Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2010, 2(6): 53-59
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X

Molecular docking study on Hemagglutinin protein of H1N1 virus with recommended antiviral drugs

Vijay Tripathi*, Richa Sinha, Dwijendra Kumar Gupta***

ABSTRACT

Molecular docking is routinely used for understanding drug–receptor interaction in modern drug design. The goal of protein docking is to obtain a model for the bound complex from the coordinates of the unbound component molecules. Current docking methods evaluate a vast number of docked conformations by simple functions that measure surface complementarily. Many proteins undergo small side chain or even backbone movements on binding of different ligands into the same protein structure. This is known as induced fit and is potentially problematic for virtual screening of databases against protein targets. In this report we investigate the limits of the flexible protein approximation used by the docking program, AutoDock, through cross-docking using protein structures of influenza hemagglutinin. Here, we describe the suitability of antiviral drugs recommended against influenza for the docking and 3D structure prediction of hemagglutinin protein of the novel influenza A virus H1N1. The 3D structure of the macromolecular complex resulting from the protein-ligand association is a very useful basis to understand its specific functions. Homology model of hemagglutinin protein was constructed using MODELLER 9v6 and the model was energy minimized and validated using Gromacs to obtain a stable structure, which was further used for 3D structure prediction and docking through molecular docking studies using Autodock. The active sites were analyzed by the program Surface Racer.

Keywords: Protein-ligand docking, Novel influenza virus, Hemagglutinin, Antiviral drugs.

INTRODUCTION

Nowadays, molecular docking approaches are routinely used in modern drug design to help understand drug–receptor interaction. Most biological processes are known to take place through protein-ligand interactions. The three-dimensional structure of a protein-ligand complex could serve as a significant source of understanding the way proteins interact with each other and perform biological functions. Therefore, knowing the detailed structure of protein-ligand complexes at the atomic level has been very important issues in biological sciences. However this detailed structure analysis is not an easy job. In the protein data bank [1] where

experimentally determined 3D structures of proteins are stored, most of the protein structures are a single protein chain and only a small fraction (about 10 %) of the structures correspond to protein-ligand complexes. Two key elements are basically required for performing protein-ligand docking; an efficient conformational search algorithm and an accurate free energy function. The free energy function should be logically precise so that it can discriminate the native-like association of two constituent molecules from a variety of non-native associations. The search algorithm must be capable of exploring extensively the huge conformational space and can find conformations with free energy values near to the overall minimum [2]. There is always some imprecision in a free energy function. To overcome this discrepancy the search algorithm, instead of generating the conformation with lowest energy, it generates multiple low energy conformations. Based on the appropriate scoring function several structures are then selected, and proposed as candidates for the native-like structures for the complex.

One of the most well-known search algorithms is “Simulated Annealing (SA)” [3,4]. The major advantage of SA is its ability to avoid becoming trapped at local minima. However, since this method generates only one conformation at a time, the conformational multiplicity cannot be directly implemented. Another powerful search algorithm is “Conformational Space Annealing (CSA)” [5]. The CSA is based on the genetic algorithms (GA). The major advantage of the CSA is that it can produce various low-energy solutions. Genetic algorithm evolves the population of possible solutions through genetic operators to a final population, optimizing a predefined fitness function. AutoDock [6] is a similar suite of programs developed using genetic algorithms. The quality of the solutions usually depends on the preliminary genes, the number of evolutionary occurrence, and the fitness function to choose the more favorable conformers. Taking our previous studies into consideration the hemagglutinin protein of the H1N1 virus was found to be highly susceptible to mutations and has evolved in a phylogenetically significant way. Hence for further docking studies the HA protein has been preferred.

HA is a type I membrane protein consisting of 566 amino acids. Addition of carbohydrates to HA has been shown to be crucial for the accurate protein folding in the endoplasmatic reticulum [7]. The HA0, which is post-translationally modified, is then cleaved by cellular proteases into two subunits, HA1 (328 residues) and HA2 (221 residues), but remain linked by a disulphide bridge. The cleavage of HA0 is a prerequisite for the conformational change in the HA which occurs upon low pH and this change is essential for release of the viral genome into the cytosol [8]. There are three receptor-binding sites, one buried on each HA1 subunit, and they are protected and inaccessible to antibodies. Five antigenic sites have been identified using monoclonal antibodies on HA. These sites cover much of the surface of the globular head and binding of antibodies results in neutralisation of the virus.

Antiviral drugs are prescription medicines (pills, liquid or an inhaler) with activity against influenza viruses, including swine influenza viruses. Two classes of antivirals are currently available—the M2 ion channel inhibitors (i.e., the two adamantanes, amantadine and rimantadine) and the neuraminidase inhibitors (i.e., oseltamivir and zanamivir). Oseltamivir is a neuraminidase inhibitor, serving as a competitive inhibitor towards sialic acid, found on the surface proteins of normal host cells. By blocking the activity of the neuraminidase, Oseltamivir prevents new viral particles from being released by infected cells. Both rimantadine and the similar drug amantadine are derivates of adamantane. Amantadine is the organic compound known formally as 1-aminoadamantane. The mechanism of Amantadine's antiviral activity involves interference with a viral protein, M2 (an ion channel, which is required for the viral particle to become "uncoated" once taken inside a cell by endocytosis). Zanamivir was the first neuraminidase inhibitor commercially developed. It works by binding to the active site of the

neuraminidase protein, rendering the influenza virus unable to escape its host cell and infect others.

In most docking studies, conformational changes occur on ligand binding. This may only involve small side chain rotations to maximize interactions with the ligand [9], 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. 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. Hence knowledge of the protein–ligand interactions of the hemagglutinin protein with the specific antiviral drugs may give an important insight into the binding interactions and relativeness of the drug for the present pandemic swine flu. We first explain the methods with reference to our previous data calculated followed by finding the best docking results alongwith an analysis of active sites and cavities and then cumulating the results obtained so far to reach a conclusion for a potential drug target.

MATERIALS AND METHODS

Sequence retrieval:

Influenza A virus protein sequences were downloaded from NCBI. The FASTA sequence of the hemagglutinin protein was obtained from (http://www.ncbi.nlm.nhm.gov/). The target was identified by pdb-BLAST using the representative HA of the new strain (A/New Mexico/AF1900/2008(H1N1) as query (gene bank no. ACH69193.1). Following BLASTp query, an Influenza A Virus with PDBid: 1rvx_A was selected as target for docking with the drugs

Homology modeling of hemagglutinin:

In protein structure prediction, homology modeling, also known as comparative modeling, is a class of methods for constructing an atomic-resolution model of a protein from its amino acid sequence (the "query sequence" or "target") [10,11,12]. For the analysis of conserved regions between target and template sequences the CLUSTAL *W* server (www**.**ebi.ac.uk/clustalw/*) with* 1rvx_A as target was chosen from PDB BLAST hits**.** The obtained Model was validated using Gromacs using GROMOS 96 force fields and final Energy minimized using AutoDock to obtain stable structure for further studies.

Molecular dynamics:

Molecular dynamics (MD) is a form of computer simulation in which atoms and molecules are allowed to interact for a period of time under known laws of physics, by approximations of known physics, giving a view of the motion of the atoms. MD probes the relationship between molecular structure, movement and function. It was originally conceived within theoretical physics, but is applied today mostly in materials science and modeling of biomolecules. It employs algorithms from computer science and information theory. Gromacs was used for molecular dynamics of modeled protein.

Identification of cavities:

A cavity (or void) is an interior empty space that is not accessible to the solvent probe. It has no mouth openings to the outside bulk solution. The cavities in the receptor were mapped to assign an appropriate active site, the basic feature used to map the cavities were the surface mapping of the receptor and identifying the geometric voids as well as scaling the void for its hydrophobic characteristics. Hence all the cavities that were present in the target protein were identified and ranked based on their size and hydrophobic surface area. The Surface Racer program was used to analyze the cavities in the protein interior which are inaccessible to solvent from outside.

Considering the dimensions and hydrophobic surface area, cavity-1 was found to be the best void as an active site.

Identification of active sites

The active site of the target protein contains the catalytic and binding sites. The structure and chemical properties of the active site allow the recognition and binding of the substrate. The active site in many enzymes can be inhibited or suppressed by the presence of another molecule. The active site is usually a small pocket at the surface of the enzyme that contains residues responsible for the substrate specificity (charge, hydrophobicity, steric hindrance). PASS (Putative Active Sites with Spheres) software (http://www.ccl.net/cca/software/UNIX/pass/overview.shtml) used for active sites prediction of modeled protein.

Docking approaches:

Docking is the process by which two molecules fit together in 3D space. Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand—protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure. For the molecular docking analysis the Autodock version 3 has been used involving genetic algorithm. The Lamarckian genetic algorithm (LGA) was selected for the ligand conformational search. Autodock is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock program was developed using genetic algorithms. Genetic algorithm evolves the population of possible solutions through genetic operators to a final population, optimizing a predefined fitness function.

RESULTS AND DISCUSSION

A confidence level sequence identity gives a reliable alignment between the target sequence and template structure. Our PDB BLAST Hit for hemagglutinin target sequence gave a best hit score with 1rvx_A with an E-value of 5e-159, bit score of 556 with atomic resolution of its X-ray crystal structure obtained from diffraction studies being 2.20 Å and observed R-value of 0.227. Structurally conserved regions for the Model and the template were determined by multiple sequence alignment. According to Verify_3D 87.04% of the residues had an average 3D-1D score > 0.2 in model generated by modeler. Thereafter optimized structure was obtained by removing water molecules from the modeled structure. The generated 3D structure was further energy minimized using the Gromos96 forcefields in Gromacs. The energy minimized model gave an average total energy of -1.31519+06 Kcal/mol and an average root mean square deviation (rmsd) of 409.453. The other parameters are given in table 1.

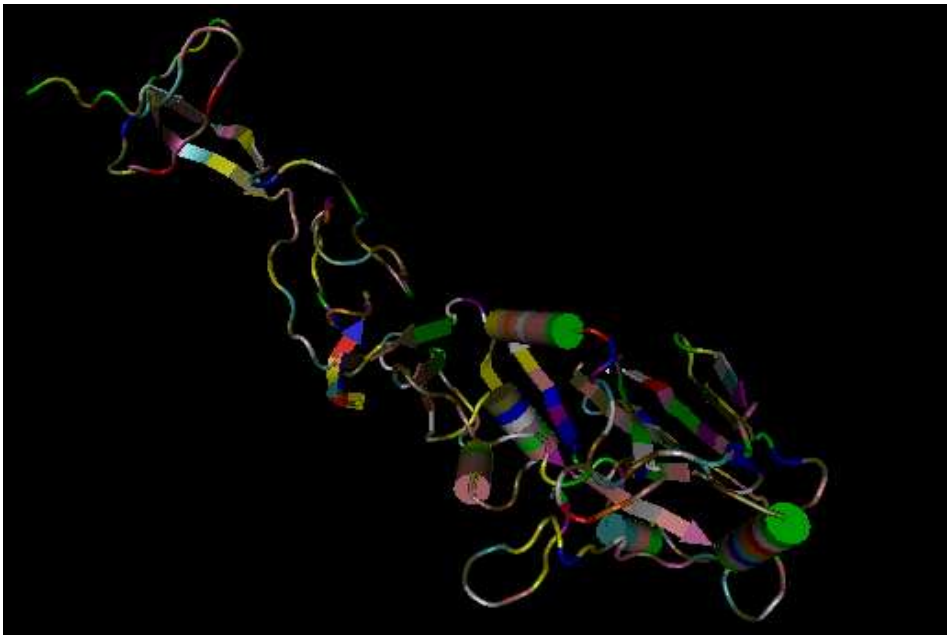

Fig 1: Modeled Protein of HA

Table 1: **Result of energy minimization**

Parameter	Average	RMSD	Fluctuation	Drift	Total drift
Potential energy	-1.60913e+06	1577.17	1572.95	-498	-399.39
Kinetic energy	293937	1511.17	1485.67	-1196.12	-959.286
Total energy	-1.31519+06	409.453	117.548	-1694.12	-1358.6
Pressure in bars	597.911	98.8168	0	-427.367	-342.746
Temperature	301.578	1.55055	1.5243	-1.22722	-0.984223
Heat capacity	12.4723 J/mol k (factor=2.64346e-05)				

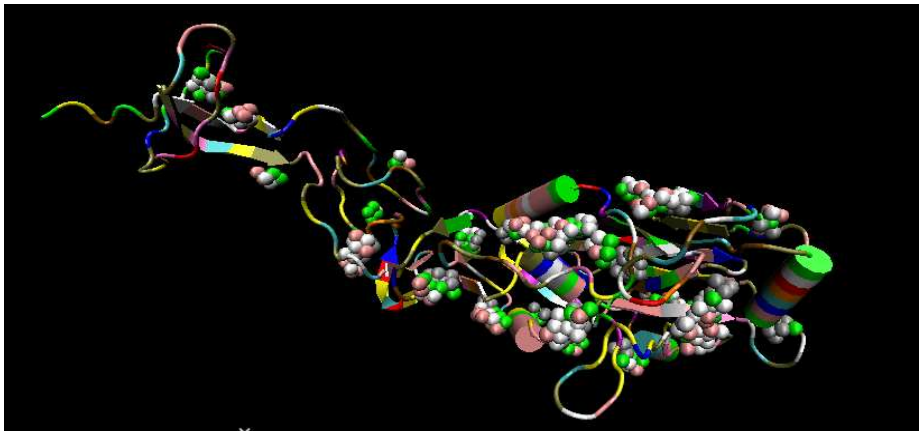

Fig 3: Active sites of protein HA (shown as beads)

The prediction of active sites using the modeled protein was done using the PASS software. The active sites have been shown in the figure 3 below. Small bead like structures are the active sites. The cavities predicted by the program Surface Racer resulted in 1 major cavity of 2 Å radius and 66 cavities of 1 Å radius. The single large cavity with its amino acid sequence and positions is shown in table 2 below.

Table 2: Residues in the single 2Å cavity

ATOM	GROUP	RESIDUE	POSITION
ATOM	CG2	VAL	131
ATOM	O	PHE	143
ATOM	NH1	ARG	145
ATOM	NH2	ARG	251

The ligand molecules were docked with the target protein using Autodock. Docking of amantadine with the target protein gave the best docking results with the minimum energy of -6.05 and 2 hydrogen bonds. The two hydrogen bonds of the protein with the ligand were formed with LYS 22 and GLU 21. The docking analysis of the four antiviral drugs is shown below.

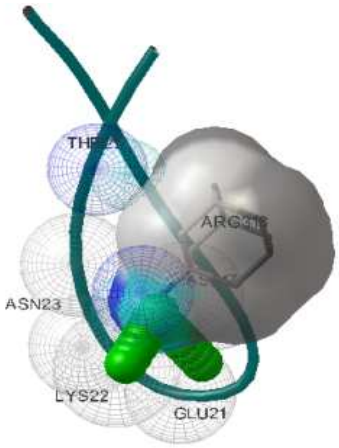

Fig 4: Docking with Amantadine

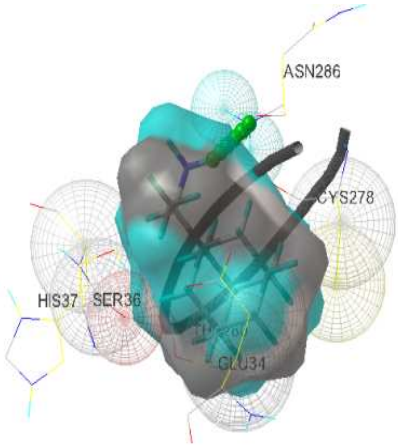

Fig 5: Docking with Rimantadine

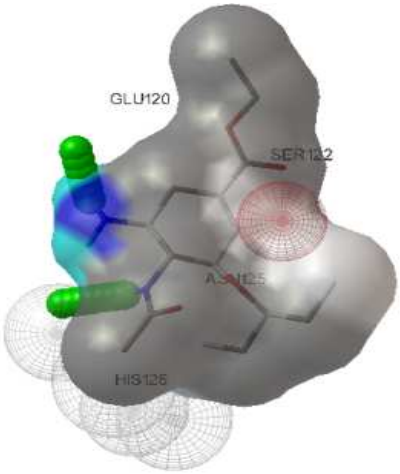

Fig 6: Docking with Tamavir

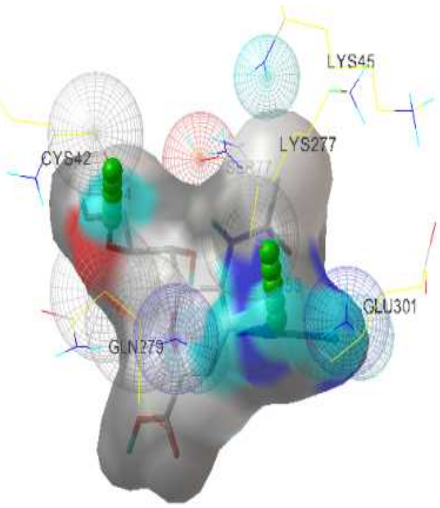

Fig 7: Docking with Zanamivir

Docking Data with Rimantadine

Rank	Run	Docked energy	RMSD
1	3	-6.45	98.36
2	4	-6.26	98.18
3	1	-6.17	98.15
4	9	-6.05	92.76
5	10	-5.81	92.62
6	7	-5.71	92.66
7	6	-6.04	97.53
8	5	-6.03	126.49
9	8	-6.01	126.50
10	2	-5.60	94.92

Docking Data with amantidine

Rank	Run	Docked Energy	RMSD
1	1	-5.31	95.35
2	10	-5.16	95.68
3	4	-5.06	97.41
4	8	-5.03	97.36
5	3	-4.87	97.81
6	5	-5.05	110.03
7	9	-4.89	109.01
8	7	-4.98	99.81
9	2	-4.88	102.12
10	6	-4.84	118.16

Docking Data with Tamavir

Rank	Run	Docked Energy	RMSD
1	3	-7.57	101.34
2	7	-5.91	103.79
3	5	-5.89	118.13
4	6	-5.73	131.38
5	1	-5.67	131.41
6	9	-5.66	101.81
7	2	-5.58	114.12
8	8	-5.57	129.00
9	10	-5.49	125.18
10	4	-5.45	126.48

Docking Data with Zanamivir

Rank	Run	Docked energy	RMSD
1	6	-5.01	125.19
2	9	-4.75	127.50
3	4	-4.69	127.62
4	8	-4.69	114.63
5	7	-4.64	124.59
6	1	-4.62	111.32
7	3	-4.56	107.08
8	10	-4.25	96.61
9	5	-4.01	128.89
10	2	-3.69	120.32

CONCLUSION

In this study of docking analysis of hemagglutinin protein with antiviral drugs, the amantadine which is an adamantane drug derivative is one of the most recent potent drug targets for H1N1. In this work, we have constructed a 3D Model of hemagglutinin, using the MODELLER software and obtained a refined model after energy minimization by Gromacs. The final refined model was further assessed by AutoDock program, and the results show that the model was stable and reliable.

REFERENCES

[1] H.M. Berman et. al. *Nucleic Acid Research* **2000**, 28, 235-242.
[2] K. Lee et. al., *Int. J. Mol. Sci.* **2008**, 9, 65-77
[3] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, *J. Chem. Phys*., **1953**, 21, 1087-1092.
[4] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Science* **1983**, 220, 671-680.
[5] J. Lee, H.A. Scheraga, S. Rackowsky, *J. Comput. Chem*., **1997**, 18, 1222-1232.
[6] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.*, **1998**, 19, 1639-1662.
[7] R.W. Doms, R.A. Lamb, J.K. Rose, A. Helenius, *Virology*, **1993**. 193(2), 545-62.
[8] H.D. Klenk, R. Rott, M. Orlich, J. Blodorn, *Virology*, **1975**. 68(2), 426-39.
[9] R. Najmanovich, J. Kutter, V. Sobolev, M. Edelman, Proteins, **2000**, 39, 261.
[10] S.Y. Chung, S. Subbiah, Structure ***1996***, 4, 1123–27.
[11] B. John, A. Sali, Nucleic *Acids Res* **2000**, 31(14), 3982-92.
[12] A. Sali, T.L Blundell, J. *Mol. Biol.* **2003**. 234, 779-815.