Crystal structure of chikungunya virus nsP2 cysteine protease reveals a putative flexible loop blocking its active site

Manju Narwal, Harvijay Singh, Shivendra Pratap, Anjali Malik, Richard J. Kuhn, Pravindra Kumar, Shailly Tomar

PII: S0141-8130(18)30011-4
DOI: doi:10.1016/j.ijbiomac.2018.05.007
Reference: BIOMAC 9608

To appear in:

Received date: 1 January 2018
Revised date: 1 May 2018
Accepted date: 2 May 2018

Please cite this article as: Manju Narwal, Harvijay Singh, Shivendra Pratap, Anjali Malik, Richard J. Kuhn, Pravindra Kumar, Shailly Tomar , Crystal structure of chikungunya virus nsP2 cysteine protease reveals a putative flexible loop blocking its active site. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Biomac(2017), doi:10.1016/j.ijbiomac.2018.05.007

Crystal structure of chikungunya virus nsP2 cysteine protease reveals a putative flexible loop blocking its active site

Manju Narwal[1¶], Harvijay Singh[1¶], Shivendra Pratap[1], Anjali Malik[1], Richard J. Kuhn[2], Pravindra Kumar[1], Shailly Tomar[1*]

[1]Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee-247667, Uttarakhand, India.

[2]Markey Center for Structural Biology and Purdue Institute for Inflammation, Immunology and Infectious Disease, Purdue University, West Lafayette, IN 47907, USA.

[¶]Contributed equally to this article

***Corresponding author: Shailly Tomar**

Email address: shailfbt@iitr.ac.in, shaiprav@gmail.com

Abstract

Chikungunya virus (CHIKV), a mosquito-borne pathogenic alphavirus is a growing public health threat. No vaccines or antiviral drug is currently available in the market for chikungunya treatment. nsP2pro, the viral cysteine protease, carries out an essential function of nonstructural polyprotein processing and forms four nonstructural proteins (nsPs) that makes the replication complex, hence constitute a promising drug target. In this study, crystal structure of nsP2pro has been determined at 2.59 Å, which reveals that the protein consists of two subdomains: N-terminal protease subdomain and C-terminal methyltransferase subdomain. Structural comparison of CHIKV nsP2pro with structures of other alphavirus nsP2 advances that the substrate binding cleft is present at the interface of two subdomains. Additionally, structure insights revealed that access to the active site and substrate binding cleft is blocked by a flexible interdomain loop in CHIKV nsP2pro. This loop contains His548, the catalytic residue, and Trp549 and Asn547, the residues predicted to bind substrate. Interestingly, mutation of Asn547 leads to three-fold increase in K_m confirming that Asn547 plays important role in substrate binding and recognition. This

study presents the detailed molecular analysis and signifies the substrate specificity residues of CHIKV nsP2pro, which will be beneficial for structure-based design and optimization of CHIKV protease inhibitors.

Keywords

Chikungunya virus; cysteine protease; active site blockade; FRET assay; plus-strand RNA virus.

1. Introduction

Chikungunya virus (CHIKV) is a reemerging arthropod-borne virus, which belongs to the *Alphavirus* genus of the *Togaviridae* family. CHIKV was first isolated from the Makonde Plateau of the southern province of Tanzania in 1952 [1], where it was believed to be maintained in a sylvatic transmission cycle among the mosquitoes and non-human primates of the region [2]. CHIKV, which is spread majorly by *Aedes* sp. of mosquitoes, *Aedes aegypti* and *Aedes albopictus*, is an etiological agent of Chikungunya fever in humans. It causes an abrupt onset of fever upon infection, which is characterized by maculopapular rash, incapacitating arthralgia, myalgia, headache and chill [3]. Chikungunya fever is also characterized by high viraemic content with the concomitant irregularities such as intensified lymphopenia and thrombocytopenia [1]. Despite its first emergence in the early 1950s, CHIKV reemerged in the mid of first decade of twenty-first century in several islands of the Indian Ocean where epidemic of an unprecedented magnitude broke out. These outbreaks were found to be associated with an unexpected mutation (A226V) in the E1 structural protein of the virus, which enhanced the viral infectivity in one of its major carrier mosquito *A. albopictus* [2]. Indian Ocean CHIKV outbreak was succeeded by many outbreaks in various geographic locations in Indian subcontinent and Caribbean Sea Islands [4]. A large number of patients were infected from CHIKV in Southern regions of India and a minor CHIKV outbreak in several parts of North-East India was reported in 2011 [5, 6]. In 2013 again, large numbers of Chikungunya infections were reported in the various Caribbean islands and French Guiana [4].

CHIKV belongs to the genus alphavirus, which includes other pathogenic viruses like Sindbis virus (SINV), Semliki Forest virus (SFV), Ross River virus (RRV), the western, eastern and Venezuelan equine encephalitis virus (VEEV) etc. These are enveloped, positive-sense ssRNA viruses that are responsible for causing diseases in humans and animals. The cryo-electron microscopy (cryo-EM) structures of alphavirus have revealed that the virus particles are spherical in shape of 65–70 nm in

diameter having icosahedral symmetry and a triangulation number four (T=4) [7]. The genome comprises of an approximately 11.8 kb (+)-sense ssRNA with a 5' cap and a 3' poly (A) tail [8]. Upon viral entry into the host cell, the 5' two-third of the viral genome RNA is directly translated into the nonstructural polyprotein precursor nsP1234, which is further processed by the proteolytic catalysis of virus encoded cysteine protease (nsP2) into separate nonstructural proteins (nsPs) viz. nsP1, nsP2, nsP3 and nsP4. These proteins with the help of host factors contribute to architect the replication complexes in order to replicate the viral genome and transcribe the 26S subgenomic viral RNA. The subgenomic RNA codes for the structural polyprotein containing capsid (CP), E3, E2, 6K & E1 [9, 10].

Each nsP plays a distinct, defined and indispensable role in the replication of viral genome. nsP1 is a membrane associated protein which possesses S-adenosyl-L-methionine (SAM)-dependent methyltransferase (MTase) and guanylyltransferase (GTase) activities to catalyze the viral RNA capping reaction [11-15]. nsP1 is also believed to interact with nsP4 and facilitate the initiation of minus-strand RNA synthesis and target the replication complex to the plasma membrane by loading it to the endoplasmic reticulum [16-21]. nsP2 is a multifunctional protein, which is composed of two domains. The N-terminal domain possesses RNA helicase, RNA dependent 5'triphosphatase and nucleoside triphosphatase activity [22-24]. The C-terminal half of nsP2 is a cysteine protease, which cleaves the nonstructural polyprotein precursor nsP1234 into different nsPs in a specific and sequential manner [25-30]. Besides, a nuclear localization sequence is present in nsP2 that is responsible for partially transporting nsP2 into the nucleus of virus infected host cells [31]. nsP3 is composed of two domains, a highly conserved N-terminal macrodomain and a less conserved C-terminal domain [32,33]. Recent studies have shown that macrodomain of nsP3 possesses ADP-ribosylhydrolase activity which is pivotal for viral life cycle [34, 35]. nsP4, the RNA dependent RNA polymerase is the key enzyme responsible for catalyzing the synthesis of the viral RNAs, and it also possesses terminal adenylyltransferase activity (TATase) activity [36, 37].

Viral cysteine proteases are a class of protease enzymes, which cleave viral polyproteins at specific site through the catalytic mechanism that involves nucleophilic cysteine thiol in the catalytic dyad or triad. CHIKV nsP2pro, the alphavirus cysteine protease belongs to the peptidase family C9 of clan CA, cleaves and processes the nonstructural polyprotein into various nsPs [38]. Active site of the enzyme contains a catalytic dyad composed of a Cys and a His residue. CHIKV nsP2pro has 40 % sequence identity with VEEV nsP2pro (PDB ID: 2HWK) and 44 % sequence identity with SINV nsP2pro (PDB ID: 4GUA). Alphavirus nsP2pro structure revealed that it contains a novel fold that does not structurally resemble any of the cysteine proteases but the relative positioning of the catalytic Cys and His dyad in the active site resembles the position of catalytic residues in the papain-like cysteine proteases [28, 39].

Thus, CHIKV nsP2pro being virus specific is a potential antiviral drug target. This study reports the bacterial expression, purification, enzyme activity, effect of Asn547 mutation on substrate binding and catalysis, crystallization and atomic structure determination of CHIKV nsP2pro (471-791 residue) at 2.59 Å. The enzymatic activity of purified nsP2pro protein was confirmed by using a FRET based proteolytic assay. Crystal structure reveals that it contains an N-terminal subdomain possessing catalytic residues and a C-terminal subdomain that structurally resembles the SAM-dependent MTase domain. Structural analysis revealed the presence of a variable loop barricading the substrate binding cleft present at the interface of the two subdomains. This variable loop contains Trp549 and Asn547 and substrate docking studies for VEEV nsP2pro indicate their role in substrate recognition [27]. Significance of this residue in active site accessibility and modulation by possibly restricting the entry of substrate in the active site has previously been hypothesized by Russo et al., 2010 [27]. In this study, the crystal structure of CHIKV nsP2pro clearly reveals that the flexible loop containing Asn547 restricts the entry of substrate and substitution of Asn547 to Ala leads to a reduction in enzymatic activity, specifying the importance of Asn547 in substrate recognition and binding. Further, the structural comparison of CHIKV nsP2pro with the crystal structures of VEEV and SINV nsP2pro illustrates the molecular specifics of the substrate binding cleft at the subdomain interface.

2. Materials and methods

2.1. Reagents

Restriction enzymes NdeI, XhoI, Dpn I, T4-DNA ligase and phusion polymerase were purchased from NEB (New England Biolabs, Massachusetts, USA). Oligonucleotides for gene cloning and site-directed mutagenesis were obtained from Integrated DNA Technology (Coralville, USA). MiniPrep plasmid isolation kit was used for plasmid isolation (Qiagen, USA). The *E. coli* expression vector pET 28c (Novagen) containing CHIKV nsP2pro gene with N-terminal His tag (6x His) followed by thrombin protease site was used for over-expression of the enzyme. For protein purification, HisTrap HP Ni-Sepharose column and Hi-load 16/60 Superdex-200 size exclusion column were obtained from GE healthcare (LC, United Kingdom). AKTA Prime plus system from GE Healthcare was used for protein purification. Amicon ultra protein concentrators were purchased from Millipore (Darmstadt, Germany). CD spectrophotometer model J-1500 (JASCO, USA) equipped with Peltier thermostat was used for circular dichroism analysis. For crystallization, PEG ION screens were obtained from Hampton Research (Hampton Research Inc. CA, USA) and all other chemicals of analytical grade were obtained from commercial sources. The plasmid pET28c containing CHIKV nsP2pro was taken as template for site directed mutagenesis. For nsP2pro FRET assay fluorogenic oligopeptide substrates were purchased from

Biolinkk (New Delhi, India). Fluorescence signal was monitored in multimode-plate reader, Cytation 3 (BioTek Instruments, Inc. Vermont, USA).

2.2. CHIKV nsP2pro cloning and expression

The DNA fragment encoding CHIKV nsP2pro (residues 471-791) was PCR amplified from CHIKV cDNA (nsP1-nsP3 gene fragments). Oligonucleotides 5'-GATTCT CATATGATACATTCCAAAATAAAGCCAACG-3' (forward) containing *NdeI* restriction endonuclease site and 5'-GATTCTCTCGAGTTAGAAGGCTGCATTCAGTTGATTG-3' (reverse) containing a *Xho I* restriction endonuclease site, were used as primers in PCR reaction. The resultant PCR product was subcloned into *NdeI* and *XhoI* restriction sites of pET28c vector. The ligated product was transformed into freshly prepared *E. coli* DH5α competent cells by $CaCl_2$ derived heat shock method [40]. Kanamycin resistant transformants were selected and grown in LB medium supplemented with 50 µg/ml kanamycin by incubating the culture overnight at 37°C. Overnight 5 ml culture was used for plasmid isolation using a MiniPrep plasmid isolation kit and the clone was confirmed by restriction-enzyme digestion and PCR gene amplification. Further, DNA sequencing in both the directions using T7 promoter and T7 terminator primers was done using isolated plasmid (pET28c-nsP2pro) to confirm CHIKV nsP2pro DNA insert.

For protein expression, *Rossetta* (DE3) strain of *E. coli.* was transformed using pET28c-nsP2pro plasmid. LB broth, supplemented with 50 µg/ml kanamycin and 35 µg/ml chloramphenicol was used for inoculation of *Rossetta* (DE3) containing pET28c-nsP2pro. Culture was grown at 37°C to an optical density of 0.6 at 600 nm (OD_{600}). Then 0.4 mM isopropyl-β-1-thiogalactopyranoside (IPTG) was added for protein induction. Induced culture was further grown for ~12 h at 18°C. The bacterial cells were finally harvested by centrifugation at 6°C and cell pellets were stored at -80 °C for further use.

2.3. Site directed Mutagenesis

Asn547 was mutated to Ala to verify the functional significance of this residue. Additionally, the catalytic dyad residue Cys478 was mutated to Ala to confirm its role and use it as a negative mutant for nsP2pro assay. For making mutations, complementary mutagenic oligonucleotides were designed to introduce alanine at the targeted site; forward primer- 5'-TAAAGCCAACGTTGCTTGGGCTAAG AGCTTGG-3' and reverse primer- 5'-CCAAGCTCTTAGCCCAAGCAACGTTGGCTTTA-3' were used to introduce Cys478Ala mutation while forward primer- 5'- TATTACGCGGATGCGCACTGGGATAA TAGG-3' and reverse primer- 5'- CCTATTATCCCAGTGCGCATCCGCGTAATA-3' were used to introduce Asn547Ala mutation. PCR amplification was carried out using pET28c-CHIKVnsP2pro as a template for making these mutations. Following PCR, the parent plasmid was digested by *DpnI* restriction

enzyme and the digested product was transformed into XL-1 blue competent cells. Kanamycin resistant transformants were selected, picked and grown overnight at 37°C in LB medium supplemented with kanamycin. DNA sequencing of the mutant plasmid was done to confirm mutations. For expression of mutants, protocol similar to the expression of the wild type CHIKV nsP2pro was followed.

2.4. CHIKV nsP2pro Purification

For purification, cell pellet was resuspended in ice cold lysis buffer containing 50 mM Tris, 500 mM NaCl, 5 % glycerol, and 10 mM Imidazole at pH 7.5. After lysis, DNase I was added to a final concentration of 0.01 mg/ml along with 10 mM $MgCl_2$ & 0.5 mM dithiothreitol (DTT). French press (Constant Systems Ltd, Daventry, England) was used for cell disruption and the obtained cell lysate was centrifuged at 13,500 g for 45 min at 6°C to separate supernatant from cell debris. The supernatant was loaded on a pre-equilibrated 5 ml HisTrap column. CHIKV nsP2pro was eluted from the HisTrap affinity column by running a linear gradient of 10-300 mM imidazole in elution buffer containing 50 mM Tris (pH 7.5), 250 mM NaCl, 5 % glycerol and 300 mM imidazole. For analyzing the purity of samples, eluted fractions were run on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The nsP2pro protein containing fractions were pooled, put inside dialysis bag and were dialyzed for ~ 3 hours against the dialysis buffer containing 50 mM Tris (pH 7.5), 5 % Glycerol, 0.5 mM DTT and thrombin (thrombin:protein ratio 1:50). To remove uncleaved His-tag nsP2pro and the His-tag produced during dialysis, sample was reloaded on the pre-equilibrated 5 ml HisTrap column. The unbound fractions containing nsP2pro with His-tag were collected and concentrated for the next purification step. Further, the partially purified nsP2pro sample was loaded on a pre-equilibrated HiLoad Superdex 75 16/600 size-exclusion chromatography column using AKTA purifier system. AKTA was operated at 6°C at a flow rate of 1 ml/min. The size-exclusion column was calibrated using low-molecular-weight protein markers consisting of albumin (66 kDa), ovalbumin (~45 kDa), trypsin (~23 kDa) and lysozyme (~14 kDa) and the standard curve was made for estimating the molecular weight of purified nsP2pro. The homogeneity of gel-filtration eluted fractions was further analyzed on 12% SDS–PAGE. The fractions containing pure protein sample were pooled and concentrated to ~10 mg/ml using an Amicon Ultra-15 concentrator with a cutoff of ~10 kDa. Concentration of protein was measured by UV absorbance spectroscopy at 280 nm with calculated molar extinction coefficient of 5.0880×10^4 M^{-1} cm^{-1} for nsP2pro. The same purification protocol was employed for purification of nsP2pro mutants.

CD measurements of CHIKV nsP2pro (WT) and mutant (Asn547Ala) were performed in 20 mM sodium phosphate buffer (pH 7.6, 100 mM NaCl, 4% glycerol). All spectral measurements were carried out using a quartz cuvette of 0.1 cm path length. All scans were recorded with a scan speed of 50 nm/min. Effect of

temperature on the secondary structure was studied by observing thermal spectra at different temperature intervals (20˚C- 90˚C).

2.5. FRET based proteolytic assay

A FRET based proteolytic assay was used to characterize the protease activity of purified nsP2pro using fluorogenic oligopeptide substrate DABCYL-DRAGG↓YIFSE-EDANS-NH_2 containing 4-(4-dimethylaminophenyl-azo)benzoic acid (DABCYL) and 5-[(2-aminoethyl)amino] naphthalene-1-sulfonic acid (EDANS) at the N-terminus and C-terminus respectively. The amino acid sequence of substrate peptide was derived from the cleavage site present between nsP3-nsP4 (¾) non-structural proteins of CHIKV. The numbering was assigned to the substrate residues in which the P1 and P1′ positions represent the residues at scissile bond [41].

The proteolysis reaction of the substrate peptide by nsP2pro was performed in the reaction buffer 20 mM Bis-Tris-Propane, pH 7.5 using 96-well black microplate (Corning Inc., USA) at room temperature by varying the concentration of substrate peptide. The enzyme reaction was initiated by addition of 1 μM purified protein from the stock of concentrated protein. The fluorogenic peptide substrate was excited using a wavelength of 340 nm and the emission spectrum was monitored at 490 nm in the multimode-plate reader, Cytation 3.The cleavage of peptide substrate lead to dissociation of the FRET pair and decrease in energy transfer that in turn lead to the increase in the fluorescence signal which was monitored over a period of time. The inner filter effect was negligible as the measurements of the fluorescence signal was carried with low concentrations of fluorogenic substrate, where the intensity of fluorescence was linearly proportional to the concentration of fluorescent peptide substrate. All the reactions were conducted in duplicates. Purified CHIKV nsP2pro Cys478Ala mutant was used as a negative control to confirm the protease activity of CHIKV nsP2pro. Chymotrypsin was used as a positive control in the FRET assay. Additionally, a fluorogenic peptide substrate (DABCYL-GAEEWSLAIE-EDANS-NH_2) which does not contain nsP2pro cleavage site was also used as a negative control in nsP2pro enzyme assay.

2.6. Determination of reaction rate and kinetic values

The catalytic activity of nsP2pro was measured by following the hydrolysis of the fluorogenic substrate peptide. The initial reaction velocity was calculated by performing the reaction in duplicate. The initial velocity (V_i) was determined by monitoring the change in the RFU over time. Fluorescence extinction coefficient (FEC) was used to determine the amount of product formed over the time. The curve of RFU, obtained from varied substrate concentrations was plotted as a function of time and

slope of the curve was assigned as FEC. The initial velocity [V_i] at each substrate concentration was plotted against [S]. Data were fit to the Michaelis−Menten equation [$V_0 = (V_{max}[S])/(K_m + [S])$] using Graph pad prism software for the determination of the kinetic parameters including V_{max}, K_m and k_{cat}/K_m values.

2.7. Optimal pH of CHIKV nsP2pro

Effect of pH on the proteolytic activity of nsP2pro was determined by assaying the enzymatic activity at different pH ranging from pH 5.5 to 11.0, using different buffer systems: Sodium cacodylate buffer (pH 5.5 - 6.5), Bis-tris propane buffer (pH 7.0 - 9.5) & CAPS buffer (pH 10.0 – 11.0). The relative fluorescence unit at all different pH conditions were measured and relative protein activity was calculated against each of the pH values. The pH experiments were performed in duplicates.

2.8. CHIKV nsP2pro crystallization

For crystallization purpose commercially available screens i.e Crystal Screen I & II, PEG ion I & II, Index, Salt and Crystal Screen Cryo were purchased from Hampton Research. For manual optimization of conditions, highest purity chemicals were purchased from Sigma Aldrich. The solution were made in high grade nuclease free water and filtered through 0.22 µm filter (from Millipore India). Crystallization reagents were maintained at 4°C. Crystallization of CHIKV nsP2pro was tried by sitting drop vapor diffusion method. The N-terminal 6x-His-tag of the protein was removed by cleavage with thrombin and the purified protein was concentrated to ~8 mg/ml in 25 mM Tris buffer (pH 7.5), 5 % Glycerol, 100 mM NaCl, and 0.5 mM DTT. Initial crystal hits were obtained in a condition containing 1.6 M sodium citrate tribasic dihydrate (pH 6.5) at 4°C. The crystals obtained were small in size and diffracted to low resolution. To improve size and quality of the crystals, manual crystallization trays were setup at 4°C. 1 µl of the purified protein (~8 mg/ml) was mixed with 1 µl of well solution with different sodium citrate molarities and pH. Mountable crystals of CHIKV nsP2pro appeared in 0.6 M sodium citrate tribasic dihydrate (pH 6.5) in a month. Prior to mounting, crystals were soaked for 30 s in cryoprotective solution containing 5 % glycerol, 0.6 M sodium citrate tribasic dihydrate pH 6.0 and flash-frozen in liquid nitrogen.

2.9. Data collection and structure refinement

The diffraction data was collected at 100 K from a single large crystal using a Bruker Microstar copper rotating anode X-ray generator (CuKα wavelength = 1.54 Å). The images were collected on MAR345 image plate detector. The crystal to detector distance was kept 200 mm and images were

collected with exposure time of 10 min and an oscillation width of 1˚ per image. The crystal diffracted to a maximum resolution of 2.59 Å.

For structure solution, diffraction data were indexed, integrated, scaled and merged using the HKL2000 program suite [42]. The initial phases for CHIKV nsP2pro were obtained by molecular replacement with MOLREP of the CCP4 suite [43]. The crystal structure coordinates of CHIKV nsP2pro (PDB ID: 3TRK) were used as a search model for structure solution using molecular replacement method. All ligands and waters coordinates were removed from the search template. This model provided sufficient phase estimates for subsequent model building and yielded a solution with four molecules per asymmetric unit. The reflections within the resolution range 74.59-2.59 Å were selected for refinement. The rigid body refinement was followed by iterative cycles of restrained atomic parameter refinement including TLS refinement with REFMAC5 and PHENIX [44, 45]. The repetitive cycles of model rebuilding based on σA-weighted 2Fo–Fc and Fo–Fc maps were performed by employing COOT [46]. Finally, the water molecules were added to the peaks contoured in the Fo-Fc electron density difference Fourier map at a 3σ cutoff level which was simultaneously satisfying density contoured at 1σ in the 2Fo-2Fc map. The stereochemical properties of the refined structure model of CHIKV nsP2pro were analyzed using the PROCHECK program [47]. Structural analysis of the refined model and the preparation of figures were done using the PyMOL visualization tool [48]. Structural comparison and superimposition of CHIKV nsP2pro with previously determined nsP2pro structures was carried out by using PyMol visualization tool as well as COOT [46, 48].

3. Results

3.1. Purification of nsP2pro

CHIKV nsP2pro was produced recombinantly using bacterial expression system and purified to homogeneity by immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (Fig. 1A). Molecular weight determined from the chromatogram obtained for the size exclusion chromatography indicates that purified CHIKV nsP2pro is monomeric in solution as calculated from the standard graph of protein molecular weight markers run on Superdex 75 gel-filtration column. A single band of ~37 kDa corresponding to the molecular weight of CHIKV nsP2pro was observed on SDS-PAGE. Both the CHIKV nsP2pro mutant (Cys478Ala and Asn547Ala) were purified using the IMAC and size-exclusion chromatograpy. The purity of samples was assessed by analyzing samples on 12% SDS-PAGE (Fig. S1).

The far-UV CD spectroscopy was used to assess whether the site directed mutagenesis has affected the secondary structure elements of nsP2pro. The CD spectra of nsP2pro (WT) and mutant (Asn547Ala) under native conditions (20 mM sodium phosphate pH 7.6, 20˚C - 40˚C) shows two minima bands at 209

nm and 222 nm wavelength which is characteristic to α-helices and the presence of one positive maxima band at 197 nm depicts the presence of β-sheets in nsP2pro (Fig. S2 A & S2B). However, at temperature higher than 40˚C, both nsP2pro wild-type and mutant started to lose secondary structure element. The CD analysis reveals that single residue mutation does not induce any structural change in the protein.

3.2. Trans-cleavage nsP2pro proteolytic assay using fluorogenic peptide

For testing the protease activity of purified protein, a FRET based approach which provides operational convenience and robustness, was chosen. This assay was used to provide real time analysis of proteolytic cleavage of substrate peptide by CHIKV nsP2pro. The *in vitro* protease activity was monitored using the fluorogenic peptide substrate DABCYL-DRAGG↓YIFSE-EDANS-NH_2. Upon the catalytic action of CHIKV nsP2pro, the peptide ¾ gets cleaved into two fragments. This separates the quencher DABCYL from the fluorophore EDANS and leads to reduction in the FRET signal thereby increasing the fluorescence of EDANS at 490 nm. Similar protease assays have been used for alphavirus proteases for high throughput protease inhibitor screening [49, 50, 51]. The proteolytic reaction for nsP2pro was carried out in 100 µl reaction volume containing 20 mM Bis-Tris-propane buffer (pH 7.5) as the enzyme assay buffer in 96-well plate. The efficient cleavage of the substrate peptide ¾ resulted in an increase in the fluorescence signal with time confirming that the purified CHIKV nsP2pro is enzymatically active (Fig. 1B). A fluorogenic peptide substrate (DABCYL-GAEEWSLAIE-EDANS) which does not contain nsP2pro cleavage site was also used as a negative control. No increase in fluorescence with time was observed with this peptide confirming that the observed activity is of CHIKV nsP2pro and also the specificity of purified protein towards the fluorogenic substrate peptide corresponding to the ¾ cleavage site of CHIKVnsP2 (Fig. 1B). Cys478Ala, the active site mutant was used a negative control. CHIKV nsP2pro protease activity was not detected for the mutant further confirming that the observed activity is of the purified CHIKV nsP2pro. Proteolytic activity of Asn547Ala CHIKV nsP2pro mutant was assessed using the above described assay method. RFU plotted as a function of time for Asn547Ala, showed reduction in the enzymatic activity in comparison to that of the wild type CHIKV nsP2pro (Fig. 1C).

3.3. Determination of kinetic parameters

For the calculation of kinetic parameters, V_i of the enzyme at different substrate concentrations was determined. All the values were normalized by subtracting the readings obtained with the same reaction having no enzyme. The fluorescence extinction coefficient (FEC) was calculated to be 198 RFU/µM for the FRET substrate. This FEC value was used to determine the amount of the product formation in each reaction. The kinetic parameters K_m and k_{cat}/K_m for the wild type CHIKV nsP2pro were

calculated to be 10.17±1.24 μM and 4.08×10^2 M^{-1} sec^{-1} respectively. Active site mutant Cys478Ala showed insignificant increase in fluorescence, confirming indispensable role of cysteine as an active site residue (Fig. 1B). Proteolytic activity assay of Asn547Ala mutant was performed as of the wild type, which resulted in three-fold increase in K_m value. The K_m and k_{cat}/K_m value for Asn547Ala mutant were calculated to be 29.55±3.93 μM and 1.4×10^2 M^{-1} sec^{-1} respectively.

3.4. Optimal pH of CHIKV nsP2pro

To investigate the pH dependence of the reaction of substrate peptide corresponding to the ¾ cleavage site in alphavirus nonstructural polyprotein with CHIKV nsP2pro activity was measured by varying the pH of the enzyme assay buffer ranging from pH 5.5 to pH 11. The graph of fluorescence plotted as a function of buffer pH shows a bell-shaped pH profile which suggest that enzyme is partially active at a pH lower than the physiological pH and maximum activity was observed at pH 7.5. At pH 6.5, the relative enzyme activity reduced up to 43% which further lowered to 17% at pH 6.00 (Fig. 1D). Thus, the optimal pH of CHIKV nsP2pro for efficient cleavage of the substrate peptide corresponding to the ¾ site is 7.0 to 8.0.

3.5. Crystallization and overall structure

Purified and active CHIKV nsP2pro was concentrated to ~ 8 mg/ml and diffraction-quality crystals were obtained using 0.6 M sodium citrate tribasic dihydrate pH 6.0 as precipitant at 4˚C. From a single crystal, a complete data set was collected and structure was determined at 2.59Å. CHIKV nsP2pro crystal belongs to the orthorhombic space group $P2_12_12_1$ with unit cell parameters a=87.04, b=158.96, c=158.88, and $\alpha=\beta=\gamma=90^{\circ}$ with four molecules per asymmetric unit. The estimated Matthews coefficient was 3.76$Å^3$ Da^{-1}, and solvent content was 67.0%. The corresponding refinement statistics are shown in Table 1.

The three dimensional structure of CHIKV nsP2pro was determined using molecular replacement method by taking the unreported crystal structure of the single subunit of CHIKV nsP2pro as a search model (PDB ID: 3TRK). The final model was refined to a final R factor value of 21.0 % and R*free* value of 24.20 % with more than 90 % of the residues in the most favored region of the Ramachandran plot. CHIKV nsP2pro is observed to have four molecules in the asymmetric unit in the crystal form. The final model of CHIKV nsP2pro contained 1284 amino acid residues, 10 glycerol molecules and 342 H_2O molecules. A total of 10 glycerol molecules are found bound at certain sites in the final refined structure of CHIKV nsP2pro. This implicates solvent accessibility and less rigidity of CHIKV nsP2pro allowing the entry and binding of glycerol molecules at various sites including the substrate binding site S2.

3.6. Monomeric structure

The overall structure of each monomer of CHIKV nsP2pro is similar to the crystal structures of nsp2pro reported from VEEV and SINV [28, 39] but detail comparative analysis indicates structural variation in the active site and the substrate binding groove. Alphavirus nsP2 characteristically consists of a papain-like cysteine protease subdomain that is connected to the C-terminal SAM-dependent MTase-like subdomain [28]. Each monomeric chain of CHIKV nsP2pro structure consists of 321 residues starting from Phe471 to Gly791.

The N-terminal protease subdomain, which is apparently smaller in size consists of 123 residues from Phe471 to Glu593, is dominantly helical and contains 7 α-helices and 4 β-strands. Whereas the C-terminal subdomain is comparatively larger in size and extends from Glu625 to Gly791. It contains 5 α-helices and 7 β-strands aligned with each other (Fig. 2). Both the domains are connected through a random coil spanning from Asp594 to Met624. The protease subdomain contains both the catalytic dyad residues Cys478 and His548, which are primarily located near the subdomain interface. The catalytic residue Cys478 is positioned at the N-terminus of α1-helix, and the second catalytic residue His548 along with a completely conserved Trp549 is located in the loop region between β1 and β2-strands. Structural alignment with DALI program identified a papain-like protease MERS PLpro (PDB ID: 4PT5) that showed low similarity with the CHIKV nsP2pro with a Z score of 2.9 and 7 % identity. In addition, secondary structure composition and topological arrangement of CHIKV nsP2pro is substantially different from other cysteine protease structures deposited in the PDB. Thus, proteases of the cysteine proteinase superfamily (SCOP) are observed to have very less similarity with CHIKV nsP2pro. This indicates that CHIKV nsP2pro and nsP2pro from other members of the alphavirus genus contains a novel cysteine protease fold.

The C-terminal MTase-like subdomain of CHIKV nsP2pro extends from Arg606 to Gly791. Tertiary structural comparison indicates that this domain is structurally identical to the proteins that belong to the SAM-dependent MTase superfamily of enzymes. DALI program has identified many MTases structurally similar to the MTase-like subdomain of nsP2pro, notably, ribosomal RNA large subunit MTase (PDB ID: 3DOU) with a Z score of 14.1, FtsJ MTase (PDB ID: 1EJ0) with a Z value 13.5 and MTase (PDB ID: 2OYO) with a Z value 11.3. Sequence similarity of CHIKV nsP2pro MTase-like subdomain is 21 % with the heat shock protein FtsJ RNA MTase of *E. coli* and 19 % with the NS5 MTase of dengue [28]. However, the biological function of MTase-like subdomain of alphavirus nsP2pro is vaguely understood.

3.7. Catalytic domain and active site

Crystal structure of CHIKV nsP2pro reveals that the catalytic dyad residues Cys478, which is positioned at the N-terminal of a α1-helix and His548, present in the loop between β1 and β2-strands are contributed by the protease subdomain (Fig. 2). The active site comprising of the catalytic dyad Cys478 and His548 is formed at the interface of the two subdomains. Structural comparisons of the active site residues of CHIKV nsP2pro with various cysteine proteases indicate that despite the difference in the fold, the orientation of Cys478 and His548 is similar to the catalytic dyad of papain (PDB ID: 1PPN). However, comparison of the catalytic dyad residues with VEEV nsP2pro (PDB IDs: 2HWK and 5EZQ) shows positioning of these residues to be slightly different from the conformation observed in CHIKV (Fig. 3A, 3B, 3C, 3D & 3E). The distance between the catalytic dyad residues is less in the crystal structure of CHIKV nsP2pro as compared to that of VEEV nsP2pro (PDB IDs: 2HWK and 5EZQ). Structural comparison of crystal structure of native nsP2pro from VEEV (PDB IDs: 2HWK and 5EZQ) with the VEEV nsP2pro in complex with peptide-like E-64d (PDB ID: 5EZS) showed that the conformation of catalytic His546 is dependent on the availability of substrate peptide in the active site because His546 flips inward & nearer to Cys477 in the VEEV nsP2pro E-64d inhibitor adduct (Fig. 3F).

Three residues immediately after each of the catalytic dyad residues are observed to be conserved in alphavirus nsP2pro sequence (Fig. S3). These active site motifs, $_{478}$CWAK/R$_{481}$ and $_{548}$HWDN$_{551}$ are specific to nsP2pro cysteine protease of alphaviruses (Fig. 4A & 4B). The bulky residue, Trp549 of $_{548}$HWDN$_{551}$ motif is part of the glycine specificity motif (GSM), which is essential for binding to the substrate cleavage site in numerous cysteine proteases including alphavirus nsP2 having a completely conserved Gly residue at the P2 position in substrate peptide [28]. Structural comparison of VEEV nsP2pro structure shows that the S2 subsite is defined by a conserved Trp549 in CHIKV nsP2pro. Additionally, in CHIKV nsP2pro (PDB ID: 4ZTB), a glycerol molecule is present in the active site and it makes an H-bond with Trp549 (Fig. 4C). This interaction of the glycerol molecule with Trp549 in the active site seems to be consequently positioning His548 in an orientation similar to His546 of VEEV nsP2pro-E-64d complex (5EZS). Henceforth, the orientation of His548 in CHIKV nsP2pro structure is proposed to be appropriate in assisting and positioning the substrate peptide for the nucleophilic attack by Cys478. To ascertain the effect of glycerol on the function in CHIKV nsP2pro, the FRET based proteolytic assay for purified protein was carried out in the presence of varying concentration of glycerol. As expected concentration of glycerol higher than 10% v/v, was found to be impeding nsP2pro protease activity supporting the above observation (Data not shown).

3.8. Substrate binding cleft

Crystal structure of alphaviral nsP2pro in complex with the substrate peptide is not available till date. Therefore, the structural and functional aspect of the substrate binding region of alphavirus nsP2pro is based on the results obtained by docking of substrates into the active site of nsP2pro protein structure [28, 52]. Crystal structure of VEEV nsP2pro and substrate docking studies have revealed that the substrate binding pocket of nsP2pro, which is a deep pronounced groove traversing the active site can accommodate a five residue long peptide [27, 28]. Structural analysis of the three dimensional structure of CHIKV nsP2pro reveals that substrate binding site is present beneath the variable loop connecting β1 and β2 strands at the interface of the protease subdomain and the MTase-like subdomain (Fig. 2) and the active site region is observed as a closed tunnel. The five residue junction region P4-P1' (^{P4}RAGGY$^{P1'}$) corresponding to the ¾ cleavage site of nsP2pro docked into the active site of VEEV nsP2pro identified the substrate binding residues [26]. The substrate binding residues near the active site in CHIKV nsP2pro were identified by its structural comparison with VEEV nsP2pro structure. Table 2 compares the residues that participate in the substrate binding of CHIKV nsP2pro with that of VEEV nsP2pro. The N-terminal protease subdomain contributes the catalytic and the major substrate binding residues (Asn476, Cys478, Asn547, and Tyr544 etc) while the MTase-like subdomain contributes only the substrate binding residue (Met707, Asp711 etc). This observation indicates that residues from the MTase-like subdomain of CHIKV nsP2pro do not necessarily participate in the catalytic machinery but do play a pivotal role in substrate recognition and binding.

3.9. Active site blockade by flexible loop

A surface view around the active site of nsP2pro in CHIKV crystal structure reveals that the loop present between β1 and β2-strands that possesses His548, the catalytic residue along with the loop present between β7-strand and α9-helix of the MTase-like subdomain are closing the access to the active site and thus, acting like a gate for substrate entry (Fig. 5A, 5B). Interestingly, the active site is in open conformation in the crystal structures of all other alphavirus (VEEV and SINV) nsP2pro determined till date. In the crystal structure of CHIKV nsP2pro a close conformation of the loop is observed and detailed structural analysis revealed that this loop is closer to the MTase-like domain hinting towards its regulatory role in substrate entry. Additionally, this dynamic loop possesses a conserved residue, Asn547 which is at a distance of 4.1 Å and 5.1 Å from Leu670 of the MTase-like domain in 4ZTB and 3TRK respectively (Fig. 5C). This 4.1 Å distance in 4ZTB is appropriate for making van der Waals interaction. The positioning of variable loop as well as Cα carbon of loop residue Asn 547 is essentially the same in the previously determined but not reported crystal structure of CHIKV nsP2pro with PDB ID 3TRK Structural comparisons of alphavirus nsP2pro crystal structures reveal high conformational variability in the loop between β1 and β2-strands that contains Asn547 residue. Additionally, in SINV nsP2pro this

loop is much longer than VEEV and CHIKV as SINV nsP2 contains seven extra residues in this loop (Fig. 3C & Fig. S3). The RMSD for residues corresponding to active site variable loop (544 to 549 in 4ZTB) were calculated for all the structures. The loop showed deviation among all the structures and with reference to 4ZTB, it was found to be 0.32 Å, 1.33 Å, 1.48 Å and 2.08 Å for 3TRK, 5EZS, 2HWK and 4GUA structures respectively. Reduced distance between Asn547 and Leu670 in crystal structures of CHIKV nsP2pro structures (PDB IDs: 4ZTB & 3TRK) and the observed deviation of this loop in different structures points towards the conformational flexibility of the loop between β1 and β2-strands. Furthermore, the temperature factor (B factor) analysis shows that the loop present between β1 and β2-strands has elevated B factor in the crystal structure of CHIKV nsP2pro (PDB ID: 4ZTB). This conformational variability and flexibility of the loop containing catalytic residue His548 in CHIKV nsP2pro specifies its role in regulating the entry of substrate in the active site of CHIKV nsP2pro.

Moreover, Asn547 of CHIKV nsP2pro, which is found to be sequentially conserved as Asn545 in VEEV nsP2pro and partially conserved in other alphaviruses (~56% identity), is proposed to be a critical substrate binding residue [27]. Structural insights into VEEV nsP2pro-E64d inhibitor complex and docking of its substrate peptide in the active site of VEEV nsP2pro (PDB ID: 2HWK) showed that Asn545 plays a critical role in regulating the binding of substrate peptide at the active site [27, 53]. Molecular docking study of CHIKV nsP2pro has also revealed the significance of Asn547 as a substrate binding residue which makes hydrogen bond to the P4 Arg of the substrate peptide through backbone amide. Thus, it can be put forth that Asn547 possibly plays a central role in regulating the entry of substrate in the active site based on the conformational variability of the loop containing this residue in the structure of CHIKV nsP2pro. Furthermore, Leu670 present in the MTase-like subdomain loop, that also contributes to the active site barricading is highly conserved (~81 %) in alphaviruses (Fig. S3). Similarly, it has been hypothesized for VEEV nsP2pro that the entry of substrate is a regulated function and is controlled by the residues Asn544 and Asn545 present in the dynamic β-hairpin of the protease subdomain and Leu665 (corresponding to CHIKV nsP2pro Leu670) from the loop of MTase-like subdomain [27].

3.10. Effect of Asn547Ala and Cys478Ala mutation

Alphavirus nsP2pro is a cystiene protease, which contains a catalytic dyad; cysteine and histidine in its active site. Crystal structure analysis of CHIKV nsP2pro has pointed towards the structural flexibility of the loop present between β1 and β2-strands. This loop is highly dynamic in nature as evident by its elevated B factor. The loop contains catalytic residue His548 as well as Asn547 and Trp549, the residues predicted to bind the incoming substrate. Previous protein-peptide docking and MD simulation

studies of VEEV nsP2pro has asserted on the significance of Asn547 as a substrate binding residue, which stabilizes the enzyme-substrate complex by making a backbone to backbone interaction to one of the substrate subsite. Further VEEV nsP2pro studies have advanced that the mutation of this residue results in minor change in the K_m and K_{cat} of the enzyme [53]. To investigate the functional role of Asn547 in substrate binding in CHIKV nsP2pro, the site-directed mutagenesis was performed and Asn547 was replaced with Ala. Kinetic analysis of this mutant with substrate peptide ¾ was assessed by FRET based proteolytic assay. The Asn547Ala mutation led to a three-fold increase in $K_{m.}$ and reduction in $K_{cat}/K_{m.}$(Fig 6). This drop in catalytic efficiency hints toward the functional substrate specificity role of Asn547 in CHIKV nsP2pro. Active site mutant of CHIKV nsP2pro (Cys478Ala) completely abolished the proteolytic activity of the enzyme.

4. Discussion

CHIKV is an important arthropod borne virus with substantial impact on global health and cause a major viral disease that requires the development of antiviral drug to combat Chikungunya disease. Virus specific proteases have particularly become an attractive drug targets for viral diseases, ever since the potent protease inhibitor specific to HIV-1 protease have been licensed and protease inhibitor for hepatitis C virus have entered the clinical trials [54, 55]. The C-terminal domain of nsP2 (nsP2pro) is a papain-like cysteine protease that plays a key role in processing of the viral non-structural polyprotein inside the infected host cell and leads to the formation of individual virus replication proteins (nsP1, nsP2, nsP3 and nsP4) [25, 52]. Till date, no drug is commercially available for the treatment of Chikungunya infection. CHIKV specific viral enzymes, including two of the proteases the capsid protease and the nsP2pro, are potential target for antiviral drug development and discovery [47, 56]. In this study we have described the detailed crystal structure analysis of CHIKV nsP2pro. Our investigation gained insights into conformational variability of the active site and the understanding of molecular determinants of CHIKV nsP2pro enzymatic activity and substrate specificity.

Crystal structure CHIKV nsP2pro is observed to have four molecules in the asymmetric unit and each subunit consists of two subdomains; the N-terminal protease subdomain and the C-terminal MTase-like subdomain. The catalytic pocket of the enzyme is positioned at interface of the protease and the MTase-like subdomains. The surface view of the crystal structure of CHIKV nsP2pro revealed that the active site at the interface is closed because of high flexibility of the loop present between β1 and β2 strands of the protease subdomains. This loop between β1 and β2 strands contains Asn547 residue that has high conformational variability and elasticity, due to which Asn547 seems to be regulating the access

of the active site by specifically binding to the substrate subsite. Previous molecular docking and simulation studies of VEEV nsP2pro with the substrate peptides showed that this β-hairpin might restricts the entry of the substrate into the active site. This observation was based on the highly dynamic nature of the β-hairpin between β1 and β2 strands and the interaction of the β-hairpin residue, Asn545 with the docked peptide in VEEV nsP2pro [27].

Structural comparison analysis of the active site cleft of CHIKV nsP2pro with that of VEEV nsP2pro revealed the substrate binding position and has led to the identification of the active site groove residues that participate in the substrate binding (Table 2). The prefixes P and S refer to the substrate peptide residues and the complementary protease binding sites respectively [41]. In CHIKV nsP2pro structure these sites appear as shallow depression in a deep long groove at the interface of the protease and the MTase-like subdomains (Fig. 5A, 5B). There are number of residues in the substrate binding cleft along with the catalytic His548 and Cys478 that contribute to each of these subsites (Table 2). It has been observed that most of the residues are conserved in these subsites in the alphaviruses (Fig. S3). The residues of S1' and S1 subsites of CHIKV nsP2pro are exactly same as in VEEV nsP2pro protease while few differences are observed in the residues of the S2, S3 and S4 subsites in two structures [27, 28].

Three dimensional structural comparison of CHIKV nsP2pro with VEEV nsP2pro reveals that the P1 residue of substrate binds to the S1 subsite of nsP2pro, which is formed by catalytic Cys478 and nearby residues Asn476, Val477and Trp479. The S2 subsite is defined by Trp549, a highly conserved residue in alphaviruses along with Asn547, Trp479 and Tyr512. The conservation of Trp549 in alphavirus nsP2pro and its interaction with Gly at P2 position makes CHIKV nsP2pro a GSM specific cysteine protease [57]. The P3 Ala of CHIKV substrate peptide binds to S3 subsite residue of nsp2pro which are Asn547, Tyr544 and Met707, while the S4 subsite is defined by Trp549, Gln706 and Asp711. Structural analysis revealed that the access to the active site and substrate binding groove is blocked by a dynamic loop present in the protease subdomain between the β1 and β2 strands. This loop contains a substrate binding residue Asn547, side chain of which is flexible and open to attain more than one conformation, as evident by its high B factor. Previously concluded molecular docking and biochemical studies have revealed the significance of Asn547 as a substrate binding residue in VEEV nsP2pro and CHIKV nsP2pro. In the crystal structure of CHIKV nsP2pro this residue is found coming towards Leu670 of the MTase-like domain and closing the access of active site. To examine the role of this residue in substrate recognition and binding Asn547 was mutated to Ala which results in a three-fold drop in K_{cat}/K_m. This reduction in catalytic efficiency may account for the specificity of Asn547 as a substrate binding residue, which is involved not only in main chain interactions with the substrate but it's side chain might also participate in binding interaction to the substrate subsites. These observations reflect that conformational flexibility of the loop connecting β1 and β2 strands might be responsible for regulating the access to the

active site residue Cys478, which lies just beneath this variable loop in the crystal structure of CHIKV nsP2pro.

Moreover, recent structural studies of VEEV nsP2pro in complex with E-64d revealed an important molecular interaction at the interface of two subdomains. In this structure, Asn475 which is in polar contact with carbonyl oxygen of Asp507 of the protease subdomain makes an H-bond with Arg662 of the MTase-like subdomain and aligns itself to assist the nucleophilic attack by the catalytic cysteine (PDB ID: 5EZS) and thus proposed to act like a transition state stabilizing residue [53] (Fig. 7A). In CHIKV nsP2pro, Asn476 is homologous toAsn475 of VEEV and is found to make polar contact with the backbone carbonyl oxygen of Asp509 (Fig. 7B & Fig. S3). Sequence alignment of CHIKV nsP2pro with VEEV revealed that Arg662 of VEEV nsP2pro is substituted by Thr665 in CHIKV nsP2pro. However, an adjacent residue Glu669 seems to be playing a role similar to Arg662 of VEEV nsP2pro, as Glu669 is positioned in such a way that it makes molecular contact with Asn476 in CHIKV nsP2pro structure (Fig. 7C). This structural analysis suggests that Asn475 might potentially play a role of a transition state stabilizing residue in CHIKV nsp2pro. This Asn residue is also found in other viral papain-like cysteine proteases like SARS PLpro, MERS PLpro and corona viral protease, where its function has been established as the transition state stabilizing residue [58-60].The molecular interactions at the interface of the two subdomains near the active site specify the regulatory role of interface residues in the substrate binding and enzymatic activity of alphavirus nsP2pro. Additionally, Trp549, a highly conserved S2 subsite residue of the alphavirus nsP2pro is located in a shallow depression at the interface between the two subdomains and makes significant contribution in substrate binding. In CHIKV nsP2 crystal structure, the side chain of Trp549 is H-bonded to a glycerol molecule present near the active site at the substrate binding subsites P2/P3, which might mimic the molecular interaction with that of actual substrate (Fig. 4C).

These observations prophesies the conformational flexibility of substrate binding region. The flexibility of this loop close to the main catalytic residue His548 with low sequence conservation indicates its functional role and significance in accommodating three different virus specific substrate peptides with sequence variations.

5. Conclusion

The crystal structure analysis of CHIKV nsP2pro shows that the access of the active site to substrate peptide is blocked by a flexible loop. The FRET based proteolytic assay performed in this study using recombinantly purified enzyme with proper controls demonstrates that enzyme is active on its fluorogenic cognate substrate peptide at physiological pH and activity reduced significantly with the

decreasing pH. Structure analysis of the dynamic loop present between β1 and β2 strands revealed the presence of Asn547 which possibly makes interaction with the incoming substrate peptide and also regulates the active site accessibility. This observation was further corroborated by mutating of Asn547 which led to a significant reduction in catalytic efficiency. The presence of Asn547 along with catalytic residue His548 and the conserved Trp549in the variable loop leads to the hypothesis that this loop of CHIKV nsP2pro influences the binding of substrate peptides and regulates the entry of the substrate in the active site.

Structural insight into the active site, the substrate binding subsites and the conformational flexibility of the variable loop specify structural determinants at the interface of the two subdomains, which can be exploited for structure-based rational drug design for discovery of nsP2pro inhibitory compounds as antivirals.

Availability of supporting data

The atomic coordinates and structure factor amplitudes are available in the Protein Data Bank repository (http://www.rcsb.org) with accession number 4ZTB.

Acknowledgement

Authors thank the Macromolecular Crystallographic Facility (MCU) at IIC, Indian Institute of Technology, Roorkee, India. MN & AM thanks University Grant Commission (UGC). HS & SP thanks the Ministry of Human Resource Development (MHRD) Govt. of India for senior research fellowships (SRF).

Competing interests

The authors declare that they have no competing interests.

Funding

This project was financed by Department of Biotechnology (DBT) Government of India (Grant ref. number BT/PR9670/MED/29/807/2013).

Author's Contributions

ST, PK and RJK conceived the work and evaluated the results. MN, HS & AM conducted the research. HS, MN & SP performed structural analysis & investigation. HS wrote the original draft of manuscript.

ST supervised the work and edited the paper. PK & RJK reviewed the final manuscript. All authors have read and approved the final manuscript.

References

1. S.D. Thiberville, N. Moyen, L. Dupuis-Maguiraga, A. Nougairede, E.A. Gould, P.Roques, X. de Lamballerie, Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy, Antiviral Res. 99 (2013) 345–370.

2. C. Saisawang, P. Sillapee, K. Sinsirimongkol, S. Ubol, D.R. Smith, A.J. Ketterman, Full length and protease domain activity of chikungunya virus nsP2 differ from other alphavirus nsP2 proteases in recognition of small peptide substrates, Biosci. Rep. 35 (2015) 196.

3. B.A. Pastorino, C.N. Peyrefitte, L. Almeras, M. Grandadam, D. Rolland, H.J. Tolou, M. Bessaud, Expression and biochemical characterization of nsP2 cysteine protease of Chikungunya virus, Virus. Res. 131 (2008) 293–298.

4. W.V. Bortel, F. Dorleans, J. Rosine, A. Blateau, D. Rousset, S. Matheus, I. Leparc-Goffart, O. Flusin, C. Prat, R. Cesaire, F. Najioullah, Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe, Euro. Surveill. 19 (2014) 20759.

5. P. Ray, V.H. Ratagiri, S.K. Kabra, R. Lodha, S. Sharma, B.S. Sharma, M. Kalaivani, N. Wig, Chikungunya Infection in India: Results of a Prospective Hospital Based Multi-Centric Study, PLoS ONE 7 (2012) e30025.

6. S.A. Khan, P. Dutta, R. Topno, J. Borah, P. Chowdhury, J. Mahanta, Chikungunya outbreak in Garo Hills, Meghalaya: an epidemiological perspective, Indian. J. Med. Res. 14 (2015) 591-597.

7. J. Jose, J.S. Snyder, R.J. Kuhn, A structural and functional perspective of alphavirus replication and assembly, Future Microbiol. 4 (2009) 837-856.

8. A.H. Khan, K. Morita, M.M.C. Parquet, F. Haseben, E.G. Mathenge, A. Igarashi, Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site, J. Gen. Virol. 83 (2002) 3075–3084.

9. E.G. Strauss, C.M. Rice, J.H. Strauss, Complete nucleotide sequence of the genomic RNA of Sindbis virus, Virology 133 (1984) 92–110.

10. E.G. Strauss, C.M. Rice, J.H. Strauss, Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon, Proc. Natl. Acad. Sci. U S A 80 (1983) 5271–5275.

11. T. Ahola, L. Kaariainen, Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl- GMP, Proc. Natl. Acad. Sci. U S A 92 (1995) 507–511.

12. S. Mi, V. Stollar, Expression of Sindbis virus nsP1 and methyltransferase activity in Escherichia coli, Virology 184 (1991) 423–427.

13. R.K. Cross, Identification of a unique guanine-7-methyltransferase in Semliki Forest virus (SFV) infected cell extracts, Virology 130 (1983) 452–463.

14. H.L. Wang, J. O'Rear, V. Stollar, Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity, Virology 217 (1996) 527–531.

15. S.T. Tomar, M. Narwal, E. Harms, J.L. Smith, R.J. Kuhn, Heterologous Production, Purification and Characterization of Enzymatically Active Sindbis Virus Nonstructural protein nsP1, Protein. Expression. Pur. 79 (2011) 277-284.

16. C.L. Fata, S.G. Sawicki, D.L. Sawicki, Alphavirus minus-strand RNA synthesis: identification of a role for Arg183 of the nsP4 polymerase, J. Virol. 76(2002) 8632–8640.

17. C.L. Fata, S.G. Sawicki, D.L. Sawicki, Modification of Asn374 of nsP1 suppresses a Sindbis virus nsP4 minus-strand polymerase mutant, J. Virol. 76 (2002) 8641–8649.

18. A. Salonen, L. Vasiljeva, A. Merits, J. Magden, E. Jokitalo, L. Kaariainen, Properly folded nonstructural polyprotein directs the Semliki Forest virus replication complex to the endosomal compartment, J. Virol. 77 (2003) 1691–1702.

19. S.G. Sawicki, D.L. Sawicki, L. Kaariainen, S. Keranen, A Sindbis virus mutant temperature-sensitive in the regulation of minus-strand RNA synthesis, Virology 115 (1981)1 61–172.

20. Y. Shirako, J.H. Strauss, Requirement for an aromatic amino acid or histidine at the N terminus of Sindbis virus RNA polymerase, J. Virol. 72 (1998) 2310–2315.

21. Y.F. Wang, S.G. Sawicki, D.L. Sawicki, Sindbis virus nsP1 functions in negative-strand RNA synthesis, J. Virol. 65 (1991)985–988.

22. M. Rikkonen, J. Peranen, L. Kaariainen, ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2, J. Virol. 68 (1994) 5804–5810.

23. L. Vasiljeva, A. Merits, P. Auvinen, L. Kaariainen, Identification of a novel function of the alphavirus capping apparatus: RNA 5-triphosphatase activity of Nsp2, J. Biol. Chem. 275 (2000) 17281–17287.

24. G.M. de Cedron, N. Ehsani, M.L. Mikkola, J.A. Garcia, L. Kaëaëriaëinen, RNA helicase activity of Semliki Forest virus replicase protein NSP2, FEBS Lett. 448 (1999) 19-22.

25. L. Vasiljeva, L. Valmu, L. Kääriäinen, A. Merits, Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2, J. Biol. Chem. 276 (2001) 30786-30793.

26. Mayuri, T.W. Geders, J.L. Smith, R.J. Kuhn, Role for Conserved Residues of Sindbis Virus Nonstructural Protein 2 Methyltransferase-Like Domain in Regulation of Minus-Strand Synthesis and Development of Cytopathic Infection, J. Virol. 82 (2008)7284.

27. A.T. Russo, R.D. Malmstrom, M.A. White, S.J. Watowich, Structural basis for substrate specificity of alphavirus nsP2 proteases, Journal. Mol. Graph. Modell. 29 (2010) 46-53.

28. A.T. Russo, M.A. White, S.A.Watowich, The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease, Structure 14 (2006) 1449–1458.

29. N. Garmashova, R. Gorchakov, E. Frolova, I. Frolov, Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription, J. Virol. 80 (2006) 5686-96.

30. N. Garmashova, R. Gorchakov, E. Volkova, S. Paessler, E. Frolova, I. Frolov, The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff, J. Virol. 81 (2007) 2472-84.

31. J. Peranen, M. Rikkonen, P. Liljestrom, L. Kaariainen, Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2, J. Virol. 64 (1990) 1888–1896.

32. G. Li, M.W. La Starza, W.R. Hardy, J.H. Strauss, C.M. Rice, Phosphorylation of Sindbis virus nsP3 in vivo and in vitro, Virology 179 (1990) 416-27.

33. H. Malet, B. Coutard, S. Jamal, H. Dutartre, N. Papageorgiou, M. Neuvonen, T. Ahola, N. Forrester, E.A. Gould, D. Lafitte, F. Ferron, The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket, J. Virol. 83 (2009) 6534-45.

34. L. Eckei, S. Krieg, M. Bütepage, A. Lehmann, A. Gross, B. Lippok, A.R. Grimm, B.M. Kümmerer, G. Rossetti, B. Lüscher, P. Verheugd, The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases, Sci. Rep. 7 (2017) 41746.

35. R.L. McPherson, R. Abraham, E. Sreekumar, S.E. Ong, S.J. Cheng, V.K. Baxter, H.A. Kistemaker, D.V. Filippov, D.E. Griffin, A.K. Leung, ADP ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus replication and virulence, Proc. Natl. Acad. Sci. U S A. 114 (2017) 1666-1671.

36. J.K. Rubach, B.R. Wasik, J.C. Rupp, R.J. Kuhn, R.W. Hardy, J.L. Smith, Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. Virology 384 (2009) 201-8.

37. S. Tomar, R.W. Hardy, J.L. Smith, R.J. Kuhn, Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity, J. Virol. 80 (2006) 9962-9.

38. N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the peptidase database, Nucleic. Acids. Res. 38 (2010) D227-33.

39. G. Shin, S.A. Yost, M.T. Miller, E.J. Elrod, A. Grakoui, J. Marcotrigiano, Structural and functional insights into alphavirus polyprotein processing and pathogenesis, Proc. Natl. Acad. Sci. U S A. 109 (2012) 16534-9.

40. H. Inoue, H. Nojima, H. Okayama, High efficiency transformation of Escherichia coli with plasmids. Gene 96 (1990) 23-8.

41. I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochem. Biophys. Res. Commun. 27 (1967) 157-62.

42. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods. Enzymol. 276 (1997) 307-26.

43. A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Crystallogr. 30 (1997) 1022-5.

44. G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta. Crystallogr. Sect. D: Biol. Crystallogr. 53 (1997) 240-55.

45. P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta. Crystallogr. Sect. D: Biol. Crystallogr. 66 (2010) 213-21.

46. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta. Crystallogr. Sect. D: Biol. Crystallogr. 60 (2004) 2126-32.

47. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.

48. W.L. Delano, The PyMOL molecular graphics system, DeLano Scientific, San Carlos, CA (2008).

49. M. Aggarwa, R. Sharma, P. Kumar, M. Parida, S. Tomar, Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. Sci Rep. 5 (2015) 14753.

50. M. Aggarwal, S. Dhindwal, P. Kumar, R.J. Kuhn, S. Tomar, trans-Protease activity and structural insights into the active form of the alphavirus capsid protease, J. Virol. 88 (2014) 12242-53.

51. H. Singh, R. Mudgal, M. Narwal, R. Kaur, V.A. Singh, A. Malik, M. Chaudhary, S.Tomar, Chikungunya virus inhibition by peptidomimetic inhibitors targeting virus-specific cysteine protease, Biochimie 149 (2018) 51-61.

52. S. Dhindwal, P. Kesari, H. Singh, P. Kumar, S. Tomar, Conformer and pharmacophore based identification of peptidomimetic inhibitors of chikungunya virus nsP2 protease, Journal. Biomol. Structure. Dyn. 2 (2016)1-8.

53. X. Hu, J.R. Compton, D.H. Leary, M.A. Olson, M.S. Lee, J. Cheung, W. Ye, M. Ferrer, N. Southall, A. Jadhav, P.J. Glass, Kinetic, Mutational, and Structural Studies of the Venezuelan Equine Encephalitis Virus Nonstructural Protein 2 Cysteine Protease, Biochemistry 55 (2016) 3007-19.

54. J.T. Hsu, H.C. Wang, G.W. Chen, S.R. Shih, Antiviral drug discovery targeting to viral proteases, Curr. Pharm. Des. 12 (2006)1301-14.

55. D. Lamarre, P.C. Anderson, M. Bailey, P. Beaulieu, G. Bolger, P. Bonneau, M. Bös, D.R. Cameron, M. Cartier, M.G. Cordingley, A.M. Faucher, An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, Nature 426 (2003) 186-9.

56. R, Sharma, B, Fatma, A. Saha, S. Bajpai, S. Sistla, P.K. Dash, M. Parida, P. Kumar, S. Tomar, Inhibition of chikungunya virus by picolinate that targets viral capsid protein, Virology 498 (2016) 265-76.

57. A. Golubtsov, L. Kääriäinen, J. Caldentey, Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis, FEBS lett. 580 (2006) 1502-8.

58. J. Lei, J.R. Mesters, C. Drosten, S. Anemüller, Q. Ma, R. Hilgenfeld , Crystal structure of the papain-like protease of MERS coronavirus reveals unusual, potentially druggable active-site features, Antiviral. Res. 109 (2014) 72-82.

59. K. Ratia, A. Kilianski, Y.M. Baez-Santos, S.C. Baker, A. Mesecar, Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease, PLoS Pathog. 10 (2014) 1004113.

60. K. Ratia, K.S. Saikatendu, B.D. Santarsiero, N. Barretto, S.C.Baker, R.C. Stevens, A.D. Mesecar, Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme, Proc. Natl. Acad. Sci. 103 (2006) 5717-22.

Figure legends

Fig. 1. CHIKV nsP2pro purification and protease assay. (A) The chromatogram of gel-filtration results obtained on loading protein sample on Superdex75 column; a single major peak was obtained at around 75 ml that contained purified protein. SDS-PAGE profile of purified CHIKV nsP2pro is shown as an inset, Lane M: Pre-stained molecular weight markers, Lane nsP2: Purified CHIKV nsP2pro without His-tag. **(B)** The plot of fluorescence values resulting from the substrate peptide cleavage as a function of time for purified CHIKV nsP2pro enzyme. The concentration of nsP2 was kept constant (1 μM) while concentration of the substrate ranged from 5 to 30 μM. The control reactions were; a reaction with 5 μM of fluorognic substrate peptide without addition of purified protein and a reaction using control peptide (5 μM) with fluorogenic properties similar to nsP2pro experimental peptide but not having the nsP2pro specific cleavage sequence. **(C)** Processing of fluorogenic substrate peptide ¾ by wild type CHIKV nsP2pro (WT) and Asn547 mutant of CHIKV nsP2pro (Asn547Ala). The concentration of both WT and mutant (Asn547Ala) was kept constant (1 μM) while substrate concentration was varied from 1 μM to 10 μM. The fluorescence signal resulting from the cleavage of substrate peptide is shown as a function of time. **(D)** Effect of pH ranging from 4.5 to 9.5 on the proteolytic activity of nsP2pro. The relative activity was calculated at different pH by taking the activity at pH 7.5 as 100%.

Fig. 2. Crystal structure of CHIKV nsP2pro. The N-terminal protease subdomain is depicted in magenta color containing both the catalytic dyad Cys478 and His548 colored by atom types. The C-terminal MTase-like subdomain is colored in cyan and the linker connecting the two subdomains is shown in blue color. Variable loop in protease domain designated as 'loop between β1 and β2 strands' possess catalytic His548 and substrate binding residue Trp549 and Asn547.

Fig. 3. Structural comparison of the active site residues of CHIKV nsP2pro with other alphaviral nsP2pro. CHIKV nsP2pro (4ZTB) is depicted by magenta color and numbering of the residues is done according to respective PDB IDs. **(A)** Superimposition of CHIKV nsP2pro structures (4ZTB with 3TRK). **(B)** Superimposition of CHIKV nsP2pro (4ZTB) with papain (PDB ID: 1PPN). **(C)** Superimposition of CHIKV nsP2pro (4ZTB) with SINV nsP2pro (4GUA). Length of variable loop in SINV nsP2pro is higher than that of CHIKV nsP2pro. **(D)** Superimposition of CHIKV nsP2pro (4ZTB) with VEEV nsP2pro (2HWK). **(E)** Superimposition of CHIKV nsP2pro (4ZTB) with VEEV nsP2pro (5EZQ). **(F)** CHIKV nsP2pro (4ZTB) superimposed with VEEV nsP2pro-E-64d complex (5EZS).

Fig. 4. Active site motif residue comperison and glycerol binding. **(A)** Conserved residues of active site motif $_{478}$CWAK$_{481}$ and $_{548}$HWDN$_{551}$ of CHIKV nsP2pro are superimposed with VEEV nsP2pro. Green color represents CHIKV nsP2pro (4ZTB), magenta color represent VEEV nsP2pro (2HWK) and blue color represent VEEV nsP2pro (5EZS). **(B)** Conserved residues of active site motif $_{478}$CWAK$_{481}$ and $_{548}$HWDN$_{551}$ of CHIKV nsP2pro are superimposed with SINV nsP2pro. Green color represents CHIKV nsP2pro (4ZTB) and blue color represent SINV nsP2pro (4GUA). **(C)** CHIKV nsP2pro (4ZTB) Trp549 is H-bonded to the glycerol molecule (GOL 4) mimicking the molecular interaction of substrate peptide with substrate subsite S2. Catalytic residue Cys478 and His548 are also shown in magenta stick.

Fig. 5. CHIKV nsP2pro surface view and structural comparison of subdomain interface. **(A)** CHIKV nsP2pro monomer (4ZTB) showing various subsites represented with different colors. Subsite S1′ is represented by orange, subsite S1 is represented by green and subsite S2 is represented by red color. Zoom panel shows positioning of the active site residues with respect to the gate closing the peptide substrate entry into the groove present at the interface of the two subdomains. **(B)** Back view of the tunnel showing remaining substrate binding subsites S3 and S4; Yellow represents subsite S3 and blue represents subsite S4. **(C)** CHIKV nsP2pro zoom in view of the variable loop present between β1 and β2 stands showing Asn547 of protease subdomain and Leu670 of the MTase like subdomain along with molecular bond length between both the residues. A 3σ |Fo|-|Fc| omit map for the interface region is shown as a blue wire cage. Interface of CHIKV nsP2pro is superimposed from two different PDBs (4ZTB and 3TRK). Asn547 of variable loop from 4ZTB and 3TRK is shown by green and orange stick, respectively and their corresponding distances are shown from Leu670 of the MTase-like domain.

Fig. 6. Effects of the Asn547Ala mutations on K_m. The interaction between the substrate and Asn547 was predicted from the flexibility of variable loop of the active site in the crystal structure of CHIKV nsP2pro. Kinetic parameters for the proteolytic activity of wild type and Asn547Ala mutant of CHIKV nsP2pro were measured using a FRET assay in 20 mM Bis-Tris-Propnae reaction buffer (pH 7.5) at room temperature and a fluorogenic substrate corresponding to ¾ cleavage site of CHIKV.

Fig. 7. Comparison of molecular interactions at subdomain interface of CHIKV nsP2pro with that of VEEV nsP2pro. **(A)** VEEV nsp2pro (PDB ID: 5EZS): Molecular interaction (H-bond) of Asn475 with Asp507 and Arg662. **(B)** and **(C)** CHIKV nsP2pro Chain B: Molecular interaction of Asn476 with Asp509 and Glu669 respectively.

Table 1. Data collection and refinement statistics of CHIKV nsP2pro (4ZTB)

Data collection	
Space group	$P2_12_12_1$
Unit Cell Parameter Cell dimensions *a* (Å) *b* (Å) *c* (Å)	 87.04 158.96 158.88
Resolution range(Å)	47.21-2.59 (2.68-2.59)
Total Reflection	165501
Unique Reflection Completeness (%) R_{sym} (%)[a] (Last Shell) CC1/2 (%) Mean *I/σ(I)* (Last Shell) Matthews coefficient ($Å^3$ Da^{-1}) Solvent content (%) Multiplicity (Last Shell)	56663 85 (61.65) 0.16 (0.682) 97.2 (86.2) 3.75 (1.40) 3.76 67.0 2.9 (1.5)
Refinement	
No. of residues	1284
Water	342
B factor ($Å^2$) (Wilson) $R_{cryst\ (\%)}$ $R_{free\ (\%)}$	36.72 21.0 24.2
RMS deviation	
Bond lengths (Å)	0.016
Bond angles (°)	1.68
Ramachandran Plot	
Residues in favored region (%) Residues in outlier region (%)	95 0.31

[a] $R_{sym} = \Sigma_{hkl} \Sigma_i^n{}_{=1} | I_{hkl,i} - [overbar] I_{hkl} | / \Sigma_{hkl} \Sigma_i^n{}_{=1} | I_{hkl,I}$

Table 2. Substrate subsite and corresponding substrate binding residues of CHIKV nsP2pro and VEEV nsP2pro

Substrate residue of CHIKV nsP2pro	Substrate residue of VEEV nsP2pro	Substrate subsite	CHIKV nsP2pro residues	VEEV nsP2pro residues	Contributing Subdomain
P1'-Tyr	P1'-Tyr	S1'	Ala 475	Ala 474	Protease
			Asn 476	Asn 475	Proteaes
			Lys 481	Lys 480	Protease
			Cys 478	Cys 477	Protease
			His 548	His 546	Protease
P1-Gly	P1-Ala	S1	Asn 476	Asn 475	Protease
			Cys 478	Cys 477	Protease
			Trp 479	Trp478	Protease
			His 548	His 546	Protease
			Asn 547	Asn 545	Protease
				Leu 665	MTase-like
P2-Gly	P2-Gly	S2	Tyr 512	Cys 477	Protease
			Trp549	Trp 547	Protease
			Trp 479	Trp 478	Protease
				Ala 509	Protease
P3-Ala	P3-Ala	S3	Tyr 544	Trp547	Protease
			Met 707	Met 702	MTase-like
			Asn 547	Ile698	MTase-like
P4-Arg	P4-Asp	S4	Asn 547	Ser 513	Protease
			Trp 549	Trp 547	Protease
			Gln 706	Met 702	MTase-like
			Asp 711	Lys 706	MTase-like

Fig. 1

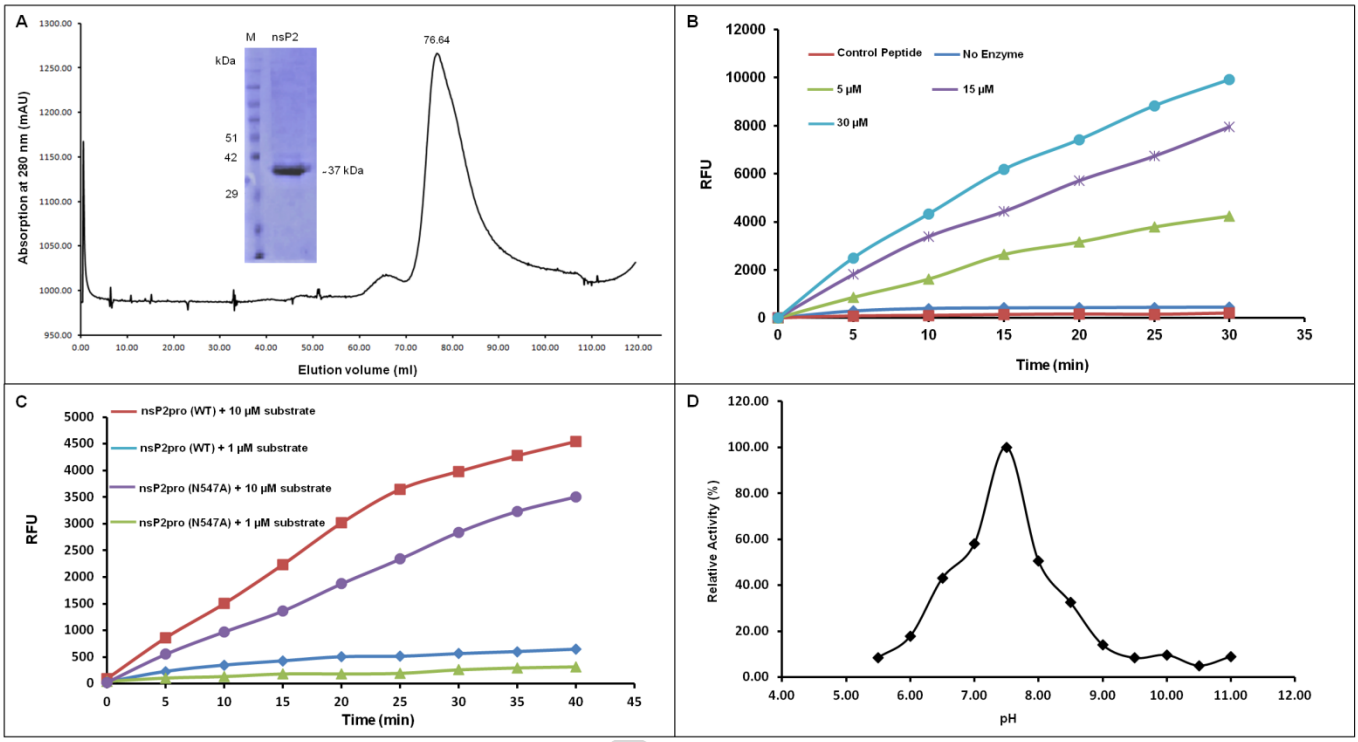

Fig. 2

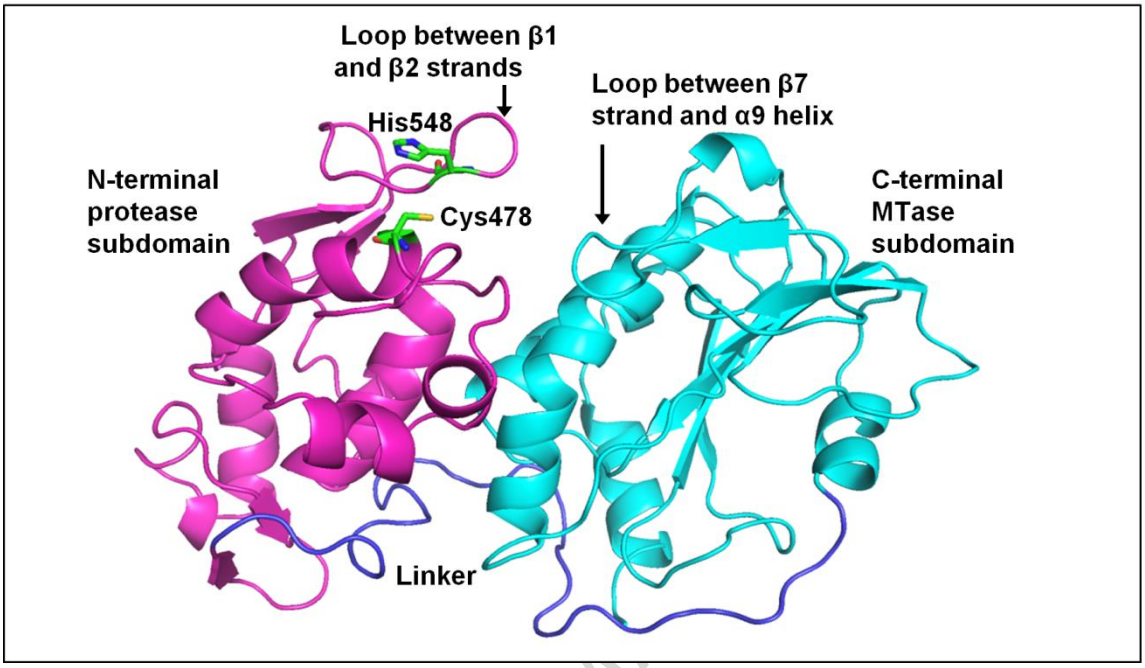

Fig. 3

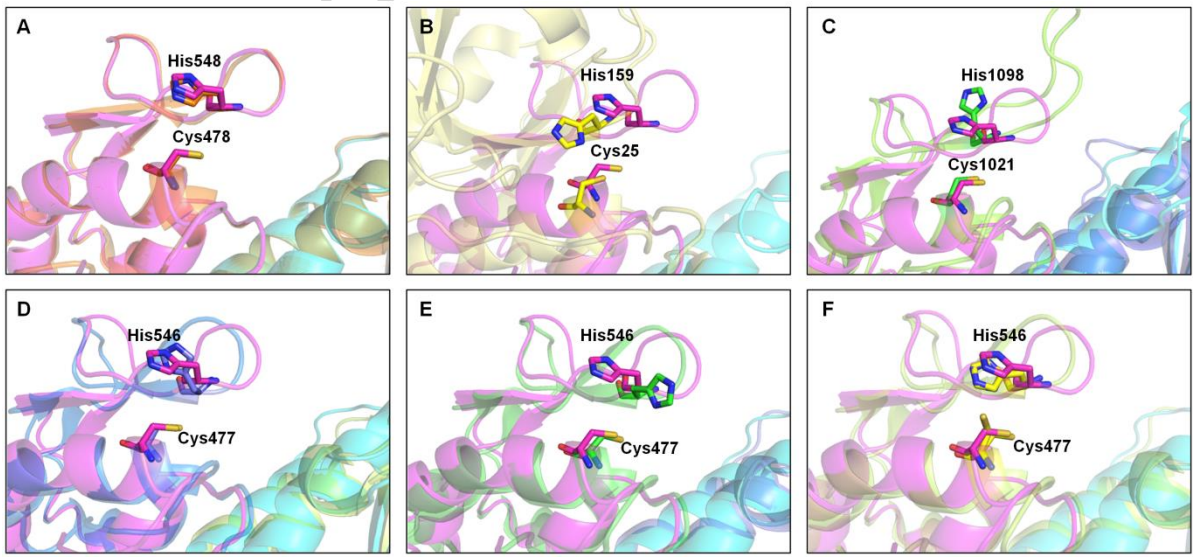

Fig. 4

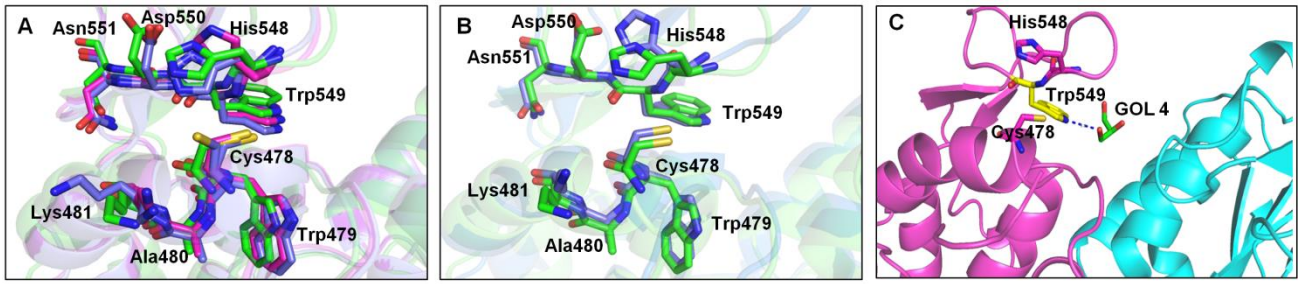

Fig. 5

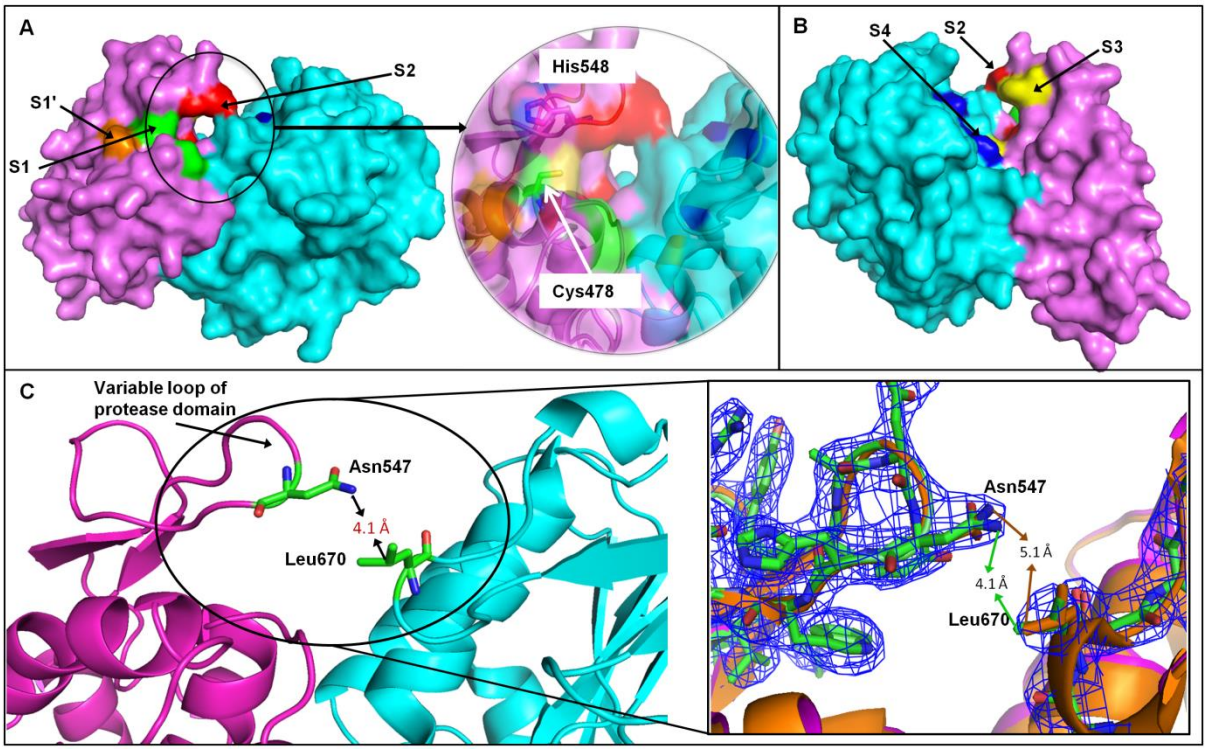

Fig. 6

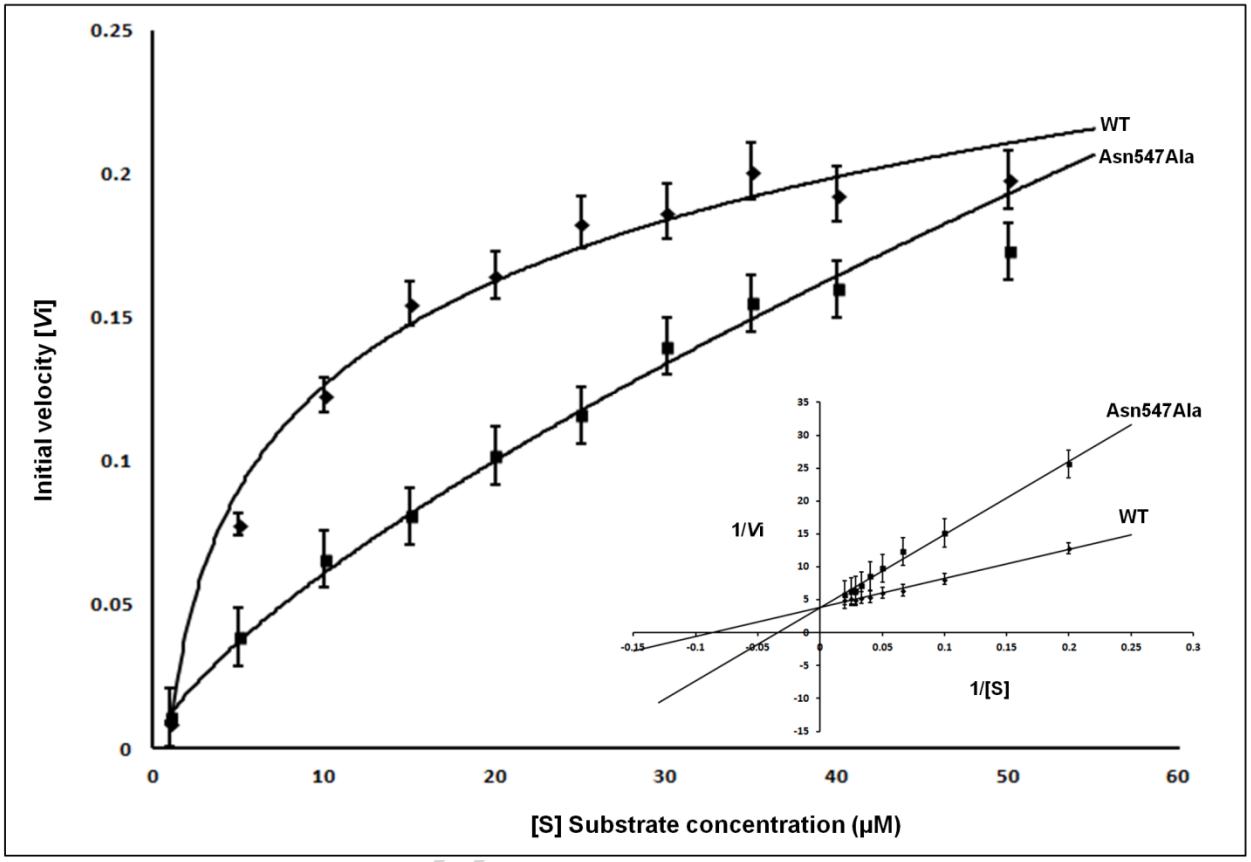

Fig. 7