Virology 515 (2018) 223–234

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

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

ARTICLE INFO

Keywords:
Alphavirus
Chikungunya
Capsid protein
Hydrophobic pocket
Surface plasmon resonance

ABSTRACT

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 1.2×10^{-3} M and 0.2×10^{-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; Sourisseau 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;

* Corresponding author.
E-mail addresses: shailfbt@iitr.ac.in, shaiprav@gmail.com (S. Tomar).

https://doi.org/10.1016/j.virol.2017.12.020
Received 23 July 2017; Received in revised form 3 December 2017; Accepted 20 December 2017

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; Vasiljeva et al., 2000). The mature alphavirus virion (70 nm in diameter) has envelop 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 has genome of approximately ~ 11.7 kb, capped at the 5′ end and poly-adenylated 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; Ryman 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 (Gaedigk-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; Gaedigk-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 alphavirus 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 GDSG motif that contains the active site serine nucleophile is also conserved among alphavirus CPs ($_{211}$GDSG$_{214}$ 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.