THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 14, pp. 12475–12482, April 8, 2011


# Crystal Structure of a Viral Protease Intramolecular Acyl-enzyme Complex

## *INSIGHTS INTO cis-CLEAVAGE AT THE VP4/VP3 JUNCTION OF TELLINA BIRNAVIRUS**[S]

Received for publication, October 28, 2010, and in revised form, January 10, 2011 Published, JBC Papers in Press, February 2, 2011, DOI 10.1074/jbc.M110.198812

**Ivy Yeuk Wah Chung and Mark Paetzel**[1]

*From the Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada*

**Viruses of the *Birnaviridae* family are characterized by their bisegmented double-stranded RNA genome that resides within a single-shelled non-enveloped icosahedral particle. They infect birds, aquatic organisms, and insects. *Tellina virus 1* (TV-1) is an *Aquabirnavirus* isolated from the mollusk *Tellina tenuis*. It encodes a polyprotein ($NH_2$-pVP2-X-VP4-VP3-COOH) that is cleaved by the self-encoded protease VP4 to yield capsid precursor protein pVP2, peptide X, and ribonucleoprotein VP3. Here we report the crystal structure of an intramolecular (*cis*) acyl-enzyme complex of TV-1 VP4 at 2.1-Å resolution. The structure reveals how the enzyme can recognize its own carboxyl terminus during the VP4/VP3 cleavage event. The methyl side chains of $Ala^{830}$(P1) and $Ala^{828}$(P3) at the VP4/VP3 junction point into complementary shallow and hydrophobic S1 and S3 binding pockets adjacent to the VP4 catalytic residues: nucleophile $Ser^{738}$ and general base $Lys^{777}$. The electron density clearly shows that the carbonyl carbon of $Ala^{830}$ is covalently attached via an ester bond to the Oγ of $Ser^{738}$. A highly ordered water molecule in the active site is coordinated in the proper position to act as the deacylating water. A comparative analysis of this intramolecular (*cis*) acyl-enzyme structure with the previously solved intermolecular (*trans*) acyl-enzyme structure of infectious pancreatic necrosis virus VP4 explains the narrower specificity observed in the cleavage sites of TV-1 VP4.**

Viruses belonging to the family *Birnavirdae* infect vertebrates, mollusks, insects, and rotifers (1). The family can be further divided into three genera: *Avibirnavirus*, *Entomobirnavirus*, and *Aquabirnavirus*. Birnaviruses that are of significant economic interest include the *Avibirnavirus* infectious bursal disease virus, which infects chickens, and the *Aquabirnavirus* infectious pancreatic necrosis virus (IPNV)[2] which infects salmon (2, 3). *Tellina virus-1* (TV-1) is an *aquabirnavirus* that shares similar morphological and physicochemical properties with IPNV (4).

Birnavirus carries a bi-segmented (segment A and B) double-stranded RNA genome that is encapsulated in an icosahedral capsid of 60 to 70 nm in diameter (1, 5). In segment A, the largest open reading frame encodes a polyprotein that has the conserved arrangement of $NH_2$-pVP2-VP4-VP3-COOH with the exception of gene X, which exists between pVP2 and VP4 in TV-1 (Fig. 1) and *Blotched snakehead virus* (BSNV) (6, 7). The polyprotein is processed by the self-encoded protease viral protein 4 (VP4) to generate the capsid precursor protein pVP2, VP4 itself, and ribonucleoprotein VP3 (6, 8–10). Further processing of pVP2 yields the capsid protein VP2 and additional peptides (6, 11, 12). Protein VP3 associates with the genome to form a ribonucleoprotein complex (13). The exact positions of the major cleavage sites within segment A have been confirmed by mass spectroscopy or N-terminal sequencing for infectious bursal disease virus (14), IPNV (10), BSNV (7), *Drosophila* X virus (DXV) (15), and TV-1 (6). In general, small uncharged residues such as alanine and serine are most frequently found at the P3, P1, and P1′ positions of these birnavirus polyprotein cleavage sites.

The VP4 protease from birnavirus is a serine endoprotease that utilizes a serine-lysine catalytic dyad mechanism as opposed to the classical serine-histidine-aspartate catalytic triad mechanism (16–18). Other proteases that utilize the serine-lysine dyad mechanism include the following bacterial proteases: type 1 signal peptidase (19), UmuD (20), LexA (21), Lon protease (22), and signal peptide peptidase (SppA) (23). Besides the unusual feature of utilizing a lysine ϵ-amino group as a general base, these enzymes are also unusual in that their serine Oγ nucleophile attacks the scissile bond from the *si*-face rather than the *re*-face as is seen in most serine proteases.

TV-1 was first isolated from the sand dwelling marine bivalve mollusk *Tellina tenuis* (24). Manifestations of the disease include a thin and chalky shell as well as a pale yellow digestive gland. TV-1 VP4 is encoded as part of the 1114-residue long segment A polyprotein. It cleaves after amino acid residue positions 512, 618, and 830 to yield capsid precursor protein pVP2, peptide X, VP4 itself, and ribonucleoprotein VP3 (Fig. 1*A*) (6, 8). Additional cleavage sites were found at the C-terminal side of amino acids 451, 492, and 499 within the pVP2 region (Fig. 1*A*) (6). The cleavage sites identified can be described by the consensus Ala-*X*-Ala ↓ (Fig. 1*B*) (6). Site-directed mutagenesis studies on TV-1 VP4 are consistent with residue $Ser^{738}$ acting

* This work was supported in part by grants from the Canadian Institute of Health Research (to M. P.), the National Science and Engineering Research Council of Canada (to M. P.), the Michael Smith Foundation for Health Research (to M. P.), and the Canadian Foundation of Innovation (to M. P.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2.

*The atomic coordinates and structure factors (code 3P06) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

[1] To whom correspondence should be addressed: South Science Bldg., 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada. Tel.: 778-782-4230; Fax: 778-782-5583; E-mail: mpaetzel@sfu.ca.

[2] The abbreviations used are: IPNV, infectious pancreatic necrosis virus; TV-1, *Tellina virus 1*; BSNV, blotched snakehead virus; VP4, viral protein (protease) 4 in the poly-protein (pVP2-VP4-VP3) coded for in birnavirus; r.m.s., root mean square.

as the nucleophile and residue $Lys^{777}$ serving as the general base (6). Like other VP4s, TV-1 VP4 belongs to the evolutionary clan SJ according to the MEROPS protease data base (25). However, the amino acid sequence of TV-1 VP4 is more divergent from the rest of the VP4s and is classified into family S69 instead of family S50 in which the IPNV, infectious bursal disease virus, BSNV, and DXV VP4s have been assigned.

The first VP4 structure solved was from BSNV; however, no substrate was bound in its active site (26). In contrast, the crystal structure of IPNV VP4 revealed an intermolecular (*trans*) acyl-enzyme complex (27). Here we report the structure of VP4 from TV-1 at 2.1-Å resolution with the C terminus binding into its own active site forming an intramolecular (*cis*) acyl-enzyme complex.

## MATERIALS AND METHODS

*Cloning*—The DNA region encoding full-length TV-1 VP4 (residue 619–830, Swiss-Prot accession number Q2PBR5) was first amplified by polymerase chain reaction (PCR) using Vent DNA polymerase (New England Biolab). The forward primer has the sequence 5′-AGGCCCATATGGCCGACAGGCCCATGATC-3′ and the reverse primer has the sequence 5′-CTTGTGAAGGCGGCCGCTCATGCTTGCGCCACGTTCTTTCCGGAGGAGAA-3′. The restriction enzyme-digested PCR product was then cloned into restriction sites NdeI and NotI of plasmid $pET28b^{+}$. This allowed for the incorporation of an amino-terminal histidine tag when the recombinant protein was expressed. The plasmid was then transformed into NovaBlue cells (Novagen) for plasmid isolation. The sequence of the DNA insert was verified by DNA sequencing. The N terminus of the expressed protein has an extra 21 amino acids (MGSSHHHHHHSSGLVPRGSHM) that includes the $His_6$ affinity tag and the linker region. The calculated molecular mass for this construct is 24,745 Da (233 amino acids, calculated isoelectric point, 9.9).

*Protein Purification*—Cells for protein overexpression were prepared by transforming the plasmid containing the gene for TV-1 VP4 described above into *Escherichia coli* strain *Tuner* (*DE3*) followed by selection on LB plates with 0.05 mg/ml of kanamycin. For overexpression, a cell pellet from 100 ml of overnight culture grown in LB medium supplemented with 0.05 mg/ml of kanamycin was used to inoculate each liter of M9 minimal medium. Four liters of culture were grown for each batch of protein expressed and purified. After shaking at 37 °C for 8 h, a mixture of L-amino acids (100 mg of lysine, phenylalanine, and threonine; 50 mg of isoleucine, leucine, and valine) and 60 mg of selenomethionine were added to each liter of culture. Each liter of culture was induced with 0.5 ml of 1 M isopropyl β-D-1-thiogalactopyranoside after 15 min. The induced cultures were allowed to grow overnight at 25 °C and were then harvested by centrifugation at 9,110 × *g* for 7 min. The cell pellet was stored at −80 °C for 15 min to facilitate cell lysis. The frozen cell pellet was re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 10% glycerol, 1 mM dithiothreitol (DTT), 7 mM magnesium acetate, 0.1% Triton X-100, 1 units/ml of benzonase, and 0.2 mg/ml of lysozyme) and incubated at 4 °C overnight with gentle agitation. The cell debris was removed by centrifugation at 28,964 × *g* and the clear supernatant was loaded onto a 5-ml nickel-nitrilotriacetic acid metal affinity column (Qiagen). A step gradient containing 100, 300, and 600 mM imadizole in standard buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10% glycerol, and 1 mM DTT) was used to elute the histidine-tagged VP4. Fractions positive for VP4 were then applied to a 5-ml SP-Sepharose FF cation-exchange column equilibrated with standard buffer and eluted stepwise with 100, 300, and 500 mM NaCl in standard buffer. Fractions positive for VP4 were pooled and loaded onto a size exclusion column (HiPrep 16/60 Sephacryl S-100 HR) equilibrated with crystallization buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10% glycerol, and 1% β-mercaptoethanol). The size exclusion column was connected to a Pharmacia AKTA Prime™ system that pumped at a flow rate of 0.7 ml/min. Fractions with pure VP4 were pooled and concentrated using a Millipore centrifugal filter (10 kDa cutoff). The concentrated sample was incubated with chymotrypsin overnight at 4 °C and then applied to the same size exclusion column mentioned above. Details of this limited proteolysis procedure are described elsewhere (28). Purified VP4 was concentrated to ~40 mg/ml for crystallization trials.

*Crystallization*—The crystal used for data collection was obtained using the hanging-drop method at room temperature (~23 °C). On a coverslip, 1 μl of VP4 was mixed with 1 μl of reservoir reagent and 1 μl of 0.2 M urea as an additive. To aid in crystal nucleation, this drop was seeded with 1 μl of selenomethionine-labeled TV-1 VP4 crystals from an older drop. The drop was allowed to reach vapor equilibrium via incubation over 1 ml of reservoir reagent in a grease-sealed chamber. The optimized reservoir condition was 21% PEG8000, 0.55 M ammonium sulfate. The cryosolution contained 70% of the buffer reservoir and 30% glycerol. The crystal was transferred into the cryosolution, flash-cooled in liquid nitrogen, and then subjected to diffraction analysis. These hexagonal crystals belong to space group $P6_422$, have unit cell dimensions of 59.1 × 59.1 × 208.1 Å with one molecule in the asymmetric unit, a Matthews coefficient of 2.1, and a solvent content of 42.1%.

*Data Collection*—Data collection was carried out at the Canadian Light Source on beam line 08ID-1 at a wavelength of 0.9789 Å with 0.5 degree oscillations and each image was exposed for 5 s. The distance between the crystal and the detector was 350 mm.

*Structure Determination and Refinement*—The diffraction images were indexed and integrated using the program MOSFLM and scaled to 2.1 Å using SCALA (29). The structure was solved by single-wavelength anomalous dispersion (f′, −7.11; f″, 4.77) using the program AutoSol from the Phenix program suite (30–32). The coordinates from AutoSol were refined using rigid body and restrained refinement in REFMAC5 (33). The model from AutoSol was 90% complete. The remaining 10% of the molecule was built manually in COOT (34) using a difference density map generated by REFMAC5. Restraints for the acyl-enzyme linkage were prepared using the REFMAC5 library definitions (35). The final round of restrained refinement in REFMAC5 included TLS (5 TLS groups) (36, 37).

*Structural Analysis*—Residue contact and hydrogen bond analysis was performed using the program CONTACT in the

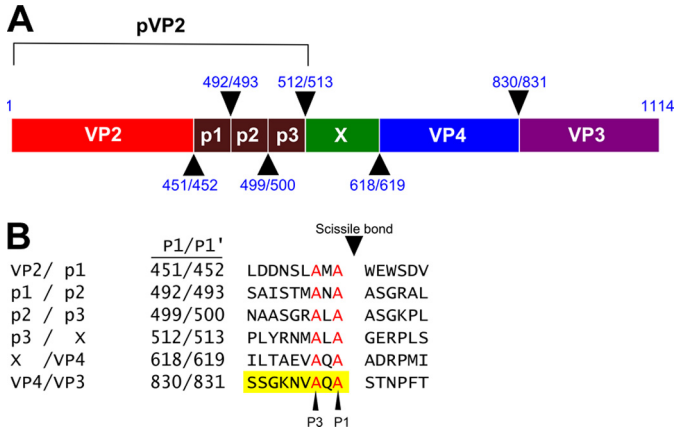


FIGURE 1. **The TV-1 VP4 protease cleavage sites.** *A*, the genomic arrangement of TV-1 segment A. The *arrowheads* indicate the sites of cleavage with the P1 and P1′ residue numbers listed. *B*, a list of known TV-1 VP4 cleavage sites (6). The amino acids in the VP4/VP3 junction that are visible in the crystal structure are highlighted in *yellow*.

CCP4i crystallography software suite (29). The root mean square deviation (r.m.s. deviation) values reported for protein superpositions were calculated using SUPERPOSE (38). The secondary structure assignments were performed using STRIDE (39). The accessible surface area analysis was performed with Surface Racer 5.0 (40).

*Figure Preparation*—The structure figures were generated in PyMOL (41). The protein sequence alignment was prepared using ClustalW (42) and ESPript (43). The electron density map used in Fig. 3 was prepared from the mtz file generated in REFMAC5 using program FFT from the CCP4i crystallography software suite (29).

## RESULTS

*Structure Solution*—The amino terminus of TV-1 VP4 (residue 619) is defined by its X/VP4 cleavage site and its carboxyl terminus is defined by its VP4/VP3 cleavage site (residue 830) (Fig. 1). We have successfully cloned, expressed, and purified full-length wild-type TV-1 VP4 protease (residue 619–830) with an amino-terminal $His_6$ affinity tag. Limited proteolysis was used to remove disordered amino-terminal residues to promote crystallization. We have crystallized and solved the structure of TV-1 VP4 using single wavelength anomalous dispersion phasing methods and refined the structure (residues 637–830) to 2.1-Å resolution (Table 1).

*Overall Protein Architecture*—The overall dimension of TV-1 VP4 is ~38 Å × 34 Å × 39 Å. Electrostatic analysis reveals a significant amount of positively charged surface, consistent with its theoretical isoelectric point of 9.8 (residues 619–830). This is a unique characteristic of TV-1 VP4, the theoretical isoelectric point of VP4 from other birnaviruses ranges from 4.7 to 6.6. TV-1 VP4 has an $\alpha/\beta$-fold consisting of 16 $\beta$-strands ($\beta$1–$\beta$16), four $\alpha$-helices ($\alpha$1–$\alpha$4), and two $3_{10}$ helices ($\eta$1–$\eta$2) (Fig. 2). $\beta$-Strands 1 to 9, 15, and 16 form two $\beta$-sheets that are predominantly arranged in an anti-parallel fashion. This $\beta$-sheet region houses the substrate binding groove and specificity pockets. One $3_{10}$-helix and one $\alpha$-helix ($\alpha$1) are also found in this region. Adjacent to this $\beta$-sheet platform is a parallel $\beta$-sheet made up of three short strands ($\beta$10, $\beta$13, and $\beta$14).

**TABLE 1**
**Data collection, phasing, and refinement statistics**

| | |
|---|---|
| **Crystal parameters** | |
| Space group | $P6_422$ |
| *a*, *b*, *c* (Å) | 59.1, 59.1, 208.1 |
| **Data collection statistics** | |
| Wavelength (Å) | 0.97893 |
| Resolution (Å) | 52.0-2.1(2.2-2.1)[a] |
| Total reflections | 154,167 (9,938) |
| Unique reflections | 13,466 (1,841) |
| $R_{merge}$ [b] | 0.107 (0.350) |
| Mean $(I)/\sigma(I)$ | 14.5 (4.3) |
| Completeness (%) | 99.8 (98.5) |
| Redundancy | 11.4 (5.4) |
| **SAD phasing** | |
| Number of sites | 7 (out of a possible 7) |
| Figure of merit (49.7-2.5 Å) | 0.46 |
| **Refinement statistics** | |
| Protein molecules (chains) in absorbance units | 1 |
| Residues | 194 |
| Water molecules | 47 |
| Total number of atoms | 1503 |
| $R_{cryst}$ [c]/$R_{free}$ [d] (%) | 18.2/23.3 |
| Average *B*-factor (Å$^2$) (all atoms) | 28.4 |
| R.m.s. deviation on angles (°) | 1.979 |
| R.m.s. deviation on bonds (Å) | 0.024 |

[a] The data collection statistics in brackets are the values for the highest resolution shell.
[b] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection, and $\langle I(hkl)\rangle$ is the mean intensity of that reflection.
[c] $R_{cryst} = \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure-factor amplitudes, respectively.
[d] $R_{free}$ is calculated based on 5% of the reflections randomly excluded from refinement.

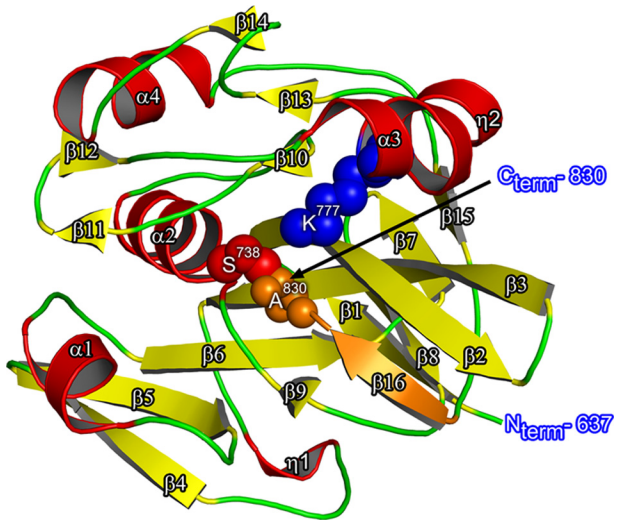


FIGURE 2. **The TV-1 VP4 protease protein fold.** There are 4 $\alpha$-helices (shown in *red*, $\alpha$1–4), 16 $\beta$-strands (shown in *yellow*, $\beta$1–15, and *orange*, $\beta$16), and two $3_{10}$ helix (shown in *red*, $\eta$1,2). The nucleophile, $Ser^{738}$, is shown as *red spheres* and the general base, $Lys^{777}$, as *blue spheres*. The last five residues ($Val^{826}$–$Ala^{830}$) at the C terminus of VP4 are shown in *orange* with the last residue ($Ala^{830}$, shown as *orange spheres*) forming an intramolecular (*cis*) acyl-enzyme with the nucleophile, $Ser^{738}$ (shown as *red spheres*).

This $\beta$-sheet is surrounded by three $\alpha$-helices ($\alpha$2–$\alpha$4) and a $\beta$-hairpin ($\beta$11 and $\beta$12). The nucleophilic $Ser^{738}$ is located at the amino-terminal end of $\alpha$2 and the general base $Lys^{777}$ is part of $\alpha$3. The C-terminal residues (809–830) of TV-1 VP4 wrap around the perimeter of the molecule such that it ultimately arrives with $Ala^{830}$ (the P1 residue for the VP4/VP3 cleavage site and the last residue in the expressed construct) directly adjacent to the $Ser^{738}$ nucleophile. The C-terminal residues pack tightly against the VP4 surface with $\beta$15 hydrogen bond-

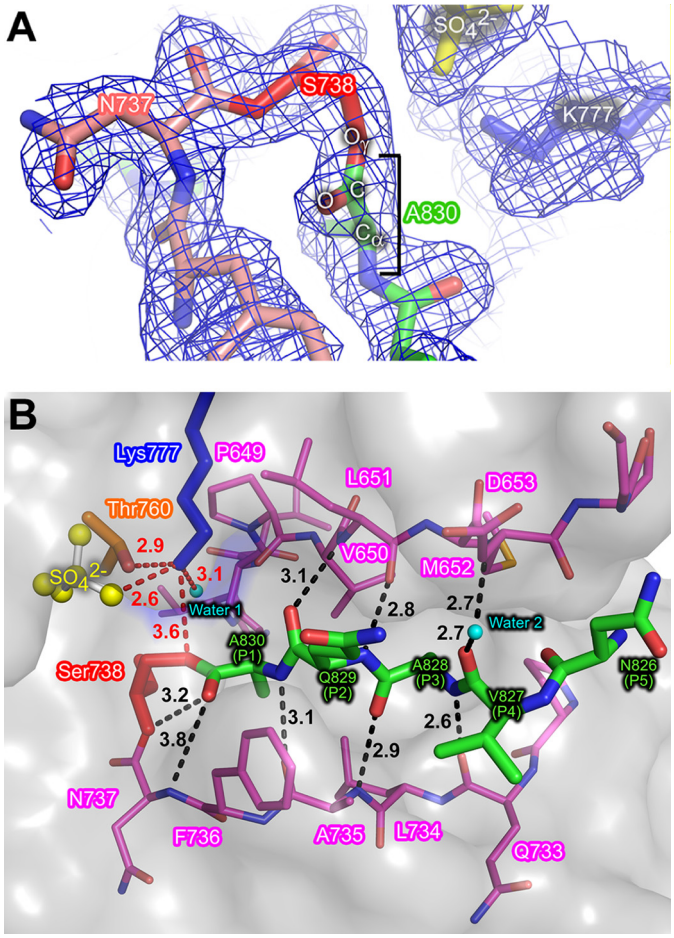


FIGURE 3. **An intramolecular acyl-enzyme intermediate reveals the enzyme-substrate binding interaction in TV-1 VP4 protease.** *A*, an electron density map ($2F_o - F_c$ contoured at 1.2$\sigma$) for atoms in and near the active site. The acyl-enzyme ester linkage is formed between the side chain Oγ of $Ser^{738}$ (shown in *red*) and the carbonyl carbon of the last residue, $Ala^{830}$ (shown in *green*). Residues near the C terminus are shown in *green* and residues leading up to the nucleophile, $Ser^{738}$, are show in *salmon*. The general base, $Lys^{777}$, is shown in *blue* and the sulfate ion adjacent to the active site is shown in *yellow*. *B*, the C terminus of TV-1 VP4 (shown as *green sticks*) binds into its own active site forming an intramolecular (*cis*) acyl-ester linkage between the carbonyl carbon of the last residue, $Ala^{830}$, and the Oγ of $Ser^{738}$ (shown in *red*). The residues that line the substrate binding groove are shown as *magenta sticks* with hydrogen bonds shown as *dashed black lines* buried in a semitransparent molecular surface. Hydrogen bonds formed between the active site residues are shown in *red*.

ing with β7 and β16 hydrogen bonding with β2. The loop that connects β15 and β16 passes under the β-hairpin extension made up of β2 and β3 (Fig. 2).

*An Intramolecular (cis) Acyl-enzyme Intermediate Revealed for the VP4/VP3 Junction*—There is clear continuous electron density from the $Ser^{738}$ Oγ to the carbonyl carbon of $Ala^{830}$ (the C terminus of VP4, P1 residue for the VP4/VP3 cleavage site) (Fig. 3*A*). The electron density at the active site therefore shows that this structure is an intramolecular (*cis*) acyl-enzyme complex or intermediate for the TV-1 VP4/VP3 cleavage site. The covalent linkage between the side chain hydroxyl Oγ of $Ser^{738}$ and the main chain carbonyl of $Ala^{830}$ fits the electron density with a trigonal planer geometry expected for an ester linkage (44, 45).

*Cleavage Site Recognition Groove and Specificity Pockets*—The last three residues of TV-1 VP4 ($Ala^{830}$(P1) to $Ala^{828}$(P3) of the VP4/VP3 cleavage site (46)) are stabilized by anti-parallel β-sheet type hydrogen bonding interactions with residues 651 to 653 of β-strand 2 and parallel β-sheet type hydrogen bonding interactions with residues 733 to 737 of β-strand 9 (Fig. 3*B*). The β-sheet style interaction between the C-terminal residues and the VP4 substrate binding groove are extended with the help of an ordered water molecule (water 2). This ordered water forms hydrogen bonds with the main chain carbonyl oxygen of the P4 residue ($Val^{827}$) and the main chain nitrogen of $Asp^{653}$ (Fig. 3*B*). The shallow and uncharged S3 binding pocket that accommodates $Ala^{828}$ (P3) is formed by residues $Val^{650}$, $Leu^{651}$, $Met^{652}$, $Phe^{663}$, $Pro^{732}$, $Gln^{733}$, and $Leu^{734}$. The P1 residue ($Ala^{830}$) of the VP4/VP3 junction fits into a pocket made up of atoms from $Leu^{648}$, $Pro^{649}$, $Val^{650}$, $Leu^{734}$, $Ala^{735}$, $Phe^{736}$, $Ser^{738}$, and $Trp^{739}$. There are no binding pockets for the P2, P4, or P5 residues as these side chains are pointing away from the substrate binding surface.

*Catalytic Residues*—This wild-type VP4 structure provides an opportunity to observe the atomic position of the catalytic residues during the acyl-enzyme intermediate stage of the reaction cycle for the VP4/VP3 cleavage event, just before the ester bond within the acyl-enzyme intermediate is attacked by a catalytic (deacylating) water. Results from a previous mutagenesis study suggested that TV-1 VP4 utilizes a serine/lysine catalytic dyad (6). The proposed nucleophilic Oγ of $Ser^{738}$ is oriented toward the Nζ of the proposed general base $Lys^{777}$ (Fig. 3*B*). However, the presence of a sulfate bound just adjacent to $Lys^{777}$ has steered the Nζ of $Lys^{777}$ slightly away from the $Ser^{738}$ Oγ. This sulfate sits in a positively charged pocket just adjacent to the general base $Lys^{777}$ and is coordinated by the Nζ of $Lys^{777}$, the Oγ1 of $Thr^{760}$, the Nδ2 of $Asn^{771}$, and the main chain amide nitrogens of $Leu^{772}$ and $Leu^{773}$. The Nζ of the general base $Lys^{777}$ is also coordinated via hydrogen bonds to the Oγ1 of $Thr^{760}$ and a highly ordered and barrier water (water 1) (Fig. 3*B*). The Nζ of $Lys^{777}$ is completely buried (0.0 $Å^2$ of accessible surface area) within the active site of this acyl-enzyme intermediate structure (the average accessible surface area for the 10 other lysine Nζ within TV-1 VP4 is 50.7 $Å^2$). The covalently bound C-terminal carbonyl oxygen of $Ala^{830}$ points into an oxyanion hole assembled from the main chain NH of $Ser^{738}$ and $Asn^{737}$.

## DISCUSSION

### *Comparative Analysis of the Overall Tertiary Structure of VP4 Proteases from TV-1, IPNV, and BSNV*

Due to a low level of sequence identity, the TV-1 VP4 is categorized into a separate protease family (S69) from the other VP4 proteases (S50). Despite this low level of sequence identity, a comparison of the BSNV, IPNV, and TV-1 VP4 structures reveals an overall conservation of tertiary structure (Fig. 4). In quantitative terms, TV-1 VP4 shares 20% sequence identity with BSNV VP4 and has a main chain superposition r.m.s. deviation of 3.5 Å. For IPNV VP4, the sequence identity with TV-1 VP4 is only 12% and the main chain superposition r.m.s. deviation is 7.1 Å.

A distinguishing feature of TV-1 VP4 is the presence of an α-helix between β-strands 5 and 6, a region that forms a loop structure in both IPNV and BSNV (Figs. 2 and 4). Another unique feature of the TV-1 VP4 structure is the presence of two

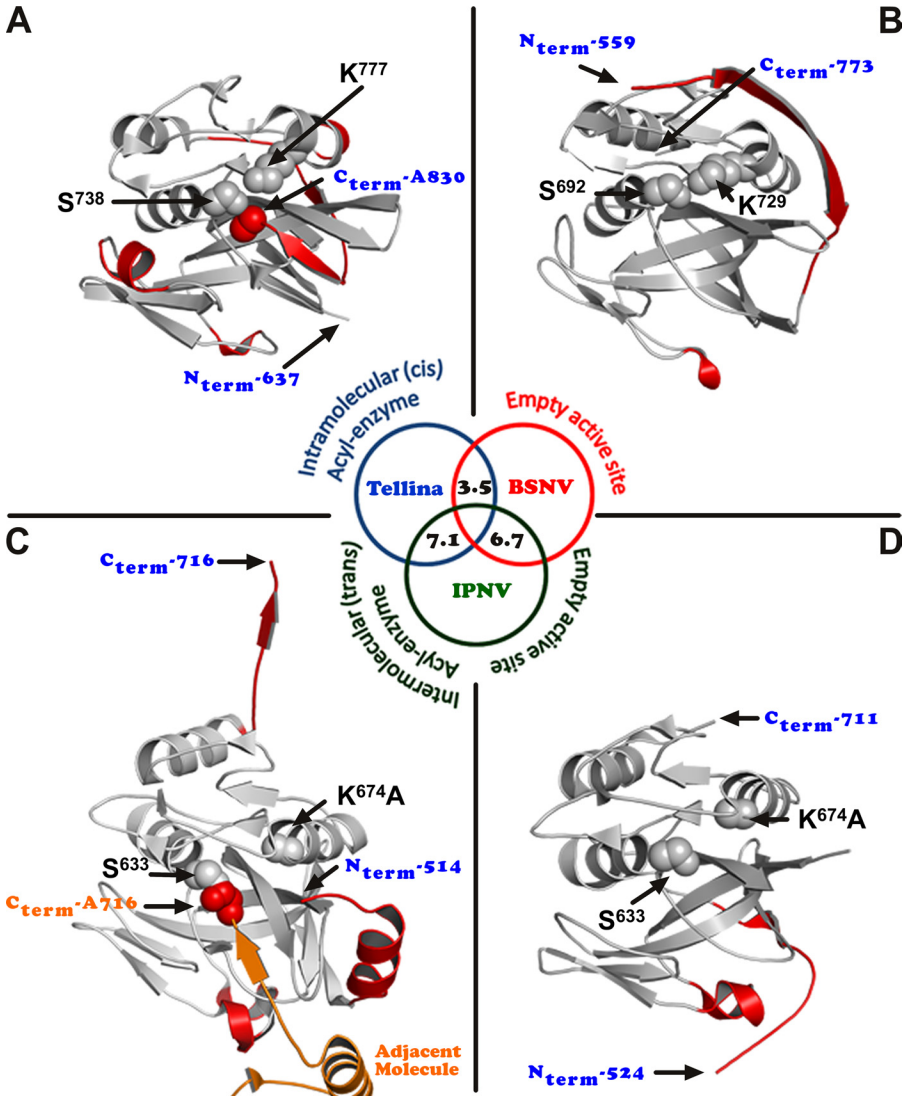


FIGURE 4. **Comparison of VP4 protease structures.** The protein folds of: *A*, TV-1; *B*, BSNV (2GEF) (26); *C* and *D*, IPNV VP4s (2PNL and 2PNM) (27) are shown. All secondary structural elements that are superimposable are shown in *gray*, those that are unique to each protease are shown in *red*. In TV-1 VP4 (*A*), an intramolecular (*cis*) acyl-enzyme complex is formed between the serine nucleophile, $Ser^{738}$ (shown as *gray spheres*), and the last residue, $Ala^{830}$ (shown as *red spheres*). In IPNV VP4 (*C*), the serine nucleophile (shown as *gray spheres*) forming an intermolecular (*trans*) acyl-enzyme complex with the last residue ($Ala^{716}$, shown as *red spheres*) of an adjacent molecule (shown in *orange*) (27). The structures of BSNV VP4 (*B*) and IPNV VP4 (*D*) have empty substrate binding sites. The *numbers* in the *overlapping segment of the circle* denotes the main chain r.m.s. deviation between VP4s.

$\beta$-strands ($\beta$15 and 16) that lead to the C terminus of TV-1 VP4 and form part of the main $\beta$-sheet region (Figs. 2 and 4). In the intermolecular (*trans*) acyl-enzyme intermediate structure of IPNV VP4 (Fig. 4*C*), the C terminus is extended away from the structure and bound to an adjacent active site in a neighboring VP4 molecule. The IPNV VP4 acyl-enzyme structure also has an $\alpha$-helix near the N terminus. This helix packs against and orders the $\beta$-hairpin extension that makes up part of the VP4 substrate binding groove. The shorter construct of IPNV VP4 is lacking much of the above described N-terminal helix region and therefore has a somewhat disordered $\beta$-hairpin region (Fig. 4*D*). A unique characteristic of the BSNV VP4 is the long N-terminal $\beta$-strand that makes an intimate $\beta$-augmentation interaction with another VP4 molecule in the asymmetric unit (Fig. 4*B*); however, this proposed dimerization has yet to be detected in solution.

A sequence alignment of five different VP4s in the *Birnavirus* family reveals only nine invariant residues: $Pro^{649}$, $Val^{650}$, $Pro^{716}$, $Ser^{738}$, $Thr^{760}$, $Gly^{761}$, $Lys^{777}$, $Gly^{786}$, and $Leu^{789}$ (TV-1 numbering, supplemental Fig. S1). $Pro^{649}$, $Val^{650}$, $Ser^{738}$, $Thr^{760}$, $Gly^{761}$, and $Lys^{777}$ are found in or near the active site. $Pro^{716}$ is found adjacent to the S3 binding pocket. Each of the invariant active site residues aligns well when the BSNV, IPNV, and TV-1 VP4 structures are superimposed (supplemental Fig. S2).

### *Comparative Analysis of TV-1 and IPNV VP4 Protease Enzyme/Substrate Interactions*

Crystal structures of both TV-1 VP4 and IPNV VP4 have revealed covalent acyl-enzyme intermediates. The bound substrate in the TV-1 VP4 structure is its own C terminus, the VP4/VP3 junction, forming an intramolecular (*cis*) ester link-

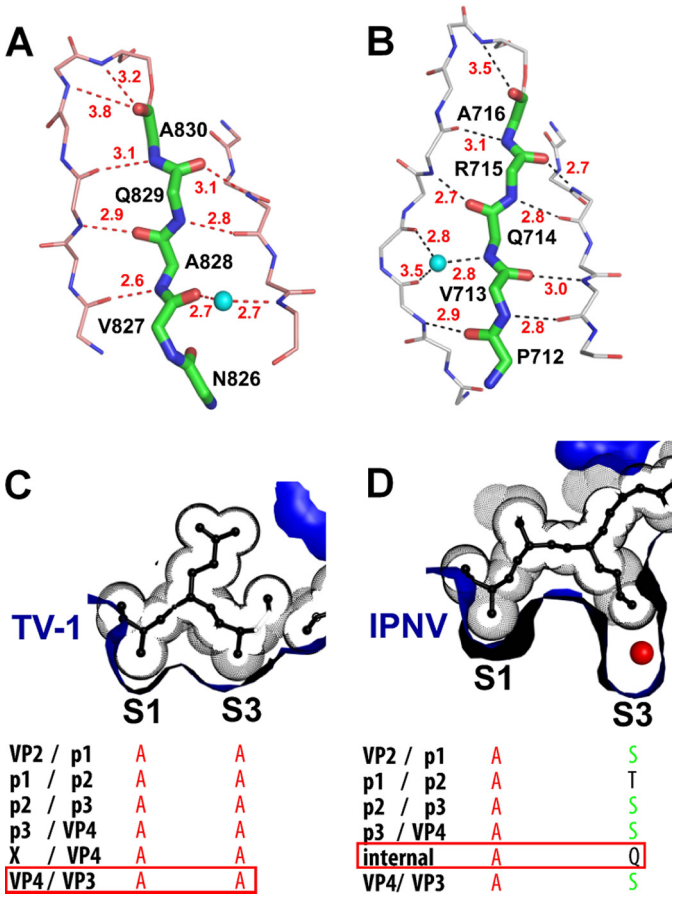


FIGURE 5. **Comparison of TV-1 and IPNV VP4 substrate binding grooves and specificity binding pockets.** *A*, the TV-1 VP4 binding groove is shown with the main chain carbon atoms colored in *salmon*. The main chain carbon atoms for the VP4/VP3 cleavage site (residues P1 to P5) are colored in *green*. An ordered water that is involved in the substrate/binding groove interaction is shown as a *cyan sphere*. The main chain hydrogen bond distances (*dashed lines*) between the enzyme and the substrate are shown. *B*, the IPNV VP4 binding groove shown in a similar fashion to *panel A*, except that the main chain carbon atoms are colored in *white*. *C* and *D* depict a comparison of TV-1 and IPNV substrate specificity binding pockets shown in cross-section. The amino acid residues within the bound substrate are shown as ball-and-stick (*black*) with the van der Waals spheres shown as *dots surrounding* the respective residues. The binding pockets are shown as *blue surfaces*. The P1 and P3 residues for known cleavage sites are shown *below* their binding pockets. Residues that are frequently found at these sites are shown in color. *C*, the VP4/VP3 cleavage site seen in the TV-1 VP4 intramolecular (*cis*) complex presented here is *boxed in red*. *D*, the internal cleavage site seen in the previously solved IPNV VP4 intermolecular (*trans*) complex is *boxed in red* (27).

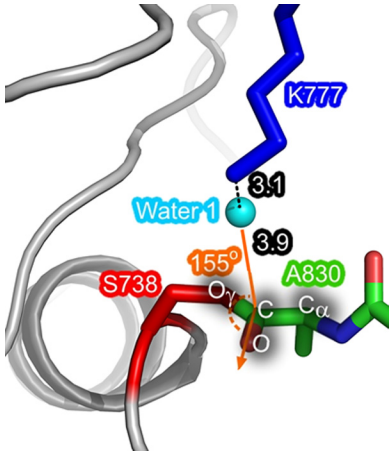


FIGURE 6. **A potential deacylating water in TV-1 VP4 protease.** The serine nucleophile, Ser$^{738}$ (*red*), lysine general base, Lys$^{777}$ (*blue*), and the C terminus (*green*) of TV-1 VP4 are shown. A potential deacylating water (water 1) is shown as a *sphere* (*cyan*). The atoms that define the trigonal planar geometry of the ester linkage between the Ser$^{738}$ Oγ and the Ala$^{830}$ carbonyl carbon are labeled. The angle ($O_{ester} - C_{ester} - O_{water}$) of attack on the ester carbonyl by the potential deacylating water (water 1) is given. The distance from water 1 to the Nζ of the activating lysine general base is given, along with the distance from the water to the ester carbonyl carbon (in Å).

age (Fig. 4*A*), whereas in the previous IPNV VP4 structure the bound substrate is an internal cleavage site near the C terminus forming an intermolecular (*trans*) ester linkage with an adjacent molecule (Fig. 4*C*). A comparison of the substrate binding grooves with bound substrate reveals a similar hydrogen bonding pattern spanning residues P1 to P5 with an average hydrogen bonding distance of 3.0 Å for both TV-1 and IPNV VP4s (Fig. 5, *A* and *B*). Both substrate binding grooves utilize a water molecule within the interaction, but on opposite sides of the substrate (Fig. 5, *A* and *B*).

All of the TV-1 cleavage sites contain an alanine at the P1 and P3 positions (Fig. 5*C*) (6). The same is observed for the P1 position in the IPNV cleavage sites; however, the residues at the P3 position in the IPNV cleavage sites are more variable with serine occurring most frequently and a glutamine at this position in the IPNV VP4 internal cleavage site (Fig. 5*D*) (10). Analysis of the molecular surfaces for the S1 and S3 specificity pockets within the TV-1 and IPNV acyl-enzyme structures reveals structural reasons for the broader specificity seen at the P3 position in the IPNV cleavage sites. The S1 and S3 binding pockets of TV-1 VP4 are shallow and hydrophobic, complementary to the alanine methyl group side chains seen at the P1 and P3 positions (Fig. 5*C*). The S1 binding pocket of IPNV VP4 is also shallow and hydrophobic, but the S3 pocket is deep and hydrophilic allowing it to accommodate a greater variety of residues at the P3 position (Fig. 5*D*). In fact, the bottom of this pocket is lined with a number of water molecules, which may play a role in adjusting the fit for a variety of side chains.

### *The Catalytic Machinery of TV-1 VP4 Protease*

*Oxyanion Hole*—During serine protease-catalyzed cleavage of scissile peptide bonds, a negatively charged tetrahedral intermediate develops, which is thought to be neutralized by a feature of the enzyme called an oxyanion hole. It is usually composed of main chain amide NH groups acting as hydrogen bond donors to the scissile bond oxyanion (47). The TV-1 VP4 acyl-enzyme structure shows that the main chain amide nitrogen of the serine nucleophile (Ser$^{738}$) is within hydrogen bonding distance to the scissile carbonyl oxygen. It is possible that the main chain amide nitrogen of Asn$^{737}$ may also contribute to oxyanion stabilization, but the distance and orientation for this interaction is less than optimal (Fig. 3*B*). It is possible that the TV-1 VP4 oxyanion hole is only completely formed during the transition states, when the tetrahedral oxyanion is present. Additional stabilization may be provided by the positive dipole of α-helix 2, which is oriented toward the oxyanion hole (Figs. 2 and 6).

*Potential Deacylating Water*—Having an acyl-enzyme intermediate complex of TV-1 VP4 protease affords us the opportunity to identify the location of a potential "deacylating" (also called the "catalytic," "hydrolytic," or "nucleophilic") water. For

a water to function as the nucleophile during the deacylation step of a proteolytic 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 (the so-called Bürgi angle of ~107°) (48). An examination of the TV-1 VP4 protease active site region reveals an ordered and buried water molecule (water 1) positioned in approximately the correct position to serve as the deacylating water. It is coordinated by hydrogen bonds to $Ser^{738}$ Oγ, $Lys^{777}$ Nζ, $Pro^{649}$ O, and $Thr^{760}$ Oγ1 (Fig. 6). This is the only active site water seen in the correct position to serve as the deacylating water. It is unlikely that the bound sulfate would be occupying the position for a deacylating water in that it would not be at the correct distance or angle for attack on the ester carbonyl.

*Trapping the Intramolecular (cis) Acyl-enzyme Intermediate in a Wild-type VP4 Protease Active Site*—Although the deacylation step in the serine protease mechanism is often rate-limiting, trapping the acyl-enzyme within a crystal structure has been very challenging. Acyl-enzyme structures have been reported for serine proteases using: (i) short peptide substrates and low pH (44, 49), (ii) small ester substrates using flash cooling (50, 51), (iii) small molecule inhibitors such as β-lactams (52, 53), and (iv) small peptide substrates using pseudo-steady state conditions (45).

We have found that viral proteases such as VP4, which essentially cleave themselves out of the middle of a viral polyprotein, are helpful in studying the acyl-enzyme intermediate stage of serine protease catalysis in that the enzyme itself contains the specificity residues for the C-terminal cleavage-site (Fig. 1). Previously, we used an active site mutation in the IPNV VP4 protease to trap an intermolecular (*trans*) acyl-enzyme intermediate (Fig. 4*C*) (27). Mutating the lysine general base to an alanine and having the internal VP4 cleavage site available at the C terminus of each VP4 molecule along with a high local concentration within the crystallization drop helped to drive the reaction backwards one step to generate the intermolecular (*trans*) acyl-enzyme complex. The stabilization of this complex was likely due to the lack of a general base that is required to activate a deacylating water. In the work presented here we have trapped the intramolecular (*cis*) acyl-enzyme complex for the VP4/VP3 cleavage site in the TV-1 VP4 protease using a wild-type active site. This directly demonstrates the ability of VP4 to cleave in *cis.* The construct was designed such that the C-terminal residue is the P1 residue of the VP4/VP3 cleavage site. It is likely that a sulfate ion available from the optimized crystallization condition and bound adjacent to $Lys^{777}$ has stabilized the protonated positively charged state of the ϵ-amino group of the lysine general base ($Lys^{777}$), thus preventing it from activating the deacylating water. The sulfate binding site could potentially be utilized in inhibitor design.

Superposition of all structures of VP4 shows a significant difference in the region corresponding to the sulfate binding site in TV-1 VP4, particularly residues $Asn^{771}$ to $Leu^{773}$, which form part of a loop that bridges β-strand 12 and α-helix 3 (Figs. 2 and 7). In TV-1 VP4, the side chain of $Leu^{772}$ points away from the enzyme surface. In IPNV and BSNV VP4, the corresponding residues ($Ile^{668}$ and $Val^{723}$, respectively) are pointing into

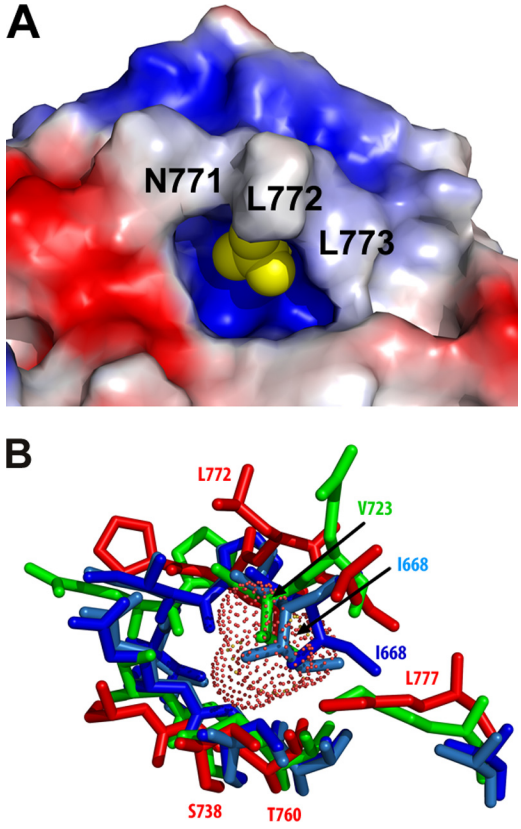


FIGURE 7. **A sulfate binding site in TV-1VP4 protease.** *A*, the molecular surface for the sulfate binding site in TV-1 VP4 is shown (positive electrostatic surface is shown in *blue*, negative in *red*) with the sulfate ion rendered in van der Waals spheres (*yellow*). *B*, a superposition of VP4 structures (IPNV (*blue*, acyl-enzyme; *light blue*, empty active site), BSNV (*green*) and TV-1(*red*)) showing the residues that surround the TV-1 VP4 sulfate binding site. The position for the sulfate is shown as a *dotted surface*.

the would-be sulfate binding site. Interestingly, the optimized crystallization conditions for the IPNV VP4 intermolecular (*trans*) acyl-enzyme structure also had sulfate present (0.4 M $Li_2SO_4$) but no electron density was seen in the active site that would correspond to a sulfate ion. Because the lysine general base was mutated to an alanine in this IPNV VP4 construct, there may not have been the necessary complementary charge for the sulfate to bind. From the collection of VP4 structures available so far, it does not appear that binding of the substrate induces a change in the conformation that would facilitate sulfate binding. IPNV VP4 was solved in the presence and absence of bound substrate yet in both cases the would-be sulfate binding site appears to be occluded by the side chain of $Ile^{668}$ (Fig. 7).

A low pH may have further stabilized the protonated state of the TV-1 VP4 lysine general base thus preventing it from activating a deacylating water. The pH of the TV-1 VP4 crystallization drop is acidic (~pH 5.0). The pH of the crystallization conditions for the previously solved VP4 structures (IPNV and BSNV) were more basic (pH 8.5).

## *Conclusion*

The structure of the TV-1 VP4 protease allows for the identification of the substrate binding pockets (S1 and S3) and a structural rationalization of the cleavage site specificity. The acyl-enzyme complex has allowed us to identify a potential

deacylating water in the VP4 protease mechanism. A subsidiary positively charged pocket adjacent to the active site has also been identified. Despite low sequence identity, a comparison with the IPNV and BSNV VP4 structures reveals that the protein architecture is conserved in this family of proteases. This is the first structure of a *Birnavirus* VP4 in the form of an acyl-enzyme complex with the Ser/Lys active site intact and also the first to show intramolecular (*cis*) cleavage. Taken together, this work provides structural insights for rational antiviral drug design for birnaviruses, using VP4 protease as a target.

*Acknowledgments—We thank the staff at the macromolecular beamline 08ID-1, Canadian Light Source, Saskatoon, Canada, for technical assistance with data collection. The Canadian Light Source is supported by Natural Sciences and Engineering Research Council of Canada, National Research Council, Canadian Institutes of Health Research, and the University of Saskatchewan. We thank Bernard Delmas for providing the initial TV-1 cDNA.*

## REFERENCES

1. Dobos, P., Hill, B. J., Hallett, R., Kells, D. T., Becht, H., and Teninges, D. (1979) *J. Virol.* **32,** 593–605
2. Delmas, B., Kibenge, F. S. B., Leong, J. C., Mundt, E., Vakharia, V. N., and Wu, J. L. (2005) *Virus Taxonomy: The Eighth Report of the International Committee on Taxonomy of Viruses,* Elsevier, Amsterdam
3. Dobos, P. (1995) *Virology* **208,** 19–25
4. Underwood, B. O., Smale, C. J., Brown, F., and Hill, B. J. (1977) *J. Gen. Virol.* **36,** 93–109
5. Espinoza, J. C., Hjalmarsson, A., Everitt, E., and Kuznar, J. (2000) *Arch. Virol.* **145,** 739–748
6. Nobiron, I., Galloux, M., Henry, C., Torhy, C., Boudinot, P., Lejal, N., Da Costa, B., and Delmas, B. (2007) *Virology* **371,** 350–361
7. Da Costa, B., Soignier, S., Chevalier, C., Henry, C., Thory, C., Huet, J. C., and Delmas, B. (2003) *J. Virol.* **77,** 719–725
8. Casañas, A., Navarro, A., Ferrer-Orta, C., González, D., Rodríguez, J. F., and Verdaguer, N. (2008) *Structure* **16,** 29–37
9. Sánchez, A. B., and Rodriguez, J. F. (1999) *Virology* **262,** 190–199
10. Petit, S., Lejal, N., Huet, J. C., and Delmas, B. (2000) *J. Virol.* **74,** 2057–2066
11. Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S., Delmas, B., and Rey, F. A. (2005) *Cell* **120,** 761–772
12. Da Costa, B., Chevalier, C., Henry, C., Huet, J. C., Petit, S., Lepault, J., Boot, H., and Delmas, B. (2002) *J. Virol.* **76,** 2393–2402
13. Hjalmarsson, A., Carlemalm, E., and Everitt, E. (1999) *J. Virol.* **73,** 3484–3490
14. Lejal, N., Da Costa, B., Huet, J. C., and Delmas, B. (2000) *J. Gen. Virol.* **81,** 983–992
15. Chung, H. K., Kordyban, S., Cameron, L., and Dobos, P. (1996) *Virology* **225,** 359–368
16. Ekici, O. D., Paetzel, M., and Dalbey, R. E. (2008) *Protein Sci.* **17,** 2023–2037
17. Paetzel, M., and Dalbey, R. E. (1997) *Trends Biochem. Sci.* **22,** 28–31
18. Botos, I., and Wlodawer, A. (2007) *Curr. Opin. Struct. Biol.* **17,** 683–690
19. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (2002) *J. Biol. Chem.* **277,** 9512–9519
20. Ferentz, A. E., Walker, G. C., and Wagner, G. (2001) *EMBO J.* **20,** 4287–4298
21. Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M., Kim, B., Little, J. W., and Strynadka, N. C. (2001) *Cell* **106,** 585–594
22. Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A., and Gustchina, A. (2004) *J. Biol. Chem.* **279,** 8140–8148
23. Kim, A. C., Oliver, D. C., and Paetzel, M. (2008) *J. Mol. Biol.* **376,** 352–366
24. Hill, B. J. (1976) in *Wildlife Diseases* (Page, L. A., ed) pp. 445–452, Plenum Press, New York
25. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010) *Nucleic Acids Res.* **38,** D227–33
26. Feldman, A. R., Lee, J., Delmas, B., and Paetzel, M. (2006) *J. Mol. Biol.* **358,** 1378–1389
27. Lee, J., Feldman, A. R., Delmas, B., and Paetzel, M. (2007) *J. Biol. Chem.* **282,** 24928–24937
28. Chung, I. Y., and Paetzel, M. (2011) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 157–160
29. Collaborative Computing Project No. 4 (1994) *Acta Crystallogr. Sect. D* **50,** 760–763
30. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1948–1954
31. Grosse-Kunstleve, R. W., and Adams, P. D. (2003) *Acta Crystallogr. D Biol. Crystallogr.* **59,** 1966–1973
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40,** 658–674
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
34. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
35. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2184–2195
36. Painter, J., and Merritt, E. A. (2006) *J. Appl. Crystallogr.* **39,** 109–111
37. Painter, J., and Merritt, E. A. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62,** 439–450
38. Maiti, R., Van Domselaar, G. H., Zhang, H., and Wishart, D. S. (2004) *Nucleic Acids Res.* **32,** W590–4
39. Frishman, D., and Argos, P. (1995) *Proteins* **23,** 566–579
40. Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) *J. Comput. Chem.* **23,** 600–609
41. DeLano, W. L. (2002) *The PyMOL Molecular User's Manual*, DeLano Scientific, San Carlos, CA
42. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23,** 2947–2948
43. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15,** 305–308
44. Katona, G., Wilmouth, R. C., Wright, P. A., Berglund, G. I., Hajdu, J., Neutze, R., and Schofield, C. J. (2002) *J. Biol. Chem.* **277,** 21962–21970
45. Radisky, E. S., Lee, J. M., Lu, C. J., and Koshland, D. E., Jr. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 6835–6840
46. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157–162
47. Ménard, R., and Storer, A. C. (1992) *Biol. Chem. Hoppe Seyler* **373,** 393–400
48. Bürgi, H. B., Dunitz, J. D., and Shefter, E. (1973) *J. Am. Chem. Soc.* **95,** 5065–5067
49. Wilmouth, R. C., Clifton, I. J., Robinson, C. V., Roach, P. L., Aplin, R. T., Westwood, N. J., Hajdu, J., and Schofield, C. J. (1997) *Nat. Struct. Biol.* **4,** 456–462
50. Alber, T., Petsko, G. A., and Tsernoglou, D. (1976) *Nature* **263,** 297–300
51. Ding, X., Rasmussen, B. F., Petsko, G. A., and Ringe, D. (1994) *Biochemistry* **33,** 9285–9293
52. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998) *Nature* **396,** 186–190
53. Wilmouth, R. C., Westwood, N. J., Anderson, K., Brownlee, W., Claridge, T. D., Clifton, I. J., Pritchard, G. J., Aplin, R. T., and Schofield, C. J. (1998) *Biochemistry* **37,** 17506–17513