Trans-protease activity and structural insights into the active form of the alphavirus capsid protease

Megha Aggarwal, Sonali Dhindhwala, Pravindra Kumar, Richard J. Kuhn and Shailly Tomara

aDepartment of Biotechnology, Indian Institute of Technology Roorkee, Roorkee-247667, India
bDepartment of Biological Sciences and Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907, USA

Running Title: Structure of active capsid protease from Aura virus

Corresponding author. Tel.: (91) 01332-285849; fax: (91)-1332-273560
E-mail address: shailfbt@iitr.ac.in, shaiprav@gmail.com (S. Tomar)

ABSTRACT

The alphavirus capsid protein (CP) is a serine protease that possesses cis-proteolytic activity essential for its release from the nascent structural polyprotein. The released CP further participates in viral genome encapsidation, nucleocapsid core formation followed by its attachment to glycoproteins and virus budding. Thus, protease activity of the alphavirus capsid is a potential anti-alphaviral target to arrest capsid release, maturation and structural polyprotein processing. However, the discovery of capsid protease
inhibitors has been hampered due to the lack of a suitable screening assay and crystal structure in its active form. Here we report the development of a trans-proteolytic activity assay for Aura virus capsid protease (AVCP) based on fluorescence resonance energy transfer (FRET) for screening protease inhibitors. Kinetic parameters using fluorogenic peptide substrates were estimated, and a $K_M$ value was found to be $2.63 \pm 0.62 \mu M$, and a $K_{cat}/K_M$ value was $4.97 \times 10^4 M^{-1} min^{-1}$. Also, the crystal structure of the trans-active form of AVCP has been determined to 1.81 Å resolution. Structural comparisons of the active form with the crystal structures of available substrate-bound mutant and inactive blocked forms of the capsid protease identifies conformational changes in the active site, the oxyanion hole and the substrate specificity pocket residues, which could be critical for rational drug design.

IMPORTANCE

The alphavirus capsid protease is an attractive antiviral therapeutic target. In this study, we have described the formerly unappreciated trans-proteolytic activity of the enzyme and for the first time have developed a FRET based protease assay for screening capsid protease inhibitors. Our structural studies unveil the structural features of the trans-active protease which has been previously proposed to exist in the natively unfolded form (1). The different enzymatic forms have been structurally compared to reveal conformational variations in the active and substrate binding site. The flexible active site residue Ser218, the disordered C-terminal residues after His261, and the presence of a water molecule in the oxyanion hole of AVCPΔ2 reveal the effect of the C-terminal Trp267 deletion on enzyme structure. New structural data reported in this study along with the fluorogenic
Assay will be useful in substrate specificity characterization, high-throughput protease inhibitor screening and structure-based development of antiviral drugs.

Alphaviruses are arthropod borne, enveloped, positive-sense, single stranded RNA viruses that are the causative agent of a range of serious human and livestock diseases. Infection in humans can cause fever, rash, encephalitis and polyarthritis. The members of genus alphavirus include Chikungunya virus (CHIKV), Eastern, Western and Venezuelan equine encephalitis virus, Aura virus, Semliki Forest virus (SFV) and Sindbis virus (SINV). Due to the epidemic outbreak of CHIKV in 2005-2006 in the Indian subcontinent, it is considered to be a re-emerging pathogen and a potential public health threat. Currently no antiviral drug or vaccine is available against alphaviruses.

The 49S and 26S RNA of alphaviruses are translated into the non-structural and structural proteins in the infected cell, respectively. The structural polyproteins are CP, E3, E2, 6K and E1 (2). The alphavirus CP is present at the amino-terminus of the structural polyprotein and has been reported to be a chymotrypsin-like serine protease. The first step in structural polyprotein processing is the autocatalytic cleavage of CP to release itself from rest of the polyprotein (3, 4). This released CP performs multiple functions in the virus life cycle including the formation of capsomers by intermolecular interactions with other CP monomers, encapsidation of the genomic RNA to form the nucleocapsid cores, and interaction with the cytoplasmic domain of glycoproteins that is essential for the virus budding process (5, 6, 7, 8, 9, 10).

The alphavirus CP consists of two major domains. The N-terminal domain is highly disordered, rich in basic amino acids, and functions during encapsidation of viral
genomic RNA (5, 11, 12, 13). The C-terminal domain possesses cis-autoproteolytic activity and is inactivated after cleavage of the scissile bond between Trp-Ser residues (AVCP residues: W267-S268). The Trp-Ser residues containing the scissile bond are conserved among all alphaviruses and are present at the C-terminal end of CP. After cleavage, the free carboxylic group of conserved Trp267 at P1 position interacts with the catalytic triad and remains bound in the S1 pocket of the CP. This bound carboxyl terminus tryptophan residue in the active site inactivates the protease and blocks further trans-cleavage activity (3, 14, 15, 16, 17). The active site molecular architecture and the catalytic triad are well conserved among all serine proteases including alphavirus CPs (AVCP: His144, Asp163 and Ser218) and exhibit similar spatial architecture in the active site (3, 17, 18). Additionally, the GDSG motif containing the active site nucleophilic serine residue that is well conserved in chymotrypsin-like serine proteases is also found to be completely conserved in the CP of alphaviruses (AVCP: 215GDSG219) (3, 14, 15, 17).

The CP is reported to not possess trans-activity since it is an auto-proteolytic enzyme that auto-inhibits to give a turn-over number of 1 (3, 18). However, investigations by Morillas et al. have revealed that truncations of 1-7 C-terminal residues of the SFV CP, including deletion of the ultimate C-terminal tryptophan, restores the enzymatic activity of protein (1). The truncated enzyme was found to possess high esterase activity after deletion of the highly conserved C-terminal tryptophan. However, until now the trans-cleavage protease activity of the alphavirus CP in which trans-cleavage of the peptide bond containing tryptophan at P1 position has not been demonstrated. Furthermore, they concluded from the study that correct folding of the
tertiary structure of SFV CP is dependent on the presence of conserved Trp residue at the
C-terminus (1). The findings of this study were intriguing and persuaded us to probe the
trans-peptidase activity and structure of the pre-cleavage form of another member of
alphavirus genus, AVCP.

The crystal structures of the post-cleavage form of CP having the C-terminal
tryptophan bound in the catalytic site have been reported from different members of the
alphavirus genus including Aura virus. Additionally, the crystal structure of catalytically
inactive S215A variant of Sindbis virus capsid protein (SCP) is also available, in which
two additional C-terminal P1' and P2' residues (Ser265 and Ala266) are also bound near
the active site (14). As the structure consists of substrate residues in the specificity
pocket, this structure is considered an enzyme-substrate complex form of CP in which the
catalytic triad residue Ser215 has been mutated to alanine to abolish the self-cleavage
activity. Until now, the structures available for the alphavirus CP were of the inactive
state with a blocked catalytic site (3, 15, 17, 19). Therefore, the development of structure-
based antiviral drug design strategies targeting the CP proteolytic activity has been absent
due to the unavailability of the structure of the unblocked active enzymatic form and the
lack of a CP protease assay.

The C-terminal domain of CP also participates in the budding process by
interacting with the glycoproteins through a hydrophobic pocket that lies on the capsid
surface (20). This hydrophobic pocket has also been proposed to be involved in capsid-
capsid interaction required for nucleocapsid formation (21). The crystal structure of SCP
containing a dioxane molecule in the hydrophobic pocket suggested the use of dioxane
derivatives for targeting the capsid and E2 glycoprotein interaction for antiviral
development (22). Some dioxane derivatives that prevent CP-glycoprotein interaction by binding to the CP hydrophobic pocket lead to defects in virus budding (23, 24). However, specific antiviral molecules that target the capsid protease activity and block the initiation of polyprotein processing have not yet been identified.

In this study, the structural analysis of the C-terminally truncated protease domain from Aura virus (AVCPΔ2) is described and a biochemical analysis of the trans-protease activity of the protein is reported. The high resolution atomic structure of AVCPΔ2, which represents the unblocked catalytically active form of enzyme, has been determined. This is the first report of the trans-proteolytic activity and the crystal structure of the active form of the alphavirus CP. We discuss the conformational changes in the protein catalytic site, the oxyanion hole and the substrate specificity pockets upon transition from catalytically active to fully processed and cleaved (subsequent to auto-catalytic cleavage) inactive protein via a substrate bound complex intermediate. The observation provides insight into the properly folded structure of the trans-active form and new possibilities for structure-based antiviral drug design targeting the protease activity of alphavirus CP.

**MATERIALS AND METHODS**

**Construction of expression plasmid.** The AVCPΔ2 construct (residue 110-265) was prepared in which the N-terminal disordered region and the last two residues at the C-terminus of the protease domain were deleted. Aura virus genomic cDNA was used as the template for polymerase chain reaction (PCR) amplification of the DNA fragments encoding AVCPΔ2. The oligonucleotides 5'-
CTGGAATT CATATGGCCCTGAAATTTGAAGCCGAC -3' (forward) and 5'-
CTAGAATCTC GAGCTATACAGTATCTTCGTGGGTGG -3' (reverse) containing
NdeI and XhoI sites, respectively, were used in the PCR reaction. These primers were
designed on the basis of GenBank accession no. NP_819015.1. The PCR amplified DNA
fragment was purified using the PCR purification kit (Qiagen, USA) according to the
manufacturer’s instructions. The purified PCR fragment and pET28c vector containing
the Tobacco etch virus (TEV) protease cleavage site were digested with NdeI and XhoI
restriction enzymes. The digested products were separated on a 1 % agarose gel and
purified using the DNA gel extraction kit (Qiagen, USA). Restriction enzyme digested
plasmid and PCR products were ligated using T4 DNA ligase. DH5α (DE3) cells were
transformed with the ligation mixture by the heat shock method (25). The transformed
cells were plated on Luria-Bertani (LB) agar plates containing 50 μg/ml kanamycin and
incubated overnight at 37 ºC. The obtained colonies were picked and grown overnight at
37°C in LB broth containing 50 μg/ml of kanamycin. Plasmids isolated from 5 ml culture
using a MiniPrep plasmid isolation kit (Qiagen, USA) were screened by PCR and were
assayed by restriction enzyme digestion for the presence of AVCPΔ2 insert. This insert
was sequenced in both directions using T7 forward and T7 reverse primers to confirm the
identity of the pET28c-AVCPΔ2 plasmid.

Expression and Purification of AVCPΔ2. For protein production, the cloned
recombinant pET28c-AVCPΔ2 plasmid was transformed into E. coli strain Rosetta
(DE3). Colonies of the transformed cells were grown in LB broth supplemented with
kanamycin (50 μg/ml) and chloramphenicol (35 μg/ml) at 37°C to an optical density of
0.4 at 600 nm (OD600). At this point, the temperature was reduced to 18°C and the culture
was allowed to grow to an OD_{600} of 0.8. Protein expression was then induced using 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the induced culture was grown overnight at 18°C. The cells were harvested by centrifugation and the obtained cell pellet was stored at -80°C till further use.

The purification procedure for AVCPΔ2 was very similar to the published protocol for the native AVCP (26). Briefly, the cell pellet from a 1 liter culture was resuspended on ice in 30 ml of purification buffer (50 mM Tris-HCl pH 7.6, 15 mM imidazole and 100 mM NaCl) and the cells were disrupted using a cell disruptor (Constant Systems Ltd, Daventry, England). The cell lysate was subjected to centrifugation at 4°C. The N-terminal His-tagged AVCPΔ2 was purified using Ni-NTA beads (Qiagen, USA) and elution was done in 250 mM imidazole. The N-terminal His-tag from AVCPΔ2 was cleaved by incubating the purified protein with TEV protease overnight and the sample was simultaneously dialyzed against the dialysis buffer (50 mM Tris-HCl pH 7.6, 20 mM NaCl) at 4°C. After His-tag cleavage, the protein sample was reloaded onto Ni-NTA column to remove His-tagged TEV protease and uncleaved His-tagged AVCPΔ2. The flow-through containing AVCPΔ2 protein without His-tag was concentrated and loaded onto pre-equilibrated HiLoad Superdex 75 16/60 size-exclusion chromatography column (GE Healthcare) using an ÄKTA purifier (GE Healthcare), which was operated at a flow rate of 0.5 ml/min at 4°C. Gel-filtration fractions were run on a 15% SDS-PAGE to analyze protein purity. The fractions containing pure protein were pooled and concentrated to 15 mg/ml using a 3 kDa cutoff Amicon Ultra-15 concentrator (Millipore, Bedford, Massachusetts, USA). The concentration and yield of purified protein was estimated by UV-Vis spectroscopy at 280 nm using an extinction
coefficient method. The yield of AVCPΔ2 from 1 liter bacterial culture was ~ 15 mg, which is slightly higher as compared to the native AVCP (26).

**Trans-protease activity assay.** The proteolytic activity assay for AVCPΔ2 was performed by a FRET-based proteolytic assay using 5-[(2'-aminoethyl)-amino] naphthelenesulfonic acid (EDANS) and 4-[[4-(dimethylamino) phenyl] azo] benzoic acid (Dabcyl) as the FRET pair. The peptide having Dabcyl and EDANS at the N and C termini, respectively, was procured from Biolinkk, New Delhi, India. The sequence of the substrate peptide was derived from the cleavage site of CP containing the conserved Trp-Ser scissile bond for protease cleavage. The cleavage analysis was performed in 20 mM HEPES buffer (pH 7.0) (reaction buffer) at room temperature. The purified AVCPΔ2 was incubated with the reaction buffer and the fluorogenic peptide substrate was added to a final concentration of 1 μM. The reaction was carried out in total volume of 1 ml in an eppendorf. At various time points, the reaction mixture was transferred to the fluorescence cuvette for steady state fluorescence measurements with a Fluorolog-3 Spectrofluorimeter LS55 (HORIBA Jobin Yvon Spex). The fluorogenic peptide substrate was excited at a wavelength of 340nm and the emission spectrum scanning from 440 to 600 nm was recorded. The increase in the fluorescence was monitored at specific time intervals. For kinetic studies, different peptide substrate concentrations ranging from 0.6 to 16 μM were used. The initial velocity (Vi) at different substrate concentrations was calculated. The kinetic parameters were calculated by Lineweaver Burk Plots.

**Far-UV circular dichroism spectrum.** To estimate the secondary structure of AVCPΔ2, circular dichroism (CD) analysis of the purified protein was performed using a PC controlled Applied Photophysics (Model Chirascan, UK) spectropolarimeter. CD
spectra were recorded in a quartz cell of 1 mm at a bandwidth of 1.0 nm between 180 to 260 nm and time per point 0.25 s at 25°C. The instrument was subjected to constant nitrogen purge. The CD experiment was performed using 0.1 mg/ml of purified protein in 20 mM potassium phosphate buffer at pH 7.5. Three scans were recorded and averaged. The baseline for the blank buffer was subtracted from the average protein spectrum and the secondary structure was determined by the online server Dichroweb using the CONTIN method (27).

**Crystallization and data collection.** Purified and concentrated protein was crystallized using the sitting drop vapor diffusion method. Protein crystallization was done in a 96 well sitting drop crystallization tray (Hampton Research, Aliso Viejo, CA) with 1 μl of protein (15 mg/ml in 50 mM Tris-HCl pH 7.6, 20 mM NaCl) and 1 μl of the reservoir solution equilibrated against 50 μl of the reservoir buffer. AVCPΔ2 crystals were obtained in 5 days using 0.2 M Sodium citrate tribasic dihydrate and 20 % w/v Polyethylene glycol (PEG) 3,350 at 20°C. The composition of cryo-protectants was also optimized for high resolution data collection. Prior to data collection, the obtained crystal was soaked in mother liquor containing 15 % (v/v) glycerol as cryoprotectant. X-ray data were collected using a MAR 345 imaging-plate system with Cu Kα radiation generated by a Bruker–Nonius Microstar H rotating-anode generator operated at 45 kV and 60 mA. The data were collected under cryogenic conditions at a wavelength of 1.54 Å. Diffraction data were processed and scaled using the MOSFLM package (28). The data collection and processing statistics are summarized in Table 1.

**Structure solution and refinement.** The structure determination was carried out using the molecular replacement method by taking the crystal structure of the native
AVCP that has the C-terminal tryptophan residue bound to the active site as the search model (PDB ID: 4AGK). For this purpose, the MOLREP program from the CCP4 software suite was used and Refmac5 program was used for the restrained refinement (29, 30). The analysis of the electron density map and manual model building were carried out using the COOT program (31). The stereo-chemical properties of the refined structure model of AVCPΔ2 were analyzed using the PROCHECK program (32).

Structural analysis of the refined model and the preparation of figures were done using the PyMOL visualization tool (33). For the analysis of dimer interface contacts, the PISA (Protein Interfaces, Surfaces and Assemblies) web server was used (34).

**Protein structure accession number.** Structure coordinates for the AVCPΔ2 have been deposited in the Protein Data Bank with accession code 4UON.

**RESULTS**

**Purification of active AVCPΔ2.** The carboxyl terminal Trp267 of AVCP remains bound to the S1 specificity pocket blocking the entry of substrate for trans-cleavage (17). This blocked enzyme represents the post-cleavage form of CP. Deletion of the conserved Trp267 is anticipated to free the blocked active site, making it accessible to the peptide substrate for trans-protease cleavage (Fig. 1A). This trans-active protease would represent the pre-cleavage state of auto-proteolytic alphavirus CP. Therefore, the AVCPΔ2 (110-265) construct with deleted C-terminal Glu266 and Trp267 residues from the AVCP domain (110-267) was designed and cloned in a bacterial expression vector. The recombinant AVCPΔ2 protein having a TEV protease removable 6x His affinity tag at the N-terminus was produced in soluble form in *E. coli*. Protein from the soluble
fraction was purified using Ni\(^{+2}\) affinity and size exclusion chromatography. The N-terminal His\(_6\)-tag was removed from the purified protein using TEV protease. SDS-PAGE analysis showed the presence of a single protein band of \(~17\) kDa confirming sample purity and homogeneity of purified AVCP\(\Delta\) (Fig. 1B). A standard curve using gel-filtration molecular weight markers was prepared and the average molecular mass of the major elution peak of purified protein was calculated using the calibration curve. The estimated molecular weight of the protein was calculated to be \(~17\) kDa suggesting that AVCP\(\Delta\) is a monomer in solution.

**Characterization of trans-protease activity.** For assessment of the trans-protease activity of AVCP\(\Delta\), the fluorogenic peptide substrate Dabcyl-Gly-Ala-Glu-Glu-Trp↓Ser-Leu-Ala-Ile-Glu(EDANS) was used. In this fluorometric assay, EDANS is the fluorophore donor (F) having an excitation wavelength at 340 nm and a maximum emission wavelength at 490 nm. DABCYL is the non-fluorescent quencher (Q) having a maximum absorption at 470-520 nm wavelength. The activity of AVCP\(\Delta\) was measured by FRET analysis at different time points. The donor fluorophore EDANS was excited at 340 nm and emission spectra were recorded by scanning from 440 to 600 nm. To perform this experiment a fixed concentration of substrate (1 \(\mu\)M) was used and the fluorescence data measurements were made at specific time intervals. An increase in the intensity of fluorescence with time corresponds to a decrease in the FRET signal, which demonstrates the cleavage of the peptide substrate (Fig. 1C). The effect of pH on the trans-protease activity of AVCP\(\Delta\) was determined by performing the fluorometric protease assay in the pH range 4.5 - 9.5. The pH optimal for the peptide substrate (DABCYL)-GAEEW↓SLAIE (EDANS) was 7.0 (data not shown). To determine the kinetic
parameters, the calculated Vi was plotted against different substrate concentrations (Fig. 1D). A Lineweaver-Burk Plot was used to determine the value of $K_M$ for the given fluorogenic peptide substrate. The value of $K_M$ was found to be $2.63 \pm 0.62 \mu M$ and the value for catalytic efficiency $K_{cat}/K_M$ for the AVCPΔ2 trans-protease activity was found to be $4.97 \times 10^4 M^{-1} \text{min}^{-1}$. The purified inactive form of AVCP containing Trp267 at the C-terminus that blocks the active site, was used as a negative control in all the experiments and showed negligible increase in fluorescence intensity over time (data not shown).

**Secondary structure analysis using CD.** The far-UV CD spectrum was used to determine the secondary structure content in the purified AVCPΔ2. The analysis of the CD data was performed using the CONTIN method on the Dichroweb web server. The deconvolution of data for AVCPΔ2 indicates a structural content of 2.6 % $\alpha$-helix, 40.5 % $\beta$-sheet, 19.2 % $\beta$-turn and 37.6 % random coil (data not shown). For comparative structural analysis, the CD data of native AVCP was also analyzed. The result was comparable to the secondary structure of native AVCP, which shows 2.8 % $\alpha$-helix, 42.1 % $\beta$-sheet, 19.1 % $\beta$-turn and 36 % random coil (data not shown). These results suggest that active AVCPΔ2 has an appropriate secondary structure, which is similar to that of native AVCP. Thus, it was expected that AVCPΔ2 possessing trans-capsid protease activity, would have a tertiary fold very similar to the inactive, post-cleavage state of AVCP.

**Crystallization of AVCPΔ2.** To determine the tertiary structure of AVCPΔ2, the purified protein was concentrated to ~ 15 mg/ml and crystallization attempts were made. Diffraction quality crystals were obtained by using 0.2 M Sodium citrate tribasic
dihydrate and 20 % w/v Polyethylene glycol 3,350 at 20°C. From single crystal, a complete data set was collected at 1.81 Å (Table 1). AVCPΔ2 crystals belong to monoclinic lattice with space group \( P2_1 \), however, inactive AVCP crystals reported earlier also belonged to the monoclinic lattice but in space group \( C2 \) (17). Two molecules per asymmetric unit are found in AVCPΔ2 whereas inactive AVCP has one monomer per asymmetric unit. The estimated Matthews coefficient was 3.15 Å³Da⁻¹ and solvent content was 60.97 % for two molecules of AVCPΔ2 per asymmetric unit.

**Overall structure of AVCPΔ2.** The structure of AVCPΔ2 was determined at 1.81 Å resolution and refined to a final Rfactor of 18.5 % and Rfree of 21.8 % with more than 95% residues in the most favored region of the Ramachandran plot. High resolution data show clear electron density for the N-terminal six residues, which were disordered and thus not visible in the native AVCP crystal structure (17). Similar to the other already known chymotrypsin-like serine proteases, the structure of each monomer of AVCPΔ2 consists of two sub-domains with each domain made up of 6-7 stranded, antiparallel, \( \beta \)-barrel structure as shown in Fig. 2A. The catalytic triad consisting of residues His144, Asp166 and Ser218 is placed at the interface of these two domains. The surface view of the monomer depicting different specificity pockets, the oxyanion hole and the active site is shown in Fig. 2B. The overall fold within each monomer of AVCPΔ2 is very similar to the native AVCP with root mean square deviation (RMSD) of 0.33 Å on Ca atoms. On the other hand, comparison of the native AVCP with the native AVCP with bound dioxane in the E2 binding hydrophobic pocket, showed a RMSD of 0.13 Å (17). This indicates that the main-chain conformational variations are more pronounced between the
native and active form as compared to the two forms (apo and ligand bound) of native AVCP.

At the C-terminus of AVCPΔ2, the electron density for the last four residues (262-265) is missing which indicates conformational flexibility of these residues in the absence of the last two residues (Fig. 2C). In native AVCP, the C-terminal loop takes a β turn at residue His261 to allow Trp267 to bind the substrate specificity pocket. In AVCPΔ2, Trp267 is truncated so the loop movement is flexible and no longer restricted due to the binding of Trp267 at the S₁ specificity pocket. The movement and flexibility of loop is crucial for proper positioning of the scissile peptide bond relative to the active site and oxyanion hole for cleavage. Thus, this loop movement towards the active site seems to play an important role in the binding of Trp267 to the active site and the auto-proteolytic property of CP. Moreover, the C-terminal loop, which is missing from unblocked form and found in the Trp bound native state, shows different conformations in two chains of the SCP substrate bound intermediate form (Fig. 2C) (14). This suggests that the loop is highly flexible in three different enzymatic states and becomes restrained in position after proteolytic cleavage.

In AVCPΔ2, the structure of active form of the enzyme, three glycerol molecules (used as cryoprotectant) are clearly visible in the electron density map of both the chains and show interactions with different amino acid residues. Superposition of the crystal structure of AVCP the inactive state, and AVCPΔ2 the active form, shows that a glycerol molecule binds exactly at the same position where Trp267 is present in the S₁ specificity pocket of AVCP. Another glycerol molecule in both chains of AVCPΔ2 occupies a cavity by making polar interactions with the main chain of Gly147 and Ile149, side chain
of Glu168 and two water molecules. Gly214 in chain A and Tyr131, His196, Asp217 and Arg220 in chain B interact with the third glycerol molecule through H-bonding. As glycerol has been proven to be a protein structure stabilizing agent, hence glycerol used in soaking the crystal before cryo-freezing and data collection might be influencing the structure of AVCPΔ2 through stabilization (35, 36, 37, 38).

Dimeric interactions. This study shows the monomeric nature of protein in solution and the crystallographic dimer in the asymmetric unit. In the AVCPΔ2 structure, both the monomers are in a tail-to-tail contact through their C-terminal sub-domains. Previously, a number of dimers and trimers were found in the SCP and the SFV CP that were arranged in head-to-tail and tail-to-tail fashion in different crystal forms (3, 14, 19). In head-to-tail dimers, the N- and C-terminal sub-domains are in contact as in the SCP S215A mutant crystal structure (14). In head-to-head dimer formation, the N-terminal sub-domains are in contact as in the SFV CP wild-type I crystal type (19).

The dimeric interface area in AVCPΔ2 as calculated using the PISA web server was found to be ~ 375 Å², comprising approximately 4.5 % of the total solvent accessible area for each monomer that was consistent with previous reports in SCP (14, 15). The dimer interaction involves two sheets VII and VIII (190GFYNW194 and 197GAVQFS202) and two loops (188PE189 and 224DNS226) of both the monomers. The two chains are held together by a few hydrogen bonds and mostly non-bonded interactions (14, 19).

These contacts are facilitated by the contribution of 11 residues from each subunit that are common to both the monomers. The residues Pro188, Glu189, Gly190, Phe191, Tyr192, Asn193, Gly197, Ala198, Asp224, Asn225 and Ser226 from both monomers...
comprise the dimer interface. These mainly involve hydrophobic interactions along with hydrogen bond formation by residue Asn225 of one monomer with Phe191 and Tyr192 residues of the other monomer (Fig. 2D). The hydrophobic interactions were found to play a more important role in dimer formation as compared to the hydrogen bonding. However, no evidence is available for the biological relevance of this dimer. Moreover, the mutations that lead to crystallographic dimer disruption show no adverse effect on viral replication (14). Thus, two molecules in the asymmetric unit do not represent a physiological dimer and this is consistent with the AVCP being a monomer in solution.

Conformational switching in the hydrophobic pocket. The electron density corresponding to the N-terminal residues from 110-115 (plus two extra residues at the N-terminal left after His-tag cleavage) is present in both chains of the AVCPΔ2 structure. These residues were disordered in the previously reported inactive AVCP and in the dioxane-bound AVCP structures (17). These residues in AVCPΔ2 form an arm that projects out of the structure. A detailed analysis of AVCPΔ2 structure shows that this N-terminal arm interacts and binds in the hydrophobic pocket which is formed between the two β-barrel sub-domains of the neighboring symmetry monomer molecule (Fig. 3A and 3B). The hydrophobic residues Leu111 and Phe113 of the N-terminal arm bind at the pocket between Tyr183, Trp250 and Phe169. Other residues that line this cavity are Met135, Glu136, Lys138, Ser162, and Met167. Fig. 3C shows various interactions that stabilize the binding of the N-terminal arm in this pocket. Earlier crystallographic reports of the alphavirus capsid have also reported that this hydrophobic pocket is occupied by similar binding of the hydrophobic N-terminal arm residues from neighboring CP molecules (14, 21).
Based on structural and mutational studies, it is hypothesized that this hydrophobic pocket plays a role at two stages in the virus life cycle. Firstly, the binding of the N-terminal arm residues to the hydrophobic pocket in crystal structures of two different alphaviruses including AVCPΔ2 and SCP suggests a biological significance of this interaction (14, 21). Furthermore, mutational studies of capsid residues 108 and 110 in the SCP have revealed the role of this interaction in capsid assembly (21). Thus, it is proposed that binding of the N-terminal arm in the hydrophobic pocket of neighboring CP contributes to the formation of nucleocapsid core by linking capsid proteins together before RNA binding and core assembly. Additionally, sequence analogy of cytoplasmic domain of E2 (cdE2) and the N-terminal arm of CP, mutational, modeling and structural studies have suggested that cdE2 and CP hydrophobic pocket interactions are crucial for virus budding process (17, 20, 22). A conformational switching in the hydrophobic pocket has been proposed that changes the role of this pocket from core assembly to virus budding (19).

The structural comparison of the hydrophobic pocket in the pre-cleavage state AVCPΔ2 (with bound N-terminal arm) and the post-cleavage state AVCP (with unoccupied hydrophobic pocket) visibly illustrate conformational changes in the hydrophobic pocket (Fig. 4A). It is presumed that the slightly wider pocket with the bound N-terminal arm in AVCPΔ2 represents the conformation likely to promote core assembly. Whereas the narrower pocket with somewhat two distinct sub-pockets in inactive AVCP is perhaps engaged in interaction with cdE2 desirable for virus budding. As shown in Fig 4B, minor variations in most of the side chains of residues (Glu136, Met135, Lys138, Ser162, Met167, and Phe169) lining the hydrophobic pocket direct this...
conformational variation. However, further investigations are required to confirm this
presumption of conformational switching in the hydrophobic pocket.

The catalytic triad. The alphavirus CP has a fold similar to the chymotrypsin-like serine proteases and the conserved catalytic triad residues Ser218, His144 and
Asp166 of AVCP form the active site at the interface between the two β-barrel sub-domains (Fig. 2A) (3, 14, 15, 17, 19). The polar interactions of Ser218 and His144 with
Trp267, and the complementary shape of the S₁ specificity pocket and side chain of
Trp267 likely fix the position of the completely conserved P₁ residue Trp267 in the active
site for scissile bond cleavage. In both chains of the AVCPΔ2 structure, which represents
the pre-cleavage state, Ser218 is found to be present in two alternate conformations
showing its flexibility, and gets fixed upon entry of substrate (Fig. 5A). However in the
post-cleavage state, represented by AVCP, Ser218 does not show flexibility as the
carboxyl terminal Trp267 produced on scissile bond cleavage makes polar contacts with
the Ser218 side chain (17). The Ser218 main chain shows displacement of only 0.4 Å
between the active pre-cleavage and the inactive post-cleavage states. His144 and
Asp166 show similar polar interactions with each other and with no alternate
conformations in various enzymatic states. Only a minor difference of 0.4-0.6 Å is
detectable in the side chain of His144. The Asp166 residue is partially exposed to the
solvent and interacts with two conserved water molecules in both enzymatic forms.

The oxyanion hole. The oxyanion hole, indispensable for proteolytic activity,
plays a role in the substrate binding and stabilization of the tetrahedral transition
intermediate containing the oxyanion through polar interactions with the main chain
amides of the conserved GDSG motif (39, 40). Interestingly, the crystal structure of
AVCPΔ2 shows altered conformation for its oxyanion hole (Fig 2C). This has resulted due to a flip of the peptide bond between Pro215 and Gly216 as shown by Fig. 5B. The oxyanion hole residues Gly216 and Gly219 show backbone conformational differences in both structure forms. A displacement of up to 2.9 Å can be observed in Gly213-Gly216 as shown in Fig 5B. Thus, it reveals the role of the oxyanion hole main chain in substrate binding. The Ser218 residue conformational change does not impart any effect on the alteration of the oxyanion hole conformation as can be seen from the mutant S215A crystal structure of SCP and an equivalent substitution in subtilisin (14, 41).

Superposition of crystal structures of AVCPΔ2, the active pre-cleavage state and AVCP, the inactive post-cleavage state, shows that the carboxylate oxygen atom of Trp267 in AVCP is just 1.9 Å away from the carboxyl oxygen of Pro215 in AVCPΔ2 indicative of steric hindrance in substrate binding. These conformational changes in residues from Gly213-Gly217 change the overall geometry of the specificity pocket as shown in the surface view in Fig. 5C. Hence, it can be hypothesized that region 213-217 has an altered conformation in the absence of substrate in the specificity pocket of the pre-cleaved state and changes during the substrate binding to prevent the steric hindrance between Pro215 and the incoming substrate.

The active form, AVCPΔ2 contains a water molecule in the oxyanion hole (Fig. 6A). Interestingly, this water molecule is conserved in Chain A of the SCP substrate bound mutant S215A, which is considered as an intermediate form (Fig. 6B), whereas in its Chain B, Gly213 (corresponds to Gly216 in AVCP) main chain NH atom forms an H-bond with carbonyl oxygen atom preceding scissile bond (Fig. 6C). This interaction was absent in Chain A of SCP substrate bound mutant S215A as the loop that brings the...
substrate inside the substrate binding pocket is highly flexible and doesn’t reach the oxyanion hole in Chain A. Post-cleavage AVCP also doesn’t contain water molecules in the oxyanion hole and shows similar interactions with the C-terminal carboxylate as that of Chain B of the SCP substrate-bound form (Fig. 6D). Previously, Chain A of substrate bound form in SCP was considered as early enzymatic state in which substrate is loosely bound and is more similar to chymotrypsin without substrate (14). Hence, the unblocked active AVCPΔ2 represents the pre-cleavage state of the alphavirus CP prior to cis-catalytic activity. Consequently, it can be concluded that the substrate scissile bond replaces a water molecule present in the oxyanion hole of AVCPΔ2 during proteolysis.

The substrate specificity pockets. For comparative structural analysis, the structure of AVCPΔ2 was superimposed onto the inactive AVCP as well as onto the substrate bound SCP mutant S215A structures. A series of conformational changes to variable extent in different specificity pockets are revealed. As seen in substrate bound SCP crystal structure, the substrate specificity pockets S1, S4, S1’, S2’ and S4’ make molecular contacts with the bound substrate residues: Trp264 the P1 residue, Thr261 the P4 residue, Ser265 the P1’ residue, Ala266 the P2’ residue and Pro268 the P4’ residue respectively (14). Inspection of these substrate specificity pockets in AVCP crystal structure indicates significant conformational changes in the main-chain backbone and also in the side chain of substrate binding residues (Fig. 7).

The most notable conformational changes are seen both in the backbone and amino acid side chains of the S1 specificity pocket. The S1 specificity pocket includes residues Trp194, Ile208-Ser218 (includes active site Ser residue), Ile232-Asn238 and Thr243-Leu245. The largest deviation between the active AVCPΔ2 and inactive AVCP is

21
in the main chain of residues Ile208-Gly216. The substrate bound SCP S215A mutant representing the intermediate form, shows a conformation rather similar to the C-terminal Trp bound inactive AVCP state. This confirms the conformational flexibility of the specificity pocket between active and inactive Trp bound forms. Gly213 and Gly214 show deviation up to 2.9 Å between AVCP and Chain A of active AVCPΔ2. The deviation is different between the two chains and found to be 3.4 Å between AVCP and Chain B of AVCPΔ2 (Fig. 2C, 7A). Moreover, as described above, Pro215 and Ser218 show side chain conformational flexibility in the active and inactive forms of AVCP.

Apart from some minor deviations found at region Ile232-Gly236 among the two structural forms, a displacement of up to 0.9 Å in main chain can be seen at Ala237-Thr243. This displacement may be significant, as this region lines the substrate specificity pocket. This difference can be accounted by the presence of the C-terminal residues in AVCP native structure. Due to the presence of Trp in the AVCP crystal structure bound at the S1 specificity pocket, the C-terminal residues are restricted and their interaction with neighboring residues forms a compact cavity. In AVCPΔ2, these interactions are absent, the C-terminal loop is flexible, and does not involve in such interactions. This suggests a flexible S1 specificity pocket in the absence of Trp267.

The S1 specificity pocket residue Leu234 in AVCP is conserved among alphavirus CPs. Leu234 main chain carbonyl O forms an H-bond with Ser218 side chains in all the different forms of the CP. This also forms an H-bond with the NH atom of the main chain preceding Trp267 (Fig. 7B). This suggests the importance of Leu234 in holding the catalytic triad at a proper position and providing a favorable connection between the active site and substrate. The Leu234 backbone shows deviation of less than
0.4 Å between the truncated and native states of AVCP, which could be due to the absence of Trp267 as it interacts with Leu234. Ser214 (conserved in other serine proteases) which corresponds to Leu234 in alphavirus CP, is very crucial for serine protease activity as it forms an H-bond with the catalytic triad residue Asp102 (chymotrypsin residue). Also mutational studies show a reduction in catalytic activity upon Ser214 mutation as the consequence of displacement of its backbone (42). Thus, the position at Leu234 is very important for substrate binding and hence crucial for catalysis.

Another region of substantial flexibility lies in the S4 specificity pocket that comprises of Ala237, Ser246 and His261 residues out of which Ala237 shows a backbone deviation of 0.6 Å while Ser246 shows side chain flexibility. In both forms, it is present in a dual conformation that contributes the major region of flexibility (Fig. 7C). This Ser246 side chain shows polar contacts with Leu234, Gly235, Val247, Thr259 and Thr264 in both inactive and active AVCP; one additional interaction with His261 is present in inactive form and is absent in the active AVCPΔ2. His261 also shows deviation in the main-chain backbone (0.8 Å) between the two forms of AVCP. It is probably due to the change in His261 position between different enzymatic states to bring the scissile bond near the active site. The P2 and P3 residues Glu266 and Val265 are present in a loop which is oriented outwards. Glu266 forms ionic interaction with catalytic triad residue His144. However, in the active AVCPΔ2 crystal structure Val265 is found to be disordered.

Likewise, the S1' pocket (including residues Met129, His144 and Val145) does not show much variation (~ 0.4 Å) in the backbone conformation. Ser265, the P1' residue in SCP substrate bound form shows H-bonding with Ile126, which corresponds to
Met129 in AVCP. Met129 shows polar interaction with the main chain of Val121. Again the significant variations lie in the $S_4'$ and $S_2'$ specificity pockets. The $S_4'$ pocket residues Asn123 and Lys127 are displaced from the position by 0.5 Å and 0.9 Å, respectively. This pocket consists of Asn120 and Asp124 in SCP, however the $P_4'$ residue is hydrophobic in both (Ile271 in AVCP and Pro268 in SCP). Ile128 and Met129 residues in $S_2'$ pocket shows significant deviation of 0.6 Å and 0.4 Å respectively (Fig. 7D). The $P_2'$ residue Ala266 in the SCP substrate bound form shows interaction with water molecules.

DISCUSSION

In last two decades extensive investigation on the alphavirus CP has been performed. Most of the structure-function analyses have focused on characterization of the inactive form of CP (C-terminal Trp bound form) using X-ray crystallographic, cryo-EM, mutational and biochemical studies (43, 44, 45). In 2008, the enzymatic characterization of the esterase activity of the truncated form of SFV CP was reported (1). Up until now, no study has reported in vitro trans-protease activity of the alphavirus CP. It was proposed that the conserved Trp residue at the C-terminus is required for proper folding and stabilization of CP. It was moreover proposed that the active capsid produced by truncating the conserved C-terminal Trp is natively an unfolded protein. As a result, the three dimensional structure of the active form of CP has not been determined.

As a consequence, the protease activity of alphavirus CP has not been targeted for antiviral drug discovery. In this study, recombinant active AVCP in which two C-terminal $P_2$ and $P_1$ residues Glu266 and Trp267 have been removed (AVCPΔ2), was
produced and purified. Purified protein was further used for characterization of the trans-protease activity. A fluorometric assay was developed for screening of protease inhibitors against alphavirus capsid protease. The kinetic studies showed $K_{cat}/K_M$ value of $4.97 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ for the trans-protease activity of AVCPΔ2. This result indicates that the protease domain of truncated CP is fully active and possesses catalytic properties similar to other chymotrypsin-like serine proteases.

The crystal structure of this active form of the Aura virus CP (AVCPΔ2) has been determined. This is the active pre-cleavage state preceding cis-catalytic activity. Thus, there are three different forms of the alphavirus CP: the active state (this work), the substrate bound intermediate form (14) and the inactive C-terminal Trp-bound form (17).

It is likely that the protein undergoes conformational rearrangements during structural polyprotein processing, viral replication and budding. The significant conformational changes in the three different forms have been analyzed and reported in this study. The novel structure of the trans-active form of AVCPΔ2 provides structural insight into the catalytic site and substrate binding pockets revealing a range of conformational flexibilities. These conformational variations should be taken in account for structure based drug designing of anti-alphaviral molecules targeting CP activity.

Additionally, conformational rearrangements are observed by comparison of the pre-cleavage and the post-cleavage states of AVCP in the CP hydrophobic pocket (20, 21). The conformational changes may be responsible for the switching mechanism of the hydrophobic pocket in binding to N-terminal arm for core assembly and then its interaction with cdE2 glycoprotein during virus budding.
Structural comparisons show that the specificity site pockets are highly flexible. The active site residue Ser218 side chain is in a dynamic state and exhibits different conformations. The Leu234 backbone has an important role in substrate binding and maintaining the position of the catalytic triad. The oxyanion hole contains a water molecule in the truncated form and Chain A of substrate bound form; however, it is missing from Chain B of the substrate bound form and native enzymatic state. This might be due to the formation of an H-bond between the scissile bond and the oxyanion hole residues during proteolysis. Moreover, the last C-terminal segment is highly flexible until the substrate attaches to the pocket and undergoes proteolysis. This loop is missing in the unblocked form, flexible in the substrate bound form and static in C-terminal Trp-bound native state. Knowledge of the structure of the truncated unblocked alphavirus CP allows for a comprehensive analysis of capsid structures at different stages. The presence of glycerol in the active site of AVCPΔ2 where the P$_1$ residue Trp267 binds opens the possibility of designing glycerol-based alphavirus CP inhibitors. The reported structure highlights the previously unseen conformational changes and provides a foundation for structure-based design of antiviral compounds that will block the initial step of alphavirus structural polyprotein processing.

ACKNOWLEDGEMENTS

This research was supported by Defence Research and Development Organization (DRDO), Government of India. We thank the Macromolecular Crystallographic Facility (MCU) at IIC, IIT Roorkee for data collection, structure determination and refinement. MA thanks CSIR for financial support. R.J.K. acknowledges support from the NIH.
through NIGMS award GM56279. We thank Todd W. Geders for reading the manuscript critically and for helpful suggestions.

REFERENCES


**FIGURE LEGENDS**

**FIG 1** Purification and trans-proteolytic activity analysis for AVCPΔ2. (A) Schematic representation of inactive and active forms of alphavirus CP. In the inactive form Trp267 shows auto-inhibitory action on enzyme activity after autoproteolysis (left). In the absence of Trp267, the protein acquires its catalytic property as the substrate can easily access the active site (right). (B) Gel filtration chromatography and SDS-PAGE analysis of major peak fractions show the protein (AVCPΔ2) purified to homogeneity. The protein is monomeric in nature as determined by gel filtration profile. (C) Analysis of In vitro trans-proteolytic activity of AVCPΔ2. The enzymatic cleavage assay was carried out using 15 μg of protein in HEPES buffer (20 mM, pH 7.0) by the addition of 1 μM fluorogenic peptide substrate at room temperature. The hydrolysis of the peptide substrate was measured at different time intervals. The excitation was done at 340 nm and emission spectrum for each time point is shown as scan from 450 nm to 600 nm. (D) The initial velocity (μM/min) was calculated for increasing concentration of the substrate. Different substrate concentrations ranging from 0.6 μM to 16 μM were used (shown on X-axis). Experiment was done in triplicate and the values represent the average data. All the data were normalized using the same reaction without the enzyme.
FIG 2 Crystal structure of AVCPΔ2 and its comparison with other enzymatic forms of
the alphavirus CP. (A) Overall structure of the monomer contains two β-barrel
subdomains consisting of the catalytic triad in between the cleft. The catalytic triad
residues are presented in sticks. The coloring is done on the basis of secondary structure.
(B) Surface view of AVCPΔ2 showing different pockets and regions involved in
catalysis, coded with different colors. (C) Both chains of active (Chain A in green and
Chain B in yellow color) and substrate-bound form (Chain A in blue and Chain B in
gray) were aligned along with native AVCP (pink). Circle 1 (black) shows the variation
in the C-terminal region in two chains of substrate bound form, while this region is absent
from the active AVCP and remains intact in native AVCP. Circle 2 (blue) shows the loop
flexibility in S1 specificity pocket. (D) Close-up view of the dimeric interaction in
AVCPΔ2. The residues involved in crystallographic dimer formation are shown in sticks.
Asn225 of one chain forms H-bonds with Phe191 and Tyr192 of the other chain. Chain B
residues are labeled with an apostrophe sign. Chain A is shown in green while chain B is
in yellow color.

FIG 3 The hydrophobic pocket of AVCPΔ2 with the bound N-terminal arm of
neighboring molecule. (A) The crystal structure of AVCPΔ2 dimer with the symmetry
related molecules. The hydrophobic pocket of each molecule occupied with the N-
terminal arm of the neighboring symmetry related molecule is shown. The dotted square
highlights one such binding. (B) The zoomed view of hydrophobic pocket of one of the
AVCPΔ2 subunit with bound N-terminal arm of the neighboring subunit is displayed in
the surface view and (C) The cartoon view with all the polar interactions shown with dotted lines and the interacting water molecules are shown as spheres.

**FIG 4** Structural comparison of the hydrophobic pocket between native and active AVCP (A) The surface view and (B) the superposed view of the hydrophobic pocket of active AVCPΔ2 (green in color) over inactive AVCP (pink in color) crystal structure to show the differences in the side chain conformation of some of the residues that occur at this hydrophobic pocket to accommodate the N-terminal arm of the neighboring subunit.

**FIG 5** Comparative studies of native and active AVCP for structural differences in the catalytic triad and S₁ specificity pocket. (A) The catalytic triad residues from different enzymatic states: native AVCP (pink), active AVCPΔ2 Chain A (green) and AVCPΔ2 Chain B (yellow) show conformational change in Ser218 side chain. The catalytic triad residues and Trp267 are shown as sticks. The interactions of Ser218 and His144 with Trp267 are shown as red dash line. (B) The differences at the S₁ specificity pocket of AVCPΔ2 (green in color) are shown in superposition with native AVCP (pink in color) crystal structure with bound Trp267 at the active site. The dotted red lines highlight the difference at region Gly213-Gly216. The red colored dotted box highlights the flipped peptide bond at Pro215 and Gly216. (C) The differences in the volume of S₁ specificity pocket encircled by a dotted red circle are shown in surface view.

**FIG 6** Oxyanion hole of different enzymatic forms of alphavirus CP. (A and B) Oxyanion hole of AVCPΔ2 and Chain A of substrate bound SCP occupy a water
molecule. However, (C and D) represent the H-bonding of oxyanion hole residues with carbonyl oxygen atom preceding scissile bond in chain B of substrate bound form and native AVCP respectively.

**FIG 7** Structural comparison of substrate specificity pockets of native and active AVCP.

(A) Specificity pocket S₁ of AVCPΔ2 chain A (green) and chain B (yellow) is compared with that of native AVCP (pink). The main chain backbone at Gly213, Gly214 region is showing major difference in both chain A (2.9 Å) and chain B (3.4 Å) as compared to native form. Gly236 interacts with Val265, Asn238 with Thr210 and Leu245 with His261 in the native form which are absent in AVCPΔ2. (B) Leu234 main chain forms H-bond with Ser218 as well as Trp267. (C) Comparison of the S₄ pocket shows deviation in backbone as well as side chains. Ser246 interaction with His261 and Thr264 is found in inactive AVCP, but absent from the active one. (D) S₄' and S₂' pockets comparison in active and inactive AVCP demonstrates the displacement of 0.5, 0.9 and 0.6 Å in Asn123, Lys127 and Ile128 respectively.

**Table 1.** Data collection and refinement statistics for the active form of AVCPΔ2.

<table>
<thead>
<tr>
<th>Crystallographic Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P21</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a=53.56, b=71.35, c=61.51</td>
</tr>
<tr>
<td></td>
<td>α=90.00, β=114.33, γ=90.00</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>56.08 – 1.81</td>
</tr>
<tr>
<td>Completeness (%) (Last Shell)</td>
<td>97.6 (83.6)</td>
</tr>
<tr>
<td>Rmerge† (%) (Last Shell)</td>
<td>5.6 (51.0)</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Mean $I/\sigma(I)$ (Last Shell)</td>
<td>15.1 (3.0)</td>
</tr>
<tr>
<td>No. of observed reflections (Last Shell)</td>
<td>147227 (17067)</td>
</tr>
<tr>
<td>No. of unique reflections (Last Shell)</td>
<td>38673 (4783)</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>Matthews coefficient ($\text{Å}^3 \text{Da}^{-1}$)</td>
<td>3.15</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>60.97</td>
</tr>
<tr>
<td>Multiplicity (Last Shell)</td>
<td>3.8 (3.6)</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Residues</td>
<td>306</td>
</tr>
<tr>
<td>Water molecules</td>
<td>400</td>
</tr>
<tr>
<td>$R_{\text{cryst}}$ (%)</td>
<td>18.5</td>
</tr>
<tr>
<td>$R_{\text{free}}$ (%)</td>
<td>21.8</td>
</tr>
<tr>
<td>Average $B$-factor ($\text{Å}^2$)</td>
<td>23.01</td>
</tr>
<tr>
<td>r.m.s.d on bond lengths (Å)</td>
<td>0.013</td>
</tr>
<tr>
<td>r.m.s.d on bond angles (Å)</td>
<td>1.434</td>
</tr>
</tbody>
</table>

**Ramachandran plot (%)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Favoured</td>
<td>95.5</td>
</tr>
<tr>
<td>Allowed</td>
<td>4.5</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.0</td>
</tr>
</tbody>
</table>

† $R_{\text{merge}} = \frac{\sum_{\text{hkl}} |I(hkl) - \langle I(hkl) \rangle|}{\sum_{\text{hkl}} I(hkl)}$, where $I(hkl)$ is the $i$th observation of reflection $hkl$ and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations $i$ of reflection $hkl$. 

38