Structural and Functional Insights into Animal Circoviruses and Gyroviruses

Rebecca Barnewall

A dissertation submitted in partial fulfilment of the requirements for the degree of

Bachelor of Animal Science (Honours)

School of Veterinary and Animal Sciences

Faculty of Science,

Charles Sturt University

September 2017

Rebecca Barnewall 11525012

Abstract

Circoviridae and *Anelloviridae* viruses are present worldwide and can infect a large range of host species. New virus species are constantly being discovered and assigned to these families; whilst others that are likely to be assigned to these families, remain to be officially classified taxonomically. The virus families *Circoviridae* and *Anelloviridae* possess a circular genome, enclosed in a small and simple isohedral capsid.

This study provides determination of an expression method for *Gyrovirus* constructs CAV VP1 X and VP1 XL and low reslution

These findings indicate that there is now a known expression method for CAV constructs VP1 X and VP1 XL, and BCV X. Although expression trails for PCV-3 and AfaCV were unsuccessful, these initial trials will provide a base line for future trials.

The key aims of this honours thesis were to characterise the structure of BCV X with and without DNA using crystallography, design and test recombinant expression constructs of Starfish Cap protein and test recombinant expression of PCV-3 and CAV.

Successful expression and x-ray diffraction of BCV X + DNA VLP was achieved however the resolution achieved was not high enough to model the complex. Unfortunately expression of both PCV-3 and AfaCV was unsuccessful in this study due to time restraints and/or cell line issues. Achievement of expression of CAV VP1 X and

CAV VP1 XL and subsequent solubilisation will provide a basis for expression and solubilisation of the two CAV constructs in the future studies. Overall, the preliminary results achieved in this study will be the basis for future work in the laboratory.

These findings indicate that there is now a known expression method for CAV constructs VP1 X and VP1 XL, and BCV X. Although expression trails for PCV-3 and AfaCV were unsuccessful, these initial trials will provide a base line for future trials.

Contents

List of tables

List of figures

Acknowledgements

I would like to thank my supervisors Professor Jade Forwood, Professor Shane Raidal and fellow laboratory colleagues. With special mention of my primary supervisor Professor Jade Forwood, for his constant support, mentoring and insight into protein analysis. I thank my fellow laboratory colleagues Kate Smith, Sophie Tsimbalyuk, Sarah Williams and Shubhagata Das for their support, encouragement and friendship throughout my studies. Special thank you, to Shubhagata Das for training me in the laboratory procedures and helping me with my experiments. Thank you also to Emily Cross for ordering in required supplies and work in the laboratory making buffers and the like.

I would also like to thank Charles Sturt University for the Honours Scholarship and The Graham Centre for Operating Honours Scholarship. I am extremely grateful for the support of these scholarships during my honours year.

Finally I would like to thank my housemates and parents for their support, friendship and love throughout the last year. It has been a long eventful 9 months with many challenges that have helped me prosper and develop.

Certificate of Authorship

I Rebecca Barnewall, hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgement is made in the dissertation. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged.

I agree that the dissertation be accessible for the purpose of study and research in accordance with the normal conditions established by the University Librarian for the care, loan and reproduction of the thesis.*

…….

……..

Signature Date

Rebecca Barnewall 11525012

List of Units

˚C	degrees Celsius
bp	base pairs
cm	centimetre
g	grams
g/L	grams per litre
g/mL	grams per millilitre
h	hour
kb	kilobase
kDa	kilodalton
L	litre
M	molar
mAU	milli absorbance unit
mg	milligrams
mg/mL	milligrams per millilitre
min	minute
mL	millilitre
mm	millimetres
mM	millimolar
ng	nanograms
nt	nucleotide
rpm	revolutions per minute
rpm	revolutions per minute
s	seconds
V	volts
v/v	volume per volume
w/v	weight per volume
µg/mL	micrograms per millilitre
µL	microlitre

List of Abbreviations

3D	Three dimensional
AA	Amino acid
AfaCV	Asterias forbesi-associated circular virus
ARM	Arginine Rich Motifs
ATPase	Adenosine triphosphatase
BCV	Bat Associated Circovirus
BFDV	Beak and Feather Disease Virus
Cap	Capsid protein
Cap	Gene encoding the capsid protein
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CAV	Chicken Anaemia Virus
CRESS-DNA	Circular Rep-encoding ssDNA Viruses
Cryo-EM	Cryo Electron Microscopy
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
FLPC	Fast protein liquid chromatography
GST	Glutathione-S-transferase
GTPase	Guanosinetriphosphatase
GuHCl	Guanidine Hydrochloride
HIS	Histidine
ICTV	International Committee on Taxonomy of Viruses
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LB	Luria-Bertani
LIC	Ligation independent cloning
mRNA	Messenger Ribonucleic acid
NaCl	Sodium Chloride

NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
OD	Optical density
ORF	Open reading frame
PBFD	Psittacine Beak and Feather Disease
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCV-1	Porcine Circovirus 1
PCV-2	Porcine Circovirus 2
PCV-3	Porcine Circovirus 3
pH	Potential of Hydrogen
PMWS	Post-weaning multisystemic wasting syndrome
PRRS	Porcine reproductive & Respiratory syndrome
RCR	Rolling circle replication
Rep	Gene encoding the replication protein
Rep	Replication associated gene/protein
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
ssDNA	Single stranded Deoxyribonucleic acid
SSWD	Sea Star Wasting Disease
VLP	Virus like Particle
VP (1-3)	Viral Protein (1-3)
PCVAD	Porcine circovirus associated diseases
PNDS	Porcine dermatitis and nephrophathy syndrome
IHC	Immunohistochemistry

1.0 Literature Review

1.1 Introduction

The virus families *Circoviridae* and *Anelloviridae* possess a circular genome, enclosed in a small and simple isohedral capsid (1-3). The family Circoviridae has undergone recent classification and now contains two genera, *Circovirus* and *Cyclovirus*. Prior to 2017, family differences in genomic organisation led to classification into two genera, *Circovirus* and *Gyrovirus* (3-5), however, the latter was reassigned under the Anelloviridae family (6). The genomic organisation of Circoviruses, Cycloviruses and Gyroviruses is represented in Figures 1.1, 1.2 and 1.3 respectively.

Circoviridae and *Anelloviridae* viruses are present worldwide and can infect a large range of host species. New virus species are constantly being discovered and assigned to these families, whilst others that are likely to be assigned to these families, remain to be officially classified taxonomically (6-9). As both families can infect a range of animals, wild to agricultural, the economic impact associated with *Circoviridae* and *Anelloviridae* diseases are significant (6, 10). Porcine Circovirus 2 (PCV-2) is one of the top three economically significant viral infections of swine farming and is a precursor for more significant swine diseases such as Porcine reproductive & Respiratory syndrome (PRRS) and mycoplasmal pneumonia(10) . Other *Circoviridae* and *Anelloviridae* also cause economic impact through both; direct losses and control strategy expenses (11, 12) .

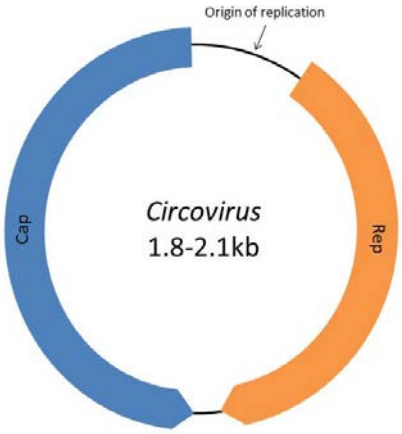


Figure 1.1Schematic diagram of Circovirus

Non-enveloped, round, isohedral symmetry, about 17 nm in diameter. *Circovirus* capsid consists of 12 pentagonal trumpet-shaped pentamers. Genome is ssDNA of about 1800 to 2000bp. Replication is through double-stranded intermediates; the Rep protein initiates rolling circle replication (RCR). The viral DNA and complementary DNA of replicative intermediate encode two mRNAs for the Rep and Capsid proteins (13).

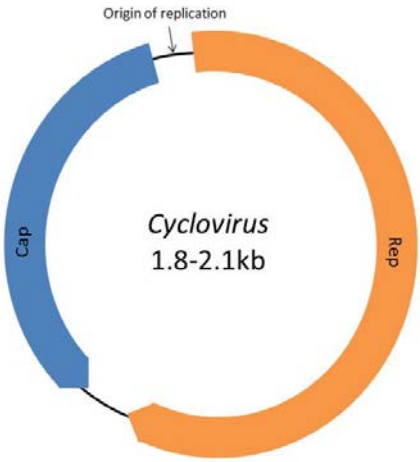


Figure 1.2 Schematic diagram of Cyclovirus

Non-enveloped, round, isohedral symmetry, about 17 nm in diameter. Cyclovirus capsid consists of 12 pentagonal trumpet-shaped pentamers. Genome is ssDNA of about 1800 to 2000 bp. Replication is through double-stranded intermediates. The replication (Rep) protein initiates and terminates rolling circle replication (RCR). The viral DNA and complementary DNA of the replicative intermediate encode two mRNAs for the Rep and Capsid proteins (14).

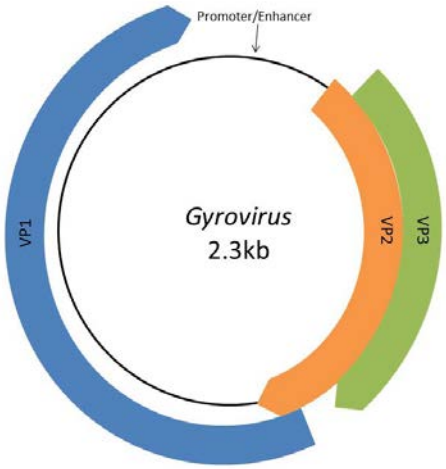


Figure 1.3 Schematic diagram of *Gyrovirus*

Non-enveloped, round, isohedral symmetry, about 25 nm in diameter. *Gyrovirus* capsid consists of 12 pentagonal trumpet-shaped pentamers. Genome is ssDNA of about 2300bp. Replication is through double-stranded intermediates; a Rep-like protein initiates rolling circle replication (RCR). mRNA is transcribed by cellular RNA polymerase II and encodes three proteins: VP1 (structural protein), VP2 and VP3 (15).

The *Circoviridae* family represent an expanding assembly of viruses that have early origins and whose scope is ever-increasing, that may soon include all vertebrates (12, 16). Circovirus-like sequences have been identified in fossil sequences suggesting circoviruses may represent a form of evolution relating to *Nanoviridae* and *Geminiviridae* families (4, 17-20).

1.2 Nomenclature/ Taxonomy

Circoviridae are comprised of a covalently closed circular single stranded deoxyribonucleic acid (ssDNA) monopartite genome ranging from approximately 1.7 to 2.1 kb (4, 9, 21). Virions are nonenveloped, 12-35nm in diameter, and display

icosahedral symmetry. Replication occurs by rolling circle replication (RCR) (9, 10). Phylogenetic and genomic difference led to the *Circoviridae* family being classified into two genera, *Circovirus* and *Cyclovirus* (6). Species are assigned to a given genus based on genomic organisation, with particular focus on the location of origin of replication relative to coding regions (21) (Figure 1.1 & 1.2). Threshold for members of the family *Circoviridae* is 80% genome-wide nucleotide sequence identity (6, 21).. New discoveries of further species in the genus remain to be included in the taxonomic classification (4).

Coding regions of Circoviruses are arranged in an ambisense orientation within the genome organisation, creating two intergenic regions. The larger intergenic region is between the 5' ends of the two major open reading frames (ORFs) Rep and Cap, whilst the shorter region exists between the 3' ends of the genes (4). The aambisense viruses contain at least one gene segment that is positive, and one part that is negative in polarity. The 5' part of ambisense ribonucleic acid (RNA) is of positive polarity containing an ORF that can be translated directly, whilst the section containing negative polarity contain a ORF only (22).

Circoviruses encode two proteins Cap and Rep. In the case of PCV-2, the Cap protein, with a molecular mass of 28 kDa, is encoded by ORF 2, and encapsidates the genome of1,767 bases. The Cap protein of PCV-1 and PCV-2 are 66% identical. The N terminus of Cap circovirus proteins is highly basic, and interacts with the packaged DNA (9).

Anelloviridae contain 10 genera of vertebrate-infecting viruses and other than *Circoviridae*, represents the only other animal viruses that contain covalently closed, circular ssDNA genomes. Members harbour a conserved genetic organisation, with a coding region containing a major ORF, an overlapping ORF2, and additional ORFs-, and untranslated regions (23) (Figure 1.3). The genus *Gyrovirus* encompasses only one virus, Chicken anaemia virus (CAV). Here the capsid consists of a single type of structural protein VP1. The protein possess both structural and function roles, with motifs associated with RCR encompassed in the protein (4, 9) The N terminus of ORF 2 interacts with packaged DNA while the C terminus of the protein consists of motifs associated with RCR (9).The noncoding region of the CAV genome is partly Guanine Cytosine rich and able to form putative hairpin structures as a negative sense genome organisation. The nontranscribed region of the genome contains transcription initiation and termination signals. Tandemly arranged within this nontranscribed region is four or five 19 nucleotide repeats with which promoter enhancer activity is associated (4).

1.3 Replication

Rolling circular replication is a mechanism adopted by some plasmids that represent one of the simplest initiation strategies (9, 24, 25). The replication process is characterised by a singular initiation mechanism, which relies on sequence-specific cleavage. This occurs at the nick site of the double strand origin of one of the parental DNA strands by an initiator Rep protein (Figure 1.4) (25).

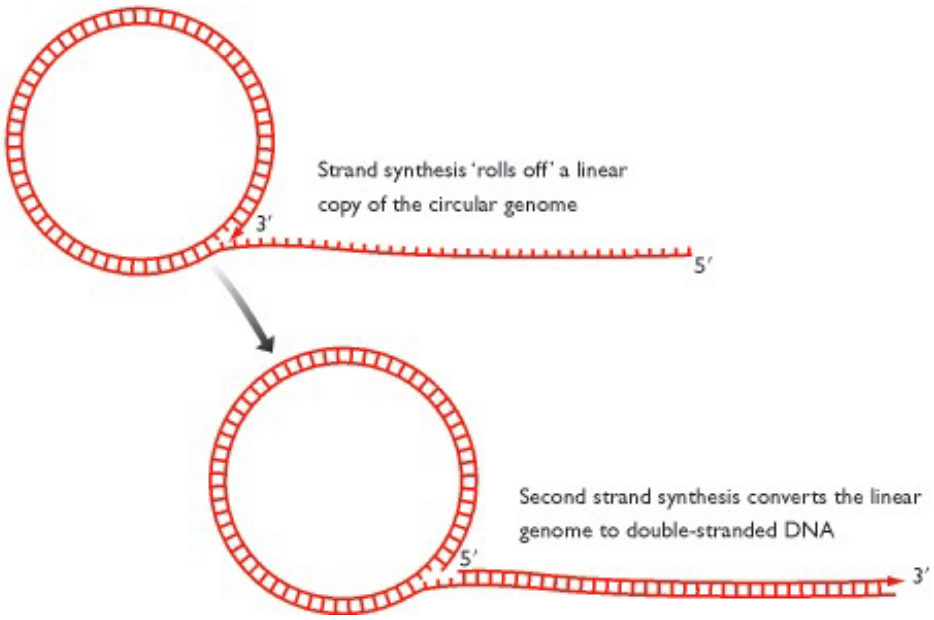


Figure 1.4 Rolling Circle Replication

A replication process, which involves continual synthesis of a polynucleotide, that is 'rolled off' a circular template molecule (26).

1.3.1 Rep protein

The Rep gene of Circoviruses and Gyroviruses harbours amino acid functional motifs linked to RCR. The functional domain of the Rep protein is conserved and contains an endonuclease domain and P-loop (27, 28).The DNA-binding domains may also play a significant role in specificity and binding of substrates. Nuclear magnetic resonance (NMR) has been employed to resolve some of the structural elements for the Rep protein of PCV-2 (27, 29, 30).The structurally resolved endonuclease domain of Rep protein of PCV-2 has elucidated tyrosine or serine rich motifs to play a role in metal ion activation, DNA binding and ssDNA cleavage. Previous structural and sequence analysis has shown that there are three motifs near the N-terminus of the

endonuclease domain which is conserved across circoviruses, geminiviruses and nanoviruses (27, 31).

The second structural domain of the Rep protein is a P-loop motif which is conserved amongst circoviruses and crucial for PCV replication (32-35). Activity involving ATPase binding of this domain was detected in PCV (30), with a similar consensus ATPase/GTPase binding motif detected in geminiviruses which is apparently essential for viral genome replication (36).

It should be noted that Gyroviruses do not possess Rep protein rather a Rep-like protein that is involved in initiation of RCR (15).

1.3.2 Cap protein

Capsid proteins associate together to enclose the genome, forming an outer shell, thereby serving to protect the viral genome (37, 38). These macromolecular assemblies are formed during replication. The assemblies can range from large non-enveloped spherical capsid virions with icosahedral regularity to smaller cytoplasm and nucleus localised complexes (1). In addition to structural roles, the Cap protein plays many important roles in viral infection including virus attachment, cell entry, uncoating to release the genome into the target cell and packaging of newly formed viral particles (37-39).. In the case of PCV-2, the Cap protein binds receptors including glycosaminoglycan's (GAGs), heparin sulphate and chondroitin sulphate B, present in many cell types (40).

Open reading frame 2 of PCV-2 encodes the capsid protein, with a molecular mass of 27.8kDa (41). Cap is considered to be the most immunogenic protein of the circovirus PCV-2. (33)

1.4 Crystallography

Three dimensional atomic structures provide insight into viral life cycles and development of antiviral drugs. X-ray crystallography and Cryo Electron Microscopy (Cryo-EM) have been used to determine the atomic structure of viruses. Both techniques require relatively large quantities of protein, which can either be purified directly from their host or from tissue culture (42, 43). Once the protein is purified, conditions that may induce crystallisation of the protein are screened using a common technique the hanging drop. Crystals are generally then optimised from these screens, and diffracted using X-rays, generated by either a rotating anode, or a synchrotron. Collection of the reflection data can be used to produce a model of the protein that nest fits an electron density map (1, 44).

One limitation of X-ray crystallography is that it requires molecules to be arranged in a consistent conformation within a lattice. This, however, may not be possible if the protein is flexible, or exists in multiple conformations (45, 46). Though, X-ray crystallography has been a routine technique for uncovering structural information of important biomolecules and cellular processes, the technique is limited by the growth of well-ordered 3D crystals.. Therefore complementary techniques such as cryo-EM are becoming more popular with recent advances in detectors, allowing low angstrom resolutions to be achieved, without the need for generating prior crystals (43, 45, 46)

Instead of diffraction of crystals, Cryo-EM determines structures by computationally combining images of many individual macromolecules in identical or similar conformations (43, 46). In this technique, samples of purified molecules in solution are applied to an electron microscopy (EM) grid covered with a thin holey carbon film and blotted by a filter paper to remove majority of solution. With the molecules undergoing a plunge-freezing process, the molecules embedded in a line layer of vitreous ice are imaged using an electron microscope. The final 3D reconstructions is a density map that can be inferred in the same way as electron density maps determined by X-ray crystallography(46). The major advantage of Cryo-EM over X-ray crystallography is instead of ordered crystals, molecules in solution are used (43, 45, 46).

Khayat et al (2011) published the 2.3 angstrom structure of PCV-2 providing the first structural description for a *Circovirus* family member. The choice of a PCV-2 consensus sequence was made in order to pursue the crystal structure for what was most representative of the PCV-2 population (44). Other than BFDV and PCV-2 (1), evidence of crystal structures of other viruses incorporated in *Circoviridae* family is lacking in published journal articles.

1.5 Phylogenetic Relationships

Members of the family *Circoviridae* are phylogenetically different from family members of *Anelloviridae,* figures 1.5 and 1.6 respectively. Both families comprise of animal viruses containing circular ssDNA genomes however there is some notable difference between the two families (6, 21, 23). *Circoviridae* are evolutionary related to

eukaryotic viruses that possess a circular ssDNA genome which encode a Rep protein. This Rep contains a conserved endonuclease and helicase (47-50). In contrast, *Anelloviridae*, include viruses from a separate lineage of ssDNA replicons based on genomics and structural features (6, 23)

Distinct differences also occur within the *Circoviridae* family resulting in two genera for further classification (6, 51, 52). *Circoviridae* viruses have been identified in vertebrates and invertebrates, though in most cycloviruses the definitive host is unknown (21). The Rep protein in the circovirus genomes is encoded on the virion sense strand and the Cap protein is the complementary strand, whilst the gene orientation is opposite in cycloviruses. Additionally, the intergenic region at the 3' end of the ORFs in the circovirus genome is relatively larger when compared to those in cyclovirus genomes. Furthermore, in the majority of cycloviruses, the genes encoding the Rep and Cap proteins overlap at their 3' ends (53, 54). It should be noted that *Cyclovirus* genomes are not grouped by the type of organism from which they are identified (19, 21, 26).

The *Anelloviridae* family, containing 10 genera, houses a progressively divergent variety of viruses of many sizes and pathogenicity. Complete genomes range phylogenetic tree enables identification of genera. Based on the ICTV taxonomic classification, Anelloviridae must not diverge more the 56% from genus and 35% from species in order to be classified in this family(23).

A Maximum likelihood (ML) phylogenetic tree was constructed using publicly available genome isolated for, from GenBank, to show the relationship between study target viruses and other viruses within the families *Circoviridae* and *Anelloviridae,* therefore highlighting why the specific viruses; CAV, PCV-3, BCV and AfaCV were chosen for this study (Figure 1.5). Due to CAV in the past being classified in *Circoviridae*, until only recently, it seemed the most appropriate candidate for studying the similarities and difference between the two ssDNA animal virus families, Circoviridae and Anelloviridae. Starfish circovirus (AfaCV) with its lack of research and highly divergent genome was chosen to include in the study as more research into this virus may result in it too being reclassified. Both viruses PCV-3 and CAV have economic importance as they both infect production animals. The recent founding of PCV-3 overseas highlighted the need for research to be conducted to increase knowledge of this virus and like PCV-1 and PCV-2, develop a vaccine in the near future. Of the 27 circoviruses classified to date, BCV makes up 8 of those viruses within the *Circoviridae* family. This shows the highly divers nature of the virus within a family.

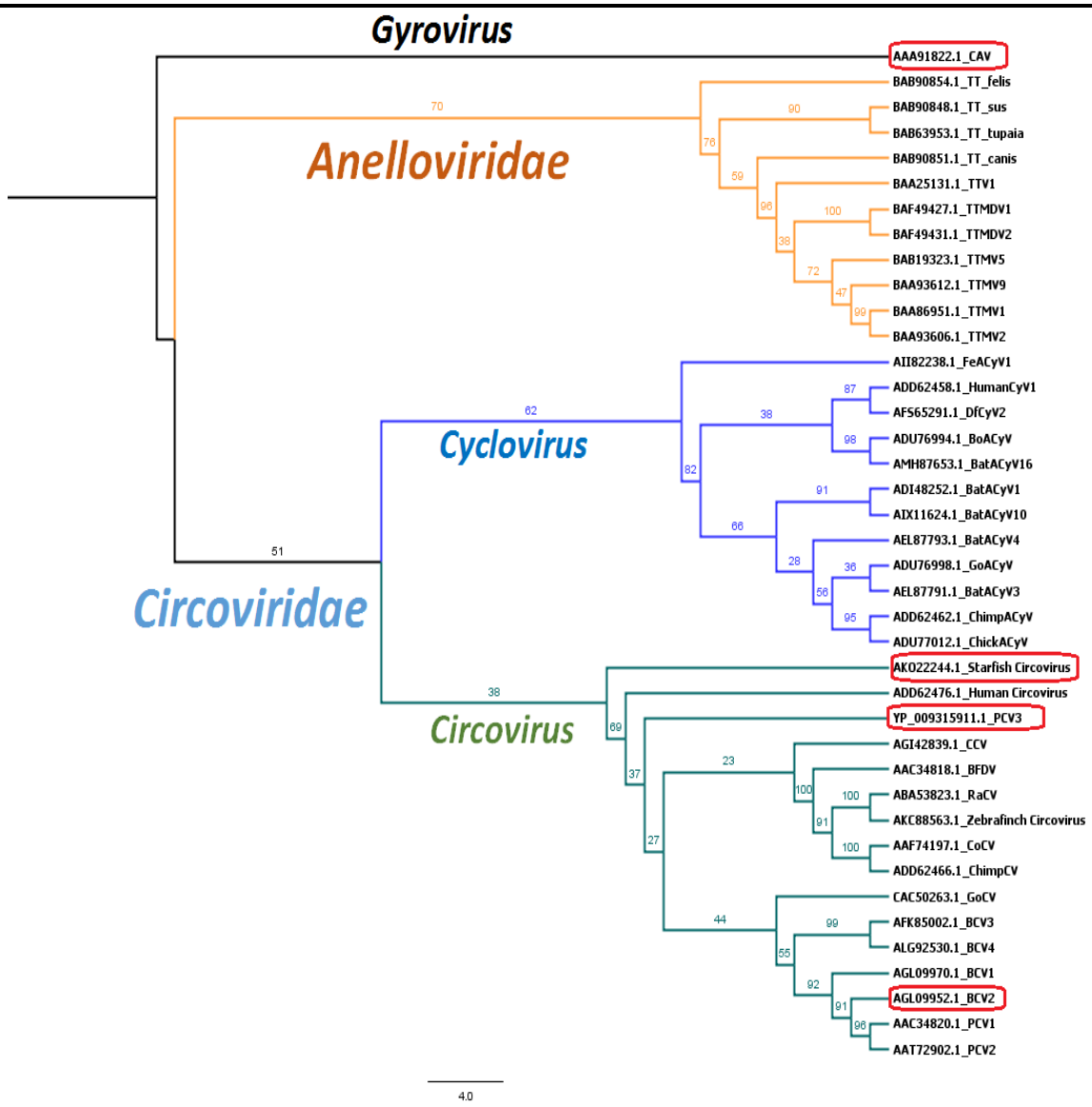


Figure 1.5 Phylogenetic tree representing relationship between *Anelloviridae* and *Circoviridae*

Virus species circled in red indicate the viruses which are included in this study. The numbers that are depicted on the branches are bootstrap percentages that are the designated confidence of the topographical relationship between particular virus species. Number next to the virus name is the Genbank accession numbers.

1.5.1 Porcine Circovirus (PCV)

Porcine circovirus associated diseases (PCVAD) are clinically displayed by post weaning mulisystemic wasting syndrome (PMWS), reproductive failure, respiratory and enteric disase and porcine dermatitis and nephrophathy syndrome (PNDS) (55). Unlike PCV-2 and PCV-3, PCV-1 is not attributed to a disease (4) Since its discovery in the early 1900s, with seroprevalance levels near 100% Porcine Circovirus 2 has been one of the most present swine viruses within the domestic population (56). Porcine Circovirus 2 is the etiological agent of PMWS, and is an essential component of PCVAD, though its role in PNDS is yet unknown (55, 57, 58). In a recent study by R. Palinski et al. (55), PCV-3 was identified in sows that had died acutely with PDNS-like clinical signs .

Divergence and genomic similarities of the capsid and replicase proteins of all 3 porcine circoviruses has recently been uncovered (59). The genomes of PCV-1 and PCV-2 are similar, with 80% homology in the origin of replication and 82% homology of Rep gene. However, there is a high amount of sequence diversion found in the Cap genes with less than 62% homology between PCV-1 and PCV-2 (4). Whereas the cap and rep of PCV-3 is only 37% identical to that of PCV-2. Interestingly, the cap and rep proteins of PCV-3 are 55% identical to those of bat associated circovirses (55). The genomic organisation of the 3 porcine circoviruses are represented in Figure 1.7 in schematic form.

In a recent study by R. Palinski et al. (55), it was found that PCV-3 is commonly found within the US swine population and may play an etiologic role in both reproductive failure and PDNS. Within the study immunohistochemistry (IHC) analysis of sow tissue samples found PCV-3 antigens in kidney, skin, lung and lymph node samples localised in typical PDNS lesions (55). Due to the high economic impact of PCV-2, the novel circovirus PCV-3 warrant further study to clarify its significance and role in PCVAD.

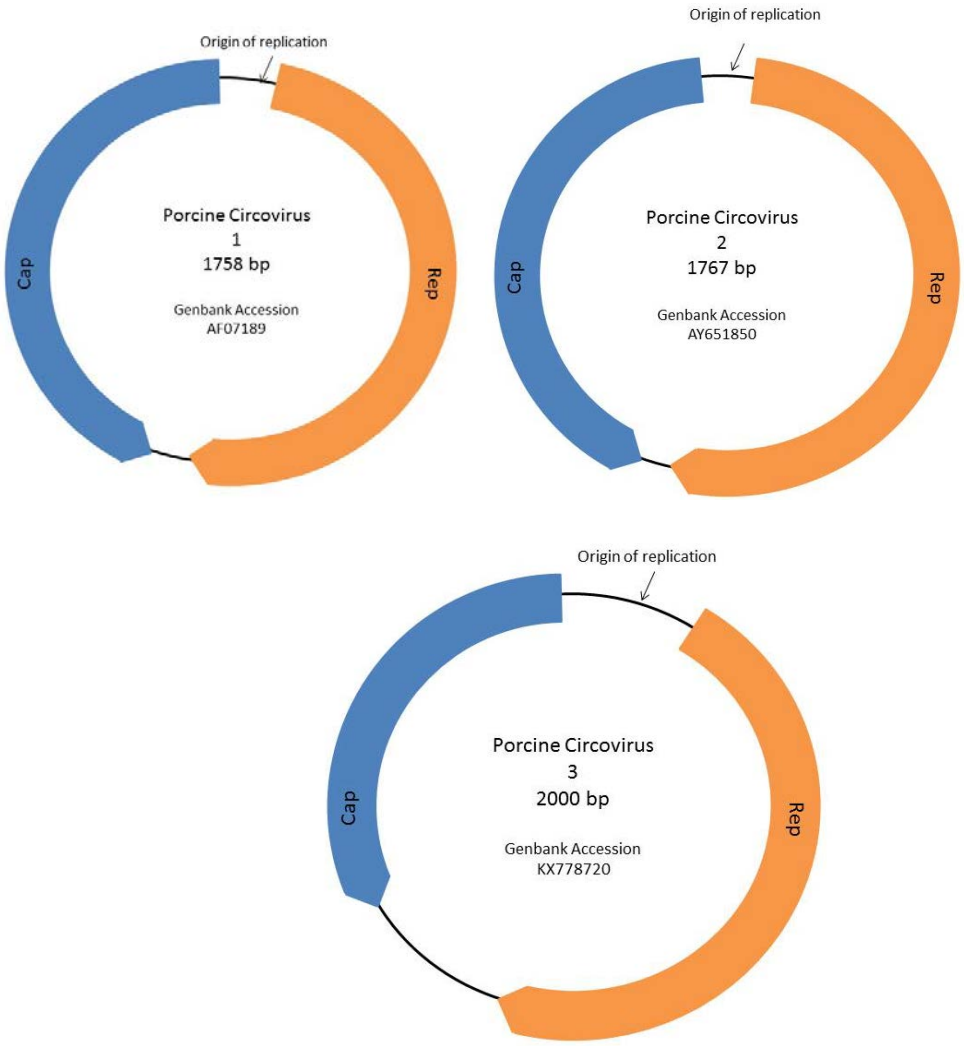


Figure 1.6 Schematic genome organisations of PCV-1, PCV-2 and PCV-3

Figure shows replication protein (Rep) and capsid protein (Cap). Genome is ssDNA for PCV-1, PCV-2 and PCV-3 with lengths of 1758 bp, 1767 bp and 2000 bp in size, respectively (54, 59). The GenBank accession numbers for the individual viruses are depicted within the specific diagrams.

1.5.2 Starfish associated circular virus (AfaCV)

Seas stars (*Asterozoa*) are keystone predators in marine enviroments, essential for structuring invertebrate communities ciritical for ecosystem health. Asterias forbesi-

associated circular virus (AfaCV), a ssDNA virus was discovered in a Forbes sea star displaying symptoms of sea star wasting disease (SSWD). The genome organisation of AfaCV is similar to that of members of the family *Circoviridae*, as it displays circular ssDNA encodin a Rep protein (60). Surveys based on PCR and viral metageonmics indicate that AfaCV is not clearly associted with SSWD, but represents the first circular Rep-enconding ssDNA (CRESS-DNA) virus detected in echinoderms (60, 61). The presence of the Cap and Rep protein shown in Figure 1.8 display a clear link to the *Circoviridae* family (62). Future studies are needed to determine the prevalance of AfaCV and other CRESS-DNA viruses amongst other sea star species and aquatic animals.

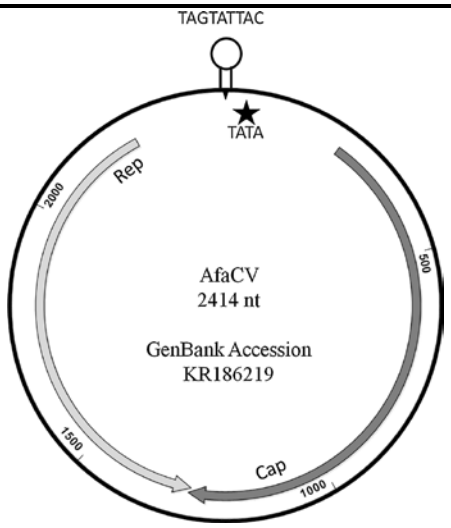


Figure 1.7 Schematic genome organisation of Asterias forbesi-associated circular virus (AfaCV)

Figure shows replication protein (Rep) and a putative capsid ORF (Cap). Genome is ssDNA of approximately 1,206 base pairs (60).

1.5.3 Bat Associated Circovirus (BCV)

Bats (order *Chiroptera*), the second most diverse group of mammals, and the only flying mammal, are known natural reservoirs of many emerging viruses (63, 64). Many diseases carried by Bats can potentially be passed on and cause disease in humans and other animals (63-65). During recent disease studies involving faecal collection from bats, five full-length circular ssDNA genomes were recovered. The five sequences shared a similar genomic organisation to *Circovirus* or more accurately the recently proposed *Cyclovirus* genera of the *Circoviridae* family (Figure 1.9) (66).

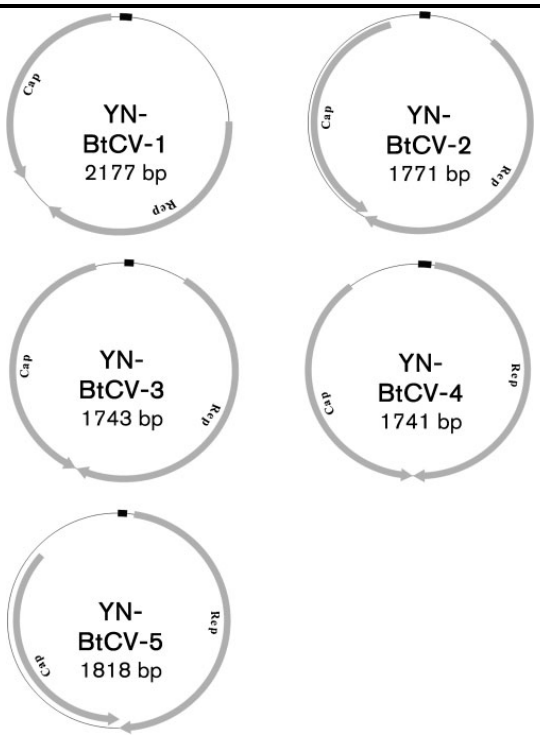


Figure 1.8 Schematic organisation of the five novel ssDNA genomes recovered from bat faeces in China

The two major ORFs encoding the Rep and Cap proteins are shaded. Locations of the potential stem-loop structures are marked by a filled box (66).

Cycloviruses are characterised by organised genomes that contain two major ORFs in opposite directions, (65). Whereas, characteristics of the genera *Circovirus* include, C*ap* located in the positive strand and the larger *Rep* located on the minus strand. Studies show that of the five ssDNA genomes recovered from bat faecal samples, some sequences are characteristic of *Cyclovirus* whereas others are characteristic of *Circovirus* (64, 67)

The highly basic arginine-rich domain (ARM) at the N terminus, which is typical of *Circovirus* Cap proteins, was found in YN-BtCV-2, -3, -4 and -5. The genomic structures of YN-BtCV-2, -3, -4 and -5 were similar to the newly proposed genera cyclovirus (66, 67). Further studies are required however to confirm X. Ge et al. (66) experimental findings.

1.5.4 Chicken anaemia virus (CAV)

Anelloviridae Gyrovirus, Chicken anaemia virus (CAV) has a high incidence of virus infection however, clinical disease is uncommon. Young chicks become infected with vertical (egg) transmission from infected breeders. Breeders do not show clinical signs of infection due to age resistance (68). First signs of acute disease can be seen in progeny birds between 7 and 14 days old (10, 69). Associated with lymphoid depletion, immunosuppression and developmental abnormalities are common of CAV infection (70). Severe anaemia associated with CAV is of concern due to the immunosuppressive nature of this clinical sign (71).

The viral genome consists of a circular, non-enveloped negative sense ssDNA of 2.3 kb that contains three major ORFs encoding three protein; VP1, VP2 and VP3 (Figure 1.10) (10, 72). Viral protein 1 (VP1) is 51kDa in size and acts as a capsid protein whilst VP2 is 24 kDa in size and acts as a structural/scaffold protein essential in the assembly of the virus (72, 73). Viral protein 3 (VP3), also referred to as apoptin is 13kDa in size and is important in cell pathogenesis and apoptosis (68, 72, 74).

Chicken anaemia virus (CAV) induces cell death of chicken lymphoblastoid T cells via apoptosis. Immunoassays and DNA analysis have confirmed that VP3 of CAV genome is in fact expressed in CAV-infected cells (74). Studies conducted have also shown that apoptin (VP3) is involved in CAV replication. This has been confirmed through the trial of Virus-like Particle (VLP) production of CAV with and without VP3 present (68).

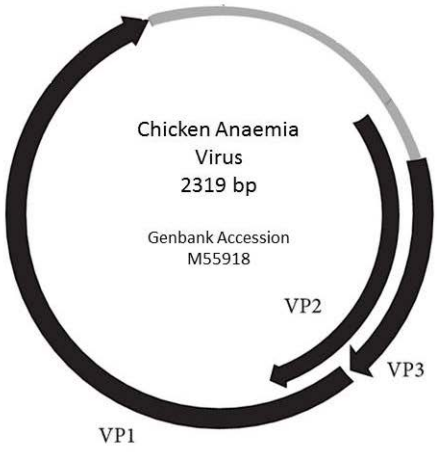


Figure 1.9 Schematic genome organisation of Chicken Anaemia Virus (CAV)
Three major ORFs, VP1, VP2 and VP3 are displayed above. Chicken Anaemia Virus genome is 2319 bp in size (75).

1.6 Mortality

Mortality rates vary widely between members of the both genera *Circovirus* and *Gyrovirus*. Mortality rates of Porcine Circovirus 2 are up to 15%, believed to be much higher for Porcine Circovirus 1 related diseases due to PCV-1 not being associated with clinical signs (10). Little is known about the mortality rates of both Bat and Starfish circoviruses. This may be due to more emphasis being placed on the *Circoviridae* infected production animals due to the associated economic loss. In *Gyrovirus*, chicken anaemia virus reportedly has a mortality rate of 10% though some believe this to be higher due to the increased immune suppression and increase chance of secondary infection by another virus (10, 71).

1.7 Transmission, Diagnosis &Treatments

Circoviruses and Gyroviruses are host specific or are believed to have a narrow host range with faecal-oral transmission being the most common. However, vertical transmission has been reported in some cases of infection (4, 51). In general, diagnosis of *Circoviridae* diseases is based on detections of the virus via culture, polymerase chain reaction (PCR) in situ hybridisation or detection of antibodies. However, some clinical signs may aid in diagnosis (4, 10, 76, 77). Development of enzyme-linked immunosorbent assay (ELISA) for the detection of serum with CAV antibodies has been achieved but is not widely used due limitations of the test. The test is specific to CAV therefore limits other condition that can be tested in the same ELISA (78). Earlier in 2017, multiplex quantitative real-time polymerase chain reaction was found to be a rapid

and reliable method for the detection and differentiation of PCV-2 and PCV-3 in the field (79).

There are few known treatments for *Circoviridae* infection in animals therefore prevention and/or vaccinations represent the best treatment method for all (4, 58, 80).

1.8 Prevention methods

Prevention of infection with CAV can be achieved with the build-up of antibodies within the breeder flocks before they come to an age of point of lay. This will avoid vertical transmission of the virus and should avoid faecal oral transmission as birds will possess maternal derived antibodies (MDA) at hatch. Seroconversion is a period of time in which antibodies develop and become detectable. This can be achieved through the use of live vaccines (10).

Prevention of circoviruses, such as PCV-1 and 2 has reportedly only been achieved through vaccination or sows and piglets. Introduction of PCV-2 vaccines into North America in 2006 and consequently other countries has resulted in a dramatic fall of mortality associated with PCV-2 (56, 81). Concern; however have been raised over the change in genotype of the virus, believed to be as a result of the vaccine. This has been shown by published works and field work showing the failure of the PCV-2 vaccine in light of the emerging genotypes. Some believe this failure in vaccine and adaptation of the virus is a result of conditioning of the virus due to the widespread use of the PCV-2 vaccine(56)

1.9 Conclusion and project aims

The main objective of the following experiments was to:

1. Characterise the structure of BCV X with and without DNA using crystallography
2. Design and test recombinant expression constructs of Starfish Cap protein
3. Test recombinant expression of PCV-3
4. Test recombinant expression of CAV

Structural and Functional Insights into Animal Circoviruses and Gyroviruses

Bec Barnewall

2.0 Research Paper

2.1 Introduction

The virus families of *Circoviridae* and *Anelloviridae* possess a circular genome, enclosed in a small and simple isohedral capsid (1-3). Present worldwide ad infecting a large range of host species, new virus specie are constantly being discovered and assigned to these families (6-9).Viruses from both families can infect a range of animals, both domestic and wild, and the economic impact associated with *Circoviridae* and *Anelloviridae* diseases are significant (6, 10) There are few known treatments for *Circoviridae* and *Anelloviridae* infection in animals therefore prevention and/or vaccinations represent the best treatment method for all (4, 58, 80).

The recent reclassification of the Circoviridae family into two genera, Circovirus and Cyclovirus highlights out ever-growing knowledge on virus diversity within these families. Prior to 2017, Gyrovirus were grouped within the Circoviridae family(3-5), however, this genus was recently reassigned under the Anelloviridae family (6). The genomic organisation of Circoviruses, Cycloviruses and Gyroviruses is represented in Figures 1.1, 1.2 and 1.3 respectively.

This study will focus on the expression and purification of the structural capsid protein in animal viruses. These viruses will include; Chicken Anaemia Virus (CAV), Bat Associated Circovirus (BCV), Porcine Circovirus 3 (PCV-3) and Asterias forbesi-associated circular virus (AfaCV). The literature review provides an assessment and

analysis of a wide range of studies that have been undertaken in relevant areas of virology, and structural biology as well as identifying gaps in the literature that will be addressed in this study.

2.2 Materials and Methods

2.2.1 Materials

All media, buffers and solutions were prepared using reverse osmosis, deionised water (See Table 2.1 and Table 2.2). Unless otherwise stated all chemicals and reagents were purchased from Sigma-Aldrich. Equipment used is outlines in Table 2.3. All proteins were inserted in the vector outline in Table 2.4.

Table 2.1 Buffers used for protein analysis and purification

Buffer	Composition	pH (if applicable)
Caps Buffer A	20mM CAPS; 500mM NaCl; 50mM Imidazole	10.5
Caps Buffer B	20mM CAPS; 500mM NaCl; 500mM Imidazole	10.5
Caps Buffer C	20mM CAPS, 500mM NaCl; 200mM L-Arginine	10.5
Phosphate Buffer	50mM Phosphate Buffer (0.0466 moles Na_2HPO_4 + 0.0034 moles NaH_2PO_4)	
HIS (Nickel) Buffer A	50mM Phosphate buffer; 300mM NaCl; 20mM Imidazole	8
HIS (Nickel) Buffer B	50mM Phosphate buffer; 300mM NaCl, 500mM Imidazole	8
GST Buffer A	50mM Trizma; 125mM NaCl	8
Exchange Buffer	0.5M Trizma, 125mM NaCl	pH to difference pH with HCl and NaOH
Caps Buffer A+ L-arginine	20mM CAPS; 500mM NaCl; 50mM Imidazole;500mM L-arginine	10.5

8M Urea	8M Urea	
GuHCl	1M GuHCl, 0.5% Triton	
Caps Buffer A high salt	20mM CAPS; 500mM NaCl; 50mM Imidazole;1.5M NaCl	10.5
HIS (Nickel) Buffer A + 8M Urea	50mM Phosphate buffer; 300mM NaCl; 20mM Imidazole; 8M Urea	
GST Buffer A + 2M Urea	50mM Trizma; 125mM NaCl; 2M Urea	
HIS (Nickel) Buffer A + 200mM L-Arg	50mM Phosphate buffer; 300mM NaCl; 20mM Imidazole; 200mM L-Arginine	
6M GuHCl	6M GuHCl	
1M GuHCl solubilisation buffer	1M GuHCl, 0.5% Triton, 50mM Trizma, 250mM NaCl	
GST Buffer A + Urea + SDS	50mM Trizma; 125mM NaCl; 8M Urea; 1% SDS	8
GST Buffer A + 1% SDS	50mM Trizma; 125mM NaCl; 1% SDS	8
HIS (Nickel) Buffer A + 1% SDS	50mM Phosphate buffer; 300mM NaCl; 20mM Imidazole; 1% SDS	8
HIS (Nickel) Buffer B + 1% SDS	50mM Phosphate buffer; 300mM NaCl, 500mM Imidazole; 1% SDS	8
20× BOLT© MES SDS Running Buffer	1-5% Sodium Lauryl sulphate	
Biorad Precision Plus Protein ™ Standards	30% (w/v) glycerol; 2% SDS; 62.5mM Trizma; 50mM DTT; 5mM EDTA; 0.02% NaN_3; 0.01% Bromophenol blue	6.8

Table 2.2 Reagent Stock Solution Compositions

Stock Solution	Composition
20× NPS	0.5M(NH4)2SO4; 1M KH2PO4; 1M Na2HPO4
50×5052	10% glucose; 10% αNlactose; 15% glycerol
NaCl	2.852M NaCl
Expression Media/LB Media	5% Tryptone; 2.5% Yeast
Coomassie Blue Stain	0.2% Coomassie brilliant blue R-250; 40% Ethanol; 10% Glacial acetic acid
Destain Solution	10% Glacial acetic acid; 10% Ethanol

Table 2.3 Equipment

Equipment	Use
GE FPLC	Affinity and size exclusion chromatograph; Fast protein liquid Chromatography
Beckman Centrifuge	Centrifuge to separate samples
Ratek Orbital	Aeration and mixing of samples
Orbital Incubator	Temperature control for starter culture with mixing
-80˚C Freezer	Storage of samples; purified proteins, competent cells, glycerol stocks
-20˚C Freezer	Storage of samples; cell pellets, plasmids stocks, reagents, soluble cell extracts
Precast 4-12% BIS TRIS gel	SDS-PAGE
BOLT© Mini Gel Tank	SDS-PAGE

Australian Synchrotron Puck	Transport of protein crystals to Australian Synchrotron and storage of samples at beam line
Tomy Autoclave	Sterilisation of rubbish, tips, tubes, flasks, buffers and reagents
Bio Rad gel imaging system	Imaging of SDS-PAGE gels
Heat block	Denaturation of protein samples to be analysed by SDS-PAGE
Vortex	Mix solution
Agar Plates	Plate out transformed cells in order to isolate colonies
Spreaders	Spread transformed cells onto agar plates
23˚C Incubator	For crystal plates
37˚C Incubator	For growth of transformed cells overnight

Table 2.4 Primary Structure of Target Proteins

Protein	Amino Acid Sequence	kDa	pI
BCV NLS (1-243)	MVYRRRRGRGRRARPMSSLGRLLYRKPWLMHPRFRARYRWRRKN GITNLRLTRQVELWVPKDAANASFYVNHYTFDLDDFIPAGTQLNSSP LPFKYYRIRKVKVEFQPRLPITSPFRGYGSTVPILDGAFVTPATGESD PIWDPYINFSGRHVIRTPAWYHKRYFTPKPLIDGNTGFFQPNNKQNA LWFPNKQGQNIQWSGLGFAMQKGNEAYNYQVRFTLYVQFREFDLF NNKYTAHMDVPL	29	10.63
BCV X (42-243)	KNGITNLRLTRQVELWVPKDAANASFYVNHYTFDLDDFIPAGTQLNS SPLPFKYYRIRKVKVEFQPRLPITSPFRGYGSTVPILDGAFVTPATGE SDPIWDPYINFSGRHVIRTPAWYHKRYFTPKPLIDGNTGFFQPNNKQ NALWFPNKQGQNIQWSGLGFAMQKGNEAYNYQVRFTLYVQFREFD LFNNKYTAHMDVPL	23	9.54
CAV VP1 (1-447)	RRARRPRGRFYAFRRGRWHHLKRLRRRYKFRHRRRQRYRRRAFR KAFHNPRPGTYSVRLPNPQSTMTIRFQGVIFLTEGLILPKNSTAGGYA DHMYGARVAKISVNLKEFLLASMNLTYVSKLGGPIAGELIADGSKSQ AAENWPNCWLPLDNNVPSATPSAWWRWALMMMQPTDSCRFFNH PKQMTLQDMGRMFGGWHLFRHIETRFQLLATKNEGSFSPVASLLSQ GEYLTRRDDVKYSSDHQNRWRKGEQPMTGGIAYATGKMRPDEQQ YPAMPPDPPIITSTTAQGTQVRCMNSTQAWWSWDTYMSFATLTALG AQWSFPPGQRSVSRRSFNHHKARGAGDPKGQRWHTLVPLGTETIT	52	10.64

	DSYMGAPASEIDTNFFTLYVAQGTNKSQQYKFGTATYALKEPVMKS DSWAVVRVQSVWQLGNRQRPYPWDVNWANSTMYWGSQP		
CAV VP1 X (51-447)	PGTYSVRLPNPQSTMTIRFQGVIFLTEGLILPKNSTAGGYADHMYGA RVAKISVNLKEFLLASMNLTYVSKLGGPIAGELIADGSKSQAAENWP NCWLPLDNNVPSATPSAWWRWALMMMQPTDSCRFFNHPKQMTLQ DMGRMFGGWHLFRHIETRFQLLATKNEGSFSPVASLLSQGEYLTRR DDVKYSSDHQNRWRKGEQPMTGGIAYATGKMRPDEQQYPAMPPD PPIITSTTAQGTQVRCMNSTQAWWSWDTYMSFATLTALGAQWSFP PGQRSVSRRSFNHHKARGAGDPKGQRWHTLVPLGTETITDSYMGA PASEIDTNFFTLYVAQGTNKSQQYKFGTATYALKEPVMKSDSWAVV RVQSVWQLGNRQRPYPWDVNWANSTMYWGSQP	45	9.44
CAV VP1 XL (MMC aa 30-42 + CAV VP1 X 51-447)	YRRRRRHFRRRRF[51]PGTYSVRLPNPQSTMTIRFQGVIFLTEGLILPK NSTAGGYADHMYGARVAKISVNLKEFLLASMNLTYVSKLGGPIAGEL IADGSKSQAAENWPNCWLPLDNNVPSATPSAWWRWALMMMQPTD SCRFFNHPKQMTLQDMGRMFGGWHLFRHIETRFQLLATKNEGSFS PVASLLSQGEYLTRRDDVKYSSDHQNRWRKGEQPMTGGIAYATGK MRPDEQQYPAMPPDPPIITSTTAQGTQVRCMNSTQAWWSWDTYM SFATLTALGAQWSFPPGQRSVSRRSFNHHKARGAGDPKGQRWHT LVPLGTETITDSYMGAPASEIDTNFFTLYVAQGTNKSQQYKFGTATY ALKEPVMKSDSWAVVRVQSVWQLGNRQRPYPWDVNWANSTMYW GSQP	47	9.95
PCV-3 (1-214)	RHRAIFRRRPRPRRRRRHRRRYARRRLFIRRPTAGTYYTKKYSTMN VISVGTPQNNKPWHANHFITRLNEWETAITFEYYKILKMKVTLSPVIS PAQQTKTMFGHTAIDLDGAWTTNTWLQDDPYAESSTRKVMTSKKK HSRYFTPKPLLAGTTSAHPGQSLFFFSRPTPWLNTYDPTVQWGALL WSIYVPEKTGMTDFYGTKEVWIRYKSVL	25	10.93
AfaCV NLS (1-347)	MPRNTRRRTGRSRRRRTLRRGYRKFTGGRGITKEHDRTNIYYKKRM PRFKKRRWRRFSKKVQNVSEKDLGTQTVVFNDQIETGDFFNSDGLI TGSATLCLYSMNGSEVPYNDLAVMMNNLNFAAPTVDLGTTKYRTSK VIFKSAVFDLTLTNTSYDRIGGVNNIEAAYSLEVDIHTVVIKKGARDLD FISGTLAEKGLKTIDKMIEVAEKEVRPIGGDDSKIFNTTKAKFNRGGTL WDVPHALSLFGMKILSKKKYFIPGGATITYQMRDAKRRVMVPQFNM LGVNKPGWTKFLYITYKLVPGVGVIGQDENVPTQIRTRLSVGVTRKY SYKIEGQNDVRDNYNPDL	39	10.18

Table 2.5 Description of Vector

Vector	Description
pMCSG21	Vector encodes spectinomycin resistance, T7 promoter and lac operon. Also encodes N-terminal HIS tag with TEV cleavage site

Table 2.6 E. coli strain used

Escherichia coli strain	Description
Rosetta 2 (DE3)	Cell has chloramphenicol resistance and expression is driven by a T7 promoter which is repressed until IPTG induction

2.2.2 Methods

2.2.2.1 DNA methods

2.2.2.1.1 Transformation of competent E. coli cells

Transformation of competent *E. coli* [Rosetta 2(DE3)] (Novagen, USA) cells was achieved through the heat shock method. One microliter (1 µL) of plasmid vector was added to 50 µL of competent cells and incubated on ice for 30 min. Cells were then placed in a water bath at 42 ºC for 45 s, to achieve heat shock, and returned to ice for a further 2 min. Two hundred and twenty five microliters (225 µL) of room temperature Luria-Bertani (LB) media was added to the cells and this suspension was incubated 1 h at 37 ºC and 225 rpm. After recovery, 50 µL of transformed cells were selectively cultured on LB agar containing the antibiotic that complemented the resistance of the plasmid vector. Agar plates were then incubated overnight at 37 ºC.

2.2.2.1.2 Amplification and Extraction of Plasmids

Extraction of plasmids occurred from QIAprep Miniprep kits as per manufacturer's instructions.

2.2.2.2 Large-scale Recombinant Protein Expression Methods

In order to produce the large amount of protein required for crystallography, recombinant DNA technologies and expression methods were utilised. Two methods of protein expression procedures were utilised. This was due to the difference in protein yields obtained from different protein.

2.2.2.2.1 Starter cultures

In order to perform large scale expression, high cell densities in the log growth phase are required to be inoculated into expression media in large baffled flasks. High cell numbers was achieved by starter culture that was prepared using single colony from transformation plates. Isolated colonies were extracted from the transformation plate by touching with a pipette tip. This tip was then placed in a 50 mL falcon tube containing 5 mL LB broth containing the appropriate antibiotic. This starter solution was grown overnight at 37 ˚C at 225 rpm.

2.2.2.2.2 Auto-Induction Methods

The auto-induction method automatically induces protein expression. This is achieved as *E. coli* cells saturate the growth media and convert energy source from glucose to lactose. As suggested by the name, this method does not require constant monitoring, addition of inducing agents nor heating to initiate expression. That is expression will commence automatically as *E. coli* cell take up lactose as an energy source. This results in inhibition of the repressor at the lac operon sequence, allowing T7 polymerase to commence gene expression.

For an overnight culture, 500 µL of the appropriate antibiotic stock was added to each 2 L sterile baffled flask. To which; 25 mL of 20× NPS, 10 mL of 5052, 500 µL of $MgSO_4$ and 470 mL of LB expression media were also added. The flasks were covered with sterile foil and incubated at room temperature on a platform orbital for 24 h at 90 rpm.

The cells were then harvested via centrifugation at 18˚C and 6000 rpm, with the supernatant decanted into waste media bottles and autoclaved. The cell pellet is resuspended in buffer corresponding to the affinity tag encoded by the plasmid, i.e. HIS-tagged protein were resuspended in HIS (Nickel) Buffer A, Bat Circovirus is an exception and is resuspended in Caps Buffer A. Resuspended cells were places in a 50 mL Falcon, so that each Falcon tube contained cells resuspended from 2 L of expression media. Tubes were appropriately labelled and stored in -20 ˚C freezer.

2.2.2.2.3 IPTG Method

Where production of high yields was not successful through auto-induction method as described in 2.2.2.2.2, isopropyl-β-D-thiogalactopyranoside (IPTG) methods was used. This technique involved induction of protein expression by addition of IPTG at an optical density (OD) of 0.6-0.8.

For an overnight culture, 500 µL of the appropriate antibiotic stock was added to each 2 L sterile baffled flask. To which 30 mL of NaCl (5 g/30 ml) and 470 mL of LB expression media were also added. The flasks were covered with sterile foil and incubated on a platform orbital at 37 ˚C and 90 rpm until the content reached an OD_{600} between 0.6 and 0.8. Protein expression was then induced by addition of 500 µL of 1 M IPTG. This is an inducer that binds the lac repressor and enables the T7 polymerase to being protein expression. Induced cells were incubated at room temperature and 90 rpm overnight.

Similar to the auto-induction methods, following expression, the cells were then harvested via centrifugation at 18 ˚C and 6000 rpm, with the supernatant decanted into waste media bottles and autoclaved. The cell pellet is resuspended in buffer corresponding to the affinity tag encoded by the plasmid, i.e. HIS-tagged protein were resuspended in HIS (Nickel) Buffer A, Bat Circovirus is an exception and is resuspended in Caps Buffer A. Resuspended cells were places in a 50 mL Falcon, so that each Falcon tube contained cells resuspended from 2 L of expression media. Tubes were appropriately labelled and stored in -20 ˚C freezer.

2.2.2.3 Protein Purification Methods

2.2.2.3.1 Solubilisation of inclusion bodies (82)

The formation of inclusion bodies, especially in *E. coli,* is often the result of the protein being insoluble or fast expression that leads to improper folding and formation (83). Solubilisation of inclusion bodies can be achieved through addition of detergent to the buffer and altering the pH. Resuspended cells, retrieved from the -20˚C freezer were subject to two freeze/thaw cycles. Following this 1 mL of lysozyme (20 mg/mL, Sigma Aldrich, USA) was added to lyse the cell wall, followed by the addition of 10 µL DNase (5 µ/mL, Invitrogen, USA). This was added to degrade the DNA and reduce viscosity of the sample. Samples were incubated at room temperature for 30 mins. The sample extract was centrifuged at 12000 rpm for 30 min at 18 ˚C in order to remove cell debris. Supernatant was then tipped off into a 50 mL falcon tube and stored in the cold store. The cell pellet in the bottom of the centrifuge tube was then resuspended in a solubilisation buffer, see Table 2.1 (buffers Caps Buffer A + L-Arginine to His (Nickel)

Buffer B + 1% SDS). Resuspended pellet was then incubated at room temperature for at least 1 hr.

After incubation The sample extract was centrifuged at 12000 rpm for 30 min at 18 ˚C in order to remove cell debris. Further clarification of supernatant was achieved by 0.45 µL syringe filter. If successful solubilisation occurred, which was check with SDS-PAGE, the supernatant could then be loaded onto one of the below chromatography purifications.

During the purification process sample were removed for analysis by SDS-PAGE. In this case, 10 µL of the resuspended sample after incubation with DNase and lysozyme(WC) were added to eppendorf tube containing 90 µL of 2× Tris-Glycerine SDS-PAGE buffer. Ten microliters of pellets, that resulted after each centrifugation was also added to eppendorf tube containing 90 µL of 2× Tris-Glycerine SDS-PAGE buffer. A volume of 20 µL of all supernatants that were produced during the above solubilisation procedure were added to eppendorf tube containing 90 µL of 2× Tris-Glycerine SDS-PAGE buffer.

2.2.2.3.1.1Ethanol wash when using Guanidine Hydrochloride as a detergent (84)

Due to guanidine HCL precipitating with SDS, it is necessary to remove the former before undergoing a SDS-PAGE analysis of the sample. Removal is achieved by washing the sample with ethanol. A volume of 225 µl of cold (0˚-4˚C) 100% ethanol is added to 25 µl of sample, which contains the protein of interest. The sample is then mixed by vortexing, and chilled for 10 mins at -20˚C. Centrifuge the sample in

microcentrifuge tube at ~ 15,000 rpm for 5 mins, at 4˚C. Careful withdrawal of the supernatant after centrifugation is required as it may be difficult to see the pellet. Resuspend the pellet in 250 µl of 90% ethanol and mix thoroughly by vortexing. Centrifuge the sample in microcentrifuge tube at ~ 15,000 rpm for 5 mins, at 4˚C. The supernatant was then withdrawn by pipetting and the pellet resuspended in 2× SDS Buffer. The sample was then prepared, as per 2.2.2.5, for analysis by SDS-PAGE.

2.2.2.3.2 Nickel Affinity Chromatography for Purification of His-tagged Proteins

Purification of His-tagged proteins from *E. coli* extracts was achieved through nickel affinity chromatography. Firstly, suspended cells were subject to two freeze/thaw cycle. Following this 1 mL of lysozyme (20 mg/mL, Sigma Aldrich, USA) was added to lyse the cell wall, followed by the addition of 10 µL DNase (5 µ/mL, Invitrogen, USA). This was added to degrade the DNA and reduce viscosity of the sample. Samples were incubated at room temperature for 30 mins. The sample extract was centrifuged at 12000 rpm for 30 min at 18 ˚C in order to remove cell debris. Further clarification of supernatant was achieved by 0.45 µL syringe filter.

Using a superloop and AKTA FLPC the soluble cell extract was then injected onto a pre-equilibrated HIS-trap column, in HIS (Nickel) buffer A. Washing of the column with 10 column volumes of HIS (Nickel) buffer A removed unbound proteins which was then eluted with by applying a gradient elution of HIS (Nickel) buffer B for 5 column volumes. Fractions of eluted protein were assessed by Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and combined into a 50 mL Falcon tube.

During the purification process sample were removed for analysis by SDS-PAGE. In each case, 10 µL samples of whole cell (WC) were added to eppendorf tube containing 90 µL of 2× Tris-Glycerine SDS-PAGE buffer. For the soluble extract (SE) and flow through (FT), 10 µL samples were added to eppendorf tubes containing 40 µL 2 × Tris-Glycerine SDS-PAGE buffers. All remaining elution samples of 20 µL were added to 20 µL 2 × Tris-Glycerine SDS-PAGE buffers.

2.2.2.3.3 Nickel Affinity Chromatography for Purification of Bat Circovirus

Similar to the above protocol, Bat circovirus was purified by nickel affinity chromatography, though changes in buffers and reagents should be noted.

Freeze/thawing of Bat samples were not undertaken due to the addition of FastBreak Cell Lysis Buffer (1X; Promega. Following addition of 1 mL FastBreak Cell Lysis Buffer (1X; Promega), for every 10 mL cell suspension, 1 mL of lysozyme (20 mg/mL) was added to lyse the cell wall, followed by the addition of 10 µL DNase (5 µ/mL). This was added to degrade the DNA and reduce viscosity of the sample. Samples were incubated at room temperature for 30 mins. The sample extract was centrifuged at 12000 rpm for 30 min at 18 ˚C in order to remove cell debris. Further clarification of supernatant was achieved by 0.45 µL syringe filter.

Using a superloop and AKTA FLPC the soluble cell extract was then injected onto a pre-equilibrated His-trap column, in Caps Buffer A. Washing of the column with 10 column volumes of Caps Buffer A removed unbound proteins which was then eluted

with by applying a gradient elution of Caps Buffer B for 5 column volumes. Fractions of eluted protein were assessed by SDS-PAGE and combined into a 50 mL Falcon tube.

Samples were collected for analysis by SDS-PAGE as above (see 2.2.2.5)

2.2.2.3.4 Size Exclusion Chromatography

Size exclusion chromatography was undertaken to further purify sample. This was achieved with AKTA FPLC using s200 26/60 column from GE Healthcare. To achieve appropriate sensitivity a sample of no more that 12mL was loaded using a superloop. This was done onto a pre equilibrated column in GST A Buffer, at a flow rate of 2.5 mL/min. In the case of bat sample, Caps Buffer C was used to equilibrate and run size exclusion chromatography.

Size exclusion works on the principle of separation based on size of proteins. That is the larger proteins are eluted early whilst the smaller protein are slowed down as they get trapped within the bead of the column.

Fractions were collected once the UV level reached an absorbance of 20 mAU, with samples being collected for analysis by SDS-PAGE. Protein fraction samples of 20 µL were added to 20 µL 2x Tris-Glycerine SDS Page buffer. Collected fractions containing pure protein were combined and concentrated using a 10 kDa 15 mL Amicon centrifuge tube as per manufacturers' protocol. The samples was concentrated to concentrations of 10 mg/mL as measure on Nanodrop machine and stored in 25 µL aliquots at -80 ˚C.

2.2.2.4 Protein Crystallisation Methods

2.2.2.4.1 Sparse-matrix hanging-drop vapour diffusion method

Preliminary crystallisation trails were carried out via the sparse-matrix hanging-drop vapour-diffusion method. This was performed by adding 300 µL of commercial crystallisation screen into the corresponding reservoirs of pre-greased48 well plates supplied by Hampton Research. Crystal screen used included; Crystal Screen 1 (Hampton), Crystal Screen 2 (Hampton), Peg Ion 1 (Hampton), Peg Ion 2 (Hampton), Proplex 1 (Molecular Dimensions), Proplex 2 (Molecular Dimensions), Pact Premier 1 (Molecular Dimensions), Pact Premier 2 (Molecular Dimensions). The reservoir solution was mixed in a 1:1 ratio of protein to reservoir solution onto a coverslip and inverted over the corresponding reservoir.

In the case of Bat circovirus trial, DNA was mixed with purified Bat X (8 mg/mL) in a 1:2 ratio of DNA to protein. This mixture was then mixed in a 1:1 ratio of protein DNA solution to reservoir solution onto a coverslip. On the same coverslip a drop of protein (without DNA added) and reservoir solution mixed in a 1:1 ratio was also labelled with the coverslip then being inverted on the corresponding reservoir.

2.2.2.4.2 Optimisation of Crystal Screening

If small crystals were identified from sparse-matrix hanging-drop vapour-diffusion screening, optimisation plates were set as described in 2.2.2.4.1. The reservoir containing the crystal screen was however replaced with a solution containing modified

concentrations of the buffer, precipitant and salt. This was done in an attempt to improve the crystal quality, size and more importantly its diffraction.

2.2.2.4.3 Preparation of Protein crystal for X-ray Diffraction Data Collection

Cooling of crystals first required identification of a suitable cryoprotectant in order to prevent the crystal from being damaged by ice crystals. To find these suitable conditions, a cryoprotectant was screened based on the condition the protein crystal formed in, contained difference percentages of glycerol 20-30%. An appropriate cryoprotectant was chosen by using the minimum amount of glycerol required to protect the crystal during freezing, as too much glycerol can dissolve the crystal and/or affect diffraction. The protein crystal was then extracted from the hanging drop, by mounting on a Hampton crystal loop, placed in cryoprotectant and flash frozen with liquid nitrogen. The crystal loop, containing the protein crystal was then placed into an Australian Synchrotron crystal puck which was transferred to a liquid nitrogen transport dewar for transport to the Australian Synchrotron.

2.2.2.4.4 Collection, Processing and Structure Solving of X-ray Diffraction Data

Diffraction data was collected at the Australian Synchrotron on the MX1 and MX2 macromolecular crystallography beamline (source-03BM1 and 03ID1, respectively). Data was processed using iMosFlm (85), where the images of diffraction were indexed and integrated; and then scaled, phased and refined using CCP4 suite, using AIMLESS, Phaser and REFMAC respectively (86-88).

2.2.2.5 Protein Visualisation by SDS-PAGE

In order to assess the purity and composition of protein samples, SDS-PAGE was undertaken. Preparation of samples was achieved through addition of 2× SDS loading buffer with a 1:9 dilution for whole cell, 1:4 dilution for soluble extract and FPLC flowthrough and a 1:1 dilution for all other samples. Samples were then heated to 100 ˚C for 2 min using a heating block, then vortexed for 1 minute. Centrifugation was then undertaken for 1 min at 10,000 rpm to make sure sample was at the bottom of the eppendorf tube before loading.

A 12 well precast gel (4-12% BIS TRIS) was rinsed with deionised water ad clamped into a BOLT^{TM} electrophoresis unit. BOLT running buffer (Novex) of 15 mL was diluted in 300 mL deionised water. The solution was then poured into both side of the BOLT tanks, ensuring sample wells of the precast gel were covered with solution. Protein marker, Biorad Precision Plus Protein ^{TM} Standards (20 µL), was loaded into the first loaded into the first well. Protein samples were then loaded into subsequent well, 10 µL each. Once samples were loaded, electrophorese was undertaken for 30 min at 160 V.

Following electrophoresis, the gel was removed from the pre-cast plate and washed in deionised water to remove excess running buffer. In order to visualise the proteins, a staining process was then undertaken. The gel was stained by incubating with Coomassie stain for 5 min. The stain was then removed and the gel washed with deionised water to remove excess stain. The gel was places in de-stain solution for 30 min, with the de-stain buffer being replaced every 30 min until all background

Coomassie stain was removed, revealing protein bands. The gel was then place in deionised water. Imaging of the gel was achieved by using Bio Rad imaging system with the Coomassie stain protein gel setting.

2.3 Results

In order to characterise the structure of capsid proteins from Bat, Starfish, Porcine (PCV-3) and CAV, these genes encoding these proteins were synthesised and codon optimised by GenScript (Hong Kong). All target genes for the study were cloned into the pMCSG21 plasmid vector at *SspI* site (89)(see Figure 2.1) using ligation independent cloning (LIC) (91). Ligation independent cloning relies on the generation of large single-stranded overhangs (or “sticky ends”) by the T4 DNA polymerase. This method does not require specific DNA restriction on the gene insert or ligation and is highly efficient and cost effective (92).

Recombinant expression of the Cap targets was performed using transformed *E. coli* BL21 (DE3-Rosetta 2), and one of two different expression systems, IPTG or auto-induction, as per methods 2.2.2.2.3 and 2.2.2.2.2, respectively. In all cases, expression from the pMCSG21 vectors produced a recombinant 6-His fusion protein with a TEV cleavage site, allowing efficient purification of the target protein.

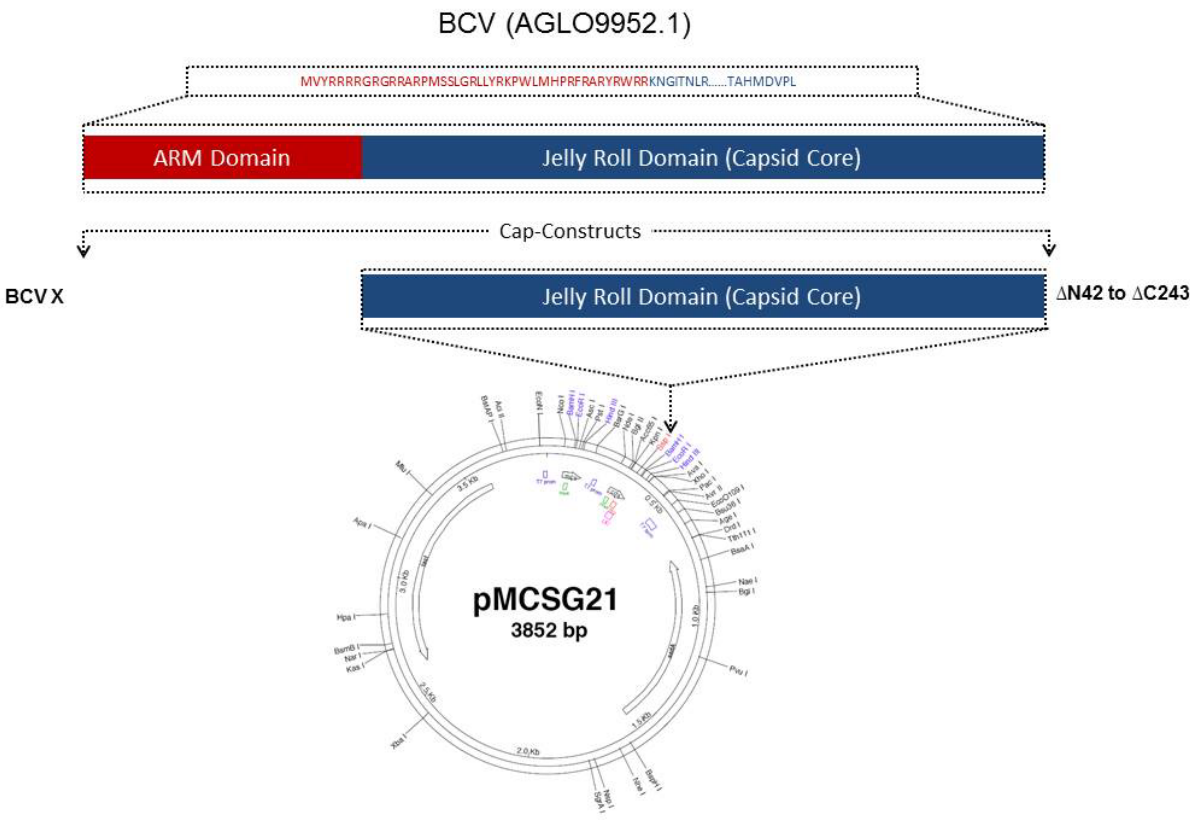


Figure 2.1 Schematic comparison of BCV and BCV X, and BCV X used as an example for cloning into pMCSG21 expression vector

Schemtic diagram of the Cap protein target in the BCV part of this study, cloned into the SSP1 cloning site of pMCSG21 for recombinant expression. The horizontal bar represents the difference between the full length construct and that used in the study. The original ARM domain of BCV is depicted in a red domain. Species Genbank accession number of the genome is shown at the top of the diagram. While the numbers on the right side denotes the length of each Cap protein.

2.3.1 Bat Associated Circovirus (BCV)

Recombinant expression of the BCV Cap protein was performed using optimised expression vector (pMCSG21), *E. coli* BL21 (DE3-Rosetta 2), and expression media IPTG as per method 2.2.2.2.3. This expression media was chosen as it had already been successfully used for expression of BCV construct in the laboratory. Prior to this study it was unknown if a capsid protein without an ARM domain would form a VLP in the presence of single-stranded DNA.

The solving of the full length BCV X-ray crystal structure in our laboratory in 2017 provided a structural basis for the formation of VLPs by BCV constructs (91).

The aim of this project was therefore to characterise the structure of BCV X in the presences of single-stranded DNA, and compare its structure with the previously determined BCV crystal structure solved in the Forwood laboratory in 2016.

2.3.1.1 Expression and Purification of Bat X Circovirus

To test the ability of BCV X (Capsid protein devoid of xxxx N-terminal residues) to form VLPs in the presence of single stranded DNA, large quantities of BCVX were recombinantly expressed and purified. A band corresponding to the correct molecular weight of BCV X (23 kDa) was clearly overexpressed in the whole cell extract (See Lane 2 of Figure 2.2C). The following lane, Lane 3 of Figure 2.2C, demonstrates that the protein is mostly soluble due to its presence in the “soluble-extract’ sample, which was prepared by centrifugation of the whole cell lysate to remove insoluble material. The soluble extract was injected onto the HIS-trap affinity column using an AKTA FPLC

(Figure 2.2A), thus allowing adsorption and consequent purification. The flow through from the affinity column, lane 4 Figure 2.2C, indicates the BCV X was bound to the column. The subsequent elution peak, lane 6 Figure 2.2C corresponding to Figure 2.2A, displays a predominant band corresponding to the correct molecular weight of BCV X.

The purified sample was then injected onto an S200 size exclusion chromatography column (Figure 2.2B) to separate the different particle sizes of BCV X and any other non-specifically bound proteins. Separation of peaks in the size exclusion graph, Figure 2.2B displays separation of elution of BCV X target protein due to size. The first peak in Figure 2.2B, eluting between 100 and 150 mL is typical for that of large VLP complexes since VLP particles contain 60 copies of the protein (or some 1,380 kDa) and pass through the column quickly. This is compared to the second peak which formed between 200 and 250 mL, indicating that these particles were smaller in size thus taking longer to elute through the column. The third peak elutes after 300 mL, which from previous purifications in the laboratory indicates this peak is representative of contaminants such as buffers, peptides, and/or small molecular weight compounds.

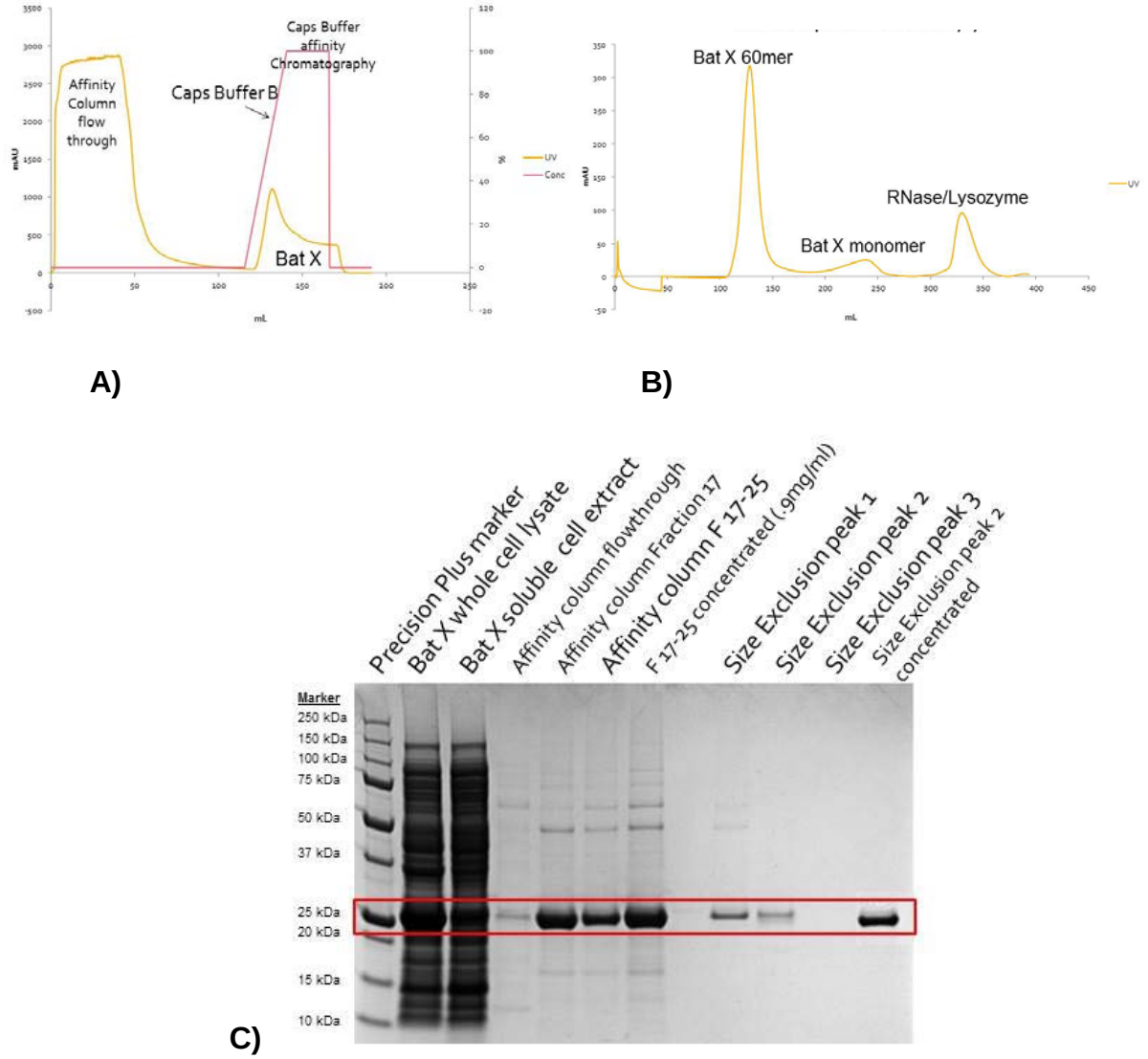


Figure 2.2 Affinity and Size exclusion chromatography of Bat X expressed via IPTG and resuspended in Caps Buffer A, with corresponding SDS-PAGE analysis

A) HIS-trap purification of Bat X UV trace (red), eluent (orange). B) Size exclusion purification of Bat X with UV trace (orange). C) SDS-PAGE visualisation of PMCSG21 Bat X; Lane 1 Marker, Lane 2 Whole Cell, Lane 3 Soluble Extract, Lane 4 Flowthrough, Lane 5 affinity column fraction 17, Lane 6 affinity column fraction 17-25, Lane 7 affinity column fractions 17-25 concentrated to 0.9 mg/ml, Lane 8 empty, Lane 9 S200 peak 1, Lane 10 S200 peak 2, Lane 11 S200 peak 3, Lane 12 S200 peak 2 concentrated. The red box outlines the calculated molecular weight of BCV X, 25 kDa.

2.3.1.1.1 Crystallography trials of Bat X

Purified BCV X from section 2.3.1.1 was used to set up a crystal screen as per method 2.2.2.4.1. A drop containing DNA was mounted next to a control drop of BCV without DNA in this hanging drop protocol. Crystals were obtained by using crystal screen Proplex 1 (Molecular Dimensions), with no optimisation prior to diffraction undertaken. Large crystals (see Figure 2.3A) were flash cooled and diffracted at the Australian Synchrotron.

The diffraction of BCV X + DNA, obtained in 0.2 M sodium acetate, 0.1 M sodium citrate pH 5, with 20% glycerol as a cryoprotectant, diffracted to a resolution of 4.3 Å (Figure 2.4). The scaling and refinement statistics for BCV X + DNA are outline in Table 2:7.

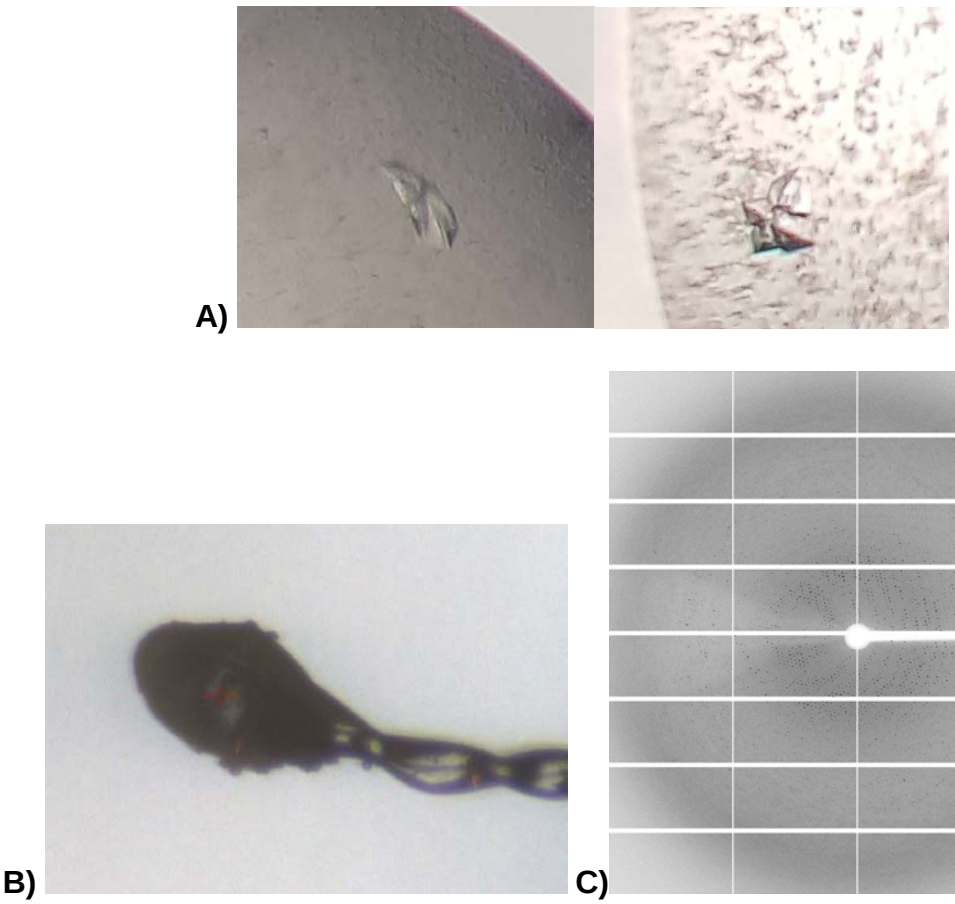


Figure 2.3 Crystals of Bat X + DNA
A) Protein crystals in hanging drop, B) mounted protein crystal, C) X-ray diffraction image

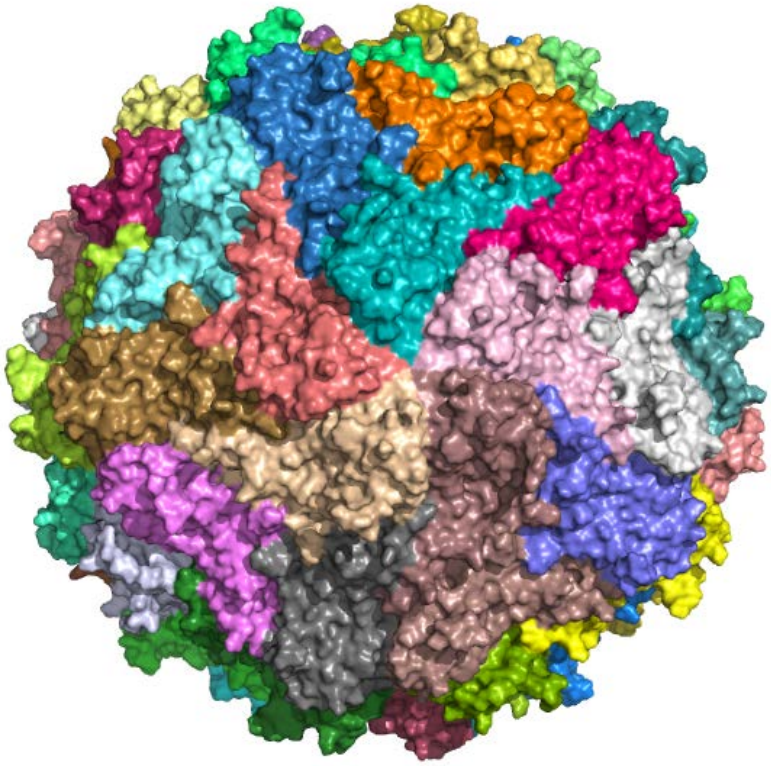


Figure 2.4 Diagram of T=1 icosahedral assembly from BCV X + DNA capsid structure as determined by X-ray

Three-dimensional surface map reconstruction of BCV X, capsid showing pentameric subunit assemblies, with each colour representing one of the 60 proteins that make up the capsid.

Table 2.7 BCV X + DNA data collection and refinement statistics

Data Type	Statistics
Wavelength	
Resolution Range	20 - 4.31 (4.462 - 4.31)
Space Group	P 42 21 2
Unit cell	285.16 285.16 239.64 90 90 90
Total reflection	132482 (13086)
Unique reflection	66241 (6543)
Multiplicity	2.0 (2.0)
Completeness (%)	98.82 (100.00)
Mean I/sigma(I)	8.80 (4.29)
Wilson B-factor	90.23
R-merge	0.07543 (0.1796)
R-meas	0.1067 (0.254)
R-pim	0.07543 (0.1796)
CC 1/2	0.992 (0.837)
CC*	0.998 (0.955)
Reflections used in refinement	66241 (6543)
Reflections used for R-free	3311 (314)
R-work	0.1905 (0.2028)
R-free	0.2753 (0.2820)
CC (work)	0.943 (0.866)
CC (free)	0.875 (0.756)
Number of non-hydrogen atoms	46950
Macromolecules	46950
Protein residues	5670
RMS (bonds)	0.010
RMS (angles)	1.34
Ramachandran favoured (%)	93.30
Ramachandran allowed (%)	6.42
Ramachandran outliers (%)	0.29
Rotamer outliers (%)	2.09
Clashscore	1.45
Average B-factor	110.61
Macromolecules	110.61

Stastics for the highest resolution sheel are show in parentheses ( )

2.3.2 Chicken Anaemia Virus (CAV)

Recombinant expression of the Cap targets of CAV VP1, CAV VP1 X and CAV VP1 XL were performed using optimised expression vector (pMCSG21), *E. coli* BL21 (DE3-Rosetta 2), and expression media IPTG. Construct CAV VP1 represent the full length protein for the capsid-like protein (VP1) of CAV. The sequence of CAV VP1 X differs from that of CAV VP1, in that the ARM Domain has been removed, whilst the CAV VP1 XL sequences consisted of the minimal ARM Domain from Major Mitchell Cockatoo (MMC) BFDV (See Figure 2.4). Prior to this study successful bacterial expression of CAV and CAV constructs was not achieved in this or any other laboratory.

The aim of this project was therefore to test recombinant expression of the three constructs.

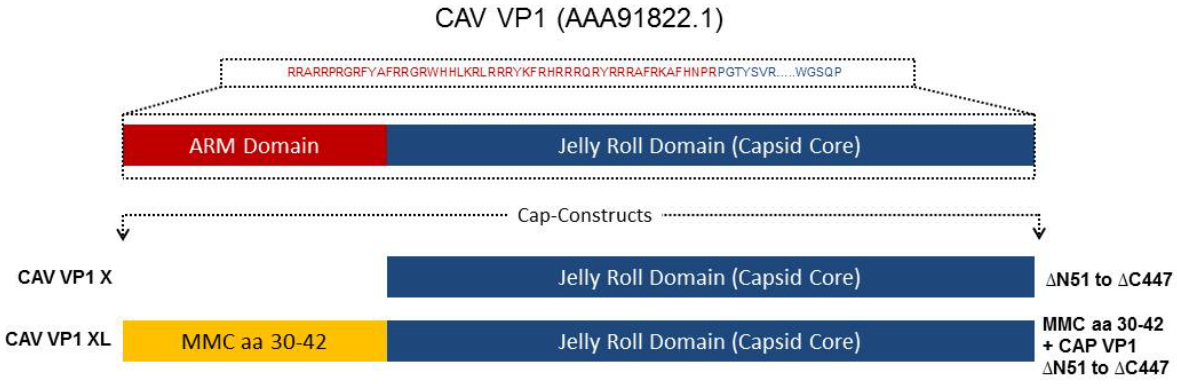


Figure 2.5 Schematic comparison of CAV VP1, CAV VP1 X and CAV VP1 XL
Schemtic diagram of the Cap protein target in the CAV part of this study for recombinant expression. Each horizontal bar represents differences from the full length construct CAP VP1.The original ARM domain of CAV is depicted in a red domain and an alternate ARM domain in CAV VP1 XL is the yellow region. Species Genbank accession number of the genome is shown at the top of the diagram. While the numbers on the right side denotes the length of each Cap protein.

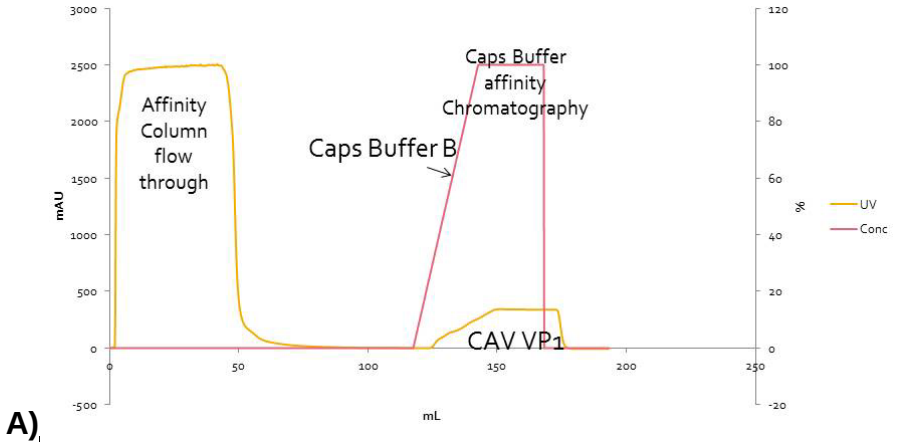
2.3.2.1 Expression and Purification Trials of CAV VP1, CAV VP1 X and CAV VP1 XL

2.3.2.1.1 Expression and Purification trials of CAV VP1

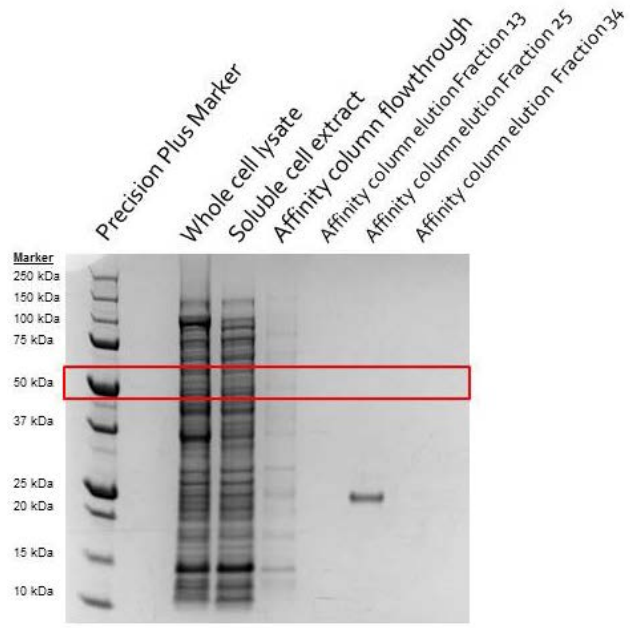
To test recombinant expression, 4 L of CAV construct CAV VP1 was expressed by IPTG and purified as per methods 2.2.2.2.3 and 2.2.2.3.2, respectively. A band corresponding to the correct molecular weight of CAV VP1 (52 kDa) was not overexpressed in the whole cell extract (See Lane 2 Figure 2.5B). The following lane, lane 4 Figure 2.5B also demonstrates the lack of over-expressed soluble protein. The soluble extract was injected onto the HIS-trap affinity column using an AKTA FPLC (Figure 2.5A). The flow through from the affinity column contained few proteins at the expected molecular weight (lane 5 Figure 2.5B). The subsequent elution peak lanes 6, 7 & 8 Figure 2.5B display the lack of a band at the correct molecular weight for CAV VP1.

Due to lack of overexpression in whole cell and production of only small quantities of purified protein achieved through affinity chromatography, further purification by size exclusion chromatography was not carried out.

Bands of overexpression at approximately 37 kDa and 15 kDa, in whole cell and soluble cell extracts respectively, are believed to be *E. coli* proteins that are commonly over-expressed. The band produced from the affinity elution at 25 kDa is believed to be a contaminate protein.



A)



B)

Figure 2.6 Affinity chromatography of CAV VP1 expressed via IPTG, resuspended in Caps Buffer A, and corresponding SDS-PAGE analysis

A) HIS-trap purification of CAV VP1 UV trace (red), eluent (orange). B) SDS-PAGE visualisation of CAV VP1; Lane 1 Marker, Lane 2 empty, Lane 3 Whole Cell, Lane 4 Soluble Extract, Lane 5 Flowthrough, Lane 6 HIS-trap fraction 13, Lane 7 HIS-trap fraction 25, Lane 8 HIS-trap fraction 34. The red box outlines the calculated molecular weight of CAV VP1, 52 kDa.

2.3.2.1.2 Expression and Purification trials of CAV X

To test recombinant expression, 4 L of CAV constructs CAV VP1 X was expressed by IPTG and purified as per methods 2.2.2.2.3 and 2.2.2.3.2, respectively. A band corresponding to the correct molecular weight of CAV VP1 X (45 kDa) was clearly overexpressed in the whole cell extract (see lane 3 Figure 2.6B). However, the lack of a band at the correct molecular weight in the soluble extract, lane 4 Figure 2.6B, highlights that the protein is not soluble. The soluble extract was injected onto the HIS-trap affinity column using AKTA FPLC (Figure 2.6A), and the subsequent elution peaks lanes 6 & 7 Figure 2.6B display a lack of a band at the correct molecular weight for CAV VP1 X.

Due to overexpression in the whole cell but production of only small quantities of purified protein achieved through affinity chromatography, further purification by size exclusion chromatography was not carried out. The construct needs to first be solubilised before carrying on with purification trials.

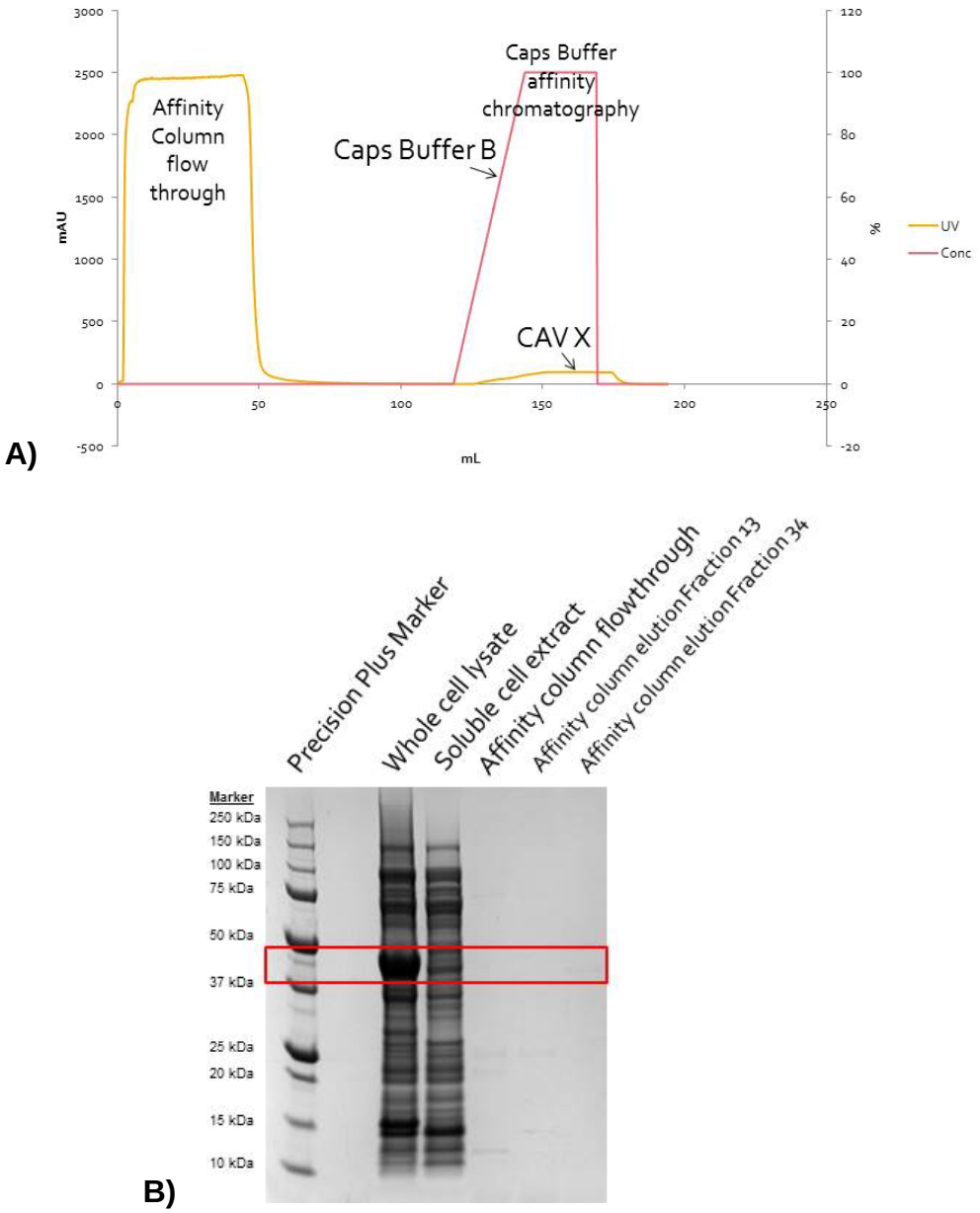


Figure 2.7 Affinity chromatography of CAV VP1 X expression via IPTG, resuspended in Caps Buffer A, and corresponding SDS-PAGE analysis

A) HIS-trap purification of CAV VP1 X UV trace (red), eluent (orange). B) SDS-PAGE visualisation of CAV VP1 X; Lane 1 Marker, Lane 2 empty, Lane 3 Whole Cell, Lane 4 Soluble Extract, Lane 5 Flowthrough, Lane 6 HIS-trap fraction 13, Lane 7 HIS-trap fraction 34. The red box outlines the calculated molecular weight of CAV VP1 X, 45 kDa.

2.3.2.1.3 Expression and Purification trials of CAV VP1 XL

To test recombinant expression of the CAV VP1 XL construct, 4 L was expressed by IPTG and purified as per methods 2.2.2.2.3 and 2.2.2.3.2, respectively. A band corresponding to the correct molecular weight of CAV VP1 XL (47 kDa) was clearly overexpressed in the whole cell extract (see lane 3 Figure 2.7B), however was not soluble, as demonstrated by the absence of a band at the correct molecular weight in the soluble extract (see lane 5 Figure 2.7B). The soluble extract was injected onto the HIS-trap affinity column using an AKTA FPLC (Figure 2.7A), thus allowing adsorption and subsequence purification. The flow through from the affinity column, lane 4 Figure 2.7B shows the majority of proteins seen in the soluble extract had bound to the column. The subsequence elution peak collections lanes 6 & 7 Figure 2.7B display a lack of a band at the correct molecular weight for CAV VP1 XL.

Due to overexpression in whole cell but production of only small quantities of purified protein achieved through affinity chromatography, further purification by size exclusion chromatography was not carried out. The construct needs to first be solubilised before further purification trials.

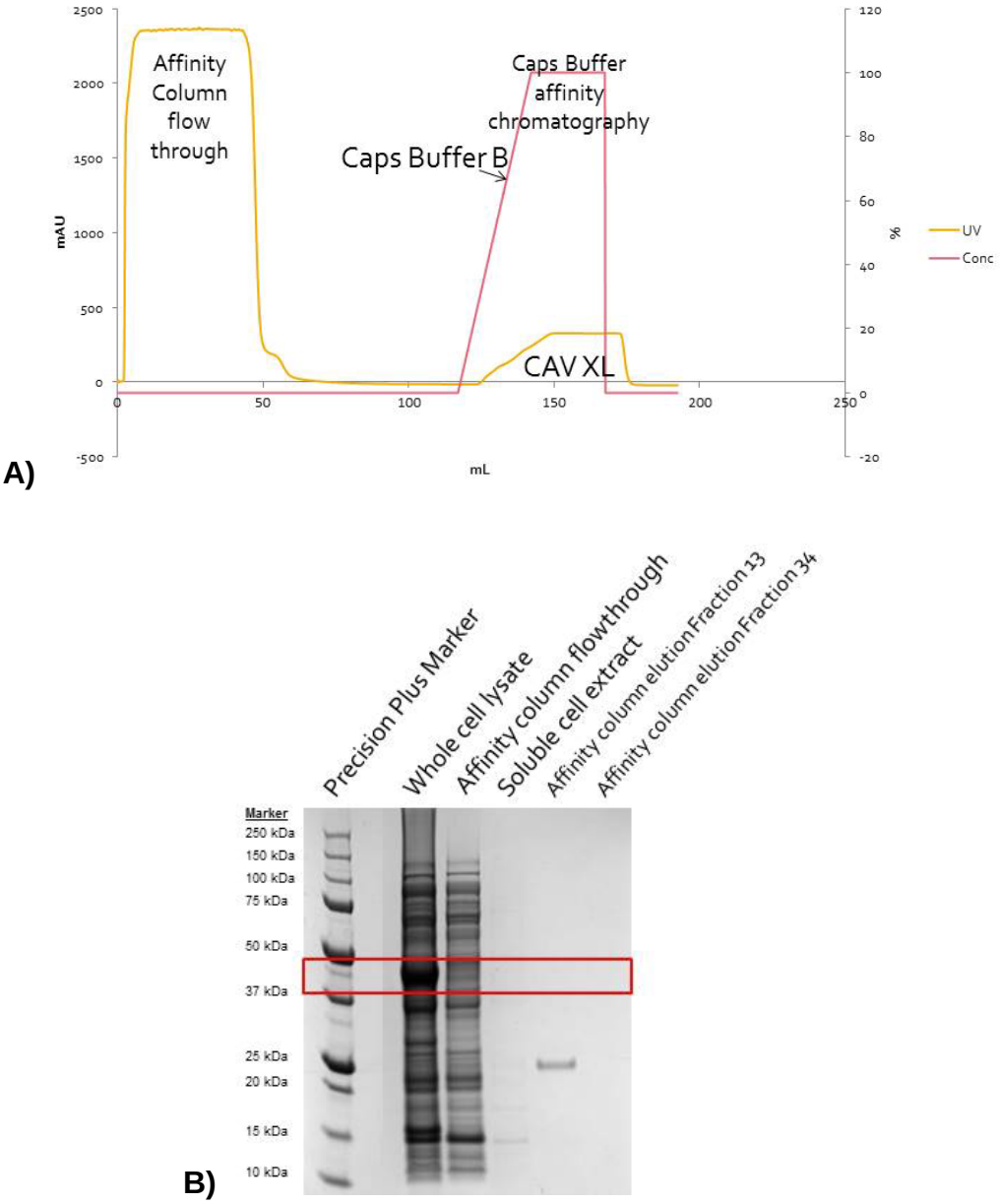


Figure 2.8 Affinity Chromatography of CAV VP1 XL expressed via IPTG, resuspended in Caps Buffer A, and corresponding SDS-PAGE analysis

A) HIS-trap purification of CAV VP1 XL UV trace (red), eluent (orange). B) SDS-PAGE visualisation of CAV VP1 XL; Lane 1 Marker, Lane 2 empty, Lane 3 Whole Cell, Lane 4 Flowthrough, Lane 5 Soluble extract, Lane 6 HIS-trap fraction 13, Lane 7 HIS-trap fraction 34. The red box outlines the calculated molecular weight ofCAV VP1 XL, 47 kDa.

2.3.2.2 Initial solubility trials of CAV VP1 X and CAV VP1 XL

As a result of the over-expression in the whole cell extracts of CAV VP1 X and CAV VP1 XL, and lack thereof for CAV VP1, it was decided to proceed with trails of solubilisation of constructs X and XL.

Firstly, solubilisation trials were conducted using buffers and detergents that had previously been used to solubilise protein in other projects in the laboratory.

2.3.2.2.1 Solubility Trial of CAV VP1 X

To trial solubility potential, small quantities of whole cell extracts of CAV VP1 X were resuspended in buffers of various pH and different detergents. The protein was incubated for 30 mins to allow break down of the whole cell extract, and a sample of the whole cell extract was then collected for SDS analysis (as per 2.2.2.5). Following centrifugation, the supernatant (soluble extract) was collected for SDS analysis (as per 2.2.2.5). The results of the trial are found in Figure 2.8. A band corresponding to the correct molecular weight of CAV VP1 X (45 kDa) is observed in the majority of lanes in Figure 2.9. The appearance of two bands running close together at approximately 45 kDa on the SDS-PAGE analysis highlight the necessity to run further experiments to make certain which band is the target protein.

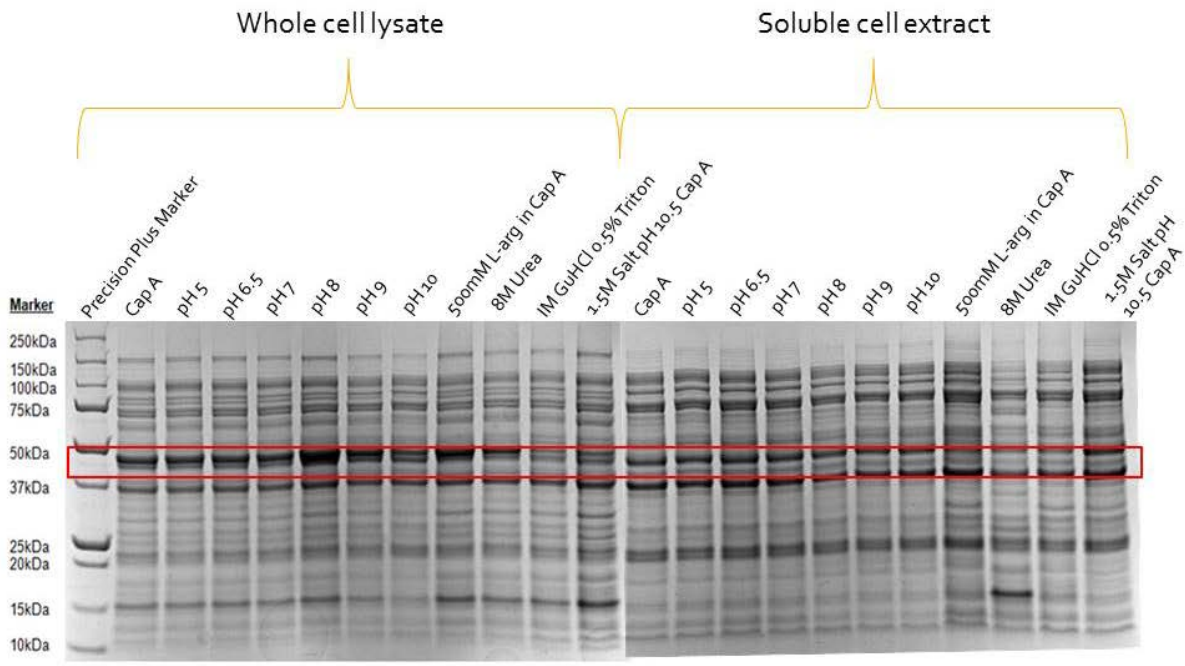


Figure 2.9 SDS-PAGE analysis of CAV VP1 X solubility trials

SDS-PAGE visualisation of CAV VP1 X; Lane 1 Marker, Lane 2 Whole Cell (WC) resuspended in Caps Buffer A, Lanes 3-8 WC resuspended in exchange buffer pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 9 WC resuspended in Caps Buffer A + L-Arginine, Lane 10 WC resuspended in 8M Urea, Lane 11 WC resuspended in GuHCl, Lane 12 WC resuspended in Caps Buffer A High Salt, Lane 13 Soluble Extract (SE) of Caps Buffer A treatment, Lanes 14-19 SE of exchange buffer treatment pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 20 SE of Caps Buffer A + L-Arginine treatment, Lane 21 SE of 8M Urea treatment, Lane 22 SE of GuHCl treatment, Lane 23 SE of Caps Buffer A High Salt treatment. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of CAV VP1 X, 45 kDa.

Solubility trials of CAV VP1 XL

Using a similar approach to CAV VP1 X, the solubility potential of CAV VP1 XL was tested against various pH and detergents. The protein was allowed to incubate for 30 mins and a sample of the whole cell extract was then collected for SDS analysis (as per 2.2.2.5). Following centrifugation, the supernatant (soluble extract) was collected for SDS analysis (as per 2.2.2.5) and the results are presented in Figure 2.9. A band

corresponding to the correct molecular weight of CAV VP1 XL (47 kDa) can be seen in the majority of the lanes in Figure 2.9. The appearance of 2 bands running close together at approximately 47 kDa on the SDS-PAGE analysis highlighted the necessity to run further experiments to ascertain which band is the target protein

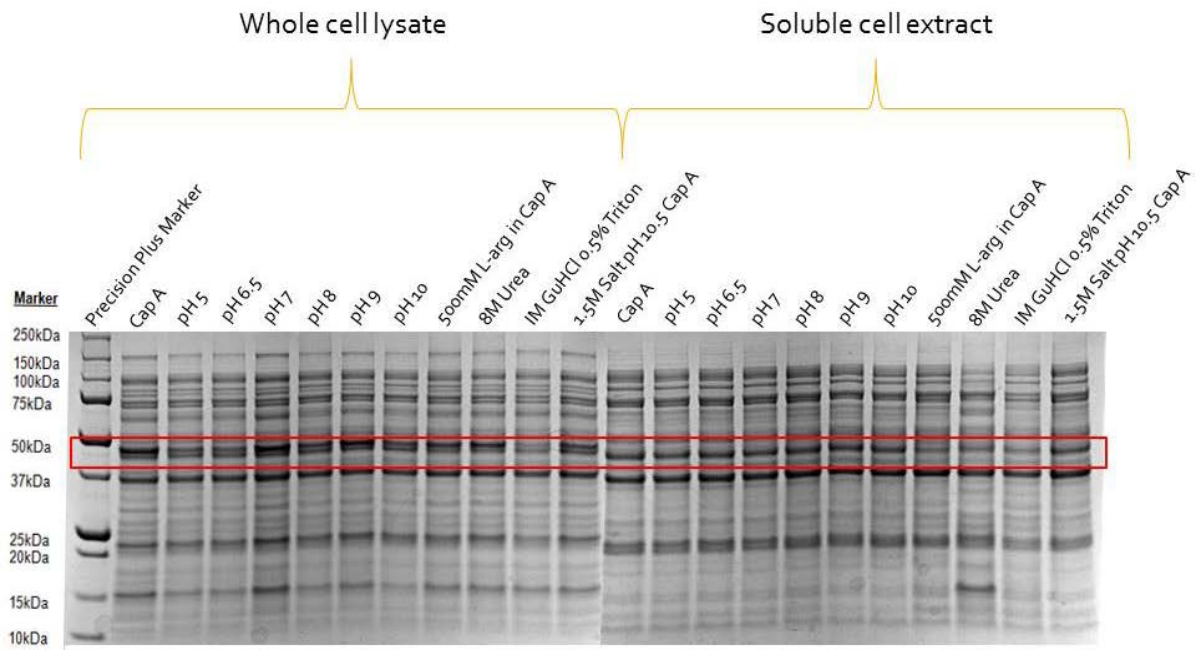


Figure 2.10 SDS-PAGE analysis of CAV VP1 XL solubility trials

SDS-PAGE visualisation of CAV VP1 XL; Lane 1 Marker, Lane 2 Whole Cell (WC) resuspended in Caps Buffer A, Lanes 3-8 WC resuspended in exchange buffer pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 9 WC resuspended in Caps Buffer A + L-Arginine, Lane 10 WC resuspended in 8M Urea, Lane 11 WC resuspended in GuHCl, Lane 12 WC resuspended in Caps Buffer A High Salt, Lane 13 Soluble Extract (SE) of Caps Buffer A treatment, Lanes 14-19 SE of exchange buffer treatment pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 20 SE of Caps Buffer A + L-Arginine treatment, Lane 21 SE of 8M Urea treatment, Lane 22 SE of GuHCl treatment, Lane 23 SE of Caps Buffer A High Salt treatment. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of CAV VP1 XL, 47 kDa.

2.3.2.3 Trials to determine if CAV VP1 X and CAV VP1 XL are within inclusion bodies

Whilst the above approach was aimed at characterising the solubility of CAV protein, we also made attempt to resolubilise the CAV proteins from inclusion bodies. The formation of inclusion bodies, especially in *E. coli,* is often the result of the protein being insoluble or fast expression that leads to improper folding and formation (83). Solubilisation of inclusion bodies can be achieved through resuspension of the pellet with the addition of detergent and/or chaotropic agents such as urea or guanidine hydrochloride.

2.3.2.3.1 Solubility of CAV VP1 X from inclusion bodies trials

In order to determine if CAV VP1 X can be resolubilised from inclusion bodies, small quantities of the whole cell extract were resuspended in buffers containing different solubilising formulations (see Figure 2.10). Following incubation for 30 mins, the sample was centrifuged and the supernatant (soluble extract) was collected for SDS analysis (as per 2.2.2.5). The results of the trial are found in Figure 2.10. A band corresponding to the correct molecular weight of CAV VP1 X (45 kDa) was not seen in Figure 2.10. The failure of expected solubilisation by resuspension in 8 M Urea displays inability to solubilise CAV VP1 X from inclusion bodies.

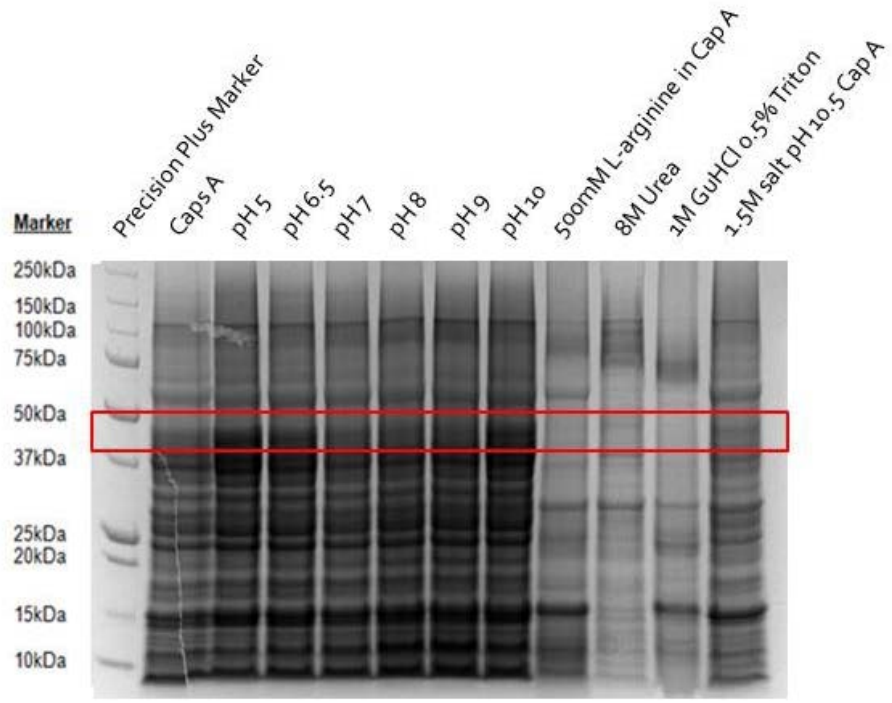


Figure 2.11 SDS-PAGE analysis of solubilisation of CAV VP1 X from inclusion bodies via resuspension of precipitate in different buffers or different detergent origin and pH

SDS-PAGE visualisation of CAV VP1 X, Lane 1 Marker, Lane 2 Precipitate resuspended in Caps Buffer A, Lanes 3-8 precipitate resuspended in exchange buffer pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 9 precipitate resuspended in Caps Buffer A + L-Arginine, Lane 10 precipitate resuspended in 8M Urea, Lane 11 precipitate resuspended in GuHCl, Lane 12 precipitate resuspended in Caps Buffer A High Salt. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of CAV VP1 X, 45 kDa.

2.3.2.3.2 Solubility of CAV VP1 XL from inclusion bodies trials

A trial to determine if CAV VP1 XL is within inclusion bodies was carried out by, using methods outlined in 2.2.2.3.1. Bacterial pellets containing CAV VP1 XL were mixed with various buffer as shown in Figure 2.11. The mixture was then vortexed and allowed to incubate for 30 mins and following centrifugation, the supernatant (soluble extract) was collected for SDS analysis (as per 2.2.2.5). The results of the trial are presented in Figure 2.11. A band corresponding to the correct molecular weight of CAV VP1 XL (47 kDa) was not observed, however solubilisation did occur in pH10 (lane 8 Figure 2.11) though the band does not correlate to CAV VP1 XL.

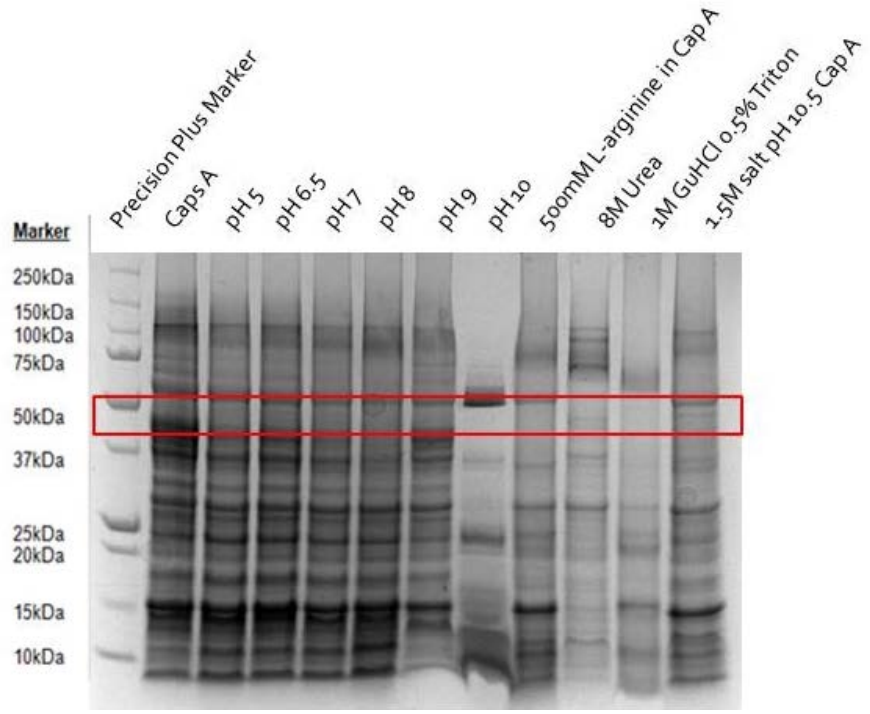
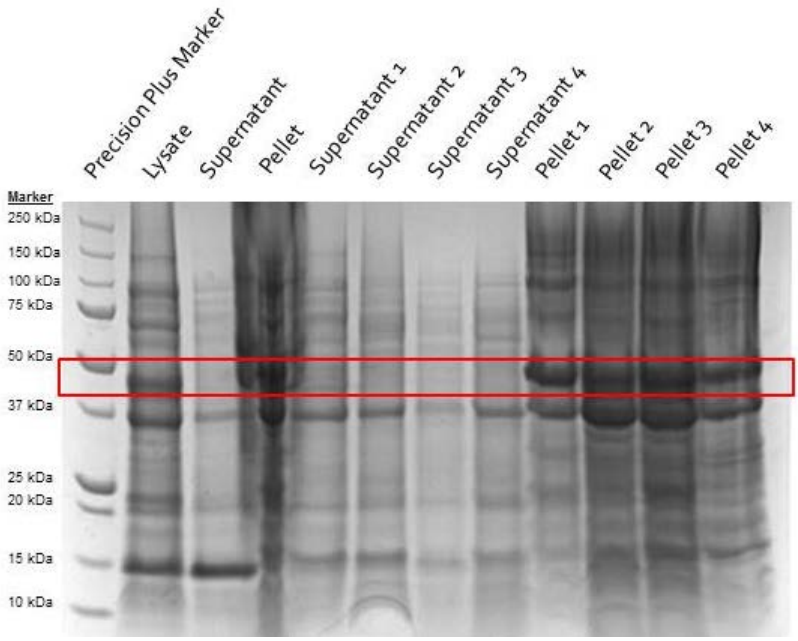


Figure 2.12 SDS-PAGE analysis of solubilisation of CAV VP1 XL from inclusion bodies via resuspension of precipitate in different buffers or different detergent origin and pH

SDS-PAGE visualisation of CAV VP1 XL, Lane 1 Marker, Lane 2 Precipitate resuspended in Caps Buffer A, Lanes 3-8 precipitate resuspended in exchange buffer pH 5, 6.5, 7, 8, 9, & 10, respectively, Lane 9 precipitate resuspended in Caps Buffer A + L-Arginine, Lane 10 precipitate resuspended in 8M Urea, Lane 11 precipitate resuspended in GuHCl, Lane 12 precipitate resuspended in Caps Buffer A High Salt. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of BCV X, 25 kDa. The red box outlines the calculated molecular weight of CAV VP1 XL, 47 kDa.

Further expression of CAV VP1 XL was undertaken and trailed for solubilisation as per 2.2.2.2.3. The solubilisation potential of CAV VP1 XL was carried out as above using different resuspension buffers (see Figure 2.12) and a band corresponding to the

correct molecular weight of CAV VP1 XL (47 kDa) was observed in all 4 pellets (lane 9-12 Figure 2.12), and not in the supernatant (soluble extract)(lane 5-8 Figure 2.12). It is possible that SDS presence in the loading dye is playing a role in solubilisation of the protein.

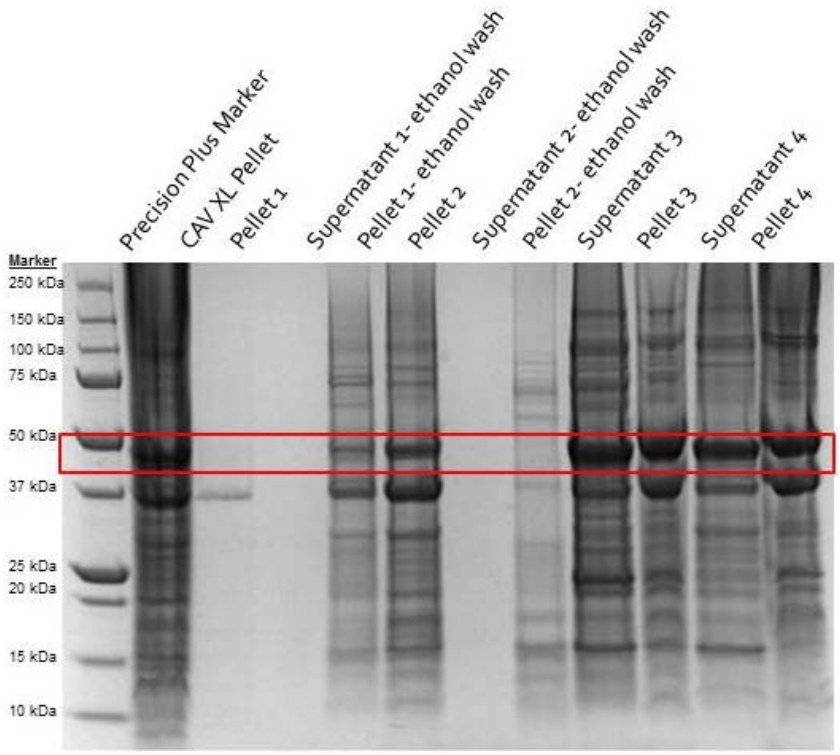


Experimental conditions correlating to the above numbers:
1. HIS (Nickel) Buffer A, 8 M Urea
2. 2 × GST Buffer A, 2M Urea pH 12.5
3. HIS (Nickel) Buffer A, 200mM Arginine
4. 250mM Arginine , 1 PBS tablet pH 11

Figure 2.13 SDS-PAGE analysis of solubilisation of CAV VP1 XL from inclusion bodies using mild solubilisation processes HIS (Nickel) Buffer A + 8 M Urea, 2× GST Buffer A + 2 M Urea pH 12.5, HIS (Nickel) Buffer A + 200 mM Arginine & 250 mM Arginine + 1PBS tablet pH 11)

SDS-PAGE visualisation of CAV XL, Lane 1 Marker, Lane 2 Whole Cell, Lane 3 Soluble Extract, Lane 4 Pellet, Lane 5 Soluble Extract (SE) of HIS (Nickel) Buffer A + 8M Urea treatment, Lane 6 SE of 2 × GST Buffer A + 2 M Urea pH 12.5 treatment, Lane 7 SE of HIS (Nickel) Buffer A, 200 mM Arginine treatment, Lane 8 SE 250 mM Arginine + 1PBS tablet pH 11 treatment, Lane 9 Insoluble extract (IE) of HIS (Nickel) Buffer A + 8M Urea treatment, Lane 10 IE of 2 × GST Buffer A + 2 M Urea pH 12.5 treatment, Lane 11 IE of HIS (Nickel) Buffer A, 200 mM Arginine treatment, Lane 12 IE 250 mM Arginine + 1PBS tablet pH 11 treatment. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of CAV VP1 XL, 47 kDa.

Following failed attempts to solubilise CAV VP1 XL from inclusion bodies an alternative approach was employed based on methods from I. Palmer and P. T. Wingfield (84) describing the preparation and extraction of proteins from inclusion bodies. Results of the trail can be seen in Figure 2.13. A band corresponding to the correct molecular weight of CAV VP1 XL (47 kDa) was observed in the supernatant (soluble extract) of both treatments number 3 and 4 (lane 9 and 11 Figure 2.13). The presence of corresponding bands of correct molecular weight of CAV VP1 XL in pellets of both condition 3 and 4 demonstrate that not all the protein has been solubilised from inclusion bodies. Importantly, solubilisation of CAV VP1 XL from inclusion bodies was demonstrated for the first time. Further studies to optimise the buffers and resolubilisation are required



Experimental condition correlating to the numbers above:
1. 6 M GuHCl
2. 1 M GuHCl, 0.5% triton, 50 mM Tris, 250 mM NaCl
3. 8 M Urea, GST Buffer A, 1% SDS, pH8
4. GST Buffer A, 1% SDS, pH8

Figure 2.14 SDS-PAGE analysis of solubilisation of CAV VP1 XL from inclusion bodies using mild solubilisation processes (6 M GuHCl, 1 M GuHCl + 0.5% Triton + 50 mM Tris + 250 mM NaCl, 8 M Urea + GST Buffer A + 1% SDS pH 8 & GST Buffer A + 1% SDS pH 8)

SDS-PAGE visualisation of CAV XL, Lan 1 Marker, Lane 2 Whole Cell, Lane 3 Insoluble extract (IE) from 6M GuHCl treatment, Lane 4 Soluble extract (SE) from ethanol wash of 6M GuHCl treated whole cell, Lane 5 IE from ethanol wash of 6M GuHCl treated whole cell, Lane 6 IE from 1 M GuHCl + 0.5%Triton + 50 mM Tris + 250 mM NaCl treated Whole Cell, Lane 7 SE from ethanol wash of 1 M GuHCl + 0.5%Triton + 50 mM Tris + 250 mM NaCl treated Whole Cell, Lane 8 IE from ethanol wash of 1 M GuHCl + 0.5%Triton + 50 mM Tris + 250 mM NaCl treated Whole Cell, Lane 9 SE from 8 M Urea + GST Buffer A + 1% SDS pH8 treatment, Lane 10 IE from 8 M Urea + GST Buffer A + 1% SDS pH8 treatment, Lane 11 SE from GST Buffer A + 1% SDS pH 8 treatment, Lane 12 IE from GST Buffer A + 1% SDS pH 8 treatment. See Table 2.1 for composition of buffers. The red box outlines the calculated molecular weight of CAV VP1 XL, 47 kDa.

2.3.3 Porcine Circovirus 3 (PCV-3)

Recombinant expression of the Cap target of PCV-3 was performed using optimised expression vector (pMCSG21), *E. coli* BL21 (DE3-Rosetta 2), and two different expression medias, IPTG or auto-induction, as per methods 2.2.2.2.3 and 2.2.2.2.2, respectively. Prior to this study it was unknown if PCV-3 can be expressed recombinantly in *E. coli*.

The aim of this project was therefore to test recombinant expression of PCV-3.

2.3.3.1 Auto-induction expression and purification trials of PCV-3

To test recombinant expression, 8 L of PCV-3 was expressed by IPTG, as per method 2.2.2.2.3, with 4L being resuspended in Caps Buffer A and the remaining 4L in HIS (Nickel) Buffer A, before both being purified as per method 2.2.2.3.2. A band corresponding to the correct molecular weight of PCV-3 (25 kDa) was not overexpressed in the whole cells (see lanes 2 & 8 Figure 2.14C). The lack of significant bands at the correct molecular weight in the soluble extract, lanes 3 & 9 Figure 2.14C, demonstrates that the protein is not highly expressed in *E. coli*. The soluble extract was injected onto the HIS-trap affinity column using AKTA FPLC (Figure 2.14A & B), thus allowing adsorption and subsequence purification. The flow through from the affinity column, lanes 4 & 10 Figure 2.14C, shows little difference from the soluble extract, demonstrating little from the soluble extract has bound to the column. The subsequence affinity elution's lanes, 5 & 11 Figure 2.14B, display a lack of a band at the correct molecular weight for PCV-3, most likely owing to the poor expression.

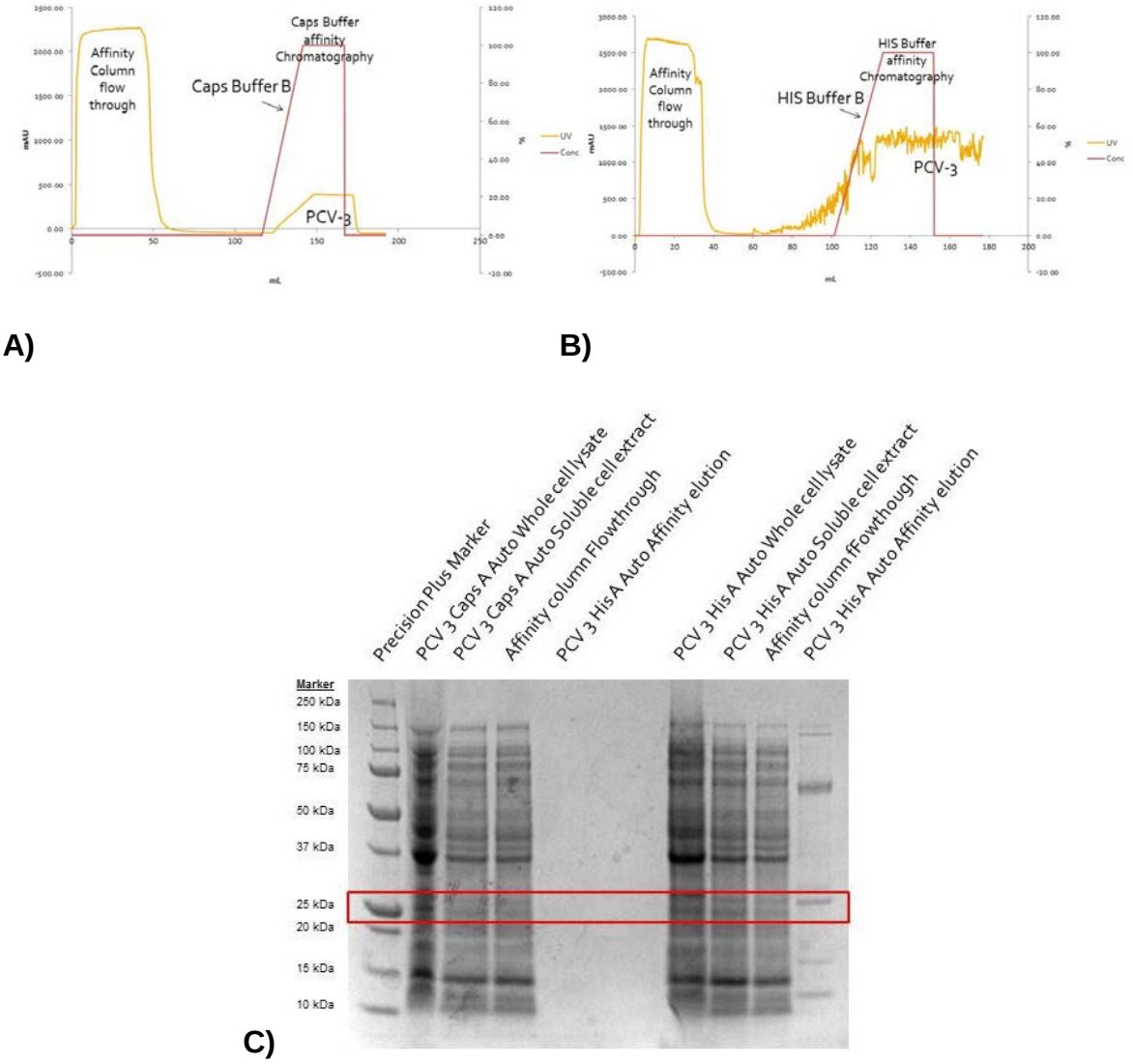
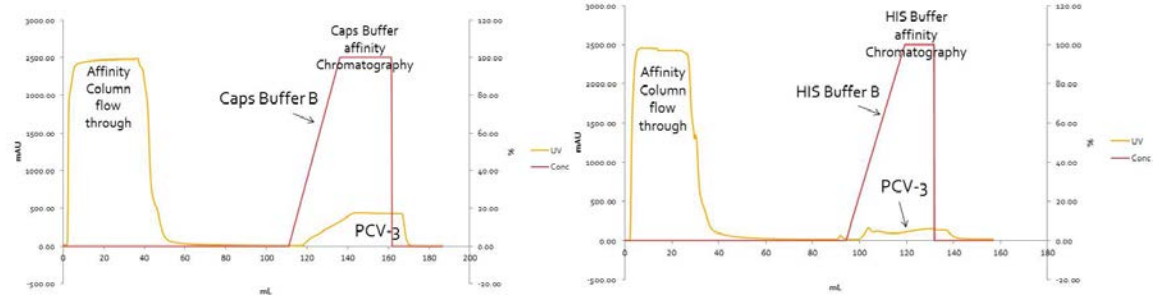


Figure 2.15 Affinity chromatography of PCV-3expressed via Auto-induction, resuspended in Caps Buffer A and HIS (Nickel) Buffer A respectively, with corresponding SDS-PAGE analysis

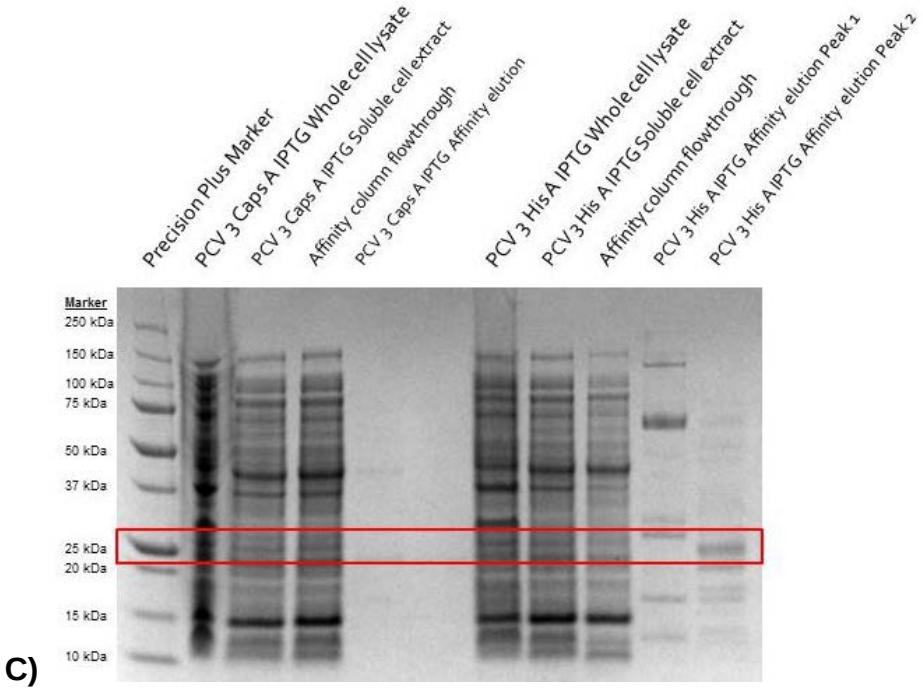
A) HIS-trap purification of PCV-3 expressed via Auto-induction in Caps Buffer A UV trace (red), eluent (orange). B) HIS-trap purification of PCV-3 expressed via Auto-induction in HIS (Nickel) Buffer A UV trace (red), eluent (orange). C) SDS-PAGE visualisation of PCV-3; Lane 1 Marker, Lane 2 Caps Buffer A Whole Cell, Lane 3 Caps Buffer A Soluble Extract, Lane 4 Caps Buffer A flowthrough, Lane 5 Caps Buffer A HIS-trap elution, Lane 6&7 empty, Lane 8 HIS (Nickel) Buffer A Whole Cell, Lane 9 HIS (Nickel) Buffer A Soluble Extract, Lane 10 HIS (Nickel) Buffer A Flowthrough, Lane 11 HIS (Nickel) Buffer A HIS-trap elution. The red box outlines the calculated molecular weight of PCV-3, 25 kDa.

2.3.3.2 IPTG expression and purification trials of PCV-3

To test recombinant expression, 8 L of PCV-3 was expressed by auto-induction, as per method 2.2.2.2.2, with 4L being resuspended in Caps Buffer A and the remaining 4L in HIS (Nickel) Buffer A, before both being purified as per method 2.2.2.3.2. A band corresponding to the correct molecular weight of PCV-3 (25 kDa) was not overexpressed in the whole cells (see lanes 2 & 8 Figure 2.15C). The lack of significant bands at the correct molecular weight in the soluble extract, lanes 3 & 9 Figure 2.15C, demonstrates that the protein was not over-expressed. The soluble extract was injected onto the HIS-trap affinity column using an AKTA FPLC (Figure 2.15A & B) with the flow through from the affinity column, lanes 4 & 10 Figure 2.15C, showing little difference from the soluble extract. The subsequent affinity elution's ,lanes 5 & 11 Figure 2.15C, display a lack of a band at the correct molecular weight for PCV-3.



A) B)



C)

Figure 2.16Affinity chromatography of PCV-3 expressed via IPTG, resuspended in Caps Buffer A and HIS (Nickel) Buffer A, with corresponding SDS-PAGE analysis

A) HIS-trap purification of PCV-3 expressed via IPTG in Caps Buffer A UV trace (red), eluent (orange). B) HIS-trap purification of PCV-3 expressed via IPTG in HIS (Nickel) Buffer A UV trace (red), eluent (orange). C) SDS-PAGE visualisation of PCV-3; Lane 1 Marker, Lane 2 Caps Buffer A Whole Cell, Lane 3 Caps Buffer A Soluble Extract, Lane 4 Caps Buffer A flowthrough, Lane 5 Caps Buffer A HIS-trap elution, Lane 6&7 empty, Lane 8 HIS (Nickel) Buffer A Whole Cell, Lane 9 HIS (Nickel) Buffer A Soluble Extract, Lane 10 HIS (Nickel) Buffer A Flowthrough, Lane 11 HIS (Nickel) Buffer A HIS-trap Peak 1, Lane 12 HIS (Nickel) Buffer A HIS-trap Peak 2. The red box outlines the calculated molecular weight of PCV-3, 25 kDa.

2.3.4 Starfish Circovirus (AfaCV)

Recombinant expression of the Cap target of AfaCV were performed using optimised expression vector (pMCSG21), *E. coli* BL21 (DE3-Rosetta 2), and expression media auto-induction, as per method 2.2.2.2.2. Prior to this study it was no known if AfaCV could be expressed recombinantly and due to the highly divergent sequence of AfaCV from typical circoviruses sequences, little structural information was also known.

The aim of this project was therefore to design and test recombinant expression of the construct encoding AfaCV.

2.3.4.1 Expression and purification trials of Starfish circovirus

To test recombinant expression, 4 L of AfaCV was expressed by auto-induction and purified, as per methods 2.2.2.2.2 and 2.2.2.3.2 respectively. A band corresponding to the correct molecular weight of AfaCV (39 kDa) was not overexpressed in the whole cell extract (See Lane 3 Figure 2.16B). Similar to PCV-3, the lack of a band at the correct molecular weight (lane 4 Figure 2.16B), demonstrates the protein was not over-expressed. The soluble extract was injected onto the HIS-trap affinity column using an AKTA FPLC (Figure 2.16A) and the subsequent elution peaks, see lanes 6, 7 & 8 Figure 2.16B, display a lack of a band at the correct molecular weight for AfaCV.

Due to lack of overexpression and production of only small quantities of purified protein achieved through affinity chromatography, further purification by size exclusion chromatography was not carried out.

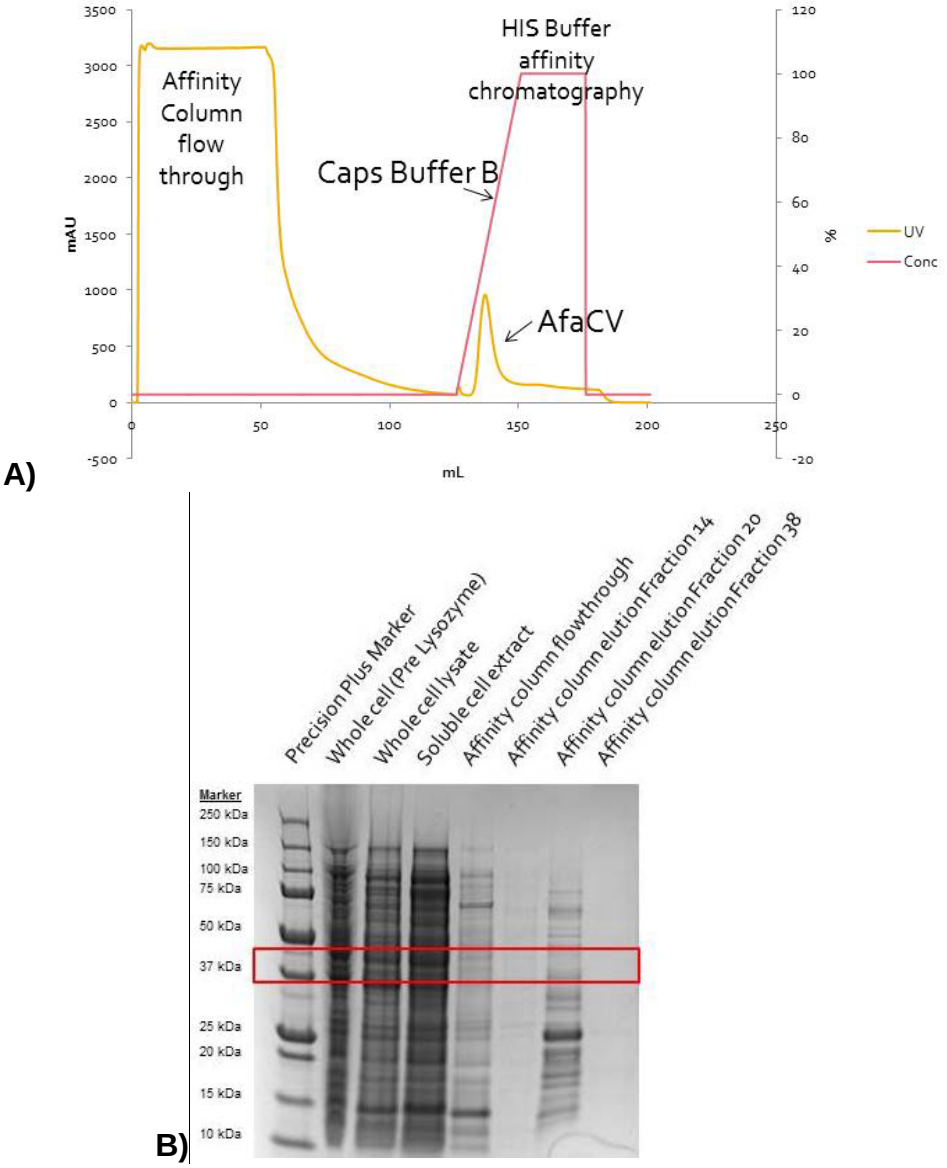


Figure 2.17 Affinity chromatography of Starfish NLS expressed via Auto-induction and resuspended in HIS (Nickel) Buffer A, with corresponding SDS-PAGE analysis

A) HIS-trap purification of Starfish NLS UV trace (red), eluent (orange). B) SDS-PAGE visualisation of Starfish NLS; Lane 1 Marker, Lane 2 Whole Cell prior to addition of lysozyme, Lane 3 Whole Cell, Lane 4 Soluble Extract, Lane 5 Flowthrough, Lane 6 HIS-trap fraction 14, Lane 7 HIS-trap fraction 20, Lane 8 HIS-trap fraction 38. The red box outlines the calculated molecular weight of AfaCV, 39kDa.

2.4 Discussion

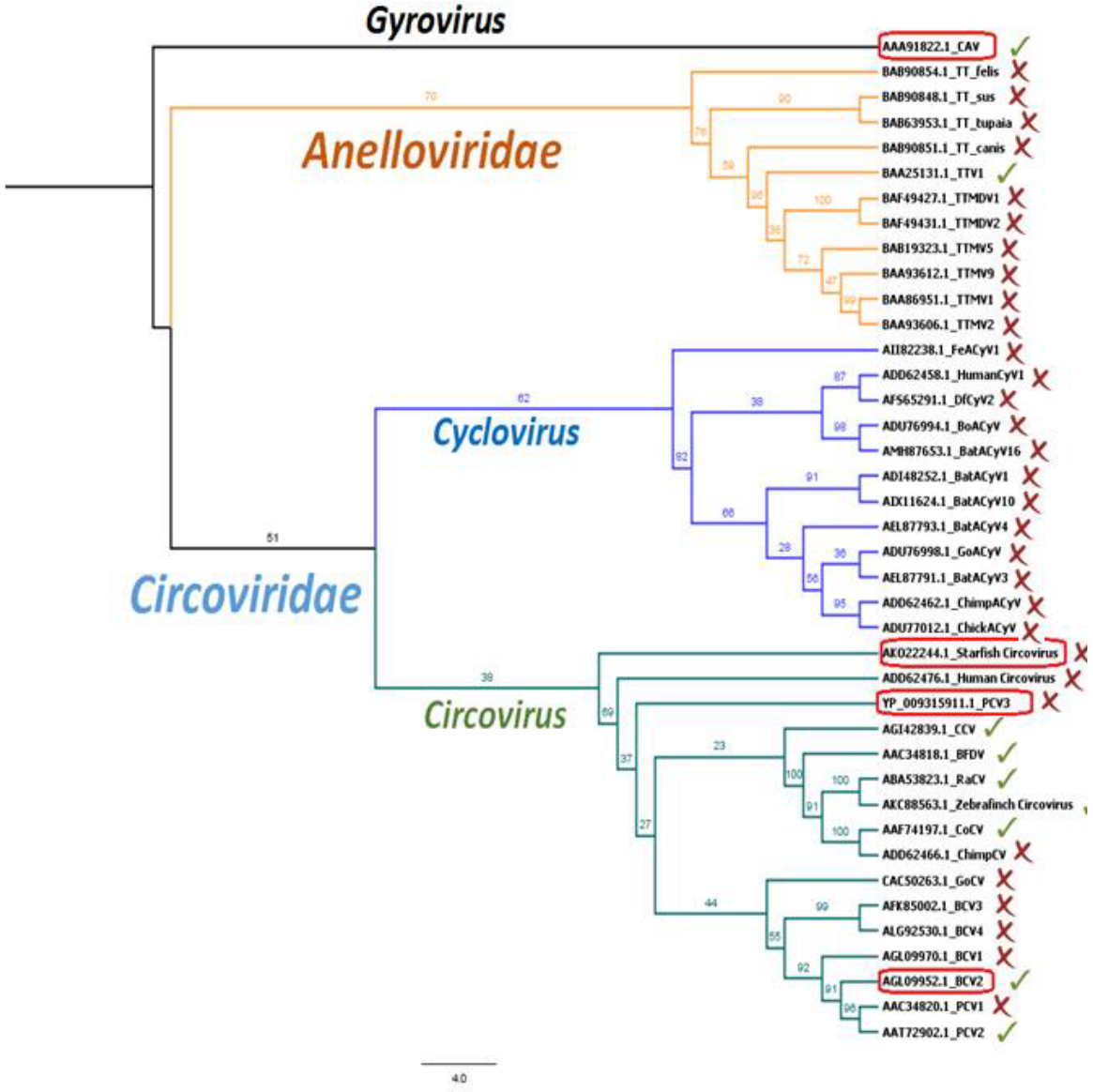


Figure 2.18 Phylogenetic tree outlining which *Circoviridae* and *Anelloviridae* have a been successfully expressed

Green tick indicates the virus proteins that have been successfully expressed. Red crosses indicate the virus protein that have not been successfully expressed.

As displayed in Figure 2.17, successful expression of Anelloviridae and Circoviridae is sparse. The novelty of these two virus families means there is a lack of knowledge and/or studies are in their initial phases.

2.4.1 Bat Associated Circovirus (BCV)

The solving of the full length BCV X-ray crystal structure in our laboratory in 2017 provided a structural basis for the formation of VLPs by BCV constructs. Recombinant expression of BCV X was successfully achieved through those methods used to solve BCV.

The BCV X + DNA complex structure was solved to a resolution of 4.3 Å. Positive internal densities, shown in Table 2.7, show DNA may be present inside the VLP. These positive internal densities were no present in the previously solved BCV X VLP, indicating this is due to the addition of DNA prior to hanging drop crystallography being undertaken. Due to the low resolution however this cannot be modelled.

2.4.2 Chicken Anaemia Virus (CAV)

All CAV constructs were difficult to express in the initial stages of the project. The initial mass expression trail of all three CAV constructs showed that CAV VP1 X and CAV VP1 XL expressed in the Rosetta 2 (DE3) cell line (see Figure 2.6 & Figure 2.7). Whilst CAV VP1 did not express in Rosetta 2 (DE3) cell line (see Figure 2.5). This small scale expression provides a basis for expression on a large scale.

As CAV VP1 X and CAV VP1 XL were insoluble, common laboratory solubilisation techniques were applied but however failed (see Figure 2.8 & Figure 2.9). The common solubilisation technique of resuspension in 8M Urea may have failed due to its effectiveness being modulated by pH and ionic strength. Harsher solubilisation techniques were applied to solubilise the construct from inclusion bodies, with the addition of 1% SDS proving to be effective in solubilising some of CAV VP1 XL from inclusion bodies. There is however still improvement to be made in order to increase the solubility of CAV VP1 XL. Extracting proteins by this method result in denaturation (unfolding) of the protein by the strong detergent and pH. Protein denaturation, in the case of pH, occurs due to the ionisation of side chains, generally the higher the pH the less residual structure retained by the protein. Sodium dodecyl sulphate (SDS) is a common and effective protein denaturation detergent as it binds directly to the protein. Denaturation of the protein for extraction from inclusion bodies then requires refolding, with refolding yields often being hindered by the denaturation processes. Protein can be purified denatured and renatured, however the main danger of working with an unfolded protein is it vulnerability to chemical modification and sensitivity to contaminating protease activity (82, 84). Small scale solubilisation provides a basis for solubilisation on a large scale, with the SDS solubilisation method already being adopted by others in the laboratory.

2.4.3 Porcine Circovirus 3 (PCV-3)

Expression of PCV-3 was problematic with IPTG and auto-induction expression methods failing in both Caps Buffer A and HIS Buffer A. This was unusual as PCV-3

was predicted to express similar to BCV as the cap and rep proteins of PCV-3 are 55% identical to those of bat associated circovirses (55). Both expression methods IPTG and auto-induction failed to express PCV-3 in Caps Buffer and HIS Buffer.

2.4.4 Starfish associated circular virus (AfaCV)

Initial expression of AfaCV was unsuccessful via auto- induction in HIS Buffer A. This may be due to AfaCV being highly divergent from other circoviruses, as demonstrated by the phylogenetic tree Figure 2.17 (60).

Expression in all experiments may have been hindered with a suspected flaw in the Rosetta 2 (DE3) cell line. Due to the wait on the new cell line this theory is yet to be tested, however other experiments in the laboratory are having difficulty with expression.

2.4.5 Future studies

2.4.5.1 Bat Associated Circovirus (BCV)

Future experiments will focus on gaining higher resolution to model internal DNA take up. This will be achieved through recombinant expression and trailing and optimising crystal screen conditions.

2.4.5.2 Chicken Anaemia Virus (CAV)

Now that CAV VP1 X and CAV VP1 XL have been successfully expressed and solubilised on a small scale, this should be applied to a large scale trial. Trialling of refolding techniques to gain the highest yield of solubilised refolded protein can then be undertaken. If this is successful, crystallisation trails can begin in an attempt to solve X-ray crystal structures.

Further experiments are required in order to achieve expression of CAV VP1. Trials may include IPTG expression method in HIS Buffer A, and failing this auto-induction with both Caps and HIS buffers.

2.4.5.3 Porcine Circovirus3 (PCV-3)

Further experiments are required to test if the lack of experession was due to an issue with the Rosetta (DE3) cell line or if a different cell line all together is needed for successful expression of PCV-3.

2.4.5.4 Starfish associate circular virus (AfaCV)

Further experiments are required to achieve expression, trialling both IPTG and auto induction with Caps Buffer and HIS Buffer. Failing this it is recommended to trial different cell lines instead of Rosetta 2 (DE3).

2.4.6 Conclusion

The key aims of this honours thesis were to characterise the structure of BCV X with and without DNA using crystallography, design and test recombinant expression constructs of Starfish Cap protein and test recombinant expression of PCV-3 and CAV.

Successful expression and x-ray diffraction of BCV X + DNA VLP was achieved however the resolution achieved was not high enough to model the complex. Unfortunately expression of both PCV-3 and AfaCV was unsuccessful in this study due to time restraints and/or cell line issues. Achievement of expression of CAV VP1 X and CAV VP1 XL and subsequent solubilisation will provide a basis for expression and solubilisation of the two CAV constructs in the future studies. Overall, the preliminary results achieved in this study will be the basis for future work in the laboratory.

These findings indicate that there is now a known expression method for CAV constructs VP1 X and VP1 XL, and BCV X. Although expression trails for PCV-3 and AfaCV were unsuccessful, these initial trials will provide a base line for future trials.

References

1. Sarker S, Terron MC, Khandokar Y, Aragao D, Hardy JM, Radjainia M, Jimenez-Zaragoza M, de Pablo PJ, Coulibaly F, Luque D, Raidal SR, Forwood JK. 2016. Structural insights into the assembly and regulation of distinct viral capsid complexes. Nat Commun 7:13014.

2. International Committee on Taxonomy of Viruses. 2016. Virus Taxonomy. https://talk.ictvonline.org/taxonomy/. Accessed July 2017.

3. Zhang XX, Liu SN, Xie ZJ, Kong YB, Jiang SJ. 2012. Complete genome sequence analysis of duck circovirus strains from Cherry Valley duck. Virol Sin 27:154-64.

4. Mankertz A. 2010. Circoviruses. In Mahy BWJ, van Regenmortel MHV (ed), Desk Encylopedia Animal and Bacterial Virology. Academic Press,

5. Mankertz A. 2008. Animal Viruses: Molcular Biology. In Mettenleiter TC, Sobrino F (ed), Horizon Scientific Press,

6. Rosario K, Breitbart M, Harrach B, Segales J, Delwart E, Biagini P, Varsani A. 2017. Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. Arch Virol 162:1447-1463.

7. Wildlife Health Australia. 2014. Psittacine Beak and Feather Disease (PBFD) https://www.wildlifehealthaustralia.com.au/Portals/0/Documents/FactSheets/Avian/Psittacine%20Beak%20and%20Feather%20Disease%20(PBFD)%20Jun%202014%20(2.1).pdf. Accessed June 2017.

8. Australian Government: Department of the Environment and Heritage. 2004. Beak and feather disease (psittacine circoviral disease). http://www.ea.gov.au/biodiversity/threatened/tap/beakandfeather/. Accessed May 2017.

9. Crowther RA, Berriman JA, Curran WL, Allan GM, Todd D. 2003. Comparison of the Structures of Three Circoviruses: Chicken Anemia Virus, Porcine Circovirus Type 2, and Beak and Feather Disease Virus. J Virol 77:13036-13041.

10. Todd D. 2000. Circoviruses: immunosuppressive threats to avian species: a review. Avian Pathol 29:373-94.

11. Franzo G, Cortey M, Olvera A, Novosel D, Castro AM, Biagini P, Segales J, Drigo M. 2015. Revisiting the taxonomical classification of Porcine Circovirus type 2 (PCV2): still a real challenge. Virol J 12:131.

12. Gillespie J, Opriessnig T, Meng XJ, Pelzer K, Buechner-Maxwell V. 2009. Porcine Circovirus Type 2 and Porcine Circovirus-Associated Disease. Journal of Veterinary Internal Medicine 23:1151-1163.

13. Swiss Institute of Bioinformatics. 2008. Circovirus. http://viralzone.expasy.org/118?outline=all_by_species. Accessed July 2017.

14. Swiss Institute of Bioinformatics. 2008. Cyclovirus. http://viralzone.expasy.org/7296?outline=all_by_species. Accessed September 2017.

15. Swiss Institute of Bioinformatics. 2008. Gyrovirus. http://viralzone.expasy.org/117. Accessed July 2017.

16. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A. 2011. Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J Gen Virol 92:1302-8.

17. Gibbs MJ, Weiler, G. F,. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombines with a vertebrate-infecting virus. Proc Natl Acad Sci U S A Biol Sci 96:8022-8027.

18. Cheung AK. 2003. Transcriptional Analysis of Porcine Circovirus Type 2. Virology 305:168-180.

19. Niagro FD, Forsthoefel, A. N, Lawther, R. P, Kamalanathan, L., Ritchie, B. W, Latimer, K. S, Lukert, P. D,. 1998. Beak and feather disease and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. Arch Virol 143:1723-1744.

20. Bassami MR, Berryman, D., Wilcox, G. E, Raidal, S. R,. 1998. Psittacine Beak and Feather Disease Virus Nucleotide Sequence Analysis and Its Relationship to Porcine Circovirus, Plant Circoviruses and Chicken Anaemia Virus. Virology 249:453-459.

21. International Committee on Taxonomy of Viruses (ICTV). 2017. Circoviridae. https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/circoviridae. Accessed September 2017.

22. Nguyen M, Haenni A-L. 2003. Expression strategies of ambisense viruses. Virus Research 93:141-150.

23. International Committee on Taxonomy of Viruses (ICTV). 2017. Anelloviridae. https://talk.ictvonline.org/ictv-reports/ictv_9th_report/ssdna-viruses-2011/w/ssdna_viruses/139/anelloviridae. Accessed September 2017.

24. ML Lepherd PC, GB Hunt, PC Thomson and KL Bosward. 2011. Assessment of the short-term systemic effect of and acute phase response to mulesing and other options for controlling breech flystrike in Merino lambs. The Journal of the Australian Veterinary Association LTD 89:19-26.

25. Ruiz-Maso. J A, C. M, L. B-R, M. E, M. C, G. DS. 2015. Plasmid rolling-circle replication. Microbiology Spectrum 3.

26. Brown TA (ed). 2002. Genomes, 2nd edition. Oxford: Wiley-Liss, https://www.ncbi.nlm.nih.gov/books/NBK21113/figure/A8120/. Accessed

27. Vega-Rocha S, Gronenborn B, Gronenborn AM, Campos-Olivas R. 2007. Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circovirus structures. Biochemistry 46:6201-12.

28. Gibbs MJ, Smeianov VV, Steele JL, Upcroft P, Efimov BA. 2006. Two families of rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes. Mol Biol Evol 23:1097-100.

29. Willems H, Hofmeister, R., Reiner, G. 2014. Complete Genome sequences of two porcine circovirus type 2 field isolates bearing an unusual sequence duplication in the rep gene. Genome Announcements 2.

30. Cheung AK. 2012. Porcine circovirus: transcription and DNA replication. Virus Res 164:46-53.

31. Vega-Rocha S, Byeon IJ, Gronenborn B, Gronenborn AM, Campos-Olivas R. 2007. Solution structure, divalent metal and DNA binding of the endonuclease domain from the replication initiation protein from porcine circovirus 2. J Mol Biol 367:473-87.

32. Cheung AK. 2015. Specific functions of the Rep and Rep proteins of porcine circovirus during copy-release and rolling-circle DNA replication. Virology 481:43-50.

33. Fort M, Sibila M, Nofrarias M, Perez-Martin E, Olvera A, Mateu E, Segales J. 2010. Porcine circovirus type 2 (PCV2) Cap and Rep proteins are involved in the

development of cell-mediated immunity upon PCV2 infection. Vet Immunol Immunopathol 137:226-34.

34. Mankertz A, Hillenbrand B. 2001. Replication of porcine circovirus type 1 requires two proteins encoded by the viral rep gene. Virology 279:429-38.

35. Llyina TV, Koonin EV. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA rpelication encoded by diverse replicons from eubacteria, eucaryotes and arachaebacteria. Nucleic Acids Res 20:3279-3285.

36. Akbar Behjatnia SA, Dry IA, Ali Rezaian M. 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. Nucleic Acids Res 26:925-931.

37. Mangala Prasad V, Willows SD, Fokine A, Battisti AJ, Sun S, Plevka P, Hobman TC, Rossmann MG. 2013. Rubella virus capsid protein structure and its role in virus assembly and infection. Proc Natl Acad Sci U S A 110:20105-10.

38. Cheng S, Brooks CL. 2013. Viral Capsid Proteins Are Segregated in Structural Fold Space. PLoS Comp Biol 9.

39. Lucas W. 2010. Viral Capsids and Envelopes: Structure and Function. doi:10.1002/9780470015902.a0001091.pub2.

40. Misinzo G, Delputte PL, Meerts P, Lefebvre DJ, Nauwynck HJ. 2006. Porcine circovirus 2 uses heparan sulfate and chondroitin sulfate B glycosaminoglycans as receptors for its attachment to host cells. J Virol 80:3487-94.

41. Peng Z, Ma T, Pang D, Su D, Chen F, Chen X, Guo N, Ouyang T, Ouyang H, Ren L. 2016. Expression, purification and antibody preparation of PCV2 Rep and ORF3 proteins. Int J Biol Macromol 86:277-81.

42. Liu Z, Guu TS, Cao J, Li Y, Cheng L, Tao YJ, Zhang J. 2016. Structure determination of a human virus by the combination of cryo-EM and X-ray crystallography. Biophys Rep 2:55-68.

43. Egli M. 2010. Diffraction Techniques in Structural Biology. Nucleic Acid Chemistry.

44. Khayat R, Brunn N, Speir JA, Hardham JM, Ankenbauer RG, Schneemann A, Johnson JE. 2011. The 2.3-angstrom structure of porcine circovirus 2. J Virol 85:7856-62.

45. Zanotti G. 2016. Cryo-EM and X-Ray Crystallography: Complementary or Alternative Techniques? NanoWorld Journal 2:22-23.

46. Cheng Y. 2015. Single-Particle Cryo-EM at Crystallographic Resolution. Cell 161:450-7.

47. Llyina RV, Koonin, E.V. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicaons from eubacteria, eucaryotes and archaebacteria. Nucleic Acid Research 20.

48. Koonin EV. 1993. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPase including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acid Research 21:2541-2547.

49. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479-480:2-25.

50. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol 157:1851-71.

51. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjango JB, Peeters M, Gross-Camp ND, Muller MN, Hahn BH, Wolfe ND, Triki H, Bartkus J, Zaidi SZ, Delwart E. 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. J Virol 84:1674-82.

52. Delwart E, Li L. 2012. Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. Virus Res 164:114-21.

53. Viruses ICoTo. 2017. Circoviridae Genus: Cyclovirus. https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/circoviridae/660/genus-cyclovirus. Accessed September 2017.

54. Viruses ICoTo. 2017. Circoviridae Genus: Circovirus. https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/circoviridae/659/genus-circovirus. Accessed September 2017.

55. Palinski R, Pineyro P, Shang P, Yuan F, Guo R, Fang Y, Byers E, Hause BM. 2017. A Novel Porcine Circovirus Distantly Related to Known Circoviruses Is Associated with Porcine Dermatitis and Nephropathy Syndrome and Reproductive Failure. J Virol 91.

56. Franzo G, Tucciarone CM, Cecchinato M, Drigo M. 2016. Porcine circovirus type 2 (PCV2) evolution before and after the vaccination introduction: A large scale epidemiological study. Sci Rep 6:39458.

57. S Krakowka JE, F McNeilly, C Waldner, G Allan. 2005. Features of porcine circovirus-2 disease: correlations between lesions, amount an ddistribution of virus, and clinical outcome. Journal of Veterinarian Diagnostic Investigation 17:213-222.

58. Finsterbusch T, Mankertz A. 2009. Porcine circoviruses--small but powerful. Virus Res 143:177-83.

59. Phan TG, Giannitti F, Rossow S, Marthaler D, Knutson TP, Li L, Deng X, Resende T, Vannucci F, Delwart E. 2016. Detection of a novel circovirus PCV3 in pigs with cardiac and multi-systemic inflammation. Virol J 13:184.

60. Fahsbender E, Hewson I, Rosario K, Tuttle AD, Varsani A, Breitbart M. 2015. Discovery of a novel circular DNA virus in the Forbes sea star, Asterias forbesi. Arch Virol 160:2349-51.

61. Munang'andu HM, Mugimba KK, Byarugaba DK, Mutoloki S, Evensen O. 2017. Current Advances on Virus Discovery and Diagnostic Role of Viral Metagenomics in Aquatic Organisms. Front Microbiol 8:406.

62. Chiappetta CM, Cibulski SP, de Sales Lima FE, Varela AP, Amorim DB, Tavares M, Roehe PM. 2017. Molecular Detection of Circovirus and Adenovirus in Feces of Fur Seals (Arctocephalus spp.). EcoHealth 14:69-77.

63. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z. 2012. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. J Virol 86:4620-30.

64. Han HJ, Wen HL, Zhao L, Liu JW, Luo LM, Zhou CM, Qin XR, Zhu YL, Liu MM, Qi R, Li WQ, Yu H, Yu XJ. 2017. Novel coronaviruses, astroviruses, adenoviruses and circoviruses in insectivorous bats from northern China. Zoonoses Public Health doi:10.1111/zph.12358.

65. de Sales Lima FE, Cibulski SP, Bello AGD, Mayer FQ, Witt AA, Roehe PM, d'Azevedo PA. 2015. A Novel Chiropteran Circovirus Genome Recovered from a Brazilian Insectivorous Bat Species. Genome Announcements 3:1393-1415.

66. Ge X, Li J, Peng C, Wu L, Yang X, Wu Y, Zhang Y, Shi Z. 2011. Genetic diversity of novel circular ssDNA viruses in bats in China. J Gen Virol 92:2646-53.

67. de Sales Lima FE, Cibulski SP, Dos Santos HF, Teixeira TF, Varela AP, Roehe PM, Delwart E, Franco AC. 2015. Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. PLoS ONE 10:e0118070.

68. Wang Y, Song X, Gao H, Wang X, Hu Y, Gao Y, Qi X, Qin L, Lin H, Gao L, Yao S, Han C, Wang X, Chen H. 2017. C-terminal region of apoptin affects chicken anemia virus replication and virulence. Virol J 14:38.

69. Smuts HE. 2014. Novel Gyroviruses, including Chicken Anaemia Virus, in Clinical and Chicken Samples from South Africa. Adv Virol 2014:321284.

70. Lima FE, Cibulski SP, Dos Santos HF, Teixeira TF, Varela AP, Roehe PM, Delwart E, Franco AC. 2015. Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. PLoS ONE 10:e0118070.

71. Adair BM. 2000. Immunopathogenesis of chicken anemia virus infection. Developmental and Comparitive Immunology 24:247-255.

72. Wani MY, Dhama K, Latheef SK, Singh SD, Tiwari R. 2014. Correlation between cytokine profi le, antibody titre and viral load during sub-clinical chicken anaemia virus infection. Veterinarian Medicine 59:33-43.

73. Noteborn MH, Koch G. 1995. Chicken anaemia virus infection: molecular basis of pathogenicity. Avian Pathol 24:11-31.

74. Noteborn MHM, Todd D, Verschueren AJ, de Gauw HWFM, Curran WL, Veldkamp S, Douglas AJ, McNulty MS, Van der Eb EB, Koch G. 1993. A Single Chicken Anemia Virus Protein Induces Apoptosis. J Virol:346-351.

75. Zhang X, Liu Y, Ji J, Chen F, Sun B, Xue C, Ma J, Bi Y, Xie Q. 2014. Identification of a chicken anemia virus variant-related gyrovirus in stray cats in china, 2012. Biomed Res Int 2014:313252.

76. Lee DA. 2012. Porcine Circovirus Associated Diseases. http://www.dpi.nsw.gov.au/__data/assets/pdf_file/0003/436827/Porcine-Circovirus-Associated-Disease.pdf. Accessed May 2017.

77. Noteborn MH, Verschueren CA, Van Roozelaar DJ, Veldkamp S, Van Der Eb AJ, de Boer GF. 1992. Detection of chicken anaemia virus by DNA hybridization and polymerase chain reaction. Avian Pathol 21:107-18.

78. Todd D, Mawhinney KA, Graham DA, Scott ANJ. 1999. Development of a blocking enzyme-linked immunosorbent assay for the serological diagnosis of chicken anaemia virus. J Virol Methods 82:177-184.

79. Kim HR, Park YR, Lim DR, Park MJ, Park JY, Kim SH, Lee KK, Lyoo YS, Park CK. 2017. Multiplex real-time polymerase chain reaction for the differential detection of porcine circovirus 2 and 3. J Virol Methods 250:11-16.

80. Allan GM, Ellis, J. A.,. 2000. Porcine circoviruses: a reivew. J Vet Diagn Investig 12:3-14.

81. International Federation for Animal Health Europe. 2016. Circovirus Fact Sheet. http://www.ifaheurope.org/food-producing-animals/success-stories/circovirus.html. Accessed May 2017.

82. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. 2015. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Fact 14:41.

83. Garcia-Fruitos E. 2010. Inclusion bodies: a new concept. Microb Cell Fact 9:80.

84. Palmer I, Wingfield PT. 2004. Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. Curr Protoc Protein Sci Chapter 6:Unit 6 3.

85. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr 67:271-81.

86. Evans P. 2006. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62:72-82.

87. Evans PR. 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr 67:282-92.

88. Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69:1204-14.

89. Stols L, Zhou M, Eschenfeldt WH, Millard CS, Abdullah J, Collart FR, Kim Y, Donnelly MI. 2007. New vectors for co-expression of proteins: structure of Bacillus subtilis ScoAB obtained by high-throughput protocols. Protein Expr Purif 53:396-403.

90. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. 2006. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Expr Purif 47:446-54.

91. Das S. 2017. Insights into the molecular, evolutionary and structural biology of circoviruses. Charles Sturt Univeristy, Unpublished.

92. Stevenson J, Krycer, J.R., Phan, L., Brown, A.J.,. 2013. A Practical Comparison of Ligation-Independent Cloning Techniques. PLoS ONE 8:e83888.

93. Journal of Virology. 2017. Instructions to authors. jvi.asm.org. Accessed August 2017.

Appendix 1- DNA sequences of target virus proteins

BCV NLS

243 residues 729 base sequence

ATGGTGTATCGCCGCCGCCGCGGCCGCGGCCGCCGCGCGCGCCCGATGA
GCAGCCTGGGCCGCCTGCTGTATCGCAAACCGTGGCTGATGCATCCGCGCTTTCG
CGCGCGCTATCGCTGGCGCCGCAAAAACGGCATTACCAACCTGCGCCTGACCCG
CCAGGTGGAACTGTGGGTGCCGAAAGATGCGGCGAACGCGAGCTTTTATGTGAAC
CATTATACCTTTGATCTGGATGATTTTATTCCGGCGGGCACCCAGCTGAACAGCAG
CCCGCTGCCGTTTAAATATTATCGCATTCGCAAAGTGAAAGTGGAATTTCAGCCGC
GCCTGCCGATTACCAGCCCGTTTCGCGGCTATGGCAGCACCGTGCCGATTCTGGA
TGGCGCGTTTGTGACCCCGGCGACCGGCGAAAGCGATCCGATTTGGGATCCGTAT
ATTAACTTTAGCGGCCGCCATGTGATTCGCACCCCGGCGTGGTATCATAAACGCTA
TTTTACCCCGAAACCGCTGATTGATGGCAACACCGGCTTTTTTCAGCCGAACAACA
AACAGAACGCGCTGTGGTTTCCGAACAAACAGGGCCAGAACATTCAGTGGAGCGG
CCTGGGCTTTGCGATGCAGAAAGGCAACGAAGCGTATAACTATCAGGTGCGCTTT
ACCCTGTATGTGCAGTTTCGCGAATTTGATCTGTTTAACAACAAATATACCGCGCAT
ATGGATGTGCCGCTG

BCV X

201 residues 603 base sequence

AAAAACGGCATTACCAACCTGCGCCTGACCCGCCAGGTGGAACTGTGGGT
GCCGAAAGATGCGGCGAACGCGAGCTTTTATGTGAACCATTATACCTTTGATCTGG
ATGATTTTATTCCGGCGGGCACCCAGCTGAACAGCAGCCCGCTGCCGTTTAAATAT
TATCGCATTCGCAAAGTGAAAGTGGAATTTCAGCCGCGCCTGCCGATTACCAGCC
CGTTTCGCGGCTATGGCAGCACCGTGCCGATTCTGGATGGCGCGTTTGTGACCCC
GGCGACCGGCGAAAGCGATCCGATTTGGGATCCGTATATTAACTTTAGCGGCCGC
CATGTGATTCGCACCCCGGCGTGGTATCATAAACGCTATTTTACCCCGAAACCGCT
GATTGATGGCAACACCGGCTTTTTTCAGCCGAACAACAAACAGAACGCGCTGTGG

TTTCCGAACAAACAGGGCCAGAACATTCAGTGGAGCGGCCTGGGCTTTGCGATGC
AGAAAGGCAACGAAGCGTATAACTATCAGGTGCGCTTTACCCTGTATGTGCAGTTT
CGCGAATTTGATCTGTTTAACAACAAATATACCGCGCATATGGATGTGCCGCTG

CAV VP1 NLS

447 residues 1341 base sequence

CGCCGCGCGCGCCGCCCGCGCGGCCGCTTTTATGCGTTTCGCCGCGGCC
GCTGGCATCATCTGAAACGCCTGCGCCGCCGCTATAAATTTCGCCATCGCCGCCG
CCAGCGCTATCGCCGCCGCGCGTTTCGCAAAGCGTTTCATAACCCGCGCCCGGG
CACCTATAGCGTGCGCCTGCCGAACCCGCAGAGCACCATGACCATTCGCTTTCAG
GGCGTGATTTTTCTGACCGAAGGCCTGATTCTGCCGAAAAACAGCACCGCGGGCG
GCTATGCGGATCATATGTATGGCGCGCGCGTGGCGAAAATTAGCGTGAACCTGAA
AGAATTTCTGCTGGCGAGCATGAACCTGACCTATGTGAGCAAACTGGGCGGCCCG
ATTGCGGGCGAACTGATTGCGGATGGCAGCAAAAGCCAGGCGGCGGAAAACTGG
CCGAACTGCTGGCTGCCGCTGGATAACAACGTGCCGAGCGCGACCCCGAGCGCG
TGGTGGCGCTGGGCGCTGATGATGATGCAGCCGACCGATAGCTGCCGCTTTTTTA
ACCATCCGAAACAGATGACCCTGCAGGATATGGGCCGCATGTTTGGCGGCTGGCA
TCTGTTTCGCCATATTGAAACCCGCTTTCAGCTGCTGGCGACCAAAAACGAAGGCA
GCTTTAGCCCGGTGGCGAGCCTGCTGAGCCAGGGCGAATATCTGACCCGCCGCG
ATGATGTGAAATATAGCAGCGATCATCAGAACCGCTGGCGCAAAGGCGAACAGCC
GATGACCGGCGGCATTGCGTATGCGACCGGCAAAATGCGCCCGGATGAACAGCA
GTATCCGGCGATGCCGCCGGATCCGCCGATTATTACCAGCACCACCGCGCAGGG
CACCCAGGTGCGCTGCATGAACAGCACCCAGGCGTGGTGGAGCTGGGATACCTA
TATGAGCTTTGCGACCCTGACCGCGCTGGGCGCGCAGTGGAGCTTTCCGCCGGG
CCAGCGCAGCGTGAGCCGCCGCAGCTTTAACCATCATAAAGCGCGCGGCGCGGG
CGATCCGAAAGGCCAGCGCTGGCATACCCTGGTGCCGCTGGGCACCGAAACCAT
TACCGATAGCTATATGGGCGCGCCGGCGAGCGAAATTGATACCAACTTTTTTACCC
TGTATGTGGCGCAGGGCACCAACAAAAGCCAGCAGTATAAATTTGGCACCGCGAC
CTATGCGCTGAAAGAACCGGTGATGAAAAGCGATAGCTGGGCGGTGGTGCGCGT

GCAGAGCGTGTGGCAGCTGGGCAACCGCCAGCGCCCGTATCCGTGGGATGTGAA
CTGGGCGAACAGCACCATGTATTGGGGCAGCCAGCCG

CAV VP1 X

396 residues 1188 base sequence

CCGGGCACCTATAGCGTGCGCCTGCCGAACCCGCAGAGCACCATGACCAT
TCGCTTTCAGGGCGTGATTTTTCTGACCGAAGGCCTGATTCTGCCGAAAAACAGCA
CCGCGGGCGGCTATGCGGATCATATGTATGGCGCGCGCGTGGCGAAAATTAGCG
TGAACCTGAAAGAATTTCTGCTGGCGAGCATGAACCTGACCTATGTGAGCAAACTG
GGCGGCCCGATTGCGGGCGAACTGATTGCGGATGGCAGCAAAAGCCAGGCGGC
GGAAAACTGGCCGAACTGCTGGCTGCCGCTGGATAACAACGTGCCGAGCGCGAC
CCCGAGCGCGTGGTGGCGCTGGGCGCTGATGATGATGCAGCCGACCGATAGCTG
CCGCTTTTTTAACCATCCGAAACAGATGACCCTGCAGGATATGGGCCGCATGTTTG
GCGGCTGGCATCTGTTTCGCCATATTGAAACCCGCTTTCAGCTGCTGGCGACCAA
AAACGAAGGCAGCTTTAGCCCGGTGGCGAGCCTGCTGAGCCAGGGCGAATATCT
GACCCGCCGCGATGATGTGAAATATAGCAGCGATCATCAGAACCGCTGGCGCAAA
GGCGAACAGCCGATGACCGGCGGCATTGCGTATGCGACCGGCAAAATGCGCCCG
GATGAACAGCAGTATCCGGCGATGCCGCCGGATCCGCCGATTATTACCAGCACCA
CCGCGCAGGGCACCCAGGTGCGCTGCATGAACAGCACCCAGGCGTGGTGGAGCT
GGGATACCTATATGAGCTTTGCGACCCTGACCGCGCTGGGCGCGCAGTGGAGCTT
TCCGCCGGGCCAGCGCAGCGTGAGCCGCCGCAGCTTTAACCATCATAAAGCGCG
CGGCGCGGGCGATCCGAAAGGCCAGCGCTGGCATACCCTGGTGCCGCTGGGCA
CCGAAACCATTACCGATAGCTATATGGGCGCGCCGGCGAGCGAAATTGATACCAA
CTTTTTTACCCTGTATGTGGCGCAGGGCACCAACAAAAGCCAGCAGTATAAATTTG
GCACCGCGACCTATGCGCTGAAAGAACCGGTGATGAAAAGCGATAGCTGGGCGG
TGGTGCGCGTGCAGAGCGTGTGGCAGCTGGGCAACCGCCAGCGCCCGTATCCGT
GGGATGTGAACTGGGCGAACAGCACCATGTATTGGGGCAGCCAGCCG

CAV VP1 XL

409 residues 1227 base sequence

TATCGCCGCCGCCGCCGCCATTTTCGCCGCCGCCGCTTTCCGGGCACCTA
TAGCGTGCGCCTGCCGAACCCGCAGAGCACCATGACCATTCGCTTTCAGGGCGTG
ATTTTTCTGACCGAAGGCCTGATTCTGCCGAAAAACAGCACCGCGGGCGGCTATG
CGGATCATATGTATGGCGCGCGCGTGGCGAAAATTAGCGTGAACCTGAAAGAATT
TCTGCTGGCGAGCATGAACCTGACCTATGTGAGCAAACTGGGCGGCCCGATTGCG
GGCGAACTGATTGCGGATGGCAGCAAAAGCCAGGCGGCGGAAAACTGGCCGAAC
TGCTGGCTGCCGCTGGATAACAACGTGCCGAGCGCGACCCCGAGCGCGTGGTGG
CGCTGGGCGCTGATGATGATGCAGCCGACCGATAGCTGCCGCTTTTTTAACCATC
CGAAACAGATGACCCTGCAGGATATGGGCCGCATGTTTGGCGGCTGGCATCTGTT
TCGCCATATTGAAACCCGCTTTCAGCTGCTGGCGACCAAAAACGAAGGCAGCTTTA
GCCCGGTGGCGAGCCTGCTGAGCCAGGGCGAATATCTGACCCGCCGCGATGATG
TGAAATATAGCAGCGATCATCAGAACCGCTGGCGCAAAGGCGAACAGCCGATGAC
CGGCGGCATTGCGTATGCGACCGGCAAAATGCGCCCGGATGAACAGCAGTATCC
GGCGATGCCGCCGGATCCGCCGATTATTACCAGCACCACCGCGCAGGGCACCCA
GGTGCGCTGCATGAACAGCACCCAGGCGTGGTGGAGCTGGGATACCTATATGAG
CTTTGCGACCCTGACCGCGCTGGGCGCGCAGTGGAGCTTTCCGCCGGGCCAGCG
CAGCGTGAGCCGCCGCAGCTTTAACCATCATAAAGCGCGCGGCGCGGGCGATCC
GAAAGGCCAGCGCTGGCATACCCTGGTGCCGCTGGGCACCGAAACCATTACCGA
TAGCTATATGGGCGCGCCGGCGAGCGAAATTGATACCAACTTTTTTACCCTGTATG
TGGCGCAGGGCACCAACAAAAGCCAGCAGTATAAATTTGGCACCGCGACCTATGC
GCTGAAAGAACCGGTGATGAAAAGCGATAGCTGGGCGGTGGTGCGCGTGCAGAG
CGTGTGGCAGCTGGGCAACCGCCAGCGCCCGTATCCGTGGGATGTGAACTGGGC
GAACAGCACCATGTATTGGGGCAGCCAGCCG

PCV3

213 residues 639 base sequence

CGCCATCGCGCGATTTTTCGCCGCCGCCCGCGCCCGCGCCGCCGCCGCC
GCCATCGCCGCCGCTATGCGCGCCGCCGCCTGTTTATTCGCCGCCCGACCGCGG
GCACCTATTATACCAAAAAATATAGCACCATGAACGTGATTAGCGTGGGCACCCCG
CAGAACAACAAACCGTGGCATGCGAACCATTTTATTACCCGCCTGAACGAATGGGA
AACCGCGATTACCTTTGAATATTATAAAATTCTGAAAATGAAAGTGACCCTGAGCCC
GGTGATTAGCCCGGCGCAGCAGACCAAAACCATGTTTGGCCATACCGCGATTGAT
CTGGATGGCGCGTGGACCACCAACACCTGGCTGCAGGATGATCCGTATGCGGAA
AGCAGCACCCGCAAAGTGATGACCAGCAAAAAAAAACATAGCCGCTATTTTACCCC
GAAACCGCTGCTGGCGGGCACCACCAGCGCGCATCCGGGCCAGAGCCTGTTTTT
TTTTAGCCGCCCGACCCCGTGGCTGAACACCTATGATCCGACCGTGCAGTGGGGC
GCGCTGCTGTGGAGCATTTATGTGCCGGAAAAAACCGGCATGACCGATTTTTATG
GCACCAAAGAAGTGTGGATTCGCTATAAAAGCGTGCTG

Starfish NLS

347 residues 1041 base sequence

ATGCCGCGCAACACCCGCCGCCGCACCGGCCGCAGCCGCCGCCGCCGCA
CCCTGCGCCGCGGCTATCGCAAATTTACCGGCGGCCGCGGCATTACCAAAGAACA
TGATCGCACCAACATTTATTATAAAAAACGCATGCCGCGCTTTAAAAAACGCCGCT
GGCGCCGCTTTAGCAAAAAAGTGCAGAACGTGAGCGAAAAAGATCTGGGCACCCA
GACCGTGGTGTTTAACGATCAGATTGAAACCGGCGATTTTTTTAACAGCGATGGCC
TGATTACCGGCAGCGCGACCCTGTGCCTGTATAGCATGAACGGCAGCGAAGTGCC
GTATAACGATCTGGCGGTGATGATGAACAACCTGAACTTTGCGGCGCCGACCGTG
GATCTGGGCACCACCAAATATCGCACCAGCAAAGTGATTTTTAAAAGCGCGGTGTT
TGATCTGACCCTGACCAACACCAGCTATGATCGCATTGGCGGCGTGAACAACATT
GAAGCGGCGTATAGCCTGGAAGTGGATATTCATACCGTGGTGATTAAAAAAGGCG
CGCGCGATCTGGATTTTATTAGCGGCACCCTGGCGGAAAAAGGCCTGAAAACCAT
TGATAAAATGATTGAAGTGGCGGAAAAAGAAGTGCGCCCGATTGGCGGCGATGAT
AGCAAAATTTTTAACACCACCAAAGCGAAATTTAACCGCGGCGGCACCCTGTGGGA
TGTGCCGCATGCGCTGAGCCTGTTTGGCATGAAAATTCTGAGCAAAAAAAAATATT

TTATTCCGGGCGGCGCGACCATTACCTATCAGATGCGCGATGCGAAACGCCGCGT
GATGGTGCCGCAGTTTAACATGCTGGGCGTGAACAAACCGGGCTGGACCAAATTT
CTGTATATTACCTATAAACTGGTGCCGGGCGTGGGCGTGATTGGCCAGGATGAAA
ACGTGCCGACCCAGATTCGCACCCGCCTGAGCGTGGGCGTGACCCGCAAATATA
GCTATAAAATTGAAGGCCAGAACGATGTGCGCGATAACTATAACCCGGATCTG

Appendix 2- Journal of Virology Guidelines (93)

References

In the reference list, references are numbered in the order in which they are cited in the article (citationsequence reference system). In the text, references are cited parenthetically by number in sequential order. Data that are not published or not peer reviewed are simply cited parenthetically in the text (see section ii below). (i) References listed in the References section. The following types of references must be listed in the References section: • Journal articles (both print and online) • Books (both print and online) • Book chapters (publication title is required) • Patents • Theses and dissertations • Published conference proceedings • Meeting abstracts (from published abstract books or journal supplements) • Letters (to the editor) • Company publications • In-press journal articles, books, and book chapters • Data sets • Code Provide the names of all the authors and/or editors for each reference; long bylines should not be abbreviated with “et al.” All listed references must be cited in the text. Abbreviate journal names according to the PubMed Journals Database (National Library of Medicine, National Institutes of Health; available at http://www.ncbi.nlm.nih.gov/nlmcatalog/journals), the Instructions to Authors June 2017, Instructions to Authors Journal of Virology jvi.asm.org 11 primary source for ASM style (do not use periods with abbreviated words). The EndNote output style for ASM Journals’ current reference style can be found here; click “Open” and then “Download and Install” to save it to your EndNote Styles folder (it should replace any earlier output styles for ASM journals [all ASM journals use the same reference style]). Note that DOIs are not needed for most references, but see the exception for reference 17 below. ASM copy editors will automatically insert DOIs on all references in the CrossRef and PubMed databases during copyediting. URLs for government reports and other references not indexed in these database should be provided if desired; URLs or DOIs for citations of database accession numbers and code/software should be provided by you.