Crystallographic analysis of birnavirus VP4 proteases

by

Ivy Yeuk Wah Chung

B.Sc., Simon Fraser University, 2005

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2012

Approval

Name: **Ivy Yeuk Wah Chung**

Degree: **Doctor of Philosophy (Molecular Biology and Biochemistry in the Faculty of Science)**

Title of Thesis: **Crystallographic analysis of birnavirus VP4 proteases.**

Examining Committee:

Chair: Jack Chen, Associate Professor

Mark Paetzel
Senior Supervisor
Associate Professor

Michel Leroux

Supervisor
Professor

Peter J. Unrau

Supervisor
Associate Professor

Edgar C. Young
Internal Examiner
Associate Professor
Department of Molecular Biology and Biochemistry

Martin J. Boulanger

External Examiner
Associate Professor, Biochemistry and Microbiology
University of Victoria

Date Defended/Approved: December 12, 2012

Partial Copyright Licence

Abstract

Birnaviruses have a bi-segmented double-stranded RNA genome residing within a single-shelled non-enveloped icosahedral particle. There are economic reasons to study birnaviruses and prevent their propagation as many of the birnaviruses are pathogenic to species that are consumed by humans (salmon, yellowtail fish, chicken, and clam) and some are commercially farmed. Protease VP4 cleaves the polyprotein (NH_2-pVP2-VP4-VP3-COOH) of birnavirus into components required for virion assembly. It utilizes a serine-lysine (Ser/Lys) catalytic dyad that is less characterized than the Ser/His/Asp classical catalytic triad; thus, there are also scientific interests in studying its mechanism. The crystal structure of *Infectious pancreatic necrosis virus* (IPNV) revealed acyl-enzyme complexes suggesting that VP4 proteases could be used to trap different stages of the reaction mechanism. Here, I present the results of the crystallography analysis on VP4 proteases from *Tellina virus* 1 (TV-1), and *Yellowtail ascites virus* (YAV). These structures provided insights on how VP4 proteases interact with the substrates and how the polyprotein is cleaved; thus, will aid in the design of anti-birnavirus compounds.

TV-1 was first isolated from the sand dwelling marine bivalve mollusk *Tellina tenuis* (clam). Manifestations of the disease include a thinner and chalkier shell as well as a pale yellow digestive gland. The structure of TV-1 VP4 was solved to 2.1 Å resolution revealing an intramolecular (*cis*) acyl-enzyme complex which demonstrates how the enzyme recognizes its own carboxy-terminus during the VP4/VP3 cleavage event. To our knowledge, this is the first time that an intramolecular acyl-enzyme has been observed within a protease crystal structure.

YAV infection leads to ascites in yellowtail fish (*Seriola quinqueradiata*), which is popular in sushi. The existence of a previously proposed internal cleavage site within VP4 was confirmed using protein and fluorometric peptide cleavage assays as well as capturing it in two acyl-enzyme structures. The native active site structure (2.5 Å resolution) revealed both the acyl-enzyme and product bound states. The lysine mutant structure (2.3 Å resolution) revealed the acyl-enzyme and empty binding site states of VP4, which allows for the observation of structural changes upon substrate or product binding.

Keywords: Enzyme mechanisms, Protease, Acyl-enzyme, Serine-lysine catalytic dyad mechanism, Birnavirus.

Dedication

This thesis is dedicated to my late father who had given me the opportunity of an education.

Acknowledgements

I would like to thank my supervisor, Dr. Mark Paetzel, for giving me the opportunity to work in his lab and supervisory committee members Dr. Michel R. Leroux, Dr. Peter J. Unrau and Dr. Edgar C. Young for all their constructive feedbacks. My research was great expedited by the help of our lab manager, Deidre de Jong-Wong, who prepares reagents and keeps the lab immaculate. I would also like to express my gratitude to Dr. Anat Feldman and Dr. Jaeyong Lee who laid the foundations for my thesis project and their supervision. It was a great pleasure to work in a lab staffed with creative and supportive individuals.

Table of Contents

Approval ii
Partial Copyright Licence iii
Abstract iv
Dedication vi
Acknowledgements vii
Table of Contents viii
List of Tables xi
List of Figures xii
Glossary xv

1. Motivation for investigating VP4 protease structure, function and mechanism. 1
1.1. VP4 protease 1
1.2. Related proteases. 3
1.3. Classification of VP4 3
1.4. VP4 structures 7
1.4.1. BSNV VP4 7
1.4.2. IPNV VP4 9
1.5. The proposed Ser/Lys dyad mechanism of VP4 protease. 12
1.6. Effect of pK_a on Ser/Lys dyad activity 18
1.7. Why is VP4 protease important? 19
1.8. Rational inhibitor design 20
1.9. Research objectives 21
1.9.1. Using X-ray crystallography to investigate birnavirus polyprotein processing 21
1.9.1.1. Goals and potential problems. 21
1.9.2. Capturing different stages of the catalysis mechanism to help in inhibitor development. 22

2. Birnaviruses 23
2.1. Classification 23
2.2. Virion properties 25
2.3. Viral life cycle 28
2.3.1. Mechanism of entry 28
2.3.2. Viral assembly 28
2.4. Examples of birnaviruses 29
2.4.1. Causative agent of Gumboro disease – *Infectious bursal disease virus* (IBDV) 29
2.4.2. Anoxia sensitivity inducing birnavirus – *Drosophila X virus* (DXV) 31
2.4.3. Causative agent of infectious pancreatic necrosis – *Infectious pancreatic necrosis virus* (IPNV) 33
2.4.4. A birnavirus isolated from bivalve mollusk – *Tellina virus 1* (TV-1) 35
2.4.5. An ascites causing birnavirus – *Yellowtail ascites virus* (YAV) 37
2.4.6. A birnavirus isolated from a cell line originated from an apparently healthy carrier fish – *Blotched snakehead virus* (BSNV) 39
2.5. Protein in birnaviruses 41

2.5.1. Viral protein 1 (VP1) 41
2.5.1.1. Structure 41
2.5.1.2. Self-priming activity 43
2.5.2. Viral protein 2 (VP2) 48
2.5.2.1. Structure 48
2.5.2.2. Capsid particle and pVP2 processing 50
2.5.3. Viral protein 3 (VP3) 55
2.5.4. Viral protein 4 (VP4) 59
2.5.5. Viral protein 5 (VP5) 60
2.5.5.1. IPNV VP5 60
2.5.5.2. IBDV VP5 61

3. Crystal structure of a viral protease intramolecular acyl-enzyme complex - insights into *cis*-cleavage at the VP4/VP3 junction of *Tellina* birnavirus 64
3.1. Introduction 65
3.2. Materials and methods 65
3.2.1. Cloning 65
3.2.2. Protein purification 65
3.2.3. Crystallization 68
3.2.4. Data collection 69
3.2.5. Structure determination and refinement 69
3.2.6. Structural analysis 71
3.2.7. Figure preparation 71
3.3. Results 72
3.3.1. Structure solution 72
3.3.2. Overall protein architecture 74
3.3.3. An intramolecular (*cis*) acyl-enzyme intermediate revealed for the VP4/VP3 junction 76
3.3.4. Cleavage-site recognition groove and specificity pockets 78
3.3.5. Catalytic residues 78
3.4. Discussion 79
3.4.1. Crystallization strategies 79
3.4.2. Comparative analysis of the overall tertiary structure of VP4 proteases from TV-1, IPNV and BSNV 82
3.4.3. Comparative analysis of TV-1 and IPNV enzyme/substrate interactions 87
3.4.4. The catalytic machinery of TV-1 VP4 90
3.4.4.1. Oxyanion hole 90
3.4.4.2. Potential deacylating water 91
3.4.5. Trapping the intramolecular (*cis*) acyl-enzyme intermediate in a wild-type VP4 protease active site 92
3.5. Conclusion 95

4. Crystal structures of *Yellowtail ascites virus* VP4 protease: trapping an internal cleavage site *trans* acyl-enzyme complex in a native Ser/Lys dyad active site 96
4.1. Introduction 96

4.2. Materials and methods 96
4.2.1. YAV VP4 constructs 96
4.2.2. Protein expression and purification 99
4.2.3. YAV VP4 full-length self-cleavage assay 102
4.2.4. Fluorometric peptide cleavage assay 102
4.2.5. Crystallization 102
4.2.6. Data collection 103
4.2.7. Structure solution and refinement 104
4.2.8. Structural analysis 106
4.3. Results 106
4.3.1. Self-cleavage at the YAV VP4 internal cleavage site 106
4.3.2. Self-cleavage at the YAV VP4 internal cleavage site promotes crystallization 107
4.3.3. Overall architecture and active site 109
4.3.4. Different C-terminal conformations 111
4.3.5. *Trans* acyl-enzyme and enzyme-product complexes in a native active site 111
4.3.5.1. Hydrogen bonding network. 113
4.3.5.2. Substrate specificity pockets. 115
4.3.6. Substrate binding groove: empty vs. bound 117
4.3.7. Aqueous channel leads to active site 117
4.4. Discussion 118
4.4.1. Ser^{633} is the YAV VP4 nucleophile, but could Thr^{655} Oγ function as a nucleophile in the absence of Ser^{633} Oγ ? 118
4.4.2. Interpreting mutagenesis results for non-catalytic residues 119
4.4.3. *Cis* vs. *trans* cleavage at the YAV VP4 internal cleavage site 121
4.5. Conclusion 122

5. Concluding remarks 124
5.1. Contribution of thesis project 124
5.1.1. Capturing interactions at the polyprotein cleavage sites. 124
5.1.2. Capturing the intermediate steps in reaction mechanism of VP4 proteases 125
5.1.3. Capture the first intramolecular (*cis*) acyl-enzyme complex of a viral protease in PDB. 126
5.2. Future studies 126
5.2.1. The importance of residues downstream of the scissile bond. 126
5.2.2. Involvement of VP4 in pVP2 processing. 127
5.2.3. Enzyme-substrate interactions at the pVP2/VP4 junction. 127
5.2.4. Capturing other intermediate steps in the reaction mechanism. 128

References 129

Appendices 142
Appendix A. Standard curve of HiPrep 26/60 Sephacryl S-100 high resolution. 143

List of Tables

Table 1-1. VP4 sequence and structural information. 5

Table 1-2. Identity and similarity between different VP4 sequences. 5

Table 1-3. Standard pK_a values for functional groups involved in the VP4 mechanism. 19

Table 2-1. Diseases associated with birnaviruses. 24

Table 2-2. Genome and capsid size of birnaviruses. 27

Table 2-3. Table of features for VP1. 44

Table 2-4. Sequence identity and similarity between different VP1s. 45

Table 2-5. Table of features for pVP2. 52

Table 2-6. Sequence identity and similarity between different pVP2s. 53

Table 2-7. Table of features for VP3. 57

Table 2-8. Sequence identity and similarity between different VP3s. 58

Table 2-9. Table of features for VP5. 62

Table 2-10. Sequence identity and similarity between different VP5s. 62

Table 3-1. Data collection, phasing and refinement statistics for TV-1 VP4 crystal. 70

Table 4-1. Data collection and refinement statistics for YAV VP4 crystals. 105

List of Figures

Figure 1-1. Gene products of *Infectious bursal disease virus* (IBDV) segment A and B. 2

Figure 1-2. Radial phylogenetic tree of VP4 proteases. 4

Figure 1-3. VP4 sequence alignment. 6

Figure 1-4. Structure of BSNV VP4. 8

Figure 1-5. The intermolecular acyl-enzyme complex of IPNV VP4 captured at the internal cleavage site. 10

Figure 1-6. The structure of IPNV VP4 with an empty binding groove. 11

Figure 1-7. The generalized mechanism of Ser/Lys proteases. 14

Figure 1-8. The Schechter and Berger nomenclature of peptidase specificity. 15

Figure 1-9. Free energy diagram for a two-step covalent catalysis mechanism. 16

Figure 1-10. List of polyprotein cleavage sites in birnaviruses. 17

Figure 2-1. The genomic arrangement of segment A in birnaviruses. 26

Figure 2-2. The genomic arrangement of segment A and a list of polyprotein cleavage sites in IBDV. 30

Figure 2-3. The genomic arrangement of segment A and a list of polyprotein cleavage sites in DXV. 32

Figure 2-4. The genomic arrangement of segment A and a list of polyprotein cleavage sites in IPNV. 34

Figure 2-5. The genomic arrangement of segment A and a list of polyprotein cleavage sites in TV-1. 36

Figure 2-6. The genomic arrangement of segment A and a list of polyprotein cleavage sites in YAV. 38

Figure 2-7. The genomic arrangement of segment A and a list of polyprotein cleavage sites in BSNV. 40

Figure 2-8. Structure of IBDV VP1. 42

Figure 2-9. Protein sequence alignment of VP1. 46

Figure 2-10. Structure of IBDV VP2 monomer. 49

Figure 2-11. Structure of T13 IBDV viral particle formed by trimers of VP2. 51

Figure 2-12. Protein sequence alignment of pVP2. 54

Figure 2-13. Structure of VP3 in IBDV. 56

Figure 2-14. Protein sequence alignment of VP3. 59

Figure 2-15. Protein sequence alignment of VP5. 63

Figure 3-1. Purification scheme. 67

Figure 3-2. Crystals of TV-1 VP4. 69

Figure 3-3. TV-1 VP4 protease cleavage sites 73

Figure 3-4. The TV-1 VP4 protease protein fold. 75

Figure 3-5. An intramolecular acyl-enzyme intermediate reveals the enzyme-substrate binding interaction in TV-1 VP4. 77

Figure 3-6. Size exclusion chromatography of VP4 before and after cation exchange. 81

Figure 3-7. Comparison of VP4 structures. 83

Figure 3-8. A sequence alignment of birnavirus VP4 proteases. 85

Figure 3-9. Conservation of the active site region in VP4 proteases. 86

Figure 3-10. Comparison of TV-1 and IPNV VP4 substrate binding groove and specificity binding pockets. 88

Figure 3-11. A potential deacylating water in TV-1 VP4 protease. 91

Figure 3-12. Sulphate binding site in TV-1VP4. 94

Figure 4-1. A schematic of YAV segment A polyprotein processing by VP4 protease and activity assays for truncated YAV VP4 (509-716). 98

Figure 4-2. Purification scheme of YAV VP4. 101

Figure 4-3. Crystals of YAV VP4. 103

Figure 4-4. Intermolecular (*trans*) acyl-enzyme complex, product complex and empty active site. 108

Figure 4-5. The YAV VP4 protein fold. 110

Figure 4-6. Binding groove and active site interactions seen in *trans* acyl-enzyme complexes and *trans* enzyme-product complexes within a native active site of YAV VP4 protease. 112

Figure 4-7. YAV VP4 protease structure with a mutant active site (K674A) reveals a comparison of acyl-enzyme complex and empty substrate binding groove. 114

Figure 4-8. Changes in substrate specificity binding pockets between bound and unbound states. 116

Figure 4-9. A channel leads to the active site of YAV VP4 protease. 118

Figure 4-10. Mapping of non-catalytic residue site-directed mutants that affect catalysis. 121

Glossary

Å	Symbol for Ångströms which equals to 10^{-10}m.
Acyl-enzyme intermediate	An intermediate formed during enzyme catalysis in which the active site residue is acylated to the substrate.
Asymmetric unit	Define as the smallest unit that can be used to generate a unit cell using symmetry operators.
B factor	Also known as the "temperature-factor" which is an indicator of the degree to which the electron density is spread out due to the thermal motion of the atoms in the lattice. A higher B-factor can indicate higher mobility or error in the model.
β-augmentation	A type of β-sheet/strand interaction in which the β-strands involved are contributed by two different molecules. The folding of these molecules are independent of the β-augmentation formation.
Birnavirus	A non-enveloped virus with a bi-segmented double-stranded RNA genome encapsulates in an icosahedral capsid.

Birnavirus polyprotein	A protein encoded by the larger open reading frame of segment A in the birnavirus genome and contains proteins pVP2, VP4, and VP3. In TV-1 and BSNV, it also encodes peptide X between pVP2 and VP4. The polyprotein is about 1000-residue-long.
***Blotched snakehead virus* (BSNV)**	A birnavirus that infects blotched snakehead fish (*Channa lucius*) and is the type species of blosnavirus.
Bürgi angle	The angle of attack (~107^{o}) on a carbonyl by a nucleophile as determined by Bürgi and Dunitz using crystallographic data.
Completeness	The total number of reflections measured, expressed as a percentage of the total number of reflections expected for a given resolution.
Crystallographic refinement	An iterative process whereby the R-factors (values that indicates the agreement between the molecular model and the crystallographic data) were improved by means of adjusting the atomic coordinates of the molecular model to match the crystallographic data.
***Drosophila X virus* (DXV)**	An *entomobirnavirus* that causes sensitivity to anoxia in flies.

Figure of merit (FOM)	In the context of SAD phasing in crystallography, it is an estimate of phase quality which ranges from 0 to 1. An FOM above 0.45 is very good for SAD phasing whereas a FOM of 0.25-0.45 is acceptable.
***Infectious bursal disease virus* (IBDV)**	An *avibirnavirus* that destroys lymphoid cells in the bursa of Fabricius in chickens.
***Infectious pancreatic necrosis virus* (IPNV)**	An *aquabirnavirus* that causes disease in various *salmonid* fish.
Matthews coefficient	Also known as specific volume (Vm) and defined by the equation: $\frac{Volume\ of\ unit\ cell}{Molecular\ Weight\ of\ the\ macromolecule * Z * X}$ Where Z is the number of asymmetric units in the unit cells and X is the number of molecules in the asymmetric unit. The unit for Matthews coefficient is $Å^3$/dalton.
Molecular replacement	A method to obtaining phase information using phase information provided by a model from a homologous molecule.
PEG	The abbreviation for polyethylene glycol which is a precipitant that is commonly used in crystallization trials.

Peptide bond	A chemical bond formed between the carboxyl group and amino group of adjacent amino acids which links the individual amino acids into polypeptide/protein.
Redundancy	It is defined as the average number of independently measured reflection in a crystallographic dataset and is calculated by the equation: $Redundancy = \frac{Number\ of\ measured\ reflections}{Number\ of\ unique\ reflections}$
***re*-face**	In the case of protease cleavage, it refers to the face of the peptide with carbonyl oxygen (O), amide Nitrogen (N) and C_α going in a clockwise direction.
R_{merge}	Measures the agreement among multiple measurements of the same reflections and is defined by the equation: $R_{merge} = \sum_{hkl}\sum_i \lvert I_i(hkl) - \langle I(hkl)\rangle \rvert \,/\, \sum_{hkl}\sum_i I_i\,(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl)\rangle$ is the mean intensity of that reflection.
RMSD	The acronym for root mean square deviation. In crystallography, it is a measure of how good the final crystallographic model fits the expected bond lengths and bond angles.

SAD	The abbreviation for single-wavelength anomalous diffraction, a method to obtain initial phases by measuring diffraction data at a single wavelength near the absorption edge of a heavy-atom such as selenium. In the absence of anomalous scattering, the intensity of the Friedel pair is equal but have opposite phases. In the presence of anomalous scattering, the intensity is no longer equal nor having opposite phases.
***si*- face**	In the case of protease cleavage, it refers to the face of the peptide with carbonyl oxygen (O), amide Nitrogen (N) and C_α going in a counter-clockwise direction.
Solvent content	Defined as the fraction of the unit cell taken up by the solvent and is calculated by the equation: 1.23/Vm (specific volume).
Space group	Indicates the symmetry of the unit cell in a crystal. The capital letter in the space group indicates the lattice type and the other symbols represents the symmetry operations that can be used on the asymmetric unit to generate the complete content of the unit cell.
***Tellina virus 1* (TV-1)**	A birnavirus isolated from the sand dwelling marine bivalve mollusk *Tellina tenuis* (clam).

Unit cell	The smallest repeating unit that can be use to generate a crystal by using translation.
Viral protein 2 (VP2)	A capsid protein encodes by segment A of the birnavirus genome and translates as part of the polyprotein.
Viral protein 3 (VP3)	A protein encodes by segment A of the birnavirus genome and translates as part of the polyprotein. It is known to associate with the capsid protein VP2, the double-stranded RNA genome and the self-encoded RNA-dependent RNA polymerase, VP1.
Viral protein 4 (VP4)	A protease encodes by segment A of the birnavirus genome and translates as part of the polyprotein. It is responsible for cleaving the polyprotein into its constituent components
X-ray	A type of electromagnetic radiation with wavelength between 0.01 - 10 nm.
***Yellowtail ascites virus* (YAV)**	An *aquabirnavirus*, YAV and YAV-like viruses are members of marine birnavirus (MABV) which infect both fish and shellfish. YAV infection leads to ascites in yellowtail fish (*Seriola quinqueradiata*), which is popular in sushi, thus causing significant losses to the fish-farming industry.

1. Motivation for investigating VP4 protease structure, function and mechanism.

1.1. VP4 protease

Viral protease VP4 is a Ser/Lys dyad protease that is encoded within the birnavirus genome. Birnavirus encodes a polyprotein, a single polypeptide, on which the capsid precursor protein (pVP2), VP4 and inner capsid protein (VP3) are found (**Figure 1-1**) [1,2,3]. Protease VP4 is responsible for the cleavage of the polyprotein and its processing is required for the assembly of the virus.

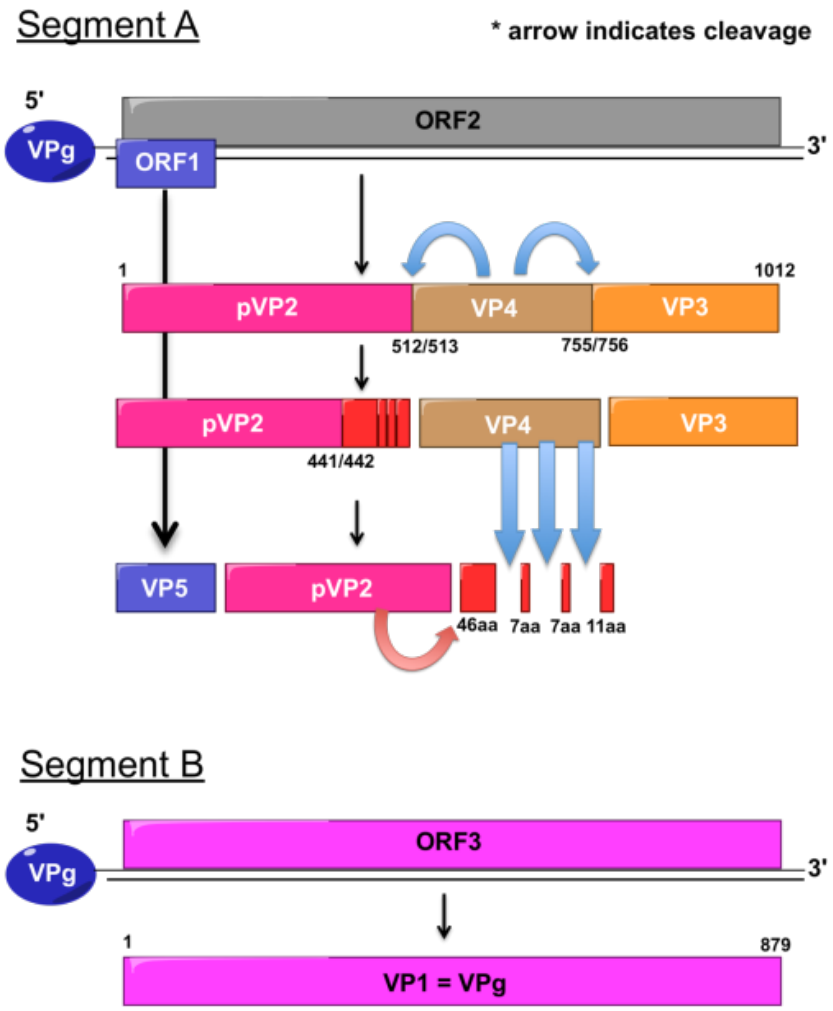


Figure 1-1. Gene products of Infectious bursal disease virus (IBDV) segment A and B.

There are two open reading frames (ORFs) in segment A of the IBDV genome. The larger ORF (ORF2) encodes a polyprotein that contains the capsid precursor protein (pVP2), viral protease (VP4) and inner capsid protein (VP3). Protease VP4 cleaves the polyprotein to yield: pVP2, VP4, and VP3 (blue arrows). The smaller ORF codes for VP5 which is shown to induce apoptosis in DF-1 cells. The C-terminal end of pVP2 is further processed by VP4 to yield additional peptides (blue arrows) [1]. Cleavage between residues 441 and 442 of the polyprotein releases the mature capsid protein and this cleavage is mediated by the autoproteolytic activity of pVP2[2]. Segment B has only one ORF which codes for an RNA-dependent RNA polymerase (VP1) that can exist as a free polymerase or link to the RNA genome (VPg) [3]. The number, type and order of cleavage events may differ in the various species of birnavirus, although the overall maturation process is likely to be similar to that seen in IBDV.

1.2. Related proteases.

Examples of Ser/Lys dyad proteases include type I signal peptide[4], signal peptide peptidase (SppA) [5], UmuD family of peptidases (UmuD[6], LexA [7], lambda repressor[8]) and Lon protease[9]. Type I signal peptidase and SppA are involved in signal peptide processing on proteins exported from the cytoplasm. Proteases UmuD and LexA participate in the bacterial SOS response triggers by DNA damage[6]. The lambda repressor prevents the phage from entering the lytic cycle by repressing promoters associated with lytic functions [10-12]. Lon protease participates in protein quality control and degradation of misfolded proteins [13].

1.3. Classification of VP4

According to the MEROPS protease database, the VP4s from IPNV, *infectious bursal disease virus* (IBDV), *Blotched snakehead virus* (BSNV), *Drosophila X virus* (DXV) and *Tellina virus 1* (TV-1) belong to the evolutionary clan SJ [14]. 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 other VP4s have been assigned to (**Figure 1-2**). A list of VP4 features, PDBs, sequence comparison, sequence alignment can be found in **Tables 1-1**, **1-2** and **Figure 1-3**.

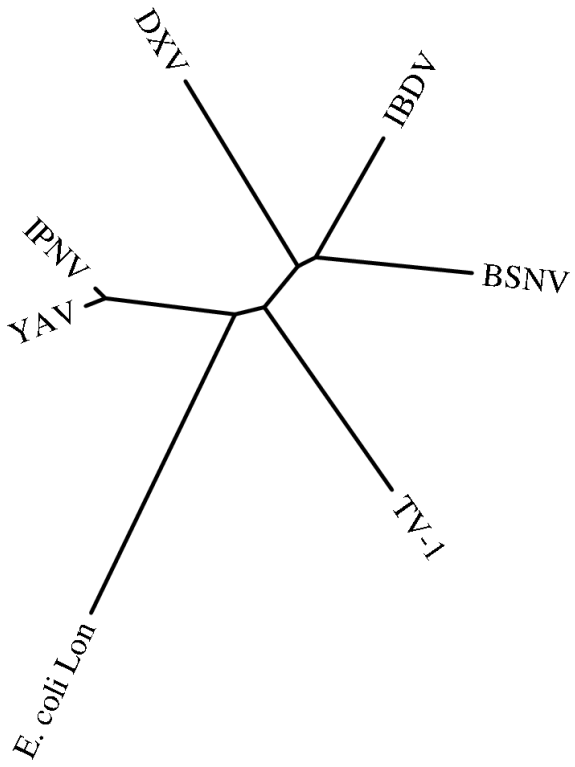


Figure 1-2. Radial phylogenetic tree of VP4 proteases.

The radial phylogenetic tree was generated using the Phylogeny.fr server following the "One Click" workflow (program MUSCLE for alignment, program Gblocks for curation, program PhyML for phylogeny, and program TreeDyn for tree rendering) [15-21]. The UniProt accession number for the VP4 sequences are listed in **Table1-1** and the UniProt accession number of *E.coli* Lon protease is P0A9M0.

Virus name	UniProt accession number	Sequence length (AA.)	Size (KDa)	pI	PDB ID	Reference
BSNV	Q8AZM0	234	25.2	5.64	2GEF [24]	Da Costa *et al.* [25]
IPNV	Q703G9	226	24.0	4.75	1)2PNM [26] 2)2PNL [26]	Galloux *et al.* [27]
YAV	P89521	226	24.1	4.86	1) 3R0B 2) 4HHC	Suzuki *et al*. [28]
IBDV	P25220	243	26.5	6.66		Bayliss *et al.* [29]
DXV	Q96724	223	24.5	5.86		Chung *et al.* [30]
TV-1	Q2PBR5	212	22.5	9.83	3P06	Nobiron *et al.* [31]

Table 1-1. VP4 sequence and structural information.

The sequences were obtained from UniProt. Values for sequence length, size and pI were calculated using ProtParam [22, 23].

Identity (%) / Similarity (%)	BSNV	IPNV	YAV	IBDV	DXV
IPNV	17.7/30.0				
YAV	17.3/29.2	81.0/88.1			
IBDV	20.5/33.2	18.1/27.8	18.1/28.6		
DXV	19.6/31.5	15.9/29.3	14.2/29.3	21.3/27.7	
TV-1	19.5/32.2	11.8/27.6	10.5/23.7	15.8/27.9	18.9/30.3

Table 1-2. Identity and similarity between different VP4 sequences.

The protein sequence alignment was generated using ClustalW [32] and the sequence identities and similarities were calculated using the sequences identities and similarities web server (URL:http://imed.med.ucm.es/Tools/sias.html) using the following equation:

$$\%\ \text{identity or similarity} = 100\left(\frac{\text{Identical positions}}{\text{Length of the alignment}}\right)$$

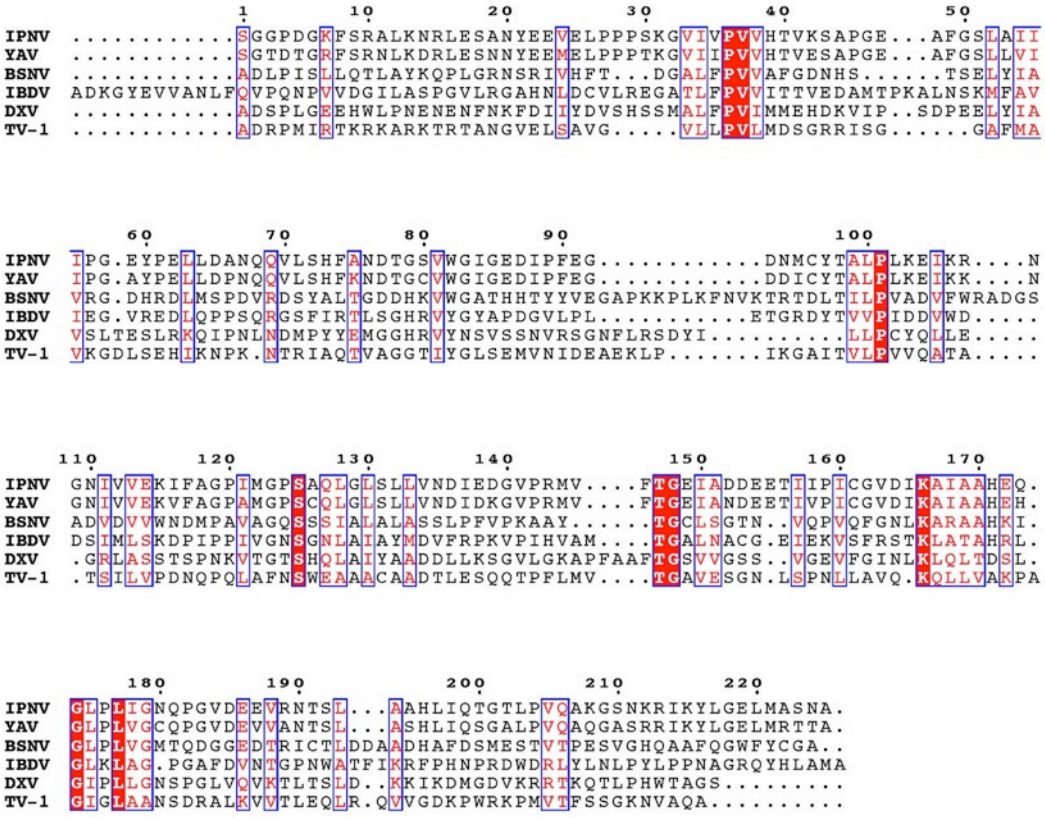


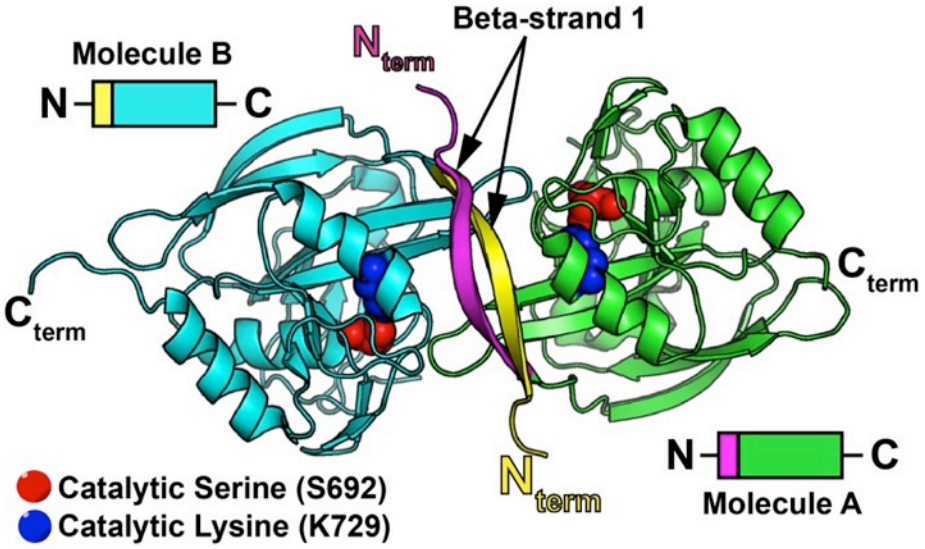
Figure 1-3. VP4 sequence alignment.

The VP4 sequence alignment was performed using the program ClustalW[32] and the graphical output was generated using the program ESPript[33]. Residues that are conserved are shown as white text highlighted in red and semi-conserved residues are shown as red text in blue boxes.

1.4. VP4 structures

1.4.1. BSNV VP4

In 2006, Feldman *et al.* solved the structure of BSNV VP4 with a C-terminal truncation. The structure revealed an empty active site (**Figure 1-4**)[24]. The VP4 protein fold consists of 13 β-strands, one β -hairpin, 3 α-helices and one 3_{10} helix. The structure can be divided into two domains with domain I composes mainly of β-strands and domain II composes of α-helixes with intervening β-strands. The serine nucleophile Oγ (Ser^{692}) is within hydrogen bonding distance to the lysine general base Nζ (Lys^{729}) (**Figure 1-4**). The ε-amino group of the lysine general base is also coordinated by the Oγ of Thr^{712}. Superposition of active sites from other Ser/Lys proteases (e.g. signal peptidase, Lon protease and LexA from *E. coli*) revealed that they also have a hydroxyl from a serine or threonine within hydrogen bonding distance to the lysine general base[24]. Dimerization was observed in the crystal structure of BSNV VP4 with the N-terminal β-strands (β1, **Figure 1-4**, magenta and yellow) from adjacent molecules hydrogen bonding with each other in an antiparallel β-augmentation fashion. No dimers were detected in solution for the construct used in crystallization.



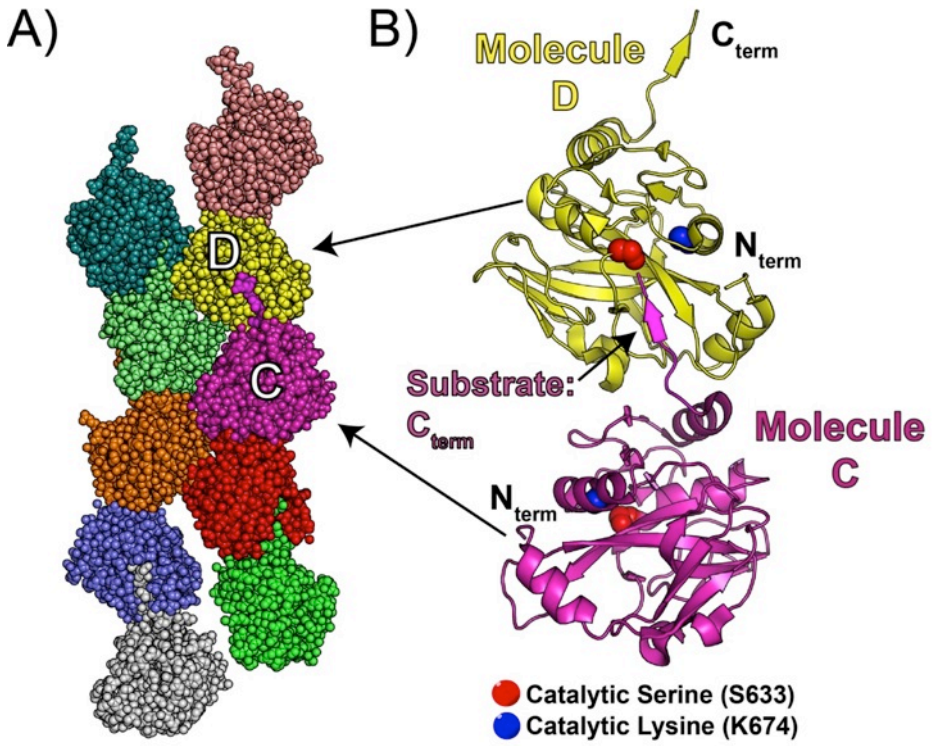
PDB Code	Resolution(Å)	UniProt accession Number	Reference
2GEF	2.2	Q8AZM0[25]	Feldman *et al.* [24]

Figure 1-4. Structure of BSNV VP4.

Each asymmetric unit is composed of two molecules of VP4 which are shown in yellow/cyan and magenta/green[24]. The N-terminal β-strand (β1, Magenta and yellow) from each molecule hydrogen bond with each other in an antiparallel fashion to form a homodimer. The serine nucleophile (Ser^{692}) is shown as red spheres and the lysine general base (Lys^{729}) is shown as blue spheres. The figure was prepared using Pymol[35].

1.4.2. IPNV VP4

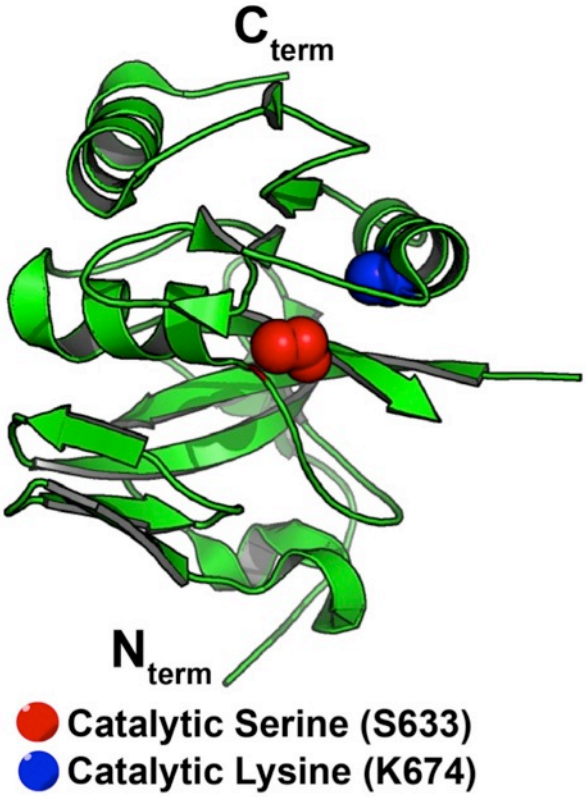
In 2007, Lee *et al*. published two crystal structures of IPNV VP4 with one forming an intermolecular acyl-enzyme complex (**Figure 1-5**) and the other as an empty enzyme (**Figure 1-6**) [26]. The serine nucleophile (Ser^{633}) is shown as red spheres and the lysine general base (Lys^{674}) is shown as blue spheres in **Figures 1-5** and **1-6**. The lysine general base was mutated to an alanine. The acyl-enzyme complex structure was captured at the internal cleavage site with the C-terminus (**Figure 1-5B**, magenta) of VP4 bound in the active site of an adjacent molecule forming a long chain of molecules in the crystal. There are 10 molecules in the asymmetric unit in this crystal form (**Figure 1-5A**). The overall architecture and fold of IPNV VP4 is similar to that of BSNV. A major difference is found at the N-terminus where an α-helix is found in place of a β-strand. The presence of the substrate, the C-terminus of an adjacent VP4, allowed for the identification of the S1, S3, S5 and S6 substrate binding pockets. The substrate is stabilized by anti-parallel hydrogen bonding with the top of the binding groove and by parallel hydrogen bonding with the bottom of the binding groove. In the oxyanion hole, the main chain amide nitrogen of the nucleophile (Ser^{633}) is within hydrogen bonding distance to the carbonyl carbon of residue P1. Alpha helix 2 locates immediately after Ser^{633}, its partial positive dipole may also contribute to oxyanion stabilization.



PDB Code	Resolution(Å)	UniProt accession number	Reference
2PNL	2.2	Q703G9[27]	Lee *et al.* [26]

Figure 1-5. The intermolecular acyl-enzyme complex of IPNV VP4 captured at the internal cleavage site.

A) The C-terminus of each VP4 molecules binds into the active site of an adjacent molecule forming a long chain of VP4 molecules in the crystal[26]. There are 10 molecules in the ASU and each is shown in a different colour. **B)** Two molecules from the ASU are shown in yellow (Molecule D) and magenta (Molecule C). The C-terminus of molecule C serves as the substrate for molecule D (yellow). The serine nucleophile (Ser^{633}) is shown as red spheres and the lysine general base (Lys^{674}), mutated to alanine, is shown as blue spheres. The figure was prepared using Pymol[35].



PDB Code	Resolution(Å)	UniProt accession number	Reference
2PNM	2.3	Q703G9 [27]	Lee *et al.* [26]

Figure 1-6. The structure of IPNV VP4 with an empty binding groove.

The construct used for generating this crystal consisted of residue 524-716 of the polyprotein which is shorter than that used for the acyl-enzyme structure[26]. The serine nucleophile (Ser^{633}) is shown as red spheres and the lysine general base (Lys^{674}), mutated to alanine is shown as blue spheres. The figure was prepared using Pymol[35]

1.5. The proposed Ser/Lys dyad mechanism of VP4 protease.

Protease VP4 utilizes a covalent two-step catalysis mechanism typical of hydrolase that use a serine nucleophile. The classic example of this class of enzyme is chymotrypsin. The standard accepted mechanism utilized by these enzymes included two tetrahedral transition states and one acyl-enzyme intermediate (**Figure 1-7, 1-9**). The first evidence for the two-step mechanism came from kinetic studies. Using p-nitrophenyl acetate as a substrate, Hartley *et al.* demonstrated that chymotrypsin cleaves its substrate in two steps [36]. The first step involves the quick release of the p-nitrophenyl group from the substrate leading to a yellow colour change. The second step is much slower and required a water to react with the enzyme to release the covalently bound acetate group and regenerate the enzyme for another reaction cycle. The two-step reaction mechanism can be thought of as occurring in two stages with transition state steps in between each stage. The first stage is the acylation of the enzyme to form a covalent acyl-enzyme intermediate (**Figure 1-7**,Steps 1-4). Stage two involves the deacylation of the acyl-enzyme to release the product and free enzyme (**Figure 1-7**,Steps 5-6). The following reaction mechanism for serine protease was adapted from the mechanism of chymotrypsin and the proposed mechanism of signal peptidase I, and is also consistent with the structural and biochemical evidence so far collected for VP4 protease [37, 38]. Stage one begins with the binding of the substrate to the active site and substrate binding groove (**Figure 1-7**,Step 1). The general base (blue) abstracts a proton from the $O\gamma$ serine nucleophile (red) thereby activating it for a nucleophilic attack on the P1 carbonyl carbon of the bound substrate (**Figure 1-7**,Step 2). Residue P1 refers to the residue immediately preceding the scissile bond (pink) which is the bond cleaved by the protease [39](**Figure 1-7**). This nomenclature was first used by Schechter and Berger in 1967 to describe the peptidase specificity of papain (**Figure 1-8**). The nucleophile attacks from the *si*-face of the scissile bond in VP4 protease whereas in chymotrypsin it attacks from the *re*-face [40]. The protonated general base is stabilized by a negative charge on an adjacent aspartic acid in chymotrypsin. Unlike the imidazole group of histidine in chymotrypsin, a counter charge is not required for the ε-amino group of lysine to carry out effective proton transfer [41]. The nucleophilic attack leads to the

formation of a tetrahedral transition state oxyanion on the P1 residue, a state with high free energy (**Figure 1-7**,Step 3; **Figure 1-9**). The hydrogen bond donors in the oxyanion hole (green) form hydrogen bonds with the oxyanion and stabilize it. In both chymotrypsin and VP4 proteases, backbone amides form part of the oxyanion hole. Next, a proton is transferred from the general base to the P' end of the substrate. Residue P' refers to the residues after the scissile bond (**Figure 1-8**). The peptide bond between the P1 and P1' residues is then cleaved and the P' side of the product is released. The P end of the substrate remained covalently attached to the serine nucleophile $O\gamma$ and forms an acyl-enzyme complex that is at a local free energy minima (**Figure 1-7**,Step 4; **Figure 1-9**). Stage two of the mechanism begins with the deprotonation of a catalytic (or deacylating, or nucleophilic) water to become OH^-. It then attacks the ester carbonyl carbon on the P1 residue resulting in the formation of a second tetrahedral transition state oxyanion (**Figure 1-7**,Step 5, **Figure 1-9**). Finally, the protonated general base donates a proton back to the $O\gamma$ of the serine nucleophile and the P end of the product is released (**Figure 1-7**,Step 6). A list of polyprotein cleavage sites in birnaviruses is shown in **Figure 1-10**.

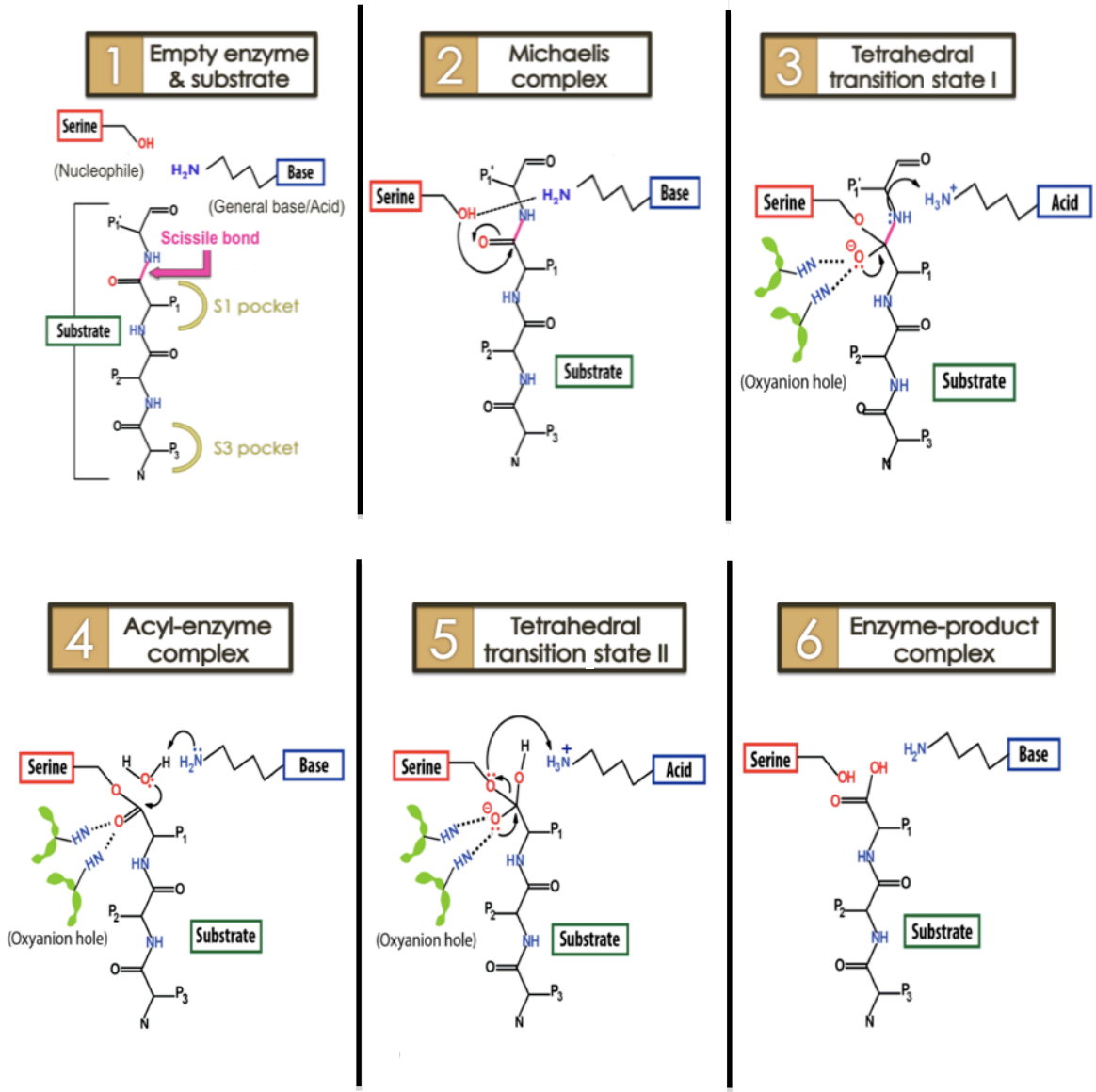


Figure 1-7. The generalized mechanism of Ser/Lys proteases.

This generalized mechanism was adapted from the Ser/His/Asp triad mechanism of chymotrypsin[38] and the Ser/Lys dyad mechanism proposed by Paetzel *et al.* [37]. The stages involved are empty enzyme and substrate (**step 1**), Michaelis complex (**step 2**), tetrahedral transition state I (**step 3**), acyl-enzyme complex (**step 4**), tetrahedral transition state II (**step 5**) and enzyme-product complex (**step 6**).

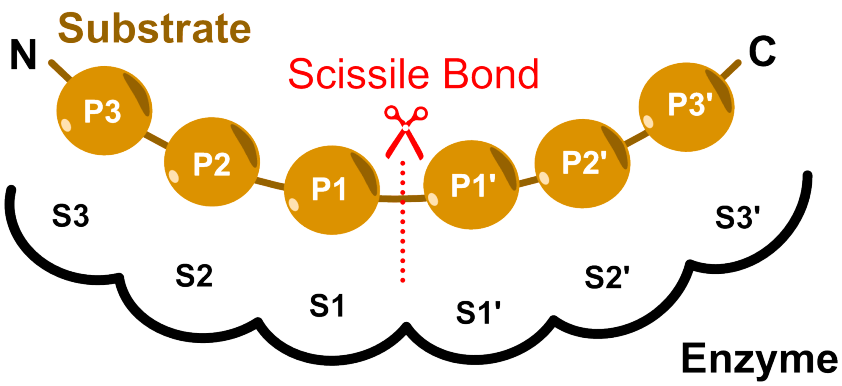
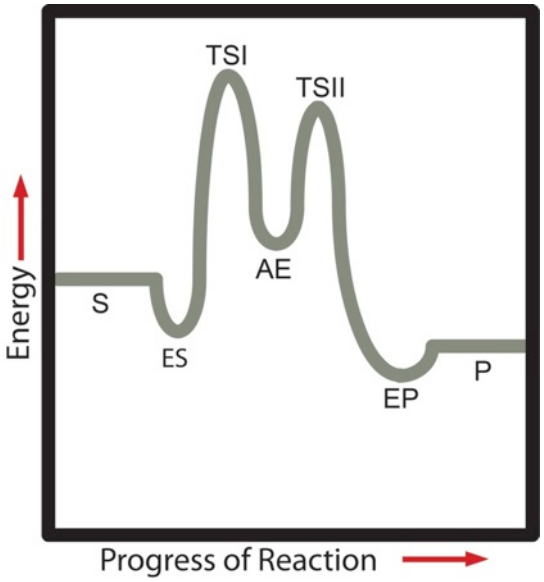


Figure 1-8. The Schechter and Berger nomenclature of peptidase specificity.

This nomenclature was first used by Schechter and Berger in 1967 to describe the peptidase specificity of papain[39]. The P1 residue is located immediately before the scissile bond, the bond cleaved by the peptidase. The "P" stands for peptide. The residues are numbered from the scissile bond towards the N-terminus. Residue P1' lies right after the scissile bond and they are numbered from the scissile bond towards the C-terminus. The subsites accommodating these residues have the same numbers as the peptides they accommodate but the "P" is replaced with an "S" to denote subsite.

A)



B)

$$E + S \rightleftharpoons ES \overset{P_1}{\rightleftharpoons} TS\ I \rightleftharpoons AE \rightleftharpoons TS\ II \rightleftharpoons EP \overset{P_2}{\rightleftharpoons} E$$

Figure 1-9. Free energy diagram for a two-step covalent catalysis mechanism.

A) The free energy diagram for a two-step covalent catalysis mechanism that involves the following steps: enzyme (E) and substrate (S), Michaelis complex (ES), tetrahedral transition state I (TSI), acyl-enzyme (AE), tetrahedral transition state II (TSII), enzyme-product complex (EP) and product (P) and enzyme. **B**) A simple equation of the reaction is shown.

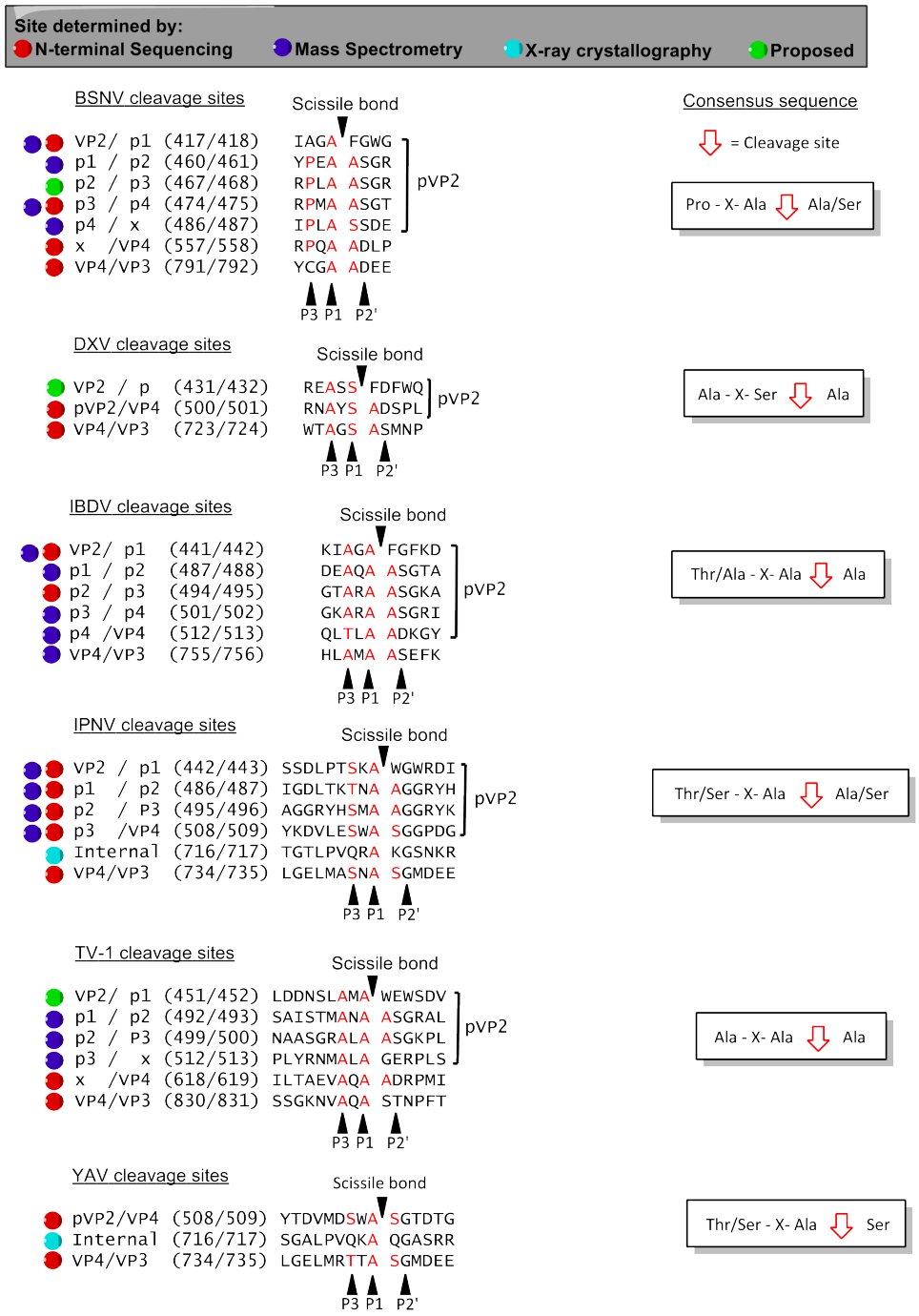


Figure 1-10. List of polyprotein cleavage sites in birnaviruses.

Figure 1-10. List of polyprotein cleavage sites in birnaviruses. (Continued from last page).

The polyprotein cleavage sites are shown for *Blotched snakehead virus* (BSNV) (Uniprot accession: Q8AZM0) [25], *Drosophila X virus* (DXV) (Uniprot accession: Q96724) [30], *infectious bursal disease virus* (IBDV) (Uniprot accession: Q9WI42) [42], IPNV (Uniprot accession: Q703G9) [43, 44], TV-1 (Uniprot accession: Q2PBR5)[31], and *Yellowtail ascites virus* (Uniprot accession: P89521)[45]. The coloured circles next to the cleavage sites denote the method of identification. The amino acids that match the consensus sequence are shown in red. The figure was prepared using ChemBioDraw Ultra 12.0.

1.6. Effect of pK_a on Ser/Lys dyad activity

The proper function of catalytic residues are linked to their protonation states which are governed by their pK_a [46]. Protease VP4 utilizes a Ser/Lys dyad mechanism. In order for lysine to function as a general base, its ε-amino group needs to reside in the deprotonated state. The pK_a of a lysine ε-amino group exposed to the solvent is ~10.5 so it remains protonated at physiological pH (**Table 1-3**). The enzyme solves this problem by providing an environment that depresses the pK_a [41, 47]. In general, the pK_a decreases when the group is placed in close proximity to a positive charge[47]. This favours the deprotonated state and helps to limit charge repulsion. The opposite is true when the residue is situated near a negative charge. The other situation that results in a shift in pK_a is to situate the titratable functional group in a hydrophobic environment; the pK_a shifts to favour a charge neutral state. Formation of favourable hydrogen bonds generate energy which can be used to lower the pK_a of hydrogen bond acceptors and increase the pK_a of the hydrogen bond donors [46].

Kinetic data show that the apparent pK_a for the lysine general base in signal peptidase 1, Lex A repressor and bacterial signal peptidase A (unpublished data), is indeed depressed to a value near 8.5 to 9.0 [7, 48, 49]. It has been difficult to measure directly the pK_a of lysine general bases by NMR, but it is being attempted.

Residue	pKa
Serine hydroxyl (nucleophile)	~ 13
Lysine ε-amino group (general base)	~ 10.5
Threonine hydroxyl	~ 13
Water	~ 14

Table 1-3. Standard pK_a values for functional groups involved in the VP4 mechanism.

The values listed are for solvated surface conditions, not those buried in the active site [41, 47].

1.7. Why is VP4 protease important?

Although none of the birnaviruses have so far been reported to cause disease in humans, many of these viruses are pathogenic to species that are consumed by humans (salmon, trout, yellowtail fish, chicken, and clam) and some are commercially farmed. Thus, there are economic reasons to study this family of viruses and prevent their propagation. Viral protease VP4 is a potential drug target for birnavirus as its inhibition will prevent cleavage of the polyprotein into components required for virion assembly. Study on VP4 cleavage events will also give insights into polyprotein processing as it substrates are the cleavage sites on the polyprotein. This is of value to the study of viruses in general. There are also scientific interests in studying VP4 as they utilize a serine-lysine (Ser/Lys) catalytic dyad that is more recently discovered and less characterized than the Ser/His/Asp classical catalytic triad. Also, the VP4 structure of *Infectious pancreatic necrosis virus* (IPNV) revealed an intermolecular (*trans*) acyl-enzyme complex with the C-terminus of an adjacent VP4, the native substrate, suggesting that VP4 proteases could be used as models for the study of acyl-enzyme complexes and other stages in the catalytic mechanism.

1.8. Rational inhibitor design

The discovery and development of inhibitors for VP4 protease would be a first step in the development of anti-viral compounds against birnavirus. X-ray crystallographic study yields detailed structural information on how the enzyme interacts with the substrate, product or inhibitor. It can also be used to capture the different stages in the reaction mechanism (**Figure 1-9**). Crystal structures are therefore essential in rational drug design [50]. As mentioned above, protease VP4 uses a Ser/Lys dyad catalytic mechanism. Various serine protease inhibitors function by trapping the intermediate steps in the reaction mechanism. Transition state analogue inhibitors form stable adducts with the enzyme thereby mimicking the tetrahedral transition state. Peptidyl aldehydes, chloromethyl ketones (CMK) and borate based compound are examples of oxyanion transition state analogue inhibitors [38, 51-53]. Acyl-enzyme intermediates are more stable compared to the transition states as they have lower free energies (**Figure 1-9**)[54]. Azapeptides inhibit chymotrypsin-like enzyme by forming acyl-enzyme intermediates [38, 55]. The transition from acyl-enzyme intermediate to tetrahedral transition state II requires the activation of a deacylating water. Inhibitors for TEM-1 β-lactamase were developed based on the displacement of a deacylating water [56, 57]. Water molecules also stabilize the interaction between the enzyme and the substrate by bridging the hydrogen bonding network. The potency of an existing human immunodeficiency virus (HIV) protease inhibitor was enhanced by the presence of a cyclic urea carbonyl oxygen that mimicked a hydrogen-bonded water in the active site [58]. Other inhibitors exploit the interactions between the enzyme and the substrate to enhance the binding affinity [59]. Some of the serine hydrolases are targets for marketed drugs, these include: β-lactamase, serine-type D-Ala-D-Ala carboxypeptidase and thrombin[60]. Using the information from crystal structures of VP4 protease we could use any of the above approaches to develop anti-birnavirus compounds.

1.9. Research objectives

1.9.1. Using X-ray crystallography to investigate birnavirus polyprotein processing

VP4 protease cleaves between pVP2 and VP4 as well as between VP4 and VP3 and the sequences of the cleavages site have been experimentally determined [25, 30, 31, 42-25]. There are two ways that VP4 can cleave the polyprotein: intermolecularly (*trans*) and intramolecularly (*cis*). In an intermolecular (*trans*) cleavage event, one molecule of VP4 cleaves another polyprotein molecule. In an intramolecular (*cis*) cleavage event, VP4 cleaves the same polyprotein as the enzyme itself resides. Residues on the P side of the substrate are the major determinant in substrate specificity as they interact with the binding groove and the specificity pockets. In contrast, the P' side of the substrate are less conserved and are in general do not contribute much to substrate specificity. The full-length VP4 protease (i.e. free from the polyprotein, no pVP2 or VP3 attached) contains residues on the P' side of the pVP2/VP4 junction at its own N-terminus and residues from the P side of the VP4/VP3 junction at its own C-terminus.

1.9.1.1. Goals and potential problems.

We would like to use X-ray crystallography to capture the structural interactions required for VP4 to cleave the polyprotein in both an intramolecular (*cis*) and in an intermolecular (*trans*) fashion. This thesis will focus on observing the cleavage events near the C-terminus of VP4 (i.e. VP4/VP3 cleavage site and the internal cleavage site).

Because the *cis* and *trans* cleavage events are potentially in competition with one another, producing ordered crystals might be difficult. We took advantage of the fact that some species of VP4 have an internal cleavage site and some species do not. Modeling studies suggest that those VP4 which are truncated at the internal cleavage site (i.e. end at the P1 residue of the internal cleavage site) are too short to form a *cis* interaction, therefore a *trans* interaction would predominate. YAV VP4 contains an internal cleavage site and we used both a native active site version and a construct with the lysine general base mutated to alanine to try to trap the *trans*-acyl enzyme. The internal cleavage site in YAV was previously proposed based on sequence conservation but was never been observed[45]. We will attempt to confirm this cleavage site by VP4 protein *trans* cleavage

assays and determine if the enzyme cleaved at the internal cleavage site is still active using VP4 protein *trans* cleavage assays and fluorometric peptide assays.

Other VP4 species lack an internal cleavage site. Modeling studies suggest that the VP4 C-termini (i.e. ending at the P1 residue of the VP4/VP3 cleavage site) of these VP4 are long enough to form a *cis* interaction. TV-1 VP4 protease is such an enzyme, and we attempted to trap it in a *cis* complex. This is a very unique situation which favours the formation of the *cis* complex in that there is a very high local effective concentration of the product (i.e. VP4/VP3 cleavage site). In other words, the specificity residues are covalently attached to the enzyme.

1.9.2. Capturing different stages of the catalysis mechanism to help in inhibitor development.

As mentioned earlier, many designed protease inhibitors mimic the different stages of the catalytic mechanism. When designing inhibitors, it is useful to have structural information for each stage of the reaction mechanism. In this project we will try to capture the empty active site, acyl-enzyme complex and product-enzyme complex.

2. Birnaviruses

2.1. Classification

Birnavirus derived its name from the Latin prefix “bi” meaning two and "rna" which highlight the bi-segmented (Segment A and B) double-stranded nature of its RNA genome [61]. Viruses belonging to the family *Birnaviridae* infect vertebrates, mollusks, insects and rotifers [62-69]. None of the know birnaviruses infect mammals[70]. However, birnavirus-like virions have been found in the feces of humans and animals (rats, guinea pigs, cattle, swine). These viruses are termed “picobirnaviruses” as they vary in size, number and length of genomic segments from the true birnaviruses[71]. The family of *Birnaviridae* is divided into four genera: *Avibirnavirus*, *Entomobirnavirus*, and *Aquabirnavirus*, and *Blosnavirus*, infecting birds, insects, aquatic organisms and blotched snakehead fish respectively. A list of diseases associated with birnaviruses is shown in **Table 2-1**.

Virus name	Host	Disease name/Symptoms	Reference
IBDV	Chicken	Gumboro Disease	Cosgrove [72]
TV-1	*Tellina tenuis*	Chalkier and thinner shell, pale yellow digestive gland	Buchanan [64]
IPNV	Trout, Salmon Halibut	Infectious pancreatic necrosis	Moss *et al.* [69] Christie *et al*. [68] Biering *et al.* [73]
YAV	Yellowtail fish	Yellowtail ascites disease	Hirayama *et al.* [62]
DXV	*Drosophila*	Anoxia sensitivity	Teninges *et al.* [63]
Eel Virus	European eel	skin tumour	Schwanz-Pfitzner *et al*. [74]

Table 2-1. Diseases associated with birnaviruses.

2.2. Virion properties

The genome of birnavirus is enclosed in a single-shelled non-enveloped icosahedral (T=13) capsid of ~ 60 to 70 nm in diameter (**Table 2-2**) [70]. The genome is approximately 6 kbp in length with segment A ranges from 3.0-3.6 Kbp and approximately 2.7-3.2 kbp in segment B (**Table 2-2**) [25, 28, 30, 31, 75-79]. Segment A has two open reading frames (ORFs) with the largest one encoding for a polyprotein. It consists of two structural proteins pVP2 and VP3 and a viral protease VP4 (**Figure 2-1**). Protein pVP2 is further processed to yield the mature capsid VP2 and additional peptides [44, 80]. The polyproteins of *Tellina virus 1* (TV-1) and *Blotched snakehead virus* (BSNV) also code for peptide X which is located between pVP2 and VP4[31]. In BSNV, *Drosophila X virus* (DXV), *Yellowtail ascites virus* (YAV), *Infectious bursal disease virus* (IBDV) and *Infectious pancreatic necrosis virus* (IPNV), an alternate ORF in segment A codes for VP5 (**Figure 2-1**) [25, 81]. An internal ORF of 165 codons is also found in segment A of TV-1 (**Figure 2-1**) [31]. Segment B encodes a RNA-dependent RNA polymerase, VP1.

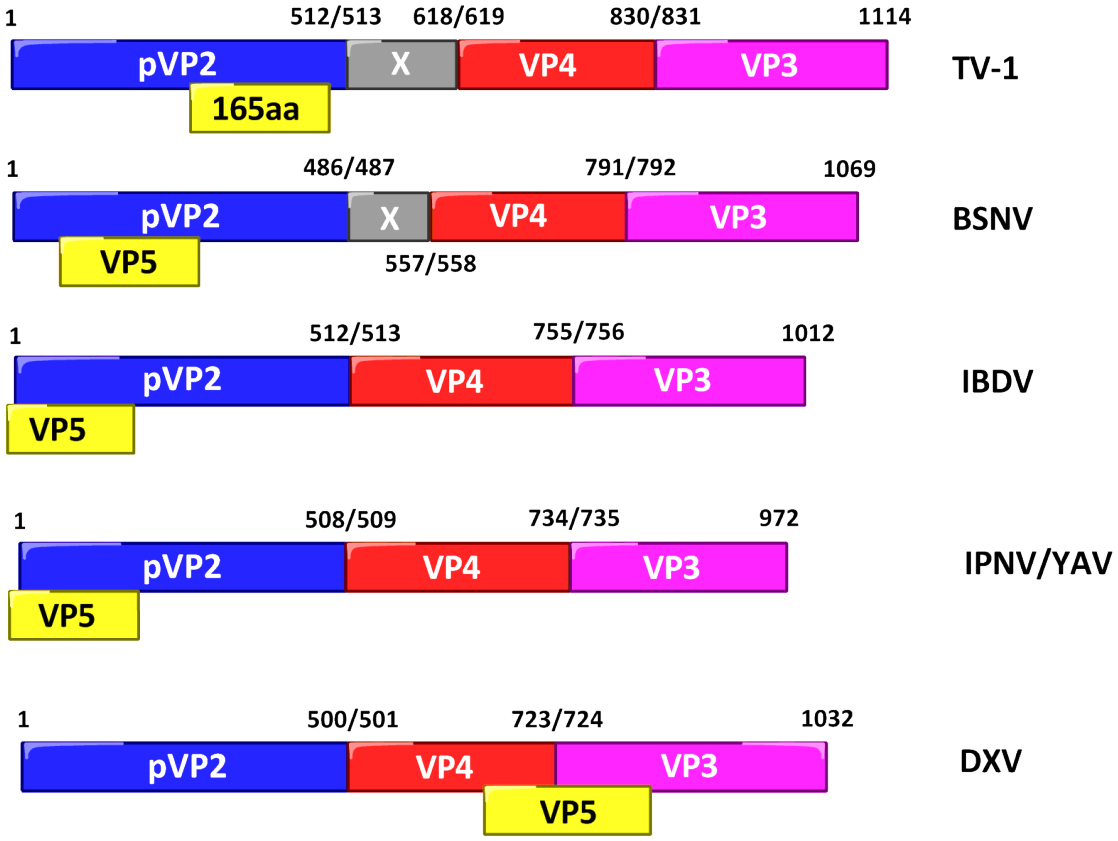


Figure 2-1. The genomic arrangement of segment A in birnaviruses.

The schematics of birnaviruse polyproteins are shown for BSNV (Uniprot accession: Q8AZM0) [25], DXV (Uniprot accession: Q96724) [30], IBDV (Uniprot accession: Q9WI42) [42], IPNV (Uniprot accession: Q703G9) [43, 44], TV-1 (Uniprot accession: Q2PBR5)[31], and YAV (Uniprot accession: P89521)[45]. The P1/P1' residues at the cleavage sites are shown above and below their respective junctions [31, 82]. The figure was prepared using ChemBioDraw Ultra 12.0.

Virus name	Refseq accession number	Diameter of capsid (nm)	Segment A (bp)	Segment B (bp)
BSNV	NC_005982.1 NC_005983.1	57	3429	2750
IBDV	NC_004178.1 NC_004179.1	65-70	3183	2715
IPNV	NC_001915.1 NC_001916.1	57-74	3097	2784
YAV	NC_004168.1 NC_004176.1	62-69	2976	2735
TV-1	AJ920335.1 AJ920336.1	59	3579	2817
DXV	NC_004177.1 NC_004169.1	59	3360	3243

Table 2-2. Genome and capsid size of birnaviruses.

The RefSeq accession numbers [67,83-86,61, 63], and the genome length[25, 75-78, 28, 30, 31,] are listed here for BSNV, IBDV, IPNV, YAV, TV-1 and DXV.

2.3. Viral life cycle

2.3.1. Mechanism of entry

The capsid structure of IBDV is evolutionarily related to that of rotavirus, a double-stranded RNA virus [87]. Rotaviruses enter the cell via endocytosis with dependency on cholesterol, the release of a membrane-perforating peptide, and the depletion of calcium ion [88-90]. Indeed, study on the cell culture adapted IBDV (caIBDV) which has mutations found in the capsid protein VP2, supports the notion of entry via an endocytotic pathway [91]. Infection by caIBDV was greatly reduced in the presence of drugs that sequester cholesterol (Nystatin), deplete cholesterol (Methyl-β cyclodextrin) or inhibit its synthesis (Progesterone) implicating the involvement of cholesterol. In IBDV, calcium ion was suggested to maintain the stability within and between VP2 capsid trimers [27, 92, 93]. The attachment of the caIBDV was also enhanced in the presence of $CaCl_2$ [91]. The primary receptor for IBDV has yet to be identified, but surface IgM from a chicken bursal lymphoma-derived DT40 cell line has been identified as a putative receptor that supported cell surface attachment[94]. The endocytosis was demonstrated to involve vacuolar H^+ ATPase (V-ATPase) positive vesicles by using bafilomycin A1, an inhibitor of V-ATPase, which abrogated the entry of caIDBV [91]. V-ATPase mediates the efflux of calcium ion into the cytosol thereby lowering the endosomal calcium concentration. The low calcium environment of the endosome was suggested to trigger the release of pep46 (residues 442-487 of the polyprotein), a maturation product of pVP2, from the viral particle [27]. This amphipathic peptide was shown to deforms and induce pores on synthetic membranes. Thus it was proposed to aid in virus penetration into the cytosol.

2.3.2. Viral assembly

A study on the morphogenesis of IPNV throughout its infective cycle in Chinook salmon embryo (CHSE)-214 cells revealed that two forms of viral particles exist [70]. The earliest particle is the pro-virion with the capsid composed of both mature and immature viral polypeptides. It is non-icosahedral, non-infectious and is 65-68 nm in diameter. The polyprotein and pVP2 were found to associate with the pro-virions. Maturation occurs when most of the remaining viral precursors within the capsid undergo proteolytic

cleavage. In IBDV, at least three of these pVP2-dervied peptides remained associated with the virion as detected using mass spectrometry and N-terminal sequencing methods[42]. Investigations of IBDV also suggests that the maturation of pVP2 to VP2 is driven by the initial preassembly of the premature virions [80]. The mature viral particle becomes icosahedral, more compact (60 nm in diameter) and infectious.

2.4. Examples of birnaviruses

2.4.1. Causative agent of Gumboro disease – Infectious bursal disease virus (IBDV)

Infectious bursal disease virus (IBDV) is an *avibirnavirus* that attacks the bursa of Fabricius, an important lymphoid organ of chicken [95, 95]. It was initially referred to as avian nephrosis and was first reported by Cosgrove in 1962 in Gumboro, Delaware of USA, hence its other name Gumboro disease [72]. The disease is highly contiguous as it can spread readily by direct contact as well as through contaminated water and feed[96]. In addition, the contaminated building remained infectious after 122 days. Antibodies against IBDV were isolated in emperor penguin chicks and adult Adélie penguins in the Antarctic as well as wild birds in Japan (e.g. peregrine falcon, jungle crow, and European widgeon) [97, 98]. The virus has also been isolated from turkey[99]. IBDV displays tropism for dividing pre-B lymphocytes in the bursa of Fabricius in chicken [95, 95]. Chicks 3-6 weeks of age are the most vulnerable to the disease as the bursa of Fabricius is at its maximal stage of development. However, signs of the disease are rarely developed in birds older than 6 weeks. Clinical signs of the disease include distress, depression, ruffled feathers, anorexia, diarrhea, trembling, and dehydration which usually lead to substantial mortality[100]. The surviving birds may become immunosuppressed which increase their susceptibility to secondary infections and interfere with vaccinations against other viruses (e.g. Newcastle virus, Marek's disease and infectious bronchitis) [101]. Serotype 1 of IBDV is pathogenic and is divided into three antigenic subgroups categorized according to their virulence: 1) classic (0% mortality) ; 2) variant (10-50% mortality); and 3) very virulent (50-100% mortality)[71,101] . In contrast, Serotype 2 of IBDV is non-pathogenic. The genomic arrangement in segment A and a list of IBDV polyprotein cleavage sites are shown in **Figure 2-2**.

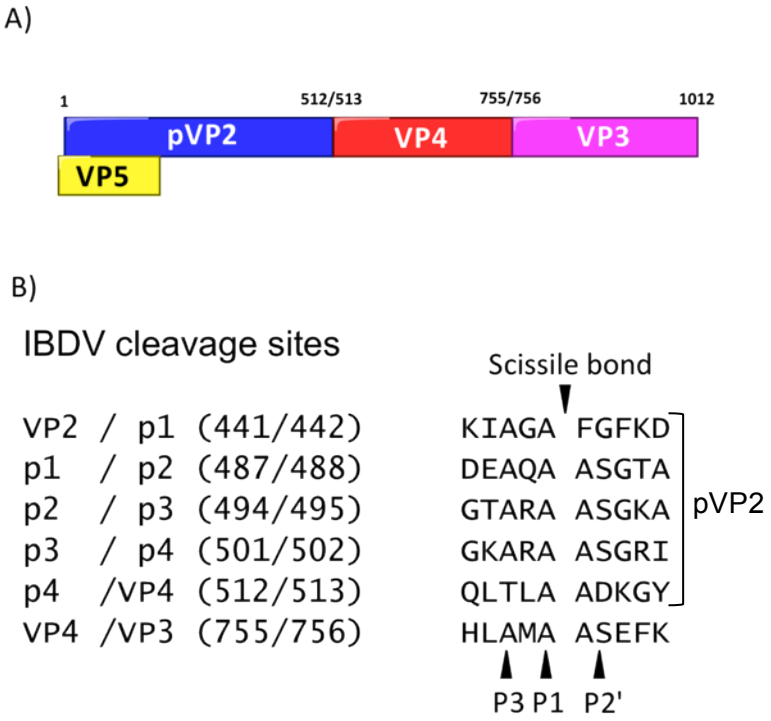


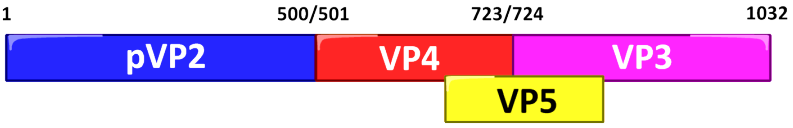
Figure 2-2. The genomic arrangement of segment A and a list of polyprotein cleavage sites in IBDV.

The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of known polyprotein cleavage sites [1, 42].

2.4.2. *Anoxia sensitivity inducing birnavirus – Drosophila X virus (DXV)*

Drosophila X virus (DXV) is an *Entomobirnavirus* that induces anoxia sensitivity and death in *Drosophila Melanogaster* (*D. Melanogaster*) [61, 63]. It was first isolated and characterized by Teninges *et al.* in 1979 as a contaminant from the uninfected control passage of *D. Melanogaster* during a study on sigma virus, an rhabdovirus known to induce CO_2 sensitivity in *Drosophila* [63, 102]. Upon examination under the negative contrast electron microscopy, Teninges *et al.* observed no rhabdoviruses but instead found large icosahedral particles showing similar morphological features to IPNV, IBDV and TV-1 [63]. In these infected dead files, viral particles were detected in the cytoplasm of the digestive tract, brain, thorax, muscles, ovaries, testis and Malpighian tubules. The virus can be transmitted via contact as uninfected files became infected after cohabitation with the infected ones. DXV was also detected in *Culicoides* species [103]. Injection of DXV resulted in CO_2 sensitivity and death in files. Death also occurred after exposure to nitrogen for 15 min at 12°C in the infected flies. Thus, Teninges *et al.* assumed that the absence of oxygen caused death in these infected flies and termed this symptom "anoxia sensitivity". The genomic arrangement in segment A and a list of DXV polyprotein cleavage sites are shown in **Figure 2-3**.

A)



B)

DXV cleavage sites

		Scissile bond		
VP2 / p	(431/432)	REASS	FDFWQ	pVP2
pVP2/VP4	(500/501)	RNAYS	ADSPL	
VP4 /VP3	(723/724)	WTAGS	ASMNP	
		P3	P1 P2'	

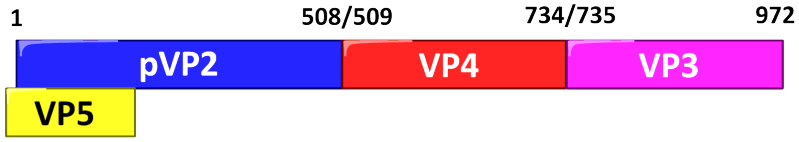
Figure 2-3. The genomic arrangement of segment A and a list of polyprotein cleavage sites in DXV.

The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of proposed and known cleavage sites in DXV [30].

2.4.3. Causative agent of infectious pancreatic necrosis – *Infectious pancreatic necrosis virus (IPNV)*

Aquabirnavirus infectious pancreatic necrosis virus (IPNV) is highly contagious and causes infectious pancreatic necrosis in salmonid fish with mortality rate ranges from 10 to 90% [71]. It was first described by M'Gonigle in 1940 as acute catarrhal enteritis in salmonid fingerlings [104]. The disease was later renamed to infectious pancreatic necrosis (IPN) by Wood *et al*. in 1955 [105]. High concentration of IPNV was detected in the water during outbreaks of IPN and feces of carrier brook trouts [106-108]. Germ-line (vertical) transmissions have also been reported in brook trout and rainbow trout possibly via the ovarian or seminal fluids [109, 110]. The disease tends to affect trout fingerlings shortly after they start feeding in freshwater and Atlantic salmon smolts after they transfer to sea water cages [71]. The infection becomes subclinical as the age of the fish increases but might persist over the lifetime of the fish. Symptoms of the disease include necrosis of acinar cells in the pancreas, abdomen swelling, mild to moderate bilateral exophthalmos, pale gills, cutaneous haemorrhages and swimming in a frantic corkscrew fashion. The polyprotein cleavage sites of IPNV can be described by the consensus motif of (Ser/Thr)-X-Ala↓(Ala/Ser) with the exception of the internal junction (Gln-Arg-Ala↓Lys) and between amino acid 442 and 443 (Ser-Lys-Ala↓Trp) (**Figure 2-4**). The genomic arrangement in segment A and a list of IPNV polyprotein cleavage sites are shown in **Figure 2-4**.

A)



B)

IPNV cleavage sites

		Scissile bond	
VP2 / p1	(442/443)	SSDLPTSKA	WGWRDI
p1 / p2	(486/487)	IGDLTKTNA	AGGRYH
p2 / P3	(495/496)	AGGRYHSMA	AGGRYK
p3 /VP4	(508/509)	YKDVLESWA	SGGPDG
Internal	(716/717)	TGTLPVQRA	KGSNKR
VP4/VP3	(734/735)	LGELMASNA	SGMDEE
		P3 P1	P2'

pVP2

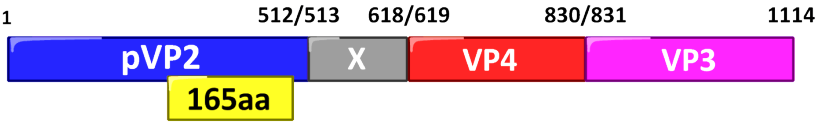
Figure 2-4. The genomic arrangement of segment A and a list of polyprotein cleavage sites in IPNV.

The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of known polyprotein cleavage sites in IPNV [43, 44].

2.4.4. A birnavirus isolated from bivalve mollusk – Tellina virus 1 (TV-1)

Tellina virus 1 (TV-1) was first described by Buchanan in 1973 who noticed a group of sand dwelling bivalve mollusks *Tellina tenuis* with thinner and chalkier shells[64]. The digestive glands of these mollusks were pale yellow instead of the usual dark brown. Further examination revealed a high proportion of the cells in the digestive gland, which produce the shell precursor material, were necrotic and harboured membrane-bound inclusion bodies with icosahedral viral particles of ~70nm in diameter. Similar viral particles were absent in the apparently healthy population. The virus was later isolated and characterized by Hill in 1976 [111]. TV-1 was shown to be antigenically distant from BSNV, IPNV, DXV and IBDV through cross-neutralization assays[67, 86, 112]. It expresses a non-structural polypeptide, X, between pVP2 and VP4 which is also found in BSNV [113] . Segment A of TV-1 VP4 codes for a polyprotein of 1114 amino acids. The self-encoded VP4 protease cleaves after amino acid residue positions 512, 618, and 830 to yield capsid precursor protein pVP2, peptide X, VP4 itself and VP3 (**Figure 2-5**) [31]. The cleavage sites identified can be described by the consensus motif Ala-X-Ala↓Ala with the exceptions of the VP2/p1 and p3/X junctions [31]. The genomic arrangement in segment A and a list of polyprotein cleavage sites in TV-1 are shown in **Figure 2-5**.

A)



B)

TV-1 cleavage sites

		Scissile bond ▼	
VP2/ p1	(451/452)	LDDNSLAMA ▼ WEWSDV	pVP2
p1 / p2	(492/493)	SAISTMANA ▼ ASGRAL	
p2 / P3	(499/500)	NAASGRALA ▼ ASGKPL	
p3 / x	(512/513)	PLYRNMALA ▼ GERPLS	
x /VP4	(618/619)	ILTAEVAQA ▼ ADRPMI	
VP4/VP3	(830/831)	SSGKNVAQA ▼ STNPFT	
		P3 P1 P2'	

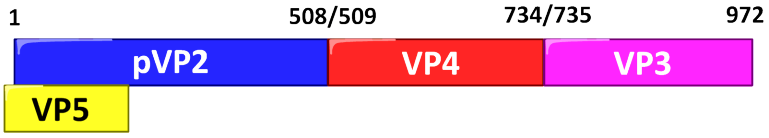
Figure 2-5. The genomic arrangement of segment A and a list of polyprotein cleavage sites in TV-1.

The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of proposed and known cleavage sites in TV-1 [31].

2.4.5. An ascites causing birnavirus – Yellowtail ascites virus (YAV)

In early summer of 1983, Sorimachi *et al.* reported ascites among yellowtail fingerlings in a farm in the Seto Inland Sea of Japan [85]. The causative agent was later identified to be *Yellowtail ascites virus* (YAV). YAV and YAV-like viruses are members of marine birnavirus (MABV) which infects both fish and shellfish [114]. YAV infection leads to ascites in yellowtail fish (*Seriola quinqueradiata*), which is popular in sushi, thus causing significant losses to the fish-farming industry [62]. YAV-like viruses have been isolated from many fish species including Japanese flounder (*Paralichthys olivaceus*), red sea bream (*Pagrus major*) and tiger puffer (*Takifugu rubripes*) [85, 115-117]. MABV infections have been reported in shellfishes jack knife clam (*Sinonovacura constricta*) and Japanese pearl oyster (*Pinctada fucata*) [118-120]. Segment A of YAV codes for a polyprotein of 972 amino acids. The self-encoded VP4 protease cleaves at the C-terminal side of resides 508 and 734 to yield pVP2, VP4 and VP3 [45] (**Figure 2-6**). The cleavage sites at the pVP2/VP4 and VP4/VP3 junctions can be described by the consensus (Ser/Thr)-X-Ala↓Ser (**Figure 2-6**)[45]. Previous site-directed mutagenesis studies have identified Lys^{674} as the general base but failed to confirm Ser^{633} as the nucleophile. An internal cleavage site is found within the VP4 region of IPNV, and since the sequence identify of Segment A between IPNV and YAV is high (~ 90%), an internal cleavage site was proposed at equivalent sites in YAV. However, the existence of the site was not confirmed. The genomic arrangement in segment A and a list of proposed and known YAV polyprotein cleavage sites are shown in **Figure 2-6**.

A)



B)

YAV cleavage sites

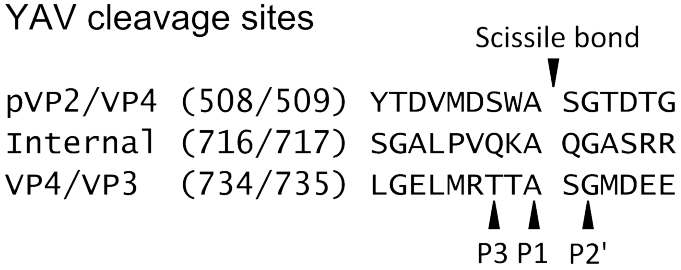


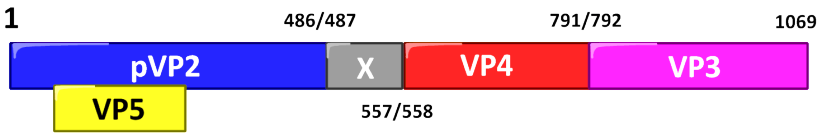
Figure 2-6. The genomic arrangement of segment A and a list of polyprotein cleavage sites in YAV.

The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of proposed and known polyprotein cleavage sites in YAV [45].

2.4.6. A birnavirus isolated from a cell line originated from an apparently healthy carrier fish – Blotched snakehead virus (BSNV)

In 1999, John *et al.* characterized a virus isolated from a warm-water fish cell line developed from the caudal peduncle of blotched snakehead fish (*Channa lucius*) and was given the name blotched snakehead virus (BSNV) [67]. BSNV is the type species of genus *Blosnavirus* belonging to the family *Birnavirdae* [25]. It infects blotched snakehead fishes found in peat swamps and forest streams of South-east Asia and are being sold in live fish market and aquarium fish trade [121, 122]. They feed on fishes, prawns, crabs and occasionally shrimps[123]. Its head resembles that of a snake with large canine teeth and scales on the gular region of the head. Its body is elongated and cylindrical with dark blotches. Segment A of BSNV codes for a polyprotein of 1069 amino acids (**Figure 1-1, 2-7**)[25]. All of the polyprotein cleavage sites can be described by the Pro-X-Ala↓(Ala/Ser) motif (**Figure 2-7**) with the exception of the VP4-VP3 junction (Cys-Gly-Ala↓Ala) and between amino acid 417 and 418 (Ala-Gly-Ala↓Phe). The genomic arrangement in segment A and a list of proposed and known polyprotein cleavage sites in BSNV are shown in **Figure 2-7**.

A)



B)

BSNV cleavage sites

		Scissile bond
VP2/ p1	(417/418)	IAGA FGWG
p1 / p2	(460/461)	YPEA ASGR
p2 / p3	(467/468)	RPLA ASGR
p3 / p4	(474/475)	RPMA ASGT
p4 / x	(486/487)	IPLA SSDE
x /VP4	(557/558)	RPQA ADLP
VP4/VP3	(791/792)	YCGA ADEE

pVP2

P3 P1 P2'

Figure 2-7. The genomic arrangement of segment A and a list of polyprotein cleavage sites in BSNV.

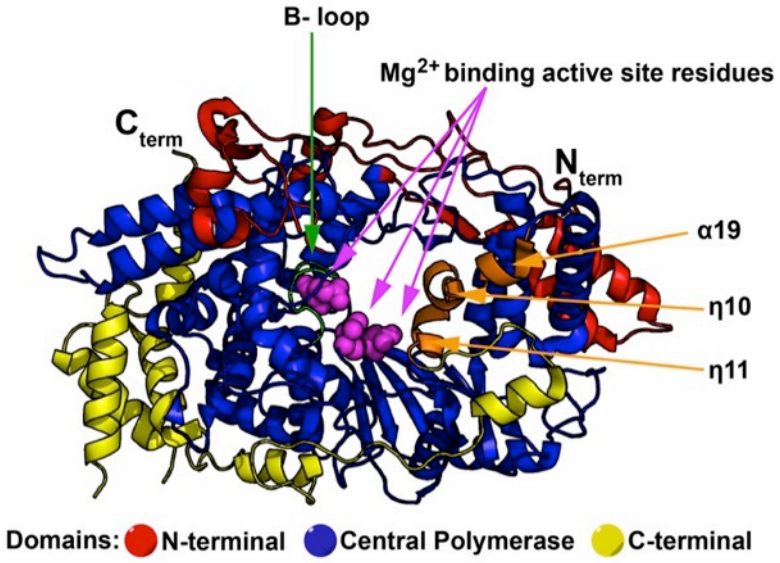
The genomic arrangement of segment A is shown in **A)** with the amino acid positions labeled at the cleavage junctions. **B)** A list of proposed and known polyprotein cleavage sites in BSNV [25].

2.5. Protein in birnaviruses

2.5.1. Viral protein 1 (VP1)

2.5.1.1. Structure

The VP1 of IBDV (**Figure 2-8**) consists of three domains made up of 37 helices (24 α-helices and 12 3_{10} helices) and 17 β-strands[124]. On one side of the catalytic cleft, residues 1-168 form the N-terminal domain (**Figure 2-8,** red) which serves to bridge the finger and thumb domains. The classical "fingers", "palm", and "thumb" subdomains are found in the central polymerase domain (**Figure 2-8,** blue) consisting of residues 169-657. The C-terminal domain (residue 658-845, **Figure 2-8,** yellow) assumes the shape of a horseshoe and is located on the opposite side of the catalytic cleft where it bridges the fingers and palm domains. A protrusion (**Figure 2-8**, orange) formed by α19, η10, and η11 of the C-terminal domain partially blocks the active site cleft. Its location is reminiscent of the C-terminal initiation platforms found in bacteriophage ϕ6 and hepatitis C virus RDRPs [125, 126]. The central polymerase domain shares similar overall architecture with those found in positive-strand RNA picorna- and caliciviruses. Comparison with other RDRPs revealed a permutation of the sequence motifs A, B and C in IBDV where the motif is found in the order of C-A-B instead of the usual A-B-C [124]. In the active site, two Mg^{2+} bind to Asp402 of motif C, and Asp416 and Glu-421 of motif A where the catalytic metal ions are expected to bind (**Figure 2-8**, magenta). In the absence of VP3, the B loop (residues 483-490, **Figure 2-8**, green) protrudes into the active site cleft preventing the binding of the template and incoming nucleotides. However, the B loop is displaced from its initial position in the complex structure of VP1 and peptide GRLGRWIRTVSDEDLE which corresponds to the VP1-binding domain of VP3. Equivalent loops are found in polymerases of foot-and-mouth disease virus (FMDV) and bacteriophage ϕ6. Pre-incubation of VP1 with VP3 or VP3 C-terminal peptide enhanced the replication of IBDV-derived ssRNA+ template in a concentration dependent manner. A list of VP1 features, PDBs, sequence comparison, and sequence alignment can be found in **Tables 2-3, 2-4** and **Figure 2-9**.



PDB Code	Resolution(Å)	UniProt accession number	Reference
2R70	2.7	Q82629[3]	Garriga *et al*.[124]

Figure 2-8. Structure of IBDV VP1.

The central polymerase domain (blue) houses the classical "fingers", "palm", and "thumb" subdomains [124]. It is flanked by the N-terminal domain (red) and the C-terminal domain (Yellow). Other key features of this polymerase are highlighted in colour: the B-loop (green), magnesium binding active site residues (Asp 402, Asp416 and Glu421, magenta) and a protrusion formed by α19, η10, and η11 (orange). Figure was prepared using Pymol[35].

2.5.1.2. Self-priming activity

The transcription of IPNV occurs in the cytoplasm using a semi-conservative and strand-displacement mechanism[127, 128]. Unlike other dsRNA viruses, the RNA synthesis in birnavirus is initiated using a protein primer [129]. The protein-priming function is achieved by the self-guanylylation of VP1 and its mechanism was studied using IPNV [130, 131]. The guanylyalation site was mapped to the N-terminal domain and is independent of its polymerase activity as the essential active site mutants remained self-guanylylation competent. At initiation, the hydroxyl group of tyrosine or serine residues in the protein-primer molecule becomes covalently attached to the α-phosphate of a guanosine monophosphate (GMP) moiety via a phosphodiester bond[76, 131-133]. This allows for further guanylylation to yield VP1-GG which was proposed to bind to the -CC sequence on viral RNA template at the 3' end [131, 134]. Consequentially, VP1 is linked to the genome (VPg). The RNA-dependent RNA polymerase (RDRP) can also exist as both a free polypeptide (VP1) [127].

Virus name	UniProt accession number	Sequence length (AA.)	Size (KDa)	pI	PDB ID	Reference
BSNV	Q8AZL8	867	96.8	6.79		Da Costa *et al.*[25]
IPNV	P22173	845	94.5	6.14	1)2YI8 [135] 2)2YI9 [135] 3)2YIA [135] 4)2YIB [135]	Duncan *et al.*[78]
YAV	Q8BAD4	845	94.4	5.86		Zhang *et al*.[79]
IBDV	Q9Q6Q5	881	97.9	8.34	1)2PGG [137] 2)2QJ1 (to be published, Garriga *et al.*), 3)2R72 [124] 4)2R70 [124] 5)2PUS [124]	Boot *et al.* [3]
DXV	Q91CD5	997	112.8	8.44		Shwed *et al*.[76]
TV-1	Q2PBR4	894	99.5	6.65		Nobiron *et al.*[31]

Table 2-3. Table of features for VP1.

The sequences were obtained from UniProt. Values for sequence length, size and pI were calculated using ProtParam [22, 23].

Identity(%) / similarity(%)	BSNV	IPNV	YAV	IBDV	DXV
IPNV	46.4/58.6				
YAV	46.3/58.0	95.7/98.1			
IBDV	49.1/60.3	44.9/58.9	44.8/58.7		
DXV	25.5/35.6	25.0/35.0	24.3/34.7	25.5/35.9	
TV-1	29.1/41.4	28.9/40.7	28.7/40.6	29.3/40.8	21.4/32.7

Table 2-4. Sequence identity and similarity between different VP1s.

The protein sequence alignment was generated using ClustalW [32] and the sequence identities and similarities were calculated using the sequences identities and similarities webserver (URL:http://imed.med.ucm.es/Tools/sias.html) using the following equation:

$$\%\ \text{identity or similarity} = 100\left(\frac{\text{Identical positions}}{\text{Length of the alignment}}\right)$$

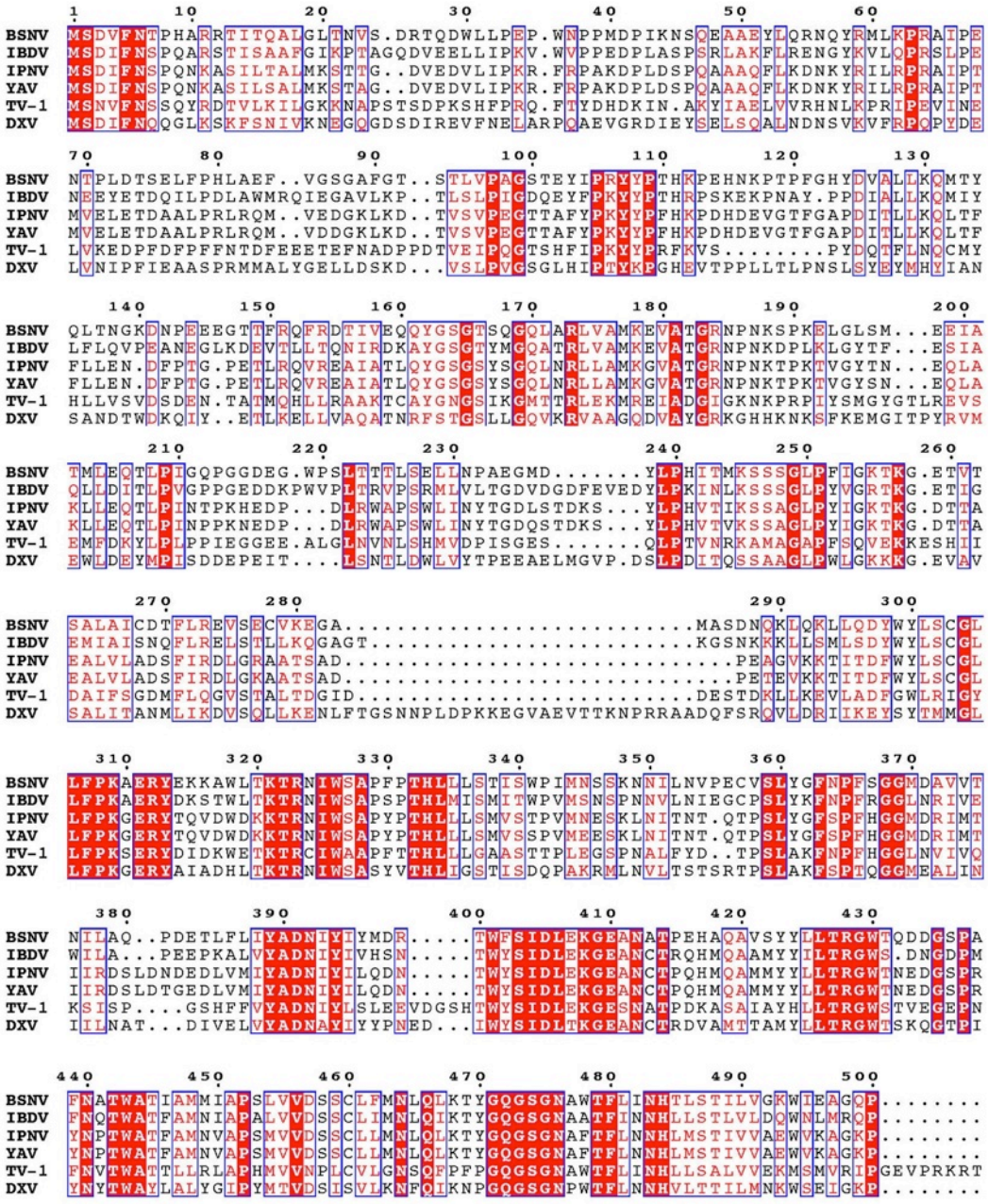


Figure 2-9. Protein sequence alignment of VP1(Part I).

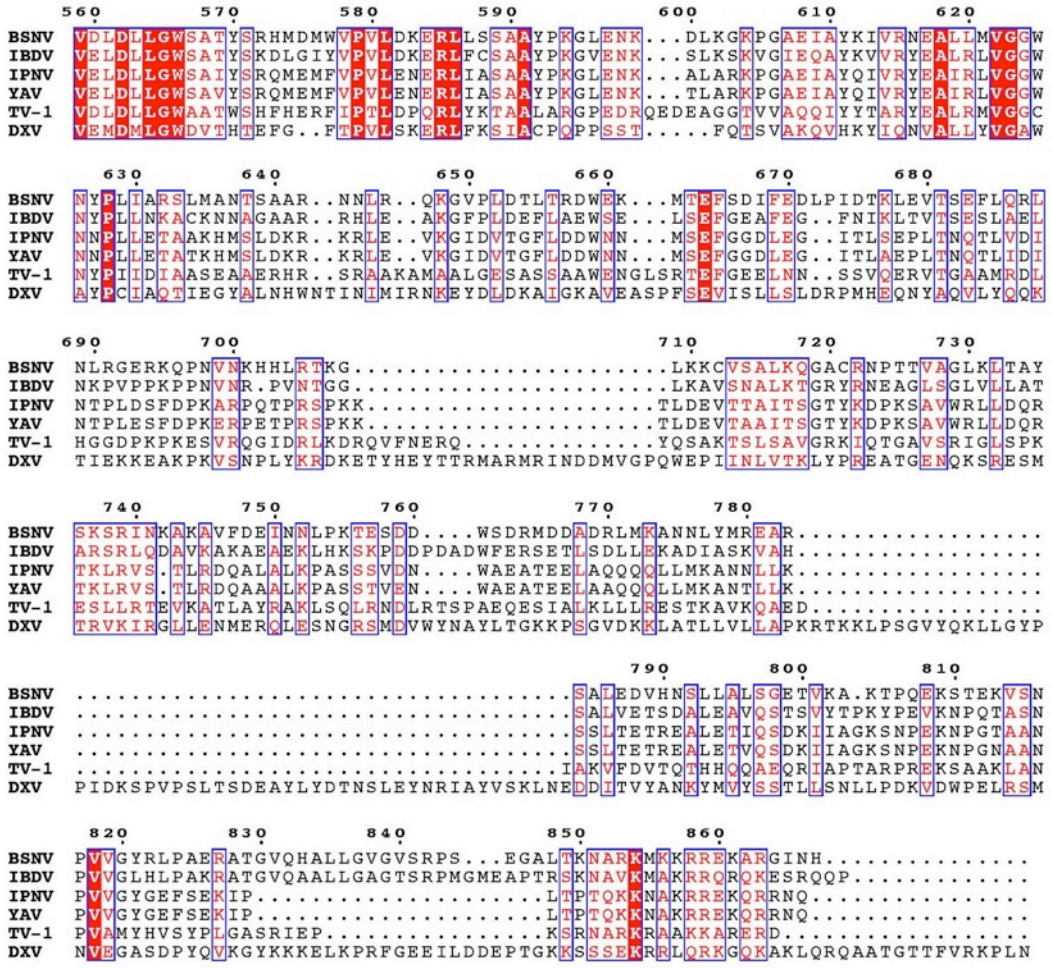


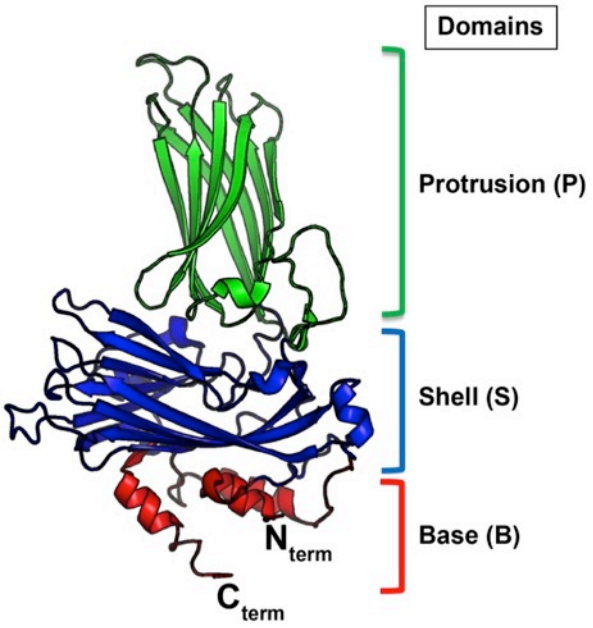
Figure 2-9. Protein sequence alignment of VP1(Part II).

The protein alignment was done using ClustalW [32] and graphical output was generated using ESPript [33]. Residues that are conserved are shown as white text highlighted in red and semi-conserved residues are shown as red text in blue boxes.

2.5.2. Viral protein 2 (VP2)

2.5.2.1. Structure

Each monomer of VP2 (**Figure 2-10**) composes of three domains and three of the monomers come together to form the basic building block of the capsid (**Figure 2-11**) [87, 92]. Three α -helices comes together to form the base (B) domain (red) (**Figure 2-10**). A "Swiss roll" β-barrel (residues 35-173 and 348-376) stacks on top of the B domain to form the shell (S) domain (blue). The protrusion (P) domain (green) position roughly perpendicular to the S domain where yet another "Swiss roll" β-barrel (residue 202-341) is found. A list of VP2 features, PDBs, sequence comparison, and sequence alignment can be found in **Tables 2-5, 2-6** and **Figure 2-12**.



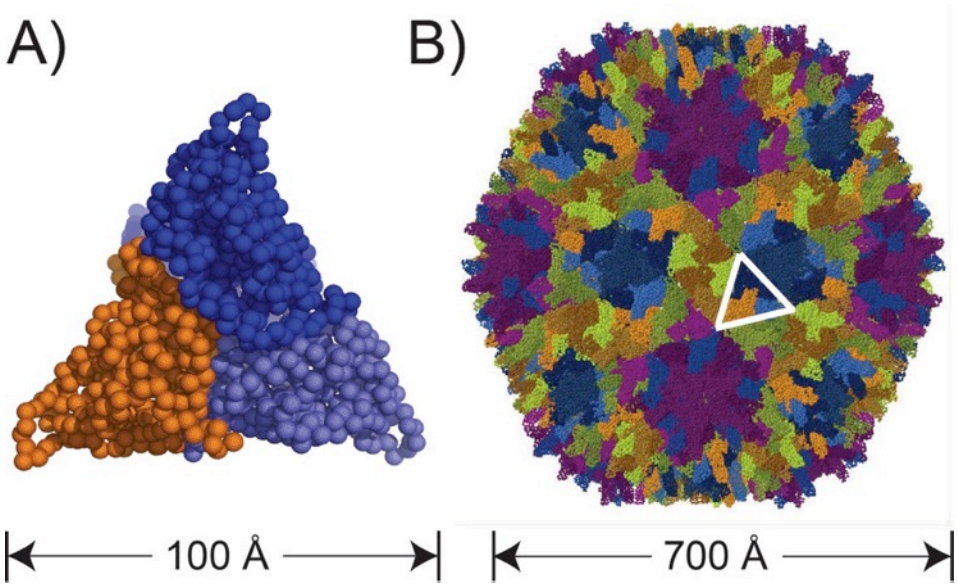
PDB Code	Resolution(Å)	UniProt accession number	Reference
1WCD	3.0	P15480 [144]	Coulibaly *et al*. [87]

Figure 2-10. Structure of IBDV VP2 monomer.

VP2 consists of three domain: the protrusion domain (P, green), the shell domain (S, blue) and the base domain (B, red)[87]. Figure was prepared using Pymol[35].

2.5.2.2. Capsid particle and pVP2 processing

The T=13 icosahedral capsid (**Figure 2-11**) of birnavirus composes of VP2 (**Figure 2-10**) [92]. However, when IBDV VP2 was expressed alone in insect cells, it formed dodecahedral T-1 particles [83]. Expression of IBDV polyprotein (pVP2-VP4-VP3) in a baculovirus system resulted in the formation of pVP2 (a precursor of VP2) tubules, but very few virus-like particles (VLPs) were detected [80]. In contrast, expression of the polyprotein with a green fluorescent protein (GFP) fused to the C-terminus of VP3 led to self-assembly of VLPs. These VLPs showed the same diameter and apparent geometry as the virions in addition to pVP2 processing. The presence of GFP was suggested to mimic the interaction with VP1. The pVP2 tubules mentioned above are likely to be homologous to the type I tubules which are also consist of pVP2 and were detected in birnavirus infected cells [63, 74, 138-141]. Mass spectrometry and N-terminal sequencing on purified IBDV virions revealed the presence of four small peptides (residues 442-487, 488-494, 495-501, 502-512) derived from pVP2 (residues 1 -512) processing [42]. Three of these peptides (442-487, 488-494, 502-512) were found to associated with the virion and two of which (442-487, 502-512) are essential for virus viability. In IBDV, the cleavage at position 441 which generates the matured VP2 capsid protein was mediated by the autoproteolytic activity of pVP2 [2].



PDB Code	Resolution(Å)	UniProt accession number	Reference
1WCE	7.0	P15480 [144]	Coulibaly *et al*. [87]

Figure 2-11. Structure of T13 IBDV viral particle formed by trimers of VP2.

A) A space filling model of VP2 trimer. Three molecules (orange, dark blue, purplish blue) of VP2 come together to form a trimer, the basic building blocking for the capsid [87]. **B)** Each monomer in the asymmetric unit is shown in a different colour and is shown using a space filling model. A single trimer is highlight with a white outline. Figure was prepared using Pymol [35] and Jmol Version 12.2.15[142].

Virus name	UniProt accession number	Sequence length (AA.)	Size (KDa)	pI	PDB ID	Reference
BSNV	Q8AZM0	486	51.7	8.29		Da Costa *et al.*[25]
IPNV	Q703G9	508	55.7	5.64	3IDE [148]	Galloux *et al.*[27]
YAV	P89521	508	55.7	5.63		Suzuki *et al.*[28]
IBDV	P15480	512	54.4	5.50	1)1WCD [87] 2)1WCE [87] 3)2DF7 [92] 4)2GSY [93] 5)2IMU (To be publish, Galloux *et al.*), 6) 3FBM [2]	Spies *et al.* [144]
DXV	Q96724	500	54.2	5.14		Chung *et al.*[30]
TV-1	Q2PBR5	512	54.6	6.68		Nobiron *et al.* [31]

Table 2-5. Table of features for pVP2.

The sequences were obtained from UniProt. Values for sequence length, size and pI were calculated using ProtParam [22, 23].

Identity(%) / Similarity(%)	BSNV	IPNV	YAV	IBDV	DXV
IPNV	39.7/52.5				
YAV	38.8/51.7	90.2/93.5			
IBDV	46.6/58.7	41.0/55.0	41.8/53.6		
DXV	35.4/49.2	31.1/47.9	31.1/47.5	33.5/50.2	
TV-1	36.9/50.5	31.2/44.0	31.0/44.4	33.8/50.9	30.0/44.5

Table 2-6. Sequence identity and similarity between different pVP2s.

The protein sequence alignment was generated using ClustalW [32] and the sequence identities and similarities were calculated using the sequences identities and similarities webserver (URL:http://imed.med.ucm.es/Tools/sias.html) using the following equation:

$$\%\ \text{identity or similarity} = 100\left(\frac{\text{Identical positions}}{\text{Length of the alignment}}\right)$$

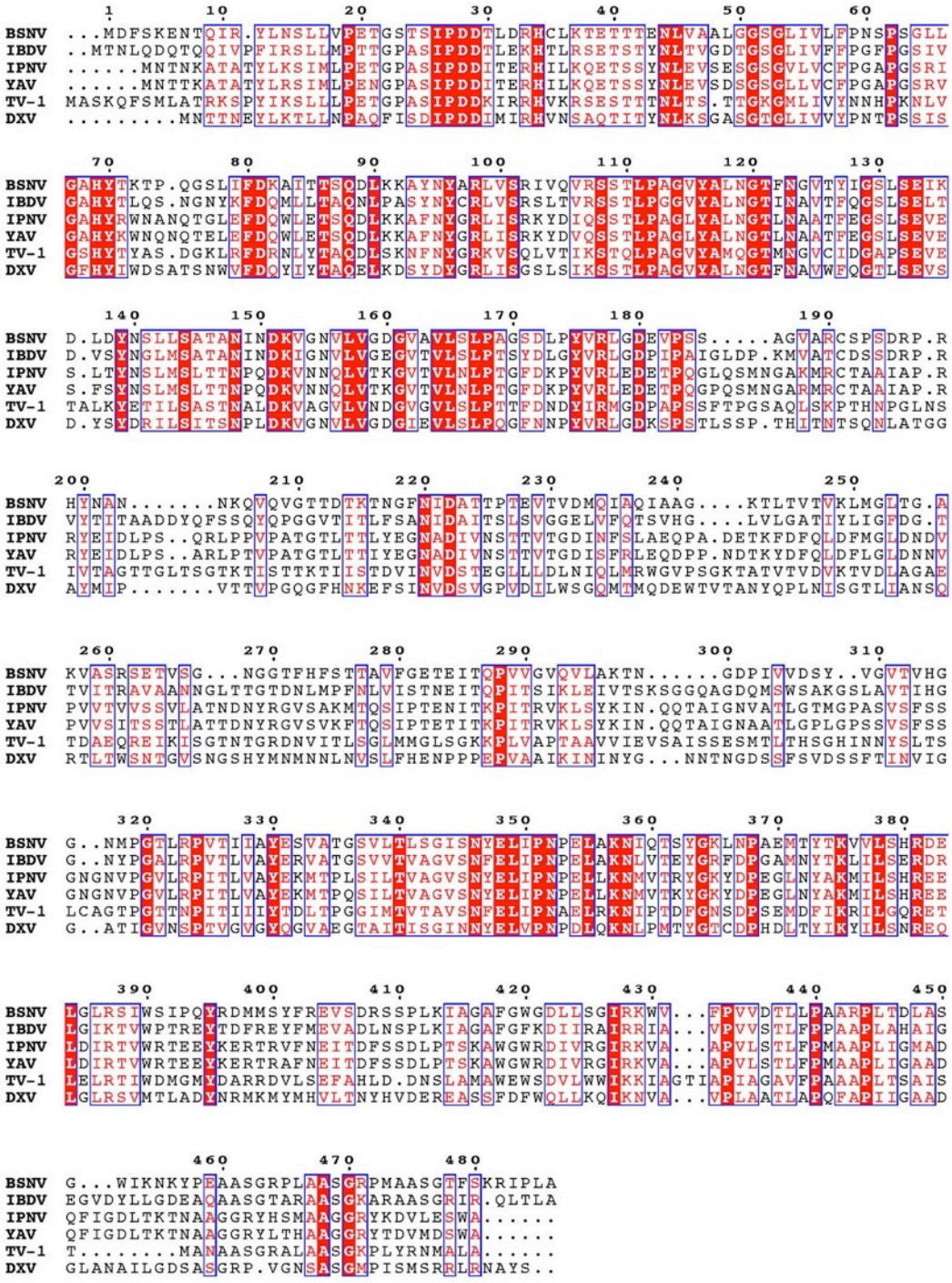


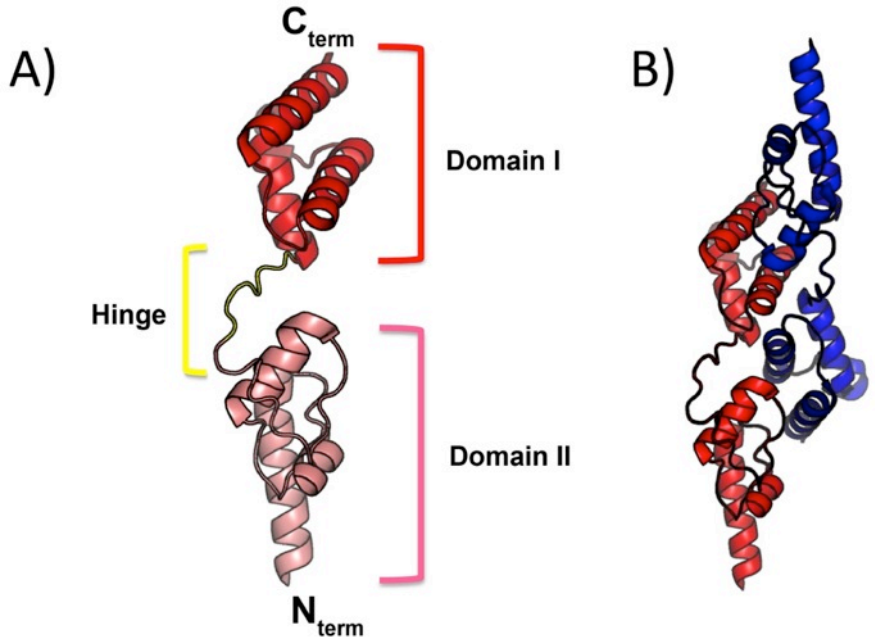
Figure 2-12. Sequence alignment of pVP2.

Figure 2-12. Protein sequence alignment of pVP2 (Continued from last page).

The protein alignment was done using ClustalW[32] and graphical output was generated using ESPript [33]. Residues that are conserved are shown as white text highlighted in red and semi-conserved residues are shown as red text in blue boxes.

2.5.3. Viral protein 3 (VP3)

The dimeric structure of IBDV VP3 (**Figure 2-13**) was solved by Casanas *et al.* in 2008. Dimerization was also detected using chromatographic and analytical sedimentation equilibrium ultracentrifugation methods [34]. Previous studies have demonstrated that oligomerization of VP3 is important for proper capsid assembly in IBDV[145]. It has been suggested that VP3 dimers associate to form this higher oligomeric state [34]. The oligomerization domain has been mapped to a 24-amino-acid long region (residues 224-247) near the C-terminus which overlaps partially with the VP1 RNA polymerase binding domain [145]. However, VP3 variants lacking this region still form dimers suggesting that other regions on the protein is responsible for dimerization [34]. Protein VP3 consists exclusively of α helices connected by loops and is divided into two structural domains. Domain one consists of α helices 1-3 (α 1-3, **Figure 2-13A**, pink) and is connected to domain two (α 4-6, **Figure 2-13A**, red) by a long hinge region (**Figure 2-13A, yellow**). The overall tertiary structure is maintained mostly by hydrophobic interactions and the residues involved are conserved among birnaviruses. The second domain of VP3 shares structural homology with transcription factors sigma factor σ70 of *Thermus themophilus* (PDB:1IW7) [146] and Set2 SRI domain (PDB:2C5Z) [147]. Indeed, the C-terminal tail of VP3 in IBDV has been shown to remove structural blockage in the active site of VP1, the self-encoded RNA polymerase [124]. In addition, VP3 of IBDV aids in capsid assembly by 1) providing a scaffold for capsid polypeptide assembly[145], 2) recruiting RNA-dependent RNA polymerase (VP1) into the capsid[148], and 3) interacting with the dsRNA genome[149, 150]. A list of VP3 features, PDBs, sequence comparison, and sequence alignment can be found in **Tables 2-7, 2-8** and **Figure 2-14**.



PDB Code	Resolution(Å)	UniProt accession number	Reference
2R18	2.3	P25220 [29]	Casanas *et al.* [34]

Figure 2-13. Structure of VP3 in IBDV.

A) Each monomer of IBDV VP3 consists of two domains[34]. Alpha-helices 1-3 forms domain one (α 1-3, pink) and is followed by a long and flexible hinge region(yellow) that connects it to domain two which made up of α -helices 4-6 (α 4-6, red). **B)** The dimeric structure of IBDV VP3 with one monomer shown in red and the other in blue. Figure was prepared using Pymol[35].

Virus Name	UniProt accession number	Sequence length (AA.)	Size (KDa)	pI	PDB ID	Reference
BSNV	Q8AZM0	278	30.9	5.78		Da Costa *et al*. [25]
IPNV	Q703G9	238	27.0	6.54		Galloux *et al.* [27]
YAV	P89521	238	27.0	8.61		Suzuki *et al*. [28]
IBDV	P25220	257	28.8	6.05	1)2R18 [34] 2)2Z7J [34]	Bayliss *et al*. [29]
DXV	Q96724	309	35.4	6.64		Chung *et al*.[30]
TV-1	Q2PBR5	284	31.7	6.15		Nobiron *et al.* [31]

Table 2-7. Table of features for VP3.

The sequences were obtained from UniProt. Values for sequence length, size and pI were calculated using ProtParam [22, 23].

Identity(%) / Similarity(%)	BSNV	IPNV	YAV	IBDV	DXV
IPNV	30.1/39.7				
YAV	30.1/39.7	80.3/88.7			
IBDV	33.1/41.5	32.3/42.4	31.1/42.8		
DXV	16.2/29.1	12.3/23.3	12.3/21.4	14.2/27.2	
TV-1	20.1/29.4	17.2/26.8	18.9/28.9	18.6/25.8	16.4/27.0

Table 2-8. Sequence identity and similarity between different VP3s.

The protein sequence alignment was generated using ClustalW [32] and the sequence identities and similarities were calculated using the sequences identities and similarities webserver (URL:http://imed.med.ucm.es/Tools/sias.html) using the following equation:

$$\%\ \text{identity or similarity} = 100\left(\frac{\text{Identical positions}}{\text{Length of the alignment}}\right)$$

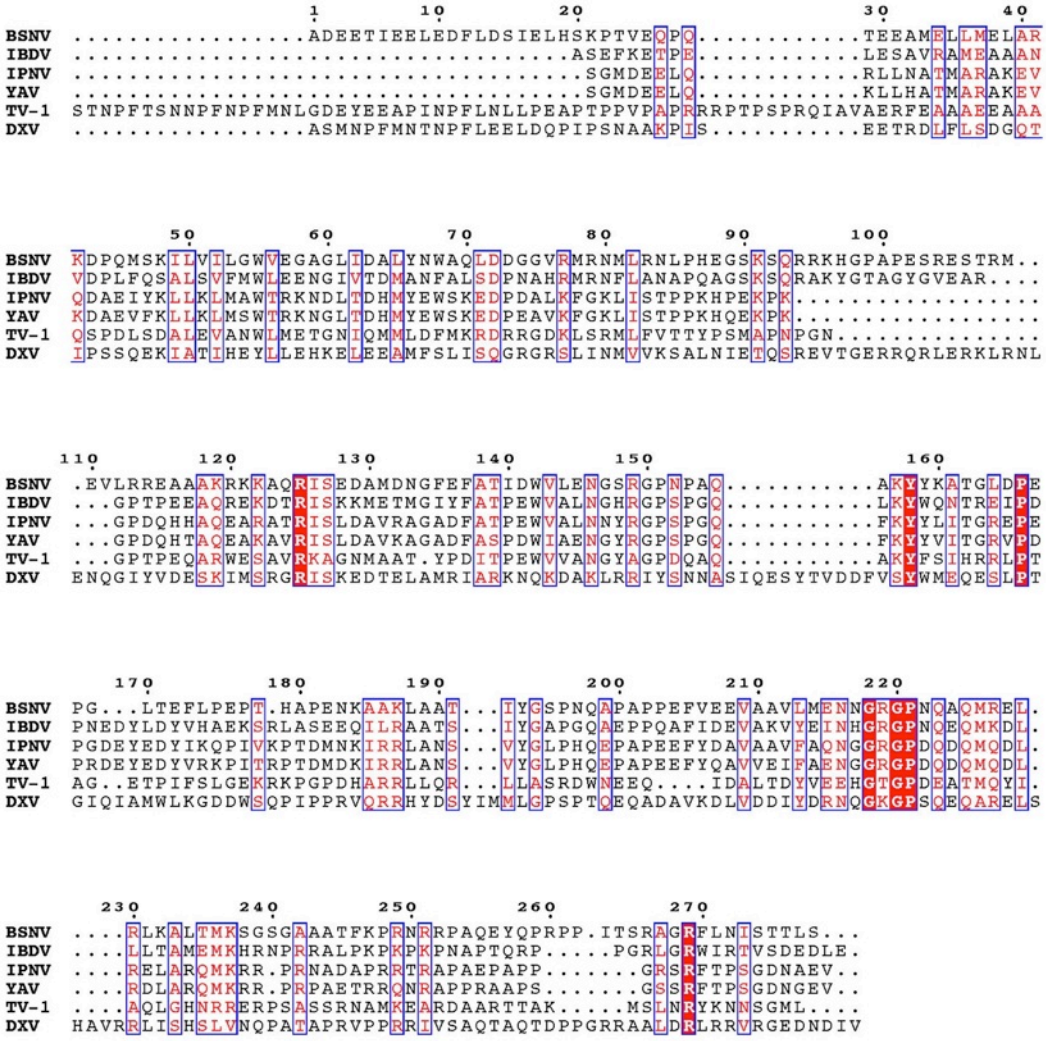


Figure 2-14. Protein sequence alignment of VP3.

The protein sequence alignment was done using ClustalW[32] and graphical output was generated using ESPript[33]. Residues that are conserved are shown as white text highlighted in red and semi-conserved residues are shown as red text in blue boxes.

2.5.4. Viral protein 4 (VP4)

Please refer to chapter 1 section 1.1 to 1.5 for description of VP4s.

2.5.5. Viral protein 5 (VP5)

An alternate ORF in segment A codes for VP5 in IPNV, IBDV, BSNV, DXV and YAV. Although they are named VP5, their amino acid sequence varies greatly (**Table 2-9, Figure 2-15**). The functions of VP5 in IPNV and IBDV have been characterized but VP5 from other members are yet to be determined. A list of VP5 features, PDBs, sequence comparison, and sequence alignment can be found in **Tables 2-9, 2-10** and **Figure 2-15**.

2.5.5.1. IPNV VP5

Protein VP5 in IPNV is a 17KDa anti-apoptotic protein [81]. Protein VP5 was able to prevent cell death caused by UV irradiation in CHSE cells. Consistent with this observation, VP5 was shown to enhance or maintain the functional half-life of survival factor Mcl-1 and control the expression of certain viral protein early in the replication cycle in CHSE cells. It contains four Bcl-2 homology (BH) domains (BH1-4) which were first identified on a membrane protein found in follicular lymphomas suffering from a t(14;18) chromosomal translocation [151]. Members of the Bcl-2 family can either promote or inhibit apoptosis via the intrinsic apoptosis pathway [152, 153]. Upon receiving a apoptotic stimulus, the pro-apoptotic BH3-only members (Bid, bad, and bim) interact with the anti-apoptotic members (Bcl-2, Bcl-XL) and neutralize their effects [154]. This triggers the oligomerization of multi-domain pro-apoptotic members (Bax/Bak) which permeabilize and form pores or voltage-dependent anion channel on the outer mitochondrial membrane causing the efflux of various apoptotic mediators. One of these mediators is cytochrome c, its release leads to the formation of apoptosome which activates the caspase protease cascade for apoptosis.

2.5.5.2. IBDV VP5

VP5 in IBDV is a highly basic protein of 17 KDa and is rich in cysteines [155]. The VP5 of IBDV was shown to interact with the voltage-dependent anion channel 2 (VDAC2) in the mitochondria to induce apoptosis in DF-1 (immortal chicken embryo fibroblast) cells [155]. Unlike the VP5 of IPNV, VP5 of IBDV lacks a Bcl-2 domain but instead carries a PEST motif [156]. This motif is rich in proline[P], glutamic acid[E], serine [S], and threonine[T] and is predominately found on short-lived proteins that are being targeted to the ubiquitin protein degradation pathway [157]. Interestingly, the anti-apoptotic Bcl-2 protein (Bcl-XL) was also shown to close VDAC on liposomes whereas the opposite is true for the pro-apoptotic Bcl-2 proteins (Bax and Bak) [158]. The opening of VDACs allows for the efflux of cytochrome c, a major apoptosis mediator. Indeed, the apoptosis induced by IBDV VP5 was shown to be accompanied by the release of cytochrome c [155, 159]

Virus name	UniProt accession number	Sequence length (AA.)	Size (KDa)	pI	Reference
BSNV	Q8AZL9	139	15.3	10.6	Da Costa *et al.* [25]
DXV	P0C748	237	27.1	10.5	Chung *et al.* [30]
IPNV	Q8JK01	148	17.5	10.3	Cutrin *et al.* [160]
YAV	Q7T402	133	15.6	9.6	Suzuki *et al.* [28]
IBDV	P0C751	145	16.5	6.8	Lombardo *et al.* [161]

Table 2-9. Table of features for VP5.

The sequences were obtained from UniProt. Values for sequence length, size and pI were calculated using ProtParam [22, 23].

Identity(%) / Similarity(%)	BSNV	DXV	IPNV	YAV
DXV	10.5/19.0			
IPNV	16.2/29.2	11.8/18.6		
YAV	11.3/21.8	11.4/17.3	75.0/80.4	
IBDV	12.3/21.9	11.4/21.1	23.7/34.9	24.7/34.9

Table 2-10. Sequence identity and similarity between different VP5s.

The protein sequence alignment was generated using ClustalW [32] and the sequence identities and similarities were calculated using the sequences identities and similarities webserver (URL:http://imed.med.ucm.es/Tools/sias.html) using the following equation:

$$\%\ \boldsymbol{identity\ or\ similarity} = \mathbf{100}\left(\frac{\boldsymbol{Identical\ positions}}{\boldsymbol{Length\ of\ the\ alignment}}\right)$$

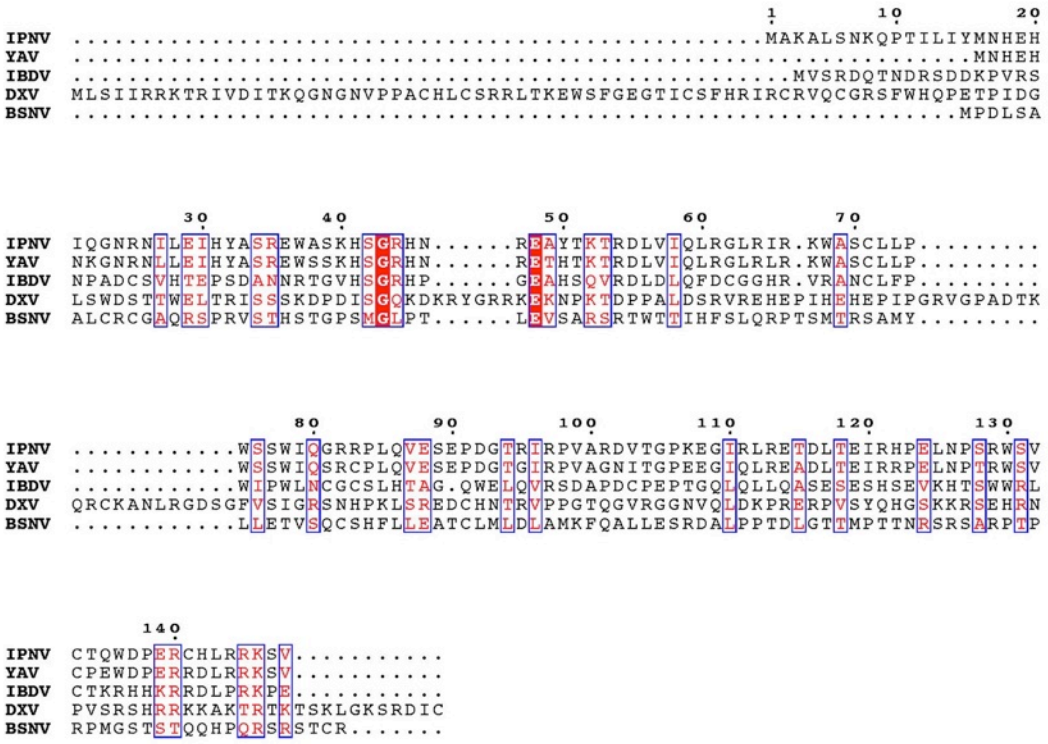


Figure 2-15. Protein sequence alignment of VP5.

The protein sequence alignment was done using ClustalW[32] and graphical output was generated using ESPript[33]. Residues that are conserved are shown as white text highlighted in red and semi-conserved residues are shown as red text in blue boxes.

3. Crystal structure of a viral protease intramolecular acyl-enzyme complex - insights into *cis*-cleavage at the VP4/VP3 junction of *Tellina* birnavirus.

The materials in this chapter were published in the Journal of Molecular Biology, Acta Crytallographica and a book chapter. The references are provided below.

1) **Crystal structure of a viral protease intramolecular acyl-enzyme complex. Insights into *cis*-cleavage at the VP4/VP3 junction of *Tellina* birnavirus.**
   Chung, I.Y. & Paetzel, M.
   J. Biol. Chem. 286, 12475-12482 (2011)

2) **Expression, purification and crystallization of VP4 protease from *Tellina virus 1.***
   Chung, I.Y. & Paetzel, M.
   Acta Crystallographica F 67, 157-160 (2011)

3) ***Tellina virus 1* VP4 peptidase.**
   Chung, I.Y. & Paetzel, M.
   Handbook of Proteolytic Enzymes 3nd Edition Elsevier Ltd. ISBN 978-0-12-382219-2 Chapter 780, pages: 3525-3529 (2012)

3.1. Introduction

As mentioned in chapter 1.3, TV-1 VP4 is more divergent from the rest of the birnavirus VP4s. Thus, a crystal structure of *Tellina virus* VP4 will give us a better picture of what structural features are being conserved. Another interesting question is how VP4 protease can cleave intramolecularly. A VP4 construct with an intact C-terminus might give us the answer. However, previous attempts failed to crystallize the full-length BSNV VP4 whereas IPNV VP4 has an internal cleavage site near the C-terminus and formed intermolecular acyl-enzyme complexes at this site. TV-1 VP4 was the next good candidate as the full-length construct expressed as a soluble protein and its C-terminus is long enough to reach its own serine nucleophile.

3.2. Materials and methods

3.2.1. 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'AGGCCCATATGGCCGACAGGCCCATGATC3'and the reverse primer has the sequence 5' CTTGTGAAGGCGGCCGCTCATGCTTGCGCCACGTTCTTTCCG GAGGAGAA3'. 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 twenty-one amino acids (MGSSHHHHHHSSGLVPR GSHM) that includes the 6x histidine-affinity tag and the linker regions. The theoretical molecular mass for this construct is 24,745 (233 amino acids, theoretical pI: 9.9).

3.2.2. Protein purification

Cells for over-expression were made by transforming the plasmids into *Escherichia coli* strain *Tuner (DE3)* followed by selection on LB plates with 0.05 mg/mL

kanamycin. For over-expression, a cell pellet from 100 mL of overnight culture grown in LB media supplemented with 0.05 mg/ml of kanamycin was used to inoculate each liter of M9 Minimal media. Four liters of cultures were grown. After shaking at 37^{o}C for 8 hours, 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 the culture was induced with 0.5 mL of 1M Isopropyl β-D-1-thiogalactopyranoside (IPTG) after 15 minutes. The induced cultures were allowed to grow overnight at 25^{o}C and were then harvested by centrifugation at 9,110 mg for 7 minutes. The cell pellet was stored at -80^{o}C for 15 minute 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 U/mL benzonase, and 0.2 mg/mL lysozyme) and incubated at 4^{o}C overnight with gentle agitation. The cell debris was removed by centrifugation at 28,964 xg and the clear supernatant was loaded onto a 5 mL nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity column (Qiagen) (**Figure 3-1**). A step gradient containing 100mM and 600mM of 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 were eluted stepwise with 100mM and 500mM NaCl in standard buffer. Fractions positive for VP4 were pooled and loading 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 the Pharmacia ÄKTA Prime^{TM} 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^{o}C and was then applied to the same size exclusion column mentioned above. Details of this limited proteolysis procedure are described elsewhere[162]. Purified VP4 was concentrated to ~ 40 mg/mL for crystallization trials.

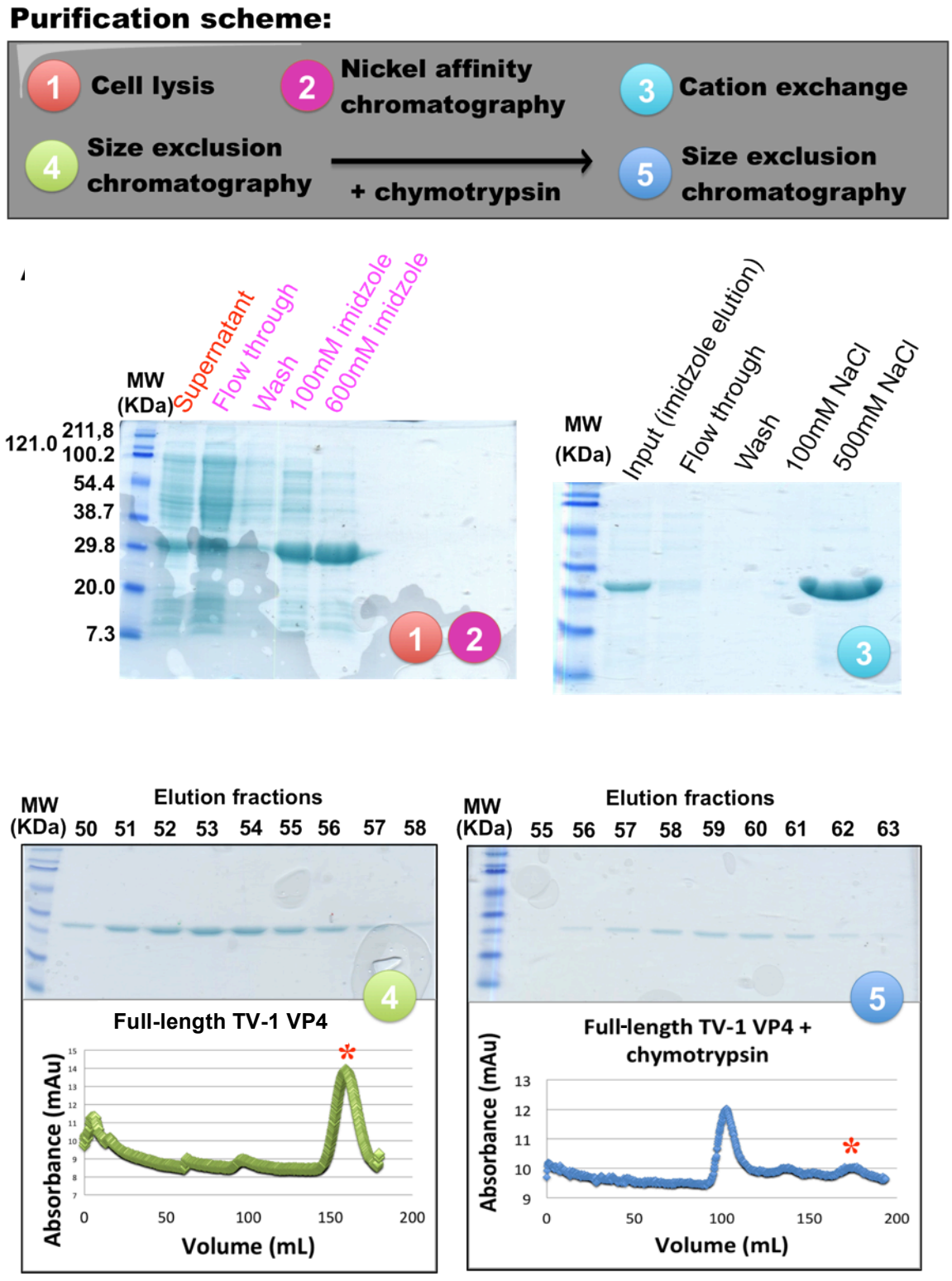


Figure 3-1. Purification scheme.

Figure 3-1. Purification scheme (Continued from last page).

The VP4 protease from TV-1 was purified from crude cell lysate (**Step 1**) using nickel column chromatography **(Step2)**. The protein was further purified using cation exchange chromatography (**Step 3**) and size exclusion chromatography (**Step 4**). Fractions positive for VP4 was then concentrated and subjected to limited proteolysis by chymotrypsin. A second size exclusion chromatography (**Step 5**) was preformed to remove chymotrypsin. The red stars on the size exclusion elution profile denote the VP4 peaks. The full-length TV-1 VP4 eluted at ~160 mL whereas the chymotrypsin-treated VP4 eluted at ~175 mL.

3.2.3. Crystallization

The crystal used for data collection was obtained using the hanging-drop method at room temperature (~23oC). On a coverslip, 1 µL of VP4 was mixed with 1 µL of reservoir reagent and 1µL of 0.2M urea as an additive (**Figure 3-2**). To aid in crystal nucleation, this drop was seeded with 1µL of selenomethionine-labelled TV-1 VP4 crystals from an older drop. The drop was allowed to reach vapour equilibrium via incubation over 1 mL of reservoir reagent in a grease-sealed chamber. The optimized reservoir condition was 21% PEG8000, 0.55M ammonium sulphate. The cryo-solution contained 70% of the buffer reservoir and 30% glycerol. Crystal for data collection was transferred into the cryo-solution, 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 x 59.1 x 208.1 Å with one molecule in the asymmetric unit, a Matthews coefficient of 2.1 and a solvent content of 42.1%. The pH of the drop is ~ 5.0 as determined using pH paper strips.

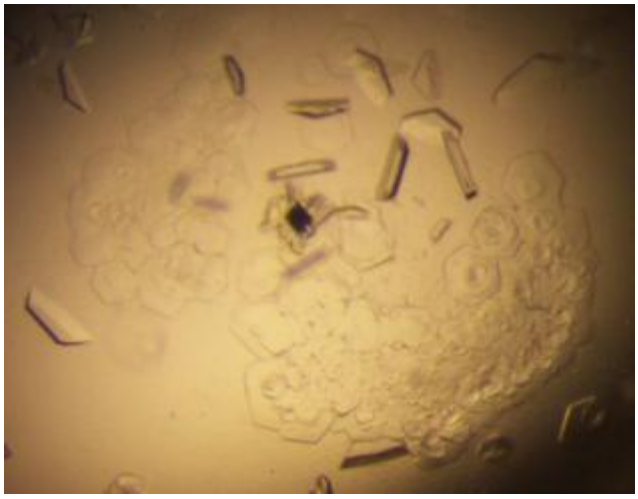


Figure 3-2. Crystals of TV-1 VP4.

The hexagonal crystals ($P6_422$) of TV-1 VP4 grew in the presence of 21% PEG8000, 0.55M ammonium sulphate with urea as an additive.

3.2.4. 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^o oscillations and each image was exposed for 5 sec. The distance between the crystal and the detector was 350mm.

3.2.5. Structure determination and refinement

The images were indexed and integrated using the program MOSFLM and scaled to 2.1 Å using SCALA[163]. The structure was solved by single-wavelength anomalous dispersion (SAD) (f': -7.11, f'': 4.77). The structure was solved using AutoSol from the Phenix program suite [164-166]. The coordinates from AutoSol were refined using rigid body and restrained refinement in REFMAC5[167]. The model from AutoSol was 90% complete and the remaining 10% of the molecule was built manually in coot [168] using a difference density map generated by REFMAC5. Restraints for the acyl-enzyme linkage were prepared using the REFMAC5 library definitions[169]. The final round of restrained refinement in REFMAC5 included TLS (5 TLS groups)[170, 171]. A table of crystallographic data statistics is shown in **Table 3-1.**

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	154167 (9938)
Unique reflections	13466 (1841)
R_{merge} [b]	0.107 (0.350)
Mean (*I*)/σ (*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 A.U.	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)	17.1
RMSD on angles (º)	1.979
RMSD on bonds (Å)	0.024

Table 3-1. Data collection, phasing and refinement statistics for TV-1 VP4 crystal

Table 3-1. Data collection, phasing and refinement statistics for TV-1 VP4 crystal. (Continued from last page).

[a] The data collection statistics in brackets are the values for the highest resolution shell.

[b] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle| \,/\, \sum_{hkl}\sum_i I_i\,(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl)\rangle$ is the mean intensity of that reflection.

[c] $R_{cryst} = \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big| \,/ \sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.

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

3.2.6. Structural analysis

Residue contact and hydrogen bond analysis was performed using the program CONTACT in the CCP4i crystallography software suite [163]. The root mean square deviation (RMSD) values reported for protein superpositions were calculated using the Superpose webserver[172]. The secondary structure assignments were performed using the STRIDE webserver[173]. The accessible surface area analysis was performed with Surface Racer 5.0[174].

3.2.7. Figure preparation

The structure figures were generated in Pymol[35]. The protein sequence alignment was prepared using the ClustalW [32] and ESPript webservers [33, 175]. The electron density map used in Figure 2-5 was converted from the mtz file generated in REFMAC5 using program FFT from the CCP4i crystallography software suite[163].

3.3. Results

3.3.1. Structure solution

The amino-terminus of TV-1 VP4 (residue 619) is defined by its X/VP4 cleavage site and its carboxy-terminus is defined by its VP4/VP3 cleavage site (residue 830) (**Figure 3-3**). We have successfully cloned, expressed and purified full-length wild-type TV-1 VP4 protease (residue 619-830) with an amino-terminal 6xHis 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 SAD phasing methods and refined the structure (residues 637-830) to 2.1 Å resolution.

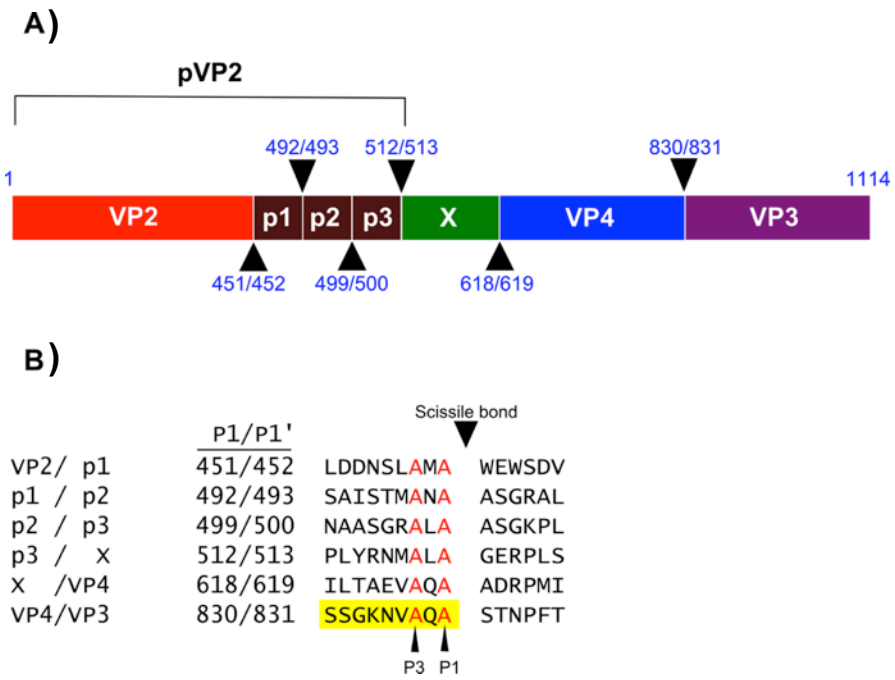


Figure 3-3. TV-1 VP4 protease cleavage sites

A) The genomic arrangement of TV-1 segment A. The arrow heads indicate the sites of cleavage with the P1 and P1' residue numbers listed. **B)** A list of known TV-1 VP4 cleavage sites [31]. The amino acids in the VP4/VP3 junction that are visible in the crystal structure are highlighted in yellow.

3.3.2. Overall protein architecture

The overall dimensions of TV-1 VP4 is approximately 38 Å x 34 Å x 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 the VP4 from other birnaviruses ranges from 4.7 to 6.6. TV-1 VP4 has an α/β fold consisting of sixteen β-strands (β1-β16), four α-helices (α1- α4) and two 3_{10} helices (η1- η2) (**Figure 3-4**). β-strands 1 to 9, 15, and 16 form two β-sheets that are predominantly arranged in an anti-parallel fashion. This β-sheet region houses the substrate binding groove and specificity pockets. One 3_{10}-helix and one α-helix (α1) are also found in this region. Adjacent to this β-sheet platform is a parallel β-sheet made up of three short strands (β10, β13 and β14). This β-sheet is surrounded by three α-helices (α2- α4) and a β-hairpin (β11 and β12). The nucleophilic Ser^{738} is located at the amino-terminal end of α2 and the general-base Lys^{777} is part of α3. The C-terminal residues (809-830) of TV-1 VP4 wrap around the perimeter of the molecule such that it ultimately arriving with Ala^{830} (the P1 residue for the VP4/VP3 cleavage site and last residue in the expressed construct) directly adjacent to the Ser^{738} nucleophile. The C-terminal residues pack tightly against the VP4 surface with β15 hydrogen bonding 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 (**Figure 3-4**).

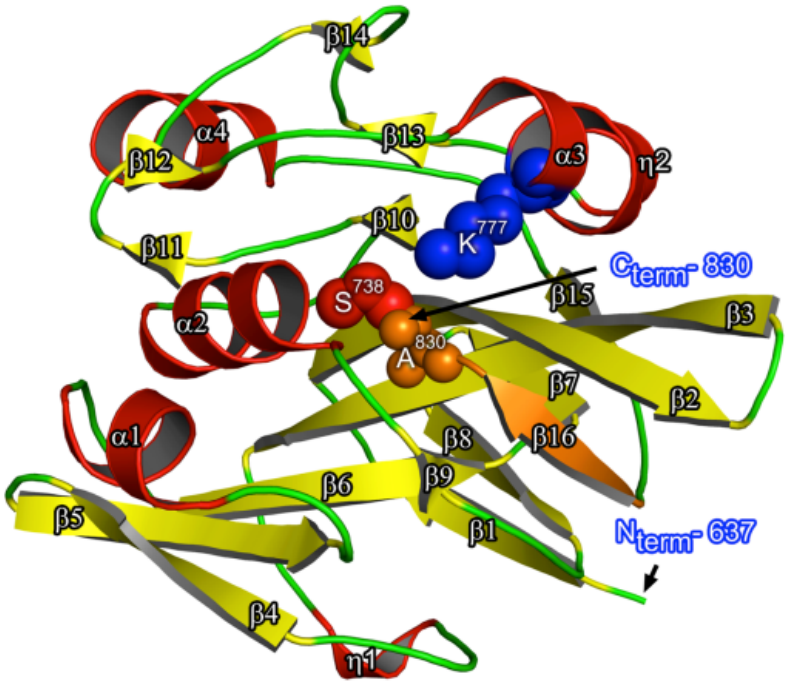


Figure 3-4. The TV-1 VP4 protease protein fold.

There are 4 α-helices (shown in red, α 1-4), 16 β-strands (shown in yellow, β1-15 and orange, β16) and two 3_{10} helix (shown in red, η1,2). The nucleophile, Ser^{738}, is shown as red spheres and the general base, Lys^{777}, is shown 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).

3.3.3. 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) (**Figure 3-5A**). The electron density at the active site therefore shows that this structure reveals 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 [176, 177].

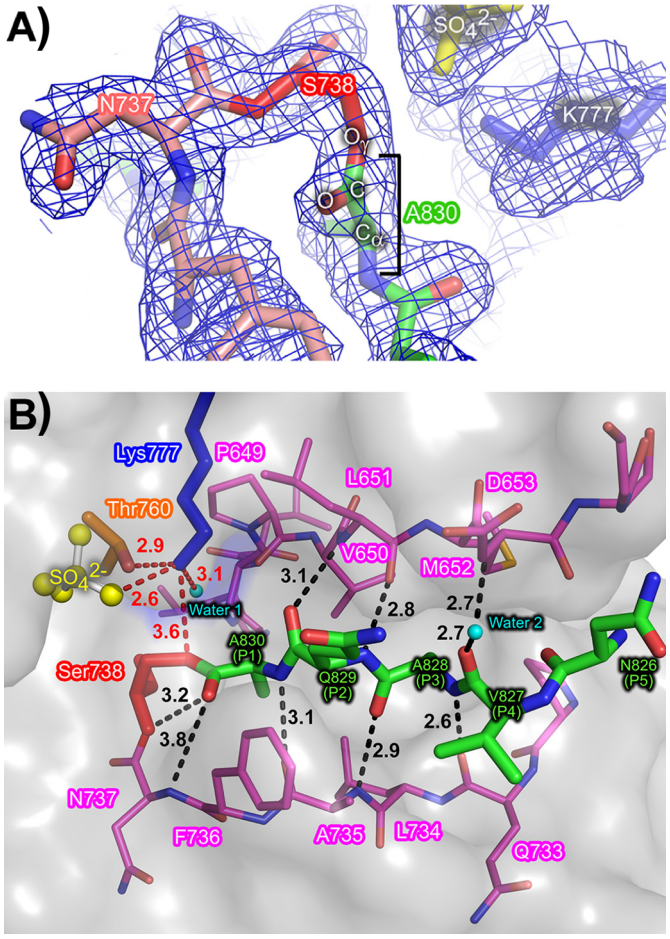


Figure 3-5. An intramolecular acyl-enzyme intermediate reveals the enzyme-substrate binding interaction in TV-1 VP4.

A) The $2F_o - F_c$ electron density map (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 sulphate 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 Oγ of the nucleophile, 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.

3.3.4. 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[39]) are stabilized by anti-parallel hydrogen bonding with residues 651 to 653 of β-strand 2 and by parallel hydrogen bonding with residues 733 to 737 of β-strand 9 (**Figure 3-5B**). 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} (**Figure 3-5B**). The shallow and uncharged S3 binding pocket which 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 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 (**Figure 3-10C**).

3.3.5. 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 previous mutagenesis study suggested that TV-1 VP4 utilizes a serine/lysine catalytic dyad. The proposed nucleophilic Oγ of Ser^{738} is oriented towards the Nζ of the proposed general base Lys^{777} (**Figure 3-5B**). However, the presence of a sulphate bound just adjacent to Lys^{777} has steered the Nζ of Lys^{777} slightly away from the Ser^{738} Oγ. This sulphate 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 order and barrier water (water 1) (**Figure 3-5B**). 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 other 10 lysine Nζ's within TV-1 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 amine groups of Ser^{738} and Asn^{737}.

3.4. Discussion

3.4.1. Crystallization strategies

The full-length N-terminally tagged TV-1 VP4 construct expressed in the soluble fraction and initially purified on a nickel affinity column successfully with most of the bacterial contaminants removed. However, size exclusion chromatographic analysis revealed that VP4 eluted in the void volume where proteins greater than 100 KDa would be expected to elute (**Figure 3-6**). Since the theoretical molecular mass of TV-1 VP4 is approximately 25 KDa, this suggested that TV-1 was either aggregated or in a homotetrameric state or larger. Crystallization screens with this protein did not produce any hits. To increase the purity of the sample, the protein eluted from the size exclusion column was applied to a cation exchange column. The sodium chloride eluted samples were pooled and subjected to size exclusion chromatography for a second time. The second size exclusion chromatography step revealed that TV-1 VP4 now eluted within a volume consistent with a monomeric nature (**Figure 3-6**). Initial attempt to crystallize the full-length monomeric TV-1 VP4 failed. Thus, limited-proteolysis was performed to remove flexible regions that might possibly be disrupting crystal packing. A variety of proteases and conditions were used. The chymotrypsin treated TV-1 VP4 gave a single major band on the SDS-PAGE gel after one hour (**Figure 3-6**, insert). Thus, chymotrypsin was chosen for proteolytic treatment of this protein which was then purified by size-exclusion chromatography yet again and concentrated and set up for crystallization screens. The chymotrypsin treated monomeric TV-1 VP4 yielded hexagonal crystals from Hampton Research Crystal Screen I condition 30 with 30% PEG 8000, 0.2 M ammonium sulphate and the optimized condition was 21% PEG8000, 0.55M ammonium sulphate. The reproducibility of these crystals was low with a success rate of around 5-10% for the native TV-1 VP4. The success rate was increased by seeding with crystals from an older drop. These crystals appeared around one week after crystal plating and can grow over months. The native crystal diffracted to 2.4 Å resolution using a rotating anode X-ray source. To obtain phase information, a batch of

selenomethionine-labelled TV-1 VP4 was purified. The selenomethionine-labelled VP4 crystallization drops were seeded with native crystals. The resulting selenomethionine-labelled VP4 crystals were more reproducible than the native crystals with a success rate of ~ 80-90%. To increase the size of these hexagonal crystals, the optimized condition was screen against the additive screen from Hampton Research. The selenomethionine-labelled crystals grew bigger in the presence of urea. These crystals appeared 2-3 days after crystal plating. However, these crystals dissolve in less than a week and require continuous crystal plating to provide the seeds for further crystallization trials.

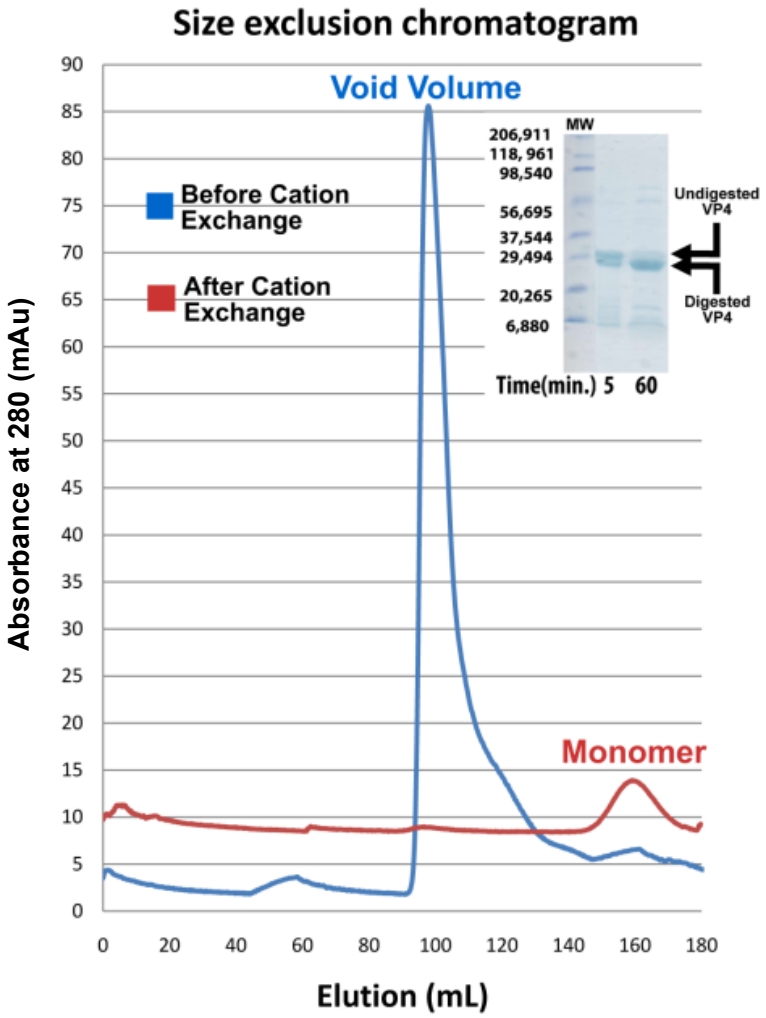


Figure 3-6. Size exclusion chromatography of VP4 before and after cation exchange.

The elution fractions collected from nickel-column were subjected to size exclusion chromatography (blue). The void volume fractions (label on the graph as void volume) were collected and pass through a cation exchange column. The elution fractions from cation exchange column were again subjected to size exclusion chromatography (red) and eluted as a monomer (label on graph as monomer). **Gel insert:** lane1, molecular weight marker; lane 2, TV-1 VP4 + chymotrypsin after 5 min; lane 3 - TV-1 VP4 + chymotrypsin after 1 hour.

3.4.2. 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 (**Figure 3-7**). In quantitative terms, TV-1 VP4 shares 20% sequence identity with BSNV VP4 and has a main chain superposition RMSD of 3.5 Å. In IPNV VP4, the sequence identity with TV-1 VP4 is only 12% and the main chain superposition RMSD is 7.1 Å.

A distinguishing feature of TV-1 VP4 is the presence of an α-helix between β-strand 5 and 6, a region that undertakes a loop structure in both IPNV and BSNV (**Figure 3-7**). The two β-strands (β15 and 16) that lead to the C-terminus of TV-1 VP4 form part of the main β-sheet region. In the intermolecular (*trans*) acyl-enzyme intermediate structure of IPNV (**Figure 3-7C**), the C-terminus is extended away from the structure and bound to an adjacent active site in a neighbouring VP4 molecule of the crystal, and an α-helix is seen at the N-terminus. The shorter construct of IPNV VP4 is lacking much of this region of the enzyme (**Figure 3-7D**). A unique characteristic of the BSNV VP4 is the long N-terminal β-strand that makes an intimate β-augmentation interaction with another VP4 molecule in the asymmetric unit (**Figure 3-7B**); however, this proposed dimerization has yet to be detected in solution.

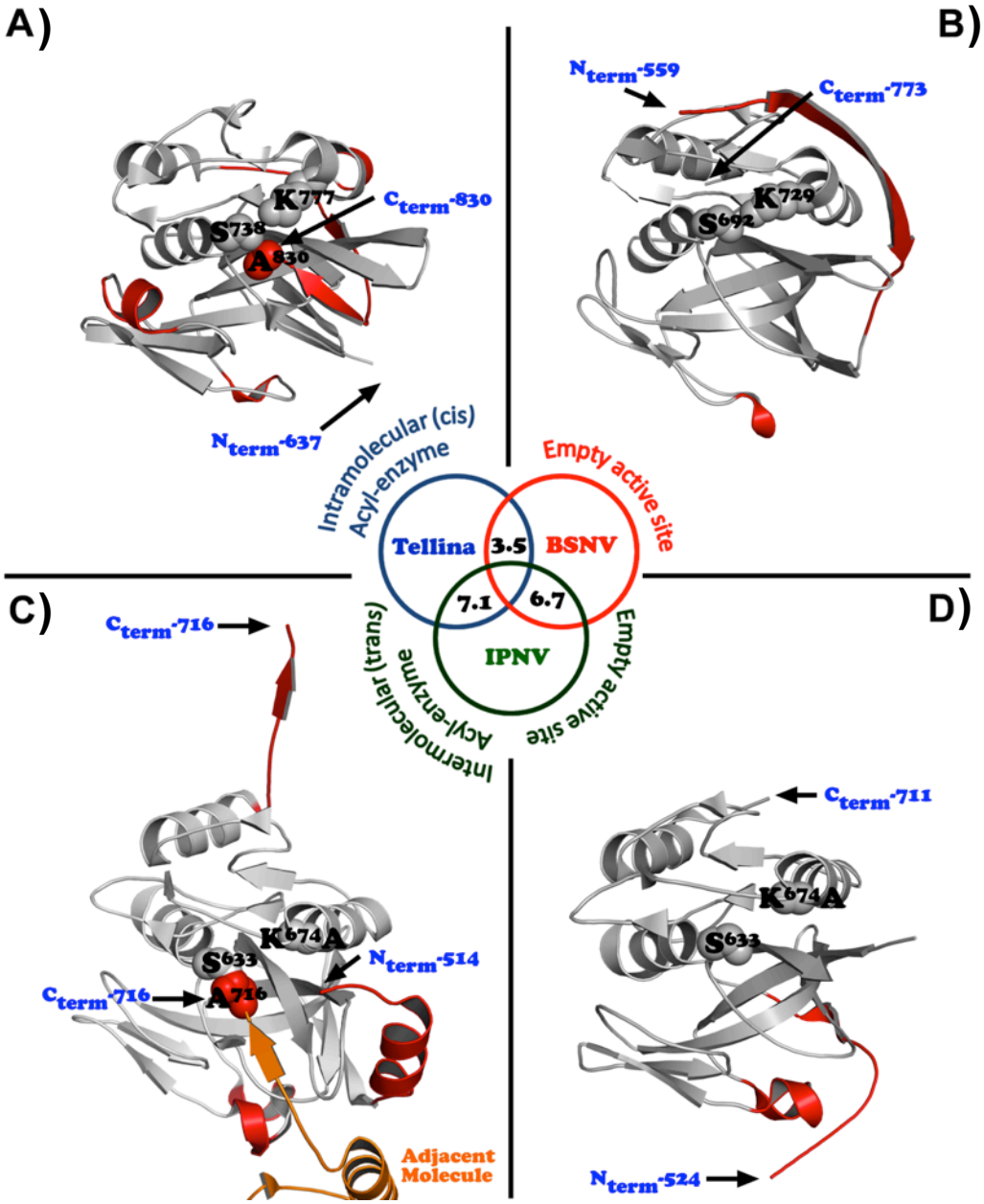


Figure 3-7. Comparison of VP4 structures.

Secondary structure diagram of **A)** TV-1 , **B)** BSNV (2GEF, chain B) [24],**C)** and **D)** IPNV VP4s(2PNL, chains C and D; 2PNM, chain A) [26]. All superimposable secondary structures are shown in grey, 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 grey spheres), and the last residue, Ala^{830} (shown as red spheres). In IPNV VP4 **(C)**, the serine nucleophile (shown as grey 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). The numbers in the overlapping segment of the circle denotes the main chain RMSD between VP4s.

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, **Figure 3-8**). Pro^{649}, Val^{650}, Ser^{738}, Thr^{760}, Gly761 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 (**Figure 3-9**).

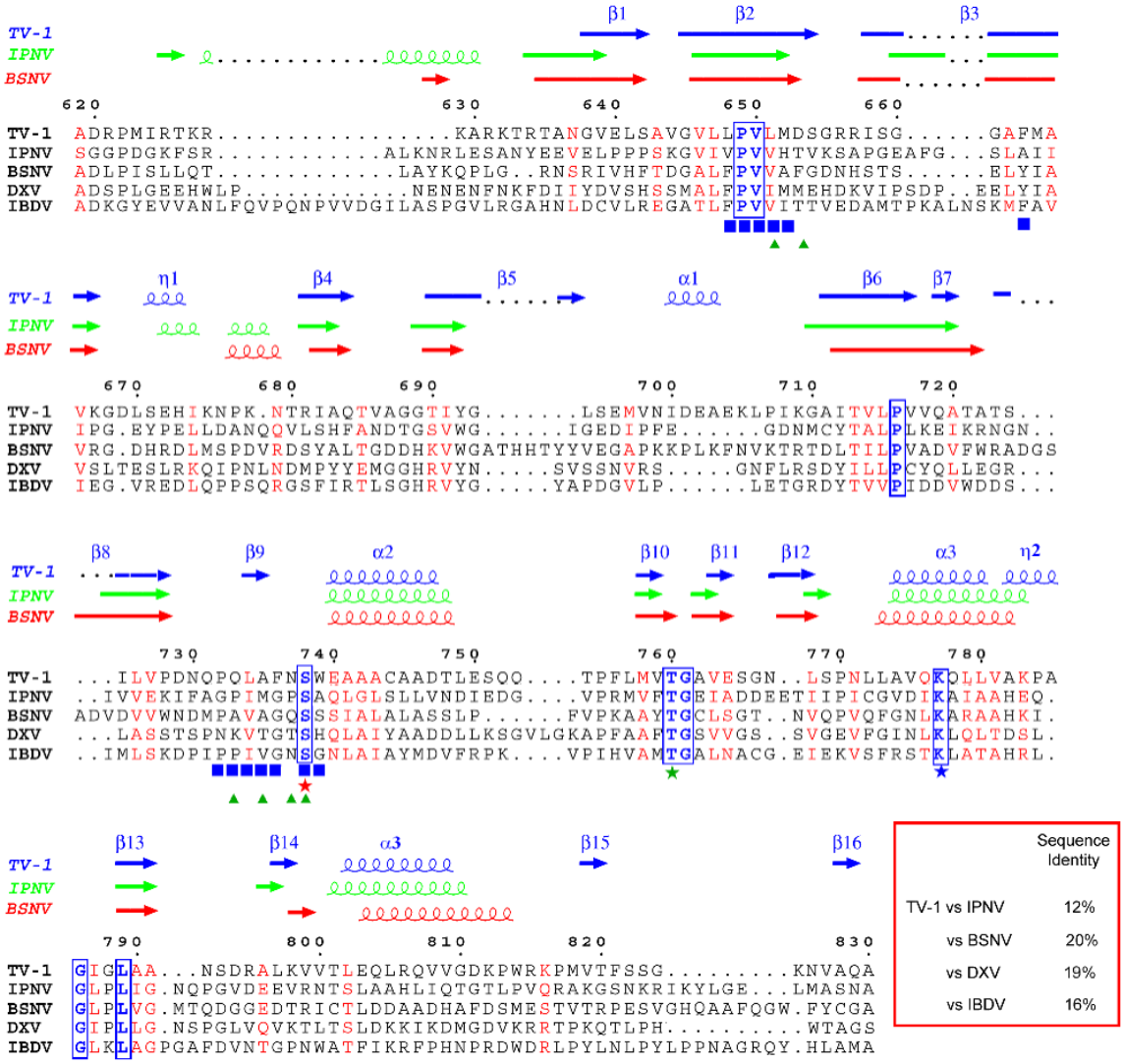


Figure 3-8. A sequence alignment of birnavirus VP4 proteases.

The VP4 sequences used for the alignment were obtained from the Swiss-Prot data bank (accession numbers are shown in parentheses): *Tellina virus 1*(TV-1; Q2PBR5)[31], *Infectious pancreatic necrosis virus* (IPNV;P90205)[27], *blotched snakehead virus* (BSNV, Q8AZM0)[25], *Drosophila X virus* (DXV; Q96724)[30], and *Infectious bursal disease virus* (IBDV; P15480)[29]. The secondary structure of TV-1 (blue), IPNV (green), and BSNV (red) are shown above the sequences with each element numbered sequentially. Residues that are involved in the catalytic mechanism are denoted by stars; nucleophile (Ser^{738}), red star; general base (Lys^{777}), blue star; general base-coordinating residue (Thr^{760}), green star. Invariant residues are coloured and boxed in blue. Residues that contribute to protease-substrate interactions are denoted by symbols (main chain substrate hydrogen bonding partners, green triangles; residues forming the S1 and S3 binding pockets, blue squares).

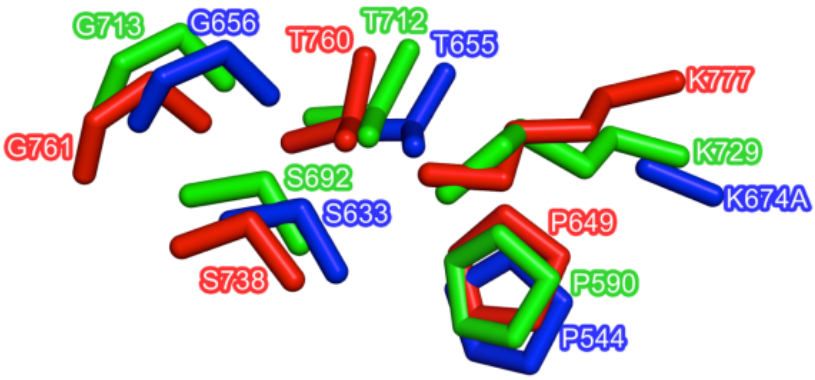
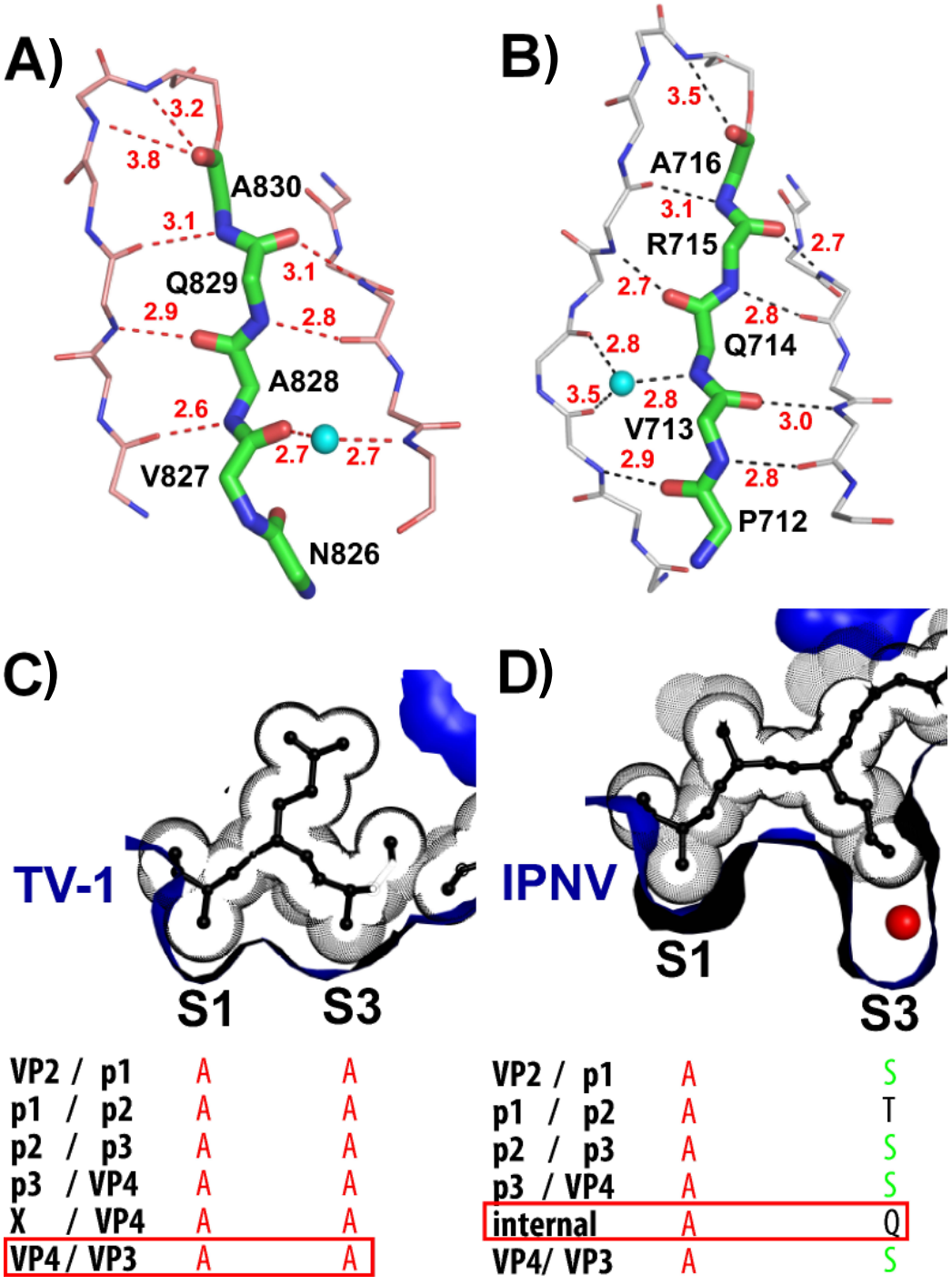


Figure 3-9. Conservation of the active site region in VP4 proteases.

Superposition of the invariant active site residues from TV-1 (red), IPNV (blue, 2PNL - chainA [26]) and BSNV VP4 proteases (green, 2GEF - chain A[24]) are shown.

3.4.3. Comparative analysis of TV-1 and IPNV 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 linkage (**Figure 3-7A**), 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 (**Figure 3-7C**). 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 (**Figure 3-10A, B**). Both substrate binding grooves utilize a water molecule within the interaction, but on opposite side of the substrate (**Figure 3-10A, B**).



VP2 / p1	A	A
p1 / p2	A	A
p2 / p3	A	A
p3 / VP4	A	A
X / VP4	A	A
VP4 / VP3	A	A

VP2 / p1	A	S
p1 / p2	A	T
p2 / p3	A	S
p3 / VP4	A	S
internal	A	Q
VP4/ VP3	A	S

Figure 3-10. Comparison of TV-1 and IPNV VP4 substrate binding groove and specificity binding pockets.

Figure 3-10. Comparison of TV-1 and IPNV VP4 substrate binding groove and specificity binding pockets (Continued from last page).

A) The TV-1 VP4 binding groove is shown with the main chain carbon atoms coloured in salmon. The main chain carbon atoms for the VP4/VP3 cleavage site (residues P1 to P5) are coloured in green. A 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 coloured 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 colour. **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 [26].

All of the TV-1 cleavage sites contain an alanine at the P1 and P3 positions (**Figure 3-10C**)[31]. 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 (**Figure 3-10D**)[43]. 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 (**Figure 3-10C**). 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 (**Figure 3-10D**). 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.

3.4.4. The catalytic machinery of TV-1 VP4

3.4.4.1. 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 nitrogens acting as hydrogen bond donors to the scissile bond oxyanion [178]. The TV-1 VP4 acyl-enzyme structure shows that the main chain amide 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 the oxyanion stabilization, but the distance and orientation for this interaction is less than optimal (**Figure 3-5B**). It is possible that the oxyanion hole is completely formed only during the transition state, when the oxyanion is present. Additional stabilization may be provided by the positive dipole of α-helix 2 which is oriented toward the oxyanion hole (**Figure 3-4**).

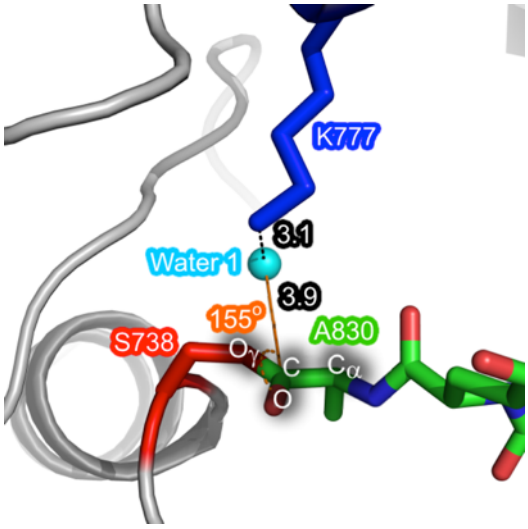


Figure 3-11. A potential deacylating water in TV-1 VP4 protease.

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 are labelled. 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 activating lysine general base is given along with the distance from the water to the ester carbonyl (in Å).

3.4.4.2. 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°)[179]. 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 (**Figure 3-11**). This is the only active site water seen in the correct position to serve as the deacylating water. It is

unlikely that the bound sulphate 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.

3.4.5. 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 [176, 180], (ii) small ester substrates using flash cooling [181, 182], (iii) small molecule inhibitors such as β-lactams [183, 184], and (iv) small peptide substrates using pseudo-steady state conditions [177].

We have found that viral proteases such as VP4 that 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 (**Figure 3-3**). Previously, Lee *et al.* used an active site mutation in the IPNV VP4 protease to trap an intermolecular (*trans*) acyl-enzyme intermediate (**Figure 3-7C**) [26]. 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 demonstrates directly 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 sulphate ion available from the optimized crystallization condition and bound adjacent to Lys^{777} has stabilized the protonated positively charged state on the ε-amino group of the lysine general base and thus preventing it from activating the deacylating water. The sulphate 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 sulphate binding site in TV-1 VP4, particularly the region corresponding the residues Asn^{771} to Leu^{773}, a loop that bridges β-strand 12 and α-helix 3 (**Figure 3-4**)**.** 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 the would-be sulphate binding site. Interestingly, the optimized crystallization conditions for the IPNV VP4 intermolecular (*trans*) acyl-enzyme structure also had sulphate present (0.4M Li_2SO_4) but no electron density was seen in the active site that would correspond to a sulphate ion. Since 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 sulphate 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 sulphate binding in that the IPNV VP4 was solved in the presence and absence of bound substrate yet in both cases the would-be sulphate binding site appears to be occluded by the side chain of Ile^{668} (**Figure 3-12**).

A low pH may have further stabilized the protonated state of the 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 much more basic (pH 8.5).

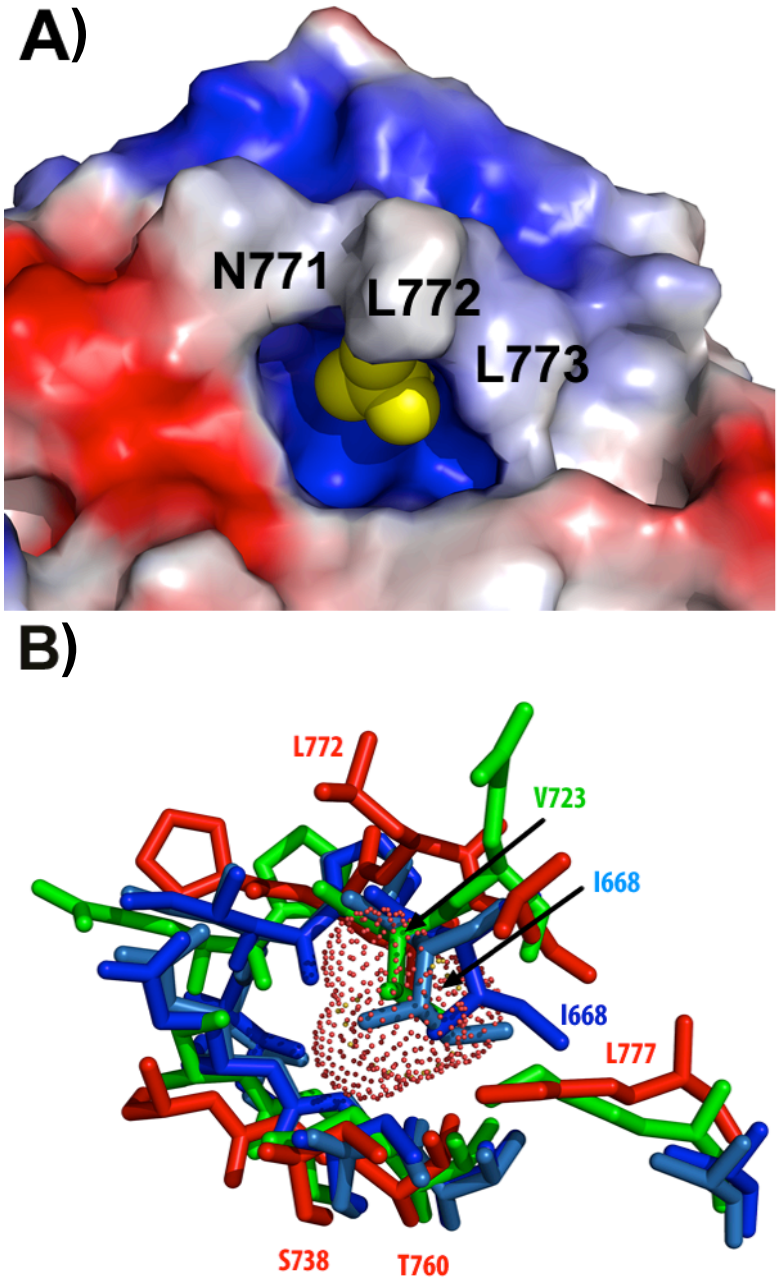


Figure 3-12. Sulphate binding site in TV-1VP4.

A) The molecular surface for the sulphate binding site in TV-1 VP4 is shown (average positive electrostatic surface is shown in blue, negative in red) with the sulphate ion rendered in van der Waals spheres (yellow). **B)** A superposition of VP4 structures (IPNV(blue, acyl-enzyme; light blue, empty active site)[26], BSNV (green)[24] and TV-1(red)) showing the residues that surround the TV-1 VP4 sulphate binding site. The position for the sulphate is shown as a dotted surface.

3.5. 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 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.

4. Crystal structures of *Yellowtail ascites virus* VP4 protease: trapping an internal cleavage site *trans* acyl-enzyme complex in a native Ser/Lys dyad active site

4.1. Introduction

Opposite to the situation of TV-1, the VP4 protease from YAV is highly homologous to that of IPNV (~80% identical). There exists an internal cleavage site in IPNV VP4 but equivalent site has yet to be confirmed in YAV. MCA-peptide and protein cleavage assays will be performed to determine if this site exists. Since IPNV VP4 formed acyl-enzyme complex at the internal cleavage site, crystallization of YAV VP4 is also a feasible strategy to answer this question. Previous mutagenesis studies on YAV VP4 showed that certain residues can alter polyprotein processing despite they are not in the active site. A structure of YAV VP4 will help to explain this phenomenon.

4.2. Materials and methods

4.2.1. YAV VP4 constructs

The cDNA for segment A of YAV was generously provided by Dr. Syunichirou Oshima. The full-length YAV VP4 construct (residues: 509-734, from the pVP2/VP4 cleavage site to the VP4/VP3 cleavage site) was amplified using a forward primer with the sequence: 5'- GGA CTC CAT GGC CAG CGG CAC AGA CAC TGG G -3' and a reverse primer with the sequence: (5'- CTG GCC TCG AGT GCA GTT GTT CTC ATT AGT TCC CC -3'). This DNA region was amplified using the polymerase chain reaction with Vent DNA polymerase (New England BioLab). The amplicon was cloned into the restriction sites *NcoI* and *XhoI* of plasmid pET28b^{+} (Novagen) using T4 DNA ligase (Fermentas). The ligation mix was transformed into NovaBlue *Escherichia coli* cells

(Novagen) for plasmid isolation. This creates a full-length VP4 construct that contains residues 509-734 from segment A of YAV, with two additional residues at the N-terminus [Met-Ala] and eight additional residues at the C-terminus: (Leu-Glu, and then a 6 x His affinity tag) (**Figure 4-1A**). The truncated YAV VP4 construct was amplified using the same forward primer as mentioned above along with a reverse primer with the following sequence 5'- CTG GCC TCG AGT GCT TTC TGC ACT GGT AGT GCT CC -3'. This DNA region was amplified using the polymerase chain reaction with Pfu DNA polymerase (Fermentus). The amplicon was cloned into the restriction sites *NcoI* and *XhoI* of plasmid pET28b^{+} (Novagen) using T4 DNA ligase (Fermentas). The ligation mix was transformed into NovaBlue *Escherichia coli* cells (Novagen) for plasmid isolation. This creates a VP4 that contains residue 509-716 from segment A of YAV, with the same additional residues at the termini as described for the full-length construct. To generate the active site mutant (K674A) constructs, a site-directed mutagenesis reaction, using Pfu DNA polymerase (Fermentus), was performed using the primers: 5'- TGC GGT GTA GAC ATC GCA GCC ATC GCC GCC CAT -3'and 5'- ATG GGC GGC GAT GGC TGC GAT GTC TAC ACC GCA -3'. The NCBI reference sequence for segment A of YAV indicates an asparagine at position 616 (a surface residue), but our sequencing indicates an aspartic acid residue at this position. We used site directed mutagenesis to change the sequence to match the reference sequence (Asn616). PfuUltra™ polymerase (Stratagene) along with the primers: 5'- AAA GAG ATC AAG AAG AAC GGA AAC ATC GTG GTG -3' and 5'- CAC CAC GAT GTT TCC GTT CTT CTT GAT CTC TTT -3' were used in the reaction. The sequences of all VP4 genes were verified by DNA sequencing (GenBank accession number NC_004168, UniProt accession number P89521).

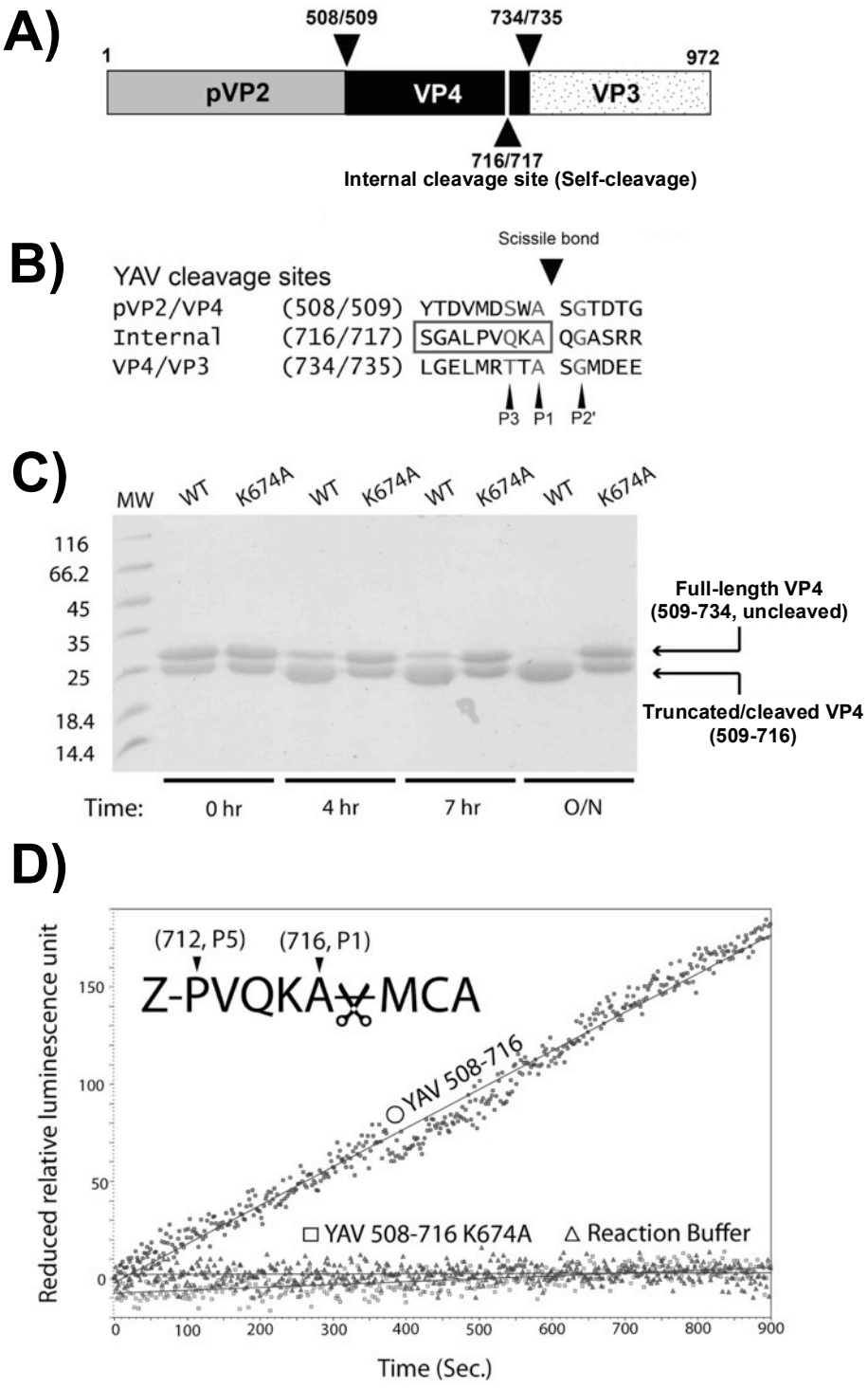


Figure 4-1. A schematic of YAV segment A polyprotein processing by VP4 protease and activity assays for truncated YAV VP4 (509-716).

Figure 4-1. A schematic of YAV segment A polyprotein processing by VP4 protease and activity assays for truncated YAV VP4 (509-716) (Continued from last page).

(A) The genomic arrangement of segment A in YAV is shown. The primary cleavage sites of cleavage are denoted by arrow heads with the P1 and P1' residue numbers listed. (**B)** A list of known major YAV VP4 protease cleavage sites. The amino acids preceding P1' in the internal cleavage site are visible in the crystal structure and are boxed in red. **(C)** To demonstrate cleavage at the internal cleavage site (after residue 716), the active truncated YAV VP4 (509-716) was incubated with full-length YAV VP4 lysine mutant (509-734 K674A) for 0, 4, 7 hours, and over-night at 23°C. As a control, the same experiment was repeated with an inactive truncated lysine mutant (509-716 K674A). A 20µL aliquot of the reaction was ran on a SDS-PAGE gel and was visualized by Coomassie stain. **(D)** A fluorometric peptide (Z-PVQKA-MCA) bearing residues P5-P1 of the VP4 internal cleavage site (residues 712-716) was incubated with active truncated YAV VP4 (○, 509-716), inactive truncated YAV VP4 (□, 509-716 K674A) and reaction buffer (Δ, 20mM MES pH6.5) alone.

4.2.2. Protein expression and purification

The expression vectors containing the YAV VP4 constructs were transformed into *Escherichia coli* strain *Tuner (DE3)* following by selection on Luria-Bertani (LB) agar plates supplemented with 50 μg/mL of kanamycin (KAN). Six liters of cultures were grown for each batch of protein purified. Ten milliliters of overnight culture was inoculated into each liter of LB/KAN media. The cultures were incubated at 37°C while shaking for 4 hrs before induction with 0.5 mL of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG). The induced cultures grew overnight at 25°C. Cells were harvested by centrifugation at 9,110 xg for 7 minutes. To facilitate cell lysis, the cell pellet was stored at -80°C for 15 minute. The frozen cell pellets was then completely re-suspended in lysis buffer (50 mM Tris-HCl buffer pH 8.0, 10% glycerol, 1 mM dithiothreitol (DTT), 7 mM magnesium acetate, 0.1% Triton X-100 and 1 U/mL benzonase). The cells were sonicated three times at 30% amplitude for 5 seconds with a 10 second rest between each of the pulses (Fisher Scientific Sonic Dismembrator Model 500). The sonciated cells were then lysed using an Avestin Emulsiflex-3C cell homogenizer (1250 psi for 3 minutes). Cell debris was removed by centrifugation at 28,964 x g for 20 minutes (**Figure 4-2**). The resulting supernatant was applied to a 5 mL nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity column (Qiagen). The column was equilibrated with 5 column volumes of standard buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 10% glycerol, and 1 mM DTT). A step gradient containing 100 mM, 300 mM and 600 mM of imidazole in standard buffer was used to elute the histidine-tagged protein. Fractions positive for VP4 were pooled, concentrated to 5 mL, and

loaded onto a size-exclusion chromatography 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 flow rate of the buffer was controlled by an ÄKTA Prime™ system (Pharmacia) set at 0.7 mL/min. Fractions with pure VP4 were pooled and concentrated to 30 mg/mL using a Millipore centrifugal filter with a molecular weight cut-off of 10 KDa for crystallization trials. Protein concentration was measured with a Nanodrop UV spectrophotometer (Themo Scientific) using an extinction coefficient of 9970 M^{-1} cm^{-1} as calculated by the Protparam server based on the amino acid sequence[22].

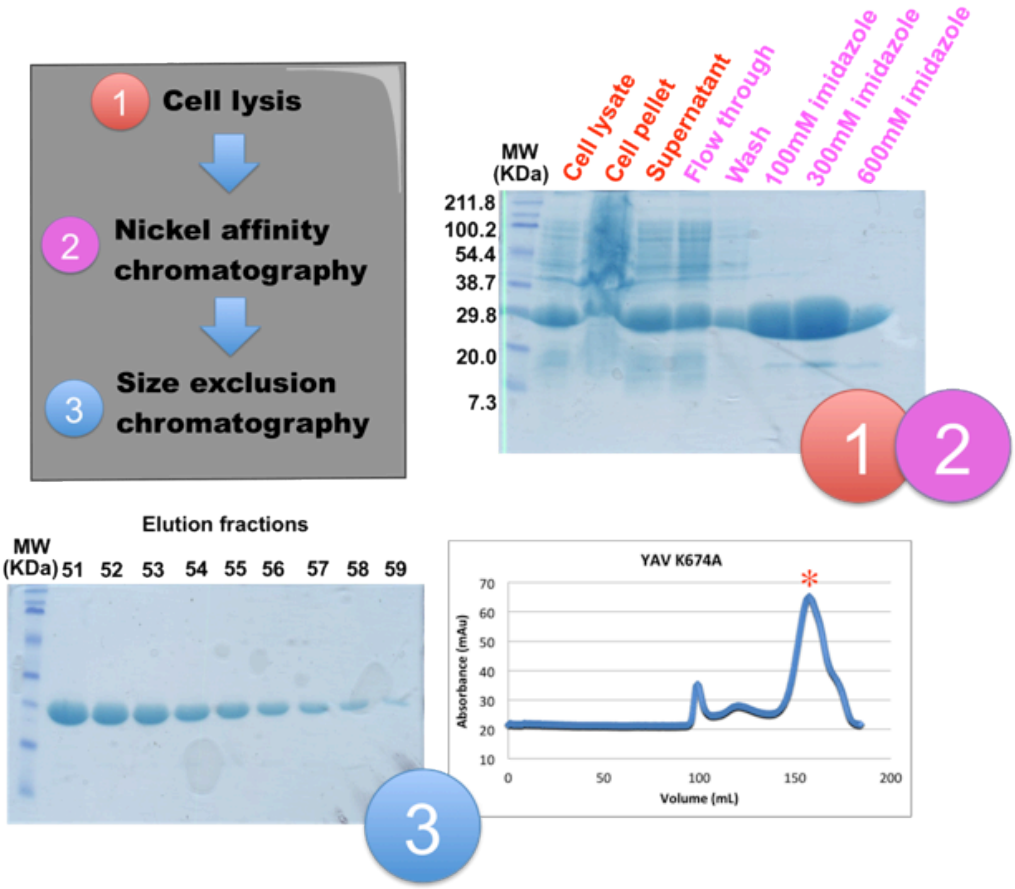


Figure 4-2. Purification scheme of YAV VP4.

All YAV VP4 constructs were purified using purification scheme shown above. The SDS-PAGEs and chromatogram for YAV VP4 K674A are presented for reference. The protein was purified from the crude cell lysate (1, red) using nickel affinity chromatography (2, magenta) followed by size exclusion chromatography (3, blue). The red star on the size exclusion elution profile denotes the peak of YAV VP4 K674A.

4.2.3. YAV VP4 full-length self-cleavage assay

1.5 µL of truncated VP4 (YAV VP4 509-716 or YAV VP4 509-716, K674A, 30 mg/mL) was added to 120 µL of the reaction buffer; 20 mM MES pH 6.5 (the same buffer and pH as the crystallization conditions) and 1.5 µL of full-length YAV VP4 (residues: 509-734, 30 mg/mL) with a mutant active site (K674A). The reaction was carried out at 23 °C and 20 µL aliquots we taken at times: 0, 4, 7 hours and overnight, mixed with an equal volume of 2X sample buffer and loaded onto a 15% SDS-PAGE gel which was visualized with Coomassie Blue stain.

4.2.4. Fluorometric peptide cleavage assay

The fluorometric peptide substrate Z-PVQKA-MCA (Z , benzyloxycarbonyl; MCA, 4-methylcoumaryl-7-amide) was synthesized by Lifetein. The peptide was dissolved in 100 % DMSO at a 50 mM stock concentration and every reaction had a final DMSO concentration of 1 % (v/v). Each 100 µL reaction contained 100 µM VP4 protease (YAV VP4 509-716 or YAV VP4 509-716, K674A) and 500 µM fluorometric peptide in reaction buffer (20 mM MES pH 6.5). The control reaction contained the fluorometric peptide in reaction buffer. The assay was run in triplicate on a 96 well µClear black plate from Greiner bio-one. The fluorescence (in relative luminescence units, RLU) was monitored at 37 °C using a SpectraMax® M5 Multi-Mode Microplate Reader (Molecular Devices) with an excitation wavelength at 380 nm and emission wavelength at 460 nm.

4.2.5. Crystallization

All crystallization trials were carried out at room temperature (~296K). The crystals used in the diffraction studies were grown using the sitting-drop vapour diffusion method. For the native active site VP4, a 1 µL protein sample (30 mg/mL) was mixed with 1 µL of reservoir reagent and allowed to reach vapour equilibrium with 1 mL of reservoir reagent in a tape-sealed chamber. The initial crystallization condition was obtained from the Hampton Research Crystal Screen I condition 23 (30% PEG 400, 0.1 M sodium HEPES pH 7.5 and 0.2 M magnesium chloride, **Figure 4-3A**). The optimized condition gave rod-shaped crystals that grew out of precipitation in 35% PEG2000, 0.1M MES pH6.5, and 0.30 M magnesium chloride **Figure 4-3B**. These monoclinic crystals

belonged to space group P1 2_1 1 and have unit cell dimensions of 41.6 x 64.3 x 187.7 Å (β = 95.8°) with five molecules in the asymmetric unit (43% solvent, Matthews coefficient of 2.14). For the active site mutant, 1μL of the native VP4 (30 mg/mL) was added to a 30 μL aliquot of the active site mutant (30 mg/mL) immediately prior to crystal plating. Three μL of this VP4 mixture was mixed with 3 μL of the mother liquor and plated as described for the native active site VP4. The crystals used in diffraction study for the active site mutant grew out of 25% PEG 2000, 0.1 M MES pH 6.5, and 0.45 M magnesium chloride(**Figure 4-3C**). These cubic crystals had unit cell dimensions of 273.5 x 273.5 x 273.5 Å with two molecules in the asymmetric unit (solvent content 73%, Matthews coefficient of 4.54) and belonged to space group F 4_1 3 2.

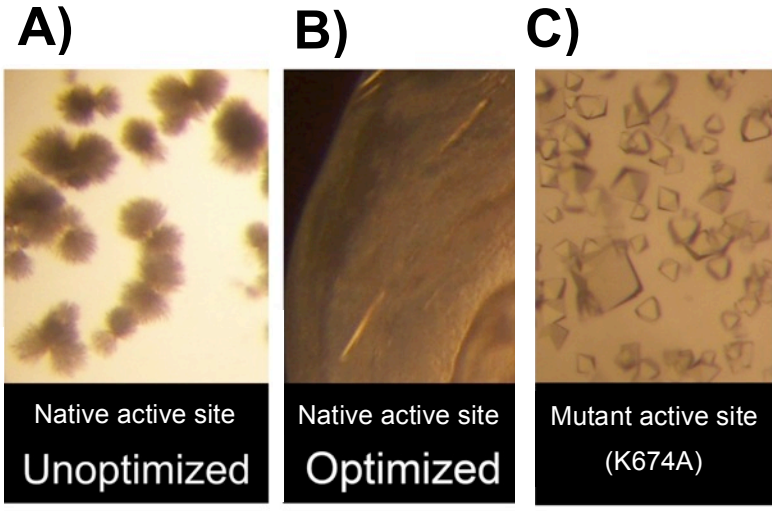


Figure 4-3. Crystals of YAV VP4.

A) Crystals of native active site VP4 from the initial hit. **B)** Crystals of native active site VP4 grew in the condition used for diffraction study. **C)** Crystals of lysine mutant (K674A) grew in the condition used for diffraction study.

4.2.6. Data collection

Although the two crystals forms grew out of very similar crystallizations conditions, the cryo-solvent conditions for the two crystals were significantly different. The cryo-solution for the native active site VP4 was 30% 2-Methyl-2,4-pentanediol (MPD) in mother liquor (35% PEG2000, 0.1M MES pH6.5, and 0.3 M magnesium

chloride), whereas 20 % glycerol in the mother liquor (25% PEG 2000, 0.1 M MES pH 6.5, and 0.45 M magnesium chloride) was used for the active site mutant. The crystal of the native active site VP4 was transferred into cryo-solution immediately prior to mounting onto a loop with cooper base. Diffraction data was collected at the macromolecular X-ray diffraction data collection facility in the Molecular Biology and Biochemistry Department of Simon Fraser University which includes a MicroMax-007 Microfocus X-ray generator, Osmic Confocal VariMax High Flux optics, an R-AXIS IV++ image plate and an X-stream2000 cryosystem. The program CrystalClear was used for data collection. The crystal-to-detector distance was 220 mm. The crystal was exposed for 3 minutes with an oscillation angle of 0.5° and an X-ray wavelength of 1.5418 Å. The crystal of the active site mutant was transferred into cryo-solution and immediately mounted on a loop with cooper base then flash-cooled with liquid nitrogen prior to transport to the Canadian Light Source (CLS) for data collection at beamline 08ID-1. The crystal-to-detector distance was 310 mm. The crystal was exposed for 6.4 seconds with an oscillation angle of 0.4° and an X-ray wavelength of 0.97949 Å (**Table 4-1**).

4.2.7. Structure solution and refinement

The indexing and integration for both datasets were carried out using MOSFLM [185] and scaling was performed with SCALA [186] from the CCP4i suite [163]. The program POINTLESS was used to confirm space group assignment [186]. Molecular replacement was employed to obtain phase estimates. For the native active site VP4, the solution was found using the program MOLREP[187] utilizing the IPNV VP4 (chain A of pdb: 2PNM) as the search model. Rigid body, and restrained refinement were performed using REFMAC5[167, 188]. For the active site mutant, the coordinates from the native active site VP4 were used as a molecular replacement search model in the program PHASER[166]. Rigid body and restrained refinements were performed using REFMAC5[167, 188]. Editing and fitting of atomic coordinates were performed using the program Coot[168].

	Native active site PDB: 3R0B	Mutant active site (K674A) PDB: 4HHC
Crystal parameters		
Space group	P1 2_1 1	F 4_1 3 2
a,b,c (Å)	41.6, 64.3, 187.7	273.5, 273.5, 273.5
α, β, γ (º)	90.0, 95.8, 90.0	
Data collection statistics		
Wavelength (Å)	1.5418	0.9795
Resolution (Å)	64.3 – 2.5(2.6 – 2.5)[a]	55.8 – 2.3(2.4 – 2.3)[a]
Total reflections	122204 (13215)	343613 (45664)
Unique reflections	33009 (4043)	39388 (5637)
R_{merge} [b]	0.088 (0.278)	0.105 (0.249)
R_{pim} [c] (all I+ & I-)	0.052 (0.178)	0.037 (0.093)
Mean (*I*)/σ (*I*)	10.5 (3.8)	13.8 (6.5)
Completeness (%)	95.9 (80.7)	100.0 (100.0)
Redundancy	3.7 (3.3)	8.7 (8.1)
Refinement statistics		
Protein molecules (chains) in A.U.	5	2
Residues	1014	405
Water molecules	281	354
Total number of atoms	7920	3376
R_{cryst} [d] / R_{free} [e] (%)	17.1 / 24.7	17.1/ 19.3
Average B-factor (Å^2) (all atoms)	25.3	35.5
RMSD on angles (º)	1.454	1.742
RMSD on bonds (Å)	0.015	0.014

Table 4-1. Data collection and refinement statistics for YAV VP4 crystals.

[a] The data collection statistics in brackets are the values for the highest resolution shell.

[b] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle| \,/\, \sum_{hkl}\sum_i I_i\,(hkl)$, 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_{pim}(precision-indicating merge) $= \sum_{hkl}(1/n_{hkl} - 1)^{1/2}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where n is the number of observed hkl reflections.

[d] $R_{cryst} = \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big| \,/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.

[e] R_{free} is calculated using 5% of the reflections randomly excluded from refinement.

4.2.8. Structural analysis

The STRIDE server was used to assign secondary structure elements [173]. The protein-protein interactions we analyzed using the PROTORP server [189]. The channel diameter was calculated using the MOLE server [35, 189-191].

4.3. Results

4.3.1. Self-cleavage at the YAV VP4 internal cleavage site

Based on sequence analysis it was previously proposed that YAV VP4 protease (residues 509-734 of the VP2-VP4-VP3 segment A polyprotein, **Figure 4-1 A, B**) would contain an internal cleavage site (between residues 716/717) near its C-terminus, in addition to its primary cleavage sites at 508/509 and 734/735[45]. Consistent with this proposal, we have observed that a truncated YAV VP4 construct (residues 509-716) with a C-terminal 6xHis-tag self-cleaved the tag upon incubation.

The full-length wild-type YAV VP4 enzymes (residues: 509-734) was not possible to isolate due to low level expression. However the same construct with the active site mutation K674A was purified in high yield. Therefore, we were able to test the ability of a truncated YAV VP4 with a native active site to cleave the full-length YAV VP4 protein at the internal cleavage site in a *trans* fashion. We found that the truncated YAV VP4 (509-716) was active and fully processed the full-length YAV VP4 (residues: 509-734, K674A) after an over-night reaction (**Figure 4-1C**). In contrast, no processing was observed when an active site mutant (K674A) version of YAV VP4 509-716 was used in the reaction (**Figure 4-1C**).

In addition, we used a fluorometric peptide assay to show that the YAV VP4 truncated at the internal cleavage site (509-716) is active. The peptide sequence used corresponded to the residues: PVQKA, which are residues 712 to 716 of the YAV polyprotein (P5 to P1 of the VP4 internal cleavage site). This peptide is modified at the C-terminus with a 4-methylcoumaryl-7-amide group (MCA). Upon cleavage of the amide bond between the MCA group and C-terminal residue, the fluorophore 7-amino-4-methylcoumain (AMC) is released which results in a fluorescent emission signal at 470

nm when excited at 380 nm [192]. We found that the YAV VP4 509-716 enzyme was active and that it recognized and cleaved at the internal cleavage site sequence. In contrast, the VP4 509-716 K674A active site mutant or the reaction buffer controls showed no activity (**Figure 4-1D**).

4.3.2. Self-cleavage at the YAV VP4 internal cleavage site promotes crystallization

As described above, the native active site version of YAV VP4 with a C-terminal affinity tag slowly cleaves itself after Ala716, the P1 residue for the internal cleavage site near the C-terminus of VP4. The resulting processed protein produces monoclinic crystals with five molecules in the asymmetric unit that diffracted to 2.5Å resolution (**Figure 4-4A**). An active site mutant version (K674A) of YAV VP4 produces cubic crystals with two molecules in the asymmetric unit that diffracted to 2.3Å resolution (**Figure 4-4B**). Interestingly, crystals of the active site mutant only appear when a small amount of native VP4 is added to promote cleavage at Ala716. The final refined structures display electron density for residues 515 to 716 in each molecule of the asymmetric unit.

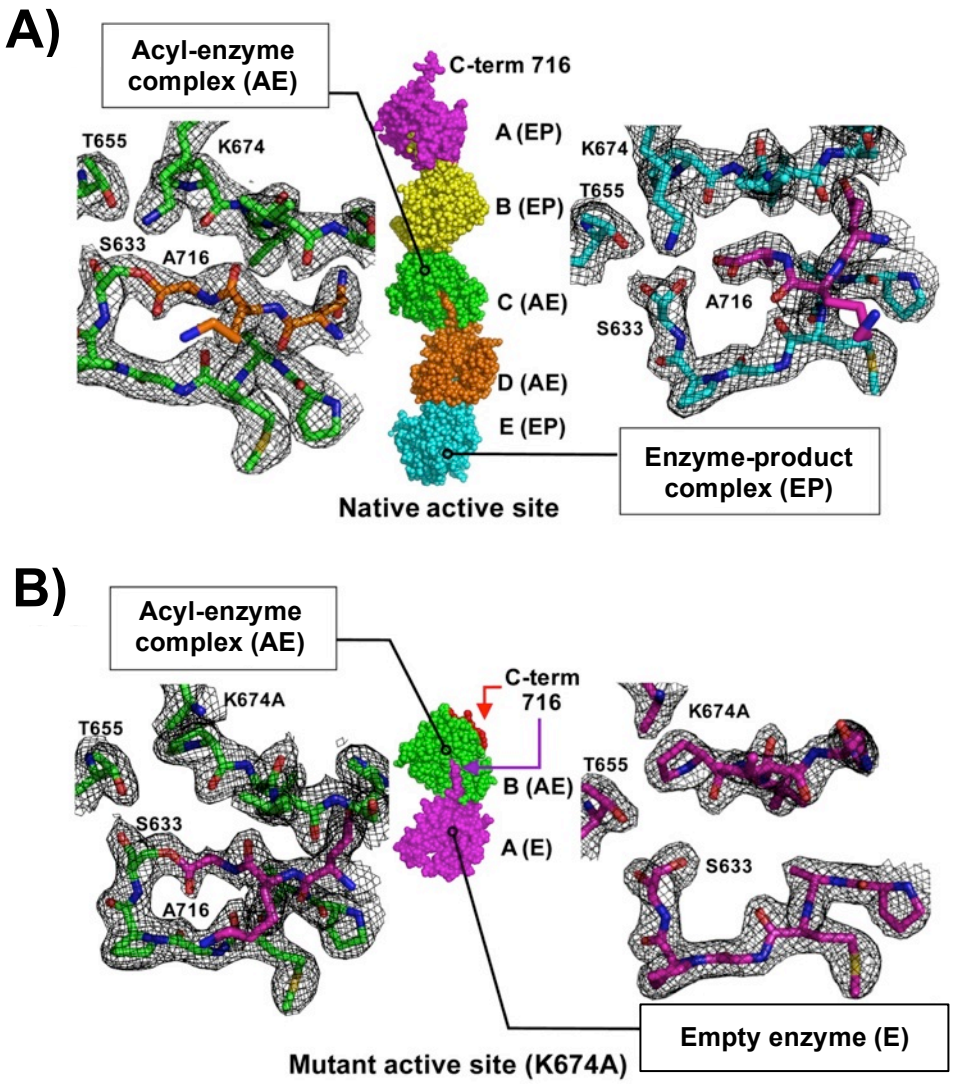


Figure 4-4. Intermolecular (trans) acyl-enzyme complex, product complex and empty active site.

Figure 4-4. Intermolecular (trans) acyl-enzyme complex, product complex and empty active site (continued from last page).

YAV VP4 protease is able to cleave at an internal cleavage site (after Ala^{716}) near its own C-terminus (Ala^{734}, the VP4/VP3 junction). Two crystal structures of YAV VP4 (with and without an active site mutation) reveals the internal cleavage site bound within the active site of a neighbouring VP4 molecule. Analysis of the electron density within each molecule of the asymmetric unit reveals three different enzyme states of the protease reaction cycle (empty active site, acyl-enzyme and product complex). The packing of YAV VP4 protease in the asymmetric unit for each structure is shown along with representative $2F_o$ - F_c electron density maps (contoured at 1 sigma) at the active site for representative types of complexes. **(A)** In the monoclinic crystal of native active site YAV VP4, five molecules of VP4 are in the asymmetric unit with the C-terminus of one molecule bound in the active site of its neighbour. The C-terminal carbonyl carbon of molecules D and E form an intermolecular acyl-enzyme complex (AE) with the active site nucleophilic serine Og of molecules C and D respectively. An enzyme-product (EP) complex is formed between molecule pairs E/A, A/B, and B/C. **(B)** In the cubic crystal of mutant YAV VP4 (Lys^{674} general base mutated to alanine) two molecules are in the asymmetric unit. The C-terminal carbonyl carbon of molecule A forms an intermolecular acyl-enzyme complex (AE) with the nucleophilic serine Og of molecule B. The active site of molecule A remains in the unbound state (E) because the C-terminus of molecule A (shown as red spheres) folded back onto itself instead of in an extended conformation.

4.3.3. Overall architecture and active site

Consistent with the theoretical isoelectric point of 4.9, the surface of YAV VP4 is predominantly negatively charged. The protease has a α/β protein fold consisting of thirteen β-strands, four α-helices and two 3_{10}-helices (**Figure 4-5A**). β-strands 1-7 form a platform on which the binding pockets for substrate recognition are constructed. Most of these β-strands interact in an anti-parallel fashion. YAV VP4 is unique in that it has a disulfide bridge (Cys^{588}-Cys^{604}) that links β-strand 5 and 6 (**Figure 4-5A**). No disulfide bonds are observed in the other VP4 proteases. Two 3_{10}-helices ($\eta1$, $\eta2$) reside on a loop between $\beta3$ and $\beta4$. Strands $\beta8$, $\beta11$, and $\beta12$ form a parallel β-sheet flanked by three α-helices ($\alpha2$, $\alpha3$, $\alpha4$) and a β-hairpin ($\beta9$, $\beta10$). This region hosted the catalytic dyad with the serine nucleophile (Ser^{633}) found immediately preceding α-helices 2 ($\alpha2$) and the lysine general base (Lys^{674}) arriving from α-helices 3 ($\alpha3$). The Lys^{674} side chain has clear electron density and is very ordered within the five molecules of the wild-type active site structure. The Nζ for Lys^{674} has an average B-factor of 20.2 $Å^2$. The Nζ for Lys^{674} is almost completely buried; its average accessible surface area is 3.4 $Å^2$. The average accessible surface area for the 9 other lysine Nζ atoms in chain A is 52.2 $Å^2$. The Nζ of Lys^{674} is coordinated by the O$\gamma1$ of Thr^{655} which resided between β-strands 8

and 9, the Oγ of Ser^{633} (the nucleophile), the main chain carbonyl oxygen of Cys^{669}, the carboxylate oxygen of Ala^{716} (C-terminal residue, for the product bound complexes), and a buried water (a proposed deacylating water)(**Figure 4-6**). For the acyl-enzyme structures (chain C and D, **Figure 4-6**) the ester carbonyl oxygen (Ala716 O) is pointed towards the oxyanion hole (Ser633 NH), away from the general base lysine Nζ. The buried water, which we propose is the deacylating or catalytic water, is coordinated by: Thr^{655}Oγ1, Ser^{633}Oγ, Lys^{674}Nζ, and Pro^{544}O (**Figure 4-6**).

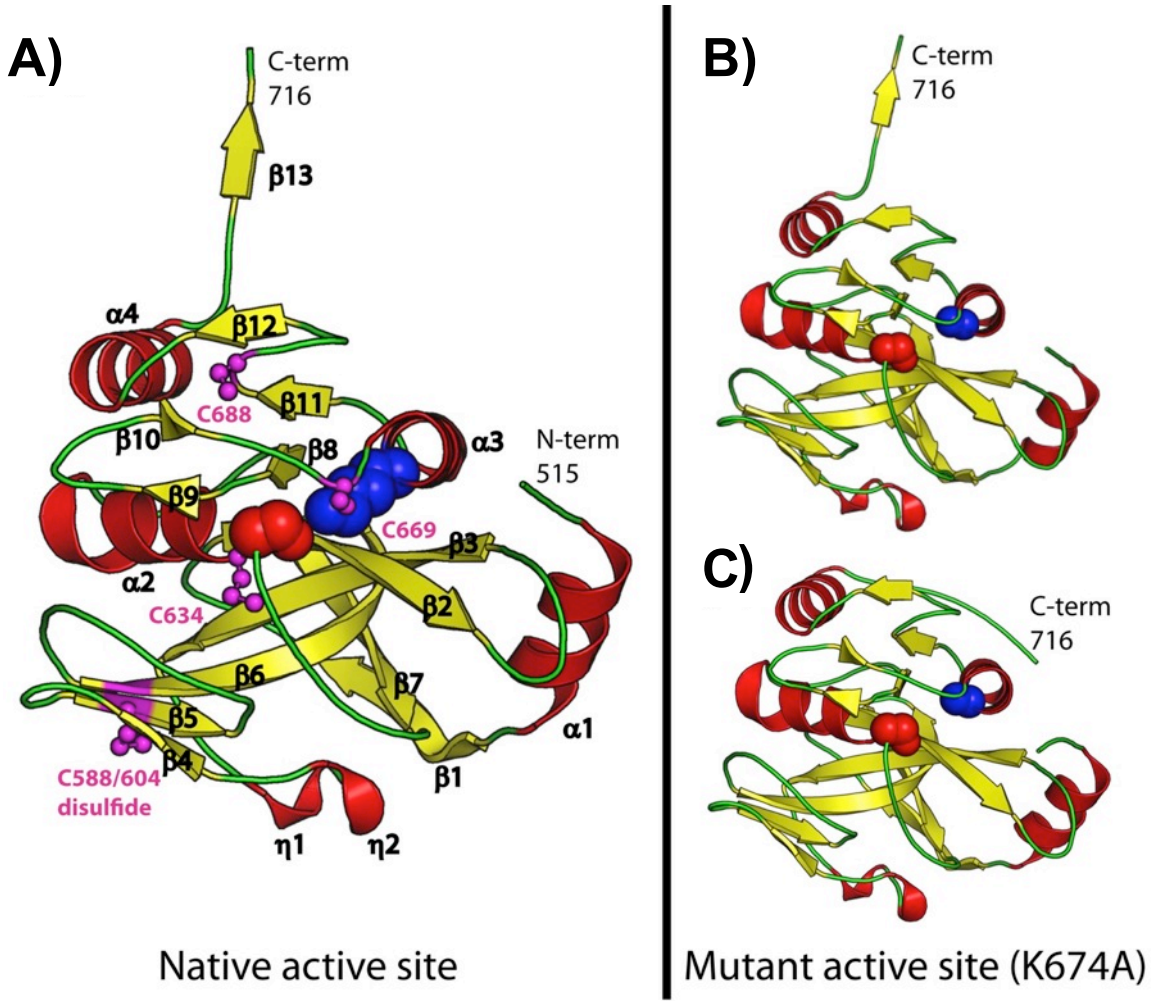


Figure 4-5. The YAV VP4 protein fold.

The protein fold for a representative protein chain in the native active site structure. (**A**) The native active site structure is shown in (**A**). The β-strands are depicted as yellow arrows and α-helices as red coils. The catalytic active site residues Ser^{633} (red) and Lys^{674} (blue) are shown as spheres. The cysteine residues are shown in magenta ball-and-stick. (**B**) The protein fold for the two different VP4 conformations observed in the mutant active site (K674) structure. The C-terminus, the main difference in the structures, is labeled.

4.3.4. Different C-terminal conformations

In the native active site structure, each protein chain in the asymmetric unit has a C-terminus that is extended (**Figure 4-4A, 4-5A)**. In contrast, the active site mutant (K674A) structure reveals two separate C-terminal conformations, one that is extended and another conformation where the C-terminus packs against the surface of the enzyme (**Figure 4-4B, 4-5B)**.

4.3.5. Trans acyl-enzyme and enzyme-product complexes in a native active site

In the asymmetric unit of the native active site structure, each substrate binding groove is occupied by the last six residues (residues 711-716) of the adjacent molecule, thus stabilizing the extended conformation for each C-terminus (**Figure 4-4A, 4-5A**). The average buried surface area between the enzyme and the neighbouring molecules is 894 $Å^2$. Analysis of the electron density reveals that two of the five VP4 molecules in the asymmetric unit form an acyl-enzyme complex with the neighbouring VP4 molecule (**Figure 4-4A).** The $O\gamma$ of the serine nucleophile (Ser^{633}) is covalently linked to the carbonyl carbon of Ala^{716}, the P1 residue for the internal cleavage site. The electron density is consistent with a trigonal planar geometry that is expected for an ester bond (**Figure 4-4A**).

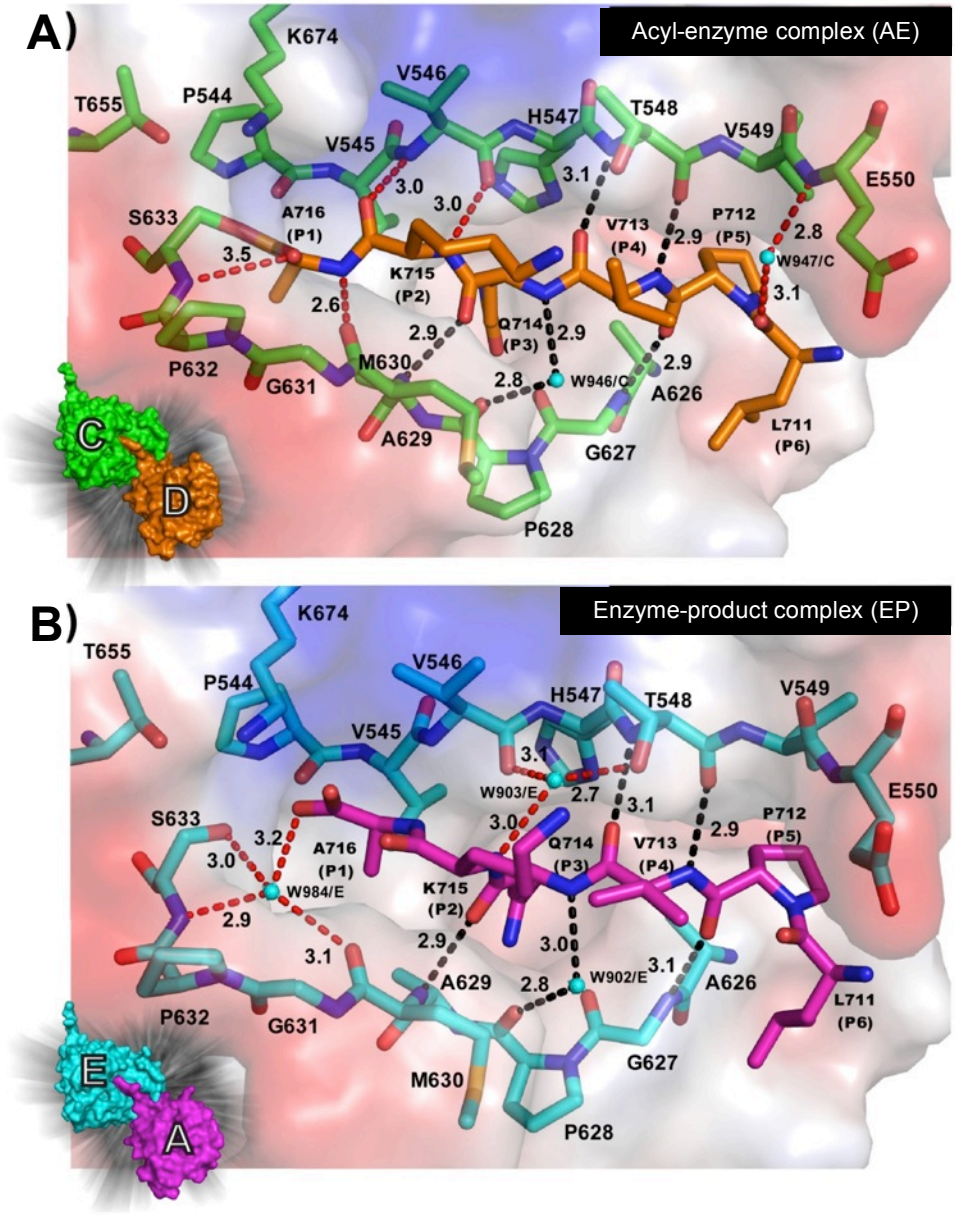


Figure 4-6. Binding groove and active site interactions seen in trans acyl-enzyme complexes and trans enzyme-product complexes within a native active site of YAV VP4 protease.

(A) The acyl-enzyme complex is shown with the substrate coloured in orange (molecule D) and the enzyme coloured in green (molecule C). **(B)** The enzyme-product complex is shown with the product coloured in magenta (molecule A) and the enzyme coloured in cyan (molecule E). The surface diagrams for the respective molecules are shown at the left-hand corner. Water molecules are shown as cyan spheres. The hydrogen bonds are shown as dashed lines. The black coloured hydrogen bonds represent those that are conserved in each molecule in the asymmetric unit. The red coloured hydrogen bonds are those that vary from molecule to molecule in the asymmetric unit, depending on the acylation state. The hydrogen bonding distances are given in ångströms.

4.3.5.1. Hydrogen bonding network.

The P6-P1 residues (711-716) are stabilized by hydrogen bonding interactions in an anti-parallel fashion with residues 546-550 on one side of the binding groove, and interact in a parallel fashion with residue 627-630 on the other side of the binding groove (**Figure 4-6**). Five of the hydrogen bonds involved in linking the C-terminus of one molecule to the binding groove of the neighbouring molecule are conserved in both the native active structure and the mutant active site structure (**Figure 4-6, 4-7**). These are: $Thr^{548}O$-$Val^{713}N$, $Thr^{548}N$-$Val^{713}O$, $Gly^{627}N$-$Pro^{712}O$, $Pro^{628}O$-$Gln^{714}N$ (via water), and $Met^{630}N$-$Gln^{714}O$ (depicted in black dotted lines in **Figure 4-6**). The acyl-enzyme complexes reveal hydrogen-bonding interactions between the P2 and P1 residues of the internal cleavage site with residues that line the binding groove near the catalytic residues. These hydrogen bonds are: $Ala^{716}O$-$Ser^{633}N$, $Ala^{716}N$-$Met^{630}O$, $Lys^{715}O$-$Val^{546}N$, and $Lys^{715}N$-$Val^{546}O$ (**Figure 4-6**). The hydrogen bond between $Ala^{716}O$ (P1 residue) and $Ser^{633}N$ (main chain nitrogen of nucleophile) is the one hydrogen bond observed in the oxyanion hole. There are no other hydrogen-bond donors near the Ala^{716} carbonyl oxygen. It is possible that the positive dipole of α-helix 2 (where Ser^{633} resides) contributes to the oxyanion stabilization.

Several water molecules participated in the hydrogen bonding network within the substrate binding groove. The water that bridges the main chain amide nitrogen of Gln^{714} (P3) to the carbonyl oxygen of Pro^{628} (**Figure 4-6, 4-7**) is seen in each of the observed reaction steps. A water that bridges $Glu^{550}N$ to $Leu^{711}O$ is only observed in the acyl-enzyme complexes. A water molecule bridges the hydrogen bond between the $O\gamma$ of Thr^{548} to the carbonyl oxygen of Val^{546} and the amide nitrogen of Lys^{715}, in some of the acyl-enzyme and enzyme-product complexes (**Figure 4-6**). The most interesting water seen in the enzyme-product complexes is the one that resides within the oxyanion hole (**Figure 4-6B**). This water, which is displaced in the acyl-enzyme complex, makes hydrogen bonding interactions with the carboxylate oxygen of Ala^{716} (P1 residue), the $O\gamma$ and main chain nitrogen of Ser^{633} (nucleophile), and the main chain carbonyl oxygen of Met^{630} (**Figure 4-6B**).

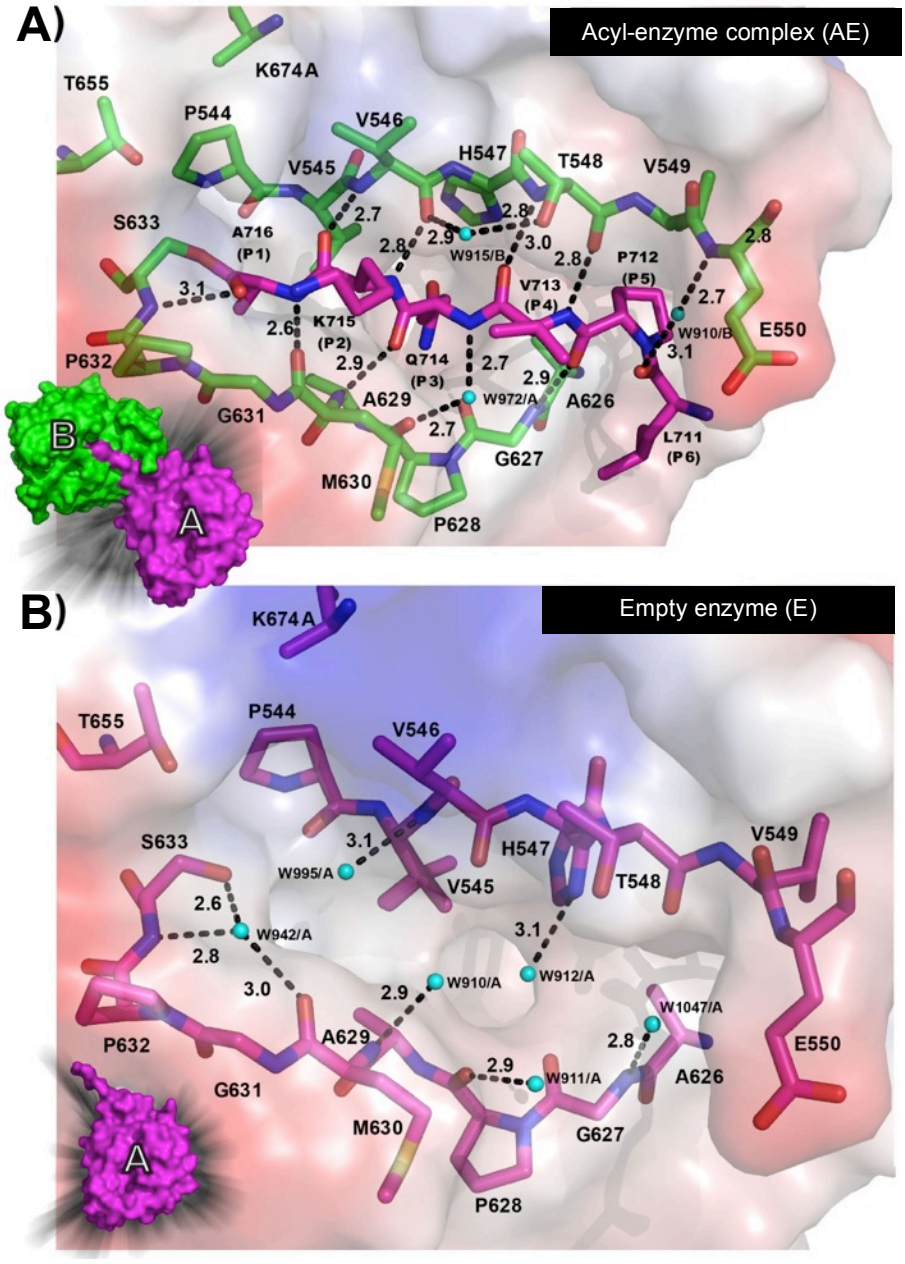


Figure 4-7. YAV VP4 protease structure with a mutant active site (K674A) reveals a comparison of acyl-enzyme complex and empty substrate binding groove.

(A) The acyl-enzyme complex is shown with the substrate coloured in magenta (molecule A) and the enzyme coloured in green (molecule B). **(B)** The empty-enzyme structure (molecule A, magenta) is shown. The surface diagrams for the respective molecules are shown at the left-hand corner. Water molecules are shown as cyan spheres. The hydrogen-bonds are shown as black dashed lines. The hydrogen bonding distances are given in ångströms.

4.3.5.2. Substrate specificity pockets.

The major substrate specificity pockets are the S1 and S3 (**Figure 4-6, 4-7**). Residues: Ile^{543}, Pro^{544}, Val^{545}, Leu^{561}, Ile^{563}, Ile^{592}, Ala^{629}, Met^{630}, Gly^{631}, Pro^{632}, Ser^{633} and Cys^{634} form a continuous surface (the S1 pocket) for the binding of the P1 residue side chain (Ala^{716}). Residues: Val^{545}, His^{547}, Ser^{559}, Leu^{560}, Leu^{561}, Gln^{576}, Gly^{591}, Ala^{607}, Leu^{608}, Pro^{609}, Val^{624}, Phe^{625}, Ala^{626}, Gly^{627}, Pro^{628}, and Ala^{629} form a continuous surface (the S3 pocket) for the binding of the P3 residue side chain (Gln^{714}). A cleft that is open on one side accommodated the P5 residue (Pro^{712}). The side chain of Leu^{711} (the P6 residue) points into a shallow pocket near the end of the substrate binding groove. There are no binding pockets for the P2 and P4 residues as their side chains are pointing away from the substrate binding groove.

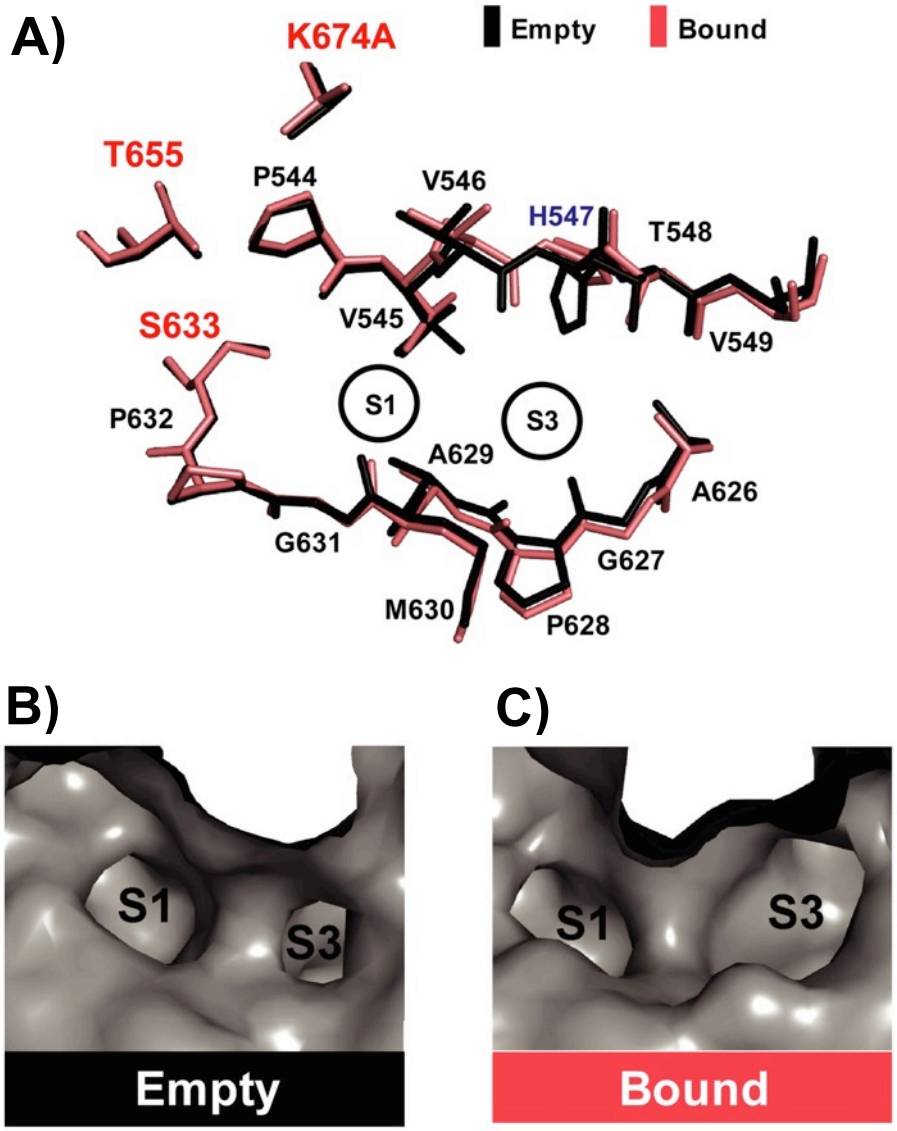


Figure 4-8. Changes in substrate specificity binding pockets between bound and unbound states.

The active site mutant of YAV protease crystallized with two molecules in the asymmetric unit, one molecule in the bound state (acyl-enzyme) and the other with an empty active site. **(A)** A superposition of the residues that make up the substrate binding groove of YAV VP4 protease in the empty (black) and bound state (salmon) are shown. The position of the S1 and S3 binding pockets are denoted by open black circles. The solvent accessible surfaces for the empty **(B)** and the bound states **(C)** of the VP4 substrate binding pockets are shown.

4.3.6. Substrate binding groove: empty vs. bound

The crystal structure of the active site mutant of YAV VP4 reveals an acyl-enzyme complex and an empty active site in the same asymmetric unit (**Figure 4-8**). This gives us the opportunity to observe structural differences within the substrate binding groove during these distinctly different enzyme states. One difference is seen in the loop between β-strands 2 and 3, and resides adjacent to the substrate binding groove (**Figure 4-5**). In the absence of a bound peptide (P6-P1 from the neighbouring molecule), this loop is more disordered with clear density observed only for the main chain atoms. With bound peptide, this turn became ordered and showed clear density for both main chain and side chain atoms. Another clear difference between the bound and unbound states of VP4 is seen within the S3 specificity binding pocket (**Figure 4-8**). The side chains of His^{547} and Val^{545} both show a change in rotamer conformation which results in a larger S3 binding pocket in the bound state (**Figure 4-8**). The main chain atoms that make up the rim of the S3 pocket have also moved to widen the opening in the bound state.

4.3.7. Aqueous channel leads to active site

Surprisingly, the S1 binding pocket in YAV VP4 is not fully enclosed within the acyl-enzyme or product complex structures; rather it is open to the surface of the enzyme via a channel (**Figure 4-9**). This fenestration leads from the protein surface, adjacent to the substrate binding groove, to an area near the proposed deacylating water within the active site. The channel is formed by residues: Ile^{592}, Ala^{629}, Met^{630}, Gly^{631} and Cys^{634}. The interior of the channel is filled with hydrogen bond donors and acceptors. In the native active site structure, no water molecules are found within the channel but ordered water molecules are coordinated at the entrance near Glu^{594}, Asp^{595}, Gly^{631} and Gln^{635}. In the active site mutant structure, water molecules are found throughout the channel in the VP4 with the empty binding groove but are absent in the acyl-enzyme containing VP4. Superposition of these molecules reveals no significant change in the position of the residues that form the channel. The narrowest point in the channel has a diameter of 2.2 Å. Although the diameter of a water molecule is approximately 2.8 Å, there are examples of water channels with narrow passage ways of approximately this dimension. For example, a selectivity filter of around 2 Å in diameter

has been reported in *E. coli* aquaporin Z, a water channel protein [193, 194]. It is possible a deacylating (catalytic water) could utilize such a channel to position itself in the correct trajectory for attack on the ester bond of the acyl-enzyme.

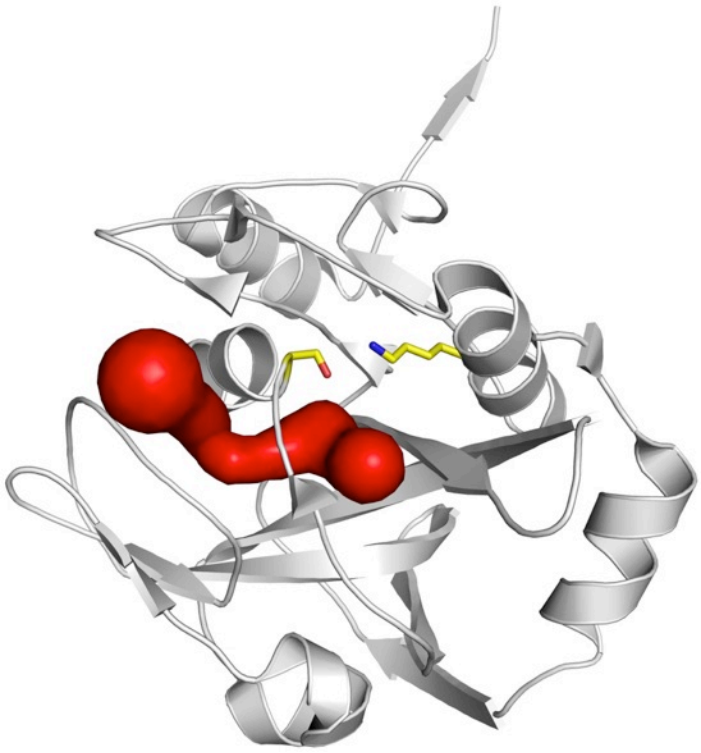


Figure 4-9. A channel leads to the active site of YAV VP4 protease.

A channel (red surface) found adjacent to the substrate binding groove, leads from the enzyme surface to the S1 substrate binding pocket.

4.4. Discussion

4.4.1. Ser^{633} is the YAV VP4 nucleophile, but could Thr^{655} Oγ function as a nucleophile in the absence of Ser^{633} Oγ ?

The Ser/Lys catalytic dyad mechanism is conserved among all Birnavirus VP4 proteases so far studied. Therefore based on sequence alignments, Ser^{633} and Lys^{674} were predicted to be the nucleophile and general base respectively in YAV VP4 [43, 45, 195].

Consistent with these predictions, in a previous mutagenesis study the processing of the YAV segment A polyprotein was completely abolished when Lys^{674}, the proposed general base, was mutated to an aspartic acid[45]. In addition, we show here that mutating Lys^{674} to alanine also produces a dead enzyme. However, previous mutagenesis experiments did not conclusively show that Ser^{633} plays the role of the nucleophile. Both pVP2/VP4 and VP4/VP3 fragment from the YAV segment A polyprotein were detected when Ser^{633} was mutated to a proline. Interestingly, similar results were also seen in IBDV mutagenesis experiments [1, 196]. Our crystal structures of YAV VP4 *trans* acyl-enzyme complexes clearly support the initial hypothesis of Ser^{633} being the nucleophile and Lys^{674} serving as the general base. The acyl-enzyme complexes directly reveal electron density linking the $O\gamma$ of Ser^{633} to the carbonyl carbon of Ala^{716} (the P1 residue of the internal cleavage site) (**Figure 4-4**).

Interestingly, Thr^{655} is the only titratable residue, beside Ser^{633}, within the vicinity of the general base Lys^{674} (**Figure 4-4, 4-6 and 4-7**). A serine or threonine is observed in a similar position in all other known proteases utilizing a Ser/Lys mechanism[197]. If it were possible for the hydroxyl of Thr^{655} to function as the nucleophile in the absence of a Ser^{633} hydroxyl, this would put the attack on the scissile bond in the opposite direction (*re*- rather than *si*-). This would also possibly affect the deacylation step, potentially allowing for a single cleavage event and explaining the pVP2/VP4 and VP4/VP3 precursors still being observed for the S633P mutant.

4.4.2. Interpreting mutagenesis results for non-catalytic residues

Previous studies showed that site-directed mutations in YAV VP4 non-catalytic residues affect polyprotein processing [45]. A cleavage product that was slightly smaller than the YAV segment A polyprotein was observed when Ile^{543} was mutated to a glycine. The side chain of Ile^{543} makes up part of the S1 binding pocket (**Figure 4-10**). Replacing Ile^{543} with a smaller residue like glycine would enlarge the S1 pocket and allow for bulkier side chains to bind. Since S1 and S3 are the major specificity pockets present in the binding groove, such a mutation in the S1 might alter the substrate specificity. Thus, the enzyme is likely to remained functional but with the ability to cleave the polyprotein at a wider variety of location. In a V686Q mutant, both the polyprotein and pVP2 were observed[45]. This suggested the cleavage at the VP4/VP3 junction was affected.

Residue Val[686] does not lie in the active site nor does it reside near the substrate specificity pockets. Instead, it is located adjacent to the C-terminal α-helix with its main chain carbonyl oxygen forming a hydrogen bond with the amide nitrogen of Thr[655], a residue that coordinated the Nζ of the general base[45]. The main chain interaction at this position is observed in all other VP4 structures solved to date[24, 26, 195]. Since this mutant still generated a fragment consistent in size to pVP2, it is unlikely that it affected the catalytic dyad. Amino acid sequence alignment analysis revealed that uncharged residues such as valine, isoleucine, leucine, and alanine are found at this position in other birnaviruses. Replacement of Val[686] with a more polar residue like glutamine could affect the packing around the C-terminus, the substrate at the VP4/VP3 junction. Moreover, Leu[705] which resided on the conserved α-helix 4 near the C-terminus is located across from Val[686]. The proper packing of this last α-helix is likely required for the appropriate presentation of the VP4/VP3 cleavage site.

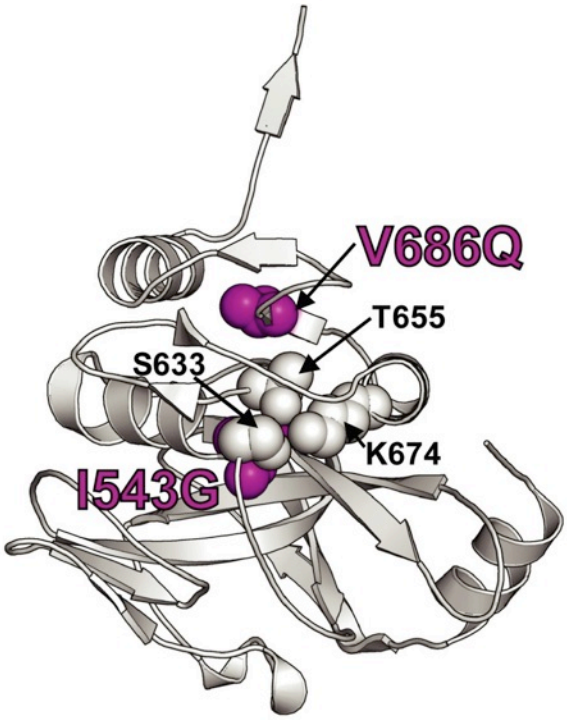


Figure 4-10. Mapping of non-catalytic residue site-directed mutants that affect catalysis.

The VP4 active residues directly involved in catalysis are shown as spheres (Ser^{633}, Lys^{674}, Thr^{655}). Polyprotein cleavage assays conducted by Imajoh *et al.* [45] have shown that non-catalytic residue mutations I543G and V686Q can significantly affect catalysis (shown as magenta spheres).

4.4.3. Cis vs. trans cleavage at the YAV VP4 internal cleavage site

All of the molecules in the crystal structures presented here reveal intermolecular (*trans*) associations between cleavage site and active site. But is this what happens in solution? In solution, the internal cleavage site processing likely only occurs via an intermolecular (*trans*) reaction in that Ala^{716} would not be able to reach the active site in an intra-molecular (*cis*) fashion without significant unfolding of secondary structure. The previous TV-1 VP4 structure, which does not have an internal cleavage site, reveals that the VP4/VP3 cleavage site (734/735 in the YAV polyprotein sequence) is just long enough to form the intramolecular (*cis*) acyl-enzyme complex[195], but the corresponding residue to 716 in YAV is too short to allow for intramolecular (*cis*) cleavage. The

intermolecular processing of the internal cleavage site is further supported by our *trans*-cleavage assay of the full-length YAV VP4 and our observation that the active site mutant enzyme (K674A) would not crystallize until a small amount active wild-type enzyme was added which allowed for cleavage at the internal cleavage site of the inactive mutant enzyme.

How do these intermolecular acyl-enzyme complexes in VP4 protease form and why are they stable enough to be seen in the crystal structure? With the mutant enzyme, we needed to add a small amount of wild-type enzyme to cleave at the internal cleavage site before crystals would form. This suggests that the internal cleavage site product is presented to the VP4 substrate binding groove and the reaction is being turn backwards one step to form the acyl-enzyme. With the general base absent, there would be no suitable deprotonated functional group in the vicinity to activate the deacylating water, thus stabilizing the acyl-enzyme complex. It is not easy to explain how the acyl-enzyme is stabilized in the native active site enzyme. The crystals were formed at pH 6.5, most likely well below the pK_a of the lysine general base, possibly this slowed the deacylation step. It is also possible that the internal cleavage site might be a regulatory mechanism for the VP4 protease. The glutamine found at the P3 position of the internal cleavage site is significantly longer and more polar than the residue seen at the P3 position at the pVP2/VP4 and VP4/VP3 cleavage sites (Ser^{506} and Thr^{732} respectively). This allows for the Gln^{714} to make multiple hydrogen bonding interactions with residues at the bottom of the S3 pocket. It is possible these additional interactions slow the release of the internal cleavage site from the VP4 binding groove and therefore act as an internal inhibitor that functions after the main cleavage sites are processed. Interestingly, there have been previous reports of VP4 self assembles into tubules [138].

4.5. Conclusion

In this study we have shown that YAV VP4 contains an internal cleavage site (716/717) near its C-terminus (residue 734). We have shown that the resulting truncated VP4 is still active. We present for the first time the structure of a Ser/Lys protease *trans* acyl-enzyme complex with a native substrate and native active site. The acyl-enzymes structure proves directly the role of Ser^{633} as the nucleophile in the YAV VP4 reaction

mechanism. A structure of a YAV VP4 mutant with the lysine general base changed to alanine is also presented. Together the two structures reveal the substrate binding site for this protease in three separate stages of the protease reaction cycle: empty active site, acyl-enzyme complex and enzyme-product complex. This structural insight helps interpret mutagenesis results and will be of value in the design of anti-birnavirus compounds and in the basic understanding of the Ser/Lys dyad catalytic mechanism. In addition, we have discovered that this protease contains a unique feature: a channel that runs adjacent to the substrate binding groove that leads to the proposed deacylating water.

5. Concluding remarks

5.1. Contribution of thesis project

5.1.1. Capturing interactions at the polyprotein cleavage sites.

The proper processing of the polyprotein is required for the assembly of birnavirus. We have successfully captured the cleavage at the VP4/VP3 cleavage site in TV-1 VP4 in form of an acyl-enzyme complex and at the internal cleavage site in YAV VP4 in forms of an empty enzyme, acyl-enzyme and enzyme-product complexes. We proposed that TV-1 VP4 is a good candidate for capturing an intramolecular (*cis*) acyl-enzyme complex as its C-terminus (i.e. VP4/VP3 junction) is long enough to reach its own serine nucleophile and it does not have an internal cleavage site to compete for the active site. Indeed, the crystal structure of TV-1 VP4 showed clearly that the serine nucleophile of TV-1 VP4 formed an ester bond of trigonal planer geometry with the last residue of its own C-terminus (i.e. VP4/VP3 junction) as expected from an acyl-enzyme complex. The structure revealed clear continuous electron density for the linkage between the nucleophile serine $O\gamma$ and the main chain carbonyl carbon of the last residue, with full occupancy. This shows directly that cleavage at the VP4/VP3 junction can occur intramolecularly (*cis*). I also sought out to confirm the presence of a previously proposed internal cleavage site in YAV VP4. Its presence was confirmed by 1) capturing it as an intermolecular (*trans*) acyl-enzyme complex, 2) showing cleavage of a MCA-fluorometric peptide substrate carrying the sequence of the internal cleavage site and 3) showing cleavage of the full-length VP4 into a fragment consistent with the size of VP4 truncated at the internal cleavage site. The crystal structure of the lysine mutant yielded both an empty enzyme and an acyl-enzyme complex. This has allowed us to compare the structural changes upon substrate binding. The presence of the substrate/product also allowed us to identify the specificity pockets and analyse the interactions in the substrate binding grooves (e.g. oxyanion hole, deacylating waters, and hydrogen bonding network within the active site and substrate binding groove). The

native active site structure reveals clear electron density for a mix of both product bound complexes and acyl-enzyme complexes. One of the biggest accomplishments in this thesis is that I have managed to elucidate the structure of acyl-enzyme complex for both YAV VP4 and TV-1 VP4 in the presence of a native active site. This has never been accomplished before. The previous IPNV acyl-enzyme structure was solved by using a construct with the lysine general based mutated to an alanine, which helped stabilize the acyl-enzyme.

5.1.2. Capturing the intermediate steps in reaction mechanism of VP4 proteases.

One of our initial hypotheses is that VP4 proteases from birnavirus might have a high propensity to form acyl-enzymes and might be used as models for the study of acyl-enzyme and enzyme-product complexes. Consistent with this, acyl-enzyme complexes were found in all VP4 structures with the substrate/product bound (e.g. IPNV, YAV and TV-1). In addition to the acyl-enzyme complexes, an empty enzyme and product-enzyme complexes have also been captured in YAV VP4. Acyl-enzyme complexes were formed in the presence of native and lysine mutant active sites in VP4 proteases. All acyl-enzyme complexes were captured at pH values below the theoretical pK_a of lysine, which potentially maintained the lysine in the protonated state and prevented it from activating the deacylating water. Perhaps, the high local effective concentration of the product helped to drive the reaction in a reverse direction. In the case of the active site mutant structures, the absence of a lysine general base would have prevented the deprotonation of the deacylating water resulting in the stabilization of an acyl-enzyme complex. Although we do not know the actual pK_a values for the catalytic residues at each stage of the reaction, we might assume that the pK_a of the would-be deacylating water is likely to be higher than that of the nucleophilic serine hydroxyl, it is possible that the low pH prevented or decreased the activation of the deacylating water in the native active site. In the case of the TV-1 structure, a sulphate ion was bound within a positively charged pocket just adjacent to the lysine general base. It is possible that the ionic interaction between the sulphate the ε-amino group of the general base stabilized the protonated form of the general base and prevented it from activating the deacylating water. No other VP4 structures have a similar sulphate binding site.

5.1.3. Capture the first intramolecular (cis) acyl-enzyme complex of a viral protease in PDB.

As discussed in Chapter 1, structural information gained from acyl-enzyme complexes can be used for inhibitor design. However, acyl-enzyme complexes are transient species in the catalytic mechanism and are difficult to capture. Many of the acyl-enzyme structures reported for serine-utilising hydrolases are enzyme-inhibitors complexes. In 1992, Strynadka *et al.* reported an acyl-enzyme intermediate structure of Penicillin G and RTEM-1 β- lactamase from *E. coli* [198]. In 2002, Katona *et al*. reported an acyl-enzyme structure of porcine pancreatic elastase with an N-acetylated tripeptide inhibitor[176]. Interestingly, VP4 proteases form acyl-enzyme complexes with their native substrates, the internal cleavage site in IPNV and YAV and with the VP4/VP3 junction in TV-1 as reported here. A search in the PDB revealed that VP4 proteases constitute all of the acyl-enzyme complexes reported for viral proteases. The VP4s from IPNV and YAV form intermolecular (*trans*) acyl-enzyme complexes. This makes the intramolecular (*cis*) acyl-enzyme complex of VP4 from TV-1 the only viral protease structure of this kind in the PDB and will be of value to all those interested in viral poly-protein processing and those interested in understanding acyl-enzyme formation and covalent catalysis in general.

5.2. Future studies

5.2.1. The importance of residues downstream of the scissile bond.

The substrate / product bound structures of VP4 have revealed how the enzymes interact with the substrates up to the P1 residue at the cleavage site. Residues downstream of the scissile bond also show some degree of conservation. Given that YAV VP4 cleaved a MCA-peptide substrate that lacked these residues, they do not appeared to be essential for recognition. FRET-based peptide cleavage assay can be carried out to determine how changes on the P' side of the substrate affect the rate of the reaction [199].

5.2.2. Involvement of VP4 in pVP2 processing.

Protease VP4 cleaves the polyprotein at the primary cleavage sites (i.e. pVP2/VP4 and VP4/VP3 junctions). In IBDV, the autoproteolytic activity of pVP2 was shown to be responsible for the final cleavage that gives rise to the mature capsid protein VP2 [2]. The S1 and S3 binding pockets are the major determinant of substrate specificity. In spite of this, the P1 and P3 residues at this cleavage site are the same as that at the VP4/VP3 junction, a known target of VP4 protease. The same can be said for DXV, IPNV and TV-1. In this regard, a question remains as to whether similar strategies are employed by the other birnaviruses to process pVP2. If this is true for other birnaviruses, how does VP4 distinguish between this site and the primary cleavage sites? Perhaps, it is dependent on the sequence at which the sites are exposed to VP4 as pVP2 processing was reported to happen later in virion assembly[80]. Other cleavage sites exist within the pVP2 region, comparison of these cleavage sites with the primary cleavage sites (i.e. pVP2/VP4 and VP4/V3 junctions) revealed that they again share similar residues at the P1 and P3 positions. Hence, VP4 was proposed to process pVP2 at these sites but such processing is yet to be confirmed [1,25]. Indeed, VP4 has been found to associate with the virion [70]. A MCA-peptide assay with peptides carrying sequences of these sites can be used to determine if they are targets of VP4 protease.

5.2.3. Enzyme-substrate interactions at the pVP2/VP4 junction.

To date, all the substrate/product bound structures of VP4 are either at the internal or VP4/VP3 cleavage sites. In the intermolecular acyl-enzyme structures of IPNV and YAV, the C-termini (i.e. the substrate/product) are extended and bound into the active site of an adjacent molecule. In contrast, the intramolecular acyl-enzyme structure of TV-1 showed that the C-terminus wrapped around the parameter of enzyme before entering the substrate binding groove. The way in which the substrate presents itself at the pVP2/VP4 cleavage site remains to be elucidated. A crystal structure of a construct that spans pVP2 and VP4 but with the lysine general base mutated to an alanine might give insight to this complex.

5.2.4. Capturing other intermediate steps in the reaction mechanism.

Here we show that we can capture the empty enzyme, acyl-enzyme and product-enzyme complexes. Transition state analogs would be required to trap the transition states. Active site or cleavage site mutations will be required to trap the substrate-enzymes complexes.

References

1. Lejal, N., Da Costa, B., Huet, J. C. & Delmas, B. 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 (2000).

2. Irigoyen, N. *et al*. Autoproteolytic activity derived from the infectious bursal disease virus capsid protein. *J. Biol. Chem.* **284**, 8064-8072 (2009).

3. Boot, H. J., ter Huurne, A. A., Peeters, B. P. & Gielkens, A. L. Efficient rescue of infectious bursal disease virus from cloned cDNA: evidence for involvement of the 3'-terminal sequence in genome replication. *Virology* **265**, 330-341 (1999).

4. Paetzel, M., Karla, A., Strynadka, N. C. & Dalbey, R. E. Signal peptidases. *Chem. Rev.* **102**, 4549-4580 (2002).

5. Kim, A. C., Oliver, D. C. & Paetzel, M. Crystal structure of a bacterial signal Peptide peptidase. *J. Mol. Biol.* **376**, 352-366 (2008).

6. Ferentz, A. E., Walker, G. C. & Wagner, G. Converting a DNA damage checkpoint effector (UmuD2C) into a lesion bypass polymerase (UmuD'2C). *EMBO J.* **20**, 4287-4298 (2001).

7. Luo, Y. *et al*. Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **106**, 585-594 (2001).

8. Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M. Crystal structure of the lambda repressor C-terminal domain provides a model for cooperative operator binding. *Cell* **101**, 801-811 (2000).

9. Botos, I. *et al*. 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 (2004).

10. Maurer, R., Meyer, B. & Ptashne, M. Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor. *J. Mol. Biol.* **139**, 147-161 (1980).

11. Meyer, B. J., Maurer, R. & Ptashne, M. Gene regulation at the right operator (OR) of bacteriophage lambda. II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and cro. *J. Mol. Biol.* **139**, 163-194 (1980).

12. Meyer, B. J. & Ptashne, M. Gene regulation at the right operator (OR) of bacteriophage lambda. III. lambda repressor directly activates gene transcription. *J. Mol. Biol.* **139**, 195-205 (1980).

13. Ekici, O. D., Paetzel, M. & Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci.* **17**, 2023-2037 (2008).

14. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227-33 (2010).

15. Dereeper, A., Audic, S., Claverie, J. M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **10**, 8 (2010).

16. Dereeper, A. *et al*. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465-9 (2008).

17. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).

18. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-552 (2000).

19. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704 (2003).

20. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539-552 (2006).

21. Chevenet, F., Brun, C., Banuls, A. L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).

22. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. in *The Proteomics Protocols Handbook, Humana Press* (ed John M. Walker) 571-607, 2005).

23. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71-5 (2012).

24. Feldman, A. R., Lee, J., Delmas, B. & Paetzel, M. Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism. *J. Mol. Biol.* **358**, 1378-1389 (2006).

25. Da Costa, B. *et al*. Blotched snakehead virus is a new aquatic birnavirus that is slightly more related to avibirnavirus than to aquabirnavirus. *J. Virol.* **77**, 719-725 (2003).

26. Lee, J., Feldman, A. R., Delmas, B. & Paetzel, M. Crystal structure of the VP4 protease from infectious pancreatic necrosis virus reveals the acyl-enzyme complex for an intermolecular self-cleavage reaction. *J. Biol. Chem.* **282**, 24928-24937 (2007).

27. Galloux, M. *et al*. Infectious bursal disease virus, a non-enveloped virus, possesses a capsid-associated peptide that deforms and perforates biological membranes. *J. Biol. Chem.* **282**, 20774-20784 (2007).

28. Suzuki, S., Kimura, M. & Kusuda, R. The complete nucleotide sequence of the polyprotein and VP5 gene of a marine birnavirus. *Fish. Sci.* **64**, 428-433 (1998).

29. Bayliss, C. D. *et al*. A comparison of the sequences of segment A of four infectious bursal disease virus strains and identification of a variable region in VP2. *J. Gen. Virol.* **71 (Pt 6)**, 1303-1312 (1990).

30. Chung, H. K., Kordyban, S., Cameron, L. & Dobos, P. Sequence analysis of the bicistronic Drosophila X virus genome segment A and its encoded polypeptides. *Virology* **225**, 359-368 (1996).

31. Nobiron, I. *et al*. Genome and polypeptides characterization of Tellina virus 1 reveals a fifth genetic cluster in the Birnaviridae family. *Virology* **371**, 350-361 (2008).

32. Larkin, M. A. *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

33. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308 (1999).

34. Casanas, A. *et al*. Structural insights into the multifunctional protein VP3 of birnaviruses. *Structure* **16**, 29-37 (2008).

35. DeLano, W. L. The PyMOL Molecular User's Manual, DeLano Scientific, San Carlos, CA. (2002).

36. Hartley, B. S. & Kilby, B. A. The reaction of p-nitrophenyl esters with chymotrypsin and insulin. *Biochem. J.* **56**, 288-297 (1954).

37. Paetzel, M., Dalbey, R. E. & Strynadka, N. C. The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target. *Pharmacol. Ther.* **87**, 27-49 (2000).

38. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501-4524 (2002).

39. Schechter I & Berger A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157-162 (1967).

40. James M.N.G. in (ed Bond J.S., B. A. J.) 1-8, Portland, Brookfield, VT, 1994).

41. Paetzel, M. & Dalbey, R. E. Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.* **22**, 28-31 (1997).

42. Da Costa, B. *et al*. The capsid of infectious bursal disease virus contains several small peptides arising from the maturation process of pVP2. *J. Virol.* **76**, 2393-2402 (2002).

43. Petit, S., Lejal, N., Huet, J. C. & Delmas, B. Active residues and viral substrate cleavage sites of the protease of the birnavirus infectious pancreatic necrosis virus. *J. Virol.* **74**, 2057-2066 (2000).

44. Galloux, M. *et al*. Peptides resulting from the pVP2 C-terminal processing are present in infectious pancreatic necrosis virus particles. *J. Gen. Virol.* **85**, 2231-2236 (2004).

45. Imajoh, M., Goto, T. & Oshima, S. Characterization of cleavage sites and protease activity in the polyprotein precursor of Japanese marine aquabirnavirus and expression analysis of

generated proteins by a VP4 protease activity in four distinct cell lines. *Arch. Virol.* **152**, 1103-1114 (2007).

46. Harris, T. K. & Turner, G. J. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85-98 (2002).

47. Gutteridge, A. & Thornton, J. M. Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30**, 622-629 (2005).

48. Paetzel, M. *et al*. Use of site-directed chemical modification to study an essential lysine in Escherichia coli leader peptidase. *J. Biol. Chem.* **272**, 9994-10003 (1997).

49. Isom, D. G., Castaneda, C. A., Cannon, B. R. & Garcia-Moreno, B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5260-5265 (2011).

50. Nienaber, V. L. *et al*. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* **18**, 1105-1108 (2000).

51. Sarubbi, E. *et al*. Peptide aldehydes as inhibitors of HIV protease. *FEBS Lett.* **319**, 253-256 (1993).

52. Tate, S. S. & Meister, A. Serine-borate complex as a transition-state inhibitor of gamma-glutamyl transpeptidase. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 4806-4809 (1978).

53. Wolf, W. M. *et al*. Inhibition of proteinase K by methoxysuccinyl-Ala-Ala-Pro-Ala-chloromethyl ketone. An x-ray study at 2.2-A resolution. *J. Biol. Chem.* **266**, 17695-17699 (1991).

54. Bernstein, N. J. & Pratt, R. F. On the importance of a methyl group in beta-lactamase evolution: free energy profiles and molecular modeling. *Biochemistry* **38**, 10499-10510 (1999).

55. Gupton, B. F., Carroll, D. L., Tuhy, P. M., Kam, C. M. & Powers, J. C. Reaction of azapeptides with chymotrypsin-like enzymes. New inhibitors and active site titrants for chymotrypsin A alpha, subtilisin BPN', subtilisin Carlsberg, and human leukocyte cathepsin G. *J. Biol. Chem.* **259**, 4279-4287 (1984).

56. Strynadka, N. C., Martin, R., Jensen, S. E., Gold, M. & Jones, J. B. Structure-based design of a potent transition state analogue for TEM-1 beta-lactamase. *Nat. Struct. Biol.* **3**, 688-695 (1996).

57. Maveyraud, L. *et al*. Crystal Structure of 6R-(Hydroxymethyl)penicillanateComplexed to the TEM-1 alpha-Lactamase from Escherichia coli:Evidence on the Mechanism of Action of a Novel InhibitorDesigned by a Computer-Aided Process. *J. Am. Chem. Soc* **118**, 7435-7440 (1996).

58. Lam, P. Y. *et al*. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380-384 (1994).

59. Kwong, A. D., Kauffman, R. S., Hurter, P. & Mueller, P. Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nat. Biotechnol.* **29**, 993-1003 (2011).

60. Robertson, J. G. Mechanistic basis of enzyme-targeted drugs. *Biochemistry* **44**, 5561-5571 (2005).

61. King, A. M., Lefkowitz, E., Adams, M. J. & Carstens, E. B. in *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses* (Elsevier Science, 2011).

62. Hirayama, T., Nagano, I., Shinmoto, H., Yagyu, K. & Oshima, S. Isolation and characterization of virulent yellowtail ascites virus. *Microbiol. Immunol.* **51**, 397-406 (2007).

63. Teninges, D., Ohanessian, A., Richard-Molard, C. & Contamine, D. Isolation and Biological Properties of Drosophila X Virus. *J. Gen. Virol.* **42**, 241-254 (1979).

64. Buchanan, J. S. *Electron microscopic observations of virus-like particles in the digestive gland of a marine bivalve mollusc Tellina tenuis.* Ser. VIth Annual Meeting of the Society for Invertebrate Pathology, 1973).

65. Hirai, K. & Shimakura, S. Structure of infectious bursal disease virus. *J. Virol.* **14**, 957-964 (1974).

66. Comps, M., Mari, J., Poisson, F. & Bonami, J. R. Biophysical and biochemical properties of an unusual birnavirus pathogenic for rotifers. *J. Gen. Virol.* **72 (Pt 6)**, 1229-1236 (1991).

67. John, K. R. & Richards, R. H. Characteristics of a new birnavirus associated with a warm-water fish cell line. *J. Gen. Virol.* **80 (Pt 8)**, 2061-2065 (1999).

68. Christie, K. E., Havarstein, L. S., Djupvik, H. O., Ness, S. & Endresen, C. Characterization of a new serotype of infectious pancreatic necrosis virus isolated from Atlantic salmon. *Arch. Virol.* **103**, 167-177 (1988).

69. Moss, L. H.,3rd & Gravell, M. Ultrastructure and sequential development of infectious pancreatic necrosis virus. *J. Virol.* **3**, 52-58 (1969).

70. Villanueva, R. A., Galaz, J. L., Valdes, J. A., Jashes, M. M. & Sandino, A. M. Genome assembly and particle maturation of the birnavirus infectious pancreatic necrosis virus. *J. Virol.* **78**, 13829-13838 (2004).

71. MacLachlan, N. J. & Dubovi, E. J. in *Fenner's Veterinary Virology* (Academic Press, 2010).

72. Cosgrove, A. S. Effects of a new nitrofuran feed additive on production efficiency in chicken. *Avian Dis.* **6**, 385 (1962).

73. Biering, E., Nilsen, F., Rodseth, O. M. & Glette, J. Susceptibility of Atlantic halibut Hippoglossus hippoglossus to infectious pancreatic necrosis virus. *Dis. aquat. org* **20**, 183-190 (1994).

74. Schwanz-Pfitzner, I., Ozel, M., Darai, G. & Gelderblom, H. Morphogenesis and fine structure of eel virus (Berlin), a member of the proposed Birnavirus group. *Arch. Virol.* **81**, 151-162 (1984).

75. Brown, M. D. & Skinner, M. A. Coding sequences of both genome segments of a European 'very virulent' infectious bursal disease virus. *Virus Res.* **40**, 1-15 (1996).

76. Shwed, P. S., Dobos, P., Cameron, L. A., Vakharia, V. N. & Duncan, R. Birnavirus VP1 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif. *Virology* **296**, 241-250 (2002).

77. Duncan, R., Nagy, E., Krell, P. J. & Dobos, P. Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions. *J. Virol.* **61**, 3655-3664 (1987).

78. Duncan, R., Mason, C. L., Nagy, E., Leong, J. A. & Dobos, P. Sequence analysis of infectious pancreatic necrosis virus genome segment B and its encoded VP1 protein: a putative RNA-dependent RNA polymerase lacking the Gly-Asp-Asp motif. *Virology* **181**, 541-552 (1991).

79. Zhang, C. X. & Suzuki, S. Comparison of the RNA polymerase genes of marine birnavirus strains and other birnaviruses. *Arch. Virol.* **148**, 745-758 (2003).

80. Chevalier, C., Lepault, J., Erk, I., Da Costa, B. & Delmas, B. The maturation process of pVP2 requires assembly of infectious bursal disease virus capsids. *J. Virol.* **76**, 2384-2392 (2002).

81. Hong, J. R., Gong, H. Y. & Wu, J. L. IPNV VP5, a novel anti-apoptosis gene of the Bcl-2 family, regulates Mcl-1 and viral protein expression. *Virology* **295**, 217-229 (2002).

82. Vancini, R. *et al*. Espirito Santo virus: a new birnavirus that replicates in insect cells. *J. Virol.* **86**, 2390-2399 (2012).

83. Caston, J. R. *et al*. C terminus of infectious bursal disease virus major capsid protein VP2 is involved in definition of the T number for capsid assembly. *J. Virol.* **75**, 10815-10828 (2001).

84. Dobos, P. & Roberts, T. E. The molecular biology of infectious pancreatic necrosis virus: a review. *Can. J. Microbiol.* **29**, 377-384 (1983).

85. Sorimachi M, H. T. Characteristics and pathogenicity of a virus isolated from yellowtail fingerlings showing ascites. *Fish Pathol*, 231–238 (1985).

86. Dobos, P. *et al*. Biophysical and biochemical characterization of five animal viruses with bisegmented double-stranded RNA genomes. *J. Virol.* **32**, 593-605 (1979).

87. Coulibaly, F. *et al*. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell* **120**, 761-772 (2005).

88. Dormitzer, P. R., Nason, E. B., Prasad, B. V. & Harrison, S. C. Structural rearrangements in the membrane penetration protein of a non-enveloped virus. *Nature* **430**, 1053-1058 (2004).

89. Chen, J. Z. *et al*. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10644-10648 (2009).

90. Sanchez-San Martin, C., Lopez, T., Arias, C. F. & Lopez, S. Characterization of rotavirus cell entry. *J. Virol.* **78**, 2310-2318 (2004).

91. Yip, C. W., Hon, C. C., Zeng, F. & Leung, F. C. Cell culture-adapted IBDV uses endocytosis for entry in DF-1 chicken embryonic fibroblasts. *Virus Res.* **165**, 9-16 (2012).

92. Lee, C. C. *et al*. Crystal structure of infectious bursal disease virus VP2 subviral particle at 2.6A resolution: implications in virion assembly and immunogenicity. *J. Struct. Biol.* **155**, 74-86 (2006).

93. Garriga, D. *et al*. The 2.6-Angstrom structure of infectious bursal disease virus-derived T=1 particles reveals new stabilizing elements of the virus capsid. *J. Virol.* **80**, 6895-6905 (2006).

94. Luo, J. *et al*. Surface IgM on DT40 cells may be a component of the putative receptor complex responsible for the binding of infectious bursal disease virus. *Avian Pathol.* **39**, 359-365 (2010).

95. Hirai, K., Funakoshi, T., Nakai, T. & Shimakura, S. Sequential changes in the number of surface immunoglobulin-bearing B lymphocytes in infectious bursal disease virus-infected chickens. *Avian Dis.* **25**, 484-496 (1981).

96. Benton, W. J., Cover, M. S. & Rosenberger, J. K. Studies on the transmission of the infectious bursal agent (IBA) of chickens. *Avian Dis.* **11**, 430-438 (1967).

97. Ogawa, M. *et al*. Seroprevalence of infectious bursal disease virus in free-living wild birds in Japan. *J. Vet. Med. Sci.* **60**, 1277-1279 (1998).

98. Gardner, H., Kerry, K., Riddle, M., Brouwer, S. & Gleeson, L. Poultry virus infection in Antarctic penguins. *Nature* **387**, 245 (1997).

99. Razmyar, J. & Peighambari, S. M. Isolation and characterization of a very virulent Infectious bursal disease virus from turkey. *Acta Virol.* **53**, 271-276 (2009).

100. Murphy, F. A. in *Veterinary Virology* (Academic Press, 1999).

101. Mahgoub, H. A. An overview of infectious bursal disease. *Arch. Virol.* (2012).

102. L'Heritier, P. The hereditary virus of Drosophila. *Adv. Virus Res.* **5**, 195-245 (1958).

103. Adams, J. R. & Bonami, J. R. in *Atlas of invertebrate viruses* 435-442 (CRC Press, Boca Raton, Florida, 1991).

104. M'Gonigle, R. H. Acute catarrhal enteritis of salmonid fingerlings. *Trans. Amer. Fish. Soc.* **70**, 297-303 (1940).

105. Wood, E. M., Snieszko, S. F. & Yasutake, W. T. Infectious pancreatic necrosis in brook trout. *AMA Arch. Pathol.* **60**, 26-28 (1955).

106. Ahne, W., Kelly, R. K. & Sclotfeldt, H. J. in *Fish health protection strategies* (eds Lillelund, K. & Rosenthal, H.) 19-67 (Federal Ministry for Research and Technology, Hamburg/Bonn, 1989).

107. Desautels, D. & MacKelvie, R. M. Practical aspects of survival and destruction of infectious pancreatic necrosis virus. *J. Fish. Res. Bd. Can.* **32**, 523-531 (1975).

108. Billi, J. L. & Wolf, K. Quantitative comparison of peritoneal washes and feces for decting infectious pancreatic necrosis (IPN) virus in carrier brook trout. *J. Fish. Res. Bd. Can.* **16**, 1459-1465 (1969).

109. Wolf, K., Quimby, M. C. & Bradford, A. D. Egg-associated transmission of IPNV virus of trouts. *Virology* **21**, 317-321 (1963).

110. Dorson, M. & Torchy, C. in *Fish and Shellfish Pathology* (ed Ellis, A.) 251-261 (Academic Press, Lodon, 1985).

111. Hill, B. J., 1976. Properties of virus iosolated from the bi-valve mollusk Tellina tenuis. In: Pagr, L.A. (Ed.), Wildlife Diseases. Plenum Press, New York, pp. 445-452 (1976).

112. Hill, B.J., Way, K. Serological classification of infectious pancreatic necrosis (IPN) virus and other aquatic birnaviruses. *Annu. Rev. Fish Dis.* **5**, 55–77 (1995).

113. Underwood, B. O., Smale, C. J., Brown, F. & Hill, B. J. Relationship of a virus from Tellina tenuis to infectious pancreatic necrosis virus. *J. Gen. Virol.* **36**, 93-109 (1977).

114. Hosono N, Suzuki S, Kusuda R. Genogrouping of birnaviruses isolated from marine fish: a comparison of VP2/NS junction regions on genome segment A. *J Fish Dis* **19**, 295–302 (1996).

115. Kusuda R, Kado K, Takeuchi Y, Kawai K. Characteristics of two virus strains isolated from young Japanese flounder Paralichthys olivaceus. *Suisanzoshoku* **37**, 115–120 (1989).

116. Nakajima K, Maeno Y, Arimoto M, Inoue K, Sorimachi M. Viral deformity of yellowtail. *Fish Pathol* **28**, 125–129 (1993).

117. Kamakura M, Suzuki S, Kusuda R. Characterization of birnavirus isolated from diseased tiger puffer. Ann Meet Jpn Soc Fish Pathol, Mie, Japan. , 14 (September 1995).

118. Suzuki S, Utsunomiya I, Kusuda R. Experimental infection of marine birnavirus strain JPO-96 to Japanese pearl oyster Pinctada fucata. *Bull Mar Sci Fish Kochi Univ* **18**, 38-41 (1998).

119. Suzuki S, Kamakura M, Kusuda R. Isolation of birnavirus from Japanese pearl oyster Pinctada fucata. *Fish Sci* **64**, 342–343 (1998).

120. Suzuki S, Nakata T, Kamakura M, Yoshimoto M, Furukawa Y, Yamashita Y, Kusuda R. Isolation of birnavirus from Agemaki (jack knife clam) Sinonovacura constricta and survey of the virus using PCR technique. *Fish Sci* **63**, 563–566 (1997).

121. Courtenay, W. R. & Williams, J. D. in *SNAKEHEADS (Pisces, Channidae) - A Biological Synopsis and Risk Assessment* 75-76 (U.S. Geological Survey, Gainesville, Florida, USA, 2004).

122. Ng, P.K.L., and Lim, K.K.P. Snakeheads (Pisces: Channidae): Natural, history, biology and economic importance. , 127 (1990).

123. Rainboth, W. J. in *FAO Species Identification Field Guide for Fishery Purposes: Rome, Italy, Food and Agriculture Organization of the United Nations (FAO)* 265, 1996).

124. Garriga, D. *et al*. Activation mechanism of a noncanonical RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20540-20545 (2007).

125. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**, 235-240 (2001).

126. Leveque, V. J. *et al*. Identification of a C-terminal regulatory motif in hepatitis C virus RNA-dependent RNA polymerase: structural and biochemical analysis. *J. Virol.* **77**, 9020-9028 (2003).

127. Dobos, P. Protein-primed RNA synthesis in vitro by the virion-associated RNA polymerase of infectious pancreatic necrosis virus. *Virology* **208**, 19-25 (1995).

128. MALSBERGER, R. G. & CERINI, C. P. Characteristics of Infectious Pancreatic Necrosis Virus. *J. Bacteriol.* **86**, 1283-1287 (1963).

129. Pan, J., Lin, L. & Tao, Y. J. Self-guanylylation of birnavirus VP1 does not require an intact polymerase activity site. *Virology* **395**, 87-96 (2009).

130. Calvert, J. G., Nagy, E., Soler, M. & Dobos, P. Characterization of the VPg-dsRNA linkage of infectious pancreatic necrosis virus. *J. Gen. Virol.* **72 (Pt 10)**, 2563-2567 (1991).

131. Xu, H. T., Si, W. D. & Dobos, P. Mapping the site of guanylylation on VP1, the protein primer for infectious pancreatic necrosis virus RNA synthesis. *Virology* **322**, 199-210 (2004).

132. Dobos, P. In vitro guanylylation of infectious pancreatic necrosis virus polypeptide VP1. *Virology* **193**, 403-413 (1993).

133. Spies, U. & Muller, H. Demonstration of enzyme activities required for cap structure formation in infectious bursal disease virus, a member of the birnavirus group. *J. Gen. Virol.* **71 (Pt 4)**, 977-981 (1990).

134. Magyar, G., Chung, H. K. & Dobos, P. Conversion of VP1 to VPg in cells infected by infectious pancreatic necrosis virus. *Virology* **245**, 142-150 (1998).

135. Graham, S. C. *et al*. The N-terminus of the RNA polymerase from infectious pancreatic necrosis virus is the determinant of genome attachment. *PLoS Pathog.* **7**, e1002085 (2011).

136. Zhang, X. & Oglesbee, M. Use of surface plasmon resonance for the measurement of low affinity binding interactions between HSP72 and measles virus nucleocapsid protein. *Biol. Proced. Online* **5**, 170-181 (2003).

137. Pan, J., Vakharia, V. N. & Tao, Y. J. The structure of a birnavirus polymerase reveals a distinct active site topology. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7385-7390 (2007).

138. Granzow, H. *et al*. A second form of infectious bursal disease virus-associated tubule contains VP4. *J. Virol.* **71**, 8879-8885 (1997).

139. Crane, M. S. *et al*. First isolation of an aquatic birnavirus from farmed and wild fish species in Australia. *Dis. Aquat. Organ.* **43**, 1-14 (2000).

140. Espinoza, J. C., Hjalmarsson, A., Everitt, E. & Kuznar, J. Temporal and subcellular localization of infectious pancreatic necrosis virus structural proteins. *Arch. Virol.* **145**, 739-748 (2000).

141. Ozel, M. & Gelderblom, H. Capsid symmetry of viruses of the proposed Birnavirus group. *Arch. Virol.* **84**, 149-161 (1985).

142. Jmol: an open-source Java viewer for chemical structures in 3D**.** http://www.jmol.org/ **.**

143. Coulibaly, F., Chevalier, C., Delmas, B. & Rey, F. A. Crystal structure of an Aquabirnavirus particle: insights into antigenic diversity and virulence determinism. *J. Virol.* **84**, 1792-1799 (2010).

144. Spies, U., Muller, H. & Becht, H. Nucleotide sequence of infectious bursal disease virus genome segment A delineates two major open reading frames. *Nucleic Acids Res.* **17**, 7982 (1989).

145. Maraver, A. *et al*. The oligomerization domain of VP3, the scaffolding protein of infectious bursal disease virus, plays a critical role in capsid assembly. *J. Virol.* **77**, 6438-6449 (2003).

146. Vassylyev, D. G. *et al*. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **417**, 712-719 (2002).

147. Vojnic, E., Simon, B., Strahl, B. D., Sattler, M. & Cramer, P. Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *J. Biol. Chem.* **281**, 13-15 (2006).

148. Lombardo, E. *et al*. VP1, the putative RNA-dependent RNA polymerase of infectious bursal disease virus, forms complexes with the capsid protein VP3, leading to efficient encapsidation into virus-like particles. *J. Virol.* **73**, 6973-6983 (1999).

149. Tacken, M. G., Peeters, B. P., Thomas, A. A., Rottier, P. J. & Boot, H. J. Infectious bursal disease virus capsid protein VP3 interacts both with VP1, the RNA-dependent RNA polymerase, and with viral double-stranded RNA. *J. Virol.* **76**, 11301-11311 (2002).

150. Kochan, G., Gonzalez, D. & Rodriguez, J. F. Characterization of the RNA-binding activity of VP3, a major structural protein of Infectious bursal disease virus. *Arch. Virol.* **148**, 723-744 (2003).

151. Cleary, M. L., Smith, S. D. & Sklar, J. Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. *Cell* **47**, 19-28 (1986).

152. Yang, J. *et al*. Prevention of apoptosis by Bcl-2: release of cytochrome c from mitochondria blocked. *Science* **275**, 1129-1132 (1997).

153. Oltvai, Z. N., Milliman, C. L. & Korsmeyer, S. J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* **74**, 609-619 (1993).

154. Ola, M. S., Nawaz, M. & Ahsan, H. Role of Bcl-2 family proteins and caspases in the regulation of apoptosis. *Mol. Cell. Biochem.* **351**, 41-58 (2011).

155. Li, Z. *et al*. Critical role for voltage-dependent anion channel 2 in infectious bursal disease virus-induced apoptosis in host cells via interaction with VP5. *J. Virol.* **86**, 1328-1338 (2012).

156. Liu, M. & Vakharia, V. N. Nonstructural protein of infectious bursal disease virus inhibits apoptosis at the early stage of virus infection. *J. Virol.* **80**, 3369-3377 (2006).

157. Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267-271 (1996).

158. Shimizu, S., Narita, M. & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* **399**, 483-487 (1999).

159. Liu, J., Wei, L., Jiang, T., Shi, L. & Wang, J. Reduction of infectious bursal disease virus replication in cultured cells by proteasome inhibitors. *Virus Genes* **35**, 719-727 (2007).

160. Cutrin, J. M. *et al*. Restriction fragment length polymorphisms and sequence analysis: an approach for genotyping infectious pancreatic necrosis virus reference strains and other aquabirnaviruses isolated from northwestern Spain. *Appl. Environ. Microbiol.* **70**, 1059-1067 (2004).

161. Lombardo, E., Maraver, A., Espinosa, I., Fernandez-Arias, A. & Rodriguez, J. F. VP5, the nonstructural polypeptide of infectious bursal disease virus, accumulates within the host plasma membrane and induces cell lysis. *Virology* **277**, 345-357 (2000).

162. Chung, I. Y. & Paetzel, M. Expression, purification and crystallization of VP4 protease from Tellina virus 1. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **67**, 157-160 (2011).

163. Collaborative Computing Project No. 4 (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect. D* **50**, 760–763.

164. Adams, P. D. *et al*. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948-1954 (2002).

165. Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1966-1973 (2003).

166. McCoy, A. J. *et al*. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

167. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255 (1997).

168. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).

169. Vagin, A. A. *et al*. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184-2195 (2004).

170. Painter, J. & Merritt, E. A. *TLSMD* web server for the generation of multi-group TLS model. *J. Appl. Cryst* **39**, 109-111 (2006).

171. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 439-450 (2006).

172. Maiti, R., Van Domselaar, G. H., Zhang, H. & Wishart, D. S. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* **32**, W590-4 (2004).

173. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579 (1995).

174. Tsodikov, O. V., Record, M. T.,Jr & Sergeev, Y. V. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* **23**, 600-609 (2002).

175. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. *Bioinformatics* **15**, 305–308 (1999).

176. Katona, G. *et al*. X-ray structure of a serine protease acyl-enzyme complex at 0.95-A resolution. *J. Biol. Chem.* **277**, 21962-21970 (2002).

177. Radisky, E. S., Lee, J. M., Lu, C. J. & Koshland, D. E.,Jr. Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6835-6840 (2006).

178. Menard, R. & Storer, A. C. Oxyanion hole interactions in serine and cysteine proteases. *Biol. Chem. Hoppe Seyler* **373**, 393-400 (1992).

179. Bürgi, H. B., Dunitz, J. D. & Shefter, E. Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.* **95**, 5065-5067 (1973).

180. Wilmouth, R. C. *et al*. Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. *Nat. Struct. Biol.* **4**, 456-462 (1997).

181. Alber, T., Petsko, G. A. & Tsernoglou, D. Crystal structure of elastase-substrate complex at -- 55 degrees C. *Nature* **263**, 297-300 (1976).

182. Ding, X., Rasmussen, B. F., Petsko, G. A. & Ringe, D. Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase. *Biochemistry* **33**, 9285-9293 (1994).

183. Paetzel, M., Dalbey, R. E. & Strynadka, N. C. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature* **396**, 186-190 (1998).

184. Wilmouth, R. C. *et al*. Inhibition of elastase by N-sulfonylaryl beta-lactams: anatomy of a stable acyl-enzyme complex. *Biochemistry* **37**, 17506-17513 (1998).

185. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.* (1992).

186. Evans, P. Scaling and assessment of data quality. *Acta Cryst.* **D62**, 72-82 (2006).

187. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22-25 (2010).

188. Murshudov, G., Vagin, A. & and Dodson, E. in *in the Refinement of Protein structures, Proceedings of Daresbury Study Weekend.*, 1996).

189. Reynolds, C., Damerell, D. & Jones, S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics* **25**, 413-414 (2009).

190. Petrek, M. *et al*. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **7**, 316 (2006).

191. Petrek, M., Kosinova, P., Koca, J. & Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15**, 1357-1363 (2007).

192. Zimmerman, M., Ashe, B., Yurewicz, E. C. & Patel, G. Sensitive assays for trypsin, elastase, and chymotrypsin using new fluorogenic substrates. *Anal. Biochem.* **78**, 47-51 (1977).

193. Savage, D. F., O'Connell, J. D.,3rd, Miercke, L. J., Finer-Moore, J. & Stroud, R. M. Structural context shapes the aquaporin selectivity filter. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17164-17169 (2010).

194. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O'Connell, J. D.,3rd & Stroud, R. M. Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z. *PLoS Biol.* **1**, E72 (2003).

195. Chung, I. Y. & Paetzel, M. Crystal structure of a viral protease intramolecular acyl-enzyme complex. Insights into *cis*-cleavage at the VP4/VP3 junction of Tellina birnavirus. *J. Biol. Chem.* **286**, 12475-12482 (2011).

196. Rodriguez-Lecompte, J. C. & Kibenge, F. S. Site-directed mutagenesis of Avibirnavirus VP4 gene. *Virology* **292**, 241-246 (2002).

197. Paetzel, M., Dalbey, R. E. & Strynadka, N. C. 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 (2002).

198. Strynadka, N. C. *et al*. Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution. *Nature* **359**, 700-705 (1992).

199. Ekici, O. D. *et al*. Profiling the substrate specificity of viral protease VP4 by a FRET-based peptide library approach. *Biochemistry* **48**, 5753-5759 (2009).

Appendices

Appendix A.

Standard curve of HiPrep 26/60 Sephacryl S-100 high resolution.

	MW (KDa)	Log MW	Elution vol.	Ve-Vo	Vt-Vo	Kav *
Blue dextran			97.02		222.98	
Albumin	67	1.8261	115	18		0.081397
Ovalbumin	43	1.6335	129	32		0.145574
Chymotrypsinogen A	25	1.3979	160	63		0.282581
Ribonuclease A	14	1.1367	181	84		0.378106

Table A. Protein standard in the Amersham Bioscience's LMW Calibration Kit and values used in constructing the standard curve.

* The K_{av} is calculated using the following equation:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

where V_e is the elution volume of the protein

V_o is the void volume (the elution volume of blue dextran)

V_t is the total bed volume

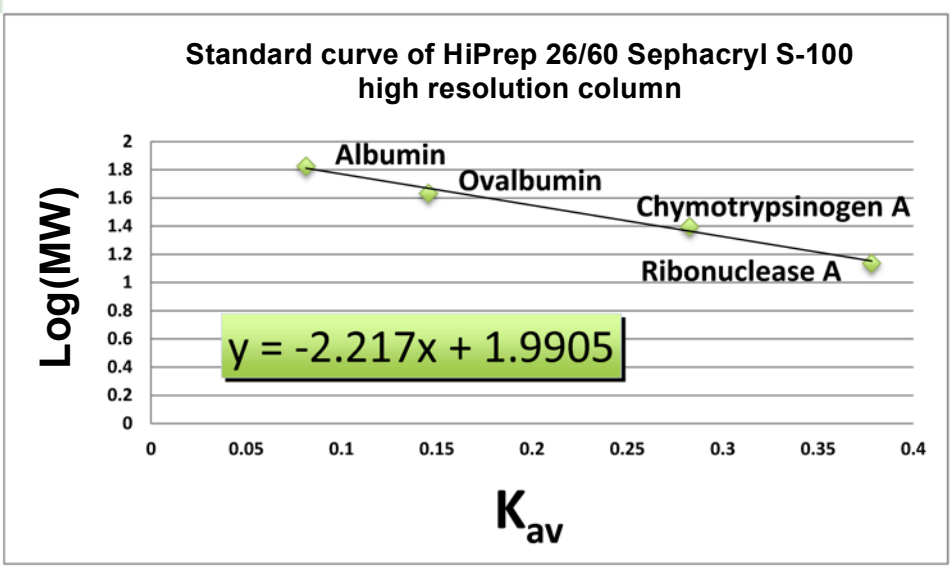


Figure A. Standard curve of HiPrep 26/60 Sephacryl S-100 high resolution column.

The standard curve was constructed by running proteins of known molecule weight (MW) and blue dextran (Amersham Bioscience's LMW Calibration Kit, **Table A**) on a HiPrep 26/60 Sephacryl S-100 high resolution column. The column was connected to an ÄKTA Prime™ system (Pharmacia) and the flow rate was 1 mL/min. The loop size was 5 mL and the injection volume was 3 mL. The concentration of blue dextran was 1 mg/mL and each of the protein standards has a concentration of 3 mg/mL. The elution volumes are plotted as K_{av} on the X axis and the Log of MWs are plotted on the Y axis. The equation for the standard curve is shown in the green box.