BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access

www.bioinformation.net

Hypothesis

Volume 8(3)

In silico modeling of ligand molecule for non structural 3 (NS3) protein target of flaviviruses

Shefali Agnihotri*, Ranjana Narula, Kaishiv Joshi, Sandeep Rana & Maneet Singh

Department of Bioinformatics, ADI Biosolution, Mohali, India 160059; Shefali Agnihotri - Email: agnihotri.shefali02@gmail.com Phone: +91-9780270139; *Corresponding author

Received December 30, 2011; Accepted January 07, 2012; Published February 03, 2012

Abstract:
Flaviviruses are small, enveloped RNA viruses which cause a variety of diseases into animals and man. Despite the existence of licensed vaccines, yellow fever, Japanese encephalitis and tick-borne encephalitis also claim many thousands of victims each year across their vast endemic areas. A number of studies have already revealed that the non-structural NS3 serine protease is required for the maturation of the viral polyprotein and thus is a promising target for the development of antiviral inhibitors. Hence, the 3D structure of NS3 protein was modeled using homology modeling by MODELLER 9v7. Validation of the constructed NS3 protein models were done by PROCHECK, VERYFY3D and through ProSA calculations. Ligands for the catalytic triad (H51, D75, and S135) were designed using LIGBUILDER. The NS3 protein's catalytic triad was explored to find out the interactions pattern for inhibitor binding using molecular docking methodology using AUTODOCK Vina. The interactions of complex NS3protein-ligand conformations, including hydrogen bonds and the bond lengths were analyzed using Accelrys DS Visualizer software. Hence, from this observation, the novel molecule designed was observed to be the best ligand against the NS3 protein of flavivirus. This molecule may prove to be a potential identity in modulating disease manifestation for all the selected flavivirus members.

Keywords: NS3 protein, homology modeling, virtual screening, docking, ligand.

Abbreviations: NCBI; National Centre for Biotechnological Information, BLAST; Basic Local Alignment Search Tool, DOPE, Discrete optimized protein energy; GROMOS96, GROningen MOlecular Simulation package, SAVS; Structure Analysis and Validation Server.

Background:
Flaviviruses are small, enveloped RNA viruses which are generally transmitted by arthropods to animals and man. Birds and mammals are the principal vertebrate hosts for flaviviruses **[1].** These flaviviruses all share a similar genomic organization and replication strategy, and yet cause a range of distinct clinical diseases in humans **[2].** Dengue virus causes an estimated 50 million cases of febrile illness each year, including an increasing number of cases of hemorrhagic fever. West Nile virus, which recently spread from the Mediterranean basin to the Western hemisphere, causes thousands of sporadic cases of encephalitis annually. Despite the existence of licensed vaccines, yellow fever, Japanese encephalitis and tick-borne encephalitis also claim many thousands of victims each year across their vast endemic areas. Antiviral therapy could potentially reduce morbidity and mortality from flavivirus infections, but no effective drugs are currently available **[3].** The viruses within the Flaviviridae family are associated with significant public health and economic impacts worldwide. Of the 3 genera in this family, the Flavivirus genus is the largest, composed of 53 species divided into 12 groups. The 4 most common species causing human disease include the Japanese

© 2012 Biomedical Informatics

encephalitis virus (JEV), Murray Valley encephalitis virus (MVEV), St. Louis encephalitis virus (SLEV), and the West Nile virus (WNV). **[4]** A number of studies have already revealed that the non-structural NS3 serine protease is required for the maturation of the viral polyprotein and thus is a promising target for the development of antiviral inhibitors **[5].** The ~11 kb flavivirus RNA genome is a positive-sense, single stranded,5`-capped RNA ((+)saran) that is released into the cytoplasm immediately following cell entry. It encodes a single, large polyprotein, which is proteolytically processed to yield three structural proteins (envelope, E; membrane precursor, Pram; and cased C) and seven non-structural (NS) proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5). **[6]** The 7 nonstructural proteins are vital for replication of the Flaviviridae**. [4]** NS3 is a multidomain protein, with an N-terminal NS3Pro **[6].** In this in-silico study, we have developed molecule inhibitor of NS3pro for 22 species of genus flavivirus using structure based drug designing. The interaction between NS3 protein and inhibitor were studied by docking methods using Auto Dockvina. The interactions of complex NS3protein-ligand conformations, including hydrogen bonds and the bond lengths were analyzed using Accelrys DS Visualizer software .We hope, this Drug will get success to clear out all the phases of clinical trial and it will be effective drug in the cure of flavivirus diseases.

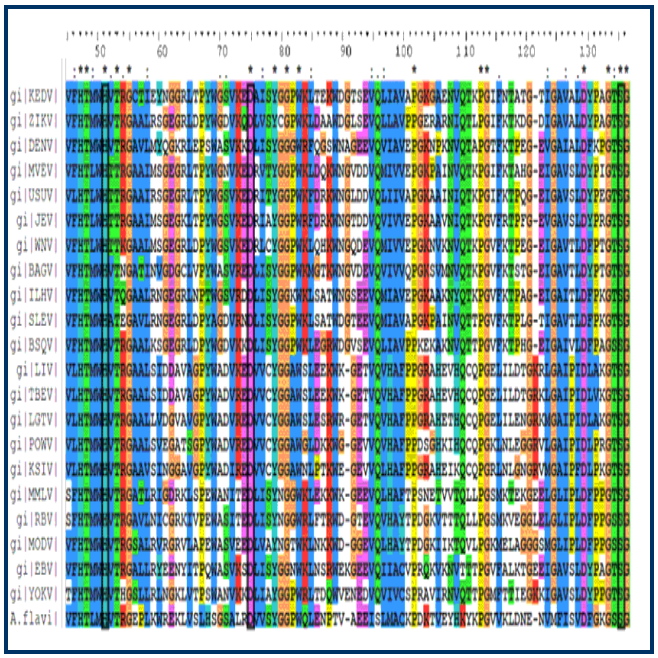

Figure 1: The sequence alignment between NS3 proteins of 22 species. All three major amino acids forming the catalytic Triad (H51, D75, and S135) have been highlighted**.**

Methodology:

Sequence alignment

The protein sequence of NS3 of 22 species was obtained by NCBI database (http://www.ncbi.nlm.nih.gov/) showing in given **Table 1 (see supplementary material)**. Using the Protein –protein blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) through NCBI, the homologous structure ofMVEV NS3 was identified, which was used as template for the homology modeling. Multiple sequence alignment of the aminoacid sequences of 22 species were performed with the online version of CLUSTALW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) program to identify the set of conserved residues in the alignment **(Figure1).**

Protein homology modeling

The homology modeling was carried out using the Modeller (http://www.salilab.org/modeller/) 9v7 program. The target and the template sequences were aligned using Modeller 9v7, a comparative protein modeling program, was used for homology modeling to generate the 3-D structures of NS3 protein for 22 species. Final homology model was selected on the basis of MOLPDF, DOPE score GA341 score.

Loop Refinement

The alignment between target and template sequence contains gaps. These gaps results for the loops in the 3d structure. So forfurther refinement of 3d models, loop refinement step was performed by using Modeller (http://www.salilab.org/modeller/) 9v7program and the best model was selected on the basis of molpdf value.

Model optimization and evaluation

The protein models of 22 species generated by homology modeling often produce unfavorable bond lengths, bond angles,torsion angles and bad contacts. Therefore, it was essential to minimize the energy to regularize local bond and angle geometry as well as to remove bad contacts. Energy mininmisation were done with the GROMOS96 (Scott *et al*., 1999) force field by implementation of Swiss-PdbViewer (http://www.expasy.org/spdbv). After the optimization procedure, the 3D models of NS3 were verified by using PROCHECK (Laskowski *et al*., 1993) Program of Structural Analysis and Verification Server (SAVES).The quality of models was also validated by ProSA (Wiederstein *et al*., 2007) (Sippl, 1993) server(https://prosa.services.came.sbg.ac.at/prosa.php), a web server for Protein Structure Analysis.

Active site identification

The active sites were revealed on the basis of previous studies. The aminoterminal domain contains the serine protease catalytic triad consisting of amino acid residues H51, D75, and S135 and the substrate-binding pocket is contained within NS3 protein **[7].**

Virtual Screening of the Lead

Ligand library was downloaded from Accelrys DS Visualizer software ("http://accelrys.com"). Then lead molecule was screened out from 2930 molecules on the basis of molecular properties. The fragment "1 H-1, 2, 4-triazole" was identified onthe basis of "Lipinski's Rule of Five" and may therefore represent suitable starting point for evolution of good quality lead compounds. MolSoft, OSIRIS and Molinspiration were used to design drug molecule on computer to define its molecular structure **Table 2 (see supplementary material).**

Rigid Docking

Hex 4.5 was used for the purpose of docking of the lead with the target molecule. The lead compound "1 H-1, 2, 4-triazole" and MVEV NS3 protein were opened in the Hex (http://hex.loria.fr/) and ligand was attached to the residue on the minimum distance position to the active site position .The

© 2012 Biomedical Informatics

docking controls were activated with default parameters .The ligand was hence docked to the receptor protein **(Figure2).**

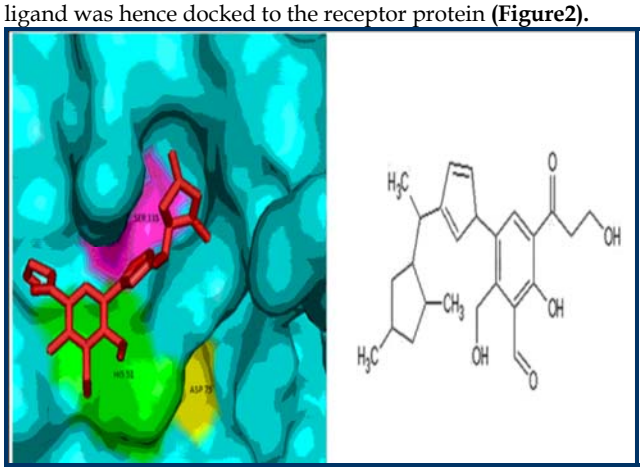

Figure 2: View of the surface of the catalytic traid HIS 51 (green), ASP 75 (yellow) and SER 135 (pink) shows docking with inhibitor 1 H-1, 2, 4-triazole (red).

Generation and optimization of Ligand

Ligbuilder was used for the generation of the ligands. The conformation of the pre-placed "seed" ensuring the binding affinity decides the manner that ligands would be grown with Ligbuilder software. Novel ligands had been developed with Ligbuilder(http://mdl.ipc.pku.edu.cn/drug_design/work/ligbuilder.html) v1.2 software. We developed 200 novel ligands for the inhibitorysite in NS3 protein. Virtual screening, an *insilico* tool for drug discovery, has been widely used for lead identification in drug discovery programs. Out of 200 novel ligands generated, 10 ligands were selected on the basis of maximum binding affinity measured in kcal/mol. The selected 10 ligands were then analyzed for drug- relevant properties based on "Lipinski's rule of five" and other drug like properties of valid structures using OSIRIS Property Explorer (http://www.organicchemistry.org/prog/peo/), Mol soft: Drug- Likeness and molecular property explorer (http://www.molsoft.com/mprop/). On the basis of binding affinity and drug like properties, one ligand that passed all of the screening tests was taken for further molecular docking study.

Protein-ligand docking

The docking of ligands to the catalytic triad of NS3protein for 22 species was performed using AutoDock Vina software.

Docking was performed to obtain a population of possible conformations and orientations for the ligand at the binding site. Using the software, polar hydrogen atoms were added to the NS3protein and its nonpolar hydrogen atoms were merged. Allbonds of ligands were set to be rotatable. All calculations for protein-fixed ligand-flexible docking were done using theLamarckian Genetic Algorithm (LGA) method. The grid box with a dimension of 20 x 20 x 20 points was used around thecatalytic triad to cover the entire enzyme binding site and accommodate ligands to move freely. The best conformation was chosen with the lowest docked energy, after the docking search was completed. The interactions of complex NS3protein-ligandconformations, including hydrogen bonds and the bond lengths were analyzed using Accelrys DS Visualizer software ("http://accelrys.com").

Discussion:

The sequence of the 22 species protein encoding for NS3 was retrieved from NCBI Database .The similarity searches was performed by protein- protein blast. The 100% similarity was found in MVEV among 22 species. So MVEV(PDB entry2WV9) , was used as template for protein homology modeling .The predicted 3D structure of NS3 protein was generated by Modeller and the structure with the lowest DOPE scores were selected . The alignment between target and template sequence contains gaps .So loop refinement step was also performed by using Modeller. The best models were selected on the basis of molpdf value. The modeller generated models statistically analyzed by structure analysis and validation server (SAVS).The structure submitted were validated and zero bad contacts was used for the further process at lead target prediction .The final protein structures selected after analysis in SAVS. The Rigid docking was performed by HEX in which protein and ligand were opened in docking software and was attached to the residue on the minimum distance positon to the active site position .The ligand was hence docked to the receptor protein. The Ligbuilder tool was used for the inhibitor generation. After generation of the lead molecule it was then screened for its activity and its drug likeness. Web based tools like Molinspiration, and OSIRIS property explorer were used for this purpose. Molinspiration uses sophiscated Bayesian statistics to compare structures of the representative ligands active on the particular target site. In OSIRIS we draw chemical structures to calculate various drug revelant properties. The binding pattern analyzed by AUTODOCK, is used to predict small molecule to the receptors of known 3D structure. The ligand and target protein were given as input and the flexible docking was performed. The negative and low value of ΔGbind indicates strong favorable bonds between NS3 protein and the novel ligand indicating that the ligand was in its most favourable conformations. The interactions of complex NS3protein-ligand conformations, including hydrogen bonds, sigma and pi bonds were analyzed using Accelrys DS Visualizer software ("http://accelrys.com") **Table 3 (see supplementary material)**

Conclusion:

Although there are still no specific vaccines or chemotherapeutic regimes for prevention and treatment for flavivirus based diseases like dengue, hemorrhagic fever, encephalitis etc. In recent years there has been substantial progress in our understanding the life cycle of flavivirus, the various stages of which represent potential targets for the development of novel antiviral drugs. NS3 protein is a particularly interesting molecular target for antiviral compounds because of it central role in all the viral life cycle of flaviviruses. A challenging aspect in the search for potent, selective antiviral drugs that interfere with multifunctional NS3 is the design of appropriate assays for druggable sites that are relevant for viral replication *in vivo*. Inhibitors of NS3pro should also be of great benefit in combating infections by other Flaviviruses, as well as Japanese encephalitis virus and West Nile virus. The development of such drugs requires a more informed structure-based drug discovery program. A further consideration includes the cost of drug synthesis which should not make the price of the final product prohibitive to poor patients in developing countries. The best highly active lead

© 2012 Biomedical Informatics

compound was docked into the active site of 22 species of flaviviruses. Thus, we hope that the lead molecules generated from this structure based drug designing of NS3 protein would be helpful in identifying structurally diverse compounds with desired biological activity for the successful treatment of various types of diseases in flaviviruses.

References:

[1] Shellam GR *et al, Rev sci tech. Off.in. Epiz* 1998 **17**: 231
[2] Fernandez-Garcia MD *et al. Cell Host Microbe*. 2009 **5**: 318 [PMID: 19380111]
[3] Sampath A & Padmanabhan R, *Antiviral Res.* 2009 **81**: 6. [PMID:18796313]
[4] Beth K. Schweitzer *et al. Labmedicin*. 2009 **40**: 8
[5] Bollati M *et al. Antiviral Res.* 2010 **87**: 125 [PMID:19945487]
[6] Yamshchikov VF & Compans RW, *J Virol*. 1995 **69**: 1995. [PMID:7884844]
[7] Mueller NH *et al. Antimicrob Agents Chemother*. 2008 **52**: 3385 [PMID: 18606844]

Edited by P Kangueane
Citation: Agnihotri *et al*. Bioinformation 8(3): 123-127 (2012)

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.

BIOINFORMATION

open access

Supplementary material:

Table 1: List of 22 species of family Flaviviridae

Name of species	Abbreviation	Protein Id's
Aedes flavivirus	AflAVI	YP_003084129.1
Bagaza virus	BAGV	YP_002790883.1
Bussuquara virus	BSQV	YP_001040004.1
Dengue virus type 2	DENV2	NP_739587.2
Entebbe bat virus	EBV	YP_950477.1
Ilheus virus	ILHV	YP_001040006.1
Japanese encephalitis virus	JEV	NP_775670.1
Kedougou virus	KEDV	YP_002790882.1
Kokobera virus	KOKV	YP_001040007.1
Langat virus	LGTV	NP_740299.1
Louping ill virus	LIV	NP_740726.1
Modoc virus	MODV	NP_619758.1
Montanamyotis leukoencephalitis virus	MMLV	NP_775649.1
Murray Valley encephalitis virus	MVEV	NP_722535.1
Powassan virus	POWV	NP_775520.1
Rio Bravo virus	RBV	NP_776076.1
St. Louis encephalitis virus	SLEV	YP_001008348.1
Tick-borne encephalitis virus	TBEV	NP_775507.1
Usutu virus	USUV	YP_164814.1
West Nile virus	WNV	YP_001527884.1
Yokose virus	YOKV	NP_872627.1
Zika virus	ZIKV	YP_002790881.1

Table 2: Ligand Properties

Descriptors	Lead compound Scores
LogP	-1.41
Solubility	-1.4
Molecular Wt	73.0
TPSA	36.081
nON	3
nOHNH	3
nviolations	0
nrotb	0
Volume	72.604
Number of HBA	3
Number of HBD	3
Mutagenic	No
Tumorigenic	No
Irritant	No
Reproductive	No

Table 3: Active sites which are highlighted (RED) shows interaction with target protein NS3

Name of Species	Active sites	Number of bonds
AFLAVI	His, Asp, Ser	1 H
BAGV	His, Asp, Ser	1 Pi
BSQV	His, Asp, Ser	1 Pi
JEV	His, Asp, Ser	3H,1 Pi
DENV2	His ,Asp, Ser	2H
EBV	His ,Asp, Ser	1 H
KEDV	His, Asp ,Ser	1H
MVEV	His, Asp, Ser	1H
USUV	His, Asp ,Ser	1H
WNV	His, Asp, Ser	1H,1 Pi
YOKV	His, Asp, Ser	3 Pi , 1 sigma
ZIKV	His, Asp, Ser	2Pi

© 2012 Biomedical Informatics