Structure Prediction and Active Site Analysis of New H1N1 Neuraminidase:Target for Antiviral Drug Design

VIVEK CHANDRAMOHAN[1*], MANJUNATH DAMMALLI[1], MAHANTESH BIRADAR[1], RICHARD L JAYARAJ[2], PARESH NATH KUNDU[1] and B. S. GOWRISHANKAR[1]

[1]Department of Biotechnology, Siddaganga Institute of Technology, Tumkur- 572103, Karnataka, India

[2]Department of Biotechnology, Periyar University, Salem, Tamilnadu, India

vivek.bioinf@gmail.com

Abstract: The H1N1 viral envelope protein neuraminidase encoded by NA gene plays a key role in the pathogenesis of swine flu. The active site of the neuraminidase protein is targeted by presently available antiviral drugs. The influenza virus often proves to be resistant to currently available drugs, due single amino acid substitutions conferred by the mutations in the gene coding for neuraminidase protein. The latest Influenza A virus A/Perth/262/2009(H1N1) sequence with accession number ADJ67981 was selected from NCBI. The BLAST program was used to identify the best template structure, which was found to be 3NSS_A. Sequence alignment was carried out with the template and query sequence, the identity and similarity was found to be 81.9% and 82.6% respectively. Homology modeling was performed using Accelrys Discovery Studio 3.5 software, the model with the lowest energy was then assessed for stereochemical quality and side-chain environment. The PDF energy and DOPE score of the best modeled structure was 2090.1682 and -43752.3632 respectively. Further active site optimization of the modeled protein was performed by molecular dynamics. The key active site residues which are crucial for further docking studies were ascertained.

Introduction

In earlier decades, trial and error method was widely used to design most of the antiviral drugs, after screening a huge number of druggable compounds. Gradually, this approach has been largely replaced by rational drug design, in which a target viral protein is pinpointed for the drug design[1]. *In-silico* computational approach can be applied to get a descriptive three dimensional structure of the protein and to select a target site in the identified protein[2]. In influenza viruses, NA surface protein plays a vital role in release of the virus from the host cell during the budding stage[3]. From the time when the crystal structure of NA was determined, it is used as a target protein for many drug compounds. Zanamivir and Oseltamivir

are two such FDA approved anti-influenza drug compound which targets the highly conserved NA active site of H1N1 Influenza virus, which is made up of 8 functional residues (R118, D151, R152, R224, E276, R292, R371 and Y406)[4,5]. The latest outbreaks of H1N1 and reports of Zanamivir and oseltamivir resistant strains have necessitated the need to find the effective alternatives to the existing anti-influenza drugs[6]. Oriental medicinal herbs with antiviral activity are currently in the spotlight as a complementary or alternative medicine.

Molecular modeling tools are utilized in generating new candidate drug molecules within a short period of time. After the generation of new probable drug candidates, the drug and target protein interaction dynamics can be assessed by carrying out docking analysis. The knowledge so obtained is applied to predict the strength of association or binding affinity between the two molecules based on scoring functions. A database of potential drug molecules and target protein structure serve as inputs for the docking analysis. The success of the docking process is determined by scoring function and search algorithm[7], which aids in determining the compatibility between the drug and its target protein. This approach is being used extensively to predict the geometries of different bimolecular complexes[8]. Scoring function predicts the strength of the binding affinity between ligand and the protein based on the complex geometry[9]. The search algorithm analyzes the drug molecule for different binding positions with its target molecule, each binding position, which is termed a "pose" and used to generate the snapshot of interactions[10].

In the present study, latest sequence of the gene coding for Neuraminidase was employed to build the N1 simulation structure by homology modeling. The model has been checked for high reliability by analysing score and Ramachandran plot. The currently built model is believed to have enormous scope for further docking analysis and to find out best probable drug like compounds with less side effects.

Experimental

From the literature, the viral envelope protein neuraminidase was identified as target protein in swine flu pandemic. The latest Influenza A virus A/Perth/262/2009(H1N1) sequence with accession number ADJ67981 was selected from NCBI (National Center for Biotechnology Information) for *in silico* analysis through homology modeling. Around 465 amino acids were found in the target sequence and protein belongs to glycoside hydrolase family.

The physicochemical properties of the protein were generated using CLC genomic workbench 5.1 by applying protein report protocol[11]. The characteristic values are as follows: Molecular weight-51.18 kDa, isoelectric point-6.07, aliphatic index-73.742, hydrophobic residues-345 and Hydrophilic residues-172. Secondary structure prediction was performed using GOR tool[12]. The structural details were found to be: α helix-228 aa, extended strand-135 aa and random coils-374.

Homology model construction

The Discovery Studio v3.5 was used for homology model construction. The homologous structures for the protein ADJ67981 were searched through NCBI – Blast (Basic local alignment search tool). The parameters of the applied algorithm are (BLOSUM62; E-threshold, 10) using pdbaa server [13]. Homologous template structure (the 2009 pandemic H1N1 neuraminidase N1) was selected for model building with PDB ID: 3NSS_A.

Sequence alignment[13] was carried out to identify the conserved regions by aligning target with the template structure. The aligned sequences were used for the model construction,

which was built using "Build homology model" protocol in DS 3.5. The protein model was generated by MODELER which was originally developed by Sali[14].

Protein simulation and validation

Modeled structure was refined by CHARMm forcefield[15] in DS Modeling protocol, which provides powerful mechanics and dynamics protocols for studying the energetics and motion of molecules, from small ligands to multi-component physiological complexes. Accelrys CHARMm forcefield was used throughout the simulation studies. Constraint was applied to allow only binding site and ligand to be flexible during the simulation. Potential energy of the modeled protein was analyzed before and after minimization by using calculates energy and minimization protocol respectively in DS 3.5[16].

The stereochemical quality of the structures was validated using 'PROCHECK'[17] and 'Verify 3D'[18]. Quality factors for the protein models were calculated using 'ERRAT2'[19]. Modeled structure was submitted to Protein Model Database[20], a repository for 3D protein models obtained by structure prediction methods.

Active site prediction

The active sites of the protein were predicted using DS 3.5, which is based on the receptor cavity method ("Eraser" algorithm)[21]. This study reveals the key residues in the target protein which are responsible for ligand binding, which are present in the active site or elsewhere.

Results and Discussion

Homology modeling methods use structural templates that have the highest sequence homology with the target protein. Homologous proteins were identified by scanning the protein sequence of ADJ67981 against 3D structures deposited in protein data bank using PDB BLAST. The 3NSS_A was found be the best template structure for Neurmaninidase protein. The target and template structure (3NSS_A) were aligned using sequence alignment protocol. The percentage of identity and similarity was found to be 81.9% and 82.6% respectively. The final alignment was carefully evaluated and was found to match the conserved residues fairly well (Figure 1). The 3D model of ADJ67981 was generated with MODELER protocol in DS 3.5. Out of five models generated during this process; the best model based on the lowest DOPE (Discrete Optimized Protein Energy) score and PDF (Probability Density Function Energy) energy was selected. The best model named as H1N.M0005 had a lowest PDF total energy of 2090.1682 and lowest DOPE score of -43752.3632, the remaining values are shown in the Table 1. MODELER is helpful in understanding the relationship between the binding-activity data and the binding conformational structure. Superimposition of PM0078710 model (Green color) over template (3NSS, Chain A in Blue color) is as shown in Figure 2 and best modeled structure is shown in Figure 3. The homology model of ADJ67981 was determined with identification of α-helices, β-sheets and coils as shown in Table 2.

Protein simulation and validation

Potential energy of the best modeled protein was checked before and after minimization by using calculates energy and minimization protocol respectively in DS 3.5 and the corresponding energy values were found to be 19438.52246 Kcal/mol and -24582.66832 kcal/mol. The standard dynamics cascade protocol was performed using DS 3.5 and the results are shown in Table 3.
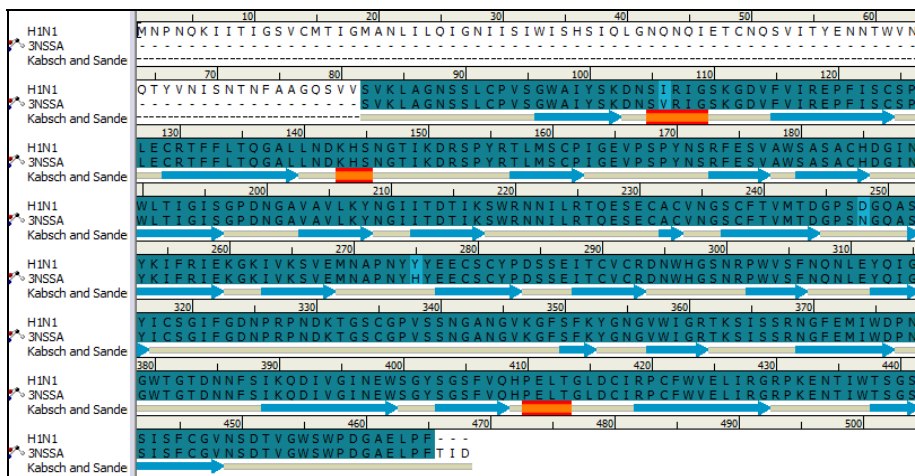
Figure 1. Sequence alignment of neuraminidase protein with template 3NNS_A. Deep green color shows conserved residue in all sequences (for color picture see online article)

Table 1. Results of homology modeling

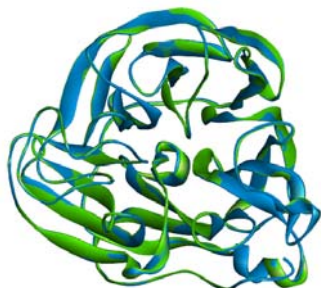| Name	PDF Total Energy	DOPE Score
H1N1.M0005	2090.1682	-43752.363281
H1N1.M0004	2170.0876	-43834.226563
H1N1.M0002	2769.6562	-42873.843750
H1N1.M0001	2844.8459	-42657.457031
H1N1.M0003	2996.1367	-42681.691406

Figure 2. Superimposition of PM0078710 model (Green) over template (3NSS, Chain A) (Blue)

Figure 3. Modeled protein PM0078710 in Solid display style by Discovery studio 3.5

Table 2. Comparison between model of neuraminidase based on highest homologue (3NNS). The nomenclature of α-helices and β-sheets

Modeled Structure		Template Structure (3NNS)	
Residues	Secondary structure	Residues	Secondary structure
GLY96 - LYS102	β Sheet	GLY96 - LYS102	β Sheet
SER105 - LYS111	A-Helix	SER105 - LYS111	A-Helix
PHE115 - CYS124	β Sheet	PHE115 - CYS124	β Sheet

Contd…

CYS129 - LEU139	β Sheet	CYS129 - LEU139	β Sheet
LYS143 - SER145	*B*- Helix	LYS143 - SER145	*B*- Helix
THR157 - PRO162	β Sheet	THR157 - PRO162	β Sheet
ARG173 - VAL177	β Sheet	ARG173 - VAL177	β Sheet
SER180 - HIS185	β Sheet	SER180 - HIS185	β Sheet
TRP190 - SER196	β Sheet	TRP190 - SER196	β Sheet
VAL203 - TYR208	β Sheet	VAL203 - TYR208	β Sheet
ILE212 - LYS217	β Sheet	ILE212 - LYS217	β Sheet
ALA232 - CYS233	β Sheet	ALA232 - CYS233	β Sheet
CYS238 - ASP244	β Sheet	CYS238 - ASP244	β Sheet
SER252 - ILE258	β Sheet	SER252 - ILE258	β Sheet
ILE263 - GLU268	β Sheet	ILE263 - GLU268	β Sheet
SER280 - PRO283	β Sheet	SER280 - PRO283	β Sheet
ILE288 - VAL291	β Sheet	ILE288 - VAL291	β Sheet
PRO302 - PHE306	β Sheet	PRO302 - PHE306	β Sheet
TYR312 - TYR316	β Sheet	TYR312 - TYR316	β Sheet
SER350 - LYS352	β Sheet	SER350 - LYS352	β Sheet
VAL357 - ARG361	β Sheet	VAL357 - ARG361	β Sheet
ASN369 - ASP376	β Sheet	ASN369 - ASP376	β Sheet
ILE389 - TRP399	β Sheet	ILE389 - TRP399	β Sheet
SER403 - GLN408	β Sheet	SER403 - GLN408	β Sheet
PRO410 - THR413	*C*-Helix	PRO410 - THR413	*C*-Helix
ARG419 - GLY429	β Sheet	ARG419 - GLY429	β Sheet
THR438 - VAL448	β Sheet	THR438 - VAL448	β Sheet
		PHE466 - ASP469	Coil

Table 3. Result for standard dynamics values

| Name	Angle Energy	Electro static Energy	Potential Energy	Van der Waals Energy	Bond Energy	Total Energy
Conformation 1	3,345.72	-26,139.4	-21,089.1	-2,410.5	2,000.67	-15,907.4
Conformation 2	3,401.44	-26,176.8	-21,100.2	-2,418.25	1,970.6	-15,907
Conformation 3	3,405.4	-26,285.6	-21,090.4	-2,367.68	2,098.48	-15,907
Conformation 4	3,409.82	-26,226.9	-21,106.8	-2,416.58	2,046.92	-15,908.1
Conformation 5	3,435.73	-26,383.5	-21,112.3	-2,326.99	2,010.27	-15,911.7
Conformation 6	3,398.27	-26,298.4	-21,168.6	-2,409.34	2,047.37	-15,912.2
Conformation 7	3,396.31	-26,309.6	-21,137.7	-2,376.59	2,083.36	-15,914.3
Conformation 8	3,446.41	-26,402.7	-21,205.4	-2,375.45	2,046.3	-15,917.6
Conformation 9	3,377.26	-26,406.2	-21,218.1	-2,363.65	2,072.95	-15,920.4
Conformation 10	3,446.44	-26,441.1	-21,206.8	-2,379.43	2,045.28	-15,924.3

The stereochemical properties for the best modeled protein were verified using 3D profiles verify algorithm available in DS 3.5. The verified score was observed as 187.16 that fall between expected high score of 175.081 and expected low score of 78.7864. The graphical representation of 3D profile score is shown in Figure 4. Model validation was carried out using SAVS server.

To verify the protein model, the coordinates of the protein model were submitted to PROCHECK. The percentage of residues in the most favorable regions of Ramachandran

plot was 88.8% and of those falling in the disallowed region was 0.0%. The remaining plot statistics are as shown in Figure 5. The volume of protein was assessed using PROVE (PROtein Volume Evaluation), the characteristics of protein volume were analysed using *Z*-Score mean 1.379, *Z*-Score RMS 35.182 and remaining values are as shown in Figure 6.

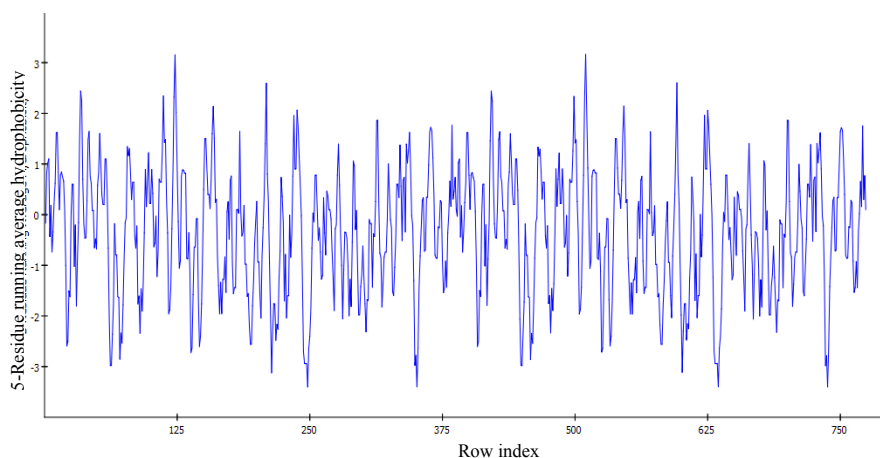**Figure 4.** 3D profiles verify graph for modeled protein (PM0078710)

Modeled structure (PM0078710)   Template structure (3NNS)
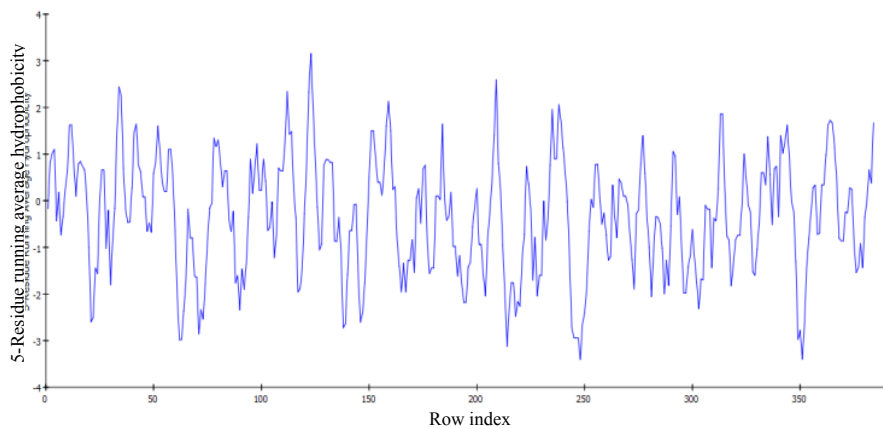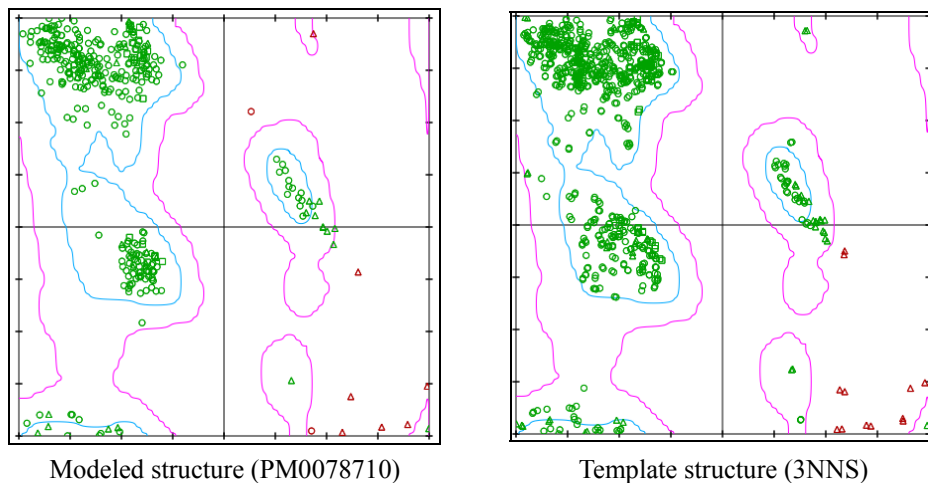
Figure 5. Comparison of the evaluation of the modeled structure and highest homologue structure (3NNS)

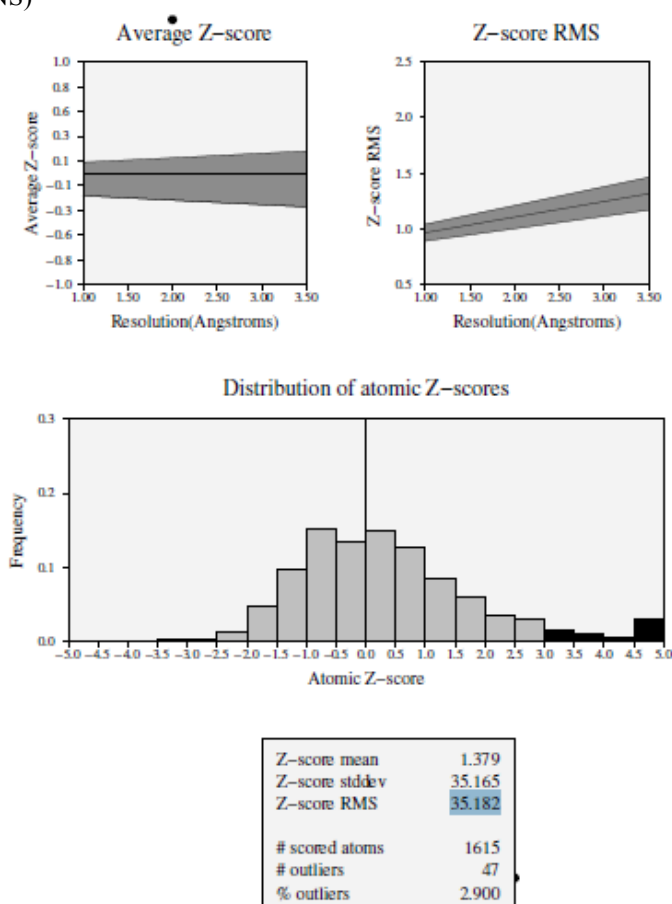| Z–score mean | 1.379 |
Z–score stddev	35.165
Z–score RMS	35.182
# scored atoms	1615
# outliers	47
% outliers	2.900

Figure 6. Main chain parameters for the model generated by PROVE

Modeled structure was deposited to Protein Model Database which is a repository for 3D protein models obtained by structure prediction methods. The modeled protein can be retrieved from PMDB using accession number PM0078710.

Active site prediction

Based on the receptor cavity method 64 amino acids were found in ligand binding pocket. 17 hydrophobic residues were found in the 5Å region of the best ligand binding pocket, which are PHE115l, VAL116, ILE117, PHE121, PHE132, PHE133, LEU134, ALA138, ILE149, LEU158, ALA178, ALA181, CYS279, VAL346, VAL424, LEU426 and ILE427. These key amino acids are involved in the hydrophobic interaction with ligand compounds (Figure 7).
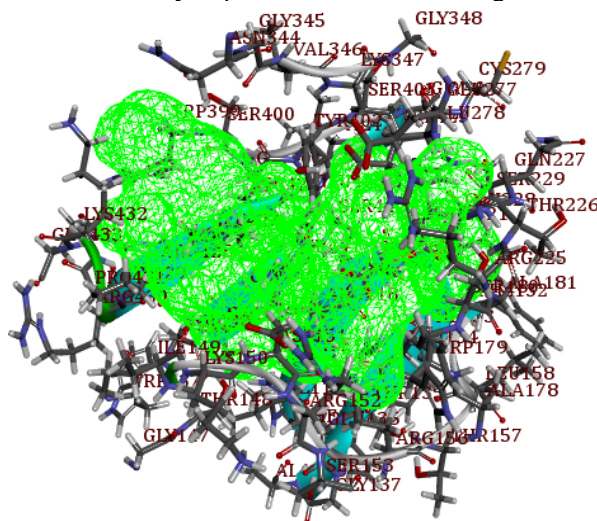

Figure 7. Active site of modeled protein PM0078710 by Discovery studio 3.5 Note: green color indicates binding area of the active site of modeled protein (PM0078710)

Conclusion

The incidence of viral disease has increased dramatically over the past few years and the worldwide spread of H1N1 influenza A virus is a matter of serious concern. The viral envelope protein neuraminidase was identified as target protein in swine flu pandemic. Two antiviral agents, Zanamivir (Relenza) and Oseltamivir (Tamiflu) have been shown to help prevent or reduce the effects of swine flu. Swine flu viruses may often mutate (change) and are easily transmissible among humans and become ineffective against currently available treatments. Hence there is a need to search for novel natural compounds which are active against neuraminidase and with fewer side effects. *In silico* targeting of the mutated neuraminidase protein requires a valid 3D protein model to assess the binding pattern and interaction with the ligands. In the present study the latest Influenza A virus A/Perth/262/2009(H1N1) sequence with accession number ADJ67981 was selected from NCBI (National Center for Biotechnology Information) for *In silic*o analysis through homology modeling. 3NSS_A was chosen as template for protein modeling and 5 models were generated using DS 3.5. The best model with lowest PDF total energy (2090.1682) and DOPE score -43752.3632 was chosen. Further model validation studies showed that the modeled protein is a reliable target. *In silico* drug targeting using this model will assist in screening and identifying a potent drug for H1N1 swine flu.

References

1. Noble D and Colatsky T J, *Emerging Therapeutic Targets*, 2000, **4**, 39-49.
2. Ekins S, Mestres J and Testa B, *Brit J Pharmacol.*, 2007, **152(1)**, 21-37; DOI:10.1038/sj.bjp.0707306
3. John Cartar B and Venetia Saunders A, *First*, 2007, 317-322.
4. Von Itzstein M, Pegg M S, Dyason J C, Wen-Yang Wu, Gaik B Kok, Betty Jin, Tho Van Phan, Mark L Smythe, Hume F White, Stuart W Oliver, Peter M Colman, Joseph N Varghese, Michael Ryan D, Jacqueline M Woods, Richard Bethell C, Vanessa J Hotham, Janet M Cameron and Charles R Penn, *Nature*, 1993, **363(6428)**, 418-423; DOI:10.1038/363418a0
5. Aoki F Y, Boivin G. and Roberts N A, *Anti-viral Therapy,* 2007, **12**, 603-616.
6. Manden J, Kääriäinen L and Ahola T, *Appl Microbiol Biotechnol*, 2005, **66(6)**, 612-621; DOI:10.1007/s00253-004-1783-3
7. Kitchen D B, Decornez H, Furr J R and Bajorath J, *Nature Rev Drug Discov.,* 2004, **3**, 935-949; DOI:10.1038/nrd1549
8. Irawin Kuntz D, Elaine Meng C and Shoichet Brain K, *Accounts Chem Res.*, 1994, **27(5),** 117-123; DOI:10.1021/ar00041a001
9. Jain A N, *Curr Protein Pept Sci.,* 2006*,* **7(5),** 407-420.
10. Shoichet B K, Kuntz I D and Bodian D L, *J Comput Chem.,* 2004, **13(3)**, 380-397; DOI:10.1002/jcc.540130311
11. Sander H, Eefje, Schrauwen J A, Linster Martin, Salin Chutinimitkul, Emmie de Wit, Vincent J Munster, Erin M Sorrell, Theo M Bestebroer, David F Burke, Derek J Smith, Guus F Rimmelzwaan, Albert D M E Osterhaus and Ron A M Fouchier, *Science,* 2012, **336(6088)**, 1534-1541.
12. Garnier J, Gibrat J F and Robson B, *Doolittle Ed.*, 1996, **266**, 540-553.
13. Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E, *Nucleic Acids Res.*, 2000, **28(1)**, 235–242; DOI:10.1093/nar/28.1.235
14. Smith S W, Overbeek R, Woese C R, Gilbert W and Gillevet P M, *Comput Appl Biosci.,* 1994, **10(6)**, 671-675.
15. Fiser A and Sali A, *Bioinformatics*, 2003, **19(18)**, 2500-1; DOI:10.1093/bioinformatics/btg362
16. Brooks B R, Bruccoleri R E, Olafson B D, States D J, Swaminathan S and Karplus M, *J Comput Chem.,* 1983, **4(2)**, 187–217; DOI:10.1002/jcc.540040211
17. Fletcher R and Reeves C M, *Comput J.,* 1964, **7(2)**, 149- 154; DOI:10.1093/comjnl/7.2.149
18. Laskowski R A, MacArthur M W, Moss D S and Thornton J M, *J Appl Cryst.,* 1993, **26**, 283–291.
19. Asad Ullah Khan, Mohd Hassan Baig and Gulshan Wadhwa, *Bioinformation*, 2011, **5(8),** 331-335
20. Tiziana Castrignano, Paolo D'Onorio De Meo, Domenico Cozzetto, Ivano Giuseppe Talamo and Anna Tramontano, *Nuc Acid Res.*, 2006, **34(suppl 1),** D306-D309; DOI:10.1093/nar/gkj105
21. Venkatachalam C M, Jiang X, Oldfield T and Waldman M, *J Mol Graph Model*, 2003, **21(4)**, 289–307; DOI:10.1016/S1093-3263(02)00164-X