

Crystal Structure of a Novel Viral Protease with a Serine/Lysine Catalytic Dyad Mechanism

Anat R. Feldman¹†, Jaeyong Lee¹†, Bernard Delmas² and Mark Paetzel¹*

¹Department of Molecular Biology and Biochemistry
Simon Fraser University
South Science Building, 8888
University Drive, Burnaby
British Columbia, Canada
V5A 1S6

²Unité de Virologie et
Immunologie Moléculaires
Institut National de la
Recherche Agronomique
F-78350 Jouy-en-Josas, France

The blotched snakehead virus (BSNV), an aquatic birnavirus, encodes a polyprotein (NH₂-pVP2-X-VP4-VP3-COOH) that is processed through the proteolytic activity of its own protease (VP4) to liberate itself and the viral proteins pVP2, X and VP3. The protein pVP2 is further processed by VP4 to give rise to the capsid protein VP2 and four structural peptides. We report here the crystal structure of a VP4 protease from BSNV, which displays a catalytic serine/lysine dyad in its active site. This is the first crystal structure of a birnavirus protease and the first crystal structure of a viral protease that utilizes a lysine general base in its catalytic mechanism. The topology of the VP4 substrate binding site is consistent with the enzymes substrate specificity and a nucleophilic attack from the *si*-face of the substrates scissile bond. Despite low levels of sequence identity, VP4 shows similarities in its active site to other characterized Ser/Lys proteases such as signal peptidase, LexA protease and Lon protease. Together, the structure of VP4 provides insights into the mechanism of a recently characterized clan of serine proteases that utilize a lysine general base and reveals the structure of potential targets for antiviral therapy, especially for other related and economically important viruses, such as infectious bursal disease virus in poultry and infectious pancreatic necrosis virus in aquaculture.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: serine/lysine catalytic dyad; viral protease; polyprotein; birnavirus; lysine general base

*Corresponding author

Introduction

Proteases of viral origin have evolved to process virally encoded polyproteins during replication, in co- and post-translational steps. These proteases are essential for viral replication, assembly and infectivity and are therefore potential targets for antiviral drugs.^{1,2}

Viruses of the *Birnaviridae* family infect animal species belonging to vertebrates, mollusks, insects and rotifers and are characterized by their bi-segmented double-stranded RNA genome (segments A and B).³ Segment A encodes two overlapping reading frames and the larger open

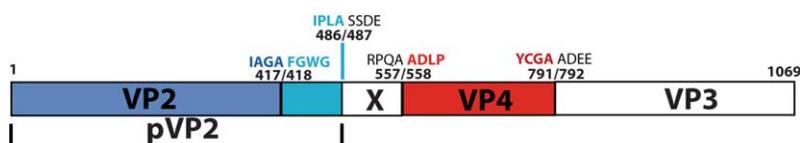
reading frame encodes a polyprotein (NH₂-pVP2-VP4-VP3-COOH; **Figure 1**). The polyprotein is processed through the proteolytic activity of VP4 to generate the proteins pVP2 and VP3.⁴ During virus assembly pVP2 is further processed by VP4 to generate the capsid protein VP2 and structural peptides. Recently, the crystal structure of the infectious bursal disease virus particle (IBDV, an avian birnavirus) was solved at a resolution of 7 Å.⁵ VP2 was the only component seen in the structure of the virus icosahedral capsid.

The blotched snakehead virus (BSNV) was initially identified from a cell line developed from the blotched snakehead fish (*Channa lucius*) caudal peduncle.⁶ Initial analysis of this birnavirus indicated that its genome encodes a polyprotein harboring an additional peptide, named X, between the pVP2 and VP4 reading frames.⁷ VP4 cleaves its own N and C termini in the polyprotein at amino acid positions 557–558 and 791–792, respectively. The cleavage site between pVP2 and the 71 amino acid long peptide X junction has been mapped between residues 486 and 487. These cleavage sites

† A.R.F. and J.L. contributed equally to this work.

Abbreviations used: BSNV, blotched snakehead virus; IPNV, infectious pancreatic necrosis virus; IBDV, infectious bursal disease virus; DXV, Drosophila X virus; VP4, viral protein (protease) 4 in the polyprotein (pVP2-X-VP4-VP3) coded for in the birnavirus.

E-mail address of the corresponding author: mpaetzel@sfu.ca



protease is shown in red. The P4 to P4' sequences for the VP2/pVP2, pVP2/X, X/VP4, and VP4/VP3 cleavage sites are shown.

Figure 1. The linear arrangement of the BSNV polyprotein. VP2, a viral capsid protein, is shown in blue. The propeptide of VP2 is shown in cyan. The X peptide and VP3 are shown in white. VP4

and those involved in the processing of pVP2 can be described by the consensus Pro-X-Ala↓(Ala/Ser) motif⁷ (Figure 1).

Site-directed mutagenesis studies on VP4 protease have shown that the conserved catalytic residues serine 692 and lysine 729 in BSNV are essential for polyprotein processing.⁷ Similar results were found for the VP4 protease in infectious pancreatic necrosis virus (IPNV) (serine 633 and lysine 674)⁴ and IBDV (serine 652 and lysine 692).^{8,9} The VP4 protease of the *Birnaviridae* family therefore are proposed to utilize a serine/lysine catalytic dyad mechanism to catalyze the processing of the polyprotein,^{4,8,9} which distinguish it from other viral proteases. Evolutionarily, VP4 protease belongs to the clan SJ and family S50 in the MEROPS protease data-bank.¹⁰ Other characterized proteases with a Ser/Lys catalytic dyad mechanism include: signal peptidase,^{11–13} LexA,¹⁴ cIλ,¹⁵ UmuD',¹⁶ and Lon protease.¹⁷

Here we report the first structural characterization of a birnavirus protease and the first viral protease to utilize a lysine general base in its catalytic mechanism.

Results and Discussion

Structure solution

After optimization of the VP4 protease construct for crystallization we obtained single and ordered cubic crystals sufficient for diffraction data collection. The crystallizable construct is 217 residues in length (residues 558–773). We have solved the structure of BSNV to 2.2 Å resolution using single-wavelength anomalous dispersion (SAD) phasing methods. The refined model includes residues 559 to 769 except for a small mobile loop (residues 637 to 648) located at the VP4 surface.

Overall protein architecture

BSNV VP4 protease is approximately 38 Å × 50 Å × 36 Å in size with a predominantly negatively charged surface consistent with its calculated isoelectric point. VP4 displays a protein fold, which is made up of two separate domains. Domain I (residues 559 to 693; light blue in Figure 2) is predominately β-sheet in structure. Domain II (694 to 769; dark blue, purple and gold in Figure 2) is α/β in structure. There are a total of 13 strands incorporated within two β-sheets and one

β-hairpin. There are three α-helices (gold in Figure 2) and one ₃10-helix (612–614, not shown for clarity) in the structure. A mixed but mostly antiparallel twisted β-sheet contains eight strands (light blue), a parallel β-sheet contains three short strands (dark blue), and a short β-hairpin (purple) contains the remaining two strands. The largest β-sheet is highly twisted and forms an open β-barrel structure, which provides the scaffold for the substrate binding site. The α-helices pack against the parallel three-stranded β-sheet and the β-hairpin in domain II. A sequence alignment of BSNV VP4 with the proteases from three other birnaviruses (IBDV, IPNV and Drosophila X virus (DXV)) is presented in Figure 3. BSNV VP4 is 23% identical in sequence to IBDV VP4, 18% identical in sequence to IPNV VP4 and 18% identical in sequence to DXV VP4. There are a number of deletions in the VP4 from IBDV, IPNV and DXV in comparison to the BSNV VP4.

Active-site architecture

The structure of BSNV VP4 reveals that the putative nucleophile Ser692 O^γ and proposed general base Lys729 N^ε are within hydrogen bonding distance (3.1 Å) and therefore consistent with previous site-directed mutagenesis studies⁷ which showed that these residues are essential for activity (Figure 4). Ser692 resides just before helix α1 and Lys729 resides within helix α2. Lys729 N^ε is also coordinated by Thr712 O^{γ1} (4.0 Å in molecule A, 3.3 Å in molecule B) and Pro590 O (3.3 Å in molecule A, 4.3 Å in molecule B) (Figure 4), the position of the N^ε being slightly different in the two molecules in the asymmetric unit. The N^ε of Lys729 is almost completely buried. The accessible surface area (ASA) of Lys729 N^ε is 5.5 Å² (average for molecules A and B) as compared to the average ASA for the N^ε of other lysine residues in BSNV VP4, which is 51.4 Å² (average for ten other lysine residues in BSNV VP4, molecules A and B). When a cleavage site (substrate) binds into the binding pocket of VP4, the N^ε of Lys729 would likely become completely buried. This suggests that the ε-amino group of Lys729 is able to become deprotonated and function as a general base due to a lowering of its pK_a value resulting from its burial within the protein.

An important component of the catalytic machinery in serine proteases is the oxyanion-hole that works by neutralizing the developing negative charge on the scissile carbonyl oxygen atom during the formation of the tetrahedral intermediates.

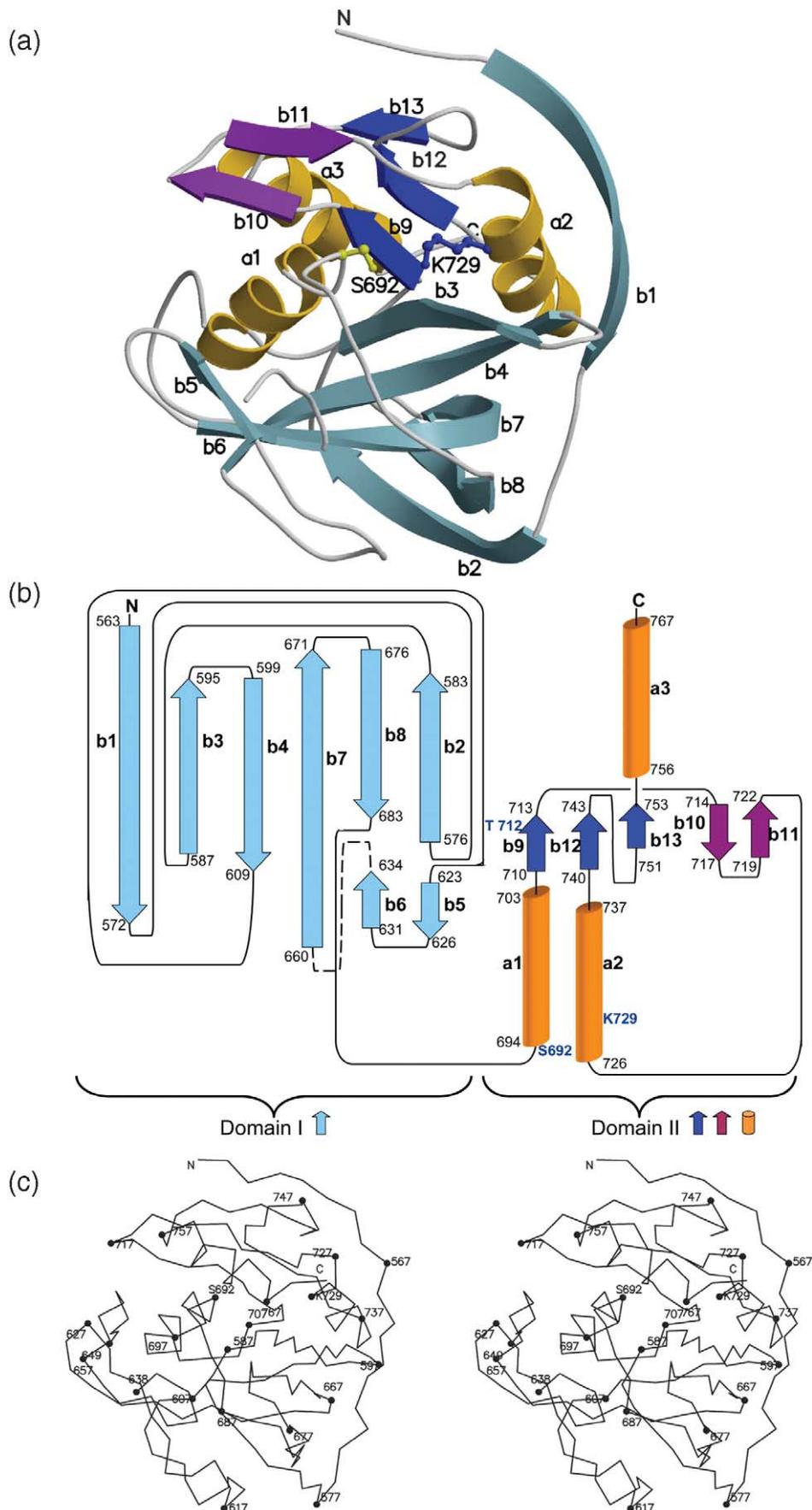


Figure 2. The protein fold of BSNV VP4 protease. (a) A ribbon diagram of VP4. The N-terminal mostly antiparallel β -sheet domain (domain I) is in light blue. The C-terminal α/β -domain is shown in dark blue (three-stranded parallel β -

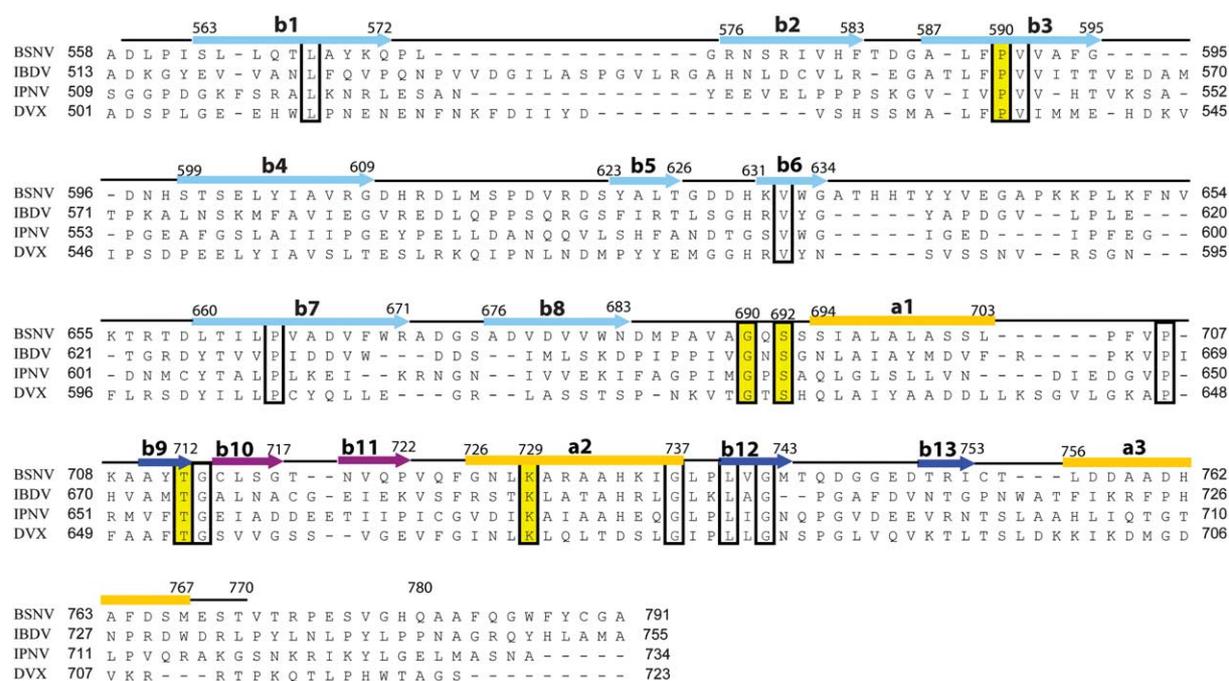


Figure 3. A sequence alignment of birnavirus VP4 proteases. The sequences were obtained from the Swiss-Prot data bank with the accession numbers BSNV (Q8AZM0), IBDV (P15480), IPNV (P90205) and DXV (Q96724). The secondary structure and numbering of BSNV is shown above the alignment. The secondary structure is colored the same as for Figure 2. The conserved and catalytic residues are boxed.

Typically oxyanion holes are formed by two main-chain amide hydrogen atoms that serve as hydrogen bond donors to the developing oxyanion. From the conservation seen in the sequence alignment with other birnavirus VP4 proteases (Figure 3) it would be likely that Ser692 and Gly690 would be the residues that contribute to the oxyanion hole. Yet based on the structure of VP4, with no substrate or inhibitor bound, the residues most likely to contribute atoms to the oxyanion hole are Ser692 and Gln691 whose main-chain NH are pointing towards the binding site (Figure 4). The Gly690 NH is pointing away from the binding site. It appears that the only other possible hydrogen bonding donor that could contribute to the oxyanion hole would be the Gln691 side-chain if it were able to make large changes in its χ_1 and χ_2 angles to position its side-chain $N^{\epsilon 2}$ over the developing transition state oxyanion. Also it is possible that substrate binding may induce conformation changes in the binding site such that Gly690 N would be a contributor to the oxyanion stabilization.

A possible candidate for a deacylating (catalytic) water (Wat270 in molecule A, Wat134 in molecule B; cyan in Figure 4), is within 3.8 Å of Lys729 N^{ζ} . This water (B factor of 26.74 Å²) is also within hydrogen bonding distance to Thr712 $O^{\gamma 1}$ (2.6 Å), Ser692 O^{γ} (2.7 Å), Ser692 O (2.9 Å) and Pro590 O (2.8 Å). The other ordered water molecules observed near the active site region would likely be displaced by the substrate. For a water to serve as a deacylating or

catalytic water in the deacylation step of the hydrolytic reaction it needs to be positioned such that it is within hydrogen bonding distance to a general base for activation as well as at a suitable angle of approach with respect to the carbonyl of the ester intermediate (so-called Bürgi angle¹⁸ of $\sim 107^\circ$). By superimposing the active sites of the VP4 structure onto the active site of the inhibitor-bound acyl-enzyme structure of *Escherichia coli* type I signal peptidase, whose nucleophilic serine is covalently bound to a β -lactam inhibitor,¹³ we can approximate the distance to the ester carbonyl carbon and the Bürgi angle. Wat270 is equidistant between Lys729 N^{ζ} (3.8 Å) and the ester carbonyl carbon (3.8 Å) of the inhibitor and has an angle of approach to the ester carbonyl carbon atom (Bürgi angle) of 150.2° .

Surface analysis reveals that the largest pocket on the surface of VP4 corresponds to an extended bent crevice on the VP4 surface that incorporates the catalytic residues (Ser692 and Lys729) at one end. The dimensions of the crevice are consistent with the binding of an extended polypeptide and the topology of the surface suggests that this crevice is the S1/S3 binding pocket. This pocket has a molecular surface of 372 Å² and contains atoms from the following residues: 589–593, 600, 602, 604, 683, 685–692, 714, 723–725, 729 (Figure 4). The sizes of the S1 and S3 binding pockets are consistent with the size of the side-chains at the P1 and P3 positions in the cleavage site (Schechter & Berger nomenclature¹⁹). The pos-

sheet), purple (β -hairpin), and gold (α -helices). The α -helices and β -sheets are numbered sequentially. The nucleophile Ser692 and general base Lys729 are shown in ball-and-stick format. (b) A schematic diagram of the protein topology of VP4. (c) A stereo diagram of the C^α trace of VP4. The position of every tenth residue is labeled, as well as the position of Ser692 and Lys729.

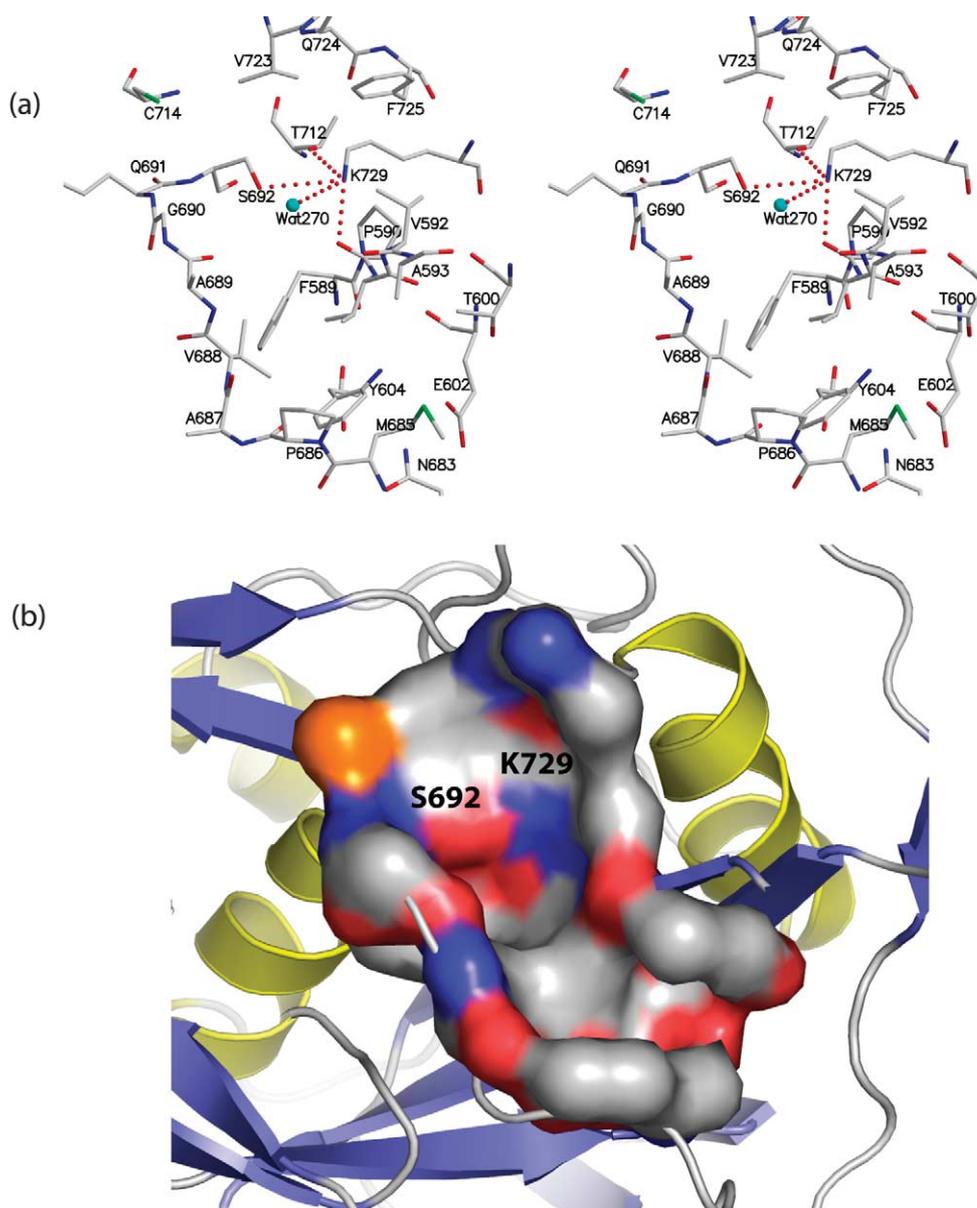


Figure 4. The active-site and substrate binding site of BSNV VP4. (a) A stereo diagram of the VP4 active site. The atoms that contribute to the binding site and active site are shown in stick (red, oxygen; blue, nitrogen; white, carbon). The hydrogen-bonding environment around Lys729 (the general base) is shown by red dots. A potential catalytic or deacylating water molecule (Wat270) is shown in cyan. (b) The largest pocket on the surface of VP4 as analyzed by the program CASTp³⁹ corresponds to the substrate binding site (shown in molecular surface; red, oxygen; blue, nitrogen; white, carbons). The position; orange, sulfur of the catalytic residues Ser692 and Lys729 are marked.

ition of the binding-site relative to the nucleophilic serine hydroxyl group shows that Ser692 O^γ attacks from the *si*-face of the scissile bond. This unusual feature has also been seen in other Ser/Lys proteases such as signal peptidase,^{13,20,21} Lex-A-like proteases^{14–16} and Lon protease.¹⁷

Comparison to other proteases utilizing a Ser/Lys dyad mechanism

Despite limited sequence identity the VP4 protease shows similarities in the binding site/active site region with many of the other Ser/Lys proteases. It was previously predicted from bioin-

formatics analysis that there would be structural similarities between VP4 and Lon protease.⁹ VP4 (protease family S50) is different from Lon protease (protease family S16) in that it does not require ATP for activity and that it does not form hexameric complexes. The crystal structures of the proteolytic domain of Lon protease from *E. coli* (PDB: 1RRE¹⁷), *Methanococcus jannaschii* (PDB: 1XHk²²) and *Archaeoglobus fulgidus* (PDB: 1Z0B²³) have been solved. Interestingly, a search for similar structures using the DALI server²⁴ reports only one protein with significant similarity in structure, the proteolytic domain of the *A. fulgidus* Lon protease (PDB: 1Z0B) (Figure 5(a)). It has a positional root-mean-square

deviation of 2.4 Å on 121 superimposed C α atoms. Figure 5(b) shows the superposition of the proteolytic domains of *E. coli* (PDB: 1RRE), *M. jannaschii* (PDB: 1XHK) and *A. fulgidus* (PDB: 1Z0B) Lon protease on BSNV VP4.

VP4 and Lon are similar in that the lysine general base comes from an α -helix (helix a2 in VP4; Figure 3) adjacent to the binding site rather than from the strand (strand b3 in VP4) that creates the mouth of the binding site, as seen in signal peptidase^{13,20,21} and Lex-like proteases.^{14–16} A comparison of the protein folds of signal peptidase and VP4 is shown by superposition of the *E. coli* signal peptidase active site on the VP4 active site (Figure 5(c)). Previous analysis has shown the catalytic core of UmuD' (and other LexA-like proteases) and signal peptidase to be similar.¹²

Despite only 16% sequence identity, the *E. coli* Lon protease^{17,22} is the most similar of the Lon protease structures to VP4 in the active site region (Figure 6(a)), although a superposition of VP4 and *E. coli* Lon reveals interesting differences, such as the size of the binding pocket. The binding pocket

of VP4 is significantly larger and more extended than that of Lon protease (Figure 6(a)). A direct comparison between the general base lysine conformation and environment in *E. coli* Lon protease and VP4 protease is made difficult by the fact that in the Lon structure the serine nucleophile has been mutated to an alanine and therefore cannot hydrogen bond with the lysine ϵ -amino group.¹⁷ In addition, the Lon structure has a sulfate ion, most likely from the crystallization conditions, interacting with the lysine ϵ -amino group, which could influence the conformation of the lysine side-chain.

Despite very different protein folds (Figure 5(c)), the strands that line the binding site region and represent most of the characterized hydrogen bonding interaction with inhibitors and substrates in signal peptidase²¹ and LexA repressor¹⁴ are superimposable with similar strands in the proposed binding site region of VP4 protease (Figure 6(b) and (c)).

A previous comparison of the micro-environment of the lysine general-base in signal peptidase and LexA-like proteases revealed a difference in the orientation of the N ζ atom of the lysine.²⁰ These

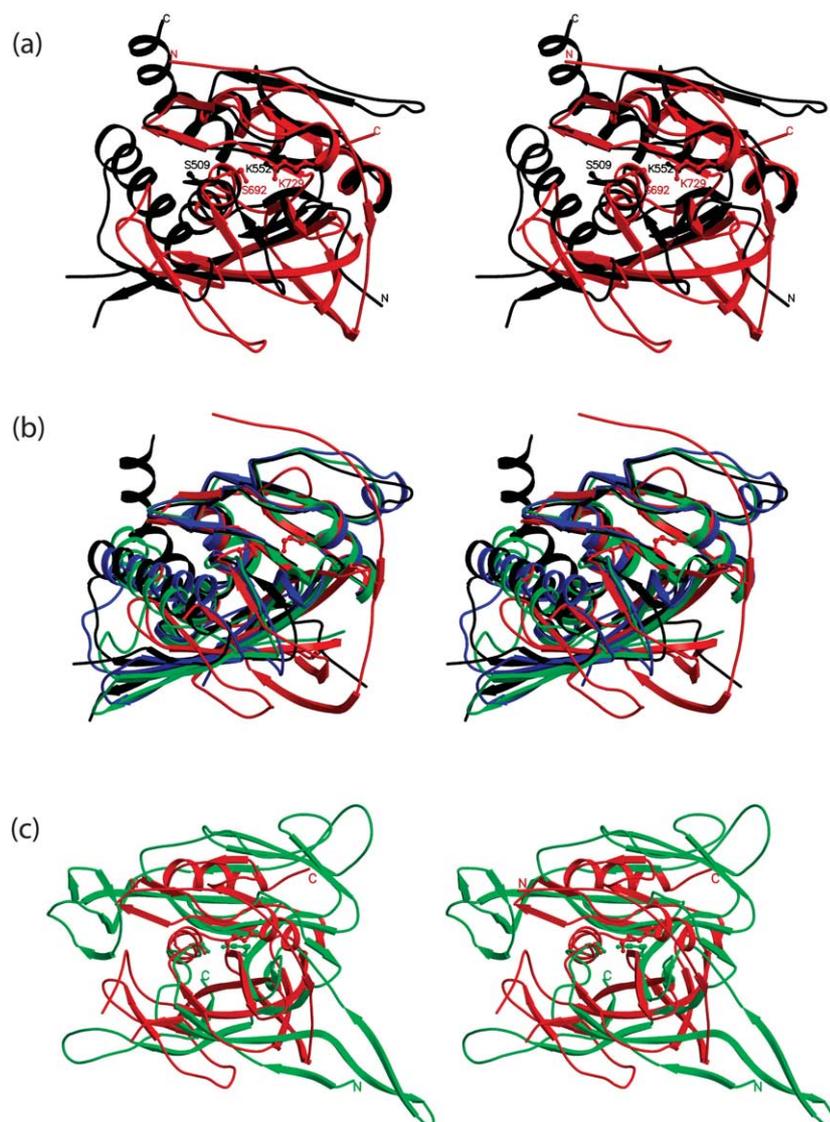


Figure 5. A comparison of the protein folds in the Ser/Lys proteases VP4 and Lon protease and signal peptidase. (a) The superposition of VP4 protease (red) and the proteolytic domain of the *A. fulgidus* Lon protease (black; PDB: 1Z0B). (b) The superposition of the proteolytic domains of *E. coli* (blue; PDB: 1RRE), *M. jannaschii* (green; PDB: 1XHK) and *A. fulgidus* (black; PDB: 1Z0B) Lon protease on VP4 (red). (c) The superposition of *E. coli* type I signal peptidase (green; PDB: 1B12) on VP4 (red).

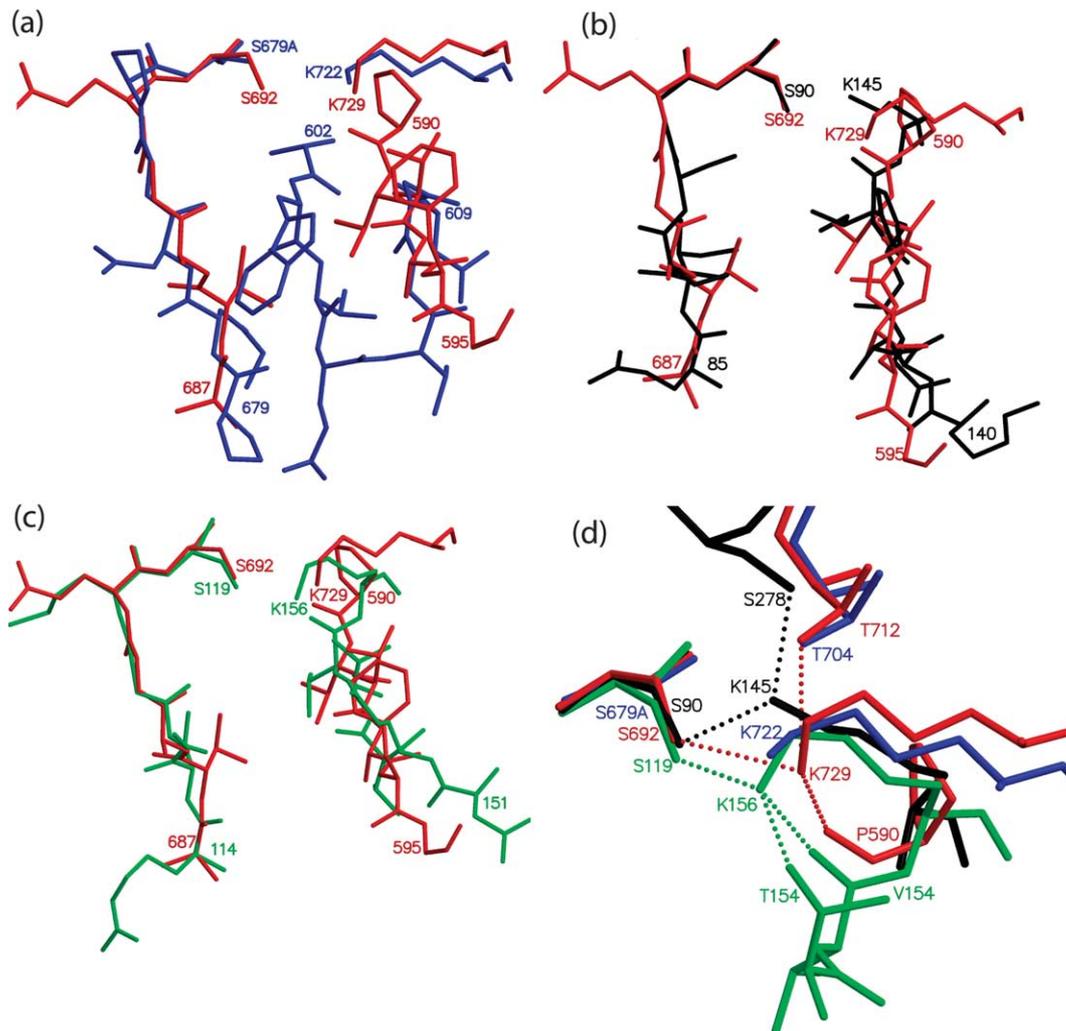


Figure 6. Comparison of the VP4 protease active site and binding site region with other Ser/Lys proteases. (a) A stick diagram of the superposition of the binding site region of BSNV VP4 protease (red) and *E. coli* Lon protease (blue; PDB: 1RRE). (b) A stick diagram of the superposition of the binding site region of BSNV VP4 protease (red) and *E. coli* signal peptidase (black; PDB:1KN9). (c) A stick diagram of the superposition of the binding site region of BSNV VP4 protease (red) and *E. coli* LexA protease (green; PDB: 1JHE). The side-chains of the catalytic residues and the end residues of the strands and loops that line the binding site regions are labeled. (d) A superposition of the catalytic residues of BSNV VP4 (red), *E. coli* signal peptidase (black), *E. coli* Lon protease (blue), *E. coli* LexA (green). Dotted lines represent hydrogen-bonding interactions.

different conformations are stabilized by a hydrogen bond to a hydroxyl group of a second (non-nucleophilic) serine or threonine, which is located in a different orientation with respect to the N^{ζ} atom of the lysine. Interestingly, relative to the nucleophile position (Ser O_{γ}), the VP4 lysine general base (Lys729, N^{ζ}) is in a position more similar to that seen in the LexA-like proteases than in signal peptidase.^{14–16,20} (Figure 6(d)).

Model of the X-VP4 cleavage site bound to the VP4 binding site

Because the binding site region of VP4 is similar to that of signal peptidase and LexA we are able to model in a peptide substrate (the X-VP4 cleavage site, P1' (Ala558) to P3 (Pro555)) into the binding site of VP4 (Figure 7), using as a guide the co-crystal

structures of signal peptidase with a non-covalently bound peptide-based inhibitor,²⁰ a covalently bound inhibitor complex with signal peptidase¹³ and LexA protease with its substrate cleavage site in its binding site.¹⁴ This model helps in the description of potential cleavage-site/binding-site interactions. The model is consistent with the P1–P3 residues of the substrate cleavage site region making antiparallel β -sheet interactions with the main-chain atoms in the residues from β -strand b3 (residues Val592 and Phe594) as well as the main-chain atoms from residues in the loop that leads to the nucleophilic residue Ser692 (residues Ala689, Gln691 and Ser692). The hydrogen bonding contacts observed between the docked in and energy minimized X-VP4 cleavage site peptide and the VP4 binding site include: Ala557 O (P1)···Ser692 N, Ala557 O (P1)···Gln691 N, Ala557 N (P1)···Ala689

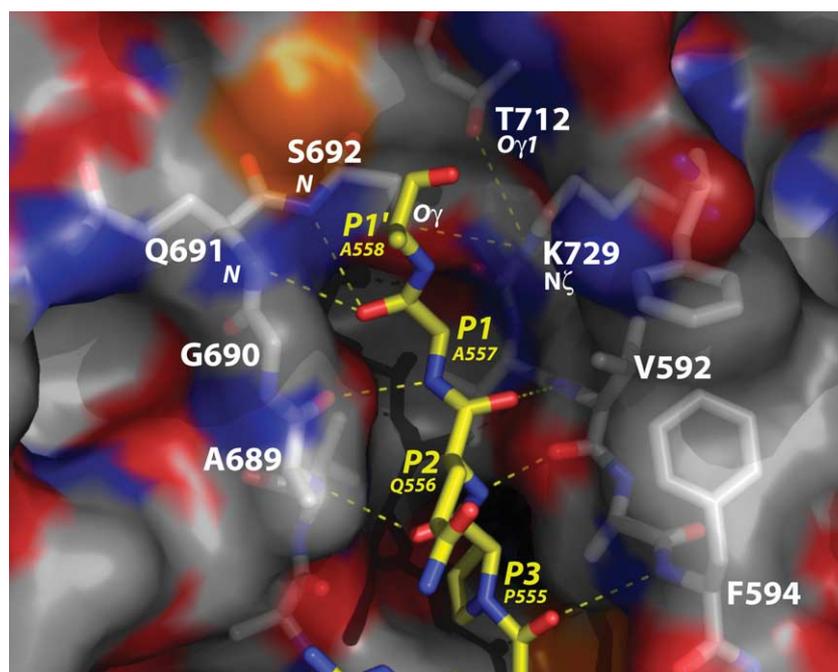


Figure 7. A model of the X-VP4 cleavage site (stick; yellow, carbon; blue, nitrogen; red, oxygen) docked into the binding site of BSNV VP4 (semi-transparent molecular surface; white, carbon; blue, nitrogen; red, oxygen). The positions of the P1' (A558) to P3 (P555) residues of the cleavage site are labeled in yellow. The positions of the residues that line the VP4 protease binding site pocket and residues and atoms critical to VP4 activity are labeled in white.

O, Gln556 O (P2)···Val592 N, Gln556 N (P2)···Val592 O, and Pro555 O (P3)···Ala689 N. The side-chains of P1 and P3 would point into the S1 and S3 binding pockets. The side-chains for the P2 and P1' residues of the cleavage-sites would be pointing away from the VP4 surface and into solvent (Figure 7), consistent with the variability seen in these residues, for example the Phe at the P1' position at the 417/418 cleavage site (Figure 1). The binding cleft bends just after the location of the P3 residue, consistent with the presence of a proline at this position in the consensus cleavage site (Figures 1 and 7).

Protein–protein interaction in the crystal reveals an N-terminal β -strand cross-over and interactions with the active-site

The VP4 crystal structure reveals that the two molecules in the asymmetric unit show an interesting extensive protein–protein interaction (Figure 8). The N-terminal β -strand (b1) makes a cross-over antiparallel β -sheet interaction with the neighboring VP4 molecule in the crystal (residues 562–569). In addition, a β -hairpin extension (β -strands b3 and b4, residues 592–600) from β -sheet 1 (light blue in Figure 2) reaches out to make contacts with the binding site of its neighboring molecule. The side-chain carboxylate from Asp596 sits right above the oxyanion hole of its neighboring molecule (Figure 8). Overall, the protein–protein interaction between molecules A and B within the asymmetric unit buries 1453 Å² of surface per molecule (14.4% of the total protein surface). The level of buried surface predicts that VP4 would exist as a dimer,²⁵ yet attempts to detect a dimer in solution have so far been unsuccessful. Furthermore, preliminary

mutagenesis experiments involving residues at the asymmetric unit interface show no effect on VP4 activity (data not shown).

Concluding Remarks

Structural analysis of the BSNV VP4 protease has allowed us to see for the first time a viral protease that has evolved the unusual serine/lysine dyad catalytic machinery, which has previously been seen in bacterial proteases and amidases. The structure now allows us to ask more detailed questions about the VP4 catalytic mechanism such as the role of Thr712 in its interaction with the lysine general base (Lys729). The solubility and stability of this construct of VP4 protease may make possible the measurement of the pK_a value for the VP4 lysine general base (Lys729 ϵ -amino group) with and without bound peptides by NMR titrations. Co-crystallization with peptides will be instrumental in confirming the cleavage-site–binding site interactions we have observed from our docking and modeling studies. The structure of VP4 gives us the relative spatial arrangement for the VP4 amino and carboxy termini relative to the binding pocket orientation and may help in addressing the questions pertaining to the order and timing of cleavage events and also whether this protease catalyzes the cleavage of its viral polyprotein in an intramolecular (*cis*) or an intermolecular (*trans*) reaction. The extensive protein–protein interaction seen in the VP4 crystal packing has so far not been shown to be of functional significance but it will continue to be explored in future studies.

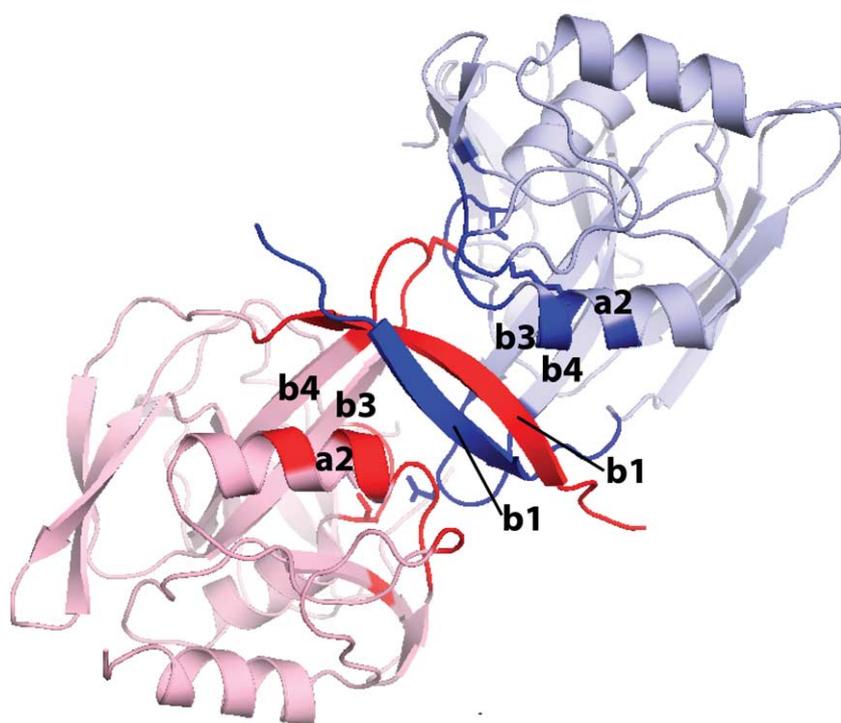


Figure 8. An extensive protein–protein interaction is observed in the asymmetric unit of the VP4 crystal structure. Molecule A is rendered as a light pink ribbon and molecule B is rendered as a light blue ribbon. Those residues making intermolecular contacts are highlighted in dark red (molecule A) and dark blue (molecule B). The secondary structural elements involved in the interaction are labeled. The N-terminal β -strand of each molecule (b1) forms anti-parallel hydrogen bonding interactions (dark red and dark blue). The carboxylate group from Asp596, which resides on a loop between strands b3 and b4 (residues 595–599), sits right above the oxy-anion hole of its neighboring molecule in the crystal.

Materials and Methods

Cloning, mutagenesis, expression and purification

The cloning of the entire open reading frame of BSNV VP4 (residues 558–791) with a methionine N-terminal residue preceding Ala558 in pET-28b+ was described.⁷ To promote crystallization this original expression construct was altered by site-directed mutagenesis to change the codon encoding Pro774 to a stop codon thereby arresting translation at Arg773. The new construct therefore eliminated 18 amino acid residues at the C terminus of VP4. The Quick Change kit (Stratagene) was used for site-directed mutagenesis. The resulting expressed protein is a 217 residues long polypeptide (residues 558–773). Its calculated molecular mass is 23,441 Da and the theoretical *pI* value is 5.7.

The truncated vector construct described above (pET28_bsnvVP4_558-773) was transformed into the *E. coli* strain Tuner (DE3). For overexpression of selenomethionine incorporated protein, an overnight culture was prepared in M9 minimal media. 10 ml of this overnight culture was used to inoculate one liter of M9 minimal media with kanamycin (0.05 mg/ml) and grown for 7 h. The following amounts of amino acids were then added to the culture as solids (100 mg of lysine, phenylalanine, threonine; 50 mg of isoleucine, leucine, valine; 60 mg of L-selenomethionine). After 15 min the expression was induced by adding 0.5 ml of 1 M IPTG. Cells were induced for 3 h and the pellet was kept frozen at -80°C .

Frozen cell pellets from one liter of induced culture (2 g wet weight) were completely resuspended in buffer A (50 mM Tris–HCl (pH 8.0), 10% (v/v) glycerol, 10 mM DTT, 7 mM magnesium acetate, 0.1% (v/v) Triton-X100). The cells were lysed by overnight incubation at 4°C after adding lysozyme (0.2 mg/ml) and benzonase (1 unit/ml). The lysate was centrifuged at 14,000 rpm for 30 min

and the clear supernatant was decanted for ammonium sulfate fractionation.

Ammonium sulfate (AS) fractionation was carried out at 30, 40, 50 and 60% (w/v) AS concentrations. After adding the AS, the supernatant was incubated on ice for 30 min, then, the precipitates were pelleted by centrifugation at 14,000 rpm for 30 min. Pellets were dissolved in approximately 1 ml of buffer B (20 mM Tris–HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA, 10% glycerol, 10 mM DTT). The BSNV VP4_558-773 precipitated at the 40% and 50% AS concentrations. These fractions were pooled and dialyzed against buffer A.

The dialyzed sample (~ 2 ml) was applied to a 0.7 ml Q-Sepharose anion exchange column equilibrated with buffer B. The protein was then eluted from the column using a NaCl step-gradient.

The pooled fractions containing BSNV VP4_558-773 were then applied to a size-exclusion column (HiPrep 16/60 Sephacryl S-100 HR) equilibrated in buffer C (20 mM Tris (pH 8.0), 100 mM NaCl, 10% glycerol, 1% (v/v) β -mercaptoethanol) and ran at 1 ml/min; 3 ml fractions were collected.

SDS-PAGE analysis of the collected fractions demonstrated a single band corresponding to VP4_558-773. The protein was concentrated to 23 mg/ml using a Millipore centrifugal filter (5000 cut off).

Crystallization

Crystallization was carried out by the sitting drop vapor diffusion method at 21°C . 1 μl of the protein sample was mixed thoroughly with 1 μl of the reservoir solution and the mixture was equilibrated against 1 ml of the reservoir solution. The protein was at a concentration of 23 mg/ml suspended in buffer C. Optimized crystallization reservoir solution conditions consisted of 0.1 M Tris–HCl (pH 8.5), 25% (w/v) PEG2000 MME, 60 mM $\text{Mg}(\text{OAc})_2$. Cryo-solvent conditions were the same as the crystallization conditions supplemented with 20%

glycerol. Crystals were flash cryo-cooled in liquid nitrogen.

Data collection

Diffraction data were collected on the selenomethionine incorporated crystals at beamline X26C, at Brookhaven National Laboratory, National Synchrotron Light Source using a Q4 CCD detector and a Nonius CAD4 Goniometer. The crystal-to-detector distance was 200 mm. Data were collected with 1° oscillations, and each image was exposed 30 s. Data were processed to 2.2 Å resolution with the program HKL2000.²⁶ See Table 1 for data collection statistics.

Structure determination and refinement

Single-wavelength anomalous dispersion (SAD) phasing was carried out using a data set collected at the peak wavelength (0.978047 Å) with the program autoSHARP.²⁷ Values for f' and f'' were found to be -6.16 and 4.06, respectively. Eight of the ten possible selenium sites were found. To improve the quality of the electron density

maps we used the program SOLOMON,²⁸ run within autoSHARP, to perform a density modification based on solvent flipping. The program Arp/Warp²⁹ automatically built approximately 80% of the polypeptide chains. The rest of the model was built and fit using the program XFIT within the suite XTALVIEW.³⁰ The structure has been refined using the programs CNS³¹ and Refmac5³² to an R_{cryst} of 22.9% and an R_{free} of 27.9%. The average B value for all atoms is 27.6. There are 401 residues and 400 water molecules in the model with two molecules in the asymmetric unit. The refined model includes residues 559 to 769 for each molecule. There is no visible electron density in either molecule in the asymmetric unit for residues 637 to 648. These residues are located at the surface of the molecule and presumably form a mobile loop. There is also no electron density for the N-terminal three residues and the last five C-terminal residues. The stereochemistry of the structure was analyzed with the programs PROCHECK.³³ For further data collection, structure solution and refinement details see Table 1.

Structural analysis

The secondary structural and protein topology analysis was performed with the programs PROMOTIF³⁴ and HERA,³⁵ respectively. The programs SUPERIMPOSE³⁶ and SUPERPOSE³⁷ were used to superimpose molecules for comparing the different molecules in the asymmetric unit and for comparison to other proteases. The program CONTACT within the program suite CCP4³⁸ was used to measure the hydrogen bonding and van der Waals contacts within the enzyme. The program CASTp³⁹ was used to analyze the molecular surface and measure the substrate binding site. The program SURFACE RACER 1.2⁴⁰ was used to measure the solvent accessible surface of the protein and individual atoms within the protein. A probe radius of 1.4 Å was used in the calculations. The Protein-Protein Interaction Server† was used to analyze the interactions between the molecules in the asymmetric unit in the VP4 crystal.

Modeling of the X-VP4 cleavage site into the VP4 protease binding site

A model of the X-VP4 cleavage site (P1' (A558) to P4 (R554)) bound to the VP4 binding site was prepared using the structure of the covalently bound penem inhibitor complex with signal peptidase¹³ (PDB: 1B12) to guide the position of the P1 Ala residue of the substrate. The positions of the P2, P3 and P4 residues were guided by the structure of the cleavage site of LexA¹⁴ (PDB: 1JHE) and the co-crystal structure of a non-covalently bound peptide-based inhibitor with signal peptidase²¹ (PDB: 1TT7D). The model was prepared using the program XFIT within the suite XTALVIEW.³⁰ The model was energy-minimized using the program CNS.³¹

Figure preparation

Figures were prepared using the programs Molscript,⁴¹ Raster3D⁴² and PyMol‡.

Table 1. Data collection, phasing and refinement statistics

<i>A. Crystal parameters</i>	
Space group	I23
a, b, c , (Å)	144.05
α, β, γ (deg.)	90
<i>B. Data collection</i>	
Wavelength	0.97804
Resolution (Å)	33.9–2.2
Number of reflections	416,066 (21,626)
Number of unique reflections	25,311 (5219)
R -merge ^a	0.059 (0.256)
Mean $\langle I \rangle / \sigma(I)$	76.9 (14.3)
Completeness	100.0 (100.0)
Redundancy	16.4 (14.4)
Anomalous completeness	95.8 (78.7)
Anomalous multiplicity	6.3 (2.1)
<i>C. Phasing statistics</i>	
Number of sites	8
Number of acentric reflections	38,524
Overall FOM acentric ^b	0.21053
Number of centric reflections	1988
Overall FOM centric ^b	0.01738
<i>D. Refinement statistics</i>	
Protein molecules (chains) in au	2
Residues	401
Water molecules	400
Total number of atoms	3398
$R_{\text{cryst}}^c / R_{\text{free}}^d$ (%)	22.6/27.9
Average B -factor (Å ²) (all atoms)	27.6
Estimated overall coordinate error (Å)	0.150
rms deviation on angles (deg.)	1.70
rms deviation on bonds (Å)	0.016

The data collection statistics in parentheses are the values for the highest resolution shell (2.28 Å to 2.20 Å).

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity, and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.

^b FOM, figure of merit = $\langle |\sum P(\alpha) e^{i\alpha} / \sum P(\alpha)| \rangle$, where α is the phase angle and $P(\alpha)$ is the phase probability distribution.

^c $R_{\text{cryst}} / R_{\text{free}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

^d R_{free} is calculated using 5% of the reflections randomly excluded from refinement.

† <http://www.biochem.ucl.ac.uk/bsm/PP/server/>

‡ <http://pymol.sourceforge.net/>

Protein Data Bank accession codes

Atomic coordinates and experimental structure factors of BSNV VP4 protease have been deposited with the RCSB Protein Data Bank⁴³ under the accession code 2GEF.

Acknowledgements

This work was supported in part by a Canadian Institute of Health Research operating grant (to M.P.), a National Science and Engineering Research Council of Canada operating grant (to M.P.), a Michael Smith Foundation for Health Research Scholar award (to M.P.), a Canadian Foundation of Innovation grant (to M.P.), and a Cystic Fibrosis Foundation Postdoctoral Fellowship Award (to A.F.). B.D. acknowledges support from the ACI Microbiologie for the French MRT. We thank Dr Clemens Vornrhein for his kind help with auto-SHARP. We thank Dr Annie Heroux at the Brookhaven National Laboratory NSLS beam line X26C. We thank Dr R. M. Sweet and all the RapidData 2005 staff. We thank Charles S. Bond for his help in obtaining the Linux version of the program HERA.

References

- Babe, L. M. & Craik, C. S. (1997). Viral proteases: evolution of diverse structural motifs to optimize function. *Cell*, **91**, 427–430.
- Tong, L. (2002). Viral proteases. *Chem. Rev.* **102**, 4609–4626.
- Dobos, P., Hill, B. J., Hallett, R., Kells, D. T., Becht, H. & Teninges, D. (1979). Biophysical and biochemical characterization of five animal viruses with bisegmented double-stranded RNA genomes. *J. Virol.* **32**, 593–605.
- Petit, S., Lejal, N., Huet, J. C. & Delmas, B. (2000). Active residues and viral substrate cleavage sites of the protease of the birnavirus infectious pancreatic necrosis virus. *J. Virol.* **74**, 2057–2066.
- Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S. *et al.* (2005). The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell*, **120**, 761–772.
- John, K. R. & Richards, R. H. (1999). Characteristics of a new birnavirus associated with a warm-water fish cell line. *J. Gen. Virol.* **80**, 2061–2065.
- Da Costa, B., Soignier, S., Chevalier, C., Henry, C., Thory, C., Huet, J. C. & Delmas, B. (2003). Blotched snakehead virus is a new aquatic birnavirus that is slightly more related to avibirnavirus than to aquabirnavirus. *J. Virol.* **77**, 719–725.
- Lejal, N., Da Costa, B., Huet, J. C. & Delmas, B. (2000). Role of Ser-652 and Lys-692 in the protease activity of infectious bursal disease virus VP4 and identification of its substrate cleavage sites. *J. Gen. Virol.* **81**, 983–992.
- Birghan, C., Mundt, E. & Gorbalenya, A. E. (2000). A non-canonical lon proteinase lacking the ATPase domain employs the Ser-Lys catalytic dyad to exercise broad control over the life cycle of a double-stranded RNA virus. *EMBO J.* **19**, 114–123.
- Rawlings, N. D. & Barrett, A. J. (1999). MEROPS: the peptidase database. *Nucl. Acids Res.* **27**, 325–331.
- Paetzel, M. & Dalbey, R. E. (1997). Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.* **22**, 28–31.
- Paetzel, M. & Strynadka, N. C. (1999). Common protein architecture and binding sites in proteases utilizing a Ser/Lys dyad mechanism. *Protein Sci.* **8**, 2533–2536.
- Paetzel, M., Dalbey, R. E. & Strynadka, N. C. (1998). Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature*, **396**, 186–190.
- Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M. *et al.* (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell*, **106**, 1–10.
- Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M. (2000). Crystal structure of the lambda repressor C-terminal domain provides a model for cooperative operator binding. *Cell*, **101**, 801–811.
- Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R. & Hendrickson, W. A. (1996). Structure of the UmuD' protein and its regulation in response to DNA damage. *Nature*, **380**, 727–730.
- Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F. *et al.* (2004). The catalytic domain of Escherichia coli Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J. Biol. Chem.* **279**, 8140–8148.
- Burgi, H. B., Dunitz, J. D. & Shefter, E. (1973). Geometric reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* **95**, 5065–5067.
- Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Paetzel, M., Dalbey, R. E. & Strynadka, N. C. (2002). Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser-Lys dyad mechanism. *J. Biol. Chem.* **277**, 9512–9519.
- Paetzel, M., Goodall, J. J., Kania, M., Dalbey, R. E. & Page, M. G. (2004). Crystallographic and biophysical analysis of a bacterial signal peptidase in complex with a lipopeptide-based inhibitor. *J. Biol. Chem.* **279**, 30781–33090.
- Im, Y. J., Na, Y., Kang, G. B., Rho, S. H., Kim, M. K., Lee, J. H. *et al.* (2004). The active site of a lon protease from *Methanococcus jannaschii* distinctly differs from the canonical catalytic Dyad of Lon proteases. *J. Biol. Chem.* **279**, 53451–53457.
- Botos, I., Melnikov, E. E., Cherry, S., Kozlov, S., Makhovskaya, O. V., Tropea, J. E. *et al.* (2005). Atomic-resolution crystal structure of the proteolytic domain of *Archaeoglobus fulgidus* lon reveals the conformational variability in the active sites of lon proteases. *J. Mol. Biol.* **351**, 144–157.
- Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucl. Acids Res.* **26**, 316–319.
- Ponstingl, H., Henrick, K. & Thornton, J. M. (2000). Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins: Struct. Funct. Genet.* **41**, 47–57.
- Otwinowski, Z. (1993). Denzo. In *Denzo* (Sawyer, L., Isaacs, N. & Baily, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
- Vornrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2005). Automated structure solution with auto-

- SHARP. In *Crystallographic Methods* (Doublié, S., ed.), Humana Press, Totowa, NJ.
28. Abrahams, J. P. & Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallog. sect. D*, **52**, 30–42.
 29. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244.
 30. McRee, D. E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165.
 31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
 32. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D*, **57**, 122–133.
 33. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
 34. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
 35. Hutchinson, E. G. & Thornton, J. M. (1990). HERA—a program to draw schematic diagrams of protein secondary structures. *Proteins: Struct. Funct. Genet.* **8**, 203–212.
 36. Diederichs, K. (1995). Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. *Proteins: Struct. Funct. Genet.* **23**, 187–195.
 37. Maiti, R., Van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucl. Acids Res.* **32**, W590–W594.
 38. C.C.P. no. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
 39. Liang, J., Edelsbrunner, H. & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.
 40. Tsodikov, O. V., Record, M. T., Jr & Sergeev, Y. V. (2002). Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* **23**, 600–609.
 41. Kraulis, P. G. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
 42. Meritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
 43. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.

Edited by M. Guss

(Received 19 November 2005; received in revised form 10 February 2006; accepted 15 February 2006)
Available online 6 March 2006