

Proposed lead molecules against Hemagglutinin of avian influenza virus (H₅N₁)

Tannistha Nandi^{1,*}

¹Department of Biotechnology, Jaypee Institute of Information Technology University, Noida (UP) 201307 India; Tannistha Nandi* - E-mail: tannistha3@yahoo.com; Phone: 91-120-2400972-972; Fax: 91-120-2400986; * Corresponding author

received December 21, 2007; accepted January 05, 2008; published January 27, 2008

Abstract :

Human infection with avian influenza H5N1 is an emerging infectious disease characterized by respiratory symptoms and a high fatality rate. Hemagglutinin and neuraminidase are the two surface proteins responsible for infection by influenza virus. Till date, neuraminidase has been the major target for antiviral drugs. In the present study we chose hemagglutinin protein as it mediates the binding of the virus to target cells through sialic acid residues on the host cell-surface. Hemagglutinin of H5 avian influenza (PDB ID: 1JSN) was used as the receptor protein. Ligands were generated by structure-based *de novo* approach and virtual screening of ZINC database. A total of 11,104 conformers were generated and docked into the receptor binding site using 'High Throughput Virtual Screening'. We proposed potential lead molecules against the receptor binding site of hemagglutinin based on the results obtained from *in silico* docking and hydrogen bond interaction between the ligand and the 1JSN protein molecule. We found sialic acid derivative 1 to be the lead molecules amongst the ligands generated by structure based *de novo* approach. However the molecules obtained from ZINC database were showing better docking scores as well as conserved hydrogen bond interactions. Thus we proposed ZINC00487720 and ZINC00046810 as potential lead molecules that could be used as an inhibitor to the receptor binding site of hemagglutinin. They could now be studied *in vivo* to validate the *in silico* results.

Keywords: lead molecules; hemagglutinin; avian influenza virus; ligand; virtual screening

Background:

Influenza virus belongs to the *Orthomyxoviridae* family, which consists of four genera: Influenza A virus, Influenza B Virus, Influenza C virus and Thogotovirus [1]. Influenza A virions are enveloped and contain eight segments of single-stranded, negative-sense RNA, which encode 11 proteins [1]. The Influenza A viral envelope has the surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA), as the major antigenic determinants of influenza viruses. There are 16 known HA antigenic subtypes (H1 to H16) all of which are found in aquatic birds; however sustained epidemics in humans have been limited to H1, H2 and H3 subtypes [2]. Nine NA subtypes are known (N1 to 9) of which only the N1 and N2 subtypes are circulating in human populations [1]. NA functions to release viruses attached to the host cell receptors and allow progeny virions to escape; thus, facilitating virus spread. Hemagglutinin (HA) is responsible for binding of virions to host cell receptors and for fusion between the virion envelope and the host cell [3]. HA molecules are homotrimers consisting of a globular head and a stalk; the head is made up of HA1 subunit and contains the receptor binding cavity. The stalk consists of the HA2 subunits and part of HA1 [3]. HA from avian and human influenza viruses preferentially bind sialic acid molecules with specific oligosaccharide side chains, alpha 2, 3 and alpha 2, 6 linkages, respectively [1].

In the recent past, avian influenza virus particularly H5N1 has become a major concern as an emerging respiratory virus. This virus was normally endemic in aquatic birds, but they crossed the species barrier in 1997 in Hongkong

[4, 5] leading to 6 human fatalities among 18 isolated cases. As of 14th December 2007, 340 cases and 208 deaths due to H5N1 have been reported from all over the world to the World Health Organization.

Currently, two licensed influenza antiviral agents are recommended for use in the United States during the 2007-08 influenza season: oseltamivir and zanamivir. These are chemically related antiviral medications known as neuraminidase inhibitors that have activity against both influenza A and B viruses [6]. However, no inhibitor molecule is available against the HA protein of influenza virus till date.

In the present study, we analyzed the binding site of avian H5 hemagglutinins (HAs) from avian influenza virus to human cell receptor analogs. We studied the docking interaction of derivatives of sialic acid, galactose and N-acetyl glucosamine molecules at the active site of avian H5 hemagglutinins. In addition, we had chosen molecules from a non-commercial small molecules database (ZINC) and studied their docking interaction at the active site of H5. From this study, we proposed lead molecules that competitively bind the active site of HA, prevent its attachment to the host cell surface moieties, and thus might serve as a good antiviral drug to control influenza virus infection.

Methodology:

Protein structure

Three-dimensional structure of hemagglutinin H5 (PDB ID: 1JSN) from influenza A virus strain A/DUCK/SINGAPORE/3/97 solved by X-ray crystallography at 2.4Å was retrieved from the Protein Data Bank [7]. It consists of two protein chains: chain A (325 residues), chain B (176 residues) and a ligand (Chain M) with o-sialic acid (SIA), d-galactose (GAL) and n-acetyl-d-glucosamine (NAG). The molecule 1JSN contains 216 waters of crystallization.

Protein structure preparation

The protein structure 1JSN was fixed by assigning bond orders, adding hydrogens, identifying chain information, hetero groups and water molecules. All the water molecules were deleted except for the 25th HOH molecule, which was found to stabilize the interaction of the ligand (SIA, GAL and NAG – chain M) at the binding site of chain A (Figure 1a). Finally, the hydrogen bonds were optimized and the structure was minimized to 0.30 RMSD.

De novo ligand building

We used Ligbuilder 1.2 to build ligand molecules from SIA, GAL and NAG within the binding pocket of 1JSN [8]. The protein molecule 1JSN was kept rigid, however the flexibility of the ligand molecule was considered. We defined the starting point or seed structure by retaining the core structure of the ring of SIA, GAL and NAG and replacing the groups attached to the core of the ring with H atoms (Figure 2a, Figure 2b, Figure 2c). These H atoms served as growing sites for building ligand molecules. Fragments from the library containing simple hydrocarbon chains, amines, alcohols, and even single rings were added to these growing sites. We generated a set of 500 ligand molecules each from SIA, GAL and NAG. The parameters used in generating ligand molecules were as follows: MW: 160-480; logP: -0.4 to 5.6; Hbond donor 1-5; Hbond acceptor: 1 to 10. Babel 1.6 an open source program was used to convert PDB format to mol2 format and vice versa.

Virtual screening of ligands

ZINC database containing over 4.6 million compounds in ready-to-dock, 3D formats was used for virtual screening [9]. ZINC database was searched using the SMILE script (C1CCOCC1) of tetrahydropyran structure (Figure 2d) and 900 lead molecules were retrieved.

Ligand preparation

Ligands obtained from ZINC database and those built using Ligbuilder1.2 were prepared for docking by generating states at a target pH of 7.0 ± 2.0 . Force field used was OPLS_2005. For each state of ligand, a maximum of 32 stereoisomers were generated by retaining the specific chirality, of which only one low energy ring conformation was chosen per ligand. This work was carried out using Schrodinger software.

Docking of ligands

The prepared ligands were docked into the binding site of 1JSN using the GLIDE module of Schrodinger. The grid was generated around this receptor binding site of 1JSN (Figure 1b) to define the space for docking of our ligand molecules.

For prepared sets of ligands, GlideScore (GScore) was given by: $GScore = a * vdW + b * Coul + Lipo + Hbond + Metal + BuryP + RotB + Site$, where, vdW = van der Waals energy; $Coul$ = Coulomb energy; $Lipo$ = Lipophilic contact; $Hbond$ = Hydrogen-bonding; $Metal$ = Metal-binding; $BuryP$ = Penalty for buried polar groups; $RotB$ = Penalty for freezing rotatable bonds; $Site$ = Polar interactions in the active site and the coefficients of vdW and $Coul$ were: $a = 0.065$, $b = 0.130$ for Standard Precision (SP) Glide

Results and discussion:

Most attention has been directed toward the type A influenza viruses because they alone have the potential for major antigenic shift and resultant pandemic spread. Hemagglutinin protein of influenza A virus is a viable target for the discovery and development of small molecule inhibitors of virus growth. In the present study, we analyzed the receptor-binding site of hemagglutinin chain A (1JSN) for its affinity to various ligand molecules.

According to a previous study 191 amino acid sequences of HA from influenza virus A belonging to all known HA subtypes were compared. The emphasis was given on functional sites (receptor-binding cavity with its right and left edges) and degree of their conservation in each subtype. It was shown that despite low degree of sequence similarity, the active site is well conserved [10]. Also it was established by earlier studies that GLN residue plays an important role in the host range restriction [11]. In accordance to them our results showed the core of 1JSN, H5 receptor binding site consisted of GLU 186 (OE1), GLN 222 (OE1), and VAL (O) 131 residues. These correspond to atom number 1485, 1764 and 1033 in the pdb entry of 1JSN. We retained only the 25th water molecule at the receptor binding site for docking studies as it was found to stabilize the interaction of the ligand molecule.

We had strictly followed Lipinski's rule of five while designing and screening the ligands. We used a molecular weight range of 160-480 in generating the ligands because of Lipinski's second rule with regard to drugs size selectivity on their permeation across cellular membranes [12]. We selected a range of logP from 0.4 to 5.6 because most drugs have a logP value around 3 [13]. We generated a total number of 11,104 ligand conformers (2476 from 500 Sialic acid, 3516 from 500 Galactose, 2537 from 500 NAG and 2575 from 900 lead molecules from ZINC database) for studying the docking interaction at the receptor binding site of HA.

We had screened the obtained ligand conformers by high throughput virtual screening for extremely accurate binding mode predictions using GLIDE module. GLIDE module was used as it provided consistently high enrichment at every level. Each of the ligand conformer was tested at the binding site and a docking score was obtained. We observed that hydrogen bond was the stabilizing force in this protein-ligand interaction. Hence, we computed the hydrogen bond interactions of the ligand with the receptor binding site of 1JSN protein chain A.

We chose the best two ligands each from SIA, GAL and NAG derivatives (Figure 3). Our results showed that

amongst the ligands designed *de novo* GAL derivatives had the highest docking scores, followed by NAG and SIA derivatives. The docking score was computed by considering parameters like van der Waals energy, Coulomb energy, lipophilic contact, Hydrogen-bonding, penalty for buried polar groups, penalty for freezing rotatable bonds and polar interactions in the active site. Based on the docking scores, it was clear that GAL derivatives were better as inhibitors than NAG derivatives, which in turn were better than SIA derivatives.

At this point, we computed the atomic interaction (hydrogen bond) of these ligands with the residues present at the receptor binding site of 1JSN. It was observed a hydrogen bond exists between atom 1764 from GLN 222 of chain A and sialic acid derivative 1. This interaction was also found between atom 1764 from GLN 222 of chain A and chain M of 1JSN. This *in silico* docking indicated that the relative position of the hydrogen-bonding site of the ligand is reasonably well matched to the receptor-binding site. Thus sialic acid derivative 1 could be one of the lead molecules which could competitively bind to the receptor-binding site of HA (chain A). Another significant observation was that NAG derivative 1 & 2 forms hydrogen bonds with the atom 1486(OE2) of GLU186 of chain A. This observation was also supported by a previous study stating influenza virus utilizes sialic acid as an initial receptor [14].

Studies show that the receptor for haemagglutinin is the terminal sialic acid residue of host cell surface sialyloligosaccharides, while sialidase (neuraminidase) catalyses the hydrolysis of terminal sialic acid residues from sialyloligosaccharides. Extensive crystallographic studies of both these proteins have revealed that the residues that interact with the sialic acid are strictly conserved. Therefore, these proteins make attractive targets for the design of drugs to halt the progression of the virus [15].

However the docking studies of ligand molecules obtained by screening of ZINC database gave higher docking scores compared to the ligands obtained from SIA, GAL and NAG derivatives. The best docking scores were obtained for 2-(hydroxymethyl)-6-(4-methylphenoxy)-tetrahydropyran-3, 4, 5-triol (ZINC00487720) and 2-(p-tolylaminomethyl) tetrahydropyran-2, 3, 4, 5-tetrol (ZINC00046810) (Figure 3). The atom number 1764 (OE1 GLN A 222) was found to be involved in stabilizing the ligand interaction at the receptor-binding site through hydrogen bond. Again this was found to be conserved between chain M - chain A of 1JSN, ZINC00487720- chain A and ZINC00046810 - chain A. Thus both from docking scores and conserved hydrogen bond interactions noted we found the ZINC compounds to be potent lead molecule which could be a better inhibitor than sialic acid derivatives.

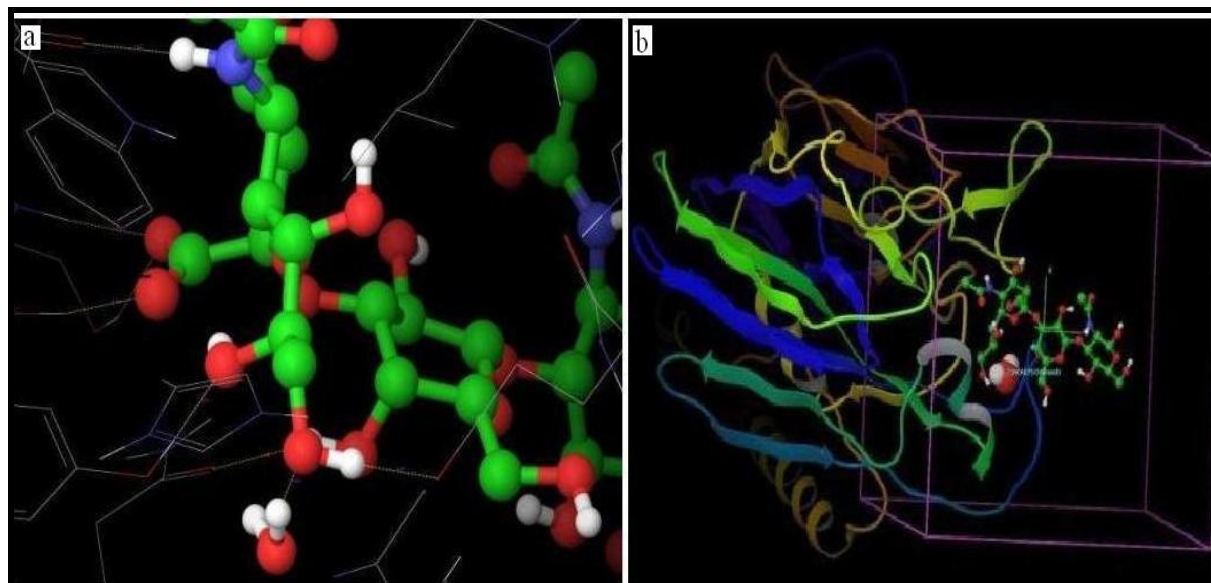


Figure 1: (a) The hydrogen bonds between the ligand, 25th HOH and chain A receptor binding site were represented by dotted yellow lines. The atom wise interactions were listed below (atom numbers were taken from PDB ID: 1JSN). (b) Grid generation step: the figure shows the grid box generated at active site of chain A (ribbons) in presence of 25th HOH (red and white) and the chain M (ligand - ball and stick model).

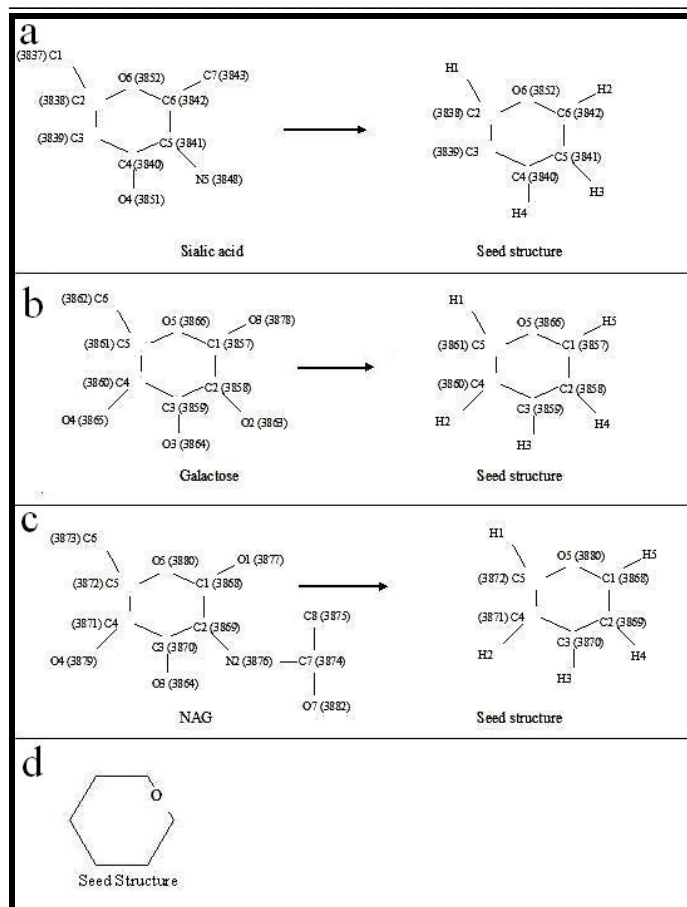


Figure 2: The carbohydrate molecules and the corresponding seed structures are shown in a, b and c. The tetrahydropyran structure is shown in d.

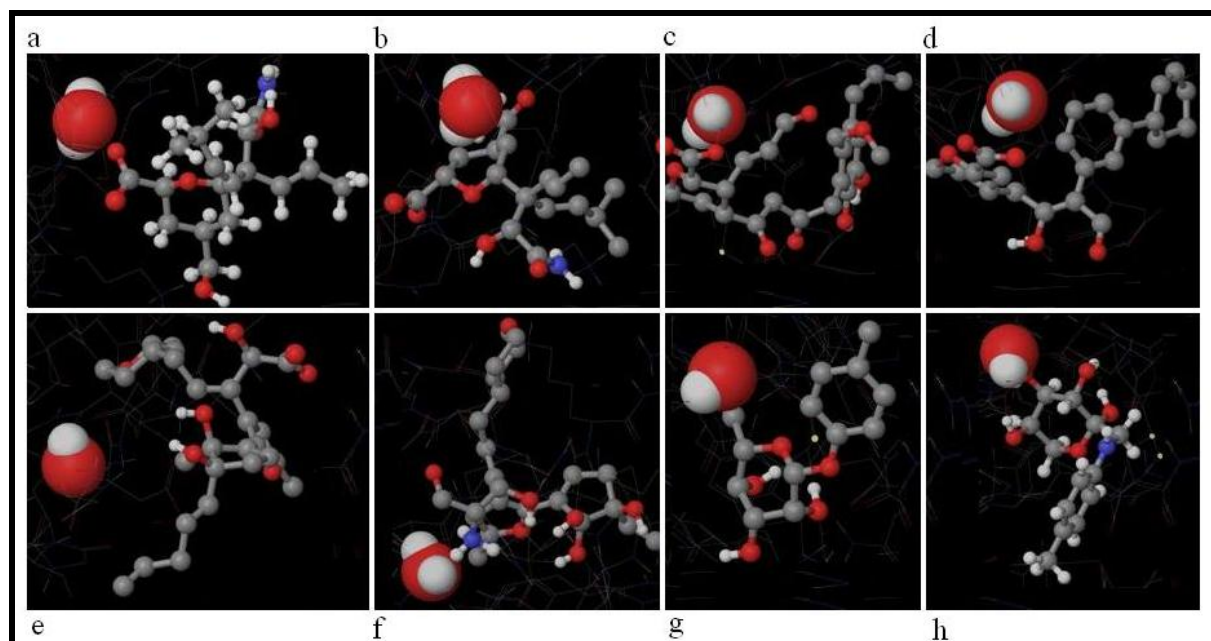


Figure 3: The GlideScore (docking score) is -4.76 for Sialic acid derivative 1 (a) and -4.72 for Sialic acid derivative 2 (b). The GlideScore is -5.73 for Galactose derivative 1 (c) and -5.61 for Galactose derivative 2 (d). The GlideScore is -5.55 for NAG derivative 1 (e) and -5.48 for NAG derivative 2 (f). The GlideScore is -6.59 for ZINC00487720 (g) and -6.57 for ZINC00046810 (h).

It was shown that the monosaccharide sialic acid cannot account for the molecular determinant of virus receptor-binding specificity in the context of the whole sialyloligosaccharide receptor. On the other hand, neuraminidase inhibition efforts using sialylmimetics have resulted into two antiinfluenza drugs, zanamivir and oseltamivir, which have been shown to reduce both the severity and duration of influenza illness. Nevertheless, the usage of monomeric sialylmimetics requires relatively large amounts of expensive compounds, which may also induce virus resistance and side effect. Therefore, it is critical to develop new antiinfluenza drugs and improve the current antiinfluenza drugs [16].

Thus current means of prevention of influenza virus infection are not entirely satisfactory because of their limited efficacy. A recent study reports CIO(2) gas could therefore be useful as a preventive means against influenza in places of human activity without necessitating evacuation [17]. Another study reports aptamer to be a promising candidate for treatment and prophylaxis of influenza virus infections [18].

Conclusion:

Hemagglutinin is the primary determinant of host range because of its role in host cell recognition and attachment. Molecules obtained by screening ZINC database were found to be better inhibitors than those derived from the NAG, GAL and SIA molecules. Hence we proposed ZINC00487720 and ZINC00046810 as potent lead molecules as they could competitively inhibit the binding of HA to the host cell receptor. The results obtained from the present study provided a step toward the preparedness of pandemic flu. The proposed lead molecules could be either used alone or in combination with neuraminidase inhibitors to treat viral influenza infection after *in vivo* testing.

Acknowledgement:

TN is thankful to IIIT University for all the support and facilities. TN is also thankful to Dr. Shashidhar N Rao from

Schrodinger, US for his contributions. TN is grateful to Prof Tan Tin Wee, NUS for his help.

References:

- [01] L. G. Ross & K. Subbarao, *Clin Microbiol Rev.*, 19: 614 (2006) [PMID: 17041137]
- [02] R. A. Fouchier, *et al.*, *J Virol.*, 79: 2814 (2005) [PMID: 15709000]
- [03] R. G. Webster, *et al.*, *Clin Microbiol Rev.*, 56: 152 (1992) [PMID: 1579108]
- [04] C. Bender, *et al.*, *Virology*, 254: 115 (1999) [PMID: 9927579]
- [05] K. Y. Yuen, *et al.*, *Lancet*, 351: 467 (1998) [PMID: 9482437]
- [06] A. E. Fiore, *et al.*, *MMWR Recomm Rep.*, 56: 1 (2007) [PMID: 17625497]
- [07] Y. Ha, *et al.*, *Proc Natl Acad Sci.*, 98: 11181 (2001) [PMID: 11562490]
- [08] R. Wang, *et al.*, *J. Mol. Model.*, 6: 498 (2000)
- [09] J. J. Irwin, *et al.*, *J Chem Inf Model*, 45: 177 (2005) [PMID: 15667143]
- [10] A. Kováčová, *et al.*, *Virus Genes*, 24: 57 (2002) [PMID: 11928990]
- [11] A. Vines, *et al.*, *J Virol.*, 72: 7626 (1998) [PMID: 9696865]
- [12] C. Rauch & A. Pluen, *Eur Biophys J.*, 36: 121 (2007) [PMID: 17211622]
- [13] E. Michalsky, *et al.*, *BMC Bioinformatics*, 6: 122 (2005) [PMID: 15943884]
- [14] D. S. Dimitrov, *Nat Rev Microbiol.*, 2: 109 (2004) [PMID: 15043007]
- [15] J. C. Wilson & M. von Itzstein, *Curr Drug Targets*, 4: 389 (2003) [PMID: 12816348]
- [16] X. L. Sun, *Curr Med Chem.*, 14: 2304 (2007) [PMID: 17896978]
- [17] N. Ogata & T. Shibata, *J Gen Virol.*, 89: 60 (2008) [PMID: 18089729]
- [18] C. Cheng, *et al.*, *Biochem Biophys Res Commun.*, 15: 670 (2008) [PMID: 18078808]

Edited by W. Cuff

Citation: Nandi, Bioinformatics 2(6): 240-244 (2008)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.