Crystal structure of the 3C protease from Southern 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 large 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. On the basis of capsid sequence comparisons the FMDV serotypes are grouped into two phylogenetic clusters, with O, A, C, and Asia 1 in one, and the three Southern African Territories serotypes, (SAT-1, SAT-2 and SAT-3) in another, a grouping pattern that is broadly, but not rigidly, reflected in $3C^{pro}$ amino acid sequences. 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 Southern 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 large 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. On the basis of capsid sequence comparisons the FMDV serotypes are grouped into two phylogenetic clusters, with O, A, C, and Asia 1 in one, and the three Southern African Territories serotypes, (SAT-1, SAT-2 and SAT-3) in another, a grouping pattern that is broadly, but not rigidly, reflected in $3C^{pro}$ amino acid sequences. 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 serotypes (O, A, C, Asia1 and three Southern 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 in their capsid proteins compared 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. Cell entry in infected hosts is followed immediately by translation of a large open reading frame in the viral RNA. This yields a polyprotein precursor of over 2,000 amino acids that is processed into fourteen distinct capsid and

non-structural proteins for virus replication. The majority of this processing is done by the virus-encoded 3C protease ($3C^{pro}$), which cleaves the precursor at ten distinct sites. FMDV $3C^{pro}$ may also assist infection by proteolysis of host cell proteins and has RNA-binding activity that is important for initiation of replication of the viral RNA [reviewed in (Curry et al. 2007b)].

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

Sequence analysis has shown that while variation within FMDV $3C^{pro}$ does not rigidly reflect that observed with capsid proteins, the SAT_type 3C proteases generally form a distinct cluster (van Rensburg et al. 2002). Mapping of the sequence variation between different FMDV serotypes onto the structure of $A10_{61}$ $3C^{pro}$ indicated that the peptide-binding face of the protease is completely conserved among the non-SAT serotypes (which are 91-97% conserved in amino-acid sequence), supporting the notion that identification of inhibitors of the protease might aid the development of broad spectrum antiviral drugs (Birtley et al. 2005; Curry et al. 2007a). This structure should therefore serve as a useful

77 model for the 3C protease from this group of viruses. However, the same comparison
78 suggested the presence of at least two amino acid differences on the peptide-binding
79 surfaces between $A10_{61}$ $3C^{pro}$ and the corresponding 3C sequences from SAT serotype
80 viruses.

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

86

87 **MATERIALS AND METHODS**

88 **Cloning and mutagenesis:** We used the polymerase chain reaction (PCR) to amplify the
89 coding regions for the FMDV 3C proteases of sub-types SAT2/GHA/8/91 (Accession No.
90 AY884136), SAT1/NIG/5/81 (Accession No. AY882592), SAT1/UGA/1/97 (Accession No.
91 AF283456), and SAT2/ZIM/7/83 (Accession No. AF540910). In each case the reaction was
92 performed using DNA primers (Table 1) that introduced 5′ *Xho*I and a 3′ *Hind*III restriction
93 sites into the PCR products. These served to facilitate ligation into a version of the pETM-11
94 vector that had been modified to insert a thrombin cleavage site immediately downstream
95 of the N-terminal His tag (Birtley & Curry 2005). DNA ligations were performed using the
96 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.

Details of the particular modifications made to expressed proteins are given in the Results and Discussion section.

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) which have 80%, 92%, 82% and 85% amino acid sequence identity respectively with the $3C^{pro}$ from FMDV $A10_{61}$ (Fig. 1). 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

163 single additional Gly residue appended to the N terminus of the protease polypeptide. To
164 ensure the solubility of the SAT-type 3C proteins, we introduced to all constructs a C142A
165 substitution to remove a surface-exposed Cys that had been shown previously to be
166 responsible for protein aggregation (Birtley & Curry 2005; Birtley et al. 2005). (The C95K
167 mutation also introduced to eliminate aggregation of $A10_{61}$ FMDV $3C^{pro}$ (Birtley & Curry
168 2005) was not needed here because residue 95 is an Arg in the SAT 3C proteases used in
169 this study). In addition, the active site nucleophile was eliminated from all constructs by
170 incorporation of a C163A substitution to prevent adventitious proteolysis in highly
171 concentrated samples of purified $3C^{pro}$. For consistency with our earlier naming scheme
172 these SAT2/GHA/8/91, SAT1/NIG/5/81, SAT1/UGA/1/97, and SAT2/ZIM/7/83 3C
173 constructs will be referred to as SAT2/G-g3C^{pro}(1-208), SAT1/N-g3C^{pro}(1-208), SAT1/U-
174 g3C^{pro}(1-208), and SAT2/Z- g3C^{pro}(1-208) respectively.

175 The $3C^{pro}$ proteins from all four SAT sub-types yielded soluble protein that was purified
176 first by metal-affinity chromatography and then, following thrombin cleavage of the
177 N-terminal His tag, on a gel filtration column (see Materials and Methods). Of the four,
178 SAT1/N-g3C^{pro}(1-208) appeared to be the most soluble and could be concentrated to
179 20 mg/mL (see Table 2). The other three variants exhibited some precipitation during gel
180 filtration, indicated by a void peak containing aggregated $3C^{pro}$, which was about one-third
181 of the area of the monomeric peak. They also had lower apparent solubility limits and could
182 be concentrated to ~6 mg/mL [SAT2/G-g3C^{pro}(1-208)] or ~11 mg/mL [SAT1/U- g3C^{pro}(1-
183 208), 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 unsuccessful. 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 (Table 2) 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 some soluble protein, no useable crystals were obtained.

206 **Structure of SAT2/G-g3C^{pro}(1-208):** A complete dataset to 3.2 Å resolution was obtained
207 from crystals of SAT2/G-g3C^{pro}(1-208). The crystals belong to space-group $P3_2$ and have a
208 long c-axis (318.5 Å). The diffraction data were phased by molecular replacement using a
209 search model based on the crystal structure of type $A10_{61}$ FMDV 3C^{pro}, which is 80%
210 identical in amino-acid sequence to SAT2/G-g3C^{pro}(1-208) (see Materials and Methods).
211 This gave an unambiguous solution with a log likelihood gain of 1495 (McCoy et al. 2007),
212 revealing five molecules in the asymmetric unit. Though of modest resolution, the initial
213 electron density maps (Fig. 2a) were of sufficient quality to guide adjustment of the initial
214 molecular replacement model prior to multiple interleaved rounds of refinement and
215 model building. Because of the limited resolution and non-crystallographic symmetry,
216 refinement was performed using group B-factors and non-crystallographic restraints.
217 Model building was done conservatively – amino acid side-chains were truncated to the C_β
218 atom in cases where there was no indicative electron density. The final model of SAT2/G-
219 g3C^{pro}(1-208) contains residues 7-207 for all five chains and has an R_{free} of 27.2% and good
220 stereochemistry; full data collection and refinement statistics are given in Table 3.

221 As expected, given the high level of amino acid sequence identity with $A10_{61}$ 3C^{pro} (80%),
222 FMDV SAT2/G-g3C^{pro}(1-208) adopts the same trypsin-like fold (Fig. 2b), which has been
223 described in detail elsewhere (Birtley et al. 2005; Sweeney et al. 2007). Superposition of
224 the five molecules in the asymmetric unit shows that they are highly similar to one another
225 (Fig. 1 and 2c) – the pair-wise root mean square deviation in C_α positions between chains is
226 0.2-0.3 Å. The largest differences are observed in the longest surface-exposed loops, the E_1-
227 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. 2c). 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. 2d). 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.

ACKNOWLEDGEMENTS

We thank staff on beamline ID23 at the ESRF for assistance with data collection.

247

248 **REFERENCES**

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

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

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

257 Bastos AD, Haydon DT, Forsberg R, Knowles NJ, Anderson EC, Bengis RG, Nel LH, and Thomson GR.
258 2001. Genetic heterogeneity of SAT-1 type foot-and-mouth disease viruses in southern Africa.
259 *Arch Virol* 146:1537-1551.

260 Birtley JR, and Curry S. 2005. Crystallization of foot-and-mouth disease virus 3C protease: surface
261 mutagenesis and a novel crystal-optimization strategy. *Acta Crystallogr D Biol Crystallogr*
262 61:646-650.

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

266 Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J,
267 Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, and Warren GL. 1998. Crystallography &
268 NMR system: A new software suite for macromolecular structure determination. *Acta*
269 *Crystallogr D Biol Crystallogr* 54:905-921.

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

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

275 Curry S, Roque-Rosell N, Zunszain PA, and Leatherbarrow RJ. 2007b. Foot-and-mouth disease virus
276 3C protease: recent structural and functional insights into an antiviral target. *Int J Biochem Cell*
277 *Biol* 39:1-6.

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

280 Engh RA, and Huber R. 1991. Accurate bond and angle parameters for x-ray protein-structure
281 refinement. *Acta Crystallogr A Foundations Crystallogr* 47:392-400.

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

285 Goldschmidt L, Cooper DR, Derewenda ZS, and Eisenberg D. 2007. Toward rational protein
286 crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci*
287 16:1569-1576.

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

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

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

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

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

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

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

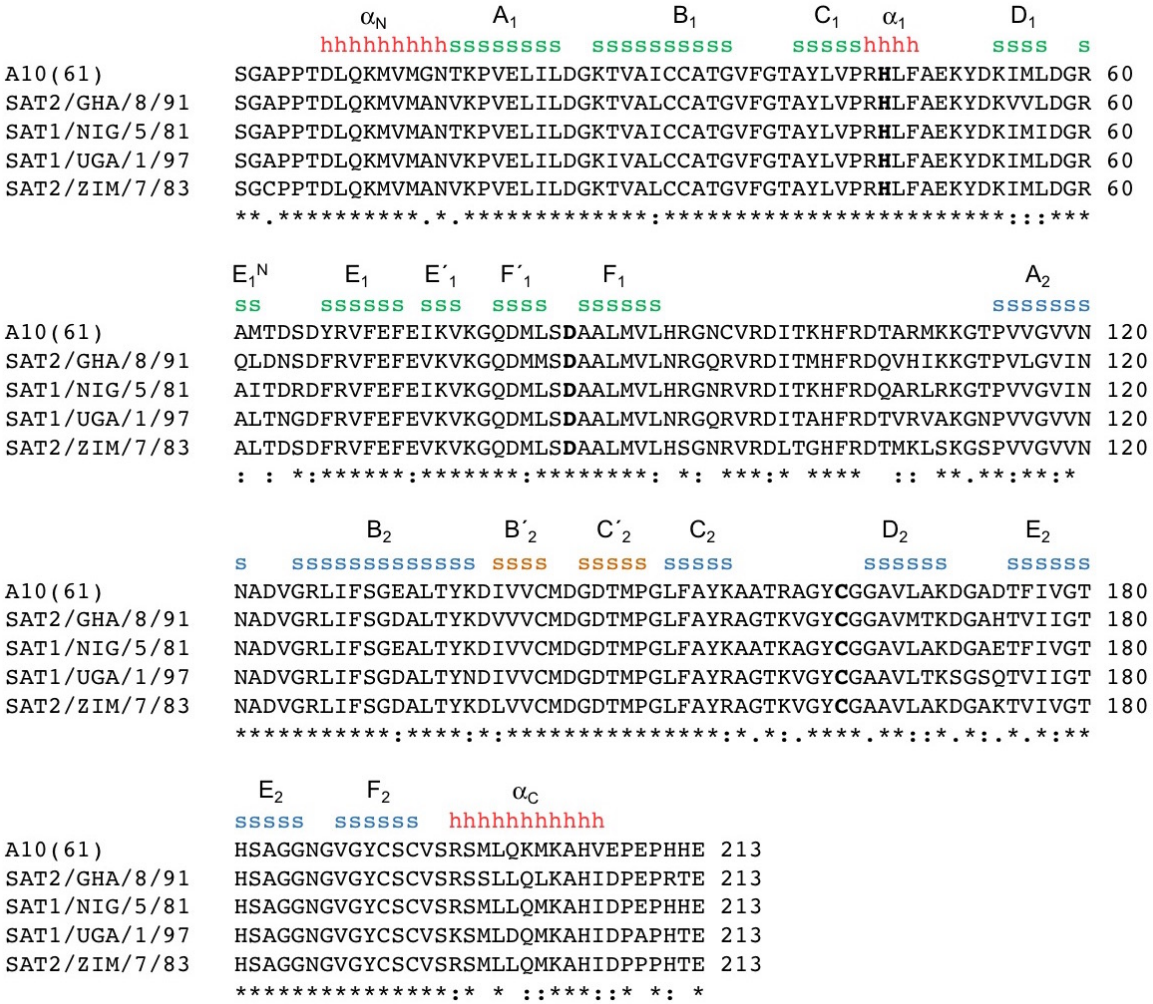
306 van Rensburg H, Haydon D, Joubert F, Bastos A, Heath L, and Nel L. 2002. Genetic heterogeneity in
307 the foot-and-mouth disease virus Leader and 3C proteinases. *Gene* 289:19-29.

308 Yin J, Bergmann EM, Cherney MM, Lall MS, Jain RP, Vederas JC, and James MN. 2005. Dual modes of
309 modification of hepatitis A virus 3C protease by a serine-derived beta-lactone: selective
310 crystallization and formation of a functional catalytic triad in the active site. *J Mol Biol* 354:854-
311 871.

312 Zunszain PA, Knox SR, Sweeney TR, Yang J, Roque-Rosell N, Belsham GJ, Leatherbarrow RJ, and
313 Curry S. 2010. Insights into cleavage specificity from the crystal structure of foot-and-mouth
314 disease virus 3C protease complexed with a peptide substrate. *J Mol Biol* 395:375-389.
315

316

317 **FIGURES**



318

319 **Figure 1: Amino acid sequence alignment of $A10_{61}$ $3C^{pro}$ with the 3C proteases from**

320 **the four SAT serotypes used in this study.** Secondary structure features are indicated (h

321 = α-helix; s = β-strand), and coloured and labelled as in Fig. 2b (consistent with the naming

322 scheme used in (Birtley et al. 2005).

323

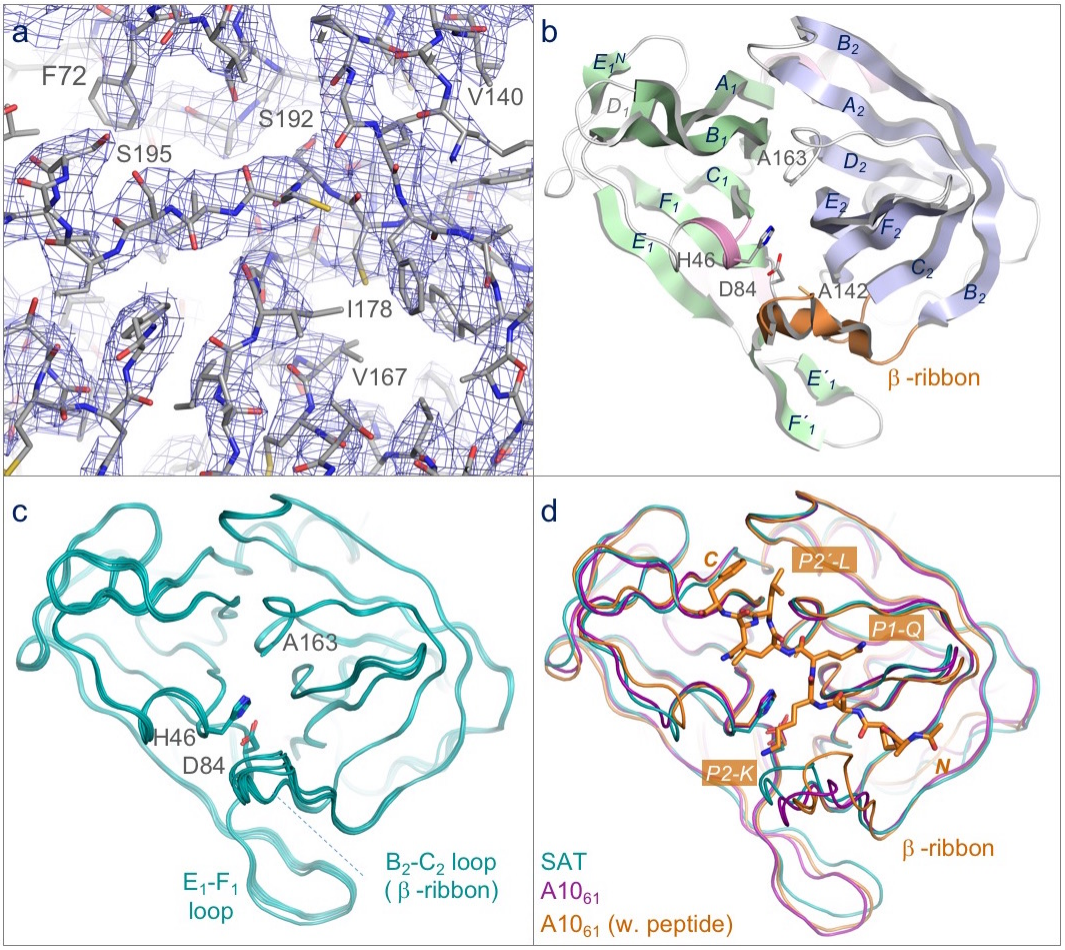


Figure 2: Structure of the 3C protease from the SAT2/GHA/8/91 serotype FMDV.
(a) Section of the 3.2 Å resolution electron density map (blue chicken wire) calculated with phases from the final refined model, which is shown as sticks coloured by atom type: grey – carbon; red – oxygen; blue – nitrogen; yellow – sulphur. (b) Overall structure of SAT2/G-g3C^{pro}(1-208), with secondary structure elements indicated. The N- and C-terminal β-barrels are coloured green and blue respectively. (c) Superposition of the five molecules of SAT2/G-g3C^{pro}(1-208) in the asymmetric unit of the crystal, shown in ribbon representation. (d) Comparative superposition of SAT2/G-g3C^{pro}(1-208) (teal) with $A10_{61}$ 3C^{pro} in the absence (purple; PDB 2J92)) and presence (orange; PDB 2WV4)) of a peptide substrate (shown in stick representation).

336

337 **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

338
339

340 **Table 2:** Protein yields and solubilities

Protein	Yield (mg per L of culture)	Maximum concentration (mg/mL)	Aggregation
SAT1/N-g3C^{pro}(1-208)	7.5	19.8	-
SAT1/U- g3C^{pro}(1-208)	1.2	11.9	+++
SAT2/Z- g3C^{pro}(1-208)	2.2	11.3	++
SAT2/G-g3C^{pro}(1-208)	2.5	5.7	++
SAT2/G-g3C^{pro}(1-207h)	2.1	7.2	+
SAT2/G-g3C^{pro}(1-207)	1.7	5.6	+++

341

342
343

344 **Table 3:** 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

345

346 [1]Values for highest resolution shell given in parentheses

347 [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

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

349 respectively.

350 [3]$R_{work} = 100 \times \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$.

351 [4]R_{free} is the R_{model} calculated using a randomly selected 5% sample of reflection data that

352 were omitted from the refinement.

353 [5]RMS, root-mean-square; deviations are from the ideal geometry defined by the Engh and

354 Huber parameters (Engh & Huber 1991).

355