International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:6, No:12, 2012

Chikungunya Protease Domain–High throughput Virtual Screening

Surender Singh Jadav, Venkatesan Jayaprakash, Arijit Basu, and Barij Nayan Sinha

***Abstract*—**Chikungunya virus (CHICKV) is an arboviruses belonging to family Tagoviridae and is transmitted to human through by mosquito (*Aedes aegypti* and *Aedes albopictus*) bite. A large outbreak of chikungunya has been reported in India between 2006 and 2007, along with several other countries from South-East Asia and for the first time in Europe. It was for the first time that the CHICKV outbreak has been reported with mortality from Reunion Island and increased mortality from Asian countries. CHICKV affects all age groups, and currently there are no specific drugs or vaccine to cure the disease. The need of antiviral agents for the treatment of CHICKV infection and the success of virtual screening against many therapeutically valuable targets led us to carry out the structure based drug design against Chikungunya nSP2 protease (PDB: 3TRK). Highthroughput virtual screening of publicly available databases, ZINC12 and BindingDB, has been carried out using the Openeye tools and Schrodinger LLC software packages. Openeye Filter program has been used to filter the database and the filtered outputs were docked using HTVS protocol implemented in GLIDE package of Schrodinger LLC. The top HITS were further used for enriching the similar molecules from the database through vROCS; a shape based screening protocol implemented in Openeye. The approach adopted has provided different scaffolds as HITS against CHICKV protease. Three scaffolds: Indole, Pyrazole and Sulphone derivatives were selected based on the docking score and synthetic feasibility. Derivatives of Pyrazole were synthesized and submitted for antiviral screening against CHICKV.

***Keywords*—**Chikungunya, nsP2 protease, ADME filter, HTVS, Docking, Active site.

I. Introduction

CHIKUNGUNYA virus (CHIKV) (family *Togaviridae*, genus *Alphavirus*) is transmitted by *Aedes aegypti* and *Aedes albopictus* mosquitoes. It causes an acute febrile illness associated with severe joint pain that can persist for a long time even after viral clearance. A large outbreak of chikungunya has been reported in India between 2006 and 2007[1]. Due to changing patterns of vector distribution, abundance in response to climate change and increased vector-human contact, CHIKV is regarded as a potential worldwide public health problem, with no preventive or therapeutic means available. Infection is transmitted via infected mosquitoes [2]. In the vertebrate host, transient viremia and dissemination occur as virus is released from cells that later lyses. Infection with seroconversion in the absence of clinical disease is common, but disease can be incapacitating and, in cases of encephalitis, occasionally fatal. Virus is eliminated by the immune system but arthritis or central nervous system impairment may persist for weeks [3,4]. The enveloped virions are spherical, 60 to 70 nm in diameter with a positive-sense, monopartite, single-stranded RNA genome, ca. 11.7 kilobases long. The lipid-containing envelope has two (rarely three) surface glycoproteins that mediate attachment, fusion, and penetration. The icosahedral nucleocapsid contains capsid protein and RNA. Virions mature by budding through the plasma membrane [3,4]. CHIKV viruses attach to cells, probably via interactions between E2 and a poorly defined family of cellular receptors found on many vertebrate and invertebrate cells. Entry is thought to take place in mildly acidic endosomal vacuoles where glycoprotein spikes undergo conformational rearrangements and an acid-dependent fusion event (principally a function of E1) delivers genomic RNA to the cell cytoplasm. Viral replication occurs in the cytoplasm. Initial translation of virion RNA produces a polyprotien that is proteolytically cleaved into an RNA polymerase. Transcription of the virion RNA through a negative-strand RNA intermediate produces a 26S positive-strand mRNA which encodes only the structural proteins, as well as additional 42S RNA, which is incorporated into progeny virions. Translation from the 26S mRNA (which represents the 3′ one-third of genomic RNA) produces a polyprotein that is cleaved proteolytically into three proteins: C, PE2, and E1; PE2 is subsequently cleaved into E2 and E3. Envelope proteins formed by posttranslational cleavage are glycosylated and translocated to the plasma membrane. Virion formation occurs by budding of preformed icosahedral nucleocapsids through regions of the plasma membrane containing E1 and E2 glycoproteins. Genomic RNA is capped and polyadenylated and serves as mRNA for nonstructural proteins (e.g., RNA-dependent RNA polymerase) which are encoded in the 5′ two-thirds of the genome. Complementary (antisense) RNA, made from genomic RNA, serves as a template for progeny genomic RNA. A subgenomic mRNA representing the 3′ one-third of the genome encodes the structural proteins [3-5, 13].

CHIKV is enveloped, single stranded positive sense RNA virus having genome of≈12,000 nt, encoding four non-structural (ns1–4) and three main structural proteins (C, E2 and E1) with organization as: 5′-cap-(non-structural proteins)-(junction region)-(structural proteins)-Poly (A) tail-3′. Prototype viruses like Sindbis Virus (SINV) and Semliki Forest Virus (SFV) of this family have been extensively studied. Open Reading Frame1 encodes a polyprotien precursor termed as nsP1234 and processed into different non

Surender Singh Jadav is with Department of Pharmaceutical Sciences, Birla Institute of Technology Mesra, Ranchi, Jharkhand, India (e-mail: surenderjs@gmail.com).

Venkatesan Jayaprakash is with Department of Pharmaceutical Sciences, Birla Institute of Technology Mesra, Ranchi, Jharkhand, India (e-mail: venkatesanj@bitmesra.ac.in).

Arijit Basu is with Department of Pharmaceutical Sciences, Birla Institute of Technology Mesra, Ranchi, Jharkhand, India (e-mail: abasu@bitmesra.ac.in).

B. N. Sinha is with Department of Pharmaceutical Sciences, Birla Institute of Technology Mesra, Ranchi, Jharkhand, India (e-mail: bnsinha@bitmesra.ac.in).

structural proteins (nsP1, nsP2, nsP3 and nsP4) after stepwise proteolytic cleavages. The 3′-terminal one-third of the genome encodes for viral structural polyprotein which is expressed from a separate subgenomic mRNA and cleaved co-translationally and posttranslationally into structural proteins C, E1, E2 [1]. Alphavirus nsP4, the viral RNA-dependent RNA polymerase (RdRp) and the processing intermediates are mature products of nsP123 and are essential components of the viral RNA replication and transcription complexes [5, 13]. Alphaviral nsP2 is a multifunctional protein [3-6]. Nucleoside triphosphatase, helicase, and RNA-dependent 5′-triphosphatase activities have been located in the N-terminus of the protein while the proteolytic domain has been mapped to its C-terminal part [3-6]. The nsP2 protease is responsible for cleavages in the non-structural polyprotein [7]. Helicase seems to be essential for the function of viral RdRp in positive sense RNA viruses [8]. In addition, it may be involved in capping, RNA translocation, genome packaging, protection of RNA at replication center, modulating RNA-protein interactions etc.

Function of cytoplasmic capping enzyme, nsP1, is necessary since all viral RNAs are synthesized in the cytoplasm, and host-capping enzymes are restricted to the nucleus. The enzymatic reaction involves a covalent link between 7-methyl-GMP and nsP1, whereas eukaryotic capping enzymes form a covalent complex only with GMP. Capping mediated by nsP1 would consist of the following reactions: GTP is first methylated and then forms the m7GMp-nsP1 complex, from which 7-methyl-GMP complex is transferred to the mRNA to create the cap structure. Palmitoylated nsP1 is remodeling host cell cytoskeleton, and induces filopodium-like structure formation at the surface of the host cell. The two separate domains in nsP2 has different biological activities. The N-terminal section is part of the RNA polymerase complex and has RNA trisphosphatase and RNA helicase activity. The C-terminal section harbors a protease that specifically cleaves and releases the four mature proteins. Non-structural protein, nsP3 is essential for minus strand [14] and subgenomic 26S mRNA synthesis and nsP4 is a RNA dependent RNA polymerase. It replicates genomic and antigenomic RNA by recognizing replications specific signals. Transcribes also a 26S subgenomic mRNA by initiating RNA synthesis internally on antigenomic RNA. This 26S mRNA encodes for structural proteins [5-8].

A. CHIKV nsp2 Protease Catalytic Mechanism or Proteolytic Activity

The Protease domain of CHIKV nsp2 is organized into two domains. Sequence based analyses have shown helicase domain at the N-terminus and protease domain at C-terminus of nsP2. The CHIKV nsP2 protease has two catalytic residues **Cys1013** and **His1083**, which were involved in proteolytic activity of nsP2 polyprotein [9-12]. In all other alphavirus domains the Tryptophan residue is present next to the catalytic histidine and it is highly conserved, unlike others here **Trp1084** residue is present and it is highly conserved in CHIKV protease Subsites. The catalytic cystiene is situated at the N terminus of an α helix, and the catalytic histidine is located on β strand [13].

II. Materials and Methods

All computational analysis was carried out on a Red Hat 5.0 Linux platform running on a Dell Precision workstation with Intel core 2 quad processor and 8GB RAM.

A. Protein Preparation

X-ray crystal structure of CHIKV nsP2 protease PDB: 3TRK) downloaded from www.rcsb.org. The protein (PDB: 3TRK) was prepared using the Protein Preparation Wizard. Preprocessed (Bond orders were assigned, Hydrogens were added, Metals were treated, and Water molecules were deleted) and heterostate for co-crystallized ligand was generated using Epik, protonation state and optimization of H-bonding of the protein side chains were assigned using Protassign. Energy minimized (impref minimization) using OPLS2001 force field.

B. Receptor Grid generation

Receptor Grid has been prepared with default parameters and without any constrains. Site specified with catalytic residues Cys1013 & His1083 and a conserved residue Trp1084.

C. Database Preparation

Drug-like property sub-set from publicly available databases ZINC12 (http://zinc.docking.org) and protease inhibitors from BindingDB (http://www.bindingdb.org) was downloaded. Eliminating the molecules with undesirable properties reduced volume of the database. This was done with Openeye Filter Tool using Drug-like filter parameter file.

D. Ligand Preparation

Using Ligprep utility of Schrodinger Suite with default parameters, the ligands in filtered database were prepared for docking. OPLS 2005 force field was employed to minimize the ligands.

E. Ligand-Docking

The ligand-docking was performed with **Glide** module in Schrodinger LLC. HTVS protocol in Glide was employed to enrich the database. Top 10% HITS were then subjected to docking with XP (Extra Precision) protocol implemented in Glide. Top 100 molecules were manually analyzed to pick up molecules with different scaffold and synthetic feasibility. Picked HITS were then served as query to build a library of closely related molecules based on shape based screening with Openeye's vROCS tool. The HITS from the library were then manually analyzed to pick up synthetically feasible candidates.

III. Results and Discussions

A huge chemical database with around 10,173,210 (10,090,210 from drug-like property sub-set of ZINC12 + 83,000 protease inhibitors from BindingDB) was subjected to ADME/TOX filter using Openeye's FILTER tool. This has

reduced the volume of database to 5,000,000. A two-tyre docking approach was employed. The first-tyre utilizes fast docking algorithm to quickly identify the molecules that fits in to the binding pocket. This led to the drastic reduction in time and computational workload. HTVS protocol in Glide module was employed to achieve this task and top 10% of the HITS (enriched database: 5000,000 molecules) were taken to the second-tyre. Extra-precision protocol (XP-protocol) in Glide was then employed to carry-out the second-tyre docking. Top 100 HITS from the second-tyre docking was analyzed manually.

Molecules were picked based on structural diversity (different scaffold), no. of interaction & nature of interaction with receptor and synthetic feasibility of the molecule in our laboratory infra-structure. Table I summarizes the list of molecules picked with their ZINC ID, structure and their docking sore, while Table II summarizes their interaction with the active residues. The approach adopted is depicted in Fig. 1.

We observe that ligands with bulkier groups are not accommodated in S1' pocket due to the steric clashes with S3 (**Tyr1079**, **Trp1084 & Met1238)** & S4 (**Ser1048**) residues. This makes very few ligands and their functionalities to reach the S1' pocket and mostly getting accommodated in S2 and S3 having good interaction with the residues lining it.

The top ten molecules were selected in such a way that a portion of the molecule or functional group is getting accommodated in S1' pocket with reasonable interaction with the residues lining it. Fig. 2 shows 2D plot of docked conformation of Ligand ZINC67680487 with 3TRK. Portion of indole ring is accommodated in S1' pocket and the phenyl ring of indole nucleus exhibit hydrophobic interaction with side chain of catalytic residue **Cys1013**. Also the molecule exhibits five H-bond interactions with **Tyr1047**, **Tyr1079**, **Trp1084** and **Asn1246**. We observe that majority of the ligands in the top HITS contain a specific architecture that interacts with above-mentioned residues. A common link –NH-C(=O)-C(=X)- present in the HIT molecules ensures the above mentioned interaction possible. Also it is observed that –C(C=X)- portion may be in open chain or portion of aromatic/hetero-aromatic ring. In the listed HITS, we found in general it is a part of phenyl or small heterocyclic ring.

Few HITS & their analogues identified through this study were synthesized and submitted for antiviral screening and results are awaited.

IV. Conclusion

The current study employed HTVS approach for identifying few novel HITS against recently released X-ray crystal structure of CHICKV protease (3TRK). We identified a common link required for interaction with residues lining S1', S3 and S4 pockets. Few HITs and Their analogues were synthesized and submitted for antiviral screening. The experimental result will substantiate the fact that the common link proposed might be a potential pharmacophoric feature for designing protease inhibitors against CHICKV.

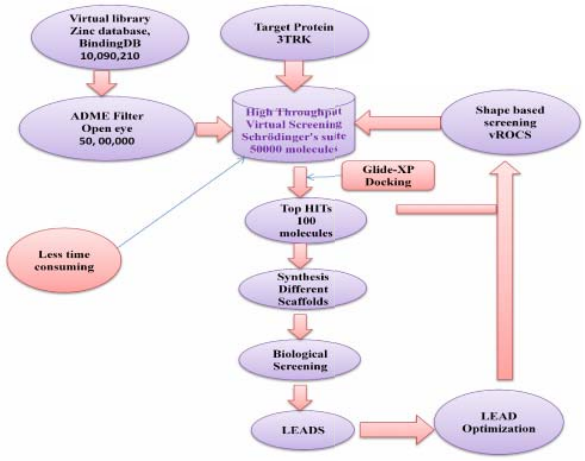

Fig. 1 ADME Filter and High throughput Virtual Screening against CHIKV nsP2 Protease (PDB: 3TRK)

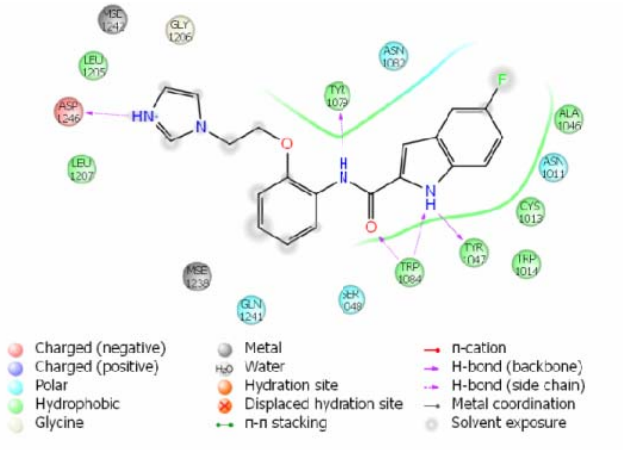

Fig. 2 2D plot of docked conformation of ligand ZINC67680487 with 3TRK

TABLE I
Docking Score and Code of Top10 HITs

Code	Docking Score	xLogP
ZINC67680487	-9.609	2.15
ZINC58049973	-9.564	2.83
ZINC67802867	-9.496	1.40
ZINC65395467	-9.390	0.58
ZINC67921389	-9.384	1.14
ZINC63656937	-9.381	0.47
ZINC55133607	-9.194	0.67
ZINC67680547	-9.187	3.06
ZINC67973564	-9.151	1.97
ZINC36594806	-9.145	2.00

Acknowledgment

Authors are gratefully acknowledged to Department of Biotechnology (DBT), India for financial support during this period, New Indigo-Era net for the grant. We are also thankful

to Department of Pharmaceutical Sciences, Birla Institute of Technology Mesra for providing infrastructure and facilities.

REFERENCES

[1] C. Lahariya and S. K. Pradhan, "Emergence of Chikungunya virus in Indian subcontinent after 32 years: a review", Journal of Vector Borne Diseases, vol. 43, pp. 151-160, 2006.
[2] S. A. Talawar and H. S. Puja, "An outbreak of chikungunya epidemic in south India Karnataka", International Journal of Research & Reviews in Applied Sciences, Vol. 5, pp. 229-234, 2010.
[3] J. H. Strauss, "The alphaviruses: gene expression, replication, and evolution", Microbiology Reviews, pp. 58491-562, 1994.
[4] T. P. Monath, "The Arboviruses: epidemiology and ecology", Vol. 5, CRC Press, Boca Raton, FL.
[5] J. A. Lemm, T. Rumenapf, E. G. Strauss, J. H. Strauss and C. M. Rice, "Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus and plus-strand RNA synthesis", EMBO J, vol. 13, pp. 2925–2934, 1994.
[6] W. R. Hardy and J. H. Strauss, "Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans", Journal of Virology, vol. 63, pp. 4653–4664, 1989.
[7] K. H. Kim, T. Rumenapf, E. G. Strauss and J. H. Strauss, "Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus pathogenesis in invertebrate hosts", Virology, vol. 323, pp. 153–163, 2004.
[8] G. Kadare and A. L. Haenni, "Virus-encoded RNA helicases", Journal of Virology, vol. 71, pp. 2583–2590, 1997.
[9] Andrew, T. R. Mark, A. W. Stanley and J. Watowich, "The Crystal structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease", Structure, vol. 14, pp. 1449-1458, 2006.
[10] E. Mossessova and C. D. Lima, "Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast". Molecular Cell, vol. 5, pp. 865-876, 2000.
[11] Andrew, T. R. Mark, A. W. Stanley and J. Watowich, "Structural basis for substrate specificity of Alphavirus nsP2 proteases, Journal of Molecular .Modeling, vol. 29, pp. 46-53, 2010.
[12] K. D. Singh, P. Kirubakaran, S. Nagarajan, S. Sakkiah, K. Muthuswamy, D. Velmurugan and J. Jeyakanthan, "Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease", Journal of Molecular modeling, vol. 18, pp. 39-51, 2012.
[13] Y. A. Karpe, P. P Aher and K. S. Lole, "NTPase and 5'-RNA Triphosphotase Activities of Chikugunya Virus nsP2 Protien", Journal of PLoS ONE, vol. 6, pp. 22336, 2010.
[14] H. Malet, B. Coutard, S. Jamal, H. Dutartre, N. Papgeorgiou, M. Neuvonen, T. Ahola, N. Forrester, E. A. Gould, D. Lafitte, F. Ferron, J. Lescar, A. E. Gorbalenya, X. de.Lamballerie and B. Canard, "The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsp3 Macro domains define a Conserved Adenosine binding pocket", Journal of Virology, pp. 6534–6545, 2009.