Conformational Plasticity of 2A Proteinase from Enterovirus 71

Qixu Cai†, Muhammad Yameen†¶, Weihua Liu†, Zhenting Gao‡, Yaozong Li‡,
Xuanjia Peng‡, Yaxian Cai‡, Caiming Wu¹, Qian Zheng¹, Jian Li²#, Tianwei Lin¹#  

¹State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, China; ²State Key Laboratory on Lead Compound Research, Wuxi AppTec Co., Ltd., Shanghai, China.  

#To whom correspondence should be addressed.  

Tianwei Lin: Address: Xiang’an South Road, Xiang’an District, Xiamen 361102, China. Tel.: +86-592-2184992; Fax: +86-592-2184992. Email: twlin@xmu.edu.cn  

Jian Li: Address: 288 Fute Zhong Road, Waigaoqiao Free Trade Zone, Shanghai 200131, China. Tel: +86-21-50464625; Fax: +86-21-50461000. Email: jian_li@wuxiapptec.com  

†These authors contribute equally to this work.  

¶current address: Department of Chemistry, Government College University Faisalabad, Punjab, Pakistan  

Running title: The Crystal Structure of EV71 2A Proteinase  

Word count: abstract: 190; text: 4699.
Abstract

2A proteinase (2A<sub>pro</sub>) is an enterovirally encoded cysteine protease that plays essential roles in both the processing of viral precursor polyprotein and the hijack of host cell translation and other processes in the virus life cycle. The crystallographic studies of 2A<sub>pro</sub> from enterovirus 71 and its interaction with the substrate are reported here. EV71 2A<sub>pro</sub> was comprised of an N-terminal domain of four-stranded antiparallel $\beta$-sheet and a C-terminal domain of six-stranded antiparallel $\beta$-barrel with a tightly bound zinc atom. Unlike other 2A<sub>pro</sub> structures, there is an open cleft across the surface of the protein in an open conformation. As demonstrated by the crystallographic studies and modeling of the complex structure, the open cleft could be fitted with the substrate. On comparison 2A<sub>pro</sub> of EV71 to those of the human rhinovirus 2 and coxsackievirus B4, the open conformation could be closed with a hinge motion in the bII2 and cII $\beta$-strands. This was supported by molecular dynamic simulation. The structural variation among different 2A<sub>pro</sub> structures implicates a conformational flexibility in the substrate-binding cleft. The open structure provides an accessible framework for the design and development of therapeutics against the viral target.
Introduction

Enteroviruses are a family of single stranded, positive sense RNA virus of Picornaviridae (1). Although most of the enterovirus-associated diseases are mild and asymptomatic, some members in the family can cause severe diseases and death, especially in the young and immunocompromised. Enteroviruses are the leading cause of aseptic meningitis which in turn is the most common infection in the central nervous system (2). Enterovirus 71 (EV71) is an important pathogen besides polioviruses of the family. It is emerging as the most significant neurotropic enterovirus in some area of the world in outbreaks and epidemics of hand, foot and mouth disease (HFMD) (3-5). This virus circulates in US and 26% of the adults tested in a study had antibody (6). An epidemic of enterovirus 71 infection in 1998 in Taiwan resulted in over 400 cases of neurologic diseases and an estimated of 1.5 million of HFMD and herpangina (7). The outbreaks of EV71-associated diseases have also been reported in the United States, Australia, Sweden, Japan, Bulgaria, Hungary, Hong Kong, Malaysia, Vietnam, and China (8-16). It has been associated with a variety of clinical diseases, including HFMD, herpangina, aseptic meningitis, encephalitis, and even fatal pulmonary edema or hemorrhage (17). Enterovirus can also cause severe chronic diseases, such as dermatomyositis, polymyositis, dilated cardiomyopathy, and diabetes mellitus (2).
Upon infection, a polyprotein is translated from the single open reading frame in the genome of an enterovirus and is processed into mature proteins by virally encoded 2A and 3C/3CD proteinases (3C\textsuperscript{pro}/3CD\textsuperscript{pro}) (1). 2A protease (2A\textsuperscript{pro}) is an enzyme that cleaves at its own N-terminus, at the junction between VP1 and 2A of the polyprotein. Besides its essential role in processing the viral proteins, 2A\textsuperscript{pro} shuts off the host’s cap-dependent protein production by cleaving the elongation factors eIF4GI/II for the synthesis of viral proteins (18, 19). The 2A\textsuperscript{pro} of coxsackievirus B3 and other similar viruses in the enterovirus family cleaves dystrophin, a cytoskeletal protein, and is directly linked to dilated cardiomyopathy (20). 2A\textsuperscript{pro} from polioviruses also interferes with the nuclear traffic (21-23) and hijacks the splicing and transcription machinery (24). It is of great importance to understand the structural basis of 2A\textsuperscript{pro} function to assist in the design and development of therapeutics for the treatment of associated diseases. However, there is a gap in our knowledge of 2A\textsuperscript{pro}, as its current structure reveals no space for substrate-binding. Herein we determined the crystal structures of EV71 2A\textsuperscript{pro} and its complex with the substrate. The structure revealed an open substrate-binding cleft connecting from S5 site to the active site, then extending to a possible S3’ site. Comparative studies with other 2A\textsuperscript{pro} and molecular dynamic simulation indicated that the open conformation could be closed with a hinge motion in the bII2 and cII β strands. The open cleft provides an accessible template for docking chemical compounds and implicates a useful system for structure-based design and development of therapeutics to treat associated diseases.
Materials and Methods

Protein Expression and Purification

The cDNA encoding EV71 2A<sup>pro</sup> was synthesized based on EV71 strain E2004104-TW-CDC (GenBank Accession No. EF373576). It was amplified by PCR and inserted into pGEX-4T-1 vector (GE Healthcare) to generate a plasmid containing the coding sequence of N-terminal GST fusion 2A<sup>pro</sup>. A single mutation was made at the active site nucleophile (C110A). The cDNA encoding 2A<sup>pro</sup> substrate peptide corresponding to the VP1-2A cleavage junction (ITTL*GKFG) was also inserted in a pGEX-4T-1 vector (GE Healthcare) to generate a plasmid containing the coding sequence of N-terminal GST fusion 2A<sup>pro</sup> substrate. Following verification of its DNA sequence, *Escherichia coli* strain BL21 (DE3) pLysS (Novagen) was transformed with the vector pGEX-4T-1-2A<sup>pro</sup> and pGEX-4T-1-2A<sub>sub</sub>. The bacterial cultures were grown at 37°C to an OD<sub>600</sub> of 0.6, and 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) was added to induce the 2A<sup>pro</sup> or its substrate expression at 20°C for 12 hours. Bacteria were subsequently harvested. For the purification of EV71 2A<sup>pro</sup>, the bacteria were lysed by ultrasonication in lysis buffer (PBS, pH 7.4). The lysate was centrifuged at 36,000g for 30 min to remove cell debris. The supernatant was loaded onto a glutathione sepharose column (GE Healthcare). After washing with PBS pH 7.4 for several column volumes, the GST tag was removed by thrombin in PBS at 22°C.
overnight. Superdex 75 16/60 column (GE Healthcare) was pre-equilibrated with a buffer containing 50mM Tris-HCl pH8.0, and then the 2A<sup>pro</sup> was loaded onto the column and the fractions containing 2A<sup>pro</sup> were collected. The protein was concentrated to 10mg/mL by a Centricon Centrifugal Filter Unit (MWCO 3,000) for crystallization. For the purification of EV71 2A<sup>pro</sup> substrate, the GST fusion 2A<sup>pro</sup> substrate was loaded onto a glutathione sepharose column (GE Healthcare) and the GST tag was removed by thrombin. The 2A<sup>pro</sup> substrate was collected and mixed with the purified 2A<sup>pro</sup> at a molar ratio of 1:3 for two hours. Then the complex was concentrated to 10mg/mL by a Centricon Centrifugal Filter Unit (MWCO 3,000) for crystallization.

For enzymatic activity assay, the native sequence of EV71 2A<sup>pro</sup> was inserted in pET-28a vector (Novagen) to generate a plasmid containing the coding sequence of the protein with a C-terminal 6×His tag. The vector was transformed into *Escherichia coli* strain BL21 (DE3) pLysS (Novagen) for protein expression with the same protocol for pGEX-4T-1-2A<sup>pro</sup>. The cells were harvested and lysed in PBS pH7.4, and the supernatant was loaded onto a Ni-NTA column, and eluted with 20mM Sodium phosphate, pH7.4, 500mM NaCl, 150mM Imidazole. The protein solution was concentrated to 6mg/mL and stored in -80°C for enzymatic activity assay.
Enzymatic Activity Assay

A fluorescent peptide with the sequence of Dabcyl–KSRTAITTLGKFGQQSGE-Edans was employed as the substrate for the 2Apro enzymatic and inhibition assays based on the fluorescence resonance energy transfer (FRET) effect. The sequence is derived from the VP1/2A junction of the polypeptide. The fluorescence was monitored at 500 nm with the excitation wavelength set at 340 nm using a fluorescence spectrophotometer (Cary Eclipse). The experiments were carried out in 20mM Tris-HCl, pH 7.5, and 100mM NaCl at 30°C. The measurements in the first 5 min were used to calculate the initial rates. The program of GraphPad Prism 5 (GraphPad Software, Inc.) was used to calculate $K_M$ and 95% confidence intervals. The $K_M$ value was expressed as the best-fit value ±SEM.

Protein Crystallization

Crystallization of EV71 apo-2Apro and complexes with substrate were carried out at 16 °C by hanging-drop vapor-diffusion. The 2Apro crystals were made with a well solution containing 0.1M HEPES, pH 7.5, 20% 2-propanol, and 10% PEG4000.
Substrate peptides were generated for crystallographic studies of complex structures. The sequence for the recombinant peptide was GSITTLGKFG. The sequences for the ten synthetic peptides of eleven residues were:

- SRTAITTLGKF
- SRTAITTMGKF
- SRTAITTRGKF
- SRTAITFLGKF
- SRTAITFMGKF
- SRTTITTLGKF
- SRTTITMGKF
- SRTTITTRGKF
- SRTTITFLGKF
- SRTTITTRGKF

The crystals of 2Apro/substrate complex were obtained in conditions similar to those of apo-2Apro.

X-ray Data Collection, Processing and Structure Determination

The crystals were transferred into a solution containing 0.1M HEPES, pH7.5, 20% 2-propanol, 10% PEG4000, and 25% PEG400 as cryoprotectant prior to flash cooling to 100 K for data collection. X-ray diffraction data were collected using a Rigaku rotating anode X-ray generator (λ = 1.5418 Å), equipped with a MAR Research 345mm imaging plate and processed with the automar package from Marresearch GmbH. Diffraction data were also obtained under synchrotron radiation, but the quality was not significantly improved and the resolution was not extended.

The structure of EV71 2Apro was determined by molecular replacement, employing the program Phaser (25) with the crystal structure of HRV2 2Apro (PDB ID: 2HRV) as
the search model. Manual model building and refinement were performed with Coot (26), CNS (27, 28), and Refmac5 (29, 30). To determine the $2A^{\text{pro}}$/substrate complex structure by molecular replacement, the refined $2A^{\text{pro}}$ structure was used as the initial phasing model. The substrate is located in the 2mFo-DFc electron density map. The complex structure was refined with Refmac5. The statistics for data collection and refinement are in Table 1. The presentations were made with PyMOL (31).

Modeling the Interaction between $2A^{\text{pro}}$ and its Substrate

The catalytic triads were superposed between the complex structure of EV71 $2A^{\text{pro}}$, with the partial substrate and the crystal structure of a $3C^{\text{pro}}$ from coxsackievirus A16 (CVA16), in complex with its substrate of 11 amino acid residues (P10-P1) (Protein Data Bank accession code: 3SJ9). Poly(Ala) of the P6-P3 moiety from the CVA16 $3C^{\text{pro}}$ complex structure was built onto the substrate sequence (P2-P2') in the EV71 $2A^{\text{pro}}$. After the residues were mutated to those of the native substrate sequence for EV71 $2A^{\text{pro}}$, an initial model for the P6-P2' substrate was generated by Schrödinger Prime (Schrödinger) in the substrate-binding cleft of the EV71 $2A^{\text{pro}}$ structure. The model complex structure was further optimized using Prime (Schrödinger) MM-GBSA with default settings that freeze $2A^{\text{pro}}$ atoms, but allow the substrate atoms to adjust during the energy minimization.

Accession Numbers
The coordinates of the crystal structures were deposited into the Protein Data Bank. The accession numbers are 4FVB and 4FVD.

Results

The Kinetic Properties of EV71 2Apro

The over-expressed and purified EV71 2Apro was employed to investigate the kinetic properties of the enzyme by a method based on fluorescence resonance energy transfer (FRET) (32). Derived from the VP1/2A junction of the polypeptide (Fig. 1A), the substrate is a fluorescent peptide with the sequence of Dabcyl-KSRTAITTLGKFGQQSGE-Edans. The optimal condition for the catalytic activity was identified in a two dimensional grid of varying both pH and NaCl concentrations. As shown in Fig. 1B, the best condition for the proteolytic activity of EV71 2Apro is 20mM Tris-HCl, pH7.5, and 100mM NaCl at 30°C. A $K_M$ of $6.46 \pm 0.77$ µM was derived and the 95% confidence interval of $K_M$ is in between 4.82 and 8.10 µM (Fig. 1C).

The Structure of EV71 2Apro

The sequence identity for 2Apro of EV71 and coxsackievirus A16, two leading causative agents in HFMD, is 97%. That between EV71 and other coxsackieviruses A and B, including that of echoviruses, is about 75%. Polioviruses have a lower identity of about 60%. In contrast, the sequence identity with 2Apro of human rhinoviruses,
another group of picornaviruses encoding the enzyme, is below 40%. The sequences around the active site and zinc-binding residues are highly conserved among 2A²⁰³ from different viruses (Fig. 2A).

An EV71 2A²⁰⁷ mutant, in which the active site Cys110 was mutated to Ala, was crystallized at pH 7.5, the optimal pH. The crystal structure was determined to 1.90 Å resolution by molecular replacement with 2A²⁰⁸ from human rhinovirus 2 (HRV2) (33) as the initial phasing model (Table 1). The electron density is well defined for residues 7 to 144. The fold of EV71 2A²⁰⁸ is similar to those of HRV2 2A²⁰⁹ (33) and coxsackievirus B4 (CVB4) (34) with an N-terminal domain of four-stranded antiparallel β-sheet with a small α helix, which is a modified chymotrypsin fold (33), and a C-terminal domain of six-stranded antiparallel β-barrel with a tightly bound zinc atom (Fig. 2B).

There is a cleft across the surface of EV71 2A²⁰⁸ with a constriction at one end where the active site residues are located in the current structure (Fig. 3A). This is in contrast to the other structures of 2A²⁰⁸. The active site of EV71 2A²⁰⁸ consists of a catalytic triad including His21 (His18 in HRV2 2A²⁰⁹ and His21 in CBV4 2A²⁰⁹), Asp39 (Asp35 in HRV2 2A²⁰⁹ and Asp39 in CBV4 2A²⁰⁹), and Cys110 (Cys106 in HRV2 2A²⁰⁹ and Cys110 in CBV4 2A²⁰⁹) (Fig. 2). His21 functions as the general base, while
Cys110 is the nucleophile which is replaced by Ala in this structure. The active site residues can be superposed well between EV71 and HRV2 2A\textsuperscript{pro} (Fig. 2C).

In the Protein Data Bank, CVB4 2A\textsuperscript{pro} structure is represented by 17 structures (PDB ID: 1Z8R). The root-mean-square-deviation (RMSD) between Ca atoms in the 17 structures is in between 1.6 to 3.8 Å. Most comparative studies for CVB4 2A\textsuperscript{pro} herein are made with the first set of coordinates in 1Z8R. The RMSD between the Ca atoms of CVB4 and EV71 2A\textsuperscript{pro} is 9.0 Å if the structures are aligned by the LSQ routine in coot (26). Better alignment could be obtained with the SSM (secondary-structure matching) routine (35) in coot and the corresponding core RMSD is 1.35 Å, which is the RMSD between the aligned Ca atoms. For comparison, the RMSD between the Ca atoms of HRV2 and EV71 2A\textsuperscript{pro} is 3.7 Å by LSQ and the core RMSD is 1.14 Å by SSM. It is indicative that the core structures of CVB4 and EV71 2A\textsuperscript{pro} are similar.

The variations in CVB4 2A\textsuperscript{pro} is a reflection of the dynamic nature of the enzyme (34). Since one of the catalytic triad, Cys110, were in a region that was not well defined (34), no further comparison on the active site residues from CVB4 2A\textsuperscript{pro} was made.

The cleft is long enough to accommodate P5-P1 residues of the substrate, using the nomenclature by Schechter and Berger (36), and connected to the S’ sites across the active site (Fig. 3A). If the active site is placed on the right, the south rim of the cleft is formed by the C-terminal domain with the active site residue Cys110 under the rim
in the narrow part of the cleft. While the N-terminal domain forms the northeast floor
of the cleft along with the active residues of His21 and Asp39, the north rim is formed
by the bII2-cII loop from the C-terminal domain (Fig. 2B and 3A). The S1 site is a
shallow binding pocket with the oxyanion hole (33) underneath. It appears to be able
to accommodate a side chain of moderate size. In the substrate-binding cleft, the S2
site is considerably larger and deeper under the north rim. This is not obvious in the
other two 2A<sup>pro</sup> structures where the bII2-cII loop moves southward to compress the
space (33, 34) (Fig. 3A and 3C). Although there is a clear opening in the
substrate-binding cleft, it is still significantly narrower than the S1 and S2 sites of
3C<sup>pro</sup>, the other cysteine protease encoded in the EV71 genome (Fig. 3A and 4A).
This is so because the bII2-cII loop is folded over the S2 site and active residues of
His21 and Asp39, while the β-ribbon at the similar location in 3C<sup>pro</sup> does not (37) (Fig.
3A, 4A and 4B). If the substrate-binding cleft in EV71 2A<sup>pro</sup> is described as a canyon,
then the S1, S2, S1’ and active sites in 3C<sup>pro</sup>, form a basin. The distance from S1 to S2
across the canyon is about 10 Å (Fig. 3A) while the distance across the basin in the
north-south direction is about 20 Å (Fig. 4A). It is indicative that the inhibitors
targeting the active site of 3C<sup>pro</sup> might not be suitable for 2A<sup>pro</sup>. The narrow cleft
around S1 and S2 site, as well as across the active site, implicates a better chance of
fitting non-covalent inhibitors in 2A<sup>pro</sup>. The S3, S4 and S5 sites are characterized by
general like-features that can be fitted with amino acid sidechains along a groove.
Another significant feature of the EV71 2A<sup>pro</sup> substrate-binding is the uneven, but
accessible potholed contour surrounding the S3'-possible site (Fig. 3B), which is absent in 3Cpro (37) (Fig. 4). This difference is due to the absence of a similar αI-βI loop in 3Cpro, which covers the region (Fig. 4B).

An Open Conformation

Despite the overall similarity in fold, the north rim of the substrate-binding clefts of EV71, HRV2, and CVB4 2Apro, reveals striking differences in conformation (Fig. 3A and 3C). If the C-terminal domains are superposed, the bI2-cI loop of EV71 2Apro in the north rim would have to slide southward to overlap with the loops in HRV2 and CVB4 2Apro. In addition, the tip of the loop would have to bend further downward to those of CVB4 2Apro where the substrate-binding cleft is all but inaccessible to substrate (Fig. 3A and 3C). The distance between the tips of the bI2-cI loops in EV71 and HRV2 2Apro is about 3 Å. A similar distance between those in EV71 and CVB4 2Apro is about 7 Å. The movement of the bI2-cI loop originates at a hinge motion of the bI2 and cI β strands, which unfolded part of the fully extended β structures in EV71 2Apro to make a kink in HRV2 and CVB4 2Apro (Fig. 3A and 3C). The consequence of this closure of the cleft is that the substrate peptides can no longer be fitted into the current HRV2 and CVB4 2Apro structures because of space constraint (Fig. 3A). In light of the fact that the bI2-cI loop of HRV2 2Apro is also variable between the two copies of the enzyme in the asymmetric unit of the crystal and the residues in this region are of higher temperature factors (with an average of...
about 65 Å$^2$ for Cα atoms) and weaker electron density than other regions of HRV2
2A$^{pro}$ structure (33) and CVB4 2A$^{pro}$ is a substrate-free solution structure, it is
conceivable that the bII2-cII loop of 2A$^{pro}$ is a region susceptible to conformational
change in response to the substrate binding, with the conformations found in the
structures of HRV2 and CVB4 2A$^{pro}$ for substrate-free state, while the conformation
found in EV71 2A$^{pro}$ represents an open conformation for binding the substrate, which
is supported by the structural characterization of 2A$^{pro}$/substrate complex (following).
It is known that crystal packing has induced conformational changes in the structural
elements directly involved in the inter-molecular interaction in 3C$^{pro}$ (37-40). The
molecular contacts between EV71 2A$^{pro}$ in the crystal were investigated. There are
two types of inter-molecular contacts in EV71 2A$^{pro}$ crystal (Fig. 5A). Neither of
these contacts was directly associated with the bII2-cII loop. One of the contacts
interacted with the N-terminal domain although it was in the vicinity of the bII2-cII
loop in the C-terminal domain (Fig. 5A). The closest distance between the bII2-cII
loop and the neighboring N-terminal domain is 5 Å. The other contact involved the
direct protrusion of the neighboring His25 into the substrate-binding cleft but it
interacted at the south, rather than the north rim of the cleft (Fig. 5A). In addition, its
dislodgement by the substrate did not alter the opening of the cleft (Fig. 5B). This
indicated that the adoption of a different conformation by the bII2-cII loop was not
the artifact of crystal packing. It might be argued that this interaction with the
neighboring His25 could represent an initial event in the 2A pro interaction with substrate, however, in which the binding of a sidechain from the substrate into the cleft would alter the dynamics of the structure to favor the open conformation and to make the cleft fully accessible to the substrate.

Molecular Dynamic Simulation

To evaluate the flexibility of the bII2-cII loop, an 8 nanosecond (ns) molecular dynamic (MD) simulation was carried out. EV71 2A pro was placed in a periodic water box containing 150mM NaCl to generate the MD trajectory. The RMSD for the Cα atoms, with respect to the first MD frame obtained after solvent equilibration, shows that the system became thermodynamically stable after approximately 3 ns in the MD simulation (Fig. 6A). The Cα distance between residues Glu88 and Pro107 was used as a measurement for the opening of the substrate-binding cleft. Glu88 is the residue at the tip of the bII2-cII loop. Pro107 is at the other side of the cleft in EV71 2A pro. This distance varied dynamically in between 8.5~16 Å in a periodical manner due to the movement of the bII2-cII loop (Fig. 6B).

The MD simulation with the same setting was also carried out for the HRV2 2A pro (PDB ID: 2HRV). The result shows the HRV2 2A pro is much more stable than EV71 2A pro (Fig. 6C), although the movement between Cα of the equivalent Glu84 and Pro103 could be as big as 12 Å. The Cα distance between residues Glu88 and Pro107
in the EV71 2Apro crystal structure is 11.1 Å, while the compatible distance in HRV2 2Apro crystal structure is 8.2 Å (Fig. 6D).

Interaction with the Substrate

Eleven substrate peptides were employed for the crystallization of the complex. After many trials, a complex was obtained by co-crystallization with the Cys110Ala mutant and an octopeptide derived from the VP1-2Apro junction of the precursor polypeptide. The structure was determined to 1.66 Å resolution in the same space group of the peptide-free enzyme. The EV71 2Apro structure in the complex closely resembles the apo structure with an RMSD of 0.24 Å. The average B factor was also similar for the bII2-cII loop in about 25 Å² for Cα atoms. The substrate interaction dislodged but did not completely displace the interaction from the neighboring His25 in the substrate-binding cleft (Fig. 5B). Probably as a consequence of this His25 interaction that interferes with the substrate binding, the density was not fully connected (Fig. 7A), and only 4 residues of the substrate with a sequence of TL/GK were modeled in the electron density at S2-S2’ sites (Fig. 7A). P2 residue, a Thr, forms hydrogen bonds with residues Arg93 and Gln95, and is placed snugly in a pocket which can accommodate sidechains larger than that of Thr. S1 is a shallow binding site, and its specificity towards Leu seems mainly due to space constraint around it. However, the space is open and can accommodate other residues in the upward direction, which probably explains the lack of specificity at S1 (41). A Gly residue at the P1’ site is
probably due to the fact that the active site of the enzyme is highly constricted, and only the smallest residues can be accommodated. No interaction could be identified for P2’, probably because the electron density for its side chain is not well defined (Fig. 7A and 7B).

**P6-P2’ Substrate Binding Model**

Only partial structure of the substrate was visualized in the complex. To get a better understanding of the 2A\textsuperscript{pro} interaction with the substrate, a model is generated for the full peptide binding at the cleft. Since both 2A\textsuperscript{pro} and 3C\textsuperscript{pro} are cysteine proteinases, it is conceivable that the structure of P3-P6 moiety for 2A\textsuperscript{pro} could take reference from the counterpart in 3C\textsuperscript{pro}. The complex structures of EV71 2A\textsuperscript{pro}, with the tetrapeptide substrate and the complex structure between coxsackievirus A16 (CVA16) 3C\textsuperscript{pro} and its P1-P10 substrate (PDB ID: 3SJ9), were superposed on the catalytic triads and the P1-P2 moieties. The poly(Ala) derived from the P6-P3 moiety in 3C\textsuperscript{pro} was built onto the P2-P2’ substrate structure in the 2A\textsuperscript{pro} complex structure and the initial model was made with the sequence of TAITTL/GK, the sequence at the P1/P2 junction of the polyprotein. After energy minimization, the P6-P2’ substrate model of EV71 2A\textsuperscript{pro} was generated (Fig. 8A). It is apparent that there are only small changes in the P1-P2 structure of the substrate in 2A\textsuperscript{pro}. This suggests that the addition of other residues and the removal of His25 from the crystal packing would induce minimal perturbation to the peptide. The model substrate fit reasonably well into the substrate-binding cleft.
without any unfavorable geometry and clash between enzyme and the substrate (Fig. 8A). The modeling of the substrate peptide onto the substrate-binding cleft in HRV2 2A<sup>pro</sup> would result in clashes between the peptide and the enzyme, especially around P2 (Fig. 8B). It is indicative that there would be a conformational change in the substrate-binding cleft of 2A<sup>pro</sup> for the recognition of the substrate.

**Discussion**

2A<sup>pro</sup> plays multiple roles in the enterovirus infection. On one hand, it is a vital component in the life cycle of the virus, as it cleaves its own N-terminus during the virus replication. On the other hand, it is essential for the viral usurpation or interference of host functions, such as the shutdown of cap-dependent translation of host proteins to facilitate the cap-independent translation of viral polyprotein (18, 19), the cleavage of dystrophin by that of coxsakievirus B3 (20), as well as the inhibition of nuclear traffic and splicing and transcription machineries by poliovirus 2A<sup>pro</sup> (21-24). It is important to understand the structural basis of 2A<sup>pro</sup> function and develop a framework for the design of therapeutics.

There are characteristic differences in the substrate-binding clefts between the two proteinases encoded in the enterovirus genome. This provides the structural basis for designing specific 2A<sup>pro</sup> and 3C<sup>pro</sup> inhibitors. The β-ribbon in 3C<sup>pro</sup> is peninsula-like with S2, S3 and S4 sites around it, but does not fold over the S2 site (37). This results...
in a more exposed architecture around the active sites of S1 and S2. This more exposed architecture might be the reason that P2 is highly variable among 3C<sub>pro</sub> substrates (42), and the reason for more tolerance to chemical groups taking different binding positions at S2, without diminishing the potency of a 3C<sub>pro</sub> inhibitor (37). In contrast, the folding of bII2-cII loop in 2A<sub>pro</sub> over the S2 site and active site residues makes the binding cleft more confined.

In this crystal structure, a neighboring His<sub>25</sub> binds inside the substrate-binding cleft, but not at the bII2-cII loop. It is not unreasonable to suggest that this interaction from a neighboring His<sub>25</sub> might be analogous to an initial event in the substrate interaction with the enzyme. As the bII2-cII loop can vary among different conformations, as was demonstrated by MD simulation and CVB4 2A<sub>pro</sub> structure, an interaction by the sidechain of the substrate peptide in the south rim of the substrate-binding cleft would tip the balance and facilitate the opening of the substrate-binding cleft, which lends support to the notion that the substrate binding induces the conformational change.

The more confined substrate-binding cleft as compared to that in 3C<sub>pro</sub> provides more constraint for fitting chemical compounds and a better chance to develop non-covalent inhibitors targeting EV71 2A<sub>pro</sub>. Another distinct feature of the EV71 2A<sub>pro</sub> structure, is the potholed contour surrounding the S3’-possible site, which can be targeted for fitting compounds bridged across the active site. As there are no
specifically designed 2A\(^{pro}\) activity-inhibitors to date, the EV71 2A\(^{pro}\) structure provides a framework for the rational design and development of therapeutics to successfully target enterovirus-associated diseases.

Acknowledgments

The research was supported by the funding from National Basic Research Program of China (973 program, 2012CB724500), from the Open Research Fund of the State Key Laboratory of Cellular Stress Biology, Xiamen University (SKLCSB2012KF003), and from the 111 Project of Education of China (NO. B06016). The crystallographic data collection at Beamline BL17U1 at Shanghai Synchrotron Radiation Facility is gratefully acknowledged.
References


development of Coot. Acta crystallographica. Section D, Biological

Nature protocols 2:2728-2733.


Figures legends

FIG. 1. Enzymatic kinetic evaluation of EV71 2A<sup>pro</sup>. (A) Diagramatic representation of EV71 polyprotein. The substrate peptide sequence for 2A<sup>pro</sup> was derived from the VP1/2A junction of the polypeptide. The cleavage site for 2A<sup>pro</sup> is indicated by a star (*) and the cleavage sites for EV71 3C<sup>pro</sup> are indicated by triangles (△). (B) Optimization of the proteolytic reaction conditions. Two dimensional grid searches by varying both pH and NaCl concentrations were carried out. The optimal buffer condition is 20mM Tris-HCl, pH 7.5, and 100mM NaCl. (C) Measurement of K<sub>M</sub> value for EV71 2A<sup>pro</sup> at 20mM Tris-HCl pH 7.5, and 100mM NaCl. Each data point was measured in triplicates. A K<sub>M</sub> of 6.46 ± 0.77 µM was derived and the 95% confidence interval of K<sub>M</sub> is in between 4.82 and 8.10 µM.
FIG. 2. Sequence and structure of 2A\textsuperscript{pro}. (A) Multiple sequence alignment of 2A\textsuperscript{pro}. All sequences were from UniProt database. EV71 represents human enterovirus 71; CVA16 stands for coxsackievirus A16; CVA9 is coxsackievirus A9; CVB3 is coxsackievirus B3; EC1 is echovirus 1; PV1 is poliovirus 1 and HRV2 stands for human rhinovirus 2. The residues identical among all viruses are in red, and those identical among enteroviruses are in purple. The active site residues are in boxes. The residues in interaction with zinc are indicated by a star underneath. The sequence identities are indicated. The secondary structures assignments for EV71 2A\textsuperscript{pro} are shown on top of the alignment, and those for HRV2 2A\textsuperscript{pro} are at the bottom. (B) Ribbon diagram of EV71 2A\textsuperscript{pro}. The N-terminal domain is in green and the C-terminal domain is in blue. The zinc ion is represented as a purple sphere. The active site residues are shown. (C) The superposition of the active site residues between EV71 and HRV2 (cyan) 2A\textsuperscript{pro}. The residue numbering is based on the sequence of EV71 2A\textsuperscript{pro}. Residue 110 of the catalytic triad in EV71 2A\textsuperscript{pro} was mutated to Ala, while the corresponding residue in HRV2 2A\textsuperscript{pro} is Cys.
FIG. 3. Comparison of 2A<sup>pro</sup>. (A) Surface rendering of 2A<sup>pro</sup>. The N-terminal domain is in green and the C-terminal domain, except the bII2-cII loop, is in blue. (Left) EV71 2A<sup>pro</sup>. There is an extensive cleft across the surface of the enzyme for substrate binding with a constriction at the east end. The bII2-cII loop (yellow) forms the north part of the substrate-binding cleft with catalytic residues under the constriction. The width of the cleft is about 10 Å. (Middle) HRV2 2A<sup>pro</sup>. The bII2-cII loop (red orange) moves southward and the cleft becomes narrower and is closed over the constriction. The substrate peptides can no longer be fitted. (Right) CVB4 2A<sup>pro</sup>. The bII2-cII loop (magenta) blocks the access to the cleft from exterior and the substrate cannot be fitted. (B) A 90° view of EV71 2A<sup>pro</sup> from (A). A possible S3’ site is of a defined cavity and, together with S2’ site, well suited for fitting the small molecules. (C) Superposition of bII2 and cII β strands and bII2-cII loop in ribbon diagram. Parts of the bII2 and cII β strands in EV71 2A<sup>pro</sup> change to coil-like structures in HRV2 and CVB4 2A<sup>pro</sup> (circled), with color coding as in A. As a result, this part of the structure behaves like a hinge that moves the bII2-cII loop southward. The tip of bII2-cII loop, in CVB4 2A<sup>pro</sup>, bends further downward.
FIG. 4. Comparison between EV71 2A<sub>pro</sub> and 3C<sub>pro</sub>. (A) Surface rendering of EV71 3C<sub>pro</sub>. The N-terminal domain is in green and the C-terminal domain is in blue. The active, S1, S2 and S1’ sites form a basin. The distance in the north-south crossing of the basin is about 20 Å. (B) Comparison of the active site residues and substrate-binding clefts between 2A<sub>pro</sub> and 3C<sub>pro</sub> in ribbon diagrams. The 3C<sub>pro</sub> is in light gray. The bII2-cII loop in 2A<sub>pro</sub> folds over the S2 site while aI-bI loop in 3C<sub>pro</sub> covers the S’ sites.
FIG. 5. Inter-molecular interactions in the EV71 2A pro crystal. (A) There are 2 molecular contacts in 2A pro crystal. The neighboring molecule in red orange ribbons interacts with the N-terminal domain although it is close to the bII2-cII loop. The closest distance between the bII2-cII loop and the neighboring N-terminal domain is 5Å. The molecule in cyan ribbons interacts with the south rim of the substrate-binding cleft with residues His25 protruding into the binding cleft. (B) Binding of the substrate (in sticks) dislodges the neighboring His25 but there is no movement in neither the south or north rim of the substrate-binding cleft. The area of this view is boxed in A.
FIG. 6. The cleft width varies during MD simulation. (A) Variation of RMSD for Cα atoms during the simulation. The system reaches equilibrium in 3 ns. (B) The distances between Cα atoms of Glu88 and Pro107 during the simulation. A smooth line is created to highlight the trend. (C) The distance variations between Cα atoms of Glu84 and Pro103, two residues across the clefts in Chain A (left) and Chain B (right), of the HRV2 2Apro crystal structures during MD simulation. Both systems are more stable than EV71 2Apro, although the change can be over 12 Å. (D) The distance between Cα atoms of Glu84 and Pro103 across the cleft in HRV2 2Apro (orange) is 8.2 Å and that of Glu88 and Pro107 in EV71 2Apro (yellow) is 11.1 Å in the crystal structures. During the MD simulation, the shortest distance between the Cα atoms of Glu88 and Pro107 of EV71 2Apro is 8.5 Å at 2.41 ns (cyan) and the longest distance is 16 Å at 6.38 ns (blue).
FIG. 7. Complex structure of EV71 2A\(^{pro}\) and the substrate in stereo views. (A)
The electrostatic surface potential of EV71 2A\(^{pro}\) contoured at between -10 kT/e (red) and +10 kT/e (blue) with the substrate and its electron density map (2mF\(_{o}\)-DF\(_{e}\) at 0.7 \(\sigma\) in blue chickenwire). (B) Interactions between EV71 2A\(^{pro}\) and the substrate. The hydrogen bonds or ionic interactions are indicated as dash lines.
FIG. 8. Model complex structures. (A) Modeling of the full substrate peptide in the binding cleft. Cα atoms in the model substrate are in cyan and those in the crystal structure are in yellow. P2-P2' of the model substrate matches closely to the crystal structure. The space constraint towards the exterior of the cleft is open for P1, which might explain the large variation among P1 residues. (B) Modeling of substrate peptide onto HRV2 2A\textsuperscript{pro}. There are steric clashes around P2 site, highlighted by an oval. In order to bind the substrate in a similar fashion to EV71 2A\textsuperscript{pro}, the conformation of the bII2-cII loop would have to change.
Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>2Apro</th>
<th>2Apro/Substrate complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) (Å)</td>
<td>86.613</td>
<td>86.566</td>
</tr>
<tr>
<td>(b) (Å)</td>
<td>44.265</td>
<td>43.991</td>
</tr>
<tr>
<td>(c) (Å)</td>
<td>52.036</td>
<td>51.764</td>
</tr>
<tr>
<td>(\beta) (°)</td>
<td>112.58</td>
<td>112.21</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-1.90 (1.97-1.90)</td>
<td>50-1.66 (1.72-1.66)</td>
</tr>
<tr>
<td>(R_{sym}) or (R_{merge})</td>
<td>0.0345 (0.2617)</td>
<td>0.0386 (0.4443)</td>
</tr>
<tr>
<td>(I / I_0)</td>
<td>11.9 (2.1)</td>
<td>11.5 (1.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>92.4 (96.0)</td>
<td>93.3% (85.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>1.93 (1.84)</td>
<td>3.66 (3.29)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.9</td>
<td>1.66</td>
</tr>
<tr>
<td>No. reflections</td>
<td>12668</td>
<td>18996</td>
</tr>
<tr>
<td>(R_{work} / R_{free})</td>
<td>0.194/0.240</td>
<td>0.190/0.226</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1069</td>
<td>1098</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Water</td>
<td>115</td>
<td>133</td>
</tr>
<tr>
<td>B-factors (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.3</td>
<td>26.3</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>38.0</td>
<td>27.0</td>
</tr>
<tr>
<td>Water</td>
<td>36.3</td>
<td>37.4</td>
</tr>
<tr>
<td><strong>Ramachandran statistics (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored</td>
<td>93.38</td>
<td>97.08</td>
</tr>
<tr>
<td>Allowed</td>
<td>6.62</td>
<td>2.19</td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>R.m.s. deviations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.022</td>
<td>0.028</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>2.02</td>
<td>2.10</td>
</tr>
</tbody>
</table>

Values in parentheses are for highest-resolution shell.