpolation techniques, linear regression analysis and correlation coefficient respectively.

2.1 Geometric Interpolation

Interpolation techniques are mathematical computations that employ geometrical relationships or cellular regression. Geometrical interpolations exploit the ways of subdividing a cube. There are four geometrical interpolations, trilinear, prism,

pyramid, and tetrahedral. In this work, we focused to explain the trilinear interpolation.

Basically, trilinear interpolation is the multiple application of the linear interpolation (Fule, 1990). Therefore, we start with the linear interpolation, and then extend to bilinear interpolation and trilinear interpolations.

2.1.1 Linear Interpolation

A linear interpolation is depicted in Figure 6; a point P on the curve between the lattice point P_0 and P_1 is to be interpolated. The interpolated value $P_c(x)$ is linearly proportional to the ratio of $(x - x_0)/(x_1 - x_0)$, where $(x_1 - x_0)$ is the projected length of the line segment connecting points P_0 and P_1, and $(x - x_0)$ is the projected distance of the line connecting points P and P_0, as shown in the follow equation.

$$P_c(x) = P(x_0) + [(x - x_0)/(x_1 - x_0)][P(x_1) - P(x_0)] \qquad (19)$$

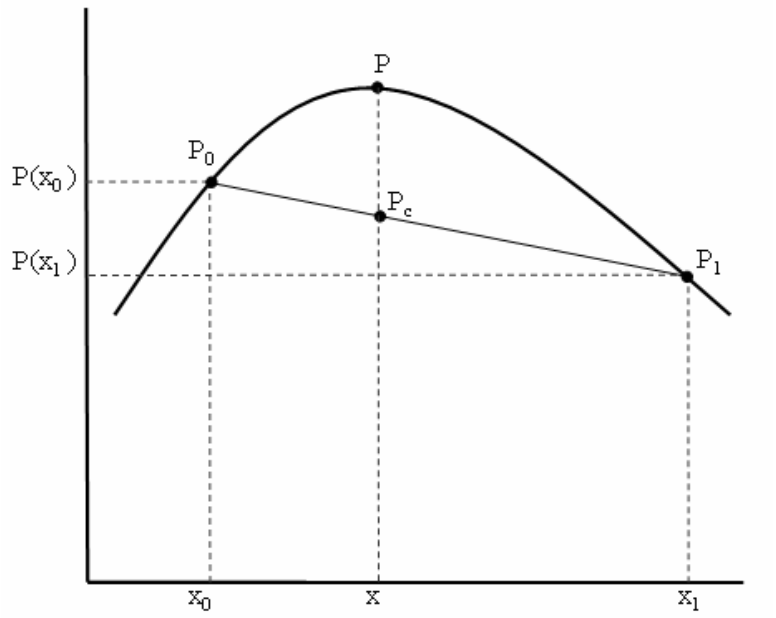


Figure 6 Linear interpolation

2.1.2 Bilinear Interpolation

In two dimensions, we have a function of two variables P(x, y) and four lattice points $P_{00}(x_0, y_0)$, $P_{01}(x_0, y_1)$, $P_{10}(x_1, y_0)$, and $P_{11}(x_1, y_1)$ as shown in Figure 7. To obtain the value for point P, we first hold y_0 constant and apply the linear interpolation on lattice points P_{00} and P_{10} to obtain P_0.

$$P_0 = P_{00} + (P_{10} - P_{00})[(x - x_0)/(x_1 - x_0)] \tag{20}$$

Similarly, we calculate P_1 by keeping y_1 constant.

$$P_1 = P_{01} + (P_{11} - P_{01})[(x - x_0)/(x_1 - x_0)] \tag{21}$$

After obtaining P_0 and P_1, we again apply the linear interpolation to them by keeping x constant.

$$P_{(x,y)} = P_0 + (P_1 - P_0)[(y - y_0)/(y_1 - y_0)] \tag{22}$$

Substituting equation 20 and 21 into equation 19, we obtain

$$\begin{aligned} P_{(x,y)} = {} & P_{00} + (P_{10} - P_{00})[(x - x_0)/(x_1 - x_0)] \\ & + (P_{01} - P_{00})[(y - y_0)/(y_1 - y_0)] \\ & + (P_{11} - P_{01} - P_{10} + P_{00})[(x + x_0)/(x_1 - x_0)] \\ & \times [(y - y_0)/(y_1 - y_0)] \end{aligned} \tag{23}$$

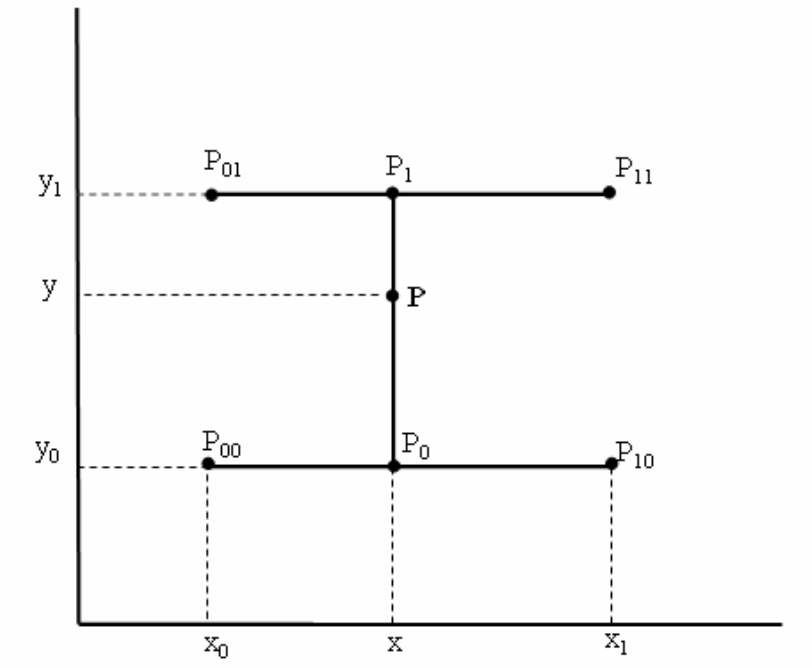


Figure 7 Bilinear interpolation

2.1.3 Trilinear Interpolation

The trilinear equation is derived by applying the linear interpolation seven times (see Figure 8). The general expression for the trilinear interpolation is given in equation 21 to compute value of point P.

$$P_{(x,y,z)} = c_0 + c_1\Delta x + c_2\Delta y + c_3\Delta z + c_4\Delta x\Delta y + c_5\Delta y\Delta z + c_6\Delta z\Delta x + c_7\Delta x\Delta y\Delta z \quad (24)$$

where Δx, Δy and Δz are the relative distances of the point with respect to the starting point P_{000} in the x, y, and z directions, respectively, as shown in equation 25-27.

$$\Delta x = (x - x_0)/(x_1 - x_0) \quad (25)$$

$$\Delta y = (y - y_0)/(y_1 - y_0) \quad (26)$$

$$\Delta z = (z - z_0)/(z_1 - z_0) \quad (27)$$

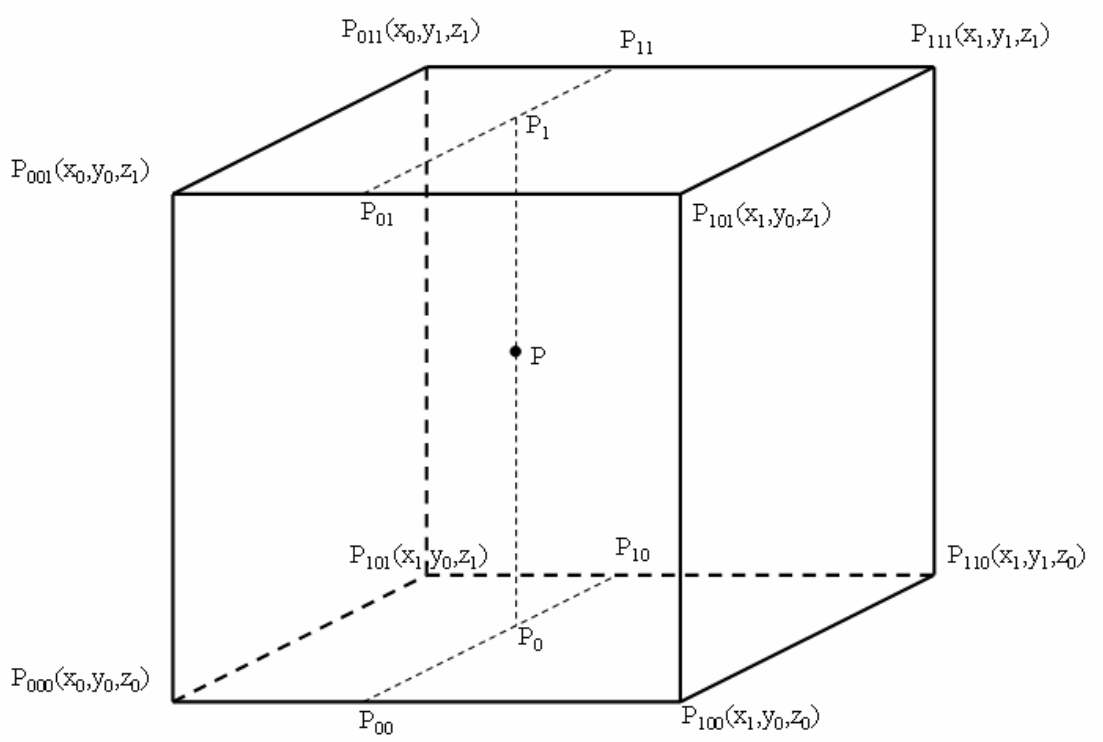


Figure 8 Trilinear interpolation.

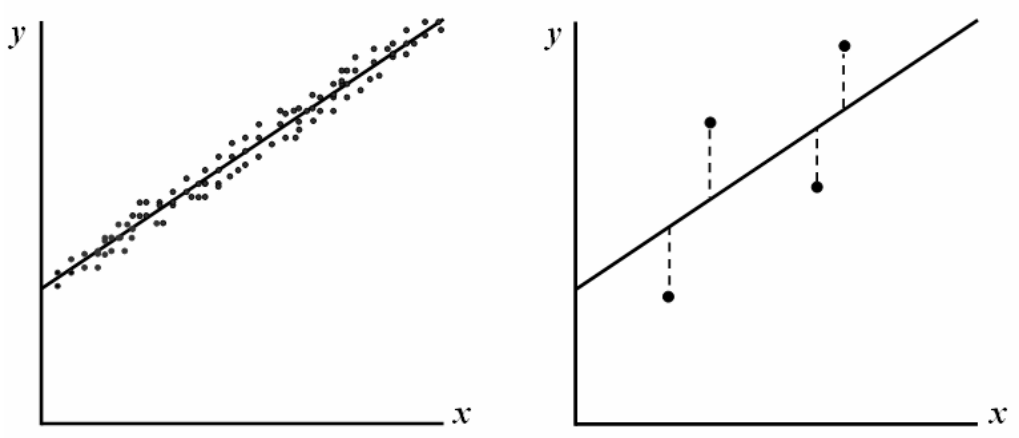
Coefficients c_j , when j= 0,1,2,3,...,7 are determined from the values of the vertices.

$$
\begin{aligned}
&c_0 = P_{000};\ c_1 = (P_{100} - P_{000});\\
&c_2 = (P_{010} + P_{000});\ c_3 = (P_{001} - P_{000});\ c_4 = (P_{110} - P_{010} - P_{100} + P_{000});\\
&c_5 = (P_{011} - P_{001} - P_{010} + P_{000});\ c_6 = (P_{101} - P_{001} - P_{100} + P_{000});\\
&c_7 = (P_{111} - P_{011} - P_{101} - P_{110} + P_{100} + P_{001} + P_{010} - P_{000});
\end{aligned}
\tag{28}
$$

2.2 Linear Regression Analysis

A mathematical procedure has been used finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets of the points from the curve see example in Figure 9. The linear least squares fitting technique is

the simplest and most commonly applied form of linear regression and provides a solution to the problem of finding the best fitting straight line through a set of points. In fact, if the functional relationship between the two quantities being graphed is known to within additive or multiplicative constants. The formulas for linear least squares fitting were independently derived by Gauss and Legendre.



a) b)

Figure 9 Least squares fitting; a) Example of the best linear fitting curve, b) The vertical offsets from a line.

Vertical least squares fitting proceeds by finding the sum of the squares of the vertical deviations R^2 of a set of n data points from a function f.

$$R^2 \equiv \sum [y_i - f(x_i, a_1, a_2, ..., a_n)]^2 \tag{29}$$

The condition for R^2 to be a minimum is that

$$\frac{\partial (R^2)}{\partial a_i} = 0 \tag{30}$$

For $i = 1,2,3,\ldots,n$ For a linear fit,

$$f(a,b) = a + bx \tag{31}$$

So that,

$$R^2(a,b) \equiv \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 \tag{32}$$

$$\frac{\partial(R^2)}{\partial a} = -2\sum_{i=1}^{n} [y_i - (a + bx_i)] = 0 \tag{33}$$

$$\frac{\partial(R^2)}{\partial b} = -2\sum_{i=1}^{n} [y_i - (a + bx_i)]x_i = 0 \tag{34}$$

These lead to the equations

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{35}$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \tag{36}$$

Solving equations 35-36 is given

$$a = \frac{\bar{y}(\sum_{i=1}^{n} x_i^2) - \bar{x}\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{37}$$

$$b = \frac{(\sum_{i=1}^{n} x_i y_i) - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{38}$$

Subunit *a* and *b* values from equation 37-38 into equation 31 will be give the fitted values *f(a,b)*.

2.3 Correlation Coefficient

The correlation coefficient, sometimes also called the cross-correlation coefficient, is a quantity that gives the quality of a least squares fitting to the original data. To define the correlation coefficient, first consider the sum of squared values SS_{xx}, SS_{xy} and SS_{yy} of a set of n data points (x_i,y_i) about their respective means,

$$r^2 = \frac{SS_{xy}^2}{SS_{xx}SS_{yy}} \tag{39}$$

When

$$\begin{aligned} SS_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2 \end{aligned} \tag{40}$$

$$\begin{aligned} SS_{yy} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= \left(\sum_{i=1}^{n} y_i^2\right) - n\bar{y}^2 \end{aligned} \tag{41}$$

$$\begin{aligned} SS_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= \left(\sum_{i=1}^{n} x_i y_i\right) - n\overline{xy} \end{aligned} \tag{42}$$

In theory the correlation coefficient falls between 0 and 1, where 1 corresponds to a perfect correlation and zero corresponds to total disorder. As depicted in the Figure 10, the examples of the linear fitting curve of two data that gave the different correlation coefficient.

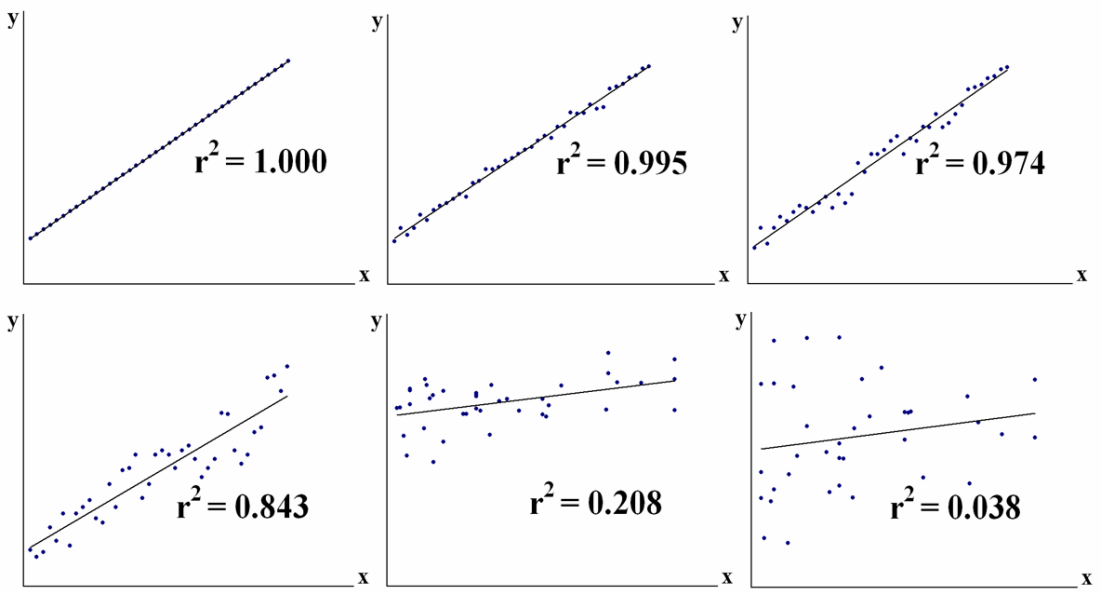


Figure 10 Example of the linear fitting curve of different correlation coefficient values.

Computational Details of Calculations

1. Part I: Investigate the Feasibility of Using AutoDock 3.0.5 Program as a Tool of Virtual Screening

The aim of this work is to assess the usability of the nonprofit free AutoDock program; we performed docking studies on 24 NA complexes. Set of NA-inhibitors complexes known K_i (inhibitory constant) or IC_{50} (inhibitory concentration 50%) values were taking from literature review which have been reported by Wang and Wade in 2001. Before docking calculation, AutoDock requires pre-calculated grid maps for each atom type present in the ligand being docked. This helps to make the docking calculations extremely fast. These maps are calculated by AutoGrid. A grid map consists of a three dimensional lattice of regularly spaced points, surrounding (either entirely or partly) and centered on some region of interest of the macromolecule under study. Typical grid point spacing varies from 0.2 Å to 1.0 Å, although the default is 0.375 Å (roughly a quarter of the length of a carbon-carbon single bond). Each point within the grid map stores the potential energy of a 'probe' atom or functional group that is due to all the atoms in the macromolecule, are shown in Figure 4. Therefore, before docking with AutoDock program will be prepared protein and ligand for calculate the interaction energy.

1.1 Preparation of 24 NA and inhibitor complexes coordinates

The X-ray structures of the 24 co-crystal of inhibitors and neuraminidases (NA type N2 and N9) were obtained from the Protein Data Bank as shown in Table 1. The structures of the target proteins and the ligand molecules were saved in PDB format. The AutoDock tools package was employed to generate the docking input files and analyze the docking results. The docking procedure was prepared the enzyme grid by removing ligand and water molecules from protein complexes. Then the protein structures were added with only polar hydrogen atoms follow by adding Kollman charges (charges of amino acid). The charged proteins were then solvated using the AutoDock3.0 "addsol" module. Those enzymes were ready as an input for enzyme grid generating. Another separated routine was to

prepare a ligand by adding hydrogen, charges, and assigning the rotatable bonds for the ligand. We applied Gasteiger charge (charges of small compounds) and maximum number of rotatable bonds on ligands throughout the experiments. Both preprocessed ligand and enzyme grid were required for generating of the dock parameter file where required parameters were set. Most NA inhibitors (NIs) currently known are sialic acid (Neu5Ac) analogues (Gubreva *et al.*, 2000; Kim *et al.*, 1998) and known IC_{50} values of a set of NA inhibitor complexes, are shown in Table 1

Table 1 PDB code (NA type), ligand name, ligand structures and experimental inhibitory potencies of the 24 NA complexes.

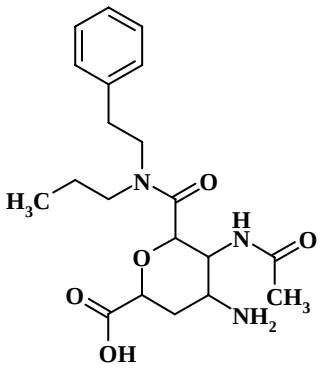
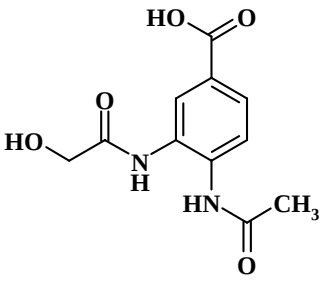
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
1	1BJI (N9)	G21		8.70
2	1ING (N2)	ST5		2.40

Table 1 (Continued)

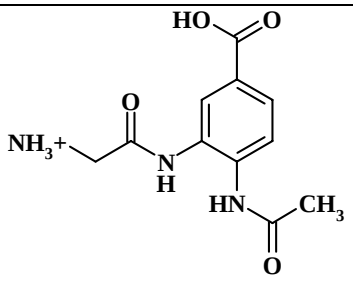
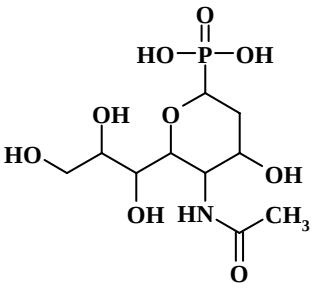
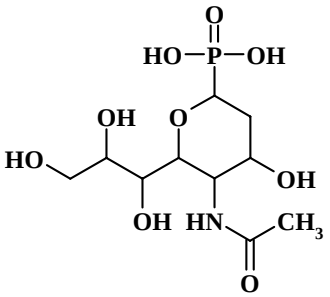
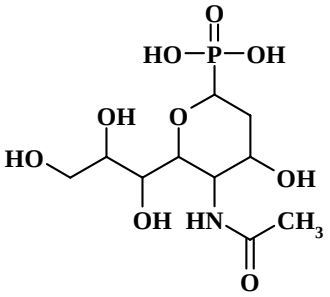
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
3	1INH (N2)	ST6		2.30
4	1INW (N2)	aPANA		2.70
5	1INX (N2)	ePANA		4.70
6	1INY (N2)	ePANA		3.16

Table 1 (Continued)

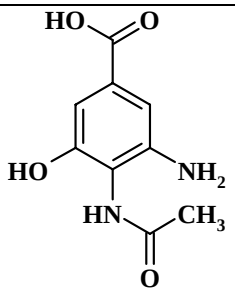
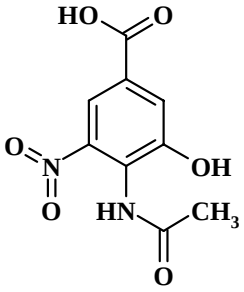
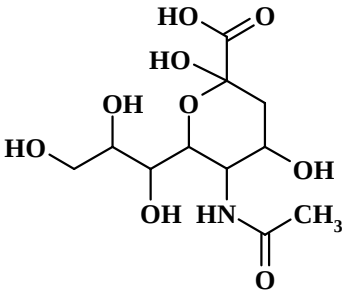
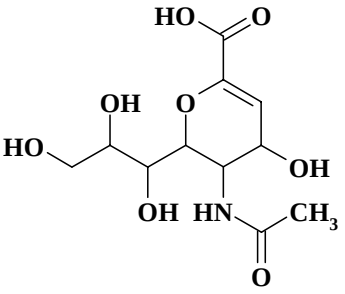
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
7	1IVC (N2)	ST2		1.70
8	1IVD (N2)	ST1		3.12
9	1MWE (N9)	α-Neu5Ac		1.70
10	1NNB (N9)	Neu5Ac2en		4.70

Table 1 (Continued)

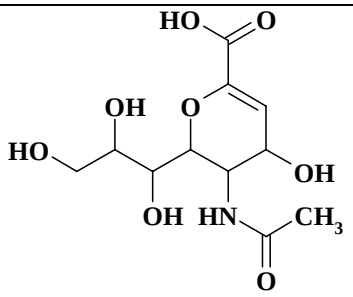
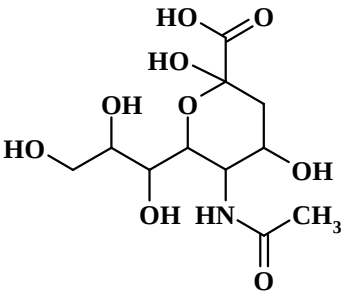
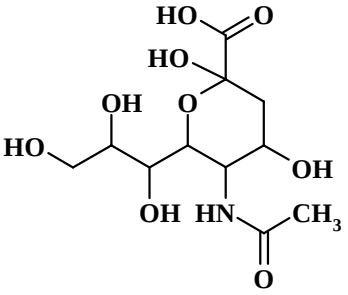
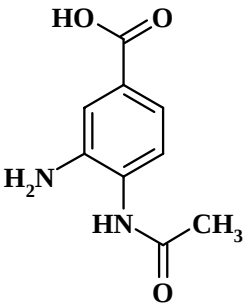
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
11	1NNC (N9)	GNA		8.70
12	2BAT (N2)	α-Neu5Ac		2.70
13	2QWB (N9mutant)	α-Neu5Ac		0.70
14	1IVE (N2)	ST3		1.40

Table 1 (Continued)

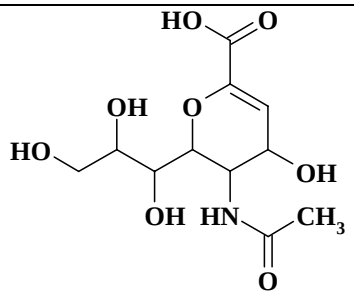
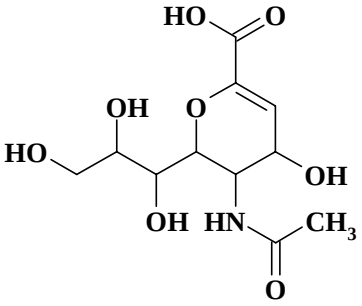
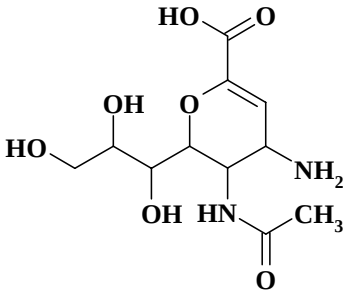
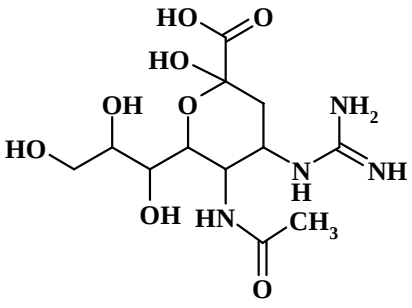
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
15	1IVF (N2)	Neu5Ac2en		4.82
16	2QWC (N9mutant)	Neu5Ac2en		3.39
17	2QWD (N9mutant)	4AM		4.00
18	2QWE (N9mutant)	GNA		6.96

Table 1 (Continued)

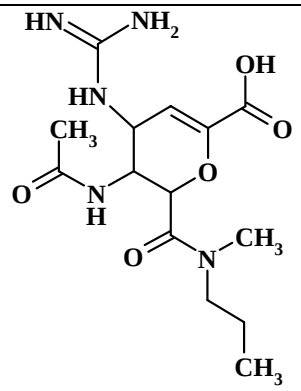
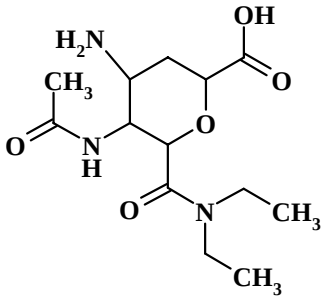
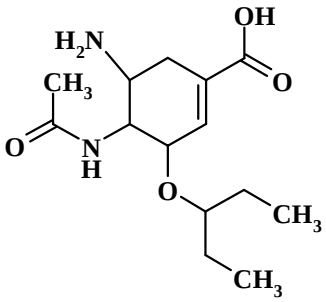
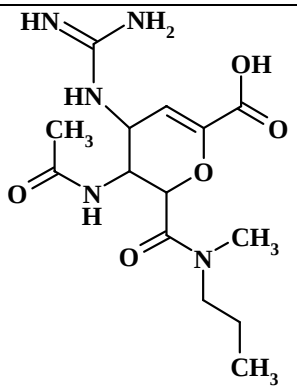
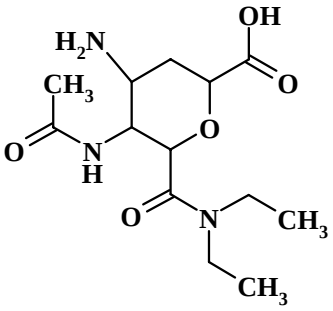
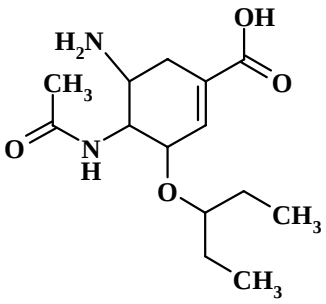
No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
19	2QWF (N9mutant)	G20		5.28
20	2QWG (N9mutant)	G28		3.64
21	2QWH (N9mutant)	G39		4.89

Table 1 (Continued)

No.	PDB code (NA type)	Ligand name	Ligand structure	pIC_{50}
22	2QWI (N9)	G20		7.70
23	2QWJ (N9)	G28		6.64
24	2QWK (N9)	G39		8.70

1.2 Validation of the method

In this part, we calculated separating into two steps. In the first step, we were reproduced the 24 co-crystals of NA and inhibitors. After that, the grid maps that results the lowest root-mean-square deviation (rmsd) values (hereafter referred to the best rmsd values) between predicted docked conformation and the x-ray structures

were selected to do the cross-validation experiment step. In the cross-validation step, we docked all NA inhibitors (NIs) into each NA binding pocket. The correlation coefficients between predicted values and experimental data will be used to assess the ability of each method. In this work, we validated in three methods as following.

1.2.1 Method I: Enzyme grid maps generated with grid spacing 0.375 Å.

The first step of this method were tried to reproduce the 24 co-crystal complexes using AutoDock program version 3.0.5 in order to calculate docked energy, binding free energy and the root-mean-square deviation (rmsd) value. In the reproduction of 24 co-crystals of NA and inhibitors, enzyme grid maps were generate with grid spacing 0.375 Å (the default parameters of the AutoGrid program) and vary grid box dimensions in four type (90×90×90, 100×100×100, 110×110×110, and 120×120×120 points). As the center of the common grids, we used the center on ligand that had been removed from the binding site of the target protein under consideration. When calculated the enzyme grid maps were completed, the AutoDock program could run. Among several parameters set in the dock parameter file, the optimization algorithm that generated the best conformations by the Lamarckian genetic algorithm in AutoDock 3.0.5 was chosen to search the best 50 conformations and orientation of space of the inhibitors while keeping the protein structure rigid. In this study used searching the best 50 conformations because had been reported that it's the good results for docking with AutoDock program. After that, the grid maps that yielded the best rmsd values between docked conformation and the x-ray structures were selected to do the cross validation step.

The second step was cross-validated 24 co-crystals in order to see the calculated free energy of binding and docked energy prediction when ligands did not belong to the corresponding NA. In this step, all NIs were fully optimized using the semi-empirical method (PM3) with GAMESS program to search the optimize structure in free space before docking to macromolecules.

1.2.2 Method II: The Van der Waals coefficients and the electrostatic energy will be scaled by 0.1000 and 0.0100 respectively.

The ΔG terms on the right-hand side of equation 10 in the original AutoDock 3.0 scoring function are coefficients empirically determined using linear regression analysis from a set of 140 protein-ligand complexes with known binding constants. For the Van der Waals coefficient (ΔG_{vdW}) will be scaled by 0.1485, the hydrogen bonding coefficient (ΔG_{hbond}) will be scaled by 0.0656 and the electrostatic energy coefficient (ΔG_{elec}) will be scaled by 0.1146. For loss of torsional degrees of freedom upon binding coefficient (ΔG_{tor}) and the ligand desolvation free energy coefficient (ΔG_{tor}) will be scaled by 0.3113 and 0.1711 respectively. However, the calculation by using original scoring function in AutoDock program for method I did not give a good correlation between predicted values and experimental data. Then, this method tested changing the scaling coefficients of the Van der Waals and the electrostatic energy term in the AutoDock scoring function. For the Van der Waals coefficients, we scaled with 0.1000. The electrostatic energy scaled with 0.0100. Those coefficients have been reported that success on QSARs study for NA systems (Wang and Wade, 2001). In the practice, these parameters were changing in GPF file which is input file for generate grid maps.

1.2.3 Method III: Enzyme grid maps generated by varying grid point dimension and grid spacing

In this method, the calculations were done by using the similar procedure in the first method. But, this method we reproduced each ligand in to the native protein to calculate docked energy, binding free energy and the root-mean-square deviation (rmsd) value by using new grid map parameters. For method III, enzyme grid maps generated with grid box dimensions of 90×90×90, 100×100×100, 110×110×110, 120×120×120 points and each grid box dimensions were varying grid point spacing of 0.275 Å, 0.325 Å, 0.375 Å, 0.425 Å, and 0.475 Å. The same as

method I, the grid maps that yielded the best rmsd values between docking and the x-ray structures of each complex has been selected to do the second step.

In addition, the many output of 24 co-crystal reproduction with AutoDock program were rescored using six different scoring functions in the FRED program, and then the best docked conformation that lowest binding score form 50 docked conformations is considered. Comparisons between the lowest binding score docked conformations by FRED scoring functions and the best rmsd docked conformation.

In the second step, the cross validation 24 co-crystals were calculated free energy of binding and docked energy prediction when ligands did not belong to the corresponding NA by using gird maps that lowest the rmsd from the first step. After that, all docking results from AutoDock program were rescored using various scoring functions in FRED program. Finally, the enzyme grid map that show the highest correlation between pIC_{50} of data set prediction and pIC_{50} of the experiment data have been selected to virtual screening step. In the virtual screening will be used the best pocket and the suitable procedure to screen the small compounds on database (ChemeiBase, Kasetsart University) or other unknown set.

1.3 Prediction of Inhibitory Potencies

Predicted pIC_{50} can be calculated from the binding free energy or dock energy value (ΔG) as followed:

$$\Delta G = -RT \ln IC_{50} \tag{43}$$

When pIC_{50} = -log IC_{50} (Masukawa *et al.,* 2003).

2. Part II: Virtual screening using AutoDock and FRED program

A database of around 30,000 compounds was obtained from ChemieBase database. The chemical structures in the Thai plants are collected in this database. The 2D structure of small compounds on database were generated 2D structure using ISIS Draw program in mol format and converted 2D to 3D structures in mol2 format using CORINA program. After that, all compounds were fully optimized using the semi-empirical method (PM3) with GAMESS program.

2.1 Preparation of the screening compounds library

The optimize structures of small compounds were taken from ChemieBase database (30,000 compounds) and the benzoic acid derivatives (47 compounds) which synthesized by Dr. Warinthorn Chavasiri, Chulalongkorn University. The coordinates of unknown set were saved in mol2 format. We assign the maximum number of rotatable bonds for all ligands using the autotors module of the AutoDock program.

2.2 Docking of unknown set compounds

From Part I, the results of validation methods were found that the third method gave the best result between the prediction values and experimental data. The data set from PDB code 2QWI showed highest correlation coefficient base on PLP scoring function (see in Results and Discussion section). Therefore, the protein target PDB code 2QWI and procedure of third method were selected to do virtual screening for anti-influenza drugs.

All ligands were docked into pocket of NA, using enzyme grid maps of PDB code 2QWI. After that, the docking results from AutoDock program were rescored using various scoring function in FRED program. The compound of top rank in PLP scoring function is considered to test biological activity assay.

RESULTS AND DISCUSSION

1. Part I: Investigate the Feasibility of Using AutoDock Program as a Tool for Virtual Screening

To date, a number of crystal structures of NA and inhibitor complexes have been released, in which different kinds of NA inhibitors (NIs) bound to the active site in a similar pattern (Lew *et al.*, 2000; Zheng *et al.*, 2006). Investigation on these structures also showed that there are many co-crystals NA-inhibitor complexes which is virtually no change in the orientation of the side chain amino acids of the NA active site, except for some minor conformational changes in Glu276 and Arg224, Figure 11 (Moscona, 2005). From these reason, this studies validated methods to find the best protein pocket from 24 protein structures of NA and a suitable procedure for virtual screening for anti-influenza drugs. Moreover, we want to assess the feasibility of using AutoDock program as a tool for virtual screening. Therefore, in this part the protein structure that gave the best result of data set between the predicted values and experimental data were selected to doing virtual screening in Part II.

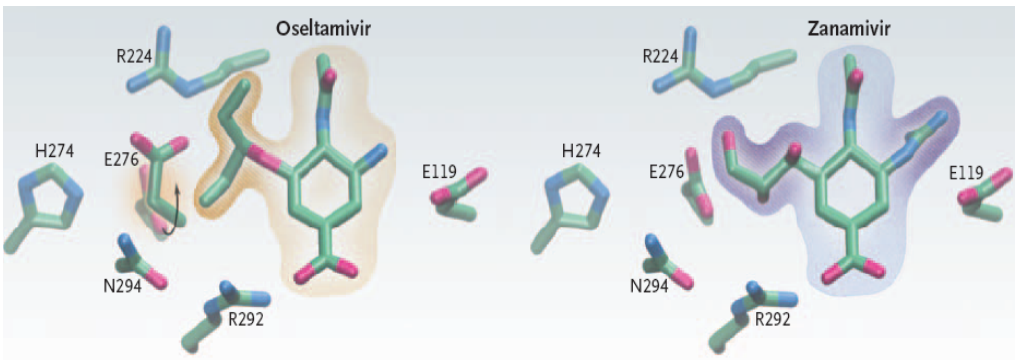


Figure 11 The neuraminidase active site changes shape to create a pocket for oseltamivir and zanamivir.

Source: Moscona (2005)

1.1 Validation of the method

In this part, the results of each method were explained as in the following.

1.1.1 Method I: Enzyme grid maps generated with grid spacing 0.375 Å

For method I, AutoDock program has been used to reproduce the 24 co-crystals of NA and inhibitors. The reproduction used grid spacing 0.375 Å with out varied, of 90×90×90, 100×100×100, 110× 110×110 and 120×120×120 box dimensions in order to calculate binding free energy, docked energy and rmsd values. Results of reproduction are shown in Table 2. The rmsd varied from 0.78 to 1.84 Å. The correlation coefficient (r^2) of experimental inhibitory potencies (pIC_{50}) and predicted pIC_{50} of binding free energy which is converted by equation 22 was 0.54 as shown in Figure 12. However, the correlation of experimental inhibitory potencies (pIC_{50}) and predicted pIC_{50} of docked energy was 0.34 as shown in Figure 13. The grid maps having the best rmsd value of reproduction have been selected to do the cross-validation step.

In Table 2, column 1 shows the PDB code of protein structure from Protein Data Bank. Column 2 is inhibitor name which it's corresponding with protein structure PDB code in column 1. Experimental pIC_{50} were taken from report of Wang and Wade (2001). Column 4, 5 and 6 show predicted binding free energy, docked energy and rmsd values. Finally, column 7 shows a suitable grid point (herein represent with one grid dimension) for each protein structure at grid spacing 0.375 Å.

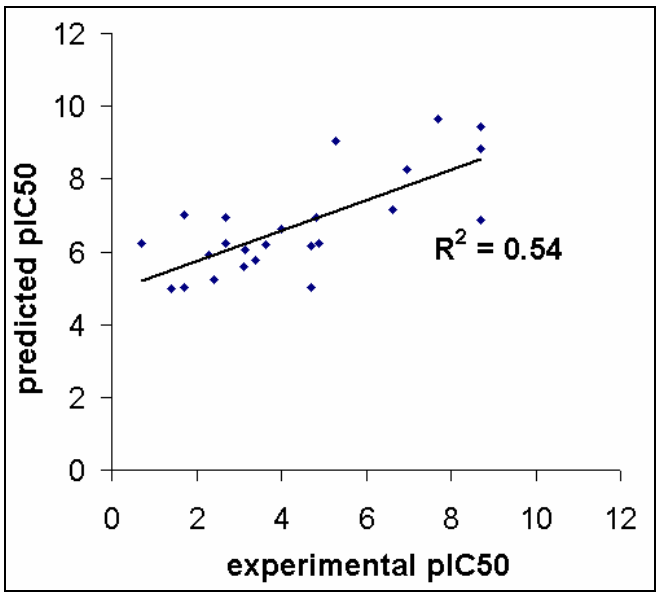


Figure 12 Plot of experimental pIC_{50} values and predicted pIC_{50} of 24 co-crystal reproduction given by lowest binding free energy, method I.

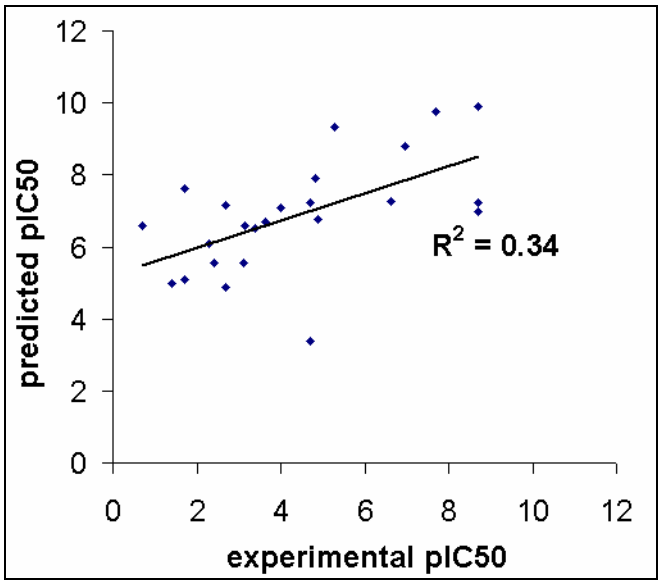


Figure 13 Plot of experimental pIC_{50} values and predicted pIC_{50} of 24 co-crystal reproduction given by lowest docked energy, method I.

Cross-validation step of method I, when we plots between experimental inhibitory potencies (pIC_{50}) and predicted pIC_{50} from docking all NIs into each of binding pocket of NA 24 PDB code (hereafter called data set and represent with protein PDB code). The correlation coefficients given by lowest binding free energy and docked energy are summarized in Table 3. In Table 3, most data set from 24 protein binding pocket which represent with protein PDB code showed the correlation coefficients (r^2) lower than 0.60. The r^2 values ranging between 0.04 and 0.66 based on predicted binding free energy and ranging between 0.02 and 0.63 based on docked energy. The data set from PDB code 2QWI was highest correlation in this set was 0.66 in term of binding free energy and 0.61 in term of docked energy. A second rank, the data set from PDB code 2QWD showed correlation was 0.63 in term of docked energy and 0.62 in term of binding free energy. All of the other data set gave very poor correlation yielded in the both of binding free energy and docked energy term. From these results in cross-validation step, we found that data set from protein structure PDB code 2QWI gave the best correlation between predicted values and experimental data ($r^2 = 0.66$). However, when compared these results with the previous results of reproduction in this method. We found the correlation of co-crystals reproduction step showed that AutoDock failed to reproduce for these systems because the r^2 values were lower than 0.60 (0.54 and 0.34). For this reason, we concluded that enzyme grid maps affinities in method I did not a suitable grid maps parameter for virtual screening for anti-influenza drugs. Therefore, calculation of enzyme grid maps affinity was improved to increase correlation.

Table 3 Correlation between the experimental pIC_{50} and the predicted pIC_{50} of cross-validation step in method I.

PDB code	Correlation coefficients (r^2)	
	Binding free energy	Docked energy
1BJI	0.57	0.24
1ING	0.20	0.13
1INH	0.52	0.45
1INW	0.45	0.28
1INX	0.36	0.18
1INY	0.49	0.49
1IVC	0.39	0.39
1IVD	0.04	0.02
1IVE	0.31	0.17
1IVF	0.57	0.48
1MWE	0.57	0.52
1NNB	0.53	0.48
1NNC	0.58	0.51
2BAT	0.39	0.40
2QWB	0.45	0.42
2QWC	0.57	0.48
2QWD	0.62	0.63
2QWE	0.53	0.59
2QWF	0.47	0.51
2QWG	0.54	0.45
2QWH	0.50	0.33
2QWI	0.66	0.61
2QWJ	0.56	0.43
2QWK	0.57	0.48

1.1.2 The Van der Waals coefficients and the electrostatic energy will be scaled by 0.1000 and 0.0100 respectively

In this method, we chose three protein structures PDB code by random from data set of protein structure PDB code that gave high correlation coefficient between experimental data and predicted values in cross-validation step of method I. The protein structure PDB code 1BJI, 2QWC, and 2QWD were selected to generate the new grid maps by new scaling coefficients of Van der Waals and electrostatic energy (0.1000 and 0.0100, respectively). When docked all NIs into three binding pockets. The correlation coefficient of data set from PDB code 1BJI, 2QWC and 2QWI are shown in Table 4, in the bracket were r^2 values from method I and PDB code in column 1 represent the data set from each protein PDB code. Data set of three PDB code showed poor correlation. Therefore, we concluded that this method did not a suitable method for NA system.

Table 4 Correlation between the experimental pIC_{50} and the predicted pIC_{50} from method II.

	Correlation coefficients (r^2)	
PDB code	Binding free energy	Docked energy
1BJI	0.38 (0.57)	0.33 (0.24)
2QWC	0.38 (0.57)	0.32 (0.48)
2QWI	0.44 (0.66)	0.49 (0.61)

1.1.3 Enzyme grid maps generated by varying grid point dimension and grid spacing

In this method, the enzyme grid map were generated using five different values of grid spacing point of 0.275 Å, 0.325 Å, 0.375 Å, 0.425 Å, and 0.475Å for each of grid point dimensions of 90×90×90, 100×100×100, 110× 110×110 and 120×120×120 point, were similar with method I. But, method I did not varied grid

spacing (fixed at 0.375 Å). The results of 24 co-crystal reproduction from method III are shown in Table 5. The best rmsd ranging between 0.48 Å and 1.35 Å. Suitable grid point and spacing of each protein structure are shown in column 2 and 3 respectively.

When plotted between experimental pIC_{50} and predicted pIC_{50} which is converted by equation 22. It showed the correlation coefficients of 0.72 and 0.71 bases on binding free energy and docked energy respectively as shown in Figure 14 and Figure 15. These correlation values are acceptable ($r^2 \geq 0.60$). Moreover, when we compared the rmsd value between Table 2 and Table 5, we found that the most rmsd values in Table 5 are lower than those in Table 2. These results showed that the box size and grid spacing inside the docking box were very important for generate a suitable grid maps of each protein structure before docking and increase the accuracy of docking results. When compared the correlation coefficients of reproduction step from method I and method III. We found that the method III gave r^2 values higher than method I. These results indicated that AutoDock was competent in reproducing the experimentally found binding position and conformation of NIs.

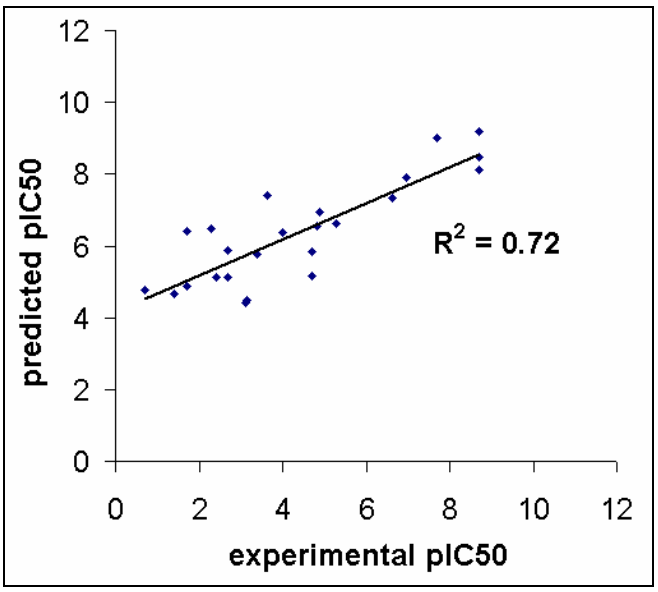


Figure 14 Plot of experimental pIC_{50} values and predicted pIC_{50} of 24 co-crystal reproduction given by lowest binding free energy, method III.

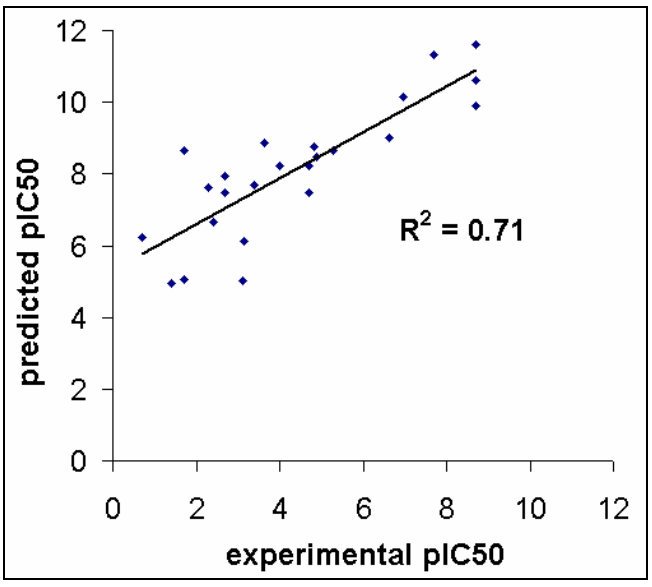


Figure 15 Plot of experimental pIC_{50} values and predicted pIC_{50} of 24 co-crystal reproduction given by lowest docked energy, method III.

In method III, the docked conformations or docking results from 24 co-crystals of reproduction which generated by AutoDock program were tested by

scoring with scoring function's FRED program (Chemgauss2, Chemscore, PLP, Screenscore, Shapegauss and consensus). For each complex have fifty docked conformations which cover the entire binding pocket as shown in Figure 16. The best docked conformation from fifty docked conformations of each docking complex was selected by the lowest binding score of each scoring function in FRED program. When we know the best docked conformation, we could take the binding free energy and docked energy from AutoDock scoring function and plots the predicted pIC_{50} binding free energy and docked energy of prediction with the experimental pIC_{50}. The correlation coefficient (r^2) showed in Table 6.

Table 6 Correlation between the experimental pIC_{50} and the predicted pIC_{50} of reproduction step in method I using FRED selected the best docked conformations.

	Correlation coefficients (r^2)	
Scoring function	Binding free energy	Docked energy
Chemgauss2	0.54	0.62
Chemscore	0.56	0.57
PLP score	0.70	0.69
Screenscore	0.67	0.73
Shapegauss	0.32	0.42
consensus	0.61	0.64

In Table 6, scoring functions still produced correlation coefficient higher than 0.60 including PLP, Screenscore and consensus scoring function. These scoring functions showed a fair correlation between prediction values and experimental data (r^2 = 0.61 - 0.73). PLP scoring function gave the best correlation coefficient of 0.70 bases on binding free energy and Screenccore scoring function gave the best correlation coefficient of 0.73 bases on docked energy. All of other scoring function gave poor correlation. Moreover, when we considered cluster rank of the lowest binding score conformer (hereafter referred to the best binding score) by

various scoring function in FRED program and the lowest rmsd conformer (herein referred to the best rmsd), we are found that these docked conformation does not coincide. However, these scoring function were selected the best docked conformer in the same cluster rank with the best rmsd conformer as shown in Figure 17. The best rmsd conformer are superimposed, best binding score of scoring function in FRED program and the x-ray structure were found that the orientation of each ligand docking pose within a targeted binding site are very similar with x-ray structure pose, as shown in Figure 18 and Figure 19. In Figure 19 compared the position of G20 from AutoDock of the best rmsd conformer, the best binding score conformer using PLP score in FRED program and x-ray structure in active site of NA PDB code 2QWI. Moreover, Figure 19 showed the important amino acid residues in binding pocket of NA. These results showed that we can use various scoring function in FRED program to select the best docked conformation from the many docked conformation from AutoDock program for NA system.

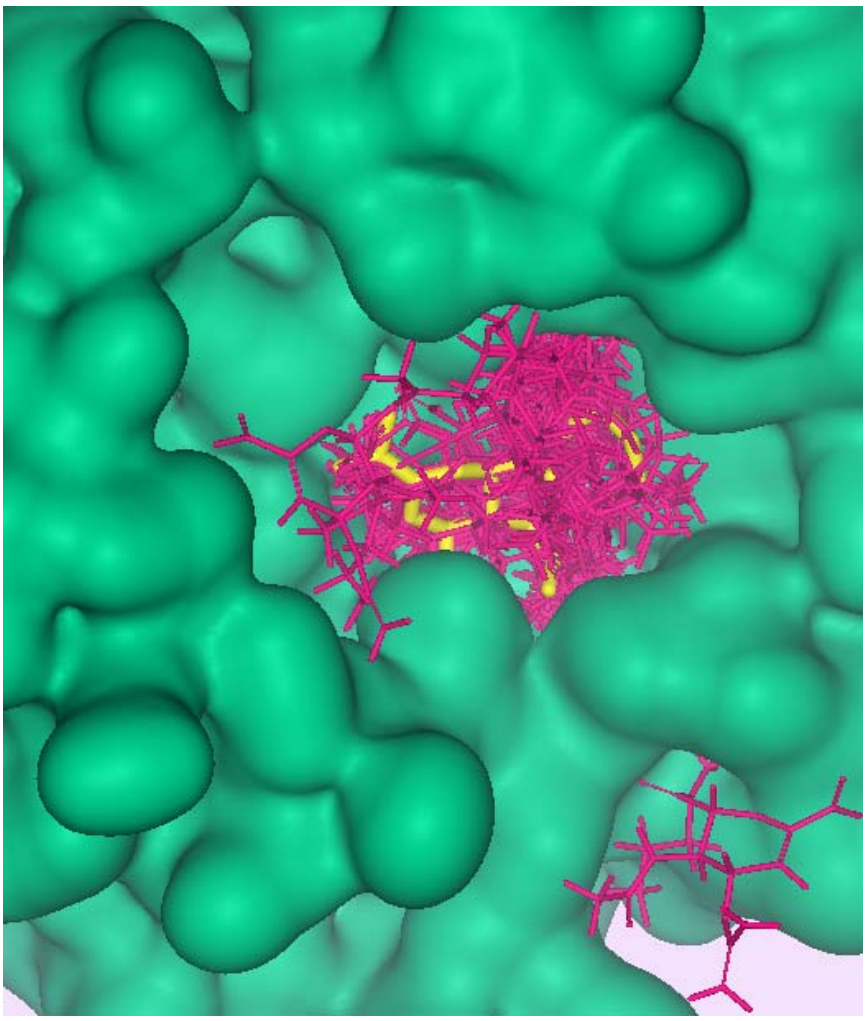


Figure 16 The fifty docked conformations ensemble of G20 in active site of NA PDB code 2QWI generated by AutoDock program (pink), and x-ray structure of inhibitor (yellow).

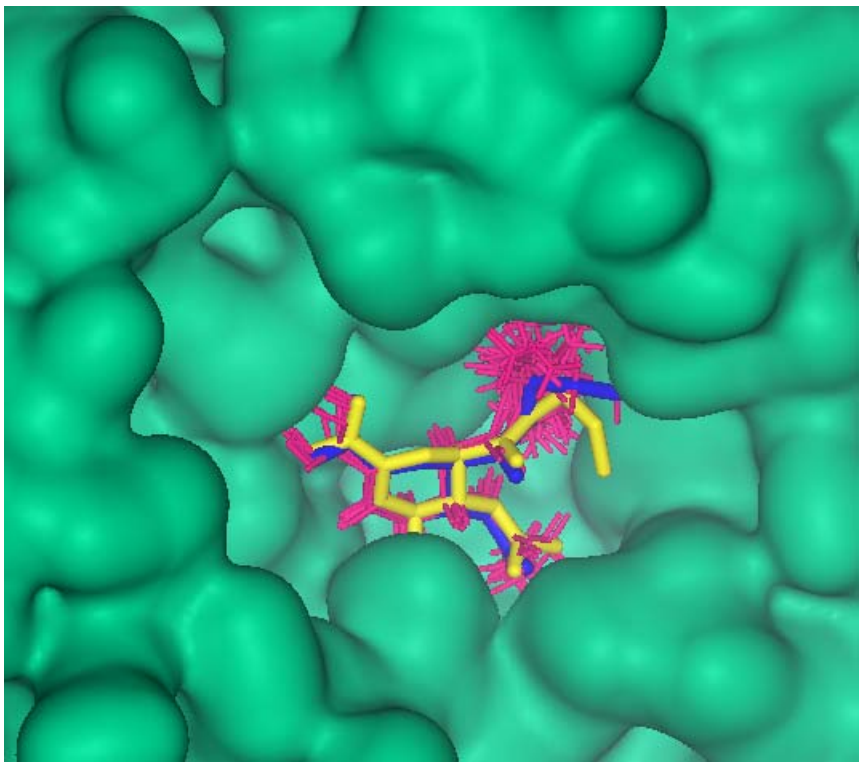


Figure 17 The orientation of docked conformations (pink) from AutoDock in the same cluster rank with the best rmsd docked conformations (blue) and x-ray structure of inhibitor (yellow) in binding pocket of NA.

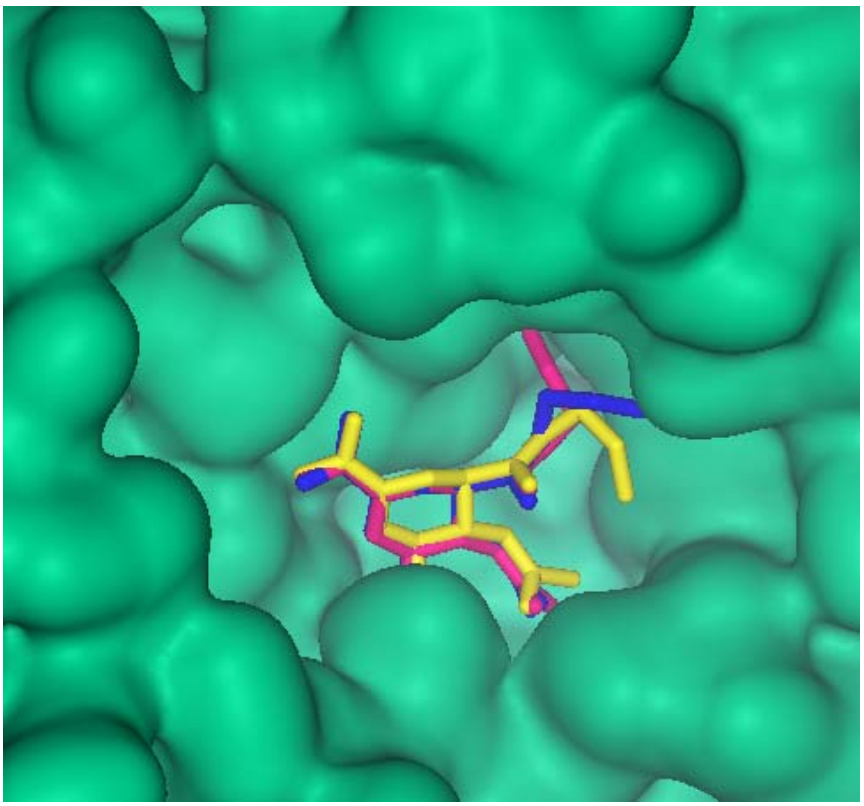


Figure 18 Comparison the position of G20, best binding score docked conformation (pink), the best rmsd docked conformation (blue) and x-ray structure of inhibitor (yellow) in binding pocket of NA.

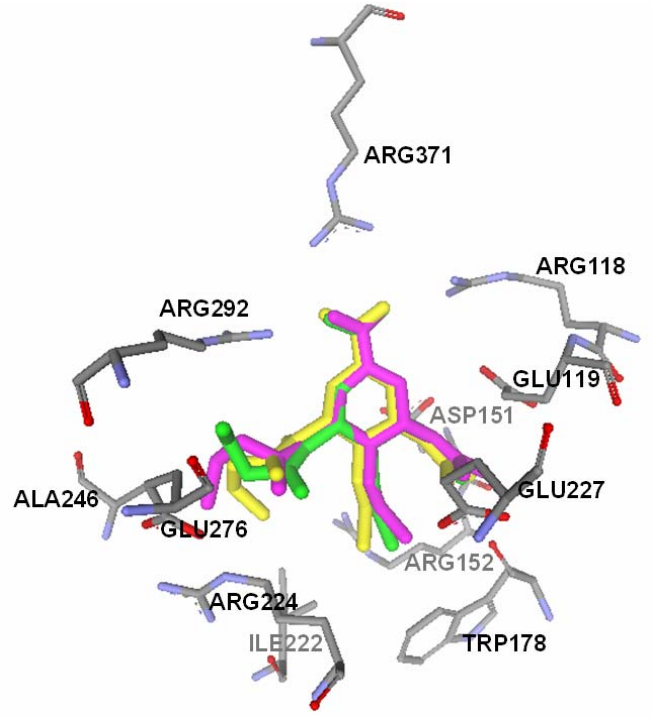


Figure 19 Comparison of the position of G20 calculates by AutoDock 3.0.5 program, best rmsd conformer (pink), the best binding score conformer using PLP score in FRED program (green) and x-ray structure (yellow) in active site of NA.

Doing the same as in the method I (calculated grid maps using 0.375 grid spacing), enzyme grid maps that yield the best rmsd value of reproduction in this method, as shown in Table 5, is selected to do the cross-validation.

In cross-validation step of method III, the results of all NIs docked into each protein binding pocket called data set from each PDB code (in these Table represented with protein structure PDB code). When we plotted between experimental pIC_{50} and predicted pIC_{50} of each data set. The correlation coefficients are shown in Table 7 - 9.

Table 7 Correlation between the experimental pIC_{50} and the predicted pIC_{50} of cross-validation step in method III.

PDB code	correlation coefficients (r^2)	
	Binding free energy	Docked energy
1BJI	0.48	0.60
1ING	0.24	0.14
1INH	0.49	0.49
1INW	0.35	0.51
1INX	0.30	0.42
1INY	0.35	0.53
1IVC	0.23	0.50
1IVD	0.32	0.43
1IVE	0.21	0.45
1IVF	0.37	0.44
1MWE	0.49	0.49
1NNB	0.48	0.56
1NNC	0.54	0.54
2BAT	0.47	0.52
2QWB	0.47	0.54
2QWC	0.49	0.56
2QWD	0.51	0.58
2QWE	0.55	0.61
2QWF	0.53	0.64
2QWG	0.53	0.65
2QWH	0.13	0.44
2QWI	0.63	0.61
2QWJ	0.54	0.53
2QWK	0.46	0.51

Table 7 showed the correlation coefficients given by lowest binding free energy and docked energy respectively. Most data set gave the poor correlation, except for data set from PDB code 2QWI gave correlation higher than 0.60 in the both of predicted binding free energy and docked energy term (0.63 and 0.61respectively). All of other data set, some data set gave correlation higher than

0.60 from only term of docked energy such as data set from PDB code 1BJI, 2QWE, 2QWF and 2QWG.

Moreover, the results of all data set from cross-validation step were rescored by six scoring function in FRED program to select the best docked conformation from fifty docked conformation. Doing similar with reproduction step in this method (method III), when we know the best docked conformation, we could take the binding free energy and docked energy from AutoDock scoring function and plots the predicted pIC_{50} binding free energy and docked energy with the experimental pIC_{50}. The correlation coefficients (r^2) given by using six scoring function in FRED program selected the best docked conformation are shown in Table 8.

In Table 8, most data set gave poor correlation, except for some data set gave correlation higher than 0.60 in the both of predicted binding free energy and docked energy term. Data set from PDB code 1BJI gave r^2 = 0.61 and 0.66 by Chemgauss2 scoring function, 1NNB are showed 0.60 and 0.61 by consensus score, 2QWF are showed high r^2 base on PLP and Shapegauss scoring function, and finally 2QWI was good correlation base on consensus score and lowest docked energy. However, these results did not identify that this method was a suitable procedure for virtual screening for anti-Influenza drugs. Due to, there are many data set, *i. e.*, data set from PDB code 1BJI, 1NNB, 2QWFand 2QWI are showed the correlation values higher than 0.60 and these data set gave the best correlation by different scoring function in FRED program (see Table 7 and Table 8). When compared these results with previous results in reproduction step. We found that these results do not coincide with previous results in reproduction step. However, data set from PDB code 2QWI gave correlation higher than 0.60 from all step in method III.

Filnally, when we plotted between experimental pIC_{50} and binding score of six scoring functions of FRED program. We found that some data set of NA gave a good correlation between the predicted binding score from scoring functions in FRED program and experimental inhibitory potencies (pIC_{50}). The correlation coefficients given by various scoring function in FRED program are listed in Table 9.

In Table 9, among of six scoring function in FRED program, PLP scoring function gave the best agreement between its binding score and experimental inhibitory potencies (pIC_{50}) with correlation coefficient of 0.71 from data set of PDB code 2QWI as shown in Figure 20. Moreover, there are some data set gave r^2 values higher than 0.60 such as data set from PDB code 1BJI ($r^2 = 0.70$), 1MWE ($r^2 = 0.61$), 2BAT ($r^2 = 0.64$), 2QWF ($r^2 = 0.65$) and 2QWJ ($r^2 = 0.61$) based on PLP scoring function, and the data set from PDB code 2QWB ($r^2 = 0.64$) and 2QWE ($r^2 = 0.68$) based on Screenscore scoring function. For all of other data set and other scoring functions gave vary poor correlations. When compared these results with previous results in third method. We found that the data set from protein structure PDB code 2QWI gave highest correlation values in the both of correlation plots by using lowest energy of AutoDock scoring function and by using the best binding score of PLP scoring function in FRED program with experimental pIC_{50}. Those results indicated that grid maps affinities of protein structure PDB code 2QWI from method III was a suitable grid maps parameter for NA system because its gave the best correlation between predicted values and experimental data. Therefore, enzyme grid maps of PDB code 2QWI and scoring procedure by PLP scoring function in FRED program were selected to do virtual screening procedure in the Part II. Moreover, those results indicated that we can used AutoDock program as a tool for virtual screening when using combine with PLP scoring function in FRED program.

Table 9 Correlation between experimental activities (-log IC_{50}) and binding scores of 24 protein-ligand complexes by various scoring functions in FRED program.

PDB code	Correlation coefficients (r^2)					
	Chemgauss2	Chemscore	PLP	Screenscore	Shapegauss	Consensus
1BJI	0.40	0.01	0.70	0.43	0.45	0.40
1ING	0.24	0.07	0.34	0.14	0.38	0.25
1INH	0.30	0.05	0.43	0.19	0.36	0.30
1INW	0.32	0.11	0.41	0.09	0.33	0.25
1INX	0.23	0.19	0.25	0.17	0.40	0.20
1INY	0.44	0.05	0.57	0.55	0.40	0.40
1IVC	0.38	0.11	0.19	0.08	0.29	0.25
1IVD	0.29	0.12	0.30	0.00	0.42	0.40
1IVE	0.17	0.00	0.02	0.02	0.41	0.38
1IVF	0.32	0.29	0.44	0.33	0.43	0.42
1MWE	0.40	0.04	0.61	0.48	0.49	0.56
1NNB	0.13	0.13	0.10	0.06	0.06	0.16
1NNC	0.12	0.16	0.09	0.05	0.18	0.21
2BAT	0.32	0.03	0.64	0.49	0.44	0.55
2QWB	0.35	0.04	0.54	0.64	0.36	0.42
2QWC	0.34	0.04	0.47	0.50	0.40	0.38
2QWD	0.30	0.02	0.54	0.48	0.55	0.49
2QWE	0.39	0.06	0.52	0.68	0.53	0.50
2QWF	0.31	0.02	0.65	0.56	0.48	0.52
2QWG	0.37	0.03	0.52	0.45	0.44	0.45
2QWH	0.33	0.07	0.32	0.33	0.37	0.33
2QWI	0.33	0.01	0.71	0.64	0.45	0.48
2QWJ	0.35	0.10	0.61	0.39	0.49	0.36
2QWK	0.33	0.10	0.55	0.24	0.43	0.49

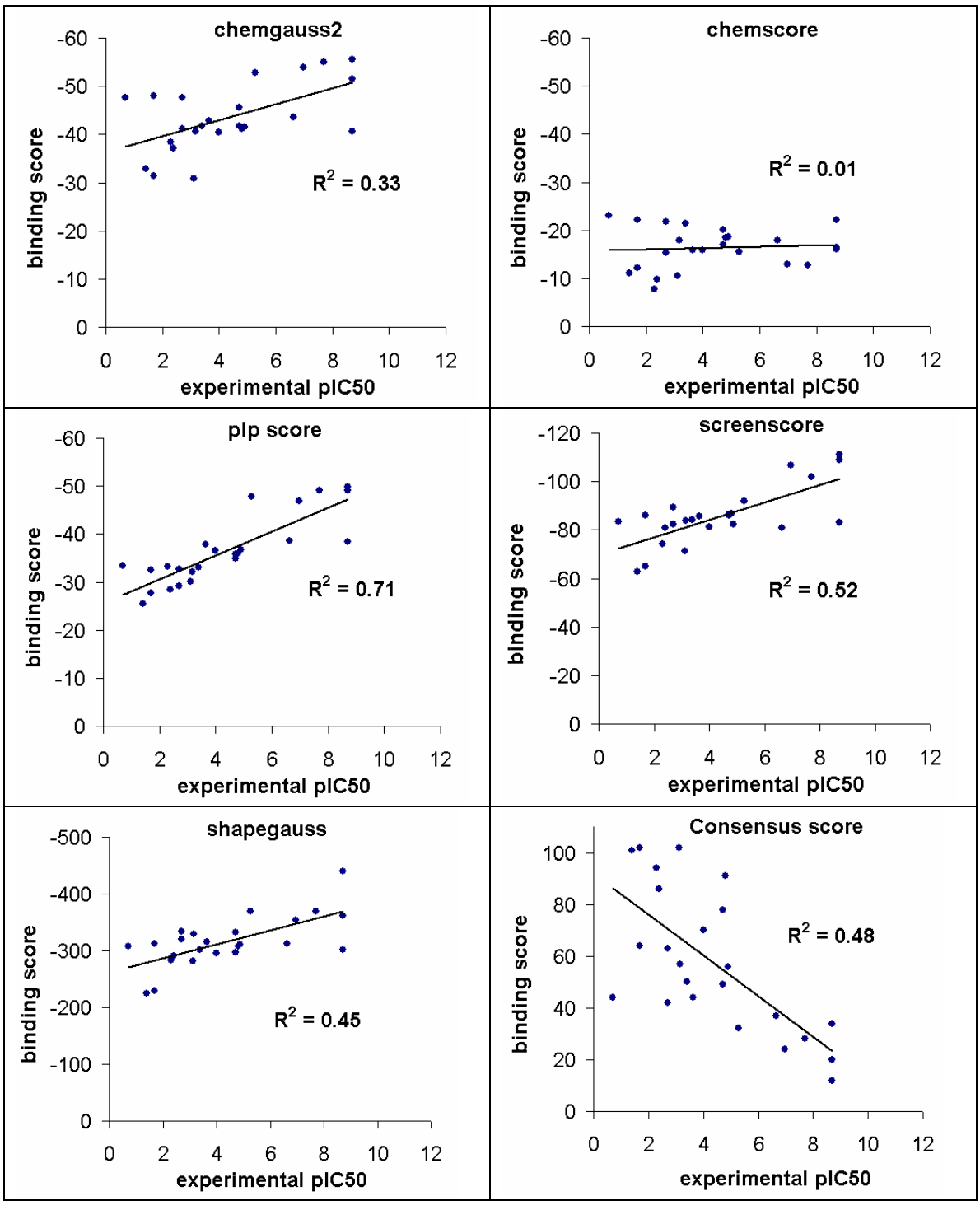


Figure 20 Plot of experimental pIC_{50} values versus binding score in FRED program from data set of protein structure PDB code 2QWI.

2. Part II: Virtual screening using AutoDock and FRED program

The ligand structures of unknown set from ChemieBase (30,000 compounds) and benzoic acid derivatives (47 compounds, represent with C1-C47 which is known structure but it can not disclose at present) docked into pocket of NA by using enzyme grid maps 2QWI (grid point 100 and grid spacing 0.275). Then, docking results from AuotDock 3.0 were scored by various scoring function in FRED program.

Docking results of unknown set from ChemieBase, we chose the hit compounds by the lowest binding score with the top rank of PLP scoring function, as shown in Table 10. Hit compounds were separated into 2 groups. The first group, we considered from binding score lower than -40 as a reference because it is binding score of highest inhibitory potencies (pIC50 = 8.70, oseltamivir (G39)) from 24 NIs. The second group, we considered from binding score lower than -25 as a reference because it is binding score of lowest experimental inhibitory potencies drugs (pIC50 = 0.70, α-Neu5Ac) from 24 NIs. The CAS and PLP binding score of hit compounds in the first group are shown in Table 11. These hit compounds did not test the neuraminidase inhibitory assay. The detail of hit compounds in the second group did not show in this thesis because it has many compounds.

Table 10 Number of hit compounds with PLP score separated into 2 groups.

Binding score	Number of hit compounds
≤ -40	258
≤ -25	11,297

Table 11 The CAS of hit compounds with PLP score lower than -40.

No.	CAS	Binding score	No.	CAS	Binding score
1	130486-19-0	-78.77	48	638-95-9	-44.78
2	125215-19-2	-63.95	49	70-18-8	-44.73
3	24219-97-4	-63.39	50	61-33-6	-44.69
4	5545-89-1	-62.90	51	1898-66-4	-44.62
5	13832-70-7	-55.23	52	4098-40-2	-44.57
6	188915-23-3	-54.63	53	28500-01-8	-44.42
7	33043-22-0	-54.11	54	57-97-6	-44.42
8	66756-57-8	-50.97	55	59-05-2	-44.20
9	69-23-8	-50.87	56	29219-79-2	-44.14
10	G21	-49.80	57	94285-27-5	-44.02
11	84-97-9	-49.45	58	29462-18-8	-43.98
12	70-18-8	-49.24	59	57-97-6	-43.87
13	G20	-49.18	60	175991-74-9	-43.81
14	GNA	-49.07	61	51-55-8	-43.79
15	23564-05-8	-48.64	62	75695-93-1	-43.78
16	13832-70-7	-48.36	63	32873-72-6	-43.76
17	13832-70-7	-48.31	64	589-68-4	-43.63
18	23564-05-8	-47.85	65	98260-60-7	-43.54
19	259098-17-4	-47.82	66	79-57-2	-43.50
20	G20	-47.77	67	14320-70-8	-43.43
21	522-17-8	-47.55	68	24258-38-6	-43.38
22	142435-70-9	-47.44	69	57574-09-1	-43.34
23	113000-69-4	-46.94	70	14255-77-7	-43.30
24	GNA	-46.87	71	67-43-6	-43.23
25	2009-24-7	-46.78	72	35367-38-5	-43.18
26	53994-73-3	-46.73	73	1862-41-5	-43.17
27	128-33-6	-46.72	74	70-18-8	-43.12
28	658106-2	-46.71	75	234430-58-1	-43.07
29	145840-77-3	-46.59	76	1707-75-1	-43.01
30	1898-66-4	-46.55	77	10172-03-9	-42.99
31	53994-73-3	-46.51	78	144407-50-1	-42.97
32	160564-55-6	-46.27	79	559-70-6	-42.97
33	33049-41-1	-46.14	80	55285-14-8	-42.86
34	482-36-0	-46.11	81	76824-35-6	-42.85
35	113467-91-7	-46.06	82	51707-55-2	-42.74
36	145746-34-5	-45.99	83	548776-51-8	-42.74
37	50370-12-2	-45.96	84	4707-32-8	-42.69
38	138175-83-4	-45.90	85	99663-29-3	-42.63
39	14291-41-9	-45.89	86	508-02-1	-42.60
40	56-65-5	-45.89	87	55-56-1	-42.59
41	18412-96-9	-45.67	88	83-88-5	-42.57
42	76824-35-6	-45.43	89	31251-03-3	-42.57
43	2009-24-7	-45.30	90	458-37-7	-42.56
44	188915-21-1	-45.25	91	989-51-5	-42.56
45	21416-53-5	-45.02	92	72967-11-4	-42.47
46	98534-99-7	-44.96	93	189181-20-2	-42.46
47	87276-92-4	-44.83	94	6197-30-4	-42.43

Table 11 (Continued)

No.	CAS	Binding score	No.	CAS	Binding score
95	56-65-5	-42.42	142	482-67-7	-41.38
96	148113-82-0	-42.38	143	61080-23-7	-41.38
97	6988-81-4	-42.36	144	121362-89-8	-41.36
98	98192-65-5	-42.35	145	51-55-8	-41.34
99	458-37-7	-42.34	146	2673-40-7	-41.32
100	2468-21-5	-42.32	147	13739-04-3	-41.29
101	36734-19-7	-42.31	148	29908-03-0	-41.27
102	215377-69-8	-42.28	149	64052-89-7	-41.24
103	54794-74-0	-42.27	150	72058-36-7	-41.24
104	83-88-5	-42.27	151	14320-68-4	-41.22
105	144707-50-6	-42.24	152	906-33-2	-41.19
106	73213-15-7	-42.24	153	58-73-1	-41.19
107	176744-93-7	-42.22	154	59669-26-0	-41.18
108	103055-07-8	-42.21	155	506-32-1	-41.17
109	64706-54-3	-42.20	156	259098-19-6	-41.16
110	85233-19-8	-42.14	157	474806-18-3	-41.16
111	5178-19-8	-42.12	158	41787-68-2	-41.15
112	5041-82-7	-42.09	159	604-80-8	-41.14
113	69-53-4	-42.06	160	145701-12-8	-41.14
114	537-12-2	-42.04	161	26543-89-5	-41.14
115	56-65-5	-42.04	162	63675-72-9	-41.14
116	176744-93-7	-41.98	163	125409-95-2	-41.11
117	2131-66-0	-41.96	164	480-10-4	-41.10
118	160564-56-7	-41.95	165	63675-72-9	-41.07
119	293320-58-8	-41.94	166	744209-94-7	-41.07
120	76824-35-6	-41.93	167	83-88-5	-41.07
121	672948-77-5	-41.91	168	59-43-8	-41.06
122	71697-59-1	-41.87	169	989-51-5	-41.03
123	1898-66-4	-41.87	170	76824-35-6	-41.03
124	101-31-5	-41.86	171	83306-18-7	-41.02
125	35028-02-5	-41.82	172	55038-30-7	-41.02
126	564-25-0	-41.76	173	495-86-3	-40.99
127	90101-16-9	-41.74	174	5085-54-1	-40.97
128	145746-02-7	-41.73	175	58962-34-8	-40.95
129	22254-24-6	-41.68	176	69-53-4	-40.93
130	482-35-9	-41.61	177	23564-05-8	-40.92
131	1898-66-4	-41.60	178	623577-74-2	-40.92
132	83-88-5	-41.57	179	1200-22-2	-40.90
133	53-60-1	-41.55	180	139-70-8	-40.89
134	128039-91-8	-41.53	181	84-21-9	-40.88
135	56-89-3	-41.51	182	24939-16-0	-40.88
136	113467-93-9	-41.50	183	101007-06-1	-40.87
137	17073-30-2	-41.49	184	14259-45-1	-40.86
138	161050-58-4	-41.45	185	144407-49-8	-40.84
139	132-66-1	-41.43	186	564-25-0	-40.84
140	51-55-8	-41.42	187	21829-25-4	-40.81
141	86-24-8	-41.40	188	19895-95-5	-40.80

Table 11 (Continued)

No.	CAS	Binding score	No.	CAS	Binding score
189	144341-75-3	-40.78	236	60-87-7	-40.21
190	118-10-5	-40.78	237	559-70-6	-40.20
191	117-68-0	-40.77	238	162049-30-1	-40.17
192	28274-10-4	-40.76	239	482-35-9	-40.17
193	147-24-0	-40.73	240	70-18-8	-40.16
194	989-51-5	-40.73	241	70-18-8	-40.15
195	84-97-9	-40.71	242	59-30-3	-40.13
196	3736-59-2	-40.71	243	133-03-9	-40.12
197	80-08-0	-40.69	244	125410-00-6	-40.10
198	144707-51-7	-40.67	245	482-35-9	-40.10
199	30918-54-8	-40.67	246	59182-85-3	-40.09
200	160564-54-5	-40.67	247	113-42-8	-40.05
201	104319-96-2	-40.67	248	531-44-2	-40.04
202	5302-74-9	-40.67	249	51-34-3	-40.03
203	59-30-3	-40.66	250	57-96-5	-40.03
204	76824-35-6	-40.66	251	109905-45-5	-40.03
205	58-64-0	-40.65	252	131-543-22-1	-40.02
206	160096-57-1	-40.62	253	118-10-5	-40.00
207	302-79-4	-40.62	254	69-53-4	-40.00
208	26543-89-5	-40.61	255	59-05-2	-39.99
209	81943-62-6	-40.59	256	22839-47-0	-39.96
210	59703-84-3	-40.58	257	27836-64-2	-39.93
211	79-57-2	-40.55	258	G39	-39.93
212	56070-94-1	-40.54			
213	60-92-4	-40.54			
214	74991-91-6	-40.53			
215	53-8	-40.53			
216	32222-06-3	-40.50			
217	149596-53-2	-40.50			
218	989-51-5	-40.48			
219	22611-36-5	-40.46			
220	152-72-7	-40.45			
221	104340-86-5	-40.42			
222	517-44-2	-40.41			
223	74-79-3	-40.40			
224	6873-15-0	-40.39			
225	127-47-9	-40.38			
226	217960-81-1	-40.31			
227	169392-11-4	-40.30			
228	81659-82-7	-40.30			
229	160564-61-4	-40.29			
230	50370-12-2	-40.28			
231	83-88-5	-40.28			
232	53994-73-3	-40.25			
233	164015-39-8	-40.24			
234	59756-60-4	-40.23			
235	24775-57-3	-40.22			

In case of benzoic acid derivatives 47 compounds, we chose 5 hit compounds from the lowest binding score with the top rank of PLP scoring function, as shown in Table 12. Five hit compounds were submitted for neuraminidase inhibition assay.

Table 12 PLP binding score of 5 compounds.

No	Compound	Binding score
1	C47	-38.73
2	C32	-34.98
3	C26	-32.76
4	C21	-31.76
5	C30	-31.30

From neuraminidase inhibition assay, the experimental data showed that these compounds have inhibitory activity NA with IC_{50} values from 2.19 mM to 4.13 mM, as shown in Tale 13. These results indicated that hit compounds from virtual screening with AutoDock program combine with PLP scoring function in FRED program can inhibit enzyme neuraminidase.

Table 13 The experimental results of neuraminidase inhibition assay of 5 hit compounds.

No	Compound	Experimental IC_{50} (mM)
1	C 47	2.34
2	C 32	2.19
3	C 26	2.42
4	C 21	4.08
5	C 30	4.13

After that these hit compound were submitted to test the anti-viral activity assay or drug sensitivity assay. In the experimental of drug sensitivity assay, these compound tested inhibitory activity for viral in 3 subtype of influenza viral, H5N1,

H3N2 and H1N1. For each subtype of influenza viral tested by using viruses in 3 levels, 100, 25 and 5 TCID50/0.1 ml. The results of experimental showed that compound C47 have inhibitory activity against 3 subtypes of influenza viral, as shown in Table 14 and Figure 21 - 23.

Table 14 The experimental results of drug sensitivity assay of 5 hit compounds.

Drug	Drug concentration (μM/ml)	% inhibition, variation of viruses titer treated with drug								
		A/Thailand/1 (KAN-1)/2004 H5N1			A/New Caledonia/20/99 H1N1			A/Fujian/411/01 H3N2		
		100	25	5	100	25	5	100	25	5
	100	ND	ND	ND	ND	ND	ND	ND	ND	ND
	50	97	93.6	93.6	91.5	97.2	90.8	94.4	94.6	84.5
	25	93.6	58.3	79.5	90.8	91.2	90.1	95.3	94.4	88.7
C 47	12.5	<10	<10	<10	32.8	40.5	90	12	40.4	73
	100	ND	<10	<10	ND	ND	ND	ND	ND	ND
	50	ND	<10	<10	ND	<10	<10	ND	<10	<10
	25	ND	<10	<10	ND	<10	<10	ND	<10	<10
C 26	12.5	ND	<10	<10	ND	<10	<10	ND	<10	<10
	100	ND	<10	<10	ND	<10	80	ND	ND	ND
	50	ND	<10	<10	ND	49.8	70.8	ND	<10	<10
	25	ND	<10	<10	ND	47.9	59.2	ND	<10	<10
C 30	12.5	ND	<10	<10	ND	3.97	<10	ND	<10	<10
	100	ND	<10	<10	ND	<10	10.2	ND	14.9	22.6
	50	ND	<10	<10	ND	<10	11.7	ND	17.8	<10
	25	ND	<10	<10	ND	<10	<10	ND	15.4	37.9
C 21	12.5	ND	<10	<10	ND	<10	20.7	ND	<10	32.1
	100	ND	ND	ND	ND	ND	ND	ND	ND	ND
	50	ND	<10	<10	ND	<10	34.2	ND	3	43.8
	25	ND	<10	<10	ND	<10	33	ND	<10	32
C 32	12.5	ND	<10	<10	ND	<10	<10	ND	<10	21.3

Note ND is not detection.

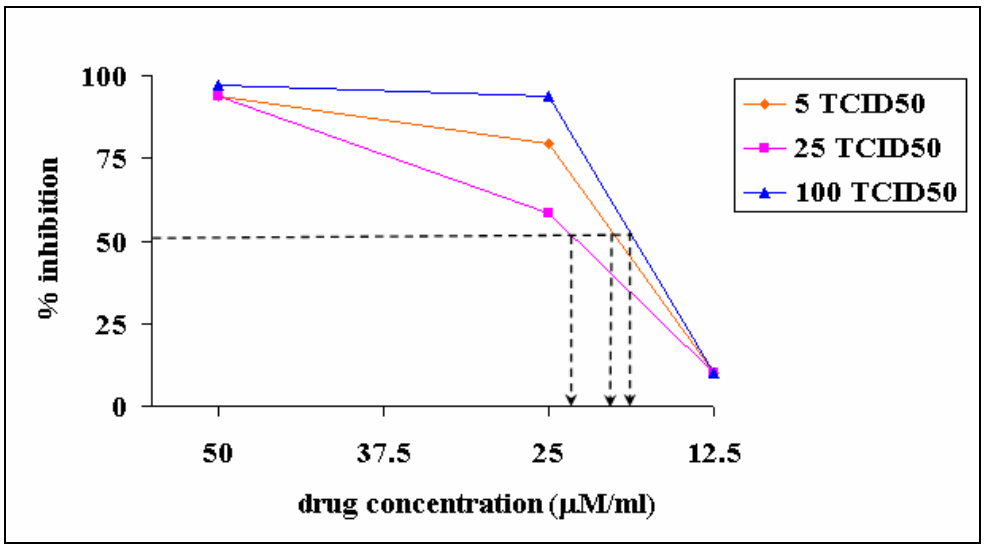


Figure 21 Plot of % inhibition versus drug concentration of C47 in subtype H5N1 of influenza virus.

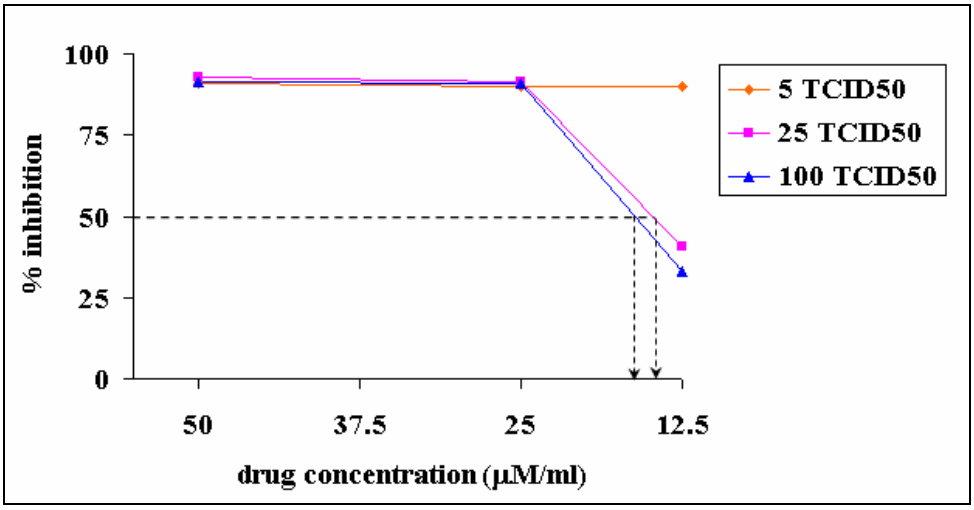


Figure 22 Plot of % inhibition versus drug concentration of C47 in subtype H1N1 of influenza virus.

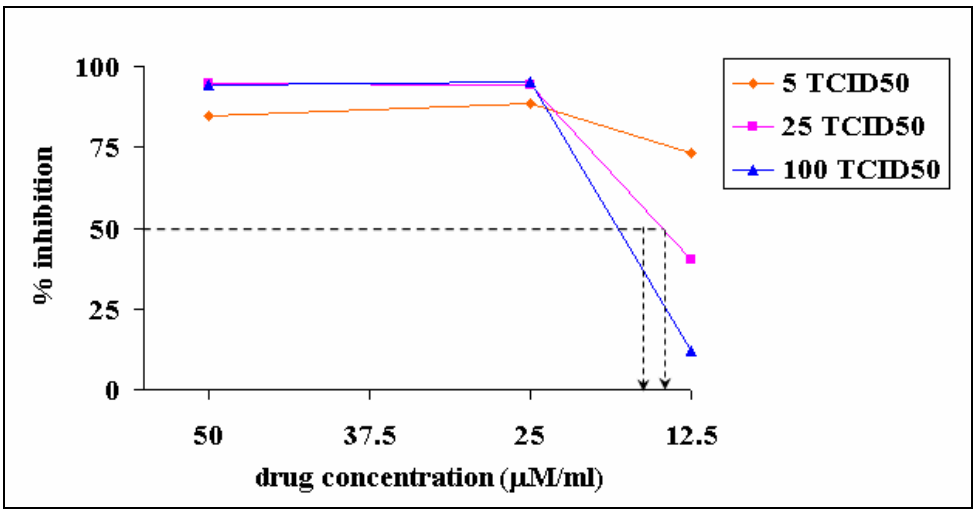


Figure 23 Plot of % inhibition versus drug concentration of C47 in subtype H3N2 of influenza virus.

CONCLUSION

We studied the feasibility of using AutoDock as a tool for virtual screening by doing cross-validation of 24 co-crystals of NA and inhibitors in three methods. For the method I, AutoDock performed fairly well in term of co-crystal reproduction (rmsd from 0.78 to 1.84) with free all rotatable bonds. The correlation coefficient (r^2) of predicted binding free energy and pIC_{50} of the experimental data was 0.54 and 0.34 base on docked energy. In cross validation, most data set showed poor correlation. For the method II, we chose by random some protein PDB code (1BJI, 2QWC, 2QWI) to do reproduction of co-crystals and cross validation. The correlation did not increase when compare with results in method I. Method III, AutoDock performed the co-crystal reproduction better than method I. The best rmsd was between 0.48 Å and 1.35 Å. The correlation coefficients between pIC_{50} free energy of prediction and pIC_{50} experimental data was 0.72 and 0.71 bases on free energy of binding and docked energy respectively. These results indicated that AutoDock was competent in reproducing the experimentally found binding position and conformation of NIs. Moreover, in the method III, the results of reproduction 24 co-crystals were tested by scoring with scoring function's FRED program (ChemGauss2, ChemScore, PLP, ScreenScore, ShapeGauss and Consensus). We found that three scoring functions, *i.e.,* PLP, Screenscore and Consensus showed a fair correlation between prediction values and experimental data (r^2 = 0.61 -0.73) which showed that we can use various scoring function in FRED program to select the best docked conformation from the many AutoDock results for NA system. Results of cross validation, data set from PDB code 2QWI gave the highest correlation 0.71 by PLP scoring function. Therefore, the protein complex PDB code 2QWI was the best pocket for this system.

When we screened compound on ChemieBase (around 30,000 compounds) by using enzyme grid maps of PBD code 2QWI. We presented list of hit compounds which did not test the neuraminidase inhibition assay. For benzoic acid derivatives 47 compound, we found that 5 compounds were submitted for neuraminidase inhibition

assay. These compounds showed inhibitory activity NA with IC_{50} values from 2.19 mM to 4.13 mM, and compound C47 showed inhibitory activity against 3 subtypes of influenza viral (H5N1, H3N2 and H1N1).

LITERATURE CITED

Abagyan, R. and M. Totrov. 2001. High-throughput docking for lead generation. **Curr. Opin. Chem. Biol.** 5(4):375-382.

Ahn, K.-H. and H.J. Halpern. 2007. Comparison of local and global angular interpolation applied to spectral-spatial EPR image reconstruction. **Med. Phys. Medical Physics** 34(3):1047-1052.

Anderson, A.C. 2003. The process of structure-based drug design. **Chem. Biol.** 10: 787-797.

Atigadda, V.R., W.J. Brouillette, F. Duarte, S.M. Ali, Y.S. Babu, S. Bantia, P. Chand, N. Chu, J.A. Montgomery, D.A. Walsh, E.A. Sudbeck, J. Finley, M. Luo, G.M. Air and G.W. Laver. 1999. Potent Inhibition of Influenza Sialidase by a Benzoic Acid Containing a 2-Pyrrolidinone Substituent. **J. Med. Chem.** 42(13):2332-2343.

Babu, Y.S., P. Chand, S. Bantia, P. Kotian, A. Dehghani, Y. El-Kattan, T.-H. Lin, T.L. Hutchison, A.J. Elliott, C.D. Parker, S.L. Ananth, L.L. Horn, G.W. Laver and J.A. Montgomery. 2000. Bcx-1812 (rwj-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. **J. Med. Chem.** 43(19):3482-3486.

Bajorath, J. 2002. Integration of virtual and high-throughput screening. **Nat. Rev. Drug Discov.** 1: 882-894.

Bamford, M.J. 1995. Neuraminidase inhibitors as potential anti-influenza drugs. **J. Enzym Inhib.** 10(1):1-16.

Birch, L., C.W. Murray, M.J. Hartshorn, I.J. Tickle and M.L. Verdonk. 2003. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. **J. Comput. Aided Mol. Des.** 16(12):855-869.

Bodian, D.L., R.B. Yamasaki, R.L. Buswell, J.F. Stearns, J.M. White and I.D. Kuntz. 1993. Inhibition of the fusion-inducing conformational change of influenza hemagglutinin by benzoquinones and hydroquinones. **Biochemistry** 32(12):2967-2978.

Boehm, H.-J. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. **J. Comput. Aided Mol. Des.** 8(3):243-256.

Brenk, R., L. Naerum, U. Graedler, H.-D. Gerber, G.A. Garcia, K. Reuter, M.T. Stubbs and G. Klebe. 2003. Virtual Screening for Submicromolar Leads of tRNA-guanine Transglycosylase Based on a New Unexpected Binding Mode Detected by Crystal Structure Analysis. **J. Med. Chem.** 46(7):1133-1143.

Brooijmans, N. and I.D. Kuntz. 2003. Molecular recognition and docking algorithms. **Annu. Rev. Biophys. Biomol. Struct.** 32335-373.

Bursulaya, B.D., M. Totrov, R. Abagyan and C.L. Brooks, III. 2004. Comparative study of several algorithms for flexible ligand docking. **J. Comput. Aided Mol. Des.** 17(11):755-763.

Buzko, O.V., A.C. Bishop and K.M. Shokat. 2002. Modified AutoDock for accurate docking of protein kinase inhibitors. **J. Comput. Aided Mol. Des.** 16(2):113-127.

Cao, T. and T. Li. 2004. A combination of numeric genetic algorithm and tabu search can be applied to molecular docking. **J. Comput. Biol.** 28(4):303-312.

Carlson, H.A. and J.A. McCammon. 2000. Accommodating protein flexibility in computational drug design. **Mol. Pharmacol.** 57(2):213-218.

Chand, P., Y.S. Babu, S. Bantia, N. Chu, L.B. Cole, P.L. Kotian, W.G. Laver, J.A. Montgomery, V.P. Pathak, S.L. Petty, D.P. Shrout, D.A. Walsh and G.M. Walsh. 1997. Design and Synthesis of Benzoic Acid Derivatives as Influenza Neuraminidase Inhibitors Using Structure-Based Drug Design. **J. Med. Chem.** 40(25):4030-4052.

Charifson, P.S., J.J. Corkery, M.A. Murcko and W.P. Walters. 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. **J. Med. Chem.** 42(25):5100-5109.

Chen, H.M., B.F. Liu, H.L. Huang, S.-F. Hwang and S.-Y. Ho. 2007. SODOCK: swarm optimization for highly flexible protein-ligand docking. **J. Comput. Chem.** 28(2):612-623.

Chen, H., P.D. Lyne, F. Giordanetto, T. Lovell and J. Li. 2006. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. **J. Chem. Inf. Model.** 46(1):401-415.

Chen, Y.Z., X.L. Gu and Z.W. Cao. 2001. Can an optimization/scoring procedure in ligand-protein docking be employed to probe drug-resistant mutations in proteins. **J. Mol. Graphics Modell.** 19(6):560-570.

Claussen, H., C. Buning, M. Rarey and T. Lengauer. 2001. FlexE: Efficient molecular docking considering protein structure variations. **J. Mol. Biol.** 308(2):377-395.

Diller, D.J. and R. Li. 2003. Kinases, Homology Models, and High Throughput Docking. **J. Med. Chem.** 46(22):4638-4647.

Doman, T.N., S.L. McGovern, B.J. Witherbee, T.P. Kasten, R. Kurumbail, W.C. Stallings, D.T. Connolly and B.K. Shoichet. 2002. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. **J. Med. Chem.** 45(11):2213-2221.

Eldridge, M.D., C.W. Murray, T.R. Auton, G.V. Paolini and R.P. Mee. 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **J. Comput. Aided Mol. Des.** 11(5):425-445.

Enyedy, I.J., Y. Ling, K. Nacro, Y. Tomita, X. Wu, Y. Cao, R. Guo, B. Li, X. Zhu, Y. Huang, Y.-Q. Long, P.P. Roller, D. Yang and S. Wang. 2001. Discovery of Small-Molecule Inhibitors of Bcl-2 through Structure-Based Computer Screening. **J. Med. Chem.** 44(25):4313-4324.

Ewing, T.J.A., S. Makino, A.G. Skillman and I.D. Kuntz. 2001. DOCK 4..0: search strategies for automated molecular docking of flexible molecule databases. **J. Comput. Aided Mol. Des.** 15(5):411-428.

Gohlke, H. and G. Klebe. 2002. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. **Angew. Chem. Int. Ed.** 41(15):2644-2676.

Goodsell, D.S., G.M. Morris and A.J. Olson. 1996. Automated docking of flexible ligands: applications of AutoDock. **J. Mol. Recognit.** 9(1):1-5.

Greer, J., J.W. Erickson, J.J. Baldwin and M.D. Varney. 1994. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. **J. Med. Chem.** 37(8):1035-1054.

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

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

Halperin, I., B. Ma, H. Wolfson and R. Nussinov. 2002. Principles of docking: an overview of search algorithms and a guide to scoring functions. **Proteins.** 47(4):409-443.

Hatakeyama, S. 2005. Neuraminidase inhibitors as potential antiviral agents for pandemic influenza viruses. **Bio. Clinica.** 20(14):1252-1256.

Hoffmann, D., B. Kramer, T. Washio, T. Steinmetzer, M. Rarey and T. Lengauer. 1999. Two-Stage Method for Protein-Ligand Docking. **J. Med. Chem.** 42(21):4422-4433.

Huang, R.T., R. Rott and H.D. Klenk. 1981. Influenza viruses cause hemolysis and fusion of cells. **Virology** 110(1):243-247.

Jain, A.N. 2003. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. **J. Med. Chem.** 46(4):499-511.

Kamionka, M., T. Rehm, H.-G. Beisel, K. Lang, R.A. Engh and T.A. Holak. 2002. In Silico and NMR Identification of Inhibitors of the IGF-I and IGF-Binding Protein-5 Interaction. **J. Med. Chem.** 45(26):5655-5660.

Kim, C.U., W. Lew, M.A. Williams, H. Wu, L. Zhang, X. Chen, P.A. Escarpe, D.B. Mendel, W.G. Laver and R.C. Stevens. 1998. Structure-Activity Relationship

Studies of Novel Carbocyclic Influenza Neuraminidase Inhibitors. **J. Med. Chem.** 41(14):2451-2460.

Klebe, G. 2000. Recent developments in structure-based drug design. **J. Mol. Med.** 78: 269-281.

Knegtel, R.M.A., I.D. Kuntz and C.M. Oshiro. 1997. Molecular docking to ensembles of protein structures. **J. Mol. Biol.** 266(2):424-440.

Kontoyianni, M., L.M. McClellan and G.S. Sokol. 2004. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. **J. Med. Chem.** 47(3):558-565.

Kramer, B., M. Rarey and T. Lengauer. 1999. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. **Proteins** 37(2):228-241.

Krovat, E. M., T. Steindl and T. Langer. 2005. Recent advances in docking and scoring. **Curr. Comput. Aided Drug Des.** 1: 93-102.

Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge and T.E. Ferrin. 1982. A geometric approach to macromolecule-ligand interactions. **J. Mol. Biol.** 161(2):269-288.

Lew, W., X. Chen and C.U. Kim. 2000. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. **Curr. Med. Chem.** 7(6):663-672.

Lyne, P.D. 2002. Structure-based virtual screening: an overview. **DDT** 7: 1047-1055.

Schneider, G. and H.J. Böhm. 2002. Virtual screening and fast automated docking methods. **DDT** 7: 64-70.

MacDonald, S.J.F., R. Cameron, D.A. Demaine, R.J. Fenton, G. Foster, D. Gower, J.N. Hamblin, S. Hamilton, G.J. Hart, A.P. Hill, G.G.A. Inglis, B. Jin, H.T. Jones, D.B. McConnell, J. McKimm-Breschkin, G. Mills, V. Nguyen, I.J. Owens, N. Parry, S.E. Shanahan, D. Smith, K.G. Watson, W.-Y. Wu and S.P. Tucker. 2005. Dimeric Zanamivir Conjugates with Various Linking Groups Are Potent, Long-Lasting Inhibitors of Influenza Neuraminidase Including H5N1 Avian Influenza. **J. Med. Chem.** 48(8):2964-2971.

Marrone, T.J., J.M. Briggs and J.A. McCammon. 1997. Structure-based drug design: Computational advances. **Annu. Rev. Pharmacol. Toxicol.** 37: 71-90.

Masukawa, K.M., P.A. Kollman and I.D. Kuntz. 2003. Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. **J. Med. Chem.** 46(26):5628-5637.

McConkey, B.J., V. Sobolev and M. Edelman. 2002. The performance of current methods in ligand–protein docking. **Curr. Sci.** 83: 845-856.

McGann, M.R., H.R. Almond, A. Nicholls, J.A. Grant and F.K. Brown. 2003. Gaussian docking functions. **Biopolymers** 68(1):76-90.

McNally, V.A., A. Gbaj, K.T. Douglas, I.J. Stratford, M. Jaffar, S. Freeman and R.A. Bryce. 2003. Identification of a novel class of inhibitor of human and Escherichia coli thymidine phosphorylase by in silico screening. **Bioorg. Med. Chem. Lett.** 13(21):3705-3709.

Medina-Franco, J.L., A. Golbraikh, S. Oloff, R. Castillo and A. Tropsha. 2005. Quantitative Structure-activity Relationship Analysis of Pyridinone HIV-1 Reverse Transcriptase Inhibitors using the k Nearest Neighbor Method and QSAR-based Database Mining. **J. Comput. Aided Mol. Des.** 19(4):229-242.

Miteva, M.A., W.H. Lee, M.O. Montes and B.O. Villoutreix. 2005. Fast Structure-Based Virtual Ligand Screening Combining FRED, DOCK, and Surflex. **J. Med. Chem.** 48(19):6012-6022.

Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew and A.J. Olson. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **J. Comput. Chem.** 19(14):1639-1662.

Morris, G.M., D.S. Goodsell, R. Huey and A.J. Olson. 1996. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. **J. Comput. Aided Mol. Des.** 10(4):293-304.

Moscona, A. 2005. Oseltamivir resistance - Disabling our influenza defenses. **N. Engl. J. Med.** 353(25):2633-2636.

Murray, C.W., C.A. Baxter and A.D. Frenkel. 1999. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. **J. Comput. Aided Mol. Des.** 13(6):547-562.

Najmanovich, R., J. Kuttner, V. Sobolev and M. Edelman. 2000. Side-chain flexibility in proteins upon ligand binding. **Proteins** 39(3):261-268.

Oda, A., K. Tsuchida, T. Takakura, N. Yamaotsu and S. Hirono. 2006. Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. **J. Chem. Inf. Model.** 46(1):380-391.

Oprea, T.I. 2002. Virtual screening in lead discovery: A viewpoint**. Molecules** 7: 51-62.

Peng, H., N. Huang, J. Qi, P. Xie, C. Xu, J. Wang and C. Yang. 2003. Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening. **Bioorg. Med. Chem. Lett.** 13(21):3693-3699.

Perez, C. and A.R. Ortiz. 2001. Evaluation of docking functions for protein-ligand docking. **J. Med. Chem.** 44(23):3768-3785.

Rarey, M., B. Kramer, T. Lengauer and G. Klebe. 1996. A fast flexible docking method using an incremental construction algorithm. **J. Mol. Biol.** 261(3):470-489.

Sandak, B., H.J. Wolfson and R. Nussinov. 1998. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. **Proteins** 32(2):159-174.

Sears, P. and C.H. Wong. 1999. Carbohydrate mimetics: a new strategy for tackling the problem of carbohydrate-mediated biological recognition. **Angew. Chem. Int. Ed.** 38(16):2301-2324.

Seifert, M.H.J., K. Wolf and D. Vitt. 2003. Virtual high-throughput in silico screening. **Biosilico** 1: 143-149.

Sherman, W., T. Day, M.P. Jacobson, R.A. Friesner and R. Farid. 2006. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. **J. Med. Chem.** 49(2):534-553.

Shoichet, B.K. 2004. Virtual screening of chemical libraries. **Nature** 432: 862-865.

Sotriffer, C.A., H. Gohlke and G. Klebe. 2002. Docking into Knowledge-Based Potential Fields: A Comparative Evaluation of DrugScore. **J. Med. Chem.** 45(10):1967-1970.

Stahl, M. and M. Rarey. 2001. Detailed Analysis of Scoring Functions for Virtual Screening. **J. Med. Chem.** 44(7):1035-1042.

Taylor, R.D., P.J. Jewsbury and J.W. Essex. 2002. A review of protein-small molecule docking methods. **J. Comput. Aided Mol. Des.** 16(3):151-166.

Teramoto, R. and H. Fukunishi. 2007. Supervised Consensus Scoring for Docking and Virtual Screening. **J. Chem. Inf. Model.** 47(2):526-534.

Vangrevelinghe, E., K. Zimmermann, J. Schoepfer, R. Portmann, D. Fabbro and P. Furet. 2003. Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking. **J. Med. Chem.** 46(13):2656-2662.

Venkatachalam, C.M., X. Jiang, T. Oldfield and M. Waldman. 2003. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. **J. Mol. Graphics Modell.** 21(4):289-307.

Verdonk, M.L., V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor and P. Watson. 2004. Virtual screening using protein-ligand docking: Avoiding artificial enrichment. **J. Chem. Inf. Comput. Sci.** 44(3):793-806.

Verdonk, M.L., J.C. Cole, M.J. Hartshorn, C.W. Murray and R.D. Taylor. 2003. Improved protein-ligand docking using GOLD. **Proteins** 52(4):609-623.

Verkhivker, G.M., D. Bouzida, D.K. Gehlhaar, P.A. Rejto, S. Arthurs, A.B. Colson, S.T. Freer, V. Larson, B.A. Luty, T. Marrone and P.W. Rose. 2000. Deciphering common failures in molecular docking of ligand-protein complexes. **J. Comput. Aided Mol. Des.** 14(8):731-751.

Wang, R., Y. Lu, X. Fang and S. Wang. 2004. An Extensive Test of 14 Scoring Functions Using the PDBbind Refined Set of 800 Protein-Ligand Complexes. **J. Chem. Inf. Comput. Sci.** 44(6):2114-2125.

Wang, R., Y. Lu and S. Wang. 2003. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. **J. Med. Chem.** 46(12):2287-2303.

Wang, T. and R.C. Wade. 2001. Comparative Binding Energy (COMBINE) Analysis of Influenza Neuraminidase-Inhibitor Complexes. **J. Med. Chem.** 44(6):961-971.

Warren, G.L., C.W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff and M.S. Head. 2006. A Critical Assessment of Docking Programs and Scoring Functions. **J. Med. Chem.** 49(20):5912-5931.

Zavodszky, M.I., P.C. Sanschagrin, L.A. Kuhn, R.S. Korde and L.A. Kuhn. 2003. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. **J. Comput. Aided Mol. Des.** 16(12):883-902.

Zheng, M., K. Yu, H. Liu, X. Luo, K. Chen, W. Zhu and H. Jiang. 2006. QSAR analyses on avian influenza virus neuraminidase inhibitors using CoMFA, CoMSIA, and HQSAR. **J. Comput. Aided Mol. Des.** 20(9):549-566.

APPENDICES

APPENDIX A: Theoretical Background

Molecular docking

An important method widely used in drug discovery is molecular docking (Halperin *et al.*, 2002). Molecular docking is a computational method attempting to predict the three-dimensional structures of receptor-ligand complexes and to evaluate the relative affinity of these bound ligands. Consequently, the aims of this method are to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity of their correct poses (Krovat *et al.*, 2005). In molecular docking, two proposes are needed in its procedure. They are docking algorithm used to generate a large number of possible structures and scoring function used to identify which are the most interest (Leach, 1996; Böhm, 2002; Brooijmans and Kuntz, 2003).

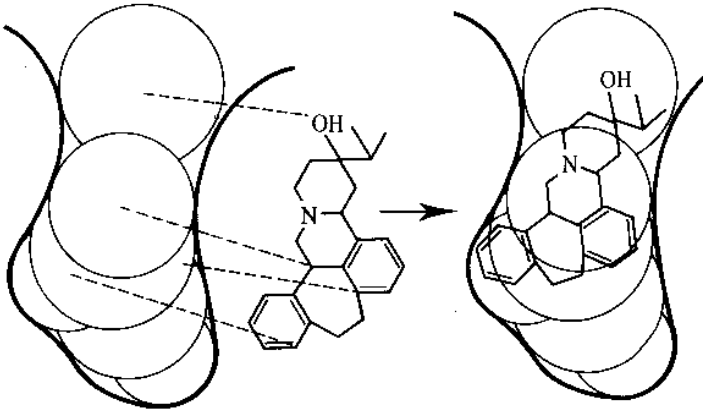
1. Docking algorithm

The first step in molecular docking method is to search the configurational and conformational degrees of freedom by using docking algorithm. The docking algorithm can be categorized by using degree of freedom. In rigid docking, the search algorithm finds the various position of ligand in the active site of receptor using translational and rotational degrees of freedom. To perform conformationally flexible docking, torsional degree of freedom of ligand is added to explore it in the active site of receptor.

1.1 Rigid docking

Rigid docking fits a rigid ligand into a rigid active site. The usual approach of this algorithm has been to generate a discrete model of the binding site consisting a set of points that define its spatial extensions with special emphasis on surface properties. Atoms of ligands are then matched onto these receptor points. The first docking program using this technique is DOCK program (Kuntz *et al.*, 1982). DOCK

program represents a binding site as overlapping spheres. Ligands are then matched to sphere centers by using clique technique, as shown in Appendix Figure A1.



Appendix Figure A1 DOCK algorithm. Atoms are matched to a binding site represented as a collection of overlapping spheres.

Source: Leach (1996)

1.2 Conformationally flexible docking

Docking by using rigid docking is fast but highly inaccurate for ligands with rotatable bonds, therefore the flexibility of ligands have been considered in docking algorithm, as called conformationally flexible docking. It can be divided in three types, i. e. systematic, stochastic and deterministic searches.

1.2.1 Systematic search

This algorithm is based on a grid of values for each formal degree of freedom. Each of these grid values is explored in a combinatorial fashion during a search. An example of a systematic search is the anchor-and-grow or incremental construction algorithm. In this algorithm, the ligand conformers are not an external pre-processing step. Their conformers are generated within the boundaries of the binding site. Ligands are generally divided into rigid (anchor) and flexible parts. Conformers are generated by growing the ligand from an anchor fragment. The procedure is described as follows. Firstly, one or more anchors with flexible parts are

defined by perception of rotatable bonds. Secondly, the anchor fragment is docked into the active site. Finally, the flexible parts are added sequentially with systematic scanning of torsional angles. The anchor-and-grow or incremental construction algorithm is implemented in several docking programs such as DOCK, FlexX, and Hammerhead.

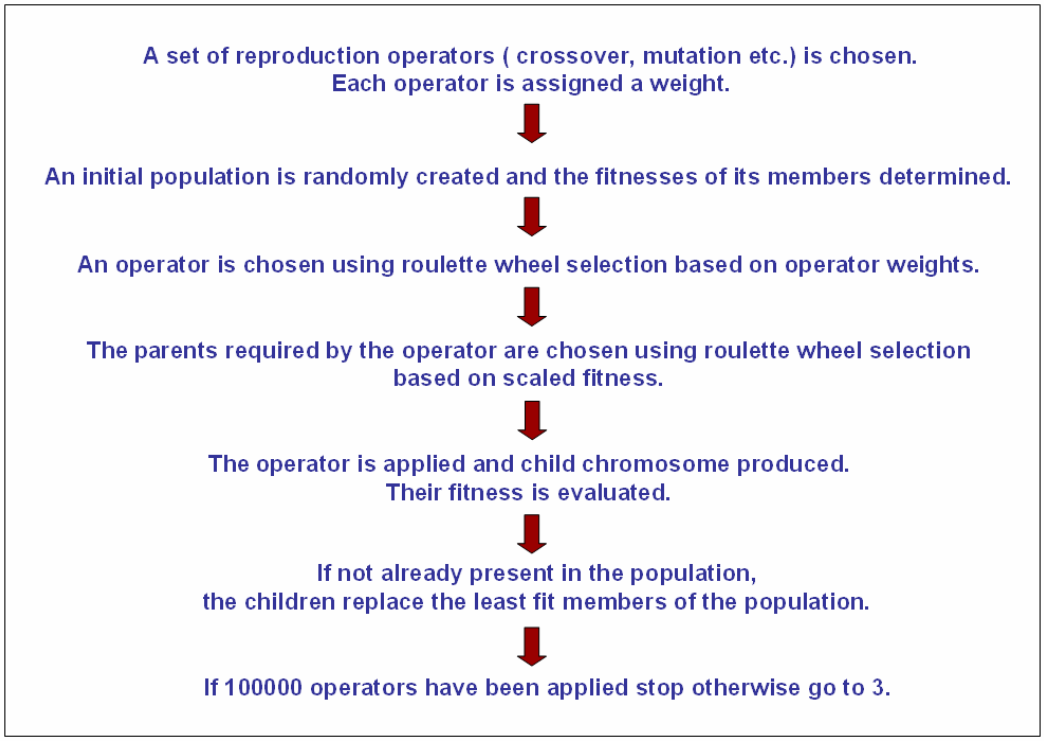
1.2.2 Stochastic search

Because of the combinatorial fashion in incremental construction algorithm, it is appealing idea to search the whole orientational and conformational space in one process, e.g. stochastic search. Stochastic search algorithms make random changes, usually changing one degree of freedom of the system at a time. Advantage of the stochastic search algorithms is the less complicated data structure requirement. Examples of stochastic searches are Monte Carlo (MC) methods and evolutionary algorithms.

In Monte Carlo (MC) algorithms, a whole ligand is usually represented by a string of real value variables describing translation, rotation and variable torsional angles. Random changes in these variables form the basis of these algorithms. After each move, the structure is minimized, and the energy of the new structure is determined. Several docking programs using the Monte Carlo (MC) algorithms are AutoDock, ICM, MCDOCK, ProDock, and PRO_LEAD.

Evolutionary algorithms are stochastic methods using to find the energy minimum. Evolutionary algorithms involve generation of an initial population of ligand conformations. They did not include crossover process, a process that swaps large regions of the “parents” during evolution. When these algorithms include crossover process, it is known as genetic algorithms (GA). The common feature of evolutionary and genetic algorithms is a cyclic variation-selection process. “Parents” breed “offspring”, and the best of each generation according to a fitness measure becomes the parent of the next optimization cycle. The general procedure of GA is

shown in Appendix Figure A2. The examples of programs using these algorithms are GOLD, GA-DOCK, and AutoDock.



Appendix Figure A2 Genetic algorithm (GA) procedure.

1.2.3 Deterministic search

In deterministic searches, the initial state determines the move that can be made to generate the next state, which generally has to be equal to or lower in energy than the initial state. Examples of deterministic searches are energy minimization methods and molecular dynamics (MD) simulations.

2. Scoring function

The evaluation and ranking of predicted ligand conformations is a crucial aspect of structure-based virtual screening. Even when binding conformations are correctly predicted, the calculations ultimately do not succeed if they do not differentiate correct poses from incorrect ones, and if 'true' ligands cannot be identified. So, the design of reliable scoring functions and schemes is of fundamental

importance. Free-energy simulation techniques have been developed for quantitative modeling of protein–ligand interactions and the prediction of binding affinity. However, these expensive calculations remain impractical for the evaluation of large numbers of protein–ligand complexes and are always not accurate. Scoring functions implemented in docking programs make various assumptions and simplifications in the evaluation of modelled complexes and do not fully account for a number of physical phenomena that determine molecular recognition for example, entropic effects. Essentially, three types or classes of scoring functions are currently applied: force field-based, empirical and knowledge-based scoring functions.

2.1 Force field-based methods

The force field-based scoring functions use the molecular mechanics force fields for the estimation of binding affinity. The interaction energies (van der Waals and electrostatic) between the receptor and ligand are computed. The overestimation of complex stability can be partially ascribed to the neglect of solute entropic terms. This scoring function can also added the empirical terms to take into account the entropy and solution changes. The AMBER and CHARMM force field are widely used as a scoring function in several docking program. Examples of the programs implemented the force field-based scoring functions are DOCK and AutoDock programs.

2.2 Empirical scoring functions

The empirical scoring functions are the most widely utilized in current drug design/discovery software. The empirical scoring functions are based on the assumption that the binding free energy ($\Delta G_{binding}$) is a sum of interactions multiplied by weighting coefficients (ΔG_i), as shown in equation (54). The energy contributions in scoring function usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and binding entropy.

$$\Delta G_{binding} \approx \sum \Delta G_i f_i(r_l, r_{l'}) \qquad (1)$$

Where each f_i is a simple geometrical function of the ligand coordinate r_l and the receptor coordinate $r_{l'}$. Most empirical scoring functions are derived by evaluating the functions f_i on a set of protein-ligand complexes and fitting the coefficients (ΔG_i) to experiment binding affinities of these complexes. The relative weight of each energy contribution depends on the training set. General problems of empirical scoring functions are that they are difficult to know what each term exactly accounts for and to access where errors come from. Binding free energy estimations can only be successful if the molecules make similar interactions to the ones in the training set complexes. Examples of the programs implemented the empirical scoring functions are FlexX, GOLD and LUDI.

2.3 Knowledge-based methods

The knowledge-based scoring functions use the structural data of protein-ligand complexes to observe frequencies of atom-atom interactions. Force and potentials are collected to get a score for their binding affinities, as called potential of mean force (PMF). The major different with the empirical scoring function is that no binding data are needed. This is an advantage of knowledge-based scoring functions that they can be derived from large training set. Such knowledge-based approaches have their foundation in the inverse formulation of the Boltzmann law, as shown in equation (2).

$$E_{ijk} = -kT\ln(p_{ijk}) + kT\ln(Z) \qquad (2)$$

Where the energy function (E_{ijk}) is called PMF for a state defined by the variables i, j, and k, p_{ijk} is the corresponding probability density, and Z is the partition function. The second term of the sum is constant at constant temperature T and does not have to be regarded, since $Z = 1$ can be chosen by definition of a suitable reference state leading to normalized probability densities, p_{ijk}. Variables i, j, and k can be chosen to be protein atom-types, ligand atom-types, and their inter-atom distance. The reference state is defined as the state where protein and ligand do not interact each other. PMF and DrugScore are the examples of knowledge-based scoring functions.

Up to now, there is no single scoring function that can correctly rank every protein-ligand complex because the relative contribution of different protein-ligand interactions may vary between structural families (Krovat *et al.*, 2005). The scoring functions still need the improvements to enhance the reliability of discrimination correctly docked from misdocked conformation or consensus scoring that is a combination of two or three scoring function has been proposed (Kontoyianni *et al.*, 2005).

3. Parameterization Method 3

Parameterization Method 3 (PM3) is a semi-empirical model based on the basic NDDO (Neglect of Diatomic Differential Overlap) theory by Michael Dewar at the University of Texas, Austin. Semi-empirical models based on the Hartree–Fock set of ideas. As this is not the forum for developing the ideas of Hartree-Fock theory, the derivation of the Roothaan-Hall equations.

PM3 works very well, for instance, when nitro-derivatives, extensively parameterized in the MBSP, are calculated. PM3 performs sometimes better for geometries that are guessed with high precision.

APPENDIX B: Neuraminidase Activity Assay

1. Neuraminidase Assay

1. Set up neuraminidase activity assay

2. Control: Neuraminidase (NA) from Clostridium perfringens (1 munit/μl). Various concentration (1, 2, 4, 8 μl) of NA were used to determine the optimal concentration for the assay.

3. In microtiter plate, add 50 μl reaction buffer + 50 μl NA (variuos concentration) or 50 μl water (negative control)

4. Prepare Amplex red/horseradish peroxidase/galactose oxidase/fetuin (50 µl Amplex red in DMSO, 10 µl HRP, 100 µl galactose oxidase, 250 µl fetuin, 4.59 µl buffer) solution

5. Add 50 µl of Amplex red solution (4) in each well

6. Incubate 30-60 min at 37 ^{0}C

7. Measure the fluorescence (at 560 nm or 640 nm or 670 nm) or absorbance at 530-560 nm.

2. Neuraminidase Inhibition Assay

1. Use 4 µl of NA as optimized in NA assay in tested wells and positive control well.

2. set positive control (no drug) by add reaction buffer 96 µl to NA to be 100 µl.

3. For drug containing well, add drug at various concentration by % volume of the positive control well (0%-no drug, 25%-drug 24 µl, 50% drug –drug 48 µl, 100% -drug 96 µl).

4. Add 50 µl of Amplex red solution (4) in each well.

5. Measure the fluorescence (at 560 nm or 640 nm or 670 nm) or absorbance at 530-560 nm.

6. Result shown, % reduction or inhibition calculated by comparing percentage of the Fluorescence level in each tested well with positive control well.

APPENDIX C: Oral Presentations and Poster Contributions to Conferences

Oral Presentation and Proceedings

1. Sukhontip Thaomola, Chak Sangma. AutoDock as a Tool for Virtual Screening for Anti-Influenza Drugs. The 10th Annual National Symposium, On Computational Science and Engineering (ANSCSE 2006), Mar 23-25, 2006, Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand.

Poster Contribution to Conferences

1. Sukhontip Thaomola, Chak Sangma. anti-HIV-1 RT Resistance Folds Prediction by AutoDock 3.0 and FRED Program.3 Science and Technology of Thailand (STT.32), October 10-1, 2006, QUEEN SIRIKIT NATIONAL CONVENTION CENTER (QSNCC), Bangkok, Thailand.

CURRICULUM VITAE

PERSONAL DATA

Name:	Miss Sukhontip Thaomola
Date of Birth:	August 22, 1981
Place of Birth:	Yasothon, Thailand
Nationality	Thai
Religion	Buddhism

EDUCATION

: YEAR	INSTITUTION	DEGREE
2000-2003	Thaksin University	B.Sc. (Physics)

SCHOLARSHIP

: YEAR	GRANT SOURCE
2004-2005	The Postgraduate on Education and research in Petroleum and Petrochemical Technology (ADB-MUA)