Investigating the Mechanism of Genome Encapsidation for Avian Nephritis Virus

Conor David Haydon

PhD

University of York

Biology

September 2019

Abstract

Avian astroviruses (AAstVs) cause growth suppression syndromes that place large economic burdens on the global poultry industry. The vast inter-serotype antigenic diversity of these +ssRNA viruses has hindered the production of vaccines, and the emergence of increasingly pathogenic strains means an improved understanding of AAstV biology is urgently required to develop novel control strategies. A growing number of +ssRNA viruses are reported to use packaging signal (PS)-mediated assembly, where multiple RNA stem-loops (SL) across the genome act cooperatively to facilitate efficient genome encapsidation and virion assembly. This mechanism and its sequence-specific characteristics are apparently conserved between strain variants, potentially presenting an ideal antiviral target. Therefore, the existence of PSs in avian nephritis virus (ANV), a prominent AAstV, was explored.

To guide expression of ANV capsid protein (CP) for identification of PSs by RNA SELEX, the CP structure was investigated by computational modelling. ANV CP showed significant structural homology to human astrovirus CP, and an RNA SL was predicted to interact with the interior surface of the capsid by binding-specific substructure homology to panicum mosaic virus. Expression of CP was trialled in plant, avian-derived, and bacterial systems, with intact CP only generated in *E. coli* which was then used for further study. Whilst magnesium ions induced capsomer-like formation by ANV CP, these appeared assembly-incompetent as neither additional magnesium nor viral RNA promoted capsid assembly. In the absence of protein suitable for SELEX, ANV was recovered by reverse genetics for an alternate NGS approach to identify enriched nucleotides within virions. RNA structure analysis at three identified positions revealed a PS-like consensus SL motif. Synonymous mutation of the most conserved SL resulted in delayed accumulation of extracellular virions, indicating an involvement in encapsidation. This thesis presents the first evidence for PS-mediated assembly by ANV, which may be a mechanism employed across the *Astroviridae*.

Table of Contents

List of Tables

List of Figures

Acknowledgements

A big thank you to all those who have helped and supported me over the last four years, including everyone in the AVO group at Pirbright past and present, and those who have enjoyed my 'lively' electronic music preferences whilst working alongside me in the lab.

I would especially like to thank all of my supervisors for their time and guidance – particularly Ashley for teaching me everything I could need to know in the lab, giving me the skills and tools to tackle any experiment or dilemma that came my way. I have truly appreciated the countless hours you've spent going through my work and ideas, I would be surprised if you hadn't questioned my sanity on several occasions. Thank you Reidun and Eric, your incredible cross-discipline expertise gave valuable perspective to help keep the project on track numerous times, and for helping implement bioinformatics analyses that a molecular biologist could only dream of, you've truly earnt the title of 'mathemagicians'.

I would also like to thank Toby for his involvement in the project, for direction when all seemed to be failing, and for making access to the NGS facilities at Pirbright possible with the assistance of the picornavirus group. Pippa, I really appreciate the numerous hours you spent looking for astroviruses under the electron microscope. Thanks to George and Keith at the John Innes Centre, for kindly letting me play with the *Nicotiana* plants in their greenhouses, and to Victoria from the Agri-Food and Biosciences Institute, for astrovirus-related chats and sending me the virus isolates and antibodies, without which this project would not have been possible.

Finally, I would like to thank my family and friends, I'm sure they would like to see and hear from me once again – particularly after the arduous writing up period, which has kept me locked away from almost all social contact! In this case I think it is Sara who needs the biggest thanking, her support and understanding has been incredible, not only has she accompanied me on every high and low over the last four years, but she made sure that I was eating only the best home-cooked meals whilst writing. I'm sure finishing this thesis will be as relieving to her as to me!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

1 Introduction

1.1 Viral genome encapsidation

All organisms encompassing genetic material, whether eukaryotic, prokaryotic or viral, need to condense and package their nucleic acids in order to protect and contain them, whilst simultaneously being able to readily access them for transcription and translation when required. Eukaryotes and prokaryotes utilise histones, histone-like proteins, DNA-binding proteins and an array of other proteins and factors to condense DNA with length in the order of centimetres down to micrometre-size[1, 2]. The condensing of nucleic acid is also a vital step in the life cycle of viruses, where the cognate genome, or segmented genome, needs to be packaged inside of its respective capsid of virally-encoded proteins, so that it is protected from host and environmental degradation as well as able to disseminate in order to deliver its genomic cargo and further its replication. Encapsidation must be rapid and efficient in order to outpace host defence mechanisms, to specifically package the cognate genomic material over excess background host nucleic acids, and to produce multiple viable infectious virions.

1.2 Mechanisms of encapsidation

1.2.1 DNA viruses

The mechanism of packaging for large DNA viruses, in particular double-stranded DNA (dsDNA) viruses, such as the *Herpesviridae* and tailed bacteriophages is through the insertion of the genome into a pre-assembled procapsid. The procapsid is readily formed by capsid protein (CP) and scaffold proteins[3], then the genome is actively introduced through a dodecameric protein portal complex at one of the procapsid vertices by a virally encoded ATP powered motor protein which initiates upon recognition of a specific sequence or structure present in the viral genome[4-8]. The assembly of the procapsid in the absence of nucleic acid shows that all of the information for assembly is encoded within the protein components, i.e. CP (in some cases a combination of major and minor CPs) and scaffold proteins. It is also possible that the portal protein, that in many viruses forms an integral part of the capsid shell, plays a role in nucleation of procapsid assembly[9, 10].

1.2.2 RNA viruses

Many RNA viruses do not form procapsids or encode nucleic acid packaging motors, and so rely on alternative methods in order to encapsidate their genome. The packaging of genomic material in small icosahedral RNA viruses occurs from direct interaction of the nucleic acid with its CPs in a cooperative way.

For a long time, this process had been thought to be purely reliant on non-specific, electrostatic interactions. In this model, positively charged residues on the interior of the capsid are thought to cause condensation and collapse of the negatively charged nucleic acid genome upon binding[11]. This view was informed by a direct correlation observed between genome size and the net charge of the highly basic N- or C-terminal arms of the CP in positive-sense single-stranded RNA (+ssRNA) viruses[12]. It was further supported by experiments on heteroencapsidation and anion-induced capsid formation, where assembly of some +ssRNA viral capsids could be achieved *in vitro* using non-cognate genomes, or by increasing the ionic strength to adequately high levels, respectively[13, 14]. However, spontaneous assembly of capsid around nucleic acid based on electrostatic interactions alone would yield many defective particles and does not explain the observed specificity of the CP to discriminate between cognate genomic and excess background host-cell nucleic acids[15, 16]. It also ignores the efficiency of capsid assembly. For example, capsid assembly in bacteriophage MS2 in the absence of genetic material takes many days, whilst taking only hours in the presence of genetic material[17, 18], suggesting that the genomic RNA plays essential roles in virion formation.

1.2.3 Packaging signal-mediated assembly

Analysing the mechanisms behind genome-promoted assembly revealed that the genomic RNA is instrumental in switching the conformation of the MS2 capsomer, a protein dimer, from a symmetric to an asymmetric form upon contact with an RNA stem-loop (SL)[19]. In the fully formed capsid, there are 60 asymmetric CP dimers located around the 12 particle 5-fold axes, and 30 symmetric CP dimers centred on the 2-fold axes. Statistically, the symmetric and asymmetric form of the CP dimer can interconvert in solution, but this reaction is unfavourable[20]. It is, however, promoted by binding of the MS2 translational operator (TR)[21]. TR binding to a CP dimer initiates capsid assembly, and TR therefore acts as a packaging signal (PS). Studies on the TR indicated that MS2 CP could regulate expression of the viral replicase gene in a molecular switch-like mechanism, where replicase translation is inhibited, and assembly of viral particles is promoted, by CP binding to the TR

SL, giving TR a dual-function as both a translational repressor and as a PS[22]. Analysis of TR binding to the CP dimer revealed that the determinants of binding impose minimal sequence specificity[23, 24], implying that there could be many other SLs in the genome that trigger this effect. Some binding competent variants were known[25, 26], and structural studies revealed many more contacts between genomic RNA and CP at defined positions in MS2[27-30], as had been seen more widely in other +ssRNA viruses[31-37]. This posed the question whether these contacts were random, or if there was a common motif between these SLs, in which case they could all be considered as PSs.

In order to address this in MS2, Hamiltonian Paths Analysis (HPA) was developed[38]. HPA is a combinatorial method that can be used to identify SLs in the secondary structure of a packaged viral genome that are likely to be in contact with the inner capsid shell. Using knowledge of distances between potential binding sites as a constraint set, it interrogates different possible combinations of SLs for those that are consistent with capsid geometry. Using this approach for MS2 it was possible to show that one hemisphere of the capsid is more likely to be occupied by PSs, where organisation is more constrained than the other, and a list of putative PSs were derived[39]. This list agreed well with asymmetric cryo-electron microscopy (Cryo-EM) reconstructions[39, 40], showing computational modelling could be used to predict PS motifs. The collective action of these PSs in capsid formation is known as PS-mediated assembly, a paradigm shift from the electrostatic assembly model[41].

This poses the question of how PSs promote capsid formation, which has been addressed from different perspectives. One aspect of genome packaging is the compaction of the nucleic acid in order to fit the capsid shell. To investigate this, the collapse in hydrodynamic radius (R_h) of genomic RNA upon addition of cognate CP has been studied for MS2 and a plant virus, Satellite Tobacco Necrosis virus (STNV), using single-molecule fluorescence correlation spectroscopy studies. These studies showed that fluorescently labelled genomes of MS2 and STNV undergo a rapid condensation at limiting (low) CP concentrations, and then gradually increase to the R_h expected of properly formed virions[42, 43]. However, no collapse in R_h was observed when non-cognate CP was added to the labelled RNAs, or when non-viral RNAs were assayed. In these cases, particles and CP aggregates were slowly formed instead. This data suggests that the presence of sequence or structure-specific interactions between the genome and its respective CP promote particle assembly through a mechanism known as two-stage assembly, in which the genomic RNA collapses in R_h upon contact with CP, and then gradually expands to the

radius of the fully formed capsid upon recruitment of additional CPs. However, PS-mediated assembly does not always involve this two-stage assembly mechanism, as discussed below.

A second approach to understanding the mechanism of PS-mediated assembly is via modelling. In mathematical graph theory, a Hamiltonian path is a connected path that visits each vertex of a graph exactly once[44]. Considering a viral capsid as a polyhedral shape, as in Caspar Klug theory[45], in which faces represent capsomers, and then connecting these faces in the sequential order in which the capsomers are recruited to the growing capsid shell, results in a Hamiltonian Path representation of capsid assembly (Figure 1.1).

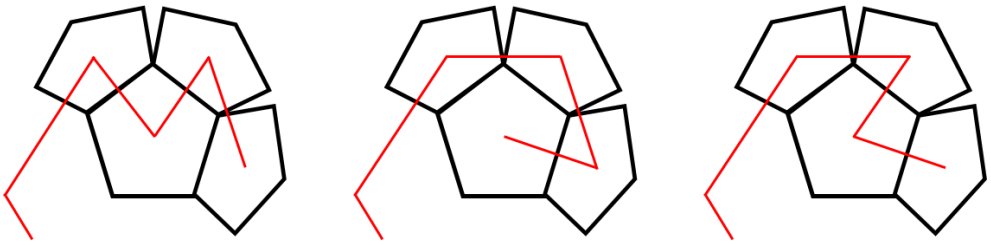


Figure 1.1: A simplified model of capsid assembly following three possible Hamiltonian paths. Genomic RNA (red) specifically interacts once at each RNA-CP binding site on CP subunits (shown simplified as pentamers) in a connected pathway, known mathematically as a Hamiltonian path. Three examples of the different paths an RNA could take are shown.

This concept can therefore be used to study virus assembly. In particular, this approach has been used to address the question of how viruses solve a viral equivalent of a Levinthal's paradox. Levinthal's paradox in protein folding refers to the conundrum of how polypeptide chains fold efficiently into their native, functional state without exploring all combinatorially possible options. For viruses, this becomes the question of how to select the most efficient assembly pathways among the immense number of possible ones. Hamiltonian paths were used to interrogate the outcomes of stochastic simulations of virus assembly that model the formation of a dodecahedral shell from pentagonal units (CP proxy), each of which can interact with PSs with either high (green), intermediate (blue) and low (red) affinity (Figure 1.2)[44, 46, 47]. This analysis revealed that the variation in CP affinities results in selective packaging of viral RNAs against a backdrop of cellular competitor RNAs along the most efficient assembly pathways. It also revealed that this

effect can only be observed at low CP concentrations, typical of a viral infection *in vivo* when CP is initially being produced, but is masked if assembly is considered at higher concentrations. As many capsid assembly studies use high concentrations of CP, this provides an explanation for why the effects of PS-mediated assembly are not observed in some *in vitro* assembly experiments.

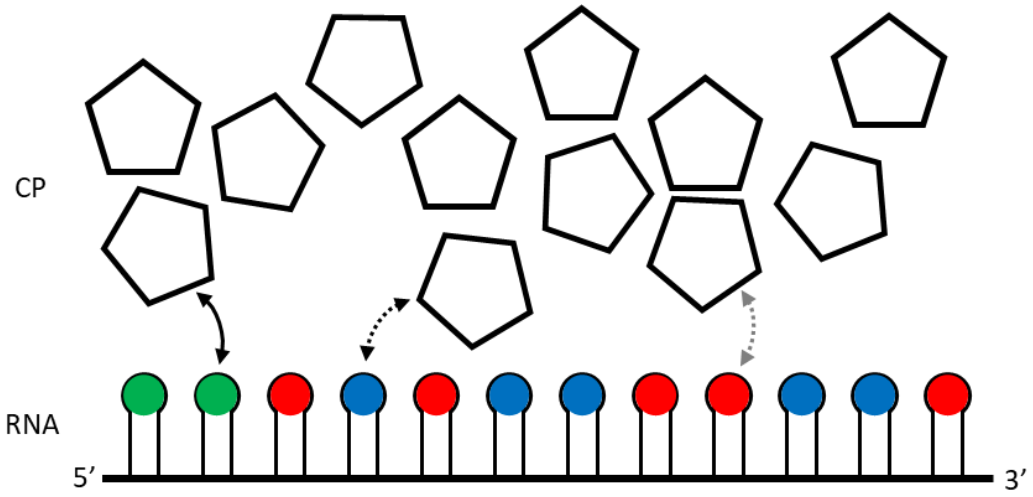


Figure 1.2: PS-mediated assembly reduces the complexity of virus formation. Assembly of CP units onto the genome is directed by PSs with heterogenous binding affinity. As shown by protein ramp modelling, nucleation of capsid assembly occurs at the highest affinity PSs, helping to direct assembly down a limited number of potential pathways and prevent kinetically trapped assembly intermediates. 12 example PSs throughout an RNA molecule are represented with strong (green), intermediate (blue) and weak (red) CP affinities. Examples of potential sites for reversible assembly initiation are represented by arrows, with a solid arrow indicating a more likely nucleation site when low concentrations of CP are present.

Modelling of genome encapsidation for RNAs with PSs of heterogenous affinity to CP, when in the presence of gradually accumulating CP and high levels (300:1) of competitor RNAs, akin to that seen *in vivo*[48,38], showed that PS-mediated assembly of virions was highly effective under these conditions[15, 49]. The accumulation of CP over time decreased the number of kinetically trapped assembly intermediates, as well as the number of potential assembly pathways the CP could take, demonstrating how PS-CP interactions could reduce the complexity of virus formation. The increased efficiency was caused by a reduction in

assembly initiation sites, in a manner where only the PSs with the highest binding affinity for CP promoted nucleation, the remaining PSs were then able to direct recruitment of CP in a controlled and ordered manner to rapidly complete virus assembly. Highly efficient genome packaging at low RNA and CP concentrations is a feature likely preferred by viruses, as the more incorrectly formed viral particles and free CP/RNA produced, the more likely these pathogen-associated molecular patterns (PAMPs) are to stimulate host immune responses, thus restricting viral replication. Therefore, it appears that PS-mediated assembly may provide several significant benefits to a virus: selection of cognate RNAs for packaging, efficient and rapid formation of infectious progeny virions, and evasion of host-cell detection mechanisms.

1.2.4 Different variants of PS-mediated assembly in nature

Whilst the example of MS2, a virus that uses PS-mediated assembly, is discussed in detail above, PS-mediated assembly is not constrained solely to the specifics of this example. For instance, PS interaction with CP can promote capsid formation in other ways than the conformational switching of a dimer from a symmetric to an asymmetric form, and it is not always the case that viral genomes must undergo a collapse in R_h prior to assembly. For example, many viruses package their genomes simultaneously as they are transcribed by the RNA polymerase, in replication complexes, and this co-transcriptional packaging would not require a collapse of the genome prior to packaging. However, PSs can still have an impact on the formation of the capsid shell in these viruses as discussed below. A comprehensive presentation of PS-mediated assembly and viral examples from different families can be found in a recent review article[50].

1.2.4.1 The plant virus STNV

The CP dimer-switching mechanism of MS2 is rare in virology. Capsomers are more commonly seen to extend N- or C-terminal arms into the capsid interior, which result in electrostatic repulsion when these units attempt to assemble to form the capsid shell. In these cases, PS interactions with CP act by overcoming this barrier. An example of this is STNV, in which interactions with PSs initiate a conformational switch, from assembly-incompetent monomers to assembly-competent trimers. This occurs by PS-mediated neutralization of positive charge in the CP N-terminal arms, allowing CP-CP interactions to overcome electrostatic repulsion that restricts their higher-order structure formation[51, 52]. Crystal structures of STNV virus-like particles (VLPs) reveal a close association of multiple RNA SLs with the CP, derived from packaged CP-encoding mRNAs[35]. An interdisciplinary

approach, based on a bespoke bioinformatics analysis of RNA-based systematic evolution of ligands by exponential enrichment (SELEX) data identified multiple dispersed PSs in this virus[53]. These stem-loops with a consensus loop motif (AXXA, in which X is any nucleotide), are repeated 30 times throughout the genome, and were shown to bind to STNV CP. When CP was added to a short genomic fragment encompassing 5 putative PSs (the 5' 127 nucleotides of the STNV genome), trimeric CP formation and subsequent particle assembly was promoted[42, 52]. Resolution of these aptamer-assembled particles by X-ray crystallography revealed that the disordered N-terminal basic arm of the CP had become more ordered, and was bound to the stem-loop structures within the RNA[51]. This suggests a mechanism in which STNV PSs are able to trigger a conformational change in the CP, by promoting order within the N-terminal arm, thus enabling assembly-competent trimers to form. The PS-mediated assembly mechanism is well understood in this system. Using insights into the PS code, it was possible to demonstrate that the PSs can be optimised to create substrates with better assembly properties that outcompete wild type virus[54].

1.2.4.2 Hepatitis B virus

A similar phenomenon occurs in the C-termini of the Hepatitis B virus (HBV) CP, demonstrating that PS-mediated encapsidation is not a feature solely attributed to the group IV +ssRNA viruses[55]. Recent work on hepatitis B virus (HBV), a virus which encapsidates a +ssRNA copy of its genome as well as its polymerase for subsequent dsDNA synthesis within the capsid prior to infection[56], was found to contain PSs when SELEX and bioinformatics approaches were applied[55]. The significance of the identified consensus PS was examined using single-molecule fluorescence correlation spectroscopy, showing that dye-labelled putative PSs were able to promote HBV VLP formation, whereas the capsid alone self-assembled poorly under the same conditions[55]. Similar to STNV, the HBV capsid protein also contains a basic-residue rich arm that extends into the cavity of the virion, suggesting that neutralisation of this positively-charged region by PSs is required to overcome the repulsive electrostatic interactions between CPs. This mechanism is supported by data showing that removal of the basic arm domain allows VLPs to readily form in the absence of genomic RNA[57].

As the +ssRNA genome of HBV needs to be transcribed into a dsDNA form within the capsid shell, having many strong CP-interacting secondary structures present could prohibit efficient polymerase activity. The results of the SELEX and bioinformatics analysis are in line with this, showing that significantly fewer PS-like sites are present in HBV compared to the

other viruses studied[55]. One hypothesised benefit of HBV using a PS-mediated encapsidation is that it would promote an ordered assembly pathway, preventing the genome from potentially becoming entangled in a way that could inhibit conversion of the genome to dsDNA by the co-packaged polymerase, i.e. via early termination or blockage of polymerase translocation.

1.2.4.3 Hepatitis C virus

PSs were also identified in hepatitis C virus (HCV), another enveloped virus[58]. Identification of putative PSs in the genome of HCV was done using a SELEX and bespoke bioinformatics approach to identify and align aptamers to the HCV genome[58]. This approach revealed a conserved loop motif that was present throughout the genome at multiple locations. Mutational validation of these putative PSs showed that abolishing any single PS did not have significant effect on viral replication and particle formation, however mutation of multiple putative PSs resulted in decreased infectious titres of HCV[58], indicating that PS-CP interactions in HCV act cooperatively to facilitate assembly of the capsid around the genome.

1.2.4.4 Picornaviruses

Further examples of PS-mediated assembly are found in the picornaviruses. Cryo-EM reconstructions of the human parechovirus 1 (HPeV1) virion revealed significant electron density around the internal 5-fold vertices of the capsid, similar to that seen with STNV and MS2 previously[35, 36], suggesting evidence for PS-mediated encapsidation[37]. As predicted, a consensus stem-loop motif was identified by SELEX and bioinformatics analysis[59], and the PS-CP interaction was characterised to atomic detail[37]. Mutational studies validated the observed RNA-CP interactions, finding that mutation of CP residues directly in contact with the PSs resulted in million-fold drops of viral titre. Similarly, when PS motifs were mutated, a decrease from zero to a million-fold in viral titre was observed, confirming that these regions in the CP and HPeV1 genome were critical for particle assembly[59].

For HPeV1 CP, similarly to STNV, RNA binding promotes oligomerisation of CP into pentameric capsomers[59]. These PS-CP interactions are also thought to subsequently stabilise the virion, with RNA bridging across the pentamer-pentamer contacts. Interestingly, structural data show that similar RNA-CP interactions are present across the *Parechoviridae*[60, 61]. With the SELEX aptamers identified for HPeV1 showing widespread matches to other *Parechovirus* genomes, it suggests that a common mechanism for

encapsidation is likely conserved across the viral family, akin to that seen in the *Leviviridae*[59, 62].

Collectively, these data suggest that encapsidation of viral RNA by means of PS-mediated assembly may be a broadly utilised mechanism in viruses, with multiple SL structures distributed across the genomic RNA directing the recruitment and addition of CP conformers at defined positions to form the capsid[43, 52, 58, 59, 62-65]. In each virus studied, the encoded PSs are seen to have heterogeneous affinities for the CP and vary around a consensus structure and sequence that is conserved across strain variants. In all cases the additive effect of multiple PSs mediates efficient cooperative assembly of the particle[19, 52, 62, 66].

1.2.5 Presentation of PSs within viral RNA

Significant structural differences are seen between viral RNA when in virions or CP-free environments[67-69], suggesting that only a subset of the PS SLs may actually exist stably in solution. It is likely that alternate RNA structures with a lower free energy overshadow the SLs of PSs, as supported by folding of 22/30 of the putative PSs in STNV[53]. One solution to this is kinetically-driven folding of the RNA, where CP binding to the few available PSs induces the refolding of neighbouring structures that mask additional PSs, by offsetting the unfavourable energies in a sequential manner[69]. However, unless a high affinity PS site is stable in solution to initiate nucleation, such as the MS2 TR, or a cluster of cooperatively acting PS are present, then alternate mechanisms in which to fold and present PSs are required.

Some viral families, including the *Picornaviridae* and *Bromoviridae*, are thought to package their genomes as they are transcribed from the polymerase in replication complexes. This co-transcriptional packaging is supported by observations of direct interaction between viral proteins involved in genome replication with CP and RNA, which is hypothesised to collectively facilitate genome encapsidation in a structure which has been termed an assemblysome[59, 70-72]. Studies on HPeV1 infer that PS-mediated encapsidation of the viral genome may also be facilitated in a co-transcriptional manner, as addition of genome-length viral RNA to CP does not promote particle formation[37]. This suggests that particular local secondary structures (i.e. PSs) may only be presented in nascent genomes and captured by CP as they are synthesised by the polymerase. In summary, data indicate that PSs with weak short-range local interactions in their structures may only be induced and captured during genome synthesis and/or the assembly process. In these cases, PS-

mediated assembly does not follow the two-stage mechanism of MS2 and STNV, but PSs still play crucial roles during assembly.

1.2.6 Methods to identify genome encoded packaging signals

The studies discussed above highlight the difficulty in identifying encoded PSs by secondary structure analysis alone[69]. This is exacerbated by the sparse nature of the sequence motif, and the variation around a consensus motif sequence that is required for variation of the PS affinities for CP.

Several techniques have been developed in order to investigate and identify the presence of capsid-interacting RNA sequences and structures encoded in viral genomes. One of the most successful being RNA-based SELEX, coupled with bioinformatics analysis. RNA SELEX has been previously used as a tool to isolate high affinity RNA ligands to numerous target proteins[73-76]. The first instance of using CP as a selection target was for bacteriophage MS2, resulting in the identification of aptamers with a consensus sequence matching that of the known CP binding site within the genome, the TR[77]. Since then, the technique has been adapted and applied to multiple other viral capsids, including those with no prior RNA-CP binding data available. This has led to the identification of putative PSs encoded within the genomes of MS2[65], GA[62], STNV[53], HCV[58], HBV[55] and HPeV1[59], using cognate CP or capsomers as selection targets.

The RNA SELEX approach works by exposing the inside of immobilised CP or capsomers to a random RNA aptamer library. Aptamers that bind the viral capsid are enriched through repeated selection cycles to identify aptamers with affinity to the capsid protein. Fully assembled capsids are typically also present in solution to sequester any aptamers with affinity to the exterior capsid surface. The resulting aptamer population is then subjected to next-generation sequencing (NGS), and bioinformatics analysis is used to identify putative RNA sequences and structures within the ensemble of SELEX hits. These are then mapped back to the viral genome to identify putative PS targets for validation[53].

As SELEX aims to identify aptamers with affinity to the CP interior, the resulting positive hits are expected to share some degree of similarity to putative PSs. PSs are seen to be degenerate in both structure and sequence, varying around a consensus SL motif, and the same holds for the aptamers. Therefore, bioinformatics analysis is required in order to identify common sequence motifs and secondary structures shared by both the RNA aptamers and the PSs encoded in the viral genome. Identified aptamers are aligned against

the viral genome using a 1 nucleotide (nt) sliding window, and non-contiguous matches with a probability equivalent to a contiguous match of a given length (the Bernoulli score) are then calculated for each alignment[78], measuring the goodness of fit[79]. Alignments with a sufficiently high Bernoulli score, for example 12, equating to a 1:10,000 probability, are selected and the fold of the genomic sequence at these positions is predicted to reveal whether a common stem-loop motif is present at these locations across the genome.

A NGS approach was also recently developed as an alternate technique to SELEX for the identification of multiple conserved putative PSs across the genome of foot-and-mouth disease virus (FMDV)[80]. By analysing deep sequencing data derived from both packaged and unpackaged viral RNA populations, clusters of constrained variation could be identified in the packaged population, suggesting that they may have a functional importance for genome encapsidation. Analysis of these regions indicated that they formed RNA stem-loops sharing a common motif. This was confirmed by synonymous mutations disrupting the structures of the identified SLs, demonstrating that the NGS approach had indeed identified putative PSs[80]. Using NGS to identify putative PSs is possible due to the quasispecies nature of RNA viruses. The lack of a proof-reading function in the virally-encoded polymerase allows frequent introduction of errors, leading to the production of a highly varied population of related genomes[81]. Based on the PS-mediated assembly hypothesis, mutations affecting PS motifs or structures within the genome will in turn produce RNAs that are either less efficient at packaging or packaging incompetent. Therefore, these mutants will be selected against during virion formation and be less frequent in the packaged genome population when compared to the unpackaged population. This NGS approach provides a simple methodology for the identification of putative PSs and can easily be transferred to other RNA viruses that are able to be grown and purified *in vitro*. It is an alternative to defining CP selection targets via VLP or capsomer systems, as is required for SELEX.

1.3 Genome packaging in an avian virus

To date, studies on PS-mediated assembly have only been performed in viruses infecting bacteria, plants and mammals. With the exception of FMDV, there is no investigation into genome-encoded PSs within veterinary viruses, specifically those that infect the avian species. Therefore, the avian astroviruses were selected as a target for study, because they present similarities to features commonly exhibited in viruses utilising a PS-mediated encapsidation mechanism, as discussed further below.

1.4 The *Astroviridae*

The family *Astroviridae* is composed of 2 genera, *Mamastrovirus* and *Avastrovirus*, which infect mammalian and avian species, respectively. There are currently 19 genotypes of *Mamastrovirus* and 3 of *Avastrovirus* officially recognised by the International Committee on Taxonomy of Viruses (ICTV 2018). These are classified as *Mamastrovirus* 1-19 and *Avastrovirus* 1-3, respectively[82]. Astrovirus (AstV) strains are currently classified based on the host species they were isolated from[83]. However, there is growing evidence that some astroviruses might not be strictly species specific and may emerge from zoonotic transmissions; a recent study detected both avian and mammalian astroviruses (AAstV and MAstV, respectively) in non-human primates from a range of locations[84]. Similarly, reports of a HAstV-like virus in a captive chimpanzee[85], amphibian-related astrovirus in faecal matter of rhesus monkeys[86], and astroviruses distantly related to mink and ovine astroviruses in human stool[87] are present in the literature. There is also evidence of cross-species transmission into avian species with a report identifying an astrovirus genetically similar to the Mamastroviruses in faecal samples of a clinically healthy avian host[88], and evidence to support the transmission of Avastroviruses into mammals, including humans[89, 90]. Collectively, these reports suggest the *Astroviridae* may have a broad host range potential, albeit producing asymptomatic infection in most cases, or that prolonged close-contact can produce recombinant astrovirus strains capable of productive cross-species transmission[82, 91-94].

The vast diversity of the *Astroviridae* is driven by mutation and recombination. Non-synonymous mutations commonly occur in the ORF2 gene[95], and recombination events between HAstV strains are most frequently seen at the ORF1b-ORF2 junction[96-101], with 6.6% of HAstV strains predicted to be recombinant[102]. Evidence of intertypic recombination between several mammalian and avian astrovirus strains has also been observed[96, 103-111]. Therefore, the high recombination rates, as well as the zoonotic potential of astroviruses, suggest that classification based solely on host species may not be an accurate method.

1.4.1 Astrovirus disease

Astroviruses are primarily transmitted via the faecal-oral route[112, 113], infecting the small intestine and causing enteric disease worldwide. AstVs were first identified in humans[114, 115] and shortly thereafter in a range of other animals, including chickens[116], pigs[117], dogs[118], mice[119], turkeys[120], cats[121], ducks[122], sheep[123], cows[124], mink[125], and more recently in bats[126], sea lions[110], pigeons[94] and geese[127]. However, this is not a complete list of host species and

more AstVs infecting other animals are likely to be discovered, especially with the increasing sampling of the viromes of different animal species using NGS techniques[128].

The Mamastroviruses predominantly cause gastroenteritis[129], with a typical human astrovirus (HAstV) infection producing mild watery diarrhoea lasting 1-4 days. The virus can also be associated with vomiting, fever and abdominal pain in some cases[130]. HAstV is the second leading causative agent of paediatric acute gastroenteritis after rotaviruses[131, 132], with a previous report estimating that there are almost 4 million cases of HAstV-associated diarrhoea each year in the USA alone[133]. HAstV infection typically clears without the need of medical intervention, however, symptoms can be prolonged and more severe in the immunocompromised and elderly with systemic spread and the development of neurological syndromes such as meningitis and encephalitis[134, 135]. Similar symptoms have also been observed in other mammalian species infected with AstVs recently[136-144]. HAstV infection appears to be widespread, with ~90% of children seropositive for antibodies to at least one of the 16 known HAstV genotypes by the age of 5[145-148], with seroprevalence observed to decrease in healthy adults, suggesting a loss of immunity over time[148]. Immunity appears to be genotype specific, given that susceptibility to sequential HAstV infections have been detected[149, 150]. It is thought that many AstV infections are asymptomatic, potentially due to pre-existing antibodies, and therefore remain largely unreported[87, 151-153].

The Avastroviruses, like their mammalian counterparts, cause diarrhoea and can cause extra-intestinal disease, producing a diverse range of pathologies in avian species. Young birds are most susceptible to AAstVs, with infection typically occurring in the first few weeks of life via the faecal-oral route[154-157], and there is increasing evidence to suggest some AAstVs are transmitted vertically from hen to chick through eggs also[158, 159]. Clinical disease varies by strain and serotype, causing pathologies such as enteritis[160], hepatitis[161] and nephritis[155, 162, 163]. Chicken astrovirus (CAstV) and avian nephritis virus (ANV) have been linked to poor hatchability and growth suppression syndromes in broilers such as runting-stunting syndrome (RSS), gout, 'white chicks' condition, chick nephropathy and interstitial nephritis[155, 159, 162-165]. Some strains also cause mortality[163], and the emergence of highly pathogenic strains which cause outbreaks with up to 40% mortality have recently been reported[159, 166, 167]. Similarly to CAstV and ANV, Turkey astrovirus (TAstV) is linked to growth depression, and also to poult enteritis complex or syndrome (PEC and PES) and poult enteritis mortality syndrome (PEMS) in young turkeys[168, 169]. Whilst Duck astrovirus (DAstV)

can cause fatal hepatitis in ducklings resulting in up to 50% mortality[122]. Infection with Avastroviruses may also lead to immunosuppression, leaving poultry more susceptible to opportunistic secondary infections[170-173]. These examples underscore the need to control these pathogens.

Astroviruses are widespread and found globally[91, 174-176] and the prevalence of AAstVs in poultry flocks can be very high or ubiquitous for some strains[177-186], making the control and prevention of *Avastrovirus* infection an important area of focus. However, the exact economic impact of astroviral disease on the poultry industry remains unquantified[115].

There are currently limited control measures for astrovirus diseases, vaccines and antibodies developed against astroviruses only provide partial protection[187, 188], chemotherapeutics are inefficacious, and elimination of environmental contamination is difficult to achieve. Thorough sanitation and several weeks resting between flocks are the only effective control methods available at this time[189, 190].

1.4.2 The astrovirus virion

Astroviruses are small, 28-41 nm non-enveloped viruses, which encapsidate their +ssRNA genome within a T=3 icosahedral capsid. The term astrovirus was first used due to the star-like appearance (*astron* meaning star in Greek) of newly discovered small round viruses in the faeces of hospitalised patients suffering with gastroenteritis[191]. However, only about 10% of the population of AstV virions actually exhibit the characteristic 5 or 6 pointed star-like morphology when observed by electron microscopy, and this morphology may represent an intermediate step of viral dissasembly[168, 192, 193]. Interestingly, not all astrovirus strains display the star-like morphology, with the AAstVs ANV, CAstV and DAstV only seen as small spherical viruses, lacking any prominent surface structure.

The AstV capsid is composed of 180 copies of the CP[194]. Cryo-EM reconstructions of the HAstV virion show a continuous inner layer (core), surrounded by an outer layer of globular densities (spike), that are disconnected or weakly associated with the capsid. In its immature form there are 90 globular spike dimers on the surface of the capsid which are reduced to 30 upon maturation through proteolysis[194] (Figure 1.3). In the mature capsid form, there are triangular plateaus along the edges of the two-fold icosahedral axis and depressions at the three- and five-fold axes of the capsid, and the globular spike domains appear to have weak linkages to the core, suggesting that they may be flexible or loosely associated. The AstV capsid is highly resilient, with virions maintaining infectivity in as low

pH as 3, and after treatment with various detergents, chloroform, ethyl ether, trypsin and most alcohols[195-197]. Additionally, particles are stable at 60 °C for 5 minutes, at room temperature for 48 hours, and at 5 °C for 1 week[195]. This inherent resistance promotes environmental survival of AstV particles and makes effective removal of contamination challenging.

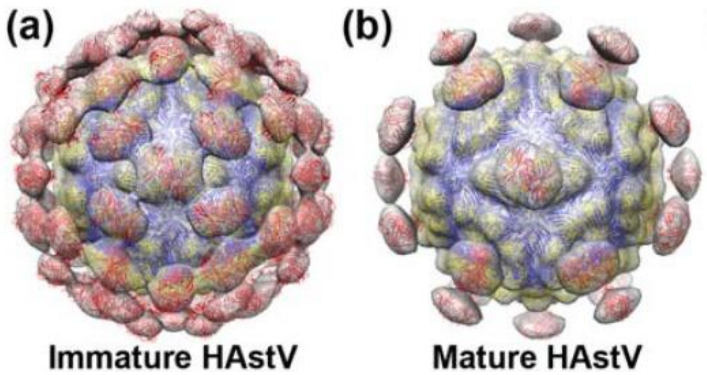


Figure 1.3: Cryo-EM reconstructions of the human astrovirus virion. (**A**) The immature HAstV virion, containing 90 globular spikes. (**B**) The mature HAstV virion after trypsin proteolysis, with 30 globular spikes. Figure adapted from[198].

1.4.3 Astrovirus genome organisation, ORFs & proteins

The AstV genome consists of a +ssRNA that varies in size between 6.2-7.7 kbp with a 5' viral genome-linked protein (VPg) and 3' poly-A tail[199, 200]. The 3' end of almost all astroviral genomes contain a conserved secondary structure motif, the stem-loop II motif (s2m), which is seen in other +ssRNA viral families, the *Caliciviridae*, *Picornaviridae* and *Coronaviridae*[201], and is hypothesised to interact with viral and host proteins required for RNA replication. However, this structure is absent in TAstV-2, bat AstVs, rat AstVs and HAstV-MLB1 astroviuses[202-204]. Terminal untranslated regions (UTRs) flank three overlapping open reading frames (ORFs), ORF1a, ORF1b and ORF2 (Figure 1.4). The 5' UTR of AstVs is typically 11-85 nt long, whilst the 3' UTR varies in length between mammalian and avian astroviruses, at either 80-85 or 130-305 nt, respectively[192, 205]. Computational analysis of the HAstV UTRs suggests the presence of putative host protein binding sites which may promote a genomic RNA secondary structure conformation suitable for replication, or aid in recruiting proteins involved in the viral replication complex[206].

ORF1a and 1b are located at the 5' end of the genome and encode the non-structural proteins (nsP) which are translated from the genomic RNA as two polyproteins, nsP1a and nsP1ab. nsP1a is thought to be cleaved into a variety of peptides, evidenced by

experimental data and the presence of conserved predicted motifs; including a serine protease with trypsin-like activity, a bipartite nuclear localisation signal (NLS), 5-6 transmembrane helices, an immunoreactive element, a putative VPg, and a putative death domain[161, 199, 202, 207-212]. Downstream of the putative VPg motif is a hypervariable region, with insertions and deletions in this area associated with cell culture adaption[213] and efficient RNA synthesis[214]. The putative NLS is only present is some astroviral strains[212] and localises to the nucleus in infected cells[215]. However, the NLS is located within the putative VPg[210] and so its function, if any, remains uncertain. Bioinformatics analysis identified a putative death domain towards the C-terminal end of nsP1a with structural similarities to proteins which are known to induce caspase activity, and expression of ORF1a alone in CaCo-2 cells could promote apoptosis[209], suggesting this may be a functional domain.

ORF1b which encodes nsP1b, the viral RNA-dependant RNA-polymerase (RdRp)[199, 216], overlaps with ORF1a by 10-148 nts and by 12-45 nts in MAstVs and AAstVs respectively[153]. ORF1b is accessed from ORF1a by a retrovirus-like (-1) ribosomal frameshift (RFS) which is triggered by a highly conserved RNA SL motif following a slippery heptameric sequence (AAAAAAC)[199, 200]. However, an alternate translation strategy may be used by AAstVs, as some strains have predicted stop codons in ORF1a upstream of the SL structure, suggesting another frameshift structure may be formed or a different mechanism for ORF1b translation is used[217, 218].

ORF2 is located at the 3' end of the genome and encodes the precursor capsid protein which is translated from poly-A tailed subgenomic mRNA of 2-2.5 kb in size with short 5' and 3' UTRs[219]**.** The ORF2 mRNA is transcribed from an internal subgenomic promoter within the antigenome, which is suggested to be located around the ORF2 start codon[153, 212]. Sequence similar to the conserved subgenomic promotor is also present at the 5' end of the genome, suggesting it has a role in the transcription of genomic RNA also[212].

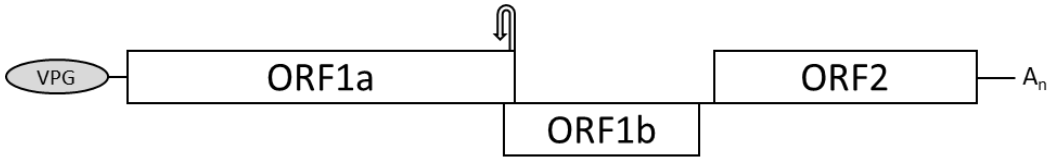


Figure 1.4: Typical genome organisation of the Avastroviruses. Schematic diagram of the organisation of the three open reading frames within the AAstV genome, the flanking UTR's, 5' VPg and 3' poly-A tail, as well as the ribosomal frameshift structure located between ORF1a and 1b.

1.4.4 Astrovirus capsid proteins

Recently, a number of structural studies focused on the *Astroviridae* have established the crystal structures of the ORF2 encoded CP and show that there are two main structural domains, the core and the spike (Figure 1.3: blue/green and red, respectively), in astrovirus CPs. Structures for the core domains of HAstV-1 and -8, and the spike domains of HAstV-1, -2, -8 and TAstV-2 have been reported[193, 194, 220-222]. The tertiary structure of the HAstV core domain is strikingly similar to that of hepatitis E virus (HEV), despite a low ~18% sequence identity between them[193, 194, 220]. This may suggest a potential evolutionary relationship between these viruses[223, 224]. In contrast, the AstV spike domain structures are dissimilar to that of HEV, and are highly variable even between AstV strains, suggesting they interact with different host proteins[194, 221].

The core domain of the CP is located in the N-terminal half of the ORF2 precursor protein (Figure 1.5), which is more conserved than the C-terminal spike domain across the *Astroviridae*[225]. A region of repeating small and basic amino acids, predominantly arginine and serine or arginine and lysine, is found within the first ~70 amino acids of the N-terminal domain of all astroviral CPs[226, 227], and is thought to be a putative RNA recognition domain by analogy to other +ssRNA viruses[53, 228-234]. The C-terminal half of the CP precursor forms the spike proteins on the particle[194]. The high sequence variability in this region is indicative of a domain that is under selective pressure from the immune system[235, 236] and correlates with the differing tropisms of the *Astroviridae*[225, 237]. Finally, a conserved acidic domain at the C-terminus of the CP contains multiple putative caspase cleavage sites, and may be a substrate for host caspase cleavage that promotes the processing of the capsid for assembly and release, as this region is not present in the extracellular virion[238, 239].

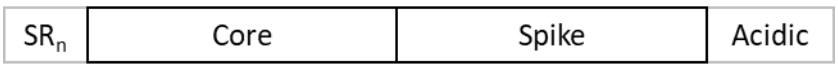


Figure 1.5: Domain architecture of the astrovirus capsid protein precursor. Schematic representation of the astrovirus precursor capsid protein, highlighting the respective positions of the basically-charged disordered N-terminus (typically consisting of repeating serine and arginine residues), the capsid core domain, the spike domain, and the acidic C-terminal domain.

Although uncommon in enteric RNA viruses[240, 241], an enterotoxin-like activity of the astrovirus capsid was identified when turkey poults orally administered with TAstV VLPs

developed VLP-induced diarrhoea akin to that seen in astrovirus infection[242]. TAstV infection triggers the rearrangement of F-actin and redistribution of sodium-hydrogen exchangers (NHE2 and NHE3) from the membrane to cytoplasm in jejunal tissues of experimentally infected turkey poults, suggesting infection provokes osmotic diarrhoea[243]. Similarly, infection of cell monolayers with HAstV VLPs or UV-inactivated virions increases epithelial barrier permeability, correlating with a decrease of actin stress fibres and disruption of occludin at cellular tight-junctions[244]. These data suggest that astrovirus-induced diarrhoea is linked to the viral capsid protein itself and independent of inflammation and cell death caused by viral replication in contrast to other gastrointestinal pathogens[245], reminiscent of cholera toxin induction of diarrhoea[246]. In addition to enterotoxin-like properties, a 30 amino acid peptide in the N-terminus of the HAstV-1 capsid core domain with sequence homology to human neutrophil defensin-1 has been shown to bind complement proteins C1q and MBL, inhibiting activation of the classical and lectin pathways of the complement system, respectively[247, 248]. This potentially explains the mild levels of inflammation and cellular damage observed in the gut during infection[173].

1.5 Astrovirus replication

There is limited information available concerning the replication of astroviruses, with most of what we understand derived from studies on HAstV genotypes 1-8 and TAstV-2. This leaves it unclear how other MAstVs and AAstVs may compare. One major pitfall in astrovirus research is the lack of robust cell culture systems and animal models in which astroviruses can be isolated and characterised, with just some strains of HAstV, ANV, DAstV, CAstV, canine AstV and murine AstV able to grow *in vitro*, albeit poorly, depending upon strain[91, 211, 237, 249-253]. For several AAstVs, including CAstV, TAstV, ANV and DAstV, successful virus isolation has been achieved in embryonated eggs[254-256]. However, cell culture systems will be vital to developing our understanding of AAstV biology.

In addition, several reverse genetics systems have been developed for the *Astroviridae*, including HAstV-1[257], ANV-1[211], porcine astrovirus[258] and chimeric HAstV-1/8 containing a combination of the HAstV-1 non-structural and HAstV-8 structural genes[259]. The advent of reverse genetics systems for the *Astroviridae* is an exciting development, which provides the key tools needed in order to study the molecular biology of these viruses.

1.5.1 Receptor binding and entry

Entry of astroviruses is poorly understood and the host receptors are currently unknown. It is estimated that HAstV has an attachment half-time of about 10 minutes, indicating high

affinity receptor binding, or an abundance of host-cell attachment factors[260]. However, different virus strains are likely to have different receptor requirements. Data from HAstVs show that some cell lines are permissive for binding and entry of certain serotypes but not others, for instance CaCo-2 cells are susceptible to infection with all 8 HAstV genotypes, whilst BHK-21 and HTC-15 cells only support HAstV-2 and HAstV-1 infection, respectively[237]. Putative conserved carbohydrate-binding pockets were identified on the spike proteins of HAstV and TAstV, and addition of non-sialic sugar molecules such as heparin, heparan sulphate and dextran sulphate was seen to reduce HAstV-8 binding *in vitro*, whilst infection with HAstV-1 was unaffected[221, 261]. Further work is needed to fully elucidate the role of carbohydrates in astroviral attachment, and their importance for each strain. Furthermore, HAstV receptor binding may trigger the disruption of cellular tight-junctions, as increased epithelial barrier permeability and subsequent diarrhoea are associated with the viral binding stage of its life-cycle[244].

HAstV particles have been observed in coated pits and vesicles during infection by electron microscopy, suggesting astrovirus may enter cells by clathrin-mediated endocytosis[262]. This is supported by data showing reduced viral entry *in vitro* when clathrin-coated pit formation and endocytosis are inhibited or disrupted, whilst virus binding to the host cell was unaffected[260, 262]. AstV entry may also be dependent upon the actin cytoskeleton and the presence of cholesterol in the cell membrane, as drugs targeting the disruption and removal of these, respectively, impair HAstV infection[260].

Data suggest that HAstVs require activation of the extracellular signal-regulated kinase (ERK) 1/2 pathway for a productive infection[263]. The pathway is activated within 15 minutes of viral binding to the cell and can be achieved even with inactivated virus or VLPs[260, 263]. Inhibition of the ERK 1/2 pathway substantially reduces astrovirus production, affecting both genomic and subgenomic mRNA accumulation as well as viral protein expression[263]. Activated ERK may phosphorylate and activate the RdRp, as seen for coxsackievirus[264], or it may contribute to formation of clathrin-coated endosomes as seen for the internalisation of lactoferrin receptors[265]. However, the exact function of activated ERK in relation to HAstV infection is currently unknown.

After binding, astrovirus entry into the cell, and genome release into the cytoplasm, occur within around 130 minutes[260]. Whilst the mechanism for uncoating is unclear, disassembly of capsid and genome release into the cytoplasm is thought to be mediated by acidification of late endosomes[262]. The importance of this mechanism in replication was shown for

HAstV-8, as inhibition of endosome maturation by Rab7 knock-down drastically reduced infection[260]. The stabilizing effect of divalent cations observed on astrovirus virions may also suggest a possible mechanism for genome release[266], as the low concentrations of cations found in the cytoplasm of intestinal cells may destabilize the capsid and promote virus uncoating.

1.5.2 Transcription and translation

Although the exact location of genome release is unknown, the viral genomic RNA serves as an mRNA template for the initial translation of the non-structural proteins including the RdRp. Like other +ssRNA viruses, such as the *Caliciviridae,* initiation of translation may be dependent on a VPg cap[267], this is supported by protease treatment of the HAstV genome prior to transfection leading to reduced replication of the virus[268]. Additionally, the region of astroviral nsP1a/4 containing the putative VPg and hypervariable region, when phosphorylated, was found to oligomerise and interact with the viral RdRp[269]. nsP1a/4 was also located near the endoplasmic reticulum (ER), colocalised with viral RNA and host protein CD63[270], supporting the hypothesis that this is indeed the putative astroviral VPg[208].

The first proteins translated upon genome uncoating are nsP1a and nsP1ab from ORF1a and ORF1a/b, respectively. The ribosomal frameshift mechanism between the two overlapping reading frames allows translation of the nsP1ab polyprotein 25-28% of the time[271] and processing releases nsP1b, the viral RdRp[272]. The nps1a polyprotein is self-cleaved by the virally-encoded serine protease and by a cellular protease into at least 5 non-structural peptides[231, 273-275]. Together, the products of ORF1a and 1b work to facilitate modulation of the cell and viral genome replication.

Little mechanistic detail about astrovirus replication is available. However, it has been shown that replication complexes assemble in subcellular membrane compartments, thought to be ER-derived[208, 276]. These membrane structures may not be exclusively derived from the ER as there is evidence to suggest that membranes from other cellular organelles may be recruited into the replication complexes[277]. Host genes linked to cholesterol and fatty acid synthesis, as well as phosphatidylinositol and inositol phosphate metabolism were found to be important in HAstV replication, as silencing of these in Caco-2 cells significantly reduced the transcription, translation and production of virions[277]. The manipulation of these genes and their products may be linked to the formation and regulation of replication complexes, as seen for multiple other viruses[278-282]. However, their exact mechanisms of function in astroviral infection remain unknown.

Transcription of the antigenome within replication complexes is detected around 6-12 hours post-infection (hpi)[276, 283]. This is then used as a template for positive-sense subgenomic and genomic RNA synthesis, occurring immediately after anti-genome production, with transcription of subgenomic RNA estimated at levels 10-fold greater than the genomic RNA, and the anti-genome representing just 0.7-4% of the total positive-sense RNA produced[284].

The host gene encoding an RNA helicase (*DDX23*) was found to be important in HAstV replication, as knock-down of *DDX23* reduced viral yield by ~40%[277]. +ssRNA viruses similar in size to astroviruses typically encode a helicase domain[285]. Unusually, however, data suggest that HAstV does not encode its own RNA helicase and instead recruits it from the host, as previously suspected from comparative sequence analysis of astroviral genomes[210, 286]. The UTRs of HAstVs may also play a role in viral replication via host factor recruitment, such as for replication complex formation, as seen for various other RNA viruses[287]. It was observed in CaCo-2 cells that PTB/hnRNP1 binds to a conserved helical structure in the 3' UTR of HAstV and plays an important role in the replication of the virus, as siRNA knock-down of PTB reduced the production of negative-sense HAstV RNA by ~90%[288]. *In silico* analysis also identified additional putative recognition sites for SRSF and hnRNPE2 proteins in the UTRs of the HAstVs[206]. Binding of these may act to function in a similar manner to that observed for poliovirus, whereby they aid in promoting translation initiation[289].

1.5.3 Assembly and egress

During HAstV infection, expression of ORF2 CP and virus assembly occur alongside genome replication within subcellular double-membrane vesicle compartments[123, 208, 290]. CP is translated around 8-12 hpi, coinciding with a large increase of sub-genomic RNA around this time[239, 283]. The location of assembly may vary between strains, as clusters of HAstV particles are observed near the nucleus[208, 291], whereas crystalline arrays of virus are instead seen in lysosomes, autophagic vesicles and the cytoplasm during ovine AstV infection[123].

Translation of ORF2 in HAstV-8 generates an ~90 kDa capsid precursor (VP90) which is cleaved intracellularly at multiple putative caspase recognition motifs in the acidic C-terminal region by host caspases-3 and -9 to produce a ~70 kDa capsid protein (VP70)[238, 239]. 180 copies of capsid protein assemble to form the immature virion, and both virions comprised of VP70 or VP90 have been observed intracellularly[231, 238, 283, 290, 292]. Dimers and possible trimers of HAstV-8 VP90 have also been observed in infected CaCo-2 cells[290],

suggesting that the capsid protein readily forms intermediate structures which may then facilitate formation of the particle, akin to HEV[293, 294].

VP90 associates with membrane structures most likely by its highly acidic C-terminus, as VP70, which lacks this domain, is not membrane associated[290]. The C-terminal region is also protected from trypsin cleavage when capsid protein is membrane bound, further supporting this[290]. This suggests that caspase cleavage releases membrane-bound capsid protein, and may promote virion assembly and release from the cell, as observations suggest that extracellular astrovirus capsids are formed exclusively from VP70, and not the VP90 precursor[231, 283, 292]. The function, if any, of the ~8 kDa C-terminal region released upon VP90 caspase cleavage is unknown. However, it is highly conserved between HAstV strains suggesting it may have additional roles in infection[198].

Recent work identified the activation of the host cell eIF2α-ATF4-CHOP pathway during HAstV-1 infection[276]. Sensing of AstV dsRNA by PKR was seen to promote phosphorylation of eIF2α, which induces translational block, in the late stages of HAstV infection 12-16 hpi. However, dsRNA was detected as early as 6 hpi, suggesting PKR was suppressed by HAstV via an unknown mechanism. Phosphorylated eIF2α subsequently activated the eIF2α–ATF4 pathway. The study found that CHOP, a target of ATF4, was then upregulated in HAstV-1 infected cells, activation of which induces caspase activity, and was linked to the release of HAstV-1 progeny into the extracellular media[276]. These data suggest that astroviruses are able to specifically modulate the innate immune response to viral infection, by suppressing and then activating the eIF2α-ATF4-CHOP pathway in order to benefit replication and egress, respectively.

The non-lytic release of HAstVs from infected cells occurs around 12 hpi and is not accompanied with significant cell death[239]. It is thought egress may be mediated by membrane destabilization or export through membrane vesicles[198, 272]. Consistent with this is recent data suggesting a viroporin-like activity of a newly identified protein (XP), which is encoded in ORFX, a hidden gene within ORF2[295]. Egress of HAstV particles appears to require the activity of caspases-3 and -7[276], and although the processing of the capsid is also caspase-dependent, the cleavage and release events appear to be independent of one another[209, 238, 239].

Assembled CP within extracellular virions is further cleaved into 3 smaller units by trypsin-like proteases[291], causing structural changes, to form the mature virion and increase

infectivity by 10^5-fold *in vitro*[292]. For HAstV-8, trypsin cleaves VP70 into 3 smaller molecules of 34 and 27/25 kDa (Figure 1.6), representing the capsid core and spike domains, respectively[291]. This proteolytic processing releases 60 of the 90 globular spike dimers present on the surface of the particle[193] to give rise to the mature virion morphology. Unusually among astroviruses, ANV appears able to form mature infectious virions in the absence of trypsin *in vitro*, suggesting capsid maturation by proteolysis is either not required, or that a different maturation mechanism is used[200].

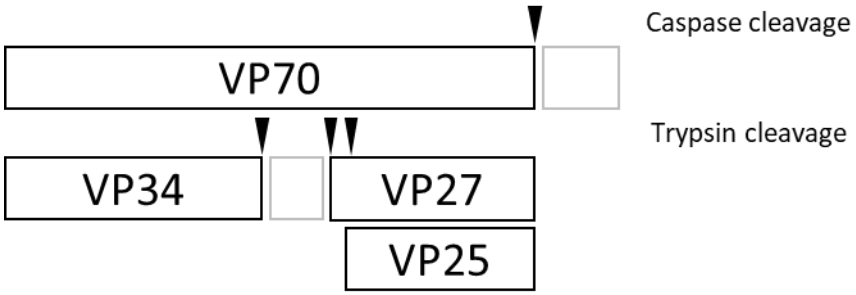


Figure 1.6: Proteolytic processing of the HAstV capsid precursor. Schematic diagram of the proteolytic processing pathway for the HAstV-8 capsid. The full-length precursor (VP90) is cleaved at multiple sites in its C-terminus by caspases to remove the acidic region, forming VP70 which is then able to form immature virus particles capable of exiting the cell. These immature particles are processed by an extracellular trypsin-like protease to produce the capsid core (VP34) and two spike variations (VP25 and VP27), which comprise the mature, highly infectious HAstV-8 particles[291, 296].

1.6 Genome packaging mechanism of astroviruses

Viral packaging enzymes and nucleoproteins seem to be absent in astroviruses, suggesting encapsidation of genomic RNA could be PS-mediated[64]. This hypothesis may be further supported by the presence of highly conserved stem-loop structures (RFS and s2m) at the start of the replicase gene (ORF 1b)[199, 200] and in the 3' UTR[201], respectively, as well as a highly basic N-terminal arm within the CP across the *Astroviridae*, which are features seen in other +ssRNA viruses that exploit a PS-mediated assembly mechanism[39, 65, 68]. The importance of the basic N-terminal region in the CP for genome packaging was displayed in a study investigating the capsid of HAstV, where removal of residues 11-30 reduced viral titre by half, and removal of residues 31-50 caused a 1000-fold reduction in infectious virus when compared to the wild type produced in cells[231].

The work presented in this thesis investigates the genome encapsidation mechanism of the *Astroviridae*. Given the similarities with +ssRNA viruses that are known to use a PS-mediated encapsidation mechanism, such as conserved RNA SLs in the genome and a highly basic N-terminal arm of the CP, the hypothesis that the *Astroviridae* also use this mechanism will be explored, using the model avian astrovirus ANV G-4260. To achieve this, a protein expression system to produce VLPs for RNA SELEX and a reverse genetics system in which to undertake mutation studies of putative PSs were constructed.

In Chapter 2, details of the experimental procedures used throughout this thesis are presented. In Chapter 3, the structure of the ANV CP was investigated using homology modelling to reveal similarities to the other known AstV structural proteins. In addition, a domain on the interior of the CP was predicted to interact with SLs of RNA based on binding-specific substructure homology.

Having computationally defined the structural domains within the CP of ANV, Chapter 4 investigates the expression of AAstV CP in a range of systems, with attempts to assemble VLPs guided by the initial modelling. Capsomer-like structures could be formed by ANV CP, however, ANV VLPs were not produced.

Chapter 5 shows the construction of a reverse genetics system for ANV, from which infectious virus was recovered. ANV was then characterised and methods to purify viral particles were investigated and optimised.

Finally, as VLPs or ANV virions suitable for RNA SELEX could not be produced in Chapters 4 and 5, respectively, an alternate approach to identifying PSs was investigated. In Chapter 6, a NGS approach was used to identify nts that were enriched in the genomes of ANV virions when compared to intracellular viral RNA. The structures of putative PSs were identified in these conserved regions, and the most conserved PS, PS1, was then disrupted using the reverse genetics system, to recover PS1 mutant ANV. The role of PS1 in genome encapsidation was then validated using viral growth assays, demonstrating that mutation of PS1 resulted in a delayed accumulation of extracellular virus.

Chapter 2

2 Materials and Methods

2.1 Materials

2.1.1 Virus isolates

Isolates of the avian astroviruses, ANV strain G-4260, CAstV 612 and CAstV 11672 were a kind gift of Victoria Smyth, Agri-Food and Biosciences Institute (Northern Ireland).

2.1.2 Cell lines

BSRT7: Cell line derived from BHK21 cells which constitutively express bacteriophage T7 polymerase, a kind gift of A. Easton (University of Warwick)[297].

LMH: Immortal cell line derived from diethylnitrosamine-induced primary hepatocellular carcinoma of a male leghorn chicken (ATCC® CRL-2117™)[298].

DF-1: A spontaneously transformed chicken fibroblast cell line from East Lansing line chickens (ATCC® CRL-12203™)[299].

Chicken Kidney cells (CK): Primary cells derived from the kidneys of 2-3 week-old SPF Rhode Island Red chickens[300].

Chick Embryo Fibroblasts (CEF): Primary cells derived from 9 day-old chicken embryos from SPF eggs (VALO Line, BioMedia).

2.1.3 Media and buffers

BES medium: 1x EMEM (Sigma), 10% Tryptose Phosphate Broth (Gibco), 0.2% Bovine Serum Albumin (Sigma), 20 mM N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES) (Sigma), 0.4% Sodium Bicarbonate (Sigma), 2 mM L-Glutamine (Sigma), 100 U/ml Penicillin/Streptomycin (Gibco) and 25 U/ml Nystatin (Sigma).

Blocking buffer: 5% skimmed milk (Sigma) in PBS.

Boiling mix: TruPAGE™ LDS Sample Buffer (Sigma) and TruPAGE™ DTT Sample Reducer (Sigma).

CEF growth medium: 1x E199 (Sigma), 10% Tryptose Phosphate Broth (Gibco), 5% FBS (Sigma), 0.2% Sodium Bicarbonate (Sigma), 100 U/ml Penicillin/Streptomycin (Gibco) and 1 µg/ml Amphotericin B (Gibco).

CEF maintenance medium: 1x E199 (Sigma), 10% Tryptose Phosphate Broth (Gibco), 2% FBS (Sigma), 0.2% Sodium Bicarbonate (Sigma), 100 U/ml Penicillin/Streptomycin (Gibco) and 1 µg/ml Amphotericin B (Gibco).

CK cell growth medium: 0.8x EMEM (Sigma), 8.85 mM N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) (Sigma), 8.8% Tryptose Phosphate Broth (Gibco), 8.8% Newborn Bovine Serum (NBS) (Sigma), 1.6 mM L-glutamine (Sigma), 8.8 U/ml Penicillin/Streptomycin (Gibco) and 40 U/ml Nystatin (Sigma).

Complete DMEM: 1x DMEM (Sigma), 10% FBS (Sigma), 100 U/ml Penicillin/Streptomycin (Gibco) and 1 µg/ml Amphotericin B (Gibco).

Complete EMEM: 1x EMEM (Sigma), 10% FBS (Sigma), 1 mM Non-essential Amino Acids (Gibco) and 1 mM Sodium Pyruvate (ThermoFisher).

Extraction buffer: 1 mM $MgSO_4$ and 1 mM $NaPO_4$, pH 7.4.

LB Broth (Invitrogen).

MMA: 10 mM MES (2-[N-morpholino]ethanesulfonic acid) pH 5.6, 10 mM $MgCl_2$ and 100 µM Acetosyringone.

PBSa: 0.2 g/L KCl, 0.2 g/L KH_2PO_4, 8 g/L NaCl and 1.15 g/L Na_2HPO_4 (Sigma, 56064C), pH: 7.2-7.5.

1x TBE: made from 10x UltraPure™ TBE Buffer (Invitrogen), diluted in Super-Q water (SQW).

TBS-T: 50 mM Tris-HCl (pH 7.4) (Sigma), 150 mM NaCl (Sigma), 0.1% Tween 20 (Sigma).

TN buffer: 50 mM Tris-HCl (pH 7.4) (Sigma), 100 mM NaCl (Sigma).

TNS buffer: 50 mM Tris-HCl (pH 7.4) (Sigma), 300 mM Na_2SO_4 (Sigma).

TruPAGE™ TEA-Tricine SDS running buffer (Sigma).

2.1.4 Chemicals, reagents and kits

2.1.4.1 Chemicals and reagents

Ambion™ Recombinant RNase A (Invitrogen).

Ammonium sulphate (Sigma).

Ampicillin 100 µg/ml (Sigma, A5354).

Attachment Factor Protein (1X) (Gibco).

Chloramphenicol 25 µg/ml (Sigma, R4408).

EDTA (Invitrogen).

GelRed® Nucleic Acid Gel Stain (Biotium).

Geneticin™ Selective Antibiotic (G418 Sulfate) (Gibco).

IGEPAL® CA-630 (Sigma).

Isopropylthio-β-galactoside (IPTG) (Invitrogen).

Kanamycin 50 µg/ml (Sigma, K0254).

Lipofectamine® 2000 Reagent (Invitrogen).

Opti-MEM (Gibco).

PageBlue™ Protein Staining Solution (Thermo Scientific).

Polyethylene glycol 6000 (Sigma).

Precision Plus Protein™ Kaleidoscope™ Prestained Protein Standards (Bio-Rad)

Quick-Load Purple 2-Log DNA Ladder (NEB).

Restriction enzymes (NEB).

Ribonuclease H (Invitrogen).

RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen).

VECTASHIELD® Mounting Medium with DAPI (Vector laboratories).

2.1.4.2 Kits

GoTaq® DNA Polymerase (Promega, M7122).

Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, 28903470).

Monarch® DNA Gel Extraction Kit (NEB, T1020S).

NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621S).

Q5® High-Fidelity DNA Polymerase (NEB, M0491S).

QIAGEN Plasmid Midi Kit (QIAGEN, 12145).

QIAprep Spin Miniprep Kit (QIAGEN, 27106).

ReliaPrep™ RNA Cell Miniprep System (Promega, Z6011).

SuperScript™ III First-Strand Synthesis System (Invitrogen, 18080085).

SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Invitrogen, 12574018).

SuperScript™ IV First-Strand Synthesis System (Invitrogen, 18090050).

T4 DNA Ligase (Thermo Scientific, 15224041).

Zero Blunt™ TOPO™ PCR Cloning Kit, with pCR™-Blunt II-TOPO™ Vector (Invitrogen, 450245).

2.1.4.3 Consumables

Gene Pulser® electroporation cuvettes (Bio-Rad), 0.1 cm gap width.

HisPur™ Cobalt Resin (Thermo Scientific).

iBlot™ 2 Transfer Stacks, PVDF, mini (Invitrogen).

Ni-NTA Agarose resin (Thermo Scientific).

Poly-Prep® Chromatography Columns (Bio-Rad).

TruPAGE™ precast 4-20% poly-acrylamide gels (Sigma).

2.1.5 Primers and probes

	Name	Sequence
pCRII vector-specific primers	M13a_F	TGTAAAACGACGGCCAGT
	M13a_R	CAGGAAACAGCTATGACC
ANV GSP primers	GSP_dT	TTTAAAAGTTAGCCAATTC
	IntR1	GGTGTCAATTGAATCTGC
	IntR2	GTTATTCTCAAGTCAAGAGG
	IntR3	CAGCCGTACAAACAACCACG
	IntF1	TCAATGACACAATGAAGACC
	IntF2	GAACCAGTCATCCAACAG
	IntF3	CATGCAATAATATCAGCGTGT
	IntF4	CGAATAGATGGGATGGC

Table 2.1: Primers for ANV G-4260 isolate cloning and sequencing.

Location	Primer	Sequence
3' UTR	Forward	TTGAGAAGGGGTGGACCGTA
	Probe	6-FAM-CAGCAACTGACTTTC-MGB
	Reverse	TCCCAGCTTTCCTGTACCCT
ORF1a	Forward	TCCTTCCCAAGTAAGGCCGATAA
	Probe	6-FAM-TGTTGATGGTGGTTCTGCTGGAGTTGGCT-MGB
	Reverse	TGGCCTGCTGTGTAGATGTAGT

Table 2.2: Primers and probes for qPCR of ANV G-4260. Primer-probe set for the 3' UTR was as described by Smyth *et al*[301].

Name	Sequence
Oligo(dT)$_{20}$	TTTTTTTTTTTTTTTTTTTT
ANV-GSP_F2	CCGAATAGATGGGATGGCTTCGGC
ANV-GSP_R2	AAAAGTTAGCCAATTCAAAATTAATTCAAATAATGAAAAGCCC
ANV-Mid1_R	CGAAGTCAAGCATTTTGG
ANV-Mid2_F	GGACTATCCTGAACTCACTGC

Table 2.3: Primers for synthesis of cDNA/dsDNA for next generation sequencing.

	Name	Sequence
Mutant-specific	WT_F	GGTATGCTGCGTGAT
	WT_R	AGTCTTGCATACCCA
	CH_F	GGGATGCTCCGGGAT
	CH_R	AGCCTGGCGTATCCC
	RT_F	GGTATGCTACGTGAC
	RT_R	AGTCTCGCGTACCCC
	ED_F	GGTATGCTGAGAGAT
	ED_R	TCTCTTGCATACCCA
pANV generic	PS1-sequencer_F	TCTCTGGCACTTGAG
	PS1-sequencer_R	GCCAAGTGTTGTGGA

Table 2.4: Primers for pANV PS1 mutant insert confirmation and sequencing.

2.1.6 Antibodies

2.1.6.1 Primary antibodies

ANV positive chicken-antiserum (EF84/50-289) (a kind gift of V. Smyth, AFBI), diluted 1:500 for immunofluorescence and 1:2000 for western blot.

Mouse Anti-Penta·His (Qiagen, 34660), diluted 1:2000 for western blot.

Mouse Anti-HA.11 (Covance, MMS-101P), diluted 1:2500 for western blot.

2.1.6.2 Secondary antibodies

Goat Anti-Chicken IgY (H+L) Alexa Fluor® 488 (Invitrogen, A-11039), diluted 1:500 for immunofluorescence.

Donkey Anti-Chicken IgG (H+L) IRDye® 800CW (LI-COR, 926-32218), diluted 1:10,000 for western blot.

Goat Anti-Mouse IgG IRDye® 680RD (LI-COR, 926-68070), diluted 1:10,000 for western blot.

2.1.7 Plasmids

Plasmid	**Antibiotic resistance**
pCR™Blunt II (Invitrogen)	Kanamycin
pET23a (A. Panjwani, The Pirbright Institute)	Ampicillin
pEAQ (G. Lomonossoff, John Innes Centre)	Kanamycin
pcDNA3.1(+) (ThermoFisher)	Ampicillin
pUC19 (ThermoFisher)	Ampicillin
pSTBlue-1, containing CAstV VF07-13/7 ORF2 (V. Smyth, AFBI)	Ampicillin and Kanamycin
pOET-1, containing CAstV VF08-3a ORF2 (V. Smyth, AFBI)	Ampicillin

Table 2.5: Plasmids used within this thesis.

2.1.8 Sequences

T7 promotor: TAATACGACTCACTATAG

Synthetic DNA, containing the HDR (dotted box) and T7 terminator (grey highlight):
AAAAAAAAAAAAAAAAAAAATTCCTGCGGCCGGGTCGGCATGGCATCTCCACCTCCTCGCGGTCCG
ACCTGGGCATCCGAAGGAGGACGTCGTCCACTCGGATGGCTAAGGGAGAGCTCGGATCCGGCTGC
TAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT
TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG

Nucleotide sequences (GenBank ID): ANV G-4260 (AB033998.1), CAstV 612 (JN582317.1), CAstV 11672 (JN582327.1), VF07-13/7 (GU222331.1), VF08-3a (GU222332.1).

Protein structures (PDB ID): HAstV-1 capsid core (5EWN), TAstV-2 spike (3TS3), HAstV-2 spike (5KOU), PMV capsid protein (4V99), HEV capsid core (2ZZQ).

2.2 Methods

2.2.1 Nucleic acid methods

2.2.1.1 Preparation of encapsidated RNA by RNase A digestion

IGEPAL® CA-630 (Sigma) was added to samples to a 0.5% final concentration to release any potential membrane-bound RNA from cellular debris, prior to the addition of RNase A to 10 µg/ml (Ambion). Reactions were incubated for 30 minutes at 37 °C, before addition of BL + TG buffer from the ReliaPrep™ RNA Cell Miniprep System (Promega) to inhibit RNase A activity. RNA purification then proceeded as below (2.2.1.2).

2.2.1.2 RNA isolation

RNA extractions were completed using the ReliaPrep™ RNA Cell Miniprep System (Promega) following the manufacturer's instructions. Briefly, samples were treated as the smallest cell input option (1×10^2 - 5×10^5) and were mixed with 100 µl of BL + TG buffer by vortexing for 5 seconds. Next, 35 µl of 100% isopropanol was added and mixed by vortexing for 5 seconds. The resulting suspension was transferred to a Minicolumn and centrifuged at 12,000 *x g* for 1 minute and flow through was discarded. The membrane was washed with 500 µl of RNA wash solution and centrifuged at 12,000 *x g* for 1 minute with the flow through discarded. Next, contaminating DNA was removed by addition of 30 µl DNase I incubation mix to the Minicolumn membrane and incubated for 15 minutes at room temperature. The membrane was then washed sequentially with 200 µl of column wash

solution and 500 µl of RNA wash solution and centrifuged at 12,000 *x g* for 1 minute after each solution was added. The Minicolumn was then placed into a new collection tube and a final 300 µl of RNA wash solution was added and centrifuged at 12,000 *x g* for 2 minutes. The Microcolumn was transferred to an elution tube and 12 µl nuclease-free water added to the membrane and centrifuged at 12,000 *x g* for 1 minute. Purified RNA was stored at -80 °C.

2.2.1.3 First-strand cDNA synthesis

cDNA was synthesised from purified RNA using SuperScript® III reverse-transcriptase (Invitrogen) following manufacturer's instructions. Reactions containing 1.5 µl RNA, 1 µl 10 mM dNTPs, 1 µl 50 µM random hexamers (Invitrogen) and 9.5 µl nuclease-free water were incubated at 65 °C for 5 minutes to denature any secondary structures, then cooled on ice for 1 minute. To this, 4 µl 5x buffer, 1 µl SuperScript III reverse transcriptase, 1 µl 0.1 M DTT, and 1 µl RNaseOUT were added. Reactions were then incubated at 25 °C for 5 minutes followed by 50 °C for 40 minutes for cDNA synthesis, and a final 70 °C for 15 minutes to terminate the reaction, with the resulting cDNA stored at -20 °C.

2.2.1.4 One-step reverse transcription and PCR

For the cloning of avian astrovirus isolates, the SuperScript® III One-Step RT-PCR with Platinum® Taq DNA Polymerase kit (Invitrogen) was used. Reactions contained 25 µl 2x reaction mix, 2 µl SuperScript™ III RT/Platinum™ Taq enzyme Mix, 1.5-2.5 µl of genomic RNA derived from AstV isolates, primers for both reverse transcription and PCR, and nuclease-free water to 50 µl. Reverse transcriptions were primed with either 20 pmol of gene-specific primer (GSP) (section 2.1.5) or 50 pmol of oligo$(dT)_{20}$, PCR was primed with either 20 pmol of forward-sense GSP (section 2.1.5) or 100 ng of random hexamers. For these single-tube RT-PCR reactions, initial reverse transcription was performed at 45 °C for 30 minutes for GSP and oligo(dT) primed reactions, or at 25 °C for 10 minutes and then 45 °C for 30 minutes when using random hexamer primed reactions. After the completion of cDNA synthesis PCR was subsequently performed on these reactions, with an initial denaturation step of 94 °C for 2 minutes, followed by 40 cycles of denaturation at 94 °C for 15 seconds, primer annealing at an appropriate Tm (55-66 °C) for 30 seconds, elongation at 68 °C for 2 minutes 30 seconds, and a final 5 minute elongation step at 68 °C.

2.2.1.5 Polymerase chain reaction

For cloning and generation of products for sequencing, Q5® High-Fidelity DNA Polymerase (NEB) was used. Reactions contained 10 µl 5x reaction buffer, 200 µM dNTPs, 0.5 µM of each forward and reverse primer, < 1,000 ng template DNA, 0.5 µl Q5 polymerase enzyme and nuclease-free water to 50 µl. PCR was performed using an initial denaturation at 98 °C for 2 minutes, followed by 25 cycles of denaturation at 98 °C for 20 seconds, primer annealing at an appropriate Tm for 20 seconds, and elongation at 72 °C for 25 seconds per kilobase amplified, with a final extension step at 72 °C for 2 minutes.

For analytical PCR assays, GoTaq® Green Master Mix (Promega) was used. Reactions contained 10 µl 2x GoTaq® master mix, 1 µM of each forward and reverse primer, < 200 ng template DNA and nuclease-free water to 20 µl. For colony PCR, colonies were picked with sterile pipette tips and dipped into the aliquoted GoTaq PCR reaction mix 2-3 times to introduce template DNA, the same tip was then used to inoculate fresh LB antibiotic cultures for overnight growth. PCR was performed with an initial denaturation at 95 °C for 3 minutes, followed by 18 cycles of denaturation at 95 °C for 20 seconds, primer annealing at an appropriate Tm for 20 seconds, and elongation at 72 °C for 30 seconds per kilobase amplified, with a final extension step at 72 °C for 1 minute.

2.2.1.6 *In vitro* RNA transcription

RNA was synthesised *in vitro* from cDNA downstream of the T7 promotor within pANV, using T7 RNA polymerase (NEB). Reactions contained 0.5 mM of each NTP, 1 µg template DNA, 1 µl RNaseOUT, 1 µl DTT, 2µl T7 RNA Pol, 10x reaction buffer and nuclease-free water to 20 µl. Reactions were incubated at 37 °C for 1 hour, and subsequent RNA was purified as above (2.2.1.2).

2.2.1.7 Restriction enzyme digestion

Restriction endonuclease digest reactions contained 5 units of enzyme (NEB), 0.5 µg DNA, 2.5 µl 10x buffer (as recommended by the manufacturer for the enzyme), and nuclease-free water to 25 µl. Reactions were incubated for 15-60 minutes at the recommended temperature for the enzyme used. Where double-digestion using enzymes with different optimal temperatures was performed, reactions were incubated for 30 minutes at the lower temperature followed by 30 minutes incubation at the higher temperature. Reactions involving DpnI were first incubated at 37 °C for 20 minutes and then 80 °C for 20 minutes to inactivate the enzyme.

2.2.1.8 Agarose gel electrophoresis

Agarose gels (typically 0.8-1% w/v) were prepared by heating agarose in 1x TBE buffer using a microwave until all agarose was in solution. This was then cooled before addition of GelRed® Nucleic Acid Gel Stain (Biotium) to a 1x final concentration and after mixing thoroughly the gel was cast. After the gel had fully set, it was placed into a Sub-Cell® (Bio-Rad) electrophoresis tank containing 1x TBE and combs were removed. DNA samples were mixed with appropriate volumes of 6x loading dye and loaded into the wells with the 2-Log DNA ladder (NEB) used as a size standard on the same gel. Samples were electrophoresed at 100 V for 30-60 minutes and imaged using the Gel Doc EZ imaging system (Bio-Rad).

2.2.1.9 Gel purification of DNA

For gel purification, DNA in agarose gels was visualised using a Safe Imager™ 2.0 Blue Light Transilluminator (Invitrogen) and cDNAs were size selected using a scalpel to isolate desired fragments. A single band of the expected size was extracted for cDNAs amplified using GSPs, while multiple bands, with preference for fragments of 1-2 kb, were taken from amplifications using random hexamers. DNA was extracted from agarose plugs using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) following the manufacturer's instructions. Briefly, 500 µl of capture buffer type 3 was added to the agarose plug in a 1.5 ml Eppendorf to achieve a ratio of 100 µl:100 mg minimum and incubated at 60 °C with occasional vortexing to dissolve the gel. This sample mix was then added to a MicroSpin Column and centrifuged at 16,000 *x g* for 1 minute with the flow-through discarded. 500 µl of wash buffer type 1 was added to the column and centrifuged at 16,000 *x g* for 1 minute, flow-through was discarded and the column then centrifuged again at 16,000 *x g* for 1 minute. DNA was eluted into a clean 1.5 ml Eppendorf tube by adding 40 µl of elution buffer type 6 to the membrane and centrifugation at 16,000 *x g* for 1 minute.

2.2.1.10 DNA ligation

cDNA derived from the AstV isolates was cloned into pCRII using the TOPO® TA Cloning® Kit (Invitrogen) to enable sequencing of AstV inserts. Following the manufacturer's instructions, 4 µl of AstV cDNA, 1 µl Salt Solution and 1 µl TOPO® vector were mixed and ligation reactions were incubated at room temperature for 30-60 minutes or at 4 °C overnight.

Ligation of linear DNA with sticky or blunt ends was performed with T4 DNA ligase (ThermoFisher). Ligation reactions were formed from 100 ng of linear vector DNA, 1:2 molar ratio of insert DNA, 2 µl 10x T4 DNA ligase buffer, 5 U of T4 DNA ligase and nuclease-free water to a final volume of 20 µl. Reactions were incubated at 22 °C for 1 hour, or at 16 °C overnight for complicated ligations.

The NEBuilder® Assembly Tool was used to design primers to PCR amplify AstV cDNA inserts with homologous overlapping ends for seamless cloning into plasmid vectors using the NEBuilder® HiFi DNA Assembly Cloning Kit (NEB). Following the manufacturer's instructions, linear vector:insert DNA at a molar ratio of 1:2, 10 µl of 2x NEBuilder HiFi DNA Assembly Master Mix and nuclease-free water to 20 µl were mixed. These reactions were incubated at 50 °C for 1 hour to ligate inserts into the vector.

The resulting ligation products were then transformed into competent *E. coli* cells (Section 2.2.1.12).

2.2.1.11 Preparation of electrocompetent *E. coli*

50 µl of DH5α cells (Invitrogen) were used to inoculate 10 ml of LB (with this and all subsequent growth steps in absence of antibiotic selection) and incubated with shaking for 2 hours at 37 °C. This starter culture was used to inoculate a 100 ml overnight culture (in LB). The overnight culture was then used to inoculate 1 litre of LB and grown to OD_{600} 0.5-0.6. All subsequent steps were done on ice and in a cold room where possible. The 1 litre culture was split into 6 x 200 ml falcon tubes and incubated on ice for 30 minutes before centrifugation at 3,825 *x g* for 15 minutes at 0 °C in a JLA-16.250 rotor. The supernatant was decanted and 50 ml of sterile ice-cold deionised water was used to re-suspend each pellet, resuspended pellets were then made up to 160 ml each with sterile ice-cold deionised water and centrifuged at 1,875 *x g* for 15 minutes at 0 °C. The supernatants were decanted and the pellets were re-suspended and pooled in a total of 600 ml sterile ice-cold deionised water and split into 3 equal aliquots before centrifugation at 1,875 *x g* for 15 minutes at 0 °C. The supernatants were discarded and the pellets were re-suspended in a total of 40 ml sterile, ice-cold 10% glycerol and subsequently centrifuged at 2,000 *x g* for 15 minutes at 0 °C in a Beckman S4180 rotor. Finally, the supernatant was decanted and the pellet was re-suspended in 2 ml of ice-cold 10% glycerol which was then aliquoted into 50 µl volumes and flash frozen in a dry-ice ethanol bath before storage at -80 °C.

2.2.1.12 Transformation of *E. coli*

Ligation products from cloning reactions were diluted 1:4 in nuclease-free water and incubated on ice until cold, 2 µl volumes of these dilutions were then added to 50 µl aliquots of electrocompetent *E. coli* DH5α cells thawed on ice, before transfer to 0.1 cm Gene Pulser electroporation cuvettes (Bio-Rad) and transformation at 1.8 kV (200 Ohms, 25 microFd) using a Gene Pulser® (Bio-Rad). Transformed cells were recovered in 350 µl of SOC at 37 °C for at least 1 hour with shaking. The entire volume of bacterial suspension in SOC was then spread onto LB antibiotic selection plates and incubated at 37 °C overnight. The following antibiotics (or combinations thereof) were used for selection: Kanamycin 50 µg/ml, Ampicillin 100 µg/ml, Chloramphenicol 25 µg/ml. Transformant colonies were picked using a clean pipette tip, subject to colony PCR (section 2.2.1.5), and inoculated on to a fresh selective LB agar plate or into 5 ml overnight cultures for purification of plasmid DNA.

2.2.1.13 Purification of plasmid DNA (Miniprep)

Plasmid DNA was purified from overnight bacterial cultures using the QIAprep spin Miniprep kit (QIAGEN) following manufacturer's instructions with several modifications. Bacteria were pelleted from 3 ml of overnight culture by centrifugation at 3,000 *x g* for 5 minutes, pellets were resuspended in 250 µl buffer P1, followed by addition of 250 µl buffer P2 and inversion of the tube 3-5 times, until complete lysis of cells is observed (via LyseBlue). Then 350 µl of buffer N3 was added and the tube inverted gently to mix 3-5 times. The lysate was clarified by centrifugation at 21,000 *x g* for 10 minutes with the resulting supernatant transferred to a QIAprep spin column, which was then centrifuged at 12,000 *x g* for 1 minute. 750 µl of buffer PE was added to the column and centrifuged at 12,000 *x g* for 1 minute, the flow-through was discarded and the column centrifuged again at 12,000 *x g* for 1 minute. The column was transferred to a clean 1.5 ml Eppendorf tube and 40 µl of EB was added directly to the column membrane and incubated at room temperature for 1 minute prior to centrifugation at 12,000 *x g* for 1 minute. Purified plasmid DNA was then kept at 4 °C, or -20 °C for long-term storage.

2.2.1.14 Purification of plasmid DNA (Maxiprep)

Plasmid DNA was purified from overnight cultures using the QIAprep spin Maxiprep kit (QIAGEN) following manufacturer's instructions. Bacteria were pelleted from 100 ml of overnight culture by centrifugation at 4,000 *x g* for 10 minutes, pellets were resuspended in

10 ml buffer P1, followed by addition of 10 ml buffer P2 and inversion of the tube 3-5 times, until complete lysis of cells is observed (via LyseBlue). Then 10 ml of buffer N3 was then added and the tube inverted gently to mix 3-5 times. The lysate was clarified by centrifugation at 4,000 *x g* for 60 minutes with the resulting supernatant transferred to a QIAprep column equilibrated with 10 ml buffer QBT, which was then drained by gravity-flow. 2 x 30 ml of buffer QC was added to the column and drained by gravity-flow, the flow-through was discarded. The column was transferred to a clean 50 ml Falcon tube, and 15 ml of buffer QF was added directly to the column membrane and drained by gravity-flow. DNA was precipitated by addition of 0.7 volumes of room-temperature isopropanol and centrifugation at 4,000 *x g* for 60 minutes at 4 °C. The supernatant was then carefully decanted and the pellet washed with 5 ml room-temperature 70% ethanol, before centrifugation at 4,000 *x g* for 60 minutes at 4 °C. The supernatant was discarded and the pellet air dried for 5 minutes before resuspension in 500 µl of nuclease-free water.

2.2.1.15 Sequencing and analysis

Purified plasmid DNA containing AstV inserts were submitted to SourceBioscience or GATC for sequencing. Plasmid DNA was made to 100 ng/µl for sequencing and where custom primers (Section 2.1.5) were provided these were at a concentration of 3.2 pmol/µl. Sequencing of AstV isolate cDNAs within the pCII-TOPO vector was performed using the M13a forward and reverse primer sites flanking the TOPO insertion site of the vector. Due to the large size of some AstV inserts, internal GSPs were used to achieve complete sequencing coverage in addition to sequencing from the M13 flanking sites (Section 2.1.5). Sequence data was then assembled into contigs using SeqMan Pro (DNASTAR lasergene), using the default assembly parameters. For template-based contig assembly, the published ANV G-4260 genome sequence (GenBank: AB033998.1) or partial genome sequences of CAstVs 612 and 11672 (GenBank: JN582317.1 and JN582327.1, respectively) were set as the reference sequence.

2.2.2 *In Silico* modelling of the ANV capsid protein

The SWISS-MODEL webserver[302] was used to predict the tertiary structures of the ANV G-4260 capsid core and spike domains from the full ORF2 primary amino acid sequence (GenBank ID: AB033998.1) by homology modelling against structures in the SWISS-MODEL template library, derived from experimentally determined structures in the protein data bank.

Superimposition of protein structures was performed using the FATCAT pairwise alignment webserver[303]. The alignment model was set to flexible and used to compare the predicted ANV core and solved HAstV-1 core domain (PDB ID: 5EWN) structures, the predicted ANV spike and solved TAstV-2 spike (PDB ID: 3TS3) structures, the solved HAstV-1 and HEV core domains (PDB ID: 5EWN and 2ZZQ, respectively), as well as the predicted ANV core and solved PMV capsid structure (PDB ID: 4V99). The PDB structure files generated by SWISS-MODEL for the predicted ANV domains were used as templates for the pairwise alignments.

The COACH webserver[304] was used to predict potential ligands and their binding sites of the modelled ANV core and spike protein domains, using the SWISS-MODEL predicted PDB structures as query templates.

All predicted structures and superimpositions were analysed and imaged using PyMOL[305]. The surface electrostatic potential of the predicted ANV core and spike was calculated using the built-in tool ('Protein contact potential (local)') on PyMOL. Images were captured using the command 'ray 1000, 1000' and saved using the command 'png *FileName*, dpi=600'.

2.2.3 Cell culture methods

2.2.3.1 Routine passage

Adherent cell lines were grown to 80-90% confluence prior to passage. To subculture, media was removed from flasks and the cells were washed once with PBS. Cells were detached by incubation with minimal volumes of 0.25% trypsin-versene at 38.5 °C for 1-2 minutes. Detached cells were then resuspended with media containing serum and seeded into new flasks with fresh media at an appropriate seeding density for that cell line.

Details for primary cell and cell line origins and culture medium composition are found in sections 2.1.2 and 2.1.3, respectively.

2.2.3.2 Mammalian cell culture

BSRT7 cells were grown in complete DMEM supplemented with 1 mg/ml Geneticin (Gibco), in a 5% CO_2 atmosphere at 38.5 °C. Cells were sub-cultured at a seeding rate of 1:20-1:30.

2.2.3.3 Avian cell culture

CK cell preparation: CK cells were prepared from 2-3 week old SPF Rhode island red chickens by Cell culture services at the Pirbright Institute following the method of Hennion and Hill[300]. After seeding, primary CK cells were grown for 3 days in CK growth media at 37 °C in a 5% CO_2 atmosphere, CK media was then removed and cells washed once with PBS before addition of 1x BES media to maintain the cells for up to an additional 5 days.

LMH cells were grown in complete EMEM (Gibco) at 38.5 °C in a 5% CO_2 atmosphere. Cells were sub-cultured at a rate of 1:4-1:6. All flasks used to culture LMH cells were pre-treated with 1x attachment factor (Gibco) for at least 30 minutes, with excess attachment factor removed before addition of cells and media.

DF-1 cells were grown in complete DMEM (Sigma) at 38.5 °C in a 5% CO_2 atmosphere and sub-cultured at a rate of 1:10-1:20.

CEF preparation: Bodies of decapitated 9 day-old chicken embryos, derived from SPF eggs (VALO Line, BioMedia) were washed with PBSa and then homogenised through a 10 ml needless syringe. The resulting tissue was washed twice in PBSa prior to the addition of Trypsin. Tissue was incubated with trypsin for 4 minutes, with stirring, after which the supernatant was carefully decanted and transferred into NBCS to neutralise the trypsin. This process was repeated as necessary until cells no longer dissociated from the tissue, typically 3-4 cycles. Pooled supernatant was filtered through a coarse gauze filter and then centrifuged for 10 minutes at 524 *x g*. The resulting pellets were then resuspended in fresh warm CEF growth medium and passed through a 40 µm pore filter (Swinnex), before being counted and seeded into flasks with fresh CEF growth medium. After seeding, primary CEF cells were grown to 80-90% confluence in CEF growth medium at 38.5 °C in a 5% CO_2 atmosphere, CEF growth medium was then removed and cells washed once with PBS before addition of 1x CEF maintenance medium.

2.2.4 Expression of recombinant astrovirus capsid protein

2.2.4.1 Mammalian and avian expression systems

Mammalian and avian cells were transfected using Lipofectamine® 2000 Reagent (Invitrogen) following the manufacturer's instructions. Expression vector plasmid DNA (2.5 µg) was mixed with 250 µl of Opti-MEM (Gibco), whilst 10 µl of Lipofectamine reagent was separately mixed with 250 µl of Opti-MEM per transfection. These suspensions were

incubated for 5 minutes and then mixed together before a further 30 minute incubation. Cells of ~80-90% confluency in 6 well plates were washed with PBS and given 2 ml fresh media per well prior to addition of the DNA-lipofectamine complexes, no further media changes occurred. Cells were then harvested 48-72 hours post-transfection for analysis.

2.2.4.2 Bacterial expression systems

Starter cultures (5 ml) of *E. coli* strain Rosetta™(DE3)pLysS cells transformed with pET23a protein expression vectors were grown overnight in LB at 37 °C with shaking. Starter cultures were used to inoculate 50 ml volumes of LB and were grown to an OD_{600} of 0.6-0.8, cultures were then induced using Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.4 mM. Cultures were induced for 5 hours at 30 or 37 °C in a shaking incubator set to 200 rpm before cells were harvested by centrifugation. Cells were kept under antibiotic selection with 25 mg/ml Chloramphenicol and 100 mg/ml Ampicillin during both growth and expression in LB.

2.2.4.3 *In planta* expression systems

Nicotiana benthamiana plants used in this study were grown for 5-6 weeks in greenhouses maintained at 24 °C and 70% relative humidity with 16 hours of light per day.

Agrobacterium tumefaciens carrying the pEAQ expression vector containing recombinant AstV capsid genes were grown and used to infiltrate leaves of *Nicotiana Benthamiana* plants to direct expression of the recombinant viral protein within leaves as described previously[306]. Briefly, *A. tumefaciens* harbouring the pEAQ-AstV expression vectors were grown to stable phase in LB containing Rifampicin (50 μg/ml). *A. tumefaciens* cells were pelleted by centrifugation at 4,000 *x g* and then resuspended in MMA to 0.4 OD_{600} and incubated at room temperature for 1 hour before pressure infiltration through small grazes into the air spaces of fully expanded *N. benthamiana* leaves using a needless syringe. Transduced leaves were harvested 7 days post-agroinfiltration for protein isolation and analysis.

2.2.5 Purification of recombinant astrovirus protein

2.2.5.1 Purification of AstV capsid protein expressed in *E. coli*

Cultures of *E. coli* BL21 pLysS Rosetta cells (50 ml) expressing recombinant ANV CP were centrifuged at 3,000 *x* g for 10 minutes, the supernatant was discarded and the pellet re-suspended in 3 ml of TN buffer containing cOmplete™, EDTA-free Protease Inhibitor

Cocktail (Roche). The suspended cells were then split into 2 x 1.5 ml Eppendorf tubes and lysed by sonication (Diagenode Bioruptor) in an ice water bath for 15 minutes at medium power for 30 seconds on, 30 seconds off. The lysate was then clarified by centrifugation at 16,000 x g at 4 °C for 20 minutes before purification of the His-tagged recombinant AstV capsid protein by cobalt affinity chromatography (Section 2.2.5.3).

2.2.5.2 Purification of AstV capsid protein expressed in *N. benthamiana*

Protein was extracted from agroinfiltrated *Nicotiana benthamiana* leaves as described by Saunders and Lomonossoff[306]. Leaves were homogenised in 3 volumes of extraction buffer containing cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche) by mechanical blending. Homogenate was then cleared of insoluble debris through Miracloth and the supernatant further clarified by centrifugation at 12,000 x *g* for 30 minutes. The resulting homogenate was either processed for nickel affinity chromatography (Section 2.2.5.3) to isolate His-tagged recombinant AstV CP, or for sucrose gradient centrifugation to isolate VLPs (Section 2.2.8.3).

2.2.5.3 Cobalt and nickel affinity chromatography

Cobalt resin (ThermoFisher) or nickel resin (ThermoFisher) was prepared for purification of recombinant His-tagged AstV capsid protein by washing 3 times in TNS buffer containing 5 mM Imidazole, resin was then incubated with clarified cell lysate for 60 minutes at room temperature mixing on a roller. Lysate and resin was transferred to Poly-Prep chromatography columns (Bio-Rad) and the lysate was allowed to drain through the resin matrix under gravity flow. Once all lysate had passed through the resin bed it was washed twice with 20 bed volumes of TNS buffer containing 5 mM Imidazole and once with 20 bed volumes of TNS buffer containing 10 mM imidazole. Purified protein was then eluted from the resin using 1 bed volume of TNS buffer supplemented with 300 mM imidazole. The resulting purified protein was stored at 4 °C and used for analysis by SDS-PAGE (Section 2.2.5.4) and transmission electron microscopy (Section 2.2.8.4).

2.2.5.4 SDS-PAGE and Western Blotting

For SDS-PAGE analysis, protein samples were mixed with 2x boiling mix (TruPAGE™ LDS and DTT sample buffer, Sigma) and boiled at 85 °C for 5 minutes, then loaded on to TruPAGE™ precast 4-20% poly-acrylamide gels (Sigma). Samples were electrophoresed using TruPAGE™ TEA-Tricine SDS running buffer (Sigma) at 180 V. Gels were then either stained with PageBlue (ThermoFisher) for 1 hour and washed with deionised water, or transferred

onto a PVDF membrane using the iBlot® 2 Gel Transfer system (ThermoFisher) at 23 V for 6 minutes. Membranes were blocked for at least 30 minutes in 5% skimmed milk in PBS (blocking buffer), and then incubated with primary antibody diluted in blocking buffer, either for 1 hour at room temperature or overnight at 4 °C with shaking. Primary antibody was removed, and membranes washed 3 times with TBS-T for 5 minutes. LI-COR compatible fluorescent secondary antibodies were diluted in blocking buffer and incubated with membranes at room temperature with shaking for one hour whilst covered in tin foil to minimise light degradation of the fluorophore. The secondary antibody was removed and membranes washed with TBS-T as before, prior to rinsing membranes once with PBS. Membranes were imaged using the Li-COR Odyssey and analysis was performed using the Image Studio™ Lite software (Li-COR).

2.2.6 Recovery and propagation of ANV from the reverse genetics system

2.2.6.1 Recovery of ANV in BSRT7 cells

To recover infectious ANV virions, the reverse genetics plasmid containing the full-length cDNA clone of the ANV genome under the control of a T7 promoter was transfected into BSRT7 cells using lipofectamine 2000 (Section 2.2.4.1). BSRT7 cells constitutively express the T7 RNA polymerase to drive transcription of the ANV genome from the plasmid cDNA clone. At 48 hours post-transfection, extracellular culture media was collected and centrifuged at 1,500 x *g* for 10 minutes with the resulting supernatant filtered using a 0.22 µm syringe filter (Millipore) to remove debris from this extracellular virus preparation.

Intracellular virus was also harvested from BSRT7 cells, cells were detached from flasks using 0.25% trypsin-versene and pelleted at 1,500 x *g* for 10 minutes, before resuspension in minimal volumes of 2% serum DMEM and homogenisation using a TissueLyser II (QIAGEN) at 28 Hertz for 1 minute. Extracellular and intracellular virus preparations from transfected BSRT7 cells were then used to inoculate confluent CK cells using neat or 10-fold serial dilutions of virus suspensions. The inoculum was removed 4 hpi and cells were washed with PBS and overlaid with BES medium for continued incubation until cytopathic effect (CPE) was observed.

2.2.6.2 Propagation of ANV in LMH and CK cells

ANV was further propagated in LMH and CK cells using clarified and filtered extracellular media harvested from ANV infected CK cells (Section 2.2.6.1) showing high levels of CPE,

typically 72-96 hpi. Inoculation of T75 and T150 flasks of LMH or CK cells was performed using 0.5- and 1-ml volumes of media from ANV infected CK cells respectively.

2.2.7 Characterisation of recovered ANV

2.2.7.1 Viral titration by plaque assay

To determine the titre of ANV, confluent CK cells in 6-well plates were inoculated with 250 µl of 10-fold serially diluted virus suspension (in BES) from infected media of CK cells for 1 h. The inoculum was removed and the cells were overlaid with 2 ml BES medium supplemented with 1% agar to limit virus diffusion. At 72 hpi, monolayers were fixed by addition of 10% formaldehyde for 1 hour, this was carefully removed along with the agar-BES plugs, using a flat-edged spatula, and the CK cell monolayer was then stained using 0.1% crystal violet for 20 minutes before rinsing with water, drying and imaging to quantitate the titre of virus stocks as plaque-forming units (pfu)/ml.

2.2.7.2 Viral titration by $TCID_{50}$

To determine the 50% tissue culture infectious dose of ANV, confluent CK cells in 24-well plates were inoculated in quadruplicate with 100 µl of 10-fold serially diluted virus suspension from infected media of CK cells in BES medium. Presence of CPE in each well was reported at 48, 72, 96 and 120 hpi and $TCID_{50}$ was calculated using the Spearman and Kärber algorithm[307].

2.2.7.3 Quantitative RT-PCR

Quantitative-PCR (qPCR) for ANV was performed based on the method of Smyth *et al*[301]. cDNA for qPCR was prepared from purified RNA as described earlier (Sections 2.2.1.2 and 2.2.1.3). qPCR reactions were set up in triplicate for each cDNA sample with each reaction comprising 10 µl 2x ABsolute Blue qPCR Low ROX master Mix (ThermoFisher), primers to a final concentration of 400 nM, FAM dye-labelled probe with MGB quencher to a final concentration of 120 nM, 2 µl of either template, standard or control DNA and nuclease-free water to 20 µl. Where the published method targeted a region within the 3' UTR using degenerate primers[301], an additional primer and probe set used here were designed to specifically target a 106 bp region in the ORF1a of ANV (Section 2.1.5). 2 µl of nuclease-free water was used for the negative control reactions and the recombinant ANV reverse genetics plasmid was used to create standards, ranging from $1 \times 10^8 - 1 \times 10^3$ copies per 2 µl in tenfold dilutions. Reactions were run using an Applied Biosystems 7500 Real-Time PCR

System; starting with initial denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 sec, and then primer annealing and template amplification at 60 °C for 45 sec. Results were analysed using the manufacturer's software to calculate absolute copy numbers based on the recombinant ANV reverse genetics system plasmid reference standard.

2.2.7.4 Immunofluorescence

CK and LMH cells were grown to 80-90% confluency on 13 mm glass coverslips in 24 well plates. Where LMH cells were used, coverslips were pre-treated with 1x attachment factor (Gibco) as described previously (Section 2.2.3.3). Confluent cells were infected with virus at an MOI of ~0.02 (calculated from $TCID_{50}$) and fixed at 72 hpi using 4% paraformaldehyde in PBS for 1 hour. Cells were then washed in PBS and treated with 0.1% Triton X-100 for 15 minutes and washed again with PBS. The cells were blocked using 0.5% bovine serum albumin in PBS for 30 minutes and then incubated with primary antibody for at least 60 minutes. Excess primary antibody was removed by 3 x 5 minute washes in PBS before incubation with the secondary antibody for 60-90 minutes. Coverslips were then washed in PBS as for the primary antibody, rinsed with de-ionised water and carefully blotted vertically against tissue to pull residual liquid off the cells and dry the bottom of the coverslip before mounting. Coverslips were mounted on VECTASHIELD® containing 4',6-diamidino-2-phenylindole (DAPI) to stain nuclei, and sealed in place using clear nail varnish. Cells were observed with a Leica SP5 confocal laser scanning microscope using the 63x objective lens at 1.4 magnification and 8 bits resolution, and illuminated with the 405, 488 and 561 nm lasers. Images were captured and analysed using Leica's LAS AF software.

2.2.8 Virus purification

2.2.8.1 Concentration of extracellular ANV virions

Extracellular media from infected CK cells was harvested and cellular debris was removed by centrifugation at 2,000 *x g* for 10 minutes and filtration of the resulting supernatant through a 0.22 µm syringe filter unit (Millipore). Extracellular ANV was precipitated by addition of Polyethylene glycol (PEG) 6000 to the clarified supernatant to a final concentration of 8% and incubated for 1 hour at room temperature with continuous mixing until fully dissolved. PEG precipitated virus was pelleted by centrifugation at 4,500 *x g* for 1 hour at 4 °C, then gently re-suspended in minimal TN buffer over 1-3 hours to concentrate the virus.

2.2.8.2 Sucrose purification of ANV

ANV virions within extracellular media of infected CK cells were subject to concentration by centrifugation onto a 60% sucrose cushion (in PBSa) at 130,000 *x g* for 2 hours at 4 °C using a Beckman Coulter SW55Ti or Sorvall Surespin 630 rotor depending on volumes to be processed. The interface above the 60% sucrose was then carefully extracted and diluted in PBSa, before centrifugation through a 20, 40 and 60% sucrose step-gradient (in PBSa) at 130,000 *x g* for 2 hours at 4 °C. Fractions for analysis were then taken from the top of the tube in 1 ml volumes using a pipette.

2.2.8.3 Sucrose gradient purification of CAstV 11672 capsid protein

10-50% sucrose (in TN buffer) gradients were preformed using a Gradient Master™ (BioComp) following the manufacturer's instructions. The soluble lysate of *N. benthamiana* leaves agroinfiltrated to express recombinant CAstV 11672 CP was carefully loaded onto the top of the preformed sucrose gradient and centrifuged at 136,800 *x g* for 1.5 hours at 4 °C using a Thermo Scientific™ TH-641 Swinging Bucket Rotor. Gradient fractions were the harvested for analysis by puncturing the tubes near the bottom with a needle and collecting 1 ml fractions under gravity flow.

2.2.8.4 Electron microscopy

For negative-stain examination of protein structures and virions, 7 µl of purified protein or viral sample (Sections 2.2.5.3, 2.2.8.1 and 2.2.8.2) was applied to freshly glow discharged carbon formvar 200 mesh copper grids and incubated for 2 minutes. Excess liquid was removed by blotting with filter paper before grids were stained with 2% uranyl acetate for 60 seconds, blotted once again and air dried. Grids were imaged at 100 kV using an FEI T12 transmission electron microscope, and micrographs were collected using a TemCam-F214 camera (TVIPS).

For examination of ultra-thin cellular cross-sections, infected cells on Thermanox™ coverslips (ThermoFisher) were fixed sequentially with 2% glutaraldehyde followed by 1% osmium tetroxide for 60 minutes each. Fixed samples were dehydrated with ethanol at concentrations of 70% for 30 minutes, 90% for 10 minutes, then 100% for 3 x 10 minutes. Samples were infiltrated with propylene oxide and epoxy resin (Agar Scientific) at a 1:1 ratio for 60 minutes, then with 100% resin at 60 °C overnight, from which 70 nm sections were cut from resin blocks using a UC7 Ultramicrotome (Leica). Ultra-thin cross-sections

were stained with uranyl acetate and lead citrate then imaged as for negative-stain preparations.

2.2.9 Sample preparation for next-generation sequencing

2.2.9.1 Viral RNA sample preparation and purification

To prepare packaged and total populations of ANV RNA for NGS and bioinformatics analysis, 18 x T150 flasks of confluent chicken kidney (CK) cells were infected with reverse genetics recovered ANV that had been passaged in CK cells 9 times. Virus was harvested from the flasks of infected CK cells at 72 hpi. For packaged RNA, the extracellular media from infected CK cells (~400 ml) was clarified by centrifugation at 3,000 *x g* for 10 minutes, and the supernatant was then filtered using a 0.22 µm syringe filter. Virions were then concentrated by PEG precipitation (Section 2.2.8.1), with the resulting pellets resuspended in 1 ml TN buffer containing 1% IGEPAL® CA-630. Total intracellular RNA was prepared from cells of 4 x T150 flasks of infected CKCs, these were harvested by scraping, pooled and then pelleted after extracellular media had been removed. RNA was purified from the packaged and total fractions using the ReliaPrep™ RNA Cell Miniprep System (Section 2.2.1.2), except that the DNase I step was omitted and all samples were kept on ice where possible with final elution of RNA in 12 µl of nuclease-free water.

2.2.9.2 Genome-length cDNA synthesis

cDNA for NGS was synthesised from purified RNA (Section 2.2.9.1) using SuperScript® IV (SSIV) reverse-transcriptase (Invitrogen) following the manufacturer's instructions. Reactions contained 11 µl purified RNA, 1 µl 10 mM dNTPs and 1 µl 50 µM Oligo $d(T)_{20}$ (Invitrogen) and were incubated at 65 °C for 5 minutes, then cooled on ice for 1 minute. To this, 4 µl 5X SSIV buffer, 1 µl 100 mM DTT, 1 µl RNaseOUT and 1 µl SuperScript IV reverse transcriptase were added. Reactions were then incubated at 50 °C for 10 minutes with a final incubation at 80 °C for 10 minutes. Template RNA was removed from first strand cDNA prior to PCR by adding 1 µl RNase H (Invitrogen) and incubating the reaction at 37 °C for 20 minutes.

2.2.9.3 Genome-length cDNA amplification

Q5® High-Fidelity DNA Polymerase (NEB) was used for genome-length PCR as described above (Section 2.2.1.5), using 2 µl of the RNase H-treated SSIV reaction as template DNA (Section 2.2.9.2). The PCR reaction was adjusted to have an initial denaturation at 98 °C for

1 minute, followed by 40 cycles of denaturation at 98 °C for 20 seconds, primer annealing for 20 seconds at 65 °C, and elongation at 72 °C for 5 minutes, with a final extension step at 72 °C for 2 minutes.

2.2.9.4 Purification of genome-length DNA

Genome-length PCR products (Section 2.2.9.3) were size selected using agarose gel electrophoresis (Section 2.2.1.8). DNA was purified from the gel slice using the Monarch® DNA Gel Extraction Kit (New England Biolabs) following the manufacturer's instructions. Four volumes of Monarch Gel Dissolving Buffer were added to the cut agarose gel slices in 1.5 ml Eppendorf tubes and heated at 50 °C until fully dissolved. The sample mix was then added to the MicroSpin Column and centrifuged at 16,000 *x g* for 1 minute, flow-through was discarded, and 200 µl of DNA Wash Buffer was added and centrifuged at 16,000 *x g* for 1 minute. Flow-through was again discarded and the column centrifuged a second time at 16,000 *x g* for 1 minute. Finally, DNA was eluted into a clean 1.5 ml Eppendorf tube by adding 6 µl of DNA Elution Buffer to the centre of the matrix, incubated for 1 minute at room temperature and then centrifuged at 16,000 *x g* for 1 minute.

Quality control was performed on 1:10 dilutions of each eluted DNA prior to NGS using the Qubit Fluorometer (Invitrogen).

2.2.10 Next-generation sequencing and bioinformatic data analysis

2.2.10.1 Illumina Library preparation

To prepare DNAs for sequencing, samples were diluted to 0.2 ng/µl in a total volume of 1 µl. Library preparation was then performed using the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's instructions with a Hamilton NGS STAR (Hamilton Robotics), as per official Illumina scripts. Libraries were sequenced on an Illumina MiSeq platform using a v2 300-cycle cartridge to produce paired end reads of around 150 bp each.

2.2.10.2 Sequence data processing

The quality of the raw paired-end reads generated by the MiSeq were first assessed using FastQC (v0.11.5)[308]. Each read was trimmed to remove the first 15 and last 5 bases from the 5' and 3' ends, respectively, using PrinSeq-Lite (v0.20.4)[309]. Reads were further trimmed using Sickle (v1.33)[310] to remove sequences with a Phred quality score below q38 and with read lengths less than 70 bp in size.

Trimmed reads were then aligned to the ANV G-4260 reference sequence (GenBank: AB033998.1) using BWA-MEM[311], and the resulting .SAM file was converted into a .BAM file using SAMtools (v1.2)[312]. The Linux Grep command was used to extract all sequences in the alignment with a mismatch equal to or less than 4 nucleotides to the reference sequence, via the NM:i: tag, creating a new .BAM file in which the potentially misaligned sequences had been removed from the contigs.

The SAMtools "sort" utility was used to order the sequence alignments to that of the reference genome and then the "mpileup" utility was run, generating read coverage, and the variants at each genomic position. Finally, an in-house R script, eleano.R (written by Rocio Enriques-Gasca, The Genome Analysis Centre, Norwich), was run to count the quantities of each base (A, T, G, C) at each genome position, and output the results in a chart formatted text file.

2.2.10.3 Entropy calculations

The equation to calculate Shannon-Wiener diversity (H') at each genome position in the sequence alignments was adapted from a previous study on the population diversity of FMDV[80]:

$$H' = -\sum_{b \in \{A,C,G,T\}} p_b log_2(p_b)$$

Where p_b represents the proportion of each base (A, C, G, T) present.

In order to analyse data within the chart formatted text file generated by the eleano.r script, the equation was converted into an excel function, and the entropy (y) for each nucleotide at each position calculated using:

y[A,C,G,T] = **IF(X=0,0,X*(LOG(X,2)))**

Where X represents the occurrence of a specific nucleotide (A, C, G, T) in the ensemble of variants at a particular genome position as a percentage. This function was run four times at each genome position, for each individual nucleotide, to give y[A], y[C], y[G] and y[T], respectively.

The total entropy for each genome position (H') was then calculated using:

H' = **-SUM(y[A]+y[C]+y[G]+y[T])**

The total entropy at each genome position in the packaged population was then subtracted from that of the total population, to give the difference in Shannon's entropy scores.

2.2.10.4 Mutation rate calculations

To calculate the difference in mutation rate at each genome position between the packaged and total populations, the number of nucleotide variants at each individual position in a population, relative to that of the reference sequence, were calculated as a percentage. The percentage of variants at each position in the packaged population was then subtracted from that of the total population, to give the difference in mutation rate.

Chapter 3

3 *In silico* modelling of the ANV capsid protein structure

3.1 Astrovirus capsid proteins

Encapsidation of the viral genome is essential to the production of infectious progeny virions. Studies on several +ssRNA viruses have shown that rapid and selective encapsidation of the genome is achieved through a PS-mediated assembly mechanism[43, 52, 58, 59, 62-65], in which the CP and the RNA play crucial roles. The *Astroviridae* share several features with viruses known to utilise PS-mediated assembly, including a basically-charged N-terminal extension in the CP and highly conserved RNA SLs within the genome (RFS and s2m), suggesting this viral family may utilise a similar mechanism of genome packaging. Capsids of the *Astroviridae* are formed from the product of the ORF2 gene which is translated from a subgenomic RNA to produce a CP precursor of 72-90 kDa, depending on the strain of virus. This precursor has a common structure across the *Astroviridae*, containing a highly basic N-terminal region (residues ~1-70) leading to a conserved N-terminal domain representing the capsid core, then a variable domain representing the spike, ending with an acidic C-terminal region containing multiple putative caspase cleavage sites. Crystal structures for the HAstV-1 and -8 capsid core domains, as well as the HAstV-1, -2, -8 and TAstV-2 spike domains have recently been determined[193, 194, 220-222]. These data show there is high structural homology between the core and spike domains of different HAstV isolates[194, 220, 222], whereas, pronounced structural variation in the spike domain is seen between the HAstVs and TAstV, correlating to the significant difference in primary sequence between these strains[202, 221].

Excluding the spike protein of TAstV-2, there are no solved crystal structures for the CPs of the AAstVs, limiting the knowledge to use as a starting point for investigation into the genome encapsidation mechanism of ANV. To begin addressing this, in the absence of a solved ANV CP structure, a computational approach was used to identify, map and predict the tertiary structures of domains within the capsid protein by homology modelling with known astrovirus structures.

3.2 Homology modelling of the ANV capsid protein

3.2.1 Identification of structural domains encoded within ANV ORF2

To begin investigating the structure and functional domains of the ANV capsid protein, comparative analysis on the ORF2 encoded protein sequence was performed using the SWISS-MODEL structure prediction software[302]. This programme performs searches using BLAST and HHblits to find related sequences and structures to use as templates to model prediction of structures for proteins without solved structures. SWISS-MODEL identified two main domains within the ANV CP, likely representing a capsid core and a spike region, consistent with the structures of other reported astrovirus CPs[225] (Figure 3.1).

```
  1 MAGGATAPAGAKPKQPKQKQKKPSSQARKKPSQKQKAMKPVKQELRKVEKQVRVLKARTN
 61 GPKVNDTMKTTVTVGTLVGQTQSGLNRQLRVSFNPLLMKSTEGGSTTPLSIRASMYEMWK
121 PLSVEIFATPLSGFSSVVGSVGFMVITLNGLEASADSIDTIKARRHVQMALGRPYRLKLS
181 ARELAGPREGWWLVDTSEAPADAYGPAVDLMLAYATENLLGTSSGSTTSYTGTLWQVEMR
241 VTYAFSTYNPKPGLQTLVSQSITGGQTVTIQPSPDDGSLIMTTNSQQVLALLTPRVAGQR
301 KGKSQTIWAIAGSAVDAAATVLGPWGYLLKGGFWLVRLIFGGSSARNTTTRQYQIYPSVE
361 SALTDQPIFGNSTGTQSVTVPICHITEVVNPNAESNNLPPPTTGAQPQPQPPAPIEEILL
421 PLAELTGQPGVPPLYTFDGSSYTPPTNWLGSTILLTGIPAHKRVTGNLAKFGVTNLQMSK
481 VAATALEIYDFTDFGVFFGTGSYLSEGGIHTGKTLIYSLMSGQTPNPWLAANQSGTTWYM
541 PSWAGFPQPGQGDYFLQMQDVTDTTTHTTSVNVYFLVAYRQSRRLIAFFNTGGTARPAPT
601 SMLCLYNVDCGRAPQTPYPTFQSTLQSLNQIGVDAKSDPDSDDDISLAGSVIGDEFDSVD
661 HLERERELMRRLRDLDLRRFQI
```

Figure 3.1: Domains of the ANV ORF2 encoded capsid protein. The domains encoded within ANV ORF2 (GenBank ID: AB033998.1) identified by SWISS-MODEL are shown. Amino acids 1-64 represent the highly basic unstructured N-terminus, followed by the capsid core domain (65-391, dark purple) which is separated from the spike domain (417-627, lilac) and acidic C-terminus (628-683) by a flexible region (392-416). Unstructured regions of the protein are light grey and previously predicted caspase cleavage sites[238] are underlined.

3.2.2 Structure prediction of the ANV CP domains

The sequence homology of the ANV ORF2 encoded CP to those of other AstVs with solved crystal structures was sufficient to allow template-based modelling of predicted structures for the ANV CP spike and core domains.

3.2.2.1 Spike domain

The closest predicted structural homologue of the ANV spike domain was the spike (P2) domain of TAstV-2 (Figure 3.2), despite low sequence similarity. Whilst there is a similarity in overall topology between the two avian astrovirus spike domains, there is variation on the outer surface, particularly at the putative receptor-binding site, where a shallow depression is located[221]. This may infer that even though the capsids of these two Avastroviruses may be closely related, they potentially have very different receptor tropisms, consistent with *in vitro* data[192]. The structure of the HAstV-1 spike monomer is also shown to highlight the significant structural variation to that of the avian astrovirus spike domains (Figure 3.2 D), as may be expected for the low sequence homology and differing host tropisms between them (Table 3.1).

3.2.2.2 Proteolytic cleavage of the capsid protein

By analogy to the studied HAstVs, the astrovirus capsid is required to undergo one or more cleavages by various host caspases in order to trigger release of the particle from the cell[238, 239]. Previous work has identified 3-4 putative caspase sites in the capsid protein of ANV[238] (Figure 3.1). After modelling the spike domain of ANV, it appears that the first putative site (DVTD) is unlikely to be a viable target for cleavage as it is found buried internally within the structure of the spike and is likely non-accessible. If the structure modelled here is indeed similar to the crystal structure of TAstV-2, then the latter 3 sites are potentially viable targets for host caspases, with the 'DSDD' motif being most optimal for trimming the greatest amount of the acidic region from the C-terminus of the spike domain.

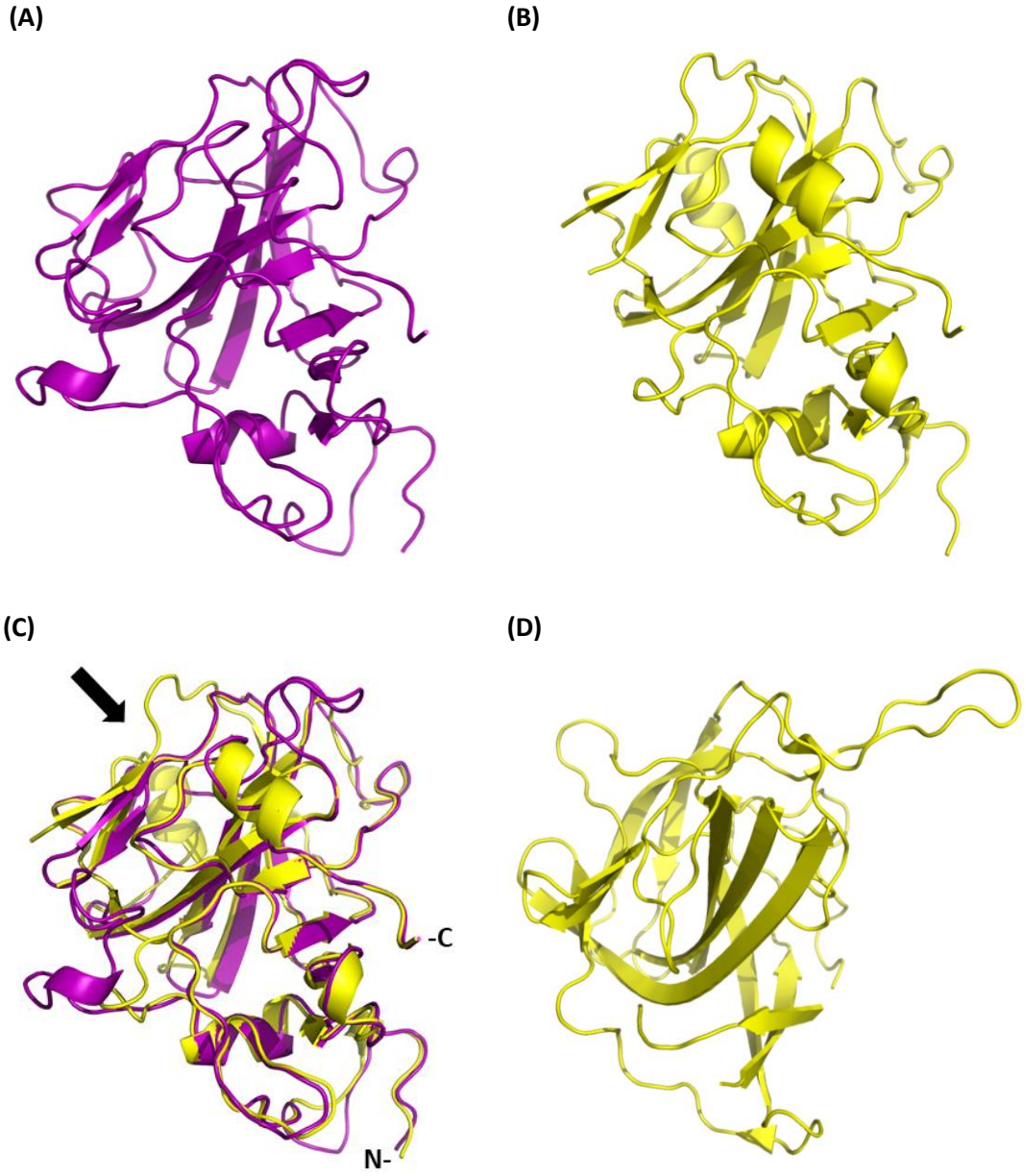


Figure 3.2: Structural comparison of ANV, TAstV-2 and HAstV spike domains. (**A**) The structure of the ANV spike monomer was computationally predicted using SWISS-MODEL, based on (**B**) the known crystal structure of the TAstV-2 spike monomer (PDB ID: 3TS3). (**C**) Superimposition of the predicted ANV capsid spike and known TAstV-2 spike. The structure alignment has 187 equivalent positions with a root-mean-square deviation (RMSD) of 1.53 Å, without twists, indicating close positioning of the polypeptide backbone atoms. The putative receptor site of the TAstV-2 spike is indicated by an arrow. (**D**) The solved crystal structure of the HAstV-1 spike monomer (PDB ID: 5EWO) is shown, highlighting the structural dissimilarity to the AAstV spikes.

	BLASTp results Vs. ANV-G4260 ORF2 (1-683)	
	Query cover (position)	*Sequence identity*
HAstV-1 (AAC34717.1)	51% (25-367)	30.62%
TAstV-2 (NP_987088.1)	94% (21-666)	33.13%
CAstV VF08-56 (AFK92942.1)	75% (36-528)	37.65%
HEV (P29326)	46% (1-250)	19.71%

Table 3.1: Sequence homology between CPs of ANV, HAstV, TAstV and CAstV as well as HEV.

3.2.2.3 Core domain

The closest predicted structural homologue to the ANV capsid core domain was for the core (S and P1) domains of HAstV-1 (Figure 3.3). This *in silico* modelling approach showed high homology between the predicted ANV and solved HAstV-1 capsid core structures, despite low amino acid sequence conservation (Table 3.1). This structural homology was particularly notable in the S domain which contains a jelly-roll β-barrel fold that is located on the inside of the capsid shell. Indicating that this could be a conserved structure within the AstV CP that may be present across the *Astroviridae* regardless of host species (Figure 3.3 C). Interestingly, the capsid core of HEV was identified as a significant structural homologue of ANV, consistent with previous data for HAstV[194]. Superimposition of the known HEV and HAstV-1 crystal structures (Figure 3.3 D) suggest the HEV core is less structurally similar to HAstV-1 than the predicted ANV core (Figure 3.3 C). The structural homology between the *Astroviridae* and HEV could suggest a deep evolutionary link between these groups. If the function of this core domain is also conserved then it may mean data from studies of HEV capsid assembly could be applicable to astroviruses.

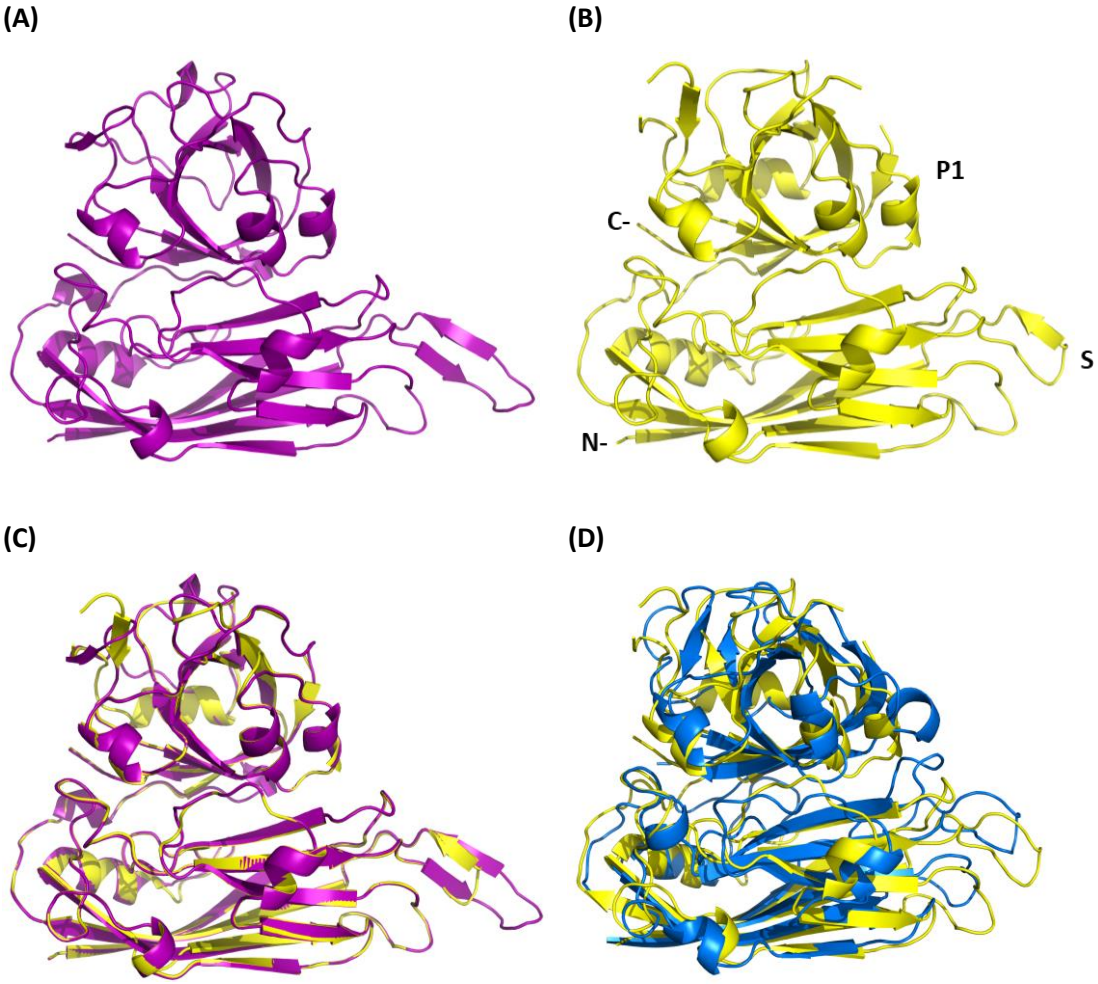


Figure 3.3: Structural comparison of ANV, HAstV-1 and HEV capsid core domains. (**A**) The structure of the ANV core domain was computationally predicted using SWISS-MODEL, based on (**B**) the known crystal structure of the HAstV-1 core (PDB ID: 5EWN). The lower half of the displayed structure, containing the jelly-roll β-barrel fold, forms the S domain, whereas the top half containing the squashed β-barrel fold makes up the P1 domain. (**C**) Superimposition of the predicted ANV capsid core and known HAstV-1 core. The structure alignment has 312 equivalent positions with a root-mean-square deviation (RMSD) of 0.99 Å, without twists, indicating highly similar positioning of the polypeptide backbone atoms. (**D**) Superimposition of the HEV capsid core (blue) onto the HAstV-1 core for structural comparison. The structure alignment has 292 equivalent positions with a root-mean-square deviation (RMSD) of 3.04 Å, without twists, indicating close positioning of the polypeptide backbone atoms.

3.2.2.4 The basic N-terminal region

In a number of viruses that utilise a PS-mediated assembly mechanism, such as STNV, HBV and HPeV1, there is a basically-charged unstructured region present at either the N- or C-terminal end of the CP[37, 51, 55, 57, 78]. The N-terminal region of astrovirus CP (approximately residues 1-70, depending on strain), whilst variable in sequence between strains, is abundant in basically-charged residues irrespective of host species. The N-terminal sequence of AstV CPs generally consist predominantly of repeating serine and arginine residues, however this region in ANV contains repeating glutamine and lysine residues (Figure 3.4). Because of the position of the basic N-terminal sequence in the core domain, it could suggest that this region forms unstructured extensions which may protrude into the capsid interior. This would be consistent with the involvement of this region in either electrostatic or sequence-specific RNA-CP interactions during encapsidation of genomic RNA by analogy to viruses that utilise a PS-mediated assembly mechanism[37, 51, 55, 57, 69, 78]. Furthermore, the basic N-terminal region is not proteolytically cleaved from the CP during particle assembly[231], and is therefore part of the AstV capsid shell, consistent with an important function in the mature virion, which was further supported by *in vitro* studies showing that complete or partial deletion of this region significantly reduced viral titre[231]. *In vitro* data also show that this basic region of the AstV capsid is dispensable for VLP formation[266], suggesting it is not essential for CP-CP interactions. Taken together these data may indicate that the N-terminal basic region of AstV CP has a role in genome packaging.

```
ANV       ---MAGGATAPAGAKPKQPKQKQKKPSSQARKKPS---------- 32
HAstV     ---MASKSNK--QVTVEVSNNGRNRSKSRARSQSR----GRDKSV 36
TAstV     MAAMADKVVVKKTTT-----RRRGRSNSRSRSRSRSRTK-KTV 39
CAstV     ---MADKAVVVKKTTIQKPRRARQRSSSRGRSRSRSRSRVRKV 42
             **.       ..     . : : .*:.*.:
ANV       ------QKQKAMKPVKQELR-------KVEKQVRVLKARTNGPKV 64
HAstV     KITVNSRNRARRQPGRDKRQSSQRVRNIVNKQLR--KQGVTGPKP 79
TAstV     KIIEKKPEKSILKKIDQAER-------RDAKQLRRIRKKVQGPPV 77
CAstV     KIVETAPKPILK--VRPKQT-------KILRKIRKLERKTNGPLV 78
                  :                     :::*  .  . **
```

Figure 3.4: Sequence homology of the basic N-terminal arm between AstVs. Clustal Omega alignment of the N-terminal regions from ANV, HAstV, TAstV and CAstV. Positively charged residues (arginines and lysines) are highlighted in blue. '*' denotes full conservation of the amino acid at that position, ':' for conservation of strongly similar residue properties, and '.' for conservation of residues with weakly similar properties. Sequences used as listed in Table 3.1.

3.2.3 The interior surface of ANV CP is highly basic

As RNA molecules have a high negative charge, it may be hypothesised that regions of the CP in close contact to PSs will have an overall positive charge in order to counter electrostatic repulsion at these sites. To investigate this, the surface electrostatic potential of the modelled ANV capsid core structure was calculated with PyMOL[305]. Modelling showed a strongly basic character on the surface of the jelly-roll β-barrel fold within the lower half of the displayed structure (S domain) (Figure 3.5), which is located on the interior surface of the capsid, as determined by solved cryo-electron microscopy structures of HAstV[194, 220]. This positioning of the basic surface could indicate that it is involved in close-range interactions with genomic RNA, such as acting to stabilise the backbone of RNA within stem-loop structures, though the electrostatics analysis data shown here does not conclusively identify a putative PS-CP interaction site.

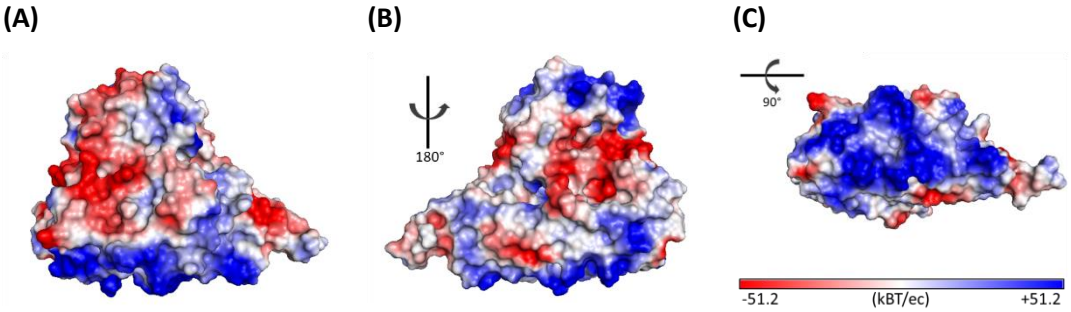


Figure 3.5: Prediction of surface electrostatic potentials in the predicted ANV capsid protein structure. The surface electrostatic potential of the predicted ANV capsid core domain was calculated using PyMOL and superimposed on views of the capsid protein monomer presented from different angles. (**A**) and (**B**) show the capsid core side-on rotated through 180° to show the front and back of the same monomer, respectively. (**C**) View of the capsid core from the perspective of the capsid shell interior (A rotated 90°). Surface charge: Blue = positive, White = neutral, Red = negative.

3.2.4 Predicted ligands of the ANV capsid protein

Earlier studies on HAstV CP assembly identified the requirement of Mg^{2+} in the formation of HAstV VLPs[266]. Therefore, as predicted structures of the ANV capsid core and spike proteins had been modelled, the COACH webserver[304] was used to computationally predict any

possible ligands, by comparison of binding-specific substructures from protein function databases, that may interact with these structures in order to identify any key molecules that may be involved in ANV CP assembly. Potential ligands of the core domain include Zn^{2+}, SO_4^{2-} and Ca^{2+} ions, albeit with low C-scores (0.13-0.14 out of 1) indicating a poorer prediction reliability and thus may not be meaningful. In comparison, when the known HAstV-1 core structure was analysed, only Ca^{2+} was identified as a divalent cation ligand (C-score of 0.20). Next, the structures of the predicted ANV and known HAstV-1 spikes were examined by the COACH webserver, Mn^{2+}, Mg^{2+} and Ca^{2+} were identified as potential ligands for both of the structures (C-scores between 0.16-0.23), indicating that these may be more accurate predictions than those for the core domain. These data could suggest that the CP of ANV may follow a similar divalent cation (i.e. Mg^{2+}) promoted mechanism of VLP assembly to that of HAstV-1[266], even with the highly dissimilar spike domain topology, and it is plausible that ANV may even utilise different ions to HAstV for capsid assembly and stability, such as the Zn^{2+} or SO_4^{2-} identified here.

Unexpectedly, the highest scoring result from binding-specific substructure comparison to the modelled ANV core was for a nucleic acid ligand, in the form of an RNA hairpin binding to the highly basic inner surface of the ANV CP (Figure 3.6). This prediction was based on the crystal structure of panicum mosaic virus (PMV) (PDB ID: 4V99), solved by X-ray diffraction analysis, where icosahedrally ordered genomic RNA hairpins ~17 nts in size were identified bound at sites across the twofold A-B subunit interfaces on the interior of the capsid[34]. To confirm the validity of the prediction, the structural homology of the PMV CP and modelled ANV capsid core were investigated (Figure 3.7). Comparison of the two proteins revealed high levels of structural similarity, with the PMV CP appearing to be a structural equivalent to the astroviral CP S domain. Moreover, high homology in structure at the site of genomic RNA interaction is observed between the ANV and PMV CPs, suggesting these distinct viral proteins could share a conserved function.

This data, alongside with the surface electrostatic predictions, further support the hypothesis that this region of the ANV capsid may be important for the sequence-specific binding of viral RNA, and therefore that ANV may use a PS-mediated assembly mechanism for encapsidating its genome.

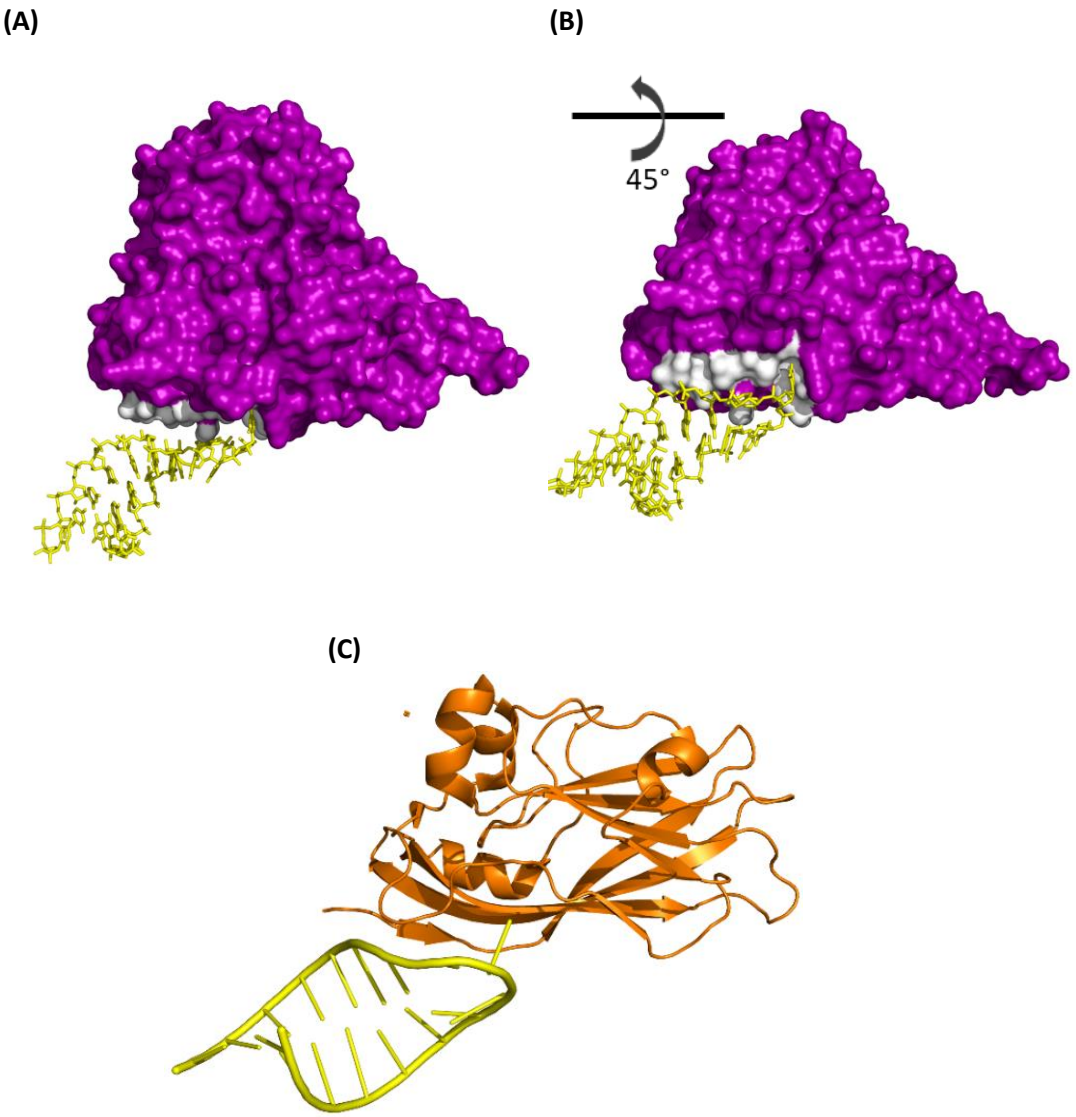


Figure 3.6: Predicted RNA ligand and binding site for the ANV capsid core. (**A**) Model of the predicted ANV capsid core-RNA complex, extrapolated from the solved crystal structure of PMV virions. RNA (atomic model) is in yellow and protein surface is shown in purple. (**B**) rotated 45°. With predicted RNA binding residues corresponding to N65, T67, K69, L122, S123, V124, E125, L179, R240 and A244 shown in white. The confidence score (C-score) of the prediction is 0.33, the C-score ranges from 0-1, where a higher score indicates a more reliable prediction. (**C**) The solved crystal structure of the PMV CP-RNA complex[34] (PDB ID: 4V99), identified as a ligand-binding structural homolog to the modelled ANV capsid core. RNA is shown in yellow with a generic sequence of $U_2A_2UAU_5AU_5$, derived from the best fit to the electron density maps, from which COACH used as a template model.

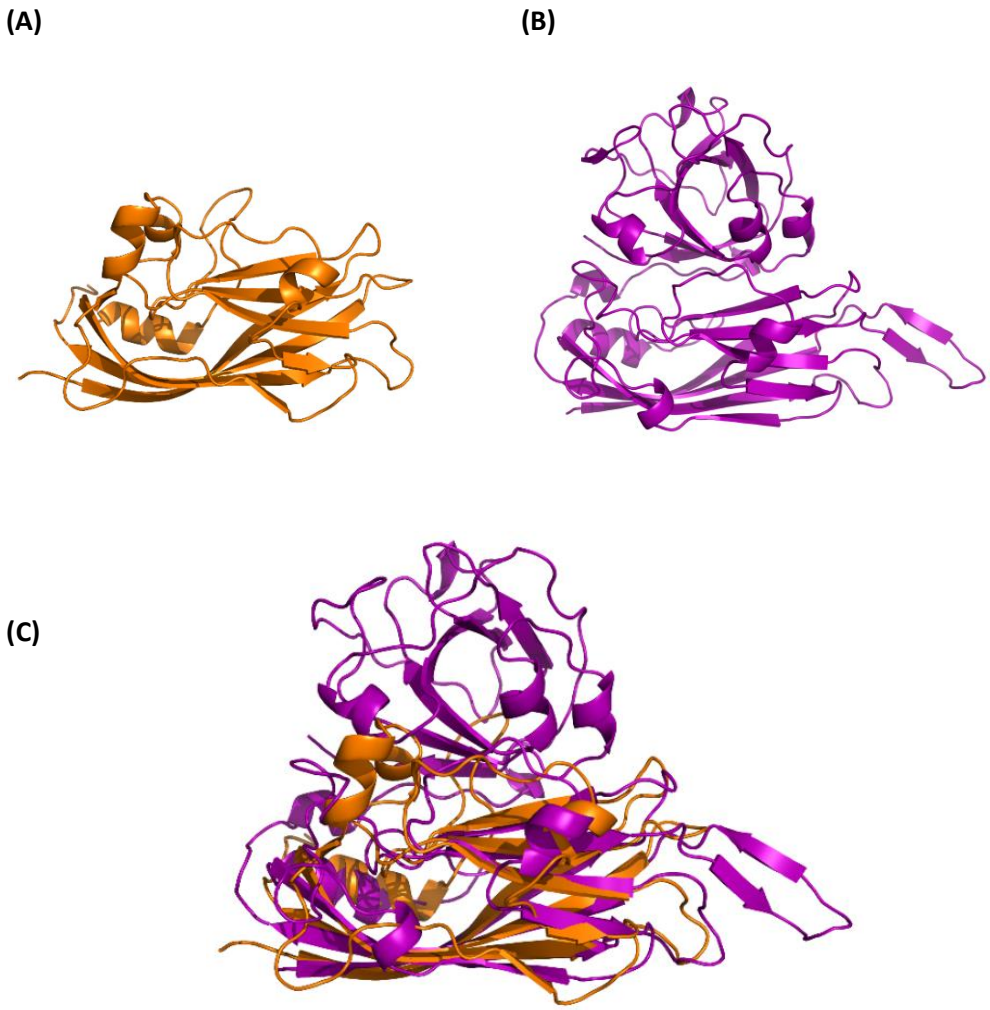


Figure 3.7: Structural comparison of the PMV and ANV core domains. (**A**) The known crystal structure of the PMV capsid monomer (PDB ID: 4V99). (**B**) The modelled ANV capsid core. (**C**) Superimposition of the PMV and ANV CPs. The structure alignment has 159 equivalent positions with a root-mean-square deviation (RMSD) of 3.06 Å, without twists, indicating close positioning of the polypeptide backbone atoms. The PMV CP appears to closely resemble that of the astrovirus core S domain, with a high structural homology on the capsid interior.

3.3 Summary

3.3.1 Predicted structure of the ANV capsid protein

Collectively, the computational homology modelling of the ANV CP presented here has identified the relative locations of the core and spike domains within ORF2, predicted their respective tertiary structures and indicated putative ligands, most importantly of which is an RNA SL interaction located at a region of basic charge on the surface of the core domain.

The availability of solved crystal structures for the human and turkey astrovirus structural proteins enabled homology modelling of the ANV CP, revealing that the capsid structure of ANV is likely to be similar to that of the other *Astroviridae* (Figure 3.2 and Figure 3.3). Whilst the ANV capsid core shows high structural homology to that of the known HAstV-1 core, it may be expected that this domain would show a greater level of similarity to those of the other avian astroviruses, such as TAstV-2 or CAstV, based on sequence homology (Table 3.1). However, no structures for the capsid core protein of either TAstV or CAstV have been produced to verify this hypothesis. There are limitations to using homology modelling however, as absence of structural templates with significant sequence identity can lead to predictions using sub-optimal templates and result in potentially unreliable data. These issues might be the case for the predicted ANV spike structure presented here as the sequence and structure of this domain is hypervariable within *Astroviridae*[225, 226].

The modelling data here also support the hypothesis that the *Astroviridae* may utilise a PS-mediated encapsidation mechanism, as an RNA SL structure was mapped to the basic surface regions of the ANV CP from substructure comparison of known ligand-binding proteins (Figure 3.6). The ligand interaction was derived from studies on the crystal structure of PMV, which identified the presence of multiple ordered RNA stem-loops closely associated with the CP 2-fold subunit interface[34]. Although specific PSs have not been identified in the genome of PMV, the structural data for this virus strongly indicate that it may utilise a PS-mediated assembly strategy, by analogy to MS2, STNV and HPeV1[37, 313, 314]. Furthermore, the high structural homology between the capsid cores of HAstV and ANV to that of the PMV CP could indicate that there may be conserved elements of PS-mediated assembly between these distinct viruses, such as the CP residues involved in PS interaction, or even the PS stem-loop motif sequences used.

The unstructured N-terminal extensions found in the AstV CP may also be important for PS-mediated assembly of viral particles by analogy to the CP of viruses which use this method of genome packaging. For example, these extensions may be involved in recognition of

sequence and structure-specific motifs within the viral RNA, creating a switch-like mechanism which promotes capsid assembly or structure stabilisation upon PS binding, as seen for STNV and HPeV1, respectively[51, 59]. However, data for HAstV VLP assembly in insect cells suggest that interaction of the CP basic N-terminal region with PSs may not be essential for capsid stability, as VLPs were readily assembled from both full-length and N-terminally truncated forms of CP[266]. Therefore, PS-mediated neutralization of positive charge in the ANV CP N-terminal arms may act to overcome electrostatic repulsion between the CPs that restricts their higher-order structure formation, as seen for STNV[51, 52]. Alternatively, the N-terminal extensions may help facilitate the organisation of RNA within the capsid so that the genome is packaged correctly, as hypothesised for HBV[55]. Regardless of exact function the AstV N-terminal region plays, it appears to be important for the production of infectious particles, as viral titre of HAstV is significantly reduced upon mutation of this region, implicating its role in an essential function of the viral life-cycle, such as genome encaspidation[231]. The high sequence variation in the basic arm between astroviruses (Figure 3.4) indicates that if this CP region interacts with genomic RNA in a structure- or sequence-specific manner during encapsidation, then the PSs utilised may also vary slightly by strain, a feature which is observed between the related MS2 and GA bacteriophages[62].

In summary, the results presented here for the identification and modelling of structural domains within the ANV CP will inform cloning of *Avastrovirus* ORF2 and its sub-regions into expression systems for production of CP, as well as help guide subsequent experiments on the assembly of AstV VLPs for use in PS-identification studies, such as RNA SELEX.

Chapter 4

4 Heterologous expression of avian astrovirus ORF2 capsid protein for virus-like particle formation

4.1 Control and prevention of viruses

Current control measures for astroviruses are limited, and elimination of environmental contamination is time consuming, expensive and difficult to achieve[189]. Control of viruses, both human and veterinary, is commonly achieved by vaccination with live attenuated, inactivated viruses, or non-infectious VLPs[315-318] which consist of viral capsid proteins in the conformation of the authentic virus, but lack the viral nucleic acid[319-321]. VLPs can be produced in a broad range of expression systems, including plants, bacteria, mammalian and insect cell lines[322]. VLPs are attractive vaccine candidates as they are non-replicative, cheap to produce, highly immunogenic, and able to provide protection against wild type virus[320].

However, a recently developed experimental CP-based vaccine (non-VLP) against a novel strain of CAstV only provided partial maternally-derived protection to chicks challenged with the virus[187]. The antigenic diversity between AAstVs would greatly limit the generation of cross reactive neutralising antibodies by vaccination[96, 180, 242, 244, 256], and a unique enterotoxin-like activity of the AstV capsid, identified from reports of VLP-induced pathology[242, 244], may further restrict the use of VLPs to vaccinate against AstVs. Therefore, novel means of control are needed to prevent AstV disease in poultry, which could potentially be achieved by targeting a highly conserved essential function of the virus, such as capsid assembly or genome encapsidation, which is likely to be a conserved mechanism within a virus group[62, 323].

Whilst AstV VLPs may be unsuitable for vaccination, they could however be used to improve our understanding of AstV capsid assembly and genome packaging through the use of VLPs or their assembly intermediates to perform identification of RNA-encoded PSs by RNA SELEX[53, 59]. It is likely that AstVs exploit this encapsidation mechanism as they lack identifiable genome packaging machinery and share features common in other +ssRNA viruses which utilise PS-mediated assembly[53, 55, 58, 59, 65]. Thus, to improve understanding of AstV genome packaging and capsid assembly, which could provide novel avenues for future

control strategies, a CP expression platform for ANV was established and optimised to provide VLPs for RNA SELEX screening to identify PSs within the ANV genome.

4.2 Design of astrovirus capsid expression constructs

Regions of *Avastrovirus* ORF2 capsid genes, from ANV G-6240 and CAstV 612, 11672, VF07 and VF08, were cloned into specialised expression vectors to test the expression of recombinant ORF2 CP in multiple systems. Cloning of truncated ORF2 regions was guided by homology with the solved crystal structure of the HAstV capsid core (Section 3.2.1) and similarities to the CPs of other viruses utilising PS-mediated assembly. One such similarity is the presence of a highly basic N-terminal arm in *Avastrovirus* CPs, a feature found in many CPs of viruses which use PS-mediated assembly, leading to the hypothesis that this region of the CP may be important for genome encapsidation. To ensure that the production of recombinant CP for PS studies was as similar to wild type CP as possible, truncations and epitopes for detection and purification were positioned to avoid interference with the structure of the N-terminus.

4.3 Expression and purification of avian astrovirus capsid protein

4.3.1 Plant-based expression system (*Nicotiana benthamiana*)

As plant-based expression systems have previously been used to successfully express large quantities of viral CPs capable of VLP formation[306, 324, 325], an *N. benthamiana* codon optimised variant of the full-length capsid gene of CAstV 11672 was cloned into pEAQ to produce pEAQ-CAstV-11672. In this construct, an in-frame 6-histidine (6His) epitope was introduced at the C-terminus of ORF2 for purification and detection of the expressed protein. pEAQ-CAstV-11672 was introduced into *N. benthamiana* leaves by agroinfiltration, and expression of the C-terminally His-tagged CAstV 11672 CP was assayed over a period of 7 days post-infiltration (Section 2.2.4.3). On successive days leaves were harvested and frozen prior to processing. At the end of the time course the harvested leaves were processed (Section 2.2.5.2) and the presence of CAstV 11672 CP in the resulting soluble and insoluble fractions was analysed by western blotting. From as early as 3 days post-agroinfiltration, a protein of > 75 kDa which reacted with the mouse anti-His antibody could be detected in the insoluble fraction (Figure 4.1). Whilst in the soluble fraction a similar band was only readily detected at 7 days post-agroinfiltration (Figure 4.1). This protein is consistent with the expected size of the CAstV 11672 capsid protein, at 81.6 kDa.

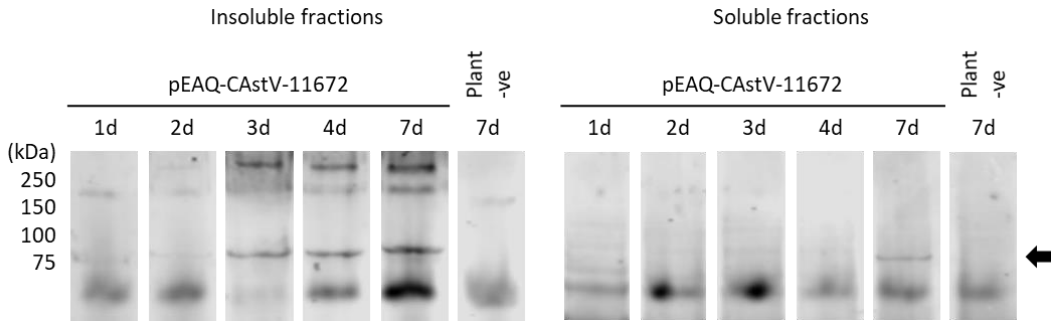


Figure 4.1: Detection of recombinant CAstV 11672 CP in *N. benthamiana*. pEAQ-CAstV-11672 was agroinfiltrated into leaves of *N. benthamiana* plants to express recombinant CAstV 11672 CP with a C-terminal 6His tag for purification. Leaves were harvested at 1, 2, 3, 4 and 7 days post-infiltration, then processed to yield insoluble and soluble fractions for western blotting. In the insoluble fraction an anti-His antibody reactive band of > 75kDa is readily detectable from 3 d post infiltration, whilst in the soluble faction this band is only weakly detected after 7 days post-infiltration. The detected band of > 75kDa is consistent with the expected size of the recombinant CAstV 11672 CP, which is calculated at 81.6 kDa.

Protein from the soluble fraction of agroinfiltrated pEAQ-CAstV-11672 plant lysate was purified using either nickel (Ni^{2+}) immobilised metal affinity chromatography (IMAC) (Section 2.2.5.3) or sucrose density gradient ultracentrifugation (Section 2.2.8.3). Nickel purification elution fractions were resolved by SDS-PAGE and stained with PageBlue. A doublet of bands were present at around ~60 kDa in the first elution fraction with a smear of smaller products below, which may be indicative of CP degradation or proteolytic processing to smaller units (Figure 4.2 A). This suggests that CP eluted from the nickel resin may have lost sequence from the N- or C-terminal regions, potentially including the hypothesised PS-interacting domain, rendering CP derived from this system unsuitable for RNA SELEX identification of PSs. Sucrose gradient purification gave rise to three bands of approximately 55, 65 and 70 kDa in the fourth fraction of the gradient which were absent from other fractions (Figure 4.2 B). This may once again be consistent with degradation or programmed proteolytic cleavage of CP, or the carry-over of contaminating cellular proteins rather than the presence of VLPs, as no His-tagged protein could be detected by western blotting. As neither purification method isolated the previously identified 81.6 kDa full-length CAstV 11672 capsid protein, alternate expression systems were trialled.

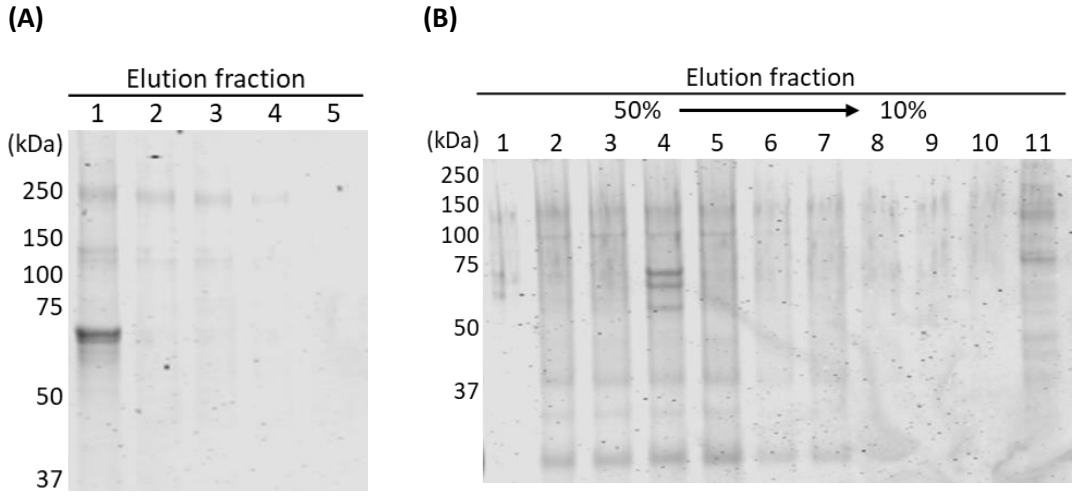


Figure 4.2: IMAC and density gradient purification of CAstV 11672 CP expressed in *N. benthamiana*. Recombinant CAstV 11672 CP was purified from the soluble fraction of agroinfiltrated *N. benthamiana* leaves 7 days-post infiltration. Elution fractions from (**A**) Ni^{2+} IMAC and (**B**) sucrose density gradient purification were resolved on 4-20% polyacrylamide gels, then stained with PageBlue. A cluster of bands of around ~60kDa were observed in the 1st Ni^{2+} IMAC elution fraction rather than the expected 81.6 kDa recombinant protein. Sucrose density gradient purification did not yield the expected 81.6 kDa recombinant CAstV 11672 CP, but three bands ranging between ~55-70 kDa in the 4th elution fraction.

4.3.2 Avian cell-based expression system (DF-1 cells)

For expression of *Avastrovirus* ORF2 in avian cells, the full-length capsid gene of ANV G-4260 as well as CAstV strains VF07 and VF08 were cloned into pCDNA3.1(+), with an in-frame C-terminal 6His tag (ANV) or HA epitope tag (CAstVs), respectively. An additional construct with an N-terminal 6His tag fused with ANV ORF2 was also produced. These expression plasmids were transfected (Section 2.2.4.1) into DF-1 cells to assess the possibility of using an immortalised avian cell line to produce correctly post-translationally modified ANV or CAstV CP.

Expression of the full-length CP of ANV containing either an N- or C-terminal 6His tag in DF-1 cells was assessed by western blotting at 72 hours post-transfection (Figure 4.3). Whilst the anti-His antibody detects a number of background bands present in the mock and transfected cells, there is a distinct band of around 75 kDa specifically detected in lysates of DF-1s transfected with pCDNA3.1(+) encoding N- or C-terminally tagged ANV ORF2. This product is consistent with the expected size of ANV CP at 73.9 kDa, suggesting that the

complete capsid gene can be successfully translated in a chicken cell line with a 6His tag at either terminus of the protein product. Interestingly, two additional bands of < 37 kDa were detected by the anti-His antibody in cells producing N-terminally tagged CP (Figure 4.3). These smaller proteins could represent proteolytically processed CP, corresponding to different forms of the core domain of the capsid, which is within the N-terminal region of ORF2 and linked to the 6His tag. The processed core domains of other astroviruses are around ~34 kDa in size supporting this hypothesis[291, 292]. No evidence of CP processing was detected with the C-terminally tagged expression construct. However, if processing of C-terminally tagged ANV CP has occurred, the apparent absence of smaller ORF2 proteins may be explained by cleavage of the 6His tag from any processed CP, which would prevent detection of these products. The CPs of HAstVs are processed at their C-terminus by host caspases which remove the C-terminal acidic region[238, 239], the corresponding region to which the 6His epitope is fused in the ANV construct. The presence of multiple putative caspase cleavage sites between the spike and acidic domains of the ANV CP supports this hypothesis (Figure 3.1), however, further investigation would be required[238].

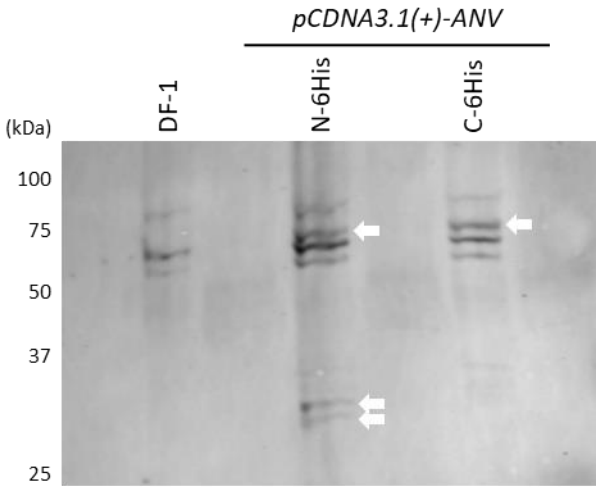


Figure 4.3: Detection of recombinant His-tagged ANV CP in DF-1 cells. Lysates of DF-1 cells transfected with expression constructs containing either N- or C-terminal 6His tag variations of ANV ORF2 were resolved by SDS-PAGE and recombinant CP detected by western blotting. For both N- and C-terminal 6His tag constructs, a specific anti-His antibody reactive band of around 75 kDa was detected which is consistent with the expected size of the recombinant ORF2 protein, at 73.9 kDa. Two bands of around 30 kDa are also present for the N-terminally tagged ORF2 protein. White arrows indicate ORF2-specific products.

Following the successful detection of recombinant ANV CP production in DF-1 cells, expression constructs for C-terminally HA-tagged ORF2 of CAstV strains VF07 and VF08 were transfected into DF-1 cells, and expression assessed at 48 hours post-transfection. Western blotting of whole cell lysates indicated that HA-tagged recombinant CP was expressed by the VF07 construct, but not the VF08 construct (Figure 4.4). Whilst HA-tagged protein was detected in cells transfected with the VF07 ORF2 construct, a 74.7 kDa band corresponding to unprocessed CP was not observed. Instead, a prominent band of ~30 kDa and 2 fainter bands of around 24 and 27 kDa were detected. Based on the location of the 6His tag, the 30 kDa band seen here likely represents the spike region of processed VF07 CP. This suggests that similarly to ANV CP, processing of the CAstV CP occurs within DF-1 cells. However, this result does not match that seen for C-terminally His-tagged ANV CP in DF-1 cells, suggesting there may either be differences in capsid processing between the AAstV strains, or in detection sensitivity of the antibodies used against the HA and 6His antigens.

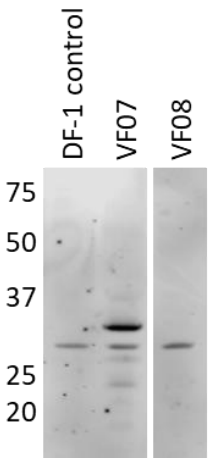


Figure 4.4: Detection of recombinant HA-tagged CAstV CP in DF-1 cells. Lysates of DF-1 cells transfected with expression constructs containing C-terminal HA-tagged CAstV VF07 and VF08 ORF2 were resolved by SDS-PAGE, and recombinant CP detected by western blotting. For the VF07 ORF2 construct, a specific anti-HA antibody reactive band of around 30 kDa was detected. However, no protein consistent with the expected size of the full-length CP was observed, and no HA-tagged protein was detected in DF-1 cells transfected with the VF08 ORF2 expression construct.

Collectively, these data suggest that expression of either ANV or CAstV ORF2 in DF-1 cells, a cell line derived from the host species of these viruses, may not be suitable for recovery of full-length recombinant CP. Processing of the recombinant CP within DF-1 cells resulted in truncated forms of the CP, for most, if not all constructs examined, akin to CAstV 11672 ORF2 expressed in *N. benthamiana* (Section 4.3.1). Therefore, any attempt to purify protein from this system would likely result in material that does not encompass the intact CP, potentially rendering it unsuitable for RNA SELEX identification of PSs.

4.3.3 Bacterial expression system (*E. coli*)

For expression in bacterial cells, the full-length ORF2 and a truncated version containing only the predicted N-terminal core domain were cloned into pET23a with a C-terminal 6His tag in each construct. Design of ORF2 truncation to isolate sequence encoding the CP core domain was done using alignment of the ANV capsid sequences. Alignment indicated high conservation from positions 1-391, with homology rapidly dropping after this region, suggesting that it may correspond to the capsid core domain of ANV. This domain prediction was further confirmed by homology modelling against the solved HAstV-1 capsid core[194], indicating that residues 65-391 were in-fact likely to contain the core domain (Section 3.2.1). Thus, the region encompassing residues 1-391 and the full-length ORF2 were cloned into pET23a (Figure 4.5 A and B).

Expression of the full-length and truncated ANV ORF2 within pET23a constructs was tested in BL21 pLysS Rosetta *E. coli* (Section 2.2.4.2). SDS-PAGE and western blot analysis show that protein of the expected size was detected in the cell lysate of transformed *E. coli*, corresponding to 43.9 and 73.9 kDa for capsid core and full-length CP, respectively (Figure 4.5 C). Levels of detection for the expressed constructs suggest that the core protein was produced in higher quantities than the full-length CP. The observed lower levels of full-length protein expression may be due to toxicity of the ANV spike region in bacteria, or improper CP folding within the prokaryotic cells leading to aggregation in bacterial inclusion bodies, or removal of the C-terminal His-tag by proteolytic processing rendering it undetectable with the antibody used.

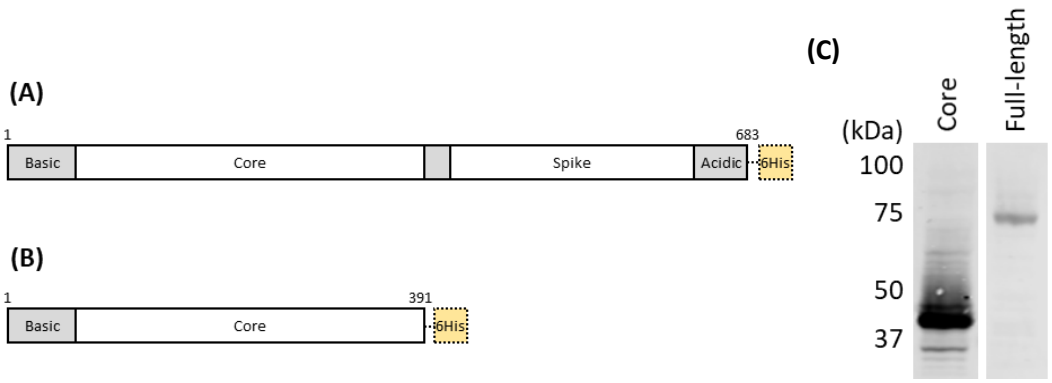


Figure 4.5: Expression of ANV ORF2 constructs in *E. coli*. *In silico* analysis of the ANV ORF2 sequence revealed that residues 1-391 correspond to the capsid core domain, determined by alignment and structure homology modelling against the known HAstV CP. Therefore, (**A**) the full-length ORF2 and (**B**) the region corresponding to the core domain were cloned into pET23a with a C-terminal 6His tag for expression of recombinant ANV CP in *E. coli*. Regions highlighted in grey indicate the unstructured regions flanking the core and spike domains, as determined by homology modelling against the known HAstV-1 (PDB ID: 5EWN) and TAstV-2 (PDB ID: 3TS3) crystal structures, respectively. (**C**) Expression of core or full-length ANV ORF2 CP in BL21 cells was induced using IPTG for 5 hours at 30 °C. The cells were pelleted and lysed by sonication, clarified lysates were then resolved by SDS-PAGE and recombinant CP detected by western blotting using antibodies targeting the 6His tag. Large quantities of protein corresponding to the 43.9 kDa capsid core were detected, whereas only low levels of the 73.9 kDa full-length CP were detected.

Having shown the production of recombinant ANV CP in bacterial cells, purification of the core and full-length CPs by IMAC using a cobalt resin (Co^{2+}) was attempted (Section 2.2.5.3). Bound recombinant protein was eluted after three sequential washes to the resin, with each collected as individual fractions. Eluted protein was analysed by SDS-PAGE and PageBlue staining to confirm purification of the expected products (Figure 4.6). A ~40 kDa band was evident in the lysate of cells expressing the ANV capsid core, which was present in the elution fractions post-IMAC. Similarly, a ~75 kDa band was detected in the clarified lysate of cells expressing the full-length CP, which was also present in the elution fractions post purification. The capsid core was noticeably more abundant than the full-length CP in the purified fractions, however basal expression of the capsid core was also higher which may account for the differences observed (Figure 4.5).

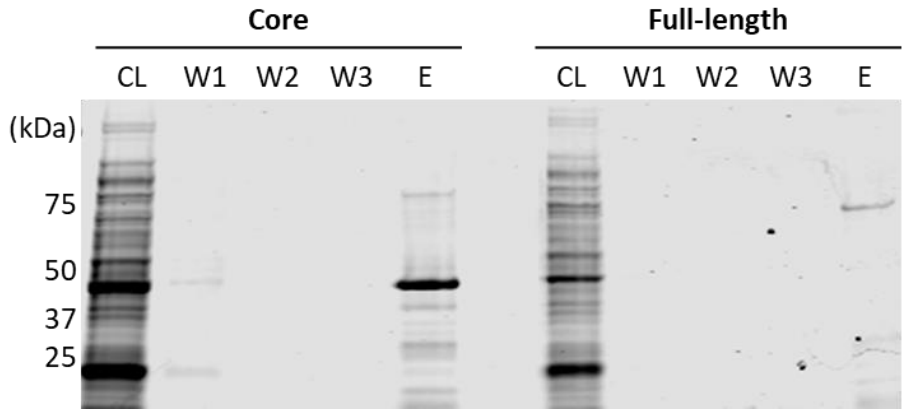


Figure 4.6: IMAC purification of recombinant ANV CP expressed in *E. coli*. Recombinant ANV CP was purified using IMAC from the lysate of BL21 cells induced to express either the core or full-length ORF2 constructs. Clarified lysates (CL) containing total cellular protein, the three washes (W1-3) and a single elution (E) fraction from Co^{2+} IMAC were resolved by SDS-PAGE for each construct and stained with PageBlue. Capsid core and full-length CP were successfully purified using IMAC. The eluted protein consists predominantly of species of the expected size for the core and full-length CP, 43.9 and 73.9 kDa, respectively.

To confirm that the ~40 kDa protein purified from the bacterial expression system was indeed a product of ANV ORF2, the band was excised and subjected to mass spectrometry analysis using nanoscale liquid chromatography coupled with tandem mass spectrometry. Peptide fragments from mass spectrometry indicated that this protein corresponded to the entire ANV capsid core region as expected (Figure 4.7).

MAGGATAPAGAKPKQPKQKQKKPSSQARKKPSQKQKAMKPVKQELRKVEKQVRVLKARTN
GPKVNDTMKTTVTVGTLVGQTQSGLNRQLRVSFNPLLMKSTEGGSTTPLSIRASMYEMWK
PLSVEIFATPLSGFSSVVGSVGFMVITLNGLEASADSIDTIKARRHQMALGRPYRLKLSA
RELAGPREGWWLVDTSEAPADAYGPAVDLMLAYATENLLGTSSGSTTSYTGTLWQVEMRV
TYAFSTYNPKPGLQTLVSQSITGGQTVTIQPSPDDGSLIMTTNSQQVLALLTPRVAGQRK
GKSQTIWAIAGSAVDAAATVLGPWGYLLKGGFWLVRLIFGGSSARNTTTRQYQIYPSVES
ALTDQPIFGNSTGTQSVTVPICHITEVVN

Figure 4.7: Mass spectrometry analysis of the IMAC-purified ANV capsid core protein. Recombinant ANV capsid core protein was expressed in BL21 cells, purified by Co^{2+} IMAC and products were resolved using SDS-PAGE. The protein band of the expected size for the ANV capsid core (43.9 kDa) was excised and subjected to mass spectrometry. Peptides consistent with the ANV CP were identified (highlighted in grey) which cover almost the entire ANV ORF2 core domain (residues 1-391), indicating that the purified protein corresponds to the intact capsid core of ANV.

4.3.3.1 Proteolytic cleavage of the ANV capsid protein

In order to inform VLP assembly studies, the capsid processing of ANV was investigated to examine whether a similar processing mechanism to that of the HAstVs is used, and thus to assess if it may be a requirement for VLP formation, purified recombinant full-length ANV CP was subjected to digestion with trypsin at various concentrations. It is well documented that the HAstV CPs undergo several cleavages by trypsin-like proteases during its maturation, drastically altering the capsid structure[193, 291, 292]. This processing appears to be highly important in HAstVs, as infectious titre is increased by over 10^5-fold after trypsin treatment[292]. However, trypsin does not appear necessary for infectivity of ANV in cell culture[211].

Treatment of purified ANV CP with trypsin (0.02%) for 1 hour at 37 °C lead to the emergence of a protein ~29 kDa in size, at higher trypsin concentrations a transition of the 29 kDa protein into one of ~27 kDa is seen (Figure 4.8). This ~27 kDa protein may represent a fully processed CP product that may be present in the mature capsid of ANV, akin to that seen for the HAstVs[291, 292], and these data indicate the ANV CP may follow a similar pattern of proteolytic processing to HAstV CPs.

To confirm the identity of the ~27 and ~29 kDa fragments as cleavage products of trypsin-treated ANV CP, bands corresponding to these fragments were excised from the gel and subject to mass spectrometry analysis. The mass spectra of these two fragments contained peptides corresponding to the C-terminal spike region of the ANV capsid protein, confirming they are cleavage products of trypsin-treated ANV CP. The peptides corresponding to the ~27 kDa product band suggest this has an N terminal truncation relative to the ~29 kDa cleavage product (residues 463-645 and 363-636, respectively), suggesting there may be multiple processing sites to release the spike region from the full ANV CP (Figure 4.9). This observation is consistent with data from TAstV-2 where pepsin-treatment of full-length CP produced a 26 kDa protein product corresponding to the spike domain of TAstV-2 CP[221]. The trypsin cleavage products seen here for ANV CP appear to follow the same pattern as for HAstV-1 CP, where two proteins are produced with the larger containing the spike domain and a partial core region, and the smaller containing only the spike domain[198, 292]. Interestingly, however, no protein equivalent of the ~34 kDa capsid core region produced by trypsin-treatment of HAstV particles[193, 291, 292] was observed after trypsin treatment of ANV CP. This may be due to the fact that the ANV CP used here

was not in an assembled virion form and thus trypsin cleavage sites within the core region were exposed allowing complete digestion of this domain.

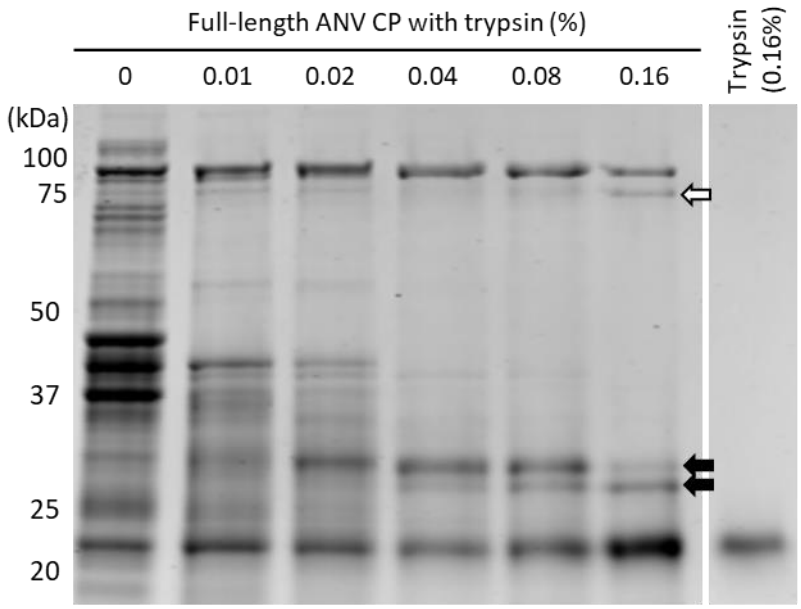
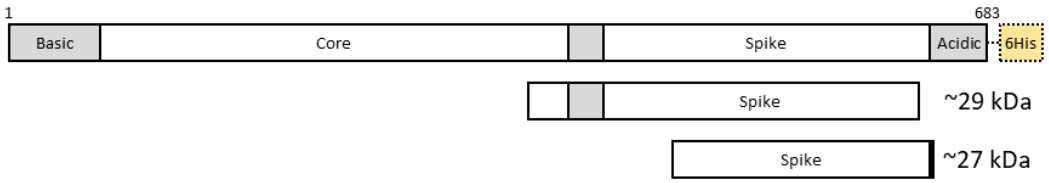


Figure 4.8: Trypsin digestion of full-length ANV capsid protein. Full-length ANV CP was purified by Co^{2+} IMAC from the lysate of BL21 cells expressing ANV ORF2. Purified CP was treated with increasing concentrations of trypsin to assess if the proteolytic processing of ANV CP was similar to HAstV CPs (Figure 1.6). Two bands of protein ~27 and ~29 kDa in size (indicated by black arrows) become apparent as the trypsin concentration is increased above 0.04 %, with the ~27 kDa protein becoming more prominent when trypsin concentration is increased to 0.16 %. A third band of ~75 kDa is also seen upon digestion with 0.16 % trypsin and may represent intact ANV CP (indicated by white arrow).

(A)

MAGGATAPAGAKPKQPKQKQKKPSSQARKKPSQKQKAMKPVKQELRKVEKQVRVLKARTN
GPKVNDTMKTTVTVGTLVGQTQSGLNRQLRVSFNPLLMKSTEGGSTTPLSIRASMYEMWK
PLSVEIFATPLSGFSSVVGSVGFMVITLNGLEASADSIDTIKARRHVQMALGRPYRLKLS
ARELAGPREGWWLVDTSEAPADAYGPAVDLMLAYATENLLGTSSGSTTSYTGTLWQVEMR
VTYAFSTYNPKPGLQTLVSQSITGGQTVTIQPSPDDGSLIMTTNSQQVLALLTPRVAGQR
KGKSQTIWAIAGSAVDAAATVLGPWGYLLKGGFWLVRLIFGGSSARNTTTRQYQIYPSVE
SALTDQPIFGNSTGTQSVTVPICHITEVVNPNAESNNLPPPTTGAQPQPQPPAPIEEILL
PLAELTGQPGVPPLYTFDGSSYTPPTNWLGSTILLTGIPAHKRVTGNLAKFGVTNLQMSK
VAATALEIYDFTDFGVFFGTGSYLSEGGIHTGKTLIYSLMSGQTPNPWLAANQSGTTWYM
PSWAGFPQPGQGDYFLQMQDVTDTTTHTTSVNVYFLVAYRQSRRLIAFFNTGGTARPAPT
SMLCLYNVDCGRAPQTPYPTFQSTLQSLNQIGVDAKSDPDSDDDISLAGSVIGDEFDSVD
HLEREREDLMRRLRDLDLRRFQI

(B)



(C)

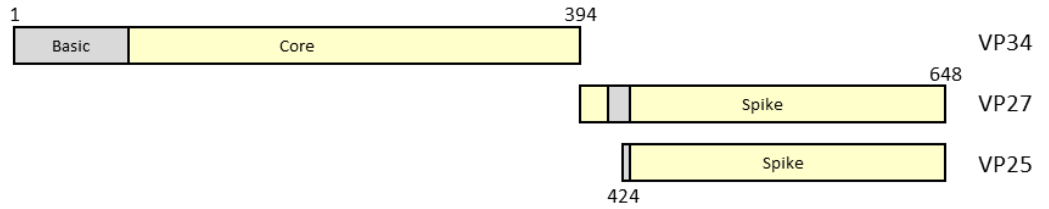


Figure 4.9: Mass spectrometry analysis of the trypsin-treated full-length ANV CP. Two products of tryptic digest on full-length ANV CP (Figure 4.8) were excised and subjected to mass spectrometry, corresponding to the bands ~29 and ~27 kDa in size. (**A**) Peptides corresponding to the ANV CP in the 29 and 27 kDa bands were identified (highlighted in grey, and outlined in red, respectively). (**B**) The identified peptides mapped to the predicted C-terminal spike region of the ANV CP, indicating the ~27 and ~29 kDa proteins represent processed forms of the ANV spike domain. (**C**) Schematic summary of the proteolytic processing pathway of HAstV-1 CP which follows a similar digestion pattern to that observed for ANV with three processed forms of CP found in mature HAstV particles. Figure adapted from[198, 292].

4.4 Formation of virus-like particles from purified ANV capsid protein

4.4.1 ANV capsid proteins form structural intermediates of capsid assembly

As purification of the recombinant capsid core and full-length ANV CPs from bacterial cells had been achieved, the ability of these proteins to form VLPs or structural intermediates of capsid assembly was evaluated by negative staining and transmission electron microscopy (TEM). In TN buffer, both the ANV capsid core and full-length CPs did not form any visible structures (Figure 4.10 A and B). Previous work had identified the importance of magnesium ions in HAstV VLP assembly[266], and so 10 mM $MgCl_2$ was added to TN buffer to investigate if ANV VLP formation required divalent cations. Addition of 10 mM $MgCl_2$ to TN buffer promoted the formation of non-uniform ~16 nm structures by both the core and full-length recombinant ANV CPs (Figure 4.10 C and D). However, no ~30 nm spherical virion-like structures were formed. Increasing concentrations of $MgCl_2$ (15 mM) did not alter the appearance of structures formed by the capsid core domain (Figure 4.10 E). Whilst in contrast, the 16 nm structures formed by full-length CP generally had more pronounced morphologies when TN buffer was supplemented with 15 mM $MgCl_2$ (Figure 4.10 F). Similar structures were formed by purified HAstV-1 capsid protein and were hypothesised to represent capsomers or T=1 VLP structures[266]. Assembly of these HAstV-1 capsomers into VLPs was magnesium dependent, and reversible upon addition of chelating agents[266]. The requirement for magnesium to promote ANV capsomer formation may also suggest that a similar mechanism is used by ANV. However, there are differences between ANV and HAstV-1, whilst 10 mM $MgCl_2$ was sufficient to promote VLP formation by HAstV-1 capsomers, only formation of capsomer-like structures was seen with the ANV CP and higher concentrations of $MgCl_2$ did not promote VLP formation by these ANV capsomers. In conclusion, these observations suggest that ANV may require different conditions to HAstV in order to promote capsid assembly by purified CP.

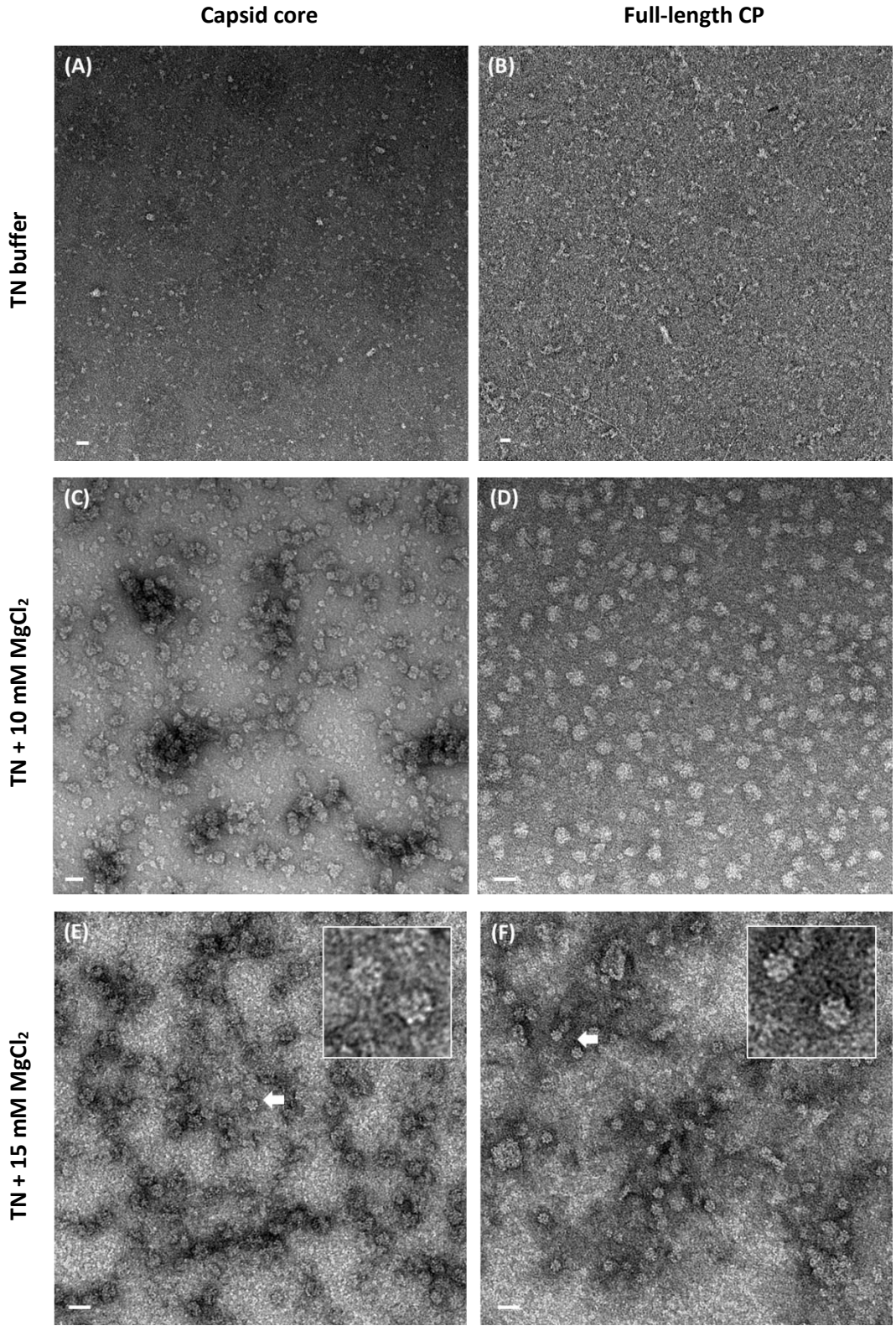

Capsid core
Full-length CP
TN buffer
TN + 10 mM MgCl2
TN + 15 mM MgCl2
(A)
(B)
(C)
(D)
(E)
(F)

Figure 4.10: $MgCl_2$ promotes the formation of 16 nm capsomer-like structures by purified recombinant ANV CPs. Structure formation by recombinant ANV capsid core and full-length CP purified over a range of $MgCl_2$ concentrations was assessed by negative stain TEM. In TN buffer (no $MgCl_2$), VLPs or assembly intermediates do not readily form by either (**A**) the core domain or (**B**) full-length ANV capsid protein. In the presence of 10 mM $MgCl_2$ ~16 nm non-uniform structures are formed by both the (**C**) ANV capsid core and (**D**) full-length CP. (**E**) In TN buffer containing 15 mM $MgCl_2$, the capsid core did not show altered morphology compared to 10 mM $MgCl_2$. (**F**) Whilst full-length CP formed more uniform ~16 nm structures in 15 mM $MgCl_2$. Scale bars = 25 nm.

4.4.2 Divalent cations are required for ANV capsomer stability

The requirement of $MgCl_2$ for the stability of capsomer-like structures formed by the purified full-length ANV CP was further confirmed by addition of chelating agents to samples prepared with $MgCl_2$. Addition of EDTA to a suspension of 16 nm ANV capsomer-like structures (purified in the presence of 10 mM $MgCl_2$) resulted in the loss of these structures, to resemble preparations of capsid protein purified in the absence of $MgCl_2$ (Figure 4.11). This result indicates that $MgCl_2$, or perhaps other divalent cations, are important for the formation and stability of these 16 nm structures.

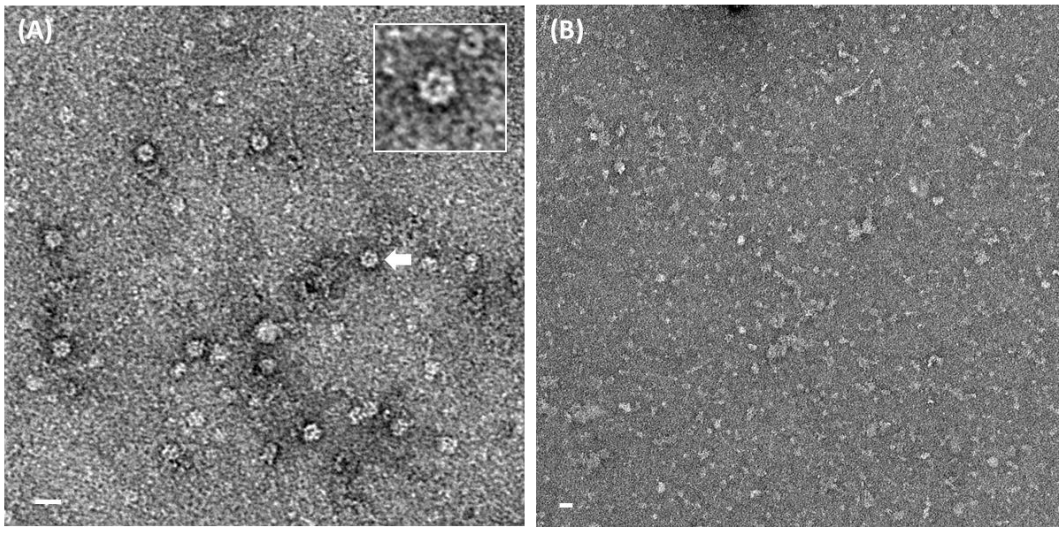


Figure 4.11: $MgCl_2$ is required for ANV capsomer-like structure stability. Stability of capsomers formed by recombinant full-length ANV CP in the presence or absence of EDTA was assessed by negative stain TEM. (**A**) Full-length recombinant ANV CP purified in TN buffer containing 10 mM $MgCl_2$ formed capsomer-like structures of ~16nm diameter. (**B**) These cog-like structures were subsequently disassembled by addition of EDTA to a final concentration of 10 mM. Scale bars = 25 nm.

As analysis of ANV CP preparations by TEM indicated that magnesium ions promote capsomer formation of the ANV CP, the ability of other divalent cations to substitute for Mg^{2+} ions was assayed. As *In silico* modelling suggested that Mg^{2+}, Zn^{2+} and Ca^{2+} ions could all potentially bind to various pockets on the ANV CP (Section 3.2.4), full-length CP was purified in TN buffer containing calcium to assess if ionic strength, or the presence of specific ions was responsible for formation the capsomer-like structures.

TEM examination of full-length ANV CP purified in TN buffer or TN containing calcium (10 mM $CaCl_2$) revealed the formation of highly irregular structures by ANV CP when Ca^{2+} ions were present (Figure 4.12). These structures do not show the cog-like morphology of those produced in the presence of Mg^{2+} (Figure 4.11 A). The structures formed in the presence of Ca^{2+} potentially represent aggregated protein rather than correctly formed assembly intermediates. This data indicates that Mg^{2+} may be preferred for assembly of ANV CP into capsomers, possibly due to its size, to maintain an assembly competent protein conformation that may be disrupted by larger ions such as Ca^{2+}.

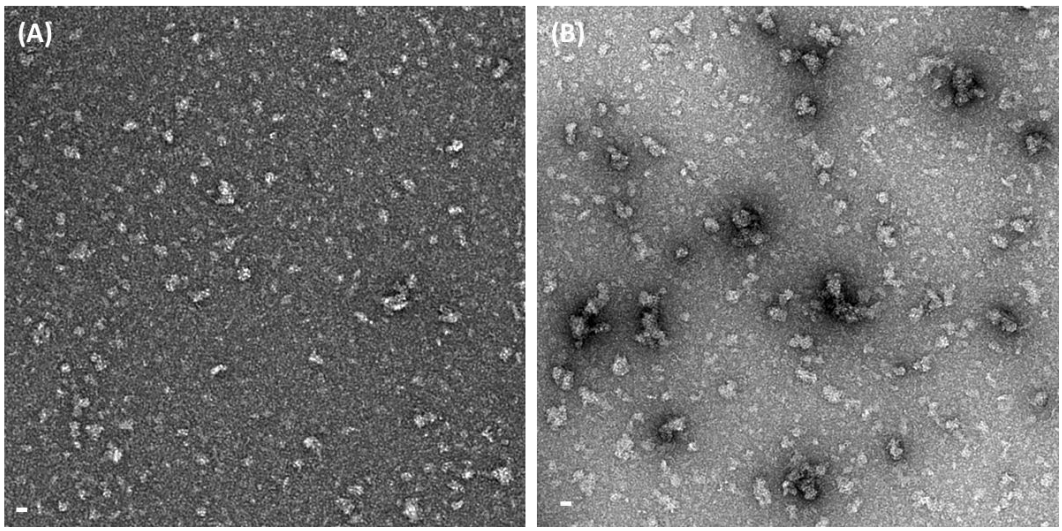


Figure 4.12: Calcium promotes less consistent ANV CP assembly than magnesium. Structure formation by recombinant full-length ANV CP purified with or without $CaCl_2$ in TN buffer was assessed by negative stain TEM. (**A**) In TN buffer with no $CaCl_2$, VLPs or assembly intermediates do not readily form. (**B**) In the presence of 10 mM $CaCl_2$, irregular non-uniform structures are formed. However, these do not resemble the cog-like morphology previously observed for CP in the presence of $MgCl_2$. Scale bars = 25 nm.

4.4.3 Proteolytic processing of ANV ORF2 may promote VLP formation

Proteolytic processing of HAstV capsid protein by trypsin-like proteases is required for virion maturation, with the cleavage of capsid protein into 3 subunits and release of 60 of the 90 dimeric spikes from the surface of the capsid shell. As recombinant ANV CP showed similar tryptic processing to HAstV CPs (Figure 4.8 and Figure 4.9), and that processing of HAstV CP by trypsin-like proteases is required for virion maturation, proteolytic processing may therefore be important or essential for proper ANV virion formation. To test this hypothesis, preparations of purified ANV CP in TN buffer were treated with trypsin at a final concentration of either 0%, 0.05% or 0.25% for 30 minutes at 37 °C. Electron microscopy showed the presence of irregular spherical particles ~30 nm in size after tryptic digest (Figure 4.13), indicating that proteolytic processing of the full-length recombinant CP may promote VLP-like structure formation, even in TN buffer lacking magnesium ions.

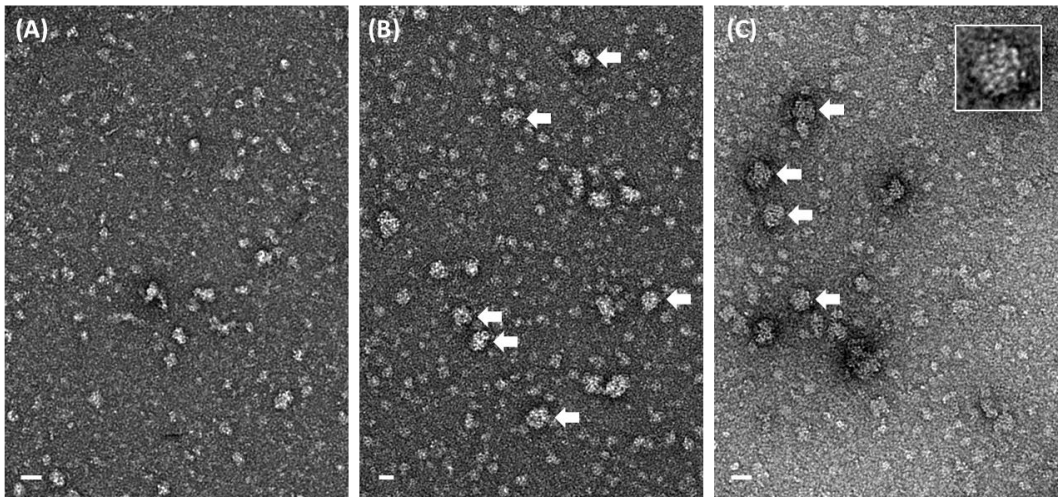


Figure 4.13: Proteolytic processing of full-length ANV CP promotes formation of 30 nm VLP-like structures. Structure formation by purified recombinant full-length ANV CP treated with trypsin in the absence of divalent cations was assessed by negative stain TEM. (**A**) In TN buffer without trypsin, CP structures do not readily form. (**B**) In the presence of Trypsin (0.05%) irregular spherical particles ~30 nm in size are formed. (**C**) At a higher trypsin concentration (0.25%) similar 30 nm particles are observed, suggesting the protein structures present are stable and resistant to further proteolysis. White arrows indicate ~30 nm particles. Scale bars = 25 nm.

4.4.4 ANV genomic RNA does not promote assembly of capsomers into virions

Virion assembly in +ssRNA viruses that use a PS-mediated assembly mechanism, requires, or is promoted by sequence-specific genomic RNA-CP interactions[19, 51, 52, 54, 55]. To investigate if ANV CP requires genomic RNA in order to promote capsid assembly, preparations of capsid core or full-length ANV CP were incubated with *in vitro* transcribed RNA from a genome-length cDNA clone of ANV (Section 2.2.1.6). In addition to RNA, NaCl in the TN buffer was replaced with Na_2SO_4 to assess if increased ionic strength would promote assembly of ANV CP, as VLPs of HEV, a structurally similar virus[194, 326], are formed more efficiently in the presence of Na_2SO_4 than with NaCl[327].

Irregular structures were seen in capsid core and full-length CP purified in Na_2SO_4 (Figure 4.14 A and B). Many capsomer-like structures, akin to those seen previously, were present in the capsid core preparation, indicating that the higher ionic strength may have a similar effect to Mg^{2+} on the CP of ANV, though only for the core domain. Addition of genomic RNA to purified ANV CP did not result in any observed morphological change in the apparent structure of either the capsid core or full-length ANV CP when compared to CP in the absence of RNA, and no virus particles were observed (Figure 4.14 C and D). The failure of genomic RNA to promote capsid assembly could indicate that the CP itself may be assembly incompetent, or require specific ions, post-translational modifications, or appropriate proteolytic processing. Alternatively, formation of the ANV capsid may only occur co-transcriptionally, with assembly of CP on nascent genomes as they are transcribed by the viral RdRp in an assemblysome, as has been reported for other viruses[59, 67, 68, 70-72].

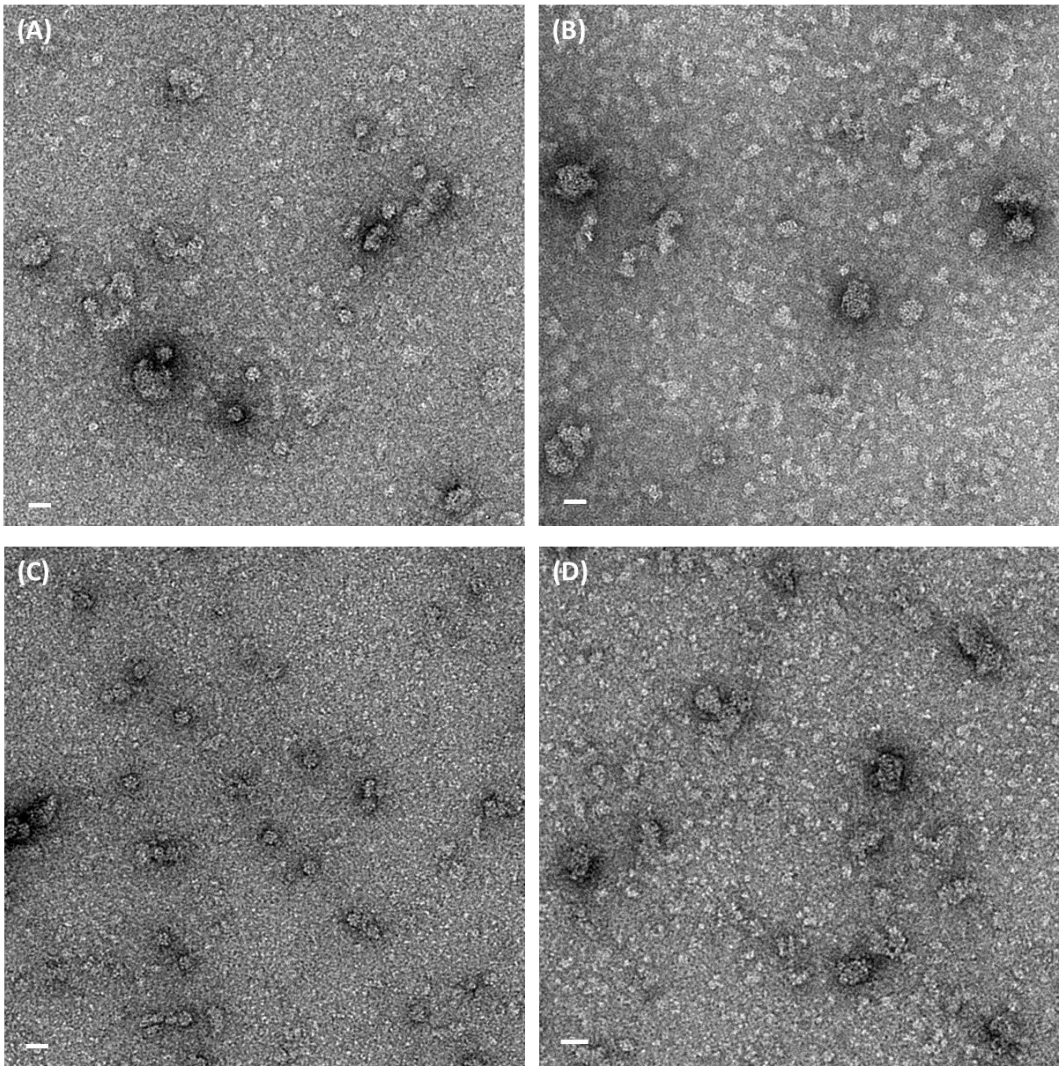


Figure 4.14: ANV RNA does not promote virion formation by ANV capsid core or full-length CP. Assembly of viral particles by addition of ANV RNA to purified recombinant capsid core and full-length ANV CP was assayed assessed by negative stain TEM. (**A**) Capsid core, or (**B**) full-length CP in TNS buffer show only the presence of capsomer-like, and irregular protein structures, respectively. Addition of *in vitro* transcribed ANV RNA (0.5 ng/µl) at an estimated substoichiometric ratio to (**C**) capsid core, or (**D**) full-length CP does not produce any noticeable change in the structures for either CP form when compared to their respective TNS-only control. Scale bars = 25 nm.

4.5 Summary

4.5.1 Expression and purification of recombinant ANV capsid protein

Expression of ORF2 CP from various AAstVs was attempted in systems ranging from plant to bacteria, resulting in the detection of recombinant ANV or CAstV CP in cellular lysate of *N. Benthamiana* plants, DF-1 cells and in BL21 bacteria (Figure 4.1, Figure 4.3 and Figure 4.5,

respectively). However, detection of intact full-length CP was not always observed, and presence of smaller than expected products suggests that either processing, or degradation of CP had occurred in these systems. The importance of purifying intact CP can be explained by analogy to the CPs of other viruses which use a PS-mediated assembly mechanism, wherein a basically-charged unstructured arm at one of either of the CP termini plays a crucial role in RNA interaction[19, 51, 55, 59]. The function of this domain varies between viruses, for example, neutralisation of the positive charge in STNV CP arms by PSs allows CP-CP interactions to overcome electrostatic repulsion and assembly-competent CP trimers to form[51, 52]. The CPs of the *Astroviridae* also contain a similar basically-charged N-terminal arm[194, 220], and removal of this region in whole, or in part, significantly reduces infectious viral titre[231], suggesting it may be critical to virus formation. However, CP can still form VLPs when the basic arm is absent[266], supporting the hypothesis that this region is specifically involved in PS recognition and or RNA binding. Thus, purification of ANV CP with an intact N-terminal region is crucial in order to study PS-mediated encapsidation by ANV. Therefore, as only the recombinant capsid core and full-length ANV CP produced in BL21 cells appeared to be intact, protein produced by this system was selected for further study.

Processing of the AAstV CP in eukaryotic expression systems could be anticipated given that HAstV CP undergoes a complex multi-step proteolytic maturation pathway during infection, requiring at least both caspase and trypsin-like activites[225, 292]. This is reflected by the different patterns of processed ORF2 products between the various expression systems investigated here, indicating that one (or more) unknown proteolytic events may have occurred (Figure 4.2, Figure 4.3 and Figure 4.4). The importance and extent of processing by trypsin-like and other proteases in the lifecycle of ANV is currently unexplored. However, a previous report suggests that ANV virions are infectious without trypsin-treatment[211], which is further supported by work presented in this thesis (Section 5.3.2). Collectively, these data suggest that ANV may follow a different maturation process to the other studied astroviruses and might not require proteolytic digestion in order to be fully infectious. However, trypsin digestion patterns of the ANV CP in this work may indicate that the CP is processed similarly to HAstV CPs (Figure 4.9). Thus, further study is needed to investigate the full requirement of protease activity by ANV CP for maturation.

4.5.2 Formation of virus-like particles from purified ANV capsid protein

Both the purified ANV capsid core and full-length CP appeared to be competent for assembly into capsomer-like structures (Figure 4.10), showing similarity to the 16 nm ring-

like structures formed by HAstV CP previously[266]. The morphologically similar HEV capsid is formed by the assembly of dimers into decameric capsomers[328], therefore, if the AstV capsid is as structurally similar as indicated (Section 3.2.2.3)[194], it would be reasonable to hypothesise that these 16 nm ANV capsomer-like structures could represent decamers of CP.

The full-length ANV CP was able to form irregular ~30 nm VLP-like structures or intermediates when digested with trypsin (Figure 4.13). However, the inability of the recombinant ANV CP to form VLPs of homogenous size and shape indicates the potential absence of key factors required for proper capsid assembly. Studies show that full-length recombinant HAstV CP can assemble into VLPs when expressed in eukaryotic cells, with or without the 70 N-terminal residues. This would imply that the assembly of VLPs for other AstVs may be possible[266]. Therefore, the inability of ANV CP to assemble into VLPs here may be a result of producing recombinant CP in a bacterial expression system, wherein the necessary post-translational modifications or processing required by CP for assembly do not occur.

An alternate explanation for the inability of recombinant ANV CP to form VLPs could be that RNA, or more specifically PSs, are required for capsid assembly. This is supported by data for HEV, where it was observed that formation of HEV T=3 particles requires the presence of viral RNA, and only smaller T=1 particles form in the absence of RNA[328]. However, ANV CP did not readily form capsids upon addition of *in vitro* transcribed ANV genomic RNA to preparations of purified CP (Figure 4.14). This would be consistent with a co-transcriptional encapsidation mechanism, where only nascent genomes are packaged by CP as they are transcribed by the RdRp in assemblysomes[37, 59, 70-72]. As genomic RNA is produced by the viral polymerase, there is opportunity for short-range structures to form in the RNA, i.e. PSs, which may be energetically unfavourable in the context of larger RNA molecules. This could explain why direct addition of ANV RNA to CP was not sufficient to promote particle assembly, hinting that active transcription of the viral genome in the presence of CP, or addition of shorter RNA fragments to CP preparations may be what is required to promote ANV VLP formation.

Chapter 5

5 Construction of a reverse genetics system for ANV

5.1 Alternate approaches to identifying putative packaging signals

As described earlier (Section 1.6), astroviruses share features with other viruses that utilise a PS-mediated assembly mechanism and lack any known machinery for packaging genomic RNA into pre-formed capsids. This suggests that the instructions for genome encapsidation may be encoded within the sequence and structure of the viral genomic RNA itself. Various approaches for the identification of PS-CP interactions have been developed over the last decade, including but not limited to RNA SELEX and NGS[53, 80]. Of these approaches, RNA SELEX coupled with bioinformatics analysis of the selected library and its matches to the viral sequence has been the most prominent method used for the identification of putative PSs in multiple viral families[37, 53, 55, 62].

For SELEX identification of putative PSs, purified assembly-competent capsid protein is required which can be derived from either capsomers, disassembled virions, or disassembled VLPs. As attempts to produce and purify ANV VLPs or assembly-competent capsomers for SELEX screening were inconclusive (Section 4.5.2), a reverse genetics system for ANV was constructed. The reverse genetics system opens up approaches for PS identification which are not available with a VLP system. For example, purified virus could be used in high resolution cryo-EM to identify RNA-CP interaction[314], or for deep sequencing of packaged and unpackaged viral genomes to identify putative PSs[80]. This NGS approach is less challenging to undertake than SELEX, as disassembly and partial reassembly of the capsid shell is not required for NGS, and only the isolation of packaged (virion) and unpackaged (cellular) viral RNA populations is needed.

A functional reverse genetics system is also required for validation of putative PSs in viral genome encapsidation, irrespective of the experimental approach used to identify them. Importantly, a cDNA clone provides a stable genetic background into which mutations can be readily introduced, and from which recovery of putative PS-mutant viruses can be attempted for subsequent comparison to the wild type virus.

To date, only a few reverse genetics systems have been developed for the *Astroviridae*, with systems for ANV G-4260, Porcine astrovirus 1 strain GX (PAstV1-GX1), HAstV-1, and a HAstV-1/8 chimera reported[211, 257-259]. The last report of a successful reverse genetics

system for an avian astrovirus was in the early 2000s[211]. In that study, two plasmid clones of ANV were constructed, one from which genome-length RNAs were transcribed *in vitro* under the control of a T7 promotor, and subsequently *in vitro* m^7G capped before transfection into chicken kidney (CK) or baby hamster kidney (BHK) cells. The other system consisted of a plasmid within which the cDNA clone of ANV was under the control of the CMV promotor, allowing this to be directly transfected into CK or BHK cells to produce RNA, subsequently recovering infectious ANV[211].

Based on this previous study[211], a similar plasmid-based system was constructed for ANV to provide a key tool for the study of this virus. ANV recovered from this system was further characterised, and methods to concentrate and purify the virus were explored in order to produce material suitable for PS identification.

5.2 Construction of a full-length cDNA clone of ANV G-4260

5.2.1 Synthesis of cDNA from astrovirus genomic RNA

To produce a plasmid-based system for the production of infectious ANV, a cDNA copy of the full-length RNA genome must be cloned and assembled, as summarised in Figure 5.1. To achieve this, viral genomic RNA was purified from an inactivated isolate of ANV G-4260 (a kind gift of V. Smyth, AFBI), and used as a template for cDNA synthesis (Sections 2.2.1.2 and 2.2.1.4). The primers used for cDNA production (Table 2.1) were positioned to produce overlapping fragments of the genome to enable subsequent construction of a genome-length cDNA clone. The resulting cDNAs were gel purified (Section 2.2.1.9) with fragments of the expected size (GSP primed reactions) extracted for ligation into pCRII using the Zero Blunt TOPO PCR cloning kit (Invitrogen). The resulting plasmids were transformed into *E. coli* DH5α and 2-3 clones selected for subsequent preparation of plasmid DNA and sequencing of the insert (Sections 2.2.1.12 and 2.2.1.13). Sequence data was checked against the sequence databases using BLAST[329] and then assembled into contigs using SeqMan Pro (DNASTAR lasergene) using the default assembly settings to validate overlaps between genome segments (Section 2.2.1.15).

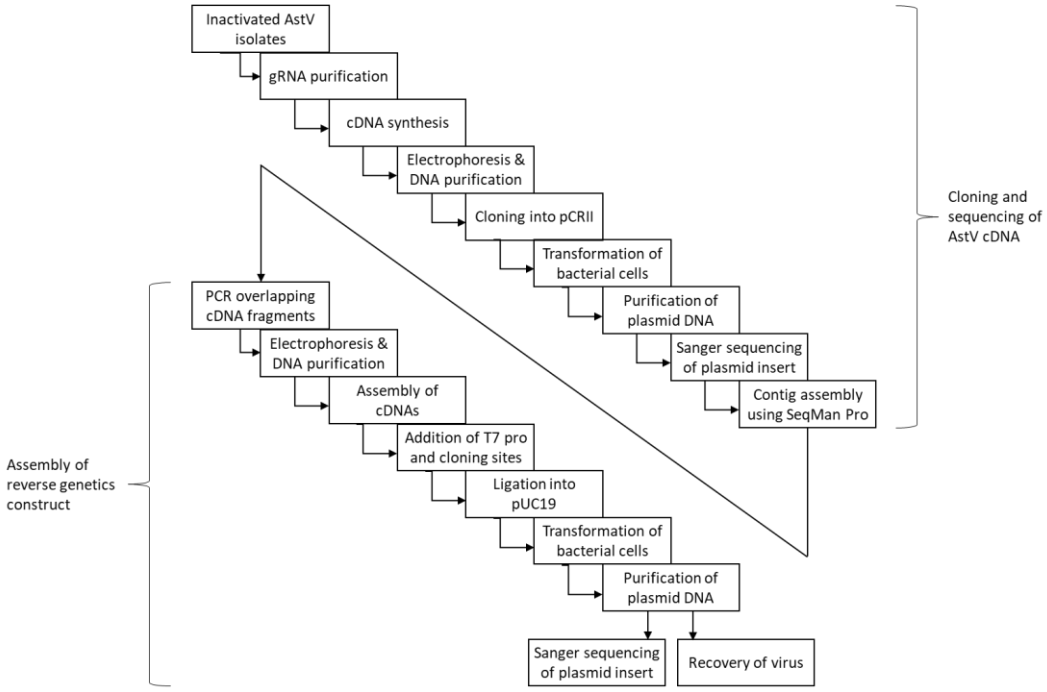


Figure 5.1: Process of cloning ANV cDNA and constructing the plasmid-based reverse genetics system. Flow chart detailing the stages involved in cloning a genome-length cDNA of ANV, and construction of a functional sequence-validated infectious reverse genetics system for ANV.

5.2.2 Assembly of a genome-length ANV G-4260 expression plasmid

Plasmids containing cloned regions of the ANV genome were used to assemble a genome-length cDNA clone of ANV for the reverse genetics construct. Four overlapping DNA segments (R3_1, R2_3, R1_8 and dT_7) that spanned the entire genome of ANV (Figure 5.2 A) were chosen and amplified by PCR to produce linear fragments. These fragments were mixed in equimolar amounts and subject to PCR to produce a single genomic length cDNA clone of ANV G-4260 (cloneANV) (Figure 5.2 B). Addition of flanking sequences for assembly into pUC19 and a 5' T7 promotor site (Section 2.1.8) to drive expression of the ANV genome to cloneANV was then performed by PCR to generate cloneANV-T7. Next, HDR-T7, a PCR-derived synthetic DNA sequence containing a 3' Poly-A_{20} to mimic the viral pA tail followed by the hepatitis delta ribozyme (HDR) and T7 terminator sequence (Section 2.1.8), was assembled into EcoRI-linearised pUC19 along with cloneANV-T7 using the NEBuilder HiFi DNA assembly kit (Section 2.2.1.10) to produce pANV (Figure 5.2 C). pANV was initially characterised by restriction digest with BglI, and by PCR using primers targeting the 5' and 3' ends of the cassette, as well as the sequence in pUC19 flanking the multiple

cloning site (Figure 5.3). The BglI digestion pattern and PCR results were consistent with those expected for correctly assembled pANV.

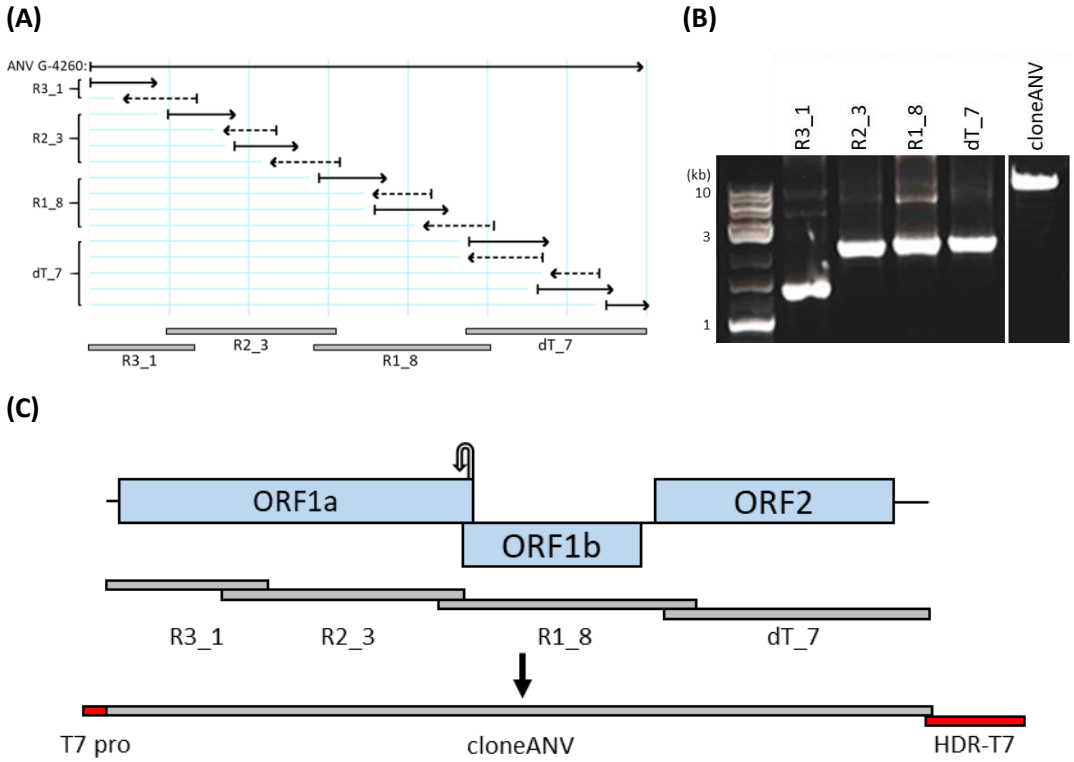


Figure 5.2: Assembly of the genome-length ANV cDNA cassette. (**A**) Overlapping cDNA segments of the ANV genome were cloned, sequenced and then assembled into contigs using SeqMan Pro (DNASTAR lasergene) using the published ANV G-4260 genome as a reference sequence (GenBank ID: AB033998.1). Sequences for the four cDNAs (shown below assembly) were generated using both flanking and internal primers due to the length of the fragments. (**B**) PCR amplification of the four cDNA fragments used for assembly of cloneANV, and the single genomic-length cDNA assembled from the four fragments by PCR. (**C**) The organisation of the ANV genome is shown, detailing the three open reading frames, UTRs and the RFS structure. The four overlapping clones generated by RT-PCR used for assembly of the genome-length cassette and the regions they represent are shown underneath the ANV genome (grey), with the number below indicating the clone. The HDR-T7 terminator and T7 promotor sequences introduced during PCR and assembly into pANV are also shown (red).

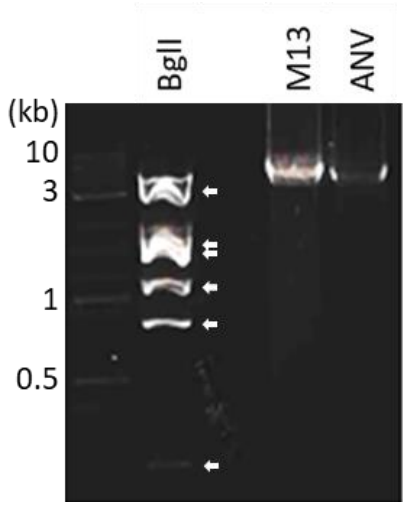


Figure 5.3: Confirmation of correct assembly of pANV. The correct assembly of cloneANV-T7 and HDR-T7 into pUC19 was confirmed by restriction digest and PCR of pANV. Digestion with BglI produced the expected DNA band pattern, with sizes of 4225, 1867, 1648, 1118, 790 and 242 nts. PCR with either the M13 primer pair targeting sequence flanking the multiple cloning site of pUC19, or primers targeting the 5' and 3' ends of the ANV and HDR insert show products of expected size, approximately 7.2 kb.

The ANV reverse genetics construct built here differs from those reported previously[211]. One such system utilised *in vitro* transcription from the T7 promoter and *in vitro* capping to produce viable RNA templates for translation after transfection. The system described here uses the T7 promoter combined with the self-cleaving HDR to produce capped full-length transcripts, without extra sequence, directly after transfection of the pANV plasmid into T7 RNA polymerase expressing cells, such as BSRT7 cells. A T7-driven method was chosen as this provides reliable expression of downstream genes. Addition of the HDR allows subsequent self-cleavage of the T7-expressed RNA, creating transcripts with a defined 3' Poly-A tail. Thus, producing RNAs that are more similar to viral RNA in infected cells with a simpler approach than those previously reported[211].

5.2.3 Sequence analysis of pANV

The cDNA construct within pANV was sequenced and compared to the previously published ANV sequence[211] to confirm the absence of mutations in ORFs that may have been introduced through reverse transcription and PCR. pANV showed 99.74% sequence homology to the published ANV G-4260 genome (GenBank ID: AB033998.1), with 18 base differences (Figure 5.4 and Table 5.1). Fourteen of these were found to be non-synonymous, and eight of which occurred in ORF1a, one in ORF1b and five in ORF2. Three were synonymous, with one in each ORF, and a single base deletion was also present at position 1 of the genome with respect to the reference sequence. Importantly, none of the identified mutations introduced a premature stop codon that would lead to early termination of translation. Additionally, no mutation within the T7 promotor/terminator or HDR sequences was found.

The sequence variation observed here may be due to selection of mutants from within the ANV quasispecies, or from genetic drift considering that the original published sequence was submitted almost 20 years ago, therefore the identified nt differences in pANV were not repaired. Genetic variation is to be expected as astroviruses are +ssRNA viruses, which are known to contain error-prone RdRps, allowing mutations to accumulate over time[330, 331]. This is consistent with the sequence diversity seen in Avastroviruses, even between closely related subtypes[332].

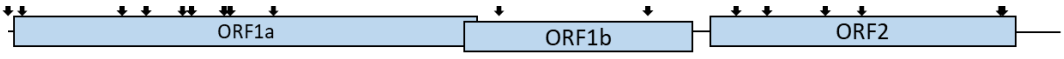


Figure 5.4: Location of mutations within the pANV construct. Schematic diagram of the ANV genome and its encoded ORFs within the pANV construct. The locations of point mutations identified by sequence comparison to the published ANV G-4260 strain (GenBank ID: AB033998.1) are indicated by arrows. A majority of the identified base variants were located within the ORF1a non-structural gene. However, one was found in the 5' UTR (position 1) which had been removed during the pANV cloning process. The remaining point mutations were located in ORF1b and ORF2, two in the polymerase and six in the capsid gene, respectively.

	Nucleotide substitution	Amino acid change vs reference genome
5' UTR	C1Δ	-
ORF1a	G48A	R11H
	C729T	P239L
	T888C	I292T
	A1127G	T372A
	A1185T	Y391F
	T1401C	F463S
	T1439C	F476L
	T1441C	-
	T1728C	V572A
ORF1b	C3209A	L37I
	G4191A	-
ORF2	C4770T	T67I
	C4972T	-
	A5357C	T263P
	T5597C	S343P
	C6513T	A648V
	G6528C	G653A

Table 5.1: Mutations present in pANV. Nucleotide substitutions and their effect on the encoded amino acids in each ORF between the pANV reverse genetics construct and that of the published wild type sequence (GenBank ID: AB033998.1) are shown.

5.3 Recovery of infectious virus from the ANV reverse genetics construct

5.3.1 Recovery of ANV in BSRT7 cells

With the presence of non-synonymous mutations in the pANV cDNA clone, the putative ANV reverse genetics construct was tested for the ability to generate infectious virus. When pANV was transfected into BSRT7 cells (Section 2.2.6.1) no cytopathic effect (CPE) was observed up to 72 hours post-transfection, which may be expected as the cell line is derived from tissue of a non-target host for ANV. In the absence of CPE, the production of ANV genomic RNA from transfected pANV was tested by RT-PCR. Total RNA was prepared from pANV-transfected and control BSRT7 cells and subject to RT-PCR (Sections 2.2.1.2, 2.2.1.3 and 2.2.1.5). Using this approach, RNA corresponding to ANV was detected in BSRT7 cells (Figure 5.5) validating the production of viral RNA from pANV by the T7 RNA polymerase constitutively expressed in BSRT7 cells.

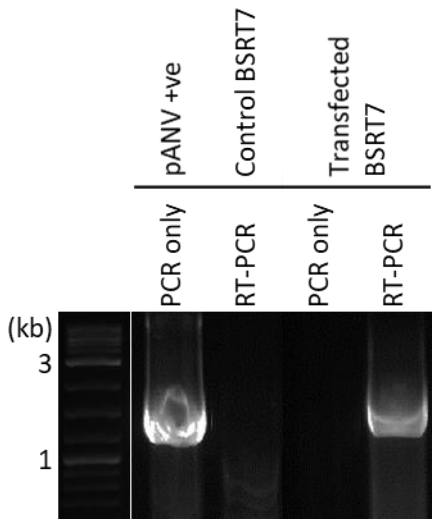


Figure 5.5: ANV genomic RNA is produced following pANV transfection into BSRT7 cells. BSRT7 cells were mock-transfected (without plasmid) or transfected with pANV, and cells were harvested 72 hours post-transfection for extraction of total cellular RNA. Purified total RNA was subject to RT-PCR or PCR (without RT) in order to confirm the specific presence of ANV genomic RNA. ANV ORF1b DNA (2.2 kb) was detected in the pANV plasmid control and no ORF1b RNA or DNA was detected in the mock-transfected BSRT7 cells. By contrast, RNA corresponding to ANV was detected in pANV transfected BSRT7 cells by RT-PCR, with no pANV plasmid DNA detected by PCR alone, confirming that RNA is transcribed from the construct within transfected BSRT7 cells.

5.3.2 Passage of ANV in CK cells

As transcription of ANV genomic RNA was detected in BSRT7 cells despite the absence of apparent CPE, the production of infectious ANV virions by BSRT7 cells was assessed. For this, primary chicken kidney (CK) cells from 2-3 week-old Rhode Island Red birds, which are derived from the normal target tissue of ANV, were infected with ANV recovered from either the lysate of transfected BSRT7 cells (intracellular virions) or the extracellular media from these cells (released virions). By 96 hpi CPE was observed in CK cells exposed to the lysate of transfected BSRT7 cells, but not in those exposed to BSRT7 extracellular media or in control uninfected CK cells (Figure 5.6 A-C). To confirm that the observed CPE was caused by ANV, total RNA was prepared from infected CK cells showing CPE and used for RT-PCR detection of ANV RNA as previously (Section 5.3.1). No ANV RNA was detected in uninfected CK cells, indicating these cells were derived from birds that did not have a pre-existing ANV infection (Figure 5.6 D). ANV RNA was readily detected in CK cells exposed to

the lysate of BSRT7 cells transfected with pANV. By contrast, ANV RNA was not detected in CK cells exposed to the extracellular media of pANV transfected BSRT7 cells, which may contain released virions (Figure 5.6 D). The presence and absence of ANV RNA was consistent with the presence or absence of CPE in cells exposed to the lysate or extracellular media of pANV transfected BSRT7 cells, indicating that ANV was the causative agent of cell death.

These data indicate that infectious ANV virions were produced from the pANV reverse genetics system after transfection into BSRT7 cells (Figure 5.5), and that these virions were capable of productively replicating in CK cells in line with previously reported systems[211]. Furthermore, the ability to recover and propagate infectious ANV virions indicates that the 18 nt differences present in the pANV construct, compared to the reference sequence, are not significantly deleterious to the viability of ANV G-4260. The absence of CPE and ANV RNA in CK cells exposed to extracellular media from transfected BSRT7 cells suggests either that very low titres of, or no, extracellular virions are produced by these cells, or that virions are retained in the cytoplasm.

To assess if any of the mutations in pANV affected the ability to produce extracellular virions, naïve CK cells were infected with ANV recovered from the cell lysate and extracellular media of infected CK cells. No signs of CPE were seen in control uninfected CK cells (Figure 5.7 A). In naïve CK cells exposed to either the lysate (Figure 5.7 B) or extracellular media (Figure 5.7 C) of infected CK cells, CPE manifested as subtotal destruction around 72-96 hpi and was observed as clusters of rounded CK cells loosely associated with the monolayer. This is consistent with the previously observed CPE in CK cells infected with BSRT7 lysate-derived ANV. RNA corresponding to ANV was also readily detected in these cells in line with the observed CPE (Figure 5.7 D). Taken together, these results indicate that the synonymous and non-synonymous mutations present in the pANV construct do not prevent ANV replication in CK cells, and that extracellular progeny virions are produced by infected CK cells.

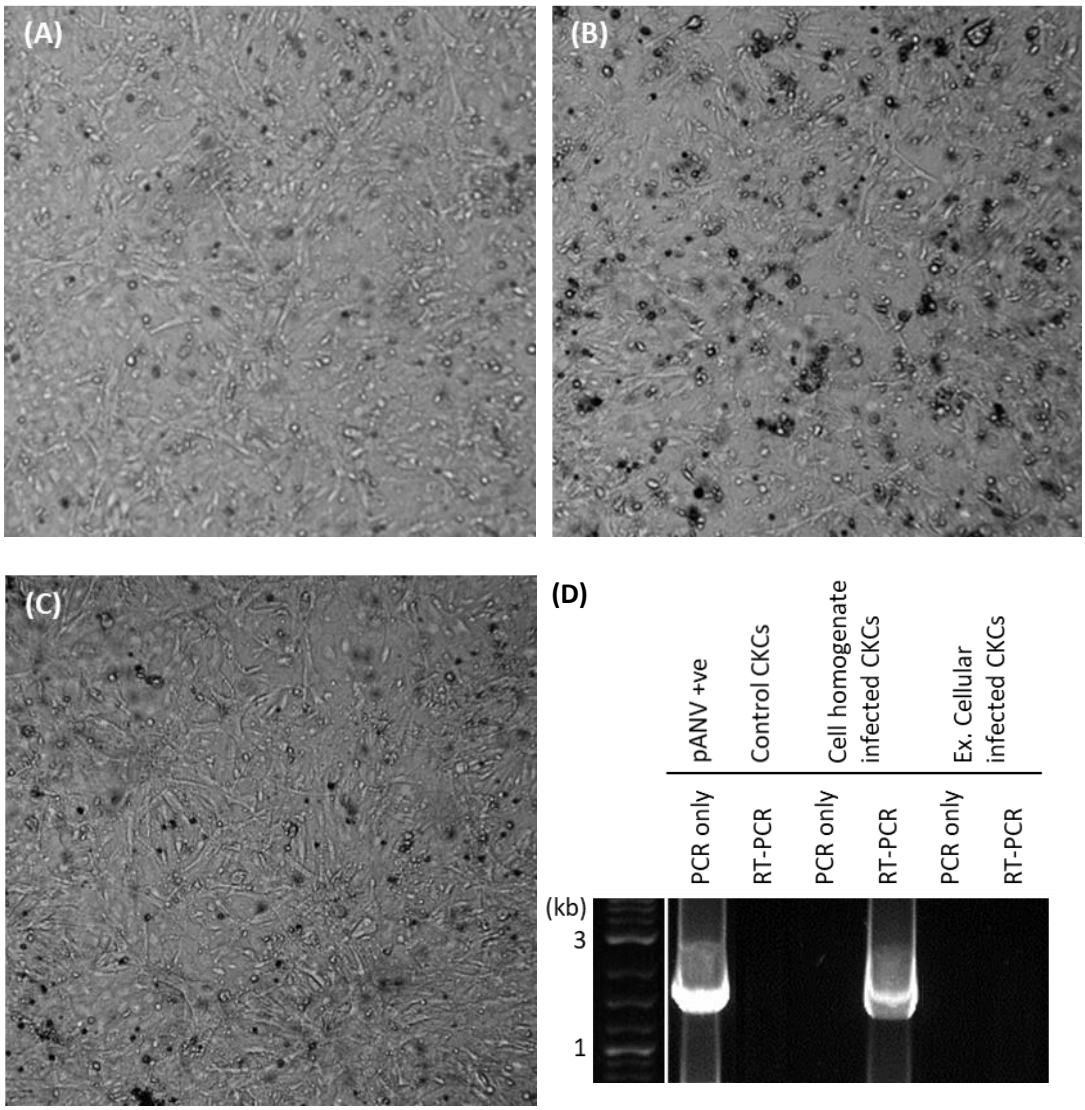


Figure 5.6: ANV recovered from BSRT7 cells is infectious to CK cells. (**A**) Uninfected CK cells are healthy and evenly distributed across the flask after 96 hours. (**B**) CK cells infected with pANV transfected BSRT7 cell lysate show signs of CPE at 96 hpi, with cells rounding and detaching from the flask. (**C**) CK cells infected with extracellular media of pANV transfected BSRT7 cells show little indication of CPE, similar to uninfected cells. (**D**) Detection of ANV ORF1b DNA/RNA (2.2 kb) in the plasmid control, uninfected CK cells and in CK cells inoculated with pANV-transfected BSRT7 cell lysate or extracellular media. RNA corresponding to ANV is detected in CK cells infected with pANV-transfected BSRT7 cell lysate by RT-PCR, but not by PCR alone, confirming that viral RNA is present and that the PCR signal is not due to residual contamination of the pANV plasmid DNA. No ANV ORF1b RNA was detected in CK cells inoculated with pANV-transfected BSRT7 extracellular media.

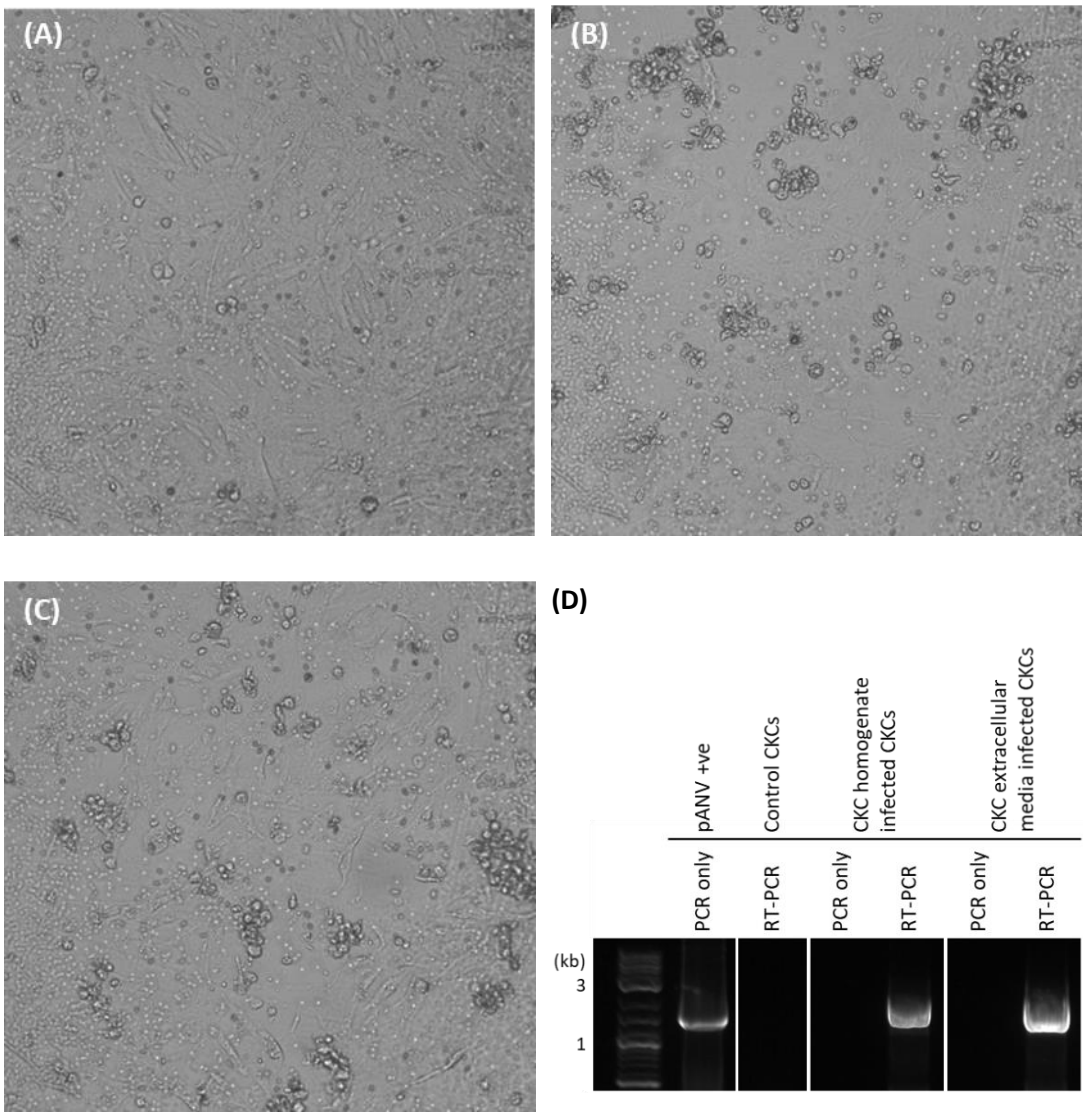


Figure 5.7: CK cells produce extracellular ANV virions. Naïve CK cells were infected with ANV from the lysate or extracellular media of infected CK cells. (**A**) In uninfected CK cells, no CPE was seen. Whilst in naïve CK cells inoculated with either (**B**) ANV infected CK cell lysate or (**C**) extracellular media, CPE was seen between 72 and 96 hpi. (**D**) ANV ORF1b DNA (2.2 kb) was detected in the positive pANV plasmid control, whilst no ORF1b RNA or DNA was detected in control CK cells by RT-PCR. ANV ORF1b RNA was specifically detected in CK cells infected using both cell lysate and extracellular media derived from ANV infected CK cells.

5.3.3 Characterisation of the reverse genetics system derived ANV

5.3.3.1 Yield of ANV from CK cells

To establish the yield of infectious ANV from infected CK cells, extracellular media stocks of ANV derived from the reverse genetics system were titrated on naïve CK cells (Section 2.2.7.1). Monolayers of naïve CK cells were exposed to 10-fold serial dilutions of released ANV for 1 hour before overlay with BES media containing 1% (w/v) agar. Infected cell monolayers were fixed and stained at 96 hpi, by which time small indistinct plaques of ANV were apparent on CK cells. Enumeration of plaques at the 10^{-3} dilution indicated an infectious titre of 2.1 x 10^5 pfu/ml (Figure 5.8 A). However, the plaque assay protocol used here for ANV appears unreliable as wide-spread CPE was commonly seen in infected CK cells overlaid with agar. Defined plaques were only occasionally observed, suggesting the concentration of agar used in the overlay was insufficient to prevent diffusion of ANV. Thus, an alternate method of quantification was required to reliably calculate the infectious titre. In order to quantitate the yield of infectious ANV in a manner that could be consistently replicated, the 50% tissue culture infective dose ($TCID_{50}$) method of quantitation was used (Section 2.2.7.2). Replicate wells of CK cells were inoculated with 10-fold serial dilutions of extracellular media stocks of ANV and used to calculate the $TCID_{50}$ value of the stock. Individual wells of CK cells were observed for CPE at 24-hour intervals and counted as positive if there were wide-spread signs of cell rounding and detachment from the well. Titres were then calculated using the Spearman-Kärber method[307]. Using this method, ANV in an extracellular media stock was quantitated at a titre of 4 x 10^5 $TCID_{50}$/ml at 96 hpi, with no increase in titre to 120 hpi (Figure 5.8 B). The $TCID_{50}$ of ANV recovered from pANV shows lower levels of infectious units compared to that achieved by Imada *et al.*, with a 10-fold difference in titre[211]. This may be due to the single freeze-thaw cycle that the ANV stocks used here were subject to prior to inoculation, which could cause a loss in infectious units due to particle disruption from ice crystal formation. The resulting titre of ANV at 96 hpi calculated as $TCID_{50}$/ml is in broad agreement with the observed plaque forming titre, indicating both methods are capable of measuring the yield of ANV to a similar accuracy.

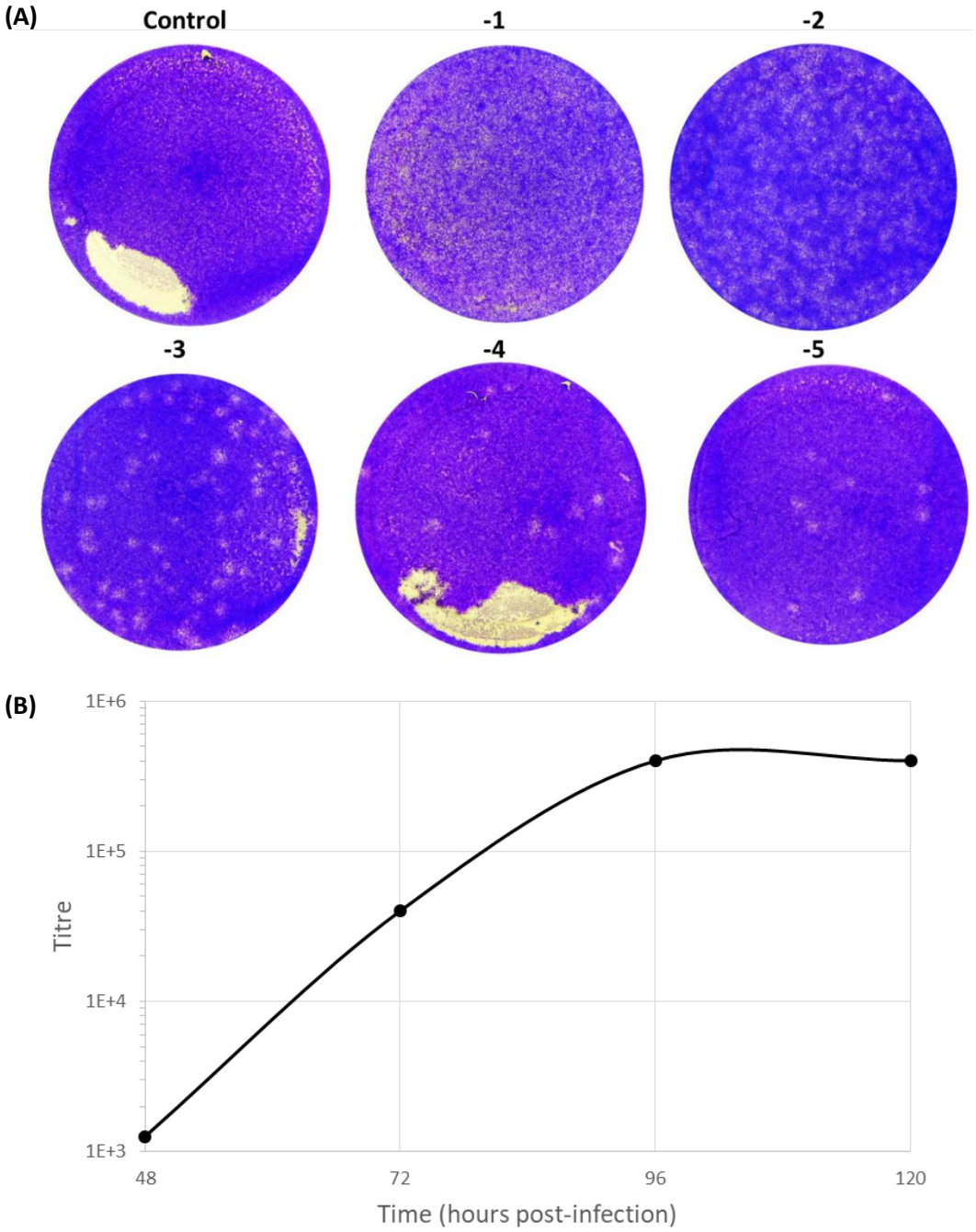


Figure 5.8: Yield of ANV from infected CK cells. (**A**) CK cells were inoculated with 10-fold serial dilutions of ANV and overlaid with BES containing 1% agar. At 96 hpi, agar plugs were removed and cells fixed in paraformaldehyde before staining using 0.1% crystal violet. Plaques formed by ANV are small and indistinct, giving titres of 2.1 x 10^5 pfu/ml when calculated from the 10^{-3} dilution. (**B**) Replicate wells of CK cells were inoculated with 10-fold serial dilutions of ANV and occurrence of CPE was reported per well for each dilution at 24 hour intervals. Infectious titre ($TCID_{50}$) was then calculated at each time point using the Spearman-Kärber method and plotted over the course of infection.

5.3.3.2 Growth kinetics of reverse genetics system derived ANV

Having shown recovery of infectious ANV from the reverse genetics system, the replication kinetics of this virus in CK cells was investigated. In order to measure the accumulation of extracellular RNase-resistant viral RNA, as a proxy for production of infectious virions, an RNase digestion protocol was first optimised.

RNase A was added to extracellular media of ANV infected CK cells 72 hpi and incubated for different lengths of time. RNA from the remaining protected genomes was then purified and quantified by RT-qPCR using a TaqMan RT-qPCR assay targeting sequence within the ANV ORF1a region (Section 2.2.7.3). After 30 minutes incubation, the titre of genome copies detected was reduced by ~25% (or 4.1 x 10^8 genome copies/ml) when compared to the RNase-absent control, with no further loss of genome signal even when the incubation time with RNase was increased to 60 minutes (Figure 5.9). This data suggests that addition of RNase A to infected cell lysate at a concentration of 10 µg/ml, with subsequent incubation at 37 °C for as little as 30 minutes, is sufficient to degrade any unprotected ANV genomic RNA, whilst leaving putatively encapsidated ANV genomes intact. Unexpectedly, there was an increase in the number of detected genomes in infected media incubated with RNase for 10 minutes compared to the untreated control. This may be due to the degradation of viral particles and RNA upon extended incubation at 37 °C, as the untreated control was incubated at 37 °C for 60 minutes in total compared to the 10 minute incubation of the RNase treatment.

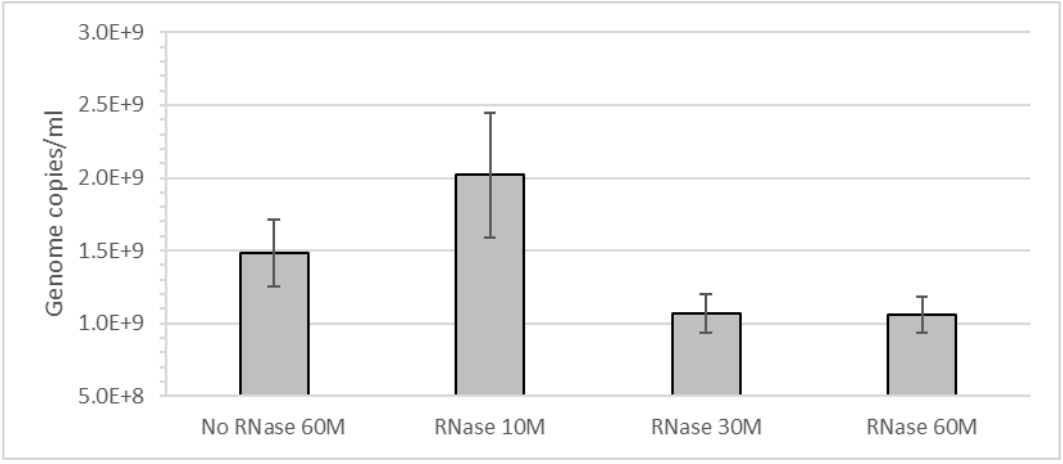


Figure 5.9: RNase A degradation of unprotected ANV RNA. Extracellular media from ANV infected CK cells was subject to RNase treatment for different lengths of time in order to determine the length of digestion required to fully degrade unprotected ANV RNA. Total RNA from each treatment was purified and genome copies of the RNase-resistant ANV RNAs were quantified by RT-qPCR, representing putatively encapsidated RNAs. The average absolute genome count from three technical replicates for each treatment is shown with the standard deviation.

With a protocol for measuring the quantity of putatively packaged ANV genomes in the extracellular media established, the release of ANV virions from infected cells over time was investigated. CK cells were infected with ANV at an MOI of ~0.13 (calculated from titres of RNase-resistant ANV RNA). Extracellular media was harvested from these cells at 24-hour intervals to assay the production of released virus. Harvested media was treated with RNase for 30 minutes (as above) before RNA purification in order to remove any non-encapsidated RNA present that could be derived from either disrupted cells or incorrectly formed particles.

Quantification revealed that extracellular ANV RNA is produced at a rate of 4.83×10^8 copies/24h during the linear phase of the growth curve (calculated from genome copy increase between 0-48 hours), reaching maximal levels of viral RNA at 72 hpi, equating to around 2×10^9 genome copies/ml (Figure 5.10). After 72 hpi, the titre of ANV RNA in the extracellular media plateaus. The 4-log difference between RT-qPCR genome copy number and previously calculated infectious titre (pfu/ml or $TCID_{50}$) could be explained by the fact that RT-qPCR only measures the quantity of RNA present, some of which may be derived from non-infectious defective particles, whereas pfu/$TCID_{50}$ measure the number of viable replication competent virions. These defective particles can account for a sizable percentage of the virion population in RNA viruses[333-335] and would not be detected by plaque assay or $TCID_{50}$ as these assays rely on the initiation of a productive infection to measure virus titre.

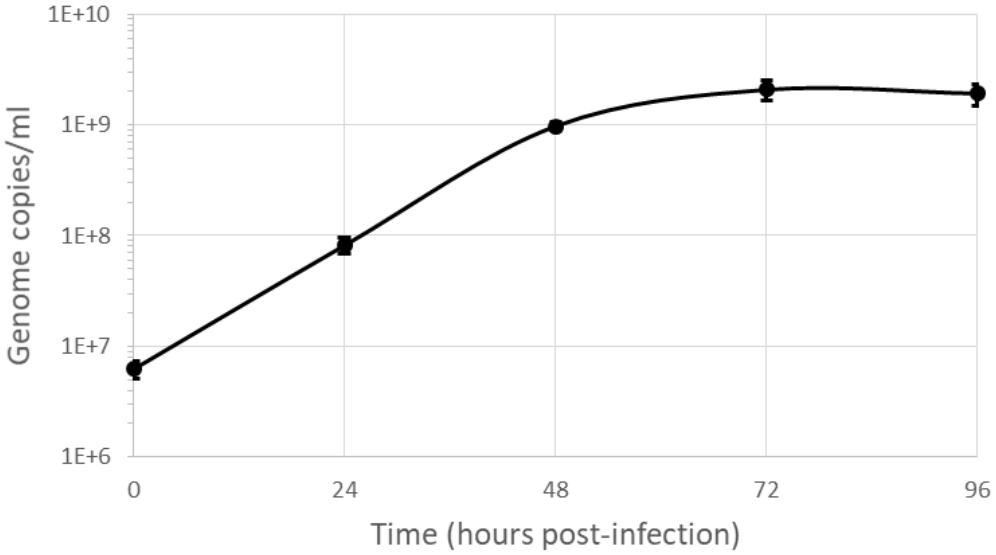


Figure 5.10: Kinetics of extracellular ANV virion production in CK cells. CK cells were infected with ANV at MOI 0.13 and extracellular media was harvested at 24h intervals to 96 hpi. Extracellular RNase-resistant ANV RNA was purified and absolute genome copies/ml quantified by RT-qPCR to determine the accumulation of extracellular ANV over time, with 0h representing the input inoculum. The average titre and the standard deviation of three technical replicates at each time point is shown.

5.3.3.3 Investigating cellular models for the cultivation of ANV

As primary CK cells are of limited availability, the ability of several other cell types to support ANV infection was assessed to establish if readily available cell lines, rather than primary cells, could support the propagation of ANV. Therefore, LMH and DF-1 cells and primary CK and CEF cells were grown to confluency and then infected with ANV. Cells were grown to similar confluency as to provide roughly equal starting numbers of cells so that the effective MOI would be comparable. Intracellular RNA was then extracted from the entirety of cells in each culture dish, harvested at 48 hpi, and viral RNA was quantitated using a TaqMan RT-qPCR assay targeting ANV ORF1a.

The highest quantity of ANV RNA was found in CK cells, followed by LMH and CEF cells which produced levels of ANV RNA 2- and 3-logs lower than CK cells, respectively (Figure 5.11). These data indicate that primary chicken kidney, immortalised chicken liver and primary chicken fibroblast cells can all support the replication of ANV, in line with *in vivo* data[336]. In contrast, an immortalised chicken fibroblast cell line (DF-1 cells), produced almost 6-logs less ANV RNA than primary CK cells, and almost 3-logs lower than their

primary CEF equivalent. BSRT7 cells infected with ANV were included as a control and showed the lowest levels of ANV RNA production (2.23 x 10^5 total genome copies).

Of the cells tested only CK cells showed signs of CPE, with no apparent CPE seen in the LMH cells in contrast to previous observations[91]. This suggests that whilst LMH cells produce comparable levels of ANV RNA, they may not be a suitable system in which to study the complete replication cycle of ANV. This is likely due to failing support of virion egress, as a previous report observed the majority of virus to be retained intracellularly in this cell line[91]. Overall, these data indicate that CK cells are the best available model for propagation of ANV, supporting high level viral RNA production and release of infectious ANV virions, consistent with *in vivo* data showing ANV to be highly prevalent in the kidneys of young chicks[336].

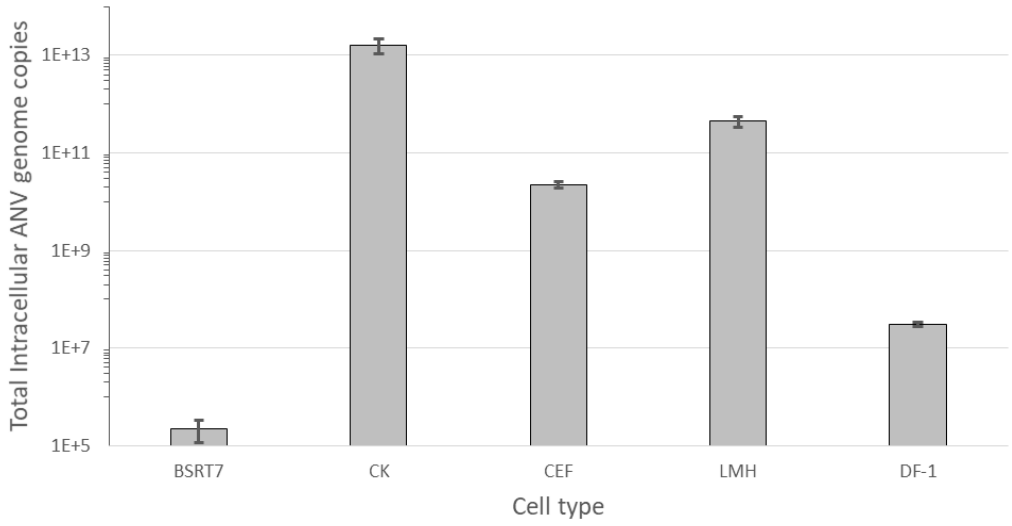


Figure 5.11: Intracellular ANV viral RNA accumulation in different cell types. Primary and immortal avian cell types were infected with ANV at similar effective MOIs. The entirety of cells in each culture dish were harvested at 48 hpi for RNA purification and RT-qPCR determination of absolute intracellular ANV genome copy numbers. The average intracellular ANV RNA titre in each cell type is shown with the standard deviation over three technical replicates.

5.3.3.4 Foci of ANV are detected in the cytoplasm of infected cells

To investigate the intracellular localisation of ANV and further confirm the identity of the virus recovered from the reverse genetics system, infected CK and LMH cells were subjected to immunofluorescence staining using a known antiserum from an ANV infected chicken (a kind gift of V. Smyth, AFBI). CK and LMH cells were infected with ANV at a MOI of ~0.02 (calculated from $TCID_{50}$) and fixed at 72 hpi before processing for immunofluorescence staining (section 2.2.7.4) The chicken anti-ANV antiserum reacted with both CK and LMH cells infected with reverse genetics system derived virus, indicating that the recovered virus shares close if not identical antigenicity to that of field circulating wild type ANV (Figure 5.12). Foci of ANV proteins are observed in the cytoplasm of infected CK and LMH cells, consistent with previous observations[211]. These foci may correspond to replication complexes within the cell, which are a characteristic of cytoplasmic replicating RNA viruses[337].

Infection appears to be more widespread in LMH cells than CK cells, as detection of protein corresponding to ANV is seen in a greater number of cells in the monolayer, whereas detection is mainly seen in clusters of cells for ANV infected CK cells. This data may suggest that LMH cells could be a more efficient system for ANV infection, as the effective ratio of infected to uninfected cells is greater than that seen in the CK cells. However, the apparent increase of ANV positive LMH cells could instead be due to intracellular retention of virus in these cells. Further investigation is required to characterise the replication of ANV in LMH cells, in order to validate if they are a suitable cellular model for the study of ANV.

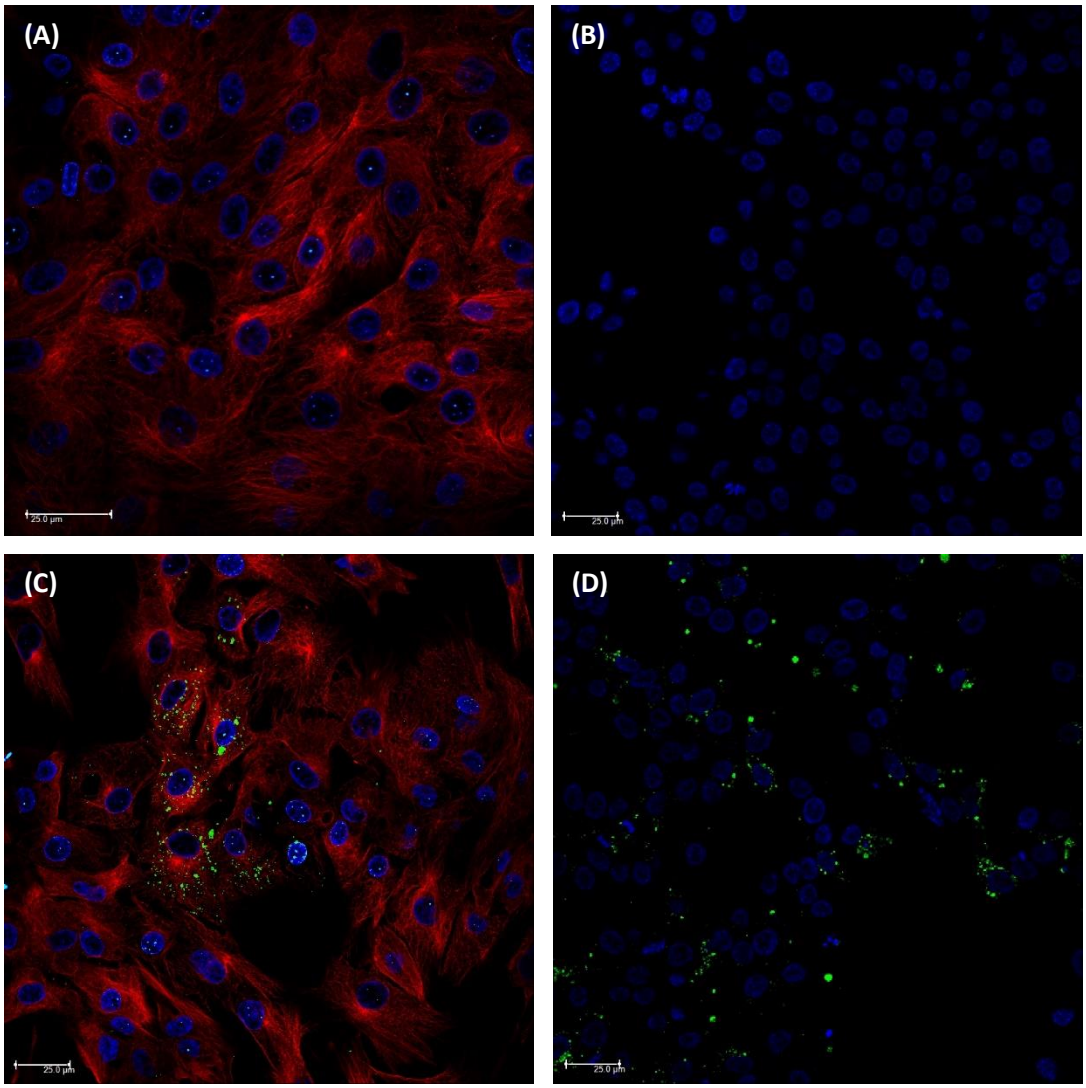


Figure 5.12: Foci of ANV are seen in the cytoplasm of infected CK and LMH cells. CK or LMH cells were mock infected or infected with ANV at an MOI of ~0.02, then fixed and processed for immunofluorescence at 72 hpi. Uninfected (**A**) CK cells and (**B**) LMH cells appear healthy and show no foci of ANV proteins (green), which are seen in both ANV infected (**C**) CK cells and (**D**) LMH cells when stained with an anti-ANV chicken serum. The tubulin network (red) of cells was stained with an anti-tubulin antibody to highlight the cytoplasm, and the nuclei of cells were stained with DAPI (Blue). Scale bars = 25 µm.

5.3.3.5 Characterisation of the anti-ANV chicken serum

To identify the putative target antigen(s) of the anti-ANV chicken serum, western blotting was used to investigate serum reactivity to denatured total protein from ANV infected cells and to recombinant CP expressed in bacteria (Section 2.2.4.2). As the lysate of infected CK cells will include products from all three of the ANV ORFs, recombinant ANV CP (core region and partial spike region, residues 1-626) expressed in BL21 cells was used as a known positive control for ANV CP.

The anti-ANV chicken serum did not detect any ANV proteins in uninfected CK cells or extracellular culture medium (Figure 5.13). By contrast, in infected CK cells the chicken antiserum detected a single protein species of approximately 37 kDa. A similarly sized band was detected in the lysate of bacteria expressing the recombinant ANV capsid, but not in control bacteria (Figure 5.13). These data suggest that an antibody within the chicken antiserum can recognise the denatured ANV CP and possibly a processed form. Previous observations indicate that astrovirus capsid protein is proteolytically processed and these processed forms typically range in size between 25-34 kDa[291, 292], consistent with the band detected by the chicken antiserum. These findings are also consistent with data from an earlier study, which identified several potential processed forms of ANV CP, including three different products between 33-43 kDa, using a rabbit antiserum raised against ANV virions[338].

As western blotting tests the ability of antibodies to recognise denatured epitopes, it is possible the chicken antiserum could contain other antibodies against alternate ANV antigens, which target their native conformation, including the non-structural proteins encoded within ORF1a or 1b. Therefore, further investigation is required to fully characterise the repertoire of targets of the anti-ANV chicken serum, in order to draw any further conclusions as to the identity of the cytoplasmic ANV foci revealed by immunofluorescence.

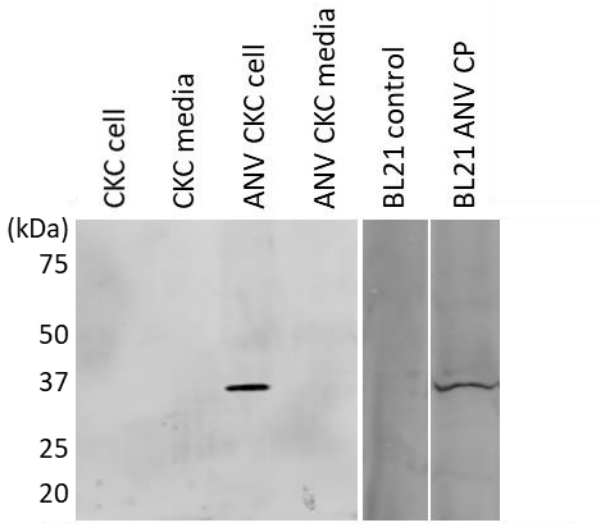


Figure 5.13: Anti-ANV chicken antiserum detects a processed form of capsid protein. Total protein from uninfected and ANV infected CK cells and from BL21 cells expressing recombinant ANV CP was prepared for SDS-PAGE and western blot analysis using the anti-ANV chicken antiserum as a primary antibody for detection. A single protein species of approximately 37 kDa was detected by the antiserum in ANV infected CK cells and BL21 cells expressing recombinant capsid protein, indicating that a western blotting-compatible antibody within the serum recognises epitopes of the ANV CP.

5.3.3.6 ANV associates with actin filaments

As cytoplasmic foci of ANV proteins had been detected by immunofluorescence of infected CK and LMH cells, infected CK cells were subjected to TEM to further investigate the intracellular localisation of ANV. CK cells were infected as previously in the immunofluorescence studies (Section 5.3.3.4) and fixed 72 hpi for embedding, ultrathin sectioning and TEM (Section 2.2.8.4).

In ultrathin sections of infected CK cells, virus-like structures of approximately 30 nm were seen (Figure 5.14), which are consistent with the expected 28 nm size of ANV virions[202, 211]. The characteristic 5/6 pointed star-like morphology of AstV particles was not observed here. Previous reports suggest that typically only ~10% of the astroviral population present this morphology, and it is thought that many of the AAstVs do not have prominent surface structures at all[168, 192-194]. These virus-like structures were predominantly located on the basal surface of the cells associated with actin filaments suggesting ANV may use actin and its associated motor proteins for transport and egress from infected cells.

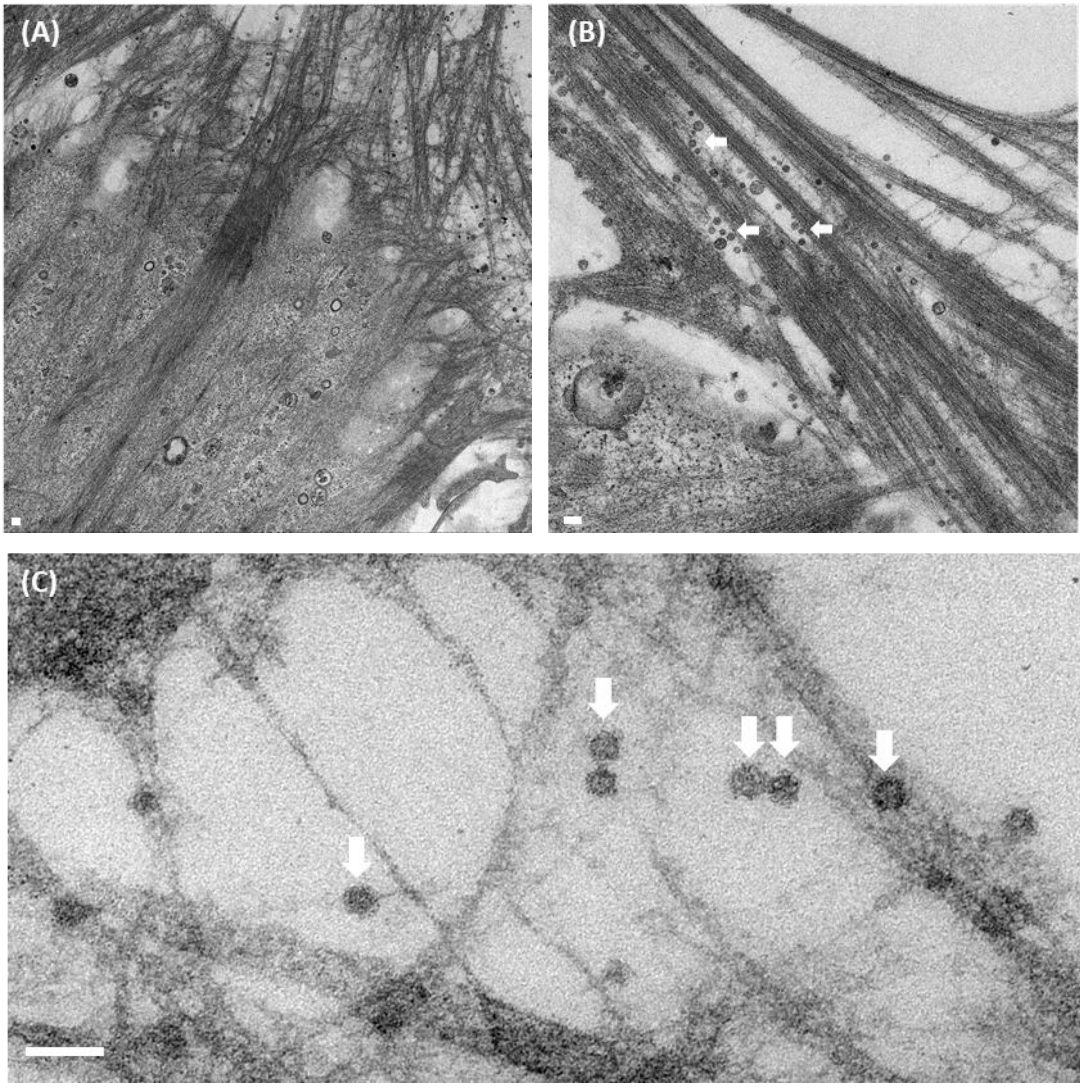


Figure 5.14: Virion-like particles are associated with actin in ANV infected CK cells. CK cells were infected with ANV at a MOI of ~0.02, fixed at 72 hpi, and then processed for ultrathin sectioning and TEM analysis. (**A**) Actin filaments located at the periphery of the cytoplasm in a CK cell are shown at low magnification. (**B**) Clusters of small spherical particles associated with the actin filaments, indicated by white arrows, are shown at medium magnification. (**C**) ~30 nm sized virion-like particles, indicated by white arrows, are seen associated with actin filaments at high magnification. Scale bars = 100 nm.

5.3.4 Preparation of infectious ANV for PS identification studies

5.3.4.1 Concentration of ANV virions by precipitation

ANV virions produced by cells infected with virus derived from the reverse genetics system are ideal alternative substrates to purified CP for identification of putative PSs. These virions can be used in NGS-based analysis to identify conserved sequences and structures within packaged RNAs. For this, highly concentrated and purified intact virus particles are required.

To concentrate ANV virions, increasing concentrations of polyethylene glycol 6000 (PEG) or ammonium sulphate were added to samples of clarified and filtered extracellular media from infected CK cells to precipitate virions. Precipitated ANV was collected by centrifugation and the resulting pellets were re-suspended in TN buffer, RNase-treated, and the number of packaged genomes then determined by RT-qPCR.

The yield of precipitated RNase-resistant viral RNA increases with rising PEG concentration, with the exception of 6% PEG, plateauing between 8% and 10% (Figure 5.15). At a final concentration of 8 or 10% PEG, around 4 times as many RNase-resistant ANV genome copies/ml were present compared to the untreated infected CK cell extracellular media (Figure 5.15). In contrast, precipitation with ammonium sulphate showed decreasing recovery of ANV RNA with increasing salt concentration. A final concentration of 20% ammonium sulphate was the most effective treatment with approximately 2.5 times as many RNase-resistant ANV genome copies present compared to the untreated infected CK cell extracellular media, lower than the maximum recovery with PEG (Figure 5.15). Thus, PEG precipitation appears to be the optimal method for the concentration of ANV from extracellular media.

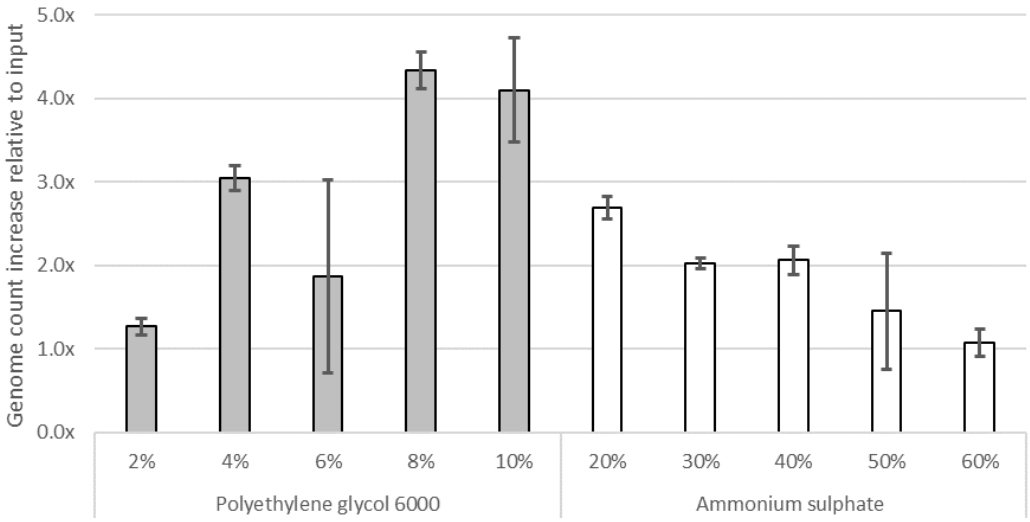


Figure 5.15: Concentration of ANV virions from extracellular media. To determine the optimal concentration of polyethylene glycol 6000 or ammonium sulphate for the precipitation of ANV particles, each was added to ANV infected CK extracellular media to a range of final concentrations. After precipitation, pellets were resuspended in 100 µl of TN buffer and RNase-treated to remove any unprotected ANV RNA. Absolute genome copy number, relative to 100 µl of untreated ANV infected CK extracellular media, was then determined by RT-qPCR for each treatment. The average recovery for three technical replicates is shown with the standard deviation of these replicates.

5.3.4.2 Concentration of ANV virions by sucrose density centrifugation

To further concentrate and purify intact ANV particles away from cellular material and improperly formed viral particles, ANV precipitated with PEG from clarified and filtered extracellular media of infected CK cells was subject to ultracentrifugation through a range of sucrose concentrations.

Resuspended ANV from PEG precipitation was layered on top of a sucrose cushion (20, 30, 40, 50, 60 or 70% in TN buffer) and subject to centrifugation at 130,000 *x g* for 2 hours at 4 °C. Pelleted material that passed through the sucrose cushion was re-suspended in TN buffer, RNA was purified from the resuspended pellet and used for RT-qPCR quantification of ANV genome copies. Quantitation of viral genomic RNA revealed that ANV pelleted through the 20 and 30% sucrose cushions (Figure 5.16), with the numbers of viral genome copies per ml increasing 73- and 33-fold, respectively, relative to the resuspended ANV PEG-precipitate. In contrast, ANV genomic RNA was not detected in the pellets that formed under cushions of 40% sucrose and above, suggesting that the buoyant density of ANV is between that of 20 and 40% sucrose under the conditions tested.

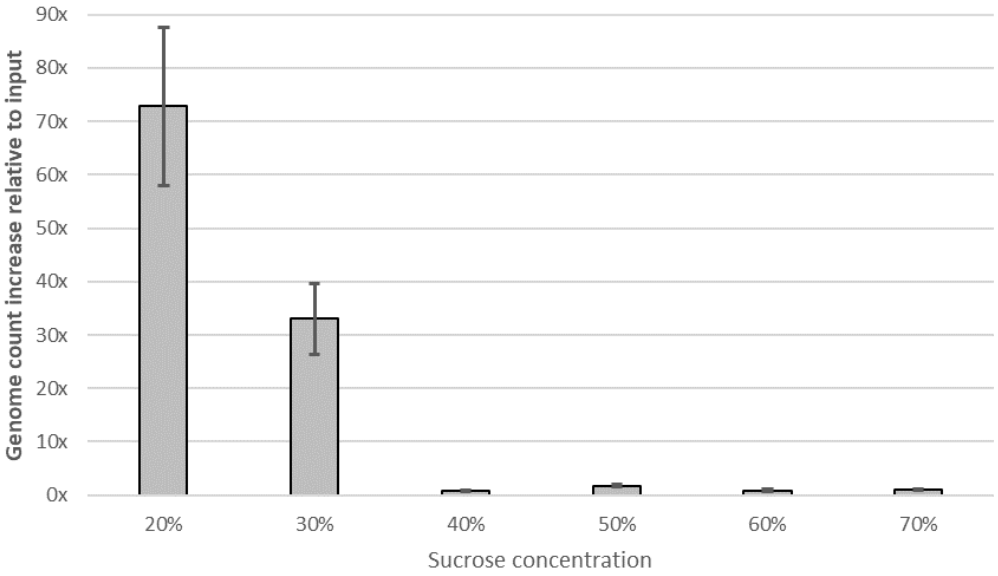


Figure 5.16: Concentration of ANV by sucrose cushion. PEG precipitated ANV from infected CK cell extracellular media was centrifuged through sucrose cushions of the indicated concentrations. Material that pelleted through the sucrose was resuspended, and the ANV genome copy number quantitated by RT-qPCR. The fold increase in absolute genome copy number is presented relative to the resuspended PEG precipitated ANV infected extracellular media which formed the input. The average genome copy number and the standard deviation of three technical replicates are shown for each treatment.

5.3.4.3 Morphology of sucrose density purified ANV

To assess if purified ANV virions are intact and a suitable substrate for PS studies, the morphology of purified ANV was analysed by TEM. Particles of ANV present in the extracellular media of infected CK cells were first concentrated by centrifugation onto a 60% sucrose cushion. The interface layer above the 60% sucrose cushion was extracted by pipette and diluted in PBS, so that the final concentration of sucrose was below 20%. This was then loaded onto, and centrifuged through, a 20/40/60% sucrose step gradient (Section 2.2.8.2). The resulting gradient was fractionated from the top, taking 1 ml volumes with a pipette. Absolute genome copies in each fraction were then determined by RT-qPCR quantification.

The highest quantities of ANV RNA were recovered from fractions 8 and 9, corresponding roughly to the 20-40% sucrose steps (Figure 5.17 A), consistent with previous data (Figure 5.16). As fraction 9 contained the peak of ANV RNA detection, the presence and morphology of ANV particles in this fraction was investigated by TEM. Particles of

approximately 30 nm size lacking any star-like morphology were present in the purified material, with some appearing hollow or partially complete (Figure 5.17 B). These particles may be putative ANV virions as their size and morphology are consistent with that of particles observed in ultrathin sections of ANV infected CK cells (Figure 5.14 C). The putative ANV particles observed in this gradient fraction were present in low numbers and aggregated clusters potentially due to the sample preparation or conditions used, suggesting that although it may be possible to purify ANV using sucrose density centrifugation, the resulting material may be unsuitable for use in SELEX or high resolution cryo-electron microscopy to identify PS-CP interactions. However, PEG or sucrose concentrated ANV particles would be suitable for identification of PSs using NGS techniques, as this technique is not sensitive to virion aggregation, and packaged viral RNA within virions has been separated from the total RNA population (including packaging defective genomes) within cells.

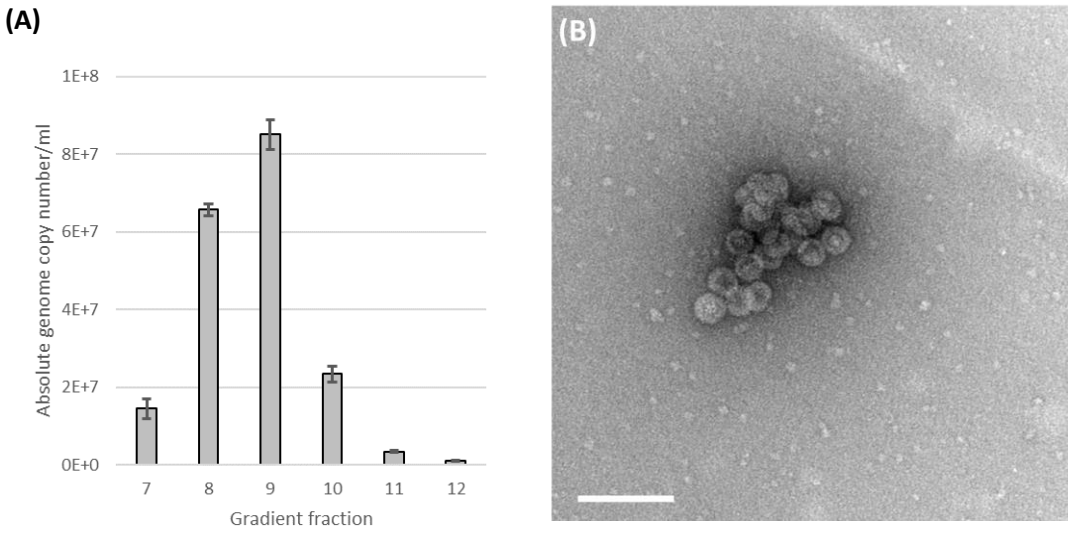


Figure 5.17: Ultrastructural analysis of sucrose purified ANV particles. ANV virions within extracellular media of infected CK cells were concentrated onto a 60% sucrose cushion by centrifugation, viral material was extracted and further purified by sucrose density gradient centrifugation. (**A**) Fractions were taken from the top of the gradient and the quantity of ANV RNA in each was determined by RT-qPCR. The average absolute genome count is shown with the standard deviation of three technical replicates. (**B**) The gradient fraction 9, containing the highest levels of ANV RNA, was imaged by TEM, revealing the presense of aggregates of ~30 nm ANV virions. Scale bar = 30 nm.

5.4 Summary

Collectively, the data presented here indicate the successful construction of a reverse genetics system for ANV G-4260, from which infectious ANV was recovered after transfection into BSRT7 cells (Figure 5.6). This system is similar to that reported by Imada *et al.,*[211] but exploits the constitutive expression of T7 RNA polymerase in BSRT7 cells to drive production of capped viral RNA for translation and virion assembly. This system has the advantage of producing authentic ANV virions in an easily transfected immortalised cell line that can then be used for infection studies in relevant primary cells (CK cells). Whilst sequencing of the ANV cDNA clone constructed here revealed 14 non-synonymous mutations with respect to the reference sequence (GenBank ID: AB033998.1), these mutations did not have a noticeably detrimental effect on the ability to recover virus from the reverse genetics system or on subsequent viral replication and pathogenesis in primary CK cells (Figure 5.7, Figure 5.8 and Figure 5.10). This suggests that together, the mutations identified in pANV represent a viable viral genotype and that the mutations relative to the reference sequence do not catastrophically impair any essential viral processes, including transcription and encapsidation. Interestingly, the majority of mutations were identified in ORF1a which may suggest the functions encoded within this protein are not as essential in cell culture in comparison to the ORF1b RdRp, in which just one non-synonymous mutation was identified.

Replication of reverse genetics system derived ANV was established in several avian cell types, including primary kidney and fibroblast cultures, as well as immortalised liver and fibroblast cells (Figure 5.11). Consistent with clinical pathology[336], the highest yield of ANV genomic RNA (as a proxy for replication) was seen in primary CK cells derived from 2-3 week old birds. Furthermore, out of the cell types assayed, CPE was seen exclusively in CK cells, in line with *in vivo* pathology data regarding the presence of gross lesions limited to the kidneys of infected chicks[336]. Whilst immortalised LMH cells also produced high levels of ANV genomic RNA, immunofluorescence staining highlighted the presence of more numerous intracellular foci of ANV proteins compared to CK cells (Figure 5.12). This may be consistent with a prior report proposing that a majority of ANV virions are retained within the cytoplasm of LMH cells[91], meaning CK cells may be the best model in which to study the complete infectious cycle of ANV, as these cells will most closely recapitulate events in the primary target tissue of ANV. Recognition of the reverse genetics system derived ANV (in immunofluorescence and western blot) by a chicken antiserum from a bird infected with a circulating field strain of ANV suggests that the cloned ANV has a similar if not identical

serology to circulating field strain wild-type ANV (Figure 5.12 and Figure 5.13). This suggests that reverse genetics system derived ANV may be a suitable substrate for identifying PSs and that data generated using this approach would be relevant to field strains.

To perform PS identification by SELEX, ANV CP must be purified away from contaminants and concentrated. Attempts to purify ANV by PEG precipitation and sucrose density centrifugation resulted in aggregated clusters of viral particles that could not be re-suspended (Figure 5.17). These purification products are therefore unsuitable for RNA SELEX as monodispersed purified virions in solution would be required to generate appropriate substrates for screening the RNA aptamer library[53], and for counter selection to remove aptamers that recognise the exterior of the capsid. However, these aggregated clusters of ANV particles represent a concentrated source of ANV, which can be used for alternative techniques to SELEX. In particular, such purified particles would be suitable for NGS studies to identify differences between packaged and unpackaged RNA populations, that could identify RNA sequences and secondary structures important for capsid assembly and virion formation, as recently described for FMDV[80].

Chapter 6

6 Identification and validation of putative packaging signals in the ANV genome

6.1 Packaging signal-mediated genome encapsidation

Packaging signals, which recruit capsid protein at defined positions across the viral RNA, and act as a scaffold for the assembly of the viral capsid around the genome in an ordered manner to achieve complete encapsidation, have been identified in a number of +ssRNA viral genomes[43, 52, 55, 58, 59, 62-65]. Identification of these RNA structures has predominantly been performed using an RNA SELEX technique[53, 55, 58, 78] (Section 1.2.6), however, a NGS approach was also shown capable of identifying putative PSs[80]. The SELEX technique of identifying putative PSs by screening RNA aptamers against viral protein requires purified assembly-competent CP derived from heterologous expression systems, either as capsomers or as partially disassembled/reassembled VLPs, or disassembled/reassembled virions. This method therefore relies on the ability to partially reassemble capsids from CP. Since attempts to produce assembly competent ANV capsomers, VLPs, or purified ANV virions were unsuccessful (Sections 4.5.2 and 5.3.4.3), an alternate approach in identifying PSs within the ANV genome must be adopted.

One suitable approach was recently described for the identification of PSs in FMDV. Using this method, subtle differences in preferred sequence at individual nt positions between viral quasispecies of packaged and unpackaged viral RNA populations were found by NGS and calculation of the entropy score for nts at each genome position[80]. Specific regions in the genomic sequence of virion RNA were found to be conserved, and secondary structure predictions revealed a common stem-loop motif within these regions that were shown to be functional PSs by mutational studies[80].

A similar NGS and entropy approach was used here for ANV to identify putative PSs within the genome. This NGS approach, coupled with bioinformatics analysis, was applied to intracellular ANV RNA and extracellular virion RNA populations produced from the reverse genetics system to identify RNA sequences and structures that were selected for in the encapsidated genomes, identifying putative ANV PSs for *in vitro* validation.

6.2 Isolation of ANV cDNA for NGS

For NGS-based identification of PSs, ANV RNA was purified from infected cells (representing both packaging competent and incompetent forms) and from PEG precipitated extracellular virions (consisting exclusively of packaging competent forms) (Section 2.2.9.1) as summarised schematically in Figure 6.1.

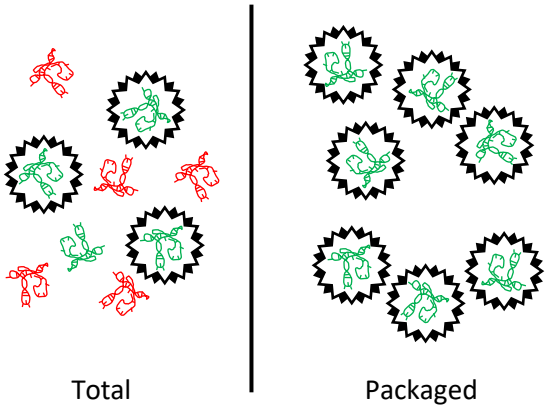


Figure 6.1: Differences in the RNA populations of ANV infected cells compared to ANV virions. Schematic representation of the quasispecies to be compared. Within infected cells, both packaging competent (green) and incompetent RNAs (red) are present, whereas within virions only the packaging competent species (green) are represented. The presence of multiple, functional PSs in the genome are presumed to determine packaging competency. Therefore, comparison by NGS of the populations may reveal differences in the sequences associated with putative PSs that promote their encapsidation.

To achieve sufficient read depth in the NGS data, enabling identification of putative PSs by conservation, a high concentration of nucleic acid input is required. Therefore, conversion of ANV RNA to cDNA and PCR amplification of full-length genomes was undertaken to maximise the number of ANV transcripts available to sequence. Purified RNA was used as a template for reverse transcription to produce cDNAs corresponding to the entire 6.9 kb genome of ANV. Initially, cDNA synthesis from total infected cellular RNA, and virion RNA from ammonium sulphate and PEG precipitated extracellular media was performed using random hexamer primers. The first-strand cDNA synthesised in this way was then amplified by PCR using primers targeting sequences at the very 5' and 3' ends of the ANV genome (Table 2.3). Using this approach, no DNA amplification was observed in RT-PCR reactions,

where total infected cellular or ammonium sulphate precipitated virion RNA was used as the template for reverse transcription (Figure 6.2 A). In contrast, a ~6.9 kb DNA, corresponding to the expected size of the ANV genome, was amplified from the PEG precipitated virion RNA sample (Figure 6.2 A). These data confirmed that full-length cDNA synthesis of the ANV genome using virion-derived RNA is possible. The lack of DNA amplification in reactions using total infected cellular RNA was unexpected and may be due to competition between excess cellular RNAs and ANV genomic RNA for random hexamer annealing, leading to the inefficient production of viral cDNA. The absence of DNA amplification from ammonium sulphate precipitated virions is unsurprising as prior data showed ammonium sulphate to poorly concentrate ANV virions (Section 5.3.4.1).

To optimise the production of genomic-length first-strand cDNA by reverse transcription, reactions were performed using alternative primers, including oligo(dT)$_{20}$ and a GSP targeting the 3' end of the ANV genome. The synthesis of genomic-length cDNA was investigated, using primers targeting sequence in the 5' and 3' ends of the ANV genome for PCR amplification. Whilst the GSP-primed reverse transcription did not produce a genomic-length ANV cDNA, random hexamer- and oligo(dT)$_{20}$-primed reverse transcription produced genomic-length ANV cDNA amplicons (Figure 6.2 B). Of the successful primers, oligo(dT)$_{20}$-primed reverse transcription appeared to be the most efficient. Smaller bands of cDNA were also observed, likely representing non-specific amplification of host-cell nucleic acid by the primer sets used.

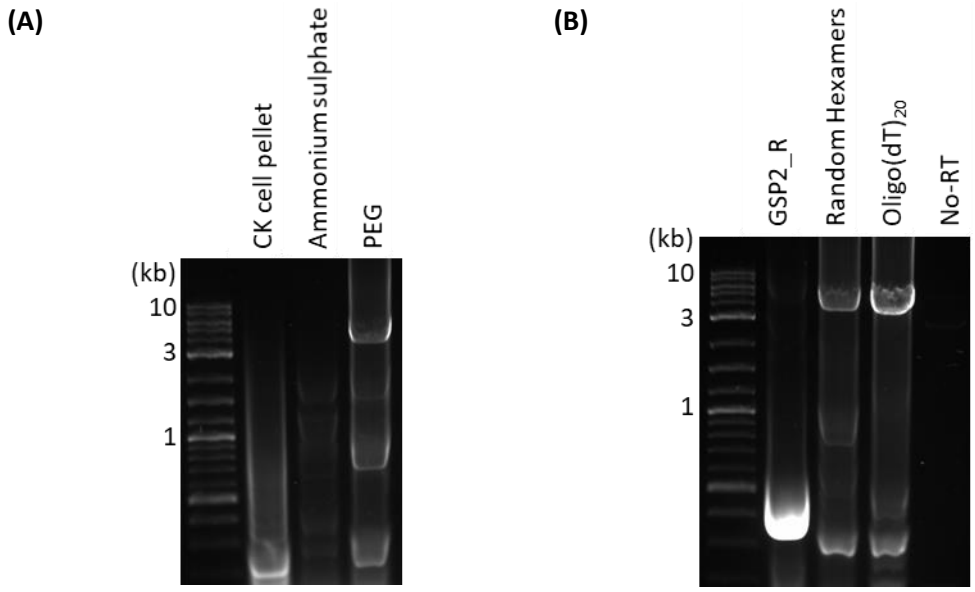


Figure 6.2: Optimisation of genome-length cDNA synthesis for ANV. (**A**) RNA was extracted from the total intracellular RNA of ANV infected CK cells, and from ammonium sulphate or PEG precipitated extracellular virions. Reverse transcription was then performed using random hexamers, and cDNA was PCR amplified using primers targeting the 5' and 3' ends of the ANV genome. A 6.9 kb DNA corresponding to genomic-length ANV was amplified using RNA derived from PEG precipitated ANV virions. (**B**) Genome-length reverse transcription efficiency of ANV RNA by either a GSP, random hexamers, or oligo(dT)$_{20}$ was assayed, using RNA extracted from PEG precipitated extracellular virions as template RNA. Random hexamer and oligo(dT)$_{20}$ primed reactions produced genomic-length ANV transcripts, which were not present in the absence of reverse transcriptase.

Using the optimised oligo(dT)$_{20}$ protocol for genome-length ANV cDNA synthesis (Section 2.2.9.2), the reactions to amplify ANV cDNA from the total intracellular RNA of ANV infected cells and extracellular virion genomic RNA were repeated and PCR amplified using primers targeting the 5' and 3' ends of the genome. Consistent with random hexamer-primed reverse transcription, a ~6.9 kb DNA product, at the expected size of the intact ANV genome, was only produced in the RT-PCR reaction containing RNA derived from PEG precipitated ANV virions (packaged), but not from total infected cellular RNA (total) (Figure 6.3 A). Therefore, further optimisation was needed to improve synthesis of genomic-length ANV DNAs from total intracellular RNA populations to enable the identification of putative PSs by NGS. Up to this point, reverse transcription product equivalent to 10% (v/v) of the PCR reaction was used as the template for amplification of genomic-length ANV DNA. As excess nucleic acid can be inhibitory to PCR, reactions were altered such that the reverse transcription products made up only 4% of the final PCR reaction volume, whilst the

amplification protocol otherwise remained unchanged (Section 2.2.9.3). Additional PCR primer pairs were also tested (Table 2.3). In one, the 3' reverse GSP_R2 was replaced with oligo(dT)$_{20}$ for genome-length amplification, whilst in others an internal reverse (Mid1_R) was paired with GSP_F2, and an internal forward primer (Mid2_F) was paired with GSP_R2. Internal primers were sited so as to produce overlapping DNAs (Figure 6.3 B). Under these conditions, products corresponding to either the ~3.5 kb halves (produced using internal oligos) or genomic-length ANV DNA (GSP oligos only) were amplified from both the total and packaged RNA populations (Figure 6.3 C). However, amplification of genomic-length ANV cDNA was unsuccessful in reactions primed with the forward sense GSP_F2 and oligo(dT)$_{20}$.

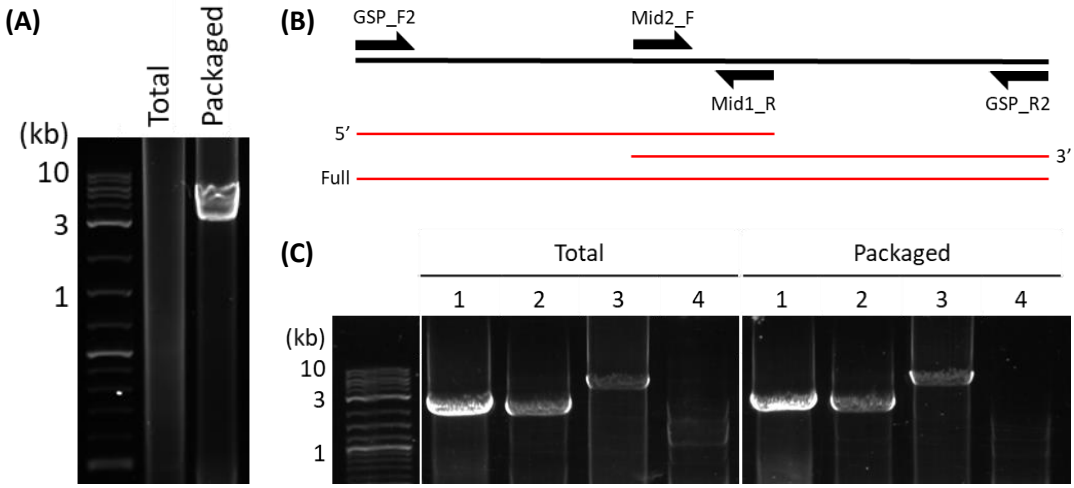


Figure 6.3: Optimisation of genome-length PCR for the ANV genome. (**A**) RNA was extracted from ANV infected CK cells (total) and PEG precipitated extracellular virions (packaged), reverse transcribed using oligo(dT)$_{20}$, and cDNA amplified by PCR using primers targeting the 5' and 3' ends of the ANV genome. Genomic-length ANV cDNA was only amplified from PEG precipitated virion-derived RNA. (**B**) Location and orientation of GSPs used for PCR amplification are shown, with the regions amplified by each pair illustrated in red below. (**C**) Reduction of the amount of first-strand cDNA synthesis products to 4% of the final PCR reaction volume resulted in the amplification of cDNAs of the expected size for both the 3.5 kb 5' (1) and 3' (2) halves of the ANV genome, as well as the 6.9 kb genomic-length transcript (3), in both the total and packaged RNA populations, that were all amplified using the primer sets in (B). No 6.9 kb ANV cDNA was amplified in the genome-length PCR reactions primed with forward sense GSP and oligo(dT)$_{20}$ (4).

cDNAs from these reactions were gel purified and extracted using the Monarch® DNA Gel Extraction Kit (Section 2.2.9.4). Recovered DNA quanity was then assessed using a Qubit Fluorometer (Invitrogen) (Table 6.1).

Qubit quantification results (ng/µl)						
Sample	**1**	**2**	**3**	**4**	**Average**	**Dilute to 0.2 ng/µl**
ANV Total 5’	15.1	15.1	14.9	14.8	14.98	1 in 73.9
ANV Total 3’	18.0	18.0	18.0	18.0	18.0	1 in 89
ANV Total 7kb	11.1	11.1	11.0	11.0	11.05	1 in 52.25
ANV Packaged 5’	13.9	13.8	13.8	13.8	13.8	1 in 68
ANV Packaged 3’	25.6	25.5	25.1	24.9	25.23	1 in 125
ANV Packaged 7kb	12.9	12.9	12.8	12.8	12.85	1 in 63.25
Water control	-	-	-	-	-	N/A

Table 6.1: Quantification of purified ANV DNAs for NGS. The concentrations of purified DNAs for the packaged and total ANV genome populations were quantified using a Qubit Fluorometer (Invitrogen). The individual dilution factors of each DNA required for the optimal NGS concentration were calculated.

6.3 NGS and bioinformatics analysis of the packaged and total ANV genome populations

6.3.1 Genome coverage and read depth in the sequenced ANV populations

Purified ANV DNAs (Figure 6.3 C) were processed to create libraries of fragments for sequencing using the Illumina MiSeq (Section 2.2.10.1). Sequence data generated by the MiSeq was then analysed and processed as described in Section 2.2.10.2. Briefly, reads with a Phred quality score > q38 were selected, equating to a probability of 1:10,000 for an incorrect base call, i.e. 99.99% accuracy. Next, the first 5 and last 15 nts for each read > 70 nt in size were trimmed using Sickle to remove lower quality sequence at either end. These trimmed reads were then assembled into contigs using the pANV cDNA clone as the reference sequence. An Rscript, eleano.r (R. Enriques-Gasca, TGAC, Norwich) was then run in order to output the ensemble of nts identified at each position along the genome, in each of the assembled sequence datasets.

Analysis of the eleano.r output for the contigs of packaged and total ANV populations revealed that coverage of the genome spanned nts 25-6884 out of a potential 6927 nts, with an average read depth of 4510.42 and 4522.52 for the packaged and total assemblies, respectively (Figure 6.4). The patterns of read depth for both populations were very similar, indicating the two datasets could be reliably compared. Further analysis was carried out for positions with a read depth of > 2000 in order to reduce potential bias from under-represented regions, particularly at the terminal regions.

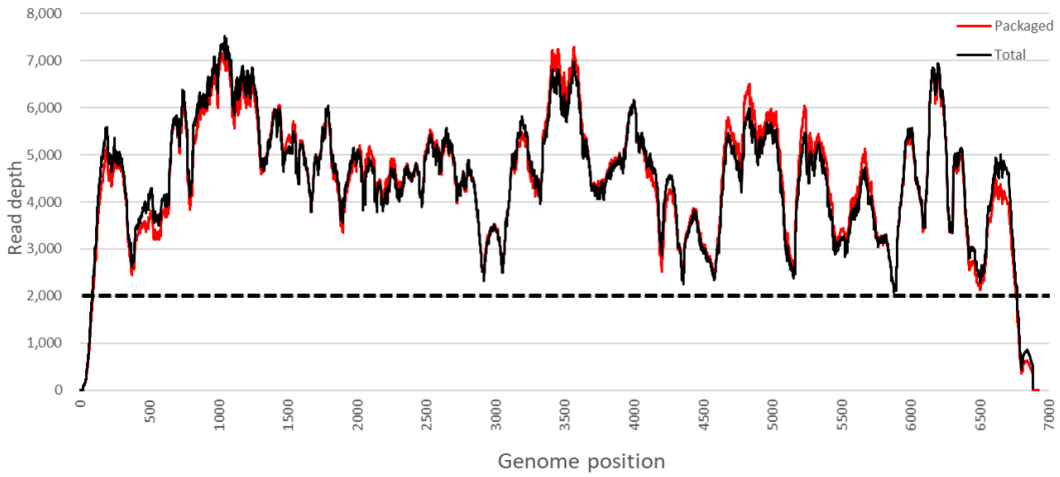


Figure 6.4: Read depth across the ANV genome from packaged and total RNA populations. High quality sequences were selected from the NGS reads, and these were then assembled into contigs for each RNA population (packaged or total). The eleano.r script was used to output the ensemble of nucleotides at each genome position in each contig, from which read depth at each position was then calculated. Read depth was plotted and compared between the packaged (red) and total (black) populations. The dashed line represents the 2000 read depth cut-off point, data above which was used for further analysis.

6.3.2 Identification of putative PSs by comparison of packaged and total ANV RNA populations

To identify conserved sequences within packaged ANV genomes that may correpsond to putative PSs, analysis of the difference in the ensemble of nts at each genome position between the packaged and total populations was attempted using entropy scores as done for FMDV[80]. Shannon's entropy was calculated to measure levels of variation at each nt position in the genome (Section 2.2.10.3). Values for the packaged population were then subtracted from those of the total population and positive values were plotted on a graph (Figure 6.5 A). Positive values represent nucelotides which are more conserved in the

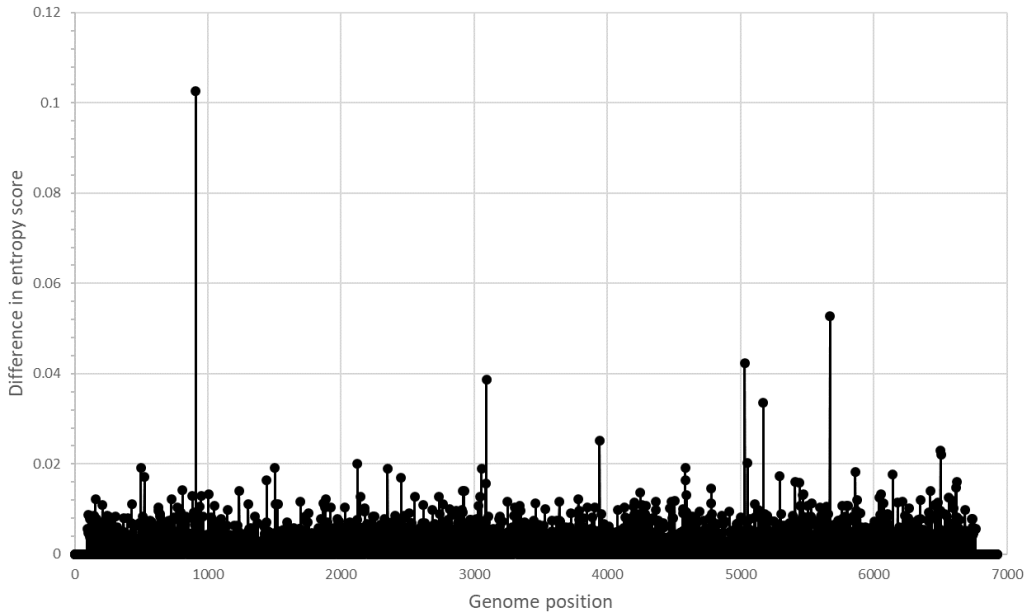
packaged population than in the total population, thus identifying regions which may be involved in genome encapsidation, and therefore contain putative PSs. Using this approach, several peaks were identified across the ANV genome. However, there was a large background noise of smaller peaks, rendering identification of the significant positions hard to achieve. The low levels of variation present at each nt position in the assembled contigs may mean that many of the entropy values are over represented, and thus are not all meaningful. With very few mutations present at each genome position (average of 1.69 and 1.77 base variants per position for packaged and total datasets, respectively) relative to the read depth (Figure 6.4), it becomes unclear whether these variations are derived from ANV mutation or from artefacts (i.e. introduced by the RT, PCR or NGS steps). This low level variation may also be a result of the ANV genetic material used for NGS, which was derived from virus recovered from the reverse genetics system, and consequently, genomes of ANV will have started as a uniform population. This, alongside the limited number of passages (~9) between recovery and sequencing, means that only a limited amount of genetic variation will have been introduced by the error-prone viral RdRp over this time.

Therefore, to reduce noise from low frequency variation, and to produce a dataset with reliable peaks for conservation analysis, the mutation rate at each genome position was instead calculated (Section 2.2.10.4). This identifies positions with significant variation, and minimises background from potential artefacts. The mutation rate, as a percentage, for each genome position with > 2000 read depth in the packaged population was subtracted from that of the total population, and positive values were plotted, which represent nts with higher conservation in the packaged population (Figure 6.5 B). This analysis identified 6 positions, where the reference sequence (deemed wild type, wt) nt is favoured in packaged genomes. These peaks were mostly consistent with those observed in the entropy score difference, however the noise from low frequency variation was greatly reduced. Analysis of each peak, numbered in the order of occurrence, shows that putative PS1 is the most conserved, with selection for the wt nt 5% more frequently in packaged genomes than those in the total population. Putative PSs 2 and 6 show > 1% more conservation of the wt nt in virions, whilst PSs 3, 4 and 5 show just 0.33, 0.73 and 0.57% greater conservation, respectively.

Upon closer inspection, all three of the positions with > 1% mutation rate difference (nts 909, 3092 and 5672) showed evidence of genetic drift or positive selection, where there

has been emergence of a second dominant nt at that position. For example, at genome position 909, 18.56% and 23.99% of sequence reads contain the new nt species in the packaged and total populations, respectively. This suggests that the wt nt at this position is enriched in the packaged population, potentially indicating a role in genome encapsidation. The mutant might have some selective advantage other than packaging in the viral lifecycle, such as adaption to cell culture.

(A)



(B)

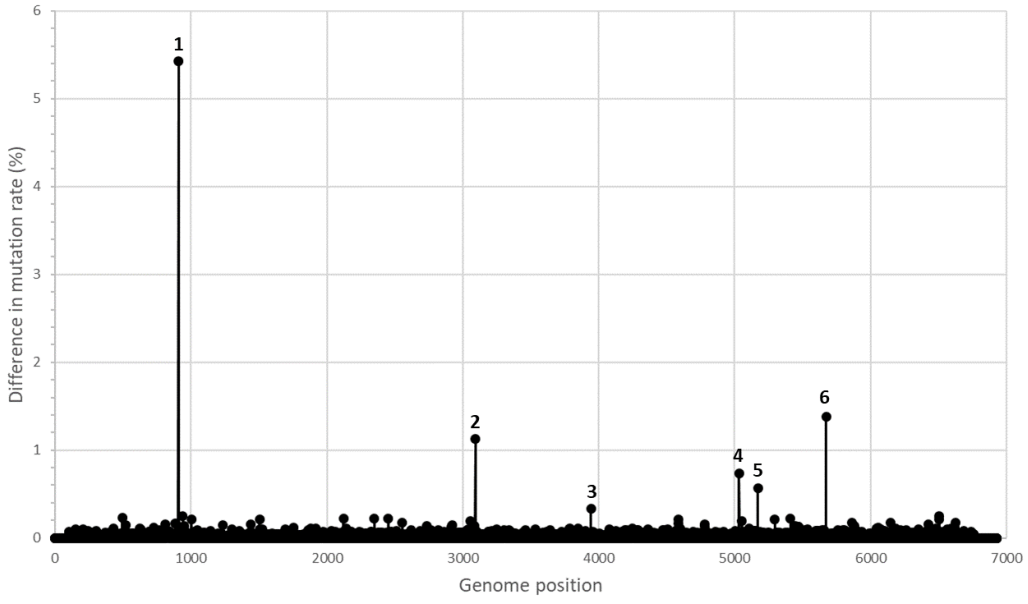


Figure 6.5: Comparison of variation at each genome position between packaged and total ANV populations. NGS reads for the populations of packaged and total ANV genomes were processed and assembled into contigs, the eleano.r script was then run on each contig to output the ensemble of nucleotides at each position along the genome. (**A**) Entropy at each position was calculated and the values for the packaged population were subtracted from that of the total population, and positive values representing nts more conserved in the packaged genomes were then plotted. Due to the low levels of variation at some of these sites, not all peaks may be meaningful. (**B**) The mutation rate at each position was calculated as a percentage, values for the packaged population were subtracted from those of the total population to give the difference in mutation rate. Positive values representing positions that are favoured in packaged genomes are plotted. This approach greatly reduced background noise from low frequency variation when compared to the entropy calculations. The putative PSs identified are labelled 1-6 in order of occurrence from 5' to 3'.

6.3.3 RNA secondary structure analysis of putative ANV PSs

To investigate the significance of the peaks identified by analysis of the mutation rate difference (Figure 6.5 B), the RNA secondary structures of the three most significant sites (PSs 1, 2, and 6), with > 1% more nt conservation in packaged genomes, were predicted using RNAstructure[339]. The ensemble of folds with negative free energy, and thus able to form energetically stable structures, were compared between the identified positions (+/- 60 nt). This revealed a similar SL motif containing a 3' GCAA sequence, as well as pyrimidine-rich (predominantly uracil) bulges in the stem located 3-5 base-pairs from the loop (Figure 6.6). PS6 does not have a full GCAA loop motif, as the last adenine nt is predicted to form part of the stem structure (denoted by brackets). However, it is still possible that the full GCAA motif in PS6 could be presented, as hydrogen bonds between base-pairs will naturally dissociate and form, particularly in the weaker A-U and G-U pairings. This could also be promoted by CP binding to the recognition motif.

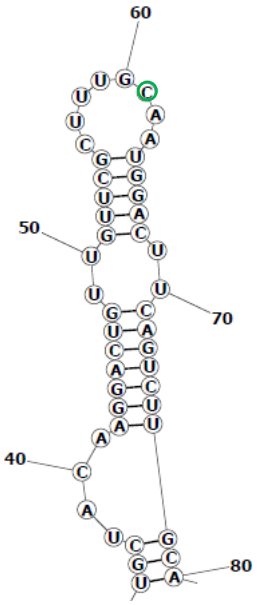
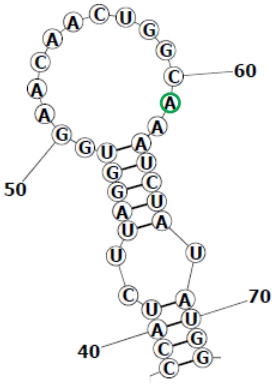
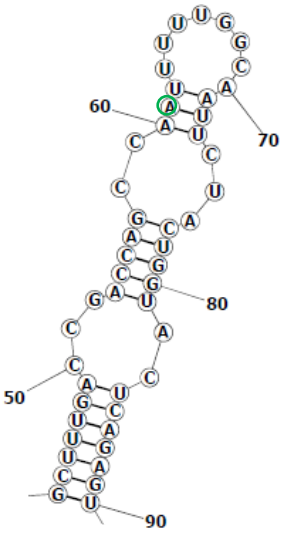
Peak	PS1	PS2	PS6
RNAstructure prediction			
Genome position	904-911	3080-3093	5674-5681
Loop motif	CUUUGCAA	GGAACAACUGGCAA	UUUUGGCA(A)
Side bulge	5 – “UU” “UU”	5 – “CU” “U”	3 – “CC” “CUA”

Figure 6.6: RNA secondary structure analysis of putative PSs with > 1% more nucleotide conservation in the packaged genomes. RNAstructure was used to predict the secondary structures of RNA at regions corresponding to the peaks of > 1% more nt conservation in packaged genomes (circled in green). A common SL motif was identified, each containing a 3’ GCAA sequence, as well as pyrimidine-rich bulges in close proximity to the loop. For PS1 and 2, the conserved nt is located in the GCAA motif, whereas in PS6 the conserved nt is predicted to base-pair in close proximity to the loop motif, stabilising its structure.

As a common SL motif was found located at each of the nt peaks with a difference of > 1% in mutation rate, the remaining 3 peaks (PSs 3, 4 and 5) with < 1% difference were then investigated to see if they also contained similar features. RNA structure analysis indicated that the nts corresponding to these peaks are not present in a loop containing the

identified GCAA motif, nor are they present in a stem of a neighbouring loop that does (Figure 6.7). However, PS3 presented a 3' CAC motif, and could potentially represent a variant of the putative PS. Overall, this data indicates that a 1% difference in mutation rate at a given nt between packaged and total populations may be the cut-off point between representing a putative PS, and noise from low frequency variation in this dataset.

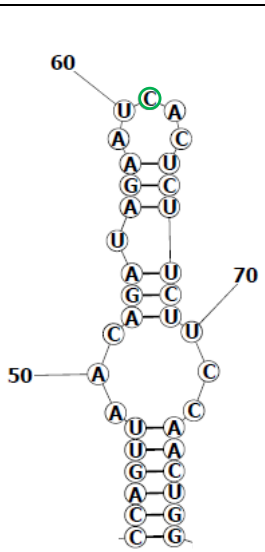
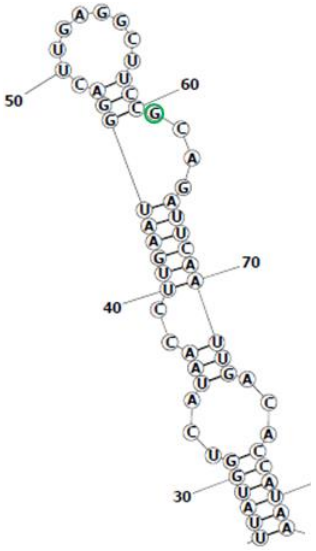
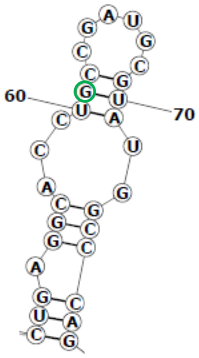
Peak	PS3	PS4	PS5
RNAstructure prediction			
Genome position	3940-3944	5021-5029	5173-5178
Loop motif	AUCAC	CUUGAGGCU	CGAUGC
Side bulge	3 – "U" " " 6 – "AAC" "UCC"	3 – " " "GCAG"	3 – "ACC" "UG"

Figure 6.7: RNA secondary structure analysis of putative PSs with < 1% more nucleotide conservation in packaged genomes. RNAstructure was used to predict the secondary structure of RNA at the remaining 3 regions corresponding to the peaks with < 1% more nt conservation in packaged genomes (circled in green). The previously identified 3' GCAA sequence motif was absent in each structure of PS4 and 5, suggesting these peaks may not represent putative PSs. However, folding of PS3 reveals a CAC motif, similar to the CAA motifs observed in PS1, 2 and 6.

Next, to investigate if there were any other SLs in close proximity to the identified putative PSs with similar features, RNAstructure was used to predict the RNA secondary structures within a sequence +/- 60 nt surrounding each peak position (Figure 6.8-Figure 6.11). This analysis shows that there may be additional putative PSs neighbouring the SLs identified by NGS and bioinformatics. In particular, adjacent to PS1 is a stem-loop containing a 3' CGA motif and a uracil-rich bulge in the stem 5 nt from the loop. This structure appears to share the key features of PS1, and therefore may represent an additional putative PS of ANV (Figure 6.8). A second SL neighbouring PS2 folds to present a similar loop sequence and pyrimidine-rich bulge as the putative PS. However, this SL contains a 3' GAAA sequence rather than the GCAA motif (Figure 6.9). RNA folding further indicates that there are two additional SLs with GCAA motifs, neighbouring PS3, even though the identified peak does not itself form a GCAA SL motif consistent with those derived from regions of > 1% greater conservation (Figure 6.10). Structural prediction of the sequence adjacent to PS6 shows yet again the presence of a second SL, containing a 3' CC(A) motif in close proximity to the putative PS, sharing at least two nts with the GCAA motif. However, as in PS6, the last adenine ribonucleotide of the motif is predicted to be involved in Watson-Crick base pairing, rather than exposed in the loop region (Figure 6.11 A). The folding of a third SL on the opposite side of PS6 is also predicted. However, a CAA motif is located at the 5' of the loop sequence, differing from the 3' consensus. Upon closer inspection of this SL, a sequence homologous to the loop sequences of PS1 and 6 is revealed, though it is predicted to be buried by base-pairing in the structure shown here. When examining the ensemble of higher free energy structures that this particular region can form, an 11 nt SL motif containing a 3' GCCA is identified (Figure 6.11 B). Furthermore, folding of the sequence surrounding PS4 and 5 did not reveal any SL structures with the putative 3' GCAA motif, or variants thereof, consistent with these sites being artefacts of low sequence variation and thus insignificant for encapsidation.

Taken together, it appears that there is at least one additional putative PS in the SLs surrounding the peaks of conservation identified by NGS, with a similar recognition motif (Table 6.2). The peaks with > 1% more conservation in packaged genomes all formed SLs with 3' GCA(A) motifs. These may represent a higher affinity version of the variants found in adjacent SLs (Table 6.2). The PSs of other studied viruses are found to contain degenerate SL motifs varying around a consensus sequence and structure[54, 55, 58, 78], some with higher affinity CP binding than others. Therefore, it is unsurprising that the putative PSs identified here for ANV follow the same pattern.

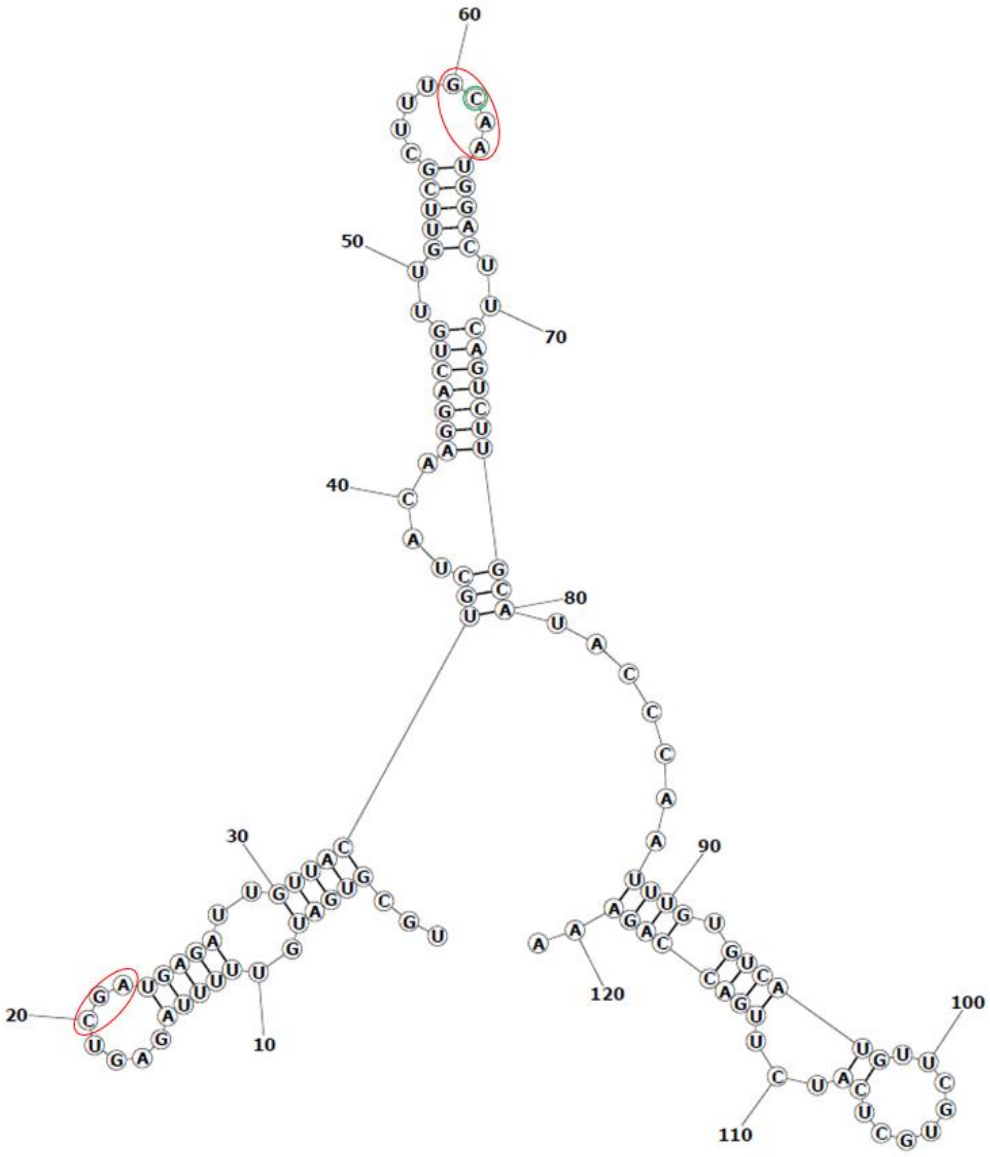


Figure 6.8: Prediction of neighbouring RNA structures surrounding putative PS1. The sequence surrounding PS1 (+/- 60 nt) was folded using RNAstructure to identify any predicted neighbouring structures that share the putative PS motif. A second SL 5' to PS1 was identified, containing a 3' CGA loop motif and a pyrimidine-rich bulge 5 nt from the loop. The putative 'GCAA' motif and its CGA variant are circled in red, and the conserved nt is shown individually circled in green.

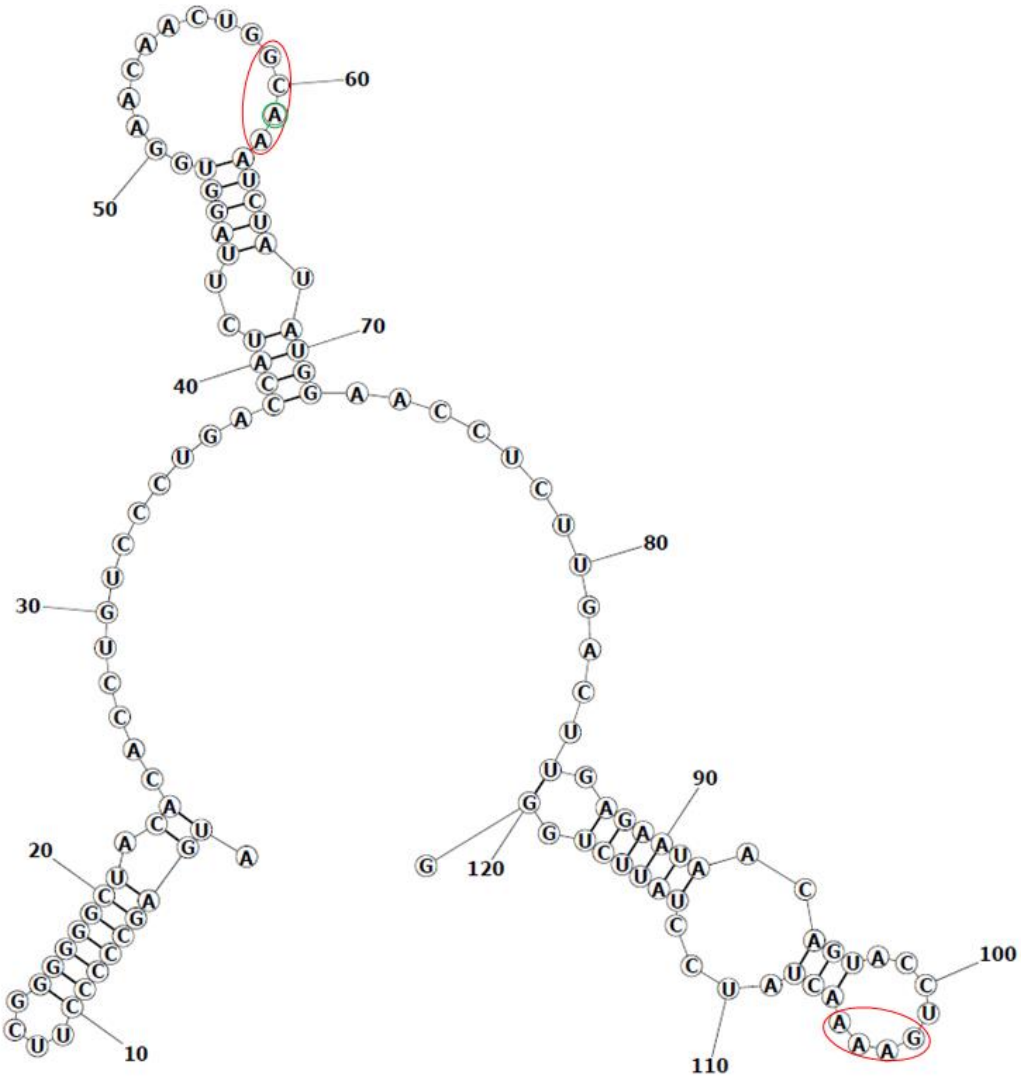


Figure 6.9: Prediction of neighbouring RNA structures surrounding putative PS2. The sequence surrounding PS2 (+/- 60 nt) was folded using RNAstructure to identify any predicted neighbouring structures that share the putative PS motif. A second SL 3' to PS2 was identified, containing a 3' GAAA motif and a pyrimidine-rich bulge 3 nt from the loop. The putative 'GCAA' motif and its GAAA variant are circled in red, and the conserved nt is shown individually circled in green.

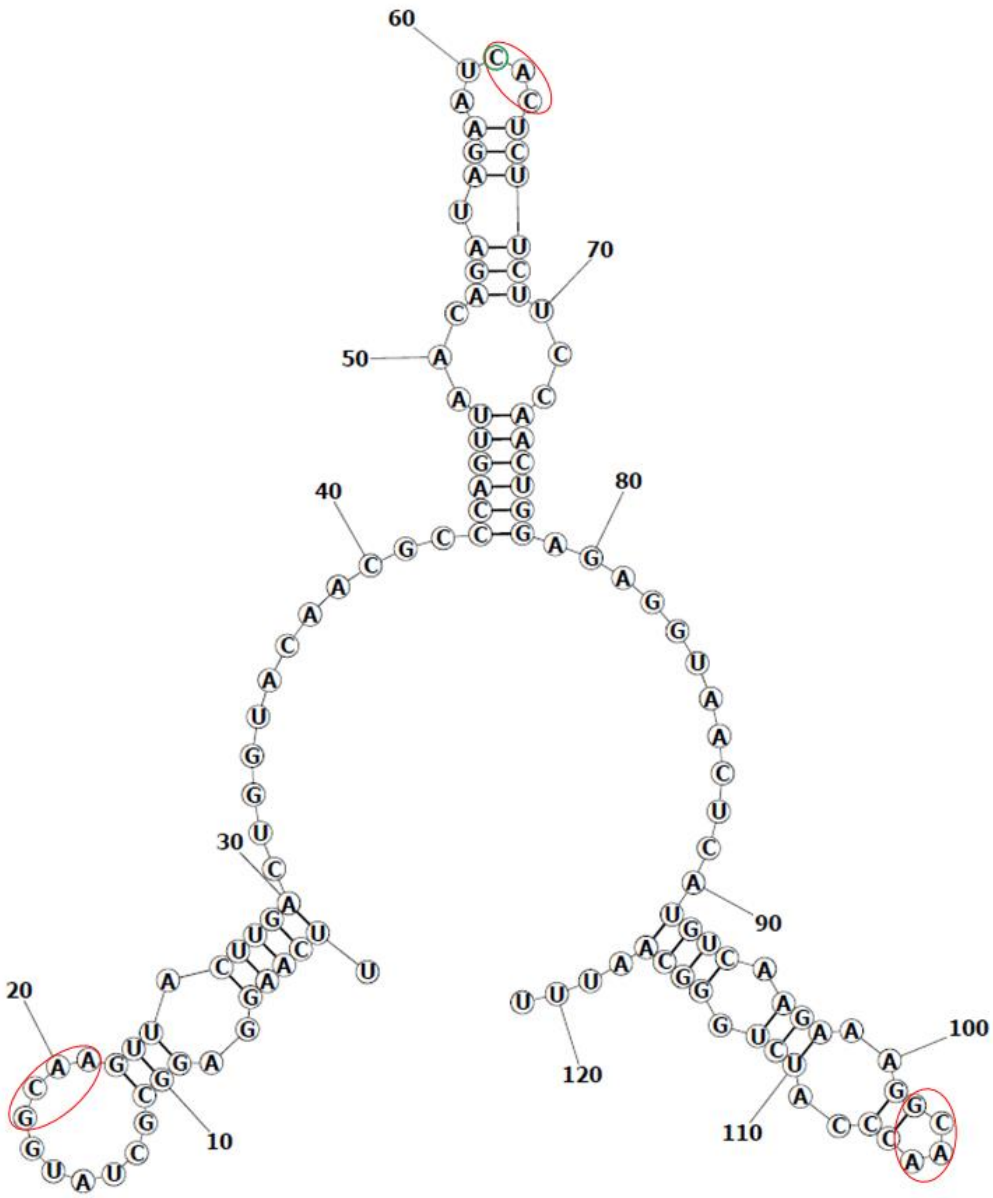


Figure 6.10: Prediction of neighbouring RNA structures surrounding putative PS3. The sequence surrounding PS3 (+/- 60 nt) was folded using RNAstructure to identify any predicted neighbouring structures that share the putative PS motif. Two SLs adjacent to PS3 were identified, both of which contain 3' (G)CAA loop motifs, however, pyrimidine-rich bulges in the stems are absent. The putative 'GCAA' motifs are circled in red, and the conserved nt is shown individually circled in green.

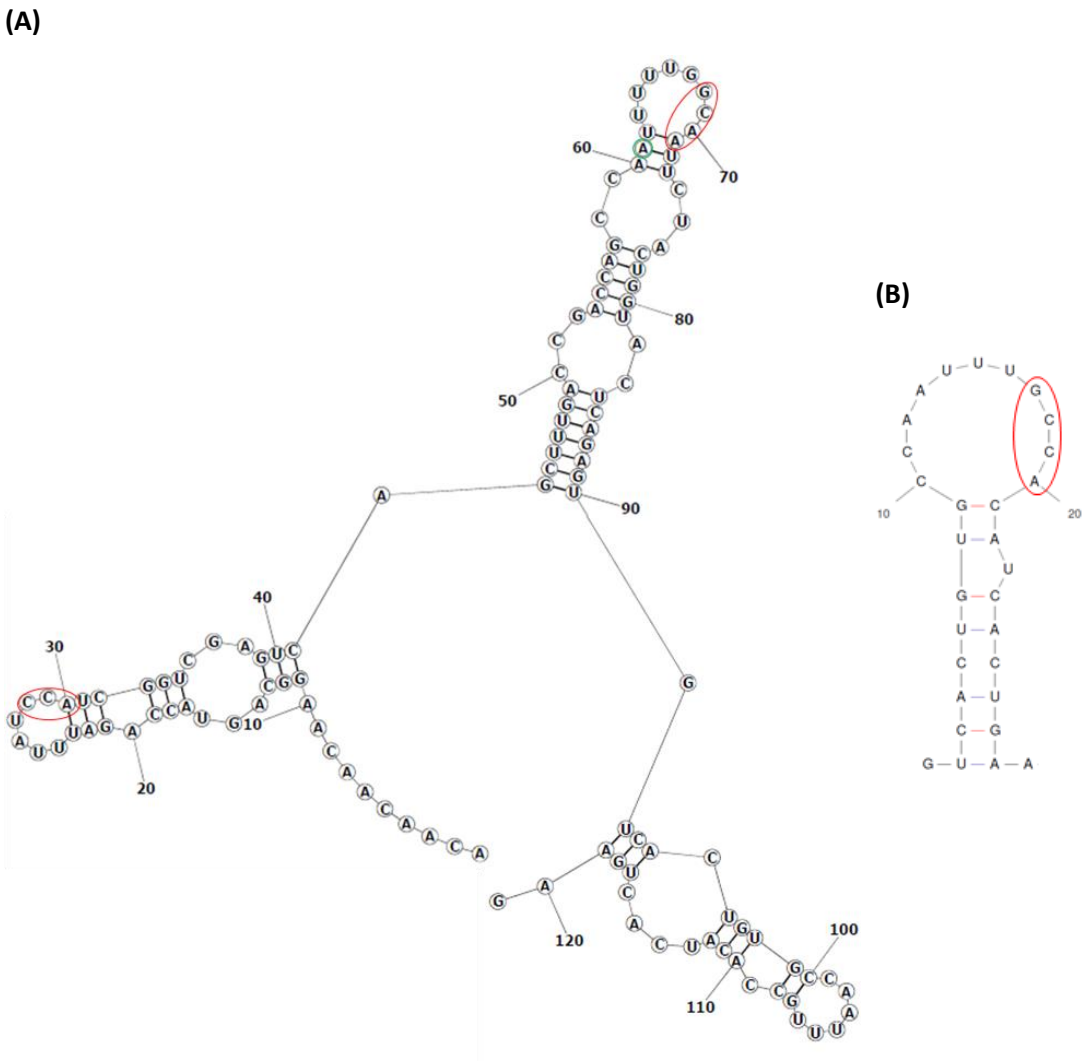


Figure 6.11: Prediction of neighbouring RNA structures surrounding putative PS6. (**A**) The sequence surrounding PS6 (+/- 60 nt) was folded using RNAstructure to identify any predicted neighbouring structures that share the putative PS motif. A second SL 5' to PS6 was identified, however, the last base of the 3' CC(A) loop motif was base-paired in the stem structure, akin to PS6 itself. A third SL is predicted to form in 3' to PS6, however, this does not contain the putative 'GCAA' motif. (**B**) Upon refolding of this region, an alternate SL is predicted to form, containing a 3' GCCA motif that was previously buried in the stem of the structure shown in (A). The putative 'GCAA' motifs are circled in red, and the conserved nt is shown individually circled in green.

Putative PS	Location of neighbouring SL	Motif
PS1	-	CUUUGCAA
	5’	GAGUCGA
	3’	UUCGUGCU
PS2	-	GGAACAACUGGCAA
	5’	UUCG
	3’	ACCUGAAA
PS3	-	AUCAC
	5’	GCUAUGGCAA
	3’	(G)CAA
PS6	-	UUUUGGCA(A)
	5’	UUAUCC(A)
	3’	CCAAUUUGCCA

Table 6.2: Sequence motifs of the putative PSs and neighbouring SLs.

6.4 *In vitro* validation of identified PSs

6.4.1 Design and construction of PS1 mutants

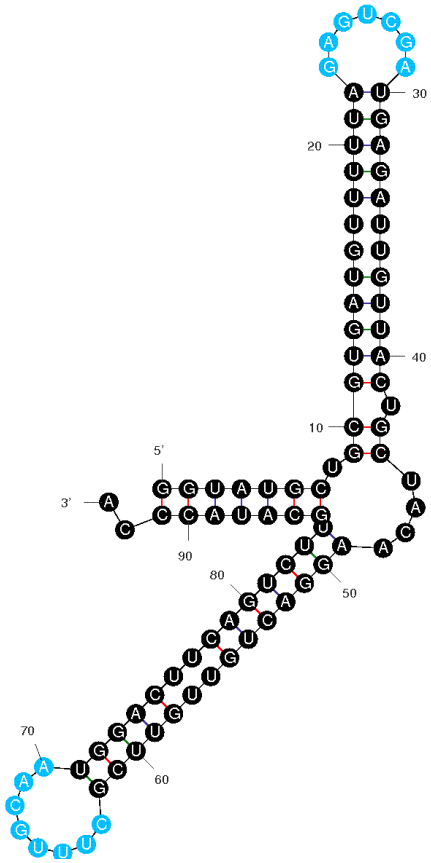
To assess the importance of the NGS-identified SLs in viral genome encapsidation, functional knock-out of the RNA structure was performed. This has previously been achieved by designing constructs with synonymous mutations to disrupt the sequence and structure of putative PSs, whilst encoding the wt amino acids. As it was not possible to synonymously disrupt the GCAA sequence motif of PS1, this wt PS was instead ablated by engineering in more energetically favourable alternative RNA structures with a lower free energy than that of the wt SL (Figure 6.12 and Figure 6.13). The identified neighbouring stem-loop, containing a 3’ CGA motif (Figure 6.8), was also simultaneously mutated, thus disrupting two potential putative PSs in this region of the genome. Disruption of multiple PSs is more likely to produce an effect on packaging efficiency due to the cooperative nature of multiple SL structures working synergistically. The designed structures incorporate the putative PS-motif, wholly or in part, into a tightly base-paired conformation to prevent the formation of the wt 36 nucleotide SL. The ensemble of alternative structures

for each mutant, with minimum free energies of up to 100% suboptimality, were also investigated to check that the putative PS loop motifs were not readily formed, confirming that each mutant had indeed knocked-down the identified PS.

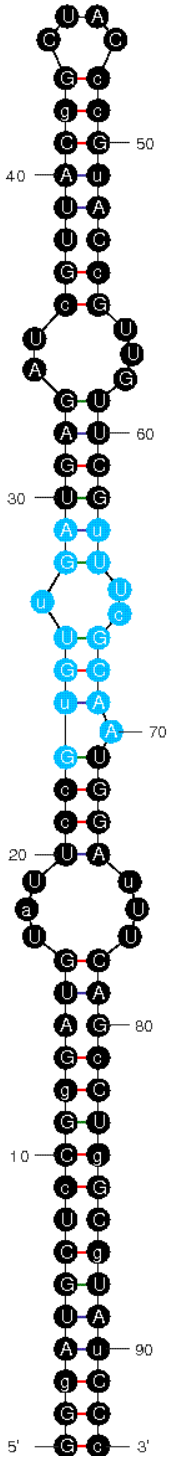
```
wt      GGTATGCTGCGTGATGTTTTTAGAGTCGATGAGATTGTTACTGCTAC 47
M1      GGgATGCTcCGgGATGTaTTccGtGTtGATGAGATcGTTACgGCTAC 47
M2      GGTATGCTGaGaGATGTcTTTcGcGTaGAcGAGATcGTTACTGCgAC 47
M3      GGTATGCTaCGTGAcGTTTTTcGcGTtGAcGAGATTGTcACcGCgAC 47
        ** ***** .* ** ** ** .* ** ** ***** ** ** ** **

wt      AAGGACTGTTGTTCGCTTTGCAATGGACTTCAGTCTTGCATACCCA  93
M1      ccGtACcGTTGTTCGtTTcGCAATGGAtTTCAGcCTgGCgTAtCCc  93
M2      ccGcACTGTTGTgCGtTTcGCgATGGACTTCtcTCTTGCATACCCA  93
M3      tcGtACcGTTGTcaGaTTcGCAATGGACTTCAGTCTcGcgTACCCC  93
         .* ** ***** .* ** **.***** ***:  ** **.** **.
```

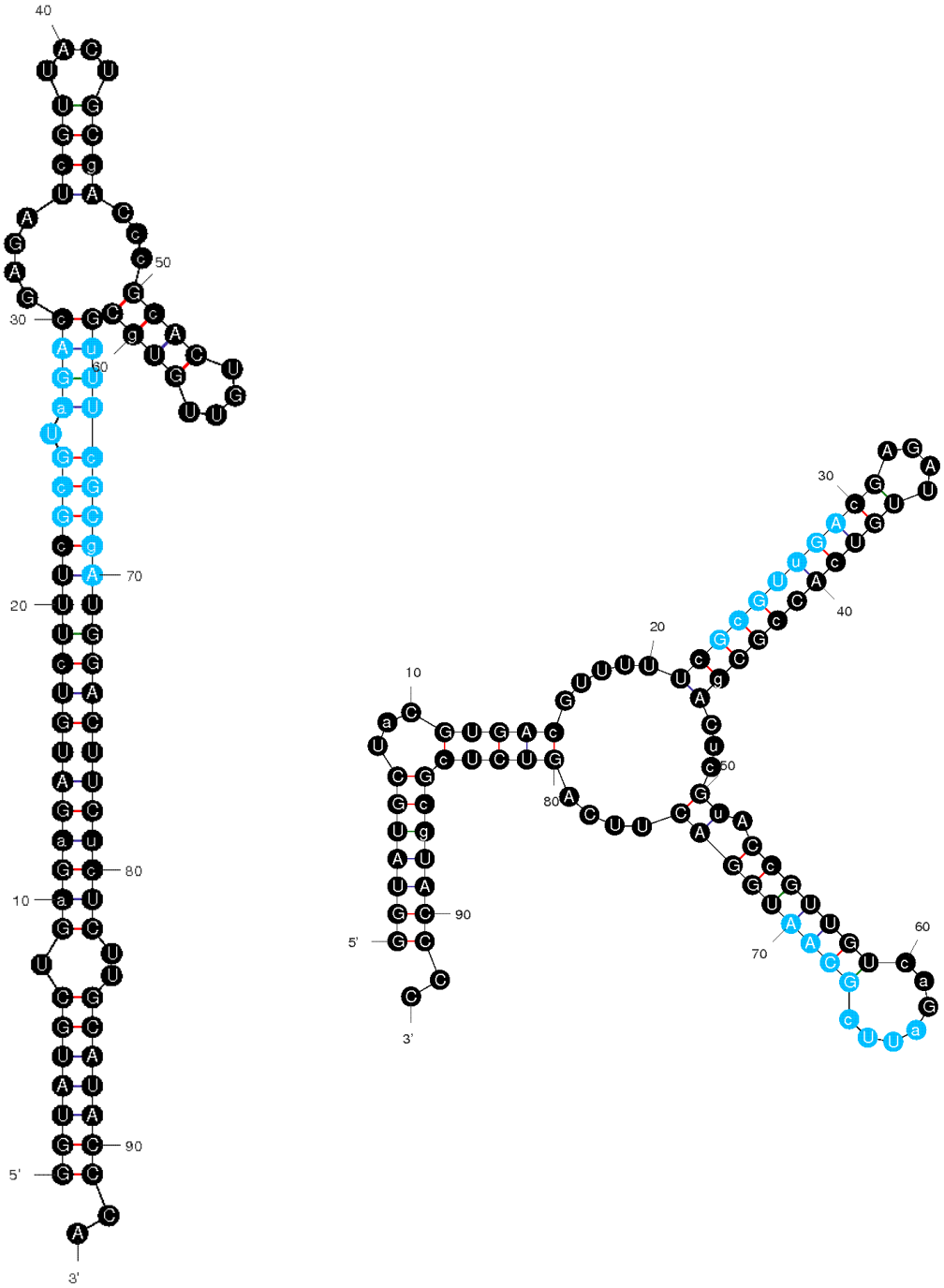
Figure 6.12: Alignment of designed PS1 mutant sequences against that of wt ANV. Synthetic DNA sequences for three mutants were designed to synonymously disrupt the sequence and structure of putative PS1 and a neighbouring SL in pANV. The alignment of the mutant PS1 sequences (mutations in lower case) with the wt sequence are shown, highlighting the locations of the PS1 loop motif (red border) and neighbouring SL loop motif (blue border).



PS1-wt



PS1-M1



PS1-M2 **PS1-M3**

Figure 6.13: Predicted structures of the wt and mutant PS1 RNAs. The minimum free energy structures of the wt and three designed mutant PS1 sequences were predicted using Mfold. These structures are shown with the loop sequences of the putative wt PS1 and neighbouring SL highlighted in blue. The free energies for each structure are -26.50, -41.00, -43.13 and -29.54 kcal/mol for the wt, M1, M2 and M3 mutants, respectively.

6.4.2 Cloning of ANV PS1 mutants

DNA corresponding to each of the mutant sequences (Figure 6.12) was synthesised (GeneArt) and cloned into the pANV reverse genetics construct as a precursor to rescuing PS1 mutant ANV (Figure 6.14 A). The wt PS1 sequence was removed from pANV by restriction digestion of flanking unique EcoRI and BstZ17I sites. NheI was additionally included in the digestion to split the pANV backbone into two sizes that could be distinguished from intact or partially digested parental plasmid. Synthetic PS1 mutant sequences were released from host plasmid backbones by EcoRI and BstZ17I digest, producing DNA fragments of wt length with complimentary ends for ligation into the digested pANV backbone. All digestion products were subjected to gel purification and size selection to ensure the use of correctly digested fragments for ligation, to incorporate mutant PS1 sequence into pANV (Figure 6.14 B). Using this approach, three ANV PS1 variant plasmid constructs (pANV-M1, pANV-M2 and pANV-M3) were constructed.

To verify the replacement of wt PS1 sequence in these constructs, the resulting plasmids were subjected to PCR using primers that targeted sequence specifically to each mutant variant (Table 2.4). These diagnostic PCRs produced products of the expected sizes from pANV-M1, -M2 and -M3, indicating the replacement of the wt PS1 sequence with mutant PS1 (Figure 6.15 A). In agreement with this, no PCR products were produced in reactions containing primers specific to the wt PS1 sequence (Figure 6.15 B), confirming the absence of both the wt PS1 sequence and of parental pANV. Correct assembly of the PS1 mutant sequences into the pANV backbone was further validated by restriction digest using EcoRI, BstZ17I and NheI, which produced the expected pattern of DNA fragments (707, 4144 and 5039 nt), indicating the correct assembly of each PS1 mutant pANV construct (Figure 6.15 C). To verify the presence of the correct mutations in each PS1 mutant pANV (pANV-M1, -M2 and -M3) and the absence of other mutations, the entire genome of ANV within these constructs was sequenced. Each mutant PS1 region contained the expected sequence, with the in-frame synonymous mutations as designed. No other mutations were identified in the remainder of the genome for each construct (Appendix 1).

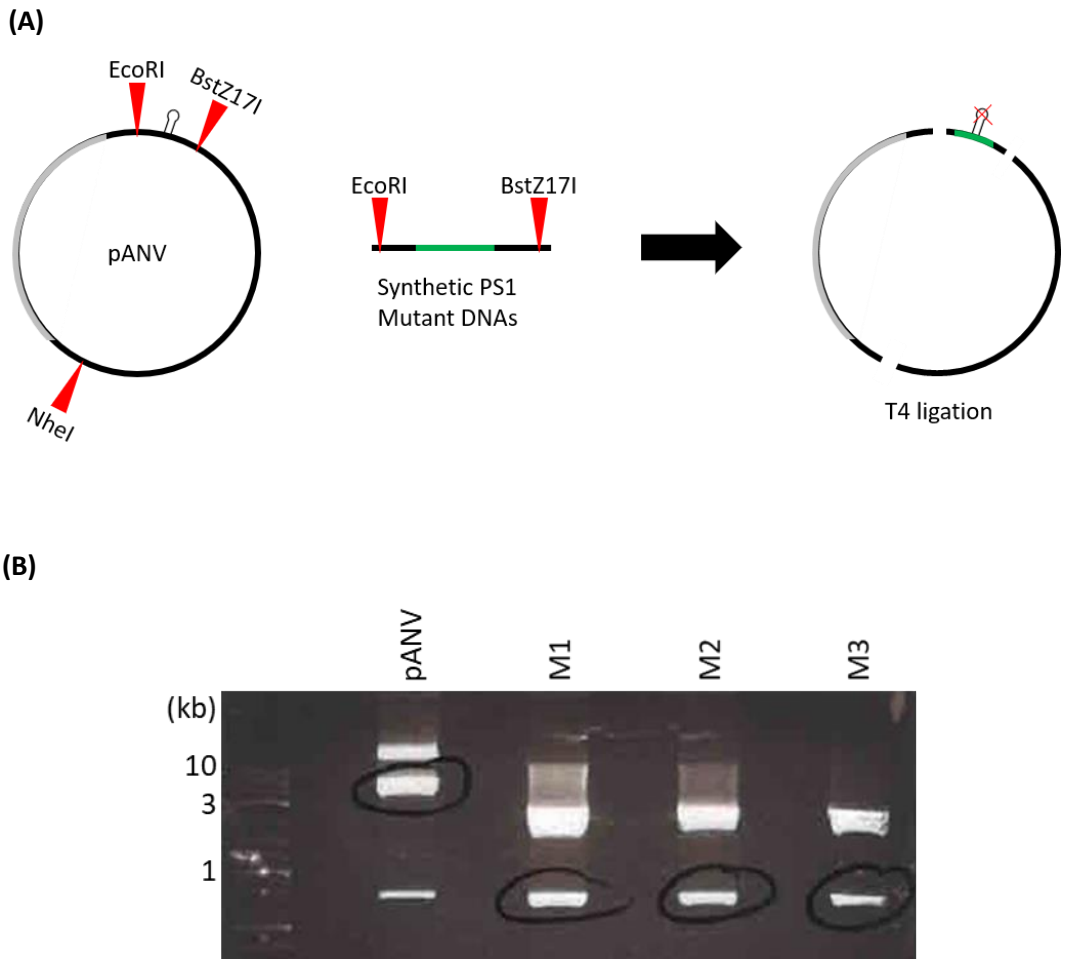


Figure 6.14: Restriction digestion of pANV and synthetic PS1 mutant sequences to isolate DNA fragments for cloning. (**A**) pANV and synthetic PS1 mutant DNAs were subjected to restriction digestion with EcoRI and BstZ17I in order to remove wt PS1 from the plasmid construct (represented by a SL) and to create complimentary overhangs for T4 ligation reactions, respectively. pANV was additionally digested with NheI to allow size selection of desired DNA fragments via electrophoresis, thus separating any partially digested or intact pANV DNA. Assembly of the DNAs containing mutant PS1 sequence (green) into the fragmented pANV backbone was completed using T4 ligase (Section 2.2.1.10), to produce pANV-M1, pANV-M2 and pANV-M3. (**B**) pANV and plasmids containing the synthetic mutant PS1 sequences were digested as outlined in (A), to remove the wt PS1 sequence and to isolate mutant DNA for cloning into the pANV construct (707 nt). This produced pANV DNA fragments distinguishable from the ~9.9 kb parental plasmid (4144 and 5039 nt) so that extracted DNAs for cloning would not contain wt PS1 sequence. DNAs for cloning of pANV-M1, pANV-M2 and pANV-M3 were excised from the gel (circled).

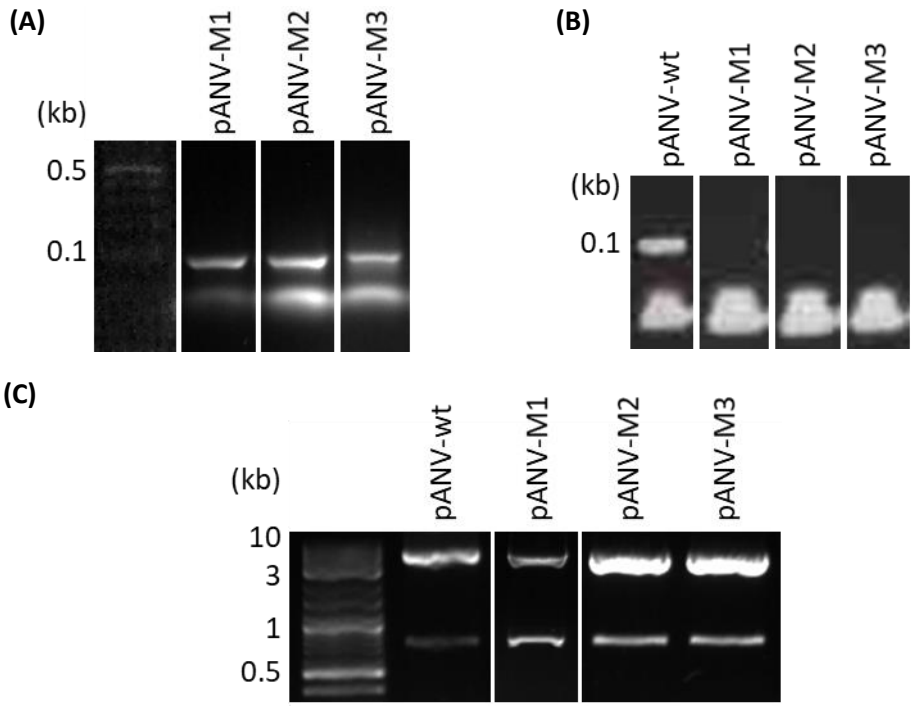


Figure 6.15: Validation of pANV PS1 mutant assemblies. Mutant PS1 DNAs were assembled into pANV to replace the wt PS1 sequence. Recovered plasmids were screened for both (**A**) the mutant insert and (**B**) wt PS1 DNA, using PCR primers specifically targeting each sequence. DNA corresponding to the PS1 mutant sequences were amplified for each mutant construct (93 nt), confirming presence of the correct insert. In contrast, primers recognising the wt PS1 sequence only amplified DNA from the parental pANV and not from the PS1 mutant constructs, confirming the region encompassing wt PS1 had been replaced in these plasmids. (**C**) Restriction digestion analysis of the pANV PS1 mutants with PCR-confirmed inserts. Digestion with EcoRI, BstZ17I and NheI resulted in the expected pattern of DNA fragments (707, 4144 and 5039 nt), consistent with correctly assembled constructs for each pANV PS1 variant.

6.4.3 Recovery of PS1 mutant ANV in BSRT7 cells

To recover PS1 mutant ANV, BSRT7 cells were transfected with one of the three PS1 variant pANV clones or with wt pANV as a positive control. Cells and media were harvested at 72 hours post-transfection for RNA extraction. These purified RNAs were subject to RT-PCR as described previously (Section 5.3.1), to confirm the transcription of PS1 mutant ANV RNA. RT-PCR demonstrated the production of RNA from all three PS1 mutant ANV constructs and the wt parental pANV. Products were specifically generated from RNA as no DNA was detected in the PCR-only controls for any transfection (Figure 6.16).

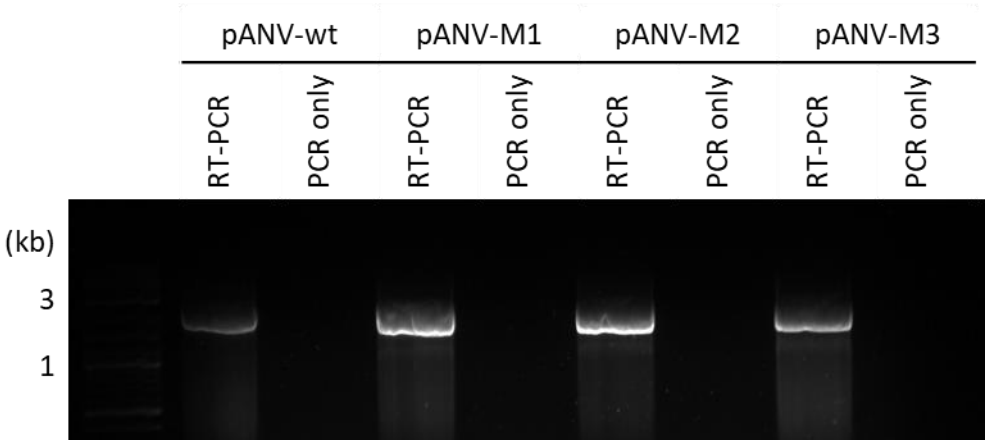


Figure 6.16: ANV genomic RNA is transcribed from PS1 mutant pANV constructs in BSRT7 cells. BSRT7 cells were transfected with wt parental or PS1 mutant pANV constructs. Cell pellets were harvested 72 hours post-transfection, from which total cellular RNA was extracted. Purified RNA was subjected to RT-PCR or PCR (without RT) to specifically confirm the presence of ANV RNA. ANV ORF1b RNA (2.2 kb) was detected in pANV, pANV-M1, pANV-M2 and pANV-M3 transfected BSRT7 cells by RT-PCR, whilst no DNA was detected by PCR alone, confirming transcription of RNA from all of the constructs.

6.4.4 Quantification of wt and PS1 mutant ANV titre

Following successful transcription of ANV RNA from each mutant PS1 construct, the production of infectious ANV was assessed. As described for the recovery of wt ANV (Section 5.3.2), naïve CK cells were inoculated with the intracellular lysate of BSRT7 cells transfected with either the wt or PS1 mutant pANV variants. At 72 hpi, comparable levels of CPE were present in all CK cells inoculated with either pANV or the mutant PS1 variants, indicating that ANV had been recovered, and that the synonymous mutations introduced to disrupt the structure of PS1 had not abolished viral replication. Extracellular media from these infected CK cells was harvested and stored to create stock inoculums for each of the ANV PS1 variants for subsequent study.

To begin quantifying any effects of PS1 mutation on ANV, assays were performed to determine the $TCID_{50}$ of each PS1 variant to quantify the titre of the recovered viruses (Section 2.2.7.2). Extracellular virus produced by infected CK cells reached titres of 4.2-7.5 x 10^5 $TCID_{50}$/ml by 96 hpi (Figure 6.17 A), indicating that production of infectious virus was not abolished by the mutations introduced to PS1 in pANV-M1, -M2 and -M3. In addition, copy numbers of RNase-resistant ANV genomes in the extracellular media stock inoculums

were determined using qPCR, following the previously described method with the addition of 0.5% IGEPAL CA-630 to RNase digestions to expose any unencapsidated ANV RNA in membrane vesicles that may otherwise be shielded from RNase digestion (Section 2.2.1.1). Using this approach, only encapsidated genomes released from infected cells as virus particles will be measured. Quantification of ANV genome using this method showed that comparable levels of RNase-resistant ANV RNA are present for the wt and each PS1 mutant, ranging from $9.16 \times 10^6 – 1.16 \times 10^7$ genome copies/ml (Figure 6.17 B). This data is consistent with results of the $TCID_{50}$ assay, and taken together, suggest that mutation of the putative PS1 structure in ANV has not had a significantly negative effect on replication of the virus.

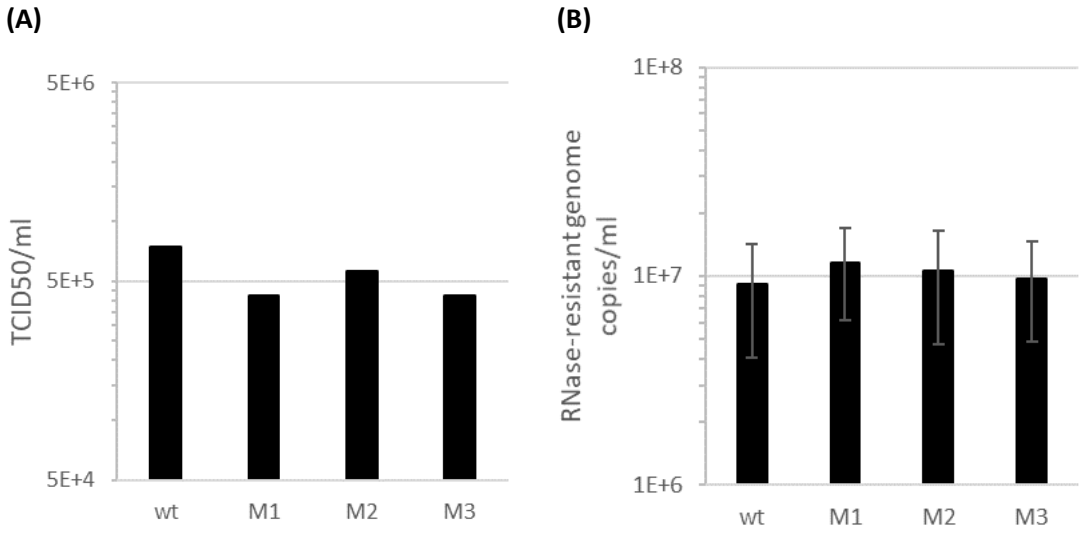


Figure 6.17: Titres of recovered wt and PS1 mutant ANV. (**A**) CK cells were inoculated with 10-fold serial dilutions of wt or mutant PS1 ANV in replicate, and occurrence of CPE was reported per well for each dilution after 96 hpi. Infectious titre was then calculated for each variant using the Spearman-Kärber method. (**B**) Stock inoculums of wt ANV and the PS1 variants, derived from extracellular media of infected CK cells, were RNase- and IGEPAL CA-630-treated in order to remove any membrane-bound and unpackaged RNA. RNA corresponding to the remaining putatively packaged genomes of ANV were purified and the absolute genome copies/ml quantified by RT-qPCR. The average titre from four biological replicates is shown with the standard deviation of technical replicates.

6.4.5 Effect of PS1 mutation on viral growth

As the titre of the recovered ANV PS1 mutants showed no significant difference to wt ANV, the growth kinetics of each ANV PS1 variant were examined to assess if these mutations had a more subtle effect on viral replication. CK cells were infected in duplicate at two

different effective MOIs of ~0.2 and ~0.02 genome copies/cell (1:5 and 1:50 cells infected, respectively), calculated from the titres of RNase-resistant ANV RNA in each stock inoculum (Figure 6.17 B). Different MOIs were used so that the number of viral replication cycles would vary between the growth assays, this could help magnify any subtle differences in growth between the PS1 mutant and wt viruses. Aliquots of extracellular media from infected CK cells were harvested at 24-hour intervals and IGEPAL CA-630- and RNase-treated before extraction of RNA, then genome copies/ml were determined for each time point by RT-qPCR. Data for ANV recovered from the pANV-M2 construct was excluded as bacterial contamination of CK cells was seen after infection with ANV-M2 stock inoculums. Growth curves were produced from the genome copy number data for each PS1 variant ANV at different time points, with different input MOIs, to assess how production of ANV virions were affected by the mutations.

At both MOIs tested, the PS1 mutant ANV-M1 grew at a rate and titre nearly identical to the wt ANV in CK cells (Figure 6.18 A and B). In comparison, the ANV-M3 PS1 mutant showed consistently lower virion production at 24 and 48 hpi, with up to a 2.8-fold reduction in released genome copies/ml compared to wt ANV infection, suggesting that the PS1 mutation in ANV-M3 may impair efficient genome packaging. Interestingly, by 72 hpi ANV-M3 achieves similar levels of extracellular virus to that of wt ANV. (Figure 6.18 A and B). The deficiency in ANV-M3 virion production was more pronounced at earlier stages of infection at the lower MOI (0.02), and overall genome copy numbers produced at 48 and 72 hpi by all viruses were up to 10-fold greater when assayed at the lower MOI (0.02) compared to the higher MOI (0.2) (Figure 6.18 A and B). Collectively, these observations suggest that infection with lower viral loads are more effective in measuring the subtle differences produced by mutation of a single PS, and that the ANV mutants assayed this way indicate that the PS1 region encompasses a sequence/structural motif involved in genome encapsidation.

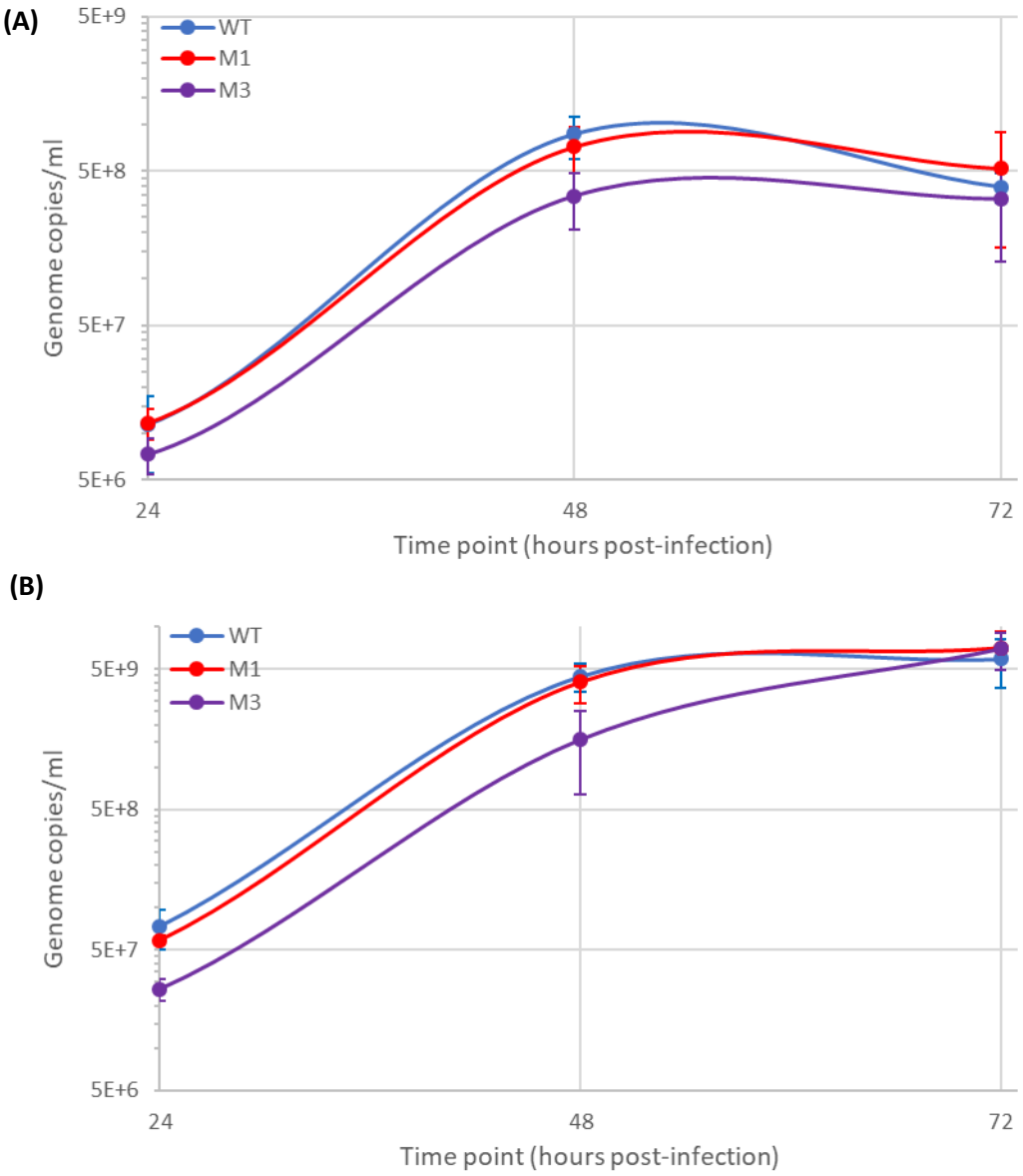


Figure 6.18: Delayed production of extracellular virions by PS1 mutant ANV infected CK cells. CK cells were infected in duplicate with either wt ANV, or PS1 mutants ANV-M1 or ANV-M3. Inoculum containing roughly equal levels of RNase-resistant ANV genome were used for each PS1 variant in two separate dilutions, to give effective MOIs of 0.2 and 0.02 genomes/cell. Aliquots of extracellular media were taken at 24-hour intervals and treated with IGEPAL CA-630 and RNase A before purification of remaining protected RNA, representing extracellular ANV virions. Genome copies/ml were determined by RT-qPCR targeting sequence within the ANV ORF1a to estimate the production of extracellular virions by each PS1 variant at each time point. Data from duplicate experiments was combined for cells inoculated at MOI 0.2 (**A**) and 0.02 (**B**), showing the average genome copies/ml and standard deviation between the six technical replicates.

As the growth kinetics of ANV-M1 were indistinguishable from wt, and those of ANV-M3 were delayed with respect to wt, the PS1 region of each virus was sequenced to check for reversion to wt sequence. Sequencing of RT-PCR products from viruses harvested at 48 hpi revealed that the PS1 sequences were identical to the expected sequences for wt, ANV-M1 and -M3, respectively, with no mutations identified in this region (Appendix 2).

To explore alternate explanations for the observed growth differences, the presence of any neighbouring putative PSs that may have been disrupted by the mutant sequences (Figure 6.12) were investigated. The sequence encompassing the mutated regions (Figure 6.12) (+/- 30 nt) was folded using Mfold, with the maximum distance between base pairs constrained to 35 so as to investigate stable local secondary structures. The minimum free energy structures revealed no additional 3' GCAA SL motifs, or variants thereof, in neither the mutant nor wt sequences adjacent to the mutated region. However, folding of a SL encompassing a 3' CAG motif was predicted for the ANV-M1 mutant (Figure 6.19). The SL was positioned at the location of the disrupted PS1 and exhibited motif sequence similarity to the identified ANV PSs. Therefore, presentation of this SL could have been adequate to replace any loss in packaging function by PS1 ablation, potentially explaining why the growth kinetics of this mutant were indistinguishable from wt. ANV-M3 does not form any subsidiary SLs with putative PS features, further supporting this hypothesis and providing an explanation for the difference in growth to both wt ANV and ANV-M1.

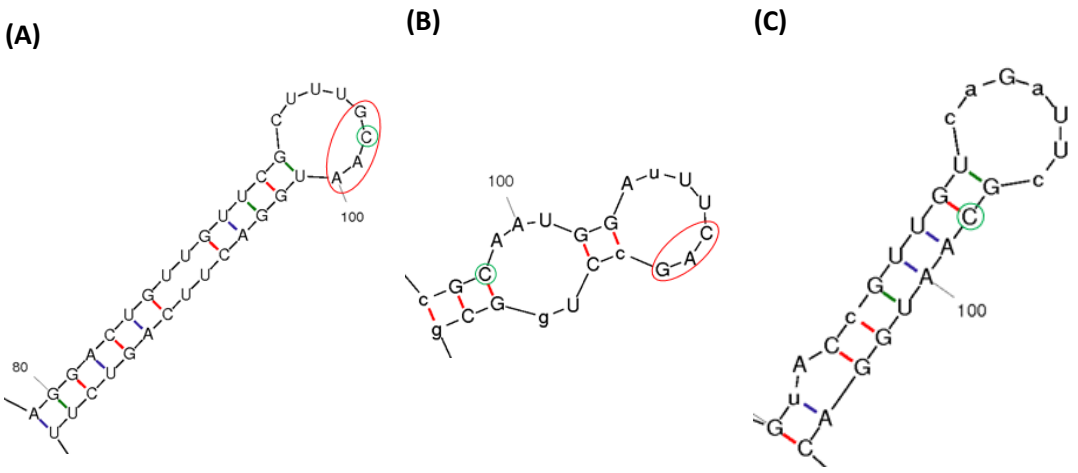


Figure 6.19: Analysis of local RNA structures surrounding mutated PS1. The sequence encompassing the mutant PS1 sequences (Figure 6.12) as well as the surrounding +/- 30 nts, were folded using Mfold. The maximum distance between base pairs was constrained to 35 in order to examine stable local secondary structures. The minimum free energy structures for (**A**) wt, (**B**) M1 and (**C**) M3 sequences reveal that a subsidiary SL containing a 3' CAG motif is present in the M1 sequence, located in the same position as the disrupted PS1 SL. The location and sequence similarity to the putative GCAA motif could indicate that this SL may act as a subsidiary PS, consistent with the growth kinetics results (Figure 6.18). Mutations introduced into the PS1 sequence are represented in lower case, and the peak nt identified by NGS is circled in green. Putative 3' PS motifs are indicated in red.

6.4.6 Genome-wide search for putative PSs with GCAA loop motif sequence

Mutagenesis experiments suggest that variations of the consensus GCAA motif could be functional. Therefore, the genome of ANV was searched for alternate SLs with the putative GCA(A)/CA(A) PS motif. A sliding-widow approach was used to fold 40 nt long frames of sequence across the genome of ANV from 5' to 3', sliding by 1 nt each frame and sampling 1000 folds at each position. The ensemble of SLs with a 3' CA(A) sequence motif, and the frequency in which they occurred was given (indicating relative stability in solution) (Appendix 3). The SLs were ranked by frequency; those in the upper 25% quartile as strong, those in the 50-75% quartile as medium, 25-50% as weak, and 0-25% as very weak (Figure 6.20). Results show the presence of 23 strongly folding SLs that exhibit a 3' CAA motif, 7 of these contained the GCAA sequence, and one of which corresponds to PS1 identified by NGS. This indicates that multiple other putative PSs may be present across the genome in ANV.

The genomic sequences encoding CP residues involved in PS binding for both STNV and HPeV1 are also found to contain PSs themselves[78, 340]. This suggests that this tight genetic embedding of PS assembly instructions and genetic code may be a characteristic of PS-mediated assembly. Therefore, as previous work identified the residues of the ANV CP predicted to interact with an RNA SL ligand (Figure 3.6), the genomic sequence at these regions was examined to establish any overlaps with the CAA loops identified here (Figure 6.20). Residues N65, T67 and K69 were found to be encoded by sequence that can fold into a SL with a 3' CAA motif sequence (genome position 4764). This demonstrates that ANV, similarly to STNV and HPeV1, may maintain a dual function within the peptide-encoding sequence involved in CP-PS binding.

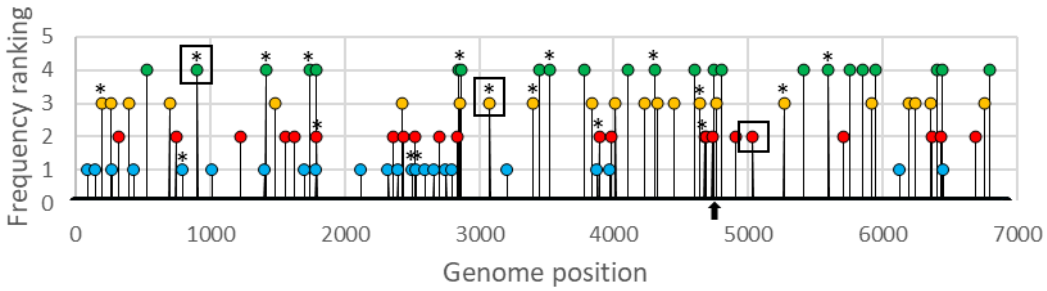


Figure 6.20: Distribution of SLs with a loop sequence containing the 3' GCAA/CAA motif. The genome of ANV was searched for SLs exhibiting a 3' GCAA/CAA loop motif sequence using a sliding-widow approach. 40 nt regions of sequence were folded across the genome, sliding by 1 nt each frame from 5' to 3' and sampling 1000 folds at each position. The frequency of occurrence for each SL with a 3' CAA sequence was given (indicating relative stability in solution). The SLs were then ranked by frequency; those in the upper 25% quartile as strong (green), those in the 50-75% quartile as medium (orange), 25-50% as weak (red), and 0-25% as very weak (blue). SLs with a GCAA loop motif sequence are marked with an asterisk, and putative PSs identified by NGS (Figure 6.5) are indicated with boxes. Additionally, the 3' CAA SL located at sequence encoding CP residues predicted to bind RNA (Figure 3.6) is indicated by an arrow.

6.5 Summary

6.5.1 Using a NGS approach to Identify putative PSs in ANV

The work described here reports the isolation of packaged and total genome populations of ANV, and the successful development of an RT-PCR protocol to synthesise and amplify genomic-length ANV transcripts (Figure 6.3 C). NGS of the two ANV populations resulted in a high level of read depth, which allowed for differences in entropy and mutation rate to be calculated at each genome position and compared between populations. However, the low frequency of variation at each genome position in the assembled contigs prevented reliable analysis using an entropy-based approach. Sites of nt variation could not be distinguished between viral mutations or artefacts that may have been introduced during RT, PCR or NGS steps, resulting in 'peaks' that may not all be meaningful (Figure 6.5 A), in contrast to the previous analysis of NGS results for FMDV quasispecies[80]. Therefore, positions showing genetic drift, and thus significant variation, were instead analysed, resulting in identification of 3 nts that were selected for > 1% more frequently in the genomes of ANV virions compared to intracellular ANV. These indicate the locations of putative PSs (Figure 6.5 B). Investigation of RNA sequence and structure at these sites revealed a common 3' GCAA motif (Figure 6.6). Further analysis of the sequences surrounding the putative PSs revealed several additional SLs with a similar motif (Figure 6.8, Figure 6.9, and Figure 6.11), supporting that this may be a putative PS motif of ANV.

6.5.2 *In vitro* validation of putative PS1 in ANV

To validate the function of the identified PSs, mutations were introduced into the pANV reverse genetics construct (Figure 6.12 and Figure 6.14 A), and recovery of PS1 mutant ANV was attempted (Figure 6.16). These mutations targeted the SL containing the most conserved nt identified in the packaged genomes when compared to the total population (selected for over 5% more frequently), corresponding to PS1 at genome position 904-911. Synonymous mutations were introduced in order to disrupt the predicted SL structure, whilst maintaining the encoded wt peptides. As it was not possible to completely abolish the wt GCAA loop motif synonymously, mutations were designed to disrupt the SL and create a more energetically stable RNA structure that would outcompete the wt SL (Figure 6.13), similarly to approaches used for HpeV1, HCV and FMDV previously[58, 78, 80]. Recovery of mutant PS1 viruses indicated that this wt SL is not crucial for viral replication (Figure 6.17). However, upon closer inspection of viral growth, it appears that disruption of PS1 in ANV appears to reduce the rate at which encapsidated viral genomes are released from

infected CK cells (Figure 6.18). This reduction in growth is only observed for one of the two mutant PS1 ANV variants (ANV-M3) during the earlier stages of infection (24 and 48 hpi), as titres matching that of the wt are observed after 72 hpi. Knock-out, or significant reduction of genome encapsidation in PS1 mutant ANV was not expected here, as PSs are thought to act cooperatively, with multiple SLs (up to 60) across the genome acting together to direct encapsidation of viral RNA[53, 62, 64]. Therefore, mutation of one, or very few PSs in the genome may have little to no observable effect on viral titre, as has been observed for other viruses previously[58, 78].

If sequence or structural changes introduced into the genome are detrimental to the production of ANV virions, then mutations to recover or replace the necessary functional elements would be selected over time. Although no reversions were identified upon sequencing the mutated regions within mutant PS1 viruses (Appendix 2), the presence of compensatory mutations at alternate sites in the genome to recover a wt-like function cannot be ruled out. However, the number of replication cycles that the viruses underwent following recovery of clonal genomic material from the cDNA constructs were limited (3 serial passages), and thus may not have given enough opportunity for substantial genetic variation to occur (that non-NGS techniques could identify).

Alternatively, there is a possibility that the putative PS1 identified here (Figure 6.6) might not be a PS of ANV. The cytosine nt in PS1 found to be > 5% more conserved in ANV virions than total cellular viral RNA may in fact be involved in an alternate SL structure, such as the stem of a neighbouring PS. This hypothesis would be consistent with the difference in viral growth kinetics between the ANV-M1 and -M3 PS1 variants, as the sequences designed to synonymously mutate the identified SL were different between them, thus these regions will form different ensembles of RNA structures. As a result, one potential fold may include that of an 'actual' wt PS in the case of ANV-M1, but not for ANV-M3, potentially explaining the difference in virion production between them. However, a more likely explanation is that the wt PS was not completely ablated in the ANV-M1 mutant.

Moreover, prediction of local secondary structures in the PS1 mutant sequences revealed folding of an alternate SL motif in the ANV-M1 sequence that closely resembles that of the putative wt PS, which was also located at the same position of the genome (Figure 6.19). This structure may act as a substitute for the disrupted PS1 SL and thus regain packaging function, which would explain why growth of ANV-M1 was indistinguishable from that of wt ANV, and why ANV-M3 was reduced in comparison. Even if the folding of this SL would

be energetically unfavourable in the context of the entire mutant sequence (Figure 6.13), the difference may be offset by the PS-CP binding free energy, as seen for STNV and MS2[39, 53], therefore allowing PS-CP binding to occur and virus assembly to continue as normal.

As multiple PSs act cooperatively to facilitate genome encapsidation, the genome of ANV was folded to identify additional SLs with motif sequence similarity to the identified GCAA PS motif sequence (Table 6.2). This revealed a number of additional sites across the genome that could fold into stable SL structures, with either a 3' CAA or GCAA sequence, and could represent additional putative PSs (Figure 6.20). Many of the stable SLs identified in ANV are seen to cluster around the start, and middle, of the capsid gene (nts 4571-6622). An apparent hallmark trait of viruses which use PS-mediated assembly is the location of PSs at the region encoding the PS-binding site of the CP[78, 340]. This appears to be an evolutionarily conserved mechanism for maintaining PS-CP binding, and thus efficient virus assembly, as mutation of the PS encoding the recognition site would result in CP unable to recognise any of the PSs throughout the genome. Therefore, the sequence and structure at this location needs to be maintained, and is likely to be conserved across strain variants[78]. This may hold true for the *Astroviridae* also, as indicated by data for ANV here (Figure 3.6 and Figure 6.20).

Together, these results indicate that a NGS and bioinformatics approach has identified putative PSs encoded within the genome of ANV, which fold into SLs presenting a motif varying around a 3' GCAA sequence.

Chapter 7

7 Discussion

7.1 Significance of astrovirus encapsidation research

The Avastroviruses are a largely unstudied viral genus, with most of what is known regarding their lifecycle derived from studies on the classical human astrovirus strains[341]. Avastroviruses infect a broad range of avian species worldwide and are found abundantly in both domesticated and wild bird populations[91, 177-186]. The emergence of *Avastrovirus* strains linked to higher clinical pathology and increased mortality highlights the need for additional research to inform prevention and control strategies[159, 166, 167]. One such area of research is the mechanism of AstV genome encapsidation and virion assembly. Packaging of viral genome into its cognate CP to form new infectious virions is a critical step in the viral lifecycle. Recent advances in this field have shown that many distinct +ssRNA viruses utilise a PS-mediated assembly mechanism, which appears to be conserved between strain variants[55, 58, 59, 62, 65, 80]. This may therefore be an attractive target for antiviral therapy, particularly in viruses with vast antigenic diversity between strains, like astroviruses, that limits the efficacy of vaccination. In this work, the genome encapsidation strategy of ANV was explored. As recombinant ANV CP could not be manipulated to form VLPs for SELEX analysis (Section 4.4), a NGS approach using reverse genetics recovered ANV was used to detect conservation within encapsidated RNAs relative to the total viral RNA, identifying 3 genome positions with preference for specific nts in the encapsidated RNA population. A consensus 3' GCAA loop motif was identified at these positions, characteristic of putative PSs (Section 6.3). A further 18 SLs with the 3' GCAA motif were then identified in the ANV genome by bioinformatics, with 7 of them forming stable structures, that could act as PSs.

7.2 PS-mediated encapsidation, risk vs reward

Use of a PS-mediated assembly mechanism places significant constraint on the viral genome. Maintaining multiple[62] SLs with a particular structure and motif throughout the RNA for packaging constrains the coding potential of the viral genome at specific sites, which in turn may limit the ability of the virus to evolve in response to selective pressures. Additionally, if PSs were mutated, compensatory mutations would be required to repair or replace the disrupted PSs[342]. These mutations would also have to be synonymous, depending on location, to maintain correct protein functionality. Therefore, the benefits of PS-mediated genome encapsidation to a virus must outweigh the costs in order for this

mechanism to be preserved over time. For example, the cooperative nature of multiple PSs throughout the genome acting together to facilitate virion formation provides a means for CP to select for the viral genome during virion assembly and prevent encapsidation of cellular RNA, the viral antigenome, or sub-genomic or degraded RNAs. Exclusion of cellular RNAs and incorrect viral RNAs may occur due to the lack of sufficient correctly positioned PSs in these RNAs, which initiate assembly nucleation and recruit all of the CP required to successfully complete the assembly pathway. Specific PS-CP interactions also help virion formation, avoiding a Levinthal's paradox effect (as seen in polypeptide-folding) (Section 1.2.3) by providing an ordered pathway for CP assembly around the genome, with the viral RNA acting as the template[40]. This programming of assembly helps prevent formation of kinetically trapped intermediates that cannot complete assembly, which might be formed by non-specific CP recruitment[44]. PS-mediated assembly also provides a solution to low CP and RNA concentrations within cells, allowing efficient assembly of CP onto the genome in a rapid and controlled manner[44], rather than leaving interactions to chance and risking detection of exposed viral RNA and improperly formed capsids by host defence mechanisms[343].

7.3 Conservation of PS-CP interactions in the *Astroviridae*

As reported previously, and observed during *in silico* modelling of AstV CP here (Section 3.2.2.3), the capsid core domains of the HAstVs share a high structural homology to one another, and to HEV[194]. The predicted structure of the ANV core region appears to be no exception, showing very high levels of similarity to the known HAstV crystal structures, which could be expected as the HAstV CP was selected as a structural template for modelling based on sequence homology to ANV. Interestingly, when these structures were searched for putative ligand binding sites via homology-driven substructure and sequence profile comparisons, a predicted CP-RNA interaction was identified (Section 3.2.4). The basis for this interaction was derived from the crystal structure of PMV, a plant virus, where ~17 nt RNA SLs from the viral genome were detected in close proximity to the 5-fold vertices of the capsid within virions[34], consistent with these SLs being PSs[36, 68]. The RNA-interacting interface of the PMV CP shares remarkable structural homology to the regions of astrovirus and HEV capsid core domains present at the interior of their capsids (Figure 3.7 and Figure 3.3). Taken together, the structural homology, alongside the identification of putative PSs within ANV (Chapter 6), suggest that both astroviruses and HEV may utilise a PS-mediated mechanism for genome encapsidation, consistent with data showing that specific RNA structures within the HEV genome have affinity for the viral CP[344].

The identification of structural homology between the CPs of AstVs, HEV and PMV may suggest that their CPs have diverged from a common ancestor[223], or that they may have all converged on a structure to achieve a common function[345], i.e. PS-mediated assembly, for which only a limited number of CP structures may be suitable. Both of these evolutionary routes present interesting hypotheses. If these structures have diverged then there is a possibility that the PSs between these distinct viruses may share some degree of homology, such as the SL structure, motif sequence, or spacing throughout the genome. However, if convergent evolution has occurred then it would indicate that there are only a limited number of CP structures capable of effectively facilitating PS-mediated assembly to form icosahedrally symmetrical capsid shells.

Studies on PS-mediated encapsidation have repeatedly shown that residues involved in the PS-CP interface, and the PS recognition motifs themselves, are broadly conserved across strain variants within a viral family[55, 58, 62, 78]. Therefore, it is also likely that PS-CP interactions across the *Astroviridae* could share similarity at a molecular level. Recent work has indirectly demonstrated that PSs may be conserved between HAstVs, as a chimeric HAstV-1 ORF1a/b and HAstV-8 ORF2 virus could be packaged by the HAstV-8 CP to produce infectious virions at comparable titres to the parental strains[259]. This data indicates that the genome of one HAstV subtype can be encapsidated by the CP of another distinct serotype, implying the mechanism and its determinants/characteristics of RNA-CP assembly are shared between strains. Additional support for a genus-wide conserved packaging mechanism comes from the identification of novel recombinant HAstV isolates derived from clinical samples, where there are numerous examples of natural ORF1a/b-ORF2 junction recombination events between strains[99, 100, 346, 347]. Recombination at the ORF2 junction is also prevalent in other MAstVs and in AAstVs, with additional evidence of inter-species recombination[96, 103-111], suggesting that the mechanics of genome encapsidation may be broadly conserved across the *Astroviridae*.

7.4 Co-transcriptional genome encapsidation

There is growing evidence in many viruses that packaging of viral RNA by PS-CP interactions occurs alongside genome synthesis in a co-transcriptional manner. In this scenario, local RNA secondary structures that are recognised by CP are only present in the nascent transcripts as they are synthesised by the polymerase, and are absent in full-length RNAs[67-69]. This mechanism has been demonstrated for HPeV1, where addition of full-length viral RNA to assembly-competent CP was insufficient to promote virus particle formation[78]. Data

generated here for ANV may also suggest that the *Astroviridae* encapsidate their genomes in a co-transcriptional manner, as addition of full-length genomic RNA to recombinant CP produced in *E. coli* did not promote virus particle formation (Section 4.4.4). Alternatively, addition of viral RNA to fixed concentration of ANV CP might have induced assembly. However, without a protein ramp similar to *in vivo* environments, CP may have bound to PSs in a disordered way and produced high levels of kinetically trapped assembly complexes instead of correctly formed ANV virions[44].

The ensemble of RNA structures formed by a genome could explain why only nascent genomes may be competent for packaging. For example, the structure of the viral RNA may transition between several conformational states during the lifecycle to switch between serving as a template for translation, replication, and for packaging. The mechanisms that control these switches are not yet fully understood and will help to further elucidate the molecular biology of a wide range of +ssRNA viruses. Global structures formed by non-encapsidated viral RNA could be expected to present structures very different to those enclosed within virions[67, 69], such as those for promoting replication and translation, e.g. ribosomal binding sites. This would mean that short-range stem-loops, such as PSs, may not be presented as they would be energetically unfavourable in the context of the complete genome. This is where co-transcriptional encapsidation may provide a solution to the problem, as when nascent genome is produced from the polymerase, close-range local RNA structures will be energetically favourable in the ensemble of folds formed by the short, incomplete transcripts. These PSs may then be captured and preserved by CP as it assembles onto the RNA. This would allow sequential formation of subsequent local RNA structures, potentially more PSs, along the genome as it is transcribed and packaged. This mechanism could also act as a regulatory switch between replication and virion assembly, as when CP concentrations are low or absent, nascent genomes would be continually produced and provide more template RNA for translation. Once sufficient quantities of subgenomic and genomic RNA are present and CP is being translated, CP will then bind to nascent genomes as they are synthesised and promote virus assembly, whilst simultaneously preventing production of excess free RNA.

If viruses, such as ANV, were to exploit a co-transcriptional encapsidation mechanism, then major PSs with high CP affinity might be expected towards the 5’ end of the genome, so as to initiate nucleocomplex formation before the length of the transcript increases to the point where long-range RNA structures become more energetically favourable, and

overshadow the PS structures. This mechanism would also provide a way of preventing CP nucleation at unwanted sites across the genome, from which only defective particles would result.

7.5 Cooperative PS-CP assembly

An increasing number of studies on +ssRNA viral genome encapsidation support the presence of multiple PSs of mixed affinity for the CP that act in a cooperative manner to facilitate assembly of the CP around the viral genome[53, 55, 58, 65, 78]. Therefore, if multiple PSs within the genome of ANV work cooperatively, and perhaps even synergistically, to achieve encapsidation, any individual PS will only have a small input towards the overall effect. Thus, mutation of just one such PS may be insufficient to have a significant impact on encapsidation efficiency, as seen in studies here (Section 6.4.4). Previous mutational studies are consistent with this, showing that multiple PSs must be disrupted past a threshold level, for example 8 in the case of HCV, before a change in packaging efficiency is observed[58]. However, mutation of particular PSs may have a more pronounced effect on encapsidation than others, which can be explained by the heterogenous affinities of PSs to CP based on the SL sequence motif, structure, and location in the assembly pathway[37].

Mutagenesis of identified putative PS1 of ANV showed only minimal effect on final viral titre upon disruption of its structure (Section 6.4.5). Quantification of released virus over time revealed that for PS1 mutant ANV-M3, accumulation of extracellular viral progeny was delayed at earlier stages of infection (24-48 hpi). By contrast, genome titres reached equal levels to those of wild type ANV after 72 hpi, indicating that the mutant virus was attenuated but still packaging competent. The observed difference in titres could be due to a slower spread of ANV-M3 through the cell monolayer, due to reduced assembly of infectious progeny virions. Towards the later stages of infection all of the susceptible cells are likely to contain virus. Therefore, the amount of potential 'viral factories' that can produce ANV become equivalent between cells infected by PS variants.

These results provide the first data to indicate that ANV uses a PS-mediated mechanism of genome encapsidation, using multiple SLs presenting a 3' GCAA motif, or a submotif thereof. Interestingly, analysis of the loop motifs present in the highly conserved RFS elements of the *Astroviridae* reveals similar 3' CAA motifs (Figure 7.1), albeit either in the context of a CCAA or RCAA motif (where R is either G or A). However, this motif is not present in all strains, including ANV. Therefore, the RFS may represent a SL with dual-function, akin to the TR of MS2, which acts as both a translational repressor and as a PS[22].

Taken together, with the data from homology modelling, it could suggest that similar PSs are widely used by the *Astroviridae*.

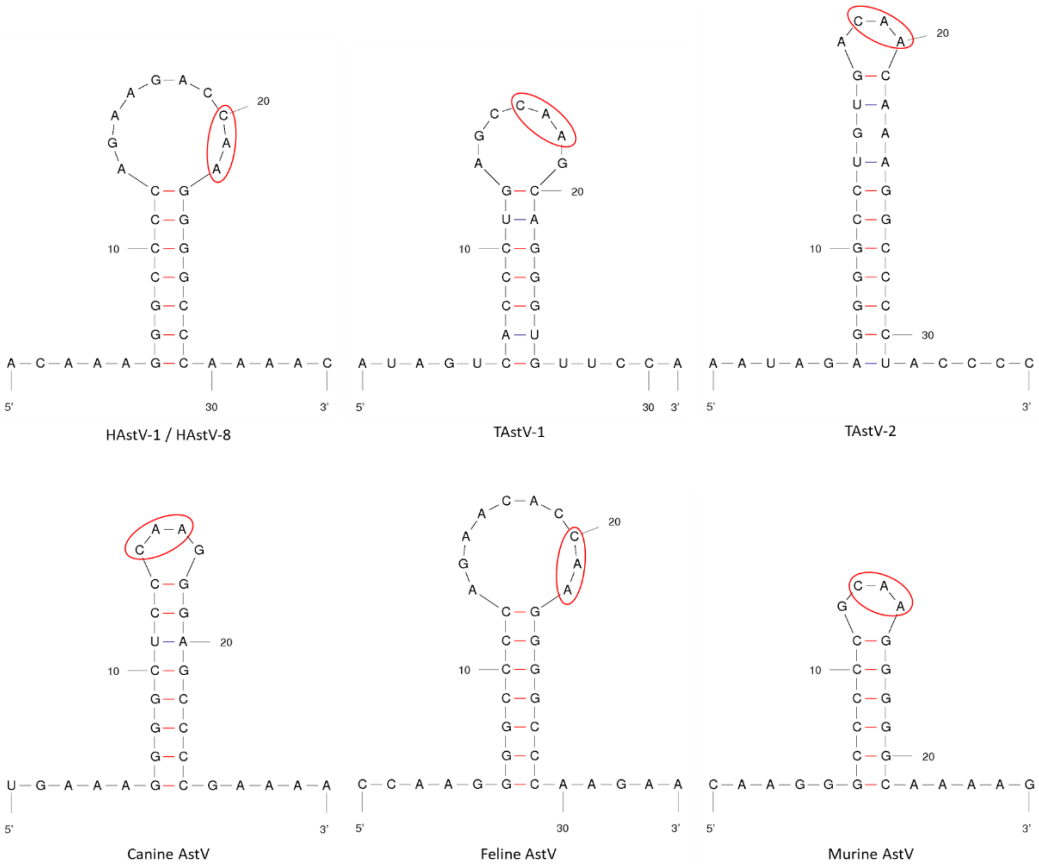


Figure 7.1: Similarities between the loop motifs of the AstV ribosomal frameshift. Sequence encoding the highly conserved ribosomal frameshift of several AstVs was folded using Mfold. The SLs show a 3' CAA loop motif, consistent with the putative PSs identified here in ANV. Structures are shown for HAstV-1 and -8 (Z25771.1 and AF260508.1, respectively), TAstV-1 (Y15936.2), TAstV-2 (AF206663.2), Canine astrovirus (KP404149.1), Feline astrovirus (KF499111.1) and Murine astrovirus (JX544743.1).

7.6 PS-CP interactions as an anti-viral therapeutic target

Computational modelling has suggested that anti-viral drugs targeting the multiple PSs present in a viral genome are less likely to generate resistance mutations compared to current strategies, and that antigenic variation across a viral genus could be potentially bypassed[348], a key limitation that conventional vaccines face[145]. Experimental data has also demonstrated that targeting capsid assembly is a viable method for inhibiting viral replication. For example, small molecular weight compounds were shown to disrupt HBV virion formation by promoting over-rapid assembly of distorted nucleocapsids, or by

kinetically trapping assembly intermediates[349, 350]. Thus, a deeper understanding of the PS-mediated assembly mechanism of ANV may yield novel strategies to control this virus, through targeting PS-CP interactions, which may also be applicable to the control of other members of *Astroviridae* with veterinary and medical importance.

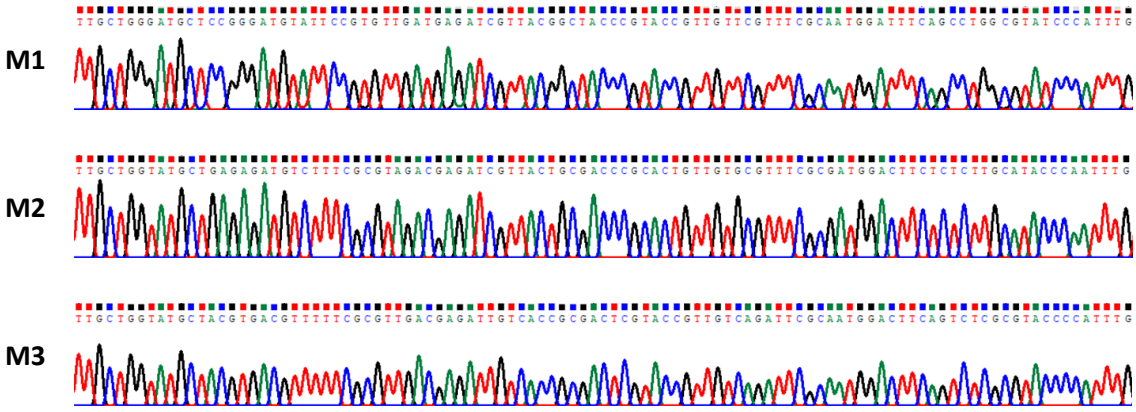
7.7 Future direction

Collectively, the work presented in this thesis indicates that ANV utilises a PS-mediated assembly mechanism for the encapsidation of its genomic RNA. However, to confirm this, further *in vitro* validation of the NGS and bioinformatics-identified putative PSs are required. With mutation of PS1 alone having a minimal effect on genome encapsidation, systematic mutation of all 3 identified putative PSs, individually, and in combination, as well as the other SLs with the identified consensus motif, could indicate whether the PSs of ANV work in a cooperative manner. Alternatively, PS-induced CP assembly assays could be developed to investigate the ability of short RNA aptamers encompassing each identified PS to promote ANV CP assembly, which would provide an affinity ranking between the PSs[51]. Elucidation of the CP residues involved in PS recognition, as well as the mechanism of PS-induced CP assembly are also important for understanding the genome encapsidation strategy used by the *Astroviridae*. Investigating this may explain why VLPs could not be formed from recombinant ANV CP in the absence of RNA. For example, as seen for STNV, there may be electrostatic repulsion between the N-terminal basic arms of CP units that prevent formation of assembly-competent higher-order structures, which only specific PS-CP interactions can overcome[51].

Finally, to validate any family- or genus-wide conservation of packaging, it would be valuable to investigate the molecular basis of genome encapsidation in other AstVs, such as the related CAstV or the more distantly related HAstV. VLP systems are already available for some HAstVs, and the factors for reversible VLP assembly established[266]. Therefore, a SELEX approach could potentially be used to identify putative PSs in these viruses. If similar PS motifs, distributions across the genome, and CP binding sites were identified between strains, then it could expose an ideal target for therapeutic intervention, that specifically targets and disrupts genome encapsidation or virion assembly. Ultimately, this work would indicate whether targeting encapsidation of the *Astroviridae* could be a viable method of control for this viral family, where the vast antigenic diversity of these viruses impedes effective control by vaccination.

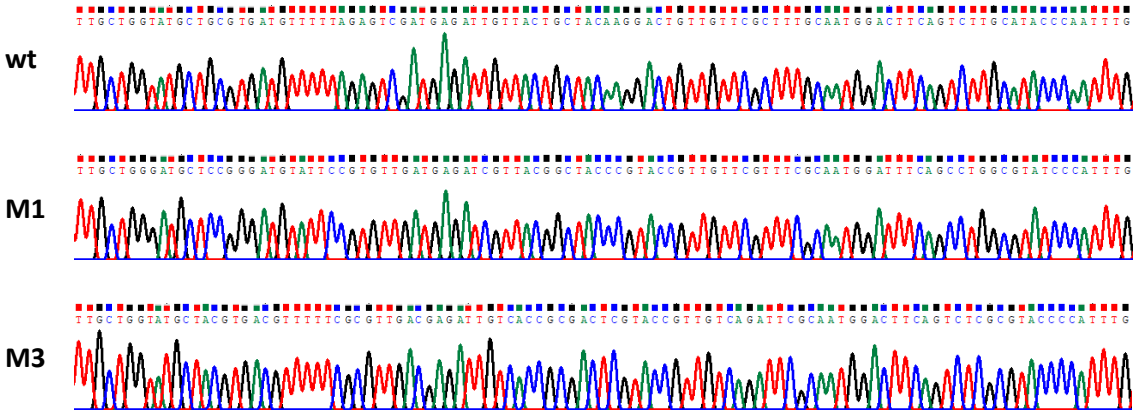
Appendices

Appendix 1



Appendix 1: Sequence analysis of PS1 mutant pANV constructs. pANV-M1, -M2 and -M3 plasmids were sequenced after correct assembly was confirmed by PCR and restriction digest analysis. Electropherogram traces indicate the expected sequences are as designed in Figure 6.12.

Appendix 2



Appendix 2: Sequence analysis of recovered PS1 mutant viruses. RNA was extracted from recovered wt and PS1 mutant viruses during the growth curve after 48 hpi (Figure 6.18). DNA fragments for sequencing were produced by RT-PCR, amplifying the mutant region and sequence flanking it. Electropherogram traces indicate that all viruses encompass the expected sequence as designed in Figure 6.12 and confirmed in Appendix 1.

Appendix 3

	Position	Loop motif	Frequency	Sum
	90	(T)TGGTCA(A)	156	156
	149	ACCCAA	116	116
*	188	TTTGCAA	286	1644
*	189	TTGCAA	129	
*	189	(T)TTGCA(A)	733	
*	190	(T)TGCA(A)	175	
*	191	TGCAA	128	
		GCAA	193	
	259	CCAA	4407	4407
	270	ACAA	343	343
	319	(T)GACTGCAGTTCA(A)	629	629
	382	(T)GAATAAAACCTGTGAACTCA(A)	983	1926
	393	(T)GTGAACTCA(A)	943	
	433	(T)GGGACA(A)	141	141
	522	AACTTAAAGACCAA	3000	6692
	531	ACCAA	3692	
	693	(T)TTAAACA(A)	373	3231
	694	(T)TAAACA(A)	2297	
	695	(T)AAACA(A)	561	
	742	(T)TGACA(A)	602	1019
	743	GACAA	224	
		(T)GACA(A)	193	
*	795	GCAA	186	186
*	904	CTTTGCAA	6267	6267
	1012	CAA	123	123
	1226	ATGTTACCAA	656	757
	1228	(T)GTTACCA(A)	101	
	1402	(T)TAATGCCA(A)	372	372
*	1411	GAGTGCAA	19166	19166
	1482	(T)TCCCA(A)	4735	4735
	1561	CAA	882	882
	1628	TTAACATCTCAA	221	802
	1629	(T)TAACATCTCA(A)	1115	
	1703	CCAA	132	132
*	1738	AGCAA	11545	11545
	1783	CAA	420	420
	1786	TCAA	7613	7613
*	1792	GCAA	964	964
	2118	AACAAATGGTCAA	113	113
	2326	CAA	155	155
	2360	AACCAA	444	444
	2394	ACGAACAA	164	164
	2426	GATCAA	2288	2288
	2435	GACCAA	636	973
		(T)GACCA(A)	337	
*	2500	(T)CAAGCA(A)	107	107

	2522	(T)TTCATTCA(A)	732	732
*	2535	GCCGGCAA	238	238
	2602	CAA	400	400
	2661	GCCTCCCCAA	197	197
	2701	(T)AAACTTCA(A)	458	458
	2756	(T)ACACA(A)	151	151
	2799	CCAAAACAACAA	210	210
	2836	(T)CATCCA(A)	752	752
	2850	AACAA	23831	23831
	2856	AACCTCAA	3934	4236
	2859	CTCAA	302	
*	2868	AGCAA	5670	10293
		AGCAACAA	4623	
	3080	(T)GGAACA(A)	231	1529
*	3082	GGAACAACTGGCAA	1108	
		AACAA	190	
	3208	CTCAA	222	222
*	3404	CCATCCAA	1164	1539
	3409	CCATCCAAGAAACCAGCAA	149	
		CAA	226	
	3447	TCCCAA	1676	6470
	3448	(T)CCCA(A)	4604	
		CCCAA	190	
*	3528	TGTGCAA	176	15767
*	3529	(T)GTGCA(A)	15591	
	3783	CCAA	12779	12779
	3839	TCCAA	2593	5190
	3840	(T)CCA(A)	2597	
	3877	(T)AGGCGTTCA(A)	103	103
*	3893	GCTATGGCAA	275	521
*	3896	(T)ATGGCA(A)	246	
	3970	(T)CATGTCA(A)	134	379
	3971	ATGTCAA	245	
	3984	CAA	1132	1132
	4016	ACAA	2389	2389
	4110	GGCCAA	6874	6874
	4230	CAA	3454	3454
*	4305	TGGGCAA	193	9887
*	4306	(T)GGGCA(A)	9694	
	4323	AGTTAATGTCAACAA	1433	5784
	4326	(T)TAATGTCA(A)	2133	
	4327	(T)AATGTCAACA(A)	188	
	4331	TCAA	2030	
	4451	(T)CCTACCTTTCCA(A)	1271	1271
	4603	TAAGCCCAA	134	6369
	4604	(T)AAGCCCA(A)	1565	
	4608	CCAA	4670	
*	4645	ACAGGCAA	1712	1712
*	4675	GAAGGCAA	743	743
	4693	CAA	995	995

	4733	CTCAA	555	958
	4734	TCAA	403	
	4745	ACAA	18034	18034
X	4764	ATGACACAA	739	1447
X	4766	(T)GACACA(A)	562	
X	4770	CAA	146	
	4804	TGGACAA	163	6766
	4805	(T)GGACA(A)	6603	
	4907	(T)GCCTCA(A)	842	842
	5037	ATTCAA	446	446
*	5267	ACACTCTGGCAA	1657	3317
*	5271	TCTGGCAA	243	
*	5272	(T)CTGGCA(A)	960	
*	5274	(T)GGCA(A)	457	
	5413	(T)GACTACCA(A)	247	11166
	5416	TACCAATAGCCAACAA	1292	
	5417	(T)ACCAATAGCCAACA(A)	6740	
	5418	CCAA	2422	
	5426	CAACAA	465	
*	5600	TCTGCAA	2211	13429
*	5601	(T)CTGCA(A)	9345	
*	5603	GCAA	1873	
	5710	(T)GCCA(A)	711	711
	5756	(T)AACA(A)	12345	12345
	5851	ACAA	22677	22677
	5918	GGTTCAA	128	1258
	5920	TTCAACAA	323	
	5921	(T)TCAACA(A)	136	
		(T)TCA(A)	671	
	5950	TCACAA	2779	7584
	5951	(T)CACA(A)	4805	
	6132	CTGGCCAA	419	419
	6200	GCAGGCTTTCCACAA	731	1496
	6205	CTTTCCACAA	765	
	6248	GATGTCACAGATACAACAA	104	1769
	6252	TCACAGATACAA	314	
	6253	(T)CACAGATACA(A)	1143	
	6258	ATACAACAA	208	
	6359	(T)CCAGCACCA(A)	123	3809
	6365	CCAACATCAA	2990	
	6368	ACATCAA	176	
	6371	TCAA	520	
	6372	(T)CAATGCTTTGTCTCTACA(A)	451	839
	6388	CAA	388	
	6407	GCTCCACAA	19439	19439
	6437	TCAA	903	903
	6443	CTCCAA	9047	9684
	6445	CCAA	637	
	6454	(T)GAATCA(A)	208	208
	6695	CAA	802	802

	6762	TCAA	4252	4252
	6801	TAGGTCAA	160	17699
	6802	(T)AGGTCA(A)	2970	
		AGGTCAATCAA	601	
	6805	TCAA	300	
	6806	CAATCAA	13668	

Appendix 3: Table of SLs in the ANV genome containing a 3' CAA motif sequence. The genome of ANV was searched for SLs exhibiting a 3' CAA loop motif sequence using a sliding-widow approach. 40 nt regions of sequence were folded across the genome, sliding by 1 nt each frame from 5' to 3' and sampling 1000 folds at each position. The frequency of occurrence for each SL with a 3' CAA sequence is given (indicating relative stability in solution). The SLs are ranked by frequency; those in the top 25% quartile as strong (green), those in the 50-75% quartile as medium (orange), 25-50% as weak (red), and 0-25% as very weak (blue). SLs with a GCAA loop motif sequence are marked with an asterisk. Additionally, a 3' CAA SL located at sequence encoding CP residues predicted to bind RNA (Figure 3.6) is marked with an X.

.

Abbreviations

A	Adenine
AAstV	Avian astrovirus
ANV	Avian Nephritis Virus
AstV	Astrovirus
ATF4	Activating Transcription Factor 4
ATP	Adenosine Triphosphate
BAM	Binary Sequence Alignment Map
BES	N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic Acid
BHK	Baby Hamster Kidney
bp	Base-pair
BSA	Bovine Serum Albumin
C	Cytosine
C1q	Complement component 1q
$CaCl_2$	Calcium Chloride
CAstV	Chicken Astrovirus
cDNA	Complimentary DNA
CEF	Chick Embryo Fibroblasts
CHOP	CCAAT-Enhancer-Binding Protein Homologous Protein
CK	Chicken Kidney
CMV	Cytomegalovirus
CP	Capsid Protein
CPE	Cytopathic Effect
Cryo-EM	Cryogenic Electron Microscopy
DAPI	4’,6-diamidino-2-phenylindole
DAstV	Duck Astrovirus
DDX23	DEAD-Box Helicase 23
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid

dNTP	Deoxyribonucleotide Triphosphate
dsDNA	Double-Stranded DNA
dsRNA	Double-Stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
eIF2α	Eukaryotic Initiation Factor 2α
EMEM	Eagle's Minimum Essential Medium
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
FAM	Carboxyfluorescein
FCS	Foetal Calf Serum
FMDV	Foot-and-Mouth Disease Virus
G	Guanine
GSP	Gene-Specific Primer
HAstV	Human Astrovirus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HDR	Hepatitis Delta Ribozyme
HEV	Hepatitis E virus
His	Histidine
hnRNP1	Heterogeneous nuclear ribonucleoprotein A1
HPA	Hamiltonian Paths Analysis
HPeV1	Human Parechovirus 1
hpi	Hours Post-Infection
ICTV	International Committee on Taxonomy of Viruses
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
kbp	Kilobase pair
kDa	Kilodalton

LB	Luria Broth
LMH	Leghorn Male Hepatoma
M	Molar
m^7G	7-Methylguanosine
MAstV	Mammalian astrovirus
MBL	Mannose-binding lectin
MGB	Minor Groove Binder
$MgCl_2$	Magnesium Chloride
µg	Microgram
µl	Microlitre
µM	Micromolar
mM	Millimolar
MMA	MES (2-[N-morpholino]ethanesulfonic acid), MgCl2 and Acetosyringone
MOI	Multiplicity of Infection
mRNA	Messenger RNA
NaCl	Sodium chloride
Na_2SO_4	Sodium sulphate
NBCS	Newborn Calf Serum
NGS	Next-Generation Sequencing
NHE2 / NHE3	Sodium–hydrogen exchanger 2/3
NLS	Nuclear localisation signal
ng	Nanogram
nm	Nanometer
nsP	Non-structural proteins
nt	Nucleotide
NTP	Nucleotide Triphosphate
ORF	Open Reading Frame
PAMPS	Pathogen-Associated Molecular Patterns
PBS	Phosphate Buffered Saline
PBSa	Phosphate Buffered Saline (Mg^{2+} and Ca^{2+} free)

PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEC	Poult Enteritis Complex
PEG	Polyethylene Glycol
PES	Poult Enteritis Syndrome
pfu	Plaque Forming Units
PKR	Protein Kinase R
PMV	Panicum Mosaic Virus
Poly-A	Poly-Adenine
PTB	Polypyrimidine tract-binding protein
PVDF	Polyvinylidene Difluoride
RdRp	RNA-dependant RNA-polymerase
RFS	Ribosomal frameshift
R_h	Hydrodynamic Radius
RMSD	Root-Mean-Square Deviation
RNA	Ribonucleic Acid
RSS	Runting-Stunting Syndrome
RT	Reverse Transcription
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
qPCR	Quantitative-PCR
s2m	Stem-loop II motif
SAM	Sequence Alignment Map
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
siRNA	Small interfering Ribonucleic Acid
SL	Stem-Loop
SPF	Specific-Pathogen-Free
SRSF	Serine and Arginine Rich Splicing Factors
+ssRNA	Positive-Sense Single-Stranded RNA
STNV	Satellite Tobacco Necrosis Virus

T	Thymine
TAstV	Turkey Astrovirus
TBE	Tris/Borate/EDTA
TBS-T	Tris-Buffered Saline-Tween
$TCID_{50}$	Median Tissue Culture Infectious Dose
TEM	Transmission Electron Microscopy
TG	1-Thioglycerol
Tm	Melting Temperature
TN	Tris-NaCl
TNS	Tris-Na_2SO_4
TR	Translational Operator
USA	United States of America
UTR	Untranslated region
UV	Ultraviolet
v/v	Volume by Volume
VLP	Virus-Like Particle
VP	Viral Protein
VPg	Viral Protein Genome-linked
w/v	Weight by Volume
wt	Wild Type

Bibliography

1. Teif V.B., Bohinc K. Condensed DNA: Condensing the concepts. Progress in Biophysics and Molecular Biology. 2011;105(3):208-22.
2. Wiggins P.A., Cheveralls K.C., Martin J.S., Lintner R., Kondev J. Strong intranucleoid interactions organize the Escherichia coli chromosome into a nucleoid filament. Proceedings of the National Academy of Sciences. 2010;107(11):4991.
3. Aksyuk A.A., Rossmann M.G. Bacteriophage assembly. Viruses. 2011;3(3):172-203.
4. Smith D.E. Single-molecule studies of viral DNA packaging. Curr Opin Virol. 2011;1(2):134-41.
5. Black L.W. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism. Virology. 2015;479-480:650-6.
6. Rixon F.J., Schmid M.F. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. Curr Opin Virol. 2014;5:105-10.
7. Rao V.B., Feiss M. The bacteriophage DNA packaging motor. Annual review of genetics. 2008;42:647-81.
8. Heming J.D., Conway J.F., Homa F.L. Herpesvirus Capsid Assembly and DNA Packaging. Advances in anatomy, embryology, and cell biology. 2017;223:119-42.
9. Motwani T., Teschke C.M. Architect of Virus Assembly: the Portal Protein Nucleates Procapsid Assembly in Bacteriophage P22. J Virol. 2019;93(9).
10. Newcomb W.W., Homa F.L., Brown J.C. Involvement of the Portal at an Early Step in Herpes Simplex Virus Capsid Assembly. J Virol. 2005;79(16):10540.
11. Hu T., Zhang R., Shklovskii B.I. Electrostatic theory of viral self-assembly. Physica A: Statistical Mechanics and its Applications. 2008;387(12):3059-64.
12. Belyi V.A., Muthukumar M. Electrostatic origin of the genome packing in viruses. Proceedings of the National Academy of Sciences. 2006;103(46):17174.
13. Bancroft J.B., Hiebert E., Bracker C.E. The effects of various polyanions on shell formation of some spherical viruses. Virology. 1969;39(4):924-30.
14. Hiebert E., Bancroft J.B., Bracker C.E. The assembly in vitro of some small spherical viruses, hybrid viruses, and other nucleoproteins. Virology. 1968;34(3):492-508.
15. Routh A., Domitrovic T., Johnson J.E. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(6):1907-12.

16. Routh A., Domitrovic T., Johnson J.E. Packaging host RNAs in small RNA viruses: an inevitable consequence of an error-prone polymerase? Cell cycle (Georgetown, Tex). 2012;11(20):3713-4.

17. Borodavka A., Tuma R., Stockley P.G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(39):15769-74.

18. Beckett D., Wu H.N., Uhlenbeck O.C. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. Journal of molecular biology. 1988;204(4):939-47.

19. Stockley P.G., Rolfsson O., Thompson G.S., Basnak G., Francese S., Stonehouse N.J., et al. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. Journal of molecular biology. 2007;369(2):541-52.

20. Perkett M.R., Mirijanian D.T., Hagan M.F. The allosteric switching mechanism in bacteriophage MS2. J Chem Phys. 2016;145(3):035101-.

21. Bernardi A., Spahr P.F. Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. Proceedings of the National Academy of Sciences of the United States of America. 1972;69(10):3033-7.

22. Beckett D., Uhlenbeck O.C. Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. Journal of Molecular Biology. 1988;204(4):927-38.

23. Dykeman E.C., Stockley P.G., Twarock R. Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. Journal of molecular biology. 2010;395(5):916-23.

24. Dykeman E.C., Twarock R. All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein. Phys Rev E Stat Nonlin Soft Matter Phys. 2010;81(3 Pt 1):031908.

25. Stockley P.G., Stonehouse N.J., Murry J.B., Goodman S.T.S., Talbot S.J., Adams C.J., et al. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. Nucleic Acids Research. 1995;23(13):2512-8.

26. Talbot S.J., Goodman S., Bates S.R.E., Fishwick C.W.G., Stockley P.G. Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex. Nucleic Acids Research. 1990;18(12):3521-8.

27. Dent K.C., Thompson R., Barker A.M., Hiscox J.A., Barr J.N., Stockley P.G., et al. The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. Structure (London, England : 1993). 2013;21(7):1225-34.

28. Toropova K., Basnak G., Twarock R., Stockley P.G., Ranson N.A. The Three-dimensional Structure of Genomic RNA in Bacteriophage MS2: Implications for Assembly. Journal of Molecular Biology. 2008;375(3):824-36.

29. Valegård K., Murray J.B., Stockley P.G., Stonehouse N.J., Liljas L. Crystal structure of an RNA bacteriophage coat protein–operator complex. Nature. 1994;371(6498):623-6.

30. Valegård K., Murray J.B., Stonehouse N.J., van den Worm S., Stockley P.G., Liljas L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. Journal of Molecular Biology. 1997;270(5):724-38.

31. Fisher A.J., Johnson J.E. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. Nature. 1993;361(6408):176-9.

32. Bottcher B., Crowther R.A. Difference imaging reveals ordered regions of RNA in turnip yellow mosaic virus. Structure (London, England : 1993). 1996;4(4):387-94.

33. Larson S.B., Lucas R.W., Greenwood A., McPherson A. The RNA of turnip yellow mosaic virus exhibits icosahedral order. Virology. 2005;334(2):245-54.

34. Makino D.L., Larson S.B., McPherson A. The crystallographic structure of Panicum Mosaic Virus (PMV). J Struct Biol. 2013;181(1):37-52.

35. Lane S.W., Dennis C.A., Lane C.L., Trinh C.H., Rizkallah P.J., Stockley P.G., et al. Construction and crystal structure of recombinant STNV capsids. Journal of molecular biology. 2011;413(1):41-50.

36. Koning R., van den Worm S., Plaisier J.R., van Duin J., Pieter Abrahams J., Koerten H. Visualization by Cryo-electron Microscopy of Genomic RNA that Binds to the Protein Capsid Inside Bacteriophage MS2. Journal of Molecular Biology. 2003;332(2):415-22.

37. Kalynych S., Palkova L., Plevka P. The Structure of Human Parechovirus 1 Reveals an Association of the RNA Genome with the Capsid. J Virol. 2016;90(3):1377-86.

38. Twarock R., Leonov G., Stockley P.G. Hamiltonian path analysis of viral genomes. Nature Communications. 2018;9(1):2021.

39. Dykeman E.C., Stockley P.G., Twarock R. Packaging Signals in Two Single-Stranded RNA Viruses Imply a Conserved Assembly Mechanism and Geometry of the Packaged Genome. Journal of Molecular Biology. 2013;425(17):3235-49.

40. Geraets J.A., Dykeman E.C., Stockley P.G., Ranson N.A., Twarock R. Asymmetric Genome Organization in an RNA Virus Revealed via Graph-Theoretical Analysis of Tomographic Data. PLOS Computational Biology. 2015;11(3):e1004146.

41. Prevelige P.E. Follow the Yellow Brick Road: A Paradigm Shift in Virus Assembly. Journal of Molecular Biology. 2016;428(2, Part B):416-8.
42. Borodavka A., Tuma R., Stockley P.G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(39):15769-74.
43. Borodavka A., Tuma R., Stockley P.G. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. RNA Biol. 2013;10(4):481-9.
44. Dykeman E.C., Stockley P.G., Twarock R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(14):5361-6.
45. Caspar D.L., Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol. 1962;27:1-24.
46. Twarock R., Bingham R.J., Dykeman E.C., Stockley P.G. A modelling paradigm for RNA virus assembly. Curr Opin Virol. 2018;31:74-81.
47. Dykeman E.C., Stockley P.G., Twarock R. Building a viral capsid in the presence of genomic RNA. Phys Rev E Stat Nonlin Soft Matter Phys. 2013;87(2):022717.
48. Eigen M., Biebricher C.K., Gebinoga M., Gardiner W.C. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. Biochemistry. 1991;30(46):11005-18.
49. Dykeman E.C., Stockley P.G., Twarock R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. Proceedings of the National Academy of Sciences. 2014;111(14):5361.
50. Twarock R., Stockley P.G. RNA-Mediated Virus Assembly: Mechanisms and Consequences for Viral Evolution and Therapy. Annual Review of Biophysics. 2019;48(1):495-514.
51. Ford R.J., Barker A.M., Bakker S.E., Coutts R.H., Ranson N.A., Phillips S.E.V., et al. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. Journal of molecular biology. 2013;425(6):1050-64.
52. Patel N., Dykeman E.C., Coutts R.H., Lomonossoff G.P., Rowlands D.J., Phillips S.E., et al. Revealing the density of encoded functions in a viral RNA. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(7):2227-32.
53. Bunka D.H., Lane S.W., Lane C.L., Dykeman E.C., Ford R.J., Barker A.M., et al. Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus:

implications for the assembly of a T=1 capsid. Journal of molecular biology. 2011;413(1):51-65.

54. Patel N., Wroblewski E., Leonov G., Phillips S.E.V., Tuma R., Twarock R., et al. Rewriting nature's assembly manual for a ssRNA virus. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(46):12255-60.

55. Patel N., White S.J., Thompson R.F., Bingham R., Weiß E.U., Maskell D.P., et al. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nat Microbiol. 2017;2:17098-.

56. Selzer L., Zlotnick A. Assembly and Release of Hepatitis B Virus. Cold Spring Harb Perspect Med. 2015;5(12).

57. Birnbaum F., Nassal M. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J Virol. 1990;64(7):3319-30.

58. Stewart H., Bingham R.J., White S.J., Dykeman E.C., Zothner C., Tuplin A.K., et al. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. Scientific reports. 2016;6:22952.

59. Shakeel S., Dykeman E.C., White S.J., Ora A., Cockburn J.J.B., Butcher S.J., et al. Genomic RNA folding mediates assembly of human parechovirus. Nature Communications. 2017;8:5.

60. Shakeel S., Westerhuis B.M., Domanska A., Koning R.I., Matadeen R., Koster A.J., et al. Multiple capsid-stabilizing interactions revealed in a high-resolution structure of an emerging picornavirus causing neonatal sepsis. Nature Communications. 2016;7:11387.

61. Zhu L., Wang X., Ren J., Porta C., Wenham H., Ekstrom J.O., et al. Structure of Ljungan virus provides insight into genome packaging of this picornavirus. Nat Commun. 2015;6:8316.

62. Dykeman E.C., Stockley P.G., Twarock R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. Journal of molecular biology. 2013;425(17):3235-49.

63. Borodavka A., Tuma R., Stockley P.G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proceedings of the National Academy of Sciences. 2012;109(39):15769.

64. Stockley P.G., Twarock R., Bakker S.E., Barker A.M., Borodavka A., Dykeman E., et al. Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. Journal of biological physics. 2013;39(2):277-87.

65. Stockley P.G., White S.J., Dykeman E., Manfield I., Rolfsson O., Patel N., et al. Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid. Bacteriophage. 2016;6(1):e1157666.

66. Gott J.M., Wilhelm L.J., Uhlenbeck O.C. RNA binding properties of the coat protein from bacteriophage GA. Nucleic acids research. 1991;19(23):6499-503.

67. Larman B.C., Dethoff E.A., Weeks K.M. Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. Biochemistry. 2017;56(16):2175-83.

68. Larson S.B., McPherson A. Satellite tobacco mosaic virus RNA: structure and implications for assembly. Curr Opin Struc Biol. 2001;11(1):59-65.

69. Rolfsson Ó., Middleton S., Manfield I.W., White S.J., Fan B., Vaughan R., et al. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. Journal of molecular biology. 2016;428(2 Pt B):431-48.

70. Jiang P., Liu Y., Ma H.C., Paul A.V., Wimmer E. Picornavirus morphogenesis. Microbiol Mol Biol Rev. 2014;78(3):418-37.

71. Liu Y., Wang C., Mueller S., Paul A.V., Wimmer E., Jiang P. Direct interaction between two viral proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus morphogenesis. PLoS pathogens. 2010;6(8):e1001066.

72. Seo J.K., Kwon S.J., Rao A.L. A physical interaction between viral replicase and capsid protein is required for genome-packaging specificity in an RNA virus. J Virol. 2012;86(11):6210-21.

73. Tuerk C., Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science (New York, NY). 1990;249(4968):505-10.

74. Tuerk C., MacDougal S., Gold L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(15):6988-92.

75. Ellington A.D., Szostak J.W. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990;346(6287):818-22.

76. Bunka D.H., Stockley P.G. Aptamers come of age - at last. Nat Rev Microbiol. 2006;4(8):588-96.

77. Schneider D., Tuerk C., Gold L. Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. Journal of molecular biology. 1992;228(3):862-9.

78. Shakeel S., Dykeman E.C., White S.J., Ora A., Cockburn J.J.B., Butcher S.J., et al. Genomic RNA folding mediates assembly of human parechovirus. Nature Communications. 2017;8(1):5.

79. Altschul S.F., Erickson B.W. A nonlinear measure of subalignment similarity and its significance levels. Bull Math Biol. 1986;48(5-6):617-32.

80. Logan G., Newman J., Wright C.F., Lasecka-Dykes L., Haydon D.T., Cottam E.M., et al. Deep sequencing of foot-and-mouth disease virus reveals RNA sequences involved in genome packaging. J Virol. 2017.

81. Lauring A.S., Andino R. Quasispecies theory and the behavior of RNA viruses. PLoS pathogens. 2010;6(7):e1001005.

82. Donato C., Vijaykrishna D. The Broad Host Range and Genetic Diversity of Mammalian and Avian Astroviruses. Viruses. 2017;9(5):102.

83. Guix S., Bosch A., Pintó R.M. Astrovirus Taxonomy. Schultz-Cherry S., editor. In: Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions. New York, NY: Springer New York; 2013. p. 97-118.

84. Karlsson E.A., Small C.T., Freiden P., Feeroz M.M., Matsen F.A.t., San S., et al. Non-Human Primates Harbor Diverse Mammalian and Avian Astroviruses Including Those Associated with Human Infections. PLoS pathogens. 2015;11(11):e1005225.

85. Wang X., Wang J., Zhou C., Yang S., Shen Q., Zhang W., et al. Viral metagenomics of fecal samples from non-human primates revealed human astrovirus in a chimpanzee, China. Gut pathogens. 2016;8:53-.

86. Xin Y.Y., Li L.L., Ao Y.Y., Xie Z.P., Li J.S., Duan Z.J., et al. A novel astrovirus identified in wild rhesus monkey feces in China. Archives of virology. 2019.

87. Kapoor A., Li L., Victoria J., Oderinde B., Mason C., Pandey P., et al. Multiple novel astrovirus species in human stool. The Journal of general virology. 2009;90(Pt 12):2965-72.

88. Pankovics P., Boros A., Kiss T., Delwart E., Reuter G. Detection of a mammalian-like astrovirus in bird, European roller (Coracias garrulus). Infect Genet Evol. 2015;34:114-21.

89. Meliopoulos V.A., Kayali G., Burnham A., Oshansky C.M., Thomas P.G., Gray G.C., et al. Detection of antibodies against Turkey astrovirus in humans. PloS one. 2014;9(5):e96934.

90. Sun N., Yang Y., Wang G.S., Shao X.Q., Zhang S.Q., Wang F.X., et al. Detection and characterization of avastrovirus associated with diarrhea isolated from minks in China. Food and environmental virology. 2014;6(3):169-74.

91. Baxendale W., Mebatsion T. The isolation and characterisation of astroviruses from chickens. Avian pathology : journal of the WVPA. 2004;33(3):364-70.
92. Lukashov V.V., Goudsmit J. Evolutionary relationships among Astroviridae. J Gen Virol. 2002;83(Pt 6):1397-405.
93. Meyer C.T., Bauer I.K., Antonio M., Adeyemi M., Saha D., Oundo J.O., et al. Prevalence of classic, MLB-clade and VA-clade Astroviruses in Kenya and The Gambia. Virology journal. 2015;12:78.
94. Kofstad T., Jonassen C.M. Screening of Feral and Wood Pigeons for Viruses Harbouring a Conserved Mobile Viral Element: Characterization of Novel Astroviruses and Picornaviruses. PloS one. 2011;6(10).
95. van Hemert F.J., Lukashov V.V., Berkhout B. Different rates of (non-)synonymous mutations in astrovirus genes; correlation with gene function. Virol J. 2007;4:25.
96. Strain E., Kelley L.A., Schultz-Cherry S., Muse S.V., Koci M.D. Genomic analysis of closely related astroviruses. J Virol. 2008;82(10):5099-103.
97. Belliot G., Laveran H., Monroe S.S. Detection and genetic differentiation of human astroviruses: phylogenetic grouping varies by coding region. Archives of virology. 1997;142(7):1323-34.
98. Wolfaardt M., Kiulia N.M., Mwenda J.M., Taylor M.B. Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. Journal of clinical microbiology. 2011;49(2):728-31.
99. Walter J.E., Briggs J., Guerrero M.L., Matson D.O., Pickering L.K., Ruiz-Palacios G., et al. Molecular characterization of a novel recombinant strain of human astrovirus associated with gastroenteritis in children. Archives of virology. 2001;146(12):2357-67.
100. Medici M.C., Tummolo F., Martella V., Banyai K., Bonerba E., Chezzi C., et al. Genetic heterogeneity and recombination in type-3 human astroviruses. Infect Genet Evol. 2015;32:156-60.
101. Ha H.J., Lee S.G., Cho H.G., Jin J.Y., Lee J.W., Paik S.Y. Complete genome sequencing of a recombinant strain between human astrovirus antigen types 2 and 8 isolated from South Korea. Infect Genet Evol. 2016;39:127-31.
102. Martella V., Pinto P., Tummolo F., De Grazia S., Giammanco G.M., Medici M.C., et al. Analysis of the ORF2 of human astroviruses reveals lineage diversification, recombination and rearrangement and provides the basis for a novel sub-classification system. Archives of virology. 2014;159(12):3185-96.

103. Pantin-Jackwood M.J., Spackman E., Woolcock P.R. Phylogenetic analysis of Turkey astroviruses reveals evidence of recombination. Virus Genes. 2006;32(2):187-92.
104. De Battisti C., Salviato A., Jonassen C.M., Toffan A., Capua I., Cattoli G. Genetic characterization of astroviruses detected in guinea fowl (Numida meleagris) reveals a distinct genotype and suggests cross-species transmission between turkey and guinea fowl. Archives of virology. 2012;157(7):1329-37.
105. Liu N., Wang F., Shi J., Zheng L., Wang X., Zhang D. Molecular characterization of a duck hepatitis virus 3-like astrovirus. Veterinary microbiology. 2014;170(1-2):39-47.
106. Lan D., Ji W., Shan T., Cui L., Yang Z., Yuan C., et al. Molecular characterization of a porcine astrovirus strain in China. Archives of virology. 2011;156(10):1869-75.
107. Nagai M., Omatsu T., Aoki H., Otomaru K., Uto T., Koizumi M., et al. Full genome analysis of bovine astrovirus from fecal samples of cattle in Japan: identification of possible interspecies transmission of bovine astrovirus. Archives of virology. 2015;160(10):2491-501.
108. Chen X., Zhang B., Yue H., Wang Y., Zhou F., Zhang Q., et al. A novel astrovirus species in the gut of yaks with diarrhoea in the Qinghai-Tibetan Plateau, 2013. J Gen Virol. 2015;96(12):3672-80.
109. Ulloa J.C., Gutierrez M.F. Genomic analysis of two ORF2 segments of new porcine astrovirus isolates and their close relationship with human astroviruses. Canadian journal of microbiology. 2010;56(7):569-77.
110. Rivera R., Nollens H.H., Venn-Watson S., Gulland F.M., Wellehan J.F., Jr. Characterization of phylogenetically diverse astroviruses of marine mammals. J Gen Virol. 2010;91(Pt 1):166-73.
111. Tse H., Chan W.M., Tsoi H.W., Fan R.Y., Lau C.C., Lau S.K., et al. Rediscovery and genomic characterization of bovine astroviruses. J Gen Virol. 2011;92(Pt 8):1888-98.
112. Kurtz J.B., Lee T.W., Craig J.W., Reed S.E. Astrovirus infection in volunteers. Journal of medical virology. 1979;3(3):221-30.
113. Midthun K., Greenberg H.B., Kurtz J.B., Gary G.W., Lin F.Y., Kapikian A.Z. Characterization and seroepidemiology of a type 5 astrovirus associated with an outbreak of gastroenteritis in Marin County, California. Journal of clinical microbiology. 1993;31(4):955-62.
114. Madeley C.R., Cosgrove B.P. Viruses in Infantile Gastroenteritis. Lancet (London, England). 1975;2(7925):124-.
115. Spalding M. Diseases of Poultry, 12th Edition 2009. p251-6.

116. Yamaguchi S., Imada T., Kawamura H. Characterization of a picornavirus isolated from broiler chicks. Avian Dis. 1979;23(3):571-81.
117. Bridger J.C. Detection by electron microscopy of caliciviruses, astroviruses and rotavirus-like particles in the faeces of piglets with diarrhoea. The Veterinary record. 1980;107(23):532-3.
118. Castro T.X., Cubel Garcia R.C., Costa E.M., Leal R.M., Xavier Mda P., Leite J.P. Molecular characterisation of calicivirus and astrovirus in puppies with enteritis. The Veterinary record. 2013;172(21):557.
119. Kjeldsberg E., Hem A. Detection of astroviruses in gut contents of nude and normal mice. Archives of virology. 1985;84(1):135-40.
120. McNulty M.S., Curran W.L., McFerran J.B. Detection of astroviruses in turkey faeces by direct electron microscopy. The Veterinary record. 1980;106(26):561.
121. Hoshino Y., Zimmer J.F., Moise N.S., Scott F.W. Detection of astroviruses in feces of a cat with diarrhea. Brief report. Archives of virology. 1981;70(4):373-6.
122. Gough R.E., Collins M.S., Borland E., Keymer L.F. Astrovirus-like particles associated with hepatitis in ducklings. The Veterinary record. 1984;114(11):279.
123. Gray E.W., Angus K.W., Snodgrass D.R. Ultrastructure of the small intestine in astrovirus-infected lambs. J Gen Virol. 1980;49(1):71-82.
124. Woode G.N., Bridger J.C. Isolation of small viruses resembling astroviruses and caliciviruses from acute enteritis of calves. Journal of medical microbiology. 1978;11(4):441-52.
125. Englund L., Chriel M., Dietz H.H., Hedlund K.O. Astrovirus epidemiologically linked to pre-weaning diarrhoea in mink. Veterinary microbiology. 2002;85(1):1-11.
126. Chu D.K.W., Poon L.L.M., Guan Y., Peiris J.S.M. Novel astroviruses in insectivorous bats. J Virol. 2008;82(18):9107-14.
127. Yuan X., Meng K., Zhang Y., Yu Z., Ai W., Wang Y. Genome analysis of newly emerging goose-origin nephrotic astrovirus in China reveals it belongs to a novel genetically distinct astrovirus. Infect Genet Evol. 2018;67:1-6.
128. Qiu Y., Chen J.-M., Wang T., Hou G.-Y., Zhuang Q.-Y., Wu R., et al. Detection of viromes of RNA viruses using the next generation sequencing libraries prepared by three methods. Virus Research. 2017;237:22-6.
129. Herrmann J.E., Taylor D.N., Echeverri P., Blacklow N.R. Astroviruses as a Cause of Gastroenteritis in Children. New England Journal of Medicine. 1991;324(25):1757-60.

130. Bosch A., Guix S., Pintó R.M. Epidemiology of Human Astroviruses. Schultz-Cherry S., editor. In: Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions. New York, NY: Springer New York; 2013. p. 1-18.

131. Walter J.E., Mitchell D.K. Astrovirus infection in children. Current opinion in infectious diseases. 2003;16(3):247-53.

132. Glass R.I., Noel J., Mitchell D., Herrmann J.E., Blacklow N.R., Pickering L.K., et al. The changing epidemiology of astrovirus-associated gastroenteritis: a review. Archives of virology Supplementum. 1996;12:287-300.

133. Mead P.S., Slutsker L., Dietz V., McCaig L.F., Bresee J.S., Shapiro C., et al. Food-related illness and death in the United States. Emerging infectious diseases. 1999;5(5):607-25.

134. Quan P.-L., Wagner T.A., Briese T., Torgerson T.R., Hornig M., Tashmukhamedova A., et al. Astrovirus Encephalitis in Boy with X-linked Agammaglobulinemia. Emerging Infectious Diseases. 2010;16(6):918-25.

135. Brown J.R., Morfopoulou S., Hubb J., Emmett W.A., Ip W., Shah D., et al. Astrovirus VA1/HMO-C: An Increasingly Recognized Neurotropic Pathogen in Immunocompromised Patients. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America. 2015;60(6):881-8.

136. Céline L.B., Michel C.K., Daniel W., Simea W., Dennis J., Rémy B., et al. Indication of Cross-Species Transmission of Astrovirus Associated with Encephalitis in Sheep and Cattle. Emerging Infectious Disease journal. 2017;23(9):1604.

137. Li L., Diab S., McGraw S., Barr B., Traslavina R., Higgins R., et al. Divergent Astrovirus Associated with Neurologic Disease in Cattle. Emerging Infectious Diseases. 2013;19(9):1385-92.

138. Blomström A.-L., Widén F., Hammer A.-S., Belák S., Berg M. Detection of a Novel Astrovirus in Brain Tissue of Mink Suffering from Shaking Mink Syndrome by Use of Viral Metagenomics. Journal of Clinical Microbiology. 2010;48(12):4392-6.

139. Bouzalas I.G., Wüthrich D., Walland J., Drögemüller C., Zurbriggen A., Vandevelde M., et al. Neurotropic astrovirus in cattle with nonsuppurative encephalitis in Europe. Journal of clinical microbiology. 2014;52(9):3318-24.

140. Seuberlich T., Wüthrich D., Selimovic-Hamza S., Drögemüller C., Oevermann A., Bruggmann R., et al. Identification of a second encephalitis-associated astrovirus in cattle. Emerging Microbes & Infections. 2016;5(1):e5.

141. Wuthrich D., Boujon C.L., Truchet L., Selimovic-Hamza S., Oevermann A., Bouzalas I.G., et al. Exploring the virome of cattle with non-suppurative encephalitis of unknown etiology by metagenomics. Virology. 2016;493:22-30.
142. Blomström A.-L., Ley C., Jacobson M. Astrovirus as a possible cause of congenital tremor type AII in piglets? Acta veterinaria Scandinavica. 2014;56(1):82-.
143. Boujon C.L., Koch M.C., Kauer R.V., Keller-Gautschi E., Hierweger M.M., Hoby S., et al. Novel encephalomyelitis-associated astrovirus in a muskox (Ovibos moschatus): a surprise from the archives. Acta Vet Scand. 2019;61(1):31.
144. Giannitti F., Caffarena R.D., Pesavento P., Uzal F.A., Maya L., Fraga M., et al. The First Case of Bovine Astrovirus-Associated Encephalitis in the Southern Hemisphere (Uruguay), Uncovers Evidence of Viral Introduction to the Americas From Europe. Frontiers in microbiology. 2019;10:1240.
145. Koopmans M.P., Bijen M.H., Monroe S.S., Vinjé J. Age-stratified seroprevalence of neutralizing antibodies to astrovirus types 1 to 7 in humans in The Netherlands. Clinical and diagnostic laboratory immunology. 1998;5(1):33-7.
146. Kriston S., Willcocks M.M., Carter M.J., Cubitt W.D. Seroprevalence of astrovirus types 1 and 6 in London, determined using recombinant virus antigen. Epidemiology and infection. 1996;117(1):159-64.
147. Mitchell D.K., Matson D.O., Cubitt W.D., Jackson L.J., Willcocks M.M., Pickering L.K., et al. Prevalence of antibodies to astrovirus types 1 and 3 in children and adolescents in Norfolk, Virginia. The Pediatric infectious disease journal. 1999;18(3):249-54.
148. Kurtz J., Lee T. Astrovirus gastroenteritis age distribution of antibody. Medical microbiology and immunology. 1978;166(1-4):227-30.
149. Guix S., Caballero S., Villena C., Bartolomé R., Latorre C., Rabella N., et al. Molecular epidemiology of astrovirus infection in Barcelona, Spain. Journal of clinical microbiology. 2002;40(1):133-9.
150. Cortez V., Freiden P., Gu Z., Adderson E., Hayden R., Schultz-Cherry S. Persistent Infections with Diverse Co-Circulating Astroviruses in Pediatric Oncology Patients, Memphis, Tennessee, USA. Emerging infectious diseases. 2017;23(2):288-90.
151. Mendez-Toss M., Griffin D.D., Calva J., Contreras J.F., Puerto F.I., Mota F., et al. Prevalence and genetic diversity of human astroviruses in Mexican children with symptomatic and asymptomatic infections. Journal of clinical microbiology. 2004;42(1):151-7.

152. Maldonado Y., Cantwell M., Old M., Hill D., Sanchez M.L., Logan L., et al. Population-based prevalence of symptomatic and asymptomatic astrovirus infection in rural Mayan infants. The Journal of infectious diseases. 1998;178(2):334-9.

153. Méndez E. and Arias C.F. Astroviruses. D.M. Knipe P.M.H., editor. In: Fields Virology. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2007. p. 981-1000.

154. Reynolds D.L., Saif Y.M., Theil K.W. Enteric viral infections of turkey poults: incidence of infection. Avian Dis. 1987;31(2):272-6.

155. Frazier J.A., Howes K., Reece R.L., Kidd A.W., Cavanagh D. Isolation of non-cytopathic viruses implicated in the aetiology of nephritis and baby chick nephropathy and serologically related to avian nephritis virus. Avian pathology : journal of the WVPA. 1990;19(1):139-60.

156. Imada T., Taniguchi T., Yamaguchi S., Minetoma T., Maeda M., Kawamura H. Susceptibility of chickens to avian nephritis virus at various inoculation routes and ages. Avian Dis. 1981;25(2):294-302.

157. Narita M., Ohta K., Kawamura H., Shirai J., Nakamura K., Abe F. Pathogenesis of renal dysfunction in chicks experimentally induced by avian nephritis virus. Avian pathology : journal of the WVPA. 1990;19(3):571-82.

158. Liu N., Jiang M., Wang M., Wang F., Zhang B., Zhang D. Isolation and detection of duck astrovirus CPH: implications for epidemiology and pathogenicity. Avian pathology : journal of the WVPA. 2016;45(2):221-7.

159. Sajewicz-Krukowska J., Pac K., Lisowska A., Pikula A., Minta Z., Kroliczewska B., et al. Astrovirus-induced "white chicks" condition - field observation, virus detection and preliminary characterization. Avian Pathology. 2016;45(1):2-12.

160. Canelli E., Cordioli P., Barbieri I., Catella A., Pennelli D., Ceruti R., et al. Astroviruses as causative agents of poultry enteritis: genetic characterization and longitudinal studies on field conditions. Avian Dis. 2012;56(1):173-82.

161. Fu Y., Pan M., Wang X., Xu Y., Xie X., Knowles N.J., et al. Complete sequence of a duck astrovirus associated with fatal hepatitis in ducklings. J Gen Virol. 2009;90(Pt 5):1104-8.

162. Imada T., Taniguchi T., Sato S., Yamaguchi S., Kawamura H. Pathogenicity of avian nephritis virus for embryonating hen's eggs. National Institute of Animal Health quarterly. 1982;22(1):8-15.

163. Shirai J., Nakamura K., Shinohara K., Kawamura H. Pathogenicity and Antigenicity of Avian Nephritis Isolates. Avian diseases. 1991;35(1):49-54.

164. Smyth V., Trudgett J., Wylie M., Jewhurst H., Conway B., Welsh M., et al. Chicken astrovirus detected in hatchability problems associated with 'white chicks'. Veterinary Record. 2013;173(16):403-4.

165. Nunez L.F.N., Parra S.H.S., Carranza C., Astolfi-Ferreira C.S., Buim M.R., Ferreira A.J.P. Detection and molecular characterization of chicken astrovirus associated with chicks that have an unusual condition known as "white chicks" in Brazil. Poultry science. 2016;95(6):1262-70.

166. Bulbule N.R., Mandakhalikar K.D., Kapgate S.S., Deshmukh V.V., Schat K.A., Chawak M.M. Role of chicken astrovirus as a causative agent of gout in commercial broilers in India. Avian Pathology. 2013;42(5):464-73.

167. Sajewicz-Krukowska J., Domanska-Blicharz K. Nearly full-length genome sequence of a novel astrovirus isolated from chickens with 'white chicks' condition. Archives of virology. 2016;161(9):2581-7.

168. Koci M.D., Seal B.S., Schultz-Cherry S. Molecular characterization of an avian astrovirus. J Virol. 2000;74(13):6173-7.

169. Pantin-Jackwood M.J., Strother K.O., Mundt E., Zsak L., Day J.M., Spackman E. Molecular characterization of avian astroviruses. Archives of virology. 2011;156(2):235-44.

170. Qureshi M.A., Yu M., Saif Y.M. A novel "small round virus" inducing poult enteritis and mortality syndrome and associated immune alterations. Avian diseases. 2000;44(2):275-83.

171. Qureshi M.A., Saif Y.M., Heggen-Peay C.L., Edens F.W., Havenstein G.B. Induction of functional defects in macrophages by a poult enteritis and mortality syndrome-associated turkey astrovirus. Avian diseases. 2001;45(4):853-61.

172. Qureshi M.A., Edens F.W., Havenstein G.B. Immune system dysfunction during exposure to poult enteritis and mortality syndrome agents. Poultry science. 1997;76(4):564-9.

173. Koci M.D., Moser L.A., Kelley L.A., Larsen D., Brown C.C., Schultz-Cherry S. Astrovirus Induces Diarrhea in the Absence of Inflammation and Cell Death. J Virol. 2003;77(21):11798.

174. Noel J., Cubitt D. Identification of astrovirus serotypes from children treated at the Hospitals for Sick Children, London 1981-93. Epidemiology and infection. 1994;113(1):153-9.

175. Oishi I., Yamazaki K., Kimoto T., Minekawa Y., Utagawa E., Yamazaki S., et al. A large outbreak of acute gastroenteritis associated with astrovirus among students and teachers in Osaka, Japan. The Journal of infectious diseases. 1994;170(2):439-43.
176. Taylor M.B., Marx F.E., Grabow W.O. Rotavirus, astrovirus and adenovirus associated with an outbreak of gastroenteritis in a South African child care centre. Epidemiology and infection. 1997;119(2):227-30.
177. Mettifogo E., Nunez L.F.N., Chacon J.L., Parra S.H.S., Astolfi-Ferreira C.S., Jerez J.A., et al. Emergence of Enteric Viruses in Production Chickens Is a Concern for Avian Health. Sci World J. 2014.
178. Ter Veen C., de Bruijn N.D., Dijkman R., de Wit J.J. Prevalence of histopathological intestinal lesions and enteric pathogens in Dutch commercial broilers with time. Avian pathology : journal of the WVPA. 2017;46(1):95-105.
179. Domanska-Blicharz K., Seroka A., Minta Z. One-year molecular survey of astrovirus infection in turkeys in Poland. Archives of virology. 2011;156(6):1065-72.
180. Todd D., Wilkinson D.S., Jewhurst H.L., Wylie M., Gordon A.W., Adair B.M. A seroprevalence investigation of chicken astrovirus infections. Avian Pathology. 2009;38(4):301-9.
181. Reynolds D.L., Saif Y.M., Theil K.W. A survey of enteric viruses of turkey poults. Avian Dis. 1987;31(1):89-98.
182. Connor T.J., McNeilly F., McFerran J.B., McNulty M.S. A survey of avian sera from Northern Ireland for antibody to avian nephritis virus. Avian Pathology. 1987;16(1):15-20.
183. Imada T., Yamaguchi S., Miura N., Kawamura H. Antibody survey against avian nephritis virus among chickens in Japan. National Institute of Animal Health quarterly. 1980;20(2):79-80.
184. McNulty M.S., Connor T.J., McNeilly F. A survey of specific pathogen-free chicken flocks for antibodies to chicken anaemia agent, avian nephritis virus and group a rotavirus. Avian pathology : journal of the WVPA. 1989;18(2):215-20.
185. Nicholas R.A., Goddard R.D., Luff P.R. Prevalence of avian nephritis virus in England. The Veterinary record. 1988;123(15):398.
186. Todd D., Trudgett J., Smyth V.J., Donnelly B., McBride N., Welsh M.D. Capsid protein sequence diversity of avian nephritis virus. Avian Pathology. 2011;40(3):249-59.
187. Sellers H., Linneman E., Icard A.H., Mundt E. A purified recombinant baculovirus expressed capsid protein of a new astrovirus provides partial protection to runting-stunting syndrome in chickens. Vaccine. 2010;28(5):1253-63.

188. Bogdanoff W.A., Perez E.I., López T., Arias C.F., DuBois R.M. Structural Basis for Escape of Human Astrovirus from Antibody Neutralization: Broad Implications for Rational Vaccine Design. J Virol. 2018;92(1):e01546-17.

189. MacLachlan, N. James. Caliciviridae and Astroviridae. Dubovi E.J., editor. In: Fenner's Veterinary Virology (Fifth Edition). Boston: Academic Press; 2017. p.497-510.

190. Reynolds D.L., Schultz-Cherry S.L. Astrovirus infections 2003. p320-6.

191. Madeley C.R., Cosgrove B.P. Letter: 28 nm particles in faeces in infantile gastroenteritis. Lancet (London, England). 1975;2(7932):451-2.

192. Koci M.D., Schultz-Cherry S. Avian astroviruses. Avian pathology : journal of the WVPA. 2002;31(3):213-27.

193. Dryden K.A., Tihova M., Nowotny N., Matsui S.M., Mendez E., Yeager M. Immature and Mature Human Astrovirus: Structure, Conformational Changes, and Similarities to Hepatitis E Virus. Journal of Molecular Biology. 2012;422(5):650-8.

194. York R.L., Yousefi P.A., Bogdanoff W., Haile S., Tripathi S., DuBois R.M. Structural, Mechanistic, and Antigenic Characterization of the Human Astrovirus Capsid. J Virol. 2015;90(5):2254-63.

195. Schultz-Cherry S., King D.J., Koci M.D. Inactivation of an astrovirus associated with poult enteritis mortality syndrome. Avian Dis. 2001;45(1):76-82.

196. Kurtz J.B., Lee T.W., Parsons A.J. The action of alcohols on rotavirus, astrovirus and enterovirus. The Journal of hospital infection. 1980;1(4):321-5.

197. McNeilly F., Connor T.J., Calvert V.M., Smyth J.A., Curran W.L., Morley A.J., et al. Studies on a new enterovirus-like virus isolated from chickens. Avian Pathology. 1994;23(2):313-27.

198. Arias C.F., DuBois R.M. The Astrovirus Capsid: A Review. Viruses. 2017;9(1):15.

199. Willcocks M.M., Brown T.D., Madeley C.R., Carter M.J. The complete sequence of a human astrovirus. J Gen Virol. 1994;75 (Pt 7):1785-8.

200. Imada T., Yamaguchi S., Mase M., Tsukamoto K., Kubo M., Morooka A. Avian nephritis virus (ANV) as a new member of the family Astroviridae and construction of infectious ANV cDNA. J Virol. 2000;74(18):8487-93.

201. Tengs T., Kristoffersen A.B., Bachvaroff T.R., Jonassen C.M. A mobile genetic element with unknown function found in distantly related viruses. Virology journal. 2013;10.

202. Koci M.D., Seal B.S., Schultz-Cherry S. Molecular characterization of an avian astrovirus. J Virol. 2000;74(13):6173-7.

203. Chu D.K.W., Chin A.W.H., Smith G.J., Chan K.-H., Guan Y., Peiris J.S.M., et al. Detection of novel astroviruses in urban brown rats and previously known astroviruses in humans. The Journal of general virology. 2010;91(Pt 10):2457-62.

204. Chu D.K., Poon L.L., Guan Y., Peiris J.S. Novel astroviruses in insectivorous bats. J Virol. 2008;82(18):9107-14.

205. Bosch A., Pintó R.M., Guix S. Human astroviruses. Clinical microbiology reviews. 2014;27(4):1048-74.

206. De Nova-Ocampo M., Soliman M.C., Espinosa-Hernández W., Velez-del Valle C., Salas-Benito J., Valdés-Flores J., et al. Human astroviruses: in silico analysis of the untranslated region and putative binding sites of cellular proteins. Molecular Biology Reports. 2019;46(1):1413-24.

207. Matsui S.M., Kim J.P., Greenberg H.B., Young L.M., Smith L.S., Lewis T.L., et al. Cloning and Characterization of Human Astrovirus Immunoreactive Epitopes. J Virol. 1993;67(3):1712-5.

208. Guix S., Caballero S., Bosch A., Pinto R.M. C-terminal nsP1a protein of human astrovirus colocalizes with the endoplasmic reticulum and viral RNA. J Virol. 2004;78(24):13627-36.

209. Guix S., Bosch A., Ribes E., Dora Martinez L., Pinto R.M. Apoptosis in astrovirus-infected CaCo-2 cells. Virology. 2004;319(2):249-61.

210. Al-Mutairy B., Walter J.E., Pothen A., Mitchell D.K. Genome prediction of putative genome-linked viral protein (VPg) of astroviruses. Virus Genes. 2005;31(1):21-30.

211. Imada T., Yamaguchi S., Mase M., Tsukamoto K., Kubo M., Morooka A. Avian nephritis virus (ANV) as a new member of the family Astroviridae and construction of infectious ANV cDNA. J Virol. 2000;74(18):8487-93.

212. Jonassen C.M., Jonassen T.T., Sveen T.M., Grinde B. Complete genomic sequences of astroviruses from sheep and turkey: comparison with related viruses. Virus Res. 2003;91(2):195-201.

213. Willcocks M.M., Ashton N., Kurtz J.B., Cubitt W.D., Carter M.J. Cell culture adaptation of astrovirus involves a deletion. J Virol. 1994;68(9):6057-8.

214. Guix S., Caballero S., Bosch A., Pinto R.M. Human astrovirus C-terminal nsP1a protein is involved in RNA replication. Virology. 2005;333(1):124-31.

215. Willcocks M.M., Boxall A.S., Carter M.J. Processing and intracellular location of human astrovirus non-structural proteins. J Gen Virol. 1999;80 (Pt 10):2607-11.

216. Lewis T.L., Greenberg H.B., Herrmann J.E., Smith L.S., Matsui S.M. Analysis of Astrovirus Serotype-1 Rna, Identification of the Viral Rna-Dependent Rna-Polymerase Motif, and Expression of a Viral Structural Protein. J Virol. 1994;68(1):77-83.
217. Kang K.I., Icard A.H., Linnemann E., Sellers H.S., Mundt E. Determination of the full length sequence of a chicken astrovirus suggests a different replication mechanism. Virus Genes. 2012;44(1):45-50.
218. Fernández-Correa I., Truchado D.A., Gomez-Lucia E., Doménech A., Pérez-Tris J., Schmidt-Chanasit J., et al. A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of Astroviridae. Scientific reports. 2019;9(1):9513-.
219. Monroe S.S., Jiang B., Stine S.E., Koopmans M., Glass R.I. Subgenomic RNA Sequence of Human Astrovirus Supports Classification of Astroviridae as a New Family of Rna Viruses. J Virol. 1993;67(6):3611-4.
220. Toh Y., Harper J., Dryden K.A., Yeager M., Arias C.F., Mendez E., et al. Crystal Structure of the Human Astrovirus Capsid Protein. J Virol. 2016;90(20):9008-17.
221. DuBois R.M., Freiden P., Marvin S., Reddivari M., Heath R.J., White S.W., et al. Crystal structure of the avian astrovirus capsid spike. J Virol. 2013;87(14):7853-63.
222. Bogdanoff W.A., Campos J., Perez E.I., Yin L., Alexander D.L., DuBois R.M. Structure of a Human Astrovirus Capsid-Antibody Complex and Mechanistic Insights into Virus Neutralization. J Virol. 2017;91(2).
223. Kelly A.G., Netzler N.E., White P.A. Ancient recombination events and the origins of hepatitis E virus. BMC evolutionary biology. 2016;16(1):210.
224. Pankovics P., Boros A., Kiss T., Engelmann P., Reuter G. Genetically highly divergent RNA virus with astrovirus-like (5'-end) and hepevirus-like (3'-end) genome organization in carnivorous birds, European roller (Coracias garrulus). Infect Genet Evol. 2019;71:215-23.
225. Krishna N.K. Identification of structural domains involved in astrovirus capsid biology. Viral immunology. 2005;18(1):17-26.
226. Jonassen C.M., Jonassen T.O., Saif Y.M., Snodgrass D.R., Ushijima H., Shimizu M., et al. Comparison of capsid sequences from human and animal astroviruses. J Gen Virol. 2001;82(Pt 5):1061-7.
227. Wang Q.H., Kakizawa J., Wen L.Y., Shimizu M., Nishio O., Fang Z.Y., et al. Genetic analysis of the capsid region of astroviruses. Journal of medical virology. 2001;64(3):245-55.

228. Rao A.L., Grantham G.L. Molecular studies on bromovirus capsid protein. II. Functional analysis of the amino-terminal arginine-rich motif and its role in encapsidation, movement, and pathology. Virology. 1996;226(2):294-305.
229. Schmitz I., Rao A.L.N. Deletions in the conserved amino-terminal basic arm of cucumber mosaic virus coat protein disrupt virion assembly but do not abolish infectivity and cell-to-cell movement. Virology. 1998;248(2):323-31.
230. Jacobs A., Hoover H., Smith E., Clemmer D.E., Kim C.-H., Kao C.C. The intrinsically disordered N-terminal arm of the brome mosaic virus coat protein specifically recognizes the RNA motif that directs the initiation of viral RNA replication. Nucleic Acids Research. 2018;46(1):324-35.
231. Geigenmuller U., Ginzton N.H., Matsui S.M. Studies on intracellular processing of the capsid protein of human astrovirus serotype 1 in infected cells. J Gen Virol. 2002;83(Pt 7):1691-5.
232. Baer M.L., Houser F., Loesch-Fries L.S., Gehrke L. Specific RNA binding by amino-terminal peptides of alfalfa mosaic virus coat protein. The EMBO journal. 1994;13(3):727-35.
233. Geigenmuller-Gnirke U., Nitschko H., Schlesinger S. Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. J Virol. 1993;67(3):1620-6.
234. Marshall D., Schneemann A. Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein. Virology. 2001;285(1):165-75.
235. Bass D.M., Upadhyayula U. Characterization of human serotype 1 astrovirus-neutralizing epitopes. J Virol. 1997;71(11):8666-71.
236. Sanchez-Fauquier A., Carrascosa A.L., Carrascosa J.L., Otero A., Glass R.I., Lopez J.A., et al. Characterization of a human astrovirus serotype 2 structural protein (VP26) that contains an epitope involved in virus neutralization. Virology. 1994;201(2):312-20.
237. Brinker J.P., Blacklow N.R., Herrmann J.E. Human astrovirus isolation and propagation in multiple cell lines. Archives of virology. 2000;145(9):1847-56.
238. Mendez E., Salas-Ocampo E., Arias C.F. Caspases mediate processing of the capsid precursor and cell release of human astroviruses. J Virol. 2004;78(16):8601-8.
239. Banos-Lara Mdel R., Mendez E. Role of individual caspases induced by astrovirus on the processing of its structural protein and its release from the cell through a non-lytic mechanism. Virology. 2010;401(2):322-32.

240. Ball J.M., Tian P., Zeng C.Q., Morris A.P., Estes M.K. Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. Science (New York, NY). 1996;272(5258):101-4.

241. Swaggerty C.L., Frolov A.A., McArthur M.J., Cox V.W., Tong S., Compans R.W., et al. The envelope glycoprotein of simian immunodeficiency virus contains an enterotoxin domain. Virology. 2000;277(2):250-61.

242. Meliopoulos V.A., Marvin S.A., Freiden P., Moser L.A., Nighot P., Ali R., et al. Oral Administration of Astrovirus Capsid Protein Is Sufficient To Induce Acute Diarrhea In Vivo. MBio. 2016;7(6).

243. Nighot P.K., Moeser A., Ali R.A., Blikslager A.T., Koci M.D. Astrovirus infection induces sodium malabsorption and redistributes sodium hydrogen exchanger expression. Virology. 2010;401(2):146-54.

244. Moser L.A., Carter M., Schultz-Cherry S. Astrovirus increases epithelial barrier permeability independently of viral replication. J Virol. 2007;81(21):11937-45.

245. Johnson C., Hargest V., Cortez V., Meliopoulos V.A., Schultz-Cherry S. Astrovirus Pathogenesis. Viruses. 2017;9(1):22.

246. Laohachai K.N., Bahadi R., Hardo M.B., Hardo P.G., Kourie J.I. The role of bacterial and non-bacterial toxins in the induction of changes in membrane transport: implications for diarrhea. Toxicon : official journal of the International Society on Toxinology. 2003;42(7):687-707.

247. Hair P.S., Gronemus J.Q., Crawford K.B., Salvi V.P., Cunnion K.M., Thielens N.M., et al. Human astrovirus coat protein binds C1q and MBL and inhibits the classical and lectin pathways of complement activation. Molecular immunology. 2010;47(4):792-8.

248. Gronemus J.Q., Hair P.S., Crawford K.B., Nyalwidhe J.O., Cunnion K.M., Krishna N.K. Potent inhibition of the classical pathway of complement by a novel C1q-binding peptide derived from the human astrovirus coat protein. Molecular immunology. 2010;48(1-3):305-13.

249. Martella V., Moschidou P., Lorusso E., Mari V., Camero M., Bellacicco A., et al. Detection and characterization of canine astroviruses. J Gen Virol. 2011;92(Pt 8):1880-7.

250. Koci M.D., Kelley L.A., Larsen D., Schultz-Cherry S. Astrovirus-Induced Synthesis of Nitric Oxide Contributes to Virus Control during Infection. J Virol. 2004;78(3):1564.

251. Cortez V., Sharp B., Yao J., Livingston B., Vogel P., Schultz-Cherry S. Characterizing a Murine Model for Astrovirus Using Viral Isolates from Persistently Infected Immunocompromised Mice. J Virol. 2019;93(13).

252. Compton S.R., Booth C.J., Macy J.D. Murine Astrovirus Infection and Transmission in Neonatal CD1 Mice. J Am Assoc Lab Anim Sci. 2017;56(4):402-11.
253. Vu D.L., Bosch A., Pinto R.M., Ribes E., Guix S. Human Astrovirus MLB Replication In Vitro: Persistence in Extraintestinal Cell Lines. J Virol. 2019;93(13).
254. Gough R.E., Borland E.D., Keymer I.F., Stuart J.C. An outbreak of duck hepatitis type ii in commercial ducks. Avian Pathology. 1985;14(2):227-36.
255. Takase K., Uchimura T., Yamamoto M., Yamada S. Susceptibility of embryos and chicks, derived from immunized breeding hens, to avian nephritis virus. Avian Pathology. 1994;23(1):117-25.
256. Smyth V.J., Todd D., Trudgett J., Lee A., Welsh M.D. Capsid protein sequence diversity of chicken astrovirus. Avian Pathology. 2012;41(2):151-9.
257. Geigenmuller U., Ginzton N.H., Matsui S.M. Construction of a genome-length cDNA clone for human astrovirus serotype 1 and synthesis of infectious RNA transcripts. J Virol. 1997;71(2):1713-7.
258. Qin Y., Fang Q., Liu H., Ji C., Chen Y., Ouyang K., et al. Construction of a reverse genetic system for porcine astrovirus. Archives of virology. 2018.
259. Sandoval-Jaime C., Guzmán-Ruiz L., López S., Arias C.F. Development of a novel DNA based reverse genetics system for classic human astroviruses. Virology. 2019;535:130-5.
260. Méndez E., Muñoz-Yañez C., Sánchez-San Martín C., Aguirre-Crespo G., Baños-Lara M.d.R., Gutierrez M., et al. Characterization of human astrovirus cell entry. J Virol. 2014;88(5):2452-60.
261. Dong J., Dong L., Méndez E., Tao Y. Crystal structure of the human astrovirus capsid spike. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(31):12681-6.
262. Donelli G., Superti F., Tinari A., Marziano M.L. Mechanism of astrovirus entry into Graham 293 cells. Journal of medical virology. 1992;38(4):271-7.
263. Moser L.A., Schultz-Cherry S. Suppression of astrovirus replication by an ERK1/2 inhibitor. J Virol. 2008;82(15):7475-82.
264. Luo H., Yanagawa B., Zhang J., Luo Z., Zhang M., Esfandiarei M., et al. Coxsackievirus B3 replication is reduced by inhibition of the extracellular signal-regulated kinase (ERK) signaling pathway. J Virol. 2002;76(7):3365-73.
265. Jiang R., Lopez V., Kelleher S.L., Lonnerdal B. Apo- and holo-lactoferrin are both internalized by lactoferrin receptor via clathrin-mediated endocytosis but differentially

affect ERK-signaling and cell proliferation in Caco-2 cells. Journal of cellular physiology. 2011;226(11):3022-31.

266. Caballero S., Guix S., Ribes E., Bosch A., Pinto R.M. Structural requirements of astrovirus virus-like particles assembled in insect cells. J Virol. 2004;78(23):13285-92.

267. Goodfellow I., Chaudhry Y., Gioldasi I., Gerondopoulos A., Natoni A., Labrie L., et al. Calicivirus translation initiation requires an interaction between VPg and eIF 4 E. EMBO Rep. 2005;6(10):968-72.

268. Velazquez-Moctezuma R., Banos-Lara Mdel R., Acevedo Y., Mendez E. Alternative cell lines to improve the rescue of infectious human astrovirus from a cDNA clone. Journal of virological methods. 2012;179(2):295-302.

269. Fuentes C., Guix S., Bosch A., Pinto R.M. The C-terminal nsP1a protein of human astrovirus is a phosphoprotein that interacts with the viral polymerase. J Virol. 2011;85(9):4470-9.

270. Zhao W., Tao X.L., Liu N., Lu H.Z., Zheng C.H., Li X.Y., et al. Host protein CD63 promotes viral RNA replication by interacting with human astrovirus non-structural protein nsP1a/4. J Gen Virol. 2019;100(4):616-28.

271. Lewis T.L., Matsui S.M. Astrovirus ribosomal frameshifting in an infection-transfection transient expression system. J Virol. 1996;70(5):2869-75.

272. Méndez E., Murillo A., Velázquez R., Burnham A., Arias C.F. Replication Cycle of Astroviruses. Schultz-Cherry S., editor. In: Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions. New York, NY: Springer New York; 2013. p. 19-45.

273. Gibson C.A., Chen J., Monroe S.A., Denison M.R. Expression and processing of nonstructural proteins of the human astroviruses. Advances in experimental medicine and biology. 1998;440:387-91.

274. Kiang D., Matsui S.M. Proteolytic processing of a human astrovirus nonstructural protein. J Gen Virol. 2002;83(Pt 1):25-34.

275. Mendez E., Salas-Ocampo M.P., Munguia M.E., Arias C.F. Protein products of the open reading frames encoding nonstructural proteins of human astrovirus serotype 8. J Virol. 2003;77(21):11378-84.

276. Isobe T., Tange S., Tasaki H., Kanamori K., Kato A., Nakanishi A. Upregulation of CHOP participates in caspase activation and virus release in human astrovirus-infected cells. J Gen Virol. 2019;100(5):778-92.

277. Murillo A., Vera-Estrella R., Barkla B.J., Méndez E., Arias C.F. Identification of Host Cell Factors Associated with Astrovirus Replication in Caco-2 Cells. J Virol. 2015;89(20):10359-70.

278. Blackham S., Baillie A., Al-Hababi F., Remlinger K., You S., Hamatake R., et al. Gene expression profiling indicates the roles of host oxidative stress, apoptosis, lipid metabolism, and intracellular transport genes in the replication of hepatitis C virus. J Virol. 2010;84(10):5404-14.

279. Perera R., Riley C., Isaac G., Hopf-Jannasch A.S., Moore R.J., Weitz K.W., et al. Dengue virus infection perturbs lipid homeostasis in infected mosquito cells. PLoS pathogens. 2012;8(3):e1002584.

280. Mackenzie J.M., Khromykh A.A., Parton R.G. Cholesterol manipulation by West Nile virus perturbs the cellular immune response. Cell host & microbe. 2007;2(4):229-39.

281. Stapleford K.A., Miller D.J. Role of cellular lipids in positive-sense RNA virus replication complex assembly and function. Viruses. 2010;2(5):1055-68.

282. Lim Y.S., Hwang S.B. Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type IIIalpha and regulates viral propagation. The Journal of biological chemistry. 2011;286(13):11290-8.

283. Monroe S.S., Stine S.E., Gorelkin L., Herrmann J.E., Blacklow N.R., Glass R.I. Temporal Synthesis of Proteins and Rnas during Human Astrovirus Infection of Cultured-Cells. J Virol. 1991;65(2):641-8.

284. Jang S.Y., Jeong W.H., Kim M.S., Lee Y.M., Lee J.I., Lee G.C., et al. Detection of replicating negative-sense RNAs in CaCo-2 cells infected with human astrovirus. Archives of virology. 2010;155(9):1383-9.

285. Gorbalenya A.E., Koonin E.V. Viral proteins containing the purine NTP-binding sequence pattern. Nucleic acids research. 1989;17(21):8413-40.

286. Jiang B., Monroe S.S., Koonin E.V., Stine S.E., Glass R.I. RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(22):10539-43.

287. Ahlquist P., Noueiry A.O., Lee W.-M., Kushner D.B., Dye B.T. Host factors in positive-strand RNA virus genome replication. J Virol. 2003;77(15):8181-6.

288. Espinosa-Hernández W., Velez-Uriza D., Valdés J., Vélez-Del Valle C., Salas-Benito J., Martínez-Contreras R., et al. PTB binds to the 3' untranslated region of the human astrovirus type 8: a possible role in viral replication. PloS one. 2014;9(11):e113113-e.

289. Bedard K.M., Daijogo S., Semler B.L. A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. The EMBO journal. 2007;26(2):459-67.
290. Mendez E., Aguirre-Crespo G., Zavala G., Arias C.F. Association of the astrovirus structural protein VP90 with membranes plays a role in virus morphogenesis. J Virol. 2007;81(19):10649-58.
291. Mendez E., Fernandez-Luna T., Lopez S., Mendez-Toss M., Arias C.F. Proteolytic processing of a serotype 8 human astrovirus ORF2 polyprotein. J Virol. 2002;76(16):7996-8002.
292. Bass D.M., Qiu S. Proteolytic processing of the astrovirus capsid. J Virol. 2000;74(4):1810-4.
293. Yamashita T., Mori Y., Miyazaki N., Cheng R.H., Yoshimura M., Unno H., et al. Biological and immunological characteristics of hepatitis E virus-like particles based on the crystal structure. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(31):12986-91.
294. Li S., Tang X., Seetharaman J., Yang C., Gu Y., Zhang J., et al. Dimerization of Hepatitis E Virus Capsid Protein E2s Domain Is Essential for Virus–Host Interaction. PLoS pathogens. 2009;5(8):e1000537.
295. Lulla V., Firth A.E. A hidden gene in astroviruses encodes a cell-permeabilizing protein involved in virus release. bioRxiv. 2019:661579.
296. Aguilar-Hernández N., López S., Arias C.F. Minimal capsid composition of infectious human astrovirus. Virology. 2018;521:58-61.
297. Buchholz U.J., Finke S., Conzelmann K.K. Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J Virol. 1999;73(1):251-9.
298. Kawaguchi T., Nomura K., Hirayama Y., Kitagawa T. Establishment and characterization of a chicken hepatocellular carcinoma cell line, LMH. Cancer research. 1987;47(16):4460-4.
299. Himly M., Foster D.N., Bottoli I., Iacovoni J.S., Vogt P.K. The DF-1 chicken fibroblast cell line: transformation induced by diverse oncogenes and cell death resulting from infection by avian leukosis viruses. Virology. 1998;248(2):295-304.
300. Hennion R.M., Hill G. The preparation of chicken kidney cell cultures for virus propagation. Methods in molecular biology. 2015;1282:57-62.

301. Smyth V.J., Jewhurst H.L., Wilkinson D.S., Adair B.M., Gordon A.W., Todd D. Development and evaluation of real-time TaqMan(R) RT-PCR assays for the detection of avian nephritis virus and chicken astrovirus in chickens. Avian pathology : journal of the WVPA. 2010;39(6):467-74.
302. Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296-W303.
303. Ye Y., Godzik A. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 2004;32(Web Server issue):W582-5.
304. Yang J., Roy A., Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics (Oxford, England). 2013;29(20):2588-95.
305. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
306. Saunders K., Lomonossoff G.P. The Generation of Turnip Crinkle Virus-Like Particles in Plants by the Transient Expression of Wild-Type and Modified Forms of Its Coat Protein. Frontiers in Plant Science. 2015;6.
307. Kärber G. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie. 1931;162(4):480-3.
308. Andrews S. FastQC: a quality control tool for high throughput sequence data (version 0.11.5). Babraham, UK. 2010.
309. Schmieder R., Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics (Oxford, England). 2011;27(6):863-4.
310. Joshi N.A., Fass J.N. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). 2011.
311. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013;1303.
312. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078-9.
313. Larson S.B., Koszelak S., Day J., Greenwood A., Dodds J.A., McPherson A. Double-helical RNA in satellite tobacco mosaic virus. Nature. 1993;361(6408):179-82.

314. Koning R.I., Gomez-Blanco J., Akopjana I., Vargas J., Kazaks A., Tars K., et al. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nature Communications. 2016;7:12524.
315. Delany I., Rappuoli R., De Gregorio E. Vaccines for the 21st century. EMBO Molecular Medicine. 2014;6(6):708-20.
316. Smith M.L., Mason H.S., Shuler M.L. Hepatitis B surface antigen (HBsAg) expression in plant cell culture: Kinetics of antigen accumulation in batch culture and its intracellular form. Biotechnology and bioengineering. 2002;80(7):812-22.
317. Paavonen J., Jenkins D., Bosch F.X., Naud P., Salmeron J., Wheeler C.M., et al. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. Lancet (London, England). 2007;369(9580):2161-70.
318. Huang D.B., Wu J.J., Tyring S.K. A review of licensed viral vaccines, some of their safety concerns, and the advances in the development of investigational viral vaccines. Journal of Infection. 2004;49(3):179-209.
319. Zeltins A. Construction and characterization of virus-like particles: a review. Molecular biotechnology. 2013;53(1):92-107.
320. Roldao A., Mellado M.C., Castilho L.R., Carrondo M.J., Alves P.M. Virus-like particles in vaccine development. Expert review of vaccines. 2010;9(10):1149-76.
321. Huang X., Wang X., Zhang J., Xia N., Zhao Q. Escherichia coli-derived virus-like particles in vaccine development. npj Vaccines. 2017;2(1):3.
322. Fuenmayor J., Godia F., Cervera L. Production of virus-like particles for vaccines. New biotechnology. 2017;39(Pt B):174-80.
323. Kim D.Y., Firth A.E., Atasheva S., Frolova E.I., Frolov I. Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. J Virol. 2011;85(16):8022-36.
324. Thuenemann E.C., Meyers A.E., Verwey J., Rybicki E.P., Lomonossoff G.P. A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. Plant Biotechnology Journal. 2013;11(7):839-46.
325. Saunders K., Sainsbury F., Lomonossoff G.P. Efficient generation of cowpea mosaic virus empty virus-like particles by the proteolytic processing of precursors in insect cells and plants. Virology. 2009;393(2):329-37.

326. Dubois R.M., Dryden K.A., Yeager M., Tao Y.J. Astrovirus Structure and Assembly. Schultz-Cherry S., editor. In: Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions. New York, NY: Springer New York; 2013. p. 47-64.
327. Yang C., Pan H., Wei M., Zhang X., Wang N., Gu Y., et al. Hepatitis E virus capsid protein assembles in 4M urea in the presence of salts. Protein Science : A Publication of the Protein Society. 2013;22(3):314-26.
328. Xing L., Li T.C., Mayazaki N., Simon M.N., Wall J.S., Moore M., et al. Structure of hepatitis E virion-sized particle reveals an RNA-dependent viral assembly pathway. The Journal of biological chemistry. 2010;285(43):33175-83.
329. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403-10.
330. Drake J.W. Rates of spontaneous mutation among RNA viruses. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(9):4171-5.
331. Ferrer-Orta C., Ferrero D., Verdaguer N. RNA-Dependent RNA Polymerases of Picornaviruses: From the Structure to Regulatory Mechanisms. Viruses. 2015;7(8):4438-60.
332. Pantin-Jackwood M.J., Strother K.O., Mundt E., Zsak L., Day J.M., Spackman E. Molecular characterization of avian astroviruses. Archives of virology. 2011;156(2):235-44.
333. Manzoni T.B., López C.B. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. Future Virology. 2018;13(7):493-503.
334. Kingsbury D.W., Portner A., Darlington R.W. Properties of incomplete Sendai virions and subgenomic viral RNAs. Virology. 1970;42(4):857-71.
335. Rott R., Scholtissek C. Investigations about the formation of incomplete forms of fowl plague virus. Journal of general microbiology. 1963;33:303-12.
336. Imada T., Yamaguchi S., Kawamura H. Pathogenicity for baby chicks of the G-4260 strain of the picornavirus "avian nephritis virus". Avian Dis. 1979;23(3):582-8.
337. den Boon J.A., Diaz A., Ahlquist P. Cytoplasmic viral replication complexes. Cell Host Microbe. 2010;8(1):77-85.
338. Mollier R.T. Molecular characterization of avian nephritis virus. Tirupati: Sri Venkateswara Veterinary University; 2014.
339. Mathews D.H. RNA secondary structure analysis using RNAstructure. Curr Protoc Bioinformatics. 2006;Chapter 12:Unit 12.6.
340. Patel N., Dykeman E.C., Coutts R.H.A., Lomonossoff G.P., Rowlands D.J., Phillips S.E.V., et al. Revealing the density of encoded functions in a viral RNA. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(7):2227-32.

341. Cortez V., Meliopoulos V.A., Karlsson E.A., Hargest V., Johnson C., Schultz-Cherry S. Astrovirus Biology and Pathogenesis. Annu Rev Virol. 2017;4(1):327-48.
342. Lok A.S., Akarca U., Greene S. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(9):4077-81.
343. Tam J.C.H., Jacques D.A. Intracellular immunity: finding the enemy within--how cells recognize and respond to intracellular pathogens. J Leukoc Biol. 2014;96(2):233-44.
344. Surjit M., Jameel S., Lal S.K. The ORF2 protein of hepatitis E virus binds the 5' region of viral RNA. J Virol. 2004;78(1):320-8.
345. Krupovic M., Koonin E.V. Multiple origins of viral capsid proteins from cellular ancestors. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(12):E2401-E10.
346. De Grazia S., Medici M.C., Pinto P., Moschidou P., Tummolo F., Calderaro A., et al. Genetic Heterogeneity and Recombination in Human Type 2 Astroviruses. Journal of Clinical Microbiology. 2012;50(11):3760.
347. Martella V., Medici M.C., Terio V., Catella C., Bozzo G., Tummolo F., et al. Lineage diversification and recombination in type-4 human astroviruses. Infection, Genetics and Evolution. 2013;20:330-5.
348. Bingham R.J., Dykeman E.C., Twarock R. RNA Virus Evolution via a Quasispecies-Based Model Reveals a Drug Target with a High Barrier to Resistance. Viruses. 2017;9(11).
349. Schlicksup C.J., Wang J.C., Francis S., Venkatakrishnan B., Turner W.W., VanNieuwenhze M., et al. Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. Elife. 2018;7.
350. Katen S.P., Chirapu S.R., Finn M.G., Zlotnick A. Trapping of hepatitis B virus capsid assembly intermediates by phenylpropenamide assembly accelerators. ACS Chem Biol. 2010;5(12):1125-36.