Analysis of the Foot-and-Mouth Disease Virus 2A-mediated Polyprotein Processing Event.

By
Lorraine Hughes B.Sc. (Hons).
Centre for Biomolecular Sciences
University of St. Andrews

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

DECLARATIONS

I, Lorraine Hughes, declare that this thesis has been composed by me, that it is a record of work carried out by me and that it has not been accepted in any previous application for any degree.

Signature:_ ___ Date: 13·08·03

I was admitted as a research student and candidate for the degree of Doctor of Philosophy in October 1998.

Signature:_ ___ Date: 13·08·03

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature of Supervisor:__ Date: 13/8/03

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the university library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published and a copy of the work may be made and supplied to any *bona fide* library or research worker.

Signature:_ ___ Date: 13·08·03

ACKNOWLEDGEMENTS

Firstly, my sincere and heartfelt thanks go to my parents,
Tom & Ena Hughes,
whose sacrifice and love have made all this possible.

Many thanks must also go to my supervisor,
Dr. Martin Ryan,
for his guidance, valuable advice and unstinting support.

Thanks to the Medical Research Council for funding my work.

Thanks to all my colleagues, who have donated materials,
shared their knowledge or just made my days brighter.

Thanks to Alex Houston for his friendship and
for sharing his expertise in DNA sequencing.

Thanks to my family and friends
for their belief and encouragement.

And, finally, special thanks to Craig Black, who has been an
invaluable source of strength.
Thank you for your friendship, love and understanding.

ABSTRACT

The 2A region of the aphthovirus foot-and-mouth disease virus (FMDV) polyprotein is only 18a.a. long. A 'primary' intramolecular polyprotein processing event mediated by 2A occurs at its own C-terminus. FMDV 2A activity was studied in artificial polyproteins in which sequences encoding reporter proteins flanked the 2A sequence such that a single, long, open reading frame was created. The self-processing properties of these artificial polyproteins were investigated and the protein products quantified. The processing products of the artificial polyprotein showed a molar excess of the protein N-terminal of 2A over the protein C-terminal of 2A, suggesting that the observed processing event is not proteolytic. Sequence analysis showed that the C-terminal portion of the much longer cardiovirus 2A region is similar to that of FMDV and there is a conserved DxExNPGP motif at the 2A/2B processing site. Database searches were performed probing for this motif and naturally occurring 2A-like sequences were identified. 2A-like sequences were found to be present in the polyproteins of other picornavirus, the polyproteins of several insect viruses, the NS34 protein of type C rotaviruses, repeated sequences of trypanosomes, the α-glucuronidase sequence of the eubacteria *Thermatoga maritima,* the carbamyl-phosphate synthetase I enzyme of the North American Bullfrog, *Rana catesbeiana*, the Mod(mdg4) 59.0 modifier protein of *Drosophila melanogaster* and the mouse mu opioid receptor variant F (MOR-1F). To investigate the mechanism of 2A-mediated processing, the 2A-like sequences were analysed for processing activity. All the 2A-like sequences of viral and trypanosomal origin demonstrated activity in artificial polyprotein systems indicating that this method of controlling protein biogenesis may not be confined to members of the *Picornaviridae.* The data presented here provide additional evidence that neither FMDV 2A nor the 2A-like sequences are autoproteolytic elements and have contributed to the development of a translational model for 2A-mediated polyprotein processing.

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

1.1. Viruses

Viruses have been known as distinct biological entities for more than a century. The discovery of new viruses continues and so the effort to characterise and classify these persists. Viruses infect all known groups of living organisms and show great biological diversity. In the endeavour to control these infectious agents and combat viral diseases, it is important to understand how viruses establish themselves in host populations. The intention of this introduction is to present some of the fundamental attributes of viruses with the cynosure being picornaviruses and in particular the translation of the aphthovirus, foot-and-mouth disease virus (FMDV), genome.

1.1.1. Discovery of Viruses

In 1892 the Russian scientist, Dimitri Ivanofsky first reported a pathogenic agent smaller than any known bacterium. He demonstrated that the causative agent of tobacco mosaic disease was retained in the sap of infected plants when passed through a filter, the Chamberland filter, that blocked the passage of all known bacteria (reviewed by Lustig & Levine, 1992). Unaware of Ivanofsky's work, the Dutch scientist, Martinus Beijerinck, independently made the same observation in 1898. He then extended the studies showing that when the filtered sap was diluted, it could regain infectivity after replication in living, growing plant tissue. These studies contributed to the development of a new concept: a filterable agent too small to be observed in the light microscope but able to cause disease by multiplying in living cells. In the same year, the German scientists, Friedrich Loeffler and Paul Frosch described and isolated the first filterable agent from

animals, the causative agent of foot-and-mouth disease, a widespread, devastating infection of cattle and other livestock. These infectious entities were mostly referred to as filterable agents and it was sometime later that the term virus was used.

Advances in the knowledge of viruses elucidated some definitive properties. A virus is a very small, infectious, obligate intracellular molecular parasite. The viral genome comprises either DNA or RNA. Within an appropriate host cell, the viral genome is replicated and directs the synthesis of other virion components, by cellular systems. Progeny virions are assembled from newly synthesised components within the host cell and are responsible for transmission of the viral genome to the next cell or organism. Subsequent disassembly of the transmitted virus particle begins the next infectious cycle.

1.1.2. Classification of Viruses

The earliest efforts to classify viruses were influenced by the types of studies being carried out at the time and so centered on perceived common pathogenic properties, transmission characteristics, host organisms and ecology. As more became known about the structure and compositions of virions, viruses were grouped on the basis of shared virion properties. As the number of discoveries of new viruses rapidly increased the need for a single, universal taxonomic scheme was highlighted. In the universal scheme developed by the International Committee on Nomenclature of Viruses (ICNV), viruses are classed by virion characteristics and grouped into orders, families, in some cases subfamilies, and genera (Strauss & Strauss, 1988).

1.1.3. Structure of Viruses

The minimum structural features for typical viruses consist of one molecule of nucleic acid integrated into a shell or coat made up of many identical protein molecules (Crick & Watson, 1956). The more complex viruses may contain several molecules of nucleic acid as well as several or many different proteins, internal bodies of definite shape, and complex envelopes with spikes that usually contain glycoproteins and lipids.

1.1.4. Genetic Strategies of Viruses

There are six basic genetic strategies utilised by viruses (*Figure 1*): Single-stranded DNA viruses; double-stranded DNA viruses; single-stranded, positive-sense RNA viruses; single-stranded, negative-sense RNA viruses; double-stranded RNA viruses and reverse transcribing viruses that have either RNA or DNA, of the positive polarity, in the virion (Baltimore, 1971). The type of genomic nucleic acid governs the molecules transcribed from the genomic nucleic acid as the virus directs the synthesis of messenger RNAs and new nucleic acid.

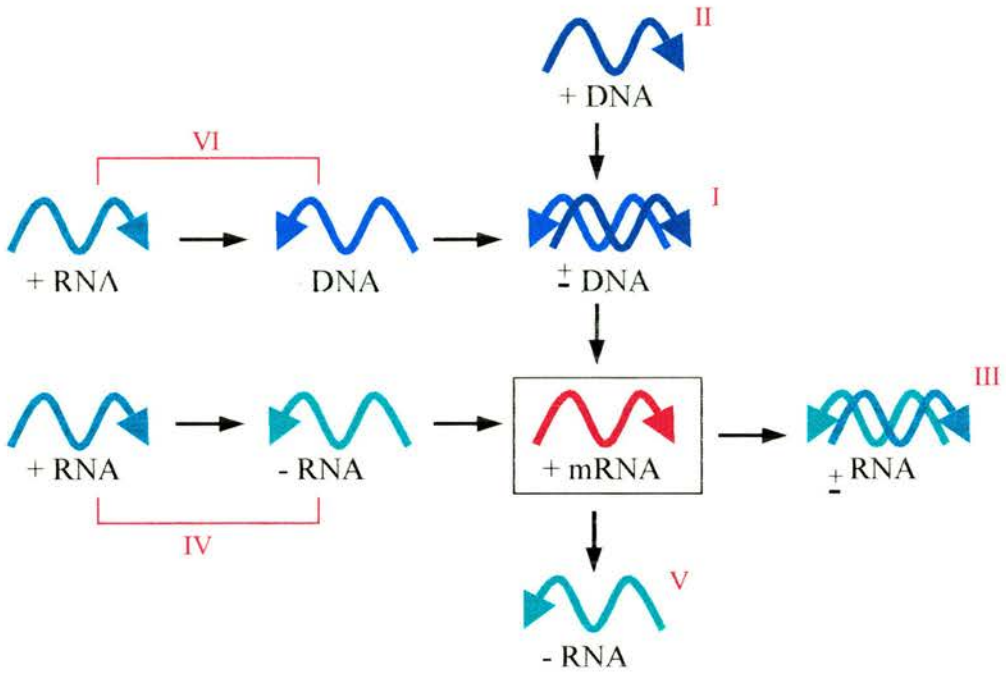


Figure 1: ***The Baltimore Classification of Viruses,*** *based on the need for all viruses to produce mRNA that can be translated by cellular ribosomes. The unique pathways from various viral genomes to mRNA define the specific virus classes on the basis of the nature and polarity of their genome.*

1.1.5. DNA Viruses

With the exception of the circular single-stranded DNA chain of circoviruses and the linear single-stranded DNA chain of the parvoviruses, the members of all the DNA virus families contain double-stranded genomes. Most of the DNA genomes are not simple linear molecules: papovavirus genomes are covalently closed, double-stranded circles; the ends of the double-stranded poxvirus genome are covalently joined and several viruses, including adenovirus, have protein molecules covalently attached to the 5' ends of their DNA strands (Murphy, 1996), (*Figure 2*).

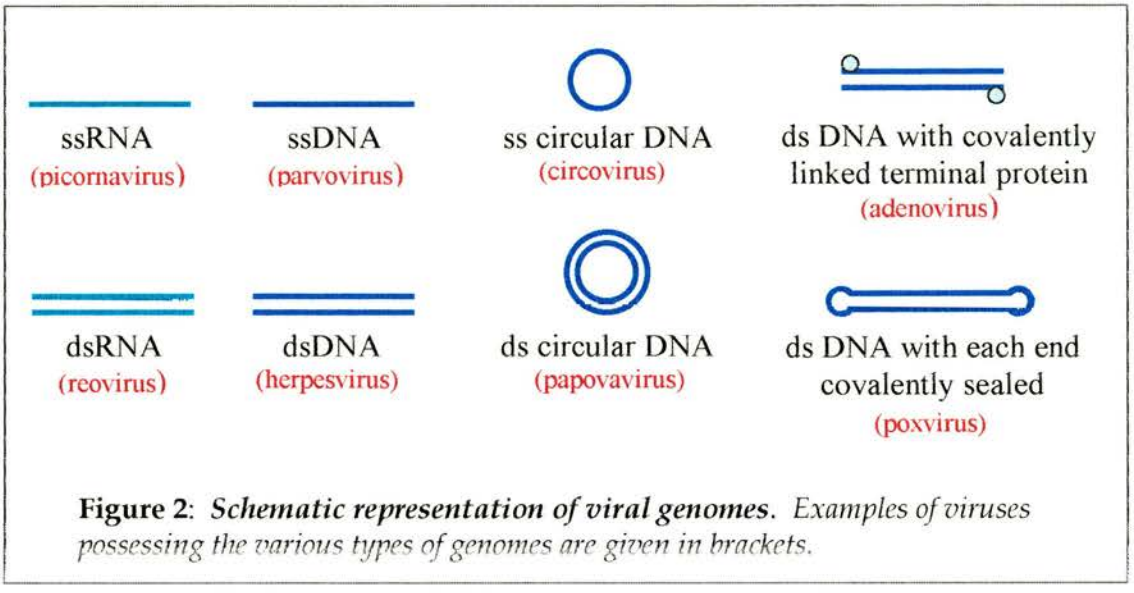


Figure 2: ***Schematic representation of viral genomes.*** *Examples of viruses possessing the various types of genomes are given in brackets.*

Each type of viral genome requires unique enzymatic tricks for its replication and in addition to the viral coat proteins must encode one or more of the enzymes needed to replicate the viral nucleic acid. The amount of information that a virus brings into a cell to ensure its own selective replication varies greatly. The DNA of the relatively large T4 bacteriophage contains about 300 genes, at least 30 of which ensure the rapid replication of the T4 chromosome in its *E. coli* host cell. T4 replication has the unusual feature that 5-hydroxymethyl-C is incorporated in place of C in its DNA. The unusual base composition of the T4 DNA makes it readily distinguishable from host DNA and selectively protect it from nucleases encoded in the T4 genome that degrade the *E. coli* chromosomes. Other T4 encoded proteins alter host cell RNA polymerase molecules so that they are unable to transcribe *E. coli* DNA and instead transcribe different sets of bacteriophage genes at different stages of infection, and in accordance with the needs of the bacteriophage.

Smaller DNA viruses, such as the monkey virus SV40 and the tiny M13 bacteriophage, carry much less genetic information and so rely heavily on host cell enzymes to carry out their DNA synthesis. Although these smaller DNA viruses tend to commandeer most of the host cell DNA replication proteins, most DNA viruses do however encode proteins that selectively initiate the synthesis of their own DNA. This selectivity is usually implemented by recognition of a particular nucleotide sequence in the virus that serves as a replication origin. This is important because a virus must override the cellular control signals that would otherwise cause the viral DNA to replicate in pace with the host cell DNA, doubling only once in each cycle.

1.1.6. RNA Viruses

Most RNA viruses have single-stranded genomes. Members of the picornavirus and togavirus families are morphologically diverse but both contain positive-sense, non-segmented genomes. The retroviruses also contain positive sense RNA, but uniquely the virion includes two copies of the genomic RNA. Rhabdoviridae, Paramyxoviridae and Filoviridae are families of the order Mononegavirales, all which have single-strand negative sense RNA genomes. The virions of orthomyxoviruses, arenaviruses and bunyaviruses have segmented genomes, they contain more than one RNA molecule (Murphy, 1996). RNA viruses have particularly special requirements for replication, as they must copy RNA molecules to reproduce their genomes. They must polymerize nucleoside triphosphates on an RNA template. Cells do not normally have enzymes to carry out this reaction, so even the smallest RNA virus must encode their own RNA-dependent polymerase enzymes for replication.

1.2. Diversity of Viral Genomes and Protein Expression

Whatever the composition and complexity of a virus genome, all must conform to the same conditions. Since viruses are obligate intracellular parasites only able to replicate inside the appropriate host cell, the genome must contain information encoded in a form that can be recognized and decoded by the particular cell parasitized. The RNA viruses employ diverse strategies to produce all their functional proteins from their relatively small genomes.

1.2.1. Polyprotein Processing

One strategy of allowing the synthesis of multiple polypeptides from a single RNA genome is to synthesize a polyprotein precursor that is proteolytically processed to form functional viral proteins. Picornaviruses, the largest family of positive sense RNA viruses, make the most dramatic use of protein processing with almost the entire positive strand RNA genome translated into a single large polyprotein (reviewed in Palmenberg, 1990). This translation strategy has been widely studied in poliovirus.

1.2.1.1. Poliovirus – a positive sense RNA virus coding for a single genome length mRNA

The single positive strand RNA genome of poliovirus is about 7400 nucleotides long. The positive strand RNA genome of poliovirus that enters the cell first serves as mRNA for protein synthesis. This is essential since replication of the viral RNA requires virus-encoded enzymes and none are present in polio virions. Like cellular mRNAs, the poliovirus genome has a poly (A) sequence at its 3' end but unlike cellular mRNA and most other viral mRNAs, it is not capped at the 5' end. Instead, poliovirus has a tiny 22 amino acid long protein called VPg (reviewed in Reuckert, 1996).

1.2.1.2. Poliovirus Polyprotein

Poliovirus is representative of viruses whose positive sense RNA genome functions as a single genome sized mRNA. When serving as mRNA, poliovirus RNA is translated into a single polyprotein from which all the viral proteins are derived by cleavage (*Figure 3*). Thus a large open reading frame occupies the vast majority of the genome. The protein-coding region in the poliovirus genome is flanked by 5' and 3' untranslated regions (reviewed in Reuckert, 1996).

1.2.1.3. Poliovirus Polyprotein Cleavages

The poliovirus polyprotein precursor does not normally exist in infected cells because proteolytic cleavage occurs at two sites before its synthesis is complete (reviewed in Palmenberg, 1990). The first part of the protein to be made is the amino terminal portion, which contains the virion proteins. The carboxyl terminal region is cleaved to yield the proteins essential for replication. During synthesis, the polypeptide composed of the structural proteins (P1) is cleaved away from the remainder of the polyprotein still nascent on the ribosome. In poliovirus this cleavage occurs between 1D and 2A and is catalyzed by the 2A proteinase. Subsequently the different functional proteins are released by cleavage events catalyzed by the 3C proteinase and its precursor $3CD^{pro}$ (*Figure 3*) (reviewed in Ryan & Flint, 1997).

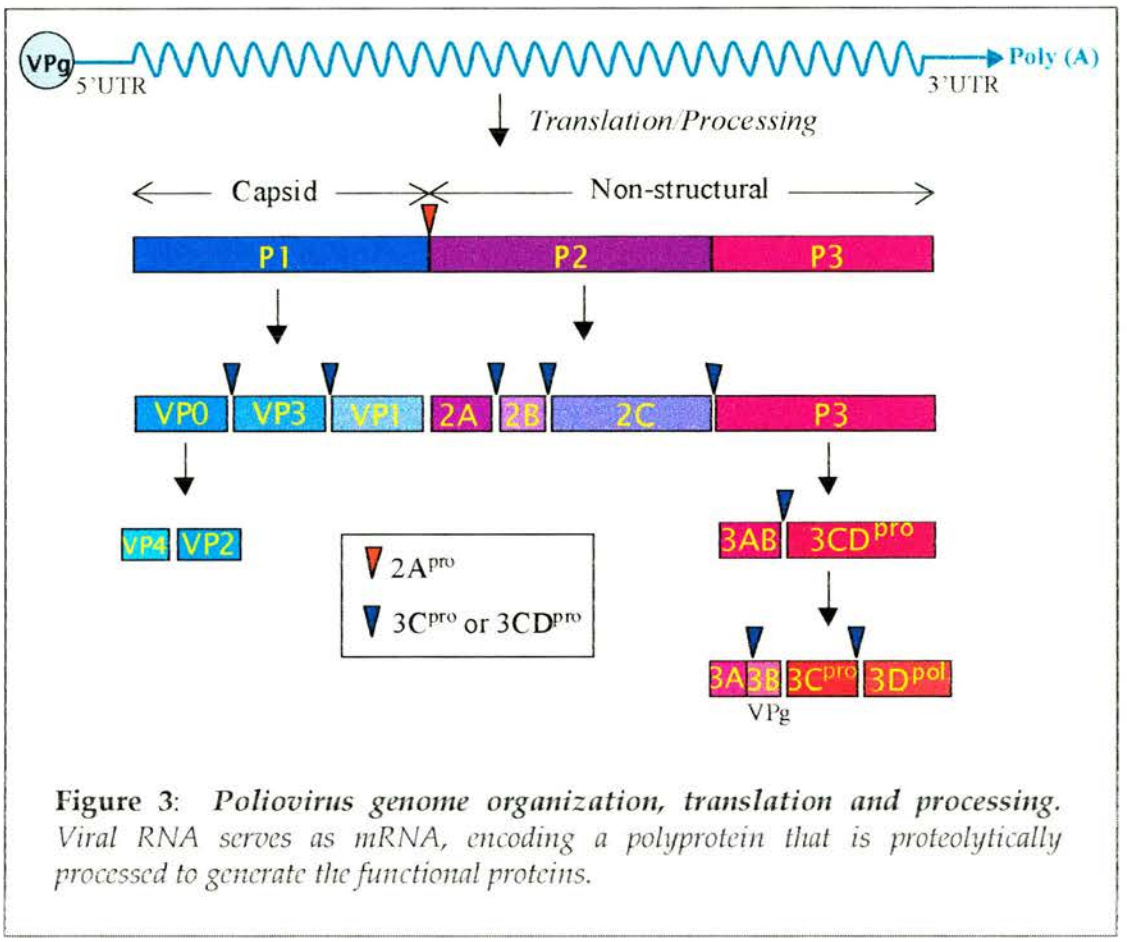


Figure 3: ***Poliovirus genome organization, translation and processing.*** *Viral RNA serves as mRNA, encoding a polyprotein that is proteolytically processed to generate the functional proteins.*

1.2.1.4. Poliovirus Gene Expression

The fact that all the poliovirus gene products are obtained by processing a single precursor implies that all the products will be synthesized in equimolar amounts. At most, one RNA-dependent RNA polymerase is required to make a new genome, whereas 60 molecules of each capsid protein are required to encapsidate it. This means that each infected cell synthesizes several unnecessary molecules of RNA-dependent RNA polymerase in order to generate enough capsid proteins for the progeny virions.

1.2.2. Subgenomic mRNA

Other positive strand RNA viruses have evolved a different mechanism of gene expression that allows structural and non-structural proteins (generally needed in greater and lesser quantities, respectively) to be produced separately. This expression strategy utilizes subgenomic mRNAs.

1.2.2.1 Togaviruses – a family of positive sense RNA viruses that code for subgenomic mRNAs

Togaviruses are among the smallest enveloped animal viruses. Their genome consists of a single strand of positive sense RNA encapsidated in an icosahedral protein shell composed of a single species of protein and enveloped by a lipid bilayer derived from the host cell plasma membrane (reviewed in Schlesinger & Schlesinger, 1996). These viruses are larger and more complex than picornaviruses and have a more elaborate strategy of gene expression, employing a combination of subgenomic mRNA and polyprotein processing to generate the functional viral proteins from a single strand of viral RNA.

1.2.2.2. Togavirus Subgenomic mRNA

Togavirus genomes are dicistronic, the cistron nearest the 5' end encodes the replication or non-structural proteins and the 3' cistron encodes the structural proteins. Early in infection the viral RNA serves as mRNA for the non-structural proteins only since the translation initiation codon for the encapsidation proteins is masked from the ribosome. Later in infection a subgenomic mRNA is synthesized using an internal transcription initiation site on the negative strand RNA (reviewed in Schlesinger & Schlesinger,

1996). The subgenomic mRNA lacks the entire replication protein cistron and the initiation codon for translating capsid is exposed (*Figure 4*).

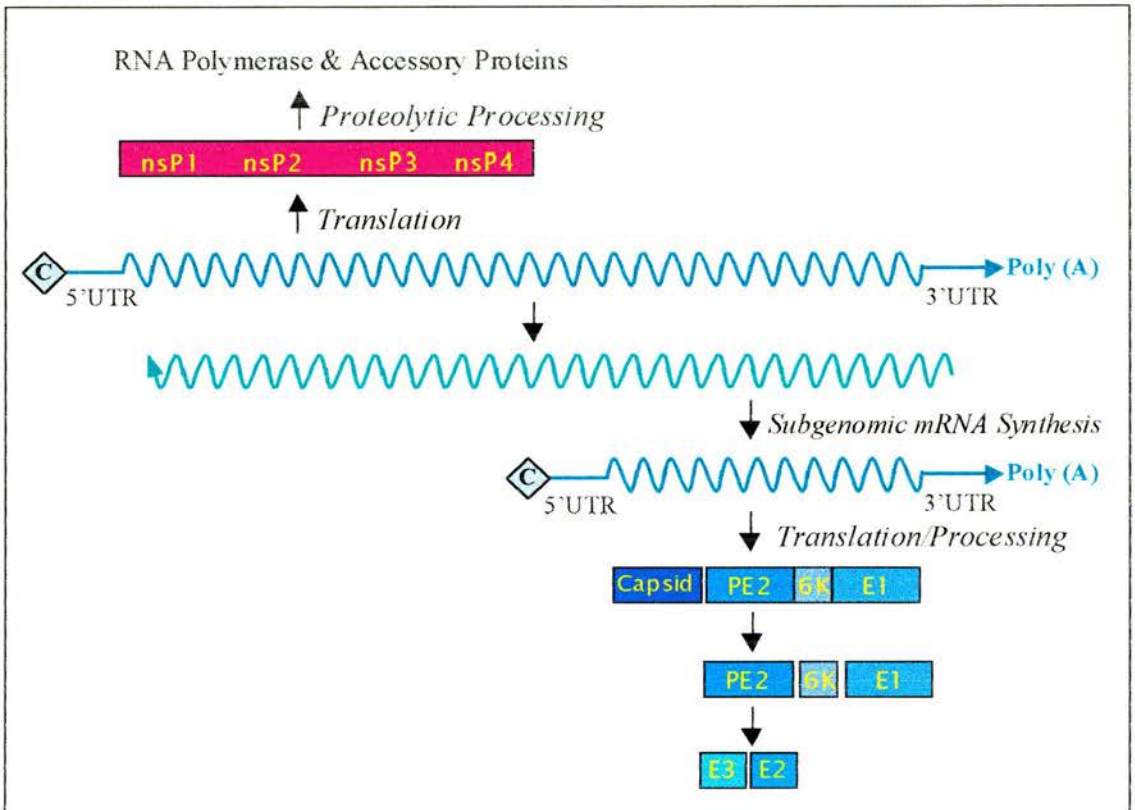


Figure 4: ***Togavirus genomic organization, translation and processing.*** *Non-structural proteins are translated directly from the viral RNA, while the structural proteins are encoded by a subgenomic mRNA copied from the negative strand RNA.*

1.2.2.3. Togavirus Polyprotein Processing

The polyproteins derived from either the genomic or subgenomic RNAs are proteolytically processed. The non-structural protein precursor is processed by a viral proteinase contained within the nsP2 region of the non-structural polyprotein (Hardy & Strauss, 1989, and Ding & Schlesinger, 1989). The structural polyprotein is processed

either by a combination of virus and host cell proteolytic activities in alphavirus or by host proteinases alone in rubivirus (ten Dam *et al.*, 1999).

1.2.2.4. Togavirus Gene Expression

The incorporation of the strategy of producing subgenomic mRNA enables the togaviruses to regulate the biogenesis of encapsidation and replicative proteins in both a temporal and quantitative manner.

1.2.3. Nested Subgenomic mRNA

Viruses with larger, single-stranded, continuous RNA genomes, such as the members of the order Nidovirales, further separate the generation of their gene products by way of a nested set of subgenomic mRNAs.

1.2.3.1. Nidovirales – an order of viruses that produce nested sets of subgenomic mRNAs

The order Nidovirales comprises of the families Arteriviridae (genus *Arterivirus*) and Coronaviridae (genus *Coronavirus* and *Torovirus*). The genomes of the arteriviruses and coronaviruses are 5' capped and polyadenylated at the 3' end (de Vries *et al.*, 1997). The genome is arranged with the non-structural proteins encoded in the first three-quarters and the 3' end encoding mostly structural proteins (*Figure 5*). The non-structural proteins are encoded by two overlapping reading frames, ORF1a and ORF1b. The downstream ORF is expressed by a frameshift event, described below. It is the proteins encoded downstrean of ORF1b that are produced from the 3'-coterminal nested set of subgenomic mRNAs. The members of the Nidovirale order can express between four (toroviruses) and nine (arteriviruses and coronaviruses) genes from subgenomic mRNAs (de Vries et al., 1997 and Snijder & Meulenberg, 1998). Generally, translation is

restricted to the unique 5' sequences not present in the next smaller RNA of the set, however there may be some exceptions (de Vries *et al.*, 1997).

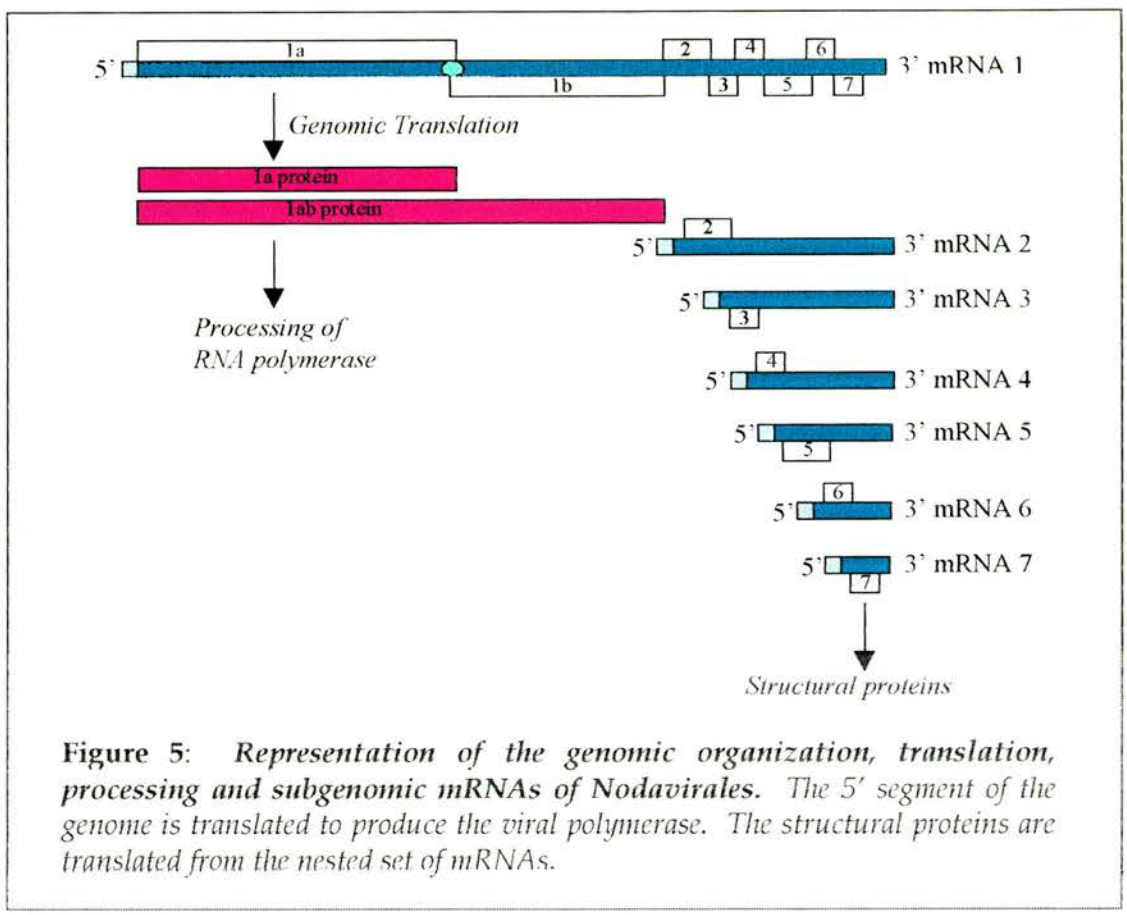


Figure 5: ***Representation of the genomic organization, translation, processing and subgenomic mRNAs of Nodavirales.*** *The 5' segment of the genome is translated to produce the viral polymerase. The structural proteins are translated from the nested set of mRNAs.*

1.2.3.2. Nidovirales - Generation of Subgenomic mRNAs

The subgenomic mRNA transcription mechanisms of arteriviruses and coronaviruses are similar (Snijder & Meulenberg, 1998). The subgenomic mRNAs are composed of a leader and a body, which are synthesized from non-contiguous sequences in the viral genome. The leader is encoded at the 5' end of the viral genome and the bodies at the 3' end. The leader and the body are separated by a conserved junction sequence encoded both in the 3' end of the leader and the 5' end of the body. It is proposed that the subgenomic mRNAs are copied from sub-genome length (-) strand RNAs which are

produced by discontinuous RNA synthesis (*Figure 6*). It is believed that the viral polymerase copies the (+) strand genome and jumps from the junction sequence of the body to the junction sequence of the leader, synthesizing the leader sequence on the 3' end of the (-) strand RNA (Snijder & Meulenberg, 1998). The (-) strand sub-genomic RNAs then serve as templates for mRNA synthesis. The nested set of mRNAs would be produced by translocation signaled by the individual junction sequences between mRNA coding regions.

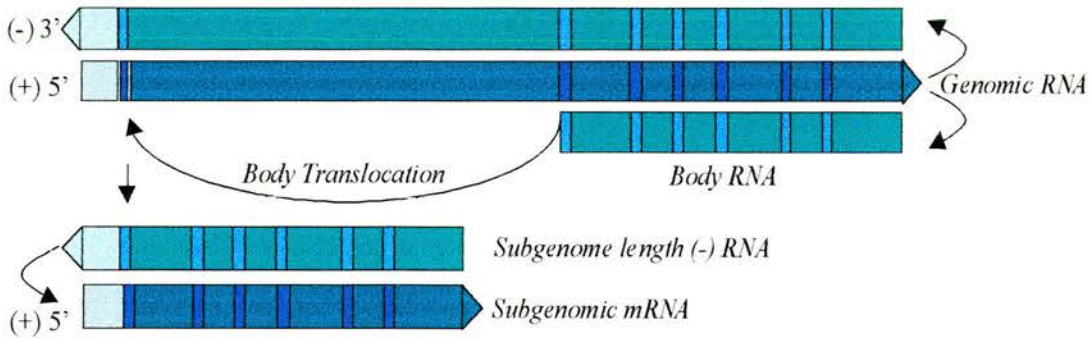


Figure 6: ***Illustration of subgenomic mRNAs formation in arteriviruses and coronaviruses.*** *The viral polymerase synthesizes a negative strand on the genomic RNA. The junction sequence (shown in blue) on the body RNA is complementary to the junction sequence 3' of the leader sequence (shown in purple) on the genomic RNA. The body translocates to the 5' end of the genomic RNA to complete synthesis of the leader sequence. The resulting subgenome length negative strand RNA is then copied to form the subgenomic mRNA. The nested set of mRNAs would be generated by the individual junction sequences, between mRNA coding regions, signaling translocation.*

1.3. Diversity of Viral Translation Strategies

In addition to the genome arrangements described previously, viruses have evolved a variety of unusual translation strategies to expand the coding capacity of their relatively small genomes, allowing the synthesis of multiple polypeptides from a single RNA genome.

1.3.1. Ribosomal Frameshifting

Ribosomal frameshifting occurs when the ribosome moves into a different reading frame in response to signals in the mRNA and continues translation in the new reading frame. This occurrence was first discovered in Rous sarcoma virus infected cells and has since been described for many other viruses, including the eukaryotic (+) strand RNA virus, infectious bronchitis virus (IBV). IBV is an avian coronavirus and is the type species of the *Coronaviridae*. As described previously (section 1.2.3), these viruses have large, single-stranded, continuous RNA genomes of positive polarity. The ribosomal frameshift site is situated within the domain encoding the non-structural proteins at a site where the two large open reading frames (ORF), 1a and 1b, overlap. The open reading frames overlap by 42 nucleotides, with ORF 1b in the –1 frame with respect to OFR 1a. A highly efficient (25%-30%) –1 ribosomal frameshift suppresses the termination codon of OFR 1a resulting in the production of a 1a-1b fusion protein (Brierley *et al.*, 1987). The frameshift signal consists of two important elements, a "slippery" sequence (UUUAAAC) at which the ribosomes actually change frame, and an RNA pseudoknot structure located just downstream of the "slippery" sequence (Brierley et al., 1989). The pseudoknot must be within 5 to 7 nucleotides of the slip site for efficient frameshifting in IBV (Brierley *et al.*, 1992). It is believed that the pseudoknot plays a role in slowing or

pausing the ribosome as it translates the "slippery" sequence, allowing realignment of the decoding tRNAs on the mRNA in a new frame (Somogyi *et al.*, 1993).

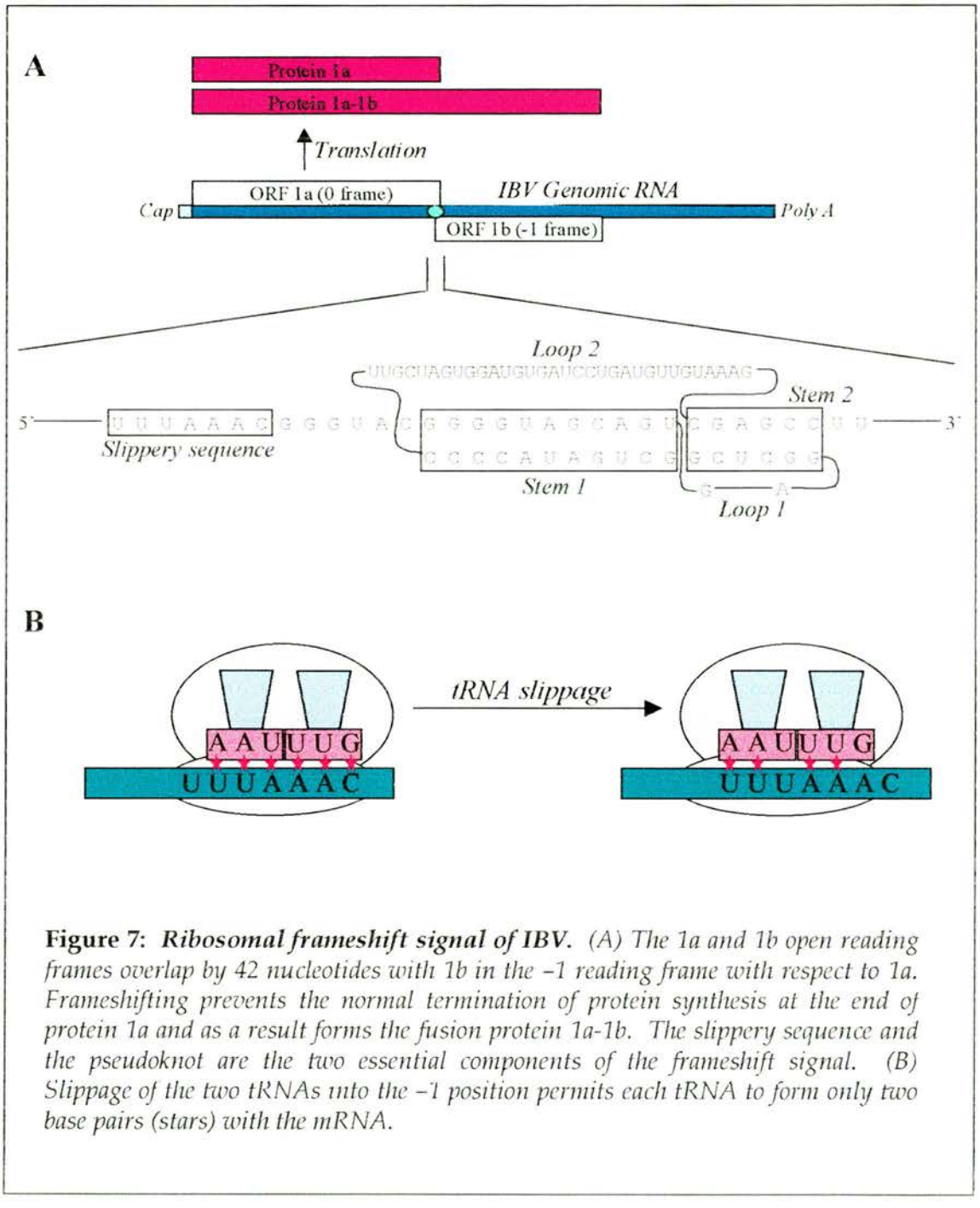


Figure 7: ***Ribosomal frameshift signal of IBV.*** *(A) The 1a and 1b open reading frames overlap by 42 nucleotides with 1b in the -1 reading frame with respect to 1a. Frameshifting prevents the normal termination of protein synthesis at the end of protein 1a and as a result forms the fusion protein 1a-1b. The slippery sequence and the pseudoknot are the two essential components of the frameshift signal. (B) Slippage of the two tRNAs into the -1 position permits each tRNA to form only two base pairs (stars) with the mRNA.*

1.3.2. Leaky Scanning

Although the majority of eukaryotic and viral mRNAs are monocistronic, with the initiation of protein synthesis usually occurring at the first AUG codon nearest the 5' end of the mRNA, some viral mRNAs that encode two (or more) proteins in overlapping reading frames have been identified. In these mRNAs initiation of protein synthesis also occurs at initiation codons downstream of the first AUG codon.

Eukaryotic mRNAs are generally thought to initiate translation by the mechanism described by the scanning model. The ribosomes and initiation factors bind to the 5' end of the mRNA and migrate in a linear manner on the mRNA until they encounter the first AUG initiation codon. In the case of some viral mRNAs it is believed that initiation occurs at the downstream AUG, in addition to the 5' AUG, because some ribosome pre-initiation complexes bypass the 5'-proximal AUG by leaky scanning and initiate protein synthesis at the second AUG codon. It is proposed that leaky scanning occurs when the first AUG codon is surrounded by nucleotide sequences which are suboptimal when compared with the consensus sequence for the initiation of protein synthesis (5'-CC(**A/G**)CCAUG**G**-3') (reviewed in Kozak, 1986) . The strongest determinants for efficient initiation are a purine residue at the –3 position and a guanine at the +4 position.

Sendai virus, which belongs to the parainfluenza virus family, is a model for leaky scanning. The mRNA that encodes the structural, P protein and non-structural, C proteins, has overlapping reading frames (Giorgi *et al.,* 1983). In fact, a total eight primary polypeptide products (C', P, C, Y1, Y2, V, W and X) are expressed from the mRNA of the P/C gene (figure 8), utilizing several translational adaptions (reviewed in

Curran *et al.*, 1998). Initiation of the C' protein occurs at an ACG codon, this is the 5' proximal start site and is otherwise in a favourable context for initiation (**GCC**ACG**G**), albeit weak initiation due to the unusual start codon. The P and C proteins are initiated from start sites downstream of the C' initiation codon. These three start sites appear to be organized in such a way as to allow access to the downstream initiation codons by leaky scanning. The downstream start site of the P protein has an AUG codon but is in a relatively poor context (**CGC**AUG**G**). The start site for the C protein, yet further downstream, is an AUG codon in a better context (**A**AG**AUGC**). With initiation at the ACG codon being inefficient, some ribosomes can continue scanning and initiate at the P protein AUG; since this is in a poor context some ribosomes can pass on to the C protein AUG which is in a better initiation context. Generally, when successive initiation codons are used in leaky scanning, they are increasingly efficient as start sites.

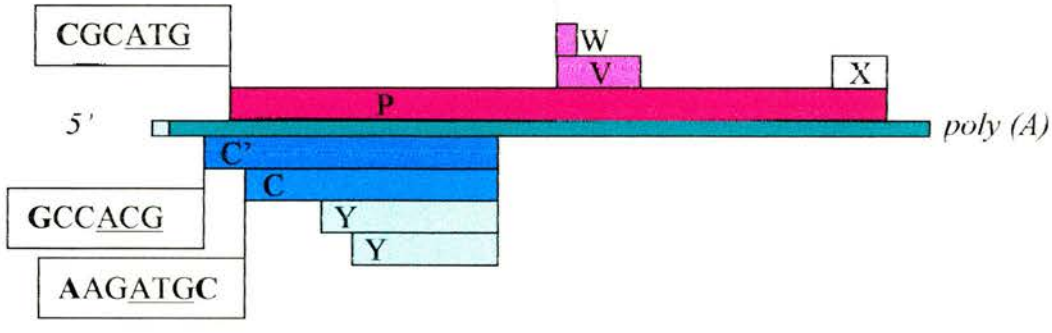


Figure 8: ***The open reading frame organization of the Sendai virus P/C gene.*** *The ORFs are shown as boxes. The initiation sites of the C', P and C proteins are increasingly efficient, permitting leaky scanning. Addition of one or two G residues at the mRNA editing site, during mRNA synthesis, produces the V or W carboxy terminus of the P protein.*

The influenza B virus RNA segment 6 is functionally bicistronic and encodes the NB and NA glycoproteins in different overlapping reading frames. The NB protein is

synthesized from the 5'-proximal AUG initiation codon and the NA protein synthesis is initiated from a second AUG codon, four nucleotides downstream (*Figure 9*). With the NB AUG initiation codon in a good context for translation initiation, it would be expected that this is where the majority of initiation events should occur. However, the NB and NA proteins are found to accumulate in roughly equimolar amounts in cells infected with influenza B virus. Consequently, it is proposed that the influenza B virus NB/NA mRNA is a unique model system in which the ribosomes, once in the region of the two initiation codons, essentially randomly choose which AUG is used. It is suggested that surrounding mRNA sequences, and the interaction of cellular factors with these sequences, may influence the selection of the initiation codon (Williams & Lamb, 1989).

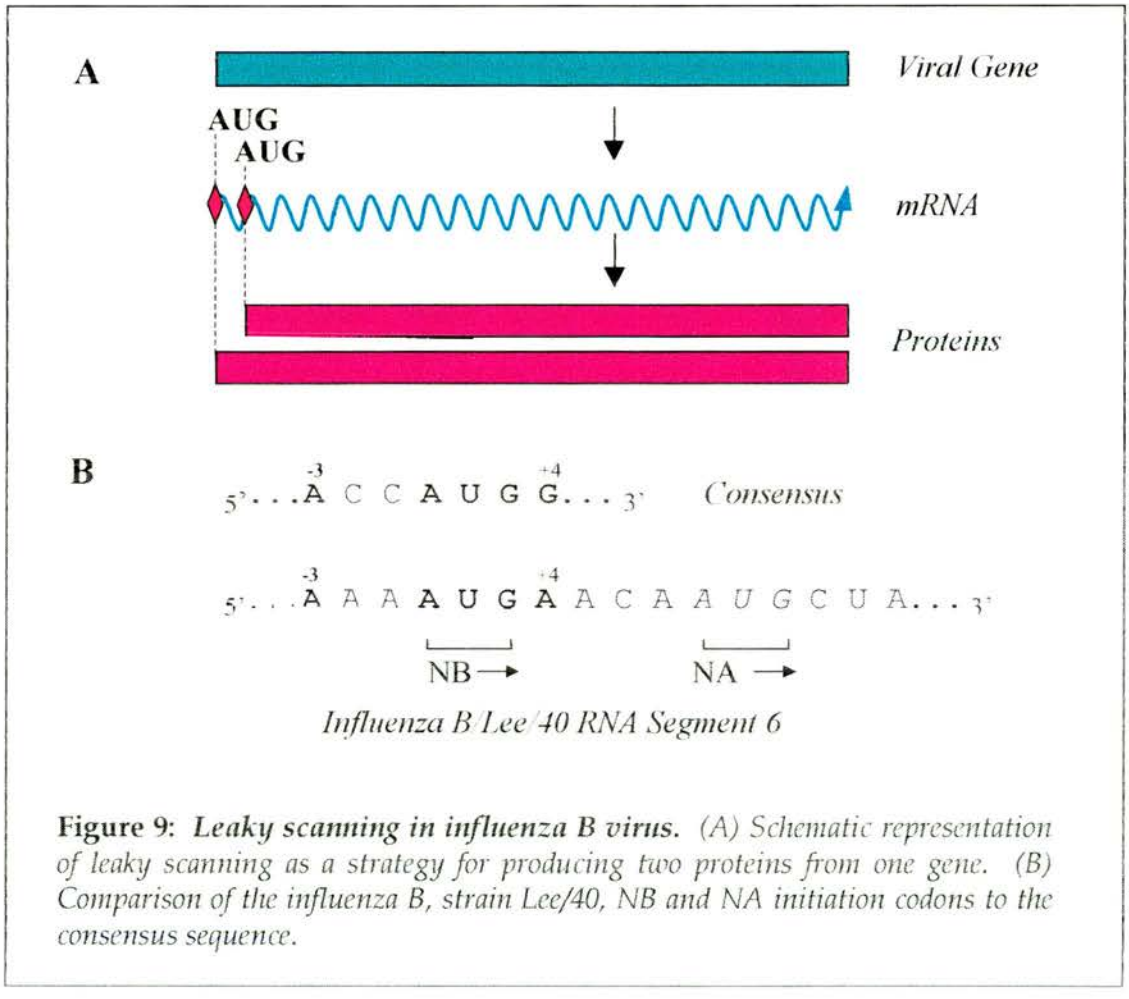


Figure 9: ***Leaky scanning in influenza B virus.*** *(A) Schematic representation of leaky scanning as a strategy for producing two proteins from one gene. (B) Comparison of the influenza B, strain Lee/40, NB and NA initiation codons to the consensus sequence.*

1.3.3. Reinitiation

Reinitiation is another strategy for producing two proteins from one mRNA. This mode of expanding the genomic coding capacity relies on termination of an upstream open reading frame with reinitiation occurring at the start codon of the downstream open reading frame. The RNA segment 7 of influenza B virus encodes two proteins, M1 and BM2, by a bicistronic mRNA. The M1 protein is translated from the 5'-proximal AUG codon while the BM2 protein AUG initiation codon overlaps with the termination codon for the M1 protein (*Figure 10*). The individual proteins are generated by a coupled translational termination-reinitiation event that occurs at the overlapping stop-start pentanucleotide sequence (UAAUG). The synthesis of the BM2 protein is dependent on the initiation and termination of the M1 protein (Horvath *et al.*, 1990).

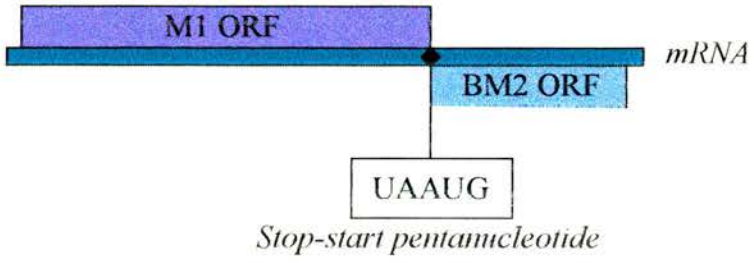


Figure 10: Bicistronic mRNA of the influenza B virus segment 7. *The initiation codon (AUG) for the BM2 protein overlaps the termination codon (UAA) for the M1 protein. A coupled translational termination-reinitiation event occurs at the stop-start pentanucleotide, yielding two discrete proteins.*

A somewhat different ribosomal termination-dependent reinitiation mechanism has been described for the expression of the protein products of the two open reading frames of the human respiratory syncytial virus M2 gene. Respiratory syncytial (RS) virus is the major causative agent for hospitalization of children <1 year of age with repiratory infection worldwide. This virus is a paramyxovirus with a negative sense RNA genome. Replication and transcription of the viral genome occur in the cytoplasm of infected cells using virus-encoded proteins. The M2 protein has a role in enhancing transcription and is encoded by the first major ORF (ORF1) of the M2 mRNA. The second major ORF (ORF2) is located towards the 3' end of the mRNA and overlaps ORF1. The protein product of ORF2, M2-2, may be involved in the control of the switch between virus RNA replication and transcription (reviewed in Ahmadian *et al.*, 2000). In human RS virus strain A2, the M2 gene is 956 nucleotides in length with the first major ORF starting at nucleotide 10 and terminating at nucleotide 592, encoding a polypeptide of 194 amino acids. Translation of the second ORF initiates, *in vitro*, at one of three AUG codons located at nucleotides 563, 569 and *581 (Figure 11)*. *In vivo* there is a preference for the AUGs located at nucleotides 563 and 569. Termination occurs at residue 832, the largest possible polyprotein being 90 amino acids in length. Clearly, the initiation codons for OFR2 are situated upsteam of the termination codon of ORF1 making the reinitiation event required for the synthesis of RS virus M2-2 distinct from the reinitation event that occurs in the expression of influenza B virus BM2. In the expression of the RS virus M2 proteins, the initiation of translation of ORF2 is coupled to the termination of translation of ORF1 (Ahmadian *et al.*, 2000). The distance between the stop codon of ORF1 and the 5'-proximal start codon of ORF2 is critical and has to be sufficiently short to allow ribosomes to reinitiate before they detach completely from the mRNA. The

mechanism that permits the reverse translocation of ribosomes, from the termination site of ORF1 to the initiation site of ORF2, is not yet known.

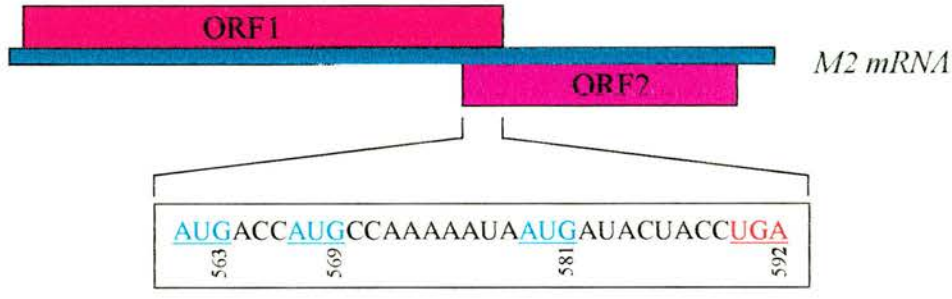


Figure 11: Illustration of the two overlapping ORFs of the human respiratory syncytial virus M2 mRNA. *The sequence of the overlapping region is shown with the ORF2 start codons in green and the ORF1 stop codon in red. The nucleotide positions of each of these in the RS virus M2 mRNA are indicated.*

1.3.4. Suppression of Termination

Protein translation usually stops at one of the three termination codons: UAG (amber), UGA (opal) and UAA (ochre). However, termination can sometimes be leaky and translation continues until a second termination codon is reached. As a result, a second protein with an extended carboxy terminus, can be produced. This phenomenon is called translational read-through or suppression of termination.

Sindbis virus and several other alphaviruses require translational read-through of a UGA codon for the synthesis of the viral RNA-dependent RNA polymerase, nsP4. The leaky stop codon interrupts the nonstructural coding region of the sindbis genome between nsP3 and nsP4 (Strauss *et al.*, 1983) (*Figure 12*). The read-through efficiency is about

10% (Li & Rice, 1993) and is required for viral replication. An important determinant of translational read-through efficiency can be the nucleotide sequence 3' to the termination codon. In sindbis virus, efficient suppression of termination requires only a cytidine residue 3' of the termination codon (Li & Rice, 1993).

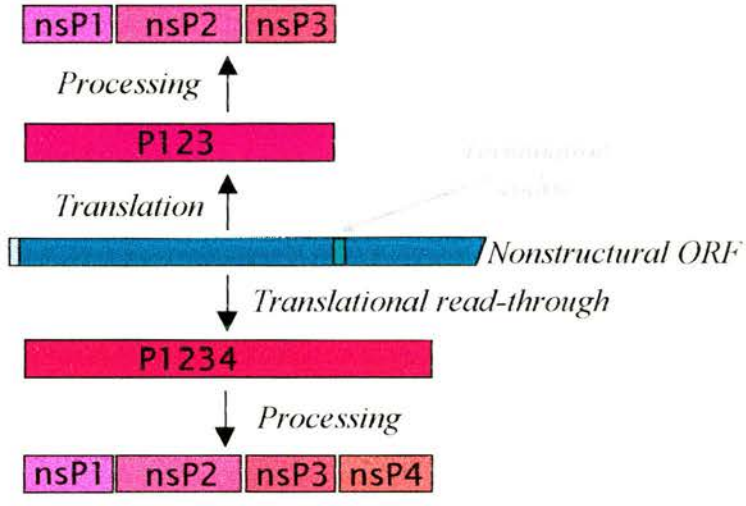


Figure 12: Suppression of termination in sindbis virus. *Suppression of the termination codon, situated between the nsP3 and nsP4 coding regions, permits translational read-through and yields the viral RNA-dependent RNA polymerase.*

Maloney murine leukemia virus is another virus that makes use of termination suppression, however, the signal is more complex than that required by sindbis virus. In murine leukemia virus and other mammalian type C retroviruses, the gag and pol coding regions are separated by an in-frame amber (UAG) termination codon. Suppression of the termination codon generates a Gag-Pol fusion protein (Yoshinaka *et al.*, 1985), that is processed to yield the functional Pol proteins (protease, reverse transcriptase and integrase). The termination codon is translated as a glutamine (Yoshinaka *et al.*, 1985).

The signal required for read-through of the *gag* termination codon consists of two distinct elements; a highly conserved eight nucleotide purine-rich sequence immediately downstream of the UAG codon and an adjacent downstream pseudoknot structure (Feng *et al.*, 1992). It is suggested that this pseudoknot may play a role analogous to those found at frameshift sites, pausing the ribosome to allow the suppressor tRNA to successfully compete with release factor at the termination site (Feng *et al.*, 1992).

The fundamental characteristic of viruses is their absolute dependence on a living host organism for reproduction; they are obligate parasites. The synthesis of viral proteins by the host protein-synthesizing machinery is the key event in viral replication. The examination of viral translation strategies has highlighted the potential to manipulate the translation machinery to produce more than one protein from a single mRNA. Ribosomal frameshifting, leaky scanning, reinitiation events and suppression of termination are alternatives to the conventional translation process which allow viruses to maximize the number of proteins encoded by their size-limited genomes. A novel mode of translational control is thought to permit the synthesis of two polyproteins from a single ORF on the genome of the picornavirus, foot-and-mouth disease virus. Analyses leading to the development of this model are presented below and the work detailed in this thesis suggests that this type of translational control is not restricted to FMDV.

1.4. Picornaviruses

The *Picornaviridae* is a family of small icosahedral viruses that contain a positive sense, single-strand RNA genome. The main genera (and their type species) are: *Enterovirus* (poliovirus), *Rhinovirus* (human rhinovirus A), *Cardiovirus* (encephalomyocarditis virus), *Aphthovirus* (foot-and-mouth disease virus), *Hepatovirus* (hepatitis A virus) and *Parechovirus* (human parechovirus). The recent inclusions in this family are: *Erbovirus* (equine rhinitis B virus), *Kobuvirus* (aichi virus) and *Teschovirus* (porcine techovirus). Members of the picornavirus family represent a diverse variety of pathogenic agents that cause some of the most medically and agriculturally significant diseases. Rhinovirus for example is an etiologic agent of the common cold in adults and children. This acute infectious disease of the upper respiratory tract is a major cause of absenteeism in schools and industry. Poliovirus is another familiar member of the picornavirus family. It belongs to the enterovirus class and produces an infectious disease of the central nervous system, which at worst can lead to permanent paralysis. Animal populations too, are susceptible to picornavirus infection. The aphthovirus, foot-and-mouth disease virus (FMDV) is one of the most virulent of pathogens ever described. It infects a variety of commercial livestock causing devastating illness.

1.4.1. Foot-and-Mouth Disease Virus Genome

Members of the picornavirus family have a similar fundamental genome arrangement to that described previously for poliovirus (*Figure 3*). However, there are some subtle and important variations between the genera. Foot-and-mouth disease virus will be described here and some of these variations will be highlighted.

The positive sense (messenger active), single-strand RNA genome of all picornaviruses, are polyadenylated at the 3' terminus and carry a small viral protein, VPg, covalently attached to the 5' end. The poly(A) tract is shortest in cardioviruses and longest in aphthoviruses. The length of the VPg protein, also termed 3B protein, varies only slightly in different viruses. Foot-and-mouth disease virus is unique in that it encodes three VPg genes (*Figure 13*) whereas the other picornaviruses encode only one. The VPg protein appears to play an important role in initiation of picornavirus RNA synthesis. Untranslated regions (UTRs), flank the protein coding region. The UTR sequences tend to be strongly conserved, with the exception of a poly(C) tract contained within the 5' UTR of aphthoviruses (*Figure 13*) and encephalomyocarditis viruses. Signals initiating translation are present at the 5' end of the plus strand and signals for initiation of RNA synthesis are present at the 3' end of the minus strand. Internal ribosome entry sites (IRESs), contained within the 5' UTRs, enable the eukaryotic ribosomes to bind directly to the internal site without first scanning from the 5' terminus.

The protein-coding region composes a single long open reading frame (ORF) which is organized into three domains. The precursor of the structural coat proteins (P1) is encoded by the 5' region of the ORF, with the precursors of the replication machinery (P2-P3) encoded towards the 3' end (*Figure 13*).

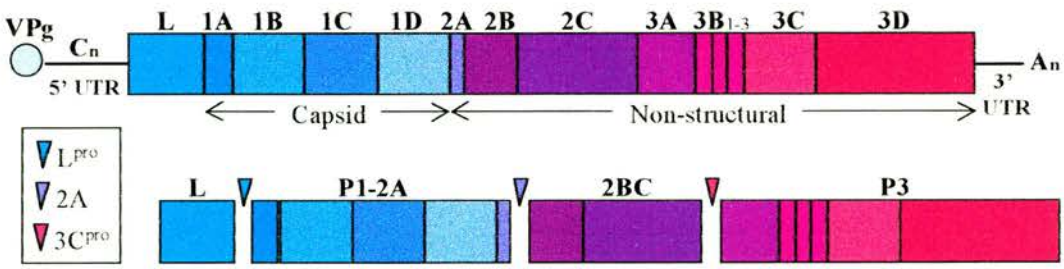


Figure 13: Schematic diagram of the foot-and-mouth disease virus genome. *The single long open reading frame is represented by the boxed area and the untranslated regions (UTRs) by a line. The 5' UTR contains a poly(C) region and the internal ribosome entry site (IRES). Primary cleavage of the polyprotein yield four products, the leader protein (L), the capsid precursor protein attached to 2A (P1-2A) and the replicative protein precursors (2BC and P3). The viral proteins that mediate these processing events are indicated by the triangles.*

1.4.2. Translation of the FMDV Genome

Translation initiation of picornavirus mRNA differs from the normal cap-dependent mechanism used in eukaryotic cells (Kozak, 1986). The cap-independent initiation at an internal AUG start codon allows the viruses to express their gene products when cellular protein synthesis has been switched off.

The protein synthesis of foot-and-mouth disease virus initiates at two start codons located 84 nucleotides apart in the same reading frame. This leads to the synthesis of two alternative N-terminal processing products of the viral polyprotein, the leader proteins Lab and Lb (Clarke *et al.*, 1985; Sangar *et al.*, 1987). The larger of the leader proteins, Lab, is initiated from the first functional AUG and contains all of the sequence of the shorter leader, Lb, plus an N-terminal extension of 28 amino acids. Both the species of leader protein exhibit similar activities (Medina *et al.*, 1993). The leader protein cleaves itself (*in cis* and *in trans*) from the N-terminus of the P1-2A capsid

protein precursor (Strebel & Beck, 1986) (*Figure 12*). It is also responsible for initiating the cleavage of the cap-binding protein complex (eIF-4F) component p220 (eIF-4G) (Devaney *et al.*, 1988). Cleavage of the p220 component underlies the inhibition of cap-dependent translation of host cell proteins.

1.4.3. FMDV Polyprotein Processing.

Although all the picornaviruses encode a single polyprotein, the processing events that yield the individual functional proteins vary significantly from one genera to the next. For FMDV, there are three primary processing events that yield four precursors L, P1-2A, 2BC and P3 (*Figure 13*). These apparent cleavage events seem to occur co-translationally as the full-length polyprotein is not normally observed in infected cells or during *in vitro* translation of FMDV RNA in rabbit reticulocyte lysates (Grubman & Baxt, 1982).

The leader protein mediates the initial FMDV primary cleavage, which liberates this protein from the remainder of the polypeptide. The second primary processing event of the FMDV polyprotein separates the capsid precursor protein from the remainder of the polyprotein. This polyprotein processing event is notably different between the genera of the picornavirus family. In enteroviruses and rhinoviruses, cleavage occurs between 1D and 2A, and is mediated by the 2A proteinase. In the cardioviruses and aphthoviruses, however, processing occurs at the C-terminus of 2A. The mechanism, by which the P1-2A precursor is released from the nascent polypeptide, remains to be fully elucidated and will be discussed later.

The third primary processing event is mediated by the 3C protease and separates the remaining P2 region (2BC) from the P3 region. The 3C protease and, in some cases its 3CD precursor (Vakharia *et al.*, 1987), catalyse all other subsequent cleavage events except the maturation cleavage of the provirion protein VP0, to generate the capsid proteins VP4 and VP2. It is suggested that the picornavirus 3C protease is structurally and evolutionarily related to the trypsin family of serine proteinases with the distinction that the active-site nucleophile serine has been replaced by cysteine (Grubman *et al.*, 1995). In addition to processing virus proteins, picornavirus 3C cleaves a number of host cell proteins (reviewed by Ryan & Flint, 1997).

1.4.4. Overview of FMDV infection cycle

On entry into the host cell, a conformational change in the virion, induced by an interaction with the host cell receptor, releases the viral genome. Once released into the cytosol, the positive-sense mRNA genome of FMDV serves as a template for translation. The initial translation process, together with polyprotein processing provides the proteins required for replication of the viral genome, such as the essential RNA –dependent RNA polymerase (3D). The first step in the synthesis of new viral RNA is to copy the incoming genomic strand to form a complementary minus-strand RNA. The minus-strand then serves as a template for the synthesis of plus-strand RNA, which occurs on the smooth endoplasmic reticulum. Initiation of plus-strand synthesis is rapid and results in multi-stranded replicative intermediates with one minus-strand and many plus stranded copies. In the early stages of infection the newly synthesised plus-strands are utilised as templates for the synthesis of more minus-stranded RNA. As the pool of plus-stranded RNA increases a greater amount is packaged into virions. The completed virus particles are ultimately released by infection-mediated disintegration of the host cell.

This overview is adapted from a comprehensive review of picornavirus infection by Reukert (1996).

1.4.5. Assembly of FMDV virions

The foot-and-mouth disease virus capsid is typical of picornaviruses, containing 60 copies of the four virion proteins, VP1 (1D), VP2 (1B), VP3 (1C) and VP4 (1A). Assembly of picornavirus virions begins when the coat precursor protein (P1) is cleaved, by the 3C protease, to form protomers of three tightly aggregated proteins. These three proteins are VP1, VP3 and VP0, the latter being a precursor of VP2 and VP4 (or 1AB). The rising concentration of these protomers, during infection, triggers their assembly into pentamers. Twelve pentamers then package the VPg-linked RNA genome forming a provirion. The provirion is not infectious until maturation cleavage occurs, yielding VP2 and VP4 from VP0, thus completing the final stages of both processing and assembly (Arnold *et al.*, 1987).

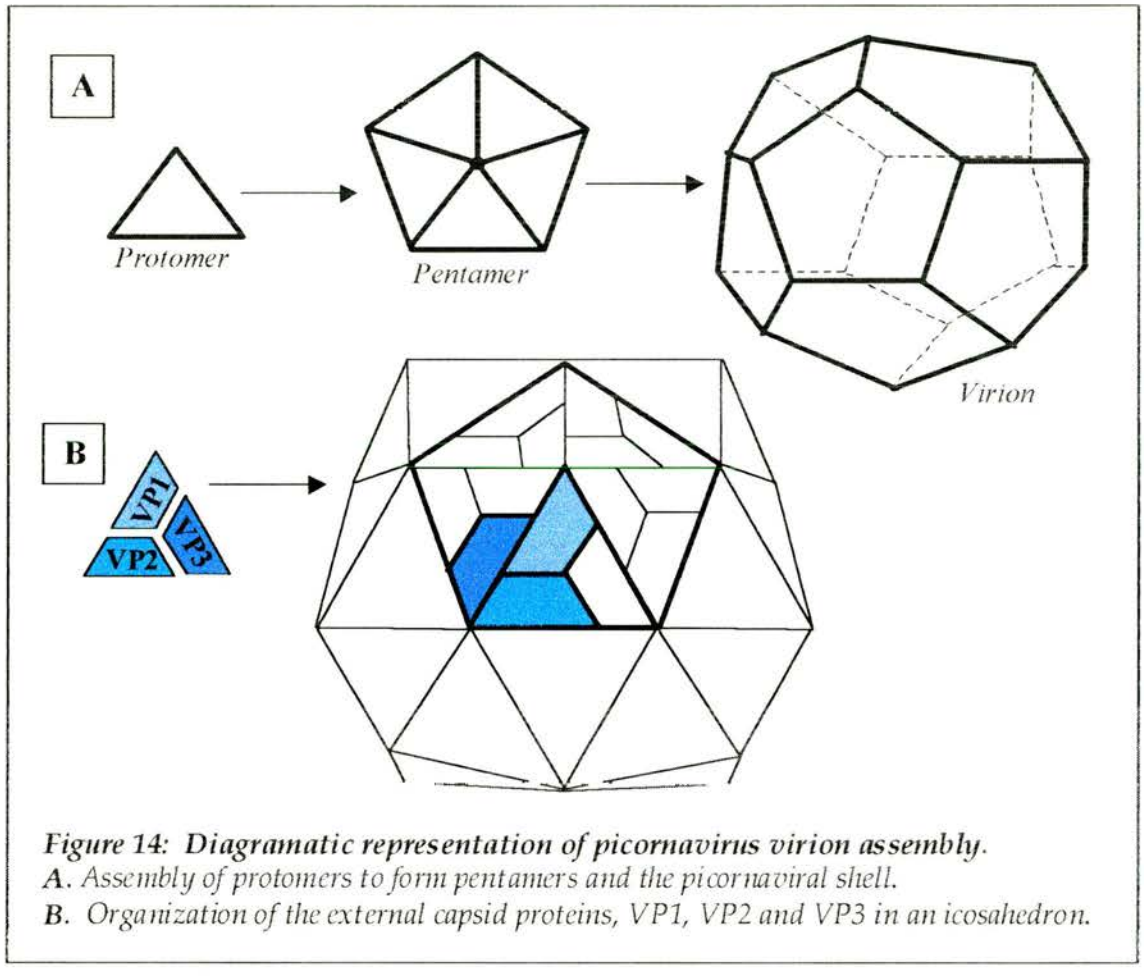


Figure 14: Diagramatic representation of picornavirus virion assembly.
A. *Assembly of protomers to form pentamers and the picornaviral shell.*
B. *Organization of the external capsid proteins, VP1, VP2 and VP3 in an icosahedron.*

1.5. Foot-and-Mouth Disease Virus 2A

The 2A region of foot-and-mouth disease virus is 18 amino acids in length. Situated between the capsid protein domain and the downstream replicative proteins, it is at a site where primary processing occurs. In foot-and-mouth disease, the 2A protein remains attached to the capsid protein precursor after the initial processing events, yielding a P1-2A processing product (*Figure 13*). Subsequent secondary processing mediated by $3C^{pro}$, or more efficiently by $3CD^{pro}$, cleaves 2A away from P1. The mechanism by which processing occurs at the 2A/2B junction, has been the focus of many studies, several of which will be summarized here.

1.5.1. FMDV 2A Active Domain

Following the identification of the L^{pro}, $3C^{pro}$ and the $2A^{pro}$ of entero- and rhinoviruses, initial suggestions that the FMDV 2A also, possessed proteolytic activity seemed reasonable. Deletion of the coding regions around 2A, in cDNA constructs, found the processing activity to be contained in a 19 amino acid region spanning 2A and containing the first residue of 2B, a proline (Ryan *et al.*, 1991).

1.5.2. FMDV 2A-mediated Cleavage of an Artificial Polypeptide

Later work, by Ryan and Drew in 1994, demonstrated that the identified 19 amino acid, 2A spanning sequence, functioned in a similar manner when inserted between two reporter genes, chloramphemicol acetyl transferase (CAT) and β-glucuronidase (GUS). In this reporter gene construct the 2A region of FMDV is in a completely foreign context, devoid of all other FMDV sequences, thus showing no requirement for L^{pro} or $3C^{pro}$. Analysis of coupled transcription-translation reactions, programmed with the reporter construct, showed that the CAT2AGUS polyprotein underwent a co-

translational, apparently autoproteolytic cleavage yielding CAT2A and GUS. However, unlike native polyprotein processing, precursors spanning the 2A/2B-cleavage site were detected (*Figure 15*). The uncleaved CAT2AGUS appeared not to undergo subsequent cleavage, remaining stable in the rabbit reticulocyte lysate system used for translation. In addition, progressive deletion of the N-terminal region of the 2A sequence highlighted that a minimum of 13 residues is required to mediate cleavage. This indicated that although the N-terminal portion of the FMDV 2A region may be influential, it is not essential for cleavage activity. The N-terminal region of 2A performs a role in forming the 1D/2A cleavage site required by the 3C proteinase, for the removal of 2A from the capsid protein precursor.

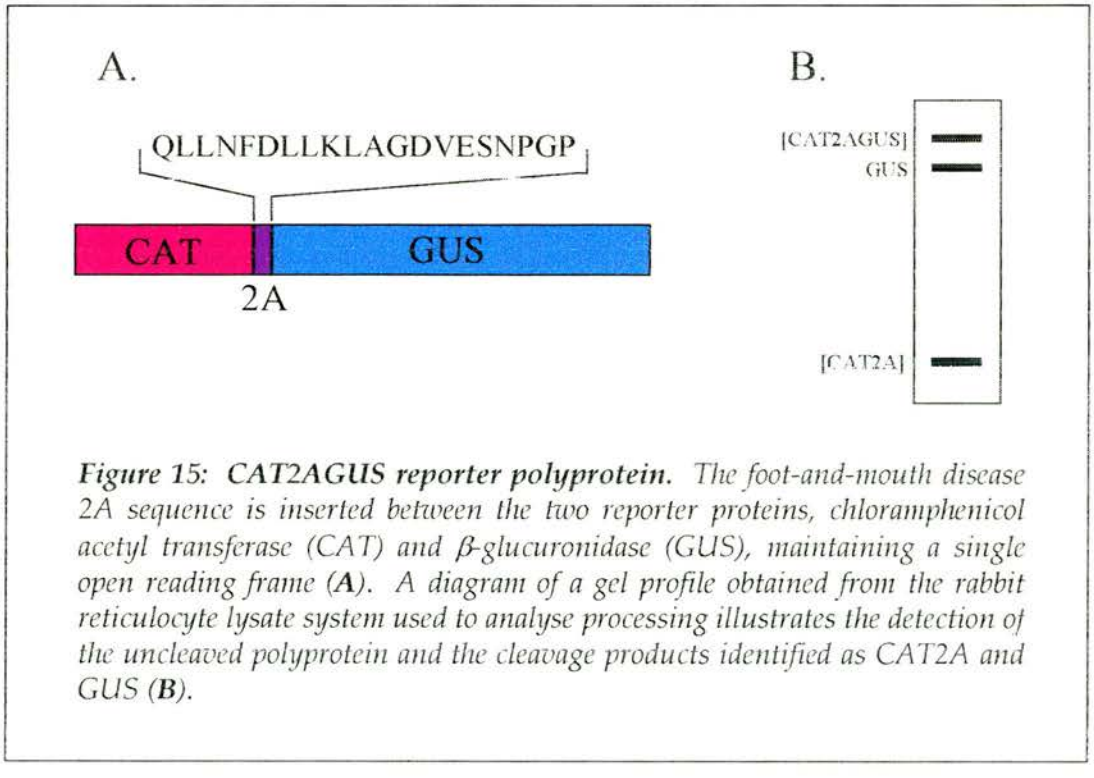


Figure 15: CAT2AGUS reporter polyprotein. *The foot-and-mouth disease 2A sequence is inserted between the two reporter proteins, chloramphenicol acetyl transferase (CAT) and β-glucuronidase (GUS), maintaining a single open reading frame (**A**). A diagram of a gel profile obtained from the rabbit reticulocyte lysate system used to analyse processing illustrates the detection of the uncleaved polyprotein and the cleavage products identified as CAT2A and GUS (**B**).*

Several theories were considered at this stage, to try and explain the co-translational cleavage event. It was suggested that perhaps the FMDV 2A region functioned as a

recognition sequence for a host cell proteinase, although this seemed unlikely due to many constraints. The host cell activity would have to be highly efficient and tightly coupled to translation to achieve the rapidity and yield of 2A/2B cleavage. The host cell proteinase would have to be conserved amongst a range of cell types and be present in the rabbit reticulocyte lysate system. In addition, the proteinase would require the ability to cleave 2A/2B in a variety of sequence contexts and act only on a proportion of recombinant polyproteins. The preferred hypotheses was that the FMDV 2A sequence either disrupts normal peptide bond formation during translation or that it possesses an entirely novel type of proteolytic activity.

1.5.3. Analogy of Cardiovirus 2A and FMDV 2A

The primary 2A/2B polyprotein cleavage of FMDV is analogous to that of cardioviruses, where it is also mediated by the 2A protein cleaving C-terminally. The cardiovirus 2A region is 150 amino acids long, making it much larger than that of FMDV. The C-terminal region of cardiovirus 2A has a considerable sequence similarity to the much shorter 2A region of FMDV. Both the foot-and-mouth disease virus 2A and the cardiovirus 2A sequences have a highly conserved –DxExNPGP- motif (where the uppercase letters denote residues with absolute sequence conservation) at their C-terminus, which has been shown to be closely involved in the cleavage activity (Hahn & Palmenberg, 1996; Donnelly *et al.*, 1997). When a 19 amino acid sequence spanning the C-terminal region of cardiovirus 2A, and the N-terminal proline residue of protein 2B, is inserted into an artificial polyprotein system, similar to that described for FMDV 2A, it mediates cleavage with high efficiency (Donnelly *et al.*, 1997).

1.5.4. FMDV 2A Mediates a Translational Effect

Further investigation of the FMDV 2A activity in the artificial polyprotein system revealed that the two cleavage products (CAT2A and GUS), were not present in equal quantities (Donnelly, 1997). Analysis of the translation profiles of the CAT2AGUS self-processing artificial polyprotein system showed considerable internal initiation within the CAT sequence in coupled transcription/translation *in vitro* systems. The N-terminally truncated forms of [CAT2A] migrated more rapidly on gels than the complete [CAT2A] protein. When these substantial amounts of truncated [CAT2A] were included in the quantification of translation products the imbalance became more marked. If the 2A region of FMDV functioned as a proteolytic element, the two products of the cleavage reaction would be expected to accumulate in a stoichiometry of 1:1. These results provided the first indication that the 2A-mediated polyprotein-processing event may not be proteolytic in nature.

In, both the rabbit reticulocyte lysate and wheat germ extract translation systems utilized, the first translation product (CAT2A) was present in higher amounts than the second product (GUS). Degradation studies showed that GUS degrades more slowly than CAT2A and that the differential rates of degradation were not to the extent of producing such differences in stoichiometry. It was concluded from this work that more CAT2A than GUS was being translated. This brought about the hypothesis that the 2A/2B-cleavage event is not caused by proteolysis but more likely to be a translational effect. A new model for 2A activity was proposed, based on the evidence that in some cases translation is terminated after 2A, producing [CAT2A] only, in others the ribosomes release [CAT2A] then continue to produce GUS, and in others translation continues as

normal to produce the full-length [CAT2AGUS] polyprotein. This model will be considered further, in terms of the results presented here, in the discussion chapter.

1.5.5. Overview of Translation

As a foundation to consider the theory that the apparent 2A-cleavage event is a translational effect, a brief summary of translation is presented here. Protein synthesis takes place in three stages, initiation, elongation and termination. Initiation results in the binding of the initiator tRNA to the start signal of the mRNA in a ribosome complex. The initiator methionyl-tRNA occupies the P (peptidyl) site on the ribosome. Elongation starts with the binding of an aminoacyl-tRNA to the A (aminoacyl) site of the ribosome. A peptide bond is then formed between the amino group of the incoming aminoacyl-tRNA and the carboxyl group of the initiating methionyl-tRNA. The resulting dipeptidyl-tRNA, translocates from the A site to the P site and the initiator tRNA molecule moves into the E (exit) site before leaving the ribosome (*Figure 16)*. These reactions are powered by the hydrolysis of GTP. A second round of elongation can proceed with the binding of another aminoacyl-tRNA to the now vacant A site.

Termination occurs when a stop signal on the mRNA is read by a protein release factor. The binding of a release factor to a termination codon in the A site activates peptidyl transferase, the enzyme normally involved with the transfer of the peptidyl-tRNA from the A to P site, so that it hydrolyses the bond between the polypeptide and the tRNA in the P site. The specificity of peptidyl transferase is altered by the release factor so that water rather than an amino group is the acceptor of the peptidyl moiety. The detached polypeptide chain is free to leave the ribosome.

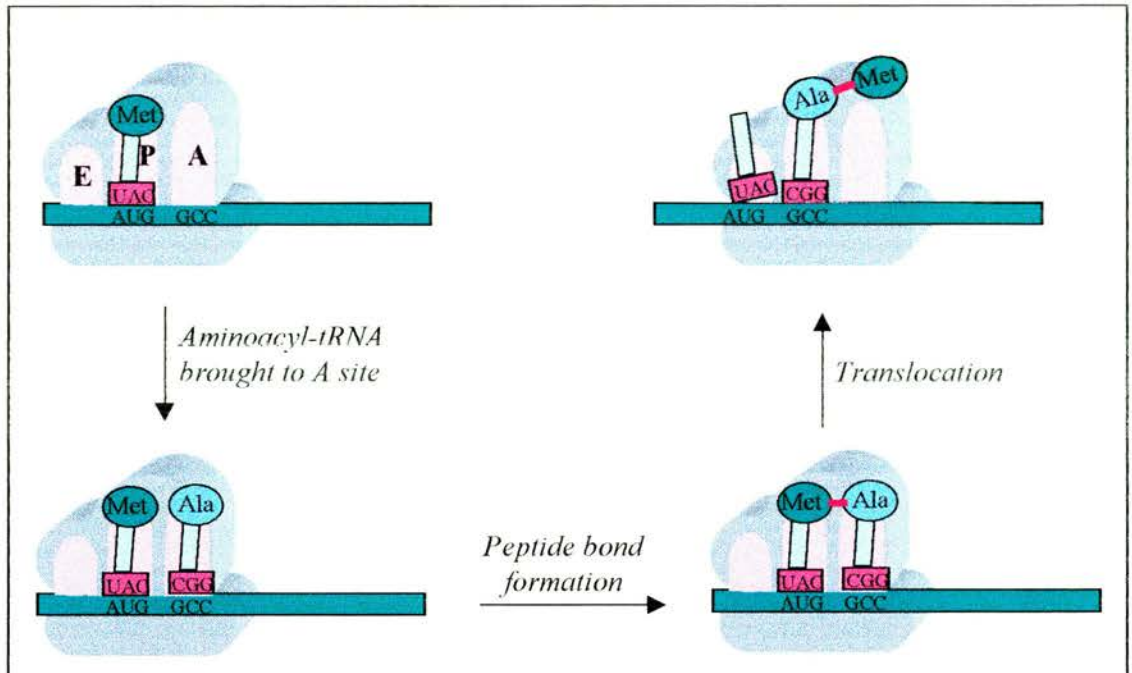


Figure 16: Translation initiation and elongation. *The initiation of protein synthesis results in the delivery of the initiator-tRNA into the P-site of the ribosome. The second aminoacyl-tRNA is positioned in the A-site and peptide bond formation occurs. Translocation brings the dipeptidyl-tRNA into the P-site and the uncharged tRNA to the exit site, with the ribosome moving three nucleotides along the mRNA. Elongation proceeds with a new aminoacyl-tRNA entering the now vacant A-site and the release of the uncharged tRNA from the exit site.*

1.5.6. The History of FMDV 2A

I have presented in this introduction an account of what was known about the processing activity of FMDV 2A at the outset of this work. The study of FMDV 2A within artificial polyprotein systems has provided insight into the mode of action of this short amino acid sequence. In constructs where the upstream or downstream contexts of the viral 2A (including the N-terminal proline of 2B) sequence were replaced, processing activity was retained. Although the activity was influenced by the upstream context, the 2A sequence itself could mediate the apparent cleavage (Ryan *et al.*, 1991). This was confirmed with

the constructs in which the 2A sequence is flanked by reporter proteins and devoid of all other FMDV sequences. Processing in this context was mediated at ~90% and any 'uncleaved' material was not subsequently cleaved upon prolonged incubation, indicating that the 2A-mediated event occurs only co-translationally (Ryan & Drew, 1994). The fact that the FMDV 2A protein does not mediate processing in prokaryotic systems suggests that the co-translational processing event is specific for 80S ribosomes or the eukaryotic environment (Donnelly *et al.*, 1997). The stoichiometric imbalance in the 'cleavage' products of the [CAT2AGUS] artificial polyprotein system provided the first indication that it may not be proteolysis that occurs at the 2A C-terminus but rather a translational effect.

1.5.7. 2A-like Sequences

The analogy of the C-terminus of cardiovirus 2A to FMDV 2A has been well documented and the –DxExNPGP- motif shown to be immediately involved in processing activity (Hahn & Palmenberg, 1996; Donnelly *et al.*, 1997). To determine whether this type of protein biogenesis is confined to the picornaviruses, or is a more widely adopted strategy, database searches were performed to find 2A-like sequences. The databases were probed for the occurrence of the –DxExNPGP- motif within protein sequences. Several 2A-like sequences were identified in these searches with origins in other picornaviruses, insect viruses, type C rotaviruses and trypanosomal, bacterial and cellular sequences. The activities of these 2A-like sequences in an artificial polyprotein system were examined using *in vitro* translation systems. A series of constructs was produced in which the 2A-like sequence is flanked by two reporter proteins, green fluorescent protein (GFP) and β-glucuronidase (GUS), maintaining a single open reading

frame. The translation profiles of these constructs were analysed and the processing activities assayed by the generation of the discrete [GFP2A] and GUS products.

At the beginning of this work it was becoming evident that the processing event that occurs at the 2A/2B junction of the FMDV polyprotein is unusual and distinct from the other previously identified proteolytic reactions characteristic of the picornaviruses. In an attempt to identify the elements of the 2A region that are required for this protein to mediate a translational effect, and to provide information on the mechanism by which the reaction takes place, 2A mutants and naturally occurring 2A-like sequences were examined. It was believed that by identifying functional 2A-like sequences and comparing them to FMDV 2A, this would increase our knowledge of the sequence requirements of an actively processing 2A protein and aid in the development of a model for 2A action. The identification of functional 2A-like sequences has also allowed speculation into the possible roles of this type of protein biogenesis in the organisms in which the sequences were found. The 2A sequence of FMDV has been used for the expression of multiple proteins for plant biotechnology and gene therapy applications, some of which will be discussed later. Evidently, understanding the 2A-mediated effect is not only significant in the biology of 2A and 2A-like sequence containing organisms but is also important in the development of 2A as a biotechnological tool.

1.6.

Aims

- Develop a reporter system to analyse the co-translational processing activity of FMDV 2A.

- Utilise the reporter system to examine FMDV 2A processing in wheat germ extract *in vitro* coupled transcription and translation systems.

- Employ the reporter system as a screen to access the processing activities of other 2A and 2A-like sequences.

- Develop reporter systems for the analysis of 2A activity in cellular expression systems.

2. EXPERIMENTAL

2.1. Cloning

2.1.1. Preparative Restriction Enzyme Digests.

All restriction enzymes were purchased from Promega. Plasmid DNA digestions were carried out under the conditions specified by the supplier. Typically 2 μg of DNA were digested with 2 units of enzyme, in a final volume of 40 μl containing 4 ml of x10 restriction buffer. The reactions were incubated at 37 °C for a minimum of 2 hours.

2.1.2. Analytical Restriction Enzyme Digests.

In general 0.2 μg of DNA were digested with 0.2 units of enzyme, in a total volume of 10 ml, including 1 ml of x10 restriction buffer. The reactions were incubated at 37 °C for 30 minutes.

2.1.3. Polymerase Chain Reactions (PCRs)

PCR was used to amplify specific segments of DNA. The PCRs were typically carried out in a final volume of 100 μl, using 20 ng of template DNA and 100 pmol of each primer. The PCRs were catalysed by TAQ DNA polymerase (Promega) in the buffer supplied by the manufacturer (50 mM KCL, 10 mM Tris-HCL pH9.0, and 0.1% Triton X-100), supplemented with $MgCl_2$ at a final concentration of 2.5 mM and dNTPs at a final concentration of 250 μM. The reactions were overlaid with mineral oil to prevent evaporation during the hot-start addition of polymerase. The reactions were initially heated to 94 °C for 5 minutes, then 85 °C for 1 minute while 2.5 units of the TAQ DNA polymerase was added. The DNA thermal cycler (Techne) was programmed to denature

at 94 °C for 1 minute, anneal at 50 °C for 1 minute and elongate at 72 °C for 1 minute, for a total of 25 cycles. A final step of elongation at 72 °C for 10 minutes was added to ensure a high yield of full-length double stranded product. All PCR reactions were purified, as directed, using the Wizard PCR Prep purification system from Promega.

2.1.4. Agarose-gel Electrophoresis.

Flat bed agarose gels of 1 % w/v agarose (Gibco) were prepared with TAE buffer (0.04 M tris-acetate, 0.001 M EDTA), which contained ethidium bromide at a final concentration of 0.5 μg/ml. DNA samples were loaded onto the gel in agarose gel loading buffer (5% [v/v] glycerol, 0.005%[w/v] bromophenol blue). Electrophoresis was carried out at 50-180 V in TAE buffer containing 0.5 μg/ml ethidium bromide. The DNA bands were subsequently visualized by illumination from an UV transilluminator.

2.1.5. Purification of DNA Fragments from Agarose Gel.

Preparative restriction digests were loaded into preparation wells of flat bed agarose gels, prepared as described above. The area in front of the band of interest was excised and filled with low-melting-point (LMP) agarose. The band of DNA was run into the LMP agarose and purified using the Wizard Prep DNA Purification system (Promega), as directed by the instructions of the manufacturer.

2.1.6. Ligations.

Ligation reactions were set up in T4 DNA ligase buffer (30 mM Tris-HCl (pH7.8), 10 mM $MgCl_2$, 10 mM DTT and 5 mM ATP) at a final volumes of 20 μl. Typically the reaction mixture consisted of 1 μl (3 Weiss units) of T4 DNA ligase (Promega), 0.5 μg of

vector DNA and an equal to 2-fold concentration of insert. The reactions were incubated overnight at 16 °C.

2.1.7. Transformation of E. coli (JM109).

Competent *E. coli* cells were either, purchased from Promega or prepared and transformed using the calcium chloride method.

A) Transformation of Promega competent JM109s

In a 1.5 ml microcentrifuge tube, 10 µl of a ligation reaction were added to 50 µl of competent cells. The tubes were flicked gently to mix the contents and placed on ice for 20 minutes. Heat-shock was performed at 42 °C in a water bath for 45-50 seconds. The tubes were then immediately placed on ice for 2 minutes. Cells were plate on agar supplemented with 100 µg/ml ampicillin.

B) Preparation and Transformation of JM109s by Calcium Chloride Method

50 ml of LB broth was inoculated with 1 ml of an overnight culture and incubated on a shaker at 37 °C until an OD_{600} of 0.3 is achieved. The cells were then cooled to 0 °C for 10 minutes in ice cold polypropylene centrifuge tubes and pelleted at 4000 rpm for 10 minutes at 4 °C. The supernatant was discarded and each pellet re-suspended in 10 ml of ice cold 0.1 M calcium chloride and stored on ice for 15 minutes. The cells were recovered by centrifugation, as above, the supernatant was discarded and the pellet re-suspended in 1 ml of ice cold 0.1 M calcium chloride. The cells were then stored on ice for a minimum of 30 minutes. To 200 µl of cells, in a 0.5 ml eppendorf, 20 µl of a ligation reaction is added and stored on ice for 30 minutes. The cells were then heat shocked in a water bath at 42 °C

for 90 seconds and immediately replaced on ice for 2 minutes prior to plating on agar containing 100 μg/ml ampicillin.

2.1.8. Mini-preparation of Plasmid DNA.

Plasmid DNA was prepared from cells using the Wizard Plus method. 5 ml of LB medium containing 100 μg/ml ampicillin was inoculated with a single colony and grown overnight at 37°C with moderate shaking. Plasmid DNA was prepared using the Wizard Plus SV Minipreps DNA Purification System (Promega), as directed by the manufacturer's protocol.

2.1.9. Maxi-preparation of Plasmid DNA.

Maxi-preparations of plasmid DNA were achieved using the Qiagen Plasmid Maxi Kit. Typically, a single colony was used to inoculate an eight hour 5 ml starter culture (5 ml of LB containing 100 μg/ml ampicillin). An overnight 100 ml culture (100 ml of LB containing 100 μg/ml ampicillin) was then inoculated with 200 μl of the starter culture. The DNA was extracted following the manufacturer's protocol and resuspended in water.

2.1.10. DNA Sequencing.

500ng of mini- or maxi-prepared DNA, together with 3.2pmol of sequencing primer in a final volume of 12μl was submitted to the Automated DNA Sequencing Unit, University of St. Andrews.

2.2.1. Coupled Transcription/Translation (T_NT) Reactions.

Proteins were expressed *in vitro* using the T_NT® Wheat Germ Extract System, according to the manufacturer's instructions (Promega). Briefly, the proteins were radiolabelled with [^{35}S]-methionine. Reactions were incubated at 30 °C for 90 minutes then stopped by the addition of an equivalent volume of 2x SDS-PAGE sample loading buffer (2 % [w/v] SDS, 20 % [v/v] glycerol, 2 % [v/v] β-mercaptoethanol, 0.2 % [w/v] bromophenol blue, 100 mM Tris, pH6.8). The reactions were analyzed, in 5 µl aliquots, on sodium dodecyl sulfate (SDS) denaturing polyacrylamide gel electrophoresis (PAGE) gels (10 % polyacrylamide).

2.2.2. Immunoprecipitation Reactions with T_NT Samples.

Prior to addition to the T_NT samples, the antibody was bound to the affinity media. The rabbit monoclonal anti-GFP antibody was bound to Protein G Sepharose (Amersham Pharmacia Biotech) and the sheep polyclonal anti-2A serum (kindly provided by G. Luke and raised against the QLLNFDLLKLAGDVESNPG, 19 amino acid of FMDV 2A) was bound to Protein A Sepharose (Amersham Pharmacia Biotech). Prior to the binding of antibody, the beads were prepared in phosphate-buffered saline (PBS). The beads were washed three times with PBS to remove the ethanol in which they were supplied. The beads were recovered from the washes by centrifugation (12,000 g) for 20 seconds. A 50 % slurry was then prepared by adding an equal volume of PBS and the beads were stored at 4 °C. To 100 µl of the 50 % bead solution, 50 µl of the anti-2A serum (to Protein G Sepharose) or 3 µl of the purified anti-GFP antibody (to Protein A Sepharose) was added. The antibody was bound to the beads for 2 hours at 4 °C with continuous mixing. The

beads/antibody complexes were washed five times with PBS containing 0.1 % [v/v] IPEGAL CA-630. The complexes were recovered between washes by centrifugation (12,000 g) for 20 seconds. The final pellet was resuspended in an equal volume of PBS containing 10 mg/ml bovine serum albumin (BSA). This was incubated for 30 minutes at 4 °C with continuous mixing. The complexes were recovered by centrifugation (12,000 g) for 20 seconds and again resuspended in an equal volume of PBS containing BSA. In 1 ml of binding buffer (50 mM Magnesium acetate, 20 mM Tris-HCl pH 8, 1 mM DTT, 0.1 % [v/v] IPEGAL CA-630, 10 mg/ml BSA), 20 µl of the 50 % solution of antibody/beads complexes was added to 7 µl of the wheat germ extract T_NT sample. This was incubated at 4 °C for 30 minutes with continuous mixing. The precipitated antigen complexes were washed five times with wash buffer (50 mM Magnesium acetate, 20 mM Tris-HCl pH 8, 1 mM DTT, 0.1 % [v/v] IPEGAL CA-630). The complexes were recovered between washes by centrifugation (12,000 g) for 20 seconds and the final pellet was aspirated to remove residual wash buffer. The final pellet was resuspended in 8 µl of 2x SDS-PAGE sample buffer and the complexes were dissociated at 95 °C for 3 minutes. The beads were pelleted by centrifugation (12,000 g) for 20 seconds and the supernatant loaded on an SDS-PAGE gel.

2.2.3. Denaturing Polyacrylamide Gel Electrophoresis (SDS-PAGE).

The gels were constructed with a 4 % polyacrylamide stacking gel and a 10 % or 12 % polyacrylamide resolving gel. The gel was run using Hoefer mini-gel apparatus. Electrophoresis was carried out at a constant current of 20 mA for one gel or 40mA for two gels, until the Bromophenol blue dye reached the bottom of the gel. Proteins were stained with Coomassie brilliant blue staining solution (0.2 % [w/v] Coomassie brilliant

blue (R-250), 20 % [v/v] methanol, 20 % [v/v] acetic acid). Gels were subsequently de-stained (20 % [v/v] methanol, 10 % [v/v] acetic acid) and dried.

2.2.4. Visualization of Radiolabelled Translation Products.

The distribution of radiolabel between the translation products, uncleaved [GFP2AGUS] and the discrete GUS and [GFP2A] products, was visualized by autoradiography for preparation of figures, or by phosphorimaging using the FUJIX Bio-imaging analyzer, BAS1000 MacBas system for isotope quantification.

2.2.5. Calculation of Cleavage Activity

The incorporation of radiolabel into the translation products was determined by densitometric analysis of the phosphorimaging profiles (Fujix BAS1000). The photo-stimulated luminescence (PSL) of each band, minus 'background' (-BG) was determined and then divided by the methionine content of the appropriate translation product (PSLcorr). The PSL-BG was determined using the three methods provided by the software (integration of profile peaks, encircling the band 'freehand' or highlighting the bands with rectangles) and the error associated with this method of determination was estimated to be ~2%. The percentage cleavage activity was calculated as:

$([GUS^{PSLcorr}]+[GFP2A^{PSLcorr}]) / ([GFP2AGUS^{PSLcorr}]+[GUS^{PSLcorr}]+[GFP2A^{PSLcorr}]) \times 100$

2.2.6. Estimation of Translational Outcomes

The proportion of ribosomes which; i) synthesize a full-length translation product, ii) synthesize only [GFP2A], and iii) synthesize both [GFP2A] then GUS was calculated as follows:

i) $[GFP2AGUS^{PSLcorr}] / ([GFP2AGUS^{PSLcorr}]]+[GUS^{PSLcorr}]+[GFP2A^{PSLcorr}]) \times 100$

ii) $([GFP2A^{PSLcorr}]-[GUS^{PSLcorr}]) / ([GFP2AGUS^{PSLcorr}]+[GUS^{PSLcorr}]+[GFP2A^{PSLcorr}]) \times 100$

iii) $([GUS^{PSLcorr}] \times 2) / ([GFP2AGUS^{PSLcorr}]+[GUS^{PSLcorr}]+[GFP2A^{PSLcorr}]) \times 100$

2.3. *Analysis of TMEV Infected Cells*

2.3.1. *Growing Cells from –70°C Stock*

Aliquots removed from –70°C storage were defrosted quickly in a water bath. The samples were centrifuged (3000 g) for 5 minutes and the supernatant discarded. The cells were resuspended in 1 ml foetal calf serum (FCS), 100 µl and 900 µl aliquots were added to small tissue culture flasks (25cm^2) containing DMEM supplemented with FCS at 10%. The cells were incubated overnight at 37°C. The medium was removed, the cells washed twice, fresh medium added and the flasks returned to the incubator.

2.3.2. *Splitting Cells*

The medium was poured off and residual medium removed by washing quickly with trypsin. The cells were washed again with trypsin, this time for 20 seconds. The trypsin was poured off leaving a tiny amount in the flask. 10 ml of medium (DMEM + FCS 5%) was added to the flask and the cells dispersed by agitation. To a fresh flask (75cm^2), containing 30 ml medium (DMEM + FCS 5%), 1 ml of the cell solution was added. The cells were maintained in a 37°C incubator.

2.3.3. Growing Virus Stock

Cells grown in DMEM medium supplemented with FCS at 5%, in a large tissue culture flask ($75cm^2$), had the medium removed and were washed twice with DMEM supplemented with FCS at 1% (DMEM + FCS 1%). 2 ml of this medium and 0.5 ml of virus stock ($2.5x10^8$ PFU) were added to the flask containing the cells. The cells were incubated in the presence of virus for 20 minutes, at 37°C, with rocking. The medium was removed, replaced with 10 ml DMEM + FCS 1% and 0.2 ml of HEPES and the cells were incubated overnight, at 37°C, with rocking. The cells in medium were transferred by pipette into a nalgene 50 ml centrifuge tube. Centrifugation (3500 g) for 5 minutes pelleted cellular material and the supernatant, containing released virus, was recovered. The cell pellet was resuspended in 5 ml DMEM and any intact cells lysed by vortexing and freeze/thawing at –70°C. Aliquots were transferred to microcentrifuge tubes, the cell debris pelleted (14,000 g, 5 minutes) and the supernatant recovered. The two supernatants were combined and stored in aliquots at –70°C. Typical virus titres achieved by this method were $5x10^8$ PFU/ml.

2.3.4. Labelling Viral Proteins

Cells from a –70°C stock were grown in a small tissue culture flask ($25cm^2$) as described above. When growth had been established the cells were split into 6 small plates (60mm) and incubated overnight at 37°C, in DMEM + FCS 10%. The medium was removed and the cells washed with PBS. 1 ml of virus stock was added to each plate and incubated for 30 minutes at 37°C. The medium was removed and the cells washed with PBS and the methionine-free (Met-free) medium. To each plate, 1 ml of Met-free medium was added and the cells were incubated for 15 minutes before the addition of 5 µl of ^{35}S-methionine. The cells were incubated at 37°C for a further 1 to 6 hours, with a

plate being removed and analysed every hour. When the plate was removed from the incubator the medium was disposed and the cells washed twice with PBS. 100 ml of lysis buffer (50mM TRIS pH7.4, 150mM NaCl, 0.1% TRITON, 1mM EDTA, 1mM PMSF and 2mM benzamidine) was added to the plate and the cells scraped and collected in an eppendorf. The cells were lysed on ice for 20 minutes, with vortexing at 5 minute intervals. The cell debris was removed by centrifugation (14,000 g, 4°C, 10 minutes) and the supernatant recovered. The proteins were denatured at 95°C for 2 minutes in protein sample buffer and analysed by SDS-PAGE, as described previously.

2.4. Analysis of Cellular Translations

2.4.1. Preparation of Labelled Cell Extract.

Monolayers of transfected Hela cells (kindly provided by P. de Felipe) were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 5% fetal calf serum. The cells were washed with methionine-free medium and incubated in methionine-free medium for 15 minutes, after which 80 μCi of [^{35}S]methionine was added. The cells were incubated for a further 150 minutes. The cells were washed twice with PBS and harvested in 1 ml of lysis buffer (40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 2 mM Benzamidine, 1 mM DTT and 0.5 % [v/v] IPEGAL CA-630). The lysate was vortexed for 10 seconds and freeze/thawed. Upon defrosting, the lysate was vortexed a further three times for 5 seconds replacing on ice for intervals of 20 seconds. The lysate was then clarified by centrifugation (14,000 g) at 4 °C for 10 minutes. The supernatant was recovered for immunoprecipitation (IP) reactions.

2.4.2. Immunoprecipitation Reactions with Cell Extracts.

Bovine serum albumin (10 mg/ml final concentration) and rabbit anti-GFP monoclonal antibody (0.5 µl) were added to the cleared cell lysate. The antibody/lysate mixture was incubated at 4 °C for 90 minutes, with continuous mixing. Protein A Sepharose beads (Amersham Pharmacia Biotech) were prepared in PBS. The beads were washed three times with PBS to remove the ethanol in which they were supplied. The beads were recovered from the washes by centrifugation (12,000 g) for 20 seconds. A 50 % slurry was prepared by adding an equal volume of PBS and stored at 4 °C. To precipitate the immune complexes 20 µl of the 50 % slurry was added to the antibody/lysate solution and this was incubated at 4 °C for 60 minutes with continuous mixing. The beads were recovered by centrifugation (12,000 g) for 20 seconds then washed five times with lysis buffer then twice with PBS. The beads were recovered after each wash by centrifugation (12,000 g, 20 seconds) and aspirated after the final wash. The final pellet was resuspended in 10 µl 4x SDS sample buffer (4 % [w/v] SDS, 40 % [v/v] glycerol, 4 % [v/v] β-mercaptoethanol, 0.4 % [w/v] Bromophenol blue, 200 mM Tris, pH6.8). Dissociation was performed at 95 °C for 3 minutes. The beads were removed by centrifugation (12,000 g) for 20 seconds and the supernatant analyzed on SDS-PAGE gels (12 % polyacrylamide). The gels were stained and the radiolabelled products visualized as described previously.

3. RESULTS

3.1. Construction and Functional Analysis of a New FMDV 2A Artificial Polyprotein

The [CAT2AGUS] artificial polyprotein system used previously to examine the processing activity of FMDV showed N-terminally truncated forms of the full-length polyprotein and [CAT2A] moiety. As described previously these shortened forms, which migrated more rapidly on gels, were attributed to internal initiation within the CAT gene. In an attempt to eliminate the complication of these truncated products from future quantification analyses the CAT gene was replaced with that of green fluorescent protein (GFP).

3.1.1. pSTA1 – The Primary Construct

Plasmid pCAT2AGUS (Ryan & Drew, 1994) was cleaved with the restriction enzymes *BamHI* and *XbaI* and the large DNA restriction fragment purified by agarose gel electrophoresis. Plasmid pGFP-N2 (Clontech) was similarly restricted and the small restriction fragment, representing the GFP gene, was gel purified. The purified restriction fragments were ligated to form pGFP2AGUS.

To allow the purification of the translation products from expression systems, a His_{x6} affinity purification tag was introduced at the N-terminus of both the GFP and GUS genes. Firstly the GFP gene was tagged using the forward oligonucleotide primer HISGFP (5' CGCGCGGGATCCACCATGGGGCACCACCACCACCACCACGGTAA-AGGAGAACTT 3') and the reverse SP6 primer (5' TATTTAGGTGACACTATAG 3'). Plasmid pGFP2AGUS was used as the template. The PCR product was doubly restricted with *BamHI* and *ApaI*, gel purified, then ligated into the large fragment of a similarly

restricted pGFP2AGUS to form pHisGFP2AGUS. Prior to the affinity tag being introduced at the N-terminus of GUS, the HisGFP2AGUS fragment was transferred into the pYES, yeast expression vector (Invitrogen), using the *BamHI* and *NotI* restriction enzymes. The resultant plasmid, pPJ1, was doubly restricted with *ApaI* and *NotI* prior to ligation with the HisGUS fragment. HisGUS was formed by PCR using the HISGUS oligonucleotide forward primer (5' CGCGCGGGGCCCCACCACCACCACCACC-ACTTACGTCCTGTAGAAACC 3'), SP6 as the reverse primer and pGFP2AGUS as the template. The PCR product was doubly restricted with *ApaI* and *NotI*. The purified PCR fragment was ligated with the similarly restricted pPJ1 large fragment.

The FMDV 2A sequence flanked by the His_{x6}-tagged GFP and His_{x6}-tagged GUS reporter genes was transferred into the cellular expression vector pcDNA3.1 (Invitrogen). Prior to the transfer of the HisGFP2AHisGUS cDNA, the multiple cloning site of the pcDNA3.1 vector was manipulated to remove the *XbaI* and *ApaI* restriction sites. The vector was restricted with *XbaI* and *ApaI*, treated with T4 DNA polymerase to produce blunt ends and religated to form pcDNA3.1x. Plasmid pPJ2 was doubly restricted with BamHI and XhoI and the restriction fragment encoding HisGFP2AHisGUS purified. The doubly His_{x6}-tagged fragment was inserted into pcDNA3.1x vector, similarly restricted, to form pcDNAPJ2.1 (*Figure 17*).

The fundamental pSTA1 plasmid is highly similar to pcDNAPJ2.1, but has the *FokI* site removed and context of the *XbaI* restriction site (at the 3' terminus of the GFP coding sequence) altered so that it is not subject to Dam methylation. The Hisx6-tagged GFP coding sequence was amplified by PCR using the T7 forward oligonucleotide primer

(5' TAATACGACTCACACTATAGGG 3'), the F4389 reverse primer (5' GCGCGCTCTAGACCCGGACTTGTATAGTTCGTCCATG-CCATGTGTAAT 3') and pcDNAPJ2.1 as the template. The PCR product was doubly restricted with *BamHI* and *XbaI*, gel purified and ligated into pcDNAPJ2.1, similarly restricted (*Figure 17*).

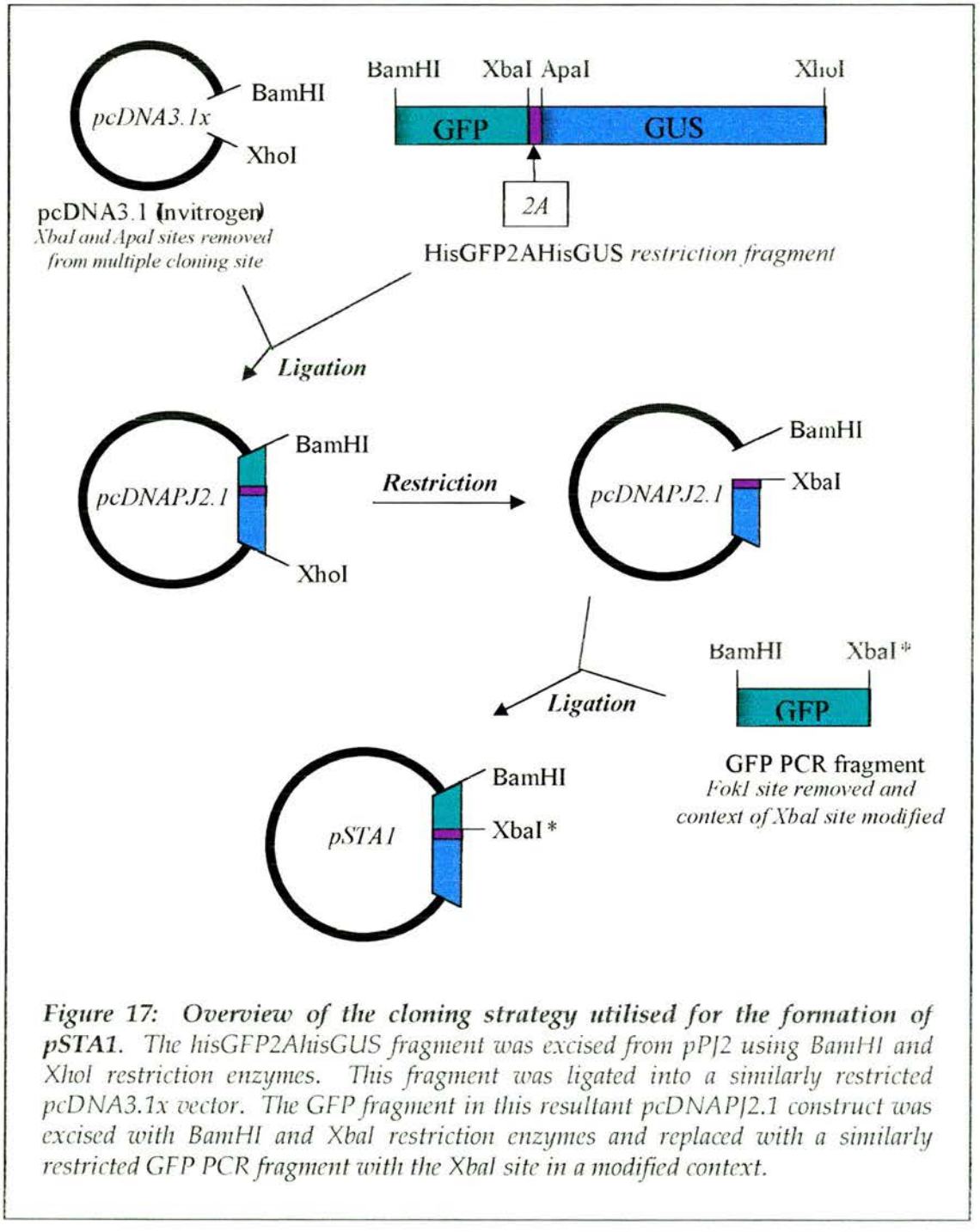


Figure 17: Overview of the cloning strategy utilised for the formation of pSTA1. *The hisGFP2AhisGUS fragment was excised from pPJ2 using BamHI and XhoI restriction enzymes. This fragment was ligated into a similarly restricted pcDNA3.1x vector. The GFP fragment in this resultant pcDNAPJ2.1 construct was excised with BamHI and XbaI restriction enzymes and replaced with a similarly restricted GFP PCR fragment with the XbaI site in a modified context.*

3.1.2. Characterization of the pSTA1 Translation Products.

The pSTA1 plasmid was used to program wheat germ extract coupled transcription/translation reactions and the radiolabelled translation products were separated by 10% SDS-PAGE. The distribution of radiolabel was visualised either by autoradiography or by phosphorimaging. The pSTA1 construct showed a gel profile similar to that observed previously with the [CAT2AGUS] system (Donnelly et al., 1997). In the [GFP2AGUS] system, the three major translation products, [GFP2AGUS], GUS and [GFP2A] were easily identifiable by size (*Figure 18*) and the problem of internal initiation experienced with the CAT gene was removed by replacing it with the GFP gene.

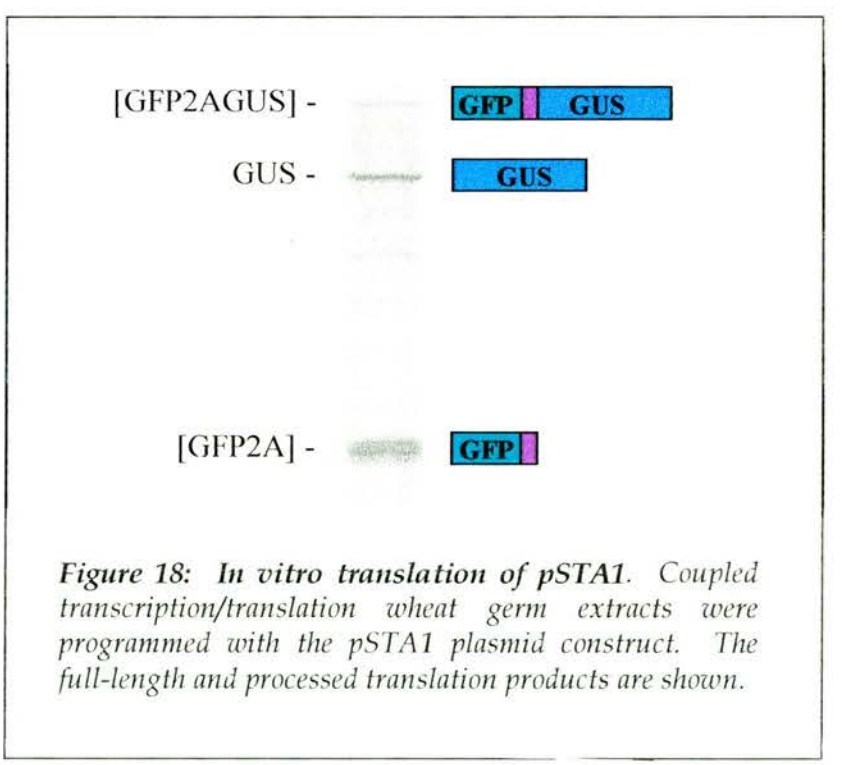


Figure 18: In vitro translation of pSTA1. *Coupled transcription/translation wheat germ extracts were programmed with the pSTA1 plasmid construct. The full-length and processed translation products are shown.*

3.1.3. Quantification of the pSTA1 Translation Products.

Quantification of the incorporation of radiolabel into each of the three translation products was achieved by densitometric analysis. The values for photostimulated light

minus background (PSL-BG) were corrected for the methionine content of the translation product (PSLcorr) to allow a direct comparison. The processing activity was calculated as the percentage of radiolabel incorporated into the GUS and [GFP2A] products compared to the total incorporation into the three forms, [GFP2AGUS], GUS and [GFP2A]. The software allowed three methods of quantification that involved peak profile searches where the major bands were identified in an automated manner, highlighting the bands in uniform rectangles or encircling the bands freehand (*Figure 19*). Multiple densitometric analyses of the same translation profile, using the alternative methods of quantification, have shown that the error associated with this estimating processing activity is ±2%. The combined error in phosphorimaging analysis of translation profiles obtained from an identical construct analysed in multiple different translation reactions (using the same batch of wheat germ extract) was calculated to be ±5% for processing activity. The greatest variation in processing activities was observed between different batches of the wheat germ extract.

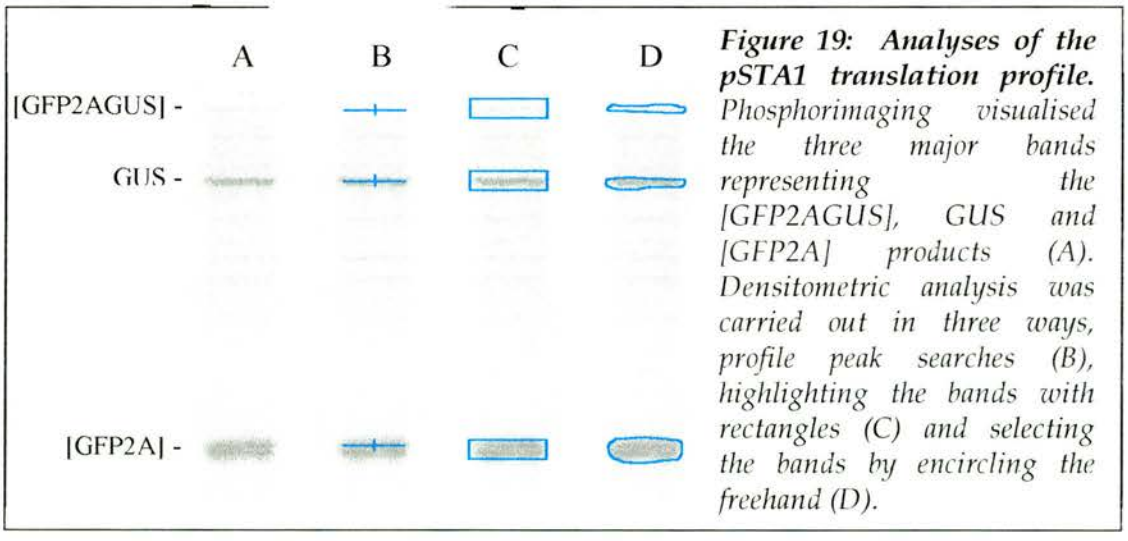


Figure 19: Analyses of the pSTA1 translation profile. *Phosphorimaging visualised the three major bands representing the [GFP2AGUS], GUS and [GFP2A] products (A). Densitometric analysis was carried out in three ways, profile peak searches (B), highlighting the bands with rectangles (C) and selecting the bands by encircling the freehand (D).*

Densitometric analysis of the translation products also highlighted a stiochiomteric imbalance between [GFP2A] and GUS consistent with previous reports for the

[CAT2AGUS] artificial polyprotein system (Donnelly, 1997; Ryan *et al.*, 1999). The product N-terminal of 2A ([GFP2A]) was always in excess of that C-terminal (GUS), with the excess varying between batches of wheat germ extract (from 2:1 to 6:1).

3.1.4. Estimation of the Translational Outcomes of pSTA1.

In an attempt to further characterize the translation profiles, an estimation of the translation outcomes was determined. The proportion of ribosomes which I) synthesize a full-length translation product [GFP2AGUS], II) synthesize the two discrete translation products [GFP2A] and GUS, and III) terminate translation after the synthesis of [GFP2A] was calculated. These proportions have varied widely between the batches of wheat germ extract programmed with the construct pSTA1. Over the period of these studies it appears that the properties of the wheat germ extracts have altered so that less of the full-length product is synthesized and there is an increased proportion of the ribosomes terminating after the synthesis of [GFP2A].

Sample	Processing Activity (%)	Ratio [GFP2A]:GUS	Proportion (%) of Ribosomes Synthesizing: [GFP2AGUS]	[GFP2A] & GUS	[GFP2A] only
1	88	2.74	15	31	54
2	91	2.77	12	32	56
3	91	1.94	13	45	42
4	97	3.84	4	25	71
5	95	5.84	5	17	78
6	95	5.13	5	18	77
7	98	6.01	2	17	81

Table 1: Analyses of the pSTA1 translation profiles. *Samples are listed in chronological order. The calculations were based on the PSLcorr values measured by densitometric analysis. Processing activity is expressed as the PSLcorr of translation products 1 and 2 as a percentage of the total PSLcorr for the three translation products. The translational outcomes are expressed as a percentage of the total initiation from the 5′ end of the polyprotein. (The percentage values have been rounded to the nearest whole number).*

3.2. Influence of N-terminus of 2B on the Activity of FMDV 2A

The 'cleavage' activity of FMDV 2A within the GFP2AGUS artificial polyprotein system is high but not complete, with full-length polyprotein detectable in the translation profiles. However, when 2A is in its native context, in a construct encoding the P1 and P2 regions of FMDV, it is completely active with no full-length polyprotein detected in the translation profile (Donnelly *et al.*, 2001a). Progressively "building-in" the sequence of 1D, N-terminal to 2A in the reporter construct, improves the cleavage activity of 2A in this environment. The longer the 1D sequence, the lower the levels of full-length polyprotein (Donnelly *et al.*, 2001b). It was of interest at this stage to assess what contribution the downstream sequences had on 2A activity. The sequence of interest was the –FFF- sequence that occurs immediately downstream of the initial proline of FMDV 2B and which is highly conserved.

3.2.1. Plasmid Constructs with 2B N-terminal Sequence.

To examine the effect, if any, of the sequences downstream of 2A in the viral genome, constructs encoding the –FFF- sequence following the initial proline of the viral 2B were formed. In order to simplify the comparison of incorporation of radiolabel, into the bands on the gel translation profiles, the GUS portion of the reporter construct was shortened to have a similar methionine content to the GFP2A portion. Constructs devoid of the –FFF- sequence but containing the shortened GUS fragment (ΔGUS) were also formed.

Plasmids encoding the polyproteins GFP2AΔGUS, GFP2AFFFΔGUS, ΔGUS2AGFP and ΔGUS2AFFFGFP were formed from constructs, pGFP2AGUS, pGFP2AFFFGUS, pGUS2AGFP and pGUS2AFFFGFP that already existed in our laboratory. The

constructs were manipulated to replace the GUS fragments with a shortened ΔGUS version (*Figure 20*). The ΔGUS fragments were generated by PCR. The template DNA for the PCR reactions was pGFP2A$_{EXT60}$HisGUS, also available within the laboratory. The forward primer was either T7 or BamΔGUS and the reverse primer was ΔGUS-Sph/Nsi. At the final stage of cloning the sequences encoding the polyproteins were transferred into the pSTA1 backbone, yielding the plasmids pSTA1/P (GFP2AΔGUS), pSTA1/PF (GFP2AFFFΔGUS), pSTA1/D (ΔGUS2AGFP) and pSTA1/DF (ΔGUS2AFFFGFP).

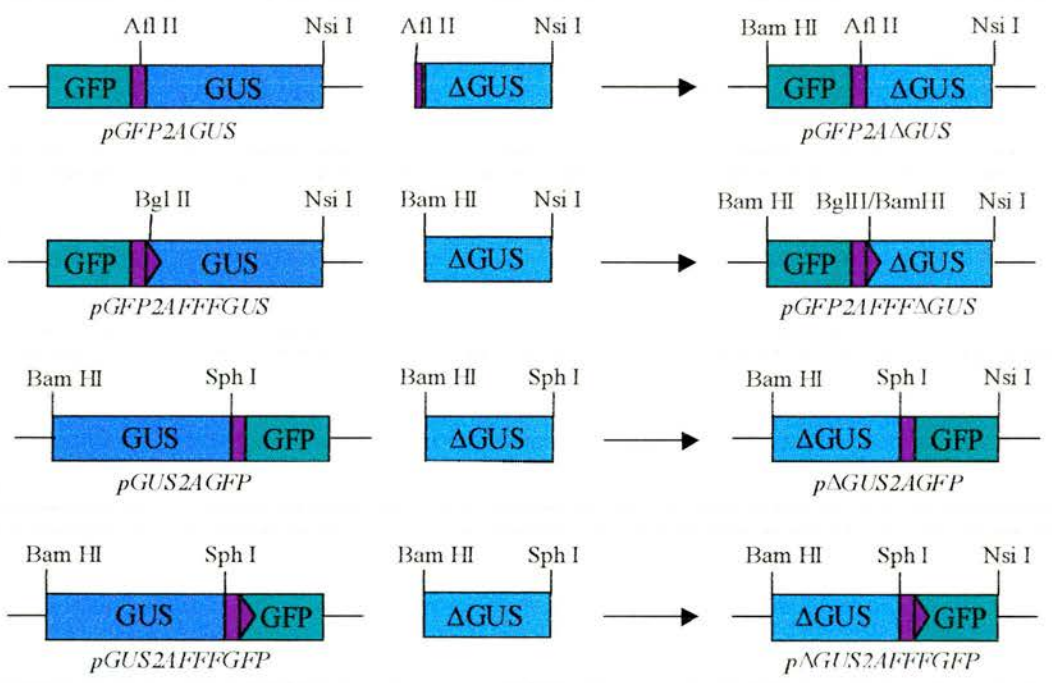


Figure 20: Cloning of the ΔGUS constructs designed to analyze the influence of -FFF- on FMDV 2A activity. *The full-length GUS fragments were removed using the restriction enzymes shown and were replaced with the appropriately restricted PCR fragments of the shortened version of GUS. The resultant intermediate constructs were in turn digested with BamHI and NsiI and transferred into the similarly restricted pSTA1 vector.*

3.2.2. Translation Profiles of the ΔGUS Constructs.

The ΔGUS constructs were used to programme wheat germ extract coupled transcription/translation reactions. The radiolabelled translation products were separated by 10% SDS-PAGE and visualized by autoradiography or phosphorimaging. The ΔGUS constructs showed gel profiles similar to that produced by pSTA1, with the three major translation products; full-length polyprotein, ΔGUS and [GFP2A] easily identifiable by size (*Figure 21*).

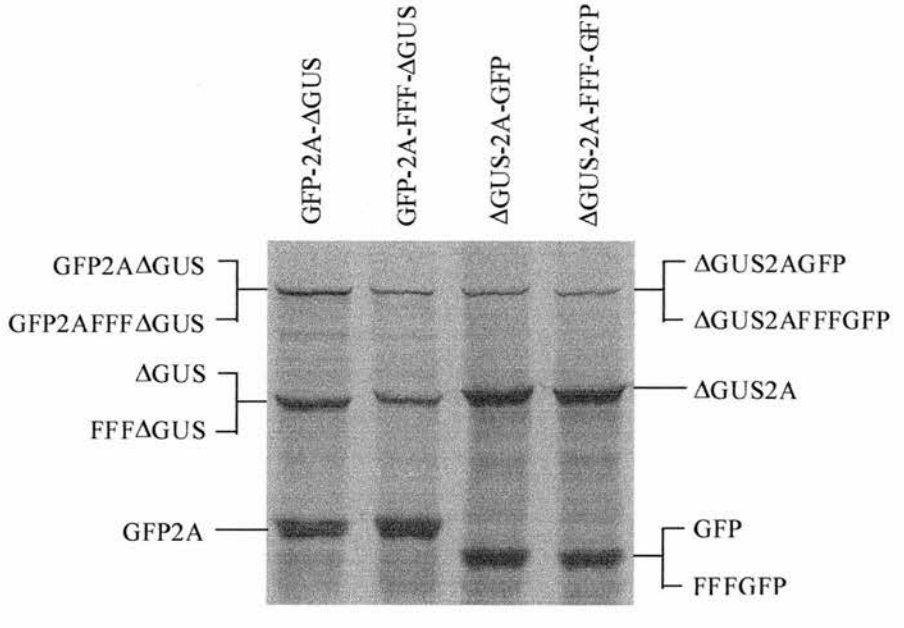


Figure 21: In vitro translation profiles of the ΔGUS constructs. *The lanes are labelled with the polyprotein that the single open reading encodes. The full-length and processed translation products are indicated.*

3.2.3. Analysis of the ΔGUS Translation Profiles.

The introduction of the N-terminal sequence of 2B into the reporter construct has dramatic effect on the translation profiles. Although the cleavage activity of 2A in each of the constructs is equally high, the presence of the -FFF-encoding sequence dramatically alters the translational outcomes in terms of the downstream translation product. Instead of translation resulting largely in the synthesis of the two discrete translation products, the -FFF-encoding sequence causes more ribosomes to terminate after synthesis of the first translation product. This enhancement of the imbalance in translation products is observed when the -FFF-encoding sequence is introduced into either genome arrangement, pGFP2AΔGUS or pΔGUS2AGFP (*Table 2*).

Plasmid	Polyprotein	Processing Activity (%)	Proportion of Synthesis (%): Full-length	TP1&TP2	TP1 only
pSTA1/P	GFP2AΔGUS	~98	1	88	11
pSTA1/PF	GFP2AFFFΔGUS	~95	5	48	47
pSTA1/D	ΔGUS2AGFP	~98	2	71	27
pSTA1/DF	ΔGUS2AFFFGFP	~98	1	47	52

Table 2: Analyses of the translation profiles generated by the ΔGUS constructs. *These constructs were designed to assess the influence of the 2B -FFF- sequence on FMDV 2A processing activity. Processing activity and the proportion of ribosomes synthesizing each of the three translational outcomes were calculated using the PSLcorr values obtained from densitometric analysis.*

The characterization of FMDV 2A presented here, together with the imbalance in translation products reported elsewhere, has shown that the presence of 2A alone can result in termination of translation after the synthesis of the N-terminal translation product (TP1). This occurrence is thought to be a result of the sub-optimal functioning

of 2A in the *in vitro* expression systems which also gives rise to the full-length polyprotein product. These outcomes of translation are not observed *in vivo* where processing is complete producing the two discrete products. In the expression of the artificial polyproteins examined here it appears that the presence of the triple phenylalanine sequence of 2B promotes termination of translation after the synthesis of 2A. The translation of this stretch of phenylalanine residues brings about a translational block in the *in vitro* system that must ultimately result in termination and dissociation of the ribosome complex. A similar role *in vivo* would provide the virus with a useful strategy of augmenting the translational control mediated by 2A, allowing the structural proteins to be synthesized in excess of the enzymatic non-structural proteins. With 60 copies of each structural protein required to encapsidate a single viral genome, combined with the limited pool of translation reagents that results from host cell shut-off, this effect may be more important when translation becomes more 'stressed' later in infection. The translational block that occurs after the synthesis of the structural proteins of FMDV may be analogous to the control of expression of early and late genes in other viruses.

3.3. The Search for Naturally Occurring 2A-like Sequences

In an attempt to determine whether the translational effect mediated by FMDV 2A is unique to picornaviruses or is a more widely adopted strategy of protein biogenesis, database searches were performed to identify any naturally occurring 2A-like sequences. The aim was to find sequences resembling FMDV 2A, clone the DNA sequences into the pSTA1 construct and screen for processing activity by the *in vitro* synthesis of the two discrete translation products, GFP2A and GUS.

3.3.1. Bioinformatics.

The databases were probed for the occurrence of the –DxExNPGP- motif, conserved among aphthoviruses and cardioviruses and known to closely associated with processing activity. Database searches were performed using the PATTINPROT facility of the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (http://www.expasy.org). This is a proteomics and sequence analysis tool that runs pattern and profile searches. The SWISSPROT (Swiss Institute of Bioinformatics), TrEMBL (European Molecular Biology Laboratory translations) and GenPept (National Center for Biotechnology Information, GenBank translations) databases were probed using this facility.

Database searches for sequences similar to that of FMDV 2A revealed that there are, indeed, other 2A-like sequences within the database. Some picornaviruses other than foot-and-mouth disease virus have 2A sequences and the porcine, bovine and human type C rotavirus have a 2A-like sequence. Some insect viruses, including those thought to be picornavirus-like, have 2A-like sequences present in their genomes. 2A-like sequences were also found in bacterial, trypanosomal and other cellular sequences (*Figure 22*).

3.3.2. 2A-like Plasmid Constructs

To analyze the processing abilities of the naturally occurring 2A-like sequences a series of constructs were made. Using the pSTA1 construct as the basis, the FMDV 2A-coding sequence was replaced individually with the coding sequence of each of the 2A-like sequences, using the PCR. Oligonucleotide primers (Oswel DNA services), encoding the 2A-like sequences, were designed to anneal to the 3' terminus of the GFP gene in pSTA1, for reverse priming of PCR. The reverse primers additionally contained an ApaI restriction site downstream of the 2A-like sequence, for subsequent cloning (*Table 2*). With the forward oligonulceotide (T7) primer and pSTA1 as the template, the PCR reactions generated fragments encoding GFP linked to the 2A-like sequence (*Figure 23*).

The PCR fragments were doubly restricted with *BamHI* and *ApaI*, gel purified and ligated with the large fragment of the similarly restricted plasmid pSTA1 (*Figure 24*). The series of constructs produced in this way encoded the 2A-like sequence flanked by the reporter proteins, green fluorescent protein (GFP) and β-glucuronidase (GUS), maintaining a single open reading frame. The integrity of the constructs was analysed by DNA sequencing.

Picornaviruses:	
FMDA10	-QLLNFDLLKLAGDVESNPGP-
FMDA12	-QLLNFDLLKLAGDVESNPGP-
FMDC1	-QLLNFDLLKLAGDVESNPGP-
FMDO1K	-QLTNFDLLKLAGDVESNPGP-
FMDVO1G	-QLLNFDLLKLAGDVESNPGP-
FMDSAT3	-QMCNFDLLKLAGDVESNPGP-
FMDSAT2	-QLLNFDLLKLAGDVESNPGP-
ERAV	-QCTNYALLKLAGDVESNPGP-
ERBV	-GATNFSLLKLAGDVESNPGP-
PTV-1	-GATNFSLLKQAGDVESNPGP-
EMCB	-HYAGYFADLLIHDIETNPGP-
EMCD	-HYAGYFADLLIHDIETNPGP-
EMCPV21	-HYAGYFADLLIHDIETNPGP-
MENGO	-HYAGYFSDLLIHDVETNPGP-
TMEGD7	-YHADYYKQRLIHDVEMNPGP-
TMEDA	-YHADYYKQRLIHDVEMNPGP-
TMEBEAN	-YHADYYRQRLIHDVETNPGP-
Type C Rotaviruses:	
Human Rotavirus	-SKFQIDRILISGDIELNPGP-
Porcine Rotavirus	-AKFQIDRILISGDIELNPGP-
Bovine Rotavirus	-SKFQIDRILISGDIELNPGP-
Insect Viruses:	
DCV	-EAARQMLLLLSGDVETNPGP-
CrPV	-FLRKRTQLLMSGDVESNPGP-
ABPV	-GSWTDILLLLSGDVETNPGP-
TaV	-RAEGRGSLLTCGDVEENPGP-
IFV	-TRAEIEDELIRAGIESNPGP-
APV	-NYPMPEALQKIIDLESNPPPG-
Trypanosome Sequences:	
APendo	-CDAQRQKLLLSGDIEQNPGP-
TSR1	-SSIIRTKMLVSGDVEENPGP-
Bacterial Sequence:	
T. maritima aguA	-YIPDFGGFLVKADSEFNPGP-
Cellular Sequences:	
D. melanogaster Mod(mdg4)	-TAADKIEGSLKMDTEGNPGP-
R.catesbeiana *CPSaseI*	-MHETKPIFTSQFHPEANPGP-
Mouse MOR-1F	-DLELETVGSHQADAETNPGP-

Figure 22: 2A and 2A-like sequences. *The sequences identified from database searching are shown with the -DxExNPGP- motif shaded.*

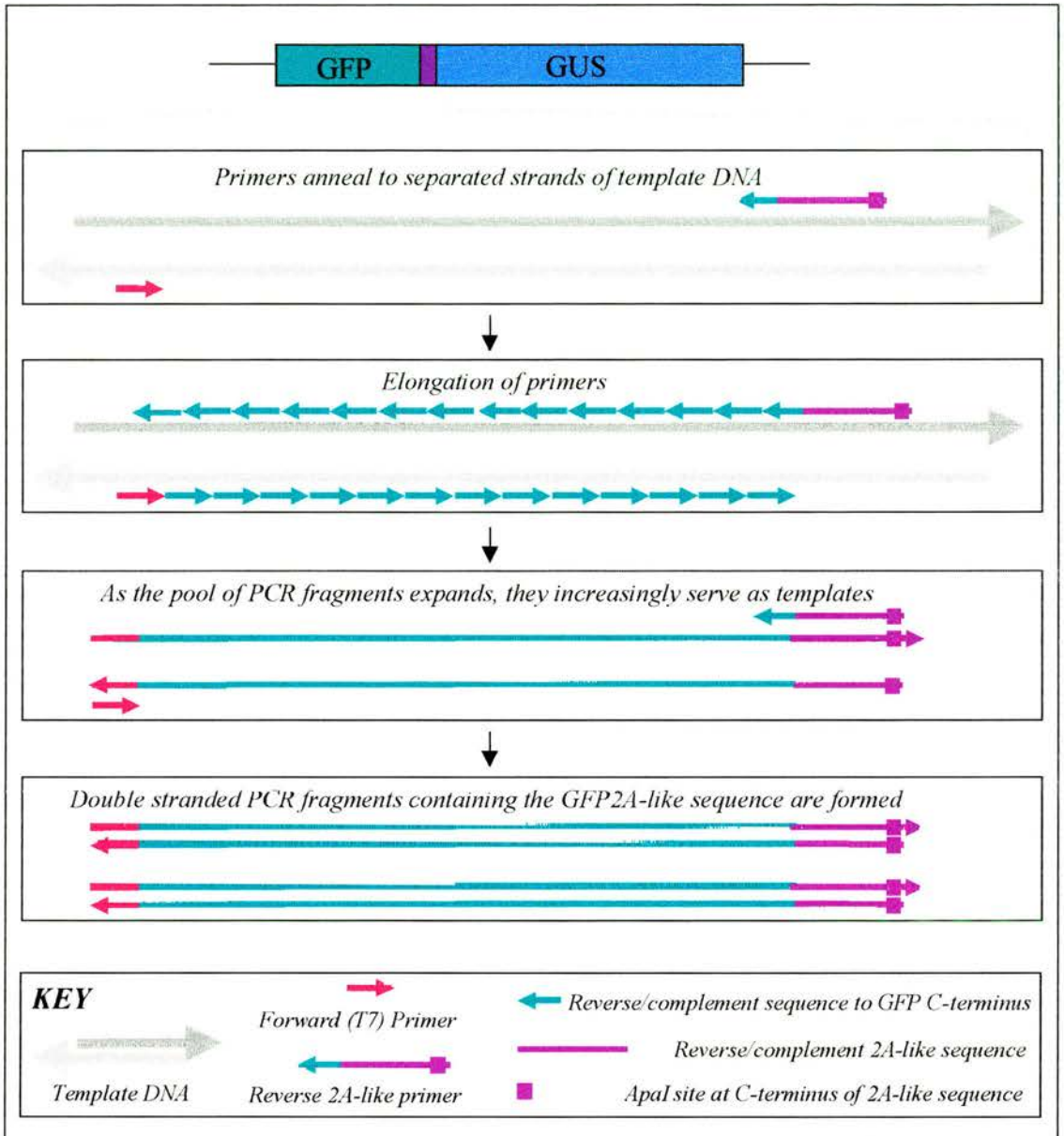


Figure 23: Schematic diagram of the technique used to generate GFP'2A-like' fragments. *The reverse primer was designed to build the 2A-like sequence on to the 3'-terminus of the GFP gene. Amplification of the GFP2A-like fragment was achieved by PCR.*

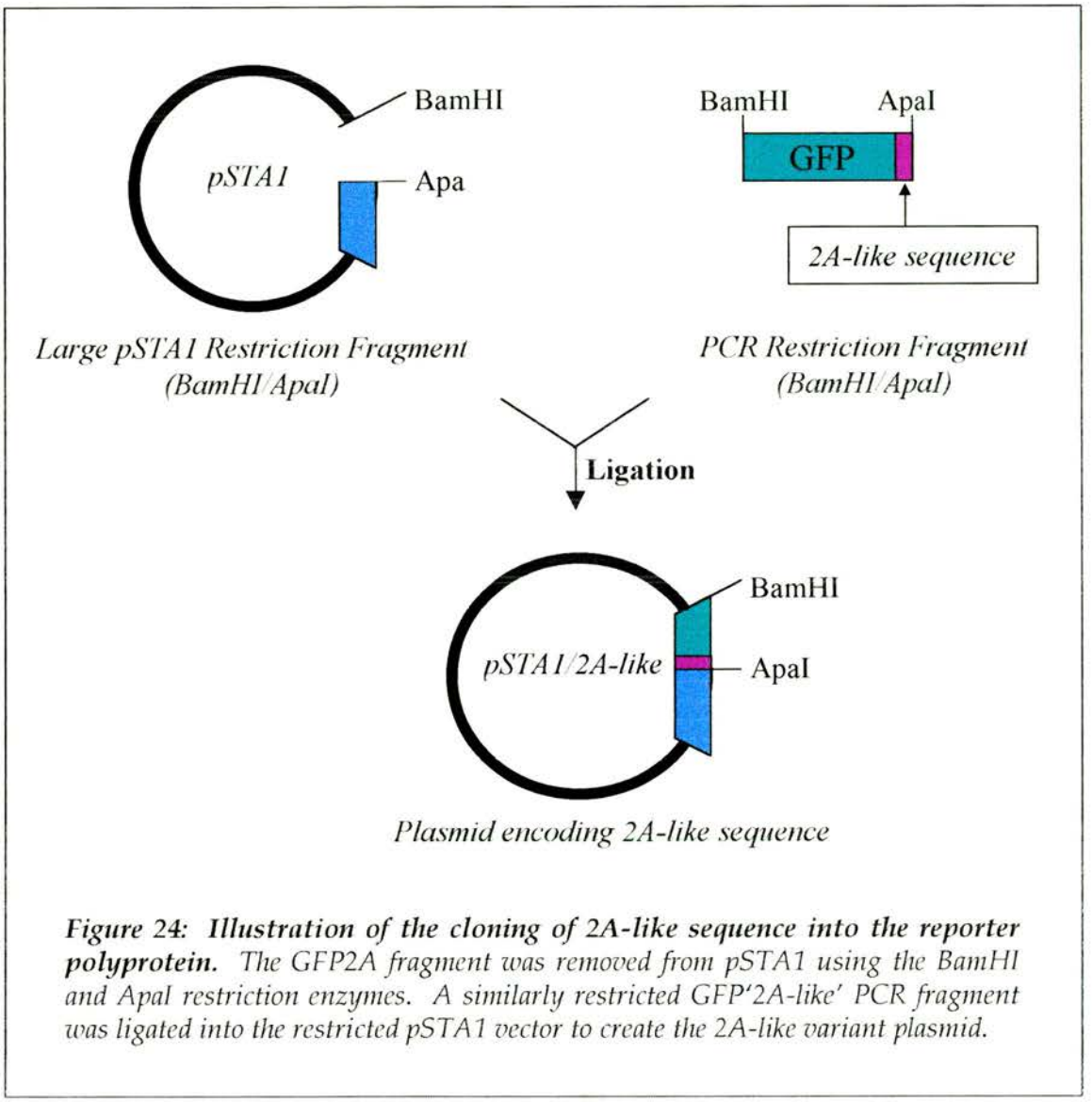


Figure 24: Illustration of the cloning of 2A-like sequence into the reporter polyprotein. *The GFP2A fragment was removed from pSTA1 using the BamHI and ApaI restriction enzymes. A similarly restricted GFP'2A-like' PCR fragment was ligated into the restricted pSTA1 vector to create the 2A-like variant plasmid.*

3.3.3. Analysis of naturally occurring 2A-like sequences.

The constructs containing the 2A and 2A-like coding sequences were used to program *in vitro* wheat germ extract coupled transcription/translation reactions. The translation products were separated by SDS-PAGE and the distribution of radiolabel visualised either by autoradiography or by phosphorimaging. Densitometric analyses were performed on the translation profiles in the manner described previously for pSTA1. To

allow comparative analysis of the activities of the 2A and 2A-like sequence, the data presented here is from a representative set of reactions using the same batch of wheat germ extract.

3.3.4. 2A and 2A-like Sequences of Mammalian Pathogens

Sequence searches using the identified 2A motif found a number of sequences belonging to the picornavirus group. Unsurprisingly the majority of these originated from strains of the aphthovirus, FMDV and the cardioviruses EMCV and TMEV. A sequence similar to FMDV 2A was also identified from equine rhinitis A virus (ERAV), formerly equine rhinovirus-1, which was recently renamed (ERAV) and reclassified to the *Aphthovirus* genus of the *Picornavirus* family (Wutz *et al.*, 1996). Another picornavirus possessing the DxExNPGP motif of FMDV 2A is the porcine teschovirus serotype 1 (PTV-1). This virus was previously named porcine enterovirus serotype 1, of the genus *Enterovirus* (Doherty *et al.*, 1999). It is now suggested that the porcine teschoviruses belong to a new genus named Teschovirus (Zell *et al.*, 2001). The final 2A-like sequence identified to have viral origin was that from the genome segment 6 of the group C rotaviruses. The DNA sequences cloned between the two reporter genes of GFP and GUS to analyse the processing activities of the mammalian viral 2A and 2A-like sequences are shown in table 3, together with the amino acids sequences which they encode.

FMDV	Q L L N F D L L K L A G D V E S N P G P cagctgttgaattttgaccttcttaagcttgcgggagacgtcgagtccaaccctgggccc
EMCV	H Y A G Y F A D L L I H D I E T N P G P cactacgctggttactttgcggacctactgattcatgacattgagacaaatcccgggccc
TMEV	Y H A D Y Y K Q R L I H D V E M N P G P taccatgctgactactacaaacagagactcatacatgatgtagaaatgaaccccgggccc
ERAV	Q C T N Y A L L K L A G D V E S N P G P cagtgtactaattatgctctcttgaaattggctggagatgttgagagcaaccccgggccc
PTV1	A T N F S L L K Q A G D V E E N P G P gccacgaacttctctctgttaaagcaagcaggagatgttgaagaaaaccccgggccc
PCRota	A K F Q I D K I L I S G D V E L N P G P gctaaattccaaatcgataaaattttaatttctggagacgtcgaattgaatcctgggccc
HCRota	S K F Q I D K I L I S G D I E L N P G P tcgaaatttcaaattgataaaattttaatttccggagatatcgaattaaatcccgggccc

Table 3: The cloned 2A and 2A-like sequences of the mammalian pathogens. *The amino acid sequences are shown together with the corresponding DNA sequence (5′-3′) cloned between the genes of the two reporter proteins GFP and GUS.*

Examination of the processing abilities of the 2A/2A-like sequences in wheat germ extract *in vitro* coupled transcription/translation reactions showed gel patterns similar to that of the FMDV 2A-containing pSTA1 construct (*Figure 25*). The full-length unprocessed material together with the N-terminal and C-terminal processing products were identified and quantitative analyses performed as described previously. It had previously been reported that the C-terminal region of the cardiovirus 2A proteins, together with the N-terminal residue of 2B, mediated cleavage in an artificial polyprotein system, with activity similar to that observed for the 2A region of foot-and-mouth disease virus (Donnelly *et al.*, 1997). The results presented here show that this region of EMCV is as active (~98%) in the [GFP2AGUS] system as FMDV 2A (~98%). The cardiovirus TMEV 2A protein showed a slightly lower level of activity (~91%) than FMDV 2A. The newly classified aphthovirus, ERAV, has a 2A sequence which is highly similar to that of FMDV (Wutz et al., 1996) and unsurprisingly showed a high

level of activity in this artificial polyprotein system (>99%). Equine rhinitis B virus, formerly equine rhinovirus-2, is the sole member of the new *Erbovirus* genus and has a 2A sequence similar to both ERAV and FMDV (Wutz *et al.*, 1996). ERBV also shows high sequence similarity in the 2A region to the porcine teschovirus-1. The Teschovirus polyprotein has a similar organization to that of the erboviruses and aphthoviruses. The PTV-1 sequence analyzed showed a similar high level of activity (~99%) to that of FMDV, EMCV and ERAV.

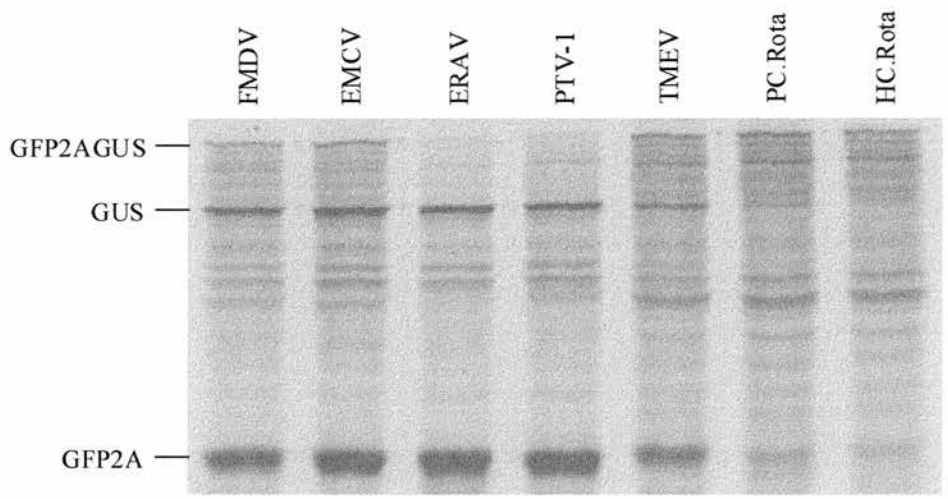


Figure 25: Translation profiles of constructs containing 2A/2A-like sequences from mammalian pathogens. *Coupled transcription/translation wheat germ extract reactions were programmed with constructs encoding the 2A/2A-like sequences in the GFP-2A-GUS artificial polyprotein system. The lanes are labelled with the origins of the 2A/2A-like sequences. The full-length translation products (GFP2AGUS), and processing products (GFP2A and GUS) are indicated.*

The 2A-like sequences from the human, porcine and bovine type C rotaviruses are present in the genome segment 6 which encodes the non-structural protein NSP3 (NS34). This sequence is highly conserved amoung the group C rotaviruses (Qian *et al.*, 1991; James *et al.*, 1999). The human (HCRota) and porcine (PCRota) 2A-like sequences tested here show activities (~50% and ~36%, respectively) much lower than those observed for the picornavirus-derived sequences. It should however be mentioned at this stage that the length of the rotavirus 2A-like sequences tested may have been too short and increasing the length of sequence cloned between the reporter proteins may lead to increased levels of processing. This phenomenon was indeed observed when the length of the FMDV 2A sequence was increased to include sequences from the C-terminus of 1D and will be discussed later.

Plasmid	Processing Activity (%)	Ratio TP1:TP2	Proportion of Synthesis (%): Full-length	TP1&TP2	TP1 only
pSTA1-FMDV	~98	4.33	3	22	75
pSTA1-EMCV	~98	4.24	3	23	74
pSTA1-TMEV	~91	3.71	12	24	64
pSTA1-ERAV	>99	5.05	0	20	80
pSTA1-PTV-1	~99	4.11	1	24	75
pSTA1-PCRota	~36	6.35	68	5	27
pSTA1-HCRota	~50	7.71	53	6	41

Table 4: Analyses of the translation profiles generated by the 2A-like sequences of mammalian pathogens. *Processing activity, molar excess of TP1 and the proportion of ribosomes synthesizing each of the three translational outcomes were calculated using the PSLcorr values obtained from densitometric analysis.*

Further analysis of the translation profiles showed that for all of the viral 2A and 2A-like sequences tested, the first translation product (TP1) was in excess of the second (TP2). The greatest excess was measured for the human group C rotavirus and the least imbalance occurred in the translation profile of TMEV (*Table 4*). When these results were compared to the values calculated for FMDV there was no evident correlation between the processing activity and the imbalance. The FMDV sequence showed higher processing activity than both HCRota and TMEV. However, the ratio of translation product 1 (TP1) to translation product 2 (TP2), for FMDV is higher than that of TMEV but lower than that of HCRota.

The viral 2A-like sequences mediate three different translational outcomes; i) complete synthesis of a full-length polyprotein, ii) synthesis of two discrete translation products and iii) synthesis of TP1 only. In an attempt to quantify these activities they were expressed as a percentage *(Table 4)*. The sequences shown to have good processing activities (FMDV, EMCV, TMEV, ERAV and PTV-1) are those which prevent the majority of ribosomes from synthesising a full-length polyprotein product. Analysis of the translation profiles show that more ribosomes terminate after the synthesis of the GFP2A moiety, with a lesser amount continuing to synthesize the discrete GUS product. The poor cleaving group C rotavirus sequences permitted the ribosomes to mostly synthesize full-length product or GFP2A, with only a small proportion synthesizing the two discrete products.

3.3.5. Insect Virus 2A-like Sequences.

Several insect viruses were identified to have 2A-like sequences within their genomes. Many of the insect virus 2A-like sequences revealed by database searches were from

viruses previously thought to be picorna-like viruses. *Drosophila* C virus (DrosC), cricket paralysis virus (CrPV) and acute bee paralysis (ABPV) are now thought to have a genome arrangement that is somewhat different from the picornaviruses (Johnson & Christian, 1998; Wilson *et al.*, 2000a; Govan *et al.*, 2000). These viruses appear to have monopartite bicistronic genomes with the replicase proteins encoded in the 5' region of the genome and the capsid proteins in the 3' region. Of the insect viral 2A-like sequences tested here, only infectious flacherie virus (IFV) is thought to have a genome arrangement similar to that of FMDV (Isawa *et al.*, 1998). Interestingly the 2A-like sequence deviates from the DxExNPGP motif, with a glycine residue in place of the aspartate (*Table 5*).

DrosC	A A R Q M L L L L S G D V E T N P G P gctgcacgtcagatgttgctcttgttatcaggagatgttgagactaaccctgggccc
CrPV	F L R K R T Q L L M S G D V E S N P G P ttcttgagaaagagaacgcaacttttgatgagtggtgatgtcgaatctaatcccgggccc
ABPV	G S W T D I L L L L S G D V E T N P G P ggctcatggactgacatattgttgttgttgtctggagatgtagaaaccaatcccgggccc
TaV	R A E G R G S L L T C G D V E E N P G P agagccgagggcagggggaagtcttctaacatgcggggacgtggaggaaaatcccgggccc
IFV	T R A E I E D E L I R A G I E S N P G P acgagggcggagattgaggatgaattgattcgtcgaggaattgaatcaaatcctgggccc
IFV(D)*	T R A E I E D E L I R A D I E S N P G P acgagggcggagattgaggatgaattgattcgtcgagacattgaatcaaatcctgggccc
APV	N T P M P E A L Q K I I D L E S N P P P G aattatcctatgcctgaggcattacaaaaaattatagacttagagtcaaatccccctcccggg

Table 5: The cloned 2A-like sequences of insect viruses. *The amino acid sequences are shown together with the corresponding DNA sequence (5'-3') cloned between the genes of the two reporter proteins GFP and GUS.*

All of the viral 2A-like sequences tested showed cleavage activity apart from a mutated sequence from infectious flacherie virus (IFV(D)*), in which the wild-type –GxExNPGP- sequence was mutated to the –DxExNPGP- motif, and the variant sequence from Acyrthosiphon pisum virus (APV). The inactive 2A-like sequences were characterized in the gel profiles by a single major band corresponding in size to the full-length unprocessed polyprotein (*Figure 26*).

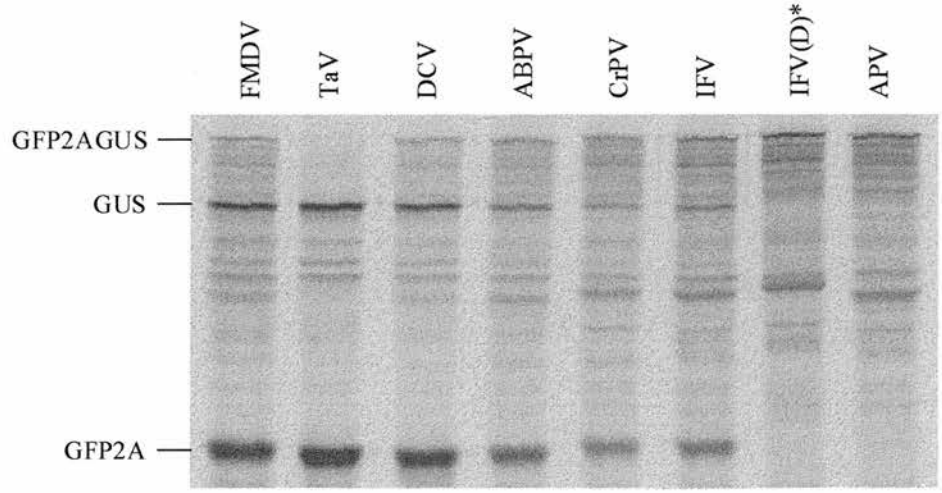


Figure 26: Translation profiles of constructs containing 2A/2A-like sequences from insect viruses. *Coupled transcription/translation wheat germ extract reactions were programmed with constructs encoding the 2A/2A-like sequences in the GFP-2A-GUS artificial polyprotein system. The lanes are labelled with the origins of the 2A/2A-like sequences, the mutated IFV sequences is marked with an asterisk. The full-length translation products (GFP2AGUS), and processing products (GFP2A and GUS) are indicated. The processing products are absent in the lanes IFV(D)* and APV, indicating inactive 2A-like sequences.*

Of the five insect virus sequences which mediated cleavage, that from Thosea asigna virus (TaV) showed the highest cleavage activity (>99%). Similar to the profile of ERAV, full-length polyprotein was barely detectable in the TaV translation profile. The

2A-like sequences from *Drosophila* C virus (DCV) and acute bee paralysis virus (ABPV) exhibited high cleavage activity (~98% and ~94% respectively), the activity of DCV being comparable to that of FMDV. The activities of the 2A-like sequences from cricket paralysis virus (CrPV) and infectious flacherie virus (IFV) had lower levels of activity than that observed for FMDV (~88% and ~87% respectively).

Plasmid	Processing Activity (%)	Ratio TP1:TP2	Proportion of Synthesis (%): Full-length	TP1&TP2	TP1 only
pSTA1-DCV	~98	4.44	3	22	75
pSTA1-CrPV	~88	3.96	14	22	64
pSTA1-ABPV	~94	6.13	7	15	78
pSTA1-TaV	>99	5.10	0	20	80
pSTA1-IFV	~87	3.99	15	21	64

Table 6: Analyses of the translational profiles from the insect virus 2A-like constructs. *Processing activity, molar excess of TP1 and the proportion of ribosomes synthesizing each of the three translational outcomes were calculated using the PSLcorr values obtained from densitometric analysis.*

Analysis of the translation profiles again highlighted the imbalance in the GFP2A and GUS translation products. The greatest imbalance was observed in the profile of ABPV (*Table 6*), where the lowest proportion of ribosomes continued to synthesize GUS after the release of GFP2A. Interestingly, the proportion of ribosomes which synthesize both TP1 and TP2 is the same for both DCV and CrPV, however, with the DCV sequence, a higher proportion terminate after GFP2A compared to more full-length product with CrPV.

3.3.6. Bacterial, Trypanosomal and Cellular 2A-like Sequences.

The remaining 2A-like sequences found in the databases come from a range of diverse sources. The cellular sequences were found in the carbamyl-phosphate synthetase I enzyme of the North American Bullfrog, *Rana catesbeiana* (Helbing & Atkinson, 1994), the Mod(mdg4) 59.0 modifier protein of *Drosophila melanogaster* (Buchner *et al.*, 2000) and the mouse mu opioid receptor variant F (MOR-1F) (Pan *et al.*, 2000). These were labelled CPS, Drosmel and Mus respectively. The bacterial 2A-like sequence, Therm, was found in the α-glucuronidase enzyme of the hyper-thermophilic bacterium, *Thermotoga maritima (*Ruile *et al.*, 1997*)*.

A further two sequences were identified by probing the databases. These 2A-like regions are present in sequences of trypanosome origin. The 2A-like sequence identified in *Trypanosoma cruzi* occurred in ORF1 of the non-LTR retrotransposon L1Tc (Martin *et al.*,1995). This ORF encodes an AP endonuclease-like sequence (APendo). The *Trypanosoma brucei* 2A-like sequence is located within the TSR-1 repeated sequence (Murphy *et al.*, 1987).

Drosmel	T A A D K I E G S L K M D T E G N P G P acagctgccgacaaaatcgagggatcctggaaaatggacaccgagggcaaccccgggccc
Mus	D L E L E T V G S H Q A D A E T N P G P gatctggaattggagacagttgggagccaccaggcagatgctgaaaccaaccccgggccc
CPS	M H E T K P I F T S Q F H P E A N P G P atgcatgagaccaaaccaatcttcacttcgcaattccaccctgaggctaaccccgggccc
TSR1	S S I I R T K M L V S G D V E E N P G P agcagtatcatccgcactaagatgctggtgtccggtgatgtggaagagaatcccgggccc
TSR1(L)*	S S I I R T K M L L S G D V E E N P G P agcagtatcatccgcactaagatgctgctgtccggtgatgtggaagagaatcccgggccc
APendo	C D A Q R Q K L L L S G D I E Q N P G P tgtgacgcgcaacgacaaaagctactgctaagcggagacattgagcagaacccagggccc
Apendo(A)*	C D A Q R Q K L L L S A D I E Q N P G P tgtgacgcgcaacgacaaaagctactgctaagcgcagacattgagcagaacccagggccc
Therm	Y I P D F G G F L V K A D S E F N P G P tatattccagattttggaggatttcttgtcaaagccgattctgagttcaatcctgggccc

Table 7: The cellular, bacterial and trypanosomal 2A-like sequences cloned into the GFP-GUS reporter polyprotein system. *The amino acid sequences are shown together with their DNA counterparts. The mutated versions of the trypanosomal sequences, which occurred during cloning and were identified by DNA sequencing, are highlighted with an asterisk.*

The 2A-like sequences cloned into the GFP-2A-GUS artificial polyprotein system are shown in table 7. The cloning of the trypanosome 2A-like sequences into the reporter system produced mutated versions of both the AP endonuclease-like (Apendo(A)*) and the TSR1 (TSR1(L)*) sequences. All of the constructs were used to programme couple transcription/translation wheat germ extract reaction and the translation profiles of the mutated sequences were examined in tandem with the database-derived sequences (*Figure 27*).

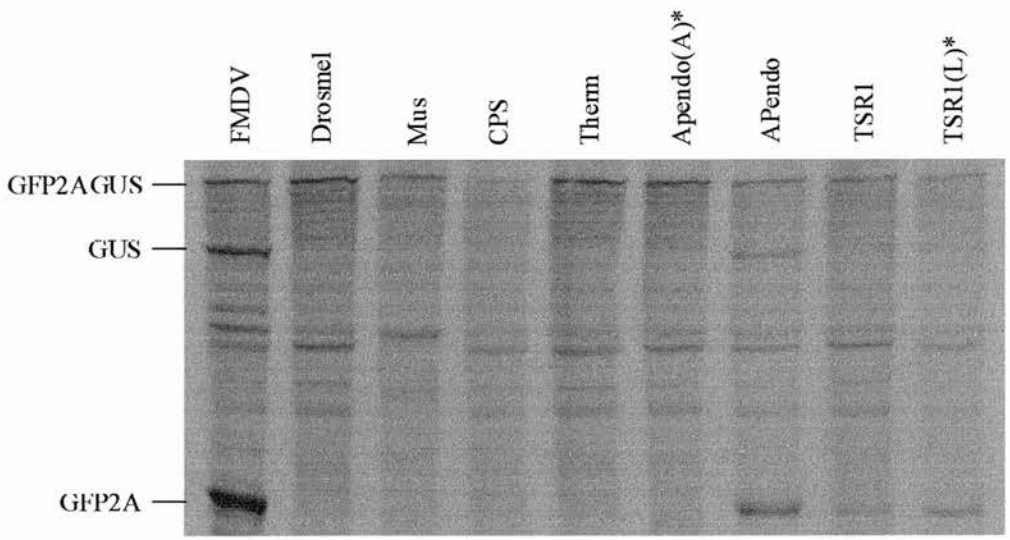


Figure 27: Translation profiles of constructs containing 2A-like sequences of bacterial, cellular or trypanosomal origin. *Coupled transcription/translation wheat germ extract reactions were programmed with constructs encoding the 2A-like sequences in the GFP-2A-GUS artificial polyprotein system. The FMDV profile is included for referance. The lanes are labelled with the origins of the 2A-like sequences and the mutated sequences are marked with asterisks. The full-length translation products (GFP2AGUS) and processing products (GFP2A and GUS) are indicated. The processing products are absent in the lanes Drosmel, Mus, CPS, Therm and APendo(A)* indicating these 2A-like sequences are inactive.*

The cellular and bacterial sequences (CPS, Drosmel, Mus and Therm) were found to be inactive. The translation profiles of these sequences are characterised by full-length polyprotein (GFP2AGUS), with no processing products (GFP2A and GUS) being detected. The trypanosomal sequences did however show some processing activity. The APendo sequence exhibited a far greater level of activity (~91%) than the TSR1

sequence (~20%). The mutated APendo(A)* sequence was inactive, while the mutated TSR1(L)* sequence was more active (~44%) than the genuine TSR1 2A-like sequence.

Plasmid	Processing Activity (%)	Ratio TP1:TP2	Proportion of Synthesis (%): Full-length	TP1&TP2	TP1 only
pSTA1-APendo	~91	5.28	10	17	73
pSTA1-TSR1	~20	2.80	84	6	10
pSTA1-TSR1(L)*	~44	2.99	63	12	24

Table 8: Analyses of the translational profiles produced by the active trypanosomal 2A-like sequences. *Processing activity, molar excess of TP1 and the proportion of ribosomes synthesizing each of the three translational outcomes were calculated using the PSLcorr values obtained from densitometric analysis.*

The analysis of the translational profiles produced by the trypanosmal 2A-like sequences (*Table 8*) indicates that the major product of the Apendo sequence is GFP2A with only a small fraction of ribosomes continuing to synthesize the GUS portion as a discrete product. The poor cleaving TSR1 sequences mostly synthesize the full-length polyprotein with very low levels of GFP2A and barely detectable levels of GUS. These trypanosomal 2A-like sequences have differing processing abilities achieving varying translational outcomes.

3.4. *Analysis of Theiler's Murine Encephalitis Virus Polyprotein Processing*

As an extension to the studies performed *in vitro* to examine the processing event that occurs at the C-terminus of the 2A region in aphthoviruses, teschoviruses, erboviruses and cardioviruses it was intended to investigate the processing events of TMEV polyprotein in BHK-21 cells. The plan was to infect the cells with the TME virus, label the protein products and analyse at hourly intervals. It was believed that this would allow the detection of any shift in protein expression, particularly whether the synthesis of capsid proteins was favoured, over the synthesis of replicative proteins, later in infection.

3.4.1. Can Picornaviruses Synthesise an Excess of Capsid Proteins?

With the knowledge that the capsids of these viruses are formed from 60 copies of each of the four structural proteins the notion that perhaps, somehow, these viruses could at some stage in infection synthesise just capsid proteins was particularly exciting. The conservation of resources after the shut-off of host cells expression would appear to be more important later in infection when materials, such as amino acyl-tRNAs, are becoming depleted. Analogous to the expression of early and late genes in more complex viruses, an ability to focus synthesis on the structural proteins without the seemingly wasteful co-synthesis of several copies of enzymatic replicative proteins may be beneficial to these picornaviruses.

3.4.2. Possible Role of 2A in Differential Expression of Structural and Non-structural Proteins.

The position of the 2A region in the genomes of the aphtho-, tescho-, erbo- and cardioviruses is at the junction of the capsid and non-capsid proteins (*Figure 28*). This location together with the fact that processing was known to occur at the C-terminus of

the 2A region, in aphtho- and cardioviruses suggested the 2A/2B junction is the perfect site for the control of differential expression of structural and non-structural proteins to take place. As our translational model for 2A-mediated processing developed and it became evident that in *in vitro* expression systems the synthesis of the protein upstream of the 2A-processing site was in excess of the protein downstream, more support for a translational control of expression emerged.

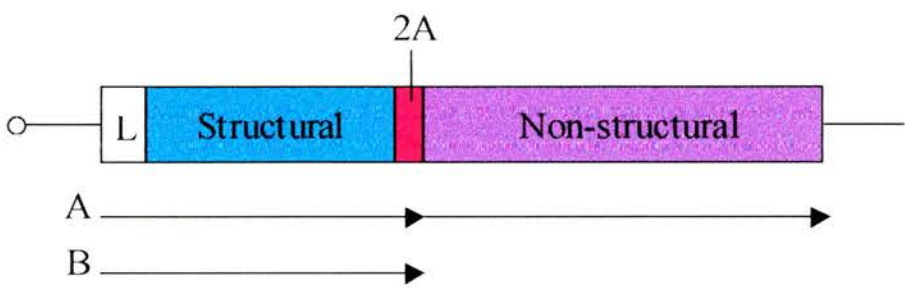


Figure 28: Location of the 2A region within the genomes of the aphth-, erbo-, tescho- and cardioviruses. *The translational outcomes determined by in vitro analysis of 2A activity are shown. Translational control at the 2A processing site may allow a switch in expression from outcome A to B, reminiscent of early and late gene expression employed by other viruses. Such a control of expression could allow the structural proteins to be produced in excess of the non-structural proteins.*

3.4.3. Incomplete 'Shut-Off' of Host Protein Expression by TMEV.

Analysis of the expression of TMEV viral proteins by the method described above was however complicated by an incomplete 'shut-off' of host cell protein synthesis. The labelling technique led to all proteins, both viral and cellular, being labelled and as a result the gel profiles obtained from SDS-PAGE contained many bands that were difficult to separate (*Figure 29*). Ultimately it was not possible to identify all the processing products of the viral polyprotein, from the vast pool of proteins present in the

gel profiles, simply on a size basis. This work was not pursued further due to the constraints of time.

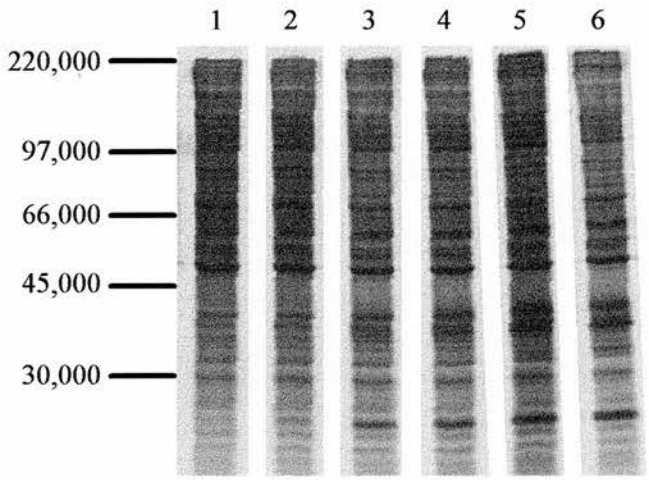


Figure 29: Gel profiles of the labelled proteins from TMEV infected cells. *The lanes are labelled 1-6 corresponding to the hours post-infection at which analysis was performed. The positions of the size markers (kDa) are shown on the left of the figure.*

3.5. The Utilization of FMDV 2A in the Expression of Multiple Proteins

The very short length of the functional FMDV 2A region together with the autonomous nature of the processing event that it mediates at its own C-terminus renders it a useful tool for a variety of biotechnological purposes. It has been known for some time that the 2A region of FMDV can function in an artificial polyprotein system devoid of all other viral sequences (Ryan & Drew, 1994). It has since been demonstrated that FMDV 2A sequences can be utilized for expression of multiple introduced proteins in plants (Halpin *et al.*, 1999) and for the development of retroviral vectors for gene therapy (de Felipe *et al.*, 1999). The 2A-like sequences identified here contribute to this toolbox of self-processing proteins. The role that the 2A/2A-like sequence is required to fulfill will govern which sequence is selected. The amount of full-length polyprotein synthesized and the ratio of the two processed products are important determinants in this selection process.

3.5.1 Examination of 2A activity in cellular expression systems.

In order to assess the processing activity of FMDV 2A in a cellular system a different artificial polyprotein organization, which would allow the processing products to be purified identically from the cell lysate, was required. The plasmid pPF19, constructed by Dr. Pablo de Felipe, was deemed to be suitable. This plasmid encodes the YFP-2A-GAL-CFP-2A-PAC fusion protein, where YFP is yellow fluorescent protein, CFP is cyan fluorescent and PAC confers puromycin resistance. In this polyprotein the CFP has an N-terminal signal sequence, GAL, which targets it to the golgi. Complete processing of the polyprotein, by the two FMDV 2A regions, should yield three translation products (YFP-2A, GAL-CFP-2A and PAC) significantly different in size to be easily separated by SDS-PAGE. Processing and targeting of the cyan fluorescent protein to the golgi and

the presence of yellow fluorescent in the cytoplasm was confirmed by fluorescent mapping of transfected *Hela* cells. This plasmid was chosen to analyse the processing activity of 2A *in vivo* because both the YFP-2A and GAL-CFP-2A products should be recognised by antibodies against FMDV 2A or GFP (since the sequences of GFP, CFP and YFP are highly similar). Purification of radiolabelled translation products by immunoprecipitation and subsequent gel separation would allow the processing products to be quantified by densitometric analysis and 2A activity to be characterised in the same manner applied to the *in vitro* reactions.

3.5.2. Detection of 2A processing activity in transfected cells.

Radiolabelled cell extracts were prepared from Hela cells transfected with the constructs, pSTA1, pSTA1/31 (Donnelly et al., 2001b) and pPF19. Immunoprecipitation reactions were performed with the anti-GFP antibody and the samples separated by 12% SDS-PAGE. Only the extract prepared from the pPF19 transfected cells yielded a positive result from the immunoprecipitation reactions (*Figure 30*). The three major bands identified in this profile correspond in size to the full-length fusion polyprotein (YFP-2A-GAL-CFP-2A-PAC) and two fusion products that arise as a result of incomplete processing of the polyprotein (YFP-2A-GAL-CFP-2A and GAL-CFP-2A-PAC). The three smaller translation products, YFP-2A, GAL-CFP-2A and PAC were not detected using this method.

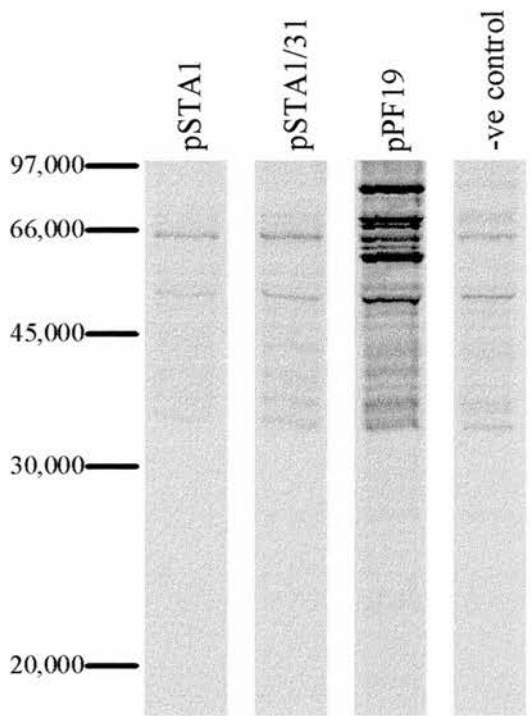


Figure 30: Gel profiles of anti-GFP immunoprecipitation reactions performed on transfected cell extracts. *The lanes are labelled with the constructs that the cells were transfected with and the positions of the protein size markers are indicated. The three major bands identified in the pPF19 profile correspond in size to the full-length fusion polyprotein (YFP-2A-GAL-CFP-2A-PAC) and two fusion products that arise as a result of incomplete processing of the polyprotein (YFP-2A-GAL-CFP-2A and GAL-CFP-2A-PAC).*

3.5.3. Efficiency of antibody-based purification of FMDV 2A translation product.

Following the results of the cellular expression system it was necessary to determine the efficiency of the anti-GFP antibody in purifying the processed translation products of the pSTA1, pSTA1/31 and pPF19 constructs. Initially, to demonstrate that 2A-mediated processing actually occurs in these constructs, they were used to programme wheat germ extract coupled transcription/translation reactions. The radiolabelled translation products were then subjected to immunoprecipitation with the anti-GFP antibody utilised in the cell extract reactions or an anti-FMDV 2A antibody. The gel profile of these reactions

shows that 2A-mediated processing occurs when the proteins of these constructs are expressed *in vitro* (*Figure 31*).

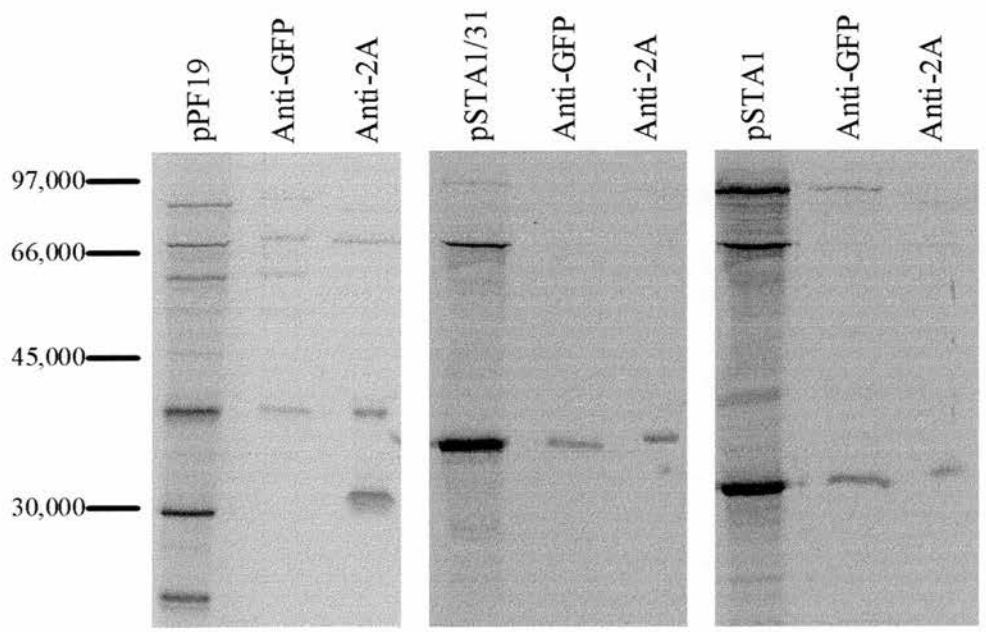
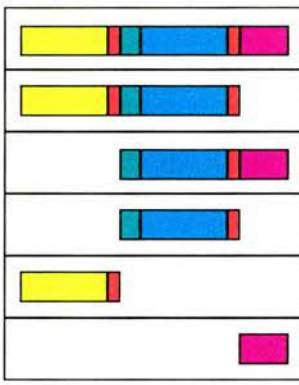


Figure 31: Gel profiles of in vitro coupled transcription/translation reactions and subsequent immunoprecipitation reactions. *The lanes are labelled with the constructs used to programme the transcription/translation reactions and the antibodies used in the immunoprecipitation reactions. The positions of the protein size markers are indicated.*

In the case of pSTA1, both the anti-2A and anti-GFP antibodies bind both the full-length polyprotein and the GFP-2A moiety. The pSTA1/31 construct which encodes 39 amino acids of FMDV 1D at the N-terminus of 2A in the GFP-Δ1D-2A-GUS polyprotein mediates a higher level of processing than pSTA1 and as a result only a very faint band representing the full-length polyprotein is detected in the *in vitro* reaction. The anti-GFP and anti-2A immunoprecipitation reactions detected the GFP-Δ1D-2A translation product of pSTA1/31. The gel profile obtained from the *in vitro* reaction programmed with pPF19 showed bands corresponding in size to all the possible translation products, full-length YFP-2A-GAL-CFP-2A-PAC, partially processed YFP-2A-GAL-CFP-2A and GAL-CFP-2A-PAC and the fully processed GAL-CFP-2A, YFP-2A and PAC. The anti-

GFP antibody purified only those products containing the CFP sequence (YFP-2A-GAL-CFP-2A-PAC, YFP-2A-GAL-CFP-2A, GAL-CFP-2A-PAC and GAL-CFP-2A), it was not efficient in binding the YFP-2A product (*summarised in figure 32*).



	kDa	Anti-GFP	Anti-2A
YFP-2A-GAL-CFP-2A-PAC	92	✓	×
YFP-2A-GAL-CFP-2A	69	✓	✓
GAL-CFP-2A-PAC	60	✓	×
GAL-CFP-2A	39	✓	✓
YFP-2A	30	×	✓
PAC	21	×	×

Figure 32: Processing products of the construct pPF19, encoding the polyprotein YFP-2A-GAL-CFP-2A-PAC. *The processing products are listed in order of size; the weights of the proteins are shown to the nearest kDa. The abilities of the anti-GFP and anti-2A antibodies to immunoprecipitate each product are indicated, where ✓ represents the presence and × the absence of a band in the gel profile.*

Similar to the immunoprecipitation reaction performed with the cell extract, immunoprecipitation of the *in vitro* reaction programmed with pPF19 yielded the three large translation products. Additionally, the *in vitro* system also precipitated the GAL-CFP-2A processing product. Immunoprecipitation of the *in vitro* translation products with the anti-2A antibody yielded bands on the gel corresponding to YFP-2A-GAL-CFP-2A, GAL-CFP-2A and YFP-2A. The translation products that do not have a C-terminal 2A sequence (YFP-2A-GAL-CFP-2A-PAC, GAL-CFP-2A-PAC and PAC) were not detected in this gel profile. This system does not seem to be as useful a technique, for the characterisation of FMDV 2A activity in a cellular environment, as was initially predicted.

3.6. Characterisation of FMDV 2A Activity in a Yeast Artificial Polyprotein System

In an attempt to determine whether FMDV 2A is active in a yeast expression system and to examine the processing mechanism, several artificial polyprotein constructs were made. These constructs addressed the question of 2A functioning in *cis* within the ribosome or in *trans* requiring cytosolic factors by utilising two pathways of protein targeting to the endoplasmic reticulum (ER) that exist in yeast.

3.6.1. SRP-independent Translocation Pathway.

The translocation pathway that is independent of the cytosolic ribonucleoprotein signal recognition particle (SRP) occurs post-translationally and relies on cytosolic chaperones. The targeted protein is exposed to the cytosol prior to being translocated into the ER lumen.

3.6.2. SRP-dependent Translocation Pathway.

In the SRP-dependent route the protein is translocated into the ER co-translationally without being released from the ribosome into the cytosolic environment. This SRP-dependent route requires the cytosolic ribonucleoprotein signal recognition particle (SRP) which recognises a signal sequence at the N-terminus of the nascent ER-targeted proteins as they emerge from the ribosome and slows translation by elongation arrest. The SRP, associated with the ribosome complex, targets it to the ER membrane. On the ER membrane the SRP interacts with its receptor and the ribosome is docked on the translocation apparatus, the translocon. Translation proceeds and the nascent chain is protected from the cytosol first by the ribosome and then by the translocon.

3.6.3. Predicted Activity of FMDV 2A within a Ribosome-Translocon Complex.

Considering that FMDV 2A is predicted to function within the ribosome during translation any 2A-containing polyprotein targeted to the ER lumen by the SRP-dependent route should be processed before complete translocation takes place. For example, in an artificial polyprotein system where 2A is flanked by a protein possessing an SRP-dependent signal sequence and a downstream reporter protein, only the first translation product would be expected to be translocated into the lumen of the ER. The second translation product, the reporter protein, would be expected to remain cytosolic as a result of the 2A-mediated interruption in the nascent peptide chain.

3.6.4. Constructs Designed to Analyse the Translocation of 2A-containing Artificial Polyproteins.

The artificial polyprotein system designed to analyse 2A activity within ribosome-translocon complexes utilizes the SRP-dependent chimeric protein $D_N\alpha F$ and the reporter protein GFP. $D_N\alpha F$ encodes yeast proalpha factor with the native signal sequence replaced by the SRP-dependent signal sequence of dipeptidyl aminopeptidase B (Dap2p). The reporter protein has the ER retention motif appended to its C-terminus so that GFP remains in the lumen if translocated. As a control a separate construct was formed with the native preproalpha factor (ppαF) as the N-terminal protein. PpαF is targeted to the ER by SRP-independent route, and if 2A requires cytosolic factors to mediate processing then only this construct would form the two discrete translation products in the yeast expression system.

3.6.5. Cloning of the Yeast Artificial Polyproteins.

The yeast studies were carried out in collaboration with Dr. Jeremy Brown and he provided PCR products of the desired genes. Initially these PCR products were cloned into the pGEM-T Easy vector (Promega). The pGUS2AGFP construct was used as the foundation on which to build the yeast artificial polyprotein constructs prior to their transfer into the yeast expression vector. The cloning strategy for the construction of the yeast artifical polyproteins began with removing the GFP portion from pGUS2AGFP with the restriction enzymes *ApaI* and *PstI*. Intermediate constructs were formed by inserting either GFP with a C-terminal ER-retention motif (HEDL) (GFPh) or this same GFP fragment with the Kar2p ER-targeting motif at its N-terminus (Kar2GFPh). These constructs, pGUS-2A-GFPh and pGUS-2A-Kar2GFPh, were subsequently restricted with the *BamHI* and *XbaI* restriction enzymes, removing the GUS portion. Fragments encoding the previously described $D_N\alpha F$ or ppαF proteins, similarly restricted, were inserted in place of the GUS region forming the following series of constructs: p$D_N\alpha$F-2A-GFPh, p$D_N\alpha$F-2A-Kar2GFPh, pPpαF-2A-GFPh and pPpαF-2A-Kar2GFPh (*Figure 33*).

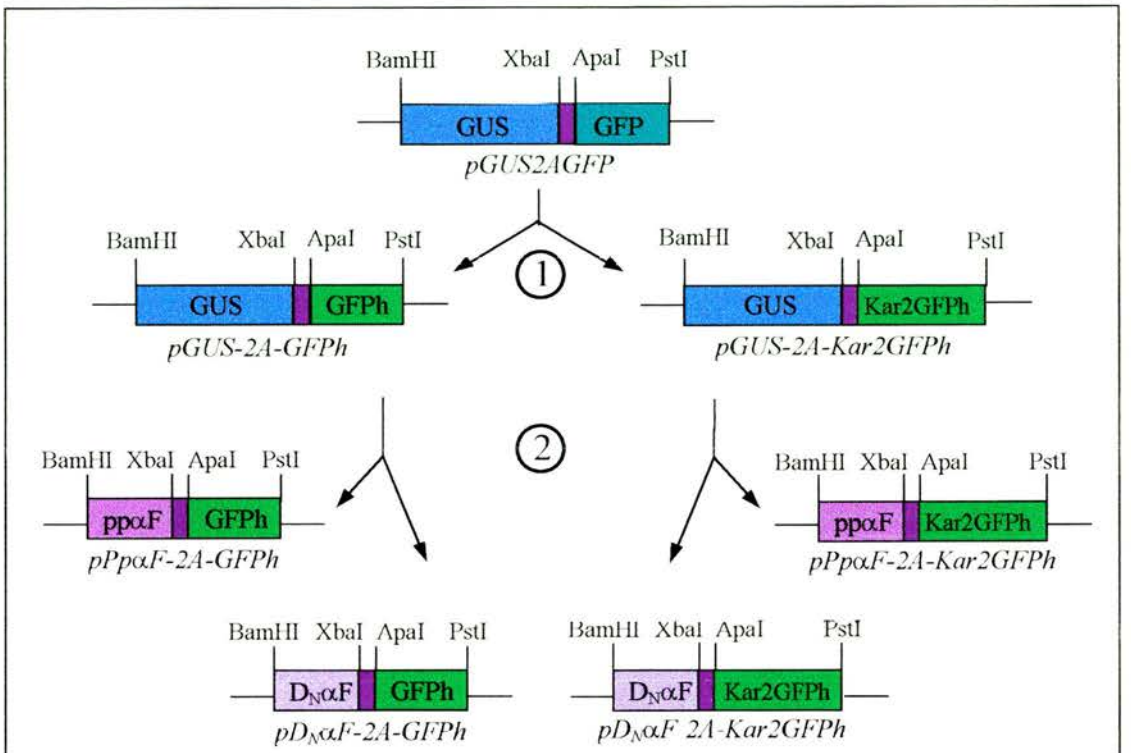


Figure 33: Schematic diagram of the cloning strategy employed to construct the FMDV 2A-containing yeast artificial polyproteins. *Using pGUS2AGFP (Donnelly et al., 2001b) as the backbone, firstly the GFP region was removed with the ApaI and PstI restriction enzymes. Similarly restricted fragments encoding GFP with a C-terminal ER-retention motif alone (GFPh) or GFP with the ER-retention motif and an N-terminal Kar2p ER-targeting sequence (Kar2GFPh) were ligated in place of the excised GFP (**1**). The GUS regions of these constructs were excised with BamHI and XbaI and replaced similarly restricted fragments encoding the yeast alpha factor. Two variants of the yeast alpha factor were utilised ($D_N\alpha F$ and ppαF), differing in the targeting signal sequence (see text for details).*

The series of constructs encoding the yeast artificial polyproteins were analysed in wheat germ extract reactions to ensure that FMDV 2A is indeed active in each of the these contexts (*Figure 34*). This was found to be the case and the entire polyproteins were then transferred into a yeast expression vector. Dr. Jeremy Brown undertook the yeast expression work and a summary of the results is presented here. FMDV 2A was shown

to be active *in vivo* in yeast. Cells transformed with the $pD_N\alpha F$-2A-GFPh and $pPp\alpha F$-2A-GFPh constructs were pulse labelled and the 2A- and GFP-containing products isolated from the cell lysates by immunoprecipitation reactions. The N- and C-terminal processing products together with some full-length polyproteins were identified (*Figure 35*). The full-length polyproteins ($D_N\alpha F$-2A-GFPh and $pp\alpha F$-2A-GFPh) and N-terminal processing products ($D_N\alpha F$-2A and $pp\alpha F$-2A) were shown to be translocated into the lumen of the ER. All the $pp\alpha F$-derived sequences possessed glycosylation sites that would be glycosylated if they entered the ER. Treatment of the cell lysates with endoglycosidase F prior to immunoprecipitation increased the mobility of the $D_N\alpha F$-2A-GFPh and $pp\alpha F$-2A-GFPh fusion proteins and the $D_N\alpha F$-2A and $pp\alpha F$-2A N-terminal processing products. Examination of the localisation of the GFP moiety by fluorescent microscopy showed that in cells transformed with either $pD_N\alpha F$-2A-GFPh or $pPp\alpha F$-2A-GFPh the GFP remained in the cytoplasm and did not gain access to the ER (*Figure 36*). Furthermore, separation of cell lysates into cytosolic and membranous fractions highlighted that the GFP in the uncleaved material was present in the membrane fractions. These results are consistent with the translation model of 2A-mediated processing occurring within the ribosome and appear to indicate that the translocon can recognize the gap in the nascent chain. Similar analysis of cells transformed with the $pD_N\alpha F$-2A-Kar2GFPh and $pPp\alpha F$-2A-Kar2GFPh constructs showed that the Kar2GFP is translocated into the ER and found associated with the membrane fractions of the cell extracts. In the case of the $pD_N\alpha F$-2A-Kar2GFPh construct it is unclear at this stage whether the translocon opens directly on presentation of pKar2GFP by ribosomes or whether the ribosomes dissociate and Kar2GFP is re-targeted.

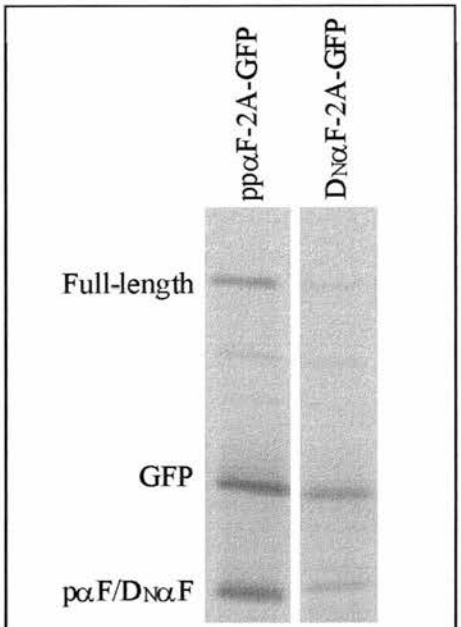


Figure 34: Translation profiles of the 2A-containing yeast artificial polyproteins. *The proteins were expressed by in vitro coupled transcription/translation reactions. Full-length proteins and processed products are indicated*

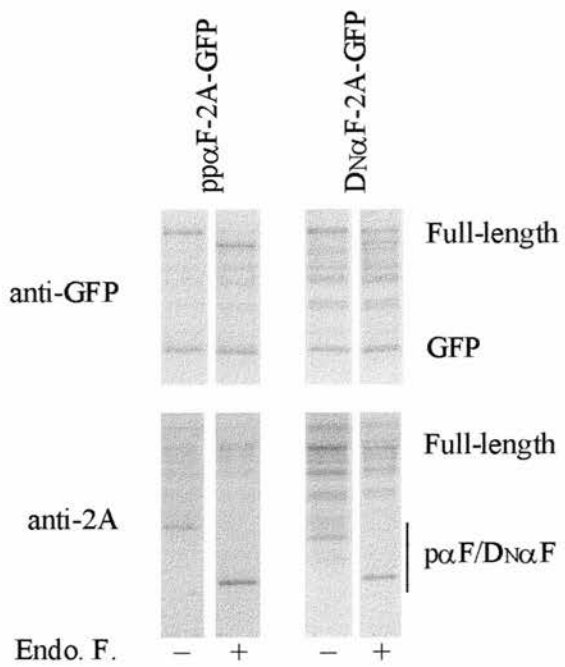


Figure 35: Immunoprecipitation of full-length and processed proteins expressed by transformed yeast cells. *Immunoprecipitations were performed with anti-GFP (upper panel) or anti-2A (lower panel). Endoglycosidae F treated (+) and untreated (-) extracts are indicated.*

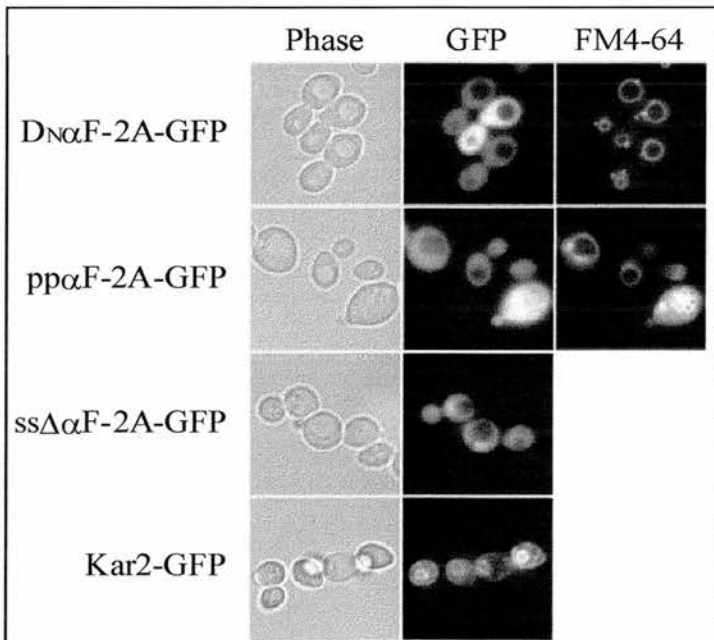


Figure 36: GFP from both $D_N\alpha$F-2A-GFP and ppαF-2A-GFP are cytosolic. *Phase images, GFP fluorescence staining of vacuolar membranes with FM4-64 are shown for cells expressing $D_N\alpha$F-2A-GFP and ppαF-2A-GFP. SsΔαF-2A-GFP and Kar2-GFP are controls representing a cytosolic protein and an ER targeted protein, respectively. The cells expressing $D_N\alpha$F-2A-GFP and ppαF-2A-GFP revealed cytosolic fluorescence with the major observable exclusion being from the vacuole.*

4. DISCUSSION

In the introduction I presented several examples of how RNA viruses employ diverse strategies to produce all their functional proteins from their relatively small genomes. Ribosomal frameshifting, leaky scanning, reinitiation and suppression of termination represent a variety of unusual translation strategies that viruses have evolved to expand their coding capacity and allow the synthesis of multiple polypeptides from a single RNA genome. I will now go on to describe what I believe to be yet another translational adaptation exploited by RNA viruses. I suggest that processing of the foot-and-mouth disease virus polyprotein at the 2A/2B junction is not a proteolytic event but rather a translational effect. I have identified other 2A and 2A-like sequences that I propose function in a similar manner to FMDV 2A and discuss their roles in the proteins of their origin. I will review the translational model for 2A activity and consider it in terms of the findings presented here.

4.1. 2A-Like Sequences

A major part of this effort to further characterize the 2A region of FMDV was to identify naturally occurring 2A-like sequences and screen them for an ability to mediate processing in an artificial polyprotein context. It was hoped that this approach would provide information on the requisites of the sequences N-terminal of the –DxExNPGP- motif known to be essential for the processing reaction mediated by FMDV at the 2A/2B junction. The 2A-like sequences recognized by database searches to contain the 2A motif are highly variable in their upstream content. The ability of these 2A-like sequences to mediate processing in the GFP-2A-GUS polyprotein system and the possible consequences of this reaction in the native protein are discussed here.

4.1.1. Picornaviruses

Foot-and-mouth disease virus was until recently the single member of the *Aphthovirus* genus of the family *Picornaviridae*. The early processing pattern of foot-and-mouth disease virus is similar to that of cardioviruses, with a processing event occurring at the C-terminus of 2A. Although the 2A regions of the cardioviruses are far larger than that of FMDV, the C-terminal regions are highly similar. The results presented here for FMDV, EMCV and TMEV are in agreement with reports that cleavage cassettes of 20 amino acids or longer will auto-process in certain heterologous contexts (Donnelly et al, 1997; Hahn & Palmenberg, 2001). The 20 amino acid segments analysed here, spanning the 2A region of FMDV or the C-terminal region of EMCV 2A, can achieve high level auto-processing of the reporter construct. The lower level of auto-processing exhibited by the TME 2A C-terminal region may be a consequence of the length of 2A region analysed. Studies on the influence N-terminal sequences on FMDV 2A activity demonstrated that as the length of 1D sequence upstream of 2A is increased, in the reporter system, the activity of 2A increased. Similarly as the length of 2A sequences was decreased, so too did activity (Donnelly *et al.*, 2001b). Corresponding results were observed in deletion mapping studies of the EMCV 2A region, constructs in which the C-terminus was shorter than 20 amino acids showed little or no processing activity.

Equine rhinitis A virus was previously named equine rhinovirus serotype 1, this was because of the mild respiratory disease it causes in horses, which resembles the common cold in man resulting from human rhinovirus infection. Characteristics such as viraemia during infection, persistent infections and the physical properties of the virus are more reminiscent of FMDV. Nucleotide sequencing of the viral genome confirmed the

identity of the virus as a picornavirus closely related to FMDV, and placed it in the aphthovirus genus (Wutz et al., 1996). The highest degree of similarity between the amino acid sequence of ERAV and FMDV occurs at the 2A region (88%). A high degree of similarity (88%) was also observed between the 2A sequence of equine rhinovirus-2 and FMDV 2A. The similarity of other ERV-2 proteins with the corresponding proteins of FMDV is low (Wutz *et al.*, 1996). This virus has been renamed equine rhinitis B virus (ERBV) and is classed in the erbovirus genus of the *Picornaviridae*. The ERAV 2A region tested here is highly active in this system and probably has an auto-processing activity in the viral polyprotein, similar to that of FMDV 2A.

Sequence analysis of a porcine enterovirus serotype 1 (PEV-1) isolate indicated that this virus has a genome organisation typical of other picornaviruses (*Figure 37*), particularly those of the cardio- and aphthovirus genera (Doherty *et al.*, 1999). Although the 2A protein of this virus has high similarity with those of FMDV (62%), ERAV (62%) and ERBV (81%), the sequence identity of the other viral proteins with representative members of the picornavirus genera is low (Doherty *et al.*, 1999). The sequence analysis of PEV-1 suggested that this virus forms a distinct branch of the picornavirus family, the new genus was named *Teschovirus* and PEV-1 renamed as porcine teschovirus-1 (PTV-1). The PTV-1 2A sequence analysed in the reporter polyprotein system was shown to be highly active and on the basis of these results it seems likely that the 2A region of PTV-1 functions, in a similar manner to FMDV 2A, to mediate auto-processing of the viral polyprotein.

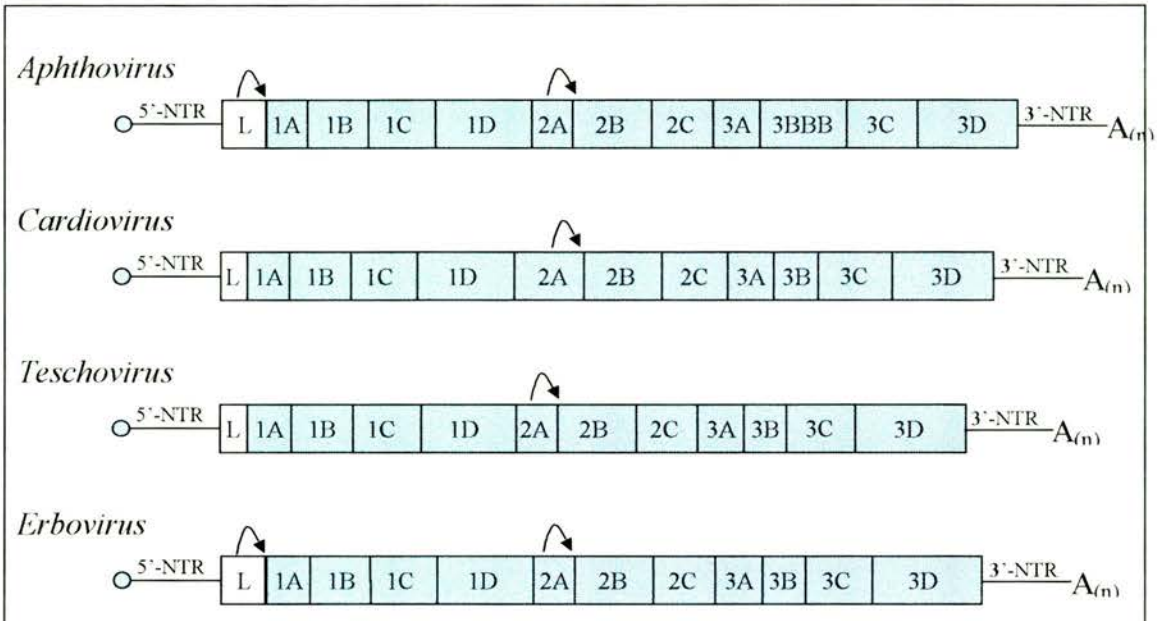


Figure 37: Genome organisation of the picornaviruses that have a 2A protein containing the DxExNPGP. *The labelled gene regions are not drawn to scale. The processing activities of the 2A proteins and the L-proteases are indicated with arrows. In the aphthovirus genus, FMDV has three copies of 3B (shown here), while ERAV does not. Non-translated regions (NTRs) flank the open reading frames. The 5' genome-liked VPg (3B) protein and 3' poly(A) tails are indicated.*

4.1.2. Rotaviruses

The other mammalian pathogens shown, by database searching, to have a 2A-like motif were the type C rotaviruses. The rotavirus genome consists of 11 segments of double-stranded RNA (dsRNA) that encode six structural and five non-structural proteins. The 2A-like region is present in the non-structural protein 3 (NSP3), which is encoded by genome segment 6. Comparisons of the NSP3 protein sequences of human, bovine and porcine type C rotaviruses (James *et al.*, 1999) has shown that the 2A-like sequence is highly conserved across the three. The NSP3 proteins are 402 amino acids long and have been compared to the group A rotavirus NSP3 (NS34) protein. The group C rotavirus NSP3 proteins are however 87-89 amino acids longer than their group A counterparts.

This C-terminal region of the NSP3 proteins of group C rotaviruses contains a dsRNA-binding domain situated downstream of the 2A-like motif (*Figure 38*). The activity of the 2A-like region analysed was far lower than that of the picornaviral 2A sequences in the reporter construct. However this may be a result of the sequences examined being shorter than the optimum length required by these proteins to exhibit higher levels of activity. This feature will be discussed later in terms of the translational model. The results presented here suggest that the type C rotaviruses could, in addition to the full-length NSP3 protein, produce the NSP3 protein minus the dsRNA-binding motif and possibly low levels of the dsRNA-binding region as a discrete product.

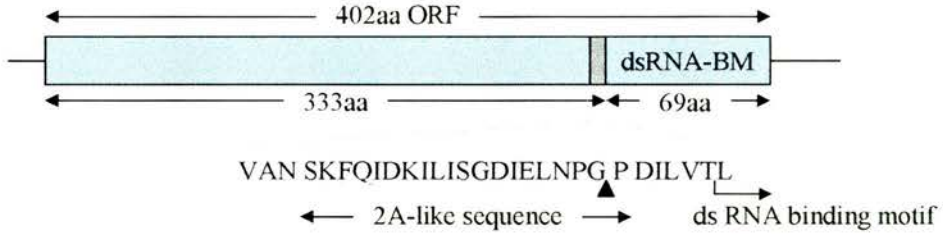


Figure 38: Genome segment 6 of the human type C rotavirus. *The 2A-like sequence is situated within the coding region of the non-structural protein, NSP3. The 2A-like sequence separates the downstream double stranded RNA binding motif from the remainder of the protein.*

4.1.3. Insect Viruses

Many insect-infecting viruses with picorna-like biophysical properties have been isolated. Genome sequence information for several of these viruses is now available, allowing 2A-like sequences to be identified. The insect virus, infectious flacherie virus (IFV), causes flacherie disease in the silkworm, *Bombyx mori*. Flacherie disease is a major factor causing serious loss of cocoon production in sericultural farms in Japan.

IFV targets the goblet cells of the midgut epithelium and multiplies in the cytoplasm. IFV has a genome arrangement typical of picornaviruses, with the structural proteins located in the N-terminal region and the replicative proteins at the C-terminus of the polyprotein. IFV has four major structural proteins, VP1, VP2, VP3 and VP4 that form an icosahedral shell. An L protein is situated at the N-terminus of the viral polyprotein and the RNA helicase, protease and RNA-dependent RNA-polymerase are encoded in the same order as the picornaviruses. The 2A-like sequence of IFV is situated at the C-terminal border of the capsid protein domain, separating the structural proteins from the non-structural proteins (*Figure 39*). The 2A-like sequence of IFV differs from the DxExNPGP motif of FMDV 2A; the aspartate residue is replace by a glycine residue, forming GxExNPGP. The IFV 2A-like sequence was active in the artificial polyprotein system utilised. It seems likely that this IFV sequence serves the same purpose as FMDV 2A in terms of polyprotein processing. It is interesting to note that the mutated IFV sequence, containing the DxExNPGP motif, is inactive in the reporter system. This finding suggests that the sequence requirements of an active 2A are more complex than possessing a DxExNPGP motif. Mutations of the aspartate residue to glutamic acid or glutamine in the FMDV 2A motif abolished activity (Donnelly *et al.*, 2001b). Similarly, mutation of this residue to histidine in the 2A motif of EMCV rendered the 2A protein inactive however, in a mutant where the aspartate residue was changed to asparagine, low levels of processing activity were detected (Hahn & Palmenberg, 1996). The requirement of a tyrosine residue, 18 amino acids upstream of the cleavage site, in EMCV 2A (Hahn & Palmenberg, 2001) reveals the requirement of a particular upstream sequence and it is possible that in IFV the upstream context supports the GxExNPGP motif in mediating processing.

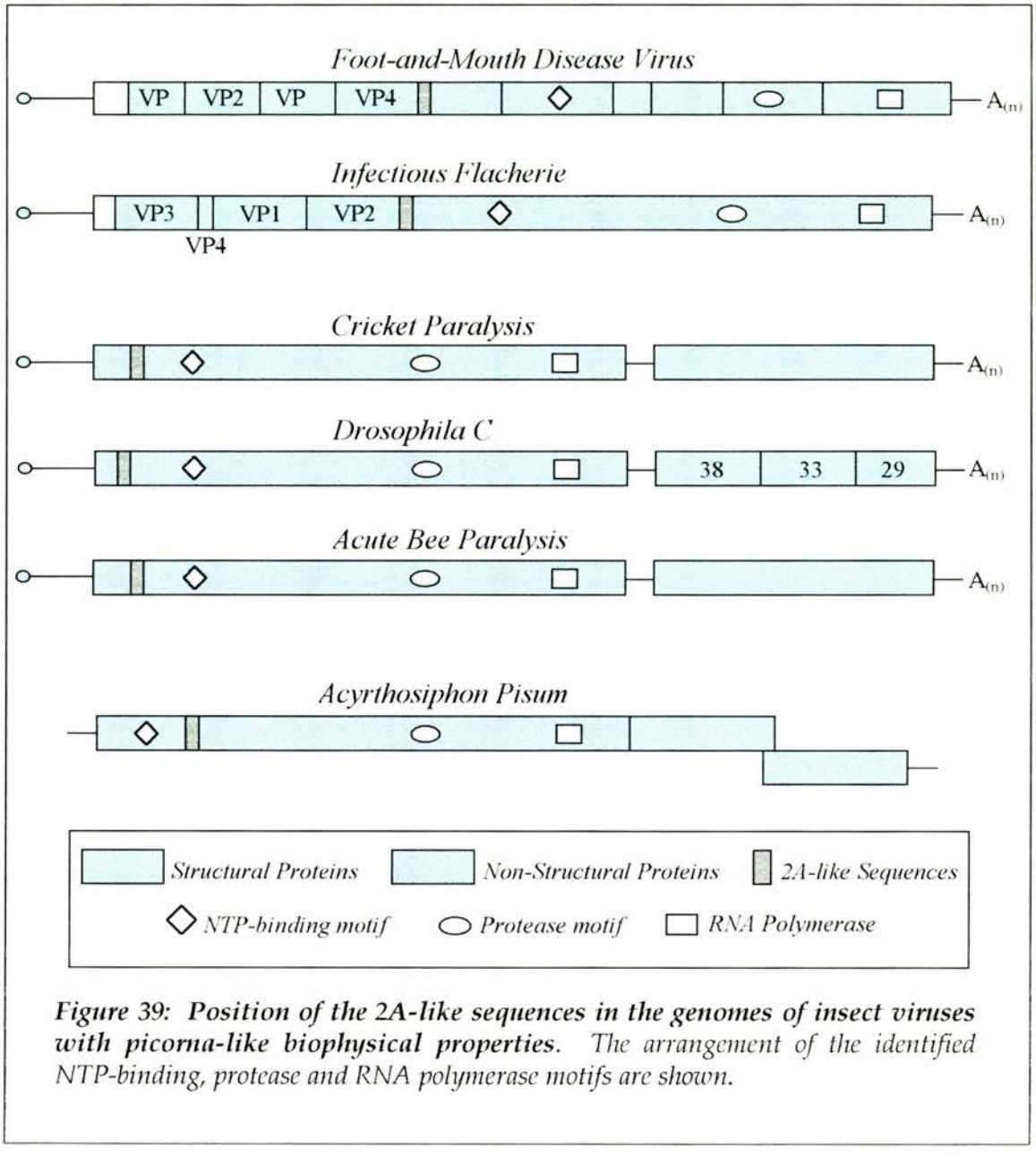


Figure 39: Position of the 2A-like sequences in the genomes of insect viruses with picorna-like biophysical properties. *The arrangement of the identified NTP-binding, protease and RNA polymerase motifs are shown.*

Several of the insect-infecting viruses with picorna-like biophysical properties have, upon genome sequencing, been shown to have a genome organisation that is quite different from picornaviruses (*Figure 39*). The genomes of these viruses are monopartite and bicistronic with the replicase polyprotein encoded by a 5'-proximal ORF and capsid polyprotein by a 3'-proximal ORF. The insect viruses, cricket paralysis virus (CrPV),

Drosophila C virus (DCV) and acute bee paralysis virus (ABPV) have this type of genome organisation. A recent report by Wilson *et al.* (2000a) has provided important information on the mechanism employed by cricket paralysis virus to synthesise two distinct polyproteins from a single mRNA. The upstream, ORF1, encoding the non-structural polyprotein, is separated from the downstream, ORF2, encoding the structural polyprotein, by a non-translated intergenic region (IR) of 189 nucleotides. The two open reading frames are preceded by internal ribosome entry sites, the upstream IRES being present in the 5'-nontranslated region and the downstream IRES in the intergenic region. These IRES elements function independently to control the translation of ORF1 and ORF2. In the systems utilised to study the IRES elements the IR IRES was consistently more active than the 5'NTR IRES, providing a mechanistic explanation for previous reports that the viral structural proteins are expressed in excess of the non-structural proteins in infected cells. The separation of the structural proteins from the non-structural proteins in the cricket paralysis virus genome together with the differential gene expression of these proteins, achieved by IRES elements of different strengths, is seen as an economical strategy for gene expression. This strategy facilitates the production of the relatively large amounts of the structural proteins required for the formation of the viral capsid while allowing the non-structural proteins, which function enzymatically in the replication of viral genome, to be produced sparingly. The control of gene expression in DCV and ABPV is suggested to be similar to that summarised here for CrPV (Johnson & Christian, 1998; Govan *et al.*, 2000).

The initiation of translation mediated by the IR IRES occurs via an unusual mechanism. Analysis of the N-terminal amino acid of CrPV ORF2 has shown that this residue is an alanine encoded by a GCU codon. This GCU codon and the preceding CCU triplet are

important for IR IRES activity (Wilson *et al.*, 2000a). The downstream ORF of DCV similarly lacks an initiating methionine codon and the N-terminal capsid protein begins with an alanine (Johnson & Christian, 1998). A mechanism for this unusual mode of translation initiation has been documented for CrPV (Wilson *et al.*, 2000b). In brief, ribosomes bind directly to a pseudoknot-like structure in the IR IRES, positioning the CCU triplet in the ribosomal P-site. The tRNA is delivered to the empty ribosomal A site and a translocation-type event then moves the GCU into the ribosomal P-site without peptide bond formation. The tRNA molecule carrying the second amino acid of the capsid protein can then occupy the now vacant A-site and translation can proceed.

The 2A-like sequences identified in the insect viruses, CrPV, DCV and ABPV occur in open reading frame 1, which encodes the replicase polyprotein. Domains with nucleotide binding (helicase), protease, and RNA-dependent RNA polymerase motifs have been identified in these viruses and they occur in a similar arrangement to that observed for the picornaviruses. The 2A-like sequences are situated in the N-terminal region of the non-structural polyprotein. The DCV and ABPV sequences analysed possess several common residues and were more active in the reporter polyprotein than the CrPV 2A-like sequence. All three of these 2A-like sequences were however capable of mediating high levels of processing and this model predicts that N-terminal products of 97aa (DCV) and 166aa (CrPV and ABPV) would be formed during translation of the viral non-structural ORF.

A region resembling the 2A sequence of FMDV was detected in the genome sequence of a virus isolated from the pea aphid, *Acyrthosiphon pisum*. However, the 2A-sequence of *Acyrthosiphon pisum* virus (APV) differed from the FMDV DxExNPGP motif at the

processing site, having DxExNPPPG instead. The organisation of the genome of APV shows some resemblance to that of DCV. It is monopartite and bicistronic, with replicase proteins encoded by the 5' region of the genome and capsid proteins encoded by the 3' region. However, there is a significant difference between these two viruses. In the APV genome the two ORFs overlap slightly, with the 3'-proximal ORF thought to be translated by a –1 ribosomal frameshift. Sequence analysis of the APV genome, reported by van der Wilk *et al.* (1997), showed that ORF1 contains domains resembling the replicase proteins; helicase, protease and RNA-dependent RNA-polymerase, in the order characteristic of picornaviruses. The capsid proteins, of which four have been identified (34K, 23K, 24K and 66K), are encoded by the 3' end of ORF1 and ORF2. It is predicted from sequence analyses that the major capsid protein, 34K, is encoded by the ORF1 3' region and that the 23K and 24K capsid proteins are breakdown products of the 34K protein. The 66K capsid protein is thought to be produced by a –1 frameshift event at the ORF1/ORF2 overlap region. A potential slippery sequence and stem-loop structure, elements required for frameshifting, have been identified at this site (van der Wilk *et al.*, 1997). The sequence resembling FMDV 2A is encoded downstream of the helicase domain and it was initially thought that if this sequence was active it could be responsible for separating this protein from the others encoded downstream in ORF1. The APV sequence tested in the artificial polyprotein system was found to be inactive. Considering the discrepancy between this sequence and the identified 2A motif this result was not surprising.

Thosea asigna virus (TaV) belongs to a virus family distinct from the picornaviruses and the picorna-like insect viruses. TaV belongs to the *Betatetravirus* genus of the family *Tetraviridae*, which infect lepidopteran insects. The TaV genome is monopartite with

the RNA-dependent RNA polymerase gene and capsid precursor protein gene on a 6.5 kb ssRNA molecule. The capsid precursor protein gene is also repeated on a subgenomic RNA molecule of 2.5 kb (Pringle *et al.*,1999). The TaV 2A-like sequence is present within the capsid precursor protein-coding region, at a site where processing has been reported to take place (Pringle *et al.*, 1999; Pringle *et al.*, 2001). Processing yields three proteins from the single ORF of the capsid protein domain. The N-terminal protein, p17, has unknown function and the other two proteins are the large (L) and small (S) capsid proteins. The 2A-like sequence is situated at the C-terminus of p17 with the second proline residue of the DxExNPGP motif being the first residue of the L protein, reminiscent of the FMDV 2A/2B junction (*Figure 40*). The sequencing studies of Pringle *et al.* (1999), together with the evidence presented here suggest that the 2A-like sequence of TaV functions in the viral capsid domain as it does in the reporter system to mediate processing of the polyprotein.

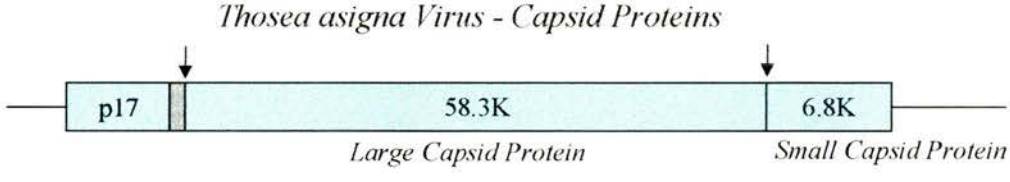


Figure 40: Position of the 2A-like sequence in the genome of TaV. *Arrows indicate the processing sites. Processing yields the large and small capsid proteins together with the p17 protein of unknown function.*

The 2A-like sequences examined here do not form an exhaustive list of the 2A-like sequences now present within the databases, in fact as time passes and more sequences become available it is predicted that the list will increase. The recently sequenced *Perina nuda* picorna-like virus (PnPV) possess two 2A-like sequences (Wu *et al.*, 2002).

Interestingly, this virus has a 2A-like region within the capsid protein domain, analogous to TaV, and has a second 2A-like sequence at the junction between the capsid and replicative proteins, similar to the picornaviruses (*Figure 41*). N-terminal sequencing of the CP2 protein has shown that this protein begins with a proline residue, predicted to be from the –NPG/P- motif of the first 2A-like sequence and indicating that this 2A-like sequence is functionally active. Based on the position of the second 2A-like sequence, it is predicted that this 2A-like sequence functions in a similar manner to the active 2A/2A-like sequences of the aphth-, cardio-, tescho-, erbo- and infectious flacherie viruses. The presence of two functionally active 2A-like sequences has never before been reported and at present remains a characteristic confined to the Perina nuda picorna-like virus.

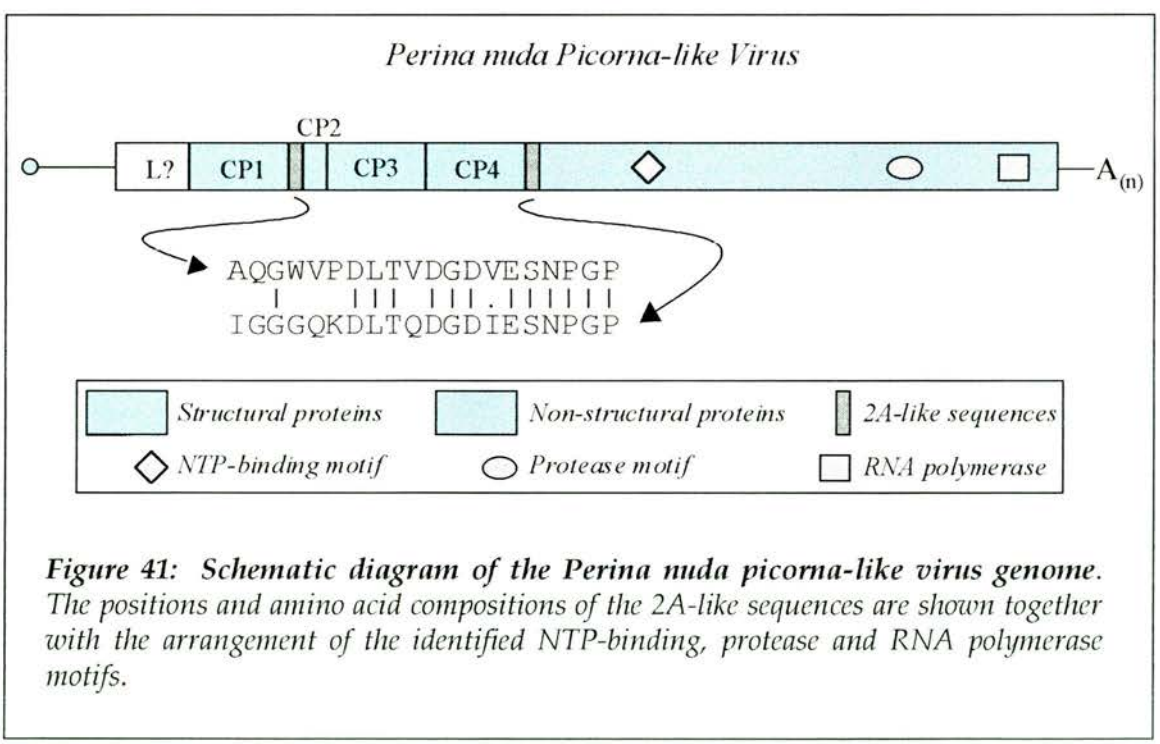


Figure 41: Schematic diagram of the Perina nuda picorna-like virus genome. *The positions and amino acid compositions of the 2A-like sequences are shown together with the arrangement of the identified NTP-binding, protease and RNA polymerase motifs.*

4.1.4. Trypanosome Repeated Sequences

The only other functional 2A-like sequences were identified in trypanosomal sequences. Trypanosomes are single-celled parasitic organisms. *Trypanosoma brucei* is the causative agent of the human African sleeping sickness, the name of which well describes the complete fatigue accompanying persistent infection with this aggressive parasite. This organism is well documented in several good textbooks for its ability to evade the host immune response (see Wolfe 1993 or Watson *et al*, 1988). The parasite enters the bloodstream of the host during an insect bite from the tsetse fly and multiplies. During the seven days or so that it takes the body to mount an immune response to the parasite, some of the trypanosomes change their surface antigens and are not recognised by the antibodies. The persistent infection causes lethargy and leads to death as the trypanosomes invade the central nervous system. In Africa and South America, trypanosomes frequently infect cattle, lowering the yield of milk and meat and increasing susceptibility of the animals to other diseases. The *Trypanosoma brucei* 2A-like sequence was identified in a protein showing homology with reverse transcriptase (RT). The coding sequence of this protein is flanked by different separate halves of a transposable element termed ribosomal insertion mobile element (RIME) (Murphy *et al.*, 1987). The 2A-like sequence is present in the N-terminal region of the RT-like protein (*Figure 42*). Similarly, the 2A-like sequence of *T. cruzi* was identified in a non-LTR retrotransposon (L1Tc) that disrupts a RIME. L1Tc has three main ORFs, the first of which encodes a protein with similarity to the human AP endonuclease protein. The 2A-like sequence is situated at the N-terminus of this protein (*Figure 42*). These trypanosomal 2A-like sequences may play a role in liberating the identified proteins from their fusion proteins.

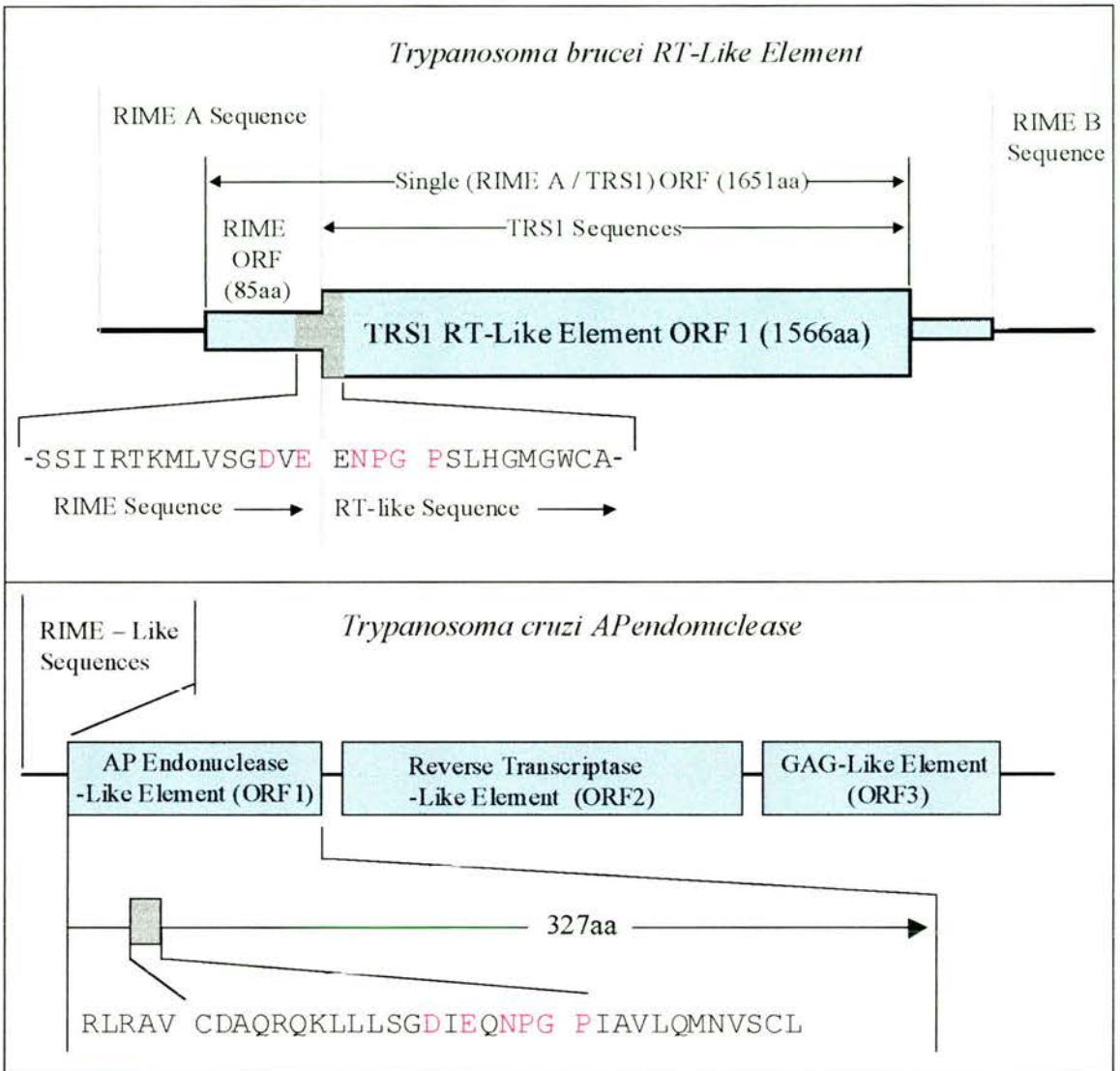


Figure 42: The positions and sequences of the T. brucei TRS1 and T.cruzi Apendo 2A-like sequences shown together with their predicted processing sites.

4.1.5. Cellular Sequences

The 2A-like sequences with cellular origin were all found to be inactive in the assay system utilised. The 2A-like sequence identified in the carbamyl-phosphate synthetase 1 (CPS1) protein of the North American bullfrog, Rana catesbeiana, differed from the 2A motif at the first position, with the aspartic acid residue being replaced with a histidine. The 2A-like sequences identified in the Mod(mdg4) 59.0 modifier protein of *Drosophila melanogaster,* the mouse mu opioid receptor variant F (MOR-1F) and the α-glucuronidase enzyme of the hyper-thermophilic bacterium, *Thermotoga maritima* all possessed the DxExNPGP motif but still were inactive. These results support the theory that a DxExNPGP motif alone is not sufficient to mediate processing and that the assistance of a critical upstream sequence of optimum length is also required. The sequences studied here were based on the length of the FMDV 2A sequence that was required to demonstrate processing in the artificial polyprotein system. However, the cellular sequences examined may have been too short and it is possible that if the 2A-like sequences were extended N-terminally processing may be detected in this artificial polyprotein system.

4.1.6. Similarities and Differences of the 2A-Like Sequences

Comparison of the naturally occurring 2A-like sequences which are able to mediate processing shows that they are largely, highly variable. One similarity is that all the active 2A-like sequences have either a valine or isoleucine residue in position 7 (see *Figure 43*). The naturally occurring non-processing 2A-like sequences do not possess one of these residues at this position. The glycine at position 9 is also highly conserved across the active 2A(-like) sequences with the only exceptions being the alanine of IFV and the histidines of the cardioviruses. A general clustering of hydrophobic residues

with aliphatic side chains occurs in the region of positions 11 to 14 in the active 2A(-like) sequences. In N-terminal deletion studies involving FMDV and EMCV 2A sequences, lengths of 2A up to and including these positions have been shown to be required for processing (Donnelly *et al.*, 2001b; Hahn & Palmenberg, 2001). It is difficult to make any other generalisations between the highly variable primary structures of the active 2A-like sequences.

Active 2A-like Sequences																			
FMDV	Q	L	L	N	F	D	L	L	K	L	A	G	D	V	E	S	N	P	G	P
TMEV	Y	H	A	D	Y	Y	K	Q	R	L	I	H	D	V	E	M	N	P	G	P
EMCV	H	Y	A	G	Y	F	A	D	L	L	I	H	D	I	E	T	N	P	G	P
ERAV	Q	C	T	N	Y	A	L	L	K	L	A	G	D	V	E	S	N	P	G	P
PTV1		A	T	N	F	S	L	L	K	Q	A	G	D	V	E	E	N	P	G	P
PRotaC	A	K	F	Q	I	D	K	I	L	I	S	G	D	V	E	L	N	P	G	P
HRotaC	S	K	F	Q	I	D	K	I	L	I	S	G	D	I	E	L	N	P	G	P
IFV	T	R	A	E	I	E	D	E	L	I	R	A	G	I	E	S	N	P	G	P
TaV	R	A	E	G	R	G	S	L	L	T	C	G	D	V	E	E	N	P	G	P
CrPV	F	L	R	K	R	T	Q	L	L	M	S	G	D	V	E	S	N	P	G	P
DrosC		A	A	R	Q	M	L	L	L	L	S	G	D	V	E	T	N	P	G	P
ABPV	G	S	W	T	D	I	L	L	L	L	S	G	D	V	E	T	N	P	G	P
TSR1	S	S	I	I	R	T	K	M	L	V	S	G	D	V	E	E	N	P	G	P
TSR1(L)*	S	S	I	I	R	T	K	M	L	L	S	G	D	V	E	E	N	P	G	P
APendo	C	D	A	Q	R	Q	K	L	L	L	S	G	D	I	E	Q	N	P	G	P
Inactive 2A-like Sequences																			
Therm	Y	I	P	D	F	G	G	F	L	V	K	A	D	S	E	F	N	P	G	P
Drosmel	T	A	A	D	K	I	E	G	S	L	K	M	D	T	E	G	N	P	G	P
Mus	D	L	E	L	E	T	V	G	S	H	Q	A	D	A	E	T	N	P	G	P
CPS	M	H	E	T	K	P	I	F	T	S	Q	F	H	P	E	A	N	P	G	P
APV	N	T	P	M	P	E	A	L	Q	K	I	I	D	L	E	S	N	P	P	P	G

Figure 43: Comparison of the amino acid compositions of the 2A-like sequences. *The hydrophobic residues are shown in red. All the active 2A-like sequences have a valine or isoleucine at position 7; the inactive 2A-like sequences do not.*

The primary structure of these 2A and 2A-like sequences influence the higher structural levels of the protein. The 2A regions of FMDV and the cardioviruses (EMCV and TMEV) have previously been reported to have aliphatic residues occurring at intervals supporting helical conformations (Donnelly et al., 1997; Hahn & Palmenberg 2001). Modelling of the 2A regions used in these studies have shown that for EMCV and TMEV a four-turn helix would have one face composed of large hydrophobic residues with strong stacking potential and another face of highly charged residues. Similar modelling of the FMDV 2A region also demonstrates a clustering of hydrophobic residues on the face of a helix (Hahn & Palmenberg, 2001). Hydrophobic faces usually indicate an affinity for other hydrophobic surfaces and it is possible that the 2A sequences provide a framework for the interaction of the native polyprotein with the translating ribosome thus creating the correct environment for the observed translational effect.

Foot-and-mouth disease 2A has now been examined and found to be active in a range of protein contexts both here in the GFP-2A-GUS, YFP-2A-GAL-CFP-2A-PAC and $D_N\alpha$F-2A-GFPh artificial polyproteins and elsewhere for a variety of biotechnological purposes. It has been demonstrated that the 2A peptide is active both in the wheat germ extract cell-free expression system and also in mammalian and yeast cells. Indeed, the FMDV 2A peptide has been found to be active in mammalian, plant and fungal cells (Mattion *et al.*, 1996; de Felipe *et al.*, 1999; Varnavski *et al.*, 2000; Halpin *et al.*, 1999, Suzuki *et al*, 2000). FMDV 2A is not however active in prokaryotic cells (Donnelly et al., 1997). It appears that the co-translational processing event mediated by the FMDV 2A peptide is specific to the eukaryotic translation machinery.

4.2.1. In Vitro Activity of FMDV 2A

The multiple translation products produced from all of the FMDV 2A and 2A-like constructs described here are generated from a single long ORF. The imbalance in the accumulation of the two processing products seen consistently with all of the active 2A-like sequences is reminiscent of the initial observation made in the profiles of constructs containing FMDV 2A. The N-terminal translation product (TP1) is always in excess of the product C-terminal of 2A (TP2). It was these results that gave the first indications that the 2A-mediated processing event may not be a proteolytic event. The elimination of several factors which may disrupt the stoichiometry of the to two 'cleavage' products have led us to believe that the excess of TP1 is a result of the increased synthesis of this product (Donnelly *et al.*, 2001a). The translational model that has been developed for the mechanism of 2A action accounts for the imbalance of the two processing products in the artificial polyprotein system while allowing for the near 1:1 ratio observed when FMDV

2A is in its native polyprotein context (Donnelly et al., 2001a). The *in vitro* reactions programmed with the 2A/2A-like series of constructs appear to indicate that there are two parts to the 2A-mediated translational effect. There is a pseudo-termination event, which when it fails yields the full-length polyprotein product. This does not occur in FMDV infected cells and appears to be a product of sub-optimal functioning of the 2A region in the *in vitro* system. This appears to be due to the length of the FMDV sequence being too short. Increasing the length of the FMDV sequence, by adding sequences of the 1D region on to the N-terminus of 2A vastly reduces the amount of unprocessed material detected in the *in vitro* reactions. A successful pseudo-termination event can be followed by two distinct outcomes. Firstly, the ribosome either stalls or dissociates consequently forming only TP1. This is the part that would account for the excess of the GFP2A moiety over the GUS product detected in the *in vitro* translation system. The final outcome occurs when the pseudo-termination reaction is followed by the second part of the 2A action, a pseudo-initiation event. This complete 2A action results in the synthesis of TP1 (GFP2A) followed by the synthesis of TP2 (GUS) without the formation of a peptide bond between the two proteins.

4.2.2. A Translational Model for FMDV 2A Activity

A translational model for FMDV 2A activity was proposed by Ryan *et al.* (1999) and suggests that the 2A protein may have esterase activity rather than that of a proteinase. In this model the 2A protein mediates an attack on the ester linkage between the nascent peptide and the tRNA moiety. The model accounts for the three outcomes observed in the artificial polyprotein system (*Figure 44*). First, the peptide bond formation proceeds throughout the length of the polyprotein to give the uncleaved [CAT2AGUS] (*Figure 44, steps 1-4*). Second, ribosomal dissociation occurs at the C-terminal of 2A, to yield

[CAT2A] only (*Figure 44, steps 1 & 5*), accounting for the excess of the first translation product (TP1). Third, that ribosomal translocation from the last codon of TP1, to the first codon of TP2, occurs without peptide bond formation to yield the discrete second translation product (*Figure 44, steps 1,2,6,7 & 8*). Finally the model also accounts for the discrepancy that when 2A is in the artificial system there is an observed imbalance in the ratio of cleavage products but in its native polyprotein context the cleavage products are produced in equimolar amounts.

4.2.3. Mechanism of FMDV 2A Activity

The model that we believe represents the mode of 2A action is based on the inhibition of the formation of a peptide bond between the terminal glycine residue of 2A and the initial proline of 2B (-NPG/P-). The peptide bond between the proline and glycine at the C-terminus of 2A is synthesised as normal. Translocation of the peptidyl-tRNA into the P site, mediated by eEF2, allows ingress of the prolyl-tRNA. The nucleophilic attack by the prolyl-tRNA amide nitrogen on the peptidyl-tRNA carbonyl carbon is inhibited by 2A, preventing peptide bond formation. Hydrolysis of the ester linkage between the peptide chain and the tRNA releases the nascent peptide from the ribosome (*Figure 45*). The prolyl-tRNA in the A site is then translocated to the P site and translation of the downstream product continues as if a peptide bond had been formed.

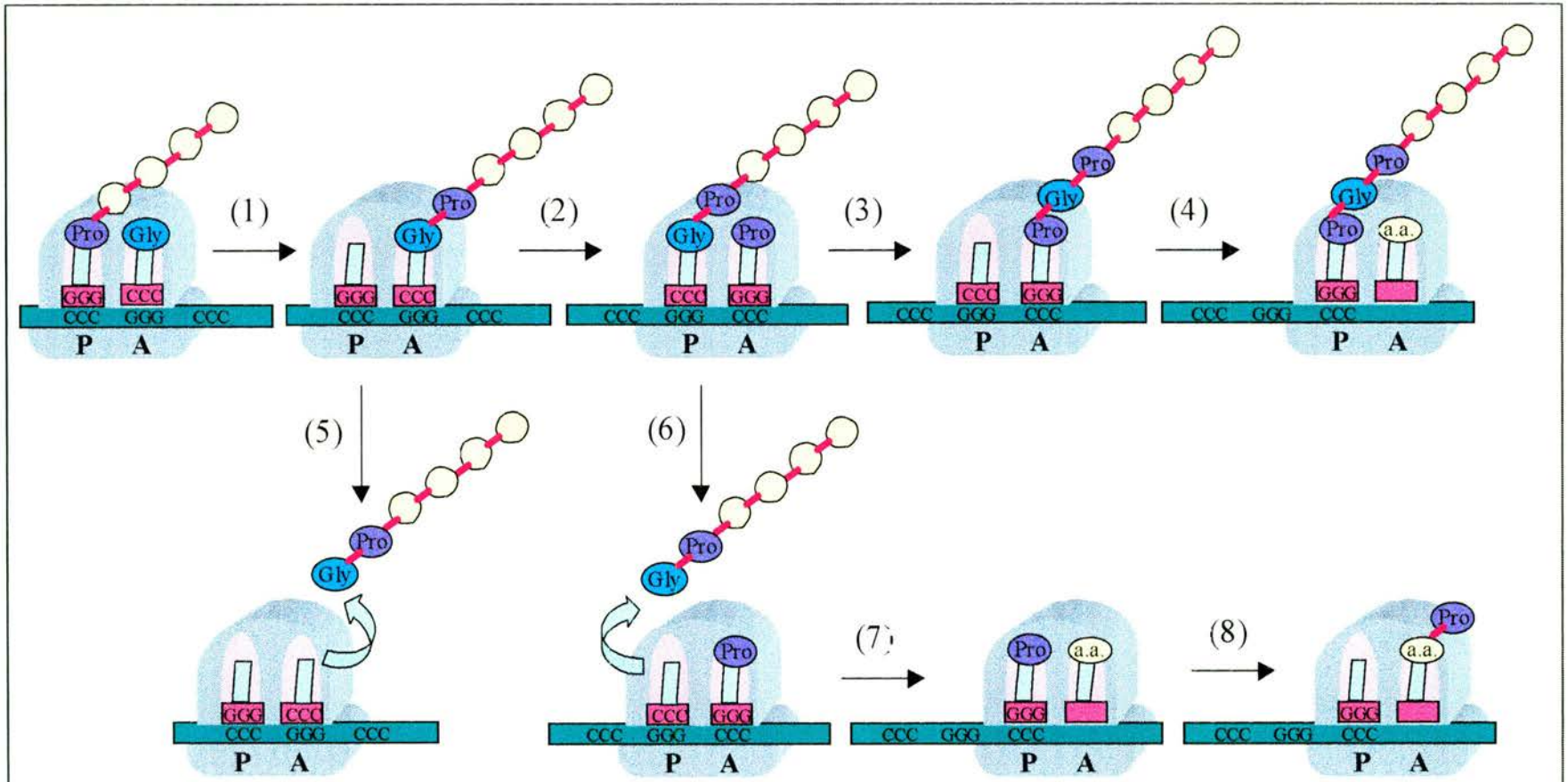


Figure 44: Schematic diagram of the model proposed for polyprotein cleavage mediated by FMDV 2A-induced esterase activity. *Shown are the three outcomes of translation which occur when 2A is in an artificial polyprotein system. I) A full-length polyprotein is produced (steps 1-4). II) The first translation product only is synthesized, followed by dissociation of the ribosome complex (steps 1 & 5). III) The first translation product is released and synthesis continues to yield the second translation product (steps 1,2,6,7 & 8). See text for detailed discussion.*

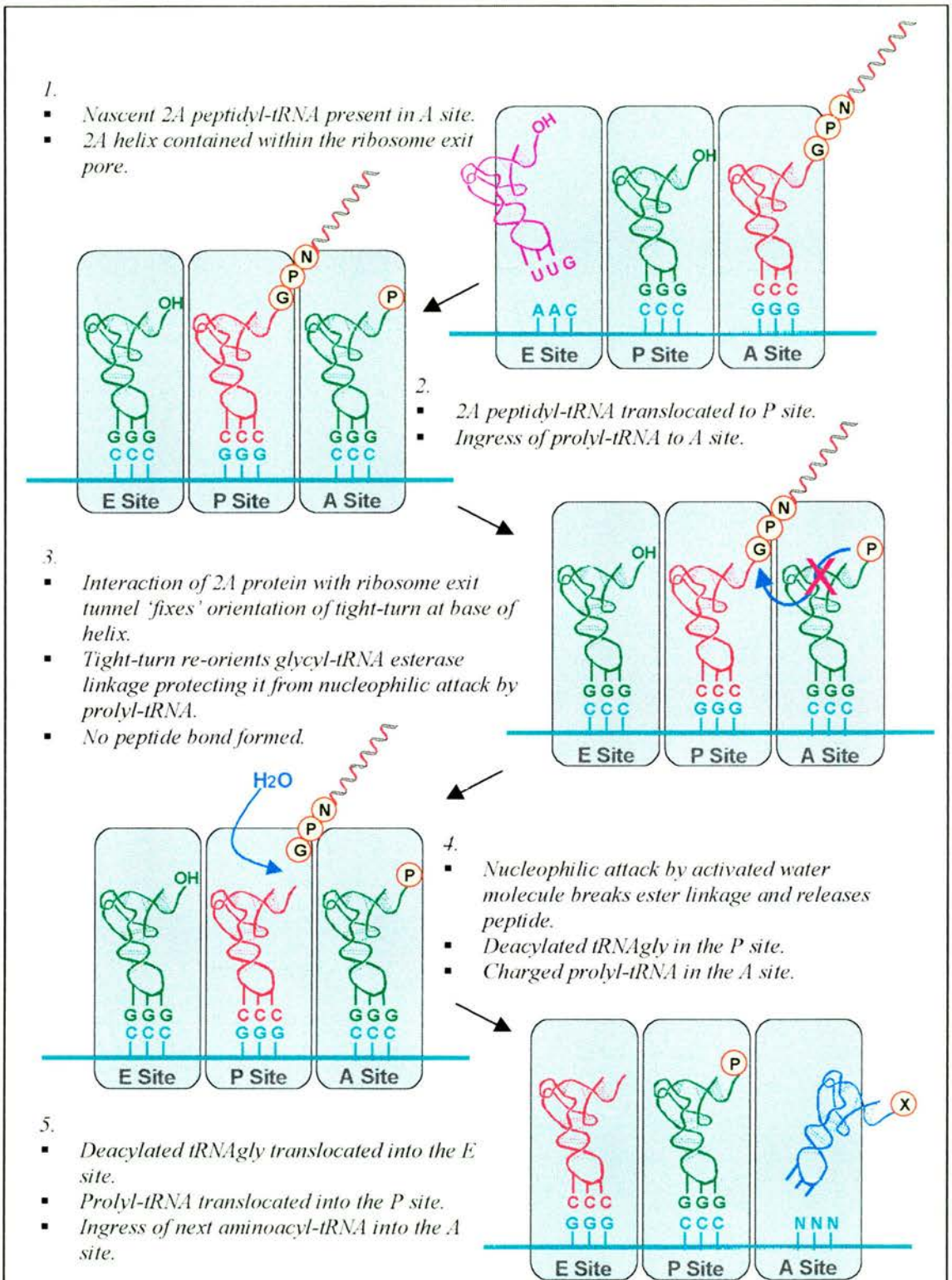


Figure 45: The mechanism by which FMDV 2A mediates a translational effect that yields two protein products from a single open reading frame.

4.2.4. The Nascent Peptide

The inactive 2A-like sequences identified by the GFP-2A-GUS artificial polyprotein screening system have shown us that the sequence upstream of the recognised – DxExNPGP- processing site motif is an important factor in the activity of 2A and 2A-like peptides. Viral sequences known to influence the translational machinery of the host cell have been described previously. The notion of a nascent peptide interacting with the exit tunnel of the ribosome to induce ribosomal pausing and inhibit peptidyltransferase activity has been well documented for both the nascent peptide control of translation of chloramphenicol resistance genes (Rogers & Lovett, 1994; Gu *et al.*, 1994; Harrod & Lovett, 1995) and ribosomal hopping within the bacteriophage T4 *gene 60* (Weiss *et al.*, 1990 and reviewed by Farabaugh, 1996).

In the case of the inducible chloramphenicol resistance genes, translation is attenuated until the inducer (chloramphenicol) stalls a ribosome in a leader sequence located at the 5' end of the regulated transcripts. Stalling causes a localised change in the structure of the mRNA exposing a normally sequestered ribosome-binding site, allowing translation of the resistance determinant. Ribosomal stalling occurs at a leader codon that is preceded by the coding sequence of a leader peptide. It is proposed by Harrod & Lovett (1995), that the leader peptide extends from its attachment to peptidyl-tRNA in the P site into the peptidytransferase centre, situated in the most interior portion of the ribosomal exit tunnel, to inhibit pepidyltransferase activity and bring about a translational pause.

The bacteriophage T4 DNA topoisomerase is a multiple sub-unit enzyme consisting of polypeptides encoded by genes *39*, *52* and *60*. Gene *60* encodes an 18kd sub-unit that is required for tight complex formation of the three sub-units. Within the coding region of

gene *60* there is an internal 50-nucleotide untranslated region that persists in the mRNA. The coding gap separates the first 46 codons of the gene *60* from the last 114 codons, it is the combination of the two translated regions that yield the 18kd protein. Protein sequencing of the gene *60* product has shown that gene is translated up to a GGA glycine codon, the codon before an in-frame UAG nonsense codon, then continued to be decoded 50 nucleotides downstream, starting immediately after another GGA glycine codon (reviewed by Farabaugh, 1996). The factors required for the ribosomal hop have been identified as, a pair of matched "take-off" and "landing" codons and the distance between them, a stop codon and hairpin structure at the beginning of the gap, and an upstream nascent polypeptide sequence. Mutational analysis of the 5' coding segment by Weiss *et al.* (1990) identified a 16 amino acid region of the nascent peptide that is critical for gap bypass. They suggest that this region of the nascent peptide be situated in the ribosomal exit tunnel when the "take-off" codon is situated at the decoding sites. The interaction of the nascent peptide with the ribosome exit tunnel may have functional consequences that destabilise the P site tRNA binding and activate bypass of the gap region.

We believe that in the case of the active 2A and 2A-like proteins the sequences upstream of the processing site form an important interaction within the ribosome exit tunnel to create an environment that promotes 2A activity at the decoding sites. The nascent peptide serves the function of orientating the tight-turn (-NPG-) at the base of the helix such that it prevents the formation of the next peptide bond but permits hydrolysis of the linkage between the nascent peptide and its tRNA. This theory is supported by the finding that it is not only the identified –DxExNPGP- motif that it required for a 2A-like sequence to be active, the sequence upstream of this motif is also critical to the functionality of the 2A protein. The incomplete processing activity of FMDV 2A and the

active 2A-like sequences (and indeed the observed inactivity of the other 2A-like sequences) in the *in vitro* reactions performed here may be due to the length of the sequences examined. Within the exit tunnel of the ribosome, the nascent peptide may interact at several or just a few crucial positions (*Figure 46*). Increasing the length of the 2A sequence may allow more interactions to occur or permit a single critical interaction, thus allowing the protein to better orient the tight-turn at the processing site and in turn increase the processing activity of the 2A protein. Increasing the length of the FMDV sequence, by extending it N-terminally to include regions of the 1D protein, improves the processing efficiency of 2A in the GFP-2A-GUS artificial polyprotein (Donnelly *et al.*, 2001b). Additionally, in the native polyprotein context, no full-length polyprotein spanning the processing site was detected (Donnelly *et al.*, 2001a).

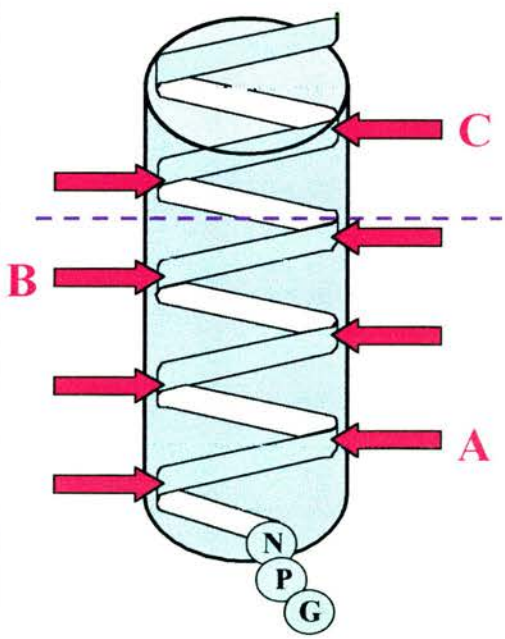


Figure 46: Schematic diagram of the nascent peptide within the exit tunnel of the ribosome. *Interactions between the nascent peptide and the ribosome may occur along the length of the exit pore, shown by the arrows. These interactions will orient the tight-turn (-NPG) at the base of the helix. Where a 2A protein requires interaction at sites A, B and C for activity, the optimum length of the 2A sequence would include these sites. It may be the case for some of the 20 amino acid sequences examined here, that they extend only to the dashed line and an important stabilising interaction a position C is lost. This could result in sub-optimal functioning of the 2A protein.*

4.2.5. The Role of Proline at the Processing Site

In order for the 2A-mediated translational effect to form two discrete products from a single open reading frame, peptide formation between the 2A peptidyl-tRNA and the prolyl-tRNA must be prevented. The proposed inhibition of peptidyltransferase activity by the nascent 2A peptide is one factor, another is the absolute requirement of a proline residue at the N-terminus of the downstream translation product. Mutation of this proline to alanine, serine, isoleucine or phenylalanine completely abolish processing activity (Donnelly *et al.*, 2001b). Proline differs from the other common amino acids in that its side chain is bonded to the backbone nitrogen atom as well as the α carbon atom. The resulting cyclic structure markedly influences protein architecture with it often found in bends of folded proteins. Glycine and asparagine also have a propensity for forming reverse turns in which the polypeptide chain abruptly changes direction, these are the three residues found at the C-terminus of 2A. The five-member ring structure of proline prevents rotation about the Cα-N bond, markedly restricting conformation and increasing steric hindrance. Consequently, proline has a lowered nucleophilicity compared to that of the other common amino acids. The elimination of processing activity that accompanies the mutation of the downstream proline is probably due to the replacement amino acid out-competing the hydrolysis of the 2A peptidyl-tRNA ester linkage. Instead of the formation of TP1 a peptide bond is formed and translation continues to form full-length product only.

4.2.6. Hydrolysis versus Peptide Bond Formation

Nissen *et al.* (2000) have suggested that during translation, hydrolysis of the peptidyl-tRNA in the P site is prevented prior to the delivery of the next appropriate aminoacyl-tRNA to the A site by the conformation of the ribosome when the A site is empty. They

suggest that when the A site is vacant the orientation of the peptidyl-tRNA in the P-site is such that it is prevented from attack. When the aminoacyl–tRNA binds in the A site a conformational change occurs precisely orientating the alpha amino group from the A site aa-tRNA and the carbonyl carbon from the P site peptidyl-tRNA. Peptide bond formation proceeds and the re-orientated peptidyl-tRNA is susceptible to hydrolysis. It has previously been suggested that the C-terminal –NPG- residues of the 2A nascent peptide could serve to orient the 2A peptidyl-tRNA substrate to disfavour peptide bond formation and promote hydrolysis (Ryan *et al.*, 1999). However, this would occur when the prolyl-tRNA occupies the A site. The 2A peptide may increase the rate of hydrolysis mediated by the ribosome or act itself as a hydrolytic element, activating a water molecule.

This translational model that has been developed for the processing activity exhibited by the 2A peptide of FMDV relies on the hydrolysis reaction that releases the first translation product from its attachment to tRNA, out-completing the formation of a peptide bond between the peptidyl-tRNA and prolyl-tRNA substrates. This is what occurs when 2A is in its native environment or when it is extended N-terminally with sequences of 1D, no full-length polypeptide is detected. However, in the *in vitro* reactions where only 20 amino acids of 2A are present, full-length polyproteins are detected showing that hydrolysis does not always out-compete peptide bond formation.

4.2.7. Product Imbalance In Artificial Systems

Several factors that may influence the translational effect mediated by FMDV 2A have been presented. The predicted helical structure of the 2A protein and the tight turn at the C-terminus (Donnelly *et al.*, 1997; Ryan *et al.*, 1999; Hahn & Palmenberg, 2001) are

thought to be crucial for the correct alignment of the nascent protein in the ribosome complex. Furthermore the kinetics of translocation compared to cleavage of the ester linkage is considered important in determining the translational outcome of the system. The imbalance in translation products, observed in the artificial polyprotein *in vitro* expression systems, is thought to arise as a result of sub-optimal functioning of the 2A protein. Situations that may account for this inefficiency are discussed here.

It is possible that the 2A and 2A-like sequences examined here are shorter than the length required for the nascent protein to correctly align in the ribosome complex. Misalignment of the peptidyl-tRNA in the P site may prevent both the hydrolysis of the ester linkage and attack of the prolyl-tRNA to form the peptide bond. This could lead to stalling of the ribosome at the processing site and upon analysis result in the TP1:TP2 imbalance. With 2A in its native context the alignment of the peptidyl-tRNA in the P site would favour hydrolysis with the translocation of the prolyl-tRNA to the P site resulting in a 1:1 ratio of the two translation products.

Alternatively, the excess of the N-terminal translation product may arise as a result of unstable translation complexes being formed. Unstable translation complexes may be formed either by the hydrolysis reaction occurring prematurely when the 2A peptidyl-tRNA is still in the A site, or by a decreased rate of translocation of the prolyl-tRNA into the P site after release of the first translation product. Dissociation of these unstable translation complexes would result in the formation of TP1 only.

Since the translocation reaction is catalysed by eEF2, the levels of this elongation factor should also be considered. When eEF2 activity is normal, translocation of the peptidyl-

tRNA from the A to P site allows ingress of prolyl-tRNA into the A site, release of TP1 by the 2A esterase activity and subsequent translocation of the prolyl-tRNA into the P site. Continuation of translation synthesizes TP2 as a second discrete translation product. In situations where eEF2 translocation activity is abnormally low, attack on the peptidyl-tRNA ester linkage may occur in the ribosomal A site releasing TP1 from the ribosome. The resultant complex, with deacylated tRNAs in both sites of the ribosome, could lead to subunit dissociation analogous to normal termination of translation (Ryan *et al.*, 1999).

Noteworthy when considering the rate of translocation, are the findings of Svitkin and Agol (1983). When examining the translation of cardiovirus RNA in Krebs-2 cell-free extracts their appeared to be a 'translational barrier' in the central region of the genome, preventing synthesis of the translation products downstream of the viral 2A. These Krebs-2 cell extracts were found to have only low levels of eEF2 activity and upon the addition of eEF2 the synthesis of the proteins C-terminal of 2A was greatly enhanced. The cardiovirus 2A regions examined here have shown translation profiles similar to that produced by FMDV 2A and agree with the proposal of Ryan *et al.*, (1999), that the model described for FMDV 2A activity, may also apply to cardiovirus 2As. This model can account for the observation of a 'translational barrier', by Svitkin and Agol (1983). The low eEF2 activity in the Krebs cell extract may lead to cardiovirus 2A attacking its own ester linkage to $tRNA^{gly}$ in the A-site of the ribosome. The translational barrier could be due to ribosome release in this region. Supplementation of the Krebs-2 cell extract with eEF2 promotes translocation and the cardiovirus 2A-mediated esterase activity occurs when the peptidyl-tRNA is in the P site allowing synthesis of the downstream products.

It is intriguing to think that the susceptibility of the 2A-mechanism to factors such as eEF2 may represent an elegant twist in the translation of viral mRNA, so that in the latter stages of infection only the capsid polyprotein is produced. This would permit the generation of the vast excess of capsid proteins (60 copies of each) required to encapsidate a single viral genome without the wasteful 1:1 production of the enzymatic replicative proteins.

4.2.8. The Role of the –FFF- Sequence of FMDV 2B

The nucleotide sequences immediately downstream of the FMDV 2A/2B processing junction (–NPG/P-) encodes a string of three phenylalanine residues. The presence of this sequence in artificial polyprotein constructs influences the translational profiles from *in vitro* expression systems. The imbalance in the two protein products is greatly enhanced with TP1 being in far greater excess of TP2. This –FFF- encoding sequence is not required for 2A function, as has been demonstrated here with the constructs devoid of this 2B region. Additionally this sequence is not present immediately downstream of the 2A-like sequences examined here. The –FFF- sequence appears to have a translational function that is independent of, but operates alongside, that of FMDV 2A. The translational profiles of the FFF-containing constructs suggests that on some occasions, when the 2A protein has successfully mediated the translational effect that releases TP1, the continued synthesis that would produce TP2 is prevented by the presence of the –FFF- encoding sequence. I have previously reported the activity of the –FFF- sequence in an artificial polyprotein construct where it is flanked by the reporter proteins GFP and GUS, with the 2A sequence omitted (GFP-FFF-GUS). Gel analysis of the *in vitro* translation profile generated by this construct identified two bands, one corresponding in size to the full-length polyprotein and another corresponding in size to GFP.

Additionally, the presence of –FFF- downstream of an inactive 2A mutant in the GFP-GUS artificial polyprotein system (GFP-2A*-FFF-GUS) yielded a similar *in vitro* translation profile. Again, both the full-length and GFP-2A* products were identified with no FFF-GUS detectable. All these studies indicate that the –FFF- encoding sequence is capable of blocking translation. Stalling of the translation apparatus at this site combined with the method of analysis would lead to the detection of the upstream product of the polyprotein examined or where an active 2A is present, the increased imbalance in processing products. The role of the –FFF- encoding sequence in the translation of the viral genome is unclear at this stage. However, the location of the –FFF- encoding sequence immediately downstream of the 2A/2B processing junction and C-terminal to the structural proteins suggest it may be involved in producing of the vast excess of capsid proteins required for encapsidation, while conserving resources by blocking translation of the replicative proteins.

4.2.9. Yeast Studies Support Translational Model

Using the yeast artificial polyprotein constructs described previously, it has been possible to demonstrate *in vitro* that the FMDV 2A protein is active in yet another protein context. The application of these constructs *in vivo* has for the first time shown the FMDV 2A protein to possess processing activity in yeast (de Felipe, Hughes, Ryan & Brown, accepted for publication in J. Biol. Chem.). This has demonstrated that the features of higher eukaryotic cells that allow 2A to promote scission of the polypeptide chain at its own C-terminus are conserved in this 'simpler' organism. This finding opens up a whole new 'toolbox' for use in the examination of 2A function. Hopefully, in the future, mutational analysis will be able to be performed in yeast by phenotypic screening allowing a shift away from the laborious method of *in vitro* analysis.

The exploitation of two pathways that exist in yeast for the targeting of proteins to the ER lumen has provided us with useful insight as to where the 2A-mediated protein scission occurs. One pathway is reliant on chaperones to target proteins that have been synthesised in the cytoplasm and is termed the SRP-independent pathway. The polyprotein PpαF-2A-GFPh was designed to examine the function of 2A in this pathway. Based on the model for 2A activity it was predicted that during the synthesis of this protein, scission would occur releasing two discrete proteins into the cytoplasm. The first translation product, ppαF-2A, carrying the signal sequence for this pathway would be targeted to the ER lumen and GFP would remain in the cytoplasm. This was indeed the case. The second targeting pathway is the SRP-dependent pathway and as the name suggests is dependent on the signal recognition particle (SRP). In this pathway the signal sequence at the N-terminus of the nascent peptide is recognised and bound by SRP as it emerges from the ribosome. This event triggers an arrest in translation until the whole complex is targeted to the ER membrane and SRP interacts with the SRP receptor. Release of the SRP and docking of the ribosome complex onto transmembrane receptor proteins allows the co-translational targeting of the protein into the lumen (*Figure 47*).

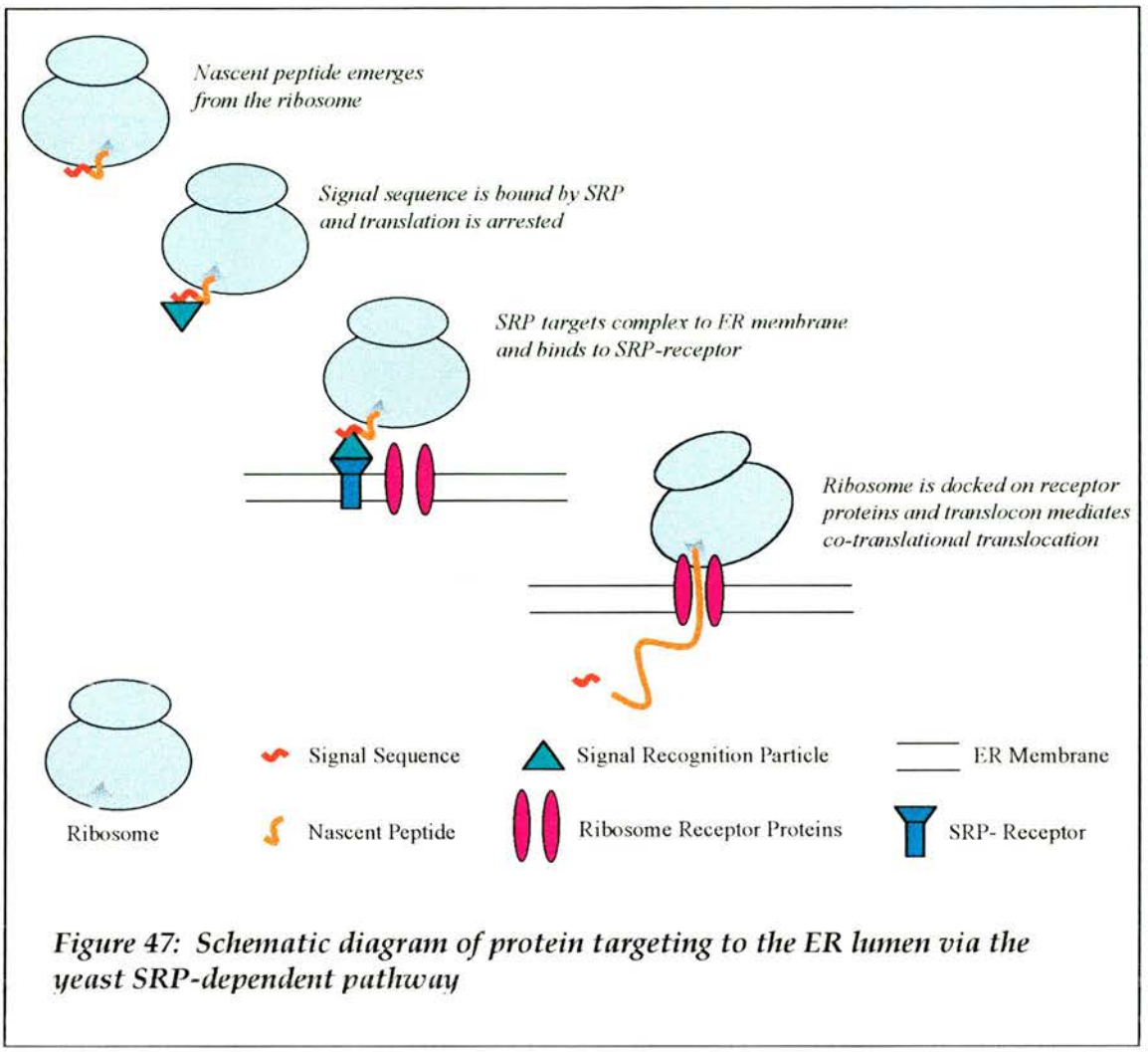


Figure 47: Schematic diagram of protein targeting to the ER lumen via the yeast SRP-dependent pathway

The $D_N\alpha$F-2A-GFPh polyprotein was designed to examine the function of 2A in the SRP-dependent pathway. The first translation product ($D_N\alpha$F-2A), carrying the signal sequence, was targeted to the lumen while GFP protein was found in the cytoplasm. The activity of the 2A peptide in this context, with translation tightly coupled to translocation, demonstrated that processing occurs in the ribosome with no requirement for extra-ribosomal cytosolic factors. The fact that the GFP protein is not found in the lumen suggests that the translocon recognises the 2A-mediated interruption in the peptide chain and prevents translocation of the product C-terminal to 2A (*Figure 48*). The study of 2A

activity in yeast supports the translational model for 2A-mediated processing, where the break in the nascent chain is proposed to occur within the ribosome.

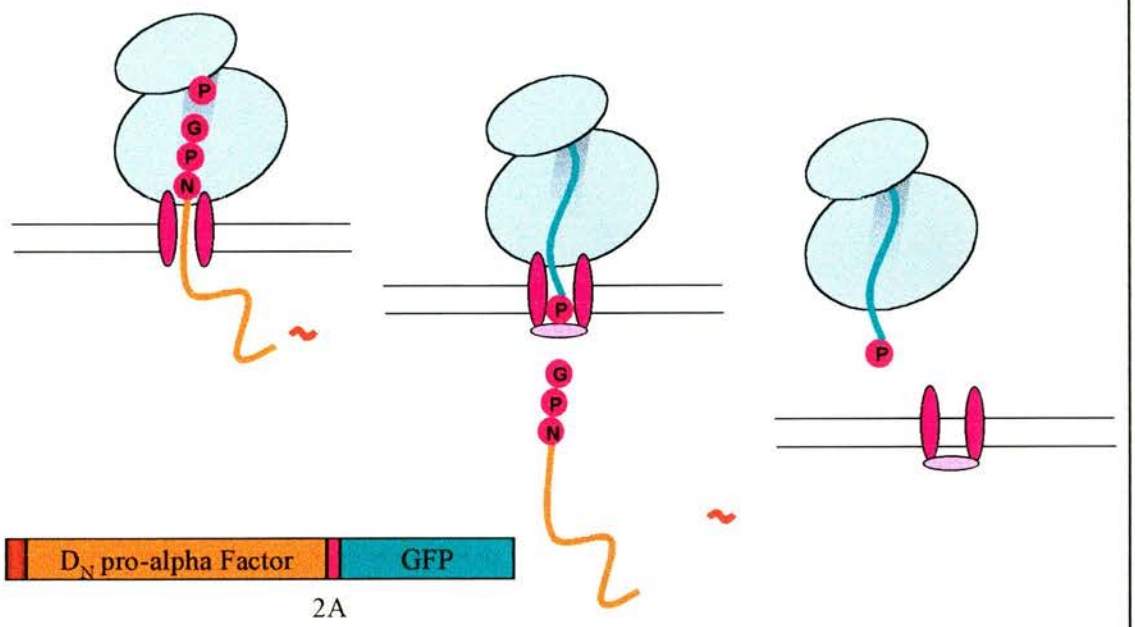


Figure 48: Graphic representation of the co-translational translocation of the construct pDNαF-2A-GFPh. *Subsequent to the targeting of the ribosome complex to the ER membrane by the SRP, translation of the nascent peptide proceeds coupled to translocation. The N-terminal protein carrying the signal sequence is targeted into the ER lumen. FMDV 2A mediates a translational effect within the ribosome. The 2A-mediated interruption in the peptide chain is detected by the translocon and the C-terminal protein, devoid of a signal sequence, is excluded from the lumen and instead released into the cytoplasm.*

The conclusions that have been made from the work presented here are:

- FMDV 2A is active in several artificial polyprotein context (e.g. GFP-2A-GUS, YFP-2A-GAL-CFP-2A-PAC and $D_N\alpha F$-2A-GFPh)
- FMDV 2A is active in wheat germ extracts, mammalian cells and yeast cells.
- The 20 amino acid long FMDV 2A sequence examined in the GFP-2A-GUS system is not optimal and full-length polyprotein is detected in *in vitro* reactions as a result of the sub-optimal functioning of 2A.
- The GFP-2A-GUS artificial polyprotein system can be used for screening 2A-like sequences for processing activity.
- Active 2A-like sequences are present in other viruses and in trypanosomal sequences.
- The 2A peptide does not mediate proteolysis but rather a translational effect that is characterised by the unequal synthesis TP1 and TP2.
- There are two parts to active 2A/2A-like sequences. Part A, is the conserved –DxExNPGP- motif. Part B, is the highly variable upstream sequence that forms interactions with the exit tunnel of the ribosome.
- The yeast data has provided important proof that the translational effect mediated by FMDV 2A occurs within the ribosome.

4.4 Summary

The study of FMDV 2A in both the *in vitro* reactions and a relatively simple organism such as yeast has provided us with real insights into the mode of action. The artificial polyprotein constructs possessing the 20 amino acid 2A and 2A-like sequences have highlighted the inefficiency of the 2A-mediated translational effect when the length of the viral sequences are too short. It is entirely possible that if the length of all the 2A/2Alike sequences were increased, in a similar manner to building in sequences of 1D at the N-terminus of FMDV 2A, the efficiency of processing would increase. The GFP-2A-GUS artificial polyprotein system has been successfully utilised as a simple, relatively rapid, screening method to determine active and inactive 2A-like sequences. Most of the viral sequences tested here have been shown to possess processing activity similar to FMDV 2A and provide support for the translational model of 2A activity. The analysis performed on these and other 2A and 2A-like sequence has provided useful information on the processing abilities of the proteins that may prove useful when determining which sequence to use for a particular biotechnological function. There are currently over 40 publications detailing successful applications of the 2A region of FMDV to various biotechnological roles. Analyses of the 2A and 2A-like sequences have allowed predictions to be made as to the control of biogenesis of the protein from which the sequences were identified. Understanding the mechanism of this processing event may assist in the control of FMDV and indeed the other viruses that use this strategy to generate individual functional proteins from a single open reading frame. The importance of which cannot be underestimated by anyone who witnessed the most recent, devastating, outbreak of foot-and-mouth disease virus in the United Kingdom.

5. REFERENCES

Ahmadian, G., Randhawa, J. S., & Easton, A. J. (2000). Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. EMBO J. *19*, 2681-2689.

Arnold, E., Luo, M., Vriend, G., Rossman, M.G., Palmenberg, A.C., Parks, G.D., Nicklin, M. J. H., & Wimmer, E. (1987). Implications of the picornavirus capsid structure for polyprotein processing. Proc. Natl. Acad. Sci. USA. *84*, 21-25.

Baltimore, D. (1971). Expression of animal virus genomes. Bacteriol. Rev. *35*, 235-241.

Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V.C., Brown, T. D. K., & Inglis, S. C. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. *6*, 3779-3785.

Brierley, I., Digard, P., & Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell *57*, 537-547.

Brierley, I., Jenner, A. J., & Inglis, S. C. (1992). Mutational analysis of the "slippery sequence" component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. *227*, 463-479.

Buchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., & Dom, R. (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in *Drosophila*. Genetics *155*, 141-157.

Clarke, B. E., Sangar, D. V., Burroughs, J. N., Newton, S. E., Carrol, A. R., & Rowlands, D. J. (1985). Two initiation sites for foot-and-mouth disease virus polyprotein *in vivo*. J. Gen. Virol. *66*, 2615-2626.

Crick, F. H. C., & Watson, J. D. (1956). Structure of small viruses. Nature *177*, 473-475.

Curran, J., Latorre, P., & Kolakofsky, D. (1998). Translational gymnastics on the Sendi virus P/C mRNA. Sem. Virol. *8*, 351-357.

De Felipe, P., Martin, V., Cortes, M. L., Ryan, M., & Izquierdo, M. (1999). Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. Gene Therapy, *6*, 198-208.

Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E., & Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. J. Virol. *62*, 4407-4409.

De Vries, A. A. F., Horzinek, M. C., Rottier, P. J. M., & de Groot, R. J. (1997). The genome organization of the Nidovirales: Similarities and differences between arteri-, toro-, and coronaviruses. Sem. Virol. *8*, 33-47.

Ding, M., & Schlesinger, M. J. (1989). Evidence that Sindbis virus nsP2 is an autoprotease which processes the virus non-structural polyprotein. Virol. *171*, 280-284.

Doherty, M., Todd, D., McFerran, N., & Hoey, E. M. (1999). Sequence analysis of a porcine enterovirus serotype 1 isolate: relationships with other picornaviruses. J. Gen. Virol. *80*, 1929-1941.

Donnelly, M. L. L. (1997). A study of the 2A Region of Aphthoviruses., Ph.D. Thesis, Department of Chemistry, University of St. Andrews, St. Andrews.

Donnelly, M., Gani, D., Flint, M., Monaghan, S., & Ryan, M.D. (1997). The cleavage activity of aphtho- and cardiovirus 2A proteins. J. Gen. Virol. *78*, 13-21.

Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., & Ryan, M. D. (2001a). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J. Gen. Virol. *82*, 1013-1025.

Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., & Ryan, M. D. (2001b). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J. Gen. Virol. *82*, 1027-1041.

Farabaugh, P. J. (1996). Programmed ribosomal frameshifting. Microbiol. Rev. *60*, 103-134.

Feng, Y., Yuan, H., Rein, A., & Levin, J. G. (1992). Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight nulceotide purine-rich sequence immediately downstream of the *gag* termination codon followed by an RNA pseudoknot. J. Virol. *66*, 5127-5132.

Giorgi, C., Blumberg, B. M., & Kolakofsky, D. (1983). Sendi virus contains overlapping genes expressed from a single mRNA. Cell *35*, 829-836.

Govan, V. A., Leat, N., Aiisopp, M., & Davison, S. (2000). Analysis of the complete sequence of acute bee paralysis virus shows that it belongs to the novel group of insect-infecting RNA viruses. Virol. *277*, 457-463.

Grubman, M. J., & Baxt, B. (1982). Translation of foot-and-mouth disease virion RNA and processing of the primary cleavage products in rabbit reticulocyt lysate. Virol. *116*, 19-30.

Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W., & Piccone, M. E. (1995). Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus. Virol. *213*, 581-589.

Gu, Z., Harrod, R., Rogers, E. J., & Lovett, P. S. (1994). Anti-peptidyl transferase leader peptides of attenuation-regulated chloramphenicol-resistance genes. Proc. Natl. Acad. Sci. USA. *91*, 5612-5616.

Hahn, H., & Palmenberg, A. C. (1996). Mutational analysis of the encephalomyocarditis virus primary cleavage. J. Virol. *70*, 6870-6875.

Hahn, H., & Palmenberg, A. C. (2001). Deletion mapping of the encephalomyocarditis virus primary cleavage site. J. Virol. *75*, 7215-7218.

Halpin, C., Cooke, S. E., Barakate, A., Amrani, A. E., & Ryan, M. D. (1999). Self-processing 2A-polyproteins – a system for co-ordinate expression of multiple proteins in transgenic plants. The Plant J. *17*, 453-459.

Hardy, W. R., & Strauss, J. H. (1989). Processing of the nonstructural polyproteins of Sindbis virus: Nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. J. Virol. *63*, 4653-4664.

Harrod, R., & Lovett, P. S. (1995). Peptide inhibitors of peptidyltransferase alter the conformation of domains IV and V of large subunit rRNA: A model for nascent paptide control of translation. Proc. Natl. Acad. Sci. USA. *92*, 8650-8654.

Helbing, C. C. & Atkinson, B. G. (1994). 3,5,3'-Triiodothyronine-induced carbamyl-phosohate synthetase gene expression is stabilized in the liver of *Rana catesbiana* tadpoles during heat shock. J. Biol. Chem. *269*, 11743-11750.

Horvath, C. M., Williams, M.A. & Lamb, R.A. (1990). Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. EMBO J. *9*, 2639-2647.

Isawa, H., Asano, S., Sahara, K., Iizuka, T., & Bando, H. (1998). Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses. Arch. Virol. *143*, 127-143.

James, V. L. A., Lambden, P. R., Deng, Y., Caul, E. O., & Clarke, I. N. (1999). Molecular characterization of human group C rotavirus genes 6, 7 and 9. J. Gen. Virol. *80*, 3181-3187.

Johnson, K. N., & Christian, P. D. (1998). The novel genome organization of the insect picorna-like virus *Drosophila* C virus suggest this virus belongs to a previously undescribed virus family. J. Gen. Virol. *79*, 191-203.

Kozak, M. (1986). Regulation of protein synthesis in virus infected animal cells. Adv. Virol. Biol. *31*, 229-292.

Li, G. & Rice, C. M. (1993). The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. J. Virol. *67*, 5062-5067.

Lustig, A., & Levine, A. J. (1992). One hundred years of virology. J. Virol. *66*, 4629-4631.

Martin, F., Maranon, C., Olivares, M., Alonso, C., & Lopez, M. C. (1995). Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from Trypanosoma cruzi: Homology of the first ORF with the Ape family of DNA repair enzymes. J. Mol. Biol. *247*, 49-59.

Mattion, N. M., Harnish, E. C., Crowley, J. C., & Reilly, P. A. (1996). Foot-and-mouth disease virus 2A protease mediates cleavage in attenuated Sabin 3 poliovirus vectors engineered for delivery of foreign antigens. J. Virol. *70*, 8124-8127.

Medina, M., Domingo, E., Brangwyn, J. K., & Belsham, G. J. (1993). The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. Virol. 194, 355-359.

Murphy, F.A. (1996). Virus taxonomy. In *Fields Virology*, 3rd edn, pp. 15-57. Edited by B. N. Fields, D. M. Knipe, & P. M. Howley. Philadelphia: Lippincott-Raven.

Murphy, N. B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M., & Pays, E. (1987). Trypanosoma brucei repeated element with unusual structural and transcriptional properties. J. Mol. Biol. *195*, 855-871.

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science. *289*, 920-930.

Palmenberg, A. C. (1990). Proteolytic processing of picornaviral polyprotein. Annu. Rev. Microbiol. *44*, 603-623.

Pan, Y-X., Xu, J., Bolan, E., Chang, A., Mahurter, L., Rossi, G., & Pasternak, G. W. (2000). Isolation and expression of a novel alternatively spliced mu opioid receptor isoform, MOR-1F. FEBS Lett. *466*, 337-340.

Pringle, F. M., Gordon, K. H. J., Hanzlik, T. N., Kalmakoff, J., Scotti, P. D., & Ward, V. K. (1999). A novel capsid expression strategy for *Thosea asigna* virus (*Tetraviridae*). J. Gen. Virol. *80*, 1855-1863.

Pringle, F. M., Kalmakoff, J., & Ward, V. K. (2001). Analysis of the capsid processing strategy of *Thosea asigna* virus using baculovirus expression of virus-like particles. J. Gen. Virol. *82*, 259-266.

Qian, Y., Jiang, B., Saif, L. J., Kang, S. Y., Ojeh, C. K., & Green, K. Y. (1991). Molecular analysis of the gene 6 from a porcine group C rotavirus that encodes the NS34 equivalent of group A rotaviruses. Virol. *184*, 752-757.

Reuckert, R. R. (1996). *Picornaviridae:* The viruses and their replication. In *Fields Virology*, 3rd edn, pp. 15-57. Edited by B. N. Fields, D. M. Knipe, & P. M. Howley. Philadelphia: Lippincott-Raven.

Rogers, E. J., & Lovett, P. S. (1994). The *cis*-effect of a nascent peptide on its translating ribosome: influence of the *cat-86* leader pentapeptide on translation termination at leader codon 6. Mol. Microbiol. *12*, 181-186.

Ruile, P., Winterhalter, C., & Liebl, W. (1997). Isolation and analysis of a gene encoding α-glucuronidase, an enzyme with a novel primary structure involved in the breakdown of xylan. Mol. Microbiol. *23*, 267-279.

Ryan, M. D., Donnelly, M., Lewis, A., Mehrotra, A. P., Wilkie, J., & Gani, D. (1999). A model for non-stoichiometric, cotranslational protein scission in eukaryotic ribosomes. Bioorganic Chem. *27*, 55-79.

Ryan, M.D. & Drew, J. (1994). Foot-and-Mouth Disease Virus 2A oligopeptide mediated cleavage of an artificial polyprotein. EMBO J. *13*, 928-933.

Ryan, M. D., & Flint, M. (1997). Virus encoded proteinases of the picornavirus super-group. J. Gen. Virol. *78*, 699-723.

Ryan, M.D., King, A.M.Q., & Thomas, G.P. (1991). Cleavage by foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J. Gen. Virol.* 72, 2727-2732.

Sangar, D. V., Newton, S. E., Rowlands, D. J., & Clarke, B. E. (1987). All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs. Nucleic Acids Res. *15*, 3305-3315

Schlesinger, S., & Schlesinger, M. J. (1996). *Togaviridae:* The viruses and their replication. In *Fields Virology*, 3rd edn, pp. 15-57. Edited by B. N. Fields, D. M. Knipe, & P. M. Howley. Philadelphia: Lippincott-Raven.

Snijder, E.J., & Meulenberg, J. J. M. (1998). The molecular biology of arteriviruses. J. Gen. Virol. *79*, 961-979.

Somogyi, P., Jenner, A. J., Brierley, I., & Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA Pseudoknot. Mol. Cell. Biol. *13*, 6931-6940.

Strauss, E. G., Rice, C. M., & Strauss, J. H. (1983). Sequence coding for the alphavrus nonstructural proteins is interrupted by an opal termination codon. Proc. Natl. Acad. Sci. USA. *80*, 5271-5295.

Strauss, J. H., & Strauss, E. G. (1988). Evolution of RNA viruses. Ann. Rev. Microbiol. *42*, 657-683.

Strebel, K., & Beck, E. (1986). A second protease of foot-and-mouth disease virus. J. Virol. *58*, 893-899.

Suzuki, N., Geletka, L. M., & Nuss, D. L. (2000). Essential and dispensable virus-encoded replication elements revealed by efforts to develop hypoviruses as gene expression vectors. J. Virol. *74*, 7568-7577.

Svitkin Y. V., & Agol, V. I. (1983). Translational barrier in central region of encephalomyocarditis virus genome. Modulation by elongation factor 2 (eEF-2). Eur. J. Biochem. *133*, 145-154.

ten Dam, E., Flint, M., & Ryan, M. D. (1999). Virus-encoded proteinases of the *Togaviridae*. J. Gen. Virol. *80*, 1879-1888.

Vakharia, V. N., Devaney, M. A., Moore, D. M., Dunn, J. J., & Grubman, M. J. (1987). Proteolytic processing of foot-and-mouth disease virus polyproteins expressed in a cell-free system from clone-derived transcripts. J. Virol. *61*, 3199-3207.

van der Wilk, F., Dullemans, A. M., Verbeek, M., & Van den Heuvel, J. F. J. M. (1997). Nucleotide sequence and genomic organization of Acyrthosiphon Pisum Virus. Virol. *238*, 353-362.

Varnavski, A. N., Young, P. R., & Khromykh, A. A. (2000). Stable high-level expression of heterologous genes in vitro and in vivo by noncytopathic DNA-based Kunjin virus replicon vectors. J. Virol. *74*, 4394-4403.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. (1988). Molecular biology of the gene, 4th Ed., pp767-771. The Benjamin/Cummings Publishing Company, Inc., California.

Weiss, R. B., Huang, W. M., & Dunn, D. M. (1990). A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene *60*. Cell. *62*. 117-126.

Williams, M.A. & Lamb, R.A. (1989). Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins. J. Virol. *63*, 28-35.

Wilson, J. E., Powell, M. J., Hoover, S. E., & Sarnow, P. (2000a). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. Mol. Cell. Biol. *20*, 4990-4999.

Wilson, J. E., Pestova, T. V., Hellen, C. U. T., & Sarnow, P. (2000b). Initiation of protein synthesis from the A site of the ribosome. Cell. *120*, 511-520.

Wolfe, S. L. (1993). Molecular and cellular biology, pp. 816-817. Wadsworth Publishing Company, California.

Wu, C. Y., Lo, C. F., Huang, C. J., Yu, H. T., & Wang, C. H. (2002). The complete genome sequence of *Perina nuda* picorna-like virus, an insect-infecting RNA virus with a genome organization similar to that of the mammalian picornaviruses. Virol. *294*, 312-323.

Wutz, G., Auer, H., Nowotny, N., Grosse, B., Skern, T., & Kuechler, E. (1996). Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses. J. Gen. Virol. *77*, 1719-1730.

Yoshinaka, Y., Katoh, I., Copeland, T. D., & Stephen, O. (1985). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA. *82*, 1618-1622.

Zell, R., Dauber, M., Krumbholz, A., Henke, A., Birch-Hirschfeld, E., Stelzner, A., Prager, D., & Wurm, R. (2001). Porcine teschoviruses comprise at least eleven distinct serotypes: molecular and evolutionary aspects. J. Virol. *75*, 1620-1631.