Crystal structure of the 3C protease from South African Territories type 2 foot-and-mouth disease virus

Jingjie Yang, Eoin N Leen, Francois F Maree, Stephen Curry

The replication of foot-and-mouth disease virus (FMDV) is dependent on the virus-encoded 3C protease ($3C^{pro}$). As in other picornaviruses, $3C^{pro}$ performs most of the proteolytic processing of the polyprotein expressed from the single open reading frame in the RNA genome of the virus. Previous work revealed that the $3C^{pro}$ from serotype A –one of the seven serotypes of FMDV – adopts a trypsin-like fold. Phylogenetically the FMDV serotypes are grouped into two clusters, with O, A, C, and Asia 1 in one, and the three South African Territories serotypes, (SAT-1, SAT-2 and SAT-3) in another. We report here the cloning, expression and purification of 3C proteases from four SAT serotype viruses (SAT2/GHA/8/91, SAT1/NIG/5/81, SAT1/UGA/1/97, and SAT2/ZIM/7/83) and the crystal structure at 3.2Å resolution of $3C^{pro}$ from SAT2/GHA/8/91).

Crystal structure of the 3C protease from South African Territories type 2 foot-and-mouth disease virus

Jingjie Yang[1], Eoin N. Leen[1#], François F. Maree[2] and Stephen Curry[1*]

[1]Department of Life Sciences, Sir Ernst Chain Building, Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom.

[2]Transboundary Animal Disease Programme, Agricultural Research Council, Onderstepoort Veterinary Institute, Private Bag X05, Onderstepoort, 0110, South Africa.

[3]Department of Microbiology and Plant Pathology, Faculty of Agricultural and Natural Sciences, University of Pretoria, Pretoria, 0002, South Africa.

[#]Present address: Astbury Building, University of Leeds, Woodhouse, Leeds LS2 9LU, UK

*Corresponding author: s.curry@imperial.ac.uk

Keywords: foot-and-mouth disease virus, 3C protease, crystal structure, structural conservation, drug target.

18

ABSTRACT

The replication of foot-and-mouth disease virus (FMDV) is dependent on the virus-encoded 3C protease ($3C^{pro}$). As in other picornaviruses, $3C^{pro}$ performs most of the proteolytic processing of the polyprotein expressed from the single open reading frame in the RNA genome of the virus. Previous work revealed that the $3C^{pro}$ from serotype A – one of the seven serotypes of FMDV – adopts a trypsin-like fold. Phylogenetically the FMDV serotypes are grouped into two clusters, with O, A, C, and Asia 1 in one, and the three South African Territories serotypes, (SAT-1, SAT-2 and SAT-3) in another. We report here the cloning, expression and purification of 3C proteases from four SAT serotype viruses (SAT2/GHA/8/91, SAT1/NIG/5/81, SAT1/UGA/1/97, and SAT2/ZIM/7/83) and the crystal structure at 3.2 Å resolution of $3C^{pro}$ from SAT2/GHA/8/91).

INTRODUCTION

Diseases caused by RNA viruses are often difficult to control because of the high mutation rate and the continual emergence of novel genetic and antigenic variants that escape from immune surveillance. The degree to which immunity induced by one virus is effective against another is largely dependent on the antigenic differences between them. Foot-and-mouth disease virus (FMDV) is an example of an antigenically variable pathogen that infects many species of cloven-hoofed animals, such as cattle, sheep, pigs and goats, and remains a potent threat to agricultural livestock (Sutmoller et al., 2003). Although FMD vaccines made from chemically inactivated virus particles are in widespread use, control of the disease remains difficult. This is because the vaccines provide only short-lived protection and the virus occurs as seven clinically indistinguishable serotpyes (O, A, C, Asia1 and three South African Territories serotypes: SAT1, SAT2 and SAT3), each of which has multiple, constantly evolving sub-types (Knowles & Samuel, 2003). Viruses belonging to the SAT serotypes display appreciably greater genomic and antigenic variation compare to serotype A and O viruses (Bastos et al., 2001; Bastos et al., 2003; Maree et al. 2011), possibly due to their long term maintenance within African buffalo (*Syncerus caffer*). Constant surveillance of circulating strains is required to ensure that vaccine stocks remain effective.

In common with other members of the picornavirus family, FMDV has a single-stranded, positive-sense RNA genome with just one open reading frame. Cell entry in infected hosts is followed immediately by translation of the viral RNA, which yields a polyprotein precursor of over 2,000 amino acids that has to be processed into fourteen distinct capsid and non-

54 structural proteins for virus replication. The majority of this processing is done by the
55 virus-encoded 3C protease ($3C^{pro}$), which cleaves the precursor at ten distinct sites
56 (reviewed in Curry et al., 2007).

57 Crystallographic analysis of the $3C^{pro}$ from a type A FMDV (sub-type $A10_{61}$) showed that,
58 similar to other picornavirus 3C proteases, it adopts a trypsin-like fold consisting of two β-
59 barrels that pack together to create a centrally-located Cys-His-Asp/Glu catalytic triad in
60 the active site (Allaire et al., 1994; Matthews et al., 1994; Mosimann et al., 1997; Birtley et
61 al., 2005; Yin et al., 2005). Subsequent studies on FMDV $3C^{pro}$ complexed with peptides
62 derived from the viral polyprotein work revealed that substrate recognition is achieved by
63 conformational changes primarily involving the movement of a β-ribbon (residues 138-
64 150) that helps to secure the position of cognate peptides in relation to the active site of the
65 protein (Sweeney et al., 2007; Zunszain et al., 2010).

66 Mapping of the sequence variation between different FMDV serotypes onto the structure of
67 $A10_{61}$ $3C^{pro}$ indicated that the peptide-binding face of the protease is completely conserved
68 among the non-SAT serotypes (which are 91-97% conserved in amino-acid sequence),
69 supporting the notion that identification of inhibitors of the protease might aid the
70 development of broad spectrum antiviral drugs (Birtley et al., 2005; Curry et al., 2007).
71 This structure should therefore serve as a useful model for the 3C protease from this group
72 of viruses. However, the same comparison suggested the presence of at least two amino
73 acid differences on the peptide-binding surfaces between $A10_{61}$ $3C^{pro}$ and the
74 corresponding 3C sequences from SAT serotype viruses.

To provide a more complete picture of the structural variation between FMDV 3C proteases from different serotypes, we set out to determine the crystal structure of $3C^{pro}$ from at least one SAT serotype virus. We report here the cloning and expression of $3C^{pro}$ from four distinct SAT1 and SAT2 viruses and the crystal structure of the $3C^{pro}$ from a SAT2 serotype virus (SAT2/GHA/8/91).

MATERIALS AND METHODS

Cloning and mutagenesis: We used the polymerase chain reaction (PCR) to amplify the coding regions for the FMDV 3C proteases of sub-types SAT2/GHA/8/91 (Accession No. AY884136), SAT1/NIG/5/81 (Accession No. AY882592), SAT1/UGA/1/97 (Accession No. AF283456), and SAT2/ZIM/7/83 (Accession No. AF540910). In each case the reaction was performed using DNA primers (Table 1) that introduced 5´ *Xho*I and a 3´ *Hind*III restriction sites into the PCR products. These served to facilitate ligation into a version of the pETM-11 vector that had been modified to insert a thrombin cleavage site immediately downstream of the N-terminal His tag (Birtley & Curry, 2005). DNA ligations were performed using the Roche Rapid Ligation Kit according to the manufacturer's instructions.

Site-directed mutagenesis was performed with the Quikchange method (Stratagene), using KOD polymerase (Novagen). All DNA sequences were verified by sequencing.

Protein Expression and Purification: All SAT-type 3C proteases were expressed in cultures of BL21 (DE3) pLysS *E. coli* (Invitrogen) grown in lysogeny broth (LB) at 37°C with shaking at 225 rpm. Protein expression was induced for 5 hours by the addition of 1 mM

isopropyl β-d-1-thiogalactopyranoside (IPTG) once the optical density at 600 nm reached 0.8-1.0. Cells were harvested by centrifugation at 4550 g for 15 min at 4°C and frozen at –80 °C.

The volumes given below are appropriate for processing the pellet from 1 L of bacterial culture. Cell pellets were thawed on ice and re-suspended in 30 mL Buffer A (50 mM HEPES pH7.1, 400 mM NaCl, 1 mM β-mercaptoethanol) supplemented with 0.1% Triton X-100 and 1 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. Cells were lysed by sonication on ice and lysates clarified by centrifugation at 29,000 g for 20 min at 4°C. Protamine Sulphate (Sigma) was added to 1 mg/ml final concentration to precipitate nucleic acids, and lysates were then centrifuged again at 29,000 g for 20 min. The supernatant was filtered using a 1.2 μm syringe filter and incubated for 90 minutes at 4°C with slow rotation in 1 mL bed volume of TALON metal affinity resin (Clontech) pre-equilibrated with buffer A. This slurry was applied to a gravity-flow column and the TALON beads washed three times with 50 mL of Buffer A supplemented with 0, 5 and 10 mM imidazole respectively. His-tagged 3C proteins were eluted in 20 ml of Buffer A containing 100 mM imidazole, followed by a final wash with 10 ml of Buffer A containing 250 mM imidazole. To remove the His tag the eluted protein was mixed with 100 units of bovine thrombin (Sigma) and dialysed for 16 hr at 4°C in 4 L of Buffer A supplemented with 2 mM $CaCl_2$. Cleaved protein was then re-applied to TALON resin to remove the cleaved His tag and other contaminants. The untagged protease was recovered in the flow through, concentrated using Vivaspin concentraters (3 kD MWCO) (Sartorius Stedim Biotech) and further purified by gel filtration using HiLoad 16/60 Superdex 75 gel filtration column (Amersham Bioscience) in Buffer A supplemented with 1 mM EDTA and 0.01% sodium

azide at flow rate of 0.5 ml/min. Peak fractions were pooled, concentrated and stored at –80 °C. Protein concentrations were determined from absorbance measurements at 280 nm using extinction coefficients calculated with the ProtParam tool (Gastiger et al., 2005).

Crystallisation and structure determination: Crystallisation trials with purified SAT-type $3C^{pro}$ were performed at 4°C and 18°C using protein concentrations in the range 5-10 mg/mL. Initial screens were done by sitting drop vapour diffusion using a Mosquito crystallisation robot (TTP Labtech). Typically in each drop 100 nl of protein was mixed with 100nl taken from the 100 µL reservoir solution. Trials were performed with the following commercial screens: Crystal screen 1 and 2, and PEG/Ion (Hampton Research); Memstart, Memcys, JCSG+, and PACT (Molecular Dimensions); Wizard 1 and 2 (Rigaku Reagents).

Crystals of g3C-SAT2-G(1-208) for data collection were washed in the mother liquor (15% (w/v) PEG-8000, 0.09 M Na-cacodylate pH 7.0, 0.27 M Ca-acetate, 0.01 M Tris pH 8.5, 0.08 M Na-thiocyanate) supplemented with 20% (v/v) glycerol, and immediately frozen in liquid nitrogen in a nylon loop. X-ray diffraction data were processed and scaled with the CCP4 program suite (Collaborative Computer Project No. 4, 1994), and phased by molecular replacement using the coordinates of type $A10_{61}$ FMDV $3C^{pro}$ (PDB ID 2j92; (Sweeney et al., 2007)) as a search model in Phaser (McCoy et al., 2007). The search model was edited to delete side-chains (to the C_β atom) for all residues that differed with g3C-SAT2-G(1-208) and to remove all the atoms in the β-ribbon (residues 138-150), since these have been observed to vary in structure between different crystal forms (Sweeney et al., 2007). Model building and adjustments were done using Coot (Emsley et al., 2010);

crystallographic refinement was performed initially with CNS (Brünger et al., 1998) and completed using Phenix (Adams et al., 2010).

RESULTS AND DISCUSSION

Protein expression and crystallisation: We engineered bacterial expression plasmids for FMDV 3C proteases from four SAT sub-types: SAT2/GHA/8/91, SAT1/NIG/5/81, SAT1/UGA/1/97, and SAT2/ZIM/7/83 (see Materials and Methods). In doing so we were guided by the lessons learned from work to express and crystallise subtype $A10_{61}$ FMDV $3C^{pro}$, which suggested that preserving the N terminus of the protein but truncating the C terminus by up to six residues would be optimal for solubility and crystallisation (Birtley & Curry, 2005). Accordingly, for each SAT sub-type we generated expression constructs that add a thrombin-cleavable His tag to the N terminus of residues 1-208 of the 213 amino acid 3C protease; following thrombin cleavage there is a single additional Gly residue appended to the N terminus of the protease polypeptide. To ensure the solubility of the SAT-type 3C proteins, we introduced to all constructs a C142A substitution to remove a surface-exposed Cys that had been shown previously to be responsible for protein aggregation (Birtley & Curry, 2005; Birtley et al., 2005). (The C95K mutation also introduced to eliminate aggregation of $A10_{61}$ FMDV $3C^{pro}$ (Birtley & Curry, 2005) was not needed here because residue 95 is an Arg in the SAT 3C proteases used in this study). In addition, the active site nucleophile was eliminated from all constructs by incorporation of a C163A substitution to prevent adventitious proteolysis in highly concentrated samples of purified $3C^{pro}$. For consistency with our earlier naming scheme these SAT2/GHA/8/91, SAT1/NIG/5/81,

SAT1/UGA/1/97, and SAT2/ZIM/7/83 3C constructs will be referred to as SAT2/G-g3C^{pro}(1-208), SAT1/N-g3C^{pro}(1-208), SAT1/U- g3C^{pro}(1-208), and SAT2/Z- g3C^{pro}(1-208) respectively.

The 3C^{pro} proteins from all four SAT sub-types yielded soluble protein that was purified first by metal-affinity chromatography and them following thrombin cleavage of the N-terminal His tag, on a gel filtration column (see Materials and Methods). Of the four, SAT1/N-g3C^{pro}(1-208) appeared to be the most soluble and could be concentrated to 20 mg/mL. The other three variants exhibited some precipitation during gel filtration, indicated by a void peak containing aggregated 3C^{pro}, which was about one-third of the area of the monomeric peak. They also had lower apparent solubility limits and could be concentrated to ~6 mg/mL [SAT2/G-g3C^{pro}(1-208)] or ~11 mg/mL [SAT1/U- g3C^{pro}(1-208), and SAT2/Z- g3C^{pro}(1-208)].

In crystallisation trials we only obtained crystals from the 3C^{pro} of a single sub-type: SAT2/G-g3C^{pro}(1-208). These exhibited a variety of habits but the largest were needle-shaped and were typically 10 μm wide and up to 300 μm long. In initial diffraction tests on beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF) showed that the crystals belonged to a trigonal spacegroup and diffracted to a resolution limit of 2 Å. Unfortunately, for reasons that remain unclear, efforts to reproduce these crystals proved troublesome. In subsequent trials diffraction was limited to ~3 Å.

We used mutagenesis to engineer modifications to the SAT2/G-g3C^{pro}(1-208) construct in the search for better crystals. Although alterations to trim the C-terminus by one residue [in SAT2/G-g3C^{pro}(1-207)], or to add back a single His residue [in SAT2/G-g3C^{pro}(1-207h)]

– strategies that had been useful when working with type $A10_{61}$ $3C^{pro}$ (Birtley & Curry, 2005) — both yielded soluble protein and SAT2/G-g3C^{pro}(1-207h) produced crystals, there was no improvement in the resolution of the diffraction.

In a further effort to enhance crystal quality, we used the Surface Entropy Reduction prediction server (Goldschmidt et al., 2007) to design additional SAT2/G-g3C^{pro}(1-208) mutants. We made four different mutants, each containing the following pairs of substitutions: (i) K110T/K111Y (ii) K110Y/K111T; (iii) K51A/K54Y; (iv) K51T/K54S. Of these, only the K51A/K54Y mutant gave protein that was as soluble as wild-type. The K110T/K111Y and K51T/K54S double-mutants produced significantly larger void peaks during purification by gel filtration chromatography, while the K110Y/K111T double-mutant appeared almost entirely aggregated under these conditions. For the three surface-entropy mutants that did yield soluble protein, no useable crystals were obtained.

Structure of SAT2/G-g3C^{pro}(1-208): A complete dataset to 3.2 Å resolution was obtained from crystals of SAT2/G-g3C^{pro}(1-208). The crystals belong to space-group $P3_2$ and have a long c-axis (318.5 Å). The diffraction data were phased by molecular replacement using a search model based on the crystal structure of type $A10_{61}$ FMDV $3C^{pro}$, which is 80% identical in amino-acid sequence to SAT2/G-g3C^{pro}(1-208) (see Materials and Methods). This gave an unambiguous solution with a log likelihood gain of 1495 (McCoy et al., 2007), revealing five molecules in the asymmetric unit. Though of modest resolution, the initial electron density maps were of sufficient quality to guide adjustment of the initial molecular replacement model prior to multiple interleaved rounds of refinement and model building (Fig. 1a). Because of the limited resolution and non-crystallographic symmetry, refinement

was performed using group B-factors and non-crystallographic restraints. Model building was done conservatively – amino acid side-chains were truncated to the C_β atom in cases where there was no indicative electron density. The final model of SAT2/G-g3C^{pro}(1-208) contains residues 7-207 for all five chains and has an R_{free} of 27.2% and good stereochemistry; full data collection and refinement statistics are given in Table 1.

As expected, given the high level of amino acid sequence identity with A10$_{61}$ 3C^{pro}, FMDV SAT2/G-g3C^{pro}(1-208) adopts the same trypsin-like fold (Fig. 1b), which has been described in detail elsewhere (Birtley et al., 2005; Sweeney et al., 2007). Superposition of the five molecules in the asymmetric unit shows that they are highly similar to one another (Fig. 1c) – the pair-wise root mean square deviation in C_α positions between chains is 0.2-0.3 Å. The largest differences are observed in the longest surface-exposed loops, the E_1-F_1 loop in the N-terminal β-barrel and the B_2-C_2 loop known as the β-ribbon in the C-terminal β-barrel (Fig. 1c). These are also the regions of greatest difference between SAT2/G-g3C^{pro}(1-208) and A10$_{61}$ 3C^{pro}; (overlay of the two structures yields an overall rms deviation in C_α positions of ~0.6 Å) (Fig. 1d). The flexibility of the β-ribbon, which shifts in position to aid peptide binding, has been noted before (Zunszain et al., 2010) and clearly it plays a similar role in SAT-type 3C proteases.

CONCLUDING REMARKS

The results reported here provide a template structure of a SAT-type FMDV 3C protease that should be of value in directing molecular investigations of this group of proteases.

Although it is frustrating that higher-resolution diffraction data were not obtained, given that initial crystals of SAT2/G-g3C^{pro}(1-208) diffracted to 2 Å, this should be possible with further optimization. Likewise, since soluble 3C^{pro} was found to be purified from three other SAT-type viruses – notably SAT1/NIG/5/81 – crystal structures for these proteases may well also be achievable. This work has applications in the use of reverse genetics approaches or the design of empty virus capsids to target antigenically significant subtypes in the design of regional vaccines for the control of FMD (Porta et al., 2013).

ACKNOWLEDGEMENTS

We thank staff on beamline ID23 at the ESRF for assistance with data collection. This work was supported by the award of a Wellcome Trust studentship (to E.N.L.) (reference: 083248/2/07/2).

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213-221.

Allaire M, Chernaia MM, Malcolm BA, and James MN. 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369:72-76.

Bastos AD, Haydon DT, Forsberg R, Knowles NJ, Anderson EC, Bengis RG, Nel LH, and Thomson GR. 2001. Genetic heterogeneity of SAT-1 type foot-and-mouth disease viruses in southern Africa. *Arch Virol* 146:1537-1551.

Bastos AD, Anderson EC, Bengis RG, Keet DF, Winterbach HK, and Thomson GR. 2003. Molecular epidemiology of SAT3-type foot-and-mouth disease. *Virus Genes* 27:283-290.

Birtley JR, and Curry S. 2005. Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. *Acta Crystallogr D Biol Crystallogr* 61:646-650.

Birtley JR, Knox SR, Jaulent AM, Brick P, Leatherbarrow RJ, and Curry S. 2005. Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. *J Biol Chem* 280:11520-11527.

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, and Warren GL. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* D54:905-921.

Collaborative Computer Project No. 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50:760-763.

Curry S, Roque-Rosell N, Sweeney TR, Zunszain PA, and Leatherbarrow RJ. 2007. Structural analysis of foot-and-mouth disease virus 3C protease: a viable target for antiviral drugs? *Biochem Soc Trans* 35:594-598.

Emsley P, Lohkamp B, Scott WG, and Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.

Engh RA, and Huber R. 1991. Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Crystallogr.* A47:392-400.

Gastiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, and Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press Inc., 571-607.

277 Goldschmidt L, Cooper DR, Derewenda ZS, and Eisenberg D. 2007. Toward rational protein
278 crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci*
279 16:1569-1576.

280 Knowles NJ, and Samuel AR. 2003. Molecular epidemiology of foot-and-mouth disease virus. *Virus*
281 *Res* 91:65-80.

282 Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA,
283 McElroy HE, Gribskov CL, and Worland S. 1994. Structure of human rhinovirus 3C protease
284 reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor
285 polyprotein. *Cell* 77:761-771.

286 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ. 2007. Phaser
287 crystallographic software. *J Appl Crystallogr* 40:658-674.

288 Maree FF, Blignaut B, Esterhuysen JJ, de Beer TA, Theron J, O'Neill HG, and Rieder E. 2011. Predicting
289 antigenic sites on the foot-and-mouth disease virus capsid of the South African Territories types
290 using virus neutralization data. *J Gen Virol* 92:2297-2309.

291 Mosimann SC, Cherney MM, Sia S, Plotch S, and James MN. 1997. Refined X-ray crystallographic
292 structure of the poliovirus 3C gene product. *J Mol Biol* 273:1032-1047.

293 Porta C, Xu X, Loureiro S, Paramasivam S, Ren J, Al-Khalil T, Burman A, Jackson T, Belsham GJ, Curry S et al. .
294 2013. Efficient production of foot-and-mouth disease virus empty capsids in insect cells following
295 down regulation of 3C protease activity. *J Virol Methods* 187:406-412.

296 Sutmoller P, Barteling SS, Olascoaga RC, and Sumption KJ. 2003. Control and eradication of foot-
297 and-mouth disease. *Virus Res* 91:101-144.

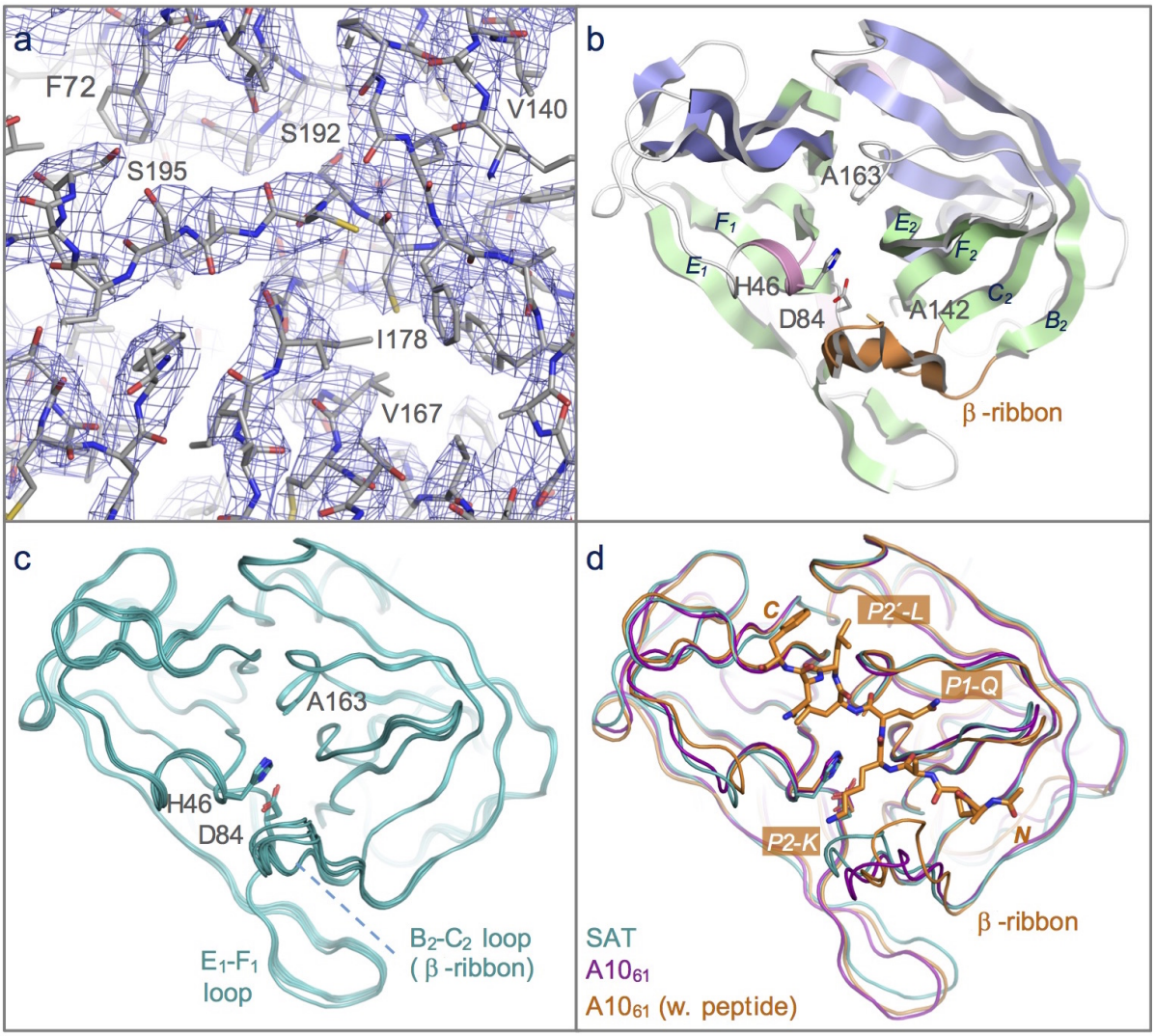
298 Sweeney TR, Roque-Rosell N, Birtley JR, Leatherbarrow RJ, and Curry S. 2007. Structural and
299 mutagenic analysis of foot-and-mouth disease virus 3C protease reveals the role of the beta-
300 ribbon in proteolysis. *J Virol* 81:115-124.

301 Yin J, Bergmann EM, Cherney MM, Lall MS, Jain RP, Vederas JC, and James MN. 2005. Dual modes of
302 modification of hepatitis A virus 3C protease by a serine-derived beta-lactone: selective
303 crystallization and formation of a functional catalytic triad in the active site. *J Mol Biol* 354:854-
304 871.

305 Zunszain PA, Knox SR, Sweeney TR, Yang J, Roque-Rosell N, Belsham GJ, Leatherbarrow RJ, and
306 Curry S. 2010. Insights into cleavage specificity from the crystal structure of foot-and-mouth
307 disease virus 3C protease complexed with a peptide substrate. *J Mol Biol* 395:375-389.
308

309

310 **FIGURE**



311

312 **Figure 1: Structure of the 3C protease from the SAT2/GHA/8/91 serotype FMDV.**

313 (a) Section of the 3.2 Å resolution electron density map (blue chicken wire) calculated with
314 phases from the final refined model, which is shown as sticks coloured by atom type: grey –
315 carbon; red – oxygen; blue – nitrogen; yellow – sulphur. (b) Overall structure of SAT2/G-
316 g3C^{pro}(1-208), with secondary structure elements indicated. (c) Superposition of the five
317 molecules of SAT2/G-g3C^{pro}(1-208) in the asymmetric unit of the crystal, shown in ribbon
318 representation. (d) Comparative superposition of SAT2/G-g3C^{pro}(1-208) (teal) with A10$_{61}$

319 $3C^{pro}$ in the absence (purple; PDB 2J92)) and presence (orange; PDB 2WV4)) of a peptide
320 substrate (shown in stick representation).
321

322 **TABLE 1:** DNA primers for cloning and mutagenesis

	SAT2/GHA/8/91
Forward	GATGATCTCGAGGAAGTGGCGCTCCGCCGACCGAC
Reverse	CATGCCAAGCTTATGGGTCAATGTGTGCTTTGAGTTGGAGCAGGCTCGACCGTG
C142A-for	GGACCAAGGTTGGATACGCTGGAGGAGCCGTCATGAC
C142A-rev	GTCATGACGGCTCCTCCAGCGTATCCAACCTTGGTCC
C163A-for	CATACAAAGATGTTGTCGTCGCCATGGACGGTGAACACCATGC
C163Arev	GCATGGTGTCACCGTCCATGGCGACGACAACATCTTTGTATG
	SAT1/NIG/5/81
Forward	GATGATCTCGAGGAAGTGGAGCGCCACCCACCGAC
Reverse	CATGCCAAGCTTAAGGGTCGATGTGTGCCTTCATC
C142A-for	GCCACCAAAGCTGGTTACGCTGGAGGAGCCGTTCTTG
C142A-rev	CAAGAACGGCTCCTCCAGCGTAACCAGCTTTGGTGGC
C163A-for	CCTACAAAGACATCGTAGTGGCTATGGATGGTGACACCATGC
C163Arev	GCATGGTGTCACCATCCATAGCCACTACGATGTCTTTTGTAGG
	SAT1/UGA1/97
Forward	GATGATCTCGAGGAAGCGGTGCGCCACCGACCGAC
Reverse	CATGCCAAGCTTATGGGTCGATGTGGGCTTTCATC
C142A-for	GGACCAAGGTAGGTTACGCTGGGGCGGCCGTACTGAC
C142A-rev	GTCAGTACGGCCGCCCCAGCGTAACCTACCTTGGTCC
C163A-for	GTACAACGACATCGTCGTCGCCATGGACGGCGACACCATG
C163Arev	CATGGTGTCGCCGTCCATGGCGACGACGATGTCGTTGTAC
	SAT2/ZIM/7/83
Forward	GATGATCTCGAGGAAGCGGAGCCCCACCGACCGAC
Reverse	CATGCCAAGCTTAAGGGTCGATGTGGGCCTTCATC
C142A-for	GGGACCAAAGTTGGATACGCTGGAGCCGCTGTTCTCG
C142A-rev	CGAGAACAGCGGCTCCAGCGTATCCAACTTTGGTCCC
C163A-for	CCTACAAAGACCTAGTCGTTGCTATGGACGGTGACACCATGC
C163Arev	GCATGGTGTCACCGTCCATAGCAACGACTAGGTCTTTGTAGG

323

324

325 **TABLE 2:** Crystallographic data collection and model refinement statistics for SAT2 $3C^{pro}$.

DATA COLLECTION	
Space-group	$P3_2$
a, b, c (Å)	54.0, 54.0, 318.5
α, β, γ (°)	α = β = 90; γ = 120
Resolution range (Å)	53.1-3.2 (3.37-3.2)
No. of independent reflections	17053
Multiplicity[1]	2.7 (2.7)
Completeness (%)	99.3 (99.5)
I/σ_I	5.7 (1.7)
R_{merge} (%)[2]	11.6 (42.4)
MODEL REFINEMENT	
No. of Non-hydrogen atoms	7535
R_{work} (%)[3]	22.2
R_{free} (%)[4]	27.2
Average B-factor (Å²)	119
RMS deviations - Bonds (Å) [5]	0.006
RMS deviations - Angles (°)	1.1
Ramachandran plot (favoured/allowed) %	89.8/10.2
PDB Accession Code	4X2Y

326

327 [1]Values for highest resolution shell given in parentheses

328 [2]$R_{merge} = 100 \times \Sigma_{hkl}|I_j(hkl) - \langle I_j(hkl)\rangle| / \Sigma_{hkl}\Sigma_j I(hkl)$, where $I_j(hkl)$ and $\langle I_j(hkl)\rangle$ are the

329 intensity of measurement j and the mean intensity for the reflection with indices hkl,

330 respectively.

331 [3]$R_{work} = 100 \times \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$.

[4]R_{free} is the R_{model} calculated using a randomly selected 5% sample of reflection data that were omitted from the refinement.

[5]RMS, root-mean-square; deviations are from the ideal geometry defined by the Engh and Huber parameters (Engh & Huber, 1991).