Computational analysis of the molecular principle behind Boceprevir resistance due to mutation in Hepatitis C NS3/4A protease and the development of new NS3/4A protease inhibitor

A Major Project dissertation submitted

in partial fulfillment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Neha Nagpal

(2K12/BIO/017)

Delhi Technological University, Delhi, India

Under the supervision of

Prof. B. D. Malhotra

Department of Biotechnology

Delhi Technological University

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road,

Delhi-110042, INDIA

CERTIFICATE

This is to certify that the M.Tech. dissertation entitled **"Computational analysis of the molecular principle behind Boceprevir resistance due to mutation in Hepatitis C NS3/4A protease and the development of new NS3/4A protease inhibitor"**, submitted by **NEHA NAGPAL (2K12/BIO/017)** in partial fulfillment of the requirement for the award of the degree of Master of Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by her under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honoring of any other degree.

Date:

Prof. B. D. Malhotra
Department of Bio-Technology
Delhi Technological University
(Formerly Delhi College of Engineering, University of Delhi)

DECLARATION

I, **Neha Nagpal** hereby declare that the M. Tech. dissertation entitled **"Computational analysis of the molecular principle behind Boceprevir resistance due to mutation in Hepatitis C NS3/4A protease and the development of new NS3/4A protease inhibitor"**, submitted in partial fulfillment of the requirement for the award of the degree of Master of Technology in Bioinformatics, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is a record of original and independent research work done by me under the supervision and guidance of **Prof. B. D. Malhotra**, Professor, Department of Biotechnology, Delhi Technological University, Delhi. The information and data enclosed in this dissertation is original and has not formed the basis of the award of any Degree/Diploma/Associateship/Fellowship or other similar title to any candidate of any university/institution.

Date:

Neha Nagpal
M. Tech Bioinformatics
Department of Biotechnology
Delhi Technological University
ShahbadDaulatpur,
Main Bawana Road,
Delhi -42, India

ACKNOWLEDGEMENT

I would like to acknowledge the people and the institute who have given us the knowledge and light of guidance and have helped us to wade through the darkness. I express my deep gratitude to my guide Prof. B. D. MALHOTRA. I consider myself fortunate and highly indebted to him for his precious guidance and invaluable suggestions. I am thankful to him for his exceedingly helpful nature. Sincere thanks to all the faculty members of Biotechnology Department, DTU, for their kind help and cooperation.

I express my sincere gratitude to Dr. ABHINAV GROVER, for his stimulating guidance, continuous encouragement and supervision throughout the course of present work.

Parents are always a perpetual source of inspiration and encouragement. This work would not have been possible, had it not been for them.

NEHA NAGPAL
2K12/BIO/017

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Meaning
HCV	Hepatitis C Virus
SVR	Sustained Viral Response
DAA	Direct- Acting Antiviral
PIs	Protease Inhibitors
SPRINT	Serine Protease Inhibitory Therapy
MD	Molecular Dynamics
FED	Free Energy Perturbation
HTVS	High-Throughput Virtual Screening
SP	Standard Precision
XP	Extra Precision
SASA	Solvent Accessibility Surface Area
VMD	Visual Molecular Dynamics
RMSD	Root Mean Square Deviation
RMSF	Root-mean-square positional fluctuation
ROG	Radius Of Gyration
PDB	Protein Data Bank
OPLS	Optimized Potential for Liquid Simulations
ESBRI	Evaluating the Salt Bridges in Proteins

Computational analysis of the molecular principle behind Boceprevir resistance due to mutation in Hepatitis C NS3/4A protease and the development of new NS3/4A protease inhibitor

Neha Nagpal

Delhi Technological University, Delhi, India

1. ABSTRACT

The hepatitis C virus (HCV) Infection is a primary cause of chronic hepatitis that eventually progresses to cirrhosis and in some instances might advance to Hepatocellular carcinoma. According to the WHO report, HCV infects 130–150 millions people globally and every year 350000 to 500000 people die from hepatitis C virus infection. Great achievement has been made in viral treatment evolution, after the development of HCV NS3/4A protease inhibitors (Boceprevir). However, efficacy of boceprevir is compromised by the emergence of drug resistant variants. The molecular principle behind drug resistance of the protease mutants such as (V36M, T54S and R155K) is still poorly understood. Therefore in this study, we have employed series of computational strategies such as molecular dynamics simulation, solvent accessibility variation, the salt bridge interactions analysis, substrate envelope analysis and binding pocket analysis to analyze the binding of antiviral drug, boceprevir to HCV NS3/4A protease mutants.

Our results clearly demonstrate that the point mutation (V36M, T54S and R155K) in protease is associated with lowering of binding affinity between of binding between boceprevir and protease. Exhaustive analysis of the simulated boceprevir bound wild and mutant complexes revealed variations in hydrophobic interactions, hydrogen bond occupancy and salt bridge interactions. Also, substrate envelope analysis scrutinized that the studied mutations resides outside the substrate envelope which may affect the boceprevir affinity towards HCV protease but not the protease enzymatic activity. Furthermore, structural analyses of the binding site volume and flexibility show impairment in flexibility and stability of the binding site residues in mutant structures. In the last of the report, we also identify the common inhibitor of the wild and mutant proteases by parallel virtual screening through combinatorial libraries which are constructed based on boceprevir scaffold.

In order to combat boceprevir resistance, renovation of binding interaction between the Drug and protease may be valuable. The structural insight from this study reveals the mechanism of the boceprevir resistance and the results can be valuable for the design of new PIs with improved efficiency.

2. INTRODUCTION

Infection by Hepatitis C virus (HCV) is now considered as most severe and chronic blood borne infection. It is reported that HCV infects 130–150 million people globally. WHO has estimated that every year 350000 to 500000 people die from hepatitis C virus infection(WHO, 2014). HCV NS3/4A protease belonging to the family of trypsin-like serine protease is involved in proteolytic cleavage of the HCV polyprotein precursor. It is a Heterodimer formed in complex with the 181-residue N-terminal protease domain of the NS3 protein and 54-residue small NS4A peptide cofactor which subsequently leads to proteolytic cleavage at four junctions of HCV polyprotein such as NS3/4A, NS4B/5A, NS5A/B and NS4A/4B (Bartenschlager, 1999; Halfon *et al.,* 2011; Kolykhalov *et al.,* 2000; Kwong *et al.,* 2008).

For the last two decades, NS3/4A protease has emerged as a prime target for anti- Hepatitis C viral drug design and discovery. Till now, by the use of structure -based drug design (SBDD), a number of NS3/4A protease inhibitors (PIs) have been designed. NS3/4A protease inhibitors (PIs) have been divided into two major classes such as first and second generation PIs. First generation PIs includes boceprevir and telaprevir, these inhibitors are the linear α-ketoamide derivatives which bind covalently with the active site of the enzyme in the reversible manner. Boceprevir is a ketoamide HCV NS3 linear protease inhibitor developed by Merck & Co and approved by the FDA on May 13, 2011.

Although Boceprevir has shown promising results in a treatment of chronic hepatitis C genotype 1 infection, point mutations in the active site and non-active site region of NS3/4A protease has led to a major challenge of drug resistance. Several point mutations including V36M, T54S, R155K, A156S/V/T, have been reported to reduce the drug sensitivity (Richard , 2011; Susser *et al.,* 2009; Tong *et al.,* 2006; Victrelis (boceprevir), 2011). In order to overcome the drug resistance due to these mutations, a detailed atomic level study about the molecular principle of drug resistance against NS3/4A PIs will be beneficial.

Several Molecular dynamics studies have been successfully implemented to explore the molecular mechanism of resistance and influence of resistance on the conformational features of NS3/4A protease. Previous molecular dynamics and free energy perturbation (FEP) simulation studies on variants for various NS3/4A protease inhibitors have unveiled the molecular basis of resistance and explained the influence of mutation on the interaction of drug with protease residues(Guo *et al.,* 2006; Welsch *et al.,* 2008; Welsch *et al.,* 2012 ; Pan et al., 2012 ; Xue *et al.,* 2013).

However, the detailed molecular basis of conformational changes caused due to (V36M, T54S and R155K) NS3/4A protease mutation on the binding of boceprevir is still poorly understood. In the current study, we have employed a series of computational modeling techniques such all-

atom molecular dynamics simulations, solvent accessibility variation ,Substrate envelope analysis , Binding pocket analysis and Salt bridge interactions analysis on boceprevir bound wild type NS3/4A protease and three mutants, V36M, T54S and R155K . An attempt was made to depict the structural variation, functional changes and binding mode associated with mutations. After revealing the mechanism of bocepervir resistance, we aim to redesign the bocepervir PIs for expanding its spectrum against both wild and mutant protease. We generated a diverse combinatorial library based on a boceprevir scaffold and common inhibitor of the wild and mutant proteases were identified by parallel virtual screening through constructed libraries.

3. REVIEW OF LITERATURE

3.1 Hepatitis C Virus (HCV) Infection

The hepatitis C virus (HCV) infection , discovered in 1989(Choo *et al.,* 1989) is a primary cause of chronic hepatitis which eventually progresses to cirrhosis and in some instances might advance to hepatocellular carcinoma. According to the WHO report, HCV infects 130–150 million people globally and every year 350000 to 500000 people die from hepatitis C virus infection. Hepatitis C virus infection burden is worldwide. HCV prevalence are higher(>3.5%) in region such as North Africa/Middle East, East and Central Asia where as Sub-Saharan Africa, South and Southeast Asia, Central, Southern and Andean Latin America, Oceania, Caribbean, Australasia, and Eastern, Western and Central Europe are believed to have moderate infection rate ranging from 1.5%-3.5%. Tropical Latin America, North America and Asia Pacific have low prevalence (<1.5%)(Mohd *et al.,* 2013).

Most of the acute hepatitis C infected individuals are not diagnosed and remain unaware about their infection until the advanced hepatitis stage such as cirrhosis, hepatocellular carcinoma and liver transplantation. It has been reported that during the acute phase of the disease about 70%-90% of infected people unable to destroy the virus and become chronic carriers.

3.2 The HCV Virus and Genome

The hepatitis C virus (HCV) is an enveloped positive-single stranded RNA virus in *Flaviviridae* family. HCV have been classified into six major genotype (Genotype 1–6), which varies in their treatment response and worldwide prevalence. Each Genotype is further subcategorized into many strain and several subtypes (a, b, c, etc). Genotypes 1(Type 1a and1 b) virus predominately, contribute for the majority of HCV infection in Japan, South America, North America and Europe. Comparatively infection rate is lower for genotype2 than for genotype 1. Genotype 3 virus is primarily prevalent in south-east Asia. Genotype 4 viruses prevail in the central Africa, Egypt and Middle East. Genotype 5 viruses report for major infections in South Africa and genotype 6 viruses are majorly widespread to Asia (Mohd *et al.,* 2013; Simmonds *et al.,* 2005; Simmonds *et al.,* 1993).

3.3 Polyprotein processing

Genome of the HCV virus contains single open reading frame whose translation is directed by internal ribosome entry site (IRES) residing in 5' un-translated region. ORF translated product is an polyprotein of 3000 amino acids , which is divided into non-structural region (NS2–NS5B proteins) and structural region (C–p7 proteins) required for replication(Lindenbach *et al.,* 2005; Reed *etal.,* 1998) (Figure 1).

The structural and nonstructural proteins are formed by post translational cleavage by the host and viral proteases respectively. Structural proteins are consists of core protein (C) which forms nucleocapsid of the virion, the envelope glycoproteins (E1 and E2), participates in receptor binding required for viral entry into the hepatocyte and p7 protein which functions as an ion channel (Griffin *et al.,* 2003; Pavlovic *et al.,* 2003).

The HCV NS3 protein is a multifunctional enzyme where N terminal region (180 amino acid residues) folds into protease, responsible for cleavage activity at four junctions of HCV polyprotein such as NS3/4A, NS4B/5A, NS5A/B and NS4A/4B. In the two-third C-terminal region of NS3 protein, RNA helicase is resident, which is responsible for RNA folding/remodeling process (Bartenschlager *et al.,* 1999; 1993).

NS4 is a vital cofactor for NS3 proteolytic cleavage activity, which complex with NS3 and enhance polyprotein processing. NS5B is the viral RNA-dependent RNA polymerase. NS5A, phosphoprotein, functions in viral assembly and interact with 3-NTR of the HCV genome, numerous cellular proteins and other non-structural proteins.

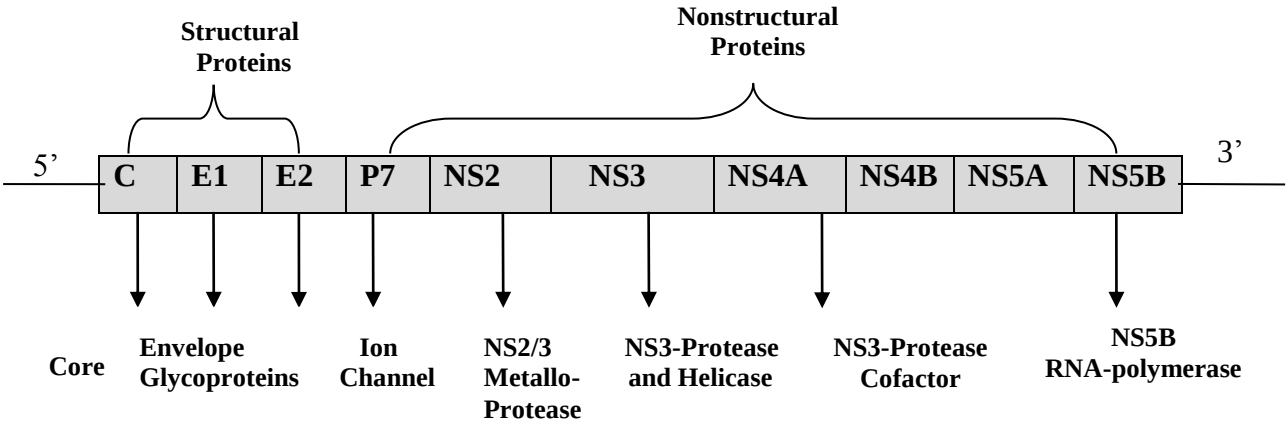

Figure 1: Schematic representation of HCV genome

3.4 HCV NS3/4A protease

HCV NS3/4A protease belongs to the family of trypsin-like serine protease is involved in proteolytic cleavage of the HCV polyprotein precursor. It is a Heterodimer formed in complex with the 181-residue N-terminal protease domain of the NS3 protein and 54-residue small NS4A peptide cofactor which subsequently leads to proteolytic cleavage at four junctions of HCV polyprotein such as NS3/4A, NS4B/5A, NS5A/B and NS4A/4B. The crystal structure of the NS3 protease consists of two β-barrel domains and two short α-helices (Figure 2). The Protease catalytic site, located between the two β-barrels, is featureless, solvent-exposed and shallow. Protease catalytic activity requires a catalytic triad including Ser-139, His-57, and Asp-81and an

oxyanion hole (Gly-137), which are located at the crevice of the two domains (Kim *et al.*, 1998;Kwong *et al.*, 1998; Love *et al.*,1996; Yan *et al.*, 1998).

NS4A cofactor, facilities the catalytic efficiency, substrate specificity and stability of the enzyme. Upon binding with the NS4A, a proper conformation of the substrate and catalytic triad is attained which enables hydrogen bond to Asp-81 and deprotonation of the Ser-139 by the His-57. Thus, NS4A is critically required for efficient catalytic turnover and its absence may diminishe the enzymatic activity (Failla *et al.*, 1995; Lin *et al.*, 1995). Hence, NS3/4A is structurally more relevant target for functional studies and drug discovery research.

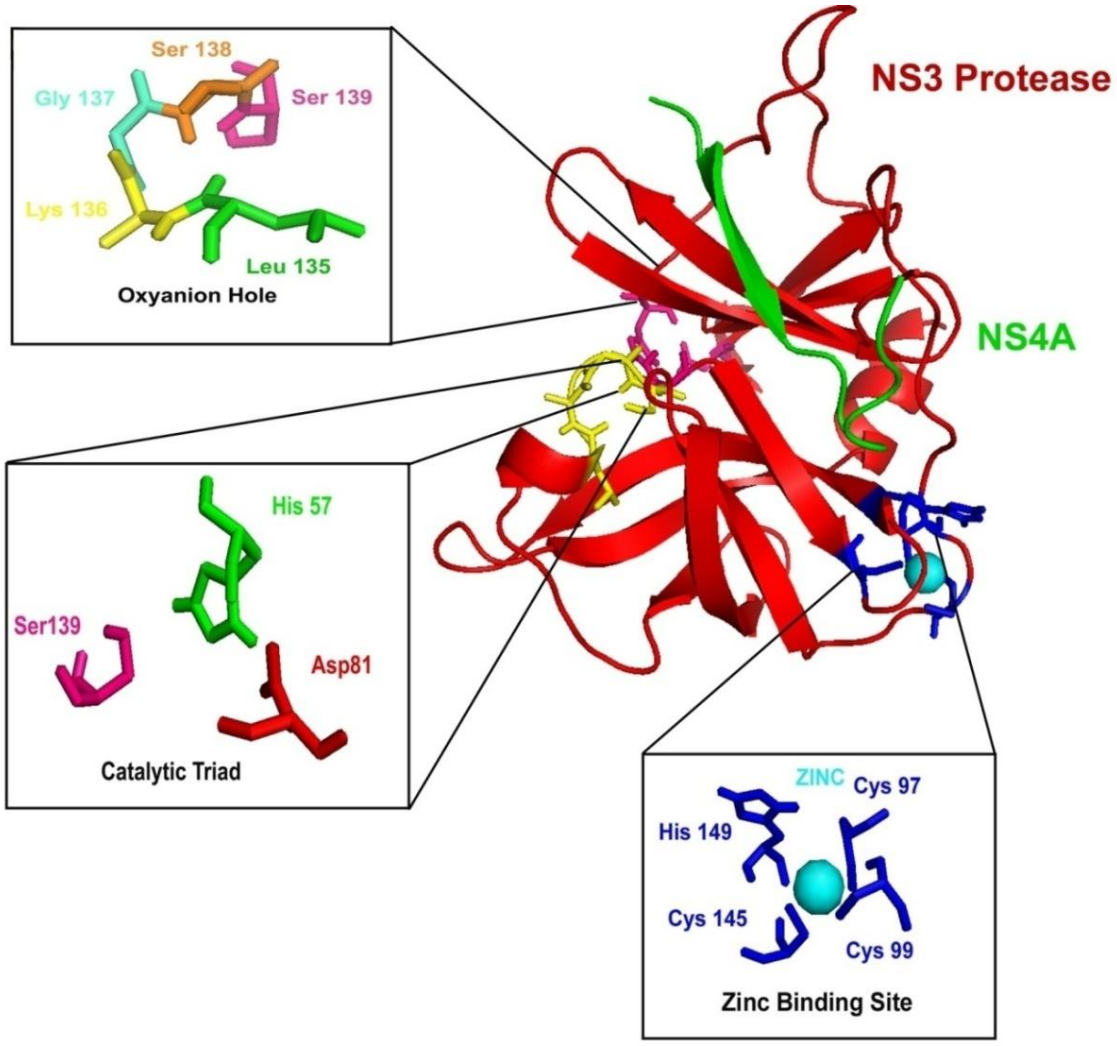

Figure 2: The Crystal structure of NS3 protease in complex with the NS4A cofactor

3.5 NS3/4A protease inhibitors

Anti-HCV direct- acting antiviral therapy (DAA) can be classified into several categories: (i) HCV NS3-4A serine protease inhibitors, (ii) HCV NS3 NTPase/helicase inhibitors, (iii) HCV NS5B polymerase inhibitors, (iv) HCV NS5A inhibitors and others.

Because of dual activities of NS3 protein, necessary for virus replication, there is huge boom in development of direct-acting antivirals (DAAs) therapies targeting NS3/4A protease. HCV NS3/4A protease inhibitors can be classified into two categories (i) first and (ii) second generation PIs. First generation PIs includes boceprevir and telaprevir which were the first PIs approved by Food and Drug Administration in 2011 and are designed specifically against HCV genotype 1, contributing towards 60% of global infections . These inhibitors are the linear α-ketoamide derivatives which bind covalently with the active site of the enzyme in reversible manner (Table 1). The second generation PIs includes macrocyclic compounds which are in clinical trials such as MK-517214 in phase II clinical trials(Harper *et al.,* 2012) and TMC43511, danoprevir and vaniprevir in phase III clinical trials(Lin *et al.,* 2009; Liverton *et al.,* 2010; Seiwert *et al.,* 2008).

Table 1: Current HCV NS3-4A protease inhibitors and status

Inhibitor name	Genotypic coverage	Company	Status
Reversible covalent inhibitor			
Incivek (telaprevir, VX-950)	1	Vertex	Approved
Victrelis (boceprevir, SCH503034)	1	Merck	Approved
Noncovalent inhibitor			
ABT-450/r	1	Abbott	Phase III
Simeprevir (TMC435)	1, 2, 5, and 6	Janssen	Phase III
Faldaprevir (BI201335)	1	Boehringer Ingelheim	Phase III
Danoprevir (RG7227)	1	Genentech	Phase II
Vaniprevir (MK-7009)	1	Merck	Phase II
MK-5172	1, 2	Merck	Phase II
Asunaprevir (BMS-650032)	1, 4	Bristol-Myers Squibb	Phase II
ACH-1625	1	Achillion	Phase II
GS-9256	1	Gilead	Phase II
ACH-2684	1, 3	Achillion	Phase II
GS-9451	1a, 1b	Gilead	Phase II
Narlaprevir/r	1	Merck	Phase II
IDX320	1, 1b, 3a, and 4a	Idenix	Phase II

3.6 Discovery of Boceprevir

Boceprevir is in a class of HCV NS3/4A protease inhibitor, developed by Schering-Plough but is now being developed by Merck . It was approved by the Food and Drug Administration on May 13, 2011 for the treatment of chronic hepatitis C (CHC) genotype 1 infection, used in combination with other medications ribavirin and peginterferon alfa. Boceprevir molecular formula is C27H45N5O5 and its molecular weight is 519.7. It has the following chemical name: (1R,5S)-N-[3-Amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3-[2(S)[[[(1,1dimethylethyl)amino] carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6,6-dimethyl-3azabicyclo[3.1.0]hexan-2(S) carboxamide. Boceprevir has the following structural formula (Figure 3).

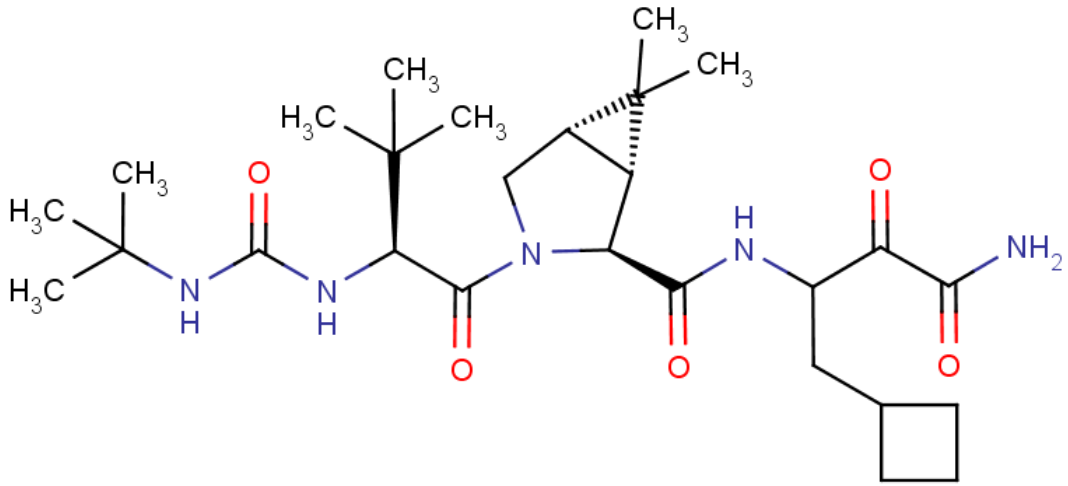

Figure 3: Structure of HCV NS3/4A protease inhibitor, Boceprevir

Boceprevir have nearly doubled the rates of sustained viral response (SVR) among treatment-experienced as well as treatment-naïve patients. Boceprevir in addition to the PEG-IFN and RBV has shown a promising improvement in SVR rate in both SPRINT-1 (serine protease inhibitory therapy) and RESPOND trials

3.7 Resistance to Boceprevir

Despite the improved efficacy of combination therapy in antiviral treatment, development of drug-resistant variants is still a concern for the long term antiviral therapies. Numerous Mutations of NS3 protease have been identified to associate with resistance in clinical studies as well as cell culture. Many of these mutations are located close to the NS3 protease active site including V36M/A/L, T54S/A, R155K/T/S/M, A156S (associated with low- to medium-level resistance), and A156Y/T (high-level resistance)(Kieffer *et al.*, 2007; Kwo *et al.*, 2010; Sarrazin *et al.*, 2007; Susser *et al.*, 2009). After SPRINT-2 and RESPOND-2 trails, Boceprevir emergent substitutions such as V36M, T54S, R155K and A156S/V/T were detected in subjects Who Did Not Achieve a (SVR) Sustained Virologic Response. T54S and R155K are the most common

mutations which were detected in the long term follow up of patients up to 2.5 years upon boceprevir treatment. Drug resistance occurs when amino acid substitution disrupts the inhibitor and protease binding while preserving substrate recognition and target function. The comparison of structural studies of boceprevir binding between wild and mutant HCV genotype 1 protease will provide a mechanistic insight to inhibitor modification to combat the drug resistance.

3.8 Molecular dynamics simulations for Bio-molecular systems

To better understand dynamic nature of proteins it is imperative to create models which describe the motions of proteins. The most widely used method for studying the interaction's among large atoms is Molecular dynamics (MD). MD is a method to simulate proteins which describes the molecular motion of system particles. In *In-silico* studies, it helps to understand the dynamic behavior of macromolecules including lipids, carbohydrates, proteins and their complex in non-aqueous (membrane) and aqueous environment.

Other applications includes refining macromolecules, exploring native conformational spaces based on experimental data, calculating free energy changes by mutations and other perturbations. MD when coupled with tools like docking virtual screening will help to address many problems concerning situations where dynamic nature of protein cannot be ignored. The situation includes ligand induced conformational changes of active site and mechanisms of highly mobile membrane proteins.

In MD simulations all atoms are assigned coordinates charges and velocities. The charges and position are used to calculate potential of system. The calculated potential helps to infer the force experienced by each atoms of simulation system. By applying Newton's law of motion for short time step, a new position and velocities is calculated for each atom. In the simulation process atoms are modeled as points of space with a given charge and mass. The calculated charges are further used in calculating electrostatic force field which is then used to find new positions of each atom using classical mechanics. The above process is repeated till the system gets the final configuration. Thus this approach helps to gather information about not only conformations of system but also the dynamic behavior of proteins.

3.9 Desmond

Desmond is a software package developed at D.E Shaw research to investigate MD simulations of biological systems. The software includes numerical techniques and parallel algorithm to accomplish high performance accuracy on multi processer systems and can also be performed on single computer. The software package also calculates energies and forces for fixed charge force fields in bimolecular simulations. It calculates both long and short range electrostatic interactions and forces. Long range interactions are calculated using Particle-Mesh-Ewald technique. The

time step is increased gradually by applying constraints which are enforced using variant of SHAKE algorithm. Force field parameters can be assigned using parameter adjustment tool. Desmond is integrated with molecular modeling environment for studying simulations of chemical and biological environment. It is also compatible with other widely used tools for trajectory viewing and analysis. Various parameters analyzed during MD simulations include RMSD, RMSF and many others. RMSD is the root mean square deviations, which calculates the standard deviation of backbone protein with respect to initial structures. RMSF value represents the standard deviation from the average atomic position over the entire MD simulation.

3.10 Molecular dynamics studies in drug resistance

Several Molecular dynamics studies have been successfully implemented to explore the molecular mechanism of resistance and influence of resistance on the conformational features of NS3/4A protease. Previous molecular dynamics studies on V55 as resistance variant for boceprevir has concluded that flexibility of protein residue might be vital for NS3 domain-domain interactions. Welsch et al also showed that on boceprevir binding to the mutant proteases, there is a loss of H-bond interactions and constriction in binding pockets(Welsch *et al.,* 2012). In 2008, By using protein-ligand docking and non-covalent residue interactions, they explained the binding mode of VX- 950 and elucidate the resistance mechanism of VX- 950 due to V36 and T54 mutations(Welsch *et al.,* 2008). Guo et al reported the free energy perturbation (FEP) simulation for A156T and D168V/Q mutant in PIs (SCH 503034, BILN 2061 and VX-950) and presented that T54 and V36 disturbs the interaction of cyclopropyl group of the VX-950 with protease residues(Guo *et al.,* 2006).

4. METHODOLOGY

4.1 Initial structures preparation

The X-RAY crystallographic structure of the NS3/4A protease bound with drug boceprevir (PDB 2OC8) was obtained from the RCSB Protein Data Bank (Prongay et al., 2007). Because of the unavailability of the mutant structures, the variants were generated from the wild structure by substituting the corresponding amino acids with the GILDE package (Friesner *et al.,* 2004). The wild and mutant unbound NS3/4A protease with empty binding site was analyzed by removing boceprevir from the wild bounded PDB structure 2OC8 and generated boceprevir bound mutant protease .The three generated mutant (V36M, T54S and R155K) and wild structure of NS3/4A protease/boceprevir complex and other three generated unbound model for wild and mutant structure of NS3/4A protease were used as a initial model for the following Molecular dynamic simulation.

Before the simulations studies, OPLS 2005 molecular mechanics force field is applied for describing the parameters for both boceprevir and wild and mutant NS3/4A protease structures. In Desmond, Protein preparation wizard was used to pre-process the structures and to fix erroneous atomic representations in crystal structure. Preprocessing includes various steps such as addition and optimization of hydrogen atoms, generation of disulphide bonds, removal of water molecules and capping of terminals. The next step involves the solvation of system in a triclinic periodic box of SPC water molecules and 19 Cl^- counter ions were added to the system for neutralization. In order to eliminate geometric problems such as unrealistic bond distances, bond angles and torsion angles, energy minimization is performed.

4.2 Molecular dynamics simulation

All the simulation studies were carried out by using the Desmond module of Schrodinger's Suite using the Maestro interface (Pikkemaat *et al.,* 2002; Shaw, 2009). Initially, steepest descent method is used for the system minimization with restraints until a gradient threshold (25kcal/mol/Å) reached. After the minimization step, Molecular dynamics simulations were performed on equilibrated systems for a time-step of 2 femtoseconds at a constant atmospheric pressure and constant temperature of 300 K. Bound as well as unbound wild and mutant structures MD simulation were carried out for 20 ns. In order to compute the stability and dynamic properties of the protein structure, trajectory files were analyzed through simulation event analysis wizard of the Desmond module. Difference in the structural stability and dynamics nature of the mutants were investigated by comparing protein properties such as RMSF, 3D backbone RMSD, and Radius of gyration (Lobanov *et al.,* 2008) with the wild protein structure. For analyzing protease-drug interactions PDBsum (Laskowski *et al.,* 1997)was

used, where acceptor and donor atoms were defined if H-bond distances are less than 3.3 Å and if acceptor-H-donor angle is greater than 90°.

4.3 Hydrogen bond occupancy

Hbond were calculated on the threshold value of Donor – Acceptor distance as 3 Å and Angle cutoff as 20 degree. Average Occupation of hydrogen bonds between the NS3/4A protease residues and boceprevir were monitored by using Analysis-hydrogen bond Tool in VMD(Humphrey *et al.,* 1996).VMD is a visualization and molecular modelling computer program, consisting of tools for sequence analysis, arbitrary graphics objects and analysis of molecular dynamic simulation results.

4.4 Salt bridge analyses

Salt bridges interactions are known to play principal role in stabilization of protein structure, therefore structural stability of the protein can be easily evaluated by presence of salt bridge interactions in them. VMD package were used for observing all set of salt-bridge pairs in the trajectory(Humphrey *et al.,* 1996). Distance between the salt bridge pair residues were figured by using ESBRI Server (Costantini *et al.,* 2008).ESBRI is web software written in PERL language used for analyzing salt bridges interactions in a protein structure, taking atomic coordinates file as input file and ESBRI output file includes all the salt bridge pair residues and distance between two residues involved in interactions.

4.5 Accessibility of Boceprevir -binding residues

Solvent Accessibility of the ligand-binding residues paves an important way in determining structural stability and flexibility of the protein. Solvent Accessibility of the binding residues can be used for computing the binding interaction between the residues and ligand. WHAT IF web server is utilized for analyzing the solvent accessibility of the binding residues in all wild and mutant NS3/4A protease(Vriend, 1990) .

4.6 Quantification of binding pocket volume

An exact determination of protein shape extracted from protein structure is required for complete understanding of the protein–ligand binding analysis and interactions. Protein structure descriptors such as surface area, volume and depth have been broadly used for investigating and comparing protein- ligand interactions. The Binding pockets volume of the wild-type structure and mutant structure were calculated by using CASTp server(Binkowski *et al.,* 2003), where probe radius were set as default value of 1.4Å

4.7 Substrate envelope analysis

It was discovered that HCV NS3/4A protease active site bound to the substrate with in a conserved consensus shape or volume which is referred as the substrate envelope. According to the Substrate Envelope hypothesis, HCV NS3/4A PIs protrudes from substrate envelope and HCV NS3/4A residues that contribute in inhibitor binding need not participate in substrate binding and that site are the source of mutation causing drug resistance.
Here in , we use substrate-envelope hypothesis to understand the drug resistance mechanism of known mutation sites such as R155K, V36M and T54S and that may help in designing the inhibitors that does not protrudes outside the substrate envelope and mimics the exact volume of the NS3/4 A substrate. We superimpose the 4B5A structure on to the active site of the boceprevir bound wild and mutant (V36M, T54S and R155K) protease by using PyMOL(Schrodinger, 2010). The simulated wild-type and mutant-boceprevir complexes were used as the reference structure for each alignment. The NS3/4A viral substrate envelope was computed using the full-length NS3/ 4A structure and product complexes 4A4B (3M5M).

4.8 Construction of screening combinatorial library

Combinatorial library was generated using the Lead grow module of VLife MDS (VLife, 2012)by substituting various chemical groups at the substitution site R1, R2, R3 site of the boceprevir scaffold.

4.9 Ligand preparation

Library Ligand dataset was prepared prior to virtual screening using the LigPrep tool of Schrödinger suite (Schrödinger, 2009) for the generation of 3-dimensional optimized molecules for which optimized potential for liquid simulations 2005 (OPLS) force field was used.

4.10 Virtual screening

Virtual screening, which has emerged as an effective filtering tool for drug designing process was used to carry out a search for novel inhibitors. Structure based virtual screening is used to compute the binding affinities and binding modes of the ligand -receptor complex. Two parallel virtual screening through the constructed library were carried out to identify the common inhibitors for the wild and mutant proteases.

4.10.1 Protein preparation

For the preparation of protein complexes, Schrödinger's protein preparation wizard is utilized which includes pre-processing, minimization and optimization of the protein structure (Sreeramulu *et al.*, 2009). Protein structures obtained by X-ray crystallography or NMR might possess some irregularities and were rectified by fixing bond orders, ionization states, steric

clashes and side chain orientations. Protein preparation includes various steps such as bad contact removal, bond length optimization, disulfide bond creation, protein terminal capping and conversion of selenomethionine to methionine. The energy minimization was performed with the OPLS force field

4.10.2 Grid generation

The cubic grid for docking calculations was generated for active site of the protease by using receptor grid generation panel. The active site amino acid residues such as His 57, Asp 81, Ser139, Ala 157 and Gly 137 that bind to boceprevir were taken into account to create the grid of dimension $10\mathring{A} \times 10\mathring{A} \times 10\mathring{A}$.

4.10.3 Docking studies

For the ligand identification, virtual screening was performed by two stage docking technique i.e. high-throughput virtual screening (HTVS) and extra precision (XP) using Glide virtual screening workflow (VSW) (Friesner *et al.*, 2004). Glide provides a rational workflow for virtual screening from HTVS to XP, enriching the data at every level such that only a few compounds need to be studied at the next higher accuracy level. Among two stages, HTVS is used for large set of ligands and XP docking is used for more accuracy. Glide docking was performed on the generated libraries and each ligand was first subjected to the high throughput virtual screening (HTVS) module of GLIDE. All the Glide docking studies were performed on Intel Core 2 Duo CPU @ 3 GHz of HP origin with 1 GB DDR RAM. All the MD simulation studies were performed in Core 2 Duo CPU E6850 @ 3 GHz of HP origin, with 3 GB DDR RAM. HTVS dock score threshold was set to -7Kcal/mol for compound screening. To obtain higher precision docking, screened compounds above the threshold values were further analysed by the XP protocol. Glide score (G score) is computed based on an empirical scoring function which combines the following parameters.

Gscore = a* vdW + b* Coul + Lipo + Hbond + Metal + BuryP + RotB + Site

where, vdW is van der Vaals force, Lipo is lipophilic contact term, Hbond is hydrogen-bond, Metal is metal-binding term, BuryP is penalty for buried polar groups, RotB is penalty for freezing rotatable bonds, Coul is Coulomb forces, Site is polar interactions at the binding site and a, b representing coefficients of vdW and Coul are set as: a = 0.065; and b = 0.130. Stability of few of the top scoring docked complexes was also inspected through MD simulation

5. RESULTS

5.1 Molecular dynamics simulation for flexibility and stability of boceprevir bound wild and mutant NS/3A protease

Root-Mean-Square Displacement of each wild and mutant protein – boceprevir complexes are illustrated in Figure 4, have showed little difference. After 4080ps, mutant V36M attained highest RMSD of 3.1 Å where as mutant T54S exhibited highest RMSD of 3.34 Å at 11913.6 ps. Wild protein – boceprevir complex showed sudden fall in RMSD of 1.60 Å at 10056ps and at this simulation time it exhibits the lower RMSD values then the mutants. From 12ns to the completion of the simulation, all boceprevir bound wild and mutants (V36M, T54S and R155K) complexes attained a similar level of deviation. According to the RMSD analysis, V36M, T54S and R155K mutant structures showed relative more fluctuations then wild structure. Mutant T54S has showed largest fluctuation then others. After 20ns long simulation, all the structures were found stable with average RMSD fluctuation within 1.0 Å. Hence, last 5ns stable trajectories were used for further energetic and structural analysis.

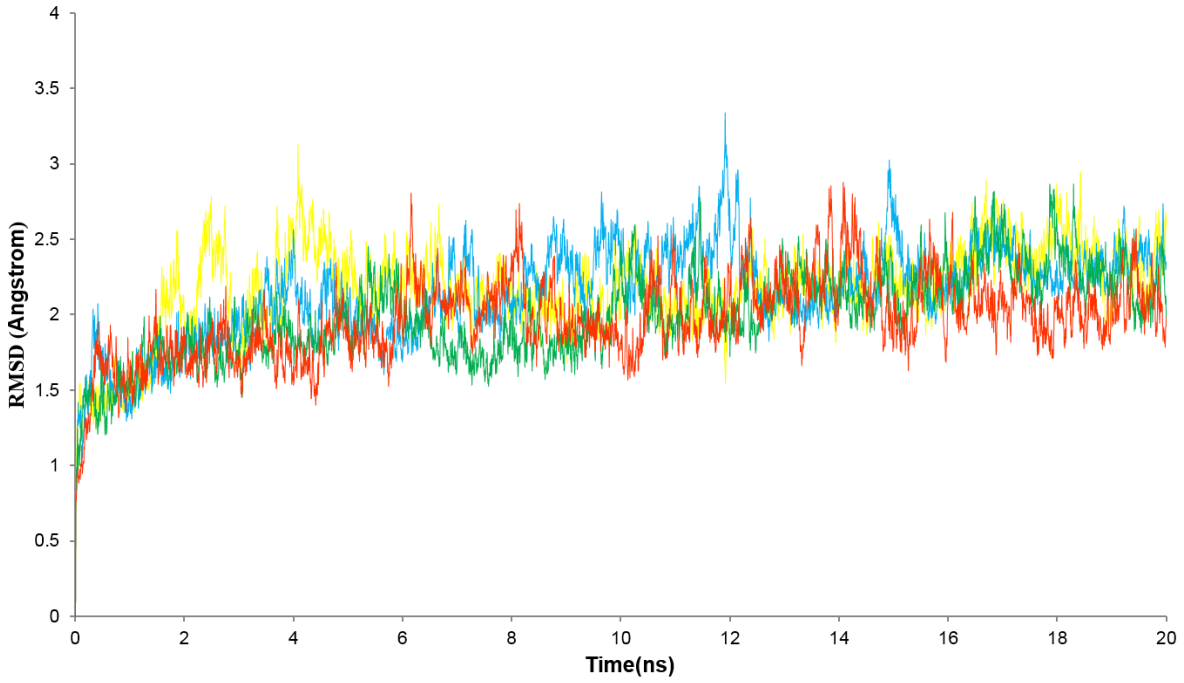

Figure 4: Graphical illustration of the RMSD trajectory of NS3/4 A protease complex with boceprevir backbone of each system (Red, Wild; Green, R155K; Blue, T54S; Yellow, V36M)

In order to investigate the effect of mutation on dynamic behavior and stability of residues, the root-mean-square fluctuation (RMSF) values of each wild and mutant NS3/4 protease system were plotted in Figure 5. In all the structures, wild as well as mutant, the region between 20 – 30 aa of the NS3/4 protease exhibits high flexible and fluctuating behavior. After the comparison of

all the RMSFs, it can be deciphered that higher degree of fluctuation was present in wild than mutant structures of NS3/4 protease.

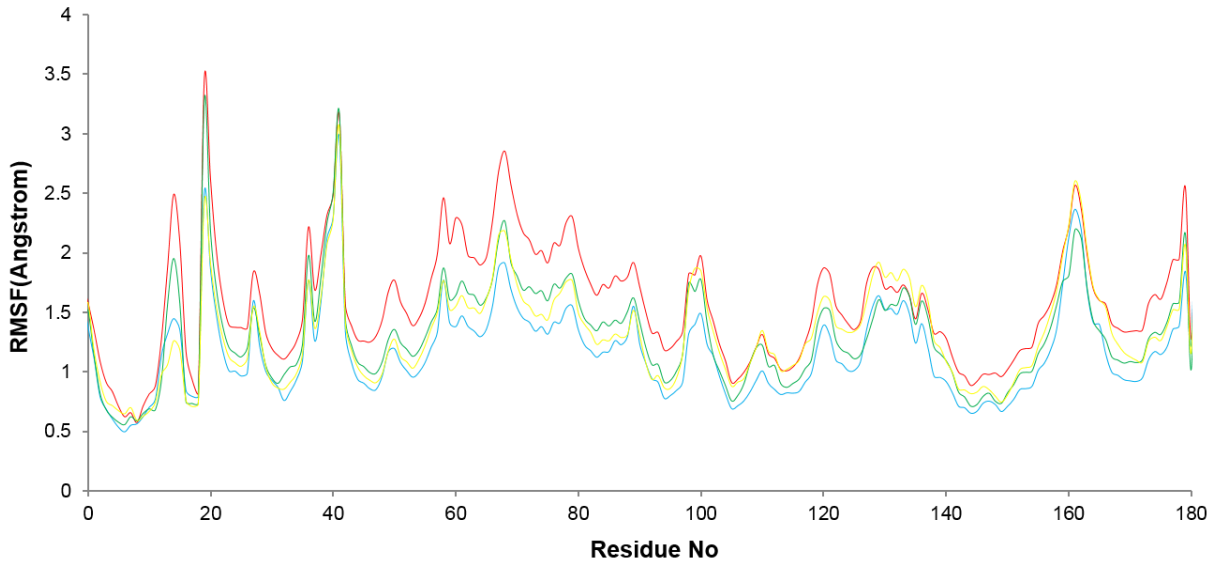

Figure 5: Graphical representation of RMSF values for all wild and mutant structure residues over time (in nanoseconds). (Red, Wild; Green, R155K; Blue, T54S; Yellow, V36M)

We studied the flexibility of amino acids involved in drug binding pockets(His 57 , Asp 81 , Ser139, Ala 157 , Gly 137) in which mutants exhibits decreased flexibility as compared to the wild (Table 2). In Figure 2, the region of 40 to 160 amino acids, wild protease contained more flexibility than mutant (V36M, T54S and R155K) .Binding residues in the mutant T45S and wild structure showed minimum and maximum RMSF value respectively whereas mutant R155K and V36M showed intermediate values. Thus this evidently explain the significance of flexibility in boceprevir binding and rigidity of binding residue in mutant structure decrease the binding affinity between the boceprevir and NS3/4 protease.

Table 2: Root-mean-square fluctuation (in Å) of boceprevir -binding residues of wild and mutant protease

Boceprevir-binding residues' Root mean square fluctuation (in Å)						
Protein structure	Thr 42	His 57	Asp 81	Gly 137	Ser 139	Ala 157
Wild–NS3/4A Protease	1.56	1.99	1.86	1.52	1.34	1.57
R155K–NS3/4A Protease	1.44	1.53	1.44	1.45	1.15	1.36
T54S–NS3/4A Protease	1.33	1.33	1.22	1.20	0.95	1.07
V36M–NS3/4A Protease	1.4	1.4	1.39	1.60	1.20	1.48

The compactness of the protein structures are represented by calculating radius of gyration of wild and mutant protease. The ROG analysis provided more detailed information about the flexibility of protein structure. ROG graph in Figure 6 indicates that throughout the simulation time, mutant structures show less Rg fluctuation than wild structure. Higher Rg fluctuation of wild protease structure explains higher flexible behavior of wild over mutant protease.

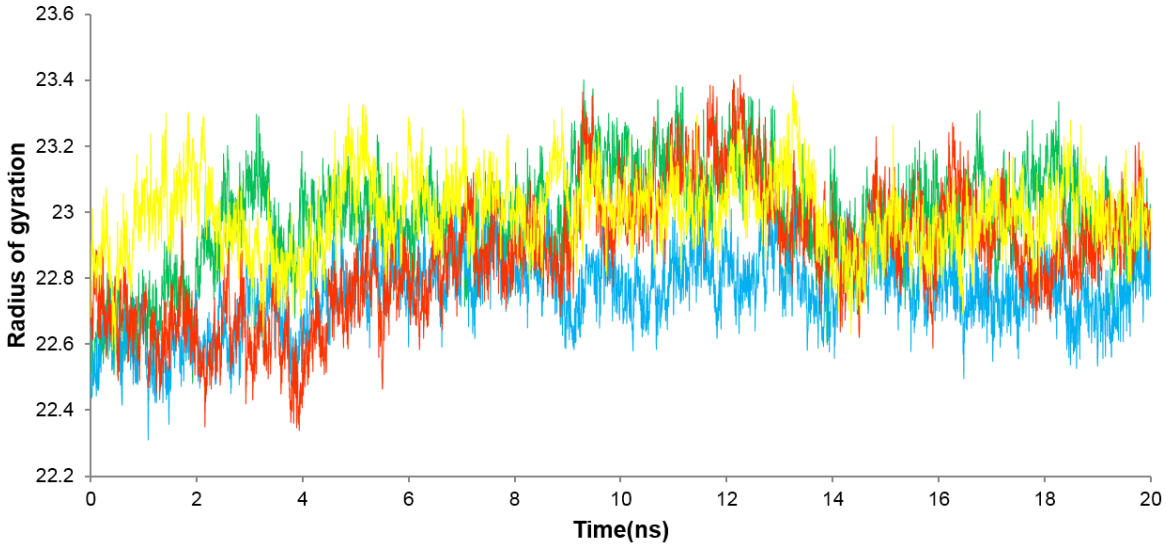

Figure 6: Graphical representation in the change in Radius of Gyration of each system during the 20ns long MD simulations

Solvent accessibility of each residues participated in functional and structural role of wild and mutants of NS3/4A protease were also calculated. According to the analysis, most of the binding residues in wild type NS3/4A protease have higher solvent accessibility than all mutant type

(V36M, T54S and R155K) NS3/4A protease (Table 3). Higher solvent accessibility in wild type shows that binding residues are more exposed for constructive interaction with the boceprevir. However in mutants, lower solvent accessibility value explains the non- availability of binding residues for favorable interactions.

Table 3: Solvent accessibility analysis at boceprevir-binding residues of wild and mutant protease

Solvent accessibility of Boceprevir binding residues (in Å)

Protein structure	Thr 42	His 57	Asp 81	Gly 137	Ala 157
Wild–NS3/4A Protease	12.7	28.4	9.1	4.0	7.9
R155K–NS3/4A Protease	3.4	30.8	7.2	3.3	7.6
T54S–NS3/4A Protease	5.0	28.0	9.0	3.7	10.4
V36M–NS3/4A Protease	5.6	27.9	7.4	4.0	5.2

5.2 Mutational effect on binding of Boceprevir with wild and mutant NS3/4A protease

Hydrogen and hydrophobic interactions have a crucial contribution in specific interactions between protein and ligand. We explored the number of intermolecular hydrophobic interactions and hydrogen bond between the boceprevir bound wild and mutant protease complex, it is depicted in Figure 7, 8(a)–(d).

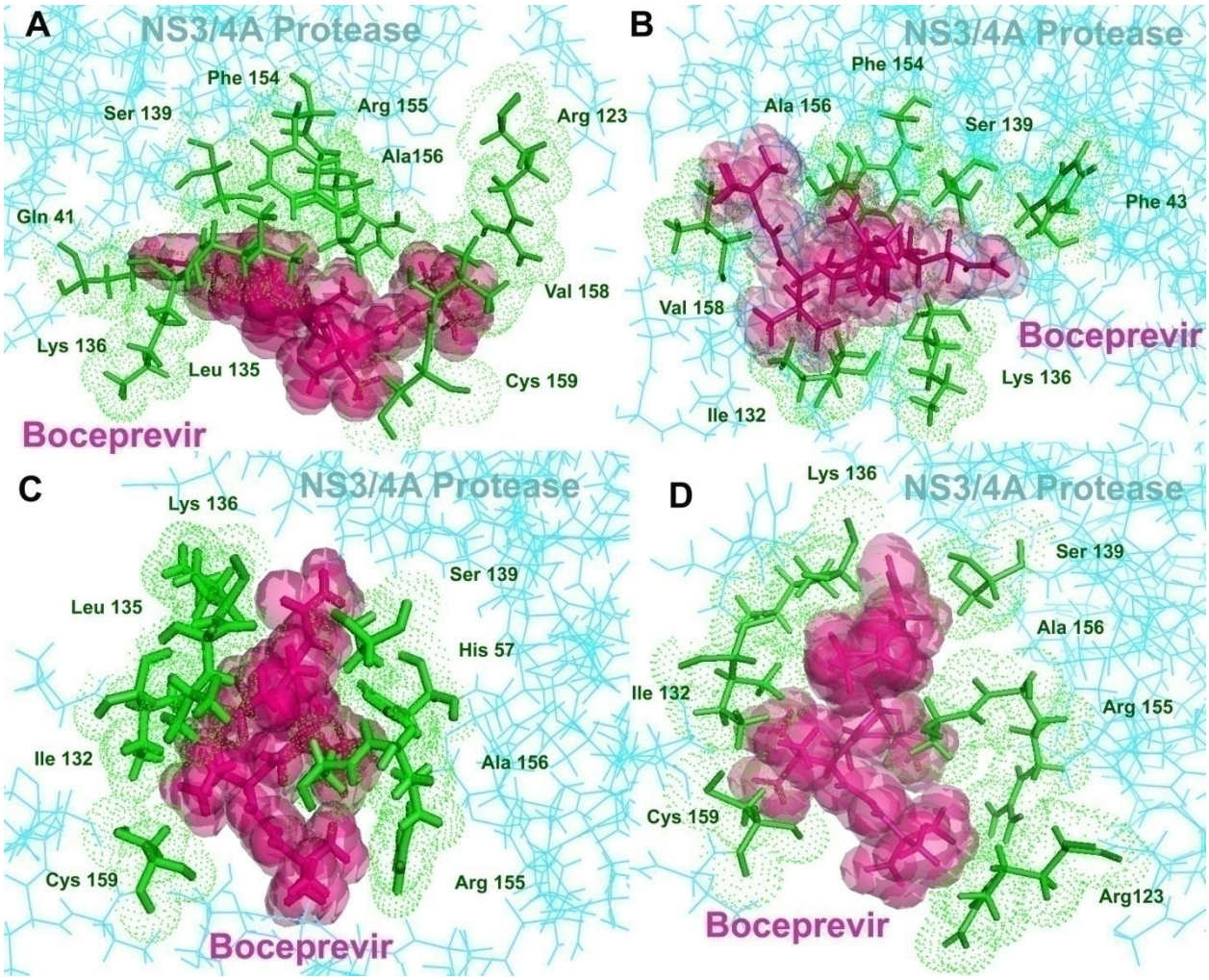

Figure 7: Hydrophobic interactions between bocepervir and NS3/4A protease complex (A) Depiction of Hydrophobic bond interactions with the boceprevir (shown in Hot pink) and wild type NS3/4A protease residues (B) R155K, (C) T54S and (D) V36M mutant type NS3/4 A protease residues (shown in Green)

Previous studies have demonstrated hydrophobic interactions and hydrophobic buried surface as a best parameter for determining binding affinity. The wild protease–boceprevir complex interactions clearly depicts that, there were six hydrogen bond interactions and ten hydrophobic interactions in between the drug boceprevir and NS3/4A protease. Boceprevir interacts with Asp 157, Gly 137, Thr 42 and His 57 via H-bonds and residues involved in hydrophobic interactions

include Arg123, Val 158 , Ala 156 , Cys 159 , Phe 154 , Ser 139 , Lys 136 , Leu 135 , Glu 41 and Arg 155.

In case of mutant protease – boceprevir complexes, there is a significant decrease in the both hydrophobic and hydrogen interactions. V36M and T54S mutant five hydrogen bond and seven four hydrophobic interactions while R155k shows six hydrogen bond and seven four hydrophobic interactions, it is depicted in Figure 7,8.

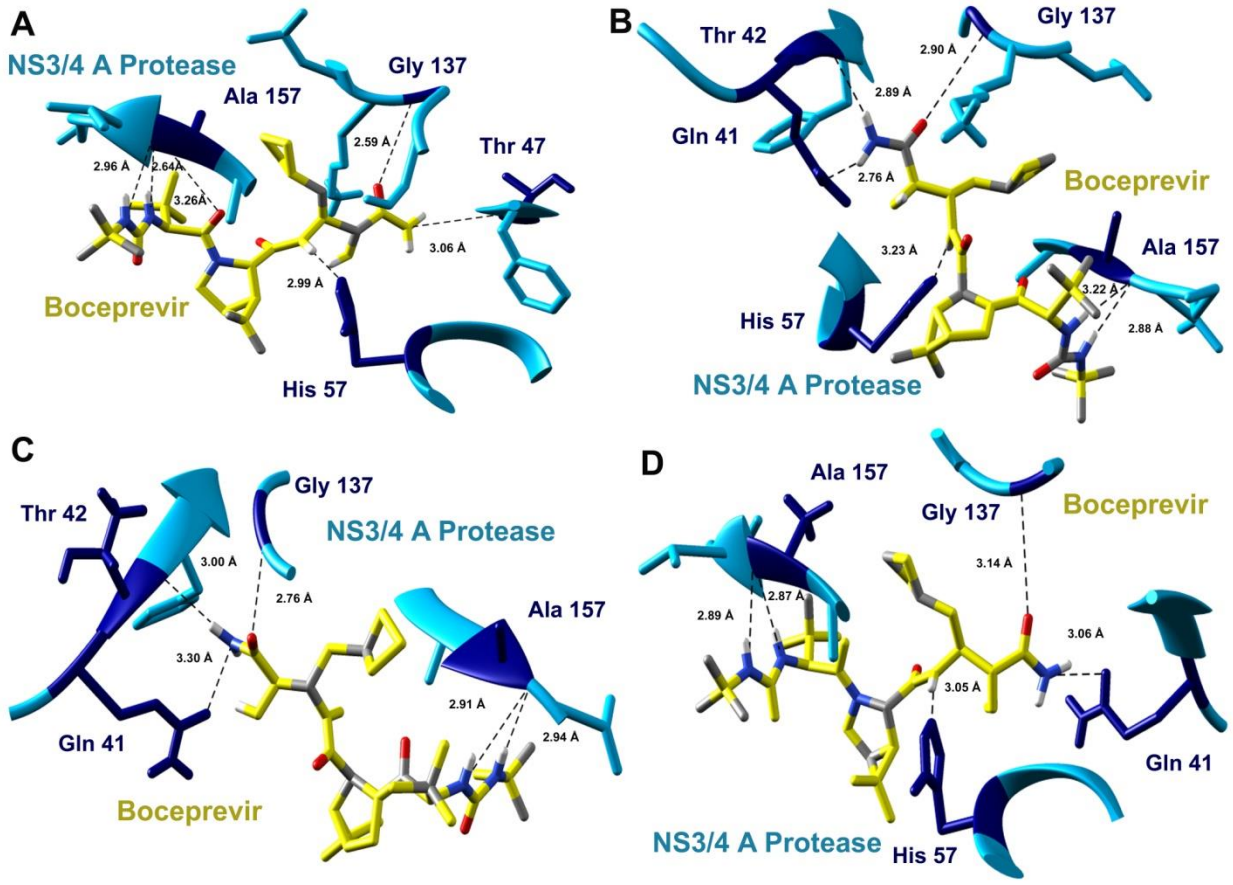

Figure 8: Hydrogen bond interactions between bocepervir and NS3/4 A protease complex (A) Depiction of hydrogen bond interactions with the boceprevir (shown in yellow) and wild type NS3/4 A protease residues (B) R155K(C) T54S (D) V36M mutant type NS3/4 A protease residues (shown in blue)

In wild protease–boceprevir complex, Asp 157 participate in three hydrogen interactions (Boceprevir (N3)----Ala 157 (o)) , (Boceprevir(N8)----Ala 157 (o)) and (Boceprevir(015)----Ala 157 (CA)) with hydrogen bond length of 2.96 , 2.64, 3.26 respectively. Compared with wild type complex, mutations (V36M, T54S and R155K) shows obvious effect on the Asp157 participation in hydrogen interaction, there is a reduction in number of hydrogen bonds and fluctuation in hydrogen bond length.

After the comparative analysis of conformation and dynamics nature of the complexes, it can be observed that amino acid substitution at position 36, 54 and 155 affect the orientation of the

NS3/4A protease active site residues as well as binding mode of the boceprevir. From the Figure 9, it can be seen that there is a fluctuation in the orientation of the active site residues (His 57, Asp 81, Gly 137 and Ser 139) in boceprevir bound wild and mutant NS3/4A protease. For investigating the influence of mutation on the conformational rearrangement of boceprevir binding site, we superimpose the simulated mutant complexes (V36M, T54S and R155K) onto the wild complex (Figure 10). It clarifies that in a mutant complexes, boceprevir have to adjust in a different position to obtain favorable interactions with the active residues.

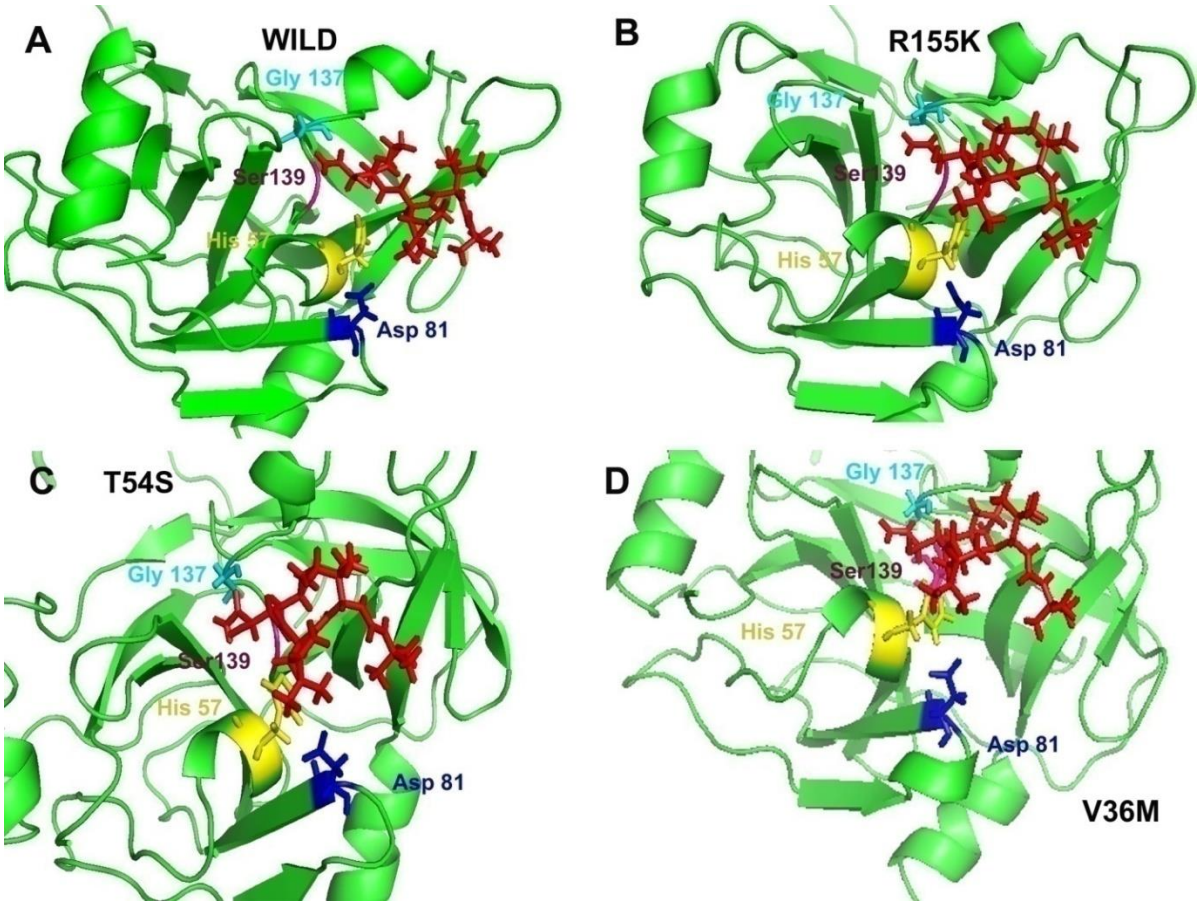

Figure 9: Structural models of simulated boceprevir (shown in Red) bound (A) wild (B) R155K(C) T54S and (D) V36M mutant NS3/4A protease

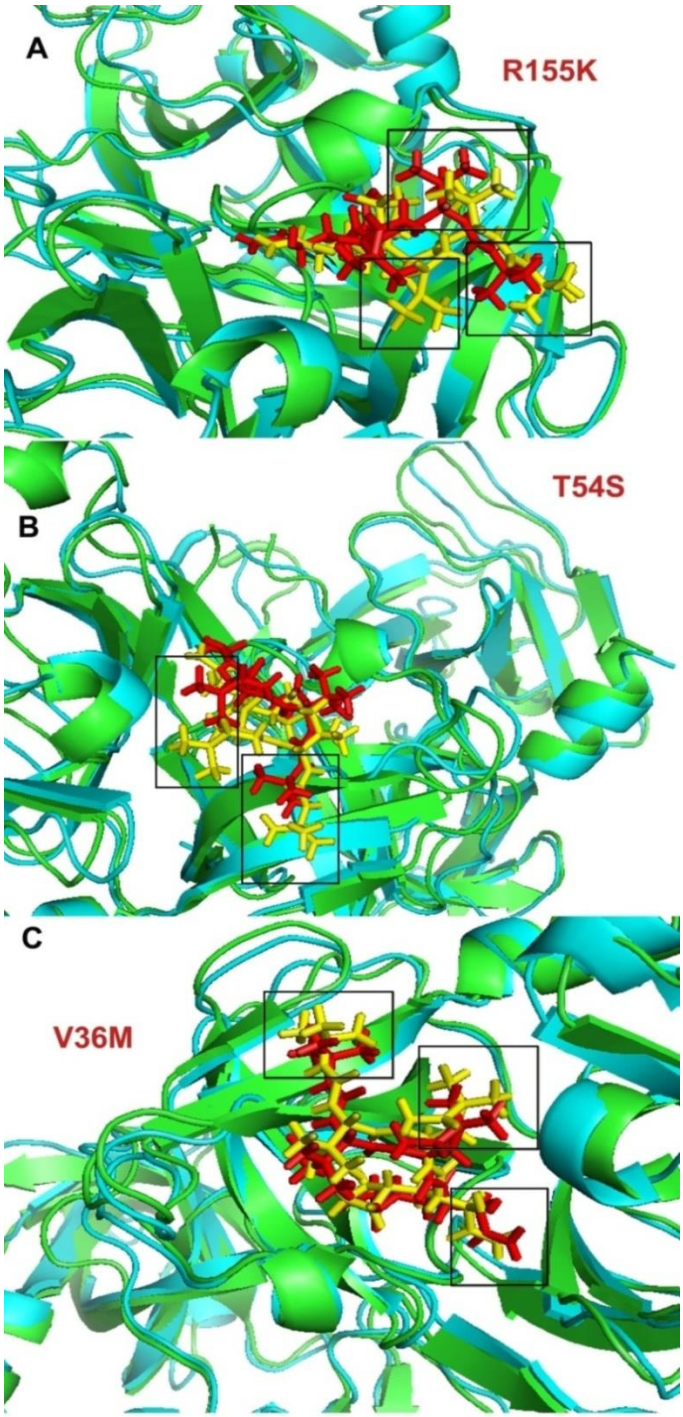

Figure 10 : Structural representation of the superimposed structure of boceprevir bound wild-type and mutant (A) R155K(B) T54S and (C) V36M NS3/4A protease .Cartoons of the Wild (cyan) and mutant (green) are shown

In order to validate the conformational rearrangement of boceprevir binding site, we calculate the distance between P2 of boceprevir and binding site residues of the wild and mutant NS3/4A protease (Table 4). From the table, it is clear that there is an increase in distance between P2 to P4 linker of boceprevir and binding site residues such as (His 57, Asp 81 and Ser 139).

Table 4: Calculated distance between P2 of boceprevir and binding site residues of the wild and mutant NS3/4A protease

Protease-Systems	**Wild**	**R155K**	**T54S**	**V36M**
His57-P2(**Å**)	10.1	11.5	10.8	11.0
Asp 81-P2(**Å**)	7.1	8.2	7.8	7.7
Ser 139-P2(**Å**)	9.0	9.5	9.2	10.4

5.3 Effect of mutation on occupancy of hydrogen bond

In the simulated wild NS3/4A protease - boceprevir complex, six hydrogen bonds were found between NS3/4A protease residues (Ala157, His 57, Gly137, and Thr 42) and Drug boceprevir viral substrate residues. We figured the occupancies of these hydrogen bonds in wild and mutant (V36M, T54S and R155K) protease - boceprevir complexes shown in Table 5.

From Table 3, it can be noted that the Wild NS3/4A protease- boceprevir complex can form hydrogen bonds with higher occupancy and were found stable throughout the simulation. However, mutant complexes (V36M, T54S and R155K) showed reduction in occupancies of hydrogen bonds. Wild protease - boceprevir complex form GLY137(N-H) ----boceprevir (O35) hydrogen bond with 48.60 % occupancy , where as mutants (V36M, T54S and R155K) show lower occupancy for this bond of 38.50 %, 30.73 %and 28.50 % respectively.

Table 5: Occupancy (%) of intermolecular hydrogen bonds participated in wild and mutant protease- boceprevir complex structures

Hydrogen residue pairs		**Occupancy (%)**			
Donor	Acceptor	Wild	R155K	T54S	V36M
Boceprevir-O35	GLY137- N	48.60%	38.50%	30.73%	28.50
Boceprevir- -N3	ALA157-O	26.34%	24.13%	25.43%	28.66%
Boceprevir-- N36	THR42-O	18.16%	4.29%	13.14%	16.84%
Boceprevir- -N8	ALA157- O	9.11%	7.36%	9.59%	6.26%
Boceprevir-- N26	HIS57- NE2	9.11%	5.16%	8.40%	15.26%

5.4 Influence of mutation on salt bridge interaction

We also investigated the stability of the protein structure by examining the salt bridge interactions in the wild and mutant structures. We obtained the salt bridge interactions in the protein structure by using Visual Molecular Dynamics (VMD) and ESBRI Server .All salt bridge interaction pairs are collectively described in the Table 6. The wild protease – boceprevir complex was stabilized by ten salt bridge interactions where as mutant protease – boceprevir complexes were found to attain less than ten salt bridge interactions.

The substitutions at position 155 by lysine, 54 by Serine and 36 by Methionine could destabilize the salt bridge interactions network in the protein. It can be observed that many of the salt bridge interactions disappears in the mutants , thus we can conclude that mutation at position 155, 54 and 36 may disrupt the stability of the NS3/4A protease

Table 6: Salt bridge analysis of boceprevir bound wild and mutant proteases

Salt Bridge Interactions Distance				
Salt Bridge Interactions	**Wild**	**R155K**	**T54S**	**V36M**
OD2 ASP25 - NH2 ARG11	2.64	2.53	2.46	2.7
OD1 ASP81- ND1 HIS57	2.82	2.67	3.17	3.1
OD2 ASP81- ND1 HIS57	3.34	_	2.75	2.9
OE1 GLU30-NH2 ARG92	3.34	2.78	_	3.3
OD2 ASP103-NH1 ARG117	2.5	3.57	_	_
OD2 ASP103-NH2 ARG117	3.46	_	_	_
OD1 ASP168- NH1ARG123	2.9	_	-	_
OD2 ASP168- NH1 ARG155	2.8	_	3.9	2.7
OD1 ASP81- NH2 ARG155	3.5	_	_	_
OE1 GLU176- NH2 ARG180	2.6	3.05	2.5	3.3
OD1 ASP168- NH2 ARG123	_	3.14	_	2.5
OD2 ASP168- NH1 ARG123	_	2.61	2.8	_
OD1 ASP25 - NH2 ARG11	_	_	3.7	_
OD1 ASP103-NH1 ARG117	_	2.98	_	_
OD1 ASP121-NH1 ARG117	_	_	3.6	_
OD1 ASP121-NH2 ARG117	_	_	2.72	_
OD1 ASP168- NH1 ARG155	_	_	2.71	_

5.5 Insight from the substrate envelope analysis

Previous studies on substrate envelope hypothesis have shown that, it is very helpful in elucidating the resistance molecular mechanism of PIs and natural substrate recognition to NS3/4A protease (Romano *et al.*, 2012; 2010; 2011; Xue *et al.*, 2012) .

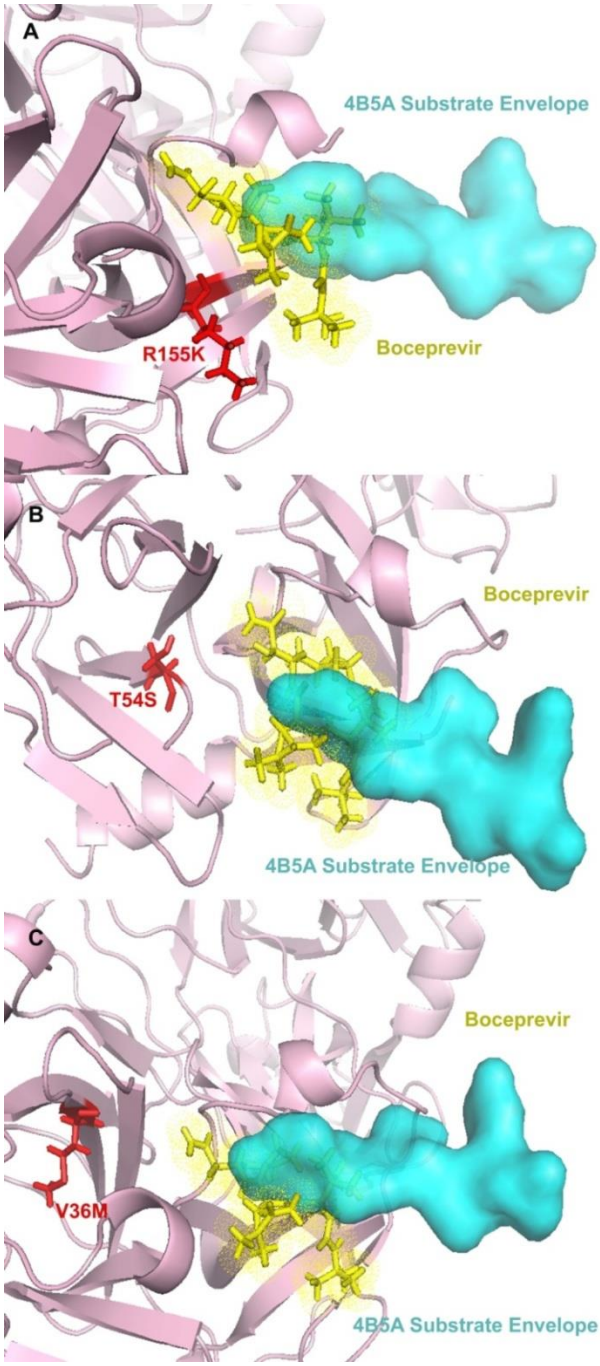

Figure 11: Stereo surface representation of the mutant NS3/4A–boceprevir complex (A) R155K, (B) T54S and (C) V36M superimposed on the 4B5A substrate.

Here, we have superimposed the substrate 4B5A onto the active site of boceprevir bound mutant (V36M, T54S and R155K) NS3/4A protease (Figure 11). In the figure 6, it is demonstrated that all the three mutation residues (V36M, T54S and R155K) has little contribution in the binding of natural substrate to the NS3/4A protease and the boceprevir binding volume protrudes from the substrate envelope volume. Thus the mutation residing outside the substrate envelope may affect the boceprevir affinity towards HCV protease but not the protease enzymatic activity and therefore, these mutations can help NS3/4A protease in escaping from the boceprevir binding without effecting substrate recognition

5.6 NS3 protease ligand-binding volume

To estimate the ability for boceprevir to bind the protease binding pocket, we analyzed the volume of the boceprevir binding pocket in these wild and mutant protease structures. Compared to the wild structure, the binding volume was smaller in the mutant structures such as V36M and T54S, where as for R155K variant, binding volume was found larger (Table 7).The variation in binding pocket volume of mutants could confer drug resistance by disrupting boceprevir-protease interaction.

Table 7: Binding pocket volume of wild and mutant protease

	Protein Structure			
	Wild	**R155K**	**V36M**	**T54S**
Surface Area	205	219	190	200
Volume(Å)	285.6	316.8	232	245

5.7 Parallel virtual screening

Our aim is to redesign boceprevir by attaching chemical groups at positions R1, R2 and R3 of the boceprevir scaffold (Figure 12),which would be effective against both wild and mutants protease. Using Lead grow, we generated a dataset of 30,000 boceprevir derivatives by anchoring different chemical groups at substitution sites.

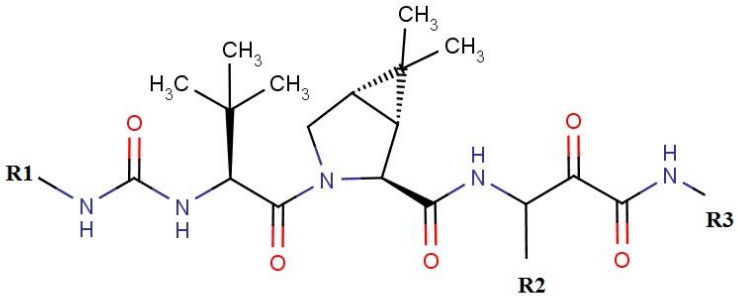

Figure 12: Diagrammatic representation of boceprevir scaffold

For all the boceprevir derivatives, 3D conformations were generated by using the LigPrep tool of Schrödinger suite and these were screened for common inhibitor for both wild and mutant protease by using Glide package. Virtual screening of NS3/4A protease active site of both wild and mutants (V36M, T54S and R155K) against the constructed library was carried out for 30,000 compounds. The active site residues that were accounted for grid generation in the protease were His 57, Asp 81, Ser139, Ala 157 and Gly 137 residues that bind to boceprevir.

100 compounds in common having high binding affinity against both wild and mutants (V36M, T54S and R155K) protease were obtained from which we report top two scoring candidates in this study(Figure13). Two of the top-ranked boceprevir derivatives and their putative molecular interactions with wild and mutants protease are shown in Figure 14-17.

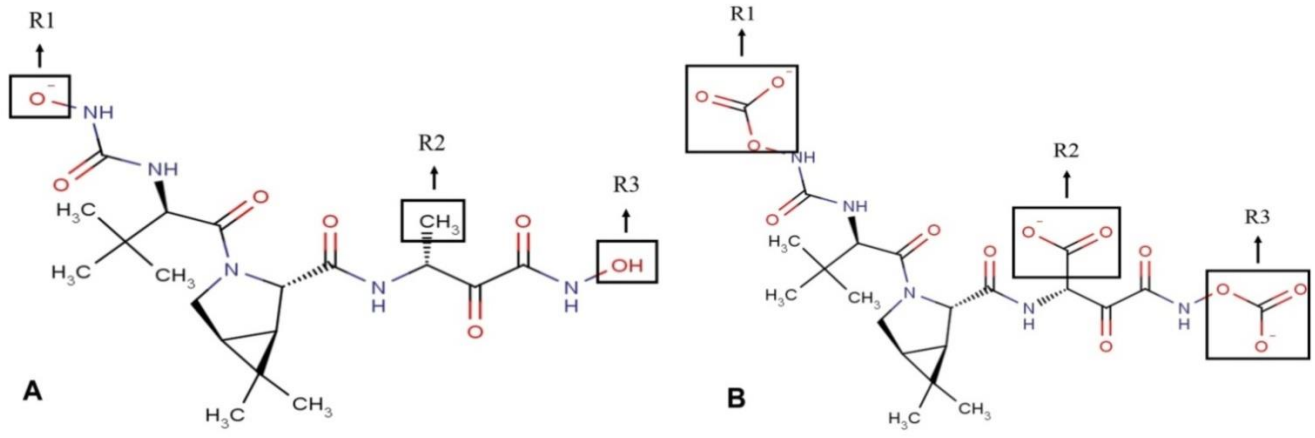

Figure 13: Diagrammatic representation of structures of top-two scoring derivatives (A) Comp1 and (B) Comp 2

Comp1 and Comp 2 showed strong binding interactions with Glide score being more than -6 Kcal/mol. Binding free energies of Comp1 and Comp 2 with wild and mutants were also significantly high; indicating a high binding affinity .Table 8 depicts the docking result in terms of XP glide score, H-bond energy, Glide energy and Glide ligand efficiency

Table 8: List of different parameters considered during the virtual screening

Properties	**Glide Docking score**	**Glide energy (kcal/mol)**	**Glide ligand efficiency**	**XP Hbond**
Wild-Comp1	-6.4	-36.51	-0.20	-3.03
R155K- Comp1	-8.8	-44.35	-0.28	-2.12
V36M- Comp1	-6.3	-50.19	-0.20	-1.63
T54S- Comp1	-6.5	-43.05	-0.21	-1.79
Wild-Comp2	-5.9	-58.43	-0.15	-2.63
R155K- Comp2	-10.0	-60.13	-0.25	-2.47
V36M- Comp2	-6.9	-56.82	-0.17	-2.77
T54S- Comp2	-7.5	-54.23	-0.19	-2.71

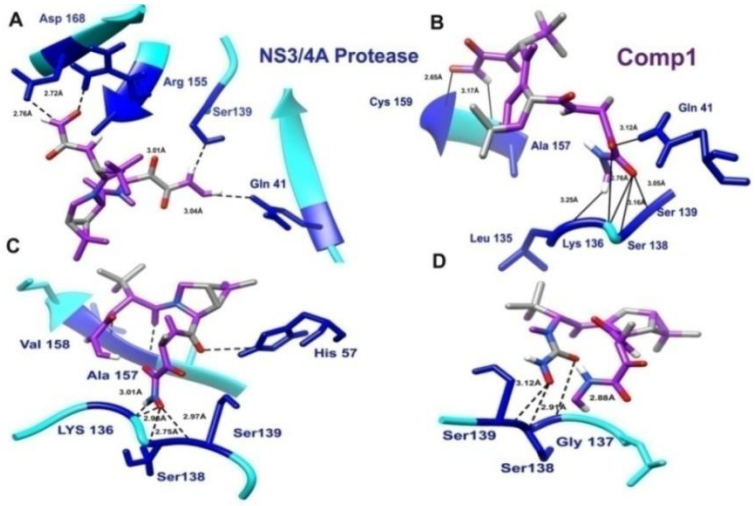

Figure 14: Hydrogen bond interactions between Comp 1 and NS3/4 A protease complex (A) Depiction of hydrogen bond interactions with the Comp 1 (shown in purple) and wild type NS3/4 A protease residues (B) R155K (C) T54S (D) V36M mutant type NS3/4 A protease residues (shown in blue)

The Compound 1 shown in Figure 12a contains oxygenated, methyl, hydroxyl side group at R1, R2 and R3 respectively. Interaction analysis of NS3/4 protease – Comp1docked complex revealed that, there is an increase in H-bonds interactions in mutant protease structure in comparison to the wild protease (Figure 14).

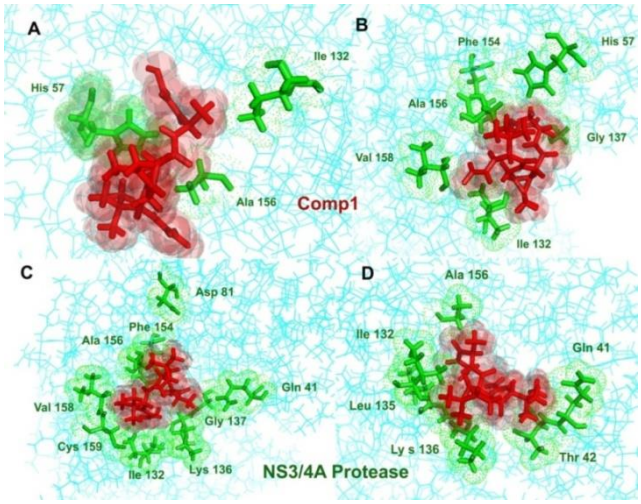

Figure 15: Hydrophobic interactions between Comp 1 and NS3/4 A protease complex (A) Depiction of Hydrophobic bond interactions with the Comp 1 (shown in Red) and wild type NS3/4 A protease residues (B) R155K (C) T54S (D) V36M mutant type NS3/4 A protease residues (shown in Green) .

Along with the hydrogen bond interactions, hydrophobic interactions also increased in mutants (V36M, T54S and R155K) protease structure. Ala 156, Phe154, Lys 136 and Val 158 are resides which showed interaction in all the mutants- comp2 docked protease (Figure 15).

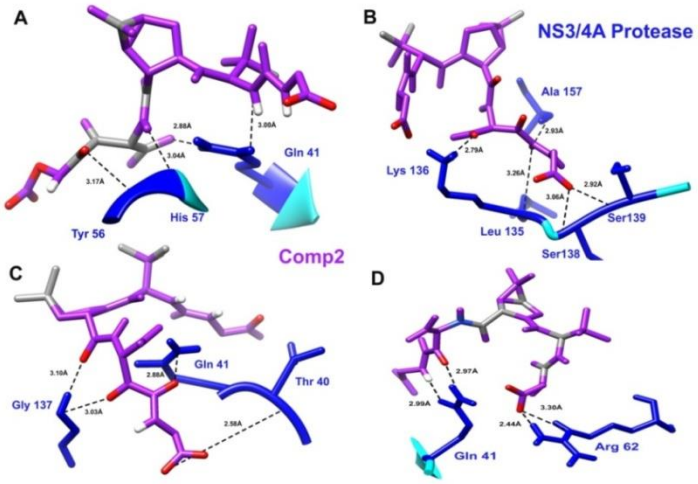

Figure 16: Hydrogen bond interactions between Comp 2 and NS3/4 A protease complex (A) Depiction of hydrogen bond interactions with the Comp 2 (shown in purple) and wild type NS3/4 A protease residues (B) R155K (C) T54S (D) V36M mutant type NS3/4 A protease residues (shown in blue)

The Compound 2 shown in Figure 12a contains poly oxygenated side group at all the substitution sites. Interaction analysis of NS3/4 protease – Comp2 docked complex revealed that, there is a increase in H-bonds interactions in mutant protease structure in comparison to the wild protease (Figure16).

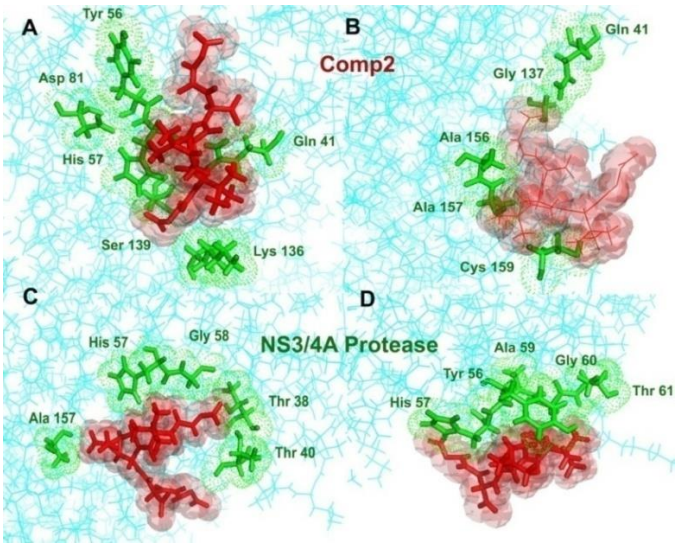

Figure 17: Hydrophobic interactions between Comp 2 and NS3/4 A protease complex (A) Depiction of Hydrophobic bond interactions with the Comp 2 (shown in Red) and wild type NS3/4 A protease residues (B) R155K (C) T54S (D) V36M mutant type NS3/4 A protease residues (shown in Green)

Along with the hydrogen bond interactions, hydrophobic interactions also increased in mutants (V36M, T54S and R155K) protease structure (Figure 17). Therefore these are the potential interactions that could be used for designing the NS3/4A protease inhibitor for both wild and mutant type protease

6. DISCUSSION AND FUTURE PERSPECTIVE

In the present work, we describe a detailed and comprehensive study of the structural features of wild HCV NS3/4A protease and drug resistant mutants to boceprevir through a series of structural bioinformatics approaches such as MD simulation, Solvent accessibility variation in Binding residues, Salt bridge interactions analysis and substrate envelope analysis
Full atomistic molecular dynamics simulations have been utilized for studying the dynamic behavior and binding of anti-viral drugs boceprevir to HCV protease. RMSD analysis of the wild and mutant complex simulation shows unstable conformation of the mutant complexes with reference to the wild complex.

The RMSFs analysis of all the active site residues provides more comprehensive information on flexibility and conformational changes of the mutant complexes. In the T54S mutant structure, flexibility of binding residues were found least where as R155K and V36M variants showed intermediate flexibility. The larger RMSFs reflect higher structural flexibility of binding residues in the wild protease than the residues in the R155K, T54S and V36M mutant structures. RMSF displacement is further supported by Solvent accessibility variation. Binding residues in wild type NS3/4A protease showed higher solvent accessibility than all mutant type (V36M, T54S and R155K) NS3/4A protease and these Shrunken SASA values of the mutant structures affects the interaction behavior of boceprevir in mutant protease structures.
Furthermore, we analyzed conformational rearrangement of the simulated mutant structures, affecting the orientation of the NS3/4A protease active site residues as well as binding mode of the boceprevir.

V36M and T54S mutations are positioned away from the ligand binding pocket and in protein interior, but mutation at position 36 and 54 displace the orientation of the side chain of these residues, therefore decreasing the binding pocket volume. R155K reside in the protease binding pocket, thereby substitution at 155 position directly influence the shape and volume of the binding pocket. These variations in binding pocket volume of mutants could confer drug resistance by disrupting boceprevir-protease interaction

In an effort to redesign the boceprevir protease inhibitor, we performed parallel virtual screening of a constructed combinatorial library of boceprevir against wild and corresponding mutant NS3/4A protease. The comparative analysis of the binding affinity and interactions of the top two compounds docked to NS3/4A protease and variants suggest that: (1) poly oxygenated group attached to the boceprevir scaffold could increase the binding affinity(2) large chemical group provoke additional interactions with the binding residues by reaching deeper into the binding pocket.

The results suggests few considerations for the designing of new NS3/4A protease inhibitors such as to reduce vulnerability to conformational rearrangement of the binding residues, the shape of the inhibitors should stabilize the binding residues conformation by enlarging the binding site and ligand should show functionality in recovering the optimized interactions between the binding residues and drug.

Substrate envelope hypothesis suggest that inhibitor designed to be confined within the substrate envelope might be able to combat drug resistance, as the mutations weakening inhibitor binding would also compromise the substrate recognition. This mechanistic insight into drug resistance will provide a valuable clue for designing and increasing the potency of inhibitors.

7. CONCLUSION

In order to reveal the mechanism of boceprevir resistance at atomic level, we considered various factors such as Solvent accessibility variation in Binding residues, Salt bridge interactions analysis and Molecular dynamic simulation. MD Simulation elicit the study of stability and flexibility of the protein by investigating RMSD, Rg, RMSF and hydrophobic interaction, hydrogen bond networks of the wild and mutant protein system. Moreover, solvent accessibility variation represents the influence of mutation on binding site residue's conformation. MD Simulation studies disclose decreased conformational mobility and the unstable RMSD of the Mutant backbone compared with wild backbone. This was further provoked by the lower flexibility of the ligand binding residues of the mutant type structures. Representative structure of the simulation result have showed a strong hydrophobic interactions between boceprevir and wild type, while the mutants showed a loss of hydrophobic interactions, which indicates that because of mutation protein-boceprevir interaction has weaken. With the hydrophobic interaction, hydrogen bond occupancy and salt bridge interactions also showed reduction in mutant complex structures.

In binding pocket analysis, variation in cavity volume is observed, V36M and T54S showed smaller cavity volume while R155K variant have larger cavity volume in comparison to the wild NS3/4A protease structure. This difference in cavity volume represents the effect of mutation on optimize interaction between boceprevir and NS3/4A protease.

The superimposition of the 4B5A substrate onto the active site of boceprevir bound mutant (V36M, T54S and R155K) NS3/4A protease demonstrated that all the three mutation residues has little contribution in the binding of natural substrate to the NS3/4A protease and the boceprevir binding volume protrudes from the substrate envelope volume. Thus the mutation residing outside the substrate envelope may affect the boceprevir affinity towards HCV protease but not the protease enzymatic activity and therefore, these mutations can help NS3/4A protease in escaping from the boceprevir binding without effecting substrate recognition.

From these data, it can be concluded that boceprevir would be active with the wild protease by showing favorable interactions, but mutants studied shall impaired the interactions, by affecting the flexibility and stability of the binding site residues. The structural insight from this study uncovers ambiguous mechanism of the boceprevir resistance and the results can be valuable for the design of new PIs with improved efficiency to combat resistance.

8. REFERENCES

1. Bartenschlager, R. (1999). The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. J Viral Hepat, ***6***, 165-181.
2. Bartenschlager, R; Ahlborn-Laake, L; Mous, J. and Jacobsen, H. (1993). Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. J Virol, ***67***, 3835-3844.
3. Binkowski, TA; Naghibzadeh, S. and Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids *Res*, ***31***, 3352-3355.
4. Choo, Q L; Kuo, G; Weiner, A J; Overby, L R; Bradley, D W and Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science, ***244***, 359-362.
5. Costantini, S; Colonna, G and Facchiano, AM. (2008). ESBRI: a web server for evaluating salt bridges in proteins. Bioinformation, **3**, 137-138.
6. Failla, C; Tomei, L. and De Francesco, R. (1995). An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. J Virol, ***69***, 1769-1777.
7. Friesner, RA; Banks, JL; Murphy, RB; Halgren, TA ; Klicic, JJ ; Mainz, DT; Shenkin, P S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem, ***47***, 1739-1749.
8. Griffin, SD; Beales, LP; Clarke, DS; Worsfold, O; Evans, SD; Jaeger, J; Rowlands, DJ. (2003). The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. FEBS Lett, ***535***, 34-38.
9. Guo, Z; Prongay, A; Tong, X; Fischmann, T; Bogen, S; Velazquez, F; Madison, V. (2006). Computational Study of the Effects of Mutations A156T, D168V, and D168Q on the Binding of HCV Protease Inhibitors. Journal of Chemical Theory and Computation,***2,*** 1657-1663.
10. Halfon, P. and Locarnini, S. (2011). Hepatitis C virus resistance to protease inhibitors. J Hepatol, ***55***, 192-206. do
11. Harper, S; McCauley, JA; Rudd, MT; Ferrara, M; DiFilippo, M; Crescenzi, B; Liverton, NJ. (2012). Discovery of MK-5172, a Macrocyclic Hepatitis C Virus NS3/4a Protease Inhibitor. ACS Med Chem Lett, ***3***, 332-336.
12. Humphrey, W; Dalke, A. and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph, ***14***, 33-38, 27-38.
13. Kieffer, TL; Sarrazin, C; Miller, JS; Welker, MW; Forestier, N; Reesink, H. W; Zeuzem, S. (2007). Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Hepatology, ***46***, 631-639.

14. Kim, JL; Morgenstern, KA; Griffith, JP; Dwyer, MD; Thomson, JA; Murcko, MA; Caron, PR. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure, ***6***, 89-100.
15. Kolykhalov, AA; Mihalik, K; Feinstone, SM. and Rice, CM. (2000). Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. J Virol, ***74***, 2046-2051.
16. Kwo, PY; Lawitz, EJ; McCone, J; Schiff, ER; Vierling, JM; Pound, D; Albrecht, JK. (2010). Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial. Lancet, ***376***, 705-716.
17. Kwong, AD; Kim, JL; Rao, G; Lipovsek, D. and Raybuck, SA. (1998). Hepatitis C virus NS3/4A protease. Antiviral *Res*, ***40***, 1-18.
18. Kwong, AD; McNair, L; Jacobson, I. and George, S. (2008). Recent progress in the development of selected hepatitis C virus NS3.4A protease and NS5B polymerase inhibitors. Curr Opin Pharmacol, ***8***, 522-531.
19. Laskowski, RA; Hutchinson, EG; Michie, AD; Wallace, AC; Jones, M L and Thornton, JM. (1997). PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci, ***22***, 488-490.
20. Lin, C; Thomson, JA and Rice, CM. (1995). A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro. J Virol, ***69,*** 4373-4380.
21. Lin, TI; Lenz, O; Fanning, G; Verbinnen, T; Delouvroy, F; Scholliers, A; Simmen, K. (2009). In vitro activity and preclinical profile of TMC435350, a potent hepatitis C virus protease inhibitor. Antimicrob Agents Chemother, ***53,*** 1377-1385.
22. Lindenbach, BD. and Rice, CM. (2005). Unravelling hepatitis C virus replication from genome to function. Nature, ***436***, 933-938.
23. Liverton, NJ; Carroll, SS; Dimuzio, J; Fandozzi, C; Graham, DJ; Hazuda, D; Vacca, JP. (2010). MK-7009, a potent and selective inhibitor of hepatitis C virus NS3/4A protease. Antimicrob Agents Chemother, ***54***, 305-311.
24. Lobanov, M; Bogatyreva, NS and Galzitskaia, OV. (2008). Radius of gyration is indicator of compactness of protein structure. Mol Biol (Mosk), ***42,*** 701-706.
25. Love, RA; Parge, HE; Wickersham, JA; Hostomsky, Z; Habuka, N; Moomaw, EW; Hostomska, Z. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. Cell, ***87,*** 331-342.
26. Mohd Hanafiah, K; Groeger, J; Flaxman, AD and Wiersma, ST. (2013). Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology*, ***57,*** 1333-1342.

27. Pan D; Xue W; Zhang W; Liu H and Yao X. (2012) Understanding the drug resistance mechanism of hepatitis C virus NS3/4A to ITMN-191 due to R155K, A156V, D168A/E mutations: a computational study. Biochim Biophys Acta, ***1820***, 1526-1534.
28. Pavlovic, D; Neville, DC; Argaud, O; Blumberg, B; Dwek, RA; Fischer, WB; & Zitzmann, N. (2003). The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. Proc Natl Acad Sci U S A, ***100***, 6104-6108.
29. Pikkemaat, MG; Linssen, AB; Berendsen, HJ. and Janssen, D. B. (2002). Molecular dynamics simulations as a tool for improving protein stability. Protein Eng, ***15***, 185-192.
30. Prongay, AJ; Guo, Z; Yao, N; Pichardo, J; Fischmann, T; Strickland, C; Madison, V. (2007). Discovery of the HCV NS3/4A protease inhibitor (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3- [2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]- 6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (Sch 503034) II. Key steps in structure-based optimization. J Med Chem, ***50***, 2310-2318.
31. Reed, K. E. and Rice, C. M. (1998). Molecular characterization of hepatitis C virus. *Curr Stud* Hematol Blood Transfus ,***62***, 1-37.
32. Richard JO Barnard, JL; Stefan Zeuzem, John M. (2011). Analysis of resistance-associated amino acid variants (RAVs) in non-SVR patients enrolled in a retrospective long-term follow-up analysis of boceprevir phase 3 clinical studies. Hepatology.
33. Romano, KP; Ali, A; Aydin, C; Soumana, D; Ozen, A; Deveau, LM. and Schiffer, CA. (2012). The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. PLoS Pathog, ***8,*** e1002832.
34. Romano, KP; Ali, A; Royer, WE and Schiffer, CA. (2010). Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. Proc Natl Acad Sci U S A, ***107***, 20986-20991.
35. Romano, KP; Laine, JM; Deveau, LM; Cao, H; Massi, F. and Schiffer, CA. (2011). Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease. J Virol, ***85,*** 6106-6116.
36. Sarrazin, C; Kieffer, TL; Bartels, D; Hanzelka, B; Muh, U; Welker, M. and Kwong, A. D. (2007). Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. Gastroenterology, ***132,*** 1767-1777.
37. Schrodinger, LLC. (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
38. Schrödinger, L. (2009). LigPrep, version 2.3. Schrödinger, LLC, New York.
39. Seiwert, SD; Andrews, SW; Jiang, Y; Serebryany, V; Tan, H; Kossen, K; Blatt, M. (2008). Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227). Antimicrob Agents Chemother, ***52***, 4432-4441.
40. Shaw, DE. (2009). Desmond Molecular Dynamics Sysytem. Research/Schrodinger.(2.2 edn. New York).

41. Simmonds, P; Bukh, J; Combet, C; Deleage, G; Enomoto, N; Feinstone, S. and Widell, A. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology, ***42***, 962-973. doi: 10.1002/hep.20819
42. Simmonds, P; Holmes, EC; Cha, TA; Chan, SW; McOmish, F; Irvine, B. and Urdea, M. S. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol, ***74*** , 2391-2399.
43. Susser, S; Welsch, C; Wang, Y; Zettler, M; Domingues, FS; Karey, U.and Sarrazin, C. (2009). Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. Hepatology, ***50***, 1709-1718. doi: 10.1002/hep.23192
44. Tong, X; Chase, R; Skelton, A; Chen, T; Wright-Minogue, J. and Malcolm, B. A. (2006). Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. Antiviral Res, ***70***, 28-38.
45. Victrelis (boceprevir). (2011).
46. VLife, M. (2012). 3.5 (2004) Molecular Design Suite. Vlife Sciences Technologies Pvt. Ltd, Pune, India.
47. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J Mol Graph, , 52-56, 29.
48. Welsch, C; Domingues, FS; Susser, S; Antes, I; Hartmann, C; Mayr, G. and Lengauer, T. (2008). Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus. Genome Biol, ***9,*** R16.
49. Welsch, C; Schweizer, S; Shimakami, T; Domingues, FS; Kim, S; Lemon, SM and Antes, I. (2012). Ketoamide resistance and hepatitis C virus fitness in val55 variants of the NS3 serine protease. Antimicrob Agents Chemother, ***56***, 1907-1915.
50. World Health Organization : Hepatitis C Fact sheet. (2014). Hepatology.
51. Xue, W; Pan, D; Yang, Y; Liu, H. and Yao, X. (2012). Molecular modeling study on the resistance mechanism of HCV NS3/4A serine protease mutants R155K, A156V and D168A to TMC435. Antiviral Res, ***93,*** 126-137.
52. Xue, W; Ban Y; Liu H. and Yao X. (2014) Computational study on the drug resistance mechanism against HCV NS3/4A protease inhibitors vaniprevir and MK-5172 by the combination use of molecular dynamics simulation, residue interaction network, and substrate envelope analysis. J Chem Inf Model, ***54***,621-633.
53. Yan, Y; Li, Y; Munshi, S; Sardana, V; Cole, JL; Sardana, M. and Chen, Z. (1998). Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. Protein Sci, ***7***, 837-847.

9. APPENDIX

RMSD values of NS3/4A protease complex with boceprevir backbone of each system

RMSD Values of Protease Systems				
TIME STEP (ns)	**WILD**	**R155K**	**T54S**	**V36M**
0	5.95E-15	4.52E-15	4.33E-15	6.10E-15
0.504	1.725207	1.396768	1.785209	1.414118
1.0032	1.701573	1.513876	1.36328	1.691642
1.5024	1.81485	1.725728	1.651861	1.747877
2.0016	1.778731	1.711693	1.754142	1.956722
2.5008	1.968009	1.810511	1.805113	2.529417
3	1.693438	1.787021	1.983291	1.816072
3.552	1.762087	1.829188	2.089468	1.768439
4.0032	1.574674	1.674868	2.501866	2.533161
4.5504	1.757096	1.845239	2.199133	2.700565
5.0016	1.771468	1.951708	1.926989	2.282336
5.5008	1.936982	2.023938	1.898622	2.039197
6	1.845342	1.802227	1.703096	2.285084
6.504	1.864194	1.807421	2.120044	2.360737
7.0032	2.116441	1.739431	2.031129	2.166123
7.5024	2.044764	1.591787	2.003641	1.971059
8.0016	2.287754	1.817975	2.004787	2.161055
8.5008	1.905619	2.209179	2.066983	1.954584
9	1.936245	1.869067	2.320121	1.966314
9.504	1.839393	2.013091	2.402554	1.956869
10.0032	1.840581	2.254605	2.443148	2.098056
10.5024	2.069154	2.053957	2.578804	2.155534
11.0016	2.250522	1.82373	2.537543	2.176546
11.5008	1.972221	2.398303	2.374682	1.959187
12	2.125385	1.723904	2.648409	2.068809
12.504	1.972184	1.913189	2.29951	2.127589
13.0032	2.236245	2.182676	2.141144	2.201732
13.5024	2.274385	2.189766	2.165328	2.087217

RMSD Values of Protease Systems				
TIME STEP (ns)	**WILD**	**R155K**	**T54S**	**V36M**
14.0016	2.072246	2.289574	2.189742	2.191796
14.5008	1.974979	2.124383	2.38181	2.096597
15	1.80874	2.082116	2.692463	2.135774
15.504	2.084649	2.580564	2.285774	2.29067
16.0032	2.16875	2.314381	2.138387	2.209922
16.5024	2.127768	2.718788	2.508159	2.207674
17.0016	2.212539	2.634686	2.571315	2.56382
17.5008	2.116195	2.42845	2.342326	2.490794
18	2.032902	2.552035	2.330455	2.737913
18.504	1.952234	2.132836	2.491352	2.264587
19.0032	2.123753	2.102116	2.241332	2.045026
19.5024	2.20098	2.030504	2.395397	2.610498
20	1.982995	1.968445	2.351352	2.655236

RMSF values for all wild and mutant structure residues over time

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
0	1.59365	1.521762	1.387437	1.582723
1	1.357015	1.14106	1.115089	1.18236
2	1.088868	0.817957	0.78912	0.915159
3	0.913142	0.681597	0.684986	0.750795
4	0.835568	0.619903	0.603967	0.715419
5	0.712029	0.577387	0.534121	0.672663
6	0.624655	0.556916	0.495629	0.646739
7	0.658331	0.623931	0.550763	0.698836
8	0.575096	0.5888	0.562939	0.591264
9	0.71334	0.650658	0.622141	0.625703
10	0.815366	0.690349	0.701301	0.665567
11	0.881806	0.681611	0.768019	0.740942
12	1.277217	0.930199	1.135371	0.975365
13	1.877898	1.567431	1.319802	1.060064
14	2.494386	1.95428	1.446978	1.26
15	2.090279	1.576083	1.352609	1.184786
16	1.175746	0.749417	0.844707	0.757017
17	0.93876	0.735784	0.793852	0.708384
18	0.827942	0.736313	0.794176	0.727958
19	3.471852	3.291687	2.513667	2.432786
20	2.626568	2.189375	1.863065	1.913331
21	1.99767	1.681455	1.462557	1.531544
22	1.615225	1.385237	1.178859	1.28925
23	1.396898	1.206903	1.009393	1.124476
24	1.376563	1.159235	1.0063	1.075778

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
25	1.374039	1.126984	0.964927	1.045324
26	1.37072	1.200682	0.989423	1.110069
27	1.841518	1.538449	1.596591	1.54647
28	1.665104	1.313124	1.22113	1.34364
29	1.332909	1.056995	1.00617	1.04482
30	1.200278	0.950973	0.929682	0.906853
31	1.140489	0.904368	0.870318	0.863516
32	1.108764	0.977923	0.758761	0.853843
33	1.166848	1.042261	0.834416	0.911887
34	1.255789	1.063323	0.931105	0.982084
35	1.442325	1.22978	1.093793	1.148408
36	2.220365	1.980439	1.76798	1.772
37	1.690169	1.426819	1.258813	1.363698
38	1.983813	1.789599	1.581245	1.623643
39	2.319695	2.251274	2.118287	2.067982
40	2.487464	2.507255	2.2953	2.277643
41	3.165893	3.189011	2.972667	3.058922
42	1.569817	1.449145	1.336016	1.411706
43	1.388614	1.222183	1.079408	1.178415
44	1.270228	1.083291	0.937457	1.039163
45	1.261303	1.045256	0.904347	0.966914
46	1.252189	0.994669	0.857226	0.927834
47	1.303496	0.989089	0.850757	0.908085
48	1.436642	1.085822	0.962996	0.983326
49	1.67397	1.271615	1.167766	1.172495
50	1.773297	1.357719	1.195179	1.273673
51	1.588665	1.245452	1.077723	1.127942

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
52	1.492513	1.191667	1.009054	1.08414
53	1.389892	1.131256	0.953959	1.029367
54	1.452636	1.181351	0.995322	1.10455
55	1.57755	1.289634	1.078286	1.218453
56	1.786083	1.419532	1.206961	1.365488
57	1.992672	1.532147	1.332341	1.49842
58	2.465549	1.87477	1.770491	1.767202
59	2.079265	1.610666	1.410899	1.522897
60	2.295083	1.643181	1.378376	1.548444
61	2.240318	1.764965	1.46993	1.639723
62	1.980847	1.662487	1.381701	1.538463
63	1.961257	1.644979	1.342052	1.535268
64	1.899737	1.561006	1.294158	1.490768
65	1.964338	1.623994	1.357456	1.611608
66	2.269347	1.779933	1.552466	1.789437
67	2.690286	2.117983	1.878939	2.15011
68	2.857753	2.268437	1.912914	2.180459
69	2.57479	1.918232	1.693028	1.94356
70	2.328163	1.808834	1.539572	1.733234
71	2.172456	1.688623	1.449711	1.596358
72	2.111474	1.716208	1.416618	1.551052
73	1.990414	1.659519	1.340622	1.468526
74	2.020609	1.682254	1.378516	1.481537
75	1.917177	1.597974	1.315418	1.435314
76	2.084453	1.736913	1.415598	1.599992
77	2.065544	1.705016	1.411245	1.660711
78	2.256281	1.790634	1.533715	1.755139

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
79	2.30586	1.82044	1.55319	1.765046
80	2.040008	1.597961	1.352616	1.493317
81	1.865964	1.449475	1.2255	1.398058
82	1.739198	1.395012	1.177545	1.331801
83	1.64563	1.34666	1.122347	1.229951
84	1.73607	1.420125	1.164854	1.268014
85	1.706871	1.386855	1.170056	1.259812
86	1.80847	1.435409	1.261045	1.316255
87	1.764923	1.410042	1.227427	1.287032
88	1.798732	1.506259	1.293298	1.310009
89	1.921607	1.623171	1.550635	1.515468
90	1.718046	1.420269	1.235513	1.310349
91	1.504696	1.207626	1.072496	1.074279
92	1.328966	1.061038	0.942483	0.93522
93	1.332992	1.062545	0.912019	0.967375
94	1.186701	0.916071	0.778351	0.865433
95	1.199347	0.92794	0.800088	0.860402
96	1.25749	1.007066	0.843512	0.946981
97	1.338825	1.144998	0.909224	1.148426
98	1.825731	1.747282	1.319844	1.684757
99	1.815968	1.675965	1.397371	1.86878
100	1.975467	1.775546	1.487529	1.845454
101	1.604186	1.411279	1.229201	1.530032
102	1.427981	1.140944	1.134425	1.337178
103	1.236611	1.010458	0.964295	1.163002
104	1.09379	0.884578	0.813445	1.019699
105	0.912811	0.756703	0.690989	0.878025

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
106	0.94045	0.802303	0.71453	0.912721
107	0.997992	0.900272	0.755844	0.943055
108	1.085001	1.073195	0.830997	1.072302
109	1.20866	1.202093	0.926547	1.20102
110	1.316474	1.224423	1.007609	1.346795
111	1.153719	1.046788	0.903347	1.173767
112	1.115775	1.053868	0.851775	1.142726
113	1.023788	0.902282	0.809162	1.008727
114	1.010704	0.870904	0.823619	1.019731
115	1.042525	0.903689	0.820611	1.054068
116	1.113326	0.94005	0.834784	1.094423
117	1.251924	1.028767	0.9134	1.23375
118	1.421818	1.101096	0.986769	1.2828
119	1.737362	1.402939	1.225784	1.496508
120	1.875503	1.534156	1.392295	1.631177
121	1.822889	1.517385	1.303828	1.599149
122	1.556227	1.267024	1.099111	1.38633
123	1.478499	1.183437	1.067024	1.357257
124	1.405299	1.15874	1.009984	1.329521
125	1.357032	1.110543	1.011018	1.33925
126	1.424649	1.143517	1.081869	1.395364
127	1.744442	1.275991	1.350215	1.587914
128	1.878865	1.441226	1.546309	1.808253
129	1.865072	1.6077	1.638822	1.920601
130	1.701906	1.517348	1.524071	1.793814
131	1.718895	1.568339	1.53136	1.832458
132	1.669963	1.564116	1.483207	1.764659

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
133	1.732767	1.706665	1.598078	1.860989
134	1.641563	1.634649	1.487263	1.787632
135	1.446765	1.399531	1.24046	1.5452
136	1.662711	1.597144	1.404632	1.726135
137	1.529041	1.459704	1.208278	1.601897
138	1.325163	1.209407	0.964306	1.262852
139	1.340119	1.153315	0.953037	1.203343
140	1.283108	1.094022	0.916641	1.084646
141	1.127544	0.991319	0.82408	0.995782
142	1.000793	0.8319	0.7122	0.855373
143	0.968841	0.789057	0.698034	0.847117
144	0.887933	0.712665	0.652053	0.815793
145	0.920859	0.725972	0.667867	0.828445
146	0.981671	0.794025	0.735245	0.876355
147	0.979001	0.821063	0.753316	0.848956
148	0.992264	0.754313	0.725362	0.795011
149	0.961618	0.73682	0.666318	0.745013
150	1.021678	0.825223	0.710186	0.804725
151	1.099955	0.893502	0.767846	0.907898
152	1.180515	0.983957	0.850348	1.016698
153	1.198101	0.997703	0.858707	1.034645
154	1.212542	1.007927	0.878723	1.053324
155	1.410247	1.154715	1.011327	1.21357
156	1.476362	1.239074	1.072076	1.331198
157	1.572584	1.369998	1.181912	1.486353
158	1.735258	1.56114	1.355329	1.680023
159	2.016282	1.763163	1.735696	1.979335

RMSF Values for all residues over time				
Residue No.	**WILD**	**R155K**	**T54S**	**V36M**
160	2.222421	1.81212	2.132659	2.237305
161	2.567934	2.196198	2.365217	2.601036
162	2.377608	2.128328	2.197643	2.431011
163	1.981563	1.697027	1.677929	2.016809
164	1.694755	1.42583	1.409313	1.707137
165	1.601119	1.344265	1.395163	1.602535
166	1.571506	1.272886	1.153688	1.543359
167	1.405477	1.126958	0.997057	1.32385
168	1.368596	1.099871	0.970375	1.230348
169	1.336903	1.074245	0.933296	1.160968
170	1.341405	1.087683	0.924075	1.124336
171	1.347834	1.079552	0.919152	1.096982
172	1.354212	1.100581	0.949008	1.081434
173	1.579248	1.279271	1.101519	1.254999
174	1.65166	1.332625	1.169386	1.289681
175	1.613399	1.316091	1.145532	1.263013
176	1.755539	1.417793	1.207965	1.380913
177	1.946518	1.575054	1.359072	1.522215
178	1.95105	1.588532	1.385773	1.525896
179	2.554342	2.162618	1.844732	2.067648
180	1.28485	1.036978	1.271111	1.182692
181	3.431213	3.393893	3.62445	3.43493

Occupancy (%) of intermolecular hydrogen bonds participated in wild protease-boceprevir complex

Donor	Acceptor	Occupancy
SegC2-GLY137-Main-N	SegC2-HU5999-Side-O35	48.60%
SegC2-HU5999-Side-N3	SegC2-ALA157-Main-O	26.34%
SegC2-HU5999-Side-N36	SegC2-THR42-Main-O	18.16%
SegC2-HU5999-Side-O33	SegC2-HIS57-Side-NE2	7.89%
SegC2-HU5999-Side-N8	SegC2-ALA157-Main-O	9.11%
SegC2-HU5999-Side-N26	SegC2-HIS57-Side-NE2	11.42%
SegC2-HU5999-Side-N36	SegC2-GLN41-Side-OE1	4.05%
SegC2-HU5999-Side-N26	SegC2-ARG155-Main-O	0.02%
SegC2-GLN41-Side-NE2	SegC2-HU5999-Side-O33	0.12%
SegC2-HU5999-Side-O33	SegC2-GLN41-Side-OE1	1.39%
SegC2-HU5999-Side-O33	SegC2-HIS57-Side-CD2	0.02%
SegC2-SER139-Main-N	SegC2-HU5999-Side-O35	0.02%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O25	0.07%
SegC2-ALA157-Main-N	SegC2-HU5999-Side-O15	0.05%
SegC2-SER139-Side-OG	SegC2-HU5999-Side-O35	0.02%

Occupancy (%) of intermolecular hydrogen bonds participated in R155K mutant protease-boceprevir complex

Donor	Acceptor	Occupancy
SegC2-HU5999-Side-N36	SegC2-GLN41-Side-OE1	3.45%
SegC2-HU5999-Side-N3	SegC2-ALA157-Main-O	24.13%
SegC2-GLY137-Main-N	SegC2-HU5999-Side-O35	38.50%
SegC2-HU5999-Side-O33	SegC2-HIS57-Side-NE2	4.03%
SegC2-HU5999-Side-N36	SegC2-THR42-Main-O	4.29%

Donor	Acceptor	Occupancy
SegC2-HU5999-Side-N26	SegC2-HIS57-Side-NE2	5.16%
SegC2-HU5999-Side-N8	SegC2-ALA157-Main-O	7.36%
SegC2-HU5999-Side-N26	SegC2-LYS155-Main-O	1.30%
SegC2-ALA157-Main-N	SegC2-HU5999-Side-O15	1.10%
SegC2-HU5999-Side-O33	SegC2-HIS57-Side-CD2	0.02%
SegC2-HU5999-Side-O33	SegC2-GLN41-Side-OE1	1.87%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O35	0.02%
SegC2-GLN41-Side-NE2	SegC2-HU5999-Side-O33	0.14%
SegC2-SER139-Side-OG	SegC2-HU5999-Side-O35	0.41%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O25	0.10%
SegC2-HU5999-Side-N36	SegC2-GLN41-Side-NE2	0.02%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O33	0.07%

Occupancy (%) of intermolecular hydrogen bonds participated in T54S mutant protease-boceprevir complex

Donor	Acceptor	Occupancy
SegC2-GLY137-Main-N	SegC2-GLY137-Main-N	30.73%
SegC2-HU5999-Side-O33	SegC2-HU5999-Side-O34	3.26%
SegC2-HU5999-Side-N3	SegC2-HU5999-Side-N4	25.43%
SegC2-HU5999-Side-N36	SegC2-HU5999-Side-N37	13.14%
SegC2-HU5999-Side-N36	SegC2-HU5999-Side-N37	9.47%
SegC2-HU5999-Side-O33	SegC2-HU5999-Side-O34	4.39%
SegC2-HU5999-Side-N26	SegC2-HU5999-Side-N27	8.40%
SegC2-HU5999-Side-N8	SegC2-HU5999-Side-N9	9.59%
SegC2-HU5999-Side-N26	SegC2-HU5999-Side-N27	0.02%
SegC2-SER139-Main-N	SegC2-SER139-Main-N	0.02%

Donor	Acceptor	Occupancy
SegC2-ALA157-Main-N	SegC2-ALA157-Main-N	0.26%
SegC2-LYS136-Side-NZ	SegC2-LYS136-Side-NZ	0.14%
SegC2-SER139-Side-OG	SegC2-SER139-Side-OG	0.38%
SegC2-HU5999-Side-N26	SegC2-HU5999-Side-N27	0.02%
SegC2-GLN41-Side-NE2	SegC2-GLN41-Side-NE3	0.02%

Occupancy (%) of intermolecular hydrogen bonds participated in V36M mutant protease-boceprevir complex

Donor	Acceptor	Occupancy
SegC2-SER139-Side-OG	SegC2-HU5999-Side-O33	3.17%
SegC2-HU5999-Side-N3	SegC2-ALA157-Main-O	28.66%
SegC2-HU5999-Side-N36	SegC2-THR42-Main-O	16.84%
SegC2-HU5999-Side-N8	SegC2-ALA157-Main-O	6.26%
SegC2-HU5999-Side-O33	SegC2-GLN41-Side-OE1	2.04%
SegC2-GLY137-Main-N	SegC2-HU5999-Side-O35	28.50%
SegC2-HU5999-Side-N36	SegC2-GLN41-Side-OE1	3.26%
SegC2-HU5999-Side-O33	SegC2-HIS57-Side-NE2	8.01%
SegC2-ALA157-Main-N	SegC2-HU5999-Side-O15	0.22%
SegC2-GLN41-Side-NE2	SegC2-HU5999-Side-O33	0.41%
SegC2-HU5999-Side-N26	SegC2-HIS57-Side-NE2	15.26%
SegC2-SER139-Side-OG	SegC2-HU5999-Side-O35	0.60%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O25	0.07%
SegC2-HU5999-Side-N36	SegC2-GLN41-Side-NE2	0.02%
SegC2-GLN41-Side-NE2	SegC2-HU5999-Side-N36	0.02%
SegC2-HU5999-Side-N26	SegC2-HIS57-Side-CD2	0.02%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O33	0.05%
SegC2-LYS136-Side-NZ	SegC2-HU5999-Side-O35	0.05%

Salt bridge analysis of wild protease-boceprevir complex

Residue 1	Residue 2	Distance
NH2 ARG A 11	OD2 ASP A 25	2.64
ND1 HIS A 57	OD1 ASP A 81	2.82
ND1 HIS A 57	OD2 ASP A 81	3.34
NH2 ARG A 92	OE1 GLU A 30	3.34
NH1 ARG A 117	OD2 ASP A 103	2.5
NH2 ARG A 117	OD2 ASP A 103	3.46
NH1 ARG A 123	OD1 ASP A 168	2.9
NH1 ARG A 155	OD2 ASP A 168	2.8
NH2 ARG A 155	OD1 ASP A 81	3.55
NH2 ARG A 180	OE1 GLU A 176	2.6

Salt bridge analysis of R155K mutant protease-boceprevir complex

Residue 1	Residue 2	Distance
NH2 ARG A 11	OD2 ASP A 25	2.53
ND1 HIS A 57	OD1 ASP A 81	2.67
NH2 ARG A 92	OE1 GLU A 30	2.78
NH1 ARG A 117	OD1 ASP A 103	2.98
NH1 ARG A 117	OD2 ASP A 103	3.57
NH2 ARG A 123	OD1 ASP A 168	3.14
NH2 ARG A 123	OD2 ASP A 168	2.61
NH2 ARG A 180	OE1 GLU A 176	3.05

Salt bridge analysis of T54S mutant protease-boceprevir complex

Residue 1	Residue 2	Distance
NH2 ARG A 11	OD1 ASP A 25	3.78
NH2 ARG A 11	OD2 ASP A 25	2.46
ND1 HIS A 57	OD1 ASP A 81	3.17
ND1 HIS A 57	OD2 ASP A 81	2.75
NH1 ARG A 118	OD1 ASP A 121	3.68
NH2 ARG A 118	OD1 ASP A 121	2.72
NH2 ARG A 123	OD2 ASP A 168	2.82
NH1 ARG A 155	OD1 ASP A 168	2.71
NH1 ARG A 155	OD2 ASP A 168	3.99

Salt bridge analysis of V36M mutant protease-boceprevir complex

Residue 1	Residue 2	Distance
NH2 ARG A 11	OD2 ASP A 25	2.76
ND1 HIS A 57	OD1 ASP A 81	3.11
ND1 HIS A 57	OD2 ASP A 81	2.91
NH2 ARG A 92	OE1 GLU A 30	3.35
NH2 ARG A 123	OD1 ASP A 168	2.58
NH1 ARG A 155	OD2 ASP A 168	2.79
NH2 ARG A 155	OD2 ASP A 81	3.68
NH2 ARG A 180	OE1 GLU A 176	3.31

Radius of Gyration of each system

Radius of Gyration of each system				
TIME STEP (ns)	**WILD**	**R155K**	**T54S**	**V36M**
0	22.7094	22.57944	22.43857	22.68167
0.504	22.51879	22.65821	22.61689	22.91936
1.0032	22.77356	22.58861	22.68524	23.05055
1.5072	22.72683	22.71842	22.64228	22.91502
2.0016	22.54111	22.77639	22.67449	23.06704
2.5008	22.69844	23.03101	22.6546	22.79641
3.0528	22.65275	23.0112	22.76098	22.8753
4.0032	22.49911	23.0966	22.60288	22.81093
4.5024	22.71118	23.09306	22.72123	23.00095
5.0016	22.90526	23.05007	22.92032	22.89367
5.5008	22.8181	22.85971	22.69523	23.0324
6	22.65925	22.88154	22.8024	23.23935
6.504	22.9134	23.01363	22.91363	22.87237
7.0032	22.98123	22.89516	23.01299	23.13961
7.5024	22.899	22.92311	22.94041	23.05042
8.04	22.8887	22.86222	22.75567	22.95725
9	22.77994	23.07484	22.73773	23.12488
9.504	23.16481	23.11219	22.83189	23.1253
10.0032	23.09709	23.0949	22.87092	23.08258
10.5024	23.02675	23.18482	22.73456	22.94088
11.0016	23.1061	23.3049	22.98108	23.13523
11.5008	23.08751	23.10525	22.81941	23.03297
12	23.05998	23.22722	22.88203	22.98492
12.504	23.15016	23.12727	22.82672	23.02931
13.0032	22.91528	22.94008	22.82237	23.22445
13.5024	23.01046	22.99966	22.66939	23.03788
14.0016	22.94309	23.04117	22.75688	22.7753
14.5008	22.79562	22.84898	22.81326	22.72013
15	22.91129	22.8954	22.82568	22.93764
15.504	23.01501	23.11571	22.73411	22.85169
16.0032	23.01218	22.95449	22.70018	22.93973
16.5024	23.00763	23.10583	22.66458	22.95024
17.0016	22.93584	22.88377	22.64428	23.04354
17.5008	22.79172	23.00103	22.70611	22.9904

Radius of Gyration of each system				
TIME STEP (ns)	**WILD**	**R155K**	**T54S**	**V36M**
18	22.77412	23.17737	22.82826	23.08319
18.504	22.81885	23.11797	22.60514	23.21619
19.008	22.91447	23.05577	22.76848	22.89382
19.5024	23.0058	22.80938	22.72068	23.01838
20	23.1567	22.90856	22.79065	23.00965

10 of XP screened compounds having lower affinity energy (kcal /mol) against wild and mutant proteases

Library ID	WILD	R155K	T54S	V36M
Boceprevir56	-6.40	-8.8	-5.2	-6.3
Boceprevir66	-5.9	-10.0	-7.5	-6.9
Boceprevir53	-5.9	-4.2	-7.9	-6.9
Boceprevir62	-5.7	-5.2	-5.2	-6.6
Boceprevir70	-5.6	-4.8	-4.9	-6.3
Boceprevir480	-5.6	-8.3	-6.0	-6.6
Boceprevir5072	-5.3	-5.4	-6.1	-6.0
Boceprevir6000	-5.3	-4.5	-5.7	-6.1
Boceprevir6023	-5.4	-4.1	-5.7	-7.9
Boceprevir4340	-5.4	-6.2	-4.8	-5.8