Structure-function insights into chikungunya virus capsid protein: Small molecules targeting capsid hydrophobic pocket

Rajesh Sharma, Pooja Kesari, Pravindra Kumar, Shailly Tomar*
Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee 247667, India

A R T I C L E   I N F O

Keywords:
Alphavirus
Chikungunya
Capsid protein
Hydrophobic pocket
Surface plasmon resonance

A B S T R A C T

The crystal structure of chikungunya (CHIKV) virus capsid protease domain has been determined at 2.2 Å. Structure reveals a chymotrypsin-like protease fold with a conserved hydrophobic pocket in CHIKV capsid protein (CP) for interaction with the cytoplasmic tail of E2 (cdE2) similar to the capsid protein of other alphaviruses. Molecular contacts between CP-cdE2 were determined by fitting structures of CHIKV CP and cdE2 into the cryo-EM map of Venezuelan equine encephalitis virus (VEEV). Binding of (S)-(+)Mandelic acid (MDA) and Ethyl 3-aminobenzoate (EAB) to the hydrophobic pocket of CP was evaluated by molecular docking. Surface plasmon resonance (SPR) and fluorescence spectroscopy experiments confirmed MDA and EAB binding to the CP. The binding constants ($K_D$) obtained from SPR for MDA and EAB were $1.2 \times 10^{-3}$ M and $0.2 \times 10^{-9}$ M, respectively. This study adds to the understanding of chikungunya virus structural proteins and may serve as the basis for antiviral development against chikungunya disease.

1. Introduction

Chikungunya fever is a re-emerging, arthropod-borne viral disease, reported recently in Central and South America, apart from Asia and Africa (Abdelnabi et al., 2017; Cavrini et al., 2009; Morrison, 2014). In the countries of Indian subcontinent and many other parts of the world, the chikungunya viral fever outbreaks are observed almost every year during monsoon and post-monsoon season when the mosquito density increases. The main vectors responsible for chikungunya virus (CHIKV) transmission are Aedes aegypti and Aedes albopictus. In humans, CHIKV infection is associated with a range of symptoms from mild fever, musculoskeletal pain, rashes to persistent arthralgia. Clinical manifestations of CHIKV disease include lymphopenia, severe dermatological lesions, encephalitis and neonatal encephalopathy caused due to fetal transmission during pregnancy ( Larrieu et al., 2010; Souriisseau et al., 2007). The disease was first reported in 1953, in Makonde plateau from the serum of a febrile patient, a place at the border near Mozambique and Tanzania (Ross, 1956). Since its first incidence, numerous outbreaks occurred in various parts of the world. In 2004, major outbreak occurred in Kenya that had spread the disease in Mayonette, Comoros, Madagascar, La Reunion Island, South East Asia, Europe and West Africa (Sergon et al., 2008, 2007). In 2005, CHIKV outbreak in La Reunion Island affected one-third of its population (Gerardin et al., 2008; Soumahoro et al., 2011). In India, CHIKV epidemic in 2006, had affected more than 1.4 million people (Kumar and Gopal, 2010; Mavalankar et al., 2008;). CHIKV has been listed as category C pathogen by US National Institute of Allergy and Infectious diseases (NIAID) in 2008 (Schwartz and Albert, 2010). Unfortunately, to date, no antiviral drug or vaccine is commercially available against CHIKV infection.

CHIKV belongs to genus Alphavirus and family Togaviridae. Alphaviruses are enveloped, positive-sense single-stranded RNA viruses transmitted by mosquitoes. The genus alphavirus comprises of 29 members that infect a range of animal hosts such as humans, rodents, fish, horse etc. Other members of this genus include many human and animal viruses like Semiliki Forest virus (SFV), Sindbis virus (SINV), Ross River virus (RRV), Western equine encephalitis virus (WEEV), Venezuelan equine encephalitis virus (VEEV) etc. These are further classified into Old World and New World viruses based on the mechanism employed for shutting the host transcription off, mortality rate and disease presentation. The Old World viruses include CHIKV, SINV and SFV etc. utilize the nsp2 protein to down regulate the host cell transcription, have low mortality rate and cause arthralgia. The New World viruses include VEEV, WEEV etc. utilize their CP to down regulate the host cell transcription, have high mortality rate and cause encephalitis (Garmashova et al., 2007; Hahn et al., 1988).

Cryo-electron microscopy (cryo-EM) and crystallographic studies of various alphaviruses and their proteins provide crucial details about the distribution and molecular organization of the virion structural components (Choi et al., 1996, 1997; Lee et al., 1996; Mancini et al., 2000;
Zhang et al., 2002, 2005; Mukhopadhyay et al., 2006; Voss et al., 2010; Sherman and Weaver, 2010; Kostyuchenko et al., 2011; Long et al., 2015). The viral RNA genome is encapsidated by 240 copies of CP that forms the nucleocapsid core (NC) (Schmaljohn and McClain, 1996; Vasileva et al., 2000). The mature alphanvirus virion (70 nm in diameter) has an envelope derived from host cell membrane, embedded with 80 spikes in \( T = 4 \) icosahedral symmetry. Each spike is made up of trimers of E1 and E2 (surface glycoproteins) heterodimers (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994). Alphaviruses, including CHIKV, have genomes of approximately \( \sim 11.7 \) kb, capped at the 5′ end and polyadenylated at the 3′ end. The 5′-two-thirds and the 3′-one-third of the alphavirus genome encodes for non-structural and structural polyproteins, respectively. The non-structural polyprotein is produced by translation of the genomic 49 S RNA that via minus-strand RNA intermediate transcribes the 26S sub-genomic RNA. Upon translation, the 26 S sub-genomic RNA produces structural polyprotein. The processing of non-structural polyprotein leads to the formation of individual mature non-structural proteins: nsP1, nsP2, nsP3 and nsP4. nsP1 is the capping enzyme, nsP2 is the viral helicase and protease, and nsP4 is RNA dependent RNA polymerase, which has intrinsically disordered N-terminal domain (Strauss and Strauss, 1994; Shirako et al., 2000; Tomar et al., 2006, 2011).

The C-terminal one-third region of the viral RNA genome is translated to form the structural polyprotein, which upon processing by capsid protease and host proteases yields structural proteins in order of CP-E3-E2-6K-E1. Structural proteins are required for virion entry, nucleocapsid assembly and virus budding from the host cell membranes. The CP (SinV) is divided into three regions: region I (residues range: 1–80), region II (residues range: 81–113) and region III (residues range: 114–264). The regions I and II are part of the N-terminal domain of CP and are involved in encapsidation of the genomic RNA (Hong et al., 2006). The region III is part of the C-terminal domain, which is responsible for the serine protease activity of CP. The CP has cis-proteolytic activity that cleaves itself from the nascent structural polyprotein precursor (Choi et al., 1991; Tong et al., 1993; Hahn and Strauss, 1990; Aliperti and Schlesinger, 1978). The remaining structural polyprotein has a signal peptide sequence at the N-terminus, which helps in its translocation to the endoplasmic reticulum (ER) and Golgi bodies (Garoff et al., 1978). The E1-E2 heterodimer, self-assembles on the virus surface to form trimeric spikes (Mukhopadhyay et al., 2006; Soonsawad et al., 2010; Strauss and Strauss, 1994). The mature virion envelope is embedded with 80 trimeric spikes which are organized in \( T = 4 \) icosahedral symmetry (Von Bonsdorff and Harrison, 1978). The virus enters the host cell via receptor mediated endocytosis that involves the interactions of E2 glycoprotein with the host cell receptors. The cryo-electron microscopy studies have found that the N-linked glycosylation sites on the E2 glycoprotein is responsible for binding to heparin sulphate (Knight et al., 2009; Ryma et al., 2007). 6 K is a small structural protein having the size of 6000 Da. Small number of copies (up to 30) are incorporated into virions (Gaëdigk-Nitschko and Schlesinger, 1990; Lusa et al., 1991). It is an important structural component and undergoes palmitoylation at conserved cysteine residues, which are crucial for the formation of infectious particles (Liljestrom et al., 1991; Gaedicke-Nitschko and Schlesinger, 1991, 1990). In mammalian cells, 6 K protein is associated with membrane modification by forming cation-selective ion channels (Sanz et al., 2003, 1994; Melton et al., 2002). The E1 envelope glycoprotein is involved in the formation of icosahedral shell of the virus particle and membrane fusion during entry of virus in host cell. The crystal structures of the p62-E1 heterodimer and of the mature E3-E2-E1 glycoprotein complexes along with their fitting into the cryo-EM map of CHIKV revealed insights into the mechanism of low-pH-triggered membrane fusion (Voss et al., 2010; Pletnev et al., 2001).

The CP is a multi-functional protein consisting of two domains, namely the N-terminal RNA binding domain and the C-terminal protease domain (Choi et al., 1991; Melancon and Garoff, 1987; Strauss and Strauss, 1994). The N-terminal domain being less conserved and intrinsically disordered with high degree of positive charge, is responsible for binding to RNA genome, shutting host transcription off and dimerization of CP (Lulla et al., 2013; Owen and Kuhn, 1996). The C-terminal protease domain of alphanvirus CP is highly conserved and is a chymotrypsin-like serine protease. It possesses cis-proteolytic activity that cleaves at the W/S scissile bond and separates CP from the structural polyprotein (Choi et al., 1991; Melancon and Garoff, 1987). After the cis-proteolytic cleavage, the C-terminal tryptophan residue of CP remains bound to the S1 specificity pocket of the protease and blocks the protease activity (Aggarwal et al., 2012; Choi et al., 1991). Recently in-vitro trans-cleavage activity using fluorogenic peptide containing CP protease site have been reported for Aura virus capsid protease (AVCP Δ2) and CHIKV CP (Aggarwal et al., 2014, 2015). Structural studies have revealed that the basic molecular architecture of the active site and the catalytic triad are highly conserved among the serine proteases including the CP of alphaviruses (CHIKV CP: His139, Asp161 and Ser213) (Aggarwal et al., 2012; Choi et al., 1991). Furthermore, the GDG motif that contains the active site serine nucleophile is also conserved among alphavirus CPs (E21GDSG214 in CHIKV CP) (Aggarwal et al., 2012; Choi et al., 1996, 1991). Alphavirus CP (1–261) is ~29 kDa protein and contains nuclear localization signal (NLS) and nuclear export signal (NES) for nuclear-cytoplasmic trafficking. In CHIKV, the CP amino acid stretches from residues 60–99 and 143–155 have been mapped as NLS and NES, respectively (Thomas et al., 2013).

Molecular interactions between the cdE2 and the hydrophobic pocket present in the C-terminal CP domain plays crucial role in the virus budding process (Lee et al., 1996; Kim et al., 2005). Besides molecular contacts between cdE2-CP, the cdE1-CP interaction has also been reported to be important for the budding process (Barth et al., 1992). Combinations of X-ray crystallographic and cryo-electron microscopy (cryo-EM) structural studies have provided crucial insights into the possible mode of interactions between the cytoplasmic tails of glycoproteins (E1 and E2) with CP (Lee et al., 1996, Zhang et al., 2011). Structural analysis of CP and glycoprotein structures fitted into the cryo-EM density map of alphavirus revealed that the conserved Pro405 residue of Aura virus cdE2 makes molecular contacts with the conserved hydrophobic pocket of CP (Aggarwal et al., 2012). Dioxane bound in the hydrophobic pocket of CP-dioxane complex has been proposed to structurally mimics the hydrophobic residue of cdE2. The crystal structure studies of Aura virus CP dioxane complex have proposed that dioxane mimics the pyrrolidine ring of Pro405 residue of Aura virus cdE2 (Lee et al., 1998; Aggarwal et al., 2012; Lopez et al., 1994; Owen and Kuhn, 1997). Moreover, recent findings have demonstrated that binding of picolinic acid (PCA) to the hydrophobic pocket of CHIKV CP inhibits chikungunya virus replication (Sharma et al., 2016). This suggested that heterocyclic ring compounds similar to dioxane and PCA have potential to bind the hydrophobic pocket and disrupt its interaction with the viral glycoproteins.

This study aims to structurally characterize CHIKV CP, investigate the CHIKV CP-cdE2 glycoprotein interactions, and identify molecules that bind into the conserved CP hydrophobic pocket, which may potentially block the viral budding process. In this work, the crystal structure of CHIKV CP has been determined at 2.2 Å resolution. The crystal structure of CHIKV CP and the structure models of glycoproteins were fitted into the cryo-EM density map of VEEV to analyze CP-cdE2 interactions. Based on the molecular interaction, small drug-like heterocyclic molecules were docked into the CHIKV CP hydrophobic pocket in the same position as the conserved Pro404 residue of cdE2 loop. The binding of molecules was experimentally evaluated using surface plasmon resonance (SPR) and fluorescence spectroscopy. The availability of CHIKV CP structure paves way for design and development of inhibitors based on the identified small heterocyclic drug-like molecules targeting the hydrophobic pocket of CHIKV CP.
W245A mutation; Forward primer- 5′-CTCTCGGTGGTGACCGCGAATA
The following complementary mutant primers were designed to make
residue in the hydrophobic pocket of CHIKV CP was mutated to Ala.

2. Materials and methods

2.1. Expression and Purification

The methods for expression and purification of CHIKV CP have been
described previously (Sharma et al., 2016). Briefly, the gene encoding
the carboxyl terminal protease domain (CHIKV CP, residues 106–261)
was cloned in pET28c vector with an N-terminal TEV protease cleavable
His-tag. For performing intrinsic fluorescence, the conserved Trp245
residue in the hydrophobic pocket of CHIKV CP was mutated to Ala.
The following complementary mutant primers were designed to make
W245A mutation; Forward primer- 5′-CTCTCGGTGGTGACCGCGAATA
AAGACATTGTC-3′ and Reverse primer- 5′-GACAACTGTCAATTTCGCC
GTCACCACCGAG-3′. The pET28c-CHIKV CP was used as the tem-
plate for doing PCR reaction followed by Dpn I restriction enzyme di-
gestion of the PCR product to remove the template plasmid. The di-
gested PCR product was used to transform XL-1 blue competent cells.
Both the native CHIKV CP and the mutant W245A proteins were
overexpressed in Escherichia coli strain Rosetta (DE3) using isopropyl
β-D-1-thiogalactopyranoside (IPTG) and was puriﬁed by Ni-NTA afﬁnity
chromatography (Fig. 1A). Further, TEV protease was added to remove
the His-tag and then puriﬁed protein fraction was loaded onto HiLoad
16/600 prep grade Superdex 75 size-exclusion chromatography column
(GE Healthcare) for further puriﬁcation and molecular weight estima-
tion (Fig. 1A). CHIKV CP puriﬁed protein was concentrated and used for
crystallization trials.

2.2. Crystallization and data collection

The sitting drop vapor diffusion method was used for protein crys-
tallization. The protein concentration used for the crystallization was ~
8 mg/ml in 50 mM Tris–HCl pH 7.6, 20 mM NaCl and 5% glycerol.
Different matrix crystal screens were used for the optimization of
crystal growth. The ratio of protein and reservoir buffer used for setting
up crystallization trays was 1:1 against the 100 µl reservoir buffer.
Crystals were obtained at 20 °C in 0.1 M BisTris (pH 6.5), 20% w/v
polyethylene glycol (PEG) 3350 and 0.1 M octyl β-D-glucopyranoside
(BOG); the latter was used as an additive for improving crystal quality.
Use of BOG during crystallization signiﬁcantly decreased the mosaicity
and improved the size of the CHIKV CP crystals. The initial hits grew
gradually and full size crystals grew over the course of 4 weeks. The
obtained rectangular shaped crystal of CHIKV CP was diffracted at
home source (Fig. 1B). The cryoprotectant composition was also
optimized for improving the resolution. The crystal was equilibrated for
2 min in cryoprotectant solution composed of mother liquor containing
25% (v/v) ethylene glycol. Data were collected at the home source with
MAR 345 imaging plate system using Cu Kα radiation generated by a
Bruker-NioniusMicrostar H rotating-anode generator operated at 45 kV
and 60 mA (Fig. 1C). The data were collected at cryogenic conditions
(100 K) at a wavelength of 1.54 Å. The diffraction data were processed,
integrated and scaled using HKL2000 package (Otwinowski and Minor,
1997).

2.3. Structure solution and reﬁnement

The crystal structure of CHIKV CP was determined by molecular
replacement method using poly-Ala model of chain A of Semliki Forest
virus capsid protein (SFCP) (PDB ID: 1VCP) as the search model for
structure solution. The CP of CHIKV shares ~ 93% sequence homology
with SFCP. The MOLREP program from the CCP4 software suite was
used for the molecular replacement (Vagin and Teplyakov, 1997), and
the phenix.refine program (Afonine et al., 2012; Adams et al., 2010)
was used for reﬁnement. The electron density map analysis and manual
model building were carried out using the COOT program (Emsley and
Cowtan, 2004). The partial model was taken through several rounds of
reﬁnement in phenix.refine, followed by reﬁtting in the electron-den-
sity maps. Model rebuilding was done by examining Sigma-A-weighted
maps (Terwilliger, 2001). Structural analysis of the reﬁned model and
the preparation of ﬁgures were done using the PyMOL visualization tool
(The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger)
(DeLano, 2002). The ﬁnal reﬁned model was validated using MOL-
PROBITY server (Davis et al., 2007). Surface assembly analysis was
carried out by PISA (Shoemake et al., 2009). The data collection and
reﬁnement statistics are summarized in Table 1.

Protein structure accession number. Structure coordinates for
CHIKV CP have been deposited in the Protein Data Bank under acces-
sion number 5H23.

2.4. Molecular modeling

Homology models of chikungunya virus E1 and E2 glycoprotein
were generated using the program MODELLER and Swiss-Model (Sali
and Blundell, 1993; Schwede et al., 2003). On the basis of sequence
similarity and structural completeness, the VEEV structure (PDB ID:
3JOC) was used a template for building E1 and E2 models (Zhang et al.,
2011). Clustal omega program was used for multiple sequence
alignment of query with template sequence (http://www.ebi.ac.uk/Tools/msa/clustalo/). The quality of the models were evaluated by the Ramachandran plot (%), allowed 3, favored 97.

<table>
<thead>
<tr>
<th>Crystallographic data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector distance (mm)</td>
<td>170</td>
</tr>
<tr>
<td>No. of images</td>
<td>223</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a = 32.3, b = 39.1, c = 59.3</td>
</tr>
<tr>
<td></td>
<td>a = 92.6°, β = 91.18°, γ = 103°</td>
</tr>
<tr>
<td>Cell Volume (Å³)</td>
<td>75073.19</td>
</tr>
<tr>
<td>Resolution range (Å) (Last Shell)</td>
<td>50-2.2 (2.24-2.20)</td>
</tr>
<tr>
<td>Completeness (%) (Last Shell)</td>
<td>91.6 (77.6)</td>
</tr>
<tr>
<td>Mean I/σ(I) (Last Shell)</td>
<td>0.06 (0.22)</td>
</tr>
<tr>
<td>Total no. of reflections</td>
<td>31669</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>13451</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Matthews coefficient (Å² Da⁻¹)</td>
<td>2.29</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>49.0</td>
</tr>
<tr>
<td>Multiplicity (Last Shell)</td>
<td>2.4 (2.2)</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>32.9</td>
</tr>
</tbody>
</table>

### 2.5. Identification and docking of small heterocyclic molecules

The small heterocyclic molecules structurally similar to PCA were identified from PubChem database. AutoDock 4.2 was used for carrying out docking of identified small molecules into CHIKV CP hydrophobic pocket. Docking was conducted using Windows 2007 on a HPxw8400 workstation. The 3D structures of ligands for the CHIKV CP protein docking were retrieved from PubChem Compound Database. The name of the ligands and their molecular formulas are given in Table 2. Hydrogens atoms and Gasteiger charges were added. For dioxane, picolinic acid, MDA and EAB the Gasteiger charges added were 1.9874, 0.9877, 0.9876, and 1.9875, respectively. The atomic potential grid map was calculated by AutoGrid 4 with a spacing of 0.375 Å and the Grid box dimensions were 34 Å x 34 Å x 34 Å. Other docking parameters were set as default. Obtained results were manually analyzed using PyMOL (DeLano, 2002).

### 2.6. Surface plasmon resonance (SPR)

SPR experiments were performed using BIACORE T200 optical biosensor (GE Healthcare, USA) with research grade CMS sensor chips. SPR binding and kinetic studies were carried out in Phosphate buffer saline (10 mM phosphate buffer with 2.7 mM KCl and 137 mM NaCl pH 7.4) supplemented with 0.5% DMSO. Additionally, 0.1% Tween20 (P9416-Molecular Biology grade, Sigma) was added in the PBS buffer to reduce non-specific binding and for suitable solubility of analytes. This buffer was used for preparation of test samples from the stock solutions of analytes and also as a control buffer sample for performing SPR experiments. The changes in the refractive index as a function of time at constant flow rate were recorded. The net increase in the refractive index as compared to the control recorded over time, corresponds to the amount of ligand that binds to CHIKV CP. This net change is recorded in response unit (RU). The data were collected using Biacore control software and analysis was done using Biacore evaluation software.

Standard amine-coupling method was used for the immobilization of CHIKV CP protein. CHIKV CP (100 μg/ml) dissolved in low salt buffer (0.1 M Sodium acetate buffer) with pH 4.5 was immobilized to a CMS sensor surface. After the activation of flow cell surface by flowing 1:1 mixture of 100 mM N-ethyl-N-(dimethylaminopropyl)-carbozidime hydrochloride (EDC) and 100 mM N-hydroxysuccinimide (NHS), the CHIKV CP protein was injected. CHIKV CP protein was added to the activated flow cell surface using a flow rate of 10 μl/min for 7 min and the residual activated carboxyl methyl groups were blocked by injecting ethanolamine. The adjacent flow cell was used as a reference for doing all the SPR measurements that contains no ligand and blocked in the similar way as activated flow cell.

### 2.7. Fluorescence spectroscopy

Intrinsic fluorescence experiments for purified CHIKV CP protein were performed using a Hitachi F-4600 fluorescence spectrophotometer. The quartz cuvette of dimensions 10 mm x 4 mm was used for recording the fluorescence spectra, at 25 °C. To measure the fluorescence intensity of CHIKV CP, the protein sample (10 μM) were excited
at 280 nm wavelength and the emission spectra were recorded in the range of 295–550 nm. Each spectra obtained was the average of three scans. The buffer containing 50 mM Tris, pH 7.6 and 20 mM NaCl was used for diluting the stock solutions of protein and ligand molecules. The effect of MDA and EAB on the CHIKV CP intrinsic fluorescence was studied. The emission spectra was recorded upon addition of MDA (0.5, 2.5, 5.0, 7.5 and 10 mM) and EAB (6.25, 12.5, 50, 100, 200 and 400 μM) to CHIKV CP (10 μM), after incubation for 15 min. The obtained spectra were analyzed and compared with the native CP and the mutant W245A protein spectra.

3. Results and discussion

3.1. 3D structure of CHIKV CP

3.1.1. Crystal structure determination

CHIKV CP protein was crystallized in the triclinic space group (P1) with two molecules per asymmetric unit. The estimated Matthews coefficient was 2.29 (Å³ Da⁻¹) which corresponds to 49.0% solvent content. The structure of CHIKV CP was determined by diffraction data extending to 2.2 Å resolution. BLASTp results showed that CHIKV CP sequence shares highest (~ 93%) sequence homology with sequence data extending to 2.2 Å resolution. BLASTp results showed that CHIKV CP sequence shares highest (~ 93%) sequence homology with sequence data extending to 2.2 Å resolution. BLASTp results showed that CHIKV CP sequence shares highest (~ 93%) sequence homology with sequence data extending to 2.2 Å resolution. BLASTp results showed that CHIKV CP sequence shares highest (~ 93%) sequence homology with sequence data extending to 2.2 Å resolution.

The α-β-capsid protease domain of CHIKV CP was built into the electron density map of the CP crystal structure. The CP sub-domains were modeled using phenix.autobuild (Table 1). The β-sheets of N-terminal sub-domain 1 [β1(114–119), β2 (122–130), β3 (133–137), β4 (143–144), β5 (155–157), β6 (162–166)] and C-terminal sub-domain 2 [β7 (185–189), β8 (192–197), β9 (200–204), β10 (216–219), β11 (224–233), β12 (237–246), β13 (249–253), β14 (259–260)] arrange to form a Greek key motif, which is a key characteristic of chymotrypsin-like serine proteases. In addition to this, the N-terminal sub-domain 1 also contains two helices: α1 (146–151) and α2 (168–174) (Fig. 2A). Each sub-domain is stabilized by the presence of salt bridges. Five salt bridges (Glu116-Lys118, Lys118-Asp145, Lys136-Glu163, His139-Asp161, and Asp212-Glu234) are formed near the N-terminus.

3.1.2. Monomer architecture

The overall assembly of protease domain of CHIKV consists of two β-barrel sub-domains similar to crystal structures of other alphavirus CPs. The sub-domains 1 and 2 consist of six and eight β-strands, respectively linked via a 10 residues linker loop region (Ala175–Glu184) (Figs. 2A, 2B and 2C). The β-sheets of N-terminal sub-domain 1 [β1(114–119), β2 (122–130), β3 (133–137), β4 (143–144), β5 (155–157), β6 (162–166)] and C-terminal sub-domain 2 [β7 (185–189), β8 (192–197), β9 (200–204), β10 (216–219), β11 (224–233), β12 (237–246), β13 (249–253), β14 (259–260)] arrange to form a Greek key motif, which is a key characteristic of chymotrypsin-like serine proteases. In addition to this, the N-terminal sub-domain 1 also contains two helices: α1 (146–151) and α2 (168–174) (Fig. 2A).

Each sub-domain is stabilized by the presence of salt bridges. Five salt bridges (Glu116-Lys118, Lys118-Asp145, Lys136-Glu163, His139-Asp161, and Asp148-Lys151) are confined to sub-domain 1 whereas four salt bridges (His180-Asp248, Arg200-Glu234, Arg215-Asp212 and Arg223-Asp219) are confined to sub-domain 2 (Table 3). These salt bridges are responsible for stability and compact sub-domain structure of CHIKV CP. The Cα RMS deviation between both CHIKV CP monomers A and B was 0.24 Å.

3.1.3. Substrate binding site architecture

The spatial architecture of the substrate binding pocket is composed of conserved catalytic triad of serine protease family including alphavirus CPs. As a conserved feature of serine proteases, CHIKV CP contains a catalytic triad formed of His139, Asp161 and Ser213 residues (Fig. 3A). The triad is positioned in the cleft between the two β-barrel sub-domains (Figs. 2C and 3A). Additionally, the CHIKV CP is also composed of Glu116-Lys118, Lys118-Asp145, Lys136-Glu163, His139-Asp161, and Asp148-Lys151 motifs having key active site serine residue (Fig. 3A). This feature is conserved among alphavirus CPs and other members of chymotrypsin-like serine proteases. The residues Ser213 and Gly216 are involved in forming the oxyanion hole. Alphavirus capsid protease cleaves the scissile bond between the W/S site and

![Fig. 2. Crystal structure of CHIKV CP. (A) Cartoon representation of CHIKV CP crystal structure with detail secondary structure elements of the two sub-domains. The sub-domain 1, the sub-domain 2 and the linker region are represented in blue, magenta and green color, respectively. (B) Surface representation of the sub-domain 1, sub-domain 2 and the linker region of CHIKV CP crystal structure. The color scheme is same as in the cartoon view representation. (C) The surface view of CHIKV CP showing highly conserved catalytic triad and hydrophobic pocket in red and blue color, respectively.]

![Table 3. CHIKV CP intra-subdomain salt bridges.](image-url)
leaves the terminal Trp261 residue in the substrate binding pocket further inactivating the capsid protease.

3.1.4. Dimer architecture

CHIKV CP crystalizes as a dimer in the asymmetric unit whereas it is a monomer in solution. Previous studies have reported the crystallographic dimer in the crystal structures of alphavirus CPs from Semliki Forest (PDB: 1VCP), Sindbis (PDB: 1WYK) and Aura virus (PDB: 4UON) (Aggarwal et al., 2014; Choi et al., 1997, 1991). In the crystal structure of CHIKV CP, a crystallographic dimer with similar dimeric interface as in the crystal structures of SFV CP (SFCP), Sindbis virus CP (SCP) and Aura virus CP (AVCP) was observed. The CHIKV CP crystallographic dimer is in tail-to-tail contact through the C-terminal sub-domains as observed in previously determined crystal structures (Aggarwal et al., 2014; Choi et al., 1991, 1995, 1997). However, in SFCP a head-to-head dimeric contact, involving the interaction between the N-terminal sub-domains of chain B and C was also observed.

The AB dimeric interface area in CHIKV CP was found to be ~441.6 Å², encompassing 5.5% of total solvent-accessible area (SAA) as calculated using PISA. The previous reports show that the dimeric interface area in SCP, AVCP and SFCP are ~409 Å², ~375 Å² and ~374 Å², respectively. This suggests that in CHIKV CP, a large part of C-terminal subdomain of one monomer interacts with the C-terminal subdomain of another monomer. The contacts between two monomers are facilitated by involvement of 13 residues from each monomer. The dimer interface include residues Pro183, Glu184, Gly185, Tyr186, Tyr187, Asn188, Thr189, Gly192, Ala193, Asp219, Asn220, Lys221 and Arg223 from each monomer. The CHIKV CP dimer was stabilized by interactions taking place at the monomeric interface. The residues Tyr186, Asn188 and Asn220 of monomer A are involved in forming H-bonds with Tyr186 and Asn220 of monomer B (Figs. 3B, 3C, Fig. 4 and Table 4). Apart from this, hydrophobic interactions also play a crucial role in CHIKV CP dimer stability.

The region from Pro-Glu-Gly-Tyr-Tyr-Asn, and residues Gly192, Ala193 and Asn220 are highly conserved among CHIKV CP, SFCP, SCP and AVCP (Fig. 4). However, the Tyr186 in CHIKV CP is replaced by His192, Phe188 and Phe191 in SFCP, SCP and AVCP, respectively (Fig. 4). In SCP, a mutation at conserved Phe188Gly resulted in the absence of crystallographic dimer (Choi et al., 1996). Additional mutational and structural studies need to be carried out for analyzing the biological significance of the conserved residues at the dimer interface.

3.2. CHIKV CP-cdE2 interactions

The cryo-EM structure of VEEV at a resolution of 4.4 Å provided the structural insights of viral surface assembly for mature virion formation (Zhang et al., 2011). The VEEV structural analysis shows that the cytoplasmic domain of E2 glycoprotein (cdE2) makes molecular contacts with the hydrophobic pocket of the capsid protein. The helix-loop-helix motif of cdE2 interacts with the hydrophobic pocket of capsid protein (Supplementary Fig. S1). The E2 protein of CHIKV is 423 amino acids long and consists of three regions, an ectodomain of 364 residues followed by a transmembrane helix (TM) of 26 residues and a 33 residues long cdE2 domain. Additionally, E2 is associated with two N-linked glycans at positions Asn263 and Asn345 (Sun et al., 2013). Though, the exact Cd-cdE2 interactions are not clear until now. To predict the interactions taking place between Cd-cdE2 at the viral surface of CHIKV, we fitted the crystal structure of CHIKV CP and the modeled cdE1 and cdE2 glycoproteins into cryo-EM density map of VEEV (Fig. 5A). The correlation coefficient of fitting of CHIKV cdE2 and CHIKV CP into VEEV cryo-EM map was 0.67. The cdE2 is highly conserved among the alphavirus family and through its helix-loop-helix conformation, it interacts with the hydrophobic pocket (Figs. 5B and 5C). The conserved Tyr (Tyr400 in CHIKV) is located at the end of the helix-I of the helix-loop-helix motif of cdE2. The studies in SINV proposed that this conserved Tyr400 residue of cdE2 is responsible for making contacts with CP (Lee et al., 1996; Sun et al., 2013). However, in the predicted CHIKV cdE2 model, the Tyr400 is oriented away from the hydrophobic pocket of CP and makes hydrophobic contacts with Leu411 and Leu412 of another helix (helix-II). Similar orientations of Tyr400 residue and disulfide bond (Cys396-Cys417) is seen in the cryo-EM structure of VEEV (Zhang et al., 2011).

Structural analysis of CHIKV cdE2 unveils three conserved cysteine residues (Cys396, Cys416, and Cys417) in the helix-loop-helix motif (Fig. 5D). The Cys396 residue of helix-I of helix-loop helix motif forms disulfide bond with Cys417 lying in the loop region at the end of helix-II (Fig. 5C). These conserved cysteine residues (Cys416 and Cys417) play a critical role in interactions with CP (Jose et al., 2012). These two residues have also been suggested to undergo palmitoylation (typical feature of membrane proteins) that helps in anchoring and orienting the cdE2 at the inner side of membrane that facilitate its interactions with CP (Jose et al., 2012). Mutations of these conserved Cys residues results in virus assembly defects and affects the entry of virus in host cells, thereby, decreasing the virus infection (Jose et al., 2012). The loop region of the helix-loop-helix motif of cdE2 contains seven residues (402Leu-Val408). The residues of the loop region of cdE2 interact with residues of the hydrophobic pocket of CHIKV CP. The Pro404 lying in the loop region gets buried into the hydrophobic pocket and makes hydrophobic contacts with the surrounding residues (Fig. 5E). In the crystal structure of CHIKV CP, ethylene glycol used as a cryo-protectant was found bound in the hydrophobic pocket, occupying the same position as Pro404 of cdE2. Similar hydrophobic contact of the conserved Pro405 of Aura virus cdE2 with the hydrophobic pocket of AVCP has...
earlier been suggested for Aura virus (Aggarwal et al., 2012).

3.3. Structural comparison of alphavirus capsid proteins

Structural comparison of CHIKV CP with the known crystal structures of alphavirus capsid proteins was done for detailed structural analysis. The sequence based homology search for CHIKV CP against PDB database using NCBI BLASTp tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) mostly gave hits of alphaviral capsid crystal structures (ranges from 63% to 93% identity). Structural homolog search using DALI server revealed the highest structural similarity of CHIKV CP with SFCP (PDB: 1VCP) having identity 93%, Z-score 29.2 and RMSD 0.7 Å, followed by AVCP (PDB: 4AGJ, with identity 69%, Z-score 28.7 and RMSD 0.9 Å), SCP (PDB: 1WYK, with identity 67%, Z-score 28.5 and RMSD 0.9 Å) and VEEVCP (PDB: 1EP5, with identity 63%, Z-score 27.8 and RMSD 0.9 Å). Following the highest structural similarity with Alphavirus CP’s (SFCP, AVCP, SCP and VEEV), DALI server also revealed structural similarity to the other members of serine protease family like chymotrypsin, flavivirus NS3 protease, 3C like proteinase, glutamyl endopeptidase, chymase and elastase. Structural similarity of CHIKV CP with flavivirus protease indicates the evolution of flavivirus and alphavirus protease from a common ancestry of chymotrypsin-like serine proteases. Superimposition of CHIKV CP with related alphavirus CP structures and structural homology results from DALI reveals the β-barrel structure is highly conserved among various members of serine protease family. Although the overall structural architecture and active site conformation of alphavirus enzymes is typically conserved, slight variation in the position of three loop regions lying on the surface of the protein was observed. Fig. 3C depicts the three loop regions showing high Cα RMS deviation. Region I lies in-between the β1 and β2 involves residues His119-Lys122 (Circle 1). This region in SFCP has been reported to be involved in making head-to-head contacts in hexamer and pentamer formation in SFCP. Region II is part of the linker-loop connecting the β-sheets of two sub-domains. This region is of variable length and embraces diverse conformations. The length of this inter-domain linker-loop in CHIKV CP, AVCP, SFCP, SCP and VEEV CP is of 10, 18, 10, 13 and 11 amino acid residues long stretch, respectively (Circle 2). A part of this inter-domain loop region contributes to formation of the conserved hydrophobic pocket where the tail of E2 glycoprotein binds. The region III is the loop region connecting β12 (237–246) and β13 (249–253). The length of this loop region is two residues long in CHIKV CP (residues: 247–248) and SFCP (residues: 253–254). When compared to the other alphavirus CPs, this loop region was longer in AVCP, VEEVCP and comparable to SCP (Circle 3). This flexible loop region interacts with cdE2 glycoprotein and is also involved in dimer interface contacts (Choi et al., 1997).

Table 4

<table>
<thead>
<tr>
<th>Monomer A</th>
<th>Monomer B</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O; Tyr186</td>
<td>ND2; Am220</td>
<td>3.1</td>
</tr>
<tr>
<td>N; Tyr186</td>
<td>OD1; Am220</td>
<td>3.4</td>
</tr>
<tr>
<td>OD1; Am188</td>
<td>N; Tyr186</td>
<td>3.5</td>
</tr>
<tr>
<td>OD1; Am220</td>
<td>N; Tyr186</td>
<td>3.1</td>
</tr>
<tr>
<td>ND2; Am220</td>
<td>O; Tyr186</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Fig. 4. Multiple sequence alignment of CHIKV CP with CP of other alphaviruses. The N-terminal RNA binding domain of CHIKV CP is shown with dotted box. Identical and similar residues are background highlighted in red and yellow colors, respectively. The solid black color squares under the amino acids represent residues of the hydrophobic pocket. The highly conserved catalytic residues (His139, Asp161 and Ser213 CHIKV CP) are represented by solid black circles below the residues. The conserved amino acid residues involved in crystallographic dimer formation are represented by upward black arrows below the residues. The residues involved in polar interactions at the dimeric interface are represented by solid black triangles below the residues. The conserved GDSG motif (211GDSG214 in CHIKV CP) is shown within the solid box.
3.4. Identification and molecular docking of small heterocyclic molecules

The crystal structural studies of capsid protease in complex with dioxane from Sindbis (PDB ID: 1WYK) and Aura virus (PDB ID: 4AGJ) (Aggarwal et al., 2012; Choi et al., 1996; Lee et al., 1998) and studies of alphavirus inhibition by heterocyclic compounds targeting the conserved hydrophobic pocket (Kim et al., 2005; Aggarwal et al., 2017, 2012; Sharma et al., 2016) have revealed that the hydrophobic pocket of CP is a potential antiviral drug target. Therefore, in this study PubChem database was searched for drug-like heterocyclic compounds similar to dioxane, picolinic and piperazine. The small molecules containing heterocyclic ring were identified for potentially targeting the CP-cdE2 interactions. These small molecules were docked into the hydrophobic pocket of CHIKV CP crystal structure. Based on the docking score, MDA and EAB showed better binding tendency to CHIKV CP hydrophobic pocket than dioxane and PCA. Table 2 shows the molecular structure and docking score of the compounds. Detail molecular interaction of MDA and EAB compounds in the hydrophobic pocket of CHIKV CP is shown in Figs. 6A and 6B. The predicted binding pose of MDA and EAB in the CHIKV CP hydrophobic pocket are shown in panel C and D of Fig. 6 respectively. The binding affinities of these two molecules were further experimentally evaluated using SPR and fluorescence spectroscopy.

3.5. Surface plasmon resonance

The ligand (CHIKV CP protein) was immobilized on the flow cell 4 (FC-4) and the final immobilization level was 2771 RU. Flow cell 3 (FC-3) was selected as a reference to minimize the unwanted drift and systematic noise. For binding studies, the identified MDA and EAB molecules were injected over the immobilized CHIKV CP, at concentrations ranging from 0.0625 to 1 mM and 0.125–1 mM, respectively. The responses achieved upon binding of the two analyte, from different concentration range of MDA and EAB were used to measure the binding kinetic parameters for the interaction with CHIKV CP. For kinetic studies, the contact time of MDA and EAB binding to the CHIKV CP protein was monitored for 180 s with the flow rate of 30 µl/min. The responses of reference cell obtained for the control buffer blanks were flat, demonstrating the stability of the base line during the experiment. Kinetic parameters for binding of MDA and EAB to the immobilized CHIKV CP protein were measured. The data obtained were evaluated by the Biacore Evaluation software (Biacore T200 Evaluation Software, Version: 2.0) (Figs. 7A and 7B). The data were fitted as per the global fitting of the kinetic model with two state model. The values of the rate constants (Kd) obtained from SPR for binding of MDA and EAB to CHIKV CP were 1.2 × 10−3 M and 0.2 × 10−9 M respectively.
3.6. Effect of MDA and EAB on CHIKV CP intrinsic fluorescence

The intrinsic fluorescence of CHIKV CP (residues 106–261) and CHIKV CP W245A mutant were measured in its native form and upon binding of MDA and EAB at varied concentrations. As suggested from docking studies of dioxane, MDA and EAB are predicted to bind the hydrophobic pocket of CHIKV CP with higher affinities than dioxane and PCA (Sharma et al., 2016). The binding of MDA and EAB to CHIKV CP hydrophobic pocket results in decrease of fluorescence intensity with the increase in concentration as shown in Figs. 7C and 7D. Both the MDA and EAB quench the CHIKV CP fluorescence and shift the emission maxima to the longer wavelengths i.e. exhibiting red shift. It is suggested that the binding of MDA and EAB in the CHIKV CP hydrophobic pocket further exposed the buried Trp245 to the solvent and results in red shift in the emission spectra. Fluorescence studies involving CHIKV W245A mutant, results in no change in the fluorescence spectra in the presence of MDA and EAB confirming that MDA and EAB bind to the hydrophobic pocket and the residue Trp245 shows red-shift upon the addition of MDA and EAB. Fluorescence studies further validated the SPR results that MDA and EAB bind to the hydrophobic pocket of CHIKV CP.

4. Conclusion

Alphavirus capsid protein is a multifunctional protein, divided into two domains: the N-terminal RNA binding subdomain and the C-terminal protease domain. The present study, describes the crystal structure of CHIKV CP at 2.2 Å resolution and showed that the protease domain has two β-barrel sub-domains, here referred to as N-terminal sub-domain 1 (residue 111–174) and C-terminal sub-domain 2 (residue 185–261) connected together through a 10 residue long linker (residue 175–184). Belonging to the serine proteases, CHIKV CP contains a catalytic triad formed of residues His139 and Asp161 (contributed by sub-domain 1) and Ser213 (contributed by sub-domain 2). The triad is positioned in the cleft between the two β-barrel sub-domains. The key catalytic Ser213 residue along with Gly211 helps in stabilization of the backbone oxygen atom of Trp261 thereby leading to formation of oxyanion hole. The cis-proteolytic activity of the CP is responsible for cleavage from the nascent structural polyprotein during translation (peptide bond cleavage between Trp261-Ser262), leaving the terminal Trp261 residue in the substrate binding pocket, thus inactivating the CP.

The surface near the cleft contains hydrophobic residues which acts as site for binding of the cdE2 glycoprotein and is referred to as hydrophobic pocket (Lee et al., 1996; Kim et al., 2005). Structure comparison with other known alphavirus CP shows that the loop region found in the hydrophobic pocket is flexible and of variable length. The protein-protein interactions between cdE2 and CP are crucial for alphavirus budding and thus life cycle of virus. The residues present in the helix-loop-helix motif of cdE2 position the loop region in the hydrophobic pocket of CP. The Pro404 of cdE2 makes hydrophobic contacts with the residues of the CP hydrophobic pocket. Studies in past have reported the various techniques to design peptidomimetic molecules against viral proteases (Dhindwal et al., 2016; Randolph and...
obtained for the binding of MDA to the immobilized CHIKV CP protein are shown. The data were fitted by two state binding model with $K_D$ value of 0.37. The $K_D$ value obtained for the binding of MDA to immobilized CHIKV CP was $1.2 \times 10^{-3}$ M. (B) Binding and kinetic analysis of Ethyl 3-aminobenzoate (EAB). The relative responses obtained for the binding of EAB to the immobilized CHIKV CP protein are shown. The data were fitted by two state binding model with $K_D$ value of 0.77. The $K_D$ value obtained for the binding of EAB to immobilized CHIKV CP was $0.2 \times 10^{-9}$ M. (C) The intrinsic fluorescence of CHIKV CP and protein upon binding to MDA was determined. The emission spectra was recorded upon addition of MDA at different concentrations up to 10 mM (0.5, 2.5, 5.0, 7.5 and 10 mM) to the CHIKV CP (10 µM), after incubation for 15 min. The fluorescence intensity got decreased with the increase in concentration of MDA. Fluorescence studies involving the CHIKV W245A mutant (10 µM) did not show any red shift/quenching upon addition of MAD (5 mM). (D) The intrinsic fluorescence of CHIKV CP protein and upon binding to EAB was determined. The emission spectra was recorded upon addition of EAB at different concentrations up to 400 µM (6.25, 12.5, 50, 100, 200 and 400 µM) to the CHIKV CP (10 µM), after incubation for 15 min. The fluorescence intensity decreases with the increase in concentration of EAB. Fluorescence studies involving the CHIKV W245A mutant (10 µM) did not show any red shift/quenching upon addition of EAB (50 µM).

**DeGoey, 2004; Smith III et al., 1994).** Such techniques can be employed to target the cdE2-CP interactions and design mimetics against the hydrophobic pocket.

Literature shows that targeting hydrophobic pocket through di-oxane and picolinic acid (PCA) structurally mimics the binding of pyrrolidine ring of proline residue and is predicted to disrupt the cdE2-CP molecular interactions and indirectly inhibit the virus budding and replication process. Small heterocyclic compounds (MDA and EAB) were selected based on the structure similarity to PCA. Results of molecular docking of MDA and EAB suggested that these two molecules have good binding affinity for the hydrophobic pocket of CHIKV CP. Further, the molecular interactions were analyzed biophysically through SPR and fluorescence experiments. The values of the rate constants ($K_D$) obtained from SPR for binding of MDA and EAB to CHIKV CP were $1.2 \times 10^{-3}$ M and $0.2 \times 10^{-9}$ M, respectively. Both MDA and EAB quench the CHIKV CP fluorescence and exhibit red shift. Fluorescence studies involving the CHIKV W245A mutant further confirmed the role of residue Trp245 in the red-shift of emission maxima and binding of MDA and EAB to the protein. Since the cdE2-CP interaction is expected to takes place through the conserved Pro404 residue lying in the loop region of the helix-loop-helix motif of cdE2 and the residues lining the hydrophobic pocket of the CP are conserved (Supplementary Figure. S2). Thus, MDA and EAB derivative molecules designed and developed based on the complimentary shape and chemistry of the hydrophobic pocket are expected to have broad spectrum antiviral activity against alphaviruses as has been reported for picolinic acid (Sharma et al., 2016).

The study elucidates the atomic structure of CHIKV CP and the importance of the hydrophobic pocket of chikungunya CP. It is an important stepping stone for the future mutational and structural analyses of CP in complex with hydrophobic molecules binding in the pocket and paves way for structure-based antiviral development against the chikungunya disease.

**Acknowledgements**

ST acknowledges the financial support from Indian Council of Medical Research (ICMR: Ref no. BIC/12(26)/2013). Department of Biotechnology (DBT) is acknowledged by RS for his Senior Research Fellowship (SRF). Ministry of Human Resource Development (MHRD) is acknowledged by Pooja for her Senior Research Fellowship (SRF). We thank the Macromolecular Crystallographic Facility (MCU) at IIC, Indian Institute of Technology Roorkee.

**Appendix A. Supplementary material**

Supplementary data associated with this article can be found in the online version at [http://dx.doi.org/10.1016/j.virol.2017.12.020](http://dx.doi.org/10.1016/j.virol.2017.12.020).

**References**


Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W.,
Aggarwal, M., Dhandwal, S., Kumar, P., Kuhn, R.J., Tomar, S., 2014. Trans-Protease active site-directed mutations in the chimeric proteins that form the active form of the alphavirus capsid protein. J. Virol. 88, 12242–12253.
Barth, B.U., Suomalainen, M., Liljeström, P., Garo,


sites. Nucleic Acids Res. 38, 518–524.


